

METODOLOGÍA UNIFICADA PARA EL CÁLCULO DE UNA COLECCIÓN CARTOGRÁFICA MULTITEMPORAL DE OCUPACIÓN DEL SUELO EN LA DEMARCACIÓN HIDROGRÁFICA DEL SEGURA

F. Cánovas, J. Moreno, F. J. Gomariz y F. Alonso.

Instituto Universitario del Agua y del Medio Ambiente (INUAMA), Universidad de Murcia,
Campus de Espinardo, Edificio B. 30100 Murcia. fulgencio.canovas@um.es

RESUMEN

En este trabajo se describe la metodología seguida para el cálculo de una colección de cartografía de ocupación del suelo desde 1976 hasta 2006, año por año, para el ámbito de la DHS. Se ha elaborado una metodología unificada para el tratamiento de grandes volúmenes de información de una heterogeneidad considerable y para un territorio que supera los 18.000 Km². El tipo de clasificación elegida ha sido la supervisada, apoyada por una clasificación previa no supervisada. Se ha mejorado considerablemente el resultado del clasificador al añadir variables con información textural obtenidas mediante el cálculo de la función semivariograma a partir de las capas de albedo y NDVI. El escaso tiempo disponible para realizar el trabajo ha obligado al equipo a realizar una labor preliminar de diseño de una metodología a la vez eficaz y sencilla que permitiera la automatización de tareas. Los métodos han sido validados para el tratamiento de imágenes del sensor Landsat MSS, TM, ETM y ETM+. Para el tratamiento de cuatro escenas de imágenes desde 1976 a 1986 y dos escenas de 1987 a 2006, en la mayoría de los casos dos conjuntos de escenas por año. Este tratamiento de los datos ha sido posible gracias a la integración en el flujo de trabajo de programación de scripts en entorno *bash* sobre sistema operativo GNU/Linux, SIG GRASS y paquete estadístico R, ya que todos ellos funcionan totalmente integrados, permiten la entrada y salida de información y el uso compartido de variables y parámetros.

ABSTRACT

This work describes the methodology used to obtain a yearly series of land cover maps, from 1976 to 2006, of the *Demarcación Hidrográfica del Segura* (River Segura basin in South East Spain). In order to analyze this great amount of information for a large (more than 18,000 km²) and heterogeneous area, a unified methodology has been developed. It has been used a supervised classification improved by a previous not supervised classification using both spectral (4 or 6 Landsat solar spectrum bands depending on the sensor available, MSS or TM, ETM) and two textural variables. The last ones were obtained estimating the semivariogram function from the albedo and NDVI layers. Having used two images from different seasons, the total number of variables included in the classification is 12 or 16 depending on the year. The data processing has been integrated in a workflow of BASH and R scripts reading and analysing data stored in GRASS GIS. This computational approach to the problem has facilitated the ending of a project that had the constraint of a very short time to be completed.

Palabras clave: teledetección, cobertura del suelo, clasificación supervisada, información textural.

INTRODUCCIÓN

El objetivo de este trabajo es presentar una metodología unificada en un entorno GNU/Linux para la elaboración de una colección multitemporal de coberturas del suelo en el ámbito de la Demarcación Hidrográfica del Segura (DHS) desde 1976 hasta 2006.

El trabajo se enmarca en un proyecto más amplio y ambicioso que trata de analizar la evolución espacio-temporal de los usos del suelo en el ámbito de la Demarcación Hidrográfica del Segura desde 1976 hasta la actualidad con técnicas de teledetección. Forma parte de un convenio de colaboración entre el INUAMA (Universidad de

Murcia) y la Fundación Instituto Euromediterráneo del Agua (IEA).

Los primeros trabajos realizados en el marco de este convenio han sido los que trataban de conseguir una metodología unificada en entorno GNU/Linux que permitiera el procesamiento de la ingente cantidad de información espacial disponible. Piénsese que para abarcar el ámbito de la Demarcación Hidrográfica del Segura (algo menos de 19.000 Km²) son necesarias más de dos escenas de Landsat-5 y 4 escenas en el caso de los antiguos Landsat MSS. Además, la mayoría de los mapas de coberturas se han calculado con dos fechas diferentes (normalmente primavera y verano) con el objeto de mejorar la discriminación de cultivos. En

total se han procesado más de 100 escenas del satélite Landsat.

A partir del desarrollo de la metodología presentada en esta comunicación la intención del equipo de trabajo ha sido obtener una primera versión de estos mapas de cobertura de usos, 31 en total, que permitiese comprobar la adecuación de la metodología seleccionada.

METODOLOGÍA

La base metodológica del proyecto son un conjunto de técnicas de clasificación de imágenes de satélite como clasificación supervisada, clasificación no supervisada, incorporación de información textural y estacional y asignación de píxeles a clases por máxima verosimilitud.

El programa utilizado para el tratamiento digital de las imágenes ha sido GRASS debido a la experiencia del equipo de trabajo con este software y a la potencia que tiene a la hora de automatizar tareas y aplicar procedimientos de trabajo en forma de *script*. La alta integración con R permite utilizar este último programa para implementar métodos sofisticados de análisis de datos y gráficos avanzados a partir de los datos de GRASS. En concreto ha servido para aplicar métodos de análisis estadístico de la información extraída de las áreas de entrenamiento.

Georreferenciación y corrección geométrica

El primer paso en el tratamiento de las imágenes de satélite ha sido su georreferenciación, que se ha basado en la búsqueda de un conjunto de puntos de control con los que calibrar una ecuación de transformación de tipo polinómico.

Se han localizado puntos de control homogéneamente repartidos para evitar sobreestimar un área, ya que los errores se incrementarían en las áreas submuestreadas.

Finalmente se ha optado por contar con una muestra inicial de 35 puntos para escenas del sensor MSS, 50 para las del sensor TM y 120 cuando se trataba de escenas flotantes del sensor TM.

Se ha observado que el error medio cuadrático no disminuye de forma significativa al aumentar el grado de los polinomios de transformación, pero sí se incrementa el riesgo de que aparezcan distorsiones locales importantes. Por ello finalmente se ha optado por emplear transformaciones lineales.

Como criterio general, se ha tratado de obtener valores de error cuadrático entorno al tamaño del píxel.

Cálculo de reflectividad: corrección atmosférica y de iluminación

Se ha realizado la corrección atmosférica según el método del mínimo del histograma (Chavez, 1988) para todas las imágenes. De este modo la distorsión atmosférica residual que quede tras la corrección afectará, en cada imagen, del mismo modo a los píxeles de las áreas de entrenamiento utilizadas para caracterizar las clases y a los posteriormente clasificados minimizando su impacto en el error de clasificación.

Con respecto a la corrección de iluminación se ha optado por un ajuste semiempírico como el realizado en el método lambertiano de corrección C (Teillet *et al*, 1982).

Creación de capas de información textural

Se ha incorporado información textural, que representa la variación espacial del brillo de una imagen (Berberoglu *et al*, 2007). Según estos mismos autores, la utilización de este tipo de información resulta especialmente interesante en paisajes mediterráneos debido a la gran variedad y fragmentación de patrones espaciales.

Si las características espectrales del conjunto de clases seleccionado no permite garantizar la separabilidad entre estas, la información textural puede introducir información adicional que incremente la separabilidad de las clases.

En los últimos años se ha experimentado con el uso de información textural en teledetección, se trata del cálculo, píxel a píxel, de índices texturales que cuantifican los criterios texturales cualitativos tradicionales de la fotointerpretación. Para el cálculo de las capas de información textural se ha utilizado la función semivariograma como indicador de la variación espacial del brillo en torno a un píxel:

$$\gamma = \frac{\sum_{i=1}^4 (b - b_i)^2}{8}$$

donde b hace referencia al valor en la celda analizada de la banda de que se trate, b_i hace referencia al valor en las cuatro celdillas vecinas (norte, sur, este y oeste).

Este índice se ha aplicado a dos bandas obtenidas a partir de los valores de reflectividad:

- Primer componente de un ACP con las bandas de reflectividad, que aparece relacionado con un albedo medio ponderado de la imagen.
- Índice Normalizado de Vegetación (NDVI)

Clasificación supervisada

Las especiales características de este proyecto imponen la necesidad de buscar áreas de entrenamiento que hayan permanecido con el mismo uso en los últimos 30 años, para ello se han utilizado mapas de uso de suelo e imágenes de tres fechas:

- Mapa de Cultivos y Aprovechamientos de 1977 y de 2001.
- Ortofoto de 1986-87.
- Mapa del proyecto Corine-Land Cover de 1990 y 2000.
- Ortofoto de la Dirección General de Medio Natural de la CARM de 2008.

Se ha asumido que las áreas que mantenían un mismo uso en esta cartografía y ortofoto han permanecido con un uso constante durante todo el periodo de tiempo estudiado.

El conjunto de clases seleccionado a priori son un total de 550 áreas de entrenamiento que suman una superficie de 19.282 Hectáreas.

Análisis estadístico de las áreas de entrenamiento

La necesidad de cumplir las exigencias de un método de clasificación paramétrico como es la máxima verosimilitud, obliga a una comprobación exhaustiva de su cumplimiento, en particular la normalidad de las variables empleadas, en este caso las bandas corregidas atmosféricamente y la capa de semivariograma de albedo y NDVI.

En este sentido, y con carácter exploratorio, se han obtenido gráficos de normalidad para cada combinación de los valores de las bandas con las clases informacionales. Estos gráficos se han obtenido también para transformaciones de variable logarítmica e inversa.

A partir de estos gráficos resulta fácil observar la presencia de *outliers* que pueden eliminarse recalculando el mapa de áreas de entrenamiento para eliminar de estas los píxeles con dichos valores.

Una vez eliminados los *outliers* se ha determinado, a través del análisis de los gráficos, la transformación más adecuada para cada banda. En cada banda se ha buscado una transformación de compromiso entre todas las clases.

Todas estas operaciones se han realizado con el software estadístico R. Es software libre y presenta como principal ventaja para este trabajo su integración total con GRASS.

La integración de la clasificación supervisada y no supervisada se hace combinando categorías de ambos métodos empleando un algoritmo de agrupación jerárquico. Se utiliza como criterio de ordenación la divergencia entre las CE (clases espectrales) y las CI (clases informacionales). Si existe una CE con baja divergencia con todas la CI se considera una clase significativa y se añade al conjunto de clases. Un método similar ha sido probado con buenos resultados en (Chuvieco y Congalton, 1988).

El índice de divergencia se ha utilizado para evaluar la similitud o disimilitud entre clases de manera que indica la posibilidad de distinguir una de otra (Chuvieco, 2002) y (Thomas *et al*, 1987).

$$Div_{a,b} = \frac{Tr\{(V_a - V_b)(V_b^{-1} - V_a^{-1})\}}{2} + \frac{Tr\{(V_a^{-1} + V_b^{-1})(M_a - M_b)(M_a - M_b)^t\}}{2}$$

En la anterior ecuación M_a es el vector de medias de la clase a , V_a es la correspondiente matriz de varianzas-covarianzas, Tr la función traza de la matriz, t indica la matriz transpuesta y $^{-1}$ la matriz inversa.

Se trata de un test paramétrico que, por tanto, asume normalidad en las variables que caracterizan las clases y será especialmente adecuado para comprobar la separabilidad entre clases empleando el método de máxima verosimilitud.

Una vez obtenido un conjunto de clases óptimo y su caracterización estadística procede asignar cada uno de los píxeles presentes en la imagen a una de aquellas. Para esto se ha utilizado el algoritmo de máxima verosimilitud, que calcula para cada clase A la probabilidad de que un píxel x pertenezca a ella a partir de la función de densidad multivariante (Chuvieco, 2002):

$$p(x/A) = (2\pi)^{-m/2} |V_A|^{-0.5} \exp\{-0.5(X - m_A)^t V_A^{-1} (X - m_A)\}$$

donde m indica el número de bandas, V_A es la matriz de varianzas-covarianzas para la clase A ; X el vector de valores del píxel x y m_A el vector de medias de la clase A .

Validación. Error de clasificación

La tarea de validación de resultados se ha llevado a cabo analizando una serie de áreas test obtenidas del mismo modo que las áreas de entrenamiento pero que no fueron utilizadas para este. Esta forma de proceder permite obtener una estimación realista de los errores a partir del cálculo de las matrices de confusión y de los indicadores que se desprenden de ella. Como aproximación general del resultado de la clasificación se ha utilizado en Índice de Fiabilidad y el Coeficiente *Kappa*. Con este tipo de análisis se ha obtenido, no sólo una caracterización del error cometido, sino también una medida sobre la adecuación de la clasificación considerada a la realidad y de los parámetros utilizados para caracterizarlas. Puede por tanto utilizarse para definir un nuevo conjunto de clases para realizar una clasificación.

RESULTADOS

El resumen de los resultados obtenidos en la clasificación se puede observar en la tabla 1. La mayoría de las imágenes están clasificadas con un Índice de Fiabilidad y un Coeficiente *Kappa* superior a 0,61.

Tabla 1.- Distribución de las medidas de fiabilidad de los mapas de cobertura del suelo 1976-2006.

Intervalo	Í. de Fiabilidad	C. <i>Kappa</i>
< 0.5	0	7
0.51-0.60	13	8
0.61-0.70	15	16
> 0.70	3	0

A la vista de estos resultados se puede afirmar que son algo inferiores a los valores que se suelen obtener en este tipo de estudios. Concretamente, si los comparamos con un estudio realizado por Graeme G. Wilkinson en 2005 (Wilkinson, 2005) en el que se analizaban más de 500 experimentos, los valores obtenidos en nuestro trabajo son un 18% inferiores para el Índice de fiabilidad y un 11,43% para el Coeficiente *Kappa*. Finalmente se ha calculado un total de 31 mapas de coberturas del suelo.

CONCLUSIONES

La primera fase del proyecto “Evolución espacio-temporal de los usos del suelo en el ámbito de la DHS” del que este trabajo forma parte, ha sido llevada a cabo satisfactoriamente ya que se ha logrado desarrollar formalmente y escribir el código para una metodología unificada del pre-procesamiento de imágenes procedentes del satélite

Landsat. Además se ha conseguido una primera versión de los mapas de cobertura del suelo del ámbito estudiado. Los procedimientos han sido implementados en forma de *script* de manera que para un futuro inmediato se trata de ir aislando sistemáticamente las posibles fuentes de error en la clasificación. De esta forma se espera que se consiga el objetivo planteado a medio plazo: poner en producción, formando parte del Sistema de Información del INUAMA e IEA, una colección multitemporal de mapas de cobertura del suelo para la DHS, cuyas medidas de acierto/fiabilidad estén dentro de los márgenes aceptados por la comunidad científica.

BIBLIOGRAFÍA

- Berberoglu, S.; Curran, P.J.; Lloyd, C.D.; Atkinson, P.M. 2007. Texture classification of Mediterranean land cover. *International Journal of Applied Earth Observation and Geoinformation* n.9: 322-334.
- Chávez, P.S. 1988. An improved dark-object subtraction technique for atmospheric scattering correction of multispectral data. *Remote Sensing of Environment* n. 24: 459-479.
- Chuvieco, E.; Congalton, R.G. 1988. Using cluster analysis to improve the selection of training statistics in classifying remotely sensed data. *Photogrammetric engineering and remote sensing* n.54(9): 1275-1281.
- Chuvieco, E. 2002. *Teledetección Ambiental*. Barcelona, Ariel.
- Riaño, D; Salas, J; Chuvieco, E. 2000. *Corrección atmosférica y topográfica, información multitemporal y auxiliar aplicadas a la cartografía de modelos de combustibles con imágenes landsat TM. Tecnologías Geográficas para el Desarrollo Sostenible*: 222-239. Universidad de Alcalá.
- Teillet, P.M.; Guindon, B.; Goodenough, D.G. 1982. On slope-aspect correction of multispectral scanner data. *Canadian Journal of R. S.* n.58: 84-106.
- Thomas, I.L.; Ching, N.P.; Benning, V.M.; D’aguanno, J.A. 1987. A review of multi-channel indices of class separability. *International Journal of Remote Sensing* n.8(3): 331-350.
- Wilkinson, G.G. 2005. Results and implications of a study of fifteen years of satellite image classification experiments. *IEEE Transactions on Geoscience and Remote Sensing* n.43(3): 433-439.

AGRADECIMIENTOS

El desarrollo de este trabajo ha sido posible gracias al Proyecto de Investigación denominado “Evolución de los usos del suelo en la Cuenca del Segura mediante técnicas de Teledetección” financiado por la IEA.