

---

# Correlación entre patrones puntuales en tipos de fuentes de captaciones hídricas en Colombia

Correlation between specific patterns in types of sources of water sources in Colombia

Karen Alejandra Bojacá Sánchez.<sup>a</sup>  
karenbojaca@usantotomas.edu.co

Wilmer Pineda Ríos.<sup>b</sup>  
wilmerpineda@usantotomas.edu.co

Jony Romero Guerrero.<sup>c</sup>  
jonyromero@usantotomas.edu.co

---

## Resumen

En el presente estudio se analiza espacialmente la relación entre dos procesos puntuales con el fin de proponer un método para identificar la correlación entre tipos de fuentes de captaciones hídricas a lo largo del territorio colombiano. Los datos utilizados contienen dos tipos de fuentes hídricas, superficiales y subterráneas; a través del índice de dispersión se identificó que el conjunto de datos es un patrón agregado, además se verificó la hipótesis de aleatoriedad espacial completa por medio del método de cuadrantes, dicho método se utilizó para identificar la correlación entre las fuentes de captaciones hídricas superficiales y las fuentes de captaciones hídricas subterráneas para ello previamente se realizaron simulaciones para patrones poisson y aglomerados o Cluster, comprobando el método utilizado para identificar la correlación entre los patrones analizados.

## Abstract

In the present study, the relationship between two specific processes is spatially analyzed in order to propose a method to identify the correlation between types of water sources throughout Colombia, the data used contains two types of water sources, superficial and underground; through the dispersion index it was identified that the data set is an aggregated pattern, in addition, the hypothesis of complete spatial randomness was verified by means of the dequating method, said method was used In order to identify the correlation between the sources of surface water sources and the sources of underground water sources, previously simulations for poisson and cluster patterns were performed, checking the method used to identify the correlation between the patterns analyzed.

---

<sup>a</sup>Estudiante pregrado Estadística, U. Santo Tomás, sede Bogotá

<sup>b</sup>Docente Facultad de Estadística, U. Santo Tomás, sede Bogotá

<sup>c</sup>Docente Departamento de Ciencias Básicas, U. Santo Tomás, sede Bogotá

## 1. Introducción

Los recursos hídricos son los cuerpos de agua que existen en el planeta disponibles en cantidad y calidad suficientes, en un lugar y en un período de tiempo apropiados para satisfacer una demanda identificable en el ser humano.

En Colombia, debido a su ubicación geográfica y a sus condiciones de relieve representa una riqueza importante de recursos hídricos, cuando es comparada con el promedio mundial; la demanda de agua aumentó cerca del 5 % al pasar de 35.582 millones de metros cúbicos al año en 2014 a 37.308 millones en 2018. En él se pueden encontrar seis tipos de aguas, incluyendo aguas lluvias, aguas superficiales, aguas subterráneas, aguas termo minerales, aguas marinas y oceánicas y aguas de alimentación glacial.

El agua subterránea es aquella que se filtra a través de grietas y poros de las rocas y sedimentos que yacen debajo de la superficie de la tierra, acumulándose en las capas arenosas o rocas porosas del subsuelo. Aproximadamente el 75 % del territorio colombiano presenta condiciones favorables para el almacenamiento de agua subterránea. (Instituto Colombiano de Geología y Minería. 2015).

El agua superficial es aquella que se encuentran en el exterior de la tierra. Pueden ser corrientes que se mueven en una misma dirección y circulan continuamente, como los ríos, quebradas y arroyos o las que se encuentran estancadas como los lagos, lagunas y pantanos.

La falta de abastecimiento de agua en algunas zonas de Colombia, el aumento de la contaminación hídrica por vertimientos contaminantes tiene ahora una prioridad importante en los planes de las autoridades competentes de la mayoría de países.

Según el trabajo investigativo realizado por Giraldo & Poveda (2013) frente a la variabilidad espacial y temporal en almacenamiento de agua en el suelo en Colombia, se estudia la precipitación en el territorio Colombiano en el periodo Agosto de 2002 Febrero de 2008, y su relación respecto al evento el Niño. Concluyen que el modelo de humedad de suelo por medio de patrones puntuales característicos, es decir estudiados individualmente por zonas representa una buena aproximación el ciclo anual de la humedad de suelo en las diversas regiones del territorio Colombiano.

El Departamento de Epidemiología de la Universidad de California en los Ángeles (UCLA) afirma que, si dos distribuciones muestran patrones similares, sugiere que están relacionados entre sí, ya sea directa o indirectamente, la similitud entre dos distribuciones sugiere que una distribución es causa de la otra distribución, o dos distribuciones tienen la misma causa.

Además indican que en el mapa del cólera de Londres del Dr. John Snow encontraron la similitud entre la distribución de casos de cólera y la ubicación de una bomba de agua. Esto sugiere que el agua de la bomba contiene una causa de cólera.

La similitud espacial entre dos distribuciones puede ser descrita cuantitativamente por la proximidad espacial entre las distribuciones.

En la actualidad no se cuenta con un método para identificar la correlación entre dos o más patrones puntuales aglomerados que puedan ser utilizados en las diferentes áreas de conocimiento.

El objetivo de este trabajo es establecer por medio de un método la correlación entre patrones puntuales aplicado a la captación del recurso hídrico en los departamentos de Colombia.

Por lo anterior se realiza un análisis exploratorio de los datos para determinar la estructura espacial de los recursos hídricos. En la segunda sección se abarca el marco teórico de patrones puntuales, el método de cuadrantes y procesos Poisson especificando sus características, propiedades y tipos; en la tercera sección se especifica el marco metodológico, variables de estudio y la fuente de información de los datos utilizados; en la cuarta sección se presentan los resultados encontrados. Por último en la quinta sección se describen las conclusiones.

## 2. Marco Teórico

### 2.0.1. Patrones puntuales

Un proceso puntual es un proceso estocástico en el que observamos las ubicaciones de algunos eventos de interés dentro de una región delimitada. Diggle (2003) define un proceso puntual como un mecanismo estocástico que genera un conjunto de eventos contables. Diggle (2003) y Möller & Waagepetersen (2003) proporcionan definiciones adecuadas de los diferentes tipos de procesos de puntos y sus propiedades principales. Las ubicaciones de los eventos generados por un proceso de puntos en el área del estudio  $A$  se llamarán patrones de puntos estos patrones implican una medida de conteo.

El análisis de patrones puntuales se evidencia en áreas de investigación como epidemiología como lo son los de casos localizados de infecciones gastrointestinales (Brix y Diggle 2001), en ecología en árboles forestales (Rathbun & Cressie 1994b), datos geológicos, ocurrencias de terremotos (Zhung, Ogata & Vere-Jones 2002), entre otros.

Los patrones puntuales son realizaciones de experimentos aleatorios y se distingue entre patrones (completamente) aleatorios, agrupados (espacialmente agregados), y los normales.

- En el patrón aleatorio los puntos se distribuyen aleatoriamente en el espacio.
- En el patrón regular existe una distancia media entre los puntos que tiende a ser constante, la distancia media entre un evento y su vecino más cercano es más grande que la esperada bajo aleatoriedad.
- En el patrón aglomerado o cluster existen acumulación de puntos en el espacio, es decir la distancia media entre los eventos  $s_1$  y su vecino más cercano es más pequeña que la misma distancia media en un patrón espacial completamente aleatorio.

El análisis estadístico de los patrones de distribución de puntos requiere un modelo teórico de referencia base para el desarrollo de métodos formales que comprueban la significancia de los resultados exploratorios.

El modelo teórico más simple es el de aleatoriedad espacial completa, en este modelo se asume que los puntos son independientes y tienen una distribución de probabilidad uniforme sobre el área de estudio.

Las variables espaciales estiman  $\lambda(s)$  como la intensidad en una localización  $(s)$ , la cual es obtenida por métodos no paramétricos, de suavizamiento de cuadrantes, o por métodos de estimación de la intensidad donde:

$$\lambda = \text{Número de eventos que ocurren por unidad de área.}$$

### Método de cuadrantes

Es una prueba estadística para evaluar la similitud entre dos distribuciones. Se utiliza para analizar no solo un conjunto de puntos sino también la relación entre dos conjuntos de puntos.

Para probar la hipótesis de aleatoriedad espacial completa se puede utilizar el método de los cuadrantes basados en la división del dominio  $D$  en subregiones no traslapadas (llamadas cuadrantes)  $\{A_1, A_2, \dots, A_n\}$  de igual tamaño.

El conteo del número de puntos en  $n$  cuadrantes, si la media es aproximadamente igual a la varianza, es posible pensar que la distribución Poisson sea adecuada y que se cumpla la aleatoriedad espacial completa.

Para verificar el patrón de los puntos se calcula el índice de dispersión.

$$I = \frac{S_u^2}{\bar{u}} \quad (1)$$

- Si el índice de dispersión  $I = 1$  será un patrón aleatorio.
- Si el índice de dispersión  $I < 1$  será un patrón regular.
- Si el índice de dispersión  $I > 1$  será un patrón agregado.

Bajo aleatoriedad espacial completa se considera que los eventos ocurridos en cada área  $S_i$  son no correlacionados y siguen una distribución Poisson.

### Propiedades de primer y segundo orden

Las propiedades de primer orden son descritas por la intensidad del proceso, su función está dada por:

$$\lambda(x) = \lim_{|dx| \rightarrow 0} \left( \frac{E[N(dx)]}{|dx|} \right) \quad (2)$$

Donde  $dx$  define una mínima región espacial alrededor de  $x$ ,  $|dx|$  el área de la región, y  $N(dx)$  es el número de eventos ocurridos en  $dx$ .

Las propiedades de segundo orden dan a conocer el tipo de las interacciones entre los eventos del proceso. La intensidad de segundo orden de dos puntos  $x$  e  $y$  refleja la probabilidad de que ocurra cualquier par de eventos en las proximidades de  $x$  e  $y$ , respectivamente, su función está dada por:

$$\lambda_2(x, y) = \lim_{|dx| \rightarrow 0, |dy| \rightarrow 0} \left( \frac{E[N(dx, dy)]}{|dx||dy|} \right) \quad (3)$$

Donde  $dy$  es un intervalo que contiene el tiempo  $y$ ,  $|dy|$  representa la longitud del intervalo y  $N(dx, dy)$  es el número de eventos ocurridos en  $dx * dy$ .

En estas propiedades se incluye la densidad de covarianza la cual está dada por:

$$\gamma(x, y) = \lambda_2(x, y) - \lambda(x)\lambda(y) \quad (4)$$

La densidad de covarianza es el proceso puntual similar a la función de covarianza de un proceso estocástico evaluado en los reales, la densidad de covarianza será idénticamente cero.

Además se encuentra en estas propiedades la función de correlación de pares de puntos, la cual está dada por:

$$g(x, y) = \frac{\lambda_2(x, y)}{\lambda(x)\lambda(y)} \quad (5)$$

La función de correlación de pares se puede interpretar como la función de densidad de probabilidad estándar aplicada a que un evento ocurra en cada uno de las observaciones centradas en los puntos  $(x, y)$ .

### Estimación de la función de intensidad

La intensidad y la densidad espacial son parte de las propiedades de primer orden dado que miden la distribución de eventos en la región de estudio. Es importante resaltar que la intensidad y la densidad espacial no dan a conocer la interacción entre dos puntos arbitrarios, esta interacción es dada por las propiedades de segundo orden que reflejan cualquier tendencia de los eventos.

La estimación de la intensidad se puede hacer de forma no paramétrica mediante el suavizado, mediante kernel o paramétricamente, proponiendo una función específica para la intensidad cuyos parámetros se estiman al maximizar la probabilidad del proceso puntual. Si se ha observado  $n$  puntos  $\{x_i\}$ , la forma de un estimador de suavizado kernel es la siguiente:

$$\hat{\lambda}(x) = \frac{1}{h^2} \sum_{i=1}^n k\left(\frac{\|x - x_i\|}{h}\right) / q(\|x\|) \quad (6)$$

donde  $k(u)$  es una función kernel bivariada y simétrica,  $q(\|x\|)$  es una corrección de borde para compensar las observaciones faltantes que ocurren cuando  $x$  se encuentra cerca del borde de la región A. El ancho de banda  $h$  mide el nivel de suavizado. Donde los valores pequeños producirán estimaciones muy altas, mientras que los valores grandes producirán funciones muy suaves.

### Función K de Ripley

La función K de Ripley (1976) es una función de  $\lambda_2$  para procesos estacionarios e isotrópicos. También es conocida como la medida reducida del segundo momento (Cressie(1993)), y es la función reducida del segundo momento. Mide el número de eventos encontrados hasta una distancia dada de cualquier evento en particular.

Está dada por:

$$k(d) = \frac{E(Ne)}{\lambda} \quad (7)$$

Donde:

$\lambda$  = Intensidad

$Ne$  = Número de eventos adicionales a una distancia  $d$  de un evento elegido aleatoriamente

#### 2.0.2. Procesos de Poisson

El proceso Poisson da un punto de inicio natural para un desarrollo estadístico de un patrón puntual observado.

Suponiendo  $\{T_1, T_2, \dots, T_n\}$  variables aleatorias distribuidas representan los tiempos que transcurren entre una ocurrencia del evento y la siguiente ocurrencia, además estos tiempos son independientes y cada uno sigue una distribución  $\exp(\lambda)$ .

Por lo tanto se puede definir el proceso de Poisson al tiempo  $t$  como el número de ocurrencias del evento que se han observado hasta ese instante  $t$ .

De esta manera se puede definir de manera espacio temporal un proceso de Poisson así:

Sea  $\{T_1, T_2, \dots, T_n\}$  una sucesión de variables aleatorias independientes cada una con distribución  $\exp(\lambda)$  el proceso de Poisson de parámetro  $\lambda$  para un proceso continuo  $\{X_t: t \geq 0\}$ , estará definido de la siguiente manera

$$X_t = \text{máx}\{n \geq 1 : T_1 + T_2 + \dots + T_n \leq t\} \quad (8)$$

Esto significa que, para un valor de  $n \geq 1$ , todos los tiempos de interarribo a partir de  $n$ , siguen teniendo distribución  $\exp(\lambda)$ , y por lo tanto el proceso de conteo de eventos a partir del tiempo  $n$  es un proceso de Poisson.

### 2.0.3. Procesos Poisson no homogéneos

Se originan cuando el proceso no es estacionario y el parámetro  $\lambda$  del proceso de Poisson no es necesariamente una constante sino una función del tiempo y se obtiene sustituyendo la intensidad constante  $\lambda$  de un proceso Poisson por una función de intensidad variable  $\lambda(x)$ .

Estos procesos están definidos por:

- $N(A)$ : Número de eventos en la región  $A$
- $|A|$ : Área de la región  $A$
- $E(N(A)) = \lambda(A)$
- $\lambda(A) = \int_A \lambda(s) ds$

Estos procesos de Poisson no homogéneos tienen las siguientes características:

- $N(0) = 0$
- $N(A) \sim \text{Poisson}(\lambda(A))$
- Si  $A_1$  y  $A_2$  son dos regiones no traslapadas,  $N(A_1)$  y  $N(A_2)$  son independientes.
- Dado  $N(A) = n$ , las localizaciones  $\{s_1, s_2, \dots, s_n\}$  son realizaciones independientes distribuidas uniformemente.

$$f_A(s|n) = \frac{\lambda(s)}{\int_A \lambda(s) \cdot ds} \quad (9)$$

En este caso los intervalos de tiempo entre eventos sucesivos  $T_n$ ,  $n \geq 1$  ya no son independientes ni tienen distribución exponencial.

### 2.0.4. Patrones Poisson

En Huang (2011) se estudia la relación de correlación entre procesos de dos puntos. cuando el proceso puntual es Poisson, un solo coeficiente es suficiente para describir la relación de correlación.

Dado el teorema:

sea  $\{N_t: > 0\}$  el proceso de Poisson. El número esperado y la varianza de eventos en  $(0, t]$  son

$$\begin{aligned} E(N_t) &= \lambda t \\ V(N_t) &= \lambda t \end{aligned}$$

Donde  $N_1(b)$  y  $N_2(b)$  son procesos de dos puntos,  $\rho(b)$  denotando los coeficientes de correlación entre  $N_1(b)$  y  $N_2(b)$

$$\{N_1(b) = N_1^-(b) + N(b)\} \quad (10)$$

$$\{N_2(b) = N_2^-(b) + N(b)\} \quad (11)$$

Donde  $N_1(b), N_2(b)$  y  $N(b)$  son procesos de Poisson independientes con tasa  $\lambda_1, \lambda_2$  y  $\lambda$ , entonces  $N_1(b)$  y  $N_2(b)$  son procesos de dos puntos, por lo tanto el coeficiente de correlación entre  $N_1(b)$  y  $N_2(b)$  es:

$$\rho(b) = \frac{\lambda}{\sqrt{(\lambda_1 + \lambda)(\lambda_2 + \lambda)}} \quad (12)$$

## 3. Marco Metodológico

### 3.1. Tipo de estudio

El tipo de estudio a desarrollar es explicativo dado que se quiere establecer la relación entre los patrones espaciotemporales del tipo de fuentes de captaciones hídricas en Colombia ya sea subterránea o superficial, además es correlacional para lograr identificar la relación de los tipos de fuentes de agua y la ubicación en la que se encuentran.

### 3.2. Variables de estudio

Las variables a utilizar serán:

- El tipo de fuente hídrica superficial que se encuentre en el territorio colombiano analizadas espacialmente.
- El tipo de fuente hídrica subterránea que se encuentre en el territorio colombiano analizadas espacialmente.

### 3.3. Base de información de los datos

Para realizar el estudio propuesto se utilizó la base captación de agua, del Sistema de Información del Recurso Hídrico (SIRH) del Instituto de Hidrología, Meteorología y Estudios Ambientales (IDEAM). La cual contiene información del tipo de captación, ya sea fuente hídrica subterránea o superficial, las coordenadas en las que se encuentran ubicadas, el código y nombre de la zona y corporación a la que se encuentra vinculada.

### 3.4. Fuentes hídricas

Para el presente estudio se observaron en la figura 1 los puntos donde se presentan las fuentes de captaciones hídricas en Colombia.

## Fuentes hídricas en Colombia



Figura 1: Fuentes hídricas

Se presentan la totalidad de puntos de tipo de captación de agua tanto subterránea como superficial para el territorio colombiano, el número total de casos es de 13.325

#### 3.4.1. Fuentes hídricas superficial y subterránea

## Fuentes superficial y subterránea en Colombia

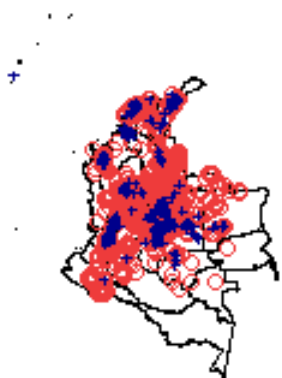


Figura 2: Fuentes superficial y subterránea

En la figura 2, se puede observar que la fuente hídrica superficial la cual se identifica de color rojo se encuentra presente en mayor número en los departamentos del territorio colombiano dado que el número

de casos es de 11.598 y la fuente hídrica subterránea se identifica de color azul y su número de casos es de 1.727.

### 3.5. Tipo de Patrón puntual

Las ubicaciones de los eventos de fuentes hídricas generados por un proceso de puntos en Colombia se llamarán patrones de puntos.

Para verificar el patrón de los puntos se calcula el índice de dispersión donde se obtuvo un valor de 1757.787 dado que el índice es mayor a uno se puede inferir que el patrón que presenta el recurso hídrico es agregado, por lo tanto existen aglomeraciones de puntos en el espacio.

### 3.6. Densidad de los patrones observados

Se puede hacer una primera descripción del patrón puntual estimando la densidad estadística espacial a partir de los datos observados.

#### 3.6.1. Intensidad de fuentes hídricas superficial

Mapa densidad de fuente hídrica superficial

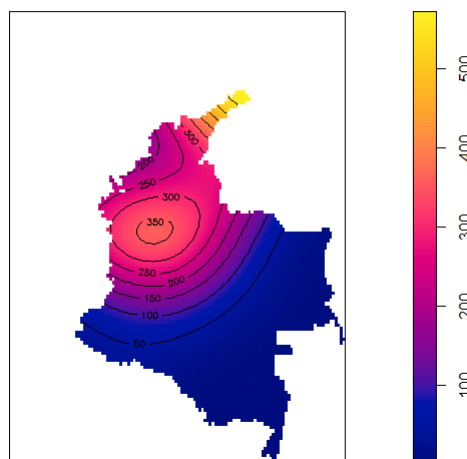


Figura 3: Densidad de captaciones hídricas superficiales

En la figura 3 se puede observar el mapa de densidad que representa gráficamente la concentración de agua superficial, se puede observar que el nivel de intensidad máximo se representa en la región Caribe, en la región Andina y Pacífica presenta niveles intermedios de intensidad de fuentes de captación hídricas superficiales.

### 3.6.2. Intensidad de fuentes hídricas subterráneas

Mapa densidad de fuente hídrica subterránea

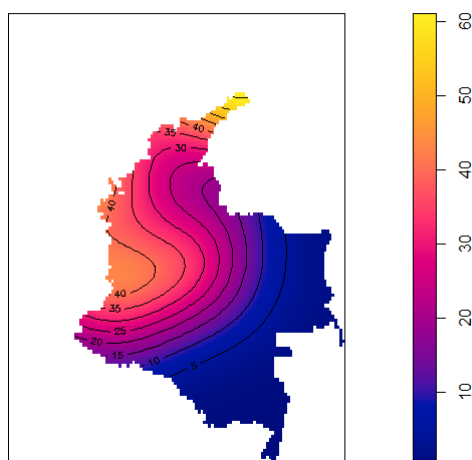


Figura 4: Densidad de captación hídricas subterránea

En la figura 4 se puede observar el mapa de densidad que representa gráficamente la concentración de agua subterránea, se puede observar que el nivel de intensidad máximo se representa en la región Caribe y Pacífica, en la región Andina se presentan niveles de intensidad intermedios de fuentes de captación hídricas subterráneas.

### 3.7. Método de cuadrantes

El análisis estadístico de los patrones de distribución de puntos requiere un modelo teórico como lo es el de aleatoriedad espacial completa, en este modelo se asume que los puntos son independientes y tienen una distribución de probabilidad uniforme sobre el área de estudio.

Para probar la hipótesis de aleatoriedad espacial completa se puede utilizar el método de los cuadrantes.

Para verificar si los datos están bajo aleatoriedad espacial se tiene la hipótesis:

$$H_0 : \text{Aleatoriedad Espacial}$$

$$H_1 : \text{No existe Aleatoriedad Espacial}$$

### 3.8. Correlación por medio de método de cuadrantes

La función de correlación se basa en las distancias dos a dos de los eventos del patrón puntual.

La manera que se plantea para estimar esta función de correlación es por medio del método de los cuadrantes, dado el número de eventos en cada subregión en la que se dividió el área de ubicación de los patrones puntuales.

## 4. Resultados

### 4.0.1. Simulación de patrones puntuales

Para realizar la verificación del método de cuadrantes propuesto se plantearon simulaciones de patrones aglomerados, generando una (o varias) realizaciones del proceso clúster en una región  $S \times T$ .

En esta región se encuentran las coordenadas para generar un proceso cluster, la intensidad ( $\lambda$ ) del proceso y el número de patrones a generar en la simulación, luego de generado el proceso se realizó el método de cuadrantes y se observó la correlación para diferentes escenarios.

Para llevar a cabo la simulación espacial de los procesos puntuales planteadas a continuación se utilizó el paquete *stpp* que proporciona herramientas para generar un proceso de puntos de agrupación tanto para procesos Poisson como Cluster además analiza las propiedades de segundo orden entre otros.

De esta librería se utilizó la función *rpcp* la cual genera una o varias realizaciones de patrones de puntos Clúster, los argumentos de la función son:

- *s.region*: Es la matriz de dos columnas que especifica las coordenadas donde se encontrarán los patrones, si este argumento falta se considera el cuadrado de la unidad.
- *t.region*: Es el vector que contiene los valores mínimo y máximo del intervalo de tiempo, si este argumento falta se considera el intervalo  $[0, 1]$ .
- *n.parents*: Se refiere a la distribución de los patrones puntuales la cual es Poisson con intensidad  $\lambda$ .
- *n.points*: Es el número de puntos para simular.
- *cluster*: Puede tomar la distribución uniforme, exponencial y normal.
- *lambda*: Es la intensidad del proceso el cual puede tomar un valor numérico, una función o una matriz.

Los diferentes escenarios fueron simulados con la siguiente función de intensidad (*lambda*).

$$\lambda(a, y, t) = ae^{-fy-t} \quad (13)$$

Donde  $a$  es una constante y es:

$$a = \frac{2000}{(1-e^{-4})(1-e^{-2})} \quad (14)$$

Se siguieron los siguientes pasos:

1. Se definió los parámetros de la función de intensidad ( $\lambda$ ), es importante observar que cada simulación de patrones cluster realizada tiene una función de intensidad diferente.
2. Se establecieron los argumentos de la función *rpcp* para crear los distintos patrones, para cada simulación se consideró un número diferente de puntos, se utilizó un cluster normal para todas las simulaciones.

3. Se crea el cuadrante con una longitud de  $10*10$  para cada patron simulado, por medio de un gráfico se observa el número de patrones en cada sub región
4. Por último se establece la correlación entre dos patrones simulados por medio del paquete de *stats* con el argumento *cor* de *Rstudio*

#### 4.0.2. Simulación de patrones Aglomerados o Clúster

En la figura 5, se pueden observar dos simulaciones de patrones, en el primer escenario donde se generaron 400 patrones puntuales con una intensidad igual a 150 y en el segundo patrón generado para este proceso Clúster tiene una intensidad igual a 50 y un número menor de puntos igual a 250.

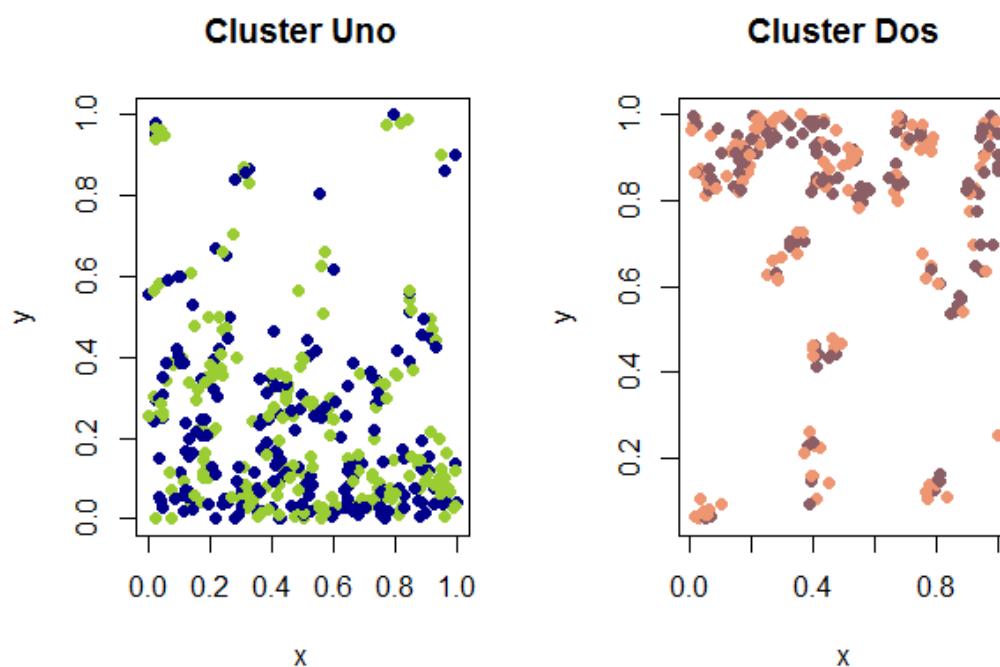


Figura 5: Patrones Aglomerados

Una vez generados los aleatorios del proceso se determino las subregiones en las que se particiona el cuadrante para este caso se hace un cuadrante de 100 subregiones, como se observa a continuación.

**Cuadrante del Cluster Uno**

7	0	0	0	0	0	0	2	2	2	
0	0	1	5	0	1	0	0	0	1	
0	0	1	0	0	0	0	0	0	0	
0	2	3	0	0	2	1	0	0	0	
4	3	1	0	1	1	0	0	5	0	
2	1	6	0	1	3	0	0	3	5	
7	1	0	8	5	1	2	3	6	4	0
10	9	1	5	10	11	4	5	0	1	
2	10	3	5	14	4	8	5	8	12	
8	13	9	21	12	15	20	24	14	15	

**Cuadrante del Cluster Dos**

8	13	20	8	10	4	8	11	0	16
12	11	0	1	17	9	9	0	1	10
0	0	0	5	0	2	0	0	0	2
0	0	6	2	0	0	0	4	2	6
0	0	0	0	0	0	0	0	7	0
0	0	0	3	10	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	5	2	0	0	0	0	1
1	0	0	3	2	0	0	5	4	0
9	0	0	1	0	0	0	0	0	0

Figura 6: Cuadrantes de patrones aglomerados

Se puede observar en la figura 6 la cantidad de patrones simulados que se encuentran por cada subregión en cada patrón cluster, tomando el el número de eventos por cuadrante, al realizar la correlación se obtuvo como resultado  $-0,2565601$ , se puede inferir que la correlación es negativa dado que la aglomeración del patrón uno es indirecta a la del patrón dos.

Se realizó una segunda simulación para generar nuevos aleatorios como se observa en la figura 7 se presentan dos simulaciones de patrones, en el cluster tres se generaron 300 patrones puntuales con una intensidad de 120 y en el cluster cuatro los patrones tienen una intensidad de 80 y un número de puntos igual a 100.

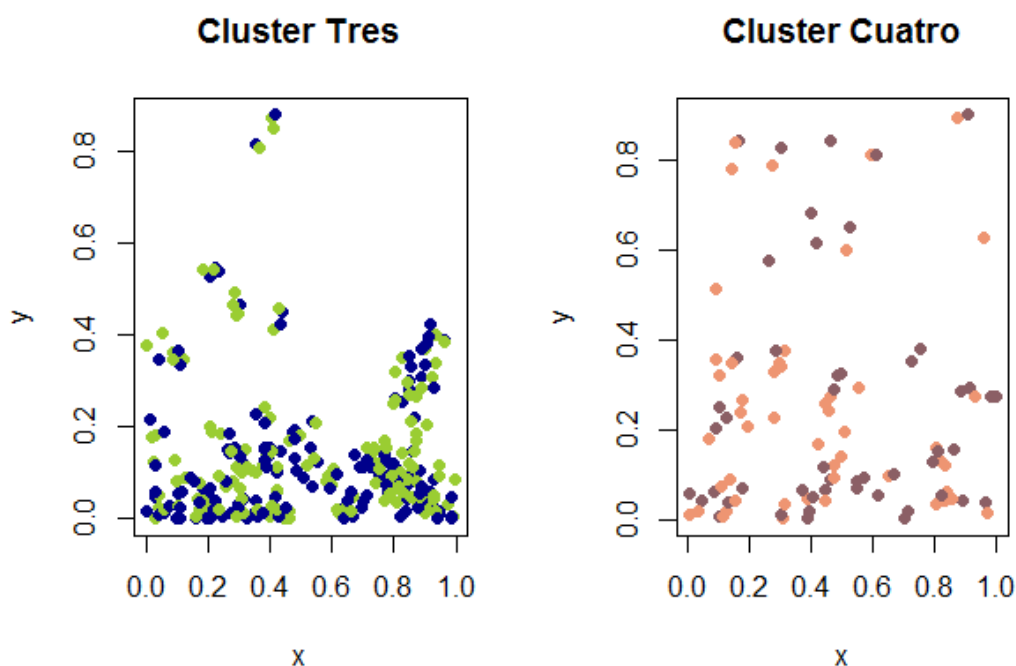


Figura 7: Patrones Aglomerados

Una vez generados los aleatorios del proceso se determinó las subregiones en las que se particiona el cuadrante para este caso se hace un cuadrante de 100 subregiones, como se observa a continuación.

**Cuadrante del Cluster Tres**

**Cuadrante del Cluster Cuatro**

0	0	0	0	0	0	0	0	0	0
0	0	0	2	3	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	1	4	0	0	0	0	0	0	0
1	0	4	1	4	0	0	0	0	2
4	3	0	0	0	0	0	7	8	
1	0	1	4	0	2	0	1	1	2
6	0	8	13	12	4	4	16	10	1
13	24	16	13	13	6	17	14	25	19

0	0	0	0	0	0	0	0	0	0
0	2	0	1	1	1	1	0	1	1
0	1	1	0	0	0	0	0	0	0
0	0	0	1	1	1	0	0	0	1
1	0	1	0	0	1	0	0	0	0
0	0	0	0	0	0	0	0	0	0
1	3	3	2	2	0	0	2	0	0
1	5	1	0	4	1	0	0	1	4
1	0	0	0	4	1	0	1	5	0
5	8	0	7	5	3	3	2	6	2

Figura 8: Cuadrantes de patrones aglomerados

Se puede observar en la figura 8 la cantidad de patrones simulados que se encuentran por cada subregión en cada patrón cluster, tomando el el número de eventos por cuadrante, al realizar la correlación se obtuvo como resultado 0.6083915, se puede inferir que estos patrones no son independientes, tienen una relación entre si.

Se realizó una nueva simulación para generar procesos aglomerados como se observa en la figura 9, en el cluster cinco se generaron 550 patrones puntuales con una intensidad de 250 y el cluster dos ya descrito anteriormente.

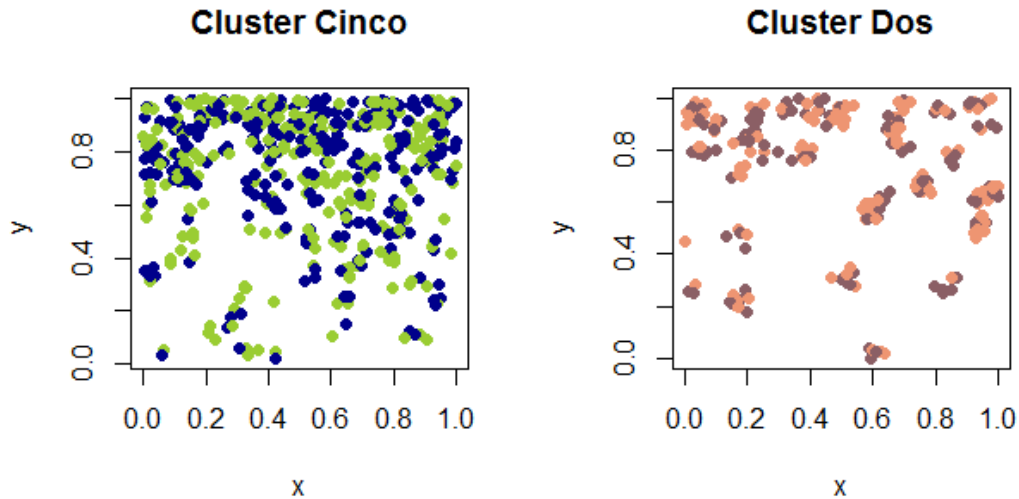


Figura 9: Patrones Aglomerados

Una vez generados los aleatorios del proceso se determinó las subregiones en las que se particiona el cuadrante para este caso se hace un cuadrante de 100 subregiones, como se observa a continuación.

**Cuadrante del Cluster Cinco**

9	2	0	1	6	1	8	1	7	1	8	8	3	2	1	0	0
1	3	1	6	3	9	1	2	1	7	8	4	1	3	1	0	0
1	4	1	8	2	3	5	8	9	1	0	4	1	4	0	0	0
5	9	0	7	4	8	7	6	5	2	0	0	0	0	0	0	0
2	3	0	2	4	3	6	6	1	0	1	0	0	0	0	0	0
0	5	0	1	1	5	4	8	5	5	0	0	0	0	0	0	0
7	1	0	0	0	4	9	1	4	1	0	0	0	0	0	0	0
0	0	1	4	1	0	4	1	0	5	0	0	0	0	0	0	0
0	0	5	1	0	0	2	0	4	0	0	0	0	0	0	0	0
2	0	1	4	2	0	0	0	1	1	0	0	0	0	0	0	0

**Cuadrante del Cluster Dos**

1	4	2	8	8	1	6	5	9	3	4	1	3	0	0	0	0
7	3	9	2	2	2	9	2	0	3	0	0	0	0	0	0	0
5	4	4	5	1	0	1	1	5	0	0	0	0	0	0	0	0
0	1	0	0	0	0	4	1	5	0	1	7	0	0	0	0	0
0	0	0	0	0	9	5	0	0	7	0	0	0	0	0	0	0
1	5	0	0	0	0	0	0	0	4	0	0	0	0	0	0	0
0	0	0	0	1	4	0	0	3	0	0	0	0	0	0	0	0
4	7	1	0	0	3	0	1	3	0	0	0	0	0	0	0	0
0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	3	3	0	0	0	0	0	0	0	0	0	0

Figura 10: Cuadrantes de patrones aglomerados

Se puede observar en la figura 10 la cantidad de patrones simulados que se encuentran por cada subregión en cada patrón cluster, tomando el el número de eventos por cuadrante, al realizar la correlación se obtuvo como resultado 0.3275267, se puede inferir que estos patrones no son independientes, tienen una relación entre si.

Se realizó una nueva simulación para generar procesos aglomerados como se observa en la figura 11, en el cluster seis se generaron 250 patrones puntuales con una intensidad de 400 y el cluster siete los patrones tienen una intensidad de 250 y un número de puntos igual a 550.

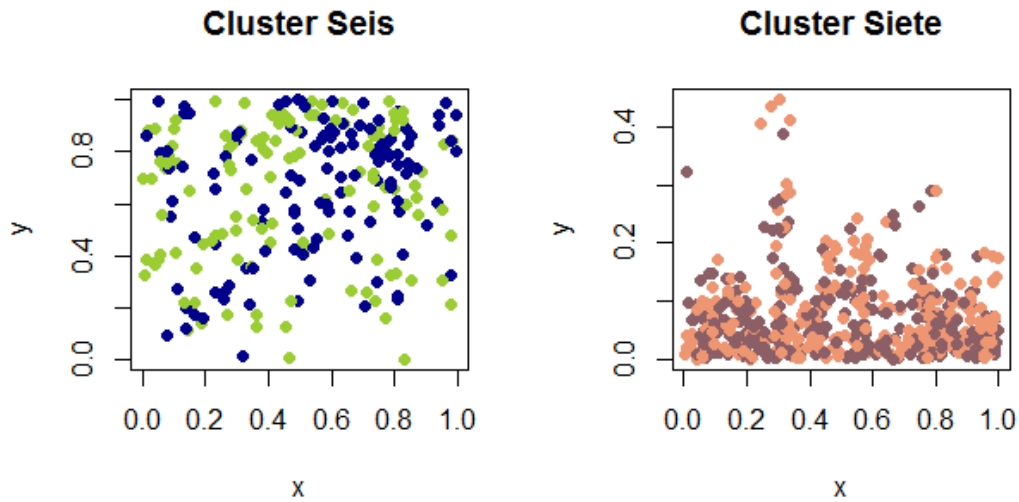


Figura 11: Patrones Aglomerados

Una vez generados los aleatorios del proceso se determinó las subregiones en las que se particiona el cuadrante para este caso se hace un cuadrante de 100 subregiones, como se observa a continuación.

**Cuadrante del Cluster Seis**

1	4	1	1	1	1	7	6	4	4	4
5	0	5	9	3	1	0	5	1	3	4
8	2	4	2	4	1	3	5	6	0	
3	1	1	1	3	2	3	6	4	1	
2	0	1	4	4	2	1	2	3	2	
1	3	5	1	2	6	1	0	1	1	
4	1	0	4	0	1	1	3	2	2	
0	4	4	1	2	0	1	3	2	1	
0	5	1	2	1	0	0	1	0	0	
1	0	0	1	1	0	0	0	1	0	

**Cuadrante del Cluster Siete**

0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0
0	0	2	2	0	0	0	0	0	0	0
1	0	0	2	0	0	0	0	0	0	0
0	0	4	9	1	4	3	2	1	0	
5	9	3	9	1	8	2	1	3	1	2
5	7	6	4	3	5	2	3	0	6	4
5	7	6	4	3	5	2	3	0	6	4

Figura 12: Cuadrantes de patrones aglomerados

Se puede observar en la figura 12 la cantidad de patrones simulados que se encuentran por cada subregión en cada patrón cluster, tomando el número de eventos por cuadrante, al realizar la correlación se obtuvo como resultado  $-0.328261$ , se puede inferir que la correlación es negativa dado que la aglomeración del patrón seis es indirecta a la del patrón siete.

Con los patrones simulados ya anteriormente descritos se realizó la correlación entre ellos descrita en la siguiente tabla:

Patrón Aglomerado	Coefficiente de correlación
1 y 3	0.650962
1 y 4	0.666022
1 y 5	-0.421089
1 y 6	-0.427714
1 y 7	0.7422724
2 y 3	-0.251913
2 y 4	-0.240920
2 y 5	0.3275267
2 y 6	0.3351206
2 y 7	-0.232696
3 y 5	-0.408894
3 y 6	-0.363001
3 y 7	0.598424
4 y 5	-0.268497
4 y 6	-0.307027
4 y 7	0.503730
5 y 6	0.378312
5 y 7	-0.311589
6 y 7	-0.328261

Tabla 1: Simulaciones patrones Aglomerados

#### 4.1. Método de cuadrantes aplicado en fuentes hídricas

Al contrastar la hipótesis de aleatoriedad espacial completa se tiene un valor p de la prueba del cuadrante el cual es menor que  $\alpha=0,05$ , por lo tanto se puede rechazar  $H_0$ , los puntos de fuentes hídricas superficial y subterránea no son completamente aleatorios desde el punto de vista espacial.

##### 4.1.1. Método de cuadrantes fuente hídrica superficial

El método de cuadrantes consiste en dividir el mapa del territorio colombiano donde se encuentran ubicadas las fuentes de captaciones hídricas superficiales en subregiones en este caso se dividieron en 100 subregiones.

Número de fuentes de agua superficial por cuadrante

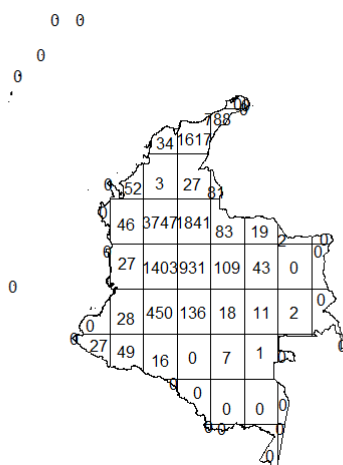


Figura 13: fuentes hídricas superficiales

En la figura 13 se puede observar el número de fuentes de agua superficial por cada subregión de cuadrantes en la que se dividió el territorio colombiano, en los departamentos de Santander y Antioquia se encuentra el mayor número de fuentes de captaciones hídricas

4.1.2. Método de cuadrantes fuente hídrica subterránea

La división del mapa del territorio colombiano donde se encuentran ubicados las fuentes de captaciones hídricas subterráneas se realizó en 100 cuadrantes al igual que las subregiones que se realizaron para las fuentes hídricas superficiales.

Número de fuentes de agua subterránea por cuadrante

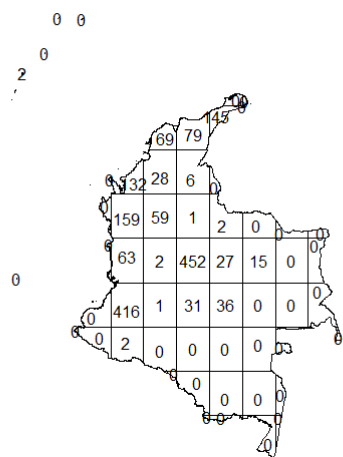


Figura 14: Fuentes subterránea

En la figura 14 se puede observar el número de fuentes de agua subterráneas por cada subregión de cuadrantes en la que se dividió el territorio colombiano, en los departamentos de Cundinamarca, Nariño y Cauca se encuentra el mayor número de fuentes de captaciones hídricas. Es relevante identificar que los departamentos como Vichada, Guainiá y Amazonas no presentan fuentes de captaciones hídricas ni superficiales ni subterráneas dada la naturaleza de la información utilizada.

Además se puede observar en las figuras 13 y 14 el número de eventos por cuadrante para los dos tipos de fuentes de captaciones hídricas, al realizar la correlación se obtuvo como resultado 0.1959365, se puede inferir que la correlación es *positiva* dado que los patrones tanto de fuentes de captaciones hídricas superficiales y subterráneas no son independientes, tienen una relación entre sí.

## 5. Conclusiones y trabajos futuros

- Se encuentra mediante las estadísticas de segundo orden que el patrón para el tipo de fuentes de captaciones hídricas en Colombia se clasifica como aglomerado, es decir, las fuentes superficiales y subterráneas se presentan en mayor cantidad en determinadas zonas.
- La metodología propuesta por medio de los cuadrantes da a conocer una forma de plantear la correlación entre dos patrones aglomerados.

Como trabajo futuro se podría explorar otro tipo de metodología basada en procesos Poisson para plantear la correlación entre dos o más patrones puntuales.

## 6. Agradecimientos

A mis padres por su constante y permanente apoyo, a mi tutor Wilmer Pineda por su orientación y enseñanza en este proceso y a mi Cotutor Jony Romero por su constante paciencia y conocimiento para culminar este trabajo.

## Referencias

- [1] BERMUDEZ,S. (2016). *Metodología para la evaluación espacio temporal de la accidentalidad vial en Bogotá: caso Avenida Boyacá*, (tesis de maestria).Universidad Nacional,Bogotá,Colombia.
- [2] DIGGLE,P.J (2003).*Statistical Analysis of Spatial Point Patterns*Vol. 1.
- [3] DIGGLE, P.J, & BRIX, A. (2001). *Predicción espaciotemporal para procesos log-gaussianos de Cox.*,Revista de la Real Sociedad de Estadística: Serie B (Metodología Estadística), núm. 63 (4), pp. 823-841. <https://doi.org/10.1111/1467-9868.00315>
- [4] GIRALDO,R (2011). *Estadística Espacial* NOTAS DE CLASE.[PDF],Recuperado de <https://www.unal.edu.co/Notas/Claseespaciotiempo>.
- [5] MÖLLER,J, & WAAGEPETERSEN,R. P.(2003). *Statistical Inference and Simulation for Spatial Point Processes*,Washington, D.C,Chapman& Hall/CRC.
- [6] OGATA, YOSIHIKO, & VERE-JONES. (2002). *Correction to Space-time model for regional seismicity and detection of crustal stress changes*,Journal of Geophysical Research.
- [7] OLAYA,R. (2015). *Modelo espacial de muertes por accidentes de tránsito en la zona urbana de Cali, Colombia durante el periodo 2004-2014*, (tesis de maestria).Universidad de Manizales,Colombia.
- [8] RATHBUN, S, & CRESSIE, A. (1994). *A Space-Time Survival Point Process for a Longleaf Pine Forest in Southern Georgia.*,Journal of the American Statistical Association, núm. 89, pp. 1164-1174.
- [9] RINCÓN,L (2012).*Introducción a los procesos estocásticos*,Universidad Nacional Autónoma de México,Facultad de Ciencias,Departamento de Matemáticas.