



# 7º CONGRESO FORESTAL ESPAÑOL

**Gestión del monte: servicios  
ambientales y bioeconomía**

26 - 30 junio 2017 | Plasencia  
Cáceres, Extremadura

---

---

7CFE01-034

---

---

Edita: Sociedad Española de Ciencias Forestales  
**Plasencia. Cáceres, Extremadura. 26-30 junio 2017**  
**ISBN 978-84-941695-2-6**

© Sociedad Española de Ciencias Forestales

## Identificación de formaciones desarboladas a partir de teledetección en Castilla y León

BENGOA, J. L.<sup>1</sup>, DEL EGIDO, F.<sup>2</sup> y MOLINA, C.<sup>3</sup>

<sup>1</sup> Junta de Castilla y León

<sup>2</sup> Dpto. de Biodiversidad y Gestión Ambiental. Universidad de León

<sup>3</sup> Actividades, Estudios y Proyecto en el Medio Ambiente, S.L.

### Resumen

La cartografía de las cubiertas vegetales es un instrumento básico para diferentes tareas de gestión administrativa del medio natural. Los elevados costes asociados a estas cartografías hacen conveniente buscar alternativas para reducirlos.

Las técnicas de clasificación de cubiertas a partir de teledetección pueden cumplir esta función para determinadas cubiertas, en particular para las que cubren extensiones importantes, considerando, en todo caso, categorías que sean suficientemente identificables por estas técnicas. Esto obliga a integrar en cada categoría comunidades que sean parecidas en términos de respuesta espectral, estructura y fenología (depende de las variables clasificadoras que se consideren).

Tanto las fuentes de información (imágenes de satélite y otras como Lidar e información orográfica y climática), como la calidad del material de entrenamiento, la metodología de procesado de datos y la elección de las categorías a identificar son fundamentales para que los resultados sean satisfactorios. En este trabajo se recogen y valoran los resultados obtenidos para Castilla y León, considerando 39 tipos de cobertura desarbolada.

### Palabras clave

Clasificación-de-cubiertas, *data-science*, árboles-de-decisión, matorrales, pastizales

### 1. Introducción

El estudio y cartografía de la vegetación, los biotopos, los hábitats o la cobertura del suelo en general han evolucionado mucho en las últimas décadas de mano de la fotografía aérea, los sistemas de posicionamiento y las herramientas para el manejo de la información geográfica. Sin embargo, la fotointerpretación sigue haciéndose manualmente: es una tarea compleja que se considera difícilmente sustituible o automatizable porque se apoya no sólo en “lo que se ve” en las imágenes aéreas, sino también en el conocimiento del territorio.

Esta reflexión, que era casi un axioma, empieza a flaquear cuando entran en juego técnicas avanzadas de análisis de datos (*machine learning*), cuando se cuenta con buenas y abundantes fuentes de información (teledetección, Lidar, etc.) y cuando se juntan estos dos ingredientes con un buen conocimiento del territorio y su cubierta vegetal.

Bajo esta premisa se presenta en este Congreso una comunicación titulada “Clasificación semiautomática de cubiertas naturales arboladas en Castilla y León”, que presenta los resultados obtenidos en la clasificación de cubiertas naturales y seminaturales mediante técnicas de clasificación supervisada apoyada en teledetección, para el conjunto de Castilla y León. En dicha comunicación no se analiza el tratamiento que se ha dado a las coberturas no arboladas, de las que nos ocupamos en ésta.

Tanto las fuentes de información como la calidad del material de entrenamiento, la metodología de procesado de datos y la elección de las categorías son fundamentales para que los resultados sean satisfactorios. En todo caso, estas técnicas tienen limitaciones que chocan de frente con la heterogeneidad de los paisajes naturales y seminaturales. Esto es especialmente patente para las

formaciones desarboladas, que constituyen un colectivo particularmente difícil de clasificar (en todos los sentidos).

Este trabajo surgió en el ámbito técnico de gestión del territorio, con unos objetivos que se distancian sensiblemente de los habituales en el estudio del paisaje vegetal del ámbito científico y educativo. Esto obliga a poner el foco en los aspectos de la cubierta vegetal que más condicionan la gestión administrativa del territorio y a trabajar con una resolución espacial acorde con las unidades territoriales que se manejan en gestión (dependientes de la propiedad del terreno, de las cubiertas - en la medida en que condicionan la gestión administrativa- y de los perímetros habituales de actuación). Los recintos SIGPAC son una referencia fundamental que viene de la gestión de la Política Agrícola Común, pero que podría extenderse a otras tareas de gestión del territorio con las oportunas adaptaciones.

Este planteamiento se distancia sensiblemente de la cartografía destinada a la planificación, con menor resolución espacial como ha ocurrido con las diferentes ediciones del Mapa Forestal de España a escalas 1:1000.000, 1:200.000 o 1:50.000. La escala 1:25.000 puede considerarse frontera entre la planificación y la gestión.

Este trabajo se enmarca dentro de una línea de colaboración entre el Instituto Tecnológico Agrario de Castilla y León (ITACYL) y la Consejería de Fomento y Medio Ambiente de la Junta de Castilla y León que, a través de su Centro de Información Territorial, ha financiado la preparación del material de entrenamiento correspondiente a las formaciones desarboladas naturales y seminaturales para este proyecto.

## 2. Objetivos

El objetivo de este trabajo es presentar algunos avances en la clasificación semiautomática de cubiertas a partir de información disponible a escala regional como la procedente de teledetección, vuelos Lidar y demás información auxiliar.

En concreto, en esta comunicación se presenta el listado de formaciones naturales y seminaturales no arboladas propuesto para este proyecto, el material de entrenamiento elaborado para una primera fase del mismo y una valoración cualitativa y cuantitativa de los resultados obtenidos.

## 3. Metodología

### Resolución temática

El estudio y clasificación de la vegetación ha dado lugar a una ingente literatura científica que constituye el punto de partida para analizar la resolución temática que queremos y podemos gestionar. Sin embargo, a la hora de decidir el número de categorías (tipos de cubierta) con el que vamos a trabajar también debemos tener en cuenta tres circunstancias: (1) trabajamos con un sistema semiautomático de clasificación basado principalmente en teledetección, (2) necesitamos material de entrenamiento suficiente de cada una de las categorías que definamos y (3) este subproyecto forma parte de uno más amplio destinado a cartografiar las cubiertas en todo el territorio (no sólo áreas naturales y seminaturales), por lo que las categorías que consideramos se integran en una clasificación general de cubiertas del suelo.

Como punto de partida decidimos trabajar con 40 a 50 tipos de formación desarbolada (y un número parecido para las arboladas). El número total de categorías consideradas en el proyecto (versiones ejecutadas con información satelital de 2016) es de 142-145 e incluye tanto las cubiertas naturales y seminaturales como las agrícolas, las artificiales y otras.

Este número de categorías es muy inferior al número de comunidades vegetales definidas en Castilla y León por lo que es necesario hacer un esfuerzo de integración. Por ejemplo, de acuerdo a la cartografía detallada de hábitats elaborada dentro de los Lugares de Interés Comunitario en Castilla y León, hay definidas cerca de 370 comunidades vegetales básicas, de las cuales 255 corresponden a formaciones naturales o seminaturales no arboladas; el número de categorías se incrementa

notablemente si tomamos como referencia las asociaciones contempladas en el *checklist* sintaxonómico (Rivas-Martínez et al., 2001; Rivas-Martínez et al., 2002). Las dos referencias principales para la elaboración de la lista de categorías contempladas en este trabajo han sido el Mapa de Vegetación de Castilla y León (López Leiva et al., 2009) y los documentos que integran la Cartografía detallada de hábitats de Castilla y León (inédito) en la que han participado dos de los autores de esta comunicación.

Cada una de las categorías consideradas en este trabajo integra comunidades diversas que hemos agrupado porque previsiblemente son parecidas en términos de respuesta espectral, fenología y estructura (estas tres propiedades de las cubiertas están estrechamente relacionadas con las variables clasificatorias utilizadas). Irremediablemente, ha sido necesario agrupar comunidades con diferencias florísticas importantes e, incluso, con significación ecológica algo dispar. Siempre que ha sido posible hemos evitado mezclar hábitats que son de interés comunitario con los que no lo son. Más adelante se hacen algunas consideraciones adicionales relacionadas con la clasificación de las comunidades vegetales, los biotopos, los hábitats o las cubiertas en general.

### Resolución espacial

La resolución espacial depende fundamentalmente de las fuentes de información que se utilizan para la clasificación. En estos años anteriores hemos utilizado imágenes de los satélites Landsat y Deimos y optamos por una resolución (tamaño de píxel) de 20 x 20 m, de forma que se clasifica individualmente el tipo de cubierta de cada píxel. En teoría, a partir de finales de 2015 iban a estar disponibles las imágenes del Sentinel 2a, lo que iba a proporcionar imágenes con bastante frecuencia y resolución de 10 m. Sin embargo, en la práctica esta disponibilidad no ha sido suficiente, por lo que se ha mantenido la resolución de los 20 m en 2016.

El posible cambio de resolución de 20 a 10 m no es necesariamente una mejora sino que deberá ser valorado a la vista de los resultados. Debe tenerse en cuenta que las formaciones naturales y seminaturales son sistemas complejos, integrados por diversos elementos (ejemplares arbóreos, arbustivos, matas, hierbas, etc.) de forma que un tipo de cubierta sólo está bien representado cuando se abarca una superficie mínima que permita integrar todos sus elementos (o la mayoría de ellos). El siguiente ejemplo ilustra esta limitación: una dehesa vista desde lejos es una dehesa; si la vemos desde muy cerca dependiendo de donde miremos podemos ver un pasto herbáceo o un 'bosque' de encina. A diferentes escalas, esto ocurre en todas las comunidades vegetales y debe ser tenido en cuenta cuando se habla de la presencia o ausencia de una determinada comunidad, biotopo o hábitat. Cuanto más heterogéneo es un tipo de cubierta (en este sentido), mayor tiene que ser la superficie mínima en la que se puede definir y con la que se tiene que trabajar.

### Nomenclatura y clasificación

La cartografía de la cubierta vegetal identifica las comunidades vegetales, biotopos o cubiertas en general de cada porción de terreno de acuerdo a un modelo de datos previamente definido, que normalmente se apoya en nomenclaturas (taxonomías) de especies y/o de comunidades. Mientras que la nomenclatura de especies cuenta con criterios bastante consensuados en el ámbito científico, la de comunidades vegetales, biotopos y hábitats es diversa en cuanto a criterios y, por lo tanto, cuenta con numerosas propuestas terminológicas. La que más se ha desarrollado en España a lo largo del siglo XX ha sido la de la escuela sigmatista que alcanza un hito significativo con la publicación del *checklist* sintaxonómico (Rivas-Martínez et al., 2001; Rivas-Martínez et al., 2002).

La preeminencia de esta escuela fitosociológica no resta valor e interés a otras alternativas para la descripción del paisaje vegetal. Algunas de estas alternativas ganan en flexibilidad evitando las rigideces de la metodología de Braun-Blanquet lo que les permite describir de forma más fiel o realista las cubiertas vegetales. También hay diferencias importantes en la terminología y en la forma de agrupar jerárquicamente las distintas unidades. Una revisión, siquiera somera, de esta materia excede los objetivos de esta comunicación; únicamente nombraremos, por su actualidad y relevancia

la iniciativa del Comité del Inventario del Patrimonio Natural y la Biodiversidad (adscrito al MAPAMA) de promover una lista patrón de hábitats terrestres basada en la nomenclatura que promueve la Agencia Europea de Medio Ambiente (AEMA): European Nature Information System (EUNIS). El European Topic Centre on Nature Conservation (predecesor del European Topic Centre on Biological Diversity) empezó a trabajar en esta clasificación en 1995 y culminó en una primera fase de la mano de la Agencia Europea de Medio Ambiente con la publicación de la versión revisada en 2004 (Davies et al, 2004). En los últimos años la Agencia Europea de Medio Ambiente le está dando un nuevo impulso con la revisión de algunas de sus clases como las de bosques y matorrales (Schaminée et al. 2012, 2013, 2014, 2015). La lista patrón española de hábitats terrestres cuenta con un primer borrador sobre el que previsiblemente se seguirá trabajando en los próximos años. Este borrador cuenta con 2.247 categorías acordes con EUNIS (en diferentes niveles jerárquicos) que, a su vez, hereda o se alimenta parcialmente de los resultados de distintas iniciativas abordadas en las últimas décadas en el ámbito Europeo como son la clasificación paleártica (Devillers et al., 1996), su predecesora Corine Biotopos (Devillers et al., 1991) y, naturalmente, la fitosociológica sigmatista. La AEMA recoge en algunas de sus publicaciones el devenir de todas estas iniciativas y las de varios estados miembros de la U.E. y su relación con clasificaciones previamente existentes (European Environment Agency, 2014, 2015; Evans, 2016).

Como se ha comentado anteriormente, no hay un criterio único para clasificar y dar nombre a las comunidades vegetales, los biotopos, los hábitats o las cubiertas en general ya que la validez de dicha sistemática depende de muchos factores. En el caso de nomenclaturas con fines cartográficos, además de los objetivos de la cartografía, son claves la resolución espacial y temática que se busca, el modelo de datos y las fuentes de información con que se elabora la cartografía. Estos últimos condicionan totalmente su resolución temática, es decir, la posibilidad de distinguir o no diferentes tipos de cubiertas.

En el proyecto que nos ocupa, las limitaciones referidas anteriormente nos han llevado a considerar un número contenido de categorías que no solo deben presentar estructura, fenología y respuesta espectral diferenciadas, sino también tener significación por su uso y, dentro de lo posible, por su valor natural.

Por la resolución espacial del proyecto, quedan fuera las coberturas que sólo se presentan en pequeñas superficies (menos de 20x20 m) y, por razones prácticas, las que ocupan superficies pequeñas en el conjunto de Castilla y León (la necesidad de disponer de material de entrenamiento de cada categoría es un condicionante importante). Con ello no se quiere restar valor e interés a dichas comunidades; simplemente se pone de manifiesto que puede haber otras alternativas para su cartografía o seguimiento.

El esfuerzo para acotar el número de categorías en una cartografía no va necesariamente en su menoscabo sino que, al contrario, puede otorgarle unos valores que muchas veces se infravaloran en cartografía de la vegetación. Es habitual que los sistemas de clasificación de comunidades vegetales, biotopos o hábitats tengan una cierta tendencia que podemos calificar de inflacionista en un intento por representar de la forma más fiel posible los distintos elementos del paisaje. Sin embargo, muchas veces, la utilidad de una cartografía mejora cuando incorpora un esfuerzo de síntesis para agrupar adecuadamente las distintas unidades. Los criterios pueden ser de significación ecológica o geobotánica (ver por ejemplo Sainz Olleros et al., 2010), de valor natural, de gestión administrativa o bien integrar diferentes perspectivas, como el caso que nos ocupa. Con ello no se pretende restar valor a la tendencia contraria, que también tiene su espacio, especialmente en un mundo en el que la óptica del *big data* está cambiando la forma de gestionar la información.

### Procesado de datos y material de entrenamiento

Esta cartografía se ha elaborado mediante técnicas de clasificación semiautomática de píxeles individuales (20x20 m) basadas en árboles de decisión utilizando algoritmos de *machine learning* y apoyadas en información de teledetección, Lidar e información auxiliar. Se trata de técnicas de clasificación supervisada: a partir del material de entrenamiento (píxeles clasificados por nosotros) se obtienen las reglas de clasificación que, posteriormente, se aplican a todos los píxeles de Castilla y

León obteniendo el correspondiente mapa ráster de cubiertas. Los detalles de esta metodología pueden consultarse en otra comunicación presentada en este mismo congreso, titulada “Clasificación semiautomática de cubiertas naturales arboladas en Castilla y León”.

Además de la adecuada elección de las variables clasificadoras y del procesado de datos, que no abordamos en esta comunicación porque se tratan en la referida anteriormente, son claves la elección de las categorías de cubiertas y el correspondiente material de entrenamiento. En los apartados siguientes nos ocupamos de estas dos cuestiones.

Se han llevado a cabo dos versiones de esta clasificación, una primera (2016-1) con 40 categorías de cubierta y sin reservar material de contraste y una segunda (2016-2) en la que se ha depurado la asignación de categorías de cubierta (se han reducido a 39) y en la que se ha reservado un 10-15% de parcelas para el contraste. En esta segunda versión se utilizó una capa Lidar adicional entre las variables clasificadoras (% de primeros retornos entre 0,25 cm y 3 m sobre el suelo aparente, que se suma a las de cobertura arbolada y altura dominante).

### Lista de categorías de cubierta

En la tabla 1 se enumeran las 39 categorías utilizadas en este trabajo (versión 2016-2). Por cuestiones de espacio, no se incluye la descripción de cada categoría, que puede consultarse en <http://bit.ly/ClasCob> (<http://www.medioambiente.jcyl.es/> → apartado “Flora y Fauna” → sección de “Clasificación de cubiertas”). En este fichero también recogemos la correspondencia con los Hábitats de Interés Comunitario de acuerdo a la interpretación utilizada en el Plan Director para la implantación y gestión de la Red Natura 2000 en Castilla y León.

Tabla 1. Categorías de cubierta con número de píxeles de entrenamiento y contraste y de polígonos totales

Tipo de cubierta	Píxeles entrenamiento	Píxeles contraste	Número polígonos
<b>Arbustedos y enebrales</b>			
Espinar caducifolio	1.512	137	48
Arbustedo con sabina mora	3.094	357	49
Enebral arbustivo	2.894	164	44
Matorral silicícola con enebros rastreros	1.706	254	35
Enebral y/o sabinar rastrero basófilo	2.269	268	36
<b>Matorrales dominados por ericáceas</b>			
Brezal alto denso	12.892	366	84
Brezal bajo claro	1.536	220	48
Brezal bajo denso	2.989	326	61
Brezal hidrófilo	2.130	267	55
Gayubar	1.921	77	35
<b>Matorrales dominados por labiadas y afines</b>			
Tomillar mixto	2.239	150	45
Tomillar-pradera	3.696	999	39
Cantuesar	1.236	99	38
Romeral	2.467	233	52
Pastizal leñoso mixto	4.663	711	49
Matorral gipsófilo	2.585	205	63
<b>Matorrales dominados por cistáceas</b>			
Jaral de estepa	1.921	284	59
Jaral de jara pringosa	2.205	162	63
Brezal-jaral (C. ladanifer)	1.921	229	36
<b>Matorrales dominados por leguminosas</b>			
Piornal alto de montaña	4.185	752	115

Tipo de cubierta	Píxeles entrenamiento	Píxeles contraste	Número polígonos
Piornal serrano	5.646	925	78
Escobonal blanco y afines	3.213	688	79
Formaciones abiertas con <i>Cytisus scoparius</i>	724	108	22
Olaguinar	3.410	427	75
Tojar	3.728	565	66
Aulagar	2.182	239	56
Cambronal	3.233	379	90
Erizonal	4.149	395	67
<b>Pastizales herbáceos</b>			
Pastizales de alta montaña crioturbados	2.950	356	59
Cervunales y otros pastizales no crioturbados de alta montaña	2.118	250	63
Cervunales de media montaña	1.586	147	48
Pastizales basófilos mesófilos	1.484	166	36
Pastizales basófilos xero-mesófilos	1.924	253	47
Vallicares y majadales	8.252	511	115
Prados	3.935	611	59
Herbazales rudero-arvenses	4.362	254	42
<b>Otros</b>			
Juncales	2.478	382	68
Roquedos y pedreras	14.630	1.433	168
Cárcavas	593	59	26

En un principio nos planteamos recoger en esta comunicación las correspondencias entre estas categorías y las comunidades vegetales básicas (CVB) definidas en la cartografía detallada de hábitats de los Lugares de Interés Comunitario de Castilla y León coordinada por la Junta de Castilla y León (trabajo en el que nos hemos apoyado), pero lo posponemos a una futura edición de este proyecto en la que nuestras categorías de cubierta estén más consolidadas. Lo mismo podemos decir respecto a las correspondencias con las categorías EUNIS del borrador de lista patrón española de hábitats terrestres.

Como se indicó anteriormente, se han ejecutado dos versiones de esta clasificación, la primera de ellas con una categoría más, que era “Suelo desnudo”. Esta categoría mostró mal comportamiento en la clasificación por presentar carácter invasivo muy fuerte: el sistema asignaba esta categoría a cubiertas que en realidad son brezales claros, o tomillares, o pastos xero-mesófilos, etc. La justificación estaba en la calidad del material de entrenamiento que no se había elaborado manualmente sino de forma sistemática, a partir del SIOSE. Tras comprobar la deficiente calidad del material de entrenamiento y las colisiones con otras tres clases de terreno desnudo (“Roquedo y pedreras”, “Cárcavas” y “Barbecho”) se optó por eliminar esta categoría en la segunda versión de esta clasificación. Los resultados cuantitativos que se ofrecen en esta comunicación corresponden a esta segunda versión.

### Material de entrenamiento

Como se indica más arriba, las formaciones desarboladas son particularmente difíciles de clasificar, entre otras cosas porque tienen tendencia a formar todo tipo de mezclas y (micro-)mosaicos. Esto dificulta la obtención de material de entrenamiento de calidad, razón por la cual hemos dedicado un particular esfuerzo a esta tarea.

La información cartográfica disponible acerca de cubiertas vegetales (comunidades vegetales, hábitats) es en realidad cartografía de sistemas complejos, porque lo habitual es que cada polígono integre varias cubiertas. Incluso, en los casos en los que figura nominalmente una sola, la realidad es

que suele haber otras acompañantes que no se recogen en la cartografía. En las primeras ediciones de este proyecto utilizamos provisionalmente información extraída de las cartografías disponibles, filtrando de forma sistemática las teselas con cubiertas más puras y aplicándoles el correspondiente buffer. Como la calidad de este material de entrenamiento no era suficiente, decidimos abordar un trabajo específico, más manual.

El material utilizado para esta edición del mapa de cubiertas (85-90% entrenamiento; 10-15% contraste) se ha obtenido mediante trabajo de campo y de gabinete. En campo se seleccionaron 441 parcelas de referencia de tamaño variable (la mayoría entre 0,79 y 5 ha) adecuadamente georreferenciadas y caracterizadas para este proyecto, repartidas en las distintas categorías y en sus respectivos ámbitos de distribución. Este trabajo nos permitió depurar la lista inicial de categorías para llegar a las 39 que se recogen en esta comunicación. A continuación se completó el material de entrenamiento en gabinete a partir de ortofotografía e información de inventarios realizados en otros proyectos (principalmente, el proyecto coordinado por la Junta de Castilla y León de Cartografía Detallada de Hábitats en los LIC, inédito). En gabinete se seleccionaron otras 1,877 parcelas de referencia. Esto hace un total de 2.318 parcelas con una superficie total de casi 6.000 hectáreas, que supone entre 22 y 168 parcelas por categoría (30-530 ha) con una media de 59 parcelas y 149 hectáreas por categoría. En la tabla 1 se recoge el número de píxeles de entrenamiento y contraste y el número total de parcelas para cada tipo de cubierta. El número total de píxeles 20x20m (entrenamiento+contraste) utilizados para este análisis es de 145.000. Este material es de buena calidad en cuanto a su georreferenciación y asignación de categorías, pero no lo es tanto en cuanto a su distribución e independencia de unos píxeles respecto a otros (la información que aportan dos parcelas de 1 ha es mayor que la que aporta una parcela de 2 ha). Por lo tanto, a la hora de valorar este material de entrenamiento no debemos perder de vista el hecho de que sólo disponemos de 22-168 parcelas de referencia por categoría. En cuanto al número de píxeles, consideramos que las categorías con menos de 1.000 píxeles de entrenamiento tienen pocas probabilidades de funcionar bien en la clasificación (tal es el caso de las formaciones abiertas con *Cytisus scoparius* y las cárcavas).

La técnica más sencilla y habitual para tener una valoración objetiva y cuantitativa de la calidad de este tipo de técnicas de clasificación consiste en contar con una sub-muestra de parcelas, independiente de la muestra de entrenamiento, para utilizarlas en un contraste *a posteriori*. A pesar de la escasa cuantía del material de referencia disponible, decidimos reservar parte de estas parcelas para su uso en el contraste porque el objetivo de este trabajo no es tanto obtener la mejor capa posible como identificar las fortalezas y debilidades de la metodología, lo que nos permitirá avanzar en la depuración de categorías y del procesado. Además de esta evaluación cuantitativa hemos llevado a cabo un análisis cualitativo del mapa obtenido.

#### 4. Resultados

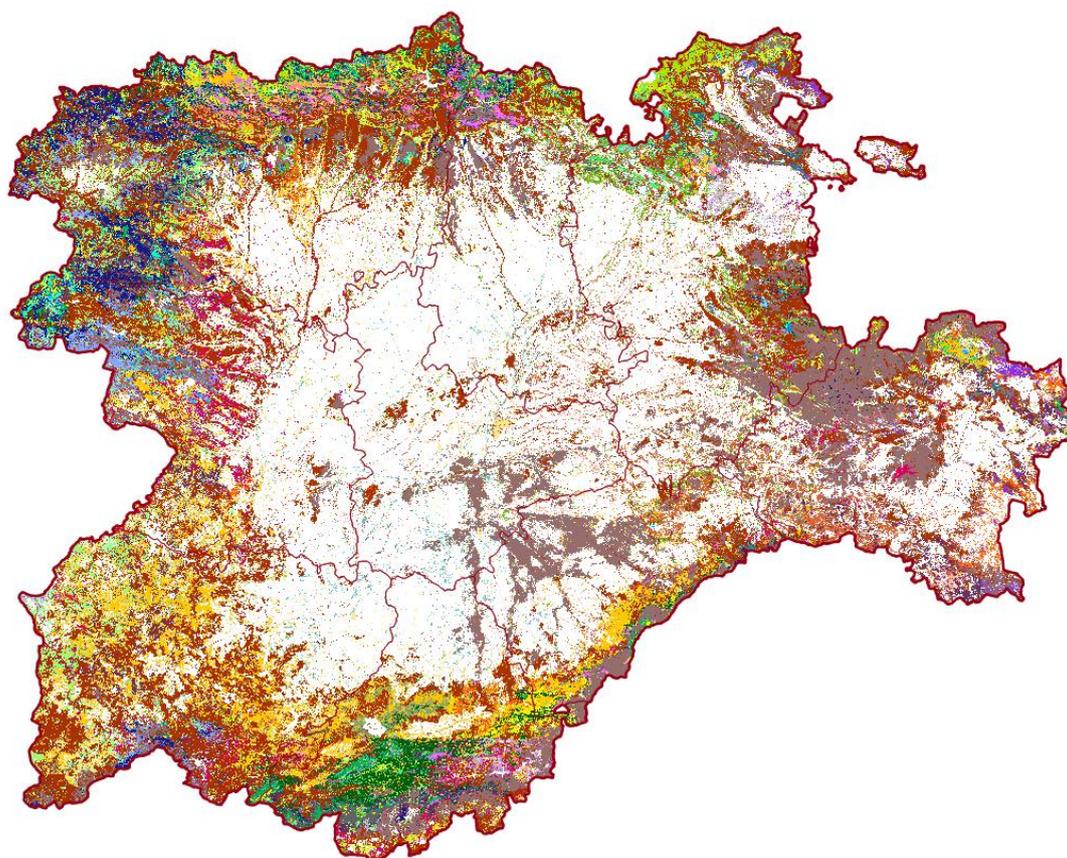
El resultado de la clasificación es un mapa ráster con píxeles de 20x20 m clasificados en alguna de las 145 categorías consideradas en el proyecto, de las cuales 39 corresponden a formaciones naturales o seminaturales no arboladas. A título ilustrativo se muestra en la figura 1 el resultado final después de agrupar las 142 clases en grandes grupos, salvo las clases de matorral y pastizal que se han mantenido desagregadas.

Se han analizado los resultados por separado para los cultivos agrícolas, con los que se ha obtenido porcentajes de acierto cercanos al 80%, para las formaciones arboladas con un 77% y para los matorrales y pastizales con el 62%. En la tabla 2 se recogen los porcentajes de acierto para cada categoría (versión 2016-2) en la cual se reservó un 10-15% de recintos para el contraste. Estos porcentajes son el complementario a 100 del error de omisión (píxeles de la categoría X que quedan clasificados en otras categorías).

Tabla 2. Categorías de cubierta y porcentaje de acierto obtenido con la muestra de contraste (versión 2016-2)

Tipo de cubierta	%	Tipo de cubierta	%
<b>Arbustados y enebrales</b>		<b>Matorrales dominados por leguminosas</b>	
Espinar caducifolio	26%	Piornal alto de montaña	39%
Arbustado con sabina mora	60%	Piornal serrano	93%
Enebral arbustivo	84%	Escobonal blanco y afines	81%
Matorral silicícola con enebros rastreros	91%	Formaciones abiertas con Cytisus scoparius	41%
Enebral y/o sabinar rastrero basófilo	97%	Olaguinar	65%
		Tojar	90%
<b>Matorrales dominados por ericáceas</b>		Aulagar	32%
Brezal alto denso	39%	Cambronal	61%
Brezal bajo claro	15%	Erizonal	82%
Brezal bajo denso	30%		
Brezal hidrófilo	38%		
Gayubar	71%	<b>Pastizales herbáceos</b>	
		Pastizales de alta montaña crioturbados	59%
<b>Matorrales dominados por labiadas y otros afines</b>		Cervunales y otros pastizales no crioturbados de alta montaña	52%
Tomillar mixto	83%	Cervunales de media montaña	46%
Tomillar-pradera	50%	Pastizales basófilos mesófilos	7%
Cantuesar	27%	Pastizales basófilos xero-mesófilos	34%
Romeral	97%	Vallicares y majadales	61%
Pastizal leñoso mixto	85%	Prados	74%
Matorral gipsófilo	53%	Herbazales rudero-arvenses	0%
<b>Matorrales dominados por cistáceas</b>		<b>Otros</b>	
Jaral de estepa	48%	Juncales	69%
Jaral de jara pringosa	83%	Roquedos y pedreras	82%
Brezal-jaral (C. ladanifer)	36%	Cárcavas	51%

Figura1. Mapa de Castilla y León clasificado



## 5. Discusión

Estos resultados muestran una capacidad predictiva bastante baja: tan solo 12 de las 39 categorías cuentan con porcentajes de acierto que podemos considerar aceptables para su aplicación práctica (superiores al 80%). Además hay bastantes tipos de cobertura con porcentajes de acierto particularmente bajos (<40%). Algunos de ellos son esperables por tener menos material de entrenamiento o ser particularmente heterogéneos (o ambos) como los espinares caducifolios. Sin embargo, hay otras categorías (como los cuatro tipos de brezal) que cuentan con más material de entrenamiento, son relativamente homogéneos internamente y, sin embargo, se clasifican mal.

Además, hay algunas cubiertas que aparecen sobredimensionadas porque presentan errores de comisión elevados (no reflejado en la tabla 2). El error de comisión correspondiente a una categoría X es el porcentaje de píxeles de otras categorías que quedan clasificados en dicha categoría X. A las categorías con error de comisión alto podemos considerarlas 'sumidero'. Algunas categorías con este problema son:

- Jaral de jara pringosa: hay un número significativo de píxeles de encinar disperso y abierto (FCC<70%), rebollar disperso (FCC<40%) y pinar de negral disperso (FCC<40%) que se clasifican dentro de esta categoría. Incluso algunos píxeles identificados como "formación arbolada dispersa con chopos" acaban clasificados en esta categoría. Creemos que además de corregir la pérdida de peso de las capas Lidar en el árbol de decisiones comentada anteriormente son necesarias otras mejoras que eviten este tipo de errores poco esperados.
- Brezal-Jaral (*Cistus ladanifer*): hay un número significativo de píxeles "Piornal alto de montaña" y "Brezal alto denso" que acaban clasificados en esta categoría. También el pasa a algunos píxeles de encinar disperso (FCC<40) y rebollar y pinar de negral dispersos y abiertos (FCC<70).

- Aulagar: hay un número significativo de píxeles de “Brezal bajo denso”, “Jaral de estepa” que acaban clasificados en esta categoría así como, en menor medida, algunas formaciones arboladas dispersas y abiertas (Qp, Pt, Qi).
- Romeral: hay un número significativo de píxeles de “Matorral gipsófilo” y “Aulagar” que acaban clasificados en esta categoría.
- Cantuesar: hay un número significativo de píxeles de “Jaral de estepa” y “Cambronal” que acaban clasificados en esta categoría, así como algunos píxeles que en realidad son formaciones dispersas y abiertas de encina.
- Vallicares y majadales: hay un número significativo de píxeles de “Escobonal blanco y afines” y “Formaciones abiertas con *Cytisus scoparius*” que acaban clasificados en esta categoría así como algunas formaciones dispersas y abiertas de encina y dispersas de rebollo.
- Juncuales: hay un número significativo de píxeles de formaciones abiertas con chopos que acaban clasificados en esta categoría.

Algunos de estos errores de comisión pueden deberse a que son categorías heterogéneas. Por ejemplo, la decisión de juntar majadales y vallicares (más berciales) en una sola categoría no está tan relacionada con su parecido en cuanto a las variables clasificadoras (respuesta espectral, fenología, etc.) sino por el hecho de que suelen presentarse formando micro-mosaicos y no es fácil separarlas cartográficamente.

Otra circunstancia que puede fomentar este efecto sumidero de algunas clases es el desequilibrio en la cuantía y distribución del material de entrenamiento: las clases que tienen una proporción muy elevada pueden tener un excesivo peso decisorio en las dicotomías del árbol de decisión.

Es importante comentar que hay diferencias muy grandes entre los porcentajes de acierto obtenidos con el material de contraste (tabla 2) y los correspondientes al de entrenamiento (no recogidos en la tabla 2, casi todos ellos cercanos al 100%). Creemos que este dato da la clave para interpretar estos resultados: el sobreentrenamiento es excesivo. Sobreentrenamiento significa que las reglas de clasificación se adaptan bien al material de entrenamiento con el que se construyen pero no funcionan igualmente bien con el resto de los píxeles. Esto suele ocurrir cuando la muestra es pequeña, cuando no es suficientemente representativa del espacio muestral y cuando el sistema tiene muchos grados de libertad (muchas variables suficientemente independientes entre sí). En estos casos, el sistema clasificatorio se adapta demasiado a la casuística particular del material de entrenamiento y, en consecuencia, no es generalizable al resto del espacio muestral.

Dentro de este mismo proyecto se ha trabajado también con formaciones arboladas y los resultados han sido muy diferentes: el porcentaje medio de acierto con éstas es del 77%, subiendo al 88% si consideramos solo las masas cerradas ( $FCC \geq 70\%$ ). La justificación de estas diferencias de resultados entre las cubiertas arboladas y las desarboladas pueden deberse a dos tipos de causas:

1. El material de entrenamiento:

- El tamaño muestral: para las formaciones arboladas el número medio de píxeles por clase de cubierta es de 9.661 mientras que para las clases desarboladas es de 3.350 (cerca de un tercio).
- La autocorrelación interna: depende principalmente del número de polígonos y su distribución. Respecto a lo primero, el número medio de polígonos de entrenamiento por clase de cubierta arbolada es de 327 mientras que para las desarboladas es de tan solo 59 (menos de la quinta parte).
- La distribución espacial: una muestra sistemática facilita la representatividad del espacio muestral mientras que una muestra dirigida puede tener déficits importantes de representatividad. La muestra arbolada es bastante sistemática mientras que la desarbolada ha seguido en parte criterios de accesibilidad para optimizar el trabajo de campo con el consiguiente sacrificio de representatividad.

2. Los criterios para la delimitación de las categorías y las características intrínsecas de cada tipo de cubierta. Como se indicó anteriormente, las formaciones naturales desarboladas son más heterogéneas y variadas que las arboladas en cuanto a que están dominadas por un

mayor número de especies y que la realidad de nuestros paisajes, en los que las frecuentes mezclas, introgresiones, micro-mosaicos, etc. hacen que muchas cubiertas sean difícilmente asignables a una sola categoría. Para categorizar este espacio muestral tan diverso en unas pocas clases es necesario reunir comunidades un tanto dispares en cada categoría y los criterios para agrupar comunidades deben revisarse para conjugar los aspectos fisionómico-estructurales y ecológicos con los relativos a su correlación con las variables clasificadoras (respuesta espectral, fenología, estructura, etc.). Esta complejidad inherente a las formaciones desarboladas no se presenta de igual forma para las arboladas (salvo para el arbolado disperso, que conjuga elementos arbolados con desarbolados).

A este análisis de los resultados cuantitativos se pueden añadir algunas observaciones cualitativas obtenidas analizando la versión 2016-1 (sin reserva de material de contraste) en 10 zonas de referencia de 5x5 km distribuidas por toda Castilla y León y en las que contamos de información de calidad sobre las cubiertas desarboladas.

En este análisis cualitativo se ha comprobado que el mencionado efecto sumidero se produce localmente también con otros tipos de cubierta como los enebrales los arbustedos de sabina mora, los erizonales y algunos tomillares mixtos. En concreto, el sistema asigna la categoría de *Juniperus phoenicea* a algunas repoblaciones fracasadas de pino carrasco y piñonero, donde apenas quedan ejemplares de estos pinos; así mismo asigna la categoría de enebral arbustivo a terrenos con arbolillos dispersos sobre un subpiso de aulagar (o incluso roquedo); y asigna las categorías de erizonal y de tomillar mixto en algunas parameras sorianas fuera del ámbito calcícola que es propio de estas comunidades.

Una característica común a los tres grupos de arbustedos es que son clases muy variables pero el número de parcelas de referencia usadas en el entrenamiento no es elevado en ninguna de ellas, por lo que el tamaño muestral no justifica su carácter invasivo. Este carácter se presenta sobre todo en zonas geográficas próximas a la ubicación de las zonas de referencia. No tenemos una explicación convincente para esto ya que las variables explicadoras utilizadas no incluyen la latitud ni la longitud geográficas; únicamente la pertenencia a diferentes escenas satelitales (no corregidas radiométricamente) podría contribuir a este sesgo. La corrección de estos problemas pasa por clases lo más homogéneas posibles, material de entrenamiento bien distribuido y, posiblemente, corrección radiométrica de las imágenes de satélite.

Tres categorías que el sistema confunde con cierta frecuencia son el tomillar mixto el pastizal xero-mesófilo y el matorral gipsófilo. Se trata de cubiertas que pueden presentar aspectos parecidos e incluso es habitual que presenten introgresiones y micro-mosaicos, por lo que era previsible este tipo de confusiones. Integrando dos o tres categorías en una super-categoría se mejorarían los porcentajes de acierto. Esta integración puede hacerse una vez entrenado el sistema si se cuenta con suficiente material de entrenamiento de cada subcategoría.

Un problema habitual en este tipo de sistemas es que las reglas de clasificación que funcionan bien en algunas zonas dejan de hacerlo si nos trasladamos a otros ámbitos geográficos. Con toda probabilidad esto ocurre en mayor o menor medida también en nuestro caso debido a que la distribución de la muestra de entrenamiento no es óptima. De forma más o menos clara lo hemos detectado con los erizonales, que quedan mejor clasificados en zonas donde hay parcelas de entrenamiento y peor donde no las hay. Una buena distribución del material de entrenamiento mejora este problema, salvo cuando el tipo de cubierta es de área geográfica restringida, como los erizonales, en cuyo caso la solución pasa por mejorar la capacidad predictiva de las otras clases con las que tiende a confundirse.

Otro elemento que puede ser un factor de distorsión en el sistema es la cubierta de líquenes sobre el suelo desnudo: se trata de algo habitual en suelo poco desarrollados, especialmente en zonas de paramera o clima continental en general, que cambia el color a muy oscuro y que es bastante variable, sobre todo como consecuencia del paso del ganado (por lo tanto no tiene una fenología definida).

## 6. Conclusiones

El proyecto de cartografía semiautomática de cubiertas nos ha demostrado que es posible llegar a resultados bastante satisfactorios para bastantes tipos de cubierta pero que no todas las categorías se dejan clasificar con igual facilidad. En concreto, la identificación de matorrales y pastizales es especialmente compleja y requiere un esfuerzo adicional que pasa por (1) elegir adecuadamente las categorías, (2) un buen material de entrenamiento y (3) un adecuado procesamiento de los datos. Creemos que hay recorrido de mejora en estos tres frentes.

## 7. Agradecimientos

Nuestro agradecimiento al Centro de Información Territorial de la Junta de Castilla y León que ha financiado la obtención del material de entrenamiento para este proyecto, dentro de la colaboración que mantiene este Centro con la Dirección General del Medio Natural y el Instituto Tecnológico Agrario de Castilla y León (ITACYL) para la elaboración de un Sistema de Información de Cubiertas.

## 8. Bibliografía

DAVIES, C. E.; MOSS, M. & HILL, M. O.; 2004. EUNIS habitat classification revised. European Environmental Agency – European Topic Centre on nature Protection and Biodiversity.

DEVILLERS, P.; DEVILLERS-TERSCHUREN J. & LEDANT J-P.; 1991. CORINE biotopes manual. Vol. 2. Habitats of the European Community. Office for Official Publications of the European Communities, Luxembourg.

DEVILLERS P. & DEVILLERS-TERSCHUREN J.; 1996. A classification of Palaeartic habitats. Council of Europe, Strasbourg: Nature and environment, No 78.

EUROPEAN ENVIRONMENT AGENCY; 2014. Terrestrial habitat mapping in Europe: an overview. Joint MNHN-EEA report. EEA Technical report No 1/2014. Lead authors: Jean Ichter (MNHN); Douglas Evans, Dominique Richard (ETC/BD).

EUROPEAN ENVIRONMENT AGENCY; 2015. Linking in situ vegetation data to the EUNIS habitat classification: results for forest habitats. EEA Technical report No 18/2015

EVANS, D.; 2016. The EUNIS habitats classification - past, present & future. *Revista de Investigación Marina*, 2012, 19(2)

LÓPEZ LEIVA, C; ESPINOSA RINCÓN, J. & BENGUA MTZ. DE MANDOJANA, J. 2009. Mapa de vegetación de Castilla y León. Síntesis 1:400.000. Junta de Castilla y León. Consejería de Medio Ambiente.

RIVAS MARTÍNEZ, S.; FERNÁNDEZ GONZÁLEZ, F.; LOIDI ARREGUI, J.; FERNANDES LOUSÃ, M & PENAS MERINO, A.; 2001. Sintaxonomical checklist of vascular plant communities of Spain and Portugal to association level. *Itinera Geobotanica*, 14: 5-341.

RIVAS MARTÍNEZ, S.; DÍAZ GONZÁLEZ, T. E.; FERNÁNDEZ GONZÁLEZ, F.; IZCO SEVILLANO, J.; LOIDI ARREGUI, J.; FERNANDES LOUSÃ, M & PENAS MERINO, A.; 2002. Vascular plant communities of Spain and Portugal: addenda to the syntaxonomical checklist of 2001. Part I. *Itinera geobotánica*, 15(1): 5-432 y part II. *Itinera geobotánica*, 15(2): 433-922.

SÁINZ OLLERO, H; RUT SÁNCHEZ DE DIOS, ANA ISABEL GARCÍA-CERVIGÓN MORALES.; 2010. La cartografía sintética de los paisajes vegetales españoles: una asignatura pendiente en geobotánica. *Ecología* Nº 23, 2010, págs. 251-275.

SCHAMINÉE, J. H. J.; CHYTRÝ, M.; HENNEKENS, S. M.; MUCINA, L.; RODWELL, J. S. & TICHÝ, L.; 2012. Development of vegetation syntaxa crosswalks to EUNIS habitat classification and related data sets. Alterra, Institute within the legal entity Stichting Dienst Landbouwkundig Onderzoek.

SCHAMINÉE, J. H. J.; CHYTRÝ, M.; HENNEKENS, S. M.; JIMÉNEZ-ALFARO, B; MUCINA, L.; RODWELL, J. S. & Data Contributors.; 2013. Review of EUNIS forest habitat classification. Alterra, Institute within the legal entity Stichting Dienst Landbouwkundig Onderzoek.

SCHAMINÉE, J. H. J.; CHYTRÝ, M.; HENNEKENS, S. M.; JANSSEN, J. A.M.; JIMÉNEZ-ALFARO, B; KNOLLOVÁ, I.; MUCINA, L.; RODWELL, J. S.; TICHÝ, L. & Data Contributors.; 2014. Vegetation analysis and distribution maps for EUNIS habitats (tasks 1, 2 & 3) Alterra, Institute within the legal entity Stichting Dienst Landbouwkundig Onderzoek.

SCHAMINÉE, J. H. J.; CHYTRÝ, M.; HENNEKENS, S. M.; JANSSEN, J. A.M.; JIMÉNEZ-ALFARO, B; KNOLLOVÁ, I.; MARCENO, C; MUCINA, L.; RODWELL, J. S.; TICHÝ, L. & Data Contributors.; 2015. Review of grassland habitats and development of distribution maps of heathland, scrub and tundra habitats of EUNIS habitats classification