



## Monográfico

### Revisiones

de la Cruz Rot M. 2006. Introducción al análisis de datos mapeados o algunas de las (muchas) cosas que puedo hacer si tengo coordenadas. *Ecosistemas*. 2006/3

(URL: [http://www.revistaecosistemas.net/articulo.asp?Id=448&Id\\_Categoria=1&tipo=portada](http://www.revistaecosistemas.net/articulo.asp?Id=448&Id_Categoria=1&tipo=portada))

## Introducción al análisis de datos mapeados o algunas de las (muchas) cosas que puedo hacer si tengo coordenadas

M. de la Cruz Rot

Depto. Biología Vegetal, E.T.S. Ingenieros Agrónomos, Universidad Politécnica de Madrid, E-2840 Madrid. ([marcelino.delacruz@upm.es](mailto:marcelino.delacruz@upm.es))

*Introducción al análisis de datos mapeados o algunas de las (muchas) cosas que puedo hacer si tengo coordenadas. El análisis más común de los patrones espaciales de puntos consiste en un test de aleatoriedad a todas las escalas. Como la aleatoriedad suele ser la excepción más que la regla en la naturaleza, debe continuarse con la modelización del proceso espacial que es capaz de generarlo. En el caso de patrones marcados pueden testarse diferentes hipótesis relativas a la asociación espacial de las diferentes marcas: independencia, etiquetado aleatorio, ausencia de correlación entre marcas, gradiente en el valor de las marcas, etc.*

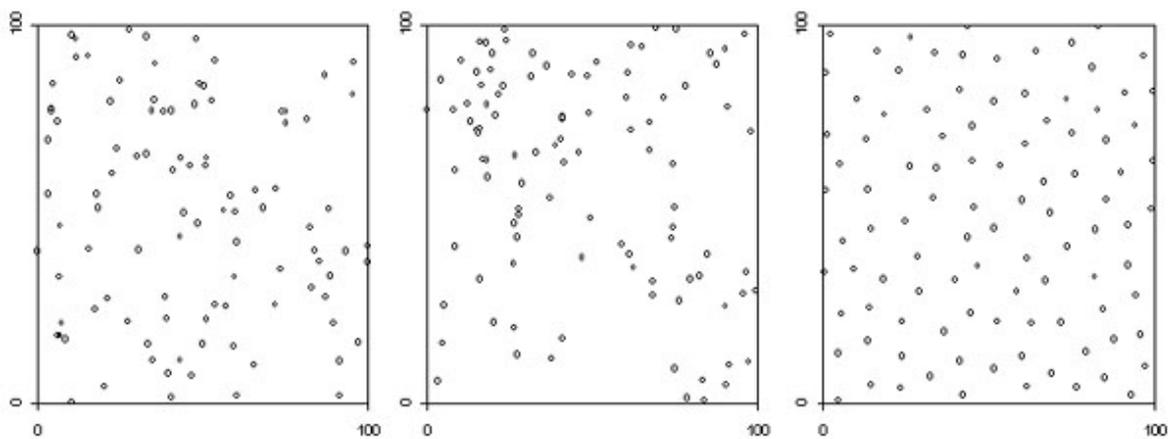
*Palabras clave: patrones marcados, independencia, etiquetado aleatorio, coordenadas*

*An introduction to spatial point pattern analysis or just a few (of many) things that could be done with mapped data. The most frequent analysis of a spatial point pattern is a test of complete spatial randomness. As randomness is more of an exception than the rule in the natural world it would be usually followed by modelling the generating spatial process. Different hypothesis can be tested for marked patterns regarding the spatial dependence between marks: independence, random labelling, correlation, gradients in the value of marks, etc.*

*Key words: spatial pattern, independence, random labelling, coordinates.*

### Introducción

Quizás la forma más fiel de reflejar la estructura espacial de una población, comunidad, o de cualquier fenómeno ecológico de naturaleza discreta (manchas de hábitat, perturbaciones, etc.) es la representación cartográfica de todos los elementos del mismo en una región geográfica concreta. En muchas ocasiones, dependiendo de la escala de estudio, tales elementos pueden describirse aceptablemente mediante sus coordenadas espaciales (x, y), generándose así un conjunto de datos que recibe el nombre de *patrón espacial de puntos* (Diggle, 2003). La metodología habitual en el estudio de estas estructuras asume que el patrón espacial de puntos de una población, comunidad, etc., es una realización concreta de un *proceso espacial de puntos* subyacente, que hay que describir, y cuyas propiedades son una buena descripción del patrón concreto. Un proceso de puntos es un proceso estocástico que 'genera' patrones de puntos aleatorios que comparten la misma estructura espacial (la ley del proceso), por ejemplo patrones de Poisson (distribución completamente al azar), regulares o agrupados (**Fig. 1**).



**Figura 1.** Ejemplos de patrones aleatorios o de Poisson (*izquierda*), agrupado (*centro*) y uniforme (*derecha*).

Una de las aplicaciones para las que se suelen emplear las técnicas de análisis de patrones de puntos es para inferir la existencia de interacciones en comunidades y poblaciones a partir del estudio del patrón espacial (la disposición) de los individuos. Su empleo se basa en la asunción de que el análisis del patrón espacial y de sus variaciones en el espacio y en el tiempo podría explicar los mecanismos subyacentes a la construcción de la estructura y al funcionamiento de la dinámica de poblaciones y comunidades (Wiegand *et al.*, 2003, Seabloom *et al.*, 2005).

La naturaleza del patrón generado por procesos biológicos puede estar afectada por la escala a la que el proceso es observado. La mayoría de los ambientes naturales muestran heterogeneidad a una escala lo suficientemente grande como para permitir la aparición de patrones agregados. A una escala menor, la variación ambiental puede ser menos acentuada y el patrón estará determinado por la intensidad y la naturaleza de las interacciones entre los individuos (Diggle, 2003). La tradicional clasificación de los patrones -como la que aparece en la mayoría de los libros de texto de ecología- en regular, aleatoria o agregada, es por lo tanto una simplificación excesiva, que aunque puede resultar útil en una primera etapa exploratoria del análisis debería dejar paso a una descripción más detallada y multidimensional basada en estadísticos funcionales o en la formulación de modelos explícitos del proceso subyacente.

La descripción de un proceso de puntos

Bajo la asunción de estacionaridad (el proceso es homogéneo o invariante a la translación) e isotropía (el proceso es invariante a la rotación), las características principales de un proceso de puntos pueden ser sumariadas por su propiedad de primer orden ( $\lambda$ , o intensidad: el número esperado de puntos por unidad de área en cualquier localidad), y por su propiedad de segundo orden, que describe las relaciones entre pares de puntos (p. ej. la probabilidad de encontrar un punto en las inmediaciones de otro). En el caso de patrones uniformes o regulares, la probabilidad de encontrar un punto en las inmediaciones de otro es menor de la que tendría un patrón aleatorio mientras que en los patrones agrupados la probabilidad es mayor. El estimador más popular de las propiedades de segundo orden es la función K de Ripley, que las estima a todas las escalas.

**La función de K de Ripley (y otras)**

La función K se define como

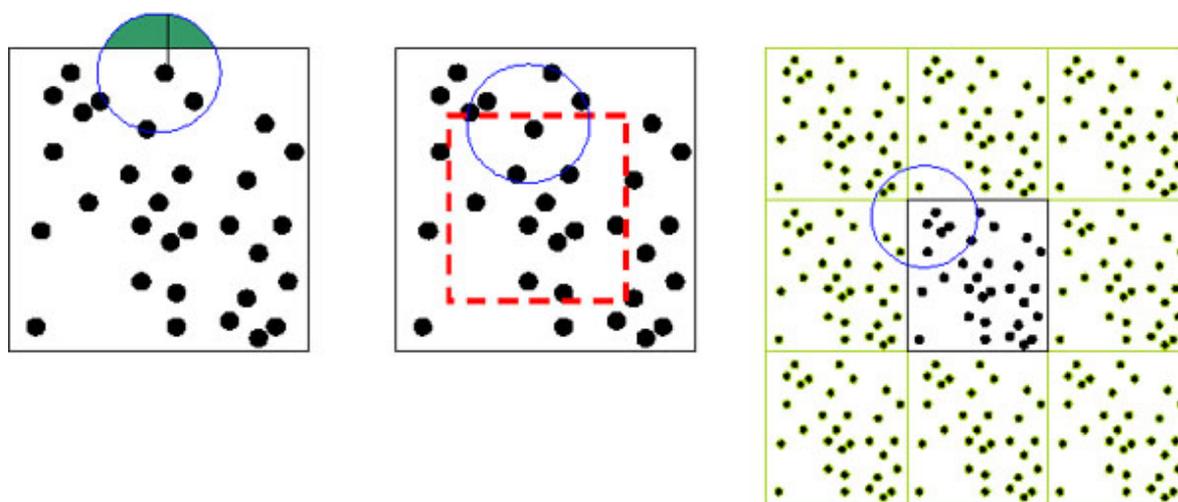
$$K(r) = \lambda^{-1} [\text{n}^\circ \text{ medio de individuos en un radio } r \text{ alrededor de cualquier individuo}]$$

siendo  $\lambda$  la densidad de individuos (nº de individuos por unidad de área). Con frecuencia se define la función K diciendo que  $\lambda K(r)$  es el número medio de individuos dentro de un círculo de radio  $r$  alrededor de un individuo 'típico' del patrón (sin contar dicho individuo central). Por lo tanto,  $K(r)$  describe las características del proceso de puntos a muchas escalas (tantas como diferentes  $r$  consideremos). La forma más sencilla de estimar  $\lambda$  y  $K(r)$  sería:

$$\hat{\lambda} = \frac{N}{A} \qquad \hat{K}(r) = \frac{1}{\hat{\lambda}} \frac{1}{N} \sum_{i=1}^N \sum_{j \neq i}^N I(d_{ij} < r)$$

donde  $N$  es el número de puntos del patrón,  $A$  la superficie del área de estudio y  $I(d_{ij} < r)$  la función indicadora, que toma el valor de 1 si la distancia entre los puntos  $i$  y  $j$  es menor que  $r$  y 0 en el caso contrario.

En la práctica, y dado que el límite del área de estudio suele ser arbitrario, es necesario introducir un factor que corrija el 'efecto borde' (**Fig. 2**). Los 'efectos borde' surgen porque los puntos que aparecen fuera de los límites del área de estudio no son tenidos en cuenta para estimar  $K(r)$  aunque se encuentren a una distancia menor de  $r$  de un punto situado dentro del área. Si no se tienen en cuenta, los efectos borde producen estimaciones sesgadas de  $K(r)$ , especialmente para valores grandes de  $r$ . Se han propuesto diferentes mecanismos y estimadores para corregir el efecto borde, como ponderar los recuentos alrededor de puntos próximos al borde (Ripley, 1988), replicar el patrón alrededor del área de estudio (Osher, 1983) o establecer bandas 'tampón' en la periferia del área de estudio que proporcionen puntos y que eviten el efecto borde (Ripley, 1988). Una revisión de los métodos de corrección del efecto borde puede consultarse en Haase (1995) y Goreaud y Pelissier (1999). De todas formas, dado que los mecanismos que corrigen el efecto borde no son perfectos, se suele recomendar no calcular  $K(r)$  más allá de  $r < 1/3$  de la longitud del lado más corto del área de estudio (Baddeley y Turner, 2005) o hasta  $r < (A/2)^{1/2}$  en el caso de áreas no rectangulares, (Dixon, 2002). Otros autores (Lancaster y Downes, 2004) han puesto de manifiesto la importancia de comprobar si el efecto borde es necesario en el contexto del estudio ecológico que se realice (no tendría sentido corregirlo, por ejemplo, en el análisis de poblaciones completas con límites naturales).

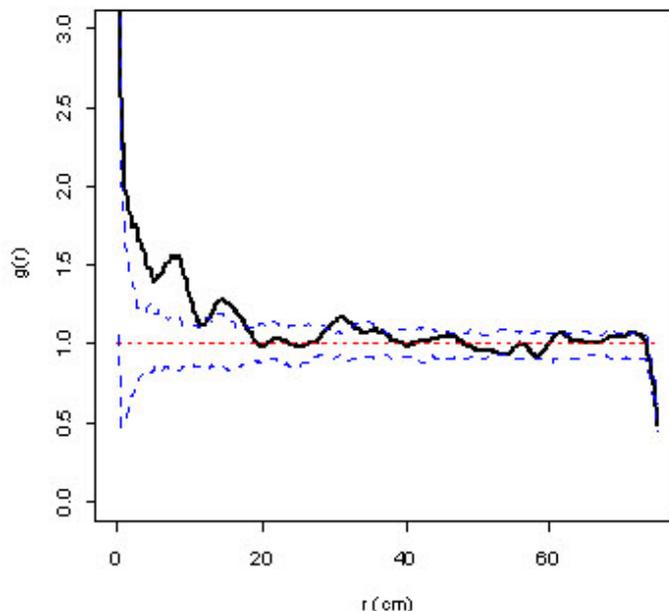


**Figura 2.** Corrección del efecto borde. **Izquierda.** Método de Ripley. A las estimaciones de los puntos próximos al borde se les da un peso proporcional a la porción del círculo que queda fuera del límite del área de estudio (para compensar por los puntos no registrados y que podrían encontrarse en las inmediaciones). **Centro.** Método del área tampón. La estimación de la función  $K(r)$  sólo se hace a partir de los puntos incluidos dentro de un área de menor tamaño que la original, de forma que los puntos que quedan en el área 'tampón' externa se pueden contar (no hay efecto borde). El grosor del área tampón debe ser igual al valor máximo de  $r$  para el que se va a estimar la función. **Derecha.** Método de la traslación (sólo posible en áreas de estudio rectangulares): el patrón se repite en todos los lados del área de estudio, asumiendo que es representativo de lo que no se ha registrado fuera.

Otras funciones que se han empleado para describir y testar patrones espaciales están basadas en la distribución de distancias entre puntos que existiría en un patrón de Poisson, como por ejemplo la función de distribución de distancias al vecino más próximo  $G(r)$ , la función de distribución de distancias a un punto fijo aleatorio (también llamada función de espacio vacío)  $F(r)$  (Diggle 1979, 2003), o la recientemente descrita función  $J$  de Van Lieshout y Baddeley (1996), una combinación de las anteriores que presenta la ventaja de ser insensible al 'efecto borde' y no necesitar mecanismos de corrección. Aunque se suele recomendar el empleo simultáneo de varias de estas funciones debido a sus propiedades complementarias (Ripley, 1981; Diggle, 2003), no suele ser la norma en los trabajos ecológicos. Una excepción interesante es el trabajo de Barot *et al.*, (1999), en el que se emplean conjuntamente  $K$ ,  $G$  y  $F$ .

Todas las funciones anteriores, incluida la función  $K$ , son en cierta forma funciones de distribución acumulada ya que, a cada escala o distancia  $r$ , todos los pares de puntos separados por una distancia *menor* que  $r$  se usan para estimar el valor de la correspondiente función. En ocasiones puede ser necesario disponer de una función que caracterice de forma no acumulativa el patrón, es decir que tenga en cuenta tan sólo los pares de puntos que se encuentran separados por una distancia *exactamente* igual o similar a la distancia  $r$ . La función de correlación de par  $g(r)$  (*pair correlation function*; Stoyan y Stoyan 1994) es la herramienta apropiada en este caso. En un patrón de Poisson,  $g(r) \equiv 1$  para todas las distancias  $r$ ; en cualquier patrón empírico, valores de  $g(r) > 1$  indican que las distancias alrededor de  $r$  son relativamente más frecuentes de lo que

serían en el caso de CSR (por ejemplo, si se tratase de valores pequeños de  $r$  indicaría agrupamiento) mientras que valores de  $g(r) < 1$  indican que las correspondientes distancias son menos frecuentes de lo esperado, lo que podría indicar algún tipo de inhibición (Stoyan y Penttinen, 2000) (**Fig. 3**). Una versión re-escalada de la función de correlación de par es la *O-ring* (Wiegand y Moloney 2004). El término *ring* hace referencia a que, a diferencia de los círculos de radio  $r$  alrededor de cada punto que se emplean para calcular la función  $K$  empírica, aquí se emplean 'anillos' (coronas circulares) de radio similar a  $r$  y grosor variable.



**Figura 3.** Función de correlación de par del patrón de puntos de la comunidad gipsófila de la figura 4. Los valores de  $g(r) > 1$  indican que las distancias entre puntos de valor alrededor de  $r$  son más frecuentes de lo que serían en un proceso CSR. En este caso, al tratarse de valores de  $r$  pequeños ( $< 20$  cm), la gráfica indica la existencia de agrupamiento. Cuando la función de correlación de par tiene sólo un pico, el tamaño medio de los agregados o manchas puede estimarse a partir de la escala para la que el valor de  $g(r)$  cae a 1 [en este caso alrededor de 20 cm, lo que coincide con la estimación a partir de  $L(r)$  en la figura 4].

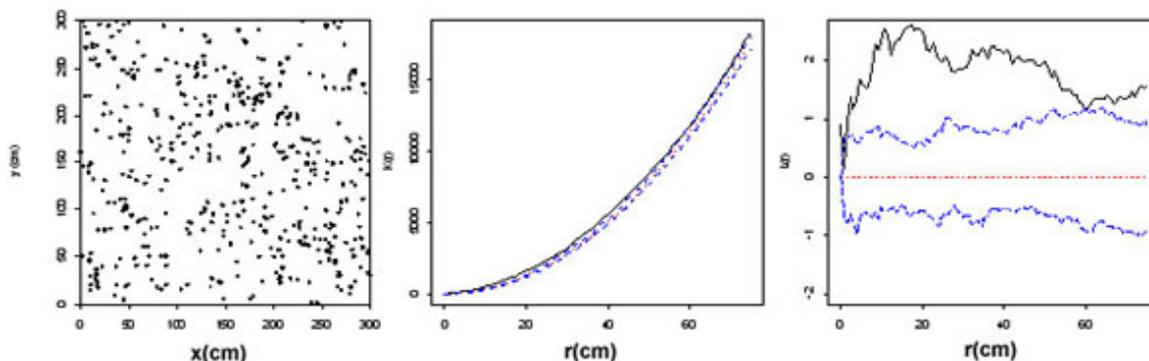
## Algunas preguntas que se pueden responder con datos mapeados

¿Cuál es la distribución de los individuos (agrupada, al azar, uniforme)?

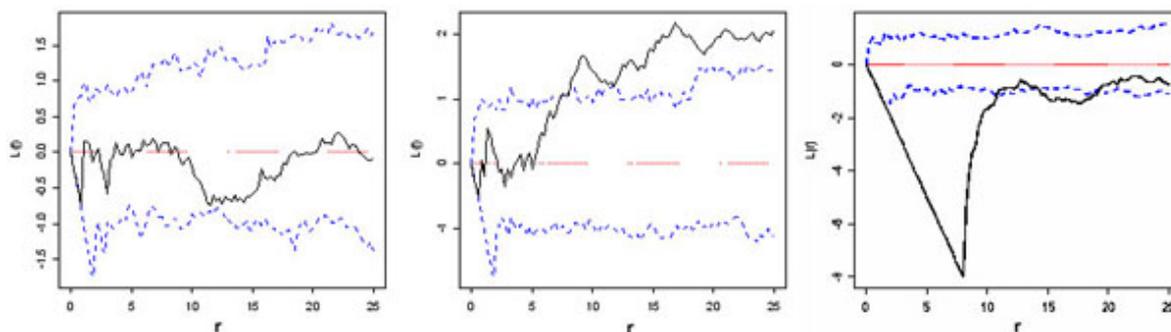
Esta es quizás la primera cuestión que se plantea a la hora de analizar la estructura espacial de un patrón de puntos, y una de los principales usos que se le ha dado a la función  $K$  en el contexto ecológico. La respuesta se basa en realizar un test de aleatoriedad espacial completa (o 'CSR', de *Complete Spatial Randomness*), o lo que es lo mismo, aleatoriedad en todas las escalas. En realidad la prueba testa si el patrón observado puede considerarse como una realización de un proceso de Poisson homogéneo. El test consiste en calcular la función  $K$  del patrón de puntos observado y compararlo con el de la función  $K$  teórica de un patrón de Poisson de la misma intensidad. Es fácil comprobar que el valor teórico esperado asumiendo CSR es siempre  $\pi r^2$  (y que asimismo en el caso de patrones agregados este valor sería mayor y menor en el caso de patrones uniformes). La comparación visual o numérica de la curva de  $K(r)$  y la de  $\pi r^2$  es una buena forma de análisis exploratorio de los datos, con valores de  $K(r) > \pi r^2$  indicando agregación y valores de  $K(r) < \pi r^2$  indicando uniformidad o regularidad a la escala considerada. Dado que en áreas de estudio 'reales', con límites definidos, se pueden producir variaciones estocásticas que causen valores mayores o menores que  $\pi r^2$  incluso para patrones de Poisson, el test se completa con la construcción de intervalos de confianza empleando el método de Monte Carlo. Se simula un número elevado de patrones de Poisson con la misma intensidad y en un área del mismo tamaño que el patrón observado. Los valores de la función de cada uno de ellos se ordenan y bien se representan en forma de 'envuelta' para cada valor de  $r$  los valores máximo y mínimo alcanzados o bien se calculan y representan gráficamente los percentiles apropiados, por ejemplo el 2.5%

y el 97.5% , para formar una *envuelta* que representa un intervalo de confianza del 95 % (Dixon, 2002, Moller y Waagepetersen, 2006). De esta forma tan sólo se consideran patrones no-CSR aquellos en los que aunque  $K(r)$  sea distinta de  $\pi r^2$ , sus valores queden por encima o por debajo de los valores límites representados por las envueltas.

En la práctica, suele emplearse con mayor frecuencia la función  $L(r) = (K(r)/\pi)^{1/2}$  que además de tener una varianza constante (Ripley, 1979), permite una interpretación más cómoda del test (**Fig. 4**, **Fig. 5**). Bajo CSR,  $L(r) = r$  y por lo tanto puede testarse si  $L(r) - r = 0$  a cada distancia  $r$ . Los valores críticos de  $L(r) - r$  se calculan también con simulación Monte-Carlo. Otras representaciones/test alternativas incluyen  $K(r) - \pi r^2$  (con valor teórico de 0) y  $[K(r)/\pi r^2] - 1$  (con valor esperado de 0 y desviaciones proporcionales al exceso o mengua de individuos).



**Figura 4. Izquierda.** Patrón espacial de puntos de una comunidad gipsófila del centro de la Península Ibérica. **Centro.** Función  $K(r)$  del patrón observado (negro continuo),  $K(r)$  teórica asumiendo CSR (rojo punteado) y envueltas con los máximos y mínimos valores obtenidos en 99 simulaciones de patrones CSR de la misma intensidad (azul, rayado). **Izquierda.** Función  $L(r) [= (K(r)/\pi)^{1/2} - r]$  del mismo patrón. Los colores y patrones de las curvas igual que en el gráfico de  $K(r)$ . La curva de la función  $L$  empírica discurre fuera del espacio definido por las envueltas obtenidas en la simulación de patrones CSR, lo que indica la existencia de agregación. En la representación de la función  $L$ , el valor máximo suele indicar el tamaño del agregado o mancha típica (en este caso, alrededor de 20 cm).



**Figura 5.** Función  $L(r) [= (K(r)/\pi)^{1/2} - r]$  de los patrones aleatorio (izquierda), agrupado (centro) y uniforme (derecha) de la figura 1. La curva negra continua representa la función  $L(r)$  empírica y la roja punteada el valor teórico de la hipótesis CSR. Las curvas azules rayadas representan los percentiles 2,5 % y 97,5 % de los valores alcanzados por la función  $L(r)$  en 999 simulaciones de patrones de Poisson de la misma intensidad que el original. En la gráfica de la izquierda la función  $L(r)$  empírica queda dentro de la banda definida por el intervalo de confianza, por lo que no se puede rechazar la hipótesis de CSR (es decir el patrón de puntos sería una realización de un proceso de Poisson). En la gráfica del centro, la  $L(r)$  empírica tiene valores mayores que los del límite superior del intervalo de confianza a partir de  $r = 8$  metros, por lo que se puede rechazar la hipótesis de CSR con un riesgo  $\alpha = 5\%$  a favor del agrupamiento a escalas mayores de 8 metros. En la gráfica de la derecha, la  $L(r)$  empírica tiene valores menores que el límite inferior del intervalo de confianza prácticamente para todas las distancias menores de 11 metros, lo que indicaría una fuerte inhibición entre los puntos a esa escala y permitiría rechazar también, con el mismo  $\alpha$ , la hipótesis de CSR.



## ¿Cómo se puede describir el patrón?

A pesar de su amplia aplicación en ecología, el test de CSR no suele tener un interés intrínseco ya que la distribución aleatoria no suele ser la norma en la naturaleza. Sin embargo su uso como herramienta exploratoria permite formular hipótesis interesantes respecto a los patrones y su génesis. El objetivo del análisis estadístico de un patrón de puntos después del test de CSR normalmente consistirá en obtener una descripción más detallada mediante la formulación de un modelo (paramétrico o no) que describa el proceso generador del patrón observado y su ajuste a los datos disponibles. En la literatura sobre estadística espacial se han desarrollado y explorado numerosos modelos que describen procesos no-CSR y que podrían tomarse como vías para continuar el análisis según los resultados del test (ver una introducción por ejemplo en Diggle, 2003).

Uno de los más célebres y con evidentes aplicaciones en ecología es el proceso de Poisson agrupado (también llamado proceso de Neyman-Scott). Este proceso incorpora una forma explícita de agrupamiento espacial y por lo tanto proporciona una base bastante satisfactoria para modelizar patrones agregados. Se basa en tres postulados (Diggle, 2003):

1. Existe un conjunto de puntos 'padres' con distribución completamente aleatoria (CSR) e intensidad  $\rho$ .
2. El número de puntos 'hijos' de cada padre sigue una distribución de Poisson con media  $m$ .
3. Los puntos 'hijos' se distribuyen alrededor de los padres siguiendo una distribución gaussiana bivariada, con media 0 y varianza  $\sigma^2$  en  $x$  e  $y$ .

Para un matemático es fácil demostrar que la función  $K$  teórica del proceso es:

$$K(r; \sigma, \rho) = \pi r^2 + \rho^{-1} [1 - \exp(-r^2/4\sigma^2)]$$

El proceso de Strauss es apropiado para modelizar repulsión o regularidad a pequeña escala. Se basa en que el número de vecinos a una distancia crítica  $\delta$  de cualquier punto es menor al esperado en caso de CSR. Se genera a partir de un proceso de Poisson homogéneo en donde una proporción  $1-\gamma$  de los puntos dentro de la distancia  $\delta$  son borrados. Una aproximación de la función  $K$  para el proceso de Strauss de núcleo blando (*soft-core*) es (Dixon, 2002):

$$K(r; \gamma, \delta) = \begin{cases} \gamma \pi r^2 & \text{si } 0 < r \leq \delta \\ \pi r^2 - (1-\gamma) \pi \delta^2 & \text{si } r \geq \delta \end{cases}$$

En el proceso Strauss de núcleo duro (*hard core*),  $\gamma = 0$ , y por lo tanto no existe ningún vecino dentro de la distancia crítica.

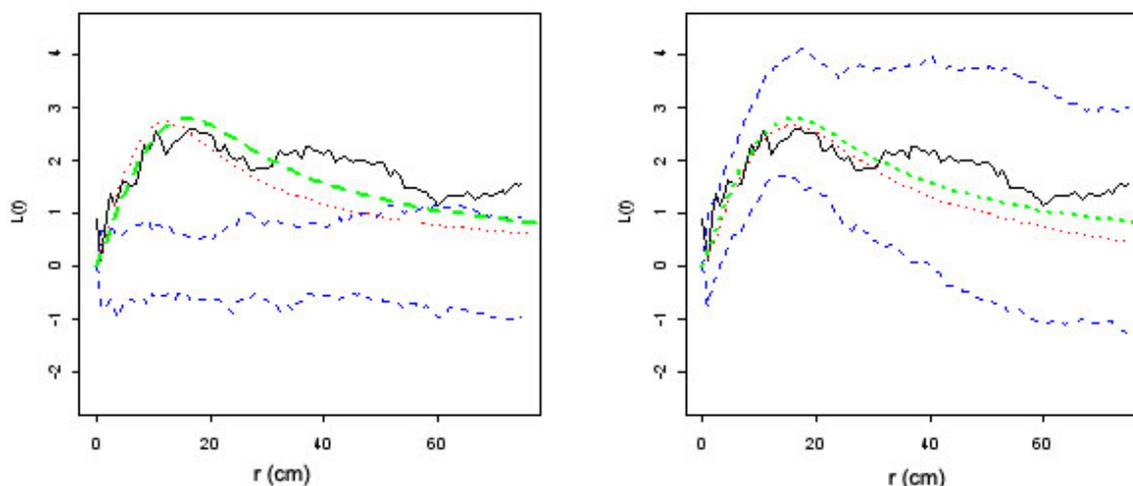
## Ajuste de modelos usando funciones sumario

Tradicionalmente el ajuste de modelos se ha realizado con métodos denominados *ad hoc* (Diggle, 2003) y basados en comparar funciones sumario empíricas y teóricas de aquellos procesos para los cuales se conoce la forma matemática de la función sumario del modelo teórico (no se conoce de todos). La idea consiste en encontrar los parámetros que minimizan las diferencias entre ambas (método del contraste mínimo, Diggle y Graton 1984). Por ejemplo, para ajustar el proceso de Poisson agrupado, Diggle (2003) sugiere minimizar la expresión:

$$D(\sigma, \rho) = \int_0^{r_0} \left\{ \left[ \hat{K}(r) \right]^c - \left[ K(r; \hat{\sigma}, \hat{\rho}) \right]^c \right\}^2 dr$$

donde  $K(r)$  es la función  $K$  del mapa de puntos observado y  $K(r; \sigma, \rho)$  es la función  $K$  teórica del proceso agrupado dados los parámetros  $\sigma$  y  $\rho$ . La constante  $r_0$  representa el límite de integración (es decir, selecciona el límite de las escalas que se ajustarán) y  $c$  es una constante de ajuste necesaria para controlar (estabilizar) la varianza de la función  $K$  empírica y asegurar una convergencia más estable en el proceso de minimización. Para el caso concreto del proceso de Poisson agrupado, Diggle (2003) recomienda un valor de  $c \leq 0.25$  y un límite de integración  $r_0$  menor de  $\frac{1}{4}$  de la longitud del lado en el caso de parcelas cuadradas (aunque a mayor o menor límite el modelo se ajustará más a efectos globales o a pequeña escala).

respectivamente). Los valores de los otros parámetros,  $\sigma$  y  $\rho$  se estiman durante la minimización de  $D(\sigma, \rho)$ . La minimización se realiza de forma análoga a una regresión no lineal, por ejemplo con el algoritmo de Nelder y Mead (1965). Los valores iniciales para la minimización de  $\sigma$  y  $\rho$  pueden obtenerse mediante 'tanteo' (representación, junto a la función K empírica, de varias funciones K teóricas con valores intuitivos de los parámetros y selección de los que produzcan un ajuste visual más aproximado; **Fig. 6**) o mediante una minimización previa de la función  $K(r; \sigma, \rho)$  con valores arbitrarios.



**Figura 6.** Ajuste de un proceso de Poisson agrupado mediante el método *ad hoc* basado en la función K de Ripley. **Izquierda.** En línea negra continua, función  $L(r)$  de la comunidad gipsófila de la figura 4. Curvas azules: envueltas de la simulación de patrones CSR de la misma intensidad. Curva roja punteada, 'tanteo' con la función teórica del proceso agrupado  $\{K(r; \sigma, \rho) = \pi r^2 + \rho^{-1} [1 - \exp(-r^2/4\sigma^2)]\}$  con valores de  $\sigma = 5.0$  y  $\rho = 3.33e-03$ . Curva verde: Función ajustada tras la minimización de la función  $D(\sigma, \rho)$  con el algoritmo de Nelder y Mead (1965). Los parámetros ajustados son  $\sigma = 6.80$  y  $\rho = 2.48e-03$ . **Derecha.** Curvas negra y verde igual que en la figura de la izquierda. Curvas azules: envueltas obtenidas tras la simulación de 99 procesos agrupados de Poisson de la misma intensidad que el original y con los parámetros ajustados en la etapa anterior. Curva roja punteada: valor medio de la función L en 99 las simulaciones.

El ajuste *ad hoc* del proceso de Strauss o de cualquier otro del que se conozca explícitamente la fórmula matemática de la función sumario que lo describe (función K o la que sea) se realiza de forma similar. La única diferencia apreciable estriba en el valor de la constante de ajuste  $c$  que mientras para otros procesos de naturaleza agregada debe seguir siendo de  $c \leq 0.25$ , para procesos de naturaleza similar a CSR o regular se establece en  $c = 0.5$  (Diggle, 2003).

A partir del modelo ajustado se simulan patrones de puntos que sirven para construir las *envueltas* igual que en el caso del test de CSR. Si la función K observada queda dentro de la banda definida por las envueltas se acepta que el patrón pueda ser descrito por el modelo ajustado (**Fig. 6**). Aunque ésta ha sido la forma tradicional del ajuste *ad hoc* en ecología, desde un punto de vista estadístico lo ideal sería emplear una función distinta a la función K (por ejemplo la G, H, J, etc) para realizar el test de Monte Carlo (Diggle, 2003).

A pesar del desarrollo reciente de métodos formales basados en el análisis de la verosimilitud, los métodos *ad hoc* siguen siendo útiles tanto por su utilidad para una rápida exploración de un rango de modelos como por el método de evaluación visual directa del ajuste del modelo que proporcionan. De hecho son los que siguen empleándose mayoritariamente en los análisis ecológicos de patrones de puntos. Interesantes ejemplos recientes son la modelización del patrón de establecimiento de árboles en zonas aclaradas del bosque tropical por Batista y Maguire (1998) o del patrón de reclutamiento de plántulas de *Pinus uncinata* en el Sistema Ibérico por Camarero *et al.*, (2005).

## Ajuste de modelos con funciones de verosimilitud y pseudo-verosimilitud

Uno de los inconvenientes de los ajustes *ad hoc* es la necesidad de conocer la forma matemática de la función sumario que se empleará en ajustar el modelo. Como se ha comentado anteriormente, no todos los procesos tienen funciones sumario teóricas conocidas. Por otra parte, la mayoría de las funciones sumario asumen que el patrón de puntos es homogéneo (es decir, de intensidad constante), cuando esto sólo suele ocurrir en escalas muy pequeñas y la existencia de gradientes ambientales condiciona la aparición de patrones *inhomogéneos* (o no estacionarios, con intensidad variable en función de la



localidad). Aunque se han propuesto algunos métodos para intentar solucionar este problema, por ejemplo, delimitando áreas de intensidad más o menos 'constante', (Pelissier y Goreaud, 2001), tales aproximaciones sólo son válidas en casos de patrones heterogéneos muy sencillos en los que se pueden delimitar objetivamente subparcelas homogéneas. No es el caso de los patrones inhomogéneos ligados a gradientes. En estos casos se ha propuesto dividir el patrón en pequeñas subparcelas de un tamaño tal que la intensidad dentro de cada una pueda ser considerada constante, realizar un test local de CSR en cada subparcela y posteriormente realizar un test global tomando como hipótesis nula un proceso de Poisson inhomogéneo (Couteron *et al.*, 2003).

Uno de los avances recientes en la modelización de patrones de puntos consiste en el empleo de métodos de inferencia basada en la verosimilitud para ajustar modelos paramétricos flexibles, que permiten liberarse de las restricciones de la homogeneidad (pueden incluir superficies *-trends-* espaciales) e incluso introducir dependencia de covariables, interacciones entre puntos y dependencia de marcas (etiquetas asociadas a cada punto). Cada modelo se especifica en función de su *intensidad condicional*. La intensidad condicional es una función  $\lambda(u, \mathbf{x})$  de la localidad espacial  $u$  y del patrón de puntos completo  $\mathbf{x}$ . De esta forma, si consideramos una región infinitesimal de área  $du$  alrededor de un punto  $u$ , la probabilidad de el proceso de puntos contenga un punto en esa región infinitesimal, dada la posición del resto de puntos fuera de esa región, es  $\lambda(u, \mathbf{x}) du$  (Baddeley y Turner, 2006). Por ejemplo, el proceso de Poisson homogéneo tiene intensidad condicional

$$\lambda(u, \mathbf{x}) = \beta$$

donde  $\beta$ , una constante, es la intensidad del proceso. En el proceso de Poisson inhomogéneo la intensidad no es constante sino una función de la localidad,  $\beta(u)$ , por lo que la intensidad condicional es

$$\lambda(u, \mathbf{x}) = \beta(u)$$

El proceso de Strauss, por el contrario, tiene como intensidad condicional

$$\lambda(u, \mathbf{x}) = \beta \gamma^{t(u, \mathbf{x})}$$

donde  $t(u, \mathbf{x})$  es el número de puntos del patrón  $\mathbf{x}$  que quedan dentro de una distancia  $\delta$  de la localidad  $u$ . Aquí  $\gamma$  es el parámetro de la interacción y  $\delta$  la distancia crítica del proceso de Strauss.

Expresados de esta forma, los modelos pueden ajustarse con cierta dificultad con el método de máxima verosimilitud o más fácilmente con el método de máxima pseudo-verosimilitud (Baddeley y Turner, 2000). Esta técnica permite ajustar cualquier modelo para el cual la intensidad condicional tenga la forma loglineal

$$\lambda(u, \mathbf{x}) = \exp [\psi^T B(u) + \varphi^T C(u, \mathbf{x})]$$

donde  $\psi$  y  $\varphi$  son los parámetros que hay que estimar. El término  $B(u)$  representa la 'tendencia espacial' (*trend*) o el efecto de covariables espaciales, mientras que el término  $C(u, \mathbf{x})$  representa las 'interacciones estocásticas' del proceso de puntos. Este término, por ejemplo, no aparece si el modelo es de un proceso de Poisson. En algunos casos puede que haya que realizar una re-parametrización del modelo para adaptarlo a la forma loglineal. Por ejemplo, la intensidad condicional del proceso de Strauss adopta la forma loglineal si se hace  $B(u) \equiv 1$  y  $C(u, \mathbf{x}) = t(u, \mathbf{x})$ , y los parámetros se toman como  $\psi = \log \beta$  y  $\varphi = \log \gamma$  (Baddeley y Turner, 2006).

A la hora de ajustar el modelo, la función  $B(u)$  se trata como la componente 'sistemática' del modelo, mientras que la interacción entre puntos [ $C(u, \mathbf{x})$ ] define la 'familia' de distribuciones del modelo, de forma análoga a las familias de los modelos lineales generalizados (*glm*), lo que permite emplear software estadístico estándar en esta tarea (Baddeley y Turner, 2000). Sin embargo este método estima sólo los parámetros canónicos de modelo ( $\psi$  y  $\varphi$ ), es decir, aquellos para los que la intensidad condicional es loglineal. Los llamados 'parámetros irregulares', como el radio  $\delta$  de la interacción en el proceso de Strauss o cualquier otro, deben especificarse a priori; su estimación puede realizarse construyendo perfiles de pseudo-verosimilitud (ver un ejemplo en la última sección del artículo).

Una de las numerosas ventajas del ajuste de modelos mediante la pseudo-verosimilitud es que se pueden hacer predicciones y generar mapas de la intensidad condicional ajustada. Más importante, sin embargo, es la posibilidad de examinar los residuos del modelo para diagnosticar su idoneidad (Baddeley *et al.*, 2005; Moller y Waagepetersen, 2006), e incluso realizar inferencia formal. Para procesos de Poisson (homogéneos o inhomogéneos), la máxima pseudo-verosimilitud es equivalente a

la máxima verosimilitud, por lo que diferentes modelos ajustados al mismo patrón de puntos se pueden comparar mediante el test de cociente de verosimilitud (*likelihood ratio test*, ver ejemplo en la última sección del artículo).

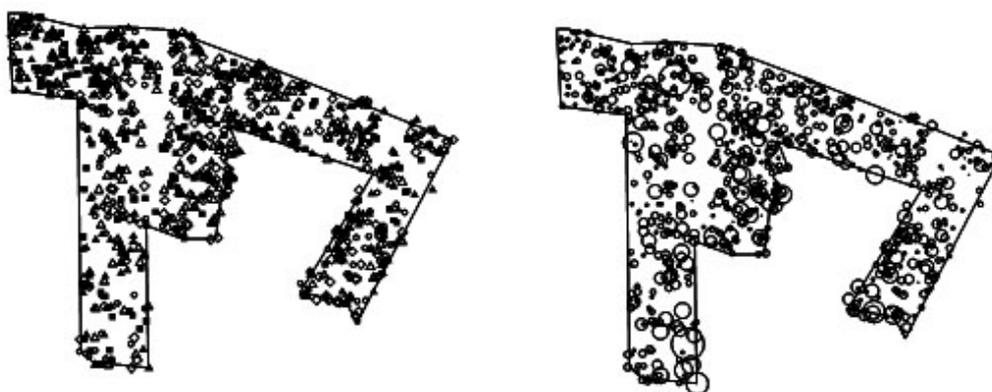
Para otros procesos, Baddeley y Turner (2006) proponen un estadístico llamado  $\Delta$  (delta), análogo a la desviación en la teoría de la verosimilitud y consistente en el doble del logaritmo del cociente de pseudo-verosimilitud (PL):

$$\Delta = \Delta(x) = 2 \left( \log PL(\hat{\theta}_1(x); x) - \log PL(\hat{\theta}_0(x); x) \right)$$

cuya significatividad estadística puede ser estimada mediante un test de Monte Carlo basado en la simulación del proceso ajustado con algoritmos como el de Metropolis-Hastings (Moller y Waagepetersen 2003, 2006).

## El caso de los puntos diversos

Con frecuencia los elementos representados por el patrón de puntos no son idénticos sino que representan diferentes clases tales como especies, sexos, clases de edad, etc. En la terminología del análisis de patrones de puntos, a esta información adicional asociada a las coordenadas se la denomina 'marca' y al patrón que la posee 'marcado'. Las marcas pueden ser de tipo discreto, como las mencionadas anteriormente o continuas (la más común el tamaño, *dbh*, etc., **Fig. 7**). Cuando los patrones de puntos son marcados, el análisis puede extenderse más allá de la caracterización del patrón global o de los patrones de cada tipo de marca para intentar responder a la pregunta de si existe dependencia entre los diferentes tipos (marcas) del patrón.



**Figura 7.** Patrones marcados. **Izquierda.** Patrón multivariado de una parcela experimental de la Estación Biológica Chamusquín (U.T.P.L, Zamora Chinchipe, Ecuador). Cada punto representa un árbol y cada símbolo un gremio forestal diferente. **Derecha.** Patrón con marcas continuas. El diámetro de cada círculo es proporcional al *dbh* de cada árbol.

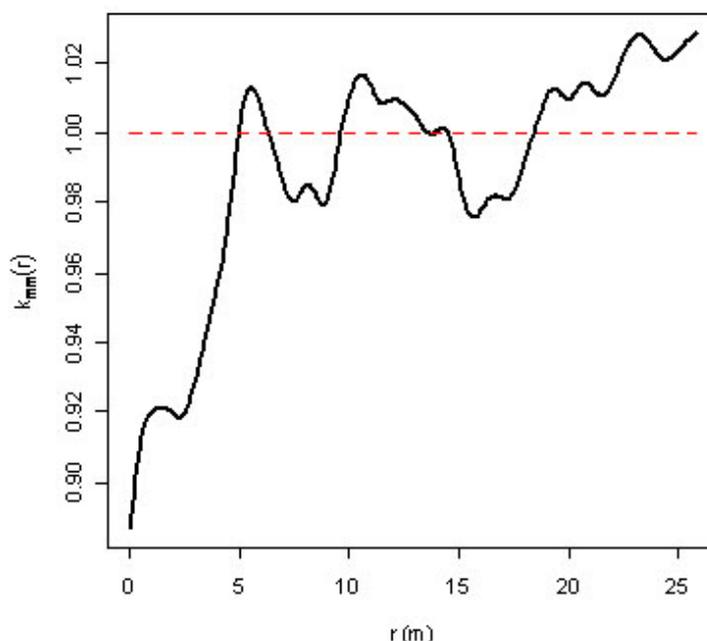
### Funciones de correlación de marcas

Cuando las marcas representan alguna variable continua (por ejemplo, los diámetros a la altura del pecho *dbh* o las áreas basales en una formación forestal) se emplean funciones que estimen la correlación entre marcas. Si  $m(0)$  es el valor de la marca en un punto 'origen' y  $m(r)$  el valor de la marca en cualquier punto situado a la distancia  $r$ , su relación se cuantifica por  $f(m(0), m(r))$ , donde  $f$  es una función 'apropiada' para testar dicha relación. Dos de las más importantes funciones son (Stoyan y Penttinen 2000):

$$f_1(m_1, m_2) = m_1 m_2$$

$$f_2(m_1, m_2) = \frac{1}{2} (m_1 - m_2)^2$$

La **función de correlación de marca** (*mark correlation function*, Stoyan & Stoyan 1994)  $k_{mm}(r)$  se obtiene dividiendo el valor medio de  $f_1(m_1, m_2)$  entre el valor medio de la marca ( $\mu^2$ ) para normalizarlo. En el caso de marcas independientes,  $k_{mm}(r) = 1$  para todas las distancias  $r$ ; valores  $< 1$  indican que la 'correlación' es menor que la media a esas distancias y valores  $> 1$  lo contrario (**Fig. 8**). Por ejemplo, en el caso de que la marca sea el *dbh*, lo normal es encontrar valores de  $k_{mm}(r) < 1$  para  $r$  pequeñas y valores oscilando alrededor de 1 para  $r$  grandes (Stoyan y Penttinen, 2000).



**Figura 8.** Función de correlación de marca para el diámetro a la altura del pecho (dbh) en la parcela experimental de Chamusquín. El valor teórico esperado en el caso de independencia entre las marcas es  $k_{mm}(r) = 1$ . Como suele ser habitual para esta variable,  $k_{mm}(r) < 1$  para  $r$  pequeñas, indicando la existencia de repulsión entre individuos de parecido diámetro a distancias cortas.

La función de **variograma de marca** (*mark variogram*, Cressie 1993)  $g(r)$  se obtiene del valor medio de  $f_2(m_1, m_2)$  para cada distancia  $r$ . Su nombre se debe al paralelismo formal con los variogramas geoestadísticos y describe las diferencias de las marcas de los puntos separados por una distancia  $r$ . En el caso de comunidades de plantas, para valores pequeños de  $r$  da información sobre las interacciones entre los individuos, mientras que para  $r$  mayores da información sobre la influencia de factores ambientales como el suelo, la humedad o la altitud (Stoyan y Penttinen, 2000).

Tanto la función de correlación de marca como la de variograma de marca son funciones de densidad locales, no acumulativas, igual que la función de correlación de par o la *O-ring*.

#### Funciones para patrones multivariados.

La herramienta más comúnmente utilizada para el análisis de patrones con marcas discretas es la función K-cruzada (*K-cross*, en ocasiones denominada también *K* bivariada o *intertipo*). La función  $\lambda_j K_{ij}(r)$ , donde  $K_{ij}(r)$  es la función K-cruzada y  $\lambda_j$  es la intensidad del patrón de tipo ' $j$ ', proporciona el número medio de puntos de tipo ' $j$ ' dentro de un radio  $r$  alrededor de cualquier punto de tipo ' $i$ '. Los estimadores de  $K_{ij}(r)$ , así como los mecanismos para corregir el efecto borde son muy similares a los de  $K(r)$  (ver por ejemplo en Diggle 2003, Dixon 2002).

En un patrón con  $n$  tipos diferentes de marcas se podrían calcular hasta  $n^2$  funciones  $K_{ij}$  diferentes. Desde un punto de vista práctico, sin embargo hay que distinguir las auténticas funciones K-cruzadas ( $K_{ij}(r)$  cuando  $i \neq j$ ) de las funciones K univariadas [ $K_{ii}(r)$ ].

Como los análisis de dependencia entre patrones se realizan siempre entre pares de marcas, independientemente del número diferente de éstas, es frecuente denominar a la función K-cruzada como K bivariada o  $K_{12}(r)$ . En el caso de patrones infinitos, sin límites,  $K_{12}(r) = K_{21}(r)$ ; sin embargo en patrones dentro de un área limitada el efecto borde afecta de forma no simétrica a

las dos funciones por lo que lo habitual suele ser emplear la función bivariada de Lotwick y Silverman (1982)  $K_{12}^*(r)$ , un estimador ponderado de los dos anteriores.

Otras funciones, como por ejemplo la J de Van Lieshout y Baddeley tienen también su versión J-cruzada. Su uso e interpretación es similar al de la función K-cruzada (Van Lieshout y Baddeley 1999). Del mismo modo, también existe una versión cruzada de la función de correlación de par,  $g_{ij}(r)$ .

Una comparación aplicada entre la función de correlación de par y la función K-bivariada sobre los mismos patrones puede verse por ejemplo en Pelissier (1998) o en Schurr *et al.*, (2004).

#### Análisis de la relación entre dos patrones de puntos.

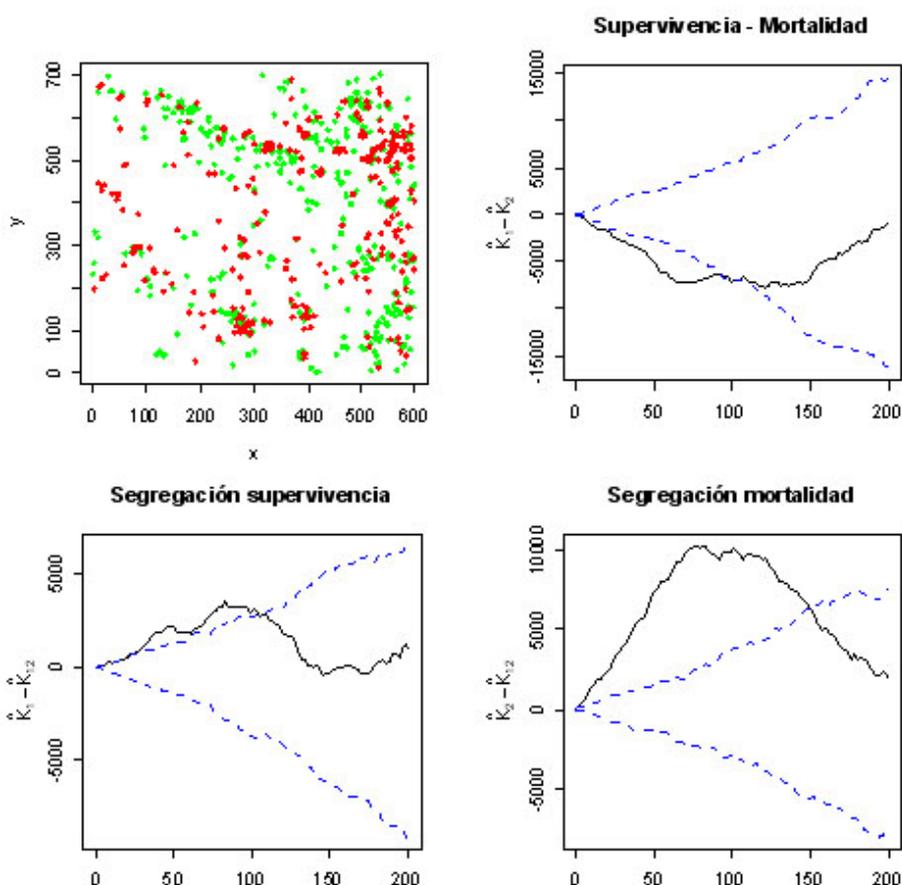
La relación entre dos procesos espaciales puede abordarse desde dos perspectivas diferentes (Dixon 2002, Goreaud y Pelissier, 2003). La perspectiva o hipótesis de independencia (Lotwick y Silverman, 1982) se plantea cuestiones relativas a la interacción entre los dos procesos. La perspectiva o hipótesis del etiquetado aleatorio (*random labelling*, planteada por Cuzick y Edwards, 1990) se plantea cuestiones relativas al proceso que asigna las etiquetas o marcas a los puntos de un patrón global. Cada hipótesis tiene aplicación en problemas ecológicos distintos.

Bajo la hipótesis de independencia (la hipótesis nula es que los dos procesos son independientes), el valor teórico esperado para  $K_{12}^*(r) = \pi r^2$ , independientemente de cómo sea la estructura espacial del patrón que porta cada tipo de marca. Por lo tanto, al igual que en el caso de la función K univariada, se puede y se suele trabajar con la función  $L_{12}^*(r) = (K_{12}^*(r)/\pi)^{1/2}$ . En el caso de que se cumpla la hipótesis nula,  $L_{12}^*(r) = r$ , por lo que se puede testar si  $L_{12}^*(r) - r = 0$  a cada distancia  $r$ . Valores de  $L_{12}^*(r) - r > 0$  indican atracción entre los dos procesos a la distancia  $r$ , valores  $< 0$  indican repulsión (ver ejemplo en la última sección del artículo). Al igual que en el caso de la función univariada, los valores críticos de  $L_{12}^*(r) - r$  se calculan también con simulación Monte-Carlo, aunque aquí es más complicado que en el caso univariado ya que las simulaciones deben mantener el patrón espacial de cada proceso individual a la vez que rompen la posible dependencia que exista entre ellos. En el caso de que los procesos individuales puedan ser descritos por modelos paramétricos, la simulación de estos permite fácilmente estimar los valores críticos del test (Dixon, 2002). Cuando la forma del área de estudio es rectangular, el método no paramétrico de desplazamiento toroidal (*toroidal shift*) es una alternativa razonable (y de hecho se usa con gran frecuencia). El método consiste en mantener constantes las coordenadas de uno de los patrones y desplazar una idéntica pequeña distancia aleatoria en las direcciones x e y todos los puntos del otro patrón. El área de estudio se trata matemáticamente como un toro (con los bordes superior e inferior conectados y los bordes izquierdo y derecho conectados) de tal forma que los puntos que con el desplazamiento 'salen' del área por un lado reingresan por el lado contrario. El desplazamiento aleatorio y el cálculo de  $L_{12}^*(r)$  se repite un número elevado de veces para obtener los valores críticos del test. He y Duncan (2000) usan este método para, entre otras cosas, analizar la asociación entre pares de especies en un bosque de abeto de Douglas y averiguar el efecto de la mortalidad sobre dicho patrón.

La hipótesis del etiquetado aleatorio se debe emplear cuando el patrón multivariado se puede considerar como resultado de un proceso jerárquico en el que un primer 'subproceso' genera el patrón de puntos y un subproceso posterior asigna marcas a los puntos (por ejemplo, sexo en especies dioicas, presencia o ausencia de epífitos en árboles, enfermedad o no en los individuos, mortalidad y supervivencia, etc.). Bajo la hipótesis de *random labelling* cada patrón individual sería una muestra aleatoria del patrón total y por lo tanto, debido a la naturaleza de la función K,  $K_{12}(r) = K_{21}(r) = K_{11}(r) = K_{22}(r) = K(r)$ , es decir, todas las funciones cruzadas serían iguales a la función K univariada del patrón completo (Dixon, 2002). Las desviaciones de la hipótesis nula de etiquetado aleatorio se evalúan mediante diferencias entre pares de funciones K.

$K_{11}(r) - K_{22}(r)$  evalúa si un patrón está más o menos agrupado que el otro (y a qué escala, **Fig. 9**).

$K_{11}(r) - K_{12}(r)$  y  $K_{22}(r) - K_{12}(r)$  evalúan la segregación de los procesos, es decir, evalúan si un tipo de punto tiende a estar rodeado por otros puntos del mismo tipo (**Fig. 9**). La inferencia se basa habitualmente en simulación Monte Carlo mediante la permutación aleatoria de las marcas sobre las coordenadas del patrón completo.

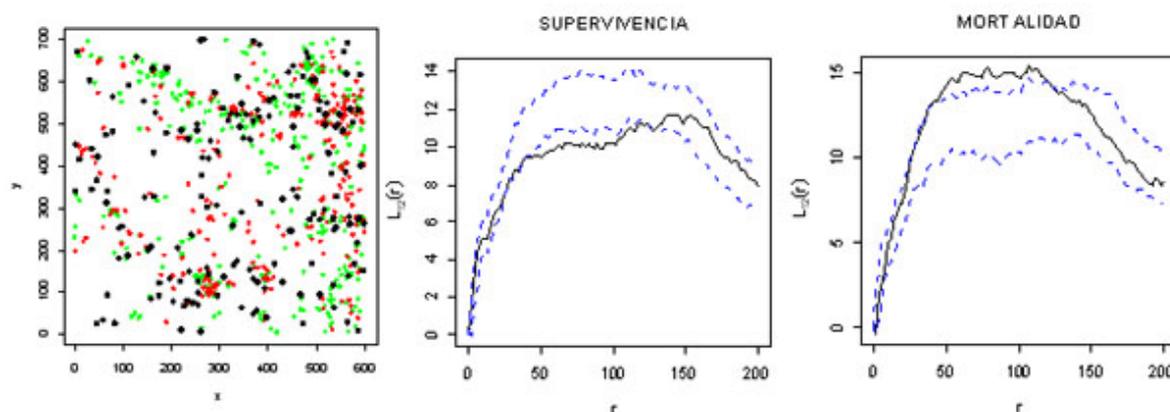


**Figura 9.** Análisis de la dependencia en patrones bivariados con la hipótesis de etiquetado aleatorio (*random labelling*). **Cuadro superior izquierdo.** Patrón bivariado de supervivencia (verde) y mortalidad (rojo) de plántulas de *Helianthemum squamatum*. Ya que este patrón de puntos se puede describir como el resultado de dos procesos secuenciales (un primer subproceso establece el patrón de plántulas; un subproceso posterior distribuye las etiquetas 'sobrevive' y 'muere' sobre el patrón anterior), la hipótesis nula adecuada para analizar la dependencia espacial entre las marcas es la del etiquetado aleatorio. La relación entre los patrones se analiza mediante la diferencia de las funciones univariadas (**cuadro superior derecho**), con lo que se puede testar si los patrones presentan diferencias en la intensidad de agrupamiento y mediante diferencias entre la función univariada de cada proceso y la bivariada (**cuadros inferiores**), lo que permite testar si existe segregación (es decir, si los puntos de cada patrón se rodean de puntos del mismo tipo con mayor frecuencia de lo esperado por azar). Las envueltas se generan en todos los casos mediante permutaciones aleatorias de las marcas sobre el patrón global de plántulas. La diferencia entre el patrón de supervivencia y mortalidad muestra que éste es más agrupado que el primero (diferencia negativa). Los test de segregación ponen de manifiesto que tanto los casos de supervivencia como los de mortalidad están segregados (están más rodeados de casos del mismo tipo), especialmente la mortalidad, lo que confirma la separación espacial de ambos fenómenos.

En alguna ocasión, en vez de emplear las diferencias entre funciones K, se ha empleado la función univariada  $L(r)$ , para analizar la estructura de cada uno de los procesos que componen el patrón bivariado por separado. Suzuki et al. (2003) analizan los patrones espaciales de mortalidad y supervivencia en una especie herbácea calculando la función L de cada uno de las marcas y comparándola con las envueltas obtenidas por permutación aleatoria de la marca sobre el patrón de puntos. Valores de  $L(r)$  mayores o menores que los límites indicados por las envueltas indican respectivamente agregación o repulsión del suceso considerado en cada caso (mortalidad o supervivencia).

En ocasiones puede ser necesario combinar ambas hipótesis. Por ejemplo, De la Cruz et al., (en preparación) evalúan la relación entre el patrón espacial de la mortalidad de plántulas de *Helianthemum squamatum* y las plantas adultas de una

comunidad gipsófila. La hipótesis nula para la relación entre las plantas adultas y las plántulas muertas debería ser la de independencia, pero dado que la hipótesis nula de la mortalidad se puede considerar como un proceso de etiquetado aleatorio sobre la población de plántulas, en vez de realizar *toroidal shift* de adultos y plántulas muertas, la simulación para la inferencia se basa en aleatorizar las marcas (mortalidad, supervivencia) sobre el patrón de plántulas (**Fig. 10**).

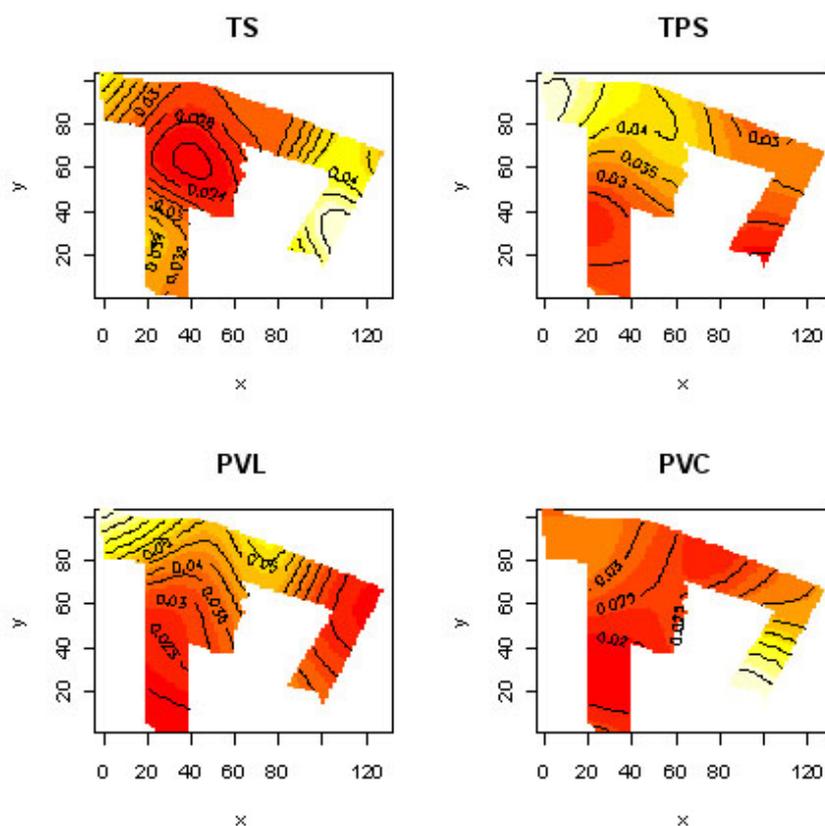


**Figura 10.** Combinación de hipótesis en el análisis de un patrón marcado. **Izquierda.** Parte de un patrón multivariado de una comunidad gipsófila. Puntos negros grandes: individuos adultos de *Helianthemum squamatum*; puntos verdes pequeños: plántulas de *H. squamatum* que sobreviven al censo post-estival; puntos rojos: plántulas que mueren. **Centro.** Función  $L$  cruzada (bivariada) para testar la asociación entre los casos de supervivencia y los individuos adultos de *H. squamatum*. **Izquierda.** Idem con los casos de mortalidad. Como los procesos que generan el patrón de adultos y los patrones de supervivencia y mortalidad de plántulas son independientes, la hipótesis apropiada para testar la asociación entre ambos es la de independencia. Sin embargo, dado que la supervivencia y mortalidad son generados por un subproceso aleatorio sobre el patrón total de plántulas, las simulaciones para generar las envueltas se basan en el etiquetado aleatorio (mortalidad, supervivencia) de dicho patrón. El patrón de supervivencia muestra repulsión con el de adultos a escalas de entre 40 y 120 cm mientras que la mortalidad está asociada a los adultos aproximadamente a la misma escala, lo que parece indicar que el territorio ya ocupado por los adultos es el más apropiado para el establecimiento de las plántulas (sin que ello suponga un incremento en la supervivencia) mientras que a partir de 40 cm de distancia de los adultos se encuentran ambientes en los que la supervivencia se ve comprometida.

Otros procesos marcados pueden modelizarse a partir de modelos empíricos. Por ejemplo, Batista y Maguire (1998) para analizar el patrón espacial de autoaclareo en bosques tropicales ajustan un modelo generalizado binomial (regresión logística) a los datos de mortalidad de árboles, empleando como predictores variables de tamaño, crecimiento y vecindad. El modelo ajustado se emplea después como 'algoritmo' para simular el proceso a partir del mapa de la cubierta forestal original.

### Un ejemplo actual: análisis de la dependencia entre marcas en un patrón multivariado complejo

El objetivo es testar la dependencia espacial entre especies forestales de los distintos gremios forestales tropicales que crecen en una parcela experimental de la Estación Biológica Chamusquín (Universidad Técnica Particular de Loja, Zamora-Chinchipe, Ecuador, **Fig. 7**): TPS: tolerantes parciales a la sombra; TS: tolerantes a la sombra; PVC: pioneras de vida corta; PVL: pioneras de vida larga. Los gremios se definen fundamentalmente en relación con la tolerancia a la luz y el tipo de crecimiento de las diferentes especies, por lo que es previsible encontrar patrones de asociación y repulsión entre los distintos gremios según su mayor o menor afinidad. El análisis de las relaciones entre las marcas de la parcela de Chamusquín se ve complicado por dos motivos: cada gremio parece seguir una tendencia espacial distinta, en ningún caso homogénea (**Fig. 11**) y la forma de la parcela, no rectangular, lo que impide emplear el método de *toroidal shift*, que sería el más fácil de aplicar para testar la hipótesis nula de independencia.



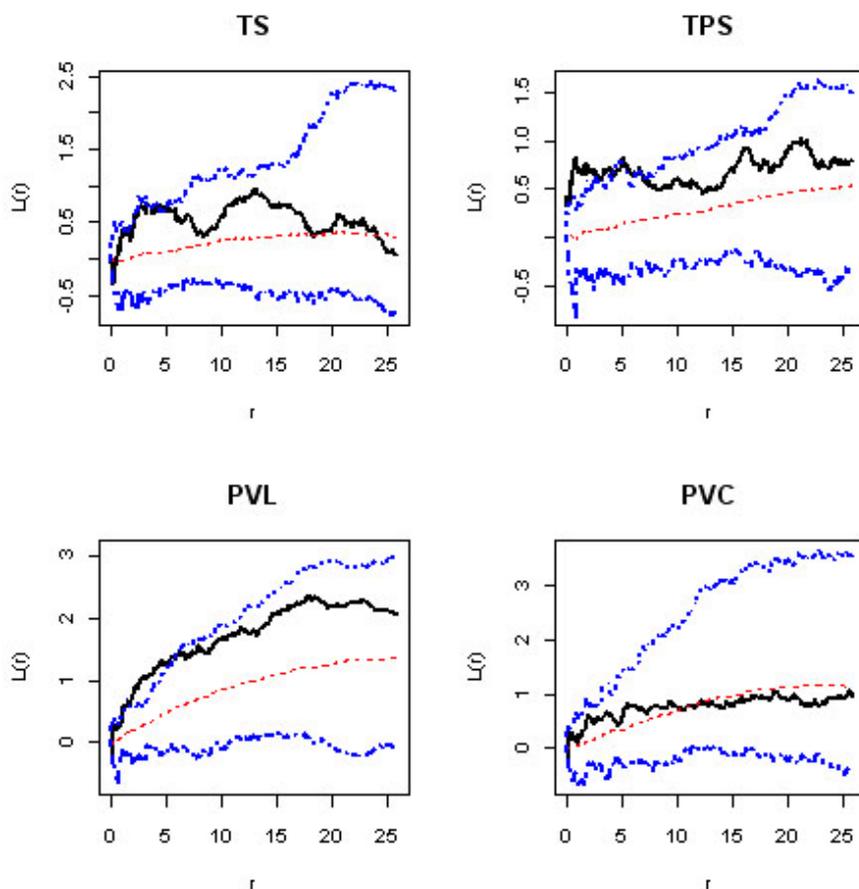
**Figura 11.** Intensidad de los patrones de los diferentes gremios forestales en la parcela de Chamusquín (estimada mediante un *kernel* gaussiano isotrópico). TPS: tolerantes parciales a la sombra; TS: tolerantes a la sombra; PVC: pioneras de vida corta; PVL: pioneras de vida larga. Se aprecia una falta de homogeneidad en todos los patrones y gradientes de variación distintos para cada uno de ellos.

Como alternativa, en primer lugar se ajustan mediante máxima pseudo-verosimilitud patrones de Poisson homogéneos e inhomogéneos con superficies de tendencia polinómicas (de primer grado y de segundo grado en  $x$  e  $y$ ) a cada patrón individual. El análisis del cambio de desviación de los modelos (al ser procesos de Poisson la máxima pseudo-verosimilitud es igual a la máxima verosimilitud y se puede emplear el cociente de verosimilitud), confirma que ninguno de los gremios tiene intensidad constante y que excepto TPS, que puede describirse mediante una superficie sencilla, los demás requieren una superficie polinómica de orden 2 (**Tabla 1**). La simulación de los modelos ajustados (**Fig. 12**) muestra que dos de ellos (PVC y TS) quedan razonablemente bien descritos con la superficie ajustada mientras que los otros dos (PVL y TPS) requieren la inclusión en el modelo de un componente de interacción que dé cuenta del 'agrupamiento' que se pone de manifiesto en las gráficas.



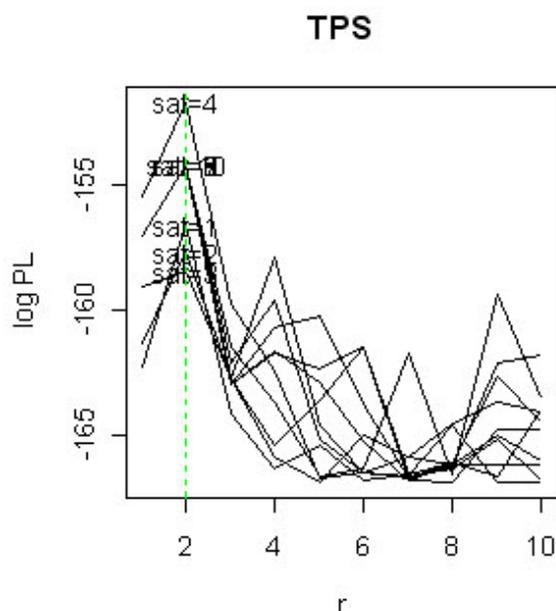
**Tabla 1.** Análisis secuencial de la desviación.

Gremio	Modelo	G.I. Resid.	Desv. Resid.	G.I.	$\Delta$ Desviación	P(> Chi )
<b>PVC</b>	homogéneo	1282	893.61			
	x + y	1280	887.66	2	5.95	0.05
	(x + y) <sup>2</sup>	1277	873.61	3	14.05	0.003
<b>PVL</b>	homogéneo	1328	1054.68			
	x + y	1326	1026.23	2	28.45	6.634e-07
	(x + y) <sup>2</sup>	1323	1012.61	3	13.63	3.462e-03
<b>TS</b>	homogéneo	1287	907.33			
	x + y	1285	905.11	2	2.22	0.33
	(x + y) <sup>2</sup>	1282	898.64	3	6.47	0.09
<b>TPS</b>	homogéneo	1301	984.35			
	x + y	1299	975.18	2	9.18	0.01
	(x + y) <sup>2</sup>	1296	975.03	3	0.15	0.99

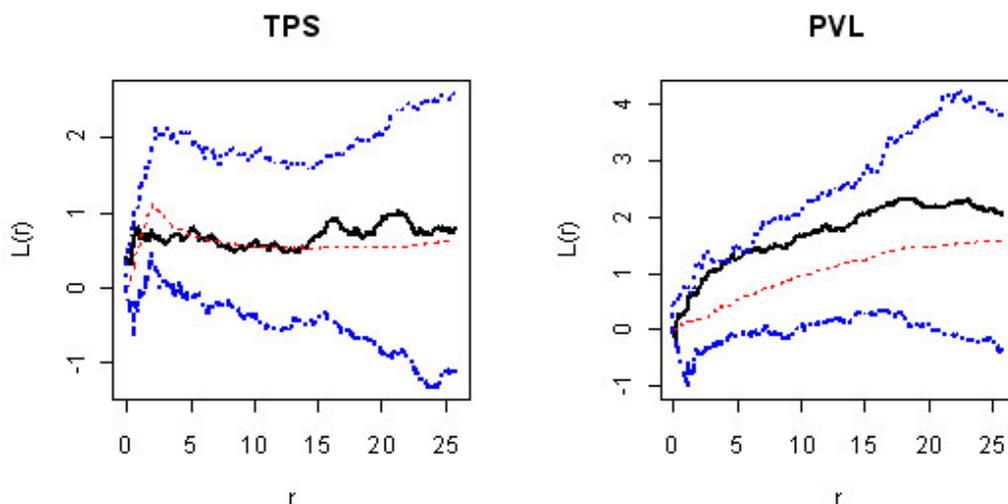


**Figura 12.** Función  $L$  (curva continua negra) de los gremios forestales de la parcela de Chamusquín. En cada caso, las líneas discontinuas azules representan las envueltas obtenidas tras la simulación de 99 modelos Poisson inhomogéneos ajustados a cada patrón, con polinomio de segundo grado en  $x$  e  $y$  y como componente de tendencia espacial (excepto en PVL, de grado 1). La línea roja discontinua representa el valor medio de todos los obtenidos en las simulaciones y puede emplearse como referencia de el valor esperado de cada modelo ajustado. Se observa cómo las funciones de los gremios TS y PVC quedan dentro del espacio delimitado por sus envueltas, por lo que se puede considerar que el modelo de Poisson inhomogéneo los describe razonablemente bien. En el caso de TPS y PVL, la función empírica sobresale por encima de las envueltas superiores, lo que sugiere la necesidad de incorporar en sus modelos una componente de interacción que dé cuenta del agrupamiento que se pone de manifiesto en las curvas.

De entre los escasos procesos que puedan modelizar agrupamiento y que actualmente estén disponibles para ser ajustados mediante máxima pseudo-verosimilitud, empleamos el proceso de saturación de Geyer (1999). Se trata de una modificación del proceso de Strauss en el que cada punto contribuye a la densidad de probabilidad del patrón en una cantidad  $\beta \gamma^{\min[s, t(x_i, X)]}$ , donde  $t(x_i, X)$  representa el número de 'vecinos íntimos' de  $x_i$  en el patrón  $X$  (vecinos dentro de un radio  $r$  establecido) y  $s$  el umbral de saturación del proceso. Si  $s$  es infinito, el proceso se convierte en uno de Strauss con parámetro de interacción  $\gamma^2$  mientras que  $s = 0$  genera el proceso de Poisson. Si  $s$  es un valor positivo cualquiera, valores de  $\gamma < 1$  describen un proceso uniforme y de  $\gamma > 1$  un proceso agregado. Independientemente de la mejor o peor interpretación biológica que se le pueda dar a los parámetros del modelo, su uso en esta ocasión se justifica en la posibilidad de ajustar un modelo paramétrico que nos permita realizar simulaciones de los procesos ajustados. Como hay dos parámetros no 'canónicos' en el modelo (el coeficiente de saturación  $s$  y el radio de interacción  $r$ ) construimos un perfil de pseudo-verosimilitud asignando diferentes combinaciones de valores para  $s$  y  $r$  y observando qué combinación proporciona el mejor ajuste (**Fig. 13**). La simulación de los modelos nuevamente ajustados para PVL y TPS (**Fig. 14**) ahora sí parece describir razonablemente dichos patrones por lo que se podrán emplear en el siguiente paso.

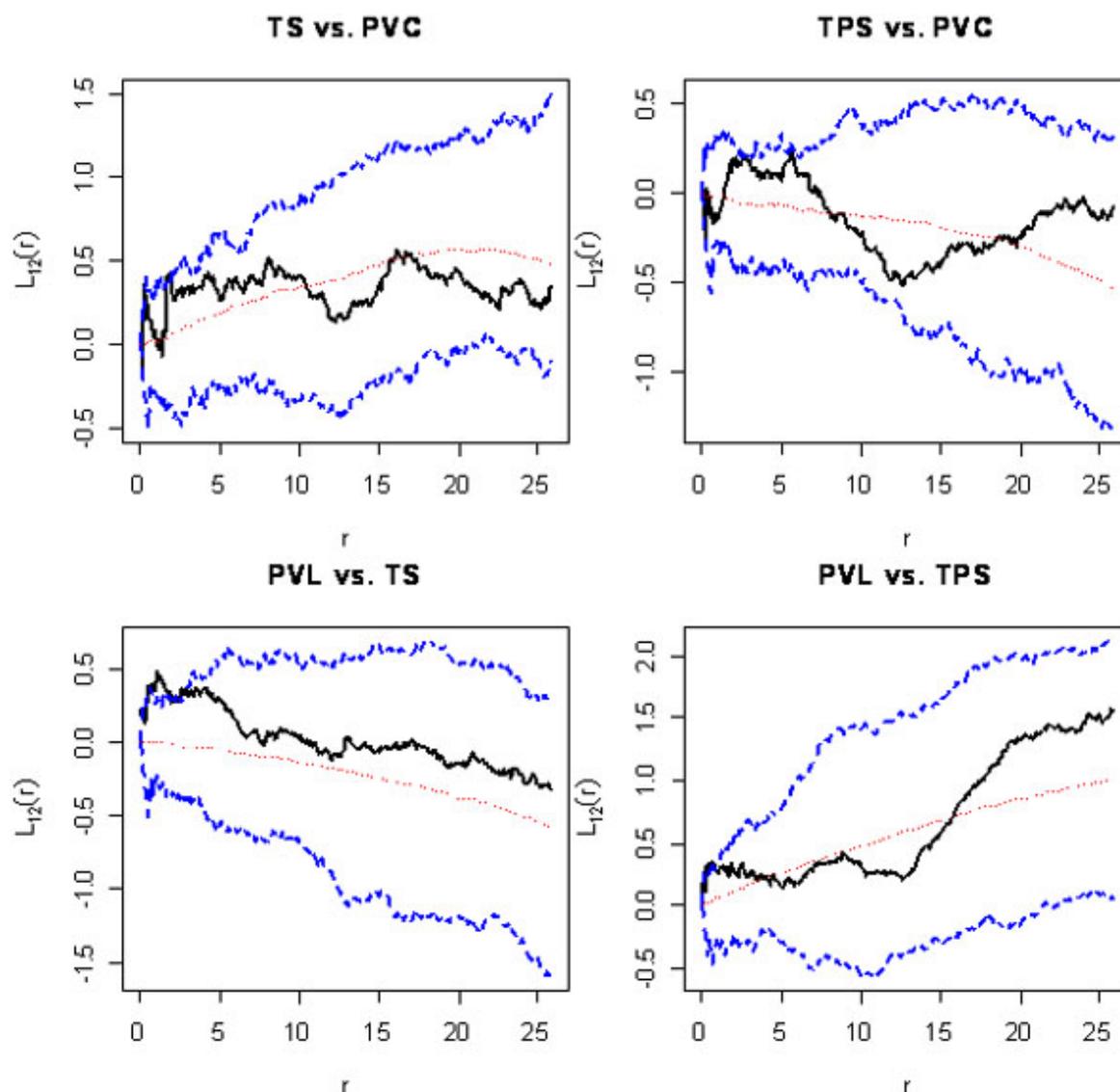


**Figura 13.** Perfil de pseudo-verosimilitud para el ajuste de un proceso de Geyer al patrón del gremio de especies tolerantes parciales a la sombra (TPS). Como el proceso de Geyer tiene dos parámetros irregulares ( $r$ , la distancia que define dos puntos como vecinos íntimos, y  $s$ , la saturación), el perfil se construye ajustando modelos para diferentes combinaciones de valores de  $r$  y  $s$  (en el ejemplo, y a partir de las gráficas de la función  $K$ , se consideró probar valores para ambos parámetros entre 1 y 10). La gráfica representa el logaritmo de la máxima pseudo-verosimilitud alcanzada por cada combinación de parámetros. El valor máximo se obtiene para  $r = 2$  y  $s = 4$ , por lo que se fijaron dichos valores en el modelo.



**Figura 14.** Función  $L$  (línea continua negra) de los gremios TPS y PVL y envueltas (líneas azules) procedentes de 99 simulaciones de los procesos de Geyer inhomogéneos ajustados en cada caso. En esta ocasión la función empírica queda dentro del intervalo definido por las envueltas por lo que se puede considerar que los modelos ajustados describen razonablemente bien cada patrón.

Dado que no se puede emplear la técnica de *toroidal shift*, para construir las envueltas que permitan testar la hipótesis de independencia, la alternativa más razonable es realizar simulaciones de cada uno de los patrones individuales (o dejar uno fijo y simular el otro) mediante los modelos que hemos ajustado y calcular la función  $K$  bivariada en cada caso (**Fig. 15**). El análisis de las funciones  $K$  cruzadas bajo hipótesis de independencia pone de manifiesto, por ejemplo, que la distribución de los pioneros de vida corta (PVC) es independiente de la de los tolerantes a la sombra (TS) y tolerantes parciales a la sombra (TPS). Por el contrario, el gremio de pioneros de vida larga (PVL) presenta atracción a corta distancia con los tolerantes a la sombra y parece casi independiente de los TPS.



**Figura 15.** Análisis de las relaciones entre patrones individuales de un patrón marcado multivariado, siguiendo la hipótesis de independencia. La línea negra continua representa la función  $L$  cruzada (o bivariada). Las líneas azules representan las envueltas obtenidas tras 99 simulaciones en cada caso. Como se trata de una parcela de contorno irregular y los patrones individuales tienen una distribución no estacionaria, no es posible realizar las simulaciones con el método de *toroidal shift*. En su lugar se han simulado 99 realizaciones de los modelos ajustados a cada tipo de patrón. La línea roja punteada representa el valor medio de las funciones simuladas. La distribución de los árboles del gremio de pioneros de vida corta (PVC) es independiente de la de tolerantes a la sombra (TS) y tolerantes parciales a la sombra (TPS). El gremio de pioneros de vida larga (PVL) presenta atracción a corta distancia con los tolerantes a la sombra y es casi independiente de los TPS.

## Dónde saber más y software

Las obras clásicas de referencia sobre análisis espacial de puntos son Ripley (1981), y Diggle (1983), algo obsoletas hoy en día ya que el análisis estadístico de patrones espaciales de puntos es un campo de investigación extraordinariamente dinámico en el que se producen avances y contribuciones prácticamente a diario. Existe una versión actualizada del libro de Diggle (Diggle, 2003) y otras revisiones más actuales, como la de Stoyan y Stoyan (1994). Una de las causas de la dificultad de penetración de los avances en el análisis estadístico de los patrones espaciales de puntos en el campo de la ecología es la inaccesibilidad y/o la escasa inteligibilidad de los trabajos sobre estadística espacial para los ecólogos (Stoyan y Penttinen, 2000) y la disponibilidad del software apropiado. Dixon (2002) es una referencia muy asequible sobre la función  $K$  escrita por un ecólogo mientras que Baddeley y Turner (2006) y Moller y Waagepetersen (2006) presentan una síntesis de las técnicas y herramientas más actuales pensando en los ecólogos y otros posibles usuarios legos en la materia. Respecto al software, se han desarrollado diversos programas autónomos para realizar los análisis más sencillos como *Passage*, *Programita* (Wiegand y Moloney, 2004), módulos de análisis espacial de *ADE4*, etc. Sin embargo el software más potente



hoy en día para el análisis y modelización de patrones espaciales de puntos está apoyado en la plataforma R (R Core Team, 2005; [www.r-project.org](http://www.r-project.org)) y en el sistema S-plus ([www.insighful.com](http://www.insighful.com)): *spatial* (Venables y Ripley 2002), *spatstat* (Baddeley y Turner, 2005; [www.spatstat.org](http://www.spatstat.org)), *splancs* (Rowlingson y Diggle, 1993, 2005) y alguno más. De entre ellos, el paquete *spatstat* es el que actualmente está experimentando una mayor tasa de desarrollo (aparecen nuevas versiones con periodicidad mensual) y el que incorpora herramientas más potentes y actuales. Todos los ejemplos de este artículo se han realizado con *spatstat*.

## Agradecimientos

A los profesores de la Universidad Técnica Particular de Loja (Ecuador) Jaime Black, Soledad Bustos, Karla Tapia, y Rafael Vicuña por compartir los datos de la parcela experimental de la Estación Biológica Chamusquin.

## Referencias

- Baddeley, A. y Turner, R. 2000. Practical maximum pseudolikelihood for spatial point patterns (with discussion). *Australian and Journal of Statistics* 42(3): 283-322.
- Baddeley, A. y Turner, R. 2005. Spatstat: an R package for analyzing spatial point patterns. *Journal of Statistical Software* 12 (6): 1-42.
- Baddeley, A., Turner, R., Moller, J. y Hazelton, M. 2005. Residual analysis for spatial point processes. *Journal of the Royal Statistical Society (series B)* 67: 1-35.
- Baddeley, A. y Turner, R. 2006. Modelling Spatial Point Patterns in R. En *Case Studies in Spatial Point Process Modelling* (eds. A. Baddeley, P. Gregori, J. Mateu, R. Stoica & D. Stoyan), pp. 23-74. Springer. Heidelberg. Alemania.
- Barot, S., Gignoux, J. y Menaut, J. 1999. Demography of a savanna palm tree: predictions from comprehensive spatial pattern analyses. *Ecology* 80: 1987-2005.
- Batista, J.L.F. y Maguire, D.A. 1998. Modelling the spatial structure of tropical forests. *Forest Ecology and Management* 110: 293-314.
- Camarero, J.J., Gutiérrez, E., Fortin, M.-J., y Ribbens, E. 2005. Spatial patterns of tree recruitment in a relict population of *Pinus uncinata*: Forest expansion through stratified diffusion. *Journal of Biogeography* 32: 1979-1992.
- Couteron, P., Seghier, J. y Chadoeuf, J. 2003. A test for spatial relationships between neighbouring plants in plots of heterogeneous plant density. *Journal of Vegetation Science* 14: 163-172.
- Cressie, N. 1993. *Statistics for Spatial Data*. Wiley, New York, .
- Cuzick, J. & Edwards, R. 1990. Spatial clustering for inhomogeneous populations (with discussion). *Journal of the Royal Statistical Society (series B)* 52: 73-104
- Diggle, P.J. 1979. Statistical methods for spatial point patterns in ecology. En *Spatial and Temporal Analysis in Ecology* (eds. Cormack, R.M. y Ord, J.K.), pp. 99-150. International Co-operative Publishing House, Fairland, Maryland, .
- Diggle, P.J. 1983. *Statistical Analysis of Spatial Point Patterns*. Academic Press. London .
- Diggle, P.J. 2003. *Statistical Analysis of Spatial Point Patterns*. 2ª ed. Arnold, London, .
- Diggle, P.J. and Gratton, R.J. 1984. Monte Carlo methods of inference for implicit statistical models. *Journal of the Royal Statistical Society (series B)* 46: 193-212.
- Dixon, P. M. 2002. Ripley's K function. En *The encyclopedia of environmetrics* (eds. El-Shaarawi, A.H. & Piergorsch, W.W.), pp. 1976-1803. John Wiley & Sons Ltd, New York, .
- Geyer, C.J. 1999. Likelihood Inference for Spatial Point Processes. En *Stochastic Geometry: Likelihood and Computation* (eds. Barndorff-Nielsen, O.E., Kendall, W.S. y Van Lieshout, M.N.M.), pp. 79-140. Chapman and Hall / CRC, Monographs on Statistics and Applied Probability, number 80.



- Goreaud, F. y Pelissier, R. 1999. On explicit formulas of edge effect correction for Ripley's K-function. *Journal of Vegetation Science* 10: 433-438.
- Goreaud, F. y Pelissier, R. 2003. Avoiding misinterpretation of biotic interactions with the intertype  $K_{12}$ -function: population independence vs. random labelling hypotheses. *Journal of Vegetation Science* 14: 681-692
- Haase, P. 1995. Spatial pattern analysis in ecology based on Ripley's K-function: introduction and methods of edge correction. *Journal of Vegetation Science* 6: 575-582.
- He, F. y Duncan, R.P. 2000. Density-dependent effects on tree survival in an old-growth Douglas fir-forest. *Journal of Ecology*, 88: 676-688.
- Huang, F. y Ogata, Y. 1999. Improvements of the maximum pseudo-likelihood estimators in various spatial statistical models. *Journal of Computational and Graphical Statistics* 8: 510-530.
- Lancaster, J. y Downes, B. 2004. Spatial point pattern analysis of available and exploited resources. *Ecography* 27: 94-102.
- Moller, J. y Waagepetersen, R. P. 2003. *Statistical Inference and Simulation for Spatial Point Processes*. Chapman and Hall/CRC, Boca Raton .
- Moller, J. & Waagepetersen, R. P. 2006. Modern statistics for spatial point patterns. *Scandinavian Journal of Statistics*. En prensa.
- Nelder, J. A. y Mead, R. 1965. A simplex algorithm for function minimization. *Computer Journal* 7: 308-313.
- Ohser, J. 1983. On estimators for the reduced second moment measure of point processes. *Mathematische Operationsforschung und Statistik, series Statistics* 14: 63- 71.
- Pelissier, R. 1998. Tree spatial patterns in three contrasting plots of as southern Indian tropical moist evergreen forest. *Journal of Tropical Ecology* 14: 1-16.
- Pelissier, R. & Goreaud, F. 2001. A practical approach to the study of spatial structure in simple cases of heterogeneous vegetation. *J. Veg. Sci.* 12, 99-108.
- R Development Core Team (2005). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, .
- Ripley, B.D. 1979. Test of 'randomness' for spatial patterns. *Journal of the Royal Statistical Society (series B)* 41: 368-374.
- Ripley, B.D. 1981. *Spatial Statistics*. Wiley.
- Ripley, B.D. 1988. *Statistical inference for spatial processes*. Cambridge University Press.
- Rowlingson, B. y Diggle, P.J. 1993. Splancs: Spatial Point pattern analysis Code in S-plus. *Computers and Geosciences* 19: 627-655.
- Rowlingson, B. y Diggle, P.J. 2005. *splancs: Spatial and Space-Time Point Pattern Analysis. R package version 2.01-16*. <http://www.r-project.org>, <http://www.maths.lancs.ac.uk/~rowlings/Splancs/>
- Seabloom, E. W., Bjornstad, O. N., Bolker, B. y Reichman, O. J. 2005. The spatial signature of dispersal and competition in successional grasslands. *Ecological Monographs* 75:199-214.
- Schurr, F. M., Bossdorf, O., Milton, S.J. & Schumacher, J. 2004. Spatial pattern formation in semi-arid shrubland: a priori predicted versus observed pattern characteristics. *Plant Ecology* 173: 271-282.
- Stoyan, D. y Stoyan, H. 1994. *Fractals, Random Shapes and Point Fields*. Wiley, Chichester, .



- Stoyan, D. y Penttinen, A. 2000. Recent Applications of Point Process Methods in Forestry Statistics. *Statistical Science* 15 (1): 61-78.
- Suzuki, R.O., Kudoh, H. y Kachi, N. 2003. Spatial and temporal variations in mortality of the biennial plant, *Lysimachia rubida*: effects of intraspecific competition and environmental heterogeneity. *Journal of Ecology* 91: 114-125.
- Van Lieshout, M.N.M. y Baddeley, A.J. 1996. A nonparametric measure of spatial interaction in point patterns. *Statistica Neerlandica* 50: 344-361
- Van Lieshout, M.N.M. y Baddeley, A.J. 1999. Indices of dependence between types in multivariate point patterns. *Scandinavian Journal of Statistics* 26: 511-532.
- Venables, W.N. y Ripley, B. D. 2002. *Modern Applied Statistics with S*. 4ª ed. Springer-Verlag, New York .
- Wiegand, T., Jeltsch, F., Hanski, I. y Grimm, V. 2003. Using pattern-oriented modeling for revealing hidden information: a key for reconciling ecological theory and application. *Oikos* 100: 209-222
- Wiegand, T. y Moloney, K. A. 2004. Rings, circles and null-models for point pattern analysis in ecology. *Oikos* 104: 209-229