# Satellite data geoprocessing to estimate PM<sub>2.5</sub> over the Megalopolis of Central Mexico

Marco Antonio MORA-RAMÍREZ<sup>1</sup>, Edgar MARTÍNEZ-LUNA<sup>1</sup> and Xochitl CRUZ-NÚÑEZ<sup>2\*</sup>

<sup>1</sup> Departamento de Fisicomatemáticas, Facultad de Ciencias Químicas, Benemérita Universidad Autónoma de Puebla, Puebla Pue. Av. 14 Sur Col. Jardines de San Manuel, Ciudad Universitaria, Puebla, Puebla, México. P.O. Box 1067, C.P. 72001 México.

<sup>2</sup> Instituto de Ciencias de la Atmósfera y Cambio Climático, Universidad Nacional Autónoma de México, Circuito de la Investigación Científica s/n, Ciudad Universitaria, 04510 Ciudad de México, México.

\*Corresponding author: xochitl.cruz@atmosfera.unam.mx

Received: 2022 September 23 Accepted: 2023 March 14

#### RESUMEN

En la Megalópolis del centro de México frecuentemente se registran niveles de partículas suspendidas finas (PM<sub>2.5</sub>) superiores a la Norma Oficial Mexicana, causando diversas enfermedades respiratorias, desde síntomas agudos hasta asma y cáncer de pulmón. Por ello, la medición de PM<sub>2.5</sub> es fundamental como primer paso para advertir a la población de los niveles de riesgo por exposición a las partículas. Desafortunadamente, la región central de México no cuenta con suficientes sitios de monitoreo, limitando la disponibilidad de datos. Este estudio aborda este problema al usar datos satelitales para desarrollar un modelo de regresión lineal múltiple. Nuestro modelo usa el espesor óptico de aerosoles (AOD por sus siglas en inglés), la húmedad relativa (RH), la temperatura (T), la capa límite planetaria (PBLH), y el índice de Vegetación de Diferencia Normalizada (NDVI) como variantes independientes para estimar las concentraciones de PM<sub>2.5</sub> en la región de estudio. Se encontró que la relación entre el AOD y las concentraciones de PM<sub>2.5</sub> estaba fuertemente influenciada por la RH y la T. Sin embargo, este efecto se compensa con una baja PBLH (< 400 m) que permite que las mediciones de AOD y las concentraciones de PM<sub>2.5</sub> con datos satelitales. Este estudio podría ayudar a mejorar el monitoreo de la calidad del aire en la Megalópolis del Centro de México al proveer mayor información temporal y espacial sobre la concentración de partículas en la atmósfera.

#### ABSTRACT

The Megalopolis of Central Mexico experiences high levels above the Official Mexican Standard (NOM) of  $PM_{2.5}$ , leading to various respiratory diseases ranging from acute symptoms to chronic illnesses such as asthma and lung cancer. It is crucial to measure  $PM_{2.5}$  levels accurately to warn the public about the risks of exposure to particulate matter. Unfortunately, the Megalopolis of Central Mexico has a shortage of monitoring sites, limiting data availability. This study addresses this issue using satellite data to develop a multiple linear regression model. Our model uses aerosol optical depth (AOD), relative humidity (RH), temperature (T), the planetary boundary layer height (PBLH), and the normalized difference vegetation index (NDVI) as independent variables to estimate  $PM_{2.5}$  concentrations in the region under study. The relationship between AOD and  $PM_{2.5}$  concentrations was found to be strongly influenced by RH and T. However, this effect is compensated for by a low PBLH (< 400 m), which enables AOD and  $PM_{2.5}$  concentrations using satellite data. This study could help improve air quality monitoring in the Megalopolis of Central Mexico by providing more spatial and temporal data on particle concentrations in the atmosphere.

# Keywords: AOD, PM<sub>2.5</sub>, NDVI, Mexico.

#### 1. Introduction

Air pollution is a significant global public health issue, responsible of 6.67 million deaths in 2019 (HEI, 2022), particularly in urban areas where people are exposed to harmful pollutants (Liang and Gong, 2020). Fine particulate matter, with an aerodynamic diameter of less than 2.5  $\mu$ m (PM<sub>2.5</sub>), is identified as the primary pollutant causing numerous health problems, including lung cancer (Cohen and Pope, 1995), heart attacks (Lee et al., 2014), asthma (Tiotiu et al., 2020), bronchitis (Zhang and Zhou, 2021), and more recently linked to a higher intensity of SARS-CoV-2 infections (Czwojdzińska et al., 2021). Therefore, it is crucial to have complete and accurate systematic monitoring of ambient PM<sub>2.5</sub> concentrations.

The Megalopolis of Central Mexico (MCM), with approximately 32 million inhabitants, comprises five states surrounding Mexico City. Each state in the MCM has its ground-based monitoring network. However, only Mexico City has provided groundbased measurements systematically over the past 15 years (Aldape and Flores, 2011; RAMA, 1996). Nevertheless, even with a more robust monitoring station network in the MCM, the number of monitoring sites may be insufficient to obtain extended coverage of aerosol distribution, including its sources and sinks, as indicated by the Ministry of Environmental Sustainability and Territorial Planning of the State of Puebla (SSAOT, 2012). In this context, satellite-retrieved data have proven to be a valuable complement to ground-based monitoring networks (Ebell et al., 2013).

The aerosol optical depth (AOD) is defined as the integral of aerosols' extinction coefficient in the vertical (Gao et al., 2021). It is an excellent proxy to characterize the degree of turbidity in the atmosphere and thus directly related to the amount of aerosol particles in the atmosphere (Kong et al., 2016). AOD can be obtained through remote sensing from the surface (Holben et al., 1998) or satellites, which provide comprehensive spatial and temporal coverage to generate information beyond the domain of in-situ surface monitoring stations.

In several instances, satellite-retrieved AOD has been correlated with  $PM_{2.5}$  concentrations on the ground with varying levels of precision (Kim et al., 2016; Kong et al., 2016; Li et al., 2015; Schaap et al., 2009; van Donkelaar et al., 2006). Additional parameters such as temperature, relative humidity, wind speed, and other pollutants have been included in other studies to improve the quality of the calculated PM<sub>2.5</sub> data (Xu and Zhang, 2020; Zhao et al., 2018), indicating that models can compensate for the PM<sub>2.5</sub> space-time gaps left by monitoring stations and enhance the predictive power (Al-Saadi et al., 2005). These studies present advantages and disadvantages related to the meteorological conditions, chemical composition, and vertical distribution of aerosols considered when deriving PM<sub>2.5</sub> concentrations from satellite-retrieved AOD (Zheng et al., 2017).

In addition to the mentioned methods and features, AOD measurements have complications; satellite measurements of AOD strongly depend on the cloudiness factor (Clark, 1983). Furthermore, the availability of satellite data may be compromised by overpass times in some cases (Bojanowski et al., 2014). AOD satellite data for this application is commonly obtained from the Moderate Resolution Imaging Spectroradiometer (MODIS), an instrument onboard Aqua and Terra satellites (Ghotbi et al., 2016). The retrieved data are processed with specific algorithms based on the surface reflectance properties of the study area (Hsu et al., 2004; Levy et al., 2013).

This study utilizes daily AOD observations from MODIS instruments on Aqua and Terra, processed with three different algorithms (Dark Target, Deep Blue, and combined DT and DB), in combination with meteorological parameters to generate a multiple linear regression model for estimating  $PM_{2.5}$ concentrations in the MCC. Various atmospheric parameter correlations are analyzed to build insights into a statistical model for calculating PM25 concentrations from satellite data. The study found that the correlation between AOD and PM<sub>2.5</sub> strongly depends on relative humidity (RH), planetary boundary layer height (PBLH), and temperature. Interestingly, the results also reveal that correlations vary spatially due to gradients in terrain elevation, which is a novel finding for the study region.

#### 2. Data and Methods

A diverse set of data for 2012 is used in this paper collected from various sources, including groundbased monitoring stations, reanalysis models, remote sensors on the surface, and onboard satellites for the area of study shown in Figure 1. The central region of Mexico (-100°W to -97°W and 18°N to 20°N), enclosed by the blue rectangle in the inset over which satellite remote sensing data are examined, is made up of six states: Mexico City, State of Mexico, Morelos, Tlaxcala, Hidalgo, and Puebla. The red areas in Figure 1 represent the urban settlements. The center of the MCM is highly populated, subject to intense anthropogenic activity, and linked to industrial activities in the north. Generally, biogenic emissions occur over the west and south associated with wildfires and emissions from Popocatepetl Volcano, respectively (Mora et al., 2017). The data sets are summarized in the subsequent subsection.

#### 2.1 Data validation

#### 2.1.1 Remotely sensed data

Satellite data from the GES-DISC (Goddard Earth Sciences Data and Information Services Center) Interactive Online Visualization ANd aNalysis Infrastructure website (GIOVANNI, 2023) were used to obtain aerosol optical depth (AOD) and other parameters. To ensure data quality, any data sets with a cloud fraction (CF) greater than 70% were excluded (Myhre et al., 2007). Table I overviews the data sets' parameters, instruments, and spatial and temporal resolutions. AOD daily data were retrieved from both MODIS Aqua (MY) and MODIS Terra (MO) satellites (Sayer et al., 2014) using three algorithms [Deep Blue (DB), Dark Target (DT), and Combined Dark Target and Deep Blue (DTDB)] designed to capture specific features applicable in the study region's spatial domain. Ground-based remotely sensed AOD measurements from the Aerosol Robotic Network (AERONET; Holben et al., 1998) were used to select the appropriate satellite AOD product. The AOD-AERONET level 3.0 product (AERONET 2021; Smirnov et al., 2000) located at the Universidad Nacional Autonomá de México (National Autonomous University of Mexico) (19.3°N, -99.18°W) was used.

The AOD-AERONET measurements were taken at 340 nm, 380 nm, 440 nm, 500 nm, 675 nm, 870 nm, and 1020 nm, and interpolation over the wavelength was performed to obtain the corresponding aerosol band AOD-AERONET at 550 nm, the same as AOD from MODIS. Several studies have shown good



Fig. 1. The blue rectangle in the inset encloses the region  $(-100^{\circ}\text{W to }-97^{\circ}\text{W}, 18^{\circ}\text{N to }20^{\circ}\text{N})$  with six states of the Megalopolis of Central Mexico. The red-colored areas correspond to populated urban centers, and the yellow markers represent the ground-based monitoring stations in the Megalopolis with a sufficiency of data (>70%).

Variable	Product name	Satellite/Instrument	Time res.	Spatial res.
AOD-DT	MYD08 D3 v6	MODIS-Aqua	D	$1^{\circ} \times 1^{\circ}$
AOD-DT	MOD08 D3 v6.1	MODIS-Terra	D	$1^{\circ} \times 1^{\circ}$
AOD-DB	MOD08 D3 v6.1	MODIS-Terra	D	$1^{\circ} \times 1^{\circ}$
AOD-DB	MYD08 D3 v6.1	MODIS-Aqua	D	$1^{\circ} \times 1^{\circ}$
AOD-DTDB	MOD08 D3 v6.1	MODIS-Terra	D	$1^{\circ} \times 1^{\circ}$
AOD-DTDB	MYD08 D3 v6.1	MODIS-Aqua	D	$1^{\circ} \times 1^{\circ}$
A-AOD		AERONET	D	columnar
CF	MOD08 D3 v7	MODIS-Terra	D	$1^{\circ} \times 1^{\circ}$
PBLH (m)	M2TMNXFLX	MERRA-2 Model	М	$0.5^{\circ} \times 0.625^{\circ}$
NDVI	MOD13C2 v006	MODIS-Terra	М	0.05 °
RH (%)	AIRS3STD v006	AIRS	D	$1^{\circ} \times 1^{\circ}$
T (°C)	AIRS3STD v006	AIRS	D	$1^{\circ} \times 1^{\circ}$
SatPM <sub>2.5</sub>	M2TMNXAER v5.12.4	MERRA-2 Model	М	$0.5^{o} \times 0.625^{o}$

Table I. Summary of the complete Data Set Used for this study in the spatial domain (-100° to -97°W, 18° to 20°N), where D and M mean daily and monthly resolutions, respectively.

agreement between satellite and land-retrieved AOD measurements (Bhaskaran et al., 2011; Tripathi et al., 2005). However, Tripathi et al. (2005) found that MODIS overestimates AOD during the Monsoon period, characterized by seasonal wind produced by the displacement of the equatorial belt, resulting in dust transport from other regions. In contrast, MODIS underestimates the AOD observed by AERONET during cold season.

Statistical analysis was performed to examine the data variance and frequency distribution and to determine whether the climatic seasons of the year for the study region influence AOD satellite measurements with MODIS. We assumed that the planetary boundary layer height (PBLH) is a determinant parameter to explain differences between MODIS and AERONET AOD retrieved. Due to the study region having different elevations, the correlation of aerosol particles with AOD varies spatially. PBLH and a proxy of PM2.5 mass concentration were retrieved from the Modern-Era Retrospective analysis for Research and Applications, version 2 (MERRA-2) model (Gelaro et al., 2017). The proxy, SatPM<sub>2.5</sub>, includes the surface mass concentrations of black carbon, organic carbon, sulfate ion, dust, and sea salt with diameters less than 2.5 µm for calculating SatPM<sub>2.5</sub> (Buchard et al., 2016).

## 2.1.2 Data integration

After selecting the appropriate satellite AOD product (SatAOD) using the process described in the previous subsection, it is necessary to synchronize the data sets of SatAOD, NDVI, PBLH, T, RH, and SatPM2. These data sets come from different instruments and satellites and have varying spatial resolutions. Therefore, a spatial synchronization process is required to compare and use them in a multiple linear regression model. This is achieved through the Inverse Distance Weighting interpolation (IDW) technique using R-code (Gimond, 2022) to interpolate and refine the grids in a mesh (Pebesma, 2004). It should be noted that this technique estimates the new points by taking the average of the nearest neighbors, and contamination at the edge of the spatial domain may occur. Therefore, to avoid this issue, a slightly larger spatial domain (at least two grid cells) than the one shown in the inset in Figure 1 is considered.

The refined raster layers obtained (5.5 km  $\times$  5.5 km) are clipped according to the extent shown in Figure 1, and verification is done to ensure that the data distribution has not been altered. It is also essential to consider the scale, as previous research has shown that the scale is crucial in the bias of regression model outputs (Paciorek, 2010). Additionally, studies have reported that the correlation between AOD and PM<sub>2.5</sub> decreases when the AOD resolution is reduced (Chudnovsky et al., 2013). Finally, the monthly mean average of variables is calculated to complete the data integration process.

## 2.2 Data correlation

To ensure that the model has desirable features such as low or no correlation among independent variables but significant influence on the dependent variable  $(PM_{2.5})$ , it is crucial to investigate the possible correlation between the independent and dependent variables before using variable data to build the model. This investigation can be carried out using the Raster package in R (Raster Package, 2023).

Local correlation is recorded for the central grid, and the operation is repeated throughout the entire domain in the raster. To use the function in two rasters simultaneously, an intermediate step involves defining a third raster that records the positions of the values of each grid, listing them from 1 to the total number of grids. The focal function then extracts the positions for which values from the two rasters are obtained.

This technique helps to investigate the spatial correlation between the independent variables (NDVI, T, RH, PBLH, AOD) and the dependent variable (PM<sub>2.5</sub>) on a monthly or quarterly season basis (MAM, JJA, SON, and DJF). It allows for identifying differences in areas of the study region due to various factors such as human activity, meteorology, soil types, and characteristics of each sub-region.

Finally, the correlation between the two variables in each grid is shown on a map, highlighting regions where the correlation between the variables is negative (-1), null, or positive (+1). This map can help to refine the model further, ensuring that it meets the desirable features required for accurate predictions of PM<sub>2.5</sub> levels.

#### 2.3 A Model to estimate $PM_{2.5}$ concentration

In summary, the study incorporates six AOD products derived from Aqua and Terra satellites, processed with three different algorithms, to find the most suitable AOD variable that explains Sat-PM<sub>2.5</sub> concentrations. The data are integrated and synchronized spatially and temporally before using them in a linear regression model that includes satellite-retrieved variables (SatAOD, NDVI, PBLH, RH, and T) in each grid and month of the simulation domain. Before model construction, spatial and temporal correlation analyses are conducted among the variables. The proposed linear regression model is expressed as follows:

$$TeoPM_{2.5_{s,t}} = \alpha_{0,t} + \alpha_{1,t}SatAOD_{st} + \alpha_{2,t}NDVI_{st} + \alpha_{3,t}PBLH_{st} + \alpha_{4,t}RH_{st} + \alpha_{5,t}T_{st}$$
(1)

The multiple linear regression model's parameters  $(\alpha 0, \alpha 1, \dots, \alpha 5)$  are adjusted to obtain the best fit with the observed data at monitoring sites ( $TeoPM_{2.5}$ ). The process begins by considering all variables as predictors and selecting the best ones using the Akaike Information Criterion (AIC). AIC is a metric used to compare the fit of multiple regression models, with a smaller AIC indicating a better fit (Sakamoto et al., 1986). The AIC penalizes models with excessive independent variables based on their ability to explain the dependent variable. It yields the lowest score for a model with minimal loss of information or the highest predictive power while minimizing the number of predictor variables. Various R-Studio libraries (stats v3.6.2, olsrr 0.5.3) are utilized to perform the analysis. The "confinity" function in R (stats v3.6.2, n.d.) estimates the confidence intervals for the parameters in the fitted regression model.

#### 2.3.1 PM<sub>2.5</sub> Ground-based measurements

To validate the performance of the regression model, ground-based  $PM_{2.5}$  concentrations for the study region were obtained from multiple sources, including the National Air Quality Information System (SINAICA, 2021) and the Mexico City's automatic air quality monitoring network (Red Automática de Monitoreo Atmosférico, or RAMA) (RAMA-CDMX, 2021; RAMA-EdoMex, 2021). 13 monitoring stations with a sufficiency rate > 70% located in Mexico City and the State of Mexico were selected. Hourly  $PM_{2.5}$  concentration data from these stations were carefully verified, and suspected outliers were removed. These ground-based  $PM_{2.5}$  data were then compared with the  $PM_{2.5}$  values estimated by the regression model (TeoPM<sub>2.5</sub>) to assess the model's efficiency.

#### 3. Results

#### 3.1 Satellite AOD validation

The correlation analysis between the six AOD products and SatPM<sub>2.5</sub> revealed a high R<sup>2</sup> value ranging between 0.73 and 0.76, except for AOD from MODIS Aqua DTDB, which had an R<sup>2</sup> of 0.59. Among the six products, AOD from MODIS Terra DT showed the highest correlation (R<sup>2</sup>=0.76), indicating its suitability for the linear regression model. In addition, the AOD grids (3×3) generated by each algorithm and platform were compared with the AERONET network's point data to gain further insights into AOD selection. Figure 2 shows the monthly correlation results of AOD from MODIS-Terra (MO) and MODIS-Aqua (MY) for each algorithm: DT, DB, DTDB, and AOD from AERONET.

Although the Aqua and Terra algorithms used for processing satellite data are similar, there is a better correlation coefficient with Aqua than with Terra. The dry-hot season, from March to May, is characterized by higher AOD values, especially in May, as seen in Figure 3.

In May, the AOD data from MODIS Terra and Aqua satellites, as well as the rescaled (0.25) AOD-AERONET data, showed unexpectedly high values compared to the expected trend. This indicates that the satellite data may capture particles in the upper atmosphere that cannot be detected with the same precision by AERONET measurements from the surface. The DB product used for processing satellite data does not differentiate between fine and coarse mode aerosols, which could explain the lack of differentiation between dry-hot season (MAM) and cold quarter (DJF). However, the presence of coarse particles from wildfires during the dry-hot season (MAM) suggests otherwise. In October, there were discrepancies between the MODIS and AERONET data trends, with (0.25) AOD-AERONET slightly below the expected trend



Fig. 2. Monthly AOD correlation between AERONET and MODIS Aqua (MY) and MODIS Terra (MO) products (DB, DT, and DTDB) for the 2012 year.



Fig. 3. Comparison of AERONET monthly AOD averages and MODIS Terra and Aqua (DB) monthly and spatial (0.25) AOD averages in 2012. The red-dotted line represents the monthly mean trend.

and AOD MODIS data above it. This could be attributed to the challenges of remote sensing measurements during the rainy season (June-August).

To refine the grids of the mesh, various interpolation algorithms were applied using the "gstat" package in R (Gimond, 2022). The Inverse Distance Weighting (IDW) method was chosen to estimate new points, considering the nearest neighbors' average value (nmax = 5). This method assigns greater weight to the closest points, thus providing a more precise estimate. Figure 4 illustrates the data integration step, displaying the  $PM_{2.5}$  data obtained from remote sensing (Sat- $PM_{2.5}$ ) before and after the mesh refinement for January 2012. The gray-colored plots reveal that the spatial distribution remained unchanged after synchronization. To avoid data contamination at the region's edge, the refined mesh (Figure 4b) was cropped according to the spatial domain extent, which was larger than the region of interest in this study.

#### 3.2 Correlation among variables

To ensure accurate predictions, a linear regression model must have independent variables that are not highly correlated with each other but rather have some correlation with the dependent variable  $(PM_{2.5})$ . Figure 5 shows the correlation among satellite-retrieved variables during the March-May 2012 quarter. During the dry-hot season, from March to May, the dependence of AOD on temperature increases, with values of +0.4, +0.5, and +0.7, respectively. Poor green vegetation profiles and wildfires, including those set for agricultural purposes in the Megalopolis of Central Mexico, can contribute to the resuspension of particles and the resulting increase in AOD. Likewise, during the dryhot season, the PBLH has a negative correlation with PM<sub>2.5</sub> that decreases from -0.4 (march), -0.3 (april) to -0.2 (may). This means as temperature increases, the height of PBLH also increases, leading to a



Fig. 4. Example of Sat-PM<sub>2.5</sub> data interpolation and distribution of values (a) before  $(0.5^{\circ} \times 0.625^{\circ})$  and (b) after mesh refinement (5.5 km × 5.5 km). The gray areas correspond to the data distribution.



Fig. 5. Correlation of satellite-retrieved variables for dry-hot season months (March, April, and May) in 2012 in the Megalopolis of Central Mexico.

decrease in the correlation between PBLH and  $PM_{2.5}$ . Similarly, the correlation between PBLH and AOD during the dry-hot season (MAM) ranges between -0.4 to -0.6, suggesting that aerosols' cooling effects in the lower PBLH can suppress the development of PBLH, as reported by Zhang et al. (2022).

Figures 6 and 7 illustrate the monthly spatial correlation between Sat-PM<sub>2.5</sub> and AOD and Sat-PM<sub>2.5</sub> and PBLH5, respectively, within the study region outlined by a black rectangle. The warm colors (red) indicate a positive correlation (+1), while the cool colors (blue) indicate a negative correlation (-1). During the dry-cold season Sat-PM2.5 and AOD, from Figure 6, exhibit a high positive correlation in the east, north, and northeast regions of the spatial domain. But, the center, south, and southwest areas present a high negative correlation. The difference in the correlation between these two regions is likely explained by the significant elevation gradient, which implies substantial differences in vegetation, soil type, and meteorology.

The spatial correlation of Sat-PM2.5 and PBLH presented in Figure 7, shows negative values during the cold season (DJF), coinciding with the lowest values of the PBLH and high episodes of particle matter, which are observed systematically in México City during the winter. In the following months, the negative spatial correlation is observed up to March, but in April and May, the spatial correlation changes to positive, which certainly means that if PBLH increases, then PM2.5 also does. This outcome underscores the significance of emissions transport from areas outside the megalopolis. It helps to clarify the persistently high ozone levels or particles observed systematically despite the local mitigation strategies employed in Mexico City during the dry hot season (MAM).

# 3.3 Ground-based PM<sub>2.5</sub> Analysis

Figure 8 presents the boxplots of annual PM<sub>2.5</sub> concentrations measured at selected monitoring sites in Mexico City during 2012. The boxplot displays the



Fig. 6. Monthly mean spatial correlation between Sat-PM<sub>2.5</sub> and AOD in the study region (-100° to -97°W, 18° to 20°N) enclosed by the black rectangle.



Fig. 7. Monthly mean spatial correlation between Sat-PM<sub>2.5</sub> and PBLH in the study region (-100° to -97°W, 18° to 20°N) enclosed by the black rectangle.



Fig. 8. Annual  $PM_{2.5}$  concentrations on ground-based monitoring sites in Mexico City. The red circles represent the extreme values (suspected outliers).

median, first, and third quartile values, while the suspected outliers are indicated with red circles. These outliers are data points that lie beyond 1.5 times the interquartile range (IQR) from the upper and lower quartiles. After removing the outliers, the lower and upper fences of the box represent the minimum and maximum values, respectively.

Figure 9 shows the hourly, daily, and monthly average concentrations of PM<sub>2.5</sub> in Mexico City. Figure 9a shows that the hourly analysis by day of the week reveals two peaks, one before noon (11 a.m.) and another in the late afternoon (7 p.m.), consistent throughout the week. However, two monitoring sites in the northern part of the city, XAL and SAG, exhibit earlier peak concentrations (around 40 minutes earlier) than other sites, with XAL having the highest PM<sub>2.5</sub> levels. Figure 9b (lower-left corner) summarizes the hourly averages by day of the week and shows two distinct peaks in PM<sub>2.5</sub> concentrations throughout the day. Figure 9d (lower-right corner) shows the averages by day of the week, demonstrating that PM<sub>2.5</sub> accumulates during the week, with the highest levels observed on Friday or Saturday. After reductions in emissions from vehicles and industries during the weekend, the lowest levels are observed on Monday. The monthly averages in Figure 9c (lower-middle part) exhibit two peaks, one in May and another during

the cold season (DEF). The latter corresponds to the maximum  $PM_{2.5}$  levels Fontes et al. (2017) observed, as thermal inversion traps pollutants. For a detailed discussion of the observed peaks and the main sources of  $PM_{2.5}$ , see Mora et al. (2017).

Figure 10 highlights specific characteristics of the quarterly hourly averages of PM<sub>2.5</sub> data. The highest concentrations of particles are observed during the dry-hot season (March to May) and the cold quarter (DEF). Precipitation during the rainy season (June-August) contributes to removing PM<sub>2.5</sub>. However, two peaks are still observed yearly in  $PM_{2.5}$  levels despite the removal effect. The first peak occurs during the dry-hot season (MAM) when intense wildfires and maximum PBLH (as shown in Figure 11b) facilitate the exchange of air masses from other territories. In contrast, another peak is observed during the cold season (DEF), and the height of the PBLH reaches a minimum, indicating that the study area is relatively isolated from the neighboring regions. This isolation results in lower dilution of pollutants and an increase in  $PM_{2.5}$ .

In Figure 11, monthly means of remotely sensed data, including PBLH, RH, T, AOD, and NDVI, are presented with the corresponding ground-based PM<sub>2.5</sub> measurments. Notably, PM<sub>2.5</sub> concentrations from MERRA-2 and ground-based sites exhibit the same qualitative trend. However, PM<sub>2.5</sub> levels from



23

9

2

ശ

0

20-10соу

CAM

q

50-

40-

ର bWs



12 hour

ശ

0

20-

23

18

12

9

0

23

40

12

9

0

23

9

4

9

0

Monday

(a)

60-50PM₂₅ 30-30-



Fig. 10. Hourly averages by quarter (March-May, June-August, September-November, December-February) of PM<sub>2.5</sub> concentrations at monitoring sites in Mexico City in 2012.



Fig. 11. Monthly mean of (a) PM<sub>2.5</sub> concentrations from ground-based monitoring sites in Mexico City and remotely sensed (b) PBLH, (c) RH, (d) NDVI, (e) Temperature, and (f) AOD for 2012 in the spatial domain of study.

ground-based sites show a different trend than AOD during the cold season (DJF).

#### 3.4 Linear regression model fit

Table II presents the monthly correlation coefficients for each variable from the multiple linear regression model (TEO  $PM_{2.5}$ ). The alpha parameters have standard errors ranging from 1e-13 to 1e-11 and p-values less than 1e-16, indicating high statistical significance.

Using the multiple linear regression model coefficients, PM<sub>2.5</sub> levels were estimated for each grid in the study area. The model, which incorporated all variables as predictors, exhibited a high R<sup>2</sup> value of 0.7501, indicating that it can account for 75.01% of the observed variability in life expectancy. The p-value of the model was significant (3.787e-10), further supporting its validity. The spatial distribution of calculated PM<sub>2.5</sub> concentrations is displayed in Figure 12. Monthly trends in the estimated  $PM_{2.5}$ values show good agreement with those obtained from the MERRA-2 model. Comparison of the estimated monthly PM2.5 concentrations with corresponding monitoring sites in Mexico City resulted in a 0.6-0.8 RMSE and a standard deviation of 0.01-0.5 µg.

The most significant discrepancies between estimated and observed  $PM_{2.5}$  concentrations occurred during the rainy season (June to August) when aerosol retrievals from satellites were challenging. Conversely, the dry-hot season (MAM) provided the most favorable conditions for estimating  $PM_{2.5}$  concentrations with the model. Estimated  $PM_{2.5}$  values in Puebla exhibited lower confidence (0.4-0.6 RMSE) than those in Mexico City.

Figure 13 depicts the monthly  $PM_{2.5}$  concentrations calculated for each grid cell (TEO  $PM_{2.5}$ ) compared to the corresponding ground-based monitoring sites, Xalostoc (XAL) in Mexico City and Ninfas (NIN) in Puebla, for the base year 2012. In general, the modeled  $PM_{2.5}$  concentrations show a similar trend to the ground-based monitoring sites in both cities, with peaks occurring during the dry-hot season (MAM) and winter. However, the model results overestimate the ground-based  $PM_{2.5}$  concentrations during the dry-hot season (MAM) and in August.

#### 4. Conclusions

The study utilized spatially and temporally synchronized satellite data to investigate the relationship between several variables and environmental  $PM_{2.5}$ concentrations. It was discovered that the correlation between AOD and  $PM_{2.5}$  is influenced by RH, PBLH, and T. Based on AERONET calibration, using the AOD DB product for studying  $PM_{2.5}$  in the Megalopolis of Central Mexico is recommended. PBLH is a crucial parameter in explaining the differences between MODIS and AERONET AOD retrieval. The spatial correlation analysis revealed a significant change in the relationship between PBLH and AOD towards the end of the dry-hot season (MAM),

Table II. Multiple linear model (TeoPM<sub>2.5</sub>) correlation coefficients.

	$\alpha_0$	$\alpha_1$ AOD	α <sub>2</sub> NDVI	α <sub>3</sub> PBHL	$lpha_4$ RH	α T
January	2.86e-10	2.00e-09	-8.21e-11	-1.90e-13	5.65e-12	-1.49e-11
February	1.18e-10	5.20e-10	1.05e-10	-1.33e-13	7.48e-12	-1.30e-11
March	7.56e-10	1.81e-09	1.78e-11	1.08e-14	9.09e-13	-1.68e-11
April	2.27e-10	1.33e-09	2.85e-10	3.59e-13	1.06e-11	-5.26e-12
May	8.84e-10	-1.81e-10	1.09e-11	1.44e-14	-2.26e-12	3.31e-12
June	2.23e-09	4.82e-09	-1.99e-09	-8.71e-13	-1.83e-11	3.56e-11
July	2.54e-09	2.23e-08	-8.42e-09	-3.57e-12	1.16e-10	6.92e-11
August	7.08e-09	6.02e-09	-2.03e-09	-5.19e-12	-2.01e-11	6.03e-11
September	7.97e-10	1.57e-09	-1.12e-09	-4.23e-13	6.36e-12	4.24e-12
October	2.19e-10	1.33e-10	-3.73e-11	4.07e-14	-4.24e-12	9.32e-12
November	-2.12e-10	8.34e-10	-7.57e-11	1.67e-13	1.00e-11	-1.40e-11
December	1.14e-09	3.10e-09	-1.81e-10	-8.18e-13	1.20e-12	-1.76e-11



Fig. 12. Monthly mean  $PM_{2.5}$  concentrations calculated from the multiple linear regression model (TeoPM<sub>2.5</sub>) for the spatial domain in this study (enclosed by the black rectangle).



Fig. 13. Monthly mean PM<sub>2.5</sub> concentrations in 2012 from multiple linear regression model (TeoPM<sub>2.5</sub>, estimated; red-dotted line) and ground-based monitoring measurements in (a) Xalostoc (XAL) in Mexico City (19.526°N, –99.082°W) and Ninfas (NIN) in Puebla (19.043°N, –98.214°W). The inset (b) summarizes the correlation between estimated PM<sub>2.5</sub> concentrations (TeoPM<sub>2.5</sub>) and all the monitoring sites in Mexico City, the State of Mexico, and Puebla.

highlighting the importance of emissions transport from regions beyond the MCM. This information provides insights into potential mechanisms that could explain persistently high ozone levels or particulate concentrations despite local mitigation strategies during the dry-hot season (MAM).

The study developed a multiple linear regression model to estimate environmental PM2.5 levels. The research hypothesis that AOD data (MODIS) can estimate PM2.5 concentrations in the megalopolis of Mexico was confirmed. However, caution must be exercised in interpreting the model results since the analysis previous to the model revealed significant changes in the correlations of AOD, T, RH, and PBLH during each of the four climatic seasons. The model produced satisfactory results during the dryhot season (MAM) but failed during the rainy season (June-August). Finally, the concentrations of PM2.5 in the study area are influenced by fire emissions, and incorporating fire data as an additional variable could enhance the model's performance during this time of year (Jaffe et al., 2008; Vega et al., 2021).

# Acknowledgments

Thanks are due to Marco Herrera and Francisco J. Solano for Air Quality data support in Puebla.

# References

- AERONET. 2021. AErosol RObotic NETwork. Available at: http://aeronet.gsfc.nasa.gov. (Accessed 2022 February 18).
- Al-Saadi, J, Szykman, J, Pierce RB, Kittaka C, Neil D, Chu DA, Remer L, Gumley L, Prins E, Weinstock L, MacDonald C, Wayland R, Dimmick F, Fishman J. 2005. Improving National Air Quality Forecasts with Satellite Aerosol Observations. Bulletin of the American Meteorological Society 86: 1249–1261. https:// doi.org/10.1175/BAMS-86-9-1249
- Aldape F, Flores-M, F. 2011. Source Apportionment of fine airborne particulate matter collected in the Mexico City Metropolitan Area. International Journal of PIXE 14: 147–160. https://doi.org/10.1142/ S0129083504000215
- Bhaskaran S, Phillip N, Rahman A, Mallick J. 2011. Applications of Satellite Data for Aerosol Optical Depth (AOD) Retrievals and Validation with AERONET

Data. Atmospheric and Climate Sciences 1: 61-67. https://doi.org/10.4236/acs.2011.12007

- Bojanowski J S, Stöckli R, Tetzlaff A, Kunz H. 2014. The Impact of Time Difference between Satellite Overpass and Ground Observation on Cloud Cover Performance Statistics. Remote Sensing 6: 12866–12884. https:// doi.org/10.3390/rs61212866
- Buchard V, da Silva AM, Randles CA, Colarco P, Ferrare R, Hair J, Hostetler C, Tackett J, Winker D. 2016. Evaluation of the surface PM<sub>2.5</sub> in Version 1 of the NASA MERRA Aerosol Reanalysis over the United States. Atmospheric Environment 125: 100–111. https://doi. org/10.1016/j.atmosenv.2015.11.004
- Chudnovsky AA, Kostinski A, Lyapustin A, Koutrakis P. 2013. Spatial scales of pollution from variable resolution satellite imaging. Environmental Pollution 172: 131–138. https://doi.org/10.1016/j.envpol.2012.08.016
- Clark WJ. 1983. Cloud cover as a factor in the utilization of landsat data for limnological research. Remote Sensing of Environment 13: 453–460. https://doi. org/10.1016/0034-4257(83)90013-5
- Cohen AJ, Pope CA. 1995. Lung cancer and air pollution. Environmental Health Perspectives 103: 219-224. https://doi.org/10.1289/EHP.95103S8219
- Czwojdzińska M, Terpińska M, Kuźniarski A, Płaczkowska S, Piwowar A. 2021. Exposure to PM<sub>2.5</sub> and PM<sub>10</sub> and COVID-19 infection rates and mortality: A oneyear observational study in Poland. Biomedical Journal 44: S25–S36. https://doi.org/10.1016/J.BJ.2021.11.006
- Ebell K, Orlandi E, Hünerbein A, Löhnert U, Crewell S. 2013. Combining ground-based with satellite-based measurements in the atmospheric state retrieval: Assessment of the information content. Journal of Geophysical Research: Atmospheres 118: 6940–6956. https://doi.org/10.1002/JGRD.50548
- Fontes T, Li P, Barros N, Zhao P. 2017. Trends of PM2.5 concentrations in China: A long term approach. Journal of Environmental Management 196: 719–732. https:// doi.org/10.1016/j.jenvman.2017.03.074
- Gao L, Yan P, Mao J, Zhang X, Zhang X, Wu Y, Jing J, Xu J, Deng X, Chi W. 2021. Ambient Atmospheric Aerosol Extinction Coefficient Reconstruction from PM<sub>2.5</sub> Mass Concentrations and Application to Haze Identification in China. Aerosol and Air Quality Research 21: 200386. https://doi.org/10.4209/AAQR.200386
- Gelaro R, McCarty W, Suárez MJ, Todling R, Molod A, Takacs L, Randles CA, Darmenov A, Bosilovich MG,

Reichle R, Wargan K, Coy L, Cullather R, Draper C, Akella S, Buchard V, Conaty A, da Silva AM, Gu W, Kim G-K, Koster R, Lucchesi R, Merkova D, Nielsen JE, Partyka G, Pawson S, Putman W, Rienecker M, Schubert SD, Sienkiewicz M, Zhao B. 2017. The Modern-Era Retrospective Analysis for Research and Applications, Version 2 (MERRA-2). Journal of Climate 30: 5419-5454. https://doi.org/10.1175/ JCLI-D-16-0758.1

- Ghotbi S, Sotoudeheian S, Arhami M. 2016. Estimating urban ground-level PM<sub>10</sub> using MODIS 3km AOD product and meteorological parameters from WRF model. Atmospheric Environment 141: 333–346. https://doi.org/10.1016/j.atmosenv.2016.06.057
- Gimond M. 2022. J Interpolation in R. In: Intro to GIS and Spatial Analysis. 2022-08-24. Available at https:// mgimond.github.io/Spatial/interpolation-in-r.html (Accessed 2022 January 18).
- GIOVANNI. 2023. GES-DISC (Goddard Earth Sciences Data and Information Services Center) Interactive Online Visualization ANd aNalysis Infrastructure. Available at: http://disc.sci.gsfc.nasa.gov/giovanni (accessed 2023 January 20)
- HEI. 2022. Air Quality and Health In Cities: A State of Global Air Report 2022. Health Effects Institute, Boston, MA: Health Effects Institute.
- Holben BN, Eck TF, Slutsker I, Tanré D, Buis JP, Setzer A, Vermote E, Reagan JA, Kaufman YJ, Nakajima T, Lavenu F, Jankowiak I, Smirnov A. 1998. AER-ONET—A Federated Instrument Network and Data Archive for Aerosol Characterization. Remote Sensing of Environment 66: 1–16. https://doi.org/10.1016/ S0034-4257(98)00031-5
- Hsu NC, Tsay S-C, King MD, Herman JR. 2004. Aerosol properties over bright-reflecting source regions. IEEE Transactions on Geoscience and Remote Sensing 42: 557–569. https://doi.org/10.1109/ TGRS.2004.824067
- Jaffe D, Hafner W, Chand D, Westerling A, Spracklen D. 2008. Interannual Variations in PM<sub>2.5</sub> due to Wildfires in the Western United States. Environmental Science & Technology 42: 2812–2818. https://doi.org/10.1021/ ES702755V
- Kim K, Lee D, Lee K-Y, Lee K, Noh Y. 2016. Estimation of surface-level PM<sub>2.5</sub> concentration based on MO-DIS aerosol optical depth over Jeju, Korea. Korean Journal of Remote Sensing 32: 413-421. https://doi. org/10.7780/kjrs.2016.32.5.2

- Kong L, Xin J, Zhang W, Wang Y. 2016. The empirical correlations between PM<sub>2.5</sub>, PM<sub>10</sub> and AOD in the Beijing metropolitan region and the PM<sub>2.5</sub>, PM<sub>10</sub> distributions retrieved by MODIS. Environmental Pollution 216: 350-360. https://doi.org/10.1016/j.envpol.2016.05.085
- Lee B-J, Kim B, Lee K. 2014. Air Pollution Exposure and Cardiovascular Disease. Toxicological Research 30: 71-75. https://doi.org/10.5487/TR.2014.30.2.071
- Levy RC, Mattoo S, Munchak LA, Remer LA, Sayer AM, Patadia F, Hsu NC. 2013. The Collection 6 MODIS aerosol products over land and ocean. Atmospheric Measurement Techniques 6: 2989-3034. https://doi. org/10.5194/amt-6-2989-2013
- Li Y, Chen Q, Zhao H, Wang L, Tao R. 2015. Variations in PM<sub>10</sub>, PM<sub>2.5</sub> and PM1.0 in an Urban Area of the Sichuan Basin and Their Relation to Meteorological Factors. Atmosphere 6: 150–163. https://doi.org/10.3390/ atmos6010150
- Liang L, Gong P. 2020. Urban and air pollution: a multicity study of long-term effects of urban landscape patterns on air quality trends. Scientific Reports 10: 18618. https://doi.org/10.1038/s41598-020-74524-9
- Mora M, Braun RA, Shingler T, Sorooshian A. 2017. Analysis of remotely sensed and surface data of aerosols and meteorology for the Mexico Megalopolis Area between 2003 and 2015. Journal of Geophysical Research: Atmospheres 122: 8705–8723. https://doi. org/10.1002/2017JD026739
- Myhre G, Stordal F, Johnsrud M, Kaufman YJ, Rosenfeld D, Storelvmo T, Kristjansson JE, Berntsen TK, Myhre A, Isaksen ISA. 2007. Aerosol-cloud interaction inferred from MODIS satellite data and global aerosol models. Atmospheric Chemistry and Physics 7: 3081-3101 https://doi.org/10.5194/acp-7-3081-2007
- Paciorek CJ. 2010. The Importance of Scale for Spatial-Confounding Bias and Precision of Spatial Regression Estimators. Statistical Science 251: 107-125. https://doi.org/10.1214/10-STS326
- Pebesma EJ. 2004. Multivariable geostatistics in S: The gstat package. Computers & Geosciences 30: 683-691. https://doi.org/10.1016/j.cageo.2004.03.012
- RAMA. 1996. Informe Anual de la Calidad del Aire en la Ciudad de México 1995. Red Automática de Monitoreo Atmosférico de la Ciudad de México. Distrito Federal: Dirección General de Prevención y Control de la Contaminación Available at http://www.aire.cdmx.gob.mx/ descargas/publicaciones/flippingbook/informe\_anual\_calidad\_aire\_1995/#p=1 (Accessed 2023 May 29).

- Raster Package. 2023. raster v3.6-14, RDocumentation powered by DataCamp. Available at https://rdocumentation.org/packages/raster/versions/3.6-14 (accessed 2023 August 31).
- RAMA-CDMX. 2021. Red Automática de Monitoreo Atmosférico de la Ciudad de México, Available at: http://www.aire.cdmx.gob.mx (Accessed 2022 January 18).
- RAMA-EdoMex. 2021. Red Automática de Monitoreo Atmosférico del Estado de México. Available at: https://rama.edomex.gob.mx/calidaddelaire (Accessed 2022 January 18).
- Sakamoto Y. Ishiguro M. Kitagawa G. 1986. Akaike information criterion statistics. KTK Scientific Publishers, D. Reidel Publising Company.
- Sayer AM, Munchak LA, Hsu NC, Levy RC, Bettenhausen C, Jeong M-J. 2014. MODIS Collection 6 aerosol products: Comparison between Aqua's e-Deep Blue, Dark Target, and "merged" data sets, and usage recommendations. Journal of Geophysical Research: Atmospheres 119: 13965-13989 https://doi. org/10.1002/2014JD022453
- Schaap M, Apituley A, Timmermans RMA, Koelemeijer RBA, De Leeuw G. 2009. Exploring the relation between aerosol optical depth and PM<sub>2.5</sub> at Cabauw, the Netherlands. Atmospheric Chemistry and Physics 9: 909-925. https://doi.org/10.5194/acp-9-909-2009
- SINAICA. 2021. Sistema Nacional de la Información de la Calidad del aire. Available at: https://sinaica.inecc. gob.mx/ (Accessed 2022 December 12).
- Smirnov A, Holben BN, Eck TF, Dubovik O, Slutsker I. 2000. Cloud-screening and quality control algorithms for the AERONET database. Remote Sensing of Environment 73: 337-349. https://doi.org/10.1016/ S0034-4257(00)00109-7
- SSAOT. 2012. Programa de gestión de la calidad del aire (ProAire) para el Estado de Puebla 2012-2020. Puebla: Secretaría de Sustentabilidad Ambiental y Ordenamiento Territorial, Dirección de Calidad del Aire y Cambio Climático. Available at: https://dsiappsdev. semarnat.gob.mx/datos/portal/proaire/10\_ProAire%20 Puebla.pdf (accessed 2023 January 18).
- Stats v3.6.2. n.d. (The R Stats Package). Available at https://rdocumentation.org/packages/stats/versions/3.6.2 (Accessed 2022 December 12).
- Tiotiu AI, Novakova P, Nedeva D, Chong-Neto HJ, Novakova S, Steiropoulos P, Kowal K. 2020. Impact

of Air Pollution on Asthma Outcomes. International Journal of Environmental Research and Public Health 17: 6212. https://doi.org/10.3390/ IJERPH17176212

- Tripathi SN, Dey S, Chandel A, Srivastava S, Singh RP, Holben BN. 2005. Comparison of MODIS and AER-ONET derived aerosol optical depth over the Ganga Basin, India. Annales Geophysicae 23: 1093-1101. https://doi.org/10.5194/angeo-23-1093-2005
- van Donkelaar A, Martin RV, Park RJ. 2006. Estimating ground-level PM<sub>2.5</sub> using aerosol optical depth determined from satellite remote sensing. Journal of Geophysical Research: Atmospheres 111: D21201. https://doi.org/10.1029/2005JD006996
- Vega E, Namdeo A, Bramwell L, Miquelajauregui Y, Resendiz-Martinez CG, Jaimes-Palomera M, Luna-Falfan F, Terrazas-Ahumada A, Maji KJ, Entwistle J, Enríquez JCN, Mejia JM, Portas A, Hayes L, McNally R. 2021. Changes in air quality in Mexico City, London and Delhi in response to various stages and levels of lockdowns and easing of restrictions during COVID-19 pandemic. Environmental Pollution 285: 117664. https://doi.org/10.1016/J.ENVPOL.2021.117664
- Xu X, Zhang C. 2020. Estimation of ground-level PM<sub>2.5</sub> concentration using MODIS AOD and corrected regression model over Beijing, China. PLoS One: e0240430. https://doi.org/10.1371/JOURNAL.PONE.0240430
- Zhang P, Zhou X. 2021. Pricing air pollution: evidence from short-term exposure to air pollution on hospitalization of acute bronchitis and chronic obstructive pulmonary disease in southwestern China. International Health 14: 572-579. https://doi.org/10.1093/ INTHEALTH/IHAB071
- Zhang X, Cai C, Hu X-M, Gao L, Xu X, Hu J, Chen H. 2022. Aerosols consistently suppress the convective boundary layer development. Atmospheric Research 269: 106032. https://doi.org/10.1016/j. atmosres.2022.106032
- Zhao R, Gu X, Xue B, Zhang J, Ren W. 2018. Short period PM<sub>2.5</sub> prediction based on multivariate linear regression model. PLoS One 13: e0201011. https://doi. org/10.1371/journal.pone.0201011
- Zheng C, Zhao C, Zhu Y, Wang Y, Shi X, Wu X, Chen T, Wu F, Qiu Y. 2017. Analysis of influential factors for the relationship between PM<sub>2.5</sub> and AOD in Beijing. Atmospheric Chemistry and Physics 17: 13473–13489. https://doi.org/10.5194/acp-17-13473-2017