



GRADO EN MATEMÁTICA COMPUTACIONAL

TRABAJO FINAL DE GRADO

**Análisis espacial de ratas y cucarachas en la
ciudad de Madrid mediante procesos
puntuales**

Autor:
Carlos ALBA GIL

Tutor académico:
Jorge MATEU MAHIQUES

Fecha de lectura: __ de _____ de 20__
Curso académico 2022/2023

Resumen

Las plagas de animales son un problema que afecta muy negativamente al medioambiente. Generan problemas de salubridad para las demás especies y pueden llegar a desestabilizar el ecosistema de las zonas afectadas. Prevenir las plagas suele ser la mejor opción, pero no es sencillo para los profesionales del control de plagas predecir nuevos eventos.

Este trabajo se centra en obtener la relación entre la presencia de ratas y cucarachas en un entorno urbano, concretamente en la ciudad de Madrid. Utilizando una base de datos sobre avistamientos de estas dos especies de animales en la ciudad, realizamos un estudio estadístico cuya finalidad es encontrar un modelo cuya generación de puntos se adecúe al patrón inicial.

Para ello, primero confirmamos que no se rigen bajo un modelo aleatorio calculando las características de primer y segundo orden. Estos indicadores no solo confirman que los patrones puntuales no muestran aleatoriedad, sino que además muestran un claro grado de agregación; lo que orienta el proyecto hacia modelos cuya principal característica es el agrupamiento de los eventos generados.

Finalmente, se obtiene un modelo que encaja con la mayoría de los patrones estudiados, mientras que para los restantes, se argumenta una posible explicación que justifique el comportamiento distinto a los anteriormente mencionados.

Palabras clave

Agrupamiento, Correlación, Cucarachas, Patrón puntual, Procesos de Cox, Ratitas

Keywords

Cockroaches, Correlation, Cox processes, Clustering, Point pattern, Rats

Índice general

| | |
|---|-----------|
| 1. Introducción | 7 |
| 1.1. Contexto y motivación del proyecto | 7 |
| 2. Metodología | 11 |
| 2.1. Proceso puntual | 11 |
| 2.2. Covariables | 15 |
| 2.3. Modelos | 16 |
| 2.3.1. Modelos de Poisson | 16 |
| 2.3.2. Modelos de agrupamiento | 17 |
| 2.3.3. Modelos de inhibición | 19 |
| 3. Resultados | 21 |
| 3.1. Tratamiento de los datos | 21 |
| 3.2. Patrón puntual | 22 |
| 3.2.1. Intensidad no paramétrica | 22 |
| 3.2.2. K de Ripley | 22 |

| | |
|---|-----------|
| 3.2.3. Función de correlación de pares | 22 |
| 3.3. Covariables | 23 |
| 3.4. Comparación de resultados | 24 |
| 3.4.1. Intensidades | 24 |
| 3.4.2. K de Ripley y correlación de pares | 24 |
| 3.5. Modelos | 25 |
| 4. Conclusiones | 27 |
| 4.1. Posibles extensiones | 28 |
| A. Lista de figuras | 31 |

Capítulo 1

Introducción

1.1. Contexto y motivación del proyecto

En el entorno urbano, diversos animales conviven de forma conjunta. Algunos, como los perros o los gatos, son animales que están ampliamente domesticados y controlados. Existen múltiples asociaciones que se encargan de controlar su presencia, y hay tecnología avanzada como los chips, que ayudan a la hora de localizarlos. También existe esta tecnología en animales que son propiedad de una granja. Los trabajadores necesitan saber si hay animales que se han perdido, ya que pueden reportar grandes pérdidas si eso ocurre. Estos dos ejemplos de especies están integradas en nuestro día a día.

Sin embargo, la fauna es amplia, y existen otras especies que, conviviendo con nosotros de la misma forma que las citadas en el párrafo anterior, generan problemas sanitarios cuando se cruzan con el ser humano. Es el caso de animales como las ratas y las cucarachas. Las estructuras de saneamiento son el ecosistema donde estos animales viven y se reproducen. El problema surge cuando en ocasiones, se avistan integrantes de este tipo de especies a pie de calle; generando un problema de salud pública por culpa de las enfermedades que propagan.

En este caso, las opciones mencionadas en el primer párrafo no surtirían un efecto notable. La tasa de reproducción es elevada y su tamaño exponencialmente inferior, por tanto sería extremadamente complicado introducir dispositivos electrónicos en éstos. Además, los lugares en donde se encuentran son usualmente inaccesibles para los controladores de plagas, lo cual dificulta de sobremanera su control.

Para facilitar el trabajo a los trabajadores, sería de gran utilidad obtener información relevante de los animales y su relación con el entorno. Existe un extenso banco de datos con

información sobre avistamientos de ambos animales desde el 2010 hasta el 2013 en la ciudad de Madrid y diversas covariables. Por tanto, sería interesante investigar sobre los distintos lugares que frecuentan en la superficie, la influencia que generan esas covariables en los animales y si una especie se comporta de una forma repelente, atractiva o indiferente en contacto con la otra.

Para ello, una opción coherente sería utilizar métodos para analizar patrones puntuales. En el amplio campo de la estadística espacial, los procesos puntuales y su estudio desempeñan un papel fundamental en la comprensión de distintos patrones y fenómenos. Estos procesos ofrecen una herramienta poderosa para analizar la distribución y la interacción de eventos en el espacio. Su aplicación abarca diversos campos, destacando su utilidad en disciplinas como la ecología o la epidemiología.

Los avances en el estudio de los procesos puntuales han dado como resultado el desarrollo de metodologías estadísticas sofisticadas que permiten analizar datos de manera más precisa y efectiva. Por ejemplo, el trabajo de P.J. Diggle, un referente en este ámbito, en su libro 'Statistical Analysis of Spatial Point Patterns' ha sentado las bases para el análisis de patrones espaciales de puntos y ha proporcionado herramientas estadísticas sólidas para estudiar fenómenos como la distribución de árboles en un bosque, la propagación de enfermedades o la distribución de crímenes en una ciudad.

J. Møller es otro investigador reconocido cuyas contribuciones han sido de vital importancia, como en el libro 'Shot noise Cox processes: Advances in Applied Probability'. Este trabajo se centra en la modelización de eventos aleatorios con intensidad espacial variable.

Además, Møller ha colaborado con el investigador R.P. Waagepetersen en trabajos como 'Statistical Inference and Simulation for Spatial Point Processes' y 'Modern spatial point process modelling and inference', que presentan enfoques modernos para el modelado e inferencia de procesos puntuales espaciales.

Por otro lado, los libros de D. Stoyan, W.S. Kendall y J. Mecke, como 'Stochastic Geometry and its Applications', han ampliado nuestra comprensión de la geometría estocástica y su posible aplicación en una amplia gama de disciplinas. Estas obras han destacado la importancia de los procesos puntuales para la descripción y el análisis de objetos y eventos en el espacio, y han influido en la forma en que se abordan problemas relacionados con la distribución de partículas o la formación de patrones fractales.

Se han realizado diversos estudios que utilizan estos procesos para resolver situaciones, como el de J.J. Abellán junto a otros investigadores en su trabajo 'Procesos puntuales como herramienta para el análisis de posibles fuentes de contaminación', cuyo estudio dio a las autoridades estatales la información y las pruebas suficientes para solucionar un brote de Legionella en la ciudad alicantina de Alcoy.

Otro de estos estudios podría ser el realizado por el tutor de este trabajo, Jorge Mateu, junto a otros compañeros llamado 'Aplicación de modelos de procesos puntuales para la caracterización espacio-temporal del régimen de incendios en el este de España '. En éste, la finalidad es modelizar posibles incendios futuros usando esta metodología.

Tomando en consideración los anteriores textos, llegamos a la conclusión de que el problema encaja a la perfección con los ámbitos y características en donde estos métodos de procesos puntuales son certeros.

Capítulo 2

Metodología

En esta sección vamos a desarrollar el fundamento teórico que hemos trabajado en este proyecto.

2.1. Proceso puntual

Definición 1 Sea $u = (u_1, u_2)$ un punto en R^2 siendo (u_1, u_2) sus coordenadas cartesianas, definimos la distancia euclídea entre 2 puntos u y v como

$$\|u - v\| = [(u_1 - v_1)^2 + (u_2 - v_2)^2]^{\frac{1}{2}} \quad (2.1)$$

Definición 2 Sea A el rectángulo de la forma $A=[a, b] \times [c, d]$ tal que para todo punto $u = (u_1, u_2)$ se cumpla que $a \leq u_1 \leq b$ y $c \leq u_2 \leq d$ definimos el disco de radio $r > 0$ con centro en u como

$$b(u, r) = \{v : \|u - v\| \leq r\} \quad (2.2)$$

Definición 3 Definimos la frontera como el disco de radio r con centro en u como

$$\partial b(u, r) = \{v : \|u - v\| = r\} \quad (2.3)$$

Los datos en forma de un conjunto de puntos distribuidos irregularmente dentro de un área aparecen en muchas situaciones relacionadas con la fauna y flora. A este tipo de conjunto de datos lo llamamos patrón espacial de puntos, donde cada uno de estos puntos se denomina evento.

De forma más técnica, podemos definir un proceso puntual espacial como un mecanismo estocástico que genera un conjunto de eventos $\{x_i\}_{i=1}^n$ en el área de observación W perteneciente a R^2 . Cuando el número de puntos en el proceso es finito, se le denomina proceso puntual finito. Por simplicidad, cuando se haga referencia en este trabajo a un 'proceso puntual finito', se usará 'proceso puntual'.

Habiendo obtenido los patrones que queremos analizar, necesitaremos calcular primero la intensidad inhomogénea de cada patrón de proceso puntual.

Definición 4 *Se define la intensidad de un proceso puntual como*

$$\lambda(u) = \lim_{|du| \rightarrow 0} \left\{ \frac{E[N(du)]}{|du|} \right\} \quad (2.4)$$

siendo N una variable de conteo y W el área de observación.

Si es un proceso estacionario, se asume que $\lambda(u)$ es constante y se representa como λ

Utilizaremos el suavizado por núcleo para obtenerla, consistente en estimar la intensidad en cada ubicación del patrón de puntos promediando la intensidad de los puntos cercanos. La estimación de intensidad en cada punto u está ponderada por la siguiente función núcleo que determina la influencia de los puntos cercanos

$$\lambda(u) = \frac{1}{e(u)} \sum_{i=1}^n \kappa_h(u - x_i) \quad (2.5)$$

siendo κ_h una función de suavizado gaussiana y

$$e(u) = \int_W \kappa_h(u - v) dv \quad (2.6)$$

un factor de corrección de borde.

Uno de los indicadores principales de la independencia de los puntos es la distancia que hay entre ellos. Si la mayoría de puntos están a una distancia pequeña, puede ser un indicador de agrupamiento. Sin embargo, cuando la distancia entre puntos suele ser regular, puede conllevar la existencia de un comportamiento de inhibición. La K-función es un indicador que ayuda a analizar esta independencia de los puntos.

Definición 5 Sea X un proceso puntual homogéneo con intensidad $\lambda > 0$, entonces $\forall r \geq 0$

$$K(r) = \frac{1}{\lambda} E[N_0(r)] \quad (2.7)$$

A esta fórmula se le conoce como K-función de X , donde $N_0(r)$ es el número de puntos que están a una distancia 'r' de otro.

Sin embargo, para poder comparar con diferentes conjuntos de datos, hemos de obtener una función empírica de K que esté normalizada. Definimos la K-función empírica como

$$\hat{K}(r) = \frac{|W|}{n(n-1)} \sum_{i=1}^n \sum_{j=1}^n 1\{d_{ij} \leq r\} e_{ij}(r) \quad (2.8)$$

siendo $|W|$ el área observada, n el número de puntos, $1\{d_{ij} \leq r\}$ la función indicatriz y $e_{ij}(r)$ un factor de corrección de borde similar a (2.6).

A esta función se le llama K de Ripley.

La K de Ripley indica el número de puntos que se encuentran a una distancia 'r' y la compara con el número de puntos que se encontrarían a esa distancia en un patrón aleatorio.

Otro indicador sería la función de correlación de pares.

Definición 6 Sea $K(r)$ la K-función respecto de r , podemos definir la función de correlación de pares como

$$g(r) = \frac{K'(r)}{2\pi r} \quad (2.9)$$

siendo $K'(r)$ la derivada de la K-función.

Sin embargo, para poder comparar con diferentes conjuntos de datos, hemos de obtener una función de correlación de pares que esté normalizada, normalmente usando un suavizado de núcleo. Definimos esta estimación como

$$\hat{g}(r) = \frac{|W|}{2\pi r n(n-1)} \sum_{i=1}^n \sum_{j=1}^n \kappa_h(r - d_{ij}) e_{ij}(r) \quad (2.10)$$

siendo $|W|$ el área observada, n el número de puntos, $e_{ij}(r)$ un factor de corrección de borde similar a (2.6), $d_{ij} = \|x_i - x_j\|$ la distancia euclídea entre los puntos i y j y κ_h la función de suavizado de núcleo con h como su ancho de banda de suavizado, representada por

$$\kappa_h(x) = \frac{1}{h} \kappa\left(\frac{x}{h}\right)$$

Para entender la diferencia entre ambos indicadores, la función de correlación de pares calcula la probabilidad de que dos puntos cualquiera de la muestra estén a una distancia r , mientras que la K de Ripley representa la cantidad de puntos que están a una distancia menor que ese r .

En el caso de que el patrón puntual sea inhomogéneo, estas funciones se han de ajustar para contemplar este hecho.

Definición 7 *Definimos la K -función inhomogénea como*

$$K_{inhom}(r) = E\left[\sum_{x_j \in X} \frac{1}{\lambda(x_j)} 1\{0 < \|u - x_j\| \leq r\} | u \in X \right] \quad (2.11)$$

mientras que su estimación se obtiene mediante

$$\hat{K}_{inhom}(r) = \frac{1}{D|W|} \sum_i \sum_{i \neq j} \frac{1\{\|x_i - x_j\| \leq r\}}{\hat{\lambda}(x_i)\hat{\lambda}(x_j)} e(x_i, x_j, r) \quad (2.12)$$

siendo D

$$D = \frac{1}{|W|} \sum_i \frac{1}{\hat{\lambda}(x_i)}$$

siendo $|W|$ el área de evaluación, $e(x_i, x_j, r)$ un peso de corrección de borde y $\hat{\lambda}(x_i)$ la estimación de la función de intensidad $\lambda(x_i)$.

La fórmula de la función de correlación de pares es la misma.

2.2. Covariables

Habiendo realizado una primera observación y tratamiento de los puntos, puede surgir las siguientes preguntas: ¿Es verdaderamente la propia especie lo que crea ese efecto de atracción, o son las características del entorno? ¿Se agrupan independientemente de los recursos naturales y mediciones de la zona o estas covariables tienen algo que ver? ¿Tienen las ratas y cucarachas los mismos hábitats predilectos y por tanto eso genera la falta de aleatoriedad en la muestra? Para ello, lo óptimo es realizar un análisis de las covariables y su posible implicación en el resultado.

Definición 8 *La intensidad proporcional a la línea de base se define como*

$$\lambda(u) = \theta b(u) \tag{2.13}$$

donde $b(u)$ es una función de línea de base conocida y θ un parámetro a estimar.

Definición 9 *La función exponencial de una covariable se define como*

$$\lambda(u) = \exp(\alpha + \beta Z(u)) \tag{2.14}$$

donde $Z(u)$ es una covariable y α y β son parámetros a estimar.

Si juntamos ambas definiciones, obtenemos el siguiente modelo.

Definición 10 *El modelo de incidencia elevada se define como*

$$\lambda(u) = b(u) \exp(\alpha + \beta Z(u)) \tag{2.15}$$

donde $b(u)$ es una función de línea de base conocida, $Z(u)$ es una covariable y α y β son parámetros a estimar.

Cuando tenemos n covariables, este modelo se expande de la siguiente forma.

Definición 11 *El modelo loglineal se define como*

$$\lambda(u) = \exp(\alpha + \beta_1 Z_1(u) + \dots + \beta_n Z_n(u)). \quad (2.16)$$

donde $Z_n(u)$ representa cada una de las covariables y los α y β son los parámetros que habremos de estimar estadísticamente.

Este modelo loglineal es el adecuado para nuestros datos, ya que tenemos diversas covariables.

2.3. Modelos

2.3.1. Modelos de Poisson

2.3.1.1. Modelo de Poisson homogéneo

El proceso de Poisson homogéneo es el fundamento sobre el cual se construye la teoría de los procesos puntuales espaciales. Es el mecanismo estocástico más simple de generación de patrones puntuales, usado como estándar de aleatoriedad espacial. Se basa en los siguientes postulados:

P1- Para $\lambda > 0$ y una región W , $N(W)$ sigue una distribución de Poisson con media $\lambda|W|$.

P2- Dado $N(W) = n$, los 'n' eventos de W forman una muestra aleatoria independiente de la distribución uniforme de W .

P3- Dadas dos regiones disjuntas A y B , $N(A)$ y $N(B)$ son independientes.

En este modelo la intensidad con la que se generan los puntos es constante a través de todo el intervalo.

2.3.1.2. Modelo de Poisson inhomogéneo

Si la intensidad varía espacialmente, nos encontramos con el modelo de Poisson inhomogéneo, basada en los siguientes postulados:

P1- $N(W)$ tiene una distribución Poisson con media $\int_W \lambda(u) du$.

P2- Dado $N(W) = n$, los 'n' eventos de W forman una muestra aleatoria independiente de la distribución uniforme de W con la función de densidad de probabilidad proporcional a $\lambda(u)$.

Suponiendo que después de investigar exhaustivamente los datos, encontramos que no siguen un modelo aleatorio, entonces se abre la puerta a otros modelos en los que los puntos son dependientes entre sí.

Existen dos grandes familias de modelos no aleatorios: de agrupamiento y de inhibición.

2.3.2. Modelos de agrupamiento

En este conjunto de modelos, los puntos tienen una dependencia positiva respecto los demás. Demuestran atracción, y se forman agrupamientos (clusters). Aplicado a este trabajo, la conclusión que se sacaría es que los integrantes de las mismas especies aparecen cercanos entre sí.

2.3.2.1. Procesos de agrupamiento de Poisson

Fueron introducidos por Neyman y Scott (1958). Estos modelos se rigen bajo un procedimiento basado en los siguientes postulados:

P1- Los eventos 'padre' forman un proceso de Poisson con intensidad ϕ .

P2- Cada padre genera un número aleatorio S de puntos 'hijo' independientes e idénticos para cada padre según una función de distribución $p_s : s = 0, 1, \dots$

P3- la posición de los puntos 'hijo' de sus respectivos puntos 'padre' son independiente e idénticamente distribuidos en base a una función de densidad de probabilidad $h(\cdot)$.

Los modelos de agrupamiento de Poisson que se van a analizar son los siguientes:

· **Matérn cluster (MatClust)**: Los puntos 'hijo' se generan alrededor de los puntos 'padre' dentro de un disco de radio arbitrario mediante la siguiente distribución

$$\rho(r) = \lambda(u) + (1 - \lambda(u))\Psi(r) \quad (2.17)$$

$\rho(r)$ representa la correlación entre dos puntos a una distancia r , $\lambda(u)$ es la intensidad inhomogénea del proceso puntual en la ubicación u , $\Psi(r)$ es la función de covarianza de Matérn, definida como

$$\Psi(h) = \frac{1}{\Gamma(\nu)2^{\nu-1}} \left(\frac{\sqrt{2\nu}h}{\beta} \right)^\nu K_\nu \left(\frac{\sqrt{2\nu}h}{\beta} \right) \quad (2.18)$$

donde ν es un parámetro que controla la suavidad de la función, β un parámetro que controla el alcance de la función, $\Gamma(\nu)$ la función Gamma y K_ν una función Bessel modificada.

· **Thomas**: Si nuestro proceso se adecúa a este modelo, significa que se la generación de puntos 'hijo' respecto de los 'padre' se ha realizado en base a una distribución gaussiana isotrópica

$$f(x|\mu, \sigma) = \frac{1}{(2\mu\sigma^2)^{\frac{1}{2}}} \exp \frac{-1}{2} \frac{(x - \mu)^T(x - \mu)}{\sigma^2} \quad (2.19)$$

donde μ es el vector de media y x el vector de entrada de 'd' dimensiones.

· **Cauchy**: En este caso, los puntos 'hijo' se generan respecto a la función de distribución bivalente de Cauchy

$$h(u) = \frac{1}{2\pi\omega^2} \left(1 + \frac{\|u\|^2}{\omega^2} \right)^{-3/2} \quad (2.20)$$

siendo ω un parámetro de escalado.

· **Variance-Gamma (VarGamma)**: Este modelo genera los puntos 'hijo' usando esta fórmula

$$h(u) = \frac{1}{2^{\nu+1}\pi\eta^2\Gamma(\nu+1)} \frac{|u|^\nu}{\eta^\nu} K_\nu \left(\frac{|u|}{\eta} \right) \quad (2.21)$$

siendo η un parámetro de escalado, ν siendo un factor que controla la forma de la densidad que debe cumplir $\nu > -1/2$ y K_ν una función Bessel modificada.

2.3.2.2. Procesos de Cox

Sabemos que, al usar procesos de agrupamiento de Poisson como modelos para procesos biológicos, es posible que se generen patrones puntuales con un alto grado de agregación por la agrupación de grupos de eventos relacionados. Otra posibilidad para este suceso podría ser la heterogeneidad ambiental. Un proceso inhomogéneo de Poisson con función intensidad $\lambda(u)$ producirá agrupamientos en zonas de alta intensidad. Sin embargo, la raíz de esa heterogeneidad podría ser aleatoria por naturaleza. A un proceso de Poisson cuya intensidad $\lambda(u)$ es aleatoria, se le conoce como proceso de Cox. Definimos un proceso espacial de Cox mediante los siguientes postulados:

P1- $\{\Lambda(u) : u \in R^2\}$ es un proceso estocástico no negativo.

P2- Si $\{\Lambda(u) = \lambda(u) : u \in R^2\}$, entonces los eventos forman un proceso inhomogéneo de Poisson con función de intensidad $\lambda(u)$.

· **Log-Gaussian Cox Process (LGCP):** Un LGCP es un proceso de Cox cuya intensidad es de la forma

$$\lambda(u) = \exp(G(u)) \quad (2.22)$$

siendo $G(u)$ un campo estocástico Gaussiano.

2.3.3. Modelos de inhibición

En este conjunto de modelos, también llamados modelos de Gibbs, los puntos tienen una dependencia negativa respecto los demás. Demuestran repulsión. Si alguno de estos modelos fuera adecuado para nuestros datos, significaría que los integrantes de la misma especie no conviven conjuntamente.

Los modelos de inhibición que vamos a analizar son los siguientes:

· **Strauss inhomogéneo:** Su función de densidad es

$$f(x) = C \gamma^{s(X,r)} \prod_{i=1}^n \beta(x_i) \quad (2.23)$$

con C constante, $\beta > 0$ es el parámetro que denota abundancia y $\gamma^{s(X,R)}$ es el parámetro de interacción elevado al número de puntos 'x' más cercanos que la distancia arbitraria de interacción 'r'

· **Matérn de inhibición:**

Está basada en la función de covarianza de Matérn, que cuantifica la correlación entre puntos.

$$\rho(h) = \lambda(u) \left(1 - \frac{1 - \exp\left(-\frac{h}{r}\right)}{\exp\left(-\frac{h}{r}\right) - \exp\left(-\frac{\sqrt{2\nu}h}{\beta}\right)} \right) \quad (2.24)$$

siendo $\lambda(u)$ la intensidad, 'r' la distancia arbitraria de interacción, ν es un parámetro que controla la suavidad de la función, β un parámetro que controla el alcance de la función de inhibición.

· **Simple Sequential Inhibition (SSI):** En este modelo, cada punto se genera independientemente dentro del área de observación. Si este punto generado está a una distancia menor de 'd' (la distancia arbitraria de interacción), se rechaza y se genera otro punto aleatorio.

De forma más teórica, este modelo se construye en base a los siguientes postulados:

Sea X_i una secuencia de 'n' eventos en W

P1- X_1 está uniformemente distribuido en W.

P2- Dado $\{X_j = x_j, j = 1, \dots, i - 1\}$, X_i está uniformemente distribuido en la intersección de W con $\{y : \|y - x_j\| \geq d, j = 1, \dots, i - 1\}$.

siendo d el diámetro de los discos que rodean a cada punto.

Capítulo 3

Resultados

Después de haber explicado los pasos que debemos realizar y la teoría que hay detrás de ellos, en este apartado finalmente se van a exponer los resultados de ejecutar el procedimiento teórico aplicado a este caso.

3.1. Tratamiento de los datos

Para facilitar el manejo, el cálculo y aumentar la granularidad, se han dividido los datos asociándolos por año. Tenemos datos de 2010 hasta 2013 de 2 especies, ratas y cucarachas. Por tanto, mayoritariamente trabajaremos con 8 subconjuntos de datos, desde ratas de 2010 hasta cucarachas de 2013.

Para evitar una repetición innecesaria de gráficas, si solo se muestran las gráficas del año 2010 de cada especie, implica que los demás años tienen resultados y conclusiones similares y por tanto no aportaría nada nuevo mostrar las 8 gráficas en la sección. En caso de existir diferencias, se mostrarían las que fueran necesarias.

Otro objeto necesario para conseguir resultados fiables es delimitar correctamente el área con la que queremos trabajar. Para ello se ha generado un polígono que encaja con la forma de la ciudad de Madrid.

3.2. Patrón puntual

Las Figuras A.1-A.3 describen los datos con los que se va a trabajar. Son una representación gráfica de los puntos geográficos donde se reportó un avistamiento.

3.2.1. Intensidad no paramétrica

En la Figura A.4 se muestra la intensidad no paramétrica tomando como ejemplo el año 2010. Las zonas con colores cálidos indican que existe una gran densidad de ocurrencias en ese espacio, mientras que las zonas de colores fríos denotan ausencia de ellas. Si analizamos el resultado, vemos que concuerda con la distribución de puntos mostradas en la Figuras A.2 y A.3.

3.2.2. K de Ripley

La Figura A.5 contiene las gráficas que muestran los valores correspondientes a 2010. La línea negra representa los valores de nuestros datos y la línea roja las de un patrón aleatorio. El entorno gris que envuelve a la línea roja muestra las bandas de confianza que definen el patrón aleatorio. El valor del ancho de banda escogido es calculado de forma que minimice el criterio de error cuadrático medio definido por Diggle (1985). Como se ve en la gráfica, los valores de nuestros puntos están muy por encima de lo que sería un patrón aleatorio. Eso nos indica que la cantidad de puntos que están a una distancia 'r' es mayor de lo esperado en un modelo aleatorio. Podemos sacar la conclusión de que estamos trabajando con unos valores que muestran un claro grado de agregación.

3.2.3. Función de correlación de pares

En la Figura A.6 podemos observar los valores correspondientes a 2010. La selección de colores de la gráfica es la misma que respecto a la K de Ripley. Nuevamente, podemos observar que los valores empíricos son muy superiores a los asociados a la línea constante con valor uno, que define un patrón aleatorio. Al igual que con la K de Ripley, esto también indica que los puntos registrados demuestran un claro grado de agregación.

De las figuras anteriores podemos sacar la conclusión de que hay un claro caso de agrupamiento entre animales de la misma especie. La siguiente pregunta que se nos puede ocurrir es, si hay influencia entre ambas especies. La solución es utilizar la versión 'cross' de los 2 métodos anteriores, lo cual daría lugar a una K de Ripley y correlación de pares bivalente. En este caso

la hipótesis nula sería la independencia. Si así lo fueran, serían similares a la línea azul en la primera gráfica y a la verde en la segunda. En la Figura A.7 observamos sendas gráficas, cuyo contenido muestra la dependencia entre ambas especies.

3.3. Covariables

Como se ha comentado anteriormente, se disponía de una extensa lista con covariables asociadas a la ciudad de Madrid. De entre todas ellas, se han seleccionado aquellas que pueden ser relevantes en la presencia y actividad de estos seres vivos:

- **Distancia al agua:** La distancia de cada punto en el mapa a el cuerpo de agua más cercano. Todo ser vivo con funciones complejas requiere de agua para sobrevivir. Es interesante saber cuánto condiciona esta distancia al comportamiento de las ratas y cucarachas.

- **Distancia a una zona verde:** La distancia a un parque o áreas naturales similares. De la misma forma que el agua, los seres vivos también necesitan comida. Las zonas verdes son perfectas para buscar alimento y refugio en caso de que escasee en las profundidades de la ciudad.

- **Distancia a una fuente:** Distancia a una fuente de agua potable. En este caso, en vez de cuerpos de agua, se mide la distancia a fuentes construidas por el ser humano. Posibles filtraciones subterráneas y la cantidad de fuentes urbanas que hay pueden ser un factor determinante.

- **Cantidad de euros anuales:** Salario anual medio de las personas que viven en la zona. Las condiciones de salubridad de una zona son proporcionales a la cantidad de dinero que recae sobre los residentes. Cuanto más dinero, más tratamiento de plagas, y menos forma para que los animales puedan reproducirse.

- **Tasa de desempleo:** Porcentaje de desempleados de la zona. Esta covariable está relacionada con la anterior.

Sus respectivas gráficas se encuentran en la Figura A.8. Utilizando estos valores, calculamos la intensidad paramétrica asociada, que nos servirá para calcular la K de Ripley y la función de correlación de pares. En la siguiente sección se comentarán los resultados y se compararán con los asociados a la intensidad no paramétrica.

3.4. Comparación de resultados

En este punto, para cada patrón de proceso puntual, tenemos la intensidad inhomogénea no-paramétrica (la calculada con la función núcleo) y la intensidad inhomogénea paramétrica (la que considera las covariables). El siguiente paso consta del estudio de estos datos frente a un modelo de Poisson. La principal característica de un modelo de Poisson es la independencia de los puntos respecto a otros. Es decir, las pasadas ocurrencias del modelo no afectan en la aparición de nuevas. Aplicado a nuestro problema, significaría que los integrantes de cada especie aparecen en lugares sin correlación con otras apariciones de otros integrantes.

3.4.1. Intensidades

En las Figuras A.9 y A.10 hemos mostrado distintas formas de representar la intensidad. La primera columna está asociada a la intensidad no paramétrica y la segunda a la paramétrica. Con un simple vistazo vemos que ese pico en el primer recuadro de la segunda columna distorsiona las representaciones posteriores. Por ello hemos añadido una tercera donde se limitan los valores que generan picos para que las siguientes gráficas contengan más información relevante.

Se puede ver ahora la diferencia de precisión entre los dos tipos de intensidades y su método de cálculo. Mientras que la primera es obtenida con una función de suavizado, la segunda presenta una distribución más realista y exacta.

Superponiendo los gráficos de la segunda fila, se observa que lo que nos describía la intensidad paramétrica encaja con lo que se describe en la no-paramétrica.

3.4.2. K de Ripley y correlación de pares

En las Figuras A.11 y A.12 se muestran estas comparaciones para ambas especies. La línea negra representa los valores de nuestros datos y la línea roja las de un patrón aleatorio. El entorno gris que envuelve a la línea roja muestra las bandas de confianza que definen el patrón aleatorio. Se concluye que incluso teniendo en consideración las distintas covariables, en las muestras se siguen apreciando claras muestras de agrupamiento.

3.5. Modelos

La Figura A.13 contiene los distintos valores de la K de Ripley dependiendo del modelo. La línea negra es la que representa el valor asociado al modelo, la roja representa nuestros datos, y la verde representa el valor asociado a una Poisson. De estos valores podemos ver indicios de qué modelos de agrupamiento no son los adecuados. Vemos como En el caso de MatClust o Thomas, su valor dista mucho de los datos en comparación con el VarGamma o el LGCP. Para poder certificar nuestra decisión, presentamos la función de correlación de pares en la Figura A.14. Aquí se observa con suma claridad que el LGCP y VarGamma tienen una silueta mucho más similar a los datos que los demás modelos. Por tanto, los tomamos para realizar pruebas para todos los años.

En estas pruebas, se confirma que el modelo de agrupamiento que encaja más con los datos es el LGCP, en la Figura A.15 vemos algunos ejemplos que justifican la decisión. El entorno gris que envuelve a la línea roja muestra las bandas de confianza que definen el modelo que estamos comparando. Sin embargo, vemos en la Figura A.16 que las cucarachas en 2012 y 2013 no tienen unos valores correctos para ninguno de los dos. Por tanto, se puede decir que en estos años, las cucarachas tienen un patrón que no se adecuaba con ningún modelo estudiado.

Finalmente, en la Figura A.17 se muestran algunas simulaciones aplicando nuestros datos a los modelos a los datos de ratas de 2010. Se observa que el LGCP genera unas agrupaciones similares al patrón inicial, mientras que los demás modelos no encajan con los datos iniciales.

Capítulo 4

Conclusiones

En este trabajo, se ha logrado cumplir el objetivo establecido de desarrollar un modelo que permita comprender la presencia de ratas y cucarachas en la ciudad de Madrid. Partiendo de una base de datos de avistamientos y unos valores de covariables, se han analizado diversos indicadores relacionados con los patrones puntuales que reforzaban la idea de la existencia de agregación en la muestra, como son la función de correlación de pares y la K de Ripley. A raíz de ello, se estudiaron varios modelos de agregación que pudieran satisfacer la distribución mostrada por la muestra, omitiendo el estudio de los modelos de inhibición.

Los resultados obtenidos revelan que las ratas siguen un patrón consistente y predecible a lo largo de los años de forma similar al modelo LGCP. Sin embargo, las cucarachas muestran la misma distribución que las ratas durante la primera mitad del tiempo y aleatoriedad en la segunda mitad. Esto sugiere que su distribución puede estar influenciada por factores variables, como cambios estacionales o condiciones ambientales específicas, que se agrandan durante los años 2012 y 2013. Este hallazgo destaca la importancia de considerar la variabilidad temporal al abordar el control de las cucarachas en la ciudad, ya que es posible la existencia de un factor crítico durante el año 2012 que haya modificado el patrón usual de comportamiento de la especie.

Al investigar posibles sucesos clave, he concluido que ese factor podría ser la publicación en octubre de 2012 de un proyecto informativo desarrollado por el ayuntamiento de Madrid llamado 'Artrópodos invasores ocasionales'. En él, se informa de diversos artrópodos y el procedimiento a seguir si el lector se encuentra con ellos. Aunque la cucaracha no sea uno de los artrópodos comentados en el artículo, está afectada directamente por el aumento de fumigaciones en lugares de baja salubridad. Este hecho podría haber generado la polarización en los resultados a partir de finales de 2012, modificando y diversificando los lugares de aparición de cucarachas.

Por tanto, queda demostrada la utilidad del mismo para los encargados de controlar a estas especies en la ciudad de Madrid. Los profesionales pueden utilizar los resultados y las predicciones generadas por el modelo para tomar decisiones sobre la asignación de recursos y el diseño de estrategias de control más eficientes, priorizando las áreas y los momentos críticos para maximizar la eficacia y minimizar los riesgos para la salud pública.

4.1. Posibles extensiones

A raíz de este trabajo, se pueden realizar otras investigaciones usando esta base de datos. Con más conocimiento de la materia, se podría haber estudiado la posibilidad de incorporar modelos espacio-temporales. Estos modelos permiten estudiar cómo la distribución de las especies evoluciona a lo largo del tiempo, teniendo en cuenta tanto la variabilidad espacial como la temporal, y permitirían identificar fluctuaciones estacionales o cambios a lo largo de los años. Otro concepto que no ha sido desarrollado en profundidad es el de la correlación entre especies. La presencia de una especie puede influir en la presencia de la otra debido a interacciones ecológicas o preferencias de hábitat similares. Una posible correlación abarataría los costes al poder ejecutarse planes conjuntos de control de plagas.

Ambas extensiones habrían alargado enormemente la longitud del proyecto y habrían aumentado significativamente la dificultad del mismo.

Otro enfoque que se podría dar es el uso de este trabajo con otras bases de datos. Si en el programa utilizado, se introdujeran otros avistamientos de especies y un polígono representativo del área a examinar, con pocas modificaciones se podría obtener información relevante sobre la nueva especie investigada.

En resumen, este estudio demuestra que los procesos de patrones puntuales son una herramienta útil en el análisis y control de diversas especies que generan problemas sanitarios en entornos urbanos. El modelo desarrollado en este trabajo proporciona una base sólida para la toma de decisiones informadas y el desarrollo de estrategias de control más efectivas en la ciudad de Madrid, y puede servir como ejemplo para abordar problemas similares en otras ciudades y contextos.

Bibliografía

- [1] A. BADDELEY, E. RUBAK y R. TURNER. *Spatial Point Patterns Methodology and Applications with R*. CRC Press, 2016.
- [2] P.M. DIXON. *Encyclopedia of Environmetrics*, Vol. 3, pp 1796–1803. 2002.
- [3] P.J. DIGGLE. *Statistical Analysis of Spatial Point Patterns*. CRC Press, 1983.
- [4] J. MØLLER. *Shot noise Cox processes: Advances in Applied Probability*, Vol. 35, pp 614-640. 2003.
- [5] J. MØLLER y R.P. WAAGEPETERSEN. *Statistical Inference and Simulation for Spatial Point Processes*. CRC Press, 2004.
- [6] J. MØLLER y R.P. WAAGEPETERSEN. *Modern spatial point process modelling and inference*, Scandinavian Journal of Statistics, Vol. 34, No. 4, pp. 643-684. Wiley, 2007.
- [7] D. STOYAN, W.S. KENDALL y J. MECKE. *Stochastic Geometry and its Applications*. Wiley, 1987.
- [8] J.J. ABELLÁN, M.A. MARTÍNEZ-BENEITO, O. ZURRIAGA, G. JORQUES, J. FERRÁNDIZ y A. LÓPEZ-QUÍLEZ. *Procesos puntuales como herramienta para el análisis de posibles fuentes de contaminación*. Gaceta Sanitaria, Vol. 16 No. 5, Barcelona, 2002.
- [9] P. ARAGÓ, P. JUAN SALVADOR, C. DÍAZ-AVALOS, L. SERRA, M. SÁEZ, D. VARGA, S. TRILLES, J. MATEU. *Aplicación de modelos de procesos puntuales para la caracterización espacio-temporal del régimen de incendios en el este de España*. Sociedad Española de Ciencias Forestales, 2013.
- [10] AYUNTAMIENTO DE MADRID. *Artrópodos invasores ocasionales*, 2012.

Anexo A

Lista de figuras



Figura A.1: Avistamientos totales de ratas y cucarachas

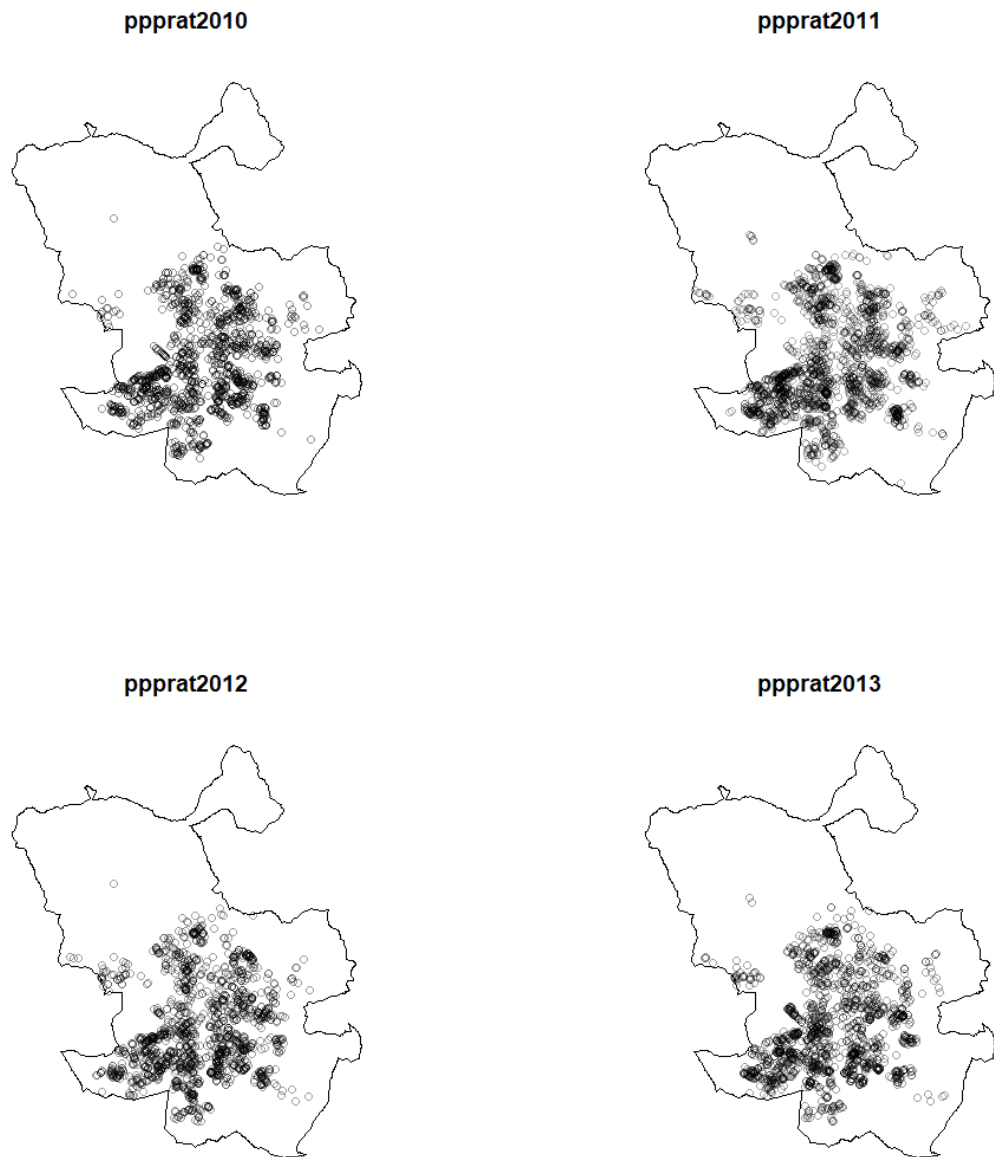


Figura A.2: Avistamientos de ratas agrupados por año

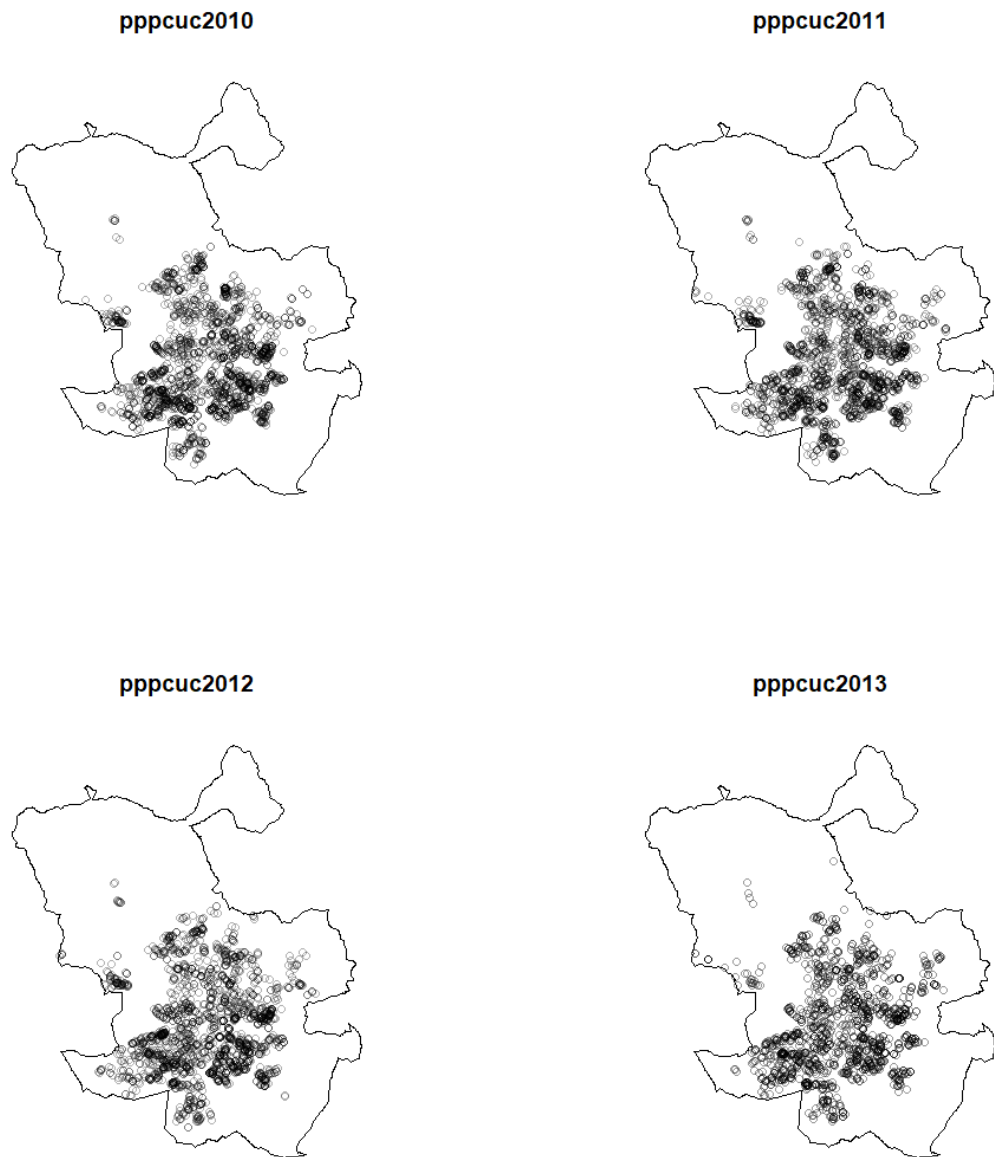


Figura A.3: Avistamientos de cucarachas agrupados por año

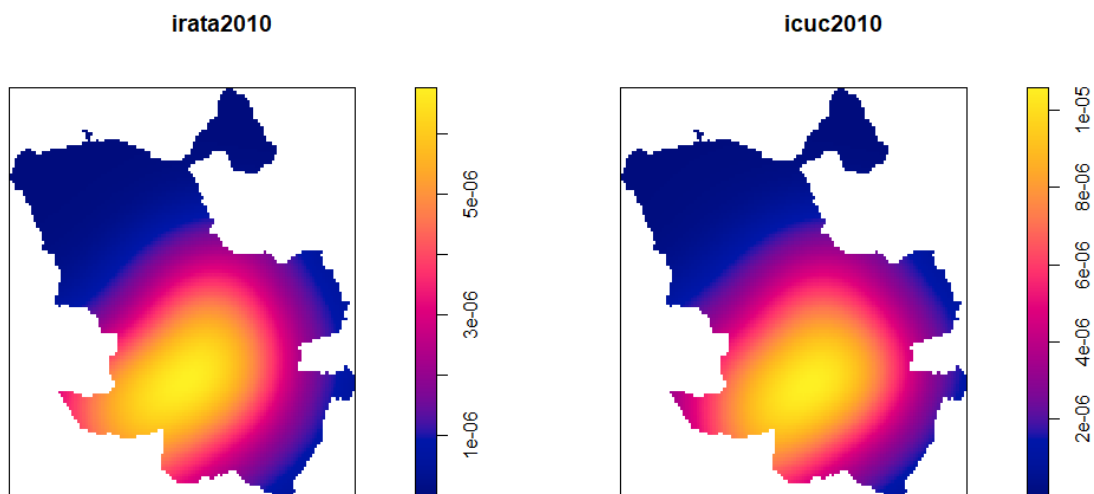


Figura A.4: Intensidad no paramétrica para las ratas y cucarachas avistadas en 2010

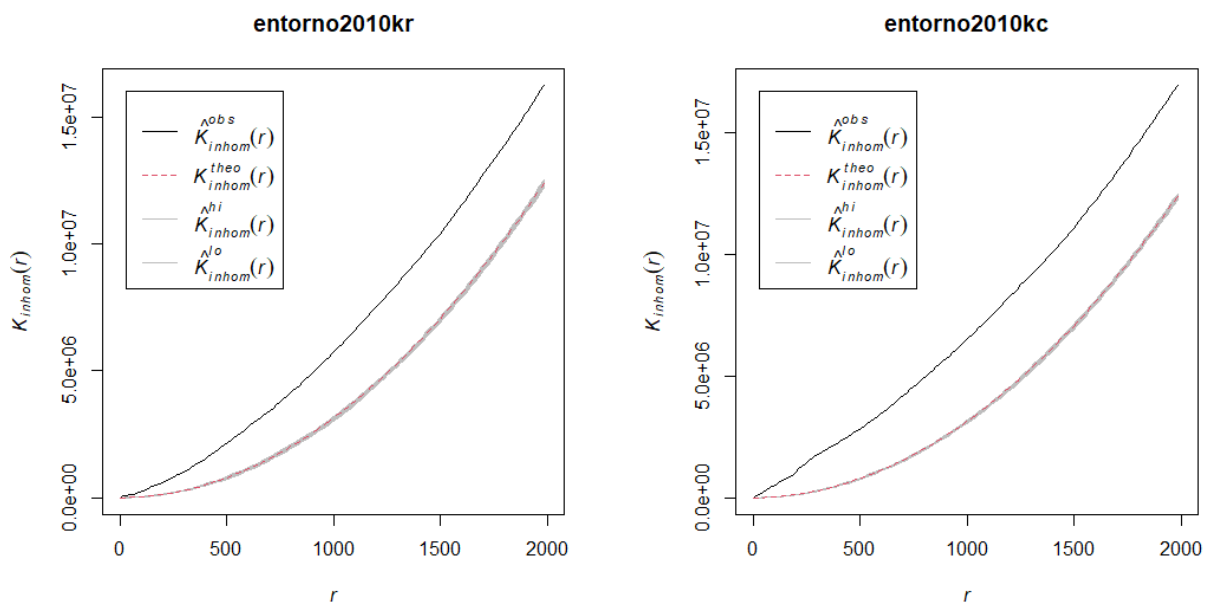


Figura A.5: K de Ripley para las ratas y cucarachas avistadas en 2010

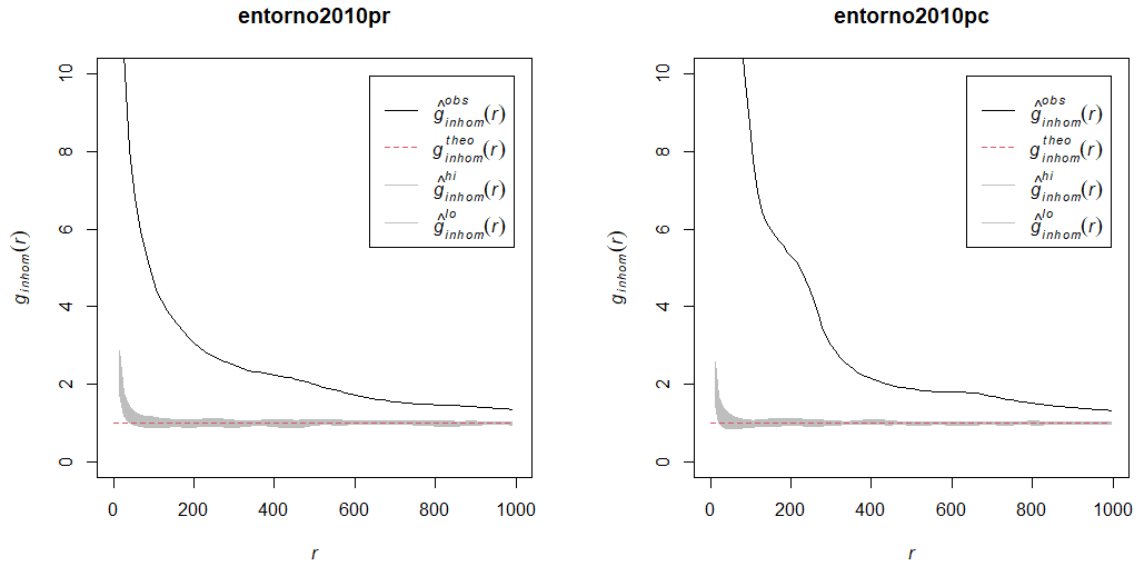


Figura A.6: Correlación de pares para las ratas y cucarachas avistadas en 2010

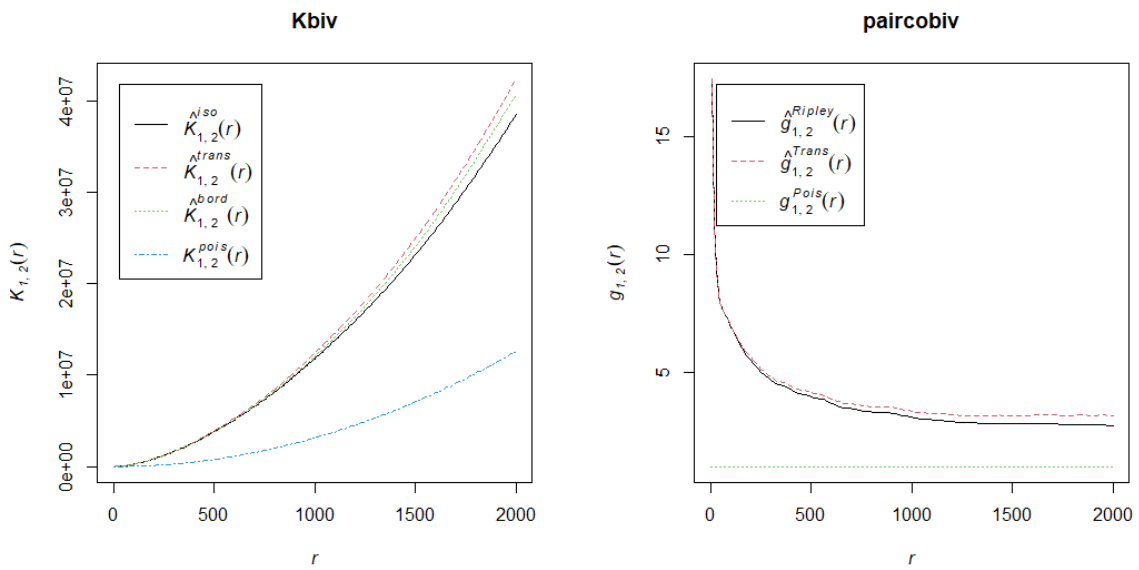


Figura A.7: K de Ripley y correlación de pares para el caso bivalente

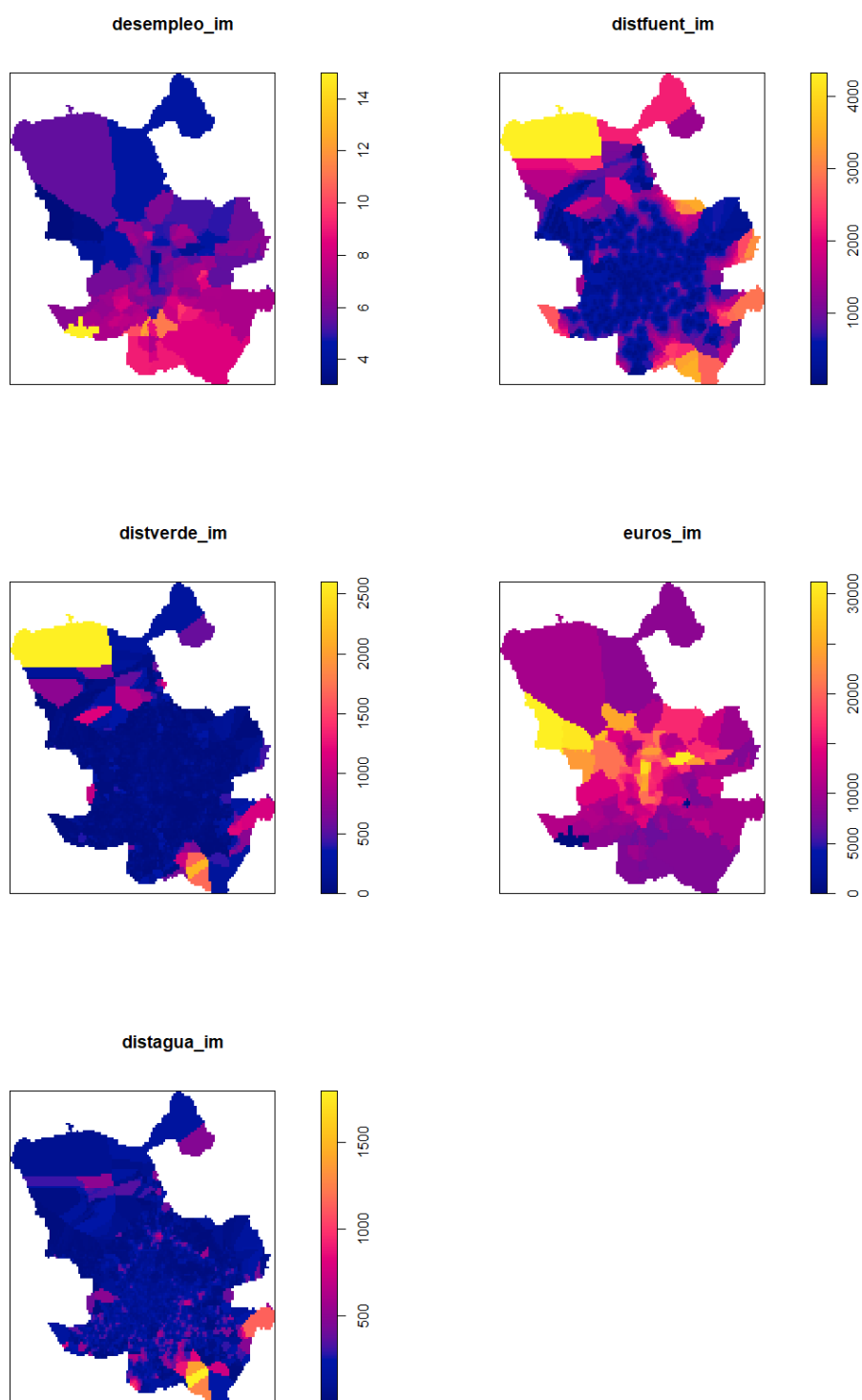


Figura A.8: Valores de las distintas covariables

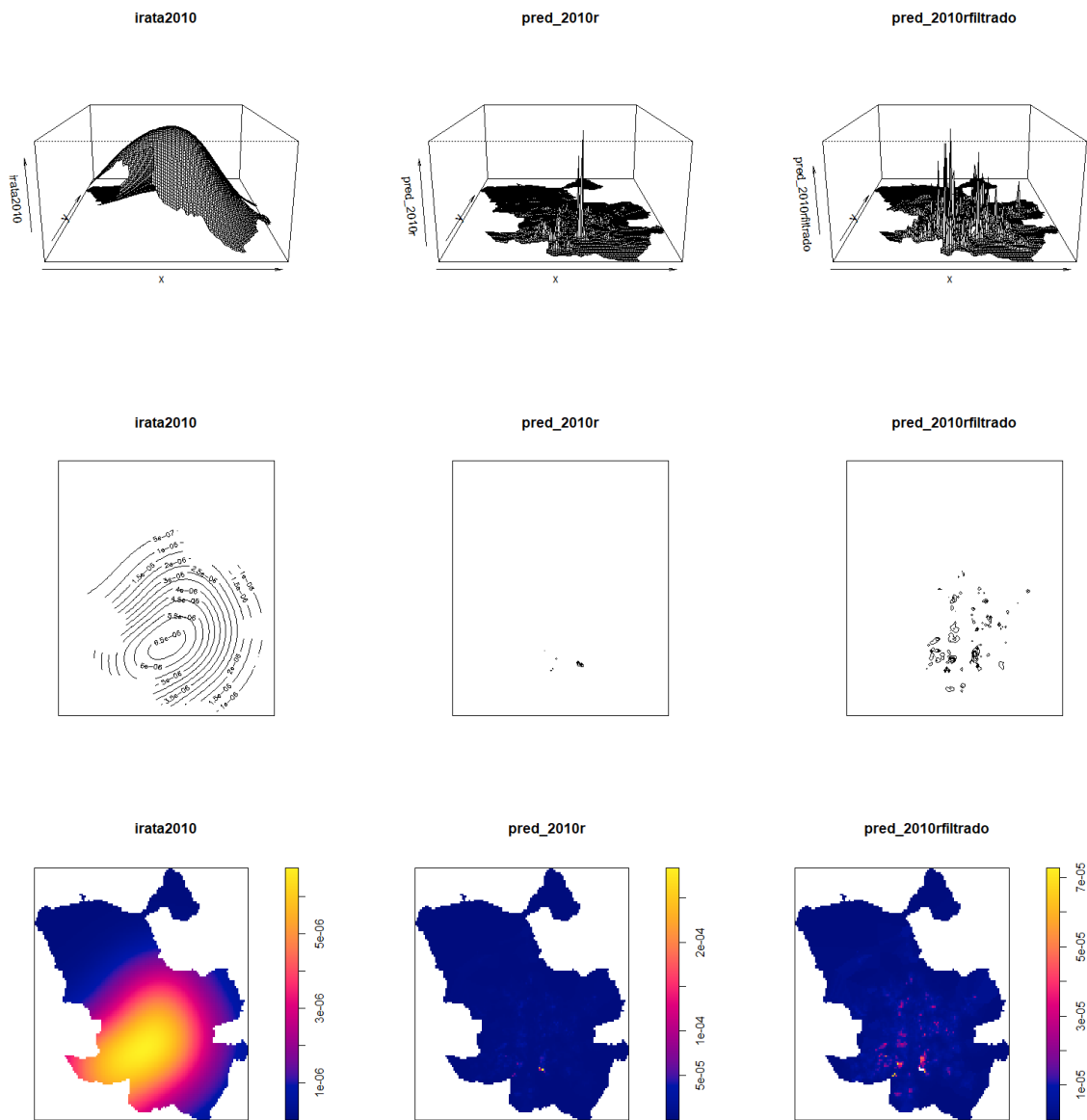


Figura A.9: Distintas representaciones de las intensidades de ratas

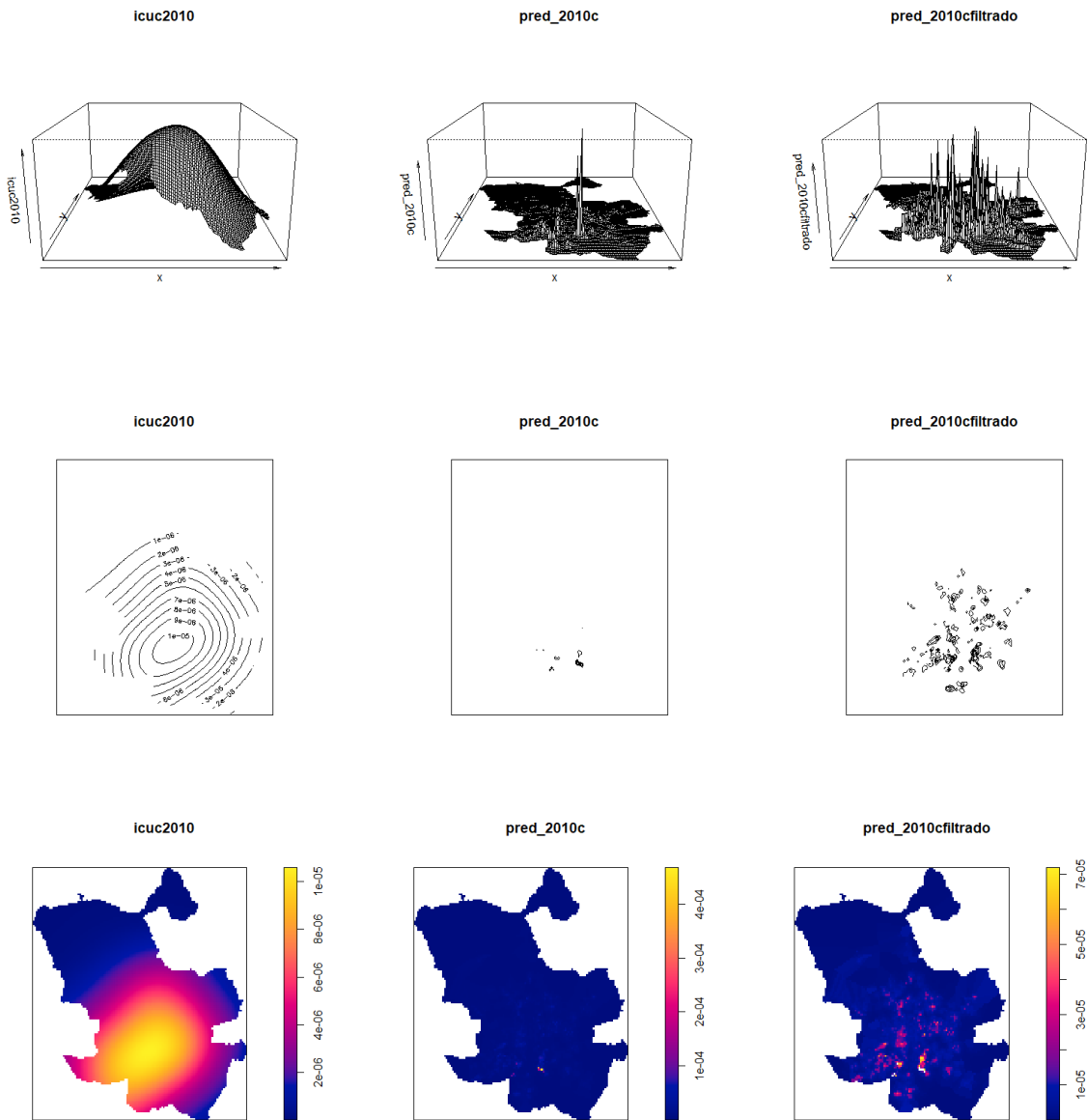


Figura A.10: Distintas representaciones de las intensidades de cucarachas

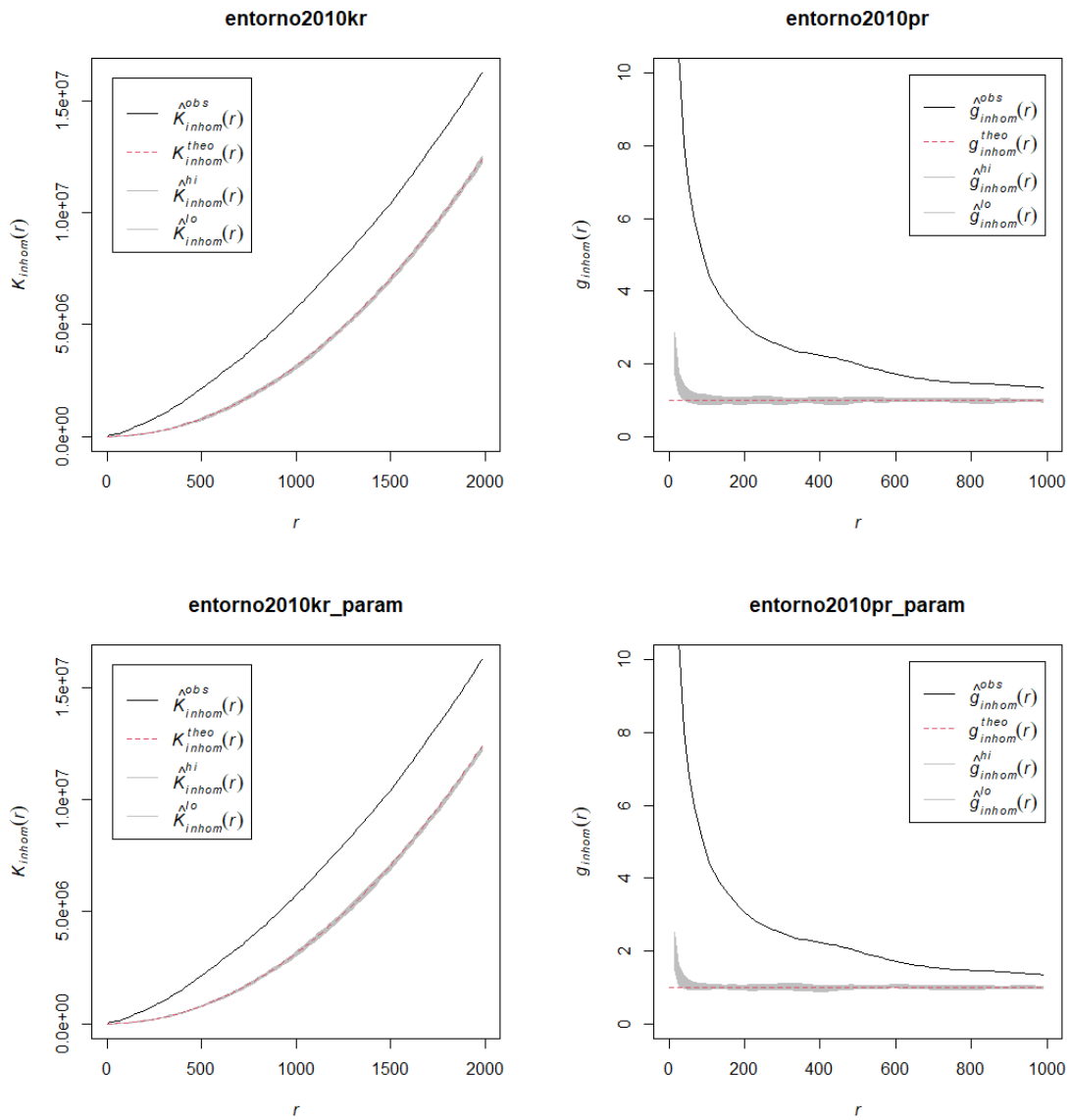


Figura A.11: K de Ripley y correlación de pares para las ratas avistadas en 2010

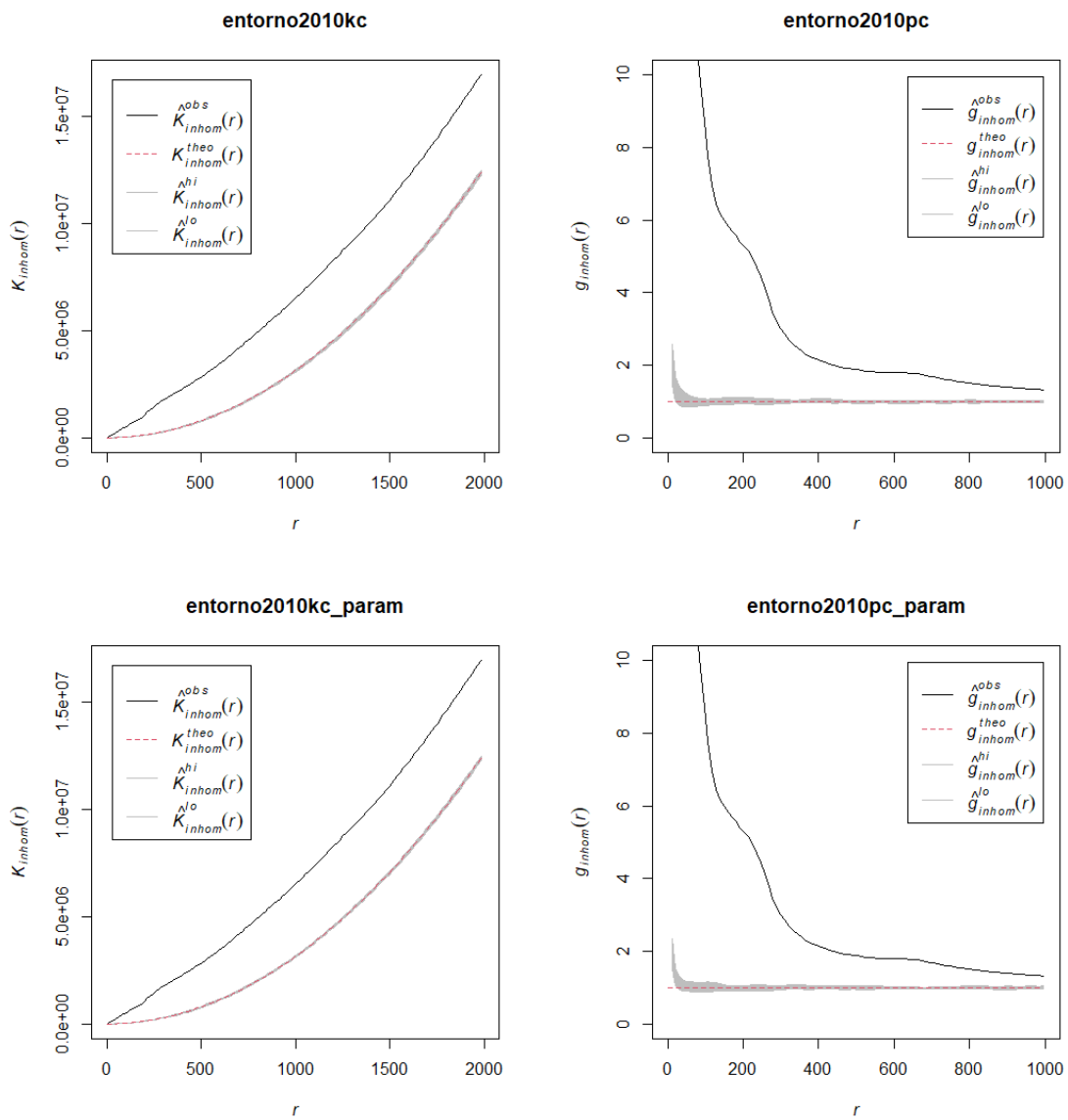


Figura A.12: K de Ripley y correlación de pares para las cucarachas avistadas en 2010

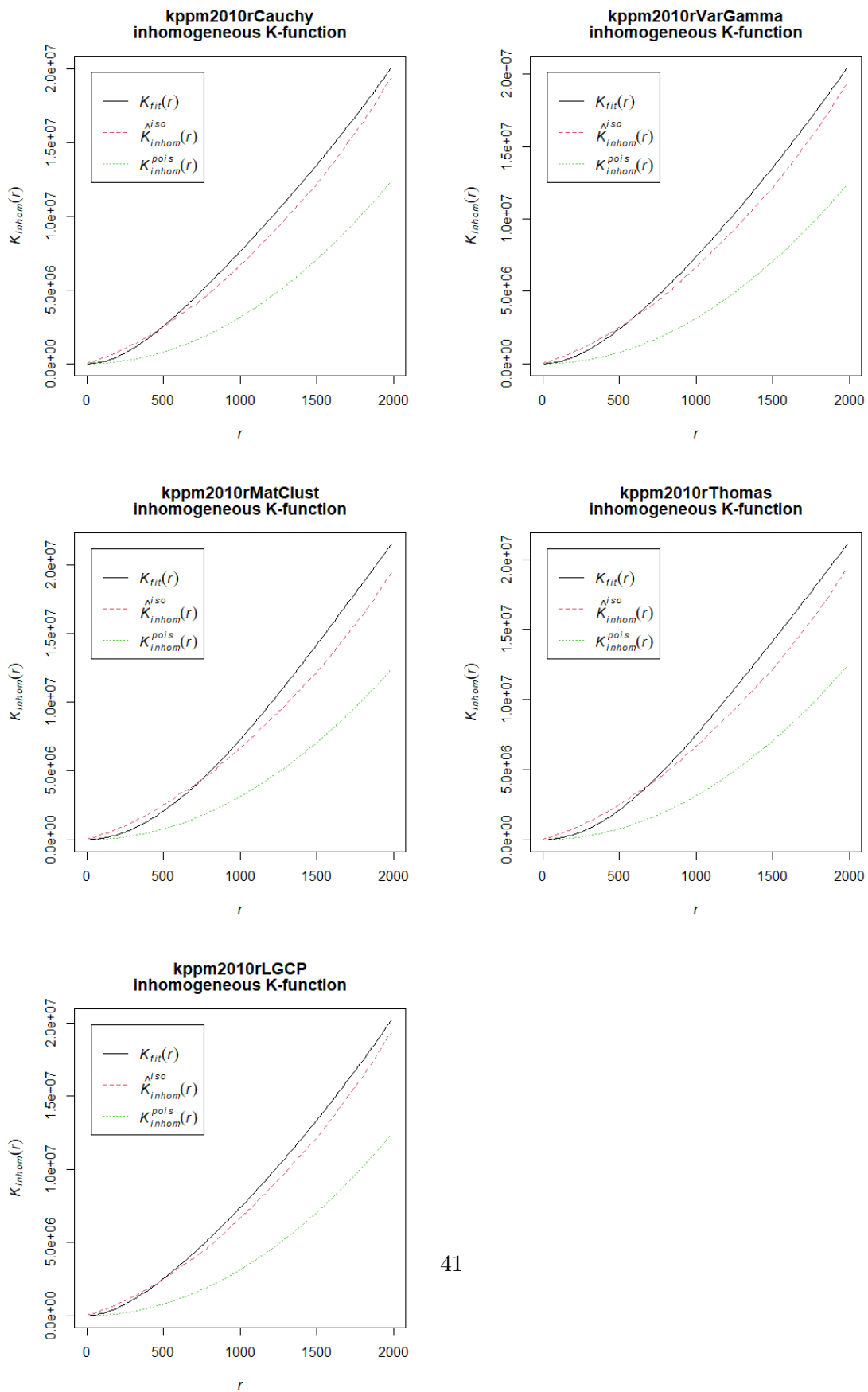


Figura A.13: K de Ripley de cada modelo de agrupamiento respecto a los datos

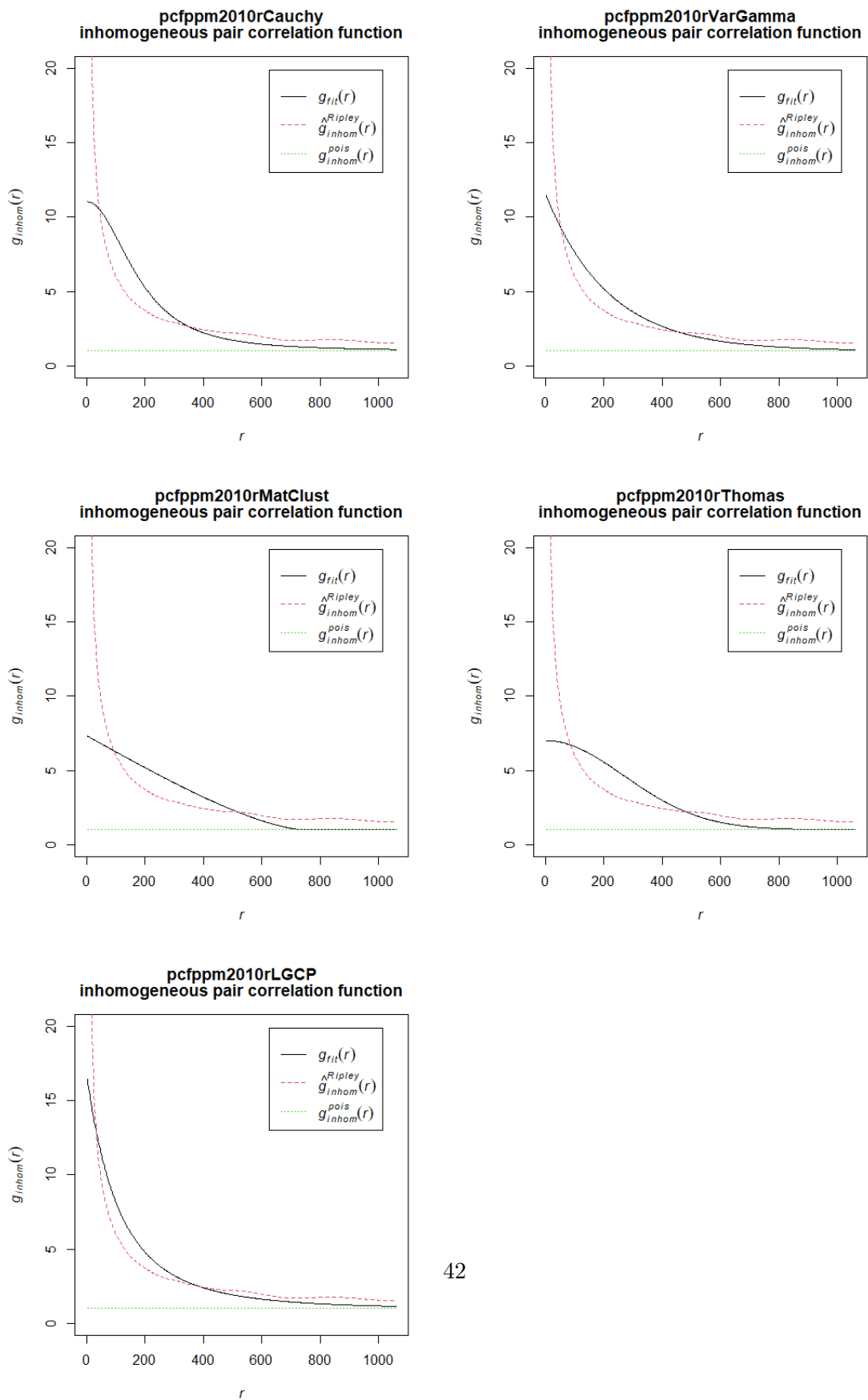


Figura A.14: Correlación de pares de cada modelo de agrupamiento respecto a los datos

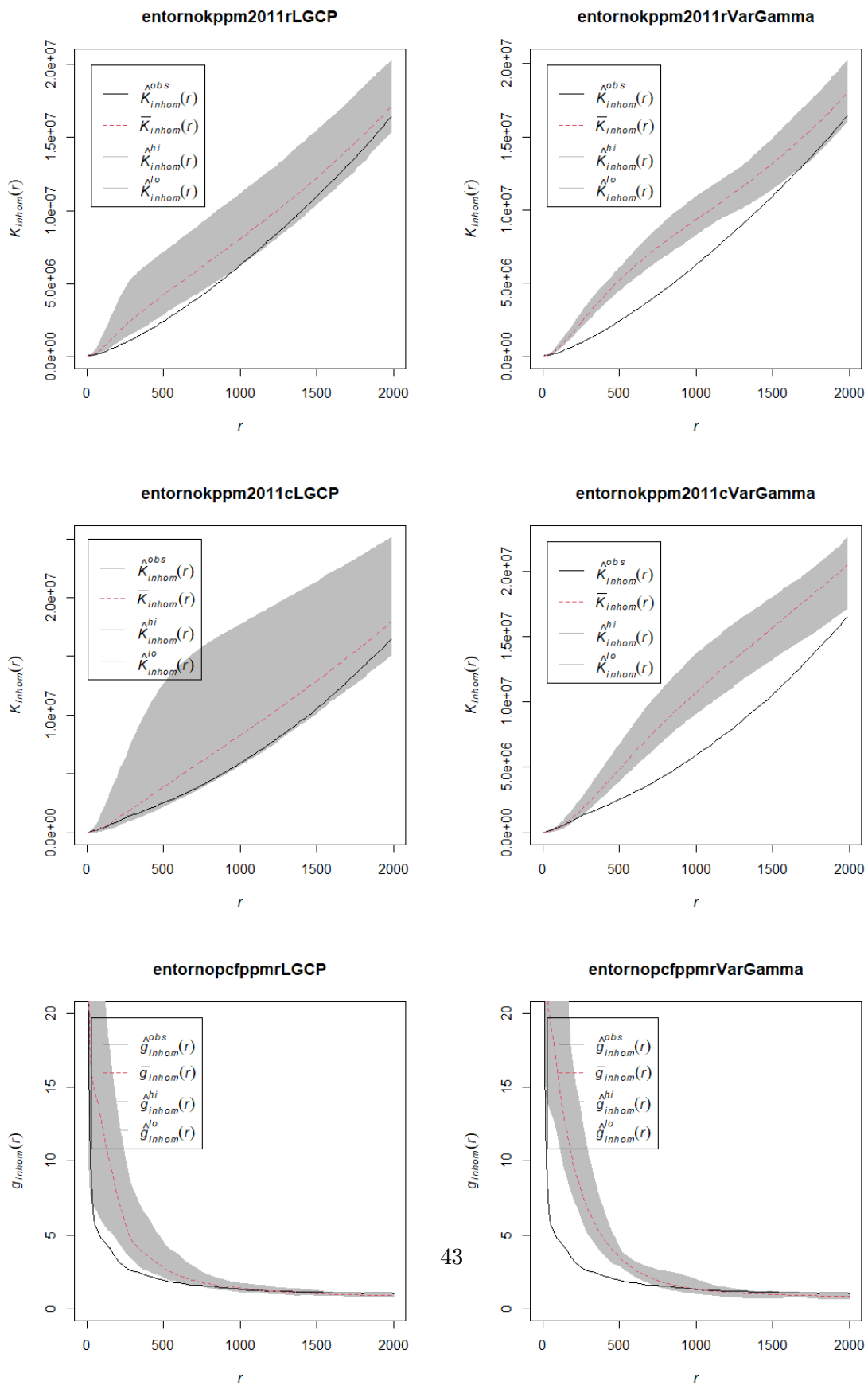


Figura A.15: Diferencias de las funciones entre el modelo VarGamma y el LGCP

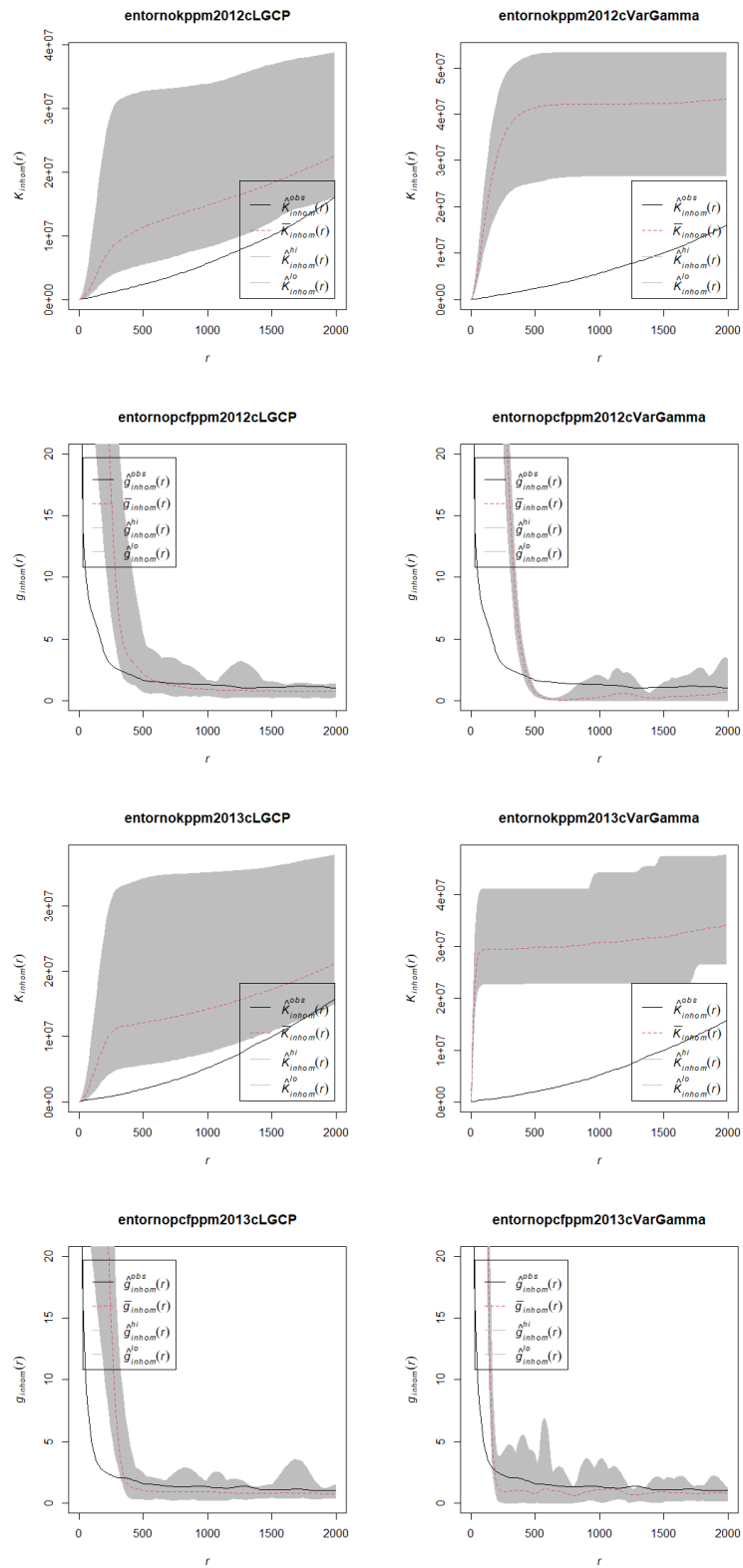


Figura A.16: Casos excepcionales de cucarachas en 2012 y 2013

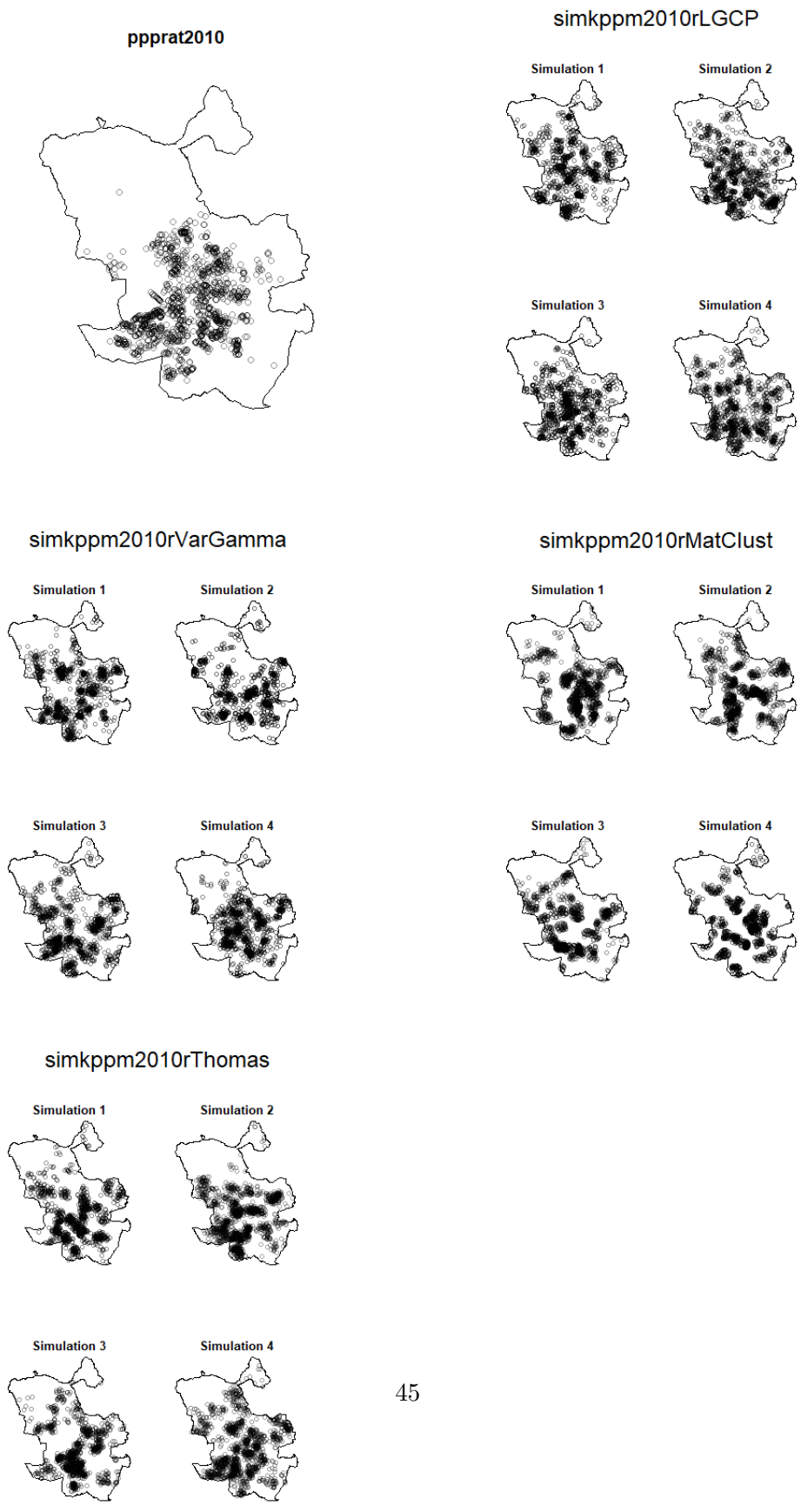


Figura A.17: Simulaciones de los datos iniciales usando distintos modelos