

Cartografía de usos y cubiertas del suelo del sureste de la Península Ibérica a partir de la clasificación de imágenes Landsat en el quinquenio 2000-2004

Juan José Vidal Macua⁽¹⁾, Xavier Pons Fernández⁽¹⁾, Alaitz Zabala Torres⁽¹⁾ y Miquel Ninyerola Casals⁽²⁾

⁽¹⁾ Departamento de Geografía de la Universidad Autónoma de Barcelona. Grupo de Investigación GRUMETS.

Emails: Juanjo.Vidal@uab.cat, Xavier.Pons@uab.cat, Alaitz.Zabala@uab.cat

⁽²⁾ Departamento de Biología Animal, Biología Vegetal y Ecología de la Universitat Autònoma de Barcelona. Grupo de Investigación GRUMETS. e-mail: Miquel.Ninyerola@uab.cat

Resumen: Se presenta la metodología utilizada en la obtención de la cartografía de usos y cubiertas del suelo para la zona oriental de Andalucía (ámbito de 29.259 km²) en el quinquenio 2000-2004, empleando el clasificador híbrido y utilizando imágenes Landsat junto con las variables auxiliares. Las áreas de entrenamiento se han obtenido de manera semiautomática a partir del SIOSE y depuradas con umbrales de NDVI. Se han ejecutado bancos de pruebas en función de las variables incluidas en el clasificador. Los mejores resultados, en cuanto a porcentaje de acierto y área clasificada, se han obtenido excluyendo la radiación solar de invierno y la banda 1. El proceso se ha realizado por separado para cubiertas naturales y urbanas, con un grado de acierto global superior al 88%, y para cultivos, con un acierto superior al 86%.

Palabras clave: cartografía de cubiertas del suelo, Landsat, clasificadores estadísticos, MiraMon

Mapping land use and cover in the southeastern Iberian Peninsula in 2000-2004 using Landsat imagery

Abstract: This paper explains the methodology used to produce land cover mapping for the east of Andalusia (29.259 km²) in 2000-2004. We used a hybrid classifier using Landsat imagery and auxiliary variables. Training areas were obtained semi-automatically from the SIOSE database and filtered with NDVI thresholds. Testbeds were conducted based on the variables included in the statistical classifier. The best results in terms of overall accuracy and classified area were obtained by excluding winter solar radiation and band 1. The process was carried out separately for natural and urbanized areas, with an overall accuracy of over 88%, and for crops with an accuracy of over 86%.

Keywords: land cover mapping, Landsat, statistical classifiers, MiraMon

1. INTRODUCCIÓN

El presente trabajo forma parte del proyecto DinaCliVe (Análisis espaciotemporal de las cubiertas del suelo y del estrés de la vegetación en la Península Ibérica a la luz de medio siglo, 1980-2025, de dinámica climática y sus anomalías) financiado por el Ministerio de Economía y Competitividad. El proyecto divide su periodo de estudio por quinquenios. Una fase principal es desarrollar una metodología sólida para la clasificación de series temporales de imágenes de satélite y grandes extensiones de terreno. Presentamos aquí una primera aproximación para la obtención de la cartografía temática en dicha fase. El objetivo del presente estudio es realizar una cartografía de usos y cubiertas del suelo del sector suroriental de la Península Ibérica a partir de imágenes Landsat (30 m de resolución), y de otras variables de carácter topo-climático, para el periodo de estudio 2000-2004 a partir de clasificadores estadísticos.

2. ÁREA DE ESTUDIO

El ámbito de trabajo (Figura 1) se sitúa en el sureste peninsular, en concreto el área que ocupa la escena 200-034 de la órbita del satélite Landsat, entre las longitudes 4° 25' W y 1° 58' E y las latitudes 38° 22' N y 36° 40' S, con una extensión continental de 29.259 km². El rango altitudinal del área de estudio va desde los 0 m en las zonas costeras, a los 3.478 m del pico Mulhacén.

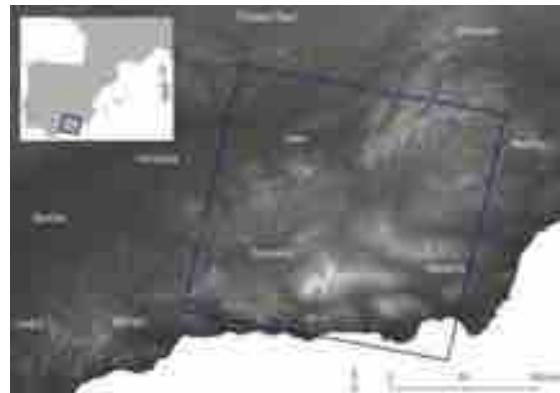


Figura 1. Ámbito de estudio

La elección de la escena ha sido motivada por su amplitud en cuanto a rango altitudinal (incluyendo zonas de elevada irregularidad topográfica y zonas, como fondos de valle, de rugosidad menos compleja) y por su heterogeneidad en cuanto al modelo de ocupación antrópico (diversidad de usos de suelo).

3. MATERIAL Y MÉTODOS

Inicialmente se seleccionan las imágenes Landsat que para el periodo 2000-2004 tienen menor recubrimiento

de nubes y posteriormente se escogen las fechas más próximas al año central del quinquenio. Las seis imágenes remotas que finalmente se utilizan son (ordenadas por mes): 12 de abril de 2001, 30 de mayo de 2001, 10 de junio de 2002, 25 de julio de 2001, 21 de agosto de 2002 y 1 de septiembre de 2003. Junto a éstas se incluyeron las variables: Modelo Digital de Elevaciones (MDE), Pendientes, Radiación solar de verano, Radiación solar de invierno y los índices NDVI (*Normalized Difference Vegetation Index*), NDWI (*Normalized Difference Water Index*) y SAVI (*Soil Adjusted Vegetation Index*) para cada una de las fechas. En total se parte de 58 variables.

Para la obtención de máscaras de la cubierta nival se ha utilizado el índice NDSI, del cual se ha obtenido un valor umbral con el que discriminar la nieve (Dozier 1989, Cea *et al.* 2006). En este ámbito de estudio dicho valor umbral se ha establecido en un valor de $NDSI > 3$. Las máscaras de nubes se han realizado manualmente.

Todas las imágenes tienen un nivel de corrección LT1 definido por el USGS. El sistema de referencia empleado es UTM WGS84. La corrección radiométrica se ha hecho utilizando el módulo CorRad de MiraMon (Pons *et al.* 2014b).

Junto al NDVI (Baret & Guyot 1991, Serra & Pons 2013) se han incluido el índice SAVI (Huete 1988, Baret & Guyot 1991, Serra & Pons 2013) y el índice de humedad NDWI, que se basa en el $NDWI_3$ utilizado en la propuesta de Ouma y Tateishi (2006) y formulado como $(B5-B4)/(B5+B4)$.

3.1 Áreas de entrenamiento y método de clasificación

El SIOSE 2005 (Sistema de Información sobre Ocupación del Suelo de España) ha sido la referencia para generar áreas de entrenamiento y validación (AEV). Las trece categorías finales de la leyenda son: bosque de coníferas, bosque de frondosas perennifolias, bosque de frondosas caducifolias, matorral (o arbollado muy disperso), pastizal, suelo desnudo o roca, láminas de agua, cultivos herbáceos de secano, cultivos leñosos de secano, cultivos herbáceos de regadío, cultivos leñosos de regadío, invernaderos y áreas urbanizadas.

Mediante sentencias SQL se han seleccionado los polígonos SIOSE que forman parte del conjunto inicial de AEV. Para ello se han identificado las coberturas más puras, aquellas que tenían un recubrimiento igual o superior al 80%. A partir del 85% en el caso de matorral y cultivos, al tratarse de cubiertas muy extendidas en el área de estudio. En el caso de las frondosas, se han seleccionado recubrimientos a partir del 35%, dado que en el ámbito de estudio su distribución es menor y más fragmentada. En la selección de coberturas de coníferas, frondosas perennifolias y frondosas caducifolias se ha tenido en cuenta que no estuvieran mezcladas entre ellas, para obtener así una respuestapectral más homogénea. La geometría de los polígonos resultantes fue contraída con un buffer interior de 45 m, con el fin de evitar la mezcla espectral que puede haber en los límites entre cubiertas, o los errores de posicionamiento geográfico (Moré *et al.* 2005, Pons *et al.* 2014a). En una segunda

etapa se depuraron los polígonos de vegetación para restringir su superficie al área ocupada por el estrato de vegetación correspondiente. Para ello, se han utilizado umbrales a partir del NDVI. Tras varias pruebas iniciales, los valores umbral de NDVI fueron: $\geq 0,63$ en coníferas, $\geq 0,60$ en frondosas perennifolias, $\geq 0,70$ en frondosas caducifolias, $\geq 0,34$ y $\leq 0,62$ para el matorral y $\geq 0,25$ en pastizal. Las áreas resultantes se fragmentaron en cuadrados de 4 hectáreas con el objetivo de que grandes polígonos pertenecientes a una cubierta concreta tuvieran una parte para el proceso de clasificación y otra para el de validación. Finalmente el 70% de estos polígonos se destina a la fase de entrenamiento, y el 30% a la de validación.

Se ha empleado un clasificador híbrido, constituido por una fase inicial de clasificación no supervisada, en la que se identifican clústeres de píxeles con un patrón estadístico similar, y una fase supervisada en la que dichos clústeres son asignados a las categorías temáticas (Chuvieco 1996, Pons & Arcalis 2012). La clasificación no supervisada se ha llevado a cabo mediante el módulo IsoMM, inspirado en el clasificador clásico IsoData (Duda & Hart 1973), y la supervisada con el ClsMix, ambos integrados en el software MiraMon. Con la intención de reducir tiempos de procesamiento y de facilitar la tarea al clasificador, se han separado por una parte las categorías de cultivos y por otra la zonas naturales y urbanizadas. De esta forma, para cada uno de estos dos grupos, en la fase no supervisada el cálculo de la distancia estadística (entre cada píxel y los centroides de los clústeres) parte de un número menor de píxeles, y las agrupaciones (clústeres) de éstos se realizan en base a una menor varianza entre variables. Las máscaras se han extraído a partir del SIOSE 2005.

Los parámetros clave en la fase del IsoMM son el mínimo número de píxeles para formar un clúster y la mínima similitud entre clases (a partir de distancia euclidiana). En cuanto al primero, un valor bajo, 50 en nuestro caso, asegura una alta representatividad de las posibles agrupaciones estadísticas, teniendo así el espacio estadístico y geográfico muy compartimentado. Y un valor bajo del segundo, en nuestro caso 1, también favorece un mayor número de clases estadísticas con entidad independiente, ya que dos clases han de ser muy parecidas entre sí para terminar fusionándose. Posteriormente, el módulo ClsMix agrupa las clases estadísticas en clases temáticas a partir de las áreas de entrenamiento. Los parámetros que monitorizan esta etapa son la fidelidad y la representatividad (Serra *et al.* 2003, Pons *et al.* 2014a, Moré *et al.* 2006). Para el primero se ha establecido un valor del 40%, lo que quiere decir que para que una clase estadística pertenezca a una temática debe de tener como mínimo el 40% de sus píxeles dentro de los polígonos de entrenamiento de esa categoría. Este umbral se ha determinado empíricamente, ya que el proceso del ClsMix emplea muy poco tiempo de cálculo. Y en cuanto a la representatividad se ha optado por el valor más bajo, 0,001 %, “neutralizando” así dicho umbral, que indica la mínima proporción que una clase estadística tendría que alcanzar en relación con el resto de clases estadísticas que forman la clase

temática. Un umbral en este sentido supondría que algunos clústeres poco representados quedaran sin asignación temática.

4. RESULTADOS Y DISCUSIÓN

Los mejores resultados, en cuanto a acierto global y por categorías, se obtienen descartando la radiación solar de invierno y la banda 1. El porcentaje de superficie clasificada es del 97,5% para cubiertas naturales/urbano y 99,5% en cultivos. El acierto global en cubiertas naturales/urbano es del 88,8% y en cultivos del 86,5% (Tablas 1 y 2). En atención a los errores por categorías, destacan los bosques de frondosas perennifolias, con un error de omisión del 25% (Tabla 1) y un error de comisión del 23%. La confusión se produce principalmente con la categoría de coníferas, lo que se trata de una situación “esperable”, dado que el bosque mixto de coníferas y frondosas perennifolias está muy extendido en la escena. Los errores en la categoría de suelo desnudo o roca (en torno al 33,4% de omisión y 19,7% de comisión), se relacionan con la incapacidad del clasificador en discriminar estas cubiertas de las de cemento u hormigón en áreas urbanizadas, con respuesta espectral muy similar. También aparece involucrado el pastizal, presente en ocasiones de forma discontinua sobre el sustrato rocoso en laderas con elevada pendiente. Pero sin duda el caso que más llama la atención es el error de omisión del 66,7% en los cultivos leñosos de regadío (Tabla 2). La categoría que más afecta a este error, es la de cultivos leñosos de secano. Dicho error podría entenderse en dos sentidos. Por un lado puede existir un desfase cronológico entre la serie temporal de las imágenes de satélite y el desarrollo de la base de datos SIOSE. También puede ser debido a que cada parcela de olivar en regadío puede tener distintas fracciones sobre este régimen, y que por lo tanto existan subparcelas (no separadas a la escala SIOSE) que no necesiten de riego. Ayuda a explicar dicho error que, según la Junta de Andalucía (Consejería de Agricultura y Pesca 2002) se consideren explotaciones de regadío a “aquellas declaraciones que rieguen, bien toda la superficie del olivar, bien solamente una parte de la misma”. En la Figura 2 se muestra el mapa final de usos y cubiertas del suelo.

5. CONCLUSIONES

La clasificación de imágenes de satélite de alta resolución espacial y espectral, como es el caso de Landsat, ha permitido la generación de una cartografía para un ámbito de gran extensión y con porcentajes de acierto superiores al 86%. El tratamiento estadístico de tal volumen de datos requiere de técnicas apropiadas. El modelo no paramétrico que supone el clasificador híbrido, compuesto por el IsoMM y ClsMix, proporciona resultados en tiempos procesamiento relativamente cortos, permitiendo lograr altos niveles de acierto y flexibilidad en la toma de decisiones. Actualmente se están realizando análisis de separabilidad entre bandas con el fin de determinar qué bandas (o coeficientes entre

éstas) pueden ayudar a distinguir categorías con mayor grado de confusión. Así mismo, cabe señalar que ya se ha comenzado a desarrollar una metodología más robusta que ayuda a concretar el conjunto de variables a utilizar mediante análisis de colinealidad, cuyo objetivo final es reducir los tiempos de procesamiento y aumentar el porcentaje de acierto global y por categorías.

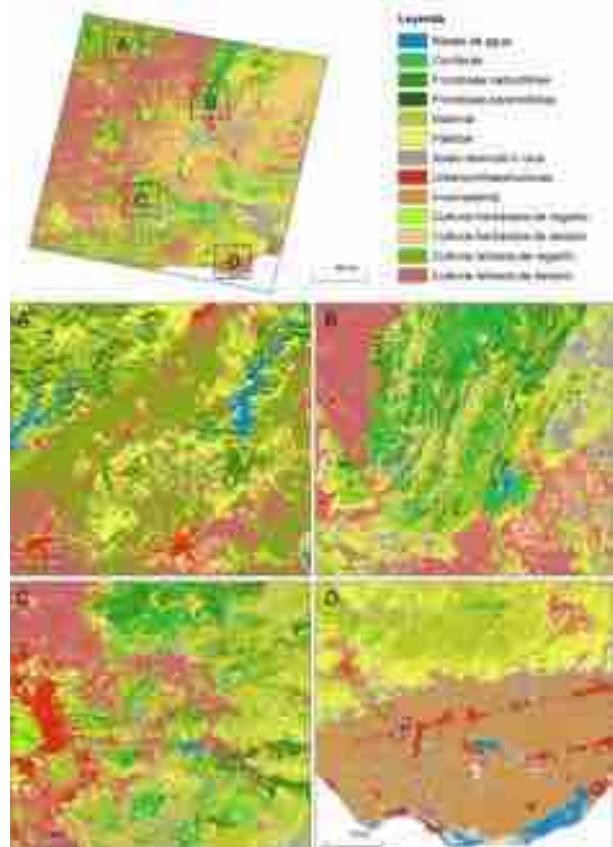


Figura 2. Mapa final de usos y cubiertas del suelo

6. BIBLIOGRAFÍA

- Baret F & Guyot G (1991). *Potentials and limits of vegetation indices for LAI and AP AR assessment*. Remote Sensing of Environment 35: 161–173
- Cea C., Cristóbal J. & Pons X. (2007). *Determinación de la superficie ni val del Pi ríneo Catalán mediante imágenes Landsat y MODIS*. In: Camacho Olmedo, M.T., J.A. Cañete Pérez, J.J. Lara Valle (Eds.) El acceso a la información espacial y las nuevas tecnologías geográficas, 65-78 (edición en CD-ROM). XII Congreso Nacional de Tecnologías de la Información Geográfica.
- Chuvieco E. (1996). *Fundamentos de Teledetección espacial*. Madrid: Rialp, 3a edición Revisada.
- Dozier J. (1989). *Spectral signature of Alpine Snow Cover from the Landsat Thematic Mapper*. Remote Sensing of Environment, 28, pp. 9-22.
- Duda, RO & PE Hart (1973) *Pattern Classification and Scene Analysis* John Wiley and Sons, New York, pp. 226-227
- Huete A.R. (1988). *A soil-adjusted vegetation index (SAVI)*. Remote Sensing of Environment 25:295-309.

- Junta de Andalucía. Empresa Pública de Desarrollo Agrario y Pesquero, Consejería de Agricultura y Pesca (2002). *El o livar a ndaluz*. Unidad de prospectiva
- McFeeters, SK (1996). *The use of the normalized difference water index (NDWI) in the delineation of open water features*. International Journal of Remote Sensing, 17: 1425-1432.
- Moré G., Pons X., Burriel J.A., Castells R., Ibàñez J.J., & Roijals X. (2005). *Diferenciación de cubiertas forestales para el MCSA a partir de la clasificación de imágenes Landsat*. Cuadernos de la Sociedad Española de Ciencias Forestales. 19: 153-162.
- Moré G., Serra P. & Pons X. (2006). *Improvements on Classification by Considering NoData Values - Application to a Hybrid Classifier to Discriminate Mediterranean Vegetation with a Detailed Legend Using Multitemporal Series of Images*. En: IGARSS 2006, IEEE Press 1, 192-195.
- Ouma YO & Tateishi R. (2006). *A water index for rapid mapping of shoreline changes of five East African Rift Valley lakes: an empirical analysis using Landsat TM and ETM+ data*. International Journal of Remote Sensing, 27: 3153-3181
- Pons X. & Arcalís A. (2012). *Diccionari terminològic de teledetecció*. Encyclopèdia Catalana i Institut Cartogràfic de Catalunya, Barcelona.
- Pons X., Sevillano E., Moré G., Serra P., Cornford D., & Ninyerola M. (2014a). *Distribución espacial de la incertidumbre en mapas de cubiertas obtenidos mediante teledetección*. Revista de Teledetección. Aceptado en agosto de 2014.
- Pons X., Pesquer, L., Cristóbal, J., González-Guerrero, O. (2014b). *Automatic and improved radiometric correction of Landsat imagery using reference values from MODIS surface reflectance images*. Int. J. Applied Earth Observation and Geoinformation, 33: 243-254
- Serra P., Pons X. & Saurí D. (2003). *Post-classification change detection with data from different sensors: some accuracy considerations*. International Journal of Remote Sensing 24, 16: 3311-3340.
- Serra P., & Pons X. (2013). *Two Mediterranean irrigation communities in front of water scarcity: A comparison using satellite image time series*. Journal of Arid Environments, 98: 41

7. ANEXO DE TABLAS

Tabla 1. Matriz de confusión de cubiertas naturales/urbano excluyendo la radiación solar de invierno y la banda 1

	Agua	Coníferas	F. caducifolias	F. perennifolias	Matorral	Pastizal	S. desnudo	Urbano	Total	Error de comisión	Exactitud del usuario
NoData	3	52	11	62	86	148	94	19	475		
Agua	27.284	0	0	0	0	7	31	10	27.332	0,2	99,8
Coníferas	0	33.690	44	2.503	233	5	1	0	36.476	7,6	92,4
F. caducifolias	0	25	3.939	164	0	16	171	0	4.315	8,7	91,3
F. perennifolias	0	2.079	104	12.463	1.172	324	46	0	16.188	23	77
Matorral	0	228	0	1.144	14.483	2.472	172	0	18.499	21,7	78,3
Pastizal	0	7	7	276	1.625	33.693	2.416	162	38.186	11,8	88,2
S. desnudo	43	0	1	8	71	821	8.714	1.188	10.846	19,7	80,3
Urbano	40	0	0	0	0	119	1.434	17.119	18.712	8,5	91,5
Total	27.370	36.081	4.106	16.620	17.670	37.605	13.079	18.498	171.029		
Error de omisión	0,3	6,6	4,1	25	18	10,4	33,4	7,5			
Exactitud del productor	99,7	93,4	95,9	75	82	89,6	66,6	92,5			
Nº total de píxeles de las áreas de test que podrían ser usados para la evaluación (A):									171.029		
Nº de píxeles de las áreas de test realmente usados para la evaluación (no sobre "nodata") (B):									170.554		
Nº de aciertos (C):									151.385		
Acuerdo global (C/A):									88,50%		
Acuerdo global (sólo los clasificados) (C/B):									88,80%		
Kappa									0,9		

Tabla 2. Matriz de confusión de cultivos excluyendo la radiación solar de invierno y la banda 1

	Herbáceos regadio	Herbáceos secano	Leñosos regadio	Leñosos secano	Invernaderos	Total	Error de comisión	Exactitud del usuario
NoData	171	479	345	368	46	1.409		
Herbáceos regadio	13.812	1.017	430	462	39	15.760	12,4	87,6
Herbáceos secano	1.681	443.030	5.994	33.114	66	483.885	8,4	91,6
Leñosos regadio	150	940	87.289	60.477	41	148.897	41,4	58,6
Leñosos secano	757	23.987	167.792	1.325.844	216	1.518.596	12,7	87,3
Invernaderos	7	83	69	164	46.867	47.190	0,7	99,3
Total	16.578	469.536	261.919	1.420.429	47.275	2.215.737		
Error de omisión	16,7	5,6	66,7	6,7	0,9			
Exactitud del productor	83,3	94,4	33,3	93,3	99,1			
Nº total de píxeles de las áreas de test que podrían ser usados para la evaluación (A):						2.215.737		
Nº de píxeles de las áreas de test realmente usados para la evaluación (no sobre "nodata") (B):						2.214.328		
Nº de aciertos (C):						1.916.842		
Acuerdo global (C/A):						86,50%		
Acuerdo global (sólo los clasificados) (C/B):						86,60%		
Kappa						0,7		