



Universidad de Buenos Aires
Facultad de Ciencias Exactas y Naturales – Facultad de Ingeniería
Maestría en Explotación de Datos y Descubrimiento del Conocimiento

Tesis de Maestría

Metodología automática para mapeo y seguimiento de la
condición de cultivos agrícolas durante la zafra a partir de
imágenes satelitales y aprendizaje automático en Uruguay

Autor: Ing. Agr. Adrián Cal
Director: Dr. Pablo Musé
Codirector: Dr. Ing. Javier Preciozzi

Tabla de Contenido

1. Resumen.....	4
2. Introducción	5
3. Antecedentes	7
4. Software	9
5. Zona de estudio y bases de datos	9
5.1 Zona de estudio	9
5.2 Bases de datos disponibles	10
5.2.1 Planes de Uso y Manejo Responsable del Suelo (DGRN-MGAP)	10
5.2.2 FUCREA y URF.....	15
5.2.3 Mapa Integrado de Cobertura/Uso del Suelo de Uruguay Año 2018	17
6. Imágenes satelitales e índices de vegetación	20
6.1 Imágenes satelitales Sentinel y Terra.....	20
6.2 Índices de vegetación y cross-ratio VH/VV	27
7. Procesamiento de las bases	29
7.1 PLANES	29
7.2 FUCREA-URF	33
7.3 COBERTURAS.....	34
8. Extracción de series temporales satelitales	36
8.1 Series temporales de Terra	36
8.2 Series temporales de Sentinel.....	40
9. Depuración de PLANES con clasificación no supervisada	48
9.1 Regionalización de PLANES	48
9.2 Depuración de PLANES mal etiquetados.....	49
10. Clasificación supervisada de las bases	60
10.1 Clasificación de PLANES con series temporales de EVI Terra	60
10.2 Clasificación de FUCREA-URF con series temporales de Sentinel.....	66
10.2.1 Depuración de puntos no agrícolas en las chacras	66
10.2.2 Experimentos evaluando bandas, índices, fechas y algoritmos.....	73
10.2.3 Clasificación de chacras.....	80
10.3 Clasificación de PLANES con series temporales de Sentinel	86
10.4 Clasificación de PLANES con modelo entrenado en FUCREA-URF	89
10.5 Clasificación binaria agrícola/no-agrícola con COBERTURAS y FUCREA-URF.....	92
11. Clasificación pixel a pixel en Google Earth Engine	95
11.1 Mapa binario agrícola/no-agrícola en GEE.....	95
11.2 Mapa de cultivos agrícolas a nivel de chacra en GEE.....	114

12. Seguimiento de la condición de cultivos agrícolas en la chacra	120
12.1. Zonificación de la chacra durante la zafra.....	120
12.2. Extracción de métricas fenológicas de la chacra.....	131
13. Conclusiones.....	137
14. Agradecimientos	139
15. Índice de figuras y tablas.....	140
16. Bibliografía	146

1. Resumen

El presente trabajo tiene como principal objetivo el desarrollo de una metodología para la identificación de cultivos agrícolas de verano en Uruguay, a partir del uso de imágenes satelitales y aprendizaje automático. Para ello, como insumo se utilizaron bases de datos con información georreferenciada de campo, e imágenes satelitales de libre acceso. De las bases, dos fueron proporcionadas por organizaciones de productores agropecuarios y son de chacras con agricultura, y otras del Ministerio de Ganadería, Agricultura y Pesca de Uruguay (MGAP). Una con información de coberturas y usos del suelo, y otra de chacras con planes de uso para conservación del suelo por erosión. Los cultivos agrícolas presentes en las bases son maíz, soja y sorgo. Las imágenes utilizadas provienen de los satélites de tipo óptico, Sentinel 2 y Terra, y de Sentinel 1, que tiene un sensor radar de apertura sintética. De estos satélites, en particular de Sentinel 1 y 2, se obtuvieron series temporales de varias bandas del espectro electromagnético.

El procesamiento de los datos se hizo a través del uso combinado de algoritmos de clasificación no supervisada, supervisada y de reducción de la dimensionalidad. De los no supervisados se evaluaron Gaussian Mixture Model (GMM), K-Means, Learning Vector Quantization (LVQ), Self Organizing Map (SOM) de Kohonen y Spectral Clustering. De los supervisados se evaluaron AdaBoost, Gradient Tree Boosting, Random Forest, Support Vector Machine (SVM) y XGBoost. El algoritmo de reducción de la dimensionalidad utilizado fue t-Distributed Stochastic Neighbor Embedding (t-SNE).

Una de las herramientas utilizadas para el procesamiento de los datos georreferenciados y las imágenes satelitales es la moderna y potente plataforma de análisis geoespacial en la nube Google Earth Engine (earthengine.google.com). Con el uso de esta herramienta se generaron mapas a nivel nacional clasificados pixel a pixel. Combinando clasificación no supervisada con el algoritmo de t-SNE se desarrolló una técnica para identificar y depurar datos mal etiquetados, así como zonas sin agricultura dentro de las chacras. Para la clasificación supervisada se generaron modelos que en la dimensión espacial trabajan clasificando puntos georreferenciados, o a nivel nacional, pixel a pixel, y a nivel de cultivo estos indican si hay agricultura o no (modelos binarios) y el tipo de cultivo específico (maíz, soja o sorgo).

De los algoritmos supervisados, Random Forest y XGBoost fueron los de mejor desempeño. Para los modelos de clasificación binaria, el mejor desempeño fue 98,5 de *F1-Score*. Para los modelos de cultivo a nivel de chacra, en maíz y sorgo, el mejor desempeño fue 75,3 y 77,5 de *F1*, y en soja 98,5, siendo este último superior al obtenido en varios trabajos internacionales.

Otro de los objetivos específicos de este trabajo fue poder realizar el seguimiento de la condición de los cultivos en la chacra durante la zafra. Para ello se desarrolló una metodología que utilizando K-Means realiza la zonificación de la chacra en dos o más zonas de acuerdo con la evolución temporal del índice NDVI desde la siembra. Además, a partir de la firma fenológica mediana de la chacra, suavizada a paso diario, y con análisis de derivada primera, se extrajeron las métricas fenológicas de mínimo, punto de inflexión, máximo, días desde mínimo a máximo, amplitud de NDVI entre mínimo y máximo. A partir de estas métricas se puede caracterizar una chacra y hacer comparaciones con otras.

Palabras claves: agricultura, aprendizaje automático, clasificación supervisada y no supervisada, Google Earth Engine, imágenes satelitales, métricas fenológicas, Sentinel, t-SNE, XGBoost.

2. Introducción

Uruguay tiene una superficie de 17.622.000 hectáreas (176.220 km²) (FAO, 2019), de las cuales 1.284.860¹ son ocupadas por cultivos de verano² (soja, arroz y maíz principalmente) y 385.700¹ por cultivos de invierno³ (trigo y cebada cervecera principalmente), representando 7,3 y 2,2% de la superficie total del país respectivamente. La superficie de cultivos de verano equivale a 12.849 km², siendo superior o similar a la superficie total de países como Jamaica, Catar, o Montenegro (10.990, 11.610, 13.810 km² respectivamente), mientras que la de cultivos de invierno (3.857 km²) es similar a la de Trinidad y Tobago (5.130 km²) (FAO, 2019).

El área agrícola de cultivos de verano creció de 300.000 hectáreas en la zafra 2000-01 (MGAP-DIEA, 2016) a 1.400.000 en 2012-13 (MGAP-DIEA, 2016). Este crecimiento fue impulsado por el cultivo de soja principalmente, y determinado por el incremento en los precios internacionales, con valores que pasaron de 200 a 700 dólares la tonelada del 2000 a 2012 (World Bank, 2019). Esta expansión del área agrícola también ocurrió en otros países de la región como Argentina, Brasil y Paraguay (Llonch, 2019). Este incremento en precios internacionales también ocurrió para otras materias primas agrícolas como maíz, trigo y arroz (International Grains Council, 2019).

Los cultivos agrícolas más importantes en superficie ocupada¹ son soja, trigo, cebada cervecera, arroz y maíz, con 995.740, 213.560, 172.140, 150.300 y 100.700 hectáreas respectivamente. La producción en toneladas para soja, arroz, trigo, maíz, caña de azúcar, cebada cervecera y sorgo, es de 2.691.000, 1.364.000, 1.108.000, 542.000, 419.000, 350.000 y 263.000 toneladas respectivamente (URUGUAY XXI, 2018).

En cuanto a las exportaciones agrícolas para 2017, soja, arroz, trigo y cebada malteada, representaron 1.189, 448, 40 y 184 millones de dólares respectivamente (URUGUAY XXI, 2018). En cuanto a personas ocupadas, la producción de cereales (excepto arroz), legumbres y semillas emplea directamente a 28.000 personas (URUGUAY XXI, 2018).

La producción agrícola hace un uso intensivo de fertilizantes, insecticidas, fungicidas y herbicidas como el conocido glifosato⁴, los cuales tienen impactos sobre el medio ambiente. En Uruguay se han registrado pérdidas continuas de colmenas de abejas desde el año 2002, siendo las mayores pérdidas en regiones donde soja y girasol son los cultivos dominantes en verano. Este fenómeno de pérdidas de colmenas se conoce como “Problema de colapso de colonias” (CCD⁵, según siglas en inglés) (Pareja et al., 2011). Algunas de las causas del CCD que manejan los investigadores es el uso de pesticidas (Frazier, Mullin, Frazier, & Ashcraft, 2011).

En un trabajo realizado por Carrasco-Letelier et al. (2014) se relevó el 73% del área de drenaje de Uruguay con algún uso agropecuario. En dicha área se analizó la calidad de agua (contenido de fósforo, oxígeno y sólidos en suspensión) y en algunas de las cuencas ubicadas en zona agrícolas se registraron valores de fósforo disuelto superiores a 25 ug/l (máximo permitido por ley⁶), mientras que en otras, niveles de oxígeno disuelto inferiores a 5 mg/l (mínimo permitido por ley). En otro trabajo realizado por Carrasco-Letelier y Beretta-Blanco (2017), estimaron

¹ Promedio 2017-2021 a partir de datos de DIEA (Dirección de Estadísticas Agropecuarias).

² Cultivos de verano: sembrados en primavera.

³ Cultivos de invierno: sembrados en otoño.

⁴ Glifosato: N-(fosfonometil) glicina (denominación IUPAC).

⁵ CCD: colony collapse disorder.

⁶ Decreto 253/79

perdidas de suelo por erosión hídrica superiores al valor tolerable (7 t/ha/año) en regiones con presencia de soja y producción lechera.

Con el objetivo de promover sistemas de producción sostenibles que hagan un uso y manejo adecuado del suelo para su conservación, por ley⁷ a partir del año 2013 en Uruguay los productores que siembren más de 50 hectáreas con agricultura deben presentar planes de uso del suelo. La realización de estos planes implica hacer rotaciones de cultivos con pasturas para que la erosión anual no supere un umbral tolerable según el tipo de suelo. En el año 2016 se presentaron 15.000 planes de uso por un total de 1.500.000 hectáreas, lo que correspondió al 96% del área sembrada para esa zafra (MGAP-DGRN, 2016).

En Uruguay la mayoría de la agricultura se realiza en condiciones de secano⁸ -con la excepción del cultivo de arroz que es bajo riego-, siendo el agua disponible para los cultivos la suministrada por lluvia. Para la zafra 2017-18, de un total de 1.200.000 hectáreas de cultivos de verano (maíz, soja, y sorgo), solo el 2% (24.000 hectáreas) se hizo con riego (Arenare, Couto, & Fontán, 2018). En el año 2017 se promulgó una ley⁹ para regular el uso de agua para riego con destino agrario. Esta ley permite la asociación de productores para desarrollar proyectos multi-prediales de riego. Es de esperar que a partir de esta ley el área bajo riego en agricultura aumente considerablemente.

Considerando la importancia de la agricultura desde el punto de vista económico, social y ambiental para Uruguay, sería útil poder desarrollar una metodología que mapee las zonas agrícolas de secano de verano y siga la evolución de la condición de los cultivos durante la zafra a partir del procesamiento y análisis de imágenes satelitales con aprendizaje automático.

Los objetivos específicos de este trabajo fueron desarrollar una metodología que permitiera:

1. Identificar y mapear los cultivos agrícolas de secano de verano durante la zafra en curso a partir de imágenes satelitales;
2. Realizar el seguimiento de la condición de los cultivos agrícolas en diferentes momentos de desarrollo y crecimiento entre siembra y cosecha;
3. Identificar la especie del cultivo presente en la chacra (maíz, soja, sorgo, etc.).

⁷ Ley N° 15.239.

⁸ Secano: sin riego.

⁹ Ley N° 19.553.

3. Antecedentes

Para el mapeo de usos y coberturas del suelo se trabaja con imágenes obtenidas por sensores a bordo de satélites de observación de la Tierra. Los dos tipos principales de satélites usados son los que tienen sensor de tipo óptico y los que tienen sensor de tipo radar de apertura sintética (SAR). Los sensores ópticos son del tipo pasivo, es decir, miden la radiación solar reflejada por la Tierra, mientras que los del tipo SAR son del tipo activo, es decir que emiten una señal o pulso hacia la Tierra y miden cuanto de esa señal vuelve hacia el sensor del satélite (ESA, 2019a). La mayoría de los trabajos que se han hecho son con imágenes de satélites del tipo óptico.

Con respecto a trabajos realizados para la identificación de usos y coberturas del suelo a partir de imágenes satelitales, en Uruguay se han hecho varios trabajos por diferentes instituciones estatales. Para empezar, mencionar los trabajos realizados con la metodología Land Cover Classification System (LCCS) de FAO (Di Gregorio & Jansen, 1998) para los años, 2000, 2008, 2011 y 2015. Este trabajo se hizo en colaboración entre el Ministerio de Vivienda, Ordenamiento Territorial y Medio Ambiente (MVOTMA) y el Ministerio de Ganadería, Agricultura y Pesca (MGAP). En estos trabajos se usó como material base imágenes satelitales Landsat 5 y 8, y se hicieron mapeos para todo el país (MVOTMA, 2019).

Por otro lado, la Dirección General de Recursos Naturales (DGRN) del MGAP viene trabajando desde el año 1996 con el uso de imágenes Landsat (5, 7 y 8), Sentinel 2 y Resourcesat 2 para el monitoreo de áreas con cultivos de invierno y verano (MGAP-DGRN, 2019a y MGAP-DGRN, 2019b).

También la Unidad de Agroclima y Sistemas de Información (GRAS) del Instituto Nacional de Investigación Agropecuaria (INIA) viene trabajando desde el año 1998 con imágenes satelitales Landsat (5, 7, y 8), NOAA-AVHRR, Suomi NPP (National Polar-orbiting Partnership) y MODIS para monitoreo del estado de la vegetación en general y de áreas con cultivos agrícolas (INIA-GRAS, 2019).

En el 2014, el Departamento de Sistemas Ambientales de Facultad de Agronomía (UDELAR), el Laboratorio de Análisis Regional y Teledetección (LART, IFEVA, Facultad de Agronomía - UBA), y el Departamento de Métodos Cuantitativos y Sistemas de Información (Facultad de Agronomía - UBA) realizaron un trabajo sobre caracterización de usos y coberturas del suelo a partir de imágenes satelitales de MODIS (Baeza, Baldassini, Bagnato, Pinto, & Paruelo, 2014).

A nivel regional, el Instituto Nacional de Pesquisas Espaciais (INPE) de Brasil viene realizando trabajos para el monitoreo de la caña de azúcar con el sistema Sugarcane crop monitoring in Brazil (CANASAT) a partir de imágenes de Landsat (5 y 7), CBERS-2 y Resourcesat I (INPE-DSR, 2019). También la Empresa Brasileira de Pesquisa Agropecuária (EMBRAPA) ha desarrollado el Sistema de Observação e Monitoramento da Agricultura no Brasil (SOMA) donde publica mapas de cultivos agrícolas obtenidos a partir de MODIS (EMBRAPA, 2019). EMBRAPA también ha desarrollado la herramienta SATVeg (Sistema de Análise Temporal da Vegetação) para la visualización de series de tiempo de índices de vegetación (EVI y NDVI) de MODIS para el Cono Sur (EMBRAPA-CNPTIA, 2019).

Con respecto a Argentina, el INTA para el período 2006-07 realizó un mapa de cobertura y usos del suelo con la metodología LCCS (Di Gregorio & Jansen, 1998) a partir de imágenes Landsat y MODIS (Volante et al., 2009). También INTA genera anualmente desde el año 2000 informes con el área de cultivos agrícolas (soja, maíz, poroto y algodón) para las provincias de Salta y Jujuy a partir de Landsat y MODIS (INTA, 2018). Por su parte el LART de la Facultad de Agronomía de la

UBA ha realizado muchos trabajos con sensoramiento remoto para identificación de áreas con cultivos con Landsat y MODIS (UBA-LART, 2019).

En Estados Unidos, en una iniciativa entre la Universidad de Maryland, Foreign Agricultural Service (FAS) del USDA y NASA, se ha desarrollado The Global Agriculture Monitoring (GLAM) (Becker-Reshef et al., 2010) para mapear y evaluar la condición de los cultivos a escala global. El sistema trabaja con imágenes de MODIS y el sensor VIIRS del satélite Suomi NPP (NPP Science Team, 2011).

En Europa, a efectos de apoyar la Política Agrícola Común (CAP) de la Unión Europea, la Agencia Espacial Europea (ESA) usa los satélites Sentinel 2 para la estimación de áreas con cultivos agrícolas. En ese sentido ha desarrollado los proyectos Sentinels for Common Agricultural Policy (ESA, 2019b) y Sentinel-2 for Agriculture (ESA, 2019c).

A partir de las imágenes satelitales se logran mapear áreas con agricultura mediante la identificación de zonas con actividad fotosintética. Es posible, a través del uso de índices de vegetación como el Índice de Vegetación de Diferencia Normalizada (NDVI), identificar la actividad fotosintética, la cual se calcula a partir de la radiación recibida desde la Tierra por el sensor montado en el satélite (NASA-Earth Observatory, 2000).

Para el procesamiento de las imágenes satelitales se usan diferentes técnicas de aprendizaje automático, como supervisadas, no supervisadas e híbridas (combinación de las anteriores). Se usan técnicas como redes neuronales artificiales, árboles de decisión, segmentación de imágenes, vecinos más cercanos, K-means, Support Vector Machine (SVM), Iso-data, entre otras (Abburu & Babu Golla, 2015).

Con respecto a técnicas supervisadas para mapeo de áreas con agricultura en Australia, sobre imágenes multitemporales de MODIS se probó la técnica de análisis de armónicos de series temporales (HANTS) y análisis de componentes principales (PCA) (Potgieter, Apan, Hammer, & Dunn, 2005). Otros autores, usando una combinación de series temporales de imágenes MODIS, Landsat 8 y Spot 4 hicieron mapeo automático de cultivos usando SVM (Waldner, Canto, & Defourny, 2015). En Ucrania, para clasificar usos del suelo de manera supervisada, se probaron redes neuronales convolucionales (1D y 2D), obteniendo mejor desempeño que usando Random Forest y logrando discriminar maíz de soja (Kussul, Lavreniuk, Skakun, & Shelestov, 2017).

Otros autores han intentado desarrollar algoritmos automáticos de clasificación con el uso de árboles de decisión con reglas definidas por expertos para mapeo de soja y maíz en Brasil a partir de MODIS, logrando una exactitud general de 87,2%, e identificando maíz de soja con la banda de infrarrojo de onda corta (Zhong, Hu, Yu, Gong, & Biging, 2016).

La tendencia actual para la identificación de cultivos agrícolas es trabajar con series temporales de imágenes para extraer métricas fenológicas que indican el momento de comienzo, desarrollo máximo y finalización de los cultivos. En ese sentido, uno de los trabajos más importantes es el que se ha hecho evaluando diferentes técnicas de suavizado de series temporales usando el filtro adaptativo Savitzky-Golay, funciones polinómicas y armónicas, y asimétricas gaussianas para extracción de métricas fenológicas (Jönsson & Eklundh, 2004). En Brasil clasificaron usos del suelo usando suavizado con función doble logística sobre series de tiempo de Enhanced Vegetation Index (EVI) a partir de imágenes Landsat 8, extrayendo métricas fenológicas y usando Random Forest (Bendini et al., 2016).

También se han probado técnicas de clasificación no supervisada segmentando imágenes de los satélites Spot 4 y 5 en diferentes categorías de usos y coberturas del suelo usando una

combinación de K-means con sistema de distribuciones de Pearson (Rekik, Zribi, Hamida, & Benjelloun, 2009).

Actualmente también se están probando técnicas para fusionar imágenes satelitales de diferentes plataformas como MODIS (buena resolución temporal pero baja espacial), Landsat 8 y Sentinel 2 (alta resolución espacial, pero baja temporal) para generar series temporales densas con alta resolución espacial y temporal, para el mapeo del progreso de cultivos a escala de predio (Gao et al., 2017).

4. Software

Para el procesamiento de los datos se utilizó los lenguajes R versión 4.0 mediante la IDE RStudio y Python 3.7 mediante la IDE Spyder. Para el manejo de la información geográfica se usó el software libre SIG QGIS 3.16.11 'Hannover'. Para la obtención y procesamiento de las imágenes satelitales se utilizó la plataforma de análisis geoespacial en la nube Google Earth Engine (earthengine.google.com). Esta plataforma permite acceder a un catálogo de imágenes de los satelitales Landsat, Sentinel (1 y 2), Terra y otros, y procesarlas en la nube usando lenguaje de programación Javascript y funciones propias de Google Earth Engine (GEE).

5. Zona de estudio y bases de datos

5.1 Zona de estudio

Ya que la agricultura de secano está presente en todo el Uruguay, la zona de estudio objetivo es todo el territorio. En éste hay zonas que son tradicionalmente agrícolas, comprendidas principalmente por los departamentos del litoral oeste (Colonia, Soriano, Río Negro y Paysandú) y otros ubicados en el centro (Durazno, Flores, Florida y San José). En la figura 1 se ve el mapa de las Regiones Agropecuarias definidas según último Censo General Agropecuario realizado en 2011 (MGAP-DIEA, 2015). Las Regiones Agropecuarias se definen de acuerdo con la producción agropecuaria predominante en cada una.

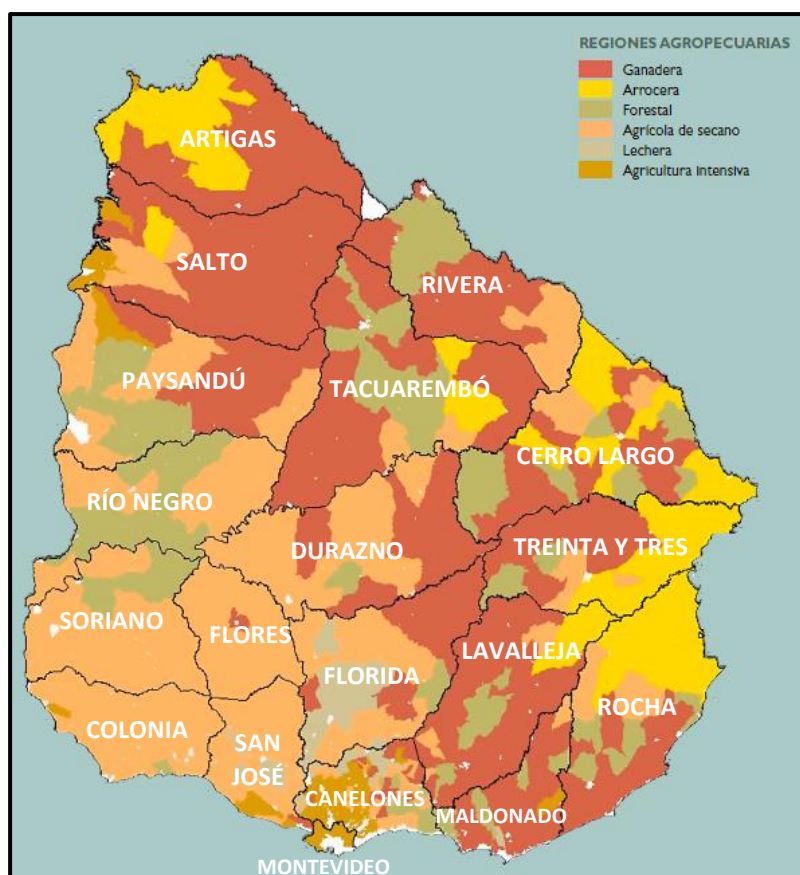


Figura 1. Mapa de Regiones Agropecuarias definidas según el último Censo General Agropecuario de 2011. Fuente: MGAP-DIEA (2015).

5.2 Bases de datos disponibles

Las bases de datos utilizadas fueron obtenidas de dos fuentes, unas de la Dirección General de Recursos Naturales (DGRN) del Ministerio de Ganadería, Agricultura y Pesca (MGAP), y otras de organizaciones privadas de productores agropecuarios como la Federación Uruguaya de Grupos CREA (FUCREA) y la empresa cooperativa Unión Rural de Flores (URF). Estas bases contienen información georreferenciada de coberturas y usos del suelo, así como chacras agrícolas con el tipo de cultivo presente según zafra. A continuación se describe cada una de las mismas.

5.2.1 Planes de Uso y Manejo Responsable del Suelo (DGRN-MGAP)

Una de las bases obtenidas de DGRN-MGAP corresponde a los Planes de Uso y Manejo Responsable del Suelo (PLANES). Todo productor agropecuario que realice agricultura en una superficie de al menos 50 hectáreas está obligado por ley a presentar ante DGRN-MGAP un **Plan de Uso del Suelo** de por lo menos un año. Este plan consiste en la realización de una rotación de cultivos agrícolas (de invierno y verano) y pasturas durante una cantidad de años con el fin de evitar que la erosión hídrica en la chacra supere determinado valor tolerable. La rotación de cultivos y pasturas tiene por objetivo conservar el suelo de acuerdo con lo establecido en el decreto reglamentario 405/008 de la ley 15.239.

La base PLANES cuenta con información de una superficie total de 1.575.000 hectáreas distribuidas entre 16.678 planes presentados entre el año 2013 y 2019 inclusive (ver figura 2). Esta base es un archivo con información geográfica de tipo vectorial en un formato conocido

como ESRI Shapefile, y puede ser manipulado con software del tipo Sistemas de Información Geográfica (SIG) como QGIS. Este archivo vectorial contiene los polígonos georreferenciados de todas las chacras presentadas para el período 2013-2019 y una tabla adjunta en formato DBF con la información del cultivo declarado a realizar en cada zafra de acuerdo con el Plan de Uso. En la tabla 1 se presenta un resumen de los campos de la base y en la figura 3 una parte de esta. En la figura 4 se ve un ejemplo de un plan conformado por varias chacras. Los datos etiquetados en dicha base no se corresponden con lo que en sensoramiento remoto se denomina “verdad de terreno”, que es cuando la etiqueta asignada es lo que se sembró en campo, sino que es la intención de siembra del productor según el Plan de Uso, el cual tiene el carácter de declaración jurada, por lo cual está comprometido a hacerlo. Puede haber circunstancias (climáticas, económicas, logísticas, etc.) donde no se cumpla con lo estipulado en el plan.

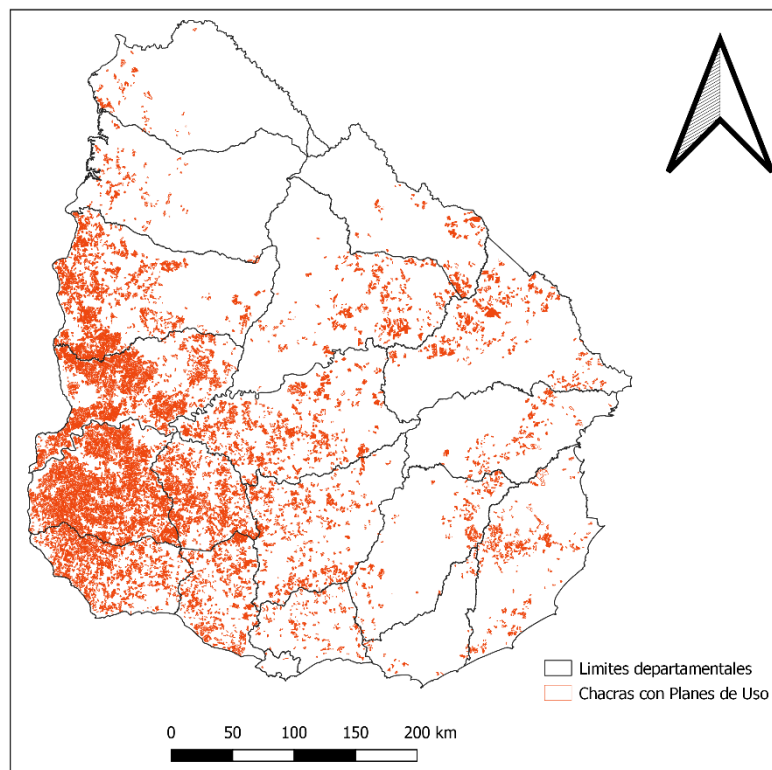


Figura 2. Mapa de distribución espacial de PLANES para el período 2013 a 2019.

CAMPO	TIPO	DESCRIPCIÓN
PLANNRO	NÚMERO ENTERO	Número identificatoria del Plan de Uso.
OI_2013	TEXTO	Cultivo a sembrar en invierno 2013.
PV_2013	TEXTO	Cultivo a sembrar en verano 2013-14.
OI_2014	TEXTO	Cultivo a sembrar en invierno 2014.
PV_2014	TEXTO	Cultivo a sembrar en verano 2014-15.
OI_2015	TEXTO	Cultivo a sembrar en invierno 2015.
PV_2015	TEXTO	Cultivo a sembrar en verano 2015-16.
OI_2016	TEXTO	Cultivo a sembrar en invierno 2016.
PV_2016	TEXTO	Cultivo a sembrar en verano 2016-17.
OI_2017	TEXTO	Cultivo a sembrar en invierno 2017.
PV_2017	TEXTO	Cultivo a sembrar en verano 2017-18.
OI_2018	TEXTO	Cultivo a sembrar en invierno 2018.
PV_2018	TEXTO	Cultivo a sembrar en verano 2018-19.
OI_2019	TEXTO	Cultivo a sembrar en invierno 2019.
PV_2019	TEXTO	Cultivo a sembrar en verano 2019-20.
AREA	NÚMERO REAL	Área en hectáreas.

Tabla 1. Campos presentes en la base PLANES.

PLANNRO	OI_2013	PV_2013	OI_2014	
1	Cereales de Invierno	Soja	Barbecho	Sorgo o M
2		Soja	Cereales de Invierno	Soja
3				
4	Cereales de Invierno	Soja	Cultivo de Cobertura	Soja
5		Soja	Pastura no Consociada	Pastura E
6	Cereales de Invierno	Sorgo o Maiz	Barbecho	Soja
7	Cereales de Invierno	Soja	Cereales de Invierno	Soja
8	Cereales de Invierno	Soja	Cereales de Invierno	Soja
9	Cereales de Invierno	Soja	Cereales de Invierno	Soja
10	Cereales de Invierno	Sorgo o Maiz	Pastura no Consociada	Soja
11		Soja	Cebada	Soja
12	Cereales de Invierno	Soja	Barbecho	Soja
13				Soja
14				Soja
15	Cereales de Invierno	Sorgo o Maiz	Barbecho	Soja
16		Soja	Cereales de Invierno	Soja
17	Cereales de Invierno	Soja	Cultivo de Cobertura	Soja
18	Cereales de Invierno	Sorgo o Maiz	Cereales de Invierno	Soja
19	Cereales de Invierno	Sorgo o Maiz	Cereales de Invierno	Soja
20	Cereales de Invierno	Sorgo o Maiz	Cereales de Invierno	Soja
21	Cereales de Invierno	Soja	Cereales de Invierno	Soja
22		Soja	Cereales de Invierno	Soja
23				

Figura 3. Ejemplo de registros del archivo DBF de la base PLANES.



Figura 4. Ejemplo de un plan con varias chacras en formato Shapefile.

En PLANES los cultivos agrícolas de verano principales por cantidad de planes y área ocupada son maíz, pasturas, soja y sorgo, como se ve en la tabla 2 y figura 5 para el período acumulado 2013-2019, siendo el cultivo dominante la soja con casi el 70% de los planes.

CULTIVO	PLANES 2013-19	
	CANTIDAD	PORCENTAJE
SOJA	50.391	69,35
SORGO O MAÍZ	8.926	12,28
PASTURA ESTABLECIDA O CONSOCIADA	8.174	11,25
PASTURA NO CONSOCIADA	3.191	4,39
SORGO	1.062	1,46
MAÍZ	690	0,95
BARBECHO	124	0,17
ARROZ	61	0,08
GIRASOL	25	0,03
COBERTURA DE VERANO	10	0,01
PAPA	8	0,01
TOTAL	72.662	100

Tabla 2. Distribución de planes por cultivo para el período acumulado 2013-19.

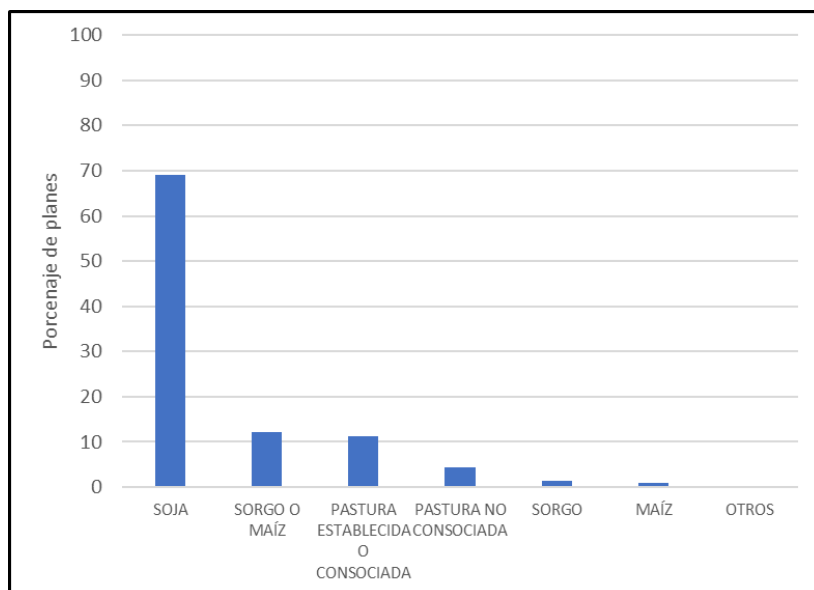


Figura 5. Gráfico de porcentaje de planes según cultivo para el período 2013-2019.

A modo de ejemplo, en la tabla 3 y figura 6 se presenta la estructura del conjunto de datos de PLANES de la zafra agrícola 2018-19. Como se ve en la tabla, el conjunto está desbalanceado, siendo el cultivo mayoritario soja con más del 60% de los datos, seguido por “pastura establecida o consociada” con más del 14% de los datos, “sorgo o maíz” con casi 10%, y “pastura no consociada”, maíz y sorgo, con menos del 10%. Es importante notar que el cultivo “sorgo o maíz” está presente, porque cuando se comenzaron a presentar los planes estaba permitido indicar si se iba a realizar sorgo o maíz, pero actualmente, es necesario especificar si se va a realizar uno u otro.

CULTIVO	PLANES		SUPERFICIE	
	CANTIDAD	PORCENTAJE	HECTÁREAS	PORCENTAJE
SOJA	6.813	61,8	677.969	67,5
PASTURA ESTABLECIDA O CONSOCIADA	1.948	17,7	142.500	14,2
SORGO O MAÍZ	1.095	9,9	94.960	9,5
PASTURA NO CONSOCIADA	758	6,9	61.965	6,2
MAÍZ	177	1,6	14.129	1,4
SORGO	232	2,1	13.065	1,3
TOTAL	11.023	100	1.004.587	100

Tabla 3. Resumen de planes según cultivo para la zafra 2018-19.

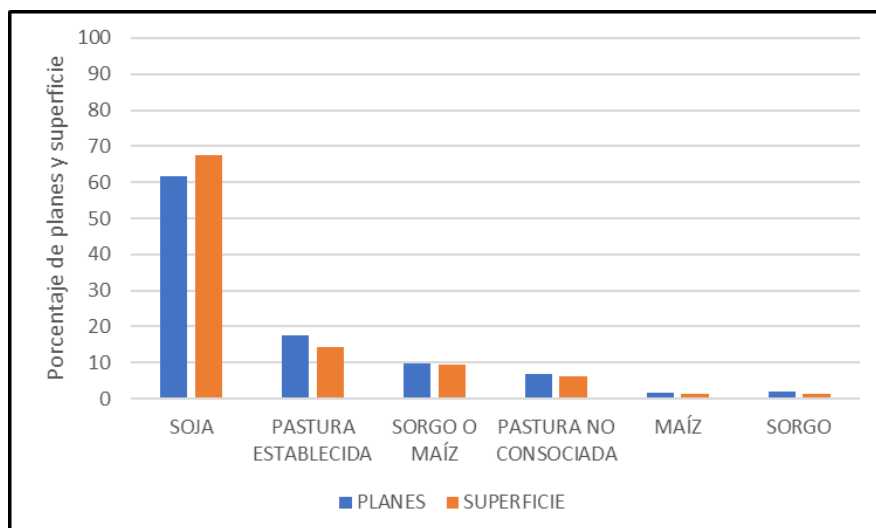


Figura 6. Gráfico de porcentaje de planes y superficie según cultivo para la zafra 2018-19.

5.2.2 FUCREA y URF

Los datos obtenidos de FUCREA y URF son bases de cultivos agrícolas de verano en formato SIG con chacras georreferenciadas, con el dato del cultivo sembrado, fecha de siembra, rendimiento, etc. En el caso de las bases de FUCREA y URF los datos etiquetados se corresponden con “verdad de terreno”, al contrario de lo que sucede con la base PLANES. A modo de facilitar el trabajo, ambas bases se fusionaron y se trabajaron como una única base. En la figura 7 se ve una chacra de la base conjunta FUCREA-URF en formato vectorial. Se trabajó con información de las zafas agrícolas 2017-18 y 2018-19, ya que a partir del año 2017 se cuenta con información continua de los satélites Sentinel. En la figura 8 se ve la distribución espacial de las chacras de la base. En las tablas 4 y 5 se ve un resumen de la información contenida en la base FUCREA- URF.



Figura 7. Ejemplo de polígonos de unas chacras de FUCREA-URF en formato Shapefile.

Como se observa en la figura 8, las chacras de FUCREA se distribuyen en los departamentos del litoral oeste (Colonia, Soriano, Río Negro y Paysandú), mientras que las chacras de URF principalmente en los departamentos del centro (Durazno, Flores, Florida y San José). De esta manera, con estas dos bases se tiene información que cubre la región tradicionalmente agrícola uruguaya.

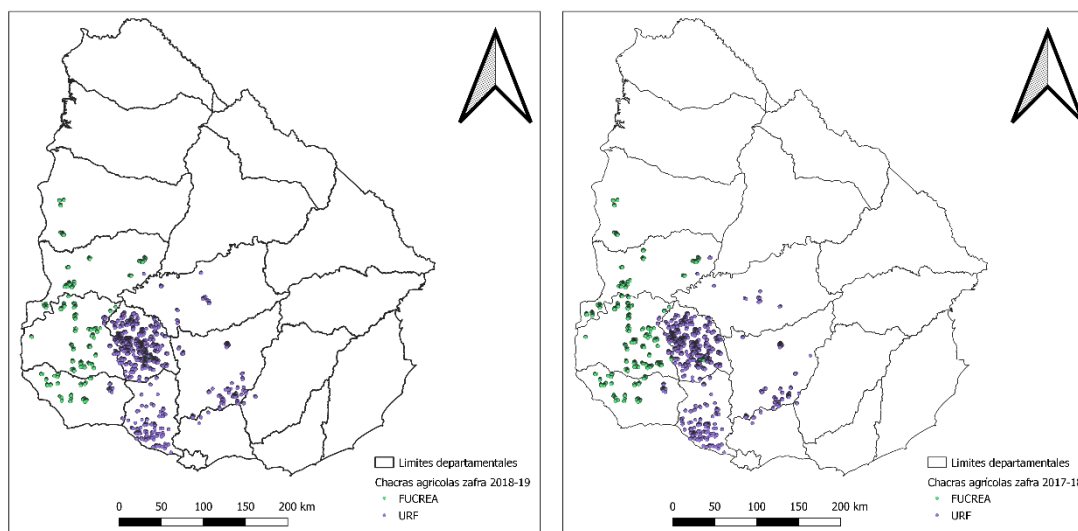


Figura 8. Mapa de chacras de FUCREA y URF para las zafas 2017-18 y 2018-19.

ZAFRA 2017-18				ZAFRA 2018-19			
EMPRESA	CULTIVO	CHACRAS	SUPERFICIE (HA)	EMPRESA	CULTIVO	CHACRAS	SUPERFICIE (HA)
URF	SOJA	770	29.392	URF	SOJA	820	31.753
FUCREA		363	16.004	FUCREA		224	10.524
SUBTOTAL		1.133	45.396	SUBTOTAL		1.044	42.277
FUCREA	MAÍZ	31	1.200	FUCREA	MAÍZ	29	1.513
URF		15	887	URF		28	1.145
SUBTOTAL		46	2.087	SUBTOTAL		57	2.658
URF	SORGO	28	1.029	URF	SORGO	52	2.138
FUCREA		1	33	FUCREA		3	126
SUBTOTAL		29	1.062	SUBTOTAL		55	2.264
TOTAL		1.208	48.544	TOTAL		1.156	47.199

Tabla 4. Resumen de las bases de FUCREA y URF para las zafas 2017-18 y 2018-19.

En la tabla 5 y figura 9 se presenta el resumen por cultivo para las zafas 2017-18 y 2018-19 de la base conjunta FUCREA-URF. Esta base está extremadamente desbalanceada, con más del 90% de los datos correspondiendo a soja, lo cual está en acuerdo con las estadísticas nacionales de siembra de cultivos. Maíz y sorgo son las clases minoritarias con valores similares en cantidad de chacras y superficie, no mayor al 5%.

CULTIVO	CHACRAS	CHACRAS (%)	SUPERFICIE (HA)	SUPERFICIE (%)
SOJA	2.177	92,0	87.673	91,5
MAÍZ	103	4,4	4.744	5,0
SORGO	84	3,6	3.326	3,5
TOTAL	2.364	100	95.743	100

Tabla 5. Resumen por cultivo de la base FUCREA-URF de las zafas 2017-18 y 2018-19.

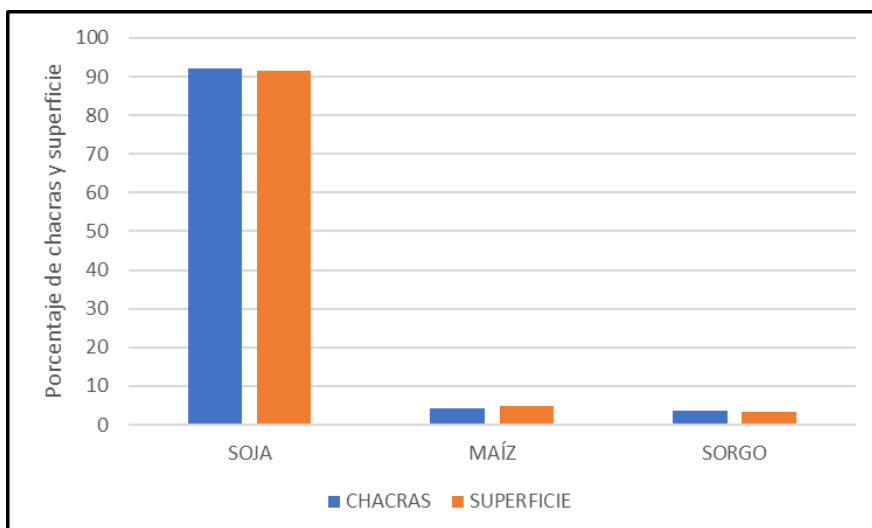


Figura 9. Gráfico de porcentaje de chacras y superficie según cultivo de FUCREA-URF de las zafas 2017-18 y 2018-19.

5.2.3 Mapa Integrado de Cobertura/Uso del Suelo de Uruguay Año 2018

El Mapa Integrado de Cobertura/Uso del Suelo de Uruguay versión 2018 (COBERTURAS) fue el resultado de una integración de diferentes productos de distintas fuentes y años por parte del grupo de trabajo de la Infraestructura de Datos Espaciales del MGAP. El mapa es “integrado” porque se hicieron análisis ad hoc de imágenes satélites y se utilizaron datos georreferenciados de otros estudios realizados con imágenes satelitales de fechas próximas y similar resolución, relevamientos de campo y registros administrativos (Petraglia, Dell’Acqua, Pereira, & Yussim, 2019). Este mapa se encuentra en formato de archivo raster TIF y en este aparecen diferentes coberturas y usos del suelo para Uruguay, como se ve en la figura 10 y tabla 6.

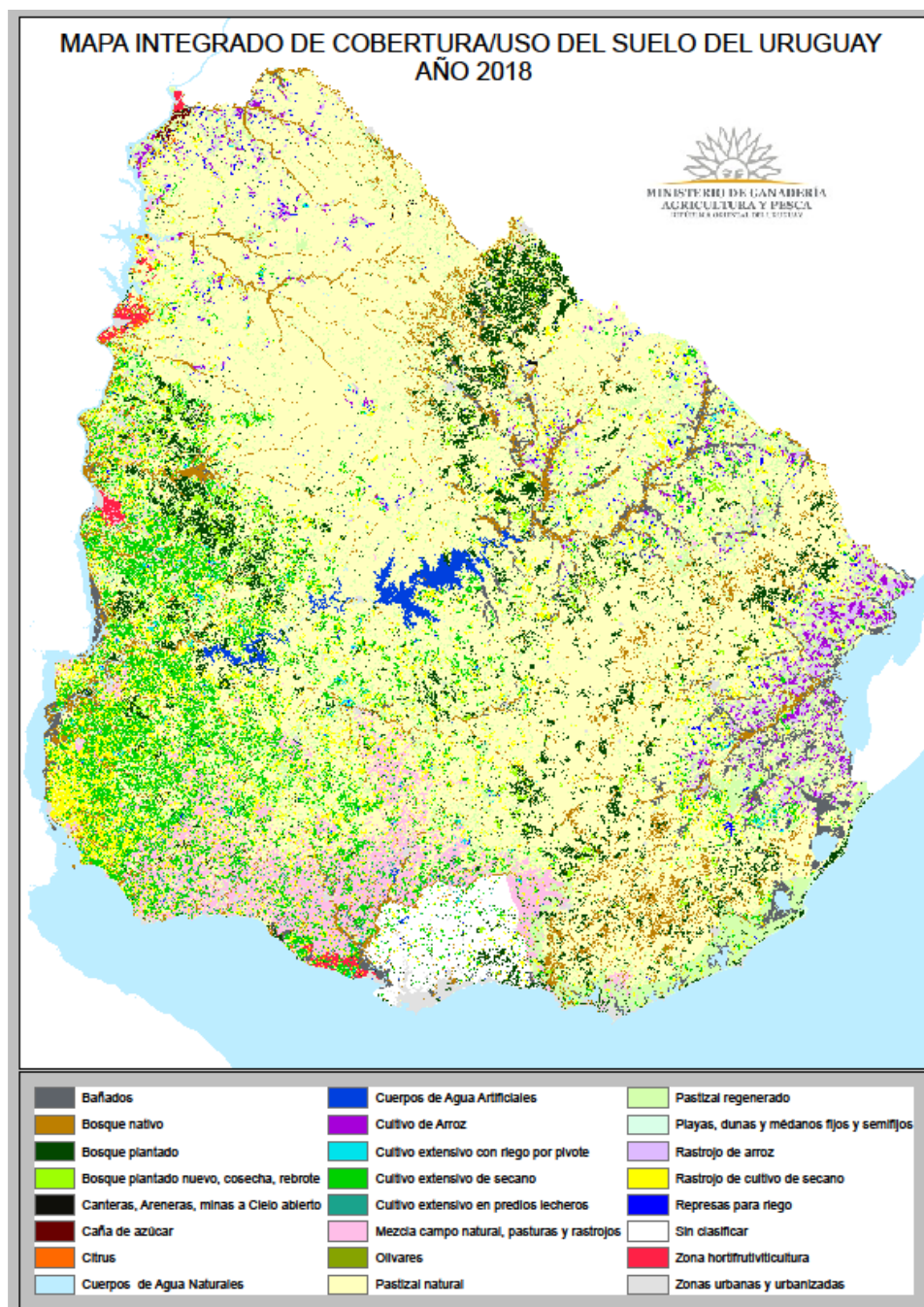


Figura 10. Mapa Integrado de Cobertura/Usos del Suelo del Uruguay Año 2018. Fuente: MGAP-DGRN (2019b).

CLASE	DESCRIPCIÓN
Zonas urbanas y urbanizadas	Área urbana, balnearios, parques, industrias, aeropuertos, campos deportivos, plantas de tratamiento de aguas, instalaciones de transporte. Puede incluir pequeñas áreas cultivos hortifrutícolas.
Canteras, areneras, minas a cielo abierto	ídem clase
Playas, dunas y medanos fijos y semifijos	Todo tipo de Playas y dunas. Los médanos fijos y semifijos tienen cobertura vegetal con diferentes densidades, pueden ser árboles, arbustos y plantas herbáceas.
Cuerpos de agua artificiales	Represas hidroeléctricas y reservas de agua para abastecimiento humano, tajamares y otros reservorios. No incluye las obras para riego que están en la categoría Represas para riego.
Represas para riego	Cuerpos de agua artificiales destinados principalmente a riego
Cuerpos de agua naturales	Ríos, arroyos, lagunas y otros cuerpos de agua naturales.
Bañados	Humedales que presentan inundación o saturación recurrente o sostenida en o cerca de la superficie, suelos con signos de hidromorfismo y la composición del tapiz herbáceo se encuentra representada en algún grado por especies hidrófilas.
Bosque nativo	ídem clase
Bosque plantado	Agrupación de todas las clases y especies de bosques provenientes de la cartografía original utilizada
Bosque plantado nuevo, cosecha, rebrote	Agrupación de las tres categorías provenientes de la cartografía original utilizada.
Pastizal natural	Pastizales naturales con poca modificación antrópica
Pastizal regenerado	Pastos naturales muy modificados, pasturas introducidas y rastrojos viejos.
Caña de azúcar	ídem clase
Citrus	Áreas de Cítricos con destino comercial.
Olivares	Plantaciones de olivos, ocasionalmente puede incluir árboles de otras especies.
Cultivo extensivo de secano	Cultivos de verano de secano principalmente soja, maíz y sorgo.
Cultivo extensivo en predios lecheros	Cultivos anuales de verano maíz y sorgo para alimentación animal.
Rastrojo de cultivo de verano	Cultivos del invierno 2018 y otros cultivos que anteceden.
Cultivo extensivo con riego con pivote	Cultivos extensivos de verano con riego.
Cultivo de arroz	Cultivo de Arroz con riego por inundación.
Rastrojo de arroz	Chacras cosechadas de zafras anteriores.
Zona hortifrutivitucultura	Zonas dedicadas a horticultura, fruticultura y viñedos que también pueden incluir otros usos en suelos descubiertos o con diversas coberturas herbáceas.
Mezcla de campo natural, pasturas y rastrojos viejos	Zonas que incluyen, forrajes, praderas sembradas, campo natural restablecido y rastrojos, con alta variación temporal, en general se asocia a la ganadería lechera.
Sin clasificar	Zonas de Canelones y Montevideo no clasificables con la resolución actual.

Tabla 6. Descripción de clases del mapa Integrado de Cobertura/Uso del Suelo de Uruguay del año 2018.

6. Imágenes satelitales e índices de vegetación

En esta sección se describen las imágenes satelitales utilizadas, sus características, de dónde se obtuvieron, y los índices de vegetación calculados que se utilizaron en este trabajo.

6.1 Imágenes satelitales Sentinel y Terra

Las imágenes satelitales son observaciones de la Tierra que se obtienen desde satélites que orbitan entorno a ésta (figura 11). Estos satélites cuentan con sensores a bordo que miden remotamente la radiación solar electromagnética reflejada (llamada reflectancia) por la superficie terrestre (figura 12). La radiación solar se compone principalmente de las porciones del ultravioleta, espectro visible (azul, verde y rojo), e infrarrojo del espectro electromagnético (figura 13). Los sensores a bordo de los satélites tienen la capacidad de medir la reflectancia de diferentes porciones o bandas del espectro electromagnético, es decir, un sensor mide la reflectancia del verde (500 a 565 nm), otro del rojo (625 a 700 nm), otro del azul (450 a 485 nm), y así otras porciones del espectro como, ultravioleta, infrarrojo, etc.

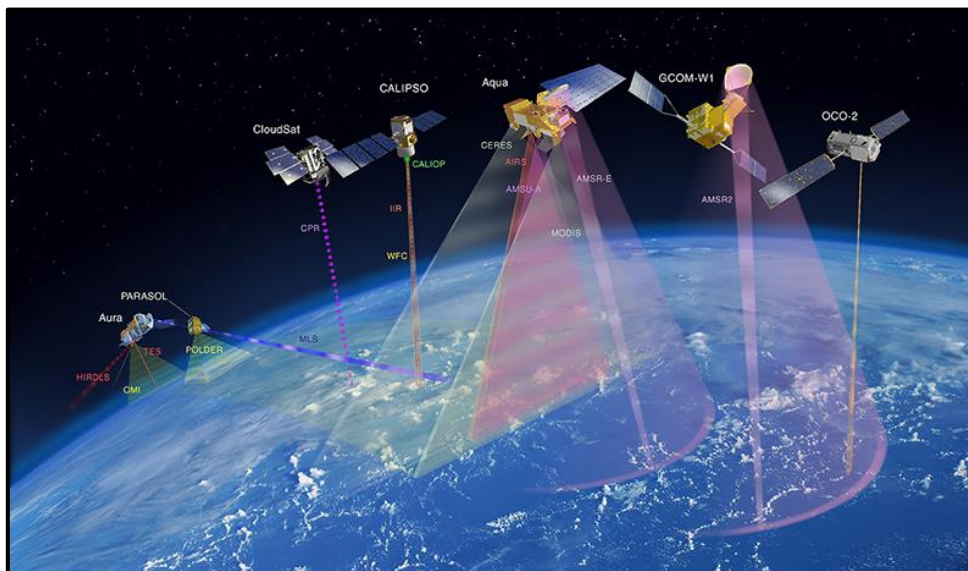


Figura 11. Satélites observando la Tierra. Fuente: https://www.universetoday.com/wp-content/uploads/2013/07/A-Train_Web.jpg

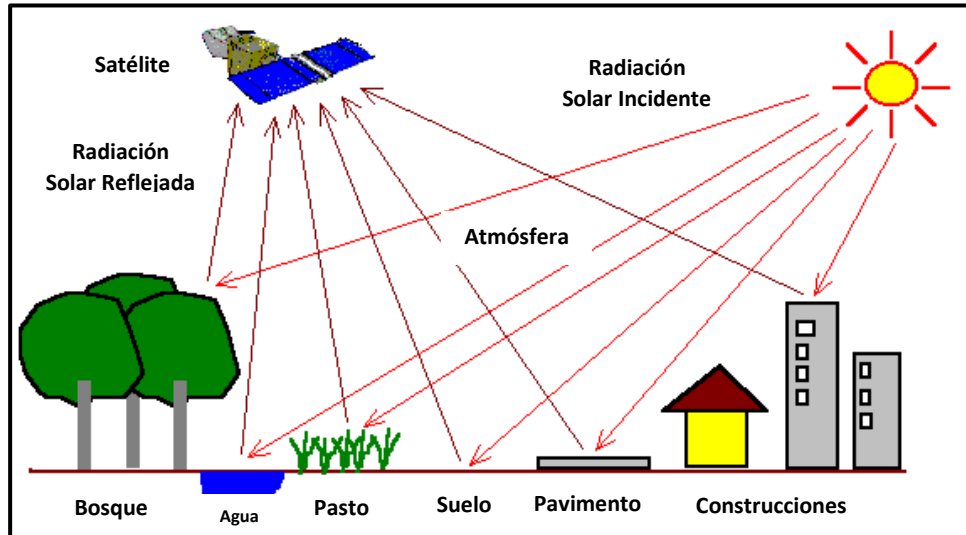


Figura 12. Radiación Solar medida remotamente por satélite. Fuente: <https://crisp.nus.edu.sg/~research/tutorial/optical.htm>

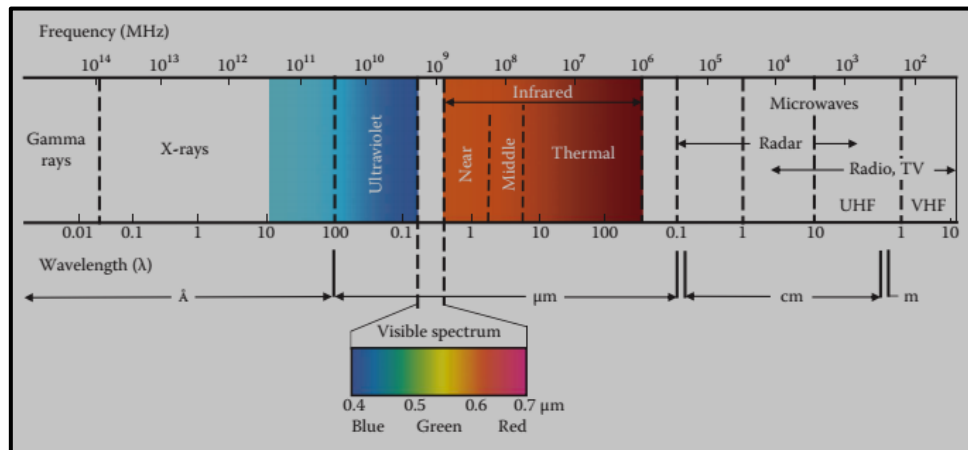


Figura 13. Espectro de radiación electromagnética. Fuente: Chuvieco (2016).

Las imágenes satelitales almacenan digitalmente la reflectancia que es medida por los sensores a bordo de los satélites en una grilla de filas y columnas, en donde la intersección de las filas y columnas es lo que se conoce como pixel (figura 14). El pixel es la unidad mínima que compone una imagen satelital. Para la grilla se conoce la localización geográfica de cada pixel y cada uno almacena la reflectancia terrestre interceptada por éste.

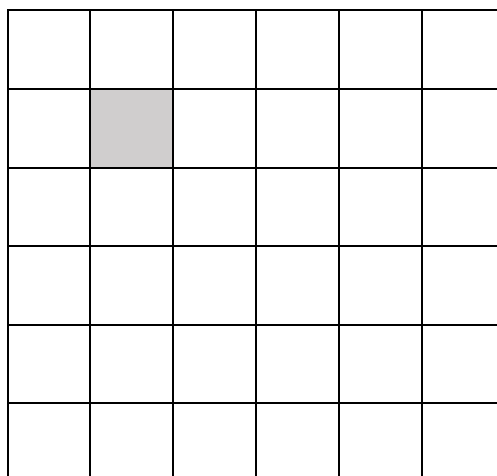


Figura 14. Grilla de imagen satelital. Pixel en gris.

En la figura 15 se ve un ejemplo de una imagen satelital Sentinel 2 y un acercamiento en una zona de la imagen, en donde se pueden ver los pixeles individuales que la componen.

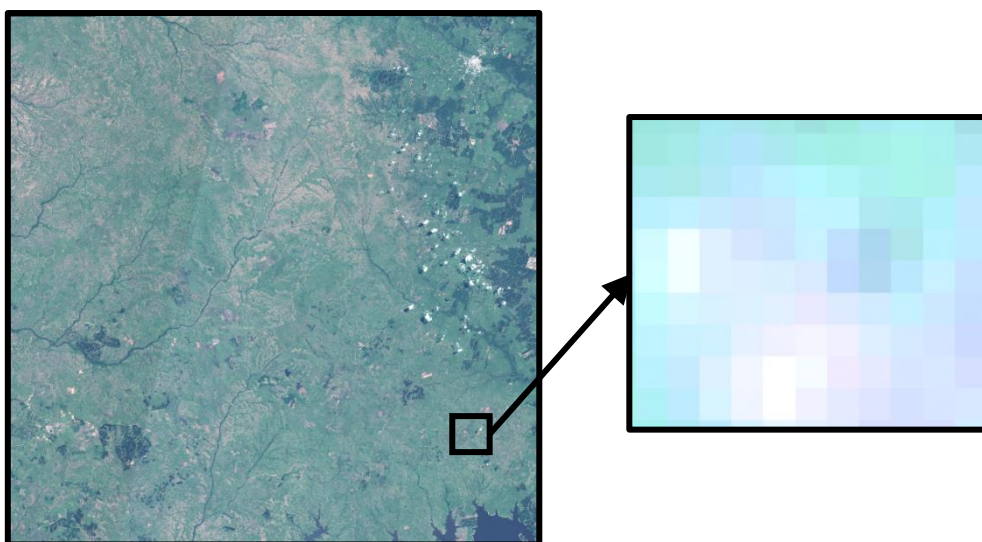


Figura 15. Imagen satelital de Sentinel 2.

Las imágenes satelitales están caracterizadas por diferentes propiedades que son: resolución espacial, resolución espectral y resolución temporal, entre otras. La resolución espacial viene dada por el tamaño del pixel, o más precisamente por el tamaño de la porción de la tierra que éste representa. Cuanto más pequeño es el pixel, mayor es el nivel de detalle de una imagen satelital. Los satélites Landsat 8, Sentinel 2 y Terra tienen un pixel de 30, 10 y 250 metros de lado respectivamente (Figura 16).

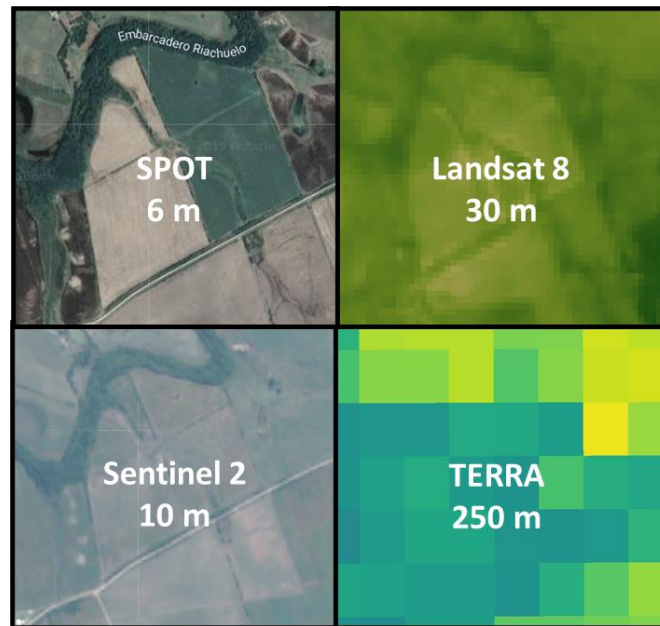


Figura 16. Comparación de tamaño de pixel para satélites Landsat 8, Sentinel 2, SPOT y Terra.

La resolución espectral se refiere a que porciones del espectro electromagnético (conocidas como bandas) pueden medir los sensores a bordo de cada satélite. Sentinel 2 tiene una resolución espectral de 10 bandas del espectro electromagnético. Esto significa que tiene sensores con la capacidad de medir la reflectancia de las siguientes bandas: azul, verde, rojo, infrarrojo cercano, rojo de borde 1 a 4, e infrarrojo de onda corta 1 y 2, como se ve en la tabla 8.

La resolución temporal es el tiempo que necesita el satélite para obtener imágenes de toda la Tierra. Para el caso de Sentinel 2a, este necesita 10 días para cubrir toda la Tierra. Esto quiere decir que para obtener dos imágenes consecutivas de una misma región, lo hace cada 10 días. Para el caso de Terra la resolución temporal es de un día. La resolución espacial y temporal están inversamente relacionadas, cuanto menor es la resolución espacial de un satélite, es decir pixeles más pequeños, imágenes de mayor nivel de detalle, mayor es la resolución temporal, más días necesita para obtener una imagen de una misma región en comparación a un satélite con menor resolución temporal como Terra.

Las imágenes satelitales usadas en este trabajo corresponden a los satélites Terra de NASA, Sentinel 1 (a y b) y 2 (a y b) de la constelación Copernicus de la Agencia Espacial Europea (ESA). En el caso de Terra se usaron específicamente imágenes del producto MOD13Q1 (versión 6 nivel 3 de procesamiento) del sensor Moderate Resolution Imaging Spectroradiometer (MODIS). Las imágenes MOD13Q1 son para períodos de 16 días consecutivos y generadas a una resolución espacial de 250 metros de pixel. El valor para cada pixel es el mejor disponible del período de 16 días según un algoritmo de selección. El criterio usado por el algoritmo es que haya baja presencia de nubes, bajo ángulo de visión y el más alto valor de EVI o NDVI (Didan, Munoz, Solano, & Huete, 2015).

Los satélites Terra y Sentinel 2 son de tipo óptico, y tienen sensores pasivos que miden reflectancia terrestre, mientras que Sentinel 1 es un satélite con sensor de tipo activo, específicamente, el sensor es un radar de apertura sintética (SAR). Los satélites de tipo SAR trabajan en la región del espectro electromagnético de las microondas, y estos funcionan emitiendo una señal o pulso que es interceptada por la superficie terrestre, y el sensor del satélite mide cuánto de esa señal interceptada es retrodispersada (backscattering) de nuevo

hacia la atmosfera (Podest, 2017). Los satélites de tipo óptico tienen el inconveniente de que cuando hay presencia de nubes, estos no pueden captar la reflectancia de la superficie terrestre, sino la reflectancia de las nubes, en cambio, para los satélites de tipo SAR, las nubes son invisibles, ya que las microondas pueden atravesarlas.

En el caso de Sentinel 1 y 2, hay imágenes disponibles desde 2014 y 2015 respectivamente, pero para este trabajo se utilizaron imágenes de 2017 a 2019 inclusive, ya que a partir de 2017 hay continuidad en la disponibilidad de las imágenes; previo a esta fecha, hay periodos donde no están disponibles las imágenes. Para el caso del Sentinel 2 se usaron imágenes Level-1C, las cuales son de reflectancia en el tope de la atmósfera (TOA). En el caso de Terra se trabajó con el período 2013 a 2019 inclusive. Todas las imágenes fueron obtenidas y procesadas en la nube a través de la plataforma de análisis geoespacial Google Earth Engine (GEE). En las tablas 7 a 10 se ven las características de los diferentes satélites mencionados. Así mismo, en las figuras 17 a 19 se ve el área de cobertura (pasadas) de los diferentes satélites.



Figura 17. Mapa de pasadas de los satélites Sentinel 2 sobre Uruguay.

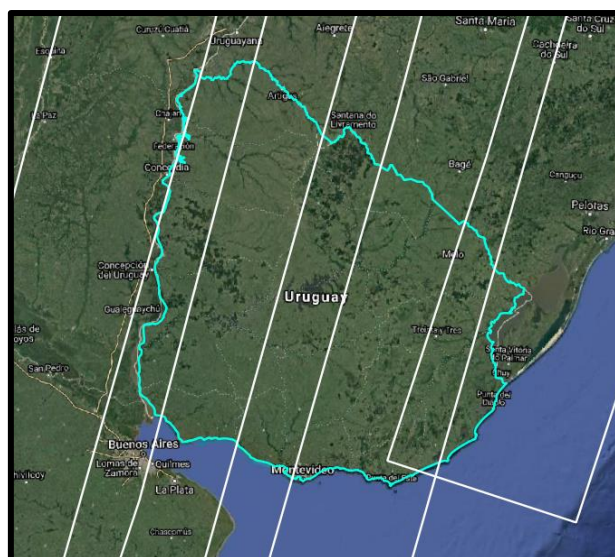


Figura 18. Mapa de pasadas de los satélites Sentinel 1 sobre Uruguay.



Figura 19. Mapa de pasadas del satélite Terra sobre Sudamérica. Fuente: https://modis-land.gsfc.nasa.gov/MODLAND_grid.html

TERRA – MODIS (MOD13Q1)			
PIXEL (metros)	BANDA	LONGITUD DE ONDA (nm)	OBSERVACIONES
250	NDVI		Índice de vegetación
	EVI		
	Rojo	645	Reflectancia de superficie
	Infrarrojo Cercano (NIR)	858	
	Azul	469	
	Infrarrojo Medio (MIR)	2.130 / 2.105 – 2.155	
	QA Detallada		Calidad
	QA Resumen		

Tabla 7. Características del Sensor MODIS a bordo del satélite Terra de NASA.

SENTINEL 2				
PIXEL (metros)	BANDA	LONGITUD DE ONDA (nm)		OBSERVACIONES
		S2-A	S2-B	
60	Aerosoles	443,9	442,3	Tope de la atmósfera (TOA)
10	Azul	496,6	492,1	
	Verde	560	559	
	Rojo	664,5	665	
20	Rojo de Borde 1 (REDGE 1)	703,9	703,8	
	Rojo de borde 2 (REDGE 2)	740,2	739,1	
	Rojo de borde 3 (REDGE 3)	782,5	779,7	
10	Infrarrojo Cercano (NIR)	835,1	833	
20	Rojo de borde 4 (REDGE 4)	864,8	864	
60	Vapor de agua	945	943,2	
	Cirrus	1.373,5	1.376,9	
20	Infrarrojo de onda corta 1 (SWIR 1)	1.613,7	1.610,4	
	Infrarrojo de onda corta 2 (SWIR 2)	2.202,4	2.185,7	
10	QA10			Calidad
20	QA20			
60	QA60			

Tabla 8. Características de los satélites Sentinel 2 (a y b) de la ESA.

SENTINEL 1			
PIXEL (metros)	BANDA POLARIMÉTRICA	LONGITUD DE ONDA (GHz)	OBSERVACIONES
10	HH	5,405	Co-polarización simple, transmisión horizontal / recepción horizontal.
	HV		Banda de polarización cruzada dual, transmisión horizontal / recepción vertical.
	VV		Co-polarización simple, transmisión vertical / recepción vertical.
	VH		Banda de polarización cruzada dual, transmisión vertical / recepción vertical.

Tabla 9. Características de los satélites Sentinel 1 (a y b) de la ESA.

	TERRA - MODIS	SENTINEL 2 (A & B)	SENTINEL 1 (A & B)
TIPO	ÓPTICO	ÓPTICO	RADAR
TAMAÑO DEL PIXEL (m)	250	10, 20, 60	10
TAMAÑO DE LA ESCENA (km)	1.200 x 1.200	100 x 100	Hasta 400 km de swath.
PERÍODO DE REVISITA (DIAS)	1	10 (CADA UNO) 5 (COMBINANDO 2A & 2B)	12 (CADA UNO) 6 (COMBINANDO 1A & 1B)
LANZAMIENTO	2.000	Sentinel 2A (2015) Sentinel 2B (2017)	Sentinel 1A (2014) Sentinel 1B (2016)

Tabla 10. Características principales de los satélites utilizados.

6.2 Índices de vegetación y cross-ratio VH/VV

En este trabajo, para las imágenes de los satélites ópticos (Terra y Sentinel 2) se usaron los siguientes índices de vegetación: Normalized Difference Vegetation Index (NDVI), Enhanced Vegetation Index (EVI) y Enhanced Vegetation Index 2 (EVI2) (Didan et al., 2015). Estos índices están relacionados directamente a la presencia de clorofila y la actividad fotosintética en las plantas. Para poder realizar la fotosíntesis éstas necesitan absorber energía de las bandas azul y rojo del espectro visible. Una planta sana, con buen desarrollo vegetal, y en activo crecimiento absorbe mucha luz azul y roja, y refleja mucho infrarrojo, mientras que plantas enfermas o estresadas, por el contrario, absorben menos azul y rojo, y reflejan menos infrarrojo, como se ve en la figura 20.

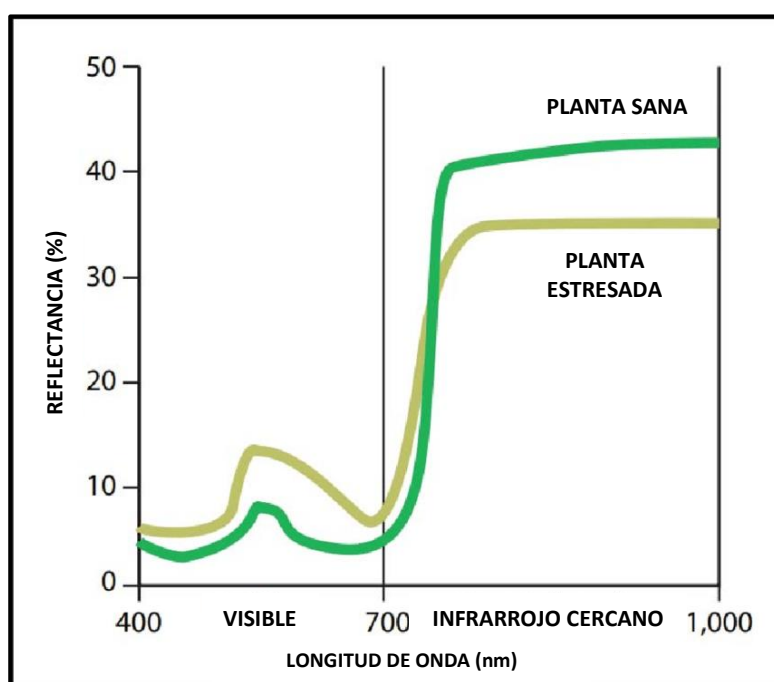


Figura 20. Gráfico de reflectancia de las plantas en el espectro electromagnético. Fuente: Niekerk et al. (2019).

Sobre este principio se calculan estos índices y se obtienen a partir de cocientes entre los valores de reflectancia de las bandas del rojo, azul, e infrarrojo cercano (NIR). El rango de valores para éstos está entre -1 a 1. Valores altos, cercanos a 1, indican una fuerte actividad fotosintética y, por lo tanto, presencia de coberturas vegetales. Valores bajos, cercanos a 0, indican muy poca actividad fotosintética, lo que puede deberse a presencia de coberturas vegetales senescentes, u otro tipo de coberturas, como espejos de agua, áreas urbanas, suelo desnudo, entre otras.

Para NDVI la fórmula de cálculo es la siguiente:

$$NDVI = \frac{(NIR - ROJO)}{(NIR + ROJO)}.$$

El índice EVI se obtiene a partir de:

$$EVI = G \times \frac{(NIR - ROJO)}{(NIR + C1 \times ROJO - C2 \times AZUL + L)},$$

donde C1, C2, G y L son coeficientes con valores de 6, 7,5, 2,5 y 1 respectivamente. El EVI, a diferencia del NDVI, tiene mejor sensibilidad cuando hay alta biomasa, y es mejor para el monitoreo de la vegetación al desacoplar la influencia del suelo y la atmósfera de la señal de la vegetación (Didan et al., 2015).

El índice EVI2 se calcula a partir de las bandas del rojo e infrarrojo cercano y un par de coeficientes, sin necesidad de usar la banda azul, como se ve en la fórmula:

$$EVI2 = \frac{2,5 \times (NIR - ROJO)}{NIR + 2,4 \times ROJO + 1}.$$

Para el caso de las imágenes MOD13Q1 éstas ya vienen con las bandas de los índices EVI y NDVI incorporadas, en cambio, para Sentinel 2 no, por lo que se calculó el NDVI, y se optó por calcular el índice EVI2 en vez de EVI, ya que se encontró para este último que los valores del índice se salían de la escala -1 a 1. Estos valores fuera de escala podrían estar explicados por el uso de la banda azul. EVI tiene un comportamiento errático sobre objetivos brillantes (nubes, nieve y hielo). Esto se debe a que la banda azul satura (Didan et al., 2015).

Para Sentinel 1, a partir de las bandas polarimétricas VH y VV, se calculó el cross-ratio VH/VV como:

$$CROSSRATIO\ VH/VV = VH - VV.$$

Este índice ha demostrado ser bueno para distinguir diferentes densidades de vegetación y tener altas correlaciones con medidas realizadas in situ de NDVI, índice de área foliar y contenido de agua de la vegetación para diferentes cultivos. Cross-ratio VH/VV ha demostrado tener potencial para monitoreo de vegetación y fenología (Vreugdenhil et al., 2018).

7. Procesamiento de las bases

En esta sección se explica el procesamiento que se realizó sobre las bases de COBERTURAS, FUCREA-URF y PLANES, previo a ser cruzadas con las imágenes satelitales de Sentinel y Terra para extraer series temporales de las diferentes bandas e índices de estos.

7.1 PLANES

Del total de los 16.678 planes para cada zafra agrícola de verano desde 2013-14 a 2018-19 inclusive, se seleccionaron solo aquellos que tuvieran asignada una etiqueta de cultivo y se dejaron fuera los que tuvieran el registro en blanco como se ve en el ejemplo de la figura 21. En la tabla 11 se ve la cantidad de planes extraídos por zafra, superficie total, superficie por cultivo y el porcentaje que ésta representa del total.

PLANNRO	OI_2013	PV_2013	OI_2014	
1	Cereales de Invierno	Soja	Barbecho	Sorgo o M
2		Soja	Cereales de Invierno	Soja
3				
4	Cereales de Invierno	Soja	Cultivo de Cobertura	Soja
5		Soja	Pastura no Consociada	Pastura E
6	Cereales de Invierno	Sorgo o Maíz	Barbecho	Soja
7	Cereales de Invierno	Soja	Cereales de Invierno	Soja
8	Cereales de Invierno	Soja	Cereales de Invierno	Soja
9	Cereales de Invierno	Soja	Cereales de Invierno	Soja
10	Cereales de Invierno	Sorgo o Maíz	Pastura no Consociada	Soja
11		Soja	Cebada	Soja
12	Cereales de Invierno	Soja	Barbecho	Soja
13				Soja
14				Soja
15	Cereales de Invierno	Sorgo o Maíz	Barbecho	Soja
16		Soja	Cereales de Invierno	Soja
17	Cereales de Invierno	Soja	Cultivo de Cobertura	Soja
18	Cereales de Invierno	Sorgo o Maíz	Cereales de Invierno	Soja
19	Cereales de Invierno	Sorgo o Maíz	Cereales de Invierno	Soja
20	Cereales de Invierno	Sorgo o Maíz	Cereales de Invierno	Soja
21	Cereales de Invierno	Soja	Cereales de Invierno	Soja
22		Soja	Cereales de Invierno	Soja
23				

Figura 21. Ejemplo de registros con y sin etiqueta de cultivo asignada de PLANES.

ZAFRA	PLANES CON ETIQUETA	SUPERFICIE (HA)					
		TOTAL	MAÍZ	SOJA	SORGO	PASTURAS	SORGO-MAÍZ
2013-14	11.464	1.144.810	9.411 (0,82%)	981.057 (85,7%)	9.655 (0,84%)	9.127 (0,80%)	131.236 (11,46%)
2014-15	12.665	1.227.280	11.183 (0,91%)	1.001.211 (81,58%)	11.871 (0,97%)	49.997 (4,07%)	145.908 (11,89%)
2015-16	12.793	1.223.932	11.510 (0,94%)	886.494 (72,43%)	11.458 (0,94%)	136.839 (11,18%)	165.927 (13,56%)
2016-17	12.619	1.176.911	8.433 (0,72%)	790.103 (67,13%)	10.481 (0,89%)	211.262 (17,95%)	147.144 (12,50%)
2017-18	12.074	1.120.843	11.120 (0,99%)	749.026 (66,83%)	10.346 (0,92%)	226.824 (20,24%)	115.260 (10,28%)
2018-19	11.047	1.009.869	14.129 (1,40%)	677.969 (67,13%)	13.065 (1,29%)	204.464 (20,25%)	94.960 (9,40%)
TOTAL	72.662						

Tabla 11. Cantidad y superficie (hectáreas) de planes con etiqueta de cultivo según zafra.

Los planes seleccionados para cada zafra fueron procesados de dos maneras diferentes según si las series temporales de datos a extraer fueran de los satélites Terra o los Sentinel (ver figura 22). En el caso de Terra, como hay imágenes disponibles desde el año 2000, se extrajeron series temporales para las seis zafas. En el caso de los Sentinel, como se mencionó en la sección 6.1, es a partir del año 2017 que hay una continuidad en las imágenes, y por lo tanto se trabajó con las zafas 2017-18 y 2018-19. Por otro lado, en el caso de Terra los píxeles son de 250 metros de lado, por lo cual, para cada plan se extrajeron series temporales solo de los píxeles enteros que estuvieran dentro de cada plan como se ve en la figura 22. Esto se hizo así para trabajar con píxeles cuya cobertura vegetal fuera homogénea, lo que se dice píxeles puros, y no una mezcla

heterogénea de diferentes coberturas vegetales. En la tabla 12 se ve la cantidad de planes y píxeles enteros extraídos por zafra para Terra.

Para el caso de los Sentinel, según la banda el tamaño del pixel es de 10 o 20 metros. En este caso, lo que se hizo fue generar para cada plan una grilla de puntos regulares y extraer series temporales por punto. Como se observa en la tabla 11, cualquiera sea la zafra, la superficie total de los planes con etiquetas supera el millón de hectáreas, y además si se observa la distribución del área entre cultivos, ésta está muy desbalanceada, siendo soja la clase mayoritaria, con porcentajes mayores al 66%, seguida por pasturas con valores superiores al 20% para las últimas zafras, sorgo-maíz con valores superiores al 9%, y las clases minoritarias de maíz y sorgo con valores entorno al 1%.

Dado que el área total de cada zafra es superior a un millón de hectáreas, y considerando el desbalanceo de las clases, se buscó un criterio para extraer una cantidad de puntos que fuera manejable y al mismo tiempo que las clases estuvieran balanceadas. Por ejemplo, extraer un punto por hectárea tendría el inconveniente de tener aproximadamente una base de series temporales de un millón de registros, lo cual sería inmanejable. Entonces, el criterio de extracción de puntos fue extraer 1 punto por hectárea para todos los planes seleccionados de los cultivos minoritarios de maíz y sorgo, y para el resto de los cultivos, extraer puntos de todos los planes de esos cultivos a diferentes densidades, es decir distancia entre puntos (figura 23), de manera de tener la misma cantidad de puntos que los cultivos minoritarios. En la tabla 13 se muestra la cantidad de puntos extraídos por cultivo y zafra.

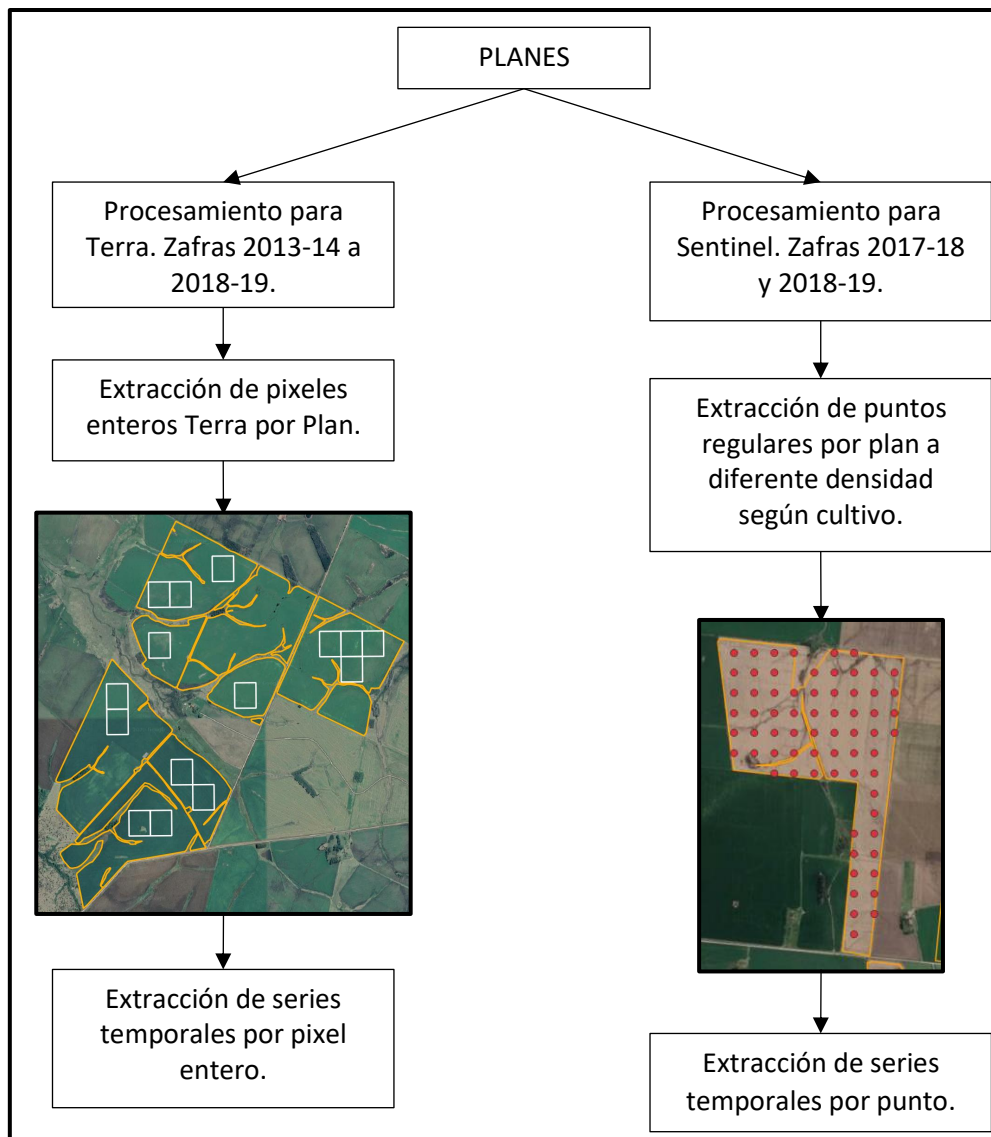


Figura 22. Flujograma de extracción de series temporales según Sentinel o Terra para PLANES.

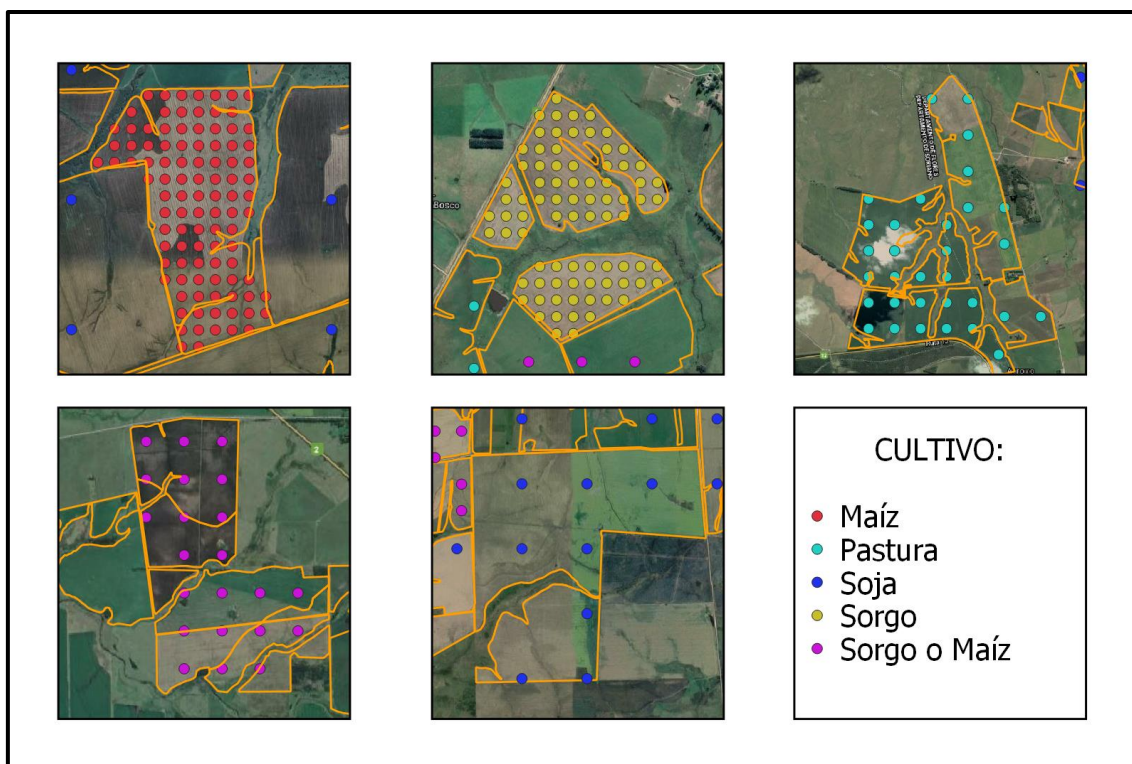


Figura 23. Puntos regulares a diferentes densidades según cultivo para extraer series temporales de Sentinel para PLANES.

ZAFRA	SERIES TEMPORALES	
	PLANES	PIXELES ENTEROS
2013-14	7.125	36.961
2014-15	7.709	39.053
2015-16	7.671	38.376
2016-17	7.511	36.230
2017-18	7.164	34.507
2018-19	6.531	30.269
TOTAL	43.711	215.396

Tabla 12. Cantidad de planes y pixeles enteros de Terra extraídos para las zafras 2013-14 a 2018-19.

CULTIVO	2017-18			2018-19		
	PLANES	HECTÁREAS	PUNTOS	PLANES	HECTÁREAS	PUNTOS
MAÍZ	121	11.125	11.164	177	14.135	14.165
PASTURA	2.649	221.248	11.100	2.492	201.424	14.268
SOJA	5.528	681.794	12.530	5.454	637.844	14.392
SORGO	192	10.352	10.395	232	13.071	13.073
SORGO-MAÍZ	1.216	115.151	11.558	1.085	94.964	13.972
TOTAL	9.706	1.039.670	56.747	9.440	961.438	69.870

Tabla 13. Cantidad de planes, hectáreas y puntos por cultivo de las zafras 2017-18 y 2018-19.

Tanto los píxeles enteros Terra seleccionados para todas las zafas, como las grillas de puntos regulares de las zafas 2017-18 y 2018-19 fueron subidas en formato Shapefile a la plataforma en la nube GEE para extraer las series temporales de las imágenes satelitales a ser procesadas posteriormente.

7.2 FUCREA-URF

Las bases de FUCREA y URF se unificaron en una base conjunta nombrada FUCREA-URF, y a partir de esta se generaron dos archivos en formato shapefile para las zafas 2017-18 y 2018-19. Para cada archivo shapefile se generó una grilla de puntos regulares cada una hectárea como se ve en la figura 24. La razón por la cual se generó esa grilla de puntos regulares fue aumentar los datos disponibles con información de cultivos para la etapa posterior de entrenamiento de modelos de clasificación. Para cada uno de los puntos de la grilla se tiene valor de ID, código de chacra y cultivo, entre otras variables. Los archivos de las chacras y los puntos regulares fueron subidos a la plataforma GEE para la posterior extracción de series temporales de los satélites Sentinel (1 y 2). Para el caso de los satélites Sentinel hay imágenes satelitales continuas a partir de la zafa 2017-18. En la tabla 14 se ve el resumen de la cantidad de puntos por cultivo y zafa. Como se ve, los datos están muy desbalanceados, con más del 90% de los datos correspondiendo a soja.



Figura 24. Chacras con grilla de puntos regulares de FUCREA-URF.

ZAFRA	CULTIVO	PUNTOS	
		CANTIDAD	PORCENTAJE
2017-18	SOJA	45.403	47,5
2018-19		42.274	44,2
SUBTOTAL		87.677	91,6
2018-19	MAÍZ	2.579	2,7
2017-18		2.074	2,2
SUBTOTAL		4.653	4,9
2018-19	SORGO	2.273	2,4
2017-18		1.065	1,1
SUBTOTAL		3.338	3,5
TOTAL		95.668	100

Tabla 14. Cantidad y porcentaje de puntos regulares por cultivo y zafa en FUCREA-URF de las zafas 2017-18 y 2018-19.

7.3 COBERTURAS

Para el mapa de COBERTURAS, primero se agruparon de manera jerárquica las 24 clases originales en 10 clases genéricas y posteriormente estas clases genéricas en dos grandes clases, agrícola y no agrícola, como se ve en la tabla 15. La clase agrícola abarca cultivos del tipo agrícola como cultivos de secano (maíz, soja, sorgo, girasol), arroz, caña de azúcar, rastrojo de arroz, entre otros. Mientras que la clase no agrícola, comprende otros usos y coberturas del suelo como cuerpos de agua, bañados, forestación, zonas urbanas, costas, entre otros. En el mapa de COBERTURAS existe la clase sin clasificar. Ésta corresponde a un área del país que representa menos del 2,6% del área del total, que por su complejidad en cuanto al tamaño de los predios agropecuarios y diversidad de usos y coberturas, los autores del mapa decidieron no clasificar. Esta clase fue descartada en este trabajo. Posteriormente se generó una grilla de puntos regulares cada 1 kilómetro para todo el país, con un total de 176.000 puntos. Esta grilla se intersectó con el mapa de COBERTURAS y se seleccionaron aquellos puntos cuya gran clase fuera no agrícola. De los 149.000 puntos no agrícola de manera aleatoria y estratificada por clase, se seleccionaron un total de 11.080 puntos (7,4% del total) como se ve en la tabla 16 y figura 25. En estos puntos están presentes todas las clases cuya gran clase es no agrícola.

CLASE	CLASE GENÉRICA	GRAN CLASE
Caña de azúcar	Agrícola	Agrícola
Cultivo de arroz		
Cultivo extensivo con riego con pivote		
Cultivo extensivo de secano		
Cultivo extensivo en predios lecheros		
Rastrojo de arroz		
Rastrojo de cultivo de verano		
Cuerpos de agua artificiales	Agua	No agrícola
Cuerpos de agua naturales		
Represas para riego		
Playas, dunas y médanos fijos y semifijos	Arena	
Bañados	Bañados	
Bosque nativo	Bosque	
Bosque plantado		
Bosque plantado nuevo, cosecha, rebrote		
Canteras, areneras, minas a cielo abierto	Minero	
Citrus	Otros	
Olivares		
Zona hortifrutivitucultura		
Mezcla de campo natural, pasturas y rastrojos viejos	Recursos Forrajeros	
Pastizal natural		
Pastizal regenerado		
Zonas urbanas y urbanizadas	Urbano	
Sin clasificar	Sin clasificar	Sin clasificar

Tabla 15. Agrupamiento jerárquico de clases de COBERTURAS.

CLASE GENÉRICA	CLASE	TOTAL
AGUA	Cuerpos de agua artificiales	466
	Cuerpos de agua naturales	295
	Represas para riego	239
	SUBOTAL	1.000
ARENA	Playas, dunas y médanos fijos y semifijos	225
BAÑADOS	Bañados	1.000
BOSQUE	Bosque nativo	858
	Bosque plantado	951
	Bosque plantado nuevo, cosecha, rebrote	191
	SUBOTAL	2.000
MINERO	Canteras, areneras, minas a cielo abierto	39
OTROS	Citrus	181
	Olivares	116
	Zona hortifrutivicultura	519
	SUBOTAL	816
RECURSOS FORRAJEROS	Mezcla de campo natural, pasturas y rastrojos viejos	288
	Pastizal natural	3.817
	Pastizal regenerado	895
	SUBOTAL	5.000
URBANO	Zonas urbanas y urbanizadas	1.000
TOTAL		11.080

Tabla 16. Cantidad de puntos filtrados por clase de COBERTURAS.

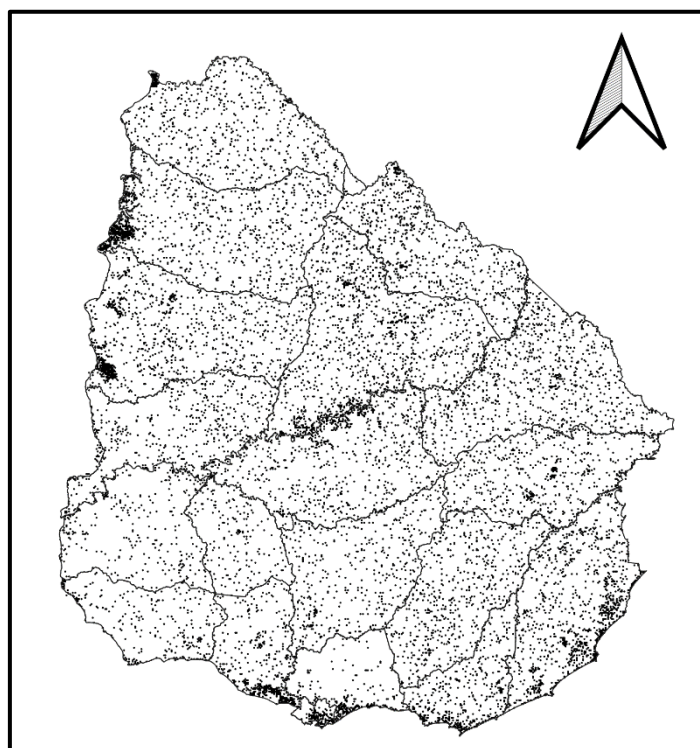


Figura 25. Mapa de distribución de puntos no agrícola seleccionados al azar de COBERTURAS.

8. Extracción de series temporales satelitales

En esta sección se explica cuál fue el procesamiento que se aplicó sobre las imágenes satelitales de Terra y Sentinel (1 y 2), y como éstas se cruzaron con las bases procesadas para extraer series temporales de las distintas bandas e índices. Posteriormente esos datos se usaron en la etapa de entrenamiento de algoritmos no supervisados y supervisados. Como se está trabajando con cultivos de verano, y estos se siembran en primavera y cosechan en otoño, el período analizado con estas imágenes para cada zafra va desde setiembre del año correspondiente a mayo del año siguiente.

8.1 Series temporales de Terra

En GEE se trabajó con imágenes del satélite Terra, específicamente se usó el producto MOD13Q1 del sensor MODIS. El período analizado con estas imágenes va desde el 30 de setiembre del año correspondiente hasta el 24 de mayo del año siguiente, es decir, que si queremos analizar la zafra de verano 2013-14, el período comienza el 30/09/2013 y termina el 24/05/2014. En la tabla 17 se ve el período temporal que abarcan las imágenes MOD13Q1. Como las imágenes abarcan períodos de 16 días, en un año hay 23 imágenes, y cada imagen abarca siempre el mismo período temporal independientemente del año.

IMAGEN MOD13Q1	PERÍODO
257	30/09 a 15/10
273	16/10 a 31/10
289	01/11 a 16/11
305	17/11 a 02/12
321	03/12 a 18/12
337	19/12 a 03/01
353	01/01 a 16/01
001	17/01 a 01/02
017	02/02 a 17/02
033	18/02 a 05/03
049	06/03 a 21/03
065	22/03 a 06/04
081	07/04 a 22/04
097	23/04 a 08/05
113	09/05 a 24/05

Tabla 17. Período que abarca cada imagen MOD13Q1 para el período analizado.

Para el período analizado, en GEE se generó lo que en sensoramiento remoto se conoce como apilado de imágenes, también llamado cubo, stack o volumen. Este apilado consiste en obtener cada una de las imágenes de EVI del período analizado ordenadas cronológicamente desde la más antigua hasta la más reciente, como se ve en el ejemplo de la figura 26. En el caso de este trabajo los apilados están compuestos de 15 imágenes, cada una de las que aparece en la tabla 17.

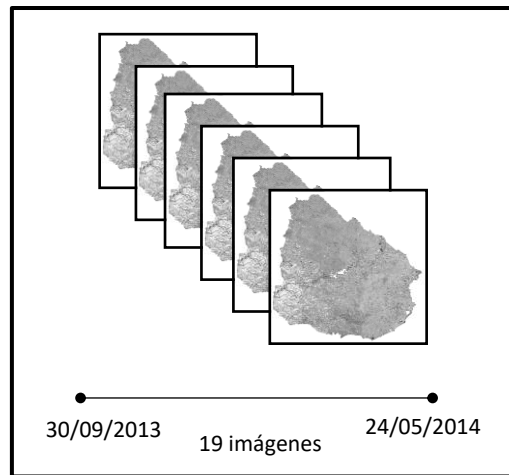


Figura 26. Apilado de imágenes MOD13Q1 para zafra 2013-14.

Una vez generado el apilado de EVI, éste se intersectó con los shapefiles de PLANES que fueron procesados y seleccionados para cada una de las zafras en la sección 7.1. Para cada plan se seleccionaron aquellos píxeles intersectados del apilado de EVI que estuvieran completamente dentro de los polígonos del plan, como se ve en la Figura 27. Esto se hizo a fines de evitar trabajar con píxeles cuya área esté parte dentro y parte fuera de la chacra, y por lo tanto no son píxeles puros, sino una mezcla de coberturas vegetales heterogéneas. De los píxeles enteros dentro de un plan, se extrajeron series temporales de EVI y se calculó la serie temporal promedio del plan como se ve en la figura 28.

En GEE se generaron los scripts para generar los apilados de EVI, la intersección con los planes y la extracción de las series temporales en archivos de formato CSV. Finalmente, en GEE se descargó para cada zafra dos archivos en formato CSV, uno que contiene las series temporales de EVI de los píxeles individuales enteros dentro de cada plan, y otro con la serie temporal promedio de todos los píxeles enteros dentro de cada plan. Por ejemplo, para el plan de la figura 27, éste está formado por 15 series temporales individuales correspondiente a los píxeles enteros en color blanco, para los cuales se calcula la serie temporal promedio como se ve en la figura 28.



Figura 27. Extracción de píxeles Terra - MOD13Q1 completamente dentro de las chacras de un plan.

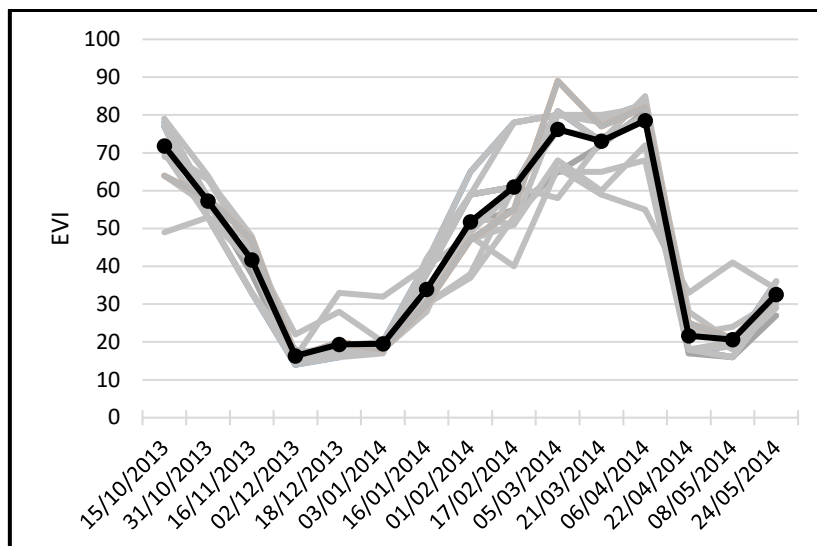


Figura 28. Gráfico de series temporales individuales (gris) y promedio (negro) de EVI de un plan para zafra 2013-14.

En la tabla 18 se ve la cantidad de series temporales extraídas de pixeles enteros y de planes por zafra para los cultivos principales (soja, pasturas, maíz-sorgo, maíz, sorgo).

ZAFRA	SERIES TEMPORALES	
	PLANES	PIXELES ENTEROS
2013-14	7.125	36.961
2014-15	7.709	39.053
2015-16	7.671	38.376
2016-17	7.511	36.230
2017-18	7.164	34.507
2018-19	6.531	30.269
TOTAL	43.711	215.396

Tabla 18. Cantidad de series temporales extraídas de pixeles enteros y de planes para las zafras 2013-14 a 2018-19.

En la figura 29 se muestra el flujograma de extracción de las series temporales de EVI para PLANES en GEE.

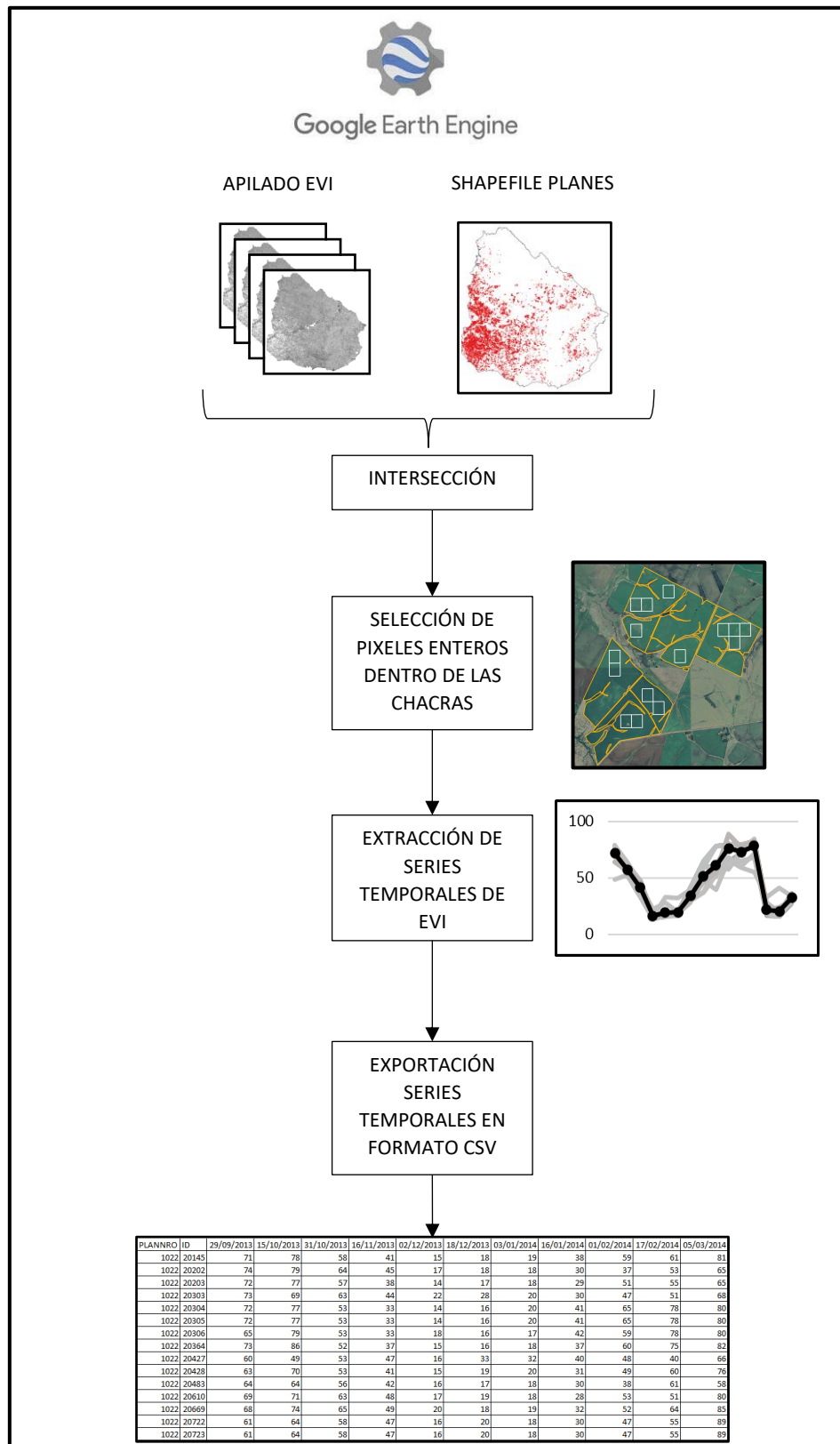


Figura 29. Flujograma de extracción de series temporales de EVI en GEE para PLANES.

8.2 Series temporales de Sentinel

En esta sección se explica el proceso de extracción de las series temporales de los satélites Sentinel 1 y 2 para las bases en la plataforma GEE. Para los Sentinel se usaron las bandas que se especifican en la tabla 19.

SENTINEL	
1	2
BANDA	
VH	AZUL
VV	EVI2
	Infrarrojo cercano (NIR)
	Infrarrojo de onda corta 1 (SWIR 1)
	Infrarrojo de onda corta 2 (SWIR 2)
	NDVI
	ROJO
	Rojo de borde 1 (REDGE 1)
	Rojo de borde 2 (REDGE 2)
	Rojo de borde 3 (REDGE 3)
	Rojo de borde 4 (REDGE 4)
	VERDE
	CALIDAD (QA60)

Tabla 19. Bandas utilizadas de Sentinel 1 y 2.

El período analizado con las imágenes de Sentinel es el mismo que para las imágenes Terra, y está comprendido entre el 30 de setiembre del año correspondiente hasta el 24 de mayo del año siguiente; por ejemplo, si estamos trabajando con la zafra 2018-19, el período va desde el 30/09/2018 hasta 24/05/2019. Definido el período a analizar, el primer paso fue construir, al igual que en el caso de las imágenes Terra, apilados de imágenes entre las fechas especificadas para extraer las series temporales para cada banda de la tabla 19. Para generar este apilado de imágenes primero se generaron imágenes sintéticas de Sentinel 1 y 2 que abarcaran el mismo periodo que el de las imágenes Terra utilizadas. Esto se hizo así, pensando a futuro que si se quisieran comparar los datos de Sentinel contra los datos de Terra, estos estuvieran en el mismo marco temporal.

En la tabla 20 vemos el período que abarca cada imagen sintética de Sentinel 1 y 2. Una imagen satelital sintética, es una imagen que se construye a partir de dos o más imágenes para la misma área, pero de distintas fechas y en donde el valor de cada pixel de la imagen sintética se obtiene como promedio, mediana, mínimo o máximo de los valores de ese pixel para las distintas fechas. Para generar cada imagen sintética de Sentinel dentro del período correspondiente de acuerdo con la tabla 20, hay que tomar todas las imágenes Sentinel que haya para ese período. En el caso de las imágenes de satélites de tipo SAR como Sentinel 1, como resultado de diferentes respuestas de retrodispersión de objetos en la superficie terrestre con la señal del satélite, las imágenes se ven granuladas. Este ruido en la imagen se conoce como ruido “sal y pimienta” o speckle (Giri Ananto et al., 2019). La presencia del ruido speckle se debe minimizar a través del aplicado de un filtro en la imagen. En el caso de Sentinel 1, cada imagen en el período comprendido es filtrada usando un filtro de mediana de tamaño 3x3. Luego para cada pixel se calcula el promedio de todas las imágenes dentro del período, obteniéndose una imagen

sintética de acuerdo a como se ve en la figura 30. Para el caso de las imágenes de Sentinel 2, estas primero se filtran usando la banda de calidad QA60 para eliminar los pixeles con nubes y sustituir su valor por NA, y luego se seleccionan los pixeles que tengan el valor más alto de todas las imágenes dentro del período, obteniéndose una imagen sintética de máximo valor, como se ve en la figura 31.

IMAGEN SINTÉTICA	INICIO	FIN
1	30-set	15-oct
2	16-oct	31-oct
3	01-nov	16-nov
4	17-nov	02-dic
5	03-dic	18-dic
6	19-dic	03-ene
7	01-ene	16-ene
8	17-ene	01-feb
9	02-feb	17-feb
10	18-feb	05-mar
11	06-mar	21-mar
12	22-mar	06-abr
13	07-abr	22-abr
14	23-abr	08-may
15	09-may	24-may

Tabla 20. Período abarcado por las imágenes sintéticas de Sentinel 1 y 2.

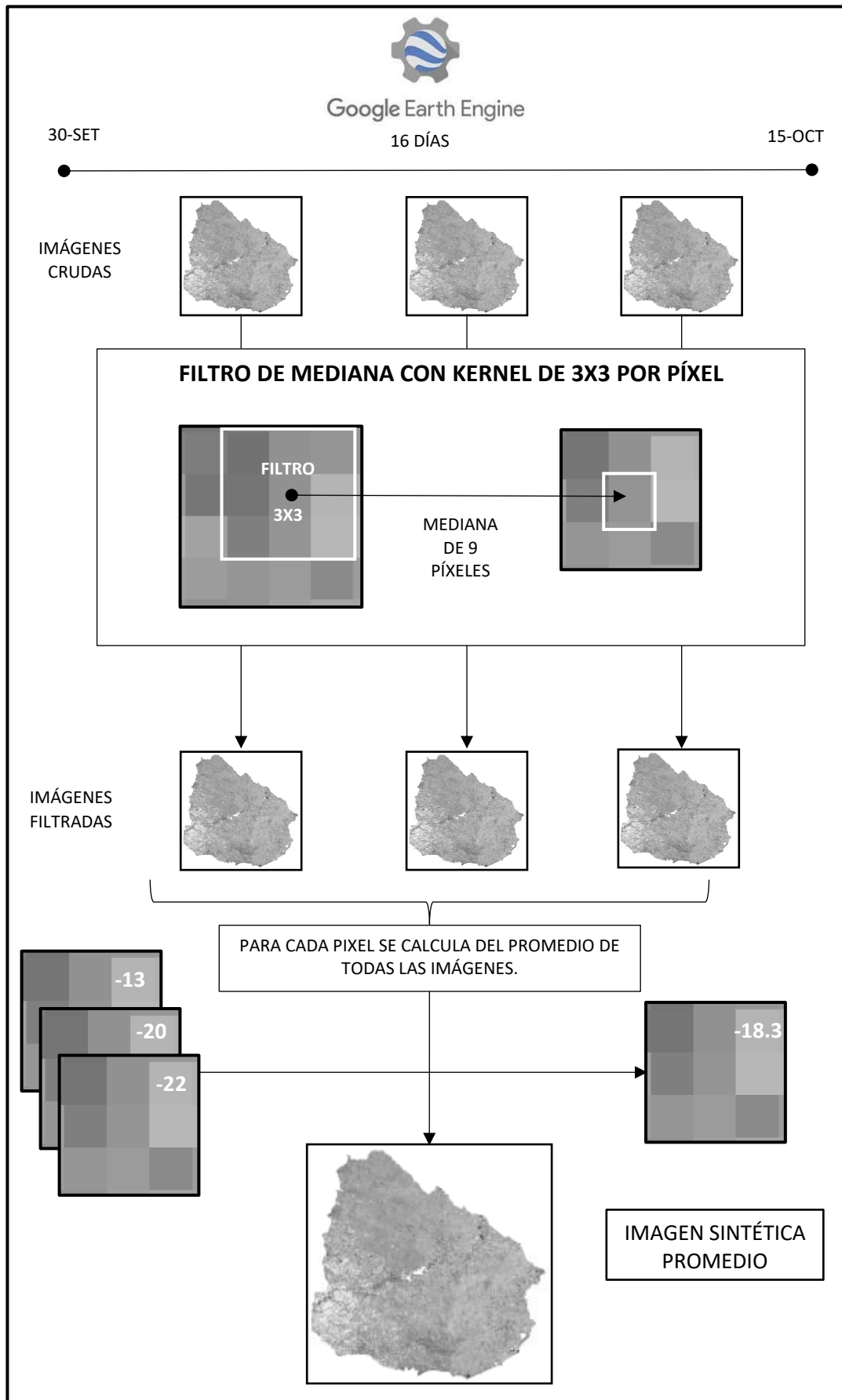


Figura 30. Flujograma de generación de imágenes sintéticas Sentinel 1 en GEE.

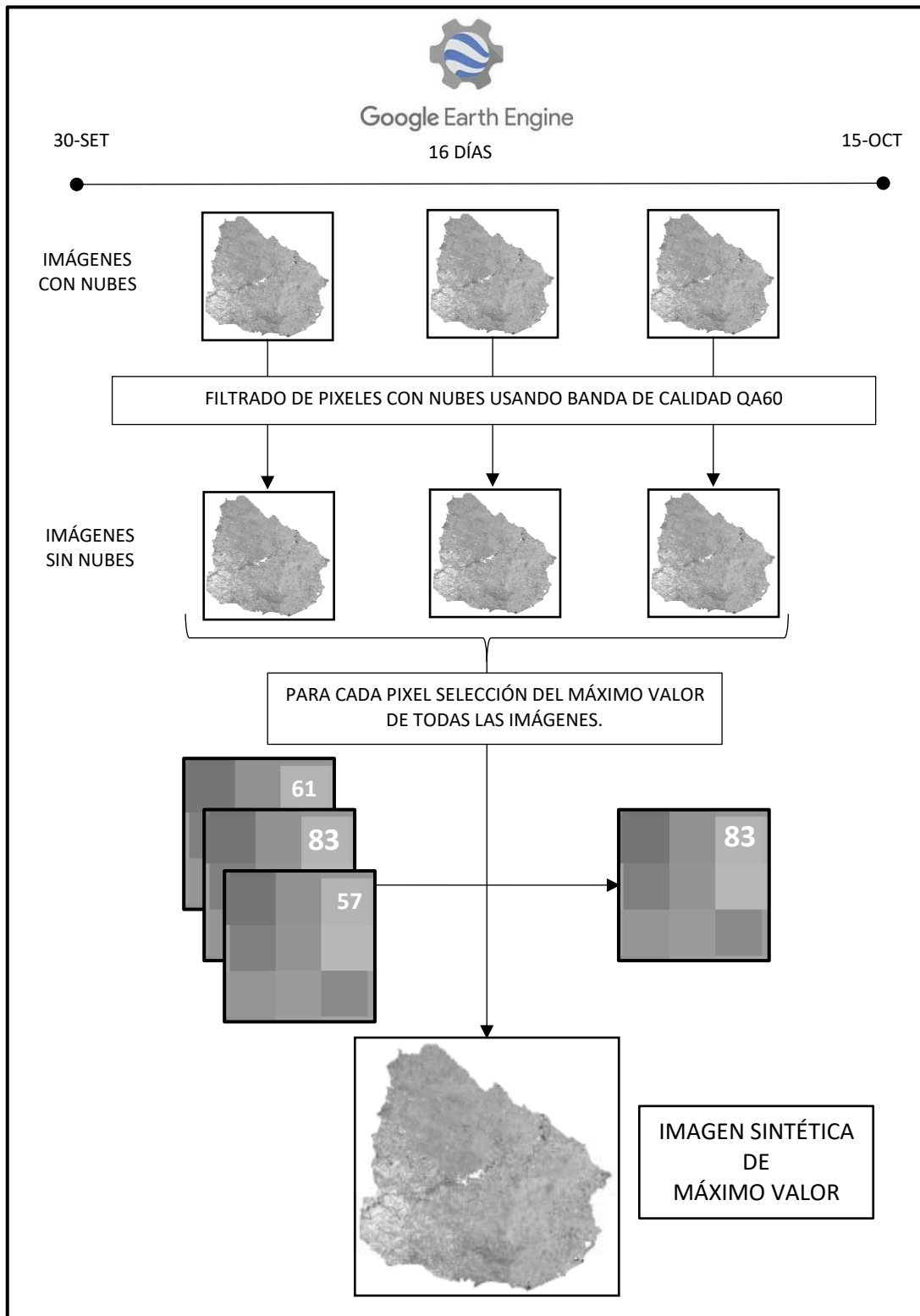


Figura 31. Flujograma de generación de imágenes sintéticas Sentinel 2 en GEE.

Una vez generadas las imágenes sintéticas para todas las bandas de Sentinel 1 y 2 especificadas en la tabla 19, hay que crear un apilado de imágenes sintéticas por banda y posteriormente se extraen las series temporales de los apilados al cruzarlos con los puntos de las bases obtenidos en la sección 7.2. En las figuras 32 y 33 se ve el proceso de generación de apilados y extracción de series temporales para Sentinel.

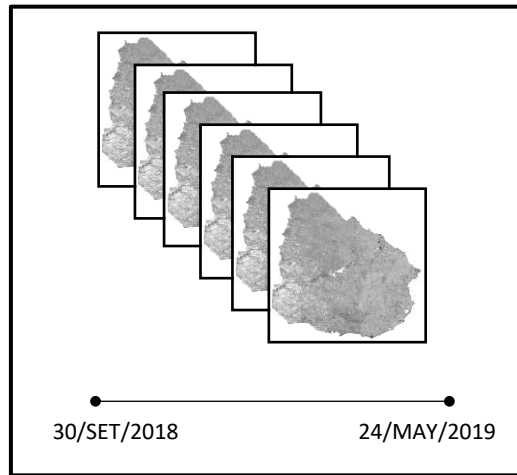


Figura 32. Proceso de generación de apilado de imágenes sintéticas de Sentinel (1 y 2).

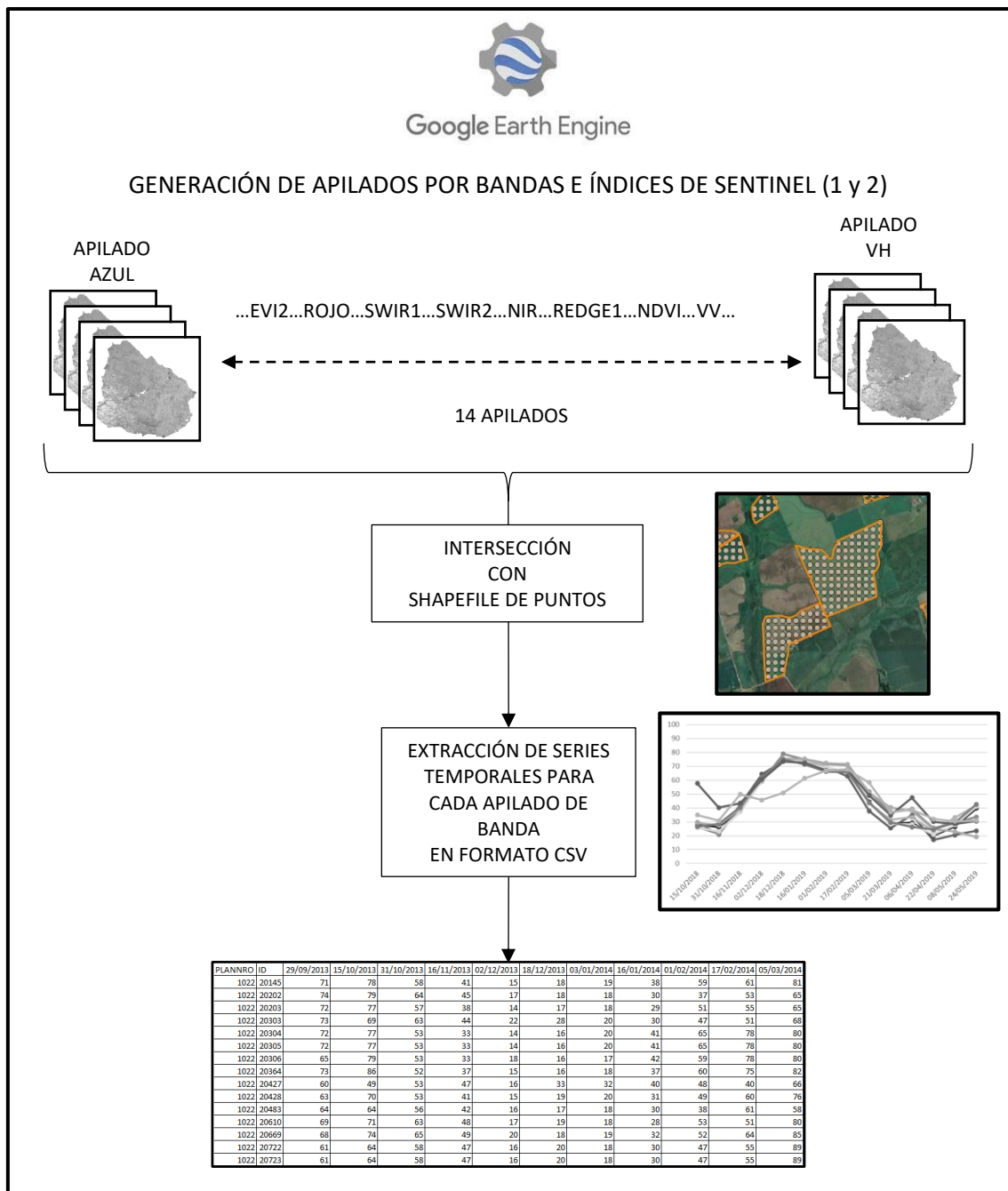


Figura 33. Flujograma de generación de apilados y extracción de series temporales por banda e índice en GEE de Sentinel (1 y 2).

Cuando se extraen series temporales de Sentinel puede haber algunas que tengan datos faltantes con valor NA para algún período. Esto puede deberse, en el caso de Sentinel 2, a que todas las imágenes disponibles para el período tengan nubes para algunos pixeles, y por lo tanto la imagen sintética correspondiente queda con valor NA para los mismos, o los datos de los pixeles de las imágenes originales no estén disponibles por un fallo en el sensor del satélite y también el valor sea NA. Dado el caso mencionado, los datos faltantes deben ser completados (imputados) de acuerdo a algún criterio, ya que algoritmos como Random Forest no funcionan si hay datos faltantes. La imputación hay que realizarla con alguna metodología, que en este caso fue interpolación lineal simple en la dimensión temporal, que consiste en tomar los valores disponibles del período anterior y posterior al dato faltante, y completar este con el promedio de ambos. En la figura 34 se ve el flujograma del proceso de imputación de las series temporales.

Para el caso de las series de Sentinel 2, este proceso de imputación de datos faltantes se realiza en dos etapas para el caso de las bandas de EVI2 y NDVI. En la primera se imputan los datos faltantes y en una segunda etapa se sustituyen aquellos valores que sean inferiores al promedio del valor anterior y posterior en la serie temporal. Esto último se realiza ya que en algunos casos puede haber un valor en la serie que sea muy inferior al de sus vecinos inmediatos. Esto se puede deber a una falla en el sensor del satélite, ya que para el caso de coberturas vegetales no es lógico para EVI2 y NDVI que haya en un breve período de tiempo una caída abrupta de un valor y este vuelva a subir abruptamente en el período siguiente como se ve en la figura 35 para la serie original. En el caso de Sentinel 1 solo se realiza la primera interpolación de la figura 34.

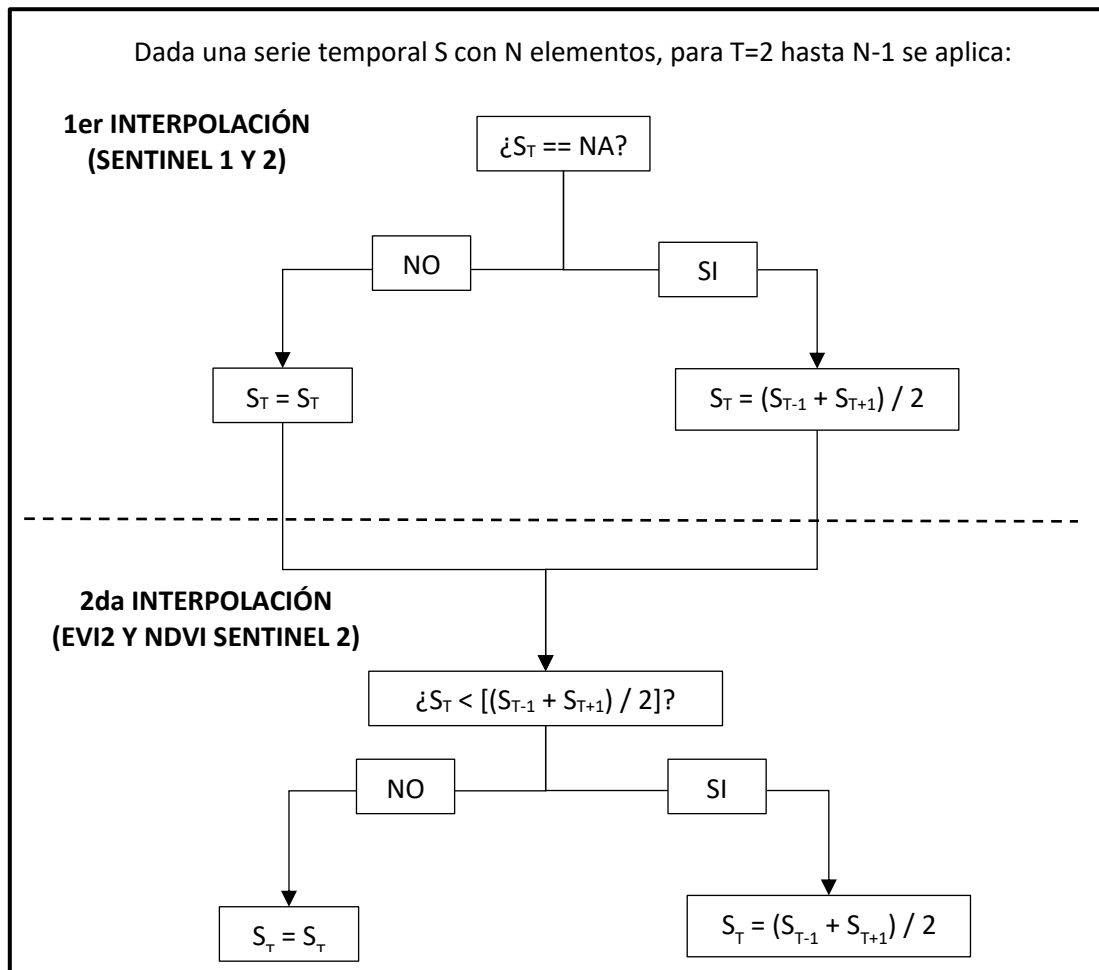


Figura 34. Flujograma de imputación de datos faltantes en series temporales.

En la figura 35 se ve una serie temporal de NDVI antes y después de aplicar el doble proceso de interpolado. En la primera interpolación se imputan los datos faltantes de la serie original. En la segunda interpolación se sustituyen aquellos valores que son inferiores al promedio de sus vecinos (valor anterior y posterior). Esto último, como se ve en la gráfica, se realiza para sustituir valores que disminuyen y aumentan su valor abruptamente en un breve período de tiempo, lo cual no es lógico de esperar en una cobertura vegetal.

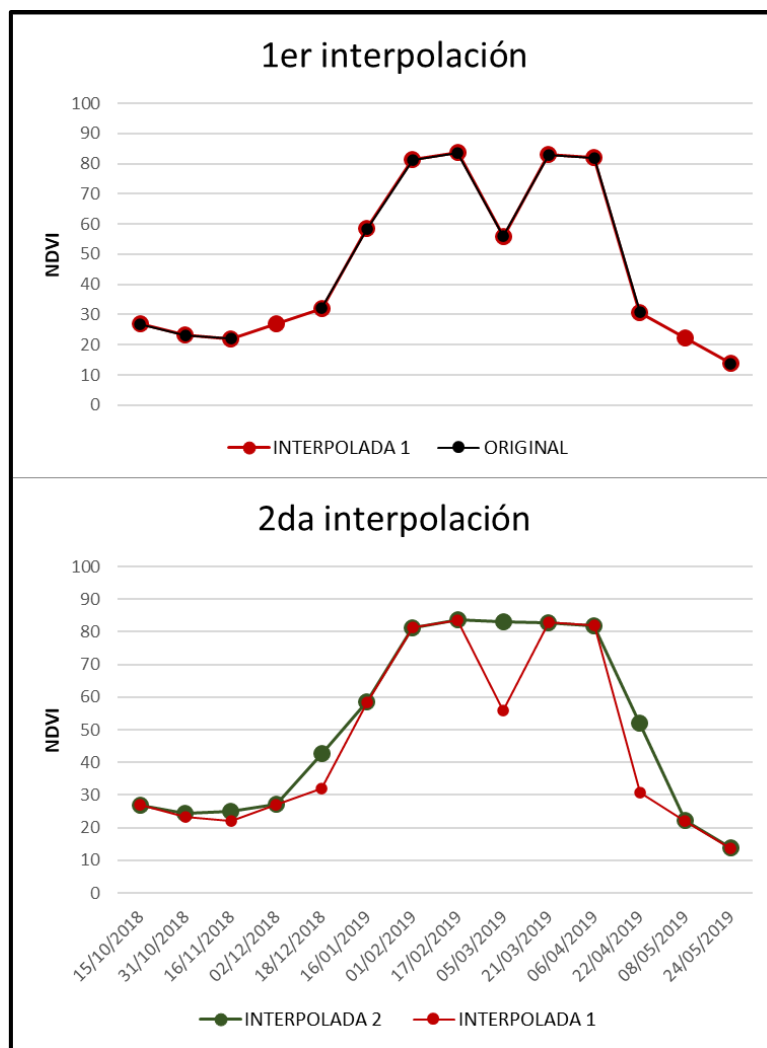


Figura 35. Gráfico de serie temporal de NDVI procesada para imputar datos faltantes.

A continuación, se presenta en la tabla 21 la cantidad de series temporales extraídas por punto según zafra y base.

ZAFRA	BASE	CANTIDAD DE PUNTOS
2017-18	PLANES DE USO	56.747
	FUCREA-URF	48.542
	COBERTURAS URUGUAY	11.080
2018-19	PLANES DE USO	69.870
	FUCREA-URF	47.126
	COBERTURAS URUGUAY	11.080

Tabla 21. Cantidad de puntos con series temporales por cultivo según base y zafra.

9. Depuración de PLANES con clasificación no supervisada

Las etiquetas de cultivo para cada plan no son lo que se denomina en sensoramiento remoto “verdad de terreno”, que es cuando para cada uno se tiene la etiqueta del cultivo que se sembró en realidad. En cambio, como ya se explicó en la sección 5.2.1, la etiqueta asignada para cada plan es lo que el productor declaró que iba a sembrar de acuerdo con una rotación de cultivos agrícolas y pasturas con el objetivo de conservar el suelo. Como no se puede afirmar que el cultivo declarado a sembrar en cada zafra haya sido el realizado, ya que, por circunstancias ambientales, económicas, logísticas, etc., puede haberse realizado otro cultivo o no sembrado nada, es necesario hacer una depuración de los datos mal etiquetados. Este proceso de depuración se hizo usando técnicas de clasificación no supervisada.

9.1 Regionalización de PLANES

Para la base PLANES, previo a aplicar técnicas de clasificación no supervisada y también supervisada, se agruparon los datos en cinco regiones de Uruguay. Esta regionalización se hizo ya que los planes están distribuidos en todo el país, entonces se consideró oportuno analizar estos agrupados por regiones más o menos homogéneas. El criterio para definir las regiones fue en base al tipo de producción agropecuaria predominante (agricultura intensiva, agrícola de secano, arrocería, forestal, ganadera y lechera) en cada departamento, de acuerdo con el último Censo General Agropecuario realizado en 2011 (MGAP-DIEA, 2015). Como se ve en el mapa de la figura 36, las regiones son: Centrosur, Noreste, Noroeste, Sureste y Suroeste. En la tabla 22 se ven los departamentos que abarca cada región.

REGIÓN	DEPARTAMENTOS
Centrosur	Canelones, Durazno, Flores, Florida, Montevideo, San José.
Noreste	Cerro Largo, Rivera, Tacuarembó.
Noroeste	Artigas, Salto.
Sureste	Lavalleja, Maldonado, Rocha, Treinta y Tres.
Suroeste	Colonia, Soriano, Paysandú, Río Negro.

Tabla 22. Departamentos por Región.

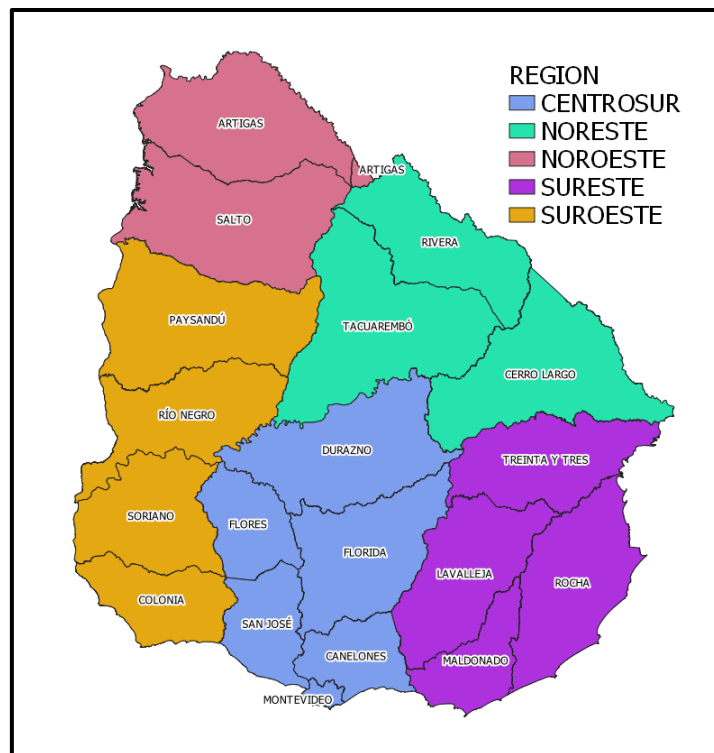


Figura 36. Mapa de regionalización del Uruguay para PLANES.

9.2 Depuración de PLANES mal etiquetados

El proceso de depuración consistió en identificar los planes mal etiquetados usando técnicas de clasificación no supervisada. Estas no requieren que los datos tengan etiquetas, sino que agrupan los datos por similitud en un espacio de características. Las técnicas evaluadas fueron: Gaussian mixture model (GMM), K-means, Self Organizing Map de Kohonen (SOM) + K-means, y Spectral clustering. Gaussian mixture model (GMM) es un algoritmo del tipo probabilístico que asume que las muestras son generadas a partir de determinada cantidad de distribuciones Gaussianas. Cada distribución es un grupo. El algoritmo para cada muestra asigna una probabilidad de pertenencia a cada grupo. El algoritmo, para cada distribución, estima la media y covariancia a partir de todas las muestras asignadas a esa distribución (Wang, Azzari, & Lobell, 2019).

K-means es un tipo de algoritmo de partición de datos en una determinada cantidad (K) de grupos. Cada dato es asignado al grupo del centroide más cercano. Habitualmente se usa la distancia euclídea entre el dato a clasificar y el baricentro de cada grupo (Jain, 2010). El algoritmo de Self Organizing Maps, o también llamado Mapas de Kohonen, es un método de análisis computacional de datos que produce mapas no lineales de datos a menores dimensiones. También puede ser visto como una técnica de agrupamiento que produce un conjunto de grupos organizados en una grilla regular (Martin et al., 2011). Por último, Spectral Clustering es una técnica que usando una matriz de similitud de los datos realiza una reducción de la dimensionalidad previo a agrupar los datos en menos dimensiones (Von Luxburg, 2007).

Como se explicó en la sección 5.2.1, en PLANES los cultivos principales son: maíz, soja, sorgo y pasturas (establecida o consociada, y no consociada). Estos se pueden dividir en dos categorías: cultivos agrícolas de secano (maíz, soja y sorgo) y pasturas. Estas categorías tienen características marcadamente diferentes en cuanto a la duración del cultivo y fenología. Los cultivos agrícolas de secano tienen un ciclo de aproximadamente seis meses, mientras que las pasturas pueden estar presentes más de un año. Esto se puede ver en la evolución del índice de

vegetación EVI o NDVI a lo largo del ciclo de cada una de estas coberturas vegetales. La evolución temporal del índice de vegetación es lo que se conoce como firma fenológica de una cobertura o uso del suelo (Baeza et al., 2014). Cada cobertura/uso del suelo tiene su firma fenológica característica a lo largo del tiempo. A modo de ilustración, la figura 37 tomada del trabajo que Baeza (2014) realizó en Uruguay, muestra las firmas fenológicas de cultivo de verano (CV), doble cultivo (DC), forestación y monte (F y M), y recursos forrajeros perennes (RFP), se observa que la evolución temporal del índice de vegetación es diferente para cada una y se pueden discernir entre sí.

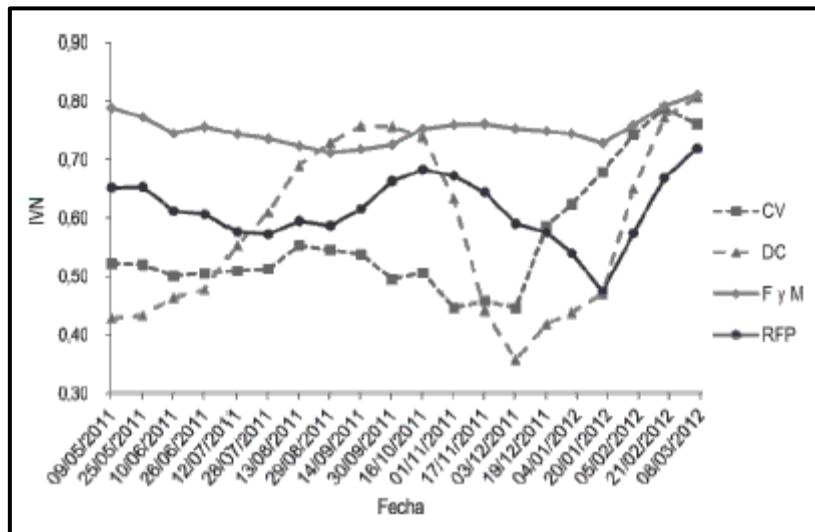


Figura 37. Gráfico de firmas fenológicas promedio de NDVI de diferentes coberturas del suelo en Uruguay (CV: cultivo de verano, DC: doble cultivo, F y M: forestación y monte, RFP: recursos forrajeros perennes). Fuente: Baeza et al. (2014).

Usando como base el mapa de COBERTURAS en GEE se extrajeron para varios puntos las series temporales de EVI2 desde octubre 2017 a mayo 2018 de las coberturas de bosque plantado (forestación), cuerpos de agua naturales, cultivo extensivo de secano (agricultura), pastizal natural, pastizal regenerado (que incluye pasturas implantadas), y zonas urbanas y urbanizadas. Para cada una se construyó la firma fenológica mediana como se ve en la figura 38. Aquí se observa algo similar a lo observado en la figura 37: las diferentes coberturas tienen su firma característica siendo fácilmente distinguibles unas de otras, salvo el caso de pastizales (natural y regenerado) que son muy similares entre sí.

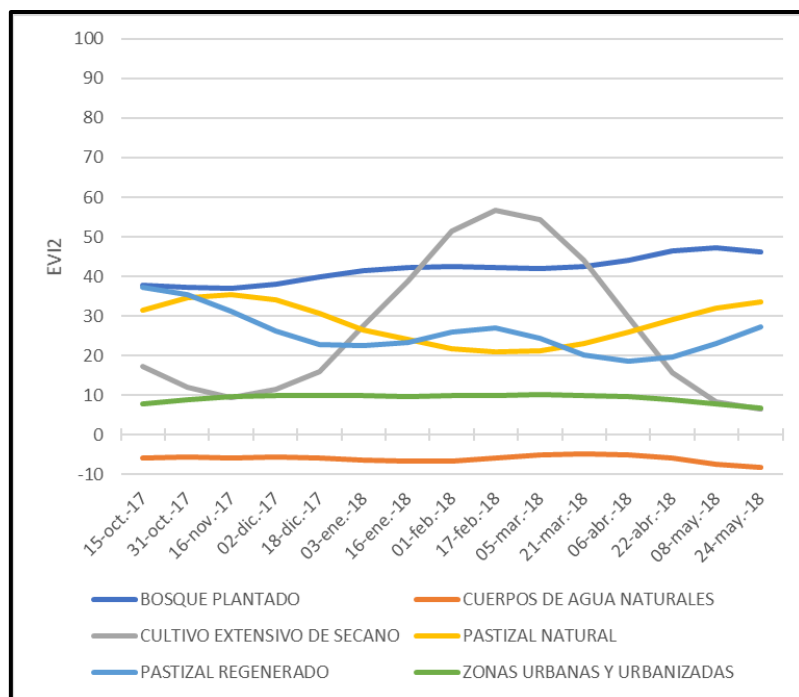


Figura 38. Gráfico de firmas fenológicas medianas de EVI2 de diferentes coberturas y usos del suelo de COBERTURAS para la zafra 2017-18.

En un estudio realizado por Marini (2008) para evaluar el avance de la frontera agrícola en la Provincia de Buenos Aires, se analizaron las firmas fenológicas de los cultivos de verano de girasol, maíz y soja, y como se observa en la figura 39 estas son muy similares entre sí, no siendo fácil la discriminación visual de estas, a diferencia de lo que sucede en los casos de las coberturas de las figuras 37 y 38, que son más contrastantes.

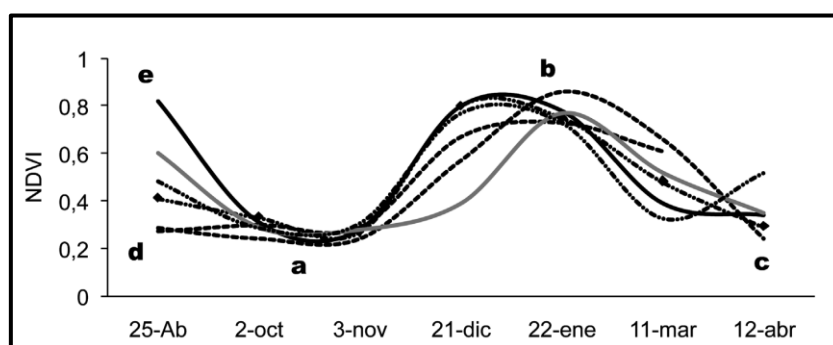


Figura 39. Gráfico de evolución de NDVI de diferentes cultivos de verano (girasol, maíz y soja) en la Provincia de Buenos Aires. Fuente: Marini (2008).

Al analizar los gráficos de las figuras 37 a 39 se observa claramente que la firma fenológica de los cultivos agrícolas tiene una forma característica que se asemeja a la de una función Gaussiana como la de la figura 40.

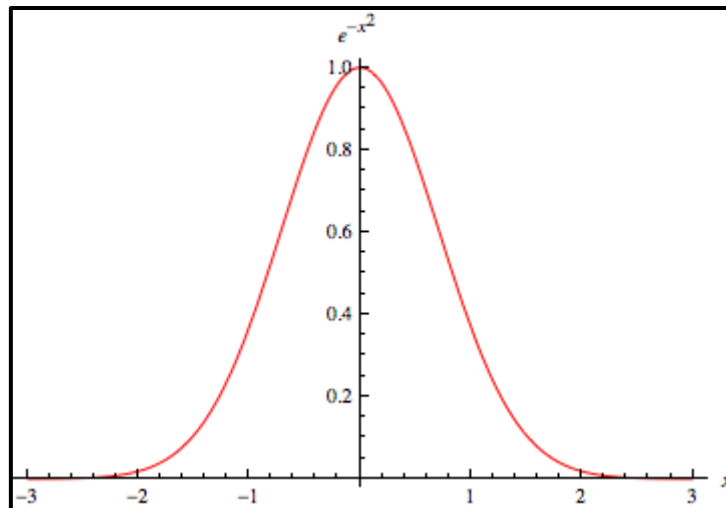


Figura 40. Función Gaussiana. Fuente:
<https://mathworld.wolfram.com/GaussianFunction.html>.

De todo lo anterior surge claramente que la firma fenológica de cultivos agrícolas es muy diferente a la de otras coberturas del suelo como agua, forestación, pasturas, zonas urbanas, etc. Esta diferencia marcada en la firma fenológica es el fundamento base que se utilizó para identificar datos mal etiquetados y depurar la base. Como se mencionó en la base PLANES hay dos grandes categorías: cultivos agrícolas y pasturas. Como ejemplo, en la figura 41 se ve la evolución de NDVI para una pastura y un cultivo agrícola de soja de la base.

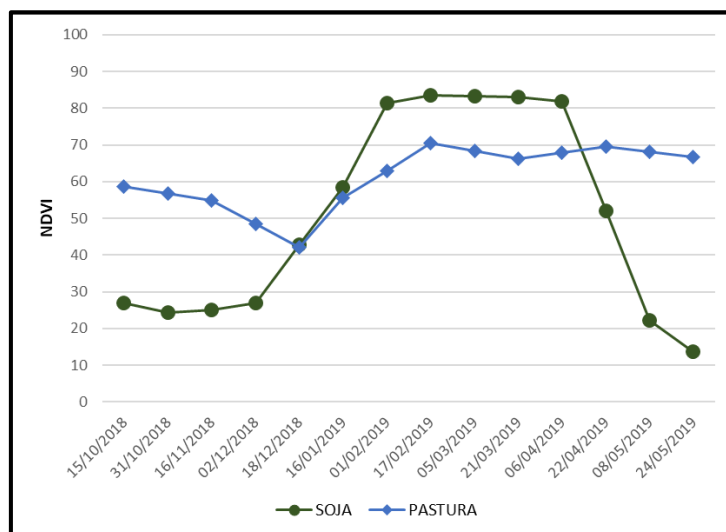


Figura 41. Gráfico de evolución de NDVI para un cultivo agrícola de secano y una pastura de la base PLANES.

Esta diferencia marcada entre las firmas fenológicas de cultivos agrícolas y de coberturas no agrícolas (en este caso pasturas), permite separar entre uno u otro. Pero no permite separar aquellos casos en que un plan declarado como agrícola y que tiene la firma fenológica típica de cultivo agrícola, pudiera el cultivo realizado no ser el declarado, por ejemplo, que el declarado sea maíz y que en la realidad se haya hecho soja, porque como se observa al analizar la figura 39, los cultivos agrícolas tienen una firma muy similar entre sí.

Para la depuración de los planes mal etiquetados se decidió trabajar con las series temporales de EVI extraídas en la sección 8.1. Los planes fueron agrupados en subconjuntos por las regiones definidas en la sección 9.1, zafra y cultivo declarado. La idea de aplicar las técnicas no

supervisadas consiste en identificar grupos de planes que tengan firmas similares entre sí, y comparar si la firma mediana de cada grupo se comporta como la de un cultivo agrícola o una pastura, y ver si ese patrón coincide con la etiqueta declarada en la base. Supongamos que un grupo de planes tiene la firma fenológica típica de un cultivo agrícola como el caso del ejemplo de la figura 39, pero todos están declarados como “pastura no consociada”; en ese caso, la firma no coincide con la etiqueta asignada, y ese es un caso de plan mal etiquetado que no se debe usar en la fase de entrenamiento de algoritmos supervisados.

Las técnicas de clasificación no supervisada se aplicaron a subconjuntos conformados por todos los planes que pertenecieran a la misma zafra, región, y con misma etiqueta de cultivo, por ej.: todos los planes de la zafra 2018-19, región Suroeste y soja, conforman un mismo subconjunto. Una vez aplicada la clasificación no supervisada a cada subconjunto, se obtiene determinada cantidad de grupos para cada uno, por ejemplo cinco, y para cada grupo se generó el gráfico de la firma fenológica mediana de EVI, como se ve en la figura 42. En esta figura, en líneas grises se muestran las firmas individuales de cada plan perteneciente a ese subconjunto y a un grupo dentro de este. En línea negra, la firma mediana del grupo, y en líneas punteadas, el percentil 5 (inferior) y 95 (superior). Posteriormente se analizó la forma de la firma mediana, para ver si correspondía a un patrón agrícola o de pastura de acuerdo con lo visto en la figura 41, y se evaluó si ese patrón coincidía con la etiqueta asignada en la base. Es importante aclarar que este método tiene un componente de subjetividad que depende del usuario que evalúa las firmas fenológicas.

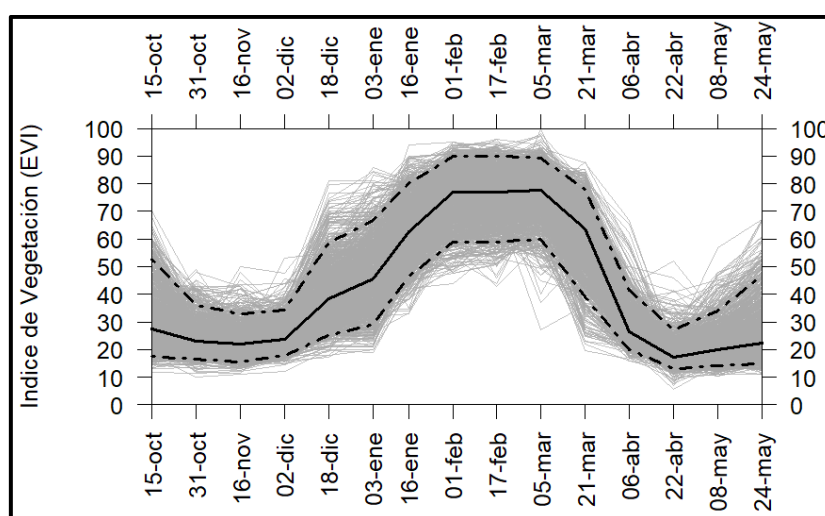


Figura 42. Gráfico de un grupo del subconjunto región Suroeste, soja y zafra 2018-19 de PLANES.

Además de aplicar la clasificación no supervisada, para poder visualizar los grupos generados en la clasificación, se aplicó una técnica de reducción de la dimensionalidad conocida como t-Distributed Stochastic Neighbor Embedding (t-SNE) (Maaten & Hinton, 2008). Esta técnica permite mapear un espacio multidimensional de características con números reales a un espacio de menor dimensionalidad, en este caso de un espacio de 15 dimensiones a uno de 2, y generar gráficos para este último. Esta técnica, en comparación a otras de reducción de la dimensionalidad más conocidas como Análisis de Componente Principales (PCA), permite obtener gráficos de más fácil visualización e interpretación. Usando la técnica de t-SNE sobre cada subconjunto clasificado se generó un gráfico en donde podemos ver todos los planes de ese subconjunto representados como puntos en el gráfico, con la etiqueta y un color según el grupo asignado por la clasificación no supervisada. En la figura 43 se observa un ejemplo del gráfico t-SNE para el subconjunto de la zafra 2016-17, región Suroeste y etiqueta soja.

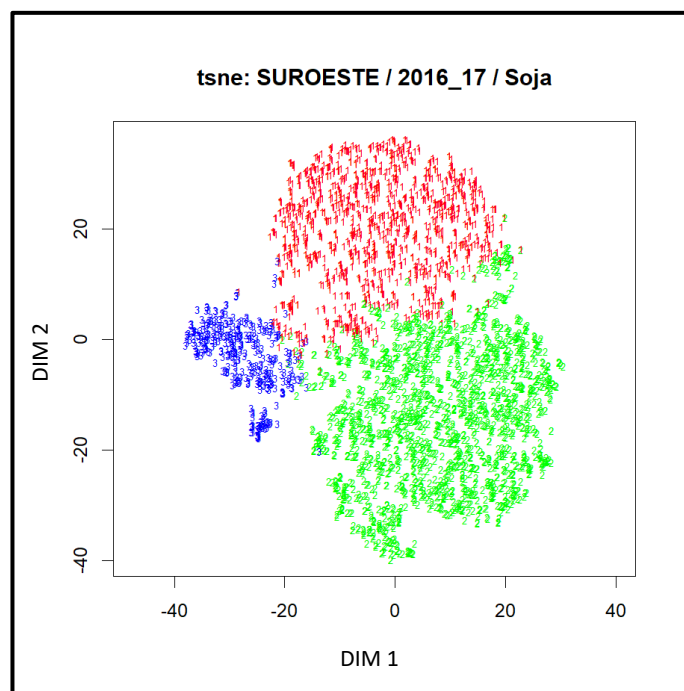


Figura 43. Gráfico de t-SNE del subconjunto de región Suroeste, soja y zafra 2016-17 de PLANES.

En el gráfico de la figura 43 se ve que ese subconjunto está conformado por tres grupos, siendo el número 3 (azul) claramente diferente a 1 (rojo) y 2 (verde), y está conformado por menos cantidad de planes. Si la clasificación aplicada a un subconjunto funciona bien, todos los datos pertenecientes a un mismo grupo deben estar conformando un aglomerado bien definido en el gráfico de t-SNE, y estar más cercanos entre sí que a puntos pertenecientes a otros grupos, como en el caso del gráfico de la figura 43.

En la figura 44 se observa el gráfico con las firmas medianas para cada grupo del subconjunto de la figura 43. En este se nota que el grupo 1 y 2 tienen la firma típica de cultivos agrícolas, mientras que el grupo 3 tiene la firma de una pastura, y precisamente todos los planes pertenecientes a ese grupo son los que se ven más separados del grupo 1 y 2 en el gráfico de t-SNE. De esta manera, los datos del grupo 3 se consideran mal etiquetados y no se usan en la etapa de entrenamiento.

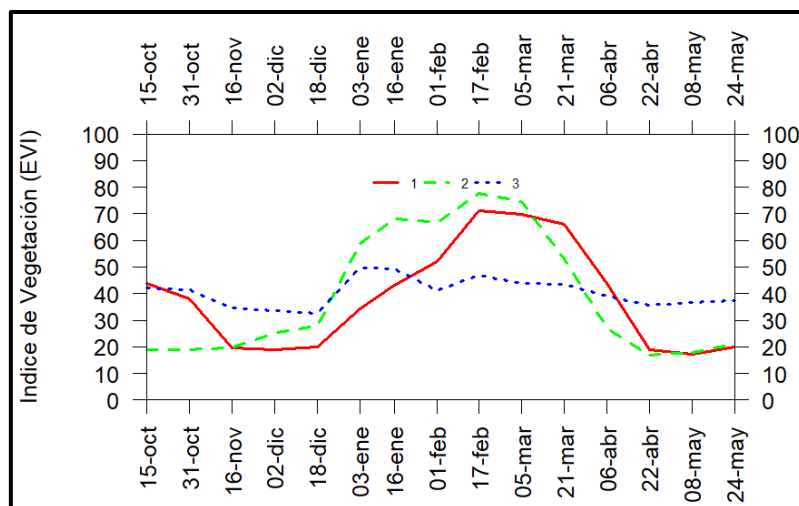


Figura 44. Gráfico de firmas medianas del subconjunto de región Suroeste, soja y zafra 2016-17.

Es así, que combinando las técnicas de clasificación no supervisada de la tabla 23, reducción de dimensionalidad como t-SNE para visualización, generando los gráficos de firmas individuales y firmas medianas por grupo, se puede identificar datos mal etiquetados como cultivos agrícolas o pasturas, y filtrarlos. En la tabla 23, para los algoritmos evaluados se indica el método usado para determinar el número óptimo de grupos para aplicar la clasificación en cada subconjunto, y además se presentan algunas observaciones.

MÉTODO	ALGORITMO	DETERMINACIÓN DE NÚMERO ÓPTIMO DE GRUPOS (K)	OBSERVACIONES
1	Gaussian mixture model (GMM)	Bayesian Information Criterion (BIC).	K=1 a 10.
2	K-Means	Método del codo.	K=1 a 50.
3	SOM + K-Means	Método Shilouette.	Número de neuronas (M) = 5 * RAIZ(N) Mapas: # Filas = # Columnas = RAIZ(M) N: cantidad de datos del subconjunto. K=1 a 10.
4	Spectral clustering	Las determinadas por el método 3.	Matriz de afinidad con 10 vecinos más cercanos.

Tabla 23. Algoritmos de clasificación no supervisada evaluados sobre los diferentes subconjuntos de región, cultivo y zafra.

En las figuras 45 y 46, usando como ejemplo el subconjunto conformado por todos los planes pertenecientes a la región Centrosur, zafra 2016-17 y con etiqueta de cultivo soja, se presentan los gráficos de t-SNE obtenidos para los cuatro métodos de clasificación aplicados. Para el caso de la clasificación con K-means, como la búsqueda del valor óptimo de grupos (K) se realizó con un máximo de 50, la clasificación resultante fue muy fragmentada, con un total de siete clases. Visto esto, en el caso de las clasificaciones realizadas con GMM y SOM+K-means se decidió hacer la búsqueda del K óptimo sobre un máximo de 10 grupos, obteniéndose tres grupos en ambos casos. Para el caso de la clasificación con Spectral clustering, el número de grupos usados fue el mismo que el obtenido con SOM+K-means.

Para el caso de la clasificación K-means, a pesar de ser muy fragmentada, se observa que se formaron grupos bien definidos, con la excepción del grupo 7 que está dividido en dos. Para el resto de los métodos, la clasificación usando GMM parece ser la peor de las tres, habiendo una mayor mezcla de los puntos pertenecientes a un grupo u otro. SOM+K-means y Spectral clustering funcionan bastante bien en discriminar los datos, pareciendo que Spectral clustering

se desempeña algo mejor. La comparación del desempeño de los métodos de clasificación se realizó a través de la exploración visual de los gráficos de t-SNE, no se usó ninguna métrica en particular para analizar la calidad de los diferentes métodos.

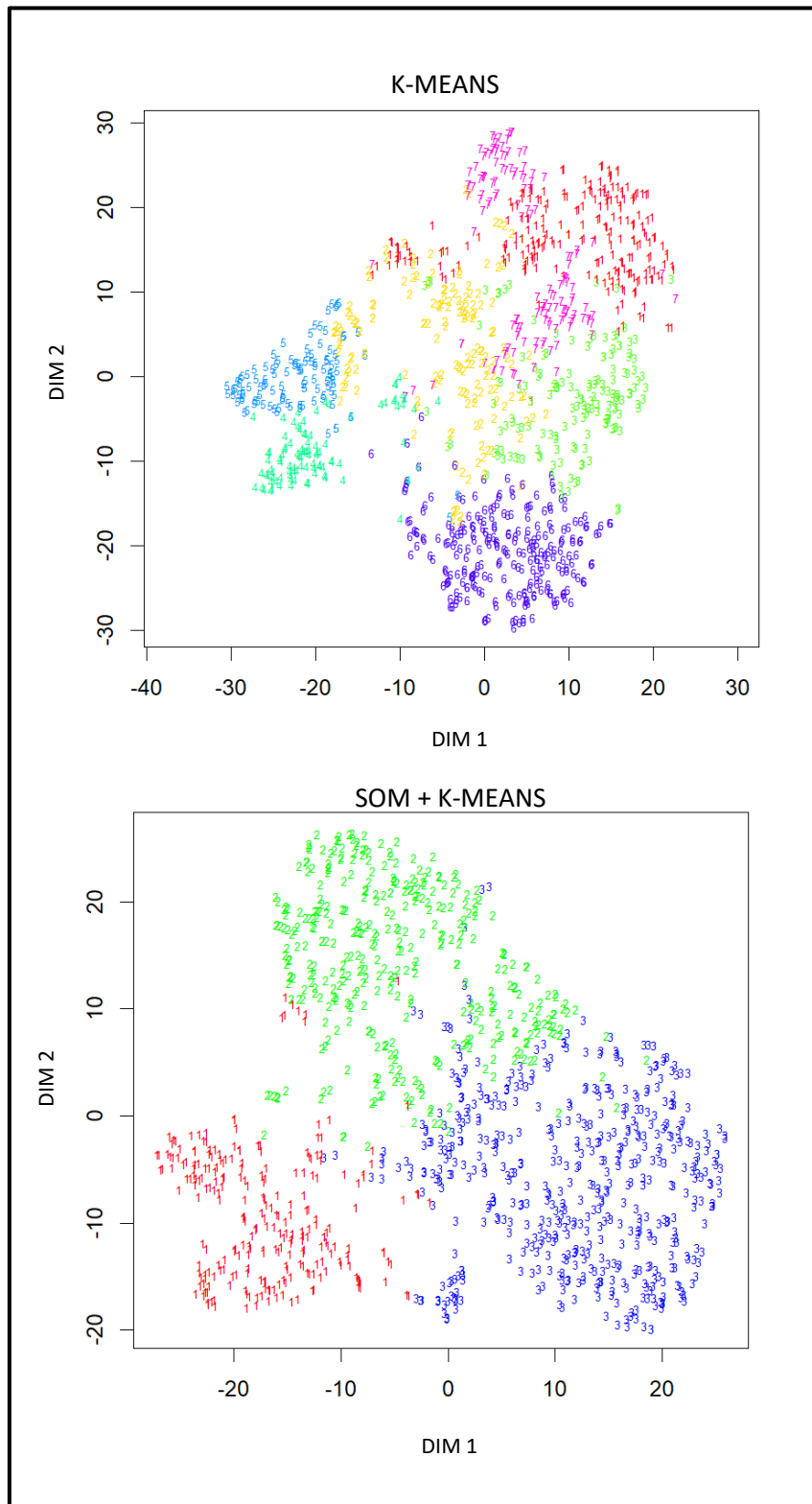


Figura 45. Gráficos de t-SNE para K-means y SOM+K-means del subconjunto Centrosur, soja y zafra 2016-17.

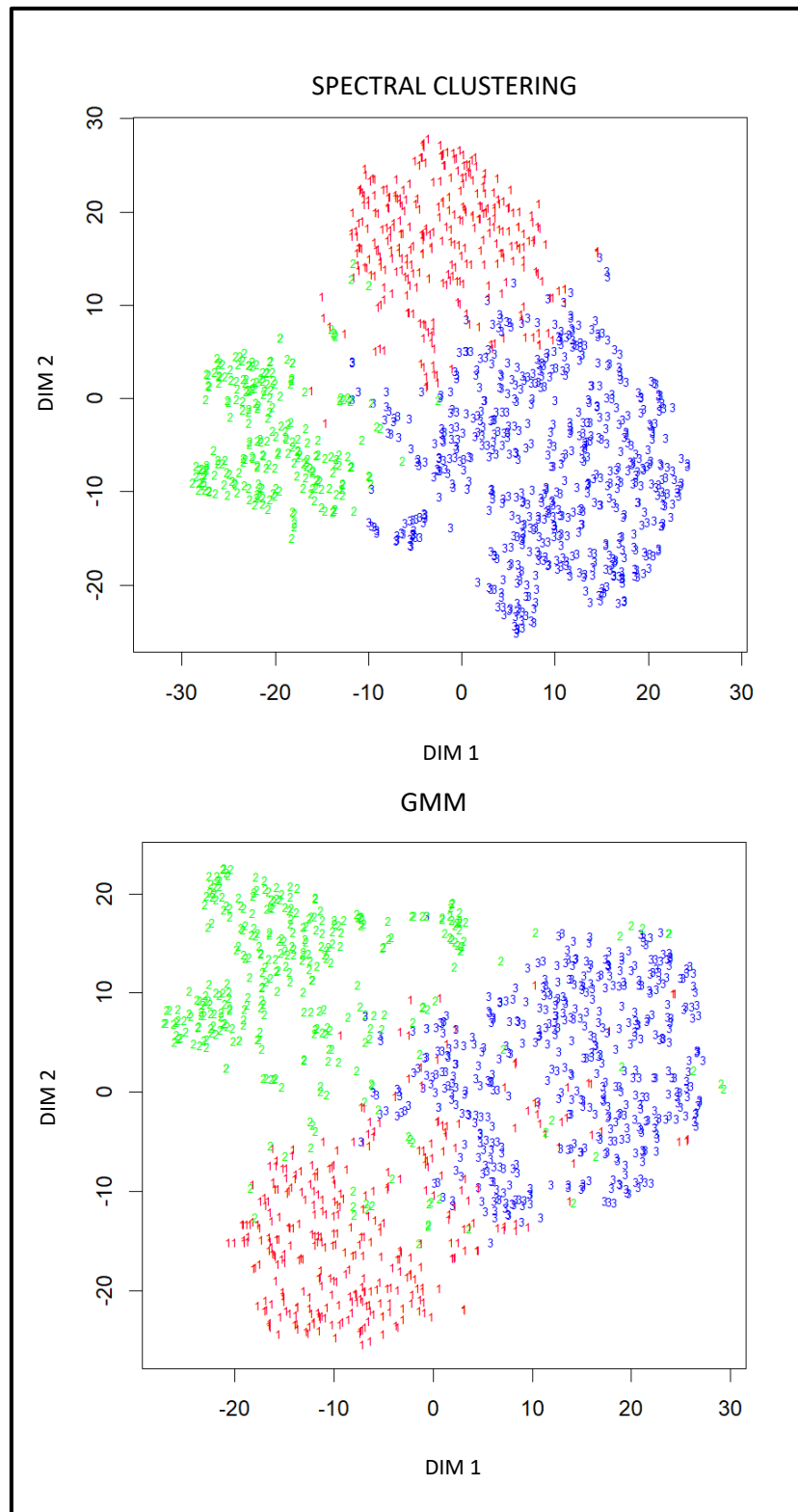


Figura 46. Gráficos de t-SNE para Spectral clustering y GMM del subconjunto Centrosur, soja y zafra 2016-17.

Los cuatro métodos de clasificación no supervisada fueron aplicados a todos los subconjuntos para las distintas combinaciones de región, zafra y etiqueta de cultivo. Considerando que el método de Spectral clustering fue uno de los que alcanzó mejor desempeño, éste se usó para filtrar planes mal etiquetados. Para cada subconjunto, se analizaron los gráficos de firmas individuales, medianas y t-SNE de la clasificación con Spectral clustering, y se registró cuáles eran

los grupos cuya firma mediana coincidía con la etiqueta de cultivo asignada en la base y cuáles no. Aquellos planes de los grupos cuya firma mediana no coincidiera con la etiqueta asignada no fueron usados en la etapa posterior de entrenamiento con algoritmos supervisados.

Una vez identificados los grupos bien etiquetados (primera depuración), se procedió a realizar una segunda depuración de los planes, que consistió en seleccionar dentro de cada grupo, aquellas firmas individuales de EVI que estuvieran entre las bandas del percentil 5 y 95, como se ve en el gráfico de la figura 47. Esto se hizo para eliminar casos atípicos (outliers), en donde su patrón difiriera de la firma mediana (percentil 50). De esta manera, como se ve en la tabla 24, de un total de 43.711 planes con etiqueta de cultivo, luego de realizar la primera depuración quedaron 35.422 planes, y luego de la segunda quedaron 14.000, es decir que la doble depuración eliminó casi un 68,5% de los planes. Estos datos con doble depuración fueron usados para entrenamiento.

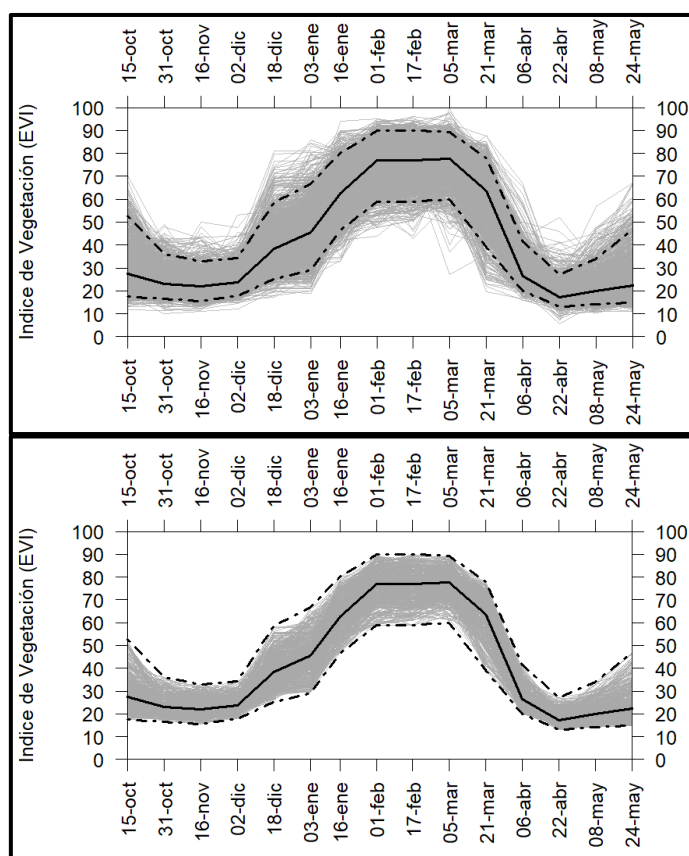


Figura 47. Gráfico de selección de firmas individuales entre percentiles 5 y 95 para un grupo de subconjunto de región Suroeste, soja y zafra 2018-19.

CULTIVO	CANTIDAD DE PLANES		
	CON ETIQUETA	1ER DEPURACIÓN	2DA DEPURACIÓN
SOJA	31.626	26.597	10.161
PASTURA	5.755	4.474	2.021
SORGO-MAÍZ	5.386	3.870	1.423
SORGO	517	250	75
MAÍZ	427	231	59
TOTAL	43.711	35.422	13.739

Tabla 24. Cantidad de planes con etiqueta de cultivo para el período 2013-2019 depurados para entrenamiento.

10. Clasificación supervisada de las bases

En esta sección se presenta el método de clasificación supervisada de cultivos que se efectuó sobre las bases utilizando las series temporales de EVI Terra y de Sentinel (1 y 2).

10.1 Clasificación de PLANES con series temporales de EVI Terra

Para las series temporales de EVI de PLANES depuradas como se detalló en la sección anterior, se aplicaron diferentes técnicas de clasificación supervisada para identificar el cultivo presente. La evaluación se hizo sobre las firmas promedio de EVI de todos los píxeles enteros de Terra que están dentro de cada plan. Como se ve en la tabla 26 los algoritmos elegidos para evaluar fueron: AdaBoost, Random Forest, Support Vector Machine (SVM) y XGBoost. Los algoritmos se ejecutaron en Python, los tres primeros perteneciendo al módulo Sklearn y XGBoost al XGB.

El algoritmo AdaBoost (Adaptative Boosting) fue propuesto por Yoav Freund y Robert Schapire (Freund & Schapire, 1997). Es un algoritmo de tipo Boosting que funciona de manera secuencial, esto es, usando como base algún clasificador débil, como puede ser un árbol de decisión pequeño, de manera iterativa en cada corrida sucesiva mediante el uso de pesos mejora la clasificación de la corrida anterior hasta obtener un clasificador robusto. La clasificación final se obtiene por voto mayoritario de los clasificadores base, ponderados por un factor de confianza aprendido durante el entrenamiento.

El algoritmo Random Forest fue desarrollado por Leo Breiman (Breiman, 2001). Este algoritmo es de tipo Bagging y funciona creando muchos árboles de decisión en la etapa de entrenamiento, cada árbol difiere en cuanto a que del conjunto usado se eligen de manera aleatoria un subconjunto de registros y un subconjunto de variables para construir cada árbol, de esta manera todos son diferentes entre sí. La clase final se obtiene por voto mayoritario de todos los árboles creados.

Support Vector Machine (SVM) fue propuesto por Vladimir Vapnik y otros (Vapnik, Guyon & Hastie, 1995). Éste, cuando se usa para clasificación a partir de los datos de entrenamiento, construye un hiperplano o conjunto de hiperplanos en un espacio de alta dimensionalidad al que usualmente se transforman los datos no linealmente mediante el uso del “kernel trick” que utiliza para separar de manera óptima, con mayor margen, los datos que pertenecen a dos o más clases excluyentes.

Finalmente, el algoritmo XGBoost (Extreme Gradient Boosting) fue creado por Tianqi Chen y Carlos Guestrin (Chen & Guestrin, 2016). Este, es el más moderno de los algoritmos evaluados.

Es un algoritmo de tipo Gradient Boosting (Friedman, 2002) que crea árboles de decisión, y a partir de estos calcula los residuales del error y estos los usa para mejorar las clasificaciones siguientes. La clasificación final es la combinación de todas las predicciones de los árboles creados.

Previo a entrenar los diferentes algoritmos, la base PLANES depurada fue barajada al azar y dividida de manera estratificada en 80% para entrenamiento y 20% para evaluación de la calidad de la clasificación. Como se observa en la tabla 25, la base está fuertemente desbalanceada, por esa razón la división en entrenamiento y evaluación se hizo de manera estratificada para que todas las clases (soja, maíz, sorgo, etc.) estuvieran presentes en ambos conjuntos en la misma proporción que en el original, y evitar que alguna clase minoritaria (maíz o sorgo) no estuviera presente en alguno de estos.

CULTIVO	PLANES		
	TOTAL	ENTRENAMIENTO	EVALUACIÓN
SOJA	10.161	8.129	2.032
PASTURA	2.021	1.617	404
SORGO-MAÍZ	1.423	1.138	285
SORGO	75	60	15
MAÍZ	59	47	12
TOTAL	13.739	10.991	2.748

Tabla 25. Cantidad de planes depurados totales en entrenamiento y evaluación según cultivo.

En observaciones de la tabla 26 se indica en qué condiciones se corrió cada algoritmo. Para el caso de los algoritmos de Random Forest, SVM y XGBoost, como el conjunto de entrenamiento está muy desbalanceado, las clases fueron balanceadas por peso, es decir, se les asignó más peso a las clases minoritarias para evitar que los algoritmos se sesguen a clasificar mejor las clases mayoritarias. Esta técnica de darle más pesos a algunas clases, en este caso las minoritarias, se conoce como Cost-Sensitive Learning (Sammut & Webb, 2017).

ALGORITMO	OBSERVACIONES
ADABOOST	300 árboles, tasa de aprendizaje= 0,025.
RANDOM FOREST	300 árboles, máxima cantidad de features=15, clases balanceadas por peso y bootstrapping.
SUPPORT VECTOR MACHINE (SVM)	Kernel lineal, coeficiente de penalización=0,025 y clases balanceadas por peso.
XGBOOST	300 árboles y clases balanceadas por peso, tasa de aprendizaje=0,025.

Tabla 26. Algoritmos de clasificación supervisada utilizados sobre el conjunto de entrenamiento en PLANES.

Una vez definidos los conjuntos de entrenamiento y evaluación, cada algoritmo de la tabla 26 se entrenó con el conjunto de entrenamiento y posteriormente se clasificó el conjunto de evaluación con cada modelo. Una vez realizada la clasificación en evaluación se construyó la matriz de confusión para cada algoritmo, y a partir de esta se calcularon las siguientes métricas de desempeño: *Precision*, *Recall* y *F1-Score* (Sharma & Mishra, 2020).

Antes de mostrar los resultados de los algoritmos evaluados, es necesario definir los conceptos de Matriz de Confusión, Falso Positivo, Verdadero Positivo, Falso Negativo, *Precision*, *Recall* y *F1-score*. La matriz de confusión (Sharma & Mishra, 2020) o también llamada matriz de error es un tipo de tabla de contingencia de doble entrada como la de la tabla 27.

Cuando se evalúa un algoritmo de clasificación supervisada como por ejemplo Random Forest, primero es entrenado con el conjunto de entrenamiento, y posteriormente es necesario evaluar qué tan bueno es el algoritmo clasificando ejemplos que no fueron usados para entrenarlo, es decir sobre el conjunto de evaluación. Esto se realiza a través de la construcción de la matriz de confusión, ésta es una matriz cuadrada, en donde en filas y columnas están presentes todas las clases del conjunto de entrenamiento (maíz, soja, sorgo, pastura, etc.), o sea que su dimensión es de $n \times n$ (n =cantidad de clases). Además, también están los totales para valores clasificados y observados. En la intersección de las columnas y filas se registra las diferentes combinaciones posibles de cultivo observado y clasificado. En la diagonal de la matriz están los casos bien clasificados, es decir, cuando lo que clasificó el algoritmo coincide con la etiqueta observada en la base; por ejemplo, la combinación pastura-pastura, nos indica que hay 395 casos de evaluación que fueron clasificados como pastura y realmente son pastura. Si tomamos la combinación pastura-soja, vemos que hay 9 casos que el modelo clasificó como soja, cuando en realidad son pastura. Es importante aclarar que no es lo mismo la combinación pastura-soja, que la combinación soja-pastura. La diferencia es sutil, pero en el primer caso, la matriz indica cuantos casos fueron clasificados como soja siendo pastura (9), y en el segundo cuantos casos del total fueron clasificados como pastura cuando en realidad son soja (2).

		CLASIFICADO					TOTAL	RECALL	F1-SCORE
		MAÍZ	MAÍZ-SORGO	PASTURA	SOJA	SORGO			
OBSERVADO	MAÍZ	1	2	0	9	0	12	8,3	12,5
	MAÍZ-SORGO	3	66	4	209	3	285	23,2	35,8
	PASTURA	0	0	395	9	0	404	97,8	98,1
	SOJA	0	13	2	2.016	1	2.032	99,2	94,1
	SORGO	0	3	0	11	1	15	6,7	10
	TOTAL	4	84	401	2.254	5	2.748		
PRECISION		25	78,6	98,5	89,4	20			

Tabla 27. Matriz de confusión de Random Forest en evaluación de PLANES.

Una vez construida la matriz de confusión, a partir de esta se pueden calcular las métricas de desempeño como *Precision*, *Recall* y *F1-Score*. *Precision* se define como el cociente entre los casos llamados verdaderos positivos (VP) sobre la suma de casos de verdaderos positivos y falsos positivos (FP). En sensoramiento remoto, *precision* es lo que se conoce como Exactitud del Usuario (User's Accuracy) (Story & Congalton, 1986).

$$Precision = \frac{VP}{FP + VP} \times 100.$$

Los VP son aquellos casos en donde lo que clasifica el algoritmo coincide con la etiqueta observada, por ejemplo, los casos que en la matriz son soja-soja, maíz-maíz, etc., es decir los valores de la diagonal de la matriz de confusión. Los FP son aquellos casos que siendo en la realidad un cultivo, el modelo lo clasifica como otro diferente; por ejemplo, siendo todos los casos maíz, el algoritmo los clasifica como soja, sorgo o pastura. Si tomamos el caso del cultivo maíz, su *precision* es de 25 ($1/4 \times 100$), porque de un total de 4 casos clasificados como maíz en entrenamiento, hay 1 que es efectivamente maíz (VP) y hay 3 casos que son maíz-sorgo (FP). En otras palabras, la *precision* indica de todos los casos clasificados como X, cuantos efectivamente son X.

El *Recall* o también llamado Sensibilidad, es la relación entre VP y la suma de VP más Falsos Negativos (FN). En sensoramiento remoto, *recall* es lo que se conoce como Exactitud del Productor (Producer's Accuracy) (Story & Congalton, 1986).

$$Recall = \frac{VP}{FN + VP} \times 100.$$

Los FN son la cantidad de casos que siendo un cultivo determinado, por ejemplo maíz, el algoritmo los clasifica como otros (soja, pastura, sorgo). Si tomamos el ejemplo del maíz, su *recall* es de 8,3 (1/12*100), porque de 12 casos que son maíz, hay 1 VP, y hay 11 FN (2 clasificados maíz-sorgo y 9 como soja). En otras palabras, el *recall* indica de todos los casos que son X, cuantos fueron clasificados como X.

El *F1-Score*, también llamado promedio armónico, relaciona *precision* y *recall*, y es otra medida de desempeño, su valor está en el rango de entre 0 y 100. Un valor de 100 indicaría una clasificación perfecta.

$$F1\ Score = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

Para obtener un valor de *F1-Score* alto, es necesario que *precision* y *recall* sean altas también. Siguiendo con el ejemplo del maíz, su *F1-Score* es de 12,5 = 2 x (25x8,3) / (25+8,3). En la tabla 28 se resumen las diferentes métricas mencionadas y su fórmula de cálculo.

METRICA	FORMULA
FALSOS NEGATIVOS (FN)	Ej.: cantidad de casos que siendo soja se clasifican como otros cultivos.
FALSOS POSITIVOS (FP)	Ej.: cantidad de casos clasificados como soja cuando en realidad son otro cultivo (maíz, pastura, sorgo).
VERDADEROS POSITIVOS (VP)	Ej.: Cantidad de casos que siendo soja son clasificados como soja.
PRECISION (P)	$P = VP / (FP + VP) \times 100$
RECALL (R) (SENSIBILIDAD)	$R = VP / (FN + VP) \times 100$
F1-SCORE	$F1-SCORE = P \times R / (P + R)$

Tabla 28. Métricas para evaluación de desempeño de algoritmos de clasificación supervisada.

En esta sección no se presentan todas las matrices de confusión de los algoritmos evaluados, sino que en las tablas 29 a 31 y figuras 48 a 50, se presentan a modo de resumen los valores obtenidos de *precision*, *recall* y *F1-score* por cultivo y por algoritmo entrenado.

Para *precision* se observa que el algoritmo que en general tuvo mejor desempeño fue Random Forest. Ahora bien, si se analiza cultivo por cultivo, algunos algoritmos se desempeñaron mejor que otros. Para el caso de maíz, XGBoost fue el mejor, con una *precision* de 50, seguido por Random Forest con 25, y SVM y Adaboost con valores de 2 y 0 respectivamente. En el caso de maíz-sorgo, a excepción de SVM, todos tuvieron valores superiores a 71, comportándose mejor Random Forest. En el caso de pastura, todos tuvieron valores superiores a 95, siendo éste el cultivo mejor discriminado, lo cual tiene sentido porque como ya se vio en el análisis de las firmas fenológicas de diferentes coberturas del suelo, ésta es muy diferente a la de un cultivo agrícola. Para soja, *precision* estuvo por encima de 86 para cualquiera de los algoritmos, siendo el mejor SVM. Y finalmente, sorgo fue uno de los peores cultivos clasificados por todos los algoritmos, siendo el mejor Random Forest con un valor de 20.

CULTIVO	PRECISION			
	ADABOOST	RANDOM FOREST	SVM	XGBOOST
MAÍZ	0,0	25,0	2,0	50,0
MAÍZ-SORGO	71,4	78,6	21,5	75,3
PASTURA	96,9	98,5	95,0	97,5
SOJA	86,3	89,4	93,8	89,6
SORGO	0,0	20,0	6,0	0,0

Tabla 29. *Precision* de los algoritmos de clasificación supervisada evaluados en PLANES.

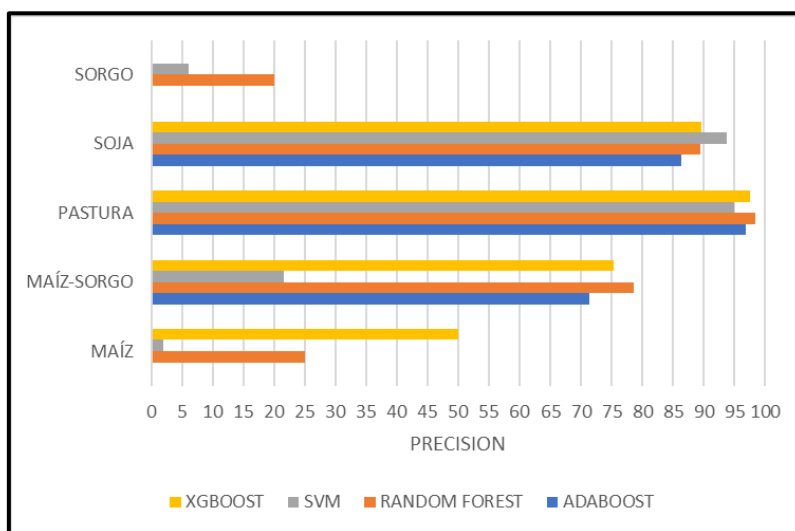


Figura 48. Gráfico de *Precision* de los algoritmos de clasificación supervisada evaluados en PLANES.

En el caso de *recall* se observa que los cultivos mejor discriminados por cualquier de los algoritmos son pastura y soja, obteniéndose valores superiores a 93, a excepción de SVM. Para el resto de los cultivos, la situación es bastante diferente, el desempeño de los algoritmos disminuye bastante. Para el caso de maíz, el mejor es SVM con un valor de 50, mientras que para Adaboost es 0, o no llega a 10 como Random Forest y XGBoost. En el caso de maíz-sorgo, el mejor también es SVM con casi 28. Finalmente, para sorgo, SVM vuelve a ser el mejor con casi 70 de *recall*, mientras que los otros algoritmos tienen valores de 0 o inferior a 7.

CULTIVO	RECALL			
	ADABOOST	RANDOM FOREST	SVM	XGBOOST
MAÍZ	0,0	8,3	50,0	8,3
MAÍZ-SORGO	1,8	23,2	27,7	24,6
PASTURA	93,8	97,8	94,1	98,0
SOJA	99,8	99,2	70,2	99,1
SORGO	0,0	6,7	66,7	0,0

Tabla 30. *Recall* de los algoritmos de clasificación supervisada evaluados en PLANES.

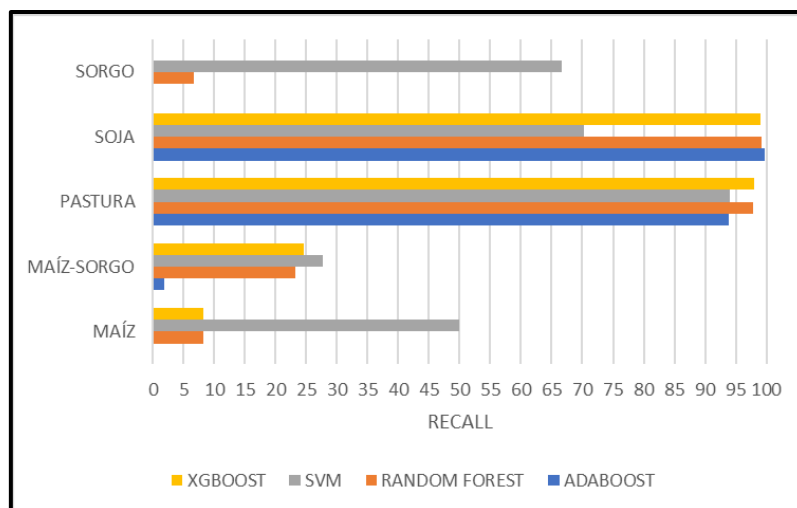


Figura 49. Gráfico de *Recall* de los algoritmos de clasificación evaluados en PLANES.

Por último, se revisaron los resultados de la métrica de *F1-Score*. Ésta está afectada por los valores de *precision* y *recall*, y habiéndose obtenido resultados malos de estas para los cultivos de maíz, maíz-sorgo y sorgo, *F1* tiene mal desempeño en estos cultivos, incluso con valores de 0, como vemos en la tabla 31 y figura 50. Para el caso de pastura y soja, *F1* tiene valores superiores a 90 para cualquiera de los algoritmos, a excepción de SVM, que obtuvo 80,3 para soja.

CULTIVO	F1-SCORE			
	ADABOOST	RANDOM FOREST	SVM	XGBOOST
MAÍZ	0,0	12,5	3,9	14,3
MAÍZ-SORGO	3,4	35,8	24,2	37,0
PASTURA	95,3	98,1	94,5	97,8
SOJA	92,6	94,1	80,3	94,1
SORGO	0,0	10,0	10,9	0,0

Tabla 31. *F1-Score* de los algoritmos de clasificación supervisada evaluados en PLANES.

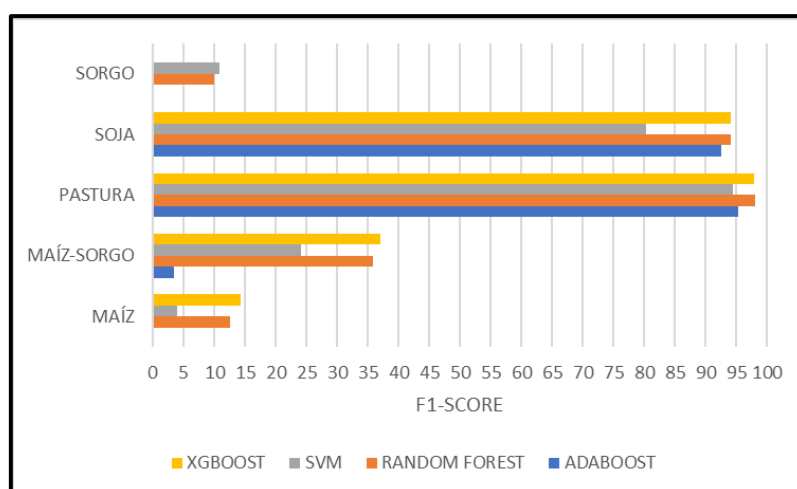


Figura 50. Gráfico de *F1-Score* de los algoritmos de clasificación supervisada evaluados en PLANES.

De los resultados anteriores se observa que los cultivos mejor discriminados por cualquiera de los algoritmos evaluados son pastura y soja, con mejor desempeño en pastura. En ambos, los

valores de *F1-Score* son superiores a 90, con la excepción de SVM para soja que apenas superó 80. Para estos cultivos, Random Forest y XGBoost son los algoritmos de mejor desempeño. En el caso de maíz, maíz-sorgo y sorgo, la situación es bastante diferente. Para estos, el desempeño es bastante malo, siendo los mejores resultados en *F1* apenas superior a 35, e incluso en algunas clases teniendo valores de 0 para algunos algoritmos, como el caso de maíz y sorgo con Adaboost y XGBoost. Este mal desempeño en estos cultivos puede estar influenciado por dos situaciones. Por un lado, éstos son clases minoritarias en la base, sobre todo maíz y sorgo. En el caso de la clase maíz-sorgo, esta representa el 10% de los planes depurados, mientras que maíz y sorgo apenas son el 1% de estos. Además, existe el problema de la confusión entre las clases, ya que tenemos la clase maíz, sorgo, y la clase maíz-sorgo, la cual es una clase artificial, ya que un plan puede ser o maíz o sorgo, pero no ambos. Como ya se mencionó en la descripción de la base en la sección 5.2.1, la existencia de esta clase se debe a que los productores agropecuarios al momento de presentar un plan de rotación de cultivos podían para el caso de que hicieran maíz o sorgo, no especificar cual iban a sembrar, pero si notificar maíz-sorgo.

10.2 Clasificación de FUCREA-URF con series temporales de Sentinel

En esta sección se usaron las series temporales de las diferentes bandas e índices de Sentinel que se extrajeron y procesaron de los cultivos de FUCREA-URF como se explicó en la sección 8.2. Esta sección se divide en tres secciones. En la primera, se explica cómo se depuraron los puntos que caen en zonas no sembradas de las chacras. En la segunda, se realizaron una serie de experimentos usando los puntos depurados dentro de las chacras para evaluar cuáles son las mejores bandas e índices, período de fechas y algoritmos supervisados a usar para clasificar cultivos. En la tercera, una vez definida la mejor combinación de bandas, fechas y algoritmos, se evalúa la clasificación a nivel de chacra.

10.2.1 Depuración de puntos no agrícolas en las chacras

Como se explicó en la sección 7.2, para cada chacra de FUCREA-URF se generó una grilla de puntos regulares cada una hectárea, y para cada punto se extrajeron las series temporales de Sentinel como fue detallado en la sección 8.2. De esta manera, una chacra que tenga 50 hectáreas tiene 50 puntos, y para cada uno se extrajeron las series temporales de las diferentes bandas e índices de Sentinel 1 y 2. Ahora bien, dentro de una chacra puede haber zonas bajas, montes de abrigo y sombra, represas pequeñas (tajamares), cursos de agua, caminos, construcciones y otros. En estas zonas no se puede sembrar. También pasa que una chacra no sea sembrada en su totalidad, sino una parte. Incluso, puede haber zonas que hayan sido sembradas, pero la emergencia del cultivo haber fallado. Algunos de los puntos regulares generados caen en algunas de estas zonas que no son agrícolas, y si se utilizan las series temporales de estos para entrenar modelos de clasificación supervisada, se está utilizando datos espurios, por lo cual esos puntos no agrícolas tienen que ser identificados y depurados.

En la figura 51 se ve un ejemplo de algunas chacras con puntos que caen en zonas no agrícolas (zonas bajas, montes de abrigo y sombra, tajamares, etc.), y en la figura 52 las series temporales de NDVI de puntos agrícolas y no agrícolas de una de las chacras anteriores.



Figura 51. Ejemplo de chacras de FUCREA-URF con algunos puntos regulares en zonas no agrícolas. 1: Monte de abrigo y sombra, 2: zonas bajas, 3: área con construcciones y caminos, 4: curso de agua.

En la chacra de la figura 52 se observa que hay puntos que caen en zonas bajas correspondientes a vías de drenajes, mientras que otros puntos caen en zonas efectivamente sembradas (agrícolas). Esto se confirma a través de los gráficos de NDVI para los puntos identificados como agrícolas y para los no agrícolas. En el caso de los puntos agrícolas, estos tienen el patrón esperado, es decir una forma similar a una función Gaussiana como se vió en la sección 9.2, mientras que los no agrícolas, no.

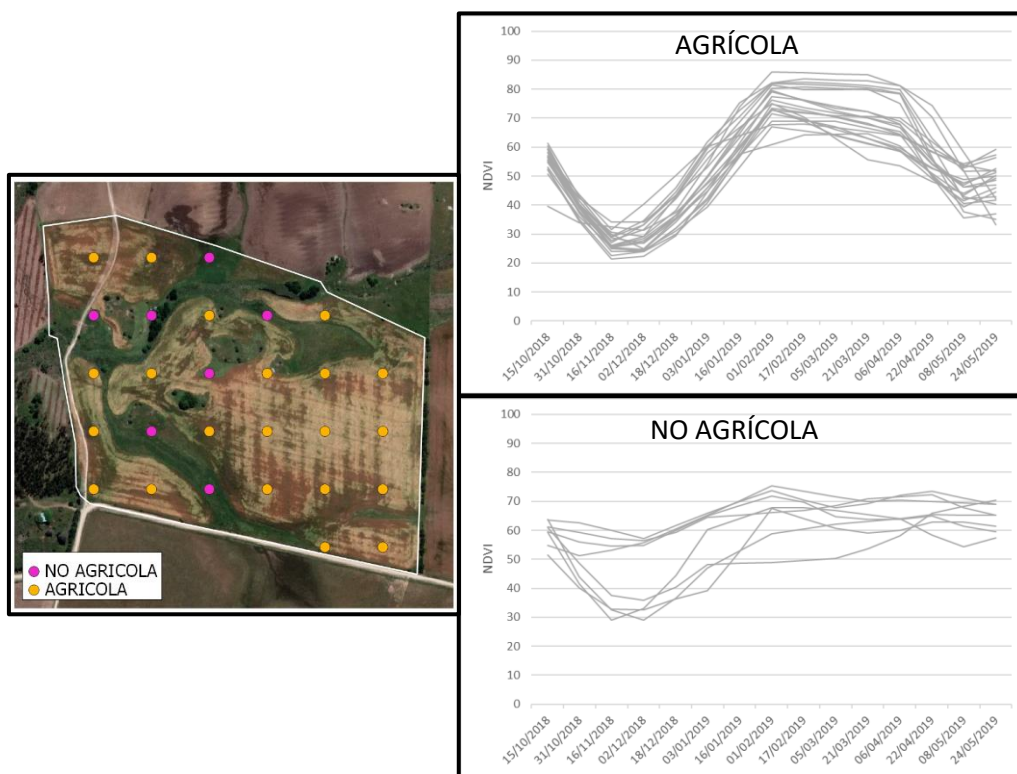


Figura 52. Series temporales de NDVI de una chacra con puntos agrícolas y no agrícolas.

El proceso de depuración de los puntos no agrícolas es similar al que se utilizó en la sección 9.2 para identificar PLANES mal etiquetados. Se utilizó de manera combinada clasificación no supervisada K-means de series temporales de NDVI, gráficos con firmas individuales y medianas de NDVI por grupo, y gráfico de t-SNE. En el flujograma de la figura 53 se ve el proceso detallado. Este se hizo para cada zafra por separado. Todos los puntos que pertenecen a una misma zafra y cultivo se toman como un subconjunto, y usando K-means se lo clasifica en K cantidad de grupos. El número óptimo de grupos se determina por el método del codo. Una vez hecha la clasificación para cada subconjunto de zafra-cultivo, se generan los gráficos de cada grupo con sus firmas individuales, los de firmas medianas NDVI de los grupos, y el de t-SNE.

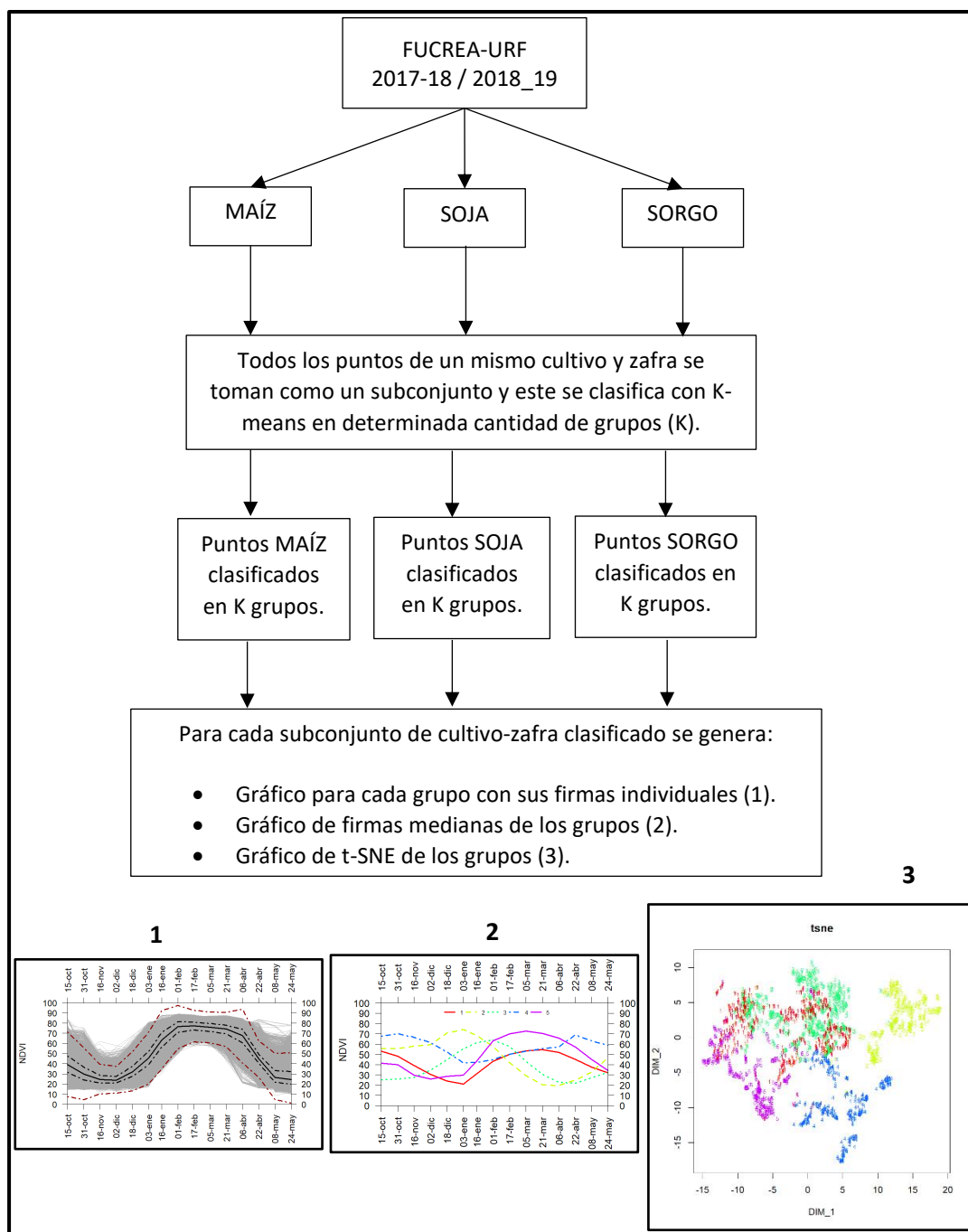


Figura 53. Flujograma de clasificación no supervisada de puntos regulares de chacras de FUCREA-URF.

Para cada subconjunto de zafra-cultivo se analiza el gráfico de cada grupo, el de firmas medianas y el de t-SNE, y se eliminan aquellos grupos que su patrón no coincide con el esperado para un cultivo agrícola. Tomando como ejemplo el subconjunto de maíz para la zafra 2017-18, en la figura 54 se muestra el gráfico de firmas medianas de NDVI y el de t-SNE. Como se ve, este subconjunto fue clasificado en cinco grupos, de los cuales el 2 y 4 no tienen firma con forma agrícola. Además, al analizar el gráfico de t-SNE, los grupos 2 y 4 (marcados con elipse roja) se ven claramente apartados del resto de los grupos. De esta manera, todos los puntos pertenecientes a los grupos 2 y 4 no se consideran agrícolas, y se eliminan para entrenamiento.

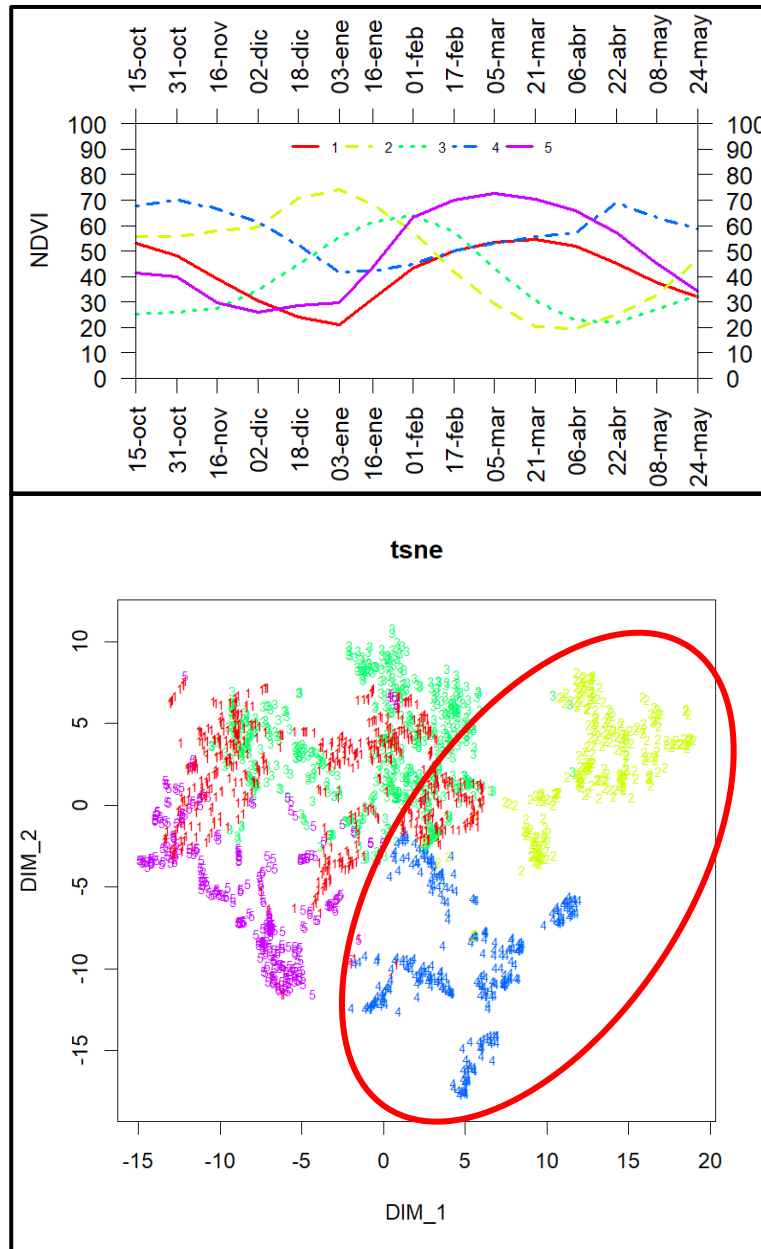


Figura 54. Gráficos de firmas medianas y t-SNE para subconjunto de maíz de zafra 2017-18.

Este proceso se repite para cada subconjunto de zafra-cultivo, de manera que al final se tiene un conjunto de datos conformados solo por aquellos puntos de los grupos con patrón agrícola. Después de realizada esta primera depuración, se realiza una segunda depuración que consiste en eliminar dentro de cada grupo aquellas firmas individuales, cuya forma se aleja de la firma mediana de su grupo, es decir que se consideran outliers. En la sección 9.2 esta depuración se hizo seleccionando todas aquellas firmas entre los percentiles 5 y 95 (P5-95), como se ve en el ejemplo de la figura 55.

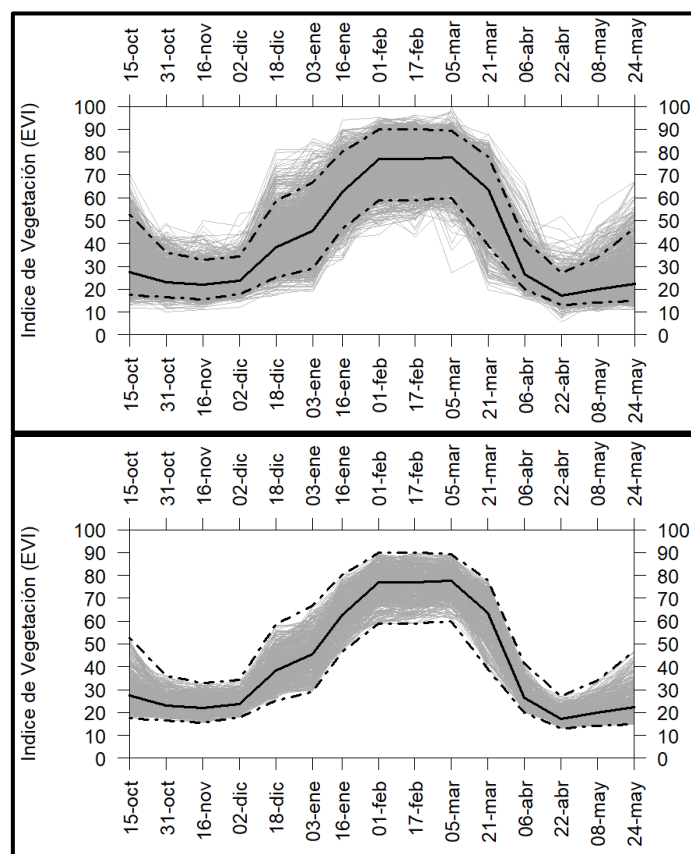


Figura 55. Gráfico de depuración de firmas individuales de PLANES entre percentiles 5 y 95 para un grupo de subconjunto de región Suroeste, soja y zafra 2018-19.

Cuando se hizo la depuración en la sección 9.2 de las firmas de PLANES usando los percentiles 5 y 95, se notó que la cantidad de firmas depuradas era importante. En el caso de PLANES, de un total de 35.422 firmas, se seleccionaron 13.739, es decir que el 61% de las firmas fueron eliminadas para entrenamiento. Debido a esto, en esta sección se decidió probar otra metodología para eliminar firmas outliers. En este caso se decidió usar el método de Tukey. Este fue introducido por Tukey en 1977 y se puede usar para datos que tenga distribuciones tanto simétricas, como asimétricas (Saleem, Aslam, & Shaukat, 2021). El método consiste, para los datos a analizar en calcular los percentiles 25 y 75, el rango intercuartílico (IQR), que es la diferencia entre los percentiles 25 y 75, y posteriormente calcular los límites superior e inferior de la distribución de la siguiente manera:

$$\text{Límite superior (LS)} = P75 + IQR \times \text{Factor}$$

$$\text{Límite inferior (LI)} = P25 - IQR \times \text{Factor}.$$

Para calcular los límites inferior y superior, como Factor multiplicador de IQR se usa generalmente valores de 1,5 o 3. En este caso se decidió usar 1,5. En la figura 56 se muestra una porción de un gráfico con las firmas individuales de una clase, los percentiles 25 y 75, y los límites inferior y superior. Para cada grupo considerado agrícola, se calculan los límites superior e inferior de cada fecha, y se construyen las líneas de límite superior e inferior. Una vez hecho esto, en cada grupo aquellas firmas que están por encima del límite superior o por debajo del límite inferior en alguna de las fechas se consideran outliers y se descartan para entrenamiento, como se muestra en el ejemplo de la figura 57.

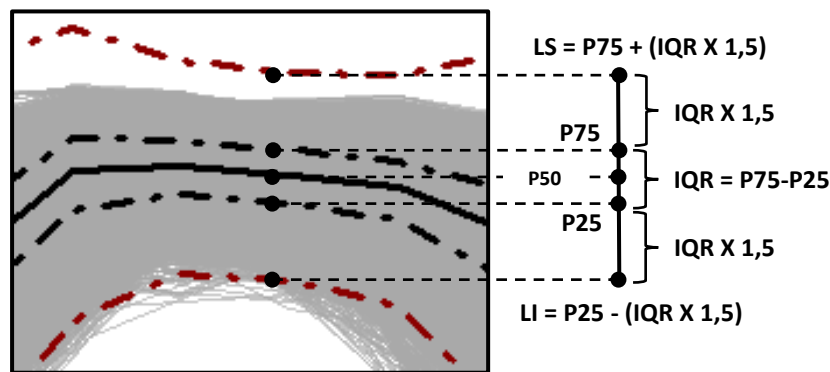


Figura 56. Gráfico de percentiles 25, 50, 75 y límites superior e inferior en un gráfico de un grupo.

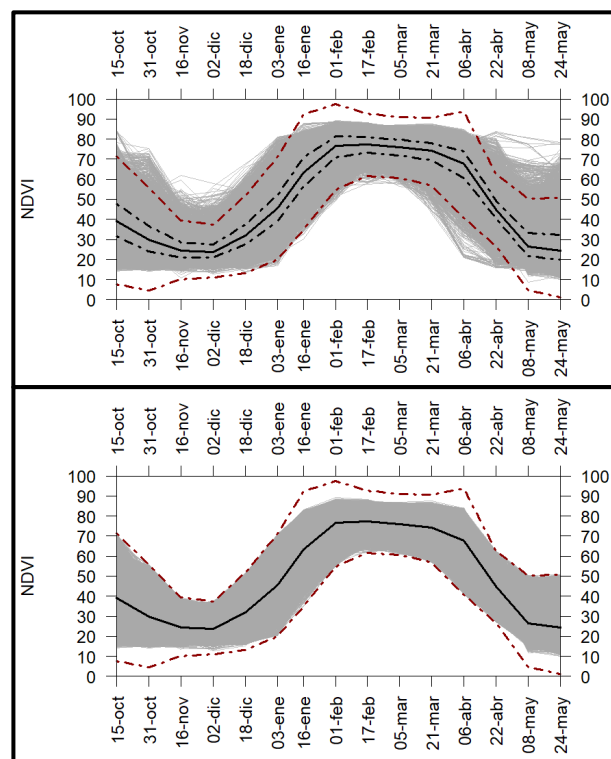


Figura 57. Gráficos de un grupo para el subconjunto de maíz de zafra 2017-18 con las líneas de los percentiles (25, 50 y 75) y límites superior e inferior.

La segunda depuración se hizo tanto con el método P5-95 y Tukey para comparar los resultados. En la tabla 32, se muestra la cantidad de puntos totales (sin depurar), la cantidad después de la primera depuración, y la cantidad de puntos después de la segunda depuración con ambos métodos. Como se ve en la tabla, de un total de 76.069 puntos, el método P5-95 selecciona 31.312 (41%) y el de Tukey 62.332 (82%). Por lo tanto, el método de Tukey es menos exigente que el de P5-95 para eliminar outliers. Visto esto, para tener más datos para entrenar se decidió usar los datos depurados por Tukey. De esta manera se dispone de 62.332 puntos para entrenamiento.

CULTIVO	PUNTOS			
	SIN DEPURAR	1ER DEPURACIÓN	2DA DEPURACIÓN	
			P5-95	TUKEY (IQR X 1.5)
MAÍZ	4.653	3.728	1.551	3.196
SOJA	87.677	70.069	28.799	57.217
SORGO	3.338	2.272	962	1.919
TOTAL	95.668	76.069	31.312	62.332

Tabla 32. Cantidad de puntos por cultivo de las zafras 2017-18 y 2018-19 de FUCREA-URF según depuración usada.

Como se observa en la tabla 32, en Tukey la base está muy desbalanceada en cuanto a la cantidad de puntos por cultivo, siendo un 91% de los datos correspondientes a soja. Por lo tanto, se decidió balancear la base, seleccionando un 10% de los datos de soja. De esta manera se tienen para trabajar, 1.919 puntos de sorgo, 3.196 de maíz y 5.721 de soja como se ve en la tabla 33. Estos puntos fueron utilizados en la siguiente sección para realizar los diferentes experimentos evaluados.

CULTIVO	PUNTOS
MAÍZ	3.196
SOJA	5.721
SORGO	1.919
TOTAL	10.836

Tabla 33. Balanceo de puntos de la base FUCREA-URF.

10.2.2 Experimentos evaluando bandas, índices, fechas y algoritmos

En la sección 8.2, para cada banda e índice de Sentinel se extrajo una tabla con las series temporales de los puntos regulares dentro de cada chacra de FUCREA-URF, teniendo un total de 3 tablas para Sentinel 1 (S1) y 12 para Sentinel 2 (S2). Un ejemplo de tabla es el extracto para el índice NDVI que se ve en la tabla 34. Los encabezados de la tabla indican la banda o índice y a qué período temporal corresponde, por ejemplo, NDVI_1015 indica que es el valor de NDVI para el período entre el 30 de setiembre y 15 de octubre (1015) y así sucesivamente. Los períodos son los correspondientes a los que aparecen en la tabla 20 de la sección 8.2. Cada fila de la tabla corresponde a la serie temporal de un punto dentro de alguna de las chacras de la base.

NDVI_1015	NDVI_1031	NDVI_1116		NDVI_0422	NDVI_0508	NDVI_0524
26,4	34,2	49,3		27,7	25,6	23,6
29,7	34,9	48,4		33,2	37,5	41,9
29,6	35,0	47,0		34,1	33,8	33,5
27,8	33,6	47,2		28,0	33,8	39,7
25,7	31,4	45,6		28,5	34,8	41,0

Tabla 34. Ejemplo de tabla de NDVI con los valores por fecha de imagen sintética.

Las tablas de las bandas e índices se pueden disponer en el formato de matriz multidimensional o también llamado tensor (ver figura 58), pero los algoritmos evaluados no tienen la capacidad de trabajar con tensores, solo pueden trabajar con matrices de dos dimensiones (filas y columnas), por lo que el primer paso previo a evaluarlos fue unir todas las tablas de las bandas e índices en una única matriz, que se denominó array multibanda, como se ve en la figura 59. De esta manera, a partir de 12 tablas para las bandas e índices de S2, y tres tablas (bandas VH,

VV y CROSS-RATIO VHVV) para S1, en donde cada tabla tiene 15 columnas, una por período de fechas analizado, se construyó el array multibanda que se compone de 225 variables o features como se ve en la tabla 35.

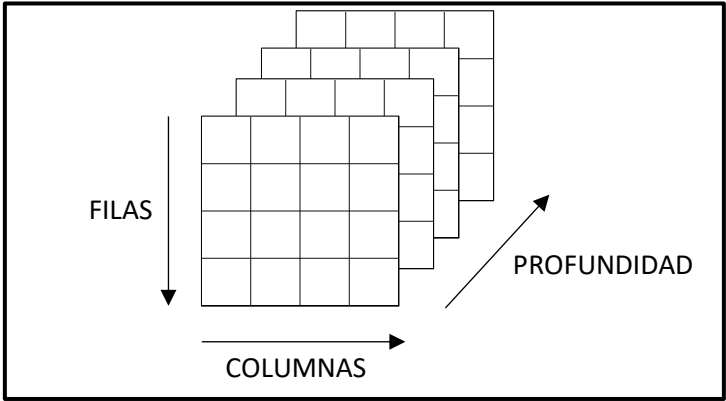


Figura 58. Matriz multidimensional o tensor.

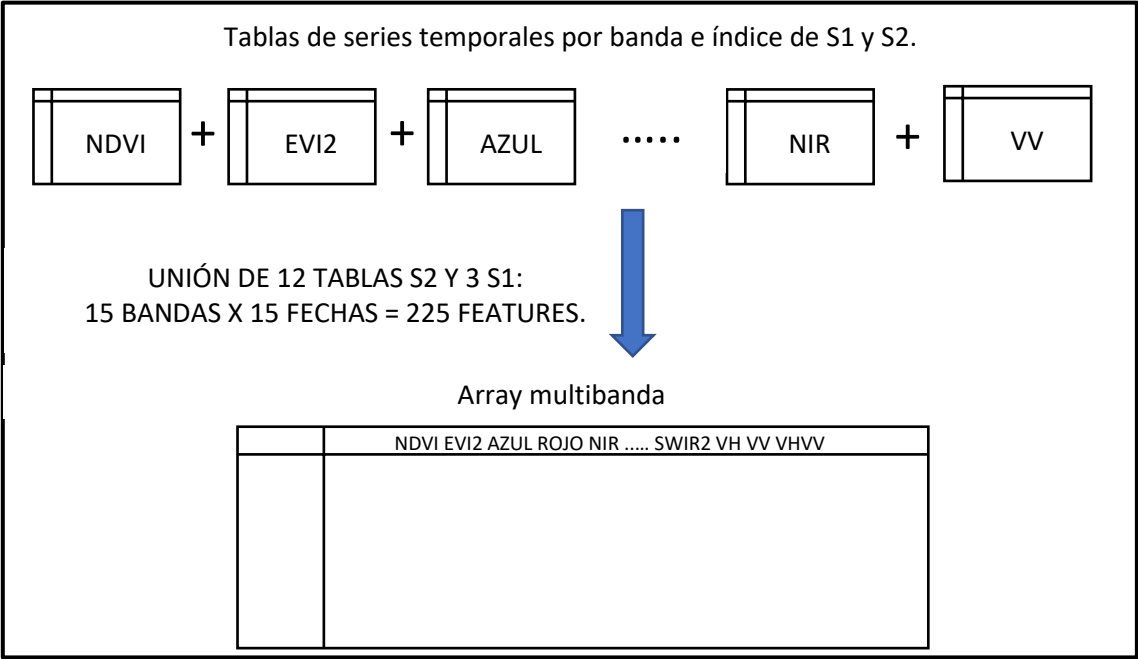


Figura 59. Unión de tablas de series temporales por banda e índice en un array multibanda.

FECH A	BANDA														
	NDVI	EVI2	GREEN	NIR	RED	BLUE	SWIR 1	SWIR 2	RED EDGE 1	RED EDGE 2	RED EDGE 3	RED EDGE 4	VH	VV	CROSS RATIO VH/VV
OCT-15	NDVI_1015	EVI2_1_015	GREEN_1015	NIR_1_015	RED_1_015	BLUE_1015	SWIR1_1015	SWIR2_1015	REDGE1_1015	REDGE2_1015	REDGE3_1015	REDGE4_1015	VH_1_015	VV_1_015	VHVV_1015
OCT-31	NDVI_1031	EVI2_1_031	GREEN_1031	NIR_1_031	RED_1_031	BLUE_1031	SWIR1_1031	SWIR2_1031	REDGE1_1031	REDGE2_1031	REDGE3_1031	REDGE4_1031	VH_1_031	VV_1_031	VHVV_1031
NOV-16	NDVI_1116	EVI2_1_116	GREEN_1116	NIR_1_116	RED_1_116	BLUE_1116	SWIR1_1116	SWIR2_1116	REDGE1_1116	REDGE2_1116	REDGE3_1116	REDGE4_1116	VH_1_116	VV_1_116	VHVV_1116
DIC-02	NDVI_1202	EVI2_1_202	GREEN_1202	NIR_1_202	RED_1_202	BLUE_1202	SWIR1_1202	SWIR2_1202	REDGE1_1202	REDGE2_1202	REDGE3_1202	REDGE4_1202	VH_1_202	VV_1_202	VHVV_1202
DIC-18	NDVI_1218	EVI2_1_218	GREEN_1218	NIR_1_218	RED_1_218	BLUE_1218	SWIR1_1218	SWIR2_1218	REDGE1_1218	REDGE2_1218	REDGE3_1218	REDGE4_1218	VH_1_218	VV_1_218	VHVV_1218
ENE-03	NDVI_0103	EVI2_0_103	GREEN_0103	NIR_0_103	RED_0_103	BLUE_0103	SWIR1_0103	SWIR2_0103	REDGE1_0103	REDGE2_0103	REDGE3_0103	REDGE4_0103	VH_0_103	VV_0_103	VHVV_0103
ENE-16	NDVI_0116	EVI2_0_116	GREEN_0116	NIR_0_116	RED_0_116	BLUE_0116	SWIR1_0116	SWIR2_0116	REDGE1_0116	REDGE2_0116	REDGE3_0116	REDGE4_0116	VH_0_116	VV_0_116	VHVV_0116
FEB-01	NDVI_0201	EVI2_0_201	GREEN_0201	NIR_0_201	RED_0_201	BLUE_0201	SWIR1_0201	SWIR2_0201	REDGE1_0201	REDGE2_0201	REDGE3_0201	REDGE4_0201	VH_0_201	VV_0_201	VHVV_0201
FEB-17	NDVI_0217	EVI2_0_217	GREEN_0217	NIR_0_217	RED_0_217	BLUE_0217	SWIR1_0217	SWIR2_0217	REDGE1_0217	REDGE2_0217	REDGE3_0217	REDGE4_0217	VH_0_217	VV_0_217	VHVV_0217
MAR-05	NDVI_0305	EVI2_0_305	GREEN_0305	NIR_0_305	RED_0_305	BLUE_0305	SWIR1_0305	SWIR2_0305	REDGE1_0305	REDGE2_0305	REDGE3_0305	REDGE4_0305	VH_0_305	VV_0_305	VHVV_0305
MAR-21	NDVI_0321	EVI2_0_321	GREEN_0321	NIR_0_321	RED_0_321	BLUE_0321	SWIR1_0321	SWIR2_0321	REDGE1_0321	REDGE2_0321	REDGE3_0321	REDGE4_0321	VH_0_321	VV_0_321	VHVV_0321
ABR-06	NDVI_0406	EVI2_0_406	GREEN_0406	NIR_0_406	RED_0_406	BLUE_0406	SWIR1_0406	SWIR2_0406	REDGE1_0406	REDGE2_0406	REDGE3_0406	REDGE4_0406	VH_0_406	VV_0_406	VHVV_0406
ABR-22	NDVI_0422	EVI2_0_422	GREEN_0422	NIR_0_422	RED_0_422	BLUE_0422	SWIR1_0422	SWIR2_0422	REDGE1_0422	REDGE2_0422	REDGE3_0422	REDGE4_0422	VH_0_422	VV_0_422	VHVV_0422
MAY-08	NDVI_0508	EVI2_0_508	GREEN_0508	NIR_0_508	RED_0_508	BLUE_0508	SWIR1_0508	SWIR2_0508	REDGE1_0508	REDGE2_0508	REDGE3_0508	REDGE4_0508	VH_0_508	VV_0_508	VHVV_0508
MAY-24	NDVI_0524	EVI2_0_524	GREEN_0524	NIR_0_524	RED_0_524	BLUE_0524	SWIR1_0524	SWIR2_0524	REDGE1_0524	REDGE2_0524	REDGE3_0524	REDGE4_0524	VH_0_524	VV_0_524	VHVV_0524

Tabla 35. Features del array multibanda por fecha y banda.

El array multibanda se construyó para las zafras 2017-18 y 2018-19. Para el conjunto de puntos depurado y balanceado en la sección anterior se extrajo la información correspondiente del array multibanda, y con ésta se realizaron varios experimentos en donde se evaluaron diferentes combinaciones de bandas e índices, períodos de fechas y algoritmos de clasificación supervisada. Los algoritmos evaluados fueron los mismos que para PLANES, es decir AdaBoost, Random Forest, SVM y XGBoost. Los puntos se dividieron de manera estratificada en 70% para entrenamiento y 30% para evaluación. De esta manera todas las clases (maíz, soja y sorgo) están presentes en ambos subconjuntos en la misma proporción que en el conjunto original.

El primer experimento que se llevó a cabo, fue determinar cuáles eran las mejores combinaciones de bandas e índices a usar para obtener la mejor clasificación de los cultivos. Para todas las bandas evaluadas, se usaron las 15 fechas disponibles. Por ejemplo, si se evalúa solo NDVI, se usan las 15 fechas de NDVI, si se evalúa NDVI + AZUL, se usan las 15 fechas de NDVI y de la banda azul, dando un total de 30 features, y así sucesivamente. Este experimento se corrió usando Random Forest con 100 árboles de decisión, max_features = “auto” y clases balanceadas por peso. Las combinaciones de bandas e índices evaluados se muestran en la tabla 36. En ésta se presentan, para las diferentes combinaciones evaluadas, los valores de *F1-Score* discriminados por cultivo, y un valor promedio de todos los cultivos. Los datos se muestran ordenados por valor promedio de *F1-Score*.

Como se observa en la figura 60 y tabla 36, se obtiene mejor resultado al usar una combinación de bandas que usar solo bandas individuales, ya que los valores más bajos de *F1* se alcanzan cuando se usa solo una banda, con la excepción del índice cross-ratio VH/VV que siendo una combinación de bandas a través de un índice, es la peor de todas, con un valor promedio de *F1* apenas superior a 70, y entre los cultivos, sorgo apenas llega a un *F1* de casi 53.

Por otro lado, la mejor combinación se obtiene al usar todas las bandas ópticas de S2, los índices de vegetación (EVI2 y NDVI) y las bandas de S1 (VH y VV), en donde el valor de *F1* es mayor a 95 para cualquiera de los cultivos. Al mismo tiempo, se observa que usar la combinación anterior más cross-ratio VH/VV, disminuye un poco el desempeño, por lo que usar este aporta información que empeora el desempeño.

Analizando el uso de las bandas individuales, las de mejor desempeño corresponden a las de radar (VH y VV), y de las ópticas, las bandas SWIR 1 y 2, mientras que la verde es la peor. Para el caso de los índices de vegetación, NDVI tiene un desempeño un poco mejor que EVI2 y el uso de todas las ópticas de S2. El uso de ambos índices combinados da un desempeño un poco mejor que el uso por separado de estos.

BANDA	F1-SCORE			
	PROMEDIO	MAÍZ	SOJA	SORGO
EVI2+NDVI+S1+S2	96,3	95,6	96,8	95,9
EVI2+NDVI+S1+S2+S1CR	96,0	95,3	96,4	95,8
EVI2+NDVI+S1	95,8	95,0	96,4	95,2
EVI2+NDVI+S1+S1CR	95,6	94,4	96,1	95,7
EVI2+NDVI+S2	95,5	95,3	96,0	94,4
EVI2+NDVI+S2+S1CR	95,3	94,3	95,8	95,1
S1+S2	94,7	94,4	95,4	93,0
S1+S2+S1CR	94,3	94,1	95,1	92,3
EVI2+NDVI+S1CR	93,9	92,6	94,8	93,2
EVI2+NDVI	93,8	92,5	94,7	93,1
NDVI	93,6	92,3	94,6	92,5
S1	93,5	92,9	94,4	91,5
EVI2	92,4	90,9	93,6	91,1
S2	92,2	90,4	93,6	90,6
S1+S1CR	90,8	90,0	92,4	86,3
VH	90,6	89,8	92,3	86,3
VV	88,6	87,9	91,1	81,1
SWIR2	86,3	87,3	88,9	74,6
SWIR1	85,9	87,4	88,5	72,5
AZUL	84,5	84,6	87,0	74,3
ROJO	83,9	84,8	86,8	69,8
NIR	83,9	80,4	87,2	77,9
REDGE4	83,8	80,8	87,0	77,7
REDGE3	83,5	80,7	86,8	76,6
REDGE1	83,4	84,5	86,8	67,4
REDGE2	82,6	79,3	86,0	75,6
VERDE	82,3	80,7	85,7	72,6
S1CR	71,9	66,6	78,4	52,8

S2: AZUL + NIR + ROJO + REDGE (1-4) + SWIR (1&2) + VERDE. S1: VH + VV. S1CR: CROSS-RATIO VH/VV.

Tabla 36. *F1-Score* en el conjunto de evaluación de diferentes combinaciones de bandas e índices del array multibanda.

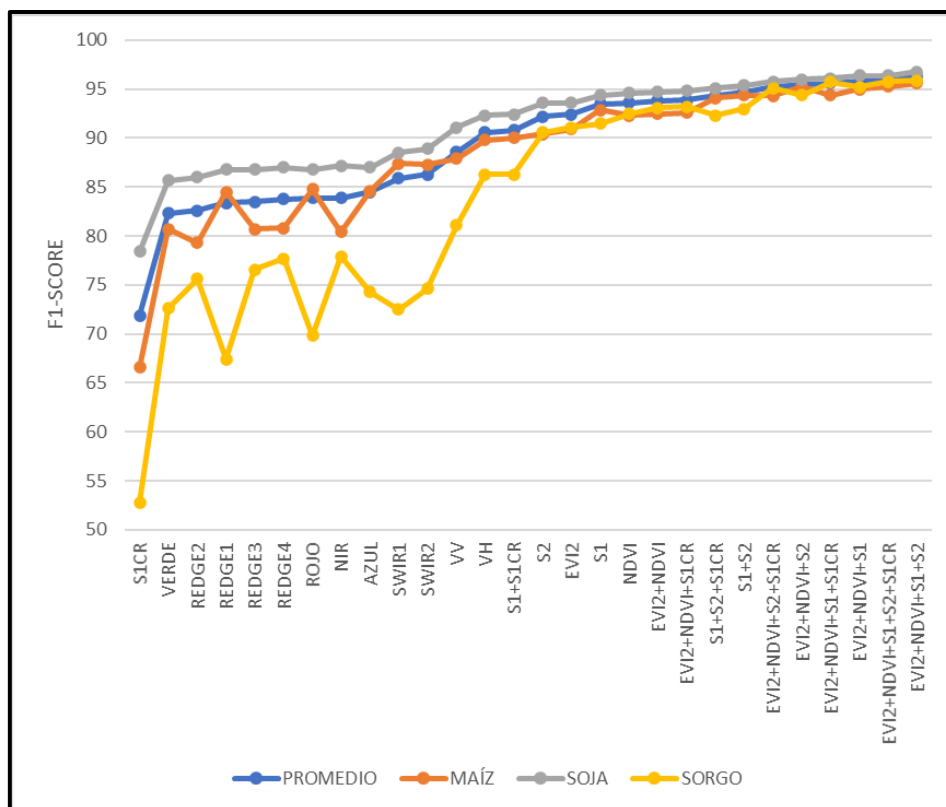


Figura 60. Gráfico de *F1-Score* en el conjunto de evaluación de diferentes combinaciones de bandas e índices del array multibanda.

El siguiente experimento fue evaluar el desempeño de Random Forest para diferente cantidad de períodos. Este experimento se realizó usando la combinación de las bandas de S2+EVI2+NDVI. Para esto, se evaluaron 9 combinaciones de períodos, todos comenzando el 15/10 y con las siguientes fechas de cierre: 16/01, 01/02, 17/02, 05/03, 21/03, 06/04, 22/06, 08/05 y 24/05. De esta manera, la primera combinación de períodos va del 15/10 y cierra el 16/01, el segundo va del 15/10 al 01/02, y así sucesivamente hasta el último, que va del 15/10 al 24/05. Como se observa en la figura 61 y en la tabla 37, el mejor desempeño para cualquier de los cultivos, es cuando el período llega hasta el 24/05, con valores de *F1* en torno a 95 para todos los cultivos. Mientras que si el período termina el 16/01, *F1* para maíz y soja está en torno a 90, mientras que sorgo baja bastante, con un valor de 78,4. A partir de la fecha de cierre del 06/04 es cuando todos los cultivos tienen un valor de *F1* superior a 90. Del gráfico de la figura 61 y la tabla 37 se observa que para sorgo, en la medida que se toman más períodos, su desempeño mejora bastante. De este experimento se concluye que, a partir del período de cierre del 6 abril en adelante, se obtiene valores de *F1* superiores a 90, lo que coincide con el período en donde los cultivos agrícolas ya están en etapa de floración.

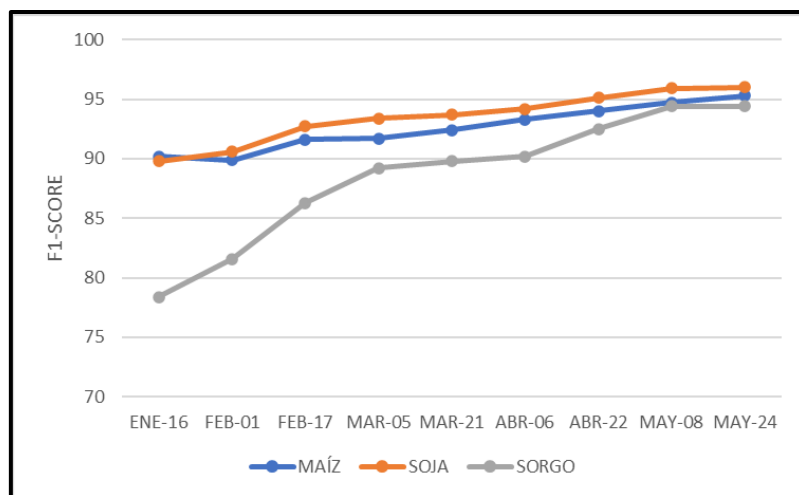


Figura 61. Gráfico de *F1-Score* según la fecha de cierre del último período utilizado.

FECHA DE CIERRE	F1-SCORE		
	MAÍZ	SOJA	SORGO
ENE-16	90,2	89,8	78,4
FEB-01	89,9	90,6	81,6
FEB-17	91,6	92,7	86,3
MAR-05	91,7	93,4	89,2
MAR-21	92,4	93,7	89,8
ABR-06	93,3	94,2	90,2
ABR-22	94,0	95,1	92,5
MAY-08	94,7	95,9	94,4
MAY-24	95,3	96,0	94,4

Tabla 37. *F1-Score* según la fecha de cierre del último período utilizado.

El siguiente experimento que se realizó fue analizar diferentes algoritmos supervisados como AdaBoost, Random Forest, SVM con diferentes kernels (lineal, polinomial, sigmoide y radial base function) y XGBoost. Para este experimento se usó la combinación de bandas S2+EVI2+NDVI. Como se ve en la figura 62 y tabla 38, para cualquiera de los cultivos, Random Forest y XGBoost son los algoritmos que mejor se desempeñan, con valores de *F1* superiores a 94, siendo XGBoost algo mejor que Random Forest. Del resto de los algoritmos, el que mejor se desempeña es SVM con kernel lineal, con valores de *F1* de 88,4 para maíz y superior a 92 para soja y sorgo. AdaBoost tiene valores de *F1* algo superior a 75 para maíz y sorgo, y de 85,5 para soja. De este experimento se concluye que el mejor resultado se tiene al usar Random Forest o XGBoost.

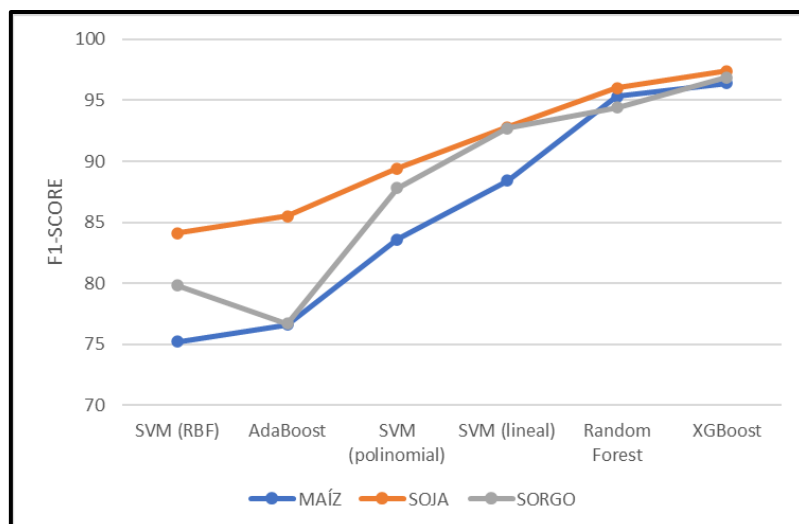


Figura 62. Gráfico de *F1-Score* según cultivo y algoritmo supervisado evaluado.

ALGORITMO	F1-SCORE		
	MAÍZ	SOJA	SORGO
XGBoost	96,4	97,4	96,9
Random Forest	95,3	96,0	94,4
SVM (lineal)	88,4	92,8	92,7
SVM (polinomial)	83,6	89,4	87,8
AdaBoost	76,6	85,5	76,7
SVM (RBF)	75,2	84,1	79,8
SVM (sigmoide)	0,0	69,2	0,0

Tabla 38. *F1-Score* según cultivo y algoritmo supervisado evaluado.

El último experimento fue evaluar el desempeño de Random Forest para la combinación S2+EVI2+NDVI, tomando un solo período a la vez e ir incrementando de a uno hasta tener los 15 períodos para la combinación analizada. Se tomó como período central de las series temporales el de cierre el 21/03, y a partir de este se fueron tomando dos, tres, hasta tener los 15 períodos, como se ve en la tabla 39. Se tomó 21/03 como período central, ya que en ese mes los cultivos empiezan a florecer o están florecidos. Si tomamos solamente los datos del período 21/03, el valor de *F1* para maíz, soja, y sorgo, es de 78, 68 y 85 respectivamente. A medida que se van incrementando los períodos utilizados, el *F1* se va incrementando, y a partir de siete períodos, el *F1* para cualquier clase es superior a 90 y se estabiliza, como se ve en la figura 63.

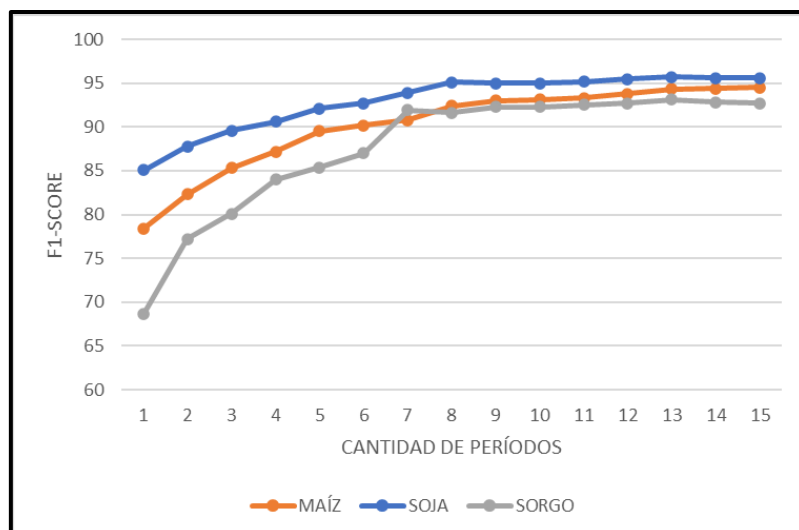


Figura 63. Gráfico de *F1-Score* por cultivo según cantidad de períodos usados para entrenar el algoritmo.

PERIODO		CANTIDAD DE PERIODOS USADOS														
INICIO	FIN	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
30-set	15-oct															
16-oct	31-oct															
01-nov	16-nov															
17-nov	02-dic															
03-dic	18-dic															
19-dic	03-ene															
01-ene	16-ene															
17-ene	01-feb															
02-feb	17-feb															
18-feb	05-mar															
06-mar	21-mar															
22-mar	06-abr															
07-abr	22-abr															
23-abr	08-may															
09-may	24-may															

Tabla 39. Cantidad y combinaciones de períodos de fechas utilizados para entrenar el algoritmo.

10.2.3 Clasificación de chacras

Los experimentos de la sección anterior sirvieron para evaluar diferentes combinaciones de bandas, períodos de fechas y algoritmos a utilizar para clasificar los cultivos de la base FUCREA-URF. Esos experimentos se realizaron con todos los puntos de la base sin considerar la pertenencia de estos a una misma chacra. En esta sección, para la combinación elegida, se quiere evaluar el desempeño al realizar la clasificación supervisada por chacra y no por puntos como se hizo en la sección anterior. En base a los análisis realizados, del array multibanda se decidió utilizar todas las columnas de todas las bandas e índices de Sentinel 2 y las dos bandas

polarimétricas de Sentinel 1, para los 15 períodos de octubre a mayo, teniendo así un total de 210 features. El algoritmo elegido fue XGBoost, que fue el de mejor desempeño en los experimentos realizados. Este se corrió con 300 árboles de decisión y una tasa de aprendizaje de 0,025.

Hay que recordar que la base de FUCREA-URF está muy desbalanceada, con 92% de las chacras siendo soja, 4,4% maíz y 3,6% sorgo, como se ve en la tabla 40.

CULTIVO	CHACRAS		HECTÁREAS	PUNTOS
	CANTIDAD	PORCENTAJE		
MAÍZ	103	4,4	4.744	4.653
SOJA	2.177	92	87.673	87.677
SORGO	84	3,6	3.326	3.338
TOTAL	2.364	100	95.743	95.668

Tabla 40. Cantidad de chacras, hectáreas y puntos por cultivo de FUCREA-URF.

Como se explicó en la sección 10.2.1, la base fue depurada para eliminar puntos de chacras sin patrón agrícola. Esto se hizo con una doble depuración, primero seleccionando grupos de puntos con firma agrícola, y posteriormente eliminando puntos con firmas outliers en cada grupo. Esta segunda depuración se hizo utilizando el método de Tukey, con un factor multiplicador de IQR de 1,5. En esta sección, a efectos de ser menos exigentes para eliminar firmas outliers, y tener más puntos para entrenamiento, se probaron los factores de 2 y 3 para multiplicar IQR, como se observa en la tabla 41. Finalmente se decidió trabajar con un factor 2, ya que de los puntos de la primera depuración elimina un 10% de estos, teniéndose un total de 68.750 puntos para entrenar.

CULTIVO	CHACRAS	PUNTOS				
		SIN DEPURAR	1ER DEPURACIÓN	2DA DEPURACIÓN		
				TUKEY (FACTOR X IQR)		
				1,5	2	3
MAÍZ	103	4.653	3.728	3.196	3.485	3.655
SOJA	2.177	87.677	70.069	57.217	63.164	67.747
SORGO	84	3.338	2.272	1.919	2.101	2.243
TOTAL	2.364	95.668	76.069	62.332	68.750	73.645

Tabla 41. Cantidad de puntos outliers depurados con diferentes metodologías de FUCREA-URF.

Después de seleccionar los puntos filtrados con factor IQR de 2, como la base está desbalanceada, para clasificación se dividieron las chacras de manera estratificada por cultivo en entrenamiento y evaluación, en porcentajes de 80 y 20 respectivamente. Posteriormente, XGBoost del conjunto de entrenamiento fue entrenado con todos los puntos de las chacras de maíz y sorgo, y un 10% de los puntos de las chacras de soja extraídos aleatoriamente. Esto se hizo así para que la cantidad de puntos estuviera balanceada y para que XGBoost corriera más rápido. En la etapa de evaluación, todos los puntos de las chacras fueron clasificados, y finalmente la clasificación final de cada chacra se determinó por voto mayoritario. Por ejemplo, si en una chacra hay 20 puntos, y 10 son clasificados soja, 5 maíz y 5 sorgo, esa se clasifica como soja, que es la clase mayoritaria. En el flujograma de la figura 64 se resume todo el proceso y en la figura 65 se muestra un ejemplo de clasificación de una chacra por voto mayoritario.

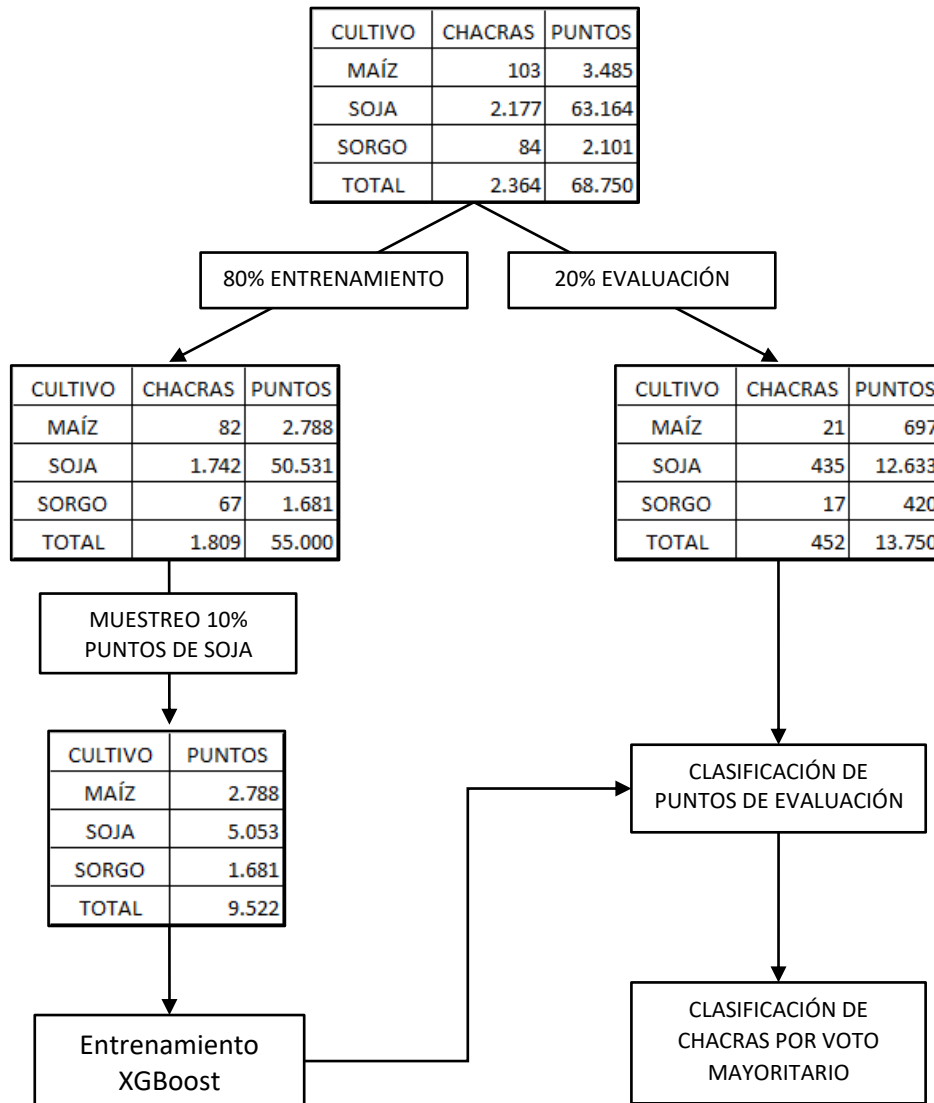


Figura 64. Flujograma de división de chacras para entrenamiento y evaluación de FUCREA-URF.



Figura 65. Clasificación de una chacra por voto mayoritario de FUCREA-URF.

Al hacer varias corridas del flujograma de la figura 64 se notó que los valores de las métricas de desempeño variaban mucho para las clases minoritarias (maíz y sorgo), lo cual puede estar explicado por la poca cantidad de chacras de estos cultivos para dividir en entrenamiento y evaluación como se ve en la tabla 41. Debido a esto, se decidió realizar 50 corridas, en donde en

cada una se repetía el proceso de la figura 64, y presentar el promedio y desvío estándar de todas las corridas de cada métrica por cultivo. En la tabla 42, se muestra la matriz de confusión de la mejor corrida, y en la tabla 43 el promedio y desvío de las métricas de todas las corridas. Así mismo, en la figura 66, se presentan los valores de $F1$ obtenidos en evaluación en cada corrida.

		CLASIFICADO			TOTAL	RECALL	F1-SCORE
		MAÍZ	SOJA	SORGO			
OBSERVADO	MAÍZ	15	3	0	18	83,3	85,7
	SOJA	2	404	2	408	99,0	98,9
	SORGO	0	2	12	14	85,7	85,7
TOTAL		17	409	14	440		
PRECISION		88,2	98,8	85,7			

Tabla 42. Matriz de confusión en evaluación de la mejor corrida de FUCREA-URF.

CULTIVO	PRECISION	RECALL	F1-SCORE
MAÍZ	85,1 (8,4)	68,2 (10,3)	75,3 (8,3)
SOJA	97,9 (0,5)	99,0 (0,5)	98,5 (0,4)
SORGO	80,9 (9,8)	74,9 (9,4)	77,5 (8,2)

Tabla 43. Valores promedio y desvío estándar de las métricas de desempeño para las 50 corridas en evaluación de FUCREA-URF.

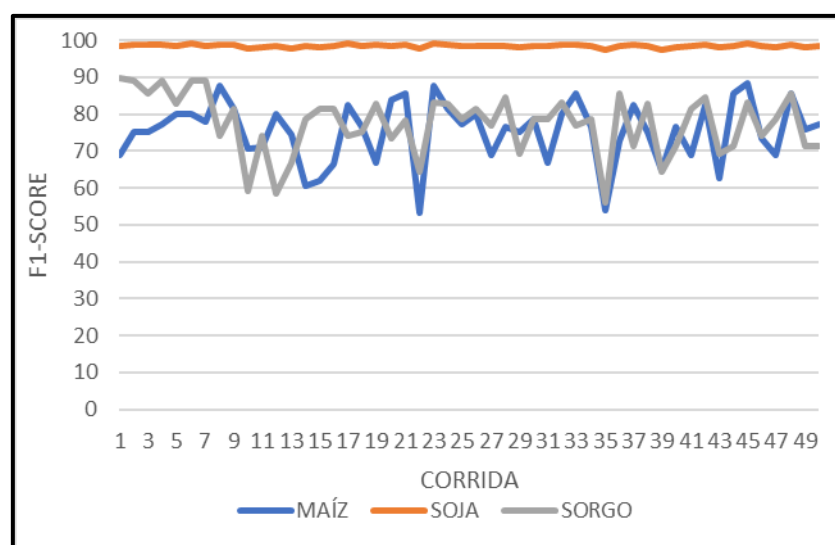


Figura 66. Gráfico de $F1$ -Score por cultivo en evaluación para las 50 corridas de FUCREA-URF.

Como se observa en la figura 66, $F1$ para soja tiene valores cercanos a 100, con poca variación entre corridas, en cambio para maíz y sorgo, $F1$ es inferior a soja, siendo muy variable, con valores entre 50 y 90 según la corrida.

Al analizar *precision* de la tabla 43 se nota que esta es muy alta para soja, con un valor de casi 98 y un desvío de 0,5, y con valores de 85,1 y 80,9 para maíz y sorgo respectivamente. Además, en maíz y sorgo, los valores de desvío son mucho mayores en comparación a soja, con valores entre 8 y 10. Esto quiere decir que el algoritmo, de aquellas chacras que clasifica como maíz, soja o sorgo, tiene mejor capacidad de acierto para soja, y además en este caso es estable y tiene poca variación entre corridas.

En el caso de *recall*, la soja tiene un valor de 99, con un desvío de 0,5, mientras que para maíz y sorgo, este es de 68,2 y 74,9 respectivamente, con un alto desvío en torno a 10. Esto indica que de aquellas chacras del conjunto de evaluación que en la realidad son maíz, soja o sorgo, el algoritmo tiene muy alto desempeño para identificar las que son soja, con poca variación entre corridas, mientras que para maíz y sorgo su desempeño es bastante inferior, con mucha variabilidad, y con peor desempeño en maíz.

Al analizar *precision* y *recall* conjuntamente se observa que estos tienen valores similares para soja (97,9 y 99) y sorgo (80,9 y 74,9), mientras que en el caso de maíz, tiene mucho mejor *precision* que *recall*, una diferencia de 16,9 puntos, lo que indica que el algoritmo en este cultivo tiene mejor capacidad para clasificar una chacra como maíz, pero es peor su desempeño para identificar todas las chacras que son maíz.

De los valores de *F1* se observa que el algoritmo elegido tiene muy alta capacidad para identificar chacras que son soja, ya que su valor es de 98,5, y a su vez es estable porque tiene un desvío de 0,4, mientras que para maíz y sorgo su desempeño comparativamente baja bastante, con valores entre 75 y 78, y con mayor variabilidad entre corridas, con valores del desvío en torno a 8.

Los valores obtenidos de *F1-Score* para cada cultivo se compararon contra valores obtenidos en artículos internacionales. La mayoría de los artículos encontrados clasifican maíz y/o soja, y solo dos además sorgo. En la tabla 44 se presentan los valores obtenidos en este y otros trabajos. Hay que tener en cuenta que los artículos difieren en las imágenes utilizadas, algoritmos de clasificación y metodología. Aun así, la comparativa sirve para tener una referencia de desempeño de la metodología desarrollada en este trabajo.

AUTOR	F1-SCORE			IMÁGENES	PAÍS	CLASIFICADOR	AÑO PUBLICACIÓN
	MAÍZ	SOJA	SORGO				
(Bargiel, 2017)	94,0			Sentinel 1	Alemania	Random Forest	2017
(LUO et al., 2021)	91,0	85,0		Sentinel 1 y 2	China	Random Forest	2021
(Vuolo, Neuwirth, Immitzer, Atzberger & Ng, 2018)	87,5	84,3		Sentinel 2	Austria	Random Forest	2018
(You & Dong, 2020)	91,0	91,0		Sentinel 1 y 2	China	Random Forest	2020
(Zhong, Hu & Zhou, 2019)	68,2			Landsat 7 y 8	Estados Unidos	Red Neuronal Convolucional	2019
(Mohite, Sawant, Pandit & Pappula, 2020)		79,0		Sentinel 1 y 2	India	Red Neuronal Profunda	2020
(Kumari, Pandey, Choudhary & Murthy, 2021)		87,2		Sentinel 1 y 2	India	XGBoost	2021
(Zhang et al., 2021)	96,1	93,9		Landsat 8 y Sentinel 2	Estados Unidos	Red Neuronal	2021
(Tian, Wu, Zeng, Zhang & Xu, 2019)	90,1			Sentinel 1 y 2	China	Random Forest	2019
(Cué La Rosa, Borges Oliveira & Feitosa, 2021)	50,0		62,3	Sentinel 1	Brasil	Red Neuronal Convolucional y Campo Aleatorio Condicional	2021
(Chamorro, Bermudez, Happ & Feitosa, 2019)	91,3	98,3	72,2	Sentinel 1	Brasil	Red Neuronal Recurrente	2019
PROMEDIO ARTICULOS	84,4	88,4	67,3				
PRESENTE TRABAJO	75,3	98,5	77,5	Sentinel 1 y 2	Uruguay	XGBoost	

Tabla 44. Comparativa de *F1-Score* de maíz, soja y sorgo para el presente trabajo y artículos revisados.

Es importante destacar que los artículos revisados son actuales, de 2017 a 2021, y estos usan las últimas metodologías y herramientas disponibles. La mayoría utilizaron imágenes de Sentinel 1 y/o 2, y uno Landsat (7 y 8). De los artículos revisados, cinco utilizaron Random Forest, cinco Redes Neuronales, otro Campo Aleatorio Condicional y otro XGBoost al igual que en el presente trabajo.

De la tabla se observa que para maíz, *F1* de la mayoría de los artículos está entre 87 y 94, con un promedio de 84,4, unos 9,1 puntos por encima del valor de 75,1 obtenido en el presente trabajo. Para el caso de soja, los artículos tienen valores entre 79 y 91, con un promedio de 88,4, mientras que el presente trabajo tiene un valor de 98,5, unos 10,1 puntos por encima. Para sorgo, en este trabajo se obtuvo 77,5, 10,2 puntos superior al promedio de 67,3 de los artículos que clasificaron sorgo. De esto se concluye que en comparación a los artículos revisados, la metodología del presente trabajo tiene peor desempeño para clasificar maíz, pero mejor desempeño para clasificar soja y sorgo.

10.3 Clasificación de PLANES con series temporales de Sentinel

En esta sección se presentan los resultados de la clasificación supervisada realizada utilizando la base PLANES con series temporales de Sentinel 1 y 2. Para esto, a pesar de tener datos desde 2013, se trabajó con los datos de las zafas 2017-18 y 2018-19, ya que como se explicó en la sección 6.1, es a partir de 2017 cuando se cuenta con una serie continua de imágenes de los satélites Sentinel.

De las chacras disponibles solo se utilizaron aquellas que su cultivo fuera maíz, soja, sorgo o pastura, dejando fuera las chacras con la etiqueta artificial sorgo-maíz. Al igual que se hizo para la base FUCREA-URF se armó un array multibanda con las mismas características, en cuanto a bandas utilizadas, y períodos de fechas. En este array están todos los puntos de las chacras de PLANES según se generaron como se explicó en la sección 7.1. En la tabla 45, según zafa y cultivo se ve la cantidad de planes, hectáreas y puntos disponibles.

CULTIVO	2017-18			2018-19		
	PLANES	HECTÁREAS	PUNTOS	PLANES	HECTÁREAS	PUNTOS
MAÍZ	121	11.125	11.164	177	14.135	14.165
PASTURA	2.649	221.248	11.100	2.492	201.424	14.268
SOJA	5.528	681.794	12.530	5.454	637.844	14.392
SORGO	192	10.352	10.395	232	13.071	13.073
TOTAL	8.490	924.519	45.189	8.355	866.474	55.898

Tabla 45. Cantidad de planes, hectáreas y puntos con series temporales de Sentinel según cultivo de las zafas 2017-18 y 2018-19 para PLANES.

Al igual que se hizo en la sección anterior, previo a entrenar el modelo de clasificación, a los datos disponibles se le aplicó una doble depuración, la primera para eliminar puntos de chacras sin patrón agrícola y sin patrón de pastura, y la segunda para eliminar firmas outliers usando un factor de 2 para multiplicar IQR. En la tabla 46 se presentan la cantidad de datos después de aplicar la doble depuración.

CULTIVO	2017-18			2018-19			TOTAL		
	PLANES	HECTÁREAS	PUNTOS	PLANES	HECTÁREAS	PUNTOS	PLANES		PUNTOS
							CANTIDAD	%	
MAÍZ	114	10.870	8.270	175	14.088	12.821	289	2,2	21.091
PASTURA	2.352	204.962	9.151	2.204	188.958	11.359	4.556	35,3	20.510
SOJA	3.794	511.668	7.178	3.865	490.048	8.431	7.659	59,4	15.609
SORGO	172	9.667	8.185	222	12.604	9.814	394	3,1	17.999
TOTAL	6.432	737.167	32.784	6.466	705.698	42.425	12.898	100	75.209

Tabla 46. Cantidad de planes, hectáreas y puntos según cultivo después de la doble depuración para PLANES.

Una vez obtenida la base depurada, se dividió el total de planes para entrenamiento y evaluación en porcentajes de 70 y 30 respectivamente. El entrenamiento se hizo con todos los puntos disponibles de cada plan del conjunto de entrenamiento, y en evaluación se clasificaron todos los puntos de los planes, y la clase final de cada plan se obtuvo por voto mayoritario, al igual que en la sección anterior. Este proceso se repitió unas 50 corridas. En la tabla 47 se presenta la matriz de confusión de la mejor corrida, y en la tabla 48, por cultivo se presentan los promedios y desvío estándar de las métricas de todas las corridas. Es importante recordar que los datos de PLANES no son verdad de terreno como en FUCREA-URF, sino que es lo que el productor agrícola declaró que iba a sembrar como ya se explicó en la sección 5.2.1.

		CLASIFICADO				TOTAL	RECALL	F1-SCORE
		MAÍZ	PASTURA	SOJA	SORGO			
OBSERVADO	MAÍZ	53	3	27	3	86	61,6	40,2
	PASTURA	28	1.288	37	14	1.367	94,2	95,4
	SOJA	90	41	2.047	119	2.297	89,1	92,6
	SORGO	7	1	15	96	119	80,7	54,7
	TOTAL	178	1.333	2.126	232	3.869		
PRECISION		29,8	96,6	96,3	41,4			

Tabla 47. Matriz de confusión en evaluación de la mejor corrida en PLANES.

CULTIVO	PRECISION	RECALL	F1-SCORE
MAÍZ	23,8 (1,6)	68,2 (4,9)	35,3 (2,1)
PASTURA	97,0 (0,5)	94,2 (0,6)	95,6 (0,6)
SOJA	96,9 (0,3)	86,3 (0,8)	91,3 (0,4)
SORGO	37,9 (1,5)	79,3 (2,8)	51,3 (1,6)

Tabla 48. Valores promedio y desvío estándar de las métricas de desempeño para las 50 corridas en evaluación de PLANES.

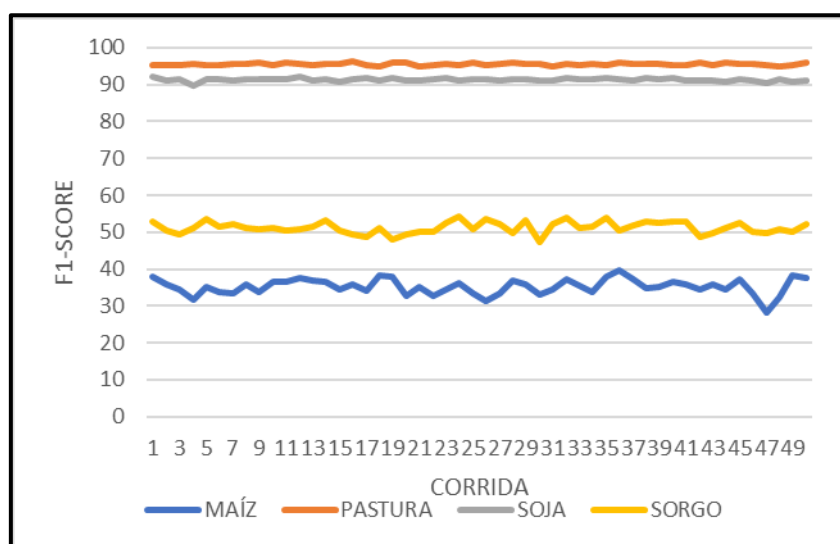


Figura 67. Gráfico de *F1-Score* por cultivo en evaluación para las 50 corridas de PLANES.

Como se observa en la figura 67, *F1* para pastura y soja está por encima de 90, con baja variación entre corridas, de menos de 1 punto como se observa en la tabla 48. Pastura es la clase de mejor desempeño, con casi 96 de *F1*. Para maíz y sorgo la situación es bastante diferente, siendo los valores de *F1* muy inferiores, en torno a 50 para sorgo, y entre 30 y 40 para maíz, con una variación en torno a 2 puntos entre corridas. Teniendo en cuenta el alto *F1* de pastura, y que

esta es la clase no agrícola, y los otros cultivos conforman la clase agrícola, claramente el algoritmo tiene un muy buen desempeño en discriminar entre cultivos agrícolas y no agrícolas.

Al analizar *precision*, se observa que para pastura y soja es de 97 puntos, mientras que para maíz y sorgo, es de 23,8 y 37,9 respectivamente. Estos bajos valores para estos cultivos, inferiores a 50, indican una sobreestimación de chacras clasificadas como maíz o sorgo. Más de la mitad de las chacras clasificadas como estos cultivos, no lo son, siendo chacras de pastura o soja, mal clasificadas.

Cuando se analiza *recall*, se ve que pastura y soja tienen los valores más altos, con 94,2 y 86,3 puntos, siendo inferiores a los valores de *precision*, pero mucho menor para soja. Para maíz y sorgo, *recall* es de 68,2 y 79,3 respectivamente. Aquí se nota que para estos dos cultivos, el desempeño en *recall* es muy superior a *precision*.

En la tabla 49 y figura 68 se presentan las métricas de desempeño en evaluación de PLANES y FUCREA-URF para los cultivos que tienen en común. Al analizar *F1*, en ambos casos se observa que el desempeño en soja es muy alto, con valores superiores a 91, pero mejor en FUCREA-URF. Para el caso de maíz y sorgo, el desempeño es bastante mejor en FUCREA-URF, siendo más del doble para maíz, y casi 30 puntos de diferencia para sorgo.

En el caso de *precision*, se observa que para soja los valores son similares, en torno a 97. Ahora, para maíz y sorgo, en FUCREA-URF esta es más del doble que en PLANES. Para el caso de *recall*, los valores entre PLANES y FUCREA-URF son muy similares, con excepción de soja, en donde en FUCREA-URF esta es 13 puntos superior a PLANES.

Hay que tener presente que se están comparando los resultados de dos bases distintas, una en que los datos son verdad de terreno como es el caso de FUCREA-URF, y otra como PLANES en que los datos no lo son. Aun así, de la tabla 49 se concluye que el modelo entrenado en la base de FUCREA-URF tiene un desempeño bastante superior al modelo de PLANES para discriminar cultivos agrícolas.

CULTIVO	PRECISION		RECALL		F1-SCORE	
	FUCREA-URF	PLANES	FUCREA-URF	PLANES	FUCREA-URF	PLANES
MAÍZ	88,5	23,8	65,9	68,2	75,1	35,3
SOJA	97,9	96,9	99,1	86,3	98,5	91,3
SORGO	81,0	37,9	77,0	79,3	78,5	51,3

Tabla 49. Comparativo de métricas de desempeño en evaluación de FUCREA-URF y PLANES.

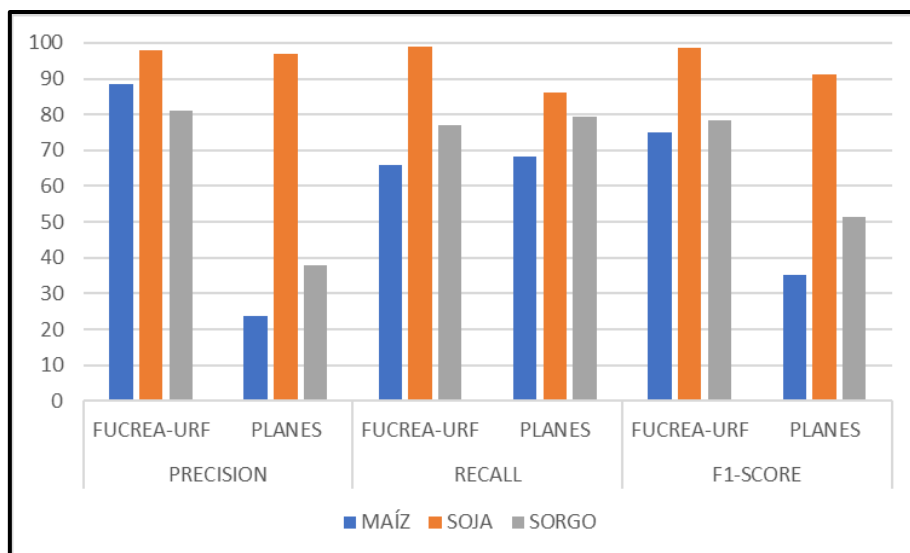


Figura 68. Gráfico de métricas de desempeño en FUCREA-URF y PLANES.

10.4 Clasificación de PLANES con modelo entrenado en FUCREA-URF

En esta sección, con el array multibanda de la base depurada FUCREA-URF se entrenó XGBoost para clasificar la base depurada PLANES. De esta, solo se utilizaron aquellos planes que su etiqueta de cultivo es maíz, soja, o sorgo, dejando fuera pastura, ya que este cultivo no está presente en FUCREA-URF. En la figura 69 se presenta el flujograma del proceso de entrenamiento y clasificación. Se entrenaron dos modelos de clasificación, uno con todos los puntos disponibles de las chacras de FUCREA-URF, al cual se llamó modelo no balanceado, y otro muestreando un 10% de los puntos de soja para generar el modelo llamado balanceado. La clasificación de los planes se hizo en dos etapas, primero clasificando cada punto dentro de estos, y luego por voto mayoritario de todos los puntos del plan se hace la clasificación final, de manera similar como se hizo en las secciones anteriores.

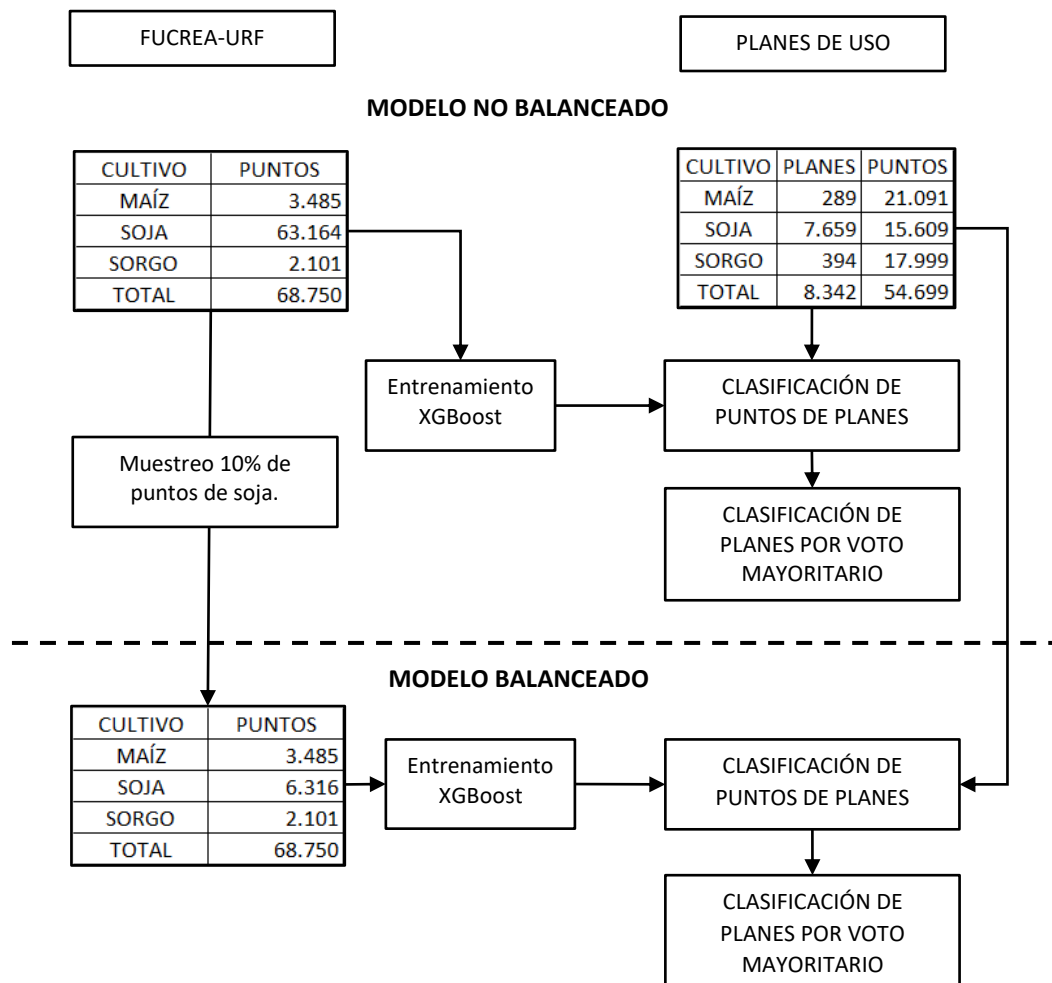


Figura 69. Flujograma de entrenamiento de XGBoost con base de FUCREA-URF y evaluación en PLANES.

En la tabla 50 se presenta la matriz de confusión sobre PLANES con el modelo balanceado, y en la 51 y figura 70 las métricas de desempeño obtenidas con el modelo balanceado y el que no. Al analizar *precision* tanto en el modelo balanceado como en el que no, se observa que soja tiene el mejor desempeño por amplio margen con valores mayores a 94. En el caso de maíz y sorgo el valor máximo es 66,7. Teniendo en cuenta el modelo con y sin balanceo, para sorgo la *precision* no se ve afectada, con valores de casi 51, mientras que para maíz la *precision* se ve muy afectada por el modelo balanceado, bajando casi la mitad, de 66,7 a 34,9.

Con respecto a *recall*, de nuevo soja es el cultivo de mejor desempeño con valores superiores a 94. Para maíz y sorgo esta es bastante inferior con un máximo de 55. Al considerar el efecto del balanceo, para estos cultivos se nota que éste mejora el *recall*, pero con mayor efecto en sorgo, en donde se multiplica por más de dos veces y media, pasando de 16,8 a 45,9.

Al analizar *F1*, se nota que el modelo balanceado es de mejor desempeño que el modelo sin, ya que para soja prácticamente no hay cambios en *F1*, con valores en torno a 96, para maíz hay una disminución de 10 puntos, pero en sorgo es donde se nota más el cambio, incrementándose *F1* casi el doble, pasando de 25,2 a 48.

		CLASIFICADO			TOTAL	RECALL	F1-SCORE
		MAÍZ	SOJA	SORGO			
OBSERVADO	MAÍZ	159	119	11	289	55,0	42,7
	SOJA	237	7.253	167	7.657	94,7	95,5
	SORGO	59	154	181	394	45,9	48,0
	TOTAL	455	7.526	359	8.340		
PRECISION		34,9	96,4	50,4			

Tabla 50. Matriz de confusión en PLANES clasificado con modelo entrenado con FUCREA-URF balanceado.

CULTIVO	PRECISION		RECALL		F1-SCORE	
	CON BALANCEO	SIN BALANCEO	CON BALANCEO	SIN BALANCEO	CON BALANCEO	SIN BALANCEO
MAÍZ	34,9	66,7	55	42,9	42,7	52,2
SOJA	96,4	94,2	94,7	98,7	95,5	96,4
SORGO	50,4	50,8	45,9	16,8	48	25,2

Tabla 51. Métricas de desempeño en PLANES con modelo entrenado en FUCREA-URF con y sin balanceo.

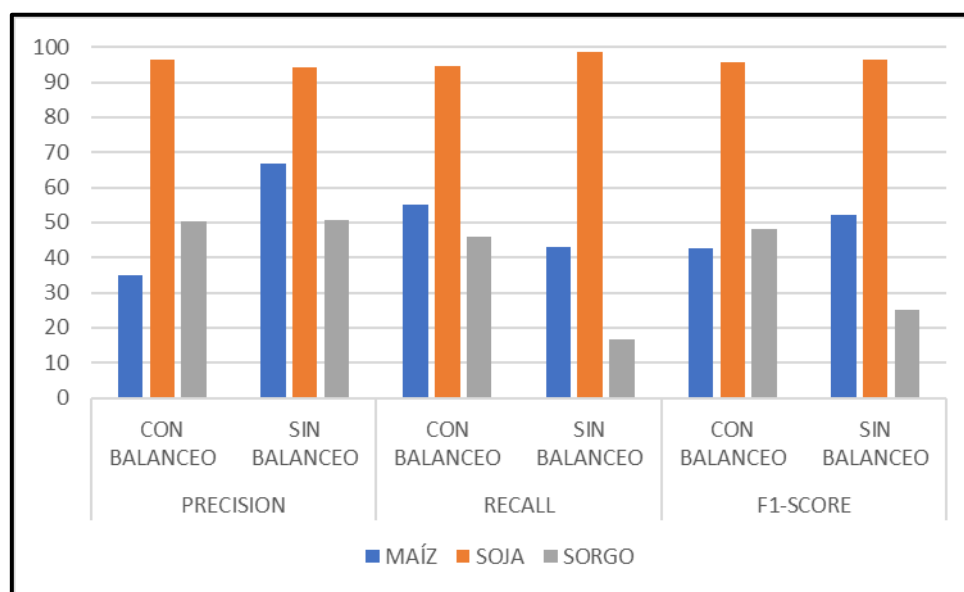


Figura 70. Gráfico de métricas de desempeño para PLANES con modelo entrenado en FUCREA-URF con y sin balanceo.

Para finalizar esta sección, en la tabla 52 y figura 71 se presentan los valores de *F1* por cultivo para los artículos revisados de la tabla 45, FUCREA-URF, PLANES, y para los modelos entrenados con FUCREA-URF y aplicado en PLANES. Empezando por soja, se observa que en comparación a los artículos revisados, tanto en FUCREA-URF, como en PLANES, *F1* es superior, con valores mayores a 91, y con el mejor desempeño en FUCREA-URF con 98,5. Para el caso de maíz, cualquiera sea la base utilizada, el desempeño es menor al valor de referencia de 84,4 de los artículos, siendo como máximo 75,3 en FUCREA-URF, y con el peor desempeño en PLANES, con 35,3. Para el caso de sorgo, el mejor desempeño se obtuvo en FUCREA-URF con un valor de 77,5, superior al 67,3 de los artículos, y el peor desempeño fue de 25,2 en el modelo sin balanceo aplicado sobre PLANES.

Al comparar el desempeño de la base FUCREA-URF contra PLANES, se observa que en la primera este es muy superior, sobre todo en maíz y sorgo. Esto como se mencionó, puede estar explicado por el hecho de que los datos de FUCREA-URF son verdad de terreno, mientras que en PLANES son los cultivos declarados a sembrar, y podría haber pasado que muchos de los cultivos declarados como maíz y sorgo, hayan sido efectivamente soja, que es el cultivo dominante.

Cuando se observan los resultados en PLANES con el modelo entrenado en FUCREA-URF, con y sin balanceo, se nota que por efecto del balanceo, el desempeño en general mejora. Si se analizan los resultados del modelo balanceado contra el desempeño del modelo entrenado y evaluado con PLANES, se nota que los valores obtenidos son similares, pero algo mejor para maíz y soja en el modelo balanceado, y algo por debajo para sorgo. Aquí de nuevo, en el modelo balanceado, sucede al igual que en PLANES, que muchos de los declarados como maíz o sorgo, son clasificados como soja.

CULTIVO	ARTICULOS TABLA 45	FUCREA-URF	PLANES	PLANES: MODELO FUCREA-URF	
				CON BALANCEO	SIN BALANCEO
MAÍZ	84,4	75,3	35,3	42,7	52,2
SOJA	88,4	98,5	91,3	95,5	96,4
SORGO	67,3	77,5	51,3	48,0	25,2

Tabla 52. *F1-Score* por cultivo de las bases de FUCREA-URF, PLANES y artículos revisados.

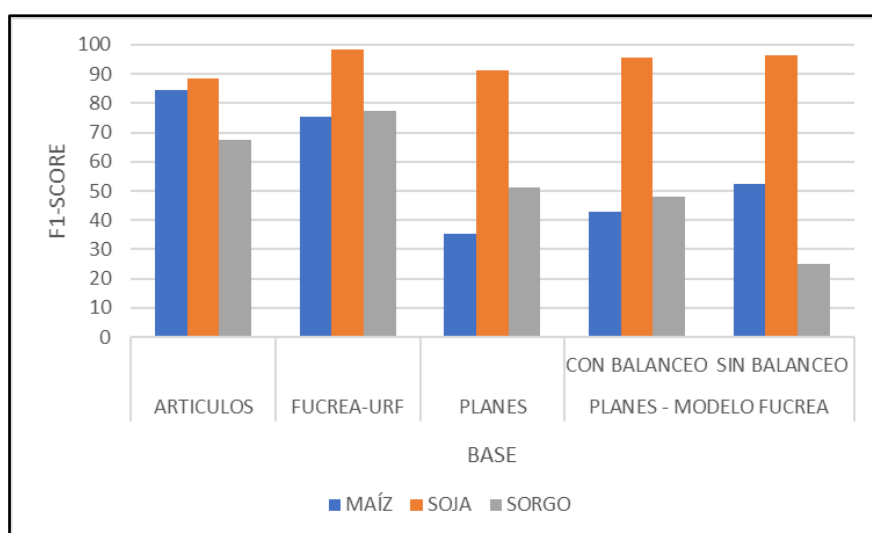


Figura 71. Gráfico de *F1-Score* por cultivo de las bases de FUCREA-URF, PLANES y artículos revisados.

10.5 Clasificación binaria agrícola/no-agrícola con COBERTURAS y FUCREA-URF

Utilizando conjuntamente las bases COBERTURAS y FUCREA-URF se entrenó XGBoost para generar un modelo que clasifica puntos georreferenciados en dos grandes categorías, agrícola y no-agrícola. Por agrícola, se entiende maíz, soja y sorgo, y por no-agrícola, las coberturas y/o usos del suelo como agua, arena, bañados, bosque, minero, recursos forrajeros, urbano y otros de la tabla 15 de la sección 7.3. En la tabla 53 se muestran las coberturas y/o usos que caen en una u otra categoría, y la cantidad de puntos para entrenar y evaluar de las zafras 2017-18 y 2018-19. En el caso de FUCREA-URF como ya se explicó, se hizo una doble depuración de los datos. Primero seleccionando grupos de puntos con firma agrícola de NDVI, y posteriormente usando Tukey con factor de 2 para multiplicar IQR se eliminaron puntos con firmas outliers. Para

COBERTURAS, después de hacer la clasificación de las series temporales de NDVI en grupos, no se encontró alguno con firma fenológica agrícola, por lo que solo se hizo la depuración de outliers con Tukey y factor de 2.

COBERTURA/USO DEL SUELO	PUNTOS		CATEGORÍA	BASE
	2017-18	2018-19		
MAÍZ	1.442	2.043	AGRÍCOLA	FUCREA-URF
SOJA	31.077	31.777		
SORGO	640	1.457		
SUBTOTAL	33.159	35.277		
AGUA	677	256	NO AGRÍCOLA	COBERTURAS
ARENA	183	199		
BAÑADOS	857	766		
BOSQUE	1.908	1.853		
MINERO	39	37		
OTROS	744	716		
RECURSOS FORRAJEROS	4.818	4.641		
URBANO	856	837		
SUBTOTAL	10.082	9.305		
TOTAL	43.241	44.582		

Tabla 53. Cantidad de puntos según cobertura/uso del suelo y zafra para entrenamiento y evaluación de modelo de clasificación binaria agrícola/no-agrícola.

Como se observa en la tabla 53, la cantidad de puntos de FUCREA-URF es más del triple que en COBERTURAS, por lo cual, previo a entrenamiento y evaluación, de la primera, se muestrearon de manera estratificada una cantidad de puntos igual a la de COBERTURAS. Después de tener los datos depurados y muestreados, XGBoost se entrenó con los datos de la zafra 2017-18 y se evaluó con los de la 2018-19. En la figura 72 se ve el flujograma de todo el proceso.

En la tabla 54 se presenta la matriz de confusión y las métricas de desempeño en evaluación. Como se observa, *F1-Score* está en torno a 93 para agrícola y no-agrícola, siendo algo superior el desempeño en la última. Al analizar *precision* y *recall*, para el primero este es de 98,3 para agrícola, unos 9,2 puntos superior a no-agrícola, mientras que para *recall* la situación es inversa, siendo 98,5 para no-agrícola y 87,9 para agrícola. De esto, se concluye que el modelo de todos los puntos que son agrícola o no-agrícola, tiene mejor capacidad para identificar los no-agrícola, mientras que de los que clasifica como uno u otro, tiene mejor capacidad para clasificar los agrícolas. De todas maneras, teniendo en cuenta la exactitud general medida con *F1*, ambas clases tienen un desempeño similar, por encima de 90, y además si se considera que el modelo fue entrenado en una zafra y evaluado con otra, se puede decir que el desempeño es muy bueno.

ENTRENAMIENTO EN ZAFRA 2017-18

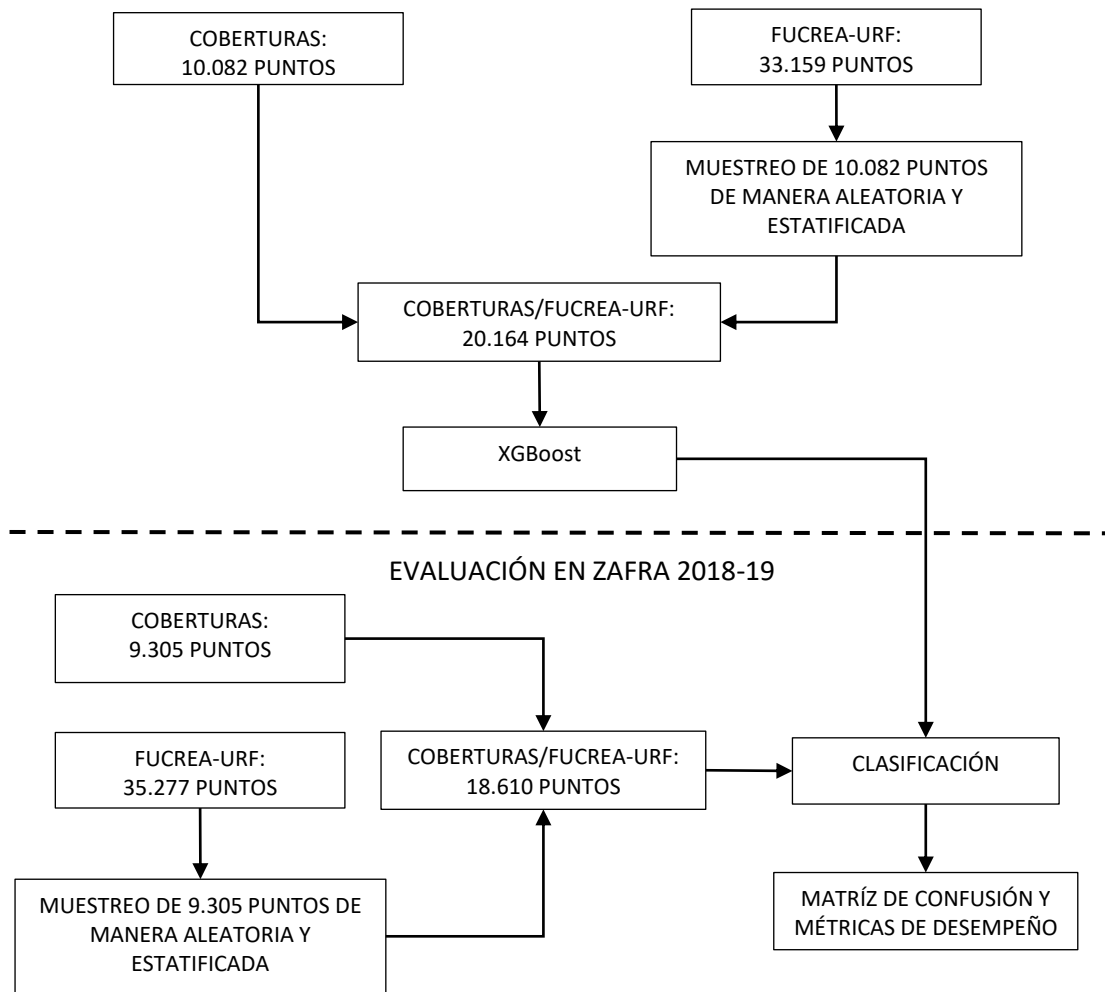


Figura 72. Flujograma de generación de modelo de clasificación binaria agrícola/no-agrícola.

		CLASIFICADO		TOTAL	RECALL	F1-SCORE
		AGRICOLA	NO AGRICOLA			
OBSERVADO	AGRICOLA	8.179	1.126	9.305	87,9	92,8
	NO AGRICOLA	138	9.167	9.305	98,5	93,6
TOTAL		8.317	10.293	18.610		
PRECISION		98,3	89,1			

Tabla 54. Matriz de confusión y métricas de desempeño de clasificación binaria agrícola/no-agrícola.

11. Clasificación pixel a pixel en Google Earth Engine

En las secciones anteriores se entrenaron modelos de clasificación a nivel de puntos georreferenciados, pero no se generaron mapas de salida, solamente se construyeron las matrices de confusión y se calcularon las métricas de desempeño. En esta sección utilizando GEE se generaron mapas clasificados pixel a pixel para todo el país. Primero se generó un mapa binario de cultivos agrícolas y no-agrícolas, y posteriormente en la zona de influencia de FUCREA-URF se identificó el cultivo específico (maíz, soja o sorgo) dentro de las chacras clasificadas como agrícolas.

11.1 Mapa binario agrícola/no-agrícola en GEE

En esta sección se detalla el proceso en GEE de la clasificación supervisada binaria para todo el país a nivel de pixel en cultivos agrícola/no-agrícola usando como datos de entrenamiento la base COBERTURAS y FUCREA-URF. Se decidió trabajar con los datos de la zafra 2017-18. Lo primero, fue crear una base conjunta de COBERTURAS y FUCREA-URF para usar en la etapa de entrenamiento y evaluación de los algoritmos elegidos. La plataforma GEE tiene un gran poder de cómputo, aun así, si el volumen de datos a procesar es muy grande pueden ocurrir errores por exceder el límite de memoria, agotar el tiempo de cálculo y otros, que determine que no se ejecuten los scripts. Para evitar eso se probaron éstos con diferentes cantidades de puntos de entrenamiento, y se observó que con 5.500 puntos éstos se ejecutaban sin problemas, y a su vez es un número adecuado para entrenar.

En las secciones anteriores las bases fueron depuradas usando Tukey con un factor multiplicador de 2 para IQR. En esta se decidió ser más exigente y usar un factor de 1, de manera que se tienen 7.728 y 23.679 puntos de COBERTURAS y FUCREA-URF respectivamente. Usando estos puntos se creó una base balanceada con un total de 11.000 puntos, 5.500 de FUCREA-URF (agrícolas) y 5.500 de COBERTURAS (no-agrícolas). En ambas bases los puntos se extrajeron aleatoriamente de manera estratificada para que todas las clases estuvieran representadas, es decir, de los 5.500 puntos de COBERTURAS, hay de las clases bañados, bosques, cuerpos de agua, recursos forrajeros, etc., y de los 5.500 de FUCREA-URF hay de maíz, soja y sorgo. Una vez generada la base conjunta, esta se dividió un 50% para entrenamiento en GEE y otro 50% para evaluación de la calidad de la clasificación. En la figura 73 se presenta el flujograma de todo el proceso. En cuanto a la distribución territorial de los puntos de entrenamiento y evaluación, en la figura 74 se ve que los no-agrícolas están en todo el territorio, mientras los agrícolas están concentrados en la zona de influencia de FUCREA-URF, que es la zona núcleo agrícola.

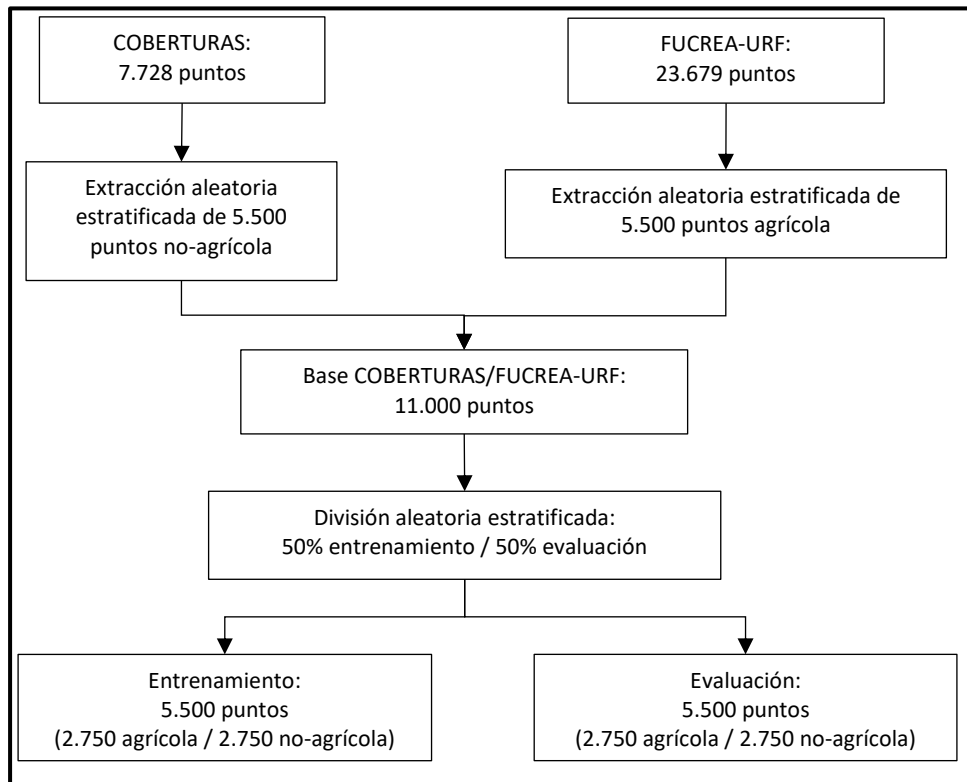


Figura 73. Flujograma de creación de la base conjunta COBERTURAS/FUCREA-URF de puntos agrícola/no-agrícola de la zafra 2017-18.

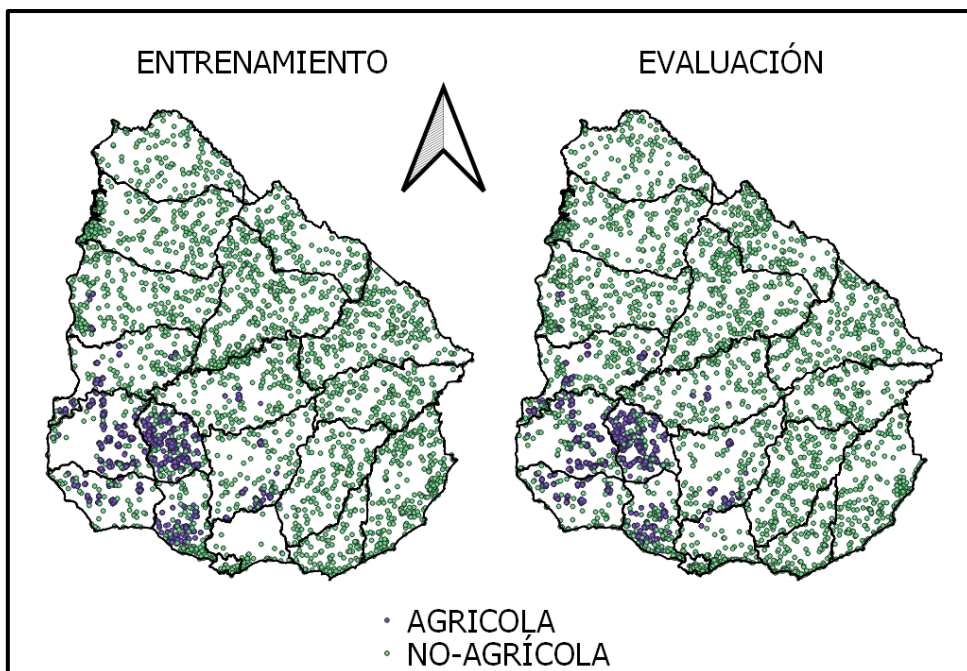


Figura 74. Mapa de distribución de puntos de entrenamiento y evaluación de la base conjunta COBERTURAS/FUCREA-URF de puntos agrícola/no-agrícola de la zafra 2017-18.

Una vez obtenidos los puntos de entrenamiento y evaluación, se corrieron en GEE los scripts de clasificación supervisada. GEE tiene disponible para usar muchos algoritmos supervisados, por ejemplo, árboles de decisión, clasificador de máxima entropía, Gradient Tree Boost, Naive Bayes, Random Forest, SVM y otros. En este caso se decidió utilizar Random Forest (RF) y

Gradient Tree Boost (GTB), este último es similar a XGBoost. Como se vio en la sección 10.2.2, estos tuvieron el mejor desempeño para identificar cultivos agrícolas.

Como fue mencionado, GEE tiene un gran poder de cómputo, no obstante, si el volumen de datos a procesar es muy grande puede excederse la memoria disponible y los tiempos de ejecución. Es importante destacar que se procesaron 440 millones de píxeles de 20 metros de lado para una superficie total de 17.622.000 hectáreas. Por lo cual, después de realizar algunas pruebas en cuanto a las bandas e índices de Sentinel (1 y 2) a utilizar, se decidió trabajar por separado con los apilados de EVI2, NDVI y el de las bandas polarimétricas VH_VV, y se descartó el uso conjunto de los índices y bandas de Sentinel (1 y 2) como se hizo en la sección 10.2.2, ya que esto excedió el poder de cómputo en GEE. Los apilados se generaron con la metodología que se explicó en la sección 8.2, pero esta tiene el inconveniente de que en el proceso de generación de las imágenes sintéticas quedan píxeles sin dato, y algoritmos como RF no funcionan con datos faltantes. Para imputar los píxeles sin dato, en cada apilado para cada píxel se aplicó en la dimensión temporal un modelo de interpolación armónica adaptado a partir de: <https://docs.google.com/document/d/1mNIRB90jwLuASO1JYas1kuOXCLbOoy1Z4NIV1qIXM10/edit>. Al aplicar este modelo se completan datos faltantes y a su vez se regenera la serie temporal original pero suavizada, como se observa en la figura 75 para el caso de NDVI. El modelo armónico tiene la siguiente fórmula:

$$p_t = \beta_0 + \beta_1 t + A * \cos(2\pi * \omega * t - \varphi) + e_t$$

En donde p_t es valor interpolado p en el tiempo t , A es amplitud, ω es frecuencia, φ es fase y e_t es un error aleatorio. Como se observa en el ejemplo de la figura 75, la serie original (NDVI) tiene dos datos faltantes, que si están presentes en la serie reconstruida (NDVI_suavizado).

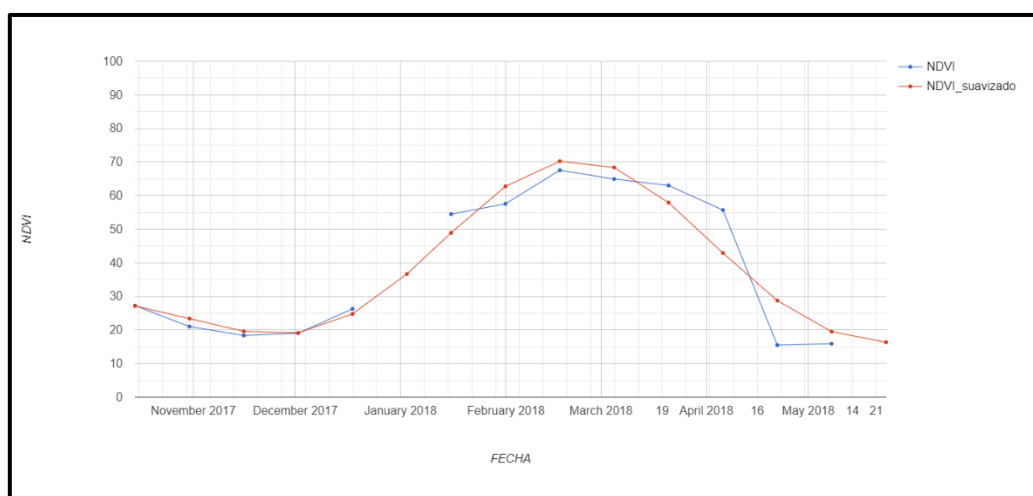


Figura 75. Imputación de datos faltantes en serie temporal de NDVI con modelo armónico.

Una vez imputados los datos faltantes, cada apilado se intersectó con los 5.500 puntos de entrenamiento y para cada uno se extrajeron las respectivas series temporales. Cada apilado está compuesto de 15 imágenes sintéticas, por lo que las series temporales se componen de 15 fechas como se ve en la figura 75. De esta manera, se cuentan con 3 juegos de series temporales (EVI2, NDVI y VH_VV). Posteriormente cada uno de esos juegos se utilizó para entrenar GTB y RF, de manera que se generaron 6 modelos de clasificación. Luego, cada apilado se clasificó pixel a pixel con sus respectivos modelos GTB y RF, y si uno de los modelos clasificaba como agrícola un pixel, este se clasificaba como agrícola, de lo contrario se clasificaba como no-agrícola. Es así como se obtuvieron un total de 3 mapas binarios de cultivos agrícola/no-agrícola. Los algoritmos

se ejecutaron con 300 árboles de decisión. En la figura 76 se muestra el flujograma de entrenamiento de los modelos de GTB y RF, y de clasificación de los apilados.

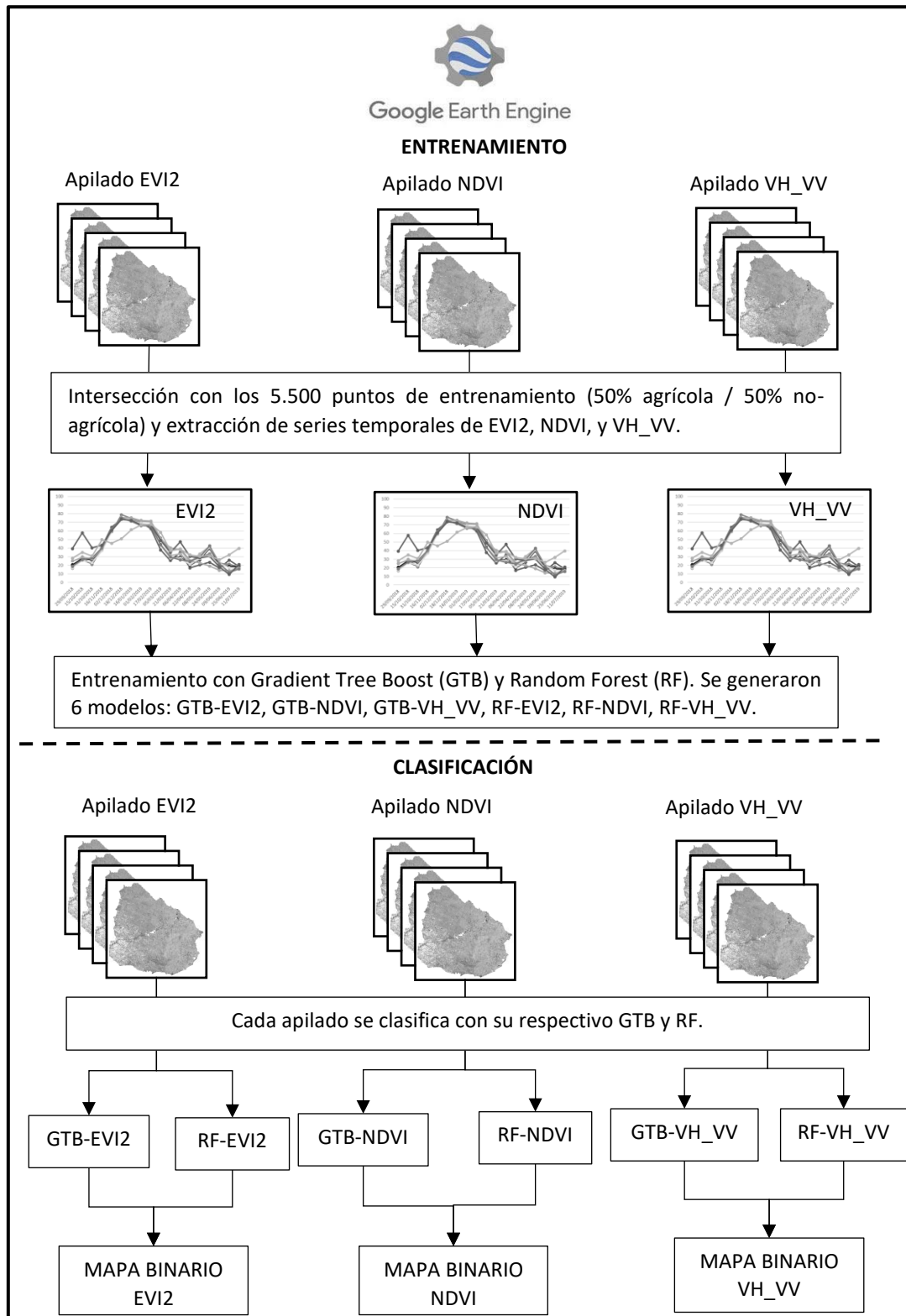


Figura 76. Flujograma de clasificación supervisada binaria agrícola/no-agrícola pixel a pixel en GEE.

En el caso de los mapas binarios EVI2 y NDVI, estos son muy parecidos, ambos obtenidos a partir de Sentinel 2 (figura 77), mientras que en el caso del mapa binario VH_VV (Sentinel 1) este

presenta bandeados que coinciden con el límite entre una pasada y otra del satélite, como se observa marcado con elipses rojas en la figura 78.

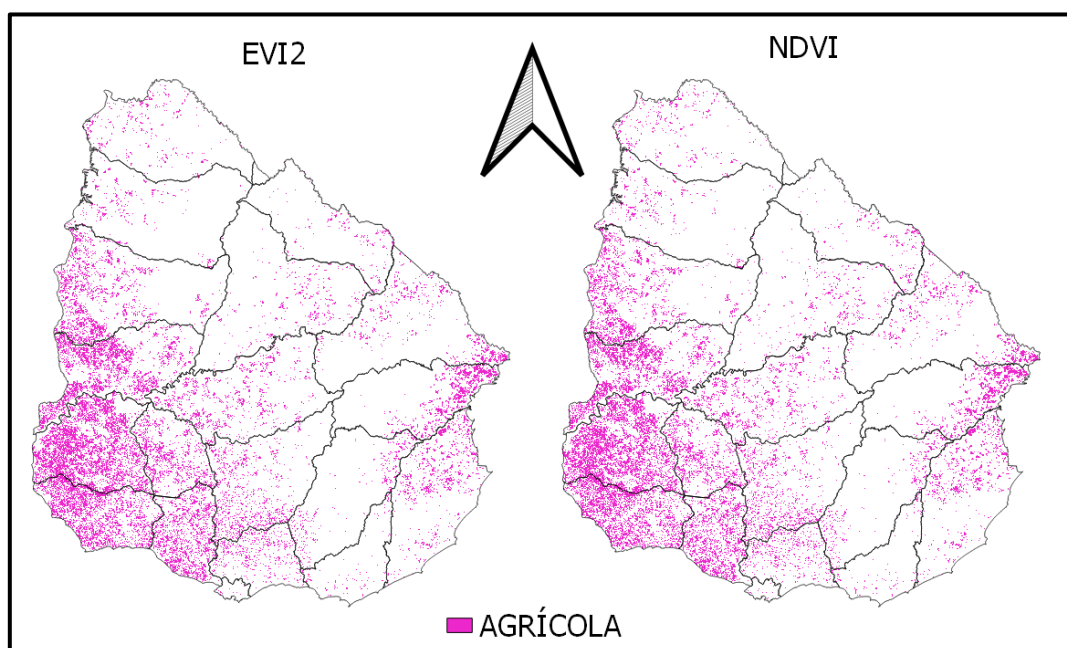


Figura 77. Mapas binarios de EVI2 y NDVI.

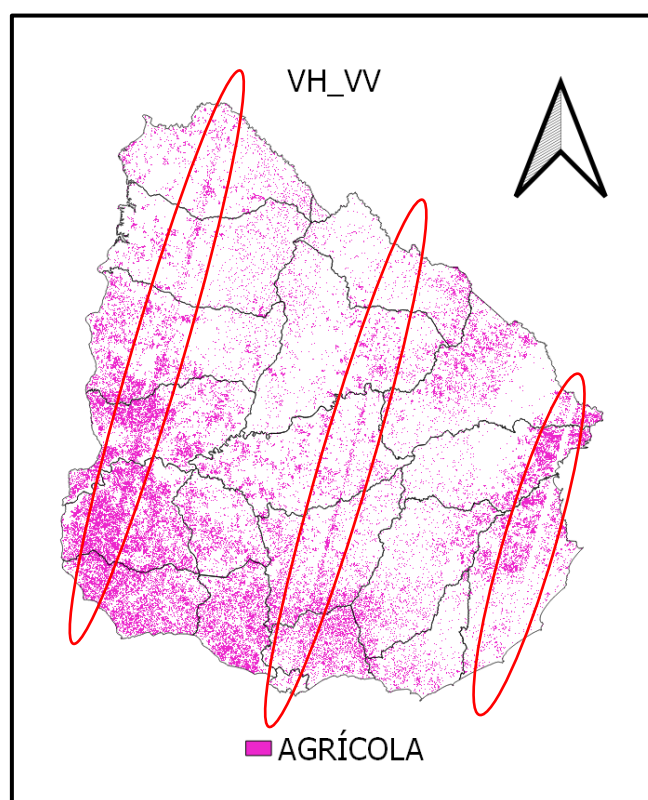


Figura 78. Mapa binario VH_VV con bandeados (elipses rojas).

En la tabla 55 se presenta el área de cada mapa binario y el área estimada por DIEA para la zafra 2017-18 (MGAP-DIEA, 2019). Es importante aclarar que el dato de DIEA no proviene de un censo, sino que se estima a partir de encuestas y no se conoce el margen de error.

FUENTE	ÁREA AGRÍCOLA (HA)
DIEA 2017-18	1.367.600
BINARIO EVI2	1.804.374
BINARIO NDVI	1.810.886
BINARIO VH_VV	2.946.207

Tabla 55. Área agrícola (hectáreas) estimada por mapas binarios (EVI2, NDVI, VH_VV) y DIEA para la zafra 2017-18.

Como se ve en la tabla 55, en cualquiera de los tres mapas binarios, el área estimada supera a la de DIEA, en el caso de EVI2 y NDVI, un 32%, y para VH_VV un 115%. El mapa binario VH_VV, tiene una sobrestimación muy grande del área y el problema del bandeo ya mencionado.

Los mapas binarios obtenidos se analizaron visualmente para evaluar la calidad de la clasificación. Este trabajo se hizo especialmente para NDVI y EVI2, ya que su área estimada es la más parecida a DIEA, y se descartó VH_VV por lo mencionado anteriormente. Al observar ambas clasificaciones, se notó que en algunos casos EVI2 clasificaba algunos pixeles como agrícola y NDVI no, y viceversa, cuando los pixeles aparentemente por la información de contexto si lo eran. Como resultado de lo anterior se decidió crear un mapa binario a partir de la fusión de los binarios de EVI2 y NDVI, en donde si un pixel había sido clasificado como agrícola en uno u otro, este se clasificaba agrícola. En la figura 79 se ve un ejemplo de lo mencionado para un par de chacras y el resultado de la fusión.

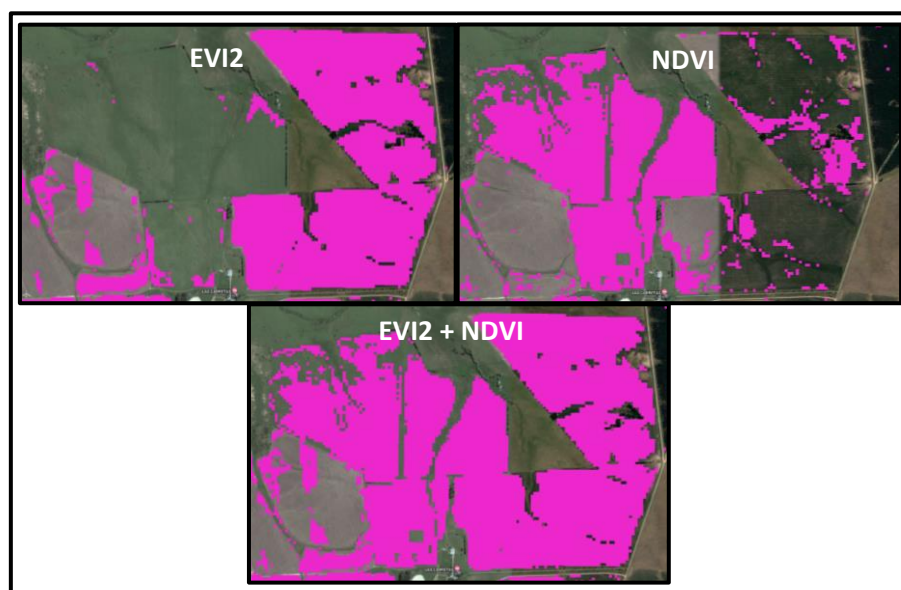


Figura 79. Ejemplo de pixeles clasificados como agrícolas por binario EVI2 o NDVI, y fusión de ambos.

El nuevo mapa binario EVI2+NDVI (binario 2) tiene un área total de 1.912.724 hectáreas, algo superior a la superficie de los mapas de EVI2 y NDVI, lo que indica que de la fusión de ambos se incrementa el área en algo más de 100.000 hectáreas. Al considerar la distribución espacial de los pixeles identificados como agrícolas en binario 2, estos se distribuyen principalmente en las regiones tradicionalmente agrícolas (agrícola de secano y arroceras), como se observa en la figura 80, donde se ve binario 2 superpuesto sobre el mapa de Regiones Agropecuarias según censo agropecuario del 2011. No obstante, puede haber pixeles agrícolas en regiones donde la actividad predominante es lechera, forestal u otra. Las regiones están definidas por la actividad

predominante, la que no es exclusiva, es decir en una región forestal puede haber agricultura, y en una agrícola puede haber forestación, y así en todas las regiones.

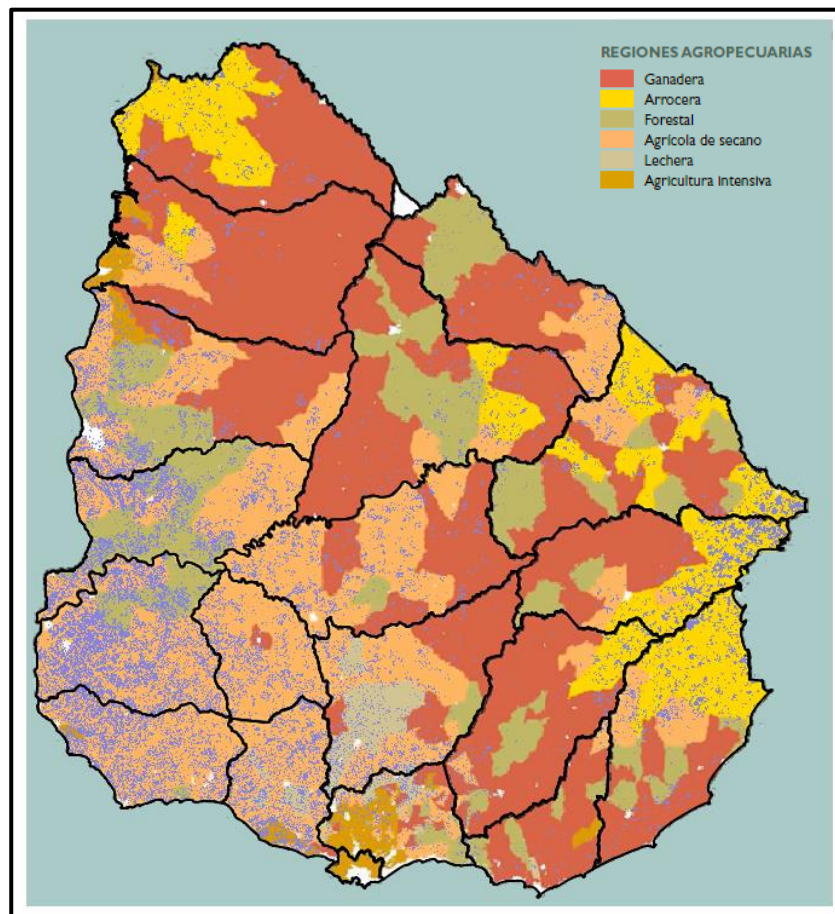


Figura 80. Mapa de binario 2 superpuesto sobre regiones agropecuarias según censo del año 2011.

En binario 2 se nota al observar el mapa que hay píxeles espurios, es decir, clasificados como agrícola cuando no lo son, como se ve claramente en el ejemplo de la figura 81, donde se indica como agrícola, píxeles en la margen de un río.



Figura 81. Ejemplo de píxeles espurios clasificados como agrícolas.

A efectos de refinar la clasificación, se evaluaron diferentes estrategias para eliminar píxeles espurios en binario 2 combinando el uso de un filtro de moda y clasificación no supervisada. En la figura 85 se muestra el flujograma de las estrategias aplicadas. Un filtro de moda es una

técnica de filtrado post-clasificación usada para reducir la fragmentación espacial conocida como efecto de “sal y pimienta”. Este, es un tipo de filtro aritmético que consiste en una ventana móvil, por lo general cuadrada y de número impares de píxeles por lado, 3x3, 7x7, etc. Esta ventana va recorriendo fila por fila y columna por columna la imagen a filtrar (ver figura 82), y sustituye el valor del píxel central de la ventana por el valor más frecuente (moda) de sus vecinos circundantes (Chuvieco, 2016). En la figura 83 se ve un filtro de moda de 3x3 y 7x7 aplicado sobre una imagen, y se observa que como resultado se generan áreas más homogéneas y con menos dispersión de píxeles de una clase y otra en comparación con la imagen original.

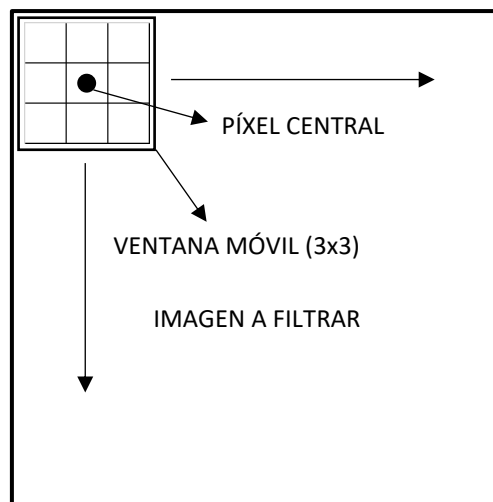


Figura 82. Filtro de moda aplicado sobre una imagen.

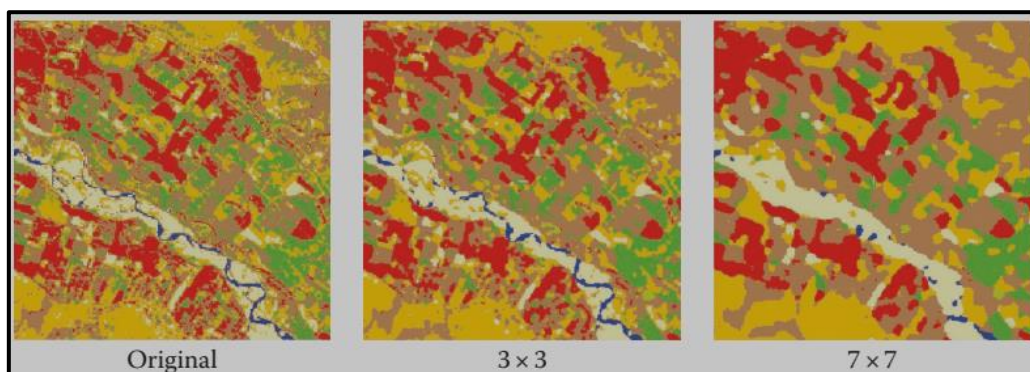


Figura 83. Ejemplo de filtro de moda con ventana de 3x3 y 7x7. Fuente: Chuvieco (2016).

Para el filtrado de la imagen binario 2, después de algunas pruebas, se decidió usar un filtro de moda de 7x7, obteniéndose el mapa binario 3. En la figura 84 se ve el efecto del filtrado sobre la misma zona de la figura 81.

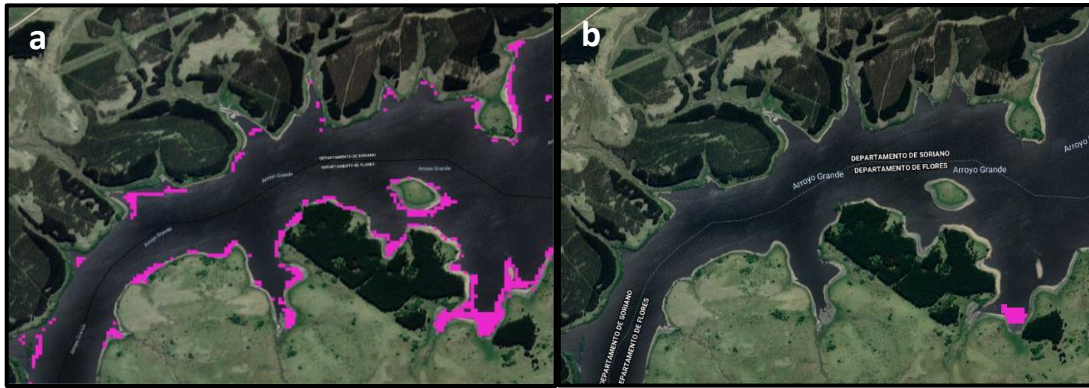


Figura 84. Efecto del filtro de moda 7x7, antes y después (a y b).

En el caso de las técnicas de clasificación no supervisadas, GEE cuenta con varios algoritmos incorporados para usar, entre los que están Cobweb, K-means, Learning Vector Quantization (LVQ), variantes de K-means, y otros. Luego de algunas pruebas, K-means y LVQ fueron los algoritmos elegidos a utilizar. LVQ es una familia de algoritmos para clasificación de patrones estadísticos que crea prototipos de aprendizaje representando regiones de grupos. Las regiones se definen por hiperplanos entre los prototipos. Son algoritmos emparentados a los mapas autoorganizados (SOM). LVQ fue inventado por Teuvo Kohonen (Nova & Estevez, 2015). Ambos algoritmos se corrieron para 5 grupos, K-means con un máximo de 300 iteraciones, y LVQ con una tasa de aprendizaje de 0,05 y 1.000 epochs.

Como se ve en la figura 85, los apilados de EVI2 y NDVI se cruzaron con el mapa binario 2 para enmascarar los píxeles clasificados como no-agrícola y usar solo los agrícolas, y cada apilado enmascarado se clasificó con LVQ y K-means, obteniéndose los mapas de grupos. En la figura 86, como ejemplo, se ve el mapa de grupos obtenido para EVI2-KMEANS, y en una zona ampliada se observa como a cada uno de los píxeles clasificados como agrícola por binario 2, K-means los agrupó en los grupos 1 a 5. Similares a este son el resto de los mapas clasificados.

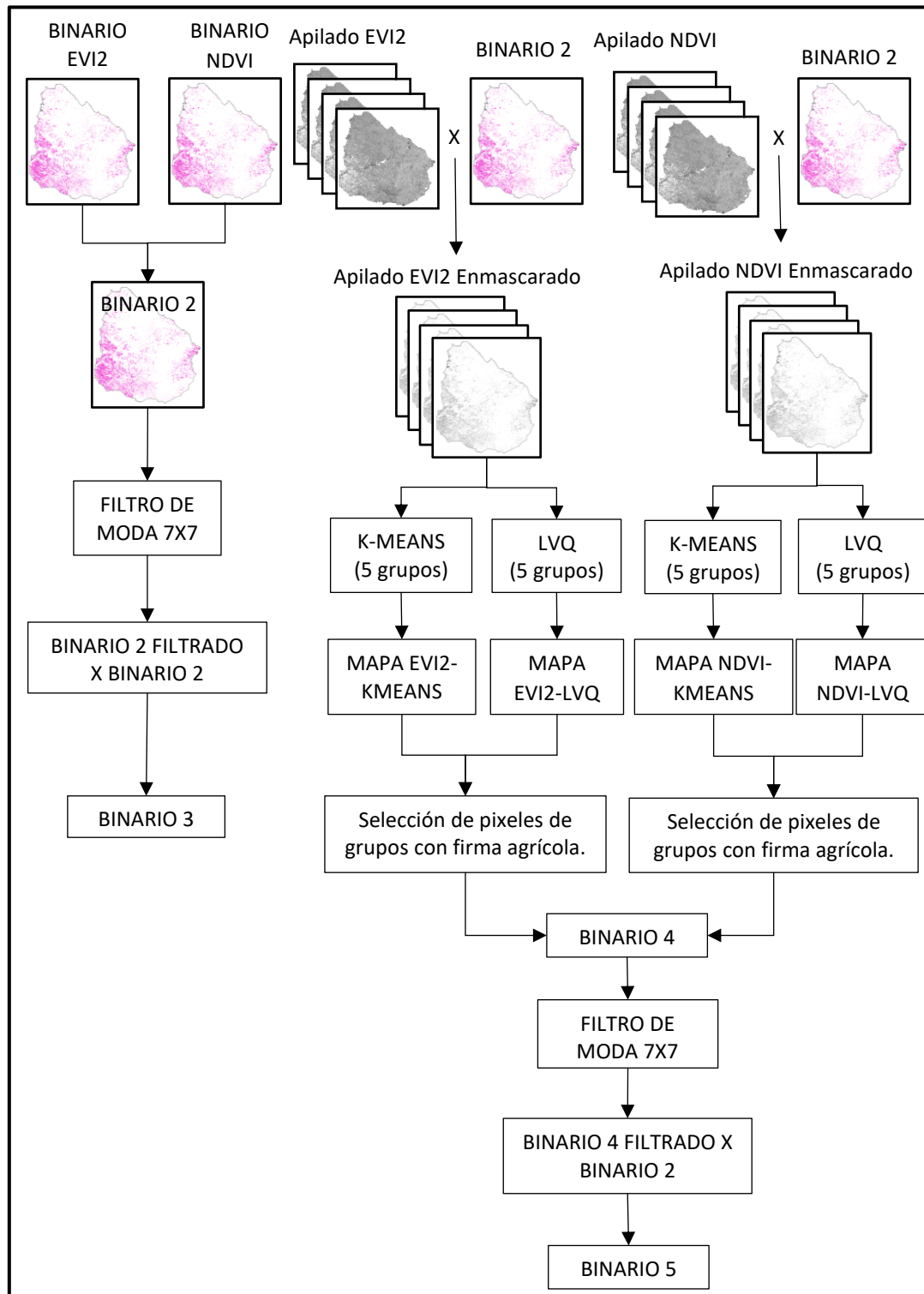


Figura 85. Flujograma de estrategias de filtrado de pixeles espurios.

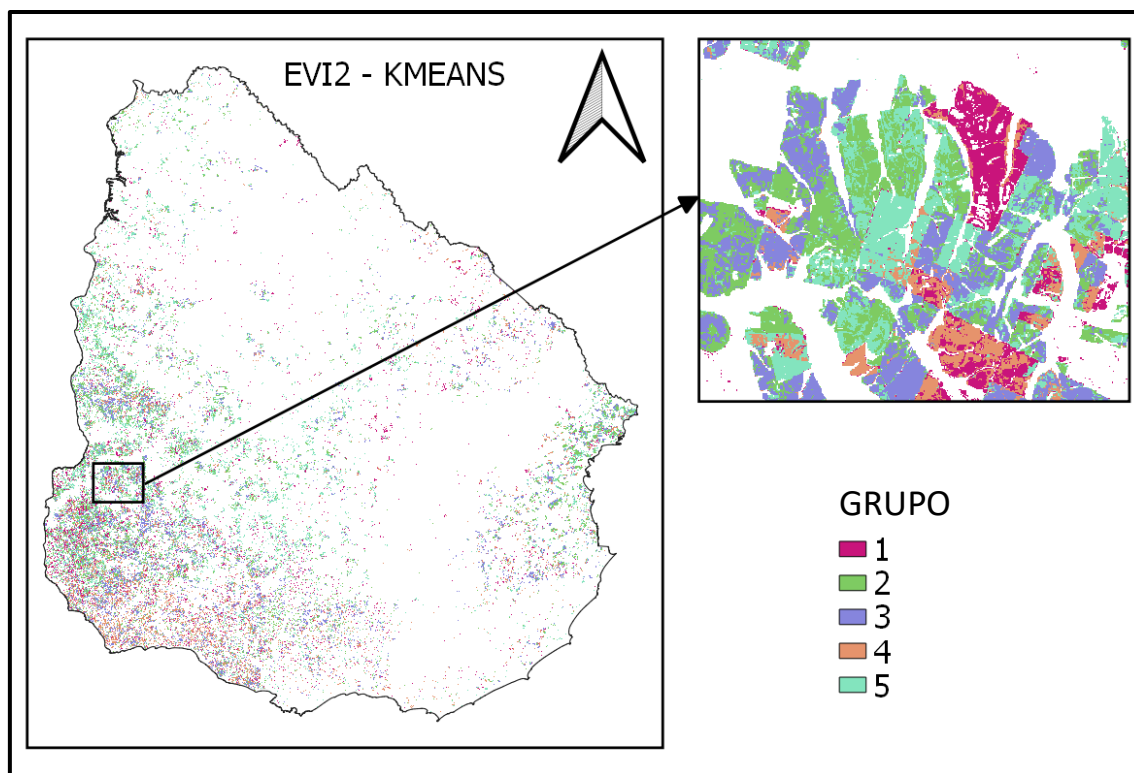


Figura 86. Mapa clasificado con K-means para el apilado EVI2 enmascarado.

Una vez clasificados los apilados enmascarados para cada combinación de índice y algoritmo (EVI2-LVQ, EVI2-KMEANS, NDVI-LVQ y NDVI-KMEANS) se generaron los gráficos de las firmas individuales y medianas de los grupos, y el gráfico de t-SNE. Como se mencionó en la sección 9.2, t-SNE es una técnica para visualizar datos con alta dimensionalidad en un espacio de menor dimensionalidad, en este caso de 15 a 2 dimensiones. Aquí se usa para evaluar visualmente la calidad de la clasificación no supervisada, ya que cada punto en el gráfico t-SNE se lo etiqueta y colorea según el grupo que le asignó la clasificación correspondiente (que se hizo en 15 dimensiones), y si la clasificación funcionó bien, los puntos pertenecientes a un mismo grupo deberían verse más cercanos entre sí que a otros, y a su vez, puede verse cuales son los grupos más similares y disimiles entre sí. Usando como ejemplo la clasificación para EVI2-KMEANS, en las figuras 87 a 89 se ven los gráficos correspondientes. Como ya se mencionó, la firma de EVI2 y NDVI de un cultivo agrícola tiene la forma similar a una función gaussiana (ver sección 9.2), es así como analizando en conjunto los gráficos de t-SNE, los de grupos, los de firmas medianas, y explorando la distribución espacial de los grupos en los mapas, se seleccionaron aquellos considerados agrícola, y se descartaron los grupos que no y todos los pixeles pertenecientes a esos, filtrando así pixeles espurios.

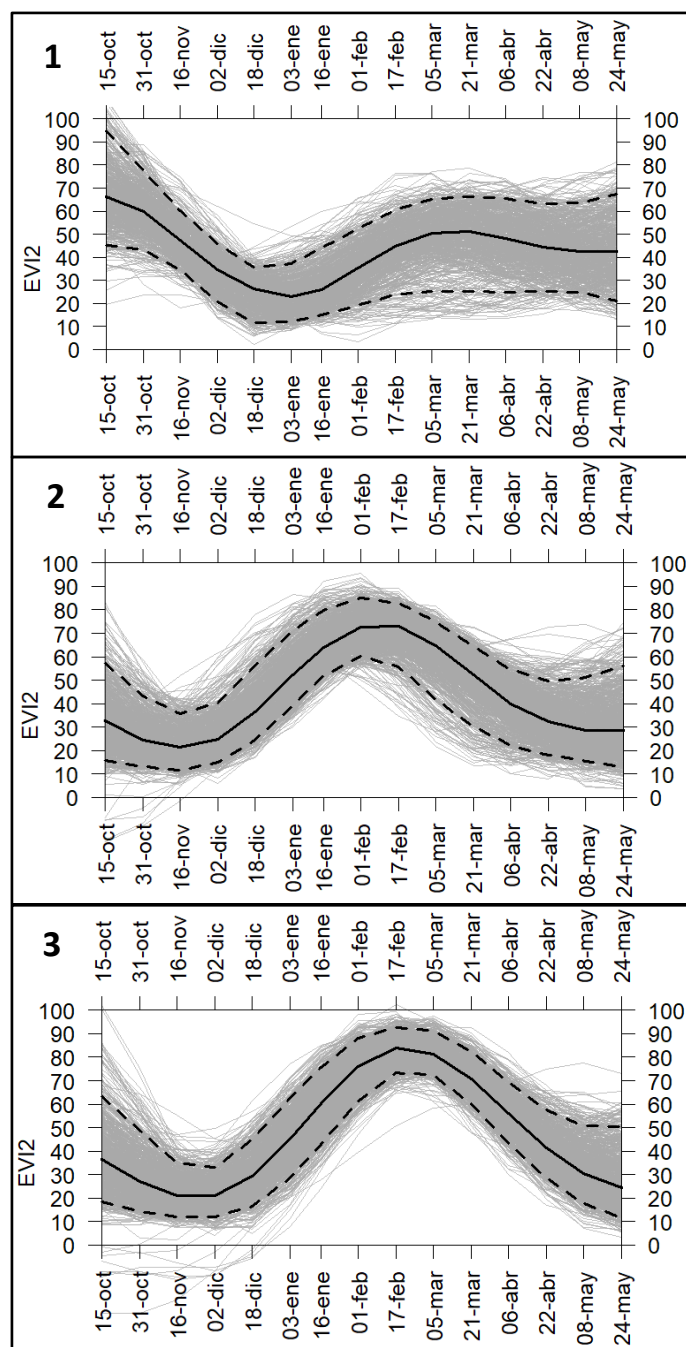


Figura 87. Firmas individuales y medianas de los grupos 1 a 3 para apilado EVI2 enmascarado clasificado con K-means.

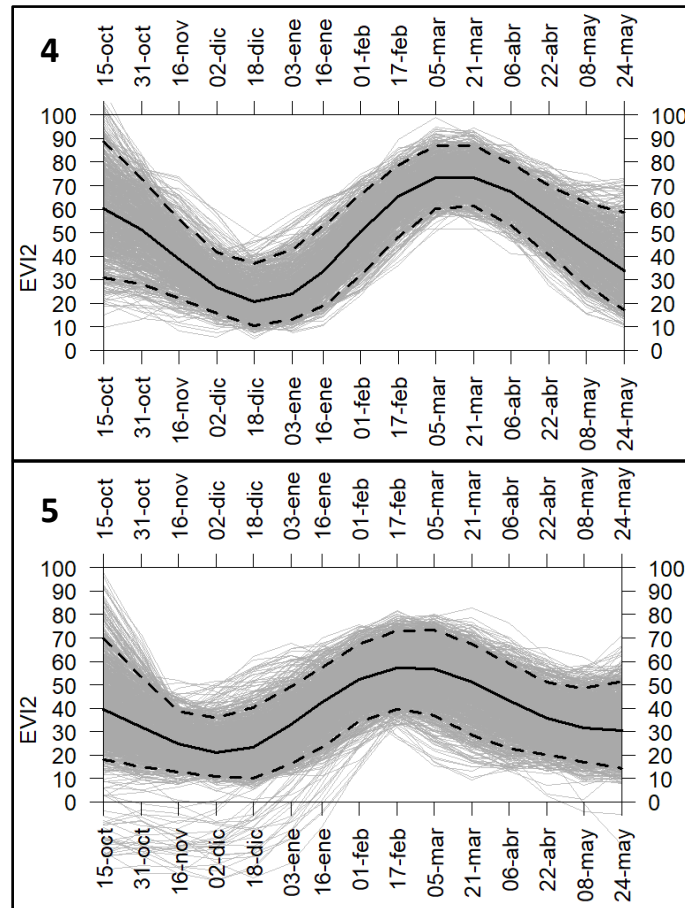


Figura 88. Firmas individuales y medianas de los grupos 4 y 5 para apilado EVI2 enmascarado clasificado con K-means.

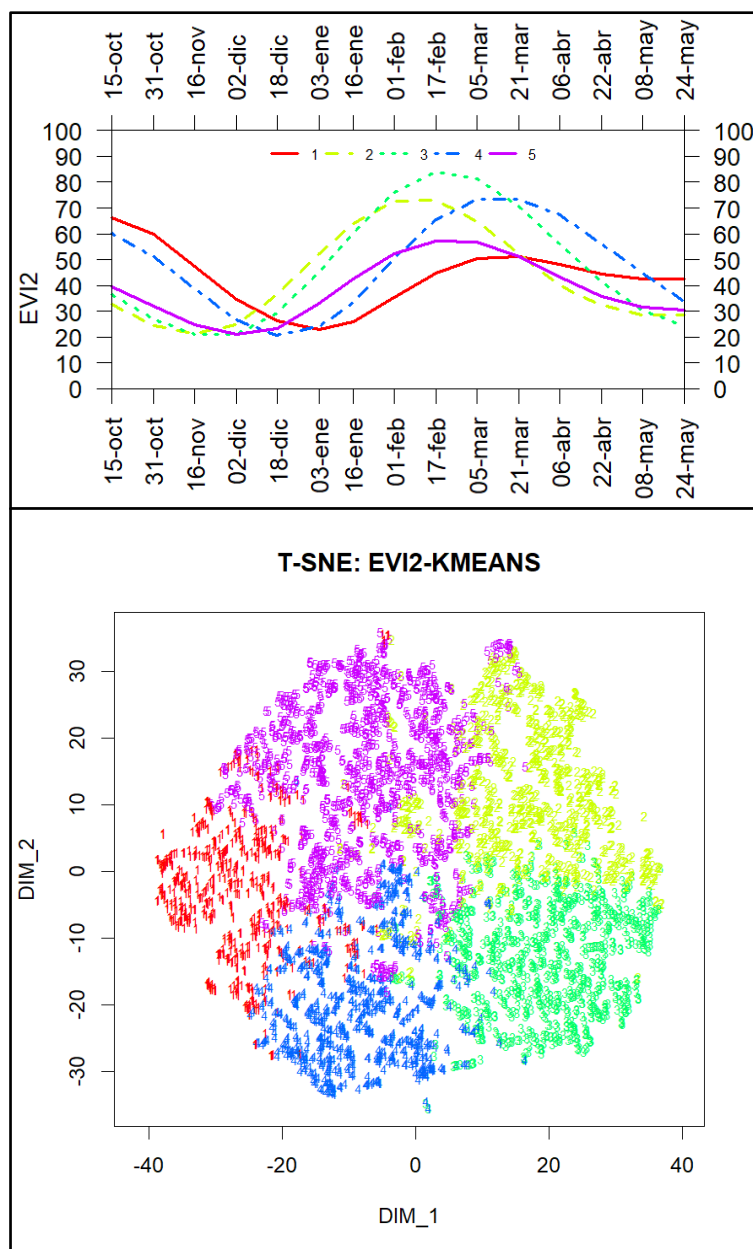


Figura 89. Firmas medianas de los grupos y gráfico t-SNE para apilado EVI2 enmascarado clasificado con K-means.

Del análisis de los gráficos y la exploración visual del mapa clasificado para EVI2-KMEANS, los grupos 2 a 5 serían agrícolas, entre otras cosas porque sus firmas medianas tienen la forma típica agrícola, mientras que el grupo 1 no, y a su vez todos los puntos de este grupo forman un conglomerado relativamente homogéneo en el gráfico t-SNE. Aun así, dentro del grupo 1 hay algunos puntos que por su ubicación en el gráfico pertenecerían más al 4 o 5, pero la mayoría de los puntos están bien agrupados en el 1. Si se observa el gráfico de firmas individuales del grupo 1 de la figura 87 parecería que algunas firmas se alejan de la firma mediana del grupo y ser más bien agrícolas, probablemente esas firmas sean los puntos entremezclados en los grupos 4 y 5.

Este mismo procedimiento se repitió para analizar los gráficos y los mapas de las otras combinaciones de índice y algoritmo, EVI2-LVQ, NDVI-KMEANS y NDVI-LVQ, descartándose los grupos sin patrón agrícola. Es importante aclarar que las clasificaciones son independientes

entre sí, por lo que lo que un conjunto de píxeles que pertenecen al grupo 1 en EVI2-KMEANS, no necesariamente van a pertenecer al 1 de NDVI-LVQ.

Esta metodología para filtrar píxeles espurios tiene un componente subjetivo que depende del usuario que lo haga, pudiendo haber situaciones que no sea tan claro definir si un grupo es agrícola o no. Una vez seleccionados los grupos agrícolas se obtuvo el mapa binario 4, al que posteriormente se le aplicó un filtro de mediana de 7x7 obteniéndose el mapa binario 5. En la figura 90 para una zona reducida se muestra la comparación de los cuatro mapas binarios, en donde se ve el efecto de la aplicación de las distintas estrategias de eliminación de píxeles espurios.

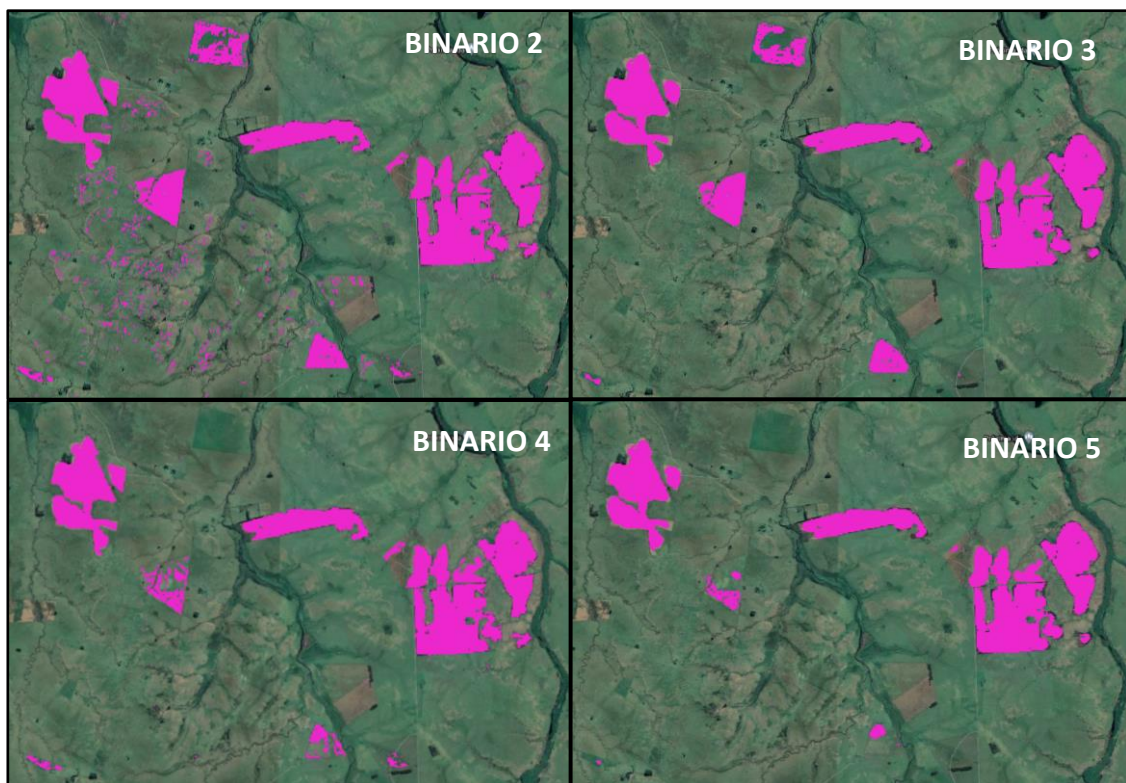


Figura 90. Comparación de los mapas binarios 2 a 5 en un área reducida.

Un aspecto para destacar en el mapa binario 2 es la calidad de la segmentación del mapa, es decir la capacidad de identificar aglomerados o zonas homogéneas de píxeles pertenecientes a la clase agrícola. Para eso, haciendo una exploración visual del mapa se notó que hay poco efecto de “sal y pimienta” como se observa en la figura 90. Por otro lado, para algunos lugares donde se tienen los polígonos de las chacras de la base FUCREA-URF y de PLANES (delimitadas en color cian), se observó que el mapa tiene una buena capacidad de identificar la ubicación y forma de las chacras como se ve en la figura 91.

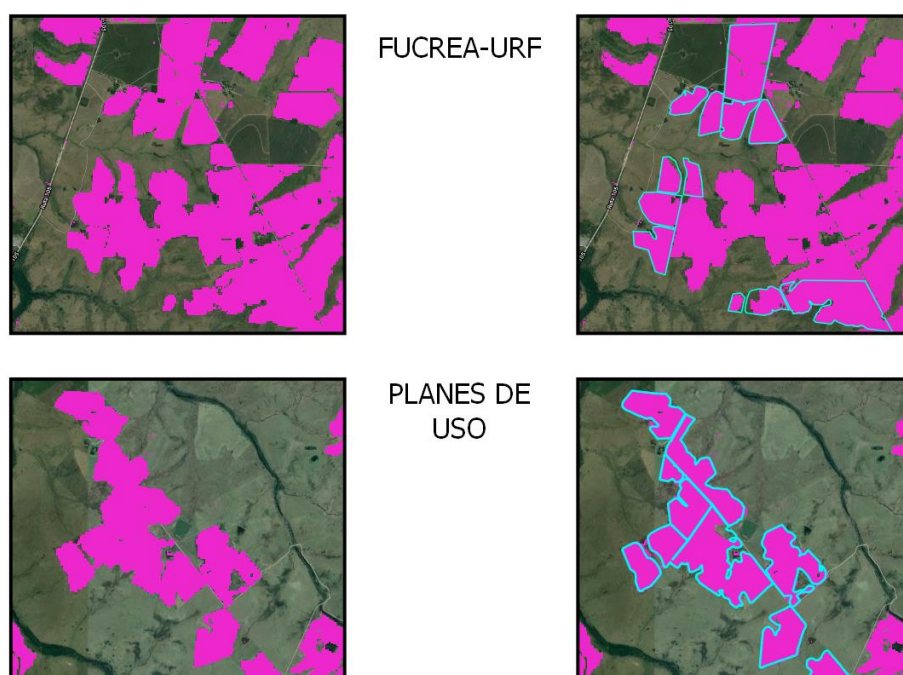


Figura 91. Comparación de delimitación de chacras en mapa binario 2 contra polígonos de chacras de FUCREA-URF y PLANES.

Una vez obtenidos los cuatro mapas binarios según las estrategias de la figura 85 se procedió a evaluar la calidad de cada uno al cruzarlos con los 5.500 puntos de evaluación, obteniéndose para cada uno la matriz de confusión y las métricas de desempeño de las clases agrícola/no-agrícola. En la tabla 56 se presenta el área de cada mapa, los valores de las métricas, y como referencia el área agrícola total (arroz, caña de azúcar, maíz, girasol, sorgo y soja) estimada según DIEA para la zafra 2017-18.

MAPA	AREA (HA)	PRECISION		RECALL		F1-SCORE	
		NO AGRICOLA	AGRICOLA	NO AGRICOLA	AGRICOLA	NO AGRICOLA	AGRICOLA
BINARIO 2	1.912.724	98,7	98,3	98,3	98,7	98,5	98,5
BINARIO 3	1.701.515	97,0	99,4	99,4	96,9	98,2	98,2
BINARIO 4	1.728.909	95,2	98,5	98,6	95,0	96,9	95,1
BINARIO 5	1.542.476	92,2	99,4	99,5	91,5	95,7	95,3
DIEA 2017-18	1.367.600						

Tabla 56. Área (hectáreas) y métricas de desempeño para los mapas binarios 2 a 5.

Como se ve en la tabla 56, para cualquiera de los mapas binarios, *F1* es superior a 95, tanto para agrícola y no-agrícola. El mayor *F1* lo obtienen los mapas binarios 2 y 3, con valores mayores a 98, pero producto del filtro de moda, el área de binario 3 es más parecida al valor de referencia de DIEA, con una superficie de 1.701.515 hectáreas, mientras que binario 2 tiene una superficie de 1.912.724 hectáreas. Aquí se ve que producto del filtro de moda se eliminaron 211.209 hectáreas entre binario 2 y 3, pero la calidad de la clasificación no se vio afectada, al tener ambos similares *F1*, por lo cual es una buena medida aplicar el filtro de moda. Al mismo tiempo, binarios 4 y 5, tienen valores de *F1* inferiores a binario 2 y 3. En el caso de binario 4, el filtro de píxeles espurios se hizo a través de clasificación no supervisada, como se explicó anteriormente y se muestra en el flujograma de la figura 85. Aquí se ve que por efecto de la estrategia aplicada, *F1* es de 95,1 y 96,9 para agrícola y no-agrícola respectivamente, y la superficie es de 1.728.909 hectáreas. En comparación a binario 2, *F1* es menor, pero aun así mayor a 95, y la superficie casi 190.000 hectáreas menos. Al comparar binario 4 contra binario 3, que tienen diferente

estrategia de eliminación de píxeles espurios, ambos mapas tienen una superficie similar, pero binario 3 tiene mejor desempeño con un *F1* superior en ambas clases. Finalmente, está binario 5 que es resultado de aplicar filtro de moda a binario 4. Al hacer esto, *F1* para no-agrícola disminuye algo respecto a binario 4, mientras que para agrícola se mantiene en valores similares. Este último mapa tiene una superficie de 1.542.476 hectáreas, un 13% más del valor de referencia de DIEA.

Con respecto a *precision* se observa que cualquiera sea el mapa binario, para la clase agrícola esta es superior a 98, y con mejor desempeño en binario 3 y 5, los cuales tienen en común filtro de moda y tuvieron valores superiores a 99. En el caso de la clase no-agrícola, binario 2 y 3 fueron los de mejor desempeño con valores de al menos 97, y en el caso de binario 4 y 5, que tienen filtro por clasificación no supervisada, tienen valores de 95,2 y 92,2 respectivamente. Aquí se ve que en binario 5 por efecto del doble filtro (clasificación no supervisada y filtro de moda), *precision* es 6,5 puntos menos respecto a binario 2. Cuando se analiza *recall*, la situación es inversa a *precision*, ya que ambas están relacionadas. Empezando por la clase no agrícola, *recall* es superior a 98 para cualquier de los binarios, pero con mejor desempeño en 3 y 5, los cuales tienen filtro de moda. Para la clase agrícola, binario 2 y 3 son los de mejor desempeño, con valores superiores a 96, y en el caso de binario 5, el cual tiene doble filtrado, es el de peor desempeño, con un *recall* de 91,5, unos 7,2 puntos menos que binario 2.

Visto y analizados los resultados de la tabla 56, surge el planteo de cuál es la clasificación más adecuada. Aquí hay que recordar, que el valor de DIEA se obtiene a partir de encuestas, no de un censo, y que no se conoce el margen de error, es decir entre qué valor máximo y mínimo está esa estimación. Si se toma como criterio de elección del mapa, el área estimada, binario 5 sería la mejor opción por tener una superficie un 13% superior al valor de referencia de DIEA, 1.542.476 contra 1.367.600 hectáreas. Pero si se presta especial atención a los valores de *precision* y *recall* de las clases no-agrícola y agrícola respectivamente, estos son de los más bajos de todos los mapas, con valores en torno a 92. En el otro extremo, está binario 2, que tiene valores de *precision*, *recall* y *F1* superiores a 98 para cualquiera de las clases, pero con la mayor sobrestimación del área de DIEA, de un 40%. Por último, están los mapas binarios 3 y 4, ambos con similar área estimada de aproximadamente 1.700.000 hectáreas, casi un 25% más que el área de DIEA, pero en el caso de binario 3 sus métricas de desempeño son mejores que las de binario 4. Teniendo en cuenta esto, la opción de binario 3 parece ser la más equilibrada en cuanto a superficie estimada y métricas de desempeño.

Además de la evaluación de la calidad de la clasificación con puntos de COBERTURAS y FUCREA-URF, se hizo otra evaluación con 2.750 puntos de COBERTURAS y 2.750 de PLANES de la zafra 2017-18. Este nuevo conjunto de evaluación no tiene ningún punto en común con los conjuntos de entrenamiento y evaluación anteriores, y ninguno de los puntos agrícola de PLANES cae en alguna de las chacras de la base FUCREA-URF. En la figura 92 se muestra la distribución de estos puntos en todo el país. Como se ve, a diferencia de los puntos agrícola de FUCREA-URF, los puntos agrícolas de PLANES tienen una distribución más amplia en todo el territorio.

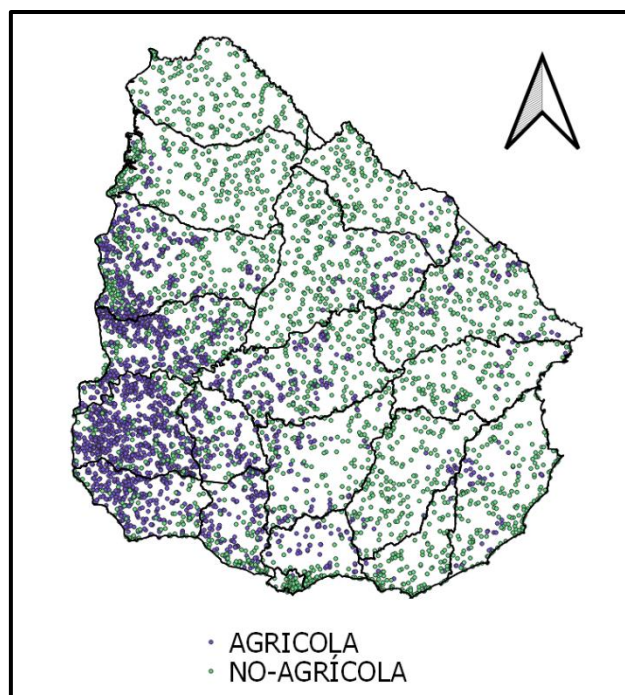


Figura 92. Mapa de distribución de puntos de evaluación de base conjunta COBERTURAS/PLANES de zafra 2017-18.

Este conjunto se usó para evaluar la calidad solo del mapa binario 3, ya que por el análisis realizado anteriormente se consideró el mapa más equilibrado de los cuatro. En la tabla 57 se presentan las métricas de desempeño en binario 3 para los conjuntos de evaluación COBERTURAS/FUCREA-URF y COBERTURAS/PLANES. Como se ve en la tabla, las métricas de COBERTURAS/PLANES son inferiores a las de COBERTURAS/FUCREA-URF, pero con valores superiores a 93. Cualquiera sea la clase, *F1* es de casi 96 en COBERTURAS/PLANES.

Para *precision* y *recall* de agrícola y no-agrícola, COBERTURAS/PLANES tiene valores algo superior a 98, mientras que para no-agrícola y agrícola valores algo superiores a 93.

CONJUNTO DE EVALUACIÓN	PRECISION		RECALL		F1-SCORE	
	NO AGRICOLA	AGRICOLA	NO AGRICOLA	AGRICOLA	NO AGRICOLA	AGRICOLA
COBERTURAS/FUCREA-URF	97,0	99,4	99,4	96,9	98,2	98,2
COBERTURAS/PLANES	93,5	98,1	98,2	93,2	95,8	95,6

Tabla 57. Métricas de desempeño en mapa binario 3 según conjunto de evaluación.

Teniendo en cuenta que los puntos de PLANES tienen una mayor distribución espacial y no están concentrados en la zona de distribución de los puntos de FUCREA-URF, que es donde están los puntos agrícolas usados para entrenar los modelos, las métricas de desempeño en COBERTURAS/PLANES son buenas y validan la calidad del mapa binario 3. También para COBERTURAS/PLANES se extrajeron de manera aleatoria 50 puntos clasificados como agrícola y 50 clasificados no-agrícola por binario 3, y para cada subconjunto se graficaron las firmas individuales de NDVI como se ve en la figura 93. Esto se hizo para explorar visualmente si la forma de las firmas individuales era agrícola o no. Como se ve, la mayoría de los puntos clasificados agrícola, su firma es agrícola, y para los puntos clasificados no-agrícola, la mayoría tiene firma no-agrícola. No obstante, se nota que hay algún punto en uno u otro subconjunto que su firma no coincide con la clasificación, es decir, siendo agrícolas, fueron clasificados no-agrícola y viceversa.

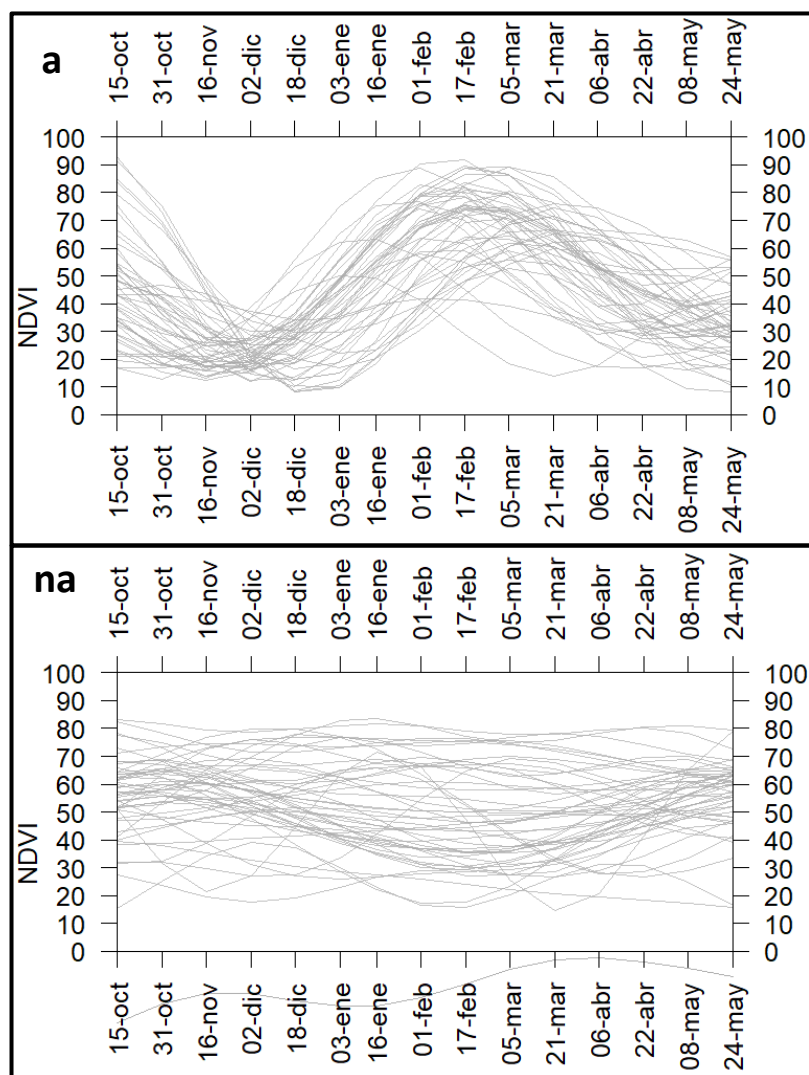


Figura 93. Gráficos de NDVI para 50 puntos aleatorios agrícolas (a) y 50 no-agrícolas (na) del conjunto de evaluación COBERTURAS/PLANES.

Visto todo lo anterior, de la clasificación binaria destaca lo siguiente:

- Los píxeles identificados como agrícolas están distribuidos principalmente en regiones agropecuarias agrícolas (agrícola de secano y arroceras) según mapa de DIEA de la figura 80.
- Buena capacidad para segmentar y agrupar píxeles como agrícolas en zonas homogéneas, con poco efecto de “sal y pimienta”, y delimitar bien la forma de las chacras como se observa en las figuras 90 y 91.
- Tomando binario 3, este sobrestima el área de DIEA en un 24% (300.000 hectáreas más). Con respecto al área total del país, binario 3 representa un 9,7% y DIEA un 7,8%.
- La métrica de $F1$ para agrícola/no-agrícola y ambos conjuntos de evaluación está por encima de 95 como se ve en las tablas 56 y 57.
- Los gráficos de NDVI de los puntos aleatorios para COBERTURAS/PLANES tienen la forma esperada para agrícola y no-agrícola.

11.2 Mapa de cultivos agrícolas a nivel de chacra en GEE

En esta sección se presentan los resultados de la clasificación supervisada para identificar el cultivo agrícola específico (maíz, soja, sorgo) a nivel de chacra. Dado que el mapa binario 2 identificó chacras en regiones agrícolas de secano y arroceras como se ve en el mapa de la figura 80, y que en la base de FUCREA-URF solo hay información de chacras de maíz, soja y sorgo, se decidió trabajar con las chacras de binario 2 presentes en los departamentos de la zona de influencia de FUCREA-URF y enmascarar el resto, como se ve en la figura 94. Aunque en los departamentos de Canelones y Lavalleja hay algunas chacras de FUCREA-URF, por ser muy pocas, se decidió excluir estos departamentos de la zona de influencia para el trabajo reportado en esta sección.

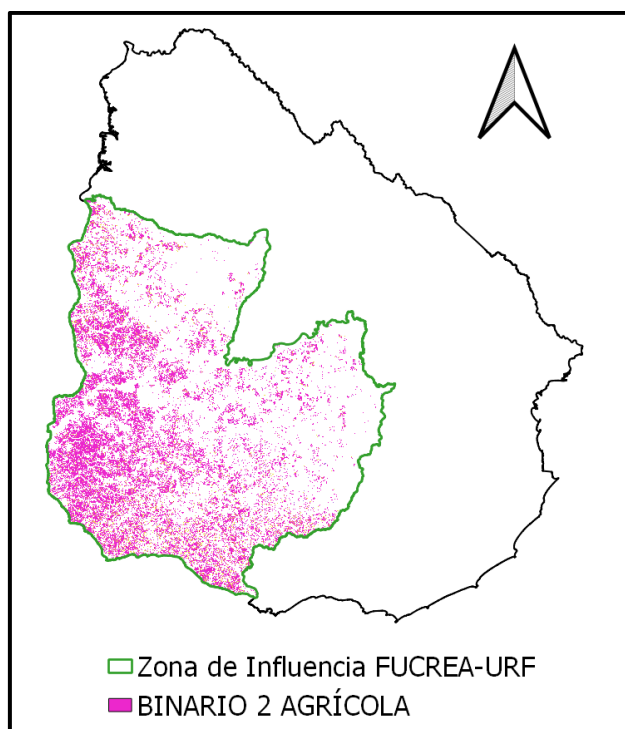


Figura 94. Mapa de chacras identificadas en binario 2 dentro de la zona de influencia de FUCREA-URF.

Una vez definida el área de trabajo, en GEE se generaron los scripts de clasificación supervisada. Al igual que en la clasificación binaria agrícola/no-agrícola se decidió trabajar con los algoritmos de Random Forest (RF) y Gradient Tree Boost (GTB), y los apilados de EVI2 y NDVI. A efectos de facilitar el proceso de clasificación, estos apilados se combinaron en un único llamado EVI2-NDVI, en este se enmascararon las chacras de binario 2 fuera de la zona de FUCREA-URF, y se obtuvo el apilado enmascarado EVI2-NDVI. De la base de FUCREA-URF se seleccionaron aleatoriamente y de manera estratificada 5.500 puntos de maíz, soja y sorgo. Esos puntos se dividieron de manera estratificada para entrenamiento y evaluación, en 70 y 30%, es decir, unos 3.850 para entrenamiento y para evaluación 1.650 (1.639 dentro de zona de influencia), como se muestra en la tabla 58 y figura 95.

CULTIVO	CANTIDAD DE PUNTOS	
	ENTRENAMIENTO	EVALUACIÓN
MAÍZ	1.056	452
SOJA	2.309	979
SORGO	485	208
TOTAL	3.850	1.639

Tabla 58. Cantidad de puntos por cultivo para entrenamiento y evaluación en la zona de influencia de FUCREA-URF.

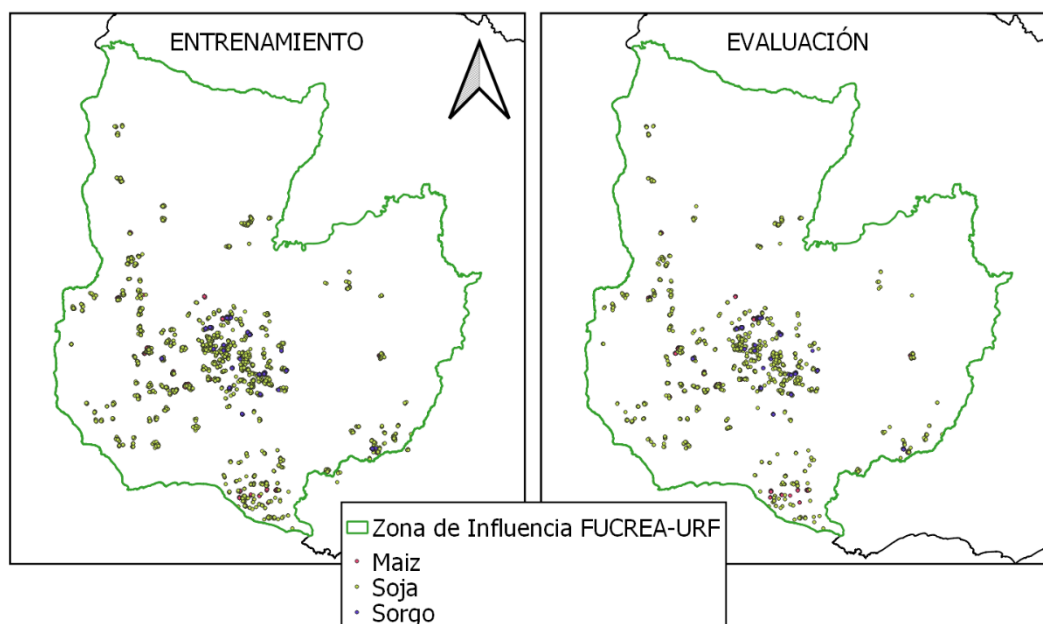


Figura 95. Mapa de distribución de puntos de entrenamiento y evaluación de cultivos en la zona de influencia de FUCREA-URF.

Posteriormente el apilado EVI2-NDVI se cruzó con los 3.850 puntos de entrenamiento y se obtuvieron las series temporales y con estas se entrenaron RF y GTB. Finalmente, el apilado enmascarado EVI2-NDVI se clasificó por separado con RF y GTB, y se obtuvieron los mapas clasificados respectivos, a los cuales se les aplicó un filtro de moda 7x7 para reducir el efecto de “sal y pimienta” y eliminar píxeles espurios de binario 2, obteniéndose los mapas GTB-MODA y RF-MODA. En la figura 96 se presenta el flujograma de todo el proceso de clasificación.

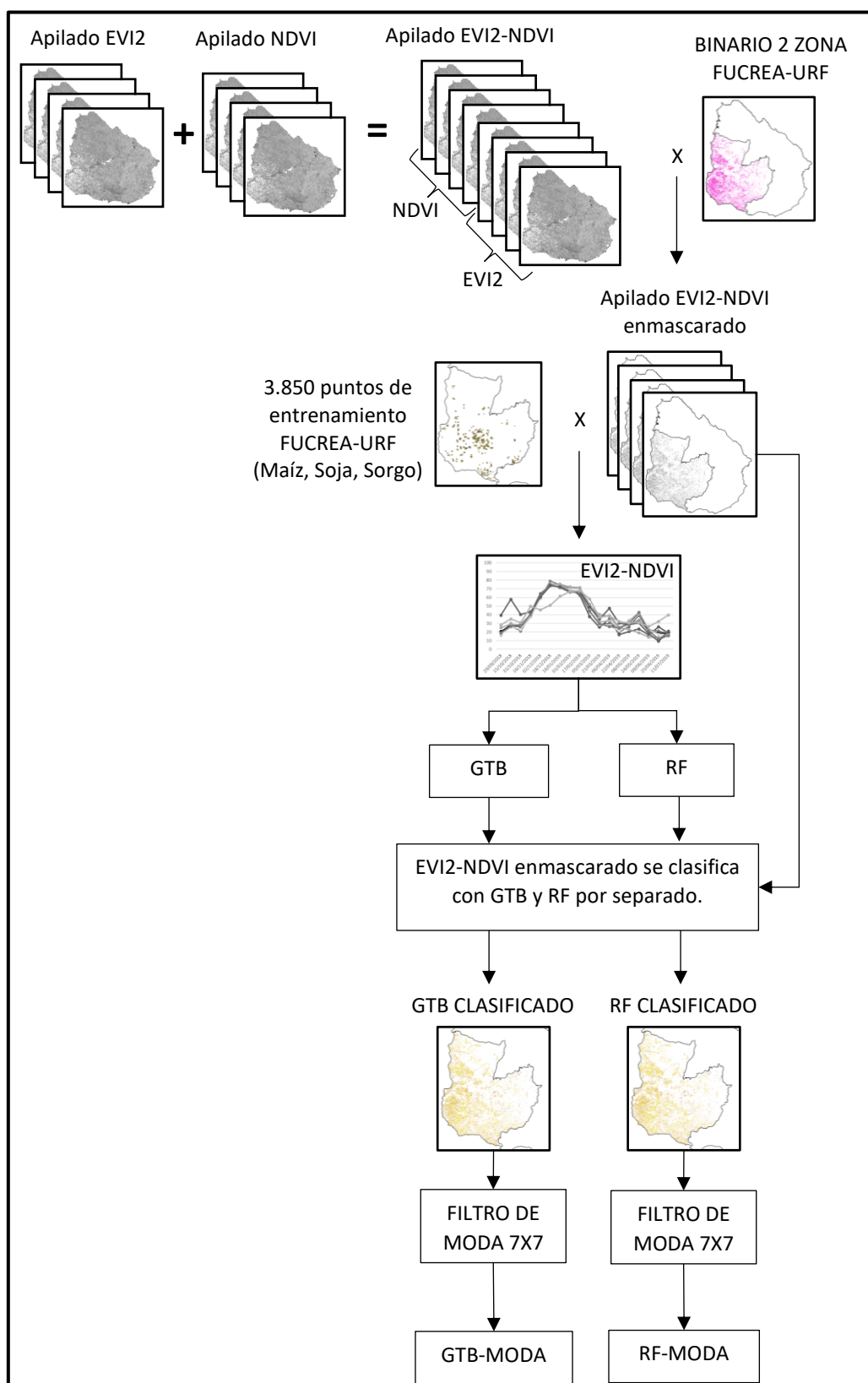


Figura 96. Flujograma de clasificación supervisada por cultivo para chacras de Binario 2 en zona de influencia FUCREA-URF.

En la figura 97, tomando como ejemplo una zona donde hay chacras de maíz, soja y sorgo de la base FUCREA-URF, se muestra el resultado de la clasificación por GTB y RF, con y sin filtro de

moda. Para empezar, se nota en las chacras de FUCREA-URF que dentro de una misma chacra hay zonas que en binario 2 fueron identificadas como agrícolas y otras no-agrícolas. Esto se debe a las siguientes causas: por un lado, dentro de una chacra hay zonas no sembrables, como ser zonas bajas, caminos internos, montes de abrigo, etc. Por otro, puede no haberse sembrado toda la chacra, solo una parte. Y para finalizar, a pesar de haberse sembrado toda una chacra, pueden existir fallos y algunas zonas donde no germinara el cultivo. Como muestra para la chacra de la figura 98 se presenta los gráficos de NDVI de las zonas agrícolas y no-agrícolas.

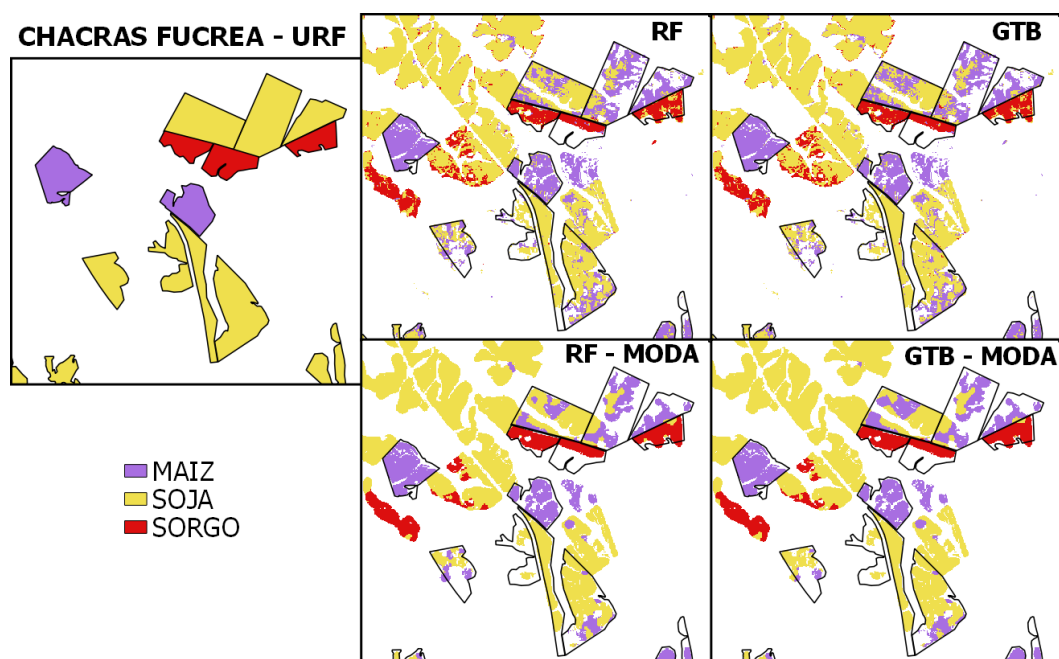


Figura 97. Ejemplo de chacras de FUCREA-URF clasificadas con GTB y RF, con y sin filtro de moda.

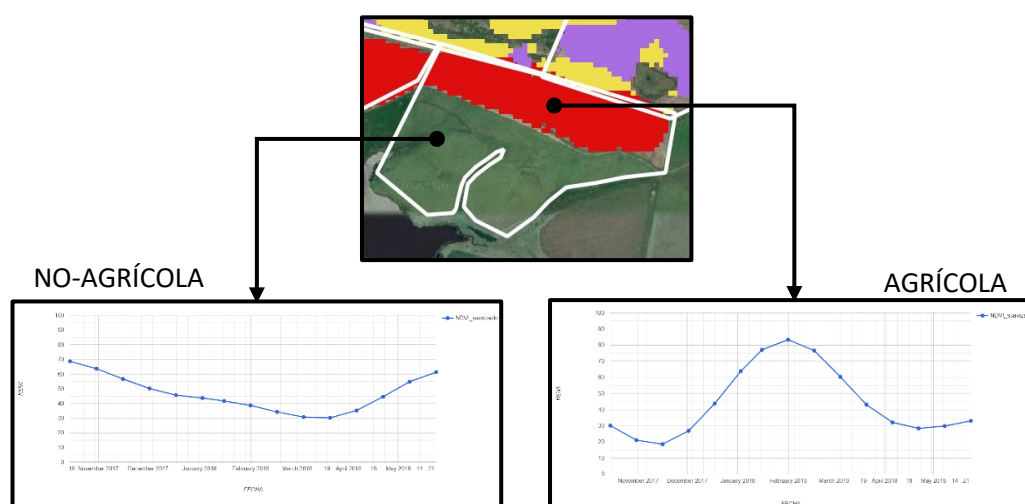


Figura 98. Gráfico de NDVI para zonas agrícola y no-agrícola de una misma chacra.

Analizando los resultados de las clasificaciones, se nota para el ejemplo de la figura 97 que las mismas logran identificar relativamente bien las chacras que tienen maíz y sorgo, mientras que para las chacras que son soja, hay una mezcla de zonas identificadas como maíz y otras como

soja dentro de las chacras. También se nota en otras zonas del mapa, que hay chacras donde hay zonas de soja-sorgo, y otras donde se mezcla maíz y sorgo. Este problema, como se ve, no lo soluciona el filtro de moda. Si bien los mapas filtrados presentan poco efecto de “sal y pimienta” y zonas homogéneas de cultivo, se da el problema de la mezcla de cultivos dentro de las zonas que se corresponden a una misma chacra como se observa en la figura 99. Esto no tiene lógica, ya que por cuestiones operativas y de logística, una chacra se siembra con un único cultivo, no se hacen mezcla de cultivos. De esta manera lo que muestra los mapas de GTB y RF son fallos de clasificación. En otros casos, parece delimitar bien zonas que son de un cultivo u otro.

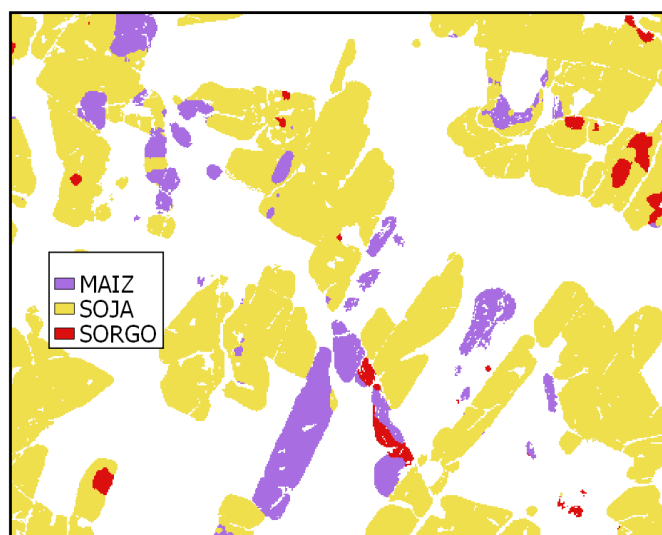


Figura 99. Ejemplo de chacras en donde se identifican zonas con 2 cultivos diferentes.

A continuación se presenta en las figuras 100, 101, y tablas 59 y 60, el área y la proporción estimada para cada cultivo agrícola en GTB-MODA y RF-MODA, y las métricas de desempeño calculadas en evaluación.

En cuanto al área estimada, se observa en la tabla 59 que tanto GTB y RF estiman áreas similares totales y por cultivo. Si se analiza el porcentaje del área agrícola ocupada por cada cultivo, esta es de más del 87% para soja, más de 8% para maíz y 3% de sorgo. Estos valores están en concordancia con los porcentajes de estos cultivos a nivel nacional según DIEA para la zafra 2017-18 como se ve en la tabla 60, con valores en DIEA de 91,7, 5,9, y 2,4% para soja, maíz y sorgo respectivamente. El área estimada total y por cultivos en la zona de influencia de FUCREA-URF supera los valores de DIEA a nivel nacional.

MAPA	ÁREA (HA)				PORCENTAJE DEL ÁREA		
	MAÍZ	SOJA	SORGO	TOTAL	MAÍZ	SOJA	SORGO
RF-MODA	115.002	1.091.414	36.685	1.243.100	9,3	87,8	3,0
GTB-MODA	101.977	1.107.918	35.981	1.245.877	8,2	88,9	2,9

Tabla 59. Áreas (hectáreas) y porcentaje del área de maíz, soja y sorgo en mapas de moda de GTB y RF en zona de influencia FUCREA-URF.

CULTIVO	ÁREA (HA)	PORCENTAJE DEL TOTAL
MAÍZ	71.200	5,9
SOJA	1.099.000	91,7
SORGO	28.800	2,4
TOTAL	1.199.000	100

Tabla 60. Área (hectáreas) y porcentajes del total, de maíz, soja y sorgo para la zafra 2017-18 según DIEA.

En cuanto a las métricas de desempeño, se observa que tanto en GTB y RF, los valores de *precision*, *recall* y *F1-Score* son similares. Analizando las métricas por cultivo, se nota que *F1* es de por lo menos 80 en cualquiera de estos, siendo soja la de mejor desempeño con valores en torno a 89, seguida por maíz con valores en torno a 83, y sorgo con valores en torno a 80. Estos valores de *F1* están indicando que estos algoritmos y con los datos de entrenamiento usados, identifican mejor las áreas de soja, luego las de maíz y por último las de sorgo.

Al analizar *precision* y *recall*, se nota que este último en cualquiera de los cultivos es inferior a *precision*, sobre todo en maíz y sorgo. La *precision* es superior a 90 para cualquiera de los cultivos, mientras que *recall* está en torno a 67 para sorgo, 74 para maíz, y 86 para soja. Estos valores indican que de los puntos de evaluación, los que los algoritmos clasifican como maíz, soja o sorgo, la mayoría lo son, pero no tienen tan buen desempeño para identificar todos los puntos que son maíz, soja o sorgo del conjunto de evaluación, por eso los valores inferiores de *recall*. Es importante mencionar que para la zafra 2017-18 el universo de datos de FUCREA-URF que se dispuso para entrenar los modelos se compone de 1.133 chacras de soja, 46 de maíz y 29 de sorgo. De esta manera se observa que el conjunto de datos está muy desbalanceado y con muy poca cantidad de chacras de maíz y sorgo, lo que no permite tener una muestra que represente la variabilidad de estos cultivos en la zona de estudio.

MAPA	PRECISION			RECALL			F1-SCORE		
	MAÍZ	SOJA	SORGO	MAÍZ	SOJA	SORGO	MAÍZ	SOJA	SORGO
GTB-MODA	94,9	92,0	99,3	74,8	86,9	67,8	83,7	89,4	80,6
RF-MODA	94,3	91,0	98,6	73,5	86,7	65,9	82,6	88,8	79,0

Tabla 61. Métricas de desempeño por cultivo de GTB-MODA y RF-MODA.

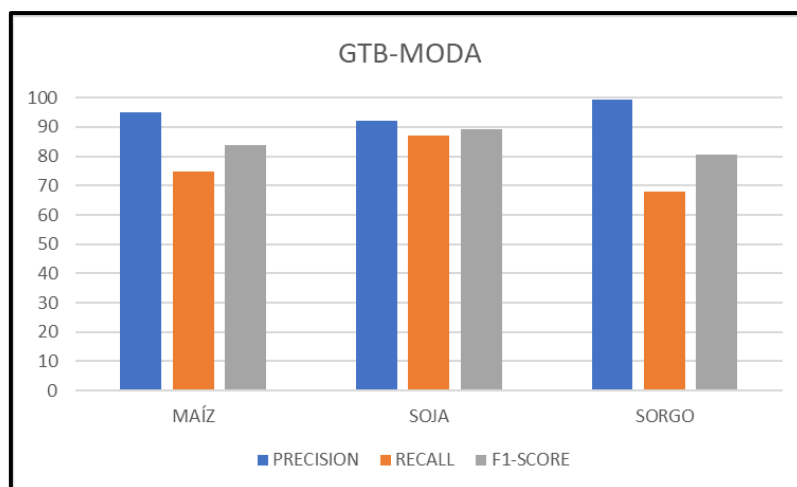


Figura 100. Gráfico de métricas de desempeño por cultivo para GTB-MODA.

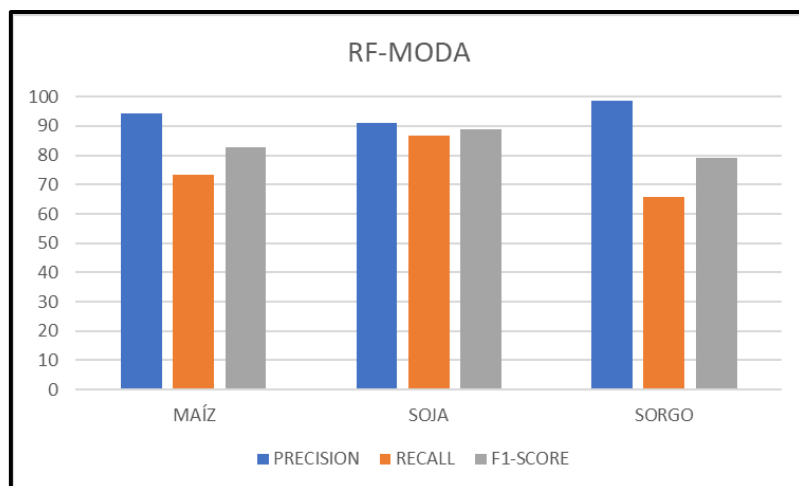


Figura 101. Gráfico de métricas de desempeño por cultivo para RF-MODA.

Visto todo lo anterior, de la clasificación supervisada de cultivos a nivel de chacra se resume lo siguiente:

- Los mapas de GTB y RF tienen poco efecto de “sal y pimienta”, y logran definir áreas homogéneas de maíz, soja o sorgo.
- Ocurre que dentro de una misma chacra hay mezcla de zonas definidas como maíz, soja o sorgo, lo cual desde el punto de vista logístico y operativo no sucede en la realidad.
- GTB y RF identifican áreas totales y por cultivo similares.
- Del área total agrícola identificada, el porcentaje ocupado por cada cultivo está en valores similares a los de DIEA.
- GTB y RF logran mejor desempeño para identificar soja, luego maíz y por último sorgo, con valores de *F1* entre 80 y 90.
- Los algoritmos usados tienen mejor desempeño en *precision* que en *recall*.
- El universo de datos usados para entrenar está muy desbalanceado hacia soja, con poca representatividad del cultivo de maíz y sorgo.

12. Seguimiento de la condición de cultivos agrícolas en la chacra durante la zafra

Uno de los objetivos específicos de este trabajo es, a partir del uso de imágenes satelitales y aprendizaje automático, poder realizar el seguimiento de la condición de los cultivos agrícolas en diferentes momentos de desarrollo y crecimiento entre siembra y cosecha. En las secciones anteriores se trabajó a escala regional de país porque lo que interesaba era poder identificar chacras con agricultura y el tipo de cultivo en estas. En esta sección interesa saber lo que ocurre dentro de la chacra durante la zafra e identificar diferentes zonas en esta a través de la evolución temporal de NDVI. Utilizando GEE se desarrolló una metodología para este objetivo. Así mismo, utilizando las series temporales de NDVI, se desarrolló una metodología para extraer las métricas fenológicas de una chacra, lo que permite determinar momentos claves del desarrollo y evolución de esta.

12.1. Zonificación de la chacra durante la zafra

Los cultivos agrícolas de verano se siembran durante la primavera y se cosechan en el otoño siguiente, por lo cual, para hacer el seguimiento de la evolución de estos durante la zafra, la ventana temporal está comprendida entre octubre y abril siguiente, según se detalla en la figura

102. En esta sección, para poder hacer un seguimiento a más corto plazo, se trabajó con imágenes satelitales sintéticas de NDVI para períodos de 10 días (decadiales) en vez de 16 días como en las anteriores. La metodología para generar estas, es la misma que se usó en la sección 8.2 y para la imputación de datos faltantes la de la 11.1 usando el modelo armónico. De esta manera, la primera imagen sintética que se generó abarca el período del 1 al 10 de octubre, la segunda del 11 al 20 de octubre, la tercera del 21 al 31 de octubre, y así sucesivamente hasta generar la última del 21 al 30 de abril, como se ve en la figura 102. En la tabla 62 se presenta el período que abarca cada imagen decadal sintética de NDVI.

IMAGEN SINTÉTICA	INICIO	FIN	DECADIAL
1	01-OCT	10-OCT	OCT-1
2	11-OCT	20-OCT	OCT-2
3	21-OCT	31-OCT	OCT-3
4	01-NOV	10-NOV	NOV-1
5	11-NOV	20-NOV	NOV-2
6	21-NOV	30-NOV	NOV-3
7	01-DIC	10-DIC	DIC-1
8	11-DIC	20-DIC	DIC-2
9	21-DIC	31-DIC	DIC-3
10	01-ENE	10-ENE	ENE-1
11	11-ENE	20-ENE	ENE-2
12	21-ENE	31-ENE	ENE-3
13	01-FEB	10-FEB	FEB-1
14	11-FEB	20-FEB	FEB-2
15	21-FEB	28-FEB	FEB-3
16	01-MAR	10-MAR	MAR-1
17	11-MAR	20-MAR	MAR-2
18	21-MAR	31-MAR	MAR-3
19	01-ABR	10-ABR	ABR-1
20	11-ABR	20-ABR	ABR-2
21	21-ABR	30-ABR	ABR-3

Tabla 62. Fechas de inicio y fin de las imágenes decadales sintéticas de NDVI.

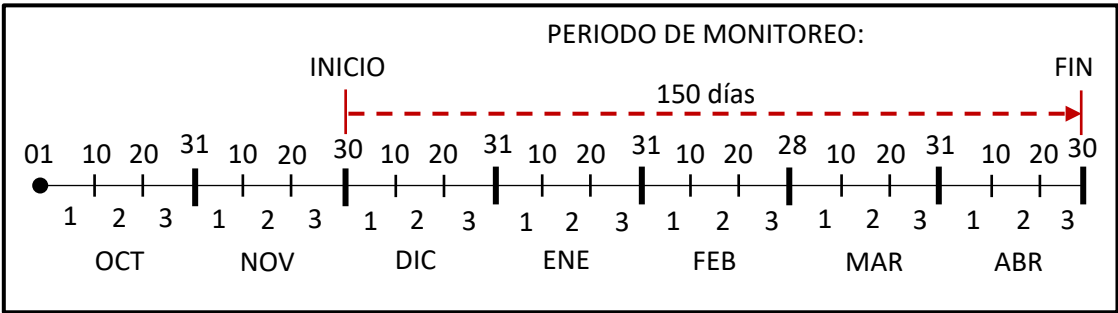


Figura 102. Ventana temporal para el seguimiento de cultivos agrícolas de verano durante la zafra.

Una vez que se generó la imagen decadal sintética DIC-1 comenzó el período de monitoreo. En ese momento, usando todas las imágenes sintéticas disponibles desde octubre a diciembre, se construyó el apilado de NDVI. Sobre éste con clasificación no supervisada, se hizo la zonificación de una chacra, y para cada zona se generó la firma fenológica mediana. En la figura 103 se

presenta el flujograma de zonificación y en la 104 el flujograma de generación de firmas fenológicas por zonas dentro de la chacra. Ambos procesos se ejecutaron en GEE.

Como se observa en la figura 103, la zonificación de la chacra se hizo de dos maneras diferentes. Una clasificando el apilado en dos zonas con el algoritmo K-Means. A esta manera se la denominó ZONIFICACIÓN BINARIA. La otra forma fue clasificando el apilado de NDVI en varias zonas. A esta se la llamó ZONIFICACIÓN MULTI-ZONA. Cuando se hizo esta última, una cuestión fue definir la cantidad de zonas a dividir la chacra. Con K-Means el usuario debe definir la cantidad de grupos (K) en el cual se deben particionar los datos, pudiendo ser el valor K elegido no óptimo. Para evitar esa circunstancia, la zonificación multi-zona se hizo con el algoritmo X-Means (Pelleg & Moore, 2000). Éste, usando la métrica Bayesian Information Criterion (BIC), determina de manera iterativa el número óptimo K para particionar un espacio de datos en determinada cantidad de grupos con K-means. El usuario debe definir cuál es el valor de K mínimo y máximo permitido, y el algoritmo entre estos límites encuentra el valor óptimo de K que maximiza BIC, y este es el utilizado para hacer la clasificación final. Para la zonificación multi-zona el valor K se buscó entre 2 y 5.

Los mapas de zonificación generados tienen algo de efecto “sal y pimienta”. Para minimizarlo se aplicó un filtro de MODA de 7x7, y como se observa esto genera zonas mejor definidas. Una vez generados los mapas filtrados de zonificación binaria y multi-zona, se generaron tres gráficos de firmas fenológicas medianas. El primero, es el gráfico de firma fenológica mediana de toda la chacra, sin tener en cuenta la pertenencia de los píxeles a alguna zona en particular. Segundo, el gráfico de firmas fenológicas medianas por zonas del mapa de zonificación binaria. Tercero, el gráfico de firmas por zonas del mapa de zonificación multi-zona. Para los gráficos de firmas medianas se consideran todas las imágenes decadales del apilado de NDVI, de manera de poder visualizar la evolución de la chacra desde la siembra. En la figura 104 se presenta el flujograma de generación de los gráficos de firmas.

El proceso de zonificación y generación de gráficos de firmas comienza cuando se tiene el apilado de NDVI de OCT-1 a DIC-1, y se repite con el siguiente apilado de OCT-1 a DIC-2, luego con el de OCT-1 a DIC-3, y así sucesivamente, hasta que se llega al último apilado de OCT-1 a ABR-3. De esta manera el proceso se ejecuta un total de 15 veces según el calendario de la figura 102.

ZONIFICACIÓN DE LA CHACRA.

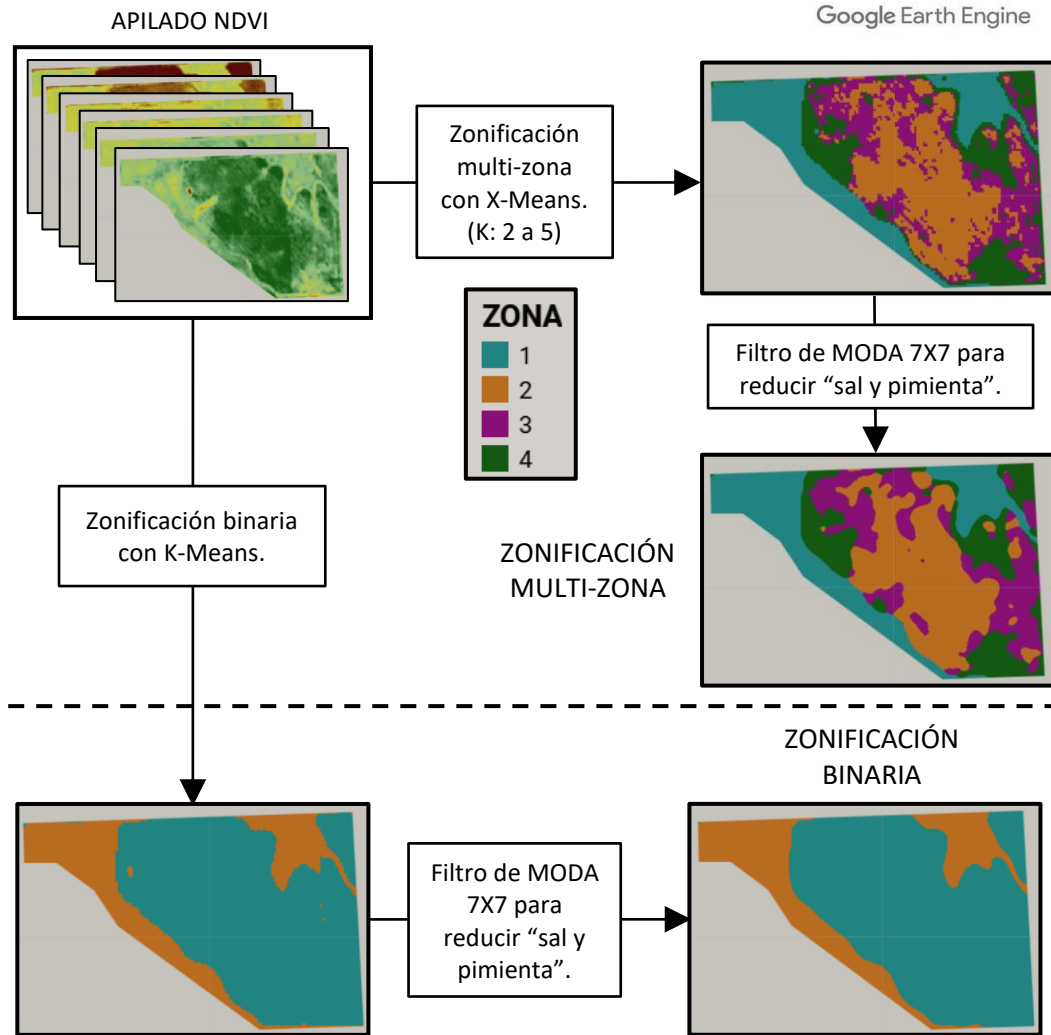


Figura 103. Flujograma de generación de mapas de zonificación de una chacra.

GENERACIÓN DE FIRMAS FENOLÓGICAS MEDIANAS.

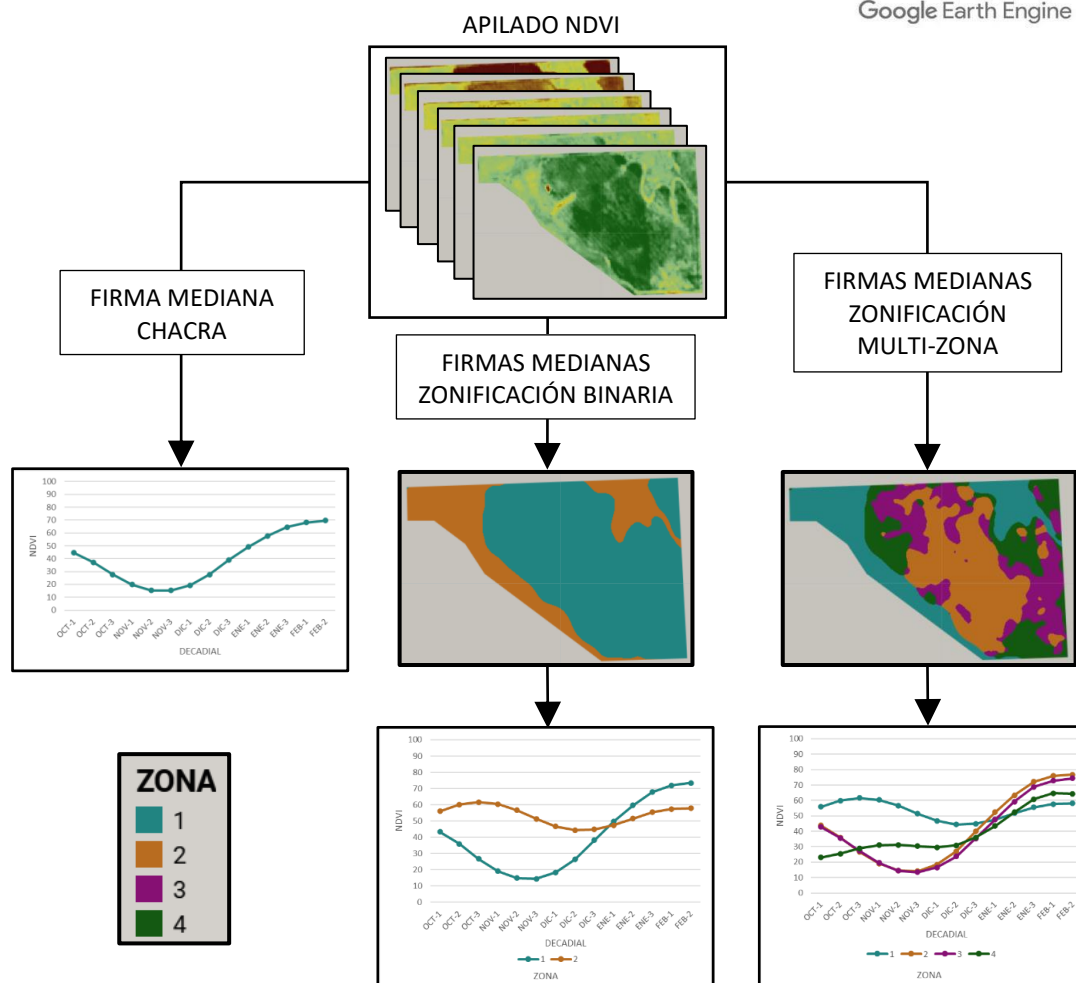


Figura 104. Flujograma de generación de firmas fenológicas medianas de una chacra.

Tomando como ejemplo una chacra de soja de la base FUCREA-URF para la zafra 2017-18 se realizó el seguimiento de esta mediante el proceso de zonificación y graficación de firmas medianas descrito anteriormente. En las figuras 105, 107, 109, 111 y 113, se presentan para cada decadal de los meses de diciembre a abril, las imágenes de color verdadero, NDVI, zonificación binaria y multi-zona de la chacra. En las figuras 106, 108, 110, 112 y 114, para la chacra se presentan los gráficos de firmas fenológicas medianas por chacra, por zonificación binaria y multi-zona, solo del decadal 2 de los meses de diciembre a abril. Es importante aclarar que cada corrida con cada apilado es independiente entre sí, en el sentido de que una zona no tiene por qué ser etiquetada con el mismo código que en una corrida anterior.

Al analizar las figuras y empezando por la zonificación binaria, se observa que esta fue muy buena para segmentar el apilado de NDVI y definir dos grandes zonas en la chacra. Cuando se observan conjuntamente las imágenes de color verdadero, NDVI y zonificación binaria, claramente estas últimas determinan muy bien los límites en donde empieza una zona u otra. A su vez, la delimitación de las zonas es muy constante a lo largo de todo el período de monitoreo, en el sentido de que los límites de las zonas varían muy poco entre una y otra corrida. Al mismo tiempo, si se observan los gráficos de las firmas medianas, sobre todo a partir de FEB-2, se nota claramente que la zona 1 corresponde al área efectivamente agrícola, mientras que la zona 2 es el área no sembrada. La forma de la firma mediana de la zona 1 es Gaussiana, mientras que la

firma de la zona 2 no. Por otro lado, ya en la zonificación binaria realizada en DIC-1 se define la zona efectivamente sembrada de la chacra.

Al analizar la zonificación multi-zona, se observan varias cosas. Para empezar, hay buena coincidencia entre las zonas definidas en binaria y multi-zona. En esta última hay una zona que es muy coincidente con la zona definida como no agrícola en binaria. Con respecto a la cantidad de zonas definidas en cada corrida, esta es variable, estando entre 3 y 4 según la corrida, y en 5 para los últimos dos decadales de abril. También se observa que algunas zonas ocupan muy poca área con relación al área total de la chacra, como por ejemplo la zona 2 en DIC-3 de la figura 105. Por último, al observar los gráficos de firmas medianas, hay zonas con firmas muy similares que el algoritmo decidió poner en zonas separadas, cuando podrían formar parte de una misma zona, como se puede observar en el gráfico de la figura 110 para las zonas 2 y 3.

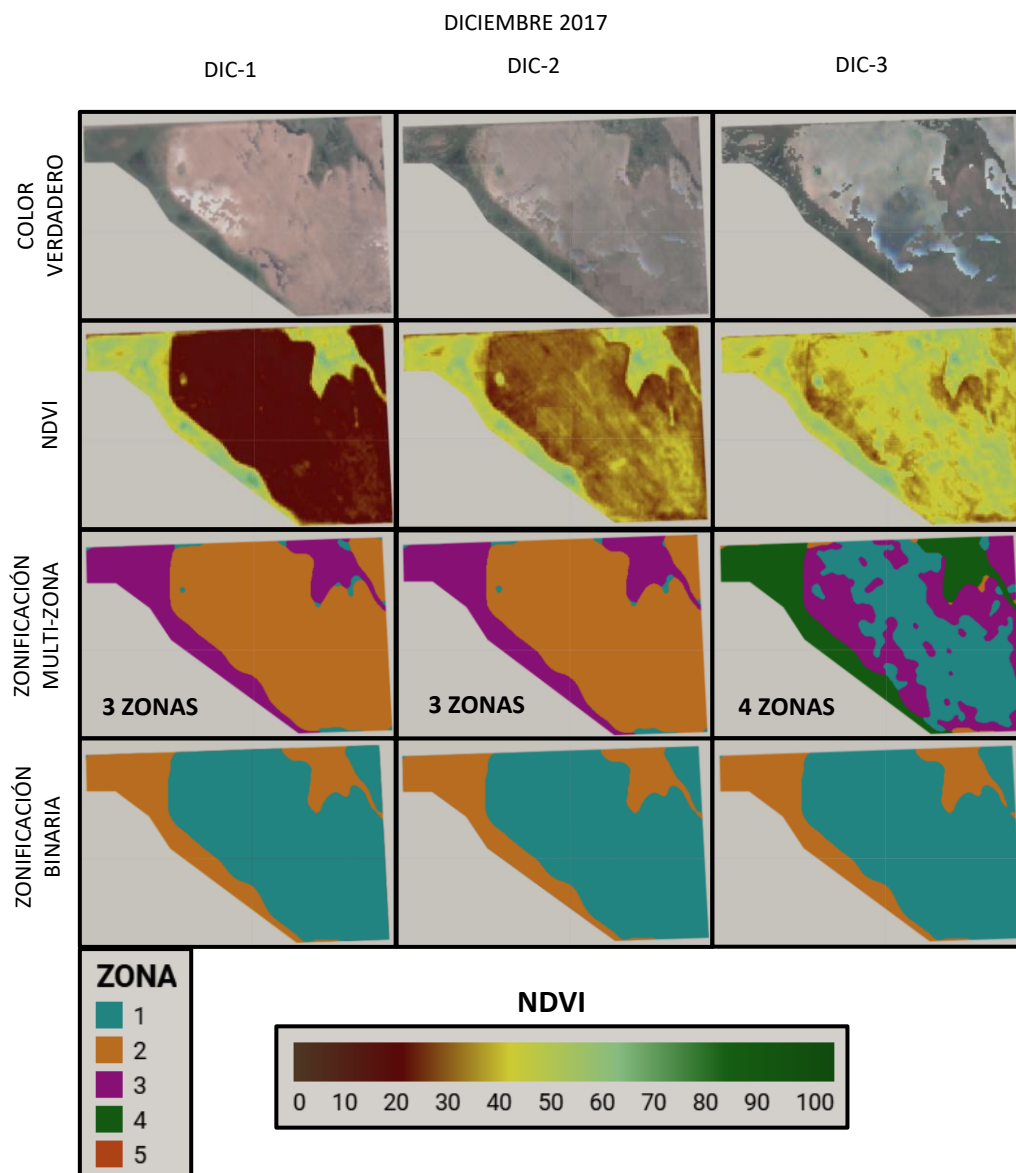


Figura 105. Mapas de zonificación de chacra de soja por decadal de diciembre de 2017.

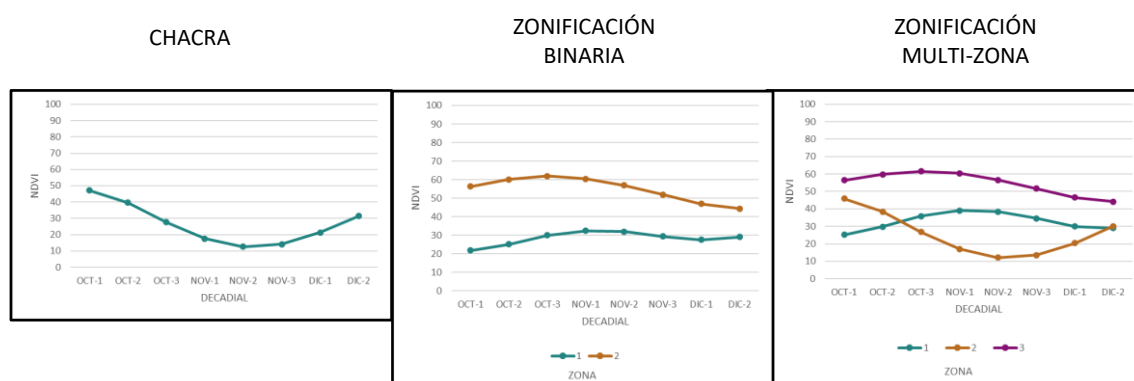


Figura 106. Gráficos de firmas fenológicas medianas de chacra de soja para OCT-1 a DIC-2 de 2017.

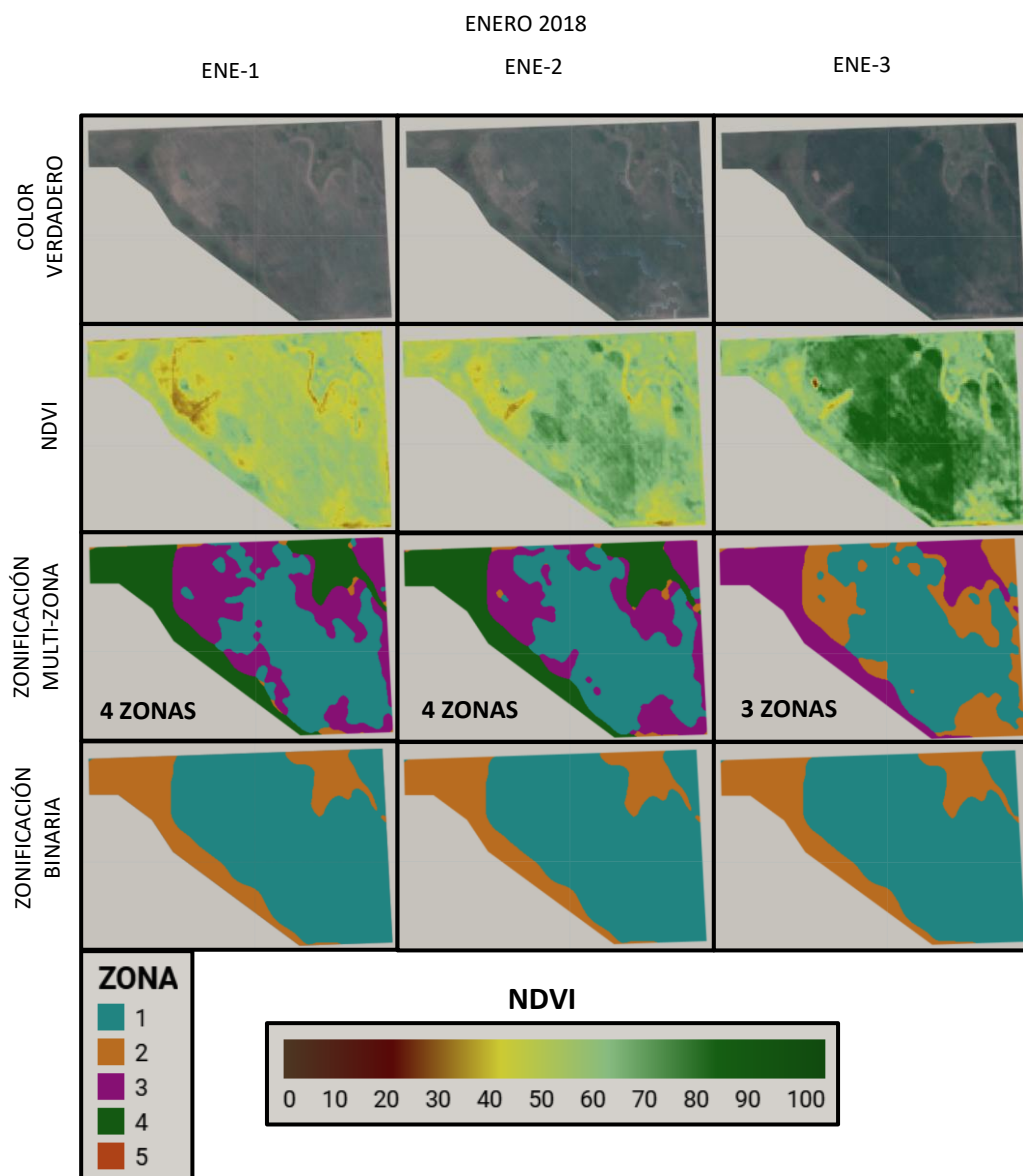


Figura 107. Mapas de zonificación de chacra de soja por decadal para enero de 2018.

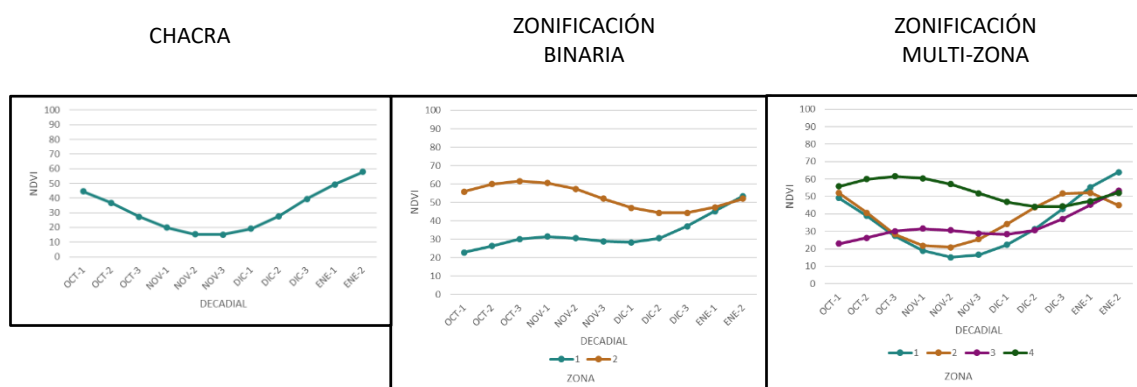


Figura 108. Gráficos de firmas fenológicas medianas de chacra de soja para OCT-1 a ENE-2 de 2018.

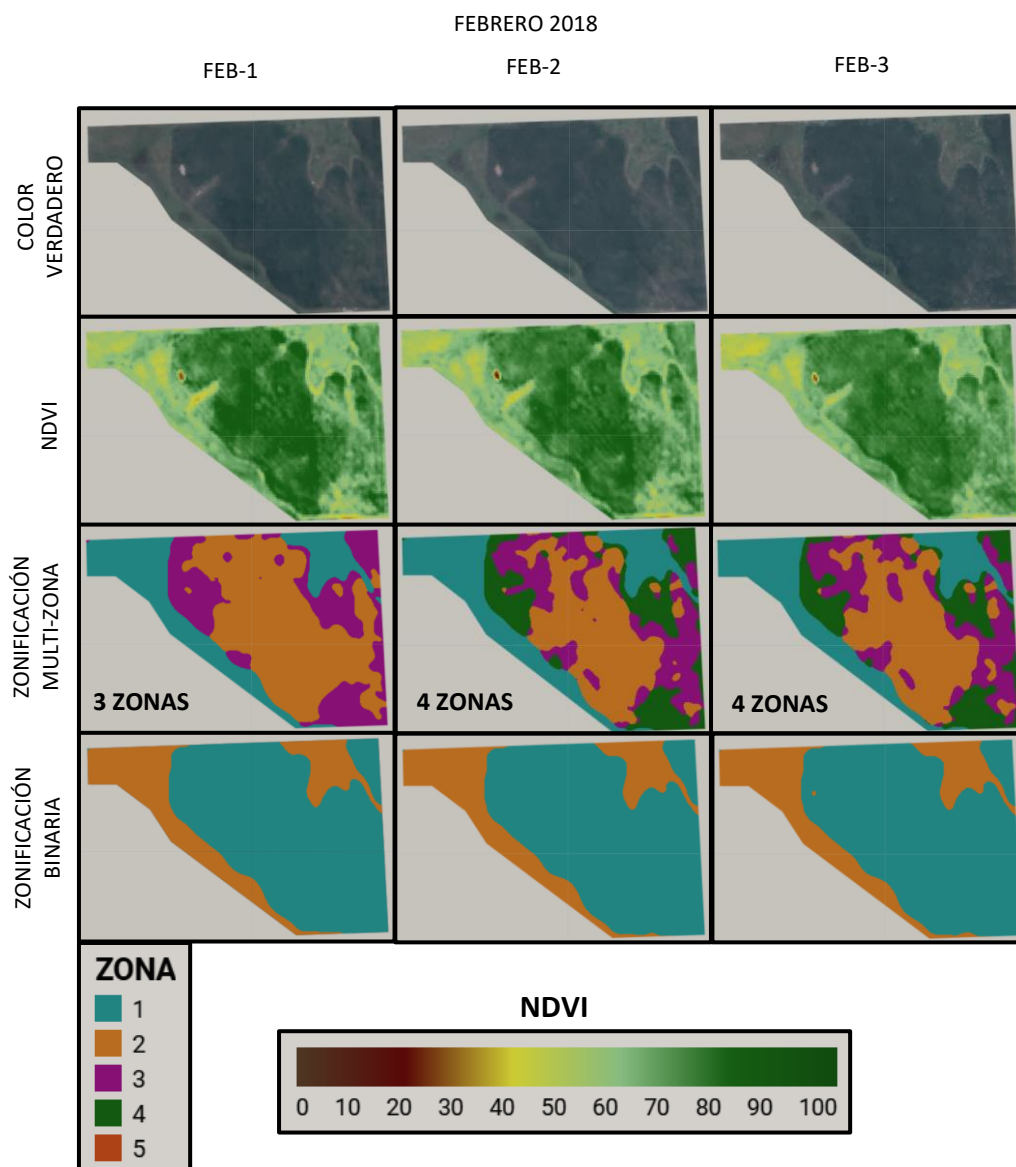


Figura 109. Mapas de zonificación de chacra de soja por decadal para febrero de 2018.

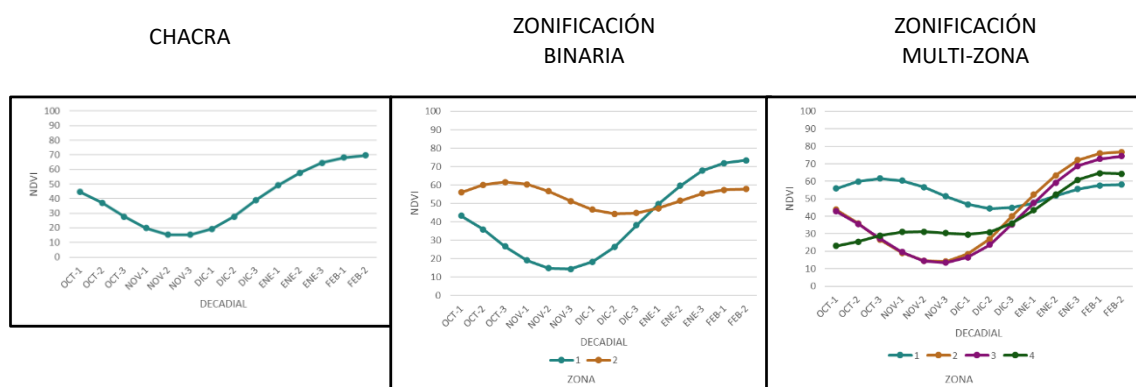


Figura 110. Gráficos de firmas fenológicas medianas de chacra de soja para OCT-1 a FEB-2 de 2018.

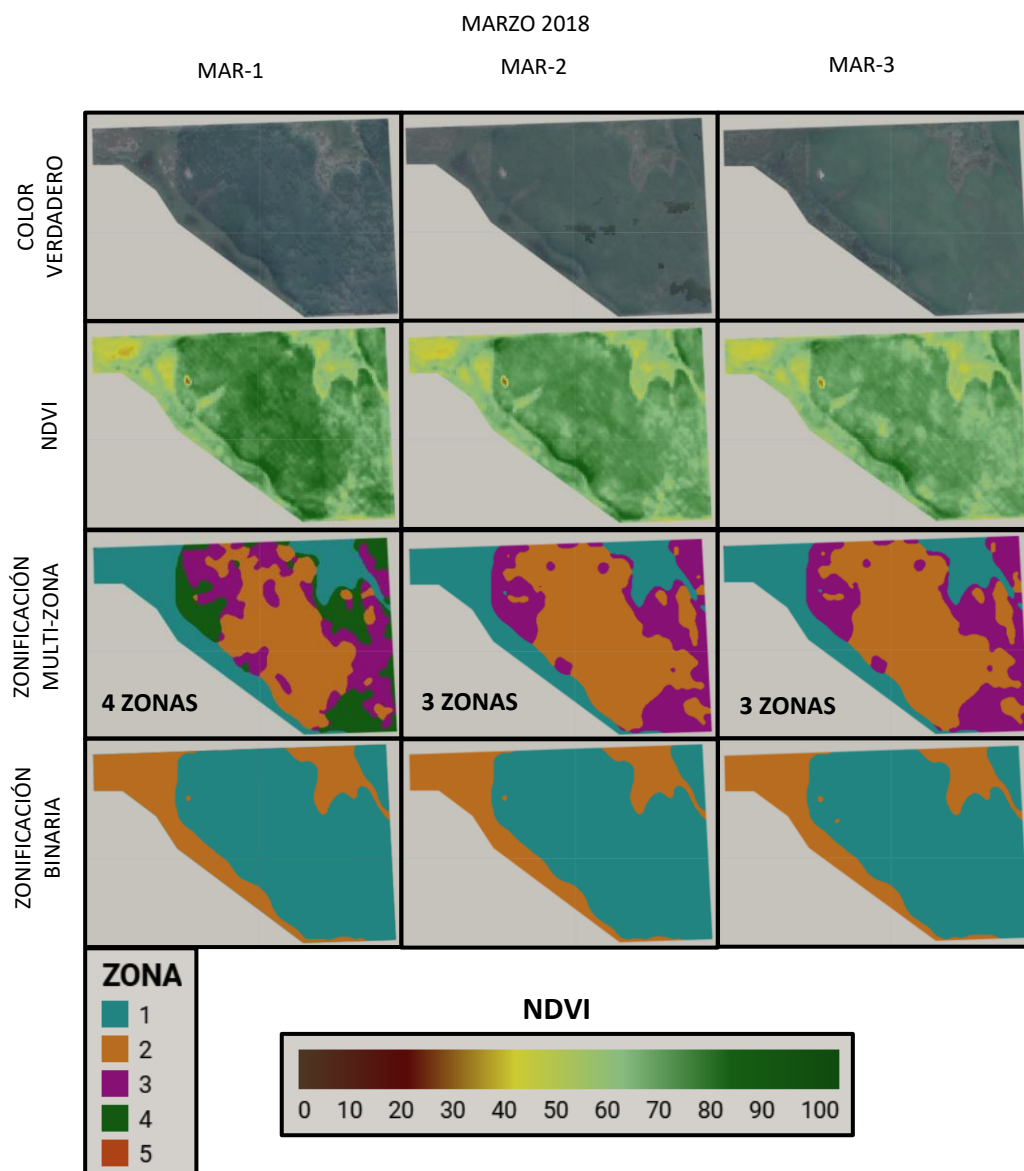


Figura 111. Mapas de zonificación de la chacra de soja por decadal para marzo de 2018.

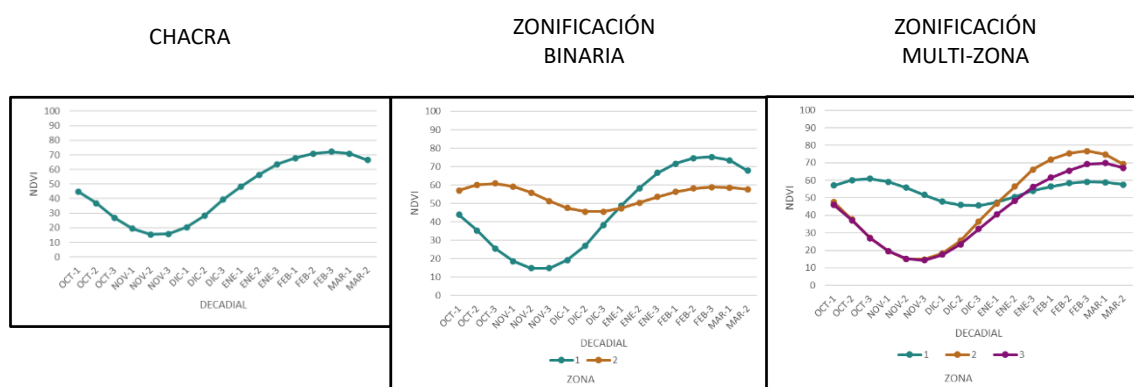


Figura 112. Gráficos de firmas fenológicas medianas de chacra de soja para OCT-1 a MAR-2 de 2018.

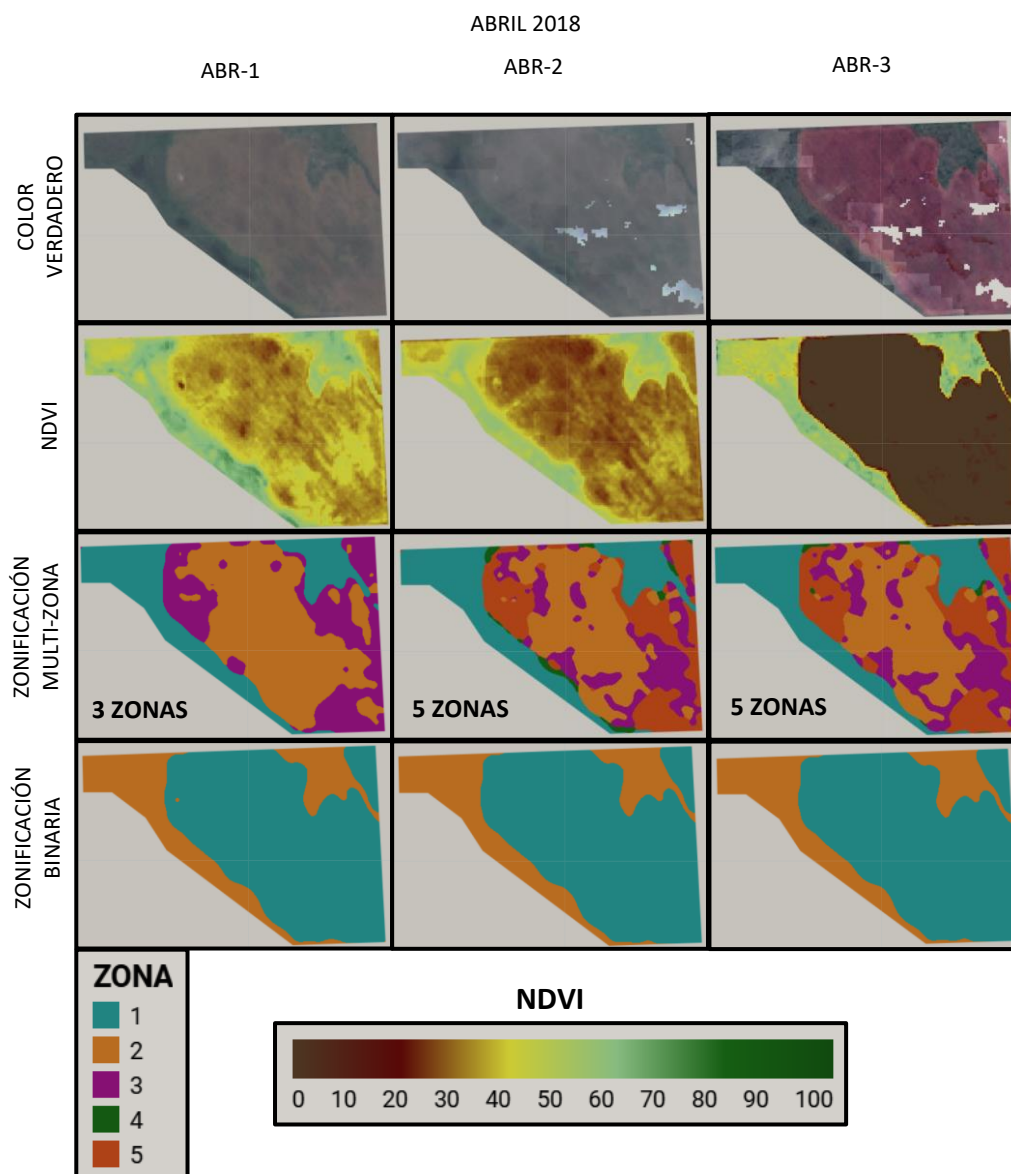


Figura 113. Mapas de zonificación de chacra de soja por decadal para abril de 2018.

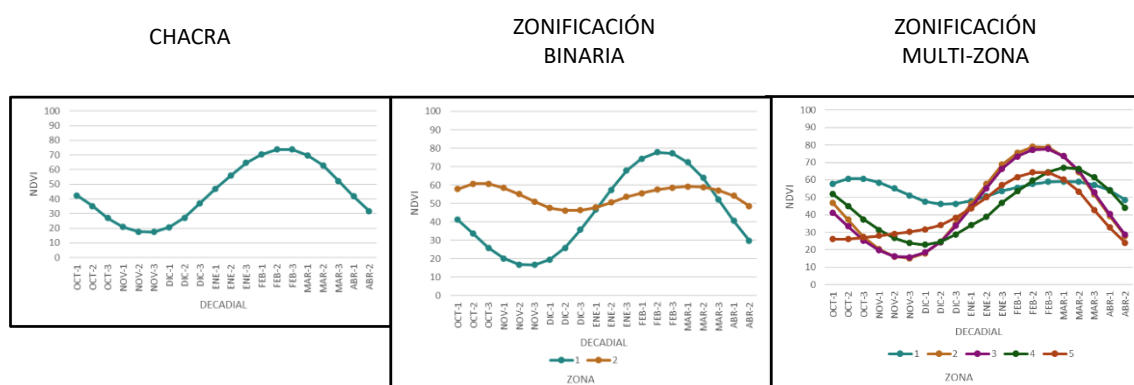


Figura 114. Gráficos de firmas fenológicas medianas de chacra de soja para OCT-1 a ABR-2 de 2018.

En la figura 115 se muestra un comparativo entre el mapa de zonificación binaria de la chacra de soja analizada y el mapa agrícola/no agrícola binario 3 que se generó en la sección 11.1. Como

se ve, binario 3 logra delimitar muy bien la zona agrícola de la chacra de soja, y tienen un gran nivel de coincidencia con el área delimitada por la zonificación binaria. Para este caso, esto demuestra que el modelo supervisado de clasificación binaria agrícola/no agrícola desarrollado a nivel nacional funciona muy bien en delimitar las chacras con agricultura.



Figura 115. Comparativa de zonificación binaria de la chacra de soja contra mapa agrícola/no agrícola binario 3.

12.2. Extracción de métricas fenológicas de la chacra

La fenología se define como el estudio de las etapas del ciclo de vida de plantas y animales según Lieth. La fenología vegetal se refiere a eventos específicos en el ciclo de vida, como por ejemplo: brotación, emergencia y senescencia foliar, floración y madurez en cultivos cerealeros. Con el uso de sensoramiento remoto muchos de estos eventos no se pueden detectar directamente a la resolución espacial de las imágenes satelitales, entonces descriptores más generales de la dinámica de la vegetación son calculados y se refiere a ellos como Fenología de la Superficie Terrestre (Land Surface Phenology (LSP)). De Beurs y Henebry definen LSP como el desarrollo espacio-temporal de la superficie con vegetación que se puede observar con sensores satelitales. Las métricas LSP están asociadas con cambios en la vegetación y comprenden el comienzo de la temporada, pico de la estación de crecimiento, senescencia o fin de la estación, largo de la estación y otros estadios de transición. A las métricas LSP también se las denomina simplemente métricas fenológicas (Zeng, Wardlow, Xiang, Hu, & Li, 2020).

La estimación de las métricas fenológicas a partir de series temporales de sensoramiento remoto comprenden tres etapas. La primera es la limpieza de los datos, la segunda es el suavizado de los datos y reconstrucción de las series temporales, y la tercera es extracción de las métricas a partir de las series temporales reconstruidas (Zeng et al., 2020). Algunos autores (Tan et al., 2011), a partir del análisis de la primera, segunda y tercera derivada de la serie reconstruida usando el software TIMESAT (Jönsson & Eklundh, 2004) extrajeron las métricas fenológicas de: comienzo del verdor, fin del verdor, comienzo de la senescencia, fin de la senescencia, máxima y mínima tasa de incremento, como se puede observar en la figura 116 para una serie temporal de EVI.

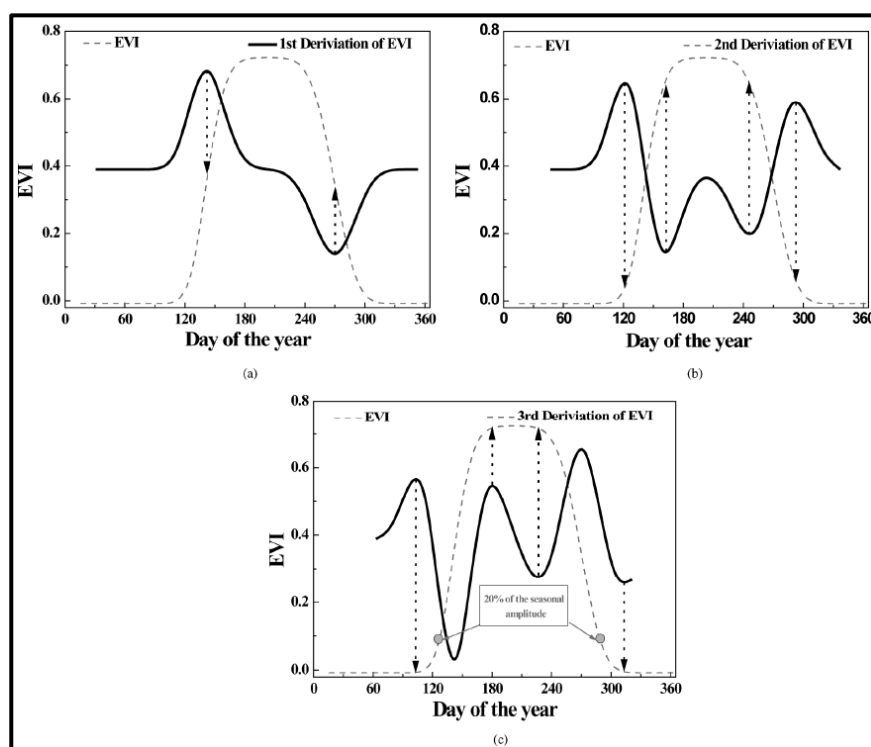


Figura 116. Ejemplo de métricas fenológicas extraídas con primera, segunda y tercera derivada de una serie temporal de EVI. Fuente: Tan et al. (2011).

En los últimos años en Estados Unidos se han venido realizando trabajos en maíz y soja para vincular las métricas fenológicas extraídas satelitalmente con las observaciones de los diferentes estadios fenológicos de los cultivos (emergencia, floración, llenado de grano, maduración, etc.) realizadas a campo, y poder desarrollar modelos de estimación de estos a partir de información satelital. Ejemplos de estos trabajos son los realizados por Bandaru (2020), Diao (2020) y Seo et al. (2019).

El último paso en el proceso de seguimiento de la evolución de la chacra durante la zafra fue estimar las métricas fenológicas de la firma mediana del área agrícola (zona 1) de la zonificación binaria realizada en la sección anterior. Como la firma fenológica mediana se construyó a partir de imágenes sintéticas decadales, se tiene un valor cada 10 días. Para poder extraer las métricas fenológicas fue necesario reconstruir la firma mediana a paso diario para poder realizar el análisis de derivadas. Para el proceso de reconstrucción de la firma mediana y la extracción de las métricas fenológicas se armó un script en R. Usando un modelo de suavizado conocido como Smoothing Splines, que está incorporado en R, fue reconstruida la firma mediana a paso diario. Una función spline es una curva construida a partir de segmentos polinomiales (cuadráticos o cúbicos) que están sujetos a condiciones de regularidad en sus uniones (Pollock, 1999). Una vez obtenida la firma mediana a paso diario, se calculó la derivada primera, y a partir del análisis de la serie reconstruida y la derivada se extrajeron las siguientes métricas fenológicas:

- Mínimo, máximo y punto de inflexión (PI).
- Fechas de mínimo, máximo y PI.
- NDVI de mínimo, máximo y PI.
- Pendiente del PI.
- Amplitud PI-mínimo: $NDVI_{PI} - NDVI_{min}$.
- Amplitud máximo-PI: $NDVI_{max} - NDVI_{PI}$.

- Amplitud máximo-mínimo: $NDVI_{max} - NDVI_{min}$.
- Días de mínimo a máximo.
- Días de mínimo a PI.
- Días de PI a máximo.

En la figura 116 se presenta el flujograma de todo el proceso de extracción de las métricas fenológicas. En la tabla 63 y figura 118 se presentan para la chacra los valores de las métricas extraídas y los gráficos de la firma reconstruida y la derivada primera.

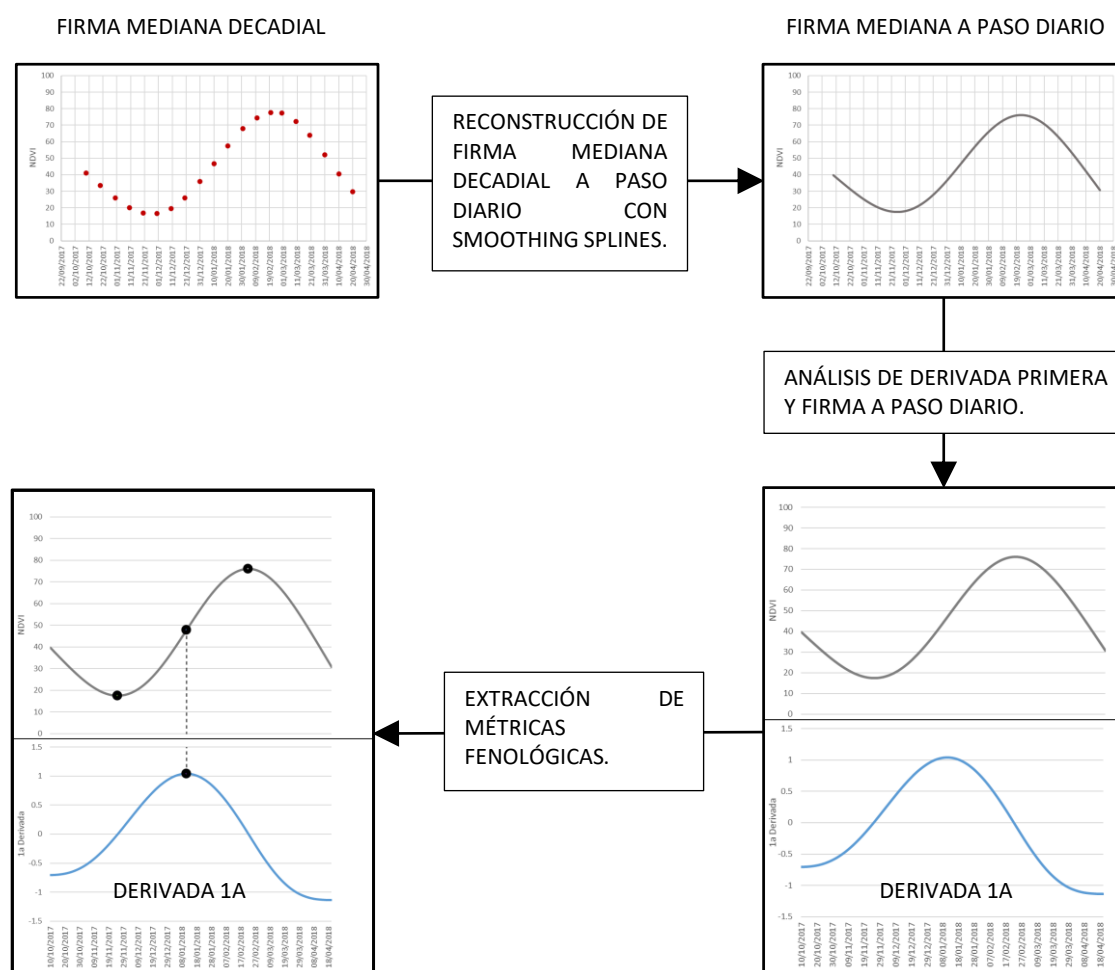


Figura 117. Flujograma de extracción de métricas fenológicas de la firma mediana de la chacra.

A partir del análisis de la tabla 63 y gráficos de la figura 118 se extrae información muy relevante. Para el ejemplo de la chacra analizada, se sabe que el mínimo ocurrió el 25/11/2017 con un NDVI de 17,5. Esta información revela que la siembra de la chacra estuvo en torno a esa fecha. El punto de inflexión, que es el punto de máxima tasa de incremento de NDVI ocurrió el 11/01/2018, 47 días después del mínimo, con un NDVI de 47,9 y una pendiente de 1,04. A partir de esta fecha, el NDVI fue incrementando su valor a tasas decrecientes hasta alcanzar el máximo. Este, lo que algunos autores llaman pico de temporada (Peak of Season) coincide con la etapa en la que los cultivos como maíz y soja ya están florecidos, ocurrió el 22/02/2018, unos 42 días después de PI, 89 días después del mínimo y con un NDVI de 76,1. De esta manera, tomando el mínimo como proxy de la fecha de siembra, la soja de la chacra analizada llegó a su máximo tres meses después de esta. A partir del máximo, el cultivo empieza a senescer, decreciendo su NDVI hasta llegar a valores de 30 el 20/04/2018. A su vez, también se calculó la amplitud, que es la

diferencia de NDVI entre los diferentes momentos claves, siendo de 30,4 puntos entre mínimo y PI, 28,3 entre PI y máximo, y de 58,7 puntos entre mínimo y máximo.

MÉTRICA FENOLÓGICA	FECHA	NDVI	PENDIENTE
MÍNIMO	25/11/2017	17,5	
PUNTO DE INFLEXIÓN (PI)	11/01/2018	47,9	1,04
MÁXIMO	22/02/2018	76,1	
	DIAS	AMPLITUD NDVI	
MÍNIMO A PI	47	30,4	
PI A MÁXIMO	42	28,3	
MÍNIMO A MÁXIMO	89	58,7	

Tabla 63. Métricas fenológicas de la firma mediana de la chacra.

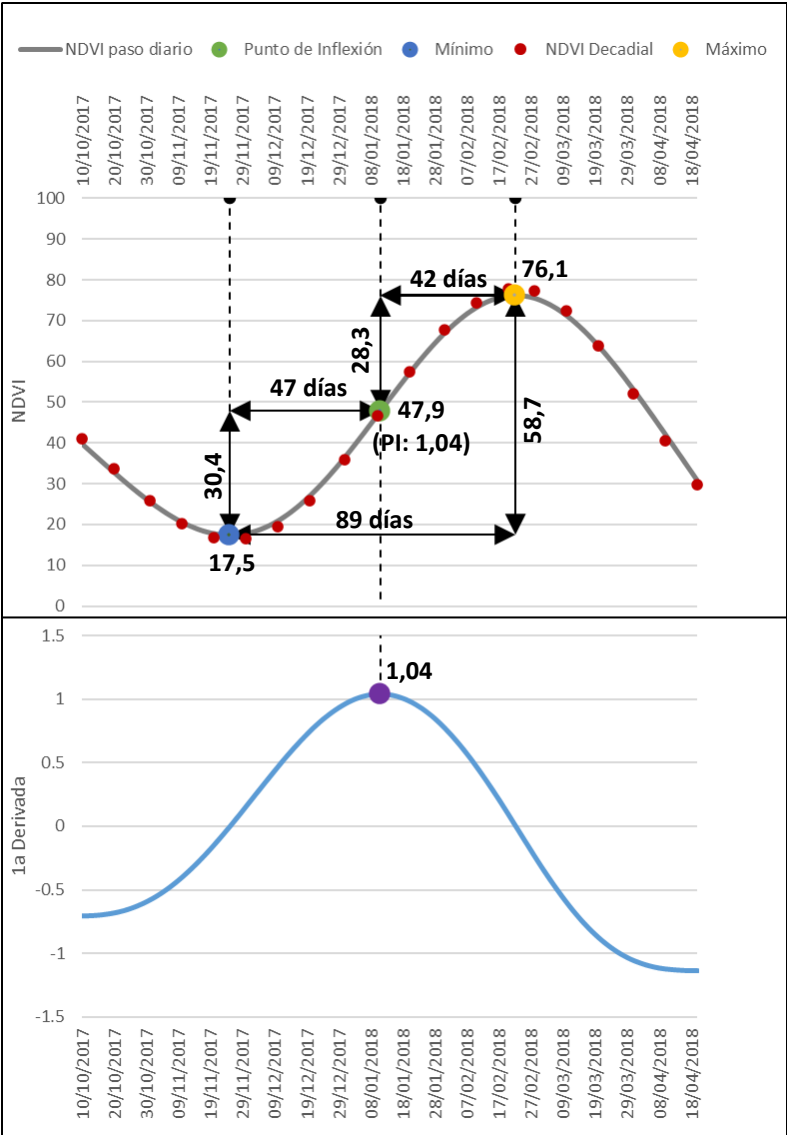


Figura 118. Gráfico de métricas fenológicas de la firma mediana de la chacra.

Es así como con el uso de las métricas fenológicas podemos caracterizar como fue la evolución del desarrollo y crecimiento de un cultivo durante la zafra. Incluso, se pueden empezar a calcular durante la zafra, para por ejemplo identificar el punto de inflexión que indica el momento a partir del cual el cultivo empieza a incrementar su NDVI a tasas decrecientes y tomar medidas

de manejo agronómico. También a través del uso de las métricas fenológicas se pueden comparar diferentes chacras de un mismo cultivo sembradas en lugares diferentes para una misma zafra, como se puede ver en el ejemplo del gráfico de la figura 119, y a través del análisis tratar de entender y explicar por qué se comportaron diferente.

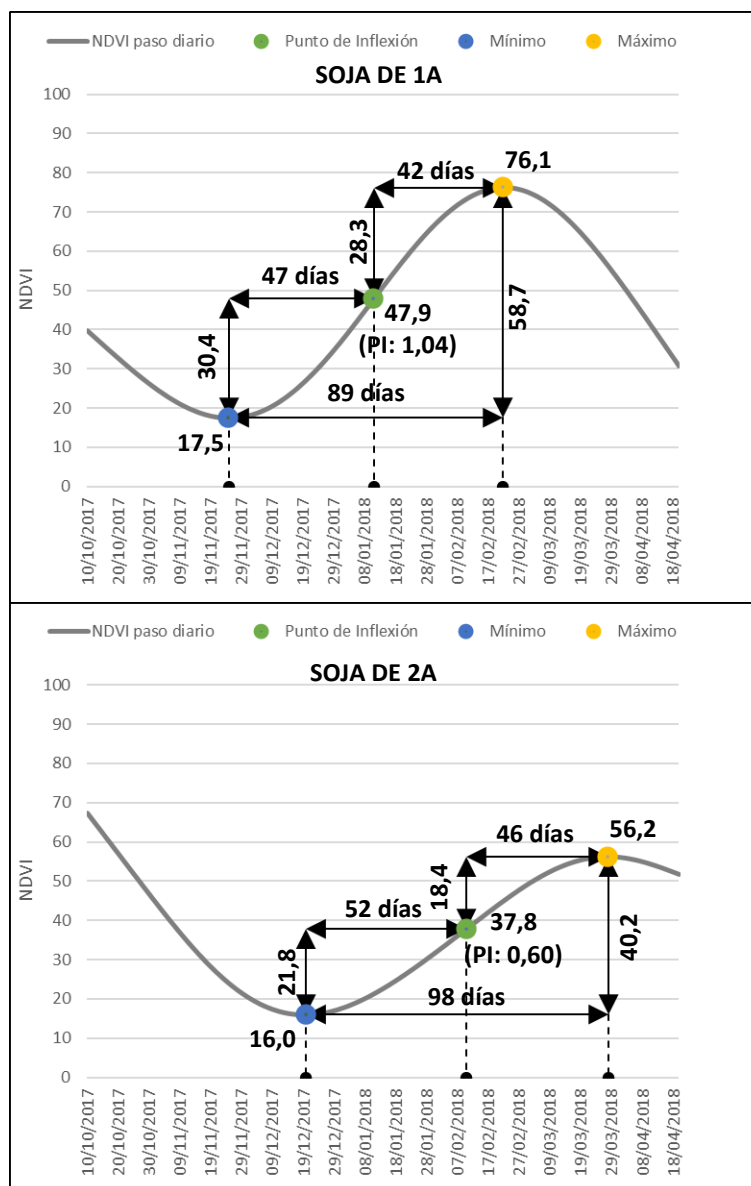


Figura 119. Gráfico comparativo de métricas fenológicas de dos chacras de soja de FUCREA-URF para la zafra 2017-18.

En la figura 119 y tabla 64 se presentan las firmas medianas y métricas fenológicas de una chacra de soja de primera y otra de soja de segunda, ambas sembradas en la zafra 2017-18. De los gráficos y tablas, se observa que hay casi un mes de diferencia entre el mínimo de una y otra chacra, siendo la soja 1ª del 25/11/2017 y del 20/12/2017 la soja 2ª. Los valores mínimos de NDVI son similares entre ambas chacras, con valores en torno a 17. Con respecto al momento de máxima tasa de incremento de NDVI (PI), aquí hay una diferencia de un mes entre la soja 1ª y la 2ª, ocurriendo el 11/01/2018 y el 10/02/2018, con NDVI de 47,9 y 37,8, y valores de pendiente de 1,04 y 0,60, respectivamente. Aquí se nota claramente que la soja 1ª incrementa su NDVI con una tasa de incremento 0,44 ($1,04 - 0,60$) más que la soja 2ª, y esto se ve reflejado en 10,1 ($47,9 - 37,8$) puntos más de NDVI en PI de la soja 1ª. Con respecto al máximo, de nuevo

hay poco más de un mes de diferencia entre las chacras, ocurriendo el 22/02/2018 en 1ª y el 28/03/2018 en 2ª. A su vez, soja 1ª alcanza un máximo de 76,1 de NDVI contra 56,2 de 2ª, unos 20 puntos de diferencia. Por último, la cantidad de días necesarios desde el mínimo para alcanzar el máximo es de 89 en soja 1ª, y unos 9 días (98) más en soja 2ª. De esta manera, las métricas fenológicas permiten ver la evolución de una chacra durante la zafra, pero también compararla contra otras, para evaluar su desempeño.

MÉTRICA	SOJA 1A	SOJA 2A
MÍNIMO	25/11/2017	20/12/2017
PI	11/01/2018	10/02/2018
MÁXIMO	22/02/2018	28/03/2018
	NDVI	
MÍNIMO	17,5	16,0
PI	47,9	37,8
MÁXIMO	76,1	56,2
	PENDIENTE	
PUNTO DE INFLEXIÓN	1,04	0,60
	AMPLITUD NDVI	
MÍNIMO A MÁXIMO	58,7	40,2
MÍNIMO A PI	30,4	21,8
PI A MÁXIMO	28,3	18,4
	DIAS	
MÍNIMO A MÁXIMO	47	52
MÍNIMO A PI	42	46
PI A MÁXIMO	89	98

Tabla 64. Comparativo de métricas fenológicas de dos chacras de soja de FUCREA-URF para la zafra 2017-18.

13. Conclusiones

Este trabajo intentó cumplir con los siguientes objetivos específicos:

1. Identificar y mapear los cultivos agrícolas de secano de verano durante la zafra en curso a partir de imágenes satelitales.
2. Realizar el seguimiento de la condición de los cultivos agrícolas en diferentes momentos de desarrollo y crecimiento entre siembra y cosecha.
3. Identificar la especie del cultivo presente en la chacra (maíz, soja, sorgo, etc.).

Para ello se accedió e integró información de diferentes bases de datos de cultivos con imágenes satelitales. De las bases, dos fueron proporcionadas por organizaciones de productores agropecuarios (FUCREA y URF) y son de chacras con agricultura (maíz, soja y sorgo). Otras dos son del Ministerio de Ganadería, Agricultura y Pesca de Uruguay (MGAP). Una con información de coberturas y usos del suelo, y otra de chacras con planes de uso para conservación del suelo por erosión. Las imágenes satelitales utilizadas fueron de los satélites ópticos Sentinel 2 y Terra, y del de tipo radar SAR, Sentinel 1.

Previo al desarrollo de los modelos de clasificación, sobre todo cuando se trabajó con los datos de Planes de Uso, fue necesario desarrollar una metodología para identificar datos mal etiquetados, ya que como se mencionó, los datos de Planes no son “verdad de terreno”, sino declaraciones de siembra a futuro por parte del productor. Los cultivos agrícolas tienen un patrón temporal de NDVI que se asemeja a la forma de una función Gaussiana, mientras que otros tipos de coberturas del suelo que no son agrícolas, no. Este patrón diferente se puede identificar combinando técnicas de clasificación no supervisada con técnicas de reducción de la dimensionalidad para visualización de datos, como t-SNE. De esta manera, esta metodología permite, de manera no supervisada, discriminar cultivos agrícolas de los que no lo son, y depurar datos para entrenamiento.

En cuanto al primer objetivo, se logró desarrollar modelos supervisados de clasificación binaria que identifican si la cobertura presente es agrícola o no. Estos modelos lograron desempeños por encima de 92 de *F1-Score*. Además, en el caso del modelo con mapeo pixel a pixel desarrollado en GEE, éste tiene una muy buena capacidad de segmentar los píxeles que son agrícolas, logrando identificar y delimitar bien las chacras agrícolas, con muy poco efecto de “sal y pimienta”. En cuanto al área determinada por este modelo, sobrestima el área de referencia de DIEA en un 24%. El área de referencia es un valor estimado a través de encuestas del cual no se conoce el margen de error. De todas maneras, el valor de área determinada por el modelo binario es razonable estando en torno a 1.700.000 hectáreas y la zona donde se ubican las chacras se corresponde con las áreas donde predomina la actividad agrícola.

El objetivo número 3 está ligado al número 1, ya que una vez identificadas las chacras con agricultura por modelo binario, se puede predecir el tipo de cultivo específico en estas. Para este objetivo los resultados obtenidos fueron dispares. Hay que tener en cuenta que las bases con información de cultivos como las de FUCREA, URF y PLANES, están muy desbalanceadas, siendo el cultivo dominante soja, lo cual tiene sentido, ya que es el más sembrado. Esto fue un problema, porque hay poca representatividad de los cultivos minoritarios maíz y sorgo para entrenar. De estos cultivos, soja fue el de mejor desempeño, con *F1-Score* de 98,5, seguido a una distancia considerable por sorgo con 77,5 y maíz con 75,3. Estos valores se compararon con promedios de referencia de artículos internacionales de 84,4, 88,4 y 67,3 para maíz, soja y sorgo

respectivamente. Soja y Sorgo estuvieron por encima de esos valores, mientras que maíz estuvo por debajo. Se considera que si la cantidad de datos de entrenamiento de maíz y sorgo se incrementara, podría mejorarse el desempeño de estos cultivos.

Para el objetivo número 2 se logró desarrollar una metodología para el seguimiento de los cultivos. Aquí, aprovechando los scripts desarrollados en GEE para extraer las series temporales de NDVI a partir de Sentinel 2, se adaptaron para hacer el seguimiento de la chacra para períodos decadales (10 días) desde la siembra. Haciendo uso de clasificación no supervisada con K-Means de los apilados de NDVI, se desarrolló una metodología que zonifica las chacras en zonas agrícolas y no agrícolas, y permite ver su evolución durante la zafra a partir de la firma mediana de NDVI. También a partir del análisis de las firmas medianas a paso diario y con análisis de derivadas, se pudo extraer las métricas fenológicas de mínimo, punto de inflexión y máximo, las cuales permiten caracterizar diferentes chacras dentro de una zafra y hacer comparaciones de evolución entre éstas.

A pesar de todo lo realizado, quedaron actividades pendientes y que pueden ser trabajos a futuro. A continuación se mencionan las que se consideran más relevantes.

En este trabajo, además de usar algoritmos de aprendizaje automático clásicos, como Random Forest y XGBoost, quedó pendiente para la clasificación de cultivos probar diferentes arquitecturas de redes neuronales, como pueden ser del tipo convolucionales, recurrentes, U-Net, Deep learning y otras existentes.

Con respecto a los índices satelitales se evaluaron EVI/EVI2 y NDVI, pero también existe la posibilidad de aprovechar más las distintas bandas del espectro electromagnético que observa Sentinel 2, y haberse calculado otros índices como Inverted Red-Edge Chlorophyll Index (IRECI), Red-Edge Position Index (REPO), los cuales aprovechan las bandas del rojo de borde (Red-Edge), que muy pocos satélites miden. También otros, como Normalized Difference Infrared Index (NDII), que combina la banda del infrarrojo de onda corta con una del rojo de borde, por citar otros ejemplos.

Las imágenes utilizadas del tipo óptico fueron las de Sentinel 2, pero también existen otros satélites ópticos como Landsat 8, y desde setiembre de 2021, Landsat 9. Estos últimos tienen una resolución espacial mayor y miden menos bandas que Sentinel 2. De todas maneras, existen técnicas que permiten fusionar y armonizar datos de diferentes satélites y aumentar así la cantidad de datos temporales disponibles a usar.

En cuanto al trabajo con métricas fenológicas para el seguimiento de cultivos, hay que explorar cruzar los datos satelitales con observaciones de estados fenológicos a campo para maíz, soja y sorgo, para desarrollar modelos locales de predicción de estado fenológico a partir de sensoramiento remoto, los cuales pueden servir de insumo para desarrollar otros modelos, como estimación de rendimiento.

14. Agradecimientos

Al Ing. Agr. Martin Dell'Acqua de la Dirección General de Recurso Naturales (DGRN) del Ministerio de Ganadería, Agricultura y Pesca (MGAP) por facilitar el acceso a la información de los Planes de Uso y Manejo Responsable del Suelo de Uruguay. Al Ing. Agr. Joaquín Echeverría por el acceso a la base de cultivos de la Federación Uruguay de Grupos CREA (FUCREA). A la Unión Rural de Flores (URF) y Verónica Hugo Ettlin de San Diego SoftWorks por el acceso a la base de cultivos LUGUS. A INIA y especialmente al Ing. Agr. MSc Agustín Giménez y la Lic. Biol. MSc PhD Guadalupe Tiscornia por el apoyo y el tiempo facilitado para realizar este trabajo. Al Dr. Pablo Musé y Dr. Ing. Javier Preciozzi por aceptar este desafío, el conocimiento y el apoyo brindado para realizar este trabajo.

Por último, agradezco especialmente a mi familia, Magdalena y nuestra hija Victoria, por el apoyo y las incontables horas que me cedieron para realizar este trabajo.

15. Índice de figuras y tablas

Figura 1. Mapa de Regiones Agropecuarias definidas según el último Censo General Agropecuario de 2011.....	10
Figura 2. Mapa de distribución espacial de PLANES para el período 2013 a 2019.....	11
Figura 3. Ejemplo de registros del archivo DBF de la base PLANES.	12
Figura 4. Ejemplo de un plan con varias chacras en formato Shapefile.	13
Figura 5. Gráfico de porcentaje de planes según cultivo para el período 2013-2019.	14
Figura 6. Gráfico de porcentaje de planes y superficie según cultivo para la zafra 2018-19.....	15
Figura 7. Ejemplo de polígonos de unas chacras de FUCREA-URF en formato Shapefile.....	15
Figura 8. Mapa de chacras de FUCREA y URF para las zafras 2017-18 y 2018-19.	16
Figura 9. Gráfico de porcentaje de chacras y superficie según cultivo de FUCREA-URF de las zafras 2017-18 y 2018-19.	17
Figura 10. Mapa Integrado de Cobertura/Uso del Suelo del Uruguay Año 2018.	18
Figura 11. Satélites observando la Tierra.....	20
Figura 12. Radiación Solar medida remotamente por satélite.	21
Figura 13. Espectro de radiación electromagnética.....	21
Figura 14. Grilla de imagen satelital.....	22
Figura 15. Imagen satelital de Sentinel 2.	22
Figura 16. Comparación de tamaño de pixel para satélites Landsat 8, Sentinel 2, SPOT y Terra.	23
Figura 17. Mapa de pasadas de los satélites Sentinel 2 sobre Uruguay.	24
Figura 18. Mapa de pasadas de los satélites Sentinel 1 sobre Uruguay.	25
Figura 19. Mapa de pasadas del satélite Terra sobre Sudamérica.	25
Figura 20. Gráfico de reflectancia de las plantas en el espectro electromagnético.	27
Figura 21. Ejemplo de registros con y sin etiqueta de cultivo asignada de PLANES.	29
Figura 22. Flujograma de extracción de series temporales según Sentinel o Terra para PLANES.	31
Figura 23. Puntos regulares a diferentes densidades según cultivo para extraer series temporales de Sentinel para PLANES.	32
Figura 24. Chacras con grilla de puntos regulares de FUCREA-URF.....	33
Figura 25. Mapa de distribución de puntos no agrícola seleccionados al azar de COBERTURAS.	35
Figura 26. Apilado de imágenes MOD13Q1 para zafra 2013-14.....	37
Figura 27. Extracción de pixeles Terra - MOD13Q1 completamente dentro de las chacras de un plan.....	37
Figura 28. Gráfico de series temporales individuales (gris) y promedio (negro) de EVI de un plan para zafra 2013-14.	38
Figura 29. Flujograma de extracción de series temporales de EVI en GEE para PLANES.....	39
Figura 30. Flujograma de generación de imágenes sintéticas Sentinel 1 en GEE.	42
Figura 31. Flujograma de generación de imágenes sintéticas Sentinel 2 en GEE.	43
Figura 32. Proceso de generación de apilado de imágenes sintéticas de Sentinel (1 y 2).....	44
Figura 33. Flujograma de generación de apilados y extracción de series temporales por banda e índice en GEE de Sentinel (1 y 2).....	45
Figura 34. Flujograma de imputación de datos faltantes en series temporales.	46
Figura 35. Gráfico de serie temporal de NDVI procesada para imputar datos faltantes.....	47
Figura 36. Mapa de regionalización del Uruguay para PLANES.	49

Figura 37. Gráfico de firmas fenológicas promedio de NDVI de diferentes coberturas del suelo en Uruguay (CV: cultivo de verano, DC: doble cultivo, F y M: forestación y monte, RFP: recursos forrajeros perennes).	50
Figura 38. Gráfico de firmas fenológicas medianas de EVI2 de diferentes coberturas y usos del suelo de COBERTURAS para la zafra 2017-18.	51
Figura 39. Gráfico de evolución de NDVI de diferentes cultivos de verano (girasol, maíz y soja) en la Provincia de Buenos Aires.	51
Figura 40. Función Gaussiana.....	52
Figura 41. Gráfico de evolución de NDVI para un cultivo agrícola de secano y una pastura de la base PLANES.....	52
Figura 42. Gráfico de un grupo del subconjunto región Suroeste, soja y zafra 2018-19 de PLANES.	53
Figura 43. Gráfico de t-SNE del subconjunto de región Suroeste, soja y zafra 2016-17 de PLANES.	54
Figura 44. Gráfico de firmas medianas del subconjunto de región Suroeste, soja y zafra 2016-17.	55
Figura 45. Gráficos de t-SNE para K-means y SOM+K-means del subconjunto Centrosur, soja y zafra 2016-17.	57
Figura 46. Gráficos de t-SNE para Spectral clustering y GMM del subconjunto Centrosur, soja y zafra 2016-17.	58
Figura 47. Gráfico de selección de firmas individuales entre percentiles 5 y 95 para un grupo de subconjunto de región Suroeste, soja y zafra 2018-19.....	59
Figura 48. Gráfico de Precision de los algoritmos de clasificación supervisada evaluados en PLANES.	64
Figura 49. Gráfico de Recall de los algoritmos de clasificación evaluados en PLANES.	65
Figura 50. Gráfico de F1-Score de los algoritmos de clasificación supervisada evaluados en PLANES.	65
Figura 51. Ejemplo de chacras de FUCREA-URF con algunos puntos regulares en zonas no agrícolas. 1: Monte de abrigo y sombra, 2: zonas bajas, 3: área con construcciones y caminos, 4: curso de agua.	67
Figura 52. Series temporales de NDVI de una chacra con puntos agrícolas y no agrícolas.	68
Figura 53. Flujograma de clasificación no supervisada de puntos regulares de chacras de FUCREA-URF.....	69
Figura 54. Gráficos de firmas medianas y t-SNE para subconjunto de maíz de zafra 2017-18... ..	70
Figura 55. Gráfico de depuración de firmas individuales de PLANES entre percentiles 5 y 95 para un grupo de subconjunto de región Suroeste, soja y zafra 2018-19.	71
Figura 56. Gráfico de percentiles 25, 50, 75 y límites superior e inferior en un gráfico de un grupo.	72
Figura 57. Gráficos de un grupo para el subconjunto de maíz de zafra 2017-18 con las líneas de los percentiles (25, 50 y 75) y límites superior e inferior.....	72
Figura 58. Matriz multidimensional o tensor.....	74
Figura 59. Unión de tablas de series temporales por banda e índice en un array multibanda. .	74
Figura 60. Gráfico de F1-Score en el conjunto de evaluación de diferentes combinaciones de bandas e índices del array multibanda.	77
Figura 61. Gráfico de F1-Score según la fecha de cierre del último período utilizado.	78
Figura 62. Gráfico de F1-Score según cultivo y algoritmo supervisado evaluado.	79
Figura 63. Gráfico de F1-Score por cultivo según cantidad de períodos usados para entrenar el algoritmo.	80

Figura 64. Flujograma de división de chacras para entrenamiento y evaluación de FUCREA-URF.	82
Figura 65. Clasificación de una chacra por voto mayoritario de FUCREA-URF.	82
Figura 66. Gráfico de F1-Score por cultivo en evaluación para las 50 corridas de FUCREA-URF.	83
Figura 67. Gráfico de F1-Score por cultivo en evaluación para las 50 corridas de PLANES.	87
Figura 68. Gráfico de métricas de desempeño en FUCREA-URF y PLANES.	89
Figura 69. Flujograma de entrenamiento de XGBoost con base de FUCREA-URF y evaluación en PLANES.	90
Figura 70. Gráfico de métricas de desempeño para PLANES con modelo entrenado en FUCREA-URF con y sin balanceo.	91
Figura 71. Gráfico de F1-Score por cultivo de las bases de FUCREA-URF, PLANES y artículos revisados.	92
Figura 72. Flujograma de generación de modelo de clasificación binaria agrícola/no-agrícola.	94
Figura 73. Flujograma de creación de la base conjunta COBERTURAS/FUCREA-URF de puntos agrícola/no-agrícola de la zafra 2017-18.	96
Figura 74. Mapa de distribución de puntos de entrenamiento y evaluación de la base conjunta COBERTURAS/FUCREA-URF de puntos agrícola/no-agrícola de la zafra 2017-18.	96
Figura 75. Imputación de datos faltantes en serie temporal de NDVI con modelo armónico.	97
Figura 76. Flujograma de clasificación supervisada binaria agrícola/no-agrícola pixel a pixel en GEE.	98
Figura 77. Mapas binarios de EVI2 y NDVI.	99
Figura 78. Mapa binario VH_VV con bandeados (elipses rojas).	99
Figura 79. Ejemplo de píxeles clasificados como agrícolas por binario EVI2 o NDVI, y fusión de ambos.	100
Figura 80. Mapa de binario 2 superpuesto sobre regiones agropecuarias según censo del año 2011.	101
Figura 81. Ejemplo de píxeles espurios clasificados como agrícolas.	101
Figura 82. Filtro de moda aplicado sobre una imagen.	102
Figura 83. Ejemplo de filtro de moda con ventana de 3x3 y 7x7.	102
Figura 84. Efecto del filtro de moda 7x7, antes y después (a y b).	103
Figura 85. Flujograma de estrategias de filtrado de píxeles espurios.	104
Figura 86. Mapa clasificado con K-means para el apilado EVI2 enmascarado.	105
Figura 87. Firmas individuales y medianas de los grupos 1 a 3 para apilado EVI2 enmascarado clasificado con K-means.	106
Figura 88. Firmas individuales y medianas de los grupos 4 y 5 para apilado EVI2 enmascarado clasificado con K-means.	107
Figura 89. Firmas medianas de los grupos y gráfico t-SNE para apilado EVI2 enmascarado clasificado con K-means.	108
Figura 90. Comparación de los mapas binarios 2 a 5 en un área reducida.	109
Figura 91. Comparación de delimitación de chacras en mapa binario 2 contra polígonos de chacras de FUCREA-URF y PLANES.	110
Figura 92. Mapa de distribución de puntos de evaluación de base conjunta COBERTURAS/PLANES de zafra 2017-18.	112
Figura 93. Gráficos de NDVI para 50 puntos aleatorios agrícolas (a) y 50 no-agrícolas (na) del conjunto de evaluación COBERTURAS/PLANES.	113
Figura 94. Mapa de chacras identificadas en binario 2 dentro de la zona de influencia de FUCREA-URF.	114

Figura 95. Mapa de distribución de puntos de entrenamiento y evaluación de cultivos en la zona de influencia de FUCREA-URF.	115
Figura 96. Flujograma de clasificación supervisada por cultivo para chacras de Binario 2 en zona de influencia FUCREA-URF.	116
Figura 97. Ejemplo de chacras de FUCREA-URF clasificadas con GTB y RF, con y sin filtro de moda.	117
Figura 98. Gráfico de NDVI para zonas agrícola y no-agrícola de una misma chacra.	117
Figura 99. Ejemplo de chacras en donde se identifican zonas con 2 cultivos diferentes.	118
Figura 100. Gráfico de métricas de desempeño por cultivo para GTB-MODA.	119
Figura 101. Gráfico de métricas de desempeño por cultivo para RF-MODA.	120
Figura 102. Ventana temporal para el seguimiento de cultivos agrícolas de verano durante la zafra.	121
Figura 103. Flujograma de generación de mapas de zonificación de una chacra.	123
Figura 104. Flujograma de generación de firmas fenológicas medianas de una chacra.	124
Figura 105. Mapas de zonificación de chacra de soja por decadal de diciembre de 2017.	126
Figura 106. Gráficos de firmas fenológicas medianas de chacra de soja para OCT-1 a DIC-2 de 2017.	126
Figura 107. Mapas de zonificación de chacra de soja por decadal para enero de 2018.	127
Figura 108. Gráficos de firmas fenológicas medianas de chacra de soja para OCT-1 a ENE-2 de 2018.	127
Figura 109. Mapas de zonificación de chacra de soja por decadal para febrero de 2018.	128
Figura 110. Gráficos de firmas fenológicas medianas de chacra de soja para OCT-1 a FEB-2 de 2018.	128
Figura 111. Mapas de zonificación de la chacra de soja por decadal para marzo de 2018.	129
Figura 112. Gráficos de firmas fenológicas medianas de chacra de soja para OCT-1 a MAR-2 de 2018.	129
Figura 113. Mapas de zonificación de chacra de soja por decadal para abril de 2018.	130
Figura 114. Gráficos de firmas fenológicas medianas de chacra de soja para OCT-1 a ABR-2 de 2018.	130
Figura 115. Comparativa de zonificación binaria de la chacra de soja contra mapa agrícola/no agrícola binario 3.	131
Figura 116. Ejemplo de métricas fenológicas extraídas con primera, segunda y tercera derivada de una serie temporal de EVI.	132
Figura 117. Flujograma de extracción de métricas fenológicas de la firma mediana de la chacra.	133
Figura 118. Gráfico de métricas fenológicas de la firma mediana de la chacra.	134
Figura 119. Gráfico comparativo de métricas fenológicas de dos chacras de soja de FUCREA-URF para la zafra 2017-18.	135
Tabla 1. Campos presentes en la base PLANES.	12
Tabla 2. Distribución de planes por cultivo para el período acumulado 2013-19.	13
Tabla 3. Resumen de planes según cultivo para la zafra 2018-19.	14
Tabla 4. Resumen de las bases de FUCREA y URF para las zafras 2017-18 y 2018-19.	16
Tabla 5. Resumen por cultivo de la base FUCREA-URF de las zafras 2017-18 y 2018-19.	16
Tabla 6. Descripción de clases del mapa Integrado de Cobertura/Uso del Suelo de Uruguay del año 2018.	19
Tabla 7. Características del Sensor MODIS a bordo del satélite Terra de NASA.	25

Tabla 8. Características de los satélites Sentinel 2 (a y b) de la ESA.	26
Tabla 9. Características de los satélites Sentinel 1 (a y b) de la ESA.	26
Tabla 10. Características principales de los satélites utilizados.	26
Tabla 11. Cantidad y superficie (hectáreas) de planes con etiqueta de cultivo según zafra.	29
Tabla 12. Cantidad de planes y píxeles enteros de Terra extraídos para las zafras 2013-14 a 2018-19.	32
Tabla 13. Cantidad de planes, hectáreas y puntos por cultivo de las zafras 2017-18 y 2018-19.	32
Tabla 14. Cantidad y porcentaje de puntos regulares por cultivo y zafra en FUCREA-URF de las zafras 2017-18 y 2018-19.	33
Tabla 15. Agrupamiento jerárquico de clases de COBERTURAS.	34
Tabla 16. Cantidad de puntos filtrados por clase de COBERTURAS.	35
Tabla 17. Período que abarca cada imagen MOD13Q1 para el período analizado.	36
Tabla 18. Cantidad de series temporales extraídas de píxeles enteros y de planes para las zafras 2013-14 a 2018-19.	38
Tabla 19. Bandas utilizadas de Sentinel 1 y 2.	40
Tabla 20. Período abarcado por las imágenes sintéticas de Sentinel 1 y 2.	41
Tabla 21. Cantidad de puntos con series temporales por cultivo según base y zafra.	47
Tabla 22. Departamentos por Región.	48
Tabla 23. Algoritmos de clasificación no supervisada evaluados sobre los diferentes subconjuntos de región, cultivo y zafra.	55
Tabla 24. Cantidad de planes con etiqueta de cultivo para el período 2013-2019 depurados para entrenamiento.	60
Tabla 25. Cantidad de planes depurados totales en entrenamiento y evaluación según cultivo.	61
Tabla 26. Algoritmos de clasificación supervisada utilizados sobre el conjunto de entrenamiento en PLANES.	61
Tabla 27. Matriz de confusión de Random Forest en evaluación de PLANES.	62
Tabla 28. Métricas para evaluación de desempeño de algoritmos de clasificación supervisada.	63
Tabla 29. Precision de los algoritmos de clasificación supervisada evaluados en PLANES.	64
Tabla 30. Recall de los algoritmos de clasificación supervisada evaluados en PLANES.	64
Tabla 31. F1-Score de los algoritmos de clasificación supervisada evaluados en PLANES.	65
Tabla 32. Cantidad de puntos por cultivo de las zafras 2017-18 y 2018-19 de FUCREA-URF según depuración usada.	73
Tabla 33. Balanceo de puntos de la base FUCREA-URF.	73
Tabla 34. Ejemplo de tabla de NDVI con los valores por fecha de imagen sintética.	73
Tabla 35. Features del array multibanda por fecha y banda.	75
Tabla 36. F1-Score en el conjunto de evaluación de diferentes combinaciones de bandas e índices del array multibanda.	76
Tabla 37. F1-Score según la fecha de cierre del último período utilizado.	78
Tabla 38. F1-Score según cultivo y algoritmo supervisado evaluado.	79
Tabla 39. Cantidad y combinaciones de períodos de fechas utilizados para entrenar el algoritmo.	80
Tabla 40. Cantidad de chacras, hectáreas y puntos por cultivo de FUCREA-URF.	81
Tabla 41. Cantidad de puntos outliers depurados con diferentes metodologías de FUCREA-URF.	81
Tabla 42. Matriz de confusión en evaluación de la mejor corrida de FUCREA-URF.	83

Tabla 43. Valores promedio y desvío estándar de las métricas de desempeño para las 50 corridas en evaluación de FUCREA-URF.....	83
Tabla 44. Comparativa de F1-Score de maíz, soja y sorgo para el presente trabajo y artículos revisados.	85
Tabla 45. Cantidad de planes, hectáreas y puntos con series temporales de Sentinel según cultivo de las zafra 2017-18 y 2018-19 para PLANES.	86
Tabla 46. Cantidad de planes, hectáreas y puntos según cultivo después de la doble depuración para PLANES.	86
Tabla 47. Matriz de confusión en evaluación de la mejor corrida en PLANES.....	87
Tabla 48. Valores promedio y desvío estándar de las métricas de desempeño para las 50 corridas en evaluación de PLANES.	87
Tabla 49. Comparativo de métricas de desempeño en evaluación de FUCREA-URF y PLANES..	88
Tabla 50. Matriz de confusión en PLANES clasificado con modelo entrenado con FUCREA-URF balanceado.	91
Tabla 51. Métricas de desempeño en PLANES con modelo entrenado en FUCREA-URF con y sin balanceo.	91
Tabla 52. F1-Score por cultivo de las bases de FUCREA-URF, PLANES y artículos revisados.....	92
Tabla 53. Cantidad de puntos según cobertura/uso del suelo y zafra para entrenamiento y evaluación de modelo de clasificación binaria agrícola/no-agrícola.	93
Tabla 54. Matriz de confusión y métricas de desempeño de clasificación binaria agrícola/no-agrícola.	94
Tabla 55. Área agrícola (hectáreas) estimada por mapas binarios (EVI2, NDVI, VH_VV) y DIEA para la zafra 2017-18.....	100
Tabla 56. Área (hectáreas) y métricas de desempeño para los mapas binarios 2 a 5.	110
Tabla 57. Métricas de desempeño en mapa binario 3 según conjunto de evaluación.....	112
Tabla 58. Cantidad de puntos por cultivo para entrenamiento y evaluación en la zona de influencia de FUCREA-URF.	115
Tabla 59. Áreas (hectáreas) y porcentaje del área de maíz, soja y sorgo en mapas de moda de GTB y RF en zona de influencia FUCREA-URF.....	118
Tabla 60. Área (hectáreas) y porcentajes del total, de maíz, soja y sorgo para la zafra 2017-18 según DIEA.	119
Tabla 61. Métricas de desempeño por cultivo de GTB-MODA y RF-MODA.....	119
Tabla 62. Fechas de inicio y fin de las imágenes decadales sintéticas de NDVI.....	121
Tabla 63. Métricas fenológicas de la firma mediana de la chacra.	134
Tabla 64. Comparativo de métricas fenológicas de dos chacras de soja de FUCREA-URF para la zafra 2017-18.	136

16. Bibliografía

- Abburu, S., & Babu Golla, S. (2015). Satellite Image Classification Methods and Techniques: A Review. *International Journal of Computer Applications*, 119(8), 20-25. <https://doi.org/10.5120/21088-3779>
- Arenare, L., Couto, P., & Fontán, M. V. (2018). *Informe sobre Riego en Uruguay. Trabajos Especiales No 354*. Recuperado de http://www.mgap.gub.uy/sites/default/files/informe_riego_en_uruguay.pdf
- Baeza, S., Baldassini, P., Bagnato, C., Pinto, P., & Paruelo, J. M. (2014). Caracterización del uso / cobertura del suelo en Uruguay a partir de series temporales de imágenes MODIS. *Agrociencia Uruguay*, 18(2), 95-105.
- Bandaru, V., Yaramasu, R., PNVR, K., He, J., Fernando, S., Sahajpal, R., ... Justice, C. (2020). PhenoCrop: An integrated satellite-based framework to estimate physiological growth stages of corn and soybeans. *International Journal of Applied Earth Observation and Geoinformation*, 92(May), 102188. <https://doi.org/10.1016/j.jag.2020.102188>
- Bargiel, D. (2017). A new method for crop classification combining time series of radar images and crop phenology information. *Remote Sensing of Environment*, 198, 369-383. <https://doi.org/10.1016/j.rse.2017.06.022>
- Becker-Reshef, I., Justice, C., Sullivan, M., Vermote, E., Tucker, C., Anyamba, A., ... Doorn, B. (2010). Monitoring Global Croplands with Coarse Resolution Earth Observations: The Global Agriculture Monitoring (GLAM) Project. *Remote Sensing*, 2(6), 1589-1609. <https://doi.org/10.3390/rs2061589>
- Bendini, H., Sanches, I. D., Körting, T. S., Fonseca, L. M. G., Luiz, A. J. B., & Formaggio, A. R. (2016). USING LANDSAT 8 IMAGE TIME SERIES FOR CROP MAPPING IN A REGION OF CERRADO, BRAZIL. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLI-B8(July), 845-850. <https://doi.org/10.5194/isprsarchives-XLI-B8-845-2016>
- Breiman, L. (2001). Random forests. *Machine learning*, 45(1), 5-32.
- Carrasco-Letelier, L., & Beretta-Blanco, A. (2017). Soil erosion by water estimated for 99 Uruguayan basins. *Ciencia e Investigacion Agraria*, 44(2), 184-194. <https://doi.org/10.7764/rcia.v44i2.1717>
- Chamorro, J. A., Bermudez, J. D., Happ, P. N., & Feitosa, R. Q. (2019). A MANY-TO-MANY FULLY CONVOLUTIONAL RECURRENT NETWORK FOR MULTITEMPORAL CROP RECOGNITION. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, IV-2/W7, 25-32. <https://doi.org/10.5194/isprs-annals-IV-2-W7-25-2019>
- Chen, T., & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 13-17-Aug, 785-794. <https://doi.org/10.1145/2939672.2939785>
- Chuvieco, E. (2016). *Fundamentals of Satellite Remote Sensing*. CRC Press. <https://doi.org/10.1201/b19478>
- Cué La Rosa, L. E., Borges Oliveira, D. A., & Feitosa, R. Q. (2021). End-to-End CNN-CRFs for Multi-date Crop Classification Using Multitemporal Remote Sensing Image Sequences. *CEUR Workshop Proceedings*, 3052.
- Di Gregorio, A., & Jansen, L. J. M. (1998). *Land Cover Classification System (LCCS): Classification Concepts and User Manual*. Roma: FAO. Recuperado de https://www.researchgate.net/publication/229839605_Land_Cover_Classification_System_LCCS_Classification_Concepts_and_User_Manual
- Diao, C. (2020). Remote sensing phenological monitoring framework to characterize corn and soybean physiological growing stages. *Remote Sensing of Environment*, 248(July), 111960. <https://doi.org/10.1016/j.rse.2020.111960>
- Didan, K., Munoz, A. B., Solano, R., & Huete, A. (2015). MODIS vegetation index user's guide

- (MOD13 series). *University of Arizona: Vegetation Index and Phenology Lab*.
- EMBRAPA-CNPTIA. (2019). Sistema de Análise Temporal da Vegetação (SATVeg). Recuperado 7 de mayo de 2019, de <https://www.satveg.cnptia.embrapa.br/satveg/login.html>
- EMBRAPA. (2019). Sistema de Observação e Monitoramento da Agricultura no Brasil (SOMABRASIL) - Portal Embrapa. Recuperado 7 de mayo de 2019, de <https://www.embrapa.br/busca-de-solucoes-tecnicas/-/produto-servico/1345/sistema-de-observacao-e-monitoramento-da-agricultura-no-brasil-somabrazil>
- ESA. (2019a). SAR missions / Copernicus / Observing the Earth / Our Activities / ESA. Recuperado 6 de mayo de 2019, de https://www.esa.int/Our_Activities/Observing_the_Earth/Copernicus/SAR_missions
- ESA. (2019b). Sen4Cap. Recuperado 6 de mayo de 2019, de <http://esa-sen4cap.org/>
- ESA. (2019c). Sentinel-2 for Agriculture. Recuperado 6 de mayo de 2019, de <http://www.esa-sen2agri.org/>
- FAO. (2019). Countries | Food and Agriculture Organization of the United Nations. Recuperado 20 de febrero de 2019, de <http://www.fao.org/countryprofiles>
- Frazier, J., Mullin, C., Frazier, M., & Ashcraft, S. (2011). Pesticides and their involvement in colony collapse disorder. *American Bee Journal*, 151, 779-781.
- Freund, Y., & Schapire, R. E. (1997). A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting. *Journal of Computer and System Sciences*, 55(1), 119-139. <https://doi.org/https://doi.org/10.1006/jcss.1997.1504>
- Friedman, J. H. (2002). Stochastic gradient boosting. *Computational Statistics & Data Analysis*, 38(4), 367-378. [https://doi.org/10.1016/S0167-9473\(01\)00065-2](https://doi.org/10.1016/S0167-9473(01)00065-2)
- Gao, F., Anderson, M. C., Zhang, X., Yang, Z., Alfieri, J. G., Kustas, W. P., ... Prueger, J. H. (2017). Toward mapping crop progress at field scales through fusion of Landsat and MODIS imagery. *Remote Sensing of Environment*, 188, 9-25. <https://doi.org/10.1016/j.rse.2016.11.004>
- Giri Ananto, W. H., Putri, A. F. S., Hadi, H. A., Hanum, D. N., Wiryawan, B. K. P., Prabaswara, R. R., & Arjasakusuma, S. (2019). Performance of various speckle filter methods in modelling forest aboveground biomass using Sentinel-1 data: case study of Barru Regency, South Sulawesi. En S. B. Wibowo, A. B. Rimba, A. A. Aziz, S. Phinn, J. T. Sri Sumantyo, H. Widiasamratri, & S. Arjasakusuma (Eds.), *Sixth Geoinformation Science Symposium* (p. 38). SPIE. <https://doi.org/10.1117/12.2549036>
- INIA-GRAS. (2019). Portal INIA Sistemas de Información y Teledetección. Recuperado 7 de mayo de 2019, de <http://www.inia.uy/gras/Comunicación/Publicaciones/Sistemas-de-Información-y-Teledetección>
- INPE-DSR. (2019). CANASAT - Sugarcane crop monitoring in Brazil by Earth observing satellite images. Recuperado 7 de mayo de 2019, de <http://www.dsr.inpe.br/laf/canasat/en/index.html>
- INTA. (2018). Monitoreo de cultivos del Noroeste Argentino a partir de sensores remotos - Campaña agrícola 2017-2018 - Cultivos extensivos de verano | Instituto Nacional de Tecnología Agropecuaria. Recuperado 7 de mayo de 2019, de <https://inta.gob.ar/documentos/monitoreo-de-cultivos-del-noroeste-argentino-a-partir-de-sensores-remotos-campana-agricola-2017-2018-cultivos-extensivos-de-verano>
- International Grains Council. (2019). IGC Grains and Oilseeds Index (GOI). Recuperado 20 de febrero de 2019, de <https://www.igc.int/en/markets/marketinfo-goi.aspx>
- Jain, A. K. (2010). Data clustering: 50 years beyond K-means. *Pattern Recognition Letters*, 31(8), 651-666. <https://doi.org/10.1016/j.patrec.2009.09.011>
- Jönsson, P., & Eklundh, L. (2004). TIMESAT—a program for analyzing time-series of satellite sensor data. *Computers & Geosciences*, 30(8), 833-845. <https://doi.org/10.1016/j.cageo.2004.05.006>
- Kumari, M., Pandey, V., Choudhary, K. K., & Murthy, C. S. (2021). Object-based machine learning

- approach for soybean mapping using temporal sentinel-1/sentinel-2 data. *Geocarto International*, 0(0), 1-19. <https://doi.org/10.1080/10106049.2021.1952314>
- Kussul, N., Lavreniuk, M., Skakun, S., & Shelestov, A. (2017). Deep Learning Classification of Land Cover and Crop Types Using Remote Sensing Data. *IEEE Geoscience and Remote Sensing Letters*, 14(5), 778-782. <https://doi.org/10.1109/LGRS.2017.2681128>
- Llonch, D. A. (2019). La soja, la otra materia prima estratégica de Sudamérica. Recuperado 20 de febrero de 2019, de <https://www.unav.edu/web/global-affairs/detalle/-/blogs/la-soja-la-otra-materia-prima-estrategica-de-sudamerica>
- LUO, C., LIU, H., LU, L., LIU, Z., KONG, F., & ZHANG, X. (2021). Monthly composites from Sentinel-1 and Sentinel-2 images for regional major crop mapping with Google Earth Engine. *Journal of Integrative Agriculture*, 20(7), 1944-1957. [https://doi.org/10.1016/S2095-3119\(20\)63329-9](https://doi.org/10.1016/S2095-3119(20)63329-9)
- Maaten, L. van der, & Hinton, G. (2008). Visualizing data using t-SNE. *Journal of machine learning research*, 9(Nov), 2579-2605.
- Martin, E., Kaski, S., Zheng, F., Webb, G. I., Zhu, X., Muslea, I., ... Kersting, K. (2011). Self-Organizing Maps. En C. Sammut & G. I. Webb (Eds.), *Encyclopedia of Machine Learning* (pp. 886-888). Boston, MA: Springer US. https://doi.org/10.1007/978-0-387-30164-8_746
- MGAP-DGRN. (2016). Informe sobre Planes de Uso de Suelos presentados en el 2016 | MGAP .. Ministerio de Ganadería, Agricultura y Pesca de la República Oriental del Uruguay. Recuperado 20 de febrero de 2019, de <http://www.mgap.gub.uy/unidad-organizativa/direccion-general-de-recursos-naturales/suelos/planes-de-uso-y-manejo-de-suelos/informe-sobre-planes-de-uso-de-suelos-presentados-en-el-2016>
- MGAP-DGRN. (2019a). Cartas temáticas | MGAP .. Ministerio de Ganadería, Agricultura y Pesca de la República Oriental del Uruguay. Recuperado 6 de mayo de 2019, de <http://www.mgap.gub.uy/unidad-organizativa/direccion-general-de-recursos-naturales/tramites-y-servicios/biblioteca-digital/cartas-tematicas>
- MGAP-DGRN. (2019b). Mapa integrado de cobertura/uso del suelo del Uruguay año 2018 | Ministerio de Ganadería, Agricultura y Pesca. Recuperado 6 de junio de 2022, de <https://www.gub.uy/ministerio-ganaderia-agricultura-pesca/comunicacion/publicaciones/mapa-integrado-coberturauso-del-suelo-del-uruguay-ano-2018>
- MGAP-DGRN. (2019c). Proyecto FAO – MGAP(Año 2000) | MGAP .. Ministerio de Ganadería, Agricultura y Pesca de la República Oriental del Uruguay. Recuperado 6 de mayo de 2019, de <http://www.mgap.gub.uy/unidad-organizativa/direccion-general-de-recursos-naturales/tramites-y-servicios/biblioteca-digital/proyecto-fao-mgap>
- MGAP-DIEA. (2015). *REGIONES agropecuarias del Uruguay* (Vol. 1).
- MGAP-DIEA. (2016). *Anuario Estadístico de DIEA 2016*. Recuperado de <http://www.mgap.gub.uy/unidad-organizativa/oficina-de-programacion-y-politicas-agropecuarias/publicaciones/anuarios-diea/anuario2016>
- MGAP-DIEA. (2019). Anuario Estadístico de DIEA 2019 | Ministerio de Ganadería, Agricultura y Pesca. Recuperado 3 de enero de 2021, de <https://www.gub.uy/ministerio-ganaderia-agricultura-pesca/comunicacion/publicaciones/anuario-estadistico-diea-2019>
- Mohite, J., Sawant, S., Pandit, A., & Pappula, S. (2020). Integration of Sentinel 1 and 2 Observations for Mapping Early and Late Sowing of Soybean and Cotton Crop Using Deep Learning. En *IGARSS 2020 - 2020 IEEE International Geoscience and Remote Sensing Symposium* (pp. 1941-1944). IEEE. <https://doi.org/10.1109/IGARSS39084.2020.9323482>
- MVOTMA. (2019). Aplicaciones de mapas temáticos. Recuperado 7 de mayo de 2019, de <https://www.mvotma.gub.uy/sistema-de-informacion-territorial/aplicaciones-de-mapas-tematicos>
- NASA-Earth Observatory. (2000, agosto 30). Measuring Vegetation (NDVI & EVI), Normalized Difference Vegetation Index (NDVI). Recuperado 6 de mayo de 2019, de https://earthobservatory.nasa.gov/features/MeasuringVegetation/measuring_vegetatio

- Nova, D., & Estevez, P. A. (2015). A review of learning vector quantization classifiers. *Neural Computing and Applications*, 25(3-4), 511-524. <https://doi.org/10.1007/s00521-013-1535-3>
- NPP Science Team. (2011). *NPP NPOESS Preparatory Project Building a Bridge to a New Era of Earth Observations*. Recuperado de <http://npp.gsfc.nasa.gov>
- Pareja, L., Colazzo, M., Pérez-Parada, A., Niell, S., Carrasco-Letelier, L., Besil, N., ... Heinzen, H. (2011). Detection of pesticides in active and depopulated beehives in uruguay. *International Journal of Environmental Research and Public Health*, 8(10), 3844-3858. <https://doi.org/10.3390/ijerph8103844>
- Pelleg, D., & Moore, A. (2000). X-means: Extending K-means with Efficient Estimation of the Number of Clusters. En *In Proceedings of the 17th International Conf. on Machine Learning* (pp. 727-734). Morgan Kaufmann.
- Petraglia, C., Dell'Acqua, M., Pereira, G., & Yussim, E. (2019). *Mapa integrado de cobertura/ uso del suelo del Uruguay, año 2018. Anuario OPYPA* (Vol. 27). Recuperado de <https://www.gub.uy/ministerio-ganaderia-agricultura-pesca/comunicacion/publicaciones/mapa-integrado-coberturauso-del-suelo-del-uruguay-ano-2018>
- Podest, E. (2017). Basics of synthetic aperture radar. *National Aeronautics and Space Administration*, 29.
- Pollock, D. S. G. (1999). Smoothing with Cubic Splines. En *Handbook of Time Series Analysis, Signal Processing, and Dynamics* (pp. 293-322). Elsevier. <https://doi.org/10.1016/B978-012560990-6/50013-0>
- Potgieter, A. B., Apan, A., Hammer, G., & Dunn, P. K. (2005). «Spying» on the winter wheat crop-generating objective planted area and crop production estimates using MODIS imagery. En *Proceedings of SSC 2005 Spatial Intelligence, Innovation and Praxis: The national biennial Conference of the Spatial Sciences Institute* (pp. 1086-1090). Spatial Sciences Institute. Recuperado de <http://research.usc.edu.au/vital/access/manager/Repository/usc:5104>
- Rekik, A., Zribi, M., Hamida, A. B., & Benjelloun, M. (2009). An optimal unsupervised satellite image segmentation approach based on Pearson system and k-Means Clustering Algorithm Initialization. *World Academy of Science, Engineering and Technology*, 35(11), 2181-2188.
- Saleem, S., Aslam, M., & Shaukat, M. R. (2021). A review and empirical comparison of univariate outlier detection methods. *Pakistan Journal of Statistics*, 37(4), 447-462.
- Sammut, C., & Webb, G. I. (2017). *Encyclopedia of Machine Learning and Data Mining*. (C. Sammut & G. I. Webb, Eds.), *Encyclopedia of Machine Learning and Data Mining*. Boston, MA: Springer US. <https://doi.org/10.1007/978-1-4899-7687-1>
- Seo, B., Lee, J., Lee, K.-D., Hong, S., & Kang, S. (2019). Improving remotely-sensed crop monitoring by NDVI-based crop phenology estimators for corn and soybeans in Iowa and Illinois, USA. *Field Crops Research*, 238(October 2017), 113-128. <https://doi.org/10.1016/j.fcr.2019.03.015>
- Sharma, A., & Mishra, P. K. (2020). State-of-the-Art in Performance Metrics and Future Directions for Data Science Algorithms. *Journal of Scientific Research*, 64(2).
- Story, M., & Congalton, R. G. (1986). Remote Sensing Brief Accuracy Assessment: A User's Perspective. *Photogrammetric Engineering and Remote Sensing*, 52(3), 397-399. Recuperado de https://www.asprs.org/wp-content/uploads/pers/1986journal/mar/1986_mar_397-399.pdf
- Tan, B., Morisette, J. T., Wolfe, R. E., Gao, F., Ederer, G. A., Nightingale, J., & Pedelty, J. A. (2011). An Enhanced TIMESAT Algorithm for Estimating Vegetation Phenology Metrics From MODIS Data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 4(2), 361-371. <https://doi.org/10.1109/JSTARS.2010.2075916>
- Tian, F., Wu, B., Zeng, H., Zhang, X., & Xu, J. (2019). Efficient Identification of Corn Cultivation Area with Multitemporal Synthetic Aperture Radar and Optical Images in the Google Earth

- Engine Cloud Platform. *Remote Sensing*, 11(6), 629. <https://doi.org/10.3390/rs11060629>
- UBA-LART. (2019). LART - Publicaciones científicas. Recuperado de <http://lart.agro.uba.ar/publicaciones/>
- URUGUAY XXI. (2018). *Oportunidades de inversión. AGRONEGOCIOS. Marzo 2018*. Recuperado de <https://www.uruguayxxi.gub.uy/uploads/informacion/b5eb6050ca43af07330c66ca7fb7ba70f0cd8189.pdf>
- Vapnik, V., Guyon, I., & Hastie, T. (1995). Support vector machines. *Mach. Learn*, 20(3), 273-297.
- Volante, J., D Collado, A., B Ferreyra, E., López, C., Navarro Rau, M., Pezzola, A., ... Al. (2009). *Monitoreo de la Cobertura y el Uso del Suelo a partir de sensores remotos*.
- Von Luxburg, U. (2007). A tutorial on spectral clustering. *Statistics and Computing*, 17(4), 395-416. <https://doi.org/10.1007/s11222-007-9033-z>
- Vreugdenhil, M., Wagner, W., Bauer-Marschallinger, B., Pfeil, I., Teubner, I., Rüdiger, C., & Strauss, P. (2018). Sensitivity of Sentinel-1 Backscatter to Vegetation Dynamics: An Austrian Case Study. *Remote Sensing*, 10(9), 1396. <https://doi.org/10.3390/rs10091396>
- Vuolo, F., Neuwirth, M., Immitzer, M., Atzberger, C., & Ng, W.-T. (2018). How much does multi-temporal Sentinel-2 data improve crop type classification? *International Journal of Applied Earth Observation and Geoinformation*, 72(June), 122-130. <https://doi.org/10.1016/j.jag.2018.06.007>
- Waldner, F., Canto, G. S., & Defourny, P. (2015). Automated annual cropland mapping using knowledge-based temporal features. *ISPRS Journal of Photogrammetry and Remote Sensing*, 110, 1-13. <https://doi.org/10.1016/j.isprsjprs.2015.09.013>
- Wang, S., Azzari, G., & Lobell, D. B. (2019). Crop type mapping without field-level labels: Random forest transfer and unsupervised clustering techniques. *Remote Sensing of Environment*, 222(May 2018), 303-317. <https://doi.org/10.1016/j.rse.2018.12.026>
- World Bank. (2019). Prospects - Commodity Markets. Recuperado 20 de febrero de 2019, de <http://www.worldbank.org/en/research/commodity-markets#1>
- You, N., & Dong, J. (2020). Examining earliest identifiable timing of crops using all available Sentinel 1/2 imagery and Google Earth Engine. *ISPRS Journal of Photogrammetry and Remote Sensing*, 161(December 2019), 109-123. <https://doi.org/10.1016/j.isprsjprs.2020.01.001>
- Zeng, L., Wardlow, B. D., Xiang, D., Hu, S., & Li, D. (2020). A review of vegetation phenological metrics extraction using time-series, multispectral satellite data. *Remote Sensing of Environment*, 237(August 2019), 111511. <https://doi.org/10.1016/j.rse.2019.111511>
- Zhang, C., Di, L., Hao, P., Yang, Z., Lin, L., Zhao, H., & Guo, L. (2021). Rapid in-season mapping of corn and soybeans using machine-learned trusted pixels from Cropland Data Layer. *International Journal of Applied Earth Observation and Geoinformation*, 102, 102374. <https://doi.org/10.1016/j.jag.2021.102374>
- Zhong, L., Hu, L., Yu, L., Gong, P., & Biging, G. S. (2016). Automated mapping of soybean and corn using phenology. *ISPRS Journal of Photogrammetry and Remote Sensing*, 119, 151-164. <https://doi.org/10.1016/j.isprsjprs.2016.05.014>
- Zhong, L., Hu, L., & Zhou, H. (2019). Deep learning based multi-temporal crop classification. *Remote Sensing of Environment*, 221(November 2018), 430-443. <https://doi.org/10.1016/j.rse.2018.11.032>