

Spatial and temporal modeling of air pollution in Mexico City Metropolitan Area

Modelado espacial y temporal de contaminantes atmosféricos en la Zona Metropolitana de la Ciudad de México

Carmina Cruz-Huerta¹; Tomás Martínez-Trinidad^{1*}; Arian Correa-Díaz²; Armando Gómez-Guerrero¹; J. Jesús Vargas-Hernández¹; José Villanueva-Díaz³; Laura E. Beramendi-Orosco⁴

¹Colegio de Postgraduados, Postgrado en Ciencias Forestales, Campus Montecillo. km 36.5 Carretera México-Texcoco, col. Montecillo. C. P. 56230. Texcoco, Estado de México, México.

²Instituto Nacional de Investigaciones Forestales, Agrícolas y Pecuarias (INIFAP), Centro Nacional de Investigación Disciplinaria en Conservación y Mejoramiento de Ecosistemas Forestales (CENID-COMEF). Av. del Progreso núm. 5, col. Barrio de Santa Catarina. C. P. 04010. Coyoacán, Ciudad de México, México.

³Instituto Nacional de Investigaciones Forestales, Agrícolas y Pecuarias (INIFAP), Centro Nacional de Investigación Disciplinaria en relación Agua, Suelo, Planta, Atmósfera (CENID-RASPA). km 6.5 margen derecha Canal de Sacramento. C. P. 35140. Gómez Palacio, Durango, México.

⁴Universidad Nacional Autónoma de México, Instituto de Geología. Circuito Interior s/n, Ciudad Universitaria. C. P. 04510. Coyoacán, Ciudad de México, México.

*Corresponding author: tomtz@colpos.mx; tel.: +52.595.952.0200 ext. 1479.

Abstract

Introduction: Large cities have air pollution problems due to the emission of polluting gases and particulate matter (PM).

Objectives: To know the intra- and inter-annual variation of pollutants (NO_x, CO, O₃, PM₁₀ and PM_{2.5}) in Mexico City Metropolitan Area and to model their spatial distribution.

Materials and methods: Data from 44 stations of the Automatic Air Monitoring Network (RAMA) were analyzed to extract information for the pollutants NO_x, O₃ and CO in the period 1986-2021, and PM_{2.5} and PM₁₀ in the periods 2000-2021 and 2003-2021, respectively. Monthly averages per station were calculated and the temporal trend of each pollutant was evaluated using the 'Theil-Sen' operator. The spatial distribution of pollutants was also modeled and the statistical performance of four interpolation methods was compared: Neural Networks, Support Vector Machine, Random Forest and Kriging Universal.

Results and discussion: NO_x and CO concentrations were high from November to January, while O₃ from April to May. The lowest concentrations of PM₁₀ and PM_{2.5} took place from July to October and the highest in May. All pollutants decreased in concentration during the period analyzed, with the most noticeable changes in NO_x (-1.28 ppb·yr⁻¹) while CO had the smallest change (-0.12 ppm·yr⁻¹). The maximum values for NO_x, O₃ and CO occurred in 1993 and for PM in 2003. The best model was Support Vector Machine, regardless of the pollutant analyzed.

Conclusion: Spatio-temporal dynamics varied among air pollutants. The analysis with spatial interpolation methods is viable and favors solution strategies to pollution problems.

Keywords: carbon monoxide; nitrogen oxides; ozone; particulate matter; Machine Learning.

Resumen

Introducción: Las grandes ciudades presentan problemas de contaminación atmosférica por la emisión de gases contaminantes y material particulado (PM).

Objetivos: Conocer la variación intra e interanual de los contaminantes (NO_x, CO, O₃, PM₁₀ y PM_{2.5}) en la Zona Metropolitana de la Ciudad de México y modelar su distribución espacial.

Materiales y métodos: Se analizaron los datos de 44 estaciones de la Red Automática de Monitoreo Atmosférico (RAMA) para extraer información de los contaminantes NO_x, O₃ y CO en el periodo 1986-2021, y PM_{2.5} y PM₁₀ en los periodos 2000-2021 y 2003-2021, respectivamente. Se calcularon promedios mensuales por estación y se evaluó la tendencia temporal de cada contaminante mediante el operador 'Theil-Sen'. También se modeló la distribución espacial de los contaminantes y se comparó el desempeño estadístico de cuatro métodos de interpolación: Redes neuronales, *Support Vector Machine*, *Random Forest* y *Kriging Universal*.

Resultados y discusión: Las concentraciones de NO_x y CO fueron altas en noviembre-enero, mientras que las de O₃ en abril-mayo. Las concentraciones más bajas de PM₁₀ y PM_{2.5} ocurrieron en julio-octubre y las máximas en mayo. Todos los contaminantes disminuyeron su concentración durante el periodo analizado, con cambios más notorios en NO_x (-1.28 ppb·año⁻¹) mientras que CO fue el de menor cambio (-0.12 ppm·año⁻¹). Los valores máximos de NO_x, O₃ y CO se presentaron en 1993 y de PM en 2003. El mejor modelo fue *Support Vector Machine*, independientemente del contaminante analizado.

Conclusión: La dinámica espaciotemporal varió entre los contaminantes atmosféricos. El análisis con métodos de interpolación espacial es factible y favorece estrategias de solución a los problemas de contaminación.

Palabras clave: monóxido de carbono; óxidos de nitrógeno; ozono; partículas sólidas; *Machine Learning*.

Introduction

Air pollutants are one of the main problems in large cities, because of high population density, increased urbanization, transportation and industrialization (Guzmán-Morales et al., 2011). The Mexico City Metropolitan Area (MCMA) is one of the most populated regions in the world; according to the latest population and housing census, the area has 20.1 million inhabitants and an average monthly movement of more than 46 million cars (National Institute of Statistics and Geography [INEGI], 2020, 2022). This emphasizes the importance of the study and measurement of air pollutants in the urban area, especially in a spatial context (Camarillo et al., 2014).

The MCMA has unfavorable conditions for air ventilation, due to the mountains surrounding the basin of the Valley of Mexico, a situation that complicates the dispersion of pollutants (Barrera Huertas et al., 2019). The MCMA covers an area of 4726.4 km² and is formed by the municipalities of Mexico City and 16 municipalities of Estado de México. As the largest urban center in the country, the study of air pollution is essential, especially to understand the spatial-temporal dynamics and its risks to human health (Guzmán-Morales et al., 2011; López et al., 2021; Navarro, 2019).

Prolonged exposure to air pollutants is harmful to the population; for example, in Mexico City, 70 to 80 % of particulate matter (PM) medium or smaller than 10 µm (PM₁₀) are made up of 13 % of toxic metals (Chow et al., 2002). On the other hand, PM_{2.5} increment affects the respiratory system of the population causing chronic lung diseases, lung cancer and respiratory infections, highlighting the importance of measuring pollutants in urban areas (Xing et al., 2016).

The effect of pollutants is not only limited to human health but also on forest ecosystems (Romieu et al., 1996). High concentrations of heavy metals from air pollution, together with PM, inhibit seed germination and affect seedling growth and development in forests. In addition, air pollution influences biochemical and physiological processes which damage cell membranes, reduce transpiration, impede the synthesis of proteins and protein acids, and inhibit plant photosynthesis (Aliyar et al., 2020; Muhammad et al., 2021).

The Automatic Atmospheric Monitoring Network (AAMN) of MCMA has 44 stations that, from 1986 to date, provide hourly information on pollutants. Unfortunately, the distribution of the stations is not homogeneous, which is an obstacle to know the exact degree of pollution in the entire region. Therefore, AAMN information has limitations to generate strategies to solve pollution problems, particularly

Introducción

Los contaminantes atmosféricos constituyen uno de los principales problemas en las grandes ciudades, producto de una alta densidad poblacional, así como del incremento en la urbanización, transporte e industrialización (Guzmán-Morales et al., 2011). La Zona Metropolitana de la Ciudad de México (ZMCM) es una de las regiones más pobladas del mundo; de acuerdo con el último censo de población y vivienda, la zona cuenta con 20.1 millones de habitantes y el movimiento de más de 46 millones de automóviles en promedio mensual (Instituto Nacional de Estadística y Geografía [INEGI], 2020, 2022). Lo anterior enfatiza la importancia del estudio y medición de contaminantes atmosféricos en la zona urbana sobre todo en un contexto espacial (Camarillo et al., 2014).

La ZMCM tiene condiciones poco favorables para la ventilación atmosférica, debido al paisaje montañoso que rodea a la cuenca del Valle de México, situación que dificulta la dispersión de contaminantes (Barrera Huertas et al., 2019). La ZMCM comprende una superficie de 4726.4 km² y está conformada por las alcaldías de la Ciudad de México y 16 municipios del Estado de México. Por ser el centro urbano más grande del país, el estudio de la contaminación atmosférica es esencial, particularmente con el fin de entender la dinámica espacio temporal y sus riesgos para la salud humana (Guzmán-Morales et al., 2011; López et al., 2021; Navarro, 2019).

La exposición prolongada de contaminantes atmosféricos es perjudicial para la población; por ejemplo, en la Ciudad de México, 70 a 80 % de las partículas sólidas (PM) medias o menores de 10 µm (PM₁₀) está constituido por 13 % de metales tóxicos (Chow et al., 2002). Por otro lado, el incremento de las PM_{2.5} afecta el sistema respiratorio de la población ocasionando enfermedades pulmonares crónicas, cáncer de pulmón e infecciones respiratorias, lo que destaca la importancia de la medición de contaminantes en las zonas urbanas (Xing et al., 2016).

El efecto de los contaminantes no solo se limita a la salud humana sino también en los ecosistemas forestales (Romieu et al., 1996). En los bosques, las concentraciones altas de metales pesados, producto de la contaminación atmosférica, en conjunto con las PM, inhiben la germinación de las semillas y afectan el crecimiento y desarrollo de las plántulas. Además, la contaminación atmosférica influye en procesos bioquímicos y fisiológicos que dañan las membranas celulares, reducen la transpiración, impiden la síntesis de proteínas y ácidos proteicos e inhiben la fotosíntesis de las plantas (Aliyar et al., 2020; Muhammad et al., 2021).

La Red Automática de Monitoreo Atmosférico (RAMA) de la ZMCM cuenta con 44 estaciones que, de 1986 a

those related to its spatial distribution. It is therefore necessary to use tools such as geostatistics, since they help to make spatial predictions, especially in those areas where there is a lack of information (Correa et al., 2023).

Traditional geostatistical techniques such as Kriging are widely used, because they allow estimating a variable in unsampled locations based on the information provided by the sample, from the adjustment of the spatial model or empirical semivariogram (Espinoza & Molina, 2014). Recently, the use of 'Machine Learning' methods have become more popular because they allow decision making or predictions based on automated learning using computational systems and algorithms capable of learning and improving from the results (Yuan et al., 2020). The combination of these tools allows studies on the spatial distribution of data without a homogeneous distribution. Therefore, the objectives of this research were: i) to know the intra- and inter-annual variation of NO_x , CO , O_3 , $\text{PM}_{2.5}$ and PM_{10} pollutants recorded in the AAMN database for the MCMA, and ii) to compare four spatial interpolation methods, including Machine Learning techniques (Neural Networks, Support Vector Machine and Random Forest) and traditional spatial interpolation (Kriging), in order to generate maps of the spatial distribution of pollutants.

Materials and methods

Study area

The MCMA is part of an endorheic basin (Figure 1) and is located in the central part of the Transverse Neovolcanic Axis, between $19^\circ 03' - 19^\circ 54' \text{ LN}$ and $98^\circ 38' - 99^\circ 31' \text{ LW}$ with an average elevation of 2240 m. The climates are temperate humid and sub-humid with summer rains and dry weather (Villalobos, 2006).

The population living in the MCMA is 20.1 million, which represents 17 % of the national population, although slightly less than half live within Mexico City (INEGI, 2020). The dominant economic activities correspond to the service sector, commerce, and industrial activities (Espejel, 2019).

Pollutant database compilation and cleansing

Pollutant database, at station level, was downloaded from the Automatic Atmospheric Monitoring Network (AAMN) (<https://datos.cdmx.gob.mx/dataset/red-automatica-de-monitoreo-atmosferico>) for the period 1986-2021 for NO_x , O_3 and CO gases; 2000-2021 for PM_{10} and 2003-2021 for $\text{PM}_{2.5}$. The AAMN database contains information on pollutant concentrations recorded at the hourly level at each monitoring station.

la fecha, proporcionan información de contaminantes cada hora. Desafortunadamente, la distribución de las estaciones no es homogénea, lo que impide conocer con exactitud el grado de contaminación en toda la región. Por tanto, la información de RAMA tiene limitaciones para generar estrategias de solución a los problemas de contaminación, particularmente en aquellos relacionados con su distribución espacial. En este sentido, es necesario el uso de herramientas como la geoestadística, ya que permite realizar predicciones de carácter espacial, particularmente en aquellas zonas donde se carece de información (Correa et al., 2023).

Las técnicas geoestadísticas tradicionales como el *Kriging* son utilizadas ampliamente, debido a que permiten la estimación de una variable en lugares no muestreados con base en la información proporcionada por la muestra, a partir del ajuste del modelo espacial o semivariograma empírico (Espinoza & Molina, 2014). Recientemente, el uso de métodos de '*Machine Learning*' han ganado mayor popularidad, porque permiten la toma de decisiones o predicciones con base en un aprendizaje automatizado mediante sistemas computacionales y algoritmos capaces de aprender y mejorar a partir de los resultados (Yuan et al., 2020). La combinación de estas herramientas permite la realización de estudios sobre la distribución espacial de datos sin una distribución homogénea. Por tanto, los objetivos de esta investigación fueron: i) conocer la variación intra e interanual de los contaminantes NO_x , CO , O_3 , $\text{PM}_{2.5}$ y PM_{10} registrados en la base de datos de RAMA para la ZMCM, y ii) comparar cuatro métodos de interpolación espacial, incluyendo técnicas de *Machine Learning* (Redes Neuronales, *Support Vector Machine* y *Random Forest*) y de interpolación espacial tradicional (*Kriging*), con el fin de generar mapas de la distribución espacial de contaminantes.

Materiales y métodos

Área de estudio

La ZMCM forma parte de una cuenca endorreica (Figura 1) y se ubica en la parte central del Eje Neovolcánico Transversal, entre los $19^\circ 03' - 19^\circ 54' \text{ LN}$ y $98^\circ 38' - 99^\circ 31' \text{ LO}$ con una elevación promedio de 2240 m. Los climas son templados húmedo y subhúmedo con lluvias en verano y clima seco (Villalobos, 2006).

La población que habita en la ZMCM es de 20.1 millones, lo que representa 17 % de la población nacional, aunque poco menos de la mitad vive dentro de la Ciudad de México (INEGI, 2020). Las actividades económicas dominantes corresponden al sector de servicios, comercio y actividades industriales (Espejel, 2019).

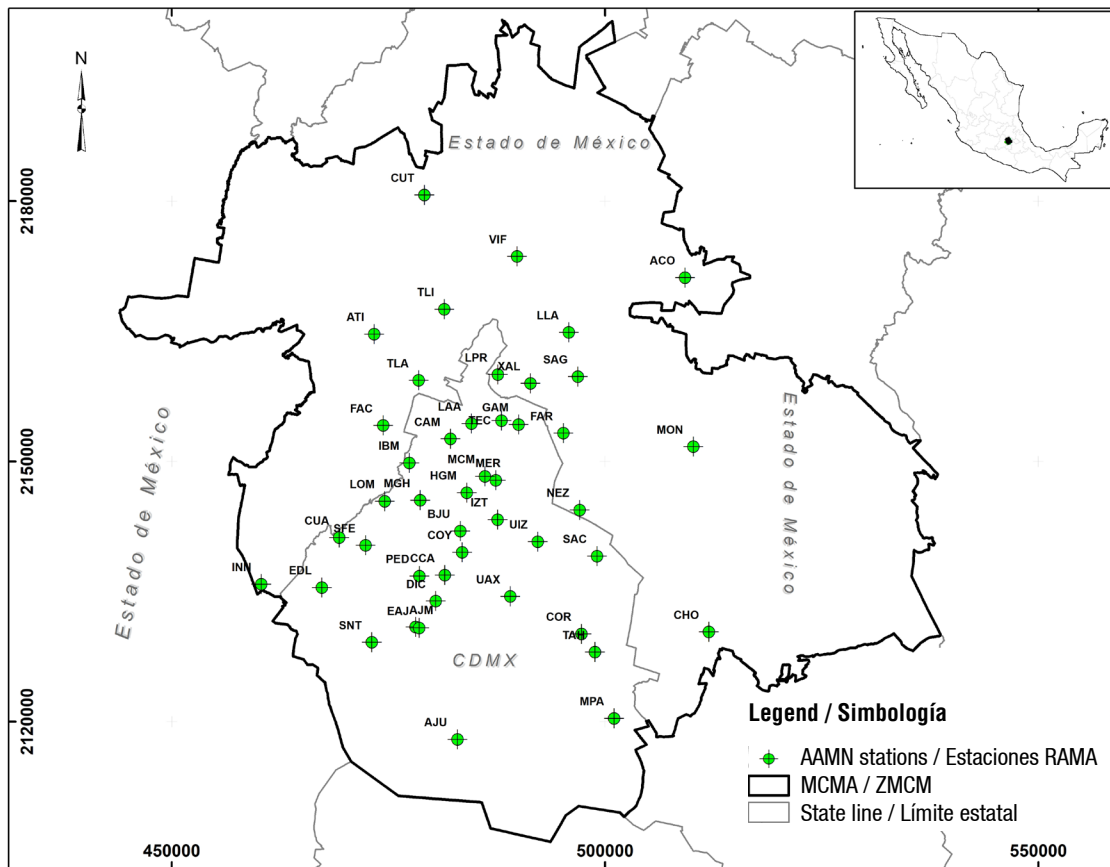


Figure 1. Study area and stations belonging to the Automatic Atmospheric Monitoring Network (AAMN) in Mexico City Metropolitan Area (MCMA).

Figura 1. Ubicación del área de estudio y de estaciones pertenecientes a la Red Automática de Monitoreo Atmosférico (RAMA) de la Zona Metropolitana de la Ciudad de México (ZMCM).

A total of 44 stations distributed in the MCMA were used; however, each pollutant had a different number of stations, since the sensors of some of them were not active on certain dates. Thus, 30, 35 and 31 stations were used for NO_x , O_3 and CO respectively, while 24 were used for $\text{PM}_{2.5}$ and PM_{10} .

The pollutant database provides information by date, time, and station. Therefore, to create a multi-annual database integrating all the stations, they were merged into a single file per pollutant using the R software. Subsequently, the data were transformed at day, month, and year levels for the corresponding analysis. The database was cleaned by removing null or erroneous data (<1 % of the total) that could affect the statistical parameters. Finally, the multi-year databases were exported in 'shape' format for further geostatistical analysis.

Descriptive statistical analysis

A descriptive statistical analysis (mean, median, minimum, and maximum) of pollutants NO_x , CO,

Compilación y depuración de base de datos de contaminantes

La base de datos de contaminantes, a nivel estación, se descargó de la plataforma Red Automática de Monitoreo Atmosférico (RAMA) (<https://datos.cdmx.gob.mx/dataset/red-automatica-de-monitoreo-atmosferico>) del periodo 1986- 2021 para los gases NO_x , O_3 y CO; 2000-2021 para PM_{10} y 2003-2021 para $\text{PM}_{2.5}$. La base de datos de RAMA contiene información de las concentraciones de los contaminantes registrados a nivel hora en cada estación de monitoreo.

Se utilizaron 44 estaciones distribuidas en la ZMCM; sin embargo, cada contaminante tuvo un número distinto de estaciones, ya que los sensores de algunas no estuvieron activos en ciertas fechas. Entonces, para NO_x , O_3 y CO se utilizaron 30, 35 y 31 estaciones, respectivamente, mientras que para $\text{PM}_{2.5}$ y PM_{10} se utilizaron 24.

La base de datos por contaminante proporciona información por fecha, hora y estación. Por tanto,

O₃, PM_{2.5} and PM₁₀ was performed; in addition, the correlation between them was evaluated using Spearman's coefficient, since the data were not adjusted to a normal distribution. The statistical analysis was performed in R version 4.0.5 (R Development Core Team, 2021).

Intra- and inter-annual variation of pollutants

Monthly average values per station and per zone were calculated to evaluate the temporal dynamics of pollutants over the course of the year. The temporal trend at the annual level was determined with the 'Theil-Sen' operator, a robust non-parametric method for obtaining temporal trends in short time series. The method fits a simple linear regression between all pairs of data and calculates the median of the slopes of all the lines (Akritas et al., 1995).

Spatial modeling methods and spatial interpolation

Spatial modeling is classified according to the following statistical techniques: a) spatial interpolation, b) spatial regression and c) Machine Learning (Chen et al., 2019; Perez et al., 2021). This study compared the predictive performance of four methods: 1) spatial interpolation (Universal Kriging) and 2) Machine Learning using supervised algorithms (Neural Network, Supported Vector Machine and Random Forest) (Castro et al., 2017; Pedrero et al., 2021). These methods were used to model the monthly and annual spatial distribution of each pollutant in the MCMA. The modeling and spatial interpolation analysis was carried out with the R software version 4.0.5 (R Development Core Team, 2021).

Interpolation with the Universal Kriging method was carried out with the 'autoKrige' function, which adjusts variograms in a grouped way from different models (spherical, exponential, and Gaussian), to finally perform the prediction with the optimal model in the areas with missing data (Estarlich et al., 2013).

For the Neural Network analysis, we used the statistical package 'R neuralnet', considering five layers of neurons arbitrarily chosen to train the model. For Supported Vector Machine, the 'ksvm' function of the 'kernlab' package was used, considering a polynomial Kernel algorithm with a penalty parameter of 25, which avoids overfitting the data (García & Lozano, 2007). Finally, Random Forest interpolation was carried out with the 'ranger' library in R, considering 1 000 trees and a single node (Espinosa-Zuñiga, 2020).

Unlike traditional interpolation methods (Universal Kriging), where the spatial location of the variables of

para construir una base de datos multianual que integrara todas las estaciones, estas se unieron en un solo archivo por contaminante mediante el *software* R. Posteriormente, los datos se transformaron a nivel día, mes y año para su análisis correspondiente. La base de datos se depuró mediante la eliminación de datos nulos o erróneos (<1 % del total) que pudieran afectar los parámetros estadísticos. Por último, las bases de datos multianuales se exportaron en formato 'shape' para su análisis geoestadístico posterior.

Análisis estadístico descriptivo

Se realizó un análisis estadístico descriptivo (media, mediana, mínimo y máximo) de los contaminantes NO_x, CO, O₃, PM_{2.5} y PM₁₀; además, se evaluó la correlación entre estos mediante el coeficiente de Spearman, dado que los datos no se ajustaron a una distribución normal. El análisis estadístico se hizo en R versión 4.0.5 (R Development Core Team, 2021).

Variación intra e interanual de los contaminantes

Para evaluar la dinámica temporal de los contaminantes en el transcurso del año, se calcularon los valores promedio mensuales por estación y por zona. La tendencia temporal a nivel anual se determinó con el operador 'Theil-Sen', un método no paramétrico robusto que permite obtener tendencias temporales en series cortas de tiempo. El método ajusta una regresión lineal simple entre todos los pares de datos y calcula la mediana de las pendientes de todas las líneas (Akritas et al., 1995).

Métodos de modelación e interpolación espacial

La modelación espacial se clasifica de acuerdo con las técnicas estadísticas siguientes: a) interpolación espacial, b) regresión espacial y c) *Machine Learning* (Chen et al., 2019; Pérez et al., 2021). En este estudio, se comparó el desempeño predictivo de cuatro métodos: 1) interpolación espacial (*Kriging Universal*) y 2) *Machine Learning* considerando algoritmos supervisados (*Neural Network*, *Supported Vector Machine* y *Random Forest*, por sus nombres en inglés) (Castro et al., 2017; Pedrero et al., 2021). Estos métodos se utilizaron para modelar la distribución espacial mensual y anual de cada uno de los contaminantes en la ZMCM. El análisis de modelación e interpolación espacial se hizo con el *software* R versión 4.0.5 (R Development Core Team, 2021).

La interpolación con el método *Kriging Universal* se realizó con la función 'autoKrige', la cual ajusta variogramas de forma agrupada de diferentes modelos (esférico, exponencial y gaussiano), para realizar finalmente la predicción con el modelo óptimo en las áreas con datos faltantes (Estarlich et al., 2013).

interest (e.g., coordinates) is used, Machine Learning modeling and interpolation uses distance matrices between pairs of points, because coordinates are correlated with each other. Finally, with the average of all data, monthly and historical pollutant distribution maps were created.

Model validation and statistical performance

Model performance was evaluated using the coefficient of determination (R^2), mean absolute error (MAE) and root mean square error (RMSE):

$$R^2 = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y}_i)^2}$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |\hat{y}_i - y_i|$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n |\hat{y}_i - y_i|^2}$$

where,

\hat{y}_i = th value obtained from the prediction

y_i = th observed value

\bar{y}_i = mean value of y_i

n = number of predicted or observed values with $i = 1, 2, \dots, n$

These metrics help identifying the best model fitting the data; low RMSE values indicate a better fit, R^2 indicates the goodness of fit of the model and MAE is a linear score and means that individual differences are weighted equally in the average (Beguin et al., 2017; Pérez et al., 2021).

Results and Discussion

Pollution historical statistics in the MCMA

The average of NO_x was 20.8 ± 0.17 ppb with a historical maximum of 111 ppb (Hangares station, inactive) and a minimum of 0.1 ppb at Milpa Alta station (Table 1). Some authors indicate NO_x decrease in the MCMA (Sandoval & Jaimes, 2002; Navarro, 2019), although the Secretariat of the Environment (SEDEMA) reports NO_x increment in 2018. These authors emphasize that mobile sources contribute more than 85 % of pollutant emissions in the MCMA.

The average CO was 1.86 ± 0.01 ppm, a value below the reference limit (9.0 ppm) of NOM-021-SSA1-202

Para el análisis de *Neural Network* se utilizó el paquete estadístico 'R neuralnet', considerando cinco capas de neuronas elegidas arbitrariamente para entrenar el modelo. Para *Supported Vector Machine*, se utilizó la función 'ksvm' del paquete 'kernlab', considerando un algoritmo de Kernel polinomial con un parámetro de penalización de 25, lo cual evita sobreajuste de los datos (García & Lozano, 2007). Finalmente, la interpolación con *Random Forest* se hizo con la librería 'ranger' en R, considerando 1000 árboles y un solo nodo (Espinosa-Zuñiga, 2020).

A diferencia de los métodos de interpolación tradicionales (*Kriging Universal*), donde se utiliza la ubicación espacial de las variables de interés (v. g., coordenadas), la modelación e interpolación para *Machine Learning* utiliza matrices de distancia entre pares de puntos, debido a que las coordenadas están correlacionadas entre sí. Finalmente, con el promedio de todos los datos, se generaron mapas de distribución de contaminantes por mes e históricos.

Validación y desempeño estadístico de los modelos

El desempeño de los modelos se evaluó mediante el coeficiente de determinación (R^2), error medio absoluto (MAE) y error cuadrático medio (RMSE):

$$R^2 = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y}_i)^2}$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |\hat{y}_i - y_i|$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n |\hat{y}_i - y_i|^2}$$

donde,

\hat{y}_i = iésimo valor obtenido de la predicción

y_i = iésimo valor observado

\bar{y}_i = valor medio de y_i

n = número de valores predichos u observados con $i = 1, 2, \dots, n$

Estas métricas permiten la identificación del mejor modelo que se ajusta a los datos; los valores bajos de RMSE indican un mejor ajuste, R^2 indica la bondad o la aptitud del modelo y el MAE es un puntaje lineal y significa que las diferencias individuales se ponderarán por igual en el promedio (Beguin et al., 2017; Pérez et al., 2021).

(Secretaría de Salud, 2021). The maximum value was 13.0 ppm for Tlalpan station, and the minimum was 0.1 ppm for Milpa Alta station. SEDEMA mentions that CO was the pollutant with the highest total absolute emission in 2018 (646434 Mg), representing 75.3 % of total air pollutant emissions, mainly product of incomplete combustion of gasoline, natural gas, oil, and other organic materials, according to that reported by Miller (2011).

O₃ is a secondary gas formed by chemical and photochemical reactions between anthropogenic and natural primary emissions of precursors nitrogen oxides (NO_x) and volatile organic compounds (VOC) or hydrocarbons (Calderón et al., 2000). O₃ had a historical mean of 33.3 ± 0.12 ppb, with a maximum concentration of 96.9 ppb for the Cerro de la Estrella (CES) station, exceeding the limits established of 90.0 ppb by NOM-020-SSA1-2021 (Secretaría de Salud, 2021).

For PM₁₀, an average of 47.7 ± 0.28 µg·m⁻³ was obtained with a maximum of 145 µg·m⁻³ for Xalostoc station, to the north of the MCMA. For this station, the main sources of suspended particulate matter are soil erosion, industries, and unpaved roads, which increase PM₁₀ concentrations (Cervantes et al., 2005). The historical minimum of PM₁₀ was 11.4 µg·m⁻³ for Investigaciones Nucleares (INN) station, to the south of the MCMA. Meanwhile, PM_{2.5} particles have a historical mean of 23.6 ± 0.14 µg·m⁻³ with a maximum for San Agustín (SAG) of 68.2 µg·m⁻³ and a minimum of 7.69 µg·m⁻³ for INN. Chow et al. (2002) mention that PM_{2.5} and PM₁₀ particles are composed of nitrates, sulfates, ammonium, organic carbon, elemental carbon, and geological material, which in large quantities are harmful to the ecosystem.

Resultados y discusión

Estadísticos históricos de contaminantes en la ZMCM

El promedio de NO_x fue 20.8 ± 0.17 ppb con un máximo histórico de 111 ppb (estación Hangares, inactiva) y un mínimo de 0.1 ppb en la estación Milpa Alta (Cuadro 1). Algunos autores indican disminución de NO_x en la ZMCM (Sandoval & Jaimes, 2002; Navarro, 2019), aunque la Secretaría del Medio Ambiente (SEDEMA) reporta aumento de NO_x en 2018. Estos autores enfatizan que las fuentes móviles son las que aportan más del 85 % de las emisiones contaminantes en la ZMCM.

El promedio de CO fue 1.86 ± 0.01 ppm, valor por debajo del límite de referencia (9.0 ppm) de la NOM-021-SSA1-202 (Secretaría de Salud, 2021). El valor máximo fue 13.0 ppm en la estación Tlalpan y el mínimo fue 0.1 ppm en la estación Milpa Alta. La SEDEMA menciona que CO fue el contaminante con mayor emisión absoluta total en 2018 (646434 Mg), representando 75.3 % del total de las emisiones contaminantes atmosféricas, producto principalmente de la combustión incompleta de gasolina, gas natural, petróleo y otros materiales orgánicos, en concordancia por lo reportado por Miller (2011).

El O₃ es un gas secundario formado por reacciones químicas y fotoquímicas entre las emisiones primarias antropogénicas y naturales de sus precursores óxidos de nitrógeno (NO_x) y compuestos orgánicos volátiles (COV) o hidrocarburos (Calderón et al., 2000). El O₃ tuvo una media histórica de 33.3 ± 0.12 ppb, con una concentración máxima de 96.9 ppb en la estación Cerro de la Estrella (CES), sobrepasando los límites

Table 1. Descriptive statistics of pollutants reported by the Automatic Atmospheric Monitoring Network (AAMN) in Mexico City Metropolitan Area.

Cuadro 1. Estadística descriptiva de los contaminantes reportados por la Red Automática de Monitoreo Atmosférico (RAMA) en la Zona Metropolitana de la Ciudad de México.

Pollutant/ Contaminante	NO _x (ppb)	CO (ppm)	O ₃ (ppb)	PM ₁₀ (µg·m ⁻³)	PM _{2.5} (µg·m ⁻³)
Average/Media	22.0 ± 0.17	1.86 ± 0.01	33.3 ± 0.12	47.7 ± 0.28	23.6 ± 0.14
Median/Mediana	20.8 ± 0.17	1.35 ± 0.01	30.8 ± 0.12	44.4 ± 0.28	22.8 ± 0.14
Minimum/Mínimo	0.10 _(MPA)	0.10 _(MPA)	5.75 _(UIZ)	11.40 _(INN)	7.69 _(INN)
Maximum/ Máximo	111.0 _(HAN*)	13.0 _(TLA)	96.9 _(CES*)	145.0 _(XAL)	68.2 _(SAG)
Period/Periodo	1986-2021	1986-2021	1986-2021	2000-2021	2003-2021
n	7042	8755	8613	4224	2682

NO_x: nitrogen oxides, CO: carbon monoxide, O₃: ozone, PM₁₀ and PM_{2.5}: medium particulate matter or smaller than 10 and 2.5 µm, respectively. MPA: Milpa Alta, HAN: Hangares, TLA: Tlalnepantla, UIZ: UAM Iztapalapa, CES: Cerro de la Estrella, INN: Investigaciones Nucleares, XAL: Xalostoc, SAG: San Agustín, *Inactive stations.

NO_x: óxidos de nitrógeno, CO: monóxido de carbono, O₃: ozono, PM₁₀ y PM_{2.5}: partículas sólidas medias o menores de 10 y 2.5 µm, respectivamente. MPA: Milpa Alta, HAN: Hangares, TLA: Tlalnepantla, UIZ: UAM Iztapalapa, CES: Cerro de la Estrella, INN: Investigaciones Nucleares, XAL: Xalostoc, SAG: San Agustín, *Estaciones inactivas.

Intra- and inter-annual variation of pollutants for MCMA

The highest NO_x concentrations were recorded in December and January (>60 ppb) and decreased during June-August (<50 ppb) (Figure 2a). Some stations were also observed to exceed 100 ppb of NO_x . These high concentrations are explained by the prevailing meteorological conditions and thermal inversions in winter, since low radiation and temperature, for example, are associated with high NO_x concentrations (Sandoval & Jaimes, 2002).

CO had no significant variations over the year (Figure 2b), because concentrations of this compound depend on emissions from automobiles and industries, which remain constant over the course of the year (Madrigal et al., 2004). O_3 (Figure 2c) the highest concentrations in April and May (>40 ppb), which coincides with the high level of solar radiation emitted in that period with different lengths of ultraviolet (UV) radiation (Wedyan et al., 2020). These, in turn, dissociate oxygen by photochemical reaction and, when available, react with other surrounding molecules (NO_x , VOC and CO) allowing the formation of ozone. Possibly, cloudiness and lower radiation in December and January decrease UV radiation and, therefore, reduce O_3 concentrations.

$\text{PM}_{2.5}$ and PM_{10} depend on the emission of particulate matter from transportation, industry, residence, commerce, and services (Popovicheva et al., 2020). The period with the lowest concentration of these particles was during the rainy season (June-August, Figure 2d-e). This information is consistent with other studies indicating that PM concentrations decrease in the rainy season, which plays a role of wet removal of particles, but increase in dry periods due to accumulation of dust in foliage (Vinasco & Nastar, 2013; Zhou et al., 2020).

Interannual variation of pollutants in the MCMA

Pollutant behavior at the interannual level was variable; for example, the maximum peak of NO_x was recorded in 1993 (Figure 3), as well as O_3 and CO. This coincides with studies that indicate that 1993 was the coldest year with the least precipitation in the MCMA (Pérez et al., 2010). These emissions exceeded the permissible pollution limits and the standards established for the central zone (Mercado et al., 1995). In that year, the highest pollution was caused by emissions from pharmaceutical companies, plastic articles and basic iron and steel industries, which emitted 64 % of atmospheric emissions in the Valley of Mexico. These companies use fossil fuels and increase CO, O_3 , NO_x and VOC (Mercado et al., 1995). Since 1995, concentrations have decreased, particularly during the last decade, even though vehicle flow and

establecidos de 90.0 ppb por la NOM-020-SSA1-2021 (Secretaría de Salud, 2021).

Para PM_{10} se obtuvo un promedio de $47.7 \pm 0.28 \mu\text{g}\cdot\text{m}^{-3}$ con un máximo de $145 \mu\text{g}\cdot\text{m}^{-3}$ en la estación de Xalostoc, al norte de la ZMCM. En esta estación, las fuentes principales de partículas suspendidas son la erosión del suelo, las industrias y los caminos sin pavimentar, que incrementan las concentraciones de PM_{10} (Cervantes et al., 2005). El mínimo histórico de PM_{10} fue de $11.4 \mu\text{g}\cdot\text{m}^{-3}$ en la estación de Investigaciones Nucleares (INN), al sur de la ZMCM. Por su parte, las partículas $\text{PM}_{2.5}$ poseen una media histórica de $23.6 \pm 0.14 \mu\text{g}\cdot\text{m}^{-3}$ con un máximo en San Agustín (SAG) de $68.2 \mu\text{g}\cdot\text{m}^{-3}$ y un mínimo de $7.69 \mu\text{g}\cdot\text{m}^{-3}$ en INN. Chow et al. (2002) mencionan que las partículas $\text{PM}_{2.5}$ y PM_{10} están compuestas por nitratos, sulfatos, amonio, carbono orgánico, carbono elemental y material geológico, que en grandes cantidades son perjudiciales para el ecosistema.

Variación intra e interanual de contaminantes en la ZMCM

Las mayores concentraciones de NO_x se registraron en diciembre y enero (>60 ppb) y disminuyeron durante junio-agosto (<50 ppb) (Figura 2a). También se observó que hay estaciones que superan las 100 ppb de NO_x . Estas altas concentraciones se explican por las condiciones meteorológicas dominantes e inversiones térmicas en invierno, ya que, por ejemplo, la radiación y temperatura bajas se asocian con concentraciones altas de NO_x (Sandoval & Jaimes, 2002).

El CO no tuvo variaciones significativas a lo largo del año (Figura 2b), debido a que las concentraciones de este compuesto dependen de las emisiones de los automóviles e industrias, las cuales se mantienen constantes en el transcurso del año (Madrigal et al., 2004). Por su parte, el O_3 (Figura 2c) presentó las mayores concentraciones en abril y mayo (>40 ppb), lo cual coincide con la cantidad elevada de radiación solar emitida en ese periodo con diferentes longitudes de radiación ultravioleta (UV) (Wedyan et al., 2020). Estas, a su vez, disocian el oxígeno por reacción fotoquímica y, cuando se encuentra disponible, reacciona con otras moléculas circundantes (NO_x , COV y CO) permitiendo la formación de ozono. Posiblemente, la nubosidad y una menor radiación en diciembre y enero disminuyen la radiación UV y, por tanto, reducen las concentraciones de O_3 .

Las $\text{PM}_{2.5}$ y PM_{10} dependen de la emisión de partículas por transporte, industria, residencia, comercio y servicios (Popovicheva et al., 2020). El periodo con menor concentración de estas partículas fue durante la temporada de lluvias (junio-agosto, Figura 2d-e). Esta información coincide con otros estudios que señalan que las concentraciones de PM disminuyen en la

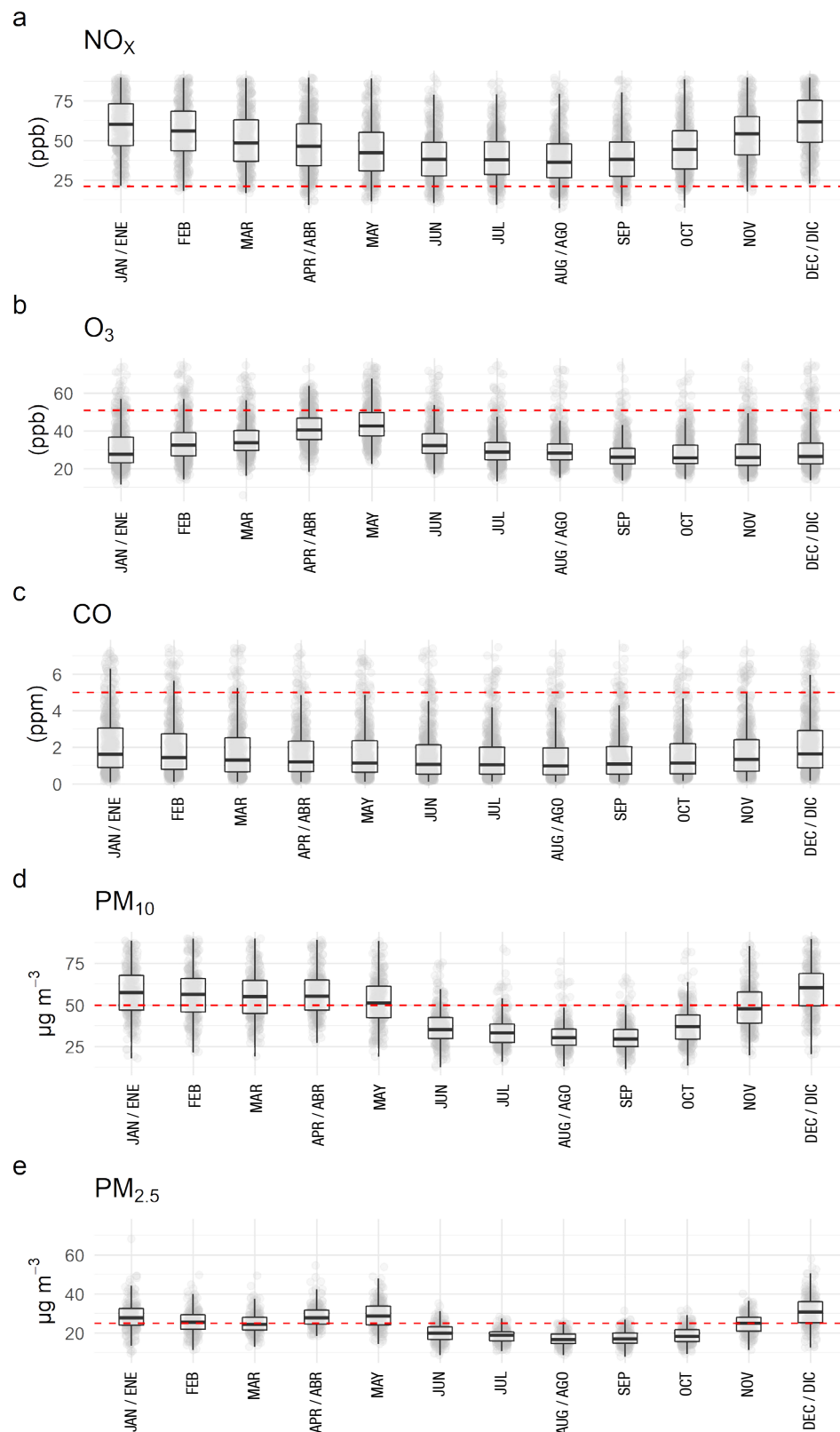


Figure 2. Monthly distribution of nitrogen oxides (NO_x), ozone (O₃), carbon monoxide (CO), particulate matter 10 (PM₁₀) and 2.5 (PM_{2.5}) in Mexico City Metropolitan Area. The dotted line indicates the permissible limit for pollutants according to NOM-023-SSA1-2021, NOM-020-SSA1-2021, NOM-021-SSA1-2021 and NOM-025-SSA1-2021.

Figura 2. Distribución mensual de los óxidos de nitrógeno (NO_x), ozono (O₃), monóxido de carbono (CO), partículas sólidas 10 (PM₁₀) y 2.5 (PM_{2.5}) en la Zona Metropolitana de la Ciudad de México. La línea punteada señala el límite permisible de los contaminantes según las normas NOM-023-SSA1-2021, NOM-020-SSA1-2021, NOM-021-SSA1-2021 y NOM-025-SSA1-2021.

number of vehicles have increased. This decrease can be attributed to improvements in automobile manufacturing, in addition to the implementation of the program to improve air quality in the Valley of Mexico (Sheinbaum, 2016).

The concentration of PM_{10} in the MCMA for the period 2000-2012 exceeded the limits of $50 \mu\text{g}\cdot\text{m}^{-3}$ indicated by NOM-025-SSA1-2021 (Secretaría de Salud, 2021). This increase is mainly attributed to emissions derived from transportation, industrial activity, and dust reincorporation from vehicle circulation (Villalobos, 2006). The lowest PM_{10} concentration was recorded in 2019 and 2020, which is probably related to the drastic closure of activities due to the COVID-19 pandemic. For example, in 2020, vehicular traffic was reduced in the MCMA, which restricted public mobility and reduced productive and industrial activities (Ale et al., 2020).

$PM_{2.5}$ exceeded the limit value established by NOM-025-SSA1-2021 and the World Health Organization (OMS, 2005: $25 \mu\text{g}\cdot\text{m}^{-3}$) in almost all the period analyzed; although in 2020, concentrations decreased, probably also because of the reduction of activities due to the social confinement derived from the COVID-19 pandemic. On the other hand, the temporal trends of NO_x , CO, O_3 , PM_{10} and $PM_{2.5}$ are negative and indicate a decrease of pollutants in MCMA (Figure 3). The pollutant with the highest decrement was NO_x ($-1.28 \text{ ppb}\cdot\text{yr}^{-1}$) and the lowest was CO ($-0.12 \text{ ppm}\cdot\text{yr}^{-1}$), while $PM_{2.5}$ and O_3 had similar slopes $-0.47 \mu\text{g}\cdot\text{m}^{-3}\cdot\text{yr}^{-1}$ and $-0.45 \text{ ppb}\cdot\text{yr}^{-1}$, respectively.

On the other hand, Spearman's coefficient indicates a high correlation between NO_x , CO, PM_{10} and $PM_{2.5}$; however, O_3 was not significantly associated with PM_{10} and $PM_{2.5}$ (Table 2). The highest correlation values were between NO_x , CO and O_3 , which could be attributed to the photochemical reaction involved in the formation of O_3 ; that is, oxygen, once available, reacts with other NO_x and CO compounds (Jenkin & Clemitshaw, 2000).

Interpolation of pollutants in MCMA

No notable difference was found in the performance of the models Kriging Universal, Supported Vector Machine, Random Forest, and Neural Network at the monthly level; however, the Supported Vector Machine method was slightly superior with R^2 values = 0.98; while the models with the lowest fit were Kriging Universal and Neural Network with $R^2 < 0.85$.

Based on the historical average of each pollutant, the Supported Vector Machine model had the best goodness of fit with R^2 greater than 0.95, except for CO, with $R^2 = 0.76$ (Table 3). For MAE and RMSE, the models with the best fit were Neural Network and Supported

temporada de lluvias, la cual desempeña un papel de remoción húmeda de partículas, pero incrementan en periodos secos por acumulación de polvo en el follaje (Vinasco & Nastar, 2013; Zhou et al., 2020).

Variación interanual de los contaminantes en la ZMCM

El comportamiento de los contaminantes a nivel interanual fue variable; por ejemplo, el pico máximo de NO_x se registró en 1993 (Figura 3), así como de O_3 y CO. Esto coincide con estudios que señalan que el año 1993 fue el más frío y con menor precipitación en la ZMCM (Pérez et al., 2010). Estas emisiones rebasaron los límites permisibles de contaminación y las normas establecidas para la zona centro (Mercado et al., 1995). En ese año se tuvo mayor contaminación como producto de las emisiones de empresas farmacéuticas, artículos plásticos e industrias básicas de hierro y acero, las cuales causaron 64 % de emisiones atmosféricas en el Valle de México. Estas empresas usan combustibles fósiles e incrementan las emisiones de CO, O_3 , NO_x y COV (Mercado et al., 1995). A partir de 1995, las concentraciones han disminuido, particularmente durante la última década, a pesar de que el flujo vehicular y el número de vehículos han incrementado. Tal disminución se puede atribuir a las mejoras en la fabricación de automóviles, además de la implementación del programa para mejorar la calidad de aire en el Valle de México (Sheinbaum, 2016).

La concentración de PM_{10} en la ZMCM durante el periodo 2000-2012 excedió los límites de $50 \mu\text{g}\cdot\text{m}^{-3}$ señalados por la NOM-025-SSA1-2021 (Secretaría de Salud, 2021). Este incremento se atribuye principalmente a las emisiones derivadas de transporte, la actividad industrial y a la reincorporación de polvos por la circulación de vehículos (Villalobos, 2006). Durante 2019 y 2020 se registró la menor concentración de PM_{10} , lo cual probablemente esté relacionado con el cierre drástico de actividades, producto de la pandemia COVID-19. Por ejemplo, durante 2020 se redujo el tránsito vehicular en la ZMCM, lo que restringió la movilidad pública y la disminución de actividades productivas e industriales (Ale et al., 2020).

Las $PM_{2.5}$ excedieron el valor límite establecido por la NOM-025-SSA1-2021 y la Organización Mundial de la Salud (OMS, 2005: $25 \mu\text{g}\cdot\text{m}^{-3}$) en la mayor parte del periodo analizado; aunque en el 2020, las concentraciones disminuyeron, probablemente también a causa de la reducción de actividades por el confinamiento social derivado de la pandemia COVID-19. Por otro lado, las tendencias temporales de NO_x , CO, O_3 , PM_{10} y $PM_{2.5}$ son negativas e indican disminución de los contaminantes en la ZMCM (Figura 3) El contaminante con mayor disminución fue NO_x

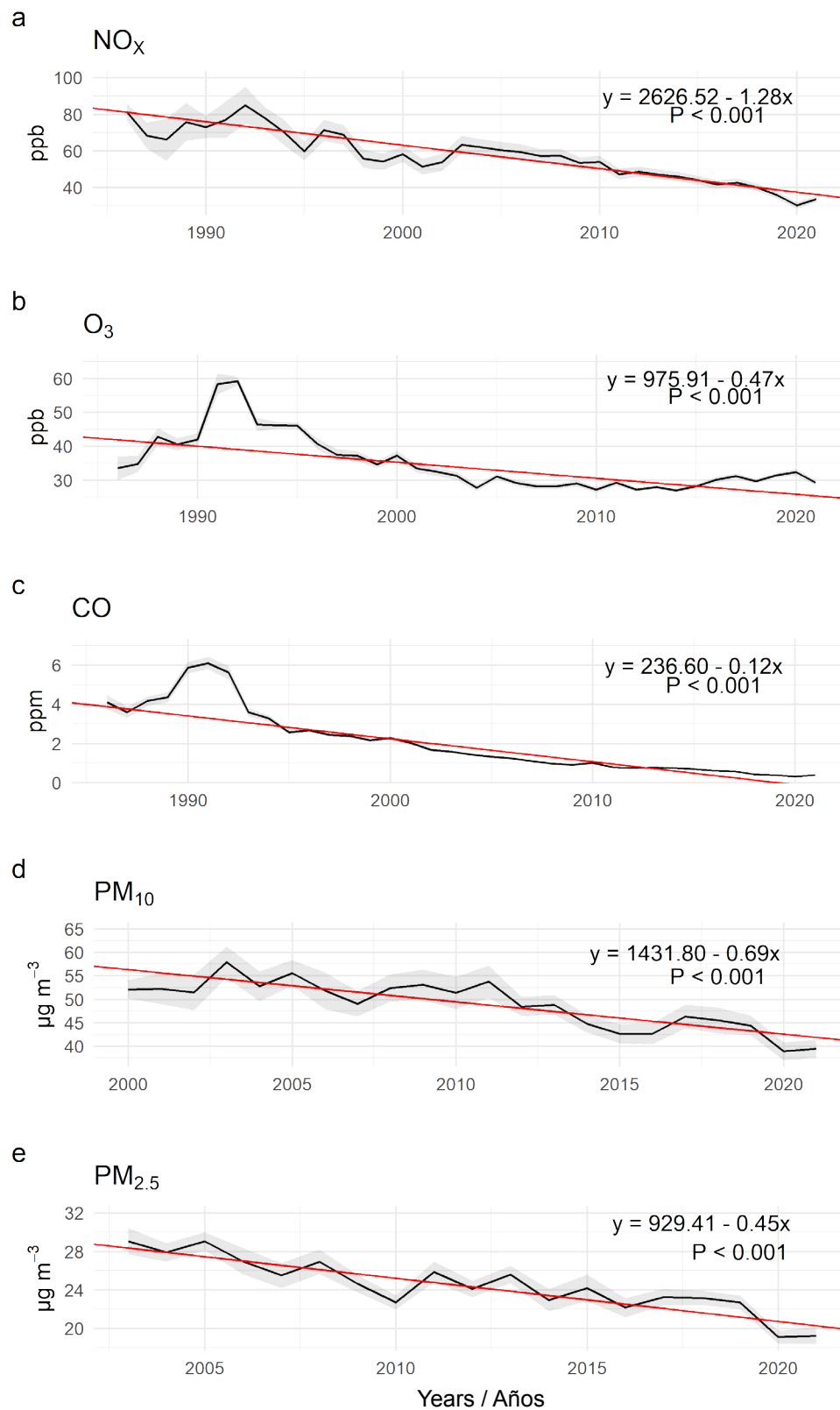


Figure 3. Interannual behavior of nitrogen oxides (NO_x), ozone (O₃), carbon monoxide (CO), particulate matter 10 μm (PM₁₀) and 2.5 μm (PM_{2.5}) in Mexico City Metropolitan Area. The light gray area represents the standard error among the stations included in the analysis.

Figura 3. Comportamiento interanual de los óxidos de nitrógeno (NO_x), ozono (O₃), monóxido de carbono (CO), partículas sólidas 10 μm (PM₁₀) y 2.5 μm (PM_{2.5}) en la Zona Metropolitana de la Ciudad de México. El área gris clara representa el error estándar entre las estaciones incluidas en el análisis.

Table 2. Spearman correlation coefficient (*P < 0.001) among pollutants analyzed in Mexico City Metropolitan Area: nitrogen oxides (NO_x), ozone (O₃), carbon monoxide (CO), particulate matter 10 (PM₁₀) and particulate matter 2.5 (PM_{2.5}).**

Cuadro 2. Coeficiente de correlación de Spearman (*P < 0.001) entre los contaminantes analizados en la Zona Metropolitana de la Ciudad de México: óxidos de nitrógeno (NO_x), ozono (O₃), monóxido de carbono (CO), partículas sólidas 10 (PM₁₀) y partículas sólidas 2.5 (PM_{2.5}).**

Pollutant / Contaminante	NO _x	O ₃	CO	PM ₁₀	PM _{2.5}
NO _x	1	0.95***	0.93***	0.85***	0.88***
O ₃		1	0.77***	0.05	-0.14
CO			1	0.79***	0.86***
PM ₁₀				1	0.86***
PM _{2.5}					1

Table 3. Model validation statistics for the interpolation of historical averages by pollutant (NO_x: nitrogen oxides, CO: carbon monoxide, O₃: ozone, PM: particulate matter) in Mexico City Metropolitan Area.

Cuadro 3. Estadísticas de validación de los modelos para la interpolación de los promedios históricos por contaminante (NO_x: óxidos de nitrógeno, CO: monóxido de carbono, O₃: ozono, PM: partículas sólidas) en la Zona Metropolitana de la Ciudad de México.

Method / Método	Statistics / Estadístico	Pollutant / Contaminante				
		NO _x (ppb)	O ₃ (ppb)	CO (ppm)	PM ₁₀ (µg·m ⁻³)	PM _{2.5} (µg·m ⁻³)
Neural Network	MAE	1.52	0.49	0.07	1.05	0.35
	R ²	0.79	0.64	0.37	0.98	0.94
	RMSE	1.58	2.76	0.07	1.08	0.36
Kriging Universal	MAE	5.99	2.39	0.48	3.25	1.39
	R ²	0.71	0.64	0.37	0.78	0.72
	RMSE	7.85	2.76	0.60	4.97	1.83
Random Forest	MAE	5.43	2.57	0.42	0.86	0.55
	R ²	0.86	0.81	0.69	0.80	0.83
	RMSE	6.72	2.85	0.54	1.18	0.86
Support Vector Machine	MAE	4.30	1.65	0.28	3.27	1.05
	R ²	0.98	0.98	0.76	0.99	0.98
	RMSE	5.59	1.96	0.34	4.74	1.44

MAE: mean absolute error, RMSE: root mean square error.

MAE: error medio absoluto, RMSE: error cuadrático medio.

Vector Machine. The Neural Network model had MAE and RMSE lower than 3.0, although R² was lower than 0.8 for NO_x, O₃ and CO. Supported Vector Machine had MAE and RMSE lower than 3.0 for all pollutants except for NO_x (MAE and RMSE between 4 and 6). On the other hand, the results indicated underestimation in the predicted values using Neural Network, Random Forest, and Universal Kriging, mainly for high values for the pollutants CO and O₃ (Figure 4). Therefore, the Supported Vector Machine model had the best characteristics for modeling and interpolating the pollutants analyzed at the monthly level, as well as the historical average.

(-1.28 ppb·año⁻¹) y el de menor CO (-0.12 ppm·año⁻¹), mientras que PM_{2.5} y O₃ presentan pendientes similares -0.47 µg·m⁻³·año⁻¹ y -0.45 ppb·año⁻¹, respectivamente.

Por otra parte, el coeficiente de Spearman señala correlación alta entre los contaminantes NO_x, CO, PM₁₀ y PM_{2.5}; sin embargo, el O₃ no mostró asociación significativa con PM₁₀ y PM_{2.5} (Cuadro 2). Los valores de correlación más altos se presentaron entre NO_x, CO y O₃, lo cual puede atribuirse a la reacción fotoquímica que existe en la formación del O₃; es decir, el oxígeno, una vez disponible, reacciona con otros compuestos de NO_x y CO (Jenkin & Clemitshaw, 2000).

Geostatistical modeling may have limitations in the selection of parameters, specially supervised Machine Learning methods. For example, Random Forest predictions can be expected to be beyond the range of observed values, if any data group presents confusion (Espinosa-Zuñiga, 2020). In the case of Neural Network, the main disadvantage is that neither its final equation nor the weights used in the model are known, making it a black box (Sheu, 2020). Kriging Universal requires an optimal variogram for the dataset to avoid extrapolations, and it is sensitive to a low number of points or a high variation between them (Shekaramiz et al., 2019). Finally, with Supported Vector Machine, Kernel and C parameters should be chosen appropriately, as they affect the complexity of the model (Cunha et al., 2022; Liu & Xu, 2014). Despite these limitations, Supported Vector Machine had the best fit for all pollutants and was therefore used to represent their spatial distribution.

Spatial dynamics of pollutants in MCMA

According to the spatial analysis, different patterns of pollutant distribution were found (Figure 5). For example, NO_x showed a circular pattern, where the area with the highest concentration was located in the central zone of Mexico City with two stations with high values (La Merced and Xalostoc > 60 ppb). O_3 had the highest concentrations at the Pedregal and Milpa Alta stations, located in the southern part of the study area. O_3 concentration decreased towards the central and northern part of MCMA, indicating a high gradient in the southwestern quadrant. These results are consistent with the study by García (2009), who indicates that the highest pollutant areas are located downwind, away from emission sources, so high O_3 concentration is located in mountainous areas south of Mexico City (Figure 5).

The highest concentration of CO was found in the northwest central quadrant, probably due to a higher use of vehicles, which are the main source of emissions. However, some authors have reported a relationship between high CO concentrations and areas where the population has higher incomes in the Mexico Valley Metropolitan Area (MCMA), especially for CO and carbon dioxide emissions (CO_2 eq) (Pérez et al., 2018).

PM_{10} and $\text{PM}_{2.5}$ concentrations were higher in the center and decreased towards the west. The station with the highest concentration is Xalostoc with values above $70 \mu\text{g}\cdot\text{m}^{-3}$ of PM_{10} and $30 \mu\text{g}\cdot\text{m}^{-3}$ of $\text{PM}_{2.5}$. This may be associated with the findings made by Cervantes et al. (2005), who affirm that the station is located in an area with evidence of soil erosion, presence of industries and unpaved roads.

Interpolación de los contaminantes en la ZMCM

No se encontró diferencia notable en el desempeño de los modelos *Kriging Universal*, *Supported Vector Machine*, *Random Forest* y *Neural Network* a nivel mensual; sin embargo, el método *Supported Vector Machine* fue ligeramente superior con valores de $R^2 = 0.98$; mientras que los modelos con menor ajuste fueron *Kriging Universal* y *Neural Network* con $R^2 < 0.85$.

Con base en el promedio histórico de cada contaminante, el modelo *Supported Vector Machine* presentó la mejor bondad con R^2 superiores a 0.95, con excepción de CO, el cual obtuvo $R^2 = 0.76$ (Cuadro 3). Respecto a MAE y RMSE, los modelos que presentaron mejor ajuste fueron *Neural Network* y *Supported Vector Machine*. El modelo *Neural Network* presentó MAE y RMSE inferiores de 3.0, aunque R^2 fue menor de 0.8 para los contaminantes NO_x , O_3 y CO. *Supported Vector Machine* presentó MAE y RMSE menores de 3.0 para todos los contaminantes excepto para NO_x (MAE y RMSE entre 4 y 6). Por otra parte, los resultados indicaron subestimación en los valores predichos utilizando *Neural Network*, *Random Forest* y *Kriging Universal*, principalmente para valores altos en los contaminantes CO y O_3 (Figura 4). Por tanto, el modelo *Supported Vector Machine* tuvo las mejores características para modelar e interpolar los contaminantes analizados a nivel mensual, al igual que el promedio histórico.

La modelación geoestadística puede tener limitantes en la selección de parámetros, en particular los métodos de *Machine Learning* supervisado. Por ejemplo, las predicciones de *Random Forest* pueden presentarse más allá del rango de valores observados, en el caso de que algún grupo de datos presente ruido (Espinosa-Zuñiga, 2020). En el caso de *Neural Network*, la principal desventaja es que no se conoce su ecuación final ni los pesos que se emplean en el modelo, convirtiéndose en una caja negra (Sheu, 2020). Por su parte, *Kriging Universal* requiere encontrar un variograma óptimo para el conjunto de datos y así evitar extrapolaciones, además de que es sensible a un número bajo de puntos o una variación alta entre ellos (Shekaramiz et al., 2019). Finalmente, con el modelo *Supported Vector Machine* se deben elegir los parámetros Kernel y C de manera adecuada, ya que estos afectan la complejidad del modelo (Cunha et al., 2022; Liu & Xu, 2014). Pese a estas limitantes, *Supported Vector Machine* presentó el mejor ajuste para todos los contaminantes y, por tanto, se utilizó para representar su distribución espacial.

Dinámica espacial de los contaminantes en la ZMCM

De acuerdo con el análisis espacial se encontraron patrones diferentes de distribución de los contaminantes

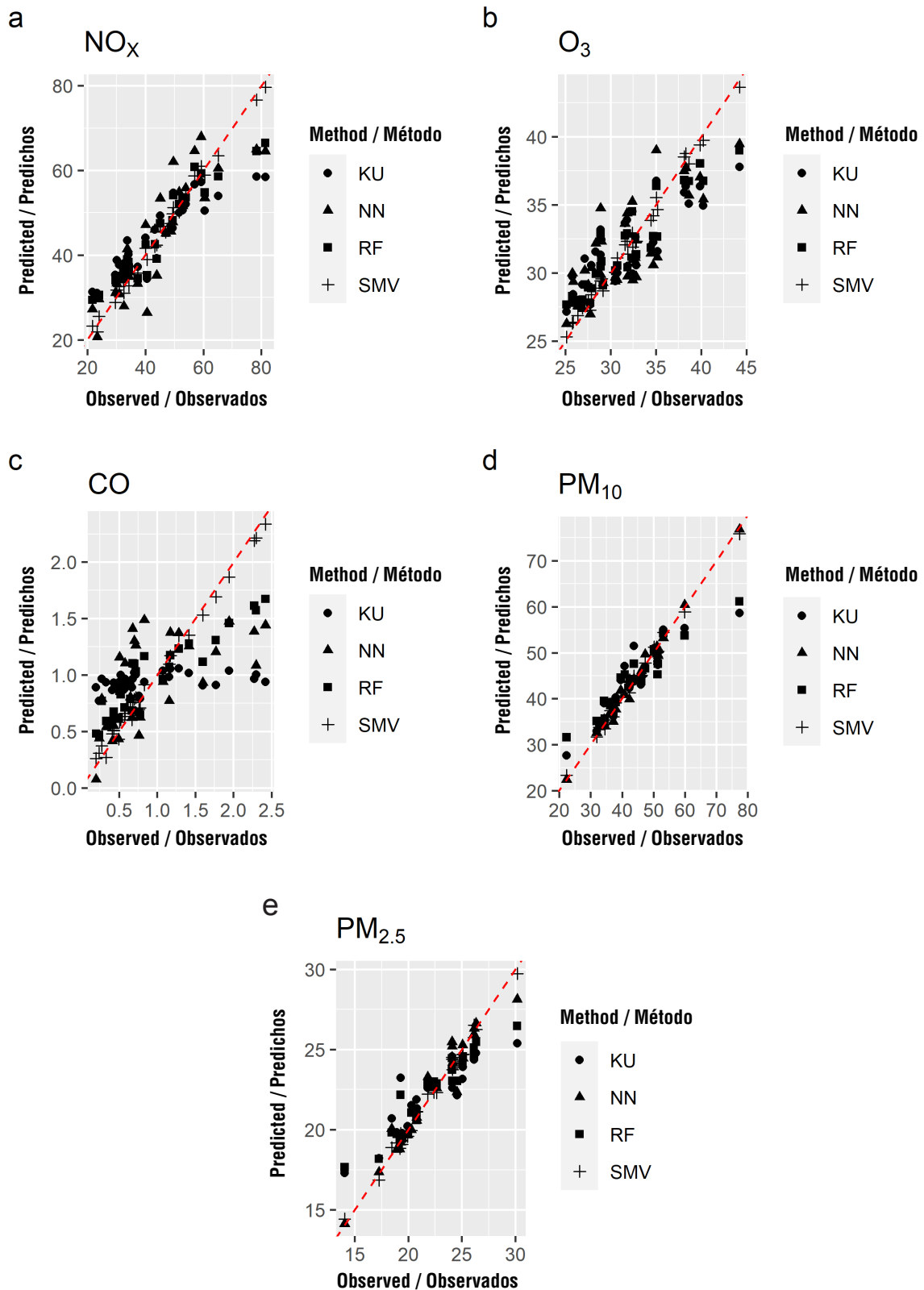


Figure 4. Observed vs. predicted values of data set by pollutants (NO_x: nitrogen oxides, CO: carbon monoxide, O₃: ozone, PM: particulate matter) using Kriging Universal (KU), Neural Network (NN), Random Forest (RF) and Support Vector Machine (SVM) in Mexico City Metropolitan Area. The dotted line indicates a 1:1 ratio.

Figura 4. Valores observados vs. predichos del conjunto de datos por contaminantes (NO_x: óxidos de nitrógeno, CO: monóxido de carbono, O₃: ozono, PM: partículas sólidas) con los modelos *Kriging Universal* (KU), *Neural Network* (NN), *Random Forest* (RF) y *Support Vector Machine* (SVM) en la Zona Metropolitana de la Ciudad de México. La línea punteada indica una relación 1:1.

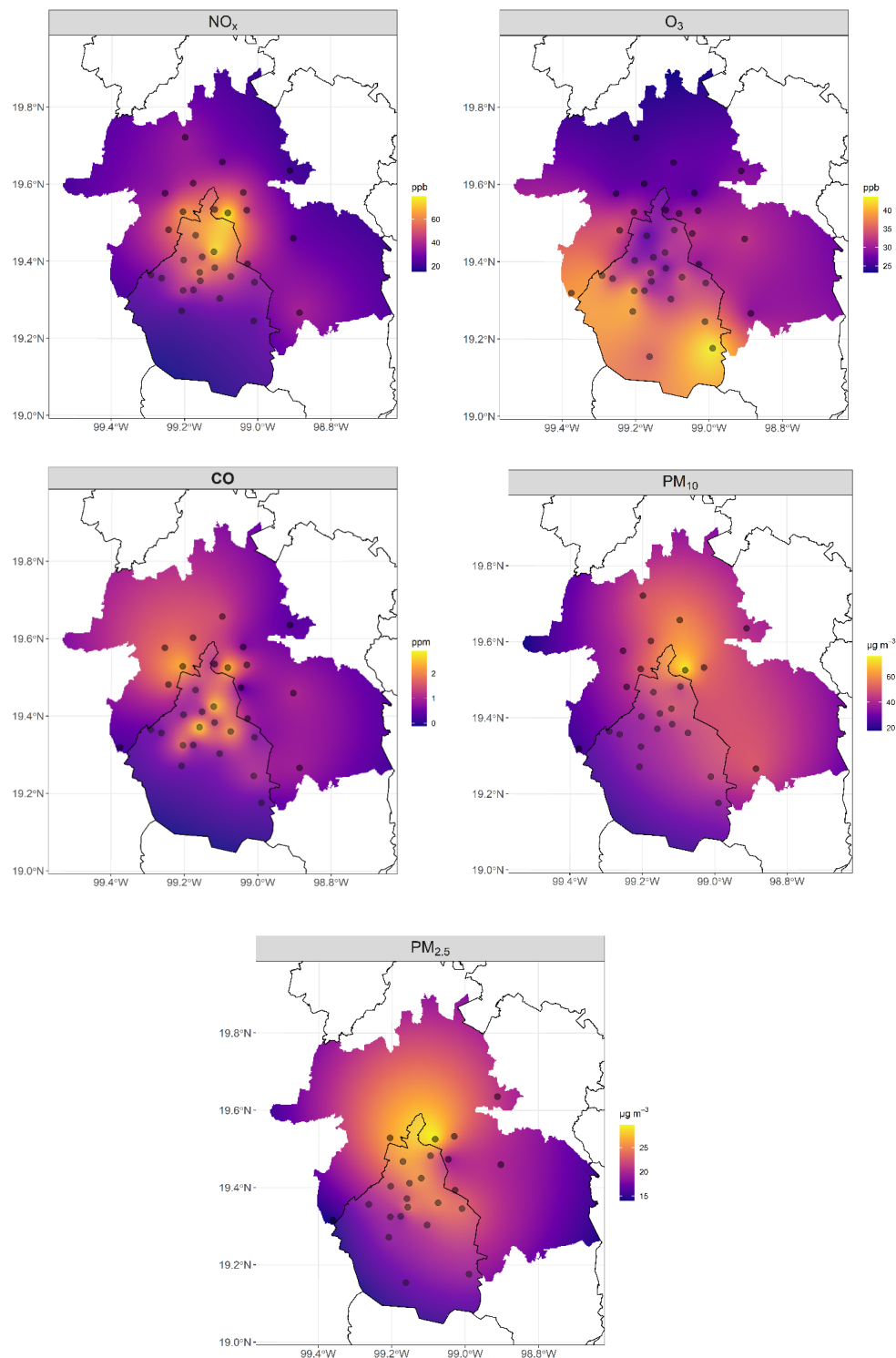


Figure 5. Spatial prediction of the average historical concentration of pollutants with the Machine Learning-Support Vector Machine method: nitrogen oxides (NO_x , 1986-2021), ozone (O_3 , 1986-2021), carbon monoxide (CO, 1986-2021), particulate matter 10 μm (PM_{10} , 2000-2021) and 2.5 μm ($\text{PM}_{2.5}$, 2003-2021) in Mexico City Metropolitan Area. The black circles represent monitoring stations belonging to the Automatic Atmospheric Monitoring Network (AAMN).

Figura 5. Predicción espacial de la concentración histórica de contaminantes con el método *Machine Learning-Support Vector Machine*: óxidos de nitrógeno (NO_x , 1986-2021), ozono (O_3 , 1986-2021), monóxido de carbono (CO, 1986-2021), partículas sólidas 10 μm (PM_{10} , 2000-2021) y 2.5 μm ($\text{PM}_{2.5}$, 2003-2021) en la Zona Metropolitana de la Ciudad de México. Los círculos negros representan estaciones de monitoreo pertenecientes a la Red Automática de Monitoreo Atmosférico (RAMA).

Conclusions

The spatio-temporal dynamics varied among the pollutants analyzed. The highest NO_x and CO concentrations were recorded from November to January, while O_3 concentration decreased in that period and increased from April to May. The lowest particulate matter concentrations occurred from July to October and the highest in May. Regardless of the pollutant, concentrations decreased in recent years; NO_x changes were most noticeable, and the lowest decrease was for CO. Key years such as 1993 or 2020 showed maximum peaks and troughs, mainly linked to the increase/closure of human activities. Although there were no notable differences among the interpolation methods, mainly among those belonging to the Machine Learning methods, the Support Vector Machine method had the best monthly and historical fit for all pollutants.

Acknowledgments

The authors would like to thank the Consejo Nacional de Humanidades, Ciencias y Tecnologías for the scholarship granted to the first author for graduate studies; as well as the Línea de Generación y Aplicación del Conocimiento (LGAC) “Mejoramiento estructural y funcional de los ecosistemas forestales” for partial funding.

End of English version

References / Referencias

- Akritas, M. G., Murphy, S. A., & Lavalley, M. P. (1995). The theil-sen estimator with doubly censored data and applications to astronomy. *Journal of the American Statistical Association*, 90(429), 170–177. <https://doi.org/10.1080/01621459.1995.10476499>
- Ale, T. M. L., Moreno, S. K., & Luque, Z. B. (2020). Perspectiva del COVID-19 sobre la contaminación del aire. *Revista de la Sociedad Científica del Paraguay*, 25(2), 155–182. <https://doi.org/10.32480/rscp.2020.25.2.155>
- Aliyar, Z. B., Shafiei, A. B., Seyedi, N., Rezapour, S., & Moghanjugi, S. M. (2020). Effect of traffic-induced air pollution on seed germination of Arizona Cypress (*Cupressus arizonica* Green) and Black Pine (*Pinus nigra* Arnold). *Urban Forestry & Urban Greening*, 55, 126841. <https://doi.org/10.1016/j.ufug.2020.126841>
- Barrera Huertas, H. A., Torres Jardón, R., Ruíz Suárez, L. G., Santos García Yee, J., Torres Jaramillo, A., Martínez Bolívar, A. P., & García Reynoso, J. A. (2019). Análisis del transporte de ozono en la cuenca atmosférica de Puebla-Tlaxcala en el centro de México. *Revista Internacional de Contaminación Ambiental*, 35(4), 869–888. <https://doi.org/10.20937/rica.2019.35.04.08>
- Beguín, J., Fuglstad, G.-A., Mansuy, N., & Paré, D. (2017). Predicting soil properties in the Canadian boreal forest with limited data: Comparison of spatial and non-spatial statistical approaches. *Geoderma*, 306, 195–205. [10.1016/j.geoderma.2017.06.016](https://doi.org/10.1016/j.geoderma.2017.06.016)

(Figura 5). Por ejemplo, NO_x presentó un patrón circular, donde el área con mayor concentración se ubicó en la zona centro de la Ciudad de México con dos estaciones con valores altos (La Merced y Xalostoc > 60 ppb). Por su parte, O_3 presentó la mayor concentración en las estaciones del Pedregal y Milpa Alta, las cuales se ubican en la parte sur del área de estudio. La concentración de O_3 disminuyó hacia la parte centro y norte de la ZMCM, marcando un gradiente alto en el cuadrante suroeste. Los resultados concuerdan con el estudio de García (2009), quien indica que las fuentes de contaminantes se encuentran vientos abajo, fuera de las fuentes de emisión, por lo que la concentración alta de O_3 se ubica en áreas montañosas al sur de la Ciudad de México (Figura 5).

La mayor concentración de CO se encontró en el cuadrante centro noroeste, debido probablemente a un mayor uso de vehículos, los cuales constituyen la principal fuente de emisiones. No obstante, algunos autores han reportado relación entre altas concentraciones de CO y áreas donde la población tiene mayores ingresos en la Zona Metropolitana del Valle de México (ZMVM), particularmente para las emisiones de CO y dióxido de carbono (CO_2 eq) (Pérez et al., 2018).

Las concentraciones de PM_{10} y $\text{PM}_{2.5}$ fueron más altas en el centro y disminuyeron hacia la parte oeste. La estación con mayor concentración es Xalostoc con valores superiores a $70 \mu\text{g}\cdot\text{m}^{-3}$ de PM_{10} y a $30 \mu\text{g}\cdot\text{m}^{-3}$ de $\text{PM}_{2.5}$. Esto puede estar asociado con lo señalado por Cervantes et al. (2005), quienes afirman que la estación se ubica en una zona con evidencias de erosión, presencia de industrias y caminos sin pavimentar.

Conclusiones

La dinámica espaciotemporal varió entre los contaminantes analizados. Las concentraciones más altas de NO_x y CO se registraron en el periodo noviembre-enero, mientras que la concentración de O_3 disminuyó en dicho periodo e incrementó en abril-mayo. Las concentraciones más bajas de partículas sólidas se presentaron en la época julio-octubre y el valor más alto en mayo. Independientemente del contaminante, las concentraciones disminuyeron en los últimos años; los cambios fueron más notorios en NO_x y la menor reducción se observó en CO. Años clave como 1993 o 2020 mostraron picos máximos y mínimos, ligados principalmente al incremento/cierre de actividades humanas. Aunque no existieron diferencias notables entre los métodos de interpolación, principalmente entre aquellos pertenecientes a los métodos de *Machine Learning*, el método *Support Vector Machine* tuvo el mejor ajuste mensual e histórico para todos los contaminantes.

- Calderón, G. D., Hernández, I. J. L., Castilla, S. L., Hernández, G. E., Barragán, M. G., Rodríguez, P. R. A., & Villegas, O. G. (2000). El ozono como molécula reactiva. Concepto actual. *Perinatología y Reproducción Humana*, 14(2), 115–123. <https://www.medigraphic.com/pdfs/inper/ip-2000/ip002f.pdf>
- Camarillo, R. P., López, A. S., Rosales, L. J. C., & Pérez, V. I. (2014). Análisis de datos de calidad del aire de la Zona Metropolitana del Valle de México mediante técnicas de agrupamiento. *Research in Computing Science*, 72, 137–150. <https://doi.org/10.13053/rscs-72-1-11>
- Castro, M., García, D., & Jiménez, A. (2017). Comparación de técnicas de interpolación espacial de propiedades del suelo en el piedemonte llanero colombiano. *Tecnura*, 21(53), 78–95. <https://doi.org/10.14483/22487638.11658>
- Cervantes, M. G. T., Bracho, L. R., & Bremauntz, A. F. (2005). Las partículas suspendidas en tres grandes ciudades mexicanas. *Gaceta Ecológica*, (74), 15–28. <https://www.redalyc.org/articulo.oa?id=53907402>
- Chen, L., Ren, C., Li, L., Wang, Y., Zhang, B., Wang, Z., & Li, L. (2019). A comparative assessment of geostatistical, machine learning, and hybrid approaches for mapping topsoil organic carbon content. *ISPRS International Journal of Geo-Information*, 8, 174. <https://doi.org/10.3390/ijgi8040174>
- Chow, J., Watson, J., Edgerton, S., & Vega, E. (2002). Chemical composition of PM_{2.5} and PM₁₀ in Mexico City during winter 1997. *The Science of the Total Environment*, 287(3), 177–201. [https://doi.org/10.1016/S0048-9697\(01\)00982-2](https://doi.org/10.1016/S0048-9697(01)00982-2)
- Correa, I. J. de J., Romero, P. J. M., Pérez, R. P., & Vázquez, A. A. (2023). Application of geostatistical models for aridity scenarios in northern Mexico. *Atmósfera*, 37, 233–244. <https://doi.org/10.20937/ATM.53103>
- Cunha, V., Magoni, D., Inácio, P., & Freire, M. (2022). Impact of self C parameter on SVM-based classification of encrypted multimedia peer-to-peer traffic. In L. Barolli, F. Hussain, & T. Enokido (Eds.), *Advanced information networking and applications* (pp. 180–193). Springer. https://doi.org/10.1007/978-3-030-99584-3_16
- Espejel, M. J. (2019). La Zona Metropolitana del Valle de México: arreglos formales y fragmentación. *Economía, Sociedad y Territorio*, 19(60), 241–271. <https://www.redalyc.org/articulo.oa?id=11162787009>
- Espinosa-Zuñiga, J. J. (2020). Aplicación de algoritmos Random Forest y XGBoost en una base de solicitudes de tarjetas de crédito. *Ingeniería, Investigación y Tecnología*, 21(3). <https://doi.org/10.22201/fi.25940732e.2020.21.3.022>
- Espinoza, E. P., & Molina, C. E. (2014). Contaminación del aire exterior Cuenca-Ecuador, 2009-2013. Posibles efectos en la salud. *Revista de la Facultad de Ciencias Médicas de la Universidad de Cuenca*, 32(2), 6–17. <https://publicaciones.ucuenca.edu.ec/ojs/index.php/medicina/article/view/883/781>
- Estarlich, M., Iñiguez, C., Esplugues, A., Mantilla, E., Zurriaga, Ò., Nolasco, A., & Ballester, F. (2013). Variación espacial de la exposición a contaminación atmosférica en la ciudad de Valencia y su relación con un índice de privación. *Gaceta Sanitaria*, 27(2), 143–148. <https://doi.org/10.1016/j.gaceta.2012.05.010>
- García, R. J. A. (2009). Influencia de la meteorología en la calidad de aire en la Zona Metropolitana del Valle de México. *TIP*

Agradecimientos

Al Consejo Nacional de Humanidades, Ciencias y Tecnologías por la beca otorgada a la primera autora para realizar estudios de posgrado; así como a la Línea de Generación y Aplicación del Conocimiento (LGAC) “Mejoramiento estructural y funcional de los ecosistemas forestales” por el financiamiento parcial.

Fin de la versión en español

- Revista Especializada en Ciencias Químico-Biológicas, 12(2), 83–86. <https://doi.org/10.22201/fesz.23958723e.2009.2.27>
- García, E., & Lozano, F. (2007). *Boosting support vector machines*. <https://elkingarcia.github.io/Papers/MLDM07.pdf>
- Guzmán-Morales, J., Morton-Bermea, O., Hernández-Álvarez, E., Rodríguez-Salazar, M. T., García-Arreola, M., & Tapia-Cruz, V. (2011). Assessment of atmospheric metal pollution in the urban area of Mexico City, using *Ficus benjamina* as biomonitor. *Bulletin of Environmental Contamination and Toxicology*, 86, 495–500. <https://doi.org/10.1007/s00128-011-0252-9>
- Instituto Nacional de Estadística y Geografía (INEGI). (2020). *Censo de Población y Vivienda*. <https://www.inegi.org.mx/programas/ccpv/2020/>
- Instituto Nacional de Estadística y Geografía (INEGI). (2022). *Vehículos de motor registrados en circulación*. <https://www.inegi.org.mx/programas/vehiculosmotor/>
- Jenkin, M., & Clemitshaw, K. (2000). Ozone and other secondary photochemical pollutants: Chemical processes governing their formation in the planetary boundary layer. *Atmospheric Environment*, 34, 2499–2527. doi: 10.1016/S1352-2310(99)00478-1
- Liu, Z., & Xu, H. (2014). Kernel parameter selection for support vector machine classification. *Journal of Algorithms & Computational Technology*, 8(2), 163–177. <https://doi.org/10.1260/1748-3018.8.2.163>
- López, F. A., Heres, D., & Marquez, P. F. (2021). Air pollution exposure and COVID-19: A look at mortality in Mexico City using individual-level data. *Science of the Total Environment*, 756, 143929. <https://doi.org/10.1016/j.scitotenv.2020.143929>
- Madrigal, U. D., Hernández, R. J. C., & Morales, M. C. (2004). Comportamiento del monóxido de carbono y el clima en la ciudad de Toluca, de 1995 a 2001. *Ciencia Ergo-Sum, Revista Científica Multidisciplinaria de Prospectiva*, 11(3), 263–274. <https://www.redalyc.org/articulo.oa?id=10411306>
- Mercado, A., Domínguez, L., & Fernández, O. (1995). Contaminación industrial en la zona metropolitana de la Ciudad de México. *Comercio Exterior*, 45(10), 766–774. <http://revistas.bancomext.gob.mx/rce/magazines/325/6/RCE6.pdf>
- Miller, B. G. (2011). 4 - The effect of coal usage on human health and the environment. In B. G. Miller (Ed.), *Clean coal engineering technology* (pp. 85–132). Butterworth-Heinemann. <https://doi.org/10.1016/B978-1-85617-710-8.00004-2>
- Muhammad, I., Shalmani, A., Ali, M., Yang, Q.-H., Ahmad, H., & Li, F. B. (2021). Mechanisms regulating the dynamics of photosynthesis under abiotic stresses. *Frontiers in Plant Science*, 11. <https://doi.org/10.3389/fpls.2020.615942>

- Navarro, A. A. (2019). Control de la contaminación atmosférica en la Zona Metropolitana del Valle de México. *Estudios Demográficos y Urbanos*, 34, 631–663. <https://doi.org/10.24201/edu.v34i3.1806>
- Secretaría de Salud. (2021). *Norma Oficial Mexicana NOM-021-SSA1-2021. Salud ambiental. Criterio para evaluar la calidad del aire ambiente, con respecto al monóxido de carbono (CO) en el aire ambiente, como medida de protección a la salud de la población*. Diario Oficial de la Federación (DOF). https://www.dof.gob.mx/nota_detalle.php?codigo=5634084&fecha=29/10/2021#gsc.tab=0
- Secretaría de Salud. (2021). *Norma Oficial Mexicana NOM-021-SSA1-2021. Salud ambiental. Criterio para evaluar la calidad del aire ambiente, con respecto al ozono (O₃), como medida de protección a la salud de la población*. Diario Oficial de la Federación (DOF). https://www.dof.gob.mx/nota_detalle.php?codigo=5633956&fecha=28/10/2021#gsc.tab=0
- Secretaría de Salud. (2021). *Norma Oficial Mexicana NOM-023-SSA1-2021. Salud ambiental. Criterio para evaluar la calidad del aire ambiente, con respecto al dióxido de nitrógeno (NO₂), como medida de protección a la salud de la población*. Diario Oficial de la Federación (DOF). https://dof.gob.mx/nota_detalle.php?codigo=5633854&fecha=27/10/2021#gsc.tab=0
- Secretaría de Salud. (2021). *Norma Oficial Mexicana NOM-025-SSA1-2021 Salud ambiental. Criterio para evaluar la calidad del aire ambiente, con respecto a las partículas suspendidas PM₁₀ y PM_{2.5} como medida de protección a la salud de la población*. https://dof.gob.mx/nota_detalle.php?codigo=5633855&fecha=27/10/2021#gsc.tab=0
- Organización Mundial de la Salud (OMS). (2005). *Guías de calidad del aire de la OMS relativas al material particulado, el ozono, el dióxido de nitrógeno y el dióxido de azufre*. https://apps.who.int/iris/bitstream/handle/10665/69478/WHO_SDE_PHE_OEH_06_02_spa.pdf;jsessionid=2AC85F7FC5F727C4C6C10E06AE3CF64B?sequence=1
- Pedrero, V., Reynaldos, G. K., Ureta, A. J., & Cortez, P. E. (2021). Generalidades del Machine Learning y su aplicación en la gestión sanitaria en servicios de urgencia. *Revista Médica de Chile*, 149(2), 248–254. <https://doi.org/10.4067/s0034-98872021000200248>
- Pérez, J. I. J., Némiga, X. A., Gaytán, J. F. M., Cedillo, J. G. G., Plata, M. Á. B., Loik, M. E., Sanabria, J. M. C. (2010). Variaciones climáticas en la Zona Metropolitana de la Ciudad de Toluca, Estado de México: 1960-2007. *Ciencia Ergo-Sum, Revista Científica Multidisciplinaria de Prospectiva*, 17(2), 143–153. <https://www.redalyc.org/articulo.oa?id=10413200004>
- Pérez, C. V., Schmelkes, E., López, C. O., Carrera, F., García, T. I. A. P., & Teruel, G. (2018). Ingreso y calidad del aire en ciudades: ¿Existe una curva de Kuznets para las emisiones del transporte en la Zona Metropolitana del Valle de México? *El Trimestre Económico*, 85(340), 745–764. <https://doi.org/10.20430/ete.v85i340.717>
- Pérez, V. Z. R., Ángeles, P. G., Chávez, V. B., Valdez, L. J. R., & Ramírez, G. M. E. (2021). Enfoque espacial para modelación de carbono en el mantillo de bosques bajo manejo forestal maderable. *Madera y Bosques*, 27(1). <https://doi.org/10.21829/myb.2021.2712122>
- Popovicheva, O., Ivanov, A., & Vojtisek, M. (2020). Functional factors of biomass burning contribution to spring aerosol composition in a megacity: Combined FTIR-PCA Analyses. *Atmosphere*, 11(4), 19. <https://doi.org/10.3390/atmos11040319>
- R Development Core Team. (2021). *R: A Language and Environment for Statistical Computing*. The R Foundation for Statistical Computing. <https://www.R-project.org>
- Romieu, I., Meneses, F., Ruiz, S., Sienna, J. J., Huerta, J., White, M. C., & Etzel, R. A. (1996). Effects of air pollution on the respiratory health of asthmatic children living in Mexico City. *American Journal of Respiratory and Critical Care Medicine*, 154, 300–307. <https://doi.org/10.1164/ajrccm.154.2.8756798>
- Sandoval, J., & Jaimes, J. L. (2002). Formación de ozono en la Ciudad de México durante una porción de primavera e invierno en cámaras de esmog exteriores. *Revista de la Sociedad Química de México*, 46(2), 180–184. <https://www.redalyc.org/articulo.oa?id=47546217>
- Secretaría del Medio Ambiente (SEDEMA). (2018). *Inventarios de emisiones de la ZMVM*. <https://www.sedema.cdmx.gob.mx/storage/app/media/DGCA/InventarioDeEmisionesZMVM2018.pdf>
- Sheinbaum, P. C. (2016). Contaminación atmosférica en la Zona Metropolitana del Valle de México. *Revista Ciencia*, 67, 70–77. https://www.revistaciencia.amc.edu.mx/images/revista/67_3/PDF/Contaminacion.pdf
- Shekaramiz, M., Moon, T. K., & Gunther, J. (2019). A note on Kriging and Gaussian processes. <https://doi.org/10.13140/RG.2.2.36631.83367>
- Sheu, Y.-h. (2020). Illuminating the Black Box: Interpreting deep neural network models for psychiatric research. *Frontiers in Psychiatry*, 11. <https://doi.org/10.3389/fpsy.2020.551299>
- Villalobos, V. M. C. (2006). *Inventario de emisiones de la ZMVM, 2006*. <http://186.96.0.232/aire/default.php?opc=Z6BhnmI=&dc=Zg==>
- Vinasco, S., Pablo, J., Nastar C., & Nastar del Rio, T. C. (2013). Variación espacial y temporal de concentraciones de PM₁₀ en el área urbana de Santiago de Cali, Colombia. *Ingeniería de Recursos Naturales y del Ambiente* (12), 129–141. <https://www.redalyc.org/pdf/2311/231130851011.pdf>
- Wedyan, G. N., Osama T. Al., & Zainab M. A. (2020). The Influence of solar radiation on ozone column weight over Baghdad City. *IOP Conference Series: Materials Science and Engineering*, 928, 072089. <https://doi.org/10.1088/1757-899X/928/7/072089>
- Xing, Y.-F., Xu, Y.-H., Shi, M.-H., & Lian, Y.-X. (2016). The impact of PM_{2.5} on the human respiratory system. *Journal of Thoracic Disease*, 8(1), E69–E74. <https://doi.org/10.3978/j.issn.2072-1439.2016.01.19>
- Yuan, Q., Shen, H., Li, T., Li, Z., Li, S., Jiang, Y., Xu, H., Tan, W., Yang, Q., & Wang, J. (2020). Deep learning in environmental remote sensing: Achievements and challenges. *Remote Sensing Environment*, 241, 111716. <https://doi.org/10.1016/j.rse.2020.111716>
- Zhou, Y., Yue, Y., Bai, Y., & Zhang, L. (2020). Effects of rainfall on PM_{2.5} and PM₁₀ in the middle reaches of the Yangtze River. *Advances in Meteorology*, 2398146. <https://doi.org/10.1155/2020/2398146>