



Patrones puntuales para los incendios en California en el 2017

Daniela Arbeláez Montoya
Jefferson Gamboa Betancur
Jean Paul Piedrahita García

Universidad Nacional de Colombia
Ciencias, Escuela de Estadística
Medellín, Colombia
2021

Índice

1. Introducción	2
2. Análisis descriptivo	2
2.1. Contorno de California	4
3. Procesos puntuales	7
3.1. Densidad cuadrática	7
4. Análisis basado en la distancia	10
4.1. Análisis de vecino más cercano promedio	10
5. Tecnicas Manuales	12
5.1. Gráfico de Patterson - Fry	12
6. Analisis de segundo orden para la incidencia de incendios en California en el 2017	12
6.1. Función K y F estimada.	13
6.2. Función G estimada y PCF.	14
7. Test para aleatoriedad completa	14
7.1. Pointwise Envelopes	15
8. Análisis Espacio-Temporal	16
8.1. Creación del objeto stpp	17
8.2. Análisis descriptivo	17
9. Estimación de la intensidad Espacio-Temporal	19
9.1. Características de segundo orden	20
9.2. Función K	20
9.3. Función Pair-Correlation	21

1. Introducción

California es uno de los lugares que tiene las temporadas de incendios forestales más mortíferas y destructivas. El conjunto de datos contiene la lista de incendios forestales ocurridos en California entre 2013 y 2020; contiene la ubicación donde ocurrieron los incendios forestales, incluido el nombre del condado, los valores de latitud y longitud y también detalles sobre cuándo comenzó el incendio forestal.

Estos datos ayudan a generar información sobre qué lugares de California están bajo amenaza de incendio, a qué hora suelen ocurrir los incendios forestales y qué tan frecuentes y devastadores son.

La base de datos con la que se va realizar el análisis de patrones puntuales fue descargada de kaggle¹.

Los paquetes a utilizar son los siguientes:

```
> library(spatstat)
> library(sf)
> library(raster)
> library(rgdal)
> library(tidyverse)
> library(maps)
> library(plot3D)
> library(rgl)
> library(stpp)
> library(rpanel)
> library(KernSmooth)
```

2. Análisis descriptivo

Para leer el archivo se utiliza la función “*read_csv*” de la librería “*tidyverse*” de la siguiente manera, visualizando la dimensión de la base de datos con la función “*dim*”.

Todos los incendios forestales registrados en la base de datos se encuentran inactivos o han sido contenidos. Debido a la gran cantidad de variables, seleccionamos algunas para visualizar un

¹<https://www.kaggle.com/ananthu017/california-wildfire-incidents-20132020>

Filas	Columnas
1636	40

poco mejor las localizaciones de incendios, tales como:

- **AcresBurned:** representa el número de acres de tierra afectadas por los incendios forestales.
- **ArchiveYear:** año en el cual se desarrolló o se produjo el incendio forestal.
- **Name:** nombre o denominación asignada al incendio forestal.
- **Counties:** nombre del condado en el cual se produjo el incendio forestal.
- **Latitude y Longitude** corresponden a las coordenadas geográficas de cada uno de los incendios forestales

```
> # Selección de variables
> firesCA <- firesCA %>% select(AcresBurned, ArchiveYear, Name,
+                                         Counties, Latitude, Longitude)
```

AcresBurned	ArchiveYear	Name	Counties	Latitude	Longitude
257314	2013	Rim Fire	Tuolumne	37.85700	-120.0860
30274	2013	Powerhouse Fire	Los Angeles	34.58559	-118.4232
27531	2013	Mountain Fire	Riverside	33.70950	-116.7288
27440	2013	American Fire	Placer	39.12000	-120.6500
24251	2013	Springs Fire	Ventura	0.00000	0.0000
22992	2013	Aspen Fire	Fresno	37.27900	-119.3180

Para saber el número de incendios que ocurrieron por año, visualizamos el número de registros con la función **table**. Se puede observar que el año 2017 fue el periodo donde más hubo incendios.

Año	Freq
2013	162
2014	117
2015	152
2016	185
2017	438
2018	316
2019	266

Se desea entonces observar y analizar el patrón de puntos de incendios forestales ocurridos durante el año 2017 en el estado de California. De esta manera realizamos el filtro correspondiente por

año de ocurrencia y además, eliminamos aquellas observaciones con errores en sus coordenadas geográficas:

```
> # Seleccionando el año 2017  
> firesCA2017 <- firesCA %>% filter(ArchiveYear == 2017, Longitude != 0,  
+                                         Latitude != 0)
```

Sin embargo, la base de datos sigue registrando seis observaciones duplicadas de acuerdo a sus coordenadas geográficas. Eliminamos dichas observaciones de la siguiente manera:

```
> # ¿Hay datos duplicados?: 6 por valores de longitud y latitud  
> sum(duplicated(firesCA2017 %>% select(Longitude, Latitude)))
```

```
## [1] 6  
  
> firesCA2017 <- firesCA2017[!duplicated(firesCA2017 %>% select(Longitude,  
+                                         Latitude)),]
```

La función **project** del paquete **rgdal** ofrece una interfaz con la librería *PROJ.4* de funciones de proyección para datos de posición geográfica. Para poder utilizar esta función debemos extraer las coordenadas de cada uno de los incendios en una matriz. Realizamos la proyección correspondiente de acuerdo a la zona UTM en la cual se encuentra el estado de California y de acuerdo al elíptope **WGS84** (Sistema Geodésico Mundial).

```
> # Seleccionando las coordenadas  
> coordFiresCA2017 <- firesCA2017 %>% select(Longitude, Latitude) %>% as.matrix()  
> # Proyección de coordenadas a UTM  
> pointsFiresCA2017 <- project(coordFiresCA2017,  
+                                     "+proj=utm +zone=11N +ellps=WGS84") %>%  
+   as.data.frame() %>% rename(X = Longitude, Y = Latitude)
```

2.1. Contorno de California

El paquete **maps** contiene la base de datos **state** que produce un mapa de los estados del territorio continental de los Estados Unidos generado a partir de los datos del Departamento del Censo del mismo país; estos datos contiene las coordenadas geográficas de los polígonos correspondientes a cada uno de los estados.

```

> # Base de datos de Estados Unidos
> UnitedStates <- map_data("state")
> # ¿Qué estados?
> names(table(UnitedStates$region))

```

```

## [1] "alabama"           "arizona"            "arkansas"
## [4] "california"         "colorado"           "connecticut"
## [7] "delaware"           "district of columbia" "florida"
## [10] "georgia"            "idaho"              "illinois"
## [13] "indiana"            "iowa"               "kansas"
## [16] "kentucky"           "louisiana"          "maine"
## [19] "maryland"            "massachusetts"       "michigan"
## [22] "minnesota"          "mississippi"        "missouri"
## [25] "montana"             "nebraska"           "nevada"
## [28] "new hampshire"       "new jersey"         "new mexico"
## [31] "new york"            "north carolina"      "north dakota"
## [34] "ohio"                "oklahoma"           "oregon"
## [37] "pennsylvania"        "rhode island"        "south carolina"
## [40] "south dakota"        "tennessee"          "texas"
## [43] "utah"                "vermont"             "virginia"
## [46] "washington"          "west virginia"       "wisconsin"
## [49] "wyoming"

```

Realizamos el filtro correspondiente al estado de California para obtener el polígono. La función **map_data** del paquete **ggplot2** convierte fácilmente los datos del paquete de **maps** en un marco de datos adecuado para trazar con **ggplot2**. La opción **Mercator** indica meridianos rectos igualmente espaciados, concordantes y rumbos rectos de la brújula.

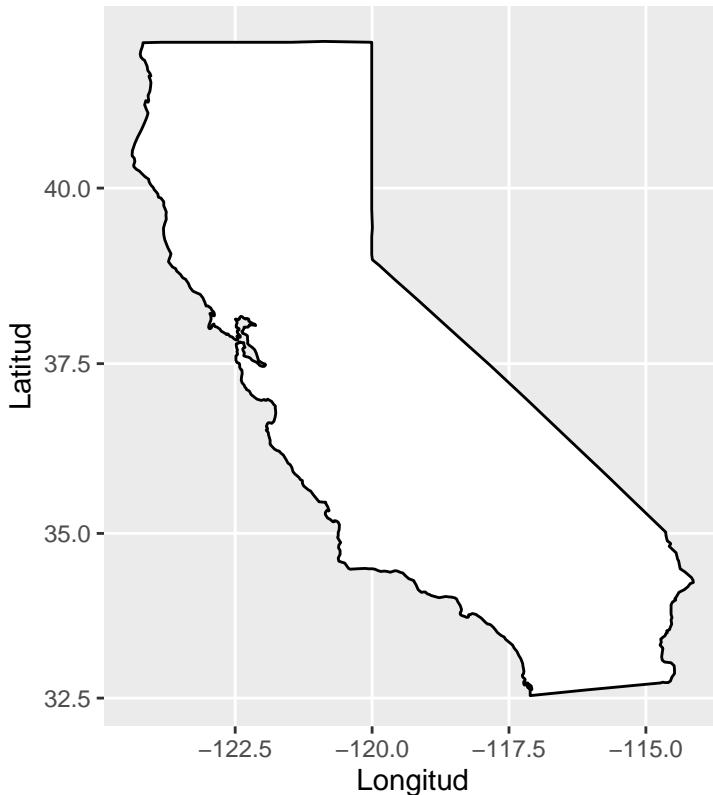
```

> # Filtramos para California
> California <- UnitedStates %>% filter(region == "california")
> ggplot(California, aes(x=long, y=lat)) +
+   geom_polygon(fill="white", colour="black") + coord_map("mercator") +
+   labs(title = "Estado de California en proyección Mercator",
+        x = "Longitud", y = "Latitud") +

```

```
+   theme(plot.title = element_text(face = "bold", hjust = 0.5, size = 12))
```

Estado de California en proyección Mercator



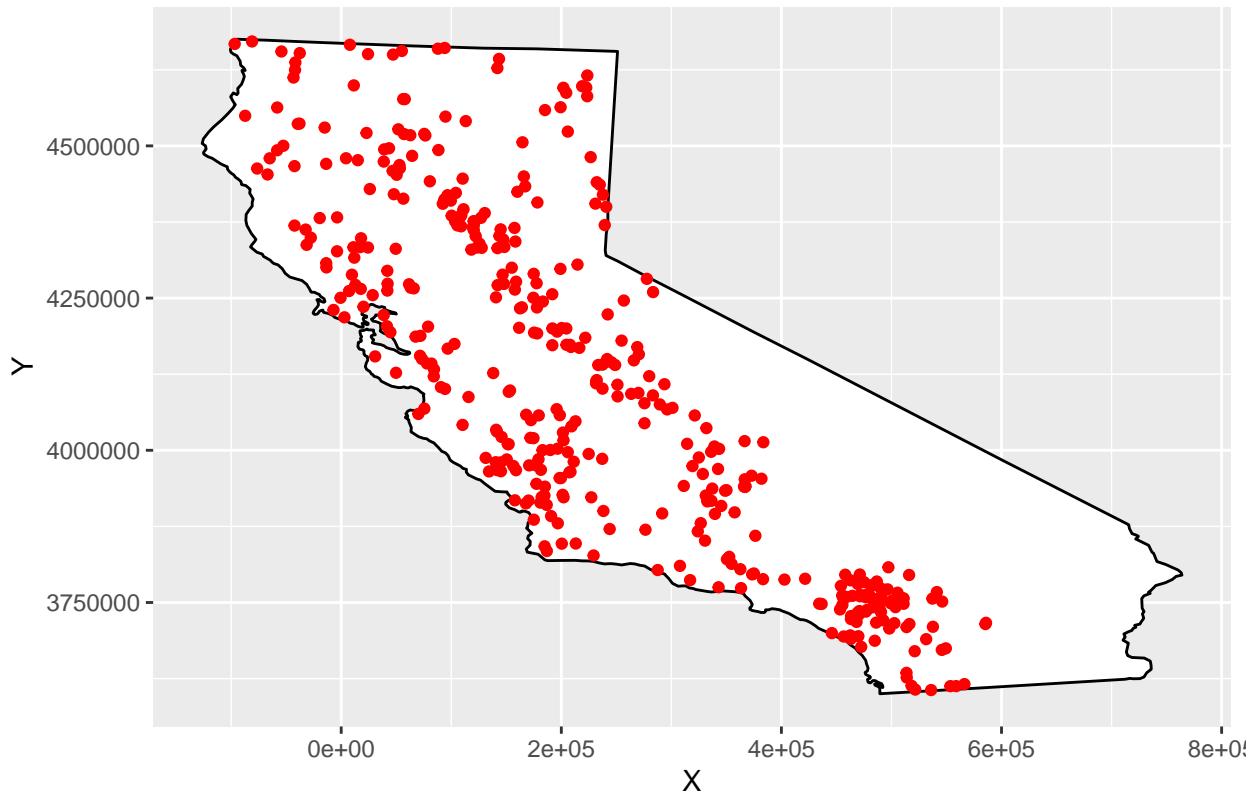
De la misma manera realizamos la proyección de las coordenadas geográficas en planas de acuerdo a la zona UTM correspondiente al estado de California.

```
> # Coordenadas del borde para proyección
> coordCA_border <- California %>% select(long, lat) %>% as.matrix()
> coordCA_border <- project(coordCA_border,
+                               "+proj=utm +zone=11N +ellps=WGS84") %>%
+   as.data.frame() %>% rename(X = long, Y = lat)
```

Luego de la proyección, podemos crear el objeto **owin** correspondiente a la ventana de observación para el análisis del patrón de puntos espacial; sin embargo, es posible que algunos puntos o incendios forestales recaigan fuera de dicha ventana de observación, algo que no sería acorde o adecuado para el análisis. Para solucionar este inconveniente utilizamos la función **inside.owin** de la siguiente manera:

Visualizamos la proyección de los puntos en nuestra ventana de observación

Localización de Incendios en California, año 2017

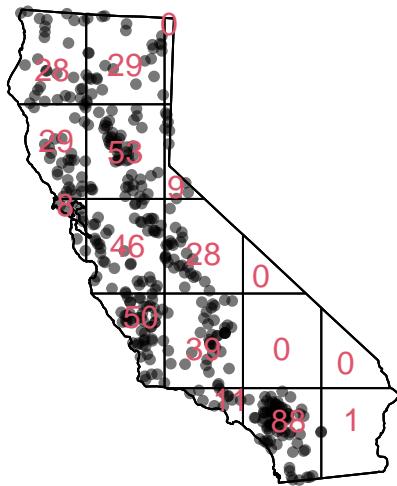


3. Procesos puntuales

3.1. Densidad cuadrática

El siguiente fragmento de código divide el estado de California en una cuadricula 5 filas, 5 columnas y luego se hace el conteo de las localizaciones que caen en cada cuadrante.

Conteo por cuadrantes



Se puede calcular la densidad de puntos dentro de cada cuadrante pero cada cuadrante esta en metros cuadrados, por ello para una visualización numérica se decidió pasar las localizaciones en km^2 de la siguiente manera:

```
> # Reescalamiento a kilometros
> FiresCA2017_ppp.km <- rescale(FiresCA2017_ppp, 1000, "km")
>
> # Conteo por cuadrantes en km
> Q <- quadratcount(FiresCA2017_ppp.km)
> # Intensidad
> Q.d <- density(FiresCA2017_ppp.km)
>
> par(mfrow = c(1,2), no.readonly = T)
> # plot de la densidad
> plot(intensity(Q, image = T),
+       main = "Densidad de incendios forestales \n en California",
+       las = 1)
> plot(FiresCA2017_ppp.km, pch = 20, cex = 0.6, col = rgb(0,0,0,.5), add = T,
```

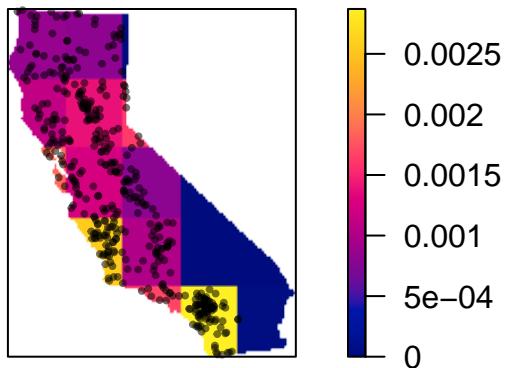
```

+      cols = "#00000088")

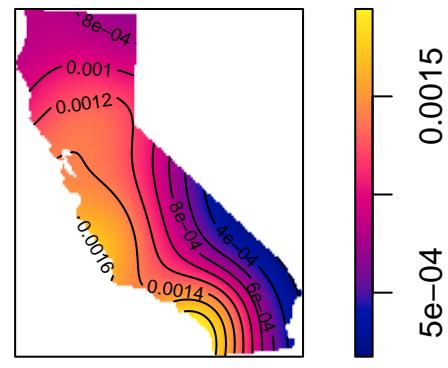
> # Intensidad estimada
> plot(Q.d, main = "Intensidad Estimada")
> contour(Q.d, add = T)

```

Densidad de incendios forestales en California

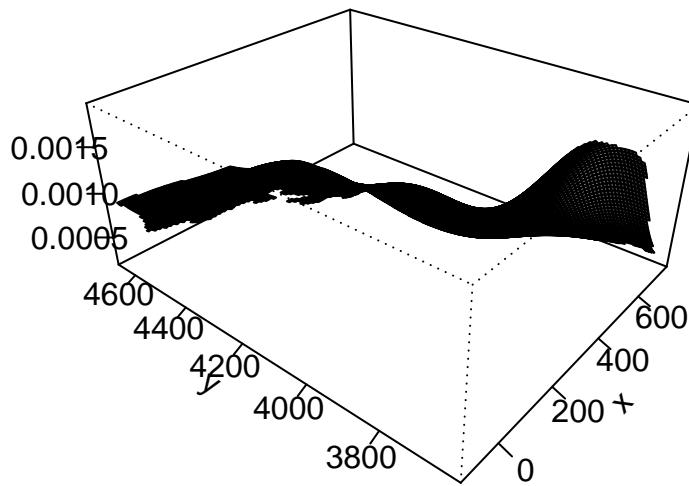


Intensidad Estimada



La intensidad espacial estimada para los incendios en California en el año 2017, proporciona una idea de lo que podemos esperar que pase en esa misma región en un futuro no muy lejano.

Superficie para la intensidad de incendios 2017



4. Análisis basado en la distancia

4.1. Análisis de vecino más cercano promedio

A continuación, calcularemos las distancias promedio de vecino más cercano (ANN) entre la ocurrencia de incendios del año 2017 en California.

Para calcular la distancia promedio del primer vecino más cercano (en acres), se establece $k = 1$:

```
## [1] 10.59479
```

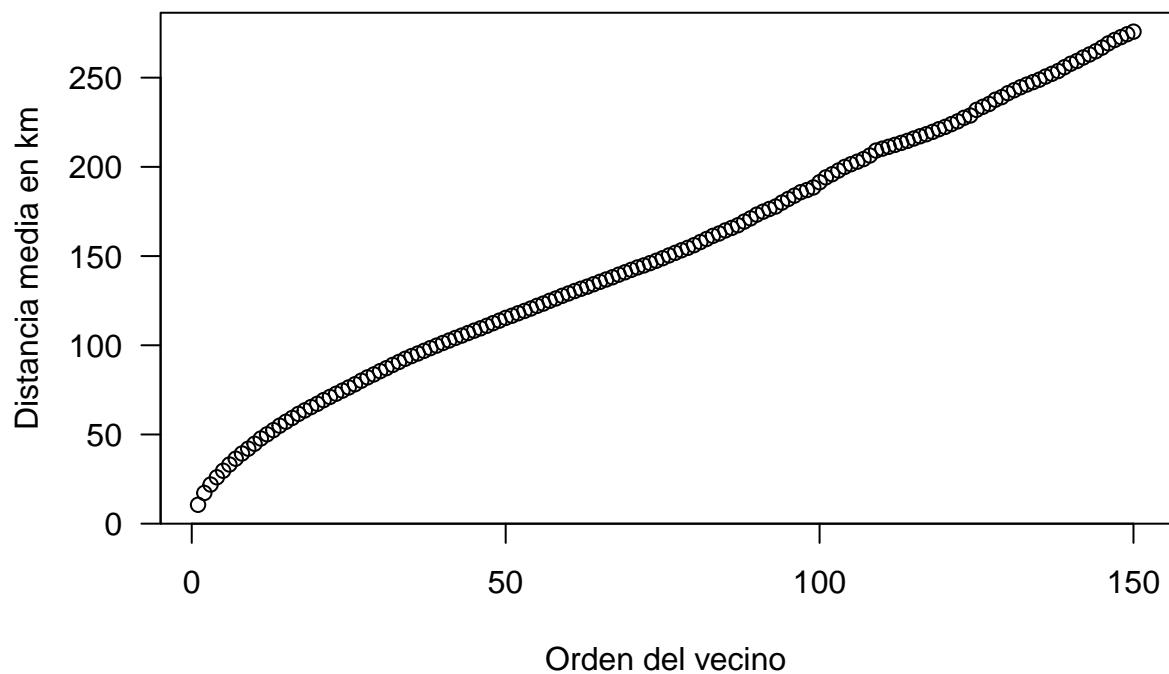
Para calcular el conjunto de distancia promedio del segundo vecino mas cercano, se usa:

```
## [1] 17.22054
```

El parámetro k puede tomar cualquier orden vecino (hasta $n - 1$ donde n es el número total de puntos).

A continuación se traza ANN como una función del orden de vecinos para los primeros 150 vecinos mas cercanos.

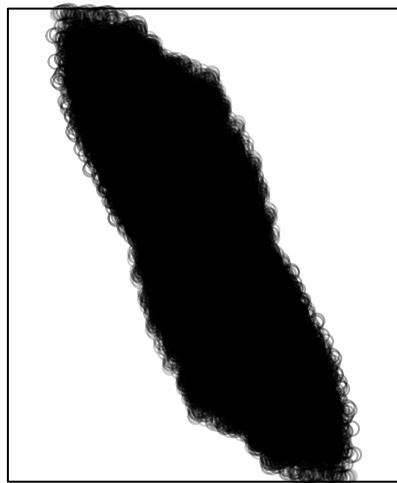
Vecinos mas cercanos



5. Técnicas Manuales

5.1. Gráfico de Patterson - Fry

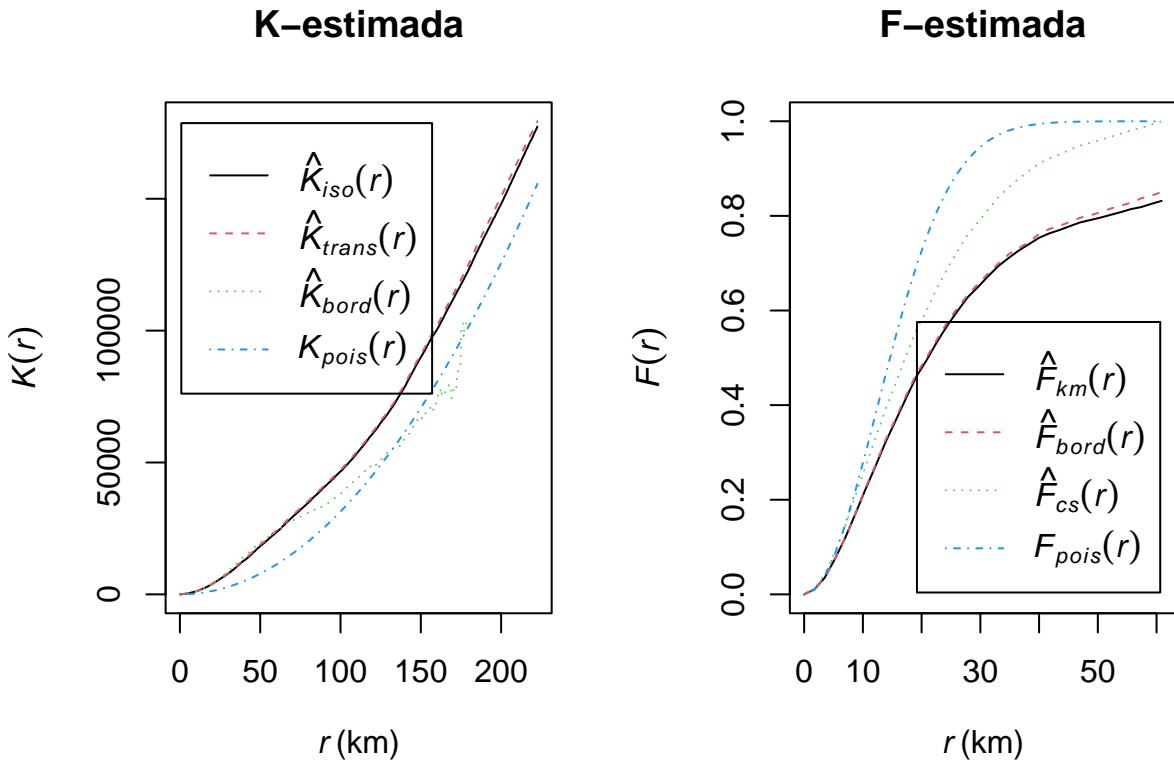
Gráfico de Fry



Al realizar el gráfico de Patterson Fry se observa que todos los puntos muestran una alta aglomeración de puntos alrededor del centro, por ello se presume un patrón de agregación.

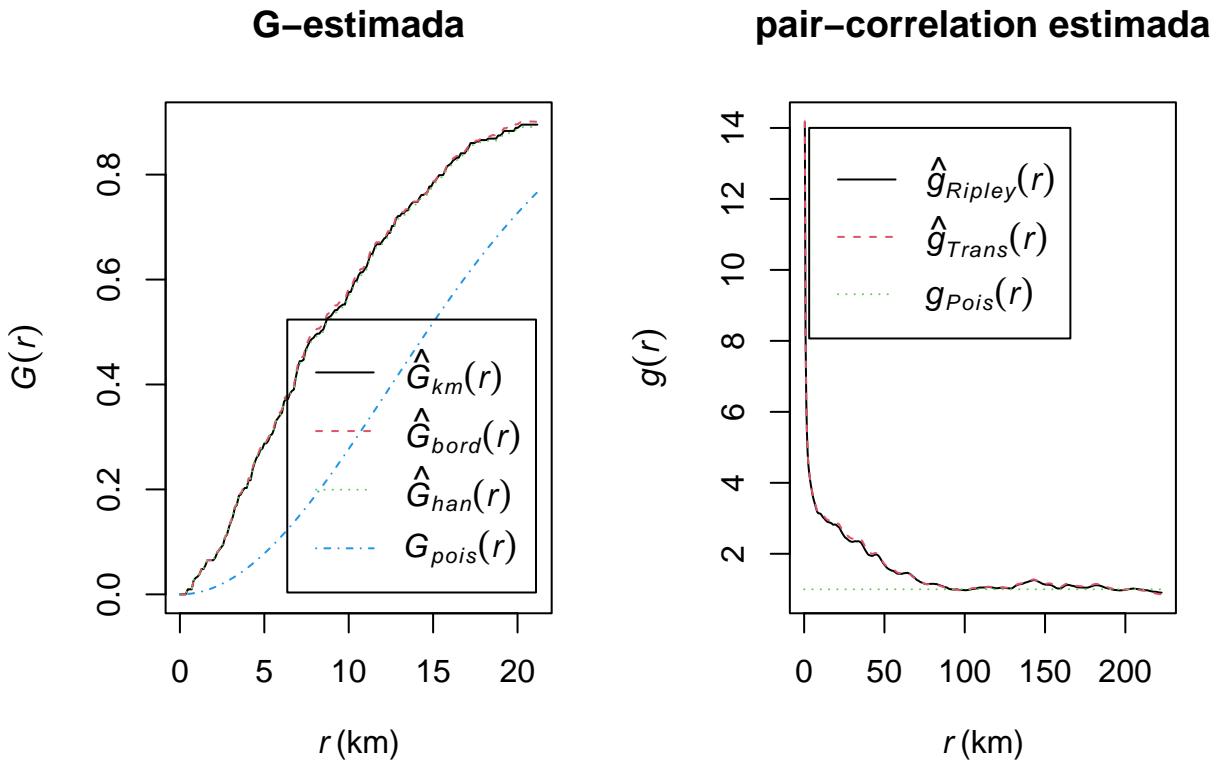
6. Análisis de segundo orden para la incidencia de incendios en California en el 2017

6.1. Función K y F estimada.



- Para la gráfica de K-estimada se ve que debido a que las estimaciones están por encima de la curva teórica, se puede pensar que se trata de un proceso de agregación.
- Cómo la función F-Estimada las curvas están por debajo de la curva teórica (Azul) se supone que es un patrón de agregación.

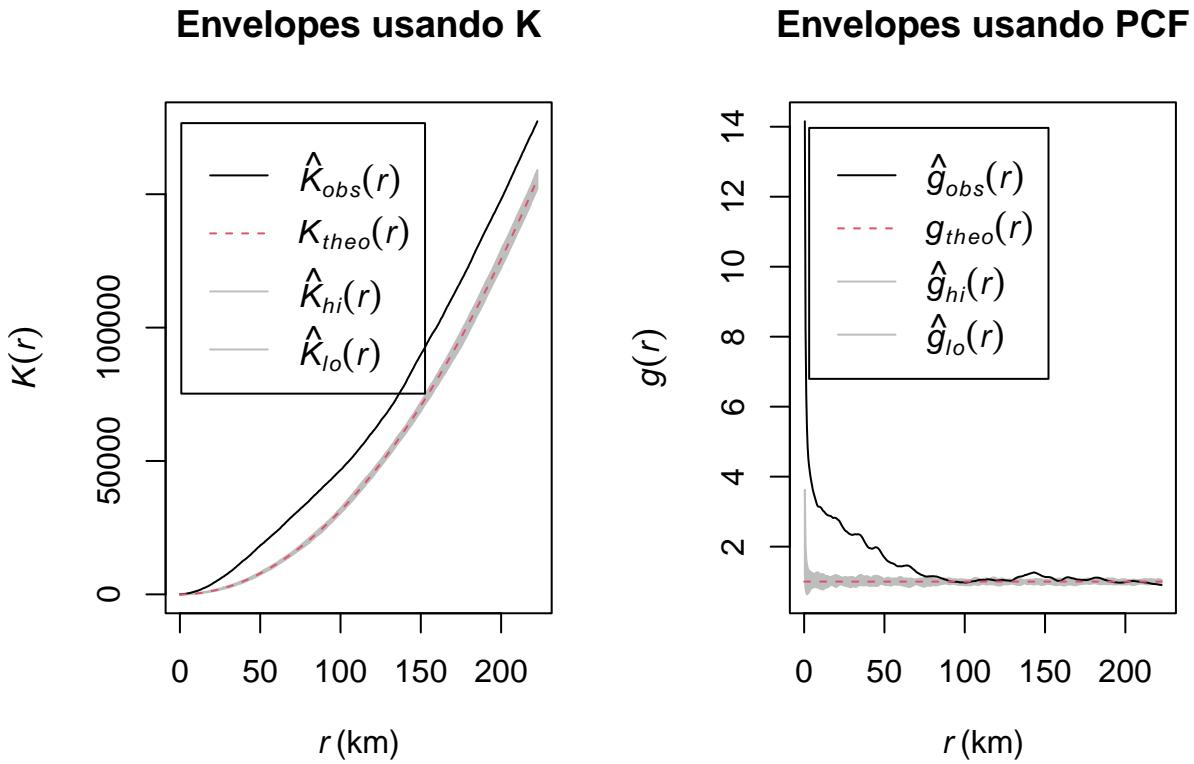
6.2. Función G estimada y PCF.



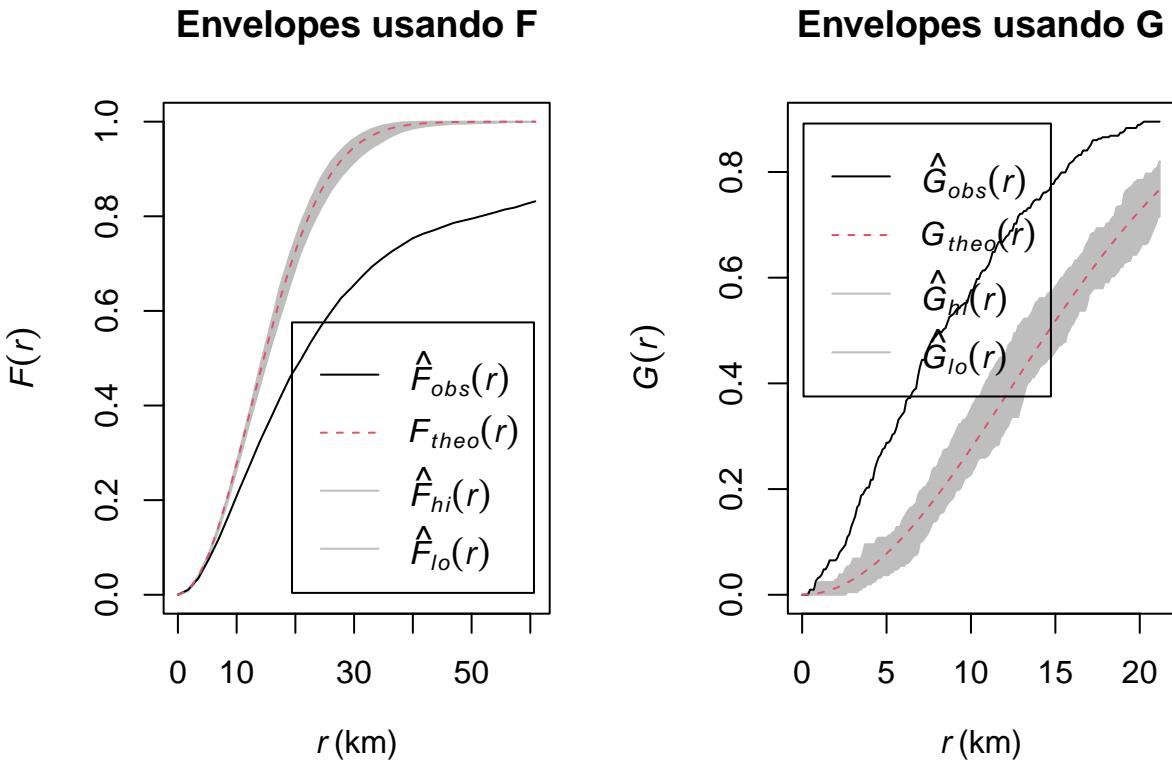
- Como las funciones estimadas para la G están por encima de la teórica se presume que el patrón es de agregación.
- Debido a que las estimaciones de la función de correlación por pares se alejan por arriba de la teórica en distancias pequeñas se puede asumir que la hipótesis del patrón puntual es de agregación.

7. Test para aleatoriedad completa

7.1. Pointwise Envelopes



- Con el test para la aleatoriedad completa se puede concluir que el patrón se aleja por arriba de la curva teórica y además la sombra gris no contiene la curva negra que es la función K-estimada, por ello se asume que es un patrón de agregación.
- Para la función PCF-estimada a distancias cortas se sobre ajusta a la curva teórica y además las sombras grises no cubren a la función estimada a distancias pequeñas intuyendo un patrón de agregación.



- Para la función F el mecanismo es al contrario, cuando la función F-estimada está por debajo de la ajustada se supone que el patrón es de agregación.
- En la función G al tener la curva estimada por encima de la teórica se supone que el patrón es de agregación.

8. Análisis Espacio-Temporal

Debemos recordar que la base de datos principal, llamada en este caso **firesCA**, contiene los registros de incendios forestales ocurridos en el Estado de California entre los años 2013 y 2019, por lo que es posible analizar la dinámica e interacción (correlación) a través del tiempo entre los eventos espaciales de los incendios. En esta última sección se desarrolla un análisis espacio-temporal de los datos, extendiendo los modelos de procesos puntuales espaciales hacia este contexto.

La base de datos contiene 187 registros duplicados, los cuales son eliminados de la misma manera que se realizó para los datos de los incendios forestales del 2017.

Para poder desarrollar el análisis, es necesario entonces adecuar correctamente los datos con los

mismos pasos hechos anteriormente, proyectando todos los puntos en coordenadas UTM.

Como se mencionó anteriormente, algunas observaciones o puntos recaen por fuera de la ventana de observación, por tanto se realiza la corrección con la función **inside.owin**

8.1. Creación del objeto stpp

La función **as.3dpoints** del paquete **spatstat** coloca a punto los datos en formato espacio-temporal. Para esto, es necesario ingresarle una matriz de tres columnas, de las cuales las dos primeras corresponden a las coordenadas planas y la tercera al tiempo.

Recordemos que un proceso aleatorio espacio-temporal es un proceso estocástico indexado en el espacio y tiempo. En este caso, los eventos que corresponde a cada uno de los incendios forman un conjunto numerable de puntos:

$$P = \{(u_i, t_i) \text{ con } i = 1, 2, \dots\}$$

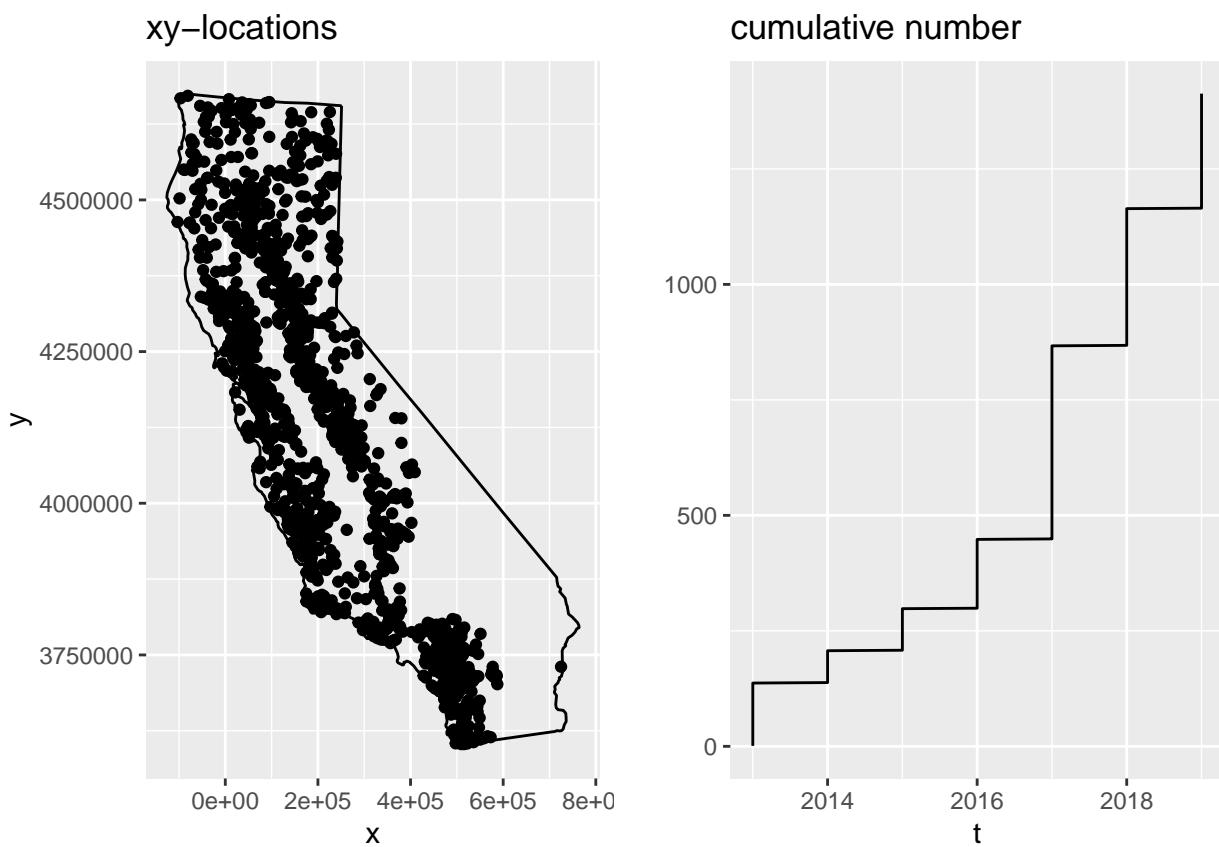
donde $u_i \in \mathbb{R}^2$ es la localización espacial de los incendios en coordenadas planas y $t_i \in \mathbb{R}$ es el tiempo de ocurrencia de cada uno de los incendios entre los años 2013 y 2019.

Es importante recordar que un proceso puntual espacio-temporal se denomina separable si las componentes de espacio y tiempo de la función de intensidad se pueden estudiar por separado.

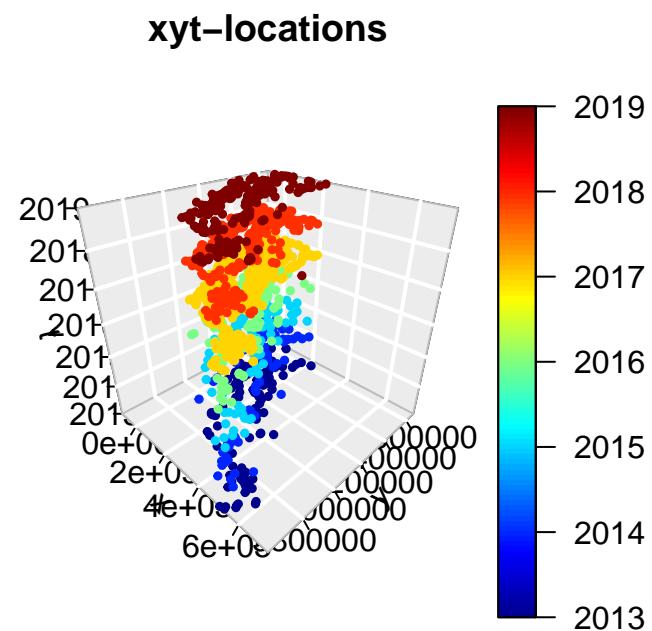
8.2. Análisis descriptivo

La función **plot.stpp**, cuyos autores son Edith Gabriel y Francisco Rodríguez, contiene diferentes formas de visualizar los datos de espacio-tiempo de acuerdo a los argumentos que se ingresen.

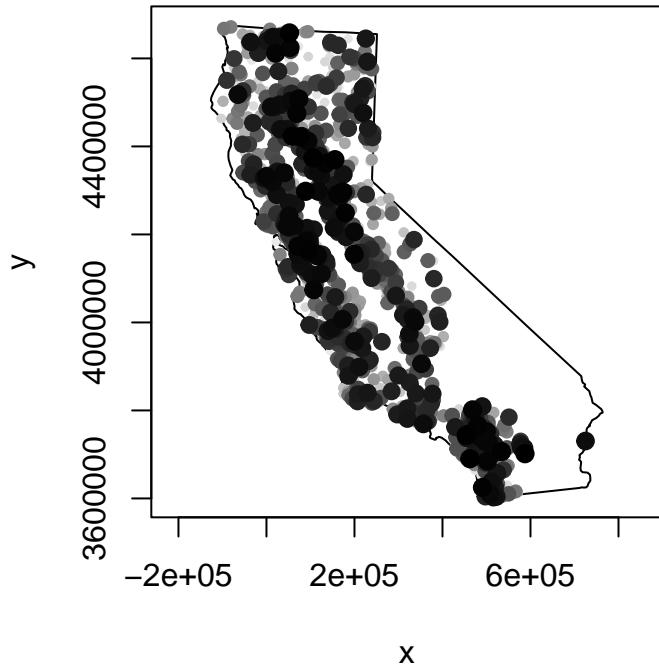
- La función utiliza por defecto **type = “projection”**, la cual traza las ubicaciones xy y la distribución acumulada de los tiempos. El argumento **style** permite adecuar el gráfico en formato **ggplot2**. Acá podemos observar todos los puntos sobre la ventana de observación, además de una especie de escala en la parte del tiempo, en donde se incrementa más por el año 2017



- La opción **type = "scatter"** permite crear un gráfico 3D del patrón de puntos.



- Si la opción **mark = TRUE**, luego el tiempo se trata como una marca cuantitativa adjunta a cada ubicación, y las ubicaciones se trazan con el tamaño y / o el color del símbolo de trazado determinado por el valor de la marca.



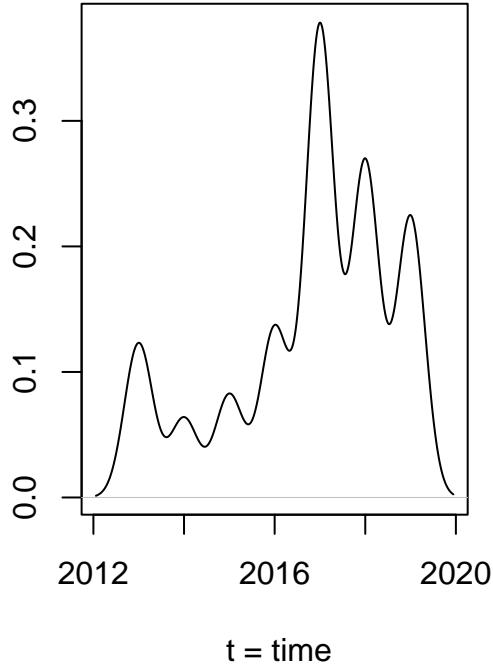
Se puede crear una animación de puntos de acuerdo al tiempo en que ocurrieron empleando la función **animation**. Adicionalmente, la función **stan** muestra datos de puntos (x , y , t) y permite resaltar dinámicamente segmentos de tiempo.

9. Estimación de la intensidad Espacio-Temporal

La primera etapa de un análisis estadístico es por lo general estimar y modelar la función de intensidad. El hecho de asumir separabilidad entre el espacio y el tiempo divide la estimación en dos componentes reduciendo la dimensión. Como ya se ha estudiado en módulos anteriores, un estimador muy sencillo e intuitivo consiste en contar, para cada localización, el número de eventos dentro de un disco centrado en dicho evento y para un radio fijo.

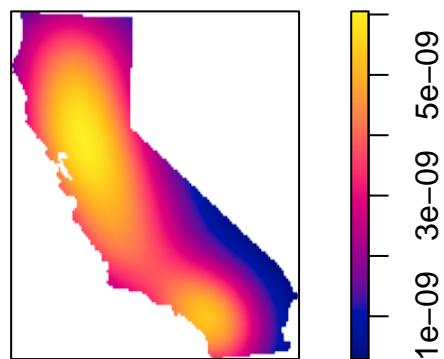
El siguiente gráfico corresponde a la estimación no paramétrica de la intensidad temporal y espacial para el patrón de incendios forestales en California.

Intensidad temporal



$t = \text{time}$

Intensidad espacial



9.1. Características de segundo orden

Las propiedades de segundo orden se utilizan para analizar y describir la estructura de correlación espacio-temporal que existe entre los eventos de un proceso puntual. Las funciones de espacio temporal no homogéneas **pair-correlation g** y **K-function** se pueden utilizar como medidas para verificar agrupación o regularidad espacio temporal, además como cuantificadores de la interacción espacio-temporal.

9.2. Función K

La **K function** se define en este caso para la estimación como el número medio de incendios dentro de un cilindro de radio r y altura $2t$ alrededor de un incendio *típico* del patrón (sin contar dicho incendio central). Por lo tanto $K(r, t)$ describe las características del proceso de puntos a muchas escalas. La interpretación se realiza en base a la superficie teórica.

9.3. Función Pair-Correlation

Dado que la función K es en cierta forma una función de distribución acumulada, en ocasiones puede ser necesario disponer de una función que caracterice de forma no acumulativa el patrón, es decir que tenga en cuenta tan sólo los pares de puntos que se encuentran separados por una distancia exactamente igual o similar a la distancia (r, t) .

Función pair-correlation espacio-temporal

