

Generación de cartografía detallada de vegetación mediante procesamiento digital de imágenes Landsat, variables orográficas y climáticas

Gerard Moré¹, Xavier Pons^{2,1}, José Ángel Burriel¹, Rafael Castells¹, Joan Josep Ibàñez¹, Xavier Roijals¹

(1) Centro de Investigación Ecológica y Aplicaciones Forestales (CREAF)
Edificio C, Universidad Autónoma de Barcelona
08193 Bellaterra

(2) Departament de Geografia, Fac. Lletres UAB
Edificio B, Universidad Autónoma de Barcelona
08193 Bellaterra

Palabras clave: Áreas de entrenamiento, Clasificación Mixta, Teledetección, Cartografía de Vegetación, Cataluña.

Resumen

En este trabajo se presenta un método de clasificación mixta de imágenes de satélite y otras variables climáticas y orográficas para la cartografía de cubiertas vegetales. Las áreas de entrenamiento se obtienen de otras bases ya existentes sobre vegetación, como es el Mapa de Hábitats de Cataluña, y son sometidas a un tratamiento estadístico para garantizar su calidad, y evitar problemas de cambios de usos de cubiertas debidos a la dinámica temporal. Los resultados expuestos aquí, aunque no definitivos, apuntan a que la metodología en su conjunto permite la creación de un mapa de vegetación con resolución espacial de 30 metros y con una leyenda que 15-20 categorías forestales (34 contando categorías de mezclas temáticas) por zona con unos porcentajes de acierto, valorados con un test independiente, de más del 85%.

1. Introducción

El Mapa de Cubiertas del Suelo de Cataluña (MCSC) es un producto digital realizado en el CREAM que describe las principales cubiertas y usos del suelo de Cataluña. Las áreas de la primera versión del MCSC se obtuvieron a partir de la fotointerpretación y digitalización sobre pantalla de ortofotografías en color 1:25 000 del año 1993 [1, 2, 3].

Actualmente se está llevando a cabo la segunda versión del MCSC, cuyo principal soporte son las ortofotografías 1:5 000 en color del período 2000/03. Sin embargo, la información obtenible con fotointerpretación de ortofotografías no permite la diferenciación de la mayoría de bosques. En este sentido, las imágenes de satélite, con una mayor resolución espectral y temporal que las ortofotografías, pueden ayudar a la discriminación de ciertas formaciones vegetales, al tiempo que el tratamiento digital de dichas imágenes permite mayores posibilidades de obtención de cartografía. Si bien es cierto que dichas imágenes tienen en muchos casos una resolución espacial menor, las ventajas comentadas permiten igualmente plantearse su utilización.

El objetivo de este trabajo es la cartografía de vegetación en el ámbito catalán utilizando clasificación digital de una serie continua de imágenes Landsat 7 ETM+ del período enero 2002 – mayo 2003, y Landsat 5 TM a partir de julio de 2003 (todas ellas con periodicidad de 16 días), así como otras variables de carácter climático y orográfico.

Existen otras experiencias donde se han utilizado imágenes de teledetección espacial para cartografiar con algún nivel de diferenciación las zonas forestales con una leyenda y extensión territorial similar a las que aquí nos ocupan, ya sea des del ámbito de la fotointerpretación (*Corine Land Cover Project* [4]), como des del punto de vista de las clasificaciones digitales [5, 6, 7, 8, 9, 10, 11].

A diferencia de la mayoría de metodologías citadas anteriormente, en este trabajo se pretende aprovechar todas las posibilidades que ofrece la resolución temporal de las imágenes de satélite con el fin de aumentar el poder de discriminación de cubiertas vegetales con comportamientos fenológicos diferenciados. En el proceso de clasificación se han incluido también variables climáticas y orográficas que contribuyen a la diferenciación de los diferentes tipos de vegetación. Otra característica del presente trabajo es la utilización de bases cartográficas digitales sobre vegetación realizadas recientemente (Mapa de Hábitats de Cataluña) como fuente para la obtención de las áreas de entrenamiento. Este proceso, que requiere un tratamiento estadístico previo, permitirá actualizar periódicamente y mediante tratamiento digital las bases utilizadas, y así mismo significa un considerable ahorro en tiempo y esfuerzo al disminuir los requerimientos en trabajo de campo.

2. Material y métodos

Las imágenes utilizadas han sido corregidas geoméricamente por el Departamento de Geografía de la UAB, utilizando un MDE siguiendo la metodología de Palà y Pons [12], implementada en el módulo CorrGeom de MiraMon [13]. Se ha colaborado con dicho departamento para la corrección radiométrica de las imágenes, utilizando el módulo CorRad de MiraMon y siguiendo la metodología descrita en Pons y Solé-Sugrañes [14], y contrastado en Salvador et al. [15].

División del ámbito catalán en zonas de estudio mediante uso de variables climáticas

Como se ha comentado, el ámbito general de estudio de este trabajo es Cataluña. Sin embargo, la superficie que esto supone es demasiado extensa para trabajar al nivel de detalle espacial y temático deseado. Una división del ámbito catalán en zonas más reducidas hace que los procesos informáticos sean más rápidos, y más cómoda la revisión visual de las imágenes y los resultados, al tiempo que evita la posibilidad de ciertas confusiones.

La distribución geográfica de la vegetación está influenciada por condiciones climáticas. Es por ello que se ha optado por realizar la división de todo el ámbito catalán utilizando variables climáticas, de modo que las nuevas áreas obtenidas sigan en cierta medida un patrón de distribución de vegetación.

Con este fin, se ha introducido en un proceso de clasificación no supervisada de tipo IsoData [16] con núcleo clasificador basado en mínima distancia euclidiana las cuatro primeras componentes principales estandarizadas de las siguientes variables, obtenidas todas ellas del Atlas Climático Digital de Cataluña [17, 18, 19, 20]:

- Temperatura media de las mínimas en Enero.
- Temperatura media de las máximas en Julio.
- Valor máximo anual en cada píxel de las Precipitaciones máximas mensuales.
- Valor mínimo anual en cada píxel de las Precipitaciones mínimas mensuales.
- Valor máximo anual en cada píxel de la Radiación máxima mensual.
- Valor mínimo anual en cada píxel de la Radiación mínima mensual.

Como resultado se obtienen nueve *clusters* que a partir de ahora se denominarán zonas climáticas y que quedan resumidas en la siguiente figura y tabla:

Código Leyenda	Descripción Zona Climática	Superficie Total (km²)
1	Temperatura Muy Alta en Verano y Media en Invierno, Precipitación Muy Baja y Radiación Alta	4 080.10
2	Temperatura Muy Alta en Verano y Media en Invierno, Precipitación Baja y Radiación Media	4 117.07
3	Temperatura Media, Precipitación Media, Radiación Alta.	4 870.63
4	Temperatura Alta, Precipitación Media en Estación Húmeda y Muy Baja en Estación Seca, Radiación Alta	5 906.49
5	Temperatura Moderadamente Baja, Precipitación Media, Radiación Baja	2 266.70
6	Temperatura Alta, Precipitación Muy Alta en Estación Húmeda y Media en Estación Seca, Radiación Media	4 502.76
7	Temperatura Media, Precipitación Muy Alta en Estación Húmeda y Alta en Estación Seca, Radiación Alta	2 836.20
8	Temperatura Muy Baja, Precipitación Muy Alta, Radiación Muy Alta	2 024.94
9	Temperatura Muy Baja, Precipitación Muy Alta, Radiación Muy Baja	1 483.27

Tabla 1: Superficie ocupada por las zonas climáticas delimitadas con IsoData.

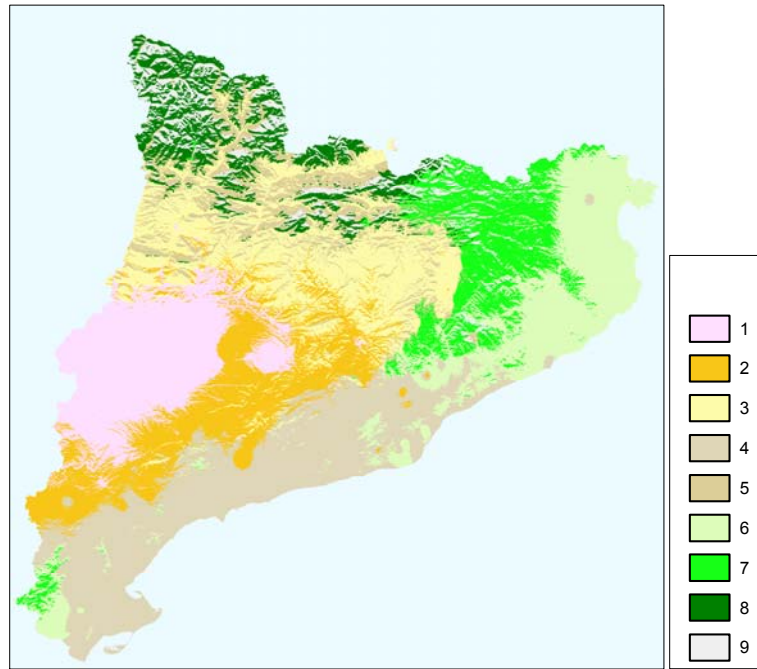


Figura 1: Zonas climáticas de Cataluña obtenidas con clasificación no supervisada IsoData (véase tabla 1 para significado de cada categoría).

Cabe matizar un par de puntos sobre estas zonas climáticas. En primer lugar, se debe tener en cuenta que las variables utilizadas en la delimitación de las zonas climáticas serán introducidas también en los posteriores procesos de clasificación de vegetación. Como se podrá comprobar más adelante, la capacidad que tienen estas variables para discriminar tipos de vegetación va más allá de las nueve zonas delimitadas en este primer proceso IsoData. Desde un punto de vista de jerarquías en los resultados de una clasificación no supervisada, se puede considerar esta delimitación de zonas climáticas como un resultado general, donde de algún modo los *clusters* obtenidos incorporan en su interior otras categorías más concretas (que intentaremos diferenciar en los procesos posteriores). En segundo lugar, aquí se presentan las zonas como unidades. Sin embargo, algunas de las áreas se han visto separadas en dos zonas diferentes, para poder adaptarse a las trayectorias que siguen las dos órbitas Landsat sobre Cataluña. En otros casos, esta división de una unidad climática en varias se ha forzado para poder beneficiarse de la mayor disponibilidad de imágenes que supone trabajar en ámbitos que compartan ambas órbitas Landsat, o incluso por problemas de nubosidad.

Detección de nubes y selección de imágenes utilizables

Aunque en teoría se dispone de todas las imágenes comprendidas entre enero de 2002 y octubre de 2003 (teniendo en cuenta tanto las imágenes Landsat 7 ETM+ como las Landsat 5 TM para fechas a partir de junio de 2003), la cantidad de imágenes realmente utilizables en cada una de las zonas queda muy reducida. Esto se debe a dos motivos. Por un lado, a la presencia de nubes y neblinas (fenómenos especialmente frecuentes durante el año 2002, uno de los más nubosos de los que existe registro) y, en menor grado, a la presencia de nieves en zonas de montaña. Por otro lado, al hecho que el modelo de corrección radiométrica todavía no funciona suficientemente bien en zonas sin radiación solar directa, con lo cual quedan prácticamente inutilizables las imágenes de invierno en zonas con relieve moderado o alto.

La cubierta nival ha sido detectada siguiendo la metodología descrita en [21], y se han aplicado máscaras con el fin de excluir esta cubierta en el resto de procesos. Algunas imágenes con nubes en poca proporción también han sido utilizadas. En estos casos se ha procedido a la detección de nubes, y a la aplicación de máscaras para su eliminación. La detección de nubes se ha efectuado mediante el uso de umbrales en las imágenes originales, previo estudio del histograma de la imagen. Las sombras de las nubes se han fotointerpretado después de la detección de nubes, puesto que los métodos automáticos no han dado todavía resultados suficientemente operativos.

Aplicación de máscara para vegetación natural

A partir del MCSC se ha creado una máscara de las cubiertas correspondientes a vegetación natural y se ha aplicado en cada una de las imágenes. Con la aplicación de esta máscara se está reduciendo la variabilidad

espectral de las imágenes y, por lo tanto, las posibilidades de confusión entre categorías. Además, dicha máscara permite la creación de una leyenda dedicada exclusivamente a vegetación natural (no cultivos, no urbano, no agua, etc.).

Preparación de las áreas de entrenamiento

Como se comentaba en la introducción, la obtención de áreas de entrenamiento es un proceso lento y al mismo tiempo clave en los procesos de clasificación de imágenes. En este trabajo se opta por la utilización de información temática existente sobre vegetación en el ámbito de Cataluña. Concretamente, se ha utilizado el Mapa de Hábitats de Cataluña, que cartografía los hábitats del territorio catalán basándose en la interpretación y adaptación de la clasificación de hábitats de la Unión Europea según lo propuesto en el documento “*CORINE Biotopes Manual*” (CBM) editado por la Comisión Europea el 1991 [22]. También ha sido utilizado el mismo Mapa de Cubiertas del Suelo de Cataluña, ya sea indirectamente en la aplicación de la máscara, así como, de forma más directa, para obtener áreas de entrenamiento de ciertas categorías (bosques de ribera).

El Mapa de Hábitats se distribuye sobre un corte cartográfico 1:50 000. La elaboración de esta cartografía empieza el año 1998 y el trabajo de gabinete prácticamente ha finalizado a finales del año 2003. La leyenda del Mapa de Hábitats proviene de la Lista de hábitats de Cataluña (LHC) y está formada por 282 categorías. La base cartográfica es de tipo vectorial e incluye información sobre el recubrimiento de cada hábitat.

El levantamiento cartográfico de la base se efectúa por fotointerpretación de pares estereoscópicos de fotografía aérea IRC 1:60 000 del Instituto Cartográfico de Cataluña (ICC) y por transcripción de los elementos fotointerpretados sobre ortofotomapas IRC 1:25 000 sobre papel. Ha contado con el soporte de bases de referencia sobre hidrografía y topografía 1:50 000, así como la prospección de campo y la consulta de documentación y cartografía existente en la zona.

El tamaño mínimo para aceptar un polígono en la fotointerpretación es de 2.25 ha. Los polígonos con anchura igual o inferior a 50 metros y mínimo de 500 metros de longitud se han representado como líneas. A continuación describimos una serie de procesos aplicados sobre esta base, con el fin de aprovechar de un modo eficiente para nuestros propósitos esta información tan exhaustiva tanto territorial como temáticamente.

En primer lugar, para garantizar el uso de áreas de entrenamiento con una máxima homogeneidad temática, se han seleccionado aquellos hábitats con recubrimiento igual o mayor al 80%. Los hábitats seleccionados han sido erosionados (módulo BuffDist de MiraMon) 60 metros en ambos lados de las fronteras de sus polígonos para reducir los posibles efectos de errores de posicionamiento geográfico, ya sean derivados de la escala cartográfica de la base, como debidos a los inevitables errores en la delimitación de hábitats naturales (fronteras borrosas).

Los polígonos erosionados han sido fragmentados a partir de cuadrados de 200 metros de lado. Esto permite que en un mismo hábitat existan píxeles utilizados en el proceso de clasificación y otros que serán reservados como verdad-terreno en una posterior validación de los resultados. Además, con ello se consigue una cierta homogeneización de la superficie que ocupa cada área de entrenamiento.

A pesar de las precauciones tomadas hasta aquí, se puede dudar aún de la calidad de las áreas de entrenamiento. El siguiente paso es una depuración de las áreas seleccionadas utilizando como criterio la variabilidad interna de las mismas. El proceso consiste en comparar, para cada categoría existente en la zona, cada uno de los polígonos con el valor patrón de media y desviación para esa categoría, calculado promediando las medias y desviaciones de todos los polígonos de la categoría. Los polígonos más extremos son filtrados de forma iterativa hasta que los parámetros de la población se estabilizan.

La figura 2 muestra un fragmento de una imagen en falso color (4+5+3). Las líneas amarillas son los bordes de los polígonos del Mapa de Hábitats, en este caso una masa de *Quercus rotundifolia* con 100% de recubrimiento. En color azul aparecen las áreas de entrenamiento erosionadas y fragmentadas. Obsérvese que las flechas amarillas señalan zonas que podríamos considerar como contaminación en la categoría. Las áreas que se han descartado mediante el filtro aparecen con trama roja.

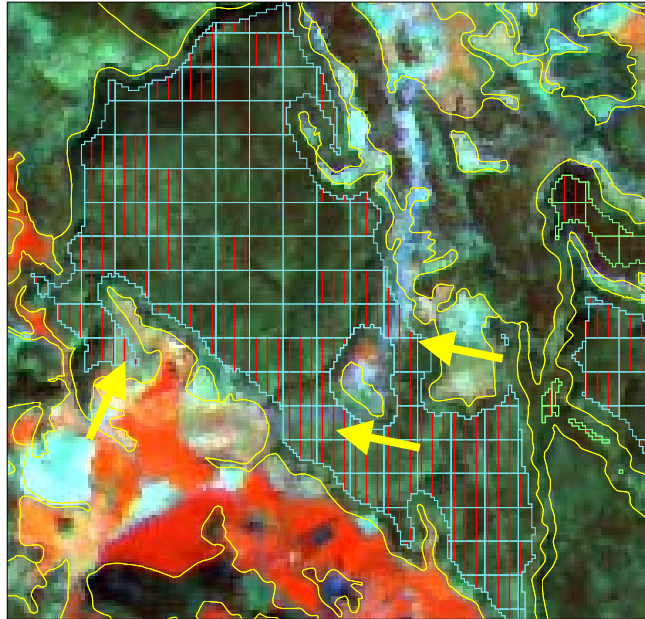


Figura 2: Composición en falso color 4+5+3 de una imagen Landsat 7 ETM+ de fecha 01-04-2003.

De todos los polígonos seleccionados finalmente se reservan, de modo aleatorio, la mitad para realizar una validación independiente de los resultados de la clasificación, dejándose la otra mitad para alimentar el clasificador.

Como se decía anteriormente, la complejidad de la leyenda del Mapa de Hábitats (282 categorías) es excesiva para nuestros fines. Algunas de estas categorías describen mezclas temáticas, aunque, dejando a parte la dificultad para delimitar fotointerpretando o en trabajo de campo este tipo de categorías, no siempre esta mezcla tiene porqué producirse a una escala subpixel. Otras categorías, se bifurcan en otras más concretas debido a la presencia de especies muy minoritarias, con un interés local por lo que refiere a la botánica, pero con tal recubrimiento que hace muy difícil diferenciarlas con imágenes de satélite. Sin embargo, en la clasificación se han introducido todas las categorías disponibles y, una vez analizada la clasificación y la matriz de confusión, se ha reducido esta leyenda original reagrupando categorías con alta confusión entre sí en una más general, o eliminando aquellas categorías de mezcla de hábitats (las cuales sólo aportan confusión).

Una doble reflexión sobre este hecho es necesaria. Por un lado y tal como se verá más adelante, el método de clasificación adoptado presupone que las clases temáticas que queremos clasificar son a menudo una suma (o un conjunto) de distintas clases espectrales. El proceso de filtrado de áreas de entrenamiento puede convertirse en pernicioso para el proceso en el sentido de estar eliminando la posibilidad de que ciertas clases espectrales se asignen a una determinada clase temática (al estar eliminando los polígonos estadísticamente más extremos de cada categoría se corre el riesgo de reducir excesivamente la variabilidad espectral que recoge cada categoría o área de entrenamiento). Por otro lado, el hecho que la leyenda del Mapa de Hábitats sea tan extensa (recordemos, 282 categorías), implica que las categorías delimitadas sean, a excepción de las ya comentadas categorías referentes a mezclas temáticas, muy puras, muy concretas. La utilización de mapas previos también tiene el inconveniente de poder introducir errores por su paulatina obsolescencia, contra la cual el control de homogeneidad radiométrica parece un buen antídoto. Todo esto nos hace advertir que el proceso de filtrado, aunque necesario, debe ser especialmente cuidadoso.

La utilización del Mapa de Hábitats y del MCSC va más allá de la obtención de unas áreas de entrenamiento para un posterior proceso de clasificación digital. El conocimiento del comportamiento espectral de cada categoría, así como la continuidad temporal en la obtención de imágenes de satélite permitirán que la actualización temporal de la cartografía sobre vegetación sea más fácil, más rápida y menos costosa.

Método de clasificación: IsoData masivo + reclasificador espectral a temático

La clasificación mixta ha sido el método utilizado hasta el momento, empleando ClsMix y, para la clasificación no supervisada, el módulo IsoData, ambos integrados en MiraMon. IsoData [16] típicamente agrupa, mediante criterio de mínima distancia, los píxeles de una escena a un centro de clase previamente ubicado según el número de

clusters especificado. Se recalculan de nuevo los centros de clase teniendo en consideración los valores de todos los píxeles que se han incorporado durante la fase anterior. Todo el proceso se repite iterativamente hasta que las asignaciones de los píxeles a los centros llegan a una situación de estabilidad suficientemente satisfactoria, o bien hasta que se llega a un número máximo de iteraciones fijado previamente por el usuario, etc.

La implementación de IsoData (MiraMon) distribuye de dos formas distintas las semillas iniciales. Por un lado, permite introducir un determinado número de semillas sobre la diagonal multivariante de los datos, repartidas de forma equidistante. El rango donde repartir semillas en cada variable se puede restringir con un criterio basado en la media y desviación de cada banda, o bien en todo el rango posible. Por otro lado, se permite introducir otro grupo de semillas de forma aleatoria en todo el espacio multivariante de la imagen. Otra particularidad interesante de la implementación de IsoData utilizada, es la posibilidad de utilizar diferentes núcleos clasificadores para investigar cuál es el más adecuado a nuestros propósitos: mínima distancia, distancia de *Manhattan*, máxima probabilidad, etc. Con ello se tiene mayor flexibilidad en lo que respecta al tratamiento más o menos paramétrico del clasificador. Por el momento se han probado dos criterios distintos para el cálculo de distancias entre píxeles y clases o *clusters*: la distancia euclidiana y la distancia *Manhattan*.

Finalmente, comentar que la implementación utilizada permite la introducción de un gran número de variables (temperatura, pluviometría, reflectancia, etc.) sin necesidad de degradar a formato byte, así como la posibilidad de obtener gran número de categorías estadísticas (hasta un máximo de 32 767).

Las variables introducidas en IsoData provienen de todas las bandas de las imágenes disponibles en cada zona (en unidades de reflectancias, excepto la banda 6, en unidades de grado centígrado) y de índices derivados de ellas (NDVI), así como de variables orográficas (modelo digital de pendientes, de resolución nominal 30m), y de variables climáticas (precipitación media anual, temperatura media de las mínimas en enero, temperatura media de las mínimas en abril, temperatura media de las máximas en julio, radiación mediana anual, todas ellas a resolución nominal 180m).

Como puede observarse, las variables introducidas cuantifican diferentes propiedades físicas del terreno, y, por lo tanto, utilizan diferentes unidades para representarse. Esto hace que las variables tengan varianzas muy dispares entre sí, hecho que no responde a su variabilidad interna real sino a la distinta magnitud que toman las unidades con la que se representa cada imagen. Es necesario, pues, escalar o estandarizar de algún modo las imágenes si no queremos que este fenómeno, totalmente irrelevante, tenga una influencia excesiva en los resultados de la clasificación.

En este sentido, se han seguido dos propuestas distintas para la estandarización de variables en la clasificación no supervisada. Por un lado, una normalización de cada imagen de modo que la imagen normalizada tenga media 0 y desviación estándar 1. Por otro lado, el uso del análisis de componentes principales (ACP). El ACP no se realiza sobre el conjunto completo de variables, sino que se escogen los mejores -los más explicativos- CP de conjuntos de variables con alta correlación entre sí (los mejores CP del conjunto de bandas del canal visible por un lado, del infrarrojo por el otro, de las variables climáticas por otro, etc.). De este modo, se minimiza la probabilidad de descartar un CP poco explicativo con relación al resto, pero con información válida para la discriminación de nuestras categorías. Ambos métodos tienen la desventaja que desperdician la información física que ofrece cada una de las imágenes. Futuras líneas de trabajo pretenden dedicarse a investigar nuevas formas de estandarización de las variables que preserven su significado físico al mismo tiempo que neutralicen el efecto de las unidades sobre la varianza de cada variable.

La segunda fase del clasificador se basa en el módulo ClsMix de MiraMon. ClsMix es un programa de clasificación mixta de imágenes. Utilizando unos umbrales de frecuencia dados por el usuario, ClsMix reclasifica cada clase espectral de una imagen obtenida con clasificación no supervisada en clases temáticas, previamente definidas estas últimas por el usuario mediante áreas de entrenamiento. ClsMix utiliza básicamente dos parámetros para realizar la reclasificación. Por un lado, la fidelidad de la clase espectral a la clase temática, que es la mínima proporción con que una clase espectral se encuentra dentro de la clase temática a la que se asignará finalmente. Esta proporción se expresa en función de la clase espectral. Por otro, la representatividad de la clase espectral en la clase temática, que es la proporción de la clase espectral dentro la clase temática (en relación con la clase temática).

La clasificación resultante se valida con una matriz de confusión (módulo MatConf de MiraMon) utilizando las áreas de entrenamiento independientes reservadas para tal fin.

3. Resultados y conclusiones

Hasta el momento la metodología aquí descrita se ha aplicado sobre dos de las zonas climáticas definidas anteriormente. Por un lado, se ha tratado la zona denominada “Temperatura Media, Precipitación Media, Radiación Alta” (código 3 de la leyenda de la figura 1 y tabla 2) y, por otro, la de “Temperatura Alta, Precipitación Muy Alta en Estación Húmeda y Media en Estación Seca, Radiación Media” (código 6 de la leyenda de la figura 1 y tabla 2). En la tabla 2 se muestra la superficie ocupada por cada una de estas zonas, las fechas de las imágenes Landsat utilizadas durante la clasificación, y un resumen de los resultados de la clasificación.

Código Leyenda	Superficie válida utilizada (píxeles)	Fechas Imágenes Landsat utilizadas en la clasificación	Número máximo de categorías clasificadas (con mezcla)	Porcentaje de acierto (ponderación sobre los clasificados)
3	7 460 696	01-04-2002, 19-05-2002, 02-08-2003, 24-09-2002	17 (34)	86.70%
6*	840 331	19-03-2003, 13-06-2002, 02-08-2003	14 (28)	86.9%

Tabla 2: Zonas Climáticas donde se ha aplicado la metodología y resultados de la clasificación. (*) Sólo el área comprendida entre órbitas.

Los resultados obtenidos hasta el momento no son aún definitivos pero apuntan ya las tendencias generales a las que conduce la metodología descrita en esta comunicación. En general, se permite discriminar de forma aceptable un número de categorías forestales que varía entre 15 y 20, y hasta 34 si consideramos también mezclas de categorías.

De las matrices de confusión se deduce que las áreas de entrenamiento obtenidas de forma automática son suficientemente válidas para utilizar en la clasificación. Las dimensiones de la matriz de confusión completa (con todas las categorías) sobrepasa los límites que permiten este documento. La tabla 3 se muestra un fragmento de una matriz para unas categorías características. En general, las categorías se muestran bien discriminables las unas de las otras, e incluso se llega a diferenciar satisfactoriamente a nivel de especie (ver, por ejemplo, el caso de los pinos, tradicionalmente muy difíciles de separar). Algunas de ellas, sin embargo, no consiguen clasificarse. Se trata de categorías con poca representatividad en la zona climática (por ello estos errores no tienen un peso importante en el porcentaje de acierto global). Los motivos por los que no aparecen se atribuyen a un excesivo filtrado de sus áreas de entrenamiento. Al ser tan poco frecuentes y tener pocos píxeles de áreas de entrenamiento, los valores patrones utilizados en el filtrado de estas categorías pueden estar desviados respecto la realidad y, consecuentemente, introducir mayor confusión.

		Verdad-Terreno						Total	%com	Exusu
		1	2	3	4	5	6			
Clasificación	0	4767	2376	384	3928	1671	6135	19261		
	1	16837	238	30	98	132	646	17981	6.36	93.64
	2	862	5832	13	84	0	12	6803	14.30	85.70
	3	215	29	2657	0	0	0	2901	8.40	91.60
	4	168	377	0	15065	2	1424	17036	11.60	88.40
	5	404	0	0	0	5408	141	5953	9.20	90.80
	6	788	18	0	1361	77	7930	10174	22.10	77.90
	Total	24041	8870	3084	20536	7290	16288	80109		
%om	29.97	34.30	13.80	26.60	25.80	51.30				
ExProd	70.03	65.70	86.20	73.40	74.20	48.70				

Tabla 3: Matriz de confusión para alguna de las categorías de la leyenda. Leyenda: 1= Otros (conjunto de 16 categorías no contempladas en esta tabla); 2= Matorrales y tomillares; 3= Prados silicícolas; 4= *Pinus halepensis*; 5= *Pinus sylvestris*; 6= *Pinus nigra*.

Los errores de omisión suelen ser considerablemente mayores a los de comisión. Obsérvese, sin embargo, que en general la mayoría de píxeles “omitidos” en una categoría no se confunden con otras sino que no se clasifican. Este hecho está íntimamente relacionado con el parámetro de fidelidad usado en ClsMix. Cuánto más alto es este parámetro, más exigente es el clasificador, más difícil es que un píxel se clasifique en una determinada categoría. El número de píxeles no clasificados puede reducirse si, a partir de los resultados de ClsMix, definimos nuevas categorías, mezcla de dos o más categorías presentes en las áreas de entrenamiento introducidas. A parte de reducir los píxeles no clasificados, estaremos ampliando la leyenda de la clasificación. Por ejemplo, en la clasificación presentada en la tabla 3, pasamos de 21 categorías y 4 685 304 píxeles clasificados a, definiendo nuevas categorías de mezcla, 34 categorías y 6 645 330 píxeles clasificados.

Un hecho a destacar es que las confusiones existentes se suelen dar entre categorías parecidas, o bien que suelen aparecer asociadas en la naturaleza (pinos y matorrales, por ejemplo). Esto puede aprovecharse para presentar los resultados cartográficos de la clasificación con distintos niveles de detalle y precisión temática (aunque a mayor detalle, menor grado de acierto general y por categoría). En lugar de intentar discriminar diferentes especies de pinos, por ejemplo, podemos generar una categoría más general que incluya a todas estas especies. La tabla 4 es la matriz de confusión de la misma clasificación que la tabla 3, pero con las categorías agrupadas. Obsérvese que el porcentaje de acierto ponderado sobre las zonas libre de NODATA aumenta del 86.7% al 93.0%.

		Verdad-terreno							Total	%com	ExUsu
		1	2	3	4	5	6	7			
Clasificación	0	2837	565	11789	476	3533	0	61	19261		
	1	6235	66	96	106	267	15	18	6803	8.30	91.70
	2	38	3316	23	19	42	0	2	3440	3.60	96.40
	3	490	0	31508	118	1063	0	2	33181	5.00	95.00
	4	68	2	82	2668	14	0	0	2834	5.90	94.10
	5	628	44	795	272	12795	0	0	14534	12.00	88.00
	6	0	0	0	0	0	22	0	22	0.00	100.00
	7	0	0	0	0	0	0	34	34	0.00	100.00
Total		10296	3993	44293	3659	17714	37	117	80109		
%om		39.40	17.00	28.90	27.10	27.80	40.50	70.90			
ExUsu		60.60	83.00	71.10	72.90	72.20	59.50	29.10			

Tabla 4: Matriz de confusión para las categorías agrupadas procedentes de la clasificación. Leyenda: 0= NODATA, 1= Coscojas, 2= Prados, 3= Pinos, 4= Robles, 5= Encinares, 6= Carrizal, 7= Suelo.

El uso del filtrado de las áreas de entrenamiento mediante el estudio de su variabilidad interna mejora sensiblemente los resultados de la clasificación. Por ejemplo, en la zona 6, utilizando una misma imagen IsoData con 310 *clusters*, con una leyenda de 16 categorías y exigiendo una fidelidad en ClsMix del 70%, el porcentaje de acierto aumenta modestamente del 86.9 (sin filtrado) a 87.1%. El hecho que el área esté dominada por la categoría *Pinus halepensis*, la cual tiene un buen porcentaje de acierto en ambos casos, nos puede enmascarar algunas diferencias existentes entre los resultados con o sin filtro. Podemos observar como el uso del filtro reduce la confusión entre ciertas categorías y al mismo tiempo permite clasificar más superficie que el uso de áreas de entrenamiento sin filtrar. Sin embargo, dado que el filtrado iterativo descarta más de la mitad de la superficie de áreas de entrenamiento, se tendrá que estudiar con más detalle si es posible mejorar los resultados variando los parámetros estadísticos de eliminación, los parámetros de estabilización de la población, y las imágenes de referencia.

4. Referencias

- [1] Mapa de Cobertes del Sòl de Catalunya, 1999. The Internet. (<http://www.creaf.uab.es/mcsc/index.htm>).
- [2] Ibàñez, J.J.; Burriel, J.A. y Pons, X., 2002. El Mapa de Cobertes del Sòl de Catalunya: Una eina per al coneixement, la planificació i la gestió del territori. Perspectives territorials, 31, pp. 10-65.
- [3] Burriel, J.A.; Ibàñez, J.J. y Pons, X., 2002. El Mapa de Cubiertas del Suelo de Cataluña: Herramienta para la gestión y la planificación territorial. Montes para la sociedad del nuevo Milenio. III Congreso Forestal Español, 3, pp. 83-89, Junta de Andalucía. Granada.

- [4] Corine Land Cover Project, 1985. Technical Guide. Direction Generale de l'Environment, Commission des Communautés européennes, Bruxelles, Belgique.
- [5] Baulies, X.; Joaquinet, M.; Pineda, L. y Tarda, A. (1991). El proyecto CORINE-Land Cover en el sector Oriental Peninsular y las Islas Baleares. IV Reunión Científica de la Asociación Española de Teledetección, pp. 244- 250. A.E.T. Sevilla.
- [6] Viñas, O. y Baulies, X., 1995. 1:250 000 Land-use map of Catalonia (32 000 km²) using multitemporal Landsat-TM data. *International Journal of Remote Sensing*, 16(1), pp. 129-146.
- [7] Land Cover Map of Great Britain, 1990. The Internet. (<http://www.ceh.ac.uk/data/lcm/index.htm>)
- [8] Fuller, R.M.; Groom, G.B. y Jones, A.R., 1994. The Land Cover Map of Great Britain: an automated classification of Landsat Thematic Mapper data. *Photogrammetric Engineering & Remote Sensing*, 60, pp. 553-562.
- [9] Wulder, M., 2002. Mapping the land cover of the forested area of Canada with Landsat data. Proceedings of 2002 International Geoscience and Remote Sensing Symposium, June 24-28, Toronto, Canada.
- [10] Homer, C.G.; Ramsey, D.R.; Edwards, T.C. Jr. y A. Falconer, 1997. Landscape Cover-Type Modeling Using a Multi-Scene Thematic Mapper Mosaic. *Photogrammetric Engineering & Remote Sensing*, 63(1), pp. 59-67.
- [11] Earth Observation for Sustainable Development of Forests, 2004. The Internet. (http://www.pfc.forestry.ca/eosd/cover/index_e.html)
- [12] Palà, V. y Pons, X., 1995. Incorporation of relief into geometric corrections based on polynomials. *Photogrammetric Engineering & Remote Sensing*, 61(7), pp. 935-944.
- [13] Pons, X., 2002. MiraMon. Sistema d'Informació Geogràfica i software de Teledetecció. Centre de Recerca Ecològica i Aplicacions Forestals, CREAF. Bellaterra. ISBN: 84-931323-4-9.
- [14] Pons, X. y Solé-Sugrañes, Ll., 1994. A Simple Radiometric Correction Model to Improve Automatic Mapping of Vegetation from Multispectral Satellite Data. *Remote Sensing of Environment*, 48, pp. 191-204.
- [15] Salvador, R.; Pons, X. y Diego, F., 1996. Validación de un método de corrección radiométrica sobre diferentes áreas montañosas. *Revista de Teledetección*, 7, pp. 21-25.
- [16] Duda, R. D. y Hart, P. E., 1973. *Pattern Classification and Scene Analysis*. John Wiley & Sons. New York.
- [17] Pons, X., 1996. Estimación de la Radiación Solar a partir de modelos digitales de elevaciones. Propuesta metodológica. En: Juaristi, J. y Moro, I. (eds.), VII Coloquio de Geografía Cuantitativa, Sistemas de Información Geográfica y Teledetección. Vitoria-Gasteiz
- [18] Ninyerola, M.; Pons, X. y Roure, J.M., 2000. A methodological approach of climatological modelling of air temperature and precipitation through GIS techniques. *International Journal of Climatology*, 20, pp. 1823-1841.
- [19] Ninyerola, M., 2000. Modelització climàtica mitjançant tècniques SIG i la seva aplicació a l'anàlisi quantitativa de la distribució d'espècies vegetals a l'Espanya peninsular. Tesis doctoral (capítol 3). Universitat Autònoma de Barcelona. (dirigida por los Drs. Xavier Pons y Joan M. Roure).
- [20] Atlas Climàtic Digital de Catalunya, 2001. The Internet. (<http://magno.uab.es/atles-climatic>).
- [21] Dozier, J., 1989. Spectral Signature of Alpine Snow Cover from Landsat Thematic Mapper. *Remote Sensing of Environment*, 28, pp. 9-22.
- [22] Cartografia dels Hàbitats de Catalunya, 1991. The Internet. (<http://www.genclat.net/mediamb/pn/2chabitats-2.htm>)