UNIVERSIDAD SANTO TOMAS
PRIMER CLAUSTRO UNIVERSITARIO DE COLOMBIA

Facultad de Estadística
Maestría en Estadística Aplicada
Anteproyecto de trabajo de grado
Noviembre 2021

Modelo bivariado espacio temporal para la explicación de las tasas del infectados y muertos del COVID en Colombia a nivel departamental

Camilo Ernesto Cano Ortega.[a]
camilocano@usantotomas.edu.co

Wilmer Darío Pineda Rios.[b]
wilmerpineda@usantotomas.edu.co

Resumen

Los modelos de datos de áreas son usados en la descripción de comportamientos de variables en espacios delimitados conocidos como vecindades. Estas usualmente son producto de la geografía o por factores políticos. Al incluir un factor temporal, se observa como una variable no solo se relaciona con sus vecinos en un periodo t, sino también con elementos de periodos anteriores. El objetivo del presente trabajo es, a partir de la teoría de los modelos espacio temporales para datos de área, lograr explicar cómo la influencia de las vecindades define la evolución de los casos de COVID-19 a nivel departamental en Colombia, del periodo del 1 de junio 2020 al 31 de enero 2021.

***Palabras clave*:** Datos de área, Modelos CAR, COVID-19, Epidemiología, Bayesiana, espacio temporal.

Resumen

Lattice data models are used to describe the behavior of variables in delimited spaces known as neighborhoods, these usually generated geographically or artificially by political factors. In addition to these neighborhoods, we include a temporal factor, we can observe how a variable is not only related to its neighbors in a period t, but also elements from previous periods. This work's main idea is based on the theory of space-time models for area data, an application of these models usually applied to epidemiology to describe the evolution of COVID-19 at the departmental level in the Colombian territory, in the period chosen from study.

***Palabras clave*:** Area Data, CAR Model, COVID-19, Epidemiology, Bayesian, temporal space.

1. Introducción

Basados en la ley de Moore, la cual afirma que las compuertas lógicas se harán más pequeñas aumentando la capacidad de procesamiento de los dispositivos, en el último siglo fueron desarrollados transistores y microchips, que permitieron pasar de máquinas de cálculo mecánicas a computadoras. En la presente revolución tecnológica, se han presentado dispositivos como el GPS que han permitido la captura de datos en tiempo real, elemento primordial para los celulares modernos y el desarrollo de las IoT (Internet of Things) (López, 2015).

[a]Estudiante Masestría Estadístia Aplicada
[b]Director de Tesis, Profesor de Estadística

Diariamente se generan numerosos datos de tipo espacial que, sin el correcto análisis, no sería posible la generación de insights para negocios, ideas innovadoras o propuestas para problemas actuales a los que se enfrenta la humanidad. Desde la rama de la estadística, la respuesta al análisis de estos datos es la aplicación del análisis espacial, el cual se encuentra dividido en tres líneas principales: geoestadística, patrones puntuales y datos de área. Para el caso de los datos de área se establece la existencia de n lugares distribuidos a lo largo de una zona de estudios, de ahora en adelante llamado vecindades, dentro de las cuales se producirán eventos medidos de manera discreta o continua.

Los modelos de datos a nivel de áreas nacen bajo el concepto de analizar un evento en un mismo lugar y cómo puede ser explicado por los predictores a partir de las influencias de los vecinos del área de estudio. Con base a esto, nacen los modelos SAR y modelos CAR los cuales se originan al trasladar la teoría de series de tiempo a un análisis espacial (Shekhar and Xiong, 2007). El planteamiento de estos modelos permite la explicación del valor actual de un área a partir de los resultados de los vecinos, así como de covariables. Es decir, no solo los vecinos sino también todas las áreas estudiadas pueden afectar en una mayor o menor medida la vecindad de interés. Esta correlación espacial y afectación tiene distintas formas de medirse, por ende, puede influir en los resultados de los modelos donde se destacan las matrices de contigüidades y las matrices de distancias. Para la medición de las influencias se encuentran los indicadores de Moran y Geary (Society, 2018b), que explican la dependencia espacial existente entre las áreas estudiadas.

En el año 2019, surge en China el virus SARS-CoV-2 conocido como Coronavirus disease 2019, previamente llamado 2019–nCov (de ahora en adelante COVID). El desconocimiento de esta enfermedad y su fácil transmisión en un mundo globalizado, permitió que evolucionara de una epidemia a una pandemia global. Este virus al afectar la salud pública especialmente de las regiones densamente pobladas, tuvo consecuencias de carácter social, económico y político que, al corte de la entrega del presente trabajo, aún se continúan presentando. Debido al avance y la cantidad de infectados, se estudiaron de manera diaria los contagios y muertes presentados en Colombia (33 departamentos, tomando Bogotá como un elemento diferente de Cundinamarca) para el periodo del 1 junio 2020 al 31 enero de 2021, en donde la variable final sería la generación de tasas de contagios y muertes, que estarán en un intervalo de (0,1).

Estos resultados se evaluaron partiendo de modelos univariados, que tienen como covariables las pruebas de COVID realizadas en cada uno de los departamentos de estudio, para la descripción de la tasa de contagios y el total de contagios para describir la tasa de mortalidad. Como un segundo ejercicio, se evaluará la influencia que tienen tanto las pruebas COVID como los casos acumulados para la generación de un modelo bivariado que explique ambas tasas.

2. Marco Teórico

La estadística al alimentarse con otras ramas como el cálculo, la teoría de conjuntos y las ciencias de la computación, ha permitido el desarrollo de muchas herramientas que harían inhabitable el mundo moderno, pues han marcado el destino de guerras, programas espaciales, programas de planeación mundial y estrategias de comercio mundial. Entre todos los posibles campos de estudios, la estadística espacial es una de las que tiene una historia loable desde sus comienzos.

Durante un brote de colera en la ciudad de Londres en 1854, el médico británico Jon Snow (uno de los precursores de la epidemiología) sentó las bases para explicar que algunos eventos no tenían independencia entre sí, dado que, al estar distribuidos sobre un espacio, podían influenciar a los elementos próximos al evento estudiado. Puestas las bases de la estadística espacial, esta se segmentó sobre tres grandes campos de estudio: la geoestadística, datos de área y patrones puntuales.

La estadística espacial está altamente relacionada con la geografía dado que, con un instrumento como un mapa, no solo se realiza la pregunta del cuánto, sino en dónde puede surgir un evento (Cressie,

1993). En el caso de datos de áreas, a diferencia de los patrones puntuales y geoestadística, su principal aplicación y fuente de información se encuentra asociada a estructuras sociales (ciudades, municipios, departamentos, gobiernos, países, etc.) Razón que explica por qué los datos de área están estrechamente relacionados al concepto de censo y actividades gubernamentales de planes de desarrollo. Todos estos planes, controles ejercidos o propuestos por entes relacionados con una forma de gobierno deben crear indicadores que permitan evidenciar la evolución de la variable objeto de estudio. En contraparte, los datos puntuales y la geoestadísticos que se apoyan en su mayoría sobre estructuras naturales donde los aspectos físicos de la tierra pueden impactar sobre los resultados de un modelo propuesto.

Un área está definida como una colección contable de lugares que contienen un conjunto de valores, que pueden tener una forma regular o irregular en sus dimensiones (Bavaud, 1998). Cada área debe buscar tener una forma de identificación indexada, que permita a través de esta identificación favorecer el desempeño de los análisis. A pesar de existir una codificación, también es importante registrar una longitud (x) y una latitud (y) que permitan referenciarlos por medio de un sistema de información geográfico.

$$D = \{(x_i, y_i) \quad \forall i \in \mathbb{N}\}$$

Al extender múltiples áreas a lo largo de una región, surge el concepto de vecino entendido como áreas aledañas que pueden influir en el área de interés (Blangiardo and Cameletti, 2015). Este concepto es de vital importancia para el análisis espacial dado que, si cada área fuera idéntica e independientemente distribuida, es posible por herramientas clásicas de muestreo o modelos, explicar un fenómeno.

Al encontrarse evidencia de una correlación espacial, la forma de abordar el problema cambia completamente, puesto que los datos de un vecino, al no ser tomados en cuenta, aumentarían el error entre el valor estimado y el real. Cada vecino se diferencia de otro mediante representación gráfica o matricial por medio del concepto de frontera. Esto dado que cada i-ésimo elemento tiene unos valores de frontera únicos e irrepetibles en la región de estudio.

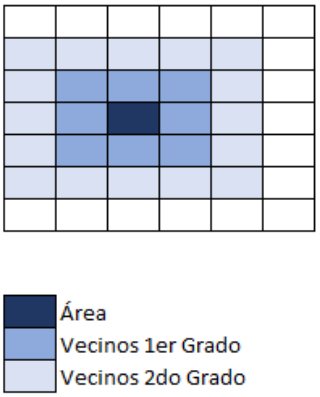

Figura 1: Estructura de vecinos según su cercania, en primer y segundo grado, Fuente: Adaptación del libro Spatial and Spatio-temporal Bayesian Models with R-INLA, Blangiardo, Cameletti Pág 176

Como se mencionó anteriormente, al estar altamente relacionadas a territorios, los datos de áreas poseen delimitaciones en donde en ocasiones mostrar interacciones con sus vecinos no presenta complejidad. Al existir correlaciones espaciales, pueden asignarse pesos a los vecinos de un área específica, en donde su información puede usarse como variables explicativas en el modelo dentro del área estudiada (Bivand et al.,

2008). Esto surge como consecuencia de tener similitudes en grupos étnicos, variables socioeconómicas o políticas de gobierno, lo que implica que una población de un área específica llegue a influenciar el desarrollo de un evento de estudio en el área de su vecino próximo.Esta influencia puede traducirse en una matriz de pesos, en donde dependiendo de la distancia que tenga un vecino con respecto al área de estudio puede estimarse su influencia.

A pesar de la importancia de estos pesos en los datos de áreas, en su mayoría las bases de toda la bibliografía se remontan a finales del siglo XX (Bavaud, 1998)(Griffith, 1993), en estos trabajos se rompe el paradigma que tenía el análisis espacial clásico, previo a los años 2000, en donde se asumía que existían pesos binarios. El análisis contemporáneo especifica que los pesos pueden variar según el área de influencia que se está estudiando, por criterios como la contigüidad, la distancia continua, vecindad más cercana, interacción espacial o bandas de distancia fija.

2.1. Indicadores de correlación Espacial

Un indicador de correlación espacial (ICA) o por sus siglas en inglés Local Indicator of Spatial Association (LISA) debe satisfacer dos requerimientos (Anselin, 1995). Cada observación estudiada debe indicar el grado de asociación espacial significativa comparada con valores similares alrededor de la observación. La suma de los indicadores espaciales de todas las observaciones es proporcional al indicador global de asociación. Formalmente, los términos I expresados por el ICA para la variable y_i, son la observación de la ubicación i como para el estadístico $L_i = f(y_i, y_{Ji})$

Donde f es la función de parámetros y y_{Ji} son los valores observados para el área j_i de i, usualmente el J_i se encuentra formalizado como la matriz de pesos promedios o de contiguidad. Los valores de y usados para el cálculo del estadístico pueden ser los datos reales, aunque se recomienda usar una estandarización, pues de esta manera evitamos el efecto de las escalas de variables en los análisis. Como requerimiento L_i, debe en lo posible inferir en la significancia de los patrones de asociación de la ubicación i. $P[L_i < \delta_i] \leq \alpha_i$,donde δ_i es el valor crítico para la asociación espacial y α_i es la significancia elegida.

Como segundo requerimiento, el ICA es una relación estadística global que se puede entender como, explicada $\sum_i L_i = \gamma\Lambda$. Donde Λ es el indicador global de asociación espacial y γ es el factor escalar. Se espera que la suma de los indicadores locales sea proporcional al indicador global. $P|\gamma < \delta| \leq \alpha$

2.1.1. Indicador I. de Moran

Desarrollado en 1950, se ha convertido en un método estándar de evaluación de correlación espacial que tiene aplicación tanto para variables continuas como discretas (Shekhar and Xiong, 2007):

$$I = \frac{N \sum_i \sum_j \mathbf{W_{i,j}}(X_i - \bar{x})(X_j - \bar{x})}{(\sum_i \sum_j \mathbf{W_{i,j}})(X_i - \bar{x})^2} \tag{1}$$

Donde N son el número de casos, $\bar{x}$ es la media de los valores de la variable X, X_i y X_j son los valores de la variable X, en las localizaciones i y j respectivamente, mientras que $\mathbf{W_{i,j}}$ son los pesos aplicados para comparar los valores.

Una propiedad que tiene el Índice de Moran, es que no depende exclusivamente de los valores de la variable X. Esta también depende de los valores de la matriz $\mathbf{W}$, debido a que si se evalua exclusivamente como 0 o 1 el estar en una frontera, habrá una diferencia si se analiza un mismo caso mediante una matriz de distancias al cuadrado entre dos puntos. La lectura del indicador es que a medida que el resultado se acerque a 1 o -1, tendrá una correlación alta, mientras que, si se encuentra cercana 0, no existe evidencia de que esos puntos estén correlacionados.

2.1.2. Indicador C. de Geary

A diferencia del índice de Moran, este método no es medido en la desviación con respecto a la media en los elementos, si no como la diferencia en cada observación:

$$C = \frac{(N-1)\sum_i \sum_j \mathbf{W_{i,j}}(X_i - X_j)^2}{2\sum_i \sum_j \mathbf{W_{i,j}}(X_i - \bar{x})^2} \tag{2}$$

Este indicador se espera que presente una variación entre 0 y 2, en donde 1 significa que no tiene correlación espacial, menor que 1 que presenta una correlación negativa. Mientras que el Índice de Moran da un indicador global de la correlación entre elementos, el método Geary es sensible a las diferencias pequeñas entre los barrios (Shekhar and Xiong, 2007).

Durante los análisis espaciales existen varios puntos a analizar para afrontar el problema de estudio (Cressie, 1993). Como regularidad o irregularidad de las localizaciones, desde el punto de vista espacial, se traduce en si las áreas de estudio presentan las mismas dimensiones. En el caso de análisis espacio temporales esto puede darse por aspectos administrativos, del estudio o factores exógenos paralelos a la investigación no será posible analizar exactamente las mismas áreas a través del tiempo.

2.2. Modelos Espaciales

La dependencia de una variable en el espacio puede surgir a partir de operaciones de procesos espaciales o como resultado de una asociación con variables espaciales dependientes. Los modelos estadísticos para datos espaciales deben tener en cuenta la variación entre las cantidades observadas en diferentes lugares. Esta variación puede ser representada a partir de la media, la correlación o una combinación de ambos (Society, 2018b) .

2.2.1. Auto regresión Espacial

Una serie de tiempo autorregresiva puede ser definida como:

$$X_t = \alpha X_{t-1} + \varepsilon_t \qquad \varepsilon_t \sim N(0, \sigma^2) \tag{3}$$

O ser definida mediante

$$\mathbf{E}(X_t)|X_{t-1}...X_{t-i}) = \alpha X_{t-1} \qquad \mathrm{Var}(X_t|X_{t-1}...X_{t-i}) = \sigma^2$$

Donde X_t asume ser un proceso gaussiano *"Un proceso estocástico se denomina "gaussiano" si todas sus funciones de distribución son gaussianas multivariables. [...]. Por lo tanto, la distribución de un proceso estocástico gaussiano está determinada solamente por la familia de esperanzas y matrices de covarianza de las funciones de distribución"*(Barbeito and Villalón, 2003). Dado que a diferencia de las series de tiempo trabajamos con vecinos, es posible considerar en el análisis los elementos t+1, considerando la asimetría hacia dos lados (Ripley, 2005), por lo tanto:

$$\begin{aligned} X_t &= \alpha X_{t-1} + \alpha X_{t+1} + \varepsilon_t \\ \mathbf{E}(X_t|X_{t-1}...X_{t-i}) &= \frac{\alpha}{2}(X_{t-1} + X_{t+1}) \\ \mathrm{Var}(X_t|X_{t-1}...X_{t-i}) &= \sigma^2 \end{aligned}$$

Esto genera la formación de los procesos Simultaneous Autoregressions (SAR) y Condicitional autoregressions (CAR).

Simultaneous Autoregressions (SAR)

A partir del desarrollo de las técnicas de series de tiempo, este conocimiento quiso trasladarse a otros ambientes especialmente a áreas relacionadas con las ciencias sociales. Uno de los primeros modelos desarrollados para el análisis espacial fue la auto regresión espacial simple, esta consiste exclusivamente en una versión rezagada de la variable dependiente y (Society, 2018b)

$$y = \rho \mathbf{W} y + \varepsilon \tag{4}$$

En donde:

W: Matriz de contigüidad construida a partir de una matriz de ponderación espacial de $n \times n$.
ρ: Parámetro espacial auto regresivo, cuando este es igual a 0, el modelo se reduce a mínimos cuadrados ordinarios.
y: Variable observada.

La matriz de pesos **W**, siempre debe estar estandarizada, en la cual la suma de sus filas debe ser igual a 1. Esta no debe ser necesariamente simétrica. Para despejar la y, dado que aparece en ambos lados de la ecuación, puede ser reescrita como:

$$y = (\mathbf{I} - \rho \mathbf{W})^{-1} \ \times \ e \tag{5}$$

De la cual podemos obtener la expresión de la varianza

$$var(y) = (\mathbf{I} - \rho \mathbf{W})^{-1} \mathbf{C} (I - \rho \mathbf{W}^T)^{-1} \tag{6}$$

En donde **C** es la matriz de varianzas y covarianzas, representada como $E(\epsilon \epsilon^t)$ donde ϵ son los errores de la variable respuesta

A pesar de que es ampliamente desarrollado, estos modelos presentan una debilidad, dado que en su origen se plantearon para organizaciones regulares e infinitas (mismas fronteras de igual tamaño, se asume mismo centroide para cada área), muy diferente a las aplicaciones reales irregulares (presentan diferentes cantidades de vecino) y finitas en el caso que se agreguen variables predictoras se genera un modelo mixto regresivo espacial autoregresivo (MRSA), dando como resultado:

$$y = \mathbf{X} \beta + \rho \mathbf{W} y + \varepsilon \tag{7}$$

En donde:
β: coeficientes de regresión
X: Matriz diseño, en donde se encuentran las variables independientes, convirtiendo $\mathbf{X}\beta$ en una combinación de un componente de tendencia lineal general (global).

Un segundo enfoque que puede modelar un SAR, es el modelo error espacial (Society, 2018b); este modelo se aplica cuando se supone una autocorrelación espacial significativa. Este modelo se define como:

$$y = \mathbf{X} \beta + \varepsilon \tag{8}$$

donde:

$$\varepsilon = \lambda \mathbf{W} \varepsilon$$

En principio es un modelo lineal estándar, pero ahora el error está compuesto por un vector ponderado de λ, la matriz $\mathbf{W}$, y $\mathbf{u}$ vector de errores *iid*.

$$\varepsilon = (\mathbf{I} - \lambda\mathbf{W})^{-1}u$$

2.3. Conditional Autoregressions (CAR)

Este modelo plantea que los valores estimados en cualquier ubicación, esta condicionada al nivel de los valores de los vecinos (Society, 2018a), este se describe de la forma:

$$\mathbf{E}(y_i|y_j, j \neq i) = \mu_i + \rho \sum_{j \neq i} w_{ij}(y_j - \mu_j) \tag{9}$$

donde:

μ es el valor esperado de la variable i. ρ es el parametro de autocorrelación espacial.

En el modelo CAR estándar, los pesos espaciales se calculan usando una función de mínima distancia, esta función puede determinarse a partir de información a priori, según el contexto del problema, aunque también puede realizarse a partir de semivariogramas o correlogramas(Society, 2018a).

En el modelo CAR la matriz de covarianza, mientras la varianza se suponga constante tiene la forma de:

$$\text{Var}(y) = (\mathbf{I} - \rho\mathbf{W})^{-1}\mathbf{I}\sigma^2 \tag{10}$$

2.3.1. Modelo Condicionales Autoregresivos Gaussianos y no Gaussianos

Los modelos condicionales autorregresivos usualmente son usados para inferencias basados en técnicas de máxima verosimilitud, pseudo máxima verosimilitud y en modelos jerárquicos. Están segmentados en dos grandes grupos los gaussianos, basados en la propiedad de los campos aleatorios de Gaussianos de Markov y los no gaussianos, llamados modelos auto logísticos para variables binarias (Gelfand et al., 2010).

CAR Gaussiano

Suponga que X_i donde $i = 1, 2, \ldots, n$, como un vector de dimensión n ; $X_i|X_j, j \neq i$ es normal con media y varianza condicional (Gelfand et al., 2010).

$$\mathbf{E}(X_t|x_{-i}) = \mu_i + \sum_{j \neq i} \beta_{ij}(x_j - \mu_j)$$

$$\text{Var}(X_t|x_{-i}) = k_j^{-1} \tag{11}$$

Modelo Condicional Intrínseco Autorregresivo

Es un modelo CAR cuya correlación en las áreas depende solamente de ϕ y $\mathbf{W}$, dado que es difícil de estimar, la versión simplificada es la correlación de $\phi = 1$ (Assunção and Krainski, 2009), esto involucraría que la matriz de covarianza no es positiva(Blangiardo and Cameletti, 2015).

$$(u_i|u_j, j \neq i) \sim N(\mu_i + \sum_{j=1}^{n} a_{i,j}(u_j - \mu_j), S_i^2) \tag{12}$$

$\phi = 1$

$S_i^2 = \frac{\sigma^2}{N(i)}$, es la varianza del área que depende del número de vecinos

$a_{ij} = 1$ si las áreas i y j son vecinas

$N(i)$ es el número total de vecindarios

Modelo Autologístico

Asumiendo X_i, $i = 1, 2, \ldots, n$, como variables aleatorias binarias, con probabilidad de éxito $\pi_i(x-1) = E(X_i|x_i)$, especifica que la media condicional está dada por (Gelfand et al., 2010):

$$logit\pi_i(x_{-i}) = \mu_i + \sum_{j \in \delta i} \beta_{ij}(x_j - \mu_j) \tag{13}$$

Modelo Autopoisson

Asumiendo X_i, $i = 1, 2, \ldots, n$ son variables aleatoria con distribución Poisson, con media condicionada $\lambda_i(x-1) = E(X_i|x_i)$ (Gelfand et al., 2010).

$$log\lambda_i(x_{-i}) = \mu_i + \sum_{j \in \delta i} \beta_{ij}(x_j - \mu_j) \tag{14}$$

2.3.2. Modelo BYM

Este modelo está basado en los modelos CAR para efectos aleatorios espaciales, en donde se expresa relación en efectos de la misma área y las áreas vecinas. En donde el efecto de b_i esta dado por la suma de $\phi_i + \theta_i$. Donde ϕ_i es el efecto que tiene sobre el área todas las áreas estudiadas, mientras que para θ_i un pequeño grupo que tienen influencia sobre el área estudiada (Rodrigues and Assunção, 2012).

El modelo Besag, York y Mollié propone un modelo de log-risks como la suma de dos efectos aleatorios de diferentes matrices de covarianzas (Martínez-Beneito and Botella-Rocamora, 2019). Por lo tanto, la estructura de covarianza de los log-risk tienen dos parametos σ_Φ^2 y σ_ϕ^2, como los efectos aleatorios espaciales y temporales. En caso de que los datos fueran altamente independientes, los efectos aleatorios presentarían una alta varianza (Martínez-Beneito and Botella-Rocamora, 2019). El modelo Besag York Mollie está dado por:

$$y_i \backsim Poisson(E_i\phi_i)$$
$$\eta_i = log(\phi_i) = b_0 + u_i + v_i$$

En donde b_0 es la tasa promedio en las vencidades, E_i los casos esperados en cada area, u_i la estructura residual espacial y v_i el residuo no estructurado.

2.3.3. Regresión Ecológica

A diferencia del modelo BYM, la regresión ecologica posee covariables que permiten explicar mejor los casos o las tasas de cada vecindad

$$y_i \backsim Poisson(E_i\phi_i)$$
$$\eta_i = b_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + u_i + v_i$$

2.4. Modelos Lineales Generalizados

Previa a la definición de los modelos espaciales deben explicarse las regresiones base para la explicación de los comportamientos de enfermedades, usualmente ligadas a distribuciones binomiales y Poisson, esto se debe a que estas distribuciones pueden usarse para el conteo de eventos o explicación de tasas (Agresti, 2015). En el caso de los modelos lineales generalizados, se pueden identificar por dos componentes, en el caso de la variable respuesta debe ser miembro de una distribución de la familia exponencial y la función de enlace debe describir como la media de la respuesta y la combinación lineal de los predictores están relacionados.

2.4.1. Modelos Poisson

Teorema 1 (Myers et al., 2012)

Se asume que la variable respuesta y_i es un conteo para las observaciones $y_i = 0, 1, 2, 3...$ un modelo razonable puede ser:

$$f(y_i) = \frac{\exp^{-\mu_i}\mu_i^{y_i}}{y_i!}, y_i = 0, 1, 2, ...$$

Donde el parámetro $\mu_i > 0$. Donde la media y la varianza son iguales. Para el caso de este modelo se asume a la existencia de una función g, la cual será llamada la función de enlace:

$$g(\mu_i) = \eta_i = \beta_0 + \beta_1 x_1 + ... + \beta_{k1} x_k =$$

Sea $i = 1, ..., n$, para $y_i \sim Bernooulli(\pi_i)$, donde $\pi_i = \frac{exp(\mathbf{x_i^\intercal}\beta)}{1+exp(\mathbf{x_i^\intercal}\beta)}$, $y_1, ..., y_n$ son independientes:

la función $g(\pi) = log\left(\frac{\pi}{1-\pi}\right)$ es llamada función logit, esta tiene un intervalo entre (0,1) con dominio en los reales de $(-\infty, \infty)$.

$$\begin{aligned} g(\pi_i) &= log\left(\frac{\pi_i}{1-\pi_i}\right) \\ &= log(\frac{\frac{exp(\mathbf{x_i^\intercal}\beta)}{1+exp(\mathbf{x_i^\intercal}\beta)}}{\frac{1}{1+exp(\mathbf{x_i^\intercal}\beta)}}) \\ &= log(exp(\mathbf{x_i^\intercal}\beta)) = \mathbf{x_i^\intercal}\beta \end{aligned}$$

Dado esto podemos definir que la regresión logistica está definida como:

$$logit(\pi_i) = log\left(\frac{\pi_i}{1-\pi_i}\right) = \mathbf{x_i^\intercal}\beta$$

2.5. Campos Aleatorios de Markov Gaussianos

Los Campos Aleatorios de Markov Gaussianos o Gaussian Markov Random Fields (GMRF), consisten en vectores aleatorios x distribuidos de manera gaussiana, el cual obedece propiedades de independencia.

Un GMRF puede usarse como una representación de una estuctura condicional independiente por medio de un grafo indirecto. Un grafo $\zeta = (\upsilon, \epsilon)$, donde υ es un grupo de nodos del grafo y ϵ es un grupo de arcos $\{i, j\}$, donde $i, j \in \upsilon$ y $i \neq j$. Un grafo está totalmente conectado si $\{i, j\} \in \epsilon$ para todo $i, j \in \upsilon$ y $i \neq j$.

Sea $x = (x_1, ..., x_n)^T$ tiene una distribución normal con una media μ y una matriz de varianza y covarianza $\sum$. Definido por u grafo $\zeta = (\upsilon, \epsilon)$, donde $\upsilon = 1, ..., n$ ϵ sean los arcos entre los nodos i y j, para x_i y x_j donde $i \neq j$. La información se encuentra intrinseca en la matriz de covarianza $\sum$ pero al ser dificil de interpretar, la matriz de precisión permite la interpretabilidad de $Q = \sum^{-1}$.

Teorema sea x normalmente distribuido con media μ y una matriz de precisión Q. Entonces para $i \neq j$

$$x_i \perp x_j | x_{-(i,j)} \Longleftrightarrow Q_{i,j} = 0$$

Es posible afirmar que los patrones de no cero de Q determina ζ, lo que permite leer Q; sean x_i y x_j condicionalmente independientes. Esto es un vector aleatrorio $x = (x_1, ..., x_n)^T \in \mathbb{R}$, un GMRF con respecto al grafo $\zeta = (\upsilon, \epsilon)$, con media del vector μ y matriz de precisión $Q > 0$, mientras que la función de la densidad tendría la forma de:

$$\pi(x) = (2\pi)^{\frac{-\pi}{2}} |Q|^{\frac{1}{2}} \exp^{\frac{-1}{2}(x-\mu)^T Q(x-\mu)} \tag{15}$$

donde $Q_{ij} \neq 0 \Longleftrightarrow i, j \in \epsilon$ para todo $i \neq j$. El siguiente teorema justifica la condición de independencia

Teorema

Sea x un campo aleatorio de Markov Gaussiano con respecto a $\zeta = (\upsilon, \epsilon)$ con una media μ y una matriz de precisión $Q > 0$. Por lo tanto:

$$\begin{aligned} Prec(x_i | x_{-i}) &= Q_{ij} \\ E(x_i | x_{-i}) &= \mu_i - \frac{1}{Q_{ij}} \sum_{j:j\sim i} Q_{ij}(X_j - \mu_j) \\ Corr(x_i, x_j | x_{-ij}) &= -\frac{Q_{ij}}{\sqrt{Q_{ij} Q_{ij}}}, i \neq j \end{aligned}$$

Los elementos de la diagonal **Q** son las precisiones condicionales de x_i dado x_{-i}, mientras que los elementos fuera de la diagonal con una escala adecuada, proporciona información sobre la correlación condicional entre x_i y x_j, dado x_{-ij}. Estos resultados se deben comparar con la interpretación de los elementos de la matriz de covarianza $\sum = (\sum ij)$. (Niño Martínez, 2018)

$$\begin{aligned} Var(x_i) &= \sum ij \\ Corr(x_i, x_j | x_{-ij}) &= -\frac{\sum ij}{\sqrt{\sum_{ij} \sum_{ij}}} \end{aligned}$$

La matriz de covarianza proporciona información sobre la varianza marginal del x_i y la correlación marginal entre x_i y x_j.

2.5.1. Propiedades de markov de un GMRF

Propiedad de pares de Markov:

$x_i \perp x_j | x_{-ij}$ si $i, j \notin \varepsilon$ and $i \neq j$

Propiedad de Markov Local:

$x_i \perp x_{-i,ne(i)} | x_{ne(i)}$ para todo $i \in v$

Propiedad de Markov Global:

$x_A \perp x_B | x_C$
Para todos los conjuntos disjuntos A, B y C donde C separa A y B, y A y B son no vacíos de v.

Donde x_{-ij} se refiere a todos los elementos i y j, significa que todos los elementos x_i y x_j que no son vecinos en el grafo, son condicionalmente independientees dado el resto (Sidén and Lindsten, 2020).

La información provista por **Q** es de altísima relevancia desde el punto de vista espacial. Esta permite mostrar el comportamiento condicional de un fenomeno por medio de los resultados de su vecino, con esto como punto de referencia, se pueden asociar a los datos de área, los cuales buscan por medio de una dependencia espacial explicar un fenomeno bajo la estructura de vecindades. Si consideramos el vector (1, ..., N) representando un área por medio de los indices, en el cual un área i se encuentra rodeada de una cantidad finita de vecinos representados como $N(i)$, de primer o segundo orden según la distancia y el tipo de frontera que compartan.

Bajo la propiedad de Markov de que el parámetro $_i$ para la i-ésima área es independiente de todos los demás parámetros, dado el conjunto de sus vecinos $N(i)$, entonces $\theta_i \perp \theta_{-i} | \theta_{N(i)}$. El punto es que la matriz de precisión Q de θ es escasa. Una suposición simplificadora común es que el campo aleatorio es estacionario. Este ejemplo espacial es también conocido como modelo CAR (Conditional Autoregressive Model).

Definición: Un GMRF $x(s)$ con respecto a v es llamado estacionario si un vector **h** y una localización s_1, s_2 de la localización ζ, $\mu(s_1 + h) = \mu(s_1)$ y $Cov(x(s_1 + h), x(s_2 + h)) = Cov(x(s_1), x(s_2))$

2.6. Estadística Bayesiana

A diferencia del enfoque frecuentista busca una función de probabilidad, la estadística bayesiana se centra en la búsqueda de una probabilidad o una creencia, este conocimiento previo será extraído de datos relacionados, encuestas, estudios previos o conocimiento de expertos (Kruschke, 2014).

Esta rama de la estadistica fue rezagada durante el siglo XX por el enfoque frecuentista, especialmente ligado a los numerosos cálculos que son requeridos en comparación con su contraparte frecuentista, aunque en años recientes por el aumento de la capacidad de procesamiento de las computadoras, han permitido un resurgimiento y un aumento de las producciones académicas y proyectos.

El teorema de Bayes es el fundamento de toda esta rama de la estadística:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \tag{16}$$

En donde la $P(A)$, es la probabilidad a priori, $P(B|A)$ la verosimilitud asociada a la a priori y $P(A|B)$ es la probabilidad a posteriori. Teniendo en cuenta esto, el teorema puede ser reescrito no para concluir sobre probabilidades si no sobre distribuciones.

$$f(x|y) = \frac{f(y|x)f(x)}{f(y)}$$

Donde:

$$f(x|y) \propto f(x)f(y|x)$$

Al reescribir el teorema encontramos a $f(x)$ como una distribución a priori, esta compuesta de datos, informes anteriores, comparativos con procesos similares o información parametrizada, este es el principal foco de discusión con la estadística frecuentista, dado que los criterios de selección y subjetividad de los análisis pueden afectar a la probabilidad a posteriori

2.6.1. Criterios Bayesianos DIC y WAIC

Uno de los criterios de comparación de modelos en el marco de la estadística bayesiana es el criterio de variación(deviance) de información (DIC) (Spiegelhalter et al., 2002).

$$DIC = \overline{D(\theta)} + pD$$

Donde, $\overline{D(\theta)} = 2log(p(y|\theta))$ donde y son los datos, θ los parámetros desconocidos del modelo y $p(y|\theta)$ la función de maxima verosimilitud. Mientras que $pD = \frac{\overline{var(D(\theta))}}{2}$. D es la medida de ajuste del modelo, mientras pD se conoce como la efectividad del número de parámetros, es una de las medidas de complejidad. El DIC obtiene la deviance a través de la función de verosimilitud, el cual buscar una incorporación de la información a priori de los parámetros.

Por otra parte el WAIC (Watabane Akaike Information Criterium) se describe como.

$$WAIC = log(\frac{\sum_{n=1}^{N} p(y|\theta)}{N}) - 2\frac{\sum_{n=1}^{N} log(p(y|\theta))}{N}$$

2.6.2. Aproximación por Integración Anidada de Laplace (INLA)

Surge como una alternativa a los métodos MC (Monte Carlo), en donde buscamos una aproximación de tipo analítica con el método de Laplace. Suponiendo que se tiene un vector de parámetros con distribución a posteriori, del vector de tamaño n, se ha fraccionado para que dar como $n = (n_1, n_2)$ y $n_2 = (n_2, \ldots, n_p)$, la densidad a posteriori quedaría escrita como:

$$p(n|y) = p(n_1 n_2|y)$$

Al aplicar la expansión de Taylor de segundo orden al logaritmo de la a posteriori, es posible redondear el condicional para n_2, mientras se maximiza el ajuste de $\hat{n}$ por medio de la función $p(n_1 n_2|y)$ (Rachev et al., 2008)

$$log(p(n_1, n_2|y)) \approx log(p(n_1, \hat{n_2})|y)) + \frac{dlog(p(n_1, n_2|y))}{dn_2} + \frac{1}{2}(n_1 - \hat{n_2})\mathbf{H}(n_1 - \hat{n_2}) \tag{17}$$

Dado que el segundo termino a la derecha es igual a cero y **H** es la matriz Hessiana, la densidad a posteriori de n_1 obtenida sería

$$\begin{aligned}
p(n_1, y) &= \int p(n_1, n_2|y)dn_2 \\
p(n_1, y) &= \int exp(log(p(n_1, n_2|y)))dn_2 \\
p(n_1, y) &= \int exp(log(p(n_1, \hat{n_2}|y))) - \frac{1}{2}(n_1 - \hat{n_2})H(n_1 - \hat{n_2}))dn_2 \\
&= p(n_1, \hat{n_2}|y) \int -\frac{1}{2}(n_1 - \hat{n_2})(-\mathbf{H})(n_1 - \hat{n_2})dn_2 \\
&\propto p(n_1, \hat{n_2}|y)|-\mathbf{H}^{-1}|^{\frac{-1}{2}}
\end{aligned}$$

Donde la última línea es conocida como el kernel de una distribución normal multivariada

2.6.3. Espacios Gaussianos latentes

Consiste en especificar una distribución para y_i que es caracterizada por un parámetro ϕ_i definida por una función de predictor estructurada, a través de una función de enlace g, donde $g(\phi_i) = n_i$, mientras que el predictor líneal está definido como:

$$n_i = \beta_0 + \sum_{m=1}^{M} \beta_m x_{mi} + \sum_{l=1}^{L} f_i(z_{ij}) \tag{18}$$

Una ventaja que tienen los modelos gaussianos latentes es su flexibilidad, dado que estos pueden adaptarse con facilidad a modelos generalizados, como a modelos dinámicos.

Haciendo una colección de todos los componentes latentes de interés para la inferencia en un grupo de parámetros θ, definido como $\theta = (\beta_0, B, f)$. Denotado como $\mathbf{\Psi = \Psi_1..\Psi_k}$ del vector de hiperpárametros k.que se asume independencia en la distribución de n observaciones dado por la máxima verosimilitud, donde cada punto y_i esta conectado a solo un elemento θ_i en un campo lantente θ (Blangiardo and Cameletti, 2015).

$$p(y|\theta, \Psi) = \prod_{i=1}^{n} p(y_i|\theta_i, \Psi)$$

Donde cada punto y_i esta conectado solo en un elemento θ_i en un espacio latente θ. Si asumimos una a priori Normal multivariada en θ con vector de media 0 y una matriz de precisión $Q(\Psi)$, con una función de densidad dado (Blangiardo and Cameletti, 2015):

$$p(y|\theta, \Phi) = 2\pi^{-\frac{n}{2}}|Q(\Psi)|^{\frac{1}{2}}exp(\frac{-1}{2}\theta^{|}Q(\Psi)\theta)$$

La función posterior resultante de θ y Ψ está dado por el producto de la máxima verosimilitud, calculado previamente:

$$
\begin{aligned}
p(\theta, \Psi|y) &\propto p(\Psi) \times p(\theta|\Psi) \times p(y|\theta, \Psi) \\
&\propto p(\Psi) \times p(\theta|\Psi) \times \prod_{i=1}^{n} p(y_i|\theta_i, \Psi) \\
&\propto p(\Psi) \times |Q(\Psi)|^{\frac{1}{2}} \times exp(2\pi^{-\frac{n}{2}}|Q(\Psi)|^{\frac{1}{2}}) \exp \frac{-1}{2}\theta|Q(\Psi)\theta + \sum_{i=1}^{n} log(p(y_i|\theta_i, \Psi))
\end{aligned}
$$

2.6.4. Aproximación a la inferencia bayesiana con INLA

La inferencia bayesiana tiene como fin encontrar las distribuciones posteriores para cada elemento del vector de parámetros y para cada vector de hiperparámetros.(Blangiardo and Cameletti, 2015)

$$
\begin{aligned}
p(\theta_i, y) = \int p(\theta_i, \Psi|y)d\Psi = \int p(\theta_i|\Psi, y)p(\Psi|y)d\Psi \\
p(\Psi_k|y) = \int p(\Psi|y)\Psi_{-k}
\end{aligned}
\tag{19}
$$

En el método la aproximación de la integral anidada de laplace asume que el modelo producirá una aproximación numérica de la posterior basada en la aproximación de método del Laplace, en donde realizará dos actividades para su cálculo, la primera el cálculo de $p(\Psi|y)$ como todas las marginales relevantes de $p(\Psi_k|y)$ y la segunda calcular los parámetros marginales de $p(\theta_i, y)$, como el calculo de los parámetros marginales de la distribución a posteriori $p(\theta_i|y)$.

Para la primera se asume que $\hat{p}(\theta_i, \Psi|y)$ como una aproximación Gaussiana dada por el método de Laplace, de $p(\theta_i, \Psi|y)$ y $\theta^*(y)$ para todo Ψ dado. La aproximación gaussiana se acerca a la real pues $p(\theta|\Psi, y)$es casi gaussiana con una distribución a priori de Campo Aleatorio de Markov Gaussianos (Gaussian Markov Random Fields) (GMFR) que puede resumirse como un vector aleatorio que sigue una distribución normal multivariante(Rue and Held, 2005), mientras que y es no informativa y presenta observaciones bien comportadas de la distribución.

$$
\begin{aligned}
p(\Psi|y) &= \frac{p(\theta, \Psi|y)}{p(\theta|\Psi, y)} \\
&= \frac{p(y|\theta, \Psi)p(\theta, \Psi)}{p(y)} \frac{1}{p(\theta|\Psi, y)} \\
&= \frac{p(y|\theta, \Psi)p(\theta, \Psi)p(\Psi)}{p(y)} \frac{1}{p(\theta|\Psi, y)} \\
&\propto \frac{p(y|\theta, \Psi)p(\theta|\Psi)p(\Psi)}{p(\theta|\Psi, y)} \\
&\approx \left. \frac{p(y|\theta, \Psi)p(\theta|\Psi)p(\Psi)}{\tilde{(p)}(\theta|\Psi, y)} \right|_{\theta=\theta^*(\Psi)} =: \tilde{(p)}(\Psi|y)
\end{aligned}
$$

La segunda tarea es mas compleja computacionalmente, dado que tiene dos caminos posibles, en la primera opción, calcular directamente las posteriores a través de la descomposición de Cholesky usando la matriz de precisión, una segunda posibilidad, es reescribir el vector de parámetros como $\theta = (\theta_i, \theta_{-i})$, usando la aproximación de Laplace se obtiene:

$$
\begin{aligned}
p(\theta_i|\Psi, y) &= \frac{p((\theta_i, \theta{-}i)|\Psi, y)}{p(\theta_{-i}|\theta i, \Psi, y)} \\
&= \frac{p(\theta, \Psi|y)}{p(\Psi|y)} \frac{1}{p(\theta_{-i})|\theta_i, \Psi, y} \\
&\propto \frac{p(\theta, \Psi|y)}{p(\Theta_{-i}|\theta i, \Psi, y)} \\
&\approx \frac{p(\theta, \Psi|y)}{\tilde{(p)}(\Theta_{-i}|\theta i, \Psi, y)}\Bigg|_{\theta_{-i}=\theta^*_{-i}(\theta_i,\Psi)} =: \hat{p}(\theta_i|\Psi, y)
\end{aligned}
$$

Donde $\tilde{(p)}(\theta_{-i}|\theta_i, \Psi_i, y)$ es la aproximación de Laplace de $p(\theta_{-i}|\theta_i, \Psi, y)$ y $\theta^*_{-i}(\theta_i, \Psi)$

2.7. Modelo Espacio Temporal

Al considerar n localizaciones en T puntos, se genera un componente espacio temporal de la siguiente manera,

$$\eta_{it} = b_0 + u_i + v_i + (\beta + \delta_i) \times t$$

En donde b_0 es la tendencia lineal, δ_i efecto diferencial, u_i y v_i errores del modelo y t el tiempo. Al expandir el modelo anterior, puede explicarse las tasas o probabilidades buscadas, mediante la interacción de las variables o los efectos, presentando el modelo de la siguiente manera:

$$\eta_{it} = b_0 + v_i + u_i + \varphi_t + \phi_t + \delta_{it}$$

Tabla 1: Tipos de interacciones para modelos espacio temporalales. Fuente: Blangiardo-Cameletti

Interacción	Interacción de parametros	Grado
I	v_i y ϕ_i	nT
II	v_i y φ_i	n(T-1) for RW1 o n(T-2) for RW2
III	ϕ_i y u_i	(n-1)T
III	u_i y φi	(n-1)(T-1) for RW1 o (n-1)(T-2) for RW2

Utilizando el paquete de INLA, se evaluaron 36 modelos con combinaciones diferentes de variables, rezagos y distribuciones.

Evaluación modelos espacio temporales univariados

Como se mencionó anteriormente, los modelos espacio temporales están determinados por la estructura

$$n_{ijy} = b_0 + u_i + v_j + Tt$$

Se generaron modelos univariados como un primer ejercicio como punto de partida y de comparación con el modelo bivariado, en el caso de describir la tasa de contagio, se generaron los siguientes modelos:

Modelos Latentes y Distribuciones A priori

Modelo BYM 2

El modelo BY2 es una reparametrización del modelo BYM, generando una combinación entre el modelos besag v* y el modelo iid u*.

$$x = \begin{pmatrix} v^* + u^* \\ u_* \end{pmatrix}$$

Donde ambos modelos tienen una precisión de hiperparámetros. La longitud de x es 2n si la longitud de u^* y v^* es n. El modelo BYM2 usa diferentes parametrizaciones donde u y v están estandarizadas para tener varianzas igual a 1. En donde τ es la precisión marginal y ϕ explica la varianza por el efecto espacial o parámetro mezclado.

$$x = \begin{pmatrix} \frac{1}{\sqrt{\tau}}(\sqrt{1-\phi}v + \sqrt{\phi}u) \\ u \end{pmatrix}$$

Para el caso de este modelo, la distribución a priori está definida como $\theta = (\theta_1, \theta_2)$:

$$\theta_1 = log(\tau)$$

$$\theta_2 = \log \begin{pmatrix} \phi \\ 1-\phi \end{pmatrix}$$

Caminata aleatoria de Orden 2

Una caminata aleatoria de orden 2 RW2 para el vector gaussiano $x = (x_1, ..., x_n)$ está construido bajo el supuesto independiente para el crecimiento de segundo orden.

$$\Delta^2 x_i = x_i - 2x_{x+1} + x_{i+2} \sim N(0, \tau^{-1})$$

La densidad de x esta diferenciada desde su n-2 orden según su incremento

$$\pi(x|\tau) \propto \tau^{(n-2)/2} \exp \frac{-\tau}{2} \sum \Delta^2 x_i^2 = \tau^{(n-2)/2} \exp \frac{-1}{2} x^T Q x$$

Donde $Q = \tau R$ y R es una matriz estructural reflejada de la estructura de los vecindarios del modelo. La distribución a priori está definida como θ.

$$\theta = \log \tau$$

En donde τ es el parámetro de precisión.

Modelo iid

El modelo Normal Wishart de 2 dimensiones es usado si se definen dos vectores de efecto aleatorio, y y v para los cuales (u_i, v_i) son iid bivariados válidos.

$$\begin{pmatrix} u_i \\ v_i \end{pmatrix} \propto N(0, W^{-1})$$

En donde la matriz de covarianza $\mathbf{W^{-1}}$ es: θ

$$\begin{pmatrix} 1/\tau_n & \theta\Phi/\sqrt{\tau_a \tau b} \\ \Phi/\sqrt{\tau_a \tau_b} & 1\tau_b \end{pmatrix} \tag{20}$$

Donde τ_a, τ_b son la precisión marginal y es el coeficiente de correlación. Para el modelo la matriz de precisión **W** es distribuida Wishart.

$$\mathbf{W} \sim Wishart_p(r, R^{-1}), p = 2$$

Con densidad

$$\pi(W) = c^{-1}|W|^{(r-(p+1))/2} \exp \frac{-1}{2} Trace(WR), r > p + 1$$

y

$$c = 2^{(rp)/2}|R|^{-r/2}\pi^{p(p-1)/4} \prod_{j=1}^{p} \Gamma((r + 1 - j)/2)$$

entonces,

$$E(W) = rR^{-1}, yE(W^{-1}) = R/(r - (p + 1))$$

Aunque la librería puede usarse hasta los efectos de los casos del 1-5, en el trabajo actual este está implementado de grado 1, para este caso la precisión marginal está definida por:

$R = [R_{11}]$

$\theta = log\tau_1$ con un parámetro

A priori no informativa

Estas son usadas en los casos que no se posee información de los parámetros de interés, de este modo toda información de la distribuación a posteriori es proporcionada a través de los datos a través de función de verosimilitud. (Trilleras Martínez, 2018). En caso de presentarse este escenario, la distribusión a priori debe ser plana, para asignar una probabilidad igual para todos los valores posibles. Una de las distribuciones no informativas mas usadas es la distribución a priodi de Jeffreys (1946) (Blangiardo and Cameletti, 2015), dada por:

$$\pi(\theta) \propto |I(\theta)|^{\frac{1}{2}}$$

Donde $I(\theta)$ corresponde al valor esperado de la matriz de información de Fisher. En el caso multivariado para el $ij - ésimo$ elemento (Trilleras Martínez, 2018).

$$\pi(\theta) \propto I_{ij}(\theta) = -E_{Y|\theta}(\frac{\partial^2 ln(L(\theta; y))}{\partial\theta_i\partial\theta_j})$$

A priori débilmente informativa

Una de las mayores criticas que se ha presentado al estadística bayesiana es el uso inapropiado de la información o de las distribuciones, en donde se argumentando que pueden influenciar los datos.Tradicionalmente las mejores alternativas se encontraba en una función Gamma, con parámetros iguales y pequeños (Gelman, 2006).

2.8. Aproximación gaussiana para $\pi(\mathbf{x}|\mathbf{y}, \phi)$ en $\mathbf{x}_\phi^{\mathbf{mode}}$

Con la aproximación gaussiana para $\pi(x|y, \phi)$ con $\pi(x|\phi)$ es un Campo Aleatorio Gaussiano de Markov con una matriz de precisión de estructura no cero como distribución previa. Se asumen datos mutuamente independientes dados por un campo latente x (Steinsland, 2003), que puede ser escrito como:

$$
\begin{aligned}
\pi(y|x) &\propto \prod_{i=1}^{N} \exp g(x_i) \\
\pi(x|y,\phi) &\propto \pi(x|\phi)\pi(y|x) \\
&\propto \exp \frac{-1}{2}\tau x^T \mathbf{Q} x + \sum_{i=1}^{N} g(x_i) \\
&\propto \exp \frac{-1}{2}\tau x^T \mathbf{Q} x + \sum_{i=1}^{N} a_i + b_i x_i + c_i x_i^2 \\
&\propto \exp \frac{-1}{2} x^T (\tau \mathbf{Q} + \mathbf{diag}(\mathbf{c}))c + \sum_{i=1}^{N} (a_i b_i x_i)
\end{aligned}
\tag{21}
$$

donde $a_i + b_i x_i + c_i x_i^2$ es la segunda aproximación de taylor para $g(x_i)$ en x_ϕ^{mode} y una $\mathbf{diag}(\mathbf{c})$es la matriz diagonal que contiene $(c_1, c_2, ..., c_n)$. Puede verse que la aploximación de la matriz de presición es $(\tau\mathbf{Q} + \mathbf{diag}(\mathbf{c}))$ y dicha aproximación es un campo gaussiano aleatorio de markov con la misma estructura condicional de la distribución previa $\pi(x|\phi)$.

2.9. COVID-19

El SARS-CoV-2 tuvo como su epicentro la provincia de Hubei en la República Popular China. Desde este punto se esparció por todo el mundo y para el 30 de enero del 2020, el Comité de Emergencia de la OMS declaró una emergencia sanitaria global. Esto basándose en el crecimiento de los casos en China y locaciones internacionales dado que fue seguida en tiempo real por el sitio web provisto por la Universidad Johns Hopkins (Velavan and Meyer, 2020).

Los coronavirus son virus largos monocatenarios positivas que infectan a los humanos y a un gran número de animales. El primero de ellos fue descrito por Tyrell y Bunoe en 1966, quienes cultivaron el virus proveniente de pacientes con resfriados comunes dándole el nombre de coronavirus. Esto basado en su morfología viral esférica con un núcleo en todo el centro y una superficie protectora, semejante a la presente en la corona solar. Estos virus están clasificados en cuatro familias alfa y beta, asociadas a mamíferos, mientras que las gama y delta son originarias de cerdos y aves. Solo 7 subtipos de coronavirus pueden infectar al ser humano, para el caso de los betas coronavirus pueden provocar enfermedad y muertes, categoría en la que se encuentra clasificado el SARS-COV-2, mientras que los alfa coronavirus son asintomáticos o medianamente infecciosos.

Aparentemente el SARS-COV-2 fue transmitido de animales a humanos en el mercado de comidas marina de Huanan en Wuhan, China. Los síntomas iniciales que permiten su detección están asociados a la neumonía, aunque existen reportes que describen síntomas gastrointestinales e infecciones asintomáticas especialmente en pacientes menores a los 6 años. El tiempo medio de incubación se encuentra en 5 días, una mediana de 3 días encontrándose en un rango de entre 0-24 días. Los pacientes infectados manifiestan fiebres constantes, tos, congestión nasal, fatiga e infección en el tracto respiratorio que puede progresar de manera agresiva, generando ahogamiento y síntomas torácicos graves correspondientes en un 75 % a neumonías. Los síntomas persistentes de la neumonía viral incluyen diminución de la saturación de oxígeno, desviación de gases en la sangre, cambios visibles a través de radiografías de tórax.

Se estima que la actual pandemia puede duplicar el número de infectados al doble cada 7 días y que cada paciente puede infectar en promedio a 2.2 o 3.58 personas. Varios estudios indican que los pacientes mayores a 60 años son la población con más alto riesgo comparado con los niños de (edad) que tienen menor probabilidad de ser infectados para los cuales el 2.2 % de los casos positivos pueden llegar a ser fatales.

Entre los estudios de modelos espacio temporales aplicados en Colombia, se encuentra la investigación realizada para la descripción del primer mes de epidemia en la ciudad de Cali, Valle del Cauca. Allí se realizó un análisis de densidad de Kernel y por medio de la función K de Ripley se verificó la existencia de patrones, (Elías-Cuartas et al., 2020). Durante el proceso de investigación de este trabajo se evidencia como los casos empiezan a trasladarse de sectores sociales con mayores ingresos y a lo largo de tiempo la cantidad de casos empiezan a trasladarse a los sectores de la ciudad con más bajo nivel socioeconómico. Otro ejemplo se dio en Pereira, Risaralda, con un enfoque totalmente diferente, mediante la aplicación de un modelo SEIR (Susceptibles, Expuestos, Infectados y Recuperados) modificado. Dentro de este se incluyeron restricciones basadas en la población que estaban en confinamiento que generaron un modelo de proyección espacio temporal.

Cabe resaltar que los escenarios propuestos, se muestra el impacto de un confinamiento más estricto frente a las medidas que se tenían en el momento de construcción del modelo (Echeverri et al., 2020). Otra técnica estudiada en Ecuador, se buscaron clústers espacios temporales en donde demostraron que la zona comercial de Guayas, al tener un gran flujo de población, fue uno de los puntos nodales para la expansión del COVID en ese país (Ballesteros et al., 2021). Se generó una investigación de patrones y predicciones t+30 para el mes de abril de 2021, que incluía técnicas de minería de texto, usando trinos para la búsqueda de patrones en las regiones, de cómo movimientos o tendencias de trabajo en casa afectaban la cantidad de casos agregados de manera regional (Tariq et al., 2021).

Por otro lado, un enfoque diferente para pronósticos mediante modelo SIR, fue realizado de manera diaria, tomando 4 fechas entre t+4, t+18, t+30 , t+60 como puntos evaluación de las predicciones de este contempló aspectos del virus como la estabilidad, la susceptibilidad y la inmunidad (Manrique Abril et al., 2020). El modelo que más se acerca a la investigación del presente trabajo fue realizado en Toronto, en el cual se desarrolló un modelo aditivo espacio temporal realizado de manera agregada, comparando cuatro modelos y el ajuste de los modelos logit con función de enlace probit, Poisson, ninomial negativa y cero inflado (Feng, 2021).

3. Metodología

3.1. Fuentes de información

En la búsqueda de las tasas de contagios y de muertes, las fuentes principales de información son el ministerio de salud a través de la plataforma de datos abiertos, en donde se encuentran la información diaria de contagios reportados, pruebas realizadas y fallecimientos. Para el caso de la tasa de contagio esta se comparará sobre los datos poblacionales para el año 2020 del Departamento Administrativo Nacional de Estadísticas (DANE).

Pruebas https://www.datos.gov.co/Salud-y-Protecci-n-Social/Pruebas-PCR-procesadas-de-COVID-19-en-Colombia-Dep/8835-5baf
Casos https://www.datos.gov.co/Salud-y-Protecci-n-Social/Casos-positivos-de-COVID-19-en-Colombia/gt2j-8ykr/data

Con excepción de las pruebas, para los modelos de contagios fue necesario realizar acumulados para el modelo planteado, esta acumulación se realizó por fecha y departamento, generando un conjunto de datos al cual nos referiremos como dataset con cerca de 8000 puntos espacio temporales. En el dataset final, se encontrara compuesto de las variables fecha, periodos y departamentos categorizados, población, casos, recuperados, muertes del día y sus respectivas acumulaciones y la pruebas acumuladas.

Dado el consumo de procesamiento que requiere cada modelo, el tiempo para el análisis se encuentra entre el periodo de junio de 2020 a enero de 2021. Como punto de partida 33 vecindarios, 32 correspon-

dientes a los departamentos administrativos de Colombia y Bogotá como un distrito especial, dado que su densidad poblacional es una de las más altas del país.

3.2. Análisis Exploratorio

Por medio de la matriz de pesos promedio, podemos conocer que se tiene en promedio 4.5 fronteras por cada departamento, excluyendo el caso de San Andrés que por su condición insular no tiene fronteras físicas con ningún otro departamento. Complementando esta información con la matriz de pesos, encontramos que el departamento con el mayor número de fronteras es Cundinamarca, naturalmente con 7 departamentos, pero al incluir Bogotá como región tendría un peso de 0.125 y con un peso medio de 0.2162.

Figura 2: Mapa vecindades, Fuente:Autoría Propia

Tasa de mortalidad

Analizando los mapas en diferentes meses, se evidencia que las zonas con índices más altos de mortalidad están en Santander, Nariño y la región Caribe. Se destaca que todos son puntos fronterizos tanto del flujo de personas como de comercio con Ecuador y Venezuela, por parte de la región caribe, al ser un puerto marítimo. En menor medida, otro clúster que se puede encontrar es la región cafetera, en donde Caldas, Risaralda y adicional Huila y Putumayo presentan valores similares de tasas de mortalidad.

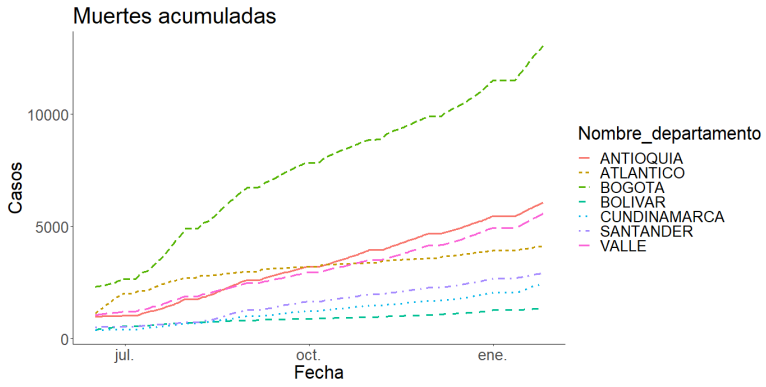

Figura 3: Muertes acumuladas Acumulados, Fuente:Autoría Propia

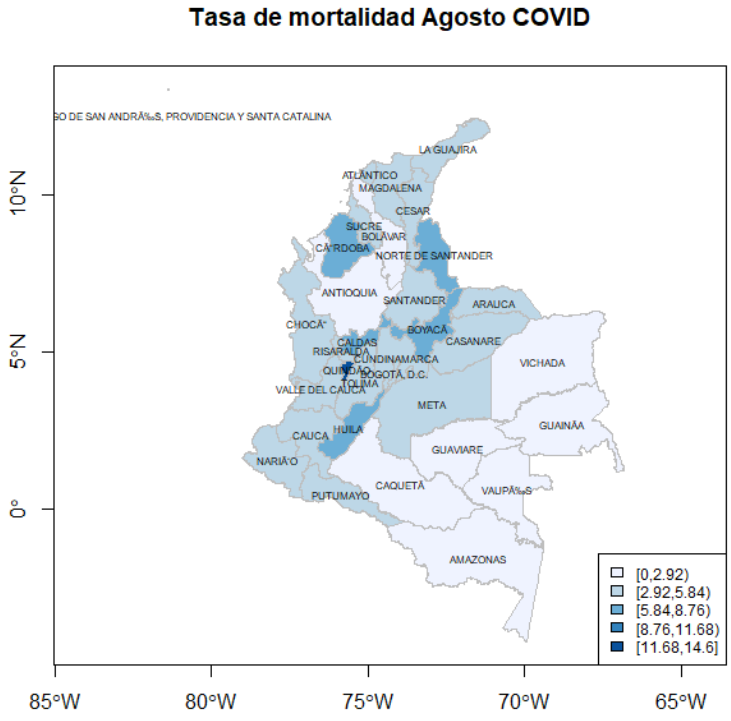

Figura 4: Tasa de Mortalidad Agosto por departamentos, Fuente:Autoría Propia

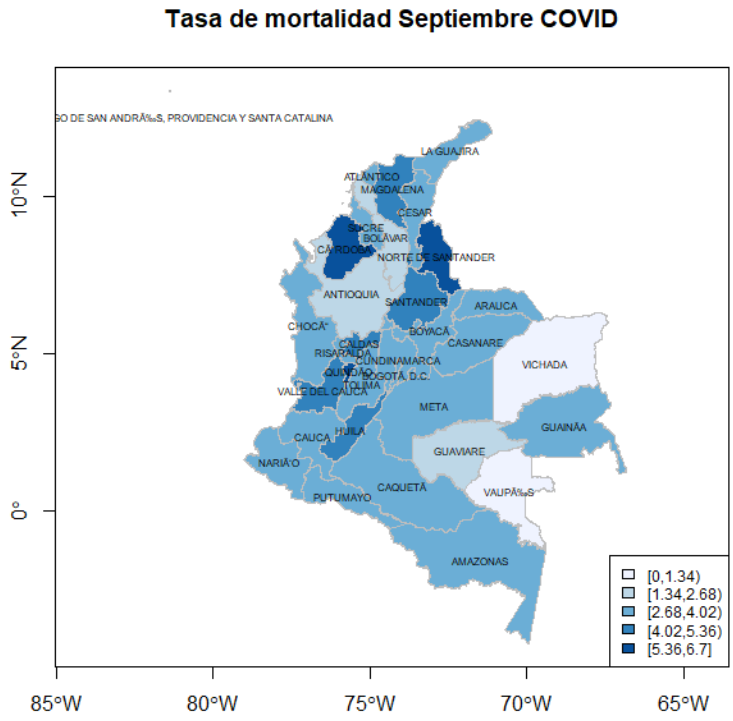

Figura 5: Tasa de Mortalidad Septiembre por departamentos, Fuente:Autoría Propia

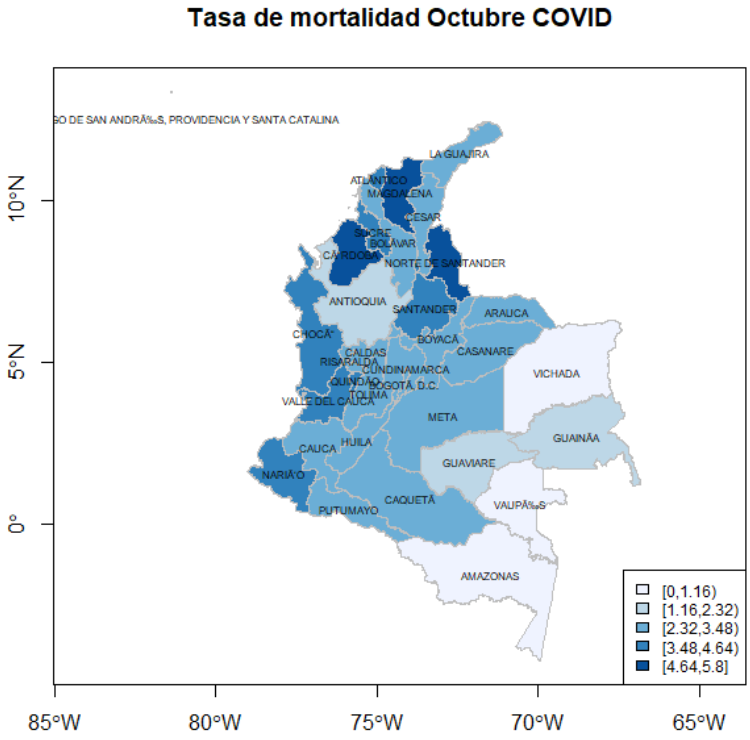

Figura 6: Tasa de Mortalidad Octubre por departamentos, Fuente:Autoría Propia

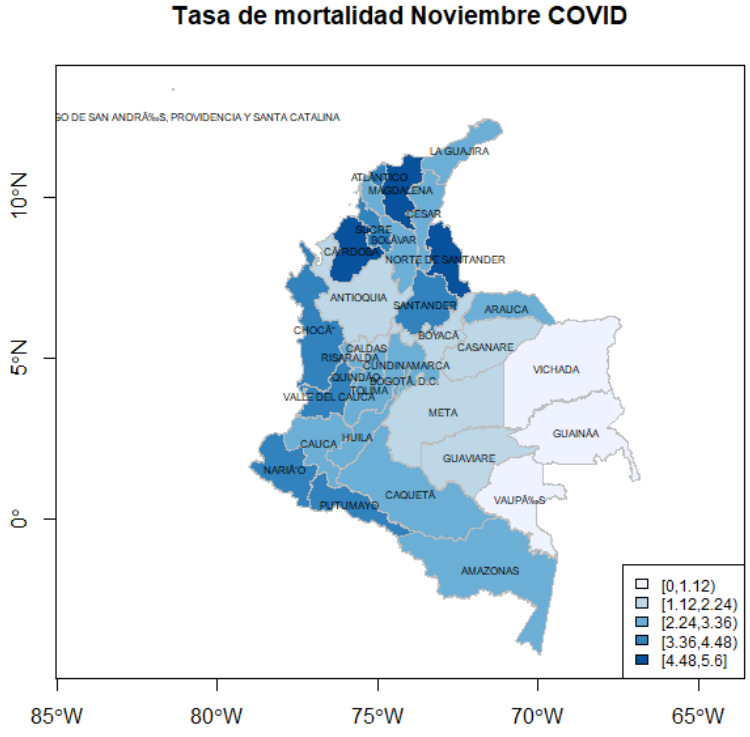

Figura 7: Tasa de Mortalidad Noviembre por departamentos, Fuente:Autoría Propia

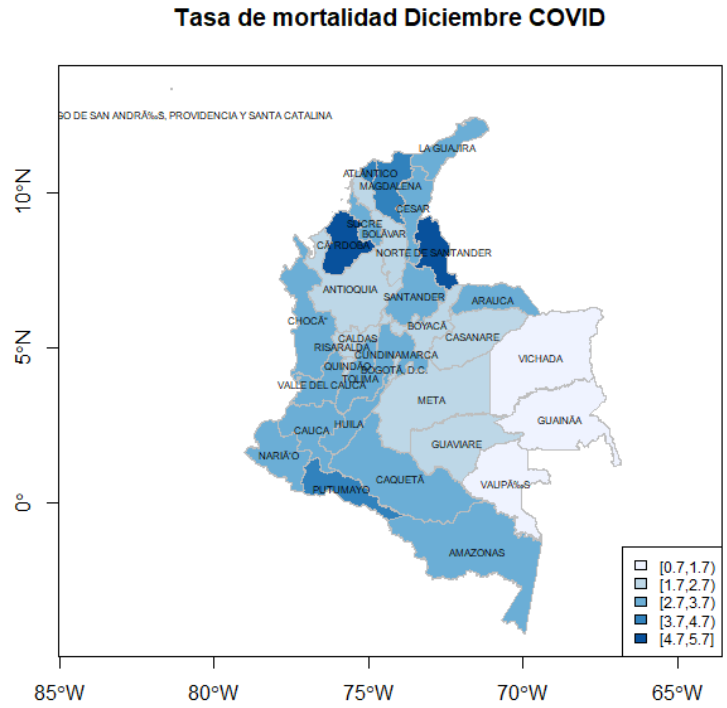

Figura 8: Tasa de Mortalidad Diciembre por departamentos, Fuente:Autoría Propia

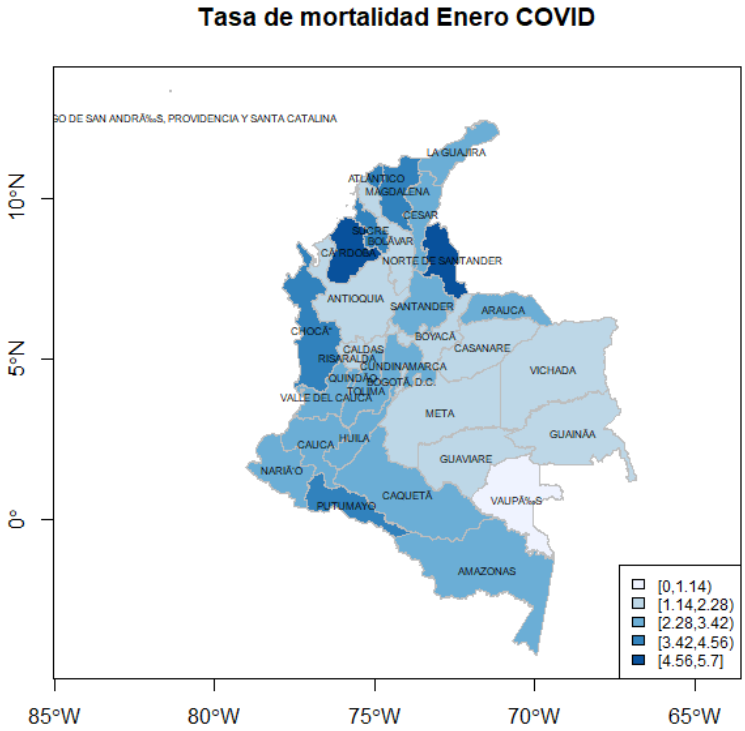

Figura 9: Tasa de Mortalidad Enero por departamentos, Fuente:Autoría Propia

De los departamentos analizados, se puede evidenciar que los departamentos de Valle, Santander y Atlántico son los que tienen una tasa más alta de mortalidad comparada con las regiones centrales. Esta tasa de mortalidad puede asociarse a las diferencias en zonas fronterizas o movilización de población flotante.

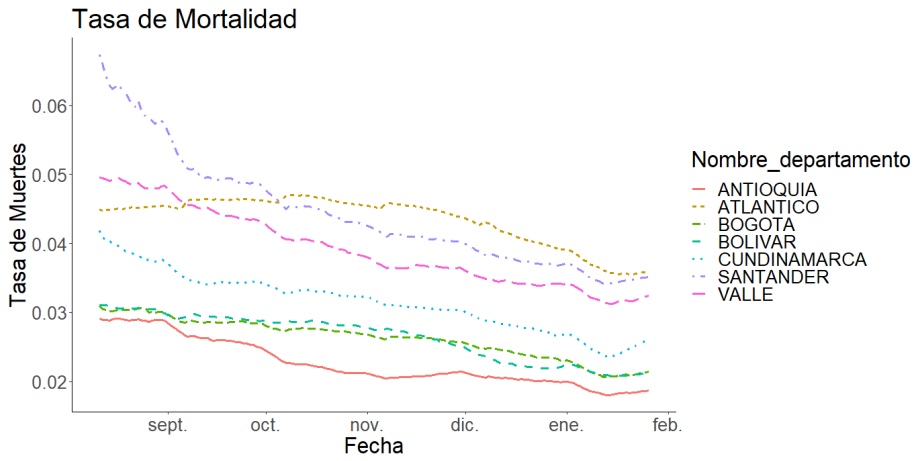

Figura 10: Tasa de Mortalidad con el indice de mortalidad más alto, Fuente:Autoría Propia

Tasa de Contagios

Las tasas de contagios en especial en Bogotá se han elevado, esto se debe a la densidad poblacional en donde tienen aproximadamente un 20 % de la población total del país, en contraste al departamento de Cundinamarca

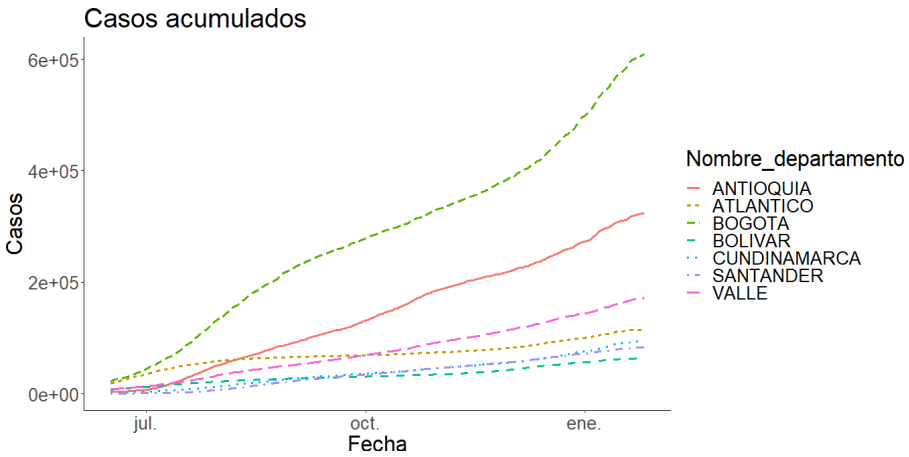

Figura 11: Casos Acumulados, Fuente:Autoría Propia

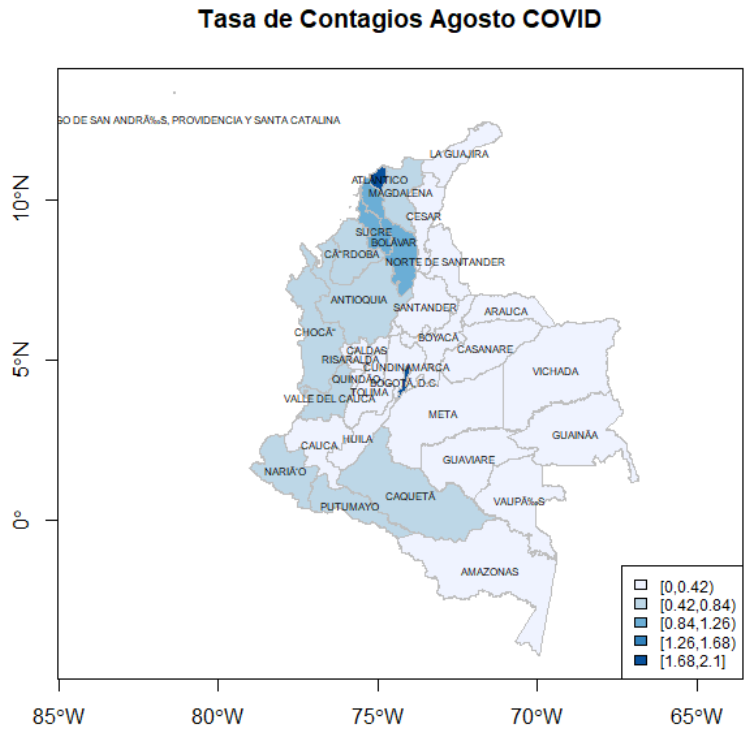

Figura 12: Tasa de Contagios Agosto por departamentos, Fuente:Autoría Propia

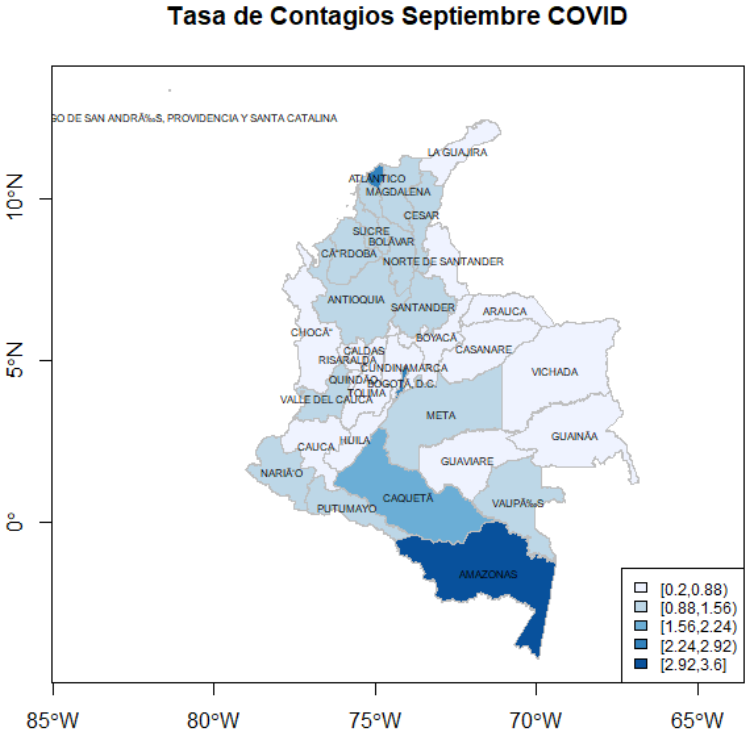

Figura 13: Tasa de Contagios Septiembre por departamentos, Fuente:Autoría Propia

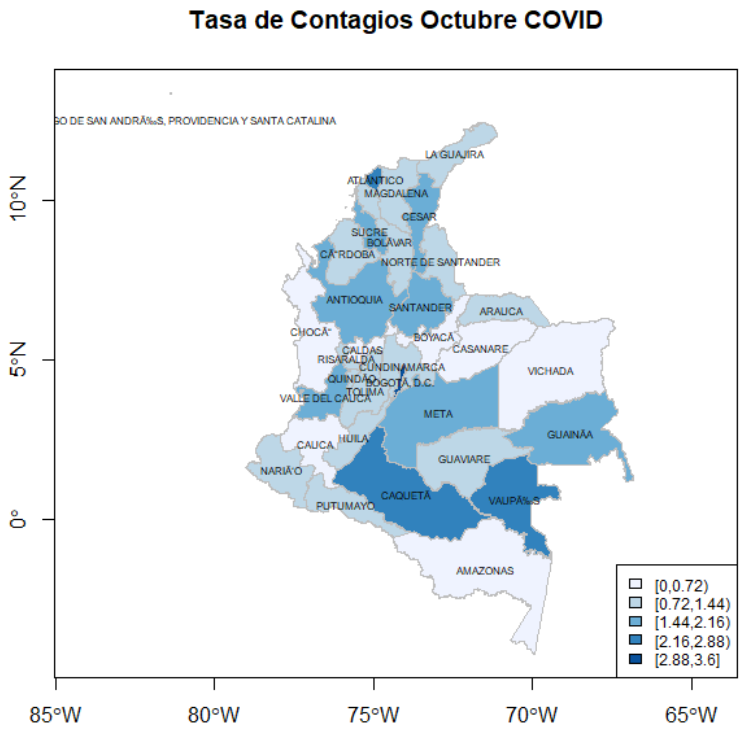

Figura 14: Tasa de Contagios Octubre por departamentos, Fuente:Autoría Propia

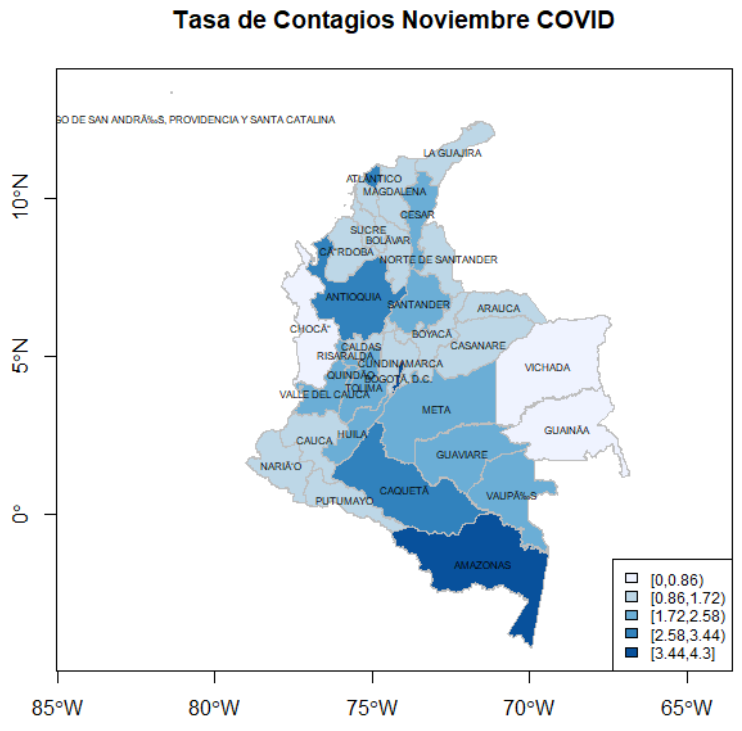

Figura 15: Tasa de Contagios Noviembre por departamentos, Fuente:Autoría Propia

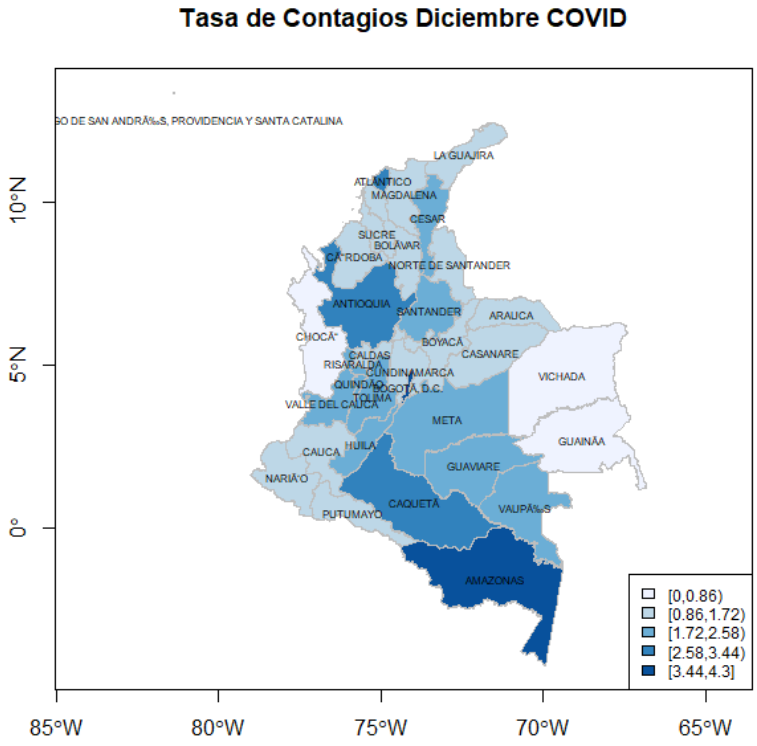

Figura 16: Tasa de Contagios Diciembre por departamentos, Fuente:Autoría Propia

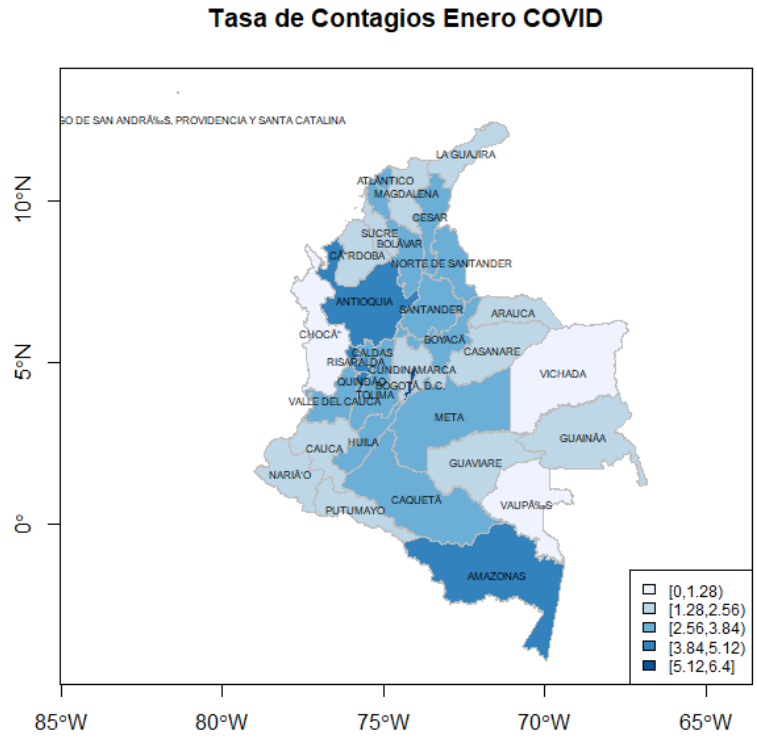

Figura 17: Tasa de Contagios Enero por departamentos, Fuente:Autoría Propia

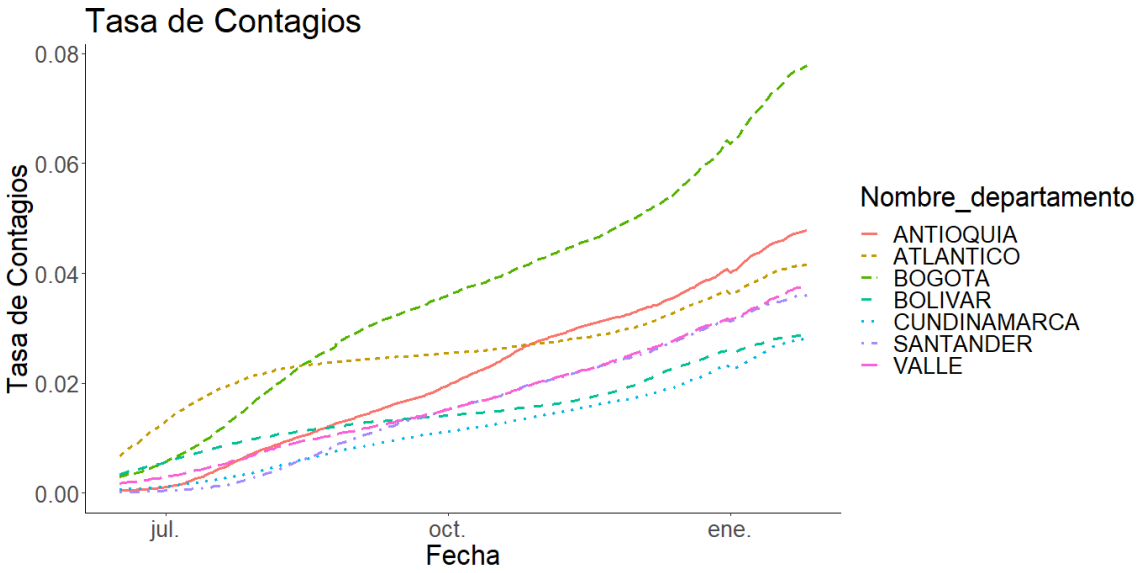

Figura 18: Tasa de Contagios departamento con tasas de contagios mayores, Fuente:Autoría Propia

Mapas relativos de contagio

Con respecto a los mapas relativos, se evidencia que la zona centro es la que posee junto con el departamento del valle el mayor número de contagios, estos aumentos en los indicadores pueden ser explicados por la influencia de Bogotá. Una de las posibles causas por parte de los residentes de la capital, en un cambio de condiciones laborales al acelerarse el proceso de teletrabajo, generaron un éxodo de población no autóctona de la ciudad a otras regiones del país, al igual que aperturas económicas y departamentales que permitieron viajes durante puentes, semanas de vacaciones o fines de semana. Esta sería la causa del aumento de la cantidad de casos esperados en esas poblaciones locales.(Gutierrez, 2021)

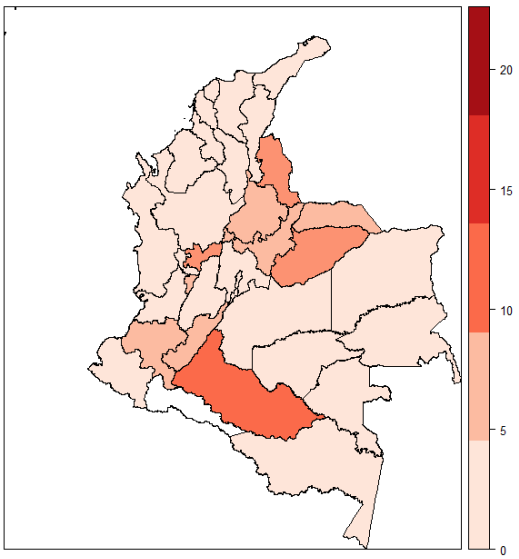

Figura 19: Mapas Relativos de Contagio Julio por Departamentos, Fuente:Autoría Propia

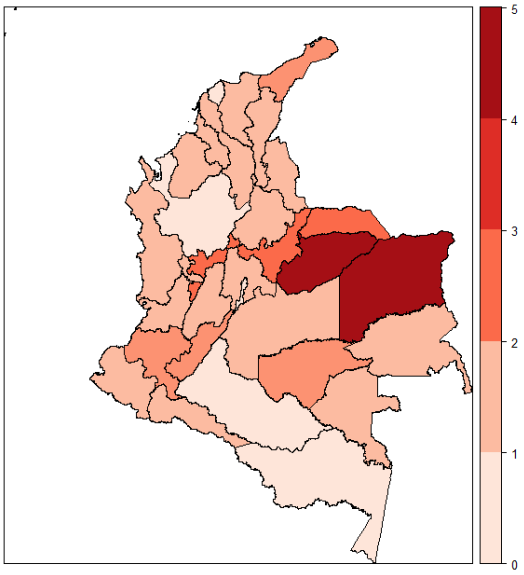

Figura 20: Mapas Relativos de Contagio Septiembre por Departamentos, Fuente:Autoría Propia

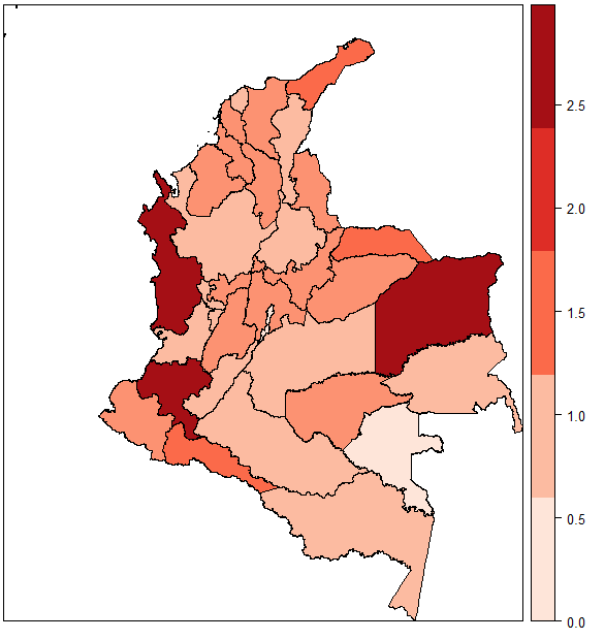

Figura 21: Mapas Relativos de Contagio Noviembre por Departamentos, Fuente:Autoría Propia

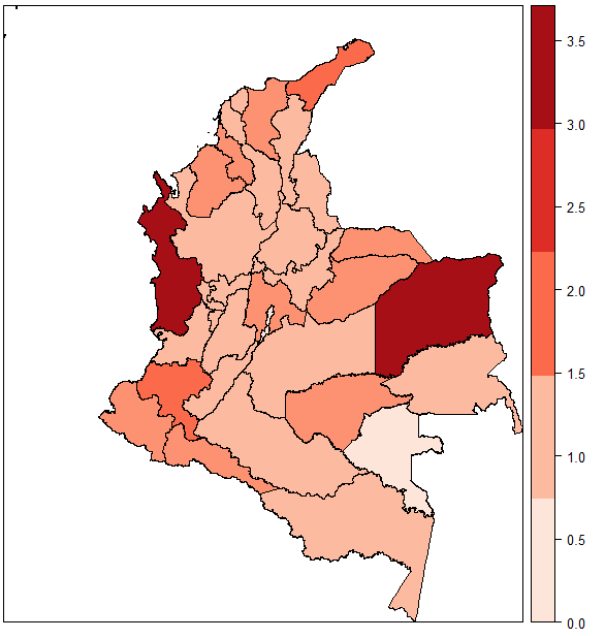

Figura 22: Mapas Relativos de Contagio Enero por Departamentos, Fuente:Autoría Propia

Mapas relativos de muertes

Este mapa se encuentra relativamente equilibrado, en su mayoría los riesgos están sobre un valor cercano a 1, los únicos puntos que se evidencian están fuera del rango son los departamentos fronterizos de las zonas de la Orinoquia y la Amazonía colombiana, esto puede explicarse por los recursos en equipos, personal y medicamentos que estas regiones tienen comparados con los departamentos y ciudades con una mayor partida presupuestal.

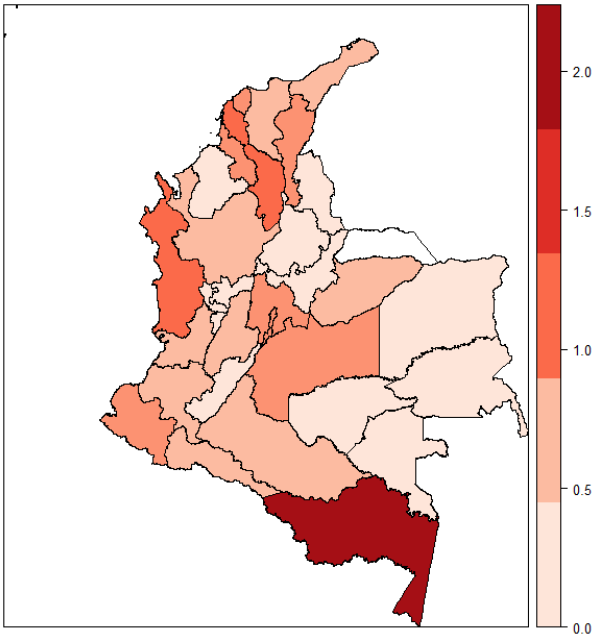

Figura 23: Mapas Relativos de Muertes Junio por Departamentos, Fuente:Autoría Propia

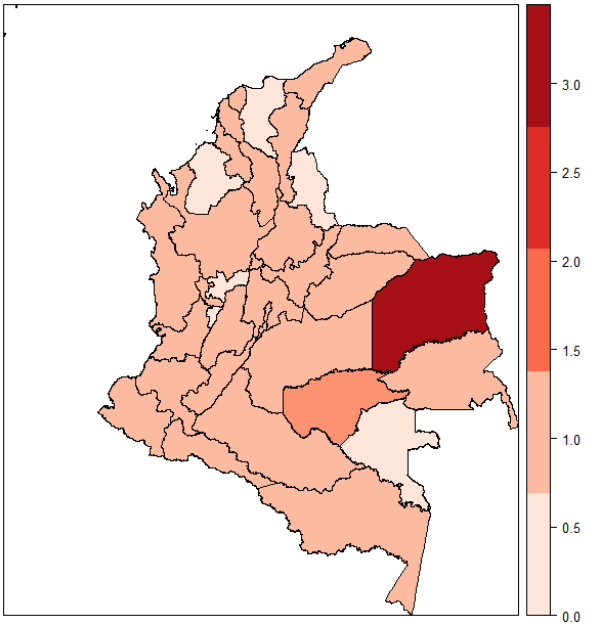

Figura 24: Mapas Relativos de Muertes Septiembre por Departamentos, Fuente:Autoría Propia

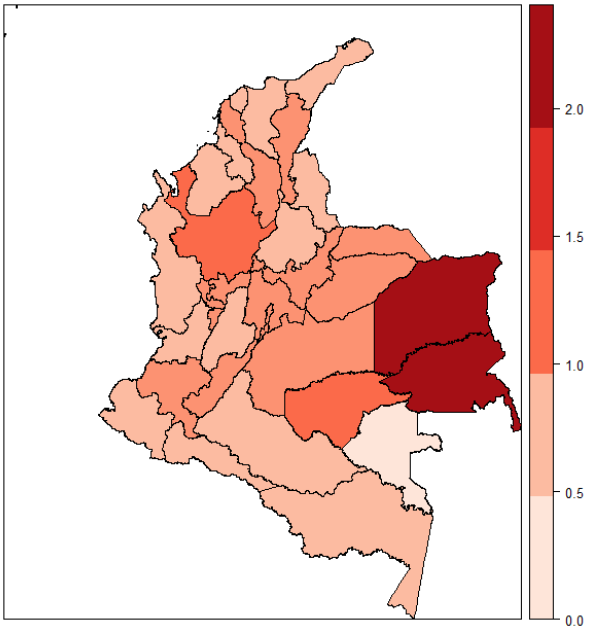

Figura 25: Mapas Relativos de Muertes Noviembre por Departamentos, Fuente:Autoría Propia

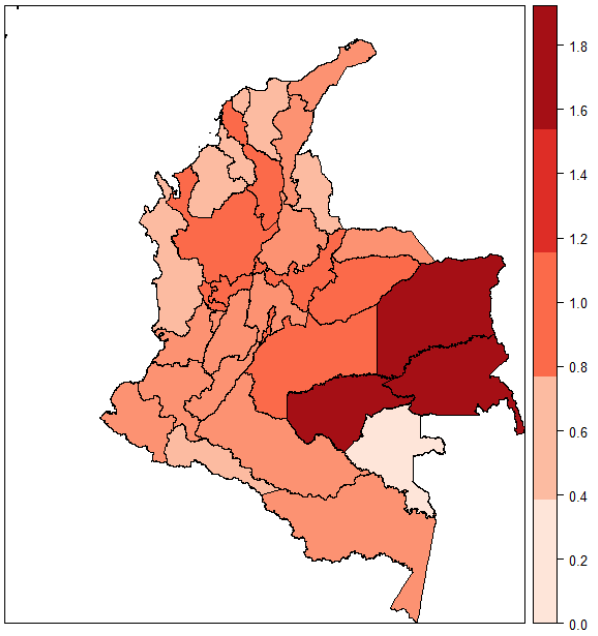

Figura 26: Mapas Relativos de Muertes Enero por Departamentos, Fuente:Autoría Propia

Función de correlación cruzada

Por medio del gráfico de correlación cruzada es posible comprobar las relaciones de dos conjuntos de datos entre ellos a lo largo del tiempo.

Función de correlación cruzada Casos-Muertes

Ratificando datos médicos, las características clínicas de los pacientes muertos por covid, con una enfermedad asociada, se encuentra en su mayoría entre los 4-20 días después de reportar síntomas, esto se muestra con la correlación cruzada en donde tiene significancia hasta los días 23.(Molina et al.)

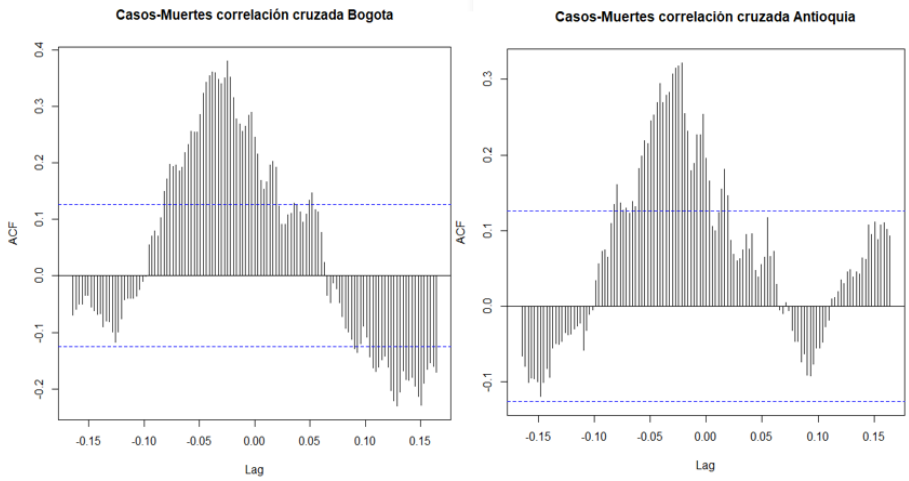

Figura 27: Correlación Cruzada Casos-Muertes Bogotá Antioquia Lag 60, Fuente:Autoría Propia

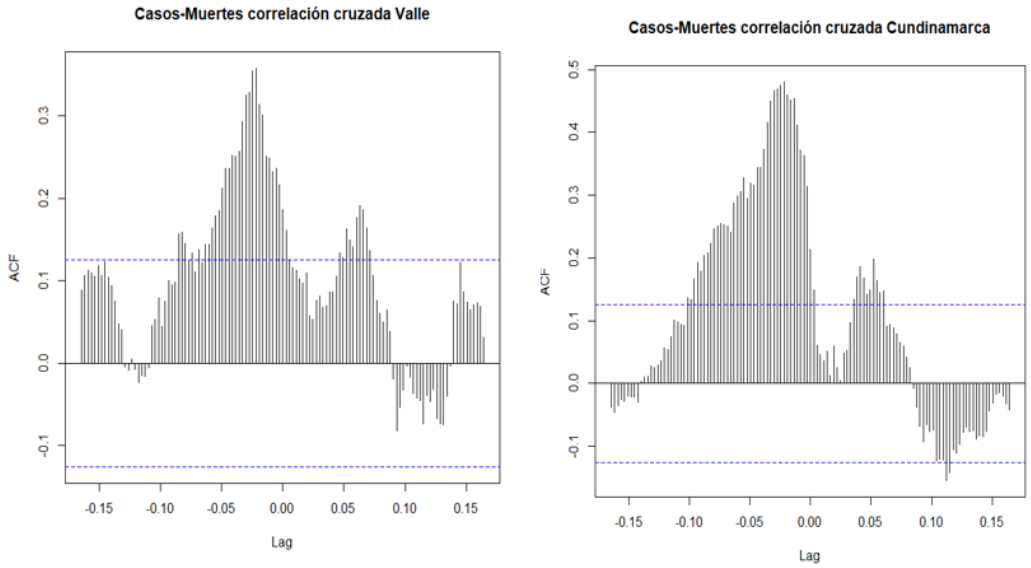

Figura 28: Correlación Cruzada Casos-Muertes Valle Cundinamarca Lag 60, Fuente:Autoría Propia

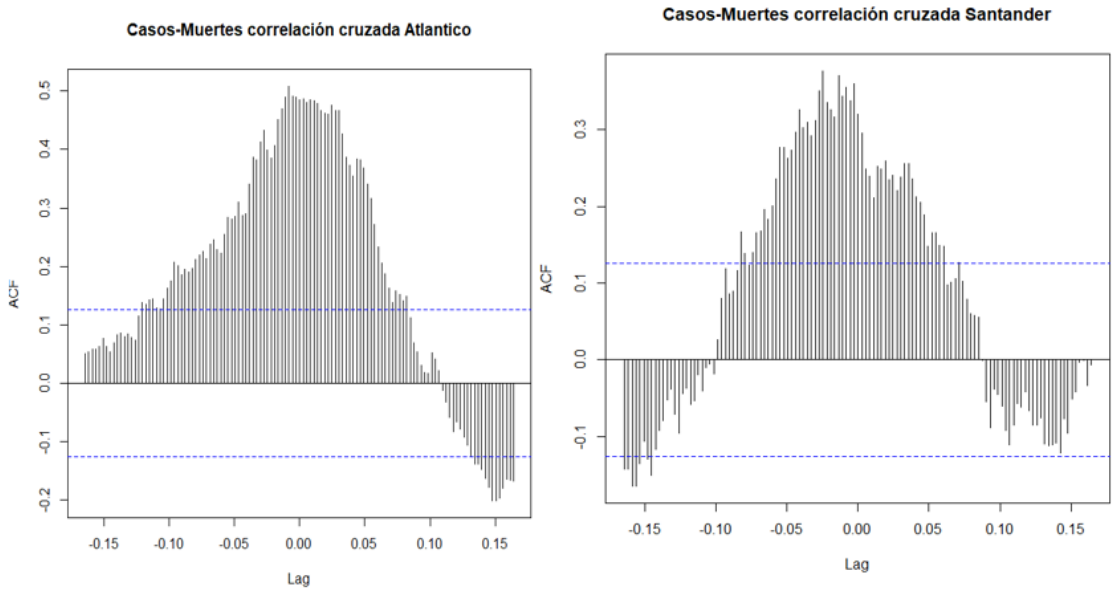

Figura 29: Correlación Cruzada Casos-Muertes Atlantico Santander Lag 60, Fuente:Autoría Propia

Función de correlación cruzada Casos-Pruebas

Para el caso de los casos reportados y las pruebas realizadas, tienen una diferencia de casi un mes, esto se debe a la cantidad de días que puede variar por paciente la presencia de síntomas y los periodos de procesamiento de pruebas, aunque se tienen alrededor de 10 días la mayor parte de la influencia.

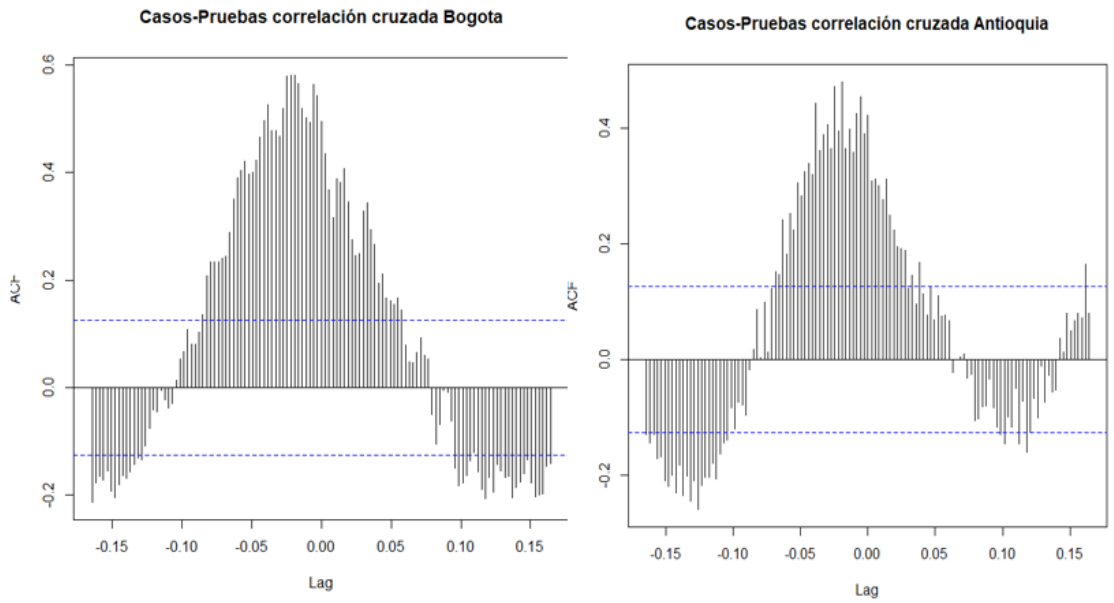

Figura 30: Correlación Cruzada Casos-Prueba Bogotá Antioquia Lag 60, Fuente:Autoría Propia

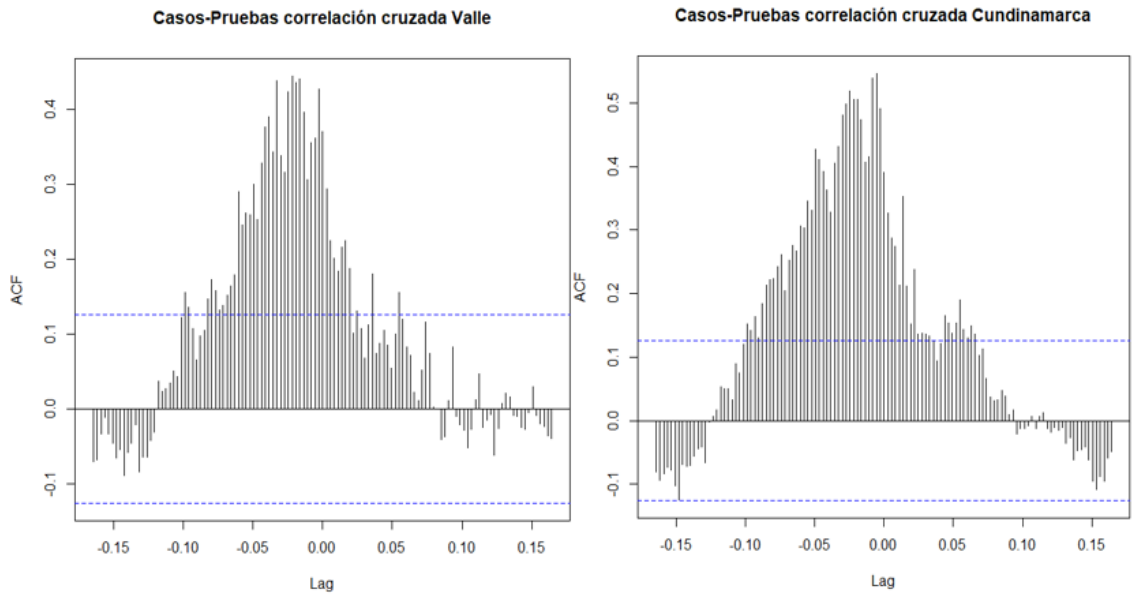

Figura 31: Correlación Cruzada Casos-Prueba Valle Cundinamarca Lag 60, Fuente:Autoría Propia

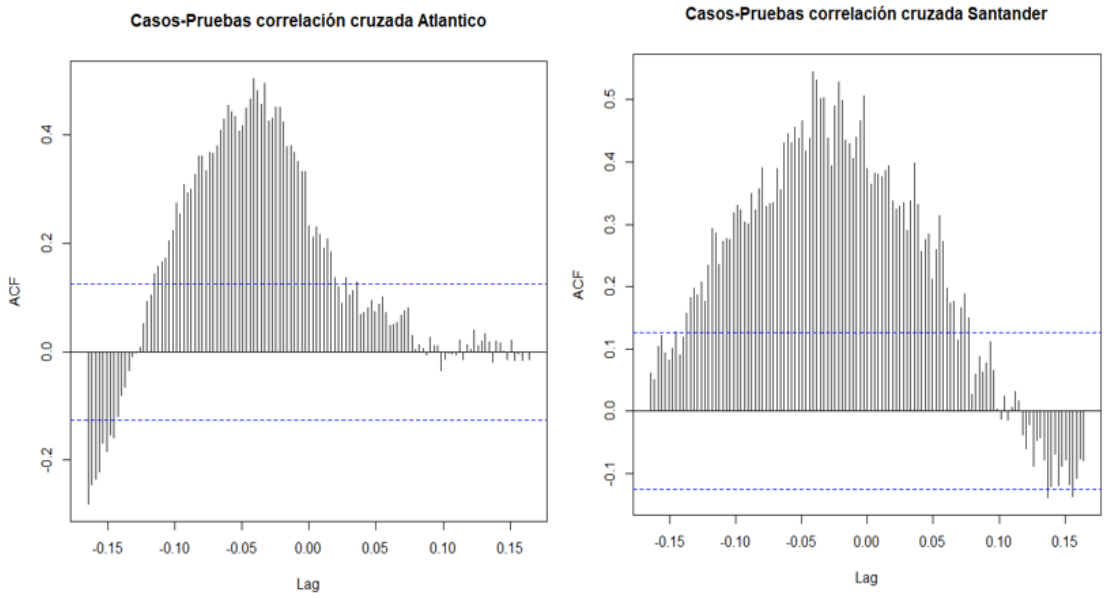

Figura 32: Correlación Cruzada Casos-Prueba Atlantico Santander Lag 60, Fuente:Autoría Propia

Autocorrelación Espacial

El primer paso para la aplicación de un modelo espacial es conocer si efectivamente existe una correlación espacial en los datos estudiados. Por lo cual el sistema de hipótesis diseñado estaría descrito como:

H_0 : Índice de Moran = 0 (No existe correlación espacial en las variables de estudio)
H_1 : Índice de Moran $\neq 0$

Para la generación de los cálculos de las correlaciones, los primeros pasos es la generación de la matriz de adyacencia, para nuestro ejercicio, esta corresponde a dimensiones de 33×33, por el supuesto de generar un vecindario especial a la ciudad de Bogotá. Esta matriz puede tener dos estilos, un Style B, en donde se tendran valores de 1 si existe una frontera y 0 en caso contrario. Desde el punto de vista del Style W, la sumatoria de los pesos será igual a 1, y si existe una frontera tomará el valor de 1/n vecindades con frontera. Para el ejercicio se realizaron los cálculos tanto por matrices de pesos, como de contigüidad, se evidencia que con un nivel de significancia del 5 % solo los meses de julio y septiembre el índice de moran, no puede rechazar la hipótesis nula.

A pesar de que el índice de morar debe variar entre -1 y 1, podemos encontrar que en los meses mas influyentes este indicador se encuentra cercano al 30 %. Dando como resultado que la influencia espacial aunque existe, puede que un indicador de Moran tan bajo, pueda afectar los parámetros espaciales de los modelos, razón por la cual se verán afectadas las interacciones espacio temporales.

Este indicador en el último trimestre del año, incremento con respecto a los meses anteriores, para el caso de enero el valor de este índice puede ser afectado por el factor vacacional, en donde las personas de los lugares más densamente poblados, llevaron nuevos casos a otros departamentos no explicando las correlaciones con tanta facilidad, comparada con meses como noviembre que al estar concentrados en un mismo sitio, el factor espacial podía explicar en parte la cantidad de contagios acumulados.

Esto puede evidenciarse de manera gráfica en el mapa de riesgos relativos, en donde se observa el cambio en los trimestres de noviembre, diciembre y enero. En el cual los dos primeros meses se encontraba estabilizado, mientras que para el último mes analizado presentan una mayor variación en los departamentos.

Tabla 2: Indicadores Mensuales de correlación espacial mediante Matriz de Frontera

Indicadores Style B					
Mes	desviación estándar	valor p	I. Moran	Valor Esperado	Varianza
Junio	2.800	0.002	0.245	-0.032	0.009
Julio	1.264	0.103	0.092	-0.032	0.009
Agosto	2.382	0.008	0.200	-0.032	0.009
Septiembre	1.199	0.115	0.091	-0.032	0.011
Octubre	1.943	0.026	0.168	-0.032	0.010
Noviembre	2.623	0.010	0.238	-0.032	0.010
Diciembre	1.970	0.024	0.171	-0.032	0.011
Enero	1.668	0.047	0.138	-0.032	0.010

Tabla 3: Indicadores Mensuales de correlación espacial mediante Matriz de Pesos

Indicadores Style W					
Mes	desviación estándar	valor p	I. Moran	Valor Esperado	Varianza
Junio	3.032	0.002	0.290	-0.032	0.011
Julio	1.431	0.076	0.119	-0.032	0.011
Agosto	2.542	0.005	0.233	-0.032	0.011
Septiembre	1.260	0.103	0.107	-0.032	0.012
Octubre	2.349	0.009	0.228	-0.032	0.123
Noviembre	3.042	0.001	0.305	-0.032	0.012
Diciembre	2.297	0.011	0.223	-0.032	0.012
Enero	0.010	0.022	0.189	-0.032	0.012

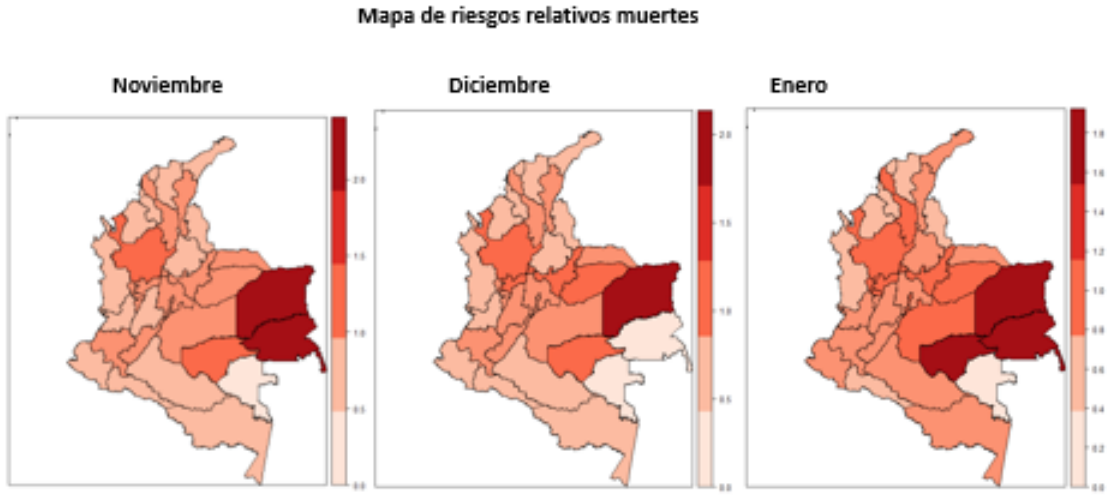

Figura 33: Trimestre final analizado por medio de mapas relativos de la tasa de mortalidad, Fuente:Autoría Propia

Al cambiar los parámetros de los índices de moran, en donde cambiamos la aleatorización por un supuesto de normalidad, encontramos conclusiones similares, en cuanto a la evolución del índice de moran a lo largo de los meses, y también para los meses de julio y septiembre no poseen significancia estadística. Desde otro punto de vista, en este caso analizando los residuales, encontramos que tiene el mismo comportamiento de los valores normalizados, llegando a las mismas conclusiones previamente mencionadas.

Tabla 4: Indicadores Mensuales de correlación espacial mediante Matriz de Frontera con Aleatorización

Indicadores Style B					
Mes	desviación estándar	valor p	I. Moran	Valor Esperado	Varianza
Junio	2.650	0.004	0.245	-0.032	0.011
Julio	1.185	0.118	0.092	-0.032	0.011
Agosto	2.215	0.013	0.200	-0.032	0.011
Septiembre	1.170	0.121	0.091	-0.032	0.011
Octubre	1.910	0.028	0.168	-0.032	0.011
Noviembre	2.577	0.005	0.238	-0.032	0.011
Diciembre	1.942	0.026	0.171	-0.032	0.011
Enero	1.625	0.052	0.138	-0.032	0.011

Tabla 5: Indicadores Mensuales de correlación espacial mediante Matriz de Pesos con aleatorización

Indicadores Style W					
Mes	desviación estándar	valor p	I. Moran	Valor Esperado	Varianza
Junio	2.864	0.002	0.290	-0.031	0.012
Julio	1.341	0.089	0.119	-0.031	0.012
Agosto	2.359	0.009	0.233	-0.031	0.012
Septiembre	1.230	0.109	0.107	-0.031	0.012
Octubre	2.312	0.010	0.228	-0.031	0.012
Noviembre	2.991	0.001	0.305	-0.031	0.012
Diciembre	2.268	0.011	0.223	-0.031	0.012
Enero	1.960	0.025	0.189	-0.031	0.012

Para el caso de septiembre esta puede explicarse a partir de pruebas no procesadas por laboratorios, un segundo factor que influyó fue la reducción de contagios en agosto, permitiendo medidas mas laxas, generando posteriormente un pico.

La principal diferencia de las tablas 4-5, con las tablas 1-2; es el método de calculo del indice de mran, en este se realiza la prueba sobre los residuales, de una regresión que solo tiene como parámetro el incercepto. dando a lugar conclusiones similares en el valor p. Nuevamente en las tablas 4-5; se diferencias, basados en el calculo de los pesos, en una se basa en el calculo de las fronteras, en el caso B siendo binaria, mientras que el W, es producto del número de vecindades del área de interés.

3.3. Modelo Espacio Temporal

Se generaron modelos univariados como un primer ejercicio como punto de partida y de comparación con el modelo bivariado, en el caso de describir la tasa de contagio, se generaron los siguientes modelos:

Modelo Tasa de Contagios

Basado en lo expuesto anteriormente, como primer paso se generó un modelo de tasa de contagios en los periodos expuestos. Se evidencia que a medida que se modificaba la forma de entrada de la covariable en los modelos, los indicadores WAIC y DIC disminuían, por lo que el mejor modelo encontrado de forma univariada para describir la tasa de contagios para el periodo estudiado es un Poisson o binomial de interacción I, con proceso idd para las pruebas para la variable rezagada de casos a 15 días. Esto puede vincularse a los tiempos de confirmación de casos por parte los laboratorios.

Tabla 6: Indicadores WAIC Modelos Tasas de Contagios

WAIC					
Modelo Casos	Pruebas	Pruebas procesos idd	Rezago 1d	Rezago 7d	Rezago 15d
Interacción I. Casos BN	97848	97426	-	-	68767
Interacción I. Casos Poisson	97684	97321	96396	95183	69469
Interacción I. Casos Binomial	97556	97197	96271	95058	69102
Interacción II. Casos BN	3420026	-	-	-	-
Interacción II. Casos Poisson	61445	61465	-	59917	81320
Interacción II. Casos Binomial	61441	-	60799	59913	3586
Interacción III. Casos BN	151339	97421	141804	139867	80289
Interacción III. Casos Poisson	159227	97297	96504	-	70043
Interacción III. Casos Binomial	159902	143081	96381	95170	69708

Tabla 7: Indicadores DIC Modelos Tasas de Contagios

DIC					
Modelo Casos	Pruebas	Pruebas procesos idd	Rezago 1d	Rezago 7d	Rezago 15d
Interacción I. Casos BN	100099	99743	-	-	70105
Interacción I. Casos Poisson	99880	99639	98698	97450	71207
Interacción I. Casos Binomial	99755	99516	98576	97327	70809
Interacción II. Casos BN	7837434	-	-	-	-
Interacción II. Casos Poisson	62061	62092	-	60496	82160
Interacción II. Casos Binomial	62057	-	61401	60493	3754
Interacción III. Casos BN	151325	99710	142584	140877	80171
Interacción III. Casos Poisson	172628	99588	98776	-	71489
Interacción III. Casos Binomial	172769	143966	98654	97408	71127

Para la comparación frente al modelo bivariado, se usará un indicador MAPE para comparar el ajuste a los datos, con respecto a la posibilidad de usar el modelo para una predicción t+1, se evidencia que estos casos tienen una medida de rezago y a medida que aumenta está, con respecto al tiempo de incubación del virus y el tiempo de procesamiento de las pruebas tanto sus indicadores WAIC, DIC como su capacidad de pronóstico, mejoran frente a los escenarios que no tuvieron rezagos.

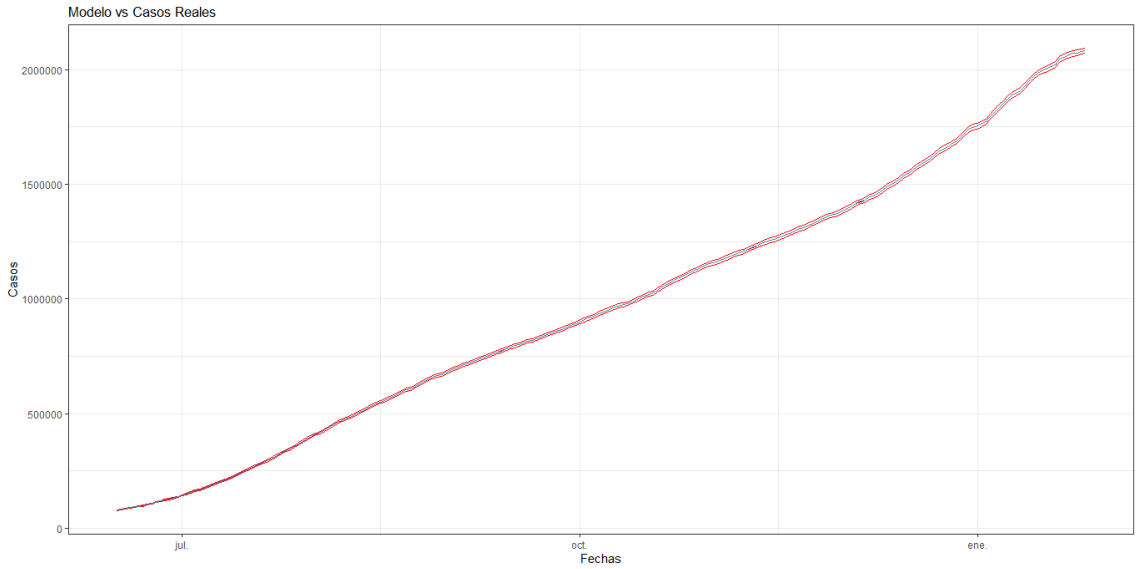

Figura 34: Comparativos Casos vs Estimaciones modelo, Limites Inferiores y superiores linea Roja, Estimación Linea Azul, Fuente:Autoría Propia

Con respecto a la comparación de la cantidad de casos pronósticados frente a los casos estimados, se observa que inclusive en bandas encontramos un excelente ajuste, aunque estas al final comienzan ha aumentar la brecha entre ellas. Por lo que es posible explicar con un excelente resultado los casos de contagios mediante un modelo univariado espacio temporal.

Tabla 8: Indicadores Predicciones t+1 Modelos Tasas de Contagios

Predicción (t+1)					
Modelo Casos	Pruebas	Pruebas procesos idd	Rezago 1d	Rezago 7d	Rezago 15d
Interacción I. Casos BN	0.2 %	inf	-	-	inf
Interacción I. Casos Poisson	0.2 %	0.2 %	0.0 %	0.2 %	1.1 %
Interacción I. Casos Binomial	0.2 %	0.2 %	0.0 %	0.2 %	1.0 %
Interacción II. Casos BN	0.2 %	-	-	-	-
Interacción II. Casos Poisson	0.2 %	97.0 %	-	97.3 %	3680.3 %
Interacción II. Casos Binomial	97.0 %	-	61401	97.3 %	3680.5 %
Interacción III. Casos BN	58.4 %	inf	inf	inf	int
Interacción III. Casos Poisson	53.1 %	19.91 %	0.2 %	-	1.16 %
Interacción III. Casos Binomial	53.1 %	0.2 %	0.2 %	0.2 %	1.1 %

A partir de los encontrados en las tablas, podemos decir que en caso de desear una buena predicción, debemos buscar los modelos que poseen rezagos en 7 o 15 días, de los cuales buscaremos modelos con respuesta Poisson o binomial, esto se debe a su desempeño en DIC, WAIC y MAPE. Con respecto a la interacción podemos decir que al ser el I. de Moran tan bajo, los efectos estructurales espaciales no afectan en mayor medida, a diferencia de los modelos temporales, en los cuales los betas son mucho más altos. Para la construcción del modelo bivariado que es costoso computacionalmente se buscarán modelos de interacción I, con variables respuesta con combinaciones Poisson y/o Binomiales. Con respecto al tipo de interacciones

Efectos Temporales Estructurados

Se empieza con el efecto del periodo de manera iid, al ver la imagen se evidencia que este efecto aunque fluctua, se encuentra siempre alrededor de 1, por lo que es posible decir que la cantidad de pacientes contagiados no son independientes día a día y dependen de días anteriores, comportamiento que podría considerarse normal, razón por la cual se eligió una caminata autoregresiva de tamaño 2, en el cual los efectos de días anteriores pueden explicar los casos del periodo de interés t.

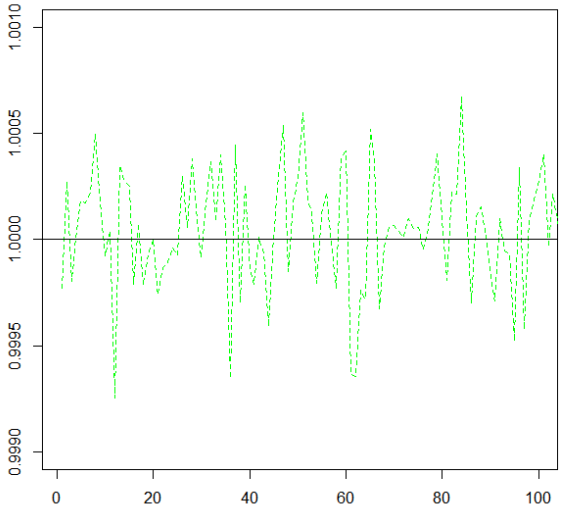

Figura 35: Efecto Fijo temporal no estructurado temporal , Fuente:Autoría Propia

Con respecto a los caso no estructurado de departamentos se evidencia una linea creciente de 0 hacia 2 en el periodo estudiando, este caso muestra un crecimiento lineal a medida que se aumentan los periodos, dando como resultado que hasta el momento de estudio se tendría una pendiente positiva, con respecto a los departamentos este fluctua a lo largo de todos los elementos por lo que no es completamente lineal y absorbe la varianza de cada vecindad.

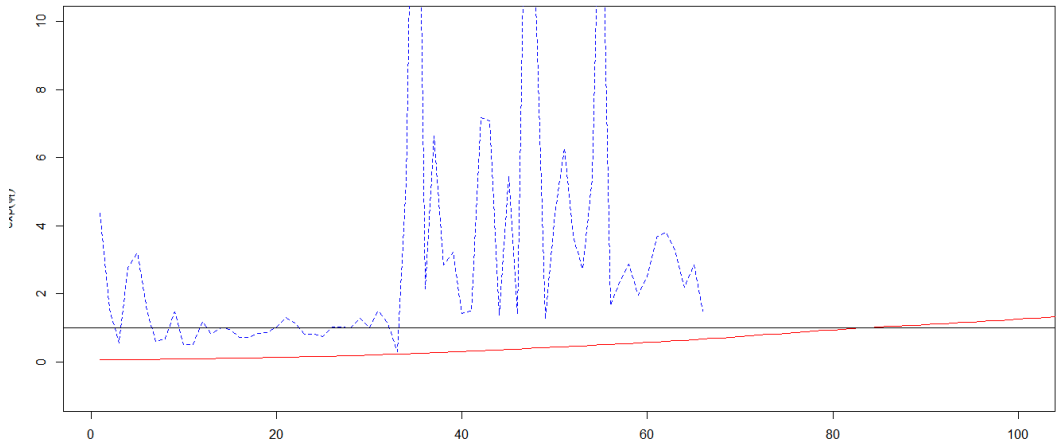

Figura 36: EF temporal no departamento estructurado temporal, Fuente:Autoría Propia

Los efectos sobre las pruebas y los un efecto temporal muestran una fluctuación, aunque tiene como media en el primer caso de 0 y la segunda cercana a uno, para el caso de las pruebas muestra la fluctuación ligada a los picos de la pandemia en donde es cercana a picos que presentan un mayor número de pruebas procesadas.

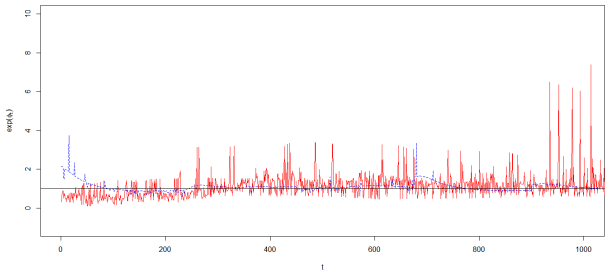

Figura 37: Efecto temporal no estructurado Efecto no estructurado Pruebas, Fuente:Autoría Propia

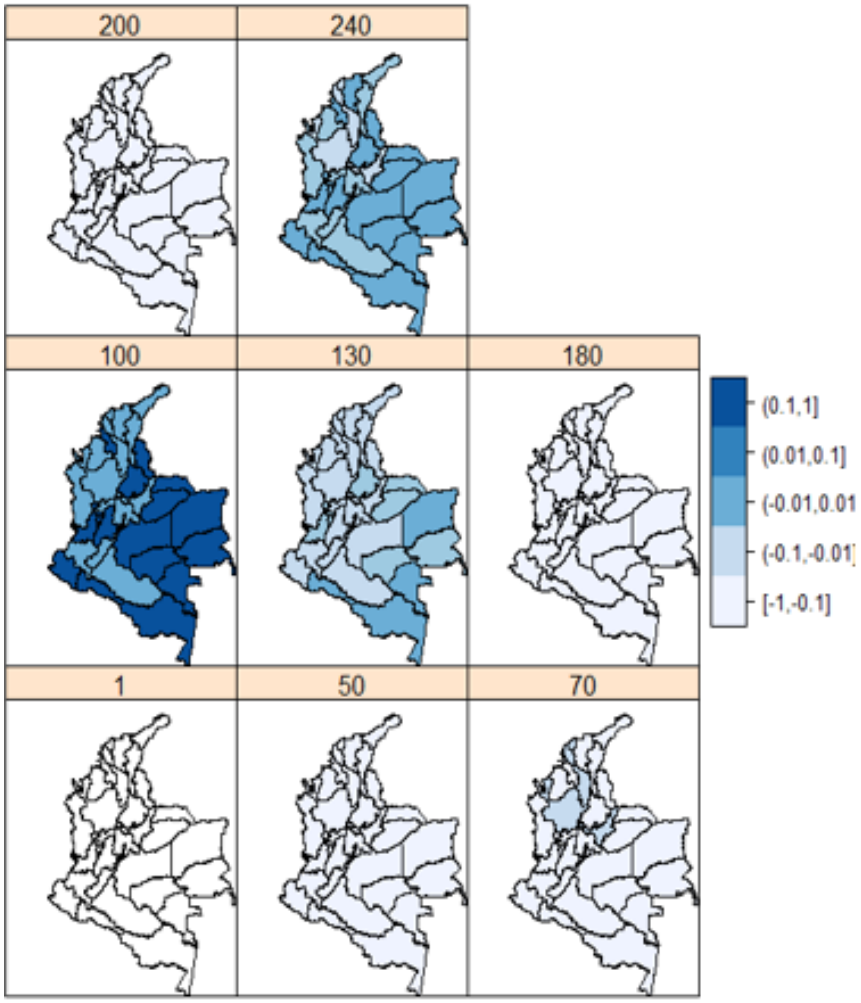

Figura 38: Tasa Contagios Modelo, Fuente:Autoría Propia

Modelo Final Contagio

Sobre el modelo final de contagio es posible concluir que:

- Los factores temporales son los elementos que mas influyen en la variable respuesta.
- Es mas importante es efecto espacio temporal que el efecto espacial, entre ambos poseen una relación de casi 4:1; mostrando que importa tanto el lugar como el momento en el qeu se hace la medición
- El tratamiento dado a las pruebas rezagadas, poseen mayor influencia que el efecto espacial.

Al aumentar el rango de tiempo en estudiado, es posible que tanto el intercepto como los ponderadores cambien, esto se debe a que la variable de contagios aún no encuentra su máximo y variables externas al modelo (Vacunación o nuevas cepas) pueden cambiar la tendencia de los datos.

Tabla 9: Efectos Modelo de Contagios Univariados (Estimaciones de media, desviación estándar y cuantiles)

Efectos Modelo Contagios Univariado					
Variable	mean	sd	0.025quant 1d	0.5quant 7d	0.975quant
(Intercept)	-4.67	0.08	-4.84	-4.67	-4.51
ID DEPTOR	3.80	0.93	2.18	3.74	5.79
ID DEPTOR	0.16	0.13	0.01	0.01	0.51
ID Periodo	127000	42400	59500	122000	224000
ID Periodo2	39700	32700	892	30300	1260000
ID DEPPER	15.80	0.72	14.30	15.80	17.20
Pruebas Res	6.05	0.17	5.72	6.04	6.40

Modelo Tasa de Mortalidad

Mientras que para el caso de muertes el mejor modelo encontrado es binomial de interacción I, con proceso idd para la variable de muertes acumuladas a 15 días. Ligado a el tiempo de incubación y muerte de los casos más agresivos de los contagiados de COVID 19.

Tabla 10: Indicadores WAIC Modelos Tasas de Contagios

WAIC					
Modelo Casos	Pruebas	Pruebas procesos idd	Rezago 1d	Rezago 7d	Rezago 15d
Interacción I Casos BN	98490	168252	74722	71087	68767
Interacción I Casos Poisson	73988	73447	72633	71553	69469
Interacción I Casos Binomial	73284	72808	72048	71054	69102
Interacción II Casos BN	85102	-	-	-	-
Interacción II Casos Poisson	85103	85217	84370	36800	81320
Interacción II Casos Binomial	3759	-	3726	36800	3586
Interacción III Casos BN	98743	-	72515	82459	80289
Interacción III Casos Poisson	8123	73892	84083	72050	70043
Interacción III Casos Binomial	83249	73252	73102	71564	69708

Tabla 11: Indicadores DIC Modelos Tasas de Mortalidad

DIC					
Modelo Casos	Pruebas	Pruebas procesos idd	Rezago 1d	Rezago 7d	Rezago 15d
Interacción I. Casos BN	98494	76378	-	72395	70105
Interacción I. Casos Poisson	75807	75353	74515	73399	71207
Interacción I. Casos Binomial	75063	74680	73905	72873	70809
Interacción II. Casos BN	85942	-	-	-	-
Interacción II. Casos Poisson	85949	6114	85212	84242	82160
Interacción II. Casos Binomial	3937	-	3902	38539	3754
Interacción III. Casos BN	98751	-	74123	82882	80171
Interacción III. Casos Poisson	81332	75543	84047	-	71489
Interacción III. Casos Binomial	82283	74895	74722	73128	71127

Tabla 12: Indicadores Predicciones t+1 Modelos Tasas de Mortalidad

Predicción (t+1)					
Modelo Casos	Pruebas	Pruebas procesos idd	Rezago 1d	Rezago 7d	Rezago 15d
Interacción I. Casos BN	20.02 %	-	inf	inf	inf
Interacción I. Casos Poisson	1.5 %	1.1 %	1.1 %	0.9 %	1.1 %
Interacción I. Casos Binomial	1.5 %	1.0 %	1.0 %	0.9 %	1.0 %
Interacción II. Casos BN	3680.6 %	-	-	-	-
Interacción II. Casos Poisson	3680.4 %	3680.0 %	9680.6 %	3680.6 %	3680.3 %
Interacción II. Casos Binomial	3680.4 %	3680.0 %	-	3680.6 %	3680.6 %
Interacción III. Casos BN	3680.6 %	inf	-	inf	int
Interacción III. Casos Poisson	3680.6 %	1.1 %	1.1 %	1.1 %	1.2 %
Interacción III. Casos Binomial	3680.6 %	1.2 %	1.2 %	1.0 %	1.1 %

Para la tasa de mortalidad fue elegido el modelo binomial de Interacción I; la metodología de selección

fue la generación de 45 modelos en donde se variaban las covariables (rezagos) y las interacciones. Posterior a obtener los criterios seleccionados (DIC, WAIC, MAPE), se buscaban los mejores resultados, sobresaliendo los modelos de Interacción I, Binomial y Poisson, los cuales entraran como comparativos contra los modelos bivariados. Estos comportamientos los encontramos de las tablas 10 a la 12, en los cuales a medida que variabamos los rezagos, generaban mejores indicadores, y las interacciones no tienen un efecto temporal estructurado, eran las que generaban mejores predicciones.

Si es comparado los casos frente a los contagios, el modelo de tasa de muertes tiene oportunidades de mejorar, esto se evidencia especialmente en las bandas en donde estas tienen una brecha mucho mayor que el modelo de contagios, aún siendo este el caso, la cantidad de muertes estimadas frente a las reales tienen un comportamiento similar, con la excepción de algunos picos, dando como resultado un modelo apropiado para ser comparado con el modelo bivariado.

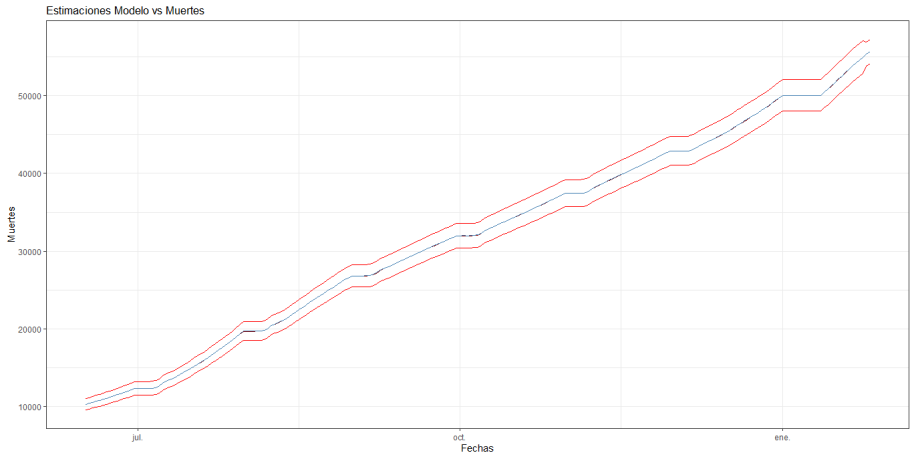

Figura 39: Comparativos Casos, Fuente:Autoría Propia

Efectos Temporales Estructurados

Al igual que en el modelo de contagios, este factor temporal se encuentra fluctuando siempre sobre 1, dando como resultado una variable que no aporta información en el modelo final.

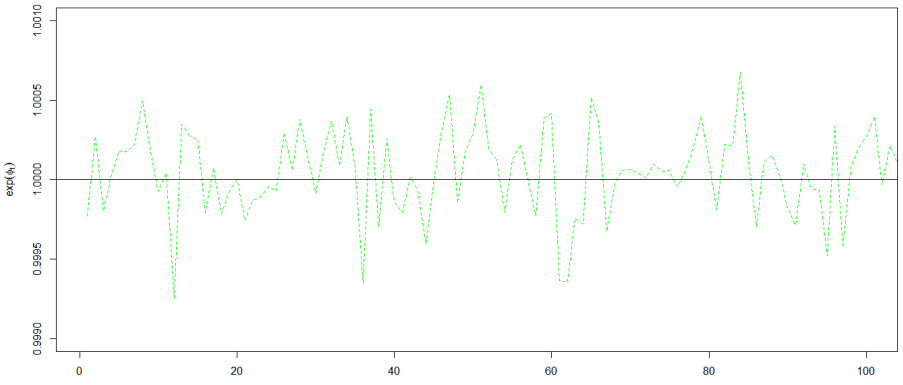

Figura 40: Efecto Fijo Temporal iid, Fuente:Autoría Propia

Con resultados similares a los encontrados en casos de estudios, el departamento es importante al fluctuar la cantidad de muertos y no es igual para todos, esto puede verse tambien revisando los mapas relativos, en donde los departamentos fronterizos y con menor presupuestos de salud presentaban una mayor tasa de muertes. Con respecto a los periodos que tienen una tendencia positiva.

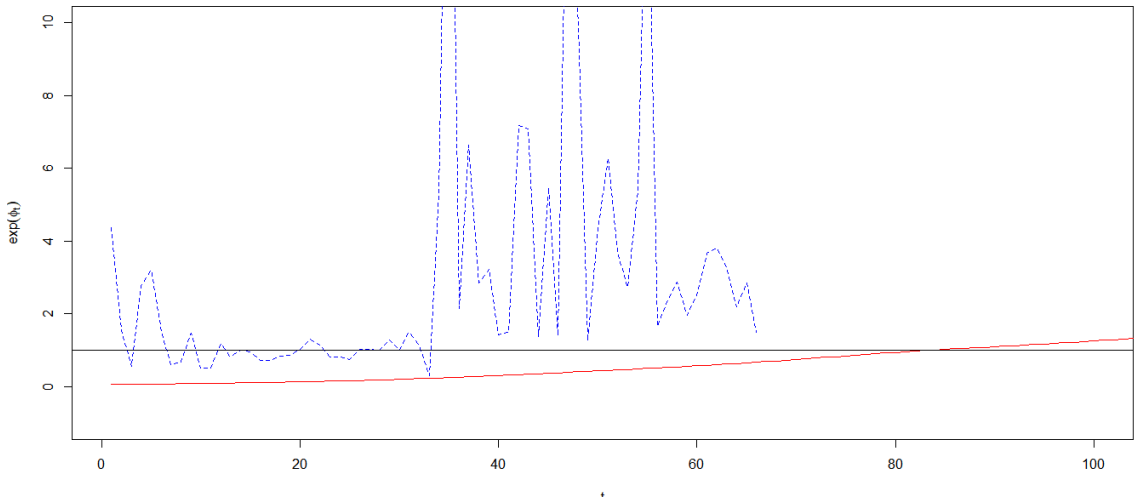

Figura 41: Efecto fijo estructurado RW2 No estructurado Departamentos, Fuente:Autoría Propia

Al comparar un proceso de casos y un efecto temporal, se encuentra que posee de manera intrinceca los picos encontrados a lo largo del periodo estudiado, en su mayoria destacados por medidas mas laxas, aperturas de fronteras departamentales y reducción en los toques de queda usados por el gobierno nacional para controlar la pandemia, un segundo elemento que no debe ignorarse es la tendencia negativa que se tiene en el efecto temporal, esto puede darse al reducir la población suceptibles de contagios y dar prioridad a la población en riesgos, esto reduce la cantidad de posibles muertes en relación a los contagiados.

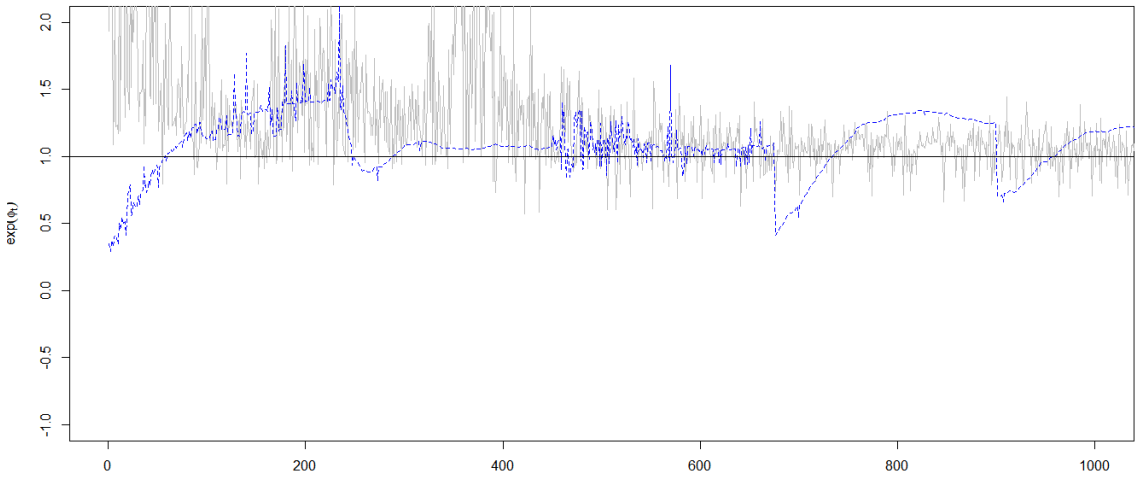

Figura 42: Comparativos Casos, Fuente:Autoría Propia

Al igual que el modelo anterior, todas las variables tienen una tendencia positiva, con excepción de la pendiente. Esto como resultado de que la cantidad de muertes es acumulada, donde las variables que mas influyen son los efectos temporales encontradas en las variables periodos y la cantidad de casos reportados.

Tabla 13: Efectos Modelo Mortalidad Univariados (Estimaciones de media, desviación estándar y cuantiles)

Efectos Modelo Mortalidad Univariado					
Variable	mean	sd	0.025quant 1d	0.5quant 7d	0.975quant
(Intercept)	-3.06	0.06	-3.19	-3.06	-2.94
ID DEPTOR	5.99	1.52	3.54	5.81	9.47
ID DEPTOR	0.23	0.20	0.01	0.17	0.73
ID Periodo	12500	40700	59400	120000	217000
ID Periodo2	38700	29200	9370	30700	116000
ID DEPPER	10.90	0.33	10.20	10.90	11.50
Casos Res	14.00	0.57	12.90	13.90	15.10

Modelo Bivariado Tasa de Contagios-Tasa de Mortalidad

Al evaluarse 6 modelos, en este caso todos bajo la interacción I, la cual dio los mejores resultados ya descritos anteriormente en el univariado. Basado en los indicadores de DIC y el WAIC, se evidencia que las distribuciones combinadas Binomial para los casos y Poisson para las cantidad de muertes registradas.

Tabla 14: Indicadores DIC y WAIC como comparativo de los Modelos Bivariados

Indicadores Modelos Bivariados			
Modelo Casos	Indicador	Rezago 7d	Rezago 15d
Interacción I Binomial-Poisson	WAIC	4479180	4131446
Interacción I Poisson-Binomial	WAIC	127015961	126530445
Interacción I Binomial-Binomial	WAIC	34475493	34370005
Interacción I Binomial-Poisson	DIC	5449022	4850258
Interacción I Poisson-Binomial	DIC	264167732	262384871
Interacción I Binomial-Binomial	DIC	246916535	46776201

Modelo Final Bivariado Tasa Contagio-Tasa Mortalidad

Con respecto al modelo bivariado, se encuentran que los casos y muertes acumuladas se encuentran con tendencia negativa al igual que el intercepto, para ir en contraparte a la variable periodo que absorbe el crecimiento de la cantidad de casos y muertes. Con respecto a los resultados se evidencia que se encuentran en una calidad inferior a los modelos univariados, no por el desempeño del modelo, en lugar de esto fue por el excelente desempeño obtenido por los modelos univariados, en donde los mejores modelos generaban

indicadores MAPEs y predicción t+1, inferiores al 1 %, describiendo con mejor exactitud la cantidad de casos de muertes y contagios en los diferentes dias y analizados.

Tabla 15: Efectos Modelo Contagios-Mortalidad Bivariado (Estimaciones de media, desviación estándar y cuantiles)

Efectos Modelo Contagios-Mortalidad Bivariado					
Variable	mean	sd	0.025quant 1d	0.5quant 7d	0.975quant
(Intercept)	-4.38268	0.10394	-4.587055	-4.38313	-4.1884
Casos Acum1	-6.24E-6	6.33E-7	-7.48E-6	-6.24E-6	-5.00E-6
Muertos Acum1	8.620E-5	3.410E-5	1.920E-5	8.620E-5	1.5300E-4
Casos Acum1	7.64E-6	0.34E-7	6.40E-6	7.64E-6	8.88
Muertos Acum	-4.1600E-4	3.410E-5	-4.830E-4	-4.160E-4	-3.490E-4
Precision ID DEPTOR	2.610000	0.661000	1.44000	2.580000	4.000000
Phi ID Periodo	0.121000	0.104000	9.190E-3	9.1200E-2	0.401000
Precision ID Periodo	8.87E4	3.64E4	3.28E4	8.40E4	1.73E5
Precision ID Periodo2	3.76E4	2.50E4	9.263	3.14E4	1.03E4
Precision ID Depper	2.93E4	2.28E4	5.07E3	2.34E4	8.97E4
Precision ID Periodo Casos Acum Res	11.932	0.43800	11.10000	11.900000	12.800000
Precision ID Pruebas Res	9.980000	0.30200	9.380000	9.990000	10.60000

Al evaluar los modelos contra los modelos bivariados, contra su contraparte que son los modelos espaciotemporales bivariados, encontramos que evaluandolos bajo los mismos parámetros de entrenamiento, los modelos univariados son más exactos aplicando un indicador MAPE. Para los casos analizados se tiene que en efecto solo un modelo bivariado tendría posibilidades de enfrentarse a los univariados, el cual es la interación Binomial Poisson, que poseen una tasa de contagios de 7 %, con respecto a la tasa de mortalidad todos los modelos analizados presentan resultados pobres en comparación con los modelos univariados.

Basados en los modelos podemos afirmar que aunque el modelo bivariado para la tasa de contagios podría ser funcional al momento de describir en mas de 1.000 puntos espacio temporales, no tiene mejor rendimiento que los modelos univariados.

Tabla 16: Comparativo del Indicador MAPE para Modelos Bivariados vs Univariados

MAPE			
Modelo Casos	Rezago	Contagios	Muertes
Interacción I Poisson	15	0.82 %	3.18 %
Interacción I Binomial	15	0.82 %	2.65 %
Interacción I Poisson	7	0.85 %	3.04 %
Interacción I Binomial	7	0.84 %	2.48 %
Interacción I Binomial-Poisson	7	6.62 %	2931.25 %
Interacción I Poisson-Binomial	7	3032.19 %	1791.44 %
Interacción I Binomial-Binomial	7	27.00 %	816.57 %
Interacción I Binomial-Poisson	15	7.73 %	2860.43 %
Interacción I Poisson-Binomial	15	3389.96 %	1744.32 %
Interacción I Binomial-Binomial	15	27.19 %	840.06 %

3.4. Conclusiones

- El mejor modelo bivariado encontrado, tuvo un desempeño inferior al compararse por medio del indicador MAPE con respecto a los modelos univariados, dando como resultado que este modelo no se ajusta de manera apropiada. Para los modelos univariados se encontro un MAPE histórico menor al 3 %, mientras que en los bivariados este indicador el menor se encontraba sobre un 7 %.

- Se evidencia que las zonas fronterizas causadas por la población flotante de estos lugares, favorecieron la expansión del COVID, otro punto adicional es que Colombia al tener una economía altamente centralizada en la zona Andina (DANE, 2022), el presupuesto y desarrollo de salud en estas zonas tienen oportunidades de mejora, mostrando como las tasas de mortalidad en los mapas cloropleticos estaban por encima de los valores esperados.

- Para los modelos univariados, se encuentra una relación con respecto a la teoría, confirmado por los gráficos de correlación cruzados, y los modelos rezagados, que 7-15 días son factores claves para describir el comportamiento de los casos o muertes. Esto muestra como los modelos están en la misma vía que la información de desarrollo del virus, en el cual su periodo de incubación se encontraba en 5 días adicionando el tiempo de procesamiento de las pruebas para el caso de contagio y para los casos de mortalidad se encontraban menor a 2 semanas en los peores casos.

- Los modelos binomiales negativos no presentan una ventaja frente a las distribuciones binomial y poisson, esto se ve refleja en los indicadores DIC y AIC univariados.

- El factor mas importante basado en los datos analizados fue el periodo, al momento de estudio la cantidad de vacunados era una variable no presente en el modelo, dado que al extender el periodo de estudio puede que los resultados no sean tan exactos como los presentados en el documento, alterando la cantidad de contagios y muertes al reducir la población suceptible de contagios.

3.5. Trabajos Futuros

Basados en las restricciones de realizarlo día a día, una ventana de tiempo en la cual pueda ingresas la cantidad de población vacunada en primera y segunda dosis, pueden tener un resultado en la generación del modelo. Esto se debe a que la adquisición cambia directamente la cantidad de personas suceptibles y las tasas de mortalidad pueden ser totalmente diferentes a las analizadas en la ventana de tiempo seleccionada para el presente trabajo. Los scripts del proyecto se encuentran almacenados en el repositorio de github, junto con las bases bases de datos utilizadas, para consulta libre de cualquier persona interesada en el proyecto: (https://github.com/ernes258/BayesTesis).

Referencias

A. Agresti. *Foundations of linear and generalized linear models*. John Wiley & Sons, 2015.

L. J. Allen, F. Brauer, P. Van den Driessche, and J. Wu. *Mathematical epidemiology*, volume 1945. Springer, 2008.

L. Anselin. Local indicators of spatial association—lisa. *Geographical analysis*, 27(2):93–115, 1995.

R. Assunção and E. Krainski. Neighborhood dependence in bayesian spatial models. *Biometrical Journal: Journal of Mathematical Methods in Bioscie0nces*, 51(5):851–869, 2009.

P. Ballesteros, E. Salazar, D. Sánchez, and C. Bolaños. Spatial and spatiotemporal clustering of the covid-19 pandemic in ecuador. *Revista de la Facultad de Medicina*, 69(1), 2021.

J. M. Barbeito and J. G. Villalón. *Introducción al cálculo estocástico aplicado a la modelización económico-financiero-actuarial*. Netbiblo, 2003.

F. Bautista, D. Palma-López, and W. Huchin-Malta. Actualización de la clasificación de los suelos del estado de yucatán. *Caracterización y manejo de los suelos de la Península de Yucatán: Implicaciones agropecuarias, forestales y ambientales*, pages 105–122, 2005.

F. Bavaud. Models for spatial weights: a systematic look. *Geographical analysis*, 30(2):153–171, 1998.

R. S. Bivand, E. J. Pebesma, V. Gómez-Rubio, and E. J. Pebesma. *Applied spatial data analysis with R*, volume 747248717. Springer, 2008.

M. Blangiardo and M. Cameletti. *Spatial and spatio-temporal Bayesian models with R-INLA*. John Wiley & Sons, 2015.

N. A. Cressie. Statistics for spatial data. Technical report, John Wiley & Sons, Inc., 1993.

DANE. Producto interno bruto por departamento. 2022.

M. G. Echeverri, A. M. Cabrera, and P. G. Echeverri. Proyección espacio-temporal del covid-19 en pereira. *TecnoLógicas*, 23(49):129–146, 2020.

D. Elías-Cuartas, D. Arango-Londoño, G. Guzmán-Escarria, E. Muñoz, D. Caicedo, D. Ortega-Lenis, A. Fandiño-Losada, J. Mena, M. Torres, L. Barrera, et al. Análisis espacio-temporal del sars-cov-2 en cali, colombia. *Revista de Salud Pública*, 22(2):1–6, 2020.

C. Feng. Spatial-temporal generalized additive model for modeling covid-19 mortality risk in toronto, canada. *Spatial statistics*, page 100526, 2021.

A. E. Gelfand, P. Diggle, P. Guttorp, and M. Fuentes. *Handbook of spatial statistics*. CRC press, 2010.

A. Gelman. Prior distributions for variance parameters in hierarchical models (comment on article by browne and draper). *Bayesian analysis*, 1(3):515–534, 2006.

D. A. Griffith. Advanced spatial statistics for analysing and visualizing geo-referenced data. *International Journal of Geographical Information Science*, 7(2):107–123, 1993.

H. Gutierrez. Al menos 906.000 personas migraron de ciudades y del campo, el año anterior. 2021.

R. P. Haining and R. Haining. *Spatial data analysis: theory and practice*. Cambridge university press, 2003.

M. F. Hernandez Villaizan et al. Análisis del número de pacientes fallecidos con cancer de mama y prostata en la poblacion colombiana por departamento en el periodo (2006-2015) mediante un modelo Poisson espacio-temporal bayesiano. 2019.

J. Kruschke. Doing bayesian data analysis: A tutorial with r, jags, and stan. 2014.

G. B. López. La geolocalización social. *Polígonos. Revista de geografía*, 27:97–118, 2015.

F. Manrique Abril, V. M. González-Chordá, O. A. Gutiérrez Lesmes, C. F. Tellez Piñerez, G. M. Herrera-Amaya, et al. Modelo sir de la pandemia de covid-19 en colombia. 2020.

M. A. Martínez-Beneito and P. Botella-Rocamora. *Disease Mapping: From Foundations to Multidimensional Modeling*. CRC Press, 2019.

L. M. C. Molina, M. J. Tejeda-Camargo, J. A. C. Clavijo, L. M. Montoya, L. J. Barrezueta-Solano, S. Cardona-Montoya, D. A. Arjona-Granados, and y. Rendón-Varela, Johnny Alexander journal=Revista Repertorio de Medicina y Cirugía, pages=45–51. Características clínicas y sociodemográficas de pacientes fallecidos por covid-19 en colombia.

R. H. Myers, D. C. Montgomery, G. G. Vining, and T. J. Robinson. *Generalized linear models: with applications in engineering and the sciences*, volume 791. John Wiley & Sons, 2012.

N. Niño Martínez. Modelo de regresión beta espacio temporal bayesiano para estimar la proporción de víctimas de desplazamiento forzado por municipio en colombia para el periodo 2002-2016. 2018.

S. T. Rachev, J. S. Hsu, B. S. Bagasheva, and F. J. Fabozzi. *Bayesian methods in finance*, volume 153. John Wiley & Sons, 2008.

B. D. Ripley. *Spatial statistics*, volume 575. John Wiley & Sons, 2005.

E. C. Rodrigues and R. Assunção. Bayesian spatial models with a mixture neighborhood structure. *Journal of Multivariate Analysis*, 109:88–102, 2012.

H. Rue and L. Held. *Gaussian Markov random fields: theory and applications*. CRC press, 2005.

S. Shekhar and H. Xiong. *Encyclopedia of GIS*. Springer Science & Business Media, 2007.

P. Sidén and F. Lindsten. Deep gaussian markov random fields. In *International Conference on Machine Learning*, pages 8916–8926. PMLR, 2020.

R. S. Society. *CAR Model*, 2018a. URL `https://www.statsref.com/HTML/index.html?car_models.html`.

R. S. Society. *SAR Model*, 2018b. URL `https://www.statsref.com/HTML/index.html?sar_models.html`.

D. J. Spiegelhalter, N. G. Best, B. P. Carlin, and A. Van Der Linde. Bayesian measures of model complexity and fit. *Journal of the royal statistical society: Series b (statistical methodology)*, 64(4): 583–639, 2002.

I. Steinsland. Parallel sampling of gmrfs and geostatistical gmrf models. Technical report, SIS-2005-003, 2003.

N. R. B. Suárez, J. E. M. Sabogal, E. J. Carmona, P. C. Rosales, F. H. L. Zambrano, M. M. Rivas, and N. O. Pérez. Análisis de asociación entre indicadores biológicos teniendo en cuenta la estructura de correlación espacial. *Modelamiento Estadístico*, page 275.

A. Tariq, T. Chakhaia, S. Dahal, A. Ewing, X. Hua, S. K. Ofori, O. Prince, A. Salindri, A. E. Adeniyi, J. M. Banda, et al. An investigation of spatial-temporal patterns and predictions of the covid-19 pandemic in colombia, 2020-2021. *medRxiv*, 2021.

A. Trilleras Martínez. Prior no informativas y su influencia en la estimación de los parámetros no estructurales de un modelo tar multivariado. *Estadística*, 2018.

T. P. Velavan and C. G. Meyer. The covid-19 epidemic. *Tropical medicine & international health*, 25(3): 278, 2020.