

## Estimación de pautas de asociación y patrones de distribución de edificaciones aisladas en espacios rurales mediante SIG y técnicas basadas en procesos puntuales.

F.B. Galacho Jiménez<sup>1</sup> y S. Reyes Corredera<sup>2</sup>.

<sup>1</sup> Departamento de Geografía-Grupo Análisis Geográfico, Universidad de Málaga. Campus de Teatinos, s/n, 29.071 Málaga.

<sup>2</sup> Departamento de Geografía-Grupo Análisis Geográfico, Universidad de Málaga. Campus de Teatinos, s/n, 29.071 Málaga.

[fbgalacho@uma.es](mailto:fbgalacho@uma.es), [sergioreyes@uma.es](mailto:sergioreyes@uma.es).

**RESUMEN:** En este trabajo se propone un método basado en las técnicas de estadística espacial y SIG para identificar la existencia de patrones geográficos de distribución y asociación de la vivienda rural. La aplicación del método propuesto se dirige a la evaluación del modo de implantación, entendiendo que la heterogeneidad espacial derivada de dicha implantación puede condicionar la estructura territorial y la coexistencia de usos, en el momento en que se llegan a formar aglomerados a través de una dinámica de densificación que surge en un proceso no definido sino aleatorio. Con la finalidad expuesta nos basamos en la caracterización de la vivienda rural como un proceso puntual no homogéneo y mediante varios estimadores como la Función  $K$  de Ripley, el análisis de punto caliente ( $G_i^*$  de Getis-Ord) y los mapas de densidades Kernel se analiza este fenómeno de amplia difusión en el medio rural y de complejas implicaciones territoriales.

**Palabras-clave:** patrones espaciales de puntos, función  $K$  de Ripley, estadística espacial, SIG.

### 1. INTRODUCCIÓN

El proceso urbano en los espacios rurales se ha venido caracterizando como un modelo de ocupación difusa por medio de edificaciones aisladas y, concretamente en lo que respecta a la consideración de una gran parte de éstas como vivienda rural recreativa. Es un fenómeno de importantes consecuencias territoriales que ha seguido una dinámica de implantación en gran medida aleatoria, no ordenada en los instrumentos de ordenación urbanística y planificación territorial; además de ser un uso expresamente prohibido en las distintas leyes del suelo. El proceso se ha venido produciendo bajo la aceptación del dejar hacer, estando condicionado por su relación con el desarrollo económico en general y, en particular por el devenir de los sistemas rurales.

Es precisamente en el declive de los sistemas rurales donde podemos encontrar una primera causa del desarrollo del urbanismo difuso en el medio rural. La vivienda rural, ya sea aislada o en relación con el núcleo rural tradicional, ha ido perdiendo su clara orientación de producción y ha pasado a jugar un papel social y económico cada vez más cercano a la vivienda urbana. Así, ha dejado de estar directamente relacionada con las actividades económicas agrícolas para pasar a ser un producto de otra actividad económica, la inmobiliaria. Se produce entonces la progresiva disgregación de la unidad residencial de la unidad de producción y se acentúa cada vez más la diferenciación entre ambas unidades, la vivienda rural cambia su funcionalidad y adquiere matices más propios de la vivienda urbana. En este proceso debe considerarse entonces la vivienda rural dentro del sistema rural bajo una nueva perspectiva que obliga a la redefinición del concepto mismo de sistema rural, ya que éste se configura ahora como un grupo de factores interrelacionados en el que la vivienda rural se ha convertido en un nuevo producto claramente discernible. El nuevo sistema rural pasa entonces a componerse de unidades de viviendas, unidades de servicios (viales de acceso, abastecimiento de agua, de electricidad, etc.), unidades funcionales (acceso a servicios básicos: centros sanitarios, educativos, etc.) y unidades de producción.

En este nuevo marco, y concretamente respecto al tratamiento de la vivienda rural en los instrumentos de planeamiento urbanístico y planificación territorial, se requiere una nueva consideración de ésta, alejada del

carácter residual que hasta ahora se le había venido dando. Si, por una parte, la distinción tradicional entre vivienda rural y urbana, realizada según planteamientos políticos y territoriales, está desapareciendo de la realidad territorial, social y económica de la zonas más dinámicas de nuestro país; por otra parte, la irrupción del proceso urbano en los espacios rurales durante las últimas décadas ha generado la formación de verdaderos entornos de aglomerados de viviendas que surgen bajo la figura del diseminado en un contexto de hábitat rural disperso. Estas nuevas formas de implantación eliminan la tradicional distinción entre vivienda rural y urbana y hacen que la consideración, que se ha venido dando a la vivienda rural en los instrumentos que regulan el proceso urbano, haya quedado obsoleta. La conceptualización de la vivienda rural debe alejarse ahora de la consideración de que es solo la casa o la estructura física y debe acercarse a que también son los servicios que requiere, la producción de esos servicios y la relación con el territorio. Al mismo tiempo, la implantación de la vivienda rural, al igual que la urbana, ya no depende de un ambiente natural sino de un ambiente creado por el hombre, con lo que eso conlleva de artificialización e intervención en el territorio.

Para abordar esta temática de investigación consideramos que el conjunto de los métodos cuantitativos de estadística espacial que tienen por objeto la detección y la descripción de patrones de distribución geográfica pueden ser herramientas útiles para la estimación de pautas de asociación, pudiendo encontrar herramientas para el análisis de patrones, para la asignación de clústeres, para la medición de distribuciones geográficas y para el modelado de relaciones espaciales.

Nuestra intención es proponer un método para identificar la existencia de patrones geográficos de distribución de la vivienda rural. Pensamos que con su aplicación se abre la posibilidad de evaluar hipótesis sobre su comportamiento y principalmente sobre el modelo de ocupación que haya podido generar el patrón observado. Nuestra hipótesis de partida es que la heterogeneidad espacial derivada de la implantación de viviendas rurales de forma aislada condicionará la estructura territorial y la coexistencia de usos, cuando lleguen a formarse aglomerados, los cuales suelen surgir bajo procesos de densificación en una dinámica aleatoria.

Si bien se puede llegar a tener una idea de un patrón general al realizar la representación cartográfica de las entidades que lo representan mediante sus coordenadas  $x$  e  $y$ , es imprescindible el cálculo de la estadística que cuantifique dicho patrón; esto posibilita la comparación de patrones para distintas distribuciones o para distintos períodos de tiempo. En este trabajo, las entidades, que representan viviendas rurales, son registradas mediante sus coordenadas proyectadas generándose así un conjunto de datos que recibe el nombre de patrón espacial de puntos. Partimos del supuesto de que la estructura que define este patrón de puntos, y sobre la que basaremos el análisis, es una realización concreta de un proceso espacial de distribución, que hay que describir en base a sus características definitorias y que nos podrá llevar a ofrecer una visión global de comportamiento territorial. Fundamentar el análisis en un proceso de puntos aleatorios que comparten la misma estructura espacial supone abordar un proceso estocástico basado en aspectos básicos de probabilidad estadística para observar cómo la dependencia espacial afecta a la inferencia estadística.

Los métodos cuantitativos de análisis espacial de patrones de puntos han sido profusamente aplicados en ecología forestal. En este contexto científico han tenido como objetivo la detección y la descripción de patrones de distribución espacial considerando la heterogeneidad espacial derivada de perturbaciones que lleva a la eliminación de árboles dominantes en el dosel forestal, lo cual puede condicionar la regeneración, la estructura de las poblaciones y la coexistencia de especies. El enfoque acerca de que la interferencia de una especie hacia otras ha sido un factor importante a la hora de determinar la estructura, dinámica y productividad de las comunidades vegetales durante mucho tiempo (Turner y Franz, 1985; Goldberg y Barton, 1992; Dalling et al., 1998; Antos y Parish, 2002; Badii et al., 2008). También han sido profusamente abordados los análisis de los patrones de variación espacial del crecimiento y la mortalidad para estudiar la competencia intraespecífica en poblaciones naturales (Kenkel, 1993; Duncan, 1991; He y Duncan, 2000). Por otra parte, en ecología forestal existe una gran diversidad de métodos para la cuantificación: estimación, tipo del patrón y rango espacial, en forma de índices que se encuentran descritos en diversas revisiones (Condés y Martínez-Millán, 1998; Dale, 1999; Liu, 2001). Otra de las aplicaciones más difundidas para las que se emplean las técnicas de análisis de patrones de puntos en ecología forestal es para medir la existencia de interacciones entre comunidades y poblaciones a partir del estudio del patrón espacial o disposición de los individuos, considerados a modo de entidades puntuales. Su aplicación se basa en la asunción de que el análisis del patrón espacial y de sus variaciones en el espacio (aunque también en el tiempo, si fuese el caso) 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; De la Cruz, 2006).

Si bien, en las aplicaciones anteriores se hace mención casi en exclusiva al comportamiento de las comunidades vegetales, pensamos que, al menos conceptualmente, se aproximan bastante a la experiencia que

nosotros pretendemos abordar. De hecho, estas técnicas también han sido aplicadas en temáticas afines a la de este trabajo como pueden ser la evaluación de la tasas de concentración del desarrollo urbano, la detección de patrones espaciales de distribución de usos urbanos, o la estimación de la densidad de vivienda como un indicador para la estimación de patrones de distribución urbanos (Mirbagheri y Matkan, 2009; Zhu et al., 2010; Kretser et al., 2008).

En el contexto descrito podemos encajar los objetivos del presente trabajo: 1) utilizar las técnicas de análisis espacial de patrones de puntos con la finalidad de analizar la existencia de interacciones entre entidades a partir del estudio del patrón espacial (la disposición) de las entidades, 2) probar diferentes técnicas para establecer pautas de asociación basados en procesos puntuales, fundamentalmente los referidos a procesos homogéneos de agrupamiento, concretamente se aborda la aplicación del análisis clúster espacial de distancia múltiple basado en la Función K de Ripley, la cual se considera un estimador muy difundido para la descripción de procesos puntuales y 3) representar modelos espaciales de patrones de distribución a partir de la información de muestreo establecida, basándonos en el descriptor anterior y mediante la aplicación de herramientas de asignación de clústeres como el análisis de punto caliente ( $G_i^*$  de Getis-Ord) y los mapas de densidades Kernel. Estos objetivos se aplicarán en un caso práctico real para ilustrar su utilidad en la inferencia de procesos territoriales y la demostración de la hipótesis de la que partimos.

## **2. METODOLOGÍA.**

Se expone a continuación la metodología aplicada, que parte con la reserva de que la mayoría de las zonas en el medio rural tienden a mostrar heterogeneidad, lo que no permitiría la aparición de patrones agregados, pero que también esto es una consecuencia dependiente de la escala y del ámbito geográfico. No obstante, esperamos que podamos encontrar dichos patrones en la zona de estudio que hemos seleccionado, un espacio de la provincia de Málaga que es muy representativo del proceso de ocupación difusa del medio rural por viviendas aisladas y donde éste se desarrolla con gran intensidad: la Comarca de la Axarquía (Yus y Robles, 2010; Galacho, 2011a y b).

Antes de exponer la metodología vamos a exponer una serie de consideraciones directamente relacionadas con las técnicas de estadística espacial. La utilización de técnicas de estadística espacial en general, y de análisis de patrones de puntos en particular, presenta tres peculiaridades fundamentales (Rozas y Camarero, 2005 y De la Cruz, 2006), que deben ser tenidas en cuenta en el diseño del método ya que determinarán su configuración: estacionariedad, isotropismo y efecto de borde. El concepto de estacionariedad implica que las entidades (puntos que representan la disposición de las edificaciones) deben estar normalmente distribuidos en el área de estudio, con la misma media y varianza, de tal manera que el proceso es homogéneo o invariante a la traslación; tal como hemos comentado, si la distribución de las viviendas rurales tiende a mostrar heterogeneidad, un patrón heterogéneo deberá mostrar diferente densidad en distintas áreas. El concepto de isotropía hace referencia a que un patrón espacial isotrópico muestra la misma intensidad en todas las direcciones (por ejemplo, viviendas equidistribuidas sobre un parcelario de regadío) por lo que el proceso sería invariante a la rotación, mientras que un patrón espacial anisotrópico (viviendas distribuidas a lo largo de un cauce o de una línea de cumbres) varía según la dirección. Finalmente, los parámetros anteriores pueden resultar condicionados por el efecto de borde, el cual vendrá determinado por la escala espacial y por tanto pueden tener influencia directa sobre los resultados del análisis del patrón de puntos, principalmente porque el análisis se fundamenta en la medición de la distancia entre viviendas, y por consiguiente, se pueden producir subestimaciones cerca de los bordes que será preciso corregir.

A continuación nos planteamos la definición del patrón espacial a analizar. Como se sabe, se puede optar por patrones espaciales simulados o por patrones espaciales reales. Ambas opciones pueden ser útiles, en nuestro caso hemos pensado que sea la realidad territorial analizada la que proporcione la lógica del modelo, optando pues por los segundos. De cualquier manera esperamos que el empleo de técnicas de análisis de patrones de puntos nos muestre las estructuras generadas a partir del análisis del patrón espacial (la disposición de las entidades) y, considerando sus variaciones en el espacio y en el tiempo, podremos observar la dinámica de la ocupación.

Antes de continuar con la exposición de la metodología, hacemos otro inciso para hacer una reflexión de una cuestión que consideramos importante para el desarrollo posterior del método. Nos hemos dado cuenta de que para comprender la estadística espacial es preciso entender la noción de patrón aleatorio, según lo cual cualquier área del plano tiene la misma probabilidad de contener un punto, una distribución de Poisson. Se parte pues, de que si el patrón de puntos mantiene una distribución aleatoria será la hipótesis nula para el análisis, esperándose con otras hipótesis alternativas que la distribución sea en agregados o que sea regular (Rozas y Camarero, 2005). En el caso de patrones agregados la probabilidad de encontrar un punto en las

inmediaciones de otro es mayor, mientras que en el patrón regular ocurre lo contrario. Esto puede estar determinado por el proceso de análisis de patrones de puntos que empleemos: bien, en base a las propiedades de primer orden de las entidades: el número esperado de puntos por unidad de área de acuerdo a su ubicación, o bien, en base a las de segundo orden: las relaciones posibles entre pares de puntos (por ejemplo, la probabilidad de encontrar puntos cercanos).

Bajo estas premisas, se realizó el inventario de edificaciones, diferenciando viviendas rurales recreativas de otras edificaciones ligadas al medio rural, procediendo a la captura de sus coordenadas  $x$  e  $y$ , que son registradas posteriormente como una capa vectorial de puntos. El muestreo resultante abarcó una superficie de unos 1.023,72 km<sup>2</sup> dando como resultado un total de 9.996 puntos. En la definición del muestreo se contemplan zonas claramente diferenciadas desde el punto de vista de sus características fisiográficas y de usos del suelo, aunque de extensiones muy diferentes. Por un lado, encontramos zonas de relieve alomado de entre 300 y 600 metros de altitud sobre el nivel del mar con uso agrícola de secano predominantemente y por otro, zonas de valle más llanas de entre 200 y 500 metros con uso de regadío. En la Figura 1 se puede observar el muestreo realizado que sirve de base para el análisis de patrones de puntos.



**Figura 1.** Zona de muestreo y mapa de puntos correspondiente al patrón real analizado. El espacio geográfico de aplicación es la comarca de la Axarquía en la zona oriental de la provincia de Málaga.

En base al muestreo anterior se construye la teoría del proceso puntual espacial siguiente, ya que representa el mecanismo estocástico más sencillo que puede generar un diseño puntual espacial y podrá aplicarse como modelo ideal de un proceso completamente aleatorio, resultando muy útil como descripción aproximada del patrón observado. Así pues, una vez registrados los puntos en la zona de muestreo se procedió a determinar estadísticamente si el patrón esperado era aleatorio (distribución de Poisson homogéneo) o difería de una distribución de Poisson de la misma intensidad (Upton y Fingleton, 1985); para ello se hicieron dos test: una prueba de bondad de ajuste que proporcionó los primeros indicios de aleatoriedad al que dimos la consideración de test riguroso y una prueba de aleatoriedad mediante simulación de Monte Carlo. Este último consiste en generar  $n$  pares de coordenadas aleatorias, siendo  $n$  igual al número de viviendas en la muestra. Se calcula el valor estadístico para cada serie de coordenadas aleatorias, repitiéndose el proceso un número determinado de veces. Los valores máximo y mínimo resultantes definirán los límites superior e inferior del intervalo de confianza estadística (Rozas y Camarero, 2005), que serán a su vez tenidos en cuenta al realizar el cálculo.

A continuación, decidimos probar diversos métodos de análisis del patrón espacial: 1) el análisis refinado de las distancia al vecino más próximo que informa sobre el patrón espacial observado como una función de la distancia  $d$  (Clark y Evans, 1954), 2) el análisis cluster espacial de distancia múltiple, que se basa

en la Función  $K$  de Ripley (Ripley, 1981) y 3) la metodología SADIE (Spatial Analysis by Distance Indices) (Perry y Dixon, 2002). Estos métodos son en cierta medida funciones de distribución acumulada ya que, a cada escala o distancia  $r$ , todos los pares de puntos separados por una distancia *menor que*  $r$  son usados para estimar el valor de la función correspondiente. Al mismo tiempo, las tres están basadas en la distribución de distancias entre puntos que mostraría un patrón de Poisson homogéneo, pero sólo la Función  $K$  de Ripley, además de trabajar con las propiedades de segundo orden en el que analizan las relaciones posibles entre pares de puntos (la probabilidad de encontrar puntos cercanos), las estima a todas las escalas y resume la dependencia espacial (agregados –clustering- de entidad o dispersión de entidad) en un rango de distancias.

Así, si un grupo de puntos se distribuye aleatoriamente en un proceso de Poisson con densidad  $\lambda$  y el número esperado de puntos en un círculo de radio  $d$  es  $\lambda\pi d^2$ , la desviación respecto a la aleatoriedad puede cuantificarse mediante la Función  $K$  de Ripley (Ripley, 1977, 1981; Upton y Fingleton, 1985) la cual determina si las entidades, o los valores asociados a las entidades, exhiben un clustering o una dispersión estadísticamente significativos en un rango de distancias, es decir, refleja el tipo, intensidad y rango del patrón espacial mediante el análisis de las distancias existentes entre todos los puntos. La función utilizada en nuestro caso es una transformación de la Función  $K$  de Ripley, donde el resultado esperado en un conjunto de puntos aleatorios equivale a la distancia de entrada. A continuación se muestra la transformación  $L(d)$ :

$$L(d) = \sqrt{\frac{A \sum_{i=1}^n \sum_{j=1, j \neq i}^n k(i, j)}{\pi n(n-1)}}$$

Donde:

$A$  es el área;  $n$  es el número de puntos;  $d$  es la distancia;  $k(i, j)$  es el peso, el cual (haya o no corrección), es 1 cuando la distancia entre  $i$  y  $j$  es menor o igual a  $d$ , y 0 cuando la distancia entre  $i$  y  $j$  es mayor que  $d$ . Cuando la corrección de borde se aplica, el peso de  $k(i, j)$  se modifica ligeramente.

En base a todo lo expuesto hasta ahora, la aplicación de la Función  $K$  se realizó con el programa ArcGis, donde encontramos en el módulo *Spatial Statistics Tools* el comando *Multi-Distance Spatial Cluster Analysis (Ripleys K Function)*. Se realizaron un total de 25 pruebas con distintas opciones de configuración, resultando como más satisfactoria a nuestro juicio, la resultante de aplicar los siguientes parámetros:

1) El número o cantidad de bandas de distancia. Se establecieron 10 (los valores más altos no parecen producir resultados mejores). Este parámetro determina el número de veces en que se incrementa el tamaño de la vecindad a la hora de analizar el conjunto de los puntos para su agrupamiento, en su aplicación lo que se especifica es el tamaño del incremento de la distancia de partida. Hemos observado que es necesaria la selección de una escala apropiada de análisis, necesitándose una banda de distancias o distancia de umbral para el análisis. Asimismo, era necesario explorar patrones espaciales en distancias múltiples y diversas escalas espaciales con la finalidad de observar cómo cambian los patrones y si estos podrían reflejar el dominio de procesos espaciales particulares. La Función  $K$  de Ripley ilustra cómo cambia el patrón de distribución: clustering espacial o la dispersión de los puntos, cuando se cambia el valor de distancia del rango de vecindad. Este estimador permite especificar la distancia a evaluar y, opcionalmente, establecer una distancia de inicio o un incremento de distancia, o ambas. Con esta información, se calcula la cantidad promedio de puntos vecinos asociados a cada punto; los puntos vecinos son aquellos que están más cerca de la distancia que se evalúa. A medida que aumenta la distancia especificada en la evaluación, cada punto tendrá por lo general más vecinos. Si la cantidad promedio de vecinos para una distancia de evaluación particular es más alta que la concentración promedio de puntos en el área de estudio, la distribución se considera en agregados a esa distancia (Mitchell, 2005).

2) La confianza estadística viene dada por lo que se denomina “sobre de confianza”. La hipótesis de aleatoriedad espacial proporcionada por la Función  $K$  de Ripley fue comprobada mediante 100 simulaciones de Monte Carlo. La Función  $K$  siempre evalúa la distribución espacial de los puntos en relación con la aleatoriedad espacial completa (CSR –complete spatial randomness-), en general, la suposición de que el patrón de puntos presenta una distribución aleatoria será la hipótesis nula para el análisis. Con la finalidad de garantizar el rango de confianza estadística para cada cálculo de la distribución aleatoria de los puntos se ejecutaron un mínimo de 99 permutaciones. Con ello se distribuía aleatoriamente el conjunto de puntos 99 veces para cada iteración, lo que coincidía con un nivel 0,05 en la prueba de significación de Monte Carlo. Después se seleccionó para cada distancia el valor  $k$  que se desvió hacia arriba y hacia abajo del valor  $k$

esperado; estos valores se convierten en los valores que marcan el intervalo de confianza, es decir, sus límites superior e inferior. Los valores  $k$  esperados suelen coincidir en forma y ubicación con la línea  $k$  esperada (denominada ExpectedK) como se muestra en la Figura 2. A pesar de que 99 permutaciones es lo mínimo aconsejable para dotar de confianza estadística al modelo (99 permutaciones es igual a un 99% de confianza estadística), existen opiniones que consideran mejor realizar 999 permutaciones (Marriott, 1979; Besag y Clifford, 1989), lo que supondría alcanzar el 99,9% de confianza estadística.

3) Existe la opción de establecer un campo de peso que represente la cantidad de entidades coincidentes en cada punto de ubicación. Nosotros hemos preferido realizar la Función  $K$  no ponderada porque hemos considerado que la asignación de pesos podría determinar los resultados esperados, ya que éstos en la Función  $K$  ponderada siempre estarán más agrupados que los resultados de la aplicación de la función sin asignación de peso. Establecer un campo de peso también influye en el cálculo del “sobre de confianza” o confianza estadística. Al no especificar un campo de peso, el “sobre de confianza” se ha creado a partir de la distribución de puntos aleatoria en el área de estudio y se ha calculado  $L(d)$  teniendo en cuenta esa distribución. Cada distribución aleatoria de los puntos se encuentra en estrecha relación con el número de permutaciones especificadas en el punto anterior.

4) Otros parámetros que se pueden incluir son la distancia a partir de la cual se iniciará el análisis de conglomerados y la distancia de incremento desde cada iteración. El valor que se introduce debe estar en unidades de mapa, es decir, deben ser valores numéricos en consonancia con el sistema de coordenadas del mapa, comúnmente en un sistema de coordenadas proyectado, no siendo válido un sistema de coordenadas geográficas. La particularidad de aplicar estos parámetros se encuentra en que se perfila el patrón en agregados con más significancia si las distancias que se incluyen son mayores. O por el contrario, a distancias menores, si la tendencia es en agregados, ésta se observa con menor evidencia. Nosotros hemos probado tres distancias 2.000, 5.000 y 10.000 metros, con incrementos variables; cantidades que vendrán determinadas por las dimensiones de la muestra utilizada.

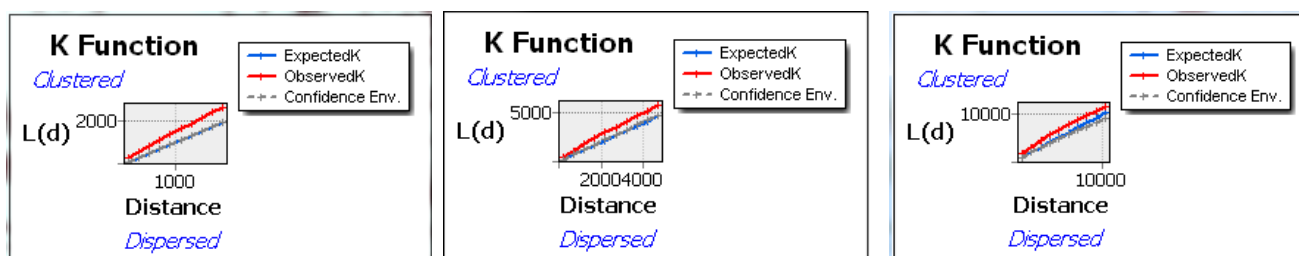
5) Como se comentó al principio, hay que tener en consideración el efecto de borde de cara a la afinación de los resultados finales. Basándonos en la extensa literatura que existe en ecología forestal para el tratamiento del efecto de borde (Getis y Franklín, 1987; Haase, 1995; Goreaud y Pélissier, 1999), hemos probado tres métodos: 1) Simular los valores de los bordes. Mediante este método se simulan una serie de puntos fuera de la zona de estudio para que el número de puntos vecinos cerca de los bordes no sea subestimado. Así, los puntos que se encuentran cerca de los bordes son duplicados como puntos “espejo”, siendo posicionados en la distancia más cercana a los límites del área de estudio dentro de la banda de máxima distancia del borde. 2) Reducir el área de análisis. Mediante esta técnica de corrección de borde se reduce el tamaño de la zona de análisis por una distancia igual a la banda de distancia más grande que haya sido utilizada en el análisis. Después de esta reducción, los puntos que se encuentran fuera de la nueva área de análisis se tendrán en cuenta solamente para la corrección del efecto de borde y no para el cálculo de las interacciones o recuentos de los puntos vecinos. 3) Fórmula de Ripley. Este método comprueba la distancia de cada punto en el borde del área de análisis y su distancia a cada uno de sus vecinos. Todos los puntos que estén más lejos del punto en cuestión son sobreestimados respecto a los anteriores. De los tres métodos mencionados hemos optado por la Fórmula de Ripley (*Ripley Edge Correction Formula*). Para todos los puntos ( $j$ ) en la zona del punto  $i$ , este método comprueba si el borde de la zona de estudio está más cerca de  $i,j$  o si está más cerca de  $i$ . Si  $j$  está más cerca, el peso adicional se da hasta el punto  $j$ . Todos los puntos que estén más lejos del punto en cuestión son sobreestimados respecto a los anteriores. Con ello, todos los puntos vecinos que estén más alejados del punto en cuestión que el borde del área de estudio se les asigna un peso extra. Es preciso reconocer que este método es más apropiado para áreas de forma rectangular o cuadrada, algo que solo en parte cumple nuestra área de estudio, no obstante, es posible matizar este aspecto si se aplica en el parámetro siguiente la definición del área de estudio.

6) Otro parámetro es la definición del área de estudio mediante la especificación de una capa con el límite de la zona estudio. Es importante tener en consideración este parámetro ya que la Función  $K$  es sensible al tamaño del área de estudio, dando resultados más homogéneos a escalas más pequeñas y viceversa. Consideramos que es conveniente introducir una capa con el límite del área de estudio porque mejoran sensiblemente los resultados. De lo contrario, la herramienta utiliza un rectángulo de encuadre mínimo que abarque el conjunto de los puntos del área de estudio, con lo que no necesariamente se producirá la alineación necesaria con los ejes  $x$  e  $y$ .

7) Finalmente se aplica la transformación  $L(d)$  que tiene por objeto linealizar la función y estabilizar la varianza con lo que el valor  $K$  esperado es igual a Distancia:

$$L(d) = \sqrt{K(d)/\pi}$$

El resultado final expuesto de modo gráfico se muestra en la Figura 2 siguiente:



**Figura 2.** Gráficos de resultados de la aplicación de la Función  $K$  de Ripley a las tres distancias seleccionadas: 2.000, 5.000 y 10.000 metros.

Como se muestra en la Figura 2, la interpretación de los resultados de la aplicación de la Función  $K$  suele realizarse representando  $L(d)-d$  frente a la distancia  $d$ , lo cual ajusta la hipótesis nula al valor cero (patrón aleatorio). Un patrón en agregados tiene lugar cuando  $L(d)-d$  frente a la distancia  $d$  es significativamente mayor que cero y un patrón regular cuando  $L(d)-d$  frente a la distancia  $d$  es significativamente menor que cero (Rozas y Camarero, 2005). De esta manera, en la lectura de los gráficos hay que atender a los conceptos *ExpectedK* y *ObservedK*, que representan los valores  $k$  esperados y  $k$  observados respectivamente (los valores resultantes se encuentran contenidos en una tabla de contingencia). Debido a que se aplica la transformación  $L(d)$ , los valores *ExpectedK* deberán coincidir con el valor de la distancia. El concepto *Confidence Env.* muestra el intervalo de confianza estadística en el que gráficamente aparecen resumidos los resultados de los valores máximo y mínimo que son lo que determinan los límites superior e inferior del intervalo de confianza. También se dispone en la tabla de contingencia mencionada de dos campos con la denominación *LwConfEnv* y *HiConfEnv*, que contienen información del intervalo de confianza para cada iteración según se haya especificado en el parámetro de cantidad de bandas de distancia. Se menciona que aparecen resumidos porque también en la tabla de contingencia mencionada se dispone de un campo *Diffk* que contiene la diferencia entre los valores  $k$  observados y los valores  $k$  esperados. Respecto a la lectura de los gráficos se observa que la línea que representa el valor  $k$  observado (*ObservedK*) está por encima de la línea que representa el valor  $k$  esperado (*ExpectedK*) para las distancias establecidas (escala de análisis), y por lo tanto la distribución es más agrupada que una distribución aleatoria es esas distancias. En el caso contrario, la distribución sería más dispersa (regular) que una distribución aleatoria a esas mismas distancias. Al mismo tiempo, en la tabla de contingencia los valores  $k$  observados son mayores que los valores del campo *HiConfEnv*, por lo el clustering espacial resultante para esa distancia es estadísticamente significativo (en los casos mostrados esto ocurre en todas las bandas de distancias establecidas menos en una, la que se corresponde con los 300 metros). En resumen, se observa que el patrón de distribución en el área de estudio tiende más a una distribución agrupada o en agregados. Esto confirma nuestra hipótesis de partida en la que se exponía que la implantación masiva de viviendas rurales en las últimas décadas está produciendo un patrón espacial de distribución en agregados, con lo que se están formando aglomerados de viviendas que surgen bajo una génesis individual.

Una vez obtenidos los resultados de la aplicación de la Función  $K$  hemos podido detectar el patrón de distribución de una forma estadísticamente fiable y con una representación gráfica de fácil interpretación. Con ello hemos respondido a la existencia o no de clustering espacial. Ahora que ya se conoce el patrón de distribución espacial de puntos, lo ideal es utilizar las herramientas de asignación de clústeres para mostrar la distribución territorial. Estas herramientas son especialmente útiles cuando se necesita mostrar la ubicación de un o más clústeres, tras el análisis estadístico de los patrones de distribución. Del conjunto de estas herramientas hemos utilizado dos: el análisis de punto caliente (*Gi\* de Getis-Ord*) y los mapas de densidades *Kernel*.

La función estadística *Gi\* de Getis-Ord* (se suele pronunciar *G-i-estrella*), que se denomina análisis de punto caliente, es de gran utilidad para identificar dónde se pueden producir clústeres espaciales significativos mediante el cálculo de la probabilidad estadística, mostrándose a través de una gradación de valores que va de valores altos (puntos calientes) a valores bajos (puntos fríos); para ello se establecen unas puntuaciones  $z$  (desviación estándar) y valores  $P$  (probabilidad de que el patrón espacial observado sea el resultado de procesos

aleatorios) que son medidas de significancia estadística e indican si se rechazará la hipótesis nula, punto por punto. Dado que partimos de un patrón ya establecido o conocido, nos interesa saber dónde se pueden producir las agrupaciones espacialmente, mostrando la probabilidad de la agrupación según su valoración estadística. Para ser considerado como punto caliente estadísticamente significativo, dicho punto debe cumplir dos condiciones: tener un valor alto y estar rodeado por otros puntos con valores altos. La conjunción espacial de puntos con estas características son especialmente significativas ya que nos muestra dónde se producen los aglomerados y con qué grado de confianza estadística. Esta función también se encuentra disponible en el programa ArcGis, módulo *Spatial Statistics Tools*. El resultado gráfico se muestra en la Figura 3.

Los resultados de la función anterior se completan con otro estadístico que calcula las magnitudes de áreas a partir de los puntos de muestreo o entidades: *Densidad Kernel*. Una vez conocidas las zonas de probabilidad de concentración con el estimador anterior a través de la localización de puntos calientes, ahora se trata de mostrar las magnitudes de las áreas donde se puedan producir los aglomerados. La estimación *Kernel* es un método no paramétrico de estimación, en el cual, en la formación de función de densidad no se imponen asunciones tan rígidas a los datos como sucede en el planteamiento paramétrico, siendo una forma útil de localizar concentraciones de puntos a través de estimaciones de densidad. Si el patrón espacial de distribución tiende a mostrarse en agregados, éste debe mostrar diferentes densidades en distintas áreas, ya que este tipo de análisis espacial calcula la densidad de puntos por unidad de área. Esta función también se encuentra disponible en el programa ArcGis, módulo *Spatial Analyst Tools*. El resultado de la aplicación de esta función se muestra en la Figura 4.

### 3. DESCRIPCIÓN Y ANÁLISIS DE RESULTADOS

En la Figura 3, se puede observar cómo en el área de estudio existen unas zonas (mostradas en rojo, y rodeadas de color naranja) que son considerados como puntos calientes y que se constituyen en los espacios de plasmación espacial del patrón de distribución de agregados con una certeza estadística del 99 %, siendo focos de concentración de viviendas. También se puede observar como ya existen áreas donde el proceso empieza a ser significativo, y aunque en la actualidad aparecen como puntos calientes con una certeza estadística del 90 %, la tendencia futura es a su consolidación. Al mismo tiempo, en dicha figura se puede observar, si bien aún como un proceso incipiente y por tanto con escasa significancia estadística, como las viviendas ya han irrumpido en el espacio natural protegido que constituye el Parque Natural de las Sierras de Tejeda y Almijara (mancha de color verde de la derecha de la imagen), lo que denota la potencia del proceso de urbanización difusa en la zona de estudio.

En la Figura 4, se muestra la densidad de las entidades en distintas zonas. Es posible observar la coincidencia de las zonas consideradas en la Figura 3 como puntos calientes con las zonas donde existe un grado de concentración significativa, considerando los valores más altos (Alto o Muy alto -representados con los colores azules más oscuros-) estimamos una superficie de 3.320 has. No obstante, si incluimos aquellos espacios donde la probabilidad de que se produzca concentración ha sido considerada como baja y se excluyen aquellos espacios donde la concentración no es significativa, la superficie aumenta hasta las 20.386 has. Esto significa que en el 19,65 % de la superficie (porcentaje respecto al total del área de estudio) está afectada por procesos de urbanización difusa.

Desde una perspectiva operativa podríamos ahora analizar las características particulares del proceso de implantación en las zonas donde los estimadores y las funciones de estadística espacial utilizadas nos indican que son áreas con mayor probabilidad de concentración de vivienda rural recreativa. El conocimiento que se tiene del área de estudio lleva a confirmar la certeza de las predicciones obtenidas y abre la posibilidad de nuevos análisis espaciales que podrían realizarse con las herramientas de análisis espacial que nos proporcionan los Sistemas de Información Geográfica.

El análisis combinado de las Figuras 3 y 4 no lleva a la demostración de que en el área de estudio ya se han formado zonas de agregados de viviendas que responden a un comportamiento territorial homogéneo. Así, tras analizar los resultados de la aplicación de las herramientas de análisis de patrones y las de asignación de clústeres podemos afirmar que más del 87,12 % de los puntos de muestreo que se corresponde con las viviendas rurales (8.696) presentan criterios de localización comunes respecto a otros elementos territoriales como otras viviendas, viales: caminos o carreteras y cauces. De hecho, el 57,75 % (5.697 viviendas) de las viviendas se encuentran situadas a una distancia media de 445 metros de otra vivienda; el 10,56 % (999 viviendas) a menos de 1 kilómetro de una carretera asfaltada y el resto, 8.896 viviendas (89,44 %), se encuentra contigua a un camino terrizo. El 3,35 % (299 viviendas) se encuentran a menos 500 metros de un cauce: río o arroyo.



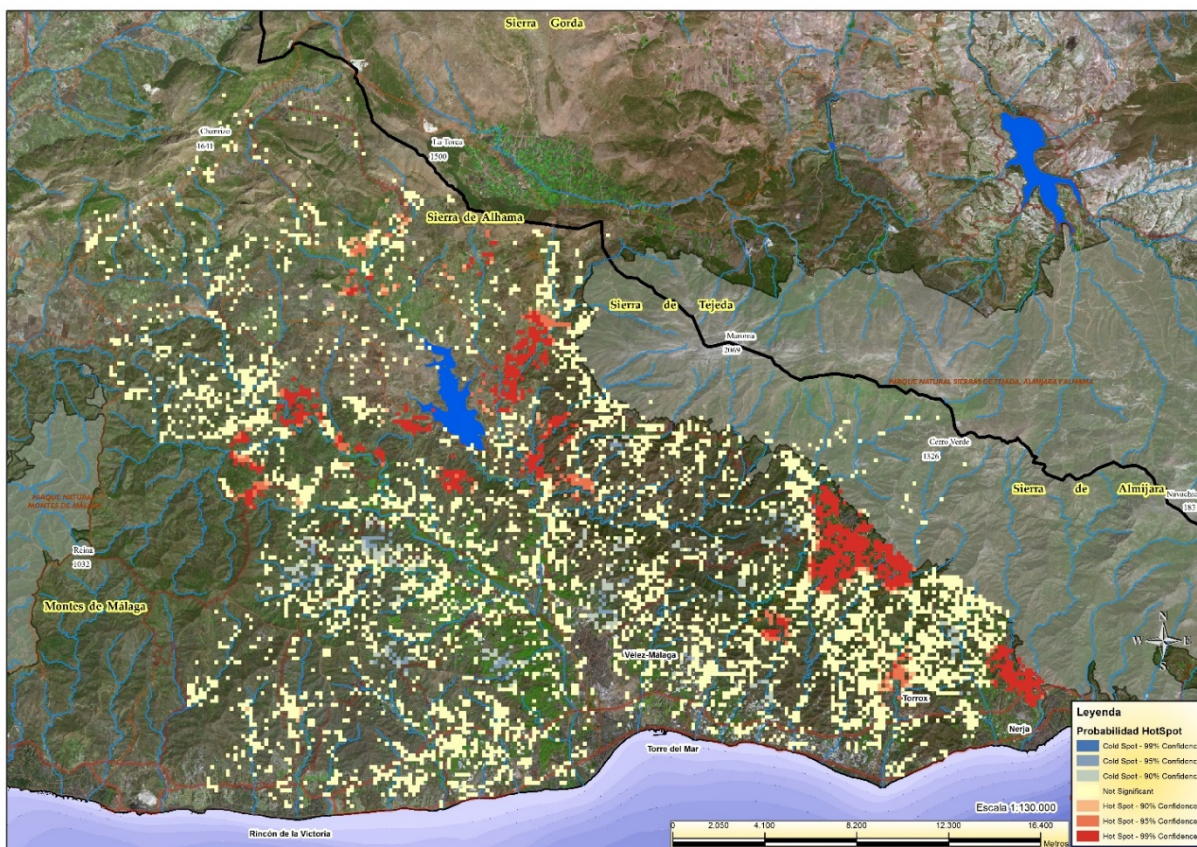


Figura 3. Mapa resultado de la aplicación de la función estadística  $G_i^*$  de Getis-Ord (Análisis Hot Spot)

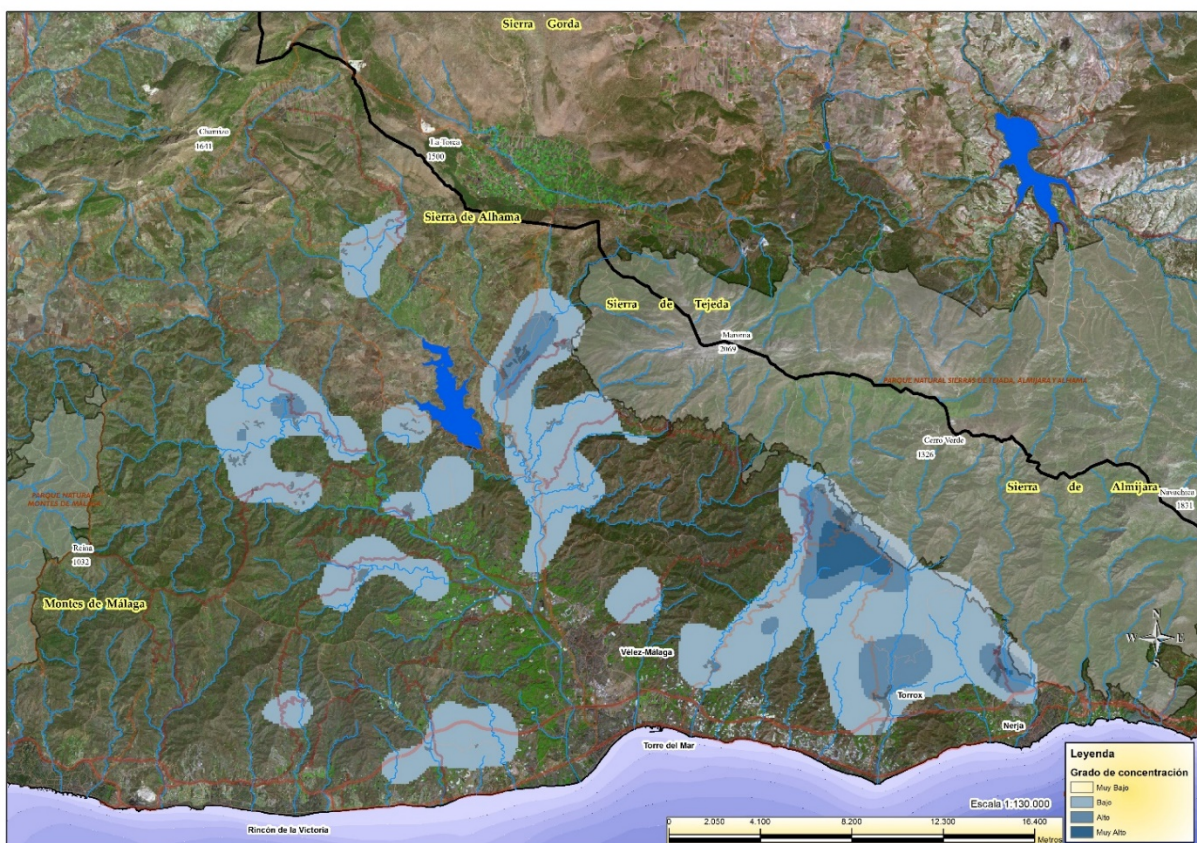


Figura 4. Mapa resultado de la localización de clústeres a partir de la aplicación de la función de Densidad Kernel.

Como se ha ido comprobando, las viviendas rurales del tipo recreativo, que son las que prevalecen en la zona de estudio, tienden a situar su emplazamiento en aquellas zonas donde ya existen otras viviendas, lo que confirma el patrón de distribución espacial en agregados descrito. Al mismo tiempo, se hace necesaria la accesibilidad a la vivienda por lo que si en una zona ya existen viales, estos actúan como factor favorecedor de la implantación.

#### 4. CONCLUSIONES

La incorporación de técnicas utilizadas en diversas disciplinas científicas, como la ecología, para el estudio de un problema de índole territorial como el que nos ocupa, si bien puede ser una aportación factible en la elaboración del método, nos obliga a no caer en simplificaciones excesivas. En un primer momento del análisis puede ser aclaratoria la tipificación de los patrones en regular, aleatoria o agregada, pero posteriormente se debe hacer una descripción más detallada y multidimensional basada en estadísticos funcionales o en la formulación de modelos explícitos del proceso subyacente. Y para ello son útiles las herramientas de asignación de clústeres con los que se consigue mostrar la distribución territorial del fenómeno analizado y además nos caracteriza en la relación con otras disciplinas científicas.

También hay que tener en cuenta, que cuando sea nuestro objeto el estimar pautas de asociación y patrones de distribución, la naturaleza del patrón generado estará determinada por la escala a la que el proceso es observado. En este sentido, la mayoría de las pruebas realizadas para el planteamiento de los patrones de distribución del problema analizado mostraban heterogeneidad a una escala lo suficientemente grande como para permitir la aparición de patrones agregados; mientras que a una escala menor, la variación puede ser menos acentuada y el patrón pasaba a estar determinado por la intensidad y la naturaleza de las interacciones entre los puntos muestreados. En este sentido, la aplicación de una herramienta clásica para analizar los patrones de distribución de puntos como la Función K de Ripley, da buenos resultados, aunque hay que tener en cuenta que su utilización produce mejores resultados cuando se trata de un proceso de puntos homogéneo, aunque existen modificaciones a esta herramienta cuando se trate de analizar entidades no homogéneas.

En este sentido, el basar nuestro método en un proceso de puntos nos proporciona un punto de inicio natural para la investigación estadística en el que buscar el patrón puntual observado, pero esto nos obliga a preguntarnos si se trata de un proceso Poisson homogéneo o no. Esto es importante porque el rechazo de la hipótesis de aleatoriedad no debe ser una sorpresa que nos pueda ocurrir. En nuestra aplicación hemos partido de hacernos la pregunta, ¿Qué tipo de patrón es?, pensando que todo el desarrollo metodológico posterior nos llevara a la demostración de que es un patrón agregado, por el contrario, si no hubiéramos podido confirmar la hipótesis de partida y nos encontramos con un patrón regular, existen un amplio número de modelos que son alternativas válidas pero el diseño metodológico debería ser otro. Es frecuente que sean realistas los modelos basados en procesos de Poisson no homogéneos, por eso en muchos trabajos se utilizan patrones espaciales simulados, en vez de patrones espaciales reales como ha sido nuestro caso. Esto supone una nueva vía de aplicación que pretendemos desarrollar en el marco de investigación que nos ha ocupado.

#### 5. BIBLIOGRAFÍA

- Antos, J.A., Parish, R. (2002): "Structure and dynamics of a nearly steady-state subalpine forest in south-central British Columbia, Canada". *Oecologia*, 130, 126-135.
- Badii, M.H., Landeros, J., Cerna, E. (2005): "Patrones de asociación de especies y sustentabilidad". *International Journal of Good Conscience*, 3(1), 632-660.
- Besag, J., Clifford, P. (1989): "Generalized Monte Carlo significance tests". *Biometrika*, 76, 633-642.
- Clark, P.J., Evans, F.G. (1954): "Distance to nearest neighbor as a measure of spatial relationships in populations". *Ecology*, 35, 445-453.
- Condés, S., Martínez-Millán, J. (1998): "Comparación entre los índices de distribución espacial de árboles más usados en el ámbito forestal". *Investigaciones Agrarias: Sistemas de Recursos Forestales*, 7, 173-187.
- Dale, M.R.T. (1999): *Spatial pattern analysis in plant ecology*. Cambridge, Cambridge University Press.
- Dalling, J.W., Hubbell, S.P., Silveira, K. (1998): "Seed dispersal, seedling establishment and gap partitioning among tropical pioneer trees". *Journal of Ecology*, 86, 674-689.

- De la Cruz, M. (2006): "Introducción al análisis de datos mapeados o algunas de las (muchas) cosas que puedo hacer si tengo coordenadas". *Ecosistemas (Revista Científica y Técnica de Ecología y Medio Ambiente)*, 15 (3), 19-39.
- Duncan, R.P. (1991): "Competition and the coexistence of species in a mixed podocarp stand". *Journal of Ecology*, 79, 1073-1084.
- Galacho, F.B. (2011a): "La irrupción del proceso urbano en los espacios rurales. Reflexiones acerca de las características de una nueva forma de relación entre territorio y economía". En Delgado, C., Juaristi, J., Tomé, S. (eds). *Ciudades y pasajes urbanos en el siglo XXI*. Santander, Ediciones de Librería Estudio, 297-314.
- Galacho, F.B. (2011b): "Implicaciones territoriales y aspectos sociales del urbanismo difuso en áreas de transición rural-urbana. Su análisis en la provincia de Málaga (España)". En *Actas del XXII Congreso de Geógrafos Españoles. Urbanismo expansivo. De la utopía a la realidad*. Alicante, Universidad de Alicante, 266-278.
- Getis, A., Franklín, J. (1987): "Second-order neighborhood analysis of mapped point patterns". *Ecology*, 68,473-477.
- Goldberg, D.E., Barton, A.M. (1992): "Patterns and consequences of interspecific competition in natural communities: A review of field experiments with plants". *American Naturalist*, 139, 771-801.
- Goreaud, F., Pélissier, R. (1999): "On explicit formulas of edge effect correction for Ripley's  $K$ -function". *Journal of Vegetation Science*, 10, 433-438.
- 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., Duncan, R.P. (2000): "Density-dependent effects on tree survival in a old-growth Douglas fir forest". *Journal of Ecology*, 88, 676-688.
- Kenkel, N.C. (1993): "Modeling Markovian dependence in populations of *Aralia nudicaulis*". *Ecology*, 74, 1700-1706.
- Kretser, H.E., Sullivan, P.J., Knuth, B.A. (2008): "Housing density as an indicator of spatial patterns of reported human-wildlife interactions in Northern New York". *Landscape and Urban Planning*, 84, 282-292.
- Liu, C. (2001): "A comparison of five-distance based methods for spatial patterns analysis". *Journal of Vegetation Science*, 12, 411-416.
- Marriott, F.H.C. (1979): "Barnard's Monte Carlo tests: how many simulations?" *Annual Review of Statistics and His Application*, 28, 75-77.
- Mirbagheri, B., Matkan, A.A. (2009): "Evaluating the concentration rate of urban áreas development using Ripley's  $K$  function in GIS". *Human Geography Research Quarterly*, 69, 51-66.
- Mitchell, A. (2005): *La Guía de Esri para el análisis SIG, Volumen 2*. Madrid, Esri Press.
- Perry, J.N., Dixon, P.M. (2002): "A new method to measure spatial association for ecological count data". *Ecoscience*, 9, 133-141.
- Ripley, B.D. (1977): "Modelling Spatial Patterns". *Journal of the Royal Statistical Society*, 39(2), 172-212.
- Ripley, B.D. (1981): *Spatial Statistic*. New York, Wiley.
- Rozas, V., Camarero, J.J. (2005): "Técnicas de análisis espacial de patrones de puntos aplicadas en ecología forestal". *Investigaciones Agrarias: Sist. Recur. For.*, 14 (1), 79-97.
- Seabloom, E.W., Bjornstand, O.N., Bolker, B., Reichman, O.J. (2005): "The spatial signature of dispersal and competition in successional grasslands". *Ecological Monographs*, 75, 199-214.
- Turner, D.P., Franz, E.H. (1985): "Size class structure and tree dispersion patterns in old-growth cedar-hemlock forest of the northern Rocky Mountains (USA)". *Oecologia*, 68, 52-56.
- Upton, G.J.G., Fingleton, B. (1985): *Spatial Data Analysis by Example. Vol.1. Point Pattern and Quantitative Data*. Chichester, Wiley

- Wiegand, T., Jeltsch, F., Hanski, I., Grimm, V. (2003): "Using pattern-oriented modeling for revealing hidden information: a key for reconciling ecological theory and application". *Oikos*, 100, 209-222.
- Yus, R., Torres, M.A. (2010): *Urbanismo difuso en suelo rústico. Deterioro ambiental y corrupción en la provincia de Málaga (El caso de la Axarquía)*. Málaga, Gabinete de Estudios de la Naturaleza de la Axarquía.
- Zhu, G., Ge, Y., Wang, H. (2010): "A modified Ripley's K function to Detecting Spatial Pattern of Urban System". Disponible en <http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=6626127&tag=1>. Consulta realizada el 5 de mayo de 2015.