

Estimación de la producción de cebada a partir de imágenes Sentinel-1, Sentinel-2 y variables climáticas

Cristian Iranzo ^{1*}, Raquel Montorio ^{1,2}, Alberto García-Martín ^{2,3}

¹ Departamento de Geografía y Ordenación del Territorio, Universidad de Zaragoza, Zaragoza, España.

² Grupo GEOFOREST-IUCA, Universidad de Zaragoza, Zaragoza, España.

³ Centro Universitario de la Defensa de Zaragoza, Academia General Militar, Zaragoza, España.

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 capaz de estimar la producción de cebada en una pequeña explotación (127 ha), localizada en el municipio de Belchite (España). Los modelos se entrenan con variables temporales adaptadas al calendario de cultivo de la cebada en la explotación estudiada. Las variables se dividen entre las creadas con información climática y las creadas con imágenes procedentes de teledetección. Se utilizan imágenes en dos rangos del espectro electromagnético, i.e., las microondas y el óptico, tomadas con los satélites Sentinel-1 y Sentinel-2, respectivamente. Los modelos se definen utilizando todas las combinaciones de variables predictoras correlacionadas con la producción mediante una regresión lineal múltiple. El modelo con mejores resultados devuelve un error en la predicción de 57,38 kg/ha (4%). La utilización de variables espectrales, derivadas del índice de vegetación radar *Cross Ratio* (CR) y el óptico *Inverted Red Edge Chlorophyll Index* (IRECI), combinadas con variables climáticas y adaptadas al calendario del cultivo, se revela como una estrategia adecuada para obtener modelos precisos.

Palabras clave: agricultura, índices de vegetación, calendario agronómico, regresión múltiple, Google Earth Engine.

Estimation of barley yield from Sentinel-1 and Sentinel-2 imagery and climatic variables

Abstract: A precise estimation of agricultural production provides relevant information for upcoming seasons, and helps in the assessment of crop losses before harvest in case of adverse situations. The objective of this work is to explore the development of a model capable of estimating barley production of a small agricultural production (127 ha) in Belchite, Spain. Variables adapted to the crop calendar of the growing barley are used to achieve that purpose. The variables have been created with weather data and remote sensing images. These images are acquired in two ranges of the electromagnetic spectrum, i.e., microwaves and optical spectral range, obtained from Sentinel-1 and Sentinel-2, respectively. Models are defined with a multiple linear regression method using all combinations of the independent variables correlated with production. The best linear regression model has a prediction error of 57.38 kg/ha (4%). The use of spectral variables, derived from radar vegetation index *Cross Ratio* (CR) and optical *Inverted Red Edge Chlorophyll Index* (IRECI), and climatic variables adapted to the crop calendar and climatic conditioning is revealed as an adequate strategy to obtain adjusted models.

Key words: agriculture, vegetation indices, crop calendar, multiple regression, Google Earth Engine.

To cite this article: Iranzo, C., Montorio, R., García-Martín, A. 2022. Estimation of barley yield from Sentinel-1 and Sentinel-2 imagery and climatic variables. *Revista de Teledetección*, 59, 61-72. <https://doi.org/10.4995/raet.2021.15099>

* Corresponding author: iranzocristian@gmail.com

1. Introducción

El crecimiento de la población mundial ha provocado un aumento en el consumo de alimentos, cuya producción debe incrementarse para abastecer las necesidades de las nuevas generaciones y garantizar la seguridad alimentaria (FAO, 2017), incluida en el segundo Objetivo de Desarrollo Sostenible (ODS) de las Naciones Unidas. La puesta en práctica de este principio requiere de explotaciones eficientes capaces de obtener la máxima producción con el mínimo de recursos posibles, minimizando el impacto sobre el medio ambiente y permitiendo así la continuidad de la explotación a largo plazo (Karthikeyan et al., 2020).

Una de las técnicas que mejora la gestión de las explotaciones es la estimación de sus producciones meses antes de la cosecha, información que facilita al agricultor la planificación de las próximas campañas y le permite calcular las pérdidas que pueden producirse durante el final de la temporada (Kalluri et al., 2003). La mejora en los programas de observación de la tierra ha permitido generar modelos de estimación de la producción mediante imágenes procedentes de la teledetección, capaces de obtener información sobre grandes superficies de una forma más rápida y económica que las medidas tomadas en campo (Steele-Dunne et al., 2017).

Los modelos con los que suelen obtenerse resultados más exactos a escala de explotación son los denominados modelos dinámicos. Simulan el crecimiento de los cultivos al final de la temporada mediante un conjunto de ecuaciones matemáticas y toda una serie de variables predictoras, relacionadas con el estado de la vegetación y las condiciones de cultivo (Kasampalis et al., 2018). Los datos procedentes de la teledetección, introducidos en los modelos a partir de índices de vegetación (en adelante IV), integran información acerca del estado de los cultivos. Su evolución durante la temporada se asocia a los cambios en la biomasa, la arquitectura del dosel vegetal y el contenido de clorofila de la vegetación, variables todas ellas relacionadas con la producción final (Pinter et al., 2003). Los modelos dinámicos que consiguen un menor error en la estimación requieren variables tomadas en campo, e.g. el total de

biomasa seca (Ameline et al., 2018), difíciles de obtener sobre explotaciones no experimentales.

Como alternativa, los modelos empíricos de regresión lineal utilizan la relación entre los datos de producción en varias temporadas y las medidas de un IV, agrupadas estas últimas entre la siembra y el momento a partir del cual se realiza la estimación (Basso et al., 2013). Cuanto mayor sea la frecuencia temporal de las medidas mejor recogerán los datos agrupados de los IV las variaciones ocasionadas por los cambios en las condiciones de cultivo (e.g. climatología y tipo de suelo).

Desde abril de 2017, las constelaciones de Sentinel-1 (S1) y Sentinel-2 (S2) desarrolladas por la Agencia Espacial Europea (ESA, por sus siglas en inglés), centradas en la región de las microondas y el óptico, respectivamente, se encuentran plenamente operativas. Ambos satélites 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 (Gibson y Power, 2000), por lo que son capaces de captar las variaciones estacionales de los cultivos. Sin embargo, su potencial para estimar la producción en pequeñas explotaciones apenas ha sido estudiado en casos reales. En cultivos de cereal existen algunas recientes excepciones (Zhao et al., 2020). Son este tipo de cultivos los que se utilizan en la explotación estudiada en este trabajo. Se compone exclusivamente por cultivos de cebada, cereal con una gran producción en España, el quinto país productor del mundo con 9,1 millones de toneladas en la campaña 2018/19 (FAO, 2020).

En este contexto, el objetivo del presente trabajo es explorar la capacidad de los modelos de regresión lineal para estimar la producción de una pequeña explotación, utilizando información climática e imágenes satelitales de resolución espacial media procedentes de las constelaciones Sentinel-1 y Sentinel-2. La escala de trabajo permite además estudiar la aplicación de variables adaptadas al calendario de cultivo en la explotación estudiada.

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. Se compone de veinte recintos (Figura 1), con una superficie total de 127 ha y 2,7° de pendiente media.

Las litologías dominantes están integradas en su mayor parte de series neógenas arcillosas y yesíferas. Los problemas de drenaje en este tipo de materiales junto a la escasa pendiente generan situaciones de inundación durante los eventos de precipitación torrencial (Longares, 1997), frecuentes bajo el clima semiárido de la zona. La precipitación anual es de 370 mm, concentrada al final de la primavera y comienzos del otoño, y una amplitud térmica anual de 17,6°C.

3. Materiales

3.1. Datos de producción

La serie de datos con la producción de cebada (*Hordeum vulgare* L. var. *Icaria*) comprende cinco temporadas de cultivo, desde 2015/16 a 2019/20. Están desagregados por recinto (i.e., una observación por parcela y campaña) e influidos por el sistema de rotación de cultivos en secano con el que se gestiona la explotación. En la primera temporada solo la mitad de los recintos producen rendimientos (≈ 64 ha), dejando el resto en barbecho para la temporada siguiente, donde los recintos cultivados en la primera entran en barbecho y así sucesivamente.

3.2. Imágenes de satélite

Las imágenes utilizadas provienen de las constelaciones S1 y S2 almacenadas en *Google Earth Engine* (en adelante GEE), entorno web que permite el procesamiento de las imágenes sin necesidad de descargarlas (Gorelick et al., 2017). La constelación S1 está compuesta por dos Radars de Apertura Sintética (SAR, por sus siglas en inglés) que captan información en la banda C. Se emplearon las imágenes tomadas en pasadas descendentes, con unos ángulos de incidencia medios de 37,5° sobre los recintos de la explotación. Las pasadas ascendentes fueron descartadas por su mayor variabilidad, con una desviación típica de 5,05° respecto a los 0,33° de las primeras. La resolución espacial de las imágenes de S1 es de 25 m, resultado de los pretratamientos integrados en la colección de GEE (ESA, 2020). De la constelación S2 se utilizaron las bandas 4, 5, 6, 7 y 8, con resoluciones espaciales entre los 10 y 20 m (Tabla 1). Se corresponden con las órbitas 51 y 94 en la tesela 30TXL.

Tabla 1. Características de los satélites utilizados. La resolución temporal refleja el período de revisita con los dos satélites en órbita y, entre paréntesis, con un único satélite.

Misión	Sentinel-1	Sentinel-2
Bandas	C	4, 5, 6, 7 y 8
Resolución espectral	5,6 cm	665-842 μ m
Resolución espacial	25 m	10-20 m
Resolución temporal	6(12) días	5(10) días

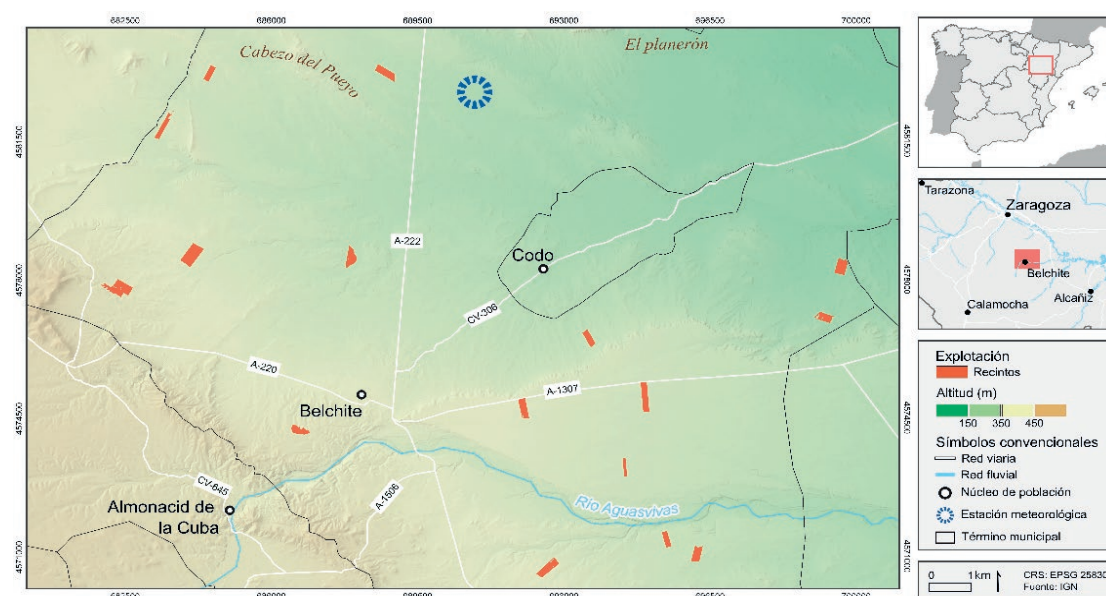


Figura 1. Localización de los recintos que componen la explotación.

El número de imágenes varía durante la serie temporal analizada (Tabla 2). En la primera temporada su número es bastante inferior al resto, consecuencia de una resolución temporal de 12 días para la constelación de S1 y de 10 días para S2. Con el lanzamiento de un nuevo satélite en cada constelación (abril de 2016 el S1-B y marzo de 2017 el S2-B) el período de revisita se redujo en 6 y 5 días, respectivamente, aumentando el número de imágenes disponibles.

Las imágenes de S2 entre el 1 de octubre de 2015 y el 27 de marzo de 2017, sin corrección atmosférica en GEE (nivel L1C), fueron corregidas con el modelo de transferencia radiativa *Second Simulation of the Satellite Signal in the Solar Spectrum* (6S) (Vermote et al., 1997), utilizando el módulo Py6s (Wilson, 2013). Se aplicó dentro de una secuencia de código en Python creada por Murphy (2020) y modificada para corregir en bucle todas las imágenes de este período. Las imágenes posteriores al 27 de marzo de 2017 se almacenan en GEE corregidas con el algoritmo Sen2Cor (nivel L2A), por lo que se realizó un análisis comparativo entre imágenes tratadas con ambos métodos para estudiar su integración conjunta. Se corrigieron con 6S imágenes L1C entre el 27 de marzo de 2017 y el 27 de marzo de 2018. A partir de los valores medios de reflectividad sobre las parcelas de la explotación entre imágenes corregidas con 6S e imágenes corregidas con Sen2Cor se calculó el RMSE en cada una de las bandas. Los bajos valores de RMSE obtenidos, inferior a 0,025 tantos por uno de reflectividad en todas las bandas analizadas (Tabla 3) descartan una variación significativa de las medidas, lo que permitió la integración conjunta de todas las imágenes en una única colección. Se descartó la

corrección de toda la colección con Py6s por su elevada demanda computacional.

Tabla 3. RMSE de los valores medios de reflectividad entre las imágenes corregidas con Sen2Cor (producto L2A de GEE) y las imágenes corregidas con Py6s.

Bandas	B4	B5	B6	B7	B8
RMSE	0,017	0,024	0,018	0,017	0,024

La influencia de las nubes y sus sombras se redujo mediante dos máscaras, aplicadas dentro de GEE a las escenas con un porcentaje de nubes inferior al 40%. Una primera máscara eliminó los píxeles con una probabilidad de nubes superior al 65% y una segunda máscara de sombras, adaptada del código de Schmitt et al. (2019), eliminó los píxeles con bajos valores de reflectividad como consecuencia de la proyección de las sombras de las nubes sobre la superficie. El código completo se puede consultar en el repositorio de Github creado por Iranzo (2020).

3.3. Índices de vegetación

Se calcularon un total de cuatro índices de vegetación (IV), utilizados en la creación de variables sobre el estado del cultivo (Tabla 4). Con la información de S1 se obtuvo el *Cross Ratio* (CR), aplicando un cociente entre las bandas con polarización VH y VV. Los valores de retrodispersión de las imágenes GEE se convirtieron a sigma (σ°) mediante la inversa del logaritmo, calculando después el cociente entre ambas bandas. A continuación se realizó la transformación logarítmica de sus valores, obteniendo resultados comparables con los estudios donde este IV ha sido aplicado (Veloso et al., 2017). Estos estudios, conducidos

Tabla 2. Número de imágenes utilizadas por temporada, misión y mes de captura. Las imágenes de Sentinel-2 seleccionadas tienen una nubosidad <40% en toda la escena, entre paréntesis se muestran las imágenes descartadas.

Temporada	2015/16		2016/17		2017/18		2018/19		2019/20	
Misión	S1	S2	S1	S2	S1	S2	S1	S2	S1	S2
Octubre	3	0	5	5	5	6	5	5	5	6
Noviembre	1	2	5	3	5	3	5	2	5	4
Diciembre	2	3	5	1	6	7	5	5	5	7
Enero	3	1	5	1	5	9	5	8	5	7
Febrero	1	1	5	4	4	1	4	9	5	6
Marzo	3	1	5	4	6	5	5	11	4	4
Abril	2	4	5	2	5	4	5	5	5	3
Total	15	12(31)	35	20(25)	36	35(37)	34	45(40)	34	37(58)

Tabla 4. Índices de vegetación calculados a partir de las imágenes de Sentinel-1 y Sentinel-2.

Índice	Ecuación	Referencia
Cross Ratio (CR)	VH/VV	Veloso et al. (2017)
Normalized Difference Vegetation Index (NDVI)	$(B8-B4)/(B8+B4)$	Rouse et al. (1974)
Normalized Difference Red Edge Index (NDRE)	$(B7-B4)/(B7+B4)$	Barnes et al. (2000)
Inverted Red Edge Chlorophyll Index (IRECI)	$(B7-B4)/(B5/B6)$	Guyot y Baret (1988)

sobre cultivos de trigo y cebada, han detectado una relación entre los valores de CR y la estructura del dosel vegetal (e.g. altura, densidad y estructura interna) y su contenido de humedad, información no correlacionada con los IV del óptico.

Los IV calculados en el espectro óptico fueron el *Normalized Difference Vegetation Index* (NDVI), el *Normalized Difference Red Edge Index* (NDRE) y el *Inverted Red Edge Chlorophyll Index* (IRECI), cuyas ecuaciones se muestran en la Tabla 4.

3.4. Datos climáticos

Se utilizaron datos diarios de temperatura y precipitación, variables con una gran influencia en la producción final de los cultivos (Monteith et al., 1977), procedentes de la estación de Belchite, gestionada por el Sistema de información Agroclimática para el Regadío (SiAR) (coordenadas ETRS89 41° 21'N – 0° 43' W, 327 m s.n.m.) (Figura 1).

4. Metodología

El estudio de la estimación de la producción en una pequeña explotación se realizó en tres etapas. En primer lugar, se generaron las series temporales con los índices de vegetación y la información climática de las cinco campañas analizadas, con las que se definió el calendario agronómico del cultivo. Después se crearon las variables predictoras, ajustadas al calendario de cultivo, y se integraron junto a la producción en una base de datos utilizando una secuencia de código en R v.4.0.3 (R Core Team, 2020). Se utilizaron para definir un conjunto de modelos de regresión lineal con todas las combinaciones posibles de variables predictoras que, junto al cálculo de los estadísticos de error, se realizaron en R con el paquete 'caret' (Kuhn, 2020). El último paso consistió en el estudio de los resultados del conjunto de modelos.

4.1. Obtención de series temporales

Los valores de cada IV a lo largo de las cinco temporadas se obtuvieron mediante dos funciones definidas en GEE (Iranzo, 2020). Al aplicarlas se genera una tabla con el valor promedio, máximo y mínimo de cada IV sobre los recintos cultivados (i.e. aquellos con una producción superior a 0) por cada una de las imágenes de satélite disponibles. Una vez calculadas, las series temporales de NDVI, NDRE e IRECI se sometieron a un filtrado posterior en una secuencia de código en R para eliminar los datos influidos por la presencia de nubes, i.e., aquellos 1) cuya diferencia respecto a la media entre el valor anterior y posterior en la serie fuera superior a 0,06 y 2) su porcentaje de nubes dentro de los recintos de la explotación fuera superior al 40%. Este último dato se integra en las propiedades de cada imagen al calcular la máscara de nubes. La serie temporal con los datos diarios de precipitación y temperatura media se generó seleccionando los datos originales de la estación del SiAR durante el período analizado, comprendido entre los meses de octubre y julio, momentos de siembra y cosecha respectivamente, durante las temporadas 2015/16 a 2019/20.

4.2. Creación del calendario agronómico

El análisis de las series temporales con información espectral y climática (Figura 2), unida a las observaciones de campo, permitió conocer las etapas más representativas en las que se divide el ciclo de cultivo de la cebada en la explotación estudiada (Tabla 5). Los días de inicio y final de cada fase se mantienen invariables durante las cinco temporadas, permitiendo relacionar las variaciones en cada una de ellas con aumentos y descensos de la producción.

La siembra comienza el uno de octubre con las primeras lluvias, prolongada hasta el quince de noviembre cuando se retrasan las precipitaciones

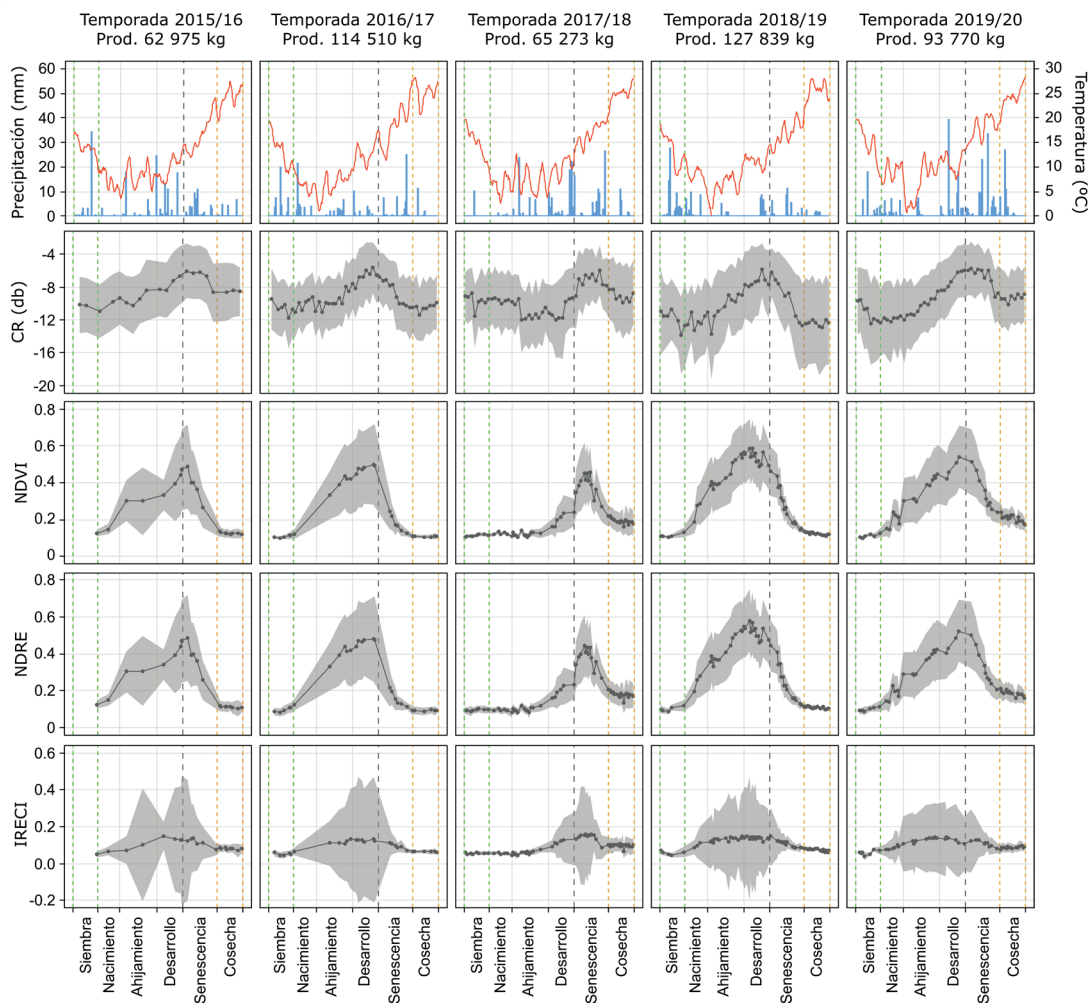


Figura 2. Serie temporal con los índices de vegetación durante las temporadas de cultivo analizadas en relación con la precipitación y la temperatura media diarias. Se muestra la desviación típica de los valores (gris), el momento final de la etapa de desarrollo utilizado como fecha límite de la predicción (15 de abril) y los periodos de siembra (verde) y cosecha (naranja).

(e.g. temporada 2017/18). La respuesta de los IV es baja en este periodo, detectando el suelo desnudo. Entre el 16 de noviembre y el 25 de diciembre

Tabla 5. Etapas del calendario agronómico de la cebada en el área de estudio utilizadas en la creación de las variables predictoras.

Fase	Período seleccionado
Siembra	1 oct - 15 nov
Nacimiento	16 nov - 25 dic
Ahijamiento	26 dic - 28 feb
Desarrollo	29 feb - 15 abr

tiene lugar la fase de nacimiento, donde los IV sufren un leve incremento debido al nacimiento de las plantas. A partir del 26 de diciembre empieza el ahijamiento de la cebada, desarrollando nuevos pies en la base del tallo que generan un incremento de la señal del CR. Esta etapa finaliza el 28 de febrero con el aumento de las temperaturas medias, dando comienzo la etapa de desarrollo. En ella tiene lugar la floración y la aparición de las espigas antes del 15 de abril, momento donde el cultivo alcanza su máximo contenido de biomasa y humedad en las temporadas con mayor producción (2016/17, 2018/19 y 2019/20). Es captado en las series temporales de los IV con un pico en sus

valores. Al contrario, en las temporadas con menor producción (2015/16 y 2017/18) las espigas aparecen después del quince de abril, en la etapa de senescencia. Se reduce así el período disponible para el llenado de los granos, disminuyendo su peso y con ello la producción final.

Durante la última etapa de senescencia la cebada pierde humedad y vigor, reduciéndose las diferencias en los valores de los IV entre temporadas. Por ello se decidió descartar la información de la etapa de senescencia a la hora de crear las variables predictivas, pues no aporta información relevante a la hora de estimar la producción.

4.3. Creación de variables predictoras

Se definieron dos grupos de variables predictoras, i.e., variables espectrales sobre el estado de los cultivos y variables climáticas. Ambas agrupan la información de las series temporales en un único dato por temporada, tomando como referencia las etapas del calendario agronómico.

Las variables sobre el estado del cultivo se crearon agrupando las medidas de los IV mediante cuatro métodos distintos, aplicados individualmente sobre cada índice, generando un total de sesenta y ocho variables (Tabla 6). Al agrupar sus valores en un único dato, se consiguió reducir la influencia provocada por la variabilidad del número de imágenes por temporada.

Los métodos utilizados fueron: 1) promedio de los valores entre la siembra y el desarrollo; 2) valor máximo, mínimo y promedio de cada una de las etapas del calendario agronómico utilizadas; 3) suma del conjunto de los promedios, máximos y mínimos de cada una de las fases del calendario de cultivo; 4) valor del percentil 95 entre las etapas de siembra y desarrollo.

Las variables asociadas al clima de la zona de estudio se crearon promediando los valores de temperatura media y sumando la precipitación en cada una de las etapas del calendario agronómico (Tabla 7). Este proceso dio como resultado cuarenta variables climáticas, a las que se añadieron dos variables adicionales: 1) el promedio de la temperatura mínima en el período teórico de vernalización, i.e. mecanismo que bloquea el desarrollo de los órganos reproductores de la vegetación durante la bajada térmica, producida aproximadamente entre el quince y el veintiocho de diciembre; 2) la suma conjunta de la precipitación en las etapas de siembra y nacimiento.

4.4. Obtención de los modelos de regresión

Los modelos de regresión lineal se crearon automáticamente relacionando las variables independientes (i.e. variables climáticas y espectrales) con la respuesta de la variable dependiente (i.e. la producción) en las cinco campañas analizadas. Se utilizaron un máximo de dos variables por modelo,

Tabla 6. Variables espectrales sobre el estado de los cultivos definidas con cada índice de vegetación. El código se completa con la abreviación del índice en el lugar de *IV*.

Variable	Código
Promedio de los valores entre la siembra y el desarrollo	IV_promedio
Promedio en la etapa de siembra, nacimiento, ahijamiento y desarrollo	IV_etapa
Suma de los promedios por estaciones (siembra-desarrollo)	sumIV
Valor máximo/mínimo en la etapa de siembra, nacimiento, ahijamiento y desarrollo	max/minIV_etapa
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

Tabla 7. Variables climáticas.

Datos	Variable	Código
Precipitación	Suma de los valores por etapas (siembra, nacimiento, ahijamiento y desarrollo)	pp_etapa
	Suma entre las etapas de siembra y nacimiento	pp_OctDic
Tª media	Promedio en las etapas de siembra, nacimiento, ahijamiento y desarrollo	tMed_etapa
Tª mínima	Promedio en la fase de vernalización	tMin_vernalización

número con el que se alcanza el grado de libertad permitido. Del conjunto de variables independientes, solamente se incluyeron las variables normales y correlacionadas con la producción mediante el coeficiente de Pearson (Tabla 8), reduciendo el coste computacional y generando un total de 136 modelos. Se consideró que existía una correlación entre la variable independiente y la producción cuando el coeficiente superara el 0,8 y su nivel de significación el 90% de confianza (i.e., un *p-value* inferior a 0,1). El umbral de significación se escogió teniendo en cuenta el reducido número de observaciones, disminuyendo la capacidad para generar correlaciones a un nivel de confianza del 95%, el valor comúnmente utilizado (Wasserman, 2004).

La bondad de los modelos para estimar la producción se evaluó a partir de una validación cruzada *leave one out* (LOOCV). Con los valores observados y estimados, resultado de la validación, se calcularon el coeficiente de determinación (R^2), el error cuadrático medio de la validación cruzada (RMSECV), la desviación predictiva residual (RPD por sus siglas en inglés, calculada como la ratio entre la desviación estándar de valores observados y el RMSECV) y el valor de significación (*p-value*) de cada modelo.

5. Resultados

Del conjunto de modelos de regresión creados se han seleccionado aquellos con un RPD superior a 2, de acuerdo con los umbrales definidos en Chang

y Laird (2002). Los modelos, como muestra la Tabla 9, explican entre el 78-98% de la varianza, con un error de estimación inferior a los 15.000 kg y significativos al nivel $p < 0,05$.

De los seis modelos seleccionados, cuatro incluyen variables espectrales (tres de ellos en combinación con variables climáticas -modelos 1, 4 y 6- y uno exclusivamente con IV -modelo 2-) y dos se componen sólo de variables climáticas (modelos 3 y 5). Entre las variables espectrales destaca el CR en la siembra, variable con la que se alcanza el mejor modelo de regresión con un RPD > 7 , un R^2 de 0,98 y un error inferior a los 4000 kg. Dentro de los IV ópticos destaca claramente el índice IRECI en la fase de ahijamiento, generando por sí solo el segundo mejor modelo (casi el 90% de la varianza explicada y un error en torno a 10.000 kg). No obstante, no hay acuerdo en la mejor métrica, siendo seleccionado el valor máximo, el mínimo y el promedio. Respecto a las variables climáticas se observa una importante variabilidad, con una mayor presencia de las variables de temperatura que de precipitación. Destaca claramente la temperatura mínima en el período de vernalización, siendo seleccionada en 3 de los 6 modelos. La temperatura máxima de la fase de desarrollo es también importante, ya que contribuye a la formación del mejor modelo de regresión. Finalmente, entre las variables de precipitación destaca la suma entre los meses de octubre-diciembre (que corresponde a las fases de siembra y nacimiento) y la precipitación durante la etapa de nacimiento.

Tabla 8. Variables correlacionadas con la producción calculadas con el coeficiente de Pearson, utilizadas en el cálculo de los modelos de regresión lineal.

Variables sobre el estado del cultivo			Variables climáticas		
Código	<i>p-value</i>	R	Código	<i>p-value</i>	R
maxNDVI_ahijamiento	0,04	0,89	tMed_ahijamiento	0,03	-0,92
maxNDRE_ahijamiento	0,05	0,88	tMin_vernalizacion	0,03	-0,91
minIRECI_ahijamiento	0,06	0,87	pp_OctDic	0,03	0,91
maxIRECI_ahijamiento	0,06	0,87	pp_nacimiento	0,08	0,83
minNDVI_desarrollo	0,07	0,85	tMed_desarrollo	0,08	0,83
NDVI_ahijamiento	0,08	0,83			
CR_siembra	0,08	-0,83			
IRECI_ahijamiento	0,09	0,82			
minCR_siembra	0,09	-0,82			
minNDRE_desarrollo	0,1	0,81			

Tabla 9. Resultado de los mejores modelos de regresión lineal, i.e., *p-value* inferior a 0,1, RMSECV por debajo de los 15000 kg y RPD superior a 2. *Ahij.* abreviatura de ahijamiento.

ID	Variables	<i>p-value</i>	RMSECV	RPD	R ²
1	tMax_desarrollo + CR_siembra	0,001	3.672,34	7,88	0,98
2	IRECI_ahij. + maxIRECI_ahij.	0,024	10.315,04	2,80	0,88
3	tMin_vernalización + pp_OctDic	0,013	11.245,56	2,57	0,89
4	tMin_vernalización + minIRECI_ahij.	0,042	11.516,95	2,51	0,87
5	tMin_vernalización + pp_nacimiento	0,015	11.870,12	2,44	0,80
6	pp_OctDic + minCR_siembra	0,040	14.368,65	2,01	0,78

6. Discusión

El mejor modelo de regresión lineal es el definido con los promedios de temperatura máxima en la etapa de desarrollo y del CR en la etapa de siembra. Una temperatura elevada en primavera favorece el desarrollo de la espiga y el posterior llenado de granos. Los valores de CR en la etapa de siembra podrían estar indicando el nacimiento temprano de la cebada, pues sus valores disminuyen a medida que comienzan a producirse las primeras lluvias y a aparecer los primeros brotes entre el suelo desnudo. Esta característica podría estar ligada a la capacidad del CR para detectar cambios en la geometría de los cultivos (Veloso et al., 2017). Las temporadas en las que se combinan una siembra temprana y temperaturas elevadas durante la etapa de desarrollo aumentan el ciclo vegetativo de las plantas, incrementando el tiempo de crecimiento y, potencialmente, el número de granos al final de la campaña. La capacidad de los datos SAR para detectar el momento de nacimiento están en línea con el trabajo de Schlund y Erasmi (2020), donde se observa la utilidad del CR en la captación de momentos concretos del ciclo de cultivo. Además, estos datos no están influidos por la nubosidad atmosférica, factor que ha invalidado el 56% de las imágenes ópticas disponibles.

Las variables sobre el estado del cultivo están presentes en cuatro de los seis modelos destacados, incluyendo los dos mejores. Se obtienen en las etapas de siembra (CR) y ahijamiento (IRECI), utilizando tanto el promedio como los valores mínimo y máximo. Es en la fase de ahijamiento donde el IRECI detecta los mayores contrastes entre las temporadas con menor y mayor producción, con un pico en la respuesta durante las segundas. El incremento se asocia con la formación de

nuevos pies en la base del tallo principal, factor que contribuye a aumentar la producción final.

Las variables relacionadas con los índices NDVI y NDRE están ausentes entre los seis mejores modelos. Llama la atención si se tiene en cuenta que el NDVI ha devuelto buenos resultados para estimar la producción (Mirasi et al., 2019), y el NDRE mejora la discretización de la vegetación sana con respecto al primero, según Frampton et al. (2013). Esta situación contrasta con el protagonismo del IRECI, en cuya formulación se incluyen las tres bandas del *red edge* y que, como muestran algunos trabajos (Mercier et al., 2019), su aplicación disminuye la saturación sobre cultivos con alta densidad vegetal, observada en las medidas de NDVI. Todo ello apunta a una posible mejora en la estimación de la producción relacionada con el uso del *red edge*. De esta forma, utilizar la constelación S2 mejoraría la capacidad de monitorización agraria respecto a los sensores sin información en el *red edge* (e.g. Landsat).

Del conjunto de variables climáticas, el promedio de temperatura mínima en la fase de vernalización es la más representada, incluida en tres de los seis mejores modelos. Las bajas temperaturas durante la vernalización son cruciales para el correcto desarrollo de la espiga, evitando que la floración coincida con la época de heladas.

La serie temporal utilizada, con cinco años de duración, no permitió implementar en el modelo de regresión una validación externa prediciendo una campaña independiente. Sin embargo, la principal limitación a la hora de aplicar este trabajo a escala regional, o sobre pequeñas explotaciones ubicadas en un entorno diferente, es la dependencia del calendario agronómico a los recintos estudiados. El cálculo de los *growing degree days* hubiera

evitado depender del calendario agronómico, pero no pudieron utilizarse por la falta de datos sobre el día de siembra. Además, el empleo de modelos de regresión lineal multivariante ha podido generar un sobreajuste de los modelos a los datos modelados. A pesar de ello, los porcentajes de RMSECV obtenidos en el presente trabajo (en torno al 4% en el mejor modelo, seguido por el 11% en el segundo) están en la línea de lo observado en otros estudios desarrollados a escala regional, y con mayores series temporales. El trabajo de Becker-Reshef et al. (2010), por ejemplo, obtuvo un 7% de RMSE, utilizando el percentil 95 de NDVI en una serie temporal con 7 años.

Los buenos resultados a la hora de modelar la producción a gran escala, combinados con el correcto funcionamiento de las técnicas de procesamiento de datos empleadas en este estudio y la continuidad de los satélites S1 y S2 más allá del 2030, deberían motivar a los gestores de pequeñas explotaciones a recabar más información sobre sus parcelas. Almacenar y conservar los datos de producción por recinto, y otros como el momento de siembra, permitirían no solo modelar la producción, sino también desarrollar mapas de rendimiento con los que detectar los sectores menos fértiles. Con esta información podrían aplicarse tratamientos específicos, reduciendo así el consumo de fertilizantes y pesticidas (Birrell et al., 1996). En el contexto general de cambio global, estas prácticas contribuyen a la preservación del medio ambiente y al cumplimiento de los ODS, especialmente a los vinculados con la seguridad alimentaria y la degradación del suelo.

7. Conclusiones

El presente estudio explora las capacidades de los sensores Sentinel-1 y Sentinel-2 para la estimación de la producción de cebada a gran escala. Los modelos de regresión lineal aplicados reflejan la sinergia existente entre variables climáticas y espectrales. La creación de variables adaptadas al calendario del cultivo es uno de los factores que más contribuye a la creación de buenos modelos estimativos a escala de explotación, concretamente la incorporación de los valores de IRECI en la etapa de ahijamiento y de los promedios de CR en la etapa de siembra, cuyas estimaciones de la producción de cebada están por encima del 78% de R^2 y por debajo del 16% de RMSECV. Los resultados obtenidos sitúan a las

constelaciones S1 y S2 como una buena opción para obtener métricas sobre el estado de los cultivos a escala de explotación.

Agradecimientos

Los autores del trabajo agradecen a D. Jesús Murria la cesión de sus datos de producción y el esfuerzo dedicado en su preparación, sin los que este trabajo no hubiera sido posible. Los autores agradecen también a los revisores del trabajo, cuyos comentarios han contribuido a la mejora en la calidad del mismo.

Referencias

- Ameline, M., Fieuzal, R., Betbeder, J., Berthoumieu, J. F., Baup, F. 2018. 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. <https://doi.org/10.1109/JSTARS.2018.2878502>
- Barnes, E., Clarke, T., Richards, S., Colaizzi, P., Haberland, J., Kostrzewski, M., Waller, P., Choi, C., Riley, E., Thompson, T., Lascano, R. J., Li, H., Moran, M. S. 2000. Coincident detection of crop water stress, nitrogen status and canopy density using ground based multispectral data. En: *Proceedings of the 5th International Conference on Precision Agriculture*. Bloomington, EE.UU., 16-19 Julio. pp 1-15
- Basso, B., Cammarano, D., Carfagna, E. 2013. Review of crop yield forecasting methods and early warning systems. En: *Proceedings of the first meeting of the scientific advisory committee of the global strategy to improve agricultural and rural statistics*. Roma, Italia, Julio (Vol. 41).
- Becker-Reshef, I., Vermote, E., Lindeman, M., Justice, C. 2010. 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. <https://doi.org/10.1016/j.rse.2010.01.010>
- Birrell, S. J., Sudduth, K. A., Borgelt, S. C. 1996. Comparison of sensors and techniques for crop yield mapping. *Computers and Electronics in Agriculture*, 14(2), 215-233. [https://doi.org/10.1016/0168-1699\(95\)00049-6](https://doi.org/10.1016/0168-1699(95)00049-6)
- Chang, C.-W., Laird, D.A. 2002. Near-infrared reflectance spectroscopic analysis of soil C and N. *Soil Science*, 167(2), 110-116. <https://doi.org/10.1097/00010694-200202000-00003>

- European Space Agency (ESA). Sentinel-1 Toolbox. *Sentinel Online*. Último acceso: 25 de Septiembre, 2020, de <https://sentinel.esa.int/web/sentinel/toolboxes/sentinel-1>
- FAO, 2017. *The Future of Food and Agriculture - Trends and Challenges*. Roma: Food and Agriculture Organization of the United Nations.
- Food and Agriculture Organization of the United Nations. Crops Statistics. *FAOSTAT*. Último acceso: 4 de Agosto de 2020, de <http://www.fao.org/faostat/en/#data/QC>.
- Frampton, W. J., Dash, J., Watmough, G., Milton, E. J., 2013. Evaluating the capabilities of Sentinel-2 for quantitative estimation of biophysical variables in vegetation. *ISPRS Journal of Photogrammetry and Remote Sensing*, 82, 83-92. <https://doi.org/10.1016/j.isprsjprs.2013.04.007>
- Gibson, P. J., Power, C. H., 2000. *Introductory remote sensing: Principles and Concepts*. Londres: Routledge.
- Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., Moore, R. 2017. Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment*, 202, 18-27. <https://doi.org/10.1016/j.rse.2017.06.031>
- Guyot, G., Baret, F. 1988. Utilisation de la Haute Resolution Spectrale pour Suivre L'état des Couverts Vegetaux. In *4th International Colloquium "Spectral Signatures of Objects in Remote Sensing"*, Aussois, 18-22 January 1988, Paris: ESA, Publication SP-287 (pp. 279-286).
- Iranzo-Cubel, C. 2020. CristianICS/sentinel-iv-gee: v0.1.0: Vegetation Index in crop rotation agricultural holding (v0.1.0). Zenodo. <https://doi.org/10.5281/zenodo.5595040>
- Kalluri, S., Gilruth, P., Bergman, R. 2003. The potential of remote sensing data for decision makers at the state, local and tribal level: Experiences from NASA's Synergy program. *Environmental Science & Policy*, 6(6), 487-500. <https://doi.org/10.1016/j.envsci.2003.08.002>
- Karthikeyan, L., Chawla, I., Mishra, A. K., 2020. A review of remote sensing applications in agriculture for food security: Crop growth and yield, irrigation, and crop losses. *Journal of Hydrology*, 586, 124905. <https://doi.org/10.1016/j.jhydrol.2020.124905>
- Kasampalis, D. A., Alexandridis, T. K., Deva, C., Challinor, A., Moshou, D., Zalidis, G. 2018. Contribution of Remote Sensing on Crop Models: A Review. *Journal of Imaging*, 4(4), 52. <https://doi.org/10.3390/jimaging4040052>
- Kuhn, M. 2020. caret: Classification and Regression Training. Versión 6.0-86. <https://CRAN.R-project.org/package=caret>
- Longares, L. A. 1997. *El paisaje vegetal en el entorno de la reserva ornitológica «El Planerón» (Belchite-Zaragoza)*. Zaragoza: Sociedad Española de Ornitología D.L.
- Mercier, A., Betbeder, J., Baudry, J., Denize, J., Leroux, V., Roger, J.-L., Spicher, F., Hubert-Moy, L. 2019. 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*. Strasbourg, Francia, 21 Octubre. pp 1114903. <https://doi.org/10.1117/12.2533132>
- Mirasi, A., Mahmoudi, A., Navid, H., Kamran, K. V., Asoodar, M. A. 2019. Evaluation of sum-NDVI values to estimate wheat grain yields using multi-temporal Landsat OLI data. *Geocarto International*, 1-16. <https://doi.org/10.1080/10106049.2019.1641561>
- Monteith, J. L., Moss, C. J., Cooke, G. W., Pirie, N. W., Bell, G. D. H. 1977. Climate and the efficiency of crop production in Britain. *Philosophical Transactions of the Royal Society of London. B, Biological Sciences*, 281(980), 277-294. <https://doi.org/10.1098/rstb.1977.0140>
- Murphy, S. 2020. Atmospheric Correction of a (Single) Sentinel 2 Image. *Github*. <https://github.com/samsammurphy/gee-atmcorr-S2>.
- Pinter, Jr., Paul J., Hatfield, J.L., Schepers, J.S., Barnes, E.M., Moran, M.S., Daughtry, C.S.T., Upchurch, D.R. 2003. Remote Sensing for Crop Management. *Photogrammetric Engineering & Remote Sensing*, 69, 647-664. <https://doi.org/10.14358/PERS.69.6.647>
- QGIS.org, 2020. QGIS Geographic Information System (Versión 3.12.3). *Open Source Geospatial Foundation*. <http://qgis.org>
- R Core Team, 2020. R: A Language and Environment for Statistical Computing. *R Foundation for Statistical Computing*. <https://www.R-project.org/>
- Rouse, J. W., Hasas, R. H., Schell, J. A., Deering, D. W., Harlan, J. C. 1974. Monitoring the vernal advancement of retrogradation of natural vegetation. *NASA/OSFC. Type III. Final Report*. Oreenbello MD. pp. 371
- Schlund, M., Erasmí, S. 2020. Sentinel-1 time series data for monitoring the phenology of winter wheat. *Remote Sensing of Environment*, 246, 111814. <https://doi.org/10.1016/j.rse.2020.111814>

- Schmitt, M., Hughes, L. H., Qiu, C., Zhu, X. X. 2019. Aggregating cloud-free Sentinel-2 images with Google Earth Engine. *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, IV-2-W7, 145-152. <https://doi.org/10.5194/isprs-annals-IV-2-W7-145-2019>
- Steele-Dunne, S.C., McNairn, H., Monsivais-Huerta, A., Judge, J., Liu, P.-W., Papathanassiou, K. 2017. Radar Remote Sensing of Agricultural Canopies: A Review. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 10, 2249–2273. <https://doi.org/10.1109/JSTARS.2016.2639043>
- Veloso, A., Mermoz, S., Bouvet, A., Le Toan, T., Planells, M., Dejoux, J.-F., Ceschia, E. 2017. Understanding the temporal behavior of crops using Sentinel-1 and Sentinel-2-like data for agricultural applications. *Remote Sensing of Environment*, 199, 415-426. <https://doi.org/10.1016/j.rse.2017.07.015>
- Vermote, E. F., Tanre, D., Deuze, J. L., Herman, M., Morcette, J.-J. 1997. Second Simulation of the Satellite Signal in the Solar Spectrum, 6S: An overview. *IEEE Transactions on Geoscience and Remote Sensing*, 35(3), 675-686. <https://doi.org/10.1109/36.581987>
- Wasserman, L. 2004. Models, Statistical Inference and Learning. En L. Wasserman (Ed.), *All of Statistics: A Concise Course in Statistical Inference* (pp. 87-96). New York: Springer. https://doi.org/10.1007/978-0-387-21736-9_6
- Wilson, R. T. 2013. Py6S: A Python interface to the 6S radiative transfer model. *Comput. Geosci.*, 51(2), 166-171. <https://doi.org/10.1016/j.cageo.2012.08.002>
- Zhao, Y., Potgieter, A. B., Zhang, M., Wu, B., Hammer, G. L. 2020. Predicting Wheat Yield at the Field Scale by Combining High-Resolution Sentinel-2 Satellite Imagery and Crop Modelling. *Remote Sensing*, 12(6), 1024. <https://doi.org/10.3390/rs12061024>