

# DIFERENCIACIÓN DE CUBIERTAS FORESTALES PARA EL MCSC A PARTIR DE LA CLASIFICACIÓN DE IMÁGENES LANDSAT

G. Moré Gómez<sup>1</sup>, X. Pons Fernández<sup>2,1</sup>, J.A. Burriel Moreno<sup>1</sup>, R. Castells Ferré<sup>1</sup>, J.J. Ibàñez Martí<sup>1</sup> y X. Roijals Lara<sup>1</sup>

<sup>1</sup> Centro de Investigación Ecológica y Aplicaciones Forestales (CREAF). Edificio C. Universidad Autónoma de Barcelona. 08193-BELLATERRA (Barcelona, España). Correo electrónico: g.more@creaf.uab.es

<sup>2</sup> Departament de Geografia, Fac. Lletres UAB, Edificio B. Universidad Autónoma de Barcelona. 08193-BELLATERRA (Barcelona, España)

## 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. La cartografía obtenida se utilizará para la ampliación de leyenda de la nueva versión del Mapa de Cubiertas de Cataluña (MCSC). Las áreas de entrenamiento se obtienen de otras bases ya existentes sobre vegetación, como son el Mapa de Hábitats de Cataluña y el mismo MCSC, 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 80%, o, a un nivel más general, una leyenda de 7 categorías (12 con categorías de mezclas temáticas) por zona con un 93.0% de porcentaje de acierto global.

Palabras clave: *Clasificación Mixta, IsoData, Cartografía vegetación*

## 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 (MCSC, 1999; IBÁÑEZ *et al.*, 2002, BURRIEL *et al.*, 2001).

Actualmente, se está llevando a cabo la segunda versión del MCSC, cuyo principal

soporte son ortofotografías 1:5.000 en color del período 2000/03. Otra comunicación en estas Jornadas explica más detalles sobre esta actualización y metodología del MCSC. 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 planearse su utilización. De hecho, la utilización de imágenes de teledetección espacial para cartografiar con algún nivel de diferenciación las zonas forestales no es ni mucho menos nueva. Simplemente como ejemplo citemos el proyecto CORINE Land Cover en el ámbito de la fotointerpretación (CORINE, 1985; BAULIES et al., 1991), el Mapa de Usos del Suelo de Cataluña en el ámbito del proceso digital de imágenes (VIÑAS & BAULIES 1995), o el Mapa de Usos del Suelo de Gran Bretaña (LCMGB, 1990; FULLER et al., 1994).

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.

La delimitación de áreas de entrenamiento por toda Cataluña mediante trabajo de campo requeriría un elevado coste en tiempo y esfuerzo. El presente trabajo utiliza bases cartográficas sobre vegetación realizadas recientemente, las cuales se propone mejorar por actualización temporal periódica y por el tratamiento digital, más objetivo. El mismo MCSC y el Mapa de Hábitats de Cataluña serán utilizados, con el tratamiento previo correspondiente, en el proceso para obtener áreas de entrenamiento de calidad de forma automática.

## 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 & PONS (1995), implementada en el módulo CorrGeom de MiraMon. 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 & SOLÉ-SUGRAÑES (1994), y contrastado en SALVADOR et al. (1996).

### 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 en buena medida 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 (DUDA & HART, 1973) 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 (PONS 1996, NINYEROLA et al., 2000; NINYEROLA, 2000; ACDC, 2201):

- 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 figura 1 y tabla 1.

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 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 presen-

cia 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 DOZIER (1989), 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

<b>Código Leyenda</b>	<b>Descripción Zona Climática</b>	<b>Superficie Total (km<sup>2</sup>)</b>
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

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 (CHC, 1991). 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 trascripción 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

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 a partir 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 verdadero 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.

Como se ha comentado, se utilizan dos indicadores para el filtrado de áreas de entrenamiento, que son la varianza y la media de cada polígono. La varianza nos informa de la pureza interna del polígono (cuanto más heterogéneo sea un polígono, más varianza tendrá). La media, en cambio, nos puede indicar diferencias para polígonos supuestamente de la misma categoría (descartaremos aquellos polígonos

cuya media difiera mucho de la media de su categoría).

Tres han sido las variables o imágenes de referencia para el filtrado de áreas de entrenamiento. Por un lado, el valor mínimo (MIN\_NDVI) y máximo (MAX\_NDVI), en cada píxel, de entre todas las imágenes NDVI utilizadas en la clasificación. Estas variables de algún modo nos informan mayormente de la densidad de vegetación en cada píxel. Por otro, un índice entre MIN\_NDVI y MAX\_NDVI (MIN\_MAX), calculado como  $100 * (MAX\_MIN) / 4$ , el cual caracteriza bastante bien la vegetación tanto caducifolia como perennifolia.

El filtrado consiste en la eliminación de aquellos polígonos más distantes de los valores patrón para cada categoría. El criterio de eliminación varía en función de si se está utilizando la varianza o la media:

Criterio 1: Para la varianza:  $x > X + a * SD$

Criterio 2: Para la media:  $x < (X - a * SD)$ , o bien  $x > (X + a * SD)$

( $x$  = valor del polígono;  $X$  = Media de los polígonos para la categoría correspondiente;

$a$  = Número de desviaciones estándar utilizadas (en este caso 2);  $SD$  = Desviación

Estándar de los polígonos para la categoría correspondiente).

Obsérvese para la varianza que no es necesario "castigar" aquellos polígonos más puros temáticamente (de aquí que sólo se corte la cola superior). En cambio, sí que nos interesa eliminar aquellos polígonos que sean diferentes entre sí, a pesar de que según el Mapa de Hábitats consten como una misma categoría.

Las iteraciones se realizan para cada una de las categorías. En primer lugar, se eliminan los polígonos utilizando el criterio 1 (varianza) de una de las imágenes de referencia. La eliminación de los extremos se repite iterativamente hasta que la población se estabiliza. Los polígonos que han superado el filtro son sometidos a un nuevo proceso iterativo utilizando el criterio 2 (media). Una vez estabilizada la población se repite todo el proceso con otra imagen de referencia, y así sucesivamente.

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 aparte la dificultad para delimitar fotointerpretando o en trabajo de campo este tipo de categorías, no siempre esta mezcla tiene porqué producirse a una escala subpíxel. 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 (DUDA & HART, 1973) 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 dis-

tancias 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 180 m).

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 de 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.

## 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.

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 (ver tabla 3). 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 a la realidad y, consecuentemente, introducir mayor confusión.

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 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. Aparte de reducir los píxeles no clasificados, estaremos ampliando la leyenda de la clasificación. Por ejemplo, en la cla-

sificación 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

Clasificación	Verdad-terreno															Total	%com	ExUsu
	2	3	5	6	7	8	11	12	13	14	15	18	19	20	21			
0	2376	85	384	93	3	3928	324	150	1671	6135	21	3312	0	48	13	<b>19261</b>	%com	ExUsu
2	<b>5832</b>	0	13	0	0	84	0	106	0	12	0	267	15	16	2	<b>6803</b>	14,30	85,70
3	0	<b>198</b>	7	1	28	0	0	10	0	0	0	34	0	0	1	<b>288</b>	31,20	68,80
5	29	34	<b>2657</b>	47	103	0	0	1	0	0	23	6	0	0	1	<b>2901</b>	8,40	91,60
6	0	0	4	<b>71</b>	5	0	0	0	0	0	0	2	0	0	0	<b>82</b>	13,40	86,60
7	0	1	1	0	<b>159</b>	0	8	0	0	0	0	0	0	0	0	<b>169</b>	5,90	94,10
8	377	0	0	0	0	<b>15065</b>	0	0	2	1424	0	71	0	0	0	<b>17036</b>	11,60	88,40
11	1	2	0	0	0	1	<b>2477</b>	2	81	0	0	12	0	0	0	<b>2650</b>	6,50	93,50
12	25	0	0	0	0	0	0	<b>152</b>	0	0	0	2	0	0	0	<b>184</b>	17,40	82,60
13	0	0	0	0	0	0	111	0	<b>5408</b>	141	0	274	0	0	0	<b>5953</b>	9,20	90,80
14	18	0	0	0	0	1361	6	0	77	<b>7930</b>	0	453	0	0	0	<b>10174</b>	22,10	77,90
15	5	0	0	0	0	0	0	0	0	0	<b>11</b>	0	0	0	2	<b>18</b>	38,90	61,10
18	207	17	18	0	9	97	106	166	51	646	0	<b>12726</b>	0	0	0	<b>14534</b>	12,40	87,60
19	0	0	0	0	0	0	0	0	0	0	0	0	<b>22</b>	0	0	<b>22</b>	0,00	100,00
20	0	0	0	0	0	0	0	0	0	0	0	0	0	<b>13</b>	0	<b>13</b>	0,00	100,00
21	0	0	0	0	0	0	0	0	0	0	0	0	0	0	<b>21</b>	<b>21</b>	0,00	100,00
Total	<b>8870</b>	<b>337</b>	<b>3084</b>	<b>212</b>	<b>307</b>	<b>20536</b>	<b>3032</b>	<b>587</b>	<b>7290</b>	<b>16288</b>	<b>55</b>	<b>17159</b>	<b>37</b>	<b>77</b>	<b>40</b>	<b>80109</b>		
% om	34,30	41,20	13,80	66,50	48,20	26,60	18,30	74,10	25,80	51,30	80,00	25,80	40,50	83,10	47,50			
ExProd	65,70	58,80	86,20	33,50	51,80	73,40	81,70	25,90	74,20	48,70	20,00	74,20	59,50	16,90	52,50			

Número total de píxeles de verdad terreno que se pueden usar en la evaluación (A) = 80.109  
 Número total de píxeles de verdad terreno que se pueden usar en la evaluación (no sobre NODATA) (B) = 60.848  
 Número de aciertos (C) = 52.742  
 Porcentaje de acierto global (C/A) = 65,80%  
 Porcentaje de acierto global (sólo los clasificados) (C/B) = 86,70%

**Tabla 3.** Matriz de confusión de la clasificación sobre la Zona 3, con IsoData con criterio de distancia Euclidiana (500 semillas sobre diagonal multivariante y 1200 aleatorias), umbral de convergencia del 0,5% y tamaño mínimo de cluster 100 píxeles, y ClsMix con fidelidad del 70% y representatividad 0.00.1 Leyenda: 0= NODATA, 2= Matorrales y tomillares, 3= Prados basófilos, 5= Prados silicícolas, 6= Landas (*Calluna vulgaris*), 7= *Genista balansa*, 8= *Pinus halepensis*, 11= *Quercus humilis*, 12= *Quercus faginea*, 13= *Pinus sylvestris*, 14= *Pinus nigra*, 15= *Pinus pinaster*, 18= *Quercus rotundifolia*, 19= Carrizal, 20= Suelos calcarios, 21= Peñascos. Categorías no clasificadas y que no aparecen en la matriz de confusión: *Coscoja*, *Brachypodium retusum*, *Cistus laurifolius*, *Quercus pyrenaica*, *Quercus ilex de tierra baja* y *Quercus ilex de montaña*



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%.

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 elimina-

ción, los parámetros de estabilización de la población, y las imágenes de referencia.

## BIBLIOGRAFÍA

- ACDC; 2001. *Atlas Climàtic Digital de Catalunya*. The Internet. (<http://magno.uab.es/atles-climatic>)
- BAULIES, X.; JOAQUINET, M.; PINEDA, L. Y TARDA, A.; 1991. El proyecto CORINE-Land Cover en el sector Oriental Peninsular y las Islas Baleares. *En: IV Reunión Científica de la Asociación Española de Teledetección*: 244-250. A.E.T. Sevilla.
- BURRIEL, J.A.; IBÁÑEZ, J.J. Y PONS, X.; 2002. El Mapa de Cubiertas del Suelo de Cataluña: Herramienta para la gestión y la planificación territorial. *En: Junta de Andalucía-S.E.C.F. (eds.), Actas del III Congreso Forestal Español-Sierra Nevada 2001, III*: 83-89. Coria Gráficas. Sevilla.
- CHC; 1991. *Cartografia dels Hàbitats de Catalunya*. The Internet. (<http://www.gen-cat.net/mediamb/pn/2chabitats-2.htm>)
- CORINE LAND COVER PROJECT; 1985. *Technical Guide*. Direction Generale de l'Environnement,

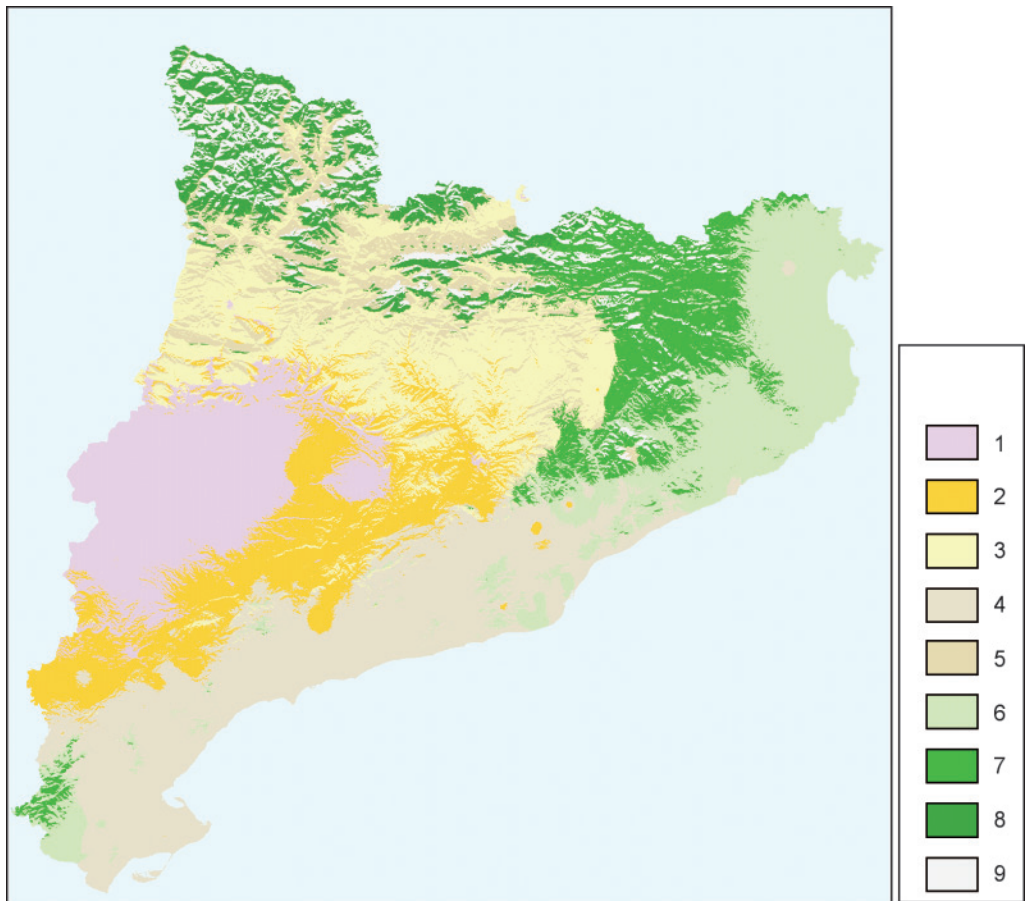
		Verdad-terreno							Total	%com	ExUsu
		1	2	3	4	5	6	7			
Clasificación	0	2837	565	11789	476	3533	0	61	<b>19261</b>		
	1	<b>6235</b>	66	96	106	267	15	18	<b>6803</b>	8,30	91,70
	2	38	<b>3316</b>	23	19	42	0	2	<b>3440</b>	3,60	96,40
	3	490	0	<b>31508</b>	<b>118</b>	1063	0	2	<b>33181</b>	5,00	95,00
	4	68	2	82	<b>2668</b>	14	0	0	<b>2834</b>	5,90	94,10
	5	628	44	795	272	<b>12795</b>	0	0	<b>14534</b>	12,00	88,00
	6	0	0	0	0	0	<b>22</b>	0	<b>22</b>	0,00	100,00
	7	0	0	0	0	0	0	<b>34</b>	<b>34</b>	0,00	100,00
Total	<b>10296</b>	<b>3993</b>	<b>44293</b>	<b>3659</b>	<b>17714</b>	<b>37</b>	<b>117</b>	<b>80109</b>			
%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				
Número total de píxeles de verdad terreno que se pueden usar en la evaluación (A) = 80.109											
Número total de píxeles de verdad terreno que se pueden usar en la evaluación (no sobre NODATA) (B) = 60.848											
Número de aciertos (C) = 56.578											
Porcentaje de acierto global (C/A) = 70,6%											
Porcentaje de acierto global (sólo los clasificados) (C/B) = 93,0%											

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

- Comission des Communautés europeennes. Bruxelles. Belgique.
- DOZIER, J.; 1989. Spectral Signature of Alpine Snow Cover from Landsat Thematic Mapper. *Rem. Sens. Env.* 28: 9-22.
- DUDA, R.D. & HART, P.E.; 1973. *Pattern Classification and Scene Analysis*. John Wiley & Sons. New York.
- FULLER, R.M.; GROOM, G.B. & JONES, A.R.; 1994. The Land Cover Map of Great Britain: an automated classification of Landsat Thematic Mapper data. *Photogrammetric Eng. Rem. Sens.* 60: 553-562.
- IBÁÑEZ, 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: 10-65.
- LCMGB; 1990. *Land Cover Map of Great Britain*. The Internet. (<http://www.ceh.ac.uk/data/lcm/index.htm>)
- MCSC; 1999. *Mapa de Cobertes del Sòl de Catalunya*. The Internet. (<http://www.crea.uab.es/mcsc/index.htm>)
- 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*. Capítulo 3. Tesis doctoral. Universitat Autònoma de Barcelona. Barcelona.
- NINYEROLA, M.; PONS, X. & ROURE, J.M.; 2000. A methodological approach of climatological modelling of air temperature and precipitation through GIS techniques. *Int. J. Clim.* 20: 1823-1841
- PALÀ, V. & PONS, X.; 1995. Incorporation of relief into geometric corrections based on polynomials. *Photogrammetric Eng. Rem. Sens.* 61(7): 935-944.
- PONS, X.; 1996. Estimación de la Radiación Solar a partir de modelos digitales de evaciones. Propuesta metodológica. En: J. Juaristi y I. Moro (eds.), *VII Coloquio de Geografía Cuantitativa, Sistemas de Información Geográfica y Teledetección*. Vitoria-Gasteiz
- PONS, X. & SOLÉ-SUGRAÑES, LL.; 1994. A Simple Radiometric Correction Model to Improve Automatic Mapping of Vegetation from Multispectral Satellite Data. *Rem. Sens. Env.* 48: 191-204.
- SALVADOR, R.; X. PONS Y DIEGO, F.; 1996. Validación de un método de corrección radiométrica sobre diferentes áreas montañosas. *Rev. Teledet.* 7:21-25.
- VIÑAS, O. & BAULIES, X.; 1995. 1:250 000 Land-use map of Catalonia (32 000 km<sup>2</sup>) using multitemporal Landsat-TM data. *Int. J. Rem. Sens.* 16(1): 129-146.

## DIFERENCIACIÓN DE CUBIERTAS FORESTALES PARA EL MCSC A PARTIR DE LA CLASIFICACIÓN DE IMÁGENES LANDSAT

G. Moré Gómez, X. Pons Fernández, J.A. Burriel Moreno, R. Castells, J.J. Ibàñez Martí y X. Roijals Lara



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