

# Análisis geoestadístico de las áreas de entrenamiento en la clasificación digital de imágenes de satélite

M. Chica-Olmo y J. P. Ruiz  
*mchica@ugr.es*

*Universidad de Granada. Dpto. Geodinámica/IACT  
Laboratorio RSGIS "Teledetección, SIG y Geoestadística"  
Avda. Fuentenueva S/N. 18071 Granada*

## RESUMEN

Este trabajo presenta una metodología aplicada al estudio de separabilidad de clases temáticas en clasificación digital de imágenes de satélite. La metodología se fundamenta en el análisis de la función variograma, que caracteriza desde el punto de vista geoestadístico la variabilidad espacial de los valores digitales. A través del cálculo y modelación del variograma se han obtenido un conjunto de parámetros que permiten caracterizar las áreas de entrenamiento de las clases temáticas y analizar la homogeneidad espacial de éstas. Con ello se pretende alcanzar una mejora de los resultados de la clasificación digital supervisada.

**PALABRAS CLAVE:** Clasificación supervisada, análisis variográfico, áreas de entrenamiento.

## ABSTRACT

This paper presents a geostatistical approach to training site characterization. The method is based on the variographic analysis of digital values registered by the sensor in the polygons established during the training stage. Different geostatistical parameters deduced from the variogram fitting are used for this spatial characterization in order to improve results of supervised classification.

## INTRODUCCIÓN

La clasificación digital es una de las etapas más representativas del tratamiento de imágenes de satélite, desde el punto de vista de la construcción de cubiertas temáticas. En ella se da especial importancia a la fase de supervisión en la que se definen las áreas de entrenamiento representativas de las clases temáticas estudiadas. El análisis exploratorio de estas clases, en función de las características radiométricas de los píxeles que componen las áreas de entrenamiento, es la base para el estudio de su separabilidad estadística y, por tanto, de ello dependerán los resultados de la clasificación.

Los estudios clásicos de clasificación digital, muestran que durante la fase de entrenamiento no se tiene en cuenta la disposición espacial de los píxeles en el interior de los polígonos; sólo se utili-

zan sus valores de radiancia espectral a los efectos de una descripción estadística (firma espectral). Este hecho bien podría llevar a una situación absurda, en la que desordenando aleatoriamente la posición espacial de los píxeles de las áreas de entrenamiento de una determinada clase, se llegaría a obtener idéntica firma espectral (estadística) que la correspondiente a su disposición original. Resulta obvio decir que la clase correspondiente a esta situación extrema no tendría igual significado temático que la original. En consecuencia, la dependencia entre valor radiométrico y posición del píxel en el interior del polígono de entrenamiento es un aspecto interesante a estudiar. Este aspecto lleva a plantearse la cuestión de si la autocorrelación espacial, o si se desea mejor la variabilidad espacial de los ND, puede constituir un criterio de mejora para la clasificación digital (Ketting y Landgreve 1975 y Salvador y San Miguel 2002).

En el presente trabajo se utiliza el procedimiento geoestadístico del análisis variográfico de las áreas de entrenamiento, con la finalidad de caracterizar la estructura de variabilidad de las distintas clases temáticas consideradas, y con base en estos resultados se estudiará su separabilidad geoestadística. A través de esta caracterización geoestadística de las clases se pretende ayudar al usuario a reducir los errores en la clasificación digital supervisada, partiendo de la hipótesis de que cualquier mejora en dicha caracterización conllevaría una mejora en los resultados de la clasificación.

## OBTENCIÓN DE DATOS

Para la realización de este estudio experimental se ha utilizado una imagen del sensor Landsat 5 TM, en la que se ha estudiado el área de la “Vega de Granada”. La fecha de adquisición corresponde al 26-08-1995, por lo que nos encontramos en un período del año marcado por la sequedad estival. La cubierta vegetal (cultivos y bosques) no es excesivamente abundante lo que unido a una baja humedad en el suelo, provoca que los aspectos geológicos (rocas y suelos) predominen radiométricamente sobre los de la vegetación. Sin duda, esto supone un problema a la hora de la caracterización de ciertas cubiertas como olivar o secano. Para la clasificación, y por tanto para definir las áreas de entrenamiento, se seleccionaron cinco clases temáticas representativas de la Depresión de Granada. Para ello, además de la información de terreno se realizó una clasificación no supervisada tipo cluster a partir de una composición en falso color TM453 (RGB), despreciando los grupos con un porcentaje inferior al 1% del tamaño de la imagen.

Además de la información radiométrica, y para evitar la correlación existente entre las distintas bandas TM, se aplicó un ACP a las bandas espectrales. Para el estudio se utilizaron las tres primeras CP, que dan información de los distintos aspectos ambientales del área de estudio.

Por medio de este análisis se definieron las clases: conífera, olivar, regadío y secano; además, se incluyó la clase urbano, por ser un ejemplo claro de textura característica, detectable a priori mediante análisis variográfico y que, sin embargo, suele presentar problemas en la clasificación. Para cada clase temática se establecieron cinco polígonos de entrenamiento. Finalmente, se utilizaron los datos correspondientes a 9 bandas (6 TM y 3 CP) para las cinco clases temáticas anteriores reconocidas cada una en cinco áreas de entrenamiento.

## ASPECTOS METODOLÓGICOS

Desde nuestro pasado IV Congreso de la AET (Sevilla, 1991), nuestro laboratorio RSGIS de la Universidad de Granada, viene presentando de forma sistemática trabajos relacionados con la integración de los métodos geoestadísticos el estudio de datos de teledetección. En todos ellos, se parte de la premisa de que los ND registrados en una imagen de satélite son una variable regionalizada  $ND(x)$ . De esta forma, la función  $ND(x)$  relaciona el valor radiométrico observado con la posición del píxel  $x$  (Chica-Olmo y Abarca 1997).

La función que caracteriza a la variable regionalizada es el variograma, definido como una función intrínseca de orden 2 que representa la mitad del momento de segundo orden, esto es:

$$\gamma(h) = \frac{1}{2} E\{[ND(x+h) - ND(x)]^2\}$$

El variograma es una función vectorial, que cuantifica la mitad de los incrementos cuadráticos medios de los ND, para parejas de píxeles distantes el módulo  $|h|$ .

La función  $ND(x)$  presenta una distribución espacial dentro de un dominio de estudio, en nuestro caso las áreas de entrenamiento, en la que destacan dos aspectos esenciales para nuestro análisis: el *carácter aleatorio*, que representa la variabilidad imprevisible y por tanto aleatoria a distancia intrapíxel (efecto de pepita), y el *carácter estructural*, que manifiesta el grado de continuidad (correlación) de los ND, y que es inherente a cada banda espectral. Ambos aspectos, son debidamente estudiado a través del variograma.

La metodología propuesta es la siguiente (Figura 1):

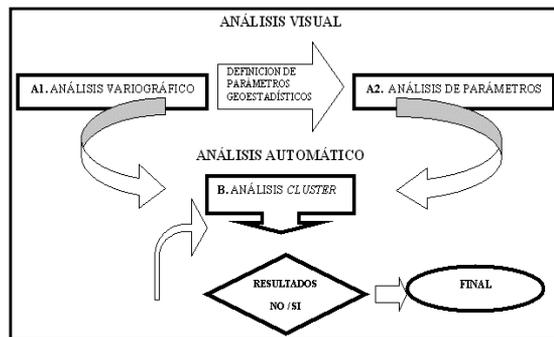


Figura 1. Diagrama de la metodología propuesta.

a) *Análisis variográfico de las áreas de entrenamiento*

Este análisis consiste en el cálculo, interpretación y modelación de los variogramas de las áreas de entrenamiento, al objeto de caracterizar las estructuras de variación de las diferentes clases temáticas. Un aspecto de gran interés reside en el estudio detallado de los parámetros geoestadísticos deducidos del ajuste de los variogramas, p.e. meseta, alcance y efecto de pepita. Así, el alcance es un indicador del nivel de correlación espacial de los valores digitales en el interior de las áreas de entrenamiento, la meseta señala la intensidad (varianza) de variación global de los valores y el efecto de pepita es la componente errática / aleatoria de variación intrapíxel. Un ejemplo puede verse en la Figura 2, en la que han representado las variogramas de la clase regadío para la componente PC3.

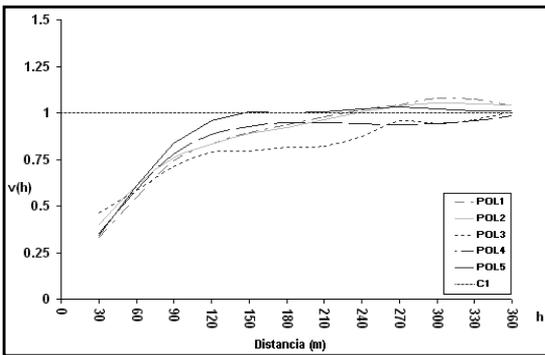


Figura 2. Variogramas de la clase regadío para la CP3.

Por medio de este análisis es fácil interpretar desde el punto de vista espacial, si algún polígono tiene un comportamiento anómalo con respecto a la tendencia general observada para su clase temática.

b) *Definición de parámetros geoestadísticos*

El análisis variográfico de las áreas de entrenamiento correspondientes a las distintas clases temáticas, permitió establecer los siguientes parámetros geoestadísticos:

- *Meseta*: valor interpretado como variabilidad máxima del variograma.
- *Pendiente*: corresponde al crecimiento del variograma en el origen. Este parámetro se interpreta como una medida de la intensidad de la variación local, por ejemplo entre píxeles distantes 30 m, 60 m, etc.

- *Efecto de pepita*: valor correspondiente a la discontinuidad del variograma en el origen.

Estos parámetros se han extraído de los variogramas de cada clase, considerando las 6 bandas TM y las 3 CP, obteniéndose un total de 27 variables para las 25 áreas de entrenamiento.

c) *Análisis de los parámetros geoestadísticos*

El análisis visual de estos parámetros mediante gráficos de dispersión, ha permitido estudiar el comportamiento anómalo de algunas áreas con respecto a la tendencia general de sus respectivas clases y, por otra parte, el nivel de separabilidad entre las clases. Hasta qué punto estos parámetros aportan información significativa para caracterizar las áreas de entrenamiento, es una cuestión que puede responderse a través de un simple análisis cluster a partir de las 27 variables geoestadísticas señaladas.

Para el análisis grupal se utilizó como medida de similitud el coeficiente de correlación de Pearson, en vez de la distancia euclidiana. Además, se consideró un método de conglomeración de grupos (vinculación inter-grupos), con la finalidad de maximizar la distancia entre los distintos grupos o clusters.

A partir del dendrograma obtenido mediante el análisis cluster, se va a evaluar la agrupación de las distintas áreas dentro de su clase temática y la separabilidad geoestadística entre éstas, como puede verse en los ejemplos representados en la Figura 3. La idea es que el usuario, antes de proceder a aplicar el algoritmo de clasificación, pueda modificar o redefinir algunas de estas áreas de entrenamiento o incluso clases, al objeto de obtener un mejor resultado en la clasificación.

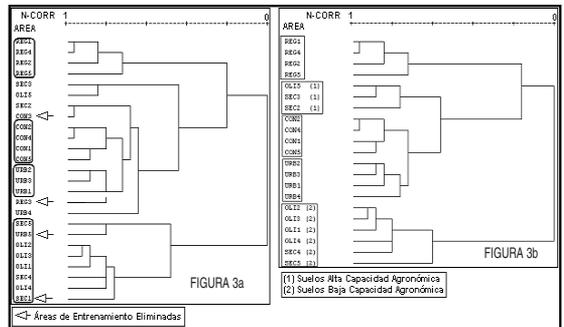


Figura 3. Análisis cluster de las áreas de entrenamiento a partir de las variables geoestadísticas.

## RESULTADOS Y CONCLUSIONES FINALES

El primer cluster realizado con toda la información (Figura 3a), muestra claramente que los parámetros geoestadísticos caracterizan a las clases temáticas, aunque se observa ciertas agrupaciones de mezcla de áreas pertenecientes a clases diferentes, principalmente secano y olivar, debido a que el substrato litológico y las condiciones topográficas se anteponen a la cubierta vegetal. En esta situación, frecuente en cualquier aplicación de teledetección, podría hacerse una redefinición temática de las dos clases indicadas, por ejemplo en el sentido de suelos con alta y baja capacidad agronómica lo que podría facilitar la interpretación de los resultados. También, puede contemplarse la eliminación de algunas áreas geoestadísticamente anómalas, como las señaladas en la Figura 3a. Con estas modificaciones se obtuvo un nuevo dendograma representado en la Figura 3b, en el que se observa una mejor separabilidad entre clases.

A fin de analizar estos resultados, se comparó la clasificación supervisada de la imagen, considerando las clases y áreas de entrenamiento iniciales, con las seleccionadas a través de este procedimiento geoestadístico exploratorio. El error global en la clasificación pasó del 16.9% al 7.9%, lo que supone una ganancia relativa en la disminución del error del 53%. Estos resultados preliminares, confirman la hipótesis de que es posible caracterizar geoestadísticamente las áreas de entrenamiento, lo que puede traducirse a través de un análisis exploratorio en una mejora de los resultados de la clasificación.

El variograma ha demostrado ser una herramienta potente para el análisis la variabilidad espacial de

los valores radiométricos en las áreas de entrenamiento. Así, los parámetros geoestadísticos utilizados para la caracterización del variograma, han definido bien la estructura de variación de las áreas de entrenamiento para las diferentes clases temáticas. Además, la facilidad de elaborar un algoritmo matemático para su extracción de estos parámetros, hace pensar en una implementación futura dentro de un clasificador.

## AGRADECIMIENTOS

Este trabajo se ha desarrollado en el marco del proyecto BTE2002-00152 financiado por el MCYT, así como en el ámbito del Grupo de Investigación RNM122 de la Junta de Andalucía.

## BIBLIOGRAFÍA

- CHICA-OLMO, M. y ABARCA, F., (1997): *Radiometric coregionalization of Landsat TM and SPOT HRV*, en International Journal Remote Sensing, vol. 19, nº 5, pp 997-1005
- KETTING, R.L. y LANDGREBE, D. A. (1975): *Classification of multispectral image data by extraction and clasification of homogeneous objects*. Symposium on Machine Classificaa-tión of Remotely Sensend Data, West Lafayette, pp. 1-11.
- SALVADOR, R. y SAN MIGUEL J. (2002): *An extension of a nonparametric clustering algorithm to derive radiometrically homogeneous objects pointed by seeds*. International Journal of Remote Sensing. Vol. 23, nº 6 pp. 1197-1205.