

Evaluación de productos de cubiertas del suelo en la península ibérica

A. Pérez-Hoyos y F. J. García-Haro

*Departament de Física de la Terra i Termodinàmica. Universitat de València.
c/ Dr. Moliner, 50. 46100 Burjassot*

Resumen

Existe una amplia variedad de productos de cubiertas del suelo a escala global y regional. Sin embargo éstos presentan grandes discrepancias entre ellos tanto técnicas como temáticas y no han sido validados convenientemente a escala regional. El objetivo de este trabajo consiste en evaluar cuatro de los productos más representativos y usados frecuentemente en la Península Ibérica, focalizando en España. Los productos seleccionados incluyen GlobCover, CORINE *land cover*, el MODIS *land cover* (MODISLC) y *Global land cover* (GLC2000). La metodología aplicada consiste en una validación directa utilizando como datos de referencia SIOSE (*Sistema de Información sobre Ocupación del Suelo en España*) y una validación indirecta mediante la intercomparación entre los distintos productos. En la validación indirecta se ha aplicado una metodología basada en la lógica difusa, que permite acomodar las discrepancias temáticas entre productos y que complementa la comparación Booleana o convencional. Para ello se define una métrica que integra en una única magnitud el solapamiento entre leyendas en base a 9 atributos independientes recogidos en el *Land Cover Classification System* (LCCS). Los resultados de la comparación indirecta muestran que la combinación CORINE-GLC2000 posee los mejores acuerdos (59.0%) mientras que MODISLC y GlobCover presentan importantes restricciones en la Península Ibérica (20.9% de acuerdo en la combinación MODISLC-GlobCover). La comparación basada en la lógica difusa ha permitido minimizar el efecto de los errores temáticos debido a las diferencias en la definición de las leyendas y detectar zonas que aunque *a priori* se consideren de desacuerdo están muy próximas temáticamente. Finalmente la validación directa ha demostrado que el producto que mejor se adecua a la realidad territorial de España es CORINE (69.5% de exactitud total) seguido de GLC200 (56.8%), mientras que GlobCover (38.7%) y MODISLC (36.1%) presentan grandes discrepancias.

Palabras clave: CORINE, GLC2000, GlobCover, MODISLC, SIOSE, cubiertas del suelo, lógica difusa, validación.

Abstract

Land cover assessment in the Iberian Peninsula

There is a wide range of land cover products at global and regional scale. Nevertheless, these products have large thematic and technical discrepancies among them, so an adequate validation in a regional level has not yet undertaken. The aim of this work is evaluate the most representative and frequently used land cover datasets in the Iberian Peninsula, focussing in Spain. The selected land cover datasets include GlobCover, CORINE *land cover*, MODIS *land cover* (MODISLC) and *Global land Cover* (GLC2000). The applied methodology consists in a direct validation with SIOSE (*Sistema de Información sobre Ocupación del Suelo en España*) as reference dataset and an indirect validation through the intercomparison between datasets. Indirect validation consists in a methodology based on fuzzy logic which allows accommodating thematic discrepancies among datasets and complements the Boolean or conventional comparison. The methodology was based on the definition of a metric which integrates in an only magnitude the overlap among legends using 9 independently attributes

* Autor para la correspondencia: Ana.Perez-Hoyos@uv.es

Recibido: 26-06-13; Aceptado: 03-10-13.

derived from the Land Cover Classification System (LCCS). Results of the indirect comparison show that CORINE-GLC2000 combination presents the best agreement (59.0%), whereas MODISLC and GlobCover have important restrictions in the Iberian Peninsula (20.9% agreement for MODISLC-GlobCover). Fuzzy logic comparison allows minimizing thematic errors in the comparison due to differences in legend definition and detects zones that a priori are in disagreement but really are very close thematically. Finally, direct validation has shown that the best land cover product in Spain is CORINE (69.5% overall accuracy) followed by GLC2000 (56.8%), whereas GlobCover (38.7%) and MODISLC (36.1%) have important discrepancies.

Key words: CORINE, GLC2000, GlobCover, MODISLC, SIOSE, land cover, fuzzy logic, validation.

Introducción

La cubierta del suelo y su estudio es de gran importancia puesto que juega un rol fundamental en las interacciones entre la biosfera, la atmósfera y la geosfera, por tanto interviene directamente tanto en procesos hídricos y de energía, como en el ciclo del carbono (Kalnay & Cai, 2003). Esta variable es uno de los inputs esenciales en diferentes tipos de modelos de cambios de usos del suelo (Verbug *et al.*, 2010), climáticos, de distribución de cultivos (You *et al.*, 2009) o modelos de transferencia suelo-vegetación-atmósfera entre otros. La cubierta del suelo es también el input de entrada en diferentes actividades de monitorización ambiental y el desarrollo de políticas medioambientales (Mayaux *et al.*, 2004). Por ello, constituye una variable de referencia utilizada en una gran variedad de aplicaciones a escala global, regional y nacional (Fritz *et al.*, 2012) y su estudio es de especial interés para una amplia variedad de usuarios, como el *United Nation's Millennium Ecosystem Assessment* (Millennium Ecosystem Assessment, 2005) o el *GMES (Global Monitoring for Environment and Security)* (Schreier & Dech, 2005).

En este contexto, es lógico que desde la aparición del primer producto a partir de imágenes del satélite Advanced Very High Resolution Radiometer se haya producido una proliferación de productos de cobertura del suelo, principalmente en la última década con el desarrollo de GLC2000 (Mayaux *et al.*, 2004), MODIS (Friedl *et al.*, 2002) y GlobCover (Bontemps *et al.*, 2011). Sin embargo, varios estudios previos (Giri *et al.*, 2005; Neumann *et al.*, 2007) han demostrado que los productos existentes presentan discrepancias notables entre ellos, puesto que han sido generados bajo la supervisión de

diferentes iniciativas nacionales e internacionales, con distintos objetivos y metodologías, y por tanto estos productos no son directamente comparables entre ellos (Herold *et al.*, 2008).

La aplicación de métodos apropiados para validarlos y cuantificar su fiabilidad es por tanto necesaria. El primer paso en la validación es la intercomparación entre productos, con el objetivo de determinar y cuantificar tanto sus ventajas como limitaciones. Esta comparación normalmente se lleva a cabo mediante una aproximación Booleana, es decir, cuando dos clases de dos mapas comparados no coinciden se asume un desacuerdo total y se asigna un valor de cero (Fritz & See, 2005). Sin embargo, este método no admite ningún tipo de flexibilidad y no tiene en cuenta los factores intrínsecos que realmente son los causantes de las discrepancias, como la incertidumbre temática o semántica (Cheng *et al.*, 2001). Ante estas limitaciones la lógica difusa (Zadeh, 1965) se ha aplicado como un método alternativo en la comparación entre productos de cubiertas del suelo, permitiendo introducir un grado de vaguedad entre las categorías evaluadas.

Conjuntamente a la comparación entre productos, la validación *per se* es una estrategia óptima en la determinación y selección del producto que reúne las características más idóneas para un objetivo específico. Aunque diferentes iniciativas de validación global de los principales productos de cobertura del suelo (Morissette *et al.*, 2002; Mayaux *et al.*, 2006; Friedl *et al.*, 2010) se han llevado a cabo, los resultados de estos estudios no son ni extrapolables a un ámbito regional ni comparables entre ellos. Por otra parte, el principal hándicap en la validación de productos es la ausencia de datos de validación, a este respecto España es un país pionero con el reciente desarrollo del Sistema

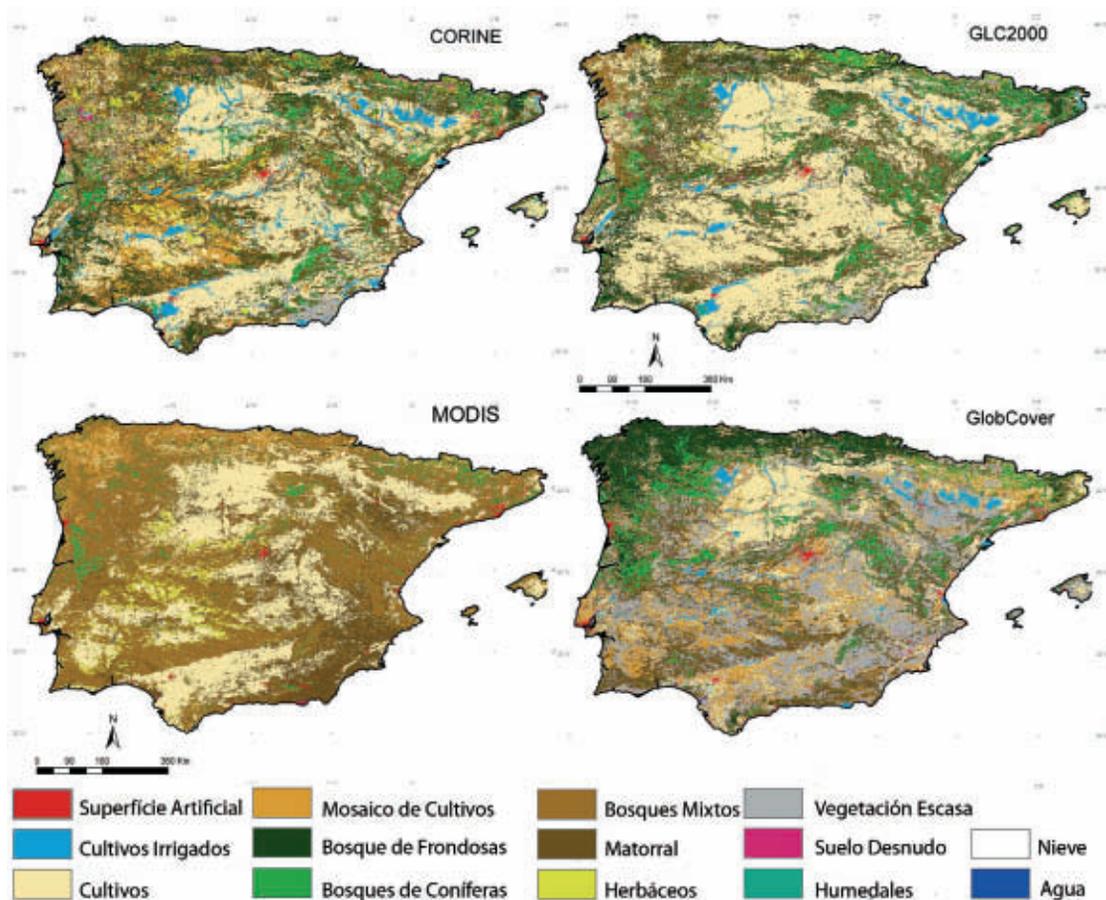


Figura 1. Mapas de cubiertas del suelo para la Península Ibérica reconciliados a una leyenda generalizada de 14 clases.

de Información sobre Ocupación del Suelo (SIOSE) que ha permitido la elaboración de información del territorio a alta resolución.

El objetivo de este trabajo consiste en explorar y evaluar el funcionamiento de los productos de cubiertas del suelo en la Península Ibérica con especial interés en España, ofreciendo una perspectiva de cuán diferentes son estos productos y qué potencial tienen para diferentes aplicaciones. Para ello se ha realizado tanto una validación indirecta en la que se han comparado los productos entre sí, como una validación directa con datos de referencia SIOSE. Los productos evaluados corresponden a cuatro de los productos de cubiertas del suelo más ampliamente utilizados a nivel global como son: CORINE, GLC2000, MODIS y GlobCover. Este trabajo amplía y ofrece sustanciales mejoras respecto a estudios previos llevados a cabo en la zona de estudio (Moreno & Chuvieco, 2009) en varios aspectos: (i) se incluye en

la evaluación un producto tan relevante como es GlobCover, (ii) ofrece un mayor análisis de las discrepancias/acuerdo entre productos al aplicar una metodología que ha demostrado ser más robusta (Pérez-Hoyos *et al.*, 2012) en otras zonas de estudio basada en la lógica difusa utilizando como marco común de comparación el *Land Cover Classification System (LCCS)* y (iii) la reciente disponibilidad de datos de alta resolución gracias a SIOSE ha permitido realizar una validación indirecta.

Datos

Productos de cubiertas del suelo

En este trabajo se han considerado cuatro de los principales productos de cubiertas del suelo a nivel global y continental (Figura 1):

Tabla 1. Características de CORINE, GLC2000, MODISLC y GlobCover.

Producto	Resolución espacial	Sensor	Fecha	Datos	Técnica de clasificación	Leyenda	Precisión
CORINE	100 m	Landsat ETM+ SPOT-4	1999-2000	Monofecha	Foto- interpretación	Jerárquica 3 niveles 44 clases	Muestreo estratificado al azar 87%
GLC2000	1 km	SPOT/VGT	Nov. 1999- Dic. 2000	Mosaicos diarios 4 bandas NDVI	Depende del socio regional No supervisada	FAO LCCS 44 clases	Muestreo estratificado al azar 69%
MODISLC	500 m	MODIS	1ª versión Nov. 2000- Dic. 2001	Composiciones bi-mensuales 7 bandas, LST y NDVI	Supervisada con árboles de decisión	IGBP 17 clases	Validación cruzada de 10 pasos 75%
GlobCover	300 m	MERIS/ Envisat	Dic. 2004- Jun. 2006	Composiciones bi-mensuales 13 bandas	Cluster No-Supervisada/ Supervisada	FAO LCCS Regional (51) Global (22)	Muestreo estratificado al azar 73%

GLC2000, MODISLC, GlobCover y CORINE. En la Tabla 1 se muestran las principales características de cada uno de los productos utilizados.

GlobCover para 2005 se obtuvo de la Agencia Espacial Europea (ESA) en cooperación con una red internacional de socios (incluyendo la EEA, FAO, GOF-C-GOLD, IGB, JRC y UNEP). Este producto presenta mayor resolución que cualquier producto global previo derivado a partir de imágenes satélite (Bicheron *et al.*, 2008). MODISLC ha sido desarrollado por la Universidad de Boston y coordinado por el MODIS *Land Team de la NASA (National Aeronautics and Space Administration)* (Friedl *et al.*, 2002). La primera versión se desarrolló en el año 2001 y es actualizada anualmente. En este estudio se ha utilizado la capa de cubierta 1 de la colección 5 (MOD12Q1 V005) para el año 2005 (Friedl *et al.*, 2010). El Global Land Cover para el año 2000 a 1 km de resolución espacial se ha implementado por el *Joint Research Centre (JRC)* de la Comisión Europea (EC) con más de 30 equipos (Bartholomé & Belward, 2005). La base de datos consiste en 19 mapas regionales armonizados en un único mapa global. El programa CORINE (*Coordination of Information on the Environment*) puesto en marcha en 1985 por la Comisión Europea y supervisado desde 1995 por la Agencia Europea del Medio Ambiente desarrolló su

primera capa de cubiertas del suelo en 1990 (CLC90). En este estudio se ha utilizado la versión del 2000 (CLC2000).

Puesto que estos productos se han llevado a cabo bajo diferentes iniciativas nacionales e internacionales y con diferentes propósitos, presentan diferencias básicas a nivel instrumental (satélite utilizado), técnico (algoritmos de clasificación) y conceptual (sistema de clasificación). A continuación se detallan las principales diferencias que pueden ser resumidas en 5 puntos:

(i) *Cubierta espacial*: GlobCover, MODISLC y GLC2000 se han diseñado de acuerdo a un marco global, mientras CORINE utiliza una aproximación regional con el objetivo de monitorizar 27 países de la Unión Europea, así como Albania, Bosnia y Herzegovina, Croacia, la ex República Yugoslava de Macedonia y Liechtenstein.

(ii) *Características del sensor y periodo de adquisición de los datos*: El desarrollo de GLC2000 se ha basado en los datos del sensor SPOT a bordo de satélite VEGETATION, que suministra datos diarios a 1 km de resolución tomados desde el 1 de noviembre de 1999 al 31 de diciembre 2000. Otras fuentes de datos, como el modelo digital de elevación, radar y datos climatológicos han sido también usados para su desarrollo. MODIS se produce anualmente utilizando la reflecti-

vidad la superficie (canales 1-7) a 500 m. GlobCover ha usado como input mosaicos bimensuales y anuales libres de nubes derivados a partir de datos MERIS (300 m) tomados en un periodo entre el 1 de diciembre de 2004 al 30 de junio de 2006. CORINE se ha elaborado a partir de imágenes de alta resolución (ej. LANDAT TM con consulta simultánea de datos auxiliares).

(iii) *Algoritmo de clasificación.* GLC2000 ha utilizado diferentes técnicas de clasificación en función de la localización geográfica, dependiendo de los requisitos y preferencias del socio que lo ha desarrollado. En el caso de Europa, se ha utilizado una clasificación no supervisada que se basa simultáneamente en patrones espectrales y temporales (Fritz *et al.*, 2003). De forma similar, GlobCover se ha beneficiado temporal y espectralmente de las composiciones bimensuales mediante un algoritmo de clúster (Arino *et al.*, 2007). MODISLC se ha producido con un conjunto de algoritmos de clasificación supervisada, basado en árboles de decisión multi-temporal conjuntamente con una técnica *boosting* que permite dar robustez a la probabilidad del píxel (Friedl *et al.*, 1999). CORINE se ha realizado mediante fotointerpretación visual en un entorno SIG (EEA, 2002).

(iv) *Sistema de clasificación o leyenda.* Tanto GlobCover como GLC2000 tienen leyendas diseñadas en el *Land Cover Classification System (LCCS)*. En el caso de GLC2000 la leyenda para el producto regional europeo está formada por 23 clases. El producto global de GlobCover está formado por 22 clases aunque en este caso hemos utilizado el producto regional que está formado por 51 clases. CORINE presenta una nomenclatura jerárquica, formada por tres niveles organizativos y teniendo 44 clases en el nivel más detallado de organización (Bossard *et al.*, 2000). MODISLC está formada por un grupo de capas con diferentes tipos de clasificaciones de cubierta. En este estudio se ha utilizado la primera capa formado por 17 clases de acuerdo al sistema de clasificación del International Geosphere Biosphere Programme (IGBP) (Loveland *et al.*, 2000).

(v) *Validación.* Los resultados de validación para los diferentes productos son: 68.6%

para GLC2000 (Mayaux *et al.*, 2006), 73% para GlobCover (Bicheron *et al.*, 2008) y 75% para MODISLC (Friedl *et al.*, 2010). La precisión de CORINE a un intervalo de confianza del 95%, es $87.0 \pm 0.7\%$ (EEA, 2006). La validación de GLC2000 se ha basado en un control de calidad con datos auxiliares y una validación cuantitativa de un muestreo estratificado aleatorio como datos de referencia. La validación de MODISLC se ha hecho con una validación cruzada con 10 particiones utilizando para ello 1860 muestras (Friedl *et al.*, 2010). Para la validación de GlobCover se ha realizado en un muestreo aleatorio con puntos distribuidos globalmente, incluyendo paisajes tanto homogéneos como heterogéneos (Bicheron *et al.*, 2008). La validación de CORINE se ha basado en un control de calidad que asegura una precisión temática $\geq 85\%$ (EEA, 2002) a partir de una validación con los datos de LUCAS (*Land Use/Cover Area frame statistical Survey*). Aunque los resultados de las validaciones se han llevado a cabo en un marco global, no son comparables entre ellos puesto que utilizan diferentes metodologías y datos de validación y además los resultados son globales, o regionales en el caso de CORINE, por tanto no se pueden extrapolar a un nivel regional como es España. Por ejemplo Cerdeira *et al.* (2006) realizaron una validación a partir de datos MODISLC en Portugal mostrando que la precisión no supera un 57%.

Datos de validación

Los datos de referencia utilizados para evaluar la precisión de las clasificaciones en España se han obtenido a partir de la base de datos SIOSE (www.siose.es). SIOSE es un proyecto enmarcado dentro del Plan Nacional de Observación del Territorio en España (PNOT) dirigido y coordinado por el Instituto Geográfico Nacional (IGN) con el objetivo de integrar las bases de datos de ocupación del suelo disponible por las Comunidades Autónomas y la Administración General del Estado y generando una base de datos de vectorial de ocupación del suelo para toda España a escala 1:25,000 con imágenes de referencia correspondientes al año 2005. SIOSE se ha llevado a

cabo con la partición de más de 25 instituciones, organizadas en grupos temáticos. Cada equipo regional produce una capa de cubiertas del suelo a partir de la fotointerpretación asistida por ordenador de imágenes SPOT5, Landsat5 TM y ortofotos.

El resultado es una base de datos orientada a objetos (OODM) interoperable y armonizada, donde la unidad de trabajo es el polígono, con una superficie lo más homogénea posible y con una unidad mínima cartografiable que varía entre 0.5 y 2 ha en función de la superficie representada. Asociado a cada polígono se han definido dos superclases denominadas de uso y cobertura. El modelo de datos SIOSE permite la asignación de uno o más usos y una cobertura a un único polígono. La cobertura de un polígono deberá ocupar siempre el 100% del mismo, sin embargo, el uso podrá diferir de este valor. Los polígonos pueden ser simples, cuando ésta sea única, y una cobertura compuesta cuando se encuentre formada por dos o más coberturas simples y/o compuestas a su vez. En función del tipo de combinación, la cobertura compuesta será una asociación (si no hay una distribución fija) o un mosaico (cuando la combinación de coberturas sigue una distribución).

La base de datos SIOSE obtenida, se ha sometido a una comprobación de campo, y dos procesos de control de calidad, uno de calidad interno llevado a cabo por los equipos regionales y uno de calidad externo realizado por el IGN, que asegura la alta calidad de los datos (Valcárcel *et al.*, 2008).

Métodos

Pre-procesado de los datos

Debido a la diferente naturaleza de los datos (resolución espacial, proyección) el primer paso para la comparación de los mimos es re-proyectarlos y co-registrarlos a un entorno común. Para ello, los productos de cubiertas del suelo han sido re-proyectados a WGS-84 y remuestreados a una resolución espacial de 1 km (0.0089286 grados) mediante un filtro de mayoría.

Validación indirecta

Comparación Convencional y Flexible

La comparación convencional o Booleana se basa en la lógica binaria, que sólo admite dos posibles opciones 0 (descuerdo) y 1 (acuerdo total). Para aplicar este método, el paso previo consiste en la armonización entre sus leyendas, que es uno de los procesos que entrañan mayor dificultad (Caetano *et al.*, 2010), puesto que como se ha indicado anteriormente los productos analizados presentan leyendas de diferente naturaleza y con un número variable de clases, por ejemplo CORINE tiene 44 clases frente a GLC2000 que presenta 23 clases. En este estudio la armonización se ha realizado creando una leyenda común de catorce clases (Tabla 2) en la que cada una de las clases de cada producto se reconcilia con cada una de las clases de la leyenda generalizada, de forma que se establece una relación de tipo uno (clase de leyenda generalizada) a muchos (clases de las leyendas de los productos). Para ello se han tenido en cuenta relaciones que otros estudios de la misma temática han establecido previamente (Giri *et al.*, 2005; Numann *et al.*, 2007). Con el objetivo de permitir cierta flexibilidad temática, típica de clases que presentan solapamientos temáticos, hemos extendido este método permitiendo cierta flexibilidad en las relaciones establecidas entre leyendas en lo que hemos denominado comparación convencional Flexible. Este método introduce un acuerdo parcial (valor de 0.5) (marcadas con negro oscuro en la Tabla 2) entre pares de clases que presentan un cierto acuerdo temático. Por ejemplo, la clase de GlobCover 110. *Mosaic forest or shrubland (50-70%)/ grassland (20-50%)* es una clase con numerosos tipos de cubierta simultáneamente que puede admitir acuerdo parcial con diferentes clases incluyendo bosque, matorral o herbáceos.

Comparación basada en la lógica difusa

En la comparación basada en la lógica Booleana se admite que hay una correspondencia

Tabla 2. Armonización de leyendas en una leyenda generalizada de 14 clases. Las clases marcadas en negrita indican flexibilidad temática. La nomenclatura de las leyendas originales se ha conservado en inglés con el objetivo de no introducir confusión en el proceso de traducción.

Leyenda generalizada	GlobCover	GLC2000	MODISIGBP	CORINE
1. Bosque de Frondosas	40. Closed to open (>15%) broadleaved evergreen or semi-deciduous forest 41. Closed (>40%) broadleaved evergreen and/or deciduous forest 50. Closed (>40%) broadleaved deciduous forest 60. Open (15-40%) broadleaved deciduous forest/ woodland Classes: 30/32/100/101/110	1. Tree cover, broadleaved evergreen, closed to open (>15%) 2. Tree cover, broadleaved deciduous, closed (>40%) 3. Tree cover, broadleaved, deciduous, open (15-40%) Classes: 6/7/9	2. Evergreen Broadleaf Forests 4. Deciduous Broadleaf Forests Classes: 5/8/9	311. Broad-leaved forest Classes: 313
2. Bosque de Coníferas	70. Closed (>40%) needleleaved evergreen forest 90. Open (15-40%) needleleaved deciduous or evergreen forest 91. Open (15-40%) needleleaved deciduous forest 92. Open (15-40%) needleleaved evergreen forest Classes: 30/32/100/101/110	4. Tree cover, needleleaved evergreen, closed to open (>15%) 5. Tree cover, needleleaved deciduous, closed to open (>15%) Classes: 6/7/9	1. Evergreen Needleleaf Forests 3. Deciduous Needleleaf Forests Classes: 5/8/9	312. Coniferous forest Classes: 313
3. Bosque Mixto	100. Closed to open (>15%) mixed broadleaved and needleleaved forest 101. Closed (>40%) mixed broadleaved and needleleaved forest <i>110. Mosaic forest or shrubland (50-70%) and grassland (20-50%)</i> Classes: 30/32/40/41/50/60/70/90/91/92	6. Tree cover, mixed leafytype, closed to open (>15%) 9. Mosaic tree-cover and other natural vegetation Classes: 1/2/3/4/5	5. Mixed Forest 8. Woody Savanna 9. Savannas Classes: 1/2/3/4	313. Mixed forest Classes: 311/312
4. Matorral	130. Closed to open (>15%) broadleaved or needleleaved, evergreen or deciduous shrubland 131. Closed to open (>15%) broadleaved or needleleaved evergreen shrubland 134. Closed to open (>15%) broadleaved deciduous shrubland Classes: 30/110/152	11. Shrubcover, closed to open (>15%), evergreen (broadleaved or needleleaved) 12. Shrubcover, closed to open (>15%), deciduous (broadleaved) Classes: 15	6. Closed Shrubland 7. Open Shrubland	322. Moors and heathland 323. Sclerophyllous vegetation 324. Transitional woodland-shrub
5. Herbáceas	120. Mosaic grassland (50-70%) and forest of shrubland (20-50%). 140. Closed to open (<15%) herbaceous vegetation (grassland, savannas or linches). 141. Closed (<40%) grassland Classes: 30/110/151	13. Herbaceous cover, closed to open (<15%) Classes: 15	10. Grasslands	231. Pastures 321. Natural grasslands
6. Vegetación Escasa	150. Sparse (<15%) vegetation 151. Sparse (<15%) grassland 152. Sparse (<15%) shrubland	14. Sparse herbaceous or shrubcover (0-15%)	16. Barren or Sparsely Vegetated	333. Sparsely vegetated areas

Tabla 2. Armonización de leyendas en una leyenda generalizada de 14 clases. Las clases marcadas en negrita indican flexibilidad temática. La nomenclatura de las leyendas originales se ha conservado en inglés con el objetivo de no introducir confusión en el proceso de traducción (cont.).

Leyenda generalizada	GlobCover	GLC2000	MODIS IGBP	CORINE
7. Suelo desnudo	200. Bare areas 201. Consolidated bare areas (hardpans, gravels, bare rock, stones, boulders) 202. Non-consolidated bare areas (sandy desert) 203. Salt hardpans	19. Bare areas 10. Tree cover; burnt (boreal forest)	16. Barren or Sparsely Vegetation	331. Beaches, dunes and sand plains 332. Bare rocks 334. Burnt areas
8. Humedales	180. Closed to open (>15%) grassland or woody vegetation on regularly flooded or waterlogged soil 185. Closed to open (>15%) grassland on regularly flooded or waterlogged soil- Fresh or brackish water	7. Tree cover; closed to open (>15%), regularly flooded, fresh or brackish; swamp forest 15. Regularly flooded (>2 month) Shrub and/ or Herbaceous cover, closed to open	11. Permanent wetlands	411. Inland marshes 412. Peat bogs 421. Salt marshes Classes: 31x Forest
9. Cultivos Irrigados	11. Post-flooding or irrigated cropland (or aquatic) 13. Post-flooding or irrigated herbaceous crops	23. Irrigated Agriculture	12. Croplands	212. Permanently irrigated land 213. Rice fields
10. Cultivos	14. Rainfed croplands 15. Rainfed herbaceous crops 16. Rainfed shrub or tree crops (cash crops, vineyards, olive tree, orchards...)	16. Cultivated and managed areas	12. Croplands	211. Non irrigated arable land 221. Vineyards 222. Fruit trees and berry plantation 223. Olive groves 241. Annual crops associated with permanent crops 242. Complex agricultural pattern
11. Mosaico de Cultivos	20. Mosaic cropland (50-70%) vegetation (20-50%) 21. Mosaic cropland (50-70%) grassland or shrubland (20-50%) 22. Mosaic cropland (50-70%) forest (20-50%) 30. Mosaic vegetation (50-70%) cropland (20-50%) 32. Mosaic forest (50-70%) cropland (20-50%)	17. Mosaic of cropland/tree cover/herbaceous cover 18. Mosaic of cropland/ shrub cover/ herbaceous cover	14. Cropland/ Natural vegetation Mosaic	243. Land principally occupied by agriculture, with significant areas of natural vegetation 244. Agro-forestry areas 231. Pastures
12. Agua	210. Water bodies	20. Water bodies	17. Water bodies	5xx. Water bodies 423. <i>Inertial flats</i>
13. Nieve	220. Permanent snow and ice	21. Snow and ice	15. Snow and ice	335. Glaciers and perpetual snow
14. Superficies Artificiales	190. Artificial surfaces and associated areas	22. Artificial surfaces and associated areas	13. Urban and built-up Lands	1xx. Artificial surfaces 422. Salines

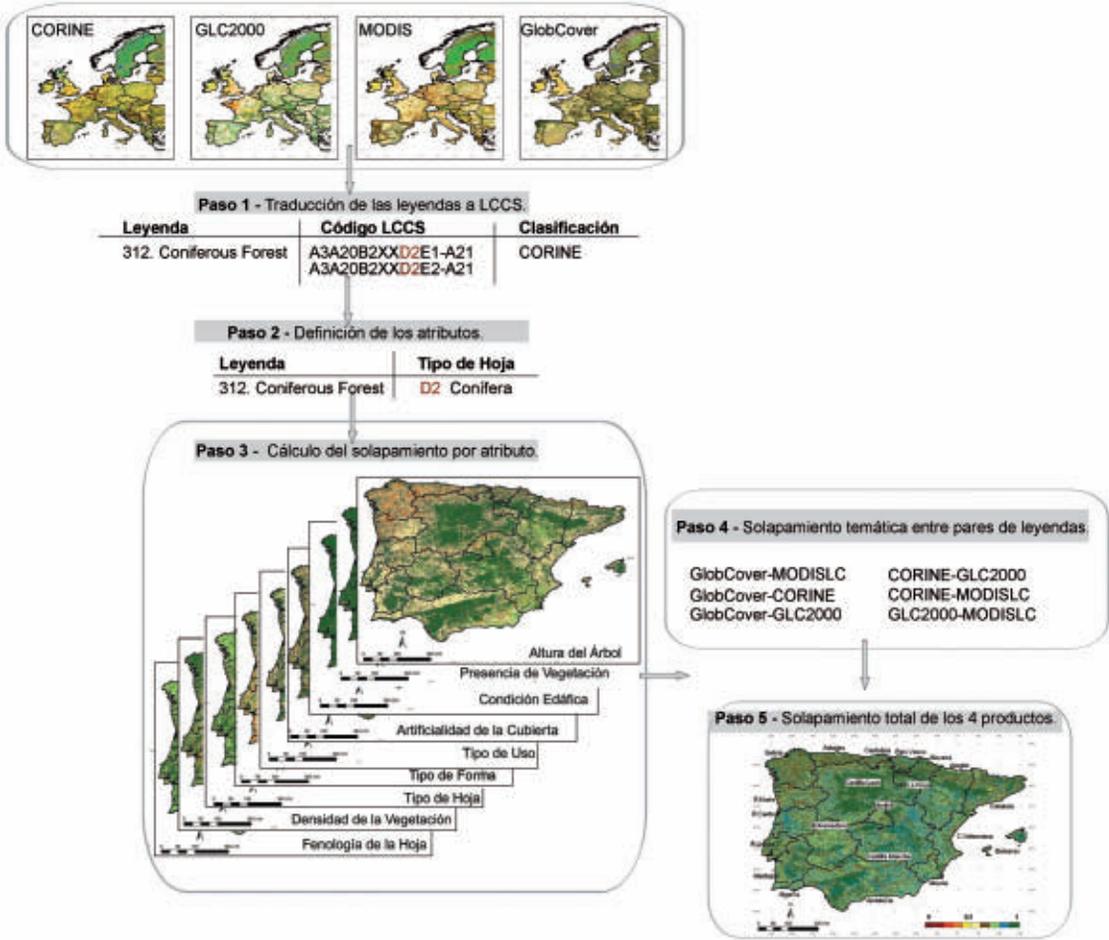


Figura 2. Diagrama de flujo del método de lógica difusa.

directa entre clases, sin embargo se ha comprobado que las leyendas presentan diferencias irreconciliables en su definición. Por tanto, en ocasiones se está hablando del mismo tipo de cobertura según diferentes productos y se presume que hay una relación directa, cuando en realidad conceptualmente son clases distintas y están definidas de diferente forma. Por ejemplo, el significado de un bosque de coníferas en GLC2000 donde la altura del árbol tiene que ser superior a 3 m es diferente de un bosque de coníferas en MODISLC, donde la altura es superior a 2 m (Fritz y See, 2008) o de uno de GlobCover donde la altura debe ser superior a 5 m. Por tanto, en este apartado se propone aplicar una metodología alternativa a la convencional basada en la teoría de la lógica difusa, puesto que este formalismo permite acomodar adecuadamente las discrepancias temáticas existentes. Se define para ello una

métrica que integra en una única magnitud el solapamiento entre leyendas en base a 9 atributos independientes recogidos en el LCCS. La Figura 2 muestra el diagrama de flujo de esta metodología. El primer paso consiste en traducir las leyendas a un sistema común como es el LCCS. Como ya se ha mencionado anteriormente GLC2000 y GlobCover utilizan este sistema de clasificación, por lo tanto existe suficiente documentación al respecto. La traducción de CORINE entraña una mayor dificultad puesto que las clases están definidas como una mezcla entre cubierta y usos del suelo (Herold *et al.*, 2008). La traducción en este caso, al igual que para MODISLC, se ha realizado utilizando documentación previa como referencia (Herold *et al.*, 2009) y con la ayuda del software LCCS distribuido por la FAO.

A partir de las leyendas traducidas se han derivado una serie de 8 atributos (descriptores

de *LCCS*) y se ha incorporado un atributo adicional de Tipos de Usos (Jansen y DiGregorio, 2002). En algunas ocasiones, se ha tenido que añadir algún código adicional con el objetivo de codificar todos los posibles casos. Por ejemplo, en el atributo Presencia de Vegetación, que según *LCCS* sólo admite dos posibles clases (Vegetación y No Vegetación), se ha incorporado Vegetación/No Vegetación para codificar aquellos casos que se produce simultáneamente la presencia/ausencia de vegetación como sucede en la clase *16. Barren or Sparsely* de la leyenda IGBP donde coexisten dos clases conceptualmente tan distintas. A continuación se detallan todos los atributos utilizados en este estudio así como sus posibles opciones de codificación:

1. *Presencia de Vegetación*: Vegetación, No Vegetación, Vegetación/No Vegetación.
2. *Condición Edáfica*: Terrestre, Acuático o inundado regularmente.
3. *Artificialidad de la Cubierta*: Natural o Seminatural y Áreas gestionadas.
4. *Tipos de Usos*: Agrícola, Forestal, Urbano, Mosaico, sin uso.
5. *Tipo de Forma*: Artificial, Herbáceo, Árbol, Matorral, Herbáceo/Matorral, Herbáceo/Árbol, Árbol/Matorral, Herbáceo/Árbol/Matorral, Artificial/Herbáceo.
6. *Tipo de Hoja*: Coníferas, Frondosas, Mixta, Sin vegetación, Sin especificación de la vegetación.
7. *Densidad de la Cubierta*: Cerrada a abierta (40-100%), Cerrada a abierta (30-100%), Cerrada a abierta (15-100%), Cerrada (60-100%), Abierta general (10-70%), Abierta (40-70%), Abierta (30-70%), Muy abierta (10-40%), Muy abierta (10-50%), Escasa (1-20%), Vegetación sin especificar (5-95%), Sin vegetación (0-4%).
8. *Fenología de la Hoja*: Perenne, Caducifolio, Sin descripción para áreas con vegetación, Sin descripción para artificial/agua/suelo desnudo.
9. *Altura del Árbol*: Suelo desnudo (0.02 m), Vegetación sin especificar (0.03-5 m), Mosaico para GlobCover (0.03-30 m), Árbol (3-30 m), Árbol GlobCover (5-30 m), Árbol MODISLC (2-30 m), Matorral (0.3-5 m), Matorral CORINE (2-7 m), Matorral MODISLC (0.3-2 m), Matorral GlobCover

(0.3-2 m), Herbáceos (0.03-1.5 m), Herbáceos CORINE (0.03-1.5 m).

Posteriormente, se ha establecido el solapamiento temático entre pares de leyendas (ej. GLC2000-CORINE), este índice tendrá unos valores que oscilarán entre 0 (clases muy diferentes) y 1 (clases conceptualmente iguales). Para ello primero se ha establecido el solapamiento por atributo de acuerdo a diferentes funciones de pertenencia de lógica difusa (Bouchon-Meunier *et al.*, 1996; Ahlqvist, 2005). El resultado se ha combinado en una única medida a partir de la media de los nueve atributos considerados. Finalmente, el solapamiento total de los cuatro productos considerados se ha establecido de acuerdo a la media entre las seis combinaciones de leyendas (ej. CORINE-GlobCover).

Validación directa

El proceso de validación directa se ha desarrollado a partir de la comparación de los productos de cubierta del suelo con los datos de referencia SIOSE. Para disponer de los datos de forma que pudieran ser utilizados en la validación ha sido necesario realizar una serie de operaciones previas en SIOSE. Primero se ha re-proyectado a WGS-84 para que fuera compatible con los productos analizados y se ha rasterizado con una resolución espacial de 50 m. Posteriormente a cada píxel que lleva asociada una leyenda SIOSE, en este caso nos hemos centrado en las cubiertas, se le ha asignado una de las catorce clases de la leyenda generalizada. Por ejemplo en el caso de tener un píxel etiquetado como I[70CNF.30MTR], lo que significa que es un polígono irregular formado con un 70% de coníferas «CNF» y un 30% de matorral «MTR», se le han asociado estos dos tipos de cubiertas simultáneamente preservado su porcentaje de ocupación. Es decir, el resultado son catorce capas correspondientes a cada una de las clases de la leyenda generalizada con el porcentaje de ocupación. Es necesario hacer un inciso en este punto puesto que SIOSE no tienen ni bosques mixtos ni vegetación escasa en su leyenda, clases que son de gran importancia en el área de estudio. Por lo que, dado que se dispone de las

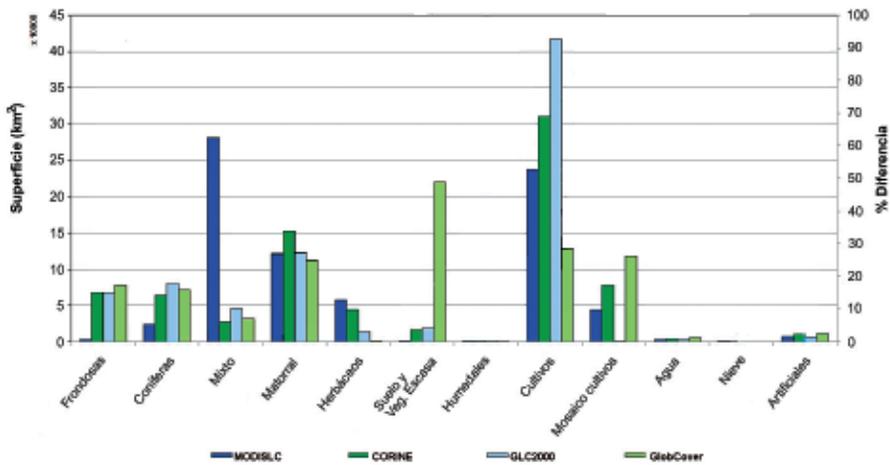


Figura 3. Porcentaje total del área para GLC2000, CORINE, MODISLC y GlobCover.

fracciones de cada clase, estas clases se han incluido a partir de la definición de bosque Mixto de CORINE (Bossard *et al.*, 2000), es decir cuando se produce simultáneamente la ocupación de bosques de coníferas y frondosas en el píxel y no se excede el 75% de cada una de las clases. Para garantizar que fuera un píxel de bosque mixto hemos ampliado esta definición considerando que la ocupación de cada uno de los tipos de bosque fuera superior al 30%. Se ha procedido de manera similar en el caso de vegetación escasa tomando para ello la definición del IGN (2002), es decir una mezcla de suelo desnudo (50-90%) y cubierta vegetal (10-50%). Por último, y para que se geo-localizara con los datos a validar se ha re-muestreado a 1 km.

Una vez se dispone de la base de datos adecuada para validar los productos globales se ha hecho un muestreo aleatorio estratificado en el que se han seleccionado alrededor de unas 90,000 muestras. El número de muestras de cada una de las clases de la leyenda generalizada se ha seleccionado de forma que fuera proporcional a la abundancia de esta clase en un promedio de los cuatro productos analizados, para ello se ha establecido un umbral mínimo de ocupación o de pureza en el píxel (valores comprendidos entre 75 y 95%) que ha ido variando en cada clase en función del número de muestras que se quisiera conseguir para cumplir este criterio.

La validación en sí se ha realizado mediante el estudio de las matrices de contingencia,

confusión o error (Stehman & Czaplewski, 1998; Foody, 2002), que son tablas de doble entrada en la que se establece la correspondencia entre las clases del producto analizado y los datos de referencia. A partir de la matriz se han derivado una serie de parámetros analíticos que nos informan sobre la precisión de la clasificación a nivel global, mediante la exactitud global (*OA-Overall accuracy*), así como diferentes medidas que nos indican la fiabilidad por clases como la exactitud de productor (*PA-Producer's accuracy*) relacionado con los errores de omisión y la exactitud del usuario (*UA-User's accuracy*) relacionada con los errores de comisión (Story & Congalton, 1986; Congalton, 1991). Además se ha calculado el índice *Kappa*, que mide la diferencia entre el acuerdo del mapa-realidad observado y el que cabría esperar aleatoriamente (Cohen *et al.*, 1960).

RESULTADOS

Comparación de superficies

La Figura 3 muestra el porcentaje total de superficie para los cuatro productos y agrupados en doce clases. La clase cultivos es la cubierta dominante para esta zona de estudio. CORINE y GLC2000 presentan porcentajes similares de esta clase, con una extensión total (incluyendo la clase mosaico de cultivos) de

44.3% (55.0%) y 46.5% (51.7%) en Portugal (España). Un porcentaje similar se ha calculado por la Oficina Estadística de la Comisión Europea (Eurostat, 2008) para el año 2007, con una superficie total de 41.2% (50.8%) en Portugal (España). En contraste, el porcentaje de las clases de cultivo para MODISLC y GlobCover son solamente de 36.6% y 31.6%.

En lo que concierne a las clases forestales, MODISLC subestima los bosques de frondosas y coníferas y sobreestima los bosques mixtos. Este resultado demuestra y confirma la existencia significativa de errores de omisión en la clase MODISLC *2.Evergreen broadleaf*, tal y como indican investigaciones previas (McCallum *et al.*, 2006) a escala global. Las discrepancias se pueden explicar parcialmente debido a la diferencia que existe en la definición de los bosques en IGBP y a la ausencia de una correspondencia directa de las clases sabanas de MODISLC. Basándonos en estudios previos (Herold *et al.*, 2008), en este estudio la clase MODISLC *8. Woody savannas* se ha relacionado con los bosques mixtos, resultando en una sobreestimación de esta clase. Aunque otros autores los relacionen con bosques de frondosas (Giri *et al.*, 2005) hemos seguido las indicaciones de Herold *et al.* (2008) puesto que en la zona de estudio hemos detectado que las zonas cartografiadas como *8. Woody savannas* ocupan principalmente la parte este de la Península Ibérica, presentando alta diversidad de bosques mixtos con muy baja presencia de bosques de frondosas (MFE50, 2001).

La extensión de la clase GlobCover *150. Sparse Vegetated* (34.5%) está sobredimensionada y es poco representativa de la realidad. En una comparación visual con una cartografía detallada (MARM, 2009), se ha detectado que esta clase presenta importantes errores de comisión principalmente en las zonas de cultivos de secano. Resultados similares se encuentran en Navarra y Murcia (García-Gigorro *et al.*, 2007) utilizando MERIS para la clasificación de la cubierta y en la validación de GlobCover en Portugal (Caetano *et al.*, 2010). Por tanto, podemos concluir que Globcover ofrece una representación poco adecuada de la cubierta en aquellas zonas ocupadas por cultivos susceptibles al estrés hídrico en zonas áridas con valores bajos de NDVI, pró-

ximos a aquellos que poseen la vegetación escasa. Esto se debe a la ausencia de la banda del infrarrojo de onda corta (*SWIR-short-wave infrared*) en el sensor MERIS, que es especialmente sensible a la vegetación, la humedad del suelo y el contenido de agua en las hojas y que mejora la discriminación de la vegetación (Clevers *et al.*, 2007).

Validación indirecta

Método Convencional y Flexible

En la Figura 4 se muestra el acuerdo entre los cuatro productos de cubiertas del suelo reclassificados a la leyenda generalizada. El acuerdo total representa un 12.7% de total de la superficie, localizándose principalmente en zonas de cultivos extensivos de secano ubicados en la meseta central y en los cultivos de regadío del valle del Ebro (Aragón y Cataluña). Estas superficies se caracterizan por presentar un paisaje homogéneo lo que permite una mejor discriminación espectral (Wu *et al.*, 2008). Zonas de alto acuerdo (26.9%) se localizan adyacentes a las áreas de acuerdo total. El método flexible Booleano permite delinear zonas de transición de muy alto acuerdo (4.2%), por ejemplo la región Centro de Portugal donde se localizan diferentes tipos de bosque, clases que temáticamente están muy próximas (ej. Bosque mixto en MODISLC y coníferas en el resto de productos).

Las zonas de desacuerdo, 15.9% para el método convencional y 9.28% para el flexible, se localizan principalmente en la zona occidental de la Península Ibérica. Estas zonas se asocian principalmente con un paisaje heterogéneo y fragmentado compuesto con una mezcla de zonas herbáceas, matorral, cultivo y bosque. Por ejemplo, el noroeste de la península, se clasifica principalmente como cultivo (GLC2000), bosques de frondosas caducifolias (GlobCover), matorral de transición (CORINE) y sabana leñosa (MODISLC).

El acuerdo total en una comparación entre pares de productos (Tabla 3), revela una buena coincidencia (59.0%) entre CORINE y GLC2000. El acuerdo disminuye drásticamente



Figura 4. Acuerdo espacial entre los diferentes productos de cobertura del suelo para una leyenda general. Izquierda (Comparación Booleana): Acuerdo total (4 de los productos), acuerdo alto (3 de los productos), acuerdo moderado (2 de los productos). Derecha (Comparación Booleana Flexible), la leyenda se extiende con valores intermedios: muy alto acuerdo (tres y medio), acuerdo sustancial (dos y medio) y acuerdo bajo (uno y medio).

Tabla 3. Acuerdo entre productos de acuerdo a una validación indirecta.

Combinación de productos	Método de lógica difusa	Método Boleano	Método Flexible
GlobCover-GLC2000	79.6	29.1	36.9
GlobCover-MODISLC	78.4	20.5	30.3
GlobCover-CORINE	77.5	29.9	36.9
GLC2000-MODISLC	78.7	36.1	43.9
GLC2000-CORINE	83.1	59.0	63.6
MODISLC-CORINE	77.0	32.6	40.1

entre otras combinaciones, principalmente entre la combinación GlobCover y MODISLC (20.5%). Las discrepancias en MODISLC se deben parcialmente a una muestra insuficiente de datos de entrenamiento del ecosistema Mediterráneo durante la clasificación (Morissette *et al.*, 2002). Aunque a nivel global la combinación GLC2000-GlobCover presenta un mayor acuerdo (55.7%), no es de esperar que estos resultados se obtengan en paisajes fuertemente antropogénicos y heterogéneos como la Península Ibérica, puesto que este tipo de paisaje complejo y de alta variación topográfica es propenso a presentar errores por falta de registro geográfico (Foody *et al.*, 2002).

Método basado en la lógica difusa

La Figura 5, muestra un mapa de solapamiento entre los cuatro productos analizados a

partir de la lógica difusa y de los atributos derivados del LCCS. Esta metodología nos permite obtener un mapa, con una gradación que oscila entre 0 (grado de acuerdo nulo) y 1 (acuerdo total), lo que nos permite analizar en mayor profundidad las discrepancias/acuerdos entre productos, así como sus causas. Asimismo permite detectar áreas que aunque en una comparación directa (método Boleano) no coincidan, son similares independientemente de la leyenda asociada. Así pues, en una comparación Booleana el resultado mostraría que dos productos presentarían una zona de desacuerdo en una zona que uno de ellos la clasifique como coníferas y otro como frondosas, sin embargo al aplicar esta metodología nos permite reconciliar estas diferencias minimizando el desacuerdo en esta zona, que presentaría un alto valor de similitud. Por ejemplo,

En la Figura 5 se puede observar que en la zona del noreste de la Península (Galicia y As-

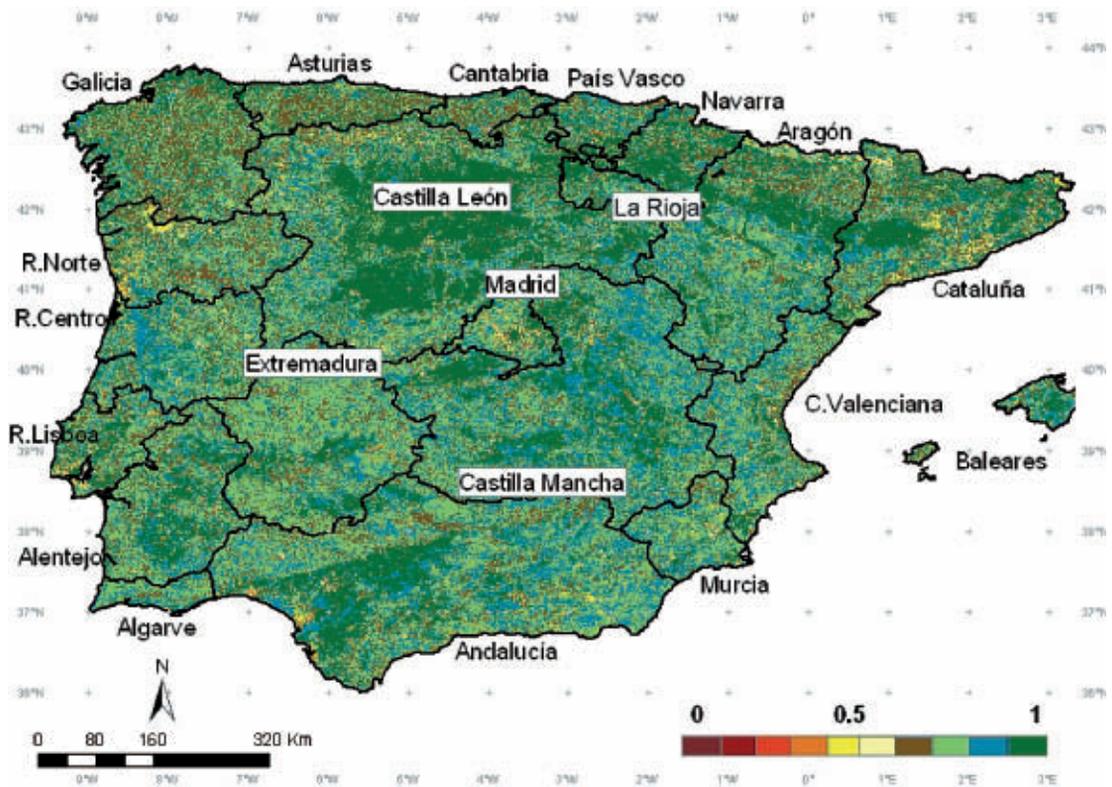


Figura 5. Mapa de solapamiento basado en la lógica difusa, representando una medida de acuerdo entre los cuatro productos.

turias) hay una zona de valores relativamente elevados (alrededor de ocho) en contraste con la Figura 4 que presenta zonas de desacuerdo. Esta zona corresponde a bosques mixtos según GLC2000 y CORINE y bosques de frondosas según GlobCover, la nueva metodología puede por tanto detectar zonas donde la discrepancia es relativamente mínima.

Las zonas de desacuerdo (zonas en amarillo) se observan principalmente en la Región Norte, Madrid, Comunidad Valenciana y Catalunya. Los principales conflictos se localizan en las zonas colindantes con las áreas metropolitanas de Madrid, Barcelona, Lisboa, Oporto y Valencia debido al cambio del paisaje producido por el crecimiento urbano. De hecho la densidad y la forma de los núcleos urbanos a escala global, como ocurre con MODISLC, son normalmente difíciles de distinguir (Friedl *et al.*, 2010). GlobCover es el producto que representa más fidedignamente el crecimiento urbano ocurrido en la última década.

Resultados parciales de la comparación basada en LCCS muestran que las clases más ge-

néricas que utilizan menos atributos en su definición (ej. Agua) están más próximas y poseen índices de solapamiento más altos (alrededor de 1) puesto que dos categorías son más similares cuanto menor es el número de vinculaciones entre ellas (Feng & Flewelling, 2004). Por tanto, los atributos que pertenecen a un nivel superior de la jerarquía de LCCS (fase dicotómica: Presencia Vegetación, Condición Edáfica y Use del Suelo) son los que originan menores discrepancias entre productos. Mientras que el resto, presentan mayores diferencias. Un ejemplo de estas discrepancias lo encontramos en las zonas forestales, que están descritas por diferentes tipos de Densidad de Vegetación incrementando la diferencia entre los distintos productos. CORINE tiene una densidad superior a 30%, mientras que MODISLC >60%, GLC2000 >15% y GlobCover define diferentes densidades según el tipo de bosque: abierto (15-40%), cerrado a abierto (>15%) y cerrado (>40%).

CORINE presenta menor ambigüedad y mayor detalle temático en la caracterización de

Tabla 4. Efecto de la resolución temática en la precisión total.

Producto	6 Clases	10 Clases	12 Clases	14 Clases
CORINE	88.57	72.88	70.68	69.55
GLC2000	80.06	61.58	57.89	56.85
MODISLC	78.32	46.13	39.66	36.15
GlobCover	77.99	50.45	47.02	38.76

Tabla 5. Efecto de la fracción sub-píxel en la precisión total.

Producto	100%	90%	80%	70%	60%
CORINE	90.2	85.82	80.66	75.32	70.59
GLC2000	88.79	82.53	76.65	70.49	64.96
MODISLC	71.71	60.76	54.64	49.02	44.37
GlobCover	73.84	64.53	59.02	53.68	48.68

las clases agrícolas. Sin embargo el principal hándicap de CORINE, comparado con los otros productos, es la ausencia de un descriptor de la fenología de la hoja en las clases de bosque. Es decir, CORINE no distingue entre hoja ca-duca y perenne, por tanto se puede concluir que este producto en estudios forestales ofrece un inventario menos detallado. GlobCover posee una leyenda muy flexible en la definición de sus clases, admitiendo una gran variedad de cubiertas simultáneamente (ej. Clase 110. *Mosaic forest of shrubland (50-70%)/grassland (20-50%)*) que aunque introduce confusión ofrece una caracterización más realista del territorio.

En la Tabla 3 se muestra el acuerdo basado en la lógica difusa, representado por la media de los datos de la imagen. Se puede observar que comparados con el método Booleano, la lógica difusa incrementa significativamente los valores de acuerdo minimizando las discrepancias temáticas, aunque el patrón es similar al que ofrece el método Booleano, mostrando que el mejor acuerdo se produce en la combinación CORINE-GLC2000.

Validación directa

La validación directa de los productos en España resulta en una exactitud global de 69.5% (kappa 0.7) en CORINE, 56.8% (kappa 0.59) en GLC2000, 38.7% (kappa 0.46) en GlobCover y 36.1% (kappa 0.5) en MODISLC. Estos

valores son menores que los resultados a nivel global. Por ejemplo, GlobCover para el producto global presenta una precisión total de 73% (Bicheron *et al.*, 2008). Sin embargo, y tal y como se ha dicho anteriormente en lugares como España, donde hay una alta variabilidad topográfica se incrementa el error geográfico, incurriendo en errores en la clasificación.

Dado las óptimas características de la base de datos de referencia, se ha evaluado el efecto de varios factores en la precisión global resultante (Tabla 4). Primero se ha evaluado como influye la resolución temática en la exactitud global. Para ello hemos evaluado la precisión considerando diferentes grados de agregación de la leyenda, seis, diez, doce y catorce clases (leyenda generalizada), resultando que la exactitud se incrementa cuanto mayor es el grado de agregación de la leyenda puesto que agregar las clases minimiza las diferencias debido a la definición de las clases y la ambigüedad temática, mejorando la exactitud (McCallum *et al.*, 2006). Sin embargo hay que tener en cuenta que la disminución de clases incurre en una pérdida de información.

Generalmente, en el proceso de validación se asume que el píxel es homogéneo y presenta un único tipo de cubierta. Dado que en este estudio disponemos de la fracción de ocupación en cada píxel también se ha evaluado el efecto que tiene la heterogeneidad sub-píxel (Tabla 5). Para ello, se ha llevado a cabo la validación considerando cada vez un umbral di-

Tabla 6. Medidas de exactitud por clase de los productos validados con SIOSE. **Nota:** EU (Exactitud del Usuario), EO (Errores de Omisión), EP (Exactitud de Productor) y EC (Errores de Comisión).

Clases	CORINE				GLC2000				GLOBCOVER				MODISLC			
	EP	EO	EU	EC	EP	EO	EU	EC	EP	EO	EU	EC	EU	EO	UA	EC
Superficies Artificiales	60.7	39.3	88.3	11.7	67.7	32.3	97.1	2.9	87.2	12.8	86.3	13.7	48.4	51.6	96.1	3.9
Cultivos Irrigados	74.9	25.1	86.2	13.8	64.5	35.5	93.0	7.0	59.2	41.0	0	100	0	100	97.8	2.2
Cultivos	97.0	3.0	74.1	25.9	96.4	3.6	63.6	36.4	72.3	27.8	74.6	25.4	79.3	20.7	65	35
Mosaico de Cultivos	0	100	0	100	0	100	0	100	0	100	0	100	0	100	0	100
Bosque de Frondosas	69.5	30.5	60	40.0	63.5	36.5	54.5	45.6	53.6	46.4	11.2	88.8	30.6	69.4	47.9	52.1
Bosque de Coníferas	75.5	24.5	60.7	39.3	74.2	25.8	51.3	48.7	68.8	31.2	54.1	46.0	34.0	66	44.2	55.8
Bosque Mixto	20.5	79.5	44.6	55.4	17.7	82.3	24.3	75.7	17.1	82.9	20.1	79.9	38.2	61.8	12.4	87.6
Matorral	65.1	34.9	43.0	57.0	46.5	53.5	36.4	63.6	25.6	74.4	5.6	94.4	8.0	92	28.9	71.1
Herbáceas	39.3	60.8	69.6	30.7	10.2	89.8	58.6	41.4	0.4	99.7	65.1	34.9	31.0	69	19.9	80.1
Vegetación Escasa	11.5	88.5	52.7	47.3	14.4	85.6	41.7	58.3	35.8	64.2	0	100	0.0	100	15.9	84.1
Suelo Desnudo	32.5	67.5	36.2	63.8	9.0	91.0	24.9	75.1	10.6	89.4	32.1	67.9	5.8	94	44.6	55.4
Humedales	53.2	46.8	80.4	19.6	77.7	22.3	60.6	36.4	0.6	99.4	30.2	69.8	2.4	97.6	80	20
Nieve	0	100	0	100	0	100	0	100	0	100	0	100	0	100	0	100
Agua	66.0	34.0	60.8	39.2	65.1	34.9	78.2	21.8	95.3	4.7	77	23	39.5	60.5	71.6	28.4

ferente de pureza que oscila desde el 50 al 100%, es decir, se ha calculado la precisión total sólo con los píxeles que superaran un umbral determinado de ocupación. En la Tabla 5 se muestra como la precisión incrementa a medida que aumenta la homogeneidad en el píxel debido a que la clasificación de píxeles puros es más sencilla (Latifovic & Olthof, 2004), puesto que en las zonas de alta fragmentación paisajística y gran heterogeneidad en el píxel es difícil captar su complejidad (Mayaux *et al.*, 2006).

En la Tabla 6 se muestran los resultados de la exactitud en un análisis por clases para la leyenda generalizada de 14 clases. En la tabla se puede observar que ciertas clases, como Mosaico de Cultivos en MODISLC, no ha sido posible evaluarlas puesto que no existen. También hay que resaltar que la clase suelo desnudo y vegetación escasa, se han validado por separado sin embargo en MODISLC conforman una única clase.

Algunas clases se caracterizan por tener alta exactitud de usuario (EU) y exactitud de productor (EP) en todos los productos lo que significa que son clases bien cartografiadas, como las superficies artificiales, cultivos, cultivos irrigados y agua. Estas clases presentan una mejor discriminación espectral y menor ambigüedad en su definición. Otras clases presentan alta EP y baja EU lo que indica una sobreestimación, como la clase de MODISLC

bosque mixto con valores de 67.3% y 9.1% respectivamente. Esta clase presenta muchos errores de comisión (90.9%) debido a que se clasifica como bosque mixto, cuando realmente corresponde a bosque de coníferas. Además, MODISLC presenta valores bajos de EP en los bosques de frondosas (9.7%) y de coníferas (33.9%) debido a la confusión de este tipo de bosques con zonas de cultivo y matorral. Tal y como se ha indicado anteriormente esto se produce en parte a la dificultad de relacionar las clases de sabana y a la diferentes definiciones de la densidad de bosques según productos.

GlobCover presenta una EP (alrededor de 0.4%) y por tanto muchos errores de omisión en la clase herbáceas, por su parte MODISLC presenta bajo EP (7%) y EU (5.6%) indicando altos valores tanto de comisión como de omisión para la clase matorral. La clase que presenta una menor EP (9%) y EU (24.9%) en GLC2000 es la clase de suelo desnudo.

Los cultivos en GlobCover están subestimados representado por una alta EU (97.8%) y una baja EP (59.2%), debido a que esta clase se confunde con vegetación escasa. La clase bosques mixtos presentan altos errores de omisión (valores entre 61.8% y 82.9%) y de comisión (valores entre 55.4% y 87.6%), esta clase se confunde normalmente con otro tipo de bosques (frondosas y coníferas) o con matorrales e incluso con herbáceas. En general, este tipo de clases presentan menos precisión

puesto que en ellas coexisten distintos tipos de cobertura con diferentes señales espectrales, por lo que es difícil delimitarlas correctamente, principalmente en productos de media o baja resolución espacial.

Conclusiones

En este trabajo se ha evaluado cuatro de los productos de cubiertas del suelo más relevantes (CORINE, GLC2000, GlobCover y MODISLC) en la Península Ibérica, haciendo especial hincapié en España. Para ello se han realizado dos validaciones: (i) una validación directa con unos datos de referencia extraídos de la base de datos SIOSE y (ii) una validación indirecta en la que se han comparado los diferentes productos entre ellos con el objetivo de detectar sus fortalezas y debilidades. A su vez, la validación indirecta se ha llevado a cabo aplicando dos metodologías, una Booleana que a su vez se ha ampliado permitiendo cierta flexibilidad temática en lo que hemos llamado comparación Flexible y otra basada en la lógica difusa a partir de unos atributos derivados del *LCCS*.

La comparación Booleana en la Península Ibérica muestra grandes discrepancias entre productos. Los mejores resultados se han obtenido en la comparación de CORINE con GLC2000 (59%), mientras que los peores se han obtenido para la combinación entre GlobCover y MODISLC (20.5%). Es decir, a nivel regional en la zona de estudio MODISLC y GlobCover presentan restricciones importantes. A grandes rasgos, se puede concluir que GlobCover sobreestima claramente la clase de vegetación escasa debido a una confusión de esta clase con zonas de cultivos, mientras que las principales discrepancias de MODISLC se deben a la dificultad en la asignación de las clases sabanas por carecer de una correspondencia directa con el resto de leyendas y con la realidad del área de estudio.

Al incrementar la flexibilidad del método (Método Flexible) se produce un ligero incremento del acuerdo, siendo más significativo el incremento en leyendas que presentan un elevado número de clases tipo mosaico, como GlobCover (ej. la combinación GlobCover-

MODISLC pasa de un 20% a un 30%). La comparación basada en la lógica difusa, ha resultado adecuada para identificar áreas donde se producen conflictos importantes entre productos, como las áreas artificiales. Además la aplicación de esta metodología ha resultado una herramienta adecuada en la localización de áreas que a priori no presentan ningún tipo de acuerdo según la comparación convencional, pero cuyas diferencias pueden resultar aceptables para una determinada aplicación. Por ejemplo, en las zonas boscosas de Galicia y Región Centro en Portugal, aunque existe una amplia variedad de tipos de cubierta, el alto grado de similitud entre clases sugiere un buen acuerdo o consenso entre productos. En esta zona, la cartografía podría considerarse, por tanto, adecuada en aplicaciones forestales.

Los resultados de la validación directa muestran que en España el producto que mejor se ajusta a la realidad del terreno es CORINE (69.5%) seguido de GLC2000 (56.8%), mientras que GlobCover (38.7%) y MODISLC (36.1%) presentan las mayores imprecisiones. Las clases agrícolas, junto con las superficies artificiales y el agua son las que se encuentran mejor caracterizadas en todas las clasificaciones, puesto que presentan una mejor discriminación espectral. Por el contrario, las clases forestales y mixtas (ej. matorral, bosque mixto) presentan los valores más bajos en exactitud del productor y del usuario.

Agradecimientos

Este trabajo se ha realizado gracias al apoyo de los proyectos LSA SAF (EUMETSAT), RESET CLIMATE (CGL2008-00381) y RESET CLIMATE (CGL2012-35831).

Referencias

- AHLQVIST, O. 2005. Using uncertain conceptual spaces to translate between land cover *categories*. *International Journal of Geographical Information Science*, Vol. 19, 831-857.
- ARINO, O., LEROY, M., RANERA, F., GROSS, D., BICHERON, P., NINO, F., BRACKMAN, C., DEFOURNY, P., VANCUTSEM, C., ACHARD, F., DURIEUX, L., BOUFG, L., LATHAM, J.,

- GREGORI, A. D., WITT, R., HEROLD, M., SAMBALE, J., PLUMMER, S., WEBER, J. L., GORYL, P., & HOUGHTON, N. 2008. GlobCover- a global land cover service with MERIS. ENVISAT SYMPOSIUM.
- BARTHOLOMÉ, E., & BELWARD, A. 2005. GLC2000: A new approach to global land cover mapping from Earth Observation data. *International Journal of Remote Sensing*, Vol. 26, 1959-1977.
- BICHERON, P., DEFOURNY, P., BROCKMANN, C., SCHOUTEN, L., VANCUTSEM, C., HUC, M., BONTEMPS, S., LEROY, M., ACHARD, F., HEROLD, M., RANERA, F., & ARINO, O. 2008. GlobCover: products description and validation report. Technical report, ESA GlobCover project.
- BONTEMPS, S., DEFOURNY, P., VAN BOGAERT, E., ARINO, O., KALOGIROU, V., & PEREZ, J.R. 2011. GLOBCOVER 2009: products description and validation report. http://due.esrin.esa.int/globcover/LandCover2009/GLOBCOVER2009_Validation_Report_2.2.pdf
- BOSSARD, M., FERANEC, J., & OTAHEL, J. 2000. Corine land cover technical guide-addendum. Technical Report 40, EEA, Copenhagen.
- BOUCHON-MEUNIER, B., RIFIQI, R., & BOUTHOREL, S. 1996. Towards general measures of comparison of objects. *Fuzzy Set System*, Vol. 84, 143-153.
- CAETANO, M., CARRÃO, H., SARMENTO, P., & NUNES, V. 2010. GlobCover Developments in Environmental Earth Observation from Space. En *Accuracy assessment of the Portuguese CORINE land cover map*, 459-467. Rotterdam: Milpress.
- CERDEIRA, C., ARAÚJO, A., CARRÃO, H., & CAETANO, M. 2006. Validação das cartografias globais de ocupação do solo, GLC2000 e MOD12Q1, para Portugal Continental. En IX Encontro de Utilizadores de Informação Geográfica.
- CHENG, T., MOLENAAR, M., & LIU, H. 2001. Formalizing fuzzy objects from uncertain classification results. *International Journal of Geographical Information Science*, Vol. 15, 27-42.
- CLEVERS, J., SCHAEPMAN, M., MÜCHER, C., DE WIT, A., ZURITA-MILLA, R., & BARTHOLOMEUS, H. 2007. Using MERIS on Envisat for land cover mapping in the Netherlands. *International Journal of Remote Sensing*, Vol. 28, 637-652.
- COHEN, J. 1960. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, Vol. 20, 37-46.
- CONGALTON, R. G. 1991. A review of assessing the accuracy of classifications of remotely sensed data. *Remote Sensing of Environment*, 37, 35-46.
- EEA. 2002. CORINE land cover update. I & CLC2000 project. Technical guidelines. Technical report, European Environmental Agency.
- EUROSTAT. 2008. Eurostat year book- Europe in Figures. Technical report, Statistical Office of the European Communities. Luxemburgo.
- FENG, C. C., & FLEWELLING, D. M. 2004. Assessment of semantic similarity between land use/land cover classification systems. *Computers, Environment and Urban Systems*, Vol. 28, 229-246.
- FOODY, G. M. 2002. Status of land cover classification accuracy assessment. *Remote Sensing of Environment*, 80, 185-201.
- FRIEDL, M. A., BRODLEY, C., & STRAHLER, A. 1999. Maximizing land cover classification accuracies produced by decision trees at continental to global scales. *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 37, 969-977.
- FRIEDL, M. A., McIVER, D. K., HODGES, J. C. F., ZHANG, D., MUCHONEY, D., STRAHLER, A. H., WOODCOCK, C. E., GOPAL, S., SCHENEIDER, A., COOPER, A., BACCINI, A., GAO, F., & SCHAAF, C. 2002. *Global land cover mapping from MODIS: algorithms and early results*. *Remote Sensing of Environment*, Vol. 83, 287-302.
- FRIEDL, M. A., SULLA-MENASHE, D., TAN, B., SCHNEIDER, A., RAMANKUTTY, N., SIBLEY, A., & HUANG, X. 2010. MODIS Collection 5 global land cover: Algorithm refinements and characterization of new datasets. *Remote Sensing of Environment*, Vol. 114, 168-182.
- FRITZ, S., BARTHOLOMÉ, E., BELWARD, A., HARTLEY, A., STIBIG, H., & EVA, H. 2003. Harmonization, mosaicking and production of the global land cover 2000 database. Technical report, Ispra, Italy: Joint Research Center.
- FRITZ, S., & SEE, L. 2008. Identifying and quantifying uncertainty and spatial disagreement in the comparison of global land cover for different applications. *Global Planetary Change*, Vol. 14, 1057-1075.
- FRITZ, S., & SEE, L. 2005. Comparison of land cover maps using fuzzy agreement. *International Journal of Geographical Information Science*, Vol. 19, 787-807.
- FRITZ, S., McCALLUM, I., SCHILL, C., PERGER, C., SEE, L., SCHEPASCHENKO, D., VAN DERVELDE, M., KRAXNER, F., & OBERSTEINER, M. 2012. Geo-Wiki: An online platform for improving global land cover. *Environmental Modelling & Software*, Vol. 31, 110-123.
- GARCÍA-GIGORRO, S., GONZÁLEZ-ALONSO, F., MERINO-DE MIGUEL, S., ROLDÁN-ZAMARRÓN, A., & CUEVAS, J. 2007. MERIS-FR potential for land use-land cover mapping in Spain. *International Journal of Remote Sensing*, Vol. 28, 1405-1412.
- GIRI, C., ZHU, Z., & REED, B. 2005. A comparative analysis of the Global Land Cover 2000 and MODIS land cover datasets. *Remote Sensing of Environment*, Vol. 94, 123-132.
- HEROLD, M., MAYAUX, P., WOODCOCK, C., BACCINI, A., & SCHMULLIUS, C. 2008. Some challenges in global land cover mapping: An

- assessment of agreement and accuracy in existing 1 km datasets. *Remote Sensing of Environment*, Vol. 112, 2538-2556.
- HEROLD, M., HUBALDN, R., & DI GREGORIO, A. D. 2009. Translating and evaluating land cover legends using the UN Land Cover Classification System (LCCS). GOCF-GOLD report series, 43, Workshop report of FAO, Rome, Italy.
- JANSEN, L., & DI GREGORIO, A. D. 2002. Parametric land cover and land-use classifications as tools for environmental change detection. *Agriculture, Ecosystems and Environment*, 91, 89-100.
- KALNAY, E., & CAI, M. 2003. Impact of urbanization and land-use change on climate. *Nature*, Vol. 423, 528-531.
- LATIFOVIC, R., & OLTFOF, I. 2004. Accuracy assessment using sub-pixel fractional error matrices of global land cover products derived from satellite data. *Remote Sensing of Environment*, 90, 153-165.
- LOVELAND, T. R., REED, B. C., BROWN, J. F., OHLEN, D. O., ZHU, Z., YANG, L., & MERCHANT, J. M. 2000. Development of a global land cover characteristics database and IGBP DISCover from 1 km AVHRR data. *International Journal of Remote Sensing*, Vol. 21, 1303-1330.
- MARM. 2013. SIG Mapa. Documentación técnica, Ministerio de Medio Ambiente Rural y Marino. Último acceso 24/06/13 <http://sig.marm.es/siga/>
- MAYAUX, P., BARTHOLOMÉ, E., FRITZ, S., & BELWARD, A. 2004. A new land-cover map of Africa for the year 2000. *Journal of Biogeography*, Vol. 31, 861-877.
- MAYAUX, P., EVA, H., GALLEGO, J., STRAHEL, A., HEROLD, M., AGRAWAL, S., NAUMOV, S., MIRANDA, E. D., BELLA, C. D., ORDYNE, C., KOPIN, Y., & ROY, P. 2006. Validation of the global land cover map. *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 44, 1728-1739.
- McCALLUM, I., OBERSTEINER, M., NILSSON, S., & SHVIDENKO, A. 2006. A spatial comparison of four satellite derived 1 km global land cover datasets. *International Journal of Applied Earth Observation and Geoinformation*, Vol. 8, 246-255.
- MFE50. 2001. Mapa Forestal de España 2001. Banco de datos de la Naturaleza. Documentación técnica, Ministerio de Medio Ambiente y Medio Rural y Marino.
- MILLENIUM ECOSYSTEM ASSESSMENT. 2005. Ecosystem and Human Well-Being: Synthesis. Island Press. Washington. Technical report, World Resources Institute.
- MORENO, M. V., & CHUVIECO, E. 2009. Validación de productos globales de cobertura del suelo en la España Peninsular. *Revista de Teledetección*, Vol. 31, 5-22.
- MORISSETTE, J. T., PRIVATTE, J. L., & JUSTICE, C. O., 2002. A framework for the validation of MODIS Land products. *Remote Sensing of Environment*, Vol. 83, 77-96.
- NEUMANN, K., HEROLD, M., HARTELY, A., & SCHMULLIUS, C. 2007. Comparative assessment of CORINE2000 and GLC2000: Spatial analysis of land cover data for Europe. *International Journal of Applied Earth Observation and Geoinformation*, Vol. 9, 425-437.
- PÉREZ-HOYOS, A., GARCÍA-HARO, J. F., & SAN-MIGUEL-AYANZ, J. 2012. Conventional and fuzzy comparisons of large scale land cover products: Application to CORINE, GLC2000, MODIS and GlobCover in Europe. *ISPRS Journal of Photogrammetry and Remote Sensing*, 74, 185-201.
- SCHREIER, G., & DECH, S. 2005. High resolution earth observation satellites and services in the next decade- a European perspective. *Acta Astronautica*, Vol. 57, 520-533.
- STEHMAN, S., & CZAPLEWSKI, R. L. 1998. Design and analysis for thematic map accuracy assessment: Fundamental principles. *Remote Sensing of Environment*, Vol. 64, 331-344.
- STORY, M., & CONGALTON, R. 1986. Accuracy assessment: a user perspective. *Photogrammetric Engineering and Remote Sensing*, Vol. 52, 397-399.
- VALCÁRCEL, N., VILLA, G., AROZARENA, A., GARCÍA-ASENSIO, L., CABALLERO, M., PORCUNA, A., DOMENECH, E., & PECES, J. 2008. SIOSE a successful test bench towards harmonization and integration of land cover/use information as environmental reference data. *In The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Science*. Beijing, vol. XXXVII, 1159-1163.
- VERBUG, P. H., NEUMANN, K., & NOL, L. 2010. Challenges in using land use and land cover data for global change studies. *Global Change Biology*, Vol. 17, 974-989.
- WU, W., SHIBASAKI, R., YANG, P., ONGARO, L., ZHOU, Q., & TANG, H. 2008. Validation and comparison of 1 km global land cover products in China. *International Journal of Remote Sensing*, Vol. 29, 3736-3785.
- YOU, L., WOOD, S., & WOOD-SICHTA, U. 2009. Generating plausible crop distribution and performance maps for Sub-Saharan Africa using a spatially disaggregated data fusion and optimization approach. *Agriculture Systems*, Vol. 99, 126-140.
- ZADEH, L. 1965. Fuzzy sets. *Information and Control*, Vol. 8, 338-353.