



# **ESCUELA POLITÉCNICA NACIONAL**

## **FACULTAD DE CIENCIAS**

### **PROCESOS PUNTUALES ESPACIALES Y SUS APLICACIONES**

### **PROCESOS PUNTUALES DE THOMAS**

**TRABAJO DE INTEGRACIÓN CURRICULAR PRESENTADO COMO  
REQUISITO PARA LA OBTENCIÓN DEL TÍTULO DE INGENIERO  
MATEMÁTICO**

**ARMANDO FRANCISCO LARA CIFUENTES**

[armando.lara@epn.edu.ec](mailto:armando.lara@epn.edu.ec)

**DIRECTORA: YANDIRA DENISSE CUVERO CALERO**

[yandira.cuvero@epn.edu.ec](mailto:yandira.cuvero@epn.edu.ec)

**DMQ, JULIO 2024**

## **CERTIFICACIONES**

Yo, ARMANDO FRANCISCO LARA CIFUENTES, declaro que el trabajo de integración curricular aquí descrito es de mi autoría; que no ha sido previamente presentado para ningún grado o calificación profesional; y, que he consultado las referencias bibliográficas que se incluyen en este documento.

---

ARMANDO FRANCISCO LARA CIFUENTES

Certifico que el presente trabajo de integración curricular fue desarrollado por ARMANDO FRANCISCO LARA CIFUENTES, bajo mi supervisión.

---

YANDIRA DENISSE CUVERO CALERO  
**DIRECTORA**

## **DECLARACIÓN DE AUTORÍA**

A través de la presente declaración, afirmamos que el trabajo de integración curricular aquí descrito, así como los productos resultantes del mismo, son públicos y estarán a disposición de la comunidad a través del repositorio institucional de la Escuela Politécnica Nacional; sin embargo, la titularidad de los derechos patrimoniales nos corresponde a los autores que hemos contribuido en el desarrollo del presente trabajo; observando para el efecto las disposiciones establecidas por el órgano competente en propiedad intelectual, la normativa interna y demás normas.

ARMANDO FRANCISCO LARA CIFUENTES

YANDIRA DENISSE CUVERO CALERO

## RESUMEN

Accidentes de tránsito son la segunda causa de muerte en Ecuador según la revista Primicias [7], hasta el año 2021 eran la octava causa algo que sorprende mucho pues en tan sólo tres años aumentó seis posiciones más. En este proyecto, se realiza un estudio de procesos puntuales en el espacio y en la red para modelar la intensidad de los accidentes de tránsito registrados en la administración zonal Eugenio Espejo entre 2017 y 2023. Utilizando los registros hasta el año 2022, se estima la intensidad y, posteriormente, con esa información, se simulan nuevos datos. Estos datos simulados se contrastan con los accidentes registrados en el año 2023 para evaluar la precisión del modelo.

Como resultado, se obtuvo una estimación de los sectores con mayor riesgo de siniestro de tránsito en la zona, se identificaron patrones y concentraciones de accidentes, así como agrupaciones. Se comparan los resultados del análisis espacial con los obtenidos del análisis en la red, destacando las fortalezas y limitaciones de cada enfoque. Además, se simula y modela un proceso de Thomas para analizar la ocurrencia de accidentes y sus patrones de agrupamiento, este proceso estocástico permite estudiar los patrones de agrupamiento y ofrece una comprensión más profunda de cómo y por qué se forman los clusters de accidentes.

**Palabras clave:** Patrón puntual, intensidad en la red, accidentes de tránsito, cluster, proceso de Thomas.

## **ABSTRACT**

Traffic accidents are the second leading cause of death in Ecuador according to Primicias magazine [7]. Until 2021, they were the eighth leading cause, which is quite surprising as in just three years, they moved up six positions. In this project, a study of point processes in space and on the network is conducted to model the intensity of traffic accidents recorded in the Eugenio Espejo zonal administration between 2017 and 2023. Using records up to 2022, the intensity is estimated and, subsequently, new data is simulated based on this information. These simulated data are compared with the accidents recorded in 2023 to evaluate the accuracy of the model.

As a result, an estimation of the sectors with the highest risk of traffic accidents in the area was obtained, patterns and concentrations of accidents, as well as clusters, were identified. The results of the spatial analysis are compared with those obtained from the network analysis, highlighting the strengths and limitations of each approach. Additionally, a Thomas process is simulated and modeled to analyze the occurrence of accidents and their clustering patterns. This stochastic process allows studying the clustering patterns and provides a deeper understanding of how and why accident clusters form.

**Keywords:** Point pattern, network intensity, traffic accidents, cluster, Thomas process.

---

# Índice general

---

<b>1. Descripción del componente desarrollado</b>	<b>10</b>
1.1. Objetivo general . . . . .	10
1.2. Objetivos específicos . . . . .	10
1.3. Alcance . . . . .	11
1.4. Marco teórico . . . . .	11
1.4.1. Proceso puntual espacial . . . . .	11
1.4.2. Proceso puntual en la red . . . . .	16
<b>2. Metodología</b>	<b>20</b>
2.1. Acerca de la base . . . . .	20
2.2. Región de estudio . . . . .	20
2.3. Análisis Espacial . . . . .	22
2.3.1. Prueba de aleatoriedad espacial completa (CSR) . . . . .	24
2.3.2. Interacción entre puntos . . . . .	25
2.3.3. Estimación de la intensidad no homogénea . . . . .	25
2.3.4. Simulación de un proceso de Thomas espacial . . . . .	28
2.3.5. Obtención de covariables . . . . .	30
2.3.6. Dependencia de covariables . . . . .	31
2.3.7. Ajuste del modelo . . . . .	32
2.4. Análisis sobre la red . . . . .	36

2.4.1. Obtención de la red . . . . .	37
2.4.2. Intensidad en la red . . . . .	37
2.4.3. Estimación de la función K . . . . .	40
2.4.4. Dependencia de covariables . . . . .	41
2.4.5. Ajuste del modelo . . . . .	43
<b>3. Resultados, conclusiones y recomendaciones</b>	<b>47</b>
3.1. Resultados . . . . .	47
3.1.1. Resultados en el espacio . . . . .	47
3.1.2. Resultados en la red . . . . .	49
3.2. Conclusiones . . . . .	52
3.3. Recomendaciones . . . . .	53
<b>Bibliografía</b>	<b>55</b>

---

## Índice de figuras

---

1.1. Proceso de agrupamiento . . . . .	15
2.1. Cantidad de accidentes y lesionados por mes . . . . .	21
2.2. Cantidad de personas fallecidas por mes . . . . .	22
2.3. Cantidad de fallecidos y lesionados por año . . . . .	22
2.4. Proceso puntual espacial . . . . .	23
2.5. Proceso puntual espacial sin las parroquias rurales . . . . .	23
2.6. Conteo de accidentes por cuadrantes . . . . .	24
2.7. Función $K$ vs $\hat{K}$ . . . . .	25
2.8. Estimación de la intensidad con corrección de borde de Diggle	26
2.9. Estimación de la intensidad con núcleo de suavizado cuártico	27
2.10. Simulación de un proceso de Thomas con la intensidad de Diggle: A la izquierda la intensidad, a la derecha la simulación del proceso . . . . .	29
2.11. Simulación de un proceso de Thomas con la intensidad de Diggle y kernel cuártico: A la izquierda la intensidad, a la derecha la simulación del proceso . . . . .	29
2.12. Obtención de covariables . . . . .	30
2.13. Covariables del proceso puntual . . . . .	31
2.14. Distancias más cercanas . . . . .	35
2.15. Red lineal de Eugenio Espejo . . . . .	38



2.16	Accidentes de tránsito ubicados sobre la red . . . . .	39
2.17	Estimación de intensidad con la corrección de borde . . . . .	40
2.18	Función $K$ vs $\hat{K}$ . . . . .	41
2.19	Función $K$ vs $\hat{K}$ inhomogénea . . . . .	41
2.20	Covariables sobre la red de Eugenio Espejo . . . . .	43
3.1.	Función de intensidad de los accidentes de tránsito en 2023	48
3.2.	Simulación de un proceso de Thomas con la intensidad es- timada . . . . .	48
3.3.	Datos reales (turquesa) vs datos simulados (amarillo) . . . . .	49
3.4.	Accidentes de tránsito de 2023 sobre la red . . . . .	50
3.5.	Simulación de proceso de Poisson inhomogéneo con inten- sidad estimada . . . . .	50
3.6.	Datos simulados (negro) vs datos reales (azul) . . . . .	51
3.7.	Comparación entre funciones de intensidad . . . . .	51
3.8.	Función $K$ estimada de los accidentes de 2023 . . . . .	52

# Capítulo 1

---

## Descripción del componente desarrollado

---

### 1.1. Objetivo general

Estudiar las propiedades y características de un Proceso puntual de Thomas.

### 1.2. Objetivos específicos

1. Identificar parroquias con la mayor cantidad de accidentes de tránsito, número de accidentes en cada año, etc.
2. Analizar patrones significativos, tendencias o agrupamientos espaciales que puedan surgir del modelo de Thomas, para obtener así una comprensión más profunda de la dinámica de los accidentes de tránsito en la administración zonal Eugenio Espejo.
3. Simular un proceso de Thomas en el espacio y comparar el proceso con los datos reales.
4. Desarrollar modelos predictivos basados en el análisis espacial y en la red vial sobre los accidentes registrados hasta el año 2022 para simular nuevos datos y comparar con los registros del año 2023, estimando así zonas con mayor riesgo de accidentes.

## 1.3. Alcance

El presente trabajo, por las limitaciones computacionales se enfocará en identificar posibles tendencias o patrones espaciales en la ocurrencia de accidentes de tránsito de la administración zonal Eugenio Espejo del cantón Quito de 2017 a 2023.

## 1.4. Marco teórico

### 1.4.1. Proceso puntual espacial

#### Conjunto localmente finito

Para cualquier subconjunto  $x \subseteq S$ , donde  $S \subseteq \mathbb{R}^d$  y  $N(x)$  es la cardinalidad de  $x$ ,  $x$  se dice localmente finito si  $N(x_B) < \infty$  para  $B \subseteq S$  Boreliano y acotado, además  $x_B = x \cap B$ .

#### Familia localmente finita

Sea  $x$  un conjunto localmente finito, entonces la familia localmente finita se define por:

$$N_{lf} = \{x \subseteq S : N(x_B) < \infty, \text{ para algún } B \subseteq S \text{ Boreliano y acotado}\} \quad (1.1)$$

A los elementos de esta familia, se los denomina *configuraciones aleatorias de puntos localmente finitos*.

#### Proceso puntual

Un proceso puntual  $X \in N_{lf}$  en  $\mathbb{R}^d$  es una configuración aleatoria localmente finita de puntos, tal que para todos los conjuntos de Borel acotados  $B \subseteq \mathbb{R}^d$ , el número de puntos de  $X$  que “caen” en  $B$ , es una variable aleatoria finita que se denotará como  $N_x(B)$  o bien  $N(x_B)$ .

## **Función de intensidad homogénea**

Baddeley [2] en el año 2010, define un proceso puntual homogéneo  $X$  como aquel cuyo número medio o esperado de puntos que se sitúan en un conjunto acotado  $B$  es proporcional al área de  $B$ , es decir:

$$\mathbb{E}(N(X_B)) = \lambda|B|$$

donde,  $\lambda$  es la *intensidad* y  $|B|$  denota el área de  $B$ .

Así pues, la *función de intensidad*  $\lambda(x)$  mide el número de puntos esperado por unidad de área (o volumen) en  $x$ , y se define como

$$\lambda(x) = \lim_{|dx| \rightarrow 0} \frac{\mathbb{E}(N(dx))}{|dx|} \quad (1.2)$$

Illian [5], menciona que si no hay gran concentración de puntos en la región  $B$ , a medida se denota

$$\Lambda(B) = \mathbb{E}(N_X(B))$$

conocido como *medida de intensidad*, y en el caso que exista gran concentración de puntos, existe una función de densidad  $\lambda(x)$  y se define la función de intensidad como

$$\Lambda(B) = \int_B \lambda(x) dx$$

## **Proceso de Poisson homogéneo**

Un proceso puntual  $X \in N_{lf}$  se dice completamente aleatorio (hipótesis de aleatoriedad espacial-CSR) si se cumplen los siguientes criterios:

- El número esperado de eventos (la intensidad  $\lambda(x)$ ) es homogéneo para todo  $B \subseteq \mathbb{R}^d$  Boreliano y acotado.
- El número de eventos en dos subregiones no superpuestas  $A_1$  y  $A_2$  son independientes.
- El número de eventos en cualquier subregión sigue la distribución de Poisson.

## Proceso puntual estacionario

Para un entero  $k$  y regiones  $\{A_i : i \in \{1, \dots, k\}\}$ , el proceso es estacionario si la distribución conjunta de  $N(A_1), \dots, N(A_k)$  es invariante ante la traslación de una cantidad arbitraria  $h$ .

## Función K de Ripley

Sea  $n$  el número de puntos en la ventana de muestreo  $W \subseteq \mathbb{R}^d$ ,  $b(x_i, r)$  un disco (o esfera) de radio  $r$  centrado en el punto  $x_i$  y sea también

$$n_i(r) = N_X(b(x_i, r) \setminus \{x_i\})$$

el número de puntos de  $N$  dentro de la distancia  $r$  desde el punto  $x_i$ , excluyendo al mismo  $x_i$ . Se define la función  $K$  de Ripley para un proceso puntual  $X$  estacionario como

$$K(r) = \frac{1}{\lambda} \mathbb{E}(N_X(b(x, r) \setminus \{x\}) | x \in X) \quad (1.3)$$

donde  $\lambda K(r)$  denota el número medio de puntos en el disco  $b(x, r)$  centrado en algún punto  $x$  (el cual no se incluye).

Para un proceso homogéneo de Poisson, la cantidad media de puntos que se encuentran dentro de la región  $b(x, r)$  es  $\lambda \pi r^2$ , por tanto, se tiene que  $K_{pois} = \pi r^2$ , luego, si  $k(r) > k_{pois}$  el proceso es consistente con agrupamiento (cluster), caso contrario se trata de un proceso regular.

Un estimador para  $K(r)$  según Adrian Baddeley [1] es:

$$\hat{K}(r) = \frac{1}{\lambda^2 |W|} \sum_i \sum_{j \neq i} \mathbf{1}_{\{\|x_i - x_j\| \leq r\}} e(x_i, x_j; r) \quad (1.4)$$

con  $e(x_i, x_j; r)$  un peso de corrección de borde.

## Estimación de la intensidad no homogénea

En el caso en el que se haya probado que el proceso puntual es inhomogéneo, la única intensidad (constante)  $\lambda$  que describe el proceso de Poisson es reemplazada por una función espacial  $\lambda(x)$ , donde  $x$  es la ubi-

cación espacial. La intensidad será diferente en cada región de la ventana, como estimador de  $\lambda(x)$ , usaremos la de corrección de borde de Diggle [1]

$$\hat{\lambda}^{(D)}(u) = \sum_{i=1}^n \frac{1}{e(x_i)} k(u - x_i) \quad (1.5)$$

donde  $k(u)$  es una función de núcleo el cual asigna más peso a los puntos de datos cercanos al punto donde se está estimando la densidad y menos peso a los puntos más alejados y  $e(u) = \int_W k(u - v)dv$  es una corrección del sesgo debido a los efectos de borde.

### **Prueba de aleatoriedad espacial completa (CSR)**

El test CSR es parte fundamental en el análisis de patrones puntuales, si se acepta la hipótesis, se puede asumir que el patrón de puntos dado es completamente aleatorio espacialmente. J. Illian y A. Penttinen [5] mencionan que si un proceso puntual es completamente aleatorio, se tiene dos consecuencias:

1. No hay necesidad de considerar un modelo más complicado; se puede utilizar el modelo simple de proceso de Poisson, con todas sus implicaciones.
2. Si no se dispone de información o datos adicionales sobre los procesos subyacentes, no es posible encontrar indicadores de una interacción interesante entre los puntos basados únicamente en la geometría del patrón observado.

Existen varios métodos para realizar este test, entre ellos el conteo de cuadrantes, medición de distancias, etc. El método que se utilizará aquí será el de los cuadrantes, el cual consta de la siguiente prueba de hipótesis:

$$\left\{ \begin{array}{l} \mathbf{H_0: El proceso puntual es completamente aleatorio.} \\ \mathbf{H_1: El proceso puntual no es completamente aleatorio.} \end{array} \right.$$

Para ello utilizaremos el estadístico de prueba  $\chi^2$  definido por J. Illian

y A. Penttinen [5] como

$$I = \frac{(k-1)s^2}{\bar{x}}$$

donde  $I$  es llamado el **índice de dispersión de puntos**,  $k$  es el número de cuadrantes o regiones,  $\bar{x}$  es el promedio de puntos por cuadrante y  $s^2$  es la varianza muestral del número de puntos por cuadrante. El índice  $I$  sigue aproximadamente una distribución  $\chi^2$  con  $k-1$  grados de libertad siempre que  $k > 6$  y  $\lambda|Q| > 1$ ,  $|Q|$  representa el área del cuadrante  $Q$ .

Por lo tanto, si  $I$  excede el valor crítico  $\chi_{k-1, \alpha}^2$  o está por debajo del valor crítico  $\chi_{k-1, 1-\alpha}^2$ , se rechaza la hipótesis de CSR con un nivel de confianza de  $1-\alpha$ .

### Proceso de cluster

M.N.M. van Lieshout [12] define un proceso de cluster en dos pasos. En el primer paso, se toma una muestra de un proceso de puntos “padre”  $Y$ . En segundo lugar, condicionado a los padres, cada uno de los puntos padres  $y_i$  genera un nuevo proceso de puntos “hijos”  $x_{ij}$  y se toma la superposición de todos los hijos.

La figura 1.1 extraída del libro [1], representa cómo se forman los procesos cluster. En la primera gráfica se generan los puntos padres, en la segunda gráfica se generan los puntos hijos, y en la tercera gráfica los puntos hijos superponen a los padres y constituyen el proceso de agrupamiento.

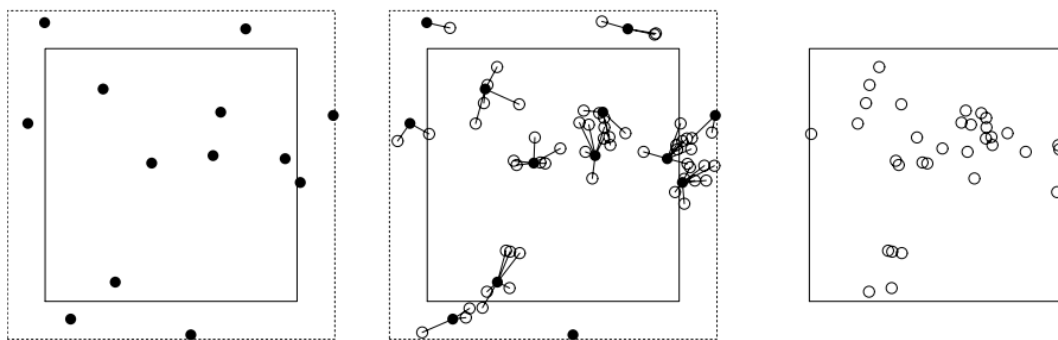


Figura 1.1: Proceso de agrupamiento

## Proceso de cluster de Neymann-Scott

El caso especial en el que los padres forman un proceso de Poisson homogéneo con intensidad  $\lambda_p$  y en el que un número cualquiera de hijos con distribución de Poisson con media  $\lambda_c$  se dispersan independientemente alrededor de su padre según una densidad de probabilidad  $f$  se conoce como un proceso de cluster de Neyman-Scott.

Sea  $X$  el proceso de Poisson homogéneo de los puntos padres sobre  $\mathbb{R}^d$ , con intensidad  $\lambda_p > 0$ . Para  $x \in \mathbb{R}^d$  (para cada punto padre), sea  $Z_x$  el proceso de Poisson inhomogéneo de los puntos hijos con intensidad  $\lambda_c > 0$  para alguna probabilidad de densidad  $f$  sobre  $\mathbb{R}^d$  tal que

$$\int_{\mathbb{R}^d} \left( 1 - \exp \left[ -\lambda_c \int_A f(y-x) dy \right] \right) dx < \infty \quad (1.6)$$

para cada  $A \subset \mathbb{R}^d$  boreliano. Entonces,  $\cup_{x \in X} Z_x$  es un proceso de Neymann-Scott Cox. Además, Lieshout [12] menciona en el **teorema 4.6** que la intensidad para cualquier proceso de Neymann-Scott Cox es  $\lambda_p \lambda_c$ .

## Proceso de Thomas

En este proceso, según M.N.M. van Lieshout [12], los hijos condicionado a los padres son ubicados alrededor de la ventana con una distribución normal con media cero y varianza  $\sigma^2$ , es decir, para  $x \in \mathbb{R}^2$

$$f(x) = \frac{1}{2\pi\sigma^2} \exp \left[ -\frac{1}{2\sigma^2} \|x\|^2 \right] \quad (1.7)$$

### 1.4.2. Proceso puntual en la red

Cuando se trabaja con redes [3], el panorama cambia significativamente, sería incorrecto sólo considerar los puntos de los accidentes, pues la función  $K$  y otras características no son las mismas en estos casos y esto se debe principalmente a que la red no es homogénea. Existen métodos computacionales para abordar esto unas más efectivas que otras, cuando se trabaja con redes la estimación de la densidad de núcleo y otras funciones como la correlación por pares no es directo como en los



casos sin red, esto implica ya una limitación y complejidad ya que diferentes ubicaciones espaciales en la red están rodeadas por diferentes configuraciones de líneas, por ello no se puede aplicar las herramientas clásicas de procesos puntuales sin red, es necesario en estos casos nuevas herramientas estadísticas y computacionales, usando por ejemplo las métricas de distancia como el camino más corto, esto también suele presentar desafíos computacionales. En resumen, al trabajar sobre la red:

- El espacio de estudio no es convexo.
- La forma de medir las distancias en la red, no es euclídea, por lo general se usa la distancia del camino más corto.
- Se modifican las definiciones de las funciones  $\lambda$  y  $K$ .
- La red no es homogénea, es decir, cada longitud  $l \in L$  tiene medida diferente.

### **Red lineal**

Se define una red lineal  $L$  como la unión de  $N$  segmentos de línea  $l_i$  tal que  $L = \bigcup_{i=1}^n l_i$  y  $N < \infty$ . Donde

$$l_i = [u_i, v_i] = \{w : w = tu_i + (1 - t)v_i, 0 \leq t \leq 1\}$$

siendo  $u_i, v_i \in \mathbb{R}^2$  los extremos de  $l_i$ .

Se asume que para  $l_i$  y  $l_j$ , con  $i \neq j$ , su intersección es vacía o es un punto extremo de ambos segmentos.

### **Intensidad en la red**

Existen muchos métodos propuestos para estimar una función de intensidad sobre la red, en este documento se hará uso del estimador propuesto por Suman Rakshit, Jorge Mateu y otros autores [11] que es una modificación de la ecuación 1.5. Sea  $x = (x_1, \dots, x_n)$  un proceso puntual sobre una red lineal  $L$ , sea  $k$  una función de núcleo bivariada, es decir, una densidad de probabilidad definida en  $\mathbb{R}^2$ , el estimador del núcleo de convolución de intensidad con la corrección de borde de Jones-Diggle es:

$$\hat{\lambda}^{JD}(u) = \sum_{i=1}^n \frac{k(u - x_i)}{c_L(x_i)}, \quad u \in L \quad (1.8)$$

Donde

$$c_L(u) = \int_L k(v - u)dv, \quad u \in L$$

### Proceso puntual aleatorio en la red

Sea  $X$  un proceso puntual en la red  $L$ , se supone:

- El número total de puntos es finito con probabilidad 1, tiene media cero y varianza finita.
- No hay puntos coincidentes múltiples (proceso puntual simple).

El proceso puntual  $X$  tiene función de intensidad  $\lambda(u)$ , con  $u \in L$  si

$$\mathbf{E}[N_X(B)] = \int_B \lambda(u)du \quad (1.9)$$

para todo  $B \subset L$  cerrado.

### Estimación de la función K

Al igual que la función de intensidad no homogénea para la red, existen muchos métodos propuestos para la estimación de la función  $K$ , en este documento se utilizará el estimador propuesto por Suman Rakshit, Adrian Baddeley y Gopalan Nair [8], el cual consiste en una corrección de la geometría en la red, es decir, aquellos segmentos de red en los cuales no se encuentran muchos eventos se les dará un peso menor, con eso se evita el sesgo en el análisis hacia ciertas regiones en donde no se registran mayor cantidad de accidentes.

Si el proceso puntual  $X$  ubicado sobre la red  $L$  tiene  $n$  puntos, se tiene la siguiente función  $K$  empírica con corrección geométrica

$$\hat{K}_L(r) = \frac{|L|}{n(n-1)} \sum_{i=1}^n \sum_{j \neq i} \frac{\mathbf{1}_{(d_L(x_i, x_j) \leq r)}}{m(d_L(x_i, x_i, x_j))} \quad (1.10)$$

donde  $m(u, r) = \text{card}(c_L(u, r))$  es el número de puntos sobre el perímetro  $c_L(u, r) = \{v \in L : d_L(u, v) = r\}$  y  $d_L(u, v) \leq r$  es la distancia del camino más corto desde la ubicación  $u$  hasta la ubicación  $v$  tal que no supera el valor del radio  $r$ ,  $m$  es el peso de compensación para la geometría de la red. Además Rakshit S. [8] menciona que si  $K_L(r) = r$  ( $r > 0$ ), entonces el proceso es de Poisson homogéneo. Así, como en el caso espacial, si la función  $\hat{K}$  está por debajo de la función  $K$  teórica, entonces el proceso es de Poisson homogéneo.

Para un proceso puntual inhomogéneo, se tiene el estimador de la función  $K$  siguiente

$$\hat{K}_L^{ih} = \frac{1}{\sum_{i=1}^n \frac{1}{\hat{\lambda}(x_i)}} \sum_{i=1}^n \sum_{j \neq i} \frac{\mathbf{1}_{(d_L(x_i, x_j) \leq r)}}{\hat{\lambda}(x_i) \hat{\lambda}(x_j) m(d_L(x_i, x_i, x_j))} \quad (1.11)$$

donde  $\hat{\lambda}(\cdot)$  es una estimación de la función de intensidad que varía espacialmente del proceso puntual en  $L$ .

# Capítulo 2

---

## Metodología

---

### 2.1. Acerca de la base

La base de datos considerada es de acceso libre y está disponible en la página de la Agencia nacional de tránsito (ANT) [4], la base de datos en consideración consta de 154,608 registros y proporciona información sobre los accidentes de tránsito en Ecuador desde enero de 2017 hasta julio de 2023.

Entre las variables presentes en la base se encuentran el año del incidente, la cantidad de personas fallecidas, la cantidad de lesionados, la parroquia, las coordenadas de latitud y longitud, el tipo de vehículo, entre otras.

### 2.2. Región de estudio

La región de estudio será la administración zonal Eugenio Espejo ubicada en el cantón Quito, hubo 33,345 registros en Quito en ese periodo, de los cuales 9451 registros corresponden a la región de estudio. En lo sucesivo, se mencionará Eugenio Espejo en lugar de Administración Zonal Eugenio Espejo para mayor brevedad.

Eugenio Espejo, cuenta con una superficie de  $75,89\text{km}^2$  y una población de 420.000 habitantes aproximadamente según el censo del 2010

(INEC) [13], está conformada por 18 parroquias: 10 urbanas y 8 rurales tales como Comité del Pueblo, Inca, Iñaquito, Nayón, etc.

Se escoge esta región debido a la gran cantidad de siniestros de tránsito registrados además de las limitaciones computacionales.

En la figura 2.1 se observan las series de tiempo de la cantidad de accidentes y lesionados por mes en la región Eugenio Espejo, el año 2017 fue el periodo con más siniestros comprendiendo entre 300 y 400 accidentes de tránsito resultando entre 100 y 200 personas lesionadas por mes como se aprecia en la gráfica.

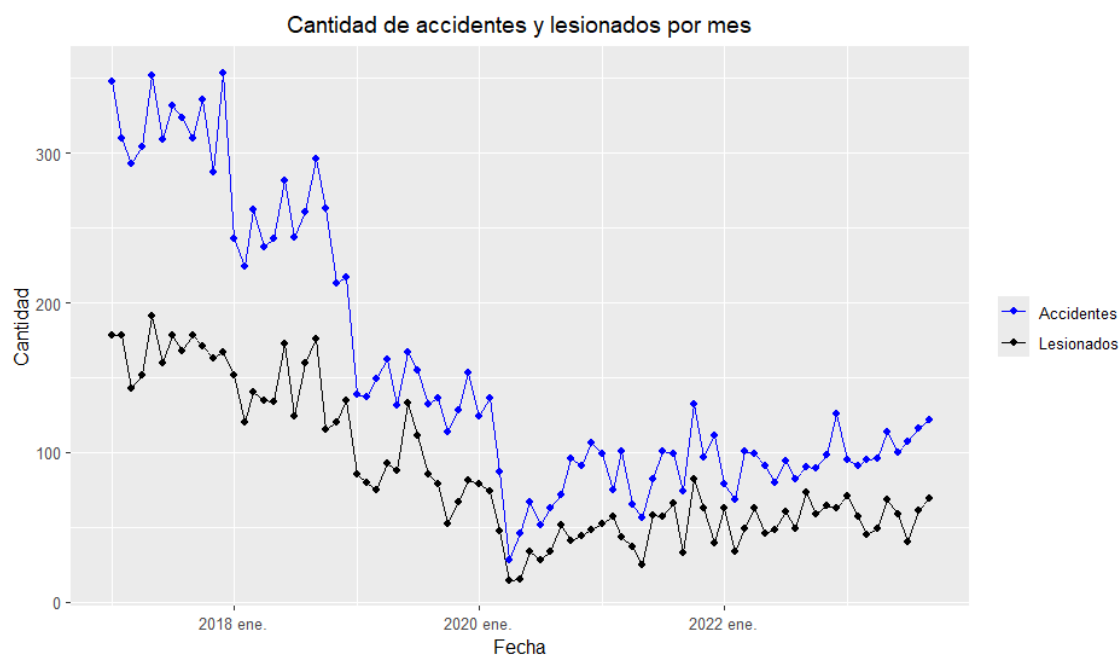


Figura 2.1: Cantidad de accidentes y lesionados por mes

en la figura 2.2 se observa en la serie temporal la cantidad de personas fallecidas en esta región, mostrando que cada año presenta un número similar de fallecidos, diciembre de 2022 siendo el mes con la mayor cantidad de eventos.

En la figura 2.3, se registraron un total de 381 personas fallecidas y 7051 personas lesionadas, siendo los años 2017 y 2019 los años con mayor cantidad de fallecidos (figura izquierda) y el año 2017 con la mayor cantidad de lesionados (figura derecha), en los accidentes de tránsito registrados en estos sectores, hubo más personas lesionadas que fallecidas, el porcentaje de fallecidos es de alrededor del 5.13. Además, la mayor cantidad de personas fallecidas y lesionadas son hombres con tasas del

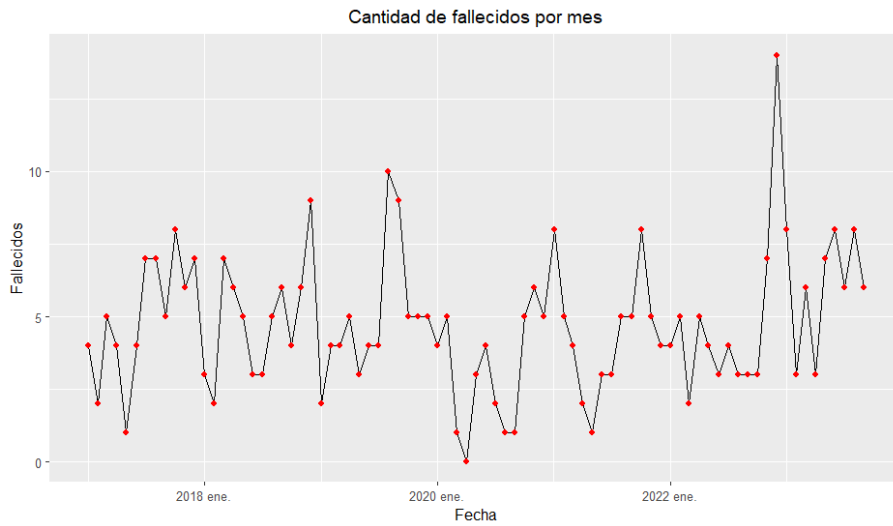


Figura 2.2: Cantidad de personas fallecidas por mes

78.48% y 65.56% respectivamente. %

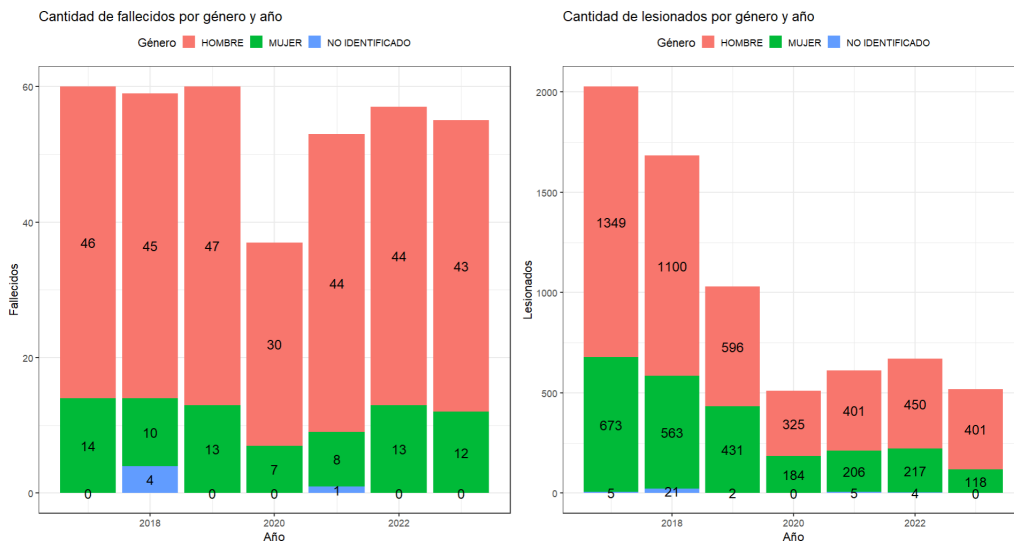


Figura 2.3: Cantidad de fallecidos y lesionados por año

### 2.3. Análisis Espacial

En la figura 2.4 se representa los siniestros de tránsito a nivel espacial. Se puede observar una concentración de eventos en las zonas urbanas, el polígono del gráfico corresponde a Eugenio Espejo, mientras que los puntos situados en él corresponden a los accidentes de tránsito.

Puesto que no hay presencia significativa de accidentes en las pa-



### 2.3.1. Prueba de aleatoriedad espacial completa (CSR)

En primera instancia, nos interesa saber si nuestro proceso puntual se trata o no de un proceso homogéneo de Poisson (CSR), para ello utilizaremos el conteo de cuadrantes, es decir, Eugenio Espejo se dividirá en subregiones rectangulares con misma área y se contabiliza el número de puntos en cada subregión.

En la figura 2.6 se muestra el conteo de accidentes por subregiones y también la intensidad respecto a éstos, notemos que la intensidad varía por cuadrante, por lo que podríamos suponer que este proceso es no homogéneo. La división de cuadrantes es de alrededor de 1000 regiones con misma área.

Asumiendo que la intensidad del proceso es homogénea, la intensidad promedio estimada es de  $0.0001109068m^{-2}$ , la razón de que el valor sea pequeño es como se mencionó antes, la mayoría de puntos se ubican en las zonas urbanas, debido a que presenta un mayor número de calles.

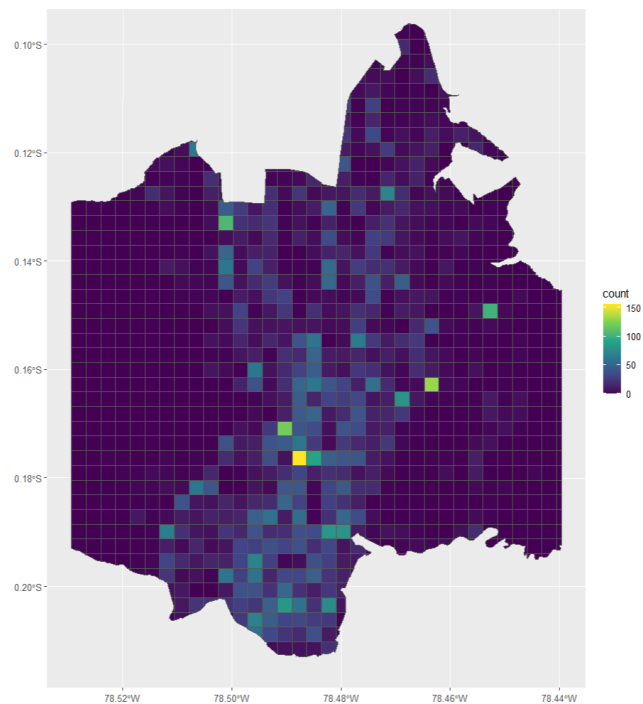


Figura 2.6: Conteo de accidentes por cuadrantes

Al realizar el test CSR utilizando el índice de dispersión de puntos en nuestra región de estudio, tenemos un valor-p menor a 0.05, por lo cual se rechaza  $H_0$ . Así, se ha probado que el proceso no es homogéneo que



respalda lo mostrado en la figura 2.6.

### 2.3.2. Interacción entre puntos

Para las interacciones entre puntos, usaremos la función  $K$  de Ripley, se muestra en la figura 2.7 la gráfica de la función  $K$  vs la función  $\hat{K}$ , como la función estimada está por encima de la teórica, se puede concluir que el proceso puntual está acotado espacialmente o es consistente con agrupamiento.

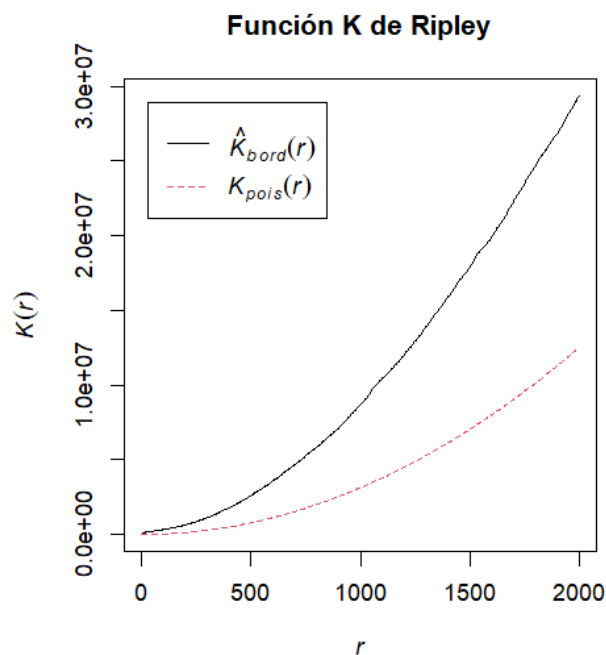


Figura 2.7: Función  $K$  vs  $\hat{K}$

### 2.3.3. Estimación de la intensidad no homogénea

Para estudiar la intensidad, se usó el estimador 1.5 que evita el efecto de borde, esto imposibilita que observaciones fuera de la región interactúen con los datos que observamos. Si sólo consideramos suavizar el ancho de banda con el valor óptimo de Diggle que se obtiene por defecto con la corrección de borde, obtenemos en la figura 2.8, la gráfica de la estimación de la intensidad de núcleo, donde la mayor interacción entre

los puntos se tiene en la parte sur de la administración zonal, lo cual contrasta la no homogeneidad del proceso.

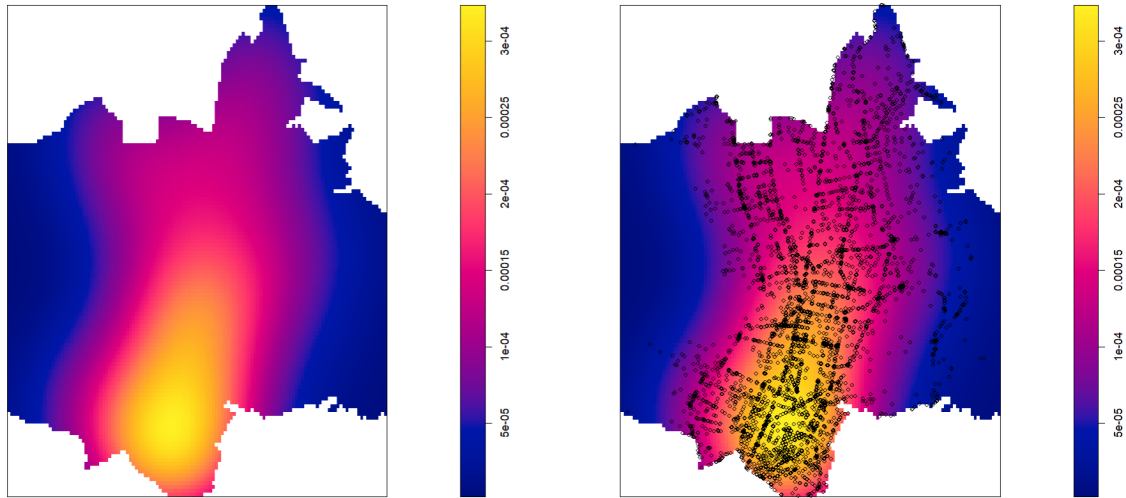


Figura 2.8: Estimación de la intensidad con corrección de borde de Diggle

Para obtener una intensidad en donde se puedan observar de mejor manera la cantidad de puntos esperados por cada unidad de área, utilizaremos adicionalmente un valor de suavizamiento para el kernel, cuando el suavizamiento es pequeño los vecindarios que se generan son pequeños, mientras que si es grande los vecindarios son extensos y puede generar un proceso homogéneo. Para elegir un valor adecuado de suavizamiento, según D.G.Rossiter [10], se suelen usar dos tipos: el Gaussiano y el cuártico, el más recomendado por el autor es el cuártico debido a su uso en muchas aplicaciones. éste se define por:

$$k(u) = \begin{cases} \frac{3}{\pi}(1 - \|u\|^2)^2 & \text{si } u \in (-1, 1) \\ 0 & \text{en caso contrario} \end{cases}$$

donde  $u \in \mathbb{R}^2$  y  $\|u\|^2 = u_1^2 + u_2^2$ .

La “unidad” está establecida por el ancho de banda  $h$ :

$$h = \frac{\|x - x_i\|}{u}$$

El núcleo cuártico asigna pesos a los eventos según su distancia del punto de interés, es decir, usamos ambos suavizamientos al mismo tiempo para mejorar la estimación de la función de intensidad, con la correc-

ción de borde de Diggle evitamos sobreestimaciones en regiones de estudio donde no se encuentran observaciones, mientras que con el suavizado cuártico asignamos pesos a los puntos de datos en función de su distancia desde el punto de interés, con ello se da más importancia en el análisis a las regiones en donde se ubican más accidentes y evitamos aquellas regiones en donde no se tengan registros o éstos sean casi nulos.

En la figura 2.9 se muestran 4 gráficos con el ancho de banda “óptimo” y algunos múltiplos de éste, la banda de ancho óptima generada por Diggle es de 1.66m, es decir, la densidad de eventos se calcula considerando un radio de 1.66 metros alrededor de cada accidente, junto con las respectivas líneas de contorno. Se observa que para  $k = 1$  y  $k = 50$ , los vecindarios son bastante pequeños por lo que es difícil visualizar la intensidad de los puntos en cada unidad de área, en cambio para  $k = 100$  y  $k = 150$ , el panorama es diferente, pues se pueden visualizar mayor concentración de puntos en ciertas regiones, sobretodo con  $k = 100$ , además podemos notar que eligiendo bandas de ancho adecuados, se obtienen gráficos de densidad mejores que la figura 2.8.

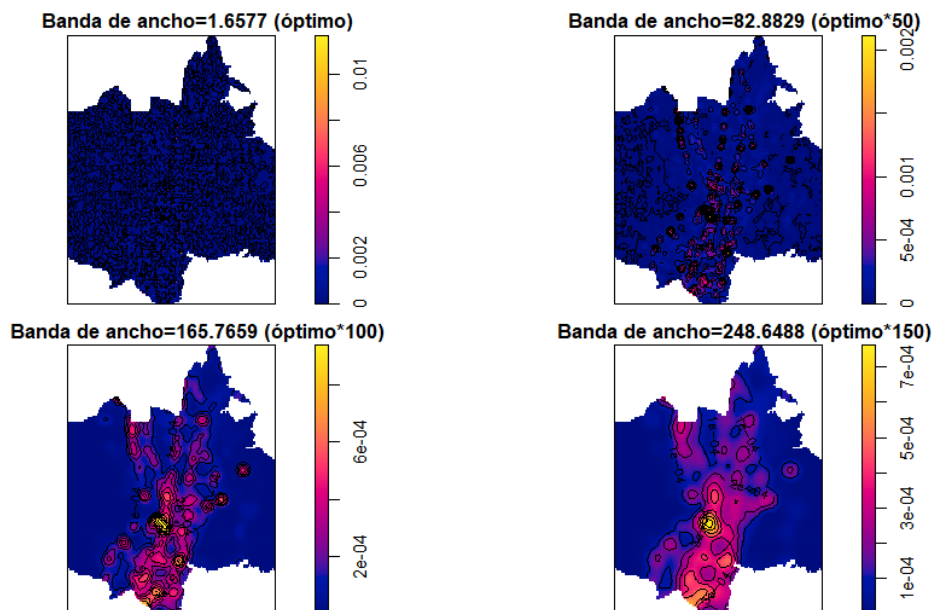


Figura 2.9: Estimación de la intensidad con núcleo de suavizado cuártico

### 2.3.4. Simulación de un proceso de Thomas espacial

A continuación se simulará un proceso de Thomas espacial sobre la ventana de muestreo de la administración zonal Eugenio Espejo, se tenían 9451 puntos ubicados sobre esta ventana, así que se realizará una simulación de puntos cercano a este valor. Para esto, se utilizará la función **rThomas** de la librería **spatstat** de R, donde el primer argumento es el valor de la intensidad del proceso de Poisson de los puntos padres, el segundo argumento es el valor de la desviación estándar  $\sigma$  de la ubicación de los puntos hijos alrededor de los padres, y el tercer argumento es la intensidad del proceso de los hijos.

Para la intensidad de los padres se usará un valor de 1, como  $\sigma$  utilizaremos un valor de 8, y como función de intensidad del proceso de los hijos se usarán dos funciones, primero sobre la función de la intensidad de la figura 2.8 que sólo consideraba el suavizamiento de banda de ancho, luego se realizará sobre la intensidad del cuarto gráfico de la figura 2.9, con ello el número medio de puntos por cluster será diferente en cada simulación.

La razón para simular con estos valores es obtener un número de puntos similar al del proceso original, por ejemplo si se trabaja con la intensidad del cuarto gráfico de la figura 2.9, la intensidad promedio de los puntos hijos es de 0.00011, obteniendo que la intensidad del proceso es  $\lambda = 1(0,00011) = 0,00011$ , dando un número esperado de 9379 puntos. Aunque existen otros métodos y librerías para simular un proceso de Thomas, se elige esta opción debido a su rapidez computacional en comparación con las alternativas.

#### Método 1

En la figura 2.10, se muestra la simulación de un proceso de Thomas con la intensidad de corrección de borde de Diggle, con esta intensidad el total de puntos generados es de 9449, como se observa hay mayor concentración de puntos en la parte baja de la ventana.

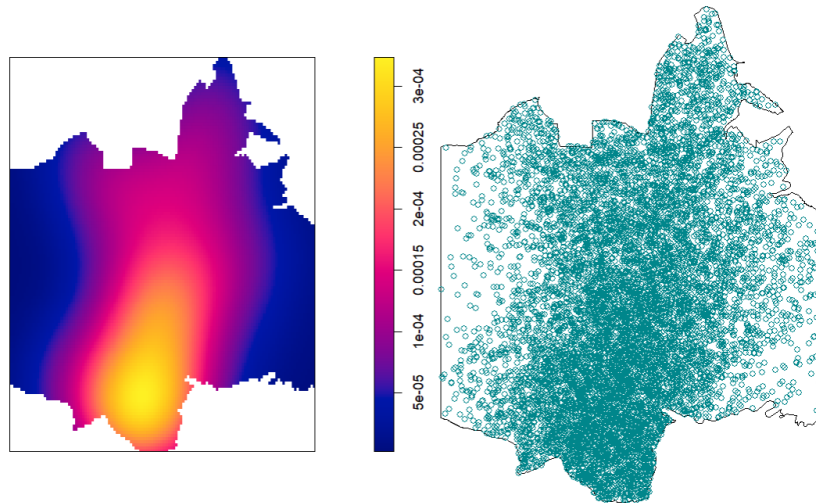


Figura 2.10: Simulación de un proceso de Thomas con la intensidad de Diggle: A la izquierda la intensidad, a la derecha la simulación del proceso

## Método 2

En la figura 2.11 se ha realizado una simulación del proceso de Thomas con el suavizamiento de núcleo cuártico y banda de ancho de 248.65, se han generado 9447 puntos y se puede presenciar de mejor manera la agrupación de los puntos, esto debido a que la función de intensidad tiene un ancho de banda y suavizamiento de núcleo óptimos.

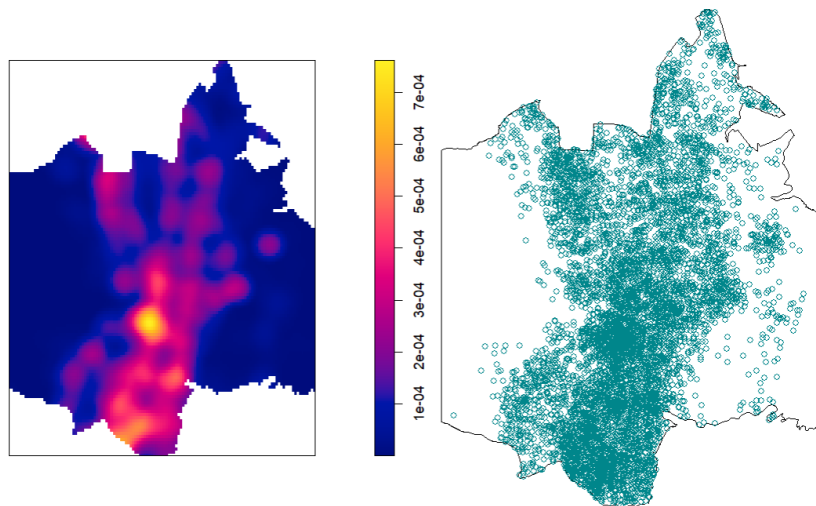


Figura 2.11: Simulación de un proceso de Thomas con la intensidad de Diggle y kernel cuártico: A la izquierda la intensidad, a la derecha la simulación del proceso

### 2.3.5. Obtención de covariables

Puesto que la base de datos de la ANT no contaba con covariables para poder realizar el ajuste de un modelo, se procedió a obtener las mismas mediante el lenguaje de programación Python utilizando la librería **osmnx** la cual sirve para trabajar con datos geospaciales provenientes de OpenStreetMap (OSM). Está diseñada para simplificar la descarga, el análisis y la visualización de datos geográficos de OSM.

Con la librería **osmnx** se puede obtener información geográfica respecto a una localización, como por ejemplo ubicaciones de restaurantes, hospitales, semáforos, lagos, carreteras, etc.

Mediante la ayuda de este paquete se obtuvieron cuatro covariables que son las ubicaciones de las paradas de buses, semáforos, estaciones de policía (UPC) y señales de pare, se decidió trabajar con estas covariables ya que podrían ser las más explicativas para este proceso.

Posteriormente, se procedió a importar estas ubicaciones a un archivo shapefile para leerlos desde el lenguaje de programación R y trabajar en conjunto con la librería **spatstat** y otras más. Se muestra en la figura 2.12 un resumen del procedimiento realizado.

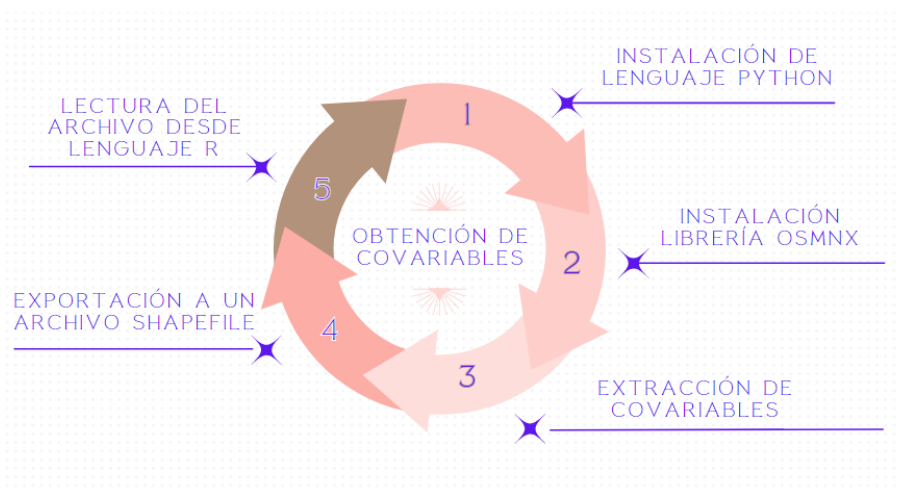


Figura 2.12: Obtención de covariables

En Eugenio Espejo, se tienen 708 paradas de buses, 1554 semáforos, 333 señales de pare y 45 estaciones de policía. En la figura 2.13 se muestran las covariables sobre la ventana de muestreo.

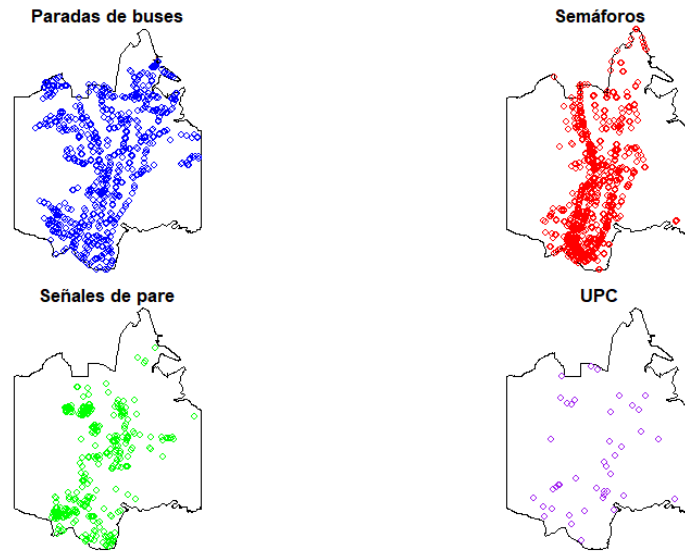


Figura 2.13: Covariables del proceso puntual

### 2.3.6. Dependencia de covariables

Una vez obtenidas las covariables, es importante saber si la intensidad de nuestro proceso es dependiente o no de éstas.

Existen varios métodos para verificar la dependencia de una covariable  $Z$  en el proceso, en este documento usaremos el test de Berman, Berman propuso dos pruebas denominadas  $Z_1$  y  $Z_2$  para la dependencia de un proceso puntual con respecto a una covariable espacial, por brevedad, no se profundizará en detalle en este test. Sin embargo, para una explicación completa de esta metodología, se sugiere revisar el libro de A. Baddeley [1] en la sección 10.5, en resumen, la prueba  $Z_1$  se basa en la suma  $T_1 = \sum_i Z(x_i)$  de los valores de la covariable en todos los puntos de datos  $x_i$ , mientras que la prueba  $Z_2$  se basa en  $T_2 = \sum_i G(Z(x_i))$  donde  $G$  es la función de distribución acumulativa de  $Z(u)$  sobre todas las ubicaciones  $u$  en la ventana, definida por

$$G(z) = \frac{1}{|W|} \int_W \mathbf{1}_{\{Z(u) \leq z\}} du$$

el test se basa en el siguiente contraste de hipótesis:

$\left\{ \begin{array}{l} \mathbf{Ho: la intensidad es independiente de la covariable.} \\ \mathbf{H1: la intensidad es dependiente de la covariable.} \end{array} \right.$

Se usará la prueba  $Z_2$  puesto que es más rigurosa y estricta. Al realizar el test de Berman, obtenemos que para las cuatro covariables el valor  $p$  es menor a 0.05. Por lo tanto, se puede concluir que el proceso es dependiente de estas covariables.

### 2.3.7. Ajuste del modelo

El modelo que se ajustará será el de Thomas, mediante la función **kppm** de la librería **spatstat** [2]. El método **kppm** ajusta un modelo log-lineal definido por

$$\log(\lambda(u)) = B(u) + \theta^T Z(u) \quad (2.1)$$

donde  $B(u)$  es una función de línea base o de compensación opcional,  $\theta$  es un vector de parámetros (coeficientes  $\beta_i$ ) y  $Z(u)$  es un vector de covariables, para  $t$  variables se tienen  $t + 1$  coeficientes a estimar.

Para un proceso de cluster donde los puntos padres forman un proceso de Poisson homogéneo con intensidad  $k$ , y los hijos una intensidad  $\mu$ , se tiene el modelo

$$\log(\mu(u)) = -\log(k) + B(u) + \theta^T Z(u) \quad (2.2)$$

Los parámetros para un proceso de Thomas se estiman haciendo coincidir la función  $k$  no homogénea estimada con la función  $k/\rho$  teórica del proceso de Thomas, donde  $\rho = 1 - e^{-\mu}$ .

A. Baddeley [1] menciona que con este método, la intensidad estimada de los nodos hijos sujetos a sus padres es

$$\mu = \exp\left(\beta_0 + \sum_{i=1}^t \beta_i Z_i(u)\right)$$

donde  $Z_1, Z_2, Z_3, \dots, Z_t$  son las covariables del proceso.

Se propondrán cuatro métodos para ajustar el modelo, y posteriormente se seleccionará aquel que muestre mayor significancia en las va-



riables.

Los registros del año 2023 se emplearán como datos de validación, mientras que los demás serán utilizados como base para el modelado. El objetivo es ajustar el modelo con estos datos y posteriormente aplicarlo sobre nuevos datos simulados y comparar su función de intensidad con los datos de validación para así evaluar su desempeño.

### Modelo 1: Ubicaciones de las covariables

En este caso, se toman las covariables presentadas en la figura 2.13, es decir,  $Z_1, Z_2, Z_3, Z_4$ , representan las ubicaciones de las paradas de buses, semáforos, señales de pare y UPC respectivamente.

Al ajustar el modelo con todas las covariables espaciales, se tiene que tanto las señalizaciones como las estaciones de policía no son significativas, por lo que las omitimos del análisis.

Se muestra en la tabla 2.1, los resultados del ajuste del modelo Thomas para este proceso.

	<b>Estimate</b>	<b>S.E.</b>	<b>CI95.lo</b>	<b>CI95.hi</b>	<b>Ztest</b>	<b>Zval</b>
(Intercept)	-7.6213	0.2078	-8.0285	-7.2141	***	-36.6804
$Z_1$	-0.0019	0.0007	-0.0032	-0.0005	**	-2.6617
$Z_2$	-0.0037	0.0006	-0.0048	-0.0026	***	-6.4566
kappa	0.00002					
scale	5.1446					

Cuadro 2.1: Resultados de la estimación

la intensidad estimada de los puntos padres es  $\hat{k} = 0,00002$  y con un radio de cluster  $\hat{R} = 5,1446$ . Luego, la intensidad estimada por este modelo es:

$$\hat{\mu}(u) = \frac{1}{0,00002} \exp(-7,6213 - 0,0019Z_1 - 0,0037Z_2)$$

### Modelo 2: Coordenadas cartesianas

En el caso de que no se disponga de covariables en la base (como se tenía al inicio), se puede ajustar un modelo con las coordenadas  $x$  e  $y$  de

los de los accidentes de tránsito, el modelo ajustado es

$$\mu(u) = \frac{1}{k} \exp(\beta_0 + \beta_1 x + \beta_2 y)$$

Las estimaciones para las coordenadas son casi nulas y por ende no significativas, obteniéndose así que la intensidad del modelo depende únicamente del intercepto con un valor de -7.1255, resultando en un proceso de Poisson homogéneo ya que para cada accidente de tránsito la intensidad tendría el mismo valor.

### **Modelo 3: Distancias más cercanas**

Para este método, se utilizarán las distancias a las ubicaciones más cercanas de las covariables de la figura 2.13, para ello se definen:

- $A_1$ : distancia a la parada de bus más cercana.
- $A_2$ : distancia al semáforo más cercano.
- $A_3$ : distancia a la señalización de pare más cercana.
- $A_4$ : distancia a la upc más cercana.

La intensidad estimada del modelo con estas covariables es

$$\mu(u) = \frac{1}{k} \exp(\beta_0 + \beta_1 A_1 + \beta_2 A_2 + \beta_3 A_3 + \beta_4 A_4)$$

En la figura 2.14 se muestran las gráficas de estas covariables

Al ajustar el modelo de Thomas sobre las cuatro covariables, se obtiene que tanto el intercepto como la distancia a la upc más cercana son significativas, es por ello que se utilizará el modelo con esta covariable para realizar el análisis espacial sobre los datos de validación.

En la tabla 2.2 se muestran los resultados de la estimación del modelo

Se observa que la intensidad estimada de los puntos padres es  $\hat{k} = 0,00001776674$  y radio de cluster  $\hat{R} = 5,26$ .

Así, utilizando la distancia a la upc más cercana se tiene como estimación del modelo

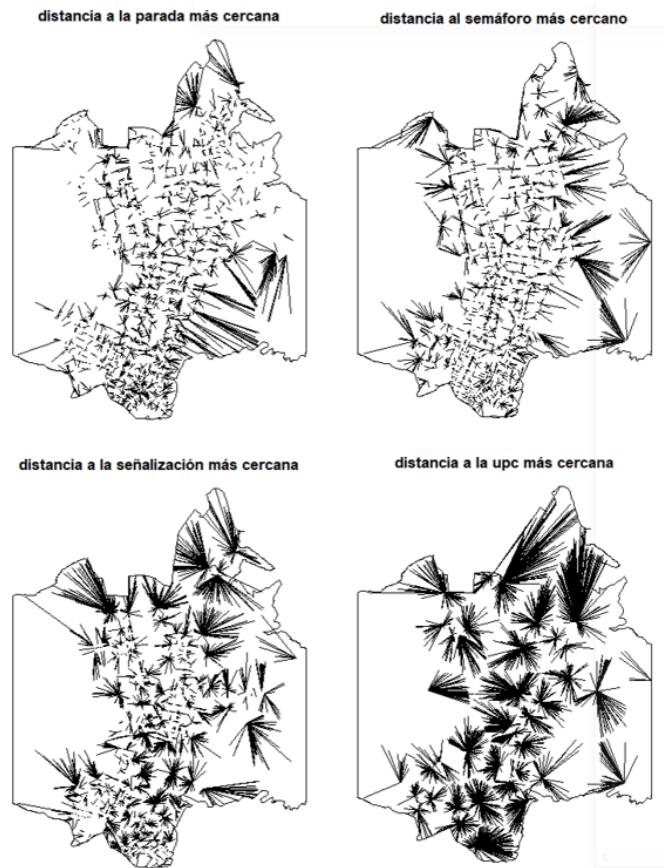


Figura 2.14: Distancias más cercanas

	<b>Estimate</b>	<b>S.E.</b>	<b>CI95.lo</b>	<b>CI95.hi</b>	<b>Ztest</b>	<b>Zval</b>
(Intercept)	-10.37169	0.1815	-10.7275	-10.0159	***	-57.1321
$A_4$	13.1496	1.9812	9.2665	17.0327	***	6.6371
kappa	0.00001776674					
scale	5.26					

Cuadro 2.2: Resultados de la estimación

$$\hat{\mu}(u) = \frac{1}{0,00001776674} \exp(-10,37169 + 13,1496A_4) \quad (2.3)$$

el coeficiente de  $A_4$  es positivo, indicando que el riesgo de accidente aumenta con la distancia desde la estación de policía más cercana.

#### **Modelo 4: Covariables más significativas**

Utilizando las covariables más significativas de los modelos anteriores se propone el siguiente modelo

$$\mu(u) = \frac{1}{k} \exp(\beta_1 Z_1 + \beta_2 A_4 + \beta_3 x)$$

es decir, trabajamos con las posiciones de las paradas de buses, distancias de los accidentes a la upc más cercana y la coordenada  $x$ . Se muestran los resultados de la estimación de este modelo en la tabla 2.3

	<b>Estimate</b>	<b>S.E.</b>	<b>CI95.lo</b>	<b>CI95.hi</b>	<b>Ztest</b>	<b>Zval</b>
$Z_1$	-0.0032	0.0006	-0.0045	-0.002	***	-5.2328
$A_4$	9.0322	2.1767	4.7659	13.2984	***	4.1495
$x$	-0.00001	0.0000004	-0.00001	-0.00001	***	-31.9802
kappa	0.00001859688					
scale	5.1035					

Cuadro 2.3: Resultados de la estimación

la intensidad estimada del proceso con este modelo es

$$\hat{\mu}(u) = \frac{1}{0,00001859688} \exp(-0,0032Z_1 + 9,0322A_4 - 0,00001x) \quad (2.4)$$

el coeficiente  $\hat{\beta}_1$  es negativo indicando que a mayor presencia de paradas de buses el riesgo de accidente se reduce,  $\hat{\beta}_2$  es positivo por lo que a medida que aumenta la distancia entre el accidente y la upc más cercana, el riesgo aumenta y  $\hat{\beta}_3$  es negativo indicando que el riesgo de accidente disminuye de este a oeste.

Se usará el modelo 4 para el posterior análisis de los resultados puesto que tiene el menor valor AIC de todos los modelos propuestos.

## 2.4. Análisis sobre la red

Se procede ahora a trabajar sobre la red, ya que tiene más sentido y el análisis es más realista por varias razones:

- Se obtiene una mayor identificación no sólo de puntos críticos sino también de segmentos de carreteras en los cuales se observe una mayor concentración de accidentes.
- Entender la red completa ayuda a mejorar la conectividad y la eficiencia del tráfico, abordando problemas que afectan a múltiples

puntos y tramos.

- Al añadir la red en el análisis, se evita que el análisis se centre en puntos que están ubicados en zonas con muy poca intervención o cantidad de registros (hay accidentes que no suceden en las vías de la carretera).

### **2.4.1. Obtención de la red**

Puesto que en la base proporcionada no se contaba con los datos de los polígonos y nodos de la red, se procedió a obtenerla de manera similar a como se obtuvieron las covariables, se procedió a obtener las calles de la región de estudio mediante la librería `osmnx` de Python, para ello se necesitaron las coordenadas exactas de la región, luego se exportó a un archivo `shapefile` para posteriormente importarlo desde el lenguaje R, el procedimiento descrito se tomó como referencia del libro de Guillermo Rodríguez López, Luis Jara Obregón y Diego Cordero Guzmán [9], específicamente en las páginas 14-18, una vez obtenida la red el siguiente paso fue “limpiarla”, eliminando los nodos sin salida que suelen representar una complejidad para el análisis no sólo en el caso de los procesos puntuales sino en otras más.

Se procede ahora a realizar el análisis del proceso puntual sobre la red, luego de ello se procederá comparar los resultados con los obtenidos anteriormente. Se muestra a continuación en la figura 2.15 las calles (redes) de Eugenio Espejo, en la que se realizará el análisis.

Se ubican sobre esta red los siniestros de tránsito, con ello se tiene una mejor representación de dichos accidentes en la figura 2.16, al igual que antes podemos visualizar que la mayor cantidad de registros se ubican en la parte baja.

### **2.4.2. Intensidad en la red**

Para el cálculo de la función de intensidad o densidad promedio de puntos en la red, al igual que en el caso espacial se pueden considerar dos casos, que la intensidad sea homogénea o no sea homogénea. Según A.Baddeley [1], un proceso puntual  $X$  en  $L$ , tiene una intensidad homo-



Figura 2.15: Red lineal de Eugenio Espejo

génea  $\lambda$  si el número esperado de puntos de  $X$  que caen en  $B \subset L$  es igual a

$$\mathbf{E}(N(X \cap B)) = \lambda l(B) \quad (2.5)$$

donde  $l(B)$  es la longitud total de  $B$ . Un estimador para  $\lambda$  es la intensidad empírica:

$$\bar{\lambda} = N(X)/l(L)$$

Si suponemos que la intensidad en la red no es homogénea, se deberá usar el estimador 1.8, al igual que en el caso espacial, existen varios métodos para escoger un ancho de banda óptimo y también métodos para suavizar el núcleo de la densidad del patrón de puntos. En este documento se trabajará con la banda de ancho para Densidad de Núcleo con Validación Cruzada de Verosimilitud [6], el ancho de banda  $\sigma$  se elige para maximizar el criterio de validación cruzada de verosimilitud de procesos puntuales, cuya estimación es:

$$\text{LCV}(\sigma) = \sum_i \log \hat{\lambda}_{-i}(x_i) - \int_L \hat{\lambda}(u) du \quad (2.6)$$



Figura 2.16: Accidentes de tránsito ubicados sobre la red

donde  $\hat{\lambda}_{-i}(x_i)$  es la estimación de la intensidad suavizada por el núcleo, dejando fuera el punto  $x_i$  con el ancho de banda de suavizado  $\sigma$ , y  $\hat{\lambda}(u)$  es la estimación de la intensidad suavizada por el núcleo en una ubicación espacial  $u$  con el ancho de banda de suavizado  $\sigma$ .

Usando este estimador, para el proceso puntual en la red la banda de ancho es de 44.48011m. Para la suavización del núcleo, nuevamente utilizaremos el cuártico, trabajando con el estimador 1.8, es decir con la corrección de Jones-Diggle, para constantes  $k = 3, 4, 5, 6$ , en la figura obtenemos las siguientes funciones de intensidad.

Se puede ver en la figura 2.17 que para  $k = 3, 4$  se puede apreciar de mejor manera la intensidad del proceso que en el caso espacial, además con  $k = 5, 6$ , la representatividad es mucho mejor ya que se ve de manera más detallada la concentración de los puntos. En conclusión, la intensidad sobre la red es mucho mejor que la espacial tomando en cuenta que el proceso puntual es de accidentes de tránsito tiene más sentido trabajar con esta intensidad.

En lo que sigue, se trabajará con la intensidad del cuarto gráfico ya que

permite apreciar de mejor manera la concentración de puntos por unidad de longitud.

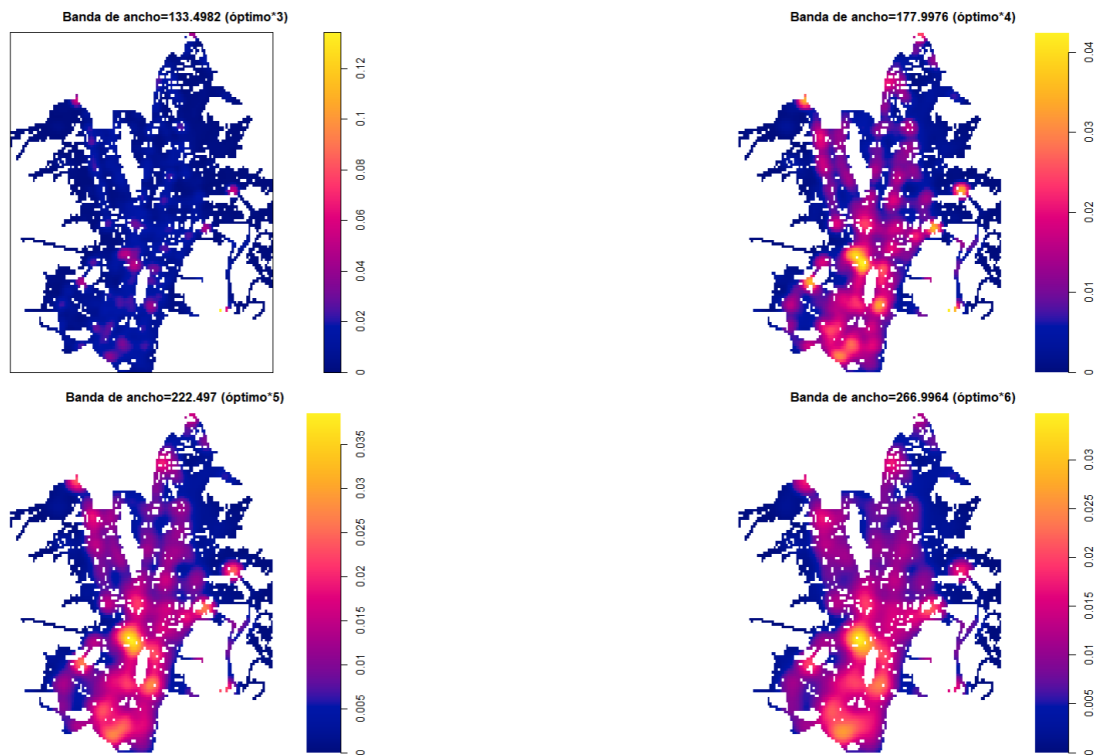


Figura 2.17: Estimación de intensidad con la corrección de borde

### 2.4.3. Estimación de la función $K$

Para la interacción entre puntos, se va a utilizar la función  $K$  para la red antes vista, puesto que no se tiene el método de los cuadrantes para verificar si el proceso es o no homogéneo, se usará el estimador 1.10, podemos observar en la figura 2.18 la gráfica de la función  $K$ , puesto que la función estimada está por debajo de la teórica el proceso es consistente con un proceso de Poisson inhomogéneo.

Además, para analizar si existen posibles agrupamientos se va usar el estimador 1.11.

Se observa en la figura 2.19 la gráfica de la función  $K$  inhomogénea, nuevamente como en el caso espacial se puede notar que hay evidencia de agrupamiento en los accidentes de tránsito.



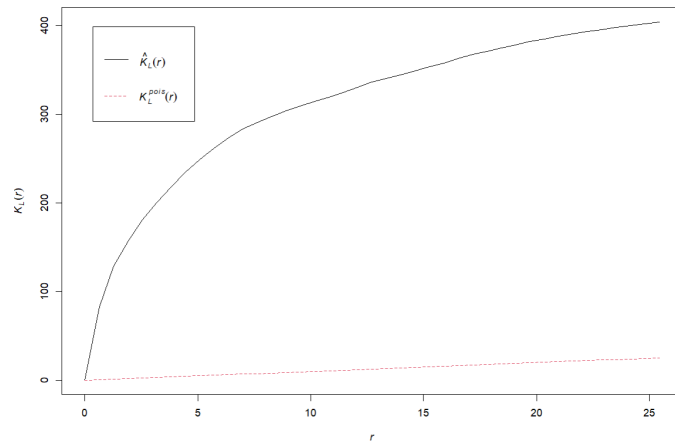


Figura 2.18: Función  $K$  vs  $\hat{K}$

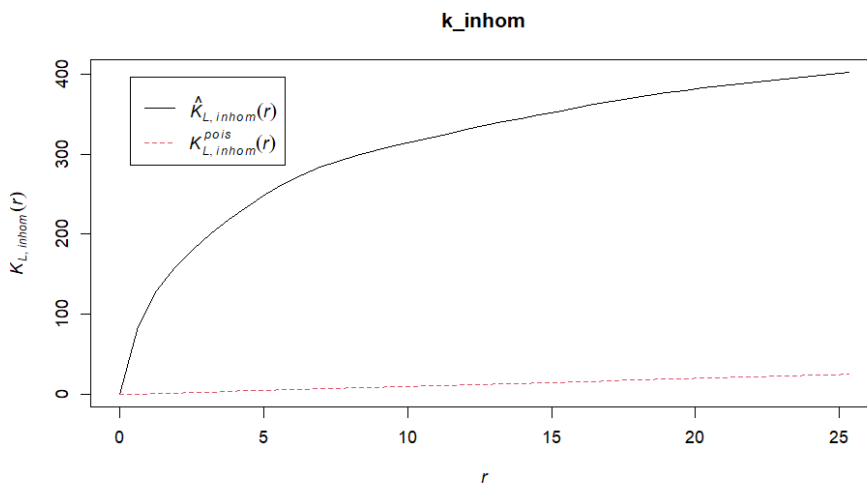


Figura 2.19: Función  $K$  vs  $\hat{K}$  inhomogénea

#### 2.4.4. Dependencia de covariables

Nuevamente, es necesario verificar si la intensidad en la red depende de las covariables proporcionadas. Para ello, se utilizará el test de Berman, pero con una modificación que consiste en agregar la red lineal  $L$ . A diferencia de otros tests, este depende explícitamente de las covariables espaciales. En este caso, se utilizará la distancia del camino más corto por defecto, ya que, como se mencionó anteriormente, la red no es homogénea ni convexa.

En la tabla 2.4 se presentan los resultados obtenidos del test de Berman, las cuatro variables son significativas para la intensidad en la red puesto que el valor-p en cada una de ellas es menor a 0.05, con ello es

razonable ajustar un modelo para el proceso puntual con estas covariables.

<b>Covariable</b>	<b>valor <math>Z_2</math></b>	<b>valor-p</b>
buses	-48.784	<2.2e-16
semaforos	-78.526	<2.2e-16
pare	-23.514	<2.2e-16
upc	-23.442	<2.2e-16

Cuadro 2.4: Test de Berman

Se muestra en la figura 2.20, las covariables espaciales sobre la red Eugenio Espejo

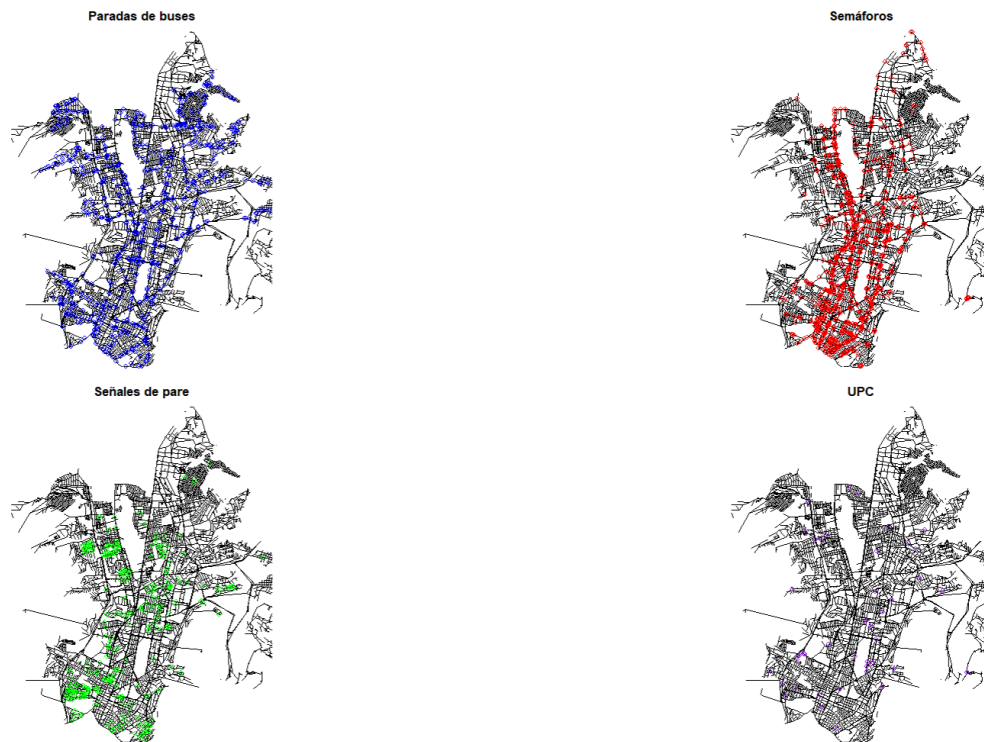


Figura 2.20: Covariables sobre la red de Eugenio Espejo

### 2.4.5. Ajuste del modelo

Hasta este momento, en el análisis de procesos puntuales sobre la red, sólo se ha implementado los modelos clásicos de Poisson homogéneo e inhomogéneo. Según A. Baddeley [3], es más complicado adaptar los modelos de agrupamiento sobre una red lineal debido a que se ha verificado que en muchas redes no existen procesos que dependan sólo de la distancia del camino más corto. Es por ello que se han propuesto nuevas medidas de distancia para estos casos.

Para ajustar un modelo de Poisson, se usará la función **lppm**, el método ajusta un modelo log-lineal definida de la misma manera que 2.1 en el caso espacial, es decir, asumimos que la intensidad tiene la forma log-lineal siguiente:

$$\log(\lambda(u)) = B(u) + \theta^T Z(u) \quad (2.7)$$

aquí,  $Z(u)$  siguen siendo covariables espaciales.

Se propondrán cuatro métodos para el ajuste del modelo, y se seleccionará el mejor de ellos para el posterior análisis sobre los datos de validación.

### Modelo 1: Ubicaciones

Se realiza el modelo de Poisson sobre la base de modelamiento con las covariables  $Z_1, Z_2, Z_3, Z_4$ , en la tabla 2.5 se muestran los resultados obtenidos

	<b>Estimate</b>	<b>S.E.</b>	<b>CI95.lo</b>	<b>CI95.hi</b>	<b>Ztest</b>	<b>Zval</b>
(Intercept)	4.0181	0.0201	-4.0576	-3.9786	***	-199.4415
$Z_1$	-0.0009	0.00007	-0.0011	-0.0009	***	-14.0355
$Z_2$	-0.0028	0.00007	-0.0029	-0.0027	***	-40.6852
$Z_3$	-0.00007	0.00002	-0.0001	-0.00003	**	-3.0764
$Z_4$	0.0004	0.00004	0.0003	0.0005	***	11.4339

Cuadro 2.5: Resultados de la estimación

Así, la estimación del modelo es:

$$\hat{\lambda}(u) = \exp(4,0181 - 0,0009Z_1 - 0,0028Z_2 - 0,00007Z_3 + 0,0004Z_4)$$

$\hat{\theta}_1$  es negativo indicando el riesgo de accidente se reduce con la ubicación de las paradas en la zona,  $\hat{\theta}_2$  es negativo por lo que el riesgo se reduce con la presencia de semáforos cerca de la zona y  $\hat{\theta}_4$  es positivo indicando que con la presencia de estaciones de policía el riesgo aumenta.

### Modelo 2: Coordenadas cartesianas

Se propone ahora el modelo

$$\log(\lambda(u)) = \theta_1x + \theta_2y$$

donde  $x$  e  $y$  son las coordenadas cartesianas de los accidentes en la zona. En la tabla 2.6 se muestran las estimaciones del modelo,  $\hat{\theta}_1$  es negativo indicando que el riesgo de accidente reduce de este a oeste mientras  $\hat{\theta}_2$  es positivo indicando que el riesgo aumenta de norte a sur.

	<b>Estimate</b>	<b>S.E.</b>	<b>CI95.lo</b>	<b>CI95.hi</b>	<b>Ztest</b>	<b>Zval</b>
$x$	-0.00005	0.000005	-0.00006	-0.00004	***	-9.3474
$y$	0.000003	0.0000004	0.000003	0.000004	***	8.2222

Cuadro 2.6: Resultados de la estimación

El modelo estimado es

$$\hat{\lambda}(u) = \exp(-0,00005x + 0,000003y)$$

### Modelo 3: Distancias más cercanas

En este caso, se trabajan con las covariables  $A_1, A_3, A_4$ , no se incluye  $A_2$  puesto que no es significativa para este modelo. En la tabla 2.7 se muestran los resultados de la estimación con estas covariables.

El modelo ajustado es:

$$\hat{\lambda}(u) = \exp(-5,8635 + 16,7005A_1 + 3,1501A_3 + 5,205A_4)$$

	<b>Estimate</b>	<b>S.E.</b>	<b>CI95.lo</b>	<b>CI95.hi</b>	<b>Ztest</b>	<b>Zval</b>
(Intercept)	-5.8635	0.0289	-5.9201	-5.8069	***	-203.0242
$A_1$	16.7005	1.1295	14.4866	18.9144	***	14.7851
$A_3$	3.1501	0.4265	2.1404	3.9861	***	7.385
$A_4$	5.205	0.2275	4.7591	5.6508	***	22.8811

Cuadro 2.7: Resultados de la estimación

Los coeficientes de  $A_1, A_3, A_4$  son positivos, por lo que a medida que incrementa la distancia con la parada de bus, señalización y upc más cercanas, el riesgo de accidente aumenta.

### Modelo 4: Covariables más significativas

Utilizando las covariables más significativas de los anteriores modelos, se propone el siguiente modelo

$$\lambda(u) = \exp(\theta_1x + \theta_2Z_1 + \theta_3Z_2 + \theta_4A_3 + \theta_5A_4)$$

con ello, en la tabla 2.8 se presentan los resultados de la estimación

	<b>Estimate</b>	<b>S.E.</b>	<b>CI95.lo</b>	<b>CI95.hi</b>	<b>Ztest</b>	<b>Zval</b>
$x$	-0.000006	0.00000005	-0.000006	-0.000006	***	-118.9022
$Z_1$	-0.0007	0.00007	-0.0008	-0.0006	***	-9.6469
$Z_2$	-0.0021	0.00007	-0.0022	-0.0019	***	-30.4305
$A_3$	3.9089	0.4268	3.0723	4.7455	***	9.1579
$A_4$	2.9409	0.2186	2.5124	3.3694	***	13.4516

Cuadro 2.8: Resultados de la estimación

así, se obtiene la intensidad estimada del proceso es:

$$\hat{\lambda}(u) = \exp(-0,000006x - 0,0007Z_1 - 0,0021Z_2 + 3,9089A_3 + 2,9409A_4) \quad (2.8)$$

$\hat{\theta}_1$  es positivo por lo que el riesgo de accidente disminuye de este a oeste,  $\hat{\theta}_2$  es negativo por lo que el riesgo disminuye a medida que aumentan las paradas en la zona,  $\hat{\theta}_3$  es negativo indicando que el riesgo se reduce a mayor cantidad de semáforos,  $\hat{\theta}_4$  es positivo indicando que el riesgo de accidente aumenta a medida que la distancia entre el vehículo y la señalización sea mayor y  $\hat{\theta}_5$  es positivo por lo que el riesgo aumenta si la distancia entre el vehículo y la upc más cercana es grande.

Se trabajará con el modelo 4 para el análisis de los resultados, pues su valor AIC es inferior a los demás modelos propuestos.

# Capítulo 3

---

## Resultados, conclusiones y recomendaciones

---

### 3.1. Resultados

#### 3.1.1. Resultados en el espacio

Para la parte espacial, se decidió trabajar con el modelo ajustado 2.4, usando este modelo se tiene que alrededor del 62% de los residuos son menores a 0.3. Se procede a realizar la comparación del modelo estimado con los datos de validación, para ello se simulará una cantidad similar de puntos a los datos de accidentes registrados en 2023 con la intensidad estimada del modelo 2.4.

Se tienen 675 accidentes de tránsito registrados en 2023, se muestra en la figura 3.1 la función de intensidad de este proceso junto con los puntos registrados, nuevamente se usó el suavizamiento cuártico para una mejor representación de la concentración de puntos por unidad de área

Realizamos una simulación de un proceso de Thomas sobre la función de intensidad estimada en el modelo, tomando  $k = 0,085$  como intensidad del proceso de los puntos padres y  $\sigma = 5$  como desviación estándar de la ubicación de los puntos hijos alrededor de sus padres. Con estos valores, se tiene un total de 712 puntos simulados. En la figura 3.2 se muestra el proceso de Thomas simulado.

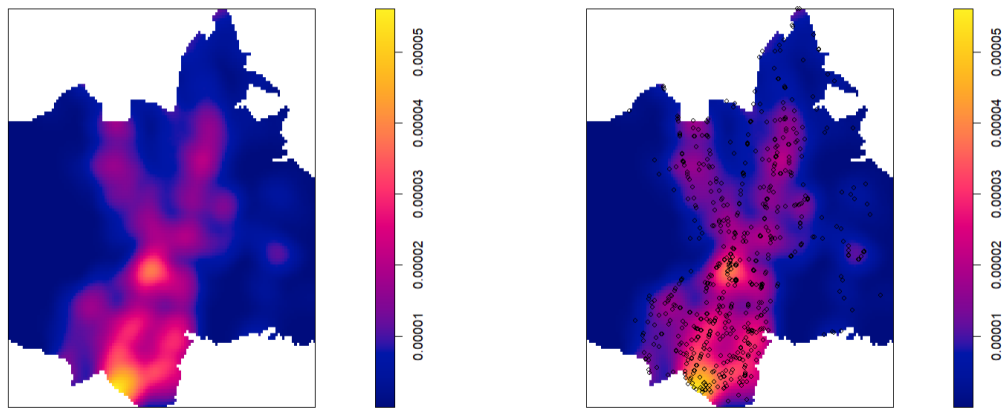


Figura 3.1: Función de intensidad de los accidentes de tránsito en 2023

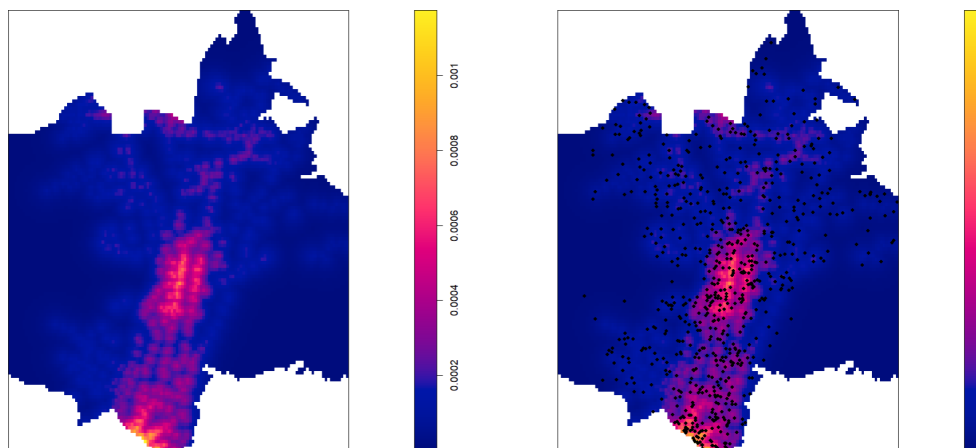


Figura 3.2: Simulación de un proceso de Thomas con la intensidad estimada

En la intensidad estimada por el modelo se puede notar que la concentración de puntos se da en la parte media y baja de la zona, además se puede ver una acumulación en zonas específicas, mientras que en la intensidad de los datos del 2023 la concentración se da en la parte baja de la zona y la representatividad es mejor, por lo que las intensidades no son similares.

En la figura 3.3 se muestra los accidentes de tránsito del año 2023 y los puntos simulados, ambos sobre la intensidad estimada con el modelo 2.4, observamos que en las zonas de mayor concentración los puntos simulados y reales están cercanos, sin embargo, en otras zonas los puntos no coinciden y están muy alejados el uno del otro, por lo que el modelo no ajusta muy bien los datos a nivel espacial sobre estas zonas.



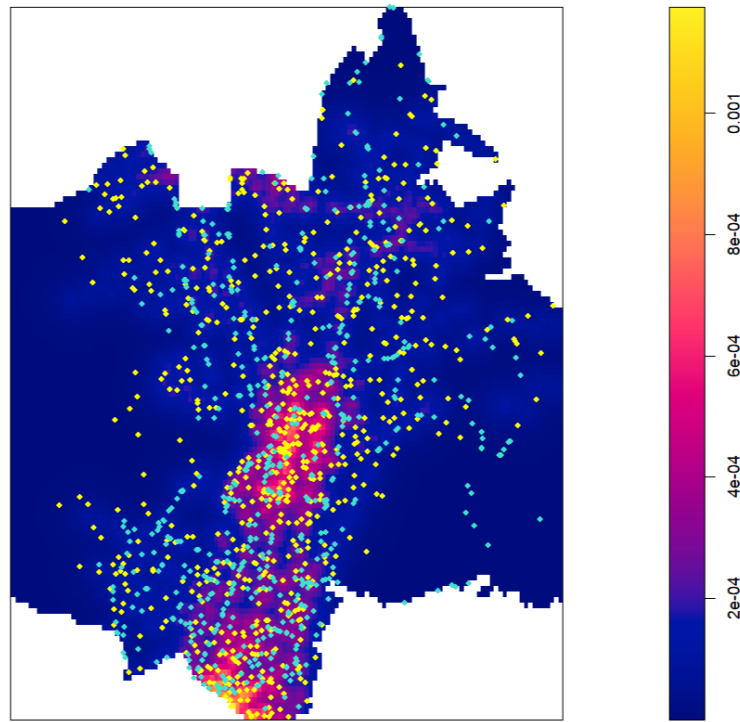


Figura 3.3: Datos reales (turquesa) vs datos simulados (amarillo)

### 3.1.2. Resultados en la red

Sobre la red, se decidió trabajar con el modelo ajustado 2.8, en la figura 3.4 se representan los accidentes de tránsito del año 2023 con su respectiva función de intensidad sobre la red, nuevamente como en el caso espacial se observa concentración de puntos en la parte baja de la zona.

Debido a que la red tiene una gran cantidad de segmentos, por las limitaciones computacionales se realizará una simulación de un proceso de Poisson inhomogéneo con la intensidad estimada del modelo ajustado y un número de puntos similar a los del año 2023.

En la figura 3.5 se han simulado 629 puntos, podemos notar que la intensidad estimada con el modelo 2.8 es similar a la intensidad con la que se trabajó en la figura 2.17 y también a la intensidad de los datos de 2023, por lo que se podría suponer que este modelo ajusta bien nuestros datos tomando en cuenta que tiene más sentido trabajar sobre la red.

Se muestra en la figura 3.6, los datos simulados y los datos del 2023

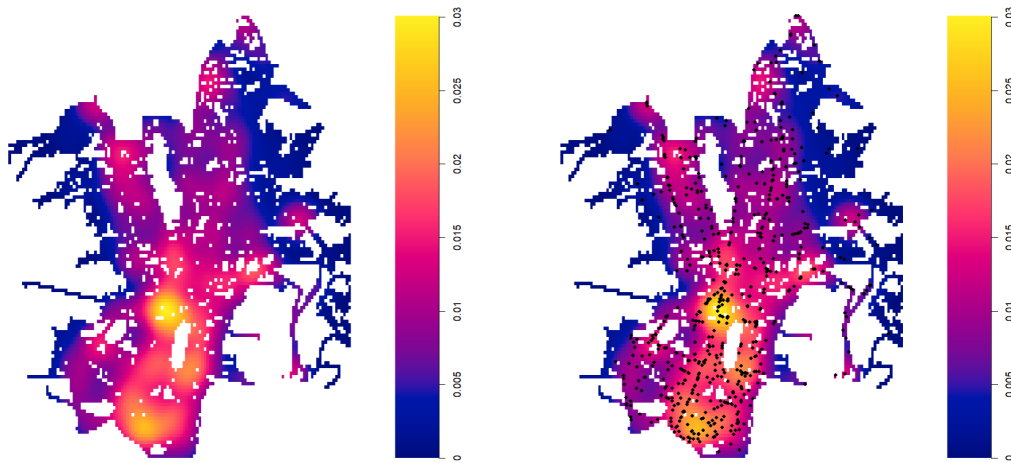


Figura 3.4: Accidentes de tránsito de 2023 sobre la red

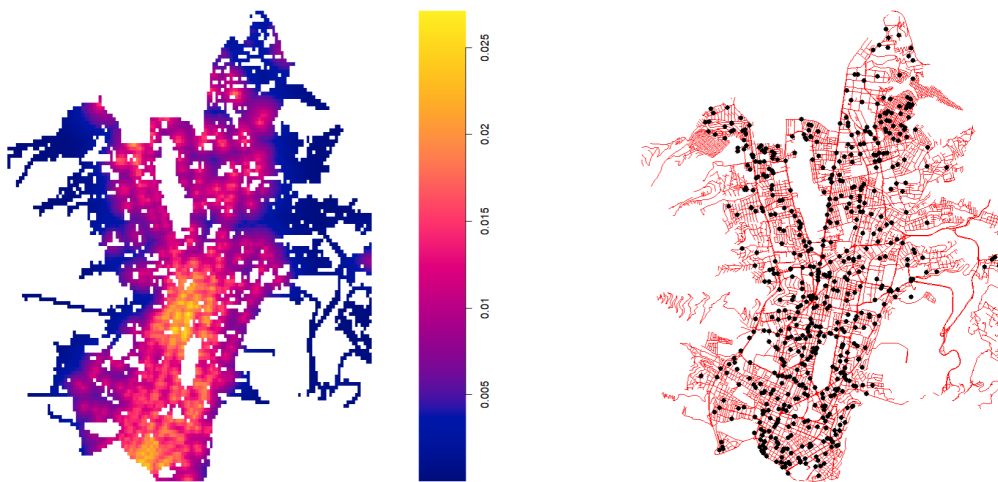


Figura 3.5: Simulación de proceso de Poisson inhomogéneo con intensidad estimada

sobre la red Eugenio Espejo con la intensidad estimada del modelo, los puntos azules corresponden a los datos del 2023 y los puntos negros corresponden a los datos simulados, podemos notar que hay una mejor aproximación de los datos simulados sobre los reales, sobretodo en las zonas donde la intensidad es mayor, por ello podríamos decir que este modelo realiza buenas predicciones de las zonas que se verán más afectadas con accidentes de tránsito en un tiempo posterior.

esto también se puede corroborar comparando las funciones de intensidad de los datos simulados con los datos reales, en la figura 3.7 se muestra en el lado izquierdo la intensidad de los datos de 2023 y en el lado derecho la intensidad de los datos simulados, se puede observar que

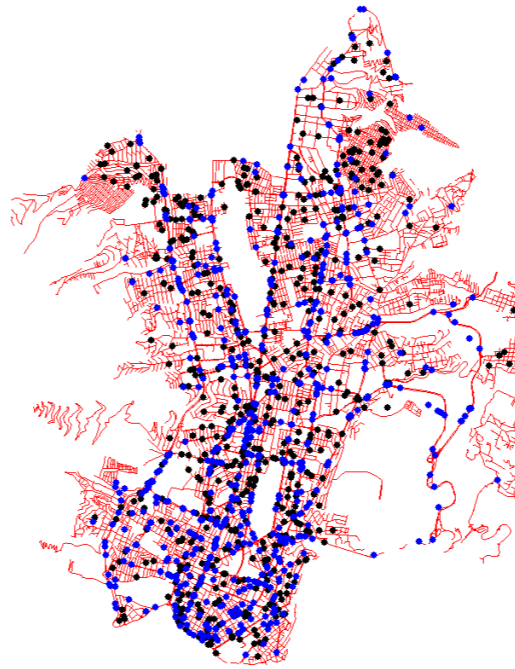


Figura 3.6: Datos simulados (negro) vs datos reales (azul)

ambas funciones son similares por lo que se podría concluir que el modelo realiza buenas estimaciones de los lugares en donde se tendrá mayor cantidad de accidentes de tránsito.

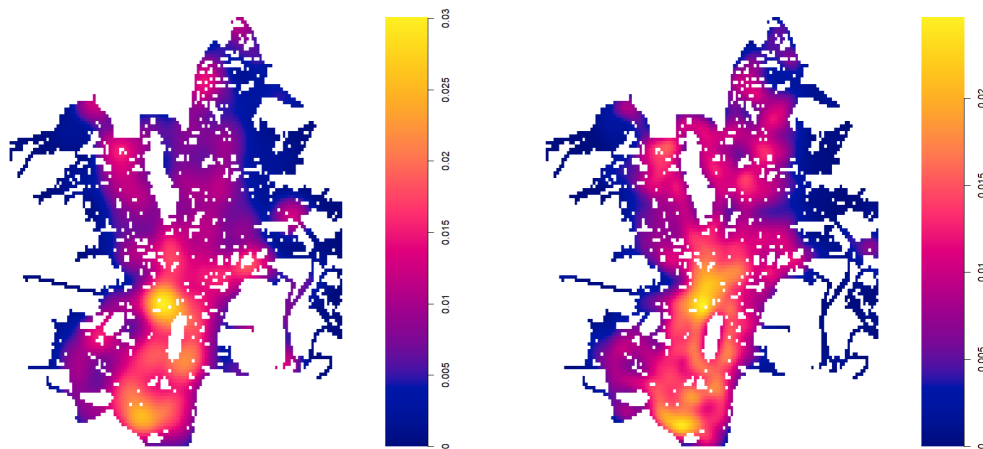


Figura 3.7: Comparación entre funciones de intensidad

Por último, podríamos verificar si con la intensidad estimada del proceso, los datos de 2023 son consistentes o no con un proceso de agrupamiento, en la figura 3.8 se representa la función  $K$  de Ryley inhomogénea de los datos de 2023 con la intensidad estimada del modelo 2.8, podemos notar que a partir de un radio  $r = 50$ , el proceso es consistente

con agrupamiento.

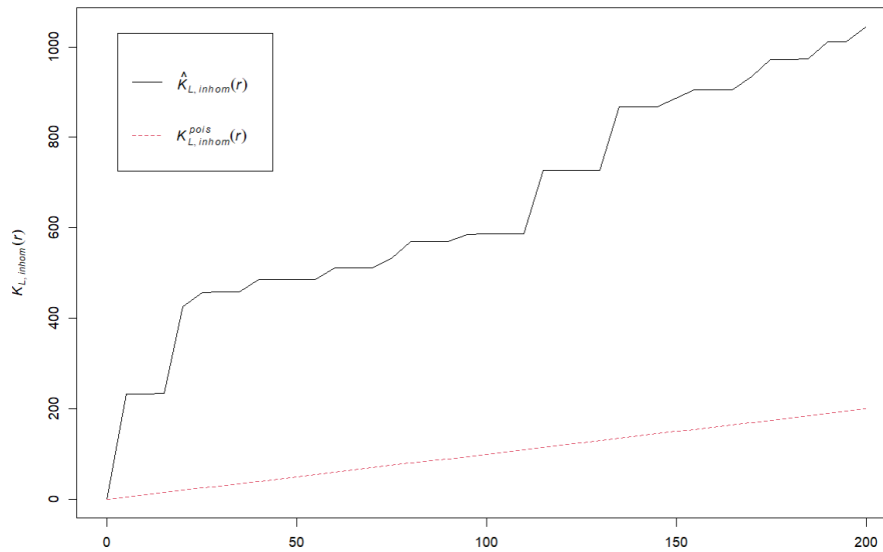


Figura 3.8: Función  $K$  estimada de los accidentes de 2023

## 3.2. Conclusiones

1. Tanto en el análisis espacial como el análisis en la red, las intensidades de todos los datos muestran una mayor concentración de eventos en los sectores de Iñaquito y Mariscal Sucre con 2243 y 1051 accidentes de tránsito respectivamente.
2. La función  $K$  de Rypley inhomogénea confirmó que los accidentes tienden a agruparse espacialmente en ciertas áreas a partir de un radio específico, validando la presencia de clusters de alta siniestralidad.
3. Se obtienen mejores estimaciones de la intensidad cuando se combinan variables de posiciones, distancias y coordenadas, obteniéndose un  $AIC$  menor en comparación a los demás modelos.
4. El análisis en la red permitió identificar no solo puntos críticos, sino también segmentos de carreteras con alta concentración de accidentes.
5. No se puede analizar la validación de los respectivos modelos puesto

que hasta el momento no se ha implementado estas herramientas en el lenguaje R para modelos de tipo Cox o modelos sobre la red.

6. La incorporación de la red de carreteras en el análisis mejoró la precisión al considerar la estructura real de las vías, proporcionando una visión más detallada de los tramos con mayor riesgo.
7. El modelo 2.4 proporciona una estimación adecuada de los datos de modelamiento, pero no identifica bien las áreas críticas al momento de simular nuevos datos y compararlos con los del 2023.
8. El modelo 2.8 permite una identificación más precisa de segmentos peligrosos, además al simular nuevos datos las estimaciones de la intensidad obtenidas son mejores que en el caso espacial cuando se las compara con los datos del 2023.
9. Al ajustar el modelo en la red, la intensidad estimada indica que los sectores de mayor riesgo de accidente de tránsito a futuro son Ñaquito, Mariscal Sucre y la zona Este de Rumipamba.

### 3.3. Recomendaciones

1. Para futuros trabajos, se podría considerar analizar la parte temporal de los datos, se podría obtener mejores estimaciones mediante un análisis espacio-temporal de los accidentes de tránsito.
2. Se podría ser más estricto y exigir que el análisis se lo realice en cada sector de Eugenio Espejo y también trabajar con un proceso puntual marcado por ejemplo incluir la marca género o tipo de vehículo.
3. Para la interacción entre puntos, la función  $K$  puede resultar difícil de interpretar correctamente, una herramienta alternativa es la función de correlación por pares que contiene contribuciones solo de distancias entre puntos iguales a un radio  $r$ , dicha función es definida por

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

4. Si se tiene a disposición, se podría añadir otras covariables que puedan ser importantes y significativas para el modelo, como por ejemplo radares de velocidad, reductores de velocidad, etc.
5. En futuros trabajos, se podría analizar si existe o no correlación espacial entre las covariables, y en el caso de existir aplicar métodos de eliminación de patrones de correlación espacial como el método LASSO.
6. En la figura [2.20](#) se puede notar que en los sectores de mayor riesgo que se estimaron con el modelo [2.8](#) no hay mucha presencia de señalizaciones ni estaciones de policía, por lo que debido a la significancia de estas variables en el modelo sería útil incrementar la cantidad y visibilidad de señales de tráfico y upc en estos sectores, también se podría desarrollar infraestructura que reduzca la velocidad de los vehículos en áreas críticas, como reductores de velocidad, rotondas y pasos peatonales elevados así como fortalecer la presencia de controles policiales en las zonas con mayor frecuencia de accidentes para hacer cumplir las leyes de tráfico.

---

## Referencias bibliográficas

---

- [1] Rolf Turner Adrian Baddeley, Ege Rubak. *Spatial Point Patterns Methodology and Applications with R*. CRC Press Taylor Francis Group, 2016.
- [2] Baddeley. *Adrian Baddeley CSIRO and University of Western Australi. Analysing spatial point patterns in R*. Copyright ©CSIRO 2010, 2010. CSIRO, 2010.
- [3] Rakshit McSwiggan Daviese Baddeley, Nair. *Analysing point patterns on networks a review*. CSIRO, 2020.
- [4] Agencia Nacional de Tránsito. Ecu911 base de emergencias. *Datos abiertos Ecuador*, 2023.
- [5] Helga Stoyan Dietrich Stoyan Janine Illian, Antti Penttinen. *Statistical Analysis and Modelling of Spatial Point Patterns*. WILEY, 2008.
- [6] Nair G. McSwiggan G., Baddeley A. *Estimation of relative risk for events on a linear network*. *Statistics and Computing*. 2019.
- [7] PRIMICIAS. *Accidentes, asesinatos y suicidios, las principales causas de muerte de niños y adolescentes*. 2024.
- [8] Nair G. Rakshit S., Baddeley A. *Efficient Code for Second Order Analysis of Events on a Linear Network*. 2019.
- [9] Cordero D. Rodríguez G., Jara L. *Tecnologías Computacionales Emergentes. Análisis de datos, Investigación e Innovación Tecnológica*. 2023.

- [10] D G Rossiter. *Tutorial: Spatial Point Pattern Analysis*. Cornell University, 2024.
- [11] M. Mehdi Moradi Greg McSwiggan Gopalan Nair Jorge Mateu Adrian Baddeley Suman Rakshit, Tilman Davies. *Fast Kernel Smoothing of Point Patterns on a Large Network using Two-dimensional Convolution*. *Isi International Statistical Review*, 2019.
- [12] M.N.M. van Lieshout. *Theory of Spatial Statistics A Concise Introduction*. CRC Press Taylor Francis Group, 2019.
- [13] Gabriela Yanguéz. Informe de revisión de gestión de la administración zonal eugenio espejo periodo 2022. 2022.