

Mapa de cultivos y superficies naturales de Castilla y León

Vicente del Blanco Medina⁽¹⁾, David A. Nafría García⁽¹⁾

⁽¹⁾ Instituto Tecnológico Agrario de Castilla y León, Ctra. Burgos-Portugal km 119, 47071-Valladolid, España. e-mail: blamedvi@itacyl.es, nafgarda@itacyl.es

Resumen: Se describe la metodología para la elaboración del *Mapa de cultivos y superficies naturales de Castilla y León*. Se trata de un mapa de ocupación del suelo obtenido a través de imágenes de satélite Deimos-1 y Landsat-8 mediante algoritmos de aprendizaje por computador. Con esta metodología se han creado los mapas de los años 2011, 2012, 2013 y 2014. El objetivo es disponer de una cartografía de ocupación del suelo con periodicidad anual que represente los cambios en los cultivos herbáceos anuales, mayoritarios en la Comunidad Autónoma. Aunque se trata de una aplicación básica de la teledetección, es un claro ejemplo de aplicación operativa.

Palabras clave: Clasificación supervisada, agricultura, teledetección, cultivos herbáceos, aprendizaje por computador.

Castilla y León crops and natural surfaces layer

Abstract: *This paper describes the methodology to create the Castilla y León crop and natural surfaces layer based on Deimos-1 and Landsat-8 satellite images using machine learning algorithms. The layer was created in 2013 and versions for 2011, 2012 and 2014 were made subsequently. The target is to update the layer yearly in order to include changes in grain crops. These crops represent the majority of the agricultural land in the region. The crops identification accuracy obtained is in general notably higher compared to natural surfaces. This project is an example of the operational use of remote sensing.*

Keywords: *Supervised classification, agriculture, remote sensing, grain crops, machine learning.*

1. INTRODUCCIÓN

Castilla y León es la mayor Comunidad Autónoma de España en términos de superficie con 94.226 km², y representa una quinta parte del Estado. La mayor parte de su territorio está compuesto por superficies de cultivos herbáceos extensivos o de vegetación natural. Tradicionalmente las bases de datos de ocupación del suelo (Mapa de cultivos y aprovechamientos de España, Sistema de Información sobre Ocupación del Suelo de España -SIOSE-, Corine Land Cover) han clasificado las zonas de cultivos herbáceos extensivos en categorías homogéneas sin especificar el cultivo o especie. Ocasionalmente se ha considerado el tipo de explotación en secano o en regadío como un atributo reseñable, que habitualmente era una fuente de conflicto. La periodicidad de actualización plurianual, unida a la dificultad de discriminar los cultivos, hacía inviable una clasificación detallada de los mismos. El interés por disponer de cartografía de ocupación del suelo de Castilla y León con el suficiente detalle en las categorías agrícolas, y con actualización anual, promovió el desarrollo de la metodología que aquí se presenta, basada en el proyecto *Crop Data Layer* del Departamento de Agricultura de Estados Unidos (Boryan, 2011).

Aunque el objetivo inicial del proyecto es disponer de cartografía de cultivos, para evitar errores de comisión, la metodología requiere que todas las tipologías posibles de ocupación del suelo estén representadas en el entrenamiento. Por tanto, el resultado final, que se muestra en la figura 1, es una clasificación general de

ocupación del suelo que incluye las superficies naturales. El nombre del proyecto -Mapa de cultivos y superficies naturales de Castilla y León- demuestra la vocación inclusiva de todo el territorio.

El mapa se ha generado para los años 2011, 2012, 2013 y 2014, aunque los resultados que aquí se presentan son los del 2014. El hecho de no disponer de imágenes Landsat-8 en los años 2011 y 2012 hace que la exactitud del resultado sea ligeramente inferior.

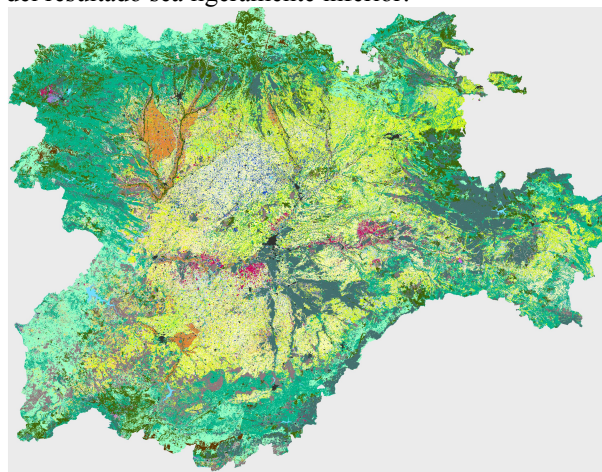


Figura 1. Vista general del Mapa de cultivos y superficies naturales de Castilla y León en 2014. Escala 1:5.500.000

2. MATERIAL Y MÉTODOS

En la metodología desarrollada además de usar imágenes de los satélites Deimos-1 y Landsat-8, se

permite la introducción de datos auxiliares que ayudan al algoritmo de clasificación a determinar la clase.

2.1. Imágenes de satélite

Para el mapa de 2014 se adquirieron 21 imágenes Deimos-1 (Tabla 1), que forman 8 coberturas completas, y 80 escenas Landsat-8 (Tabla 2) para cubrir la totalidad del territorio de la Comunidad Autónoma. Las imágenes Deimos-1 proceden del Plan Nacional de Teledetección de España (PNT) en el que participan el Instituto Tecnológico Agrario de Castilla y León (ITACyL), la Confederación Hidrográfica del Duero (CHD) y el Instituto Geográfico Nacional (IGN). Estas imágenes son suministradas con una resolución espacial de 20 m, geo-referenciadas en el Sistema de Referencia ETRS89 en proyección UTM-30N con un Error Medio Cuadrático (EMC) por componente menor o igual que un píxel.



Figura 2. Coherencia en la georreferenciación de imágenes Landsat-8 y Deimos-1 en Zona de Villanubla (Valladolid) Escala 1:25.000

Las imágenes Deimos-1 fueron seleccionadas directamente del catálogo web de Deimos Imaging intentando cubrir de la forma más eficiente posible las épocas clave en el desarrollo fenológico de los cultivos. El sensor del satélite no tiene un esquema de captura predefinido ni una dimensión de escena específica, por lo que la selección de las imágenes que afectan al territorio de Castilla y León es relativamente compleja. Las escenas Landsat-8 fueron elegidas dentro del visor EarthExplorer del USGS siguiendo criterios de complementariedad con las imágenes Deimos-1 y una reducida cubierta nubosa. Todas las escenas Landsat-8 (geo-referenciadas en WGS84/UTM con una precisión mejor o igual a 12m Circulo de Error Probable al 90%) fueron preprocesadas para ser llevadas al marco geométrico definido en el proyecto, marcado por las coberturas Deimos-1. Para ello, las bandas 2, 3, 4, 5, 6, y 7 de las escenas se sometieron a un proceso de pansharpening con la banda 8 utilizando las utilidades de tratamiento de datos geoespaciales GDAL (Geospatial Data Abstraction Library, Open Source

Geospatial Foundation) y Spectral Transformer for Landsat-8 imagery (Geosage).

La figura 2 muestra el grado de coherencia entre las escenas Landsat-8 preprocesadas y las imágenes Deimos-1 procedentes del PNT.

Las imágenes no fueron sometidas a procesos de escalado ni corrección atmosférica. Tampoco se realizó limpieza de zonas nubosas.

2.2. Datos complementarios

Para ayudar al clasificador en la determinación de la clase final, además de las imágenes de satélite se emplearon las siguientes capas:

- Clasificación de usos SIGPAC del año correspondiente.
- Precipitación media anual del periodo 1981-2010 procedente del Atlas agroclimático de Castilla y León.
- Modelo digital de elevaciones derivado del Plan Nacional de Ortofotografía Aérea (PNOA).
- Modelo digital de pendientes procedente del MDE PNOA.
- Modelo digital de orientaciones. Procedente del MDE PNOA.

Tabla 1. Fechas de las imágenes Deimos-1 utilizadas para la creación del Mapa de 2014.

	Fecha
Deimos-1	12/03/2014; 16-17/04/2014; 13-16/05/2014; 05/06/2014; 18/06/2014; 7-17/07/2014; 6-16/08/2014; 24-27/10/2014

Tabla 2. Escenas del satélite Landsat-8 y fechas utilizadas para la creación del Mapa de 2014.

Path	Row	Fecha
200	31	07/03/2014;10/05/2014;11/06/2014; 13/07/2014;14/08/2014;30/08/2014; 01/10/2014;17/10/2014
201	30-31-32	14/03/2014;01/05/2014;17/05/2014; 02/06/2014;18/06/2014;05/08/2014; 21/08/2014;24/10/2014
202	30-31-32	05/03/2014;08/05/2014;09/06/2014; 11/07/2014;27/07/2014;12/08/2014; 28/08/2014;13/09/2014
203	30-31-32	23/01/2014;12/03/2014;13/04/2014; 15/05/2014;16/06/2014;19/08/2014; 04/09/2014;22/10/2014

2.3. Entrenamiento

La clasificación se ha realizado mediante árboles de decisión generados con técnicas de aprendizaje por computador a partir de un conjunto de datos de entrenamiento.

La representatividad del entrenamiento sobre toda la casuística posible es clave para la calidad del producto final. Debido a la extensión del territorio de Castilla y León existen multitud de clases de ocupación del suelo posibles. Para este proyecto no se ha realizado un trabajo de campo específico, por lo que el conocimiento en profundidad de las características del territorio y de las bases de datos empleadas es fundamental para conseguir un conjunto de datos de entrenamiento de la mejor calidad posible.

El entrenamiento se generó a partir de las siguientes bases de datos:

- Base de datos de solicitudes de ayudas de pago único de la Política Agraria Comunitaria del año en cuestión.
- Inventario Forestal Nacional de España.
- Cartografía de habitats de Red Natura de la Consejería de Fomento y Medio Ambiente de la Junta de Castilla y León.
- Sistema de Información sobre Ocupación del Suelo de España (SIOSE).
- Digitalizaciones sobre ortofoto para casos concretos no incluidos en las bases de datos mencionadas.

El número total de clases definidas en la clasificación es de 101. Esta cifra incluye discriminaciones a nivel de especie en superficies agrícolas (diferenciando secano y regadío en algunos casos) y forestales. A partir de los recintos de entrenamiento se crea una muestra aleatoria de 100.000 casos. El número de casos seleccionados según la clase es proporcional a la extensión geográfica, aunque se establecen límites para asegurar la representatividad de las clases minoritarias.

2.3. Clasificador

A partir de los 100.000 casos se construyeron árboles de decisión con el software de minería de datos C5.0 en su versión libre (con licencia GNU/GPL) sobre Linux. El C5.0 es una mejora del modelo C4.5 desarrollado por Quinlan (1993).

Los árboles de decisión se crean en un único paso para toda Castilla y León, por tanto se da la circunstancia de que existen zonas sin imagen (especialmente con las imágenes Landsat-8) y casos en los que el territorio está afectado por nubes. No se realizaron clasificaciones independientes por áreas homogéneas porque los árboles de decisión incluyen estas posibilidades.

Los árboles de decisión generados por C5.0 son aplicados a todos los píxeles del conjunto de imágenes de satélite e imágenes auxiliares, obteniendo en un único paso la clasificación de todo el territorio de la Comunidad Autónoma.

2.4. Postprocesado

Con el fin de proporcionar un mapa de fácil interpretación para los usuarios finales, se realizan una serie de post-procesos entre los que destacan: una simplificación de la leyenda de 101 a 28 clases y una limpieza del efecto de “sal y pimienta” mediante un

cribado con las herramientas de GDAL. En la figura 3 se muestra la leyenda de cultivos definitiva y en las figuras 4 y 5 sendos ejemplos del resultado en dos zonas.



Figura 3. Leyenda de clases del mapa final

2.5 Comparación de resultados

Para la comparación de resultados se emplearon las clásicas matrices de confusión, calculadas con el software libre GRASS. Se usó como capa de contraste todo el conjunto de datos del que se extrajo la muestra de entrenamiento. Hay que tener en cuenta que solo el 0,56 % de los píxeles de esta capa fueron utilizados para la creación de los árboles de decisión.

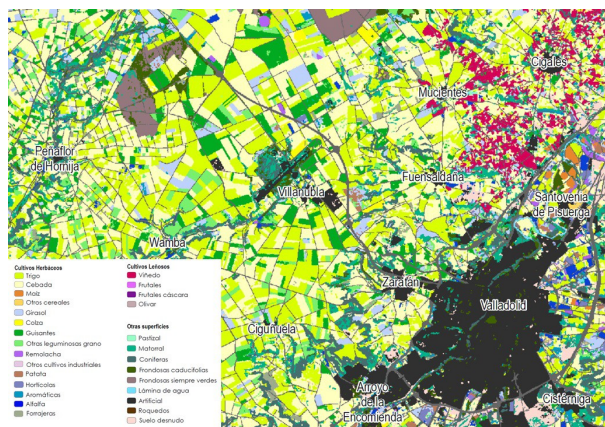


Figura 4. Extracto del Mapa de cultivos y superficies naturales. Zona de Villanubla (Valladolid)
Escala 1:375.000

Para las clases de cultivo, se realizó un control adicional utilizando la base de datos vectorial de los segmentos de la Encuesta sobre Superficies y Rendimientos de Cultivos de España (ESYRCE) del MAGRAMA y la Consejería de Agricultura y Ganadería de la Junta de Castilla y León. Cada entidad vectorial se contrastó con el valor de los píxeles de la clasificación cuyo centro estaba incluido en la forma de la entidad vectorial. Este conjunto de datos es completamente independiente a la clasificación al no ser usado en la definición de las zonas de entrenamiento.

3. RESULTADOS

3.1. Exactitud en la clasificación con respecto al entrenamiento inicial.

La Tabla 3 muestra el grado de coincidencia entre los píxeles de entrenamiento y la clasificación final. Nótese que la tabla diferencia entre la explotación en secano y regadío. En una clasificación sin esta discriminación el porcentaje de acierto sube significativamente en estos cultivos.

Tabla 3. Grado de coincidencia entre los píxeles de la capa de entrenamiento y el Mapa

Clase	Acierto %	Clase	Acierto %
Frutales	99	Girasol	91
Urbano-viales	99	Otros cereales regadío	90
Olivar	99	Alfalfa regadío	90
Aromáticas	99	Colza	90
Hortícolas	99	Cebada	87
Frutales cascara	99	Trigo	85
Viñedo	98	Girasol regadío	85
Lamina agua	98	Roquedos	80
Remolacha	96	Frondosas siempre verdes	76
Maíz	95	Alfalfa	73
Patatas	94	Guisantes	72
Otras leguminosas grano	93	Suelo desnudo	68
Pastizal	92	Trigo regadío	68
Matorral	91	Cebada regadío	62
Coníferas	91	Forrajeras	61
Frondosas caducifolias	91	Otros cereales	44

Hay que tener en cuenta que tanto en los casos de entrenamiento como en los de control pueden existir errores, al no proceder de una verdad terreno capturada ex profeso.

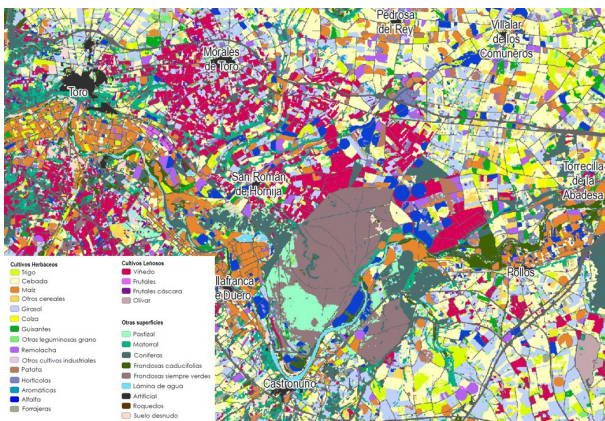


Figura 5. Extracto del Mapa de cultivos y superficies naturales. Zona de Toro (Zamora) Escala 1:375.000

3.2. Exactitud en la clasificación con respecto a ESYRCE.

La Tabla 4 muestra el grado de coincidencia entre los píxeles clasificados en las zonas de control ESYRCE con respecto al Mapa de cultivos y superficies naturales de Castilla y León.

Tabla 4. Grado de coincidencia entre los píxeles clasificados en las zonas de control ESYRCE y el Mapa

Clase	Acierto %	Clase	Acierto %
Frutales	100	Trigo	84
Viñedo	98	Alfalfa	82
Olivar	94	Otros cultivos industriales	78
Maíz	93	Coníferas	67
Remolacha	92	Guisantes	66
Aromáticas	92	Hortícolas	51
Girasol	91	Otras leguminosas grano	40
Cebada	87	Otros cereales	35
Colza	85	Frutales cascara	34
Patatas	85	Forrajeras	28

4. DISCUSIÓN

En aquellos cultivos en los que se da una cobertura vegetal homogénea los resultados arrojan un porcentaje de acierto elevado. En aquellos casos, como las superficies de vegetación natural, en los cuales existe variabilidad dentro de la misma clase, este porcentaje se reduce de forma significativa. Cabe destacar el hecho de que el método permite diferenciar cultivos de fenología muy similar como son el trigo y la cebada.

Los resultados ponen de manifiesto la necesidad de contar con un entrenamiento de calidad, amplio y bien distribuido, así como de una cantidad de imágenes elevada en las épocas de cambios fenológicos. Es necesario mejorar el entrenamiento en las zonas de vegetación natural.

El mapa creado constituye una interesante herramienta en el ámbito de la gestión del medio agrario, para ser aplicado al control de las ayudas por superficie, la realización de estudios sobre rotaciones de cultivos, estadísticas agrarias de diversa índole, estimaciones de consumos de agua, detección de irregularidades en concesiones de riego, etc. Desde el punto de vista medioambiental, este mapa de periodicidad anual podría ser una herramienta para el seguimiento y control de los cambios de cobertura del suelo.

El próximo lanzamiento del satélite Sentinel-2 hace suponer una gran oportunidad de mejora del producto debido a su resolución espacial y temporal.

Las capas generadas en el marco de este proyecto pueden descargarse libremente a través de la web del Atlas Agroclimático de Castilla y León (<http://atlas.itacyl.es>)

5. BIBLIOGRAFÍA

- Quinlan R (1993). C4.5: *Programs for Machine Learning*. Morgan Kaufmann Publishers.
- Boryan, C., Yang, Z., Mueller, R., Craig, M., 2011. *Monitoring US agriculture: the US Department of Agriculture, National Agricultural Statistics Service. Cropland Data Layer Program*. Geocarto International 26 (5), 341–358.