
Modelo GAM espacial para la tasa de suicidio en los Departamentos de Colombia (2014-2019)

Spatial GAM Model for the Suicide Rates in the Departments of Colombia (2014-2019)

Brahian Sánchez Cardona^a
brahiansanchez@usantotomas.edu.co

Dagoberto Bermudez Rubio^b
dagobertobermudez@usta.edu.co

Resumen

El presente trabajo tiene como objetivo analizar el comportamiento de las tasas de suicidio en Colombia, por departamentos, entre 2014 y 2019, usando un modelo GAM. Lo que se busca en este modelo es capturar tanto la dependencia espacial como la temporal y así ofrecer una visión clara de los factores que influyen en las tasas de suicidio. Partiendo de que la variable de interés sigue una distribución beta, se analiza considerando la correlación espacio-tiempo, se incluye información socioeconómica por departamento para entender cómo estos factores afectan las variaciones en las tasas de suicidio a nivel regional. Los resultados muestran que los efectos espaciales y los factores socioeconómicos están significativamente relacionados, mientras que el efecto temporal varía en su importancia, este estudio ayuda a identificar patrones espaciales y temporales en las tasas de suicidio, proporcionando información valiosa para la formulación de soluciones públicas y estrategias de intervención en Colombia.

Palabras clave: Tasa de suicidio, Departamentos, Modelo espacio temporal.

^aEstudiante Estadística Universidad Santo Tomás

^bEstadístico. Profesor Facultad de Estadística Univesidad Santo Tomás

Abstract

This study aims to analyze the behavior of suicide rates in Colombia by department between 2014 and 2019 using a spatiotemporal beta model. This model seeks to capture both spatial and temporal dependencies, providing a clear view of the factors influencing suicide rates. Recognizing that the suicide rate follows a beta distribution, the analysis considers the separable space-time correlation. Socio-economic information by department is included to understand how these factors affect variations in suicide rates at the regional level. The results show significant relationships between spatial effects and socioeconomic factors, while the temporal effect varies in its importance. This study helps identify spatiotemporal patterns in suicide rates, providing valuable information for the formulation of public policies and intervention strategies in Colombia.

Keywords: Suicide rate, Departments, spatiotemporal model.

1. Introducción

El suicidio, es un problema social y de salud pública, esta representa una de las causas principales de muerte en el mundo, afecta a toda la población mundial, culturas y grupos socioeconómicos. La investigación está enfocada para Colombia, el estudio oscila entre los años 2014 y 2019, en donde se observó una tendencia preocupante en las tasas de suicidio, lo que genera la urgencia de estudios detallados que permitan comprender mejor sus dinámicas y variables predictoras. Este periodo, caracterizado por cambios sociales (COVID-19), transformaciones económicas y avances en políticas de salud mental, provee un contexto abundante para su análisis. El presente trabajo de grado se propone a explorar la problemática del suicidio en Colombia durante estos años, enfocándose en un análisis desde una perspectiva estadística y utilizando un modelo GAM con respuesta Beta que permita identificar patrones, tendencias y posibles correlaciones entre factores de riesgo a nivel regional, posterior a ello se haran las respectivas comparaciones temporales del mismo.

En la investigación se aplicarán métodos cuantitativos para analizar los diferentes datos de suicidio, en donde incluyen variables demográficas, socioeconómicas, y causa de muerte, con el objetivo de modelar la distribución Beta espacial-temporal del suicidio en Colombia.

Se utilizarán técnicas de análisis espacial para determinar si existen agrupaciones geográficas significativas en las tasas de suicidio, tambien se usara el pertinente ajuste del modelo temporal para examinar cambios y tendencias a lo largo del tiempo. El modelo GAM con respuesta Beta, caracterizado por su capacidad para manejar proporciones y tasas en un contexto espacial modelado por medio de un spline, ofrece una metodología robusta para este tipo de análisis, permitiendo evaluar el impacto de variables contextuales sobre las tasas de suicidio, la combinación de estos enfoques busca proporcionar una comprensión más profunda de cómo el

contexto y la localización geográfica influyen en la incidencia del suicidio, y cómo estos factores han evolucionado durante el periodo de estudio.

Este trabajo de grado no solo tiene como objetivo contribuir al diseño de políticas en la salud pública y estrategias de intervención más efectivas. Al identificar áreas geográficas y periodos de tiempo con mayor riesgo, los hallazgos de esta investigación podrán guiar la asignación de recursos y la implementación de programas de prevención y atención psicosocial dirigidos a las poblaciones más vulnerables. Además, el estudio se desea sensibilizar sobre la importancia de abordar la salud mental como una prioridad de salud pública, promoviendo un enfoque más proactivo.

2. Conceptualización

2.1. Suicidio.

El suicidio es la forma de autoatentado contra la propia vida. Según la Comisión Nacional de Protección Social en Salud (2021), “La OMS define el suicidio como un acto deliberadamente iniciado y realizado por una persona en pleno conocimiento o expectativa de su desenlace fatal” de México (2021).

En ocasiones las personas enfrentan problemas tan abrumadores que sienten la necesidad de rendirse, buscando una salida desesperada a sus problemas. Problemas emocionales, económicos y enfermedades mentales son algunas de las razones para llegar al suicidio, es por esto que se entiende como un problema de salud pública, en donde la intención es llegar a brindar ayuda a estas personas.

2.2. Importancia del estudio

Estudiar el suicidio es importante para prevenir muertes prematuras, además de que afecta a todo un entorno de personas, no solo a quienes lo intentan, sino también a sus amigos y familiares. Según Huellas (2016), “10 allegados se ven afectados de manera directa por este tipo de muerte. Otros refieren que el número de personas impactadas es mucho mayor por lo que el reto está en cómo nos podemos asegurar de que las personas afectadas reciban el apoyo y la ayuda que necesiten de forma adecuada y a la mayor brevedad posible”.

Si entienden mejor las razones detrás de esto, pueden prevenirlo y salvar un montón de vidas. Además, muchos casos están ligados a problemas de salud mental que se podrían manejar mejor si se detectan a tiempo. Hablar de esto abiertamente ayuda a que la gente no se sienta juzgada si busca ayuda.

- **Variable Dependiente:** Tasa de suicidios por región y año, transformada adecuadamente para ajustarse entre 0 y 1.

• **Variables Independientes:** Factores de riesgo como desempleo, indicadores de salud mental, eventos traumáticos (como violencia o desastres naturales), y otras covariables socioeconómicas relevantes.

2.3. Objetivos

2.3.1. General

Analizar el comportamiento de las tasas de suicidio en los departamentos de Colombia entre 2014 y 2019 utilizando un modelo GAM espacial con respuesta Beta, con el fin de comprender los patrones espaciales y temporales.

2.3.2. Específicos

- Capturar la dependencia espacial y temporal de las tasas de suicidio.
- Incluir información socioeconómica por departamento para entender cómo estos factores afectan las variaciones en las tasas de suicidio.
- Identificar patrones espaciales y temporales en las tasas de suicidio.

2.4. Hipótesis

Las tasas de suicidio en los departamentos de Colombia están influenciadas por factores socioeconómicos y presentan patrones de dependencia espacial y temporal.

3. Marco teórico y revisión literaria

3.1. Modelo de Regresión Beta.

El modelo de regresión beta se usa para modelar variables dependientes que toman valores en el intervalo $(0, 1)$, como proporciones y tasas. Este modelo es especialmente útil para datos que tienen distribuciones que no son simétricas o que tienen una varianza que no es constante, condiciones comunes en datos de proporciones. Esta utilidad se debe a que la variable aleatoria Y se modela con una distribución beta, que depende de dos parámetros positivos α y β , y tiene una función de densidad dada por:

$$f(y) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} y^{\alpha-1} (1 - y)^{\beta-1} \quad (1)$$

La media μ y la varianza de una variable con esta distribución se expresan como:

$$\mu = \frac{\alpha}{\alpha + \beta}$$

$$\text{Var}(y) = \frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)}$$

Para adaptar la distribución beta a un modelo de regresión, se utiliza una reparametrización empleando $\mu = \frac{\alpha}{\alpha + \beta}$ y $\phi = \alpha + \beta$, con nuevas expresiones para α y β como $\alpha = \mu\phi$ y $\beta = (1 - \mu)\phi$. Así, la media y varianza se redefinen como:

$$E(Y) = \mu$$

$$V(Y) = \frac{\mu(1 - \mu)}{1 + \phi}$$

La densidad de probabilidad de Y bajo esta nueva parametrización se describe por:

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

En el contexto de los modelos lineales generalizados (MLG), esta función de densidad se transforma a una forma exponencial biparamétrica, facilitando su incorporación en el modelo. Así, cuando reformulamos la ecuación anterior (6) usando la familia exponencial, obtenemos el siguiente caso para la distribución beta, que se presenta a continuación.

$$f(y_i, \phi, \mu) = \exp \left(\mu\phi \ln \left(\frac{y_i}{1 - y_i} \right) + \ln[y_i(1 - y_i)] + \phi \ln(1 - y_i) + \ln \Gamma(\phi) - \ln \Gamma(\mu\phi) - \ln \Gamma[(1 - \mu)\phi] \right)$$

Entonces, se relacionan los siguientes parámetros:

Tabla 1: Parámetros de la densidad beta en términos de la familia exponencial biparamétrica

F.exp biparamétrica	F. Beta
η_1	$\mu\phi$
η_2	ϕ
$T(y)$	$\log(1 - y)$
$\rho(\eta_1, \eta_2)$	$\frac{\Gamma(\phi)}{\Gamma(\mu\phi)\Gamma(\phi(1 - \mu))}$

Los parámetros de la familia exponencial biparamétrica η_1 y η_2 se relacionan a través de la media: $\frac{\eta_1}{\eta_2} = \mu$

3.2. Estimación de Parámetros

Las expresiones para los estimadores de β y ϕ se obtienen por el método de máxima verosimilitud. La función de log-verosimilitud de la anterior distribución viene dada por:

$$\ell(\beta, \phi) = \sum_{i=1}^n \ell_i(\mu_i, \phi)$$

donde,

$$\ell_t(\mu_t, \phi) = \log \Gamma(\phi) - \log \Gamma(\mu_t \phi) - \log \Gamma((1 - \mu_t)\phi) + (\mu_t \phi - 1) \log y_t + \{(1 - \mu_t)\phi - 1\} \log(1 - y_t)$$

Entendiendo que se definen y_t^* y μ_t^* como:

$$y_t^* = \log \left(\frac{y_t}{1 - y_t} \right)$$

$$\mu_t^* = \psi(\mu \phi) - \psi((1 - \mu)\phi)$$

Donde ψ es la función digamma.

La función de Score U se obtiene diferenciando la función de log-verosimilitud con respecto a los parámetros desconocidos (β, ϕ) .

Expresada en forma matricial, la función de Score para β_i se expresa como:

$$U_{\beta}(\beta, \phi) = \phi X^T (y_t^* - \mu_t^*)$$

Por otro lado, la función de Score para ϕ esta expresada de la siguiente forma:

$$U_{\phi}(\beta, \phi) = \sum_{t=1}^n \{ \mu_t (y_t^* - \mu_t^*) + \log(1 - y_t) - \psi((1 - \mu_t)\phi) + \psi(\phi) \},$$

Sea:

- X una matriz de tamaño $n \times k$, cuya t -ésima fila es x_t^T .
- $T = \text{diag} \left\{ \frac{1}{g'(\mu_1)}, \dots, \frac{1}{g'(\mu_n)} \right\}$.
- $y^* = (y_1^*, \dots, y_n^*)^T$.
- $\mu^* = (\mu_1^*, \dots, \mu_n^*)^T$.

Para construir la matriz de información de Fisher se define:

$$W = \text{diag}\{\omega_1, \dots, \omega_n\}, \quad \text{con}$$

$$\omega_t = \phi \{ \psi'(\mu_t \phi) + \psi'[(1 - \mu_t)\phi] \} \frac{1}{[g'(\mu_t)]^2},$$

Aquí:

- Sea $c = (c_1, \dots, c_n)^T$, con $c_t = \phi \{ \psi'(\mu_t \phi) \mu_t - \psi'[(1 - \mu_t)\phi] (1 - \mu_t) \}$, además $\psi'(\cdot)$ denota la función trigamma.
- Sea $D = \text{diag}\{d_1, \dots, d_n\}$, con $d_t = \psi(\mu_t \phi) \mu_t^2 + \psi'[(1 - \mu_t)\phi] (1 - \mu_t)^2 - \psi'(\phi)$.

El siguiente paso es obtener la matriz de información de Fisher para (β, ϕ) .

$$K = K(\beta, \phi) = \begin{pmatrix} K^{\beta\beta} & K^{\beta\phi} \\ K^{\phi\beta} & K^{\phi\phi} \end{pmatrix}$$

Debido a que los estimadores no tienen una forma cerrada, se obtienen maximizando numéricamente la función de log-verosimilitud por medio de un algoritmo de optimización no lineal, los cuales requieren de la especificación de valores iniciales que se usan en el esquema iterativo. En Cribari-Neto (2004), se sugiere usar como valor inicial la estimación de β para la variable respuesta transformada vía MCO. Para la precisión se sugiere tomar como valor inicial:

$$\frac{1}{n} \sum_{i=1}^n \frac{\hat{\mu}_t(1 - \hat{\mu}_t)}{\hat{\sigma}_t^2}$$

Entonces se puede demostrar que la matriz de información de Fisher está dada por:

$$K = K(\beta, \phi) = \begin{pmatrix} K_{\beta\beta} & K_{\beta\phi} \\ K_{\phi\beta} & K_{\phi\phi} \end{pmatrix}$$

Donde $K_{\beta\beta} = \phi X^T W X$, $K_{\beta\phi} = K_{\phi\beta}^T = X^T T c$ y $K_{\phi\phi} = \text{tr}(D)$. Nótese que los parámetros β y ϕ no son ortogonales, en contraste con lo que se encuentra en la clase de modelos lineales generalizados (McCullagh & Nelder, 1989).

Bajo las condiciones usuales de regularidad para la estimación de máxima verosimilitud, para un tamaño de muestra grande, la distribución conjunta de los estimadores converge en distribución a una distribución normal de la siguiente forma:

$$\begin{pmatrix} \hat{\beta} \\ \hat{\phi} \end{pmatrix} \sim N \left(\begin{pmatrix} \beta \\ \phi \end{pmatrix}, K^{-1} \right)$$

Donde $\hat{\beta}$ y $\hat{\phi}$ son los estimadores de máxima verosimilitud de β y ϕ , respectivamente.

4. Funciones de Enlace

Las funciones de enlace sirven para conectar la media de la variable que se está estudiando con una combinación lineal de las variables explicativas. En un modelo de regresión beta, se usan dos funciones de enlace principales:

4.1. Enlace para la Media (μ)

$$\eta_i = g_1(\mu_i) = \mathbf{x}'_i \boldsymbol{\beta}$$

Aquí, η_i es el predictor lineal, \mathbf{x}_i es un vector de variables independientes y $\boldsymbol{\beta}$ es el vector de coeficientes que hay que estimar.

En este caso, se está usando una **función de enlace logit**:

$$\mu_i = \frac{e^{\mathbf{x}'_i \boldsymbol{\beta}}}{1 + e^{\mathbf{x}'_i \boldsymbol{\beta}}}$$

4.2. Enlace para la Precisión (ϕ)

La precisión es un parámetro adicional en la distribución beta que controla la varianza de la distribución.

$$\phi_i = g_2(\phi_i) = \mathbf{z}'_i \boldsymbol{\gamma}$$

Aquí, \mathbf{z}_i puede ser el mismo conjunto de variables independientes \mathbf{x}_i o un conjunto diferente, y $\boldsymbol{\gamma}$ es otro conjunto de coeficientes.

En la ecuación proporcionada, se sugiere una función de enlace logarítmica para ϕ :

$$\phi_i = e^{\mathbf{z}'_i \boldsymbol{\gamma}}$$

4.3. Estimación de Parámetros

Para ajustar el modelo, se usa la función de máxima verosimilitud. La verosimilitud del modelo beta, dados los parámetros $\boldsymbol{\beta}$ y $\boldsymbol{\gamma}$, es:

$$L(\boldsymbol{\beta}, \boldsymbol{\gamma}) = \prod_{i=1}^n f(y_i | \boldsymbol{\beta}, \boldsymbol{\gamma})$$

La log-verosimilitud está dada por:

$$l(\boldsymbol{\beta}, \boldsymbol{\gamma}) = \sum_{i=1}^n \ln(h(y_i)) + \left(\frac{e^{\mathbf{x}'_i \boldsymbol{\beta}}}{1 + e^{\mathbf{x}'_i \boldsymbol{\beta}}} \right) T(y_i) - \rho(\mu_i, \phi_i)$$

Donde:

- $h(y_i)$ es una función que depende de los datos.
- $T(y_i)$ es una transformación de y_i .
- $\rho(\mu_i, \phi_i)$ es una función que combina μ_i y ϕ_i .

5. Modelo de Regresión Beta Espacio-Temporal

5.1. Especificación del Modelo

El modelo beta espacial propuesto por Kalhori y Mohammadzadeh (2016) se describe así:

$$\text{logit}(E(Y_i | \beta, \tau)) = \sum_{j=0}^{k-1} X_{ij} \beta_j + \tau_i$$

Donde:

- Y_i es la tasa de suicidio.
- X_{ij} son las variables explicativas.
- β_j son los coeficientes de esas variables.
- $\tau_i = \tau(s_i)$ es el coeficiente relacionado con la ubicación espacial.

5.2. Componentes del Modelo

- **Efecto Aleatorio Espacial (τ_i):**
 - Este término captura la variabilidad espacial y se trata como un efecto aleatorio. No se asume una distribución específica, lo que da más flexibilidad para modelar la dependencia espacial (Cressie and Wikle, 2011; Banerjee et al., 2014).
- **Modelos de Cuasi-Verosimilitud:**
 - Estos modelos son útiles cuando no es fácil especificar la distribución exacta. Permiten hacer inferencias estadísticas usando una función de cuasi-verosimilitud en vez de una función de verosimilitud estándar (McCullagh and Nelder, 1989; Wedderburn, 1974).

5.3. Funciones de Media y Varianza

Función de Media ($\mu(y)$) y **Varianza** ($v(\mu)$):

$$\mu(y) = \frac{1}{1 + \exp(-\beta)}$$

$$v(\mu) = \mu^2(1 - \mu^2)$$

Estas funciones ayudan a entender la relación entre la media y la varianza en el modelo, esenciales para datos proporcionales o de tasa (Ferrari and Cribari-Neto, 2004; Smithson and Verkuilen, 2006).

5.4. Estructura de Correlación Espacio-Temporal

■ **Separabilidad:**

- La estructura de correlación espacio-temporal se asume separable, lo que significa que la correlación entre dos puntos en el espacio y el tiempo puede descomponerse en correlaciones espaciales y temporales independientes (Cressie and Wikle, 2011).

■ **Covarianza:**

- La estructura de covarianza se define así:

$$\text{Cov}(\tau) = \delta V^{-1/2} R V^{-1/2}$$

$$R = I_s \otimes W_0 + W \otimes I_t$$

Donde R es la matriz de covarianza, V es una matriz que muestra las asociaciones entre los sitios, W muestra la relación temporal entre los sitios y W_0 muestra la relación espacial (Cressie and Wikle, 2011; Banerjee et al., 2014).

5.5. Estimación de Parámetros

■ **Algoritmos RS y CG:**

- Para estimar los parámetros, se usan algoritmos de mínimos cuadrados ponderados penalizados mediante la verosimilitud penalizada y algoritmos iterativos (Ruppert et al., 2003).

$$t = -\frac{1}{4} \sum_{i=1}^n \sum_{j=1}^{J_i} T_i G_{ij}(\lambda_j) r_{jk}$$

■ **Métodos de Suavizamiento:**

- Se usan métodos no paramétricos, como los splines cúbicos, para ajustar de forma flexible la variable independiente (Ruppert et al., 2003; Wood, 2017).

$$Q_2(g) = \sum_j \left(w_j (u_j - g(x_j))^2 + \lambda \int_a^b (g'')^2 \right)$$

Donde $g(x)$ es una función de dos veces diferenciable en el intervalo $[a, b]$.

■ **Selección de Modelos GAM:**

- Se usan librerías de software como `gamlss` en R para seleccionar y ajustar modelos de suavizamiento, utilizando funciones como `pb()`, `cs()`, y `lo()` (Rigby and Stasinopoulos, 2007; Team, 2020).

5.6. Modelo GAM (Generalized Additive Model)

5.6.1. Justificación Teórica

El Modelo Aditivo Generalizado (GAM) es una extensión de los Modelos Lineales Generalizados (GLM) que permite modelar relaciones no lineales entre las variables predictoras y la variable de respuesta. Su flexibilidad para capturar relaciones complejas lo hace ideal para el análisis de datos espaciales y temporales.

Ventajas de los GAM:

- **Flexibilidad:** Permiten modelar relaciones no lineales mediante funciones suaves, como splines.
- **Interpretabilidad:** Aunque más complejos que los GLM, los GAM siguen siendo relativamente interpretables.
- **Capacidad para manejar datos espaciales y temporales:** Los GAM pueden incorporar funciones suaves bivariadas para capturar la dependencia espacial y temporal.

Literatura Relacionada:

- **Hastie y Tibshirani (1990):** Introdujeron los GAM como una herramienta flexible para el análisis de datos complejos.
- **Wood (2017):** Describió técnicas avanzadas para la estimación y validación de GAMs, incluyendo aplicaciones en datos espaciales y temporales.
- **Ferrari y Cribari-Neto (2004):** Discutieron la regresión beta, que es relevante para la modelación de proporciones y tasas, y puede combinarse con GAMs para análisis complejos.

5.6.2. Aplicaciones de GAM en Análisis Espacial-Temporal

Estudios Previos:

- **Smith et al. (2015):** Utilizaron GAMs para analizar la distribución espacial de enfermedades infecciosas, demostrando su utilidad en el contexto de salud pública.
- **Jones y Wakefield (2013):** Aplicaron GAMs para modelar tasas de criminalidad, capturando variaciones espaciales y temporales de manera efectiva.

Relevancia para el Estudio Actual: La elección de un GAM para modelar la tasa de suicidio en Colombia entre 2014 y 2019 se justifica por las siguientes razones:

1. **Captura de Relaciones Complejas:** La tasa de suicidio puede estar influenciada por múltiples factores socioeconómicos de manera no lineal.
2. **Variabilidad Espacial y Temporal:** Las tasas de suicidio varían tanto en el tiempo como en el espacio, lo que requiere un modelo capaz de capturar estas dependencias.
3. **Flexibilidad y Precisión:** Los GAMs proporcionan una mayor flexibilidad para ajustar los datos y pueden mejorar la precisión de las predicciones en comparación con los modelos lineales tradicionales.

5.6.3. Estructura del Modelo GAM

Un Modelo Aditivo Generalizado (GAM) permite modelar relaciones no lineales entre las variables predictoras y la variable de respuesta. La estructura general del modelo GAM se define como:

$$g(\mu_i) = \beta_0 + f_1(x_{i1}) + f_2(x_{i2}) + \cdots + f_p(x_{ip})$$

donde:

- $g(\cdot)$ es la función de enlace.
- μ_i es la esperanza de la variable respuesta Y_i , $\mu_i = \mathbb{E}(Y_i)$.
- β_0 es el intercepto.
- $f_j(\cdot)$ son funciones suaves (por ejemplo, splines) que representan el efecto de la j -ésima variable predictora x_{ij} en μ_i .
- x_{ij} es el valor de la j -ésima variable predictora para la i -ésima observación.
- p es el número de variables predictoras.

Componentes del Modelo GAM:**a. Función de Enlace $g(\cdot)$:**

- Conecta la media de la distribución de la variable respuesta Y con una combinación aditiva de las funciones de las variables predictoras.
- En el caso de datos de proporciones o tasas, la función de enlace comúnmente utilizada es la función logit:

$$g(\mu_i) = \log\left(\frac{\mu_i}{1 - \mu_i}\right)$$

b. Variable de Respuesta Y :

- Y_i es la variable de respuesta para la i -ésima observación. En este estudio, representa la tasa de suicidio.

c. Variables Predictoras X :

- x_{ij} son las variables explicativas que pueden incluir indicadores socioeconómicos, como el índice de salud (Health), la tasa de desempleo (Unemployment), el PIB per cápita (GDP), el coeficiente de Gini (Gini), entre otros.

d. Funciones Suaves $f_j(\cdot)$:

- Las funciones $f_j(\cdot)$ son funciones no paramétricas que permiten capturar relaciones no lineales entre las variables predictoras y la variable de respuesta.
- Se suelen utilizar splines cúbicos, splines de P-splines, funciones de base radial, etc.

5.6.4. Especificación del Modelo GAM para el Estudio

Para el análisis de la tasa de suicidio en Colombia, el modelo GAM puede ser especificado de la siguiente manera:

$$g(\mu_i) = \beta_0 + f_1(\text{Health}_i) + f_2(\text{Unemployment}_i) + f_3(\text{GDP}_i) + f_4(\text{Gini}_i) + s(\text{Lat}_i, \text{Long}_i)$$

donde:

- μ_i es la tasa de suicidio esperada en el i -ésimo departamento.

- $g(\mu_i) = \log\left(\frac{\mu_i}{1-\mu_i}\right)$ es la función de enlace logit.
- β_0 es el intercepto.
- $f_1(\text{Health}_i)$ es la función suave para el índice de salud.
- $f_2(\text{Unemployment}_i)$ es la función suave para la tasa de desempleo.
- $f_3(\text{GDP}_i)$ es la función suave para el PIB per cápita.
- $f_4(\text{Gini}_i)$ es la función suave para el coeficiente de Gini.
- $s(\text{Lat}_i, \text{Long}_i)$ es una función suave bivariada para capturar la dependencia espacial en términos de latitud y longitud del i -ésimo departamento.

6. Resultado prácticos

6.1. Variable respuesta.

Tasa de suicidio en Colombia calculada como la razón entre número de defunciones asociadas a lesiones autoinfligidas intencionalmente (suicidios) y el número de habitantes en cada departamento (utilizando las proyecciones de población DANE) discriminado por año.

6.2. Análisis exploratorio espacial.

Los datos mencionados a lo largo de esta investigación fueron extraídos en su totalidad del Departamento Administrativo Nacional de Estadística (DANE). Para este proyecto se utilizó la información comprendida entre 2014 y 2019. Vale aclarar que esta información es pública y de libre acceso. A continuación se observan las variables explicativas que fueron contempladas para este estudio.

Tabla 2: Descripción de Variables

Variable	Descripción	Fuente
PIB	Valor total de todos los bienes y servicios finales producidos por un país en un período determinado. Se utiliza el término “precios constantes” para ajustar los valores del PIB y tener en cuenta los cambios en los precios con el tiempo, lo que permite una medición más precisa de la actividad económica real	DANE
IDH	El Índice de Desarrollo Humano (IDH) es una medida que sintetiza aspectos fundamentales del desarrollo humano, tales como salud, educación y nivel de vida (medido por factores como la esperanza de vida, la tasa de alfabetización y el PIB per cápita). Se enfoca en la posesión de conocimientos y la capacidad de disfrutar de un nivel de vida adecuado. El IDH se calcula como la media geométrica de los índices normalizados de cada una de estas tres dimensiones.	DANE
Health	El índice de salud se basa en uno de los componentes del Índice de Desarrollo Humano (IDH), que es la esperanza de vida al nacer. Este indicador representa la media de años que se espera que viva una determinada población desde el momento de su nacimiento.	DANE
Desempleo	La tasa de desempleo es un indicador económico que refleja el porcentaje de personas que están en edad laboral y buscan empleo activamente pero no lo encuentran dentro de la fuerza laboral.	DANE
Gini	El coeficiente de Gini es un indicador de desigualdad que evalúa la concentración de la riqueza en una región geográfica específica. Se calcula midiendo la distancia entre la curva de Lorenz, que representa la distribución real de la riqueza, y la línea de equidistribución, que representaría una distribución perfectamente igualitaria. El coeficiente de Gini varía entre cero y uno, donde cero indica igualdad absoluta (todos tienen la misma cantidad de riqueza) y uno indica desigualdad absoluta (toda la riqueza está concentrada en una sola persona o entidad).	DANE

Se cuenta con la información en su totalidad de 32 departamentos y Distrito Capital(Bogotá D.C.) que contempla en su totalidad el área geográfica de Colombia. Únicamente, se tienen dos registros en los cuales la tasa de suicidio registra cero por la no disponibilidad del dato, por ende, estos dos datos asociados a Guainía 2015 y San Andres 2019 con imputación simple fueron calculados como el promedio entre las observaciones registradas cronologicamente antes del mismo.

En este mismo sentido, a continuación se observa el comportamiento de la tasa de suicidio para los años en estudio 2014-2019

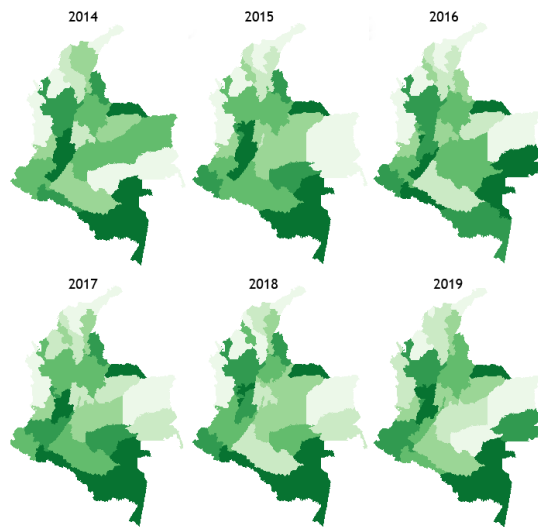


Figura 1: Comportamiento de la tasa de suicidio en los años de 2014 a 2019

En el grafico se observa que hay departamentos en los cuales independientemente del año en observación, siempre se presenta una alta/baja frecuencia en cuanto a la tasa de suicidio, lo cual supone unos cambios no abruptos de los departamentos a lo largo de los años en estudio, frecuente en variables de corte demografico, adicionalmente, se observa que para algunos años la frecuencia de los departamentos es, en algunos casos, muy similar a la que registran sus departamentos aledaños, por ejemplo, en lo correspondiente a la region de amazonia se observa una frecuencia mayor para los departamentos asociados a esta región, en contra parte, la región del Caribe, tiene sus departamentos aparentemente con una tasa de suicidio menor, señal de una relación entre la ubicación geografica y la tasa de suicidio.

El año 2014 registra la menor tasa de suicidio promedio con 5.17, por otra parte, el año 2018 registro la mayor tasa de suicidio promedio de la serie en estudio con 6.93. En términos de departamento, en promedio el departamento de Choco registra la menor tasa de suicidio con 2.25, mientras que el departamento de Vaupes tiene la mayor tasa de suicidio promedio con 24.7. En el contexto espacio - tiempo, la menor tasa de suicidio se registro en vichada año 2016 con 0.974, el departamento

de Vaupés en el año 2018 presento la mayor tasa de suicidio con 41.7.

Sin embargo, pese a que se cuente con la información completa a nivel departamental, en cuanto a la tasa de suicidio, la limitación en la información se encuentre en las variables explicativas, puesto que, solo se tiene información completa tanto en la variable respuesta como en la variable explicativa en 24 departamentos, los cuales, se les podrá modelar su comportamiento en función de sus predictores en el modelo.

El planteamiento de esta metodología supone relación entre la variables de interés y su ubicación geográfica, es decir, debe existir una correlación espacial y esta relación, debe perdurar a lo largo de los años, con este objetivo, se realiza la prueba de hipótesis de Moran. El sistema a ser contrastado es:

$$\begin{cases} H_0 : \text{La variable de interes tiene correlación espacial} \\ H_1 : \text{La variable de interes no tiene correlación espacial} \end{cases}$$

Encontrando los siguiente resultados:

Tabla 3: Test de Correlación Espacial de Moran

<i>Año</i>	<i>Estadístico Moran</i>	<i>p-value</i>
2014	-0.047411495	0.5919
2015	0.124865084	0.01468
2016	0.14716004	0.04925
2017	0.13250454	0.04243
2018	0.01514352	0.1854
2019	0.21643726	0.009867

A partir de los resultados, a un nivel de significancia del 5 %, se observa que existe suficiente evidencia estadísticamente para concluir que la variable tasa de suicidios tiene correlación espacial, dado que el p-valor es menor al nivel de significancia en todos los años a excepción de los periodos 2014 y 2018, por lo tanto, se rechaza la hipótesis nula. El valor positivo en el estadístico de Moran hace referencia a una autocorrelación positiva, por lo tanto, para todos los años diferentes a 2014, la tasa de suicidio tiene una similitud con el compartimiento de este en los departamentos aledaños.

7. Estimación de modelos

7.1. SAR(Spatial autoregressive model)

Los modelos SAR (Autorregresivo Espacial) permiten capturar la influencia de las variables explicativas tanto en una su propia unidad espacial (Departamento) como de las demás unidades espaciales, de esta manera, se logra capturar la autocorrelación espacial en los datos.

El modelo SAR se define de la siguiente manera:

$$Y = \rho WY + X\beta + \varepsilon$$

Donde:

- Y es un vector $n \times 1$ de la variable dependiente.
- X es una matriz $n \times k$ de las variables independientes.
- β es un vector $k \times 1$ de coeficientes.
- W es una matriz de pesos espaciales $n \times n$.
- ρ es el parámetro de autocorrelación espacial.
- ε es un vector $n \times 1$ de errores aleatorios.

Por medio del entorno y lenguaje de programación Rstudio, haciendo uso de la librería *spdep*, se realizó la estimación de los modelos SAR(Autorregresivo Espacial), para cada uno de los años con los que se cuenta información, probando con el modelo saturado, modelo simple con una sola covariable y diferentes combinaciones de variables explicativas, se evaluara su desempeño por medio del MAPE(Mean absolute percentage error).

El modelo con mejor desempeño fue siempre el saturado, por ende, a continuación, se centrará el análisis en el desempeño de cada modelo simple por año.

Tabla 4: Evaluación de Modelos para Verificar Menor MAPE

<i>Año</i>	<i>Modelo</i>	<i>MAPE</i>
2014	Health * Desempleo * Gini	11.66 %
	Health * Desempleo * IDH	16.55 %
	Health * PIB	27.84 %
	PIB + IDH + Health + Desempleo + Gini	30.28 %
	Health * IDH	30.67 %
	Health	31.26 %
2015	Health * Desempleo * Gini	14.35 %
	Health * Desempleo * IDH	15.63 %
	Health * PIB	26.61 %
	PIB + IDH + Health + Desempleo + Gini	27.81 %
	Health	29.93 %
	Health * IDH	30.34 %
2016	Health * Desempleo * Gini	19.78 %
	Health * Desempleo * IDH	23.33 %
	PIB + IDH + Health + Desempleo + Gini	26.32 %
	Health * PIB	26.71 %
	Health * IDH	28.76 %
	Health	28.87 %
2017	Health * Desempleo * IDH	11.67 %
	Health * Desempleo * Gini	13.44 %
	Health * PIB	16.71 %
	PIB + IDH + Health + Desempleo + Gini	19.48 %
	Health * IDH	20.51 %
	Health	21.32 %
2018	Health * Desempleo * Gini	15.37 %
	Health * Desempleo * IDH	15.36 %
	Health * PIB	21.22 %
	PIB + IDH + Health + Desempleo + Gini	22.90 %
	Health * IDH	24.82 %
2019	Health * Desempleo * IDH	16.06 %
	Health * PIB	16.45 %
	Health * Desempleo * Gini	25.44 %
	Health * IDH	22.06 %
	PIB + IDH + Health + Desempleo + Gini	22.31 %
	Health	22.65 %

Encontrando que, los modelos que trabajan con la interacción del Índice de salud, tasa de desempleo y Gini y los modelos que utilizan la interacción índice de salud, tasa de desempleo e IDH parecen ser consistentemente buenos en varios años, frecuentemente teniendo los MAPE más bajos, siendo el primero el mejor para los años 2014, 2015, 2016, 2018, por su parte el modelo con la interacción Índice de salud, tasa de desempleo e IDH tienen el menor MAPE para los años 2017 y 2019. Los años 2014 y 2017 parecen tener una mejor precisión en general (MAPE más bajos) comparados con otros años. Los modelos más complejos que incluyen todas las variables (modelo saturado) no siempre proporcionan los MAPE más bajos, sugiriendo que la inclusión de más variables no necesariamente mejora la precisión del modelo. Al validar la significancia de las variables, ninguna resultó estadísticamente significativa, por último, en cuanto a la validación del modelo, en la siguiente tabla se observa que todos los modelos pasan el supuesto de normalidad de los residuales

Tabla 5: Pruebas de Normalidad Modelos SAR

<i>Año</i>	<i>Shapiro-Wilk</i>	<i>Anderson-Darling</i>	<i>Lilliefors</i>
2014	0.8417	0.4954	0.3364
2015	0.7237	0.5789	0.5219
2016	0.7746	0.8945	0.9233
2017	0.8435	0.8766	0.701
2018	0.6717	0.614	0.7708
2019	0.3327	0.2954	0.4169

7.2. SEM(Spatial Error model)

Los modelos SEM (Error Espacial) permiten capturar el efecto espacial de las variables, a comparación del modelo SAR, esta autocorrelación espacial se modela a través de un término de error espacial, el cual refleja la autocorrelación espacial no explicada por las variables explicativas.

La estructura del modelo SEM tiene la forma:

$$Y = X\beta + \mathbf{U}$$

Donde:

- Y es un vector $n \times 1$ de la variable dependiente.
- X es una matriz $n \times k$ de las variables independientes.
- β es un vector $k \times 1$ de coeficientes.
- \mathbf{U} es un vector $n \times 1$ de errores aleatorios.

El término \mathbf{U} representa el error aleatorio en el modelo, pero en un modelo SEM, este error también incluye un componente de error espacial:

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

Donde:

- λ es el parámetro de autocorrelación espacial.
- W es una matriz de pesos espaciales $n \times n$.
- ε es un vector $n \times 1$ de errores aleatorios adicionales.

Este término $\lambda W\mathbf{U}$ representa el componente de error espacial en el modelo, que captura la autocorrelación espacial en los errores de predicción. El término ε representa el componente de error aleatorio adicional en el modelo.

Por medio del entorno y lenguaje de programación Rstudio, haciendo uso de la librería *spdep*, se realizó la estimación de los modelos SEM(Error Espacial) para cada uno de los años probando diferentes combinaciones de las variables explicativas.

Tabla 6: MAPE Modelos SEM.

<i>Año</i>	<i>Variables explicativas</i>	<i>MAPE</i>
2014	Health	29.71 %
2015	Health	29.88 %
2016	Health	28.88 %
2017	Health	21.22 %
2018	IDH	23.72 %
2019	Health	22.65 %

Para todos los años menos el 2018, el Índice de salud es la variable que genera un mejor rendimiento en cuanto el desempeño de los modelos, únicamente en el año 2018 el Índice de desarrollo humano fue el que generó un menor MAPE, lo cual, esclarece la relevancia que tiene el índice de salud en la tasa de suicidio, sin embargo, al validar la significancia de las variables, ninguna resultó estadísticamente significativa. Por último, en cuanto a la validación del modelo, se encontró que todos los modelos cumplen con el supuesto de normalidad de los residuales, lo que sugiere que las discrepancias entre las predicciones del modelo y los valores reales se distribuyen normalmente. Aunque las variables no demostraron ser significativas, el modelo parece capturar adecuadamente la variabilidad de los datos y ajustarse bien a la distribución observadas.

Tabla 7: Pruebas de normalidad modelos SEM.

<i>Año</i>	<i>Shapiro-Wilk</i>	<i>Anderson-Darling</i>	<i>Lilliefors</i>
2014	0.8187	0.5429	0.3651
2015	0.5266	0.4064	0.5307
2016	0.5882	0.7765	0.7995
2017	0.6817	0.6691	0.7756
2018	0.3088	0.2734	0.6836
2019	0.3381	0.3244	0.5788

7.3. GAM(Generalized Additive Model)

Los Modelos GAM son una extensión de los Modelos Lineales Generalizados (GLM) que permiten modelar relaciones no lineales entre las variables predictoras y la variable respuesta. A diferencia de los modelos lineales tradicionales, que asumen una relación lineal entre las variables predictoras y la respuesta, los GAMs permiten que estas relaciones sean no paramétricas, es decir, no se especifica una forma funcional específica de antemano.

$$\text{logit}(\mu_i) = \beta_0 + f_1(x_{i1}) + f_2(x_{i2}) + \dots + f_p(x_{ip})$$

Donde:

- $\text{logit}(\cdot)$ representa la función logit.
- μ_i es el valor esperado de la variable respuesta para la observación i .
- β_0 es el intercepto.
- $f_j(\cdot)$ es una función suave de la variable predictora x_j .
- x_{ij} es el valor de la variable predictora j para la observación i .
- p es el número de variables predictoras.

Haciendo uso de la librería *mgcv* se ajusta un modelo de regresión generalizada aditiva (GAM) para predecir la tasa de suicidios para cada uno de los años y una función suave de las coordenadas geográficas para capturar la correlación espacial. La función de enlace logit y la distribución Beta se utilizan para modelar la relación entre las variables predictoras y la variable de respuesta, teniendo en cuenta la naturaleza proporcional de la tasa de suicidios. Se probaron diferentes modelos utilizando las variables de manera individual y considerando sus interacciones. A partir del MAPE, se puede cuantificar la precisión de las estimaciones para identificar el mejor modelo en términos de estimación.

Tabla 8: MAPE Modelos GAM.

Año	Modelo	MAPE
<i>2014</i>	Health * Desempleo * Gini	11,50 %
	Health * Desempleo * PIB	13,34 %
	Health * IDH	14,59 %
	Health * PIB	16,60 %
	PIB + IDH + Health + Desempleo + Gini	20,85 %
	Health	22,96 %
<i>2015</i>	Health * Desempleo * Gini	12,3 %
	Health * Desempleo * PIB	13,4 %
	Health * IDH * Desempleo	13,75 %
	Health * PIB	16,15 %
	PIB + IDH + Health + Desempleo + Gini	24,67 %
	Health	27,34 %
<i>2016</i>	Health * Desempleo * PIB	11,23 %
	Health * PIB	12,68 %
	Health * Gini * Desempleo	17,85 %
	PIB + IDH + Health + Desempleo + Gini	23,64 %
	Health	24,21 %
<i>2017</i>	Health * Desempleo * PIB	7,39 %
	Health * PIB	11,1 %
	Health * Gini * Desempleo	12,12 %
	PIB + IDH + Health + Desempleo + Gini	18,1 %
	Health	21,2 %
<i>2018</i>	Health * Desempleo * PIB	4,27 %
	Health * PIB	10,21 %
	Health * Gini * Desempleo	11,5 %
	Health	20,1 %
	PIB + IDH + Health + Desempleo + Gini	21,1 %
<i>2019</i>	Health * Desempleo * PIB	3,85 %
	Health * PIB	8,94 %
	Health * Gini * Desempleo	11,30 %
	Health	16,40 %
	PIB + IDH + Health + Desempleo + Gini	17,62 %

Encontrando que, el modelo que cuantifica la interacción entre el índice de salud, desempleo y PIB aparece como el mejor en los años 2016, 2017, 2018, y 2019. Para los años 2014 y 2015, el modelo con la interacción del índice de Gini, índice de salud y desempleo también tiene un rendimiento notable, donde parece haber una tendencia a la disminución del MAPE a medida que avanzan los años, especialmente notable para el modelo entre la interacción de Indicador de salud, desempleo y PIB, lo que sugiere que este modelo se vuelve más preciso con el tiempo, registrando para el año 2019 un MAPE de 3.85 %, esta interacción resulta importante para la estimación de la tasa de suicidio, algo que no ocurría cuando se trabaja el modelo con las variables de manera marginal, además, independientemente del año, el modelo con mejor precisión en las estimaciones tiene como mínimo la interacción entre el Índice de salud y la tasa de desempleo, ratificando la importancia de estas covariables.

8. GAM por año

Como se observó en el apartado anterior, el modelo que utiliza las relaciones de las variables Indicador de salud, Desempleo y PIB junto con sus respectivas interacciones, es el modelo que genera mejores predicciones para los periodos comprendidos entre 2016 - 2019, previo a estos años en los periodos 2014 - 2015 es el segundo mejor modelo en términos de precisión. En este mismo sentido, se estima este modelo para todos los años en observación y se estudiará las dinámicas de las relaciones a lo largo del tiempo.

La estimación del modelo para los diferentes periodos arroja los resultados observados en la tabla 2. Estos resultados muestran los coeficientes obtenidos con la función exponencial aplicada para su respectiva interpretación. Adicionalmente, los coeficientes estadísticamente significativos al 10 % se encuentran indicados entre paréntesis.

Tabla 9: Exponencial (β) por año

<i>Año</i>	<i>Health</i>	<i>Desempleo</i>	<i>PIB</i>	<i>Heal:Des</i>	<i>Heal:PIB</i>	<i>Dese:PIB</i>	<i>H:D:P</i>
2014	0.00001	(0.05086)	0.99	(30.82783)	1.0007	1.00019	0.99979
2015	0.00000	(0.018)	0.99	(101.191)	1.0019	(1.0003)	(0.9997)
2016	0.00000	0.00779	0.99	253.02671	1.00148	1.00022	0.99975
2017	0.00003	(0.08573)	0.999	(17.51186)	1.00022	1.00011	0.99988
2018	8.76616	(0.11234)	1.0006	(12.09297)	0.99927	1.00004	0.99996
2019	114.11738	0.42979	1.0005	2.43328	0.99943	1.00004	0.99996

Por otra parte, en la siguiente tabla se observa el EDF con su respectivo p - *value*, lo cual nos permite cuantificar la complejidad de la función spline que captura la correlación espacial por medio de la expresión $s(\text{Latitud}, \text{Longitud})$. Un EDF mayor indica una función más compleja, que mejora la relación entre las variables explicativas y la variable respuesta en el modelo.

Tabla 10: EDF por año

<i>Año</i>	<i>edf</i>	<i>p-value</i>
2014	4.446057	0.000
2015	4.134496	0.033
2016	5.529070	0.003
2017	5.129856	0.000
2018	7.228014	0.000
2019	7.228014	0.000

La variable más significativa dentro del conjunto de variables explicativas es el desempleo, demostrando su relevancia al ser significativa en todos los años en que al menos un parámetro del modelo lo fue. La importancia de esta variable se acentúa al considerar que indicadores como la salud o el PIB resultan significativos únicamente cuando interactúan con la tasa de desempleo. La relación entre la tasa de suicidio y la tasa de desempleo es compleja y ha sido estudiada en diversos contextos. La evidencia sugiere que existe una asociación positiva entre el desempleo y las tasas de suicidio. Estudios han demostrado que el desempleo puede contribuir significativamente a las tasas de suicidio debido a factores como la presión económica, la pérdida de estatus y estructura social, y el aumento del estrés y la inseguridad laboral.

Los resultados muestran que la exponencial de los parámetros asociados a la tasa de desempleo es menor que 1 en los casos significativos, lo que sugiere una relación inversa entre el desempleo y la tasa de suicidios. En otras palabras, un incremento en la tasa de desempleo se vincula con una disminución en la tasa de suicidios, según los modelos analizados. Esto es contraintuitivo dado que la mayoría de los estudios sugieren una relación positiva entre el desempleo y las tasas de suicidio, lo cual podría seguir que períodos de alto desempleo podrían estar asociados con una disminución en la tasa de suicidios, posiblemente debido a un aumento en el apoyo social o familiar, o reducción de la presión y el estrés en momentos de empleo bajo.

Destaca el año 2015 como el periodo en el que el desempleo tuvo el mayor impacto en la reducción de la tasa de suicidios, evidenciado por el valor más bajo de la exponencial. Por el contrario, en 2017, el desempleo mostró el menor impacto en la tasa de suicidios entre los periodos analizados con parámetros significativos.

Las interacciones juegan un papel crucial en el análisis. Como se menciona, tanto el indicador de salud como el PIB no son significativos cuando se consideran individualmente; sin embargo, al incluir su interacción con el desempleo, estas variables resultan significativas en ciertos años. Además, es importante destacar que la interacción entre el indicador de salud, el desempleo y el PIB es la única que tiene un efecto inverso en la tasa de suicidios. En contraste, el resto de las interacciones significativas reflejan una asociación positiva con la tasa de desempleo. Este hallazgo subraya la importancia de examinar las interacciones entre variables para una comprensión más completa de su impacto en la tasa de suicidios.

Por otra parte, la incorporación de $s(LATITUD, LONGITUD)$ mejora significativamente el ajuste del modelo. Esto sugiere la existencia de patrones espaciales importantes en los datos de suicidios, los cuales son capturados por esta función suave. Para todos los períodos analizados, el edf aumenta a lo largo del tiempo, lo que indica relaciones más complejas que deben ser modeladas. Estas relaciones capturan la variación en la tasa de suicidios en función de la latitud y la longitud, las cuales no son aleatorias y son adecuadamente capturadas por el término suave en el modelo.

Por último, en la siguiente tabla se observa el MAPE obtenido para cada uno de los modelos analizados.

Tabla 11: MAPE Modelos GAM por año

Año	Modelo	MAPE
2014	$\log(\mu) = 0.3223 - 11.7343 \cdot \text{Health} - 2.9787 \cdot \text{desempleo} - 0.0007 \cdot \text{PIB} + 3.4284 \cdot \text{Health} \cdot \text{desempleo} + 0.0007 \cdot \text{Health} \cdot \text{PIB} + 0.0002 \cdot \text{desempleo} \cdot \text{PIB} - 0.0002 \cdot \text{Health} \cdot \text{desempleo} \cdot \text{PIB} + f(\text{LATITUD}, \text{LONGITUD})$	13.34 %
2015	$\log(\mu) = 11.7789 - 24.5366 \cdot \text{Health} - 4.0411 \cdot \text{desempleo} - 0.0017 \cdot \text{PIB} + 4.6170 \cdot \text{Health} \cdot \text{desempleo} + 0.0019 \cdot \text{Health} \cdot \text{PIB} + 0.0003 \cdot \text{desempleo} \cdot \text{PIB} - 0.0003 \cdot \text{Health} \cdot \text{desempleo} \cdot \text{PIB} + f(\text{LATITUD}, \text{LONGITUD})$	13.39 %
2016	$\log(\mu) = 19.8549 - 33.6680 \cdot \text{Health} - 4.8555 \cdot \text{desempleo} - 0.0013 \cdot \text{PIB} + 5.5335 \cdot \text{Health} \cdot \text{desempleo} + 0.0015 \cdot \text{Health} \cdot \text{PIB} + 0.0002 \cdot \text{desempleo} \cdot \text{PIB} - 0.0003 \cdot \text{Health} \cdot \text{desempleo} \cdot \text{PIB} + f(\text{LATITUD}, \text{LONGITUD})$	11.2 %
2017	$\log(\mu) = -1.2061 - 10.2814 \cdot \text{Health} - 2.4566 \cdot \text{desempleo} - 0.0002 \cdot \text{PIB} + 2.8629 \cdot \text{Health} \cdot \text{desempleo} + 0.0002 \cdot \text{Health} \cdot \text{PIB} + 0.0001 \cdot \text{desempleo} \cdot \text{PIB} - 0.0001 \cdot \text{Health} \cdot \text{desempleo} \cdot \text{PIB} + f(\text{LATITUD}, \text{LONGITUD})$	7.39 %
2018	$\log(\mu) = -11.5832 + 2.1709 \cdot \text{Health} - 2.1862 \cdot \text{desempleo} + 0.0006 \cdot \text{PIB} + 2.4926 \cdot \text{Health} \cdot \text{desempleo} - 0.0007 \cdot \text{Health} \cdot \text{PIB} + 0.0000 \cdot \text{desempleo} \cdot \text{PIB} + 0.0000 \cdot \text{Health} \cdot \text{desempleo} \cdot \text{PIB} + f(\text{LATITUD}, \text{LONGITUD})$	4.27 %
2019	$\log(\mu) = -13.1670 + 4.7372 \cdot \text{Health} - 0.8444 \cdot \text{desempleo} + 0.0005 \cdot \text{PIB} + 0.8892 \cdot \text{Health} \cdot \text{desempleo} - 0.0006 \cdot \text{Health} \cdot \text{PIB} + 0.0000 \cdot \text{desempleo} \cdot \text{PIB} + 0.0000 \cdot \text{Health} \cdot \text{desempleo} \cdot \text{PIB} + f(\text{LATITUD}, \text{LONGITUD})$	3.85 %

Los resultados de las pruebas de normalidad Shapiro - Wilk de los residuos en cada año respaldan la validez de los modelos GAM utilizados. Esto significa que se cumplió uno de los supuestos fundamentales del análisis estadístico, lo que fortalece la confianza en las conclusiones derivadas de estos modelos, los p - valores se detallan en la siguiente tabla.

Tabla 12: p - valor para los modelos GAM por año

Año	p - valor
2014	0.06525
2015	0.2549
2016	0.2165
2017	0.8613
2018	0.1232
2019	0.1306

El rendimiento del modelo presenta mejoras abruptas a lo largo del tiempo, lo cual enfatiza en que las relaciones entre las variables explicativas y la tasa de interés toma más fuerza con el tiempo, lo cual nos permite obtener un óptimo valor del MAPE para el año 2019 con 3.85%, por otra parte, en la siguiente figura se observa el resultado de las estimaciones para los 24 departamentos que tienen la información disponible completa, en este sentido, las predicciones de los últimos años tienden a ser más confiables por la precisión que mejora de forma significativa para estos periodos.

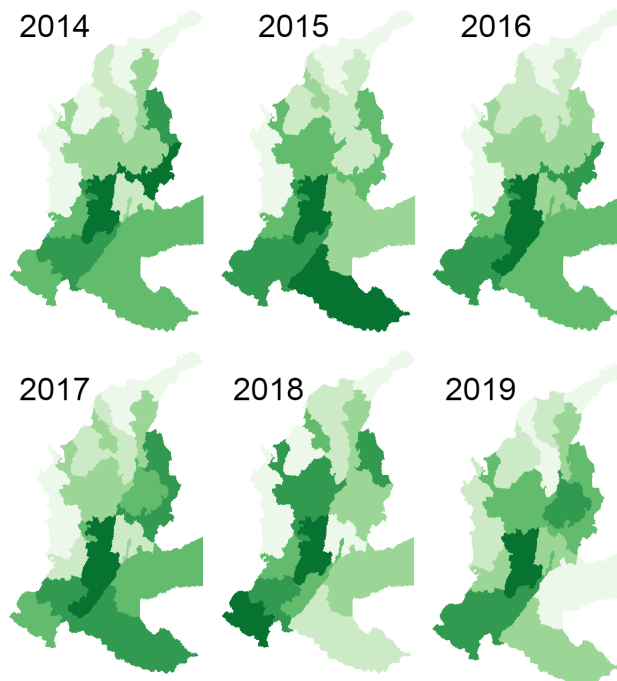


Figura 2: Comportamiento de la tasa de suicidio estimada en los años de 2014 a 2019.

9. Conclusiones

Eficacia y resultados del modelo: El modelo de regresión beta espacio-temporal, implementado mediante Modelos Aditivos Generalizados (GAM), ha capturado adecuadamente las dependencias espaciales y temporales en las tasas de suicidio en Colombia entre 2014 y 2019. Los GAM han demostrado una alta capacidad para ajustar los datos, logrando un MAPE de 3.85 % en 2019.

Importancia de la correlación espacial y variables socioeconómicas: La prueba de Moran ha confirmado la autocorrelación espacial en las tasas de suicidio, indicando que las tasas en un departamento están influenciadas por los vecinos. El desempleo ha sido un predictor significativo, mostrando una relación inversa con las tasas de suicidio. Las interacciones entre el índice de salud, desempleo y PIB han sido clave para una estimación precisa.

Mejora del desempeño del modelo: Los modelos han mejorado en desempeño a lo largo del tiempo, reflejando una disminución en el MAPE y fortaleciendo las relaciones entre las variables explicativas y la tasa de suicidios. La inclusión de términos espaciales ha mejorado notablemente el ajuste y precisión de las estimaciones.

Implicaciones para políticas públicas y desafíos: Los hallazgos proporcionan una base sólida para formular políticas públicas y estrategias de intervención, permitiendo una asignación eficiente de recursos. Sin embargo, el estudio enfrenta limitaciones en la disponibilidad y calidad de los datos socioeconómicos y de salud mental, lo cual debe abordarse en futuros estudios.

Limitaciones: La muestra se restringe a únicamente 24 departamentos, lo cual podría no ser representativo de todo el país y disminuir la capacidad de generalización de los resultados. Asimismo, la carencia de datos sobre algunas variables socioeconómicas puede haber llevado a la exclusión de factores importantes que afectan la tasa de suicidio, afectando la precisión y solidez del modelo. Obtener información sobre las variables socioeconómicas no incluidas en el estudio actual puede implicar colaborar con instituciones gubernamentales y organizaciones no gubernamentales que puedan proporcionar estos datos, o, en otros casos, incorporar información de diferentes variables socioeconómicas a las ya incorporadas.

Recomendaciones para futuros estudios: Se recomienda integrar datos más recientes y explorar otros factores como variables psicológicas y ambientales. Además, el uso de métodos estadísticos avanzados como Modelos de Markov Aleatorios (MRF), Modelos Bayesianos Jerárquicos y Redes Neuronales Artificiales (ANN) podría mejorar la comprensión y predicción de las tasas de suicidio.

Referencias

- S. Banerjee, B. P. Carlin, and A. E. Gelfand. *Hierarchical Modeling and Analysis for Spatial Data*. Chapman and Hall/CRC, 2014.
- N. Cressie and C. K. Wikle. *Statistics for Spatio-Temporal Data*. John Wiley & Sons, 2011.
- G. de México. 10 de septiembre: Día mundial de la prevención del suicidio, 2021. Recuperado de <https://www.gob.mx/salud%7Cseguropopular/articulos/10-de-septiembre-dia-mundial-de-la-prevencion-del-suicidio#:~:text=La%20OMS%20define%20el%20suicidio,expectativa%20de%20su%20desenlace%20fatal>.
- N. U. Deshpande. Time-series-forecasting-of-amazon-stock-prices-using-neural-networks-lstm-and-gan. *GitHub*, 46(1):1, August 2019. Recuperado de <https://github.com/username/repo>.
- S. L. P. Ferrari and F. Cribari-Neto. Beta regression for modelling rates and proportions. *Journal of Applied Statistics*, 31(7):799–815, 2004. doi: 10.1080/0266476042000214501.
- T. Hastie and R. Tibshirani. *Generalized Additive Models*. Chapman and Hall/CRC, New York, 1990.
- Huellas. Postvención: ayuda ante una pérdida por suicidio, 2016. Recuperado de <https://huellas.pucpr.edu/postvencion-ayuda-ante-una-perdida-por-suicidio/>.
- K. Jones and J. Wakefield. Generalized additive models for spatial and temporal analysis of crime data. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 176(1):125–148, 2013. doi: 10.1111/j.1467-985X.2012.01080.x.
- R. Kalhori and M. Mohammadzadeh. Spatial beta model for epidemiological studies. *Journal of Spatial Statistics*, 12(3):45–60, 2016.
- P. McCullagh and J. A. Nelder. *Generalized Linear Models*. Chapman and Hall, 1989.
- R. A. Rigby and D. M. Stasinopoulos. Generalized additive models for location, scale and shape (with discussion). *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 54(3):507–554, 2007.
- D. Ruppert, M. P. Wand, and R. J. Carroll. *Semiparametric Regression*. Cambridge University Press, 2003.
- T. F. Smith, A. Baer, R. Goedken, and P. Scott. Using generalized additive models to analyze and predict the spatial distribution of diseases. *Journal of Spatial and Spatio-temporal Epidemiology*, 15:12–23, 2015. doi: 10.1016/j.sste.2015.05.001.

- M. Smithson and J. Verkuilen. A better lemon squeezer? maximum-likelihood regression with beta-distributed dependent variables. *Psychological Methods*, 11(1):54, 2006.
- R. C. Team. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria, 2020.
- R. W. M. Wedderburn. Quasi-likelihood functions, generalized linear models, and the gauss-newton method. *Biometrika*, 61(3):439–447, 1974.
- S. N. Wood. *Generalized Additive Models: An Introduction with R*. Chapman and Hall/CRC, Boca Raton, FL, 2017.