



IDECOR

Ministerio de
FINANZAS



Ministerio de
**AGRICULTURA
Y GANADERÍA**



INFORME

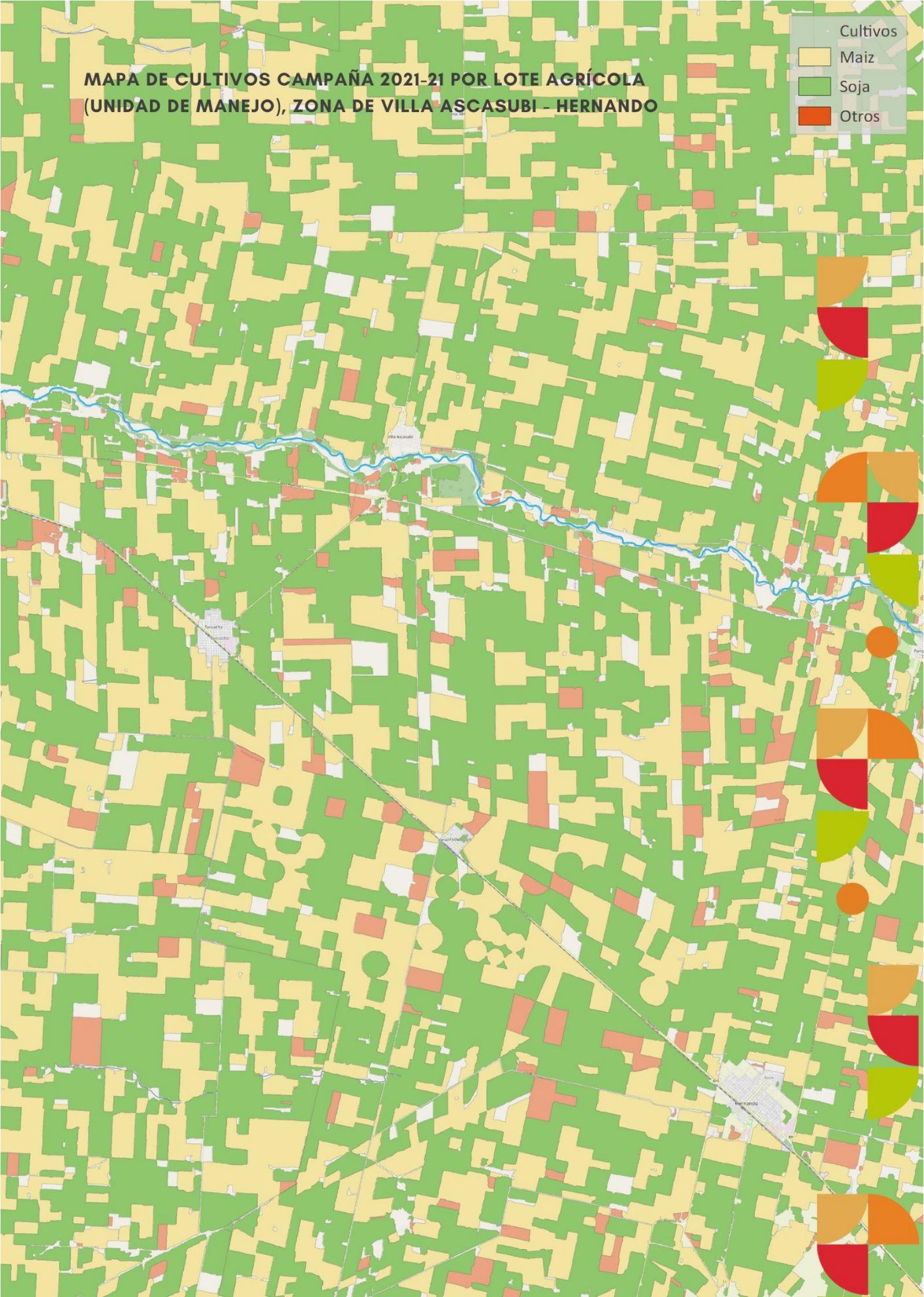
ÁREA SEMBRADA, RINDES Y PRODUCCIÓN DE SOJA Y MAÍZ

CAMPAÑA 2021/2022
PROVINCIA DE CÓRDOBA



MAPA DE CULTIVOS CAMPAÑA 2021-21 POR LOTE AGRÍCOLA (UNIDAD DE MANEJO), ZONA DE VILLA ASCASUBI - HERNANDO

- Cultivos
- Maiz
- Soja
- Otros





Contenido

Resumen Ejecutivo	4
1. Introducción	6
2. Relevamiento de datos en campo	7
3. Mapeo de cultivos estivales	11
3.1 Identificación automática de lotes agrícolas.....	11
3.2 Clasificación supervisada de cultivos estivales.....	13
3.2.1 Construcción de mosaicos	13
3.2.2 Selección de la muestra	13
3.2.3 Resultados obtenidos.....	15
4. Estimación de rendimientos por cultivo	18
4.1 Muestra utilizada.....	18
4.2 Modelos de aprendizaje computacional aplicados.....	19
4.3 Resultados obtenidos	21
5. Estimación de la producción de soja y maíz	24
Referencias Bibliográficas	28
Anexo I. Variables utilizadas en el modelo de predicción de rendimientos .	29
Anexo II - Modelos de Aprendizaje Computacional utilizados en las estimaciones	36
Créditos. Equipo de trabajo	38



Organismos participantes

Desarrollo:

- IDECOR (Infraestructura de Datos Espaciales de la Provincia de Córdoba)
Ministerio de Finanzas de la Provincia de Córdoba.
- Ministerio de Agricultura y Ganadería de la Provincia de Córdoba.

Versión y Licencia.

- MAPA DE ÁREA SEMBRADA, RINDES Y PRODUCCIÓN DE SOJA Y MAÍZ, CAMPAÑA 2021/2022, PROVINCIA DE CÓRDOBA. Está distribuido bajo una Licencia Creative Commons Atribución 4.0 Internacional. Libre para compartir, distribuir, copiar y adaptar.



- Citar como: Mapa de Área Sembrada, Rindes y Producción de Soja y Maíz Campaña 2021/2022, Provincia de Córdoba. Ministerio de Agricultura y Ganadería e IDECOR, Gobierno de la Provincia de Córdoba, septiembre 2022.

Aviso de uso

Los datos del presente trabajo fueron desarrollados por el Ministerio de Agricultura y Ganadería de la Provincia de Córdoba e IDECOR para dar soporte a la gestión de políticas públicas del sector agropecuario. En su utilización, deben tomarse las consideraciones técnicas que se derivan del presente informe.

El equipo de trabajo se desliga de cualquier uso indebido que pueda realizarse de los mapas y datos desarrollados fuera del ámbito para el cual fueron diseñados y/o sus características técnicas posibilitan, quedando estos otros usos bajo responsabilidad de los usuarios.

Resumen Ejecutivo

El presente informe detalla los resultados obtenidos y la metodología aplicada en el mapeo del área sembrada, la estimación de rendimientos y el cálculo de la producción agrícola de la Provincia de Córdoba para la campaña 2021-2022.

Los resultados indican que **7,83 millones de hectáreas fueron destinadas a la siembra** de cultivos estivales, representando el 47,5% de la superficie provincial. Del total de la tierra cultivada, **un 59,4% corresponde a soja, un 33,3% a maíz** y el 7,3% restante a otros cultivos.

Para la estimación de los rendimientos se utilizó una muestra de 2.175 datos de campo y 160 variables geográficas (estado de la vegetación, clima, etc.). Para la campaña referida, **el rendimiento promedio de soja a nivel provincial fue 31 quintales por hectárea (qq/ha)**, mientras que **para el maíz este valor se ubicó en 77 qq/ha**. El nivel de detalle del estudio permite observar la variabilidad de los rendimientos de cada cultivo sobre una grilla regular de **25 hectáreas de unidad mínima**.

La principal innovación de esta edición consiste en la estimación de las áreas cultivadas mediante un proceso de clasificación orientada a objetos, sobre la base de una metodología semiautomática de identificación de **lotes de siembra**, a partir de imágenes satelitales. Esta mejora permite una definición más precisa del área sembrada y, junto con la utilización de **técnicas de aprendizaje computacional** para la clasificación de cultivos y predicción de rendimientos, implica el uso de tecnologías y metodologías de última generación para este tipo de estudio.

Dada la escala de trabajo y la estructura del sector agrícola de la provincia, el análisis está centrado en el cultivo de soja y maíz, que en conjunto representan más del 90% de la actividad sectorial. El resto de los cultivos fueron contemplados en el mapeo de las áreas sembradas, pero no en la predicción de rendimientos ni producción.

En el cultivo de soja, los rendimientos promedios más altos se registraron en los Departamentos Marcos Juárez, Sáenz Peña y General Roca, con valores por encima de los 35 qq/ha. En cuanto al maíz, los Departamentos de mayor rendimiento promedio fueron también Marcos Juárez y Sáenz Peña, a los que se agregan Unión y Colón, todos con rendimientos promedio por encima de los 80 qq/ha.

A partir de la estimación de rendimientos y el área sembrada, se calcula la producción respectiva. En relación a la **soja**, se estimó una producción de **14,6 millones de toneladas (Tn)**; en tanto que la producción total de **maíz** se estimó en **20,2 millones de Tn**.

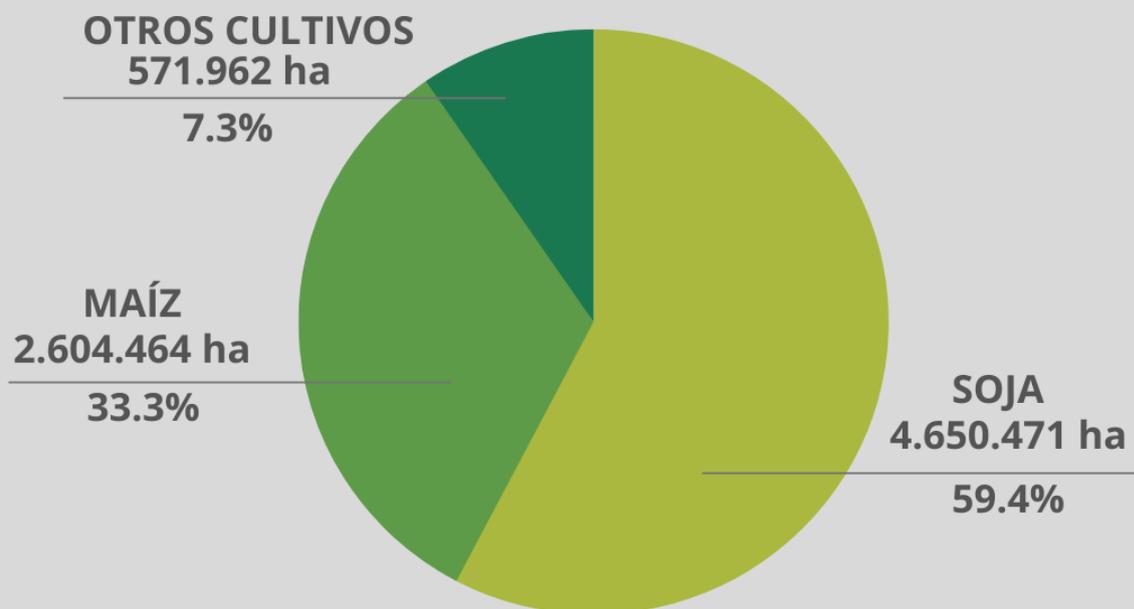


ÁREA SEMBRADA, RINDES Y PRODUCCIÓN DE SOJA Y MAÍZ

CAMPAÑA 2021/2022 PROVINCIA DE CÓRDOBA

TOTAL DE SUPERFICIE SEMBRADA

7.826.897 ha



77 QQ/HA

Rendimiento promedio de MAÍZ

31 QQ/HA

Rendimiento promedio de SOJA

20.2 MM TN

Producción de MAÍZ

14.6 MM TN

Producción de SOJA

1. Introducción

El sector agrícola de la Provincia de Córdoba ocupa un lugar preponderante en la economía local. Tomado en conjunto con la actividad ganadera, **el sector participa del 18% del producto geográfico bruto (PGB) de la provincia y del 24% de la producción agropecuaria nacional¹**. La Secretaría de Agricultura de la Provincia de Córdoba en promedio para las últimas 5 campañas gruesas, registra que el área sembrada de soja resultó de aprox. 4,6 millones de ha, con un rendimiento de 31 qq/ha y una producción de 13,8 millones de Tn. En el caso del maíz, la superficie sembrada resultó en las últimas 5 campañas, en un promedio de 2,3 millones de ha, con un rendimiento de 79 qq/ha y una producción de 16,3 millones de Tn².

En función al promedio de las últimas 5 campañas, el área sembrada en la provincia para los cultivos de soja y maíz, fue de aproximadamente 7 millones de ha, con una distribución de 32% para el maíz y 68% para la soja, lo que supone ingresos por aproximadamente 11.000 millones de dólares.

Con el fin de generar información de mayor calidad para el sector, el Ministerio de Agricultura y Ganadería de la Provincia de Córdoba, con apoyo de IDECOR, implementó a partir de la campaña 2020-2021, un programa de estimaciones agrícola basado en el uso de modernas tecnologías y enfoques metodológicos.

Esquemáticamente, el proceso se puede describir en 3 grandes etapas:

- **Mapeo del área sembrada y tipo de cultivos.** Esta tarea requiere previamente, la identificación automática de lotes de siembra, entendidos como unidad mínima de manejo agrícola y asumiendo que, dentro de cada uno, se realiza sólo un tipo de cultivo en el período analizado. Luego, a partir de una muestra de 5.084 datos de campo (tipo de cultivo sembrado en la campaña), se realizó la clasificación automática sobre imágenes satelitales, identificando el tipo de cultivo en cada lote agrícola de la provincia.
- **Predicción de los rendimientos para soja y maíz.** Mediante el ajuste de algoritmos de aprendizaje computacional, a partir de una muestra de 2.175³ datos de campo (rindes obtenidos) y 160 variables territoriales, se realiza la estimación de rindes sobre una grilla regular de 25 ha (431.835 celdas) que cubre todo el territorio cultivado.
- **Estimación de la producción.** Sobre la grilla de 25 ha, considerando las superficies sembradas y los rindes estimados para cada cultivo, se calcula

¹ Consultar en: <https://datosestadistica.cba.gov.ar/organization/0fc05944-9507-40e7-a997-b46d40cd00b0?groups=sectores-economicos>

² Consultar en: <https://agricultura.cba.gov.ar/institucional/consulta-indicadores.php>

³ Se trata de una muestra diferente a la señalada en el párrafo anterior, dado que se aplica a la resolución de un problema distinto.

la producción en cada una de las unidades de aquella, agregándose luego a nivel departamental y total provincial.

La continuidad de este programa, ya en su segunda edición⁴, fortalece la generación de información pública de calidad y a un nivel de detalle novedoso, ya que las estimaciones preexistentes estimaban la información a nivel de departamento o provincial. Este salto de calidad en la información, apoyado en la utilización de modernas técnicas y metodológicas de última generación, y en el principio de transparencia y datos abiertos, no sólo se constituye en un insumo de suma relevancia para el diseño de políticas sectoriales, sino que genera externalidades a muchas otras cadenas productivas enlazadas. Esta dinámica enriquece el proceso de toma de decisiones en el ámbito privado y fortalece la disponibilidad de información para una gestión de políticas públicas basadas en datos, orientada a brindar bienes y servicios públicos de mayor calidad para los ciudadanos. Sólo a modo de ejemplo, al conocer la distribución espacial de la producción se pueden identificar tramos de caminos rurales críticos para intensificar su mantenimiento o nuevas obras, el consumo de nutrientes del suelo que permitan calcular el balance de los mismos o identificar zonas de manejo de acuerdo a la productividad estimada a nivel predial, que permiten la implementación de estrategias de manejo diferenciales, entre otros aspectos relevantes.

2. Relevamiento de datos en campo

Las observaciones fueron recopiladas a través de la **aplicación “Rindes Collect” (Figura 1), desarrollada específicamente por IDECOR para el relevamiento georreferenciado de datos de rindes a campo.** Esta herramienta 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 en cada sitio relevado, 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.

⁴ El lector puede acceder al informe de la primera edición de este estudio, correspondiente a la campaña 2020/21, en el siguiente enlace: https://www.idecor.gob.ar/wp-content/uploads/2021/12/Informe-Cosecha-20-21_Gobierno-de-Cordoba.pdf

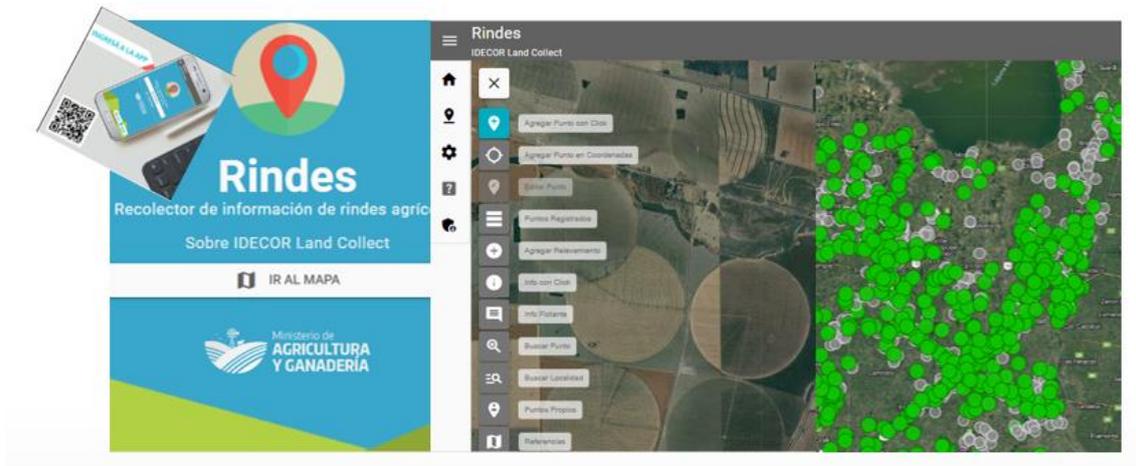


Fig. 1. Pantallas de la App Rindes.

En el relevamiento participaron los siguiente actores e instituciones:

- Secretaría de Agricultura de la Provincia de Córdoba
- Agencias Zonales del Ministerio de Agricultura y Ganadería de la Provincia de Córdoba.
- INTA Córdoba, Agencias de Extensión.
- Agentes privados y otras instituciones.

El relevamiento fue coordinado por IDECOR y el Ministerio de Agricultura y Ganadería y se obtuvieron **en total 2.247 observaciones de campo**. Luego de un proceso de depuración de observaciones espacialmente atípicas, 2.175 fueron utilizadas para la estimación de los rendimientos por cultivo y 1.968 se usaron en la estimación de la superficie sembrada de cada cultivo (además de otras muestras relevadas a dicho fin, como se detalla más adelante).

En la Figura 2 se presentan la distribución espacial de los datos recogidos y el nivel de participación de cada organismo.

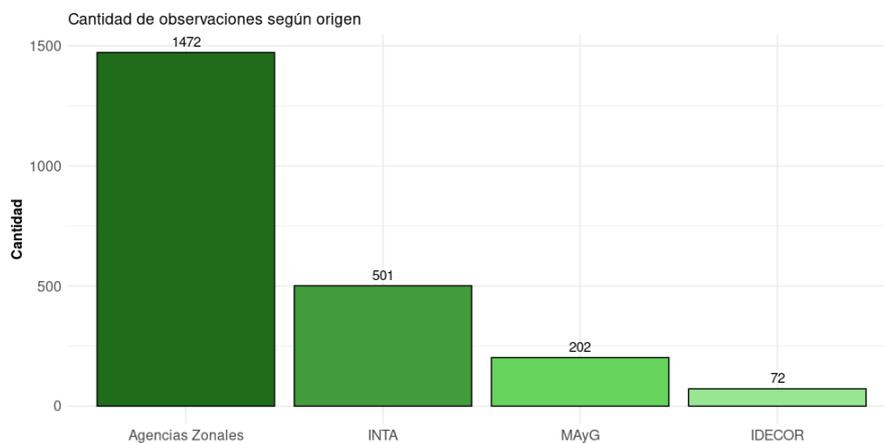
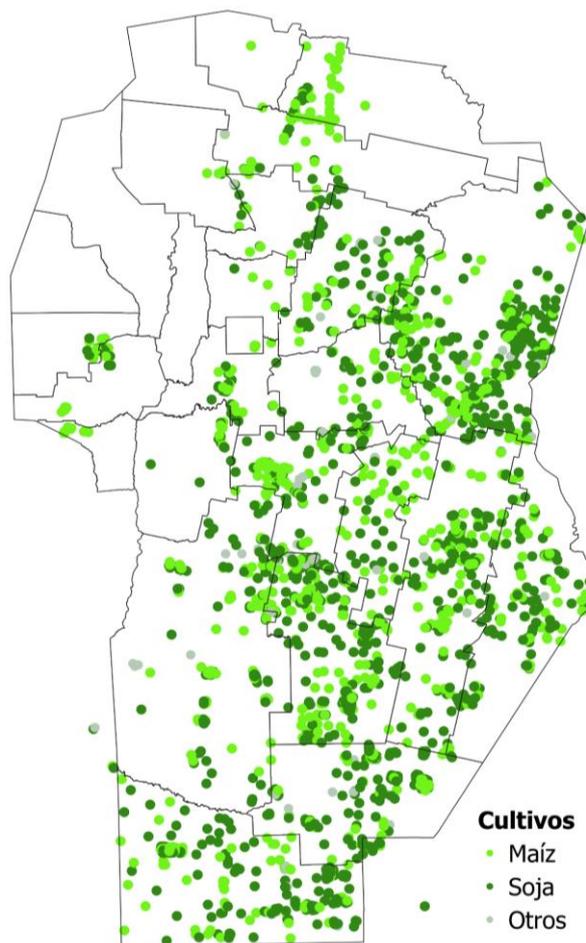


Figura 2. Distribución espacial de los datos relevados a campo (campana 2021-22) y nivel de participación de cada organismo:

Adicionalmente, en el relevamiento se informa 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 de utilidad para el tratamiento de la muestra a ser utilizada en las estimaciones, a los fines de evitar sesgos en la estimación del rendimiento, además de constituir en sí misma información de valor para identificar áreas con fecha de siembra similares y áreas afectadas por eventos adversos.

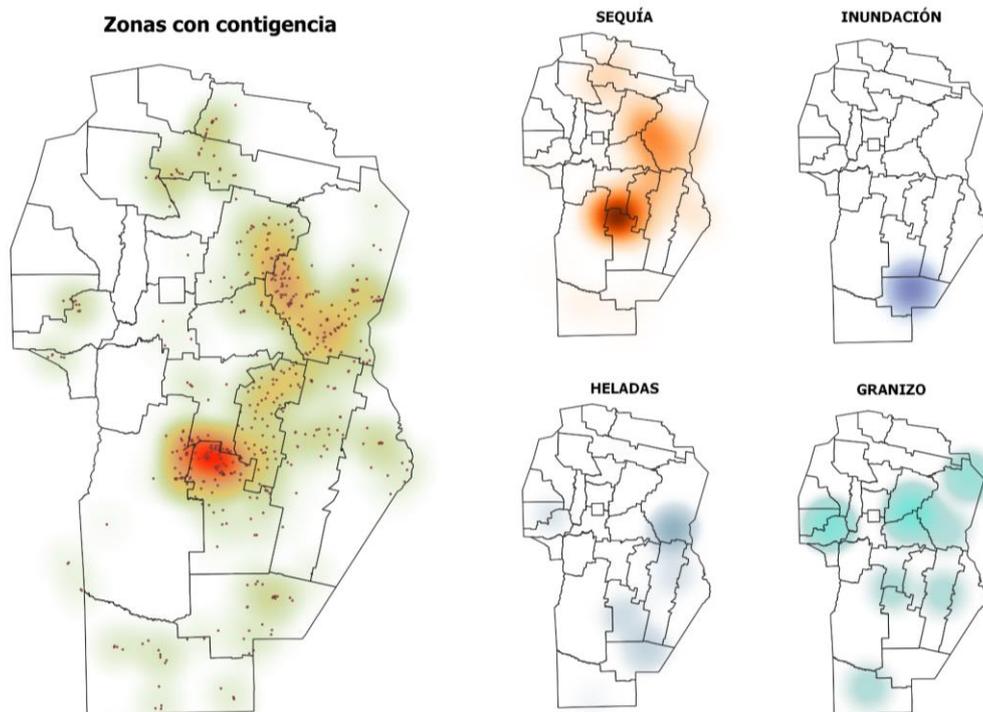
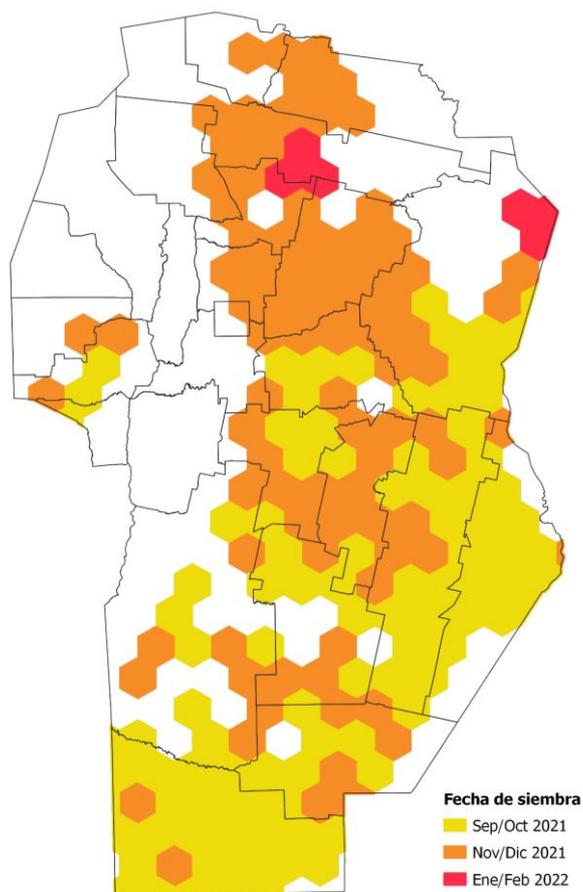


Figura 3. Mapas de distribución espacial de fechas de siembra y de eventos climáticos adversos que afectan la producción agrícola en la provincia de Córdoba, campaña 2021-22.

3. Mapeo de cultivos estivales

3.1 Identificación automática de lotes agrícolas

La determinación del área sembrada se realizó siguiendo una metodología de **clasificación de coberturas agrícolas orientada a objetos sobre imágenes satelitales** y se escogió como unidad de análisis el lote de uso agrícola.

Es importante destacar que **el lote agrícola constituye una unidad de manejo** definida por el productor para un manejo óptimo de su área productiva, por lo que en general, no coinciden divisiones parcelarias; por este motivo fue necesario realizar un proceso de segmentación buscando, a partir de imágenes satelitales y mediante el uso de algoritmos específicos, delimitar unidades espaciales con un comportamiento espectral homogéneo a lo largo de diferentes campañas agrícolas.

El área de trabajo se delimitó a partir del Mapa de Cobertura y Usos del suelo de la Provincia de Córdoba 2020-2021⁵, seleccionando las áreas correspondientes a coberturas agrícolas. A los fines de detectar comportamientos homogéneos, se trabajó con **imágenes Sentinel 2 corregidas** a reflectancia de superficie, con filtros para eliminar nubes y sombras, construyendo un mosaico de la zona de análisis cada 45 días, abarcando un período comprendido entre el 1 de enero de 2019 y el 31 de diciembre 2021. De este modo se construyeron **24 mosaicos de imágenes** sobre los que se calculó el Índice de Agua Diferencial Normalizado (NDWI, Figura 3.1) y, sobre este, 5 índices de textura (Figura 3.2): suma de varianzas, contraste, varianza, suma de promedios, disimilaridad y correlación (Haralick et. al, 1973).

A los fines de realizar un análisis multitemporal, se extrajeron los 6 índices de cada mosaico y se combinaron todos en un **único mosaico de 144 bandas**. Sobre este mosaico se aplicó el algoritmo de detección de bordes Canny (Canny, 1986), ajustando parámetros (o umbrales) de acuerdo al comportamiento de cada variable y se sumaron los resultados (Figura 3.3). De este modo se identificaron los bordes, mientras que los píxeles restantes se fueron agrupando, obteniendo áreas de comportamiento homogéneo (Figura 3.4).

⁵ Informe: <https://www.idecor.gob.ar/wp-content/uploads/2022/06/INFORME-Mapa-de-Coberturas-y-usos-del-Suelo-2020-21.pdf>. Mapa online: <https://mapascordoba.gob.ar/viewer/#/mapa/363>.

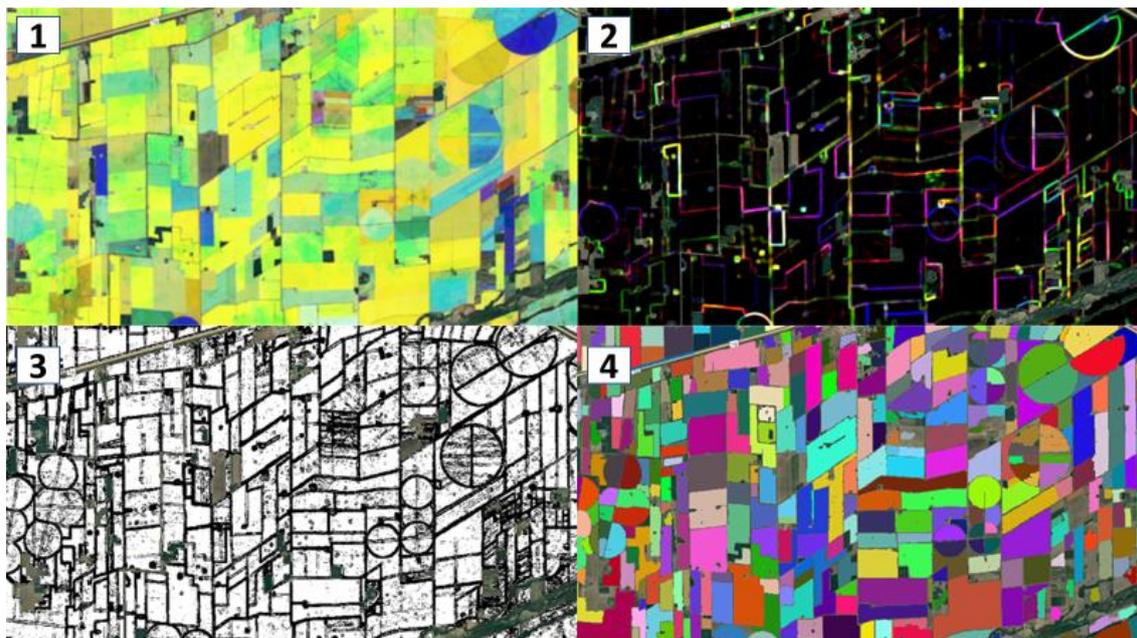


Figura 3. Imagen RGB del índice NDWI para los tres primeros períodos de análisis (1). Imagen RGB del índice varianza calculado sobre el mosaico anterior (2). Imagen binaria de bordes y áreas en negro y blanco respectivamente (3). Unidades espaciales identificadas por color (4).

Estas áreas de comportamiento homogéneo fueron vectorizadas masivamente mediante la utilización de técnicas geomáticas en QGIS (QGIS, 2021). Además, se utilizaron herramientas de análisis topológico para reducir el número de vértices y eliminar efectos de bordes. El resultado final consiste en **353.260 unidades espaciales o lotes agrícolas**⁶.



Figura 4. Vista de lotes vectorizados.

⁶ Es necesario aclarar que en lotes que excepcionalmente muestran alta variabilidad interna, se podrían generar subdivisiones menores a la unidad de manejo agrícola determinada.

3.2 Clasificación supervisada de cultivos estivales

3.2.1 Construcción de mosaicos

Se construyó una serie temporal de imágenes Sentinel 2 corregidas a reflectancia de superficie⁷, comprendiendo el **período del 1 de septiembre de 2021 al 3 de mayo de 2022**, el cual se subdividió a su vez en 10 períodos con la finalidad de incluir estadios fenológicos claves para diferenciar las especies a clasificar mediante las bandas espectrales e índices disponibles

Puesto que los períodos de tiempo utilizados pueden abarcar más de una imagen satelital para una misma región, a los fines de construir un mosaico que represente la respuesta espectral de cada especie y estadio fenológico para el período de estudio, se calculó la mediana de cada banda espectral por colección.

Sobre cada mosaico, además, se calcularon los índices espectrales EVI (Huete et al., 1994), NDRI (Gelder et al., 2009), BSI (Jamalabad & Abkar, 2004), NDVI (Rouse et al., 1973) y REP (Guyot et al., 1988).

Finalmente, dentro de cada unidad espacial, se calculó la mediana de la banda 6 (infrarrojo cercano) y de cada índice, incluyendo además el desvío estándar de los índices EVI y NDVI, haciendo un **total de 80 variables espectrales para incluir en el modelo de clasificación**. A su vez, se incluyó como factor la especie cultivada en la campaña estival 2020-2021. De esta manera, la información espectral, obtenida a nivel de píxel, es resumida a nivel de unidades espaciales o lotes agrícolas.

3.2.2 Selección de la muestra

Se analizó el comportamiento fenológico de los datos relevados a campo por especie y por zona a través de series temporales del índice NDVI, con el objetivo de contar con una muestra representativa de las diferentes clases para realizar la predicción, seleccionando los 1.968 datos mencionados anteriormente.

⁷ En la selección de imágenes se aplicó una máscara para nubes y sombras, seleccionando el valor de probabilidad con un umbral del 20 %.

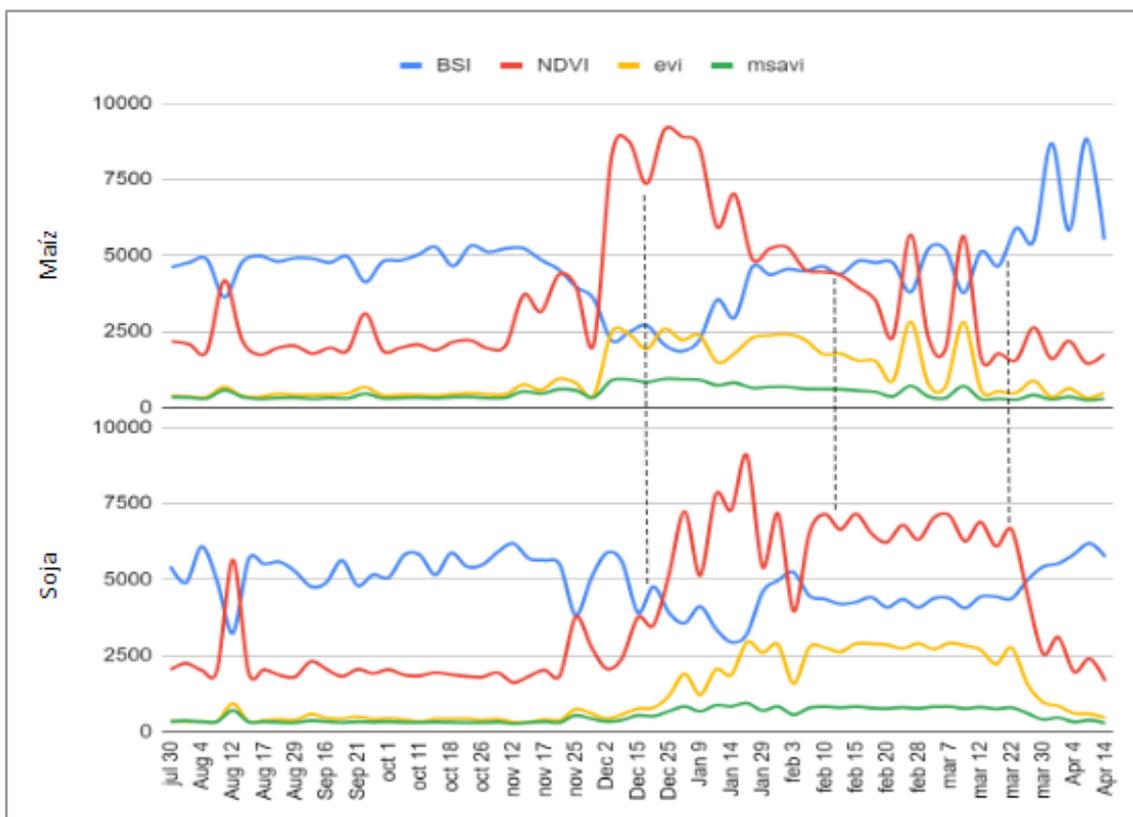


Figura 5: Comportamiento temporal de índices BSI, NDVI, EVI y MSAVI en muestras de cultivos de maíz y soja. En líneas de puntos se señalan comportamientos diferenciales entre cultivos.

Una vez depurados los datos del relevamiento ad-hoc, se incorporó un conjunto adicional de datos provistos por la Secretaría de Agricultura de la Provincia de Córdoba compuesto por más de 20.000 registros geoposicionados de la base de recetas fitosanitarias. Para la depuración de este conjunto se utilizaron los 1.968 registros seleccionados anteriormente. Con ellos se realizó una primera etapa de clasificación supervisada de los lotes con los datos provistos por Agricultura, conservando aquellos que mostraron coincidencias entre el cultivo informado y el resultado de la clasificación, obteniendo una **muestra total final de 5.084 registros**.

El modelo de clasificación seleccionado fue Random Forest⁸ (Breiman, 2001), con 44 árboles de decisión y la capacidad de ajuste de los modelos se estimó mediante un proceso de validación cruzada en 10 grupos.

⁸ En el Anexo II se puede encontrar una breve descripción de los algoritmos de aprendizaje computacional utilizados en el proceso.

3.2.3 Resultados obtenidos

Los modelos desarrollados permitieron estimar la distribución de los cultivos de maíz, soja y otras especies en toda el área cultivada en la provincia, a nivel de unidades de manejo, mostrando que la mayor proporción en la campaña 2021-2022 estuvo dedicada al **cultivo de soja, con un total de 4.650.471 hectáreas**. El **cultivo de maíz, por su parte, abarcó un área de 2.604.464 hectáreas**. También se identificaron 571.962 ha dedicadas a otros cultivos, totalizando 7.826.897 ha la superficie destinada a cultivos estivales para la campaña analizada (Fig. 6).

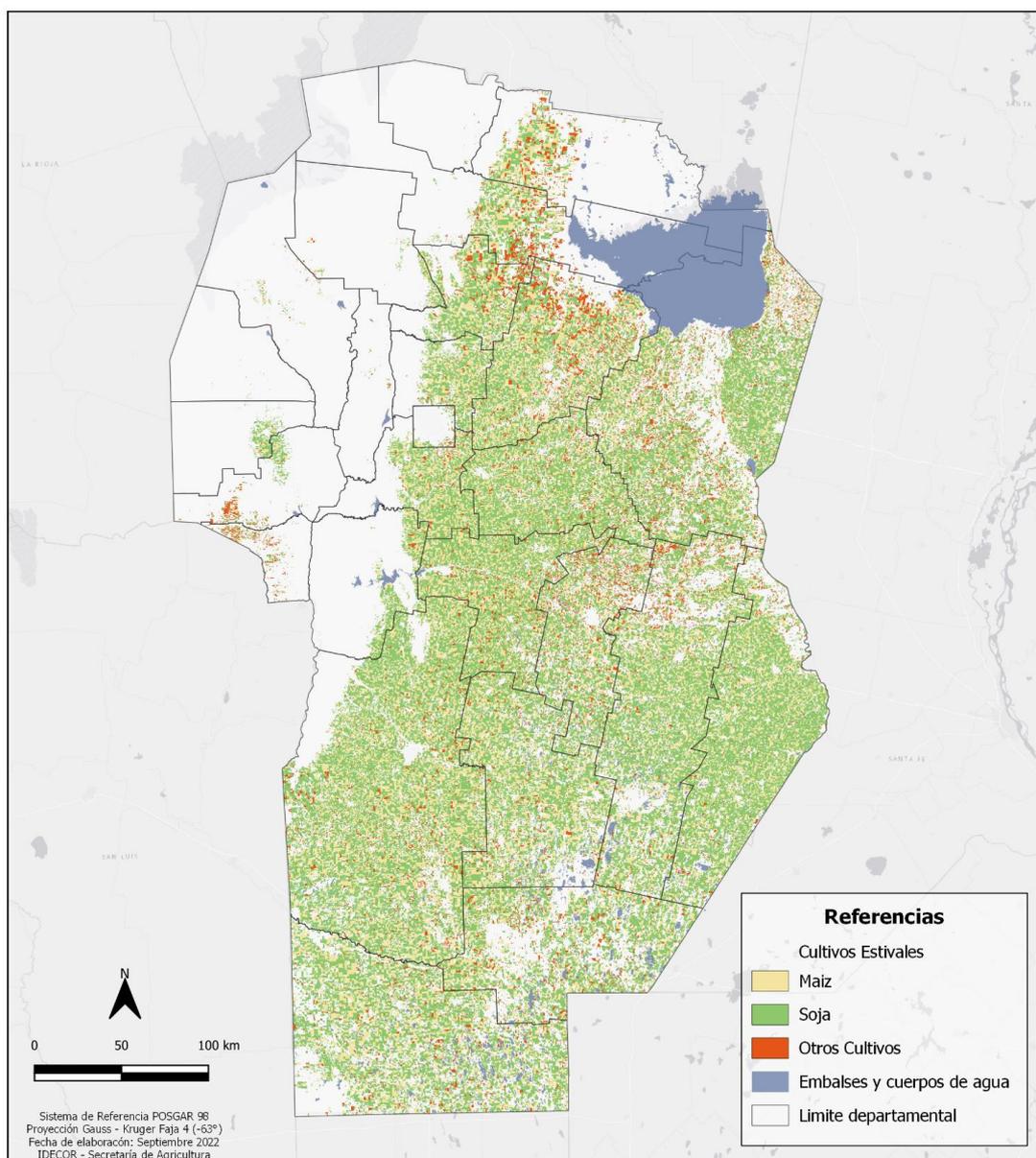


Figura 6: Distribución del tipo de cultivos campaña 2021-2022, Provincia de Córdoba.

La evaluación del desempeño mostró una **precisión global de los resultados de 92,11%**, mientras que el valor alcanzado por el coeficiente de Kappa (Bishop et al., 1975) para la clasificación fue de 0,842, indicando un alto grado de precisión en la clasificación ejecutada a través del modelo y una baja probabilidad de que estos aciertos hayan sido debidos al azar. La Tabla 1 presenta la matriz de confusión obtenida.

Tabla 1: Matriz de confusión para la clasificación de cultivos, campaña 2021-22.

Clases de cultivo		Predicción				Precisión del usuario
		Soja	Maíz	Otros	Totales	
Observación	Soja	3.183	121	17	3.321	95,84
	Maíz	161	1.303	30	1.494	87,22
	Otros	37	35	197	269	73,23
	Totales	3.381	1.459	244	5.084	
Precisión del productor		94,14	89,31	80,74	Precisión Global	92,11
Coeficiente Kappa						0,842

Se analizó la importancia de las variables en el modelo de clasificación producido, encontrando que, en general, los valores de reflectancia en la banda 6 de Sentinel 2, así como los índices de vegetación y suelo desnudo estimados para los mosaicos que van del 1 al 30 de septiembre de 2021 y del 1 al 24 de noviembre de 2021 fueron los que mayor influencia tuvieron en el modelo (Fig. 7).



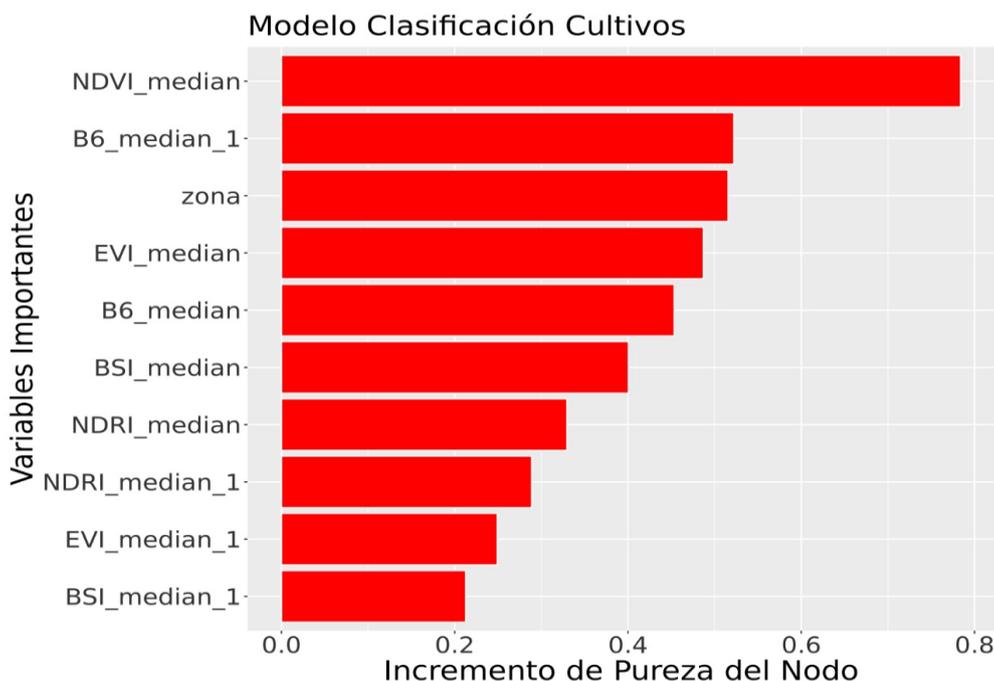


Figura 7: Importancia de las variables que mayor influencia mostraron en el modelo de clasificación de lotes por tipo de cultivo, campaña 2021-22.

Dónde:

- NDVI_median: Mediana del índice NDVI por lote para el período del 01-09-2021 al 30-09-2021.
- B6_median: Mediana de la banda 6 por lote para el período del 01-09-2021 al 30-09-2021.
- EVI_median: Mediana del índice EVI por lote para el período del 01-09-2021 al 30-09-2021.
- B6_median_1: Mediana de la banda 6 por lote para el período del 01-11-2021 al 24-11-2021.
- zona: zona de pertenencia del lote
- NDRI_median: Mediana del índice NDRI por lote para el período del 01-09-2021 al 30-09-2021.
- BSI_median: Mediana del índice BSI por lote para el período del 01-09-2021 al 30-09-2021.
- NDRI_median_1: Mediana del índice NDRI por lote para el período del 01-11-2021 al 24-11-2021.
- EVI_median_1: Mediana del índice EVI por lote para el período del 01-11-2021 al 24-11-2021.
- BSI_median_1: Mediana del índice BSI por lote para el período del 01-11-2021 al 24-11-2021.



4. Estimación de rendimientos por cultivo

Con el objetivo de estimar los rendimientos agrícolas en el área cultivada de la provincia, se construyó una grilla de predicción regular con una resolución espacial de 25 ha (celdas de 500 metros de lado). En total, la grilla cuenta con 431.835 celdas, sobre las cuales se imputaron 160 variables 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.

A partir de la muestra de rindes a campo, que cuentan con información no sólo de las 160 variables utilizadas en la estimación, sino del rendimiento en qq/ha de cada cultivo, se entrenaron diferentes algoritmos de aprendizaje computacional, seleccionando para la predicción el de mejor desempeño. Una vez identificado el mejor modelo, se estimaron los rendimientos para los cultivos de soja y maíz en toda la grilla de predicción.

4.1 Muestra utilizada

De las 2.247 observaciones recopiladas originalmente, y luego de depurar datos espacialmente atípicos, se obtuvo una **muestra final de 2.175 observaciones**, correspondientes a 908 datos de maíz (con una superficie sembrada de 58.400 ha en la muestra) y 1.267 datos de soja (con una superficie sembrada de 111.200 ha en la muestra) (Tabla 2). Como se detalló anteriormente, en la estimación de los rendimientos no se consideraron otros cultivos de menor cobertura, como maní o girasol, entre otros.

Tabla 2: Estadística descriptiva de rendimientos en qq/ha. y superficie sembrada en la muestra.

Cultivo	Obs.	Media	Mediana	Min	Max	CV(%)	P10	P90	Sup. sembrada
Soja	1.267	33	32	0	60	34	18	49	111.200
Maíz	908	76	78	0	155	33	45	105	58.400

Para la soja, la media muestral fue de 33 qq/ha, con un coeficiente de variación (cv%) del 34%. El maíz, en cambio, tuvo un rendimiento medio en la muestra de 76 qq/ha y un coeficiente de variación (cv%) del 33%.

Del análisis de la muestra surge también, que el 10% de los datos de menor rendimiento en soja tuvieron valores por debajo de los 18 qq/ha, mientras que para el maíz no superaron los 45 qq/ha. En el otro extremo, el 10% de los datos de mayor rendimiento en soja arrojaron valores por encima de los 49 qq/ha y de 105 qq/ha en el caso del maíz.





Una vez conformada la muestra, se asignó a cada observación (punto georreferenciado) el valor correspondiente de 160 variables territoriales (el detalle se incluye en el Anexo I), con el objetivo de entrenar diferentes algoritmos de aprendizaje computacional.

4.2 Modelos de aprendizaje computacional aplicados

Se entrenaron 3 modelos de aprendizaje computacional: Quantile Regression Forest (QRF) (Meinshausen, 2006), Gradient Boosting Model (GBM) (Friedman, 2001) y Support Vector Regression (SVR) (Cortes & Vapnik, 1995)⁹.

La capacidad de ajuste de los modelos se estimó mediante un proceso de validación cruzada en 10 grupos. Este proceso consiste en subdividir aleatoriamente la muestra en 10 grupos de similar tamaño, extraer uno de ellos y estimar los 3 modelos utilizando los datos de los 9 grupos restantes; posteriormente, se mide la capacidad de ajuste de cada modelo en el grupo extraído inicialmente. El procedimiento continúa de manera iterativa hasta que cada uno de los 10 grupos fue evaluado fuera de la muestra. De esta forma el procedimiento genera estimaciones del error de predicción, que son utilizados para el cálculo de indicadores del desempeño de los modelos (Tabla 3).



Tabla 3. Criterios de ajuste de los modelos predictivos.

Modelos	MAPE	MedAPE	RMSE	nRMSE	R2	CV	CD
QRF	0,20	0,11	12,14	23,76	0,80	0,20	0,20
SVR	0,25	0,15	14,18	27,75	0,62	0,24	0,25
GBM	0,25	0,16	13,42	26,26	0,75	0,24	0,25

Donde:

- 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 (out-of-sample).
- CV: Coeficiente de variación entre el valor observado y el valor predicho (out-of-sample).
- CD: Coeficiente de dispersión entre el valor observado y el valor predicho (out-of-sample).

Cómo se muestra en la Tabla 3, el algoritmo que minimiza las distintas medidas de cuantificación del error es QRF, además de explicar una mayor proporción de

⁹ Una breve reseña de las características de cada uno de estos modelos se puede encontrar en el Anexo II.



la variabilidad total del rendimiento (R²). Esta es la razón por la cual fue el modelo elegido para hacer las predicciones de los rendimientos en toda la grilla (unidades de 25 ha).

Las variables más importantes en la estimación del modelo QRF fueron las siguientes (Fig. 11).

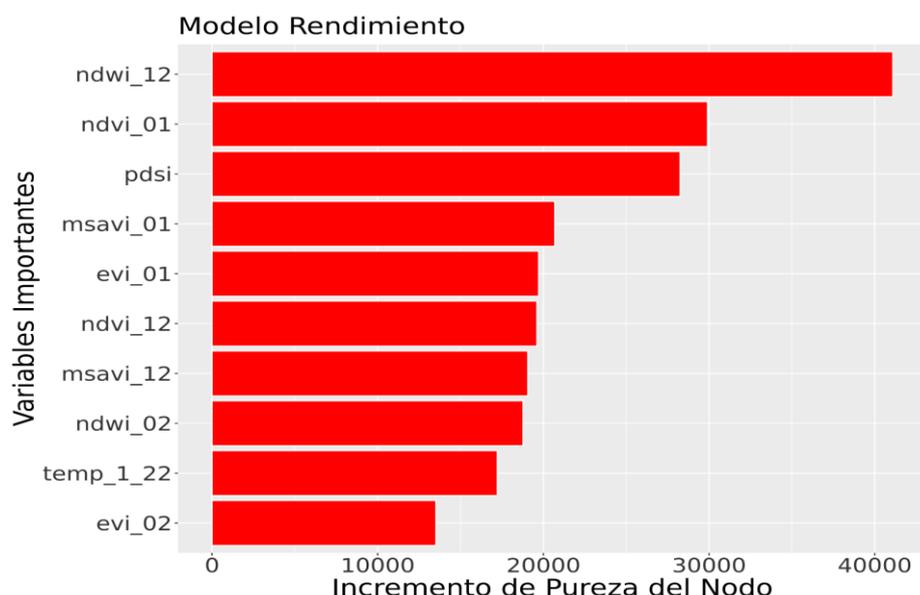


Figura 8: Importancia de las variables que mayor influencia mostraron en el modelo de estimación de rinde por tipo de cultivo, campaña 2021-22.

Donde:

- ndwi_12 = Índice de agua de diferencia normalizada modificado (MNDWI), diciembre 2021.
- ndvi_01 = Índice de vegetación de diferencia normalizada (NDVI), enero 2022.
- pdsi = Índice de Severidad de Sequía media histórica (1958-2019).
- msavi_01 = Índice de vegetación ajustado al suelo modificado (MSAVI), enero 2022.
- evi_01 = Índice de vegetación mejorado (EVI), enero 2022.
- msavi_12 = Índice de vegetación ajustado al suelo modificado (MSAVI), diciembre 2021.
- ndwi_02 = Índice de agua de diferencia normalizada modificado (MNDWI), febrero 2022.
- temp_1_22 = temperatura media enero 2022 (Red de Estaciones Meteorológicas de la Provincia de Córdoba).
- evi_02 = Índice de vegetación mejorado (EVI), febrero 2022.

En Anexo I se brindan mayores detalles de estas y las demás variables utilizadas.

4.3 Resultados obtenidos

A partir del modelo aplicado (QRF) se estimaron los rendimientos de soja y maíz en qq/ha para toda el área cultivada de la provincia, tal como puede apreciarse en las Figuras 12 y 13, resultando la siguiente estadística descriptiva de la predicción de los rendimientos (Tabla 4).

Tabla 4. Estadística descriptiva de la predicción de los rendimientos por cultivo.

Cultivo	Media	Min.	Max	CV(%)	P10	P90
Soja	31	9	58	21	23	40
Maíz	77	18	146	18	55	92

Para los rendimientos de soja, en menos del 10% de las celdas de la grilla se obtuvieron valores menores a 23 qq/ha, mientras que para el maíz menos del 10% de las celdas resultaron menores a 55 qq/ha. Para los rendimientos más elevados, el 10% de las celdas de soja superaron los 40 qq/ha; entretanto, para el maíz el 10% de las celdas superaron los 90 qq/ha. **Para la soja, la media provincial fue de 31 qq/ha**, con un coeficiente de variación (cv%) del 21%, en cambio **el maíz tuvo un rendimiento medio de 77 qq/ha** y un coeficiente de variación (cv%) del 19%.



Estimación de rendimiento Maíz (qq/ha)

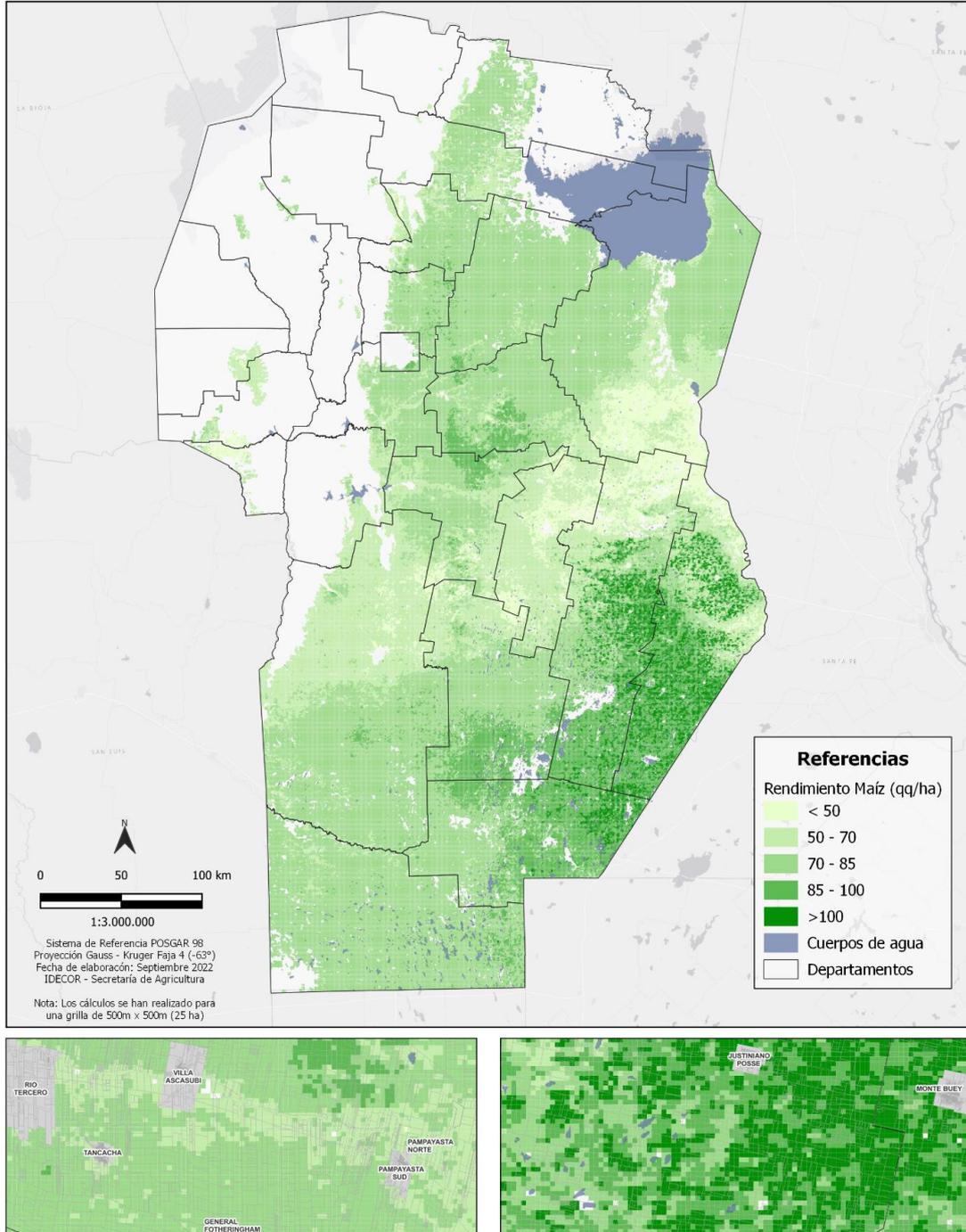


Figura 9. Mapa de rendimiento de maíz a nivel de grilla de 25 ha, campaña 2021-22, provincia de Córdoba

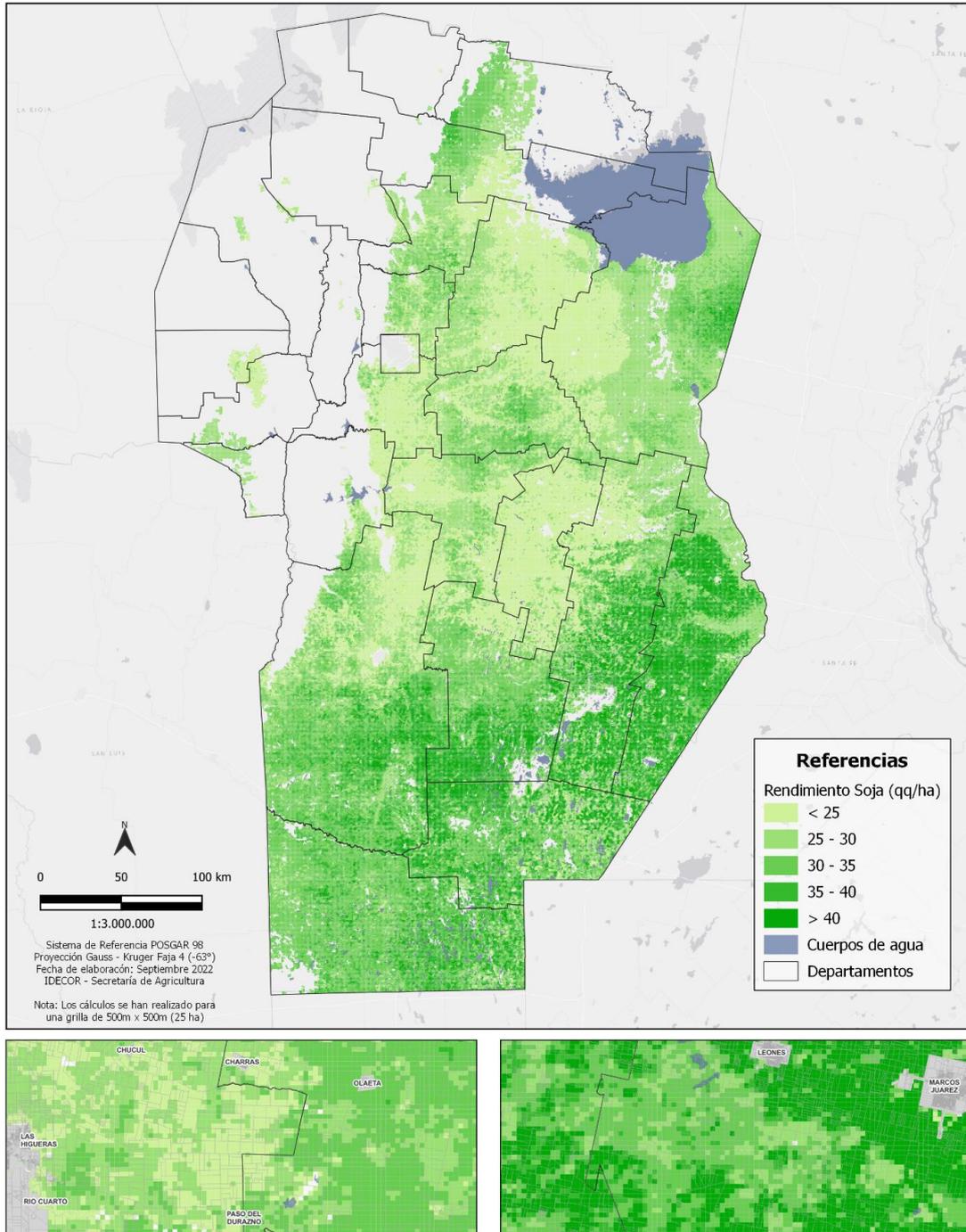


Figura 10. Mapa de rendimiento de soja a nivel de grilla de 25 ha, campaña 2021-22, provincia de Córdoba

5. Estimación de la producción de soja y maíz

El proceso de cálculo y determinación de la producción surge de multiplicar el promedio ponderado de la superficie de la celda en la unidad de manejo por el rendimiento producido en esa celda. A estos fines, se realiza una intersección espacial, entre las celdas y las unidades de manejo. En términos matemáticos, se presenta la siguiente ecuación:

$$RPPL_{C_i} = \sum \frac{sup_interseccion_celda_i}{sup_unidad\ de\ manejo} rto_celda_i$$

$RPPL_{C_i}$ = Rendimiento promedio ponderado del lote para el cultivo i

La siguiente figura (Fig. 11) ejemplifica el cálculo al realizar la intersección entre el mapa de rendimiento estimado sobre la grilla de 25 ha y el mapa de lotes, donde se identificaron los cultivos. El porcentaje corresponde a la razón entre la superficie de la intersección y la grilla, multiplicado por 100.



Figura 11. Intersección entre grilla con valores de rindes y unidades de manejo (lote agrícola).

Al contar con la producción agrícola a nivel de unidades de manejo, la información puede agregarse en unidades espaciales más grandes, como por ejemplo unidades de las cartas de suelo, unidades censales u otras administrativas, pedanía o departamentos, entre otras. A los fines del presente informe, en **Tablas 5 y 6 se presenta la producción en toneladas de soja y maíz, agregada a nivel departamental**; la Fig. 12 presenta la producción de ambos cultivos mapeada por departamento.

Tabla 5. Soja. Superficie sembrada, producción y rendimiento a nivel departamental, campaña 2021-21. Fuente: elaboración propia

Departamentos	Superficie (ha)	Producción (tn)	Rendimiento
Calamuchita	44.508	116.404	26,15
Capital	4.906	11.955	24,37
Colon	70.083	202.415	28,88
Cruz del Eje	2.849	3.687	12,94
General Roca	446.540	1.568.290	35,12
General San Martin	195.849	453.327	23,15
Ischilin	9.427	22.901	24,29
Juárez Celman	315.211	1.068.144	33,89
Marcos Juárez	465.220	1.791.644	38,51
Minas	302	-	-
Pocho	12.804	25.251	19,72
Pte Roque Saenz	271.598	1.000.496	36,84
Punilla	693	1.349	19,45
Rio Cuarto	815.493	2.590.987	31,77
Rio Primero	226.384	578.177	25,54
Rio Seco	46.582	144.622	31,05
Rio Segundo	258.201	736.169	28,51
San Alberto	14.833	30.919	20,85
San Javier	7.005	14.272	20,37
San Justo	482.621	1.401.988	29,05
Santa María	98.601	248.599	25,21
Sobremonte	433	479	11,07
Tercero Arriba	254.499	673.193	26,45
Totoral	97.062	274.231	28,25
Tulumba	64.130	183.085	28,55
Unión	444.442	1.528.920	34,40
Total Provincial	4.650.276	14.671.502	31,55

Tabla 6. Maíz. Superficie sembrada, producción y rendimiento a nivel departamental, campaña 2021-21. Fuente: elaboración propia

Departamentos	Superficie (ha)	Producción (tn)	Rendimiento
Calamuchita	24.714	171.610	69,44
Capital	2.097	16.704	79,66
Colon	45.217	365.380	80,81
Cruz del Eje	1.508	7.021	46,56
General Roca	246.187	1.927.380	78,29
General San Martin	89.472	510.239	57,03
Ischilin	5.576	33.659	60,36
Juárez Celman	202.033	1.578.034	78,11
Marcos Juárez	219.857	2.120.994	96,47
Minas	17	-	-
Pocho	4.268	21.972	51,49
Pte Roque Saenz	138.406	1.228.852	88,79
Punilla	73	227	31,10
Rio Cuarto	477.055	3.464.459	72,62
Rio Primero	173.700	1.346.693	77,53
Rio Seco	44.596	338.058	75,80
Rio Segundo	144.823	1.146.432	79,16
San Alberto	5.274	31.762	60,23
San Javier	2.440	11.688	47,90
San Justo	208.947	1.449.907	69,39
Santa María	74.162	558.151	75,26
Sobremonte	9	25	26,45
Tercero Arriba	156.086	1.146.212	73,43
Totalal	64.567	489.424	75,80
Tulumba	44.084	318.620	72,28
Unión	229.190	1.969.661	85,94
Total Provincial	2.604.358,40	20.253.165,90	77,77

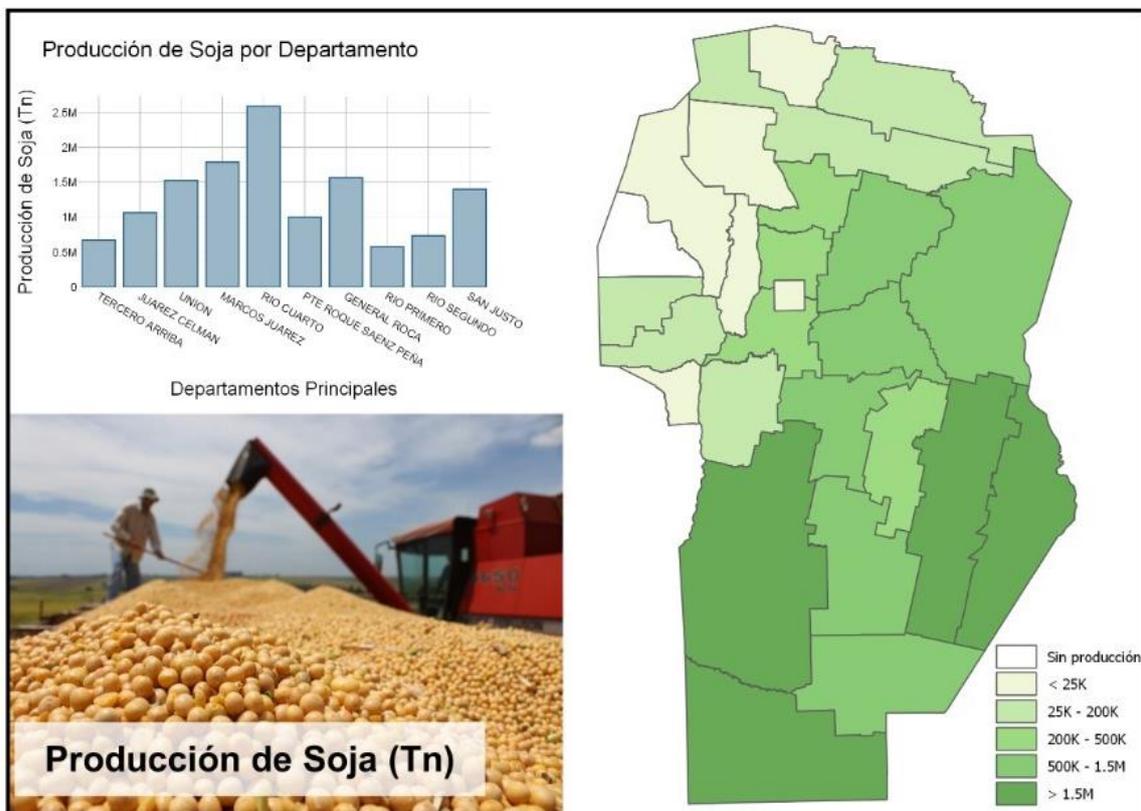
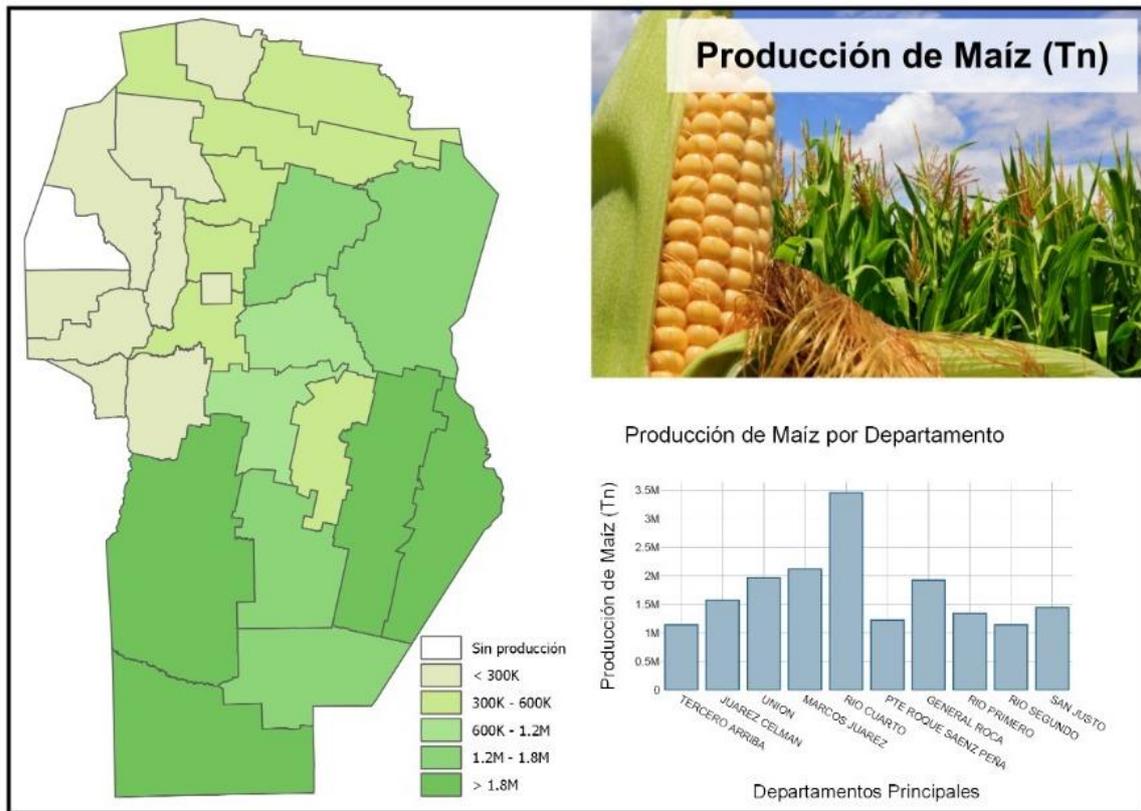


Fig. 12 presenta la producción de ambos cultivos mapeada por departamento.

Referencias Bibliográficas

Bishop, Y.M.; Fienberg S.E.; Holland P.W. (1975). *Discrete Multivariate Analysis Theory and Practice*, MIT Press, Carnbridge, Massachusetts, 557 p.

Breiman, L. (2001). Random forests. *Machine learning*, 45(1), 5-32.

Canny, J. (1986). A computational approach to edge detection. *IEEE Transactions on pattern analysis and machine intelligence*, (6), 679-698.

Cortes, C.; Vapnik, V. (1995). Support-vector networks. *Machine Learning*, 20(3), 273-297.

Friedman, JH (2001). Aproximación de funciones codiciosas: una máquina de aumento de gradiente. *Anales de estadística*, 1189-1232.

Gelder, B. K.; Kaleita, A. L.; Cruse, R. M. (2009). Estimating mean field residue cover on midwestern soils using satellite imagery. *Agronomy journal*, 101(3), 635-643.

Guyot, G.; Baret, F.; Major, D. J. (1988). High spectral resolution: determination of spectral shifts between the red and near infrared. In *ISPRS Congress*.

Haralick, R. M.; Shanmugam, K.; Dinstein, I. H. (1973). Textural features for image classification. *IEEE Transactions on systems, man, and cybernetics*, (6), 610-621.

Huete, A.; Justice, C.; Liu, H. (1994). Development of vegetation and soil indices for MODIS-EOS. *Remote Sensing of Environment*, 49, 224-234.

Jamalabad, M. S.; Abkar, A.A. (2004). Forest Canopy Density Monitoring, Using Satellite Images. XXth ISPRS Congress, Istanbul 12-23 July.

Meinshausen, N. (2006). Quantile regression forests. *Journal of Machine Learning Research*, 7(Jun), 983-999.

QGIS Development Team. (2021). QGIS Geographic Information System. Open Source Geospatial Foundation (Version 3.16). En <http://qgis.osgeo.org>

Rouse, J.W.; Haas, R.H.; Schell, J.A.; Deering, D.W. (1973): Monitoring vegetation systems in the great plains with ERTS. Third 80 ERTS Symposium, NASA SP-351, pp. 309-317.



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

Variables	Descripción
altura_median	Mediana de la altura (msnm)
pend_median	Mediana de la pendiente (%)
t_med_anual	Temperatura media anual (1970-2000) World Clim
rad_solar	Radiación solar media acumulada (1970-2000)
arrenda_2021	Arrendamiento agrícola 2021, en base a IDECOR
ndvi_mediana	Mediana de NDVI (promedio histórico 2000-2020)
evapo_medi_an	Evapotranspiración media mensual acumulada de la serie (2001-2020)
pp_med_an	Precipitación media acumulada anual histórica (1958-2019)
t_min_med	Temperatura máxima anual media (1958-2019)
t_max_med	Temperatura mínima anual media (1958-2019)
def_hidric	Déficit hídrico medio histórico (1958-2019))
pdsi	Índice de Severidad de Sequía media histórica (1958-2019)
rec_1median	Recurrencia de inundaciones (1984-2000) mediana dentro de la celda
rec_1stdev	Recurrencia de inundaciones (1984-2000) desvío estándar dentro de la celda
rec_2median	Recurrencia de inundaciones (2001-2009) mediana dentro de la celda
rec_2stdev	Recurrencia de inundaciones (2001-2009) desvío estándar dentro de la celda
rec_3median	Recurrencia de inundaciones (2010-2018) mediana dentro de la celda
rec_3stdev	Recurrencia de inundaciones (2010-2018) desvío estándar dentro de la celda
perc_agua_perm	% agua en la celda (Año hidrológico 2020/06/01 al - 2021/04/20) - IDECOR
perc_agua_aneg	% agua en la celda (Año hidrológico 2020/06/01 al - 2021/04/20) - IDECOR
n2_cob16	% sup. Cultivos anuales de secano - IDECOR
n2_cob17	% sup. Cultivos Irrigados - IDECOR
cu_moda	Moda de CU de la celda
ip_median	Mediana del Índice de Productividad
ip_stdev	Desvío estándar del Índice de Productividad

perc_pivot	Porcentaje de superficie con pivotes en la grilla
lail_11	Superficie total de hojas verdes por m2 de suelo horizontal para vegetación baja (m2/m2). Noviembre
lail_12	Superficie total de hojas verdes por m2 de suelo horizontal para vegetación baja (m2/m2). Diciembre
lail_01	Superficie total de hojas verdes por m2 de suelo horizontal para vegetación baja (m2/m2). Enero
lail_02	Superficie total de hojas verdes por m2 de suelo horizontal para vegetación baja (m2/m2). Febrero
lail_03	Superficie total de hojas verdes por m2 de suelo horizontal para vegetación baja (m2/m2). Marzo
lail_04	Superficie total de hojas verdes por m2 de suelo horizontal para vegetación baja (m2/m2). Abril
rad_ss_11	Cantidad de radiación solar que alcanza la superficie de la Tierra por m2 (J/m2). Noviembre
rad_ss_12	Cantidad de radiación solar que alcanza la superficie de la Tierra por m2 (J/m2). Diciembre
rad_ss_01	Cantidad de radiación solar que alcanza la superficie de la Tierra por m2 (J/m2). Enero
rad_ss_02	Cantidad de radiación solar que alcanza la superficie de la Tierra por m2 (J/m2). Febrero
rad_ss_03	Cantidad de radiación solar que alcanza la superficie de la Tierra por m2 (J/m2). Marzo
rad_ss_04	Cantidad de radiación solar que alcanza la superficie de la Tierra por m2 (J/m2). Abril
evap_t_11	Evaporación total acumulada en el mes de análisis (m). Noviembre
evap_t_12	Evaporación total acumulada en el mes de análisis (m). Diciembre
evap_t_01	Evaporación total acumulada en el mes de análisis (m). Enero
evap_t_02	Evaporación total acumulada en el mes de análisis (m). Febrero
evap_t_03	Evaporación total acumulada en el mes de análisis (m). Marzo
evap_t_04	Evaporación total acumulada en el mes de análisis (m). Abril
evi_11	Índice de vegetación mejorado (EVI). Noviembre
ndwi_11	Índice de agua de diferencia normalizada modificado (MNDWI). Noviembre
msavi_11	Índice de vegetación ajustado al suelo modificado (MSAVI). Noviembre
ndvi_11	Índice de vegetación de diferencia normalizada (NDVI). Noviembre



evi_12	Índice de vegetación mejorado (EVI). Diciembre
ndwi_12	Índice de agua de diferencia normalizada modificado (MNDWI). Diciembre
msavi_12	Índice de vegetación ajustado al suelo modificado (MSAVI). Diciembre
ndvi_12	Índice de vegetación de diferencia normalizada (NDVI). Diciembre
evi_01	Índice de vegetación mejorado (EVI). Enero
ndwi_01	Índice de agua de diferencia normalizada modificado (MNDWI). Enero
msavi_01	Índice de vegetación ajustado al suelo modificado (MSAVI). Enero
ndvi_01	Índice de vegetación de diferencia normalizada (NDVI). Enero
evi_02	Índice de vegetación mejorado (EVI). Febrero
ndwi_02	Índice de agua de diferencia normalizada modificado (MNDWI). Febrero
msavi_02	Índice de vegetación ajustado al suelo modificado (MSAVI). Febrero
ndvi_02	Índice de vegetación de diferencia normalizada (NDVI). Febrero
evi_03	Índice de vegetación mejorado (EVI). Marzo
ndwi_03	Índice de agua de diferencia normalizada modificado (MNDWI). Marzo
msavi_03	Índice de vegetación ajustado al suelo modificado (MSAVI). Marzo
ndvi_03	Índice de vegetación de diferencia normalizada (NDVI). Marzo
evi_04	Índice de vegetación mejorado (EVI). Abril
ndwi_04	Índice de agua de diferencia normalizada modificado (MNDWI). Abril
msavi_04	Índice de vegetación ajustado al suelo modificado (MSAVI). Abril
ndvi_04	Índice de vegetación de diferencia normalizada (NDVI). Abril
vswl_1_09	Volumen de agua en profundidades de 0 a 7 cm (m ³ /m ³). Septiembre - ERA-Land
vswl_1_10	Volumen de agua en profundidades de 0 a 7 cm (m ³ /m ³). Octubre - ERA-Land
vswl_1_11	Volumen de agua en profundidades de 0 a 7 cm (m ³ /m ³). Noviembre - ERA-Land
vswl_1_12	Volumen de agua en profundidades de 0 a 7 cm (m ³ /m ³). Diciembre - ERA-Land



vswl_1_01	Volumen de agua en profundidades de 0 a 7 cm (m3/m3). Enero - ERA-Land
vswl_1_02	Volumen de agua en profundidades de 0 a 7 cm (m3/m3). Febrero - ERA-Land
vswl_1_03	Volumen de agua en profundidades de 0 a 7 cm (m3/m3). Marzo - ERA-Land
vswl_1_04	Volumen de agua en profundidades de 0 a 7 cm (m3/m3). Abril - ERA-Land
vswl_2_09	Volumen de agua en profundidades de 7 a 28 cm (m3/m3). - ERA-Land Septiembre
vswl_2_10	Volumen de agua en profundidades de 7 a 28 cm (m3/m3). Octubre - ERA-Land
vswl_2_11	Volumen de agua en profundidades de 7 a 28 cm (m3/m3). Noviembre - ERA-Land
vswl_2_12	Volumen de agua en profundidades de 7 a 28 cm (m3/m3). Diciembre - ERA-Land
vswl_2_01	Volumen de agua en profundidades de 7 a 28 cm (m3/m3). Enero - ERA-Land
vswl_2_02	Volumen de agua en profundidades de 7 a 28 cm (m3/m3). Febrero - ERA-Land
vswl_2_03	Volumen de agua en profundidades de 7 a 28 cm (m3/m3). Marzo - ERA-Land
vswl_2_04	Volumen de agua en profundidades de 7 a 28 cm (m3/m3). Abril - ERA-Land
vswl_3_09	Volumen de agua en profundidades de 28 a 100 cm (m3/m3). Septiembre - ERA-Land
vswl_3_10	Volumen de agua en profundidades de 28 a 100 cm (m3/m3). Octubre - ERA-Land
vswl_3_11	Volumen de agua en profundidades de 28 a 100 m (m3/m3). Noviembre - ERA-Land
vswl_3_12	Volumen de agua en profundidades de 28 a 100 cm (m3/m3). Diciembre - ERA-Land
vswl_3_01	Volumen de agua en profundidades de 28 a 100 cm (m3/m3). Enero - ERA-Land
vswl_3_02	Volumen de agua en profundidades de 28 a 100 cm (m3/m3). Febrero - ERA-Land
vswl_3_03	Volumen de agua en profundidades de 28 a 100 cm (m3/m3). Marzo - ERA-Land
vswl_3_04	Volumen de agua en profundidades de 28 a 100 m (m3/m3). Abril - ERA-Land
vswl_4_09	Volumen de agua en profundidades de 100 a 289 cm (m3/m3). Septiembre - ERA-Land
vswl_4_10	Volumen de agua en profundidades de 100 a 289 cm (m3/m3). Octubre - ERA-Land



vswl_4_11	Volumen de agua en profundidades de 100 a 289 cm (m3/m3). Noviembre - ERA-Land
vswl_4_12	Volumen de agua en profundidades de 100 a 289 cm (m3/m3). Diciembre - ERA-Land
vswl_4_01	Volumen de agua en profundidades de 100 a 289 cm (m3/m3). Enero - ERA-Land
vswl_4_02	Volumen de agua en profundidades de 100 a 289 cm (m3/m3). Febrero - ERA-Land
vswl_4_03	Volumen de agua en profundidades de 100 a 289 cm (m3/m3). Marzo - ERA-Land
vswl_4_04	Volumen de agua en profundidades de 100 a 289 cm (m3/m3). Abril - ERA-Land
hs_09	Volumen de agua del suelo de 0 a 50 cm de profundidad (%). Septiembre - ERA-Land
hs_10	Volumen de agua del suelo de 0 a 50 cm de profundidad (%). Octubre - ERA-Land
hs_11	Volumen de agua del suelo de 0 a 50 cm de profundidad (%). Noviembre - ERA-Land
hs_12	Volumen de agua del suelo de 0 a 50 cm de profundidad (%). Diciembre - ERA-Land
hs_01	Volumen de agua del suelo de 0 a 50 cm de profundidad (%). Enero - ERA-Land
hs_02	Volumen de agua del suelo de 0 a 50 cm de profundidad (%). Febrero - ERA-Land
hs_03	Volumen de agua del suelo de 0 a 50 cm de profundidad (%). Marzo - ERA-Land
hs_04	Volumen de agua del suelo de 0 a 50 cm de profundidad (%). Abril - ERA-Land
vrm_02	Variación interanual de las medianas de los escenarios de rinde de maíz (adimensional). Febrero
vrm_03	Variación interanual de las medianas de los escenarios de rinde de maíz (adimensional). Marzo
vrm_04	Variación interanual de las medianas de los escenarios de rinde de maíz (adimensional). Abril
vrs_02	Variación interanual de las medianas de los escenarios de rinde de soja(adimensional). Febrero
vrs_03	Variación interanual de las medianas de los escenarios de rinde de soja (adimensional). Marzo
vrs_04	Variación interanual de las medianas de los escenarios de rinde de soja (adimensional). Abril
media_mo	Media de Contenido de materia orgánica en suelo utilizando una grilla de 3x3
std_mo	Desvío estándar de Contenido de materia orgánica en suelo utilizando una grilla de 3x3



media_p	Media de Contenido de fósforo en suelo utilizando una grilla de 3x3
std_p	Desvío estándar de Contenido de fósforo en suelo utilizando una grilla de 3x3
nitrogeno18	Contenido de nitrógeno en suelo
potasio18	Contenido de potasio en suelo
cic18	Capacidad de intercambio catiónico
media_ph	Media de pH del suelo utilizando una grilla de 3x3
std_ph	Desvío estándar de pH del suelo utilizando una grilla de 3x3
media_arcilla	Media de valores de contenido de arcilla- SH utilizando una grilla de 3x3
std_arcilla	Desvío estándar de valores de contenido de arcilla- SH utilizando una grilla de 3x3
media_limo	Media de valores de contenido de limo utilizando una grilla de 3x3
std_limo	Desvío estándar de valores de contenido de limo utilizando una grilla de 3x3
media_arena	Media de valores de contenido de arena utilizando una grilla de 3x3
std_arena	Desvío estándar de valores de contenido de arena utilizando una grilla de 3x3
pp_11_21	Precipitación promedio Noviembre - Calculado a partir de datos OMIXOM
pp_12_21	Precipitación promedio Diciembre - Calculado a partir de datos OMIXOM
pp_1_22	Precipitación promedio Enero - Calculado a partir de datos OMIXOM
pp_2_22	Precipitación promedio Febrero - Calculado a partir de datos OMIXOM
pp_3_22	Precipitación promedio Marzo - Calculado a partir de datos OMIXOM
pp_4_22	Precipitación promedio Abril - Calculado a partir de datos OMIXOM
temp_11_21	Temperatura Media Noviembre - Calculado a partir de datos OMIXOM
temp_12_21	Temperatura Media Diciembre - Calculado a partir de datos OMIXOM
temp_1_22	Temperatura Media Enero - Calculado a partir de datos OMIXOM
temp_2_22	Temperatura Media Febrero - Calculado a partir de datos OMIXOM
temp_3_22	Temperatura Media Marzo - Calculado a partir de datos OMIXOM



temp_4_22	Temperatura Media Abril - Calculado a partir de datos OMIXOM
temp_max_11_21	Temperatura Máxima Noviembre - Calculado a partir de datos OMIXOM
temp_max_12_21	Temperatura Máxima Diciembre - Calculado a partir de datos OMIXOM
temp_max_1_22	Temperatura Máxima Enero - Calculado a partir de datos OMIXOM
temp_max_2_22	Temperatura Máxima Febrero - Calculado a partir de datos OMIXOM
temp_max_3_22	Temperatura Máxima Marzo - Calculado a partir de datos OMIXOM
temp_max_4_22	Temperatura Máxima Abril - Calculado a partir de datos OMIXOM
temp_min_11_21	Temperatura Mínima Noviembre - Calculado a partir de datos OMIXOM
temp_min_12_21	Temperatura Mínima Diciembre - Calculado a partir de datos OMIXOM
temp_min_1_22	Temperatura Mínima Enero - Calculado a partir de datos OMIXOM
temp_min_2_22	Temperatura Mínima Febrero - Calculado a partir de datos OMIXOM
temp_min_3_22	Temperatura Mínima Marzo - Calculado a partir de datos OMIXOM
temp_min_4_22	Temperatura Mínima Abril - Calculado a partir de datos OMIXOM



Anexo II - Modelos de Aprendizaje Computacional utilizados en las estimaciones

Random Forest (RF) (Breiman, 2001)

Es un algoritmo que se caracteriza por ser de comportamiento anidado (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.

Consiste en generar una serie de árboles de regresión a partir de la misma muestra mediante bootstrapping. RF combina las predicciones de estos árboles mediante el promedio condicional de las predicciones de cada modelo. Esta técnica se denomina bagging y permite reducir simultáneamente el sesgo y la varianza en la predicción. Además, RF permite realizar predicciones más precisas al disminuir la correlación entre los árboles. Esto se consigue seleccionando aleatoriamente un subconjunto de variables independientes que se consideran en cada nodo de división.

Quantile Regression Forest (QRF) (Meinshausen, 2006)

El modelo QRF es una adaptación del algoritmo Random Forest (RF), que sigue la misma estrategia de ensamble, bagging, combinando árboles de decisión independientes, pero se distingue en el procedimiento de agregación de las predicciones de cada árbol. A diferencia de RF que agrega a través de la media, QRF permite realizar predicciones no sólo de la media condicionada de la variable objetivo, sino también del resto de los cuantiles. Es decir, se amplía la información brindada por Random Forest y, por lo tanto, se puede proveer de mayor información sobre la distribución total condicional de la variable objetivo, y no únicamente de la media condicionada. Esta mayor información suele utilizarse para detectar outliers (o anomalías) en el set de datos por medio de la utilización de intervalos cuantílicos. Por lo tanto, con el algoritmo QRF se puede obtener la mediana de las predicciones del bosque 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, logrando una estimación territorial del nivel de dispersión en los resultados. También surgen medidas como la desviación estándar (que se encuentra en la misma unidad de la variable objetivo) y el coeficiente de variación o dispersión que están en porcentaje. De este modo se obtiene mayor información.

Gradient Boosting Machine (GBM) (Friedman, 2001)

El algoritmo GBM es una técnica basada en árboles. En lugar de combinar las predicciones de los distintos árboles mediante bagging, GBM utiliza una técnica llamada boosting, combinada con un vector de descenso de gradiente. La



diferencia entre el bagging y el boosting consiste en la técnica de remuestreo aplicada en cada árbol. En el bagging cada observación tiene la misma probabilidad de ser replicada en la base de datos, mientras que en boosting se ponderan en función del error con el que fueron con el que fueron predichas. Cuanto mayor sea el error, mayor será la probabilidad de que una observación permanezca en la muestra de árboles posteriores, buscando enfatizar los casos difíciles. Además del boosting, GBM aplica un vector de descenso de gradiente, donde los errores se consideran como un vector con dirección y no simplemente como una magnitud. Este vector apunta a una dirección donde se minimiza la función de pérdida del algoritmo.

Support Vector Regression (SVR) (Cortes & Vapnik, 1995)

El algoritmo SVR consiste en optimizar una función convexa, minimizando el error de predicción, sujeto a un margen de tolerancia, generalmente llamado epsilon. El objetivo del algoritmo consiste en encontrar, a través de una reducción en la dimensión del problema original mediante la aplicación de kernels, la función lineal (hiperplano, vector soporte) que mejor representa los datos dentro de un margen epsilon. Es decir, consiste en estimar el modelo lineal que mejor predice las observaciones, minimizando el error sujeto a un máximo de tolerancia (epsilon).



Créditos. Equipo de trabajo

Relevamiento de datos

Personal de las Agencias Zonales de la Secretaría de Agricultura y Ganadería: General Deheza, Las Varillas, Bell Ville, Villa Huidobro, Marcos Juárez, San Francisco, El Tio, Laboulaye, Tancacha, Rio Cuarto, Santa Rosa de Rio Primero, San Agustin, Villa Maria, Villa de Maria de Rio Seco, Villa del Rosario, Dean Funes, Jesus Maria, Morteros, Oliva, Villa Dolores y Cura Brochero.

Personal del Ministerio de Agricultura y Ganadería: Constanza Álvarez, German Font, Luis Pizarro, Marcos Blanda.

Personal de las Agencias de Extensión INTA: Horacio Videla, Sebastian Muñoz, Mariela Monetti, Víctor Capuccino, Verónica Bustos, Alberto Guendulain, Enrique Yanaco, Carlos Triadani.

Equipo Técnico de IDECOR: Federico Monzani, Mariano Córdoba, Juan Pablo Carranza, Ezequiel Pozzi, Leonardo Sandon, Hernán Morales, Mario Piumetto, Agustina Michelini, José Jachuf, Carlos Salinas.





Ministerio de
**AGRICULTURA
Y GANADERÍA**

Ministerio de
FINANZAS



 mapascordoba.gov.ar  idecor.gov.ar

 idecore@cba.gov.ar