

# DISTRIBUCIÓN ESPACIAL DE MELOCOTONEROS INFECTADOS POR SHARKA EN UNA PARCELA

P. Botella<sup>1</sup>, M. Cambra<sup>2</sup> y A. López-Quílez<sup>3</sup>

<sup>1</sup>Departamento de Atención Sanitaria, Salud Pública y Sanidad Animal.  
Universidad Cardenal Herrera - CEU

<sup>2</sup>Centro de Protección Vegetal. Gobierno de Aragón

<sup>3</sup>Departament d'Estadística i Investigació Operativa, Universitat de València

## Antecedentes y objetivos

Los datos analizados en el presente trabajo corresponden al análisis espacial de los árboles enfermos por Sharka (Plum Pox Virus, PPV) en una parcela de melocotoneros. En la parcela objeto de estudio había plantados aproximadamente 8900 árboles de los que se infectaron 53 y para los cuales se dispone de las coordenadas espaciales.

Con anterioridad a la detección del virus en los árboles en estudio se eliminó la plantación de una parcela contigua por la presencia de la misma enfermedad. Este trabajo pretende descartar la completa aleatoriedad en el fenómeno de aparición de la infección en los árboles, determinar la relación de la parcela contigua inicialmente infectada con este fenómeno y buscar algún mecanismo teórico para modelizar esta infección.

El conjunto de árboles infectados puede entenderse como un *patrón puntual*, es decir, una colección de localizaciones en una superficie, que presumiblemente han sido generados de alguna forma mediante un mecanismo estocástico (Diggle, 2003). Un *proceso puntual* es un modelo probabilístico que genera colecciones de ubicaciones de fenómenos que ocurren aleatoriamente en diferentes puntos de una región. Por tanto, un proceso puntual constituye un mecanismo por el que se pueden generar patrones puntuales. A la localización de cada uno de los puntos detectados se le llama *suceso*, para distinguirlo de cualquier punto de la región de estudio.

El análisis descriptivo de un patrón puntual pretende determinar mediante diferentes técnicas si dichos sucesos presentan un patrón de agregación o *agrupado* (los sucesos se producen cerca de otros sucesos), de inhibición o *regular* (los sucesos aparecen diseminados) o de *aleatoriedad espacial completa* (los sucesos se producen con igual probabilidad en cualquier punto del espacio con independencia de dónde se hallen otros sucesos).

En un proceso puntual los sucesos pueden aparecer en cualquier punto del espacio, mientras que en el problema que nos ocupa estamos ante una serie de árboles fijados en ciertas localizaciones, de los cuales alguno podría infectarse produciendo un suceso. Los datos disponibles consisten en la ubicación de aquellos árboles infectados, pero no

disponemos de la ubicación del resto de árboles de la parcela. Por este motivo y debido a la cantidad y densidad de la distribución de árboles se ha optado por aproximar el problema mediante un proceso puntual, considerando la parcela como una región continua en la que puede aparecer un suceso en cualquier punto de la misma.

## Caracterización del patrón puntual

La independencia entre las localizaciones de los sucesos y la probabilidad uniforme en toda la región conducen al concepto de *aleatoriedad espacial completa* (AEC) generada por un *proceso de Poisson homogéneo*. Las desviaciones con respecto a esta situación pueden darse hacia una mayor regularidad, producida por procesos de inhibición, o hacia un mayor agrupamiento, creando grupos de puntos muy próximos.

Para contrastar la hipótesis de AEC en el patrón se han empleado métodos basados en distancias (Diggle et al., 1976). Para estos métodos pueden emplearse las distancias observadas entre todos los sucesos o las distancias hasta el suceso más próximo, bien sea desde la localización de un punto del patrón o desde un lugar cualquiera de la región estudiada. Las funciones de distribución de estas distancias, que denotamos por H, G y F respectivamente tienen una forma conocida bajo la hipótesis de AEC y pueden compararse mediante métodos Monte Carlo con las funciones empíricas calculadas en el patrón puntual analizado (Møller y Waagepetersen, 2004). Además se ha utilizado un contraste de AEC basado en el uso de la función F-G. Esta función calculada para nuestros datos se puede comparar con las obtenidas por simulación para un mecanismo completamente aleatorio, obteniendo un p-valor aproximado para el contraste de la hipótesis de AEC.

El análisis del patrón de melocotoneros infectados mediante los métodos de distancias descarta la hipótesis de AEC. Un estudio detallado de las representaciones gráficas apunta a un mayor nivel de agrupación en nuestros datos de lo esperado para un patrón con AEC.

## Estimación del proceso de infección

Al igual que describimos cualquier conjunto de observaciones mediante una medida de tendencia y una de dispersión, podemos describir el comportamiento de un patrón puntual mediante las propiedades de primer y segundo orden, representadas por la función de intensidad y la función K de Ripley respectivamente.

La *función de intensidad* en un punto cualquiera de la región de estudio mide la densidad de casos por unidad de área en ese punto. La estimación de esta función puede realizarse de forma puntual o mediante una suavización con una función *kernel* (Silverman, 1981). Este tipo de funciones tienen un parámetro de ancho de banda que controla el grado de suavización en el cálculo de la función de intensidad. El valor

de este parámetro representa la distancia de influencia alrededor del punto a partir de la cual se calcula la intensidad en dicho punto. En este trabajo se ha empleado la función *kernel cuártica* y se ha utilizado una distancia de 100 metros como valor del ancho de banda, que si bien no llega a ser el valor que minimiza el error cuadrático medio proporciona una función de intensidad lo suficientemente suave.

En la representación gráfica de la intensidad del proceso se observa que la zona más cercana a la parcela contigua toma valores mayores, decayendo al aumentar la distancia a dicha parcela infectada. Esto parece confirmar las sospechas de contagio directo desde esa región vecina.

La *función K de Ripley* permite captar la dependencia espacial entre los sucesos mediante la proporción esperada de pares de sucesos a una distancia determinada. Para un proceso de Poisson homogéneo la función K de Ripley tiene una expresión teórica conocida proporcional al cuadrado de la distancia. Una forma de analizar la función K calculada en los datos (empírica) es como antes mediante simulación de Monte Carlo, comparándola con la función K de los patrones simulados con AEC. Esta comparación ha permitido también rechazar la AEC para nuestros datos y obtener de nuevo indicios de un comportamiento de mayor agrupación.

## Modelización del proceso puntual

Una vez descartada la AEC en el patrón de árboles infectados, buscamos un modelo que explique la aparición de estos sucesos. Observamos en la representación gráfica del proceso puntual cómo disminuye la intensidad del proceso a medida que nos alejamos de la parcela contigua inicialmente infectada. Por este motivo, y teniendo en cuenta la hipótesis de transmisión de la enfermedad a partir de esta parcela se ha intentado modelizar la intensidad del proceso no homogéneo en un punto cualquiera en función de la distancia de dicho punto a la parcela contigua.

Una primera aproximación es considerar que la función de intensidad fuera una función lineal de la distancia a la parcela, pero no supera un contraste Monte Carlo de bondad de ajuste. Una representación gráfica de la cantidad de puntos que se encuentran a una determinada distancia indica un decaimiento exponencial, por lo que la función de intensidad podría modelizarse como una función exponencial de un predictor lineal en términos de la distancia a la parcela vecina. Partiendo de este modelo, la estimación de los coeficientes puede realizarse por máxima verosimilitud o mediante un análisis bayesiano.

La maximización de la verosimilitud no es posible directamente, pero el logaritmo de la función de verosimilitud del proceso de Poisson no homogéneo es evaluable aproximando la integral de dicha expresión por una suma en un grid de puntos. De forma equivalente a este método, se ha considerado la modelización como un modelo lineal generalizado (GLM), utilizando la aproximación de Berman y Turner (1992). La

idea básica es dividir la parcela en celdas y suponer que el número de puntos de cada celda sigue una distribución Poisson con parámetro cuyo logaritmo es una función lineal de la distancia de la celda a la frontera con la parcela conflictiva (Baddeley y Turner, 2000).

La inferencia bayesiana en el modelo anterior puede realizarse mediante simulación utilizando el programa *WinBUGS* (Spiegelhalter et al., 2002). Este análisis tiene la ventaja respecto al anterior de permitir una mayor flexibilidad al modelo de la función de intensidad. Las medias de las distribuciones posteriores obtenidas son similares a las estimaciones proporcionadas por máxima verosimilitud aproximada.

Con este modelo estimado y de nuevo utilizando simulación de Monte Carlo, comparamos las funciones K y las funciones de distancias (F, G y H) empíricas, con los límites (mínimo y máximo) de las mismas funciones calculadas a partir de las simulaciones del modelo estimado. Tras esta comparación observamos que el comportamiento de nuestro patrón encaja *aproximadamente* con el comportamiento del proceso de Poisson no homogéneo con la función de intensidad estimada. Esta función de intensidad capta la mayor agregación de los puntos cerca de la parcela inicialmente infectada y la progresiva dispersión a medida que nos alejamos de dicha parcela. Pero se observa como a distancias pequeñas, la función G empírica de nuestros datos es superior a los límites obtenidos en las simulaciones. Este hecho indica que las distancias de cada suceso al suceso más próximo no se comportan de la misma forma en los patrones simulados y en nuestros datos, es decir, encontramos en nuestros datos que a distancias pequeñas hay más parejas de sucesos de lo esperado, lo cual nos lleva a pensar en una mayor agregación de la que se puede captar con la función de intensidad exponencial propuesta.

Este comportamiento nos lleva a plantearnos la posibilidad de contemplar un *proceso cluster de Poisson*, en el que los *padres* se comportarían como un proceso de Poisson no homogéneo con función de intensidad con la forma anteriormente propuesta (mayor intensidad cerca de la parcela inicialmente infectada y decaimiento rápido de esta intensidad al alejarnos de la misma), que darían lugar posteriormente a unos *hijos* distribuidos alrededor de los padres. Un proceso de este tipo, podría explicar la cantidad de sucesos *apareados* que podemos encontrar en la exploración visual del patrón en estudio.

El análisis del modelo obtenido por la estimación de los parámetros del proceso cluster de Poisson, ha permitido comprobar que este nuevo tipo de proceso con distribución gaussiana de los hijos alrededor de los padres capta las características del proceso en estudio que no conseguía captar la modelización anterior, proporcionando de esta manera un mecanismo teórico que puede explicar la infección de estos melocotoneros, relacionando además este fenómeno con la parcela contigua inicialmente infectada.

## Agradecimientos

Este trabajo ha sido parcialmente financiado por el *INIA-Ministerio de Ciencia y Tecnología* mediante el proyecto RTA03-099.

## Referencias

- Baddeley, A. y Turner, R. (2000). Practical maximum pseudolikelihood for spatial point patterns. *Australian and New Zealand Journal of Statistics* 42:283–322.
- Berman, M. y Turner, T. R. (1992). Approximating point process likelihoods with GLIM. *Applied Statistics* 41:31–38.
- Diggle, P. J. (2003). *Statistical Analysis of Spatial Point Patterns*. London: Arnold., Second Edition.
- Diggle, P. J., Besag, J. y Gleaves, J. T. (1976). Statistical analysis of spatial point patterns by means of distance methods. *Biometrics* 32:659–667.
- Møller, J. y Waagepetersen, R. P. (2004). *Statistical Inference and Simulation for Spatial Point Processes*. Boca Raton: Chapman and Hall/CRC.
- Silverman, B. W. (1981). Density estimation for univariate and bivariate data. V. Barnett (ed.), *Interpreting Multivariate Data*, 37–53.
- Spiegelhalter, D. J., Thomas, A., Best, N. G. y Lunn, D. (2002). *WinBUGS Version 1.4 User Manual*. Cambridge: MRC Biostatistics Unit.  
URL <http://www.mrc-bsu.cam.ac.uk/bugs>