

---

# Modelo Espacio Temporal Poisson Aplicado al Hurto en Bogotá para el año 2016

Ricardo Yesid Godoy Castrillón.<sup>a</sup>  
ricardogodoy@usantotomas.edu.co

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

---

## Resumen

El uso de mapas y específicamente de los datos de área es común en estudios sobre epidemias, para este trabajo se va a aplicar el modelo espacio temporal Poisson al ámbito del hurto, así se puede identificar comportamientos de un patrón geográfico específico a lo largo del año 2016, además caracterizar regiones las cuales tendrían mayor o menor riesgo de ocurrencia o aparición del hecho y su tendencia en el transcurso de tiempo establecido.

—

**Palabras clave:** (Estadística espacio temporal).

## 1. INTRODUCCIÓN

Por tratarse de un conteo la variable a tratar en este trabajo (Hurto en Localidades de Bogotá) sería de característica discreta, la regresión lineal la cual asume supuestos de normalidad no sería la más adecuada para su modelación, aunque pertenece a la familia exponencial su rango se encuentra en los números reales, lo que lleva a tener en cuenta un modelo Poisson que tiene en la variable a explicar el número de ocurrencias del evento en un intervalo de tiempo y en una área geográfica determinada; en estos datos observados contrastados con los esperados se encuentran linealmente relacionados de la siguiente manera:

$$y_{it} \sim \text{Poisson}(E_{it}, \rho_{it})$$

Donde el valor esperado está dado por la transformación lineal:

$$\text{Log}(\rho_{i,t}) = b_0 + \mu_i + v_i + T_t$$

$Y_{i,j}$  son los casos observados en el área  $i$ ,

$E_i$  es el número de casos esperados en el área  $i$  con  $(i = 1, \dots, n)$

$b_0$  es el intercepto, cuantifica el resultado promedio en la región.

$v_i$  es el efecto específico intercambiable del área específica.

$\mu_i$  efecto en área específica estructurado espacialmente.

---

<sup>a</sup>Estudiante

<sup>b</sup>Director

$T_t$  clasificado como estructura paramétrica o no paramétrica.

Para los datos observados a nivel de área, se debe considerar la dependencia espacial dada por las vecindades.  $(s_1, \dots, s_n)$  son las áreas con bordes en común denominados "vecinos de primer orden." aquellos que comparte bordes con los vecinos de sus vecinos, "vecinos de segundo orden".

Con este trabajo se busca medir el efecto espacial y la tendencia diferencial temporal del hurto registrado en cada una de las localidades de la ciudad de Bogotá en el año 2016. También obtener un modelo espacio temporal el cual relacioné linealmente el efecto espacial y la tendencia temporal del hurto; indagar sobre el comportamiento de la variable y conocer en qué localidades de Bogotá es mayor el efecto espacial y cuales presentan tendencia temporal diferencial con valor menor y mayor al promedio general

## 2. Justificación

El hurto es un atentado al derecho que tiene la ciudadanía a hacer uso de sus bienes y pertenencias, junto con esto es función de las autoridades protegerlos ante cualquier amenaza; el artículo 2 de la Constitución colombiana señala que es función esencial del Estado proteger la vida, honra y bienes de los ciudadanos para esto cuenta con diferentes herramientas como una fuerza pública que debe tener la cooperación de la ciudadanía, y la participación activa de la sociedad civil; partiendo de lo anterior un estudio del tema utilizando técnicas estadísticas en el caso preciso un modelo espacio temporal es una contribución que realizaría la academia a la sociedad, y a la ampliación y divulgación del tema; además estas técnicas estadísticas pueden llegar a ser una herramienta para contrarrestar la delincuencia y en esta caso particular el hurto en Bogotá. Además puede ser un componente importante para el diagnóstico y la posterior evaluación de políticas en seguridad.

## 3. Antecedentes de estudio

La aplicación académica de los modelos espacio temporales se han enfocado en los estudios epidemiológicos, de estos destacamos los siguientes libros por su descripción detallada y rigurosa:

### 3.1. Bayesian disease mapping: hierarchical modeling in spatial epidemiology

(Andrew B Lawson)

En este texto se ponen a consideración conceptos y métodos propios de los modelos jerárquicos usados en epidemiología revisando diferentes aproximaciones de parámetros de estimación además el software utilizado.

### 3.2. Spatial and Spatio-temporal Bayesian Models with R-INLA.

(Marta Blangiardo y Michela Cameletti)

Este libro expone los fundamentos de la estadística bayesiana aplicada a las dimensiones espacio-temporal, teniendo en cuenta los aspectos teóricos del enfoque bayesiano; una profundización en los modelos espaciales y espacio-temporales; Ejemplos que relacionan la teoría estadística presentada a problemas de datos reales. Todo lo anterior codificado en el paquete INLA.

## 4. La Familia Exponencial

Cualquier modelo de regresión que usa linealidad describe la relación entre el valor esperado de la variable respuesta con un conjunto de variables explicativas, estos son conocidos como Modelos Lineales Generalizados (GLM), según Agresti los mencionados tienen tres componentes.(?)

### 4.1. Componente Aleatorio.

Específicamente se trata de la variable respuesta provenientes de una distribución con observaciones  $y = (y_1, y_2, \dots, y_n)^T$  independientes.

### 4.2. Predictor Lineal

Para el vector de parámetros  $\beta = (\beta_1, \beta_2, \dots, \beta_p)^T$  y una matriz  $X$  con tamaño  $n \times p$  con  $p$  variables explicativas y  $n$  observaciones, el vector de los valores esperados o predicciones es

$$\eta = X\beta,$$

donde  $\eta = (\eta_1, \eta_2, \dots, \eta_n)$  es el vector de valores esperados y cada predicción individual es dada por:

$$\eta_i = \sum_{j=1}^p \beta_j x_{ij},$$

con  $i = 1, \dots, n$  que es una combinación lineal del vector  $\beta$  con la variable respuesta  $x_i$ , de allí toma su nombre destacando la linealidad del modelo en los parámetros.

### 4.3. Función de enlace

Es la que conecta el componente aleatorio con el predictor lineal, dado  $\mu_i = E(y_i)$  con  $i = 1, \dots, n$ , la función que conecta a  $\eta_i$  con  $\mu_i$ , de la siguiente manera:  $\eta_i = g(\mu_i)$  es la función de enlace  $g(\cdot)$ , la cual debe cumplir las condiciones de ser monótona y diferenciable dos veces.

$$g(\mu_i) = \sum_{j=1}^p \beta_j x_{ij}$$

La función  $g(\cdot)$  transforma a  $\mu_i$  en su parámetro natural  $\eta_i$ . El componente aleatorio de un GLM es una variable respuesta con observaciones  $y_i$  independientes provenientes de una distribución de densidad de la forma:

$$f(y_i; \theta_i, \phi) = \exp \{ [y_i \theta_i - b(\theta_i)] / a(\phi) + c(y_i, \phi) \}$$

El conjunto de estas distribuciones es conocido como Familia Exponencial en la cual,

$\theta_i$  es el parámetro natural

$\phi$  es el parámetro de dispersión

Con el propósito de calcular el valor esperado y la varianza se aplica logaritmo natural a la expresión quedando:

$$L_i = \log(f(y_i; \theta_i, \phi))$$

$$L_i = [y_i \theta_i - b(\theta_i)] / a(\phi) + c(y_i, \phi)$$

y aplicando la primera y segunda derivada se tendría

$$\frac{\partial L_i}{\partial \theta_i} = [y_i \theta_i - b(\theta_i)]/a(\phi)$$

de la primera formula se tendría que:

$$\begin{aligned} E((y_i - b'(\theta))/a(\phi)) &= 0 \\ E(y_i) &= E(b'(\theta_i)) \\ \mu_i &= E(y_i) = b'(\theta_i) \end{aligned}$$

Y de la segunda fórmula se obtendría

$$b''(\theta_i)/a(\phi) = E[(y_i - b'(\theta_i))/a(\phi)]^2 = \text{var}(y_i)/a(\phi)^2$$

Por lo tanto

$$\text{var}(y_i) = b''(\theta_i)a(\phi)$$

#### 4.4. Verosimilitud de distribuciones pertenecientes a la Familia Exponencial.

Para n observaciones independientes la verosimilitud de las distribuciones de la Familia exponencial se puede expresar de la siguiente forma:

$$L(\beta) = \sum_{i=1}^n L_i = \sum_{i=1}^n \log f(y_i; \theta_i, \phi) = \sum_{i=1}^n \frac{y_i - b(\theta_i)}{a(\theta)} + \sum_{i=1}^n c(y_i, \phi).$$

si se hace  $\theta_i = \sum_j \beta_j x_{ij}$  y dado que  $a(\phi)$  es constante la log verosimilitud incluyendo los datos y el modelo de parámetros es:

$$\sum_{i=1}^n y_i \left( \sum_{j=1}^p \beta_j x_{ij} \right) = \sum_{j=1}^p \beta_j \left( \sum_{i=1}^n y_i x_{ij} \right)$$

con  $j = 1 \dots p$

La ecuación de verosimilitud haciendo  $\sum_j \beta_j x_{ij} = g(\mu)$  es

$$\frac{\partial L(\beta)}{\partial \beta_j} = \sum_{i=1}^n \frac{\partial L_i}{\partial \beta_j} = 0$$

Al diferenciar la log verosimilitud, se aplica la regla de la cadena.

$$\frac{\partial L_i}{\partial \beta_j} = \frac{\partial L_i}{\partial \theta_i} \frac{\partial \theta_i}{\partial \mu_i} \frac{\partial \mu_i}{\partial \eta_i} \frac{\partial \eta_i}{\partial \beta_j}$$

Puesto que  $\frac{\partial L_i}{\partial \theta_i} = (y_i \theta_i - b(\theta_i))/a(\phi)$  y que  $\mu_i = b'(\theta_i)$  y  $\text{var}(y_i) = b''(\theta_i)a(\theta)$  también se tiene que  $\frac{\partial \eta_i}{\partial \beta_j} = x_{ij}$  también  $\frac{\partial \mu_i}{\partial \eta_i}$  depende de la función de enlace, al sustituir se tiene que:

$$\frac{\partial L_i}{\partial \theta_i} \frac{\partial \theta_i}{\partial \mu_i} \frac{\partial \mu_i}{\partial \eta_i} \frac{\partial \eta_i}{\partial \beta_j} = \frac{(y_i - \mu_i)}{a(\theta)} \frac{a(\theta)}{\text{var}(y_i)} \frac{\partial \mu_i}{\partial \eta_i} x_{ij} = \frac{(y_i - \mu_i)x_{ij}}{\text{var}(y_i)} \frac{\partial \mu_i}{\partial \eta_i}$$

De esta manera la ecuación de la verosimilitud es

$$\frac{\partial L(\beta)}{\partial \beta_j} = \sum_{i=1}^n \frac{(y_i - \mu_i)x_{ij}}{\text{var}(y_i)} \frac{\partial \mu_i}{\partial \eta_i} = 0,$$

con  $j = 1, 2, \dots, p$  donde:  $\eta_i = \sum_{j=1}^p \beta_j x_{ij} = g(\mu_i)$ .

### 4.5. Matriz de covarianza para GLM.

Como se mostró anteriormente el estimador  $\beta$  es suficiente para las variables  $y_i$  y  $x_{ij}$ , adicional a esto el estimador  $\hat{\beta}$  es eficiente y se distribuye normal; para encontrar la matriz de covarianzas se utiliza la función de log verosimilitud, esta matriz es la inversa de una matriz de información denominada  $J$  la cual tiene elementos

$$E \left( \frac{-\partial^2 L_i}{\partial \beta_h \partial \beta_j} \right) = E \left( \frac{\partial L_i}{\partial \beta_h} \frac{\partial L_i}{\partial \beta_j} \right)$$

Y utilizando que  $\frac{\partial L_i}{\partial \beta_j} = \frac{(y_i - \mu_i)x_{ij}}{\text{var}(y_i)} \frac{\partial \mu_i}{\partial \eta_i}$  se llega a:

$$\begin{aligned} E \left( \frac{-\partial^2 L_i}{\partial \beta_h \partial \beta_j} \right) &= E \left( \frac{(y_i - \mu_i)x_{ih}}{\text{var}(y_i)} \frac{\partial \mu_i}{\partial \eta_i} \frac{(y_i - \mu_i)x_{ij}}{\text{var}(y_i)} \frac{\partial \mu_i}{\partial \eta_i} \right) \\ &= \frac{x_{ih}x_{ij}}{\text{var}(y_i)} \left( \frac{\partial \mu_i}{\partial \eta_i} \right)^2 \end{aligned}$$

Y dado que  $L(\beta) = \sum_{i=1}^n L_i$

$$E \left( \frac{-\partial^2 L(\beta)}{\partial \beta_h \partial \beta_j} \right) = \sum_{i=1}^n \frac{x_{ih}x_{ij}}{\text{var}(y_i)} \left( \frac{\partial \mu_i}{\partial \eta_i} \right)^2$$

Siendo  $W$  la matriz diagonal con diagonal principal:

$$w_i = \left( \frac{\partial \mu_i}{\partial \eta_i} \right)^2 / \text{var}(y_i)$$

De la matriz de información con la matriz del Modelo  $X$ . se llega a:

$$J = X^t W X.$$

La forma  $W$  y  $J$  corresponde a la función de enlace puesto que  $\frac{\partial \eta_i}{\partial \mu_i} = g'(\mu_i)$  concluyendo que para un GLM con  $\eta = X\beta$ ,  $(\hat{\beta})$  tiene una aproximación asintótica a una distribución normal  $N[\beta, (X^t W X)^{-1}]$ ; la matriz de covarianza asintótica es estimada por  $\hat{v}ar(\hat{\beta}) = (X^t \hat{W} X)^{-1}$

### 4.6. Deviance de los Modelos Lineales Generalizados.

Para un GLM con observaciones  $y_i = (y_1 y_2 \dots y_n)$  sea  $L(\mu; y)$  la función log verosimilitud expresada en términos de la media, también sea  $L(\hat{\mu}; y)$  la log verosimilitud para el modelo, y sea  $L(y; y)$  el modelo saturado es decir aquel que cuenta con todas las variables explicativas, la deviance compara el modelo escogido con el modelo saturado de la siguiente manera:

$$-2 \text{Log} \left[ \frac{\text{Maxima verosimilitud del modelo}}{\text{Maxima verosimilitud modelo saturado}} \right] = -2[L(\hat{\mu}; y)] - L(y; y)$$

Aplicando la expresión de la asintótica distribución para GLM se tendría:

$$-2[L(\hat{\mu}; y)] - L(y; y) = 2 \sum_i \frac{y_i - b(\tilde{\theta}_i)}{a(\theta)} = 2 \sum_i \frac{y_i - b(\hat{\theta}_i)}{a(\theta)}$$

Casi siempre  $a(\phi) = \phi/w_i$  y en este caso la expresión es:

$$D(y : \tilde{\mu})/\phi = 2 \sum_i w_i [y_i(\tilde{\theta}_i - \hat{\theta}_i) - b(\tilde{\theta}_i) + b(\hat{\theta}_i)]/\phi$$

$D(y; \hat{\mu})$  es conocido como deviance, ya que  $L(\hat{\mu}, y) \leq L(y; y)$  se tiene que  $D(y : \tilde{\mu}) \geq 0$  así que entre mayor sea el valor del deviance más pobre es el ajuste.(?).

## 5. Modelo de regresión Poisson

Dado que algunas variables aleatorias responden a eventos que suceden en condiciones limitadas de tiempo y espacio, como por ejemplo numero de accidentes por mes en una carretera para esto lo mas conveniente es utilizar la distribución Poisson, por su pertenencia a la familia exponencial es posible realizar un modelo lineal, la distribución de densidad Poisson se puede expresar como:

$$f(y_i; \mu_i) = \frac{e^{-\mu_i} \mu_i^{y_i}}{y_i!} = \exp(y_i \log(\mu_i) - \mu_i - \log(y_i!))$$

Si se hace  $\theta_i = \log(\mu_i)$  y  $\mu_i = e^{\log(\mu_i)}$  quedaría la siguiente expresión:

$$f(y_i; \mu_i) = \exp[y_i \theta_i - \exp(\theta_i) - \log(y_i!)]$$

En el cual, el parámetro natural es  $\theta_i$  con  $b(\theta_i) = \exp(\theta_i)$ ,  $a(\theta) = 1$ , y  $c(y_i, \phi) = -\log(y_i!)$  y como  $E(y_i) = b'(\theta)$  se tendría que

$$E(y_i) = b'(\theta) = \exp(\theta_i) = \mu_i$$

De manera análoga, se puede expresar la varianza, dado que  $\text{var}(y_i) = \beta''(\theta_i)$  se tendría que:

$$\text{var}(y_i) = \beta''(\theta_i) = \exp(\theta_i) = \mu_i$$

Lo cual es una característica propia de la distribución Poisson que su valor esperado es igual a su varianza.(?)

### 5.1. Ecuación de verosimilitud del modelo Poisson.

Para una variable aleatoria Poisson, la función de enlace es  $\eta_i = \log(\mu_i) = \sum_{j=1}^p \beta_j x_{ij}$  así que  $\mu_i = \exp(\eta)$  y  $\frac{\partial \mu_i}{\partial \eta_i} = \mu_i$  si se reemplaza en la ecuación de verosimilitud de los GLM se tendría que.

$$\sum_{i=1}^n (y_i - \mu_i) x_{ij} = 0$$

lo que muestra la suficiencia estadística  $\sum_i y_i x_{ij}$  para  $\beta$  y su valor esperado.

### 5.2. Deviance para el modelo Poisson

Dado que se tiene que  $\hat{\theta}_i = \log(\hat{\mu}_i)$  y  $b(\hat{\theta}_i) = \exp(\theta_i) = \mu_i$  de manera análoga para el modelo saturado  $\tilde{\theta}_i = \log(y_i)$  y  $b(\tilde{\theta}_i) = y_i$  la deviance para el modelo Poisson se expresa:

$$D(y; \hat{\mu}) = 2 \sum_i [y_i \log(y_i / \hat{\mu}) - y_i + \hat{\mu}_i]$$

Sí el modelo Poisson con función de enlace Log cuenta con termino de intercepto la ecuación de verosimilitud con los parámetros  $\sum_i y_i = \sum_i \hat{\mu}$  simplificando la deviance en la siguiente expresión:

$$D(y; \hat{\mu}) = 2 \sum_i y_i \log(y_i / \hat{\mu})$$

Paradigma Bayesiano En las estadística bayesiana, la probabilidad se utiliza como la medida fundamental o criterio de la incertidumbre. Dentro de este paradigma, es igualmente legítimo discutir la probabilidad de cualquier evento.

La simulación constituye una parte central de un análisis bayesiano muy aplicado, debido a la facilidad con la que las muestras a menudo pueden generarse a partir de una distribución de probabilidad. , así la función de densidad no se puede integrar explícitamente. Al realizar simulaciones, es útil considerar la dualidad entre una función de densidad de probabilidad y un histograma de un conjunto de dibujos aleatorios de la Distribución, dada una muestra lo suficientemente grande, el histograma puede proporcionar información prácticamente completa sobre la densidad y, en particular, varias muestras , momentos, percentiles y otras estadísticas de resumen proporcionan estimaciones de cualquier Aspecto de la distribución, a un nivel de precisión que puede ser estimado.gelman2013bayesian

Los problemas que enfrenta la estadística bayesiana encuentra una solución a través de métodos numéricos, lo anterior es consecuencia que las distribuciones a posterior, la mayoría de las veces son de complejidad mayor y se tiene solución cerrada de estas; de ahí la importancia de una implementación de métodos numéricos para solucionar, puesto que se requiere realizar integraciones sobre distribuciones de probabilidad que son difíciles de resolver a mano; por esta razón para conseguir inferencias sobre los parámetros de un modelo o predecir se requiere también el uso de herramientas computacionales.(correa2018introducción)

Algunos métodos de integración numérica son la cuadratura gaussiana o de Laplace (como el software INLA), el método de la integración Monte Carlo, la cual extrae muestras de la distribución de probabilidad tomando promedios que aproximen la esperanza matemática.

Vale la pena destacar que una cadena de Markov  $X_{(0)}, X_{(1)} \dots X_{(n)}$  definida en el espacio de estado  $x$  ,su distribución condicional de  $X_{(t)}$  dada  $X_{(0)}, X_{(1)}, \dots, X_{(t-1)}$  depende del antecesor  $X_{(t-1)}$  , de esta forma es condicionalmente independiente de todos los estados anteriores, tomando la siguiente expresión.

$$P(X^{(t+1)} | X^{(t)})$$

Comúnmente la anterior es conocida como probabilidad de transición; Una característica fundamental se basa en que la cadena en un estado próximo, su distribución se convierte en estacionaria, esto quiere decir que la distribución marginal de  $X(t)$  no cambia con  $t$ (?). Lo anterior requiere que una cadena de Markov tenga las siguientes características:

Irreducibilidad: Quiere decir que no importa el valor inicial  $X(0)$  la cadena llegara al estado  $x$  Una cadena de Markov  $X_1, X_2, \dots$  es irreducible si la cadena puede moverse libremente a través del espacio de estados; esto es, para dos estados cualesquiera  $x$  y  $x'$ , existe un  $n$  tal que,

$$P(x_n = x' | x_0 = x) > 0$$

Recurrencia positiva: Una cadena de Markov es recurrente positiva sí y sólo sí tiene una distribución estacionaria

Aperiodicidad (la cadena no sigue ningún ciclo ). Sí se entiende que un estado  $x$  tiene período  $d$  si  $P(X_{n+t} = x | X_t = x) = 0$  cuando  $n$  no es divisible por  $d$ , donde  $d$  es el mayor entero con esta propiedad. así que un estado  $x$  tiene período  $d = 1$  se dice que es aperiódico.

Bajo estas Condiciones de regularidad, la distribución estacionaria es también una distribución limitante en el sentido de que  $X(t)$  converge en distribución a  $\pi$

$$X^{(t)} \xrightarrow{d} \pi$$

cuando  $t \rightarrow +\infty$ , sin importar el valor de inicio. así la probabilidad  $\pi(A) = \int_A \pi(x) dx$  puede aproximarse por  $p(X(t) \in A)$  para todos los  $A \in \mathcal{X}$ .

## 6. Muestrador de Gibbs

Su objetivo es la simulación de muestras de funciones de distribución conjuntas no normalizadas, involucrando el muestreo de las distribuciones condicionales completas, buscando adicionalmente eficiencia

computacional.

Para obtener una muestra de la distribución conjunta  $p(X_1, \dots, X_d)$  el Muestreador Gibbs itera de acuerdo a este ciclo este ciclo:

Se toma una muestra  $X_1^{(i+1)}$  de  $p(X_1 | X_2^{(i)}, \dots, X_d^{(i)})$

Se toma una muestra  $X_2^{(i+1)}$  de  $p(X_2 | X_1^{(i+1)}, X_3^{(i)}, \dots, X_d^{(i)})$

Se toma una muestra  $X_d^{(i+1)}$  de  $p(X_d | X_1^{(i+1)})$

## 7. Método de Metropolis Hastings

El objetivo de este método es simular estados estacionarios  $\pi(\hat{\theta} | x)$  comenzando con un valor escogido inicial  $\theta_0$ . A continuación se describe el algoritmo:

Asignar un valor cualquiera para  $\theta(t)$ , suponiendo que  $\theta^{(t)} = \theta$ .

Generar  $\theta^*$  de  $q(\theta^* | \theta)$ . ahora  $\theta$  se define como un punto candidato y  $q$  una distribución propuesta.

Calcular

$$\alpha(\theta, \theta^*) = \min(1, \frac{\eta(\theta^*)q(\theta|\theta^*)}{\eta(\theta)q(\theta^*|\theta)})$$

Aceptar  $\theta(t+1) = \theta^*$  con probabilidad  $\alpha(\theta, \theta^*)$ .

En otro caso  $\theta^{(t+1)} = \theta$

La densidad objetivo  $\eta$  solo entra en al proceso a través del cociente  $\frac{\eta(\theta^*)}{\eta(\theta)}$  y por lo tanto no hay necesidad de conocer la constante de normalización para implementar el algoritmo, tal como lo señala Correa.

## 8. El algoritmo de aproximaciones de Laplace anidado integrado. (INLA)

El algoritmo de aproximaciones de Laplace anidado integrado. Este algoritmo de característica determinista es útil para la inferencia bayesiana, reemplaza la simulación tradicional antes mencionada, como MC y MCMC. INLA, diseñado para modelos gaussianos latentes y, en comparación con los MCMC, proporciona resultados con costos computacionales reducidos.(?)

El mencionado algoritmo determinista para la inferencia bayesiana (en lugar de la simulación, como MC y MCMC). INLA está especialmente diseñado para modelos gaussianos latentes y, en comparación con MCMC, proporciona resultados precisos en tiempos de computación más cortos.

En una camino diferente a la integración de MC vía simulación, se presenta la aproximación analítica con el método de Laplace. supóngase que se tiene la integral:

$$\int f(x)dx = \int \exp(\log(f(x)))dx$$

En la cual  $f(x)$  es la densidad de una variable aleatoria  $X$ . representamos  $\log f(x)$  a través de una expansión de la serie de Taylor evaluada en  $x = x_0$ :

$$\log(f(x)) = \log(f(x_0)) + (x - x_0) \frac{\partial \log(f(x))}{\partial x} \Big|_{x=x_0} + \frac{(x - x_0)^2}{2} \frac{\partial^2 \log(f(x))}{\partial x^2} \Big|_{x=x_0}$$

Sí se hace  $x_0$  igual al punto donde la función llega al máximo  $x^* = \operatorname{argmax}_x \log(f(x))$  entonces

$\frac{\partial \log(f(x))}{\partial x} \Big|_{x=x_0} = 0$  y la aproximación se convierte en

$$\log(f(x)) = \log(f(x^*)) + \frac{(x - x^*)^2}{2} \frac{\partial^2 \log(f(x))}{\partial x^2} \Big|_{x=x_0}$$

De esta forma la integral en general quedaría de la siguiente forma:

$$\int \log(f(x)) dx = \int \exp(\log(f(x^*)) + \frac{(x - x^*)^2}{2} \frac{\partial^2 \log(f(x))}{\partial x^2} \Big|_{x=x_0}) dx$$

Como la suma de exponentes en una exponencial se puede expresar como un producto;  $\exp(\log(f(x^*)))$  queda como una constante la cual se puede sacar de la integral

$$\int f(x) dx = \exp(\log(f(x^*))) \int \exp\left(\frac{(x - x^*)^2}{2} \frac{\partial^2 \log(f(x))}{\partial x^2} \Big|_{x=x_0}\right) dx$$

Se observa que el integrando se puede relacionar con la densidad de la distribución normal, si se hace  $\sigma_*^2 = 1 / \frac{\partial^2 \log(f(x))}{\partial x^2} \Big|_{x=x_0}$  se llega a la expresión:

$$\int f(x) dx \approx \exp(\log(f(x^*))) \int \exp\left(-\frac{(x - x^*)^2}{2\sigma_*^2}\right) dx$$

En el cual el integrando es el kernel de la distribución normal con media  $x^*$  y varianza  $\sigma_*^2$  y con mas precisión se podría decir que en intervalo  $(\alpha, \beta)$  es:

$$\int_{\alpha}^{\beta} f(x) dx \approx \sqrt{2\pi\sigma_*^2} (\phi(\alpha) - \phi(\beta))$$

## 9. RESULTADOS

### 9.1. Descripción de variables

Para presentar los resultados se decidió describir dos modelos; el primero con tendencia perimétrica en el cual se utiliza 5 variables explicativas y el tiempo, para explicar el hurto en las localidades, además del efecto espacial y el diferencial en temporal, a juicio del tesista la información que ofrece estos betas son útiles para el conocimiento del fenómeno a tratar, y para el hipotético escenario de una toma de decisiones; este modelo arroja un valor DIC de 3678.718.

El segundo modelo con interacción espacio temporal no cuenta con variables explicativas solo con el componente temporal y la respectiva interacción espacio tiempo, la justificación de este modelo se fundamenta el valor del Dic 2027.389 que es menor en comparación a otros con características similares. Blangiardo señala que si en un modelo espacio temporal se presentan covariables e interacción esta última será realizada con el residual que queda después de la modelación de los betas, en palabras del autor "Equivalentemente a lo que hemos presentado en el capítulo anterior, si covariables (por ejemplo, factores de riesgo, exposiciones, factores de confusión) están disponibles, los modelos espacio-temporales descritos anteriormente pueden incorporarlos fácilmente en el entorno de la fórmula y transformarse en regresiones ecológicas espacio-temporales. Entonces, la interpretación de los efectos espaciales, temporales y espacio-temporales sería residuales". Después de tener en cuenta las covariables incluidas en el modelo." (Traducción del tesista)

## 10. Modelo con tendencia paramétrica

Para este Modelo se utilizó una variable a explicar, que en este caso es denominada hurtos y adicional 5 variables explicativas que se explican a continuación:

Tabla 3.1

Variable	Descripción	Fuente
Hurto	Despojo a una persona de un bien cometido desde 1 de enero al 30 de Diciembre del año 2016	Dijin-Policía Nacional
Lesiones personales	Daño producido en alguna parte del cuerpo a causa de un golpe cometidas desde 1 de enero al 31 de Diciembre del año 2016	Dijin-Policía Nacional
Homicidios	Se refiere a la muerte de un ser humano causada por otra persona cometidas desde 1 de enero al 31 de Diciembre año 2016	Grupo Información de Criminalidad (GICRI) - DIJIN.
Robo de Motos	Despojo de Motocicletas	Dijin-Policía Nacional
Hurto a residencias	se refiere al apoderamiento de cosas o bienes, que pertenecen a una o varias personas que se encuentran al interior de una casa, apartamento, finca, etc.	Dijin-Policía Nacional

Tabla 1: Tabla de variables

Visualmente se puede mostrar el comportamiento de la variable hurto en las localidades a lo largo de los meses:

Con la anterior gráfica se nota que la variable en las localidades se caracterizan por un comportamiento constante durante el 2016, también es importante resaltar que el hurto se concentra más en la zona norte, y que el sur es menos proclive a este fenómeno.

A continuación se observara la variable de manera temporal:

La figura 3,2 indica que las localidades tienen un comportamiento similar en el tiempo, se nota nuevamente que la localidad de Suba, Usaquen y Rafael Uribe son aquellas que presentan mayor frecuencia de hurtos, aunque en Usaquen se comienza con frecuencia alta luego disminuye para seguir el comportamiento habitual de las demás series; también vale la pena destacar que las series temporales tienen tendencia positiva es decir incrementan a medida que pasan los meses.

Figura 1: Comportamiento espacial temporal de la variable hurto, año 2016

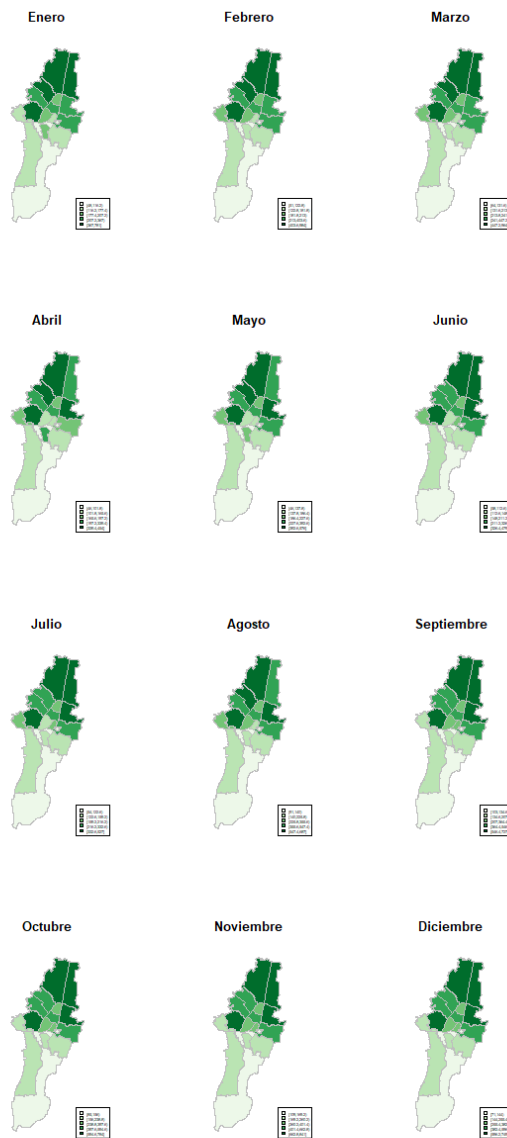
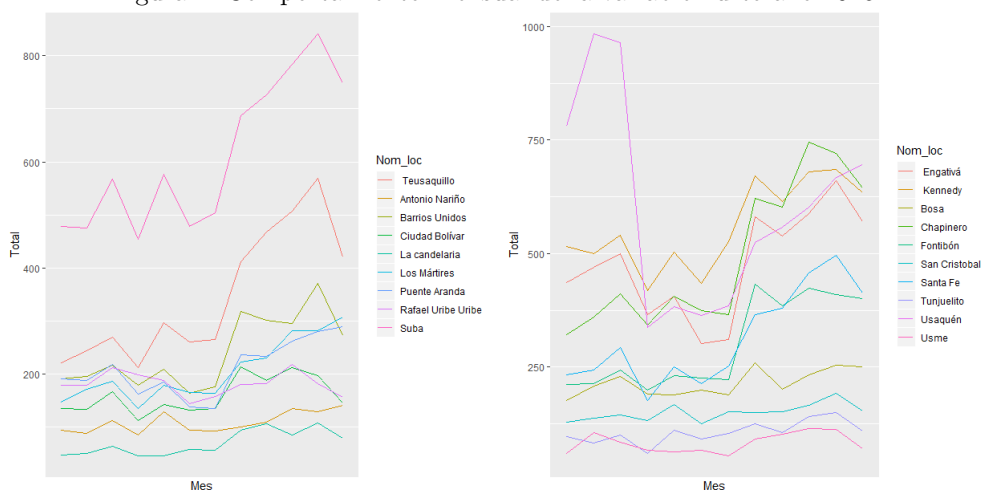


Figura 2: Comportamiento Mensual de la variable hurto año 2016



## 11. Test de Correlacion espacial

El modelo teórico exige una correlación espacial de la variable a explicar, es decir el modelo se valida a través de las pruebas de Moran; las cuales están basadas en el concepto de autocorrelación espacial de la variable a través del tiempo, consiste en la observación de los valores en áreas adyacentes que serán más similares que los esperados bajo el supuesto de independencia espacial (Cressie 1993) Vease (Bermudez y Niño) Las pruebas de hipótesis están planteadas de la siguiente manera:

$H_0$  : No posee correlación espacial la variable de interés

$H_a$  : Posee correlación espacial la variable de interés

El criterio de rechazo es  $pvalor < 0.05$

Este test que se Aplica a los diferentes estados que toma las localidades a través del 2016, rechaza todas las hipótesis nulas lo que demuestra la existencia de correlación espacial, y la posibilidad de realizar un análisis espacial.

Tabla 3.2

Mes	Pvalor	Mes	Pvalor
Enero	0.0173	Julio	0.03203
Febrero	0.01382	Agosto	0.0008249
Marzo	0.008544	Septiembre	0.0008005
Abril	0.00591	Octubre	0.002803
Mayo	0.01073	Noviembre	0.0006259
Junio	0.004943	Diciembre	0.0008478

Tabla 2: Tabla de pruebas de Morant

## 12. Resultados Modelo Espacio Temporal Paramétrico

Como se menciona en el marco teórico el modelo espacio temporal esta determinado

$$\text{Log}(\rho_{i,t}) = b_0 + \mu_i + v_i + T_t$$

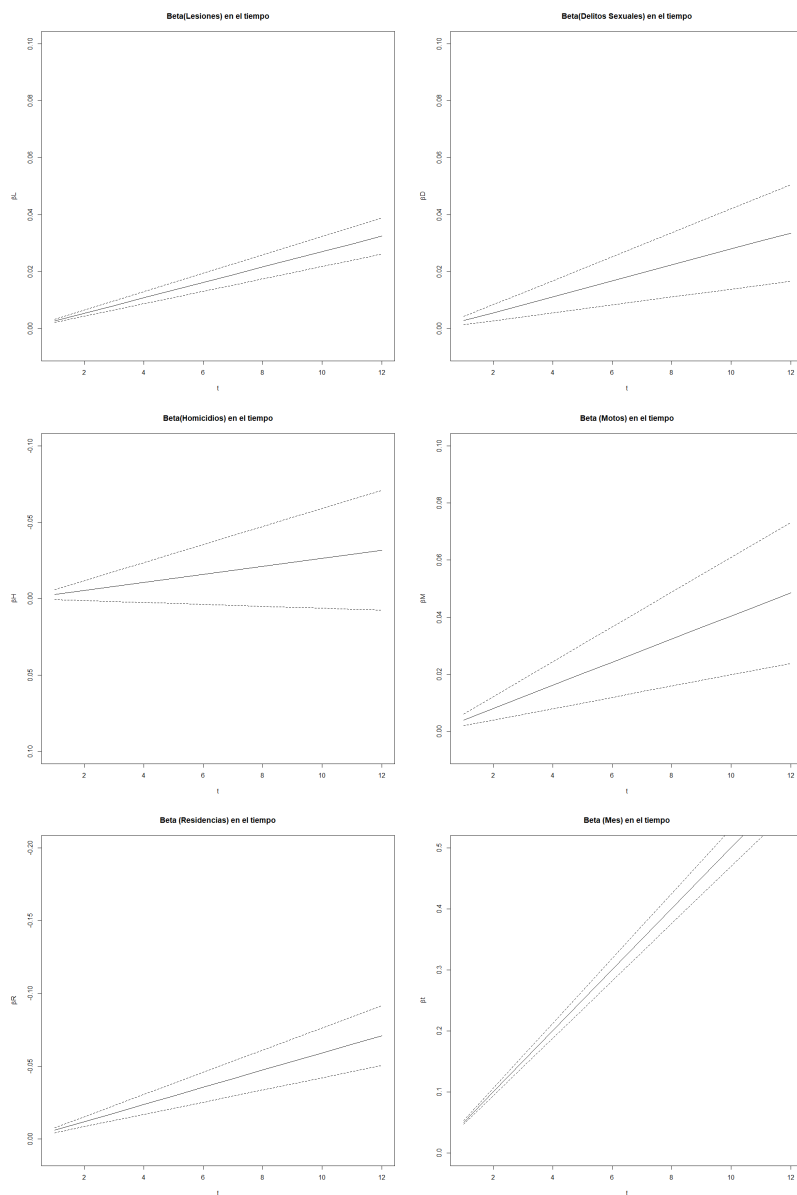
donde su primera componente son los betas provenientes de las variables explicativas estos son conocidos como efectos fijos del modelo(?)

Tabla 3.2

	mean	sd	0.025quant	0.5quant	0.975quant
(Intercept)	4.921	0.155	4.615	4.921	5.230
Lesiones	0.003	0.000	0.002	0.003	0.003
Dsexuales	0.003	0.001	0.001	0.003	0.004
Homicidios	-0.003	0.002	-0.006	-0.003	0.000
Motos	0.004	0.001	0.002	0.004	0.006
Residencias	-0.006	0.001	-0.008	-0.006	-0.004
Mes	0.049	0.002	0.046	0.049	0.052

En el periodo a evaluar, el intercepto equivale a la situación hipotética de que el valor de las otras variables fueran cero, es decir que no hubieran Lesiones, delitos sexuales, homicidios, robo a motos ni a residencias; de esta forma quedaría el conteo de hurtos en Bogotá, vale la pena destacar que la variable homicidios no resulto significativa en el modelo puesto que incluye el cero en el intervalo de credibilidad. todos los parámetros excepto Homicidios y Hurto a Residencias logran un efecto con pendiente positiva en el conteo de la variable Hurto, aunque cada factor fijo de cada variable se ve afectado por el efecto diferencial espacial especificado como  $(+i)t$ , estos Betas representan un efecto de tendencia lineal que se ve afectada por el transcurso del mes evaluado en los datos. una interpretación mas clara se observa analizando a través de los  $t : 1 : 12$  Meses, el comportamiento de los efectos fijos En la figura 3.3 se evidencia el efecto de tendencia que tiene las covariables durante el transcurso mensual del año 2016 con un 95 % de credibilidad en el intervalo; vale la pena destacar que el aporte de las variables Lesiones, hurto a residencias, robo de motos y delitos sexuales aumenta a medida que transcurrió el año 2016; también se resalta el hecho de que la variable tiempo es la que mayor pendiente presenta con un valor esperado de 0.049, lo que significa que su aporte se hace mayor en el transcurso del año, aportando mas en los últimos meses de este.

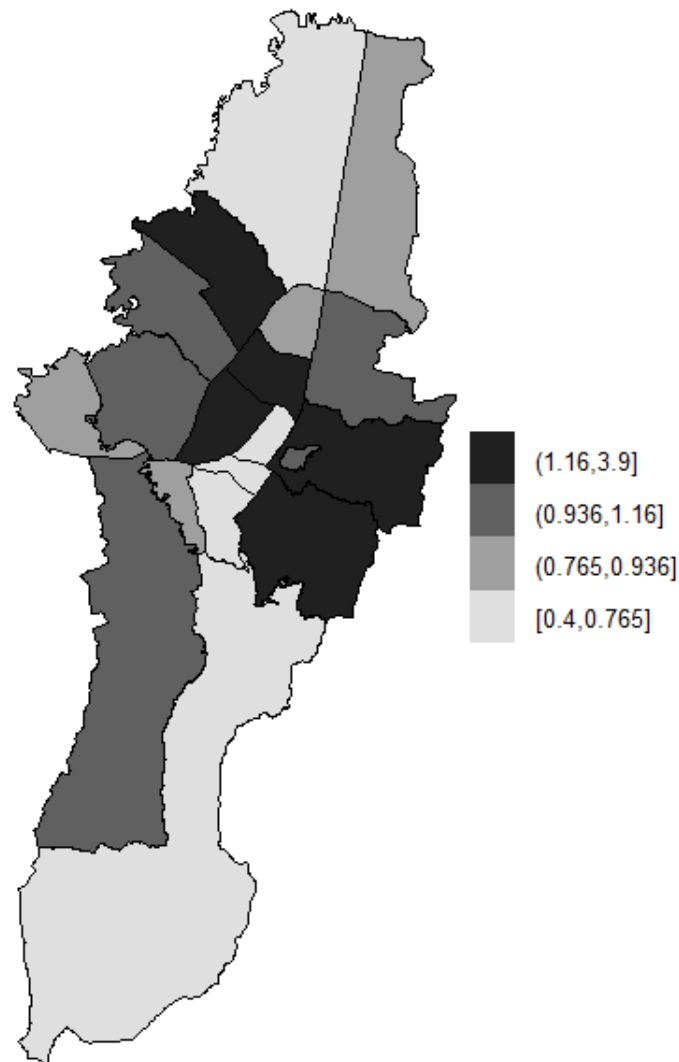
Figura 3: Tendencia temporal de la variables explicativas, año 2016



### 13. Efecto Espacial

El mencionado es obtenido por la expresión  $\zeta = \exp(\mu_i + \nu_i)$  donde  $\nu_i$  es el efecto específico de la localidad a la cual se le selecciona una prior no informativa y  $u_i$  es el efecto de área estructurado al que se asume una distribución intrínseca autoregresiva (BYM); de esta forma se representa el efecto de la localidad, más el efecto los vecinos de primer orden sobre esta. Según lo expuesto por (Bermudez y

Figura 4: Efecto espacial de la variable hurto: Año 2016



Niño) El parámetro es interpretado como el logaritmo natural del valor esperado que es afectado con la relación de área y sus vecinos, se evidencia que aumenta en la zona centro oriental y en las localidades de Engativá; aunque también es notable que en toda la ciudad este efecto es positivo lo que quiere decir que los vecinos de cada localidad aportan al conteo de hurtos en estas. lo mas destacable que se puede encontrar en el mapa de los efectos es la característica que en la periferia los efectos son menores, y se concentra en el centro.

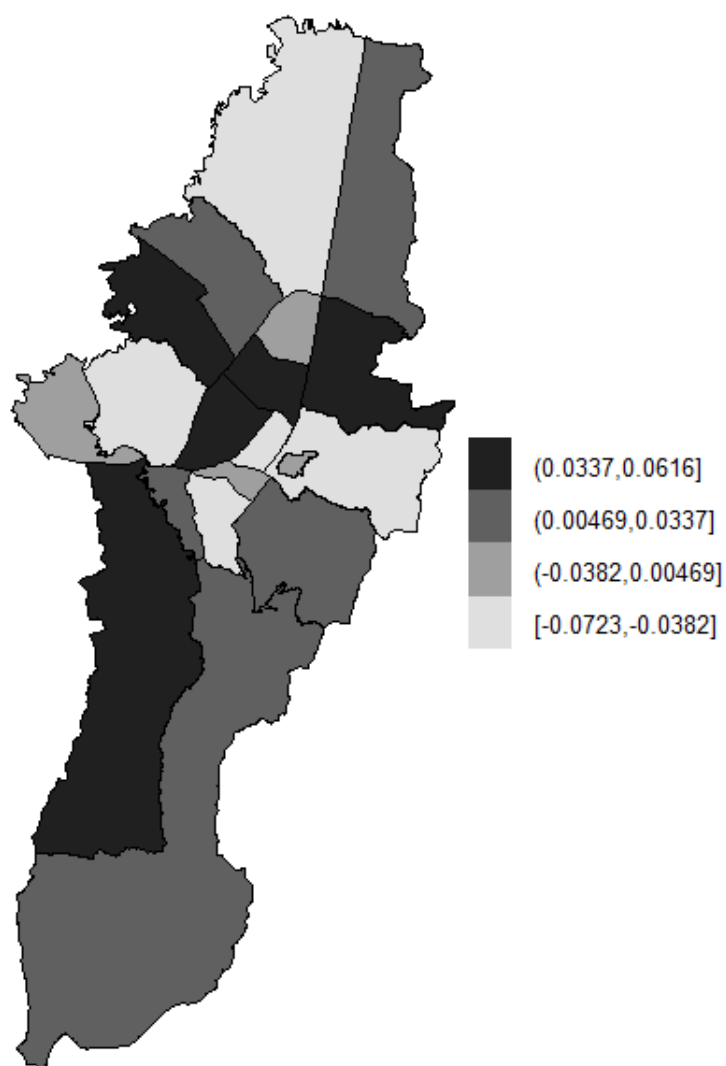
## 14. Efecto diferencial en el tiempo

Según Bernardinelli los cuales presentaron una tendencia paramétrica para el componente temporal en la que se asume que el predictor lineal puede escribirse

$$\eta_{it} = b_0 + \mu_i + \nu_i + (\beta + \delta_i)t$$

En el cual el efecto espacial principal o la tendencia lineal principal  $\beta$ , representa el efecto del tiempo global, y una tendencia diferencial  $\delta_i$  que identifica la interacción entre el tiempo y el espacio; puesto que para fines de identificación se impone una restricción de suma a cero sobre el vector  $\delta$ . Si  $\delta_i < 0$ , la tendencia específica del área es menos pronunciada que la tendencia media, mientras que  $\delta_i > 0$  implica que la tendencia específica del área es más pronunciada que la tendencia media.(?)

Figura 5: Efecto temporal de la variable hurto: Año 2016



En la zona sur específicamente en las localidades de Ciudad Bolívar y Usme se presenta un efecto diferencial superior al promedio igualmente Fontibón y Chapinero; inferior al promedio están Santa Fe

y Kennedy.

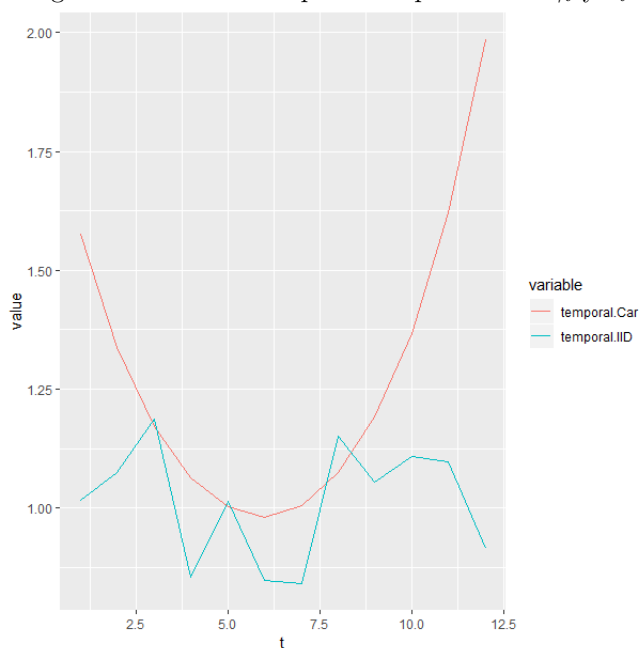
## 15. Modelo con interacción

Una extensión del modelo paramétrico con tendencia es el modelo con interacción el cual explica las diferentes tendencias a través del tiempo, el mencionado se expresa:

$$\eta_{it} = b_0 + \nu_i + \gamma_i + \phi_t + \delta_{it}$$

Adicional a los parámetros  $\beta_0$ ,  $\mu_i$  y  $\nu_i$  en este modelo se encuentra un parámetro de efecto estructural  $\gamma_t$ , y un parámetro de efecto no estructural  $\phi_t$  en el modelo en cuestión los resultados de ambos están representados en la siguiente gráfica:

Figura 6: Interacción espacio temporal entre  $\gamma_t$  y  $\nu_i$



Una tendencia similar a una ecuación cuadrática con un fuerte tendencia positiva en los últimos meses es notable para el efecto estructurado en el año 2016, mientras que el efecto no estructurado muestra algunas fluctuaciones en torno a 1.

### 15.1. Interacción

También el modelo cuenta con un vector de parámetros  $\delta$  sigue una distribución gaussiana con una matriz de precisión dada por,  $\tau_\delta R_\delta$  donde  $\tau_\delta$  es un escalar desconocido y  $R_\delta$  una matriz estructural;  $R_\delta$  puede ser factorizado por un producto kronecker de las correspondientes matrices de efectos que interactúan, Bernardetti presenta cuatro formas de interacción según la elección de efectos estructurales o no estructurales de los componentes espacial y temporal, destacando que el orden en que se escojan dichos efectos es importante.

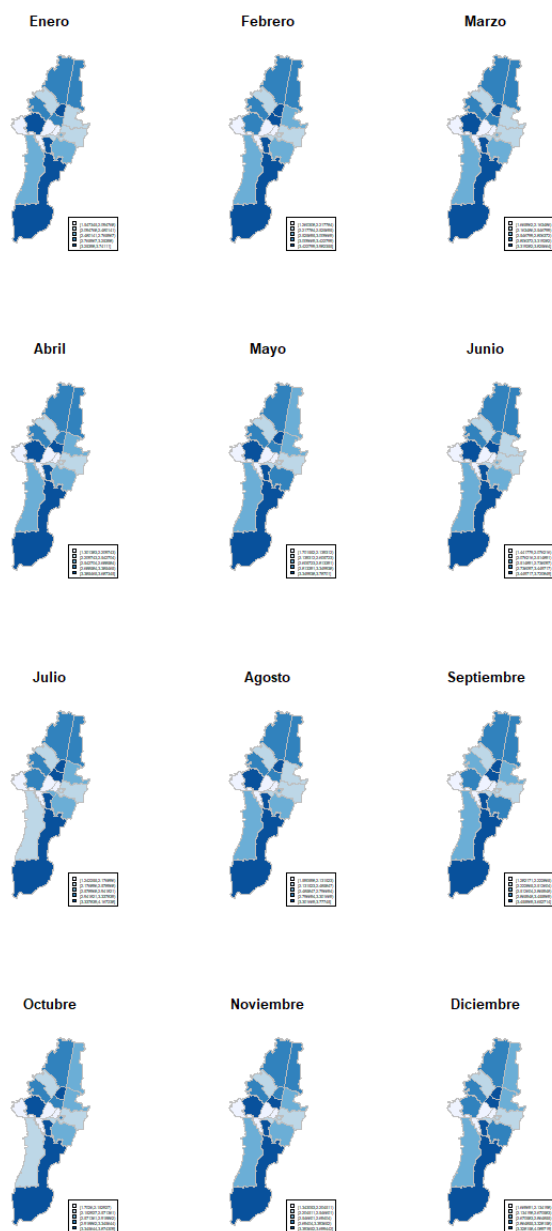
Para este modelo se escogió el efecto estructural temporal  $\gamma_t$  y el efecto no estructural  $\nu_i$  que se expresa:

$$R_\delta = R_\nu \otimes R_\gamma$$

Ambas estructuras  $R_\nu$  y  $R_\gamma$  especificada con una caminata aleatoria de segundo orden  $rw2^2$ ; de esta forma asumir que el vector  $\delta_{i1}, \dots, \delta_{iT}$  tiene una estructura autoregresiva en el componente temporal.

El gráfico muestra que el mapa en general mantiene constante es decir no se presenta una propagación espacial del fenómeno en las localidades durante el año 2016, El efecto espacial no es constante en Teusaquillo y Barrios Unidos lo que podría llevar a pensar que en estas localidades en el número de hurtos es afectado por sus vecinos de primer orden, dado que este efecto es alto; similar fenómeno ocurre en Usaquén aunque en esta la influencia tiende a disminuir

Disminuye el efecto espacial en Ciudad Bolívar y sobresale que siempre se mantiene en niveles bajos, podría decirse que el hurto obedece a dinámicas propias de la localidad. En la localidad de Usme disminuye juntamente con Ciudad Bolívar, aunque en Usme el efecto espacial es mayor

Figura 7: Interacción espacio temporal entre  $\gamma_t$  y  $\nu_i$ 

## 16. Conclusiones

Según el modelo se puede afirmar que durante el año 2016 en las localidades de Bogotá se presentó una tendencia al alza de la frecuencia de Hurtos, puesto que los betas correspondientes de la regresión Poisson aumentan con el transcurso de los meses, esto confirmaría lo observado en la descripción de los datos, también se evidencia en el gráfico de la tendencia estructural que es ascendente.

Una conclusión es que el efecto espacial es menor en la periferia y se concentra en el centro de la ciudad, en otras palabras va de la periferia al centro.

En cuanto al efecto temporal se concluye que en la zona central el hurto va en aumento en comparación a las demás localidades, se destaca la tendencia positiva en Ciudad Bolívar, y la tendencia negativa de la localidad Santa FE.

la interacción entre tiempo y espacio muestran que en el 2016 el hurto no se caracterizo por propagarse o extinguirse, aunque es un fenómeno que esta en aumento.

## Referencias

- [1] AGRESTI, A. *Foundations of linear and generalized linear models*. John Wiley & Sons, 2015.
- [2] BLANGIARDO, M., AND CAMELETTI, M. *Spatial and spatio-temporal Bayesian models with R-INLA*. John Wiley & Sons, 2015.
- [3] GAMERMAN, D., AND LOPES, H. F. *Markov chain Monte Carlo: stochastic simulation for Bayesian inference*. Chapman and Hall/CRC, 2006.
- [4] LAWSON, A. B. *Bayesian disease mapping: hierarchical modeling in spatial epidemiology*. Chapman and Hall/CRC, 2013.
- [5] LINDGREN, F., RUE, H., ET AL. Bayesian spatial modelling with r-inla. *Journal of Statistical Software* 63, 19 (2015), 1–25.
- [6] PAULA, G. A. *Modelos de regressão: com apoio computacional*. IME-USP São Paulo, 2004.
- [7] STROUP, W. W. *Generalized linear mixed models: modern concepts, methods and applications*. CRC press, 2016.