

SOFTWARE PARA EL CÁLCULO DEL SEMIVARIOGRAMA EXPERIMENTAL DE SERIES GRANDES DE DATOS

Vivas Miranda, José García y Paz González, Antonio

Facultad de Ciencias. Universidad de A Coruña. A Zapateira, 15.071. A Coruña.

SUMMARY

Software to calculate experimental semivariograms from large data sets

Geostatistics provides a set of statistical tools for incorporating the spatial coordinates of observations in data processing. Although choosing models for variograms and fitting them to data remain among the most difficult and controversial topics in geostatistics, nowadays, standard easy to use software allows to compute straightforward an experimental variogram. However, standard available software can only be used for current, limited and relatively small set of data. A computer program, SSEMI, written in C language, to calculate semivariograms from large data sets is presented. SSEMI is sufficiently versatile that it may be used to compute experimental variograms for big data series, which extent is only limited by computer memory. Additionally, a computer program, TREND, which allow remove a trend before semivariogram calculations are also presented here. Furthermore, this paper documents and discusses experience on the use of SSEMI and TREND to analyse the spatial dependence of an exhaustive data set of 82.830 soil microrelief point measurements. After trend removal, a very small nugget

effect and a short scale (range \cong 20 mm) of spatial variability were observed on the experimental semivariograms.

Key words: Geostatistics, C-language, semivariogram, exhaustive data set, soil microrelief

RESUMEN

La Geoestadística proporciona una serie de herramientas que permiten incorporar las coordenadas espaciales de los datos muestrales en el proceso de análisis de los mismos. En la actualidad, la elección de un modelo de semivariograma que se ajuste a los datos experimentales es una tarea difícil y de gran controversia en Geoestadística. Diversos programas fáciles de usar permiten calcular un semivariograma experimental. Sin embargo el software standard habitualmente disponible sólo puede ser utilizado partiendo de series de datos de entrada relativamente pequeñas. En este trabajo se presenta un programa, SSEMI, escrito en lenguaje C para el cálculo de semivariogramas a partir de series grandes de datos, cuyo tamaño viene limitado únicamente por la memoria del ordenador. También se presenta un programa, TREND, que permite retirar algunos tipos de tendencia, previo al cálculo del semivariograma experimental. Asimismo, se lleva a cabo el estudio y la discusión de un caso en el que se usan los programas SSEMI y TREND para el análisis de la dependencia espacial de una serie exhaustiva constituida por 82.830 datos puntuales de microrrelieve. Después de filtrada una tendencia lineal, se calculó el semivariograma experimental, observándose un efecto pepita prácticamente nulo y una estructura espacial con un rango pequeño (\cong 20 mm).

Palabras clave: Geoestadística, lenguaje C, semivariograma, series exhaustivas de datos, microrrelieve del suelo

INTRODUCCIÓN

Desde comienzos del siglo actual la variabilidad espacial de las propiedades del suelo ha preocupado a numerosos investigadores. MONTGOMERY (1913) llevó a cabo un estudio sobre el efecto del nitrógeno en el rendimiento del trigo a partir de datos obtenidos en 224 parcelas experimentales. Otros autores, como ROBINSON and LLOYD (1915) y PENDLETON (1919) estudiaron los errores de muestreo y las diferencias entre suelos del mismo grupo; WAYNICK (1918) analizó la variabilidad espacial de la nitrificación en el suelo; WAYNICK and SHARP (1919) estudiaron el contenido total en carbono y nitrógeno, a partir de una gran cantidad de muestras tomadas con los más distintos esquemas de muestreo, pero siempre con la preocupación de caracterizar y conocer la variabilidad espacial. Intentando unificar el método de análisis cuando se dispone de una gran cantidad de datos, HARRIS (1920) utilizó una ecuación muy próxima a la que hoy conocemos para describir la varianza mediante el krigado por bloques. Todo ello pone en evidencia la importancia concedida al estudio de la variabilidad espacial del suelo, con mucha antelación sobre el desarrollo de la geoestadística en la década de los años cincuenta (KRIGE, 1951) y sesenta MATHERON (1963, 1970)

Sin embargo, estos primeros tanteos para encontrar un método de evaluación de la variabilidad espacial, no tuvieron continuidad en los años posteriores. Por contra, se adoptaron otras técnicas, basadas en la casualización y la repetición, que, junto a los avances en el conocimiento de las funciones de distribución, llevaron a la adopción del muestreo aleatorio, lo que supuso el olvido o la minusvaloración de las coordenadas geográficas de cada uno de los puntos de una muestra compuesta. Este procedimiento, unido a la hipótesis de la distribución normal de frecuencias, ha sido y todavía es usado para asumir la independencia entre muestras y, en definitiva, permite garantizar la validez del uso de la media y de la desviación típica para representar el fenómeno estudiado, según los conceptos de la estadística clásica.

Por otra parte, el hecho de que la variable objeto de estudio presente una distribución normal no garantiza la independencia entre muestras, que únicamente puede ser verificada mediante un análisis de la autocorrelación entre las mismas. Esto es debido a que en el cálculo de la distribución de frecuencias no se considera la distancia entre las muestras tomadas en el campo. La presencia de esta dependencia espacial requiere la utilización de un tipo particular de estadística denominada Geoestadística, que permite estimar la continuidad o dependencia espacial mediante una función denominada semivariograma.

El software desarrollado para el cálculo de la dependencia espacial es muy abundante y, en general, fácilmente accesible. En efecto, ya en los primeros tiempos de la geoestadística la revista "Computer Contributions", publicada por el organismo Kansas Geological Survey presentaba regularmente no solo contribuciones teóricas sino también listados de programas completos para el cálculo de la variabilidad espacial. Posteriormente, "Computers and Geosciences" tomó el relevo, publicando en los últimos 20 años muchos programas de utilidad para el análisis geoestadístico, incluyendo algunos específicos para el estudio de las propiedades del suelo (WEBSTER and OLIVER, 1997). Entre los programas que actualmente se encuentran en la bibliografía para el cálculo de la dependencia espacial, cabe mencionar los publicados por VIEIRA (1983), ENGLUND and SPARKS (1991), DEUTSCH and JOURNEL (1992), PANNATIER (1996) y PEBESMA (1997).

Los principales objetivos de la estadística han sido predecir el patrón de variación espacial e interpolar el valor de una variable en una posición determinada, en la que no ha sido muestreada, en áreas tan diversas como son Minería, Oceanografía, Edafología, Climatología, Agricultura, Medio Ambiente, etc. Más recientemente, se ha desarrollado la simulación condicional que permite modelizar la incertidumbre.

Con frecuencia los análisis geostatísticos se llevan a cabo a partir de un número limitado, relativamente pequeño de datos muestrales. A título de ejemplo, PAZ *et al.* (1996) estudiaron la variabilidad de las propiedades generales del suelo a escala de parcela a partir de 53 muestras. Uno de los estudios más ampliamente difundidos durante los últimos años (ATTEIA *et al.*, 1994; GOOVAERTS, 1997) analizó el contenido en metales pesados de un área de 14,5 km² del Jura suizo a partir de 359 datos muestrales, siendo considerada dicha muestra un conjunto de datos exhaustivo y representativo de los datos poblacionales en el área estudiada. Por esta razón, muchos de los programas standard habitualmente utilizados, entre ellos los citados anteriormente, admiten un número limitado de datos de entrada.

Con la puesta a punto de aparatos de medida automatizados, como el equipo de rugosimetría láser, que puede medir datos a intervalos de 2 mm, y que genera cerca de 100.000 valores en 1 m² de superficie (BERTUZZI *et al.*, 1990), la capacidad de análisis de los estos programas habituales se ha tornado demasiado pequeña. De este modo se hizo necesario elaborar software específico en que el número de valores máximo analizados sea dependiente solamente de la memoria disponible en el ordenador.

Por esta razón se ha creado software específico para el análisis estructural y el cálculo del semivariograma muestral de grandes series de datos. Es necesario significar que cuando la densidad de datos experimentales es muy alta, la interpolación puede ser innecesaria; sin embargo el análisis de la dependencia espacial siempre aportará elementos para caracterizar la estructura de la propiedad estudiada.

En este trabajo se presentan los programas SSEMI, que calcula el semivariograma muestral, y TREND, que filtra una tendencia de tipo lineal, parabólico o cúbico. Estos programas pueden utilizarse para analizar series grandes de datos, de tamaño no predefinido, que viene limitado solamente por la memoria disponible del ordenador, ya que se escribieron usando algoritmos rápidos.

FUNDAMENTOS TEÓRICOS

De acuerdo con MATHERON (1963) la Geostadística se puede definir en términos descriptivos como la aplicación de la teoría de las variables regionalizadas, es decir de aquellas variables que presentan estructura espacial, a la estimación. Desde un punto de vista matemático, una variable regionalizada es simplemente una función $Z(x)$ que adopta un valor para cada punto x del espacio. Con frecuencia, en una variable regionalizada se observan dos aspectos complementarios y aparentemente contradictorios: un aspecto aleatorio, asociado con variaciones erráticas, impredecibles y un aspecto estructurado, que refleja la variabilidad espacial del fenómeno estudiado (SAMPER y CARRERA, 1990).

Una variable regionalizada puede interpretarse en términos probabilísticos como una realización de una función aleatoria, $Z(x)$. Este concepto, considerado por JOURNAL (1986) como la piedra angular en Geoestadística hay que entenderlo más como un modelo que como un concepto con significación física. En todo caso, para hacer posible la inferencia estadística resulta imprescindible introducir hipótesis adicionales acerca de $Z(x)$.

1) *Hipótesis de estacionariedad*. Se dice que una función es estacionaria de segundo orden (con frecuencia se omite este calificativo), si:

- La esperanza matemática, $E [Z(x)]$, existe y no depende de la posición de x , es decir, si m es el momento de primer orden se tiene que:

$$E [Z(x)] = m \quad (1)$$

- para toda pareja de variables aleatorias, $[Z(x), Z(x+h)]$, su covarianza existe y solo depende del valor del vector separación \mathbf{h} , es decir:

$$C(\mathbf{h}) = E[Z(x), Z(x+h)] - m^2 \quad (2)$$

en donde $C(\mathbf{h})$ es la covarianza.

Para una función aleatoria estacionaria de segundo orden existen la varianza y la covarianza y ambos momentos de segundo orden pueden ser útiles para caracterizar la variabilidad espacial.

2) *Hipótesis intrínseca*. Esta hipótesis, menos restrictiva que la anterior, es la más empleada en Geoestadística. Las funciones aleatorias intrínsecas describen fenómenos que muestran una capacidad casi ilimitada de variación, de modo que se pueden definir como aquellas cuya varianza no existe y sin embargo sus incrementos $[Z(x) - Z(x+h)]$ tienen esperanza matemática definida y varianza finita e independientes de \mathbf{x} para todo vector \mathbf{h} . Por lo tanto, en estas funciones se cumple que:

$$E[Z(x) - Z(x+h)] = m(\mathbf{h}) \quad (3)$$

$$\text{Var}[Z(x) - Z(x+h)] = g(\mathbf{h}) \quad (4)$$

Aunque no es indispensable, habitualmente se supone que $m(\mathbf{h}) = 0$

Uno de los métodos más antiguos para estimar la dependencia espacial o temporal de muestras vecinas, que tiene su origen en el análisis de series temporales es el análisis del semivariograma.

De acuerdo con las hipótesis anteriormente discutidas, una serie de observaciones efectuadas a determinada escala puede presentar una estructura espacial aleatoria (heterogeneidad estocástica) y/o variar de un modo sistemático (heterogeneidad determinística). En este último caso, es susceptible de ser descrita por un modelo llamado semivariograma. El semivariograma es uno de los momentos de segundo orden considerados en Geoestadística. Por definición la ecuación teórica del semivariograma se expresa como:

$$\gamma(\mathbf{h}) = 1/2 E[Z(x_i) - Z(x_i+h)]^2 \quad (5)$$

en donde $[Z(x_i), Z(x_i+h)]$ son pares de valores separados por un vector \mathbf{h} . Los valores de x_i e x_i+h son definidos de acuerdo con las posiciones de los datos muestrales.

El primer paso para el análisis geoestadístico consiste en verificar si el sistema de muestreo utilizado ha permitido o no poner en evidencia la existencia de dependencia espacial. Para ello se estima el semivariograma experimental, $\gamma^*(h)$, mediante la ecuación:

$$\gamma^*(h) = \frac{1}{2} \frac{1}{N(h)} \sum_{i=1}^{N(h)} [Z(x_i) - Z(x_i+h)]^2 \quad (6)$$

en donde $N(h)$ representa el número de pares de valores medidos.

Dependencia espacial, significa, en otras palabras, autocorrelación, es decir, la dependencia del valor de una variable en un punto del de sus vecinos. Esta característica está expresada en la ecuación (2) como la diferencia $[Z(x_i) - Z(x_i+h)]$. Asumiendo que la variación es independiente de la dirección, se puede utilizar en los cálculos el módulo del vector \mathbf{h} , que equivale a la distancia de separación entre muestras vecinas.

Una vez calculado el semivariograma se dispone de pares de valores de semivarianza $\gamma(h)$ y de distancias, h , que se representan gráficamente tomando como ordenadas los valores de la semivarianza y como abscisas las distancias. Al ajustar una ecuación a este gráfico se obtiene un modelo de dependencia espacial. Para propiedades que presentan dependencia espacial se espera que la diferencia entre los valores $Z(x_i) - Z(x_i+h)$ crezca con la distancia hasta un punto determinado a partir del cual se estabiliza, con un valor denominado meseta, representado por el símbolo $C_0 + C_1$, aproximadamente igual a la varianza de los datos. Esta distancia recibe el nombre de alcance, \mathbf{a} , y representa el radio de un círculo, dentro del cual los valores de la propiedad estudiada son tan similares unos con otros que estén correlacionados. El valor de la semivarianza en la intersección del eje de ordenadas se denomina efecto pepita; su símbolo es C_0 , y representa la variabilidad de la propiedad estudiada para distancias inferiores a la que separa a muestras vecinas. Así, cuanto mayor es el efecto pepita, menor es la dependencia espacial de una propiedad. Los parámetros efecto pepita (C_0), meseta ($C_0 + C_1$) y alcance, \mathbf{a} se usan en las ecuaciones que describen mediante modelos teóricos los semivariogramas; este procedimiento se discute ampliamente por VIEIRA *et al.* (1983).

Identificación y filtrado de una tendencia

En ciertos casos, la variable regionalizada no presenta estacionariedad y el semivariograma experimental no alcanza una meseta estable. Esto es consecuencia de que la función aleatoria presenta una deriva, es decir, su esperanza matemática no es constante. Si,

además, los incrementos de primer orden $[Z(x_i)-Z(x_i+h)]$ tampoco son estacionarios, se dice que dicha función aleatoria no es intrínseca. En estas circunstancias, en algunos casos, $Z(x)$ se puede expresar como la suma de una componente determinística, $m(x)$, y un residuo, $R(x)$ estacionario, con media nula y semivariograma $\gamma_R(h)$, es decir, se puede identificar una tendencia. La hipótesis de tendencia permite separar dos componentes de la función aleatoria:

$$Z(x) = m(x) + R(x) \quad (7)$$

donde $m(x)$ es la tendencia o deriva $\gamma R(x)$ es el residuo. La tendencia dependerá exclusivamente de la posición del punto, con lo cual la componente aleatoria queda plasmada mediante el residuo.

La variable regionalizada en este caso es una función de dos variables, y la tendencia tendrá que ser fijada con el fin de restarla a la variable inicial y así obtener el residuo, $R(x)$.

Las expresiones más habituales para filtrar una tendencia son de tipo lineal, parabólico y cúbico:

$$m(x,y) = A_0 + A_1x + A_2y + A_3xy \quad (8)$$

$$m(x,y) = A_0 + A_1x + A_2y + A_3x^2 + A_4y^2 + A_5xy \quad (9)$$

$$m(x,y) = A_0 + A_1x + A_2y + A_3x^2 + A_4y^2 + A_5xy + A_6x^3 + A_7y^3 + A_8x^2y + A_9xy^2 \quad (10)$$

Lógicamente, se pueden usar otras expresiones, sobre todo cuando existe un buen conocimiento del fenómeno a estudiar.

En general, la elección del modelo de tendencia a filtrar se lleva a cabo ajustando por mínimos cuadrados cada uno de los modelos considerados y eligiendo aquel que proporcione una correlación mas alta. Además, el semivariograma de los residuos debe mostrar una estructura espacial más fuerte que cuando no se considera tendencia. El siguiente paso consiste en restar a la medida efectuada en cada punto su tendencia, para obtener los residuos.

RESULTADOS Y DISCUSIÓN

Descripción de los Programas

Programa SSEMI, para el cálculo del semivariograma

SSEMI, escrito en lenguaje C, listado en el anexo I, se elaboró partiendo del programa AVAR escrito en FORTRAN por VIEIRA (1983): La diferencia entre ambos es que el nuevo programa permite el cálculo del semivariograma experimental, tratando un número mayor de datos de entrada.

Para que el programa sea operativo es necesario que calcule la semivarianza que corresponde a sucesivos intervalos de distancia (lags) entre pares de puntos y que, además proporcione el número de pares de puntos, N , usados para efectuar los cálculos en cada intervalo Δh .

La base del programa para el cálculo del semivariograma muestral es la ecuación (6). El algoritmo calcula, para cada rango de distancias considerado, h_i , la semivarianza de todos los pares de puntos de la superficie separados por una distancia comprendida dentro de dicho rango.

El intervalo de distancia, Δh , es un dato de entrada cuyo valor puede ser introducido a voluntad. De éste dependerá la precisión del valor calculado de la semivarianza en el intervalo considerado, el número de puntos del mismo y, en consecuencia el gráfico $\gamma^*(h)$. En definitiva, Δh es un parámetro del programa y puede ser considerado como el valor de la ventana de una media móvil sobre el semivariograma. El efecto de un incremento en el valor de Δh es de suavización de la curva $\gamma^*(h)$.

El valor de la semivarianza calculado en cada rango de distancia es acumulando en una lista que tiene como índice el valor medio de dicho rango; en dicha lista también figura el número de pares de puntos usados para efectuar los cálculos en cada intervalo. Este listado de salida, por tanto, contiene los valores muestrales del semivariograma, apreciándose que se trata de una serie de valores discretos.

Para ejecutar el programa SSEMI, por lo tanto, es necesario considerar solamente cuatro parámetros, de acuerdo con:

SSEMI {*archivo de entrada*} {*archivo de salida*} {*paso*} {*escala final*}

en donde,

archivo de entrada: es el archivo que contiene los datos de entrada, en formato ASCII, con 3 columnas (X, Y y Z) y un espacio entre ellas, como se puede apreciar en la tabla 1. Las dos primeras columnas contienen información sobre las coordenadas de cada posición y en la tercera se consigna el valor correspondiente de la propiedad estudiada

1.00	1.00	-0.041
2.00	1.00	0.064
3.00	1.00	-0.057
4.00	1.00	0.012
5.00	1.00	0.032
6.00	1.00	0.013
7.00	1.00	-0.017
...

Tabla 1. Ejemplo de la estructura de un archivo de entrada del programa SSEMI.

archivo de salida: archivo donde estarán los resultados del semivariograma, en formato ASCII con tres columnas. La primera es la distancia calculada desde el origen hasta el punto medio de un intervalo determinado, en la segunda se consigna el valor de la semivarianza y en la tercera el número de pares de puntos vecinos utilizados para los cálculos. En la tabla 2 se presenta un ejemplo de un archivo de salida. Nótese el elevado número de pares de puntos vecinos con que se efectúan los cálculos de la semivarianza para los sucesivos intervalos de distancia.

paso: valor del intervalo de distancia Δh .

escala final: la distancia máxima contada desde el origen a que se refieren los cálculos.

1.590864	2.451384e+00	99646
2.462831	4.190657e+00	260372
3.498359	5.529357e+00	216283
4.445881	6.899950e+00	494181
5.619194	8.285815e+00	366787
6.431048	9.291600e+00	627768
7.575309	1.042709e+01	571552
...

Tabla 2. Ejemplo de la estructura de un archivo de salida del programa SSEMI.

En síntesis, el programa SSEMI lee el archivo de la superficie (archivo de entrada), hace los cálculos y sitúa los valores de distancia, h , y semivarianza, $\gamma^*(h)$, en el archivo “*archivo de salida*”.

Programa TREND para el filtrado de tendencia

El programa TREND se escribió en lenguaje C, listado en el anexo II, a partir de otro del mismo nombre debido a VIEIRA (1983) y escrito en FORTRAN. Este programa resuelve las ecuaciones (8), (9) y (10) que permiten obtener una tendencia lineal, parabólica o cúbica respectivamente. Una vez ajustados los parámetros de estas tres expresiones por mínimos cuadrados, si se utilizan criterios puramente estadísticos, se elige aquella que proporcione una correlación más alta; sin embargo en ocasiones es preferible considerar otros criterios de retirada de tendencia, sobre todo si se conocen las leyes físicas que rigen el fenómeno estudiado.

Para ejecutar el programa TREND es necesario considerar solamente tres parámetros, de acuerdo con:

TREND {*archivo de entrada*} {*extensión del archivo de salida*} {*orden*}

en donde,

archivo de entrada: archivo que contiene los datos de la superficie, en formato ASCII, con 3 columnas (X, Y y Z) y con un espacio entre ellas. El formato y el contenido son similares a los del archivo de entrada del programa SSEMI (Tabla 1).

extensión del archivo de salida: Extensión del archivo con el mismo nombre del archivo de salida, donde se presentan tanto datos iniciales como los resultados, una vez ejecutado el programa, en formato ASCII con cinco columnas. Las tres primeras columnas de este fichero son una copia de la superficie original, o sea las columnas X, Y y Z; en la cuarta se presenta el valor de la función que se haya filtrado (lineal, parabólica o cúbica) a partir de la superficie original y en la quinta la superficie residual, que, como es lógico, se obtiene como resultado de la sustracción entre la tercera y cuarta columna. Un ejemplo de fichero de salida se presenta en esquema en la Tabla 3.

orden: número de orden de la ecuación que mejor se ajusta para el filtrado de una tendencia. 1, para ajustar una superficie lineal, 2 para una parabólica y 3 para una cúbica.

1.00	1.00	-0.041	0.400	-0.441
2.00	1.00	0.064	0.730	-0.666
3.00	1.00	-0.057	0.349	-0.406
4.00	1.00	0.012	0.349	-0.337
5.00	1.00	0.032	0.734	-0.702
...

Tabla 3. Ejemplo de la estructura de un archivo de salida del programa TREND.

El programa TREND, una vez ejecutado, imprimirá en pantalla los coeficientes de la función considerada para retirar la tendencia, es decir una de las tres expresiones (8), (9) o (10), así como los resultados del análisis estadístico.

En el archivo de salida los coeficientes están ordenados de acuerdo con la secuencia $A_0, A_1, A_2...$ Las variables estadísticas de la función ajustada que se imprimen en el mismo archivo son: coeficiente de regresión, desviación estándar, etc ...

ESTUDIO DE UN CASO

A título de ejemplo se presenta a continuación la aplicación de los programas SSEMI y TREND al análisis de la estructura de la variabilidad espacial del microrrelieve de la superficie de un suelo agrícola, a partir de una red de datos puntuales medidos con rugosímetro láser de alta resolución con un paso de medida de 2 mm x 2 mm.

Los datos experimentales se obtuvieron en laboratorio, sobre una superficie similar a la de un lecho de siembra, preparada artificialmente con una mezcla de agregados naturales de distinto calibre, en proporciones que intentan reconstruir la estructura de la superficie del suelo durante la siembra de un cultivo de cereales. Los agregados de mayores dimensiones se dispusieron aleatoriamente.

El microrrelieve se midió en la superficie inicial y otras seis veces después de la aplicación mediante simulador de las siguientes cantidades de lluvia artificial: 10, 55, 85, 130, 190 y 250 mm de lluvia producida mediante simulador.

En cada superficie experimental, de 58 cm x 58 cm, se midieron 270 perfiles separados entre sí 2 mm; cada perfil, a su vez, constaba de más de 330 puntos de medida individuales y cada punto distaba de sus vecinos menos de 2 mm. Para obviar posibles efectos de borde, no se tuvieron en cuenta los puntos más externos de la superficie

experimental. En suma, los ficheros de entrada cuya estructura espacial se pretende analizar constan de 82830 datos puntuales.

Los semivariogramas muestrales obtenidos a partir de los datos medidos con rugosímetro láser fueron estudiados por PAZ *et al.* (1998), quienes pusieron en evidencia la existencia de una dependencia espacial muy importante tanto en la superficie inicial como en las resultantes de la aplicación de cantidades crecientes de precipitación, siendo prácticamente nulas las discontinuidades en el origen. Se comprobó que los datos experimentales en todos los casos considerados estaban sesgados, tendiendo a incrementarse el valor de la cota medida con la distancia al origen. Una vez verificada la presencia de una tendencia de tipo lineal y filtrada la misma, los semivariogramas muestrales de los residuos alcanzaban una meseta estable.

Ejemplos de semivariogramas muestrales obtenidos mediante el programa SSEMI, antes y después del filtrado de una tendencia lineal mediante el programa TREND se presentan en las figuras 1 y 2. Los datos de la figura 1 corresponden a la superficie inicial y los de la figura 2 a la última de la serie de superficies escaneadas mediante rugosímetro láser, después de la aplicación de 250 mm de precipitación que degradaron el microrelieve original. En las dos figuras se puede comprobar la presencia de una deriva de tipo lineal. En efecto, se observa que en los semivariogramas calculados directamente a partir de los datos puntuales medidos con rugosímetro no se alcanza una meseta estable, lo que si ocurre con los semivariogramas muestrales de las superficies residuales, obtenidas después de retirada una tendencia lineal. También se puede observar que, en el caso estudiado, el filtrado de una tendencia lineal no modifica el valor de la discontinuidad en el origen. En suma, de la observación de los semivariogramas muestrales se infiere que, en las superficies estudiadas, el rugosímetro láser ha detectado la dependencia espacial del microrrelieve del suelo, y de los parámetros asociados al mismo, como la rugosidad superficial, a escala milimétrica.

Por otra parte, en la tabla 1 se presentan los resultados de los parámetros obtenidos mediante el programa TREND tanto en la superficie inicial como en las sucesivas superficies escaneadas con rugosímetro láser después de aplicar la precipitación que se indica. En todos los casos estudiados la tendencia encontrada, mediante un ajuste simple por mínimos cuadrados, es de tipo lineal y viene definida por una ecuación es de la forma:

$$Z(x,y) = A_0 + A_1x + A_2y$$

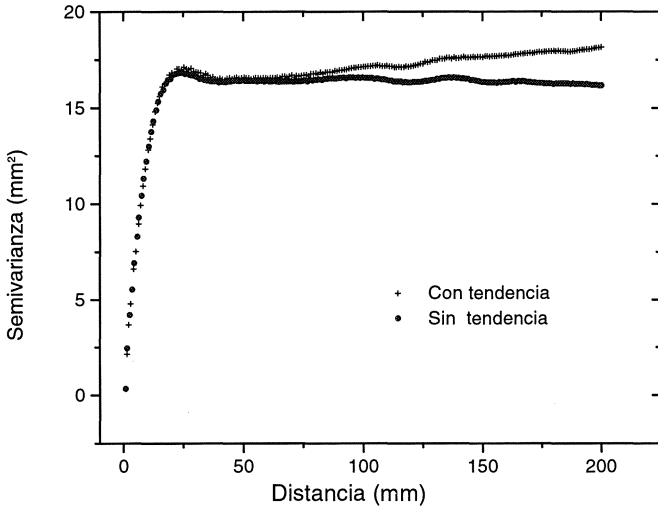


Figura 1. Semivariogramas experimentales de la superficie original antes y después de filtrada una tendencia lineal

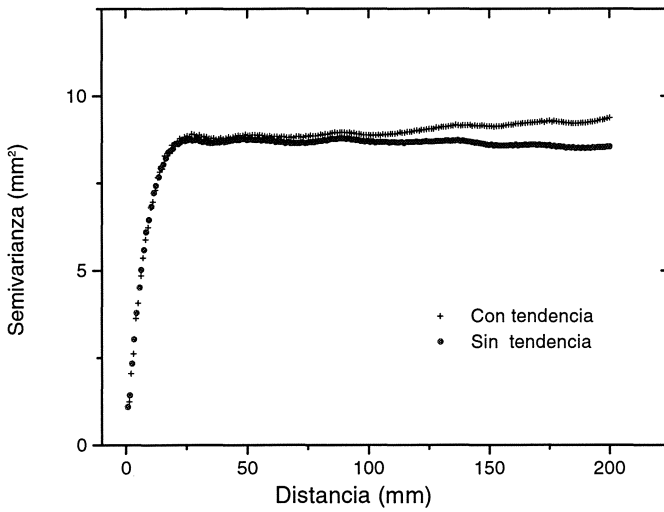


Figura 2. Semivariogramas experimentales una superficie degradada por la acción de 250 mm de precipitación antes y después de filtrada una tendencia lineal

En trabajos anteriores (PAZ y TABOADA, 1996) se había asumido, a falta de mejor información y herramientas de análisis, que en la superficie experimental los agregados estaban dispuestos al azar y, por lo tanto, dicha superficie carecía de componente orientada. Sin embargo, con el programa TREND, se comprueba la presencia una inclinación, (si bien la magnitud de la pendiente calculada es pequeña), lo que pone de manifiesto la dificultad de preparar una superficie experimental en la que el microrrelieve presente exclusivamente una componente aleatoria.

Se constata también que cada uno de los parámetros, A_0 , A_1 y A_2 presentan, en el conjunto de las siete superficies estudiadas, el mismo signo, y que el valor de los mismos no es muy diferente, independientemente de la cantidad de agua de lluvia aplicada. De este modo, la ordenada en el origen, A_0 , oscila entre 14,3319 y 28,7520 mm, la pendiente A_1 entre - 0,00517 y - 0,00974 y la pendiente A_2 entre 0,00370 y 0,00512. Estos resultados indican que la inclinación detectada se originó al construir la superficie inicial (Tabla 4).

Superficie	Precipitación (mm)	A_0	A_1	A_2
E4T1	0	19,8845	-0,00974	0,00406
E4T2	10	28,7520	-0,00743	0,00370
E4T3	55	25,8089	-0,00747	0,00421
E4T4	85	14,3319	-0,00701	0,00395
E4T5	130	14,4773	-0,00708	0,00440
E4T6	190	17,6557	-0,00617	0,00512
E4T7	250	27,0175	-0,00517	0,00451

Tabla 4. Parámetros de la tendencia calculada por regresión simple a partir de los datos puntuales de microrrelieve en sucesivas superficies.

Dada la gran densidad de puntos muestrales en el ejemplo presentado carece de sentido la interpolación geoestadística. Sin embargo, del análisis del semivariograma muestral se pueden inferir resultados importantes acerca de la variabilidad espacial del microrrelieve. Así, en las figuras 1 y 2 se comprueba como el valor de la meseta disminuye conforme evoluciona la superficie del suelo, es decir, al aumentar la precipitación. Por otra parte, el valor de la discontinuidad en el origen de los semivariogramas experimentales es prácticamente nulo. La relación entre la magnitud del efecto pepita y la varianza muestral es una medida del grado de aleatoriedad de los datos experimentales situados a distancias próximas entre sí y, dada la pequeña proporción que supone la

discontinuidad en el origen en relación con la varianza total, se pone en evidencia que existe una gran similitud entre los valores de puntos vecinos.

De los datos consignados en las figuras 1 y 2, hay que destacar, asimismo, que el alcance de la dependencia espacial se mantiene prácticamente constante a lo largo de la experiencia, sin sufrir una gran oscilación, conforme evoluciona la superficie en función de la precipitación acumulada. El alcance de los semivariogramas muestrales, una vez retirada la tendencia lineal, osciló entre 16 y 18 mm, cifra del mismo orden de magnitud que el diámetro de los agregados de mayores dimensiones utilizados para preparar el lecho de siembra artificial. Con ello, se pone en evidencia la importancia de los elementos estructurales de mayor diámetro en la configuración de las características físicas del microrrelieve del suelo.

CONCLUSIONES

Tomando como base software standard utilizado habitualmente en geoestadística, se escribieron programas en lenguaje C que, con respecto a los utilizados previamente, presentan la ventaja de que permiten el estudio del semivariograma muestral de grandes series de datos. SSEMI calcula la semivarianza experimental en función de la distancia, para intervalos predeterminados. TREND filtra los tipos de tendencia más frecuentemente encontrados

Los dos programas actualizados se utilizaron para el estudio del microrrelieve del suelo caracterizado experimentalmente mediante rugosímetro laser, lo que genera ficheros con un gran número de datos de altura, cada uno con sus respectivas coordenadas. SSEMI permitió calcular los semivariogramas muestrales apreciándose, en primer lugar, una fuerte dependencia espacial y, en segundo lugar, que la magnitud de la discontinuidad en el origen es tan pequeña que, en la práctica, se puede admitir, que los datos obtenidos describen exhaustivamente el microrrelieve. TREND permitió filtrar una tendencia lineal a los datos originales y obtener superficies residuales cuyos semivariogramas muestrales presentaban una meseta estable.

AGRADECIMIENTOS: Este trabajo se llevo a cabo en el marco del proyecto FAIR 1 CT95-0458, financiado por la Unión Europea.

BIBLIOGRAFÍA

- ATTEIA, O., DUBOIS, J. P. and WERSTER, R. (1994). Geostatistical analysis of soil contamination in the Swiss Jura. *Environmental Pollution*, 80:315-327
- BERTUZZI, P., CAUSSGNAC, J. M., STENGEL, P., MOREL, G., LORENDEAU, J. Y. and PELLOUX, G. (1990). An automated, noncontact laser profile meter for measuring soil roughness in situ. *Soil Sci.*, 149:196-178.
- DEUTSCH, C. V. and JOURNEL, A. G. (1992). GSLIB. Geostatistical software library and user's guide. Oxford University Press. Nueva York.
- ENGLUND, E. and SPARKS, A. (1991). GEO-EAS. Geostatistical Environmental Assessment Software. User's guide. ES-EPA. Las Vegas.NE. USA.
- GOOVAERTS, P. (1997). Geoestistics for natural resources evaluation. Oxford University Press. 482 pp.
- HARRIS, J.A. (1920). Practical universality of field heterogeneity as a factor influencing plot yields. *Journal Agr. Res.*, XIX(7): 279-314.
- JOURNEL, A. G. (1986). Geostatistics. Models and tools for the earth sciences. *Geol.*, 18:119-140.
- KRIGE, D. G. (1951). A statistical approach to some basic mine evaluation problems on the Witwatersrand. *J. Chem. Metall. Min. Soc. S. Afri.*, 52: 119-139.
- MATHERON, G. (1963). Principles of Geostatistics. *Economic Geology*, 58: 1246-1266.
- MATHERON, G. (1970). La Théorie des variables Régionalisées et ses Applications. *Les Cahiers du centre de Morphologie Mathématique*. Fascicule 5. Ecole des Mines de Paris. Fontainebleau. 212 pp.
- MONTGOMERY, E.G. (1913). Experiments in wheat breeding: experimental error in the nursery and variation in nitrogen and yield. U.S. Dept. Agric. Bur. Plant Indust. Bul., Washington, 269, 61 pp.
- PANNATIER, Y. (1996). VARIOWIN. Software for Spatial Data Analysis in 2D. Springer-Verlag, New York. 91 pp.

- PAZ, A., GÓMEZ, M^a. J. y TABOADA, M^a. T. (1996). Análisis geoestadístico de las propiedades generales de un suelo de cultivo. *Investigación Agraria. Producción y Protección Vegetales*, 11: 133-160.
- PAZ, A., VIVAS MIRANDA, J. G. y BERTOLANI, F. C. (1998). Caracterización geoestadística del microrrelieve del suelo a partir de datos obtenidos con rugosímetro láser. *IV Congreso Nacional del medio ambiente, Espacios Naturales, Biodiversidad y Agua*. Madrid. 144-158
- PEBESMA, E. J. (1997). *GSTAT User's Manual: Netherlands Centre for Geo-ecological research*. Faculty of environmental Sciences. University of Amsterdam. 81 pp.
- PENDLETON, R. L. (1919). Are soils mapped under a given type name by the Bureau of Soils method closely similar to one another?. *Agricultural Sciences, Davis*, 3: 369-498.
- ROBINSON, G. W. and LLOYD, W. E. (1915). On the probable error of sampling in soil surveys. *The J. Agric. Sci., Cambridge*, 7: 144-153.
- SAMPER, F. J. y CARRERA, J. (1990). *Geoestadística. Aplicaciones a la hidrología subterránea*. Centro Internacional de Métodos Numéricos en Ingeniería. Barcelona. 484 pp.
- WAYNICK, D. D. (1918). Variability in soils and its significance to past and future soil investigations. I. Statistical study of nitrification in soils. *Agricultural Sciences, Davis*, 3: 243-270. *IV Congreso Nacional del Medio Ambiente*. 145-158.
- WAYNICK, D. D. and SHARP, L. T. (1919). Variability in soils and its significance to past and future soil investigations. II Variation in nitrogen and carbon in field soils and their relation to the accuracy of field trials. *Agricultural Sciences, Davis* 4 : 121-139.
- WEBSTER, R. and OLIVER, M. A. (1992). Software review. *European J. of Soil Sci.*, 48: 173-175.
- VIEIRA, S. R., HATFIELD, J. L., NIELSEN, D. R., BIGGAR, J. W. (1983). Geostatistical theory and application to variability of some agronomical properties. *Hilgardia*. Berkeley. 51: 1-75.

ANEXO I

**LISTADO DEL PROGRAMA "SSEMI"
PARA EL CÁLCULO DEL SEMIVARIOGRAMA MUESTRAL**

SSEMI.C

/* PROGRAM SSEMI. Traducion in C language from a program to calculate a sample semi-variogram, written by Vieira (1983). Calculates the variance for different scales sizes in a surface described in an ASCII format, with three columns (X, Y and Z) */

```
#include <malloc.h>
#include <math.h>
#include <stdio.h>
void help();
void SemiH(double *, double *,double *, double *, double *,long *, double ,long ,FILE *);
struct Hstc Fiting(FILE *);
double *dvector(long);
long *lvector(long);
void free_dvector(double *);
void free_lvector(long *);
void nrerror(char *);
void Initi(double *,double *,long *);
int Nlag;
double Nstep;
void main(unsigned argc, char **argv)
{
FILE *Fin,*Fout,*Ftmp;
long N,i;
double lx,EscF;
struct Hstc H;
double *x;
double *y;
double *z;
double *vGam;
double *vDis;
long *vCnt;
/***** File Read *****/
```

```

if(argc!=5)
{
help();
exit(0);
}
Fin=fopen(argv[1],"r");
Fout=fopen("soilsv.tmp","a+");
Ftmp=fopen(argv[2],"w+");
Nlag=atoi(argv[3]);
EscF=atof(argv[4]);
Nstep=EscF/(double)(Nlag-1);
/* Fix the size of the file */
N=0;
do
{
N++;
fscanf (Fin, "%lf %lf %lf\n", &lx, &lx, &lx);
}
while (feof(Fin)==0);
x=dvector(N);
y=dvector(N);
z=dvector(N);
vGam=dvector(Nlag+1);
vDis=dvector(Nlag+1);
vCnt=lvector(Nlag+1);
/* Load data */
fseek(Fin,(long)0,SEEK_SET);
for(i=0;i<N;i++){
fscanf (Fin, "%lf %lf %lf\n", &x[i],&y[i], &z[i]);
}
Initi(vGam,vDis,vCnt);
SemiH(x,y,z,vGam,vDis,vCnt,EscF,N,Ftmp);
free_dvector(x);
free_dvector(y);

```

```

free_dvector(z);
free_dvector(vGam);
free_dvector(vDis);
free_lvector(vCnt);
fclose(Fin);
fclose(Fout);
fclose(Ftmp);
}

void SemiH(double *x, double *y, double *z, double *vGam, double *vDis, long *vCnt, double
EscF, long N, FILE *Fsai)
{
long i, j, idx;
double dy, dx, dz, dhs;
for(i=0; i<(N-1); i++)
{
for(j=i+1; j<N; j++)
{
dx=x[j]-x[i];
dy=y[j]-y[i];
dhs=sqrt(dx*dx+dy*dy);
if(dhs<=(EscF+Nstep))
{
dz=z[i]-z[j];
idx=(int)(dhs/Nstep);
vCnt[idx]++;
vDis[idx]+=dhs;
vGam[idx]+=(dz*dz);
}
}
}
for(i=0; i<Nlag; i++)
{
long cnt;
cnt=vCnt[i];

```

```

if(cnt)
fprintf(Fsai,"%f %le %ld\n",vDis[i]/cnt,vGam[i]/(2*cnt),cnt);
}
}
void Initi(double *vGam,double *vDis,long *vCnt)
{
int i;
for(i=0;i<=Nlag;i++)
{
vGam[i]=vDis[i]=0.0;
vCnt[i]=0L;
}
}
double *dvector(long nh)
{
double *v;
unsigned long tam;
tam=(nh+1)*sizeof(double);
v=(double *)malloc((unsigned long) tam);
if (!v) perror("allocation failure in dvector()");
return v;
}
long *lvector(long nh)
{
long *v;
unsigned long tam;
tam=(nh+1)*sizeof(long);
v=(long *)malloc((unsigned long) tam);
if (!v) perror("allocation failure in dvector()");
return v;
}
void perror(char error_text[])
{
void _exit();

```

```
char c;
fprintf(stderr,"Run-time error...\n");
fprintf(stderr,"%s\n",error_text);
fprintf(stderr,"...now exiting to system...\n");
_exit(0);
}
void help()
{
fprintf(stderr,"usage: Ssemi {input file} {output file} {Num. Lags} {fin.scale}\n");
}
void free_dvector(double *v)
{
free(v);
}
void free_lvector(long *v)
{
free(v);
}
```

ANEXO II

**LISTADO DEL PROGRAMA "TREND"
PARA EL CÁLCULO DEL FILTRADO DE TENDENCIA.**

TREND.C

```

/*PROGRAM TREND. Traduction in C language from a program to remove
a trend, written by Vieira (1983) */
/*Initialise */
#include <malloc.h>
#include <math.h>
#include <stdio.h>
void SLE(double *,long ,long ,double );
void help();
double *dvector(long);
void free_dvector(double *);
void nrerror(char *);
double a[15][15];
void main(unsigned argc, char **argv)
{
long j,i,k,N,jb,kb,Iord,Iord2;
double SY,SYY,SYC,SYYC,Zest,Zdev;
double SST,SSR,SSD,NDF1,AMSR,NDF2,AMSD,R2,R,F,NDF3;
double b[15],c[15],lx;
double *X,*Y,*Z;
FILE *Fin,*Fout;
char fnl[80],idx;
c[1]=1.0;
/*Read parameters */
if(argc!=4)
{
help();
exit(0);
}
Fin=fopen(argv[1],"r");

```



```

strcpy(fnl,argv[1]);
idx = strcspn(fnl, ".");
fnl[idx+1]=0;
strcat(fnl,argv[2]);
Fout=fopen(fnl, "w");
Iord=atoi(argv[3]); /* Order of fitting */
/*Read data */
i=1;
do{
fscanf(Fin,"%le %le %le\n",&lx,&lx,&lx);
i++;
}
while (feof(Fin)==0);
N=i-1;
X=dvector(N+1);
Y=dvector(N+1);
Z=dvector(N+1);
fseek(Fin,0L,SEEK_SET);
for(i=1;i<=N;i++)
fscanf(Fin,"%le %le %le\n",&X[i],&Y[i],&Z[i]);
fclose(Fin);
/*Calculate number of coefficients */
Iord2=(Iord+1)*(Iord+2)/2;
/* Zero SLE matrix */
for(i=0;i<=Iord2;i++)
{
b[i]=0.0;
for(j=0;j<=Iord2;j++)
a[i][j]=0.0;
}
/* Calculate SLE matrix */
for(i=1;i<=N;i++) /* 3 */
{
jb=1;

```

```

for (j=1;j<=Iord;j++) /* 4 */
{
for(k=1;k<=j;k++) /* 5 */
{
jb++;
kb=jb-j;
c[jb]=c[kb]*X[i];
}
jb++;
c[jb]=c[kb]*Y[i];
}
for(j=1;j<=Iord2;j++) /* 6 */
{
b[j]=b[j]+c[j]*Z[i];
for(k=1;k<=Iord2;k++)
a[j][k]=a[j][k]+c[j]*c[k];
}
}
/*Solve SLE
Calculate and save estimated values and deviations
for each observation */
SLE(b,Iord2,15,1.0e-08);
SY=0.0;
SYY=0.0;
SYC=0.0;
SYYC=0.0;
for(i=1;i<=N;i++) /* 7 */
{
jb=1;
for(j=1;j<=Iord;j++) /* 8 */
{
for(k=1;k<=j;k++) /* 9 */
{
jb++;

```

```

kb=jb-j;
c[jb]=c[kb]*X[i];
}
jb++;
c[jb]=c[kb]*Y[i];
}
Zest=0.0;
for(j=1;j<=Iord2;j++) /* 10 */
Zest=Zest+b[j]*c[j];
Zdev=Z[i]-Zest;
SY=SY+Z[i];
SYY=SYY+(Z[i]*Z[i]);
SYC=SYC+Zest;
SYYC=SYYC+(Zest*Zest);
fprintf(Fout,"%f %f %f %f %f %f\n",X[i],Y[i],Z[i],Zest,Zdev);
}
fclose(Fout);
/* Calculate error measures */
SST=SYY-SY*SY/(double)N;
SSR=SYYC-SYC*SYC/(double)N;
SSD=SST-SSR;
NDF1=Iord2-1;
AMSR=SSD/NDF1;
NDF2=N-Iord2;
AMSD=SSD/NDF2;
R2=SSR/SST;
R=sqrt(R2);
F=AMSR/AMSD;
NDF3=N-1;
/*Write TREND results to file */
printf("Surface Parameters:");
for(i=1;i<=Iord2;i++) printf("%le ",b[i]);
printf("\n");
printf("Regression %le %f %le %le\n",SSR,NDF1,AMSR,F);

```

```

printf("Deviation %le %f %le\n",SSD,NDF2,AMSD);
printf("Total Variation %le %f \n",SST,NDF3);
printf("Goodness of Fit R=%le R2=%le\n",R,R2);
}
void help()
{
fprintf(stderr,"usage: Trend {input file} {output Ext.} {Fitting order}\n");
}
/*Subroutine for solution of N simultaneous equations.
Matrix A is N x N and B is a column vector of N elements.
A is converted to the identity matrix.
B contains the solution vector. */
void SLE(double *b,long N,long N1,double ZERO)
{
long i,j,k;
double DIV,RATIO;
for(i=1;i<=N;i++) /* 100 */
{
DIV=a[i][i];
if((abs(DIV)-ZERO)<=0.0)return;
for(j=1;j<=N;j++)
a[i][j]=a[i][j]/DIV;
b[i]=b[i]/DIV;
for(j=1;j<=N;j++) /* 102 */
{
if(i!=j)
{
RATIO=a[j][i];
for(k=1;k<=N;k++) a[j][k]=a[j][k]-RATIO*a[i][k];
b[j]=b[j]-RATIO*b[i];
}
}
}
}
}

```

```
double *dvector(long nh)
{
double *v;
unsigned long tam;
tam=(nh+1)*sizeof(double);
v=(double *)malloc((unsigned long) tam);
if (!v) perror("allocation failure in dvector()");
return v;
}
void perror(char error_text[])
{
void _exit();
char c;
fprintf(stderr,"Run-time error...\n");
fprintf(stderr,"%s\n",error_text);
fprintf(stderr,"...now exiting to system...\n");
_exit(0);
}
void free_dvector(double *v)
{
free(v);
}
```