

---

# Modelos de regresión espacio-temporales Bayesianos aplicados al mercado de alojamiento AirBnB y HomeAway en Bogotá para el periodo 2015-2019

Gabriel Eduardo Moreno Veloza.<sup>1</sup>  
gabriel.moreno@usantotomas.edu.co

Wilmer Darío Pineda Ríos.<sup>2</sup>  
wilmerpineda@usantotomas.edu.co

---

## Resumen

Las plataformas digitales como Airbnb y HomeAway han crecido de manera significativa en los últimos años en el mercado de alojamiento turístico en Bogotá. El presente trabajo analiza el comportamiento de la tasa de ocupación, el número de propiedades y el ingreso total que reciben estos alojamientos por barrio, tomadas como variables dependientes, a través de modelos de regresión Beta, Poisson y Gamma espacio-temporales Bayesianos, para estimar y caracterizar posibles dinámicas de este mercado en el sector turístico de la ciudad entre el año 2015 y 2019. El estudio encuentra correlación lineal espacial y temporal entre las variables dependientes y las independientes en los años evaluados, especialmente en los barrios ubicados en las localidades de Chapinero, Teusaquillo, Santafé y Candelaria.

**Palabras clave:** Estadística espacial, Modelos Espacio Temporales, Modelo de Regresión Beta, Modelo de Regresión Poisson, Modelo de Regresión Gamma, Estadística Bayesiana.

## 1. Introducción

De acuerdo con los resultados obtenidos en la Encuesta Viajeros de Bogotá<sup>1</sup> para el año 2019, Bogotá fue visitado por 12,462,731 turistas, de los cuales 1,904,457 eran de origen internacional y 10,558,274 turistas de origen nacional, convirtiéndose en el principal destino al que se viaja en Colombia (Instituto Distrital de Turismo, 2020). Para el año 2020, el número de Prestadores de Servicios Turísticos - PST con Registro Nacional de Turismo - RNT vigente a corte de junio de 2020 fue de 4,052, de los cuales el 17 % fueron alojamientos lo que evidencia que la ciudad es el principal centro del país en oferta turística (Instituto Distrital de Turismo, 2020). Por estas razones, y por la posición privilegiada que tiene en el sector se hacía necesario analizar las posibles alternativas de hospedaje que se implementan en Bogotá para conocer si existe correlación entre variables relevantes de este mercado con otras de tipo espacial y temporal, y así, intentar de entender las preferencias de los turistas en la ciudad.

AirBnB<sup>2</sup> nació en el año 2008 en la ciudad de San Francisco, Estados Unidos, debido al sobrecupo hotelero que existía en una convención, optándose por rentar habitaciones de residencias no turísticas para cubrir la demanda; es una plataforma que ofrece espacios para hospedarse en ambientes distintos a los hoteles tradicionales; los anfitriones crean las reglas de estadía, establecen el precio y agregan fotos

---

<sup>1</sup>Estudiante de Maestría en Estadística Aplicada

<sup>2</sup>Director trabajo de grado, docente Maestría Estadística Aplicada

<sup>1</sup>Instituto Distrital de Turismo, (2020), en: <https://www.idt.gov.co/es/investigacion-viajeros-0>

<sup>2</sup>Airbnb,(2020), Pagina oficial,link página: <https://www.airbnb.com.co>

del inmueble en la plataforma, mostrando las ventajas y beneficios del inmueble. El huésped tiene una amplia variedad de opciones, entre una cama individual a una casa o condominio, pero también puede optar por alojamientos no tradicionales como casas rodantes o campings; el anfitrión, después de analizar los datos del cliente, decide rentar el inmueble o no hacerlo (Gallagher, 2018); la plataforma ofrece un seguro por incidentes, por lesiones o daños de la propiedad, pero no ofrece ningún tipo de cobertura para los huéspedes, quienes pueden ver calificaciones de otros para escoger la residencia como manera de asegurarse del lugar que quieren rentar. Airbnb ofrece la experiencia de vivir como un local, conocer a los anfitriones, en caso de que se desee, además de hacer hábitos tradicionales del lugar que se visite, e incluso vestir a la usanza del país o región.

Con sede en Austin, Texas, HomeAway fue creado en el año 2006 y opera de la misma manera que Airbnb, con la excepción de que ofrece principalmente alquileres de propiedades completas, por otro lado, se cobra una tarifa anual por suscripción, y la tarifa de alquiler es levemente mayor a la de Airbnb, esto hace que haya una menor y más selectiva oferta, lo que puede resultar atractivo para los anfitriones (Página Oficial HomeAway, 2020).

En Colombia las plataformas de este tipo no están reguladas y no son consideradas dentro del Registro Nacional de Turismo-RNT, sin embargo, y dado el número de establecimientos que existen en el país, son relevantes para evaluar su efecto en el turismo nacional. Según la plataforma airdna<sup>3</sup> que reúne bases de datos históricos e información tanto de Airbnb como de HomeAway en varios países, regiones y ciudades del mundo, existen más de 30 mil alojamientos entre activos e inactivos en Bogotá. Estas plataformas estarían activas con mayor fuerza desde el año 2015, pero existirían desde el año 2009.

En este trabajo se implementan 3 modelos espacio-temporales Bayesianos con datos de área para estimar las variables dependientes: tasa de ocupación, total de ingresos pagados en alquiler y número de propiedades por barrio, de las plataformas de alojamiento Airbnb y HomeAway en Bogotá entre el año 2015 y 2019, con el propósito de caracterizar el mercado de alojamientos alternativos de la ciudad y establecer si existe correlación con otras variables, con el espacio (barrio) y el tiempo (medido en los 5 años planteados). La implementación de modelos estocásticos se realiza dado que es un análisis de un conjunto de variables aleatorias que dependen de parámetros y están ligadas en el tiempo y el espacio, por esto la estimación debe realizarse con modelos diferentes a los tradicionales como los OLS<sup>4</sup>, además, los modelos incluyen 6 variables explicativas que pueden ayudar a entender la correlación con las variables analizadas.

En el turismo se conocen algunos trabajos donde se analizan patrones de demanda a través de análisis espacio-temporales utilizando pruebas de autocorrelación espacial (Andersson y Hoffmann, 2008; Deller, 2010; Liljenstolpe, 2011; Batista et al., 2018), otros trabajos utilizan estimaciones de densidad de Kernel y análisis de distancias euclidianas utilizando imágenes satelitales (Sun et al. 2013; Caldeira y Kastenholz, 2015; Niavis, 2020; Naimat, Wanggen y Shui, 2020), técnicas geoestadísticas de patrones puntuales (Chen et al. 2013), modelos de panel espacial dinámico, modelos de promedio móvil autorregresivo espacio-temporal (STARMA) (Yang y Zhang, 2019) y utilizando ubicación de teléfonos móviles para rastrear el movimiento de los turistas (Chen et al., 2018).

En el caso de los modelos espacio-temporales Bayesianos se conocen trabajos aplicados en el mercado de alojamiento y urbanismo (Sanchez, 2008; Gordziejczuk y Lucero, 2019; Xin, Zuolin y Rui, 2019), en el caso de la implementación de Aproximaciones de Laplace Integradas Anidadas (INLA) se observan trabajos que analizan la demanda y la elección de destinos turísticos (Zhang et al., 2018).

<sup>3</sup>La cual realiza análisis de Alquiler a corto plazo, <https://www.airdna.co/>

<sup>4</sup>Término en inglés para el concepto de Mínimos cuadrados ordinarios (OLS) es la técnica de regresión más conocida y básica en las que se correlacionan variables dependientes e independientes de manera lineal.

El presente trabajo se divide en 4 secciones, la primera, introductoria y conceptual; en la segunda, se presentan los modelos de regresión Beta, Poisson, Gamma y otros aspectos relevantes como la estadística espacial y estadística Bayesiana; la tercera, plantea la metodología espacio-temporal Bayesiana con INLA y la cuarta presenta los resultados prácticos con datos de Airbnb y HomeAway en Bogotá entre los años 2015 a 2019, tomando a los barrios catastrales como los datos de área a analizar.

## 1.1. Conceptualización

Desde su creación Airbnb se ha extendido a 191 países, y ofrece más de 4 millones de alojamientos superando a las principales marcas de hoteles combinadas. Existen varios estudios de plataformas digitales de alojamiento en ciudades turísticas a nivel mundial, incluso se han creado páginas que de forma gratuita muestran la competencia con el mercado de viviendas residenciales tales como Inside Airbnb (InsideAirbnb, 2019), evidenciando que el mercado de Airbnb no se reduce al componente vacacional sino que también puede captar huéspedes que desean quedarse periodos más largos, que es más cómodo para los clientes sin codeudores o altos niveles de documentación, así el precio final sea más costoso.

En Colombia son escasas las investigaciones sobre el impacto de dichas plataformas en la hotelería formal y en las características de la economía turística de las ciudades, e incluso en el mercado de arrendamientos para residentes. A finales del año 2017 la Asociación Hotelera y Turística de Colombia (Cotelco), publicó el primer estudio sobre el impacto de Airbnb en el mercado de alojamiento turístico en Bogotá, Cali y Cartagena buscando observar la relación entre el mercado de alojamiento formal y los alojamientos informales anunciados a través plataformas digitales; se denominan así, porque no están inscritos en el ordenamiento territorial, no está garantizada la seguridad del cliente ni la calidad del destino, también se critica el efecto sobre la contratación de personal en los hoteles y el pago de tributación desigual con alojamientos formales entre otras características (Ladevi, 2017); de acuerdo a ese estudio en Bogotá donde se realiza un análisis estadístico descriptivo, entre mayo de 2015 a abril de 2017 se reportaron más de 11.000 propiedades de las cuales 6.761 generaron ingresos y se registra una tasa de ocupación del 20 % e ingresos de US\$ 13.771.951, donde la mayoría de los alojamientos eran apartamentos y habitaciones (Ladevi, 2017); también hay portales que realizan experimentos con los datos de Airbnb para desarrollar ejercicios de aprendizaje de máquinas o aprendizaje profundo e incluso minería de datos y análisis de datos espaciales (Página web medium, 2018). Empero, son necesarios los estudios de inferencia estadística a través de modelos de regresión con datos de este tipo de plataformas de alojamiento para conocer más sobre el efecto en el mercado turístico de la ciudad, y si es posible que exista una correlación espacio-temporal con variables determinantes de este tipo de alojamiento, temas que no se han tratado con profundidad.

En este trabajo se desarrollará un modelo de regresión Beta espacio temporal Bayesiano tomando como variable dependiente la ocupación hotelera, siendo esta, una proporción o tasa resultado del cociente entre los días reservados sobre los días que el alojamiento esta disponible. Por otro lado, se plantea realizar un modelo de regresión Poisson espacio temporal Bayesiano, donde se pueda caracterizar el comportamiento del número de propiedades por barrio, y las posibles zonas donde haya una mayor correlación con la ubicación y periodos anteriores. También se implementa un modelo de regresión Gamma espacio temporal Bayesiano, para estimar los ingresos totales por barrio que pagaron los huéspedes durante los días que pernoctaron al año, es decir, el total de los ingresos que se cancelaron en los Airbnb y HomeAway por cada barrio, estableciendo el comportamiento de las ganancias y las variables que correlacionadas con los precios.

Como el análisis tiene un componente espacial, se implementa el cálculo del índice de Moran para medir la correlación espacial entre los barrios respecto a las variables analizadas, tasa de ocupación, ingresos totales y cantidad de propiedad. Para delimitar los vecindarios se utiliza una matriz de pesos espaciales, determinando cuáles y cuántas observaciones cercanas a cada barrio conforman el vecindario. En los modelos de regresión, esta matriz permite estudiar las relaciones entre las variables medidas y la estructura espacial de las propiedades en la ciudad.

## 2. Marco teórico y revisión literaria

### 2.1. Índice de Moran

En el índice de Moran se mide la autocorrelación espacial basada en las ubicaciones y los valores de las variables al mismo tiempo. En este caso, dado un conjunto de barrios y un atributo asociado el índice si el patrón está agrupado, disperso o es aleatorio (Moran, 1950) por medio de valores que oscilan entre +1 y -1, en donde +1 indica autocorrelación positiva perfecta, -1 expresa autocorrelación negativa perfecta y un valor 0 muestra la presencia de patrones completamente aleatorios en su distribución espacial, lo cual se expresa de la siguiente manera:

$$IM = \frac{n \sum_i \sum_j w_{ij} (Y_i - \bar{Y})(Y_j - \bar{Y})}{(\sum_{i \neq j} w_{ij}) \sum_i (Y_i - \bar{Y})^2} \quad (1)$$

donde  $n$  es el número de unidades geográficas en el mapa (barrios de Bogotá);  $w_{ij}$ , es la matriz de distancia que define si las áreas geográficas  $i$  y  $j$  son contiguas o no,  $Y_i$  es la variable de interés (para este caso: la tasa de ocupación, el total de ingresos y la cantidad de propiedades por barrio), y  $\bar{Y}$  es la media de la variable de interés. La hipótesis de la prueba establece que cuando:  $H_0$  : Índice Moran = 0 entonces no existe autocorrelación espacial, es decir, hay una distribución aleatoria de la variable de interés en el espacio geográfico comprendido por los barrios, esto se identifica cuando el valor  $p$  es superior 0.05. La hipótesis alternativa es  $H_1$  : Índice Moran  $\neq 0$  en la cual se rechaza la hipótesis nula y su valor  $p$  es igual o menor a 0.05, lo que significa que los barrios según las variables de interés están agrupados o dispersos, pero no de forma aleatoria, es decir, que no hay evidencia estadística para aceptar la hipótesis nula (Hidalgo, 2019).

La prueba presenta bajo el modelo nulo donde los  $Y_i$  son independientes e idénticamente distribuidas (i.i.d.) que  $IM$  tiene una distribución asintóticamente normal con media  $-1/(n-1)$  (Banerjee et al., 2004), y una varianza de la forma:

$$Var(IM) = \frac{n^2(n-1)S_1 - n(n-1)S_2 - 2S_0^2}{(n+1)(n-1)^2S_0^2} \quad (2)$$

donde,  $S_0 = \sum_{i \neq j} w_{ij}$ ,  $S_1 = \sum_{i \neq j} (w_{ij} + w_{ji})^2$ , y  $S_2 = \sum_k (\sum_j w_{kj} + \sum_j w_{jk})^2$ . El uso del  $IM$  de Moran se utiliza como una medida exploratoria en este trabajo, y no como una prueba de significancia espacial ni como un método para complementar las mediciones Bayesianas.

### 2.2. Modelo de Regresión Beta

El análisis de variables continuas que tienen como dominio el intervalo (0,1), son importantes debido a la existencia de una variada cantidad de mediciones relacionadas con la construcción de índices o tasas. Cepeda (2001) y Ferrari & Cribari-Neto (2004) proponen la distribución Beta como alternativa para modelar variables continuas en intervalo de forma clásica, mientras que Cepeda y Gamerman (2005), lo hacen desde el enfoque Bayesiano.

Ferrari y Cribari-Neto (2004) proponen la estimación por métodos clásicos basados en máxima verosimilitud, siguiendo la metodología de los Modelos Lineales Generalizados-MLG. Se proponen luego modelos no lineales doblemente generalizados con Simas (2010) y Cuervo y Achcar(2010); Cepeda (2001) y Cepeda y Gamerman (2005) desde la alternativa Bayesiana proponen una reparametrización de la Distribución Beta en función de su media  $\mu$  y dispersión  $\phi$  donde la limitante se presenta en función de la interpretación del parámetro  $\phi$ ; más adelante, los modelos con supuestos de correlación espacial son propuestos por Cepeda-Garrido et al.(2011) y recientemente se han presentado modelos de efectos fijos e incluso modelos mixtos para estructuras de datos más complejas (Gutierrez, 2014).

En el método Bayesiano propuesto por Cepeda-Cuervo (2001) para ajustar los modelos de regresión Beta, se obtienen muestras de la distribución posterior a partir de las distribuciones condicionales posteriores de dos bloques de parámetros, uno de la media, y otro del parámetro de dispersión, asumiendo distribuciones previas normales para los parámetros de regresión de dispersión y media (Cepeda y Gamerman, 2005). Las distribuciones posteriores condicionales son desconocidas y analíticamente no se podría abordar; por lo tanto, se aplican métodos computacionales como el algoritmo de Metropolis Hastings, el MCMC<sup>5</sup> y el método INLA, el cual se presenta más adelante.

Una variable aleatoria continua  $y$  que asume valores en el intervalo  $(0,1)$  y sigue una distribución condicional Beta con una función de densidad de probabilidad de la siguiente forma:

$$f(y|p, q) = \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} y^{p-1} (1-y)^{q-1}, \quad 0 < y < 1, \quad (3)$$

donde  $p, q > 0$  y  $\Gamma(\cdot)$  denota la función gamma. Con la reparametrización de la distribución Beta como una función de la media,  $\mu = E(Y)$  y el parámetro de precisión<sup>6</sup>  $\phi = p + q$ . La función de densidad Beta se puede reescribir como:

$$f(y|\phi, \mu) = \frac{\Gamma(\phi)}{\Gamma(\mu\phi)\Gamma(\phi(1-\mu))} y^{\mu\phi-1} (1-y)^{(1-\mu)\phi-1}, \quad 0 < y < 1, \quad (4)$$

donde,  $\mu = p/(p+q)$  (Ferrari y Cribari-Neto, 2004). En esta re-parametrización,  $p = \mu\phi$ ,  $q = \phi(1-\mu)$  y  $\sigma^2 = \frac{\mu(1-\mu)}{\phi+1}$ . A diferencia de la distribución gaussiana, la varianza depende de la media, que se traduce en varianza máxima en el centro de la distribución y mínima en los bordes, para respaldar la naturaleza truncada de la distribución Beta.

Después de la reparametrización  $Be(\mu, \phi)$ , un conjunto dado de observaciones o muestra aleatoria  $y_1, \dots, y_n$ , que representan proporciones, puede relacionarse con un conjunto de covariables y funciones utilizando un enfoque similar al modelo lineal generalizado, con la siguiente función de enlace, que asume el parámetro de precisión como constante:

$$g(\mu_i) = X_i^T \beta = \eta_i, \quad (5)$$

donde,  $\beta = (\beta_1, \dots, \beta_k)^T$  es un vector  $k \times 1$  de parámetros de regresión desconocidos, asumiendo que  $k < n$ ;  $X_i = (x_{i1}, \dots, x_{ik})^T$  es el vector de  $k$  covariables y  $\eta_i$  es un predictor lineal que se puede expresar,  $\eta_i = \beta_1 x_{i1} + \dots + \beta_k x_{ik}$ . En el programa estadístico R en la librería INLA se utiliza la función de enlace logit como primera opción, la cual es usada para el desarrollo del modelo en este trabajo. La función de la log-verosimilitud es  $l(\beta, \phi) = \sum_{i=1}^n l_i(\mu_i, \phi)$ , de la siguiente forma,

$$l(\mu_i, \phi) = \log \Gamma(\phi) - \log \Gamma(\mu_i \phi) - \log \Gamma((1-\mu_i)\phi) + (\mu_i \phi - 1) \log y_i + [(1-\mu_i)\phi - 1] \log(1-y_i), \quad (6)$$

el parámetro  $\mu = g^{-1}(X_i^T \beta)$  puede ser interpretado como una función de  $\beta$  y la estimación de los parámetros se realiza por el método de Máxima verosimilitud.

Una extensión del modelo empleada por Smithson y Verkuilen (2006) e introducida formalmente por Simas et al. (2010) es el modelo de regresión Beta de dispersión variable. En este modelo, el parámetro

<sup>5</sup>Por sus siglas en inglés, Markov Chain Monte Carlo, los cuales son una clase de métodos computacionales para extraer muestras de la distribución posterior articular. Estos métodos se basan en la construcción de una Cadena de Markov con distribución estacionaria la distribución posterior. Por lo tanto al muestrear repetidamente esta cadena de Markov, se obtienen muestras de la distribución posterior conjunta después de varias iteraciones, para la inferencia bayesiana tradicionalmente se utilizan el algoritmo Metropolis-Hastings y el muestreo de Gibbs.

<sup>6</sup>El parámetro  $\phi$  se conoce como parámetro de precisión ya que, para un  $\mu$  fijo, cuanto mayor  $\phi$  menor es la varianza de  $y$ ; esto significa que  $\phi^{-1}$  es el parámetro de dispersión

de precisión no es constante para todas las observaciones, sino que se modela de manera igual al parámetro de la media. Más específicamente con una distribución condicional,  $y_i \sim Be(\mu_i, \phi_i)$  independientemente,  $i = 1, \dots, n$ , y

$$g_1(\mu_i) = \eta_{1i} = X_i^T \beta, \quad (7)$$

$$g_2(\phi_i) = \eta_{2i} = Z_i^T \gamma, \quad (8)$$

donde,  $\beta = (\beta_1, \dots, \beta_k)^T$ ,  $\gamma = (\gamma_1, \dots, \gamma_h)^T$ ,  $k+h < n$ , son el conjunto de coeficientes de regresión en las dos ecuaciones,  $\eta_{1i}$  y  $\eta_{2i}$  son los predictores lineales, y  $X_i$  y  $Z_i$  son los vectores de las covariables. Como antes, ambos vectores de coeficientes son estimados por máxima verosimilitud, simplemente reemplazando  $\phi$  por  $\phi_i$  en la ecuación 5. En la propuesta desarrollada por Cepeda-Cuervo (2001),  $g_1$  es una función logit y  $g_2$  una función exponencial.

### 2.3. Modelo de Regresión Poisson

McCullagh y Nelder (1989) desarrollan los modelos lineales generalizados, donde incluyen a la Regresión Poisson como un caso especial. Hinde & Demetrio (2007) y Morales & López (2008) proponen formas de sortear los problemas de sobredispersión e implementan pruebas. La distribución Poisson condicional se establece de la siguiente forma:

$$f(y_i|\lambda_i) = \frac{\lambda_i^{y_i}}{y_i!} \exp(-\lambda_i) \quad (9)$$

Para las respuestas,  $y_i = 0, 1, 2, \dots$ , donde  $\lambda > 0$ , es el valor esperado y la varianza de  $y$ , las cuales se escriben dadas como:  $\mu = \lambda$  y  $\sigma^2 = \lambda$ .

La media esta vinculada al predictor lineal por:

$$\lambda_i = E_i \rho_i \quad (10)$$

donde  $E > 0$  es una constante conocida y  $\ln \rho_i = X_i^T \beta + u_i$  donde  $u_i \sim N(0, \sigma^2)$ . El modelo de la regresión Poisson con distribución condicional puede ser especificado de la siguiente forma:

$$f(y_i|\lambda_i) \sim \text{Poisson}(\lambda_i) \quad (11)$$

donde el parámetro de la distribución  $\lambda_i$  sigue una distribución log normal, con  $k \times 1$  vector de covariables  $X_i$ , y  $k \times 1$  vector de parámetros  $\beta$  dada la función de enlace es:

$$\ln(\lambda_i) = \ln(E_i \rho_i) \quad (12)$$

de esta forma,

$$\ln(\lambda_i) = \ln(E_i) + \ln(\rho_i)$$

donde,

$$\ln(\rho_i) = \beta_0 + \beta_1 X_1 + \dots + \beta_k X_k$$

Para este modelo la función de verosimilitud está dada de la siguiente forma:

$$l(y_i, X_i | \lambda_i) = \prod_{i=1}^n \frac{\exp(-\lambda_i) \lambda_i^{Y_i}}{Y_i!} \quad (13)$$

entonces,

$$l(y_i, X_i | \lambda_i) = \prod_{i=1}^n \exp(y_i \ln(\lambda_i) - \lambda_i - y_i!) \quad (14)$$

$$l(y_i, X_i | \lambda_i) = \prod_{i=1}^n \exp(y_i (\ln(\lambda_i) + \beta_0 + \beta_1 + \dots + \beta_n) - (E_i(\exp(\beta_0 + \beta_1 + \dots + \beta_n)) - y_i!))$$

la interpretación de los coeficientes se realiza a través de la función exponencial: elevando el intercepto  $\beta_0$  se muestra el recuento promedio de eventos en el área o período bajo estudio. De manera similar, exponenciar el genérico  $\beta_n$  devuelve el cambio en el promedio de  $y_i$  cuando  $X_n$  cambia en una unidad.

## 2.4. Modelo de Regresión Gamma

McCullagh y Nelder (1989) proponen el parámetro de media, y se considera constante el parámetro de forma. Por otro lado, Cepeda y Cuervo (2001), Cepeda y Gamerman (2005) proponen una estructura de regresión para la media y para el parámetro de forma simultáneamente, es decir, doblemente generalizados, implementando una metodología de máxima verosimilitud y bayesiana.

Una variable aleatoria  $Y$  sigue una distribución Gamma si la función de densidad de probabilidad está dada por:

$$f(y | \alpha, \lambda) = \frac{\lambda^\alpha y^{\alpha-1} e^{-\lambda y}}{\Gamma(\alpha)}, \quad 0 < y < \infty, \quad (15)$$

donde  $\alpha, \lambda > 0$ ; con esta notación la media y la varianza de  $Y$  están dadas por,  $E(Y) = \alpha/\lambda$ , y  $Var(Y) = \alpha/\lambda^2 = \mu^2/\alpha$ .

Con la reparametrización de la distribución Gamma como una función de la media,  $\mu = E(Y)$ , y el parámetro de forma  $\alpha$  dado que  $\lambda = \alpha/\mu$ , se puede escribir la función de densidad como,

$$f(y | \mu, \alpha) = \frac{1}{\Gamma(\alpha)} \left( \frac{\alpha y}{\mu} \right)^\alpha \exp(-\alpha y / \mu) \left( \frac{1}{y} \right), \quad 0 < y < \infty, \quad (16)$$

Sea  $Y_i \sim G(\mu_i, \alpha)$ ,  $i = 1, \dots, n$ , una muestra aleatoria de tamaño  $n$ . En el modelo de regresión Gamma con parámetro de forma  $\alpha$  constante, se define el siguiente modelo:

$$g(\mu_i) = \eta_i = X_i^T \beta, \quad (17)$$

donde  $g$  es la función de enlace,  $\beta = (\beta_0, \dots, \beta_p)'$  es el vector de los parámetros de la regresión,  $X_i$  son las covariables y  $\eta_i$  es un predictor lineal.

Las funciones de enlace habituales en la regresión Gamma son: la función logarítmica,  $g(\mu) = \log(\mu)$ ; la función de identidad,  $g(\mu) = \mu$ ; y la función inversa  $g(\mu) = 1/\mu$ . En los modelos lineales generalizados, el enlace canónico para la media es la función inversa. Una extensión de los modelos de regresión gamma se propone en Cepeda y Cuervo (2001), donde el parámetro de forma no es constante y se modela siguiendo una estructura de regresión, de la siguiente forma:

$$g(\mu_i) = \eta_1 i = x_i' \beta \quad (18)$$

$$h(\alpha_i) = \eta_2 i = z_i' \gamma \quad (19)$$

donde  $g$  y  $h$  son funciones de enlace,  $\beta = (\beta_0, \dots, \beta_p)'$  y  $\gamma = (\gamma_0, \dots, \gamma_k)'$  con  $p + h < n$ , son respectivamente, utilizando en ambos casos, una función logarítmica; los vectores de parámetros de la media y del parámetro de forma;  $x_i$  y  $z_i$  son respectivamente, la media y la forma de las variables explicativas,  $\eta_{1i}$  y  $\eta_{2i}$  son los predictores lineales.

Cepeda y Cuervo (2001) proponen un enfoque clásico para ajustar los modelos de regresión Gamma donde los parámetros de la media y la forma siguen las estructuras de regresión, utilizando el algoritmo de puntuación de Fisher. La función de verosimilitud se puede escribir de la siguiente forma:

$$L(\beta, \gamma) = \prod_{i=1}^n \frac{1}{\Gamma(\alpha_i)} \left( \frac{\alpha_i}{\mu_i} \right)^{\alpha_i} y_i^{\alpha_i-1} \exp \left( -\frac{\alpha_i}{\mu_i} y_i \right) \quad (20)$$

donde,  $\mu = X_i^T \beta$  y  $\alpha_i = \exp(Z_i^T \gamma)$ , y la función log-verosimilitud por:

$$l(\beta, \gamma) = \sum_{i=1}^n \left\{ -\log(\Gamma(\alpha_i)) + \alpha_i \log \left( \frac{\alpha_i y_i}{\mu_i} \right) - \log(y_i) - \left( \frac{\alpha_i}{\mu_i} \right) y_i \right\} \quad (21)$$

La parametrización de la distribución Gamma también se puede proponer de la siguiente forma,

$$\pi(y) = \frac{b^a}{\Gamma(a)} y^{a-1} \exp(-by), \quad a > 0, \quad b > 0, \quad y > 0 \quad (22)$$

donde  $E(y) = \mu = a/b$  y  $Var(y) = 1/\tau = a/b^2$ , entonces  $\mu$  es la media, y,  $\tau = (s\phi)/\mu^2$ . De esta forma,  $\phi$  es el parámetro de precisión (o  $1/\phi$  es el parámetro de dispersión) y  $s > 0$  es un escalar fijo, los cuáles generan la siguiente densidad,

$$f(y|\mu, \phi) = \frac{1}{\Gamma(s\phi)} \left( \frac{s\phi}{\mu} \right)^{(s\phi)} y^{(s\phi)-1} \exp \left( -(s\phi) \frac{y}{\mu} \right) \quad (23)$$

El predictor lineal  $\eta$  está vinculado a la media  $\mu$  mediante un enlace de registro predeterminado,

$$\mu = \exp(\eta) \quad (24)$$



## 2.5. Inferencia Bayesiana

En estadística Bayesiana todas las cantidades desconocidas en el modelo se tratan como variables aleatorias y el objetivo es estimar la distribución posterior conjunta de los parámetros. Es decir, la distribución del vector de parámetros,  $\theta$ , condicionado a los datos observados  $y$ . La forma en que se obtiene la distribución posterior se basa en el teorema de Bayes,

$$\pi(\theta|y) = \frac{\pi(y|\theta)\pi(\theta)}{\pi(y)} \quad (25)$$

donde  $\pi(y|\theta)$  es la verosimilitud de los datos  $y$  dados los parámetros  $\theta$ , los cuales toman valores en un paramétrico espacio  $\Theta$ ,  $\pi(\theta)$  es la distribución previa de los parámetros y  $\pi(y)$  es la probabilidad marginal, que actúa como una constante de normalización. La probabilidad marginal es difícil de calcular, ya que es igual a,

$$\int_{\Theta} \pi(y|\theta)\pi(\theta)d\theta \quad (26)$$

en la práctica la distribución posterior se estima sin calcular la probabilidad marginal, esta es la razón por la que el teorema de bayes se enuncia como,

$$\pi(\theta|y) \propto \pi(y|\theta)\pi(\theta) \quad (27)$$

Esto significa que la posterior se puede estimar reescalando el producto de la verosimilitud y la previa para que se integre en uno. La verosimilitud del modelo describe el proceso de generación de datos dados los parámetros, y la distribución previa refleja cualquier conocimiento previo sobre los parámetros del modelo. Cuando este conocimiento es escaso, se asumen previas no informativos para que la distribución posterior sea impulsada por los datos observados.

La distribución marginal posterior de cada elemento de  $\theta$  puede ser obtenido por la integración del resto de parámetros,

$$\pi(\theta_i|y) = \int \pi(\theta|y)d\theta_{-i} \quad i = 1, \dots, \dim(\theta) \quad (28)$$

donde  $\theta_{-i}$  representa el grupo de elementos en  $\theta$  menos  $\theta_i$ , y  $\dim(\theta)$  la extensión del vector de  $\theta$ . Las distribuciones marginales posteriores son muy útiles para resumir parámetros individuales y calcular estadísticas de resumen. Las magnitudes importantes son la media posterior, la desviación estándar, la moda y los cuantiles (Gomez-Rubio,2020).

### Criterio de Información de Desviación (DIC)<sup>7</sup>

Cuando el interés radica en la comparación entre diferentes modelos en términos de rendimiento, se puede utilizar su desviación. El Criterio de información de desviación (DIC) es una versión generalizada del criterio de información de Akaike o Criterio de información de desviación (AIC) (Rubalcaba, 2020). Desarrollado especialmente para la comparación de modelos bayesianos, es la suma de dos componentes, en el primero se cuantifica el ajuste del modelo y en el segundo se evalúa la complejidad del modelo.

---

<sup>7</sup>Se escribe DIC por sus siglas en ingles Deviance Information Criterion

Según Gelman et al. (2014), se define una media posterior como  $\hat{\theta}_{Bayes} = E(\theta|y)$  y  $k$  como una corrección de sesgo basado en datos. La medida de predicción es,

$$\widehat{elpd}_{DIC} = \log p(y|\hat{\theta}_{Bayes}) - p_{DIC}, \quad (29)$$

donde  $p_{DIC}$  es el número efectivo de parámetros, definido como,

$$p_{DIC} = 2 \left( \log p(y|\hat{\theta}_{Bayes}) - E_{post}(\log p(y|\theta)) \right) \quad (30)$$

donde la esperanza en el segundo término es una media de  $\theta$  sobre su distribución posterior. Esta es calculada usando simulaciones  $\theta^s$ , con  $s = 1, \dots, S$ , de la siguiente forma,

$$\text{computar } p_{DIC} = 2 \left( \log p(y|\hat{\theta}_{Bayes}) - \frac{1}{S} \sum_{s=1}^S (\log p(y|\theta^s)) \right) \quad (31)$$

La media posterior de  $\theta$  producirá la máxima densidad predictiva cuando sea igual que la moda y se puede producir una  $p_{DIC}$  negativa si la media posterior está lejos de la moda. Una versión alternativa de  $DIC$  utiliza una definición ligeramente diferente del número efectivo de parámetros:

$$p_{DICalt} = 2var_{post}(\log p(y|\theta)) \quad (32)$$

Tanto  $p_{DIC}$  como  $p_{DICalt}$  dan la respuesta correcta en el límite del modelo fijo y  $n$  grande y pueden derivarse de la distribución asintótica  $\chi^2$  (desplazada y escalada por un factor de  $-1/2$ ) de la log densidad predictiva. Para modelos lineales con distribuciones previas uniformes las dos medidas de tamaño de muestra efectivo se reducen a  $k$ . De estas dos mediciones,  $p_{DIC}$  es numéricamente más estable pero  $p_{DICalt}$  tiene la ventaja de ser siempre positivo. Comparado con propuestas anteriores para estimar el número efectivo de parámetros, la aproximación Monte Carlo más fácil y estable de  $DIC$  lo hicieron rápidamente popular. La cantidad real denominada  $DIC$  se define en términos de la desviación en lugar de la log densidad predictiva; por lo tanto,

$$DIC = -2 \log p(y|\hat{\theta}_{Bayes}) + 2p_{DIC} \quad (33)$$

### Campos Aleatorios Gaussianos de Markov latente (GMRF)<sup>8</sup>

El campo gaussiano latente  $x$  proporciona una herramienta flexible para modelar la dependencia temporal y espacial dentro de los datos y entre los datos y las posibles covariables. Los ejemplos típicos de modelos gaussianos latentes presentan un campo latente  $x$  de alta dimensión y un vector de parámetros  $\theta$  de baja dimensión, desde hace algunas décadas, la herramienta estándar para la inferencia Bayesiana en tales modelos de MCMC. Sin embargo, la estructura jerárquica del modelo, la alta dimensionalidad del campo latente  $x$ , y la fuerte correlación dentro de  $x$  y entre  $x$  y  $\theta$ , crean problemas para la convergencia y las propiedades de mezcla de la cadena de Markov. Para estos problemas se han planteado soluciones como utilizar un subconjunto de modelos gaussianos que no utilicen simulaciones MCMC, generando resultados similares en menos tiempo (Martino, 2007).

El núcleo de las técnicas de aproximación presentadas es una aproximación gaussiana al condicional completo del campo latente  $\pi(x|\theta, y)$  construido al emparejar la moda y la curvatura en la moda. Esto se indica como  $\tilde{\pi}_G(x|y, \theta)$ . Luego se construye una aproximación para la distribución marginal posterior de los parámetros  $\theta$  como,

<sup>8</sup>Se escribe GMRF por sus siglas en ingles Gaussian Markov Random Fields

$$\tilde{\pi}(\theta|y) \propto \frac{\pi(x, y, \theta)}{\tilde{\pi}_G(x|y, \theta)} \Big|_{x=x^*(\theta)} \quad (34)$$

donde  $x^*(\theta)$  es la configuración modal de  $\pi(x|\theta, y)$ . La aproximación en (34) es equivalente a la aproximación de Laplace de una distribución posterior marginal. Finalmente, los marginales posteriores para el campo latente se aproximan como,

$$\tilde{\pi}(x_i|y) = \sum \tilde{\pi}(x_i|y, \theta_k) \times \tilde{\pi}(\theta_k|y) \times \Delta_k \quad (35)$$

donde la suma es superior a los valores de  $\theta$  con pesos de área  $\Delta_k$  y  $\tilde{\pi}(x_i|y, \theta_k)$  son aproximaciones apropiadas a las densidades de  $x_i|y, \theta$ . El rendimiento de  $\tilde{\pi}(x_i|y)$  depende de la precisión de  $\pi(x_i|y, \theta_k)$  y del esquema de integración utilizado para calcular (35).

Un campo aleatorio gaussiano de Markov latente (GMRF)  $x = \{x_i : i \in V\}$  es un vector aleatorio gaussiano  $n = |V|$ -dimensional con independencia condicional adicional, o propiedades de Markov, es decir, asume que  $V = \{1, \dots, n\}$ . Las propiedades de independencia condicional se pueden representar mediante un grafo no dirigido  $G = (V, \varepsilon)$  con vértices  $V$  y aristas  $\varepsilon$ . Dos nodos,  $x_i$  y  $x_j$ , son condicionalmente independientes dados los elementos restantes en  $x$ , si y solo si  $\{i, j\} \notin \varepsilon$ . Entonces, se dice que  $x$  es un GMRF con respecto a  $G$ . Las aristas en  $\varepsilon$  están en correspondencia una a una con los elementos distintos de cero de la matriz de precisión de  $x$ ,  $Q$ , en el sentido de que  $\{i, j\} \in \varepsilon$  si y solo si  $Q_{ij} \neq 0$  para  $i \neq j$ . Cuando  $\{i, j\} \in \varepsilon$  decimos que  $i$  y  $j$  son vecinos, lo que se denota  $i \sim j$  (Martino, 2007).

Los GMRF tienen un amplio uso e importantes aplicaciones en análisis de series temporales estructurales, análisis de datos longitudinales y de supervivencia, modelos gráficos, regresión semiparamétrica, análisis de imágenes y estadísticas espaciales.

Una de las principales áreas de aplicación de los GMRF es la de los modelos jerárquicos (Bayesianos). Un modelo jerárquico se caracteriza por varias etapas observables y parámetros; la primera etapa consiste en supuestos distributivos para los observables condicionados en parámetros latentes; si se observa una serie de tiempo de conteo  $y$ , se puede asumir, para  $y_i$ ,  $i \in D \subset V$  una distribución de Poisson con media desconocida  $\lambda_i$ ; dados los parámetros del modelo de observación, se asume que las observaciones son condicionalmente independientes. La segunda etapa consiste en una herramienta previa para los parámetros latentes  $\lambda_i$  o, más a menudo, para una función particular de ellos (función de enlace); en el caso de la distribución Poisson se puede elegir una función de enlace exponencial y modelar las variables aleatorias  $x_i = \log(\lambda_i)$ . En esta etapa, los GMRF proporcionan una herramienta flexible para modelar la dependencia entre los parámetros latentes y, por tanto, la dependencia entre los datos observados. Esta dependencia puede ser de varios tipos: temporal, espacial o incluso espacio-temporal. La tercera etapa consta de distribuciones previas para los parámetros desconocidos  $\theta$ , estos son los parámetros de precisión en el GMRF. La distribución posterior es entonces,

$$\pi(x, \theta|y) \propto \pi(x|\theta)\pi(\theta) \prod_{i \in D} \pi(y_i|x_i) \quad (36)$$

si hay parámetros desconocidos en la verosimilitud, entonces se puede escribir de la siguiente forma, siendo el objetivo principal calcular los marginales posteriores:

$$\pi(x_i|y) = \int_{\theta} \int_{x_{-i}} \pi(x, \theta|y) dx_{-i} d\theta \quad (37)$$

donde  $x_{-i}$  es el vector de todos los  $x_j$  donde  $j \neq i$ , para cada  $i$  y marginales posteriores para los parámetros  $\theta_j$ . Dado que la integración analítica generalmente no es posible para la posterior  $\pi(x, \theta|y)$ , es común usar la inferencia basada en MCMC para estimar los marginales posteriores. Estos marginales pueden converger lentamente debido a la estructura jerárquica del problema (Rue & Held, 2005).

### Modelos condicionalmente autorregresivos (CAR)<sup>9</sup>

Los modelos condicionalmente autorregresivos (CAR) fueron introducidos por Besag (1974) y han aumentado su uso en el contexto del muestreo de Gibbs y más métodos generales de MCMC para ajustar modelos Bayesianos.

El término condicional en el modelo CAR se utiliza porque cada elemento del proceso aleatorio se especifica condicionalmente en los valores de los nodos vecinos. Sea  $Z_i$  una variable aleatoria en una  $i$ -locación, asumiendo de nuevo que la esperanza de  $Z_i$  es cero por simplicidad y sea  $z_j$  su valor realizado. El modelo CAR normalmente se especifica como,

$$Z_i | z_{-i} \sim N \left( \sum c_{i,j \neq 0}, c_{i,j}, z_j, m_{i,j} \right), \quad (38)$$

donde  $z_{-i}$  es el vector de todos los  $z_j$  donde  $j \neq i$ ,  $C$  es la matriz de dependencia espacial con  $c_{i,j}$  como el  $i, j$ -ésimo elemento,  $c_{i,i} = 0$ , y  $M$  es una matriz diagonal con elementos positivos en la diagonal  $m_{i,i}$ . Tenga en cuenta que  $m_{i,j}$  puede depender de los valores de la  $i$ -ésima fila de  $C$ . En esta parametrización la media condicional de cada  $Z_i$  está ponderada por los valores en los nodos vecinos. El componente de la varianza,  $m_{i,j}$ , a menudo varía con el nodo  $i$  y, por lo tanto,  $M$  es generalmente no estacionario. Cuando  $(I - C)^{-1}M$  es definida positiva,  $Z \sim N(0, \sum_{CAR})$ , con,

$$\sum_{CAR} = (I - C)^{-1}M \quad (39)$$

donde  $I$  es la matriz identidad y  $\sum_{CAR}$  debe ser simétrica, y por ende una matriz definida positiva, de la siguiente forma,

$$\frac{c_{i,j}}{m_{i,j}} = \frac{c_{j,i}}{m_{j,i}}, \quad \text{para todo } i, j. \quad (40)$$

si  $M$  es definida positiva, solo es necesario que  $(IC)$  tenga valores propios positivos para que  $\sum_{CAR}$  sea definida positiva. En resumen, deben cumplirse las siguientes condiciones para que  $\sum_{CAR}$  sea una matriz de covarianza CAR:  $(I - C)$  tenga valores propios positivos,  $M$  es una diagonal con elementos positivos,  $c_{i,i} = 0$  para todo  $i$ , y  $c_{i,j}/m_{i,j} = c_{j,i}/m_{j,i}$ , para todo  $i, j$  (Hoef, Hans y Hooten, 2017).

Para el caso Gausiano o autonormal se supone que,

$$Y_i | y_j, j \neq i \sim N \left( \sum_j b_{ij} y_j, \tau_i^2 \right), \quad i = 1, \dots, n \quad (41)$$

la cual se puede desarrollar de la siguiente forma,

$$p(y_1, \dots, y_n) \propto \exp \left\{ -\frac{1}{2} y' D^{-1} (I - B) y \right\}, \quad (42)$$

donde  $B = b_{ij}$  y  $D$  es la diagonal con  $D_{ii} = \tau_i^2$ . Esta expresión sugiere una distribución normal conjunta multivariada con media 0 y matriz de varianza  $\sum_y = (I - B)^{-1}D$ . Para saber si la matriz de varianza es simétrica las condiciones son las siguientes,

$$\frac{b_{ij}}{\tau_i^2} = \frac{b_{ji}}{\tau_j^2}, \quad \text{para todo } i, j. \quad (43)$$

<sup>9</sup>Por sus siglas en ingles Conditionally Autoregressive Models (CAR)

dado que  $B$  no es simétrico se vuelve sobre la matriz de distancias  $W$ , la cual se supone simétrica, se supone un set  $b_{ij} = \frac{w_{ij}}{w_{i+}}$  y  $\tau_i^2 = \frac{\tau^2}{w_{i+}}$ , de esta forma las anteriores ecuaciones son satisfechas y se produce  $p(y_i|y_j, j \neq i) = N\left(\sum_j w_{ij}y_j/w_{i+}\right)$ , y tambien se satisface a través de,

$$p(y_1, \dots, y_n) \propto \exp\left\{-\frac{1}{2\tau^2}y'(D_w - W)y\right\}, \quad (44)$$

donde  $D_w$  es una diangonal con  $(D_w)_{ii} = w_{i+}$ , se observa tambien un segundo problema,  $(D_w - W)1 = 0$ ,  $\sum_y^{-1}$  es una matriz singular, de modo que  $\sum_y$  no existe y la distribución en (38) es incorrecta, y con un poco de algebra se puede escribir,

$$p(y_1, \dots, y_n) \propto \exp\left\{-\frac{1}{2\tau^2} \sum_{i \neq j} w_{ij}(y_i - y_j)^2\right\} \quad (45)$$

los problemas de (38) pueden arreglarse redefiniendo  $\sum_y^{-1} = D_w - \rho W$  y escogiendo  $\rho$  para hacer  $\sum_y^{-1}$  no singular. Esto es garantizado si  $\rho \in (1/\lambda_{(1)}, 1/\lambda_{(n)})$ , donde  $\lambda_{(1)} < 1/\lambda_{(2)} < \dots < 1/\lambda_{(n)}$  son los valores propios ordenados de  $D_w^{-1/2}$ , y además dado que  $\text{tr}(D_w^{-1/2}W D_w^{-1/2}) = 0 = \sum_{i=1}^n \lambda_{(i)}$ ,  $\lambda_{(1)} < 0$ ,  $\lambda_{(n)} > 0$ , y 0 pertenece a  $(1/\lambda_{(1)}, 1/\lambda_{(n)})$ .

Los limites más simples pueden obtenerse reemplazando la matriz  $W$  por la escalada matriz  $\widetilde{W} \equiv \text{Diag}(1/w_{i+})W$ ; donde,  $\widetilde{W}$  no es simétrica, pero todas sus filas son estocásticas.  $\sum_y^{-1}$  puede ser escrita como  $M^{-1}(F - \alpha\widetilde{W})$  donde  $M$  es una diagonal, y si  $|\alpha| < 1$ ,  $F - \alpha\widetilde{W}$  es no singular (Banerjee, Carlin y Gelfand, 2004).

### 3. Metodología Espacio-Temporal Bayesiana con INLA

Los modelos lineales generalizados (MLG) permiten respuestas correlacionadas a través de la inclusión de efectos aleatorios en el predictor lineal, amplían la gama de posibles aplicaciones más allá de los modelos lineales generalizados ordinarios, pero implican problemas computacionales más difíciles y desafiantes, para esto, se proponen varios procedimientos de inferencia.

Un análisis Bayesiano se realiza a menudo utilizando la Cadena de Markov de Monte Carlo (MCMC), pero con la introducción de las Aproximaciones Anidadas Integradas de Laplace (INLA) el muestreo MCMC se vuelve redundante a medida que las distribuciones marginales posteriores se aproximan con precisión de forma totalmente automática. Además, INLA permite el cálculo de muchos modelos lineales generalizados Bayesianos en un período de tiempo razonable, lo que genera una mayor comparación de diferentes modelos y distribuciones previas (Roos y Held, 2011). Se compone en primera medida, de un vector de variables observadas  $y = (y_1, \dots, y_n)$ , cuya distribución se encuentra en la familia exponencial por lo regular y además la media  $\mu_i$ , de la observación  $y_i$ , está vinculada al predictor lineal  $\eta_i$ , donde las observaciones serán condicionalmente independientes, y se puede expresar de la siguiente forma,

$$\eta_i = \alpha + \sum_{j=1}^{n_f} f^{(k)}(u_{ji}) + \sum_{j=1}^{n_\beta} \beta_j z_{ji} + \varepsilon_i, \quad i = 1, \dots, n \quad (46)$$

donde  $\alpha$  es el intercepto,  $f^{(k)}$  son funciones definidas en  $n_f$  efectos aleatorios de algun vector de las covariables  $[u_{ki}]_{k=1}^{n_f}$ , los  $\beta_j$ ,  $j = 1, \dots, n_\beta$  representan el efecto de las covariables  $[z_{ji}]_{j=1}^{n_\beta}$ ; finalmente,  $\varepsilon_i$  es un término de error.

El predictor lineal puede incluir términos sobre covariables (efectos fijos) y diferentes tipos de efectos aleatorios. El vector de todos los efectos latentes se denota con  $x$  e incluye el predictor lineal, los coeficientes de las covariables, etc, de la siguiente forma:  $x = (\eta_1, \dots, \eta_n, \alpha, \beta_1, \dots)$ ; la distribución de  $y$  dependerá del vector de los hiperparámetros  $\theta_1$ .

La distribución del vector de efectos latentes  $x$  se supone que se expresa como un GMRF, el cual tendrá una matriz de precisión  $Q(\theta_2)$  y media cero, siendo  $\theta_2$  un vector de parámetros y  $\theta = (\theta_1, \theta_2)$  el vector de todos los parámetros  $\theta$ . Asumiendo que las observaciones son condicionalmente independientes debido al vector de efectos latentes y los parámetros. INLA aprovecha la estructura dispersa y las propiedades de independencia condicional de los GMRF para hacer más rápido los cálculos. La distribución posterior de los efectos latentes  $X$  puede escribirse entonces como,

$$\pi(X, \theta|y) = \frac{\pi(y|X, \theta)\pi(X, \theta)}{\pi(y)} \propto \pi(y|X, \theta)\pi(X, \theta) \quad (47)$$

donde  $\pi(y)$  es la probabilidad marginal del modelo, que es una constante de normalización y  $\pi(y|X, \theta)$  la verosimilitud. Suponiendo que  $y_i, \dots, y_n$  son independientes, dados los efectos latentes, la verosimilitud se puede escribir como,

$$\pi(y|X, \theta) = \prod_{i \in U} \pi(y_i|X_i, \theta) \quad (48)$$

donde  $U$  es el subconjunto de todos los enteros de 1 hasta  $n$  para todos los valores observados,  $y_i$  son todos los valores observados, y en el caso de los valores perdidos INLA puede calcular la distribución predictiva. La distribución conjunta de los efectos aleatorios y los hiperparámetros  $\pi(X, \theta)$  se puede reorganizar como  $\pi(x|\theta)\pi(\theta)$ , siendo  $\pi(\theta)$  la distribución previa del conjunto de parámetros  $\theta$ , (Gómez-Rubio, 2020). Entonces, porque  $X$  es un GMRF, la distribución posterior de los efectos latentes es:

$$\pi(X|\theta) \propto |Q(\theta)|^{1/2} \exp\left(-\frac{1}{2} X^T Q(\theta) X\right) \quad (49)$$

La distribución posterior conjunta de los efectos latentes y los hiperparámetros se puede escribir como:

$$\begin{aligned} \pi(X, \theta|y) &\propto \pi(\theta)|Q(\theta)|^{1/2} \exp\left(-\frac{1}{2} X^T Q(\theta) X\right) \prod_{i \in U} \pi(y_i|X_i, \theta) = \\ &\pi(\theta)|Q(\theta)|^{1/2} \exp\left(-\frac{1}{2} X^T Q(\theta) X \sum_{i \in U} \log(\pi(y_i|X_i, \theta))\right) \end{aligned}$$

INLA estima las distribuciones marginales de los efectos latentes y los parámetros  $\theta$ . Con un parámetro latente  $X_l$  se puede plantear de la siguiente forma:

$$\pi(X_l|y) = \int \pi(X_l, \theta|y) \pi(\theta|y) d\theta \quad (50)$$

de forma similar la distribución posterior marginal para un parámetro  $\theta_k$  es:

$$\pi(\theta_k|y) = \int \pi(\theta|y) d\theta_{-k} \quad (51)$$

donde  $\theta_{-k}$  es un vector de los parámetros  $\theta$  sin el elemento  $\theta_k$ .

La aproximación a la distribución conjunta posterior de  $\theta$  propuesta por Harvard Rue, Martino y Chopin (2009),  $\tilde{\pi}(\theta|y)$ , se puede utilizar para calcular las distribuciones marginales de los efectos latentes y los parámetros  $\theta$  de la siguiente forma:

$$\tilde{\pi}(\theta|y) \propto \frac{\pi(X, \theta, y)}{\tilde{\pi}_G(X|\theta, y)} \Big|_{X=X^*(\theta)} \quad (52)$$

En la ecuación anterior  $\tilde{\pi}_G(X|\theta, y)$  es una aproximación gaussiana a los efectos latentes, y  $X^*(\theta)$  representa la moda de la distribución condicional para un valor dado del vector de parámetros  $\theta$ .

La marginal posterior  $\pi(\theta_k|y)$  se puede aproximar integrando  $\theta_{-k}$  en la aproximación anterior  $\tilde{\pi}(\theta|y)$ . Del mismo modo, una aproximación a la marginal posterior  $\pi(X_i|y)$  usando integración numérica, sin embargo, es importante una aproximación a  $\pi(X_i, \theta|y)$ . Una aproximación Gaussiana es posible, pero Rue, Martino y Chopin (2009) obtienen mejores aproximaciones al recurrir a otros métodos, como la aproximación de Laplace.

### 3.1. Aproximaciones a $\pi(X_i|\theta, y)$

La aproximación a los marginales de los hiperparámetros se puede calcular marginalizando sobre  $\tilde{\pi}(\theta|y)$  para obtener  $\tilde{\pi}(\theta_i|y)$ . La presentación para los marginales de los efectos latentes requiere integrar los hiperparámetros y marginar los efectos latentes, en INLA se utiliza la siguiente información:

$$\pi(X_i|y) \simeq \tilde{\pi}(X_i|\theta^{(k)}, y) \tilde{\pi}(\theta^{(k)}|y) \Delta_k \quad (53)$$

donde  $[\theta^{(k)}]_{k=1}^K$  representa un conjunto de valores de  $\theta$  que se utilizan para la integración numérica y cada uno tiene un peso de integración asociado  $\Delta_k$ . En INLA se obtendrá estos puntos de integración colocando una cuadrícula regular sobre la moda posterior de  $\theta$  o mediante el uso de un diseño compuesto centrado en la moda posterior.

Rue, Martino y Chopin (2009) describen tres aproximaciones diferentes para  $\pi(x_i|\theta, y)$ . El primero es simplemente una aproximación Gaussiana, que estima la media  $\mu(\theta)$  y varianza  $\sigma_i^2(\theta)$ . Esto es computacionalmente económico porque durante la exploración de  $\tilde{\pi}(\theta|y)$  la distribución  $\tilde{\pi}_G(x|\theta, y)$  puede ser fácilmente estimada por marginalización.

Una segunda aproximación puede ser la aproximación de Laplace de tal forma que  $\pi(x_i|\theta, y)$  podría ser:

$$\pi_{LA}(X_i|\theta, y) \propto \frac{\pi(X, \theta, y)}{\tilde{\pi}_{GG}(X_{-i}|X_i, \theta, y)} \Big|_{X_{-i}=X_{-i}^*(X_i, \theta)} \quad (54)$$

donde la distribución  $\tilde{\pi}_{GG}(x_{-i}|x_i, \theta, y)$  representa la aproximación Gaussiana a  $x_i|x_i, \theta, y$  y  $x_i = x_{-i}^*(x_i, \theta)$  que es la moda. Como es más costoso computacionalmente, por lo que se propone una aproximación modificada de la siguiente forma:

$$\pi_{LA}(x_i|x, \theta) \propto N(x_i|\mu_i(\theta), \sigma_i^2(\theta)) \exp(\text{spline}(x_i)) \quad (55)$$

Por tanto, la aproximación de Laplace se basa en el producto de la aproximación Gaussiana y una spline cúbica  $\text{spline}(x_i)$  en  $x_i$ . La spline se estima en valores seleccionados de  $x_i$  y su función es corregir la aproximación Gaussiana.

La tercera aproximación,  $\tilde{\pi}_{SLA}(x_i|\theta, y)$  se denomina aproximación de Laplace simplificada y se basa en una expansión en serie de  $\tilde{\pi}_{LA}(x_i|\theta, y)$  alrededor de  $x_i = \mu_i(\theta)$ . De esta forma, la aproximación gaussiana  $\tilde{\pi}_G(x_i|\theta, y)$  se puede corregir para la ubicación y la asimetría (Saulau et al., 2010).

### 3.2. Procedimiento de Estimación

Todo el procedimiento de estimación en INLA consta de varios pasos. En primer lugar, la moda de  $\tilde{\pi}(\theta|y)$  se obtiene maximizando  $\log(\tilde{\pi}(\theta|y))$  en  $\theta$ ; esto se hace en un método cuasi-Newton<sup>10</sup>. Una vez que la moda posterior se ha obtenido  $\theta^*$ , las diferencias finitas se utilizan para calcular la matriz Hessiana  $H$  en la moda. Se debe tener en cuenta que  $H^{-1}$  podría ser la matriz de varianzas si la posterior es una distribución Gaussiana.

Luego,  $H^{-1}$  se descompone usando los valores propios, tal que  $H^{-1} = V\Lambda V$ . Con esta descomposición, para cada valor del parámetro  $\theta$  se puede obtener un reescalado  $z$  tal que,  $\theta(z) = \theta^* + V\Lambda^{1/2}z$ . Este cambio de escala es útil para explorar el espacio de  $\theta$  más eficientemente, ya que corrige la escala y la rotación.

Finalmente,  $\log(\tilde{\pi}(\theta|y))$  es alcanzado usando la  $z$  parametrización en dos caminos; el primero, usa una grilla regular de pasos  $h$  centrada en la moda ( $z = 0$ ) y los puntos de la grilla son considerados solamente si,

$$|\log(\tilde{\pi}(\theta(0)|y)) - \log(\tilde{\pi}(\theta(z)|y))| < \delta$$

donde  $\delta$  como umbral, la exploración se realiza primero a lo largo del eje  $z$  parametro y luego se suman todos los puntos intermedios que cumplen la condición anterior.

Esto proporciona un conjunto de configuraciones de los parámetros  $\theta$  sobre la moda posterior que se pueden utilizar en los procedimientos de integración numérica requeridos en INLA. Esto a menudo se conoce como estrategia de cuadrícula.

Alternativamente, un diseño compuesto centrado en  $\theta(0)$  se puede usar para obtener algunos puntos estratégicamente ubicados, en lugar de usar puntos en una cuadrícula regular. Esto puede ser más eficiente que la estrategia de cuadrícula a medida que aumenta la dimensión del espacio del parámetro  $\theta$ .

Para la estrategia de cuadrícula primero se explora la densidad logarítmica a lo largo de los ejes. Luego se exploran todas las combinaciones de sus valores y solo aquellas con una diferencia hasta el umbral  $\delta$ , con la densidad logarítmica en la moda, se consideran en la estrategia de integración.

Una opción más rápida que puede funcionar con un número moderadamente grande de parámetros  $\theta$ , es un enfoque empírico de Bayes que conecta la moda posterior de  $\theta|y$ . Esto sirve cuando la variabilidad en los parámetros  $\theta$  no afecta la parte posterior de los efectos latentes  $x$ . Este tipo de estimadores pueden ser útiles para ajustar modelos altamente parametrizados y pueden proporcionar buenas aproximaciones en varios escenarios.

### 3.3. Datos de Área

Los datos de área son una de las ramas de la estadística espacial, en general se usan cuando existe un lugar determinado como departamentos, municipios, comunas o barrios, la información disponible esta a nivel agregada para ese tipo de subdivisión.

Para este trabajo se incluye el modelo de Besag-York-Mollié (1991) conocido como BYM, el cual tiene en cuenta que los datos pueden estar correlacionados espacialmente y las observaciones en áreas vecinas pueden tener características más parecidas que las observaciones en áreas alejadas; incluye un efecto espacial aleatorio que suaviza los datos de acuerdo con una estructura de vecindario y un componente intercambiable no estructurado que modela el ruido no correlacionado. En entornos espacio-temporales se utilizan modelos que tienen en cuenta, no solo la estructura espacial, sino también las correlaciones temporales y las interacciones espacio-temporales.

<sup>10</sup>Los métodos cuasi-Newton son métodos que se utilizan para encontrar ceros o máximos y mínimos locales de funciones, como una alternativa al método de Newton. Se pueden usar si el jacobiano o el hessiano no están disponibles o son demasiado costosos de calcular en cada iteración (Nocedal, J., et. al., 1999)



Se pueden suponer las áreas  $i$  y  $j$  como vecinas si comparten un borde común denotado aquí como  $i \sim j$ . Esto parece razonable si las regiones tienen el mismo tamaño y están organizadas en un patrón regular. Además, se denota el conjunto de vecinos de la región  $i$  por  $\delta_i$  y su tamaño por  $n\delta_i$ . Para modelar la correlación espacial se utilizan varios modelos, entre ellos, un modelo intrínseco de GMRF, denominado Besag. La distribución condicional para  $b_i$  es:

$$b_i | b_{-i}, \tau_b \sim N \left( \frac{1}{\eta \delta_i} \sum_{j \in \delta_i} b_j, \frac{1}{\eta \delta_i \tau_b} \right) \quad (56)$$

donde  $\tau_b$  es un parámetro de precisión y  $b_{-i} = (b_1, \dots, b_{1-i}, b_{1+i}, \dots, b_n)^T$ . La media de  $b_i$  es igual a la media de los efectos sobre todos los vecinos y es proporcional al número de vecinos. La distribución conjunta de  $b_i$  es:

$$\pi(b | \tau_b) \propto \exp \left( -\frac{\tau_b}{2} \sum_{i \sim j} (b_i - b_j)^2 \right) \propto \exp \left( -\frac{\tau_b}{2} b^T Q b \right), \quad (57)$$

donde la matriz de precisión  $Q$  tiene las siguientes entradas:

$$Q_{ij} = \begin{cases} \eta \delta_i, & i = j, \\ -1, & i \sim j, \\ 0, & \text{si no} \end{cases} \quad (58)$$

El modelo es intrínseco en el sentido de que  $Q$  es singular. Para cualquier mapa 1, que denota un vector de longitud  $n$  con solo unos, es un vector propio de  $Q$  con valor propio 0; por tanto, la densidad es invariante a la adición de una constante (Riebler et al., 2016), si el mapa espacial tiene islas la definición del espacio nulo es más compleja y depende de como se definan las islas, es decir, si forman un nuevo gráfico no conectado o si están vinculadas al campo principal. El rango de deficiencia del modelo Besag es igual al número de subgrafos conectados es igual a 1 si todas las regiones están conectadas, por tanto la densidad para el modelo Besag es:

$$\pi(b | \tau_b) = K \tau_b^{(n-S)/2} \exp \left( -\frac{\tau_b}{2} b^T Q b \right), \quad (59)$$

donde  $S$  denota el número de subgrafos conectados y  $K$  es una constante. Para evitar confusiones con la intersección, se imponen restricciones de suma igual a cero en cada subgrafo conectado.

El modelo Besag solo asume un componente estructurado espacialmente y no permite una variabilidad estructurada espacialmente. Por lo tanto, el error aleatorio no estructurado o la sobredispersión pura dentro del área  $i$ , se modelará como una correlación espacial, lo que puede proporcionar parámetros estimados engañosos. Este problema se intenta compensar con el modelo BYM, el cual descompone el efecto espacial regional  $b$  en una suma de un componente espacial estructurado y no estructurado, de modo que  $b = v + u$ ; donde  $v \sim N(0, \tau_u^{-1} Q^-)$  explica la sobredispersión pura mientras que  $u \sim N(0, \tau_u^{-1} Q^-)$  es el modelo Besag, donde  $Q^-$  denota la inversa de  $Q$ , y la matriz de covarianzas de  $b$  sería:

$$\text{Var}(b | \tau_u, \tau_v) = \tau_v^{-1} S + \tau_u^{-1} Q^- \quad (60)$$

### Caminata Aleatoria de Orden 1(RW1<sup>11</sup>)

La Caminata aleatoria de orden 1 (RW1) para el vector Gaussiano  $m = (m_1, \dots, m_n)$ , esta construida asumiendo incrementos independientes:

$$\Delta m_i = m_i - m_{i-1} \sim N(0, \tau^{-1})$$

La densidad para  $x$  es derivada de  $n - 1$  incrementos:

$$\begin{aligned} \pi(m|\tau) &\propto \tau^{(n-1)/2} \exp\left(-\frac{\tau}{2} \sum (\Delta x_i)^2\right) \\ &= \tau^{(n-1)/2} \exp\left(-\frac{1}{2} x^T Q x\right) \end{aligned}$$

donde  $Q = \tau R$  y  $R$  es la matriz que refleja la estructura de vecindad del modelo. También es posible definir una versión cíclica del modelo RW1 donde se modifica el gráfico de modo que el último nodo  $m_n$  es vecino de  $m_{n-1}$  y  $m_1$ . El parámetro de precisión  $\tau$  es representado como:

$$\theta = \log(\tau)$$

y la distribución previa esta definida en  $\theta$

### 3.4. Tendencia Dinámica Paramétrica

Como se indica en el modelo BYM el efecto espacial se descompone en la suma de un componente espacial estructurado y no estructurado  $b = (v + u)$ ; para este caso  $v$  explica la sobredispersión en un barrio determinado y  $u$  el componente estructurado (Bernardinelli et al.,1995; Blangiardo y Cameletti, 2015); representando el efecto del barrio y del barrio vecino de primer orden sobre la variable a analizar.

En el caso de las estimación temporal se puede observar si entre 2015 y 2019 han ocurrido cambios en la Tasa de Ocupación, Cantidad de Alojamientos e Ingresos por barrio respecto a años anteriores. El predictor lineal se presenta de la siguiente forma:

$$\eta_{it} = b_0 + \mu_i + v_i + (\psi + \delta_i)t \quad (61)$$

donde,  $b_0$  representa el intercepto,  $\psi$  representa el efecto global del tiempo,  $t = 1, \dots, T$  representan los puntos de tiempo (expresado en meses, trimestres o años, etc.), y  $\delta_i$  la interacción entre el tiempo y el espacio. Dado que para propósitos de identificabilidad se impone una restricción de suma a cero en  $\delta = \delta_1, \dots, \delta_n$ , restricción que es similar en  $u$ ,  $\delta_i$  representa la diferencia entre la tendencia global  $\psi$  y el área -tendencia específica. 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. Se supone que  $\delta_i \sim N(0, 1/\tau_\delta)$ , pero se pueden usar otras especificaciones como una estructura autorregresiva condicional (Bernardinelli et al.,1995; Schrödle y Held,2011; Blangiardo y Cameletti, 2015); la media posterior del efecto espacial se obtiene de  $b_i = \exp(u_i + v_i)$ .

---

<sup>11</sup>Por sus siglas en inglés, Random Walk

### 3.5. Tendencia Dinámica No Paramétrica

En el modelo especificado anteriormente se impone una restricción de linealidad sobre la tendencia temporal diferencial  $\delta_i$ , sin embargo, es posible liberarlo usando una formulación dinámica no paramétrica para el predictor lineal de la siguiente forma:

$$\eta_{t,i} = b_0 + u_i + v_i + \gamma_t + \phi_t \quad (62)$$

Parámetros que son iguales a la ecuación paramétrica, con la diferencia de que  $\gamma_t$  representa el efecto temporalmente estructurado, modelado dinamicamente  $\gamma_t | \gamma_{t-1}, \gamma_{t-2} \sim N(\gamma_{t-1}, \sigma^2)$  para una caminata aleatoria de orden 1 (RW1); y el  $\phi_t$  es una prior intercambiable gaussiana:  $\phi_t \sim N(0, 1/\tau_\phi)$  (Knorr-Held, L., 2000).

#### Interacción Espacio Temporal

La interacción espacio temporal con la formulación dinámica no paramétrica se presenta de la siguiente forma:

$$\eta_{i,t} = b_0 + u_i + v_i + \gamma_t + \phi_t + \delta_{it} \quad (63)$$

En este caso el vector de parámetros  $\delta_{it}$  sigue una distribución Gaussiana con una matriz de precisión dada por  $\tau_\delta R_\delta$  donde  $\tau_\delta$  es un escalar desconocido, mientras que  $R_\delta$  es la matriz estructurada que identifica el tipo de dependencia espacial y/o temporal entre los elementos de  $\delta$ ; dicha matriz se puede factorizar como el producto kronecker de la matriz estructura de los principales efectos que interactúan. Hay cuatro formas de definir la estructura, y el orden de interacción es importante (Knorr-Held, 2000).

El tipo de interacción I asume que los efectos no estructurados  $v_i$  y  $\phi_i$  interactúan, donde la matriz estructura se escribe:  $R_\delta = R_v \otimes R_\phi = I \otimes I = I$ , porque tanto  $v_i$  como  $\phi_i$  no tienen estructura espacial ni temporal, en consecuencia tampoco la asumen en la interacción, y por lo tanto:  $\delta_{it} \sim N(0, 1/\tau_\delta)$ . La matriz  $R_\delta$  tiene un rango  $n(T)$  para una caminata aleatoria de primer orden, donde  $T$ , son los puntos de tiempo.

El tipo II combina el efecto principal temporal estructurado  $\gamma_t$  y el efecto espacial no estructurado  $v_i$ ; la matriz estructura en este caso sería:  $R_\delta = R_v \otimes R_\gamma$ , donde  $R_v = I$  y  $R_\gamma$  es la estructura de vecindad especificada, a través de una caminata aleatoria de primer o segundo orden; esto supone que para la  $i$ -ésima área el vector de parámetros  $\delta_{it}$  tienen una estructura autorregresiva en el componente del tiempo, en la cual hay independencia entre las áreas. La matriz  $R_\delta$  tiene un rango  $n(T-1)$  para una caminata aleatoria de primer orden.

El tipo III combina el efecto temporal  $\phi_t$  no estructurado y el efecto principal estructurado espacialmente  $u_i$ , la matriz de estructura sería:  $R_\delta = R_\phi \otimes R_u$  donde  $R_\phi = I$  y  $R_u$  es una estructura vecina definida a través de la especificación CAR. Esto lleva a suponer que los parámetros del  $n$ -ésimo momento  $[\delta_1, \dots, \delta_n]$  tienen una estructura espacial independiente de los otros momentos, y la matriz  $R_\delta$  tiene un rango de  $T(n-1)$ , (Blangiardo y Cameletti, 2015)

El tipo IV es el tipo de interacción más complejo, asumiendo que los efectos estructurados espacial y temporalmente  $u_i$  y  $\delta_i$  interactúan. La matriz de estructura se puede escribir como el producto de Kronecker de  $R_\delta = R_\delta \otimes R_\gamma$  y tienen un rango de  $(T-1)(n-1)$  para una caminata aleatoria de orden 1.

### 3.6. Programa Estadístico R y paquete INLA

La implementación de los algoritmos se realiza con un paquete del programa estadístico R llamado INLA, que se encuentra en un repositorio específico<sup>12</sup>, y está disponible para Windows, Mac OS X y Linux. La función principal del paquete es la que se utiliza para ajustar los modelos bayesianos utilizando la metodología INLA. Se utiliza una fórmula para especificar el modelo que se va a ajustar y puede incluir una combinación de efectos fijos y otros efectos convenientemente especificados, el tipo de efecto, los parámetros adicionales y los antecedentes de los hiperparámetros del efecto (Gómez-Rubio, 2020).

### 3.7. Metodología del trabajo

La metodología para la presente investigación incluye varios pasos entre los que se encuentra el análisis, depuración y tratamiento de datos y la aplicación del algoritmo INLA en los respectivos modelos de la siguiente forma:

En la primera parte se realizó la revisión de las fuentes secundarias y documentos propios del marco teórico. Luego de esto, se implementó el análisis y depuración de las bases de datos para las variables tanto para las 3 variables dependientes como independientes, 7 variables, y se eligió la información de los barrios ubicados en el sector catastral urbano de la ciudad, 988 barrios.

En segundo lugar se realizó un ajuste de los datos en cero que representaban el 40.2 % de las variables, tanto dependientes como independientes, dado que en un sector de la ciudad no existen establecimientos de alojamiento Airbnb o HomeAway; estos datos se cambiaron por los valores mínimos de cada variable lo que hizo factible la utilización de los modelos Beta, Poisson y Gamma sin la necesidad de implementar modelos cero inflados que pueden realizarse en trabajos futuros.

Luego de esto se realizaron pruebas de correlación espacial con la metodología de Moran, para constatar que las variables dependientes y los datos de área eran aptas para realizar los modelos espacio-temporales. Se modificaron las bases, generando variables índice, por cada variable dependiente y tipo metodología que se esperaba implementar en INLA; para este caso, modelos espacio-temporales con interacción dinámica de tipo II.

Finalmente se aplicaron los 3 modelos Bayesianos, primero con 4 variables explicativas y luego con 7 variables, seleccionando los modelos con menor Criterio de Información de Desviación (DIC). Los resultados se presentan en tablas, gráficas y mapas que describen características de las variables en la dinámica espacio-temporal, realizando un análisis comparativo por modelo.

## 4. Resultados prácticos

En esta sección se analizarán 3 modelos de regresión espacio-temporales Bayesianos (modelos Beta, Poisson y Gamma) donde las 3 variables dependientes son: Tasa de Ocupación promedio, Cantidad de establecimientos o propiedades e Ingresos Totales por barrio, para caracterizar el mercado de hospedaje tipo Airbnb en Bogotá. Esta estimación se hace para 988 barrios de la ciudad, utilizando solamente los ubicados en el área urbana del sector catastral de la ciudad entre los años 2015 a 2019. Los datos son tomados del repositorio del portal Datos Abiertos del Instituto Distrital de Turismo de Bogotá<sup>13</sup>, los cuales se ajustaron para ser agregados por año cambiando los datos en cero por los valores mínimos de cada variable que como consecuencia conllevan a un aumento en el error de las estimaciones pero al mismo tiempo permiten la utilización de los MLG de INLA<sup>14</sup> (Instituto Distrital de Turismo, 2019).

<sup>12</sup><http://www.r-inla.org>

<sup>13</sup>ver en:[https://datosabiertos.bogota.gov.co/dataset/barrios\\_aairbnb-bogota-2015-2019-mensual](https://datosabiertos.bogota.gov.co/dataset/barrios_aairbnb-bogota-2015-2019-mensual)

<sup>14</sup>En INLA existen los modelos cero inflados, pero para este trabajo no son incluidos debido a que se busca analizar el potencial de las regresiones clásicas del algoritmo.

Los resultados de los modelos evidencian que el ajuste es positivo dado que las diferencias espaciales y temporales permanecen estables.

#### 4.1. Análisis de Correlaciones Espaciales

El paso anterior a la realización de los modelos consiste saber si las variables de interés tienen autocorrelación espacial a través de la Prueba de Moran y de esta forma determinar si es viable implementar modelos espacio temporales<sup>15</sup>. La prueba de hipótesis es la siguiente:

$H_0$ : No existe autocorrelación espacial.

$H_1$ : Existe autocorrelación espacial.

De esta forma el prueba tiene dos modalidades, una en la que se puede realizar un análisis global de las variables en el tiempo y otra en la que se realizar el análisis por unidad espacial, de la siguiente forma:

Tabla 1: Prueba de Moran Global para Tasa de Ocupación

	Años	Estadístico	P.Valor	I.Moran	Esperanza	Varianza
1	2015	24.938	0.000	0.496	-0.001	0.000
2	2016	20.585	0.000	0.410	-0.001	0.000
3	2017	20.651	0.000	0.412	-0.001	0.000
4	2018	20.531	0.000	0.409	-0.001	0.000
5	2019	22.518	0.000	0.449	-0.001	0.000

Tabla 2: Prueba de Moran Global para Cantidad de Alojamientos

	Años	Estadístico	P.Valor	I.Moran	Esperanza	Varianza
1	2015	24.439	0.000	0.471	-0.001	0.000
2	2016	26.934	0.000	0.524	-0.001	0.000
3	2017	27.773	0.000	0.538	-0.001	0.000
4	2018	26.973	0.000	0.521	-0.001	0.000
5	2019	27.325	0.000	0.527	-0.001	0.000

Tabla 3: Prueba de Moran Global para Ingreso de Alojamientos

	Años	Estadístico	P.Valor	I.Moran	Esperanza	Varianza
1	2015	25.759	0.000	0.497	-0.001	0.000
2	2016	24.611	0.000	0.463	-0.001	0.000
3	2017	25.461	0.000	0.470	-0.001	0.000
4	2018	22.613	0.000	0.405	-0.001	0.000
5	2019	4.522	0.000	0.031	-0.001	0.000

En las tablas 1,2 y 3, se observan los resultados del Prueba de Moran Global, mostrando que existe evidencia estadísticamente significativa para rechazar la hipótesis nula, dado que el p-Valor es menor al 5% en las 3 tablas, por esto se puede decir que para las 3 variables dependientes que se van a analizar existe dependencia espacial global.

En el caso del análisis local, se observa de manera inicial que en algunos sectores de la ciudad habría una alta correlación entre las variables y los barrios donde están los establecimientos, pero que también

<sup>15</sup>En ningún momento se plantea que la prueba se puede utilizar en una metodología Bayesiana, solo se utiliza para hacer un reconocimiento inicial sobre la existencia de autocorrelación.

habría un efecto contagio con sus vecinos inmediatos, especialmente en algunos barrios ubicados en el centro y nororiente de Bogotá, como lo muestra más adelante la figura 1.

Esto resulta ser significativo dado que también la hotelería tradicional se ubica en estos sectores, es decir, que habría una competencia directa entre los alojamientos clásicos y alternativos, pero también podría indicar que la ubicación de los mismos está relacionado con los atractivos turísticos y las prácticas o actividades de los turistas en la ciudad, que pueden estar vinculadas con actividades de recreación o de negocios. Los barrios ubicados en las localidades de Chapinero, Santafé y Candelaria mostraría una mayor correlación o dependencia espacial.

En las localidades ubicadas en la zona sur de la ciudad también se observa autocorrelación entre barrios, posiblemente relacionado con números similares entre las tasas de ocupación, cantidades de alojamientos e ingresos, pero con valores inferiores a los mostrados en el centro-orienté de Bogotá.

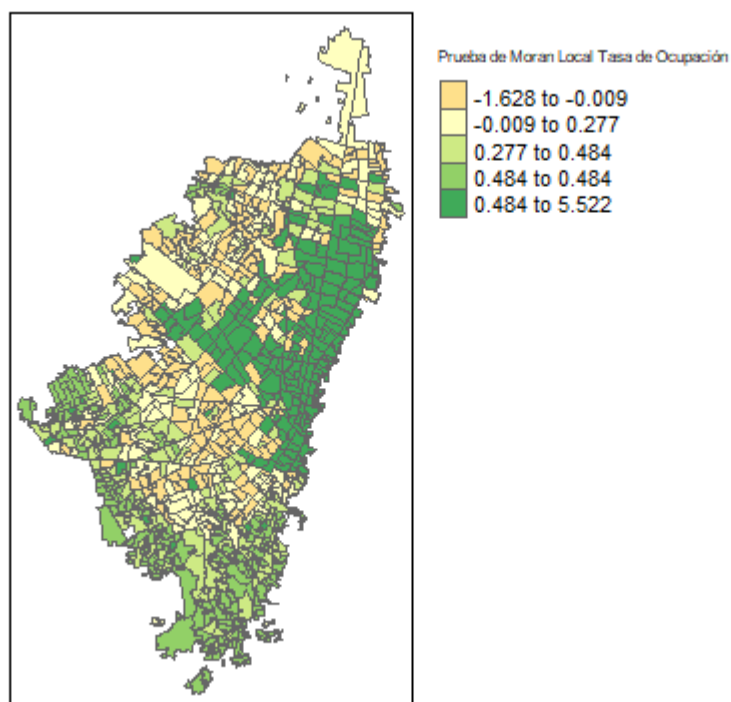


Figura 1: Prueba de Moran Local para Tasa de Ocupación

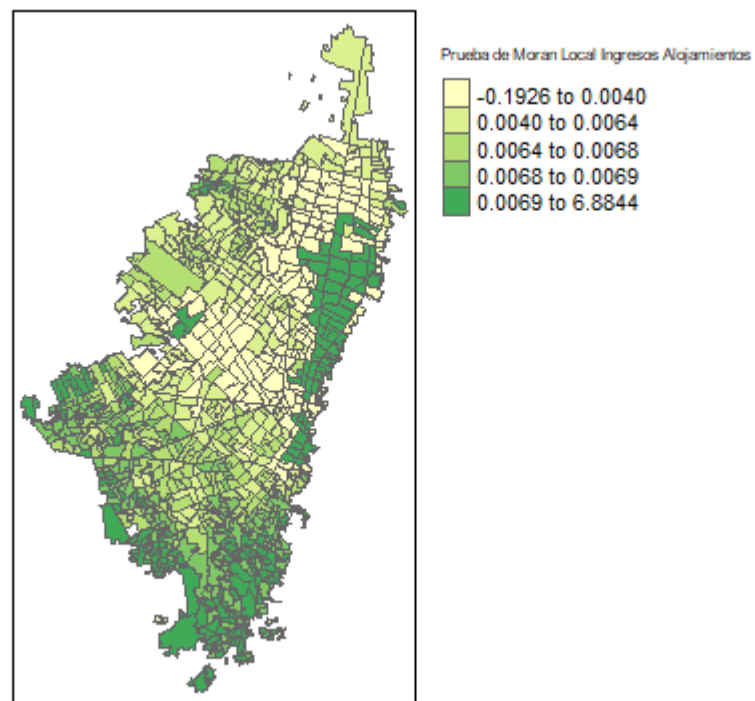


Figura 2: Prueba de Moran Local para Ingresos Totales

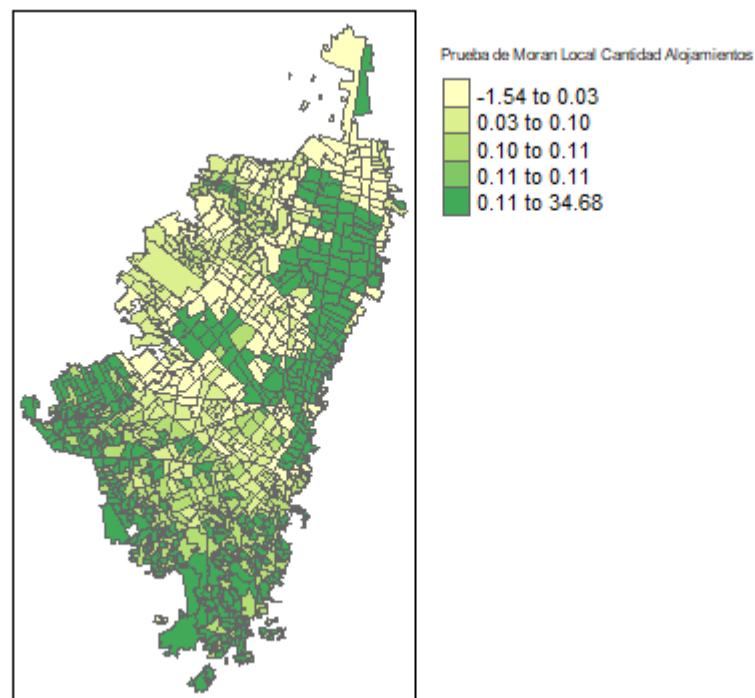


Figura 3: Prueba de Moran Local para Cantidad de Alojamientos

## 4.2. Regresión Beta Espacio Temporal Bayesiana en INLA

En esta parte se analiza la tasa de ocupación promedio anual en los establecimientos Airbnb por barrio, utilizando una Regresión Beta, que como ya se mencionó, se acopla de mejor forma a los datos numéricos continuos en escala de 0 a 1, incluyendo a los mismos. Para este modelo se incluyeron las variables independientes:

1. Vigencia: Año evaluado.
2. Deposito USD: Promedio de depósito que pagan huéspedes en dólares por barrio.
3. Limpieza USDP: Promedio de cobro de limpieza por el uso del establecimiento en dólares por barrio.
4. Media Arriendo: Promedio mensual de los arriendos que se cobran en sitios residenciales aledaños a los establecimientos por barrio.
5. Max HuespedesP: Promedio del número máximo de huéspedes que se admiten en los establecimientos por barrio.
6. ADR USDP: Promedio del cobro en establecimientos por barrio.
7. N Aloj: Tipo de Alojamiento compuesto de la siguiente forma:
  - N Aloj1: Apartamentos
  - N Aloj2: Casa
  - N Aloj3: Habitación
  - N Aloj4: Hotel
  - N Aloj5: Lugar Familiar
  - N Aloj6: No Convencional
  - N Aloj7: Propiedad Compartida

### Modelo

La estructura del modelo de regresión Beta se plantea de la siguiente forma:

$$Y_{it} \sim Be(\mu, \phi); \quad (64)$$

$$\eta_{it} = b_0 + \beta_1 Vigencia_{1i} + \beta_2 Deposito_{2i} + \beta_3 Limpieza_{3i} + \beta_4 MediaArriendo_{4i} + \beta_5 MaxHuespedes_{5i} + \beta_6 ADR_{6i} + \beta_7 NAloj_{7i} + u_i + v_i + Temporal_t \quad (65)$$

donde,  $t = 1, \dots, T$  en  $Temporal_t$ , con una estructura que puede ser paramétrica o no paramétrica.



### Valores del hiperparámetro elegido

El hiperparámetro para este modelo es el parámetro de precisión  $\phi$ , que también se puede representar como,

$$\phi = \exp(\theta), \quad (66)$$

donde la distribución previa esta definida en  $\theta$ , la cual es una distribución logGamma( $a, b$ ), con  $a$  como el parámetro de forma y  $b$  el parámetro de escala inverso y se asignan los valores  $a = 1$  y  $b = 0.1$ .

### Modelo Espacio Temporal Paramétrico

Para este caso se expresa en terminos de la razón:  $b = u_i + v_i$  por lo que estimaciones positivas indicaran que estar ubicado en un determinado barrio de Bogotá generará una mayor tasa de ocupación.

Los resultados evidencian que existe un efecto espacial más alto en los barrios ubicados en las localidades de: Usaquén, Chapinero, Santafé y Candelaria, y lo mismo ocurre en el efecto diferencial temporal, donde los cambios más altos en la tasa de ocupación ocurren en los barrios ubicados en la zona centro oriental de la ciudad; sin embargo, también se observan barrios ubicados en el corredor de la calle 26, en Teusaquillo y en la zona norte donde el efecto temporal es más alto que la media global(ver Figura 4 y 5).

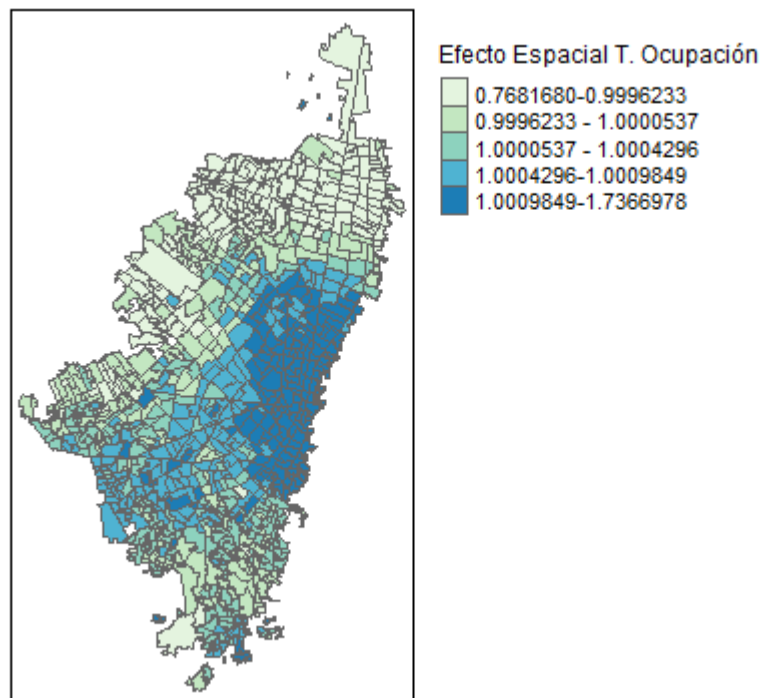


Figura 4: Estimación del Efecto Espacial para el Logaritmo de la Tasa de Ocupación por Barrio 2015 a 2019

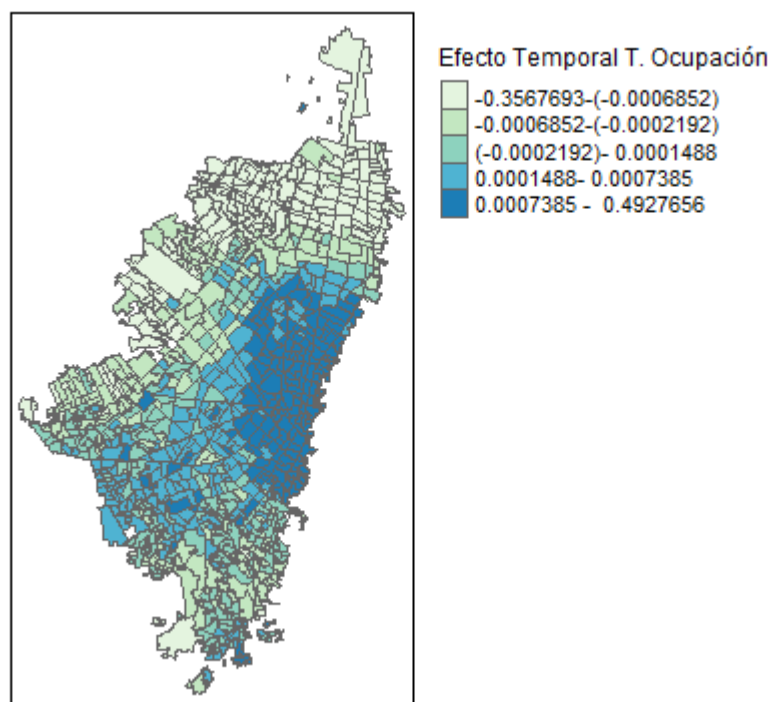


Figura 5: Estimación del Efecto Espacial y Temporal para el Logaritmo de la Tasa de Ocupación por Barrio 2015 a 2019

El efecto de tendencia lineal o el efecto fijo de las covariables se muestra en las Tablas 4 y 5 que se afecta por el cambio temporal (2015 a 2019). Se presentan dos modelos para escoger por el criterio de información de desviación (DIC) si es conveniente la inclusión de otras variables; esto se observa en la tabla 6 donde el DIC del Modelo 1 que incluyó: Vigencia, Depósito en Dólares, Limpieza en Dólares y Media del arriendo de residencias no turísticas; tiene un mayor valor que el Modelo 2 el cual incluye también las variables independientes: valor promedio del costo por alojamiento y los tipos de alojamiento como variable categórica; por lo que se escoge para analizar el Modelo 2.

En dicho Modelo se observa que las variables: cobro de depósito en Dólares, cobro de limpieza en dólares, y los tipos de alojamientos: Casa, Habitación, Hotel, No convencional y propiedad compartida, comparado con el tipo Apartamento, no son significativas. Es decir, que en estas variables no hay evidencia de interacción (el intervalo de credibilidad del 95 % para el parámetro de interacción no incluye al cero).

El intercepto y el promedio del costo del alojamiento por barrio, tienen correlación negativa, es decir que a mayor costo menores serán las tasas de ocupación; mientras que variables como: Máximo de huéspedes, Media del arrendamiento residencial en los barrios y el tipo de establecimiento: Lugar Familiar, evidencian una correlación positiva, lo que significa que la ubicación y el tipo de acomodación más amplio de establecimiento pueden correlacionarse con un aumento de tasa de ocupación.

Por otro lado, los valores de  $kld^{16}$  para todos los parámetros del modelo son cero, lo que indica que la aproximación de Laplace simplificada es apropiada para utilizarse en este caso.

<sup>16</sup>La divergencia de Kullback-Leibler ( $kld$ ) en la inferencia bayesiana, se puede utilizar como una medida de la ganancia de información al pasar de una distribución previa,  $p(\theta)$ , a una distribución posterior,  $p(\theta|y)$ ; cuando la  $kld$ , es cero significa que no hubo pérdida de información y se estarían ajustando las distribuciones en el modelo (Kullback y Leibler, 1951). Para este caso describe la diferencia entre las aproximaciones gaussianas y simplificadas o completas de Laplace para cada distribución posterior.

Tabla 4: Estimación Efecto Fijos para la Tasa de Ocupación promedio de Alojamientos Airbnb y HomeAway por Barrio 2015 a 2019 (Modelo 1)

	Media	Desviación	Percentil 2.5	Percentil 50	Percentil 97.5	Moda	kld
(Intercept)	-93.820	13.750	-120.800	-93.820	-66.840	-93.830	0.000
Vigencia	0.050	0.010	0.030	0.050	0.060	0.050	0.000
Deposito_USD	0.000	0.000	-0.000	0.000	0.000	0.00	0.000
Limpieza_USDP	0.010	0.000	0.000	0.010	0.020	0.010	0.000
Media_Arriendo	0.010	0.000	0.010	0.010	0.010	0.010	0.000

Tabla 5: Estimación Efecto Fijos para la Tasa de Ocupación promedio de Alojamientos Airbnb y HomeAway por Barrio 2015 a 2019 (Modelo 2)

	Media	Desviación	Percentil 2.5	Percentil 50	Percentil 97.5	Moda	kld
(Intercept)	-86.685	13.032	-112.230	-86.7011	-61.079	-86.730	0.000
Vigencia	0.041	0.006	0.028	0.041	0.054	0.041	0.000
Deposito_USD	0.000	0.000	-0.000	0.000	0.000	0.000	0.000
Limpieza_USDP	0.007	0.004	-0.000	0.007	0.015	0.007	0.000
Media_Arriendo	0.006	0.000	0.005	0.006	0.006	0.006	0.000
Max_HuespedesP	0.154	0.019	0.116	0.154	0.192	0.154	0.000
ADR_USDP	-0.004	0.001	-0.008	-0.004	-0.001	-0.004	0.000
N_Aloj2	0.071	0.047	-0.021	0.0714	0.163	0.071	0.000
N_Aloj3	-0.013	0.103	-0.217	-0.012	0.189	-0.012	0.000
N_Aloj4	-0.253	0.290	-0.836	-0.249	0.306	-0.242	0.000
N_Aloj5	1.181	0.489	0.221	1.181	2.143	1.180	0.000
N_Aloj6	-0.506	0.394	-1.285	-0.504	0.264	-0.501	0.000
N_Aloj7	0.114	0.125	-0.133	0.1148	0.358	0.116	0.000

Tabla 6: Criterio de Información de Desviación (DIC) para Modelos de Regresión Beta

DIC Modelo 1	DIC Modelo 2
-12513.06	-14163.09

La figura 6 muestra las gráficas de las densidades de las variables entre los percentiles 2.5 y 97.5 lo cual ayudaría a entender mejor cuales son significativas en el modelo y cuales no lo son:

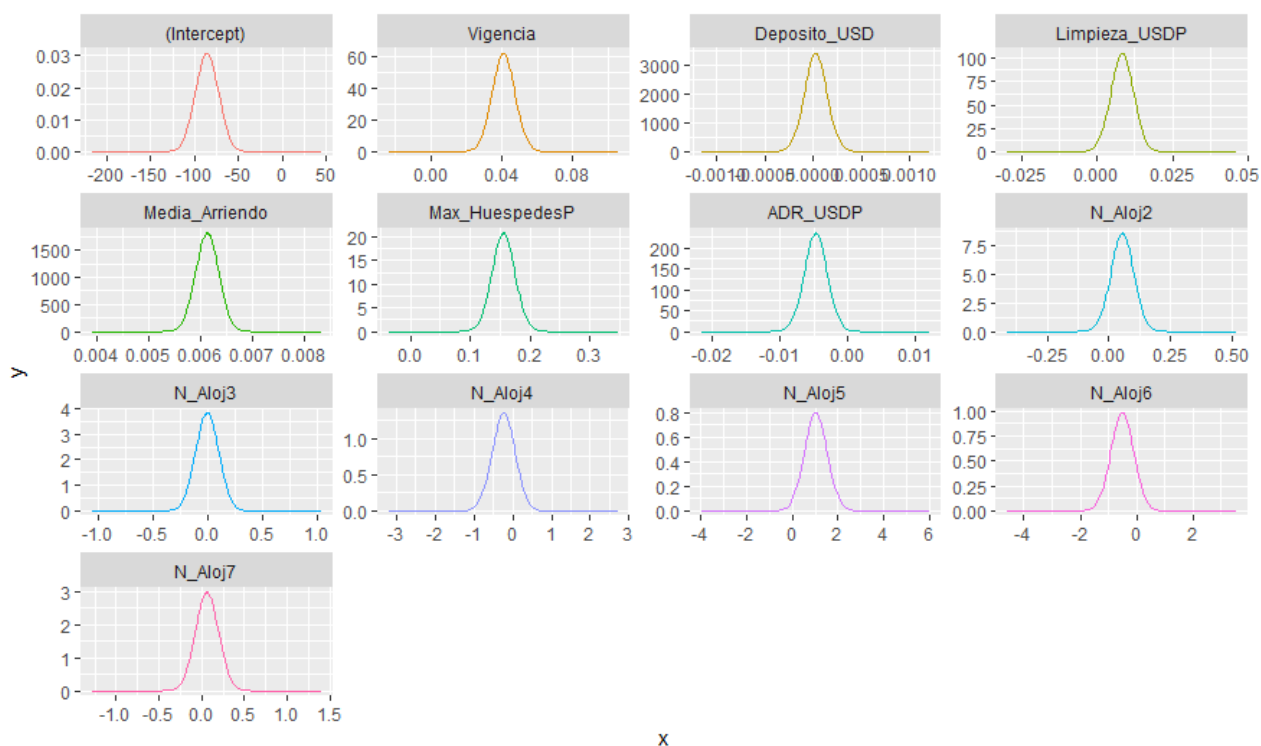


Figura 6: Densidades Modelo Beta Espacio Temporal para Tasa de Ocupación Airbnb y HomeAway por Barrio 2015 a 2019

De esta forma, las covariables Deposito y Limpieza en Dólares no serían significativas para explicar la tasa de ocupación, lo mismo ocurrirá con los tipos de establecimientos: Casa, Habitación, Hotel, No Convencional y Propiedad Compartida, los cuales, al compararse con la categoría Apartamentos, no serían significativas para estimar la ocupación en alojamientos Airbnb. Se observa el comportamiento de la media y los intervalos de credibilidad para las covariables entre el percentil 2.5 y el 97.5, mostrando cuales incluyen al cero las cuales no serían significativa para el modelo.

Las medias posteriores y los intervalos de credibilidad del 95 % de los modelos de caminatas autorregresivas y aleatorias ajustados a los datos de la tasa de Ocupación, muestran para las covariables significativas que hay un mayor grado de ajuste del modelo a los datos, y en el caso del año o Vigencia estimaciones suaves, lo que podría evidenciar la correlación temporal con una caminata aleatoria de orden 1.

### Tendencia Dinámica de Interacción Espacio Temporal No Paramétrica

La interacción espacio temporal es de tipo II, que combina el efecto temporal estructurado con el efecto espacial no estructurado, y por lo tanto, muestra una estructura autorregresiva en el componente del tiempo, en la cual hay independencia espacial entre los barrios al analizar la tasa de ocupación de los alojamientos Airbnb entre los años 2015 a 2019.

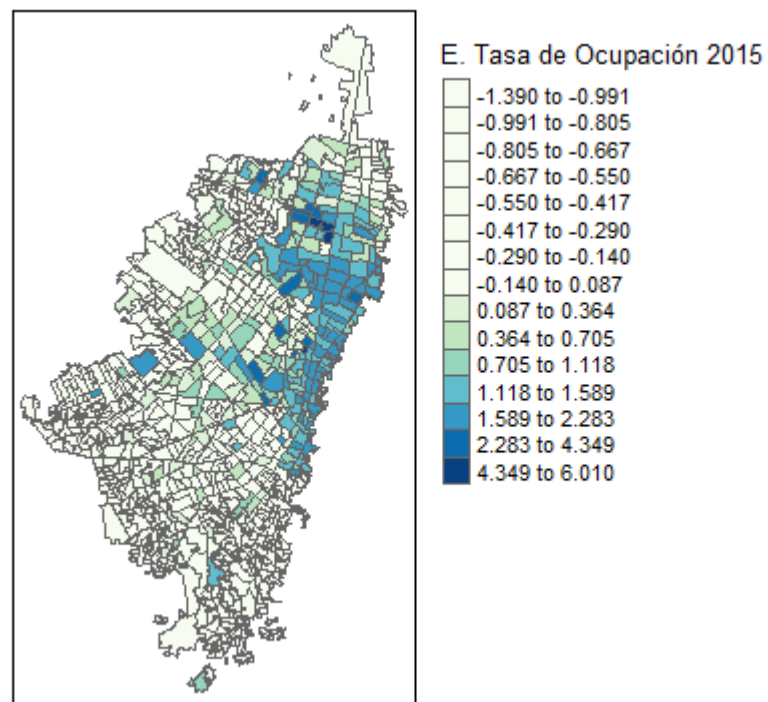


Figura 7: Estimación Interacción Espacio Temporal para el Logaritmo de la Tasa de Ocupación Airbnb y HomeAway por Barrio 2015

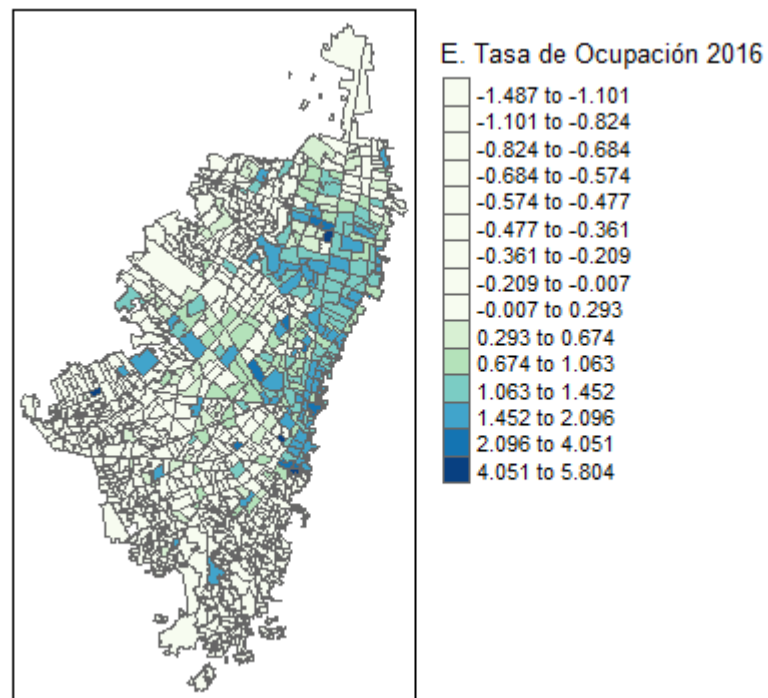


Figura 8: Estimación Interacción Espacio Temporal para el Logaritmo de la Tasa de Ocupación Airbnb y HomeAway por Barrio 2016

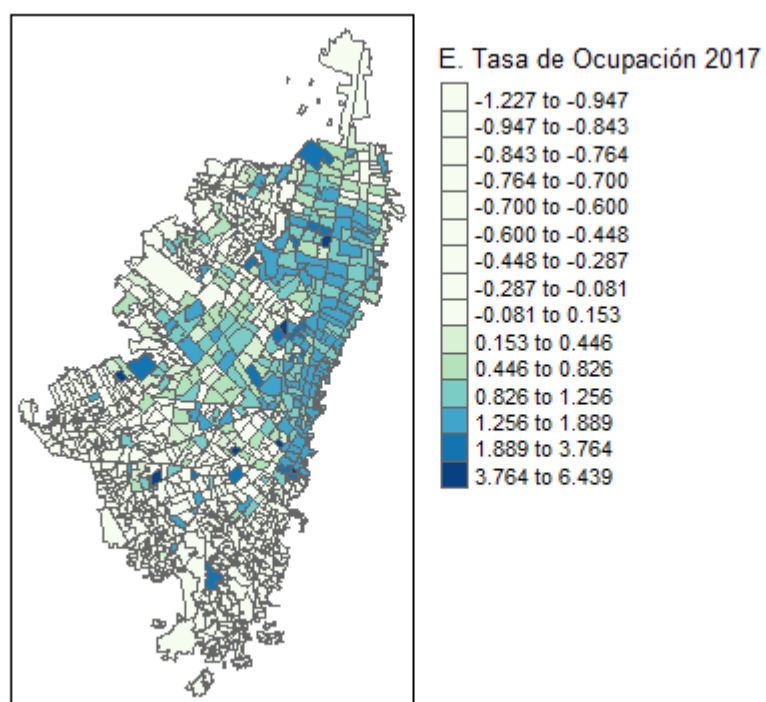


Figura 9: Estimación Interacción Espacio Temporal para el Logaritmo de la Tasa de Ocupación Airbnb y HomeAway por Barrio 2017

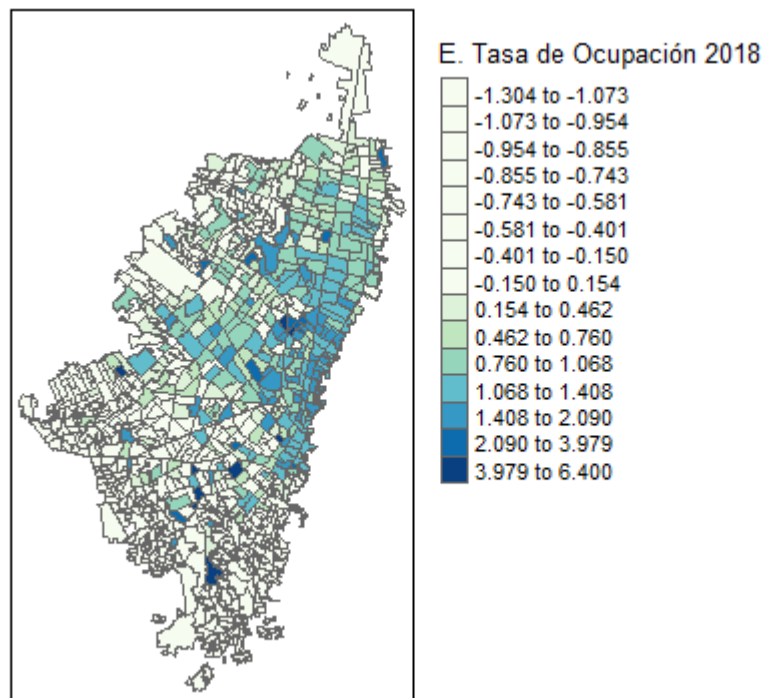


Figura 10: Estimación Interacción Espacio Temporal para el Logaritmo de la Tasa de Ocupación Airbnb y HomeAway por Barrio 2018

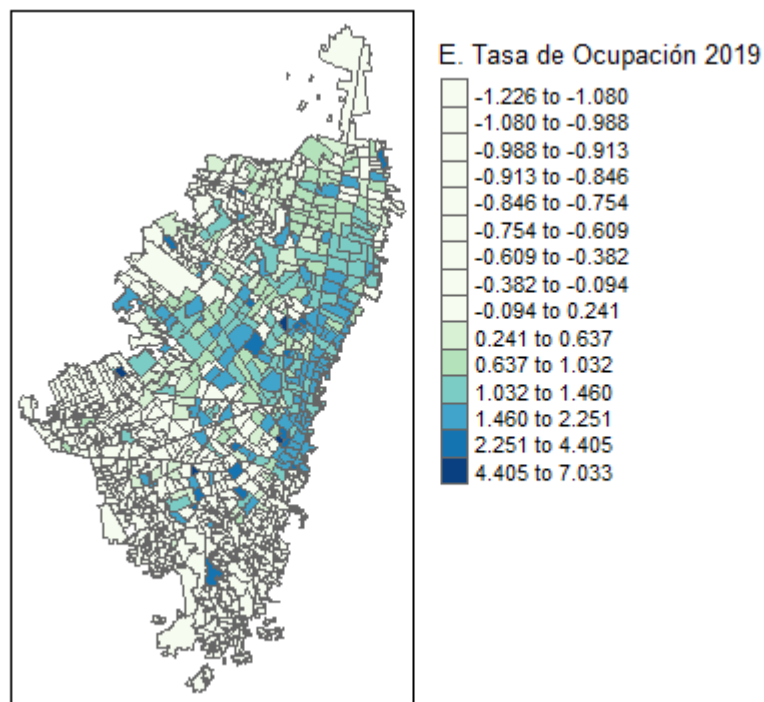


Figura 11: Estimación Interacción Espacio Temporal para el Logaritmo de la Tasa de Ocupación Airbnb y HomeAway por Barrio 2019

La figura 11 evidencia los cambios en la correlación espacio temporal con la Tasa de Ocupación para el año 2019. Para todos los años se observará una alta correlación temporal de orden 1 en los barrios ubicados en Chapinero Norte y central, también en algunos barrios de la localidad de Santafé, Candelaria y Usaquen. Sin embargo, para los años 2018 y 2019 se observa una mayor concentración en los barrios de la zona rosa de Chapinero, lo que podría estar indicando una disminución en la interacción temporal de barrios cercanos, lo que también puede estar relacionado con factores de oferta y otras comodidades como el número de huéspedes; además, el cambio en el ciclo económico también puede generar estas variaciones representadas en las diferencias temporales.

Resaltan los barrios ubicados en las localidades de Teusaquillo y Fontibón, los cuales van perdiendo intensidad entre el año 2015 y 2019 mostrando un posible emparejamiento respecto a las tasas de sus vecinos; esto evidencia también un aumento de competencia no solamente con otros establecimientos Airbnb, sino también con la oferta de hoteles existentes en el sector.

### 4.3. Regresión Poisson Espacio Temporal Bayesiana en INLA

Los modelos con Distribución Poisson se utilizan por lo regular para variables dependientes de conteo de tipo discreto. En este caso se incorpora la cantidad o el número de establecimientos Airbnb por barrio para realizar el análisis a través de esta regresión, con el objetivo de entender el crecimiento de estos alojamientos y su comportamientamiento en el tiempo y en los barrios de la ciudad. Las variables independientes aplicadas para el modelo son las siguientes:

1. Vigencia: Año evaluado.
2. Deposito USD: Promedio de depósito que pagan huéspedes en dólares por Barrio.
3. Limpieza USDP: Promedio de cobro de limpieza por el uso del establecimiento en dólares por Barrio.
4. Media Arriendo: Promedio mensual de los arriendos que se cobran en sitios residenciales aledaños a los establecimientos por Barrio.
5. Localidad: Localidad donde estan ubicados los alojamientos:
  - Localidad1: Usaquén
  - Localidad2: Chapinero
  - Localidad3: Santafé
  - Localidad4: San Cristóbal
  - Localidad5: Usme
  - Localidad6: Tunjuelito
  - Localidad7: Bosa
  - Localidad8: Kennedy
  - Localidad9: Fontibón
  - Localidad10: Engativá
  - Localidad11: Suba
  - Localidad12: Barrios Unidos
  - Localidad13: Teusaquillo
  - Localidad14: Mártires
  - Localidad15: Antonio Nariño
  - Localidad16: Puente Aranda
  - Localidad17: La Candelaria
  - Localidad18: Pafael Uribe Uribe
  - Localidad19: Ciudad Bolivar
6. Extra People USDP: Promedio cobro por personas extras que se admiten en los establecimientos por barrio.
7. ADR USDP: Promedio del cobro en establecimientos por barrio.



## Modelo

La estructura del modelo de regresión Poisson se plantea de la siguiente forma:

$$Y_{it} \sim \text{Poisson}(\lambda_{it}); \quad (67)$$

$$\eta_{it} = b_0 + \beta_1 \text{Vigencia}_{1i} + \beta_2 \text{Deposito}_{2i} + \beta_3 \text{Limpieza}_{3i} + \beta_4 \text{MediaArriendo}_{4i} + \\ \beta_5 \text{Localidad}_{5i} + \beta_6 \text{ExtraPeople}_{6i} + \beta_7 \text{ADR}_{7i} + u_i + v_i + \text{Temporal}_t \quad (68)$$

donde,  $t = 1, \dots, T$  en  $\text{Temporal}_t$ , con una estructura que puede ser paramétrica o no paramétrica.

### Valores del hiperparámetro elegido

Para especificar la distribución previa para cada parámetro  $\beta_0$  hasta  $\beta_n$ , se implementan distribuciones normales no informativas desde  $\beta_0 \sim N(0, 1)$  hasta  $\beta_n \sim N(0, 1)$

### Modelo Espacio Temporal Paramétrico

Para la Regresión Poisson el efecto espacial muestra una fuerte correlación en los barrios ubicados en las localidades de Chapinero, Usaquén, Santafé, Candelaria, Teusaquillo, Fontibón e incluso Engativá, con el número de alojamientos tipo Airbnb por barrio. Para este ejemplo se observa una mayor propagación hacia las zonas occidentales y suroccidentales de la ciudad lo que podría estar indicando la disposición de los propietarios para utilizar sus residencias como alojamientos turísticos de tipo alternativo.

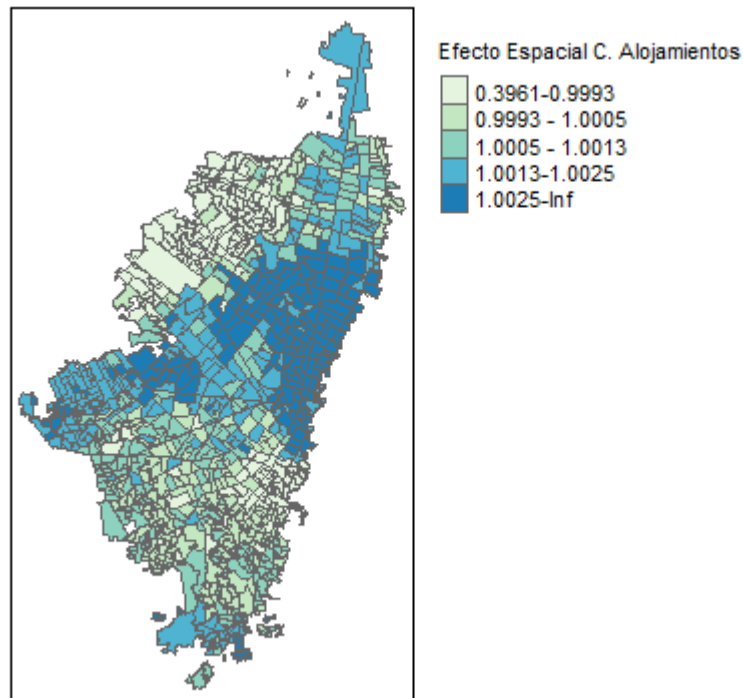


Figura 12: Estimación Efecto Espacial para el Logaritmo de la cantidad de Alojamientos por Barrio 2015 a 2019

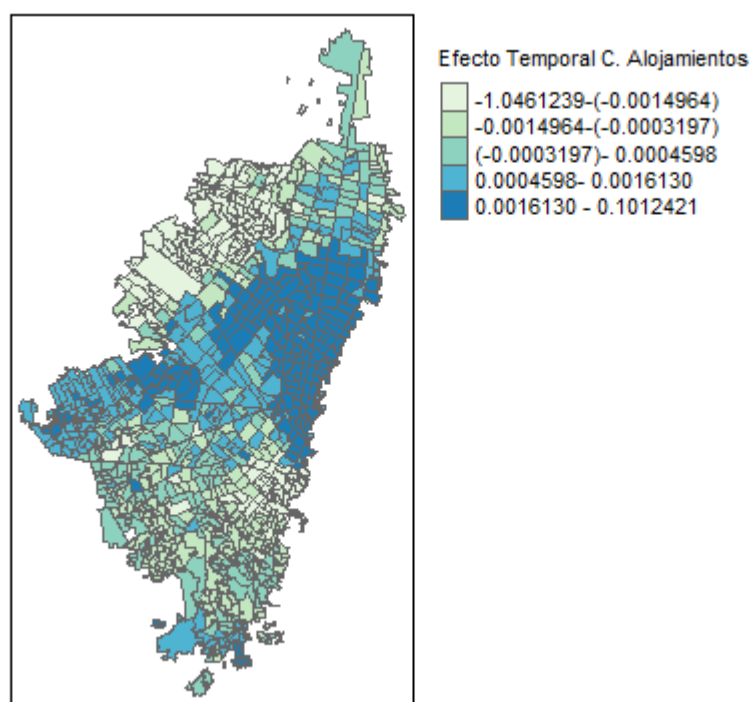


Figura 13: Estimación Efecto Temporal para el Logaritmo de la cantidad de Alojamientos por Barrio 2015 a 2019

En el caso del efecto diferencial temporal se puede evidenciar que en los barrios con mayor intensidad hay una correlación temporal más fuerte con el número de establecimientos ubicados en esas zonas. Al igual que en el efecto espacial, el temporal evidencia que hay un efecto significativo en la cantidad de barrios si se encuentran ubicados en las localidades del oriente de la ciudad: Chapinero, Teusaquillo, Candelaria y Santafé pero también en localidades como Fontibón y Kennedy; esta última localidad no sería un tradicional sector turístico, y evidenciaría cambios en las preferencias de los turistas.

En el caso del modelo de efectos fijos, se escoge entre dos alternativas a través del DIC; esto se observa en la tabla 9, donde el DIC del Modelo 1 que incluyó las variables independientes: Vigencia, Depósito en Dólares, Limpieza en Dólares y el costo por persona extra; tiene un mayor valor que el Modelo 2, el cual incluye además las variables: valor promedio del costo por alojamiento, media del arriendo de residencias no turísticas y las localidades en las que se encuentran los barrios.

Al analizar los efectos fijos del Modelo 2 se observa una correlación positiva de las variables: Media del arriendo que se cobra en los alojamientos residenciales ubicados en el barrio, es decir, que un aumento en la media de los arriendos tendría una pendiente positiva respecto al aumento de alojamientos tipo Airbnb, lo que podría evidenciar una relación entre el mercado de alojamiento para turistas y el mercado inmobiliario de residentes. Este comportamiento positivo también ocurre con el valor de las personas extra, lo que mostraría que los turistas prefieren pagar más si pueden estar ubicados en lugares estratégicos.

Tabla 7: Estimación Efecto Fijos para la Cantidad promedio de Alojamientos Airbnb y HomeAway por Barrio 2015 a 2019 (Modelo 1)

	Media	Desviación	Percentil 2.5	Percentil 50	Percentil 97.5	Moda	kld
(Intercept)	-660.580	13.10	-686.330	-660.570	-634.900	-660.560	0.000
Vigencia	0.330	0.010	0.310	0.330	0.340	0.330	0.000
Deposito_USD	0.000	0.000	-0.000	0.000	0.000	0.000	0.000
Limpieza_USDP	0.000	0.000	-0.000	0.000	0.010	0.000	0.000
ExtraPeople_USDP	0.020	0.000	0.010	0.020	0.020	0.020	0.000

Tabla 8: Estimación Efecto Fijos para la Cantidad promedio de Alojamientos Airbnb y HomeAway por Barrio 2015 a 2019 (Modelo 2)

	Media	Desviación	Percentil 2.5	Percentil 50	Percentil 97.5	Moda	kld
(Intercept)	-569.793	13.384	-596.068	-569.793	-543.539	-569.794	0.000
Vigencia	0.282	0.006	0.269	0.282	0.295	0.282	0.000
Deposito_USD	-0.000	0.000	-0.000	-0.000	-0.000	-0.000	0.000
Limpieza_USDP	0.000	0.002	-0.004	0.000	0.004	0.000	0.000
Media_Arriendo	0.003	0.000	0.002	0.003	0.003	0.003	0.000
Localidad2	0.934	0.188	0.562	0.9354	1.302	0.937	0.000
Localidad3	0.368	0.162	0.049	0.368	0.685	0.369	0.000
Localidad4	-0.139	0.119	-0.371	-0.139	0.095	-0.140	0.000
Localidad5	-0.122	0.119	-0.354	-0.123	0.111	-0.124	0.000
Localidad6	-0.073	0.165	-0.400	-0.072	0.249	-0.070	0.000
Localidad7	-0.135	0.120	-0.370	-0.135	0.103	-0.136	0.000
Localidad8	-0.074	0.116	-0.300	-0.075	0.154	-0.076	0.000
Localidad9	-0.033	0.136	-0.301	-0.0332	0.235	-0.033	0.000
Localidad10	0.023	0.118	-0.206	0.023	0.256	0.022	0.000
Localidad11	-0.002	0.114	-0.226	-0.003	0.223	-0.004	0.000
Localidad12	0.081	0.158	-0.231	0.081	0.391	0.082	0.000
Localidad13	0.591	0.204	0.189	0.592	0.990	0.593	0.000
Localidad14	0.068	0.161	-0.250	0.068	0.384	0.069	0.000
Localidad15	-0.138	0.177	-0.491	-0.137	0.206	-0.134	0.000
Localidad16	-0.170	0.137	-0.438	-0.170	0.099	-0.170	0.000
Localidad17	0.843	0.241	0.364	0.844	1.313	0.847	0.000
Localidad18	-0.144	0.128	-0.395	-0.144	0.108	-0.145	0.000
Localidad19	-0.1130	0.1147	-0.336	-0.113	0.113	-0.115	0.000
ExtraPeople_USDP	0.001	0.003	-0.005	0.0014	0.008	0.001	0.000
ADR_USDP	-0.004	0.000	-0.006	-0.004	-0.002	-0.004	0.000

Tabla 9: Deviance Information Criterion (DIC) para Modelos de Regresión Poisson

DIC Modelo 1	DIC Modelo 2
15200.73	14535.39

Al observar el comportamiento de las localidades, respecto a la localidad de Usaquén son significativas de forma positiva: Chapinero, Santafé, Teusaquillo y Candelaria, para explicar el número de alojamientos por barrio es decir, que pertenecer a dichas localidades estaría correlacionado con el aumento de alojamientos por barrio. Finalmente, el promedio del cobro en estadía por barrio, y el depósito que deben hacer los turistas por tomar el alojamiento presentan una correlación negativa, lo que implica que disminuir estos costos podría correlacionarse con el aumento de alojamientos por barrio.

Por otro lado, el *kld*, cercano a cero en todas las covariables evidencia que la utilización de la Aproximación de Laplace Simplificada sería una mejor opción que realizar el ejercicio con una metodología Gaussiana estandar.

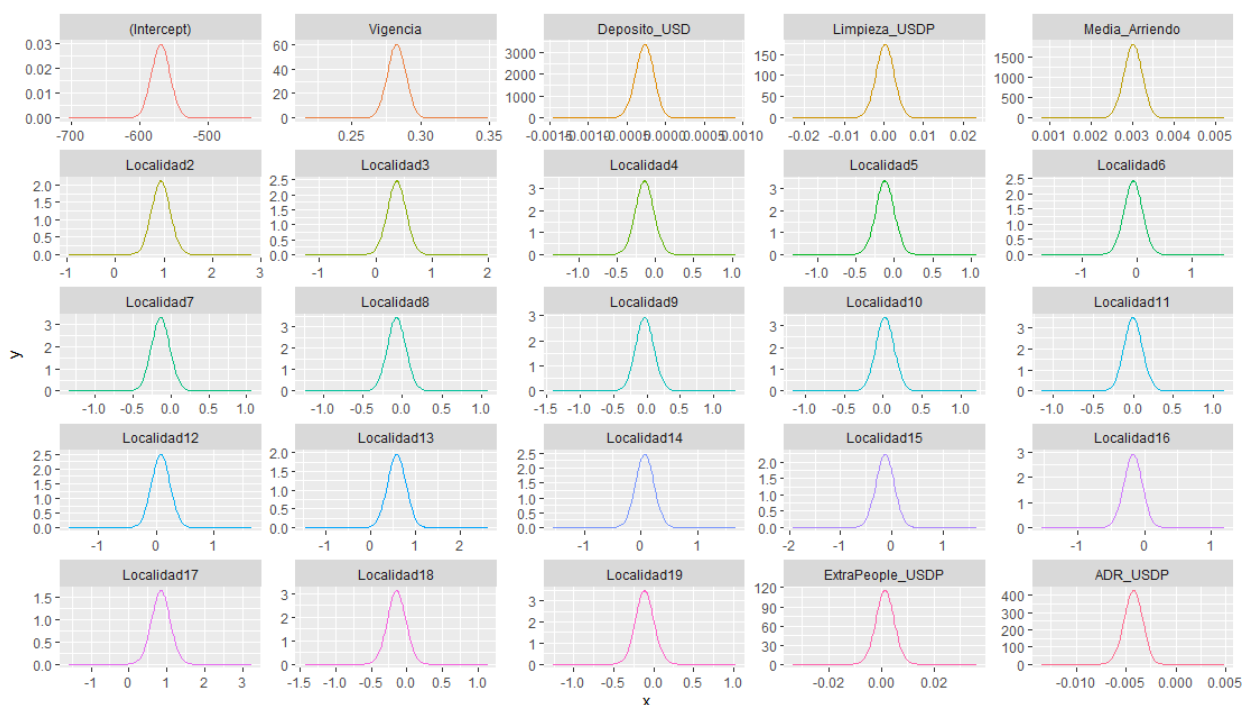


Figura 14: Densidades Modelo Poisson Espacio Temporal para la Cantidad de Alojamientos Airbnb y HomeAway por Barrio 2015 a 2019

La gráfica de densidades por covariable mostraría que las covariables: Cobro por limpieza, Cobro por persona extra, y las localidades 4,5,6,7,8,9,10,11,12,14,15,16,18, y 19, es decir las localidades que se encuentran alejadas del centro de la ciudad, no serían significativas dado que contienen al cero dentro de la distribución.

Se evidencia que las medias posteriores y los intervalos de credibilidad del 95% de los modelos de caminatas autorregresivas y aleatorias se ajustan al número de propiedades Airbnb por barrio; muestran para las covariables significativas que hay un mayor grado de ajuste del modelo a los datos y en el caso del año o Vigencia estimaciones suaves lo que podría evidenciar la correlación temporal con una caminata aleatoria de orden 1. La covariable, Localidad, evidencia intervalos muy ajustados en algunos segmentos, lo que puede estar relacionado con que la mayoría de las localidades no fueron significativas en el modelo.

### Tendencia Dinámica de Interacción Espacio Temporal No Paramétrica

Dado que se emplea una interacción dinámica de tipo II, La figura 9 evidencia que los establecimientos tipo Airbnb ubicados principalmente en la zona de Chapinero, Santafé y Candelaria, pero también en la localidad de Fontibón en cercanía al Aeropuerto, tendrían una mayor correlación con una caminata aleatoria de orden 1, es decir, tendrían dependencia temporal respecto al año inmediatamente anterior para pronosticar el número de alojamientos por barrio. Este fenómeno ocurre con una menor intensidad pero de manera general en zonas ubicadas en el suroccidente, centro y norte de la ciudad.

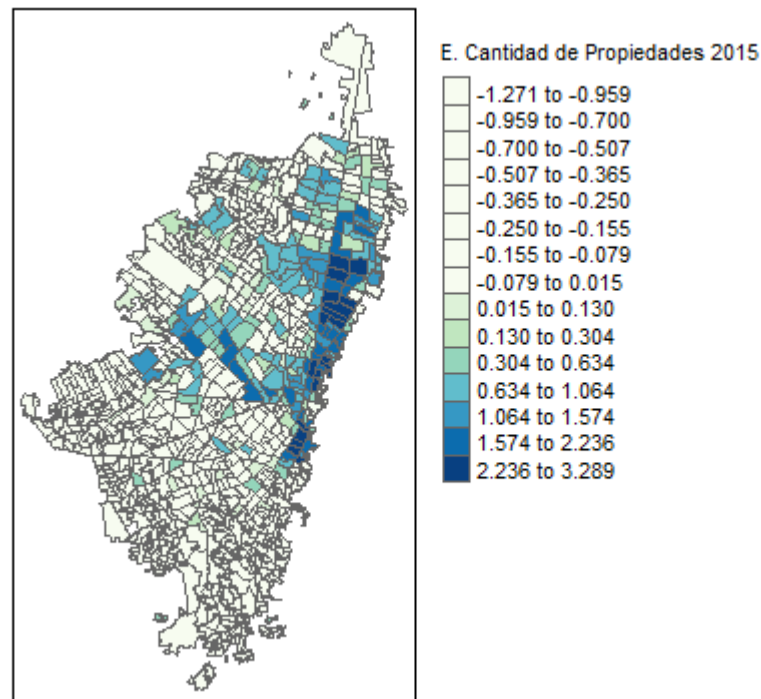


Figura 15: Estimación Temporal para el Logaritmo de la Cantidad de Propiedades Airbnb y HomeAway por Barrio 2015

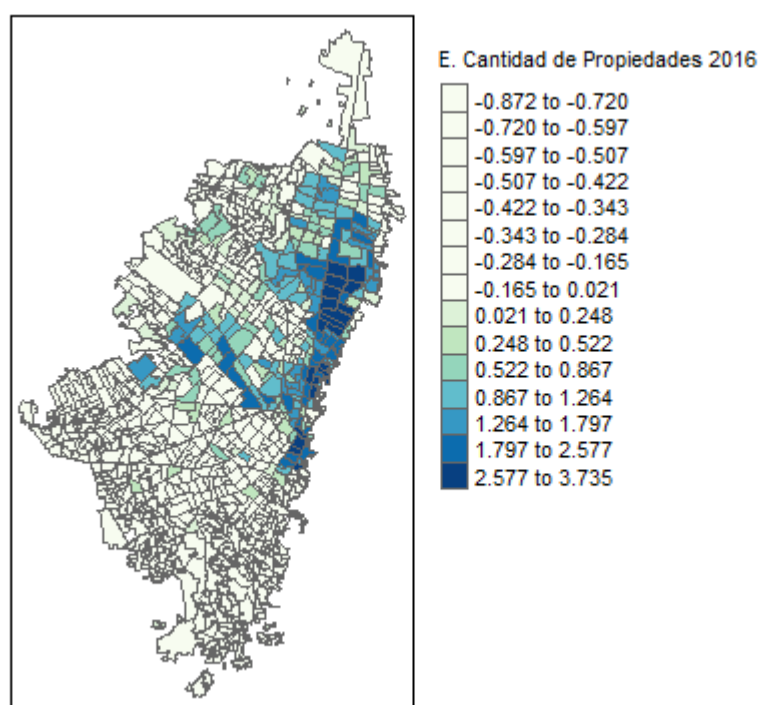


Figura 16: Estimación Temporal para el Logaritmo de la Cantidad de Propiedades Airbnb y HomeAway de Cantidad de Propiedades Airbnb y HomeAway por Barrio 2016

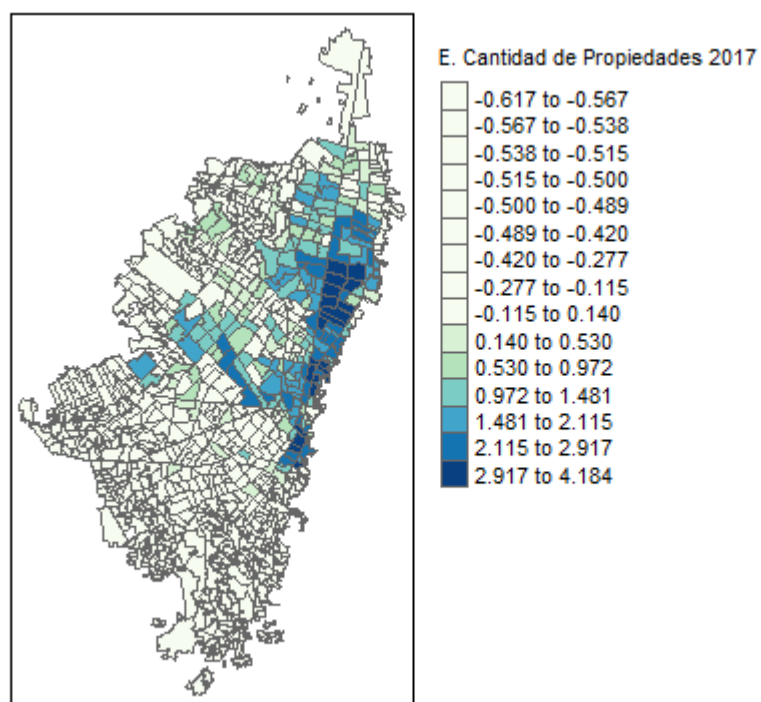


Figura 17: Estimación Temporal para el Logaritmo de la Cantidad de Propiedades Airbnb y HomeAway por Barrio 2017

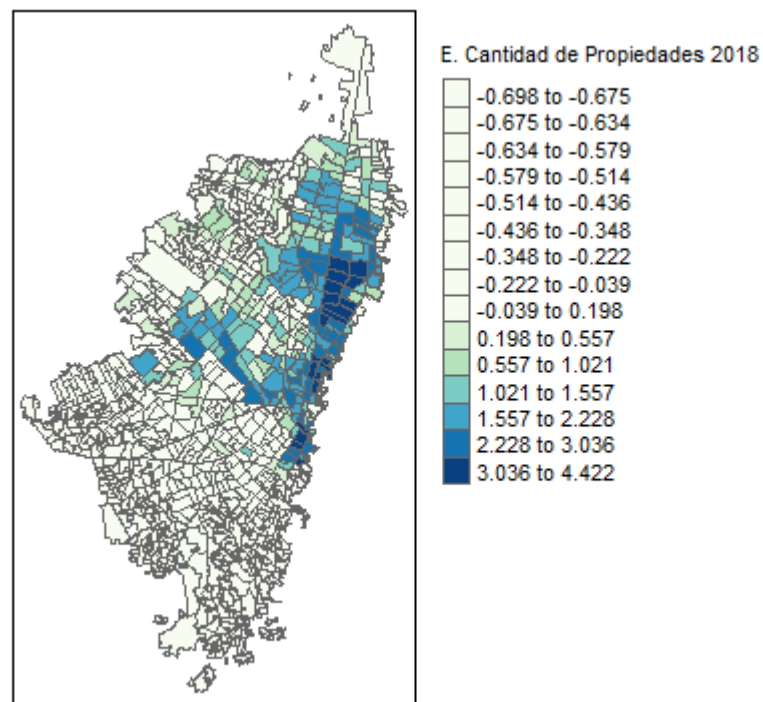


Figura 18: Estimación Temporal para el Logaritmo de la Cantidad de Propiedades Airbnb y HomeAway por Barrio 2018

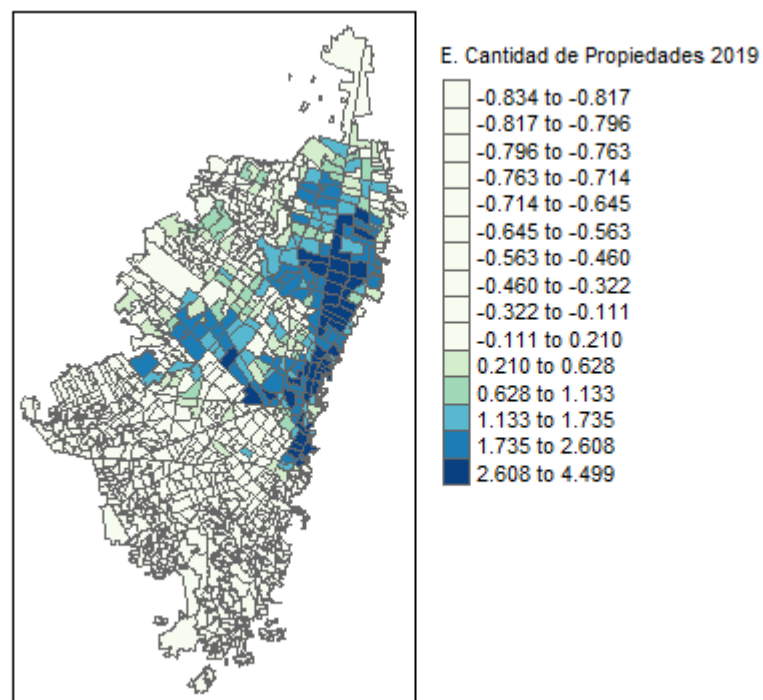


Figura 19: Estimación Temporal para el Logaritmo de la Cantidad de Propiedades Airbnb y HomeAway por Barrio 2019

A pesar de que los lugares con mayor correlación son similares, la intensidad aumenta a partir del año 2017 en las localidades de Chapinero, Santafé y Candelaria; en Teusaquillo por ejemplo, se observan cambios de intensidad para el año 2019 donde se consolida la localidad de Chapinero y se enfatizan algunos barrios de Usaquén, También se resalta la importancia que toman algunos barrios de la localidad de Suba, del sur de Fontibón y Kennedy en número de propiedades, los cuales no son tradicionales en el análisis del turismo de la ciudad.

#### 4.4. Regresión Gamma Espacio Temporal Bayesiana en INLA

Para estimar el total de ingresos de los establecimientos Airbnb por barrio, se implementa una regresión Gamma que tiene como característica una distribución entre cero (0) e infinito  $\infty$ , lo cual se aplica a valores continuos truncados a partir de cero. La flexibilidad de esta distribución es aplicada para precios y valores que no son negativos. Las covariables utilizadas en el modelo son las siguientes:

1. Vigencia: Año evaluado.
2. Deposito USD: Promedio de depósito que pagan huéspedes en dólares por Barrio.
3. Limpieza USDP: Promedio de cobro de limpieza por el uso del establecimiento en dólares por Barrio.
4. Media Arriendo: Promedio mensual de los arriendos que se cobran en sitios residenciales aledaños a los establecimientos por Barrio.
5. Localidad: Localidad donde estan ubicados los alojamientos:
  - Localidad1: Usaquén
  - Localidad2: Chapinero
  - Localidad3: Santafé
  - Localidad4: San Cristóbal
  - Localidad5: Usme
  - Localidad6: Tunjuelito
  - Localidad7: Bosa
  - Localidad8: Kennedy
  - Localidad9: Fontibón
  - Localidad10: Engativá
  - Localidad11: Suba
  - Localidad12: Barrios Unidos
  - Localidad13: Teusaquillo
  - Localidad14: Mártires
  - Localidad15: Antonio Nariño
  - Localidad16: Puente Aranda
  - Localidad17: La Candelaria
  - Localidad18: Pafael Uribe Uribe
  - Localidad19: Ciudad Bolivar



6. Extra People USDP: Promedio cobro por personas extras que se admiten en los establecimientos por barrio.
7. ADR USDP: Promedio del cobro en establecimientos por barrio.

### Modelo

La estructura del modelo de regresión Gamma se plantea de la siguiente forma:

$$Y_{it} \sim \text{Gamma}(\mu, \phi); \quad (69)$$

$$\eta_{it} = b_0 + \beta_1 Vigencia_{1i} + \beta_2 Deposito_{2i} + \beta_3 Limpieza_{3i} + \beta_4 MediaArriendo_{4i} + \beta_5 Localidad_{5i} + \beta_6 ExtraPeople_{6i} + \beta_7 ADR_{7i} + u_i + v_i + Temporal_t \quad (70)$$

donde,  $t = 1, \dots, T$  en  $Temporal_t$ , con una estructura que puede ser paramétrica o no paramétrica.

### Valores del hiperparámetro elegido

El hiperparámetro para este modelo es el parámetro de precisión  $\phi$ , que tambien se puede representar como,

$$\phi = \exp(\theta), \quad (71)$$

donde la distribución previa esta definida en  $\theta$ , la cual es una distribución  $\log\text{Gamma}(a, b)$ , con  $a$  como el parámetro de forma y  $b$  el parámetro de escala inverso y se asignan los valores  $a = 1$  y  $b = 0.1$ .

### Modelo Espacio Temporal Paramétrico

El Modelo Gamma con aplicación paramétrica, la cual indica rigidez lineal en el componente temporal, muestra que los barrios ubicados en la localidades de: Engativá, Fontibón, Kennedy y Suba, tienen una correlación más fuerte respecto a los ingresos totales recibidos y que el efecto contagio de barrios cercanos podría evidenciarse en el modelo espacial.

En el caso del efecto diferencial temporal el comportamiento es similar al espacial. Esto no necesariamente indicaría un mayor nivel de ingresos, sino cambios más fuertes en la correlación con la variable analizada respecto a los barrios vecinos y respecto a los años anteriores.

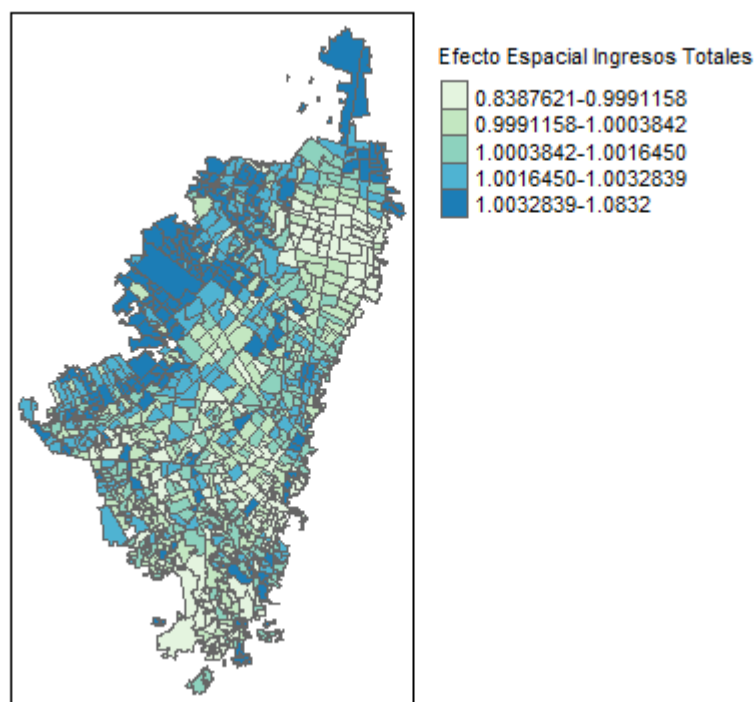


Figura 20: Estimación Efecto Espacial para el Logaritmo de la suma de Ingresos de Alojamientos por Barrio 2015 a 2019

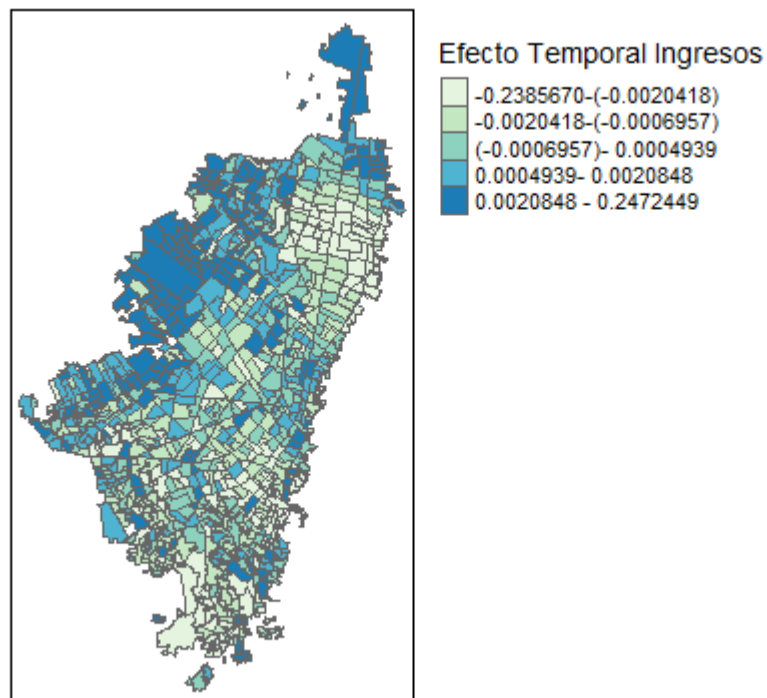


Figura 21: Estimación Efecto Temporal para el Logaritmo de la Ingresos de Alojamientos por Barrio 2015 a 2019

Al analizar los efectos fijos el Modelo 2 con un mayor número de variables presenta un menor DIC, y por eso se escoge para realizar el análisis. En dicho modelo se observa un intercepto con una alta correlación positiva con la variable ingresos totales lo que significa el aporte si las otras covariables analizadas tuvieran valores de cero. La vigencia o el año sería significativo con pendiente negativa, posiblemente relacionado con la disminución de ingresos totales en el tiempo.

Las covariables depósito en dólares, cobro de limpieza en dólares, cobro de persona extra en dólares y promedio del arriendo en dólares, tendrían una correlación positiva con la suma de ingresos de los establecimientos Airbnb por barrio, lo que podría indicar que con estos cobros los establecimientos generarían ingresos significativos y la disposición de los turistas a pagar por ubicación.

Es clave mencionar que la covariable, media del arriendo de otro tipo de propiedades no turísticas, fue incluida para intentar modelar la relación con otros precios y si bien no es posible afirmar que en los barrios donde hay alojamientos Airbnb existe un fenómeno de gentrificación, definido como reestructuración espacial de una determinada área urbana, lo cual implica el desplazamiento de los residentes de bajos ingresos que habían vivido en estos espacios (Glass, R. 1964); si es posible observar una correlación positiva con las variables analizadas: tasa de ocupación, cantidad de propiedad y suma de ingresos por barrio, estudio que podría generarse a partir del presente trabajo.

Tabla 10: Estimación Efectos Fijos para la suma de Ingresos de Alojamientos por Barrio 2015 a 2019  
(Modelo 1)

	Media	Desviación	Percentil 2.5	Percentil 50	Percentil 97.5	Moda	kld
(Intercept)	348.510	18.540	312.110	348.520	384.880	348.520	0.000
Vigencia	-0.170	0.010	-0.190	-0.170	-0.150	-0.170	0.000
Deposito_USD	0.000	0.000	0.000	0.000	0.010	0.000	0.000
Limpieza_USDP	0.040	0.000	0.030	0.040	0.050	0.040	0.000
Media_Arriendo	0.010	0.000	0.010	0.010	0.020	0.010	0.000

Tabla 11: Estimación Efectos Fijos para la suma de Ingresos de Alojamientos por Barrio 2015 a 2019  
(Modelo 2)

	Media	Desviación	Percentil 2.5	Percentil 50	Percentil 97.5	Moda	kld
(Intercept)	356.679	18.444	320.463	356.680	392.855	356.683	0.000
Vigencia	-0.175	0.009	-0.193	-0.175	-0.157	-0.175	0.000
Deposito_USD	0.001	0.000	0.001	0.001	0.002	0.001	0.000
Limpieza_USDP	0.018	0.004	0.009	0.018	0.027	0.018	0.000
Media_Arriendo	0.013	0.000	0.012	0.013	0.013	0.013	0.000
Localidad2	0.249	0.247	-0.229	0.247	0.741	0.243	0.000
Localidad3	0.339	0.169	0.009	0.338	0.673	0.336	0.000
Localidad4	0.149	0.099	-0.047	0.149	0.343	0.150	0.000
Localidad5	0.142	0.099	-0.054	0.142	0.336	0.143	0.000
Localidad6	0.168	0.138	-0.102	0.167	0.440	0.167	0.000
Localidad7	0.151	0.101	-0.048	0.151	0.349	0.152	0.000
Localidad8	0.122	0.097	-0.071	0.122	0.313	0.123	0.000
Localidad9	0.043	0.113	-0.180	0.043	0.266	0.043	0.000
Localidad10	0.110	0.103	-0.092	0.110	0.311	0.110	0.000
Localidad11	0.088	0.096	-0.101	0.088	0.276	0.089	0.000
Localidad12	-0.119	0.143	-0.398	-0.120	0.163	-0.121	0.000
Localidad13	-0.245	0.228	-0.685	-0.247	0.209	-0.252	0.000
Localidad14	0.197	0.147	-0.089	0.196	0.488	0.195	0.000
Localidad15	0.073	0.142	-0.204	0.072	0.354	0.071	0.000
Localidad16	0.052	0.112	-0.169	0.052	0.273	0.052	0.000
Localidad17	1.039	0.347	0.372	1.034	1.735	1.024	0.000
Localidad18	0.151	0.106	-0.058	0.152	0.360	0.152	0.00
Localidad19	0.134	0.096	-0.055	0.134	0.322	0.135	0.000
ExtraPeople_USDP	0.060	0.005	0.0492	0.060	0.071	0.060	0.000
ADR_USDP	0.022	0.001	0.018	0.0220	0.025	0.022	0.000

Tabla 12: Deviance Information Criterion (DIC) para Modelos de Regresión Poisson

DIC Modelo 1	DIC Modelo 2
67927.31	67363.33

La figura 22 muestra las distribuciones por covariables, encontrándose que respecto a la localidad de Usaquén solamente pertenecer a las localidades de Santafé y Candelaria serían significativas para explicar los ingresos totales por barrio. Esto difiere de los análisis anteriores dado que localidades tradicionales como: Chapinero y Teusaquillo no tendrían significancia para explicar los ingresos por barrio. También es importante observar que el,  $kld$ , al igual que los anteriores modelos esta en cero para todas las covariables, lo que permite inferir que la Aproximación de Laplace es pertinente para este modelo.

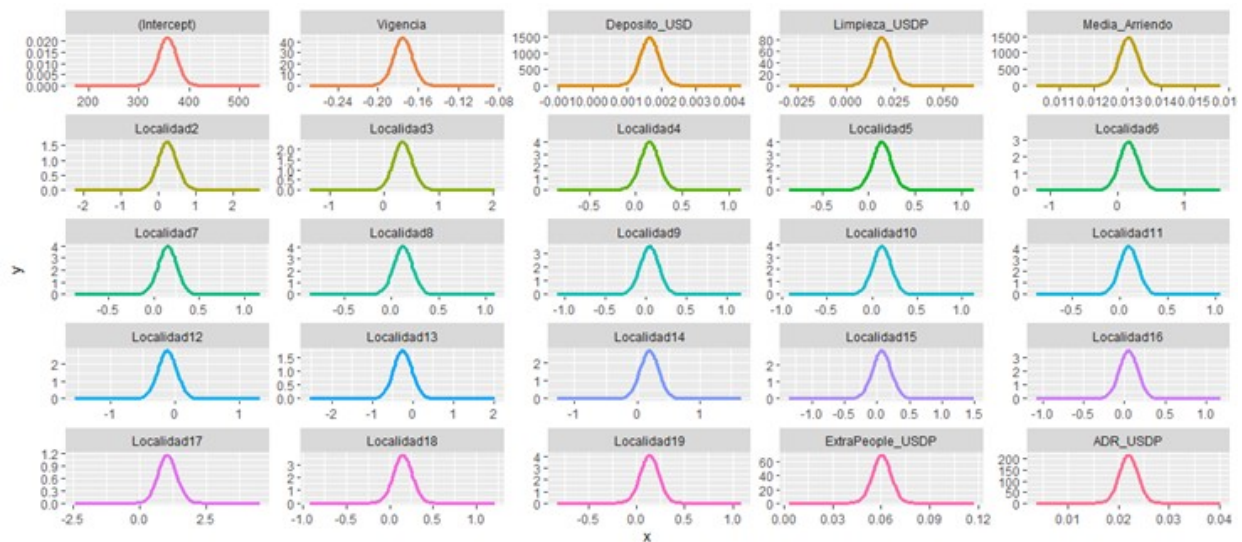


Figura 22: Densidades Modelo Gamma Espacio Temporal para los Ingresos de Alojamientos Airbnb y HomeAway por Barrio 2015 a 2019

Se observa el comportamiento de la media posterior e intervalo de credibilidad para las covariables del modelo, mostrando un ajuste en los datos de la mayoría y bandas más suaves en la vigencia o año, lo que puede evidenciar una correlación temporal con una caminata aleatoria de orden 1.

### Tendencia Dinámica de Interacción Espacio Temporal No Paramétrica

En este modelo la tendencia temporal no está definida linealmente y existe independencia entre áreas para generar las estimaciones temporales. Al observar los resultados en el mapa se observa una mayor correlación temporal en los barrios ubicados en Chapinero, Santafé, Candelaria y Teusaquillo, lo cual podría estar relacionado con la corrección que realiza el modelo sobre la tendencia temporal, pero también debido a la independencia entre vecinos. Los barrios con menos intensidad también estarían mostrando actividad de los alojamientos y la correlación temporal rezagada un año.

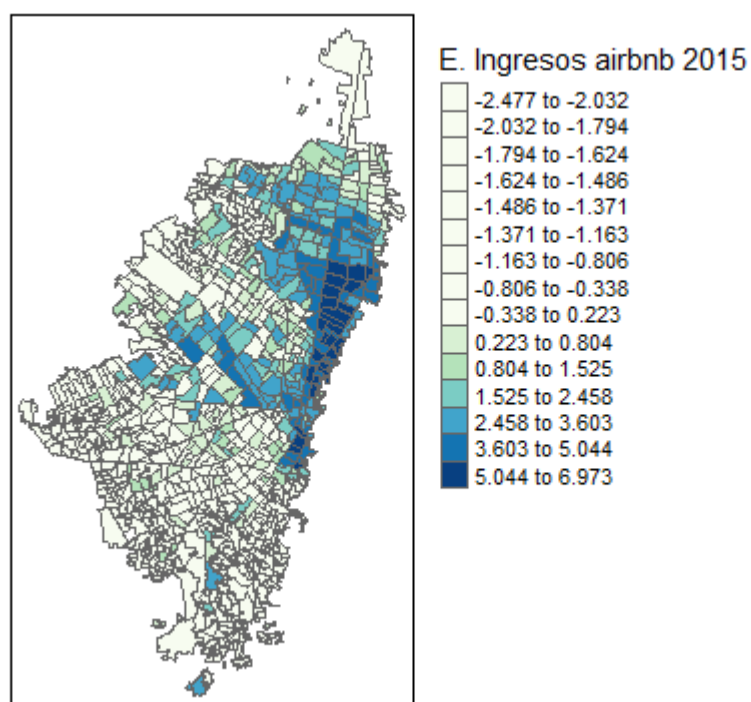


Figura 23: Comportamiento de Media Posterior e Intervalo de Credibilidad para el Logaritmo de los Ingresos Totales de Establecimientos Airbnb y HomeAway por Barrio 2015

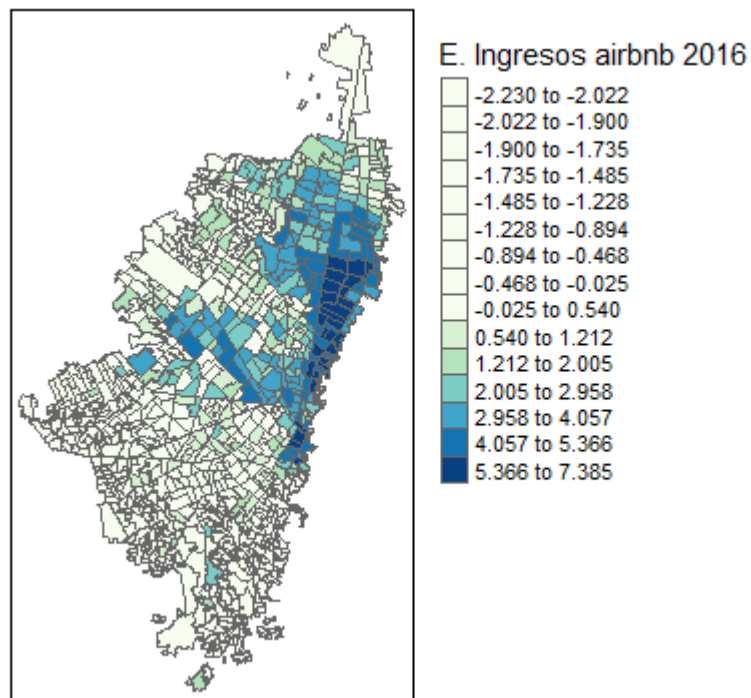


Figura 24: Comportamiento de Media Posterior e Intervalo de Credibilidad para el Logaritmo de los Ingresos Totales de Establecimientos Airbnb y HomeAway por Barrio 2016

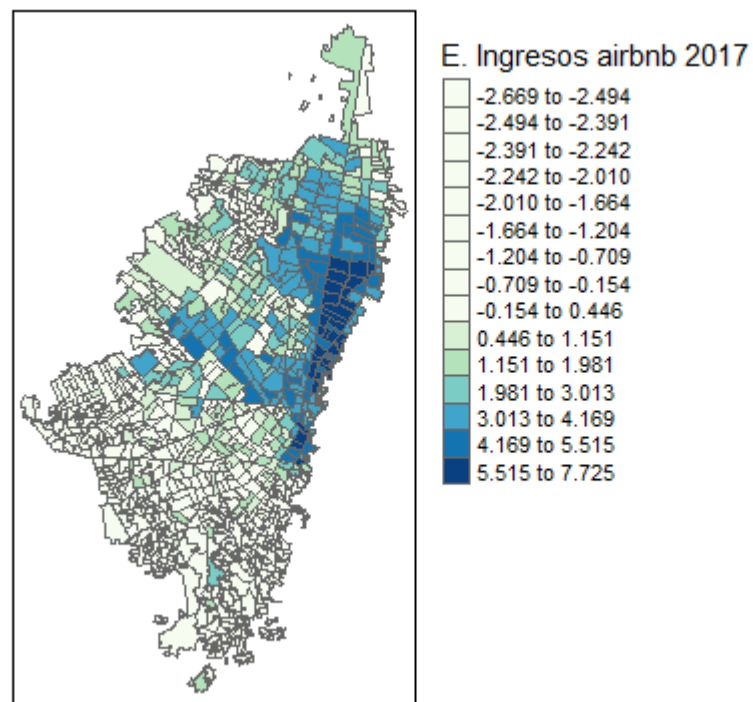


Figura 25: Comportamiento de Media Posterior e Intervalo de Credibilidad para el Logaritmo de los Ingresos Totales de Establecimientos Airbnb y HomeAway por Barrio 2017

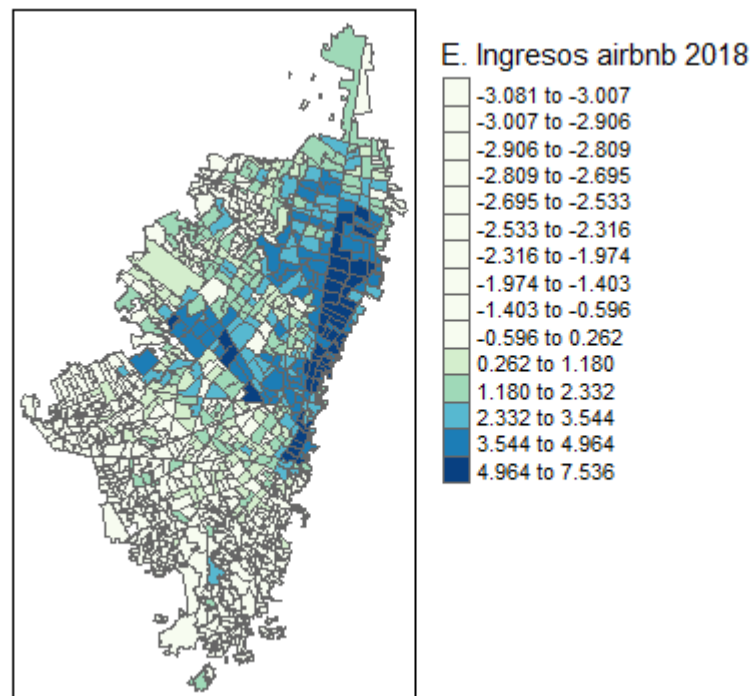


Figura 26: Comportamiento de Media Posterior e Intervalo de Credibilidad para el Logaritmo de los Ingresos Totales de Establecimientos Airbnb y HomeAway por Barrio 2018

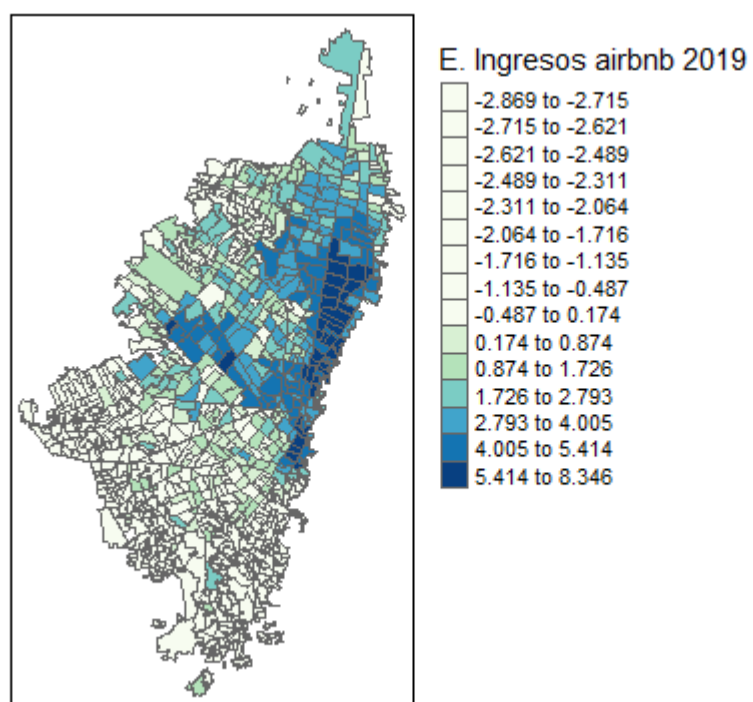


Figura 27: Comportamiento de Media Posterior e Intervalo de Credibilidad para el Logaritmo de los Ingresos Totales de Establecimientos Airbnb y HomeAway por Barrio 2019



#### 4.5. Estimación Posterior para el Análisis del Mercado de Alojamientos Turísticos en Bogotá

Los modelos espacio temporales bayesianos permiten la incorporación de varios elementos de análisis que pueden generar correlación con las variables dependientes analizadas, generando valores complementarios a los modelos tradicionales. Para el presenta análisis con los modelos INLA también es posible generar estimaciones concretas sobre el comportamiento espacial (por barrio) y temporal sobre los indicadores medidos.

En la Figura 11 se observa la estimación posterior promedio para cada variable. Respecto a la Tasa de Ocupación en el Modelo Beta la mayoría de barrios que superan el 40 % de ocupación estaran ubicados principalmente las localidades de la Candelaria, Teusaquillo, Fontibón, Usaquén, Suba y Barrios Unidos, destacandose los barrios: Escuela de Caballería, Santa Helena, San José del Prado, Julio Flores y Popular modelo, como aquellos con mayores tasas de ocupación promedio.

Al analizar las cantidades de Alojamientos tipo Airbnb el Modelo Poisson estimó que los barrios con mayor número serían: Chapinero Central, Chicó Norte, Chicó Norte II y III, Las Nieves, Granada, Santa Barbara, Maria Cristina y Antiguo Norte, ubicados en Chapinero y Usaquén, donde, en promedio habrían más de 15 alojamientos por barrio.

En cuanto los Ingresos Totales, el Modelo Gamma estimaría que los barrios con mayores rubros son: Chicó Norte, Antiguo Contry, Granada, La Cabrera, Chicó Norte II y III y Los Rosales, ubicados en la localidad de Chapinero.

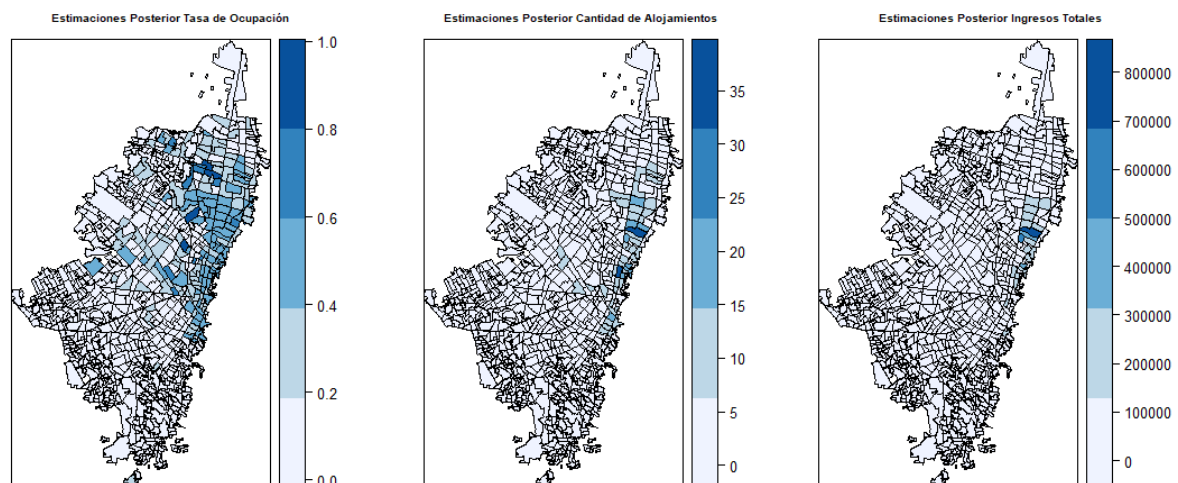


Figura 28: Estimaciones Posterior para el Logaritmo de la Tasa de Ocupación, la Cantidad de Alojamientos e los Ingresos Totales para Alojamientos Airbnb y HomeAway 2015 a 2019

En general, se puede decir que las estimaciones estarían mostrando una mayor correlación entre el lugar de ubicación, el tiempo rezagado en una unidad, las covariables que controlan y las variables dependientes analizadas. También estarían mostrando que el mercado de alojamientos turísticos en la ciudad no solo se concentra en la localidad de Candelaria y Teusaquillo, sino que también evidencian preferencias de los visitantes nacionales e internacionales por lugares menos centricos como Suba y Usaquén, pero consolidando a Chapinero como la localidad con mayor oferta, ocupación e ingresos de alojamientos alternativos turísticos.

## 5. Conclusiones y Recomendaciones

Los modelos espacio temporales otorgan una visión complementaria a mercados cuyo fin esta relacionado con actividades, zonas y lugares. Este trabajo incluye varias propuestas de covariables para explicar indicadores determinantes del sector económico del turismo a través de análisis de las variables explicativas: Tasa de Ocupación, Cantidad de Alojamientos e Ingresos totales por año y barrio.

En los tres modelos planteados, tanto en los desarrollos paramétricos como no paramétricos, se observan dinámicas de correlación espacial y temporal con propagación, especialmente en los barrios ubicados en las localidades de Chapinero, Teusaquillo, Santafé y Candelaria. Esto permite generar una descripción de las preferencias y gustos del turista que llega a Bogotá, pero además brindar un análisis del alcance y pertinencia del mercado de alojamientos alternativos en la ciudad. También permite generar insumos para un análisis en el que se pueda complementar a los alojamientos tradicionales como: hoteles, hostales y apartahoteles.

La ubicación espacio temporal especialmente para las variables de ingresos y cantidad de alojamientos estaría mostrando la fuerte relación que existe con los atractivos turísticos de la ciudad como la zona rosa de Chapinero y las zonas culturales y de negocios tradicionales de la ciudad. La tasa de ocupación estaría mostrando que lugares del norte de la ciudad ubicados principalmente en las localidades de Chapinero y Suba serían preferidos por los turistas, posiblemente relacionado con comodidades, zonas verdes y rápido acceso a centros culturales y de entretenimiento.

Zonas como Suba, Fontibón, Kennedy y Engativá las cuales no eran consideradas turísticas, muestran la diferencia de enfoques con la hotelería tradicional y la necesidad de los visitantes de conocer hábitos propios de los habitantes de la ciudad, pero también de ubicarse mejor por precios o cercanías a lugares de interés.

Las estimaciones de las distribuciones posteriores generan datos importantes sobre el posible efecto temporal y espacial en la ciudad, además estarían mostrando autocorrelación temporal de orden 1, es decir entre años, lo que también indica que es un mercado en movimiento continuo y posiblemente en crecimiento.

Las debilidades encontradas en los modelos espacio temporal Bayesianos, al igual que otros modelos Bayesianos clásicos, se traducen principalmente en el tiempo que se demora la iteración y consumo computacional. Esto puede mejorarse con nuevos ordenadores o con un desarrollo más amplio del Enfoque Varacional.

Se deberían actualizar los datos de fuentes oficiales a cortes más recientes, pero esto implicaría la compra por parte del Distrito de nueva información, lo que aún no se ha realizado. También se podrían utilizar técnicas de raspado en las páginas web de estos alojamientos alternativos para generar datos nuevos.

A manera de continuación, se propone para otros trabajos incluir estimaciones futuras sobre las variables dependientes; y la utilización de modelos truncados y cero inflados Bayesianos espacio temporales que incluyan los efectos de los lugares sin datos o con valores en cero.

## 6. Referencias

- Andersson H. y Hoffmann R. (2008) Spatial competition and farm tourism. A hedonic pricing model, American Agricultural Economics Association Annual Meeting.
- Batista F., Marín A., Rosina K., Ribeiro R., Freire S. y Schiavina M. (2018) Analysing spatiotemporal patterns of tourism in Europe at high-resolution with conventional and big data sources, *Tourism Management*, Volume 68, Pages 101-115, ISSN 0261-5177
- Banerjee S., Carlin B. y Gelfand A. (2004) Hierarchical modeling and analysis for spatial data, ISBN 1-58488-410-X.
- Banerjee S., Carlin B., y Gelfand A. (2004). Hierarchical modeling and analysis for spatial data, *Mono-graphs on Statistics and Applied Probability* 101
- Bernardinelli L., Clayton D., Pascutto C., Montomoli C., Ghislandi M., y Songini, M.(1995). Bayesian analysis of space-time variation in disease risk. *Statistics in Medicine*,14(21-22), 2433-2443.
- Besag J., York J. y Mollié A.(1991). Bayesian image restoration, with two applications in spatial statistics. *Annals of the Institute of Statistical Mathematics*, 43(1): 1-20.
- Blangiardo M, y Cameletti M.(2015). Spatial and spatio-temporal Bayesian models with R-INLA, John Wiley Sons, Ltd, p.308
- Caldeira A., Kastenholz E. (2015). Spatiotemporal behaviour of the urban multi-attraction tourist: does distance travelled from country of origin make a difference?, *Tourism Management Studies*, 11(1), en: <https://www.redalyc.org/pdf/3887/388743883011.pdf>
- Carlin, B., Banerjee, S., y Gelfand, A. (2014). Hierarchical Modeling and Analysis for Spatial Data (2nd ed.). Chapman and Hall/CRC. <https://doi.org/10.1201/b17115>
- Cepeda, E. (2001). Modelagem da variabilidade em modelos lineares generalizados, Ph.D. thesis, Universidade Federal Do Rio De Janeiro.
- Cepeda E. y Gamerman D. (2005). Bayesian methodology for modeling parameters in the twoparameter exponential family, *Estadística* 57, 93-105.
- Cepeda E. y Garrido L. (2011). Bayesian Beta Regression Models Joint Mean and Precision Modelling, <http://www.bdigital.unal.edu.co/5947/1/BayesianBetaRegresion.pdf>.
- Chen Q., Hu Z., Su H., Tang X., y Yu K. (2018). Understanding Travel Patterns of Tourists from Mobile Phone Data: A Case Study in Hainan, *IEEE International Conference on Big Data and Smart Computing (BigComp)*, 2018, pp. 45-51, doi: 10.1109/BigComp.2018.00016.
- Cuervo E, y Achcar J. (2010). Heteroscedastic nonlinear regression models, *Communications en Statistics USimulation and Computation*, 39(2), 405-419.
- Decreto Ley 897 de 2017 (2017). Por el cual se modifica la estructura de la Agencia Colombiana para la Reintegración de Personas y Grupos Alzados en Armas y se dictan otras disposiciones, Presidencia de la República de Colombia.
- Deller S. (2010). Rural poverty, tourism and spatial heterogeneity, *Annals of Tourism Research*, Vol. 37(1), p. 180-205.
- Demšar U., Harris P., Brunsdon C., Fotheringham S., y McLoone S. (2013). Principal Component Analysis on Spatial Data: An Overview, *Annals of the Association of American Geographers*, 103:1, 106-128, DOI: 10.1080/00045608.2012.689236.
- Díaz, A., y Sánchez, F. (2008). Los efectos del conflicto armado en el desarrollo social colombiano. Sánchez, Fabio. Las cuentas de la violencia. Bogotá. Universidad de los Andes-Editorial Norma.
- Ferrari, S. y Cribari-Neto, F. (2004). Beta regression for modelling rates and proportions, *Journal of*

Applied Statistics 31(7), 799–815.

Fotheringham, A., Brunson, C. y Charlton, M. (2002). Geographically weighted Regression: the analysis of spatially varying relationships. Chichester: Wiley, 196–202.

Gallagher L. (2018). La historia de Airbnb, Penguin Random House Grupo Editoria, México.

Glass, R. (1964). London, Aspects of Change. Londres, Macgibbon Kee.

Gelman A., Hwang J., y Vehtari A. (2014). Understanding predictive information criteria for Bayesian models, en: [http://www.stat.columbia.edu/~gelman/research/published/waic\\_understand3.pdf](http://www.stat.columbia.edu/~gelman/research/published/waic_understand3.pdf)

Gilks W., Richardson S., y Spiegelhalter D. (1996). Markov Chain Monte Carlo in Practice, ISBN 9780412055515 Published December 1, 1995 by Chapman and Hall/CRC, en: <https://www.routledge.com/Markov-Chain-Monte-Carlo-in-Practice/Gilks-Richardson-Spiegelhalter/p/book/9780412055515>

Gómez-Rubio V. (2020). Bayesian Inference with INLA, DOI:10.1201/9781315175584, edition: 1st Publisher: CRC Press ISBN: 9781138039872

Gordziejczuk M., y Lucero P. (2019). Turismo y calidad de vida: un estudio de autocorrelación espacial aplicado a la ciudad de Mar del Plata, provincia de Buenos Aires, Argentina, Cuadernos de geografía, Revista Colombiana de Geografía, en: <https://doi.org/10.15446/rcdg.v28n1.67275>

Gutierrez A. (2014). The Use of Working Variables in the Bayesian Modeling of Mean and Dispersion Parameters in Generalized Nonlinear Models with Random Effects, en: <http://www.tandfonline.com>

Harris P., Brunson C. y Charlton M. (2011). Geographically weighted principal components analysis. International Journal of Geographical Information Science, 25 (10), 1717–1736.

Harris P., Clarke A., Juggins S., Brunson C., y Charlton M. (2015). Enhancements to a geographically weighted principal components analysis in the context of an application to an environmental data set. Geographical Analysis, 47: 146–172.

Hidalgo G. (2019). Uso del Índice de Moran y LISA para explicar el ausentismo electoral rural en Ecuador, Revista Geográfica 160 — ISSN (impresa) 0031-0581 — ISSN (en línea) 2663-399X.

Hinde, J. y Demetrio H. (2007). Overdispersion: Models y Estimation, Computational Statistics Data Analysis Volume 27, Issue 2, 3 April 1998, Pages 151–170

Hoef J., Hans E., y Hooten M. (2017). On the Relationship between Conditional (CAR) and Simultaneous (SAR) Autoregressive Models, arXiv:1710.07000v1 [math.ST] 19 Oct 2017, en: <https://arxiv.org/pdf/1710.07000.pdf>

Instituto Distrital de Turismo de Bogotá (IDT) (2019). Base de Datos Airbnb por Barrios, en: [https://datosabiertos.bogota.gov.co/dataset/barrios\\_airbnb](https://datosabiertos.bogota.gov.co/dataset/barrios_airbnb)

Instituto Distrital de Turismo de Bogotá (IDT) (2020). Encuesta de Viajeros de Bogotá 2019, Observatorio de Turismo, en: <http://www.idt.gov.co>

Instituto Distrital de Turismo de Bogotá (IDT) (2020). Comparación de Prestadores de Servicios Turísticos 2019 vs. 2020, Observatorio de Turismo, en: <http://www.idt.gov.co>

Inside Airbnb (2019). en: <http://insideairbnb.com/about.html>

Jeffers J. (1967). Two case studies in the application of principal component analysis. Journal of the Royal Statistical Society Series C (Applied Statistics), 16 (3): 225–236.

Junsheng L., Cui L., Jinjing T., Yaofeng M., y Xiaojin W. (2019). Spatiotemporal Coupling Factors and Mode of Tourism Industry, Urbanization and Ecological Environment: A Case Study of Shaanxi, China, Sustainability 11, no. 18: 4923, en: <https://doi.org/10.3390/su11184923>

Knorr-Held L. (2000). Bayesian modeling of inseparable space-time variation in disease risk. Statistics in Medicine, 19(17–18), 2555–2567.

- Kullback S., y Leibler R. (1951) On Information and Sufficiency. The Annals of Mathematical Statistics, 22(1), p. 79–86.
- Kumar S., Lal, R., y Lloyd C. (2012). Assessing spatial variability in soil characteristics with geographically weighted principal component analysis. Computational Geosciences, 16(3), 827–835.
- Liljenstolpe C. (2011). Valuation of environmental impacts of the Rural Development Program – A hedonic model with application of GIS, 122nd EAAE Seminar Evidence-Based Agricultural and Rural Policy Making: Methodological and Empirical Challenges of Policy Evaluation.
- Liu J., Wang J., Wang S., Wang J., y Deng G. (2018). Analysis and simulation of the spatiotemporal evolution pattern of tourism lands at the Natural World Heritage Site Jiuzhaigou, China, Habitat International, Volume 79, Pages 74–88, ISSN 0197-3975, en: <https://www.sciencedirect.com/science/article/pii/S019739751830049>
- Lloyd, C. (2010). Analysing population characteristics using geographically weighted principal components analysis: a case study of Northern Ireland in 2001. Computers, Environment and Urban Systems, 34(5), 389–399.
- Martino S. (2007). approximate Bayesian Inference for Latent Gaussian Models, Doctoral Theses at NTNU, 2007:214, en: [https://ntnuopen.ntnu.no/ntnu-xmlui/bitstream/handle/11250/258135/123679\\_FULLTEXT01.pdf?sequence=1](https://ntnuopen.ntnu.no/ntnu-xmlui/bitstream/handle/11250/258135/123679_FULLTEXT01.pdf?sequence=1)
- McCullagh P., y Nelder J. (1989). Generalized Linear Models (Chapman Hall/CRC Monographs on Statistics Applied Probability). 2nd Edition, Chapman and Hall/CRC, UK.
- Morales M., y López L. (2008). Estudio de homogeneidad de la dispersion en diseños a una via de clasificación para datos de proporciones y conteo. Revista Colombiana de Estadística, 32:59–71.
- Moran P. (1950). A Test for the Serial Independence of Residuals. Biometrika, 37, 178–18, en: <http://dx.doi.org/10.1093/biomet/37.1-2.178>
- Naimat U., Wanggen W., y Shui Y. (2020). Spatiotemporal Analysis of Tourists and Residents in Shanghai Based on Location-Based Social Network's Data from Weibo, ISPRS Int. J. Geo-Inf. 9, no. 2: 70. <https://doi.org/10.3390/ijgi9020070>
- Niavis S. (2020). Evaluating the spatiotemporal performance of tourist destinations: the case of Mediterranean coastal regions, Journal of Sustainable Tourism Volume 28 - Issue 9 , en: <https://www.tandfonline.com/doi/full/10.1080/09669582.2020.1736087?needAccess=true>
- Nelder J., y Wedderburn R. (1972) Generalized linear models. Journal of the American Statistical Society, A, 135(3):370 – 384.
- Nocedal J., Wright J. (1999). Quasi-Newton Methods. Numerical Optimization. New York: Springer. pp. 192–221. ISBN 0-387-98793-2.
- Página oficial HomeAway España (2020). en: <https://www.homeaway.es/discoverycentre/sugerencias-y-recursos/lograr-resultados-satisfactorios-en-homeaway/pagos-homeaway-respuestas-a-preguntas-frecuentes>
- Portal web medium, (2018). medium.com, en: <https://medium.com/airbnb-engineering>
- Revista La Agencia de Viajes Colombia (LADEVI), (2017). en: <http://ladevi.info/articles/3266>
- Riebler A., Sorbye S., Simpson D., y Rue H. (2016). An intuitive Bayesian spatial model for disease mapping that accounts for scaling, Cornell University, en: <https://arxiv.org/abs/1601.01180v1>
- Roos M., y Held L. (2011). Sensitivity analysis in Bayesian generalized linear mixed models for binary data, Bayesian Analysis, Number 2, pp. 259–278, en: <https://projecteuclid.org/download/pdf1/euclid.ba/1339612046>
- Rubalcaba J. (2020). Cosas que conviene saber al usar AIC, DIC y otros criterios de información, en: [https://jrubalcaba.github.io/posts/information\\_criterion/](https://jrubalcaba.github.io/posts/information_criterion/) : : *text = Criterio*

- Rue H., Marino S., y Chopin N. (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations, Department of Mathematical Sciences, Norwegian University for Science and Technology, N-7491 Trondheim, Norway, en: <https://rss.onlinelibrary.wiley.com/doi/10.1111/j.1467-9868.2008.00700.x>
- Rue H., y Held L. (2005). Gaussian Markov Random Fields: Theory and Applications, volume 104 of Monographs on Statistics and Applied Probability. Chapman Hall, London.
- Sanchez, M. (2008). Análisis Espacial de Datos y Turismo Nuevas Técnicas para el Análisis Turístico. Una Aplicación al Caso Extremeño, Revista de Estudios Empresariales. Segunda época. Número: 2 (2008). Páginas: 48 - 66
- Sauleau E., Mameli V., y Musio M. (2010). Using Integrated Nested Laplace Approximation for Modeling Spatial Healthcare Utilization, arXiv:1006.3764, Cornell University, en: <https://arxiv.org/abs/1006.3764>
- Simas, A., Barreto-Souza, W., y Rocha, A. (2010). Improved Estimator for a General Class of Beta Regression Models, Computational Statistics and Data Analysis Vol: 54 (2), p. 348-366.
- Schrödle, B. y Held, L. (2011). A primer on disease mapping and ecological regression using INLA. Computational Statistics, 26, 241-258
- Sun Y., Fan H., Helbich M., y Zipf A. (2013). Analyzing Human Activities Through Volunteered Geographic Information: Using Flickr to Analyze Spatial and Temporal Pattern of Tourist Accommodation, Progress in Location-Based Services, ISBN : 978-3-642-34202-8, en: [https://link.springer.com/chapter/10.1007/978-3-642-34203-5\\_4](https://link.springer.com/chapter/10.1007/978-3-642-34203-5_4)
- Torres, A. (2011). Estudio en bloques completos vía regresión Poisson en presencia de sobredispersión, Tesis Maestría Estadística, Facultad de Estadística, Universidad Nacional de Colombia.
- Wu, Q., Cheng, J., Chen, G., Hammel, D. y Wu, X. (2014). Socio-spatial differentiation and residential segregation in the Chinese city based on the 2000 community-level census data: A case study of the inner city of Nanjing. Cities, 39, 109-119.
- Xin L., Zuolin X., y Rui L. (2019). A Spatio-Temporal Bayesian Model for Estimating the Effects of Land Use Change on Urban Heat Island, ISPRS Int. J. Geo-Inf. 8, no. 12: 522. <https://doi.org/10.3390/ijgi8120522>
- Yang Y., y Zhang H. (2019). Spatial-temporal forecasting of tourism demand, January, Annals of Tourism Research 75:106-119 en: [https://www.researchgate.net/publication/330357975\\_spatial-temporal\\_forecasting\\_of\\_tourism\\_demand](https://www.researchgate.net/publication/330357975_spatial-temporal_forecasting_of_tourism_demand)
- Zhang S., Liu X., Tang J., Cheng S., Qi Y., y Wang Y. (2018). Spatio-temporal modeling of destination choice behavior through the Bayesian hierarchical approach, Physica A: Statistical Mechanics and its Applications, Volume 512, Pages 537-551, ISSN 0378-4371, en: <https://www.sciencedirect.com/science/article/pii/S037843711830983X>