

TRABAJO FIN DE MÁSTER

ESTIMACIÓN DE LA PRODUCCIÓN DE CEBADA A PARTIR DE IMÁGENES SATELITALES Y VARIABLES CLIMÁTICAS

ESTIMATING BARLEY YIELD FROM SATELLITE IMAGERY AND CLIMATIC VARIABLES

Autor: Cristian Iranzo Cubel

Director: Dr. Alberto García Martín
Directora: Dra. Raquel Montorio Llovería

Máster Universitario en
Tecnologías de la Información Geográfica para la Ordenación del
Territorio: Sistemas de Información Geográfica y Teledetección

Noviembre de 2020



Universidad
Zaragoza

**Departamento de Geografía
y Ordenación del Territorio**



Resumen

Estimar la producción de una explotación agrícola de forma precisa permite obtener información relevante a la hora de gestionar próximas campañas y evaluar las pérdidas provocadas por situaciones sinópticas adversas antes de la cosecha. El objetivo de este trabajo es explorar el desarrollo de un modelo predictivo capaz de estimar la producción de cebada en una pequeña explotación (127 ha), localizada en el municipio de Belchite (España). Se utilizan variables adaptadas a la fenología de la cebada cultivada, creadas a partir de información climática y de imágenes procedentes de teledetección, en dos rangos del espectro electromagnético, i.e., las microondas y el óptico, con Sentinel 1 y Sentinel 2 respectivamente. Los modelos se definen con una secuencia de código en R, utilizando todas las combinaciones de variables predictoras correlacionadas con la producción. Se utiliza una regresión lineal simple y el método *Partial Least Squares Regression* (PLSR). El modelo de regresión lineal con mejores resultados devuelve un error en la predicción de 146.39 kg/ha, superior al obtenido con el método PLSR, 106.42 kg/ha. Sin embargo, el error interno del primero, RSE de 455 kg, es inferior al modelo PLSR, con un un RMSEP de 8 481 kg. Se observa una mejora en los modelos con la integración conjunta de información climática y sobre el estado del cultivo, que debe validarse utilizando una serie de años más larga.

Palabras clave: Modelo predictivo, Sentinel 1, Sentinel 2, cebada

Abstract

A precise estimation of agricultural production expose relevant information for upcoming seasons and help asses the crop losses before harvest in case of adverse situations. This dissertation seeks to develop a predictive model able to estimate the barley production of a small agricultural production (127 ha) in Belchite, Spain. Variables adapted to the phenology of the growing barley are used to achieve that purpose. They are based on weather reports and remote sensing images in two ranges of the electromagnetic spectrum; microwaves and optical range obtained from Sentinel 1 and Sentinel 2 respectively. Models are defined with a simple linear regression and the Partial Least Squares Regression method, using all the possible combinations of the predictor variables correlated with production in R code. The linear regression model with better results gives a prediction error of 146.39 kg/ha, which is over the error obtained whith the PLSR method: 106.42 kg/ha. Nevertheless, the inner error achieved with the first model (RSE of 455 kg) is lower than the one acquired with the PLSR model, with a RMSEP of 8 481 kg. The study concludes that models improve with integrating both climate and crop state variables, despite they must be validated with larger series of years.

Keywords: Predictive modelling, Sentinel 1, Sentinel 2, barley

Índice

Índice de figuras	5
Índice de tablas	6
1. Introducción	7
1.1. Justificación del trabajo	7
1.2. Objetivos	7
1.3. Estado del arte	8
2. Área de estudio	10
3. Materiales	12
3.1. Datos de producción	12
3.2. Imágenes Sentinel 1	13
3.2.1. Pretratamientos de las imágenes en GEE	14
3.3. Imágenes Sentinel 2	15
3.4. Datos climáticos	16
4. Metodología	17
4.1. Delimitación del ciclo fenológico	17
4.2. Obtención de variables espectrales sobre el estado del cultivo	19
4.2.1. Corrección atmosférica	19
4.2.2. Procesamiento en GEE	22
4.2.3. Índices de vegetación	26
4.2.4. Transformación de los IV en variables espectrales	27
4.3. Obtención de variables climáticas	28
4.4. Creación de la base de datos	29
4.5. Creación de los modelos	29
4.5.1. Cálculo del error	31
5. Resultados	31
5.1. Relación entre la producción y las variables sobre el estado del cultivo	31
5.2. Relación entre la producción y las variables climáticas	35
5.3. Modelos de predicción	36
6. Discusión	38

7. Conclusiones	40
Bibliografía	47
Anexos	48
I. Fechas de las imágenes Sentinel 1 y Sentinel 2	48
II. Función para la extracción de métricas sobre el estado de la vegetación, Sentinel 1	49
III. Función para la extracción de métricas sobre el estado de la vegetación, Sentinel 2	54
IV. Corrección atmosférica mediante Py6s	63

Índice de figuras

1.	Producción de cebada por países (izquierda) y por cultivos (derecha) durante la temporada 2018/19.	8
2.	Producción de cebada en Belchite recogida por la Cooperativa de San Martín durante las cinco temporadas analizadas.	10
3.	Climograma del entorno de Belchite	11
4.	Localización de los recintos que componen la explotación.	12
5.	Comparación entre los ángulos medio de incidencia de las pasadas descendente y ascendente de las imágenes Sentinel 1 sobre los recintos del área de estudio.	14
6.	Esquema de la teledetección SAR y distorsión en la resolución de la imagen resultante.	15
7.	Número de imágenes por temporada y sensor.	16
8.	Proceso metodológico.	18
9.	Flujo de corrección mediante Py6s.	21
10.	Proceso de extracción de un área de 10 m alrededor de los recintos.	22
11.	Esquema de la función principal para las imágenes de Sentinel 1 que genera los valores del cociente VH/VV en GEE.	23
12.	Proceso de creación de la banda con la probabilidad de sombras, utilizada en la máscara de sombras. Modificado de [61].	25
13.	Comparación entre las imágenes con máscara de nubes y con máscara de nubes y sombras en situaciones de baja cobertura nubosa (izquierda) y alta (derecha).	25
14.	Serie temporal de NDVI calculada en la temporada 2018/19 mediante las imágenes a) sin ningún filtro, b) con primer filtro de nubes en GEE, c) con la máscara de nubes y d) con la aplicación del filtro en R.	26
15.	Serie temporal con los índices de vegetación durante las temporadas de cultivo analizadas en relación con la precipitación y la temperatura media diarias.	33
16.	Relación entre las variables sobre el estado de la vegetación mejor correlacionadas con la producción y los valores de esta última durante las cinco temporadas analizadas.	35
17.	Relación entre las variables climáticas mejor correlacionadas con la producción y los valores de esta última durante las cinco temporadas analizadas.	36

Índice de tablas

1.	Representación de los distintos tipos de geomorfología en función de su superficie total, el rendimiento obtenido sobre ellos y la superficie en porcentaje dentro de los dos ciclos de producción en los que se divide la explotación.	12
2.	Datos proporcionados por el gestor de la explotación	13
3.	Características de las imágenes de Sentinel 1 empleadas en el trabajo [12].	13
4.	Bandas espectrales del sensor Sentinel 2 MSI utilizadas en el trabajo.	16
5.	Ciclo fenológico de la cebada.	19
6.	Cálculo del error cuadrático medio (RMSE) de los valores medios de reflectividad sobre la explotación entre las imágenes corregidas con Sen2Cor (producto Sentinel 2A de GEE) y las imágenes corregidas con Py6s.	20
7.	Índices de Vegetación calculados a partir de las imágenes de Sentinel 1 y Sentinel 2.	26
8.	Variables sobre el estado de los cultivos definidas con cada índice de vegetación.	28
9.	Variables climáticas.	29
10.	Variables correlacionadas con la producción mediante un nivel de significación superior al 90% de confianza y un coeficiente de correlación superior al 80%. Se dividen entre las calculadas con la R de Pearson, utilizadas en el cálculo de los modelos de regresión lineal (derecha); y la R de Spearman, utilizadas junto con las anteriores en la definición de los modelos PLSR (izquierda). Se ordenan de forma ascendente por el p -value.	30
11.	Resultado de los modelos de regresión lineal, ordenados de menor a mayor MAE. En color gris se marcan los modelos con un p -value inferior a 0.05.	37
12.	Modelos de regresión PLSR con el menor RMSEP (1°), el mayor R^2 (2°) y el menor MAE (3°). Entre paréntesis se señalan las correlaciones de las variables predictoras con el componente principal utilizado superiores a ± 0.90	38

1. Introducción

1.1. Justificación del trabajo

El crecimiento de la población mundial ha provocado un aumento en el consumo de alimentos, cuya producción debería incrementarse un 2% anual para abastecer las necesidades de la nueva generación [17]. De lo contrario, la demanda se situaría por encima de la oferta poniendo en riesgo la seguridad alimentaria, incluida en el segundo Objetivo de Desarrollo Sostenible (ODS) de las Naciones Unidas. La puesta en práctica de estos principios requiere de explotaciones eficientes capaces de obtener la máxima producción al mínimo coste, asegurando la continuidad de la actividad agrícola y ganadera [34].

Una técnica capaz de poner en marcha estas medidas con un coste bajo para el agricultor se basa en el empleo de datos procedentes de la teledetección, i.e., conjunto de técnicas que permiten adquirir información a distancia de los objetos situados sobre la superficie terrestre, utilizando la radiación electromagnética [9]. El procedimiento consiste en emplear la información de las imágenes satelitales para estimar la producción de las explotaciones meses antes de la cosecha, creando mapas de rendimiento con los que detectar las zonas menos productivas [3]. Una vez localizadas, pueden aplicarse tratamientos específicos en los sectores más deficitarios de la explotación, reduciendo los costes económicos y los niveles de contaminación al ajustar la cantidad de agua, pesticidas y fertilizantes empleados [63]. La agricultura es una de las actividades humanas que más agua consumen y mayor cantidad de CO_2 libera a la atmósfera [49], por lo que reducir las externalidades negativas de las explotaciones contribuye a preservar el medio ambiente y al cumplimiento de los ODS.

Las estimaciones de la producción por sí solas también presentan ventajas, permitiendo calcular los posible beneficios de una temporada, lo que facilita la planificación de próximas campañas y, en último término, mejora la gestión de la explotación. Un beneficio adicional que podría derivarse de estimaciones precisas sobre el rendimiento es la mejora en el peritaje de daños. Los campos de cereal pueden verse afectados por las tormentas y el granizo en la parte final de la temporada de cultivo, cuyo efecto en la producción final podría obtenerse de la diferencia entre la producción predicha y la real.

El trabajo se centra en las explotaciones con producción de cebada, (*Hordeum vulgare* L. var. *Icaria*), el sexto cereal más producido a nivel mundial por detrás de los cultivos de maíz, arroz, trigo, soja y yuca [18]. Europa es el principal productor de cebada del mundo, con 56 millones de toneladas en la campaña 2018/19 [19], mientras que España es el quinto productor mundial, con 9.1 millones de toneladas producidas en la misma fecha (Figura 1). La mayor parte de la producción se destina a la alimentación animal y la elaboración de bebidas alcohólicas [50], sectores que, junto a la agricultura, podrían estar interesados en contar con estimaciones sobre la producción de cebada.

1.2. Objetivos

El objetivo del presente trabajo es explorar la capacidad de desarrollar un modelo para estimar la producción de cebada en una pequeña explotación dos meses antes de la cosecha, utilizando para ello un conjunto de variables adaptadas al ciclo fenológico del cultivo en la zona de estudio. Estas variables se derivan de información climática e imágenes satelitales de alta resolución espacial¹.

La metodología planteada para la consecución del objetivo principal se articula en las siguientes fases, explicadas con mayor detalle en el apartado 4.

¹*High Resolution 2*, con resoluciones espaciales entre 10 y 30 m. Es la categoría en la que se encuentran las bandas espectrales utilizadas según el *Global Monitoring for Environment and Security* (GMES).

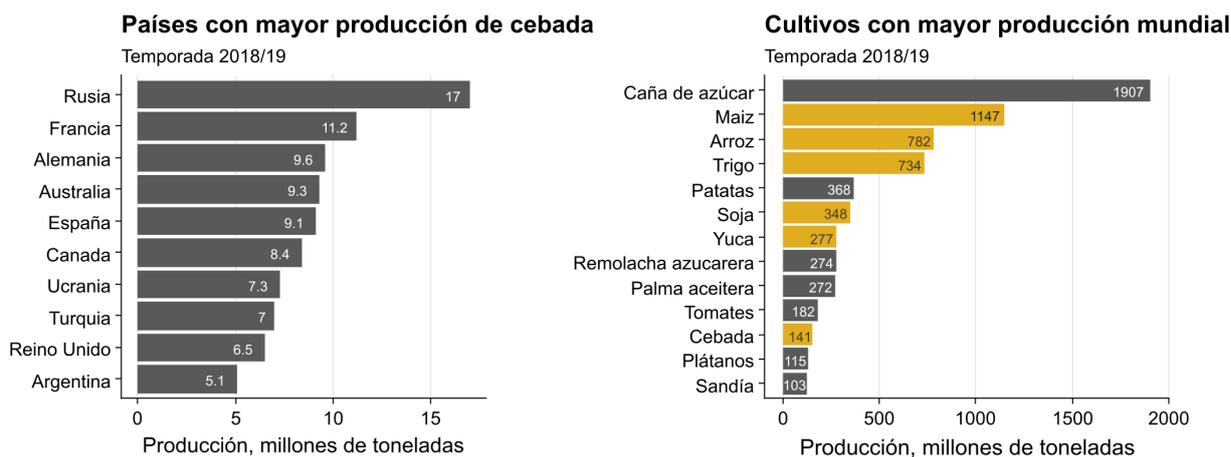


Figura 1. Producción de cebada por países (izquierda) y por cultivos (derecha) durante la temporada 2018/19. Los cultivos de cereal se muestran en color naranja. Fuente: FAOSTAT, elaboración propia.

1. Caracterización y delimitación del ciclo fenológico de la cebada en el área de estudio.
2. Definición de variables relacionadas con el estado de los cultivos mediante la extracción de información espectral sobre los campos cultivados en cada temporada, de forma automática, procesando las imágenes satelitales de las constelaciones Sentinel 1 y Sentinel 2 dentro de la plataforma Google Earth Engine (en adelante GEE).
3. Definición de variables asociadas al clima del área de estudio.
4. Creación de la base de datos combinando las variables predictoras anteriores con los cinco años de medidas de producción disponibles, entre las temporadas 2015/16 y 2019/20.
5. Ajuste automático de modelos de regresión utilizando el total de combinaciones disponibles entre las variables predictoras.

1.3. Estado del arte

En 1974, el *Large Area Crop Inventory Experiment* (LACIE), un proyecto conjunto del departamento de agricultura (USDA), la agencia espacial (NASA) y la agencia nacional para el estudio de la atmósfera y los océanos (NOAA) de Estados Unidos, demostró que la teledetección podía proveer información útil sobre el estado de los cultivos. Con esta información se consiguió estimar la producción varios meses antes de la cosecha con una precisión, a escala regional, superior a las encuestas de producción nacionales [40]. La ventaja de la teledetección respecto a las encuestas es su capacidad de obtener datos sobre grandes superficies de una forma más rápida y económica que los muestreos en campo [65], factor que compensa la pérdida de precisión con respecto a estas últimas.

Las estimaciones de la producción mediante información satelital se han basado principalmente en dos grupos de modelos, i.e., dinámicos y empíricos o estadísticos. Ambos se sustentan en la correlación observada entre la producción y una serie de índices de vegetación (en adelante IV). Su evolución durante la temporada detecta las variaciones experimentadas por la vegetación debidas a los cambios en las condiciones sobre las que se desarrolla el cultivo, e.g., climatología, tipo de suelo y nutrientes disponibles [25, 53].

Los modelos dinámicos simulan la evolución temporal de los cultivos sobre el medio en el que se desarrollan mediante un conjunto de ecuaciones matemáticas [35]. La ventaja de este tipo de modelos es su capacidad para captar las variaciones en la etapa de floración, momento crucial en el desarrollo de los cultivos. Sin embargo, utilizan un gran número de variables, muchas de las cuales deben ser

tomadas en campo. La dificultad de obtener muchas de estas variables en algunas zonas reduce la aplicación operacional de este tipo de modelos sobre explotaciones no experimentales. Algunos trabajos desarrollaron modelos generados íntegramente con información satelital, como el implementado por Idso *et al.* [32] en función de la temperatura acumulada y el porcentaje de energía reflejada por los cultivos, ambas captadas mediante teledetección. Otro problema añadido es la resolución temporal requerida, que varía desde datos diarios a semanales. Los satélites con una resolución espacial adecuada para el estudio de los cultivos a escala regional (i.e. inferior a 80m [23]) contaban con períodos de revisita próximos a los 16 días (e.g. Landsat 4, 5, 7 y 8). Además, en el rango del óptico, los problemas de nubosidad reducen el número de imágenes disponibles. Para paliar este tipo de problemas algunos estudios desarrollaron modelos dinámicos combinando información procedente de sensores en el rango del óptico y las microondas. Clevers y van Leeuwen [11] demostraron un incremento de la precisión de los modelos al estimar el índice de área foliar (LAI por sus siglas en inglés) con información procedentes de ambos espectros electromagnéticos. Más recientemente, Ameline *et al.* [1] obtuvo un error del 14 % al estimar la producción en 122 campos de maíz combinando medidas de GAI (*Green Area Index*) derivadas sensores en el rango del óptico y las microondas (Landsat 8 OLI y Sentinel 1A respectivamente).

La creación de modelos empíricos requiere de un menor número de variables pero, a diferencia de los modelos dinámicos, no pueden aplicarse sobre un área distinta de donde fueron entrenados. Los más utilizados se basan en una regresión lineal entre los valores de producción y los datos de un IV, agrupados dentro del período comprendido entre la siembra y el momento a partir del cual se realiza la predicción. La mayoría se aplican a escala regional, utilizando sensores con baja resolución espacial pero con un período de revisita diario [2, 14, 58]. Existen modelos desarrollados sobre áreas más reducidas utilizando imágenes con mayor resolución espacial, como el de Mirasi *et al.* [46], que consiguió estimar la producción sobre 1 200 ha de trigo 40 días antes de la cosecha con los datos del sensor OLI de Landsat 8 (30 m de resolución espacial y revisita cada 16 días). El error obtenido (i.e. 146 kg/ha) es inferior al de algunos modelos dinámicos recientes generados sobre una extensión similar, e.g. los 540 kg/ha de Zhao *et al.* [76] sobre cultivos de trigo. La mayoría de los modelos de regresión se definen con imágenes de la región del óptico. El espectro de las microondas ha sido menos utilizada debido a la fuerte interacción de estas longitudes de onda con factores como la humedad y la rugosidad del suelo. A escala global y regional, donde coexisten gran variedad de técnicas de cultivo y condiciones ambientales, es difícil relacionar el comportamiento de la señal con los cambios en el estado de la vegetación a lo largo del ciclo fenológico [41]. Además, las bandas derivadas del rango de las microondas son más dependientes del tipo de cultivo que las bandas del óptico. En los últimos años se ha desarrollado un nuevo IV con la región de las microondas, i.e., el Cross Ratio (CR), que mejora las series temporales derivadas sobre campos de cultivo al reducir el ruido provocado por el contenido de humedad, estabilizando los valores y aumentando su potencial para estimar la producción [69, 71].

Desde abril de 2017, las constelaciones de Sentinel 1 y Sentinel 2 desarrolladas por la *European Space Agency* (ESA), centradas en la región de las microondas y el óptico respectivamente, se encuentran plenamente operativas. Sus imágenes aumentan la resolución espacial (10-60 m), temporal (5-12 días) y radiométrica de los productos de teledetección de libre distribución. Cumplen los requisitos de monitorización agraria, i.e., un período de revisita próximo a los 10 días y una resolución espacial por debajo de los 80 m, por lo que son capaces de captar las variaciones estacionales de los cultivos y su espacialización en parcelarios más o menos homogéneos [23]. El potencial de estos satélites en la estimación de la producción apenas ha sido estudiado a un nivel de detalle en un caso real, con algunas excepciones [76]. Ambas constelaciones tienen una continuidad garantizada hasta más allá del 2030, siendo de interés su empleo en modelos predictivos y con potencial para generar productos operacionales, i.e., de forma automática a lo largo del tiempo.

2. Área de estudio

La explotación se localiza dentro del término municipal de Belchite, perteneciente a la provincia de Zaragoza, en la margen derecha de la depresión del Ebro. Forma parte de la comunidad autónoma de Aragón, el tercer productor de cebada en España por detrás de Castilla y León y Castilla-La Mancha [45]. La producción media de cebada en Belchite durante las temporadas analizadas, según los datos de la cooperativa San Martín, es de 3 650 toneladas (Figura 2).

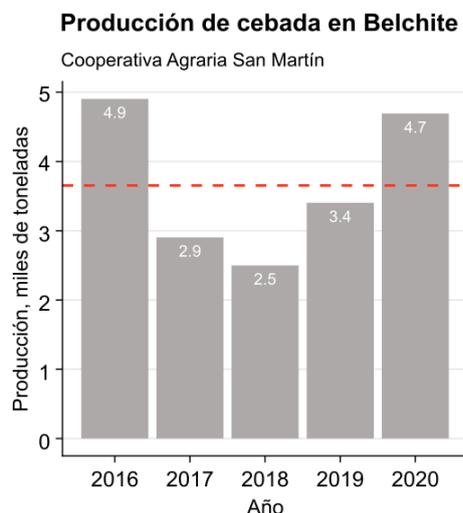


Figura 2. Producción de cebada en Belchite recogida por la Cooperativa de San Martín durante las cinco temporadas analizadas. Fuente: Cooperativa San Martín, elaboración propia.

Topográficamente es una zona llana, encajada entre los relieves tabulares de la Lomaza, al norte, y el río Aguasvivas, al sur. En cuanto a la litología, la mayor parte de este sector está compuesto por series neógenas arcillosas y yesíferas, intercaladas con depósitos de calizas y conglomerados calcáreos dispuestos de forma horizontal, generando problemas de drenaje y situaciones de inundación durante los eventos de precipitación torrencial [39]. Todo el sector septentrional está cubierto por depósitos aluviales cuaternarios, glaciares y niveles de terrazas que se extienden por la zona oriental hasta llegar al río Aguasvivas. Los problemas de drenaje y salinización derivados de este tipo de litologías y de la escasa pendiente se agravan debido al carácter semi árido del clima, clasificado como estepario en la escala de Köppen (BSk). La distribución de las precipitaciones, representada en el climograma de la figura 3, generan un verano seco, con los meses de julio y agosto con temperaturas que doblan la cantidad de precipitación. En invierno la precipitación aumenta, pero los valores más altos se producen en primavera durante el mes de mayo, seguido del mes de junio y dos meses de la estación de otoño, septiembre y octubre. Se reciben alrededor de 370 mm de precipitación anual, factor que junto a la elevada evapotranspiración, producida por el fuerte calor en los meses de verano, provoca un déficit hídrico que afecta al desarrollo de la vegetación. Respecto a la temperatura, la zona presenta inviernos fríos y veranos cálidos con una amplitud térmica anual de 17.6 °C. La altitud media de la zona, de 400 msnm, reduce ligeramente los máximos del verano y las variaciones diarias de temperatura con respecto a las zonas más bajas de la depresión del Ebro. Durante los meses de invierno y primavera la vegetación se ve afectada por las heladas, producidas por la bajada de las temperaturas por debajo de los 0 °C. Su probabilidad de ocurrencia es elevada entre diciembre y febrero, del 30.9 %, reduciéndose hasta el 5.1 % en marzo y el 1.4 % en abril.

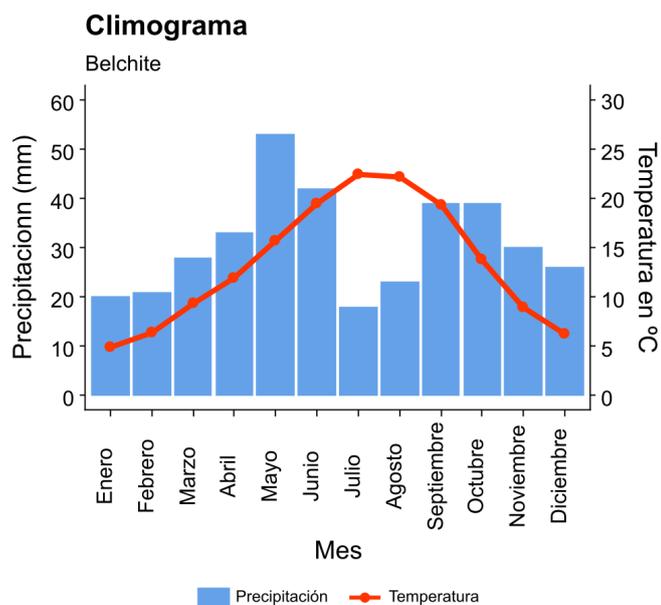


Figura 3. Climograma del entorno de Belchite. Fuente: climate-data.org, elaboración propia.

La explotación se compone de veinte recintos, representados en el mapa de la figura 4, con una superficie total de 127 ha y 2.7° de pendiente media. Se gestiona mediante un sistema de rotación de cultivos, sembrando la mitad de los recintos y dejando el resto en barbecho para la temporada siguiente. Provoca la existencia de dos ciclos de siembra en la explotación, el primero en años impares con inicio en la temporada 2015/16 y el segundo en años pares.

Este sistema provoca variaciones en las condiciones de producción derivadas de la utilización de distintos recintos dependiendo de la temporada. La más importante tiene que ver con el tipo de suelo, con una influencia directa en el rendimiento de la explotación (Tabla 1). Los recintos menos productivos (611.8 kg/ha) se sitúan sobre glaciares formados en sustratos de areniscas y arcillas miocenas, litologías que en climas áridos tienden a formar suelos con un alto porcentaje de arcillas expansivas (e.g. montmorillonita), caracterizados por un drenaje deficiente. Es la clase dominante en la explotación con 70 ha, pero se distribuye de manera desigual entre los dos ciclos de siembra. La superficie en el primer ciclo está compuesta en un 76.9 % por esta clase, mientras que en el segundo ciclo el porcentaje se reduce hasta el 32.7 %. La siguiente clase con menor rendimiento son las terrazas y glaciares cuaternarios, con 769.8 kg/ha. Los depósitos de terrazas suelen presentar una fertilidad limitada debido a su elevada capacidad de infiltración. Además, en zonas con un nivel freático próximo a la superficie, suelen aparecer problemas de encharcamiento y salinidad. Los depósitos de gravas cuaternarias son la tercera categoría con mayor superficie y con el segundo mejor dato de rendimiento por hectárea (800.4 kg/ha). Sin embargo, solamente se encuentran en el segundo ciclo, constituyendo el 26.7 % de toda su superficie. Esta misma situación se da en los recintos situados sobre depósitos aluviales de yesos (6.9 ha), con un 10.7 % de la superficie en el segundo ciclo. y las formaciones de glaciares sobre calizas terciarias (6.6 ha), con el 10.4 % del primer ciclo. Esta circunstancia provoca que el primer ciclo esté formado mayoritariamente por los sustratos con menor rendimiento. Sin embargo, el primer ciclo es el único con recintos sobre la formación de mayor rendimiento (1051.6 kg/ha), glaciares sobre calizas terciarias con conglomerados, 200 kg superior a la segunda. Las causas de su elevada productividad están relacionadas con el aporte de carbonatos, que sobre las suaves pendientes de los glaciares pueden formar calcisoles, i.e., suelos con un horizonte carbonatado favorable para la agricultura. La escasa precipitación permite a su vez que los materiales carbonatados no se filtren a capas profundas, inaccesibles para la vegetación [54]. Por otro lado, las variaciones entre temporadas producidas por el cambio en la superficie cultivada son mínimas, siendo la variación entre las hectáreas cultivadas en cada ciclo menores a 1 ha (63.55 ha en el primer ciclo y 64.43 ha en el segundo).

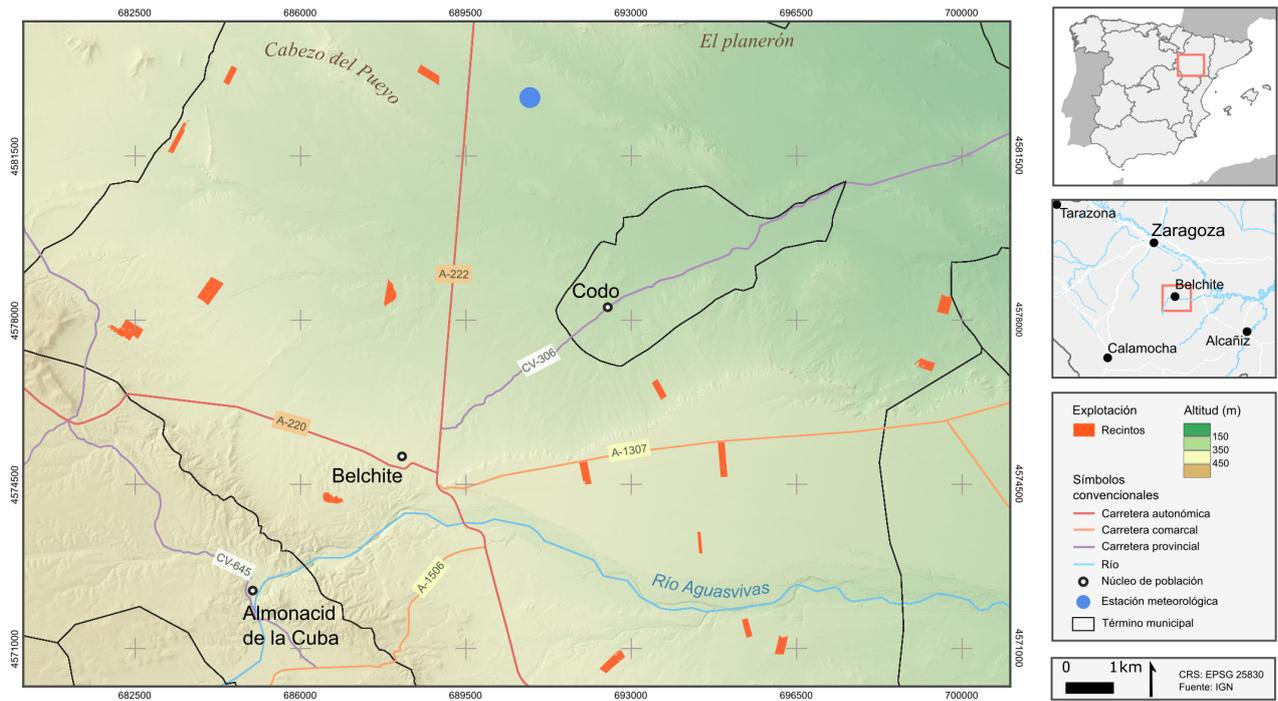


Figura 4. Localización de los recintos que componen la explotación.

Tabla 1. Representación de los distintos tipos de geomorfología en función de su superficie total, el rendimiento obtenido sobre ellos y la superficie en porcentaje dentro de los dos ciclos de producción en los que se divide la explotación. *Sup.* abreviatura de superficie, *Rend.* de rendimiento.

Geomorfología	Sup. (ha)	Rend. (kg/ha)	Sup. (%)	
			1 ^{er} ciclo	2 ^o ciclo
Glacis sobre areniscas y arcillas miocenas	70	611.8	76.9	32.7
Terrazas y glacis cuaternarios	27.3	769.8	12.6	29.8
Depósitos de gravas cuaternarias	17.25	800.4	0	26.7
Depósitos aluviales sobre Yesos	6.9	785.5	0	10.7
Glacis sobre calizas terciarias con conglomerados	6.6	1051.6	10.4	0

3. Materiales

3.1. Datos de producción

Los datos de producción han sido cedidos por el agricultor propietario de la explotación en kilos por recinto cultivado durante cinco años. Se calculan a partir de la información del rendimiento por hectárea, obtenida al final de la campaña, y los recibos proporcionados por la cooperativa tras la cosecha. En la tabla 2 se puede comprobar la estructura de la información en los cinco primeros recintos de la explotación, mostrando el sistema de rotación de cultivos empleado por el agricultor. En cada temporada los valores de producción se alternan con recintos sin cultivar (valor 0), en *barbecho* para ser cultivados en la siguiente temporada.

Tabla 2. Datos proporcionados por el gestor de la explotación. Muestra con los cinco primeros registros.

id	provincia	municipio	polígono	parcela	recinto	superficie	2016	2017	2018	2019	2020
1	50	45	510	142	1	6.9	0	13200	0	13700	0
2	50	45	504	188	1	3	0	4690	0	5250	0
3	50	45	504	13	1	3.1	0	5350	0	6190	0
4	50	45	518	96	1	5.1	5125	0	5230	0	8340
5	50	45	508	143	1	3.2	6130	0	7090	0	8620

3.2. Imágenes Sentinel 1

Los satélites activos² configurados dentro del espectro de las microondas son los denominados Radars de Apertura Sintética (SAR, por sus siglas en inglés). La misión Sentinel 1 esta compuesta por dos satélites SAR lanzados en abril de 2014 (1a) y abril de 2016 (1b). Ambos presentan una órbita circular casi polar, con 693 km de altura y 98.18° de inclinación [64]. El tiempo de revisita hasta el lanzamiento del segundo satélite era de 12 días, reduciéndose a la mitad con la puesta en órbita del Sentinel 1b (Tabla 3).

Tabla 3. Características de las imágenes de Sentinel 1 empleadas en el trabajo [12].

Banda	C
Frecuencia	5.405 MHz
Longitud de onda	5.6 cm
Modo de adquisición	Interferometric Wide Swath
Resolución en el rango	25 m
Resolución en el azimuth	25 m
Resolucion temporal	≈ 6 días
Pasada	Descendente
Polarizacion	Dual (VH y VV)
Ángulo de incidencia	$\approx 37^\circ$

Los satélites Sentinel 1 captan la retrodispersión en la banda C, con una frecuencia entorno a los 5.405 GHz y una longitud de onda de 5.6 cm. Esta región del espectro electromagnético se caracteriza por su transparencia al contenido de vapor de agua atmosférico (nubes y eventos de precipitación), además de evitar la contaminación en la respuesta espectral producida por la dispersión atmosférica y la radiación térmica de la tierra. Se han utilizado las imágenes tomadas en pasadas descendentes, con unos ángulos de incidencia medios de 37.5° y una desviación típica de 0.3° durante la serie estudiada. La reducida variabilidad de las pasadas descendentes contrasta con las ascendentes, con un mayor número de imágenes y una enorme variación en los ángulos de incidencia, con una desviación típica de 5° (Figura 5). La respuesta de la vegetación en la banda C con los ángulos de incidencia seleccionados es una combinación de la retrodispersión del suelo atenuada por la vegetación, correlacionada con la geometría del dosel vegetal, i.e., su altura, densidad y estructura interna [10, 22].

²Satélites que emiten la radiación reflejada por la superficie terrestre y detectada por el sensor.

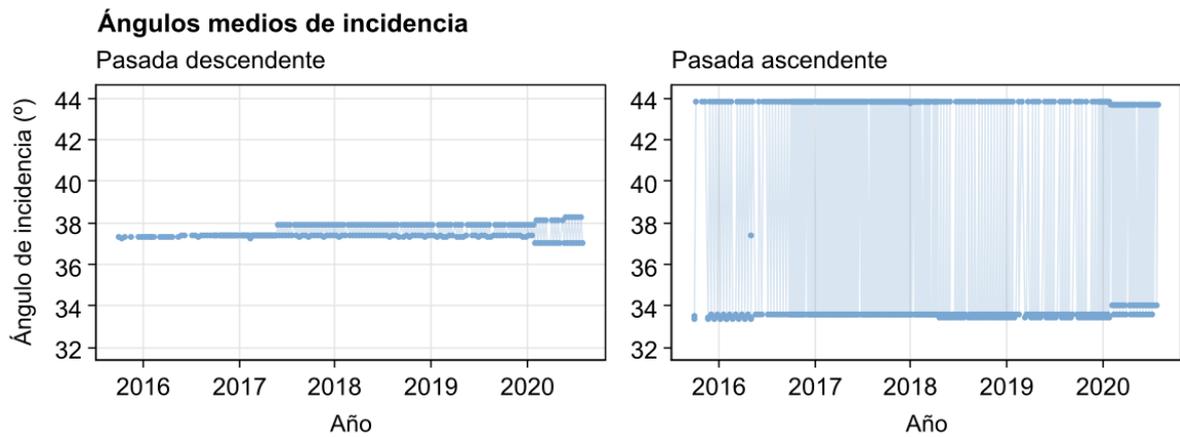


Figura 5. Comparación entre los ángulos medios de incidencia de las pasadas descendente (izquierda) y ascendente (derecha) de las imágenes Sentinel 1 sobre los recintos del área de estudio. Fuente: GEE, elaboración propia.

Se utiliza el conjunto de las imágenes Sentinel 1 con pasadas descendentes durante las cinco temporadas de cultivo analizadas, entre el 1 de octubre y el 31 de julio. El desfase en las fechas de lanzamiento es el responsable de una media de 2 imágenes mensuales en el período anterior a abril de 2016, aumentando después hasta las 4.8 imágenes mensuales (Figura 7).

3.2.1. Pretratamientos de las imágenes en GEE

Debido al proceso de obtención lateral de imágenes en los sensores SAR (Figura 6) la resolución espacial es distinta en el rango y en el azimut. Las imágenes de Sentinel 1 almacenadas en GEE están tomadas con el modo de adquisición *Interferometric Wide Swath (IW)*³, que incorpora un primer filtro espacial realizado por la ESA. Produce imágenes con una resolución espacial de 10 m (HR - *High Resolution*) homogénea en toda la escena (i.e. píxeles cuadrados), corrigiendo la resolución de las imágenes originales.

³Las imágenes se forman combinando tres subescenas tomadas en la misma trayectoria, dirigiendo el pulso emitido en diferentes direcciones del rango y del azimuth. Esta técnica se conoce como TOPSAR, *Terrain Observation by Progressive Scans SAR* [13].

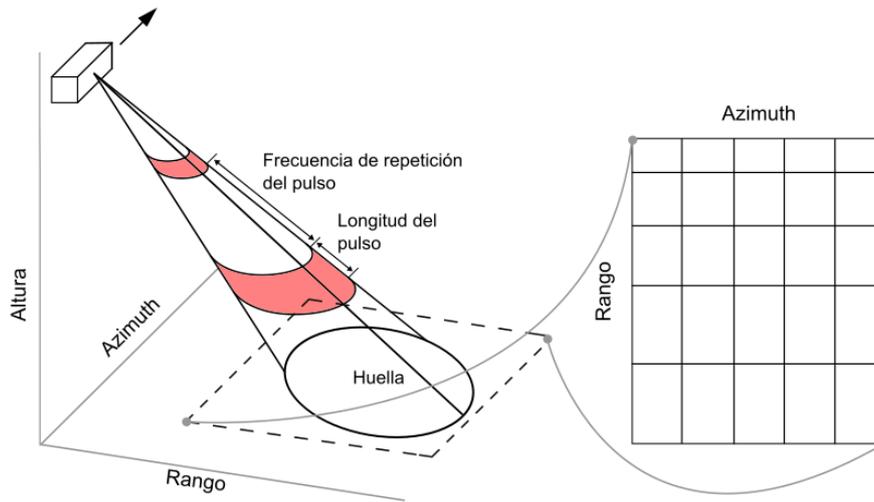


Figura 6. Esquema de la teledetección SAR (izquierda) y distorsión en la resolución de la imagen resultante (derecha). Modificado de [66].

Además de este filtro, la colección Sentinel 1 dentro de GEE incorpora una serie de tratamientos destinados a mejorar la calidad de las imágenes, aplicados mediante las herramientas desarrolladas por la ESA [15]:

1. Corrección de anomalías radiométricas, mejorando la precisión de la imagen final. Se aplica primero la herramienta `Apply Orbit File`, incorporando a los metadatos el fichero con información sobre la órbita. A continuación elimina el ruido en los márgenes de la imagen con `GRD border noise removal` y corrige las posibles alteraciones durante la combinación de las tres subescenas que forman la imagen IW (`Thermal noise removal`).
2. Calibración radiométrica, normalizando la retrodispersión en función de una unidad de superficie. Se obtiene el coeficiente σ° en geometría terreno (*ground*) referida al elipsoide, sin tener en cuenta la pendiente ni la orientación del terreno.
3. Georreferenciación en el sistema de referencia WGS84 (EPSG: 4326), utilizando el MDE de la misión *Shuttle Radar Topography Mission* (SRTM) con 30 metros de resolución. Se obtiene una imagen con geometría *Ground Range*, teniendo en cuenta la pendiente y la orientación del terreno.

El remuestreo de las imágenes con el MDE STRM de 30 m devuelve una imagen de 25 m de resolución en el rango y el azimuth. La pérdida de resolución espacial permite reducir el efecto de *speckle*, i.e., ruido coherente de las imágenes radar formado por las combinaciones aleatorias entre los elementos que componen los píxeles. Al trabajar con los valores medios de la explotación el efecto del *speckle* se reduce todavía más.

3.3. Imágenes Sentinel 2

La constelación Sentinel 2 está formada por dos satélites, 2a y 2b, lanzados en junio de 2015 y marzo de 2017 respectivamente. Presentan trece bandas dentro del espectro óptico, entre las longitudes de onda $0.3 \mu\text{m}$ y $2 \mu\text{m}$, detectando la fracción de energía solar reflejada por la superficie terrestre. Se utilizan las bandas 4-8 con una resolución espacial entre los 10 y 20 m (Tabla 4). Estas, junto a la resolución espacial de las imágenes obtenidas con Sentinel 1 (25 m), permiten trabajar con recintos superiores a 1 ha de tamaño [38].

Tabla 4. Bandas espectrales del sensor Sentinel 2 MSI utilizadas en el trabajo.

Nº de banda	4	5	6	7	8
Centro de banda (nm)	665	705	740	783	842
Anchura (nm)	30	15	15	20	115
Resolución espacial (m)	10	20	20	20	10

Se utilizan las imágenes de la cuadrícula 30TXL durante las cinco temporadas de cultivo analizadas. El período de revisita es de 10 días con un único satélite y de 5 días con ambos, generando una media de 2.1 imágenes mensuales antes del lanzamiento del segundo satélite y 5.7 imágenes con la puesta en órbita del Sentinel 2b (Figura 7). Es una media superior a las imágenes de Sentinel 1, debido a que esta última serie no cuenta con las imágenes de las pasadas ascendentes. Las fechas y el número de imágenes pueden consultarse en el anexo I.

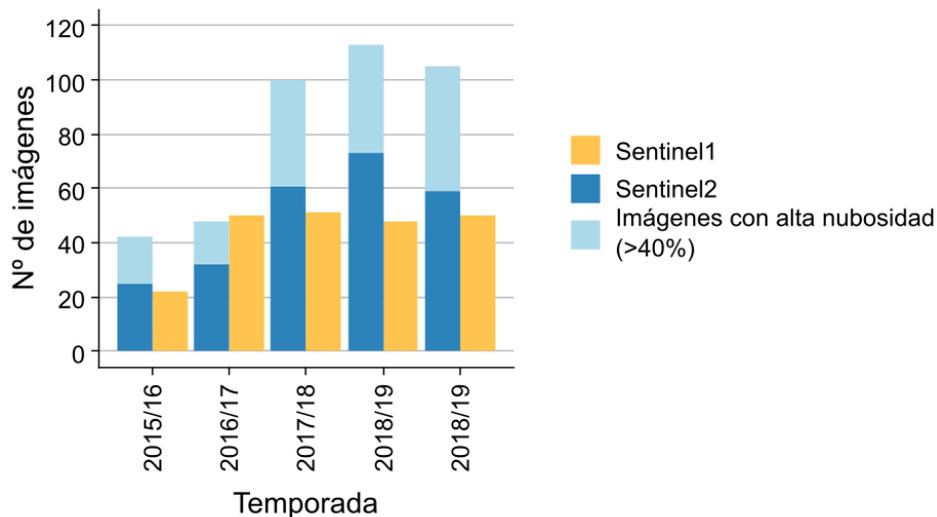


Figura 7. Número de imágenes por temporada y sensor. Fuente: GEE, elaboración propia.

A diferencia de Sentinel 1, las medidas captadas por los sensores de Sentinel 2 están afectadas por la atmósfera, siendo necesaria la corrección de las imágenes para poder relacionarlas con la evolución del estado de los cultivos (sección 4.2.1).

3.4. Datos climáticos

La cantidad de precipitación y las variaciones de temperatura en una temporada son los factores que más afectan al contenido de biomasa en los cultivos, relacionado con la producción final en una explotación [47]. Los datos diarios de temperatura y precipitación utilizados en la creación de las variables climáticas proceden de la estación de Belchite, situada en las coordenadas 41° 21'N - 0° 43' W (datum ETRS89) a 327 msnm (Figura 4). Es propiedad del ministerio de Agricultura, Pesca y Alimentación, y esta gestionada por el Sistema de información Agroclimática para el Regadío (SiAR⁴).

⁴<http://portal.mapa.gob.es/websiar/Inicio.aspx>

4. Metodología

La estimación de la producción se lleva a cabo mediante el proceso resumido en la figura 8. El primer paso es la caracterización del ciclo fenológico de la cebada en la zona de estudio, cuyas etapas sirven de base en la creación de las variables predictoras así como en la definición de la fecha límite a partir de la cual realizar la estimación, i.e., momento en el cual los cultivos alcanzan su máximo nivel de vigor vegetal. Al establecer un límite común para todas las temporadas, los desfases entre la fecha escogida (i.e. 15 de abril) y el momento real de máximo vigor en cada temporada medido con los IV pueden relacionarse con aumentos y descensos en la producción. Al definir el ciclo fenológico se delimita también el inicio y final de la temporada de cultivo en la zona de estudio, el 1 de octubre y el 31 de julio respectivamente, fechas utilizadas en la obtención de las imágenes satelitales.

El segundo paso es la obtención de la información, con la que se definen un total de 63 variables predictoras, divididas entre variables climáticas y variables sobre el estado de los cultivos. Detectan respectivamente las diferencias en el clima de cada temporada y los cambios en las condiciones de cultivo relacionadas con la gestión de la explotación. La creación de estas últimas comienza con la digitalización de la explotación, empleada en la extracción de las series temporales de IV con las que son definidas. Esta extracción se lleva a cabo mediante de forma automática con dos funciones escritas dentro de GEE, específicas para cada constelación (Sentinel 1 y Sentinel 2), sin necesidad de descargar las imágenes originales. Un paso previo a la aplicación del código en GEE es la corrección atmosférica de las imágenes de Sentinel 2 y la aplicación de dos máscaras que reducen el efecto de las nubes y sus sombras en las medidas.

Una vez obtenidas las variables predictoras se integran en una base de datos donde se relacionan con la producción en las cinco temporadas de cultivo analizadas, desde la 2015/16 a la 2019/20. A continuación se generan los modelos predictivos con todas las combinaciones de variables predictoras correlacionadas con la producción, obteniendo el dato predicho de la temporada 2019/20. De entre todos los modelos se seleccionan aquellos con el menor error interno, descartando los modelos sobre ajustados a la muestra de datos con la que se entrenan.

4.1. Delimitación del ciclo fenológico

Estimar la producción de los cultivos con imágenes satelitales de forma precisa requiere de información sobre su fenología, analizando en qué momento del ciclo fenológico se encuentra la máxima respuesta espectral de la vegetación y su relación con la producción final. En este trabajo se estudia también el grado de mejora introducido por las variables predictoras adaptadas a las fases del ciclo fenológico. La escala de trabajo permite ajustarse al ciclo de la variedad de cebada cultivada en Belchite, teniendo en cuenta además que se desarrolla bajo un sistema de rotación de cultivos en régimen de secano, i.e., con el agua procedente de la precipitación como único aporte hídrico.

La cebada es un cultivo con una gran capacidad de adaptación, producido en más de 48 millones de hectáreas en todo el mundo durante 2018 (Figura 1), desde países del trópico hasta las latitudes del Círculo Polar Ártico [18]. La cebada cultivada en la explotación analizada, pertenece a la variedad *Icaria*, de dos carreras. Se desarrolló con la colaboración de Agromonegros y la Estación Experimental de Aula Dei del CSIC, inscrita en el Registro de Variedades en marzo de 2010 [28]. Está adaptada al clima de secano presente en el área de estudio, desarrollándose en condiciones semiáridas con una buena capacidad de ahijamiento⁵, resistiendo las heladas y la sequía. Estas características permiten la siembra de la cebada con un espaciado entre semillas inferior a los 12 cm. En temporadas con buenas condiciones de cultivo este espaciado permite aumentar la densidad de pies y la cantidad de biomasa, facilitando su medición mediante teledetección [53]. Sin embargo, en temporadas de sequía extrema

⁵Desarrollo de pies secundarios en la base del tallo principal, con capacidad de generar nuevos granos que aumenten la producción.

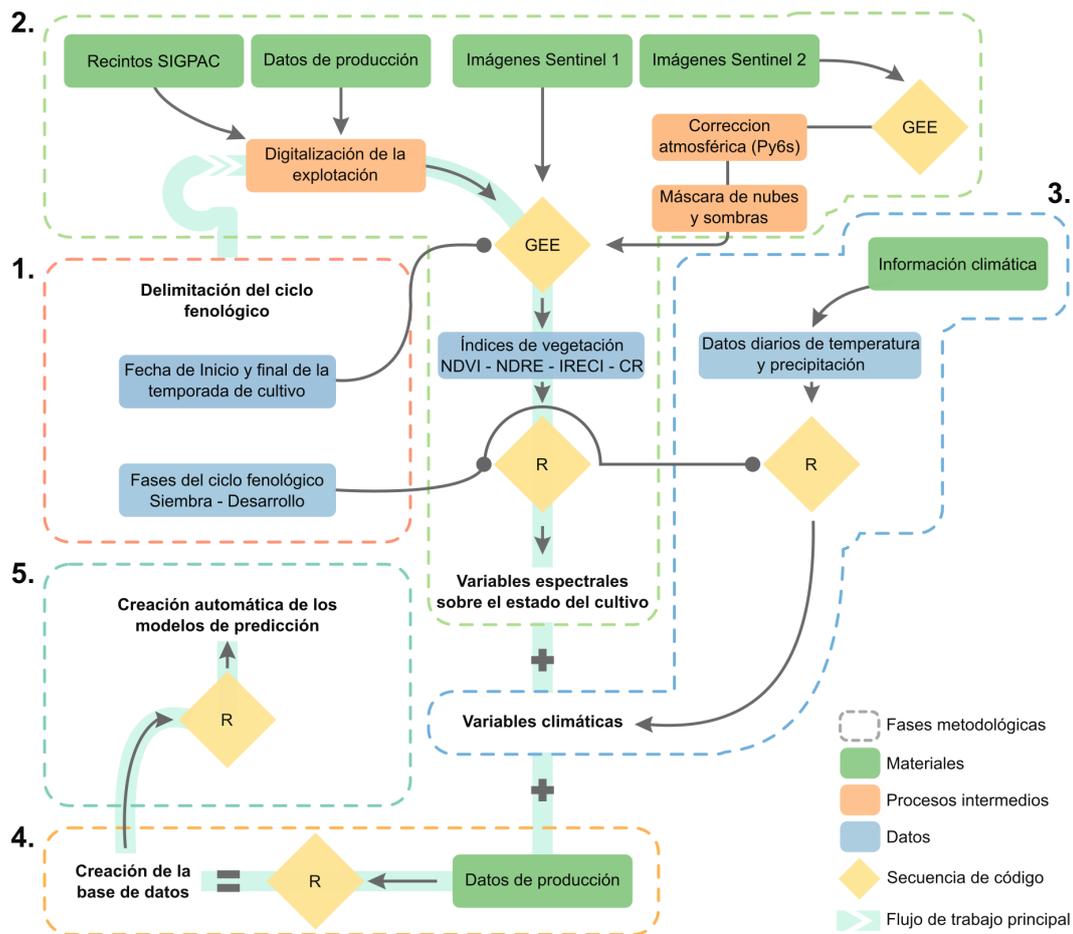


Figura 8. Proceso metodológico.

aumenta la competencia entre las plantas, reduciendo el crecimiento de la vegetación y la producción final. Las desventajas de la cebada *Icaria* son 1) su espigado precoz, que disminuye su capacidad para aprovechar las lluvias primaverales en el llenado de los granos; y 2) la tendencia al *encamado*, i.e., abatimiento de la planta en las fases finales que acerca la espiga al suelo y reduce la altura del cultivo, ralentizando la cosecha ⁶.

El ciclo fenológico de la cebada en el área de estudio, resumido en la tabla 5, comienza en octubre con la siembra de la semilla. Este proceso suele realizarse con suelos húmedos, tras un período de lluvias, aumentando la probabilidad de nacimientos. El nacimiento temprano de la vegetación permite a la planta aprovechar las lluvias otoñales para desarrollar su estructura antes de la bajada de las temperaturas en los meses de invierno. A finales de diciembre, como consecuencia del descenso térmico, se activa la *vernalización*, mecanismo que bloquea el desarrollo de los órganos reproductores hasta la subida de las temperaturas, evitando el deterioro en la estructura vegetativa de la planta durante el invierno. Si las temperaturas en esta fase no descienden lo suficiente, la cebada iniciará la fase de ahijamiento demasiado pronto, aumentando el riesgo de muerte por heladas tardías.

Durante los últimos días de diciembre y principios de enero se inicia la fase de ahijamiento, donde la vegetación desarrolla varios pies o brotes en la base del tallo, que continúa creciendo, aumentando la biomasa del cultivo. La capacidad de ahijamiento de la planta constituye un factor clave en la producción final, pues cada uno de los nuevos brotes es capaz de generar una nueva espiga, que puede traducirse en una mayor cantidad de granos. Con la reducción en la probabilidad de heladas a

⁶Las cosechadoras deben ir más despacio para recoger los granos.

principios de marzo, el crecimiento se acelera. Comienza entonces una de las etapas más importantes en el desarrollo de la planta, la floración, marcando la transición del estado vegetativo al reproductivo. El éxito de esta fase radica en el desarrollo de los órganos reproductores dentro de un intervalo de tiempo que evite una posible muerte de las flores por una helada tardía.

Entre finales de marzo y comienzos del mes de abril, una vez desarrollado el sistema reproductivo, las plantas son fertilizadas, iniciando el desarrollo del grano. Durante esta fase se alcanza el máximo contenido de biomasa y humedad en los cultivos, captado por los IV con un pico en los niveles de reflectividad dentro de la serie temporal. En la zona de Belchite este momento se produce entorno al 15 de abril, fecha elegida como el límite a partir del cual realizar la predicción. Después la respuesta de los IV comienza a descender, coincidiendo con el aumento térmico de finales de la primavera que inicia la maduración de los granos y el secado de las plantas, adoptando una coloración amarillenta. Es el comienzo de la fase de senescencia, donde la respuesta espectral de los cultivos de cereal ha demostrado no aportar información que mejore la predicción [46]. El ciclo finaliza con la cosecha de los cultivos entre los meses de junio y julio.

Las fechas de las etapas del ciclo fenológico definidas con anterioridad varían en cada temporada en función de la climatología y las condiciones de cultivo introducidas por la diferente localización de los recintos. Son captadas por las plantas, que son capaces de adaptar su desarrollo a estos factores [33].

Tabla 5. Ciclo fenológico de la cebada.

Estadio	Meses	Período seleccionado
Siembra	Octubre-noviembre	1 de oct - 15 de nov
Nacimiento	Noviembre-diciembre	16 de nov - 25 de dic
Ahijamiento	Enero-febrero	26 de dic - 28 de feb
Desarrollo de la espiga	Febrero-abril	29 de feb - 15 abr
Senescencia	Abril-junio	16 de abr - 15 de jun
Cosecha	Junio-julio	16 de jun - 31 de jul

4.2. Obtención de variables espectrales sobre el estado del cultivo

La producción de un cultivo está relacionada con el estado de la vegetación que lo compone, estado que se traduce en aumentos o descensos de su biomasa, cambios en la arquitectura del dosel vegetal y variaciones en el contenido de clorofila [76]. Estas variaciones, junto a las diferencias en el desarrollo de los cultivos provocadas por cambios en las condiciones de producción, pueden detectarse mediante series temporales de información espectral. En este trabajo, las variables sobre el estado del cultivo se obtienen combinando los valores de reflectividad en distintas bandas de las imágenes de Sentinel 1 y Sentinel 2. El resultado son una serie de índices de vegetación que concentran la información en una única banda y reducen el ruido producido por cubiertas no vegetales (e.g. suelos) [24]. Las imágenes satelitales de la región de las microondas (Sentinel 1) captan información relacionada con la estructura del dosel vegetal y su contenido de humedad [5, 42, 71]; mientras que en la región del óptico (Sentinel 2) puede derivarse el contenido de clorofila, agua y vigor vegetal de las hojas [37, 67]. A continuación se describe el proceso de corrección aplicado a las imágenes de Sentinel 2, la secuencia de código en GEE con la que se obtienen los índices de vegetación, las características de cada uno de ellos y su transformación en variables espectrales adaptadas al ciclo fenológico de la cebada.

4.2.1. Corrección atmosférica

La radiación captada por el sensor pasivo MSI de Sentinel 2 está afectada por la luz atmosférica, i.e., cantidad de energía procedente de la interacción entre la atmósfera y el flujo de radiación solar.

La mayor parte procede de la dispersión o refracción de la energía solar incidente por parte de las partículas atmosféricas en suspensión, que se suma a la radiación reflejada por las cubiertas de la superficie distorsionando los valores de reflectividad captados por el sensor. Asociar los cambios en las series temporales de IV con las variaciones del estado de la vegetación de un modo preciso solo es posible reduciendo la influencia atmosférica.

Las imágenes Sentinel 2 disponibles en GEE incorporan el nivel de procesamiento 2A, corregidas radiométrica y atmosféricamente mediante el algoritmo Sen2Cor y devolviendo valores transformados a reflectividad en superficie (BOA, por sus siglas en inglés). Este nivel de corrección está disponible en imágenes posteriores a abril de 2017. En el período anterior solo se habilita el nivel de tratamiento 1C, i.e., datos transformados a reflectividad a techo de la atmósfera (TOA, por sus siglas en inglés), sin corrección atmosférica. Dado que la serie analizada comienza en 2015, se aplicó una corrección atmosférica de forma automática directamente a las imágenes de GEE entre el 1 octubre de 2015 y el 31 de marzo de 2017. Se utiliza para ello el modelo de transferencia radiativa *Second Simulation of the Satellite Signal in the Solar Spectrum* (6s) [70], uno de los métodos de corrección más extendidos en aplicaciones operacionales. Su complejidad de cálculo es menor que otros modelos físicos más avanzados, como Sen2Cor. Se aplica utilizando el módulo Py6s [75] dentro de una secuencia de código en python, creada por Sam Murphy [48] y modificada para poder corregir en bucle todas las imágenes de la colección con nivel 1C.

Con el objetivo de comprobar la viabilidad de la integración conjunta de las imágenes corregidas con ambos métodos se calcularon las diferencias entre sus niveles de reflectividad mediante el error cuadrático medio (RMSE, por sus siglas en inglés). Se compararon imágenes 1C corregidas con Py6s en un período donde coincidían con las imágenes 2A, i.e., entre el 31 de marzo de 2017 y el 31 de marzo de 2018. Los valores de RMSE, representados en la tabla 6, no sobrepasan los 0.025 tantos por uno de reflectividad en ninguna de las bandas. El promedio de residuales, negativo, indica que la corrección 6s subestima ligeramente los valores de reflectividad de Sen2Cor. Dados los bajos valores de RMSE, la integración conjunta de ambas imágenes es viable, descartando una variación significativa entre sus valores de reflectividad capaz de introducir un sesgo en las métricas derivadas. Así, del 1 de octubre de 2015 al 31 de marzo de 2017 se utilizan imágenes corregidas con 6S, y en el resto de la serie con Sen2Cor.

Tabla 6. Cálculo del error cuadrático medio (RMSE) de los valores medios de reflectividad sobre la explotación entre las imágenes corregidas con Sen2Cor (producto Sentinel 2A de GEE) y las imágenes corregidas con Py6s.

Banda	Promedio residuales	RMSE
B2	-0.008	0.018
B3	-0.013	0.019
B4	-0.009	0.017
B5	-0.017	0.024
B6	-0.008	0.018
B7	-0.008	0.017
B8	-0.014	0.024
B8A	-0.004	0.016
B11	-0.01	0.02
B12	-0.008	0.022

La secuencia de código utilizada en la corrección de las imágenes con el método 6s se resume en la figura 9. Comienza con la definición de las funciones desarrolladas por Sam Murphy a partir del módulo Py6s, aplicadas a las bandas de la imagen por separado. La primera (`spectralResponseFunction`) calcula la respuesta espectral de las bandas de Sentinel 2, i.e., la longitud de onda central en la que captan la energía. La segunda (`toa_to_rad`) convierte los valores de reflectividad TOA de las imágenes GEE en radiancia, medida que mejora los cálculos del módulo Py6s. La tercera función (`surface_reflectance`) corrige la acción atmosférica transformando la fórmula de la radiancia obte-

nida por el sensor (ecuación 1) para despejar la reflectividad en superficie ρ , obteniendo la ecuación 2, donde L es la radiancia captada por el sensor, τ la transmisividad, E_{dir} irradiancia solar directa, E_{dif} irradiancia solar difusa y L_p el *path radiance*⁷.

$$L = \frac{\tau\rho(E_{dir} + E_{dif})}{\pi + L_p} \quad (1)$$

$$\rho = \frac{\pi(L - L_p)}{\tau(E_{dir} + E_{dif})} \quad (2)$$

Las cuatro incógnitas atmosféricas de la ecuación 2 (τ , E_{dir} , E_{dif} and L_p) se calculan con el módulo **Py6s**, utilizando la información de la base de datos de AERONET⁸ [31] a partir de la fecha de adquisición de la imagen. Se recopilan también los datos sobre la altura entre el sensor y la superficie y el ángulo cenital solar. A continuación se aplica la corrección a cada banda de la imagen con las tres funciones de [48], exportando el resultado a la nube personal de GEE. Este proceso se lleva a cabo dentro de una función principal (**conversion**), permitiendo su aplicación conjunta sobre cada una de las imágenes con nivel 1C dentro de un bucle **for**. El código íntegro puede consultarse en el anexo IV⁹.

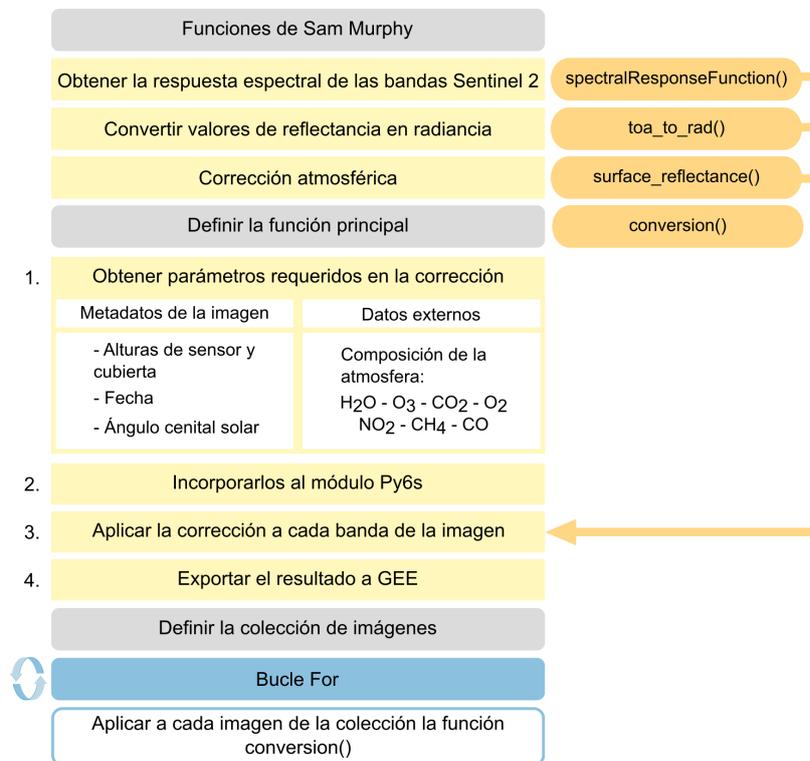


Figura 9. Flujo de corrección mediante Py6s.

⁷Cantidad de radiación captada por el sensor procedente de la dispersión atmosférica [36].

⁸<https://aeronet.gsfc.nasa.gov/>

⁹Este flujo de corrección no es aplicable a imágenes de gran tamaño debido a su lentitud de cálculo. Se aplicó un recorte sobre el área de estudio a las imágenes durante el filtrado de la colección, reduciendo su extensión y agilizando el proceso de corrección.

4.2.2. Procesamiento en GEE

La plataforma GEE es un entorno web estructurado con librerías Python y JavaScript para el análisis de datos geoespaciales. Utiliza imágenes de Earth Engine, una base de datos abierta que incorpora las colecciones de varios satélites, entre ellos Sentinel 1 y Sentinel 2 [27]. Una de sus ventajas es el procesamiento de las imágenes directamente en su base de datos, sin necesidad de descargarlas. El cálculo de los IV se ha efectuado mediante dos secuencias de código específicas para cada constelación, integrando información espacial con los recintos de la explotación junto con la fecha de inicio y final de las temporadas de cultivo.

Un aspecto clave a la hora de relacionar los valores de IV con el estado de la vegetación es evitar la inclusión en las medidas de píxeles contaminados, i.e., con información externa a los límites de los cultivos. Con el objetivo de extraer datos de las imágenes satelitales únicamente sobre el área cultivada se digitalizaron los recintos de la explotación mediante el software QGIS [56], extrayendo los polígonos SIGPAC procedentes del servicio *Web Feature* (WFS) del IDEAragón [26]. Se seleccionaron a través del código de polígono, parcela y recinto proporcionados por el agricultor (Tabla 2) generando una nueva capa en formato *shapefile*. Se eliminó un área de 10 m alrededor de los límites de los recintos, siguiendo el proceso mostrado en la figura 10, con el objetivo de reducir el ruido provocado por la vegetación localizada en la linde de los recintos. A esta capa se añadió después la información sobre la producción anual por recinto y año, utilizada en la extracción de los IV dentro de GEE para seleccionar los recintos cultivados ($\text{producción} > 0$) de cada temporada.

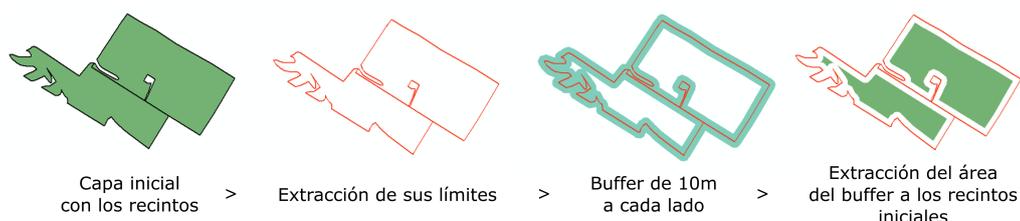


Figura 10. Proceso de extracción de un área de 10 m alrededor de los recintos.

Las dos secuencias de código utilizadas realizan cuatro acciones principales:

1. Crear la colección de imágenes e incorporar de la capa con los recintos de la explotación (sección 4.2).
2. Definir las funciones para el cálculo de los IV.
3. Definir la función principal, que devuelve la serie temporal con los valores medios del IV sobre los recintos cultivados en cada una de las temporadas analizadas.
4. Llamar a la función principal, junto al nombre de la colección y el año, mes y día de inicio y final de la serie temporal. En la función de Sentinel 2 deben definirse además el IV a calcular, obteniendo individualmente valores para los índices NDVI, NDRE e IRECI.

La función principal calcula el IV por temporadas de cultivo y únicamente sobre los píxeles de la imagen dentro de los recintos cultivados en la temporada. Este proceso se realiza de acuerdo al esquema de la figura 11. Comienza definiendo dos variables, i.e., el año de inicio y final de la serie, creadas con el año de inicio y final incluido al llamar a la función. Después crea una colección vacía que contendrá los valores del IV sobre los recintos cultivados en las temporadas de cultivo que componen la serie analizada. Una vez definidas, se crea una función de recorte y se ejecuta un bucle `for`, aplicando la siguiente secuencia una vez por temporada:

1. Seleccionar los recintos cultivados, i.e., recintos cuyo valor de producción en el año final de la temporada sea superior a 0.
2. Filtrado de la colección por la fecha de inicio y final de la temporada. Los meses y los días se definen al llamar la función. El año de inicio y final varía en cada iteración, utilizando las variables con el año inicial y final creadas en el primer paso de la función principal.
3. Recorte de las imágenes por la capa de cultivos filtrada en el primer paso de la iteración.
4. Cálculo del indicador en las imágenes de la colección filtrada (en el caso de la figura 11 el cociente VH/VV).
5. Añadir las imágenes a la colección vacía creada en el paso dos de la función principal, mediante un condicional *If*. Si la colección de imágenes creada dentro del bucle corresponde con el año inicial se integrarán a la colección, si no las imágenes se unirán a la colección ya existente.

Tras completarse el bucle se ejecuta el último paso de la función principal, representando en un gráfico el valor medio del IV en cada una de las imágenes de la nueva colección, exportando el resultado.

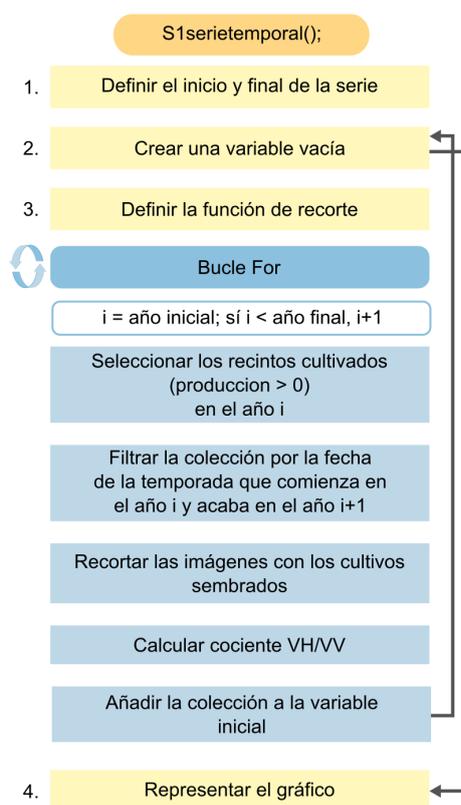


Figura 11. Esquema de la función principal para las imágenes de Sentinel 1 que genera los valores del cociente VH/VV en GEE.

Además de esta secuencia general con la que se estructuran los dos códigos, se integran varios pretratamientos adicionales en función del sensor. Dentro de la secuencia de código aplicada a las imágenes de la colección Sentinel 1 se realiza un paso previo a la obtención del CR. Las imágenes originales dentro de GEE están transformadas a decibelios (dB), por lo que la aplicación del cociente CR sobre las bandas VH y VV , con valores de signo negativo, devuelve valores positivos. Sin embargo, en los estudios donde ha sido aplicado los valores de CR son negativos [60, 64, 69, 71], pues realizan

la conversión a decibelios una vez calculado el CR. De cara a obtener valores comparables con estos trabajos, el código desarrollado en GEE integra una serie de funciones que permiten obtener valores CR negativos mediante la siguiente secuencia:

1. Transformación de las bandas con polarización VV y VH a valores sigma σ° mediante la inversa del logaritmo, $10^{dB/10}$, donde dB son los valores en decibelios de la imagen.
2. Cálculo del cociente VH/VV .
3. Transformación logarítmica de los valores del cociente mediante la conversión a logaritmo con la ecuación $10 * \log_{10}(x)$, donde x se corresponde con los valores de retrodispersión σ° . El resultado son valores de CR en decibelios con signo negativo.

El código aplicado a las imágenes de Sentinel 2 comienza definiendo la máscara de nubes y sombras. Las imágenes ópticas están afectadas por el vapor de agua que forma las nubes, desencadenando un efecto de dispersión no selectivo que produce píxeles con una reflectividad elevada en todas las bandas. La máscara de nubes utiliza la información de la colección `COPERNICUS/S2/CLOUD PROBABILITY`, con datos acerca de la probabilidad de nubes en cada píxel de la imagen, eliminando los píxeles con una probabilidad de nubes superior al 65 %. La máscara de sombras, adaptada del código de Schmitt *et al.* [61], elimina los píxeles con bajos niveles de reflectividad consecuencia de la proyección de las sombras de las nubes sobre la superficie. Al no existir información sobre el porcentaje de sombras en GEE, se crea una nueva banda en cada imagen con la probabilidad de que un píxel recoja una sombra en tantos por uno. Su obtención, resumida en la figura 12, comienza desplazando los píxeles clasificados como nubes, derivados de la colección `COPERNICUS/S2/CLOUD PROBABILITY`, sobre la posición teórica de sus sombras. Estas se calculan a partir de un vector con la altura teórica de las nubes y el ángulo de iluminación solar en la escena. Se obtiene así una nueva banda que es multiplicada por los píxeles de la imagen con una alta probabilidad de sombras:

1. Suma del valor de reflectividad en las bandas 8, 11 y 12 inferior a 0.3
2. Valor de la diferencia normalizada entre las bandas 8 (infrarrojo) y 4 (rojo) superior a -0.1, dejando fuera los píxeles de agua (< -0.1).

Todo este proceso se aplica únicamente a las imágenes con un porcentaje de nubes inferior al 40 %, descartando el resto a través de un filtro a la hora de definir la colección.

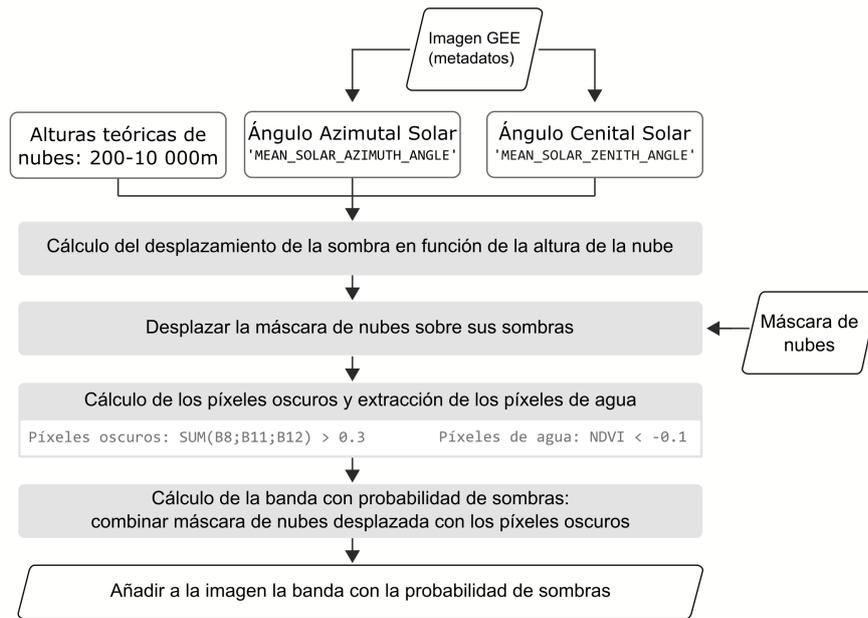


Figura 12. Proceso de creación de la banda con la probabilidad de sombras, utilizada en la máscara de sombras. Modificado de [61].

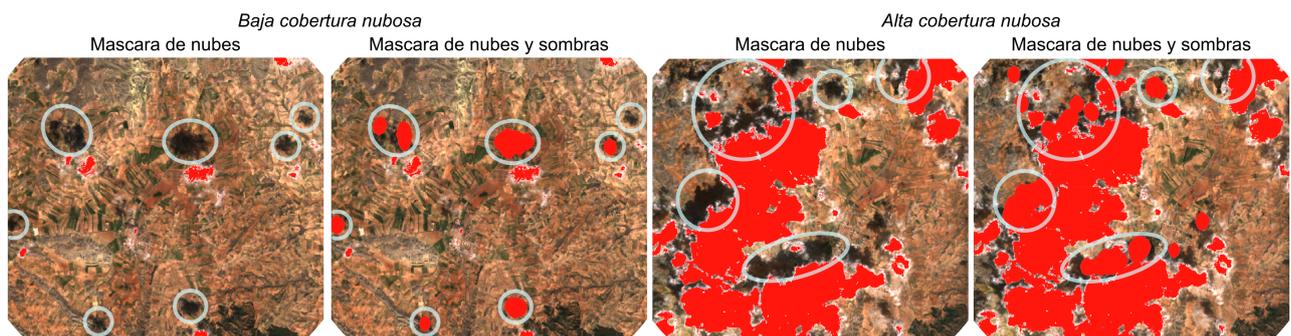


Figura 13. Comparación entre las imágenes con máscara de nubes y con máscara de nubes y sombras en situaciones de baja cobertura nubosa (izquierda) y alta (derecha).

Al aplicar la máscara de nubes y sombras se consigue eliminar una gran cantidad de píxeles afectados por la cobertura nubosa (Figura 13). Sin embargo, en situaciones con alto porcentaje de nubes las máscaras aplicadas reducen su efectividad, provocando que todavía aparezcan valores anómalos dentro de la serie. Se eliminaron mediante un filtro con código en R[57], recorriendo las series temporales y clasificando los valores entre anómalos y normales en función de dos criterios: 1) la diferencia absoluta entre el dato y la media del valor anterior y posterior no debe ser superior a 0.06; y 2) el porcentaje de nubes dentro de los recintos de la explotación en la misma fecha debe ser inferior al 40%. Después se eliminan de la serie los valores anómalos.

En la figura 14 se comprueba la efectividad del conjunto de procesos aplicados a las imágenes de Sentinel 2 mediante la evolución de la serie temporal de NDVI durante la temporada 2018/19. Las imágenes sin ningún tipo de filtro presentan un mayor número de datos contaminados por la nubosidad, con bruscos aumentos y descensos. Al aplicar el primer filtro de nubes se eliminan los valores afectados por una alta cobertura nubosa al inicio de la temporada (DOY 0-100). La máscara de nubes y sombras corrige el valor anómalo alrededor del DOY 100, pero la serie mantiene un patrón en dientes de sierra a mediados de la temporada (DOY 100-200) como consecuencia a varios datos que

todavía presentan errores. Estos se eliminan con el filtro en R, obteniendo una serie temporal mejor adaptada a las variaciones del estado de la vegetación.

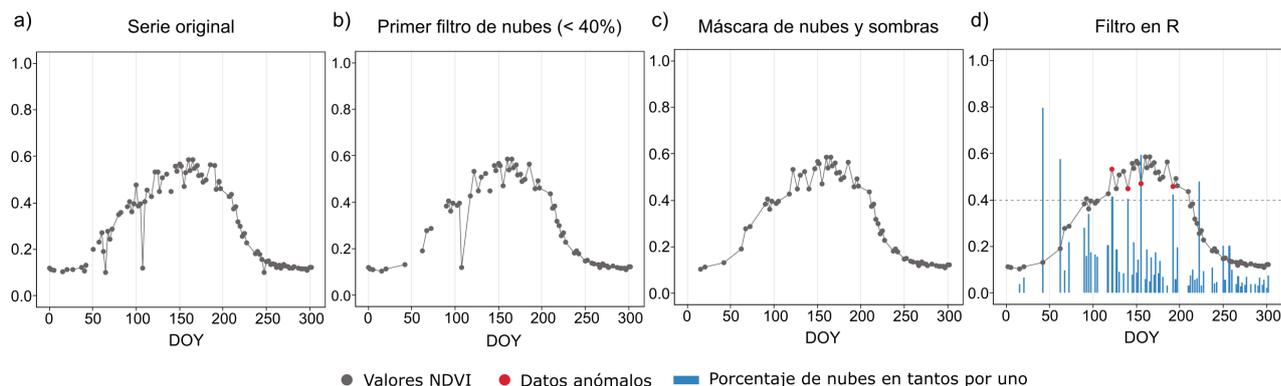


Figura 14. Serie temporal de NDVI calculada en la temporada 2018/19 mediante las imágenes a) sin ningún filtro, b) con el filtro de nubes, c) con la máscara de nubes y sombras y d) con la aplicación del filtro en R. Este último muestra los valores anómalos (rojo) según los dos criterios aplicados en el filtro de R y el porcentaje de nubes sobre los recintos de la explotación cultivados en tantos por uno. DOY: días de la temporada de cultivo comenzando en el momento de la siembra y finalizando con la cosecha.

Los códigos completos pueden consultarse en los anexos I (función Sentinel 1) y II (función Sentinel 2).

4.2.3. Índices de vegetación

Los IV calculados aplicando los códigos anteriores son el *Cross Ratio* (CR) con las imágenes de Sentinel 1 y el *Normalized Difference Vegetation Index* (NDVI), el *Normalized Difference Red Edge Index* (NDRE) y el *Inverted Red Edge Chlorophyll Index* (IRECI) con las imágenes de Sentinel 2 (Tabla 7).

Tabla 7. Índices de Vegetación calculados a partir de las imágenes de Sentinel 1 y Sentinel 2. La columna *Bandas* hace referencia al número de las bandas del sensor MSI utilizadas en la elaboración de los índices del óptico. R: región del rojo, NIR: infrarrojo próximo y RE región del *red edge*.

IV	Ecuación	Bandas	Autor
CR	VH/VV		Veloso <i>et al.</i> [69]
NDVI	$(NIR - R)/(NIR + R)$	$(B8 - B4)/(B8 + B4)$	Rouse <i>et al.</i> 1974 [59]
NDRE	$(RE3 - R)/(RE3 + R)$	$(B7 - B4)/(B7 + B4)$	Frampton <i>et al.</i> 2013 [21]
IRECI	$(RE3 - R)/(RE1/RE2)$	$(B7 - B4)/(B5/B6)$	Guyot y Baret 1988 [29]

El CR se obtiene aplicando un cociente entre las bandas con polarización¹⁰ VH y VV . Aportan información de las estructuras del cultivo horizontales (e.g. las hojas) y verticales (e.g. tallos) respectivamente. Sobre cultivos de cebada y trigo, con geometría similar, la aplicación de este cociente ha demostrado aumentar la retrodispersión del dosel vegetal (captada por la polarización doble VH) y disminuir la respuesta del suelo (gracias a la polarización vertical VV). [16, 69]. Se ha observado también que el CR es capaz de generar series más estables reduciendo el ruido provocado por los cambios en el contenido de humedad [71]. La interacción entre el pulso SAR y la superficie terrestre, basada en la permeabilidad electromagnética de las cubiertas [68], aumenta con la presencia de humedad, desencadenando una respuesta superior en las imágenes posteriores a un evento de precipitación.

¹⁰Se denomina polarización a la orientación del pulso electromagnético emitido por el sensor SAR.

En relación a los IV del óptico, el NDVI es uno de los índices que mejor reacciona al contenido de clorofila de la vegetación [21, 67, 72]. Sin embargo, su sensibilidad a la reflectividad del suelo es alta y tiende a saturar con elevados niveles de biomasa, i.e., valores de NDVI superiores a 0.8 [11]. El NDVI se calcula aplicando una diferencia normalizada entre las bandas 4 y 8 de Sentinel 2 (Tabla 7), longitud de onda del rojo (entorno a $0.645 \mu m$) y el infrarrojo próximo (entre $0.7-1.1 \mu m$) respectivamente. El comportamiento de estas dos regiones está claramente diferenciado en la firma espectral de la vegetación sana, con una absorción de energía en la longitud del rojo y reflejando gran cantidad de la radiación recibida en la longitud infrarroja. Por lo tanto, la ratio entre las dos bandas devuelve un índice capaz de medir el contenido de vegetación sana en un píxel [59], con efectividad en la estimación de la producción probada en varios trabajos [46, 55].

La absorción de la vegetación producida en el rango del rojo está asociada con el contenido de clorofila de las hojas, mientras que la reflexión en el rango del infrarrojo está correlacionada con la estructura de la hoja, con valores máximos asociados a vegetación sana con estructuras foliares óptimas. La diferencia de reflectividad entre estas dos regiones crea una zona de transición formada por una pendiente positiva desde el rojo hasta el infrarrojo próximo. El sensor MSI de Sentinel 2 cuenta con tres bandas en esta región, denominada *red edge*. Son la B5, B6 y B7, situadas respectivamente al inicio, en el momento de máxima pendiente y en el tramo final del *red edge*. La B5 comienza justo después de la B4, momento en el que la reflectividad comienza a incrementarse y disminuye la relación entre la absorción de energía y el contenido de clorofila. Según Frampton et al. [21], la B7 ofrece una mejor discretización de la señal de la vegetación sana que la B8. Por este motivo se decidió utilizar también el NDRE, índice de diferencia normalizada que reemplaza la B8 del NDVI por la B7.

El IRECI integra las tres bandas del *red edge* dentro de un cociente entre la resta de la B7 con la banda del rojo y la división entre las dos bandas restantes del *red edge*. El denominador de la ecuación pretende estimar el grado de la pendiente en el *red edge*. Con pendientes elevadas, como las presentes en la firma espectral de la vegetación sana, el cociente entre la B5 y B6 disminuye su valor, aumentando la probabilidad de obtener un alto valor de IRECI al reducir el denominador del cociente. La aplicación de este índice sobre cubiertas con cultivos ha demostrado una menor saturación en situaciones de alta densidad vegetal que el NDVI. [21, 43].

4.2.4. Transformación de los IV en variables espectrales

Una vez calculadas las series con los valores medios de los IV sobre los recintos cultivados se crean las métricas sobre el estado del cultivo. El proceso se realiza mediante una secuencia de código en RStudio [57], agrupando la información mediante seis métodos distintos:

1. Promedio de los IV entre la siembra y el desarrollo.
2. Promedio en las etapas de siembra, nacimiento, ahijamiento y desarrollo.
3. Suma de los promedios por estaciones hasta el final de la fase de desarrollo.
4. Valores máximos y mínimos en las etapas de siembra, nacimiento, ahijamiento y desarrollo.
5. Suma de los valores máximos y mínimos hasta el final la fase de desarrollo.
6. Percentil 95th entre las etapas de siembra y desarrollo, con buenos resultados al estimar la producción un mes y medio antes de la cosecha con productos de NDVI cada 16 procedentes del sensor MODIS [2].

Cada uno de ellos es aplicado a cada IV obteniendo un total de 63 variables, mostradas en la tabla 10. La suma de los promedios y de los valores máximos y mínimos durante el ciclo fenológico pretende emular a los trabajos que utilizan la suma del NDVI desde la siembra hasta el momento de

máximo desarrollo, con buenos resultados en las estimaciones de la producción a escala regional [58]. Promediar los valores por temporada o escoger el valor máximo y mínimo permite reducir la diferencia existente en el número de imágenes por temporada (Figura 7). De lo contrario, el resultado en un año podría ser inferior solamente por presentar un número menor de imágenes, independientemente de su valor de producción.

Tabla 8. Variables sobre el estado de los cultivos definidas con cada índice de vegetación. El código de cada índice se completa con su abreviación en el lugar de *IV*.

Variable	Código
Promedio entre las estaciones de siembra y desarrollo	IV_promedio
Suma de los promedios por estaciones (siembra-desarrollo)	sumIV_estacion
Promedio en la etapa de siembra	IV_siembra
:	:
Promedio en la etapa de desarrollo	IV_desarrollo
Valor máximo/mínimo en la etapa de siembra	max/minIV_siembra
:	:
Valor máximo/mínimo en la etapa de desarrollo	max/minIV_desarrollo
Suma de los valores máximo/mínimo por estaciones (siembra-desarrollo)	sumIV_max/min
Valor del percentil 95 entre las etapas de siembra y desarrollo	IV_p95

4.3. Obtención de variables climáticas

Las variables asociadas al clima del área de estudio (Tabla 5) se crearon con otra secuencia de código en RStudio, promediando los valores de temperatura media y sumando la precipitación en las etapas comprendidas entre la siembra y el desarrollo, reduciendo los datos a un único valor por temporada. Además de las variables calculadas con el proceso anterior, se definieron una serie de variables específicas con capacidad de explicar gran parte de la producción obtenida.

- *Promedio de temperatura mínima en el período teórico de vernalización* (15 diciembre- 28 febrero). Un invierno frío tiende a aumentar la biomasa de los cultivos, pues la vegetación comienza el desarrollo de sus órganos reproductivos una vez concluido el período de heladas. Además, un período de vernalización prolongado favorece el arraigo de la planta al suelo y el desarrollo de la raíz, mejorando su crecimiento posterior.
- *Promedio de temperaturas máximas en la etapa de desarrollo*. Los valores altos de temperatura en esta fase combinados con eventos de precipitación favorecen el desarrollo y llenado de los granos.
- *Precipitación durante las etapas de siembra y nacimiento* (1 octubre - 25 diciembre). Las lluvias posteriores al período de siembra generan un número mayor de nacimientos, aumentando la densidad de vegetación en los recintos cultivados.

Tabla 9. Variables climáticas.

Métrica	Variable	Código
	Suma en la etapa de siembra	pp_siembra
Precipitación	⋮	⋮
	Suma en la etapa de desarrollo	pp_desarrollo
	Suma en las etapas de siembra y nacimiento	pp_OctDic
T ^a media	Promedio en la etapa de siembra	tMed_siembra
	⋮	⋮
	Promedio en la etapa de desarrollo	tMed_desarrollo
T ^a mínima	Promedio en la fase de vernalización	tMin_vernalización
T ^a máxima	Promedio en la etapa de desarrollo	tMax_desarrollo

4.4. Creación de la base de datos

Una vez obtenidas las métricas espectrales sobre el estado del cultivo y las variables climáticas, se crea la base de datos mediante código en RStudio. Las variables se agrupan con los datos de producción en cada temporada, obteniendo una tabla cuyas observaciones o filas equivalen al número de temporadas analizadas (i.e. 5) y sus variables o columnas con el número de indicadores utilizados (Tablas 8 y 9).

4.5. Creación de los modelos

La estimación de la predicción anual en la explotación se ha realizado utilizando dos tipos de modelos, basados los primeros en la técnica de regresión lineal y los segundos en el método *partial least squares regression* (PLSR). Ambos estiman el valor de la variable dependiente y (i.e. producción) a partir del valor de la variable independiente x , resolviendo una ecuación lineal (3) compuesta por α , valor en el eje de las y de la línea con el mejor ajuste de mínimos cuadrados; β , la pendiente de la línea asociada al valor de la variable independiente x , positiva o negativa en función del signo de la correlación; ε , representa el error aleatorio del modelo [4]. Los modelos se calculan con los datos de las cuatro temporadas anteriores a la 2019/2020, un número adecuado según el estudio de Ren et al., 2008 [58], en el que se comprobó que solo eran necesarios tres años para conseguir resultados aceptables a escala regional.

$$y_i = \alpha + x_1\beta + x_2\beta \cdots + x_k\beta + \varepsilon \quad (3)$$

El cálculo de los modelos de regresión lineal se lleva a cabo con código en R, utilizando la formula `lm()` incluida en el paquete `stats` [57]. Los parámetros α y β se calculan mediante un ajuste de mínimos cuadrados, seleccionando la recta con la menor suma de residuales al cuadrado (i.e. suma de las distancias entre los valores observados y la recta del modelo al cuadrado), asegurando la mejor predicción posible. De entre todas las variables de la base de datos se utilizan aquellas que cumplen con los dos criterios siguientes:

1. Asimilables a una distribución normal, pues los modelos de regresión lineal asumen la normalidad de las variables.
2. Correlacionada¹¹ con la producción utilizando el coeficiente de correlación de Pearson (R_p), con un valor superior a ± 0.8 y nivel de significación al 90 % de confianza (i.e., con un p -value

¹¹La correlación entre dos variables se calcula con el coeficiente de correlación paramétrico de Pearson. Es una medida de su dependencia lineal, independiente de la escala de medida [20].

inferior a 0.1). El umbral escogido como significación se justifica por el reducido número de observaciones, con una capacidad reducida para generar correlaciones a un nivel de confianza del 95 %, el valor comúnmente utilizado [73].

Los modelos se definen con todas las combinaciones posibles entre las variables que cumplen ambas condiciones, representadas en la tabla 10, empleando un máximo de dos variables dependientes. Para que un modelo pueda calcularse debe existir como mínimo un grado de libertad, calculados con la ecuación $n - k - 1$, donde n es el número total de observaciones y k el número de variables predictoras [62]. Dado el número de observaciones utilizadas (i.e. 4), el modelo alcanza un grado de libertad con dos variables predictoras.

Se ha descartado la integración conjunta de dos variables dependientes correlacionadas entre sí, i.e., linealmente dependientes, al suponer una fuente de error asociada al aumento de la multicolinealidad, aportando un exceso de información redundante.

De entre todos los modelos se escogen aquellos con un *p-value* inferior a 0.1, destacando los que presentan un valor por debajo de 0.05. Es el umbral comúnmente utilizado para descartar la hipótesis nula, i.e., ausencia de relación significativa entre las variables dependiente e independientes (si la hubiera sería casual y limitada a la muestra de los datos), y aceptar la existencia de una buena modelización entre las variables [73].

Tabla 10. Variables correlacionadas con la producción mediante un nivel de significación superior al 90 % de confianza y un coeficiente de correlación superior al 80 %. Se dividen entre las calculadas con la *R* de Pearson, utilizadas en el cálculo de los modelos de regresión lineal (derecha); y la *R* de Spearman, utilizadas junto con las anteriores en la definición de los modelos PLSR (izquierda). Se ordenan de forma ascendente por el *p-value*.

Spearman			Pearson		
Código	<i>p-value</i>	<i>R</i>	Código	<i>p-value</i>	<i>R</i>
tMed_ahijamiento	0.02	-1	tMed_ahijamiento	0.03	-0.92
tMin_vernalización	0.02	-1	tMin_vernalizacion	0.03	-0.91
sumNDVI_min	0.08	0.9	pp_OctDic	0.03	0.91
NDVI_p95	0.08	0.9	maxNDVI_ahijamiento	0.04	0.89
tMed_desarrollo	0.08	0.9	maxNDRE_ahijamiento	0.05	0.88
ndvi_desarrollo	0.08	0.9	ireci_ahijamiento	0.06	0.87
ndvi_ahijamiento	0.08	0.9	maxIRECI_ahijamiento	0.06	0.87
ireci_ahijamiento	0.08	0.9	minNDVI_desarrollo	0.07	0.85
ndre_desarrollo	0.08	0.9	pp_nacimiento	0.08	0.83
ndre_ahijamiento	0.08	0.9	tMed_desarrollo	0.08	0.83
sumNDRE_min	0.09	0.8	cr_siembra	0.08	-0.83
			ndvi_ahijamiento	0.09	0.82
			minNDRE_desarrollo	0.09	0.81

El segundo tipo de modelos se ha generado con la técnica PLSR, utilizando el paquete `pls` [44]. Se basa en un modelo de regresión que utiliza los componentes principales derivados de las variables dependientes para resolver la ecuación lineal, existiendo tantas β_x como componentes se hayan integrado en el modelo. De esta forma se consigue reducir la información a varios componentes no correlacionados entre sí, evitando la multicolinealidad y permitiendo la integración conjunta de más de dos variables predictoras sin reducir el grado de libertad del modelo [8, 30]. El número de componentes principales a utilizar se define mediante el método de validación cruzada *leaveone-out cross-validation*, escogiendo el número que minimice el error estándar de la predicción (*Root Mean Squared Error of Prediction*, RMSEP), previniendo el sobreajuste del modelo [44]. Se crean modelos con todas las combinaciones de variables predictoras, seleccionando aquellas que mejor modelan la producción, i.e., con un error más bajo (sección 4.5.1). Este procedimiento, aplicado mediante una secuencia de código en R, es similar al método *interval Partial Least-Squares Regression* (iPLS)[52]. Una vez creados los modelos se obtiene la correlación entre los componentes principales utilizados y

las variables predictoras, seleccionando aquellas con una correlación superior al 0.9 como las variables con un mayor aporte a la creación de los componentes.

4.5.1. Cálculo del error

Los modelos presentan errores derivados de una relación imperfecta entre la variable dependiente e independiente, amplificadas por el ruido generado en la medición de la variable y . La calidad interna de los modelos de regresión lineal se evalúa mediante el cálculo del error estándar de los residuales (RSE, por sus siglas en inglés), un indicador de la desviación estándar de la diferencia entre el valor predicho y el observado en cada una de las observaciones con las que se define el modelo. Se calcula conforme a la ecuación 4, donde SS_r es el valor de la suma de los residuales al cuadrado y df_r son los grados de libertad del modelo.

$$RSE = \sqrt{SS_r/df_r} \quad (4)$$

En los modelos PLSR el error interno se obtiene a partir del *Root Mean Squared Error of Prediction* (RMSEP), calculado mediante la suma de los residuales al cuadrado, y el error *Residual Prediction Deviation* (RPD) [74], desviación típica de los valores observados (i.e. valores de producción durante las cuatro temporadas con las que se entrena el modelo) dividida por el RMSEP. Tanto el RMSEP y el RPD como el RSE dependen de las unidades de medida de la variable y , dificultando la comparación entre modelos con valores de producción en diferentes unidades [20]. Por ello se ha escogido un segundo indicador, el R^2 , interpretado como el porcentaje de la varianza explicado por el modelo. Se calcula dividiendo la varianza de ambos vectores, i.e., el vector con los datos predichos ($x_i\hat{\beta}_i$) y el vector con los datos observados (y_i) (ecuación 5). El modelo será más preciso cuanto mayor sea el porcentaje de la varianza y_i explicado, reduciendo la varianza de los residuales, i.e., porcentaje de varianza no explicado por el modelo. Un modelo capaz de realizar predicciones exactas tendría un R^2 del 100%.

$$R^2 = \frac{var[x_i\hat{\beta}_i]}{var[y_i]} \quad (5)$$

La calidad externa de ambos tipos de modelos se calcula mediante la media del error absoluto o *Mean Absolute Error* (MAE). Es el resultado de restar la producción observada en la temporada a predecir con el dato predicho por el modelo, devolviendo el valor absoluto. Del conjunto de modelos PLSR creados con todas las combinaciones de variables se seleccionan aquellos con el menor RMSEP, MAE y con el mayor R^2 .

5. Resultados

5.1. Relación entre la producción y las variables sobre el estado del cultivo

Los índices de vegetación calculados muestran un patrón general que caracteriza el ciclo de la cebada en la zona de estudio (Figura 15). La respuesta de los IV es baja al principio de las series, coincidiendo con la etapas de siembra y nacimiento, en las que los recintos apenas presentan vegetación y domina la señal de suelo. Comienzan a ascender durante la fase de ahijamiento, donde la aparición de nuevos pies y el crecimiento de los tallos aumenta la biomasa del cultivo. En la fase de desarrollo, entre el 26 de febrero y el 15 de abril, se alcanza el pico de respuesta en las temporadas con mayor producción (2016/17, 2018/19 y 2019/20). En las dos temporadas restantes, con una producción significativamente menor, el momento de máxima respuesta se produce en la fase de senescencia. Este último patrón

se asocia con un menor desarrollo de la vegetación durante las fases de ahijamiento y desarrollo, provocando una menor cantidad de granos y una disminución en el tamaño de la cebada que dificulta su obtención por parte de las cosechadoras, reduciendo la producción final. A partir del valor máximo, los índices comienzan a descender durante la fase de senescencia por la pérdida de humedad y el cambio de coloración en los cultivos, debido a una reducción en el contenido de clorofila. Disminuyen en mayor medida los índices del óptico (i.e. NDVI, NDRE e IRECI), que captan sobre todo el contenido de clorofila, mientras que el CR modera su descenso al detectar las propiedades relacionadas con la geometría de los cultivos, que varían en menor medida. El único cambio en la estructura del dosel vegetal se produce por el encamado de la planta al desarrollar por completo los granos, al comienzo de la etapa de senescencia. Parte de este descenso en el CR se explica también por la reducción en el contenido de humedad de la vegetación.

La desviación típica de los valores aumenta significativamente en los IV del óptico después de la fase de nacimiento, debido a las diferencias en el desarrollo de los recintos provocadas por un desfase en las fechas de nacimiento. El CR presenta una desviación típica elevada durante todo el ciclo de la cebada, sumando a las variaciones en la fecha de nacimiento de los índices del óptico las diferencias entre los recintos al inicio de la serie, con apenas vegetación, como consecuencia de la variabilidad de tipos de suelo y contenido de humedad.

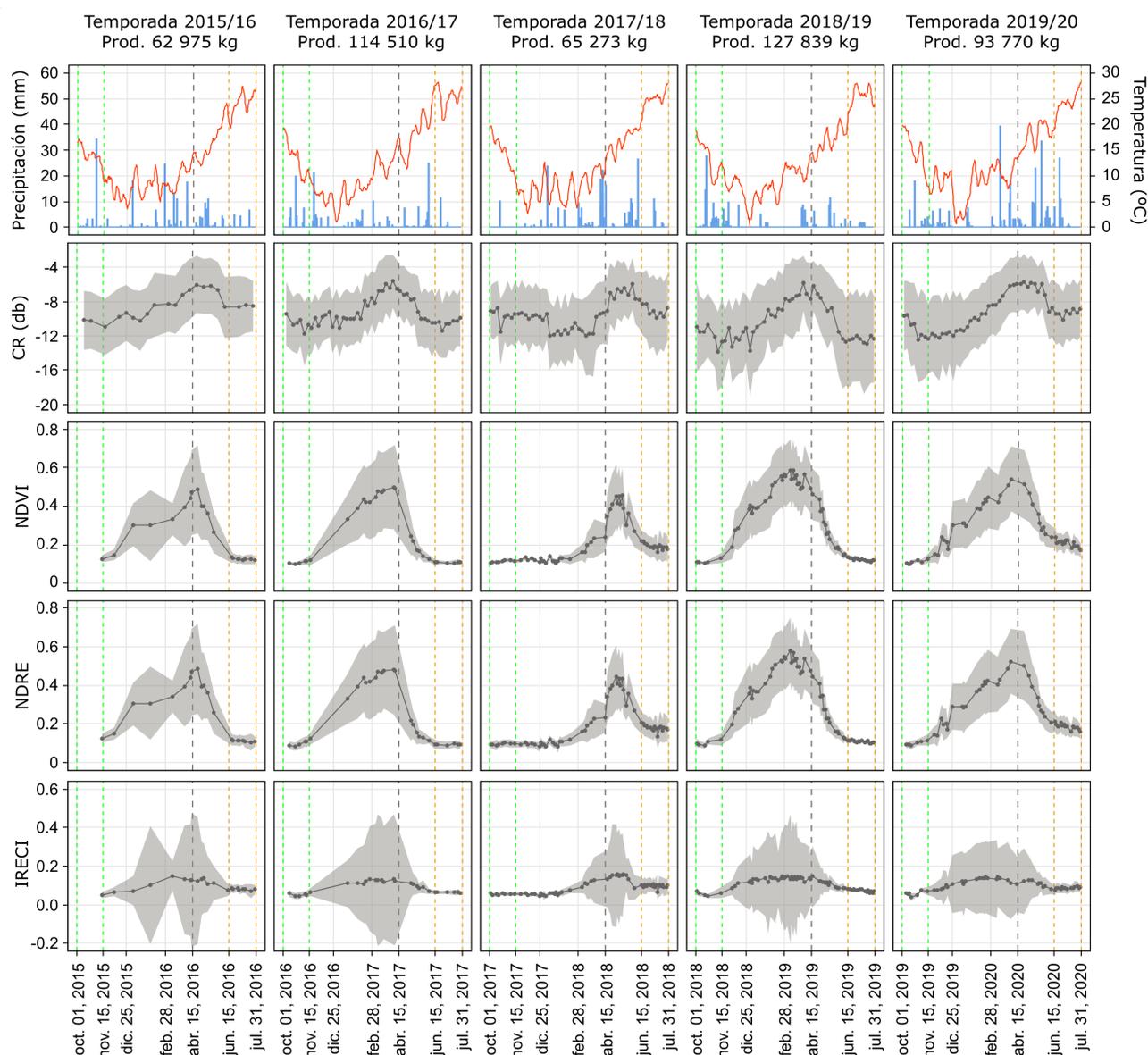


Figura 15. Serie temporal con los índices de vegetación durante las temporadas de cultivo analizadas en relación con la precipitación y la temperatura media diarias. Se muestra la desviación típica de los valores (gris), el momento final de la etapa de desarrollo utilizado como fecha límite de la predicción (15 de abril) y los períodos de siembra (verde) y cosecha (naranja).

El patrón específico de cada IV hace que las variables extraídas sean más precisas estimando la producción en momentos concretos del ciclo fenológico. En la figura 16 se representan las variables sobre el estado de los cultivos (eje de abscisas) con mejor correlación respecto de la producción (eje de ordenadas). Los valores medios de CR, a pesar de contar con una variabilidad temporal mayor que los IV del óptico¹², están correlacionados con la producción en la fase de siembra, cuyo promedio disminuye al aumentar la cantidad de producción. Durante esta etapa, las medidas de CR disminuyen al incrementarse el número de brotes de cebada nacidos antes de comenzar la fase de nacimiento, relacionado con una precipitación elevada durante el mes de octubre¹³. Se aprecia claramente durante la

¹²La elevada frecuencia temporal del CR a lo largo de toda la serie puede asociarse con frecuentes cambios en la disposición geométrica de los cultivos ocasionados por el cierzo y situaciones sinópticas convulsas [64]

¹³Es posible que los valores altos en la precipitación contribuyan a disminuir los valores del CR, reduciendo el ajuste de la correlación entre este y los valores de producción.

temporada 2017/18, donde el descenso brusco en los valores medios de CR entre la fase de nacimiento (-9dB) y ahijamiento (-12dB) marca la aparición de los primeros nacimientos, retrasados hasta comienzos de enero por la falta de precipitación en el inicio de la temporada. Sin embargo, la sensibilidad del CR en la etapa de siembra no logra captar las diferencias entre las dos temporadas con mayor y menor producción, invirtiendo sus valores, i.e., la temporada con menor producción (2015/16) presenta una media mayor que la temporada con el segundo menor registro de producción (2017/18), y lo mismo sucede entre las dos temporadas con mayor producción.

Tanto el NDVI como el NDRE presentan series temporales prácticamente idénticas. En las temporadas con menor producción, sufren una estabilización durante las fases de ahijamiento y desarrollo. Por el contrario, las temporadas con mayor producción presentan una tendencia ascendente, aumentando sus valores progresivamente desde la etapa de siembra hasta comienzos de la etapa de desarrollo. Esta diferencia hace que los valores medios de NDRE y NDVI en las fases de ahijamiento y desarrollo modelen con gran acierto la producción en las temporadas estudiadas. Sin embargo, la correlación de las variables de NDRE y NDVI en la etapa de ahijamiento con la producción muestra una separación entre las dos temporadas con menor producción (2015/16 y 2017/18), mucho mayor en el eje de abscisas que en el eje de ordenadas, indicando una mala adecuación de esta variable respecto a las temporadas con menor producción. Además invierte sus valores, i.e., el dato más bajo de producción no se corresponde con los valores más bajos de NDVI y NDRE en las fases de ahijamiento y desarrollo. Con el máximo valor de NDVI en la etapa de ahijamiento se consigue reducir la distancia entre las temporadas con menor producción. Sin embargo, este último empeora la relación entre las tres temporadas con mayor producción, provocando que el valor de la temporada 2016/17 (i.e. 114 510 kg) esté por debajo de la 2019/20 en el eje de abscisas, con menor producción (i.e. 93 770 kg). El resto de variables formadas con el NDVI y el NDRE modelan razonablemente estas tres temporadas, a excepción del percentil 95 de NDVI, con una situación similar al valor máximo de NDVI en la etapa de ahijamiento.

Los valores de IRECI presentan patrones mucho más estables que el resto de IV, sin un pico de respuesta destacado. Las cinco temporadas comienzan con una ligera subida del IRECI entre la etapa de siembra y la de ahijamiento, a partir de la cual se mantienen estables hasta la fase de senescencia. Es el aumento producido en la fase de ahijamiento el que mejor clasifica los valores de producción, con un promedio de IRECI en esta fase correlacionado positivamente con la producción. Sin embargo, el valor de la temporada a predecir se sitúa mucho más próximo de los datos de producción más altos en el eje de abscisas, una situación similar al percentil 95 de NDVI. El máximo valor de IRECI en la fase de ahijamiento representa las temporadas con valores más bajos de la misma forma que los promedios, pero empeora la relación de las tres temporadas con mayor producción, sobrepasando la tercera temporada a la segunda con el mejor dato en el eje de abscisas.

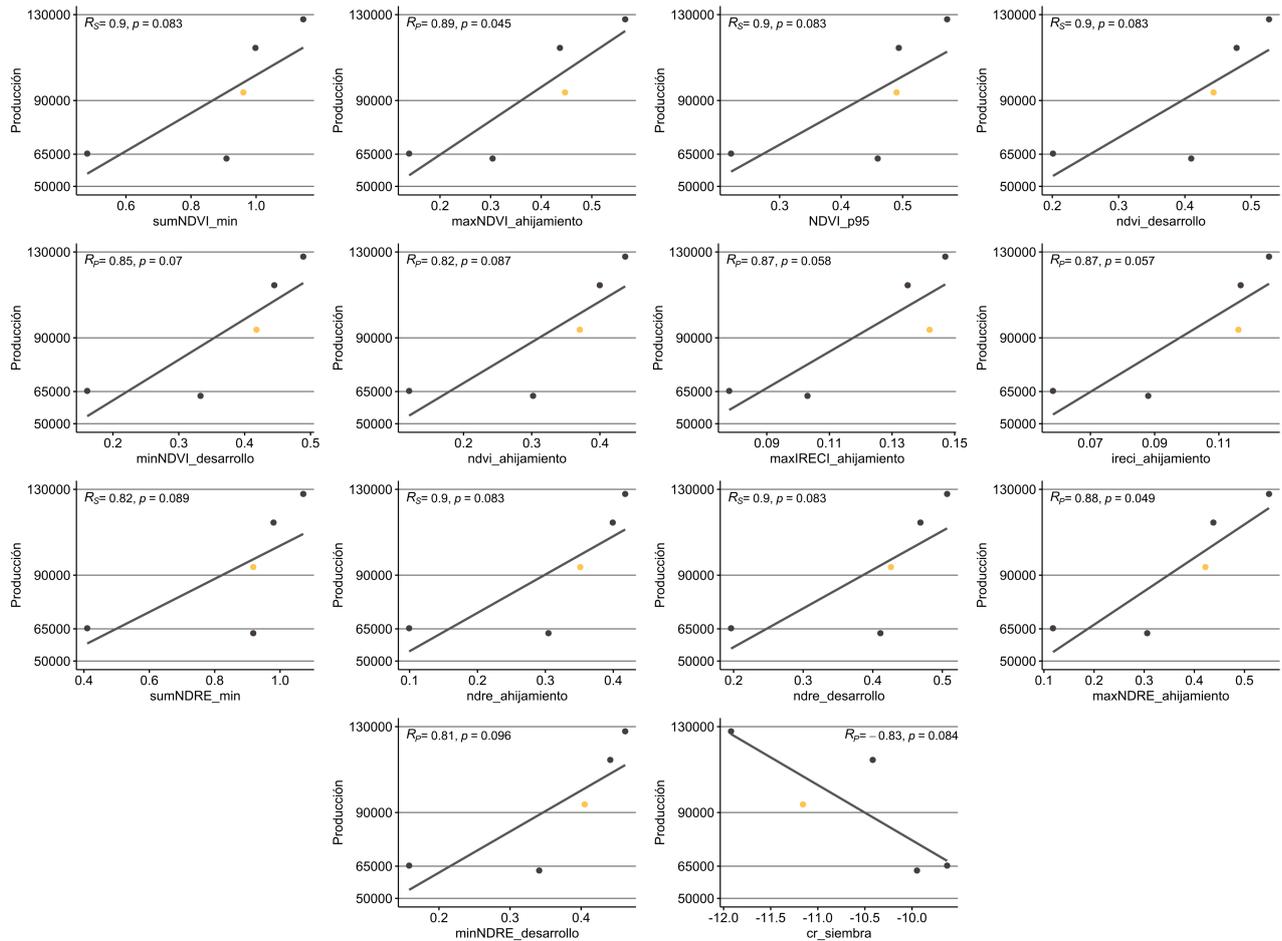


Figura 16. Relación entre las variables sobre el estado de la vegetación mejor correlacionadas con la producción y los valores de esta última durante las cinco temporadas analizadas. R_p : coeficiente de correlación de Pearson, R_s : coeficiente de correlación de Spearman. En amarillo se marca el dato de producción de la temporada a predecir.

5.2. Relación entre la producción y las variables climáticas

Las variables climáticas que mejor modelan la producción son la temperatura media de la fase de ahijamiento y el promedio de temperaturas mínimas en el período de vernalización (Figura 17). Ambas reflejan un aumento en la producción a medida que desciende la temperatura. Las temporadas con mayor producción se representan mejor con respecto a la temperatura media en la fase de ahijamiento, mientras que el promedio de las temperaturas mínimas en el período de vernalización modela adecuadamente las dos temporadas con menor producción, más próximas entre sí en el eje de abscisas. La temperatura media en la etapa de desarrollo es la que consigue una mejor representación de las temporadas con menor producción, donde un ligero incremento en la temperatura media provoca un ligero incremento en la cantidad de granos, pero empeora la correlación final al invertir la relación entre las dos temporadas con mayor producción. Las variables elaboradas con la precipitación, i.e., la suma en la etapa de nacimiento y en las etapas de siembra y nacimiento, polarizan los datos de producción entre las temporadas de mayor y menor producción, alterando además el orden de sus valores, coincidiendo la mayor cantidad de precipitación con la menor cantidad de producción en cada uno de los dos grupos (i.e. las tres temporadas con mayor producción por un lado y las dos con menor valor por otro).

La evolución de la temperatura media y la precipitación durante las series analizadas explican

algunos de los problemas que presentan los IV en la modelización de la producción final durante las dos temporadas con menor producción (2015/16 con 62 975 kg y 2017/18 con 65 273 kg), representando el valor más bajo con una mayor respuesta de los IV. Durante la temporada 2017/18, la escasa precipitación durante las etapas de siembra y nacimiento retrasó los nacimientos hasta la fase de ahijamiento, explicando los valores bajos durante las dos primeras etapas en todos los IV respecto a la temporada 2015/16. El aumento de la precipitación durante la fase de desarrollo en la 2017/18, 90 mm respecto a los 59.9 mm de la 2015/16, supuso que el 55 % de la precipitación recibida desde el momento de la siembra hasta el fin de la etapa de desarrollo se diera en esta fase. Este hecho, sumado a una mayor temperatura media en las etapas de ahijamiento y desarrollo, permitió un aumento del crecimiento de la vegetación posterior al umbral utilizado en la estimación (15 de abril), obteniendo una producción que superó a la temporada 2015/16. El problema observado en la correlación de los IV con la producción de la segunda y tercera temporada, i.e., 2016/17 con 114 510 kg y 2019/20 con 93 770 kg, se relaciona con el volumen de precipitación desde la siembra hasta el final de la etapa de desarrollo. La temporada 2016/17 recibe 155 mm mientras que la 2019/20 recibe 251 mm, favoreciendo un aumento en la cantidad de biomasa del cultivo durante esta última. Sin embargo, este incremento no se tradujo en una mayor cantidad de granos, reduciendo la relación entre la respuesta de los IV y la producción. Probablemente se produjo como consecuencia de una mayor temperatura durante la etapa de ahijamiento respecto a la 2016/17 (10.1 °C en 2020 y 8.07 °C en 2017 entre el 28 de enero y el 28 de febrero), que adelantó el proceso de reproducción y aumentó el riesgo de un deterioro en las flores como consecuencia de las heladas tardías. El incremento de las lluvias en el momento de la cosecha, muy superior al de las anteriores campañas, pudo empeorar la calidad de los granos reduciendo su peso específico y con ello la cantidad de producción.

En general, los IV agrupados según las estaciones fenológicas de la cebada modelan mejor las temporadas con mayor producción, mientras que las variables climáticas logran recoger las variaciones en las dos temporadas con menor producción.

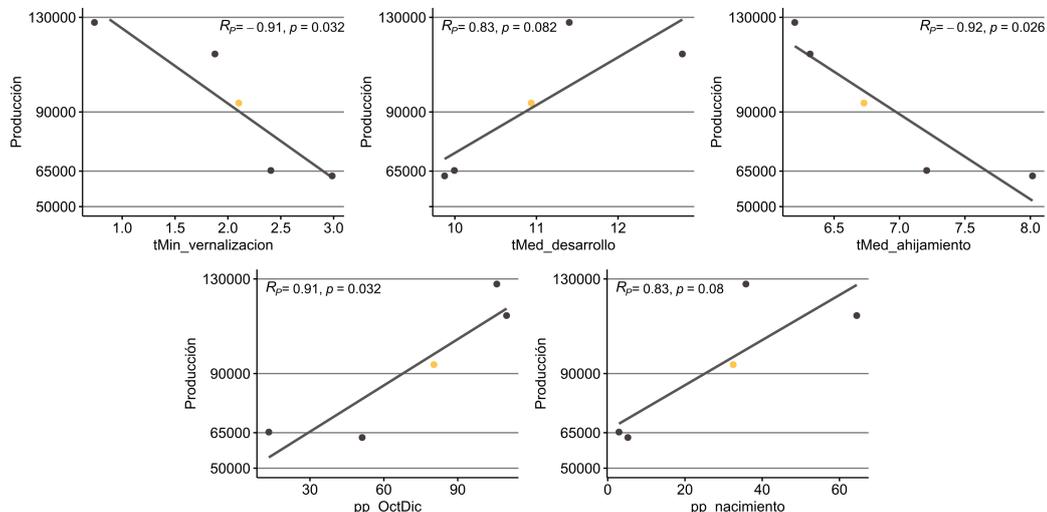


Figura 17. Relación entre las variables climáticas mejor correlacionadas con la producción y los valores de esta última durante las cinco temporadas analizadas. R_p : coeficiente de correlación calculado con Pearson. En amarillo se marca el dato de producción de la temporada a predecir.

5.3. Modelos de predicción

Del conjunto de modelos creados mediante regresión lineal con un p -value inferior a 0.1, mostrados en la tabla 11, el menor MAE al estimar la producción de 2020 (4.62 %) se obtiene con los promedios de temperatura mínima en el período de vernalización. Su RSE es el tercer valor más alto, 16 719 kg,

y la varianza explicada por el modelo es del 83 %. Los siguientes dos modelos se definen también con una única variable climática, la precipitación en las etapas de siembra y nacimiento y la temperatura media en la fase de ahijamiento. El MAE de ambos modelos asciende hasta los 6.03 % y 6.83 % respectivamente, descendiendo el RSE con respecto al primero. El valor de p -value en estos tres modelos supera el 0.05, indicando una menor probabilidad de que su precisión no sea consecuencia de un sobreajuste del modelo a los datos, i.e. relación casual entre las variables dependiente e independientes.

Al integrar los valores máximos de NDRE durante la etapa de ahijamiento más la temperatura media en esta misma fase el RSE disminuye hasta los 455 kg, el menor de todos los modelos. Se consigue también el p -value más bajo (0.008) y el R^2 más alto, siendo el modelo que más porcentaje de varianza es capaz de explicar (99 %). Por el contrario, el error alcanza el 10.62 % (146.39 kg/ha), superior a los tres primeros modelos. Este descenso en los valores de RSE junto al aumento del R^2 se observa en todos los modelos creados con la combinación de una variable climática y otra derivada de un IV, superando en todos ellos el 97 % de R^2 y manteniendo el valor de RSE por debajo de los 5 700 kg. Esta combinación se da también en los tres únicos modelos con un p -value inferior al 0.05.

Tabla 11. Resultado de los modelos de regresión lineal, ordenados de menor a mayor MAE. En color gris se marcan los modelos con un p -value inferior a 0.05. *Ahij.* abreviatura de ahijamiento.

Variables predictoras	p -value	R^2	RSE (kg)	MAE		
				kg	kg/ha	%
tMin_vernalizacion	0.087	0.83	16 719	4 329	67.64	4.62
pp_OctDic	0.086	0.84	16 600	5 650	83.09	6.03
tMed_ahij.	0.073	0.86	15 327	6 404	94.17	6.83
maxNDRE_ahij. + tMed_ahij.	0.008	0.99	455	9 955	146.39	10.62
minNDRE_desarrollo + tMed_ahij.	0.098	0.97	5 669	10 142	149.15	10.82
ndvi_ahij. + tMed_ahij.	0.078	0.98	4 490	10 201	150.01	10.88
minNDVI_desarrollo + tMed_ahij.	0.070	0.99	4 060	10 578	155.56	11.28
maxNDVI_ahij. + tMed_ahij.	0.019	0.99	1 074	11 572	170.17	12.34
cr_siembra + tMed_desarrollo	0.089	0.98	5 171	11 656	171.41	12.43
maxNDVI_ahijamiento	0.092	0.83	17 104	13 063	192.10	13.93
ireci_ahij. + tMed_ahij.	0.071	0.98	4 112	14 080	207.05	15.02
ireci_ahij.	0.094	0.82	17 348	17 634	259.32	18.81
ndvi_ahij. + minNDVI_desarrollo	0.092	0.97	5 346	18 258	268.49	19.47
maxIRECI_ahij. + tMed_ahij.	0.048	0.99	2 786	18 563	272.98	19.80

Los tres modelos PLSR seleccionados se han generado con un componente principal. El modelo PLSR con el error absoluto (MAE) más bajo (un 1.35 %) se ha obtenido con un componente principal que integra información de las dos variables climáticas mejor correlacionadas con la producción, i.e., los promedios de temperatura mínima y media en el período de vernalización y la etapa de ahijamiento respectivamente (Tabla 12). Incorpora también información sobre el estado de los cultivos a través de cuatro variables formadas con el NDRE y el NDVI, con una correlación del 90 % respecto de la producción final excepto la suma de los mínimos de NDRE entre la siembra y la etapa de desarrollo (Figura 16). Sin embargo, las que más información aportan son las dos variables climáticas, correlacionadas más de un 90 % con el componente principal. El error de este modelo reduce en 3 062 kg la cifra de error más baja obtenida con los modelos de regresión lineal. A pesar de explicar un 87 % de la varianza y de contar con un RPD del 2.8, sobrepasando el umbral establecido para considerar válido un modelo (de 2 según [6]), su RMSEP supera los 10 000 kg (11 912 kg), situación similar a la de los modelos de regresión lineal creados con una única variable, todos ellos con un p -value superior a 0.05. Este hecho, junto al reducido error externo del modelo, podría estar relacionado con un sobreajuste del modelo a los datos, impidiendo su uso fuera de la muestra de datos.

Del conjunto de modelos seleccionados, el de mayor varianza explicada ($0.89 R^2$) aumenta solo 95 kg el MAE del modelo anterior (1 362 respecto a los 1 267 del primero). Integra las dos variables climáticas de este último, ambas con el mayor aporte al componente principal utilizado. Del resto de

variables, relacionadas con el estado de los cultivos, solo conserva la suma de los valores mínimos de NDVI por temporadas (siembra - desarrollo), la única con una correlación respecto a la producción final del 90 %. El resto presentan niveles de correlación inferiores, pero consiguen una mejor diferenciación entre la temporada 2019/20 (93 770 kg) y la 2017/18 (65 273 kg) que las variables sobre el estado del cultivo integradas en el primer modelo. El RMSEP de este modelo es de 10 129 kg, demasiado elevado como para poder descartar un sobreajuste del modelo, con 3.3 de RPD.

El modelo PLSR con un RMSEP más bajo (8 481 kg) es claramente inferior al resto, con un RPD de 3.94, siendo los mejores datos de error interno, reduciendo el sobreajuste del modelo a los datos. Sin embargo, aumenta el error de la predicción (7.26 % de MAE), muy por encima de los anteriores modelos PLSR, y disminuye la varianza explicada (0.82 R²). La única variable climática en este modelo es la temperatura media en la etapa de ahijamiento, presentando la mayor relación respecto al componente principal utilizado (-0.97). Integra además el conjunto de variables creadas con los IV mejor correlacionados con la producción (Figura 16) excepto la suma de los valores mínimos de NDRE y el promedio de CR en la siembra.

Los mejores modelos de entre todos los anteriores (tanto los elaborados con regresión lineal simple como con la técnica PLSR) se consideran aquellos con un menor error interno, i.e., un menor sobre ajuste respecto a los datos con los que se entrenan. Son estos los que con mayor probabilidad podrían estimar, con buenos resultados, la producción en la zona de estudio.

Tabla 12. Modelos de regresión PLSR con el menor RMSEP (1°), el mayor R² (2°) y el menor MAE (3°). Entre paréntesis se señalan las correlaciones de las variables predictoras con el componente principal utilizado superiores a ± 0.90 . *Ahij.* abreviatura de ahijamiento, *des.* de desarrollo y *vern.* de vernalización.

Variables predictoras	N°	R ²	RPD	RMSEP (kg)	MAE		
					kg	kg/ha	%
tMed_ahij. (-0.97) NDVI_p95 maxNDVI_ahij. minNDVI_des. ndvi_ahij. ndvi_des. sumNDVI_min	13	0.82	3.94	8 481	6 811	106.42	7.26
tMin_vern. (-0.97) maxNDRE_ahij. minNDRE_des. ndvi_ahij.	8	0.89	3.3	10 129	1 362	21.28	1.45
tMin_vern. (-0.98) sumNDVI_min ndvi_des.	6	0.87	2.8	11 912	1 267	19.80	1.35

6. Discusión

El modelo de regresión lineal con mejores resultados es el definido con los valores máximos de NDRE y la temperatura media en la etapa de ahijamiento. El componente principal que devuelve el mejor modelo PLSR está integrado por la temperatura media en la etapa de ahijamiento y el conjunto de variables creadas con información espectral mejor correlacionadas con la producción (Tabla 10), a excepción del CR en la etapa de siembra y la suma de los valores mínimos de NDRE.

El modelo PLSR presenta el menor MAE, 7.26 % respecto a los 10.62 % del modelo de regresión lineal. Sin embargo, es este último el que cuenta con el menor error interno (455 kg de RSE) y el que más varianza explica (R² de 0.99). Aunque las diferencias entre el NDVI y el NDRE son mínimas, son las medidas de NDRE las que devuelven el modelo de regresión lineal más preciso. Por ello, la

integración del *red edge* podría suponer una ligera ventaja respecto de los trabajos anteriores que aplican el NDVI. Son las variables de estos dos índices (NDVI y NDRE) las que componen la práctica totalidad de las variables integradas en el modelo PLSR, a excepción de la temperatura media y dos variables obtenidas con valores de IRECI. Las etapas del ciclo fenológico más representadas en los modelos seleccionados son las fases de ahijamiento y desarrollo, observado también por Nuarsa *et al.* [51]. Las variaciones en la etapa de ahijamiento marcan el desarrollo del cultivo, pues integra la fase de floración, mientras que la fase de desarrollo se corresponde con el momento de máximo vigor vegetal, captada por los IV en el espectro óptico y cuyas variaciones se relacionan con la producción final. Por otro lado, las métricas derivadas del cociente VH/VV apenas aportan información relevante, a excepción de los promedios en la etapa de siembra. Es debido a la gran variabilidad que presentan a lo largo de las series temporales.

La comparación entre el error del mejor modelo de regresión lineal con los estudios anteriores que utilizaron la misma técnica debe realizarse teniendo en cuenta las diferencias en el área de estudio. El error obtenido por Ren *et al.* [58] fue de 214.16 kg/ha sobre una superficie de 13 millones de hectáreas, mientras que el error del modelo en este trabajo es de 146.39 kg/ha sobre una superficie entorno a las 64 ha. Por otro lado, el MAE en porcentaje es superior al modelo de Becker-Reshef *et al.* [2], calculado sobre 21 millones de hectáreas (Kansas, EEUU), con un 7%, por debajo del 10.62% obtenido en este trabajo. La mejora del error en [2] puede asociarse al mayor número de años con los que se definen los modelos, i.e., 7 años, respecto a los 3 años de [58] y los 4 años de este trabajo. Un segundo factor podría estar relacionado con el mayor número de imágenes disponibles en [2], caracterizando mejor el ciclo fenológico de los cultivos. Por este motivo podría estudiarse la integración de imágenes Landsat que aumentara las imágenes con las que realizar la predicción. Por otro lado, el coeficiente de regresión (R^2) en el trabajo de Becker-Reshef *et al.* es de 0.94, por debajo del 0.99 obtenido en este trabajo. Podría estar relacionado con la creación de variables adaptadas al ciclo fenológico de la cebada y la integración de variables climáticas.

El modelo creado por Mirasi *et al.*, definido sobre un área de 1 200 ha, cifra más próxima a la escala de este trabajo, obtuvo un error de 146 kg/ha, ligeramente inferior al error del modelo de regresión lineal definido en este trabajo (146.39 kg/ha). A raíz de estos resultados no puede afirmarse que haya una mejora significativa en la estimación de producciones utilizando modelos de regresión lineal con información adaptada al ciclo fenológico de la cebada. Sin embargo, utilizar variables específicas en función del cultivo y el aumento de la escala de trabajo podrían haber mejorado el error interno son respecto a los estudios anteriores, permitiendo asociar los cambios en la producción anual de la cebada con la variabilidad temporal y geográfica, minimizando la influencia de factores como los precios de mercado y las políticas agrícolas llevadas a cabo por los gobiernos [25]. Por el contrario, aumentar la escala de trabajo limita la aplicación del modelo y las variables utilizadas a explotaciones de la zona de estudio y sobre cultivos de trigo y cebada, similares en su comportamiento espectral y fenología.

El error obtenido con el modelo PLSR (i.e. 7.26%) es menor al obtenido por Chen y Jing [7] (i.e. 12.84%), calculado mediante la relación de información espectral procedente de Landsat 8 OLI sobre 28 zonas de entrenamiento con cultivos de trigo representativos de un área de 99 000 ha. En este estudio se concluye además que los métodos de regresión no lineal son más adecuados en la estimación de la producción de un cultivo, e.g., *artificial neural networks* (ANN). Sin embargo, los modelos PLSR presentan errores internos muy superiores a los modelos de regresión lineal con un *p-value* inferior a 0.05.

Con respecto a la integración conjunta de las variables sobre el estado de la vegetación procedentes de la región de las microondas y el óptico, no ha podido concluirse que suponga una mejora en la predicción. Ningún modelo estudiado integra variables en los dos rangos del espectro, y el único modelo de regresión lineal con un *p-value* inferior a 0.1 con presencia del CR (i.e. promedios del CR en la etapa de siembra más la temperatura media en la fase de desarrollo) devuelve un RSE de 5 171 kg y un MAE del 12.3%.

Considerando el valor de producción de la temporada a predecir (93 770 kg), entre las dos temporadas con menor y mayor producción (62 975 kg y 65 273 kg respecto a 114 510 kg y 127 839 kg), podría afirmarse que los modelos son sensibles a las variaciones de producción dentro de la serie analizada. Sin embargo, debería validarse en un mayor número de temporadas para descartar que los resultados no son fruto de una coincidencia en el ajuste de la recta de regresión, algo que de momento no parece probable dados los bajos valores de error interno en los modelos seleccionados.

7. Conclusiones

Los buenos resultados obtenidos con los modelos de regresión lineal y el menor número de variables utilizadas en su creación los sitúan como la mejor opción para estimar la producción de cebada. A pesar de obtener un error externo mayor al del modelo PLSR, el error interno de la regresión lineal es ocho toneladas más bajo que el error de este último.

Combinar la información sobre el estado de la vegetación y el estado del clima ha permitido integrar en el modelo las variaciones introducidas por el sistema de rotación de cultivos con el que se gestiona la explotación. Las variables adaptadas a la fenología del cultivo son las que mejor modelan la producción final, especialmente la temperatura media durante las fases de vernalización, ahijamiento y desarrollo.

Los resultados obtenidos sitúan a las constelaciones de Sentinel 1 y Sentinel 2 como una buena opción para obtener métricas sobre el estado de los cultivos a gran escala. Su continuidad más allá del 2030 permite utilizar sus medidas en la creación de modelos operacionales que, combinados con las técnicas de procesamiento de datos empleadas en el desarrollo de este trabajo, e.g., la plataforma de GEE, pueden crearse de forma eficiente y a bajo coste. Debido a la reducida serie de años, el modelo debería validarse con más datos para poder confirmar estas conclusiones. Una posible línea de trabajo en esta dirección es utilizar la información de estas constelaciones junto al sensor OLI de Landsat 8, estudiando la posible mejora provocada por el aumento de las imágenes satelitales.

Bibliografía

- [1] Ameline, M., Fieuzal, R., Betbeder, J., Berthoumieu, J. F. y Baup, F.: *Estimation of Corn Yield by Assimilating SAR and Optical Time Series Into a Simplified Agro-Meteorological Model: From Diagnostic to Forecast*. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 11(12):4747–4760, Dic. 2018, ISSN 2151-1535. Conference Name: IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing.
- [2] Becker-Reshef, I., Vermote, E., Lindeman, M. y Justice, C.: *A generalized regression-based model for forecasting winter wheat yields in Kansas and Ukraine using MODIS data*. Remote Sensing of Environment, 114(6):1312–1323, Jun. 2010, ISSN 0034-4257. <http://www.sciencedirect.com/science/article/pii/S0034425710000325>.
- [3] Birrell, S. J., Sudduth, K. A. y Borgelt, S. C.: *Comparison of sensors and techniques for crop yield mapping*. Computers and Electronics in Agriculture, 14(2):215–233, Feb. 1996, ISSN 0168-1699. <http://www.sciencedirect.com/science/article/pii/0168169995000496>.
- [4] Bottenberg, R. A. y Ward, J. H.: *Applied multiple linear regression*, vol. 63. 6570th Personnel Research Laboratory, Aerospace Medical Division, Lackland Air Force Base, Texas, 1963. <https://apps.dtic.mil/sti/citations/AD0413128>.
- [5] Bouman, B. A. M. y Kasteren, H. W. J. van: *Ground-based X-band (3-cm wave) radar backscattering of agricultural crops. II. Wheat, barley, and oats; the impact of canopy structure*. Remote Sensing of Environment, 34(2):107–119, Nov. 1990, ISSN 0034-4257. <http://www.sciencedirect.com/science/article/pii/003442579090102R>.
- [6] Chang, C.W., Laird, D.A., Mausbach, M.J. y Hurburgh, C.R.: *Near-Infrared Reflectance Spectroscopy–Principal Components Regression Analyses of Soil Properties*. Soil Science Society of America Journal, 65(2):480–490, 2001, ISSN 1435-0661. <https://access.onlinelibrary.wiley.com/doi/abs/10.2136/sssaj2001.652480x>, eprint: <https://access.onlinelibrary.wiley.com/doi/pdf/10.2136/sssaj2001.652480x>.
- [7] Chen, P. y Jing, Q.: *A comparison of two adaptive multivariate analysis methods (PLSR and ANN) for winter wheat yield forecasting using Landsat-8 OLI images*. Advances in Space Research, 59(4):987–995, Feb. 2017, ISSN 0273-1177. <http://www.sciencedirect.com/science/article/pii/S027311771630672X>.
- [8] Chlingaryan, A., Sukkarieh, S. y Whelan, B.: *Machine learning approaches for crop yield prediction and nitrogen status estimation in precision agriculture: A review*. Computers and Electronics in Agriculture, 151:61–69, Ago. 2018, ISSN 0168-1699. <http://www.sciencedirect.com/science/article/pii/S0168169917314710>.
- [9] Chuvieco, E.: *Teledetección ambiental : la observación de la Tierra desde el Espacio*. Ariel, Barcelona, 3ra ed., 2008.
- [10] Clemente Suárez, V.: *Análisis de mecanismos de dispersión polarimétricos para el estudio y caracterización de cultivos agrícolas*. Tesis de Doctorado, Universitat Politècnica de Catalunya, Barcelona, Feb. 2012. <https://upcommons.upc.edu/handle/2099.1/14173>, Accepted: 2012-02-13T10:24:22Z Publisher: Universitat Politècnica de Catalunya.
- [11] Clevers, J. y Van Leeuwen, H. J. C.: *Combined use of optical and microwave remote sensing data for crop growth monitoring*. Remote sensing of environment, 56(1):42–51, 1996. Publisher: Elsevier.

- [12] Collecte Localisation Satellites (CLS): *Sentinel-1 Product Definition*. Informe técnico S1-RS-MDA-52-7440, 2016. https://sentinel.esa.int/web/sentinel/user-guides/sentinel-1-sar/document-library/-/asset_publisher/1d07RF5fJMbd/content/sentinel-1-product-definition.
- [13] De Zan, F. y Monti Guarnieri, A.: *TOPSAR: Terrain Observation by Progressive Scans*. IEEE Transactions on Geoscience and Remote Sensing, 44(9):2352–2360, Sep. 2006, ISSN 1558-0644. <https://ieeexplore.ieee.org/document/1677745>, Conference Name: IEEE Transactions on Geoscience and Remote Sensing.
- [14] Dempewolf, J., Adusei, B., Becker-Reshef, I., Hansen, M., Potapov, P., Khan, A. y Barker, B.: *Wheat Yield Forecasting for Punjab Province from Vegetation Index Time Series and Historic Crop Statistics*. Remote Sensing, 6:9653–9675, Oct. 2014. https://www.researchgate.net/publication/337434713_Regional_Wheat_Yield_Estimation_by_Integration_of_Remotely_Sensed_Soil_Moisture_into_a_Crop_Model.
- [15] ESA: *Sentinel-1 Toolbox*. <https://sentinel.esa.int/web/sentinel/toolboxes/sentinel-1>, visitado el 2020-09-25.
- [16] European Space Agency: *Satellite Earth Observations In Support Of The Sustainable Development Goals (special 2018 edition)*. Informe técnico., 2018. <https://www.impel.eu/earth-observation-handbook-special-edition-2018/>.
- [17] FAO: *How to feed the world in 2050*. Informe técnico., Office of the Director, Agricultural Development Economics Division, Economic and Social Development Department., Roma, 2009.
- [18] FAO: *Food and Agriculture Organization of the United Nations, Land Resources*., 2020. <http://www.fao.org/faostat/en/#data/QC>, visitado el 2020-08-04.
- [19] Foreign Agricultural Service: *PSD Online*, 2019. <https://apps.fas.usda.gov/psdonline/app/index.html#/app/advQuery>, visitado el 2020-09-01.
- [20] Forsyth, D.: *Looking at Relationships*. En Forsyth, D. (ed.): *Probability and Statistics for Computer Science*, págs. 29–50. Springer International Publishing, Cham, Switzerland, 2018, ISBN 978-3-319-64410-3. https://doi.org/10.1007/978-3-319-64410-3_2.
- [21] Frampton, W. J., Dash, J., Watmough, G. y Milton, E. J.: *Evaluating the capabilities of Sentinel-2 for quantitative estimation of biophysical variables in vegetation*. ISPRS Journal of Photogrammetry and Remote Sensing, 82:83–92, Ago. 2013, ISSN 0924-2716. <http://www.sciencedirect.com/science/article/pii/S092427161300107X>.
- [22] Fung, A. K.: *Scattering from a Vegetation Layer*. IEEE Transactions on Geoscience Electronics, 17(1):1–6, Ene. 1979, ISSN 0018-9413. <https://ieeexplore.ieee.org/abstract/document/4071955/>, Conference Name: IEEE Transactions on Geoscience Electronics.
- [23] Gibson, P. y Power, C. H.: *Introductory remote sensing: Digital image processing and applications*. Psychology Press, Routledge, London, 2000.
- [24] Gilabert, M. A., González-Piqueras, J. y García-Haro, J.: *Acerca de los índices de vegetación*. Revista de la Asociación Española de Teledetección, 8(1):1–10, 1997, ISSN 1133-0953. <https://www.researchgate.net/publication/39195330>.
- [25] Gitelson, A. A., Viña, A., Verma, S. B., Rundquist, D. C., Arkebauer, T. J., Keydan, G., Leavitt, B., Ciganda, V., Burba, G. G. y Suyker, A. E.: *Relationship between gross primary production and chlorophyll content in crops: Implications for the*

- synoptic monitoring of vegetation productivity*. Journal of Geophysical Research: Atmospheres, 111(D08S11), 2006, ISSN 2156-2202. <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2005JD006017%4010.1002/%28ISSN%292169-8996.EXCHANGE1>, eprint: <https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2005JD006017>.
- [26] Gobierno de Aragón: *IDEAragon: Infraestructura de Datos Espaciales de Aragon*, 2020. <https://idearagon.aragon.es/portal/>, visitado el 2020-11-13.
- [27] Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D. y Moore, R.: *Google Earth Engine: Planetary-scale geospatial analysis for everyone*. Remote Sensing of Environment, 202:18–27, Dic. 2017, ISSN 0034-4257. <http://www.sciencedirect.com/science/article/pii/S0034425717302900>.
- [28] Grupo para la evaluación de nuevas variedades de cereales en España (GENVCE): *Ficha descriptiva de la variedad Icaria*. Informe técnico., 2011. <https://genvce.org/wp-content/uploads/2019/12/Icaria.pdf>.
- [29] Guyot, G. y Baret, F.: *Utilisation de la Haute Resolution Spectrale pour Suivre L'état des Couverts Vegetaux*. Spectral Signatures of Objects in Remote Sensing, 287:279, Abr. 1988, ISSN 1609-042X. <https://ui.adsabs.harvard.edu/abs/1988ESASP.287..279G/abstract>.
- [30] Hansen, P. y Schjoerring, J.: *Reflectance measurement of canopy biomass and nitrogen status in wheat crops using normalized difference vegetation indices and partial least squares regression*. Remote Sensing of Environment, 86(4):542–553, Ago. 2003, ISSN 00344257. <https://linkinghub.elsevier.com/retrieve/pii/S0034425703001317>.
- [31] Holben, B. N., Eck, T. F., Slutsker, I., Tanré, D., Buis, J. P., Setzer, A., Vermote, E., Reagan, J. A., Kaufman, Y. J., Nakajima, T., Lavenu, F., Jankowiak, I. y Smirnov, A.: *AERONET—A Federated Instrument Network and Data Archive for Aerosol Characterization*. Remote Sensing of Environment, 66(1):1–16, Oct. 1998, ISSN 0034-4257. <http://www.sciencedirect.com/science/article/pii/S0034425798000315>.
- [32] Idso, S. B., Pinter, P. J., Hatfield, J. L., Jackson, R. D. y Reginato, R. J.: *A remote sensing model for the prediction of wheat yields prior to harvest*. Journal of Theoretical Biology, 77(2):217–228, Mar. 1979, ISSN 0022-5193. <http://www.sciencedirect.com/science/article/pii/0022519379903096>.
- [33] Igartua Arregui, E., Cuesta-Marcos, A., Lasa Dolhagaray, J.M., Gracia Gimeno, M.P., Yahiaoui, S., Molina-Cano, J.L., Ciudad, F.J. y Casas Cendoya, A.M.: *Fenología y adaptación de la cebada*. En *La adaptación al ambiente y los estreses abióticos en la mejora vegetal*, págs. 159–182. Junta de Andalucía, Dirección General de Planificación y Análisis de Mercados, Servicio de Publicaciones y Divulgación., Sevilla, 2008, ISBN 978-84-8474-248-7. <https://digital.csic.es/handle/10261/80585>, Accepted: 2013-08-08T10:57:47Z.
- [34] Karthikeyan, L., Chawla, I. y Mishra, A. K.: *A review of remote sensing applications in agriculture for food security: Crop growth and yield, irrigation, and crop losses*. Journal of Hydrology, 586:124905, Jul. 2020, ISSN 0022-1694. <http://www.sciencedirect.com/science/article/pii/S0022169420303656>.
- [35] Kasampalis, D. A., Alexandridis, T. K., Deva, C., Challinor, A., Moshou, D. y Zalidis, G.: *Contribution of Remote Sensing on Crop Models: A Review*. Journal of Imaging, 4(4):52, Abr. 2018. <https://www.mdpi.com/2313-433X/4/4/52>, Number: 4 Publisher: Multidisciplinary Digital Publishing Institute.

- [36] Kaufman, Y.J.: *Aerosol optical thickness and atmospheric path radiance*. Journal of Geophysical Research: Atmospheres, 98(D2):2677–2692, 1993, ISSN 2156-2202. <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/92JD02427>, eprint: <https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/92JD02427>.
- [37] Knipling, E.B.: *Physical and physiological basis for the reflectance of visible and near-infrared radiation from vegetation*. Remote Sensing of Environment, 1(3):155–159, Jun. 1970, ISSN 0034-4257. <http://www.sciencedirect.com/science/article/pii/S0034425770800219>.
- [38] Lobell, D.B.: *The use of satellite data for crop yield gap analysis*. Field Crops Research, 143:56–64, Mar. 2013, ISSN 0378-4290. <http://www.sciencedirect.com/science/article/pii/S0378429012002754>.
- [39] LONGARES, L.: *El paisaje vegetal en el entorno de la reserva ornitológica "El Planerón" (Belchite-Zaragoza)*. Serie Investigación. 1997, ISBN 84-920441-8-7.
- [40] Macdonald, R. y Hall, F.: *Global Crop Forecasting*. Science (New York, N.Y.), 208:670–9, Jun. 1980.
- [41] McNairn, H., Champagne, C., Shang, J., Holmstrom, D. y Reichert, G.: *Integration of optical and Synthetic Aperture Radar (SAR) imagery for delivering operational annual crop inventories*. ISPRS Journal of Photogrammetry and Remote Sensing, 64(5):434–449, Sep. 2009, ISSN 0924-2716. <http://www.sciencedirect.com/science/article/pii/S0924271608000944>.
- [42] McNairn, H. y Shang, J.: *A Review of Multitemporal Synthetic Aperture Radar (SAR) for Crop Monitoring*. En Ban, Y. (ed.): *Multitemporal Remote Sensing: Methods and Applications*, Remote Sensing and Digital Image Processing, págs. 317–340. Springer International Publishing, Cham, 2016, ISBN 978-3-319-47037-5. https://doi.org/10.1007/978-3-319-47037-5_15.
- [43] Mercier, A., Betbeder, J., Baudry, J., Denize, J., Leroux, V., Roger, J.L., Spicher, F. y Hubert-Moy, L.: *Evaluation of Sentinel-1 and -2 time series to derive crop phenology and biomass of wheat and rapeseed: northern France and Brittany case studies*. En *Remote Sensing for Agriculture, Ecosystems, and Hydrology XXI*, vol. 11149 de *Proceedings of SPIE*, pág. 1114903, Strasbourg, France, Oct. 2019. International Society for Optics and Photonics, ISBN 978-1-5106-3002-4. <https://www.spiedigitallibrary.org/conference-proceedings-of-spie/11149/1114903/Evaluation-of-Sentinel-1-and--2-time-series-to/10.1117/12.2533132.short>.
- [44] Mevik, B.H., Wehrens, R. y Liland, K.H.: *pls: Partial Least Squares and Principal Component Regression*. 2020. <https://CRAN.R-project.org/package=pls>, R package versión 2.7-3.
- [45] Ministerio de Agricultura, Pesca y Alimentación (MAPA): *Cultivos herbáceos e industriales: Cereales*, 2019. <https://www.mapa.gob.es/es/agricultura/temas/producciones-agricolas/cultivos-herbaceos/cereales/>, visitado el 2020-08-09.
- [46] Mirasi, A., Mahmoudi, A., Navid, H., Kamran, K. V. y Asoodar, M. A.: *Evaluation of sum-NDVI values to estimate wheat grain yields using multi-temporal Landsat OLI data*. Geocarto International, págs. 1–16, Jul. 2019, ISSN 1010-6049. <https://doi.org/10.1080/10106049.2019.1641561>, Publisher: Taylor & Francis eprint: <https://doi.org/10.1080/10106049.2019.1641561>.
- [47] Monteith, J. L., Moss, C. J., Cooke, G. W., Pirie, N. W. y Bell, G. D. H.: *Climate and the efficiency of crop production in Britain*. Philosophical Transactions of the Royal Society of London. B, Biological Sciences, 281(980):277–294, Nov. 1977. <https://royalsocietypublishing.org/doi/10.1098/rstb.1977.0140>, Publisher: Royal Society.

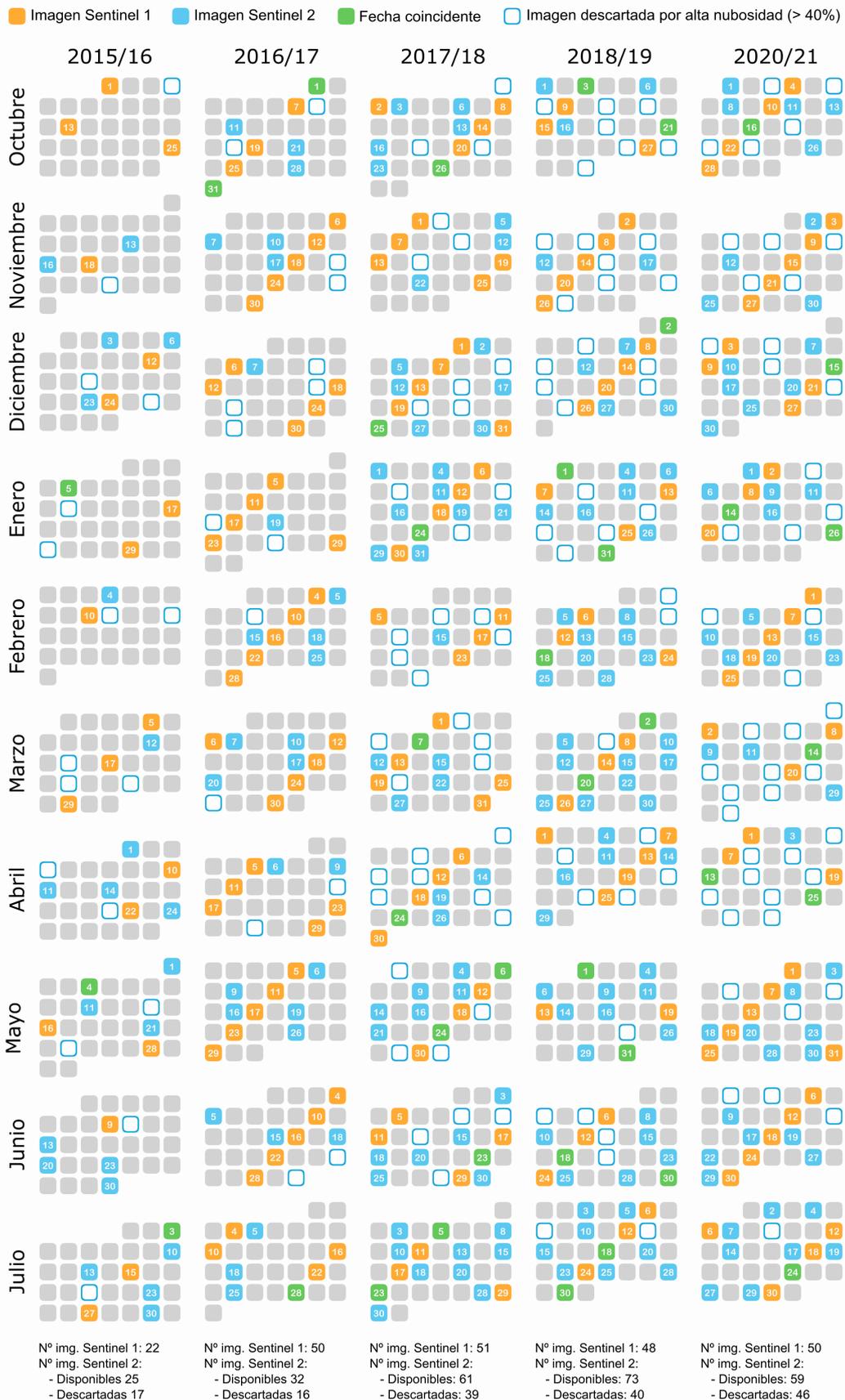
- [48] Murphy, S.: *Atmospheric Correction of a (Single) Sentinel 2 Image*. Repositorio de Github. <https://github.com/samsammurphy/gee-atmcorr-S2>.
- [49] Nakicenovic, N. y Swart, R.: *Emissions Scenarios: A Special Report of Working Group III of the Intergovernmental Panel on Climate Change*. Informe técnico., Cambridge University Press, Cambridge, 2000.
- [50] Newton, A. C., Flavell, A. J., George, T. S., Leat, P., Mullholland, B., Ramsay, L., Revoredo-Giha, C., Russell, J., Steffenson, B. J., Swanston, J. S., Thomas, W. T. B., Waugh, R., White, P. J. y Bingham, I. J.: *Crops that feed the world 4. Barley: a resilient crop? Strengths and weaknesses in the context of food security*. Food Security, 3(2):141, Jun. 2011, ISSN 1876-4525. <https://doi.org/10.1007/s12571-011-0126-3>.
- [51] Nuarsa, I., Nishio, F. y Hongo, C.: *Rice yield estimation using Landsat ETM data and field observation*. Journal of Agricultural Science, 4(3):45–56, 2012, ISSN 1916-9752. <http://ccsenet.org/journal/index.php/jas/article/view/11300/9631>.
- [52] Nørgaard, L., Saudland, A., Wagner, J., Nielsen, J. P., Munck, L. y Engelsen, S. B.: *Interval Partial Least-Squares Regression (iPLS): A Comparative Chemometric Study with an Example from Near-Infrared Spectroscopy*. Applied Spectroscopy, 54(3):413–419, Mar. 2000, ISSN 0003-7028. <https://doi.org/10.1366/0003702001949500>, Publisher: SAGE Publications Ltd STM.
- [53] Pinter, P. J. P., Jackson, R. D., Idso, S. B. y Reginato, R. J.: *Multidate spectral reflectance as predictors of yield in water stressed wheat and barley*. International Journal of Remote Sensing, 2(1):43–48, Ene. 1981, ISSN 0143-1161. <https://doi.org/10.1080/01431168108948339>, Publisher: Taylor & Francis _eprint: <https://doi.org/10.1080/01431168108948339>.
- [54] Porta, J., López-Acevedo, M. y Roquero, C.: *Edafología para la agricultura y el medio ambiente*. Ediciones Mundi-Prensa, Madrid, 3a ed., 2003, ISBN 978-1-4492-8686-6.
- [55] Prince, S. D.: *A model of regional primary production for use with coarse resolution satellite data*. International Journal of Remote Sensing, 12(6):1313–1330, Jun. 1991, ISSN 0143-1161. <https://doi.org/10.1080/01431169108929728>, Publisher: Taylor & Francis _eprint: <https://doi.org/10.1080/01431169108929728>.
- [56] QGIS.org: *QGIS Geographic Information System*. Open Source Geospatial Foundation, 2020. <http://qgis.org>, Versión 3.12.3.
- [57] R Core Team: *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2020. <https://www.R-project.org/>, R versión 4.0.2 (2020-06-22).
- [58] Ren, J., Chen, Z., Zhou, Q. y Tang, H.: *Regional yield estimation for winter wheat with MODIS-NDVI data in Shandong, China*. International Journal of Applied Earth Observation and Geoinformation, 10(4):403–413, Dic. 2008, ISSN 0303-2434. <http://www.sciencedirect.com/science/article/pii/S030324340700075X>.
- [59] Rouse, Jr., J. W., Deering, D. W., Haas, R. H., Schell, J. A. y Harlan, J.: *Monitoring the Vernal Advancement and Retrogradation (Green Wave Effect) of Natural Vegetation*. En *NASA/GSFC*, pág. 390, Greenbelt, 1974. <https://ntrs.nasa.gov/citations/19750020419>, Place: United States, North America.
- [60] Schlund, M. y Erasmi, S.: *Sentinel-1 time series data for monitoring the phenology of winter wheat*. Remote Sensing of Environment, 246:111814, Sep. 2020, ISSN 0034-4257. <http://www.sciencedirect.com/science/article/pii/S003442572030184X>.

- [61] Schmitt, M., Hughes, L.H., Qiu, C. y Zhu, X.X.: *Aggregating cloud-free Sentinel-2 images with Google Earth Engine*. En *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. IV-2-W7, págs. 145–152, Munich, Germany, Sep. 2019. Copernicus GmbH. <https://www.isprs-ann-photogramm-remote-sens-spatial-inf-sci.net/IV-2-W7/145/2019/>, ISSN: 2194-9042.
- [62] Schumacker, R. y Tomek, S.: *Linear Regression*. En Schumacker, R. y Tomek, S. (eds.): *Understanding Statistics Using R*, págs. 219–228. Springer, New York, NY, 2013, ISBN 978-1-4614-6227-9. https://doi.org/10.1007/978-1-4614-6227-9_13.
- [63] Seelan, S. K., Laguette, S., Casady, G. M. y Seielstad, G. A.: *Remote sensing applications for precision agriculture: A learning community approach*. *Remote Sensing of Environment*, 88(1):157–169, Nov. 2003, ISSN 0034-4257. <http://www.sciencedirect.com/science/article/pii/S0034425703002360>.
- [64] Sonobe, R.: *Combining ASAR-2 XSAR HH and Sentinel-1 C-SAR VH/VV Polarization Data for Improved Crop Mapping*. *Remote Sensing*, 11(16):1920, Ene. 2019. <https://www.mdpi.com/2072-4292/11/16/1920>, Number: 16 Publisher: Multidisciplinary Digital Publishing Institute.
- [65] Steele-Dunne, S. C., McNairn, H., Monsivais-Huertero, A., Judge, J., Liu, P. W. y Papathanassiou, K.: *Radar Remote Sensing of Agricultural Canopies: A Review*. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 10(5):2249–2273, May. 2017, ISSN 2151-1535. Conference Name: IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing.
- [66] Tomiyasu, K.: *Tutorial review of synthetic-aperture radar (SAR) with applications to imaging of the ocean surface*. *Proceedings of the IEEE*, 66(5):563–583, May. 1978, ISSN 1558-2256. Conference Name: Proceedings of the IEEE.
- [67] Tucker, C.J.: *Red and photographic infrared linear combinations for monitoring vegetation*. *Remote Sensing of Environment*, 8(2):127–150, May. 1979, ISSN 0034-4257. <http://www.sciencedirect.com/science/article/pii/0034425779900130>.
- [68] Ulaby, F. T., Bradley, G. A. y Dobson, M. C.: *Microwave Backscatter Dependence on Surface Roughness, Soil Moisture, and Soil Texture: Part II-Vegetation-Covered Soil*. *IEEE Transactions on Geoscience Electronics*, 17(2):33–40, Abr. 1979, ISSN 0018-9413. Conference Name: IEEE Transactions on Geoscience Electronics.
- [69] Veloso, A., Mermoz, S., Bouvet, A., Le Toan, T., Planells, M., Dejoux, J.F. y Ceschia, E.: *Understanding the temporal behavior of crops using Sentinel-1 and Sentinel-2-like data for agricultural applications*. *Remote Sensing of Environment*, 199:415–426, Sep. 2017, ISSN 0034-4257. <http://www.sciencedirect.com/science/article/pii/S0034425717303309>.
- [70] Vermote, E., Tanre, D., Deuze, J., Herman, M. y Morcette, J. J.: *Second Simulation of the Satellite Signal in the Solar Spectrum, 6S: an overview*. *IEEE Transactions on Geoscience and Remote Sensing*, 35(3):675–686, May. 1997, ISSN 1558-0644. <https://ieeexplore.ieee.org/abstract/document/581987>, Conference Name: IEEE Transactions on Geoscience and Remote Sensing.
- [71] Vreugdenhil, M., Wagner, W., Bauer-Marschallinger, B., Pfeil, I., Teubner, I., Rüdiger, C. y Strauss, P.: *Sensitivity of Sentinel-1 Backscatter to Vegetation Dynamics: An Austrian Case Study*. *Remote Sensing*, 10(9):1396, Sep. 2018. <https://www.mdpi.com/2072-4292/10/9/1396>, Number: 9 Publisher: Multidisciplinary Digital Publishing Institute.
- [72] Wall, L., Larocque, D. y Léger, P.M.: *The early explanatory power of NDVI in crop yield modelling*. *International Journal of Remote Sensing*, 29(8):2211–2225, Abr. 2008, ISSN 0143-1161. <https://doi.org/10.1080/01431160701395252>, Publisher: Taylor & Francis. eprint: <https://doi.org/10.1080/01431160701395252>.

- [73] Wasserman, L.: *Models, Statistical Inference and Learning*. En Wasserman, L. (ed.): *All of Statistics: A Concise Course in Statistical Inference*, Springer Texts in Statistics, págs. 87–96. Springer, New York, NY, 2004, ISBN 978-0-387-21736-9. https://doi.org/10.1007/978-0-387-21736-9_6.
- [74] Williams, P. C. y Sobering, D.: *Comparison of Commercial near Infrared Transmittance and Reflectance Instruments for Analysis of Whole Grains and Seeds*. *Journal of Near Infrared Spectroscopy*, 1(1):25–32, Ene. 1993, ISSN 09670335.
- [75] Wilson, R. T.: *Py6S: A Python interface to the 6S radiative transfer model*. *Comput. Geosci. - UK*, 51(2):166–171, 2013. https://www.rtwilson.com/academic/Wilson_2012_Py6S_Paper.pdf.
- [76] Zhao, Y., Potgieter, A. B., Zhang, M., Wu, B. y Hammer, G. L.: *Predicting Wheat Yield at the Field Scale by Combining High-Resolution Sentinel-2 Satellite Imagery and Crop Modelling*. *Remote Sensing*, 12(6):1024, Ene. 2020. <https://www.mdpi.com/2072-4292/12/6/1024>, Number: 6 Publisher: Multidisciplinary Digital Publishing Institute.

Anexos

I. Fechas de las imágenes Sentinel 1 y Sentinel 2



II. Función para la extracción de métricas sobre el estado de la vegetación, Sentinel 1

Acceder al código dentro de Google Earth Engine:

<https://code.earthengine.google.com/792aa6a0117cf9caf08d450bdbb97e99?noload=true>.

```
/*
=== OBTENCION DE MÉTRICAS SOBRE EL ESTADO DE UN CULTIVO ===
      SENTINEL 1
*/

// ----- CARGAR LA CAPA CON LOS RECINTOS ----- \\
var AOI = ee.FeatureCollection(
  'users/iranzocristian/explotacion_blcht_buffer20m');

// ----- FILTRAR COLECCIÓN ----- \\

// Cargar la colección Sentinel-1

var sentinel1 = ee.ImageCollection('COPERNICUS/S1_GRD')
  // Obtener imágenes con polarización VV y VH
  .filter(ee.Filter.listContains(
    'transmitterReceiverPolarisation', 'VV'))
  .filter(ee.Filter.listContains(
    'transmitterReceiverPolarisation', 'VH'))

  // Filtrar imágenes obtenidas en modo
  // 'Interferometric Wide Swath'
  .filter(ee.Filter.eq('instrumentMode', 'IW'))

  // Filtrar imágenes de alta resolución
  .filter(ee.Filter.eq('resolution', 'H'));

// Dividir la colección en función de la trayectoria

// Se selecciona la trayectoria descendente,
// presenta mejores ángulos de incidencia para analizar
// cultivos
var s1Descending = sentinel1.filter(
  ee.Filter.eq('orbitProperties_pass', 'DESCENDING'));

// .....| Cálculo del cociente VH/VV |.....\\
```

```

// Funciones para aplicar la transformación de la inversa
// del logaritmo (de dB a valores sigma lineales)

// Transformación de la banda VH

var s1vh_invlog = function(img){

    // Expresión
    var vh_log = img.expression(
        'pow(10,(vh/10))',
        {
            vh: img.select('VH') // Banda con polarizacion VH
        }).toDouble();

    // Resultado
    return vh_log.select([0],['vh'])
    // Seleccionar las propiedades a incluir en la nueva imagen
        .copyProperties(img, ['system: index', 'resolution_meters',
            'totalSlices', 'productType', 'orbitProperties_pass'
            , 'system:time_start']);
};

// Transformación de la banda VV

var s1vv_invlog = function(img){

    // Expresión
    var vv_log = img.expression(
        'pow(10,(vv/10))',
        {
            vv: img.select('VV') // Banda con polarizacion VV
        }).toDouble();

    // Resultado
    return vv_log.select([0],['vv'])
    // Seleccionar las propiedades a incluir en la nueva imagen
        .copyProperties(img, ['system: index', 'resolution_meters',
            'totalSlices', 'productType', 'orbitProperties_pass'
            , 'system:time_start']);
};

// Funcion que calcula el cociente VH/VV
// Las bandas seleccionadas son las creadas en las funciones anteriores
var s1Index = function(img){

```

```

// Expresion
var index = img.expression(
  'vh / vv',
  {
    vh: img.select('vh'),    // Banda con polarizacion VH
    vv: img.select('vv'),    // Banda con polarizacion VV
  }).toDouble();

// Resultado
return index.select([0], ['vh/vv'])
  .copyProperties(img, ['system: index', 'resolution_meters',
    'totalSlices', 'productType', 'orbitProperties_pass'
    , 'system:time_start']);
};

// Funcion que transforma el resultado del cociente a dB

var s1Trans = function(img){
// Expresion
var index = img.expression(
  '10 * log10(cr)',
  {
    cr: img.select('vh/vv'),    // Banda con el cociente
  }).toDouble();

// Resultado
return index.select([0], ['vh/vv_db'])
  .copyProperties(img, ['system: index',
    'resolution_meters', 'totalSlices', 'productType',
    'orbitProperties_pass', 'system:time_start']);
};

// ----- INICIO DE LA FUNCIÓN ----- \\
var S1serietemporal = function(coleccion,
                              a_inicio,mes_inicio,dia_inicio,
                              a_final,mes_final,dia_final){

  // Establecer los años de la serie
  var yearrangeStart = a_inicio;
  var yearrangeStop = a_final;

  // Generar la coleccion con todas las imágenes

```

```

var col;

// Funcion para recortar la coleccion
var clip = function(image){return image.clip(cultivo)};
// La variable "cultivo" se define más adelante,
// integra las parcelas cultivadas en el ciclo seleccionado

// .....| Cálculo de las series temporales |.....\

// Loop a través de los años
// Crear una serie de cada temporada de cultivo
for(var loopYear = yearrangeStart; loopYear < yearrangeStop; loopYear +=1){
  // Seleccionar la temporada de producción con la que seleccionar recintos
  var prod_temp = loopYear += 1;
  // Se suma un año, pues los valores de producción utilizados en el
  // filtro de parcelas están en el año siguiente al comienzo de la temporada

  // Se vuelve al año inicial de la temporada
  loopYear -= 1;

  // Seleccionar los campos sembrados con cebada en la temporada
  var cultivo = AOI.filter(ee.Filter.gt(prod_temp.toString(), 1));

  // Seleccionar las fechas de filtrado (de temporada en temporada)
  var start = ee.Date.fromYMD(loopYear, mes_inicio, dia_inicio);
  var end = ee.Date.fromYMD(loopYear +=1, mes_final, dia_final);

  // Filtrar la colección por fecha y recortar la colección con
  // los recintos anteriores
  var imgClip = coleccion.filterDate(start, end).map(clip);

  // CALCULO DEL COCIENTE VH/VV
  // Aplicar la inversa del logaritmo a cada banda
  var imgClip_vh = imgClip.map(s1vh_invlog);
  var imgClip_vv = imgClip.map(s1vv_invlog);

  // Combinar las dos colecciones anteriores en una sola
  var combine = imgClip_vh.combine(imgClip_vv);

  // Calcular el cociente VH/VV
  var imgCInv = combine.map(s1Index);

  // Transformar a dB
  var imgCIndB = imgCInv.map(s1Trans);

```

```

// Volver al año de inicio de la temporada
loopYear -= 1;
// Condicional IF:
// Si el bucle trabaja con el primer año de la serie crea la colección
if(loopYear == a_inicio){
    col = imgCIndB;
// El resto de años incluirá las imágenes en la colección anterior
} else if(loopYear > a_inicio && loopYear < a_final){
    col = col.merge(imgCIndB);
}
} // Fin del bucle FOR

// Crear el grafico
var chart = ui.Chart.image.series({
    imageCollection: col,
    region: AOI,
// Calcular la media de cada imagen
    reducer: ee.Reducer.median(),
    scale: 10
});
// Nombrarlo en función de los años de la serie temporal
var filename = ("des_VHVV_sum_").concat(a_inicio.toString()).concat("-")
    .concat(a_final.toString());
print(chart, filename);
};

// Aplicar la función anterior a la colección
S1serietemporal(s1Descending,2015,10,1,2020,7,31);

```

III. Función para la extracción de métricas sobre el estado de la vegetación, Sentinel 2

Acceder al código dentro de Google Earth Engine:

<https://code.earthengine.google.com/f5cf008cedae18035c7ec29e9c2daac3>.

```
/*
=== OBTENCION DE MÉTRICAS SOBRE EL ESTADO DE UN CULTIVO ===
      SENTINEL 2
*/

// ----- MÁSCARA DE NUBES Y SOMBRAS----- \\
// Parámetros generales
var max_probN = 65; // umbral a partir del cuál un píxel es considerado nube
var max_probS = 0.02; // umbral a partir del cuál se considera sombra (tantos por uno)
var ndviThresh = -0.1; // Umbral de sombras en función del NDVI
var irSumThresh = 0.3; // Umbral de sombras en función de las bandas infrarrojas

// Parámetros de extensión de la máscara sobre píxeles adyacentes
var erodePixels = 1.5;
var dilationPixels = 3;

// Alturas medias de las nubes
// Utilizadas en la proyección de sus sombras
var cloudHeights = ee.List.sequence(200,10000,250);

function dilatedErosion(score) {
// Se aplica a la capa con la probabilidad de sombras
  score = score
    .reproject('EPSG:4326', null, 20)
    .focal_min({radius: erodePixels,
      kernelType: 'circle', iterations:3})
    .focal_max({radius: dilationPixels,
      kernelType: 'circle', iterations:3})
    .reproject('EPSG:4326', null, 20);

  return(score);
}

// .....| Cálculo de la probabilidad de sombra |.....\\
// .....|          dentro de un píxel          |.....\\
```

```

// Función que proyecta las sombras de las nubes
function projectShadows(image){
  // Ángulos de iluminación
  var meanAzimuth = image.get('MEAN_SOLAR_AZIMUTH_ANGLE');
  var meanZenith = image.get('MEAN_SOLAR_ZENITH_ANGLE');

  // Banda con la probabilidad de nubes, procedente de la colección
  // "COPERNICUS/S2_CLOUD_PROBABILITY", utilizada para localizar
  // los píxeles con nubes
  var cloudMask = ee.Image(image.get('cloud_mask'))
    .select('probability').gt(max_probN);

  // Localizar píxeles oscuros (región infrarroja)
  var darkPixelsImg = image.select(['B8', 'B11', 'B12'])
    .reduce(ee.Reducer.sum());

  // Localizar píxeles oscuros asociados con agua
  var ndvi = image.normalizedDifference(['B8', 'B4']);
  var waterMask = ndvi.lt(ndviThresh);

  // Obtener píxeles de sombra
  var darkPixels = darkPixelsImg.lt(irSumThresh);

  // Crear la máscara de sombras excluyendo el agua
  var darkPixelMask = darkPixels.and(waterMask.not());
  darkPixelMask = darkPixelMask.and(cloudMask.not());

  // Localizar las sombras de las nubes, basado en la
  // geometría de iluminación (convertida en radianes)
  var azR = ee.Number(meanAzimuth).add(180).multiply(Math.PI)
    .divide(180.0);
  var zenR = ee.Number(meanZenith).multiply(Math.PI)
    .divide(180.0);

  // Localizar las sombras de las nubes
  var shadows = cloudHeights.map(function(cloudHeight){
    cloudHeight = ee.Number(cloudHeight);

    var shadowCastedDistance = zenR.tan()
      .multiply(cloudHeight); // Clasificación según altura
    var x = azR.sin().multiply(shadowCastedDistance)
      .multiply(-1); // distancia de sombras, coord. X
    var y = azR.cos().multiply(shadowCastedDistance)

```

```

        .multiply(-1); // distancia de sombras, coord. Y

    // Aplica una máscara sobre los píxeles de sombra
    // desplazando los valores de la máscara de nubes sobre sus sombras
    return ee.Image(image.get('cloud_mask'))
        .select('probability').displace(ee.Image.constant(x)
        .addBands(ee.Image.constant(y)));
});

// Agrupar las máscaras de sombras creadas en la función anterior
// dentro de una colección
var shadowMasks = ee.ImageCollection.fromImages(shadows);
var shadowMask = shadowMasks.mean();

// Crear la máscara de sombras
shadowMask = dilatedErosion(shadowMask.multiply(darkPixelMask));

// Crear la nueva banda con los píxeles sobre sombras
var shadowScore = shadowMask.reduceNeighborhood(
    {
        reducer: ee.Reducer.max(),
        kernel: ee.Kernel.square(1)
    });

image = image.addBands(shadowScore.rename(['shadowScore']));

return image;
}

// ----- INPUTS ----- \\
// .....| Recintos de la explotación |.....\\
var AOI = ee.FeatureCollection(
    'users/iranzocristian/explotacion_blcht_buffer20m');

// .....| Colección |.....\\

// UNIR COLECCIÓN DE GEE Y COLECCIÓN CORREGIDA CON GS

// Cargar la colección Sentinel 2 (nivel 2A) de GEE
var sentinel2 = ee.ImageCollection("COPERNICUS/S2_SR")
    // Filtrar por área de estudio
    .filterBounds(AOI)

    // Filtrar por fecha (comienzo al acabar la serie de

```

```

// imágenes corregidas con 6S - S2_6s)
.filterDate('2017-03-21', '2020-07-31')

// Primer filtro de nubes
.filter(ee.Filter.lt('CLOUDY_PIXEL_PERCENTAGE', 40))

// Convertir valores a reflectividad, conservando
// las propiedades de la imagen original
.map(function(img){
  var properties = img.propertyNames();
  return(img.divide(10000)
    .copyProperties(img, properties)) })

// Ordenar según la fecha de adquisición
.sort('system:time_start');

// AÑADIR LAS IMÁGENES CON EL PORCENTAJE DE NUBES
// Cargar colección con el porcentaje de nubes
var inicio = ee.Date('2015-10-01');
var final = ee.Date('2020-07-31');

var S2Clouds = ee.ImageCollection(
  'COPERNICUS/S2_CLOUD_PROBABILITY')
  .filterBounds(AOI).filterDate(inicio,final);

// Unir la imagen con el porcentaje de nubes a cada imagen de la
// colección S2 nivel 2A
var sentinel2n = ee.Join.saveFirst('cloud_mask').apply({
  primary: sentinel2,
  secondary: S2Clouds,
  condition:
    ee.Filter.equals({leftField: 'system:index', rightField: 'system:index'})
});

// Cargar la colección Sentinel-2 corregida
// Ubicación de las imágenes
var asset = 'users/iranzocristian/s2_6s/';

// Matriz con el número de imágenes
var image = ee.data.listAssets(asset);

// Seleccionar las imágenes del elemento anterior
var imageList = image['assets'];

```

```

// Obtener el número de imágenes
var num_img = imageList.length;
// Restarle uno, el iterador comenzará en el 0
num_img = num_img - 1;

// Crear una nueva lista con el id de las imágenes
// 1. Definir una lista en blanco
var imageIDs = [];
// 2. Crear iterador
var i;

// 3. Bucle for: integrar en la lista vacía cada uno
// de los id's de las imágenes contenidas en la carpeta,
// variable imageList

for(i=0;i <= num_img; i++){
    // guardar el id
    var imageid = imageList[i].id;
    // incluirlo en la lista
    imageIDs.push(ee.Image(imageid.toString()));
}

// Crear la colección de imágenes a partir de la lista anterior
var S2_6s = ee.ImageCollection.fromImages(imageIDs);

// AÑADIR EL PORCENTAJE DE NUBES
// Unir la imagen con el porcentaje de nubes a cada imagen de la
// colección
var S2_6sn = ee.Join.saveFirst('cloud_mask').apply({
    primary: S2_6s,
    secondary: S2Clouds,
    condition:
        ee.Filter.equals({leftField: 'fileID', rightField: 'system:index'})
});

// UNIR LAS DOS COLECCIONES
var S2_SR = ee.ImageCollection(S2_6sn.merge(sentinel2n));

// .....| APLICAR LA MÁSCARA DE NUBES |.....\\
// Función que crea la máscara de nubes, en función del umbral
// establecido previamente {max_probN}
var aplicar_maskN = function(img){
    // Crea la máscara con la banda del porcentaje de nubes dentro de la
    // imagen unida a ambas colecciones en los pasos previos

```

```

    var mask = ee.Image(img.get('cloud_mask'))
        .select('probability').gt(max_probN).eq(0);
    // Devuelve la imagen con la máscara aplicada
    return img.updateMask(mask);
};

// Aplicar máscara de nubes a la colección
var S2_filtroNubes = S2_SR.map(aplicar_maskN);

// .....| APLICAR LA MÁSCARA DE SOMBRAS |.....\
// Obtener la probabilidad de sombra en cada píxel
var probSombras = S2_filtroNubes.map(projectShadows);

// Obtener máscara de sombras, umbral definido previamente {max_probS}
var aplicar_masks = function(img){
var mask = img.select('shadowScore').gt(max_probS).eq(0);
return img.updateMask(mask);
};

// Aplicar máscara de sombras
var S2_maskNS = probSombras.map(aplicar_masks);

// ----- ÍNDICES DE VEGETACIÓN ----- \

// NDVI
var NDVI = function(image){
    var ndvi = image.normalizedDifference(['B8', 'B4'])
    // Ajustar los valores evitando píxeles saturados y con suelo
    .clamp(0.1, 0.8);
    return ndvi.select([0], ['NDVI'])
        .copyProperties(image, ['PRODUCT_ID',
            'system:index', 'system:time_start']);
};

// NDRE
var NDRE = function(image){
    var ndre = image.normalizedDifference(['B7', 'B4'])
    // Ajustar los valores evitando píxeles saturados y con suelo
    // .clamp(0.1, 0.8);
    return ndre.select([0], ['NDRE'])
        .copyProperties(image, ['PRODUCT_ID',
            'system:index', 'system:time_start']);
};

```

```

// IRECI
var IRECI = function(image){
  var ireci = image.expression(
    '(re3-r)/(re1/re2)',
    {
      re3: image.select('B7'),
      r: image.select('B4'),
      re1: image.select('B5'),
      re2: image.select('B6')
    });
  return ireci.select([0], ['IRECI'])
    .copyProperties(image, ['PRODUCT_ID',
      'system:index', 'system:time_start']);
};

// ----- INICIO DE LA FUNCIÓN ----- \\

var S2serietemporal = function(coleccion, IV,
                              a_inicio,mes_inicio,dia_inicio,
                              a_final,mes_final,dia_final){

  // Establecer el rango de años de la serie
  var yearrangeStart = a_inicio;
  var yearrangeStop = a_final;

  // Generar la coleccion con todas las imágenes
  var col;

  // Funcion de recorte, en función de los recintos cultivados
  var clip = function(image){return image.clip(cultivo)};
  // La variable "cultivo" se define más adelante,
  // integra las parcelas cultivadas en el ciclo seleccionado

  // .....| Cálculo de las series temporales |.....\\

  // Loop a través de los años (el código siguiente se ejecuta para cada año)
  for(var loopYear = yearrangeStart; loopYear < yearrangeStop; loopYear +=1){
    // Seleccionar la temporada de producción con la que seleccionar recintos
    var prod_temp = loopYear += 1;
    // Se suma un año, pues los valores de producción utilizados en el
    // filtro de parcelas están en el año siguiente al comienzo de la temporada

    // Se vuelve al año inicial de la temporada
    loopYear -= 1;
  }
}

```

```

// Seleccionar los campos sembrados con cebada en la temporada
var cultivo = AOI.filter(ee.Filter.gt(prod_temp.toString(), 1));

// Seleccionar las fechas de filtrado (de temporada en temporada)
// La temporada de la cebada comienza en octubre y finaliza en julio
var start = ee.Date.fromYMD(loopYear, mes_inicio, dia_inicio);
var end = ee.Date.fromYMD(loopYear +=1, mes_final, dia_final);

// Filtrar la colección por fecha y recortar la colección con
// los recintos anteriores
var imgClip = coleccion.filterDate(start, end).map(clip);

// Aplicar la función del IV llamado en la función
var imgIV = imgClip.map(IV);

// Volver al año de inicio
loopYear -= 1;

// Condicional IF:
// Si el bucle trabaja con el primer año de la serie crea la colección
// El primer año crea la colección
if(loopYear == a_inicio){
  col = imgIV;
  // El resto de años incluirá las imágenes en la colección anterior
} else if(loopYear > a_inicio && loopYear < a_final){
  col = col.merge(imgIV);
}
} // Fin del bucle FOR

// Crear el gráfico
var chart = ui.Chart.image.series({
  imageCollection: col,
  region: AOI,
  // Calcular la media de cada imagen
  reducer: ee.Reducer.mean(), //opciones= mean, median, stdDev, sum
  scale: 20
});

// Definir el nombre del gráfico en función de los años de la serie temporal
var filename = "".concat(a_inicio.toString())
  .concat('-').concat(a_final.toString());
print(chart);
}; // Fin de la función

```

```
// ----- LLAMAR A LA FUNCIÓN PRINCIPAL ----- \\  
S2serietemporal(S2_maskNS, IRECI, 2015,10,1,2020,7,31);
```

IV. Corrección atmosférica mediante Py6s

El código puede ejecutarse desde el repositorio de Sam Murphy [48] o instalando los módulos llamados en el inicio de la secuencia.

```
# Incorporación de los paquetes necesarios
import ee # Modulo de GEE
from Py6S import *
import datetime
import math
import os
import sys
sys.path.append(os.path.join(os.path.dirname(os.getcwd()), 'bin'))
from atmospheric import Atmospheric

# Inicio de Earth Engine a través de la API de Python
ee.Initialize()

# Función para extraer la respuesta espectral
def spectralResponseFunction(bandname):
    """
    Extracción de la función de respuesta espectral
    de las bandas Sentinel 2
    """

    bandSelect = {
        'B1':PredefinedWavelengths.S2A_MSI_01,
        'B2':PredefinedWavelengths.S2A_MSI_02,
        'B3':PredefinedWavelengths.S2A_MSI_03,
        'B4':PredefinedWavelengths.S2A_MSI_04,
        'B5':PredefinedWavelengths.S2A_MSI_05,
        'B6':PredefinedWavelengths.S2A_MSI_06,
        'B7':PredefinedWavelengths.S2A_MSI_07,
        'B8':PredefinedWavelengths.S2A_MSI_08,
        'B8A':PredefinedWavelengths.S2A_MSI_09,
        'B9':PredefinedWavelengths.S2A_MSI_10,
        'B10':PredefinedWavelengths.S2A_MSI_11,
        'B11':PredefinedWavelengths.S2A_MSI_12,
        'B12':PredefinedWavelengths.S2A_MSI_13,
    }

    return Wavelength(bandSelect[bandname])

# Función de conversión (reflectancia a radiancia)
```

```

def toa_to_rad(bandname):
    """
    Conversión de la reflectividad a techo de la atmósfera en radiancia
    """
    # Calcular la irradiancia solar exoatmosférica
    ESUN = info['SOLAR_IRRADIANCE_'+bandname]
    solar_angle_correction = math.cos(math.radians(solar_z))

    # Distancia Tierra-Sol (día del año, doy)
    doy = scene_date.timetuple().tm_yday
    d = 1 - 0.01672 * math.cos(0.9856 * (doy-4))
    # http://physics.stackexchange.com/
    # questions/177949/earth-sun-distance-on-a-given-day-of-the-year

    # factor de conversión
    multiplier = ESUN*solar_angle_correction/(math.pi*d**2)

    # calculo de radiancia
    rad = toa.select(bandname).multiply(multiplier)

    return rad

# Cálculo de reflectividad BOA
def surface_reflectance(bandname):
    """
    Calculo de la reflectividad en superficie a través de la
    radiancia del sensor (en función de la longitud de onda,
    específica para cada banda)
    """

    # Extraer la respuesta espectral de la banda
    s.wavelength = spectralResponseFunction(bandname)
    # Ejecutar los objetos 6s (definidos en la función principal)
    s.run()

    # Extraer las incógnitas atmosféricas
    Edir = s.outputs.direct_solar_irradiance # irradiancia solar directa
    Edif = s.outputs.diffuse_solar_irradiance # irradiancia solar difusa
    Lp = s.outputs.atmospheric_intrinsic_radiance # path radiance
    absorb = s.outputs.trans['global_gas'].upward # absorption transmissivity
    scatter = s.outputs.trans['total_scattering']\
        .upward # scattering transmissivity
    tau2 = absorb*scatter # total transmissivity

```

```

# Nota: los s.outputs son calculados automáticamente a partir de los
# objetos 6s definidos en la función de conversión principal, "conversion".

# Transformar los valores de reflectividad TOA a radiancia
rad = toa_to_rad(bandname)

# despejar la ecuación de transferencia radiativa
ref = rad.subtract(Lp).multiply(math.pi).divide(tau2*(Edir+Edif))

# Devuelve la reflectividad a BOA de una banda
return ref

# Incluir el área de estudio:
# Polígono a partir del cual se filtrara la colección
# y se recorta la imagen final

geom = ee.Geometry.Polygon([[[-0.9570796519011493,40.98197275411647],
[-0.5670650034636493,40.98197275411647],
[-0.5670650034636493,41.45919658393617],
[-0.9570796519011493,41.45919658393617],
[-0.9570796519011493,40.98197275411647]]])

# Descomentar la siguiente línea si ee.Geometry.Polygon no funciona
# geom = ee.Geometry.Rectangle(-0.996, 41.508, -0.568, 40.992)

# Modificar el polígono anterior para poder recortar la imagen final
region = geom.buffer(1000).bounds().getInfo()['coordinates']

# Inicio de la función principal
def conversion(img):

# Incorporar la fecha de la imagen
date = img.date()

# Definir las variables globales:
# Aquellas que pueden ser llamadas fuera del entorno de la función.
global toa
global info
global scene_date
global solar_z

# calcular la reflectividad a TOA

```

```

toa = img.divide(10000)

# Escribir los metadatos de la imagen
# Recopilar las propiedades
info = img.getInfo()['properties']
# Fecha: Python utiliza segundos, EE milisegundos
scene_date = datetime.datetime\
    .utcfromtimestamp(info['system:time_start']/1000)
# Angulo cenital solar
solar_z = info['MEAN_SOLAR_ZENITH_ANGLE']

# Valores sobre la composición atmosférica
# El código de las funciones se encuentra dentro
# del repositorio de samsammurphy
h2o = Atmospheric.water(geom,date).getInfo()
o3 = Atmospheric.ozone(geom,date).getInfo()
# Atmospheric Optical Thickness
aot = Atmospheric.aerosol(geom,date).getInfo()

# Altura de la superficie, a partir del MDE de la misión
# Shuttle Radar Topography mission (STRM) en GEE
SRTM = ee.Image('CGIAR/SRTM90_V4')
# Calculo de la altura media del área de estudio (geom)
alt = SRTM.reduceRegion(reducer = ee.Reducer.mean(),
    geometry = geom.centroid()).get('elevation').getInfo()
# Transformar a km, medida utilizada por Py6s
km = alt/1000

# Inicio de los objetos 6s, columna vertebral de Py6s
# A partir de la clase 6s se definen los parámetros
# requeridos por la función de transferencia radiativa

# Llamar a los objetos 6s
global s
s = SixS()

# Integrar los componentes atmosféricos
s.atmos_profile = AtmosProfile.UserWaterAndOzone(h2o,o3)
s.aero_profile = AeroProfile.Continental
s.aot550 = aot

# Calcular la geometría Earth-Sun-satellite
s.geometry = Geometry.User()
s.geometry.view_z = 0 # cálculo asumiendo vision en NADIR

```

```

s.geometry.solar_z = solar_z          # ángulo cenital solar
s.geometry.month = scene_date.month  # mes usado en la distancia Earth-Sun
s.geometry.day = scene_date.day      # día usado en la distancia Earth-Sun
s.altitudes\
    .set_sensor_satellite_level()    # Altitud del sensor
s.altitudes\
    .set_target_custom_altitude(km) # Altitud de la superficie

# Aplicar la conversión a cada banda de la imagen
# 1. Generar el objeto (imagen) a exportar
output = img.select('QA60')
# 2. Bucle de corrección: aplica la función de corrección a las bandas
# de la lista
for band in ['B2', 'B3', 'B4', 'B5', 'B6', 'B7', 'B8', 'B8A', 'B11', 'B12']:
    print(band)
    # Añade la banda corregida a la imagen a exportar
    output = output.addBands(surface_reflectance(band))
# Exportar la imagen a una carpeta en GEE
# 1. Definir parámetros de la imagen a exportar
dateString = scene_date.strftime("%Y-%m-%d")
ref = output.set({'satellite': 'Sentinel 2',
                  'fileID': info['system:index'],
                  'date': dateString,
                  'aerosol_optical_thickness': aot,
                  'water_vapour': h2o,
                  'ozone': o3})

# 2. Definir la carpeta de destino (dentro de GEE)
assetID = 'users/iranzocristian/6s_test/S2SR_'+dateString

# 3. Opciones de la imagen a exportar
export = ee.batch.Export.image.toAsset(\
    image=ref,
    description='sentinel2_atmcorr_export',
    assetId = assetID,
    region = region,
    crs = 'EPSG:4326',
    scale = 20)

# 4. Exportar la imagen
export.start()
return print("imagen "+assetID+" exportada")

# Final de la función principal

```

```

# Crear la colección a filtrar
S2 = ee.ImageCollection('COPERNICUS/S2')\
.filterBounds(geom)\
.filterDate('2015-10-01','2017-04-30')\
.filterMetadata('MGRS_TILE', 'equals', '30TXL')\
.filterMetadata('CLOUDY_PIXEL_PERCENTAGE', 'less_than', 20)\
.sort('system:time_start')\
.distinct('system:time_start')

# Definir en una lista las imágenes a filtrar
features = S2.getInfo()['features']

# CORRECCIÓN DE LA COLECCIÓN AUTOMÁTICAMENTE (bucle for)
# 1. Recorre cada imagen de la lista anterior
# 2. Obtiene su id
# 3. Se llama a la imagen de la colección GEE con el id anterior
# 4. Se aplica la función de conversión principal
for i in features:
    id = i['id']
    conversion(ee.Image(id))

# Final del Script

```
