

COVID-19 in Antioquia: A Geospatial Analysis of Case Fatality Rate

A. Garcia^a

^aFacultad de Minas, Universidad Nacional de Colombia, Medellín, Colombia

ARTICLE INFO

Keywords:
COVID-19
Geospatial analysis
Spatial regression
Antioquia

ABSTRACT


This study conducted a comparative geospatial analysis of COVID-19 case fatality rates (CFR) across municipalities in Antioquia, Colombia, employing six distinct spatial regression models. We evaluated the performance of Ordinary Least Squares (OLS), Spatial Autoregressive Model (SAR), Spatial Lag Model (SLM), Spatial Error Model (SEM), Geographically Weighted Regression (GWR), and Multiscale Geographically Weighted Regression (MGWR), each designed to address spatial dependence or heterogeneity. OLS, which disregards spatial effects, showed modest explanatory power (Adjusted $R^2 = 0.2085$). SAR and SLM incorporated global spatial autocorrelation through lagged dependent or independent variables, respectively. SEM, by modeling spatial autocorrelation in residuals, achieved the best model fit (AIC = -704.025). GWR, utilizing a fixed Gaussian kernel, allowed for local parameter variations and yielded the highest explanatory power ($R^2 = 0.560$), though with reduced degrees of freedom. In contrast, MGWR, designed for variable-specific bandwidths, exhibited a lower R^2 (0.135), suggesting potential underfitting in this specific application. Across all models, altitude and population density consistently emerged as significant predictors of CFR, while humidity and urbanization demonstrated less influence. Spatial models (SLM, SEM, GWR) significantly outperformed OLS, highlighting the critical need to account for spatial dependencies in epidemiological analyses. Among these, SEM offered the optimal balance between parsimony and fit, while GWR underscored the importance of spatial heterogeneity. These findings emphasize the value of spatially explicit modeling in informing public health responses and resource allocation strategies during pandemics.

1. Introduction

The Coronavirus Disease 2019 (COVID-19), caused by the SARS-CoV-2 virus, emerged as a global health crisis following its initial identification in January 2020. On March 11, 2020, the World Health Organization (WHO) officially declared it a pandemic, acknowledging its rapid transmission, high infectivity, and profound global impact on public health, society, and daily life [1, 2, 3]. This unprecedented scenario underscored the urgent need to understand the virus epidemiological and transmission dynamics to inform effective public health interventions and therapeutic strategies.

Understanding the spatiotemporal dynamics of infectious diseases such as COVID-19 is essential for effective mitigation and preparedness. It enables the assessment of outbreak magnitude, guides evidence-based decision-making, and supports targeted public health responses [1]. In this context, Geographic Information Systems (GIS) and spatial analysis have become indispensable tools, allowing researchers to examine disease patterns, identify spatial associations with environmental and socioeconomic variables, and explore transmission mechanisms [1]. Geography, through its integrative and holistic perspective, is particularly well-suited to analyze the complex interactions between biophysical and human factors during a pandemic [1].

A growing body of research has explored the geographic dimensions of COVID-19, including disease mapping, spatiotemporal modeling, health and social geography, environmental influences, data mining, and interactive web-based visualization [4]. These studies frequently draw upon large-scale geospatial datasets—commonly referred to as ‘big data’—sourced from official COVID-19 dashboards (e.g., Johns Hopkins University, national health ministries), anonymized mobility datasets from technology platforms (e.g., Google, Baidu, Facebook), mobile phone location data, flight itineraries, and social media streams [1, 4]. In addition, environmental (e.g., temperature, humidity) and demographic (e.g., urbanization, population density) variables are often incorporated to enrich the analytical framework [5, 6, 1, 7, 8, 9, 10, 11, 12].

 algarciach@unal.edu.co (A. Garcia)
ORCID(s):

To examine the relationships between COVID-19 incidence and potential explanatory factors, researchers have employed a range of spatial regression techniques, including Ordinary Least Squares (OLS), the Spatial Lag Model (SLM), and the Spatial Error Model (SEM) [6, 9, 13]. More sophisticated approaches such as Geographically Weighted Regression (GWR) and its extension, Multiscale Geographically Weighted Regression (MGWR), have been introduced to account for spatial non-stationarity and heterogeneous relationships across different geographic scales [8, 5, 6, 13]. These models have been instrumental in identifying key predictor variables such as population mobility, socioeconomic disparities (e.g., income, education, access to healthcare), and the role of policy interventions [7].

Among the epidemiological metrics used to assess the severity of the pandemic, the Case Fatality Rate (CFR)—calculated as the ratio of reported deaths to confirmed cases—remains a critical indicator for public health planning and communication with at-risk populations [14, 15]. However, accurately estimating CFR is methodologically challenging due to underreporting, delays in case confirmation, and the presence of asymptomatic infections [16].

Addressing the COVID-19 crisis effectively requires an interdisciplinary perspective that integrates data-driven insights, global cooperation, and anticipatory planning [1]. Data visualization and spatiotemporal modeling are vital in uncovering hidden patterns, communicating complex information, and facilitating public engagement [2]. Analyzing case data as spatiotemporal phenomena with multi-scale granularity enhances our capacity to implement effective containment and mitigation strategies [17, 18, 4].

This study evaluates spatial patterns and dependencies in Case Fatality Rate (CFR) in confirmed COVID-19 cases across the municipalities of Antioquia, Colombia, using spatial regression models including OLS, SAR, SEM, GWR and MGWR. By doing so, it seeks to contribute to the understanding of spatial health inequalities and to inform targeted public health interventions at the local level.

2. Data

2.1. Study Area

The study was conducted in the Department of Antioquia, Colombia, located between 5°30' N and 8°50' N latitude and 73°50' W and 77°00' W longitude. Antioquia encompasses an estimated total area of 62,867 km² and is composed of 125 municipalities, organized into nine subregions. The department has an estimated total population of approximately 6.903 million for 2024 [19, 20]. Figure 1 illustrates the geographical extent and municipal divisions of the study area, along with the spatial distribution of COVID-19 Case Fatality Rate (CFR).

2.2. Data Collection

Demographic and confirmed COVID-19 case data for Colombia were obtained from the Instituto Nacional de Salud (INS) [21] via the national open data portal [22]. Population data for the municipalities of Antioquia were sourced from the Departamento Administrativo Nacional de Estadística (DANE) [19].

Geographic boundaries and municipal georeferencing information were retrieved from the geoportals of DANE [19] and the Government of Antioquia [23].

Environmental data were also collected. Average municipal altitude was extracted from the NASA Shuttle Radar Topography Mission (SRTM) Digital Elevation Model (30m resolution) [24], accessed through Google Earth Engine. Annual mean temperature data were obtained from the WorldClim BIO Variables V1 dataset [25]. Annual mean relative humidity was estimated using the ERA5-Land Monthly reanalysis dataset from the European Centre for Medium-Range Weather Forecasts (ECMWF) [26]. Both climate datasets were accessed via Google Earth Engine.

The Case Fatality Rate (CFR) [15] was used as the objective variable for this study, serving as a critical indicator to understand the severity and lethality of the virus across different municipalities. The CFR was calculated as the number of known deaths divided by the total number of confirmed cases, expressed as a percentage, following the formula:

$$CFR = \frac{\text{Number of known Deaths}}{\text{Total number of Confirmed cases}} \times 100$$

The initial distribution of the municipal-level (CFR) exhibited strong positive skewness and contained outliers, as shown in Figure 2 (left), rendering it unsuitable for regression models that assume normally distributed residuals. To address this, the response variable was transformed using a logarithmic function $\log(Y + 1)$. This transformation is commonly employed to normalize skewed distributions, reduce the influence of extreme values, and accommodate zero

observations [13]. As a result, the transformed data approximates a continuous distribution with stabilized variance and residuals that more closely follow a Gaussian distribution, as illustrated in Figure 2, (right).

The predictive variables considered for this study include: municipal area, total population, population density, annual mean relative humidity, annual mean temperature, and mean altitude. The response variable, a continuous measure, is the Case Fatality Rate (CFR) [15], serving as an indicator of COVID-19-attributable mortality. The CFR was calculated as the number of known deaths divided by the total number of confirmed cases, expressed as a percentage, following the formula:

The predictive variables in this study included MEAN [1], mean temperature [1, 7, 6] and annual mean relative humidity [1, 8] as environmental geographic variables, along with urbanization [9, 7] and population density [1, 9, 10, 8, 7, 6, 11, 27, 28, 12] as socioeconomic and demographic indicators. These variables are widely recognized as predictors or influential factors in geospatial analyses of the COVID-19 pandemic

2.3. Exploratory Spatial Data Analysis

The spatial distribution of the Case Fatality Rate (CFR) reveals substantial heterogeneity across the region. To visualize this variation, Figure 3 presents a choropleth map of the CFR values.

2.4. Spatial Regression Models

2.4.1. Ordinary Least Squares (OLS):

$$y = X\beta + \varepsilon, \quad (1)$$

where y is an $n \times 1$ vector of observations, X is an $n \times k$ matrix of predictors, β is a $k \times 1$ vector of coefficients, and $\varepsilon \sim \mathcal{N}(0, \sigma^2 I)$ are i.i.d. residuals.

2.4.2. Simultaneous Autoregressive Model (SAR)

The SAR model introduces spatial lag of the dependent variable:

$$y = \rho Wy + X\beta + \varepsilon$$

where ρ is the spatial autoregressive parameter and W the spatial weights matrix [13].

2.4.3. Spatial Lag Model (SLM)

The SLM similarly incorporates spatial dependence:

$$y = X\beta + WXy + \varepsilon$$

with λ representing spatial spillover effects [13].

2.4.4. Spatial Error Model (SEM)

The SEM assumes spatial correlation in the error term:

$$y = X\beta + \lambda W\mu + \varepsilon$$

where λ controls residual spatial dependence [13].

2.4.5. Geographically Weighted Regression (GWR)

GWR captures spatial heterogeneity by allowing parameters to vary locally:

$$y_i = \beta_0(u_i, v_i) + \sum_{k=1}^p \beta_k(u_i, v_i) x_{ik} + \varepsilon_i$$

where (u_i, v_i) are coordinates and β_k are location-specific coefficients [13].

Table 1
Comparative Summary of Regression Models

Metric	OLS	SAR	SLM	SEM	GWR	MGWR
Pseudo-R ² /Adjusted R ²	0.2085	0.2458	0.2504	0.2458	0.560	0.135
AIC	-695.553	-704.025	-696.962	-699.746	11.443	482.432

2.4.6. Multiscale Geographically Weighted Regression (MGWR)

MGWR extends GWR by allowing each predictor to operate at a different spatial scale:

$$y_i = \beta_0(u_i, v_i) + \sum_{k=1}^p \beta_k(u_i, v_i; b_k) x_{ik} + \varepsilon_i$$

where b_k is a variable-specific bandwidth [13].

These models were selected to evaluate spatial heterogeneity and autocorrelation effects on CFR variable.

3. Results

3.1. Global Spatial correlation

The exploratory analysis reveals a non-random distribution of the normalized slope variable, with a positive global Moran's I value of $I = 0.14$, indicating weak but significant spatial autocorrelation ($p < 0.05$). Figure 4

3.2. Local Spatial Cluster (LISA)

The Local Indicators of Spatial Association (LISA) analysis reveals significant local clusters in the study area. As illustrated in Figure 5 high-high (HH) clusters, representing municipalities with high slope values surrounded by others with similarly high values, are primarily located in the northern region. Conversely, low-low (LL) clusters are concentrated in the central region.

3.3. Model Comparison

We evaluated six spatial regression models to analyse the determinants of the COVID-19 case fatality rate (CFR): Ordinary Least Squares (OLS), Spatial Autoregressive Model (SAR), Spatial Lag Model (SLM), Spatial Error Model (SEM), Geographically Weighted Regression (GWR), and Multiscale GWR (MGWR).

As summarized in Table 1, all spatial models outperform the OLS in terms of adjusted R² or pseudo-R². Among the global models, the Spatial Error Model (SEM) yields the best overall performance, achieving a relatively high adjusted R² (0.2458) and the lowest AIC (-699.746). This suggests that spatial autocorrelation in the error terms is a key component in explaining the distribution of CFR.

The SLM and SAR models show comparable R² values (0.2504 and 0.2458 respectively), but higher AIC values than SEM. While GWR achieves the highest pseudo-R² (0.560), its AIC (11.443) is substantially worse, indicating potential overfitting or model instability. MGWR, in contrast, shows the lowest performance in both pseudo-R² (0.135) and AIC (482.432), suggesting it is not suitable for the dataset.

Figure 6 illustrates the spatial distribution of residuals from the Spatial Error Model (SEM). The map shows no clear spatial clustering of residuals, indicating that the SEM effectively accounts for spatial autocorrelation in the data. This absence of systematic patterns in the residuals supports the model's adequacy and confirms the correction of spatial dependence, consistent with its theoretical structure. This visual evidence reinforces the statistical diagnostics, which suggest that the SEM provides the best model fit among those evaluated.

Discussion

This study analyzed the spatial variation of COVID-19 Case Fatality Rates (CFR) across Antioquia using six regression models: OLS, SAR, SLM, SEM, GWR, and MGWR. Among them, the Spatial Error Model (SEM) showed the best performance, with the lowest AIC (-704.025) and a good pseudo R² (0.2458), indicating that spatial correlation in the residuals is important and must be considered.

Although the GWR model achieved a higher pseudo R^2 (0.560), its high AIC (11.443) suggests overfitting, likely due to its complexity and limited sample size. MGWR performed worse, possibly because the multiscale structure was not strong in this dataset.

Altitude and population density were the most consistent and significant predictors of CFR. In contrast, variables like humidity and urbanization were not statistically significant. This may be due to the lack of health system variables, such as hospital capacity or ICU availability.

High CFRs in small municipalities may reflect few deaths over low case counts (e.g., 3 deaths out of 30 cases = 10%). Such instability affects model robustness. Future models should include healthcare access, comorbidities, and age structure to improve explanatory power.

References

- [1] I. Franch-Pardo, B. M. Napoletano, F. Rosete-Verges, L. Billa, Spatial analysis and gis in the study of covid-19. a review, *Science of the Total Environment* 739 (2020).
- [2] H. Turbe, V. G. Ruiz, M. Bjelogrić, J. Rochat, C. Lovis, Communicating on multivariate and geospatial data supported by ergonomics criteria: Covid-19 case, in: *Proceedings - 2020 Workshop on Visual Analytics in Healthcare, VAHC 2020*, Institute of Electrical and Electronics Engineers Inc., 2020, pp. 4–16. doi:10.1109/VAHC53729.2020.00008.
- [3] R. E. Cuomo, V. Purushothaman, J. Li, M. Cai, T. K. Mackey, A longitudinal and geospatial analysis of covid-19 tweets during the early outbreak period in the united states, *BMC Public Health* 21 (2021).
- [4] B. Lin, L. Zou, M. Yang, B. Zhou, D. Mandal, J. Abedin, H. Cai, N. Ning, Progress in understanding human-covid-19 dynamics using geospatial big data: a systematic review, 2024. doi:10.1080/19475683.2024.2418584.
- [5] Y. Chen, N. Jiang, Y. Cao, Z. Yang, X. Zhao, Visual method of analyzing covid-19 case information using spatio-temporal objects with multi-granularity, *Journal of Geographical Sciences* 31 (2021) 1059–1081.
- [6] H. Yu, J. Li, S. Bardin, H. Gu, C. Fan, Spatiotemporal dynamic of covid-19 diffusion in china: A dynamic spatial autoregressive model analysis, *ISPRS International Journal of Geo-Information* 10 (2021).
- [7] R. Ramírez-Aldana, J. C. Gomez-Verjan, O. Y. Bello-Chavolla, C. García-Peña, Spatial epidemiological study of the distribution, clustering, and risk factors associated with early cov-19 mortality in mexico, *PLoS ONE* 16 (2021).
- [8] Y. Han, L. Yang, K. Jia, J. Li, S. Feng, W. Chen, W. Zhao, P. Pereira, Spatial distribution characteristics of the covid-19 pandemic in beijing and its relationship with environmental factors, *Science of the Total Environment* 761 (2021).
- [9] I. Dutta, T. Basu, A. Das, Spatial analysis of covid-19 incidence and its determinants using spatial modeling: A study on india, *Environmental Challenges* 4 (2021).
- [10] H. Bherwani, S. Anjum, S. Kumar, S. Gautam, A. Gupta, H. Kumbhare, A. Anshul, R. Kumar, Understanding covid-19 transmission through bayesian probabilistic modeling and gis-based voronoi approach: a policy perspective, *Environment, Development and Sustainability* 23 (2021) 5846–5864.
- [11] F. Parvin, S. A. Ali, S. N. I. Hashmi, A. Ahmad, Spatial prediction and mapping of the covid-19 hotspot in india using geostatistical technique, *Spatial Information Research* 29 (2021) 479–494.
- [12] P. H. Wheeler, C. A. Patten, C. I. Wi, J. T. Bublit, E. Ryu, E. H. Ristagno, Y. J. Juhn, Role of geographic risk factors and social determinants of health in covid-19 epidemiology: Longitudinal geospatial analysis in a midwest rural region, *Journal of Clinical and Translational Science* 6 (2022).
- [13] E. Aristizábal, Libro guía del curso análisis geoespacial, Universidad Nacional de Colombia, Departamento de Geociencias y Medio Ambiente, Facultad de Minas, 2024. URL: https://edieraristizabal.github.io/Libro_AnalisisGeoespacial/intro.html, accessed: 2025-06-02.
- [14] P. Das, M. Igoe, S. Lenhart, L. Luong, C. Lanzas, A. L. Lloyd, A. Odoi, Geographic disparities and determinants of covid-19 incidence risk in the greater st. louis area, missouri (united states), *PLoS ONE* 17 (2022).
- [15] P. E. Chaparro-Narváez, Boletín técnico interactivo mortalidad por covid-19 en colombia (2021) 2021.
- [16] A. Paternina-Caicedo, N. Alvis-Guzmán, C. Duenas, J. Narvaez, A. D. Smith, F. D. L. Hoz-Restrepo, Impact of mobility restrictions on the dynamics of transmission of covid-19 in colombian cities, *International Health* 14 (2022) 332–335.
- [17] A. Y. Forkuo, T. V. Nihi, O. O. Ojo, C. N. Nwokedi, O. S. Soyeye, A conceptual model for geospatial analytics in disease surveillance and epidemiological forecasting, *International Medical Science Research Journal* 5 (2025) 30–57.
- [18] S. Roy, M. Saha, B. Dhar, S. Pandit, R. Nasrin, Geospatial analysis of covid-19 lockdown effects on air quality in the south and southeast asian region, *Science of the Total Environment* 756 (2021).
- [19] Departamento Administrativo Nacional de Estadística (DANE), Geoportal dane, 2025. URL: <https://geoportal.dane.gov.co>, accessed: 2025-05-01.
- [20] Secretaria Seccional de Salud y Protección Social de Antioquia, 2024. URL: <https://dssa.gov.co>, accessed: 2025-05-01.
- [21] Instituto Nacional de Salud (INS), Covid-19 en colombia, 2025. URL: <https://www.ins.gov.co/>, accessed: 2025-05-01.
- [22] Ministerio de Tecnologías de la Información y las Comunicaciones de Colombia, Datos abiertos colombia, 2024. URL: <https://www.datos.gov.co>, accessed: 2025-05-01.
- [23] Gobernación de Antioquia, Geoportal de la Gobernación de Antioquia, 2024. URL: <https://geodatos-gobantioquia.opendata.arcgis.com/>, accessed: 2025-05-01.
- [24] T. Farr, P. Rosen, E. Caro, R. Crippen, R. Duren, S. Hensley, M. Kobrick, M. Paller, E. Rodriguez, L. Roth, D. Seal, S. Shaffer, J. Shimada, J. Umland, M. Werner, M. Oskin, D. Burbank, D. Alsdorf, The shuttle radar topography mission, *Reviews of Geophysics* 45 (2007) RG2004.

- [25] R. Hijmans, S. Cameron, J. Parra, P. Jones, A. Jarvis, Very high resolution interpolated climate surfaces for global land areas, *International Journal of Climatology* 25 (2005) 1965–1978.
- [26] J. Muñoz Sabater, Era5-land monthly averaged data from 1981 to present, Copernicus Climate Change Service (C3S) Climate Data Store (CDS), 2019. doi:10.24381/cds.68d2bb30, accessed: 2025-05-01.
- [27] H. Yi, S. T. Ng, A. Farwin, A. P. T. Low, C. M. Chang, J. Lim, Health equity considerations in covid-19: geospatial network analysis of the covid-19 outbreak in the migrant population in singapore, *Journal of Travel Medicine* 28 (2021).
- [28] D. I. P. Putri, D. Agustian, L. Apriani, R. Ilyas, Spatial and temporal analysis of covid-19 cases in west java, indonesia and its influencing factors, *International Journal of Environmental Research and Public Health* 20 (2023).

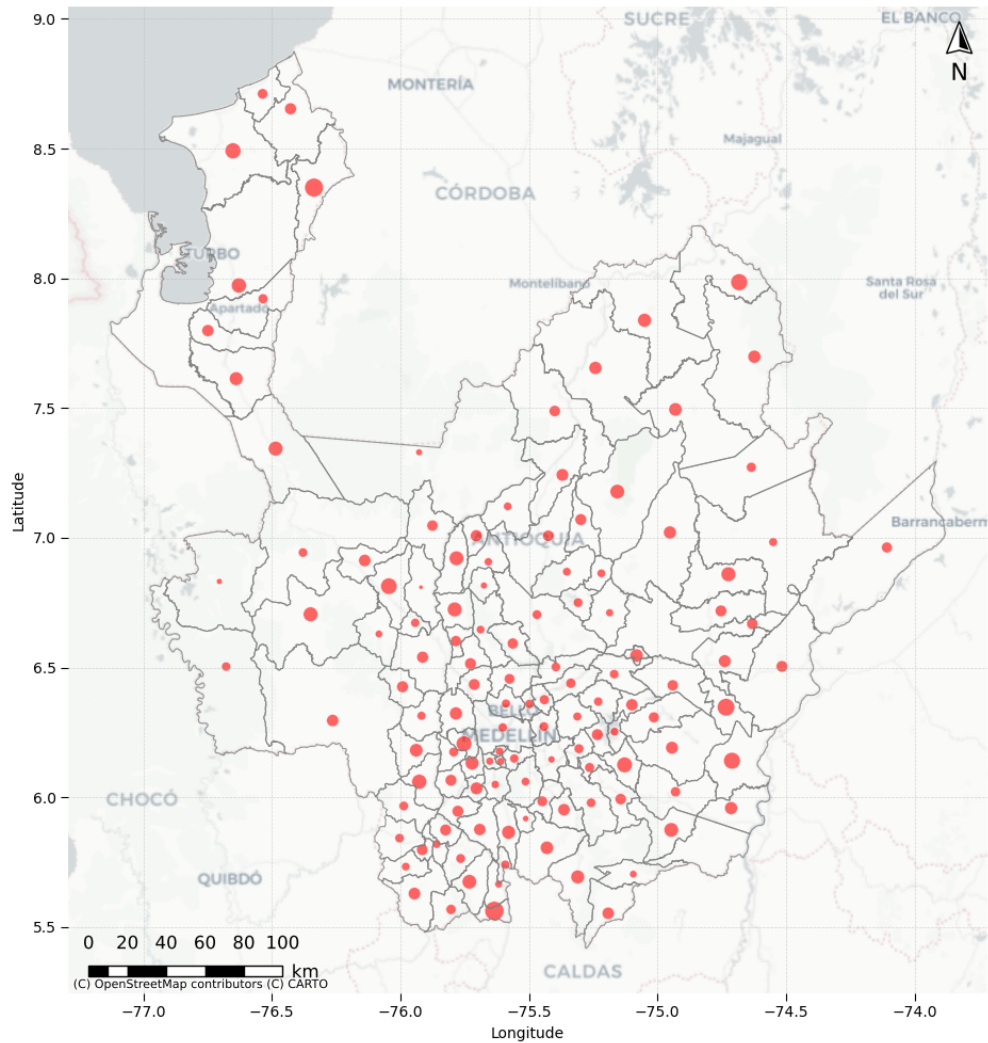


Figure 1: Map of the study area, the Department of Antioquia, Colombia, showing municipal divisions and spatial distribution of COVID-19 Case Fatality Rate (CFR).

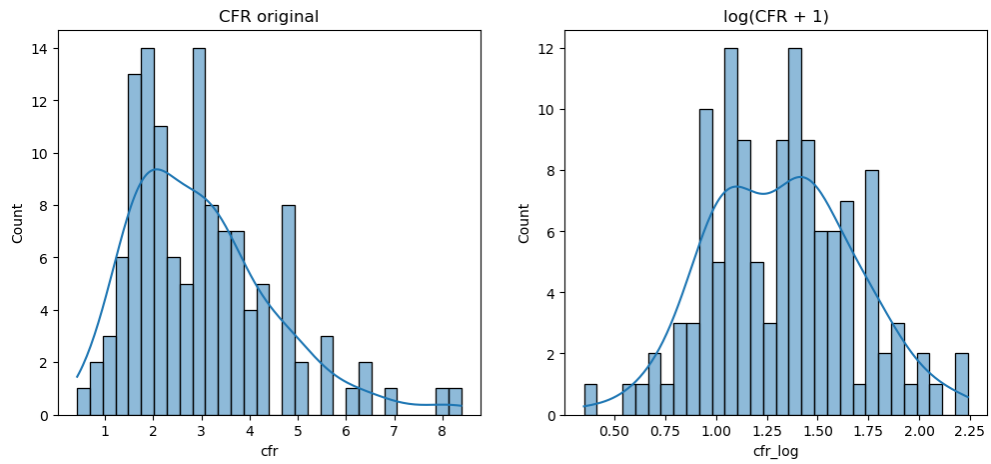


Figure 2: Distribution of Municipal-Level COVID-19 Case Fatality Rate (CFR) Before and After Logarithmic Transformation.

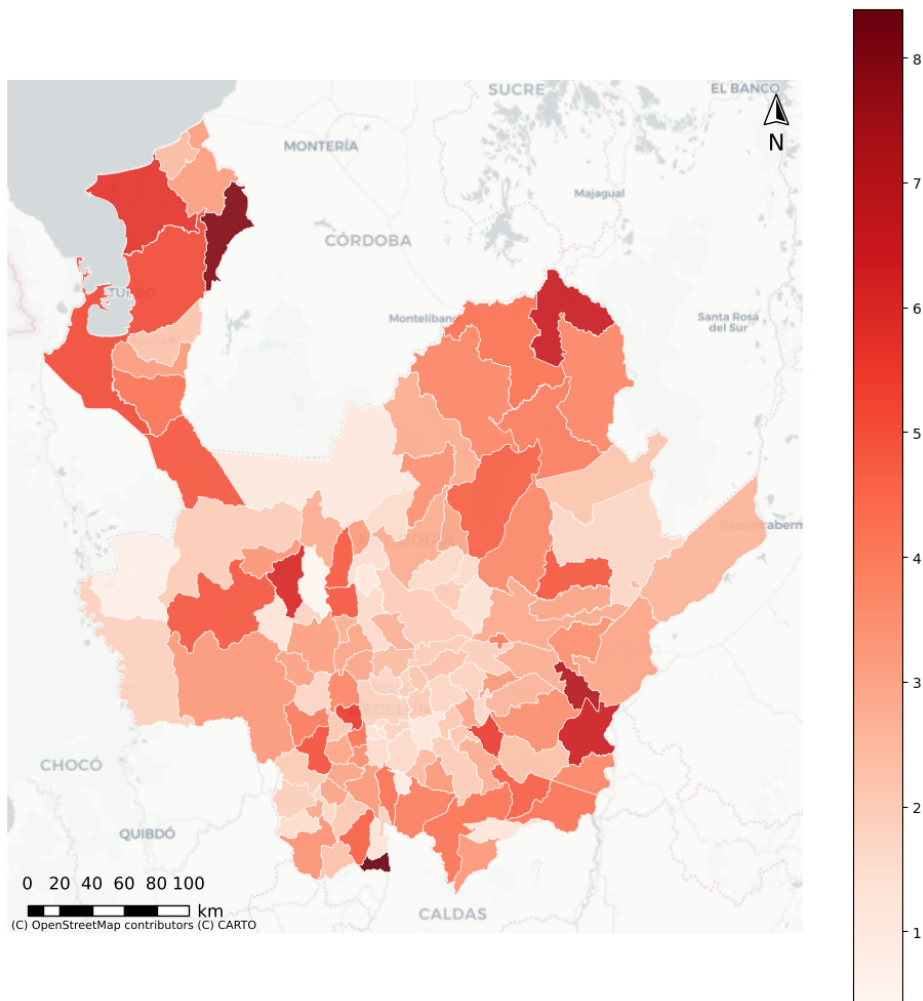


Figure 3: Spatial distribution of Case Fatality Rate (CFR) COVID-19 cases.

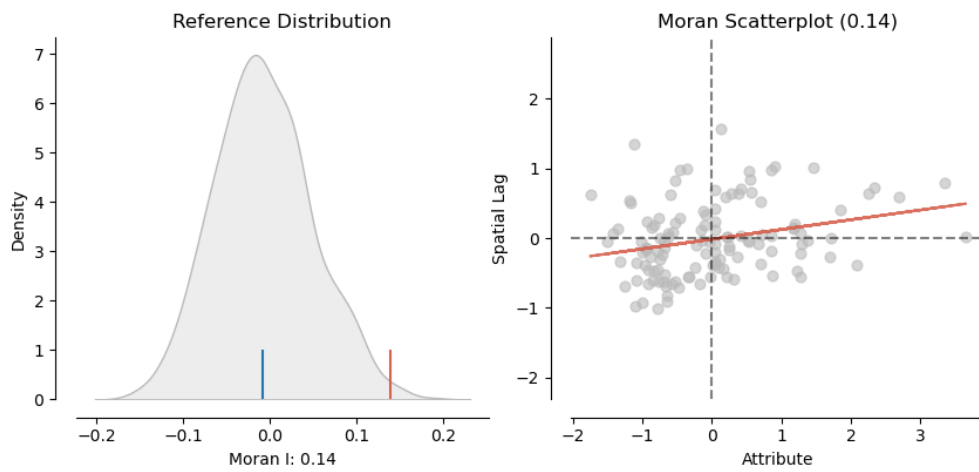


Figure 4: Reference distribution and Moran's I for the key independent variable

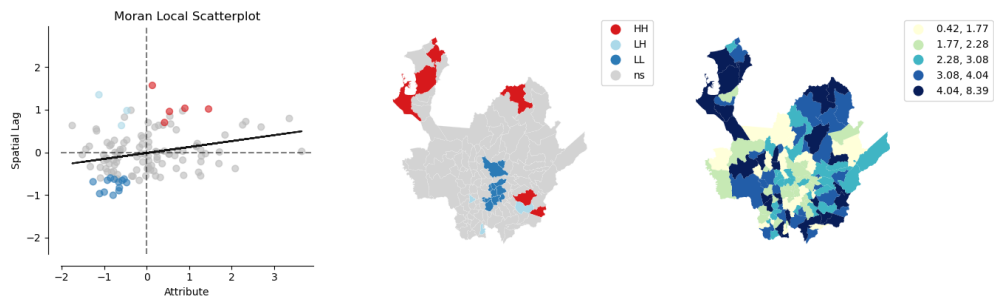


Figure 5: LISA Cluster Map of Normalized Slope

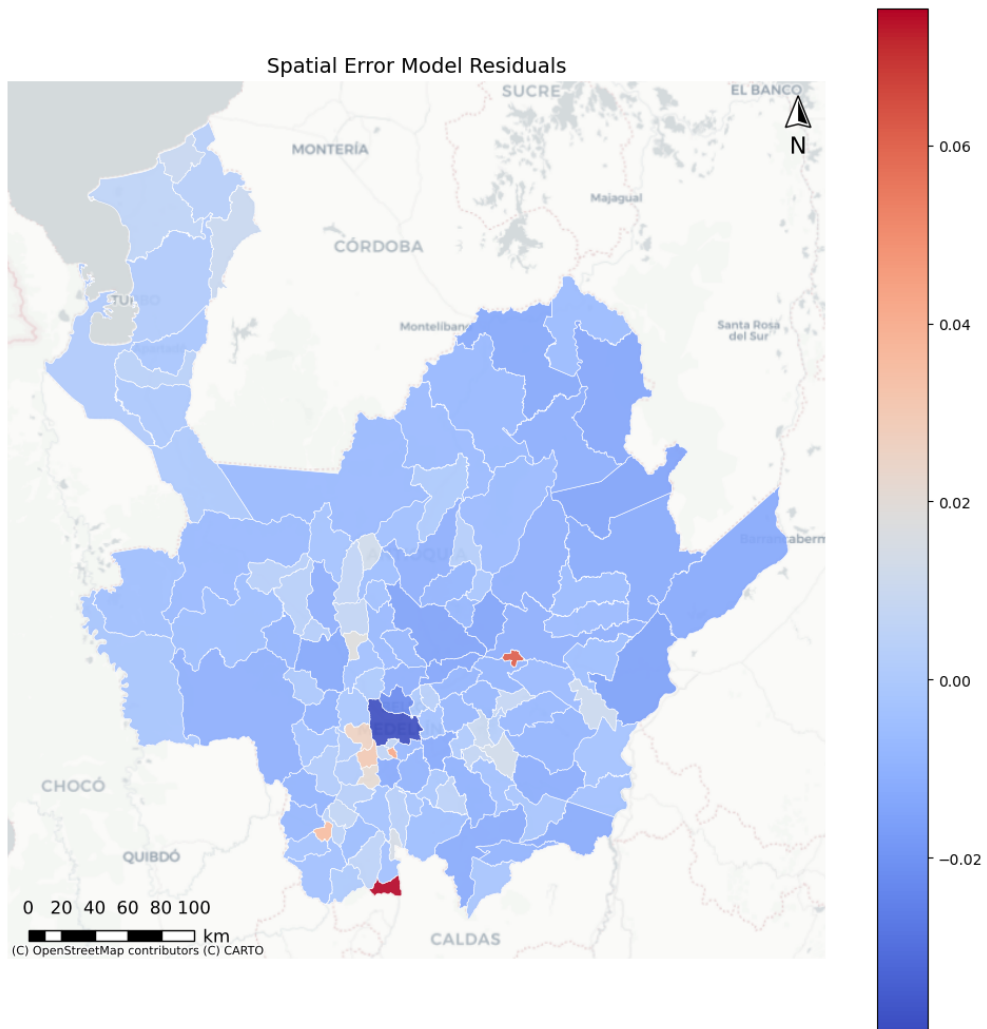


Figure 6: Spatial distribution of residuals from the Spatial Error Model (SEM). Residual clustering suggests remaining local variation not captured by the global model