

CONCEPTOS BÁSICOS DE GEOESTADÍSTICA EN GEOGRAFÍA Y CIENCIAS DE LA TIERRA: MANEJO Y APLICACIÓN

Julio González¹, Fernando Guerra² y Heriberto Gómez³

Universidad de Los Andes. Núcleo Universitario "Pedro Rincón Gutiérrez"

Departamento de Ciencias Sociales. Laboratorio de Sistemas de Información Geográfica

Recibido: octubre 2006

Aceptado: diciembre 2006

Resumen

El tratamiento estadístico de información cuantitativa ha sido realizado tradicionalmente a través de medidas descriptivas de tendencia central, de posición y dispersión o variabilidad, así como análisis de correlación y regresión. Sin embargo, en el campo de las ciencias ambientales, espaciales y ciencias de la tierra, tienden a adolecer de una referencia básica que permita relacionar las medidas obtenidas con una localización en el espacio. A partir de la segunda mitad de 1900, se generó una nueva tendencia en los estudios estadísticos para variables distribuidas espacialmente conocida como geoestadística. Dado que los programas de estadística normalmente implementados en el pregrado de las carreras de educación que se imparten en menciones como Geografía y Ciencias de la Tierra, en la Universidad de los Andes en el Táchira, aún no incorporan referencias al tratamiento estadístico de datos distribuidos espacialmente se pretende con este artículo ejemplificar las ventajas y posibilidades que conlleva el manejo y aplicación de las nociones básicas de geoestadística.

Palabras claves: geoestadística, datos distribuidos espacialmente, semivarianza.

BASIC CONCEPTS OF GEO- STATISTIC IN GEOGRAPHY AND EARTH SCIENCES: HANDLING AND APPLICATION

Abstract

The statistical treatment of quantitative information has been made traditionally through descriptive measures of central tendency, of position and dispersion or variability, as well as analysis of correlation and regression. Nevertheless, in the field of environmental sciences, space and Earth sciences, although it is important to know and to explain the general characteristics of the information, they suffer of a basic reference that allows to relate the obtained measures to a location in the space or area of study. From second half of 1900, it was generated a new tendency in the statistical studies for variables spatially distributed well-known as geo-statistic. Since the programs of statistic normally implemented in the undergraduate of the high education careers do not incorporate references to the statistical treatment in mentions as Geography and Earth Sciences, in the University of the Andes in the Táchira yet, this article to scope the advantages and possibilities that entails the handling and application of the basic of geo-statistic for the knowledge and detailed analysis of the distribution of spatial properties.

Key words: geo-statistic, spatial data distribution, semi variance.

¹ Ing. Forestal. Magíster en Suelos. Profesor adscrito al Departamento de Cs. Sociales ULA Táchira. LABSIG. Email: jtovar@ula.ve

² Geógrafo. Profesor de climatología adscrito al Departamento de Cs. Sociales ULA Táchira. LABSIG. Email: fguerra@ula.ve

³ Geógrafo. PhD en Geografía. Coord. del Laboratorio SIG adscrito al Departamento de Cs. Sociales ULA Táchira. Email: hgomez@ula.ve

1. Introducción

El inicio de un proyecto de investigación para la detección de áreas inestable en la cuenca del río Tórbes, en el estado Táchira, ha impulsado en los docentes que componen el Área de Ciencias de la Tierra, el uso frecuentes geoestadísticos y de herramientas como los sistemas de información geográfica (SIG); situación que se ha extrapolado a los estudiantes del pregrado de la carrera Educación Mención Geografía y Ciencias de la Tierra quienes reciben por su propia iniciativa un curso prebásico no formal sobre SIG.

Los estudiantes cursantes de la carrera antes mencionadas reciben, además, durante su formación profesional un curso básico de estadística donde se describen conceptos básicos de estadísticas paramétrica, como medidas de tendencia central (media, modo, mediana), medidas de dispersión (varianza, desviación típica y coeficiente de variación), incluso información sobre estadística no paramétrica, de gran utilidad para procesar la información cuantitativa y cualitativa colectada durante su labor docente.

Sin embargo, las nociones impartidas sobre sistemas de información geográfica y el análisis y procesamiento matemático y estadístico de información, que tiene obviamente una variabilidad espacial y temporal, genera en los educandos un conjunto de inquietudes y debilidades que se hacen evidentes al tener que incluir dentro del análisis estadístico variables que se caracterizan por tener: localización, anisotropía y continuidad, características inherentes a los hechos geográficos.

De manera que durante el desarrollo de estas notas se plantean algunas nociones elementales sobre geoestadística, se trata de mostrar algunas experiencias de investigaciones realizadas por los autores, que exponen la necesidad que tienen los educandos de incorporar dentro de sus análisis la ubicación de la información, concluyendo con la necesidad de incluir en los cursos básicos de estadística de la mención conceptos elementales sobre geoestadística, orientados a ser manejados posteriormente en un ambiente SIG.

2. Antecedentes

La geoestadística se refiere al conjunto de técnicas estadísticas utilizadas para el estudio de variables numéricas distribuidas en el espacio. Para Journel y Huijbregts (1978), es simplemente la aplicación de la teoría de funciones aleatorias al reconocimiento y estimación de fenómenos naturales.

La geoestadística se ha convertido en una herramienta útil para explicar la variación de propiedades en el espacio y su uso se ha generalizado recientemente donde se analizan hechos geográficos, debido a la posibilidad de conocer, entre otros, el rango de influencia de la propiedad, estimar su valor en sitios donde no existe información y por último, calcular el error de la estimación efectuada.

El uso de esta herramienta se fundamenta en el cálculo de la semivarianza y el semivariograma, los cuales constituyen las herramientas principales de un cuerpo de desarrollos matemáticos conocidos, según Royle (1980), como teoría de las variables regionalizadas, desarrollada por Matheron en la década de los sesenta (Matheron, 1963). Una variable regionalizada, de acuerdo con Ovalles (1992), no es más que una función que describe un fenómeno natural geográficamente distribuido.

Los semivariogramas son una representación gráfica en coordenadas cartesianas, donde en el eje de las ordenadas se registra la semivarianza, mientras por las abscisas se representa un vector con modulo igual al intervalo de muestreo, con dirección y sentido variables.

La semivarianza es una medida de la similitud que existe entre observaciones situadas a una determinada distancia y fue definida inicialmente por Matheron (1963) como:

$$\gamma(h) = \frac{1}{2V} \iiint (F(m+h) - F(m))^2 \partial v \quad (1)$$

En forma operacional se calcula, de acuerdo con Viera, Nielsen y Biggar (1981), de la siguiente manera:

$$y(h) = \frac{1}{2} n [Z(X_i) - Z(X+h)] \quad (2)$$

Donde:

$y(h)$ = Semivarianza

n = número de pares

$Z(X)$ = valor de la propiedad en el lugar x

$Z(X+h)$ = valor de una propiedad a una distancia h del lugar x

El semivariograma permite identificar la naturaleza de la variabilidad espacial de la propiedad, que según Burgess y Webster (1980), está compuesta por un componente espacialmente correlacionado ($\varepsilon'(x)$) y un componente al azar (ruido) no correlacionado espacialmente ($\varepsilon''(x)$). De acuerdo con Burrough (1983), esto puede expresarse de la siguiente manera:

$$Z(x) = m(x) + \varepsilon'(x) + \varepsilon''(x) \quad (3)$$

Donde:

$Z(x)$ = es el valor de la propiedad en el punto x

$m(x)$ = es el valor medio de la propiedad, el cual se asume constante en el área de muestreo.

Este planteamiento puede visualizarse en forma sencilla (figura 1), si se considera que, de acuerdo con Davis (1973), la semivarianza puede ser expresada como una diferencia, tal como se muestra en la ecuación 4, siempre y cuando se cumpla la estacionaridad de segundo orden, que se refiere a la estacionaridad del primer y segundo momento de la distribución (media, varianza y covarianza) de los valores estudiados en el área

$$Y(h) = C(o) - C(h) \quad (4)$$

Donde:

$C(o)$ = es el equivalente a la varianza de una propiedad (variable regionalizada), obtenida sobre un conjunto de localidades de muestreo.

$V(h)$ = es la covarianza de la propiedad para una distancia de separación (h).

La ecuación 4 puede representarse gráficamente, como se observa en la Figura 1, y en ella es posible observar algunos detalles. Primero: mientras más

similares sean las observaciones, menor será la semivarianza, es decir, cuando la distancia de separación entre localidades se hace muy corta, la varianza y la covarianza tienden a ser iguales y la semivarianza tiende a cero. Segundo, cuando el vector distancia (h) se hace grande, la covarianza disminuye, debido a que progresivamente existe mayor independencia entre los puntos. Simultáneamente, la semivarianza incrementa hasta alcanzar un valor teóricamente igual a la varianza de la propiedad ($C(o)$).

La distancia a la cual $\gamma(h) = C(o)$, es denominada rango. Este constituye la separación a partir de la cual, los valores de la propiedad comienzan a ser independientes unos de otros, es decir, la zona de influencia. Por otro lado, es necesario apuntar que por definición, $\gamma(h) = 0$ cuando $h = 0$, (figura 1). No obstante, en la práctica cuando $(\gamma(h))$ se aproxima a un valor positivo, el cual es denominado efecto nugget (efecto pepita) o error aleatorio.

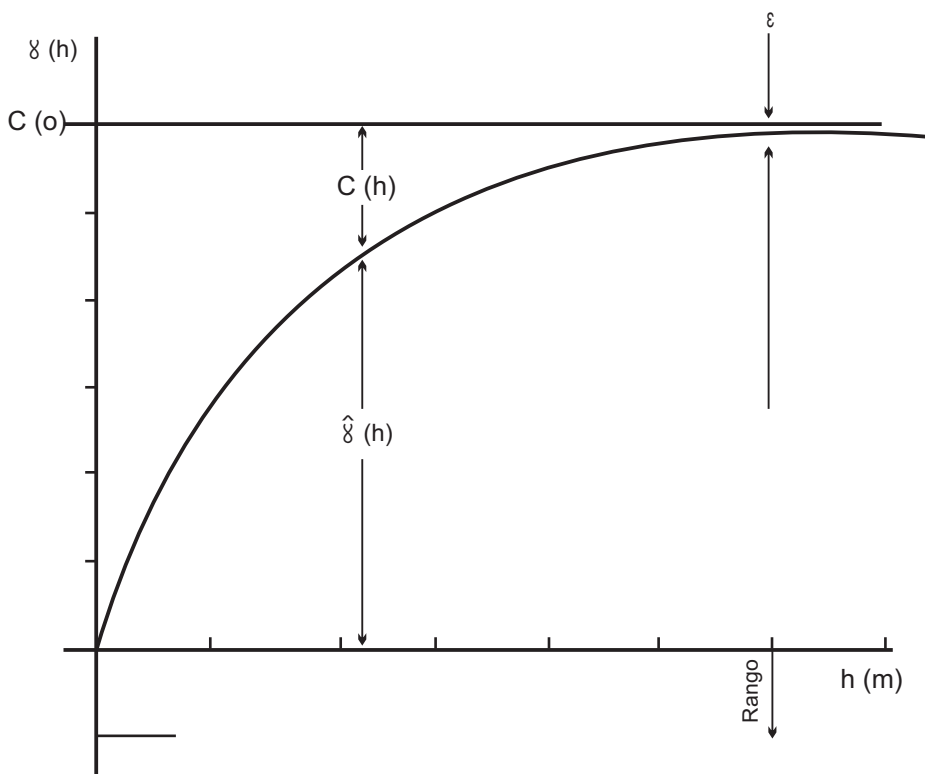


Figura 1: Semivariograma, donde $C(0)$ representa la varianza de los datos y ϵ el límite mas allá del cual $\gamma(h)$ se considera igual $C(o)$

El semivariograma permite obtener el conocimiento (cuantitativo) de cómo la variable en estudio cambia de una localidad a otra, en otras palabras, tal y como lo plantea Ovalles (1992), el semivariograma contiene toda la información referente a la variable regionalizada como es:

- El rango (zona) de influencia, que indica la región en la cual existe una dependencia espacial.
- Anisotropía, que expresa el comportamiento de la variación de la propiedad en diferentes direcciones.
- Continuidad de la variable en el espacio, expresada por el comportamiento del semivariograma en el origen.

Los puntos experimentales que conforman el semivariograma deben ajustarse a un modelo (análisis estructural), y las funciones matemáticas que se utilizan, con el objeto de explicar el comportamiento del semivariograma, deben cumplir una serie de supuestos, que solo son satisfechos por un grupo limitado de funciones.

La función matemática ajustada a los datos experimentales encontrados para la semivarianza en el semivariograma, es utilizada posteriormente, para calcular estimados de la semivarianza a diferentes distancias. El semivariograma puede ser calculado en diferentes direcciones, por eso puede evidenciar aspectos del comportamiento espacial (isotropías, anisotropías) y es la clave para describir la variación espacial de una propiedad cuantitativamente, por ello, es utilizado para realizar interpolación y predecir valores de una propiedad en áreas no muestreadas, mediante un promedio móvil ponderado, desarrollado por Krige (1951) en las minas de Sudáfrica, y popularmente conocido como *kriging* o *krigeado*.

Por medio del kriging los valores de las propiedades en estudio son calculados utilizando parámetros de los semivariogramas ajustados, mediante la siguiente ecuación:

$$Z(x_0) = \sum_{i=1}^n \lambda_i \cdot Z(x_i)$$

Donde:

- $Z(x_0)$ = estimado del atributo
 $Z(x_i)$ = muestra vecina a $Z(x_0)$
 λ_i = ponderaciones aplicadas a cada $Z(x_i)$
 n = número de muestras vecinas a $Z(x_0)$

3. Planteamientos básicos

Con la idea de dar a conocer las bondades de esta técnica estadística, se procedió a desarrollar una investigación pedológica con apoyo de la geoestadística. En efecto, los geoestadísticos han sido aplicados con éxito en estudios de suelos (ver por ejemplo, Alperín, Borges y Sarandón, 2002). Para ello se consideraron los datos de 31 barrenos recolectados en campo, tal como se destaca a continuación. Al estudiar las propiedades de los suelos, dentro de una delineación del mapa de la vegetación actual de Venezuela, ubicada a 35 km de la población de El Sombrero, estado Guárico, se colectó información con barreno (hasta 120 cm de profundidad), un mínimo de trabajo de laboratorio y observaciones de superficie.

La información recolectada para cada sitio fue organizada de manera que cada barreno es descrito por 17 atributos, de los cuales se seleccionaron para este trabajo cuatro propiedades por su relación indirecta con el régimen hídrico y por su

importancia desde el punto de manejo de suelos en el área (croma, profundidad de concreciones, porcentaje de arcilla y porcentaje de arena). En el Cuadro 1 se muestra la media, desviación standard, y coeficiente de variación para los atributos seleccionados en el área de estudio

Cuadro 1: Algunos parámetros estadísticos generales en el área de estudio.

Atributo	Media	Desviación Standard	Coefficiente de variación
Croma	3,84	1,83	47,53
Concreciones	12,72	11,31	88,94
% Arcilla	38,14	5,59	14,67
% Arena	35,45	8,2	23,0

Esos parámetros estadísticos, aunque valiosos, no suministran información sobre posición en el espacio de los datos, es decir, cómo se distribuyen los valores de las propiedades en el área estudiada. En otras palabras, los valores medios por ejemplo del contenido de arcilla y del contenido de arena para toda el área estudiada son 38,14% y 35,45% respectivamente y su desviación estándar de 5,6 y de 8,2 dando una idea del rango de variación de la propiedad, sin embargo no indican donde se ubican esos valores extremos (máximos y mínimos). Esa última información, es básica para los estudiantes de educación, mención Geografía y Ciencias de la Tierra.

Para ilustrarlo, es ampliamente conocido que la fracción arcilla, entre otros, por su elevada superficie específica y las cargas eléctricas generadas por diversos mecanismos, es determinante de muchas de las propiedades físicas, químicas, biológicas del suelo, y por ello obviamente influye sobre su calidad, de manera que se considera información importante para el diseño y aplicación de las prácticas que deben implementarse para lograr un aprovechamiento sostenible del recurso.

La información suministrada en el Cuadro 1 podría llevar al diseño de prácticas de manejo de suelos uniformes para toda el área, obviando que dentro de la misma podrían existir localidades que requieren un tratamiento diferencial.

Para expresar y visualizar el planteamiento, se muestra un análisis de superficie respuesta realizado en el área en cuestión, dicho análisis de acuerdo con Davis (1973) y Burrough (1983), constituyen una de las formas más simples para describir la variación espacial gradual de amplio rango, con el uso de un modelo de regresión polinomial que considera el valor de la propiedad (variable dependiente) en función de las coordenadas geográficas (variables independiente).

4. Resultados

Los puntos de observación (31 barrenos) se localizaron y desplegaron, seleccionado un punto de referencia y en coordenadas relativas para lograra ajustar modelos. Los modelos de regresión obtenidos para una superficie cúbica son los siguientes:

$$\% \text{Arcilla} = 84,54 - 319,33X_1 - 1114,66X_2 + 668,25X_1^2 + 2126X_2^2 - 51,67X_1 \cdot X_2 - 340,79X_1^3 - 926,58 \cdot X_2^3 - 641,94X_2^2 \cdot X_1$$

$$\% \text{Arena} = 71,26 + 27,18 X_1 + 1888,62 X_2 - 489,35 X_1^2 - 3478,29 X_2^2 - 101,49X_1 \cdot X_2 + 357,68 X_1^3 + 1352,19 X_2^3 - 1071,77 \cdot X_1 \cdot X_2 + 1375,82 X_2^2 \cdot X_1$$

En el cuadro 2 se muestra el coeficiente de correlación (r) y la F calculada para los modelos de regresión calculados, que conforman una parábola cúbica con máximos y mínimos

Cuadro 2: Información sobre el ajuste de los modelos calculados para el área de estudio.

Atributo	F. calculada	r
% Arcilla (A)	4,42**	0,82
% Arena (a)	3,40**	0,79

** Altamente significativo $\lambda = 0,01$

Tal y como se observa en el Cuadro 2, las ecuaciones de regresión calculadas pueden explicar más del 70% de la variación total del porcentaje de arcilla y de arena en el área, y parecen estar relacionadas a la morfología colinosa del paisaje. Esta variación se representa gráficamente aplicando las ecuaciones tal y como se muestra en la Figura 2.

Como puede observarse el contenido de arcilla (%) muestra una clara disminución hacia el norte del área de estudio, ello obviamente muestra que dentro del área existen lugares con variaciones contrastes de esa propiedad determinante del comportamiento del suelo y que probablemente requerirían la implementación de distintas practicas de manejo.

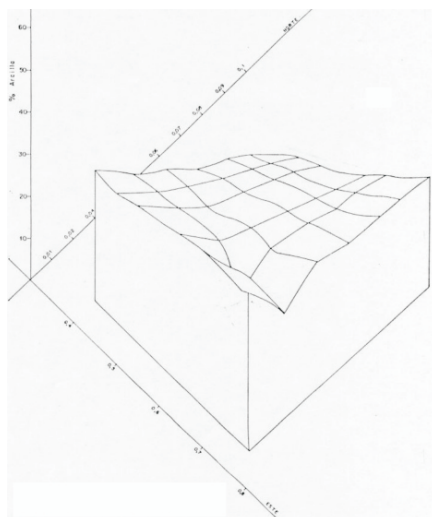


Figura 2: Representación gráfica del porcentaje de arcilla según el modelo cúbico dentro del área de estudio.

La tasa de cambio de las propiedades analizada dentro del área de estudio, implica que existe una variabilidad sistemática de las propiedades estudiadas. Según Borrough (1983), en el caso de los suelos, esta variabilidad sistemática, representa cambios graduales o marcados en las propiedades de los suelos, como una función de la interacción de los materiales parentales, el clima, la hidrología, el relieve y la actividad biológica, durante determinados lapsos de tiempo, es decir, que es el resultado de los factores, que a su vez presentan variación espacial y que pueden interactuar en forma diferencial de acuerdo a la escala de observación. Es este tipo de variación, es la que ha permitido a los prospectores obtener información sobre los suelos en forma confiable y efectivamente diferenciarlos y por supuesto extrapolar el conocimiento que tenemos sobre este cuerpo natural.

Con la información, se calculó además la semivarianza y se confeccionaron semivariogramas no diseccionados para el porcentaje de arena y arcilla, como se muestran en la Figura 3.

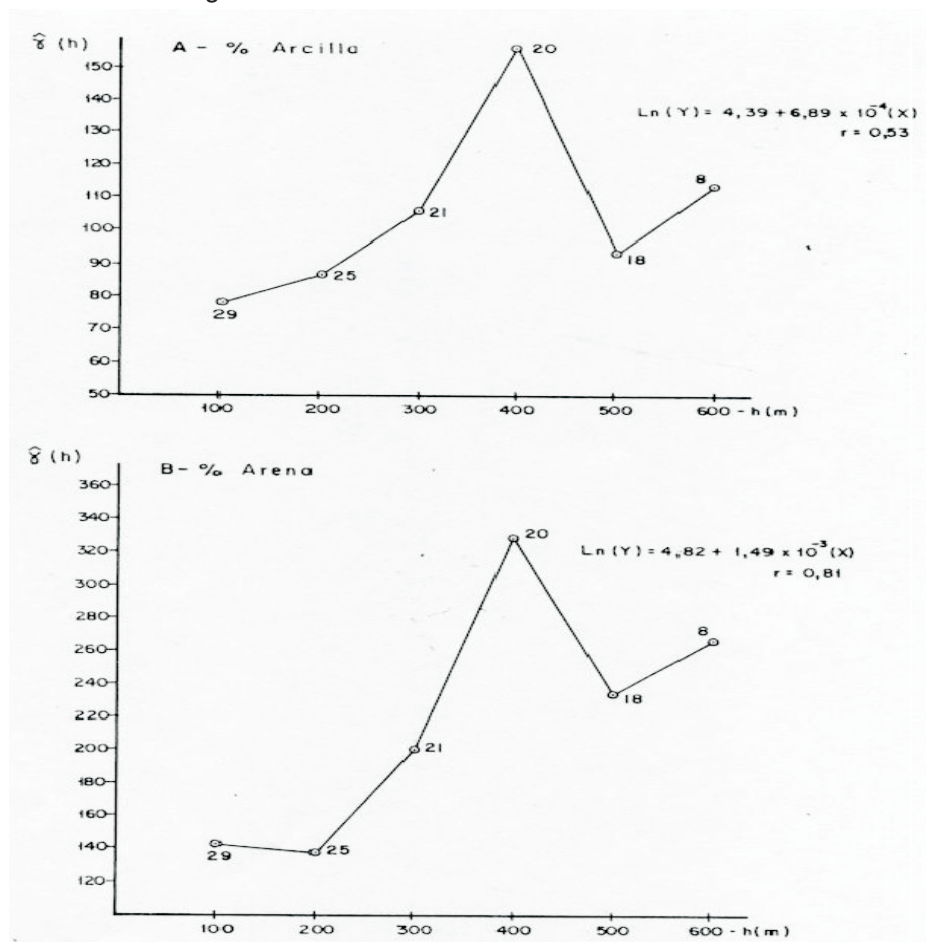


Figura 3: Semivariograma del porcentaje de arcilla (A) y del porcentaje de arena (B) de horizontes superficiales en el área de estudio.

Al hablar de dependencia espacial es importante indicar que cuanto más cercanas estén las muestras, éstas serán más similares, mientras que a mayor distancia de separación las observaciones serán menos similares (Nielsen y Alemi, 1989). De esta manera al considerar los semivariogramas mostrados en la Figura 3 es posible concluir que el área de influencia para estas propiedades (independencia de los valores) se encuentra en 400 m (rango). En efecto, se ha colectado información sobre la escala de variación y consecuentemente se ha determinado la mejor distancia para realizar un muestreo sistemático. En otras palabras, el muestreo inicial se realizó con muestras colectadas cada 100 m y la aplicación de la geoestadística nos indica que se pudo obtener esencialmente la misma información con menor número de muestras a 400 m de distancia.

Tal y como lo plantea Ovalles (1992), las muestras se toman bajo el supuesto que las características medidas en un punto también representan las áreas cercanas no muestreadas. Por supuesto como lo plantea el autor, la certeza de ese supuesto esta en función del grado de dependencia espacial que existe entre las observaciones. Si no hay dependencia espacial, la muestra solo representaría un punto. En este sentido, se ha demostrado que en la gran mayoría de los hechos geográficos, las muestras no solo representan un punto sino que tienen un área de influencia cuyo tamaño es una función directa de la dependencia espacial existente.

Los cambios de las variables pueden ser mayores en una dirección que en otra, a este fenómeno se le denomina anisotropía y es muy importante para definir las dimensiones del área de influencia de las propiedades (Camacho, Reyes y Nieto, 2006). Los semivariogramas mostrados en la Figura 3 no son direccionados de tal manera que este fenómeno no se analiza, por tal motivo, de asumirse isotropía el alcance (rango) representa el radio de un círculo.

La continuidad de la variable en el espacio, se expresa en el comportamiento del semivariograma en el origen, teóricamente pasa por el origen, cuando no es así, existe un efecto aleatorio (nugget) que de acuerdo con Ovalles (1992) es una función de la escala. En los semivariogramas de la Figura 3 el efecto nugget corresponde a menos de un 30% de la varianza total.

Este procedimiento conduce a obtener una línea polinomial, una superficie o una hipersuperficie, dependiendo si los datos se encuentran en una dimensión, dos dimensiones o se incorpora una tercera dimensión. Esta herramienta es ampliamente utilizada en geología como forma de estudiar las estructuras regionales y sus desviaciones locales y también en climatología para el estudio de la variabilidad de elementos como la temperatura estrechamente relacionada con la posición y la altitud.

5. Conclusiones

Este ejercicio académico se efectuó para demostrar en forma cuantitativa algunas de las relaciones entre los valores de las propiedades (% de arcilla y % de arena) y su ubicación en el espacio, bajo una óptica geoestadística. Sin embargo, algunos supuestos deben considerarse en la aplicación del análisis, como por ejemplo que las variables dependientes deben distribuirse normalmente (Davis, 1973 y Draper y Smith, 1981).

Sería deseable, además, que la ecuación de regresión sea determinada a partir de datos extraídos en un muestreo al azar, que agrupe la máxima información posible de la propiedad estudiada. Este punto difícil de cumplir con los datos que se utilizan en este ejemplo por que la información fue colectada a lo largo de una vía de penetración. Esta forma de colectar los datos y aunque permite

obtener información sobre la variación general y local, puede causar según Davis (1973), distorsión de la ecuación de regresión paralela al patrón de muestreo.

Finalmente, el procesamiento geoestadístico conlleva al análisis de la distribución de datos, que deben ser observados en su comportamiento espacial y en sus interrelaciones. El ejercicio académico sugiere que los estudiantes de carreras afines a Geografía y a las Ciencias de la Tierra, como estudiosos del espacio, deben ser involucrados en el conocimiento y aplicación de estas técnicas de análisis espacial que, de acuerdo con las experiencias relacionadas con la materia, se potencian con la aplicación de los sistemas de información geográfica.

Agradecimientos

Los autores agradecen al Consejo de Desarrollo Científico, Humanístico y Tecnológico (CDCHT) de la Universidad de Los Andes por el financiamiento a la siguiente investigación **Código NUTA-H-241-06-09-AA**.

Referencias Bibliográficas

- ALPERÍN, M.; BORGES, V. y SARANDÓN, R. (2002). "Caracterización espacial de los tipos de cobertura de suelo usando técnicas geoestadísticas a partir de información satelital". *Revista de la Facultad de Agronomía*. 105 (1), 45-52.
- BURGESS, T. y WEBSTER, R. (1980). "Optimal interpolation and isarithmic mapping of soil properties. I. The semivariogram and punctual kriging". *J. Soil Science*, 31: 315-331.
- BURROUGH, P. (1983). "Multiscale source of spatial variation in soil. I. The application of fractal concepts to nested levels of soil variation". *J. Soil Science*, 34: 577-597.
- CAMACHO, J.; REYES, O. y NIETO, A. (2006). "Anisotropía de esfuerzos y resistencia al corte de arenas". *Ciencia e ingeniería Neogranadina*. 54 Vol. 16 - No.
- DAVIS, J. (1973). *Statistical methods in geology*. New York: Wiley and Sons.
- DRAPER, N. y SMITH, H. (1981). *Applied regression analysis*. New York: Wiley and Sons.
- JOURNEL, G. y HUIJBREGTS, C. (1978). *Mining geostatistics*. London. Academic Press.
- KRIGE, D. (1951). "A statistical approach to some basic mine valuation problems on the Witwatersrand". *J. of the Chem., Metal. and Mining Soc. of South Africa* 52 (6): 119-139.
- MATHERON, G. (1963). "Principles of Geostatistics". *Economic Geology*. 58: 1246-1260.
- NIELSEN, D. y ALEMI, A. (1989). "Statistical opportunities for analyzing spatial and temporal heterogeneity of field soils". *Plant Soil* 115:285-296.
- OVALLES, F. (1992). "Metodología para determinar la superficie representada por muestras tomadas con fines de fertilidad". Maracay, Venezuela: FONAIAP-CENIAP-IIAG. 44p. SERIE B.
- ROYLE, A. (1980). "¿Why geostatistics?". In: P.F. Massett-jones, *Geostatistics*. New York: Mc Graw-Hill, 1-13.
- VIERA, S.; NIELSEN, D. y BIGGAR, J. (1981). "Spatial variability of field-measured infiltration". *Soil Sci. Soc. Am. J.* 45: 1040-1048.