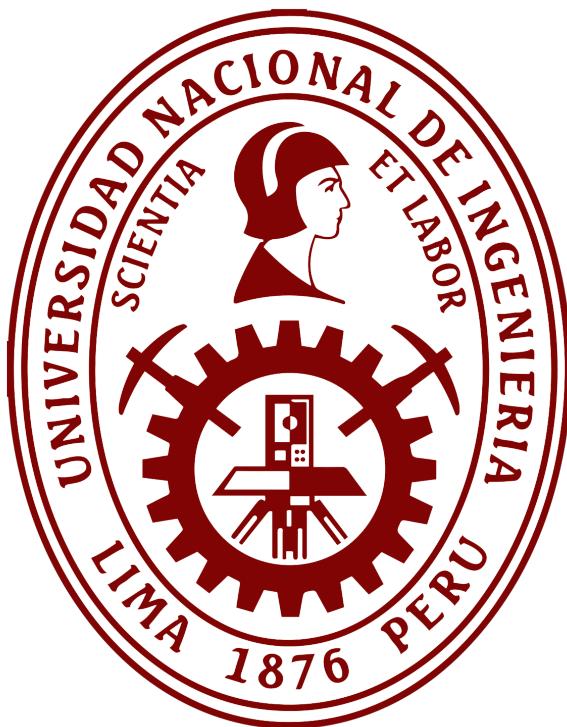


Universidad Nacional de Ingeniería

Escuela Profesional de Estadística

Análisis Espacial de la Distribución de Árboles Beilschmiedia mediante Procesos Puntuales en R



Autor:

Sofía Dextre Simangas

Curso: Estadística Espacial

Docente: Erick Chacón

Lima, Perú
Julio 2025

Índice

1. Resumen	2
2. Introducción	2
2.1. Justificación	2
2.2. Objetivo general	3
2.3. Objetivos específicos	3
3. Marco teórico	3
3.1. Conceptos generales	3
3.2. Evaluación de agrupamiento espacial	4
3.3. Estimación de intensidad	5
4. Metodología y Resultados	5
4.1. Base de datos	5
4.2. Software y herramientas utilizadas	6
4.3. Análisis exploratorio	7
4.4. Test de aleatoriedad	8
4.5. Estimación de la función de intensidad	11
4.6. Diagnóstico del modelo	13
4.7. Interacciones y transformaciones	13
4.7.1. Comparación de los 3 modelos	16
4.7.2. Comparación Log Likelihood	17
4.7.3. AIC	17
4.8. Simulación	17
5. Apéndices	19
.1. Paquetes y funciones utilizadas en R	19
A. Anexo: Código fuente en R	20

1. Resumen

Este trabajo realiza un análisis estadístico espacial de la distribución de árboles de la especie *Beilschmiedia pendula* en un bosque tropical ubicado en la Isla Barro Colorado, Panamá. Se utilizó la información contenida en los conjuntos de datos `bei` y `bei.extra`, los cuales registran la ubicación de los árboles y covariables como la elevación y el gradiente del terreno. Inicialmente, se efectuó un análisis exploratorio del patrón de puntos, así

como de las covariables espaciales. Posteriormente, se aplicaron pruebas de aleatoriedad espacial como la *función K de Ripley* y la *prueba del cuadrante*, las cuales revelaron la existencia de un patrón no completamente aleatorio, indicando inhomogeneidad espacial. Se estimó la función de intensidad del proceso puntual. Se evaluaron tres modelos: uno con

interacción entre covariables, otro con transformaciones funcionales, y uno que contiene ambos. La comparación de modelos se realizó mediante el log-verosimilitud y el criterio de información de Akaike (AIC), concluyéndose que el modelo completo presentó el mejor desempeño.

2. Introducción

Los procesos espaciales puntuales permiten modelar eventos que ocurren en ubicaciones específicas dentro de una región geográfica. Este enfoque es útil en disciplinas como la ecología o la epidemiología, donde la posición de los eventos revelan patrones que son útiles para estudios.

En este trabajo se analiza la distribución espacial de árboles de la especie *Beilschmiedia pendula* en un bosque tropical de la isla Barro Colorado, Panamá. Para ello, se utilizan herramientas estadísticas del paquete `spatstat` en R, con el fin de explorar la intensidad, la dependencia espacial y la relación con covariables ambientales como la elevación y la pendiente del terreno.

2.1. Justificación

El análisis de procesos puntuales espaciales permite modelar fenómenos donde la ubicación de los eventos es fundamental. En ecología, este enfoque es esencial para entender patrones de distribución, interacción, influencia de factores ambientales, etc.

Este estudio se justifica por la aplicación integral de conceptos teóricos como la intensidad, estacionariedad, isotropía, y herramientas como la función *K*, el test de aleatoriedad espacial (CSR), y la estimación de intensidad mediante kernels. Adicionalmente, se consideran modelos puntuales como los procesos de Poisson.

El uso del dataset `bei`, junto con el paquete `spatstat` en R, proporciona un entorno adecuado para aplicar los conceptos aprendidos en clase.

2.2. Objetivo general

Analizar el patrón espacial de los árboles de la especie *Beilschmiedia pendula* en un bosque tropical mediante técnicas de estadística espacial, aplicando modelos de procesos puntuales

2.3. Objetivos específicos

- Describir el patrón espacial de los árboles utilizando herramientas gráficas y estimación de intensidad.
- Evaluar la hipótesis de aleatoriedad espacial completa (CSR) mediante pruebas estadísticas.
- Analizar la dependencia espacial entre eventos utilizando funciones de segundo orden como la función K y la función de correlación de pares.
- Incorporar covariables ambientales en el modelo.

3. Marco teórico

3.1. Conceptos generales

Un proceso puntual espacial es un proceso estocástico cuyas realizaciones consisten en un conjunto de puntos ubicados en una región $D \subset \mathbb{R}^2$. Cada realización puede representarse como:

$$s = \{s_1, s_2, \dots, s_n\}, \quad s_i \in D$$

donde cada s_i indica la posición de un evento en el espacio. El conjunto D representa el dominio o región de observación, y suele ser una ventana rectangular o una región delimitada del plano.

Para el análisis de estos procesos, se utiliza la *medida de conteo* $N(A)$, definida como el número de puntos contenidos en una subregión $A \subset D$:

$$N(A) = \text{número de eventos en la región } A$$

A partir de esta medida, se definen funciones que permiten caracterizar la estructura espacial del patrón.

La función de intensidad de primer orden se define como:

$$\lambda(s) = \lim_{|ds| \rightarrow 0} \frac{\mathbb{E}[N(ds)]}{|ds|}$$

donde $\lambda(s)$ representa la densidad esperada de eventos por unidad de área alrededor del punto s . Si el proceso es estacionario, la intensidad es constante en todo el dominio: $\lambda(s) = \lambda$.

Estas herramientas permiten describir la forma en que los eventos se distribuyen en el espacio, ya sea de forma aleatoria, agrupada o inhibida. Este tipo de análisis es relevante en diversas disciplinas como ecología, epidemiología, geografía, entre otras.

3.2. Evaluación de agrupamiento espacial

El análisis del patrón espacial comienza con una evaluación de aleatoriedad espacial, usando herramientas como la función K de Ripley y el test de cuadrantes, que nos permitirán verificar si el proceso presenta aleatoriedad, agrupamiento o dispersión para posteriormente decidir qué modelo de intensidad utilizar.

La función $K(t)$ describe la estructura de segundo orden de un proceso puntual estacionario e isotrópico. Se define como:

$$K(t) = \frac{1}{\lambda} \mathbb{E}[N_0(t)]$$

donde:

- λ es la intensidad del proceso
- $N_0(t)$ es el número de eventos dentro de una distancia t

En el caso de un **proceso de Poisson homogéneo** (completa aleatoriedad espacial), se cumple que:

$$K(t) = \pi t^2$$

La comparación con la curva teórica πt^2 permite interpretar el patrón espacial:

- Si $\hat{K}(t) > \pi t^2$: hay **agrupamiento**.
- Si $\hat{K}(t) < \pi t^2$: hay **inhibición**
- Si $\hat{K}(t) \approx \pi t^2$: se acepta **aleatoriedad**

Test de Cuadrantes

Es una prueba de homogeneidad espacial que divide el área A en m subregiones (cuadrantes) y compara el conteo de eventos observados n_i con el valor esperado bajo aleatoriedad $e_i = \frac{n}{m}$.

El estadístico es:

$$X^2 = \sum_{i=1}^m \frac{(n_i - e_i)^2}{e_i}$$

Bajo la hipótesis nula H_0 (proceso homogéneo), se tiene:

$$X^2 \sim \chi_{m-1}^2$$

La decisión se basa en el valor p :

- Si $p < 0,05$, se rechaza H_0 : existe **variación espacial significativa**.
- Si $p \geq 0,05$, no se rechaza H_0 : el patrón es **compatible con aleatoriedad**.

3.3. Estimación de intensidad

La estimación de $\lambda(s)$ se vuelve necesaria cuando se ha descartado la hipótesis de aleatoriedad espacial (CSR), como ocurre cuando los resultados de pruebas como el test de cuadrantes o la función K de Ripley muestran agrupamiento o inhibición. En estos casos, se asume que $\lambda(s)$ no es constante, y se procede a estimarla.

Estimación No Paramétrica (Método Kernel)

Una forma común de estimar la intensidad es mediante suavizado kernel, donde la estimación en el punto s se calcula como:

$$\hat{\lambda}(s) = \sum_{i=1}^n \frac{1}{h^2} K\left(\frac{\|s - s_i\|}{h}\right)$$

donde:

- s_i son los puntos del patrón observado,
- $K(\cdot)$ es una función núcleo
- h es el ancho de banda (controla el grado de suavizado).

Este método produce un mapa continuo de intensidad, útil para detectar zonas de alta o baja concentración de eventos.

Estimación Paramétrica (Modelos de Poisson Inhomogéneo)

La intensidad puede modelarse como una función log-lineal:

$$\lambda(s) = \exp(\beta_0 + \beta_1 z_1(s) + \cdots + \beta_p z_p(s))$$

donde:

- $z_j(s)$ son las covariables
- β_j son coeficientes estimados

Si el proceso no es homogéneo se debe estimar $\lambda(s)$ usando métodos no paramétricos o modelos con covariables.

4. Metodología y Resultados

4.1. Base de datos

Este estudio utiliza las bases `bei` y `bei.extra` del paquete `spatstat.data` en R, que corresponden a un bosque tropical ubicado en la Isla Barro Colorado, Panamá.

La base `bei`, de clase `ppp`, contiene la ubicación de 3605 árboles de la especie Beilschmiedia pendula en una región rectangular de 1000×500 metros, con coordenadas en metros.

	x	y
1	11.7	151.1
2	998.9	430.5
3	980.1	433.5
4	986.5	425.8
5	944.1	415.1
6	940.5	410.4
7	950.9	405.7
8	940.8	389.2
9	945.5	379.8
10	930.5	287.1
11	931.2	372.6
12	924.2	427.8
13	936.4	453.0
14	918.3	445.4
15	915.3	409.8

La base `bei.extra` incluye covariables espaciales en formato de imagen:

- `elev`: elevación del terreno.
- `grad`: gradiente de elevación o pendiente.

	elev.x	elev.y	elev.value	elev.optional	grad.x	grad.y	grad.value	grad.optional
1	0	0	120.63	TRUE	0	0	0.25193120	TRUE
2	0	5	120.87	TRUE	0	5	0.20180990	TRUE
3	0	10	121.12	TRUE	0	10	0.16206560	TRUE
4	0	15	121.36	TRUE	0	15	0.16997370	TRUE
5	0	20	121.29	TRUE	0	20	0.20712470	TRUE
6	0	25	123.45	TRUE	0	25	0.23384130	TRUE
7	0	30	125.34	TRUE	0	30	0.20558680	TRUE
8	0	35	126.53	TRUE	0	35	0.16711140	TRUE
9	0	40	127.20	TRUE	0	40	0.15804560	TRUE
10	0	45	127.69	TRUE	0	45	0.14837400	TRUE
11	0	50	127.77	TRUE	0	50	0.15734200	TRUE
12	0	55	127.71	TRUE	0	55	0.15664740	TRUE
13	0	60	127.87	TRUE	0	60	0.15184100	TRUE
14	0	65	128.16	TRUE	0	65	0.16158140	TRUE
15	0	70	128.52	TRUE	0	70	0.18839640	TRUE

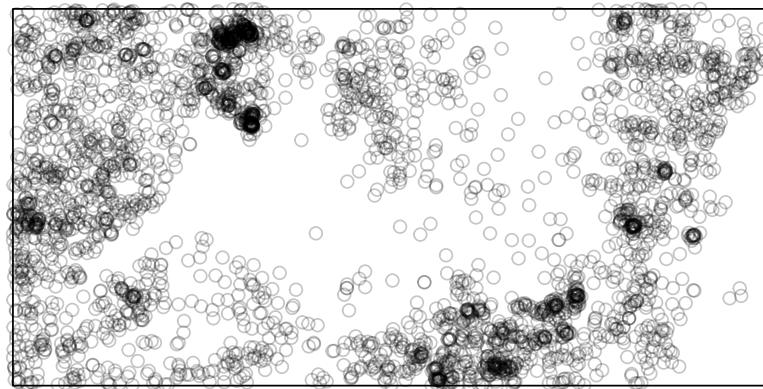
4.2. Software y herramientas utilizadas

Para el análisis estadístico de los procesos puntuales espaciales, se utilizó el software R junto con paquetes especializados del entorno `spatstat`. Estas herramientas permitieron representar, visualizar, modelar y analizar los patrones espaciales, así como ajustar modelos de intensidad y realizar inferencias.

En el Anexo .1, se detalla la lista completa de paquetes, funciones y clases utilizadas, junto con su propósito en el estudio.

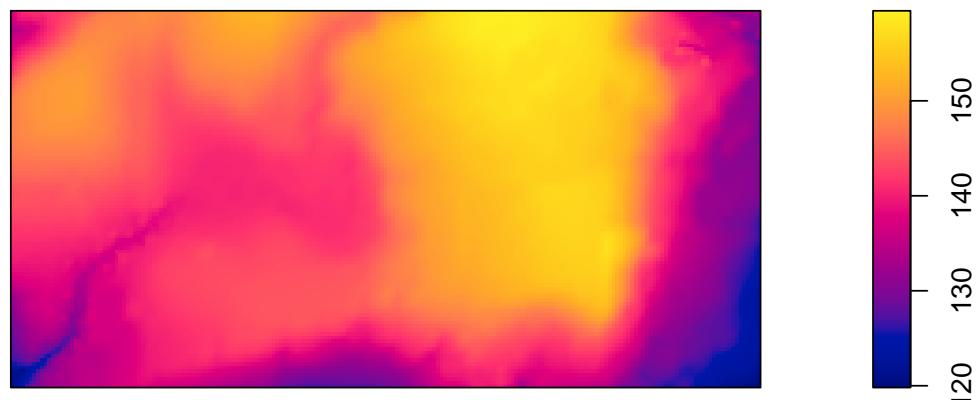
4.3. Análisis exploratorio

Distribución espacial de los árboles



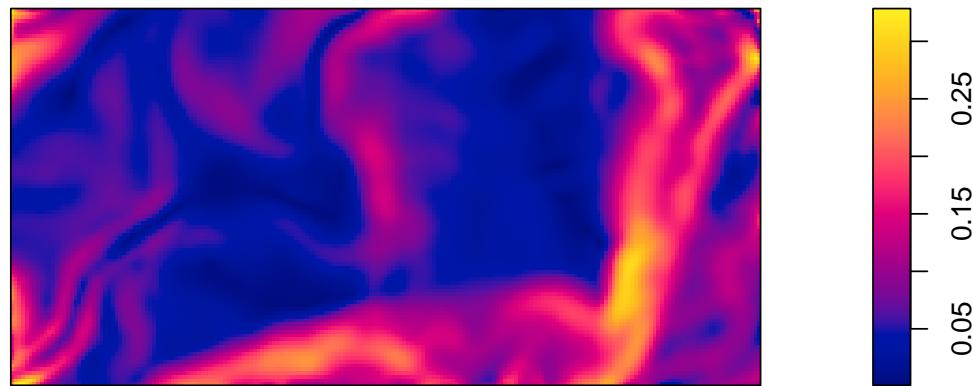
La distribución espacial de los árboles no es homogénea pues hay zonas con mayor agregación lo que sugiere clusters.

Elevación del terreno



La elevación del terreno está entre los 120 y 150 cm, además se ve que hay una zona elevada.

Gradiente de elevación (pendiente)



Este gráfico nos permite ver la pendiente de elevación, a menos pendiente el color es más oscuro, se puede ver que en general las pendientes son bajas al centro (lo que sugiere un área plana) y a los extremos hay mayor pendiente

Habiendo visto las distribuciones básicas nos preguntamos a qué tipo de proceso nos estamos enfrentando, para eso es necesario hacer los test de aleatoriedad.

4.4. Test de aleatoriedad

Test de Cuadrantes

El test de cuadrantes, divide el área de estudio en una grilla rectangular de subregiones, para este caso usaremos 5x5, es decir dividiremos la data en 25 regillas, es ideal tener un balance ya que con muy pocas celdas no se detecta variación y con muchas celdas la prueba puede ser inestable

```
> quadrat.test(bei, nx = 5, ny = 5)

Chi-squared test of CSR using quadrat counts

data: bei
X2 = 2009.9, df = 24, p-value < 2.2e-16
alternative hypothesis: two.sided

Quadrats: 5 by 5 grid of tiles
```

Hipótesis

- H_0 : El patrón espacial sigue un proceso de Poisson homogéneo (CSR).
- H_1 : El patrón espacial no es homogéneo; hay agrupamiento o inhibición.

$$X^2 = \sum_{i=1}^m \frac{(n_i - e_i)^2}{e_i}, \quad \text{donde } e_i = \frac{n}{m}$$

Resultado:

$$X^2 = 2009,9, \quad \text{gl} = 24, \quad p\text{-valor} < 2,2 \times 10^{-16}$$

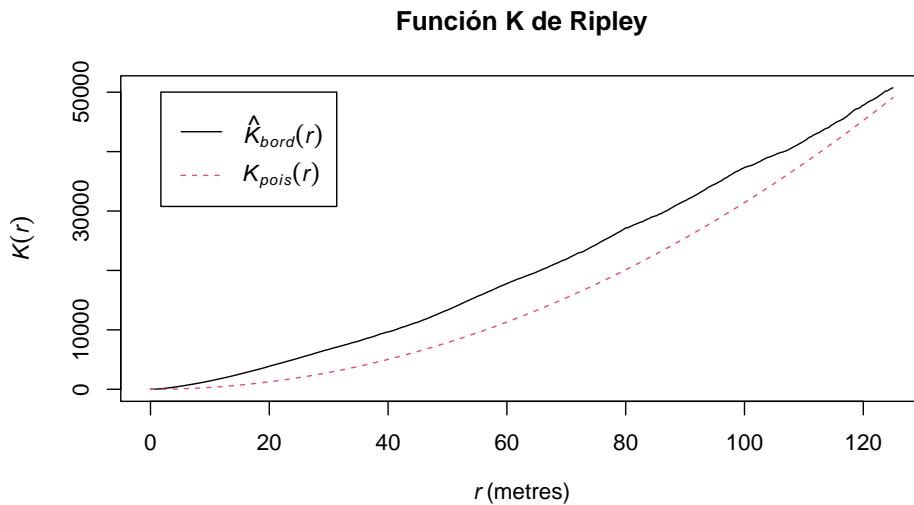
Dado que el valor del estadístico es alto y el valor p es mucho menor al nivel de significancia convencional ($\alpha = 0,05$), se rechaza la hipótesis nula por lo que el patrón no es homogéneo. Además obtenemos el cuadro

test_cuadrante

271	144.2	401	144.2	100	144.2	32	144.2	215	144.2
11		21		-3.7		-9.3		5.9	
180	144.2	217	144.2	104	144.2	34	144.2	149	144.2
3		6.1		-3.3		-9.2		0.4	
334	144.2	4	144.2	26	144.2	33	144.2	172	144.2
16		-12		-9.8		-9.3		2.3	
172	144.2	43	144.2	26	144.2	144	144.2	83	144.2
2.3		-8.4		-9.8		-0.013		-5.1	
146	144.2	89	144.2	234	144.2	338	144.2	57	144.2
0.15		-4.6		7.5		16		-7.3	

En la que cada celda representa el observado, esperado y residuo. Se pueden ver residuos muy altos sobre todo en los bordes, que indican más puntos de lo esperado.

Función K de Ripley



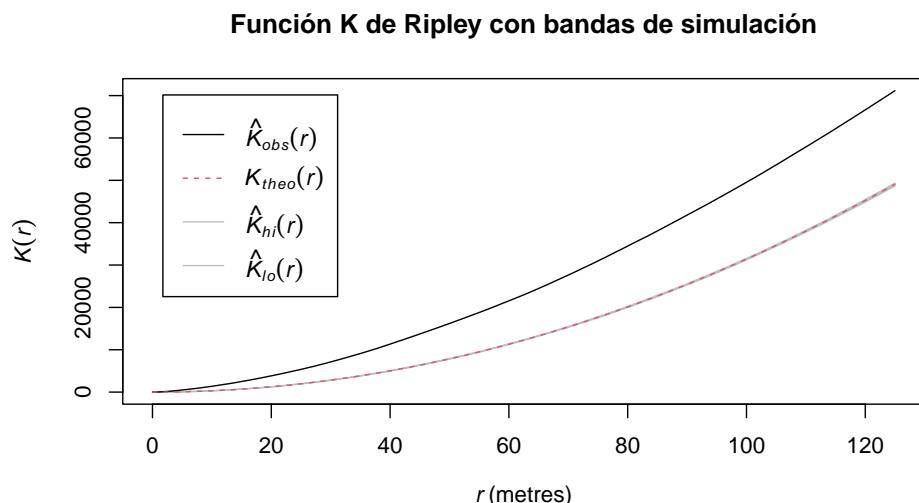
El gráfico de la función $K(r)$ muestra la acumulación esperada de puntos vecinos a distintas distancias r , con el fin de evaluar el patrón de puntos

- **Línea negra** : representa la estimación empírica $\hat{K}_{\text{bord}}(r)$, obtenida del patrón real de puntos.
- **Línea roja**: representa la función teórica $K_{\text{pois}}(r) = \pi r^2$, correspondiente al modelo de aleatoriedad espacial completa (CSR) bajo un proceso de Poisson homogéneo.

En el gráfico generado, la curva empírica $\hat{K}_{\text{bord}}(r)$ se encuentra por encima de πr^2 lo que indica que el patrón de puntos presenta agrupamiento espacial.

Agregando bandas de simulación

Además es necesario saber si dicha diferencia es significativa, esto lo podemos saber gracias a las bandas de simulación



El gráfico muestra la función $\hat{K}_{\text{obs}}(r)$ estimada a partir del patrón de puntos observado, comparada con:

- La función teórica $K_{\text{theo}}(r) = \pi r^2$ de un proceso homogéneo
- Una colección de funciones $\hat{K}_i(r)$ generadas a partir de simulaciones

La curva empírica $\hat{K}_{\text{obs}}(r)$ se encuentra por encima de las curvas generadas por las simulaciones. Esto indica que, para cada distancia r , el número de puntos vecinos es mayor al que se esperaría bajo un modelo de aleatoriedad completa.

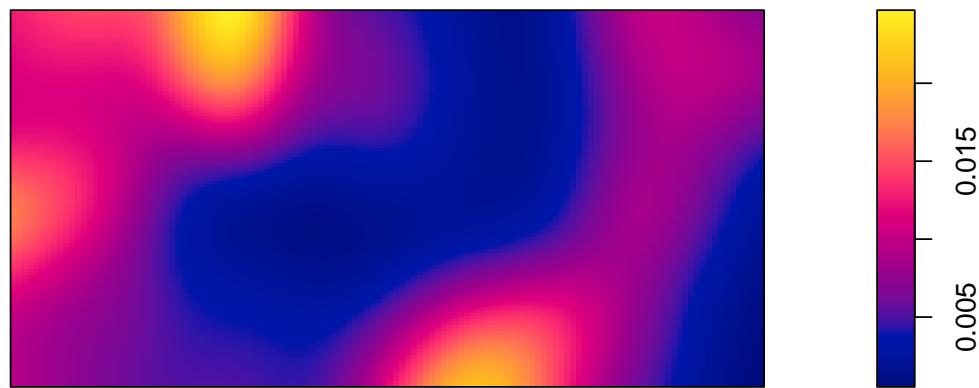
Dado que $\hat{K}_{\text{obs}}(r)$ se ubica fuera del rango de variabilidad generado por las simulaciones, se concluye que la diferencia es estadísticamente significativa.

4.5. Estimación de la función de intensidad

Método Kernel

La estimación de intensidad nos puede decir cómo varía la intensidad de eventos en el espacio

Intensidad estimada (kernel smoothing)



La estimación de la función de intensidad $\lambda(s)$ genera un mapa el cual evidencia una distribución claramente no homogénea como lo habíamos visto antes, en la que se identifican zonas con alta concentración de eventos y otras con baja o nula intensidad

Estimación paramétrica de intensidad

Se ajustó un modelo de proceso puntual de Poisson inhomogéneo con intensidad espacial $\lambda(s)$ modelada como una función log-lineal de dos covariables espaciales: la elevación del terreno (`elev`) y el gradiente o pendiente (`grad`).

Del cual obtendremos:

$$\log \lambda(s) = \beta_0 + \beta_1 \cdot \text{elev}(s) + \beta_2 \cdot \text{grad}(s)$$

Donde:

- $\lambda(s)$: intensidad del proceso en la ubicación espacial s ,
- β_0 : intercepto del modelo,
- β_1, β_2 : coeficientes asociados a las covariables.

Realizando la estimación:

```
Fitted trend coefficients:
(Intercept)      elev      grad
-8.56355220  0.02143995  5.84646680

              Estimate      S.E.    CI95.lo    CI95.hi   ztest   Zval
(Intercept) -8.56355220 0.341113849 -9.23212306 -7.89498134 *** -25.104675
elev        0.02143995 0.002287866  0.01695581  0.02592408 ***  9.371155
grad        5.84646680 0.255781018  5.34514522  6.34778838 *** 22.857313

----- gory details -----

Fitted regular parameters (theta):
(Intercept)      elev      grad
-8.56355220  0.02143995  5.84646680

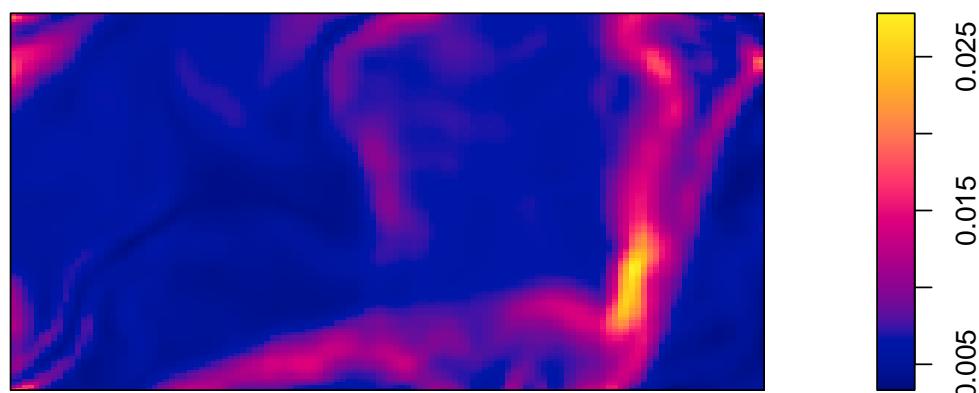
Fitted exp(theta):
(Intercept)      elev      grad
1.909398e-04 1.021671e+00 3.460097e+02
```

- El coeficiente $\hat{\beta}_1 = 0,0214$ indica que, por cada unidad adicional de elevación, la intensidad esperada del proceso aumenta en aproximadamente un 2,17% (ya que $e^{0,0214} \approx 1,0217$).
- El coeficiente $\hat{\beta}_2 = 5,8465$ indica que por cada unidad adicional de pendiente, la intensidad se multiplica por un factor de aproximadamente 346 (ya que $e^{5,8465} \approx 346$).
- Ambos coeficientes son estadísticamente significativos con valores $p < 0,001$.

En particular, la densidad de eventos (árboles) tiende a ser mayor en zonas con mayor elevación y, especialmente, con mayor gradiente del terreno.

Podemos visualizar la intensidad esperada revisando el gráfico del cálculo de los valores esperados en cada punto a partir del modelo

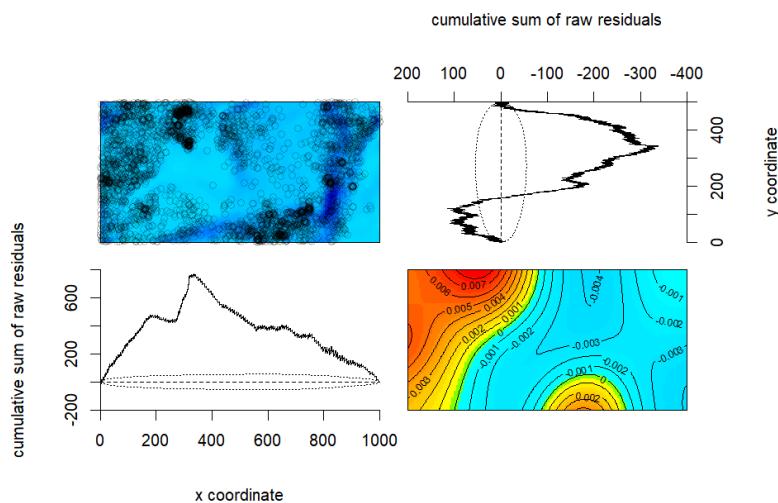
Intensidad estimada – modelo paramétrico



Por lo visto, el modelo espera más árboles en cierta zona (la parte más amarilla) además es más frecuente ver mayor intensidad en los bordes (donde hay pendiente) que en el área plana.

4.6. Diagnóstico del modelo

Podemos hacer un diagnóstico de nuestro modelo de intensidad previamente calculado con `diagnose.ppm()`, este gráfico evalúa si el modelo explica bien los datos.



- **Mapa de residuos crudos:** muestra los residuos puntuales superpuestos al patrón de árboles. Se observan zonas con residuos más intensos (más oscuros), lo cual indica que el modelo no ajusta igualmente bien en todo el espacio.
- **Suma acumulada de residuos en y :** la línea negra se desvía de las bandas de confianza, lo que indica que el modelo no explica adecuadamente la variación vertical.
- **Suma acumulada de residuos en x :** presenta una desviación considerable fuera de las bandas de confianza, sugiriendo que el modelo presenta mal ajuste en la dirección horizontal.
- **Mapa de residuos suavizados:** muestra las regiones donde el modelo sobreestima (áreas azules) o subestima (áreas rojas) la intensidad. El modelo tiende a subestimar la presencia de árboles en ciertas zonas y a sobreestimarla en otras.

El modelo ajustado con las covariables elevación y gradiente captura parcialmente el patrón espacial

4.7. Interacciones y transformaciones

Por lo visto anteriormente, nuestro modelo no predice del todo, por lo que sería bueno probar transformaciones e interacciones.

El uso de interacciones permite modelar efectos conjuntos entre covariables, es decir, capturar situaciones en las que el efecto de una covariante sobre la intensidad depende del nivel de otra.

Por otro lado, las transformaciones no lineales permiten ajustar relaciones que no siguen una forma lineal simple.

Modelo con interacciones

$$\log(\lambda(s)) = \beta_0 + \beta_1 \cdot \text{elev}(s) + \beta_2 \cdot \text{grad}(s) + \beta_3 \cdot \text{elev}(s) \cdot \text{grad}(s)$$

```
Fitted trend coefficients:
(Intercept)          elev        grad    elev:grad
-4.403426761 -0.007024363 -36.500458701  0.292813497

              Estimate      S.E.      CI95.lo      CI95.hi Ztest     Zval
(Intercept) -4.403426761 0.614763210 -5.60834051 -3.198513010 *** -7.162801
elev        -0.007024363 0.004192219 -0.01524096  0.001192235   -1.675572
grad        -36.500458701 5.261456400 -46.81272375 -26.188193651 *** -6.937330
elev:grad    0.292813497 0.036257125  0.22175084  0.363876155 ***  8.076026

----- gory details -----

Fitted regular parameters (theta):
(Intercept)          elev        grad    elev:grad
-4.403426761 -0.007024363 -36.500458701  0.292813497

Fitted exp(theta):
(Intercept)          elev        grad    elev:grad
1.223534e-02 9.930002e-01 1.406217e-16 1.340193e+00
```

- El coeficiente de `elev` es negativo pero no significativo ($p > 0,05$), sugiriendo un efecto débil o inexistente por sí solo.
- `grad` tiene un efecto negativo significativo, indicando que áreas con mayor pendiente presentan menor intensidad de eventos.
- La interacción `elev:grad` es positiva y altamente significativa ($p < 0,001$), lo cual sugiere que el efecto negativo de la pendiente se atenúa en zonas de mayor elevación.

Estos resultados apoyan el uso de un modelo con interacción

Modelo con transformaciones

$$\log(\lambda(s)) = \beta_0 + \beta_1 \cdot \text{elev}^2(s) + \beta_2 \cdot \sqrt{\text{grad}(s)}$$

```
Fitted trend coefficients:
(Intercept) I(elev^2) sqrt(grad)
-7.917083e+00 8.454618e-05 4.242841e+00

             Estimate      S.E.    CI95.lo    CI95.hi   Ztest    zval
(Intercept) -7.917083e+00 1.953440e-01 -8.299950e+00 -7.5342157393 *** -40.52894
I(elev^2)    8.454618e-05 8.088189e-06  6.869362e-05  0.0001003987 *** 10.45304
sqrt(grad)   4.242841e+00 1.691571e-01  3.911299e+00  4.5743826271 *** 25.08226

----- gory details -----

Fitted regular parameters (theta):
(Intercept) I(elev^2) sqrt(grad)
-7.917083e+00 8.454618e-05 4.242841e+00

Fitted exp(theta):
(Intercept) I(elev^2) sqrt(grad)
0.000364464 1.000084550 69.605312362
```

- La covariable `elev` fue transformada al cuadrado. Su coeficiente positivo y significativo sugiere que la intensidad de eventos aumenta con la elevación a tasas crecientes.
- El gradiente `grad` fue transformado mediante raíz cuadrada. Su coeficiente es grande y positivo, indicando que zonas con mayor pendiente tienden a concentrar más eventos.
- Todos los términos fueron estadísticamente significativos ($p < 0,001$), lo que respalda la inclusión de estas transformaciones.

El modelo capta relaciones no lineales entre la intensidad del proceso y las covariables topográficas del terreno, proporcionando un ajuste más flexible que el modelo lineal simple.

Modelo con interacciones y transformaciones

$$\log(\lambda(s)) = \beta_0 + \beta_1 \cdot \text{elev}^2(s) + \beta_2 \cdot \sqrt{\text{grad}(s)} + \beta_3 \cdot \text{elev}^2(s) \cdot \sqrt{\text{grad}(s)}$$

```
Fitted trend coefficients:
(Intercept) I(elev^2) sqrt(grad)
-7.917083e+00 8.454618e-05 4.242841e+00

             Estimate      S.E.    CI95.lo    CI95.hi   Ztest    zval
(Intercept) -7.917083e+00 1.953440e-01 -8.299950e+00 -7.5342157393 *** -40.52894
I(elev^2)    8.454618e-05 8.088189e-06  6.869362e-05  0.0001003987 *** 10.45304
sqrt(grad)   4.242841e+00 1.691571e-01  3.911299e+00  4.5743826271 *** 25.08226

----- gory details -----

Fitted regular parameters (theta):
(Intercept) I(elev^2) sqrt(grad)
-7.917083e+00 8.454618e-05 4.242841e+00

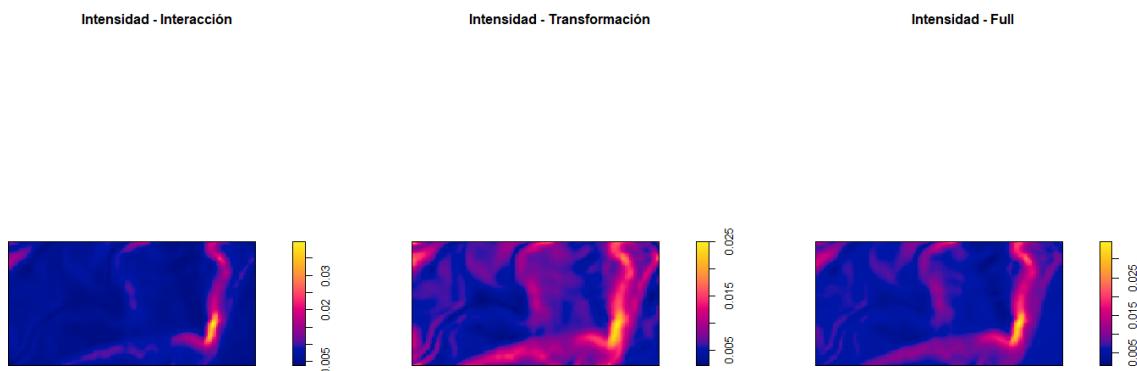
Fitted exp(theta):
(Intercept) I(elev^2) sqrt(grad)
0.000364464 1.000084550 69.605312362
```

- **Elevación al cuadrado (elev²)**: Tiene un efecto negativo; por sí sola, una mayor elevación disminuye la intensidad.
- **Raíz cuadrada del gradiente (grad)**: También tiene un efecto negativo individual sobre la intensidad.
- **Interacción elev² × grad**: Presenta un efecto positivo, indicando que cuando ambas covariables son altas simultáneamente, la intensidad puede incrementarse, compensando sus efectos individuales.

Todos los coeficientes fueron estadísticamente significativos ($p < 0,001$), lo que sugiere que la inclusión de esta interacción y transformaciones no lineales proporciona un ajuste más realista del proceso espacial.

4.7.1. Comparación de los 3 modelos

A continuación, se presentan los mapas de intensidad estimada bajo tres modelos distintos: interacción, transformación y modelo completo (transformación + interacción). Se comparan visualmente para evaluar cuál representa mejor el patrón espacial.



- **Modelo con interacción**: Muestra una concentración marcada de alta intensidad (zonas en color amarillo) principalmente en el borde derecho inferior.
- **Modelo con transformación**: Exhibe un patrón más generalizado, capturando relaciones no lineales entre las covariables y la intensidad.
- **Modelo completo (transformación + interacción)**: Combina ambos efectos, mostrando una mayor precisión en la detección de zonas con alta intensidad esperada. Se observan patrones más definidos que coinciden mejor con la distribución real de los puntos.

4.7.2. Comparación Log Likelihood

A continuación, se presentan los valores del logaritmo de la verosimilitud (logLik) para los tres modelos ajustados:

```
> logLik(modelo_int)
'log Lik.' -21110.62 (df=4)
> logLik(modelo_transf)
'log Lik.' -21078.36 (df=3)
> logLik(modelo_full)
'log Lik.' -21037.23 (df=4)
```

Modelo	logLik	Grados de libertad (df)
Modelo con interacción	-21110.62	4
Modelo con transformación	-21078.36	3
Modelo completo (transformación + interacción)	-21037.23	4

Se observa que el **modelo completo** presenta el mayor valor de log-verosimilitud (es decir, el menos negativo), lo cual indica un mejor ajuste a los datos observados.

4.7.3. AIC

Para comparar los modelos ajustados, se utilizó el criterio de información de Akaike (AIC), cuyos resultados se detallan a continuación:

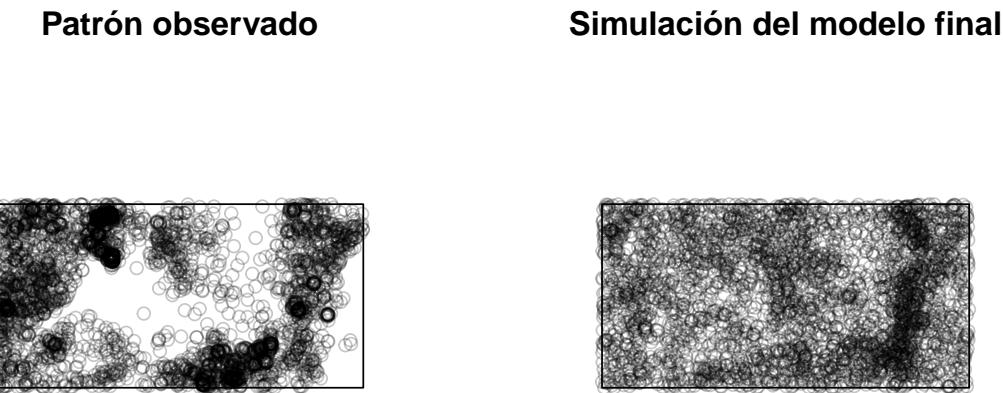
- Modelo con interacción: AIC = 42229.24
- Modelo con transformación: AIC = 42162.72
- Modelo completo (interacción + transformación): AIC = 42082.46

Dado que el modelo completo presenta el menor valor de AIC, se concluye que este ofrece el mejor balance entre ajuste y complejidad. Por lo tanto, es el modelo preferido para describir la intensidad del patrón de puntos en función de las covariables topográficas.

4.8. Simulación

Una vez ajustado el modelo de Poisson inhomogéneo con transformaciones cuadráticas e interacciones, se realizó una simulación del proceso puntual utilizando la función `simulate()` de `spatstat`. Esta función genera una realización aleatoria del patrón espacial según la intensidad estimada por el modelo ajustado. El objetivo de esta simulación es ver si el

modelo es capaz de replicar estructuras espaciales similares a las observadas en los datos.



Se observa que el modelo logra capturar la variación global de la intensidad en el espacio, reproduciendo regiones con mayor densidad de puntos y otras más dispersas. Sin embargo, la simulación presenta una distribución más suave, mientras que el patrón real contiene zonas de vacíos y agrupamientos más marcados.

Conclusiones

El presente trabajo tuvo como objetivo analizar el patrón espacial de árboles en una región geográfica utilizando procesos puntuales y estimar la función de intensidad asociada a covariables topográficas. Para ello, se emplearon herramientas estadísticas del marco de los procesos puntuales espaciales.

En primer lugar, los análisis exploratorios y las pruebas de aleatoriedad (como la función K de Ripley y el test de cuadrantes) permitieron concluir que el patrón de puntos no sigue un proceso de Poisson homogéneo, evidenciando agrupamientos y variación espacial significativa en la intensidad del proceso. Esto justificó la necesidad de un modelo de Poisson inhomogéneo.

Posteriormente, se realizaron estimaciones no paramétricas de la función de intensidad mediante *kernel smoothing*, las cuales permitieron visualizar zonas de alta y baja intensidad y luego, se ajustaron modelos paramétricos de Poisson inhomogéneo incorporando covariables como la elevación y el gradiente del terreno, considerando además transformaciones e interacciones. El modelo final seleccionado incluyó una transformación cuadrática de la elevación e interacción con la raíz cuadrada del gradiente, mostrando los mejores indicadores.

Al comparar la simulación de patrones con la data real se ve que modelo logra capturar adecuadamente la variación espacial observada.

En conclusión, se comprobó que variables topográficas como la elevación y el gradiente influyen significativamente en la distribución espacial de los árboles. El modelo ajustado no solo permite explicar dicha distribución, sino que también podría utilizarse para predecir zonas con mayor densidad esperada en escenarios similares.

5. Apéndices

.1. Paquetes y funciones utilizadas en R

Tabla 1: Paquetes y funciones utilizadas en R

Paquete	Función/Clase: para qué sirve
<code>spatstat.geom</code>	<code>ppp</code> : Representa un patrón puntual espacial; se usa para modelar la localización de los árboles. <code>owin</code> : Define la ventana de observación espacial; establece los límites espaciales del área de estudio. <code>im</code> : Representa imágenes espaciales (pixeladas); se usa con covariables como elevación y gradiente.
<code>spatstat.core</code>	<code>Kest</code> : Estima la función $K(r)$ de Ripley. <code>pcf</code> : Calcula la función de correlación por pares $g(r)$. <code>density.ppp</code> : Estima la intensidad mediante el método de núcleo (kernel). <code>envelope</code> : Genera envolventes simuladas (Monte Carlo) para evaluar hipótesis como el HPP.
<code>spatstat.explore</code>	<code>quadrat.test</code> : Aplica la prueba del cuadrante (Chi-cuadrado). <code>Gest</code> : Estima la función $G(r)$ (vecino más cercano). <code>Fest</code> : Estima la función $F(r)$ (punto aleatorio al evento más cercano). <code>Jest</code> : Estima la función $J(r)$ combinando F y G .
<code>spatstat.model</code>	<code>ppm</code> : Ajusta modelos de Poisson no homogéneos con covariables. <code>predict</code> : Realiza predicciones de la intensidad estimada. <code>diagnose.ppm</code> : Evalúa gráficamente el ajuste del modelo.
<code>spatstat.data</code>	<code>bei, bei.extra</code> : Datos de localización de árboles y covariables espaciales como elevación y gradiente.
<code>maptools</code>	<code>as.owin, readShapePoly</code> : Conversión/importación de datos espaciales (shapefiles).

A. Anexo: Código fuente en R

```

1 # -----
2 # Curso: Estadística Espacial
3 # Alumna: Sofía Dextre
4 # Descripción: Este trabajo es para la PC3 del curso, contiene el
5 # análisis
6 # puntual espacial de la data bei y bei.extra
7 #
8 # ----- LIBRERIAS -----
9 install.packages("spatstat.data")
10 install.packages("spatstat.geom")
11 install.packages("spatstat.explore")
12 install.packages("spatstat.model")
13
14 library(spatstat.data)
15 library(spatstat.geom)
16 library(spatstat.explore)
17 library(spatstat.model)
18
19 # ----- LEER DATA -----
20 data(bei)
21
22 bei.extra <- spatstat.data::bei.extra
23 str(bei.extra)
24
25 # ----- EXPLORACIÓN INICIAL: DISTRIBUCIONES -----
26
27 plot(bei, main = "Distribución espacial de los rboles ")
28 plot(bei.extra$elev, main = "Elevación del terreno")
29 plot(bei.extra$grad, main = "Gradiente de elevación (pendiente)")
30
31 # ----- TEST DE ALEATORIEDAD ESPACIAL -----
32 test_cuadrante <- quadrat.test(bei, nx = 5, ny = 5)
33 print(test_cuadrante)
34 plot(test_cuadrante)
35
36 K <- Kest(bei)
37 plot(K, main = "Función K de Ripley")
38
39 set.seed(123)
40 K_env <- envelope(bei, Kest, nsim = 99)
41 plot(K_env, main = "Función K de Ripley con bandas de simulación")
42
43 #----- ESTIMACIÓN DE LA INTENSIDAD -----
44 lambda_kernel <- density(bei)
45 plot(lambda_kernel, main = "Intensidad estimada (kernel smoothing)")
46
47 modelo_log <- ppm(bei ~ elev + grad, covariates = bei.extra)
48 summary(modelo_log)
49 plot(predict(modelo_log), main = "Intensidad estimada - modelo
50     paramétrico")
51
52 #----- DIAGNÓSTICO DEL MODELO -----
53 diagnose.ppm(modelo_log)
54 Kres <- Kres(modelo_log)

```

```

54 env_Kres <- envelope(modelo_log, Kres)
55 plot(env_Kres)
56
57 modelo_1 <- ppm(bei ~ elev, covariates = bei.extra)
58 modelo_2 <- ppm(bei ~ grad, covariates = bei.extra)
59 anova(modelo_1, modelo_log, test = "Chi")
60 simulacion <- rpoispp(modelo_log)
61 plot(simulacion)
62
63 # ----- INTERACCIONES Y TRANSFORMACIONES -----
64 modelo_int <- ppm(bei ~ elev * grad, covariates = bei.extra)
65 summary(modelo_int)
66
67 modelo_transf <- ppm(bei ~ I(elev^2) + sqrt(grad), covariates = bei.
68   extra)
69 summary(modelo_transf)
70
71 modelo_full <- ppm(bei ~ I(elev^2) * sqrt(grad), covariates = bei.extra)
72 summary(modelo_full)
73
74 AIC(modelo_int)
75 AIC(modelo_transf)
76 AIC(modelo_full)
77
78 par(mfrow = c(1, 3))
79 diagnose.ppm(modelo_int, main = "Modelo Interacci n")
80 diagnose.ppm(modelo_transf, main = "Modelo Transformado")
81 diagnose.ppm(modelo_full, main = "Modelo Full")
82
83 par(mfrow = c(1, 3))
84 plot(predict(modelo_int), main = "Intensidad - Interacci n")
85 plot(predict(modelo_transf), main = "Intensidad - Transformaci n")
86 plot(predict(modelo_full), main = "Intensidad - Full")
87
88 logLik(modelo_int)
89 logLik(modelo_transf)
90 logLik(modelo_full)
91
92 # ----- SIMULACI N PARA EL MODELO FULL -----
93 set.seed(123)
94 simulado <- simulate(modelo_full, nsim = 1)
95
96 par(mfrow = c(1, 2))
97 plot(bei, main = "Patr n observado")
98 plot(simulado[[1]], main = "Simulaci n del modelo final")
99 par(mfrow = c(1, 1))

```

Listing 1: Análisis espacial de la base bei