

## APLICACIÓN DE MEDIDAS DE VARIABILIDAD ESPACIAL PARA LA DISCRIMINACIÓN LITOLÓGICA EN IMÁGENES LANDSAT TM

F. Abarca y M. Chica-Olmo  
Laboratorio RSGIS "Teledetección, SIG y Geoestadística"  
Dpto. Geodinámica. Facultad de Ciencias. Universidad de Granada/IACT.  
Avda. Fuentenueva S/N. 18071. Granada, España.  
E-mail: rsgis@carpanta.ugr.es

**ABSTRACT:** Different measures of spatial variability (MSV) has been used in the framework of digital image classification. It is shown that these measures, when evaluated in a local window, can make important contribution to the discrimination of classes of land cover. The geostatistical methodology has been applied to the lithological classification of an interesting area, carrying out an assessment of the results.

### 1.0. Introducción

En la etapa de procesamiento digital de imágenes de satélite y, especialmente, en el campo de la clasificación digital, el análisis espacial de la información radiométrica tiene un gran interés, en particular, las características de autocorrelación espacial de los píxeles. Sin embargo, en la práctica, los algoritmos matemáticos clásicos de clasificación supervisada y no supervisada se basan mayoritariamente en la clasificación píxel a píxel, y no consideran la dependencia espectral existente entre un píxel y sus vecinos. Por esta razón, los resultados obtenidos mediante estos clasificadores podrían ser susceptibles de ser mejorados si se considera la información de autocorrelación, o de variabilidad, espacial, conjuntamente con la espectral en el proceso de clasificación.

En este trabajo se analiza mediante métodos geoestadísticos la dependencia espacial de los datos radiométricos e incluirla en los algoritmos de clasificación. Para ello se interpreta al número digital DN como una variable regionalizada (Matheron 1965), caracterizada por dos aspectos básicos, estructural y aleatorio, cuantificables mediante la función variograma. La información geoestadística resultante de carácter textural es posible considerarla como una nueva variable de interés en la estrategia de clasificación.

La metodología geoestadística que se expone está basada en el análisis de la estructura de variabilidad espacial de la radiometría (Curran 1988, Woodcock 1988, Chica-Olmo *et al.* 1991), cuyo principal objetivo es la evaluación de medidas de variabilidad espacial (MSV) para la discriminación y clasificación litológica de imágenes Landsat TM. Estas medidas que pueden ser consideradas como índices texturales se basan en el cálculo de diferentes estimadores geoestadísticos de la función variograma.

### 2.0. Area de estudio y datos de partida

La imagen utilizada para ilustrar el aspecto práctico de este trabajo cubre un área situada al SE de España, en la región de Cabo de Gata (Almería). El área corresponde a una zona donde tradicionalmente se han desarrollado investigaciones geomíneras, relacionadas con la exploración y explotación de yacimientos de Au.

Desde el punto de vista geomorfológico, el modelado está condicionado por la naturaleza volcánica de los materiales aflorantes, y por típicas llanuras de depósitos cuaternarios, producto de la erosión de estas rocas volcánicas en un clima

de carácter semiárido.

En lo referente al encuadre geológico, indicar que en el área afloran principalmente materiales volcánicos neógenos de carácter calcoalcalino, variando entre andesitas y riolitas. Estos materiales están afectados en mayor o menor grado por procesos de alteración hidrotermal favorecidos por la presencia de fracturas y fisuras, a los cuales se asocian mineralizaciones importantes de Au. Las rocas volcánicas predominantes en este sector presentan, a veces, composiciones mineralógicas relativamente parecidas, si bien, el aspecto estructural y textural ligado a los procesos de formación está bien diferenciado, encontrando intrusiones subvolcánicas, ignimbritas, coladas y flujos de material volcánico, domos, etc. Existen también afloramientos importantes de materiales terciarios, constituidos básicamente por depósitos bioclásticos. Las litologías volcánicas que afloran en el área son:

Dacitas anfibólicas	Domos y coladas. Rocas masivas con marcada foliación de flujo
Andesitas anfibólicas	Coladas e intrusiones subvolcánicas. Base formada por brechas autoclásticas.
Riolitas	Ignimbritas poco soldadas. Rocas con alteración hidrotermal.
Riolitas y dacitas anfibólicas	Domos formados por rocas masivas muy alteradas e intensamente fracturadas.
Complejo arrecifal	Calizas arrecifales y calcarenitas bioclásticas. Mioceno.
Depósitos cuaternarios	Arcillas, limos, arenas: meteorización de las rocas volcánicas

Tabla 1.- Relación y características de las rocas predominantes en el área de estudio

Para elaborar este trabajo se ha seleccionado una subescena Landsat TM tomada el 7 de julio de 1991 que cubre la zona descrita, con un área aproximada de 55 Km<sup>2</sup>, y un tamaño de 230 x 260 píxeles.

Las variables utilizadas para el cálculo de las medidas de variabilidad espacial no han sido las bandas TM originales, sino las obtenidas de la transformación de componentes principales, realizada con objeto de realzar las propiedades espectrales de las clases litológicas, y atenuar el efecto del relieve sobre los DN.

### 3.0. Aspectos metodológicos

El análisis de la regionalización de los datos es el punto inicial de cualquier estudio geoestadístico, cuya base es el cálculo, interpretación y modelización de la función variograma. Este análisis se ha desarrollado en distintas áreas de entrenamiento para la cuantificación de la variabilidad espacial de los DN, permitiendo, en cierta medida,

caracterizar espacialmente los distintos materiales que integran la zona de estudio.

### 3.1. Breve recordatorio de la función variograma

En el marco conceptual topoprobabilístico geoestadístico, el número digital de una imagen de satélite es considerado como una variable regionalizada (Curran 1988, Chica-Olmo *et al.* 1991, Atkinson 1993). Esta variable se interpreta como una función  $DN(x)$  que proporciona el número digital radiométrico  $DN$  en un pixel  $x$  localizado por sus coordenadas geográficas o por su fila y columna en la imagen. La variable regionalizada  $DN(x)$  se comporta como una función aleatoria, por lo que es posible considerar a una "imagen" como una realización particular de la función aleatoria  $DN(x)$ , que a su vez está formada por el conjunto de variables aleatorias  $\{DN(x_1), DN(x_2), \dots, DN(x_n)\}$ . Este aspecto permite considerar que, localmente, la variable regionalizada toma los valores  $\{dn(x_i), i=1, \dots, n\}$ , que son una "realización" particular de la función  $DN(x)$  de acuerdo a una ley de probabilidad de distribución espacial. Normalmente, esta ley de probabilidad espacial es desconocida, y en el caso de fenómenos naturales se tiene que inferir al menos parcialmente algunos de sus momentos de orden a partir de una única realización experimental, que en nuestro caso es el conjunto de valores numéricos que constituyen la imagen digital.

Para encontrar una solución a este problema se establecen algunas hipótesis restrictivas. La primera hipótesis es la estacionaridad de segundo orden, la cual asume que los dos primeros momentos de la función aleatoria  $DN(x)$  existen y son invariantes por traslación. Esto es, la esperanza matemática no depende de la posición del pixel  $x$ , y para parejas de valores, la covarianza existe y sólo depende del vector  $h$  de distancia que los separa. Otra situación menos restrictiva es la hipótesis intrínseca, basada en la estacionaridad de segundo orden de los incrementos de primer orden de las parejas de variables aleatorias  $DN(x)$  y  $DN(x+h)$ . Bajo esta hipótesis la esperanza matemática de los incrementos de primer orden existe y no depende de la posición del pixel  $x$ ,

$$E\{DN(x+h) - DN(x)\} = 0 \quad [1]$$

y la varianza de los incrementos de primer orden existe y no depende de los puntos  $x$  y  $x+h$ ,

$$Var\{DN(x+h) - DN(x)\} = 2\gamma(h) \quad [2]$$

donde  $\gamma(h)$  es la función variograma, que de acuerdo a la ecuación [2] se expresa como la mitad de la esperanza matemática de los incrementos cuadráticos de parejas de valores distantes  $h$ , y cuya expresión matemática es:

$$\gamma(h) = \frac{1}{2} E\{DN(x+h) - DN(x)\}^2 \quad [3]$$

de la cual se deduce que  $\gamma(h)$  es una función vectorial dependiente del módulo  $h$ , y el ángulo  $\theta$  que definen los puntos  $x+h$  y  $x$ ,  $\gamma(|h|, \theta)$

### 3.2. Medidas de variabilidad espacial (MSV)

La ecuación [3] representa la expresión clásica del variograma simple, aunque también se han propuesto otras expresiones de estimadores de la variabilidad espacial. Así, en este estudio se han considerado las siguientes funciones: variograma simple, madograma, rodograma y variogramas cruzado y pseudo-cruzado (Deutsch *et al.* 1992), las cuales

tienen gran interés en el estudio espacial de datos radiométricos.

#### i) Variograma simple

La inferencia estadística del variograma simple se realiza con el estimador de la ecuación [3]:

$$\gamma_k(h) = \frac{1}{2n(h)} \sum_{i=1}^{n(h)} \{dn_k(x_i) - dn_k(x_i+h)\}^2 \quad [4]$$

donde  $n(h)$  es el número de parejas distantes  $h$ ,  $dn_k(\cdot)$  son los valores digitales en los pixeles  $x_i$  y  $x_i+h$ , y  $k$  es la banda del sensor.

#### ii) Variograma cruzado

Cuantifica la variabilidad espacial conjunta (correlación cruzada) entre dos variables o bandas  $j$  y  $k$ ,

$$\gamma_{jk}(h) = \frac{1}{2n(h)} \sum_{i=1}^{n(h)} \{dn_j(x_i) - dn_j(x_i+h)\} \cdot \{dn_k(x_i) - dn_k(x_i+h)\} \quad [5]$$

#### iii) Variograma pseudo-cruzado

Considera las diferencias de incrementos cruzados en lugar de directos como en el caso anterior,

$$\gamma_{jk}(h) = -\frac{1}{2n(h)} \sum_{i=1}^{n(h)} \{dn_j(x_i) - dn_j(x_i+h)\} \cdot \{dn_k(x_i) - dn_k(x_i+h)\} \quad [6]$$

#### iv) Madograma

Considera el valor absoluto de las diferencias (variograma de orden 1, Matheron 1982),

$$\gamma_k(h) = \frac{1}{2n(h)} \sum_{i=1}^{n(h)} |dn_k(x_i) - dn_k(x_i+h)| \quad [7]$$

#### v) Rodograma

Corresponde a la raíz cuadrada del valor absoluto de las diferencias,

$$\gamma_k(h) = \frac{1}{2n(h)} \sum_{i=1}^{n(h)} \sqrt{|dn_k(x_i) - dn_k(x_i+h)|} \quad [8]$$

La interpretación de los variogramas de los datos radiométricos, en base a los conocimientos del fenómeno geológico estudiado, ha permitido obtener valiosa información de carácter geomínero referente al análisis de la variabilidad espacial de la información radiométrica.

Los aspectos de mayor relevancia para la interpretación de la función variograma son la *zona de influencia o alcance* que es la distancia para la cual desaparece la autocorrelación espacial entre los datos digitales, el *comportamiento en el origen*, que aporta información acerca del grado de continuidad de la variable, la *anisotropía* espacial de los números digitales, que revela comportamientos diferenciales direccionales en la variación de los valores radiométricos, y finalmente, la *superposición de estructuras*, asociadas a distintas escalas de variación (Chica-Olmo 1988).

El variograma experimental obtenido a partir de los estimadores anteriormente citados no es un verdadero variograma ya que debe cumplir determinadas condiciones de consistencia matemática, razón por la cual se debe ajustar un modelo de variograma. En particular el modelo de variograma debe ser definido positivo para garantizar la

positividad de la varianza de cualquier combinación lineal (Journel and Huijbregts, 1978).

Los modelos de variogramas más comúnmente usados en el ajuste variográfico son: esférico, gaussiano, exponencial, etc. La modelización se realiza, en general, por un proceso gráfico interactivo, escogiendo el usuario los parámetros que definen el modelo seleccionado. También existen diversos procedimientos numéricos de ajuste automático basados en el método de mínimos cuadrados ponderados.

Este análisis variográfico completo, cálculo, interpretación y ajuste, se ha llevado a cabo sobre dos variables sintéticas calculadas en las áreas de entrenamiento correspondientes a los grupos litológicos anteriormente descritos.

### 3.3. Selección de las variables de trabajo

En este estudio, a partir de la información radiométrica de las bandas TM se han obtenido dos variables adicionales deducidas de la transformación de componentes principales (PC), a fin de caracterizar y realizar propiedades espectrales de las clases litológicas estudiadas.

Las nuevas variables han sido obtenidas a partir del método Feature Oriented Principal Components Selection (FPCS), propuesto por Crosta and McM. Moore (1989). La base de este método es el examen detallado de los pesos de los eigenvectores para determinar las componentes principales que mejor se relacionan con las firmas espectrales teóricas de los objetos estudiados, en este caso, litologías.

La transformación de componentes principales de todas las bandas TM no llega a separar inequívocamente en una única componente aspectos geológicos, en nuestro caso, óxidos de hierro y la alteración hidrotermal (hydroxyl). Por esto se ha preferido seleccionar dos grupos de bandas representativas de la información espectral de los aspectos geológicos anteriormente referidos. Para el análisis de óxidos de hierro (Grupo I) se han seleccionado las bandas TM1, TM3, TM4 y TM5, y para el análisis de los minerales de alteración hidrotermal (Grupo II), las bandas TM3, TM4, TM5 y TM7.

La interpretación de los dos grupos de eigenvectores obtenidos en el cálculo de las componentes principales se ha abordado separadamente. La primera componente en ambos casos se puede interpretar como la imagen albedo, en la que todas las bandas tienen un peso positivo. Dentro del primer grupo de bandas, la PC2 contrasta los minerales ferrosos, peso alto de TM5 y bajo de TM4. La PC3 contrasta los óxidos de hierro, peso alto de TM1 y bajo TM3, al igual que la componente PC4, aunque con el signo de los pesos contrarios.

El segundo grupo de bandas resalta espectralmente los minerales de alteración. La PC2 describe el contraste entre el visible e infrarrojo próximo y el infrarrojo medio, pone también de manifiesto los minerales ferrosos, peso de TM5 alto y de TM4 bajo. La PC3 con peso alto en la TM7 y bajo en TM5 contrasta los minerales de alteración, al igual que la PC4, pero con menor intensidad.

Como resultado de este análisis, se han seleccionado una componente indicativa de cada grupo, PC2 y PC3, respectivamente, las cuales resaltan las propiedades de interés de los materiales geológicos del área de estudio. Por tanto, además de la información radiométrica TM, se disponen de estas dos variables adicionales que son susceptibles de ser analizadas espacialmente mediante la metodología expuesta, a través del cálculo de las medidas de variabilidad espacial.

## 4.0. Resultados

### 4.1. Análisis variográfico en las áreas de entrenamiento

Las dos componentes principales, PC2 y PC3, seleccionadas según el procedimiento descrito, han sido analizadas espacialmente mediante el cálculo, modelización e interpretación de las funciones variograma.

El cálculo se ha realizado en 6 áreas de entrenamiento de 500 píxeles, aproximadamente, las cuales corresponden a los 6 tipos litológicos descritos en la tabla 1. Indicar que el cálculo de la función variograma simple, omni-direccional, para las PC2 y PC3 se ha llevado a cabo para un paso de 1 píxel (30 m), y que la distancia máxima considerada corresponde a la mitad del tamaño de las áreas de entrenamiento, para así obtener un estimador robusto de la función. Asimismo, se ha utilizado el variograma medio ya que no se han observado anisotropías a escala hectométrica, aunque se han detectado anisotropías zonales importantes a escala kilométrica en el dominio la imagen completa.

El examen preliminar de los variogramas simples experimentales permite diferenciar comportamientos de la variabilidad espacial en los 6 grupos litológicos. Por otra parte, es destacable una similitud, en términos relativos, de los variogramas de las dos componentes, PC2 y PC3, que puede ser debido a un efecto proporcional entre las bandas del TM. En términos generales, los alcances varían desde valores bajos, menos de 100 metros en los materiales terciarios del complejo arrecifal, hasta relativamente altos, más de 300 metros, como es el caso de las dacitas anfibólicas. Por otro lado, las mesetas se manifiestan marcadamente diferentes entre clases; así, los materiales terciarios y andesitas anfibólicas presentan una meseta muy baja, lo que significa que son bastante homogéneos, aunque su correlación espacial no es muy alta. Por el contrario, las dacitas anfibólicas y las riolitas son más heterogéneas, presentan una mayor varianza, si bien, espacialmente son más continuas.

Las diferencias observadas en el comportamiento espacial de las clases litológicas, algunas similares en su composición como los son las dacitas y las andesitas anfibólicas, se debe en gran medida a los aspectos estructurales y texturales ligados a los procesos de formación. Así, por ejemplo, las dacitas anfibólicas que afloran como rocas masivas (domos) muestran mayor correlación espacial (alcance) que las andesitas anfibólicas formadas por brechas autoclásticas de poca extensión.

### 4.2. Cálculo de las medidas de variabilidad espacial y análisis discriminante

Con el fin de introducir la información de variabilidad espacial en la estrategia de clasificación se han calculado las medidas geoestadísticas de variabilidad espacial descritas en el apartado 3.2 sobre las componentes PC2 y PC3. Para ello se ha utilizado una ventana móvil de 7 x 7 píxeles, dimensión que ha dado los mejores resultados en las diferentes pruebas realizadas, asignando al píxel central el valor de la función. Puesto que en todos los casos se trata de funciones vectoriales se ha seleccionado el valor  $h=1$  píxel (30 m). Las funciones son: variograma simple (G), madograma (M), rodograma (R), variograma cruzado (CG) y pseudo cruzado (PG). La dirección de cálculo ha sido la media de las cuatro direcciones más importantes (NS, EW, N45W y N45E).

Las dos componentes principales, PC2 y PC3, han sido procesadas dando lugar a 8 nuevas variables (6 variogramas directos y 2 cruzados) para cada una de las seis litologías. El número de estas variables complementarias parece elevado y,

probablemente, existe en ellas información redundante. Con el fin de seleccionar las medidas de variabilidad más relevantes se ha llevado a cabo un análisis discriminante de un grupo de 205 píxeles, elegidos de 41 áreas de entrenamiento pertenecientes a las seis litologías, tomando aleatoriamente 5 píxeles de cada una.

Para cada píxel se disponen de un total de 16 variables: 6 bandas TM, 8 medidas de variabilidad espacial, y también, se han incluido, los 2 valores de la varianza estadística (V) calculados sobre las dos componentes en las ventanas móviles. Se ha preferido usar la varianza en lugar de la desviación estándar ya que las expresiones geoestadísticas son mayoritariamente cuadráticas.

De los resultados del análisis discriminante son destacables dos aspectos importantes. Por un lado, el notable incremento de los porcentajes de aciertos en la clasificación con el uso conjunto de la información espectral (bandas TM) y espacial (MSV), que en términos relativos puede llegar a más del 50%. De otra parte, se observa que la incorporación de los variogramas cruzados (correlación cruzada espacial) tiene un peso importante en la mejora de los resultados, equivalente a la obtenida a partir de las funciones univariantes (G, M y R).

#### 4.3. Clasificación supervisada: método de máxima probabilidad

A la vista de los resultados del análisis discriminante, se ha clasificado la imagen de acuerdo a la regla de decisión de máxima probabilidad utilizando la combinación de las 6 bandas TM, y 6 medidas de variabilidad espacial seleccionadas a partir del análisis discriminante como óptimas: variograma (G) y madograma (M) de cada componente, y el variograma cruzado (CG) y pseudo-cruzado (PG) de las dos componentes.

En la zona de estudio se han seleccionado 41 áreas de entrenamiento de tamaños similares, donde se conoce la realidad terreno. El objetivo es clasificar la imagen Landsat TM de acuerdo a las 6 clases litológicas predominantes en la zona, y comparar los resultados del método clásico usando sólo la información radiométrica (TM) y el método propuesto utilizando conjuntamente la información radiométrica y las medidas de variabilidad espacial (MSV)

La tabla 2 presenta la matriz de contingencia de las áreas de entrenamiento correspondiente a los resultados obtenidos en la clasificación de la imagen Landsat TM para los dos casos.

Litología	%tm	%msv	%l	pix.
Dacitas anfíbólicas	72.2	94.8	31.3	309
Andes. anfíbólicas	56.3	85.3	51.5	190
Riolitas	87.1	90.5	3.9	116
Riol. y dac. anfib.	87.3	99.3	14.0	150
Complejo arrecifal	77.1	90.7	17.6	475
M. cuaternarios	77.4	91.9	18.7	570
<b>Total</b>	<b>75.6</b>	<b>91.9</b>	<b>22.8</b>	<b>1810</b>

Tabla 2. Porcentajes de píxeles clasificados correctamente del total (pix.) a partir de las bandas espectrales TM (tm) y TM junto con las MSV (msv), indicando el incremento relativo de aciertos (%l).

A la vista de estos datos, se puede decir que el incremento en los porcentajes es muy aceptable, como se podía prever de los resultados obtenidos en el análisis discriminante. Las mejoras en los resultados de la clasificación varían en función del tipo litológico. Los valores extremos están en una mejora del 51% para las andesitas anfíbólicas y de sólo un 4% para las riolitas. En términos generales, se puede establecer el incremento medio en torno al 20%.

#### 5.0. Discusión y conclusiones

La selección de variables complementarias, PC2 y PC3 de los grupos de bandas I y II respectivamente, basada en la técnica FPCS, ha dado buenos resultados para la caracterización espectral de las litologías. El cálculo del variograma de estas componentes en las áreas de entrenamiento ofrece una primera aproximación a la separabilidad de las clases para diferentes valores de *h*, dando una primera valoración de la información espacial que puede proporcionar el cálculo de las MSV.

El análisis discriminante ha permitido la valoración de la contribución de cada una de las MSV a la separabilidad de las clases. Asimismo, ha facilitado la selección de las variables de mayor interés para la fase de clasificación supervisada.

La clasificación supervisada ha puesto de manifiesto que la incorporación de las MSV a la clasificación produce resultados más homogéneos que los obtenidos sólomente a partir de los valores radiométricos de las bandas TM.

Con los resultados obtenidos en este trabajo se concluye que el análisis de la variabilidad espacial basado en una metodología geoestadística proporciona información de interés que mejoran considerablemente los resultados de la clasificación digital por métodos clásicos.

#### Agradecimientos

La realización de este trabajo ha sido posible gracias al Proyecto de I+D DARSTIMEX (BRE2-CT93-0513) del programa BRITE/EURAM de la CE y a la financiación recibida por la DGICYT correspondiente al proyecto PB92-0925-CO2-02.

#### Bibliografía

- Crosta, A.P., and J.McM. Moore. (1989). Enhancement of Landsat Thematic Mapper Imagery for Residual Soil Mapping in SW Minasis Gerias State. Proceedings of the 7th (ERIM) .Calgary, 2-6, pp. 1173-1187.
- Chica-Olmo, M., Vives, E. y Abarca, F. (1991). Métodos geoestadísticos aplicados al análisis de variables espaciales. IV Reunión de la Asociación Española de Teledetección.
- Chica-Olmo, M. (1988). Análisis geoestadístico en el estudio de la explotación de los recursos minerales. Tesis Doctoral. Universidad de Granada.
- Cunan, P.J. (1988). The semivariogram in Remote Sensing: An Introduction. Remote Sensing of Environment. Vol. 24, pp. 493-507.
- Journel, A.G., And Huijbregts, C.J. (1978). Mining Geostatistics. Academic Press, London.
- Matheron, G. (1965). Les variables Régionalisées et leur Estimation, Masson, Paris.
- Woodcock, C.E, Strahler, A.H, and Jupp. D.L.B. (1988). The Use of Variograms in Remote Sensing: II. Real Digital Images. Remote Sensing of Environment. Vol. 25, pp. 349-379.