

Geostatística

A. Gallardo

Departamento de Ecología y Biología Animal. Facultad de Biología, Campus de Lagoas-Marcosende, Universidad de Vigo, 36310 Vigo.
Departamento de Sistemas Físicos, Químicos y Naturales, Universidad Pablo de Olavide, Ctra. Utrera km. 1, 41013 Sevilla

Geostatística. En esta mini-revisión se presentan los fundamentos de la geostatística, desde la construcción de un semivariograma hasta la interpolación de puntos en el espacio mediante la técnicas de kriging y cokriging. Mediante la revisión de los principales modelos de semivariogramas utilizados en geostatística, se presentan los parámetros que se pueden extraer del semivariograma, que son los que van a definir la continuidad espacial de la variable de nuestro interés, y los que van a ser utilizados para la interpolación de puntos no muestreados en el espacio. Además se presenta la manera de abarcar un comportamiento anisotrópico de la variable, y la asunción más importante para el uso de técnicas geostatísticas: la estacionariedad. Se discute también la utilización de más de una variable para la descripción y la estimación de valores en el espacio (variograma cruzado y cokriging), y la utilización de técnicas de jackknife para el estudio del efecto que los métodos empleados tienen sobre la interpolación de puntos (validación cruzada).

Palabras clave: Semivariograma, kriging, cokriging, anisotropía, estacionariedad, continuidad espacial.

Geostatistics. In this mini-review the fundamentals of geostatistics are described, from the basic approach to the semivariogram to spatial interpolation methods, as kriging and cokriging. The most common mathematical functions in semivariogram modelling are reviewed, with a description of the parameters extracted from these models which are used to quantify the spatial continuity of a spatial phenomenon and to interpolate values at unsampled points. The different behavior of a variable in different space directions (anisotropy), and the most important assumption in geostatistics (stationarity) are also discussed. In the final part, we explore the use of one or several other variables to improve the knowledge and the interpolation of one spatial phenomenon (cross variogram and cokriging). Finally, cross-validation, a jackknife method to check the model assumptions used in the kriging is discussed.

Key words: Semivariogram, kriging, cokriging, anisotropy, stationarity, spatial continuity.

Introducción

La heterogeneidad espacial a escala de ecosistema y de paisaje ha sido reconocida desde el comienzo del desarrollo de la ecología. La ecología del paisaje en sí no es más que el reconocimiento explícito de esta heterogeneidad. La fitosociología se desarrolló también con la intención de simplificar, en realidad convertir en unidades discretas, la heterogeneidad espacial observada en las comunidades vegetales. En contraste a la fitosociología, la ecología ha desarrollado una serie de técnicas más encaminadas al reconocimiento de la continuidad en las comunidades vegetales, sin olvidar la existencia de discontinuidades, y de todo ello ha surgido el estudio de los análisis de gradientes, fronteras, ecotonos etc. La necesidad de cuantificar estadísticamente el grado y la escala espacial en que cambian las comunidades vegetales no ha sido prioritario, entre otras cosas porque esta heterogeneidad era percibida visualmente, y por ello era espacialmente predecible. Sin embargo, los geólogos e ingenieros de minas en la segunda mitad del siglo XX se encontraron con la necesidad de desarrollar herramientas estadísticas que cuantificaran el grado y escala de variación espacial de recursos mineros que no podían percibir visualmente, pero cuyo patrón espacial necesitaban conocer para incrementar la eficiencia en la explotación de dichos recursos. Este desarrollo llevó a la creación de la geostatística, que pronto fue utilizada por otras ciencias para cuantificar la heterogeneidad espacial de variables no fácilmente perceptibles. El uso de geostatística en ecología fue introducida por primera vez de forma explícita por Robertson (1987): "*Geostatistics in ecology: interpolating with known variance*"; aunque fue desarrollada con mayor detalle por Rossi *et al.*, (1992): "*Geostatistical tools for modeling and interpreting ecological spatial dependence*". Junto a estos dos trabajos ya clásicos hay también que incluir el trabajo de Legendre y Fortin (1989): "*Spatial pattern and ecological analysis*", que aunque no es específico de geostatística, si la incluye como una de las herramientas para la descripción de patrones espaciales en ecología.

La geostatística ha permitido cuantificar la escala y grado de variación espacial de recursos para plantas y animales y su relación con la distribución de los organismos. Esta variación espacial es clave para explicar procesos ecológicos a diferentes escalas espacio-temporales. Uno de los procesos más claros es la relación entre diversidad y heterogeneidad espacial. Por ejemplo, una especie de crecimiento rápido puede llegar a sombrear y desplazar competitivamente a algún vecino de otra especie de crecimiento más lento. El resultado de la competencia, estimado en condiciones homogéneas de laboratorio, siempre sería el mismo: la especie de mayor crecimiento desplaza a su vecina de menor crecimiento. Sin embargo, en condiciones naturales algunos individuos de la especie de crecimiento más rápido podrían haber germinado en sitios (o micrositios) deficientes en recursos, igualando en crecimiento a la especie más lenta y permitiendo la coexistencia de ambas especies en un espacio relativamente pequeño. El uso de geostatística en ecología pretende en último término entender las interacciones entre las especies y sus recursos en ambientes heterogéneos.

¿Para qué es útil la geostatística?

La geostatística es una manera de describir la continuidad espacial de cualquier fenómeno natural. Con ello llegamos a conocer la forma en que varía cualquier variable continua en el espacio (patrón espacial) a una o varias escalas seleccionadas, con un nivel de detalle que permite cuantificar la variación espacial de la variable en distintas direcciones del espacio. La geostatística utiliza funciones para modelar esta variación espacial, y estas funciones son utilizadas posteriormente para interpolar en el espacio el valor de la variable en sitios no muestreados. La fortaleza de la geostatística es que esta interpolación (conocida como kriging) es considerada una estima muy robusta ya que se basa en la función continua que explica el comportamiento de la variable en las distintas direcciones del espacio, y que en contraste con otros métodos de interpolación (como por ejemplo interpolar un punto usando los valores de los puntos que le rodean ponderados por la distancia que los separa) permite asociar la variabilidad de la estima (conocido como grado de incertidumbre). La geostatística permite por tanto responder a las siguientes preguntas: ¿Cuál es el patrón espacial de mis variables de interés? ¿A qué escala se repite este patrón espacial? ¿Existe covariación espacial entre las distintas variables de interés? ¿Cuál es la mejor representación gráfica de la continuidad de mi variable? ¿Cuál es el grado de incertidumbre de estas estimas? Las respuestas a estas preguntas son siempre dependientes de la escala espacial elegida. Por ejemplo, si queremos relacionar la distribución de especies de plantas con el contenido en humedad del suelo, los resultados serán distintos si muestreamos en un cuadrado de 5 x 5 m que si lo hacemos en un cuadrado de 100 x 100 m. Nos encontraremos dos diferentes patrones espaciales (aunque el mayor engloba, y a veces oculta al menor) y dos diferentes grados de incertidumbre que podremos asociar a la respuesta a nivel de individuo a la variación en humedad (parcela pequeña) o a la respuesta de poblaciones o comunidades a dicha variación (parcela grande).

El semivariograma empírico

La función básica que describe la variabilidad espacial de un fenómeno de interés se conoce como semivariograma. El semivariograma responde a la siguiente pregunta ¿Cómo de parecidos son los puntos en el espacio a medida que estos se encuentran más alejados? Imaginemos la parcela de la **Figura 1**.

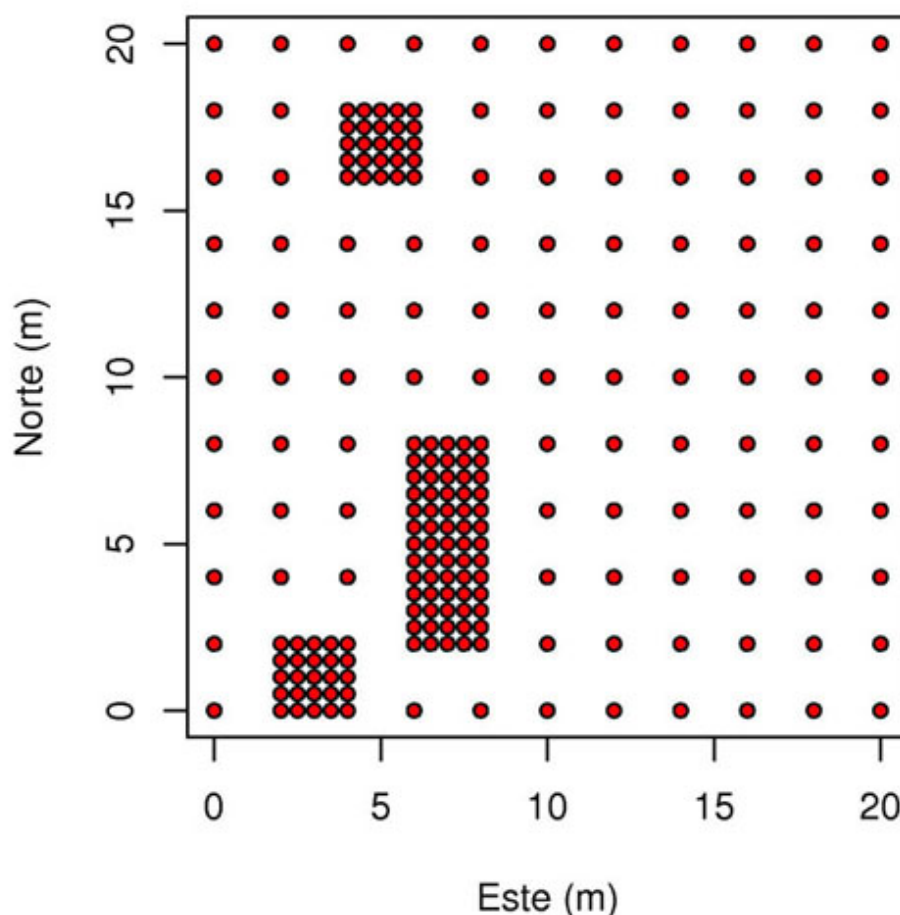


Figura 1. Localización de puntos de muestreo en una parcela de 20 x 20 metros. Las muestras fueron tomadas cada dos metros en toda la parcela y cada 50 cm en cuatro parcelas situadas aleatoriamente en un diseño anidado.

Se ha muestreado un cuadrado de 20 x 20 metros alrededor del tronco de una encina en una Dehesa de Sevilla, tomándose muestras de suelo cada 2 m, y en cuatro cuadrados aleatorios dentro de la parcela, muestras separadas por 0,5 m (Gallardo *et al.*, 2000). Para cualquier variable analizada, podemos calcular la varianza encontrada entre todos los pares de puntos separados por 0.5 m, englobando tanto los separados por esta distancia en el eje N-S como en el eje E-O. En segundo lugar podríamos calcular la varianza encontrada entre todos los pares de puntos separados por 2 m, 4 m, y así sucesivamente hasta calcular la varianza encontrada entre todos los pares separados 20 m (que serían tan solo 4). También podríamos intercalar las distancias entre muestras separadas en diagonal; la más pequeña sería la hipotenusa del cuadrado de 0,5 x 0,5 m, y la más grande la hipotenusa del cuadrado 20 x 20 m (del que sólo habría dos pares). La representación gráfica de todas estas varianzas en función de la distancia que separa a las muestras es el semivariograma (o variograma), y el cálculo de la varianza entre pares separados por intervalos de distancia se conoce como semivarianza (γ), estimada como:

$$\gamma(h) = \frac{1}{2} N_{(h)} \sum [Z_{(x)} - Z_{(x+h)}]^2$$

Donde, $\gamma(h)$ es la semivarianza para todas las muestras localizadas en el espacio separado por el intervalo de distancia h . $N_{(h)}$ es el número total de pares de muestras separados por un intervalo de distancia h . $Z_{(x)}$ es el valor de la muestra en una localización x . $Z_{(x+h)}$ es el valor de la muestra a la distancia de intervalo h desde x .

La representación del semivariograma para la concentración de materia orgánica en el suelo en el ejemplo anterior se observa en la **Figura 2**. Como se puede apreciar, esta figura no representa todos los intervalos de distancia posibles, ya que solo llega a 12 m, siendo la parcela de 20 x 20 m. Este aspecto es importante, ya que como regla general sólo se representa en el semivariograma distancias aproximadas a la mitad de la dimensión de la parcela. La razón se puede encontrar en lo ya visto, el número de pares a las distancias mayores descende y la semivarianza puede llegar a mostrar un comportamiento errático. Esto nos lleva a otra regla general: el número mínimo de pares para representar un punto en el semivariograma debe ser superior a 30. La tercera regla general es que el número de puntos en el espacio objeto de estudio no debería ser inferior a 50.

Estas reglas generales sirven de guía, pero no deben ser tomadas como un dogma insalvable. Se encuentran trabajos publicados con menos de 50 puntos en el espacio (aunque no muchos menos), y con semivariogramas que recogen distancias que superan el 50% de la dimensión de la parcela sin mostrar comportamiento anómalo.

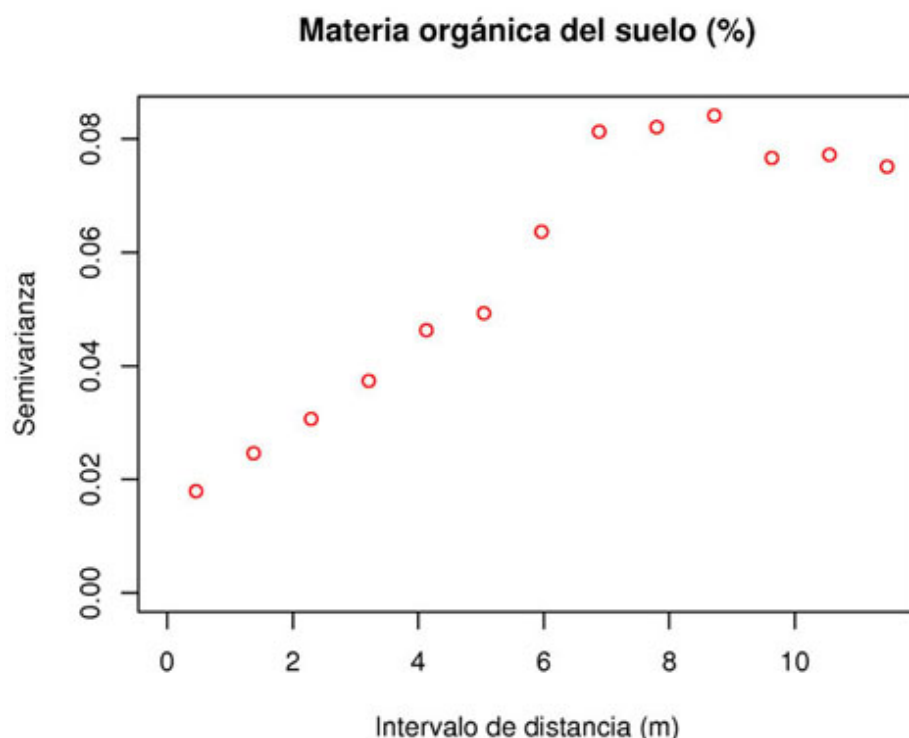


Figura 2. Semivariograma empírico mostrando la semivarianza de valores de materia orgánica del suelo separados por distancias crecientes.

La **Figura 2** ya nos da bastante información del comportamiento espacial de la variable, los valores de la muestra separados por intervalos de distancia entre 0 y 1 m son más parecidos que aquellos separados por 2 m, 3 m etc. La semivarianza aumenta a medida que aumenta la distancia que separa las muestras hasta que se llega a un intervalo de distancia aproximada de 8 m, a partir del cual la semivarianza alcanza un máximo y no aumenta más. Este nivel máximo de semivarianza debe coincidir con la varianza de la población. En este caso, los valores de semivarianza del eje Y se corresponden al logaritmo de la materia orgánica, transformación realizada para ajustar los valores a una distribución normal. Aunque la transformación de la variable no es imprescindible para construir un semivariograma, si lo es para el uso del semivariograma en la interpolación de valores de la variable, ya que todas las técnicas de interpolación son sensibles a valores extremos.

Ajuste de una función al semivariograma

El semivariograma de la **Figura 2** proporciona bastante información del comportamiento espacial de la variable. Sin embargo, es necesario ajustar una función para cuantificar el grado y escala de variación espacial. Existen numerosos modelos que se utilizan en geostatística, siendo los más comúnmente usados el modelo esférico, el modelo exponencial, el modelo gaussiano y el modelo lineal. Estos modelos se pueden observar en la **Figura 3**, en donde se han ajustado a los datos del semivariograma empírico de la **Figura 2**. El ajuste a una función permite extraer una serie de parámetros que son los que van a ser usados para la interpolación geostatística (*kriging*) y que definen el grado y escala de variación espacial. Estos parámetros son el rango (A_0), el *nugget* (C_0), el *sill* (C_0+C), y la proporción de la varianza explicada por el espacio (C/C_0+C), a menudo expresada en porcentaje (**Fig. 4**). El rango (A_0) es la distancia a la que la semivarianza deja de aumentar. El rango, por tanto, indica la distancia a partir de la cual las muestras son espacialmente independientes unas de otras, y representa el tamaño de grano o mancha que representa la variable (Paramá, 2006). Por ejemplo, en el caso de la materia orgánica de la parcela de Dehesa (**Fig. 2 y 3**), el rango es aproximadamente de unos 8 m, coincidiendo con el diámetro de la copa de una encina que ocupa una parte de la parcela, y que condiciona los niveles de materia orgánica en el suelo. El *nugget* (C_0) es la varianza no explicada por el modelo, y se calcula como la intercepción con el eje Y. Se conoce también como varianza error puesto que la varianza de dos puntos separados por 0 metros (la intercepción con el eje Y) debería ser cero. Es por ello que esta varianza está normalmente indicando variabilidad a una escala inferior a la muestreada. Además, los errores analíticos o

de muestreo también contribuyen a la aparición de la varianza error. La máxima semivarianza encontrada entre pares de puntos se conoce como *sill* y debe coincidir con la varianza de la población. $C/(C_0 + C)$ nos da el grado de variación espacial, y por tanto el grado de incertidumbre a la hora de interpolar puntos en el espacio. Un alto cociente nos indica una variable espacialmente muy predecible.

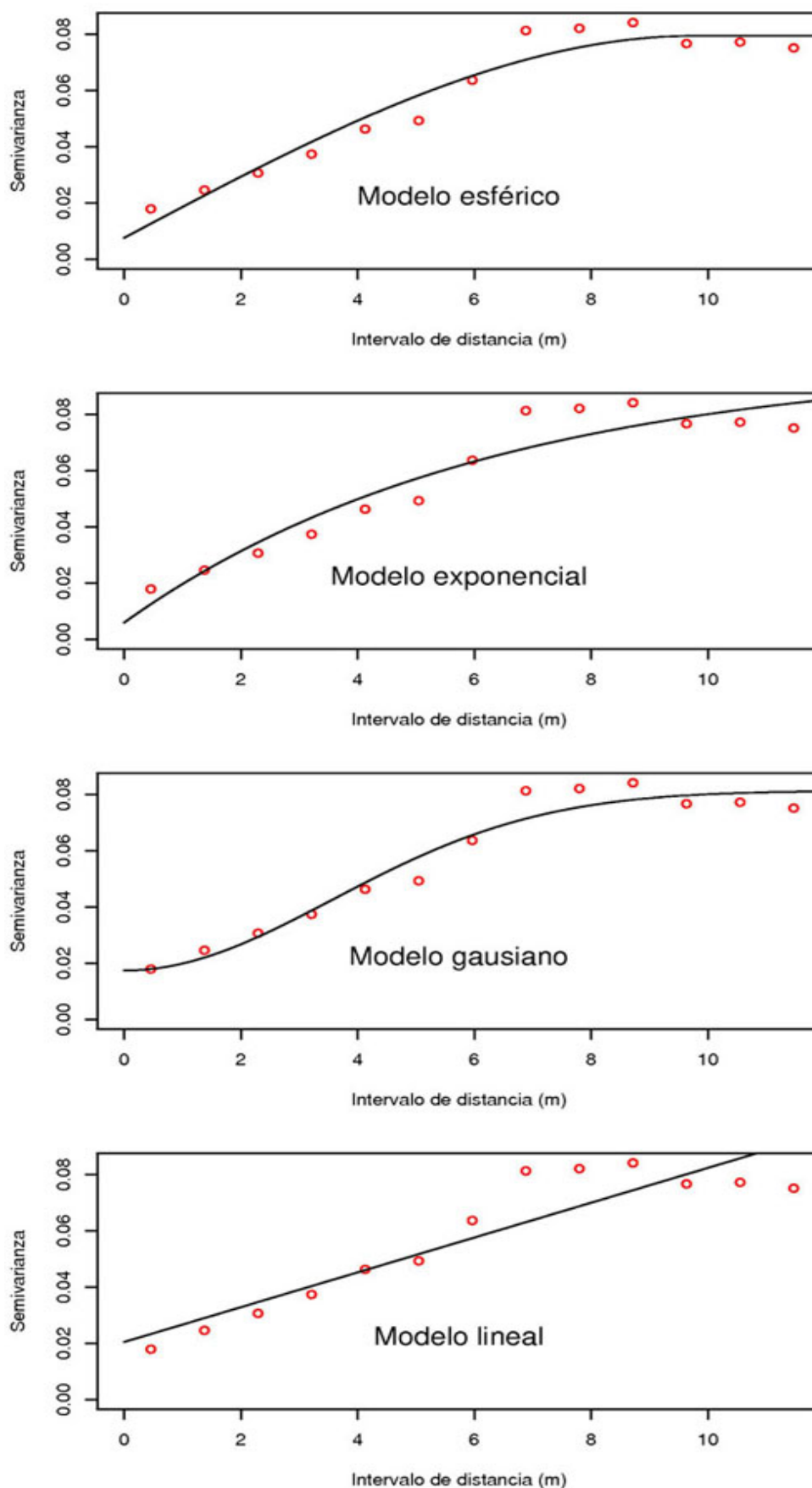


Figura 3. Ajuste del semivariograma empírico de la figura 2 a cuatro funciones matemáticas.

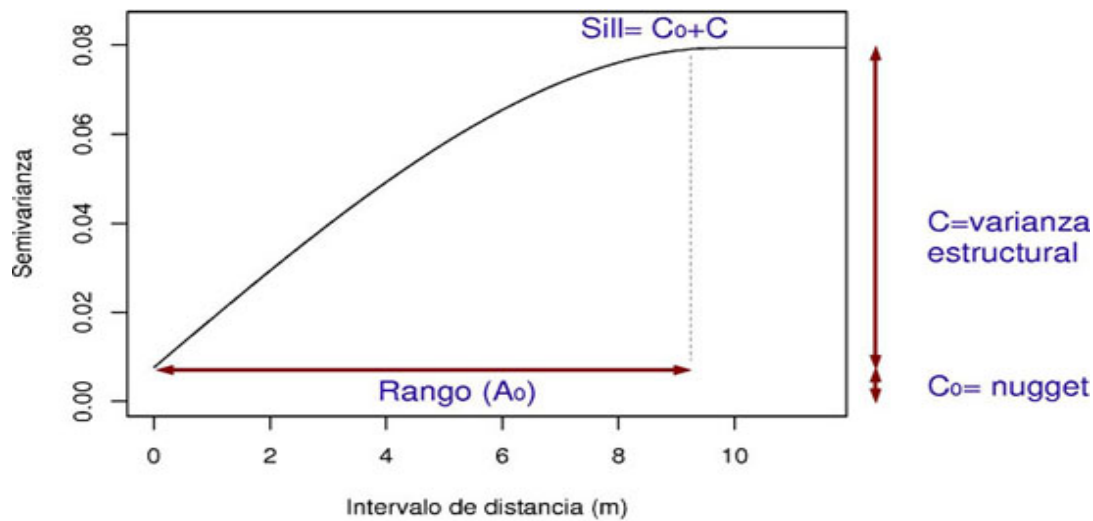


Figura 4. Parámetros utilizados en el ajuste del semivariograma a funciones matemáticas.

Las ecuaciones de los modelos más empleados son las siguientes:

Modelo esférico:

$$\gamma(h) = \begin{cases} 1.5 \frac{h}{a} - 0.5 \left(\frac{h}{a} \right)^3 & \text{si } h \leq a \\ 1 & \text{en cualquier otro caso} \end{cases}$$

Donde $\gamma(h)$ es la semivarianza en el intervalo de distancia h , y a es el rango (A_0). Este modelo tiene un comportamiento lineal a distancias de separación pequeñas cerca del origen pero se va aplanando a mayores distancias y alcanza el *sill* en la distancia a .

Modelo exponencial:

$$\gamma(h) = 1 - \exp\left(\frac{-3h}{a}\right)$$

Este modelo tiende a alcanzar el *sill* asintóticamente. El rango a es definido como la distancia a la cual el valor del variograma alcanza el 95% del *sill*.

Modelo gaussiano:

$$\gamma(h) = 1 - \exp\left(\frac{-3h^2}{a^2}\right)$$

Al igual que el modelo exponencial, el modelo tiende a alcanzar el *sill* asintóticamente, y el rango se define como la distancia a la cual el variograma alcanza el 95% del *sill*.

Modelo lineal:

$$\gamma(h) = C_0 + bh$$

Donde C_0 es el nugget y b la pendiente de la recta.

¿Cuál es el mejor modelo para mis variables?

Los criterios para seleccionar un modelo u otro dependen de los objetivos del trabajo. Si el objetivo es encontrar el modelo que mejor se ajuste al semivariograma empírico para cada variable, y no tenemos información *a priori*, podemos dejar que un determinado *software* nos ajuste automáticamente el mejor modelo basado, bien en el mínimo de la suma de los cuadrados de los residuales, bien en el R^2 de la ecuación. Si tenemos información *a priori* del comportamiento de nuestra variable, puede ser interesante realizar un ajuste manual de los modelos al semivariograma empírico, lo que permiten algunos programas de *software*. De esta forma el investigador puede fijar el *nugget*, el *sill* o el rango dependiendo del tipo de información que tenga de su variable y ajustar los parámetros de los que no tiene información. Aquí hay que notar que el ajuste automático basado en los menores residuales o en los mayores R^2 no necesariamente producen modelos con mayor significado biológico que el ajuste manual. A veces pequeñas diferencias en los menores residuales o en el R^2 no justifican la elección de un modelo u otro. Si el objetivo del trabajo es comparar los parámetros de los semivariogramas entre distintas variables o cambios en el semivariograma con el tiempo o en el espacio, la utilización de modelos diferentes resulta poco útil. Hay que tener en cuenta que, por ejemplo, los rangos del modelo esférico y el exponencial no son directamente comparables. El modelo esférico es el único que tiene un *sill* verdadero, ya que tanto el modelo exponencial como el gaussiano alcanzan el *sill* de forma asintótica, o lo que es lo mismo, no lo alcanza nunca. Estos tres modelos anteriores se conocen como modelos transicionales porque en ellos se puede estimar el *sill*, sea verdadero o no. El modelo lineal (al igual que otros modelos aquí no considerados) ni siquiera tiene *sill*, no es un modelo transicional. Los rangos (la distancia a la que se alcanza el *sill*) no son, por lo anteriormente expuesto, directamente equivalentes entre modelos. En este caso, es más conveniente elegir, cuando sea posible, un único modelo con motivo de comparar semivariogramas. El modelo esférico es el más usado, porque tiene verdadero *sill*. En segundo lugar el exponencial sobre el gaussiano, porque aunque este último refleja muy bien la continuidad espacial, la interpolación de puntos basada en este modelo es muy exigente con respecto a los valores de entrada, produciendo frecuentemente representaciones gráficas alejadas de la realidad. Por último el modelo lineal es usado para reflejar una pobre estructura espacial, o una estructura espacial cuya dimensión supera la de la parcela de estudio (por ejemplo, la parcela está dentro de un gradiente direccional).

Isotropía y anisotropía

Hasta ahora en los semivariogramas vistos se considera que la variación del valor de nuestra variable con el espacio es igual en todas las direcciones de éste (semivariograma omnidireccional). Si esto ocurre decimos que la variable tiene un comportamiento isotrópico. Pero no siempre es así, y puede ser que la variación espacial sea diferente en las distintas direcciones del espacio (anisotropía). Si tras una inspección visual sospechamos que puede ocurrir este fenómeno es interesante realizar semivariogramas considerando por separado varias direcciones del espacio (semivariogramas direccionales). Así en la **Figura 1** podríamos sólo considerar los pares de puntos separados por distancias en el eje N-S, NE-SO, E-O y NO-SE para realizar cuatro semivariogramas diferentes y observar la posible existencia de anisotropía. La construcción de semivariogramas anisotrópicos requiere un ángulo de tolerancia, de forma que todos los puntos de la parcela sean usados. Por ejemplo, si realizamos cuatro semivariogramas correspondientes a 0° , 45° , 90° y 135° , y añadimos un ángulo de tolerancia de 22.5° , todos los puntos de las parcelas son usados en uno u otro semivariograma. El resultado lo podemos ver en la **Figura 5**.

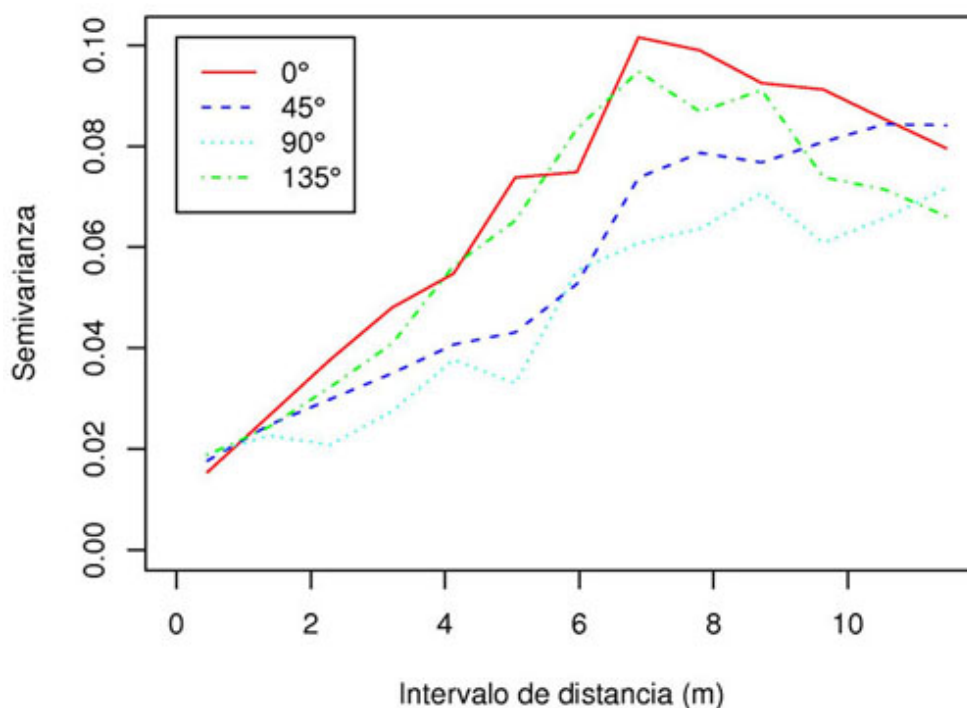


Figura 5. Semivariogramas mostrando las diferencias en continuidad espacial de la materia orgánica del suelo en cuatro direcciones del espacio.

De esta figura se extraen dos conclusiones. Primero, no existe un claro comportamiento anisotrópico en la distribución de materia orgánica en la escala muestreada, ya que los rangos y *nuggets* son aproximadamente similares. A pesar de ello, en las direcciones NE-SO (45°) y E-O (90°) el *sill* alcanza valores menores que en las otras dos direcciones, siendo también su comportamiento en las distancias mayores diferente. Si el *sill* disminuye, pero los *nugget* son similares el porcentaje de varianza explicada por el espacio ($C/C_0 + C$) también disminuye en las direcciones 45° y 90°. Siendo así, los puntos situados en esas direcciones deberían tener un peso menor a la hora de la interpolación que puntos situados en las otras dos direcciones. En segundo lugar, los semivariogramas direccionales presentan peor ajuste que el omnidireccional, lo que es de esperar debido al menor número de pares de puntos. Por tanto, si el número de puntos muestreados no es lo suficientemente grande, el aumento de información que supone los semivariogramas direccionales desaparece debido a la dificultad de encontrar modelos que recojan fielmente la variabilidad espacial de cada dirección. Cuando los rangos se mantienen aproximadamente igual en distintas direcciones del espacio pero el *sill* varía se denomina anisotropía zonal. El caso inverso se conoce como anisotropía geométrica. En esta última está más justificada la utilización de un modelo isotrópico que agrupe todas las direcciones, ya que se considera que todas estas siguen el mismo modelo básico de continuidad espacial. En la práctica es común encontrar una mezcla de ambas anisotropías.

Estacionariedad

El empleo de técnicas geostatísticas requiere la asunción de estacionariedad de segundo orden, es decir, al menos la varianza debe ser igual en las diferentes zonas del área de estudio. La falta de estacionariedad puede deberse bien a la existencia de anomalías en el espacio, bien a la existencia de una tendencia o gradiente espacial cuya dimensión es mayor que el área de estudio. La estacionariedad puede ser un problema a la hora de la interpolación de puntos en el espacio pero no justifica el abandono de la geostatística a favor de otras técnicas de interpolación (como la técnica del inverso de la distancia) ya que son igualmente sensibles a la falta de estacionariedad (Isaaks y Srivastava, 1989). Existen varias formas de evitar que la no estacionariedad de los datos afecte a la estima de puntos en la parcela. Puede que esta falta de estacionariedad se deba a la existencia detectable de dos poblaciones dentro del mismo espacio de muestreo. En este caso, lo más conveniente es dividir el espacio en estas dos poblaciones, realizar semivariogramas e interpolaciones para cada una de ellas para después unir el resultado en un único mapa. Otra aproximación al problema es restringir el radio de búsqueda de vecinos que ayuden a interpolar un valor en una zona no muestreada. Esta aproximación se basa a que en la mayoría de los casos la estacionariedad es "global" pero no se encuentra estacionariedad "local" con lo que restringiendo el uso de vecinos a distancias convenientemente cortas puede llevar a estimaciones robustas de la variable en el espacio. Por último, si la estacionariedad está provocada por una tendencia espacial más que por la existencia de dos poblaciones, se puede eliminar dicha tendencia (*detrending*) y realizar el semivariograma solo con los residuales. Hay que tener en cuenta que una tendencia espacial puede enmascarar la heterogeneidad local de nuestra área de estudio. Es esta heterogeneidad local en la que normalmente se centra nuestro interés. Si conocemos el origen de esta tendencia (por ejemplo una fuerte pendiente o un

gradiente de inundación) y podemos modelizarla (tendencia externa), tan solo hay que sustraer a los datos espaciales dicha tendencia, interpolar puntos usando el semivariograma de los residuales y añadir la tendencia al resultado final. También se puede explorar la tendencia a partir de los propios datos de la parcela (tendencia interna) mediante la sustracción de polinomios de primer y segundo grado, lo que realizan una buena parte del software geostadístico disponible. Un ejemplo del efecto que puede tener una tendencia espacial sobre la estimación de puntos se puede ver en la **Figura 6**, que se corresponde con el contenido en materia orgánica en el suelo de un bosque de inundación (redibujado de Gallardo, 2003). En el gráfico de la izquierda (a) se puede apreciar un gradiente de materia orgánica que aumenta en la dirección SO-NE. Este gradiente se debe a que la probabilidad de inundación aumenta a medida que se acerca a una masa de agua situada al NE de la parcela, lo que produce frecuentes condiciones anaeróbicas donde la materia orgánica se acumula. Sin embargo, este mapa no es del todo correcto, ya que la interpolación de los puntos se ha realizado partiendo del semivariograma con los valores brutos de la variable. El mapa refleja bien el gradiente, pero enmascara heterogeneidad local producida por pequeñas depresiones microtopográficas donde el agua y las condiciones anaeróbicas permanecen por más tiempo, siendo también probable encontrar mayor acumulación de materia orgánica. Estas condiciones sí se reflejan de forma más certera en el gráfico de la derecha (b). Aquí el mapa se ha construido a partir de un semivariograma con los residuales resultantes de sustraer a los valores brutos la recta que relaciona los valores de materia orgánica con la distancia a la masa de agua (tendencia externa). Al final de la interpolación se le ha añadido dicha tendencia, con lo cual el mapa refleja tanto el patrón de heterogeneidad global (tendencia) como la local (microtopografía).

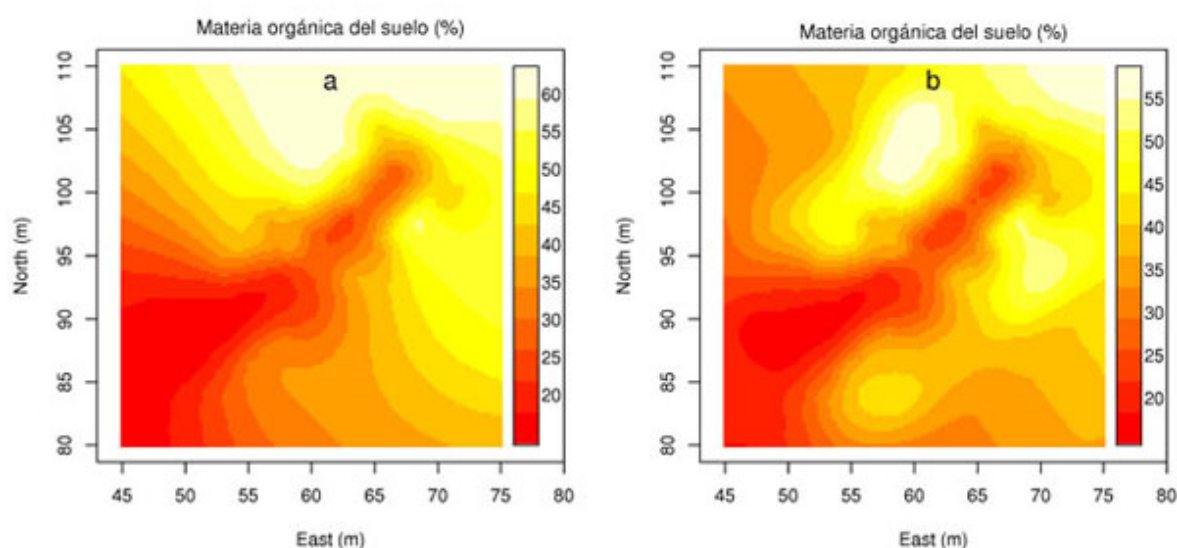


Figura 6. Mapa producto de la interpolación de valores de materia orgánica en un bosque de inundación. a) Interpolación a partir de un semivariograma construido con los valores de materia orgánica. El mapa refleja sobre todo un gradiente SO-NE. b) Interpolación a partir de un semivariograma de los residuales resultantes de sustraer una tendencia espacial (distancia a la masa de agua). El mapa refleja tanto el gradiente espacial como fuentes de heterogeneidad local.

Kriging

Kriging es la técnica de interpolación utilizada en geostatística, con la que se ha estimado los mapas de la **Figura 6**. Su fortaleza estriba en el conocimiento del comportamiento de la variable en el espacio. Así, la forma del semivariograma nos indica la capacidad predictora que tiene cada punto en función de la distancia que lo separa con otro punto. Los puntos más allá del rango, es decir cuando el semivariograma se vuelve plano, tienen la mínima capacidad predictora. A la hora de realizar un kriging hay que, además de suministrar los parámetros del semivariograma isotrópico o los anisotrópicos, decidir la estrategia adecuada para la selección de puntos para la interpolación. Si el número de valores en el espacio es suficiente, disminuir el radio de búsqueda de puntos puede ser la mejor estrategia, ya que evita problemas de estacionariedad, aunque hay que procurar que el radio de búsqueda no sea inferior al rango, ya que no se utilizaría toda la información que proporciona el semivariograma. Otra decisión a tomar es el tipo de interpolación: puntual o por bloques. Mientras que la interpolación puntual es la estima del valor de la variable en un punto del espacio, en la interpolación por bloques esta estima se corresponde con la media de un área predeterminada que rodea a ese punto. En la mayoría de los casos la interpolación por bloques (que produce un “suavizado” de las estimas) correlaciona mejor con los valores verdaderos, siendo generalmente más exacta que la interpolación puntual (Isaaks y Srivastava, 1989). El resultado final del *kriging* es un mapa con los valores interpolados de la variable. Sin embargo, a diferencia de otras técnicas, la geostatística permite que cada interpolación lleve asociado un grado de incertidumbre que puede también ser representado en el espacio (en forma de varianza o desviación estándar). Por tanto a cada punto del espacio interpolado se le puede asociar una distribución teórica, lo que además permite

la posibilidad de realizar simulaciones probabilísticas, representando el resultado del *kriging* como la probabilidad de que la variable alcance un determinado valor.

Variograma cruzado y cokriging

Hasta ahora se ha tratado de modelar la variación de un fenómeno espacial a partir de los valores de esa variable tomados del espacio. Sin embargo, es posible aprovechar la información espacial que contiene otra variable que covarie con la primera. De esta forma se puede construir un semivariograma en donde la varianza representada no sea entre puntos de la misma variable sino de una variable con respecto a otra. Esta representación se conoce como variograma cruzado. Si el variograma cruzado da una estructura interpretable, esta información puede usarse para la predicción de la primera variable en una técnica que se conoce como *cokriging*. La información que tiene una variable sobre otra es siempre menor que la que tiene una variable sobre si misma. Por ello el *cokriging* rara vez mejora la predicción del *kriging* salvo que una de las dos variables haya sido muestreada con menor intensidad que la otra. En este caso el *cokriging* puede ser muy útil. Por ejemplo, cuando todo lo demás es igual (sobre todo la textura), la humedad del suelo va a depender del contenido de materia orgánica del mismo. La humedad del suelo es relativamente fácil de medir, mientras que la materia orgánica requiere de recogida de suelos y un posterior esfuerzo analítico en el laboratorio. Una estrategia no exenta de riesgos podría ser muestrear con menos intensidad la materia orgánica para posteriormente ayudarse de la variación espacial de la humedad del suelo utilizando el *cokriging*.

Validación cruzada

Existe una forma de comprobar el efecto de todas las decisiones tomadas en los métodos de estimación de la variable en el espacio. El método se conoce como validación cruzada y consiste en eliminar un valor de la variable, calcular el semivariograma correspondiente y estimar el valor eliminado a partir de dicho semivariograma (es una técnica de *jackknife*). Si esto lo hacemos uno por uno con todos los valores de las variables, finalmente podremos representar todos los valores interpolados frente a sus valores reales (**Fig. 7**). El beneficio real de una validación cruzada es el de señal de aviso. El estudio de los resultados de la validación cruzada debe concentrarse en los aspectos negativos, como en errores muy abultados o áreas con evidente sobre o subestimación. No se debe utilizar los residuales de la validación cruzada para la mejora automática del modelo de variograma, ya que podría llevar a un modelo “mejorado” que realmente produce resultados peores. Para algunos autores es siempre preferible el ajuste de los variogramas usando información cualitativa que este tipo de ajuste automático (Isaaks and Srivastava 1989).

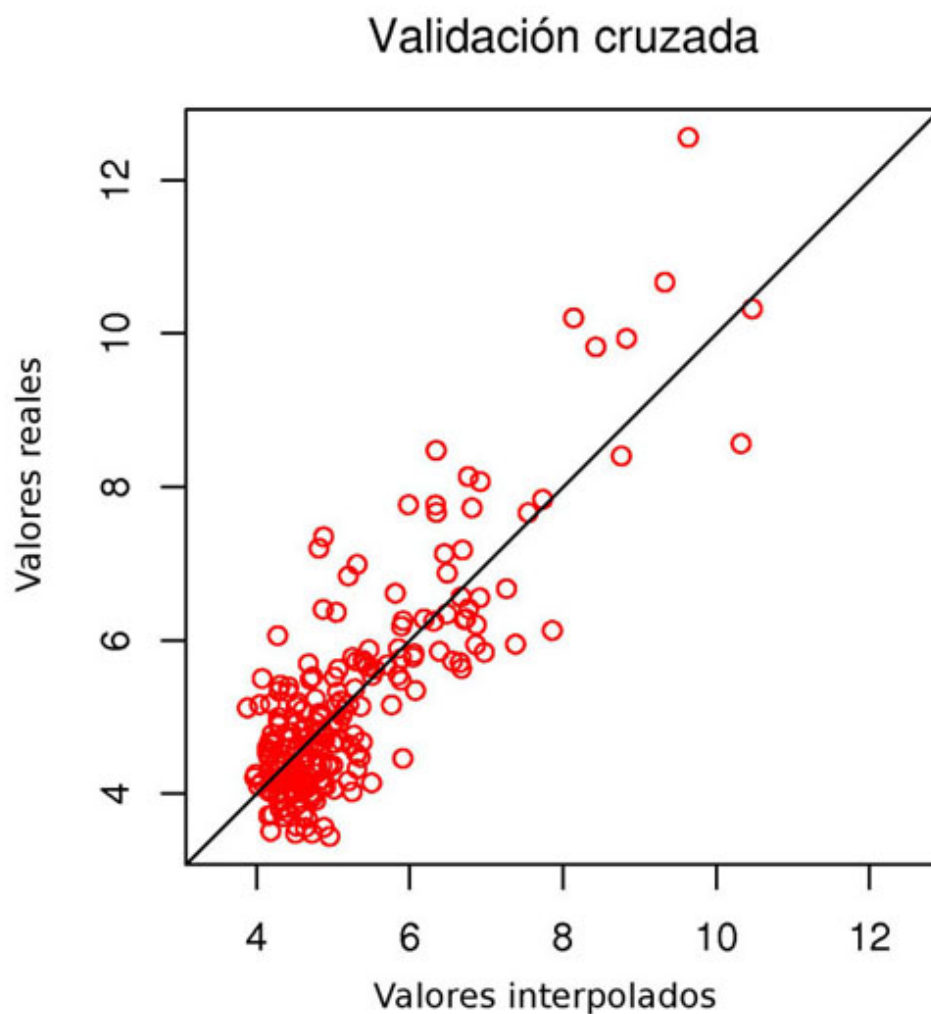


Figura 7. Relación entre los valores reales y los valores interpolados utilizando una validación cruzada.

Referencias

- Gallardo, A., Rodríguez-Saucedo J.J., Covelo, F., Fernández-Alés, R. 2000. Soil nitrogen heterogeneity in a Dehesa ecosystem. *Plant and Soil* 222: 71-82.
- Gallardo, A. 2003. Spatial Variability of Soil Properties in a Floodplain Forest in Northwest Spain. *Ecosystems* 6: 564 – 576.
- Isaaks, E.H., Srivastava R.M. 1989. An introduction to applied geostatistics. Oxford University Press, Nueva York.
- Legendre, P., Fortin M-J. 1989. Spatial pattern and ecological analysis. *Vegetatio* 80: 107-138.
- Paramá, R. 2006. *Heterogeneidad Espacial de Nutrientes del Suelo en Ecosistemas Terrestres*. Tesis Doctoral. Universidad de Vigo.
- Robertson, G.P. 1987. Geostatistics in ecology: interpolating with known variance. *Ecology* 68: 744-748.
- Rossi R.E., Mulla, D.J., Journel, A.G. y Franz, E.H. 1992. Geostatistical tools for modeling and interpreting ecological spatial dependence. *Ecological Monographs* 62: 277-314.