

## Simulación geoestadística de imágenes de satélite en teledetección

*F. Abarca, M. Chica-Olmo y E. Pardo Igúzquiza*

Dpto. Geodinámica  
Universidad de Granada

### RESUMEN

En este trabajo se describe la técnica geoestadística de simulación condicional y su aplicación en el caso práctico de la integración espacial de la banda 6 de una imagen LANDSAT TM con el resto de las bandas. Hemos considerado dos aspectos fundamentales: por un lado, el tratamiento previo de la información y definición de un modelo de variabilidad espacial que será usado en las fases de simulación no condicional y condicionamiento. De otro lado, se expone detalladamente el proceso de simulación no condicional por el método de la Integral de Fourier, punto de partida para el condicionamiento que proporcionará la imagen simulada.

### ABSTRACT

This paper reports on the application of conditional simulation techniques to the band 6 of LANDSAT TM as a method for spatial integration of this band with the rest of the bands. Two fundamental aspects have been taken into account. On the one hand, basic preprocessing of the experimental data and definition of a spatial model which will be used in the no-conditional simulation and the conditioning. On the other hand, the no-conditional simulation using the Fourier Integral Method (FIM) and then conditioning and obtention of the final simulated image with a defined spatial resolution.

### Introducción

Dado el elevado número de sensores existentes y sus diferentes resoluciones espaciales, el trabajo que desarrollamos en este trabajo tiene una fuerte componente práctica.

Abrimos aquí el capítulo dedicado a la simulación geoestadística, exponiendo su base metodológica y aplicación a un caso práctico de interés, como es, asignar a la banda TM6 la resolución espacial del resto de las bandas que componen la imagen Landsat Thematic Mapper.

Consideraremos el número digital de los píxeles definitorios de la imagen como una variable regionalizada. Dado que estamos frente a una variable de manifiesta co-

relación espacial, es susceptible de ser tratada geoestadísticamente (Chica et al. 1988). Siendo así, al computar el valor de los números digitales y a fin de establecer la nueva resolución en la banda TM6, nos encontramos ante una alternativa, ¿simulación o estimación? cada una de las posibilidades ofrecidas por la geoestadística tiene sus ventajas e inconvenientes.

Con el empleo de la simulación geoestadística se persigue el objetivo de obtener una imagen en la cual la información experimental no se ve alterada; la estimación (Chica-Olmo, 1991) por el contrario produce un efecto de suavizado no siempre deseable, consecuencia del proceso de estimación, cuyo valor puede encontrarse en la elaboración de filtros de imágenes. Por el contrario, la imagen simulada tiene unos estadísticos idénticos a la de partida, además de conservar la estructura de autocorrelación espacial cuantificada por la función variograma.

El método de simulación geoestadística que se ha introducido en el algoritmo de cálculo ha sido el Método de la Integral de Fourier (Pardo-Igúzquiza, 1993), si bien existen otras técnicas de simulación estocástica de procesos aleatorios.

### ***Prolegómenos***

Sin pretender abundar en los conceptos que concretan la base conceptual de la Geoestadística vemos interesante introducir algunas referencias a la Teoría de las Variables Regionalizadas (Matheron, 1965), para facilitar la comprensión de este trabajo.

Una variable regionalizada se define como una función  $ND(x)$  la cual devuelve el valor de una característica ND (número digital), siendo el parámetro de entrada la posición espacial del punto  $x$  de observación; de este modo, el número digital característico de la intensidad radiométrica puede ser considerado una variable regionalizada, quedando definida su localización espacial por la línea y columna que ocupa en la imagen, ésta a su vez, referenciada a un sistema de coordenadas.

Esta variable presenta dos aspectos fundamentales, por un lado, el carácter aleatorio asociado a la noción de variable aleatoria; y por otro, el carácter estructural inherente a cada fenómeno o regionalización, este último caracterizado por la función variograma.

El variograma, aféresis de semivariograma, es una función intrínseca definida como la mitad del momento de segundo orden de los incrementos de primer orden de la función aleatoria  $ND(x)$ .

$$\gamma(h) = 1/2E \{ND(x) - ND(x+h)\}^2 \quad (1)$$

Es una función vectorial cuyo argumento es el vector distancia  $h$ , y cuantifica la varianza de los incrementos cuadráticos de primer orden de la función. Esto quiere decir que a través de la información aportada por esta función puede caracterizar espacialmente la estructura de variación del fenómeno. Un estimador insesgado empleado con frecuencia para el cálculo de la función variograma a partir de los datos experimentales es el que se representa en la ecuación (2).

$$\gamma^* = \frac{NP(h)}{1/2NP_h} \sum_{i=1}^{NP_h} \{z(ND_i) - z(ND_{i+h})\}^2 \quad (2)$$

donde  $ND(x)$ ,  $ND(x_{i+h})$  son los datos experimentales observados en los puntos  $ND_i$  y  $ND_{i+h}$ ;  $NP_h$  es el número posible de parejas formadas para una distancia  $h$ .

En la Figura 1 se han representado gráficamente los variogramas direccionales de la banda 6 TM obtenidos en las cuatro direcciones frecuentemente empleadas. Se aprecia un comportamiento más o menos isótropo de la variable; en todas las direcciones la estructura de variación posee similares características, de tal suerte, que los coeficientes que caracterizan la elipse de anisotropía valen 1. Es de resaltar que los datos de ND se han transformado por requerimientos del método de simulación en su equivalente gaussiano por el método de la anamorfosis gaussiana (Chica-Olmo, 1988).

El modelo teórico variográfico al que se ha ajustado el variograma experimental es el que se relaciona en la Tabla 1. Este ajuste es necesario para operar con una función numérica continua y no con los valores estimados y discretos.

El valor del efecto de pepita representa un 22% de la varianza de la información. Este valor relativamente importante puede interpretarse de dos formas. a) Heterogeneidad en la respuesta del IR térmico de la cubierta; b) posibles desajustes en la captación y transformación digital de los datos. En este sentido, no disponemos de argumentos para discernir sobre estas dos posibilidades.

Modelo	Alcance	Meseta	An. X	An. Y
E. Pepita	0.22			
Esférico	3.5	0.28	1	1
Esférico	27.0	0.50	1	1

Tabla 1  
Modelo teórico ajustado sobre datos gaussianos

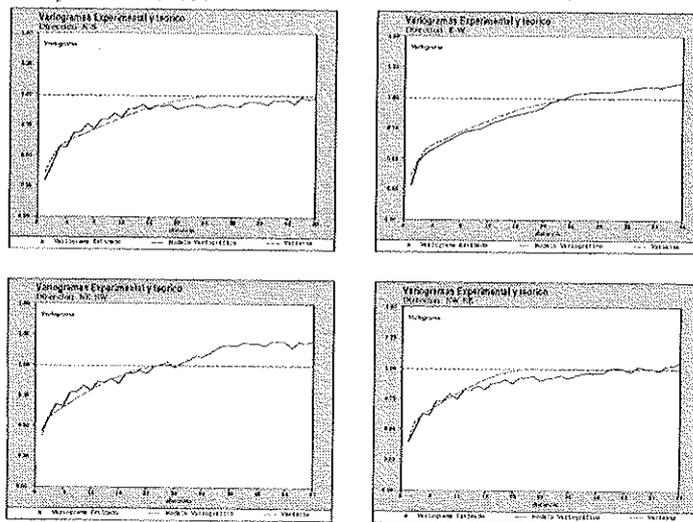


FIGURA I  
Ajuste variográfico de los datos experimentales

El análisis estructural descrito debe efectuarse en tres diferentes estadios de la simulación; en primer lugar, sobre los datos experimentales con una doble finalidad, hacer una primera aproximación de la estructura de variabilidad de la imagen, y, no menos importante, contrastar estos datos con los calculados en la imagen final simulada para verificar el éxito de la operación. Por otra parte, el análisis más preciso y concienzudo será el que se realice sobre los datos transformados gaussianos, será este el modelo que se emplee en las futuras fases de la simulación. Finalmente, y a modo de chequeo, es conveniente comprobar la semejanza de la función variograma de la imagen simulada restituida con la imagen de partida.

### Procesamiento digital

Se explica el proceso que se ha seguido para la selección de los datos experimentales. En principio, es posible tomar toda la información contenida en la banda 6 TM y arrastrar en todos los procesos la redundancia que en ella se encierra, por nuestra parte, hemos preferido eliminar esa información innecesaria en un primer paso previo al esquema de trabajo. En la Figura 2 se ha dibujado un diagrama en el que se refleja la estructura de los píxeles de 120 metros (banda 6 TM) y la relación con el resto de las bandas cuyos píxeles tienen una resolución espacial de 30 metros. Dado que hay dieciséis píxeles con valores idénticos, hemos seleccionado uno, por ejemplo el de la esquina superior izquierda. El conjunto de todos estos puntos seleccionados definen una imagen idéntica a la de partida pero de un tamaño dieciséis veces menor, está es nuestra imagen de partida.

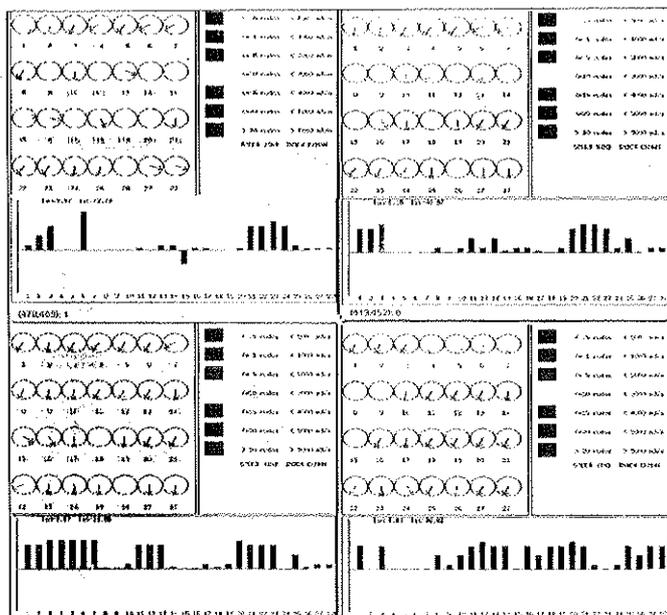


FIGURA 2

Relación entre píxeles de la banda 6 con el resto de las bandas y puntos condicionantes y simulados

En la Figura 3 se representa la imagen que hemos empleado con el propósito de ejemplificar cada una de las etapas requeridas para la consecución de una imagen simulada; se trata de una subescena de 512 x 512 píxeles de la Banda 6 TM, aunque de acuerdo con el expuesto, la imagen sobre la que hemos desarrollado este trabajo tiene una resolución de 128 x 128 píxeles.

Esta fase inicial de la que excluimos, por la razón expuesta, la información redundante contenida en los datos experimentales, será en otras aplicaciones innecesaria, por ejemplo, para la conjunción de dos imágenes con resolución espacial diferente, una MSS y una LANDSAT, no será necesario dado que no existe redundancia en la información de partida.

### Simulación de imágenes satélite

La simulación condicional de variables regionalizadas, en el caso que nos ocupa, bidimensional, tiene como principal objetivo la generación de realizaciones estocásticas de una función aleatoria  $ND(x)$ . La función aleatoria simulada ha sido calculada por el método espectral desarrollado por Shinozuka y Jan (1972) y el condicionamiento se ha llevado a cabo por la técnica de Krigeaje.

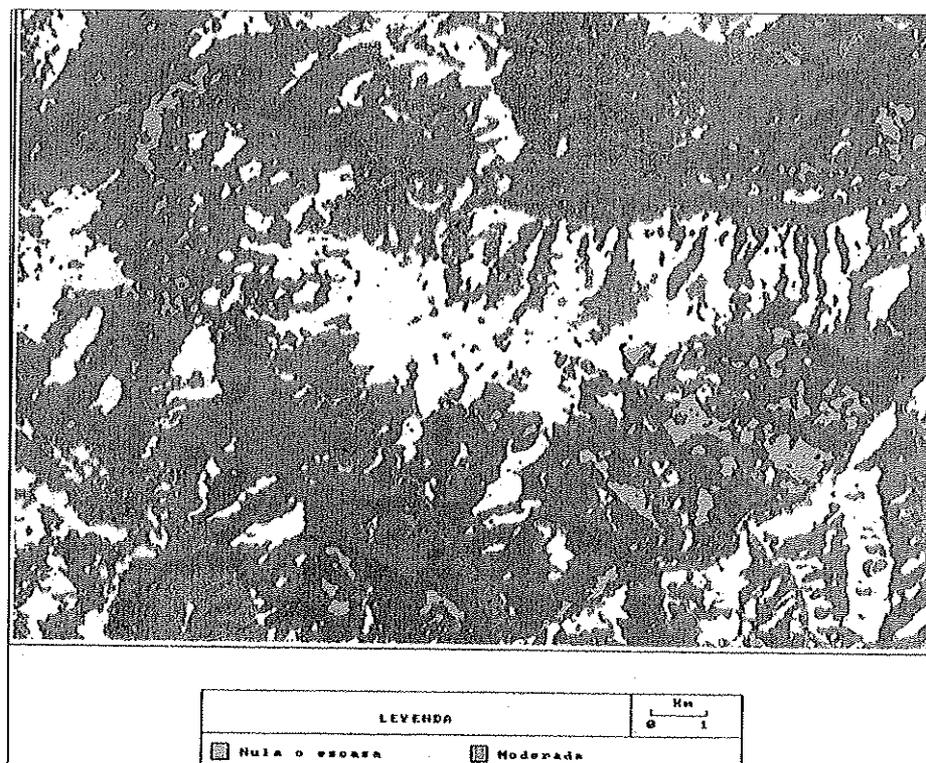


FIGURA 3  
Subescena de una imagen Landsat 6 TM

La Figura 4 presenta un esquema en el cual quedan reflejados todos los pasos a seguir en la simulación, y son éstos los que vamos a describir haciendo hincapié en los aspectos más relevantes desde el punto de vista del tratamiento de imágenes.

#### *Anamorfosis gaussiana*

La función de distribución de probabilidad de los valores experimentales es desconocida, considerando el histograma suficientemente representativo. Este histograma puede responder a cualquier modelo de función de distribución, y no necesariamente gaussiano. La mayoría de los algoritmos de simulación no condicional generan valores cuya función de distribución es gaussiana, consecuencia del Teorema Central del Límite.

Esta transformación se denomina anamorfosis inversa y se ha realizado por el método de Monte Carlo o anamorfosis gráfica. Una variable continua puede ser representada como una función gaussiana mediante anamorfosis.

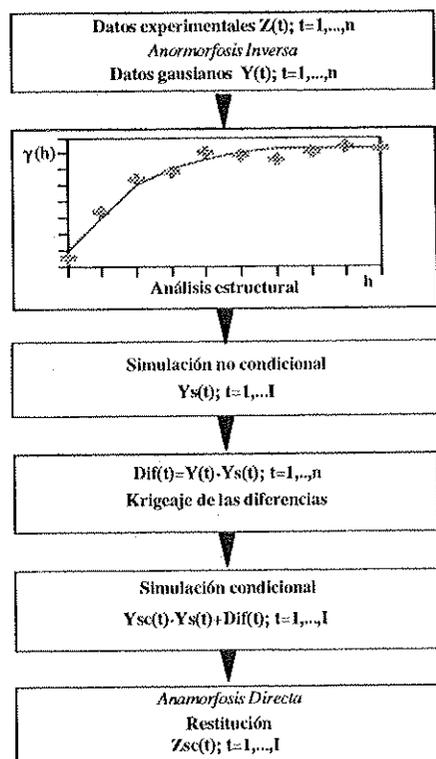


FIGURA 4  
Pasos en la Simulación

La obtención de los valores transformados mediante procedimiento gráfico se efectúa a partir del histograma de los datos experimentales, se trata de hacer corresponder a éste el histograma de frecuencias acumuladas de la ley normal.

En la Figura 5 se ha intentado mostrar la transformación asociando una variable normal  $Y_i$  de distribución  $G(y)$  a una variable  $Z_i$  de distribución  $F(z)$  y viceversa. Esta reciprocidad es viable ya que  $G$  y  $F$  son funciones monótonas creciente, pudiéndose establecer la biyección  $G(y) = F(z)$ . (Ver Chica-Olmo, 1988).

### Análisis variográfico

Como ya hemos anticipado, el análisis estructural consiste en la estimación de la función variograma y su ajuste a un modelo teórico. Este análisis es obligado realizarlo sobre los datos transformados gaussianos, el modelo que de aquí se desprenda será el que emplearemos como representativo de la estructura de variación espacial y por tanto el introducido como parámetro al algoritmo de simulación.

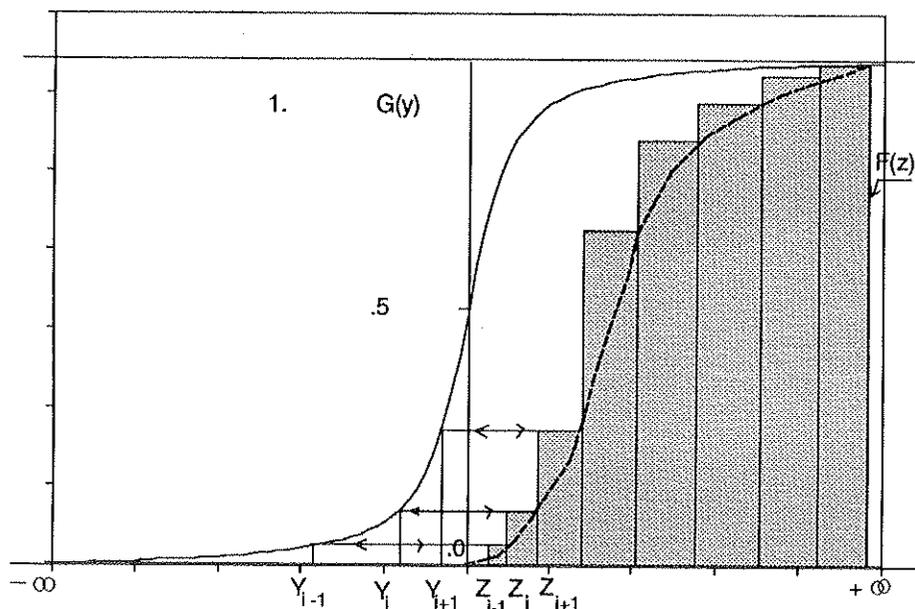


FIGURA 5  
Anamorfosis gaussiana. Método de Monte Carlo

Ya hemos relacionado los parámetros definitorios del modelo variográfico ajustado y en la Figura 1 la representación del mismo; de lo visto se infiere que la correlación espacial es extremadamente alta y, por otro lado, que el fenómeno tiene un comportamiento isotrópico. Señalar que se han ajustado dos modelos superpuestos de alcances 3,5 y 27 píxeles, respectivamente, 420 y 3240 metros. Indicaremos que en nuestro caso de estudio, la estimación de la función variograma ha sido realizada considerando toda la imagen, siendo factible, en casos no exitosos, localizar sectores de mejor estructuración. Normalmente, la función variograma manifiesta una estructura de variación más que aceptable, más fiable cuanto mayor sea el número de puntos implicados.

#### *Simulación no condicional*

La simulación geoestadística presenta numerosos métodos para la generación de realizaciones de una función aleatoria estacionaria, de estos, el más clásico es el de las Bandas Rotantes. Este método presenta algunos problemas en simulaciones en 2D, condiciones usuales en el tratamiento de imágenes. Estas irregularidades son explicadas porque la simulación multidimensional es simplificada a una simulación unidimensional sobre conjunto de rectas distribuidas de un modo homogéneo en el espacio n-dimensional, sobre éstas se obtienen las realizaciones simuladas unidimensionales.

El método que vamos a exponer bajo estas líneas es el de la Integral de Fourier, con el cual vamos a generar un conjunto finito de realizaciones distribuidas según una malla, definida ésta por la resolución espacial final que deseemos.

La representación espectral de cualquier proceso aleatorio o estocástico  $Z(k)$ , lo llamaremos indistintamente, puede ser escrita como se muestra:

$$Z(k) = \int_0^{\pi} \cos(k\omega) dv_1(\omega) + \int_0^{\pi} \sin(k\omega) dv_2(\omega) \quad (1)$$

Estos procesos,  $dv_1(\omega)$ ,  $dv_2(\omega)$ , son los límites de la transformada del seno y del coseno.

$$\begin{aligned} \sum_k Z(k) \cos(k\omega) \\ \sum_k Z(k) \sin(k\omega) \end{aligned} \quad (2)$$

Las propiedades de segundo orden de los procesos son especificados por la función de densidad espectral  $S(\omega)$  (Anderson, 1971):

donde:

$0 \leq \omega_1 \leq \omega_2 \leq \pi$ ,  $E[.]$ : esperanza matemática,  $Cov[.]$ : covarianza,  $Var[.]$ : varianza, y  $S(.)$ : función de densidad espectral.

$$E[dv_1(\omega)] = E[dv_2(\omega)] = 0$$

$$E[dv_1(\omega_2) - dv_1(\omega_1)]^2 = E[dv_2(\omega_2) - dv_2(\omega_1)]^2 = \int_{\omega_1}^{\omega_2} S(\omega) d\omega \quad (3)$$

$$Cov([dv_1(\omega), dv_2(\omega) = E[dv_1(\omega) dv_2(\omega)] = 0$$

$$Var[Z(k)] = \int_{-\pi}^{\pi} S(\omega) d\omega$$

Por otro lado, por el llamado teorema de Wiener-Khintchine queda establecido que cualquier proceso estacionario tiene una covarianza  $C(h)$  de la forma (Anderson, 1971):

$$C(h) = \int_{-\pi}^{\pi} S(\omega) e^{i\omega h} d\omega \quad (4)$$

$i = (-1)^j$ , y  $S(\cdot)$  función de densidad espectral. Y recíprocamente, dada una función representada por la última ecuación, existe un proceso estacionario con  $C(h)$  como covarianza (Cox y Miller, 1968). Ambas funciones, covarianza y función de densidad espectral contienen la misma información pero expresa de forma diferente.

Las dos posibles interpretaciones de la función de densidad espectral  $S(\omega)$  son, en primer lugar, da el porcentaje de varianza de  $Z(k)$  aportada por las componentes en el rango  $(\omega, \omega + \Delta\omega)$ . Y en segundo, da directamente las componentes en una análisis de Fourier de la función de correlación.

En la práctica, podemos obviar las integrales estocásticas y simplemente considerar  $Z(k)$  como una combinación lineal de términos sinusoidales y ortogonales (Chatfield, 1991).

Construimos un modelo de campo aleatorio discreto que es una combinación lineal de senos y cosenos que tienen fases y amplitudes aleatorias; en nuestro caso lo que nos interesa son las amplitudes. La suma de las amplitudes cuadráticas medias de las frecuencias en un intervalo dado es igual a la suma de los valores de la función de densidad espectral en este intervalo, esto es conocido como análisis espectral.

Una secuencia  $z(k)$  de  $N$  puntos pueden ser transformados en un conjunto finito de coeficientes de Fourier (Chatfield, 1991):

$$z(k) = \sum_{j=0}^{N-1} [a_j \cos(2\pi jk/N) + b_j \sin(2\pi jk/N)] \quad (5)$$

$\{k=0, \dots, N-1\}$ ,  $N$ : número de puntos de la secuencia.  
 $\{J=, \dots, N-1\}$ .

$$\begin{aligned} a_j &= \frac{1}{N} \sum_{k=0}^{N-1} z(k) \cos(2\pi jk/N) \\ b_j &= \frac{1}{N} \sum_{k=0}^{N-1} z(k) \sin(2\pi jk/N) \end{aligned} \quad (6)$$

Otra forma de representar esta secuencia discreta finita es como una serie de Fourier exponencial compleja (Hsu y Mehra, 1973):

$$z(k) = \sum_{j=0}^{N-1} A(j) e^{i(2\pi jk/N)} \quad (7)$$

$i = (-1)^{j/2}$ , and  $\{k=0, \dots, N-1\}$ . Donde:

$$A(j) = |A(j)|e^{-i\phi(j)} \quad (8)$$

Los coeficientes  $A(j)$  están relacionados con los coeficientes  $a_j, b_j$  por las ecuaciones:

$$\begin{aligned} |A(j)| &= \sqrt{a_j^2 + b_j^2} \\ \phi(j) &= \tan^{-1} (-b_j/a_j) \end{aligned} \quad (9)$$

El espectro de amplitudes discreto es la representación de la amplitud  $|A(j)|$  contra la frecuencia  $2\pi j/N$  y el espectro de fase es el gráfico de fase  $\phi(j)$  contra frecuencia  $2\pi j/N$ .

El teorema de Parseval-Rayleigh en su enunciado dice que la suma del cuadrado de las amplitudes es igual a la potencia total de  $\{z(k), k=0, \dots, N-1\}$  que puede ser identificada con la varianza (Bracewell, 1986):

$$\sum_{j=0}^{N-1} |A(j)|^2 = \frac{1}{N} \sum_{k=0}^{N-1} |z(k)|^2 = \sigma^2 \quad (10)$$

$\sigma^2$ : varianza de la secuencia  $\{z(k), k=0, \dots, N-1\}$ . El espectro de amplitudes discreto puede relacionarse directamente con la amplitud espectral discreta. Este resultado es igualmente obtenido del teorema de la transformada de Fourier de la función de correlación.

Por definición, la función de densidad espectral es la transformada de Fourier de la función covarianza.

La función de densidad espectral es una función par  $S(j) = S(-j)$ , significa que  $|A(j)| = |A(-j)|$ .

El espectro de amplitudes no puede ser aleatorio pero puede relacionarse con la densidad espectral, la cual depende del modelo de covarianza que deseemos imponer a las realizaciones.

El espectro de fase no es afectado por la estructura de la covarianza, esto significa que la podemos tomar aleatoria de una distribución uniforme entre 0 y  $2\pi$ . De este modo componemos los coeficientes complejos de Fourier:

$$\begin{aligned} \text{Parte Real: } R(j) &= |A(j)| \cos \phi(j) \\ \text{Parte imaginaria: } I(j) &= |A(j)| \sin \phi(j) \\ A(j) &= R(j) - iI(j) \end{aligned}$$

Por el cálculo de la transformada de Fourier inversa de los coeficientes complejos  $A(j)$ , las realizaciones finitas discretas  $\{z(k), \dots, N-1\}$  se obtienen con el modelo de covarianza impuesto.

Para el cálculo de la transformada de Fourier se han empleado FFT (Fast Fourier Transform) imprimiendo gran velocidad a los todos los cálculos, la imposición es que la matriz de datos a generar debe ser de filas y columnas potencias de dos (Pardo-Igúzquiza, 1991).

El resultado de esta etapa es una matriz de datos gaussianos de  $512 \times 512$ , con resolución espacial de 30 metros, que tiene la misma variabilidad espacial que la imagen de partida.

### **Condicionamiento**

Los valores que hemos simulado no coinciden con los puntos experimentales, esta fase de condicionamiento se lleva a cabo mediante el método de Krigearje (Journel y Huijbregts, 1978):

$$\begin{aligned} ND_{sc}(x) &= ND_s(x) + [ND_k^*(x) - ND_{sk}^*(x)] \\ ND_{sc}(x) &= ND_s(x) + \sum_{i=1}^N \lambda_i [ND(x_i) - ND_s(x_i)] \end{aligned} \tag{11}$$

donde:

$ND_{sc}$  = datos gaussianos simulados condicionados

$ND_s(x)$  = datos gaussianos simulados no condicionados

$ND(x)$  = datos gaussianos experimentales

$ND_k^*(x)$  valor estimado por krigearje en puntos condicionantes

$ND_{sk}^*(x)$  valor estimado por krigearje en puntos no condicionantes

$\lambda_i$  = coeficientes de krigearje

$N$  = Número de puntos de vecindad elegidos en el krigearje

### **Restitución de los valores simulados**

Esta sencilla operación se denomina anamorfosis directa y es el paso inverso al ya descrito como anamorfosis inversa. En esta fase se transforman los valores gaussianos a los equivalentes experimentales de los que partíamos. Restituidos los datos se debe cumplir que el histograma y demás estadísticos sean idénticos a los de la imagen de partida. En la Figura 6 se representa una simulación de la imagen de partida (Figura 3), siendo ésta una posible versión de la realidad, los estadísticos son coincidentes, al igual que su estructura de variabilidad.

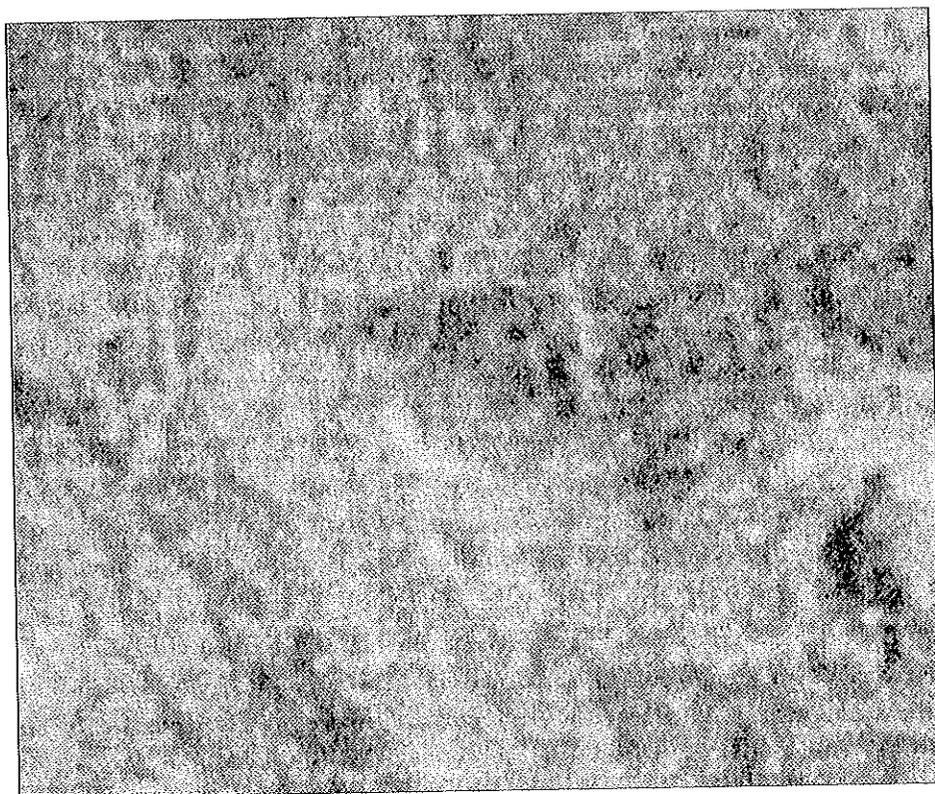


FIGURA 6  
Imagen LANDSAT TM 6 simulada

La simulación nos permite, ya definido el modelo variográfico, realizar un número infinito de repeticiones dando situaciones distintas, dependiendo del valor de la "semilla" del generador de números aleatorios asociado al espectro de fase que aludíamos en el apartado de simulación no condicional.

Los ND de los píxeles simulados presenta un marco de variación notable, estas oscilaciones van a depender de la estructura de variación espacial definida y de los datos experimentales, si bien, la estructura de autocorrelación será la misma. Si hacemos una composición RGB asociando a cada canal una imagen simulada, veremos, por un lado los puntos condicionantes en tonos grises puesto que tienen los mismos valores en las tres imágenes, y por otro, zonas con predominio de un color sobre el resto, manifestando estas diferencias de color las variaciones que se producen en la simulación.

Mientras que imágenes simuladas podemos tener tantas como deseemos, estimadas sólo hay una, siendo esta la más probable pero sin conservar los estadísticos ni

estructura de autocorrelación. Por este motivo es preciso plantearse el objeto del empleo de la imagen antes de decidirse por una imagen simulada o estimada; por supuesto, cada método tiene su campo de aplicación específico.

### **Conclusiones**

A lo largo de este trabajo hemos hablado de conjuntar la resolución espacial de distintos sensores, el caso aquí desarrollado ha sido testeado mediante un análisis de componentes principales empleando una imagen TM original y otra en la que la banda 6 era simulada. Los resultados han sido francamente alentadores.

La simulación geoestadística permite generar realizaciones de una función aleatoria imponiendo un modelo de variación espacial y una correlación con otra u otras variables (bandas), es el caso de la simulación multivariable.

Con esta metodología se abre un interesante campo de aplicaciones de la simulación en el terreno de la conjunción de resolución espacial entre información procedente de distintas fuentes.

### **Bibliografía**

- Anderson, T.W.:** 1971, *The Statistical Analysis of Time Series*; John Wiley and Sons, New York, 704 p.
- Bracewell, R.N.:** 1986, *The Fourier Transform and Its Applications*; Mc-Graw Hill International Editions, Singapore, 474 p.
- Chatfield, C.:** 1991, *The analysis of Time Series*, 4th ed.; Chapman and Hall, London, 241 p.
- Chica-Olmo, M.:** 1988, *Análisis Geoestadístico en el estudio de la explotación de los recursos minerales*. Tesis. Dpto. Geodinámica. Univ. Granada. 387 p.
- Chica-Olmo, M., Vives, E. y Abarca, E.:** 1991. *Métodos geoestadísticos aplicados al análisis de variables espaciales en teledetección*. IV Reunión Científica Asociación Española de Teledetección. p. 173-178.
- Cox, D.R. and Miller, H.D.:** 1968, *Theory of stochastic Processes*; Methuen & Co. Ltd. London, 397 p.
- Hsu, H.P. and Mehra, R.:** 1973, *Análisis de Fourier*. McGraw Hill, Bogotá, 274 p.
- Journel, A. and Huijbregts, Ch.:** 1978, *Mining geostatistics*; Academic Press, New York, 600 p.
- Matheron, G.:** 1973, *The intrinsic Random Functions and Their Applications*; *adv. appl. Prob.* v. 5, p. 439-468.
- Pardo-Igúzquiza, E.:** 1991. *Simulación Geoestadística de Variables Geológicas por Métodos Espectrales*. Tesis, Dpto. Geodinámica. Universidad de Granada. 412 p.
- Pardo-Igúzquiza, E. and Chica-Olmo, M. *The Fourier Integral Method: An Efficient Spectral Method For Simulation of Random Fields*. *Mathematical Geology*, Vol 25, nº 2, 1993. p. 177-218.