,,,,,,



Análisis espacial de los incendios forestales ocurridos en Bogotá durante el año 2021 utilizando patrones puntuales

Autores:
Michael Andres Cifuentes Mirke
Adriana Lucia Torres Villalobos

Universidad Distrital Francisco José de Caldas
Especialización en Sistemas de Información Geográfica
Bogotá, Colombia
2023

Índice

Página

IV

IV

1. Introducción **1**
1.1. Línea de investigación . . . 2
1.2. Planteamiento del problema . . . 3
1.2.1. Antecedentes del problema . . . 3
1.2.2. Pregunta de investigación . . . 4
1.3. Justificación . . . 5
1.4. Objetivos . . . 5
1.4.1. Objetivo General . . . 5
1.4.2. Objetivos Específicos . . . 5

2. Marco de Referencia **6**
2.1. Marco Conceptual . . . 6
2.2. Marco Teórico . . . 6
2.2.1. Procesos puntuales espaciales . . . 7
2.2.2. Aleatoriedad espacial completa . . . 8
2.2.3. Características de primer y segundo orden . . . 10
2.2.4. Estimación de modelos . . . 12
2.3. Marco Jurídico . . . 13
2.3.1. Decreto 2340 de 1997 - Por el cual se dictan unas medidas para la organización en materia de prevención y mitigación de incendios forestales y se dictan otras disposiciones . . . 13
2.3.2. Ley 1575 de 2012 - Ley General de Bomberos de Colombia . . . 13
2.4. Marco Geográfico . . . 13
2.5. Marco Demográfico . . . 14

3. Metodología **14**
3.1. Método . . . 15
3.2. Técnicas e Instrumentos . . . 15
3.3. Recolección de datos . . . 15

4. Desarrollo de la Propuesta **15**
4.1. Datos . . . 16
4.2. Intensidad . . . 18
4.2.1. Conteo por cuadrantes . . . 18
4.2.2. Estimación Kernel . . . 21
4.3. Correlación . . . 22

4.3.1. Función K de Ripley 23
4.3.2. Función L 24
4.3.3. Función de correlación por pares 25
4.4. Métodos basados en distancias 26
4.4.1. Test de CSR basado en distancias cortas 26
4.4.2. Gráficos exploratorios 27
4.4.3. Función G(h) y F(h) 28
4.5. Visualización de datos 28

5. Impacto y Resultados Esperados **29**

Referencias **32**

Índice de figuras

1. Mapa de Amenaza por Incendios Forestales para Bogotá a Escala 1:25000. Tomado de la página web de IDIGER. 2
2. Tipos de patrones puntuales 7
3. Ubicación de los incendios forestales ocurridos en Bogotá durante el año 2021 16
4. Número de incendios por mes y subclase en Bogotá durante el año 2021 17
5. Estimación de la intensidad de los conteos por cuadrantes y por localidades 19
6. Estimación Kernel para determinar la intensidad de los incendios ocurridos en Bogotá durante el año 2021 22
7. Función K empírica (puntos) y Función K teórica de un proceso Poisson (línea punteada) 23
8. Función L empírica (puntos) y Función L teórica de un proceso Poisson (línea punteada) 24
9. Función de correlación por pares empírica dividida por el estimador d . 25
10. Diagrama de Stienen (izquierda), y Teselación Dirichlet (Derecha) para los incendios forestales ocurridos en Bogotá durante el año 2021 27
11. Función G: función de distribución de la distancia al vecino más cercano (izquierda); Función F: función de espacio vacío 28
12. Vista Principal del Tablero de control Análisis Espacial de los Incendios Forestales en Bogotá - 2021 29

Índice de tablas

Resumen

A lo largo de este trabajo se analizan los incendios forestales ocurridos durante el año 2021 en la ciudad de Bogotá, Colombia; utilizando la metodología de patrones puntuales. A partir de la ubicación espacial y haciendo uso de la distancia euclidiana, se estima la intensidad por cuadrantes regulares y por localidades, las funciones K de Ripley y de correlación por pares g, y además, se realiza un análisis por métodos basados en distancias. En conclusión, los resultados de todas las estimaciones son consistentes con un patrón agregado así como las pruebas de homogeneidad indican una asociación positiva entre los puntos.

Palabras Clave

Incendio forestal, patrón puntual, proceso de Poisson, homogeneidad, intensidad, densidad, correlación, distancia.

1. Introducción

Los incendios son fuegos que se propagan sin control humano, cuando ocurren en la naturaleza se llaman incendios forestales; este término forestal está relacionado con una clasificación tradicional de los usos del suelo, el cual incluye cualquier zona terrestre que no sea de uso urbano ni agrícola [1]. Estos incendios pueden ser considerados como perturbaciones ecológicas de efectos discretos o difusos, graves o destructivos, producidos por fuego de origen natural o antrópico, cuya dinámica responde fundamentalmente a la concurrencia simultánea de tres o más condiciones en un mismo sitio (tipo de vegetación, cantidad de combustible, oxígeno, condiciones meteorológicas, topografía, actividades humanas, entre otras) las cuales se desarrollan sin control ni límites preestablecidos sobre terrenos con alguna clase de cobertura vegetal (nativa, cultivada o inducida), utilizando como fuente de combustible la vegetación viva o muerta y, por el riesgo que representa para los sistemas naturales o sociales, deben prevenirse y extinguirse [2].

De acuerdo con el Insituto Distrital de Gestión de Riesgos y Cambio Climático (IDIGER), en Bogotá durante el período 2010-2020 se han presentado 194 incendios forestales que han afectado 1.212,5 hectáreas, siendo el 2016, el año con mayor área afectada (378 ha), seguido por el 2014 (260 ha) y el 2010 (223 ha). Estos eventos están relacionados con las condiciones de aumento en la temperatura, especialmente durante el primer trimestre de cada año y por la incidencia del Fenómeno "El Niño", lo que genera mayor posibilidad de ocurrencia de incendio forestal [3].

Así mismo, Bogotá cuenta con la Comisión Distrital para la Prevención y Mitigación de Incendios Forestales – CDPMIF, que en conjunto con el Sistema Distrital de Gestión del Riesgo y Cambio Climático - SDGR - CC, han clasificado en tres categorías los eventos forestales dependiento de la facilidad en el control de la conflagración, el tipo de combustible involucrado, el tamaño y los atributos del área afectada:

1. **Quema:** es el fuego, con o sin control y/o límite preestablecido, originado por la actividad humana, que consume cualquier material combustible y puede afectar zonas con cobertura vegetal.

2. **Conato Forestal:** es el fuego que se propaga sin control y consume coberturas vegetales de áreas rurales o componentes de la estructura ecológica principal, cuya área de afectación es de hasta 0.5 hectáreas.

3. **Incendio Forestal:** es el fuego que se propaga sin control y consume coberturas vegetales de áreas rurales o componentes de la estructura ecológica principal, cuya área de afectación es superior a 0.5 hectáreas.

Entre los estudios realizados por IDIGER se identificó que el 31 % del territorio Bogotano se encuentra categorizado en amenaza alta, 27 % en amenaza media y 42 % en

amenaza baja. En general, las zonas de amenaza alta se caracterizan por altas pendientes y abundancia de vegetación, lo cual, unido a las altas temperaturas, genera condiciones ideales para la rápida propagación del fuego [3].

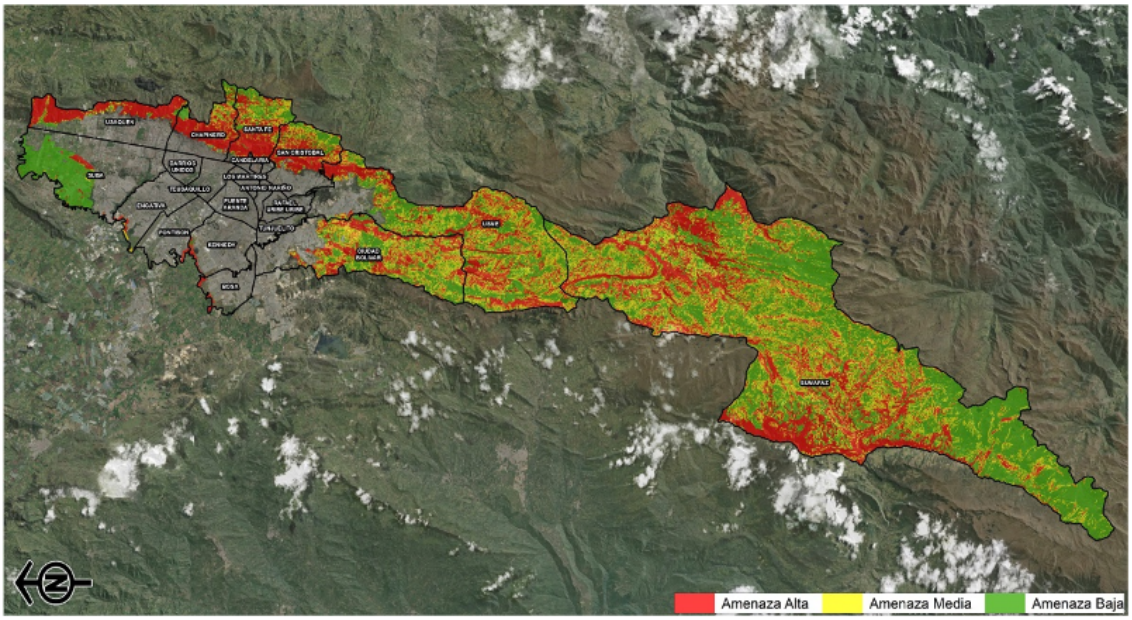


Figura 1: Mapa de Amenaza por Incendios Forestales para Bogotá a Escala 1:25000. Tomado de la página web de IDIGER.
[3]

Como se ha mencionado anteriormente, algunos de los factores que influyen en el comportamiento del fuego son la topografía, el tiempo atmosférico y el combustible. Sin embargo, los factores topográficos son inmodificables al menos en escalas cortas de tiempo; los valores de los factores meteorológicos son cambiantes, pero al contrario que los combustibles, no pueden ser alterados por el hombre. No obstante, pueden ser objeto de predicción, de ahí la importancia de su conocimiento para preveer situaciones críticas.

En el presente trabajo se analizan los incendios forestales que ocurrieron en la ciudad de Bogotá durante el año 2021 siguiendo la metodología de patrones puntuales [4], haciendo uso de los datos abiertos del Cuerpo Oficial Bomberos de Bogotá, y utilizando el lenguaje de programación de R.

1.1. Línea de investigación

El análisis por patrones puntuales se encuentra enfocado al área de geomática y estadística espacial, adscritos a la Especialización de Sistemas de Información Geográfica de la Universidad Distrital Francisco José de Caldas.

1.2. Planteamiento del problema

Son diversos los efectos que el fuego puede provocar en los ecosistemas, y pueden ser tanto positivos como negativos, por lo que es posible encontrar ecosistemas independientes al fuego que, en caso de ocurrir un incendio forestal, el impacto de este es bajo o casi nulo, dado que fisionómicamente poseen extensiones secas discontinuas haciendo que el incendio no se propague en la totalidad del área afectada. Por otro lado, los ecosistemas dependientes del fuego se construyen con base en esta perturbación, debido a que facilitan su propagación y han desarrollado características que les otorgan resistencia como el rebrote, el grosor de la corteza, la disponibilidad de bancos de semillas, entre otros [5].

Un ejemplo de lo mencionado anteriormente son los efectos en el suelo, donde la severidad del fuego genera cambios en las propiedades del mismo, disminuye la diversidad de microrganismos y acelera algunos procesos erosivos. En cuanto al contenido de materia orgánica, esta suele recuperarse aproximadamente 10 años después de la perturbación, para producir las mismas cantidades de materia orgánica que ecosistemas no alterados [6]. Así mismo, el ciclo hidrológico es afectado debido a la alteración de la escorrentía superficial y la infiltración, por pérdida de cobertura a causa del incendio [7].

Cabe resaltar que los ecosistemas terrestres en Colombia están distinguidos por ser en su mayoría sensibles al fuego, por lo que los incendios forestales pueden ser más recurrentes en el territorio, aproximadamente el 64 % del territorio nacional corresponde al combustible más disponible que son los árboles, los cuales pueden durar hasta 100 horas incinerándose, además de poseer una gran cantidad de biomasa aérea de más de 150 toneladas por hectárea.

1.2.1. Antecedentes del problema

En Colombia se han realizado diferentes estudios frente al comportamiento de los incendios forestales en los bosques naturales más susceptibles. Estos estudios corresponden a modelaciones de los incendios, donde se correlacionan diferentes variables como la cobertura vegetal, el balance hídrico, los servicios ecosistémicos, entre otros, con el fin de conocer las tendencias de los bosques [8].

Así mismo, se han usado diferentes métodos para observar la distribución de los incendios a nivel nacional como lo son las anomalías térmicas, que por medio de sensores hace posible analizar la distribución de los incendios y las zonas con mayores frecuencias de perturbaciones en diferentes periodos de tiempo. Esto proporciona información para que entidades del estado implementen diferentes planes de gestión de riesgo para prevenir los incendios o mitigar sus efectos [9].

En la literatura estadística, el estudio de los incendios forestales se ha abordado desde diferentes frentes metodológicos. Algunos de los trabajos desarrollados para analizar estos incendios han propuesto modelos de regresión lineal, en donde se tienen en cuenta

el número de eventos de fuego en la zona de interés, la variabilidad climática y factores adicionales como el balance hídrico, cobertura vegetal, bienes y servicios ecosistémicos, retamo espinoso y causalidad del fuego. Entre sus resultados, destacaron el comportamiento del ciclo hidrológico ya que tuvieron en cuenta dos años con mejor y mayor precipitación (2001 y 2011, respectivamente), donde evidenciaron la importancia de la infiltración en la ocurrencia de los incendios forestales y la dispersión del retamo espinoso. Respecto a la cobertura vegetal, mostraron la influencia que tiene el cambio de uso en el suelo, además de ser una de las causales en la pérdida de los bienes y servicios ecosistémicos provistos por los boques, junto con la propagación rápida de la especie invasora retamos espinoso que por sus habilidades de dispersión y producción de cobertura seca, es un factor de propagación de incendios forestales [8].

En 2016, estudiantes de Ingeniería Topográfica de la Universidad Distrital Francisco José de Caldas presentaron una metodología operativa con imágenes de satélite Landsat TM5 seleccionadas entre 1987 y 1995, aplicaron conceptos y procedimientos de Percepción Remota y Kriging geoestadístico para representar mapas temáticos de la estimación de amenaza de incendios forestales en los Cerros Orientales de la ciudad de Bogotá. Tuvieron en cuenta el índice de área foliar, el índice de diferencia de vegetación normalizada, la evapotranspiración, y el índice de estrés hídrico para conformar un índice de amenaza de incendios forestales de niveles bajo, medio y alto a partir de un análisis temporal. Los datos obtenidos en el procesamiento y análisis geoestadístico a partir del método de validación cruzada, sugiere que los mapas temáticos generados mediante un proceso de interpolación a una escala regional, puede ser una herramienta útil como aproximación y referencia para predecir con un alto grado de acierto las temperaturas del aire, la evapotranspiración potencial con respecto a estaciones climatológicas principales y ordinarias [10].

En febrero de 2022, dos estudiantes de la Universidad Libre del Pregrado en Ingeniería Ambiental presentaron un método para predecir la dirección de dispersión del material particulado considerando que los incendios forestales liberan este material. Tuvieron en cuenta la predominancia en el comportamiento del viento para cada mes del año, evaluaron datos de radiosondeos del período de estudio comprendido entre 2008 y 2018, y recolectaron una base de incendios forestales ocurridos en el mismo período de tiempo. Generaron mapas Skew T y Cross Section por medio del software RAOB 6.9, su predicción fue acertada en diez de doce meses para la variable dirección [11].

1.2.2. Pregunta de investigación

¿Hay alguna estructura en la forma en cómo se organizan los incendios en la ciudad de Bogotán durante el año 2021 o son eventos completamente aleatorios?

1.3. Justificación

Como es de conocimiento los incendios forestales pueden arrasar bosques y paisajes agravando la situación medioambiental y climática, esto provoca un aumento en las temperaturas, los períodos de sequía tienden a ser más prolongados y generan olas de calor, empeorando el cambio climático y este a su vez, intensifica la aparición de incendios. Por lo tanto, se hace imprescindible la aplicación inmediata de medidas que describan, caractericen y modelen el comportamiento de estos incendios para que las instituciones correspondientes puedan preparse mejor para estos escenarios.

Cabe resaltar que los bosques alto andinos son considerados ecosistemas estratégicos de alta montaña con gran importancia a nivel nacional e internacional gracias a su alto grado de biodiversidad, endemismo, oferta de servicos ecosistémicos y beneficios al ser humano, tanto a niveles regionales como globales. Por lo anterior, el monitoreo y seguimiento a los cambios de este bioma se hacen cada vez más relevantes en las estrategias de monitoreo integrado de ecosistemas del país, como lo establece la ´Estrategia para monitoreo integrado de ecosistemas de alta montaña de Colombia´ y a nivel internacional el ´Panel intergubernamental para la biodiversidad y los servicios ecosistémicos´ [12].

1.4. Objetivos

1.4.1. Objetivo General

Determinar el comportamiento de los incendios en la ciudad de Bogotá durante el año 2021.

1.4.2. Objetivos Específicos

1. Realizar un análisis exploratorio de los incendios forestales en 2021 de la ciudad de Bogotá.

2. Estudiar las propiedades de primer y segundo orden del patrón puntual.

3. Identificar la naturaleza de la distribución de los puntos.

2. Marco de Referencia

2.1. Marco Conceptual

Para comprender los términos asociados a la metodología de patrones puntuales e incendios forestales, se definen los siguientes conceptos:

- *Incendio:*. Fuego de grandes proporciones que arde de forma fortuita o provocada y destruye cosas que no están destinadas a quemarse.
- *Evento forestal:* Cualquier evento con fuego sin control que afecta zonas con cobertura vegetal y ocasiona daños ambientales, económicos y sociales y que puede ser de tres tipos: quema, conato forestal e incendio forestal.
- *Quema:* Fuego con o sin control que puede tener un límite preestablecido, que es originado por la actividad humana y consume cualquier material combustible y que puede afectar zonas con cobertura vegetal.
- *Conato forestal:* Es el fuego que se propaga sin control y consume coberturas vegetales de áreas rurales o componentes de la estructura ecológica principal cuya área de afectación es de hasta media hectárea.
- *Incendio forestal - IF:* Es el fuego que se propaga sin control y que consume coberturas vegetales de áreas rurales o componentes de la estructura ecológica principal cuya área de afectación es superior a media hectárea. Existen tres tipos: Superficiales, Subterráneos y de Copa.
- *Patrón puntual:* está formado por un conjunto de n eventos (en nuestro estudio casos son incendios forestales) que han ocurrido sobre unas determinadas coordenadas $\{(x_1, y_1), ..., (x_n, y_n)\}$ del plano.
- *Temperatura:* Grado o nivel térmico de un cuerpo o de la atmósfera.
- *Escorrentía:* Corriente de agua que rebosa su depósito o cauce natural o artificial.

2.2. Marco Teórico

Un proceso espacial corresponde a un proceso estocástico $Z(x) : x \in D$ en el que $x \in \mathcal{R}^n$ representa una ubicación en el espacio euclidiano n - dimensional, $Z(x)$ es una variable aleatoria en la ubicación x, la cual varía sobre un conjunto de índices $D \subset \mathcal{R}^n$ [13]. De acuerdo a las características del dominio D existen tres grupos: [14]:

1. **Geoestadística:** El conjunto D es fijo y continuo, por lo tanto la variable Z existe en cualquier punto del espacio D, y los puntos en el espacio no son estocásticos, en algunos casos son seleccionados siguiendo un diseño muestral o son establecidos por el investigador al inicio de la medición.

2. **Datos de área:** El conjunto D es fijo y discreto. Para este caso, las ubicaciones son represetandas por medio de regiones.

3. **Patrones Puntuales:** El conjunto D es aleatorio, los datos son generados como una realización parcial de un proceso estócastico subyacente que es en sí mismo objeto de interés.

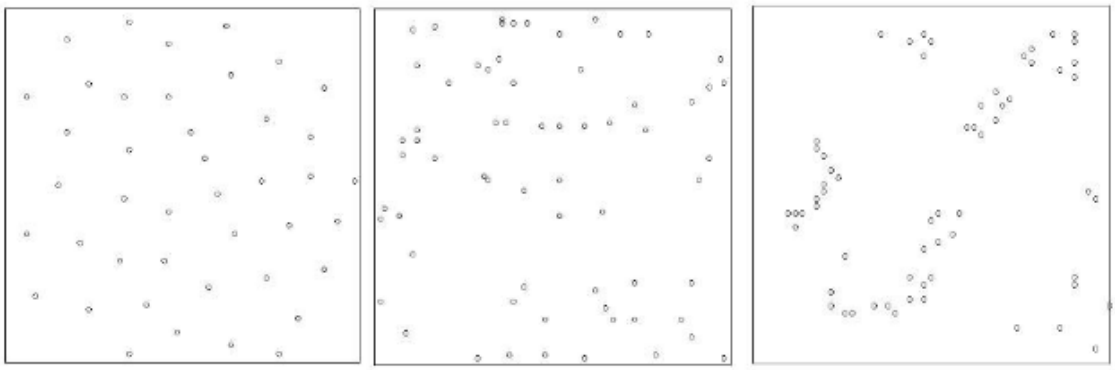
2.2.1. Procesos puntuales espaciales

Un patrón puntual es una realización concreta de un proceso puntual $X \subset R^n$, el cual es un conjunto de puntos en el espacio n-dimensional [4],

$$\mathbf{x} = \{x_1, x_2, ...x_n\}$$

donde x_i son puntos (ubicaciones) en R^n; en todo el documento se considera el caso $n = 2$.

En el análisis de patrones puntuales un paso preliminar consiste en realizar una gráfica del patrón sobre el espacio y así caracterizar el patrón cualitativamente. Dependiendo de la distribución de los puntos en el espacio el patrón se puede clasificar en regular, completamente aleatorio o agregado:



(a) Regular. (b) Aleatorio. (c) Agregado.

Figura 2: Tipos de patrones puntuales
[15]

En la Figura 2a se muestra un patrón regular (inhibición) que corresponde a la ubicación de 42 centros de células observadas bajo microscopía óptica en una sección histológica, la Figura 2b muestra un patrón aleatorio que corresponde a la ubicación de 65 árboles japoneses de pino negro en una región cuadrada en un bosque natural y la Figura 2c muestra un patrón agrupado que corresponde a la ubicación de 62 plántulas y árboles jóvenes de secuoyas de California. Estos datos fueron obtenidos del paquete spatsat de R.

2.2.2. Aleatoriedad espacial completa

En la práctica cuando se tiene un patrón puntual espacial, en principio se asume que no existe ningún tipo de configuración particular en el espacio y que el patrón es completamente aleatorio (Complete Spatial Randomness CSR), esto quiere decir que el patrón sigue un proceso de Poisson Homogéneo, el cual tiene dos principales características:

1. **Homogeneidad:** Sea $n = N(X)$ el número de puntos del patrón. Si X es un patrón puntual y A es una región en el espacio, entonces el número de puntos del patrón que caen dentro de la región A es $N(X \cap A)$. El número de eventos en cualquier región A del plano con área |A| sigue una distribución Poisson con media $\lambda|A|$,

   $$E[N(X \cap A)] = \lambda|A|,$$

   donde λ es el número medio de eventos o puntos del patrón por unidad de área y se define como la intensidad del proceso puntual [4].

2. **Independencia:** Para las regiones $A_1, A_2, ..., A_m$ que no se superponen, se cumple que $N(X \cap A_1), N(X \cap A_2), ..., N(X \cap A_m)$ son variables aleatorias independientes.

La hipótesis nula en las pruebas estadísticas para determinar si existe o no algún tipo de estructura particular subyacente, es que se cumplan las dos propiedades anteriores, es decir que exista aleatoriedad espacial completa. Las pruebas de la hipótesis CSR en general se basan en el conteo de los eventos sobre subregiones que no se superponen o alternativamente utilizan medidas de distancias entre los eventos.

Cuando no se cumple la propiedad de homogeneidad, el patrón tiene una inensidad que varía en el espacio y se llama no homogéneo. En este caso se tiene que el valor esperado de eventos en una región con área $|A|$ es:

$$E[N(X \cap A)] = Z \int_A \lambda(u)du, A \subset R^2, u \in A$$

2.2.2.1 Prueba χ^2

La idea principal de esta prueba consiste en dividir en subregiones de igual área a la región de estudio de tal manera que las subregiones no se superpongan unas con otras.

Sean $A_1, A_2, ..., A_m$, las m subregiones en las que se ha dividido la región de estudio A, entonces se satisface que $A_1 \cup A_2 \cup ... \cup A_m = A$ Dado que el número de puntos en cada subregión $n_j = N(X \cap A_j)$ para $j = 1, ...m$. Luego, si la hipótesis CSR es cierta, los n_j son realizaciones de variables aleatorias de Poisson independientes con valores esperados $\mu_j = \lambda|A_j|$, donde λ es la intensidad desconocida y $|A_j|$ el área de la región A_j, por lo tanto si todas las regiones tienen igual área $|A_j| = a$, los conteos n_j

tienen media igual a λ_a. Esta prueba puede ser utilizada para probar un buen ajuste a la distribución de Poisson asumiendo homogeneidad o para probar homogeneidad asumiendo independencia. Dados

$A_j \rightarrow$ Subregiones en los que se divide la región de estudio A con $j = 1, ..., m$

$n_j \rightarrow$ Número de puntos en la subregión A_j.

$n = \sum_j n_j \rightarrow$ Número total de puntos en la región de estudio.

$|A_j| = a_j \rightarrow$ Área de la subregión A_j

$|A| = \sum_j a_j \rightarrow$ Área total de la región de estudio.

$\hat{\lambda} = \frac{n}{|A|} \rightarrow$ Intensidad estimada de un área A.

$e_j = \hat{\lambda} a_j \rightarrow$ Valor esperado de puntos para la subregión A_j.

El estadístico de prueba se define como

$Z^2 = \Sigma_j \frac{(observado - esperado)^2}{esperado}$,

lo cual bajo la hipótesis de que los conteos en las subregiones son independientes, se reduce a

$Z^2 = \sum_j \frac{(n_j - e_j)^2}{e_j} = \sum_j \frac{(n_j - \bar{\lambda} a_j)^2}{\bar{\lambda} a_j} = \sum_j \frac{(n_j - n/m)^2}{n/m}$,

este estadístico sigue una distribución X^2_{m-1} bajo la hipótesis nula, por lo cual para valores $Z^2 > \chi^2_{(m-1),\alpha}$ para un nivel de significancia α se rechaza la hipótesis de que el número de eventos en las subregiones tienen la misma media. En esta prueba es necesario escoger cuidadosamente la forma y el número de subregiones, ya que esta prueba es sensible a estos valores influenciando los resultados.

El paquete spatstat de R cuenta con la función quadrat.test, la cual encuentra el valor-p con dos métodos [4]:

1. Chisq: Realiza el cálculo del estadístico y se obtiene el valor-p. Un requisito para usar este método es que el número de puntos en cada subregión sea mayor a 5.

2. Monte Carlo: Genera un gran número de patrones puntuales CSR (por default el número de simulaciones es 1999), se calcula el estadístico y el valor-p se determina comparando la estadística Z^2 para el patrón de puntos observado, con los valores obtenidos de las simulaciones. Con este método no es necesario que los conteos sean mayores a 5.

2.2.2.2 Métodos basados en distancias

Sea d_{ij} la distancia entre dos eventos localizados en x_i y x_j, $d_{ij} = ||x_i - x_j||$. La distancia entre el evento ubicado en x_i y el evento más cercano a este, es llamada la distancia

del vecino más cercano y se denota por d_i [14]. Un método que utiliza la distancia del vecino más cercano es la estimación del índice de Clark- Evans [4].

Esta prueba se basa en el promedio de las distancias d_i para q puntos muestreados aleatoriamente del patrón puntual (o para todos los puntos) y divide este por el valor esperado $E(D)$ para un proceso completamente aleatorio con la misma intensidad, con lo cual se obtiene un índice de regularidad. En un proceso de Poisson de intensidad λ la distancia esperada al vecino más cercano es $E(D) = \frac{1}{(2\sqrt{\lambda})}$, así el índice de Clark-Evans [4] es

$$R = \frac{\bar{d}}{E(D)} = \frac{2\sqrt{\hat{\lambda}}}{q} \sum_{i=1}^{q} d_i,$$

donde q es el número de puntos muestreados y $\lambda b = n/\|A\|$ es la estimación de la intensidad para el patrón puntual completo de n puntos en una región con área $\|A\|$. Si $R = 1$ el patrón es completamente aleatorio, $R > 1$ sugiere que es regular mientras que $R < 1$ sugiere que es agrupado. Para un patrón puntual no homogéneo se puede dar el caso $R < 1$, lo cual puede conducir a conclusiones erróneas [4].

Otro de los métodos clásicos es el índice de Hopkins-Skellam [4], el cual se define como

$$HS = \frac{\sum_{i=1}^{q} d_i^2}{\sum_{i=1}^{q} e_i^2}$$

donde $e_j = min_j\|u - x_j\|$ es la distancia de una ubicación u dentro de la ventana al punto más cercano del patrón (distancia de espacio-vacío). Si $HS = 1$ el patrón es completamente aleatorio, $HS > 1$ sugiere que es regular mientras que $HS < 1$ sugiere que es agrupado. Este índice es menos sensible a los efectos de borde y la no homogeneidad espacial [4]

2.2.3. Características de primer y segundo orden

2.2.3.1 Primer orden

El primer paso para el análisis exploratorio es la intensidad del patrón, esta medida describe el número esperado de eventos por unidad de área. La densidad empírica de los puntos está dada por la siguiente expresión:

$$\bar{\lambda} = \frac{n(x)}{\|W\|}$$

donde n(x) corresponde al número de puntos en la región de análisis y $\|W\|$ su respectiva área.

Estimación de la función de intensidad por suavizamiento de kernel

La estimación no paramétrica de la función de intensidad se realiza a través del estimador de kernel propuesto por Diggle y dado por la siguiente expresión:

$$\tilde{\lambda(u)} = \sum_{i=1}^{n} \frac{k(u-x_i)}{e(x_i)}$$

donde u es la localización dentro de la ventana de observación; W representa la Ventana de observación; $k(u)$ representa la función kernel que debe ser una densidad de probabilidad $k(u) \geq 0$ para todas las localizaciones u, para el desarrollo de este trabajo se utilizará el kernel gausiano y $e(u) = \int_W k(u-v)dv$ es la corrección por sesgo debido a efectos de borde.

Método de conteo por cuadrantes para Aleatoriedad espacial completa (CSR)

Este método consiste en dividir la ventana de observación W en subregiones $B_1, ..., B_m$ llamadas cuadrantes que, en principio, se consideran de igual área. A partir de esto, se realiza el conteo de los puntos que caen en cada subregión $n_j = n(X, \cap B_j)$ para $j = 1, ..., m$. Al tratarse de intensidad homogénea, los conteos deberían ser iguales o cercanos en cada subregión, por el contrario, los cambios fuertes en los conteos pueden sugerir tendencias espaciales.

Con el objetivo de contrastar la hipótesis de homogeneidad

H0 : La intensidad es homogénea (CSR) H1 : La intensidad no es homogénea

Se aplica la prueba X^2 (Chi-cuadrado), dado el número total de puntos $n = \sum_j n_j$, el área total de la ventana de observación $a = \sum_j a_j$, la intensidad estimada $\bar{\lambda} = n/a$ y el número esperado de conteos en el cuadrante B_j es $e_j = \bar{\lambda} a_j = n a_j / a$. El estadístico esta dado por:

$X^2_{Calculado} = \sum_j \frac{(observados - esperados)^2}{esperados} = \sum_j \frac{(n_j - e_j)^2}{e_j} = \sum_j \frac{(n_j - \bar{\lambda}_{aj})^2}{(\bar{\lambda}_{aj})^2}$

El contraste se realiza a través del valor crítico donde se rechaza la hipótesis nula si

$X^2_{Calculado} > X^2_{Crtico,gl}$,

donde gl representa los grados de libertad dados por el número de cuadrantes de igual área resultado de la partición menos uno $(n-1)$.

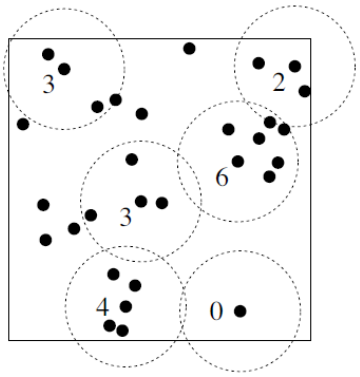
2.2.3.2 Segundo orden

Por medio de las funciones K, g, G y L se determina el comportamiento del patrón, si presenta agregaciones o dispersión en una distancia determinada.

Función K - Distancia entre puntos

La función K propuesta por Ripley (1976), permite clasificar un patrón de puntos como aleatorio, agrupado o regular, de esta manera, la función es cero si el patrón de puntos es completamente aleatorio. La figura 3 presenta la geometría de la función que busca encontrar posibles agrupaciones de un patrón de puntos dentro de una circunferencia de radio r.

$K_{inhom}(r) = \pi r^2$



[4]

Función g - Correlación por pares

La función de correlación por pares contiene contribuciones solo de distancias entre puntos iguales a r. Se obtiene a partir de la siguiente expresión:

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

donde, $K0(r)$ se calcula a través de la derivada de la función K con respecto a r

Esta función consiste en dibujar un circulo de radio r centrado en un punto del proceso y contar el número de otros puntos que caen en el círculo

Función G - Distancia al vecino más cercano

La función mide la distribución de distancias desde un evento arbitrario hasta su evento más cercano, es decir, utiliza las distancias del vecino más cercano. La estimación de esta función esta dada por la siguiente expresión:

$G_{inhom}(r) = 1 - E[\prod_{x_1 \in X \cap b(u,r)}(1 - \frac{\lambda_{min}}{\lambda_(x_i)})|X]$

Función L

La función L surge como la transformación de la función K, esta dada por la siguiente expresión:

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

2.2.4. Estimación de modelos

Modelos de Neyman-Scott para patrones agrupados es un proceso de agrupación, se forma mediante un procedimiento de dos pasos. En el primer paso, se genera un proceso de puntos "padres", en el segundo paso, cada punto "padre"y_i da lugar a un patrón aleatorio de puntos de "descendencia"x_{ij}. También se supone que el punto padre fue reemplazado por su descendencia. El conjunto que consta de todos los puntos de descendencia, independientemente de su origen, forma el proceso de punto de agrupación

X. El conjunto formado por todos los puntos de descendencia, independientemente de su origen, forma el proceso de punto de agrupación X.

2.3. Marco Jurídico

2.3.1. Decreto 2340 de 1997 - Por el cual se dictan unas medidas para la organización en materia de prevención y mitigación de incendios forestales y se dictan otras disposiciones

Artículo 3: Créase la comisión nacional asesora para la prevención y mitigación de incendios forestales adscrita al Ministerio del Medio Ambiente y cuyo objeto será el de servir de órgano asesor en materia de incendios forestales al sistema nacional para la prevención y atención de desastres y al sistema nacional ambiental, SINA.

2.3.2. Ley 1575 de 2012 - Ley General de Bomberos de Colombia

Artículo 5: Créase la Dirección Nacional de Bomberos, como Unidad Administrativa Especial del orden nacional, con personería jurídica, adscrita al Ministerio del Interior, con autonomía administrativa, financiera y patrimonio propio, cuya sede será en Bogotá, D. C.

La función del Director Nacional de Bomberos deberá ser cumplida por un Oficial de Bomberos de máximo grado de reconocida trayectoria institucional bomberil, nombrado por el Presidente de la República.

El oficial que asuma dicho cargo, tendrá el grado superior de Capitán en Jefe, las insignias y demás distintivos serán reglamentados por la Junta Nacional de Bomberos de Colombia.

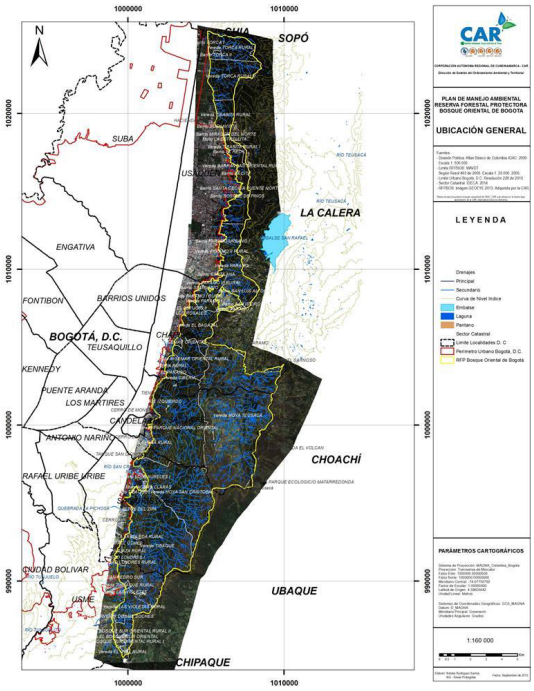
Todas las instituciones bomberiles del país, oficiales, aeronáuticos y voluntarios, así como sus miembros estarán bajo coordinación operativa de la Dirección Nacional de Bomberos de Colombia.

2.4. Marco Geográfico

La reserva forestal protectora Bosque Oriental de Bogotá es una cadena de montañas localizada en el costado oriental del casco urbano bogotano, con una extensión aproximada de 13.142,11 hectáreas, de acuerdo con la cartografía de la Resolución 463 de 2005, y se eleva desde los 2650 hasta los 3600 m.s.n.m. Incluye área rural de las localidades Usme, San Cristóbal, Santa Fe, Chapinero y Usaquén. [16]

La reserva se convierte en una barrera geográfica de la Sabana de Bogotá, colindando con los municipios de Chía, Sopó, La Calera, Choachí, Ubaque y Chipaque (Figura 1). En su interior existen 27 veredas catastrales, las cuales son: Torca I, Torca II rural,

Tibabita rural, Tibabita rural I, Barrancas oriental rural, Páramo, Páramo II rural, Páramo III rural, Páramo II, Páramo I, Páramo I rural, El Bagazal, Ingemar Oriental rural, Hoya Teusacá, Siberia, Parque Nacional Oriental, Hoya San Cristobal, Tibaque, La Arboleda rural, Chiguaza rural, Ciudad Londres rural, Tihuaque rural, Las Violetas rural, El Porvenir de los Soches, Bosque Sur Oriental rural I, Bosque Sur Oriental rural II y El Uval rural. De igual forma, dentro de la reserva se localizan cinco asentamientos informales que son: Bosque Sur Oriental, Lomitas, La Capilla, Altos de Serrezuela y Mirador del Norte. [16]



[16]

2.5. Marco Demográfico

Con una población aproximada de 7'901.653 habitantes con 3'783.875 hombres y 4'117.778 mujeres, Bogotá tienen una densidad poblacional de aproximadamente 4.146 habitantes por kilómetro cuadrado, de acuerdo con las cifras presentadas por el Departamento Administrativo Nacional de Estadística (DANE). Sin embargo, otros estudios sobre la base de los datos del censo 2005 arrojan que dentro del perímetro urbano la ciudad tiene una densidad promedio de 16.779 habitantes por kilómetro cuadrado, ubicados principalmente en las localidades de Ciudad Bolívar, Bosa, Kennedy, Engativá y Suba en donde algunos barrios llegan a alcanzar densidades superiores a los 50.000 $\frac{hab}{km^2}$.

3. Metodología

Para realizar el análisis de los incendios forestales de Bogotá se propone una metodología de patrones puntuales que describa, caracterice y determine el comportamiento de los

eventos de interés.

3.1. Método

Métodos estadísticos: La estádistica descriptiva brindará indicios sobre la tendencia, agrupación, dispersión, y comportamiento de los datos. Una vez se aplique el análisis de patrones puntuales, se podrá caracterizar las apariciones de los incendios.

3.2. Técnicas e Instrumentos

La fuente de información es secundaria ya que aparece en la página web de datos abiertos de la Unidad de Bomberos de Bogotá. En primera instancia se realizarán las técnicas de preparación de datos: limpieza, reducción, integración, normalización y transformación de datos, utilizando el software R, Python y ArcgIS 10.8. Adicionalmente, la visualización de resultados se llevará en un dashboard (tablero de control) en ArcGIS Online.

3.3. Recolección de datos

Los datos son obtenidos de la página de datos abiertos de la Unidad de Bomberos de Bogotá y contienen toda la información relacionada a los incendios atendidos por ellos para el año 2021.

Son datos de tipo espacial ya que cada evento se encuentra georreferenciado, la base además incluye variables de tipo nominal que permite identificar el tipo de servicio efectuado por el cuerpo de bomberos, y también cuenta con una variable que indica en qué fecha sucedió el evento.

Fuente: https://datosabiertos.bogota.gov.co/organization/bomberos

4. Desarrollo de la Propuesta

A lo largo del tiempo, se han desarrollado varios métodos y medidas para analizar, modelar, visualizar e interpretar las propiedades de los patrones puntuales: el uso de estadísticas descriptivas brinda un resumen de las características básicas del patrón, pero su alcance es limitado por lo que se han desarrollado técnicas más poderosas basadas en la densidad o en la distancia, según las propiedades que se estén considerando [17].

Los métodos basados en densidades se centran en las propiedades de primer orden de un conjunto de datos, es decir, la variación en las ubicaciones individuales de los puntos del

conjunto de datos en el área de interés; mientras que, los métodos basados en distancias se basan en las propiedades de segundo orden, en pocas palabras, en la interacción entre puntos, en cómo se distribuyen o en si forman alguna tendencia.

4.1. Datos

La base de datos contiene 229 registros de incendios forestales atendidos por la Unidad de Bomberos de la ciudad de Bogotá en el año 2021 (Figura 4). La información está desagregada por fecha, estación que atendió el llamado, tipo de servicio (para este caso todos corresponden a incendios forestales), subclase (Quemas, Conatos de 0 hasta 5000 m2, Incendios más de 5000 m2), origen de la causa del incendio y las coordenadas en dónde ocurrió el evento. Utilizando el software ArcGIS 10.8 se realiza la transformación de coordenadas geográficas a coordenadas planas.

En la Figura 3 se presenta la ubicación de los incendios forestales atendidos por la Unidad de Bomberos. Este mapa da un indicio de que los puntos no se encuentran totalmente dispersos en Bogotá, ni parecieran seguir un compartimiento totalmente aleatorio, al contrario, se podría intuir que existen grupos o clusters.

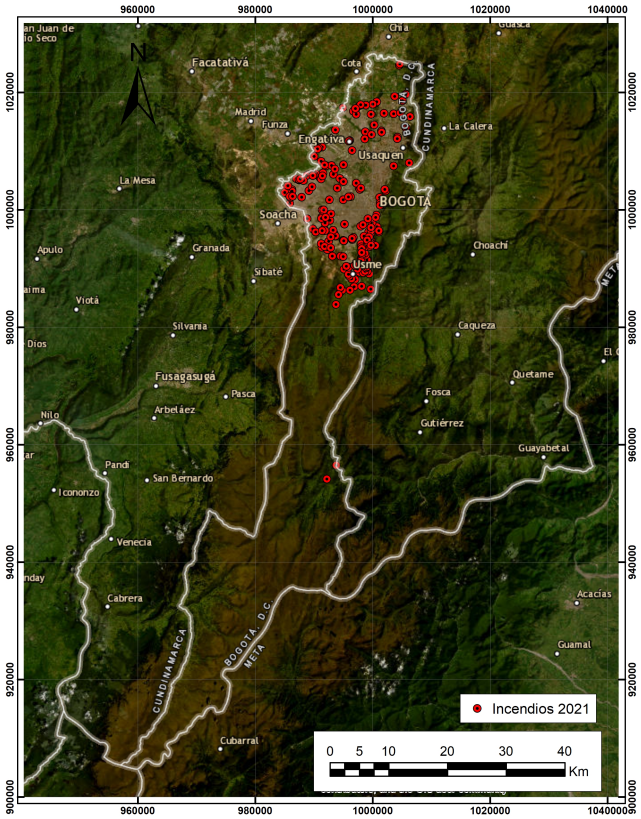


Figura 3: Ubicación de los incendios forestales ocurridos en Bogotá durante el año 2021

Mes	No. Quemas	No. Conatos de 0 hasta 5000 m2	No. Incendios más de 5000 m2	Total
Enero	37	14	1	**52**
Accidental	7	0	0	**7**
Indeterminada	24	13	1	**38**
Natural	4	1	0	**5**
Provocada	2	0	0	**2**
Febrero	42	8	1	**51**
Accidental	7	0	0	**7**
Indeterminada	31	7	1	**39**
Natural	1	1	0	**2**
Provocada	3	0	0	**3**
Marzo	10	2	0	**12**
Accidental	2	2	0	**4**
Indeterminada	5	0	0	**5**
Provocada	3	0	0	**3**
Abril	15	7	0	**22**
Accidental	4	0	0	**4**
Indeterminada	9	7	0	**16**
Natural	1	0	0	**1**
Provocada	1	0	0	**1**
Mayo	3	4	0	**7**
Accidental	1	0	0	**1**
Indeterminada	1	2	0	**3**
Natural	0	1	0	**1**
Provocada	1	1	0	**2**
Junio	2	1	0	**3**
Accidental	1	0	0	**1**
Indeterminada	1	1	0	**2**
Julio	5	1	0	**6**
Indeterminada	4	0	0	**4**
Natural	1	1	0	**2**
Agosto	14	2	0	**16**
Accidental	2	0	0	**2**
Indeterminada	11	1	0	**12**
Provocada	1	1	0	**2**
Septiembre	22	6	0	**28**
Accidental	9	1	0	**10**
Indeterminada	8	5	0	**13**
Natural	2	0	0	**2**
Provocada	3	0	0	**3**
Noviembre	9	3	0	**12**
Accidental	2	1	0	**3**
Indeterminada	7	1	0	**8**
Natural	0	1	0	**1**
Diciembre	16	4	0	**20**
Accidental	4	1	0	**5**
Indeterminada	12	3	0	**15**
Total General	**175**	**52**	**2**	**229**

Figura 4: Número de incendios por mes y subclase en Bogotá durante el año 2021

Como se enunció en el marco teórico, un patrón puntual se llama completamente aleatorio si se cumplen los siguientes requisitos:

1. El promedio de eventos por unidad de área, la intensidad $\lambda(s)$, es homogénea a lo largo de nuestra zona de interés.

2. El número de eventos en dos subregiones que no se solapan, son independientes.

3. El número de eventos en cualquier subregión sigue una distribución de Poisson

Por lo tanto, si los eventos se distribuyen uniforme e independiente a lo largo del dominio, este proceso se conoce como un proceso de Poisson Homogéneo. Esto es importante porque sirve como hipótesis nula para la investigación estadística que se desea realizar [18].

4.2. Intensidad

Conocer la intensidad de un patrón puntual es uno de los primeros y más importantes pasos en el análisis de datos. La intensidad es una característica básica descriptiva, que hace referencia al primer momento o al valor esperado, es análogo al promedio de la población. Comparada con otras propiedades de un patrón puntual, la intensidad requiere relativamente pocos supuestos en el modelo. El interés de conocer la intensidad radica en si es homogénea o no. Por ejemplo, una intensidad no homogénea puede reflejar preferencias, segregaciones, o dependencia espacial [4].

Aunque la intensidad no es el objetivo principal del análisis estadístico, sí es un componente bastante importate, debido a que la variación espacial puede fácilmente confundirse con otras características de los patrones como clustering (agrupamiento). Para establecer que un patron no es un clúster, se deben eliminar todas las explicaciones alernativas, incluida la variación espacial en la intensidad [4].

Si se sospecha que la intensidad podría no ser homogénea, se puede estimar utilizando técnicas no paramétricas como el conteo por cuadrantes o la estimación kernel.

4.2.1. Conteo por cuadrantes

Una manera sencilla para revisar si el proceso es no homogéneo es revisando si regiones de igual área contienen igual número de puntos (como debería suceder si el patrón puntual es homogéneo). Para este caso se divide la ciudad en cuadrantes de 40 m^2, y por localidades; se procede a estimar la intensidad (Figura 5)

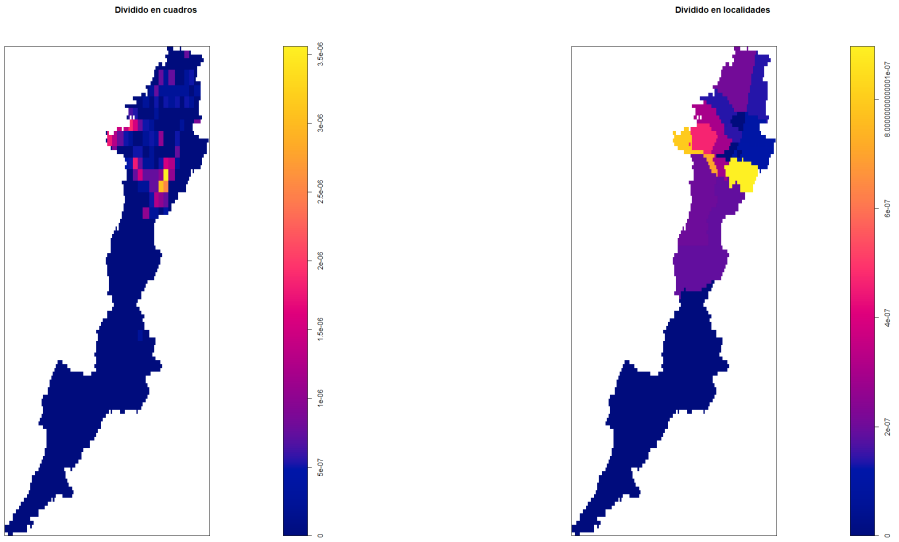


Figura 5: Estimación de la intensidad de los conteos por cuadrantes y por localidades

Si se observa por localidades, la intensidad varía significativamente, lo que nos induce a considerar que no sigue un patrón con intensidad homogénea; mientras que, haciendo el conteo por cuadrantes, pareciera que en su mayoría el patrón es homogéneo, pero existen cuadrantes con intensidad más alta que los demás, de hecho, pareciera que es un pequeño grupo el que tiene mayor intensidad, sin embargo, también nos sugiere que puede ser un patrón con intensidad no homogénea. Para determinar de forma certera la no homogeneidad de la intensidad en los conteos, se procede a realizar las pruebas CSR.

4.2.1.1 Test de homogeneidad para el conteo por cuadrantes

La prueba de ajuste estádistico, prueba la hipótesis nula de que los n puntos son distribuidos uniformemente independientes a lo largo del conjunto dominio D, en otras palabras, los conteos de sitios son variables Poisson independientes con media común [18].

Cabe resaltar que la prueba χ^2 se puede usar de dos formas: para probar el buen ajuste a la distribución Poisson asumiendo homogeneidad, o para probar homogeneidad asumiendo independencia. En este momento, el interés es la homogeneidad, por lo que se aplica la última mencionada.

Utilizando la función quadrat.test del paquete spatstat se obtiene lo siguiente:

- Divisiones por cuadrantes de 40 m^2

```
> quadrat.test(inc21PP, nx=40, ny=40, method = "Chisq")

Chi-squared test of CSR using quadrat counts

data:  inc21PP
X2 = 1848.9, df = 525, p-value < 2.2e-16
alternative hypothesis: two.sided

Quadrats: 526 tiles
```

- Divisiones por localidades

```
> quadrat.test(inc21PP, tess = BogoLoca, method = "Chisq")

Chi-squared test of CSR using quadrat counts

data:  inc21PP
X2 = 476.89, df = 19, p-value < 2.2e-16
alternative hypothesis: two.sided

Quadrats: 20 tiles
```

Asumiendo un nivel de significancia de $\alpha = 0{,}05$ para los conteos por cuadrantes de igual tamaño y por localidades, se rechaza la hipótesis nula.

Adicionalmente, la función quadrat.test permite conducir una prueba a un solo lado y computar el valor-p usando la simulación de Montecarlo en vez de la aproximación por χ^2:

- Divisiones por cuadrantes de 40 m^2

```
> quadrat.test(inc21PP, nx=40, ny=40, method = "MonteCarlo")
Conditional Monte Carlo test of CSR using quadrat
counts
Test statistic: Pearson X2 statistic

data:  inc21PP
X2 = 1848.9, p-value = 0.005
alternative hypothesis: two.sided

Quadrats: 526 tiles
```

- Divisiones por localidades

```
> quadrat.test(inc21PP, tess = BogoLoca, method = "MonteCarlo")

Conditional Monte Carlo test of CSR using quadrat
counts
Test statistic: Pearson X2 statistic

data:  inc21PP
X2 = 476.89, p-value = 0.001
alternative hypothesis: two.sided

Quadrats: 20 tiles
```

De manera similar, asumiendo un nivel de significancia de $\alpha = 0{,}05$ para los conteos por cuadrantes de igual tamaño y por localidades, considerando las simulaciones por Montecarlo, también se rechaza la hipótesis nula.

Por otro lado, si el proceso siguiera una distribución Poisson, se tendría que $E(X) = V(X) = \lambda$, por lo que se define el índice de dispersión como

$$I = \frac{V(X)}{E(X)} \approx 1$$

Para las divisiones por cuadrantes el índice arroja un resultado de 4.36, mientras que para las divisiones por localidades el índice es de 15.5, ambos valores son superiores a 1, por lo que se intuye que el patrón es agregado.

4.2.2. Estimación Kernel

La estimación Kernel es un método no paramétrico de estimación, en la formación de la función de densidad no se imponen supuestos tan extrictos a los datos como sucede con el planteamiento paramétrico [18].

Para este trabajo se utiliza el estimador corregido por Diggle, y se hace uso de la función bw.diggle, esta función utiliza el método de validación cruzada por máxima verosimilitud y se encarga de seleccionar el ancho de banda que minimiza el error [4].

En la Figura 6 (izquierda) se aprecia el resultado de la estimación por Kernel; cabe resaltar que cada celda dentro del ráster contiene un valor que corresponde con la densidad estimada en esa ubicación, cuando se visualiza en su totalidad, se pueden identificar áreas de alta y baja densidad. Dichas áreas se ven reflejadas en la Figura 6 (derecha), donde la altura refleja el valor de la densidad en esa área, como se aprecia la estimación

está concentrada en regiones pequeñas.

Los tonos en fucsia corresponden a las localidades de Suba, Bosa, Kennedy, Fontibón, Ciudad Bolívar, Tunjuelito, y Rafael Uribe Uribe; están asociados a los valores de $1x10^{-}6$ incendios por m^2, lo que quiere decir que se presenta al menos 1 incendio por km^2 en esas localidades. Por otro lado, los tonos en amarillo corresponden en gran parte a la localidad de San Cristobal y en menor parte a la localidad de Usme, tienen asociado el valor de $2{,}5x10^{-}6$ incendios por m^2, es decir que se presentan al menos 2 incendios por km^2 en estas localidades.

Existen zonas de la ciudad en donde la intensidad es cero (tonos azules), y como se aprecia en la ubicación de los incendios (Figura 3) era de esperarse porque en estas localidades no se presentó ningún evento durante el año de interés.

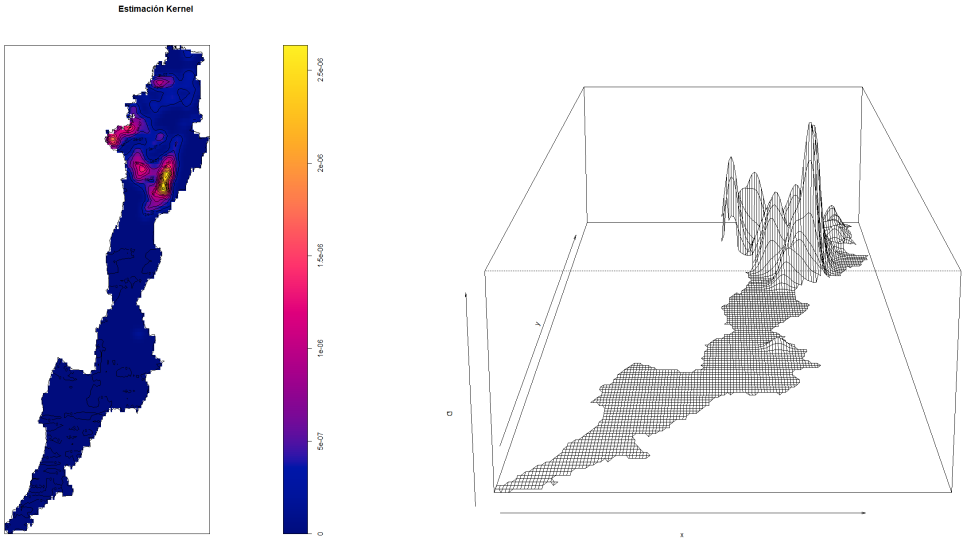


Figura 6: Estimación Kernel para determinar la intensidad de los incendios ocurridos en Bogotá durante el año 2021

4.3. Correlación

A menudo, la motivación para analizar datos de patrones de puntos es determinar si los puntos parecen haber sido colocados independientemente unos de otros, o si exhiben algún tipo de dependencia entre puntos. Cabe resaltar que la correlación es solamente un índice de asociación estadística, no una caracterización de dependencia o causalidad, y que la falta de ella no es prueba suficiente para asumir independencia [4].

4.3.1. Función K de Ripley

Quizá el estimador más sencillo de la propiedad de segundo orden de un patrón homogéneo e isotrópico sea $N(r)$, el número medio de puntos vecinos dentro de un círculo de radio r alrededor de cualquier punto típico del patrón. Para un determinado radio r, $N(r)$ estima de alguna manera la probabilidad de encontrar un punto en las inmediaciones de otro [19].

No obstante, con el objetivo de poder comparar las propiedades de segundo orden de patrones de diferente densidad, la función más utilizada sea la Función K, que es una estandarización de $N(r)$, en donde se define como:

$$K(r) = \frac{N(r)}{\lambda}$$

siendo λ la intensidad del patrón.

Utilizando la función Kest del paquete spatstat, se obtiene la siguiente gráfica:

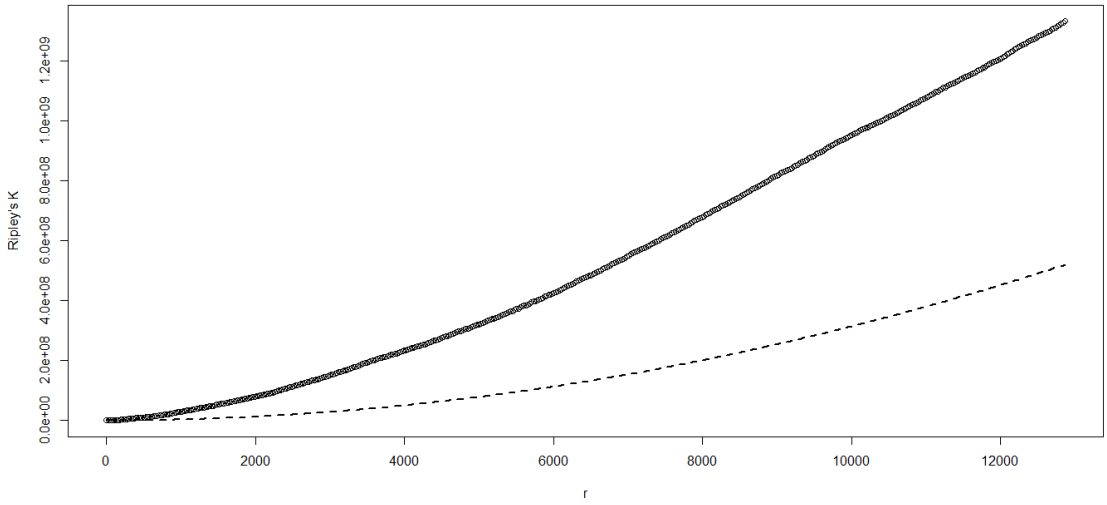


Figura 7: Función K empírica (puntos) y Función K teórica de un proceso Poisson (línea punteada)

Para interpretar la Figura 7, la función K empírica se compara con la función K teórica de un proceso de Poisson, se puede comprobar que el valor teórico de $K(r)$ en un proceso de Poisson es πr^2, lo cual sirve como punto de referencia de "no correlación"[4]. Luego, visualmente se puede observar que $\hat{K}(r) > K_{pois}(r)$, indicando que un punto típico tiene más vecinos de los que se espararía si el patrón fuera completamente aleatorio; esto es consistente con una asociación positiva entre los puntos (clustering / agrupamiento).

4.3.2. Función L

Una transformación de la Función K, es su versión centrada, $K(r) - K_{pois}(r) = K(r) - \pi r^2$. Esta función es útil y tiene una interpretación más clara para clasificar un patrón puntual como aleatorio, agregado o regular, porque la función es cero si el patrón puntual es completamente aleatorio [4].

$$L(r) = \sqrt{\frac{K(r)}{\pi}}$$

Utilizando la función Lest del paquete spatstat, se obtiene la siguiente gráfica:

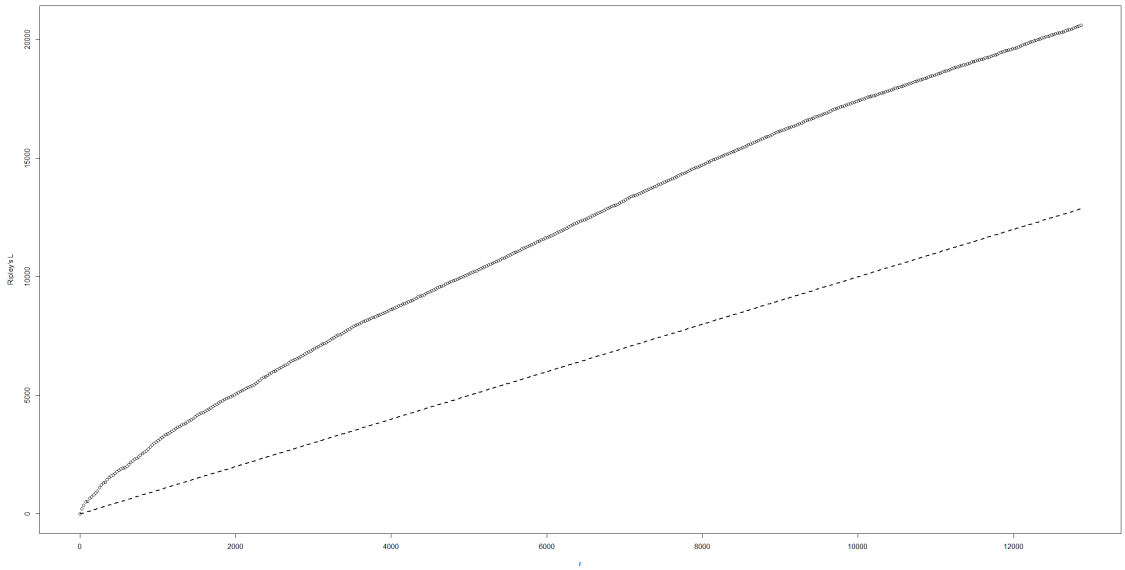


Figura 8: Función L empírica (puntos) y Función L teórica de un proceso Poisson (línea punteada)

Como se observa en la Figura 8 el patrón puntual pareciera que es consistente con un patrón agregado.

Si bien los resultados hasta ahora indican cierto grado de asociación positiva entre los puntos, vale la pena mencionar que no necesariamente se debe que sea un cluster o un patrón agregado, si se observa la ubicación de los incendios forestales para el año de interés (Figura 3), hay una gran concentración de puntos en uno de los extremos del polígono de Bogotá; como se conoce la función K es afectada por la propiedad de no homogeneidad espacial, así que se debe tener precaución con la interpretación.

4.3.3. Función de correlación por pares

La Función K es una función de distribucióna acumulada, ya que, a cada escala o distancia r, todos los pares de puntos separados por una distancia menor que r se usan para estimar el valor de la correspondiente función. En ocasiones puede ser necesario disponer de una función que caracterice de forma no acumulativa el patrón, que tenga en cuenta sólo los pares de puntos que se encuentran separados por una distancia igual a r, y es allí cuando entra la función de correlación por pares ($g(r)$) [19],

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

donde el numerador corresponde a la derivada de la Función K.

Para interpretar esta función se debe tener en cuenta que no es como la correlación estadística que va de -1 a 1, el rango de los posibles valores va de 0 a infinito, y el valor de 1 está asociado a una falta de correlación.
Utilizando la función pcf e incluyendo el estimador d del paquete spatstat, se obtiene la siguiente gráfica:

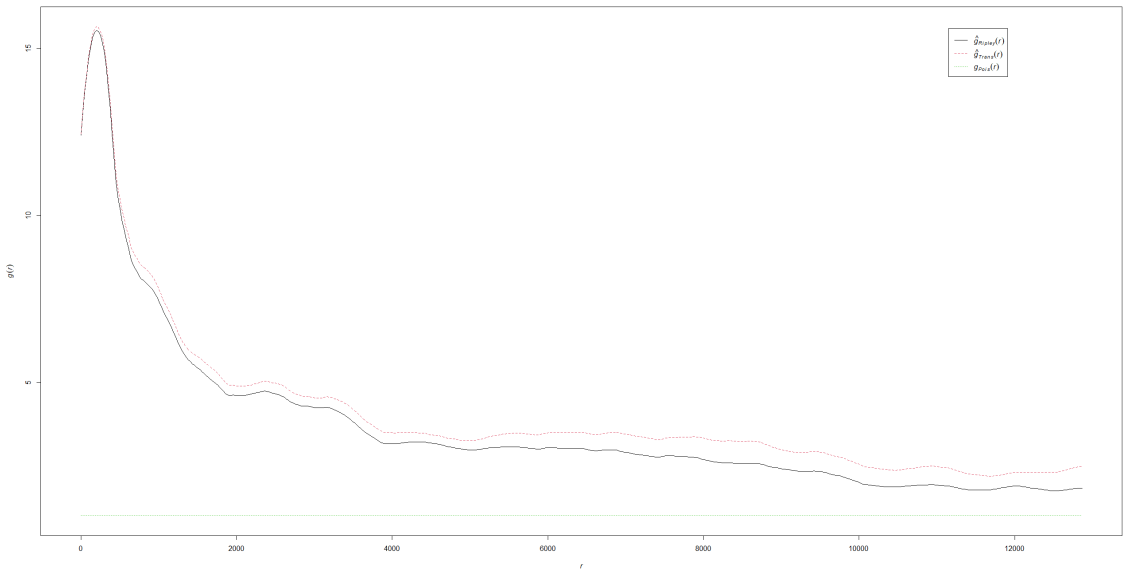


Figura 9: Función de correlación por pares empírica dividida por el estimador d

La Figura 9 sugiere que las distancias alrededor de r son relativamente más frecuentes de lo que sería un proceso aleatorio, este resultado es consistente con el agrupamiento que se viene interpretando de las anteriores funciones.

4.4. Métodos basados en distancias

Métodos estádisticos como la Función K de Ripley que mide la correlación espacial en un patrón puntual, hace parte de las herramientas más populares para analizar dependencia. Sin embargo, la corrrelación tan solo es un resumen de dependencia; se puede obtener información complementaria sobre un patrón puntual al medir espacios o distancias más cortas en el patrón, tal como la distancia al vecino más cercano [4].

4.4.1. Test de CSR basado en distancias cortas

La prueba de Clark-Evans de CSR se realiza aproximando el índice R bajo una distribución normal con media 1 y varianza $s^2 = (1/\pi - 1/4)/(m\lambda)$. Para una prueba bilateral, se rechaza la hipótesis nula de aleatoriedad espacial completa al nivel de significancia del 5 % si la estadística estandarizada $(R-1)/s$ se encuentra fuera de los límites críticos $\pm$1,96 [4].

```
 > clarkevans(inc21PP)
    naive       cdf
0.4870232 0.3580711

> clarkevans.test(inc21PP, correction = "cdf", alternative = "clustered")
Clark-Evans test
CDF correction
Monte Carlo test based on 999 simulations of CSR with
fixed n

data:  inc21PP
R = 0.35807, p-value = 0.001
alternative hypothesis: clustered (R < 1)
```

Considerando el valor-p de la prueba aplicada, se rechaza la hipótesis nula de aleatoriedad espacial completa con un nivel de significancia del 5 %.
De manera similar, se aplica la prueba de Hopkins-Skellam, la cual compara el valor del índice A a la distribución F con parámetros $(2m, 2m)$:

```
 > hopskel(inc21PP)
[1] 0.01347152
> hopskel.test(inc21PP, alternative = "clustered")

Hopkins-Skellam test of CSR
using F distribution
```

```
data:  inc21PP
A = 0.019, p-value < 2.2e-16
alternative hypothesis: clustered (A < 1)
```

Asumiendo el mismo nivel de significancia de la prueba anterior, se rechaza la hipótesis nula de aleatoriedad espacial completa. Los dos índices son consistentes con el agrupamiento de los puntos.

4.4.2. Gráficos exploratorios

El diagrama de Stienen de un patrón de puntos se obtiene dibujando un círculo alrededor de cada punto de datos, de diámetro igual a la distancia de su vecino más cercano. Los círculos se sombrean en gris si se observan sin efectos de borde, es decir, si la distancia observada al vecino más cercano es más corta que la distancia al límite de la ventana. Este diagrama puede ser útil para revelar aspectos del patrón espacial, como una tendencia en la escala espacial, un par de círculos que se tocan representan un par de puntos que son vecinos más cercanos entre sí. La fracción de área cubierta por los círculos de Stienen está relacionada con la distancia media al cuadrado del vecino más cercano [4].

Por otro lado, los mosaicos de Dirichlet son polígonos convexos (también conocidos como polígonos de Voronoi o polígonos de Stienen), y su función es dividir el espacio bidimensional en regiones disjuntas.



Figura 10: Diagrama de Stienen (izquierda), y Teselación Dirichlet (Derecha) para los incendios forestales ocurridos en Bogotá durante el año 2021

4.4.3. Función G(h) y F(h)

Como las funciones K y L, las funciones G y F son revisadas al graficar la función empírica calculada de los datos junto con la función teórica de un proceso Poisson homogéneo con la misma intensidad promedio, versus la distancia r.

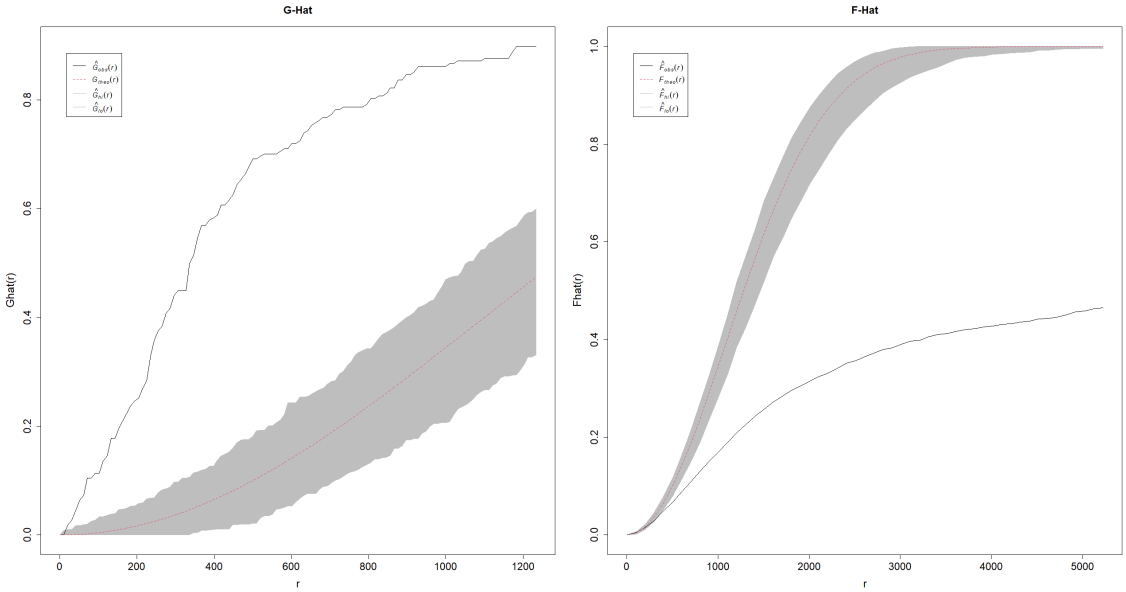


Figura 11: Función G: función de distribución de la distancia al vecino más cercano (izquierda); Función F: función de espacio vacío

En la Figura 11 (izquierda) se puede apreciar que el valor de $\hat{G}(r) = 0{,}5$ es alcanzado en $r \approx 400m$, es decir, que la mediana de la distribución de distancia al vecino más cercano se encuentra alrededor de 0.4 Km; mientras que para un proceso completamente aleatorio con la misma intensidad, la distancia mediana del vecino más cercano es de casi 1.2 Km. Además, se aprecia que la curva empírica está por encima de la curva teórica, indicando que las distancias al vecino más cercano en los datos son más cortas de lo que se esperaría si el patrón fuera completamente aleatorio. Esto es consistente con el agrupamiento de puntos (clustering).

Por otro lado, en la Figura 11 (derecha), se observa que la curva empírica está por debajo de la curva teórica, indicando que los espacios vacíos son mucho más grandes de lo que se esperaría si el patrón correspondiente fuera completamente aleatorio, esto es consistente con la agrupación de puntos (clustering).

4.5. Visualización de datos

Respecto a la visualización de datos, tal como se mencionó en la sección 3.2, se realiza un tablero de control utilizando ArcGIS Online. En el costado derecho del tablero se

encuentra un mapa de Bogotá, con la ubicación de los incendios y las siguientes capas disponibles: Límite de Bogotá, Localidades de Bogotá, Estimación de intensidad por Kernel, Valores Atípicos y Puntos Calientes; cada capa tiene disponible su correspondiente leyenda. Por otro lado, en el costado izquierdo se encuentran gráficos relacionados a los datos, como el conteo de incendios por localidades y por meses, también está disponible la clasificación según el tipo y causa del incendio. Vale la pena resaltar que el tablero está diseñado con la opción de dar click sobre el atributo de interés y este a su vez funcionará como filtro para el resto de gráficos y para el mapa disponible.

El tablero se encuentra disponible en el siguiente link `https://www.arcgis.com/apps/dashboards/260c515d14e0433bb01ccc082bac38d3`

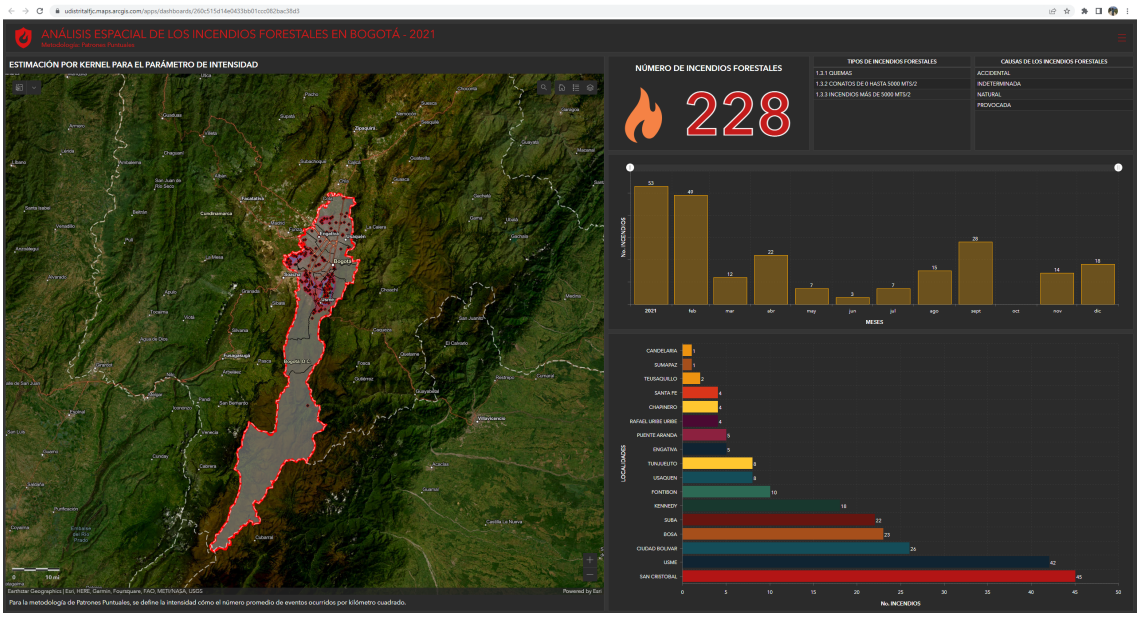


Figura 12: Vista Principal del Tablero de control Análisis Espacial de los Incendios Forestales en Bogotá - 2021

[3]

5. Impacto y Resultados Esperados

Utilizando la metodología de patrones puntuales se logra determinar que la distribución de los puntos sugiere una asociación positiva entre ellos, es decir, que los puntos posiblemente siguen un patrón agregado y no aleatorio. Sin embargo, hay que tener presente que la variación espacial puede causar que existan más puntos en un lugar que en otro, y este fenómeno puede dar la impresión de ser un patrón agregado. Por lo tanto, valdría la pena incluir información sobre covariables para determinar si esa

variación espacial se da por factores no observables o por otras causas.

Además, se puede complementar este trabajo con el ajuste de un modelo (paramétrico o no) que describa el proceso generador del patrón observado y se ajuste a los datos disponibles. Por otro lado, se debe considerar que un supuesto esencial para un modelo de patrón puntual Poisson es que los puntos de ese patrón sean independientes entre sí, lo cual resulta siendo un poco inapropiado para datos reales como los que se tienen en este trabajo. De hecho, asumiendo la asociación positiva entre los puntos, los modelos que posiblemente sean apropiados para este conjunto de datos sean el Proceso de Poisson Cluster o el Proceso de Cox.

Otro componente que se podría tener en cuenta es el tiempo, aunque para los incendios de Bogotá no existe un histórico tan antiguo (la información disponible al momento de escribir este trabajo es desde mitad de año del 2019), se podría aplicar un proceso espacio-temporal por medio de un proceso puntual de Cox con una estructura multiplicativa para la conducción de la intensidad aleatoria, incorporando información de covariables en las dos componentes (espacial y temporal). La ventaja de este modelo es la flexibilidad para el análisis estadístico y procedimientos de estimación rápidos basados en probabilidades compuestas y técnicas de simulación. En cuanto al proceso de validación del modelo, se puede seguir la metodología planteada por Moller [20] comparando con un modelo de proceso de punto de Cox log-gaussiano espacio-temporal.

Referencias

[1] J. G. PAUSAS, *Incendios Forestales una vision desde la ecologia.[En lınea] Madrid*, 2012.

[2] A. d. C. Parra Lara, F. H. Bernal Toro, D. Armenteras Pascual et al., «Incendios de la cobertura vegetal en Colombia,» 2011.

[3] *Riesgo por incendio forestal.* dirección: `https://www.idiger.gov.co/rincendiof#:~:text=3.-,Riesgo%5C%20por%5C%20Incendio%5C%20Forestal%5C%20en%5C%20Bogot%5C%C3%5C%A1,el%5C%202010%5C%20(223%5C%20ha).`.

[4] A. Baddeley, E. Rubak y R. Turner, *Spatial point patterns: methodology and applications with R.* CRC press, 2015.

[5] M. D. L. VILLERS RUÍZ, «Incendios forestales,» *Ciencias*, n.º 081, 2006.

[6] J. J. M. Sánchez, J. de las Heras Ibáñez y J. M. H. Sanz, «Impacto ecológico de los incendios forestales,» *Al-Basit: Revista de estudios albacetenses*, n.º 29, págs. 105-117, 1991.

[7] J. Mataix-Solera y C. Guerrero, «Efectos de los incendios forestales en las propiedades edáficas,» *Incendios forestales, suelos y erosión hıdrica*, págs. 5-40, 2007.

[8] K. Ocampo-Zuleta y J. Beltrán-Vargas, «Modelación dinámica de incendios forestales en los Cerros Orientales de Bogotá, Colombia,» *Madera y bosques*, vol. 24, n.º 3, 2018.

[9] D. Armenteras, F. González-Alonso y C. FRANCO AGUILERA, «Distribución geográfica y temporal de incendios en Colombia utilizando datos de anomalıas térmicas,» *Caldasia*, vol. 31, n.º 2, págs. 303-318, 2009.

[10] C. F. Jaque Espitia, C. D. Sánchez Rodrıguez, J. M. Álvarez Herrera et al., «Metodologıa para la estimación de amenaza de incendios forestales en los cerros orientales de la ciudad de Bogotá por medio de imágenes satelitales landsat tm 5 y kriging geo estadıstico,» 2017.

[11] D. A. Tirado Caraballo et al., «Evaluación de la Calidad del Aire por Material Particulado PM10 en una Zona del Barrio Ciudad Salitre en la Localidad de Fontibón (UPZ110).,» 2022.

[12] L. D. Llambı, M. T. Becerra, M. Peralvo, A. Avella, M. Baruffol y L. J. Dıaz, «Construcción de una Estrategia para el Monitoreo Integrado de los Ecosistemas de Alta Montaña en Colombia,» *Biodiversidad en la Práctica*, vol. 4, n.º 1, págs. 150-172, 2019.

[13] N. Cressie, *Statistics for spatial data.* John Wiley & Sons, 2015.

[14] O. Schabenberger y C. A. Gotway, *Statistical methods for spatial data analysis.* CRC press, 2017.

[15] I. C. Moncada Fuentes, «Análisis espacio-temporal de los accidentes de tránsito en Bogotá utilizando patrones puntuales,» *Departamento de Estadıstica*, 2018.

[16] C. Cundinamarca, «Modificación al Plan de Manejo Reserva Forestal Protectora de Bosque Oriental de Bogotá,» *Bogotá DC, Colombia. Recuperado el Agosto*, vol. 13, 2016.

[17] B. N. Boots y A. Getis, *Point pattern analysis*. SAGE publications Newbury Park, CA, 1988, vol. 10.

[18] R. Giraldo Henao, «Estadıstica espacial, notas de clase,» *Bogotá: Universidad Nacional de Colombia*, 2011.

[19] M. DE y C. ROT, «Métodos para analizar datos puntuales,»

[20] J. Møller y C. Dıaz-Avalos, «Structured spatio-temporal shot-noise Cox point process models, with a view to modelling forest fires,» *Scandinavian Journal of Statistics*, vol. 37, n.º 1, págs. 2-25, 2010.

[21] C. General, E. de Convivencia y S. Ciudadana, «Departamento Administrativo Nacional de Estadıstica,» *DANE, Colombia*, 2005.

[22] M. P. B. Castañeda, «Estadıstica espacial, notas de clase,» *Bogotá: Universidad Nacional de Colombia*, 2009.