

## Seguimiento y evaluación de la capacidad de discriminación de cultivos herbáceos en regadío a partir de imágenes de satélite en el periodo 2016-2018

Paredes-Gómez, V., Del Blanco-Medina, V., Gutiérrez-García, A. y Nafría-García, D.A.

Instituto Tecnológico Agrario de Castilla y León (ITACYL), Subdirección de Infraestructuras Agrarias, Unidad de Información Geográfica e Innovación, Finca Zamadueñas, Ctra. Burgos Km. 119, 47071, Valladolid, España.

pargomva@itacyl.es.

**Resumen:** El conocimiento de la superficie verdaderamente regada en cada campaña es un dato de gran interés para la planificación de regadíos, la estadística de rendimientos de los cultivos y la gestión del agua. El objetivo de este estudio es evaluar la capacidad de discriminar los cultivos herbáceos en regadío presentes en la comunidad de Castilla y León, mediante el Mapa de Cultivos y Superficies Naturales de Castilla y León (MCSNCyL), en los años 2016, 2017 y 2018, con condiciones agroclimáticas diferentes, siendo el 2016 un año ligeramente húmedo, el 2017 muy seco y el 2018 extraordinariamente húmedo. El método de clasificación es un algoritmo de aprendizaje automático basado fundamentalmente en imágenes de satélite y datos de entrenamiento procedentes de bases de datos de la Administración regional. El método es capaz de discriminar los principales cultivos herbáceos de secano y de regadío de Castilla y León, incluyendo los cultivos de invierno como el trigo de secano y de regadío. Fruto de este estudio se realiza una estimación de la superficie verdaderamente regada para cada uno de los tres años de estudio de este trabajo.

**Palabras clave:** Teledetección, Sentinel-2, mapa de cultivos, aprendizaje automático, regadío, área regada.

### *Monitoring and assessment of the ability to discriminate herbaceous actually irrigated crops from satellite images in the period 2016-2018*

**Abstract:** *The aim of this study is to assess the availability to distinguish irrigated arable crops in the Spanish region of Castile and León, by means of the Castile and Leon crops and natural land map (MCSNCyL), in three years with completely different meteorological pattern, considering 2016 as a mildly humid year, 2017 as a dry year and 2018 has been characterized as very humid. The land cover classification map is performed using a machine-learning algorithm based mainly on Sentinel-2 images and other ancillary data. The algorithm has been trained with data retrieved from Integrated Administration and Control System and some other land use databases available in Spain. The resulting map proves to be able to distinguish between the main rainfed crops and irrigated crops in our region. It reaches a good overall accuracy and the stability of the results makes it more feasible. Furthermore, it allow us to estimate an approximate area covering the main irrigated crops, which is not usually easily available. This latter might be an absolutely key point for policy decision makers concerning water resources management or sustainable agriculture.*

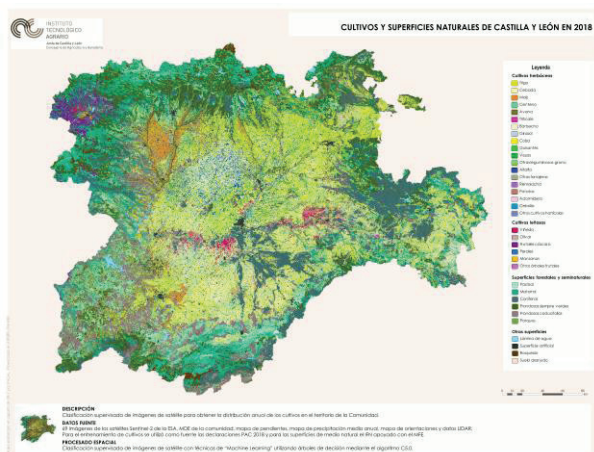
**Keywords:** *Remote sensing, Sentinel-2, crop classification, machine learning, irrigated crops, irrigated area.*

### 1. INTRODUCCIÓN

El Mapa de Cultivos y Superficies Naturales de Castilla y León (MCSNCyL, <http://mcsncyl.itacyl.es/>) es un proyecto que realiza un cartografiado anual de cultivos presentes en Castilla y León mediante técnicas de aprendizaje por computador basadas en series temporales de imágenes de satélite (Del Blanco, 2015). Este mapa de cultivos incluye entre sus clases la discriminación del tipo de explotación en secano o en regadío para aquellos cultivos que pueden ser explotados en ambos sistemas. El MCSNCyL (ver Figura 1) ha sido además propuesto como mapa de cultivos avanzado dentro de los 4 servicios propuestos como prueba de concepto en el proyecto europeo SENSAGRI (*Sentinels Synergy for Agriculture*) elaborado dentro del marco del programa EO Work programme “EO-3-2016: Evolution of Copernicus Services”.

La importancia de la implantación de esta metodología radica en la limitación que existe actualmente de conocer la superficie real de cultivos en régimen de regadío, así como su variación anual, en particular, en lo que respecta a los cultivos herbáceos. Más aún cuando la información

sobre el sistema de explotación de las bases de referencia, tanto de la PAC como de las confederaciones hidrográficas no se actualiza de una manera sistemática y en muchos casos esta característica simplemente se hereda de un año al siguiente independientemente del régimen de explotación que tenga el cultivo presente. Por este motivo, el estudio aquí expuesto proporciona una información de gran interés para la planificación y modernización de regadíos, la estadística de rendimientos de los cultivos y la gestión del agua. La estimación e identificación de cultivos de regadío lleva siendo objeto de muchos trabajos de investigación (Ambika, 2016) precisamente por la importancia que conlleva tener el mayor conocimiento posible del uso del agua en un mundo cada vez más escaso de ella. Por lo tanto, el fin último que se desea conseguir es llegar a tener un mejor conocimiento del uso del agua para la agricultura en nuestra región y poder alcanzar así con los objetivos propuestos por las directrices europeas para una agricultura y eficiencia de los recursos hídricos más sostenibles.



**Figura 1.** Mapa de cultivos y superficies naturales de Castilla y León en 2018.

## 2. MATERIAL Y MÉTODOS

### 2.1. Algoritmo de clasificación

El método de clasificación usado es de tipo supervisada y está basado en el aprendizaje automático con árboles de decisión. A partir del algoritmo inductivo See5 (Quinlan, 1992) obtenemos un modelo de predicción que sirve de herramienta para decidir qué tipo de cubierta hay basándose en las reflectancias de las imágenes de satélite y de los datos auxiliares relacionados con el tipo de cubierta y que establecemos como datos de entrada o variables independientes. De esta manera, cada píxel de la imagen es clasificado en función de su respuesta espectral de diferentes bandas y fechas de las imágenes del satélite con una aceptable cobertura nubosa.

### 2.2. Datos de entrada e imágenes de satélite

Como núcleo de los datos de entrada se han usado imágenes de satélite de moderada resolución. En 2016 se usaron imágenes de Deimos-1, LandSat 8 y Sentinel-2A, con distintas resoluciones espaciales (22 m, 30 m y 10 m de resolución espacial respectivamente) obteniendo un mapa final de 20 m. Sin embargo, en el 2017 y 2018 se usaron únicamente imágenes del par de satélites Sentinel-2 (S2A y S2B) de la ESA, lo que permitió mejorar considerablemente la resolución del mapa de clasificación a 10 m. Además, es recomendable disponer de más datos auxiliares para ayudar al algoritmo de clasificación o determinar la clase satisfactoriamente. Los datos complementarios que se han usado son: Información del uso de suelo proveniente de SIGPAC para cada año, modelo digital de elevaciones y sus derivados de pendientes y orientación, un mapa de precipitación media (1981-2010) y datos de altura de la vegetación y de fracción de cubierta provenientes de LIDAR. Todos estos datos son las variables explicativas del modelo predictivo.

Como datos de referencia para el entrenamiento del algoritmo se emplean casos conocidos procedentes de la declaración única de los beneficiarios de ayudas de la Política Agraria Comunitaria, ya que en términos generales se consideran una fuente de datos fidedigna por su alto porcentaje de cumplimiento. Además, se emplean otras bases de datos de cobertura del suelo disponibles en la región.

## 3. RESULTADOS Y DISCUSIÓN

### 3.1. Capacidad de discriminación de cultivos de invierno en secano y regadío

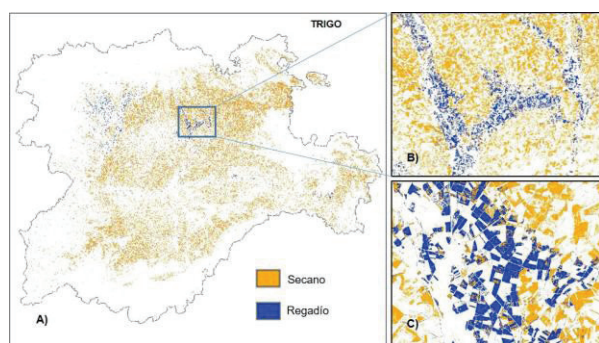
Para evaluar la calidad de la identificación de cubiertas, se elige otro conjunto de casos procedentes de las mismas fuentes que el entrenamiento, denominado conjunto de validación. Este conjunto es independiente al del proceso de aprendizaje.

La precisión del método de clasificación presentado en este estudio fue evaluada por métricas clásicas como son la precisión global, errores de comisión y omisión y el índice kappa (Caetano, 2009). La exactitud total y el coeficiente Kappa obtenidos para los mapas clasificados de los tres años de estudio fueron: 84,0% y 0,80, en el año 2016, 79,5% y 0,75, en 2017 y 83,7% y 0,81 en 2018, respectivamente (ver Tabla 1). Éstas precisiones globales se refieren a todas las clases incluidas en el mapa (más de un centenar), que incluyen además de cultivos, superficies forestales y zonas semi-naturales desarboladas (Gómez *et al.* 2018). En la tabla 1 se muestran las métricas de precisión y el porcentaje de la superficie obtenida en el mapa, de las principales clases de cultivo, tanto de secano como de regadío, para los tres años considerados en este trabajo. El índice F-Score es una media armónica de los errores de comisión y omisión.

Las medidas de exactitud de los años 2016 y 2018 son significativamente superiores a las del 2017, lo que puede deberse a diversos factores. En 2016 y 2018 hay una mayor cantidad de imágenes disponibles respecto al 2017, tal y como se explica en la sección 2.2., que mejoran el seguimiento de la evolución temporal de los cultivos y por lo tanto facilitan su identificación. Además, en el 2017, la disponibilidad de imágenes también fue inferior al 2018 porque las imágenes del satélite S2B no estuvieron disponibles hasta julio. Por otro lado, en cultivos de secano, aproximadamente el 83% de la tierra cultivada de Castilla y León, se observan precisiones más altas para estos dos años. Sin embargo, en cultivos de regadío las diferencias no siguen esa tendencia. De hecho, las medidas individuales de algunos cultivos de especial relevancia en la región, como el maíz, girasol, patatas y remolacha en regadío mejoran considerablemente en el año 2017, mostrando un mayor poder de discriminación de cultivos en regadío (ver Tabla 1). Esto lo podemos atribuir a las diferentes condiciones meteorológicas que presentaron los tres años de estudio, considerando 2016 y 2018 como un año húmedo y ligeramente húmedo, respectivamente, mientras que 2017 fue catalogado como un año muy seco, sufriendo una sequía que provocó algunos recortes de agua de riego en periodos clave del crecimiento de los cultivos. Por lo tanto, la capacidad de discriminación de cultivos en regadío mediante la metodología propuesta parece mejorar en años de sequía. Este resultado es el esperado ya que en condiciones de sequía la comparación entre el estado de la vegetación de un cultivo en regadío y uno en secano se muestran más significativas.

Es importante tener en cuenta que los cultivos de verano en régimen de regadío tienen una más fácil identificación por las condiciones agroclimáticas presentes en nuestra región, dado que cultivos como el maíz, la

remolacha y la patata, entre los más representativos, sólo se cultivan bajo regadío en dicha época del año. Sin embargo, discriminar los cultivos de invierno que se presentan tanto en secano como en regadío constituye el verdadero desafío, ya que la diferencia de respuesta espectral entre ambos regímenes en invierno no es tan acusada como en la época estival, cuando las condiciones de estrés hídrico hacen que las plantas en cultivos de secano reaccionen e inhiban su actividad fotosintética para no perder agua. Por lo tanto, desde el enfoque de la teledetección solventar esta situación es realmente complicado. Sin embargo, esta metodología ha resultado eficaz para este fin, tal y como puede verse en la Figura 2, en la que se muestra la discriminación del régimen de regadío en uno de los cultivos de invierno más representativos de la comunidad de Castilla y León, el trigo, con aproximadamente el 11% del área total de la región (ver Tabla 1). Las métricas de precisión obtenidas para el cultivo de trigo, son similares a las encontradas en la literatura (Conrad, 2010), alcanzando un índice de kappa del 0,82.



**Figura 2.** A) Parcelas de trigo de secano (naranja) y trigo de regadío (cyan) en el mapa clasificado de 2017; B y C) Detalle de las parcelas de trigo.

La discriminación entre regadío y secano que ofrece este mapa, es de especial relevancia en el contexto de la PAC, ya que interesa no sólo la identificación de cultivos sino también la verificación del cumplimiento de algún tipo de condicionalidad, ya que hay ayudas asociadas a determinados cultivos que son sólo concedidas en un determinado sistema de explotación, ya sea secano o regadío. Tal es el caso, de las ayudas asociadas a cultivos proteicos, en concreto para la alfalfa, cuyo requisito específico es que sea de secano.

Cabe destacar que hasta el momento solo se ha empleado esta técnica para la discriminación de ciertos cultivos herbáceos en regadío. Los prados y los cultivos leñosos, entre los cuales solo el viñedo es significativo en Castilla y León, no han sido evaluados por diversas razones. El viñedo en Castilla y León está orientado exclusivamente a la producción bajo figuras de calidad que imponen restricciones a la producción. En estas condiciones el riego es utilizado de forma deficitaria con el objeto de asegurar la calidad de la uva y por tanto las diferencias existentes entre el secano y el regadío son menores. Además, en el cultivo de viñedo la ratio suelo/planta es elevado y por lo tanto es más complicado detectar cambios del estado hídrico del cultivo por teledetección.

**Tabla 1.** Medidas de superficie y precisión de los principales cultivos de secano y de regadío de los mapas de clasificación 2016, 2017 y 2018.

Cultivo	nº	2016		2017		2018		
		Area mapa (%)	F-Score	Area mapa (%)	F-Score	Area mapa (%)	F-Score	
Cultivos de Secano	Trigo	1	11,95	87,12	10,56	77,11	10,47	84,78
	Cebada	5	9,82	89,17	9,56	77,53	9,16	85,52
	Centeno	6	0,71	53,54	0,72	44,78	1,17	61,57
	Avena	8	0,72	47,89	1,64	47,45	2,10	61,15
	Triticale	13	0,48	38,22	0,08	18,48	0,16	34,74
	Barbecho	21	1,92	82,17	4,97	89,17	4,50	93,57
	Girasol	33	3,87	92,32	2,83	89,92	3,42	92,67
	Colza	35	0,47	85,74	0,28	65,07	0,18	60,94
	Guisantes	40	0,38	80,46	0,52	64,73	0,47	70,77
	Habas	41	0,10	48,33	0,06	43,28	0,01	56,54
	Altramuces	43	0,62	1,10	0,13	30,61	0,08	45,82
	Garbanzos	50	0,19	54,06	0,09	61,12	0,08	71,57
	Lentejas	51	0,10	74,67	0,12	67,03	0,10	83,95
	F.Veza	52	1,11	77,99	1,53	61,48	1,53	75,52
	Yeros	53	0,14	63,14	0,11	48,64	0,08	52,92
	Alubias	54	0,13	78,07	0,16	84,18	0,11	88,85
	Alfalfa	60	0,74	82,30	0,92	75,30	1,01	85,05
	Forrajeras	61	0,63	38,36	0,19	34,15	0,13	40,54
	Esparceta	67	0,44	55,48	0,26	49,51	0,11	56,78
	Festuca	68	0,04	21,89	0,04	27,32	0,02	51,25
Raygrass	69	1,15	54,16	0,54	44,80	0,42	53,34	
Titarros	240	0,26	52,73	0,11	48,98	0,06	59,33	
Superficie Total (ha)			3.323.979		3.280.996		3.335.509	
Cultivos de Regadío	Maiz	4	1,29	96,81	1,06	97,26	1,12	98,18
	Sorgo	7	0,05	2,51	0,02	45,43	0,02	60,38
	Remolcha	82	0,27	94,90	0,28	93,94	0,20	92,32
	Patatas	94	0,26	84,98	0,20	88,14	0,14	86,73
	Adormidera	102	0,06	65,94	0,03	68,73	0,01	63,41
	Fresas	119	0,03	81,51	0,03	85,64	0,03	87,25
	Calabaza	155	0,02	0,40	0,01	64,44	0,01	80,18
	Ajo	159	0,09	46,61	0,06	50,43	0,05	57,79
	Cebolla	160	0,04	47,24	0,05	56,80	0,03	75,65
	Zanahoria	161	0,05	75,44	0,05	71,98	0,17	75,65
	Puerro	169	0,04	40,59	0,03	65,75	0,02	66,43
	Huerta	193	0,07	28,52	0,06	40,53	0,05	44,81
	Trigo Regadío	70	0,38	62,71	0,63	70,09	0,66	77,84
	Cebada Regadío	71	0,16	52,75	0,12	35,93	0,14	49,67
Alfalfa Regadío	72	0,44	77,52	0,27	76,73	0,38	84,92	
Girasol Regadío	73	0,23	71,80	0,19	73,43	0,11	63,19	
Centeno Regadío	75	0,08	25,52	0,10	32,56	0,06	61,51	
Avena Regadío	76	0,17	50,86	0,11	50,27	0,10	62,65	
Triticale Regadío	77	0,05	18,52	0,07	36,04	0,03	66,52	
Superficie Total (ha)			353.498		312.563		314.120	
Precisión global			84,0%		79,5%		83,7%	
Índice Kappa			0,80		0,75		0,81	

### 3.2. Estimación de la superficie de cultivos herbáceos realmente regados en la región

Conocer la superficie total de cultivos verdaderamente regados en cada campaña es una asignatura pendiente. Las estadísticas agrarias oficiales se nutren de las declaraciones de la PAC en las que los agricultores tienen la posibilidad de declarar cada año sus parcelas en regadío. Esta declaración obedece de forma mayoritaria a las características propias asignadas a la parcela en el SIGPAC, independientemente de que efectivamente se riegue a lo largo de la campaña. Por lo tanto, la superficie estimada mediante estadísticas derivadas de las declaraciones de la PAC debe tener la consideración de regable. Sin embargo, la superficie que obtenemos mediante teledetección refiere a un regadío “efectivo” (superficie regada), donde las imágenes de satélites detectan de forma objetiva los píxeles que se comportan como regados.

En dicho contexto, un valor añadido de este mapa, es que podemos obtener una estimación de la superficie total que hay en nuestra región dedicada a regadío. Así, la superficie mínima de regadío obtenida para 2016 ha sido

de 353.498 ha. A modo de ejemplo observamos que esta superficie de regadío “efectiva” obtenida es menor que la declarada en la PAC considerando las mismas clases (399.686 ha). En 2017 se detectaron 312.563 ha y en 2018 fueron 314.120 ha (ver Tabla 1). Teniendo en cuenta que el 2017 transcurrió como el año más seco y antes de terminar el ciclo de algunos de los cultivos de verano como el maíz y la remolacha, hubo restricciones de agua para el riego por parte de las confederaciones hidrográficas, parece razonable que la superficie detectada como regadío haya disminuido. Esto en sí mismo no disminuye la superficie de cultivos en regadío, pero sí que dificulta la detección de los mismos por teledetección en determinadas fechas claves para su detección. Además, no sorprende que no haya aumentado la superficie en regadío detectada en 2018, ya que este año fue considerado un año húmedo y la necesidad de regar no fue tan acuciante como los dos años anteriores.

#### 4. CONCLUSIONES

Las medidas de precisión obtenidas demuestran que los mapas de cobertura del suelo presentados en este trabajo son una herramienta muy útil que ayudará a las administraciones en la toma de decisiones sobre gestión del agua y la planificación de las infraestructuras de riego.

La metodología presentada demuestra ser eficaz para detectar los principales cultivos herbáceos en regadío de nuestra región. Sin embargo, se han observado algunas limitaciones identificando cultivos minoritarios en riego así como al inferir la superficie total de riego con un simple conteo de píxeles, en lugar de aplicar otros métodos, como un ajuste del área por regresión, (Gallego, 2004). Además, hay que considerar que el gran detalle temático que se ha empleado en la clasificación de estos 3 años (más de 100 clases categorizadas) hace que la precisión global alcanzada no sea en ninguno de los tres años la deseable. Por lo tanto, un ejercicio futuro es agrupar las clases para alcanzar al menos una precisión global del 85%, establecida como óptima, y un incremento en las métricas individuales de cada clase.

La determinación de la superficie verdaderamente regada sigue siendo un tema de complicada resolución debido a la gran diversidad de situaciones y a la diferente concepción del problema según su enfoque, que en muchos casos no está claramente diferenciada de la superficie regable. En cualquier caso, estas metodologías son una interesante herramienta para la monitorización de las zonas regables en lo referente a la intensidad de cultivo y uso del agua con escala de detalle, así como para la identificación de aquellas parcelas que se riegan careciendo de derecho de uso del agua y aquellas que, aun teniendo derecho de uso, no hacen uso del recurso.

#### 5. AGRADECIMIENTOS

El desarrollo de este trabajo ha sido realizado en el marco del proyecto “SENSAGRI” que ha recibido financiación del Programa de Innovación e Investigación Horizon 2020 de la Unión Europea (Contrato acuerdo de subvención nº 730074). Nuestro agradecimiento también a José L. Bengoa de la Consejería de Fomento y Medio Ambiente de la Junta de Castilla y León, por proporcionarnos

los datos de medio natural usados en la clasificación a nivel regional.

#### 6. BIBLIOGRAFÍA

- Ambika, A. K., Wardlow, B., & Mishra, V. 2016. Remotely sensed high resolution irrigated area mapping in India for 2000 to 2015. *Scientific data*, 3, 160118.
- Conrad, C., Fritsch, S., Zeidler, J., Rücker, G., & Dech, S. 2010. Per-field irrigated crop classification in arid Central Asia using SPOT and ASTER data. *Remote Sensing*, 2(4), 1035-1056.
- Caetano, M., A. Araújo, A. Nunes, V. Nunes and M. Pereira, “Accuracy assessment of the CORINE Land Cover 2006 map of Continental Portugal”. Technical Report, Instituto Geográfico Português, 2009.
- Del Blanco Medina, V. & Nafría García, D.A. 2015. Mapa de cultivos y superficies naturales de Castilla y León. XVI Congreso de la Asociación Española de Teledetección. Teledetección: Humedales y Espacios Protegidos. pp. 32-35.
- Gallego, F.J. 2004. Remote sensing and land cover area estimation, *International Journal of Remote Sensing*, 25:15, 3019-3047
- Gómez, V.P., Medina, V.D.B., Bengoa, J.L., & García, D.A.N., 2018. Accuracy Assessment of a 122 Classes Land Cover Map Based on Sentinel-2, Landsat 8 and Deimos-1 Images and Ancillary Data. In *International Geoscience and Remote Sensing Symposium*, IEEE. pp. 5453-5456.
- Quinlan, J.R. 1992. C4.5: programs for machine learning, San Mateo, CA: Morgan Kaufmann.