

Notas sobre el manejo del software geoestadístico Variowin. Parte I: Cálculo del variograma

Notes on using geostatistical software variowin. Part I: Variogram calculation

Mariana S. Santamaría*, Marta S. Malla*

RESUMEN

El término geoestadística comprende el análisis y el modelado de datos espaciales que están distribuidos en un sistema de coordenadas del espacio y/o tiempo. La teoría de variables regionalizadas, o Geoestadística, nació como una metodología inferencial para evaluar reservas de oro y se extendió rápidamente a otras observaciones espaciales donde los supuestos básicos de la estadística clásica no se cumplían. Los tres pasos básicos de un estudio geoestadístico incluyen: análisis exploratorio de datos, análisis estructural (cálculo y modelado de variogramas) y predicciones (kriging o simulaciones). El objetivo de este trabajo es proveer de una introducción detallada sobre la exploración y cálculo de variogramas utilizando el software Variowin 2.21. Este software es de distribución libre y puede obtenerse de internet. El modelado de variogramas se describe en otro trabajo (Parte II). Variowin está diseñado tanto para iniciar al principiante en el uso de los métodos geoestadísticos de una manera fácil, como para proporcionar al usuario experimentado suficiente potencia y flexibilidad para resolver problemas reales.

Palabras claves: Geoestadística, Análisis estructural, Variograma.

ABSTRACT

The term Geostatistics covers the analysis and modeling of spatial data that are distributed in a coordinate system of space and/or time. The theory of the regionalized variable, or Geostatistics, was born as an inference methodology to evaluate ore reserves and extended rapidly to other spatial observations where basic assumptions of classical statistics did not work. The three basic steps in a geostatistical study include: exploratory data analysis, structural analysis (calculation and modeling of variograms) and making predictions (kriging or simulations). This paper aims to provide a detailed introduction to the exploration and calculation of variograms using Variowin 2.21 software package. This packages is freely available from internet. Modeling of variograms is detailed in a another paper (Part II). This software Variowin is designed to make it easy for the novice to begin using geostatistical methods, as well as to provide sufficient power and flexibility for the experienced user to solve real-world problems.

Keywords: Geostatistics, Structural analysis, Variogram.

* Departamento de Matemática; Universidad Nacional del Sur, Bahía Blanca, Argentina.
E-mail:msanta@infovia.com.ar

INTRODUCCIÓN

Estas notas constituyen una guía introductoria más detallada que la dada por Malla y Santamaría (2005) sobre el manejo del software VARIOWIN versión 2,21 de Yvan Pannatier, presentado en el XXXIII Coloquio Argentino de Estadística en octubre de 2005. Este software, diseñado únicamente para la exploración y modelado de semivariogramas, permite realizar el análisis estructural de datos espaciales. Es de dominio público y puede descargarse de la dirección: <http://www-sst.unil.ch/research/variowin/index.html>.

La primera versión¹, dedicada íntegramente a la enseñanza del análisis variográfico, fue libre; sin embargo, su distribución y uso debían tener la autorización del autor, motivo por el cual no estaba disponible en la web. Posteriormente una versión comercial (Pannatier, 1996) estuvo disponible para la venta junto con el libro correspondiente. Ambos se vendían juntos, pero actualmente están fuera de edición; por ese motivo la última versión 2.21 del software es de dominio público.

Con el objetivo de exhibir el funcionamiento del software en estas notas se utiliza un conjunto de datos obtenido de la siguiente dirección de Internet: <http://www.bioss.sari.ac.uk/smart/unix/mvariog/slides/frames.htm>.


Antes de comenzar la descripción del software, cabe recordar que: la geoestadística es una herramienta imprescindible para analizar patrones de distribución espacial de datos provenientes de disciplinas tales como minería, geología, edafología, ecología y otras. Comprende un conjunto de herramientas y técnicas que sirven para analizar y predecir los valores de una variable que se muestra distribuida en el espacio o en el tiempo de una forma continua (Malla y Santamaría, 2005). Todo trabajo geoestadístico se lleva a cabo en tres etapas (Moral, 2004):


1. Análisis exploratorio de los datos: Consiste en la aplicación de técnicas descriptivas al conjunto de datos disponible, con el fin de conocer su distribución e identificar datos erróneos, pero sin tener en cuenta su distribución geográfica. Es la etapa de aplicación de la estadística clásica.
2. Análisis variográfico o análisis estructural: Está compuesto por el cálculo del semivariograma experimental y el ajuste a éste de un modelo teórico conocido. El semivariograma experimental es la herramienta geoestadística más importante en la determinación de las características de variabilidad y correlación espacial del fenómeno estudiado, es decir, permite conocer cómo la variable cambia de una localización a otra (Isaaks y Srivastava, 1989).

3. Predicciones: Consiste en realizar estimaciones de la variable en los puntos no muestreados, es decir, estimar valores desconocidos a partir, no sólo de los conocidos, sino también de su estructura de continuidad espacial. Con la información procedente de la estimación se puede realizar una serie de mapas sobre la distribución de la variable en el área de interés.

DESCRIPCIÓN DEL SOFTWARE

VARIOWIN se utiliza para el análisis espacial de datos y modelado del variograma en dos dimensiones. Consta de la siguiente colección de cuatro programas (archivos de extensión **.exe**) que deben ser ejecutados separadamente y en un cierto orden:

Prevar2D, genera un archivo de distancias **.pcf** para todos los posibles pares de datos existentes en un archivo de datos de extensión **.dat**.  **Prevar2D**

Vario2D, utiliza el archivo de comparación de a pares **.pcf** originado por el Prevar2D para hacer un análisis variográfico exploratorio en 2D.  **Vario2D with PCF**

Model, permite realizar de manera interactiva el ajuste a un modelo teórico del variograma experimental obtenido previamente por Vario2D.

 **Model**

Grid Display, sirve para exhibir archivos de grilla (formato ASCII **.grd**) como mapas de pixels.

 **Grid Display**

Los archivos de datos para VARIOWIN requieren de un formato específico (Geo-EAS, Englund y Sparks, 1991), común a varios softwares estadísticos, con extensión **.dat** y con un máximo de ocho caracteres para su nombre, caso contrario éste será truncado. La primera fila del archivo debe contener el título, la segunda, cantidad de variables incluidas las coordenadas X e Y, las siguientes, los nombres de las variables, las columnas de datos y la última en blanco. La Fig. 1 exhibe un archivo de datos basado en dicho formato.

Mediciones de backscatter			
4			
Muestra			
microwave			
1	6	6	86
2	18	6	63
...
400	234	234	97
(renglón en blanco obligatorio en todo archivo de datos)			

Fig. 1. Archivo de datos en formato Geo-EAS, backscat.dat.

¹ Corresponde a la tesis doctoral de Yvan Pannatier: *Logiciel pour l'analyse spatiale de données en 2D - Etude géologique et géostatistique du gîte de phosphates de Taiba (Senegal)*. Ph.D. thesis, University of Lausanne, Lausanne, Switzerland, 1995.

Los datos del archivo **backscat.dat** provienen de la dirección de internet mencionada anteriormente. La variable microwave del conjunto de datos representa mediciones de radiación de microondas (microwave backscatter) pertenecientes a una zona de Gran Bretaña, medida sobre una grilla de 20 x 20 puntos con intervalos entre puntos de 12 metros. A menudo la cubierta de vegetación es estimada desde aviones o satélites utilizando radares de apertura sintética. La radiación de microondas es dirigida hacia la superficie terrestre y un sensor detecta la señal reflejada, que usualmente es función de los cultivos, condiciones del suelo y humedad.

Prevar2D

Una vez creado el archivo de datos, el primer programa a ejecutar es **Prevar2D**. Este programa genera un archivo con extensión **.pcf** (pair comparison file) conteniendo todas las posibles distancias entre puntos de observaciones. Generado el archivo **.pcf** (**backscat.pcf**), éste será utilizado para los análisis subsiguientes en **VARIOWIN**, dejando de lado el archivo **.dat** (**backscat.dat**). Haciendo doble clic en el ícono **Prevar2D** se accede al programa, que inicialmente muestra una ventana con la presentación del mismo (Fig. 2).

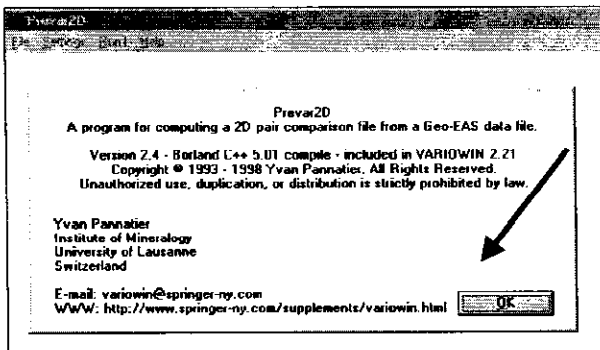


Fig. 2. Ventana de apertura del programa Prevar2D.

Presionando OK surge un cuadro de diálogo (Fig. 3), para la elección del archivo de datos, **backscat.dat**. Hacer doble clic en [...] y recorrer el árbol de carpetas hasta encontrar la que contiene dicho archivo y seleccionarlo (Fig. 4).

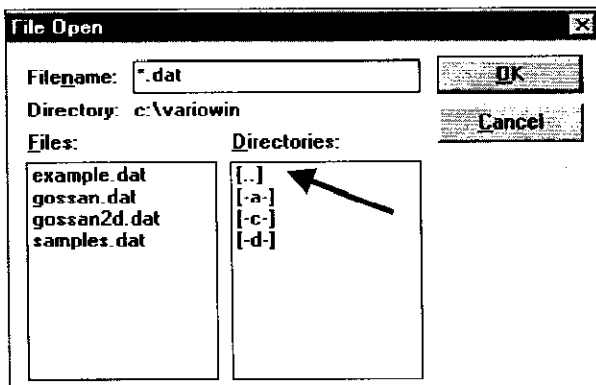


Fig. 3. Ventana para elegir el archivo de datos backscat.dat.

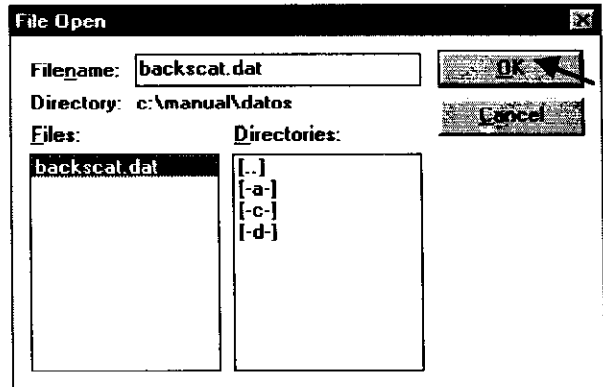


Fig. 4. Apertura de backscat.dat.

Presionando OK aparece un cuadro que contiene un resumen con la información del archivo de datos, (Fig. 5).

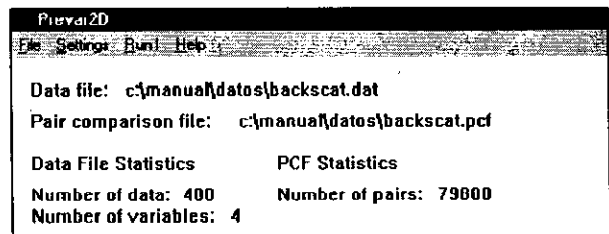


Fig. 5. Resumen de los datos del archivo backscat.dat.

Al descargar de la barra de menú **Settings/XY-Coordinates...** aparece una ventana que permite seleccionar las coordenadas X e Y para el archivo **backscat.dat** (Fig. 6).

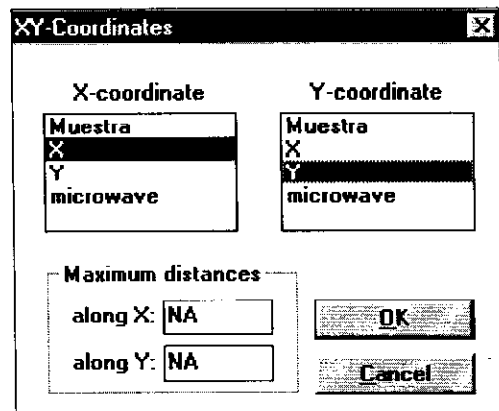


Fig. 6. Elección de las coordenadas X e Y.

Una vez elegidas las coordenadas, presionando OK se vuelve a la ventana anterior quedando habilitado el comando **Run!**, haciendo clic en él comienza el cálculo del archivo de distancias. Al final del proceso aparece en pantalla el resumen anterior incluyendo la cantidad de pares (79800) para los cuales se calcularon las distancias (Fig. 7).

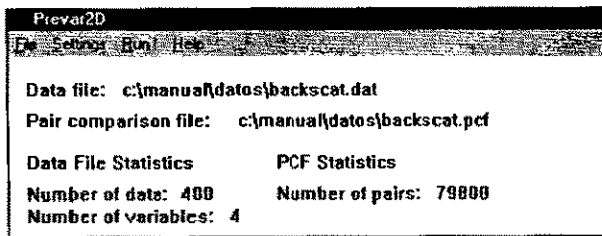


Fig. 7. Resumen de datos después de ejecutar el comando Run!

Presionando **File/Exit** se sale del programa **Prevar2D**, creándose el archivo **backscat.pcf**.

Vario2D

Haciendo doble clic en el icono **Vario2D** se accede al programa que muestra inicialmente una ventana con la presentación del mismo.

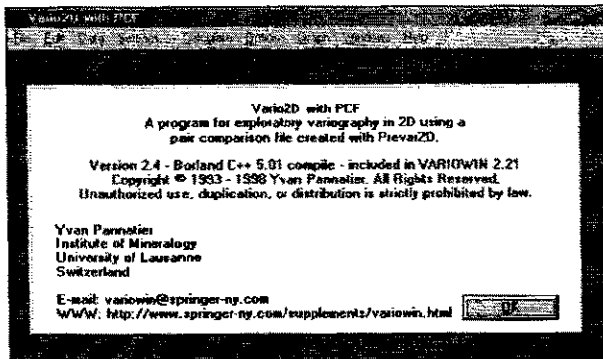


Fig. 8. Ventana de apertura del programa Vario2D.

Presionando **OK** aparece un cuadro de diálogo para la elección del archivo de comparación de pares **backscat.pcf**. Haciendo doble clic en **[..]** es posible encontrar la carpeta que contiene el archivo **backscat.pcf** y seleccionarlo (Fig. 9).

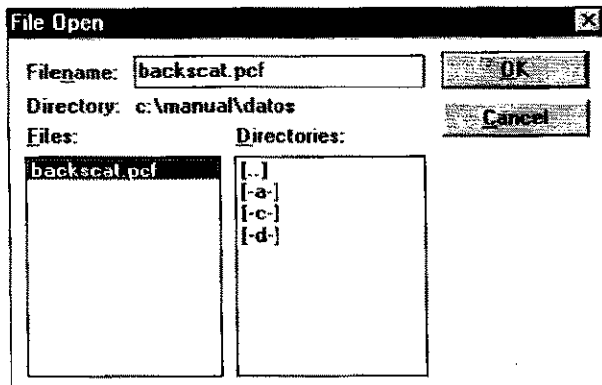


Fig. 9. Ventana para selección y apertura del archivo de datos **backscat.pcf**.

Al presionar **OK** se activan las opciones del menú principal (Fig. 10).

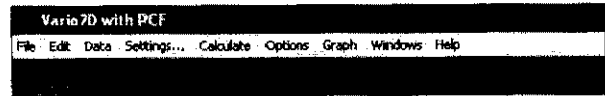


Fig. 10. Menú principal del programa Vario2D.

Al elegir **Data/Limits** se abre un cuadro conteniendo la información de los límites de las variables. La opción **Data/Map!** crea un mapa de puntos que representa la ubicación de las muestras (Fig. 11).

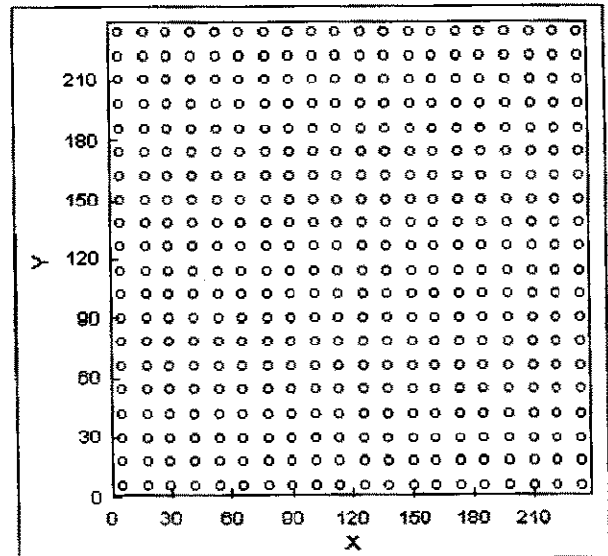


Fig. 11. Distribución de los puntos muestreados para el archivo **backscat.dat**.

Se puede cambiar el aspecto gráfico de los puntos con el comando **Settings...** (Default Symbols for Scatterplots) del menú principal (Fig. 12) o también con el comando **Graph** (Fig. 13) que permite además modificar los ejes.

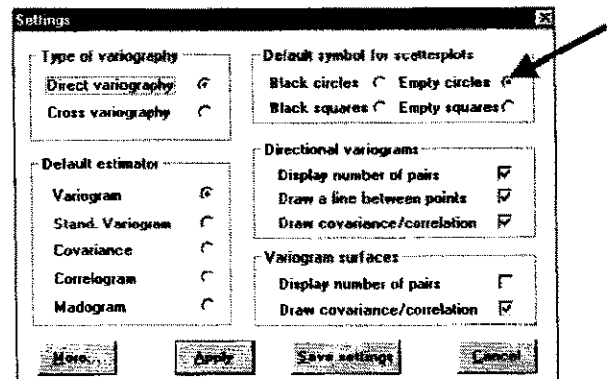


Fig. 12. Menú **Settings** para modificar el aspecto gráfico de puntos.

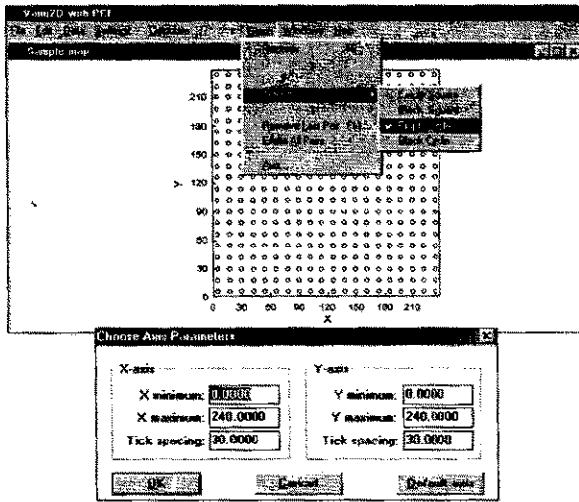


Fig. 13. Menú **Graph** para modificar aspecto gráfico de puntos y ejes.

Haciendo clic sobre cualquiera de los puntos de la Fig. 11 aparece un recuadro con toda la información correspondiente a dicho punto como se observa en la Fig. 14.

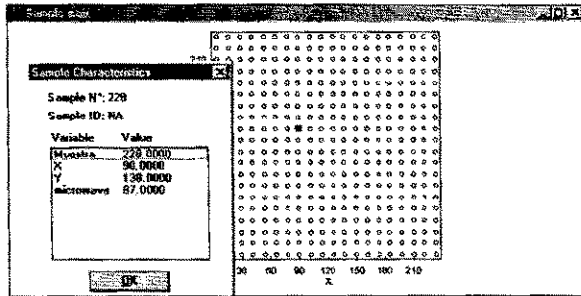


Fig. 14. Información asociada a la muestra 228.

La opción más importante del programa Vario2D es **Calculate/Directional Variogram** (Fig. 15) que permite calcular el variograma experimental.

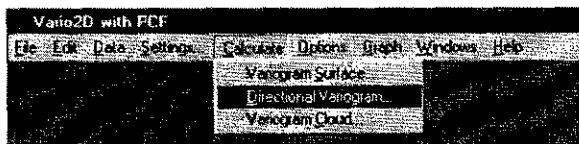


Fig. 15. Comando para el variograma direccional.

La terminología adoptada en VARIOWIN es «variograma», siendo el término correcto «semi-variograma», cuya expresión analítica es la siguiente:

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

donde h es la distancia de separación de las posiciones de las muestras x_i y $x_i + h$, $z(x_i)$ es la variable de interés en la posición i y $N(h)$ es el número de pares de datos separados por una distancia h . Graficando $g(h)$ versus h se obtiene la siguiente figura:

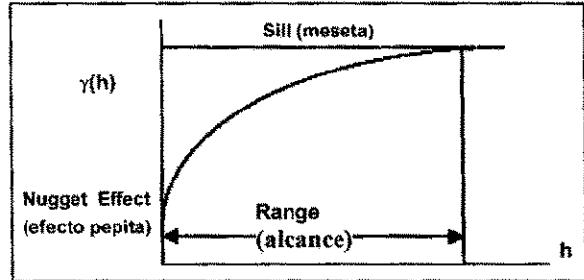


Fig. 16. Gráfico de $g(h)$ versus h .

Al seleccionar la opción mencionada **Calculate/Directional Variogram**, surge una ventana (Fig. 17) que permite elegir la variable a analizar y que por defecto asigna valores a los parámetros para el cálculo de variograma experimental. Al presionar OK aparece el gráfico del variograma experimental omnidireccional, $g(h)$.

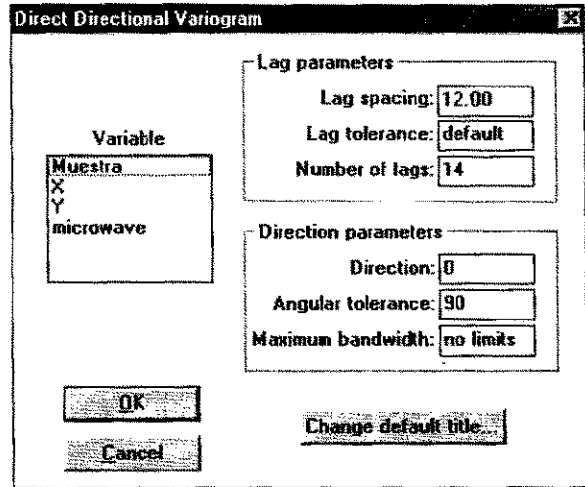


Fig. 17. Parámetros (por defecto) para el cálculo del variograma experimental.

Los valores de los parámetros pueden ser modificados a fin de obtener un variograma más apropiado teniendo en cuenta la cantidad de puntos que forman $g(h)$ y magnitud de las distancias h (Cressie, 1993, Webster y Oliver, 1990). Para encontrar estos valores es conveniente estudiar la nube de variograma. Para ello, elegir la opción **Calculate/Variogram Cloud**, surge una ventana (Fig. 18) que permite seleccionar la variable a analizar y que tam-

bién por defecto asigna valores a los parámetros requeridos (Fig. 18).

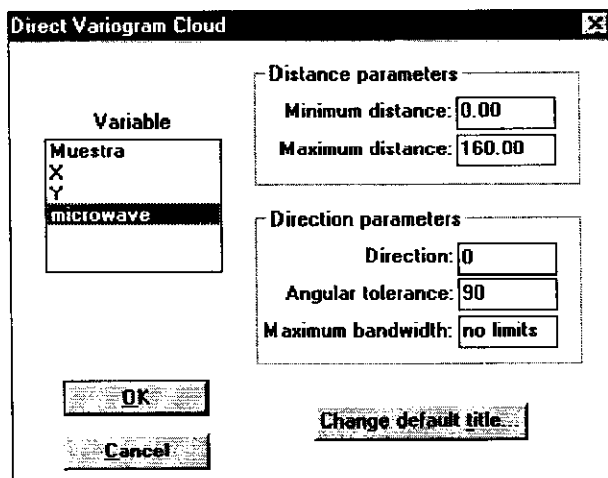


Fig. 18. Parámetros (por defecto) para la nube del variograma.

Si se presiona OK en esta última figura sin modificar los valores por defecto, se obtiene un gráfico consistente en una nube de puntos que muestra la relación entre la magnitud del vector de separación entre pares de puntos y el valor del variograma de esos pares. Los valores de los parámetros de la nube también pueden ser modificados. Para los datos del archivo `backscat.dat` se escogió el valor 161 para la distancia máxima (Maximum distance) como se observa en la nube de variograma (Fig. 19).

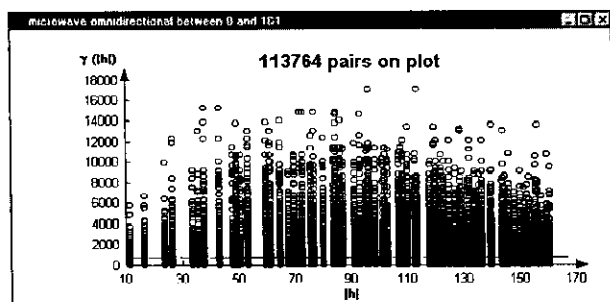


Fig. 19. Nube de variograma de la variable backscatter con distancia máxima 161.

La distancia máxima debe establecerse como aproximadamente la mitad de la distancia entre todos los pares de puntos. El software VARIOWIN no proporciona esta distancia máxima de manera inmediata, pero puede obtenerse en forma indirecta. Elegir la opción `Calculate/Variogram Cloud`, ingresar un valor exagerado, por ejemplo para los datos considerados el valor 1000, en el cuadro de texto:

Maximum distance:

VARIOWIN mostrará una advertencia indicando entre que valores varía la distancia máxima (Fig. 20).

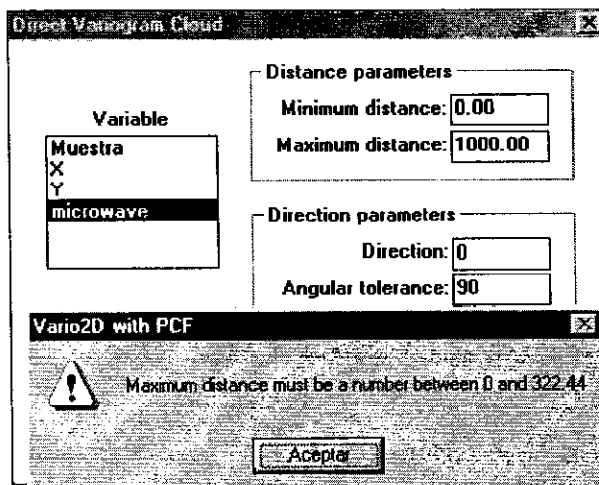


Fig. 20. Advertencia sobre la distancia máxima.

La tolerancia angular, por defecto es de 90° . Esta opción es útil para graficar las nubes de variogramas direccionales, el valor 90 indica un variograma isotrópico, es decir, sin efectos direccionales. El comando `Graph` permite modificar los símbolos, ejes e incluir o no la línea de covarianza en la nube de variograma. Los puntos en la nube y los puntos en el mapa de localización pueden relacionarse, proporcionando una herramienta útil para detectar outliers. Manteniendo visible el gráfico de la nube de variograma y accionando `Data/Map!` se obtiene sobrepuesto el mapa de localización de las muestras; al accionar `Windows/Tile` ambas figuras quedan emparejadas (Figs. 21 y 22).

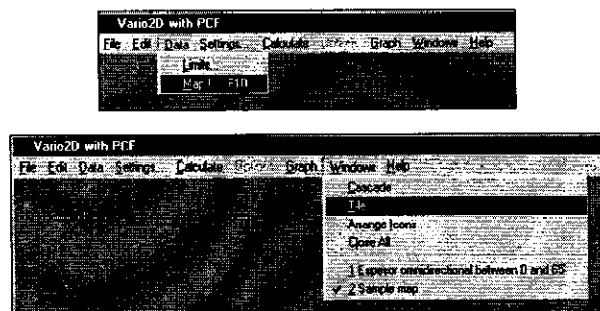


Fig. 21. Comando para conectar la nube de variograma y el gráfico de puntos.

Presionando el mouse sobre cualquier punto de la nube, en particular sobre un outlier, aparece una ventana con la siguiente información: colección de muestras que generan dicho punto en la nube de variograma, distancia de separación entre las muestras (h), valor de la variable (microwave) y valor del variograma.

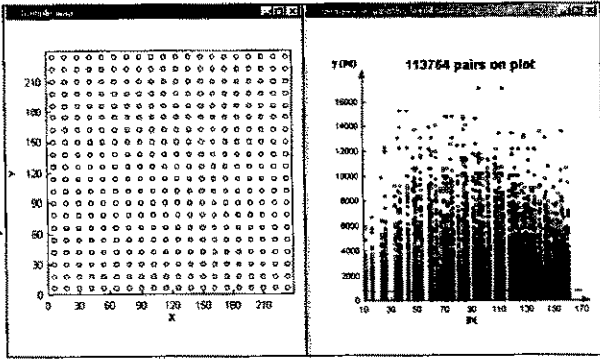


Fig. 22. Nube de variograma y gráfico de puntos conectados.

Simultáneamente en el mapa de localización se observan los puntos muestrales conectados por líneas que aparecen en la pantalla en color rojo (Figs. 23 y 24).

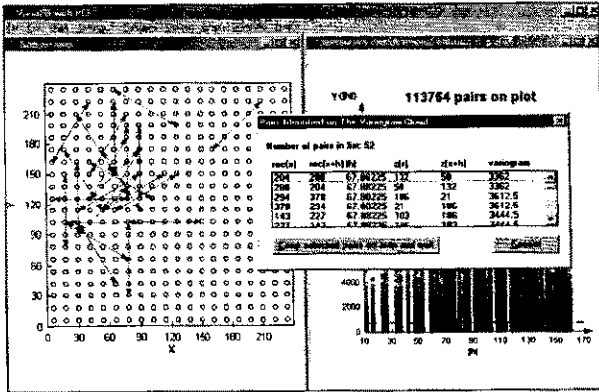


Fig. 23. Punto sobre la nube de variograma y gráfico de muestras conectadas.

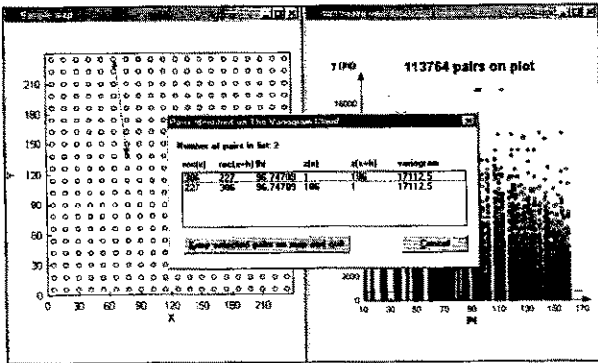


Fig. 24. Outlier en nube de variograma y gráfico de puntos conectados.

La nube puede ser salvada con extensiones de archivos gráficos mediante el comando **Edit - Copy as Metafile** o **Edit - Copy as Bitmap**. Para guardar en una tabla la información contenida en una nube de variograma presionar **File/Save as** y asignar un nombre a la nube de variograma, se crea así un archivo con extensión **.cld**, por ejemplo **backscat.cld**, Fig. 25, al que se puede acceder mediante un editor de texto (Word) (Fig. 26).

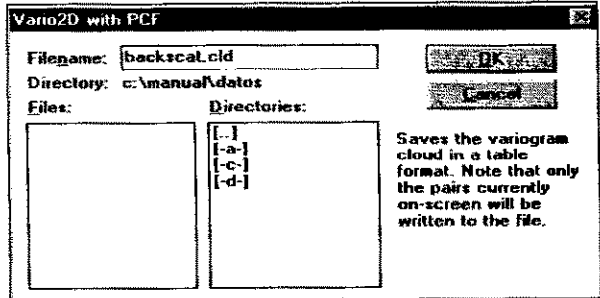
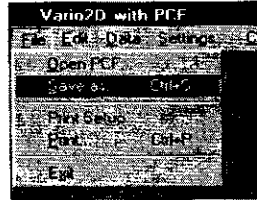


Fig. 25. Comando para salvar la nube de variograma en un archivo .cld.

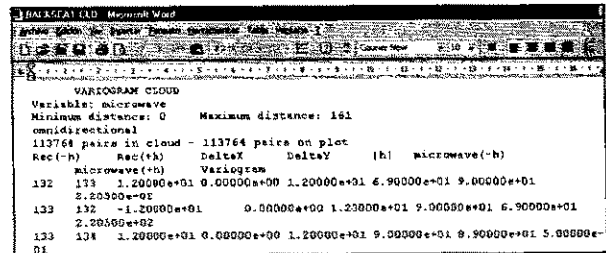


Fig. 26. Tabla conteniendo la información de la nube de variograma.

Con la información proporcionada por la nube variográfica acerca de los valores adecuados para los parámetros del variograma experimental, es posible retomar el cálculo del variograma experimental. Para ello escoger **Calculate/Directional Variogram** tal como indica la Fig. 17, reemplazar los valores por defecto por estos nuevos valores. Para los datos de la variable considerada, **microwave**, en el ejemplo del archivo **backscat.dat** se eligieron los siguientes parámetros: espaciamiento entre lags (**Lag spacing**): 12, número total de lags (**Number of lags**): 12, dirección escogida (**direction**): 0 y tolerancia angular (**Angular tolerance**): 90, obteniéndose el variograma omnidireccional de la Fig. 27. Es posible calcular distintos variogramas direccionales simplemente cambiando los valores de dirección y tolerancia angular.

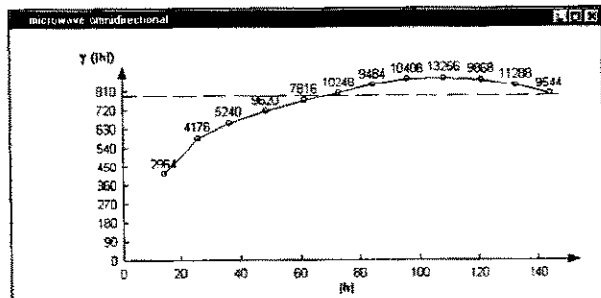


Fig. 27. Variograma experimental omnidireccional de la variable microwave.

El gráfico puede ser salvado con extensiones de archivos gráficos mediante el comando **Edit - Copy as Metafile** o **Edit - Copy as Bitmap**. El comando **Options** permite utilizar otro tipo de estimadores para realizar un análisis de correlación espacial de datos tales como variograma estandarizado, covarianza, correlograma y madograma (Fig. 28).

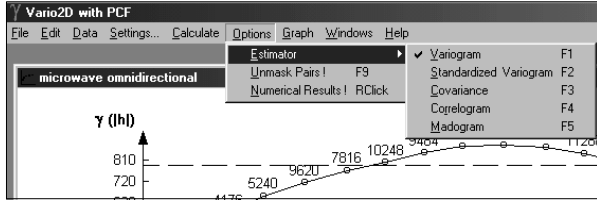


Fig. 28. Menú **Options**: distintos estimadores para realizar un análisis de correlación espacial de datos.

El gráfico de variograma puede representarse en forma de tabla presionando el mouse derecho sobre el propio gráfico o mediante **Options - Numerical Results!** (Fig. 29).

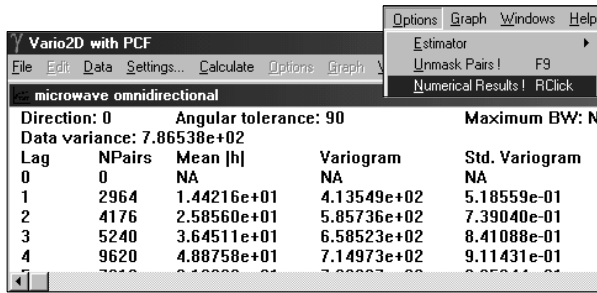


Fig. 29. Menú **Options - Numerical Results!**

Con el comando **Graph** se pueden hacer modificaciones al gráfico respecto de los ejes, símbolos, etc. Al cerrar el programa **Vario2D** la tabla obtenida mediante el comando **Options - Numerical Results!** se cierra también perdiéndose tal información. Para conservar los resultados numéricos del variograma experimental éste debe ser guardado en un archivo con la extensión **.var** (**backscat.var**), Fig. 30, al que se puede acceder mediante un editor de texto (por ejemplo Word). Este archivo será posteriormente utilizado por el programa **Model** (Fig. 31).

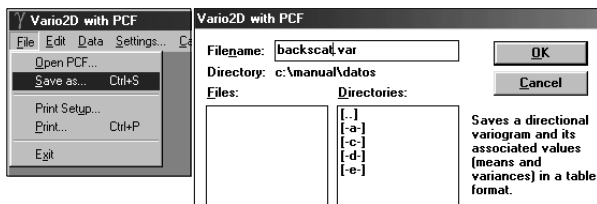


Fig. 30. Comando para salvar el variograma en un archivo **.var**.

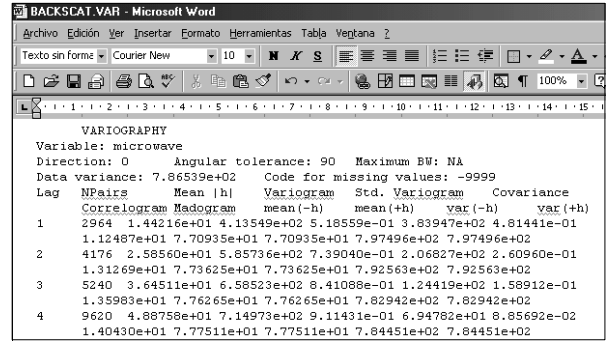


Fig. 31. Tabla conteniendo la información del variograma experimental.

Presionando el mouse sobre cada punto del gráfico del variograma experimental aparece una ventana con el h-scatterplot referido al lag elegido (da información de los pares de puntos correspondientes a dicho lag) (Fig. 32).

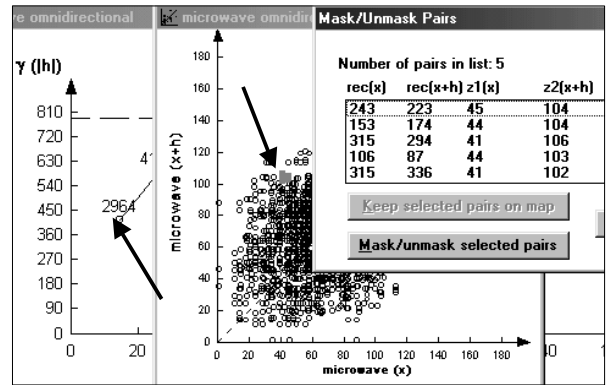


Fig. 32. h-scatterplot para lag seleccionado.

La opción **Calculate/ Variogram Surface** permite generar un mapa de variogramas, presentado en forma de pixels en escala de grises o colores (Fig. 33).

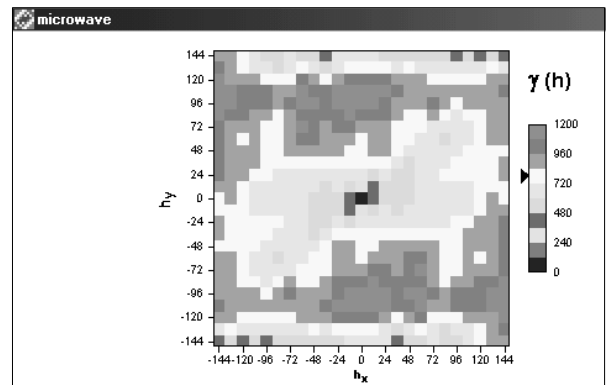


Fig. 33. Mapa de Variograma para la variable **microwave**.

Dicho mapa permite determinar las direcciones de anisotropía de la variable en estudio. El comando **Graph** permite hacer modificaciones al gráfico respecto de la paleta de colores, indicar la cantidad de

pares que forman cada valor de $g(h)$ y valor de covarianza. El comando **Options** permite utilizar otro tipo de estimadores para realizar un análisis de correlación espacial de datos tales como variograma estandarizado, covarianza, correlograma y madograma.

La información del mapa puede representarse en forma de tabla presionando el mouse derecho sobre la figura o mediante **Options – Numerical Results!** (Fig. 34).

DeltaX	DeltaY	NPairs	Mean h	Variogram	Std. Variogr.
-144	-144	64	2.03647e+02	4.49445e+02	9.95927e-01
-132	-144	72	1.95346e+02	5.71167e+02	1.27749e+0
-120	-144	80	1.87446e+02	4.40156e+02	1.02982e+0
-108	-144	88	1.80000e+02	4.88528e+02	1.14026e+0
-96	-144	96	1.73066e+02	4.56401e+02	1.10048e+0
-84	-144	104	1.66709e+02	5.62173e+02	1.31355e+0
-72	-144	112	1.60997e+02	5.12402e+02	1.16226e+0
-60	-144	120	1.56000e+02	5.43717e+02	1.15231e+0
-48	-144	128	1.51789e+02	5.85465e+02	1.24134e+0
-36	-144	136	1.48432e+02	5.54904e+02	1.13896e+0

Fig. 34. Tabla con la información correspondiente al mapa de variograma.

Para guardar toda la información contenida en la tabla del mapa de variograma, presionar **File/Save as** y asignar un nombre de archivo, para el ejemplo: **backscat.vs**, así se crea un archivo con extensión **.vs** (variogram surface) al que se puede acceder mediante un editor de texto (Fig. 35).

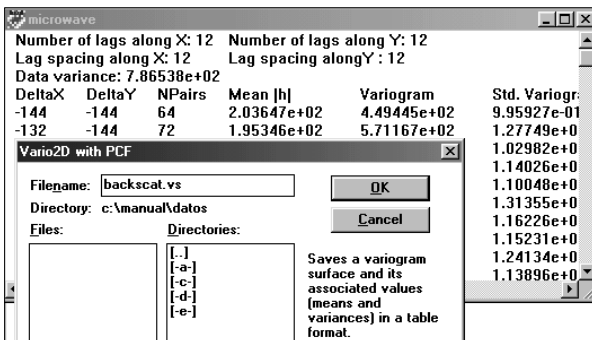


Fig. 35. Comando para salvar la tabla del mapa de variograma en un archivo con extensión **.vs**.

CONCLUSIONES

El software VARIOWIN permite un análisis estructural completo de datos espaciales. Además del análisis gráfico, brinda en forma de tablas información numérica del análisis de estimaciones resultante del cálculo del variograma experimental y de la superficie de variograma, contribuyendo a una mejor elección del modelo y parámetros a ajustar.

Variowin está diseñado tanto para iniciar al principiante en el uso de los métodos geoestadísticos de una manera fácil, como para proporcionar al usuario experimentado suficiente potencia y flexibilidad para resolver problemas reales.

Constituye un buen complemento de otros programas informáticos aplicados a la geoestadística.

BIBLIOGRAFÍA

1. Cressie, N. A. C. (1993). *Statistics for spatial data*, pp.29-104. J. Wiley & Sons.
2. Englund, E. y A. Sparks (1991). *Geo-EAS 1.2.1 User's Guide*, EPA Report #600/8-91/008 EPA-EMSL, Las Vegas, NV, (<http://www.epa.gov/ada/csmos/models/geoeas.html>).
3. Isaaks, E. H. y, R.M. Srivastava (1989). *Applied Geostatistics*, 561pp. New York Oxford, Oxford University Press.
4. Malla, M. S. y M. S. Santamaría (2005). *VARIOWIN: Software para Análisis Variográfico*, 9pp. XXXIII Coloquio Argentino de Estadística, CD-ROM.
5. Moral, F. J. (2004). Aplicación de la geoestadística en las ciencias ambientales. *Ecosistemas* (1), 11pp. <http://www.aeet.org/ecosistemas/041/revision3.htm>
6. Pannatier, Y. (1996). *VARIOWIN: Software for Spatial Data Analysis in 2D*, Springer-Verlag, New York.
7. Webster, R. y M.A. Oliver (1990). *Statistical methods in soil and land resource survey*, pp. 213-240. Oxford University Press, Oxford.