



LOS LIBERTADORES
FUNDACIÓN UNIVERSITARIA

Autocorrelación espacial y modelamiento predictivo de casos de COVID-19 en Bogotá, Colombia

Jorge Homero Wilches-Visbal, Hernán Darío Díaz-Albornoz

Introducción

Desde diciembre de 2019, la enfermedad por coronavirus 2019 (COVID-19) ha representado un desafío para los gobiernos del mundo una vez que exigía un balance entre estrategias e infraestructuras de control epidemiológico, garantía de bienestar económico así como para la administración de servicios de salud rutinarios.¹ La COVID-19 se transmite por contacto directo e indirecto (fómites) con gotas contaminadas o por aerosoles, siendo esta última vía la principal y más difícil de contener.^{2,3}

Factores ambientales como el material particulado (PM en inglés *particulate matter*), con efectos adversos para la salud humana,^{4,5} favorecen el contagio por aerosoles y el aumento de mortalidad por COVID-19.⁶⁻⁸ El material particulado, clasificado en PM10 y PM2.5 de acuerdo con su diámetro aerodinámico, sirve de protector y transporte; induce daños en los neumocitos y sobreexpresión de la enzima convertidora de angiotensina 2 que facilita el ingreso del coronavirus.⁶ Factores sociodemográficos como la densidad poblacional, el número de centros de salud, la edad, la preexistencia de comorbilidades, entre otros, también inciden en la morbimortalidad por COVID-19.⁹

Modelos epidemiológicos de autocorrelación espacial han sido utilizados para comprender mejor las rutas de propagación de las enfermedades, a través de la búsqueda de aglomerados entre factores ambientales y sociodemográficos, que ayuden a implementar medidas efectivas de mitigación del riesgo de contagio, sobre todo, en las zonas más comprometidas.^{8,10} Aunque se han reportado previamente estudios espaciales sobre COVID-19 en Colombia¹¹⁻¹³, ninguno ha empleado técnicas predictivas de aprendizaje de máquinas (ML en inglés *machine learning*) incluyendo contaminantes ambientales.

Por tanto, el propósito del presente manuscrito fue: i) establecer la existencia de autocorrelación espacial entre el número de casos por COVID-19 y la concentración de PM10 y PM2.5 y ii) determinar el mejor modelo predictivo de ML para los casos de acuerdo con variables socioambientales, en Bogotá (Colombia).

Materiales y Métodos

Se realizó un estudio descriptivo-predictivo de corte transversal, en una muestra de 111 unidades de planeamiento zonal (UPZ) del perímetro urbano,¹⁴ sobre el número de casos en función de contaminantes ambientales y variables sociodemográficas, para el periodo de observación: marzo de 2020 – febrero de 2021.

Las variables sociodemográficas (casos, densidad poblacional y centros de salud) se extrajeron o construyeron a partir de datos de las Secretarías Distritales de Salud y Planeación.^{15,16} Las variables ambientales (PM10 y PM2.5), durante el mismo periodo de tiempo, fueron tomadas de Índice Bogotano de Calidad del Aire y Riesgo en Salud (IBOCA).¹⁷

Para establecer la existencia de autocorrelación espacial entre los casos y los contaminantes, se llevó a cabo el siguiente procedimiento: a) se utilizó la técnica de interpolación espacial (IDW en inglés *inverse distance weight*) para generar capas de los valores medios de los contaminantes por UPZ; b) con estos valores y los casos, se computaron el índice global de autocorrelación espacial o de Moran (IM) bivariado y el índice de asociación espacial local (LISA en inglés *local spatial association index*), empleando la configuración Queen de orden 1.¹⁸

Para definir el modelo predictivo, se realizó lo siguiente: a) se eliminaron los valores atípicos o extremos de las variables; b) se seleccionaron las variables explicativas de mayor incidencia en la variable respuesta (casos), según los resultados de la matriz de correlación; c) se probaron tres modelos ML: árbol de decisión; bosques aleatorios y regresión lineal del paquete *sklearn*, dividiendo los datos en 80 % para entrenamiento y 20 % para prueba; d) se escogió el modelo con el mayor coeficiente de determinación (r^2) y menor error cuadrático medio (ECM) y e) se interpretaron los resultados usando el método *shap* (en inglés *shapley additive explanation*).¹⁹

Para la interpolación espacial y el cálculo del IM y el LISA se utilizaron Qgis 3.14.16 y Geoda 1.18.0.0, respectivamente. Para la predicción, fue empleado Python de Google Colab (MS Windows 10 Intel Core i7-8550U CPU, 2.00 GHz, 12 Gb).

Resultados y Discusión

La media de los casos fue de 4395,6 (DE: 3047,8); la de la densidad poblacional fue 18078,8 (DE: 10216,1) hab./km²; la de centros de salud fue 3,2 (DE: 3,1). Para los contaminantes PM10 y PM2.5, la media fue de 31,6 (DE: 7,5) y 16 (DE: 2,9) µg/m³, respectivamente.

Los coeficientes de correlación entre los casos; la densidad poblacional; centros de salud; PM10 y PM2.5 fueron, correspondientemente, 0,75; 0,41; 0,31 y 0,21, deduciéndose que las variables sociodemográficas deberían impactar más en la modelación.

En la Figura 1 se visualiza la autocorrelación espacial entre los casos y el material particulado, analizada a partir del IM y LISA.

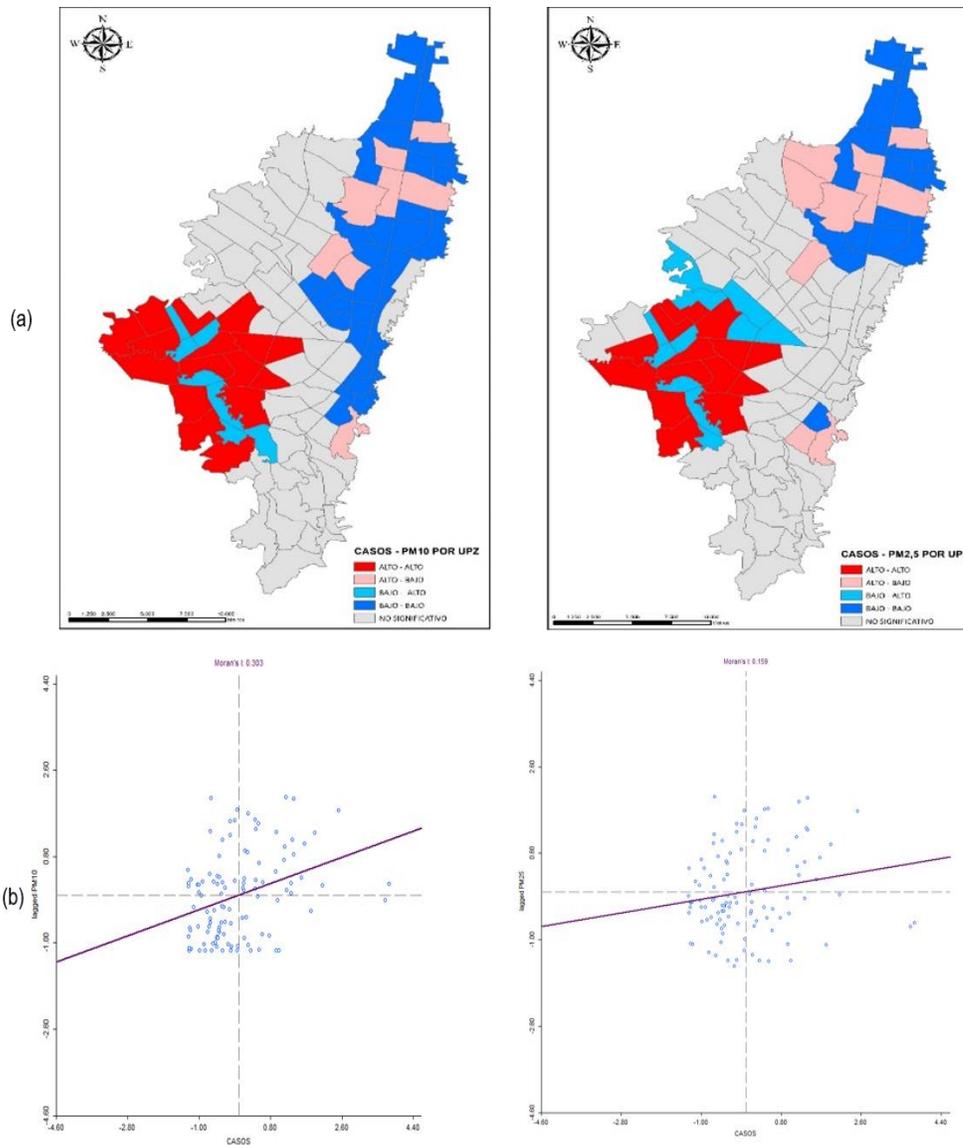


Figura 1. Gráficos de autocorrelación espacial local, de LISA (a), y global, de Moran, (b) entre los casos, PM10 y PM2.5. Fuente: elaboración propia.

El IM para el PM10 fue de 0,303 (p-valor = 0,001) mientras que para el PM2.5 fue de 0,159 (p-valor = 0,002). Esto indica la existencia de correlación espacial positiva de los casos con los PM, toda vez que $IM > 0$,²⁰⁻²² siendo superior para el PM10. Según el LISA, las UPZ del suroccidente de la ciudad, reportaron alto número de casos y altas concentraciones de PM10 y PM2.5 (agrupaciones alto-alto), mostradas en color escarlata. Por otro lado, las agrupaciones bajo-bajo (color azul) de PM2.5 y casos se localizaron en UPZ del norte de la ciudad; para PM10, se ubicaron en el norte y el oriente de esta.^{20,21} Esto puede obedecer a que la zona nororiental de la ciudad tiene menor densidad poblacional, está más arborizada por su cercanía con los cerros orientales y posee un menor nivel de contaminantes.

En la Tabla I, se consignan las métricas de los distintos modelos ML empleados para predecir los casos, en función de la densidad poblacional, los centros de salud y el PM10.

Tabla I. Métricas de modelos de predicción de casos por COVID-19 como función de la densidad poblacional y número de centros de salud y concentración de PM10 en Bogotá, de marzo de 2020 a febrero de 2021.

Modelo Predictivo de ML	Coefficiente de Determinación (r^2)	Error cuadrático medio (ECM)
Regresión Lineal	0,781	1379
Bosques Aleatorios	0,143	2735
Árbol de Decisión	0,047	2881

Fuente: elaboración propia.

De la tabla, se constata que la regresión lineal múltiple arrojó la mejor predicción, con un ECM equivalente al 31% del promedio de casos y un coeficiente de determinación que explica, aproximadamente, el 80% de la varianza de los casos. Aunque el PM2.5 se correlacionó positivamente con los casos, estuvo por debajo del PM10, además el r^2 de modelo predictivo, considerando su inclusión (0,754), disminuyó respecto a cuando se obvió (0,781). Por tal motivo, se decidió excluirlo de la modelación.

En la Figura 2 se muestra la contribución, en orden descendente, de cada una de las variables según el método *shap*.

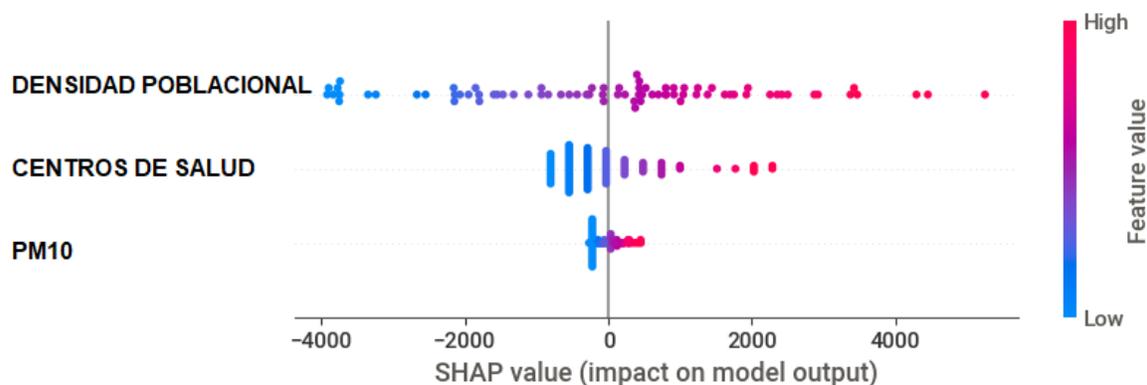


Figura 2. Interpretación del modelo predictivo de los casos mediante la metodología *shap-value*. Las variables explicativas están ordenadas descendientemente, según su impacto. Fuente: elaboración propia.

La densidad poblacional y los centros de salud contribuyeron más en el modelo y por tanto impactaron más en los casos. Como, para ambas variables, los valores altos están en el semieje positivo del *shap-value*, la relación entre estas y los casos es directamente proporcional.¹⁹ Igual sucedió para el PM10, pero en menor medida.

En zonas altamente pobladas suele ser difícil mantener el distanciamiento físico; en aquellas con mayor presencia de servicios de salud se esperan más registros y en lugares poluidos es más proclive la incidencia de enfermedades respiratorias, entre ellas, la COVID-19.

Estos hallazgos coinciden con los de trabajos hechos en Cuba,²³ Chile,²⁴ Estados Unidos²⁵ y Colombia^{11,12} donde la densidad poblacional incidió marcadamente en la cantidad de casos. Asimismo, concuerda con lo hallado en investigaciones llevadas a cabo en Italia,⁶ Austria⁷ y Francia²⁶, en las que altas concentraciones de PM10 se relacionaron con mayor cantidad de casos por COVID-19.

Las limitaciones del estudio son no haber considerado variables sociodemográficas como edad, sexo, ingresos, demás contaminantes y condiciones meteorológicas. Adicionalmente, como los datos no se recolectaron a la misma escala espacial ni estaban disponibles de tal forma, se recurrió a técnicas de interpolación y suavizado de datos.²⁷ Como perspectiva, se proyecta agregar más variables, realizar el modelamiento para decesos o utilizar *stat-models* para obtener una regresión lineal más robusta.

Conclusión

Mediante técnicas de autocorrelación espacial se verificó la asociación positiva entre casos de COVID-19 y concentración de PM10, concentrados, en unidades de planeamiento zonal del suroccidente de Bogotá.

Utilizando un modelo de regresión lineal múltiple y el método *shap* se encontró que, en orden, la densidad poblacional, el número de centros de salud y la concentración de PM10 impactaron positivamente en los casos.

Agradecimientos. Al profesor John González Veloza por sus aportes en *machine learning*, al programa y docentes de especialización en estadística aplicada de la Fundación Universitaria Los Libertadores por el conocimiento brindado.

Declaración de conflicto de intereses. Los autores declararon no tener conflicto de intereses.

Referencias

1. Prinja S, Pandav C. Economics of COVID-19: challenges and the way forward for health policy during and after the pandemic. *Indian J Public Health* [Internet]. 2020;64(6):231. Disponible en: <http://www.ijph.in/text.asp?2020/64/6/231/285631>
2. Greenhalgh T, Jimenez JL, Prather KA, Tufekci Z, Fisman D, Schooley R. Ten scientific reasons in support of airborne transmission of SARS-CoV-2. *Lancet* [Internet]. mayo de 2021;397(10285):1603-5. Disponible en: <https://linkinghub.elsevier.com/retrieve/pii/S0140673621008692>
3. Seminara G, Carli B, Forni G, Fuzzi S, Mazzino A, Rinaldo A. Biological fluid

- dynamics of airborne COVID-19 infection. *Rend Lincei Sci Fis e Nat* [Internet]. 16 de septiembre de 2020;31(3):505-37. Disponible en: <http://link.springer.com/10.1007/s12210-020-00938-2>
4. Pizzorno J, Crinnion W. Particulate matter is a surprisingly common contributor to disease. *Integr Med* [Internet]. 2017;16(4):8-12. Disponible en: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6415634/>
 5. Liu C, Chen R, Sera F, Vicedo-Cabrera AM, Guo Y, Tong S, et al. Ambient Particulate Air Pollution and Daily Mortality in 652 Cities. *N Engl J Med* [Internet]. 22 de agosto de 2019;381(8):705-15. Disponible en: <http://www.nejm.org/doi/10.1056/NEJMoa1817364>
 6. Comunian S, Dongo D, Milani C, Palestini P. Air Pollution and COVID-19: The Role of Particulate Matter in the Spread and Increase of COVID-19's Morbidity and Mortality. *Int J Environ Res Public Health* [Internet]. 22 de junio de 2020;17(12):4487. Disponible en: <https://www.mdpi.com/1660-4601/17/12/4487>
 7. Hutter H-P, Poteser M, Moshhammer H, Lemmerer K, Mayer M, Weitensfelder L, et al. Air Pollution Is Associated with COVID-19 Incidence and Mortality in Vienna, Austria. *Int J Environ Res Public Health* [Internet]. 11 de diciembre de 2020;17(24):9275. Disponible en: <https://www.mdpi.com/1660-4601/17/24/9275>
 8. Liang D, Shi L, Zhao J, Liu P, Sarnat JA, Gao S, et al. Urban Air Pollution May Enhance COVID-19 Case-Fatality and Mortality Rates in the United States. *Innov* [Internet]. noviembre de 2020;1(3):100047. Disponible en: <https://linkinghub.elsevier.com/retrieve/pii/S2666675820300503>
 9. Chaudhry R, Dranitsaris G, Mubashir T, Bartoszko J, Riazi S. A country level analysis measuring the impact of government actions, country preparedness and socioeconomic factors on COVID-19 mortality and related health outcomes. *EClinicalMedicine* [Internet]. agosto de 2020;25:100464. Disponible en: <https://linkinghub.elsevier.com/retrieve/pii/S258953702030208X>
 10. Fatima M, O'Keefe KJ, Wei W, Arshad S, Gruebner O. Geospatial Analysis of COVID-19: A Scoping Review. *Int J Environ Res Public Health* [Internet]. 27 de febrero de 2021;18(5):2336. Disponible en: <https://www.mdpi.com/1660-4601/18/5/2336>
 11. Elías-Cuartas D, Arango-Londoño D, Guzmán-Escarria G, Muñoz E, Caicedo D, Ortega-Lenis D, et al. Análisis espacio-temporal del SARS-coV-2 en Cali, Colombia. *Rev Salud Pública* [Internet]. 1 de marzo de 2020;22(2):1-6. Disponible en: <https://revistas.unal.edu.co/index.php/revsaludpublica/article/view/86431>
 12. Rocha-García R. Un Análisis Espacial Del Contagio Del Covid-19 : El Comienzo [Internet]. *Archivos de Economía*. Bogotá, Colombia; 2020. Disponible en: <https://colaboracion.dnp.gov.co/CDT/EstudiosEconomicos/518.pdf>
 13. García JM, Fuentes HJ. Análisis del COVID-19 y probabilidades de fallecimiento para los municipios de los alrededores de Bogotá. *Espacios* [Internet]. 5 de noviembre de 2020;41(42). Disponible en: <https://revistaespacios.com/a20v41n42/a20v41n42p19.pdf>

14. Infraestructura de Datos Espaciales para el Distrito Capital (IDECA). Unidad de Planeamiento [Internet]. Secretaria Distrital de Planeación. 2021 [citado 22 de mayo de 2021]. Disponible en: <https://www.ideca.gov.co/recursos/mapas/unidad-de-planeamiento-bogota-dc>
15. Secretaria Distrital de Salud (SDS). Casos confirmados de COVID-19 en Bogotá D.C [Internet]. Datos Abiertos de Bogotá. 2021 [citado 15 de mayo de 2021]. Disponible en: <https://datosabiertos.bogota.gov.co/dataset/numero-de-casos-confirmados-por-el-laboratorio-de-covid-19-bogota-d-c>
16. Infraestructura de Datos Espaciales para el Distrito Capital (IDECA). Instituciones Prestadoras de Salud Bogotá [Internet]. Secretaría Distrital de Salud. 2021 [citado 20 de mayo de 2021]. Disponible en: <https://www.ideca.gov.co/recursos/mapas/instituciones-prestadoras-de-salud-bogota>
17. (IBOCA) IB de C del A y R en S. Reporte histórico de contaminantes [Internet]. Secretaria Distrital de Ambiente. 2021 [citado 9 de mayo de 2021]. Disponible en: <http://iboca.ambientebogota.gov.co/publicaciones/176/consulta-de-datos/>
18. Anselin L. Local Spatial Autocorrelation [Internet]. Geoda: An Introduction to Spatial Data Analysis. 2021 [citado 1 de junio de 2021]. Disponible en: https://geodacenter.github.io/workbook/6a_local_auto/lab6a.html#interpretation-of-significance
19. Chen L, Yao X, Liu Y, Zhu Y, Chen W, Zhao X, et al. Measuring Impacts of Urban Environmental Elements on Housing Prices Based on Multisource Data—A Case Study of Shanghai, China. *ISPRS Int J Geo-Information* [Internet]. 10 de febrero de 2020;9(2):106. Disponible en: <https://www.mdpi.com/2220-9964/9/2/106>
20. Wu D, Liu J. Spatial and Temporal Evaluation of Ecological Footprint Intensity of Jiangsu Province at the County-Level Scale. *Int J Environ Res Public Health* [Internet]. 26 de octubre de 2020;17(21):7833. Disponible en: <https://www.mdpi.com/1660-4601/17/21/7833>
21. Kuznetsov A, Sadovskaya V. Spatial Variation and Hotspot Detection of COVID-19 Cases in Kazakhstan, 2020. *Spat Spatiotemporal Epidemiol* [Internet]. mayo de 2021;100430. Disponible en: <https://linkinghub.elsevier.com/retrieve/pii/S1877584521000290>
22. El Deeb O. Spatial Autocorrelation and the Dynamics of the Mean Center of COVID-19 Infections in Lebanon. *Front Appl Math Stat* [Internet]. 13 de enero de 2021;6. Disponible en: <https://www.frontiersin.org/articles/10.3389/fams.2020.620064/full>
23. Matamoros LZ, Campo NMS del, García LEV, Jiménez IB. Entrada de viajeros y densidad poblacional en la propagación de la COVID-19 en Cuba. *Rev Cubana Med* [Internet]. 2020;59(3):e1375. Disponible en: <http://scielo.sld.cu/pdf/med/v59n3/1561-302X-med-59-03-e1375.pdf>
24. Rivera-Córdova A. Relación densidad y pobreza con casos confirmados por coronavirus en la región metropolitana. *Rev Médica Clínica Las Condes* [Internet]. enero de 2021;32(1):81-9. Disponible en:

- <https://linkinghub.elsevier.com/retrieve/pii/S0716864020300997>
25. Wong DWS, Li Y. Spreading of COVID-19: Density matters. Xue B, editor. PLoS One [Internet]. 23 de diciembre de 2020;15(12):e0242398. Disponible en: <https://dx.plos.org/10.1371/journal.pone.0242398>
 26. Magazzino C, Mele M, Schneider N. The relationship between air pollution and COVID-19-related deaths: An application to three French cities. Appl Energy [Internet]. diciembre de 2020;279:115835. Disponible en: <https://linkinghub.elsevier.com/retrieve/pii/S030626192031312X>
 27. Barcellos C. Unidades y escalas en los análisis espaciales en salud. Rev Cuba Salud Pública [Internet]. 2003;29(4):307-13. Disponible en: http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S0864-34662003000400003