



7º CONGRESO FORESTAL ESPAÑOL

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

26 - 30 junio 2017 | Plasencia
Cáceres, Extremadura

7CFE01-523

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

© Sociedad Española de Ciencias Forestales

Clasificación semiautomática de cubiertas naturales arboladas en Castilla y León

BENGOA MARTÍNEZ DE MANDOJANA, J. L.¹, DE BLANCO MEDINA, V.² y NAFRIA GARCÍA, D. A.²

¹ Consejería de Fomento y Medio Ambiente - Junta de Castilla y León

² ITACYL - Junta de Castilla y León

Resumen

La cartografía de la vegetación, los hábitats y la cobertura del suelo en general ha 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 apoya no sólo en “lo que se ve” en las imágenes aéreas, sino también en el conocimiento del territorio. ¿Puede una máquina hacer esta misma tarea?

Los intentos de automatizar la clasificación de las coberturas flaquean, tradicionalmente, cuando el ámbito territorial crece. Sin embargo, el desarrollo de las técnicas de análisis de datos y su aplicación a la teledetección, igualmente en pleno desarrollo, están ayudando a corregir este *handicap*.

En esta comunicación se presentan los resultados obtenidos en la clasificación de cubiertas naturales y seminaturales arboladas mediante técnicas de clasificación supervisada apoyada en teledetección, para el conjunto de Castilla y León. La posibilidad de actualización anual minimizando el apoyo de campo convierte a esta línea de trabajo casi en una gallina de los huevos de oro. Creemos que estas técnicas contribuirán sustancialmente a la modernización de la cartografía de las cubiertas del suelo.

Palabras clave

Cartografía-de-la-vegetación, clasificación-de-cubiertas, teledetección, árboles-de-decisión, land-cover

1. Introducción

El estudio y cartografía de la cubierta vegetal tiene una historia relativamente larga tanto en España como en el resto de Europa y Norteamérica y la literatura científica en esta materia es muy abundante. Una de las facetas de mayor proyección de este campo es el de la cartografía de la vegetación, que cuenta con numerosos trabajos destinados a cartografías locales y muy pocos que se ocupen de grandes extensiones con cierto detalle (p. ej. provincias o comunidades autónomas completas). Estas últimas suelen venir de la mano de la administración: inicialmente la Administración General del Estado y, en las últimas décadas, también la de las Comunidades Autónomas. Las universidades, además de colaborar con las administraciones general y autonómica en estas tareas, también han abordado de forma más autónoma o bien trabajos locales o bien síntesis cartográficas a menor escala, ocupándose en estos casos de la vegetación potencial o serial.

Las imágenes aéreas y de satélite, los sistemas de información geográfica y el posicionamiento GPS han modernizado sustancialmente los trabajos de cartografía de la vegetación facilitando enormemente muchas de las tareas asociadas a estas cartografías. Sin embargo, hay un cuello de botella que se sigue haciendo de forma manual, lo que conlleva elevados costes y subjetividad no siempre suficientemente acotada: la fotointerpretación. Esto incluye el trazado de líneas y asignación de contenidos. Se trata de tareas complejas, en las que el conocimiento del territorio ayuda mucho y que difícilmente se pueden automatizar. Pero difícil no significa imposible. Hay alternativas a la fotointerpretación manual y de hecho éste es uno de los caballos de batalla tradicionales de la teledetección.

La Agencia Europea de Medio Ambiente ha impulsado diversos trabajos destinados a identificar diferentes cubiertas mediante teledetección y, vistas las dificultades, ha preferido centrarse en grupos fácilmente identificables como el suelo artificial, las zonas húmedas o la separación entre

arbolado de frondosas y coníferas. En palabras de la AEMA, *the potential of remote sensing techniques has been more clearly demonstrated for mapping the physiognomy of vegetation (life form, cover, structure, leaf and type) than for the identification of individual species or communities defined by their floristic composition* (European Environment Agency, 2014). Uno de los focos de atención de esta Agencia son los Hábitats de Interés Comunitario (HIC) y, en este campo, se muestra bastante cauta y constata la afirmación de Vanden Borre et al. (2011) sobre las dificultades de aplicación de la teledetección en este campo.

Las dificultades para la identificación de especies y para el seguimiento de hábitats de interés comunitario (HIC) mediante teledetección no son las mismas ni para todas las especies (en realidad habría que hablar de tipos de cubierta) ni para todos los HIC. Esto último es particularmente significativo ya que los HIC son tremendamente heterogéneos entre sí: algunos hacen referencia a unos biotopos muy determinados y restringidos en sus características y extensión superficial, mientras que otros corresponden a coberturas muy diversas y variadas que pueden ocupar decenas o incluso cientos de miles de hectáreas.

Esta comunicación recoge algunos resultados obtenidos en los últimos años utilizando técnicas de clasificación (*machine learning*) para la cartografía de las cubiertas vegetales (comunidades vegetales, biotopos o hábitats). Esta línea de trabajo ha dado lugar a la publicación del mapa de cultivos y superficies naturales de Castilla y León con ediciones anuales desde 2011 hasta 2016 (<http://mcsncl.itacyl.es>). La versión publicada de estos mapas no incluye el desglose interno de las distintas formaciones naturales sino que las presenta agrupadas en unidades fisionómico-estructurales (arbolado de coníferas, de frondosas, mixto, matorral, pastizal). La decisión de no publicar las categorías individuales consideradas para las formaciones naturales y seminaturales se debe a que esta cuestión está en proceso de depuración tanto en lo que se refiere a las unidades consideradas como de la cuantía y calidad del material de entrenamiento.

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 iniciada en 2014. En esta comunicación se recogen principalmente cuestiones relacionadas con las superficies naturales y seminaturales arboladas; no obstante, esta iniciativa cartográfica se lleva a cabo contemplando todas las cubiertas, tanto naturales como agrícolas, artificiales y otras. Este planteamiento global es importante para una iniciativa de este tipo y es posible gracias a la colaboración entre distintos departamentos de la Junta de Castilla y León.

La iniciativa de poner en marcha este proyecto parte del ITACYL que toma como referencia los trabajos de los servicios estadísticos del USDA (United States Department of Agriculture) que renuevan anualmente el CropScape - Cropland Data Layer (<https://nassgeodata.gmu.edu/CropScape/>).

Dentro de las superficies naturales, hay diferencias importantes entre las formaciones arboladas y las desarboladas en cuanto a su complejidad y a la información que tenemos de ellas. Estas últimas presentan una variabilidad y heterogeneidad que hacen que su tratamiento sea particularmente complejo y, para ellas, contamos con menos información precisa de la realidad terreno. Por ello hemos destinado una comunicación específica a las formaciones no arboladas que se presenta a este Congreso bajo el título "Identificación de formaciones desarboladas a partir de teledetección en Castilla y León".

2. Objetivos

La administración medioambiental es responsable de la gestión técnica de una parte del territorio (Montes de Utilidad Pública y otros), ejerce el control administrativo de los usos y actuaciones en el conjunto de los montes, tiene a su cargo el seguimiento y conservación de los valores naturales, aprueba ayudas y otros incentivos para los propietarios o titulares de los terrenos, se hace cargo de algunas tareas específicas de gestión como la defensa contra los incendios forestales, establece criterios para instrumentos de gestión como la planificación selvícola, cinegética, etc. En todas estas actuaciones hay un ingrediente común, que es el territorio y su cubierta vegetal, que las condiciona. En consecuencia, esta administración necesita información

precisa y actualizada de la cubierta vegetal del terreno. Este trabajo está destinado a contribuir, en la medida de sus posibilidades, a ese objetivo.

Este objetivo condiciona la escala, ya que la información debe obtenerse a nivel de recinto SIGPAC o equivalente, con discriminación precisa y fiable de las distintas estructuras de la vegetación y orientativa de los principales grupos de comunidades vegetales, dando preferencia en primer lugar a la criterios relacionados con la utilización de esas cubiertas y en segundo lugar y en la medida de lo posible, con su valor natural.

En esta comunicación se recogen los planteamientos del trabajo, algunas cuestiones metodológicas, y una valoración de los resultados obtenidos en la clasificación de 2016 para las cubiertas naturales y seminaturales arboladas.

3. Metodología

Planteamiento

La realidad de nuestros paisajes naturales es variada y compleja y lo habitual es que los distintos tipos de cubierta formen mosaicos y transiciones y, a su vez, los elementos que componen una cubierta se combinen en cuantías y estructuras variadas. Por lo tanto, cada entidad territorial (p. ej. polígono) es un sistema complejo que la cartografía debe reflejar de forma sintética. El camino desde la realidad del terreno hasta su representación en un mapa o base de datos no es único ni directo y requiere un ejercicio de reflexión que debe responder a varias cuestiones:

1. Los objetivos y utilidad de la cartografía. Esta cuestión se ha respondido más arriba: este proyecto tiene como objetivo disponer de una herramienta de apoyo a la gestión administrativa del territorio.
2. Los medios y las fuentes de información disponibles. Las ortofotos, la teledetección, el Lidar, el trabajo de campo, los estudios y mapas previos y auxiliares y la información que tienen las administraciones derivada de sus tareas de gestión o control administrativo son el punto de partida para cualquier cartografía de este tipo. De estas fuentes, la teledetección y el Lidar se están revelando como dos herramientas clave para la modernización de esta cartografía. Pero ello requiere que la metodología para su procesado les saque el máximo partido y esta comunicación se centra principalmente en esta cuestión.
3. A la vista de las anteriores cuestiones, el cartógrafo debe decidir las resoluciones espacial y temática (muy interdependientes) que engloban dos cuestiones clave: (1) El modelo de datos y la(s) nomenclatura(s) asociadas; y (2) La escala y demás aspectos geométricos relativos tanto a la captura de la información como a su almacenamiento.

Como se ha indicado más arriba, la fotointerpretación es un proceso complejo que suele hacerse de forma manual. Hoy en día disponemos de ortofotografías de buena resolución y con renovación periódica (cada 2-4 años), suficientes para alimentar una cartografía de todo el territorio suficientemente detallada y actualizada (al menos para las escalas y tiempos que se manejan en la gestión del medio natural). Sin embargo nuestras cartografías de la cubierta vegetal no se renuevan ni actualizan adecuadamente con esa periodicidad. En esto influyen varios factores: por un lado, las administraciones somos poco eficientes y además no somos capaces de sacar el máximo partido a la información que tenemos (en parte derivada de nuestras propias tareas de gestión y control administrativo). Por otro lado, cuando hablamos de grandes extensiones y escalas detalladas, el carácter artesano de este tipo de cartografías de la cubierta vegetal supone un *hándicap* importante. Hablamos de carácter artesano porque requiere una laboriosa fotointerpretación (habitualmente manual) y debe apoyarse en un adecuado trabajo de campo.

Esta realidad es mejorable, tanto en lo que se refiere a la gestión que hacemos las administraciones de la información que obra en nuestro poder, como de las técnicas para identificar las cubiertas y los cambios que tienen lugar en ellas de forma más automatizada.

Nuestro planteamiento fue el de utilizar técnicas de clasificación de *machine learning* para identificar las cubiertas a partir de teledetección e información auxiliar. Eso requiere responder a varias cuestiones. En primer lugar, elegir adecuadamente las categorías que queremos distinguir: si elegimos un número reducido de categorías tendremos más probabilidades de acierto pero el

producto resultante será de menor utilidad y viceversa. Por otra parte, constatamos que las técnicas de *machine learning* en general y de clasificación en particular están en continuo desarrollo, por lo que en esta comunicación recogemos algunas ideas sobre esta materia. Por último, un ingrediente imprescindible para todo esto es la disponibilidad de imágenes de satélite, cuestión a la que dedicamos el correspondiente apartado.

Antecedentes: la clasificación automática de cubiertas

La materia que se aborda en este trabajo es el punto de encuentro de tres áreas de conocimiento: las ciencias biológicas, la ciencia de datos y la teledetección (donde decimos “ciencia de datos”, nos referimos a colectivos tan dispares entre sí como los estadísticos y los programadores). Cuando analizamos los trabajos realizados desde estas áreas de conocimiento, constatamos que la forma de aproximarse a problemas similares es muy dispar y que unos ponen el foco en cosas por las que otros pasan de puntillas (y viceversa). Las razones son evidentes: no es fácil conocer con suficiente profundidad y de forma simultánea los entresijos de varios campos de la ciencia. Este no es el lugar para repasar el devenir de cómo se ha abordado el problema que nos ocupa desde cada ámbito, pero sí para buscar el punto de encuentro de todas ellas y recoger algunas aportaciones relevantes de cada una para responder a los objetivos de este trabajo.

La identificación de cubiertas a partir de la teledetección es un campo en el que se lleva trabajando al menos desde finales de los años '70 y cuenta con una extensa literatura científica. En estas cuatro décadas, la creciente disponibilidad de satélites y los avances en el procesado de datos son el combustible para un gran número de publicaciones científicas que no vamos a recoger ni analizar en esta comunicación porque queremos trasladar el foco hacia las aplicaciones prácticas. Li et al (2014) recogen algunas de estas publicaciones y comentan los distintos enfoques: desde los métodos clásicos de clasificación individual de cada pixel, hasta las técnicas espacio-contextuales que sacan partido de la textura (*texture extraction*) o las formas (segmentación), más adecuadas para trabajar con imágenes de satélite de muy alta resolución. Para resoluciones intermedias o bajas también se recogen algunas técnicas que trabajan a nivel subpixelar mediante técnica de clasificación difusa.

En la literatura científica se suelen recomendar unas técnicas de clasificación para satélites de media resolución (como los que utilizamos en este trabajo) y otras para los de muy alta resolución, pero en realidad es la comparación entre el tamaño del pixel y el de los elementos que componen las cubiertas la que da las claves para que funcionen mejor o peor unas u otras metodologías de procesado de los datos.

La extensa literatura científica no se corresponde con una esperable abundancia de aplicaciones prácticas en grandes superficies: eso significa que el salto de la ciencia a la técnica aplicada no es sencillo y que hay cosas que funcionan a pequeña escala, pero quiebran cuando se quieren trasladar a grandes extensiones. A la vista de esta triste realidad, hemos puesto el foco en ejemplos concretos de clasificación automatizada de cubiertas en ámbitos grandes con resultados satisfactorios. Los ejemplos no son abundantes y nos parecen especialmente interesantes las experiencias del United States Department of Agriculture (USDA) y del United States Geological Survey (USGS): el Cropland Data Layer (CDL), el National Land Cover Database (NLCD), el National Gap Analysis Program Land Cover Data (GAP-LCD) y el Landscape Fire and Resource Management Planning Tools (LANDFIRE). Estos programas estadounidenses están interrelacionados entre sí y son una buena referencia en este campo. Por limitación de espacio no es posible comentar estas iniciativas; no obstante, hemos dejado en <http://bit.ly/ClasCob> (<http://www.medioambiente.jcyl.es/> → apartado “Flora y Fauna” → sección de “Clasificación de cubiertas”) un documento breve en el que se comentan las principales características de estos proyectos.

Tipos de cubiertas y material de entrenamiento

En este trabajo se aborda la aproximación más sencilla y directa de los enfoques referidos anteriormente: cada pixel de terreno se clasifica en función de su respuesta espectral (diferentes bandas y diferentes fechas) y de otras variables que consideramos relacionables con el tipo de

cubierta y que disponemos para todo el territorio, como variables Lidar, orográficas, climáticas, geológicas, etc. La variable cualitativa que se quiere inferir es el tipo de cobertura.

El tipo de clasificación es supervisada y entrenamos al sistema con una muestra de píxeles para los que conocemos el tipo de cobertura, además, claro está, de las variables clasificadoras. La calidad de los resultados depende de cuatro factores clave que hay que abordar adecuadamente: las variables clasificadoras, las categorías consideradas, el material de entrenamiento y las técnicas de procesado.

Para una cartografía como la que estamos abordando, los condicionantes metodológicos expuestos a lo largo de este trabajo (teledetección, clasificación semiautomática) y el sentido práctico (objetivos y utilidad de la cartografía) son los que acaban marcando la pauta para la elección de una tipología de cubiertas que definimos a escala de píxel (20x20 m): trabajamos con clases de cubierta “estancas” aun sabiendo que la realidad no responde a este esquema simplificado. Se trata de clases que, en muchos casos, son complejas porque incluyen diferentes tipos de cubiertas (en distintos píxeles) y, en algunos casos, también porque albergan varios ingredientes dentro de un mismo píxel, acercándose a la idea de los sistemas ecológicos del National Gap Analysis Program (GAP) referido anteriormente.

A la hora de decidir la resolución temática de esta cartografía (número de tipos de cubierta considerados) hemos tenido en cuenta varios factores: En primer lugar, que los distintos ingredientes de una cartografía (fuentes de información; resolución espacial; modelo de datos; lista de categorías; utilidad de la cartografía) deben ser coherentes, equilibrados. De nada sirve una nomenclatura muy desglosada si las fuentes de información utilizadas no permiten distinguir sus categorías y resulta igualmente inútil un esfuerzo tremendo de recogida de datos de campo si no se procesa, almacena y sintetiza o representa adecuadamente, es decir, si no tiene cabida en el modelo de datos seleccionado (incluidas nomenclaturas o clasificaciones). En nuestro caso, la capacidad discriminante de las variables clasificadoras (teledetección, Lidar, etc.) son las que van a decirnos hasta donde podemos llegar. En segundo lugar, teniendo en cuenta que trabajamos con un método de clasificación supervisada, necesitamos material de entrenamiento suficiente y de calidad para alimentar al sistema y este es otro limitante práctico para este trabajo.

A la vista de estos limitantes, nos propusimos trabajar con una clasificación de cubiertas con 70-100 categorías para las cubiertas naturales y seminaturales. En términos de biodiversidad este límite podría considerarse poco ambicioso, pero en términos de capacidad predictiva de un sistema semiautomático, discriminar adecuadamente 70 u 80 tipos de cubierta naturales sería un verdadero éxito.

En esta comunicación hacemos referencia a los resultados obtenidos en dos versiones de esta clasificación ejecutadas ambas con información satelital de 2016 pero considerando distintas categorías de cobertura y con algunos cambios en las variables clasificadoras y en el proceso de cálculo (versiones 2016-1 y 2016-2). Los tipos de cubiertas contemplados se pueden organizar en cuatro bloques:

- Cubiertas naturales y seminaturales. 81 categorías en la versión 2016-1 y 74 en la 2016-2.
- Cultivos agrícolas: 59 categorías en la versión 2016-1 y 69 en la 2016-2.
- Otras: Suelo artificial (una sola categoría) y agua (una categoría).

Las cubiertas naturales se han dividido en dos grupos: terrenos arbolados (con fracción de cubierta igual o superior al 10%) y terrenos desarbolados (el resto). Se han definido 35-41 categorías arboladas y 39-40 desarboladas. Estas últimas se abordan de forma específica en otra comunicación presentada a este Congreso, por lo que a continuación nos centraremos en las arboladas.

Para elegir las categorías más adecuadas de terrenos arbolados trabajamos con los siguientes planteamientos: (1) en esta primera fase nos centramos en formaciones arboladas que están bien representadas en Castilla y León (cada categoría de esta clasificación debe contar con varias decenas de miles de hectáreas); (2) la respuesta espectral de una masa arbolada cerrada depende fundamentalmente de la especie o especies dominantes; (3) la fracción de cubierta (FCC) condiciona fuertemente la respuesta espectral de un terreno arbolado (en las formaciones arboladas abiertas o dispersas, la respuesta espectral depende de la(s) especie(s) arbolada(s), la FCC y el subpiso y acompañantes de matorral y pasto herbáceo); (4) las especies dominantes y la fracción de

cabida cubierta son dos factores de primer orden de cara a la gestión administrativa del territorio y (5) por el momento, hemos optado por trabajar sólo con categorías para las que disponemos de fuentes de información para generar material de entrenamiento sin trabajo de campo adicional. Con estos planteamientos el margen de actuación no es muy grande: las categorías consideradas se recogen en la Tabla 1. En primera instancia (versión 2016-1) se trabajó no solo con cubiertas monoespecíficas sino también con algunas mezclas frecuentes, pero finalmente (versión 2016-2) se optó por obviarlas por las razones que se detallan más adelante.

Para la FCC adoptamos hasta tres intervalos (10-40-70-100%). La consideración de uno, dos o tres intervalos depende de si están o no suficientemente representados en Castilla y León (y podemos obtener material de entrenamiento) y se agrupan intervalos cuando resulta conveniente.

A los efectos de esta clasificación la mayor parte de nuestras masas arboladas son monoespecíficas: como trabajamos a nivel de píxel individual de 20x20 m, solo pueden considerarse mezclas las que lo son pie a pie; en cambio, las alternancias de golpes, rodales o bosquetes de distintas especies, pueden ser monoespecíficas dentro de cada golpe, rodal o bosquete (y dentro de los píxeles que los componen).

Las parcelas del Inventario Forestal Nacional (IFN) son una buena fuente de información tanto para obtener material de entrenamiento para esta clasificación como para saber la frecuencia con que se presentan las distintas mezclas pie a pie. En la versión 2016-1 de esta clasificación incluimos las mezclas pie a pie, considerando como tal las parcelas en las que, dentro del radio de 25 m, hay dos o más especies que aportan al menos el 20% del área basimétrica total de la parcela. Como comentamos más adelante, incluso en el supuesto de que tuviéramos la georreferenciación exacta de las parcelas IFN, no todos los píxeles que corresponden a una parcela IFN de mezcla tienen por qué ser mezcla y, de hecho, es posible que los píxeles que efectivamente incluyen ejemplares de dos o más especies sean minoría. Esta y otras cuestiones motivaron la ejecución de una segunda versión de la clasificación (2016-2) en la que excluimos las categorías de mezcla.

El material de entrenamiento se ha obtenido, para la mayor parte de las categorías arboladas, a partir de las parcelas del IFN. La distribución de estas parcelas en el conjunto de las masas arboladas de Castilla y León es suficientemente amplia para este tipo de procesado (aunque tiene distintas intensidades de muestreo en los distintos tipos de cubierta/zonas: una parcela por cada 100-400 hectáreas). Los principales inconvenientes de esta fuente de información son su antigüedad (datan de 2002 y 2003) y que no las tenemos georreferenciadas con suficiente precisión (se hicieron sin GPS y sólo contamos con su ubicación aproximada -coordenadas aparentes- obtenida a partir de la marca hecha sobre la ortofoto). Para sortear estos inconvenientes hemos utilizado sólo las parcelas IFN que cumplen el requisito de que las características de la masa forestal correspondiente a las coordenadas aparentes de acuerdo al MFE y a la FCC estimada a partir del Lidar son coherentes con las variables dasométricas de las parcelas. Para ello hemos procesado los vuelos Lidar PNOA de 2010 y 2014 (el procesado del Lidar se ha llevado a cabo de forma que los píxeles son geoméricamente coherentes con los de las imágenes de los satélites).

Próximamente se llevará a cabo el 4º IFN en Castilla y León, lo que permitirá corregir tanto el problema de la georreferenciación como el de la antigüedad. En todo caso algunas clases con poca superficie (pocas parcelas) seguirán requiriendo trabajo específico para obtener material de entrenamiento (además es previsible que ampliemos la lista con algunas especies más minoritarias).

Para incrementar el número de píxeles de entrenamiento no hemos trabajado con las parcelas reales de radio máximo de 25 m, sino con parcelas virtuales de 50 m de radio centradas en los centros de parcela IFN. Esto aumenta ligeramente el riesgo de errores en el material de entrenamiento, pero es coherente con el error de georreferenciación y, en todo caso, creemos que compensa al proporcionar más píxeles de entrenamiento por parcela IFN (la metodología adoptada para el procesado de datos admite un cierto porcentaje de error en el material de entrenamiento sin efectos negativos importantes sobre los resultados).

Estas parcelas virtuales seleccionadas son las utilizadas para generar el material de entrenamiento, con la particularidad de que la asignación de la correspondiente categoría no se ha hecho a nivel de parcela sino de píxel 20x20 m: a todos los píxeles correspondientes a una parcela se les asigna la(s) misma(s) especie(s) dominante(s) (las del inventario dasométrico en el conjunto de la parcela), pero pueden corresponder a dos o tres intervalos distintos de FCC de acuerdo al Lidar.

Por falta de material o caducidad del mismo hay varias formaciones arboladas que queríamos incluir en este trabajo pero para las que el IFN no nos resultó útil para generar material de entrenamiento: choperas de producción (Lx), plantaciones de *Pinus radiata* (Pr), plantaciones de eucalipto (Ec, Eg), castañares (Cs), robledales de *Quercus robur* y *Q. petraea* (Qr), abedulares (Be), alcornocales (Qs) y formaciones naturales arboladas de ribera (Rb). Para algunas de ellas (Lx, Pr, Ec, Eg, Cs, Qr) hemos generado material de entrenamiento de forma manual, siendo especialmente abundante el de choperas de producción (1.184 polígonos que suman 5.273 ha, repartidos en tres intervalos de cobertura, con algo más de la mitad de estos polígonos dentro del intervalo de $FCC \geq 70\%$).

Para cada tipo de cubierta, una parte de las parcelas o recintos se utilizaron como material de entrenamiento y otra para el contraste posterior. La selección del material de contraste no se hizo sobre píxeles individuales sino sobre parcelas completas para garantizar la independencia respecto al material de entrenamiento, ya que los distintos píxeles de una misma parcela no son independientes entre sí y su uso en el contraste daría lugar a una sobreestimación de los porcentajes de acierto.

El criterio inicial (versión 2016-1) fue el de elegir de forma aleatoria aproximadamente un 30% de parcelas IFN para las categorías que contaban con más de 500 parcelas; un 20% para las que tenían entre 100 y 500 y un 10% para las categorías con 50-100 parcelas (no se reservó material de contraste para las especies para las que el entrenamiento era manual). Para la versión 2016-2 redujimos la proporción de parcelas o recintos destinados a contraste a 20% (Ps, Pt, Qp, Qi, Lx), 15% (Pn, Pp, Jt, Qf, Fs, Cs) o 10% (Ph, Pr, Ec, Eg, Qr). Ver códigos de especie en la tabla 2.

El IFN en Castilla y León cuenta con cerca de 15.000 parcelas, de las cuales hemos seleccionado 11.103 // 9.575 (versión 2016-1 // versión 2016-2) para este trabajo. Adicionalmente hemos generado 1.290 // 1.596 recintos de forma manual (Lx, Pr, Ec, Eg, Cs, Qr). Para el contraste se han reservado 2.399 parcelas IFN // 2.179 parcelas IFN + recintos manuales (versión 2016-1 // versión 2016-2). A cada parcela IFN le corresponde un número variable de píxeles (normalmente 10-20).

Las especies e intervalos de cubierta arbolada considerados en la versión 2016-1 se recogen en la tabla 1. Las especies, intervalos de cubierta arbolada y número de píxeles disponibles para entrenamiento considerados en la versión 2016-2 se recogen en la tabla 2. Del total de píxeles disponibles para entrenamiento se utilizan efectivamente una parte de ellos, tal y como se indica más adelante.

Tabla 1. Categorías de cubierta arbolada e intervalos de FCC, versión 2016-1

<i>Pinus sylvestris</i> (10-40%, 40-70%, $\geq 70\%$)	<i>Juniperus thurifera</i> - <i>Quercus</i> ($\geq 40\%$)
<i>Pinus nigra</i> (40-70%, $\geq 70\%$)	<i>Quercus ilex</i> (10-40%, 40-70%, $\geq 70\%$)
<i>Pinus pinaster</i> (10-40%, 40-70%, $\geq 70\%$)	<i>Quercus faginea</i> (10-40%, 40-70%, $\geq 70\%$)
<i>Pinus pinea</i> (40-70%, $\geq 70\%$)	<i>Quercus pirenaica</i> (10-40%, 40-70%, $\geq 70\%$)
<i>Pinus halepensis</i> ($\geq 40\%$)	<i>Quercus robur</i> (40-70%, $\geq 70\%$)
<i>Pinus radiata</i> (40-70%, $\geq 70\%$)	<i>Castanea sativa</i> (40-70%, $\geq 70\%$)
<i>Pinus sylvestris</i> - <i>P. pinaster</i> ($\geq 40\%$)	<i>Quercus ilex</i> - <i>Quercus faginea</i> (40-70%, $\geq 70\%$)
<i>Pinus pinaster</i> - <i>P. pinea</i> ($\geq 40\%$)	<i>Q. pyrenaica</i> - otros <i>Quercus</i> ($\geq 40\%$)
Otras mezclas de <i>Pinus</i> ($\geq 40\%$)	<i>Fagus sylvatica</i> ($\geq 70\%$)
<i>Pinus sylvestris</i> - <i>Quercus</i> ($\geq 40\%$)	<i>Fagus sylvatica</i> - <i>Quercus</i> ($\geq 70\%$)
<i>P. pinaster</i> o <i>P. pinea</i> - <i>Quercus</i> ($\geq 40\%$)	<i>Populus</i> de producción (10-40%, 40-70%, $\geq 70\%$)
Otros <i>Pinus</i> - <i>Quercus</i> ($\geq 40\%$)	<i>Eucaliptus</i> sp. (40-70%, $\geq 70\%$)
<i>Juniperus thurifera</i> (10-40%, 40-70%)	

Tabla 2. Categorías de cubierta arbolada y número de píxeles preseleccionados para el entrenamiento y contraste, versión 2016-2

Especie(s) dominante(s)	Intervalos de FCC			N° píxeles - disponibles				N° píxeles de contraste			
				10-40%	40-70%	≥70%	Total	10-40%	40-70%	≥70%	Total
<i>Ps: Pinus sylvestris</i>	10-40%	40-70%	≥70%	1.658	2.502	21.190	25.350	419	531	5.390	6.340
<i>Pn: Pinus nigra</i>		40-70%	≥70%		577	4.534	5.111		117	752	869
<i>Pt: Pinus pinaster</i>	10-40%	40-70%	≥70%	4.326	8.010	22.580	34.916	1.095	2.182	5.215	8.492
<i>Pp: Pinus pinea</i>		40-70%	≥70%		2.177	4.088	6.265		383	696	1.079
<i>Ph: Pinus halepensis</i>		≥40%			819		819		114		114
<i>Pr: Pinus radiata</i>		40-70%	≥70%		840	8.989	9.829		43	1.376	1.419
<i>Jt: Juniperus thurifera</i>	10-40%	≥40%		3.161	2.863		6.024	513	534		1.047
<i>Qi: Quercus ilex</i>	10-40%	40-70%	≥70%	10.379	9.941	6.296	26.616	2.507	2.479	1.661	6.647
<i>Qf: Quercus faginea</i>	10-40%	40-70%	≥70%	2.109	1.581	2.320	6.010	468	282	264	1.014
<i>Qp: Quercus pirenaica</i>	10-40%	40-70%	≥70%	5.521	5.233	14.363	25.117	1.324	1.365	3.702	6.391
<i>Qr: Quercus robur</i>		40-70%	≥70%		7.424	52.492	59.916		331	4.841	5.172
<i>Cs: Castanea sativa</i>		40-70%	≥70%		6.349	38.799	45.148		952	5.319	6.271
<i>Fs: Fagus sylvatica</i>			≥70%			3.674	3.674			638	638
<i>Lx: Populus producción</i>	10-40%	40-70%	≥70%	10.023	11.682	60.282	81.987	1.569	3.426	16.926	21.921
<i>Eg: Eucaliptus globulus</i>			≥70%			208	208			0	0
<i>Ec. E. camaldulensis</i>		40-70%	≥70%		347	813	1.160		32	77	109

Imágenes de satélite e información complementaria

El lanzamiento del Satélite Sentinel-2A a mediados de 2015, inducía a pensar que para el año 2016 el proyecto iba a poder hacer uso casi exclusivo de las imágenes de este nuevo satélite con 10 metros de resolución espacial. Sin embargo diversos problemas relacionados con la entrada en operatividad del satélite durante la primera mitad del año 2016, obligaron a adquirir tres coberturas complementarias del satélite Deimos-1 (enero, marzo y mayo) y mantener la resolución espacial de la clasificación en 20 metros. Como complemento, al igual que en años anteriores se emplearon 54 composiciones multibanda (2-3-4 y 5-6-7) de conjuntos de escenas Landsat-8 fusionadas por órbitas. Las escenas Landsat-8, de 30 metros de resolución, son sometidas a un preprocesado intenso con técnicas de pansharpening para asimilarlas a imágenes Deimos-1 de 20 metros de resolución. A partir de la segunda mitad del año 2016 se incorporaron 32 composiciones de imágenes de Sentinel-2A, bandas 2, 3, 4 y 8.

Como información complementaria se utilizaron los siguientes datos:

- Capa ráster de 20 m de resolución de la cartografía de ocupación de suelo SigPAC del 2016 generada mediante rasterización a través de GDAL.
- Cobertura del arbolado, altura dominante de la vegetación y cobertura de matorral obtenidas a través del procesado LiDAR del vuelo PNOA de los años 2010 y 2014, enmascarada en las zonas de cultivo. La cobertura de matorral se ha usado sólo en la versión 2016-2.
- Altura del terreno procedente del Modelo Digital de Elevaciones (MDE) derivado de PNOA.
- Capas de pendientes y orientaciones derivadas del MDE antes mencionado.
- Capa interpolada de precipitación media anual (30 años) procedente del Atlas Agroclimático de Castilla y León.

Procesado de datos

El procesado de los datos se ha llevado a cabo con la aplicación C5.0 en su versión libre (con licencia GNU/GPL) sobre Linux sin llevar a cabo ninguna estratificación (se procesan los datos de toda Castilla y León de forma conjunta). Esta aplicación busca para cada dicotomía el valor de qué variable clasificadora maximiza la ganancia en información.

El número total de casos de entrenamiento (cubiertas naturales, agrícolas, artificiales y agua) supera los 22 millones. Muchos de estos casos son parcialmente redundantes (no independientes) al proceder del mismo objeto geográfico (parcela agrícola, parcela IFN, recinto desarbolado de referencia, etc.). Sobre este conjunto de casos se realiza una selección aleatoria de píxeles para el entrenamiento. La representación de cada tipo de cobertura en esta selección es proporcional a su presencia en la muestra original, aunque se establece un umbral mínimo de 2.000 píxeles para asegurar la presencia de casos de entrenamiento de las cubiertas infrarrepresentadas debido a su rareza o a la escasez de entrenamiento disponible. En la generación de los árboles se emplean técnicas de boosting con 10 iteraciones.

Con el objeto de evitar el sobreajuste o sobre-entrenamiento vinculado a los casos usados y producir árboles relativamente simples que funcionen adecuadamente cuando se empleen para la clasificación de casos que no han sido objeto de entrenamiento, se realiza un podado de los árboles. En el podado se eliminan aquellas ramas que no recogen al menos dos casos de entrenamiento en su punto de división. Además se realiza una estrategia de podado estadístico comenzando desde las hojas de los árboles, que elimina las ramas que no suponen un incremento del error en un intervalo de confianza del 25%.

Una vez confeccionados los árboles, estos son aplicados píxel a píxel sobre todo el territorio de Castilla y León, obteniendo una primera versión de la capa de clasificación de cubiertas.

4. Resultados

Como resultado del procesado de datos se obtienen las reglas de clasificación que, aplicándolas a todos los píxeles de Castilla y León dan lugar a una capa en la que cada píxel tiene asignado un tipo de cubierta. 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 arbolado natural y seminatural que se han mantenido desagregadas.

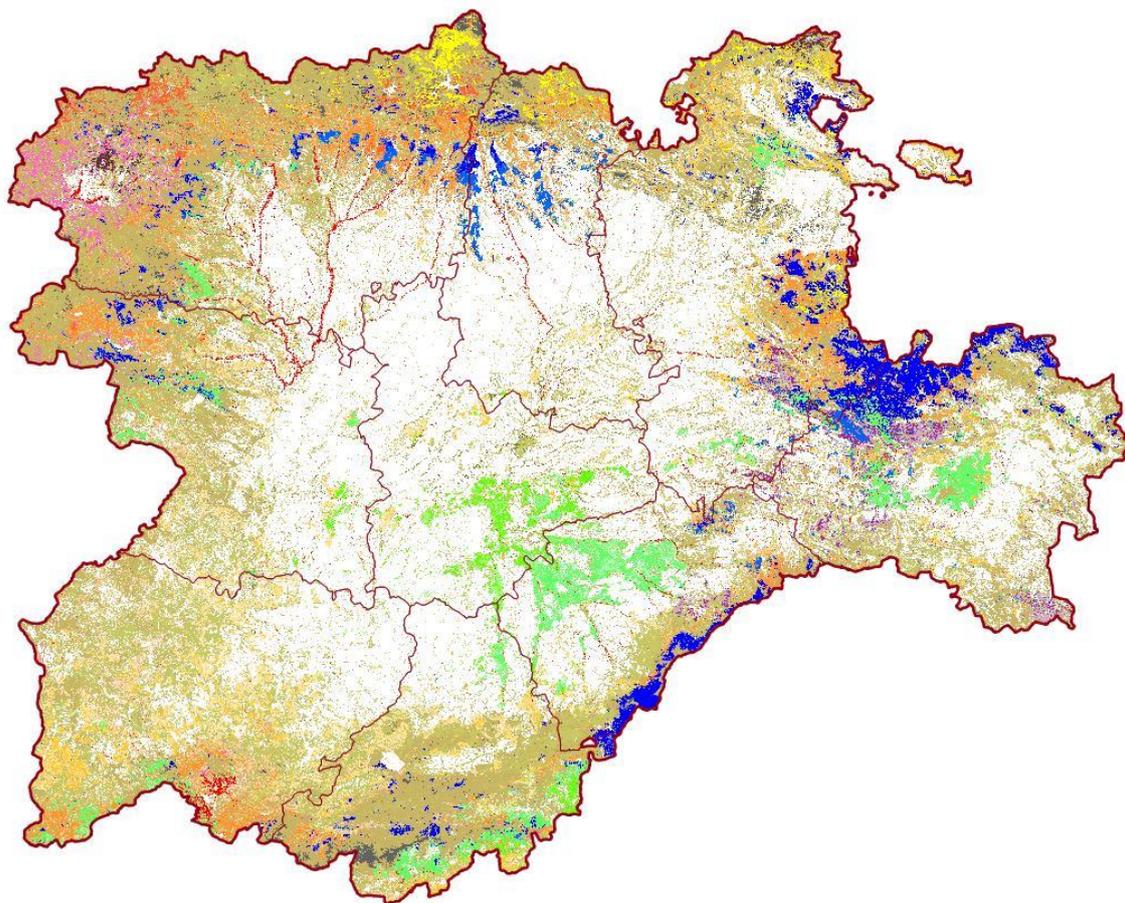


Figura1. Mapa de Castilla y León clasificado

En *machine learning*, tan importante es el producto que se genera como la validación que se haga del mismo con una muestra independiente. Teniendo en cuenta que en geografía “todo está relacionado con todo, pero lo más cercano lo está más”, se ha tenido especial cuidado para evitar en lo posible la correlación entre la muestra de contraste y la de entrenamiento.

En la tabla 3 se recoge, para cada tipo de cubierta, el porcentaje de los píxeles de contraste y de los disponibles para el entrenamiento que han sido clasificados adecuadamente.

Tabla 3. Categorías de cubierta arbolada y porcentaje de acierto (entrenamiento y contraste)- versión 2

Especie(s) dominante(s)	% acierto (contraste)			% acierto (entrenamiento)		
	10-40%	40-70%	≥70%	10-40%	40-70%	≥70%
<i>Pinus sylvestris</i>	<40%	<40%	84%	42%	40%	92%
<i>Pinus nigra</i>		<40%	67%		40%	96%
<i>Pinus pinaster</i>	<40%	45%	75%	47%	51%	85%
<i>Pinus pinea</i>		55%	87%		65%	93%
<i>Pinus halepensis</i>		88%			97%	
<i>Pinus radiata</i>		<40%	97%		51%	99%
<i>Juniperus thurifera</i>	74%	72%		86%	87%	
<i>Quercus ilex</i>	45%	52%	58%	61%	62%	71%
<i>Quercus faginea</i>	41%	18%	62%	69%	51%	91%
<i>Quercus pirenaica</i>	<40%	40%	77%	54%	44%	87%

Especie(s) dominante(s)	% acierto (contraste)		% acierto (entrenamiento)			
<i>Quercus robur</i>		<40%	94%		<40%	98%
<i>Castanea sativa</i>		<40%	92%		46%	98%
<i>Fagus sylvatica</i>			82%			100%
<i>Populus sp de producción</i>	76%	81%	97%	86%	82%	99%
<i>Eucaliptus globulus</i>						100%
<i>Eucaliptus camaldulensis</i>		75%	97%		81%	96%

Debe tenerse en cuenta que hay dos tipos de error para cada clase de cubierta X: por omisión (píxeles que perteneciendo a dicha categoría X no son clasificados como tal) y por comisión (píxeles de otras categorías que son clasificados erróneamente como X). La tabla 3 recoge el acierto complementario del error por omisión. No disponemos de una estimación completa de los errores por comisión ya que no se ha seleccionado una muestra de contraste con estos mismos criterios para todas las clases consideradas en la clasificación, pero si una aproximación considerando sólo las categorías de medio natural. En la tabla 4 se recoge el porcentaje de píxeles de las distintas clases de medio natural que han sido erróneamente clasificados en cada categoría (versión 2).

Tabla 4. Categorías de cubierta arbolada y porcentaje de error de comisión (entrenamiento y contraste)- versión 2

Especie(s) dominante(s)	% error de comisión (contraste)			% error de comisión (entrenamiento)		
	10-40%	40-70%	≥70%	10-40%	40-70%	≥70%
<i>Pinus sylvestris</i>	55%	51%	17%	36%	26%	12%
<i>Pinus nigra</i>		>60%	59%		19%	27%
<i>Pinus pinaster</i>	46%	38%	26%	27%	27%	18%
<i>Pinus pinea</i>		56%	46%		24%	31%
<i>Pinus halepensis</i>		32%			7%	
<i>Pinus radiata</i>		29%	6%		8%	7%
<i>Juniperus thurifera</i>	46%	38%		19%	19%	
<i>Quercus ilex</i>	32%	33%	37%	23%	29%	32%
<i>Quercus faginea</i>	55%	>60%	74%	28%	28%	34%
<i>Quercus pirenaica</i>	51%	41%	34%	35%	30%	23%
<i>Quercus robur</i>		45%	9%		12%	8%
<i>Castanea sativa</i>		12%	12%		12%	8%
<i>Fagus sylvatica</i>			38%			13%
<i>Populus sp de producción</i>	10%	9%	4%	5%	9%	3%
<i>Eucaliptus globulus</i>						0%
<i>Eucaliptus camaldulensis</i>		25%	49%		14%	25%

5. Discusión

Los porcentajes de acierto con la muestra de contraste son los que reflejan la calidad predictiva del sistema. Los porcentajes que se obtienen con la muestra preseleccionada para el entrenamiento son superiores a los de contraste principalmente por sobreentrenamiento: se obtienen reglas de clasificación que funcionan mejor con los píxeles utilizados para entrenar al sistema (y sus vecinos) que con el resto de píxeles. Eso puede estar relacionado con el hecho de tener poco material de entrenamiento o con una ramificación excesiva del árbol de decisión. En todo caso, las mejoras metodológicas han permitido reducir la diferencia entre estos porcentajes para la versión 2016-2

respecto a la versión 2016-1. La corrección del sobreentrenamiento pasa por encontrar la poda óptima del árbol y, acaso, por seleccionar mejor las variables clasificadoras.

Una de las conclusiones más consistentes que obtuvimos con la versión 2016-1 fue que las cubiertas de mezcla se clasifican, en todos los casos, mal, mientras que las cubiertas monoespecíficas se clasifican mejor. Esto se debe a que las parcelas IFN que habíamos identificado como mezclas no lo son en todos sus píxeles; y no sólo por los problemas de georreferenciación anteriormente referidos sino también porque lo habitual de una parcela IFN de mezcla es que incluya algunos píxeles de una especie, otros de la otra y, por fin, también algunos de mezcla. Eso significa que estábamos usando material de entrenamiento deficiente, con píxeles que asignamos a una categoría y pertenecen realmente a otra. Además, ocurre lo mismo con las parcelas de contraste, con lo que el hecho de que el sistema no asigne categorías de mezcla a todos los píxeles de estas parcelas no era, necesariamente un error. Esta fue una de las razones para llevar a cabo, este mismo año, una segunda versión de la clasificación, en la cual no consideramos las categorías de mezcla y aprovechamos para incorporar material de entrenamiento para dos de las especies que teníamos pendientes (roble y castaño).

En teoría, si las parcelas estuvieran georreferenciadas submétricamente, como sabemos las coordenadas de cada árbol respecto al centro de la parcela, podríamos saber que píxeles son mezclas verdaderas y cuáles no. Sin embargo, debido a que se trata de parcelas de radio variable, esto solo lo podemos hacer para el píxel que se solapa con la parcela de radio 10 m en la que se inventarían todos los pies.

Otra observación consistente es existencia de diferencias de acierto entre masas cerradas ($FCC \geq 70\%$), abiertas ($FCC: 40-70\%$) y dispersas ($FCC < 40\%$). Los resultados obtenidos en la versión 2016-2, en la que la calidad predictiva baja con la apertura del arbolado, son coherentes con lo esperado porque cuanto menor es la cobertura del arbolado, mayor es la proporción de elementos acompañantes en forma de subpiso y micro-mosaico con los claros entre el arbolado (que introducen distorsión o ruido). Por el momento no se ha considerado oportuno definir subcategorías dentro de clase arbolada en función de sus acompañantes, porque eso eleva sustancialmente el número de clases y queríamos conocer previamente la importancia cuantitativa de este efecto. Estos resultados apuntan hacia la necesidad de reconsiderar esta cuestión, no sin antes analizar con más detenimiento estos resultados y su comparación con los obtenidos en la versión 2016-1 de esta clasificación en la que las diferencias en el porcentaje de acierto entre intervalos de cobertura arbolada eran menores. Si bien esto está relacionado con la pérdida de peso decisivo de las variables Lidar en la versión 2016-2 frente a la 2016-1, creemos que éste no es el único causante del diferente comportamiento.

Los mejores porcentajes de acierto se obtienen para las masas cerradas de pinos, hayas, robles, castaño, chopos y eucalipto. Los peores resultados dentro de las masas cerradas se obtienen con encina y quejigo y, en general, con las masas abiertas y dispersas de todas las especies. La primera explicación que valoramos para las diferencias entre especies en lo que se refiere a estos porcentajes de acierto, fue que el material de entrenamiento de encina y quejigo era más deficiente para estas dos especies porque los defectos de georreferenciación les afectan más al ser más frecuentes sus mezclas y por quedar éstas peor recogidas en los anillos de más de 5 m o 10 m de radio de las parcelas de radio variable (en los anillos exteriores a estos círculos no se inventarían los pies de diámetro inferior a 12,5 y 22,5 cm respectivamente, muy frecuentes para encina y quejigo). Sin embargo los porcentajes de acierto obtenidos tras agrupar las clases por especies (integrar las distintas FCC) y la revisión de la matriz de confusión completa mostraron que una buena parte de clasificaciones erróneas para encina y quejigo se daban entre clases de cubierta cubierta (defecto que queda anulado al agruparlas). Esto ocurría en menor medida en la versión 2016-1 que en la 2016-2 y lleva a la conclusión de que los cambios en el procesado han llevado a que la capa Lidar de coberturas tengo menos peso en el árbol de decisión (frente a otras variables clasificadoras) en la versión 2 que en la 1. Nota: en esta comunicación no se incluye la matriz de confusión por falta de espacio ya que se trata de una matriz de 35 filas y 35 columnas si consideramos sólo las categorías arboladas y de 145x145 si las consideramos todas.

El porcentaje general de acierto es del 77% (media ponderada con el número de píxeles reservados para el contraste en cada clase). Si consideramos los resultados por intervalo de

cobertura arbolada, para masas cerradas asciende al 88%, para las abiertas desciende al 54% y para las dispersas al 49%. El porcentaje de acierto medio ponderado de la versión 2016-2 es algo mejor que el de la versión 2016-1 (no obstante, los correspondientes a algunas clases abiertas y dispersas de la versión 1 son mejores que los de la 2).

Si agrupamos intervalos de FCC para cada especie, los porcentajes de acierto se incrementan sensiblemente (porcentajes de acierto referidos a la muestra de contraste): Ps: 83%; Pn: 64%; Pt: 77%; Pp: 90%; Pr: 97%; Jt: 85%; Qi: 70%; Qf: 54%; Qp: 76%; Qr: 94%; Cs: 91%; Fs: 82%; Lx: 97%. La media ponderada con el número de píxeles de contraste de cada clase es del 86%. Omitimos los resultados correspondientes a pino carrasco y eucalipto porque la cuantía de la muestra de contraste se considera insuficiente para proporcionar resultados fiables (cercano a los 100 píxeles en estas dos categorías) y porque además el eucalipto tiene efecto sumidero bastante acentuado ya que esta clase se asigna a bastantes píxeles (proporcionalmente) que son de otras clases (error de comisión), con lo que el error de omisión puede malinterpretarse. Con esta agrupación de clases, el quejigo es la especie que peor se clasifica y creemos que los errores de georreferenciación de la muestra de entrenamiento afectan más a esta especie que a otras porque el quejigo se mezcla con bastante frecuencia con la encina y el rebollo (en Castilla y León) con lo que el material de entrenamiento puede no ser tan bueno como quisiéramos. La clase de quejigo con $FCC \geq 70$ también tiene un importante efecto sumidero.

Principales errores de omisión de la matriz de confusión:

- Pinos: Pn(C)→Ps(C): 14%; Pn(A)→Ps(C): 9%; Pn(C)→Pt(C): 10%; Pn(A)→Jt(A): 17%; Ps(C)→Pn(C): 6%; Ps(D)→Qp(D): 6%; Pt(C)→Ps(C): 5%; Ph(CA)→Pp(C): 6%;
- Quercus: Qf(A)→Qp(A): 17%; Qf(D)→Qi(D): 10%; Qf(C)→Qr(C): 7%; Qf(C)→Qi(C): 7%; Qf(A)→Qi(A): 10%; Qf(D)→Qi(D): 12%;
- Otros: Jt(CA)→Pt(C): 6%; Cs(C)→Qp(C): 7%; Fs(C)→Qp(C): 10%

Nota: se indica en primer lugar la categoría real del píxel y en segundo lugar la categoría en la que ha quedado clasificada; entre paréntesis la clase de cubierta: C: cerrado; A: abierto; D: disperso.

Para algunas formaciones arboladas dispersas, un porcentaje significativo de píxeles ha quedado clasificados en categorías no arboladas (Qp: 22%; Qf: 19%; Qi: 16%; Ps: 15%; Pt: 14%).

Al margen de los resultados que se exponen en esta comunicación, se han ensayado algunas alternativas como cambiar del análisis independiente de cada píxel que caracteriza a esta metodología a un enfoque más integrador espacialmente en el que la clasificación de cada píxel se apoya también en las características de los píxeles adyacentes. Se ha comprobado que este enfoque puede mejorar los resultados, especialmente para las cubiertas naturales y seminaturales. Sin embargo esto no se ha aplicado de forma extensiva al conjunto de la Comunidad autónoma y se apunta aquí únicamente como futura línea de trabajo.

6. Conclusiones

La principal conclusión que extraemos de los trabajos realizados es que es posible clasificar las cubiertas naturales a partir de teledetección (y otras fuentes) con un buen porcentaje de acierto (superior al 80% para especies individuales).

Llegar a unos resultados satisfactorios requiere trabajar varios frentes, como son elegir adecuadamente las clases, obtener un buen el material de entrenamiento, seleccionar bien las variables clasificadoras (imágenes de satélite, bandas, fechas, variables complementarias) y afinar el procesado de los datos.

7. Bibliografía

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).

- LI, M.; Zang, S.; Zhang, B.; Li, S. & Wu, C.; 2014. A Review of Remote Sensing Image Classification Techniques: The Role of Spatio-contextual Information. *European Journal of Remote Sensing*, 47: 389-411.
- VANDEN BORRE, J.; PAELINCKX, D.; MÜCHER, C.A.; KOOISTRA,L.; HAEST, B.; DE BLUST, G.; SCHMIDT, A.; 2011. Integrating remote sensing in Natura2000 habitat monitoring: prospects on the way forward. *J. Nat. Conserv.* 19(2), 116–125.