

# Introducción al Análisis Espacial de Datos en Ecología y Ciencias Ambientales: Métodos y Aplicaciones

FERNANDO T. MAESTRE

ADRIÁN ESCUDERO

ANDREU BONET

*(editores)*



Todos los derechos reservados. Ni la totalidad ni parte de este libro, incluido el diseño de la cubierta, puede reproducirse o transmitirse por ningún procedimiento electrónico o mecánico. Cualquier forma de reproducción, distribución, comunicación pública o transformación de esta obra solo puede ser realizada con la autorización de sus titulares, salvo excepción prevista por la ley. Diríjase a CEDRO (Centro Español de Derechos Reprográficos, [www.cedro.org](http://www.cedro.org)) si necesita fotocopiar o escanear algún fragmento de esta obra.

© Copyright by  
Universidad Rey Juan Carlos  
Servicio de publicaciones  
Los autores  
Madrid, 2008

Editorial DYKINSON, S.L.  
Meléndez Valdés, 61 - 28015 Madrid  
Teléfono (+34) 91544 28 46 - (+34) 91544 28 69  
e-mail: [info@dykinson.com](mailto:info@dykinson.com)  
<http://www.dykinson.es> <http://www.dykinson.com>

ISBN: 978-84-9849-308-5  
Depósito Legal:

*Preimpresión:*

Besing Servicios Gráficos, S.L.  
[besing@terra.es](mailto:besing@terra.es)

# Métodos geoestadísticos para el análisis de datos ecológicos espacialmente explícitos

ANTONIO GALLARDO

Área de Ecología. Departamento de Sistemas Físicos, Químicos y Naturales. Universidad Pablo de Olavide. Ctra. Utrera km. 1. 41013 Sevilla, España. E-mail: [agallardo@upo.es](mailto:agallardo@upo.es)

FERNANDO T. MAESTRE

Área de Biodiversidad y Conservación, Departamento de Biología y Geología, Escuela Superior de Ciencias Experimentales y Tecnología, Universidad Rey Juan Carlos, C/ Tulipán s/n, 28933 Móstoles, España. E-mail: [fernando.maestre@urjc.es](mailto:fernando.maestre@urjc.es)

## RESUMEN

En este capítulo se revisan los fundamentos de la geostatística, conjunto de herramientas estadísticas que permiten describir y caracterizar la continuidad espacial de variables autocorrelacionadas. Utilizando distintos ejemplos se presentan los conceptos básicos de geostatística, abarcando la geostatística univariante y su representación (semivariogramas y “kriging”), la geostatística multivariante, y las simulaciones. 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 mediante técnicas de “kriging”. Además se discuten aspectos relativos al análisis previo de datos, la estacionariedad (que es la asunción más importante para el uso de técnicas geostatísticas), y la detección de anisotropía en las variables objeto de estudio. 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”), la utilización de técnicas de “jackknife” para el estudio del efecto que los métodos empleados tienen sobre la interpolación de puntos (validación cruzada), y la utilización

de la geoestadística en un contexto multivariante mediante el modelo lineal de correogionalización. Junto con numerosos ejemplos que ilustran la utilización de estos conceptos en la práctica, cada una de las figuras de los apartados va acompañada con “scripts” o instrucciones para poder realizar los distintos análisis con el software libre R, que se encuentran en el CD acompañante junto con los datos originales utilizados en los ejemplos. El capítulo termina con una revisión de software, y algunas reseñas de páginas de interés, donde a su vez el lector puede ser redirigido a distintos sitios Web con numerosos recursos sobre geoestadística.

## 6.1. INTRODUCCIÓN

La geoestadística 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 te permite cuantificar su variación en distintas direcciones del espacio. La geoestadística emplea funciones para modelar esta variación espacial, que son utilizadas posteriormente para interpolar en el espacio el valor de la variable en sitios no muestreados. La fortaleza de la geoestadística se basa en el hecho de que esta interpolación (conocida con el nombre genérico de “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. En contraste con otros métodos de interpolación, como los métodos de triangulaciones o los basados en análisis de superficies (véase Legendre y Legendre 1998 para una revisión de los mismos), permite asociar la variabilidad de la estima (conocida como grado de incertidumbre).

La geoestadística permite responder a las siguientes preguntas: ¿Qué grado de dependencia espacial existe en mis variables de interés?, ¿Cuál es el patrón espacial de dichas variables?, ¿A que 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 mis variables? y ¿Cuál es el grado de incertidumbre de las estimas obtenidas en sitios no muestreados? Las respuestas a estas preguntas son siempre dependientes de la escala espacial elegida (véase el capítulo 2 para una revisión completa sobre la escala en ecología espacial). 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 uno de 100 x 100 m. Nos encontraremos dos patrones espaciales diferentes (aunque el mayor engloba, y a veces oculta al menor), así como con dos grados de incertidumbre distintos que podremos asociar a la respuesta de cada individuo a la variación en humedad (parcela pequeña) o a la respuesta de poblaciones o comunidades a dicha variación (parcela grande).

La heterogeneidad espacial a escala de ecosistema y de paisaje ha sido reconocida desde el comienzo del desarrollo de la ecología (Oosting 1948). La ecología del paisaje en sí no es más que el reconocimiento explícito de esta heterogeneidad (Pickett y Cadenasso 1995). La fitosociología se desarrolló también con la intención de simplificar, en realidad convertir en unidades discretas, la heterogeneidad espacial observada en las comunidades vegetales (Braun-Blanquet 1979). 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. (véase el capítulo 9 de este libro). La necesidad de cuantificar estadísticamente el grado y la escala espacial en que cambian las comunidades vegetales no ha sido prioritaria, entre otras cosas porque esta heterogeneidad era percibida visualmente, y por ello era espacialmente predecible (Terradas 2001). Sin embargo, los geólogos e ingenieros de minas, durante la segunda mitad del siglo XX, se encontraron con la necesidad de desarrollar herramientas estadísticas que cuantificaran el grado y escala de la variación espacial de recursos mineros que no podían percibir visualmente, pero cuyo patrón espacial necesitaban conocer para incrementar la eficiencia en su explotación (Journel y Huijbregts 1978). Este desarrollo llevó a la creación de la geostatística (Matheron 1963, 1970), que pronto fue utilizada por ciencias como la edafología (e. g. Burgess y Webster 1980, Webster 1985, Goovaerts 1999), la entomología (Liebhold *et al.* 1993), la nematología (Webster y Boag 1992, Ettema *et al.* 1998), la hidrología (Bárdossy y Lehmann 1998) y la fitopatología (Nelson *et al.* 1999), entre otras.

El uso de geostatística en ecología fue introducida por primera vez de forma explícita por Robertson (1987), aunque fue desarrollada con mayor detalle por Rossi *et al.* (1992). Junto a estos dos trabajos, ya clásicos, hay también que incluir el trabajo de Legendre y Fortin (1989) como uno de los

primeros que incluye a la geostatística como una de las herramientas para la descripción de patrones espaciales en ecología. Desde entonces, el uso de la geostatística en ecología ha crecido de forma notable, y se ha utilizado para analizar la estructura espacial de multitud de variables y procesos ecológicos, como la densidad, el área basal y la biomasa de especies vegetales (Biondi *et al.* 1994, Puigdefábregas y Sánchez 1996, Debouzie *et al.* 1996, Bellehumeur *et al.* 1997, Pastor *et al.* 1998), para analizar las relaciones entre la distribución de las características físico-químicas del suelo y la vegetación (Jackson y Caldwell 1993a, 1993b, Halvorson *et al.* 1994, 1995, Saetre 1999, Gallardo *et al.* 2000), para evaluar los efectos del proceso de desertificación (Schlesinger *et al.* 1996) y para evaluar las relaciones entre los patrones espaciales de distintas especies (Jonsson y Moen 1998, Maestre *et al.* 2005a), por citar sólo algunos ejemplos.

La geostatística ha permitido cuantificar la escala y grado de variación espacial de recursos para plantas y animales, así como 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 (p. ej. Jeltsch *et al.* 1998, Ettema y Wardle 2002). Así, 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 micrositos) 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.

## 6.2. ANÁLISIS EXPLORATORIO DE LOS DATOS

Tal como se ha puesto de manifiesto en la introducción del libro, un estudio preliminar y pormenorizado de los datos es un ingrediente esencial en todo análisis espacial de datos ecológicos, especialmente cuando se em-

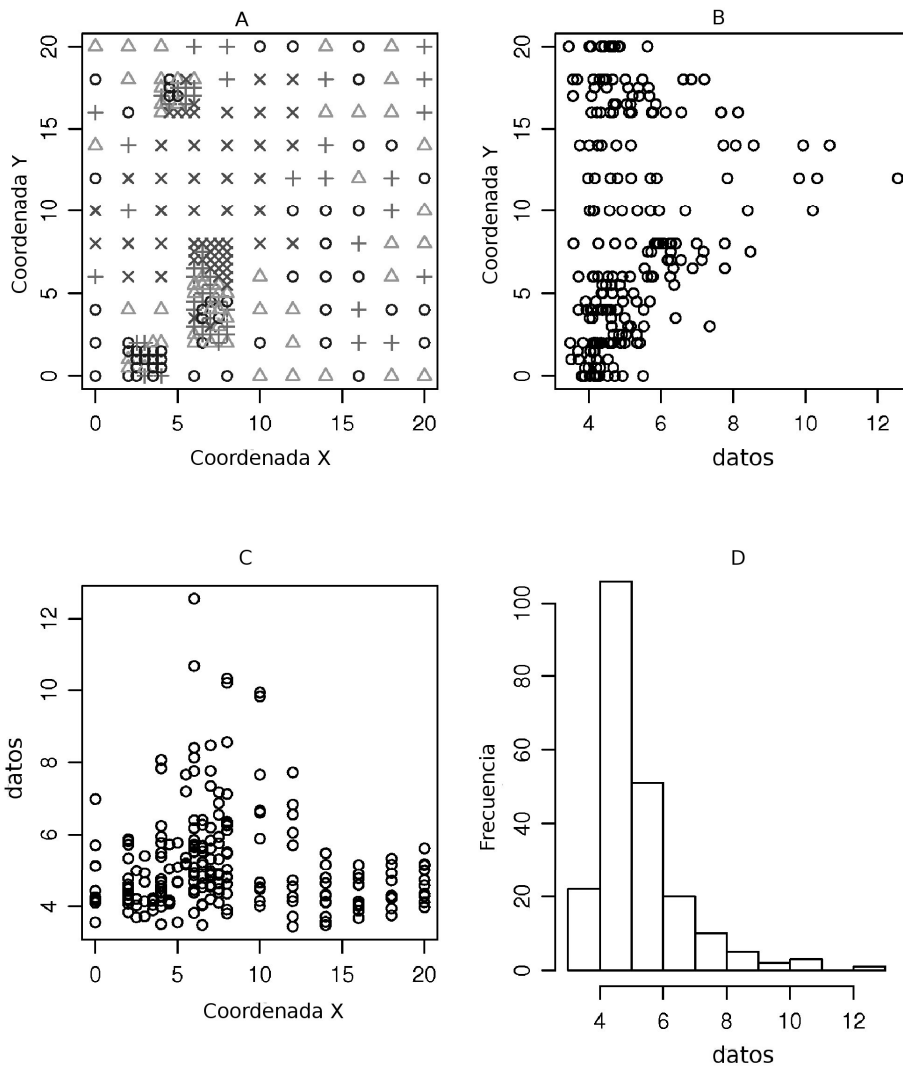
pleen técnicas geostatísticas para el mismo. El tiempo gastando en la exploración y descripción del conjunto de los datos es ampliamente recompensado por la mejora de las estimas que se obtendrán. El análisis exploratorio de los datos no tiene por qué seguir una receta de pasos estrictos, bien al contrario, es más útil perseguir intuitivamente la detección y explicación de anomalías en el espacio (Isaaks y Srivastava 1989). La imaginación y la curiosidad son las palabras claves para detectar y explicar dichas anomalías. No obstante, es recomendable que todo análisis geoestadístico empiece por un análisis exploratorio que contemple, como mínimo, los pasos descritos a continuación. Para una ampliación de los contenidos de esta sección puede consultarse Escobar (1999) y Quinn & Keough (2002).

### **6.2.1. Detección de errores**

Un primer paso antes de la utilización de herramientas descriptoras de las variables usadas debe ser la detección de errores en los datos. Estos datos pueden proceder de muestras recogidas en el campo, analizadas en el laboratorio y posteriormente introducidas en el ordenador. En cualquiera de estos pasos se puede haber cometido un error. Una forma rutinaria de hacer una primera comprobación incluiría los siguientes pasos:

- i) Ordenar los datos y examinar valores extremos. Si estos valores parecen excesivos, se debe establecer su autenticidad. Valores erráticos extremos se obtienen a veces por un movimiento del punto decimal. Las notas tomadas en el campo se deben revisar por si contienen información sobre alguna anomalía aplicable a los datos.
- ii) Localizar los valores extremos sobre un mapa. Es importante fijarse si los valores extremos están en áreas donde se dan otros valores cercanos a dicho extremo (Fig. 6.1A). Un valor extremo aislado es sospechoso de ser un error.
- iii) Comprobar errores en las coordenadas. Ordenarlas y ver si están todas dentro del rango e intervalos previstos (Figs. 6.1B y 6.1C).

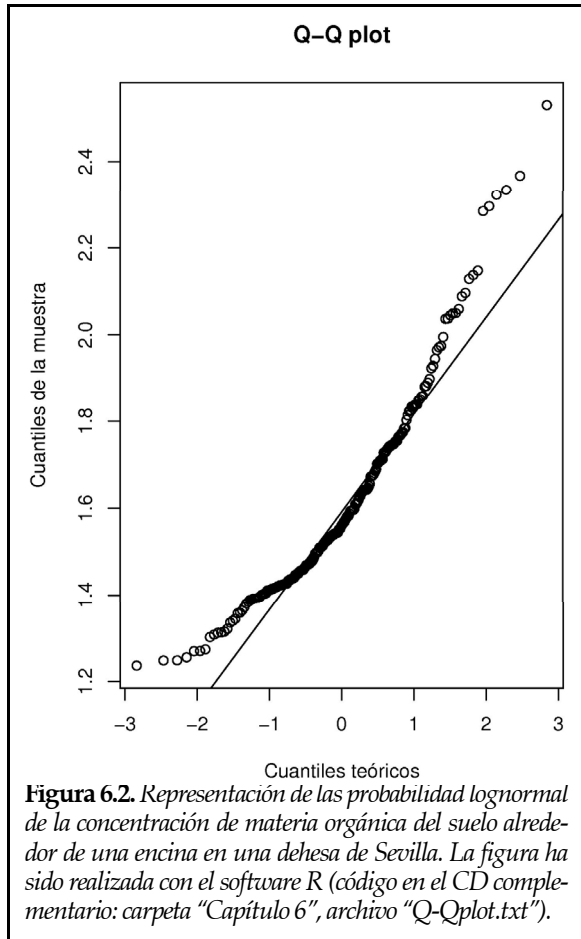
Si hay inconsistencias hay que revisar la forma en que se hizo el muestreo: ¿Fueron las coordenadas y muestras tomadas por varias personas en días diferentes? y ¿Siguieron todos la misma metodología? Preguntas de esta naturaleza ayudarán a comprobar si existen errores en nuestros datos.



**Figura 6.1.** Análisis de la distribución de los datos de materia orgánica alrededor de una encina en una dehesa de Sevilla. (A): mapa de la parcela. Los diferentes símbolos y tonos representan los cuatro cuartiles de la distribución. En (B) y (C) se representan los valores de materia orgánica con respecto a los ejes X e Y respectivamente. En (D) se representa el histograma de frecuencias. Figura realizada con el módulo GeoR del software R (se adjunta código en el CD complementario: carpeta "Capítulo 6", archivo "ExploracionDatos.txt").

## 6.2.2. Histogramas y representación de probabilidades normales

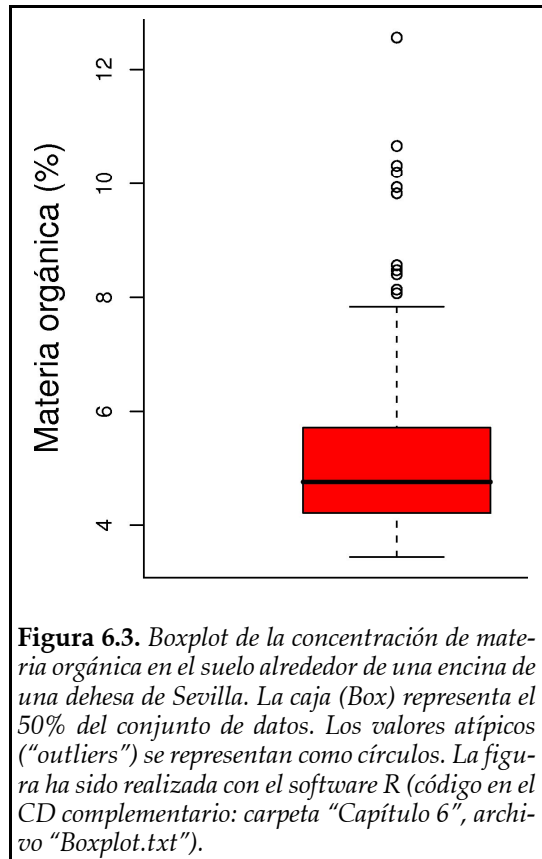
Un análisis exploratorio de los datos debe empezar por una tabla de frecuencias o su representación gráfica: el histograma (Fig. 6.1D). Es recomendable usar la misma anchura para las clases de frecuencia, de forma que la altura de cada barra sea proporcional al número de valores dentro de cada clase. Esta representación proporciona bastante información de la distribución, y puede ya permitir la detección de valores anómalos. Una segunda herramienta muy usada es la representación de probabilidades normales o lognormales con los gráficos Q-Q ("Q-Q plots"). La razón es que la estimación de valores mediante técnicas geoestadísticas funciona mejor si la distribución de los valores de los datos se acerca a una distribución Gaussiana o "normal". Una representación de probabilidad normal el eje de ordenadas se escala de forma que las frecuencias acumuladas aparecerán como una línea recta si la distribución es Gaussiana (Fig. 6.2). En esta figura se aprecia como la distribución de la materia orgánica del suelo se ajusta bien a una lognormal en los valores centrales, pero se comienza a alejar por arriba y por debajo. No es muy frecuente encontrar una distribución normal de los datos, sin embargo, una distribución relacionada, como la lognormal, puede ser una buena alternativa. Una variable está distribuida lognormalmente si la distribución del logaritmo de la variable es normal. Este tipo de representación tiene el interés adicional de que puede detectar la presencia de varias poblaciones dentro de la muestra.



Una representación de probabilidad normal el eje de ordenadas se escala de forma que las frecuencias acumuladas aparecerán como una línea recta si la distribución es Gaussiana (Fig. 6.2). En esta figura se aprecia como la distribución de la materia orgánica del suelo se ajusta bien a una lognormal en los valores centrales, pero se comienza a alejar por arriba y por debajo. No es muy frecuente encontrar una distribución normal de los datos, sin embargo, una distribución relacionada, como la lognormal, puede ser una buena alternativa. Una variable está distribuida lognormalmente si la distribución del logaritmo de la variable es normal. Este tipo de representación tiene el interés adicional de que puede detectar la presencia de varias poblaciones dentro de la muestra.

### 6.2.3. Resumen Estadístico

El análisis exploratorio de los datos puede continuar por un resumen estadístico de la distribución. Los estadísticos más utilizados son la media y la varianza, aunque estas medidas no proporcionan una completa descripción de los datos, y son muy sensibles a valores extremos. Por ello, el uso de otro tipo de estadística sumarial, como la basada en los cuantiles, son más útiles para la descripción de la variable que estemos estudiando. Los diagramas de caja (“box and whisker” o “boxplot” en inglés; Fig. 6.3) son una forma adecuada para la inspección visual de la distribución cuando se tienen tamaños muestrales por encima de ocho (Quinn y Keough 2002). En este tipo de representación se suele representar la mediana (línea que divide a la caja) junto con los cuartiles (25%



**Figura 6.3.** Boxplot de la concentración de materia orgánica en el suelo alrededor de una encina de una dehesa de Sevilla. La caja (Box) representa el 50% del conjunto de datos. Los valores atípicos (“outliers”) se representan como círculos. La figura ha sido realizada con el software R (código en el CD complementario: carpeta “Capítulo 6”, archivo “Boxplot.txt”).

de la distribución) que quedan por encima y por debajo de este descriptor. De la caja salen dos rectas que, sin que tengan una longitud superior a 1.5 veces el valor de los cuartiles que configuran la caja de la variable, llegan hasta el valor mínimo o máximo de la distribución. Además, se destacan con algún tipo de símbolo tanto los valores atípicos (“outliers”) como los extremos. Los primeros son aquellos cuyo valor es superior a 1.5 veces el valor de los cuartiles que configuran la caja de la variable, mientras que los extremos son aquellos cuyo valor es 3 veces superior al de dichos cuartiles (Escobar 1999). Los valores atípicos se alejan de una distribución normal o lognormal, pero no tienen por qué deberse a un error. Los extremos, por su parte, sí suelen considerarse como valores susceptibles de eliminarse, aunque, en último término, su carácter “extremo” debe ser evaluado cuidadosamente por el investigador.

#### 6.2.4. Descripción bivalente

En ocasiones es interesante la inspección de las diferencias en la distribución entre variables, en particular si se van a utilizar valores de una variable como predictores de valores no muestreados de una segunda variable. Una simple comparación entre histogramas y estadísticos podría ya detectar diferencias entre las distribuciones, pero, si éstas son muy parecidas, puede no revelar detalles de interés. Si se quiere realizar una comparación más detallada entre las dos distribuciones puede utilizarse un gráfico Q-Q bivalente, donde se representan los cuantiles de una variable frente a los de la otra (Fig. 6.2). Si las distribuciones de ambas variables fueran idénticas se debería observar una línea recta. Las técnicas de correlación y regresión lineal sirven también para evaluar la información que una variable tiene sobre la otra; dicha información nos puede ser de utilidad a la hora de utilizar unas variables como predictoras en el espacio de los valores de otras.

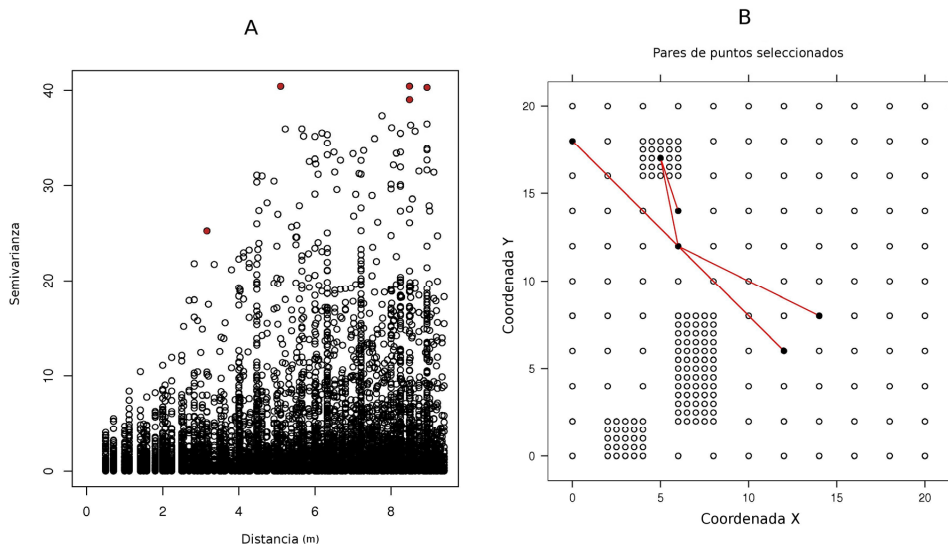
### 6.3. ANÁLISIS UNIVARIANTE: SEMIVARIOGRAMAS

Las propiedades espaciales de cada variable de interés deben ser estudiadas en un primer paso de forma individual. Al fin y al cabo, nadie tiene más información sobre la distribución espacial de una variable que ella misma. En muchas ocasiones el objetivo del trabajo acaba en este paso, aunque como se verá más adelante, otras variables también pueden proporcionar información sobre la misma. El primer paso es representar el parecido de las muestras en función de las distancias que las separan.

#### 6.3.1. Nube de pares de distancia y semivariograma

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 presentada en la Figura 6.1A, que corresponde a una parte de una dehesa de Sevilla. En esta parcela se ha muestreado una zona de 20 x 20 m situada alrededor del tronco de una encina, utilizando una malla regular en la que se tomaron muestras cada 2 m. En cuatro cuadrados localizados aleatoriamente se recolectaron muestras adicionales cada 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 solo cuatro). 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 de 20 x 20 m (del que solo 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). Este semivariograma se puede representar como una nube de puntos, representando cada punto la varianza (en realidad, la diferencia entre los dos valores al cuadrado) de un par de muestras (Fig. 6.4). La inspección de la nube de puntos es un buen momento para detectar la influencia que algunas muestras tendrán en la confección del semivariograma. En el CD adjunto se presenta el código, utilizable con el módulo *gstat* de R (Pebesma 2004; <http://www.gstat.org>), para detectar la posición de los pares de puntos que dan varianzas anormalmente altas (carpeta “Capítulo 6”, archivo “VarioNube.txt”).



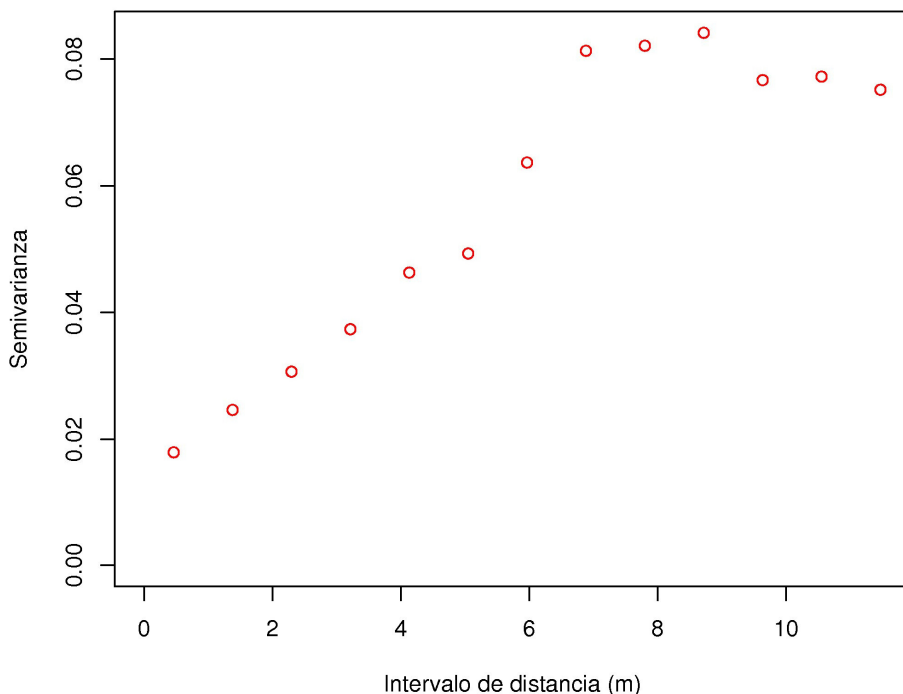
**Figura 6.4.** (A) Semivariograma de nube representando la varianza de todos los pares de muestras posibles. Los círculos rellenos destacan cinco pares de muestra que se corresponden con las distancias señaladas en el mapa de las muestras (B). La identificación de los pares de puntos y su representación en el mapa de las muestras se ha realizado con el módulo GSTAT del software R (código en el CD complementario: carpeta “Capítulo 6”, archivo “VarioNube.txt”).

El cálculo de la varianza entre pares separados por intervalos de distancia se conoce como semivarianza ( $\gamma$ ) y su estimador más utilizado es (Webster y Oliver 1990):

$$\hat{\gamma}(\mathbf{h}) = \frac{1}{2N(\mathbf{h})} \sum_{i=1}^{N(\mathbf{h})} \{z(\mathbf{x}_i) - z(\mathbf{x}_i + \mathbf{h})\}^2 \quad (6.1)$$

donde  $\hat{\gamma}(\mathbf{h})$  es la semivarianza para todas las muestras localizadas en el espacio separado por el intervalo de distancia  $\mathbf{h}$ ,  $z(\mathbf{x}_i)$  y  $z(\mathbf{x}_i + \mathbf{h})$  son los valores de la variable observados en los puntos de muestreo  $\mathbf{x}_i$  y  $\mathbf{x}_i + \mathbf{h}$  respectivamente,  $\mathbf{h}$  es la separación entre puntos de muestreo y  $N(\mathbf{h})$  es el número total de pares de muestras separados por un intervalo de distancia  $\mathbf{h}$ .

La representación del semivariograma para la concentración de materia orgánica en el suelo en el ejemplo anterior se observa en la Figura 6.5. 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 solo se representa en el semivariograma distancias aproximadas a la mitad de la dimensión de la parcela (Goovaerts 1997). La razón se puede encontrar en lo visto anteriormente; el número de pares a las distancias mayores desciende 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 (Journel y Huijbregts 1978). La tercera regla general es que el número de puntos en el espacio objeto de estudio no debería ser inferior a 50, si bien hay autores que recomiendan un mínimo de 150 puntos de muestreo para estimar el semivariograma de forma robusta (Webster y Oliver 1992). 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 (Joffre *et al.* 1996), y con semivariogramas que recogen distancias que superan el 50% de la dimensión de la parcela sin mostrar comportamiento anómalo (Ettema *et al.* 1998). No obstante, hay que tener en cuenta que semivariogramas estimados con un número bajo de puntos muestreo tendrán, por lo general, un comportamiento bastante errático (para ejemplos clarificadores véanse los trabajos de Joffre *et al.* 1996 y de Webster y Oliver 1992).



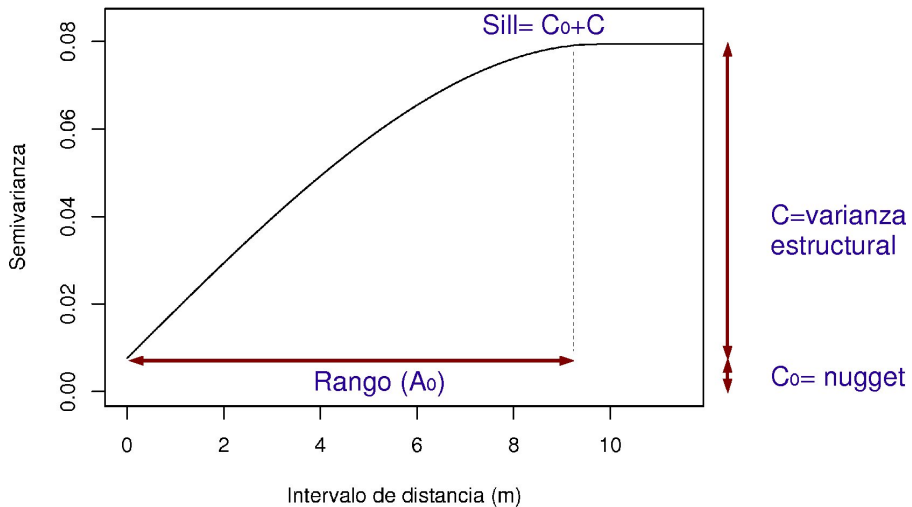
**Figura 6.5.** Representación de la semivarianza a intervalos de distancia creciente del logaritmo de la materia orgánica del suelo en una dehesa de Sevilla. La figura está realizada con el módulo *geoR* del software *R* (código en el CD complementario: carpeta "Capítulo 6", archivo "Variograma.txt").

La Figura 6.5 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 de la cual la semivarianza alcanza un máximo y no aumenta más. Este nivel máximo de semivarianza debe coincidir con la varianza de la población (Goovaerts 1997). 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.

### 6.3.2. Ajuste de una función al semivariograma

#### 6.3.2.1. Características de los modelos que se pueden ajustar al semivariograma

El semivariograma de la Figura 6.5 proporciona bastante información del comportamiento espacial de la variable. Sin embargo, es necesario ajustar una función para cuantificar el grado y escala de la variación espacial observada. 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 6.6, en donde se han ajustado dichos modelos a los datos del semivariograma empírico de la Figura 6.5. 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 (Fig. 6.6):



**Figura 6.6.** Parámetros utilizados en el ajuste del semivariograma utilizando funciones matemáticas como las presentadas en la Figura 6.7 (ver texto para una explicación más detallada). Estos parámetros se pueden calcular utilizando el código del archivo para el módulo geoR del software R incluido en el CD adjunto (Carpeta “Capítulo 6”, archivo “AjusteModelos.txt”).

i) El rango ( $A_0$ ). El rango ( $A_0$ ) es la distancia a la que la semivarianza deja de aumentar. Este parámetro, 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 (Figs. 6.2 y 6.3), 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.

ii) El "nugget" ( $C_0$ ). Este parámetro 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.

iii) La varianza estructural ( $C$ ).

iv) El "sill" ( $C_0+C$ ). La máxima semivarianza encontrada entre pares de puntos se conoce como "sill" y debe coincidir con la varianza de la población.

v) La proporción de la varianza explicada por el espacio ( $C/C_0+C$ ). Este cociente, expresado a menudo como porcentaje, nos indica la parte de la varianza que se encuentra estructurada espacialmente, 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.

Existen un gran número de modelos usados en geostatística, aunque conviene mencionar que no todas las funciones matemáticas son apropiadas para este fin cuando se quieren realizar interpolaciones utilizando "kriging". Para ello es necesario que la función a ajustar sea negativa semidefinida (Webster y Oliver 1990), lo que evitará la formación de matrices singulares durante los cálculos del "kriging" (Olea 2006). La explicación de los detalles matemáticos de esta propiedad queda fuera de los objetivos de este capítulo, aunque el lector interesado puede consultar [www.mmc.igeofcu.unam.mx/cursos/geoest/GeoEstadistica.pdf](http://www.mmc.igeofcu.unam.mx/cursos/geoest/GeoEstadistica.pdf) para una detallada explicación en castellano.

### 6.3.2.2. Principales modelos que se pueden ajustar al semivariograma

Las funciones matemáticas que son negativas semidefinidas se denominan modelos autorizados. Existe una gran variedad de estos modelos que pueden ajustarse a los semivariogramas empíricos (véase Chilès y Delfiner [1999] para una revisión exhaustiva). Las ecuaciones de los principales modelos disponibles para ajustar semivariogramas se presentan a continuación:

#### i) Modelo esférico

$$\gamma(h) = C_0 + C \left\{ \frac{3h}{2a} - \frac{1}{2} \left( \frac{h}{a} \right)^3 \right\} \quad \text{para } 0 < h \leq a \quad (6.2a)$$

$$\chi(h) = C_0 + C \quad \text{para } h > a \quad (6.2b)$$

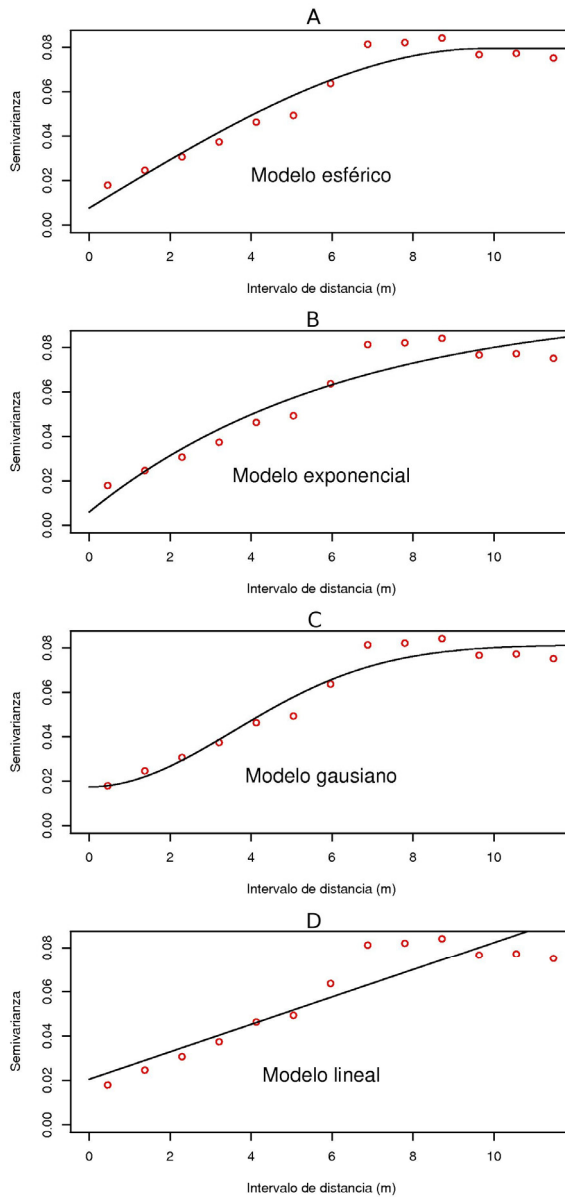
Donde  $C_0$  es el "nugget",  $C$  es la varianza estructural,  $h$  es la distancia y  $a$  es el rango ( $A_0$ ). Este modelo tiene un comportamiento lineal a distancias de separación pequeñas cerca del origen pero se va aplanando a mayores distancias y alcanza el "sill" en la distancia  $a$  (Fig. 6.7A).

#### ii) Modelo penta-esférico

$$\gamma(h) = C_0 + C \left\{ \frac{15h}{8a} - \frac{5}{4} \left( \frac{h}{a} \right)^3 + \frac{3}{8} \left( \frac{h}{a} \right)^5 \right\} \quad \text{para } 0 < h \leq a \quad (6.3a)$$

$$\chi(h) = C_0 + C \quad \text{para } h > a \quad (6.3b)$$

Los parámetros  $C_0$ ,  $C$ ,  $h$  y  $a$  de este modelo son los mismos que en el caso del modelo esférico. Si bien es poco utilizado entre los ecólogos, y es un modelo que no se encuentra disponible en todos los programas (aunque sí en el software libre *gstat*), es un modelo útil porque su curva es más gradual que la del modelo esférico, y en ocasiones permite ajustar los datos mejor que éste (Webster y Oliver 1997).



**Figura 6.7.** Ajuste del semivariograma empírico de la materia orgánica a cuatro modelos teóricos, realizado con el módulo geoR de R. El código utilizado para realizar esta figura se encuentra en el CD adjunto (carpeta "Capítulo 6", archivo "AjusteModelos.txt"). Este código también calcula el  $R^2$  del ajuste de la función y los parámetros de cada semivariograma (ver Figura 6.6)

**iii) Modelo circular**

$$\gamma(h) = C_0 + C \left\{ 1 - \frac{2}{\pi} \cos^{-1} \left( \frac{h}{a} \right) + \frac{2h}{\pi a} \sqrt{1 - \left( \frac{h^2}{a^2} \right)} \right\} \quad \text{para } 0 < h \leq a \quad (6.4a)$$

$$\gamma(h) = C_0 + C \quad \text{para } h > a \quad (6.4b)$$

Los parámetros  $C_0$ ,  $C$ ,  $h$  y  $a$  de este modelo son los mismos que en el caso del modelo esférico. Este modelo, tampoco muy utilizado en ecología, se curva de forma algo abrupta conforme se aproxima al rango. Hay que mencionar que este modelo, a diferencia del esférico y el penta-esférico, no es apropiado para el análisis de datos en tres dimensiones.

**iv) Modelo exponencial**

$$\gamma(h) = C_0 + C \left\{ 1 - \exp \left( -\frac{h}{a} \right) \right\} \quad (6.5)$$

Los parámetros  $C_0$  y  $h$  son los mismos que en el caso del modelo esférico. Este modelo tiende a alcanzar el sill asintóticamente (Fig. 6.7B). El rango  $a$  es definido como la distancia a la cual el valor del variograma alcanza el 95% del sill.

**v) Modelo gaussiano**

$$\gamma(h) = C_0 + C \left\{ 1 - \exp \left( -\frac{h^2}{a^2} \right) \right\} \quad (6.6)$$

Los parámetros  $C_0$  y  $h$  son los mismos que en el caso del modelo esférico. Al igual que el modelo exponencial, este modelo tiende a alcanzar el sill asintóticamente. El rango  $a$  se define como la distancia a la cual el variograma alcanza el 95% del sill (Fig. 6.7C). Conviene mencionar que hay autores que no recomiendan el uso de este modelo (Webster y Oliver 2001), pese a que aparece como estándar en la mayor parte de manuales de geoestadística. El motivo es que su aplicación puede derivar en inestabilidades en las ecuaciones utilizadas a la hora de realizar estimaciones mediante "krigging" (Webster y Oliver 2001).

### vi) Modelo lineal

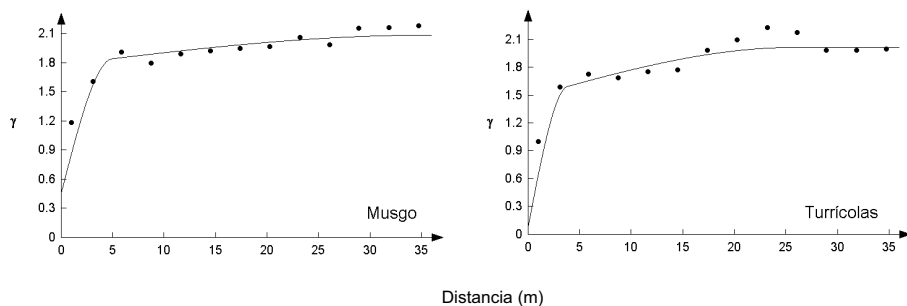
$$\gamma(h) = C_0 + bh \quad (6.7)$$

Donde  $C_0$  es el "nugget",  $h$  es la distancia y  $b$  es la pendiente de la recta (Fig. 6.7D).

### vi) Modelo potencial

$$\gamma(h) = C_0 + h^\alpha \quad \text{para } h \geq 0 \text{ y } 0 < \alpha \leq 2 \quad (6.8)$$

Donde  $C_0$  es el "nugget",  $h$  es la distancia y  $\alpha$  describe la curvatura del modelo; si  $\alpha = 1$  tenemos un modelo lineal, si  $\alpha < 1$  la curva es convexa hacia arriba y si  $\alpha > 1$  la curva es cóncava.



**Figura 6.8.** Ejemplos de semivariogramas empíricos (puntos) que muestran la heterogeneidad espacial en los valores de dos propiedades edáficas superficiales en un espartal semiárido de Alicante. Nótese como los semivariogramas muestran la presencia de estructura espacial a dos escalas diferentes. El modelo ajustado (línea sólida) fue en ambos casos un modelo esférico doble, con rangos  $a_1$  y  $a_2$  de 4,9 y 35,3 (musgo) y 3,8 y 25,4 (turrícolas). Véase el texto para los detalles. Adaptada de Maestre y Cortina (2002).

En numerosas ocasiones los semivariogramas pueden tener formas complejas que no son bien ajustadas por los modelos disponibles o bien presentar variación a distintas escalas espaciales. En estos casos la mejor alternativa pasa por combinar dos o más modelos autorizados. Por ejemplo, si la variable que estamos estudiando presenta autocorrelación a dos escalas (Fig. 6.8), podemos ajustar un modelo doble esférico (Webster y Oliver 1990):

$$\gamma(h) = C_0 + C_1 \left\{ \frac{3h}{2a_1} - \frac{1}{2} \left( \frac{h}{a_1} \right)^3 \right\} + C_2 \left\{ \frac{3h}{2a_2} - \frac{1}{2} \left( \frac{h}{a_2} \right)^3 \right\} \quad \text{para } 0 < h \leq a_1 \quad (6.9a)$$

$$\gamma(h) = C_0 + C_1 + C_2 \left\{ \frac{3h}{2a_2} - \frac{1}{2} \left( \frac{h}{a_2} \right)^3 \right\} \quad \text{para } a_1 < h \leq a_2 \quad (6.9b)$$

$$\gamma(h) = C_0 + C_1 + C_2 \quad \text{para } h > a_2 \quad (6.9c)$$

donde  $C_0$  es el “nugget” y  $C_1$  y  $C_2$  representan la varianza estructural para el primer ( $a_1$ ) y segundo ( $a_2$ ) rango respectivamente. Para ejemplos de ajustes de semivariogramas con dos o más modelos véase Maestre y Cortina (2002).

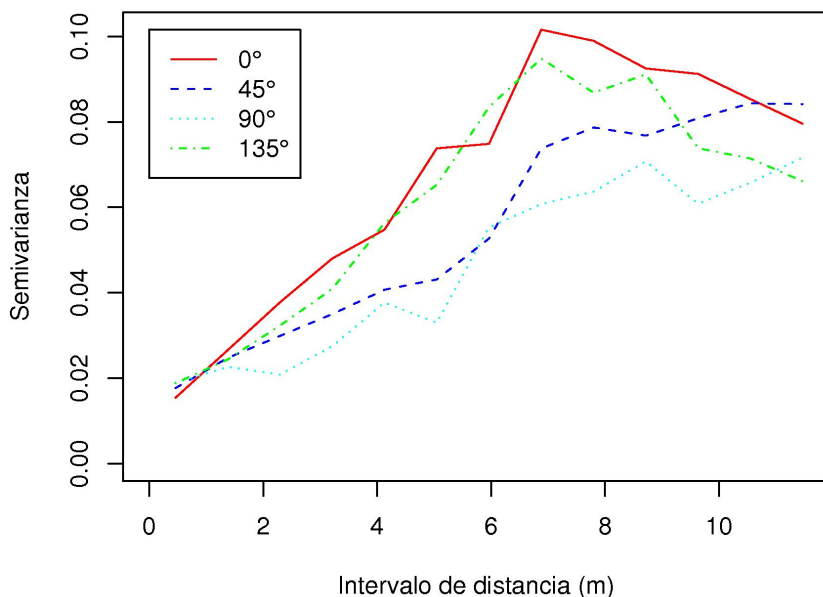
### 6.3.2.3. La elección del mejor modelo para las variables

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, utilizando como criterio el mínimo de la suma de los cuadrados de los residuales (“residual mean squares”) o el coeficiente de determinación ( $R^2$ ) de la ecuación ajustada. 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 como *variowin* o *gstat* (véase revisión de software en el apartado 6.8). 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. Conviene mencionar que el ajuste automático basado en los menores residuales o en los mayores  $R^2$  no necesariamente produce modelos con mayor significado biológico que el ajuste manual. A veces pequeñas diferencias en los menores residuales ó 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, los cambios espacio-temporales en el semivariograma de una variable dada, 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", ya que 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 los modelos lineales y potenciales se suelen utilizar para caracterizar una estructura espacial cuya dimensión supera la de la parcela de estudio, lo que ocurre, por ejemplo, cuando dicha parcela está dentro de un gradiente direccional.

### 6.3.3. Isotropía y anisotropía

En los semivariogramas vistos hasta ahora se considera que la variación del valor de una variable con el espacio es igual en todas sus direcciones (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, podríamos considerar separadamente 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  $0^\circ$ ,  $45^\circ$ ,  $90^\circ$  y  $135^\circ$ , y añadimos un ángulo de tolerancia de  $22.5^\circ$ , todos los puntos de las parcelas son usados en uno u otro semivariograma. El resultado lo podemos ver en la Figura 6.9, donde se representan semivariogramas direccionales para el contenido



**Figura 6.9.** Semivariogramas anisotrópicos de la materia orgánica del suelo en una dehesa de Sevilla. Esta figura se ha realizado con el módulo *geoR* del software *R* y el código utilizado se encuentra en el CD adjunto (carpeta “Capítulo 6”, archivo “Anisotropia.txt”).

de materia orgánica en el suelo de una dehesa de Sevilla. De esta figura se extraen dos conclusiones. En primer lugar, 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 “nuggets” 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 utilizados en su elaboración. 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. De hecho, al-

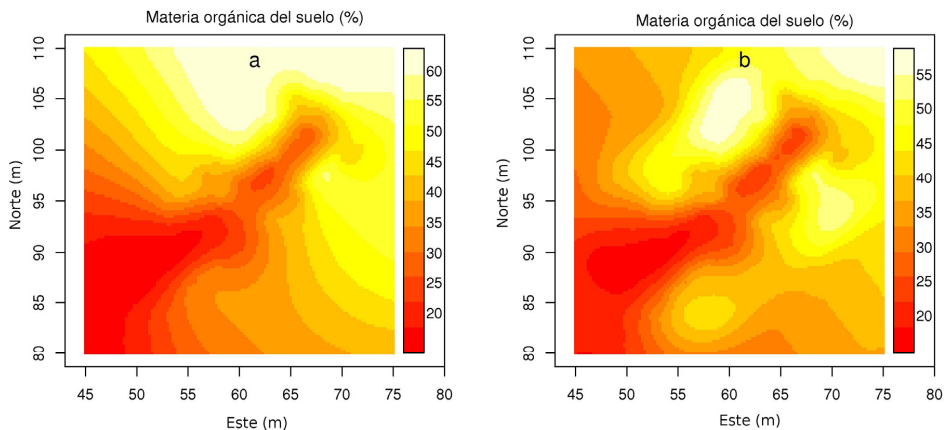
gunos autores recomiendan muestrear al menos 200 puntos de muestreo para caracterizar adecuadamente la anisotropía utilizando semivariogramas (Webster y Oliver 1992). Cuando los rangos se mantienen aproximadamente iguales en distintas direcciones del espacio pero el "sill" varía se denomina anisotropía 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 éstas siguen el mismo modelo básico de continuidad espacial. En la práctica es común encontrar una mezcla de ambas anisotropías (ver ejemplos en Isaaks y Srivastava 1989).

#### 6.3.4. 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 sea mayor que el área de estudio. La estacionariedad puede ser un problema a la hora de la interpolación de puntos en el espacio utilizando técnicas de "kriging", 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 falta de estacionariedad de los datos afecte a las estimaciones utilizando "kriging". Puede que 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 y 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, algo que te permiten la mayoría de programas de software disponibles actualmente (ver listado en el apartado 6.8). Esta aproximación se basa en que para 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, se pueden realizar 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 ésta 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 geoestadístico disponible.

Un ejemplo del efecto que puede tener una tendencia espacial sobre la estimación de puntos se puede ver en la Figura 6.10, que se corresponde con el contenido en materia orgánica en el suelo de un bosque de inundación (redibujado de Gallardo 2003a). En el gráfico de la izquierda (Fig. 6.10A) se puede apreciar un gradiente de materia orgánica que aumenta en la direc-



**Figura 6.10.** Un ejemplo de estacionariedad en los datos. El gradiente de inundación que supone la cercanía a una laguna (NE de la parcela) provoca una tendencia espacial de acumulación de materia orgánica decreciente en el eje NE-SW. 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 (raíz cuadrada de la distancia a la masa de agua). El mapa refleja tanto el gradiente espacial como fuentes de heterogeneidad local. Esta figura se ha realizado con el módulo geoR del software R y el código utilizado se encuentra en el CD adjunto (carpeta “Capítulo 6”, archivo “Estacionariedad.txt”). Modificado a partir de Gallardo (2003a).

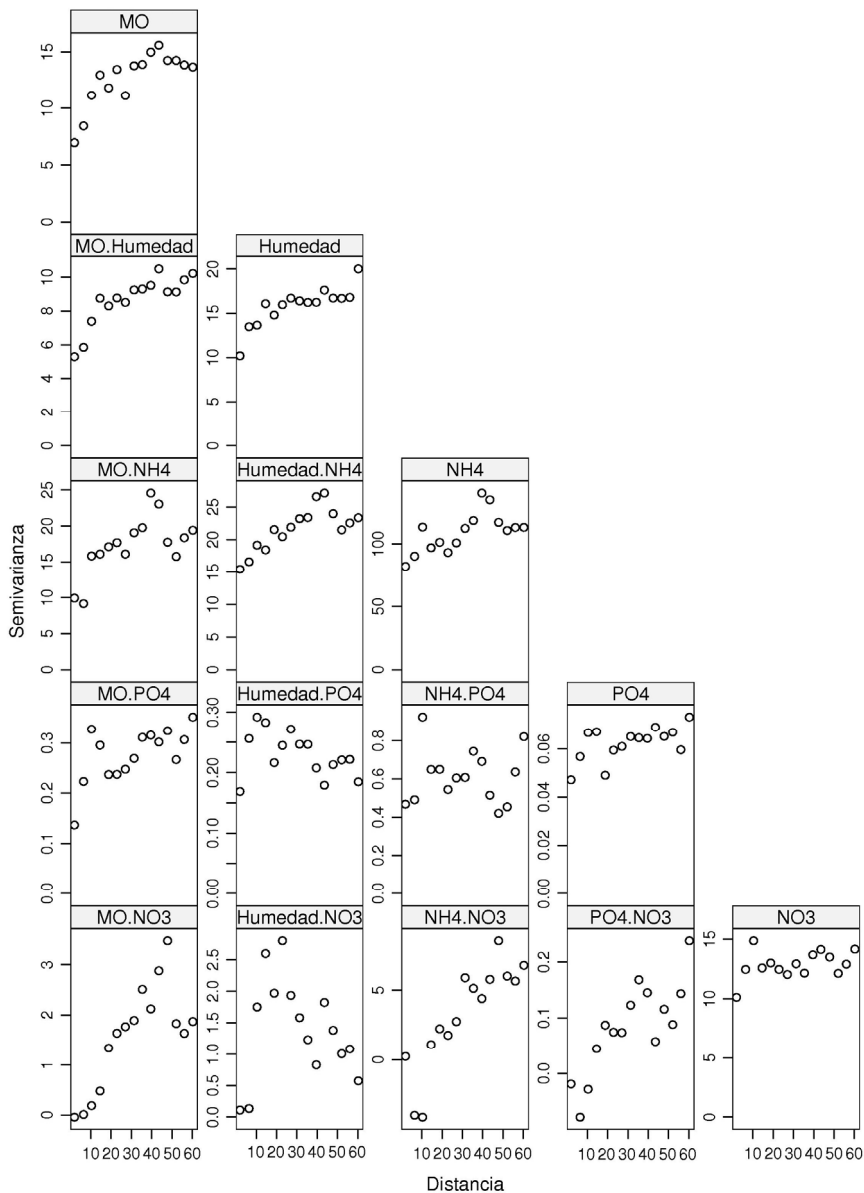
ción SO-NE. Este gradiente se debe a que la probabilidad de inundación aumenta a medida que se acerca a una masa de agua situada al NE de la parcela, lo que produce frecuentes condiciones anaeróbicas donde la materia orgánica se acumula. Sin embargo, este mapa no es del todo correcto, ya que la interpolación 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 (Fig. 6.10B). 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)

#### 6.4. ANÁLISIS MULTIVARIANTE: MODELO LINEAL DE CORREGIONALIZACIÓN

Hasta ahora se ha tratado de modelar la variación de una determinada variable a partir de sus valores evaluados de forma espacialmente explícita. Sin embargo, es posible aprovechar la información espacial que contienen otras variables que co-varíen con la primera. De esta forma se puede construir un semivariograma 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, y es representado normalmente junto a los semivariogramas simples de las variables utilizadas (Fig. 6.11). El semivariograma cruzado ( $\hat{\gamma}_{uv}$ ) puede estimarse con la ecuación (Goovaerts y Webster 1994):

$$\hat{\gamma}_{uv}(\mathbf{h}) = \frac{1}{2N(\mathbf{h})} \sum_{i=1}^{N(\mathbf{h})} \{z_u(\mathbf{x}_i) - z_u(\mathbf{x}_i + \mathbf{h})\} \{z_v(\mathbf{x}_i) - z_v(\mathbf{x}_i + \mathbf{h})\} \quad (6.10)$$

donde  $\mathbf{h}$  es la separación entre los puntos de muestreo,  $z_u(\mathbf{x}_i)$  y  $z_u(\mathbf{x}_i + \mathbf{h})$  y  $z_v(\mathbf{x}_i)$  y  $z_v(\mathbf{x}_i + \mathbf{h})$  son los valores observados en los puntos de muestreo  $\mathbf{x}_i$  y  $\mathbf{x}_i + \mathbf{h}$  para las variables  $u$  y  $v$  respectivamente, y  $N(\mathbf{h})$  es el número de comparaciones existentes a una determinada separación  $\mathbf{h}$ .



**Figura 6.11.** Semivariogramas simples y cruzados de la materia orgánica del suelo (MO), humedad, amonio (NH<sub>4</sub>), nitrato (NO<sub>3</sub>) y fósforo (PO<sub>4</sub>) en un bosque de *Pinus pinaster* en Galicia. Esta figura se ha realizado con el módulo *gstat* del software R y el código utilizado se encuentra en el CD adjunto (carpeta “Capítulo 6”, archivo “CrossVariogram.txt”).

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" (véase Goovaerts 1997 para una descripción detallada de la misma). La información que tiene una variable sobre otra es siempre menor que la que tiene una variable sobre si misma. Por ello el "cokriging" no siempre mejora significativamente la predicción de una variable, 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 menor intensidad la materia orgánica, para posteriormente, ayudarse de la variación espacial de la humedad del suelo utilizando el "cokriging".

Cuando disponemos de un conjunto de datos multivariante, es posible analizar los patrones de covariación espacial entre pares de variables utilizando una técnica geoestadística denominada modelo lineal de correogionalización (Goovaerts 1992, Webster *et al.* 1994). Ampliamente utilizada en disciplinas como la edafología (Goovaerts 1992, Goovaerts & Webster 1994, Webster *et al.* 1994), están comenzando ser utilizadas por los ecólogos en la actualidad (Maestre *et al.* 2005a). La teoría y las aplicaciones prácticas de este análisis se encuentran ampliamente descritas en Goovaerts (1997), presentándose a continuación una descripción resumida del mismo. La idea básica que subyace a este análisis es que el patrón espacial de las variables a estudiar está generado por una serie de procesos aditivos que actúan a distintas escalas. Ello implica que la forma de todos los semivariogramas individuales y cruzados es la de un modelo compuesto por el mismo tipo y número de estructuras, pero ponderado por coeficientes específicos (Goulard y Volz 1992). Para la realización del modelo lineal de correogionalización se deben seguir los siguientes pasos:

- (i) Estimación de los semivariogramas individuales y cruzados entre todos los pares de variables a estudiar.
- (ii) Selección del modelo que será ajustado a todos los semivariogramas individuales y cruzados.

- (iii) Ajustar el modelo seleccionado a todos los semivariogramas individuales y cruzados con la condición de que la matriz definida por los coeficientes  $b_{uu}^k$ ,  $b_{uv}^k$  y  $b_{vv}^k$  denominada matriz de correogionalización ( $\mathbf{B}^k$ ), sea semidefinida positiva (Goulard y Voltz 1992). Para que ello ocurra, todos sus valores propios deben ser no negativos (positivos o cero).

El modelo lineal de correogionalización permite evaluar las relaciones entre el patrón espacial de las especies a las distintas escalas definidas por los componentes de las matrices de correogionalización (tantas como estructuras distintas tenga el modelo utilizado para ajustar los semivariogramas individuales y cruzados). Para cada una de estas matrices, el modelo lineal de correogionalización permite filtrar las estructuras pertenecientes a otras escalas de variación, facilitando la comparación de los patrones espaciales de las especies a distintas escalas y proporcionando información adicional a la obtenida con los semivariogramas cruzados. Ello puede realizarse con el coeficiente de correlación estructural,  $r_{uv}^k$  calculado para cada matriz  $\mathbf{B}^k$  a partir de la siguiente ecuación:

$$r_{uv}^k = \frac{b_{uv}^k}{\sqrt{b_{uu}^k b_{vv}^k}} \quad (6.11)$$

donde  $b_{uu}^k$  y  $b_{vv}^k$  hacen referencia a los coeficientes de los semivariogramas individuales para las variables  $u$  y  $v$  respectivamente, y  $b_{uv}^k$  se refiere a los del semivariograma cruzado entre ambas.

Otra manera de resumir la información contenida en las matrices  $\mathbf{B}^k$  es mediante la aplicación de un análisis de componentes principales (Goovaerts 1992). Para ello se realiza este análisis de manera separada para cada una de estas matrices, proyectando posteriormente las correlaciones entre las variables originales y los dos primeros componentes principales en circunferencias de radio unitario denominadas círculos de correlación. Estas correlaciones pueden obtenerse mediante la ecuación (Webster *et al.* 1994):

$$c_{ij}^k = a_{ij} \sqrt{\frac{\lambda_j}{b_{ii}^k}} \quad (6.12)$$

donde  $c_{ij}$  es el valor de la correlación,  $k$  hace referencia a cada una de las matrices definidas en el modelo lineal de correogionalización,  $a_{ij}$  es el elemento  $i$  del vector propio  $j$ ,  $j$  es el  $j$  valor propio y  $b^k_{ii}$  es el elemento de  $k$  correspondiente a la variable  $i$ .

En el apartado 6.7.4 se presenta un ejemplo detallado donde se ilustra la aplicación del modelo lineal de correogionalización con un conjunto de datos reales.

## 6.5. INTERPOLACIÓN: "KRIGING"

"Kriging" es el nombre que recibe la técnica de interpolación utilizada en geostatística, con la que se ha estimado, por ejemplo, los mapas de la Figura 6.10. El término, que se acuñó en honor a un ingeniero de minas (D. G. Krige), pasó de "krigeage" (en francés) a "kriging" (en inglés), que es el término más usado, aunque se puede traducir como krigeaje en castellano (Matheron 1970).

### 6.5.1. Ventajas sobre otras técnicas de interpolación

El "kriging" es una técnica de interpolación que proporciona el mejor estimador lineal imparcial ("Best Linear Unbiased Estimator"; Schaug *et al.* 1993), suministrando además un error de estimación conocido como varianza de "kriging", que depende del modelo de variograma obtenido y de las localizaciones de los datos originales (Journel y Huijbregts 1978). Esto brinda la posibilidad de hacer análisis sobre la calidad de las estimaciones (Weerts y Bierkens 1993). Su fortaleza, y lo que lo diferencia de otras técnicas de interpolación, como el inverso de la distancia, o técnicas de triangulación, estriba en el conocimiento del comportamiento de la variable en el espacio. Al igual que en las técnicas basadas en el inverso de la distancia, la covarianza entre cualquier muestra y el punto siendo estimado generalmente decrece a medida que la muestra se aleja. Pero a diferencia del método anterior, las covarianzas calculadas para el modelo pueden provenir de una familia de funciones mucho mayor que la simple función utilizada por el inverso de la distancia. Las distancias usadas en las técnicas de "kriging" son distancias estadísticas en contraste con las

distancias geométricas utilizadas en el método del inverso de la distancia. Pero lo que realmente distingue el “kriging” de los métodos del inverso de la distancia, no es el uso de distancias estadísticas o geométricas, sino el papel que juega la agregación de puntos en el espacio. El “kriging” evita muestras redundantes, ponderando de forma distintas muestras que estén muy cerca entre sí y procedan de la misma región que muestras que estén, por ejemplo, en lados opuestos al punto que queremos interpolar. Los parámetros del semivariograma elegido tienen una importancia relevante a la hora de asignar diferentes pesos a las muestras que rodean nuestro punto a interpolar. Por ejemplo, el nugget determinará como se reparten estos pesos. Si la varianza error es muy alta, todas las muestras tenderán a tener el mismo peso en la interpolación. Si, por el contrario, la varianza error es pequeña, los coeficientes de ponderación serán muy distintos entre las diferentes muestras, muchas de las cuales (por estar muy cercanas entre ellas) serán tratadas como redundantes. El rango del semivariograma también influye en la interpolación de puntos. Los puntos más allá del rango, es decir cuando el semivariograma se vuelve plano, tienen la mínima capacidad predictora. Si el rango se alarga, estos puntos tendrían mayor peso en la interpolación de la muestra. La elección de un modelo isotrópico o anisotrópico también adquiere importancia en la interpolación de puntos en el espacio, ya que, en un modelo anisotrópico, se asignarán coeficientes mayores a las muestras situadas en el eje de mayor variación. Los detalles matemáticos del “kriging”, relativamente complejos, no se van a tratar en este apartado, remitiéndose a los lectores interesados a consultar algunos de los manuales disponibles (Goovaerts 1997, Webster y Oliver 2001).

### **6.5.2. Criterios para realizar un “kriging”**

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 falta de estacionariedad en los datos, 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.

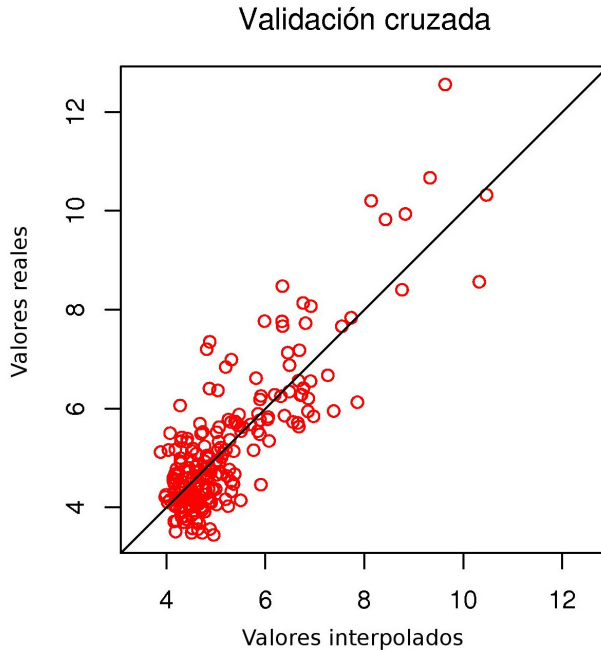
El tipo de “kriging” más usado se conoce como “ordinary kriging”. En este tipo de “kriging” la media es estimada localmente. Si se conoce la media de la variable a estimar en el área de estudio, lo que raramente ocurre, se puede utilizar el “simple kriging”. Por último, el “universal kriging” es un “ordinary kriging” donde se incluye la influencia de una tendencia espacial interna o externa a los datos (ver apartado 6.3.5).

### 6.5.3. “kriging” puntual vs. “kriging” por bloques

Otra decisión a tomar a la hora de realizar un “kriging” 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. La elección del método dependerá del objetivo concreto del estudio, pero 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). La interpolación por bloques puede ser interesante para un ecólogo cuando se intenta describir la cantidad de recursos que rodea una planta. En este caso se hace conveniente una interpolación por bloques circulares, seleccionando el área del bloque en función del tamaño de la planta (Gallardo *et al.* 2006). La posibilidad de trabajar con bloques circulares o irregulares a la hora de realizar estimas utilizando “kriging” está implementada en el módulo *gstat* del software R (<http://www.r-project.org/>).

### 6.5.4. Validación cruzada

La validación cruzada es una técnica para comparar los valores verdaderos y estimados usando solo la información disponible en nuestros datos. 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 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. 6.12). El beneficio real de una validación cruzada es el de señal de aviso. El estudio de los resultados de la valida-



**Figura 6.12.** Validación cruzada de los valores de materia orgánica del suelo de una dehesa. Esta figura, así como el procedimiento estadístico asociado, se ha realizado con el módulo `geoR` del software R y el código utilizado se encuentra en el CD adjunto (carpeta “Capítulo 6”, archivo “ValidacionCruzada.txt”).

ción 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 y Srivastava 1989). No obstante, la utilización de la validación cruzada nos puede ayudar a decidir entre diferentes estrategias de búsqueda o entre diferentes modelos.

## 6.6. SIMULACIÓN

Con técnicas de interpolación como el “kriging” se espera estimar un valor que sea lo más cercano posible al valor real. Sin embargo, indepen-

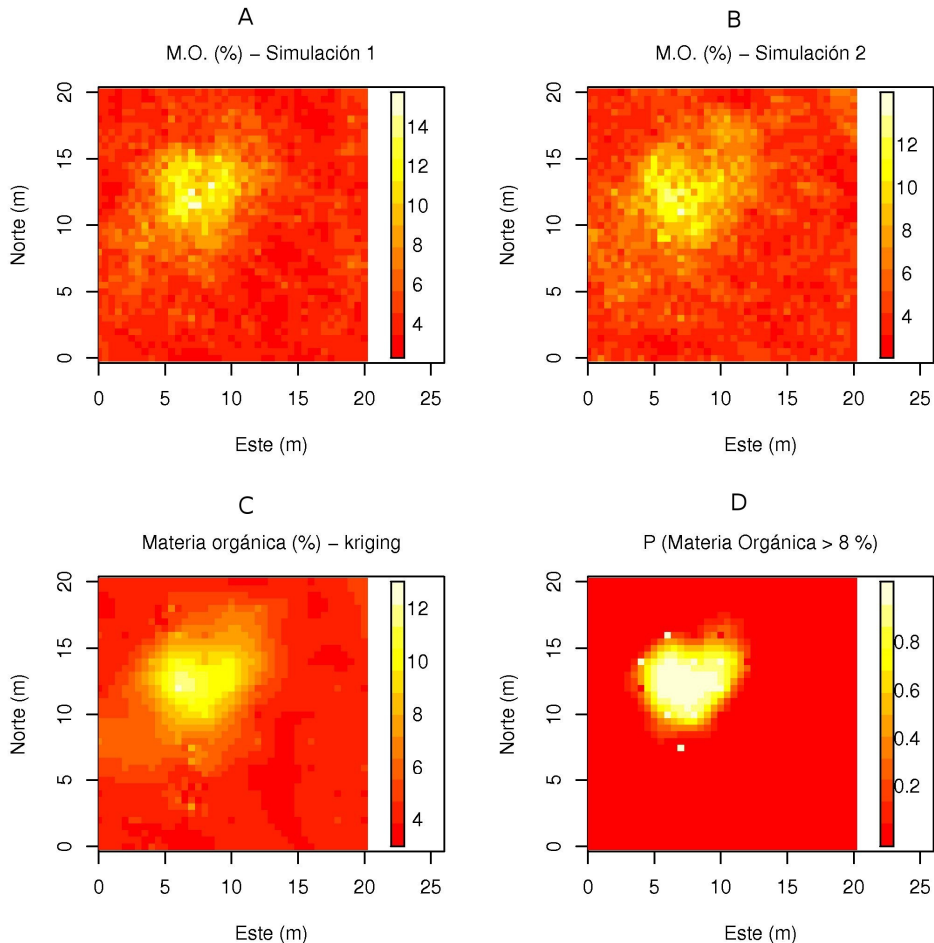
dientemente del método de estimación elegido, siempre cometeremos algún error. No es posible calcular este error exactamente, pero sí le podemos asignar una “incertidumbre”, alguna indicación de su posible magnitud. Los factores que determinan el error en la estimación son: el número de muestras cercanas, la proximidad de estas muestras (mientras más muestras más cercanas mayor será la confianza en nuestra estima), el tipo de distribución espacial de las muestras (si muestras adicionales están muy cercanas a otras ya existentes, no mejorarán significativamente la confianza en la estima), y por último, como de predecible o errática es la variable en el espacio.

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. Las simulaciones probabilísticas tienen algunas ventajas sobre la interpolación de valores del “kriging”. En primer lugar, evitan el efecto de suavizado de valores que produce el “kriging”. En segundo lugar, proporcionan medidas de incertidumbre que pueden ser de gran utilidad.

Las simulaciones de valores en el espacio más frecuentemente usadas se conocen como simulaciones condicionales gaussianas. Son condicionales porque deben hacer honor a los valores originales de la distribución, y gaussianas porque asumen que los errores en la estimación tienen una distribución normal. No siempre es así, y existe un módulo de R desarrollado para realizar simulaciones sin partir de distribuciones normales de los errores (*geoRglm* o “software for generalised linear spatial models using R”; <http://www.daimi.au.dk/~olefc/geoRglm>).

En la Figura 6.13 se presentan una comparación entre un interpolado de valores por “kriging” (Fig. 6.13C), y dos simulaciones condicionales de dichos valores (Figs. 6.13A y B), realizadas con la concentración de materia orgánica del suelo en una dehesa de Sevilla. Además se utilizan los resultados de 500 simulaciones para calcular la probabilidad de que las muestras superen valores del 8% de materia orgánica (Figura 6.13D). En todas

las figuras aparecen niveles de materia orgánica altos en la misma zona de la parcela, coincidiendo con la presencia de la copa de una encina. Sin embargo, la interpolación por “kriging” suaviza las diferencias en materia orgánica con respecto a las simulaciones, observándose además en estas últimas, valores más altos y más bajos que en el mapa resultante del “kriging”. Finalmente, en base a las simulaciones, la probabilidad de encontrar valores altos de materia orgánica se restringen al área cubierta por la encina.



**Figura 6.13.** (A y B): Interpolación de los valores de materia orgánica por simulación. (C): Interpolado mediante el uso de kriging. (D): Probabilidad de alcanzar niveles de materia orgánica superiores al 8% basada en 500 simulaciones de la concentración de materia orgánica. Esta figura se ha realizado con el módulo geoR del software R y el código utilizado se encuentra en el CD adjunto (carpeta “Capítulo 6”, archivo “Simulaciones.txt”).

## 6.7. CASOS PRÁCTICOS

### 6.7.1. Efecto de la copa de una encina sobre la distribución de nutrientes en el suelo

Este caso pertenece a un estudio realizado en una dehesa de Sevilla, donde se evaluó el efecto de una encina en distintas propiedades edáficas (Gallardo 2003b). Como se ha mencionado a lo largo de este capítulo, bajo la encina estudiada se encontraron acumulaciones de materia orgánica. Si esto es así, cabe esperar que aquellos nutrientes muy ligados a la descomposición de la materia orgánica sigan una distribución similar a esta, mientras que otros nutrientes más dependientes de minerales del suelo o independientes de la materia orgánica podrían presentar otro tipo de distribución. En base a esto se estableció una hipótesis de trabajo, presentada en la Tabla 6.1. Esperaríamos una fuerte estructura espacial asociada a la copa del árbol para el nitrógeno, ya que prácticamente no interacciona con los minerales del suelo y su ciclado se produce a través de formas orgánicas. El fósforo, por el contrario, está asociado a la materia orgánica, pero también se encuentra en equilibrio con varias formas en el suelo, y es precipitado por algunos minerales. El potasio, al igual que los dos anteriores, es un elemento esencial para las plantas, pero no forma parte de las moléculas orgánicas. Otro elemento muy interesante desde el punto de vista cuantitativo es el sodio, que no es esencial para las plantas (lo que no quiere decir que no tomen cantidades de sodio), pero sí para los animales. Por último, el litio es un elemento no esencial para las plantas. Estas características diferenciales de estos nutrientes nos hace pensar que tengan diferente grado de asociación con la presencia de una encina, y presenten estructuras espaciales diferentes con relación a ésta (Tabla 6.1).

#### 6.7.1.1. *Métodos*

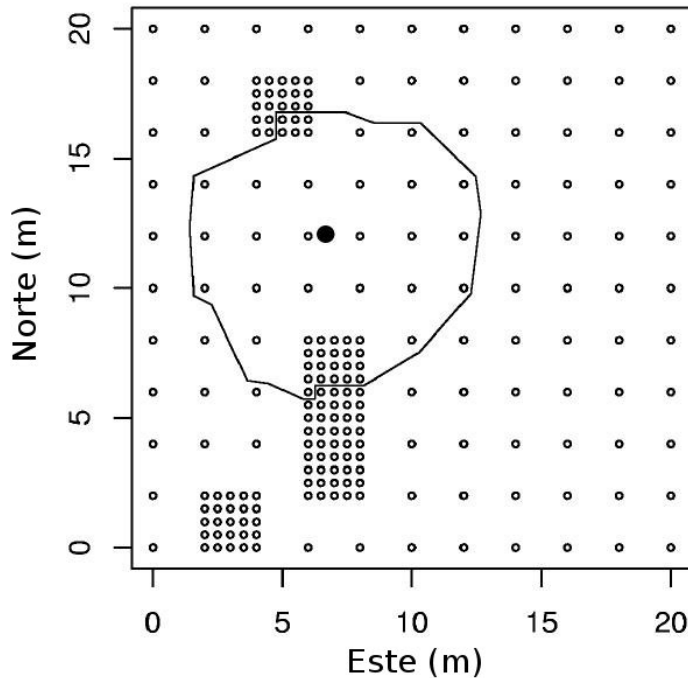
Se muestreó una parcela de 20 x 20 m que incluía una encina de unos 10 m de diámetro aproximadamente (Fig. 6.14). Las muestras se tomaron cada dos metros, en un diseño regular. Dentro de la parcela, y para recoger variabilidad espacial por debajo de dos metros, se situaron otras cinco parcelas al azar de 2 x 2 m en las que se tomaban muestras cada 50 cm. En las muestras de suelo se analizó la materia orgánica, la humedad del suelo (gravimétrica) y los nutrientes antes mencionados (para detalles ver Gallardo 1993b). Los análisis estadísticos se hicieron con el módulo *geoR* del sot-

ware R (<http://leg.ufpr.br/geoR/>). Se comprobó la normalidad de los datos, la existencia de anisotropía y la presencia de estacionariedad en los mismos.

**Tabla 6.1** Hipótesis de trabajo del ejemplo presentado en el apartado 6.7.1 (efecto de la copa de un árbol sobre distintos nutrientes en el suelo). La presencia de la cobertura arbórea de una encina debería ser más notoria sobre los nutrientes esenciales para las plantas, y sobre aquellos que interaccionen más con la materia orgánica que con los minerales del suelo.

Nutriente	Esencial para las plantas	Interacciona con la materia orgánica	Interacciona con los minerales del suelo	Estructura espacial esperada	Influencia del tamaño de la copa sobre el rango esperada
N	Sí	Sí	No	Alta	Alto
P	Sí	Sí	Sí	Media	Medio
K	Sí	No	Sí	Baja	Bajo
Na*	No	No	Sí	Baja	Nulo
Li	No	No	Sí	Baja	Nulo

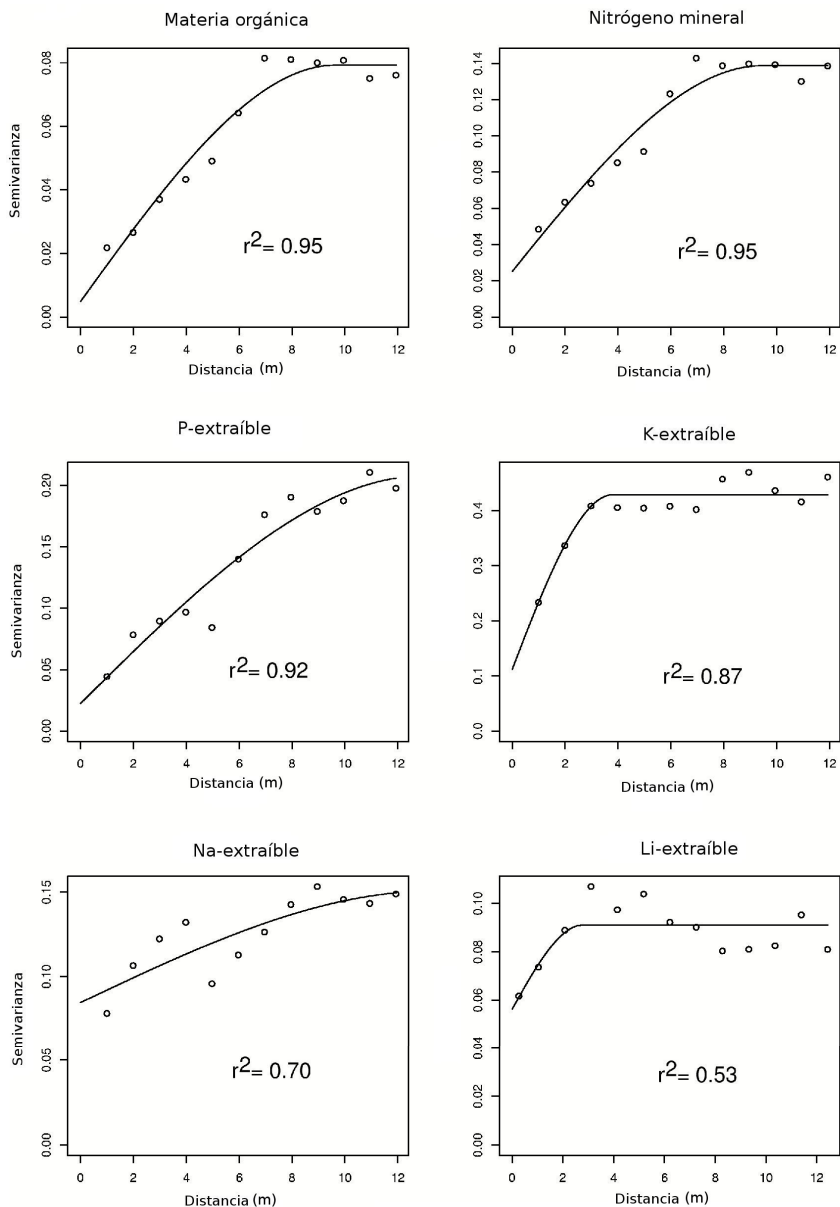
(\*) El sodio es esencial para los animales



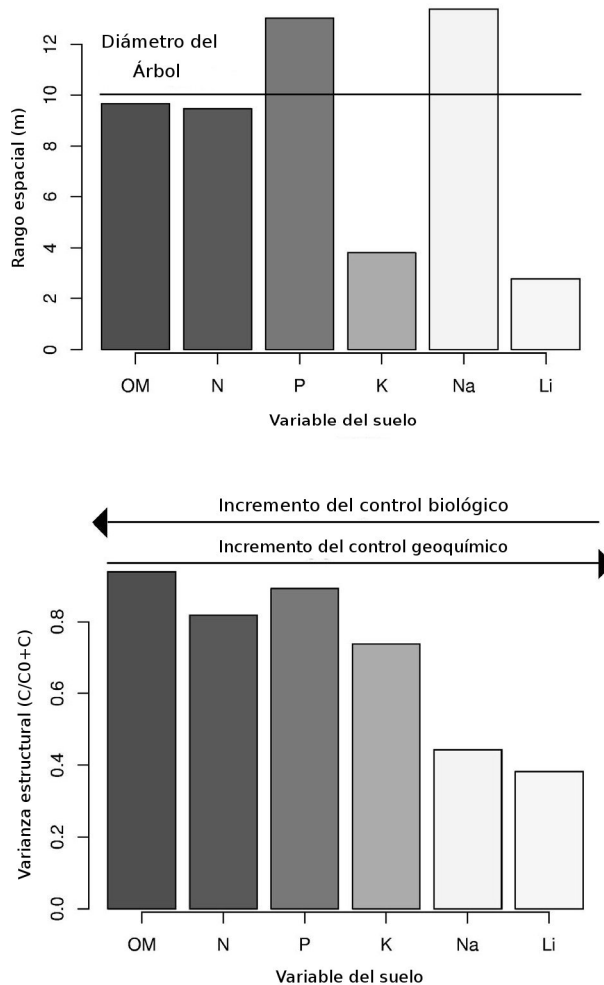
**Figura 6.14.** Diseño de muestreo en una dehesa de Sevilla. Cada punto representa una muestra de suelo. El círculo representa la extensión de la copa de una encina. Modificado de Gallardo (1993b).

### 6.7.1.2. *Resultados y discusión*

Los semivariogramas de la materia orgánica y cada una de los nutrientes de interés se encuentran en la Figura 6.15. Todos los ajustes al modelo esférico son significativos, aunque las concentraciones de Na y Li muestran un comportamiento espacial más errático. El rango espacial (la distancia a la cual las muestras están autocorrelacionadas) debería coincidir con el tamaño de la copa del árbol, si éste tiene influencia sobre la distribución de nutrientes. Eso solo se aprecia para la materia orgánica y el nitrógeno, de acuerdo con la hipótesis inicial. Fósforo y sodio muestran rangos que superan el diámetro de la encina, mientras que K y Li, posiblemente los nutrientes menos ligados a la materia orgánica, muestran rangos muy inferiores al diámetro de la encina. Si examinamos la estructura espacial, los resultados son aún más claros (Fig. 6.16). La presencia de la encina tiene tanta influencia en ésta, que el grado de dependencia espacial va disminuyendo desde la materia orgánica del suelo (muy influenciada por la acumulación de hojarasca de la encina) hasta el Li, un elemento que posiblemente circule de forma bastante independiente a la presencia de la encina. Este gradiente coincide con el aumento del control geoquímico (reacciones con minerales del suelo) frente al control biológico (a través de la materia orgánica). Así, la materia orgánica y el N están casi bajo control biológico exclusivo. La concentración de P está determinada tanto por su mineralización de la materia orgánica (control biológico) como por sus interacciones con los minerales del suelo (control geoquímico). El K no forma parte de moléculas orgánicas, pero sí circula a través de la encina (control biológico), y procede de la meteorización de los minerales del suelo (control geoquímico). El sodio no es esencial para las plantas (pero sí para los animales), y por ello la influencia de la encina en su estructura espacial es muy escasa, al igual que el Li, un elemento no esencial para las plantas. La conclusión es que la presencia de una encina impone un fuerte patrón espacial en la distribución de nutrientes del suelo, aunque el grado de dependencia variará en función de las causas que determinen la concentración de estos nutrientes.



**Figura 6.15.** Semivariograma empírico (puntos) y modelo ajustado (línea recta) para las concentraciones de materia orgánica, nitrógeno mineral, fósforo (P), potasio (K), sodio (Na) y litio (Li) en una dehesa de Sevilla (modificado de Gallardo 1993b).



**Figura 6.16.** Gráfica superior: Representación del rango espacial de la materia orgánica (OM), el nitrógeno mineral (N), el fósforo (P), potasio (K), sodio (Na) y litio (Li) extraíbles de la solución del suelo. la línea superior representa la dimensión del diámetro de la copa del árbol. Gráfico inferior: Representación de la proporción de varianza estructural encontrada en cada uno de estos elementos. Adaptado de Gallardo (1993b).

### 6.7.2. Variabilidad espacial de polifenoles totales en las hojas de una población de roble (*Quercus robur*)

Este caso está descrito en Covelo y Gallardo (2004). Los polifenoles son compuestos del metabolismo secundario cuya principal función en los tejidos de las plantas es defensiva. Las especies del género *Quercus* basan sus defensas

foliares en la acumulación de este tipo de compuestos basados en el carbono, presentando frecuentemente altas concentraciones de un número muy variable de compuestos, que pueden agruparse bajo la denominación de polifenoles totales. Aunque se conocen los cambios en el tiempo de la concentración de polifenoles en las hojas (por ejemplo Covelo y Gallardo 2001), se desconocía si dentro de una población la concentración de polifenoles foliares era espacialmente independiente. La cuestión no es trivial, ya que tiene implicaciones tanto para el muestreo como para los potenciales herbívoros. Los objetivos del trabajo eran tres: i) evaluar si la concentración de polifenoles en hoja verde presentaba estructura espacial; ii) ver si esta estructura se mantenía durante la senescencia de la hoja; y iii) analizar si esta estructura estaba relacionada con la estructura espacial de disponibilidad de luz y nitrógeno, dos de los recursos que pueden determinar la concentración de polifenoles en las hojas.

#### 6.7.2.1. Métodos

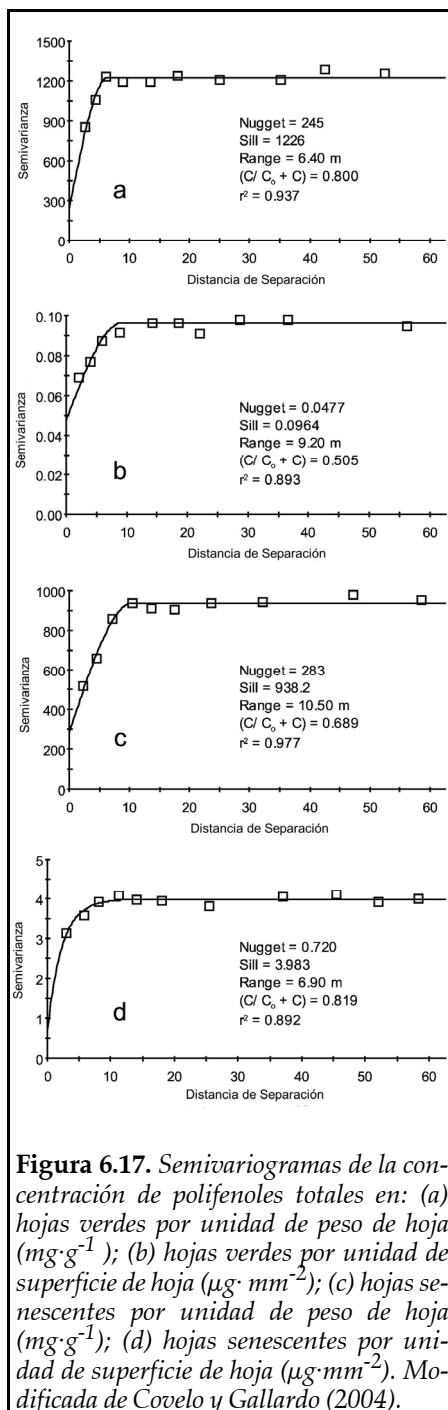
En un pinar con individuos jóvenes de roble (unos 10-20 años) se seleccionó un área de muestreo de 120 x 120 m. Dentro de este área se descartaron zonas con aparente perturbación por trabajos forestales o atravesadas por caminos. En la zona finalmente seleccionada se censaron 125 robles de tamaño superior al metro de altura. En cada uno de estos árboles se tomaron muestras de hojas a principios de verano (máxima concentración de nutrientes) y en otoño-invierno coincidiendo con la senescencia de la hoja, así como cuatro muestras de suelo, separadas un metro del tronco el árbol y en las cuatro direcciones cardinales. Además, se colocaron cuadrados adicionales para estimar la dependencia espacial a una escala menor que la distancia media entre árboles, recogiendo más de 500 muestras de suelo que fueron analizadas separadamente. En cada uno de los árboles se midió disponibilidad de luz mediante fotografía hemisférica, la concentración de polifenoles totales, nitrógeno y fósforo en las hojas, y la concentración de nitrógeno y fósforo extraíble en las muestras de suelo.

El análisis geoestadístico se hizo con el módulo *geoR* del software R. Se calcularon semivariogramas isotrópicos que se usaron posteriormente para interpolar puntos en el espacio mediante "kriging". A cada punto interpolado se le asignó una distribución normal basada en la distribución de errores de la interpolación. A partir de aquí se realizaron 500 simulaciones condicionadas a los valores de las muestras en sus localidades espaciales, de donde se extrajeron la probabilidad de que en una localidad la concentración de polifenoles supere un cierto umbral.

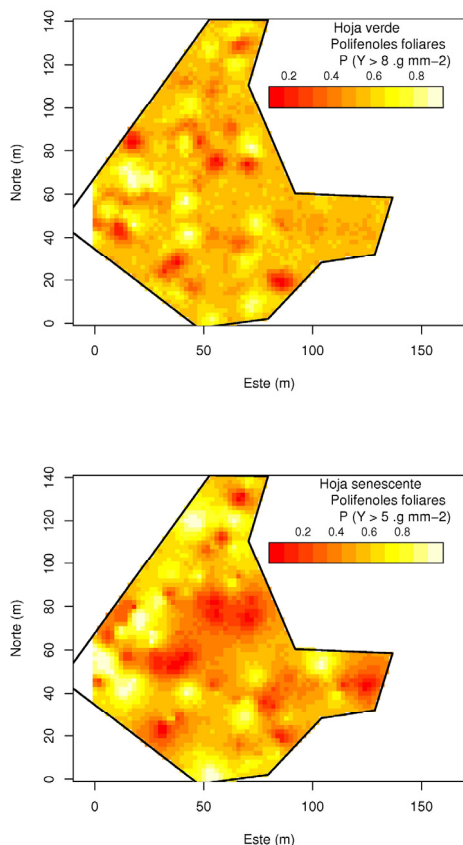
### 6.7.2.2. Resultados y discusión

La concentración de polifenoles totales en las hojas verdes y senescentes mostró estructura espacial, tanto cuando los valores se expresan por unidad de peso de hoja, como por unidad de superficie (Fig. 6.17). Comparando el porcentaje de varianza explicado por el espacio para el mismo tipo de modelo, se observa una mayor dependencia espacial en la hoja verde que en la senescente. El proceso de senescencia, por tanto afecta al patrón de distribución de los polifenoles de las hojas, lo que puede tener implicaciones importantes tanto para la herbivoría como para el ciclo de nutrientes, ya que los polifenoles están implicados en la precipitación de las proteínas, afectando tanto al tubo digestivo de los herbívoros, como a la descomposición de la hojarasca en el suelo.

También se detectó poca correlación espacial entre la distribución de la concentración de polifenoles en hoja verde y hoja senescente (Fig. 6.18). Estos resultados sugieren que los procesos que determinan la estructura espacial de los polifenoles en hojas verdes y senescentes son relativamente independientes, por ejemplo, la concentración de polifenoles mostró una correlación positiva y significativa con la disponibilidad de luz, mientras que esta relación no se apreció en las hojas senescentes (Tabla 6.2).



**Figura 6.17.** Semivariogramas de la concentración de polifenoles totales en: (a) hojas verdes por unidad de peso de hoja ( $\text{mg}\cdot\text{g}^{-1}$ ); (b) hojas verdes por unidad de superficie de hoja ( $\mu\text{g}\cdot\text{mm}^{-2}$ ); (c) hojas senescentes por unidad de peso de hoja ( $\text{mg}\cdot\text{g}^{-1}$ ); (d) hojas senescentes por unidad de superficie de hoja ( $\mu\text{g}\cdot\text{mm}^{-2}$ ). Modificada de Covelo y Gallardo (2004).

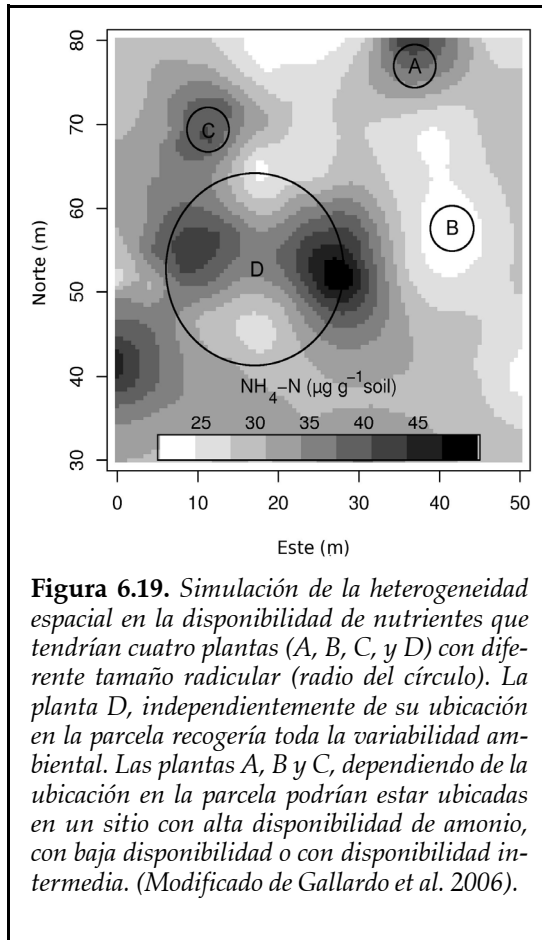


**Figura 6.18.** Probabilidad de que la concentración de polifenoles en hoja verde (gráfico superior) ó en hoja senescente (gráfico inferior) superen los valores de 8 y 5 g·mm<sup>-2</sup> respectivamente. Modificado de Covelo y Gallardo (2004).

**Tabla 6.2.** Coeficientes de correlación (*r*) entre la concentración foliar de polifenoles (en hojas verdes y senescentes), la radiación, índices de disponibilidad de nutrientes, y concentración foliar de nitrógeno (*n* = 125). (a) polifenoles por unidad de peso de hoja (mg g<sup>-1</sup>); (b) polifenoles por unidad de superficie (μg·mm<sup>2</sup>); \* = *P* < 0.05; \*\* = *P* < 0.01 y ns = *P* > 0.05.

	Polifenoles en hoja verde		Polifenoles en hoja senescente	
	(a)	(b)	(a)	(b)
Radiación total transmitida (%)	0.18*	0.31**	ns	ns
N en hoja verde (mg g <sup>-1</sup> )	0.22*	0.30**	0.23*	ns
N en hoja senescente (mg g <sup>-1</sup> )	ns	ns	ns	ns
N-NH <sub>4</sub> <sup>+</sup> (μg g <sup>-1</sup> )	ns	ns	0.27**	ns
N-NO <sub>3</sub> <sup>-</sup> (μg g <sup>-1</sup> )	ns	ns	ns	ns
Tasa de mineralización del N (μg g <sup>-1</sup> día <sup>-1</sup> )	ns	ns	ns	ns

### 6.7.3. Diferencias entre el patrón espacial del amonio y el nitrato: consecuencia a nivel de poblaciones



**Figura 6.19.** Simulación de la heterogeneidad espacial en la disponibilidad de nutrientes que tendrían cuatro plantas (A, B, C, y D) con diferente tamaño radicular (radio del círculo). La planta D, independientemente de su ubicación en la parcela recogería toda la variabilidad ambiental. Las plantas A, B y C, dependiendo de la ubicación en la parcela podrían estar ubicadas en un sitio con alta disponibilidad de amonio, con baja disponibilidad o con disponibilidad intermedia. (Modificado de Gallardo et al. 2006).

Este ejemplo se encuentra descrito en Gallardo *et al.* (2006). En él se evalúan los patrones espaciales del amonio y nitrato edáfico en varias comunidades de plantas, que van desde pastizales y matorrales y pinares hasta dehesas mediterráneas y bosques de inundación compuestos por robles y alisos. Asimismo, se evalúa la influencia que el tamaño de grano de diferentes formas de nitrógeno puede tener sobre la nutrición de plantas con sistemas radicales de distinto tamaño. La hipótesis de trabajo se puede entender a partir de la Figura 6.19. Una planta cuyo tamaño del sistema radical sea inferior al tamaño de grano que muestra el nutriente, puede caer dentro de la parcela en una zona con alto contenido en nitrógeno, o bien en otra con bajo contenido. Es decir en la

población tendríamos individuos que solo podrían explotar áreas con alto, bajo o intermedio nivel de nutrientes. Por el contrario, una planta que tenga un sistema radical de igual rango o mayor que el tamaño de grano del nutriente (el tamaño de grano medio lo define el rango espacial del semivariograma), independientemente de su localización dentro de la parcela, tendrá una alta probabilidad de explotar un área heterogénea, de modo que todos los individuos en el área tendrán similar concentración de nutrientes.

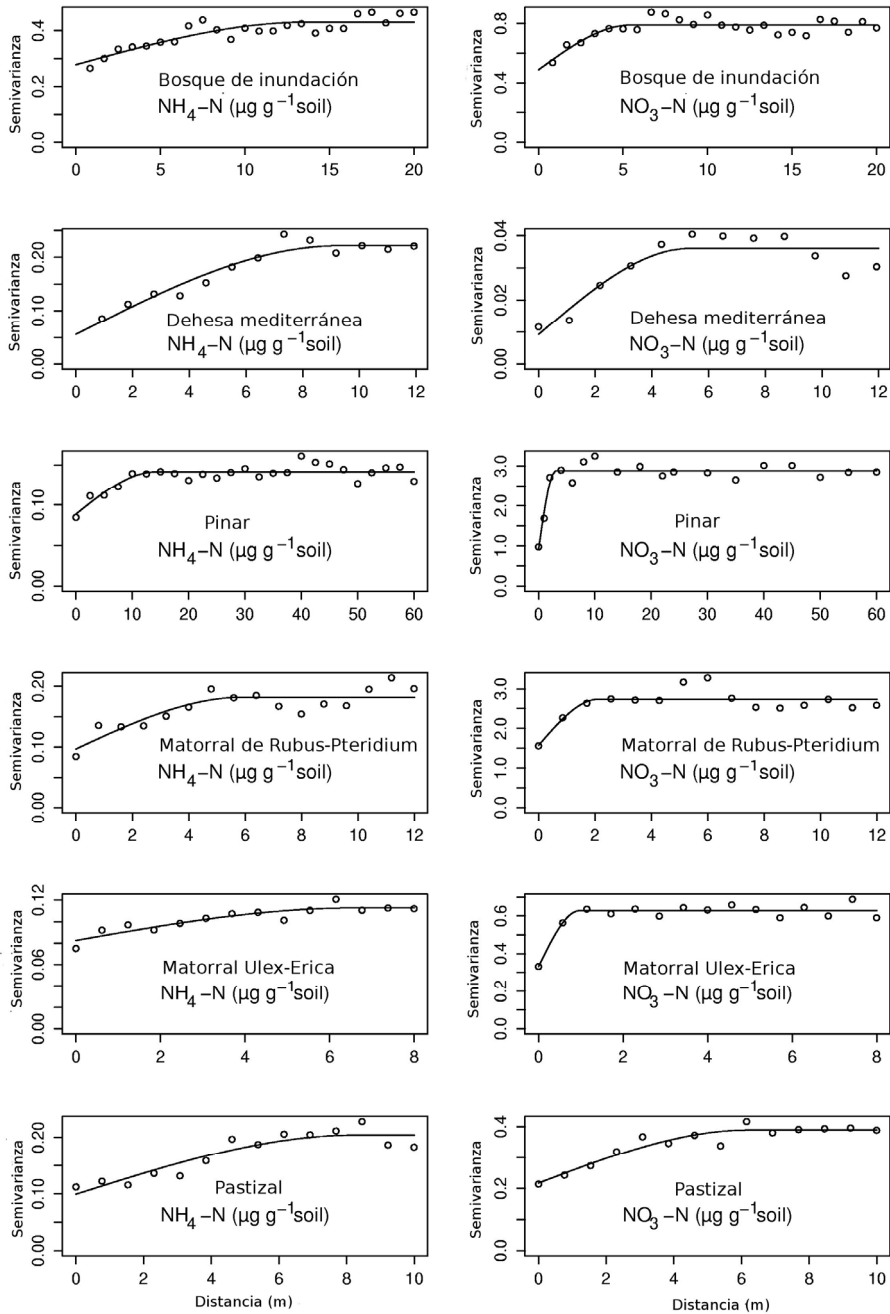
### 6.7.3.1. Métodos

Cada una de las comunidades vegetales fue muestreada en diferentes fechas y con diferentes diseños, por lo que las escalas estudiadas no coinciden en todas las comunidades. El número de puntos muestreado oscila entre 200 y 500 para cada comunidad. En cada uno de estos puntos se tomaron muestras de suelo, de donde se extrajo la concentración de amonio ( $\text{NH}_4^+$ ) y nitrato ( $\text{NO}_3^-$ ) presente en el suelo. Con estos valores se realizaron semivariogramas (utilizando el módulo *geoR* del software R). A partir de estos semivariogramas se realizaron simulaciones del valor medio de amonio y nitrato contenidos en un círculo que representa la extensión del sistema radical para poblaciones de plantas simuladas en cada una de las comunidades vegetales. Con estas simulaciones se calculó la probabilidad de que un individuo establecido en una zona de las parcelas de cada comunidad obtuviera una concentración de amonio o nitrato en el área de influencia de sus raíces superior a un umbral preestablecido. Por último se representaron la frecuencia de individuos con diferente probabilidad de alcanzar la concentración umbral de amonio o nitrato.

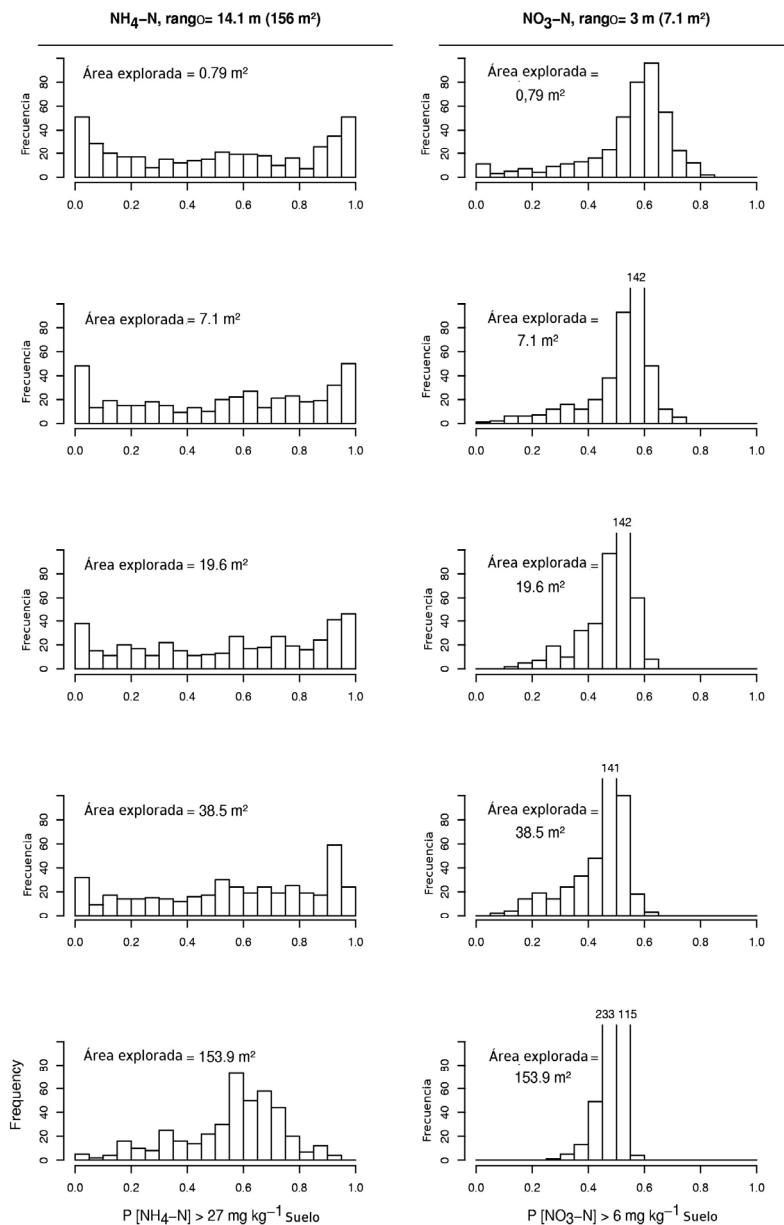
### 6.7.3.2. Resultados y discusión

El amonio presentó un rango espacial superior al nitrato en todas las comunidades (Fig. 6.20). Como todos los suelos de las comunidades eran ácidos o ligeramente ácidos, es probable que la nitrificación estuviera inhibida o disminuida, y por tanto restringida a micrositios locales favorables, lo que explica su menor rango. El amonio, por el contrario, muestra rangos amplios, que coinciden con los rangos mostrados por la materia orgánica (Gallardo *et al.* 2006).

Nuestra hipótesis principal se confirmó, tal como puede apreciarse en la Figura 6.21. En esta figura hemos simulado la probabilidad, dentro de una población simulada de plantas, de obtener una cantidad de  $\text{NH}_4$  o  $\text{NO}_3$  determinada. Cuando el sistema radical es muy inferior al tamaño de grano, la población es muy heterogénea, en el sentido de que habrá individuos con poca, mucha e intermedia probabilidad de obtener una concentración determinada de nutrientes. Sin embargo, a medida que el tamaño del sistema radical aumenta, la probabilidades de obtener una determinada concentración de nutrientes se igualan entre individuos. A igualdad de sistemas radicales, una población que tomase preferentemente nitrato (tamaño de grano más pequeño) sería siempre más homogénea (en su disponibilidad de nitrógeno) que la misma población tomando amonio (mayor tamaño de grano).



**Figura 6.20.** Semivariogramas de amonio (gráficas izquierdas) y nitrato (gráficas derechas) en seis comunidades vegetales diferentes (Modificado de Gallardo et al. 2006).



**Figura 6.21.** Frecuencia de distribución de individuos de una población con probabilidad de que el área explorada por las raíces encuentren concentraciones superiores a  $27 \text{ mg} \cdot \text{kg}^{-1}$  de amonio (gráficos de la izquierda) y concentraciones superiores a  $6 \text{ mg} \cdot \text{kg}^{-1}$  de nitrato (gráficos de la derecha). Cada gráfico simula una población de plántulas o adultos de diferente tamaño que tuviesen a su disposición la distribución de  $\text{NH}_4$  y  $\text{NO}_3$  encontrada en la comunidad de pinar.

#### 6.7.4. Análisis de los patrones de covariación espacial en una comunidad semiárida

Este ejemplo se encuentra descrito en Maestre *et al.* (2005a). En él se evalúan las relaciones en los patrones espaciales de los principales componentes perennes de un espartal semiárido situado en la provincia de Alicante (*Stipa tenacissima*, *Globularia alypum*, *Brachypodium retusum*, *Anthyllis cytisoides* y los caméfitos) utilizando varias mallas de muestreo y un modelo lineal de correogionalización. La hipótesis principal que se maneja en este estudio es que el patrón espacial de *S. tenacissima* determina el del resto de las especies presentes. Se parte de la suposición de que los cambios en las condiciones microambientales ocasionados por *S. tenacissima* (Maestre *et al.* 2001) facilitan la instalación y el desarrollo de otras especies en el área de influencia de las macollas. Se esperan encontrar asociaciones positivas entre los patrones de los distintos componentes perennes de la vegetación en los espartales, especialmente entre *S. tenacissima* y cada una de las principales especies presentes. La atribución de un papel determinante al patrón de *S. tenacissima* se justifica por su dominancia en la comunidad y por haber estado favorecida frente a las demás especies en los periodos donde fue explotada y cultivada (Barber *et al.* 1997). Los objetivos de este trabajo son los siguientes: i) evaluar la covariación conjunta entre pares de especies dentro de estas comunidades y ii) analizar el efecto de la malla de muestreo en los resultados del análisis.

##### 6.7.4.1. Métodos

El estudio se llevó a cabo en una parcela experimental de 50 x 50 m localizada cerca de Aigües de Busot, en la provincia de Alicante. Las características de la parcela, así como del muestreo realizado se han descrito en el apartado 4.6.2, por lo que no se van a volver a describir aquí. Los datos disponibles para este estudio son los valores de cobertura para las cinco especies dominantes (*S. tenacissima*, *G. alypum*, *B. retusum*, *A. cytisoides* y los caméfitos), obtenidos con cuadrados de muestreo de tamaño diferente (5 x 5, 2,5 x 2,5 y 1,25 x 1,25 m). Estos datos se encuentran en el CD adjunto (carpeta "Capítulo 4", archivo "aguas.xls"). En el mismo también puede consultarse una figura en color con el mapa digital utilizado para obtener los valores de cobertura (carpeta "Capítulo 1", archivo "Figura 1.7.bmp"). Los datos no se categorizaron ni transformaron previamente al análisis geoestadístico.

Para la realización del modelo lineal de corregeionalización se siguieron los tres pasos descritos en el apartado 6.4 (estimación de los semivariogramas individuales y cruzados entre todos los pares de especies, selección del modelo a ajustar a todos los semivariogramas individuales y cruzados y ajustar el modelo seleccionado a todos los semivariogramas individuales y cruzados). Para seleccionar el modelo que sería ajustado a todos los semivariogramas individuales y cruzados se realizó en primer lugar un análisis de componentes principales (ACP) con los datos originales de cobertura para cada una de las mallas de muestreo utilizadas (Fig. 6.22, Tabla 6.3). Posteriormente se calcularon los semivariogramas de los dos primeros vectores propios del ACP. Basándose en estos semivariogramas, se ajustaron todos los semivariogramas individuales y cruzados mediante el siguiente modelo doble esférico:

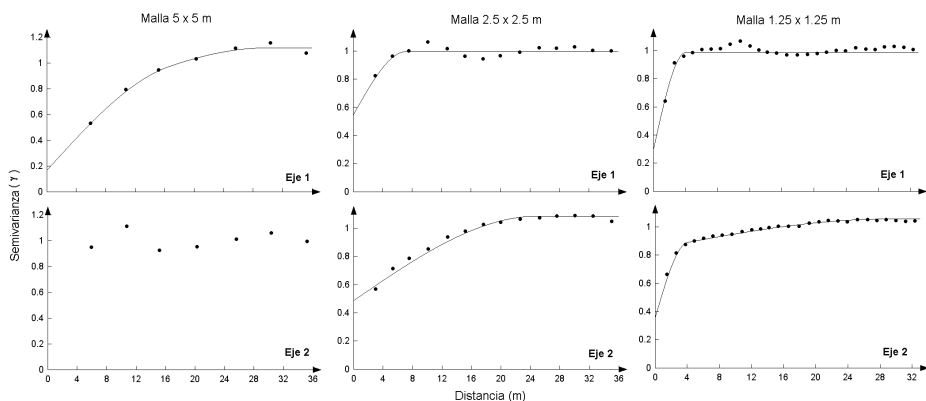
$$\gamma_{uv}(h) = b_{uv}^1 + b_{uv}^2 \left\{ \frac{3h}{2a_1} - \frac{1}{2} \left( \frac{h}{a_1} \right)^3 \right\} + b_{uv}^3 \left\{ \frac{3h}{2a_2} - \frac{1}{2} \left( \frac{h}{a_2} \right)^3 \right\} \text{ para } 0 < h \leq a_1$$

$$\gamma_{uv}(h) = b_{uv}^1 + b_{uv}^2 + b_{uv}^3 \left\{ \frac{3h}{2a_2} - \frac{1}{2} \left( \frac{h}{a_2} \right)^3 \right\} \text{ para } a_1 < h \leq a_2$$

$$\gamma_{uv}(h) = b_{uv}^1 + b_{uv}^2 + b_{uv}^3 \text{ para } h > a_2$$

donde  $a_1$  y  $a_2$  corresponden a los rangos seleccionados. Los valores de  $a_1$  y  $a_2$  para cada malla de muestreo se obtuvieron a partir de los modelos ajustados a los dos primeros componentes del ACP, mientras que los coeficientes  $b_{uv}^k$  (donde  $k = 1, 2$  y  $3$  hacen referencia a al “nugget” y a los modelos esféricos de rangos  $a_1$  y  $a_2$  respectivamente) se calcularon a partir de los datos. La estimación de estos coeficientes se realizó utilizando el algoritmo iterativo descrito en Goulard y Volz (1992).

El modelo lineal de corregeionalización utilizado, al estar compuesto por un efecto “nugget” y dos funciones esféricas con rangos  $a_1$  y  $a_2$ , da lugar a la obtención de tres matrices de corregeionalización ( $\mathbf{B}^k$ ) que permiten evaluar las relaciones entre el patrón espacial de las especies a las distintas escalas definidas por sus componentes (“nugget”, primer rango y segundo rango). Para cada una de estas matrices se calculó el coeficiente de correlación estructural,  $r_{uv}^k$  utilizando la ecuación (6.11) descrita en el apartado 6.4. Asimismo, tam-



**Figura 6.22.** Semivariogramas experimentales (puntos) y modelos ajustados (línea continua) a los dos primeros vectores propios (representados como Eje1 y Eje 2 en la figura) del análisis de componentes principales para las tres mallas de muestreo utilizadas. Las características de los modelos ajustados se describen en la Tabla 6.3. El número de pares de observaciones utilizado para calcular cada punto del semivariograma osciló entre 684 y 1044 (malla 5 x 5 m), 2964 y 13256 (malla 2,5 x 2,5 m) y 12324 y 100520 (malla 1,25 x 1,25 m).

**Tabla 6.3.** Parámetros de los modelos ajustados a los semivariogramas estandarizados de los dos primeros vectores propios del análisis de componentes principales presentados en la Figura 6.22.  $C_0$  = "nugget",  $a_1$  = primer rango (m) y  $a_2$  = segundo rango (m),  $C_1$  = varianza estructural para el primer rango,  $C_2$  = varianza estructural para el segundo rango.  $n = 100, 400$  y  $1600$  para las mallas de muestreo  $5 \times 5$  m,  $2.5 \times 2.5$  m y  $1.25 \times 1.25$  m, respectivamente.

Malla	Eje	Modelo	$C_0$	$C_1^*$	$C_2$	$a_1^{**}$	$a_2$
5 x 5 m	1	Doble esférico	0.17	0.38	0.57	16.5	30
	2	- ***	-	-	-	-	-
2.5 x 2.5 m	1	Esférico	0.45	0.55	-	6.5	-
	2	Esférico	0.49	0.60	-	25	-
1.25 x 1.25 m	1	Esférico	0.30	0.69	-	4	-
	2	Doble esférico	0.36	0.49	0.21	4	29

\* En el modelo esférico representa el único valor de varianza estructural existente.

\*\* En el modelo esférico representa el único rango existente.

\*\*\* No se encontró una estructura espacial evidente, por lo que no se ajustó ningún modelo.

bién se realizó para cada una de estas matrices un análisis de círculos de correlación utilizando la ecuación (6.12) presentada en dicho apartado.

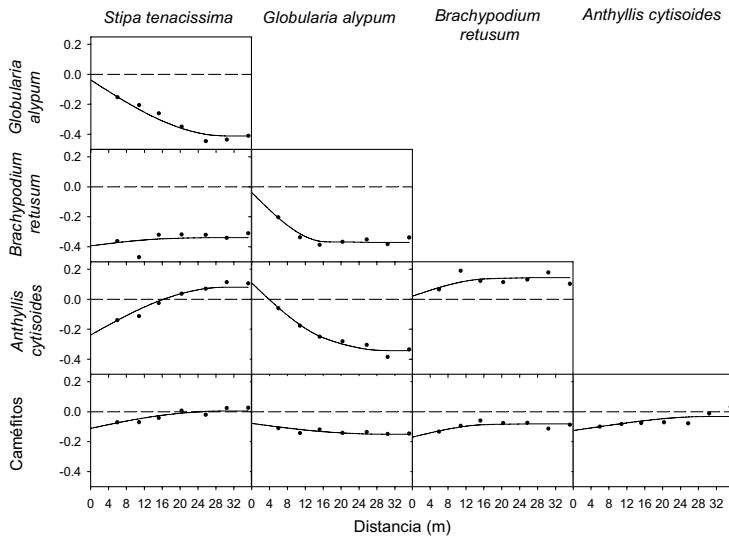
Los análisis de componentes principales se realizaron con el software SPSS 9.0 (<http://www.spss.com/es>). Los semivariogramas cruzados fueron

estimados con el programa Variowin 2.2 (<http://www-sst.unil.ch/research/variowin/index.html>), mientras que el ajuste del modelo lineal de correogionalización se realizó con el software Maple para Windows (<http://www.maplesoft.com/>).

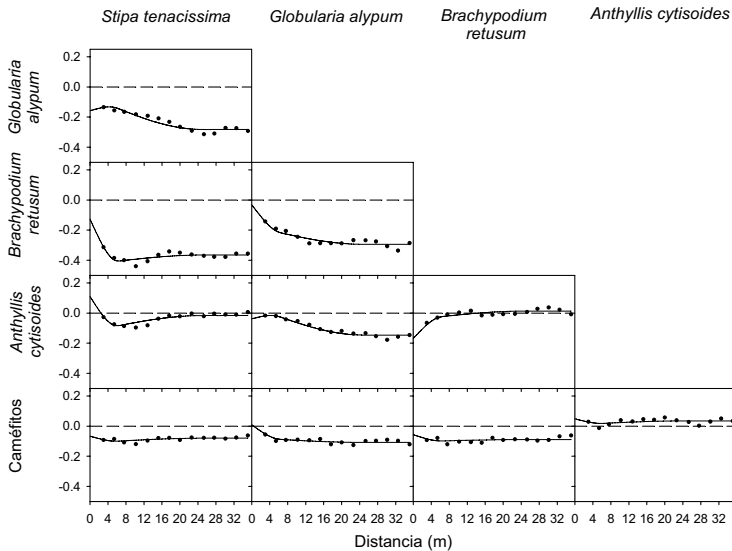
#### 6.7.4.2. Resultados y discusión

Los semivariogramas cruzados mostraron la evolución con la distancia de la covariación espacial entre las especies estudiadas (Figs. 6.23, 6.24 y 6.25). Según los resultados obtenidos con la malla de 5 x 5 m, la distribución espacial de *S. tenacissima* mostró una relación negativa a distancias pequeñas con los patrones de *A. cytisoides* y caméfitos, que llegó a ser positiva a distancias superiores a 20 m. *Globularia alypum* presentó relaciones negativas con el resto de especies. La relación negativa entre *G. alypum* y *S. tenacissima*, *B. retusum* y *A. cytisoides* se incrementó con la distancia hasta alcanzar rangos comprendidos entre los 12 y los 24 m. *Brachypodium retusum* y *A. cytisoides* presentaron una relación positiva que se incrementó con la distancia hasta estabilizarse a distancias de unos 12 m. Al aumentar la resolución, la mayor parte de las relaciones se hicieron menos evidentes, aunque se mantuvieron las negativas entre los patrones de *G. alypum* y los del resto de especies y entre las distribuciones espaciales de *S. tenacissima* y *B. retusum*. Es interesante destacar cómo las relaciones positivas entre *S. tenacissima* y *A. cytisoides* y entre *S. tenacissima* y los caméfitos desaparecieron en la malla de 1,25 x 1,25 m. La asociación entre los patrones espaciales de *A. cytisoides* y *B. retusum* se atenuó sensiblemente con el aumento de resolución, aunque no llegó a mostrarse claramente negativa en ningún caso.

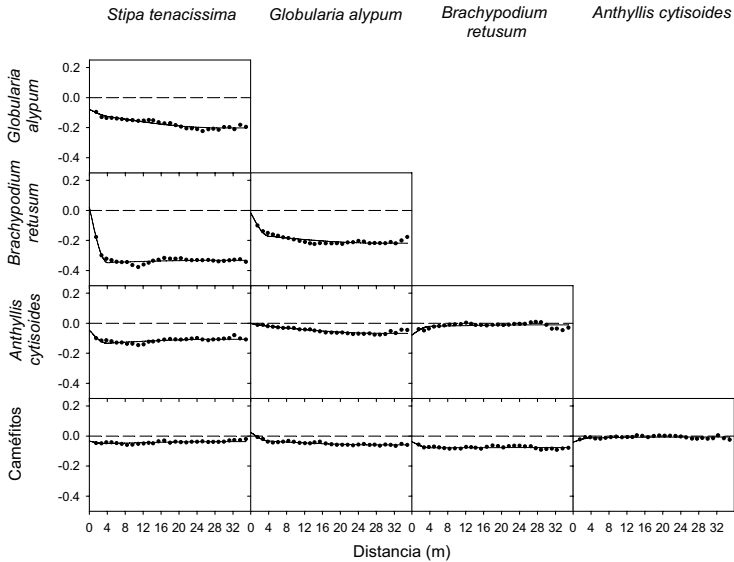
El conjunto de relaciones espaciales entre las distintas especies puede ser analizado con detalle mediante los coeficientes de correlación estructural (Tabla 6.4) y los círculos de correlación (Fig. 6.26), que permiten el análisis independiente de las correlaciones a las escalas consideradas en el modelo de correogionalización: “nugget”, primer rango y segundo rango. Las correlaciones estructurales correspondientes a la escala de segundo rango mostraron relaciones marcadamente positivas de todas las especies entre sí, excepto las combinaciones con *G. alypum* que, por el contrario, fueron siempre altamente negativas. Esta disociación entre los patrones de *G. alypum* y del resto de especies a la escala del rango mayor fue consistente para las tres resoluciones empleadas y fácilmente apreciable en los círculos de correlación.



**Figura 6.23.** Semivariogramas cruzados (puntos) y modelo de correogionalización ajustado (línea continua) utilizando la malla de muestreo de 5 x 5 m. Los valores en el eje y representan la semivarianza.



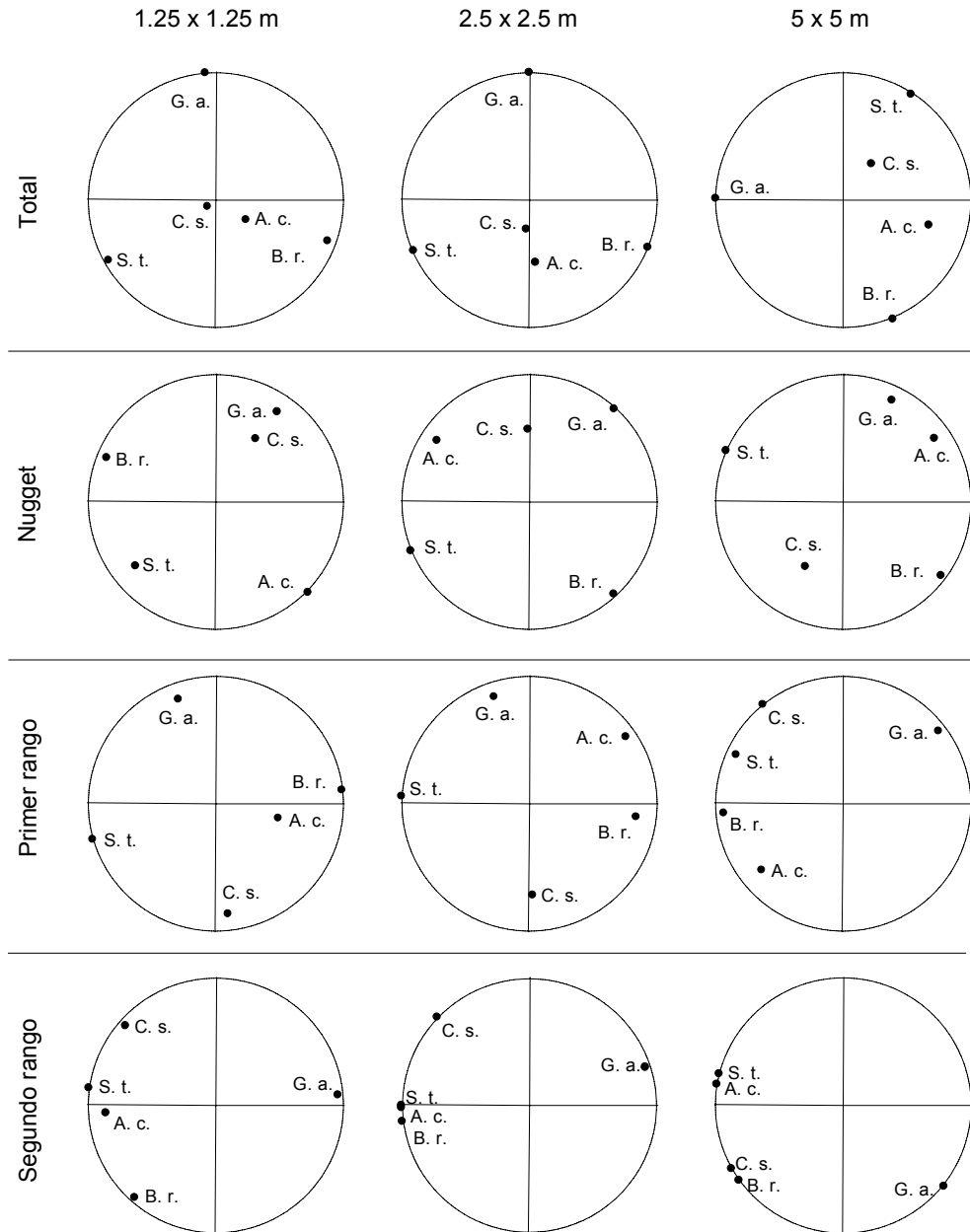
**Figura 6.24.** Semivariogramas cruzados (puntos) y modelo de correogionalización ajustado (línea continua) utilizando la malla de muestreo de 2,5 x 2,5 m. Los valores en el eje y representan la semivarianza.



**Figura 6.25.** Semivariogramas cruzados (puntos) y modelo de correogionalización ajustado (línea continua) utilizando la malla de muestreo de 1,25 x 1,25 m. Los valores en el eje y representan la semivarianza.

Las mismas relaciones positivas y negativas se observaron en la matriz correspondiente al primer rango para la malla de menor resolución. Las correlaciones estructurales para la escala definida por el primer rango en el caso de las mallas de 2,5 x 2,5 m y 1,25 x 1,25 m, correspondiente a un rango de unos 4 m, mostraron un panorama diferente. Los patrones de *B. retusum* y *A. cytisoides* se relacionaron positivamente, mientras que el patrón de *S. tenacissima* mostró correlaciones negativas con el resto de las especies, especialmente con *B. retusum* y *A. cytisoides*. Las correlaciones a distancias nulas, o escala de “nugget”, fueron mayoritariamente negativas, tal como refleja la distribución dispersa de todas las especies en los círculos de correlación de esta escala. En los círculos correspondientes a todo el conjunto de datos (marcados como “Total” en la Fig. 6.26) destacan, fundamentalmente, la disociación del patrón de *G. alypum* y, con menor intensidad, la asociación negativa entre los patrones de *B. retusum* y *S. tenacissima*.

El ajuste conjunto de un modelo de correogionalización a los datos refleja un patrón común de variación para las especies presentes. Este hecho supone la existencia de un patrón subyacente de alguna propiedad ambiental



**Figura 6.26.** Círculos de correlación para las distintas mallas de muestreo y matrices definidas en el modelo lineal de correogionalización. *S. t.* = *Stipa tenacissima*, *G. a.* = *Globularia alypum*, *B. r.* = *Brachypodium retusum*, *A. c.* = *Anthyllis cytisoides* y *C. s.* = *caméfitos*.



que determina los patrones de la vegetación o, alternativamente, refleja la dominancia de la distribución espacial de alguna de las especies –o de alguna asociación de ellas– sobre el resto de patrones. La hipótesis inicial de trabajo se corresponde más o menos con la segunda posibilidad, referida a la influencia del patrón de *S. tenacissima* sobre los demás. No obstante, los resultados de los análisis de asociación entre los patrones específicos no confirmaron esta idea. Por otra parte, estos análisis manifiestan de forma consistente la asociación positiva entre los patrones de *B. retusum* y *A. cytisoides* y la asociación negativa entre la distribución espacial de *G. alypum* y del resto de las especies. Conviene mencionar que los patrones de asociación y disociación encontrados por el modelo lineal de correogionalización con estos datos fueron, en líneas generales, concordantes con los encontrados al aplicar el método SADIE con estos datos (véase el apartado 4.6.2).

El modelo lineal de correogionalización presenta como principal atractivo la variedad y el detalle de la información que ofrece. Los semivariogramas proporcionan información sobre los cambios con la distancia en el patrón espacial de las especies por separado, así como en la covariación entre pares de especies. El modelo de correogionalización permite comprobar hasta que punto las distintas especies de la comunidad presentan un patrón de variación espacial común, así como un análisis independiente de las correlaciones espaciales entre ellas a las escalas de interés mediante las correlaciones estructurales. No obstante, hay que tener en cuenta que su base matemática es compleja, requiriendo su empleo de ciertos conocimientos previos sobre álgebra de matrices y programación, y que hasta la fecha ninguno de los programas geoestadísticos disponibles permite realizar todos los cálculos necesarios de una forma intuitiva.

## 6.8. REVISIÓN DE SOFTWARE

A continuación se presenta una lista de software, tanto libre (Tabla 6.5) como comercial (Tabla 6.6), para realizar análisis geoestadísticos. Una buena parte de los análisis presentados en este capítulo han sido realizados con los módulos (o librerías) *geoR* y *gstat*, dentro del entorno del software R. R (y sus módulos) forman parte de lo que se conoce como software libre, que implica que cualquiera puede disponer del código y mejorarlo. Este tipo de software tiene un desarrollo muy activo, y por ello, en cons-

**Tabla 6.5.** Listado de software libre y/o gratuito para realizar análisis geoestadísticos (actualizado en mayo 2007). Características: V = Variografía, K = "kriging", O = Otros métodos de estima (NN, IDW, "splines", etc.), C = "cokriging", 2D/3D. = número máximo de dimensiones con las que puede trabajar el programa (2/3), S = simulaciones y G = funciones GIS).

Nombre	URL	Plataforma	Características
Agromet	<a href="http://www.enge.ucl.ac.be/recherche/projets/agromet/">http://www.enge.ucl.ac.be/recherche/projets/agromet/</a>	Unix / Windows / C++ codes	V-K-C-2D
Cosim	<a href="http://www.ai-geostats.org/index.php?id=110">http://www.ai-geostats.org/index.php?id=110</a>	Windows / Fortran	S-2D
ExploStat	<a href="http://www.explostat.nl/">http://www.explostat.nl/</a>	Windows	V-K-O-C-2D-G
E{Z}-Kriging	<a href="http://www.ai-geostats.org/index.php?id=114">http://www.ai-geostats.org/index.php?id=114</a>	Windows	V-K-2D
GCOSIM3D	<a href="http://www.ai-geostats.org/index.php?id=108">http://www.ai-geostats.org/index.php?id=108</a>	Windows / C codes	S-3D
Geo-EAS	<a href="http://www.ai-geostats.org/index.php?id=121">http://www.ai-geostats.org/index.php?id=121</a>	Windows / Linux / Dos	V-K-2D
GeoPack	<a href="http://www.epa.gov/nrl/esd1/databases/geo-pack/access.htm">http://www.epa.gov/nrl/esd1/databases/geo-pack/access.htm</a>	Windows / Dos	V-K-C-2D
GeoDa	<a href="https://www.geoda.uiuc.edu/default.php">https://www.geoda.uiuc.edu/default.php</a>	Windows	O-2D
Geostatistical Toolbox	<a href="http://www.ai-geostats.org/index.php?id=159">http://www.ai-geostats.org/index.php?id=159</a>	DO5	V-K-C-3D
GMT	<a href="http://gmt.soest.hawaii.edu/">http://gmt.soest.hawaii.edu/</a>	Unix / C codes	O-2D
GRNN	<a href="http://www.ai-geostats.org/index.php?id=169">http://www.ai-geostats.org/index.php?id=169</a>	Windows	O-2D
GSLIB	<a href="http://www.gslib.com/">http://www.gslib.com/</a>	Windows / Unix / Fortran 77	V-K-C-3D-S
Gstat	<a href="http://www.gstat.org/">http://www.gstat.org/</a>	Linux, Windows / C codes / R	V-K-C-3D-S
ISIM3D	<a href="http://www.ai-geostats.org/index.php?id=108">http://www.ai-geostats.org/index.php?id=108</a>	Windows / C codes	S-3D
Kriging	<a href="http://www.nbb.cornell.edu/neurobio/land/OldStudentProjects/cs490-94to95/dlang/kriging.html">http://www.nbb.cornell.edu/neurobio/land/OldStudentProjects/cs490-94to95/dlang/kriging.html</a>	Unix / C codes	K-2D
$\mathbb{R}$ (geoR, geoRglm y gstat)	<a href="http://cran.r-project.org/">http://cran.r-project.org/</a>	Windows, Unix, OsX	V-K-C-3D-S
SADA	<a href="http://www.item.utk.edu/~sada/index.shtml">http://www.item.utk.edu/~sada/index.shtml</a>	Windows	V-K-O-3D-G
SAGA GIS	<a href="http://www.ai-geostats.org/index.php?id=124">http://www.ai-geostats.org/index.php?id=124</a>	Windows	V-K-O-2D-G
SGS	<a href="http://www.frieder-tonn.de/projects/sgs/index.html">http://www.frieder-tonn.de/projects/sgs/index.html</a>	Linux, C codes	K-S-2D
S-GeMS	<a href="http://sgems.sourceforge.net/">http://sgems.sourceforge.net/</a>	Windows, Linux, C++ codes	V-K-C-3D-S
Spherekit	<a href="http://www.ncgia.ucsb.edu/pubs/spherekit/">http://www.ncgia.ucsb.edu/pubs/spherekit/</a>	Unix, C codes	K-O-2D
Surface III	<a href="http://www.kgs.ku.edu/Tis/surf3/surf3Home.html">http://www.kgs.ku.edu/Tis/surf3/surf3Home.html</a>	Mac	K-O-2D
Surfit	<a href="http://surfit.sourceforge.net/index.html">http://surfit.sourceforge.net/index.html</a>	Windows, C++ codes	O
UNCERT	<a href="http://uncert.mines.edu/">http://uncert.mines.edu/</a>	Unix / C codes	V-K-O-C-2D-G
Variowin	<a href="http://www.sst.unil.ch/research/variowin/index.html">http://www.sst.unil.ch/research/variowin/index.html</a>	Windows	V-2D
Vesper	<a href="http://www.usyd.edu.au/su/agric/acpa/vesper/vesper.html">http://www.usyd.edu.au/su/agric/acpa/vesper/vesper.html</a>	Windows	V-K-2D

tante mejora. En él se pueden encontrar los últimos procedimientos desarrollados por los estadísticos. Por el contrario, muchos de los programas que se mencionan en este apartado se han dejado de desarrollar, aunque no por ello dejan de ser programas útiles (especialmente para los que se adentran por primera vez en el mundo de la geoestadística).

Los dos módulos de geoestadística disponibles en R (*geoR* y *gstat*) proporcionan algunas herramientas en común, aunque no son exactamente iguales. Por ejemplo, *geoR* no permite el “cokriging”, ni tampoco la interpolación directa en bloques (aunque con un sencillo programa se puede incluir). Por el contrario, *geoR* incluye más modelos de continuidad espacial, y utiliza las transformaciones de “box-cox” para la normalización de los datos. Además, como se ha comentado, es capaz de aplicar a la geoestadística los modelos lineales generalizados.

**Tabla 6.6.** *Software comercial disponible para realizar análisis geoestadísticos (lista actualizada en mayo del 2007).*

Software	URL
Datamine	<a href="http://www.datamine.co.uk/">http://www.datamine.co.uk/</a>
Geostokos Toolkit y ECOSSE	<a href="http://www.stokos.demon.co.uk/toolkit.htm">http://www.stokos.demon.co.uk/toolkit.htm</a>
Geostat Systems Int. Inc.	<a href="http://www.geostat.com/">http://www.geostat.com/</a>
GS+	<a href="http://www.gammadesign.com/">http://www.gammadesign.com/</a>
Genstat	<a href="http://www.vsni.co.uk/products/genstat/">http://www.vsni.co.uk/products/genstat/</a>
Surfer	<a href="http://www.goldensoftware.com/">http://www.goldensoftware.com/</a>
SAGE2001	<a href="http://www.isaaks.com/">http://www.isaaks.com/</a>
Lynx Geosystems	<a href="http://www lynxgeo.com/">http://www.lynxgeo.com/</a>
GeostatsOffice	<a href="http://www.ibrae.ac.ru/~mkanev/eng/Software.html">http://www.ibrae.ac.ru/~mkanev/eng/Software.html</a>
Isatis	<a href="http://www.geovariances.com/">http://www.geovariances.com/</a>
Systat	<a href="http://www.systat.com/">http://www.systat.com/</a>
SurGe	<a href="http://www.geocities.com/miroslavdressler/surgemain.htm">http://www.geocities.com/miroslavdressler/surgemain.htm</a>
WinGslib	<a href="http://www.statios.com/WinGslib/">http://www.statios.com/WinGslib/</a>
S+	<a href="http://www.insightful.com/">http://www.insightful.com/</a>

Del software gratuito (pero no libre) habría que destacar a *Variowin*. Aunque aparentemente no está en desarrollo, la capacidad de este programa para ajustar modelos de continuidad espacial a los semivariogramas hace que sea uno de los más intuitivos y fáciles que existen, obteniéndose con su uso ajustes muy similares a los obtenidos utilizando estimadores basados en mínimos cuadrados (Webster y Oliver 1997). Como software comercial, habría que destacar entre los disponibles al programa GS+, que va implementando nuevas características versión a versión. La última revisada por el autor incluía simulaciones. Este tipo de programas son muy fáciles de aprender y utilizar, siendo sin embargo poco versátiles. Además, siempre van por detrás de los últimos avances incluidos en R. Entre los programas comerciales también habría que citar la “versión comercial” de R, conocida como S+. Este programa, que tiene interfaz gráfica interactiva, dispone prácticamente de los mismos módulos de R, si bien su precio es bastante elevado.

## 6.9. PÁGINAS WEB DE INTERÉS

Entre todas las páginas Web disponibles, es altamente recomendable visitar la web <http://www.ai-geostats.org>, dedicada al intercambio de información sobre métodos de ecología espacial y que está especialmente centrada en la geoestadística. En ella se puede encontrar una amplia lista de programas de geoestadística, así como múltiple información y una lista de correo muy activa sobre este tema. Otra página interesante es <http://uncert.mines.edu/tutor/>, donde puede encontrarse un tutorial sobre la creación de semivariogramas con ejemplos interactivos. La página Web de *gstat* (<http://www.gstat.org>), un programa de código libre para el modelado de estadística multivariable, predicción y simulación que puede utilizarse de forma independiente o como módulo de R (y S+), es también muy interesante, ya que además de podernos descargar las distintas versiones de este versátil programa, contiene separatas, datos e incluso una lista de correo. El paquete de análisis de datos geoestadístico *geoR*, ampliamente utilizado en todo el capítulo, puede descargarse de la página Web <http://leg.ufpr.br/geoR/>. Otra página Web de interés para todo aquel que utilice el software R para el análisis de datos ecológicos espacialmente explícitos es <http://cran.r-project.org/src/contrib/Views/Spatial.html>. En esta página, mantenida por Roger Bivand, se resumen todos los módulos de R relacionados

con el análisis de datos espaciales, incluyendo un apartado para geostatística. Algunas páginas en castellano proporcionan numerosa información de utilidad. Por ejemplo en la página Web del Grupo de Modelación Matemática y Computacional de la Universidad Autónoma de México puede descargarse un completo manual sobre geostatística en pdf realizado por el Dr. Martín A. Díaz Vera ([www.mmc.igeofcu.unam.mx/cursos/geoest/GeoEstadistica.pdf](http://www.mmc.igeofcu.unam.mx/cursos/geoest/GeoEstadistica.pdf)), donde pueden ampliarse buena parte de los contenidos de este capítulo. También puede encontrarse distinta documentación en castellano sobre geostatística en la página Web <http://www.monografias.com/trabajos14/geoestadistica/geoestadistica.shtml>.

## AGRADECIMIENTOS

AG quiere agradecer a William H. Schlesinger, quien le mostró por primera vez estas técnicas al amparo del aire acondicionado en un apartamento de Las Cruces (Nuevo Méjico), en plena campaña del desierto. FTM agradece a Richard Webster toda su ayuda y paciencia durante la estancia que realizó en Rothamsted para aprender geostatística, así como todos sus consejos durante estos años. Este capítulo se ha elaborado gracias a un contrato Ramón y Cajal del Ministerio de Educación y Ciencia y a un Early Career Project Grant (ECPG 231/607), otorgado por la British Ecological Society a FTM. Este autor agradece también el apoyo de los proyectos CEFEMED (URJC-RNT-063-2), financiado por la Dirección General de Universidades e Investigación de la Consejería de Educación de la Comunidad de Madrid (DGUICAM) y la Universidad Rey Juan Carlos, INTERCAMBIO (BIOCON 06\_039), financiado por la Fundación BBVA, y REMEDINAL (S-0505/AMB/0335), financiado por la DGUICAM.