

Modeling of the PM₁₀ pollutant in Monterrey, Nuevo León, Mexico, with ARIMA, transfer functions and GARCH models

Gustavo Andrés RAMÍREZ-GÓMEZ^{1*}, Martha Elva RAMÍREZ-GUZMÁN¹, Juan Manuel ROMERO-PADILLA¹,
Antonia MACEDO-CRUZ¹, Carlos Alberto ORTIZ-SOLORIO¹, Gilberto RENDÓN-SÁNCHEZ¹,
Edgar Vladimir GUTIÉRREZ-CASTORENA² and Vicente Vidal ENCINIA-URIBE²

¹ *Colegio de Postgraduados, km 36.5 carretera México-Texcoco, Montecillo, 56264 Texcoco, Estado de México, México.*

² *Facultad de Agronomía, Universidad Autónoma de Nuevo León, Francisco I. Madero s/n, Col. Ex Hacienda El Canadá, 66050 General Escobedo, Nuevo León, México.*

*Corresponding author; email: ramirez.andres@colpos.mx

Received: March 25, 2025; accepted: June 24, 2025

RESUMEN

Las partículas menores a 10 µm (PM₁₀) son contaminantes relevantes por su impacto en la salud y el ambiente. En este estudio se analizaron datos diarios de PM₁₀ de la estación Obispedo (Monterrey, Nuevo León, México), del INECC, entre 1997 y 2014. Se incluyeron variables climatológicas como precipitación, velocidad del viento (a 2 y 10 m) y presión atmosférica, obtenidas del proyecto POWER de la NASA. Los datos se dividieron en conjuntos de entrenamiento y prueba utilizando una proporción de 80-20%. Se emplearon 3798 observaciones diarias para entrenamiento (2002 a 2012) y 950 observaciones para prueba (2012 a 2014). Se realizó un análisis descriptivo y una descomposición de series temporales para identificar tendencias y estacionalidad. Se aplicaron modelos ARIMA univariados y con funciones de transferencia, incorporando variables climáticas. Las pruebas de Dickey-Fuller aumentada y Ljung-Box validaron la estacionalidad e independencia de los residuales. El modelo ARIMA (1,1,1)(0,0,0)[365] con funciones de transferencia fue más preciso que el univariado. Se identificó una relación significativa entre las variables climáticas y la predicción de PM₁₀, lo cual respalda su uso para predicciones a corto plazo (≤ 10 días). Se recomienda aplicar modelos multivariados con más predictores y enfoques geoestadísticos para mejorar la caracterización espaciotemporal.

ABSTRACT

Particles of matter smaller than 10 µm (PM₁₀) are significant pollutants due to their impact on health and the environment. This study analyzed daily PM₁₀ data from the Obispedo station (Monterrey, Nuevo León, Mexico) provided by the INECC, covering the period from 1997 to 2014. Climatological variables such as precipitation, wind speed (at 2 and 10 m), and atmospheric pressure, obtained from NASA's POWER project, were included. Data were partitioned into training and testing sets using an 80-20% split, employing 3798 daily observations for training (2002 to 2012) and 950 observations for testing (2012 to 2014). Descriptive analysis and time series decomposition were performed to identify trends and seasonality. ARIMA models, univariate and with transfer functions incorporating meteorological variables, were applied. Augmented Dickey-Fuller and Ljung-Box tests validated the stationarity and residual independence assumptions. The ARIMA (1,1,1)(0,0,0)[365] model with transfer functions outperformed the univariate model. A significant relationship between meteorological variables and PM₁₀ predictions was identified, supporting their use for short-term forecasting (≤ 10 days). Future studies should consider applying multivariate models with additional predictors and geostatistical approaches to improve spatiotemporal characterization.

Keywords: time series, transfer functions, climate variables, Granger causality test.

1. Introduction

The rise of urban pollutants is a major concern due to their harmful effects on respiratory health, as it was seen in the deaths of Belgium (1930), Pennsylvania (1948), and London (1952) (Rodríguez et al., 2011). Furthermore, environmental pollution promotes the development of affective disorders and brain diseases, and significantly impacts mental health due to the toxicity of specific pollutants for the central nervous system (Ordóñez-Iriarte, 2020).

Due to their physical and chemical properties, particle size plays a fundamental role in solid pollutants, directly impacting climate and public health. Myridakis and Stephanou (2025) found a strong correlation between increased mortality and the presence of particles smaller than 10 μm (PM_{10}) and 2.5 μm ($\text{PM}_{2.5}$).

PM_{10} particles originate from various sources, such as dust, pollen, smoke, and soot, combined with liquid droplets of different substances (Paital and Agrawal, 2022). Their formation can result from both natural processes, such as volcanic eruptions, sandstorms, wildfires, soil erosion, or chemical reactions between gases released into the atmosphere, and anthropogenic activities, including industrial processes and the combustion of fossil fuels (Ukaogo et al., 2020).

Pollution affects the immune system and causes hormonal changes that can lead to immunosuppressive or autoimmune diseases (Huang et al., 2019). At the genetic level, exposure to solid pollutants has been observed to cause DNA protein destruction, promoting the development of various types of cancer (Vallabani et al., 2023), as well as altering proteins involved in the regulation of genotoxins, increasing their harmful effects (Badran et al., 2020).

Recently, Scapini et al. (2023) found a positive relationship between atmospheric pollutant concentrations and the spread of the SARS-CoV-2 virus, with a proportional increase in weekly cases corresponding to PM_{10} levels. On the other hand, Buoli et al. (2018) reported an increased risk of psychotropic drug prescriptions in children and adolescents for every 10 μm rise in solid pollutant concentrations. Higher pollution levels raise the risk of emergency room visits, making pollutant modeling and forecasting vital for public health and policy planning.

Additionally, studies using geostatistical techniques such as the nearest neighbor index (NNI) and nearest neighbor hierarchical clustering (NNHC) have linked high concentrations of population to an increased incidence of breast cancer (Gasca-Sánchez et al., 2021). These conditions contribute to health problems, especially in individuals with pre-existing respiratory or visual diseases. Inhaled particles can trigger both local and systemic effects by activating monocytes even at very low concentrations (ng mL^{-1}). Such activation increases the expression of adhesion molecule receptors, enhancing interactions between monocytes and endothelial cells, potentially exacerbating systemic inflammation (Quintana-Belmares et al., 2018; Leal-Iga, 2019).

Some of the most used models for forecasting atmospheric pollutant concentrations include time series models, neural networks, and multivariate adaptive regression splines (MARS), which combine the flexibility of machine learning methods with the interpretability of traditional statistical models (Alvarado et al., 2010).

In this context, the use of time series is widely employed to model the behavior and prediction of PM_{10} on a large scale. An example of this is the case of China, where 272 cities were analyzed, and a relationship was found between pollution levels and mortality from non-accidental causes such as cardiopulmonary diseases (Renjie et al., 2025).

Silva et al. (1994) used time series models that incorporate transfer functions with meteorological variables, which account for up to 40% of the mean absolute error in their predictions. With the same kind of models, Alvarado et al. (2010) recorded peak values of 240 μm in large metropolises such as Santiago de Chile.

Due to its high population density, the Mexico City Metropolitan Area, which averages PM_{10} concentrations of $75 \mu\text{g m}^{-3}$ at a specific hourly frequency (Villaseñor et al., 2000), places the most significant emphasis on air pollution monitoring. In addition to the high levels of atmospheric pollutants, this area is considered one of the most polluted urban regions in the world (Cárdenas-Moreno et al., 2021).

In the case of the city of Monterrey, located in the northern state of Nuevo León, records on PM_{10} pollution have not been as extensive as in Mexico City, where air quality has been systematically recorded for over 40 years (Raga et al., 2001). Records in Monterrey

were initiated in the early 2000s (Aguirre-López et al., 2022), and an annual PM₁₀ concentration of 61 to 80 $\mu\text{g m}^{-3}$ has been reported (Cong et al., 2019). Levels exceeding 60 $\mu\text{g m}^{-3}$ have been associated with industrial activities, particularly the operation of cement factories and food processing plants in the region. On the other hand, the Mexican Official Standard NOM-025-SSA1-2014, which regulates the permissible air quality limits for PM₁₀ at the national level, establishes maximum values of 75 $\mu\text{g m}^{-3}$ as a 24-h average and of 40 $\mu\text{g m}^{-3}$ as an annual average (SSA, 2014).

Likewise, it has been reported that PM₁₀ exhibits a stationary behavior over long periods, while over short periods it shows non-stationary characteristics with a seven-day periodicity, fitting Gamma, Weibull, and logarithmic distributions (Cárdenas-Moreno et al., 2021).

The present study aims to model and forecast the behavior of PM₁₀ with a daily frequency in the city of Monterrey, Nuevo León, Mexico, using ARIMA models with exogenous variables. Additionally, a descriptive analysis of PM₁₀ concentrations over the study period is provided, detailing significant variations, general trends, and specific episodes of elevated levels to understand and characterize the behavior of PM₁₀.

2. Materials and methods

2.1 Data

The PM₁₀ data used in this study were obtained from the National Institute of Ecology and Climate Change (INECC, 2019), with hourly information available for the period 2002 to 2014. The selected monitoring station was Obispado (code: CE), located in Monterrey, Nuevo León, with geographic coordinates 25.67598 latitude and -100.3384 longitude.

The climatological variables were extracted on an hourly basis from the POWER project (v. 2.1.17) of NASA's Langley Research Center (NASA, 2025). These variables include precipitation (PP, mm hr^{-1}), wind speed at 2 m above the ground (WS2M, m s^{-1}), wind speed at 10 m above the ground (WS10M, m s^{-1}), and atmospheric pressure (PS, kPa).

2.2 Data processing

2.2.1 Imputation and removal of outliers

Data processing and analysis were conducted using R (R Core Team, 2023). Outliers were removed using

the interquartile range method; then, the daily mean value of the pollutant was calculated, representing the average of the values recorded throughout the day.

Subsequently, missing data were imputed using the Kalman filter method, particularly suitable for time series with strong seasonality. This method utilizes a state-space model and the Kalman filter to estimate missing values, taking into account trends, noise, seasonal patterns, and underlying temporal relationships. Imputation and removal of outliers were performed with the `imputeTS` R library (Moritz and Bartz-Beielstein, 2017).

2.2.2 Yeo-Johnson transformation

Yeo-Johnson transformations are a generalization of Box-Cox transformations, designed to handle both positive and negative data using a parameter λ , which defines the transformation (Table I). The mathematical expression of the transformation varies depending on the sign of x and the value of λ , dividing into different cases according to these conditions, as shown in Eq. (1).

$$y = \begin{cases} \frac{(x+1)^\lambda - 1}{\lambda} & \text{if } \lambda \neq 0, x \geq 0 \\ \log(x+1) & \text{if } \lambda = 0, x \geq 0 \\ -\frac{(-x+1)^{2-\lambda} - 1}{2-\lambda} & \text{if } \lambda \neq 2, x < 0 \\ -\log(-x+1) & \text{if } \lambda = 2, x < 0 \end{cases} \quad (1)$$

Table I. Estimated λ values in the Yeo-Johnson transformation.

Variable	λ
PM ₁₀	0.0990325
WS2M	-0.1796091
WS10M	-0.1654505
PS	-2.847771

PM₁₀: particles smaller than 10 μm ; WS2M (WS10M): wind at 2 (10) m above ground; PS: atmospheric pressure.

where x represents the original variable to be transformed and λ is the transformation parameter. Likewise, inverse transformations (Eq. [2]) can be obtained with the best normalization library (Peterson, 2023)

$$x = \begin{cases} (\lambda y + 1)^{\frac{1}{\lambda}} & \text{if } y \geq 0, \lambda \neq 0 \\ \exp(y) - 1 & \text{if } y \geq 0, \lambda = 0 \\ 1 - (\lambda(1 - y) + 2)^{\frac{1}{2-\lambda}} & \text{if } y < 0, \lambda \neq 2 \\ 1 - \exp(-y) & \text{if } y < 0, \lambda = 2 \end{cases} \quad (2)$$

PM₁₀ and the climatological variables were transformed with the first term of Eq. (1). The corresponding λ values applied to the data are shown in Table I.

Inverse transformation of PM₁₀ and the climatological variables were obtained with the first term of Eq. (2).

2.2.3 Training and testing data

The data were partitioned into training and testing sets using an 80-20% split, resulting in 3798 data points for training. The remaining 950 data points were used to compute an annual daily average for model testing. The training set covered the period from 2002 to 2012, while the test set spanned from 2012 to 2014.

2.3 ARIMA model

2.3.1 Univariate ARIMA

An ARIMA model is a statistical model used to analyze and predict future values of a time series based on its past values. Its structure is based on the combination of autoregressive (AR), differencing (d, D), and moving average (MA) components, allowing it to capture both trends and seasonal patterns in the data, as shown in Eq. (3).

$$\begin{aligned} & \left(1 - \sum_{i=1}^p \phi_i B^i\right) \left(1 - \sum_{j=1}^P \Phi_j B^{j \times S}\right) (1 - B)^d \\ & (1 - B^S)^D Y_t = \mu + \left(1 + \sum_{k=1}^q \theta_k B^k\right) \\ & \left(1 + \sum_{m=1}^Q \Theta_m B^{m \times S}\right) a_t \end{aligned} \quad (3)$$

where Y_t represents the value of the series at time t (in this case, the PM₁₀ pollutant). The parameters d and D indicate the regular and seasonal differencing orders, respectively, while p and q represent the orders of the AR and MA terms. P and Q correspond to the orders of the AR and MA terms in their seasonal component. S defines the periodicity of the time

series, which in this case is 365, corresponding to the daily frequency, and μ represents the meaning of the output series. The coefficients ϕ_i and Φ_j are associated with the AR terms, both in their regular and seasonal versions. In contrast, θ_k and Θ_m correspond to the coefficients of the MA terms, also in their regular and seasonal variants. Finally, a_t is the error term at time t , assuming a mean of zero.

2.3.2 ARIMA with transfer functions

A general transfer function model can AR terms for the output series, MA terms, and transfer function terms, representing the dynamic relationship between the input and output variables.

The general form of a transfer function model can be expressed as Eq. (4):

$$\begin{aligned} & \Phi(B^S) \phi(B^S) \nabla^d \nabla_S^D Y_t = \\ & \Theta(B^S) \theta(B^S) e_t + \sum_{k=1}^r \left[\frac{\omega_k(B)}{\delta_k(B)} B^{d_k} X_{k,t} \right] \end{aligned} \quad (4)$$

where $\omega_k(B) = \omega_{k,0} + \omega_{k,1}B + \dots + \omega_{k,r_k} B^{r_k}$ is the numerator (or gain) polynomial for the exogenous variable $X_{k,t}$. $\delta_k(B) = 1 - \delta_{k,1}B - \dots - \delta_{k,s_k} B^{s_k}$ is the denominator (or feedback) polynomial. The term B^{d_k} represents a pure lag (d_k steps) before $X_{k,t}$ impacts y_t . If the effects of $X_{k,t}$ were instantaneous and had no memory, then $\omega_k(B)$ and $\delta_k(B)$ are reduced to constants (order 0), and sometimes $d_k = 0$; $\sum_{k=1}^r \left[\frac{\omega_k(B)}{\delta_k(B)} B^{d_k} X_{k,t} \right]$ represents the transfer component of each exogenous variable. Y_t can be explicitly written as Eq. (5):

$$\begin{aligned} & Y_t = \left[\Phi(B^S) \phi(B^S) \nabla^d \nabla_S^D \right]^{-1} \\ & \left[\Theta(B^S) \theta(B^S) e_t + \sum_{k=1}^r \frac{\omega_k(B)}{\delta_k(B)} B^{d_k} X_{k,t} \right] \end{aligned} \quad (5)$$

The inverse operator of Eq. (5) can be expanded as an infinite series in B , reflecting the dependence of y_t on its lags (ARIMA) and the lags of the exogenous variables.

On the other hand, in the “static” transfer model or ARIMAX without exogenous lags, each variable $X_{k,t}$ influences y_t only with its contemporary value (at the same instant t). No polynomials in B are included to describe possible delays or attenuations over time, so the effect of each $X_{k,t}$ is reflected instantaneously

in y_t , as shown in Eq. (6), based on the work of Box et al. (2015).

$$Y_t = \delta^{-1}(B)\omega(B)X_{t-b} + N_t \quad (6)$$

where N_t is the noise represented by the ARIMA (p, d, q) model, as shown in Eq. (7).

$$N_t = \varphi^{-1}(B)\theta(B)a_t \quad (7)$$

where a_t is the white noise of the model, which can finally be expressed as Eq. (8):

$$Y_t = \delta^{-1}(B)\omega(B)X_{t-b} + \varphi^{-1}(B)\theta(B)a_t \quad (8)$$

2.4 GARCH models

The generalized autoregressive conditional heteroskedasticity (GARCH) model is widely used in time series analysis, especially in finance, to model and predict volatility. It extends the autoregressive conditional heteroskedasticity (ARCH) model by incorporating past conditional variances into the equation. The GARCH model helps capture volatility clustering, a common feature in financial time series where periods of high volatility are followed by periods of even higher volatility, and calm periods follow each other. The volatility is modelled by the variance of the mean (μ) process (which can be modelled by an ARIMA model), and random shocks (Eq. [9]).

$$\epsilon_t = \sigma_t z_t \quad y_t = \mu + \epsilon_t \quad z_t \sim N(0,1)$$

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^q \alpha_i \epsilon_{t-i}^2 + \sum_{j=1}^p \beta_j \sigma_{t-j}^2 \quad (9)$$

where σ_t^2 is the conditional variance at time t , $\alpha_0 > 0$, $\alpha_i \geq 0$; $\beta_j \geq 0$ are parameters to be estimated; p represents GARCH terms (effect of past variances), and q represents ARCH terms (effect of past squared residuals). GARCH models are useful to model residuals of an ARIMA model.

2.5 Metrics for model selection and evaluation of prediction accuracy

The metrics used in time series analysis can be divided into two categories based on the phase in which they are applied: training and testing. During the training phase, criteria such as the Akaike information

criterion (AIC), its corrected version (AICc), and the Bayesian information criterion (BIC) are used to assess the model's fit to the observed data, penalizing complexity to avoid overfitting (Hyndman and Athanasopoulos, 2018).

On the other hand, during the testing phase, metrics are used to measure the error between the model's predictions and the training values (which were not used to fit the model). These include the mean absolute error (MAE) (Cao, 2024) and the mean squared error (MSE), along with its square root (RMSE) (Gladkova and Saychenko, 2022). These metrics help evaluate the model's ability to generate accurate predictions on unseen data, ensuring its generalization and usefulness in future scenarios.

Granger causality analysis is an econometric method based on the premise that if one variable contributes to the prediction of another, its past values must contain relevant information that improves the estimation of the second variable beyond what its past values can provide. This approach has been used in previous studies to analyze the relationship between pollutants and various economic variables (Xuan, 2024). In the context of this study, it will be applied to assess whether there is a causal relationship between meteorological variables and PM₁₀ concentration, determining to what extent meteorological conditions can influence the variability of this pollutant.

3. Results and discussion

3.1 Descriptive statistics of the time series

The variables with the highest dispersion are precipitation (PP) and PM₁₀ (Fig. 1). This is evident in the descriptive statistics (Table II), which reveal a considerable difference between minimum and maximum values. In the case of precipitation, it shows positive skewness coefficients, indicating a non-normal distribution. Additionally, its kurtosis values are also positive and high, notably significant due to the daily frequency used in this study. This contrasts with analyses based on annual averaged data, where variability tends to decrease (Villarreal-Macés and Díaz-Viera, 2018).

On the other hand, the variance of PM₁₀ is moderate to high, which could be attributed to the use of only one year of records in such analyses. These results highlight the importance of considering both

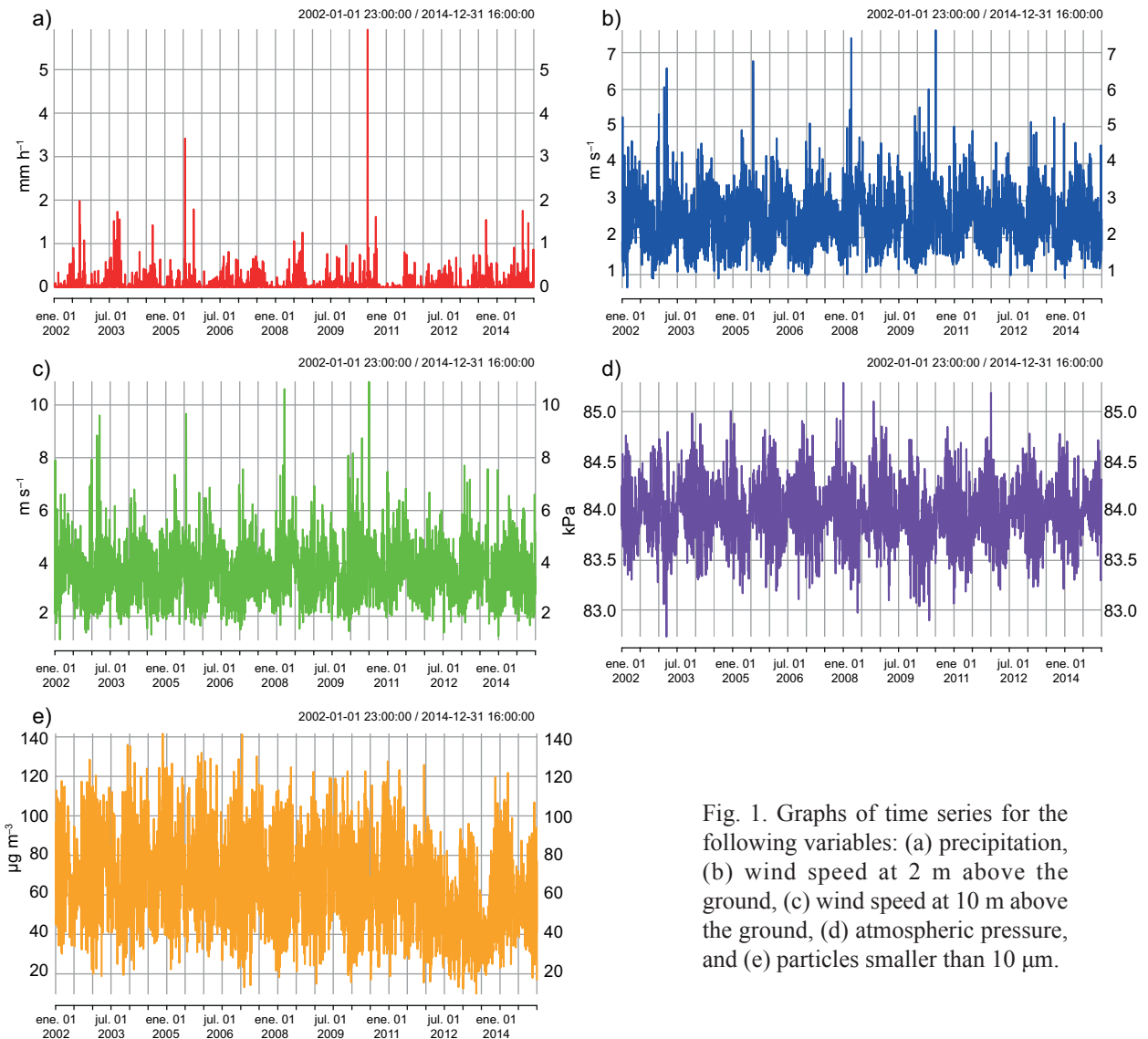


Fig. 1. Graphs of time series for the following variables: (a) precipitation, (b) wind speed at 2 m above the ground, (c) wind speed at 10 m above the ground, (d) atmospheric pressure, and (e) particles smaller than $10\ \mu\text{m}$.

Table II. Descriptive statistics of climatology variables and PM_{10} .

Statistics	PP	WS2M	WS10M	PS	PM_{10}
Minimum	0.00	0.65	1.11	82.73	9.76
Maximum	5.92	7.60	10.87	85.29	141.60
Quartile 1	0.00	1.92	2.94	83.79	49.79
Quartile 3	0.04	2.89	4.19	84.16	81.78
Mean	0.06	2.44	3.62	83.98	66.07
Median	0.00	2.38	3.53	83.98	65.54
Variance	0.0347	0.4975	0.9173	0.0853	501.7
Standard Deviation	0.186	0.705	0.958	0.292	22.4
Skewness	11.47	0.84	0.98	0.047	0.184
Kurtosis	258.43	2.25	3.12	0.46	0.404

PP: precipitation; WS2M (WS10M): wind at 2 (10) m above ground; PS: atmospheric pressure; PM_{10} : particles smaller than $10\ \mu\text{m}$.

frequency and period of observation when analyzing dispersion and variability in time series (Contreras-Arreola and González, 1999).

On the other hand, the variables WS2M and WS10M exhibit relatively close values, with WS10M consistently higher, as expected due to the well-known increase of horizontal wind speed with altitude within the atmospheric surface layer. Finally, atmospheric pressure (PS) remains relatively constant throughout the period analyzed, exhibiting the lowest skewness and kurtosis values, which suggests an approximately normal distribution (Table II). This normality assumption is statistically relevant because it allows the use of parametric statistical methods that rely on normality, such as linear regression, correlation analyses, and hypothesis testing procedures, without requiring data transformation. Additionally, this characteristic simplifies modeling and forecasting procedures, improving the reliability and interpretability of results derived from standard parametric techniques.

3.2 Decomposition of the PM₁₀ series

Before any transformation is applied to the data, the time series is decomposed into its original components

using the multiplicative method. This method captures the interaction between trend, seasonality, and random variability more precisely, improving its interpretation.

Figure 2a shows a significant decrease in PM₁₀ levels, which becomes more pronounced toward the end of 2012. However, abrupt increases are identified during 2002-2004, 2006-2008, and 2011-2012. Figure 2b also reveals a well-defined seasonal pattern in the PM₁₀ series, indicating recurring fluctuations throughout the year, likely associated with changes in meteorological conditions and other environmental factors. The random component in Figure 2c exhibits relatively constant variability, except for certain outliers, mainly in 2013 and 2014.

3.3 Breakpoints in the Time Series PM₁₀

Breakpoint analysis in time series is a technique that allows for detecting structural changes based on the lowest values of the Bayesian information criterion (BIC) and the residual sum of squares (RSS) (Bai and Perron, 1998, 2003; Zeileis et al., 2003).

Based on Figure 3, it was determined that the model with four structural breakpoints is appropriate.

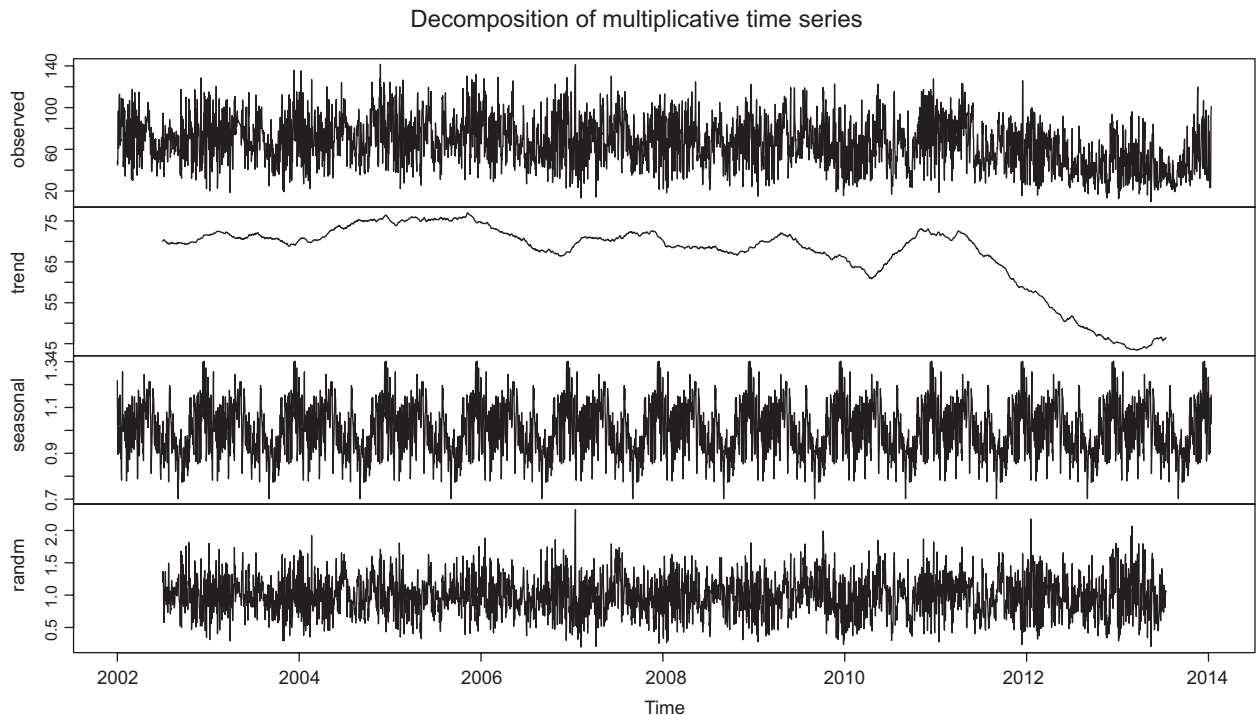


Fig. 2. Time series decomposition for PM₁₀: (a) trend, (b) seasonal, and (c) random.

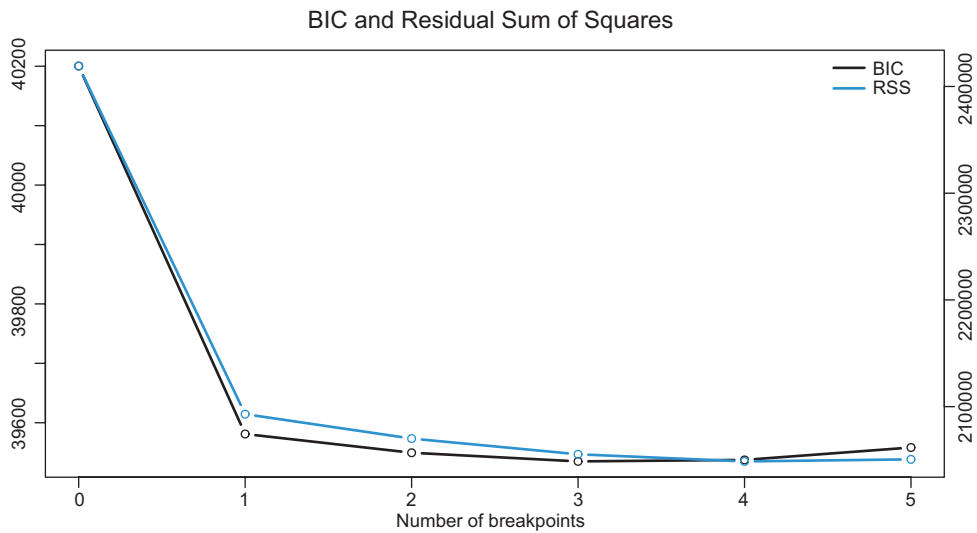


Fig. 3. Bayesian information criterion (BIC) and residual sum of squares (RSS) values for the time series of PM₁₀.

Adding a fifth breakpoint increases complexity without substantially improving model fit, making it unnecessary according to the principle of parsimony (Table III).

Based on Table IV, breakpoints were identified in 2004, 2006, 2009, and 2012. Moderate variance between segments is observed at the initial breakpoints (Fig. 4); however, variance reaches its highest level in 2009 compared to other segments. Subsequently, there is a considerable decrease in variance at the last breakpoint. This suggests a significant reduction in pollutant concentrations starting from 2006, possibly linked to changes in environmental policies or specific meteorological conditions. Nevertheless, when analyzing the last segment, an increase in pollutant concentration becomes evident. Further data are required to determine whether this upward trend will persist or if atmospheric concentrations will decrease again.

Table IV. Breakpoints based on RSS and BIC.

<i>m</i>	Break years				
1	2011				
2	2006	2012			
3	2004	2006	2012		
4	2004	2006	2009	2012	
5	2004	2006	2008	2009	2012

m: number of breakpoints; BIC: Bayesian information criterion; RSS: residual sum of squares.

Figures in bold represent the number of breakpoints in the time series selected as the best.

3.4 Transformation of time series

Based on the skewness and kurtosis statistics presented in Table II and the Jarque-Bera normality test results, which yielded significant p-values, the null hypothesis of normality is rejected. This indicates that the meteorological variables and

Table III. Model fit for different numbers of breakpoints.

Number of breakpoints	0	1	2	3	4	5
RSS	2419141	2092984	2070101	2055377	2048705	2050614
BIC	40200	39581	39550	39535	39537	39558

RSS: residual sum of squares; BIC: Bayesian information criterion. Figures in bold correspond to the lowest RSS and BIC values.

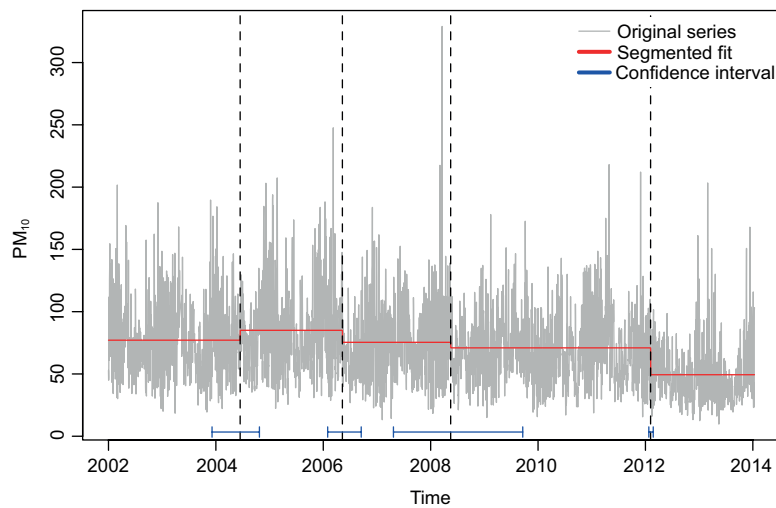


Fig. 4. Breakpoints identified in the time series of PM₁₀.

PM₁₀ pollutant variables do not follow a normal distribution.

In this context, series transformations were performed (Fig. 5) to reduce their variance, following the algorithm's recommendations for ARIMA model selection. Although these transformations did not fully achieve normality, they significantly reduced the variance, which is essential for stabilizing the time series and facilitating model fitting, as stated by Hyndman and Athanasopoulos (2018).

The autocorrelation (ACF) and partial autocorrelation (PACF) analysis of the PM₁₀ pollutant reveals specific characteristics of the time series. In the ACF (Fig. 6a), a gradual decay is observed, along with significant peaks reflecting the influence of seasonality and persistence over time. This behavior suggests a prolonged correlation between past and future values of the pollutant.

On the other hand, the PACF (Fig. 6b) shows a more limited behavior, with only a few significant initial lags, indicating that direct effects from specific lags rapidly diminish. However, certain persistent lags reflect strong seasonal patterns, consistent with the observations from the ACF.

3.5 Granger causality tests

Significant F-values were obtained for all comparisons conducted (Table V), suggesting that meteorological conditions have a significant effect on the dynamics of PM₁₀. This reinforces

the importance of including these variables in the model, consistent with Granger's (1969) hypothesis, which states:

H_0 : The lags of x do not provide significant information for predicting y .

H_1 : The past values of x contain useful information for predicting y .

3.6 ARIMA Models

3.6.1 Univariate model

The ARIMA (1,1,2)(0,0,0)[365] model fitted to the training data for the PM₁₀ variable was evaluated based on the significance of its coefficients, which were highly significant (Table VI), indicating that the model effectively captures the dynamics of the PM₁₀ time series.

The non-seasonal component of the model included an autoregressive term, indicating that the pollutant levels on any given day are directly influenced by the previous day's values. Differencing was necessary to stabilize the series and ensure stationarity. Additionally, the model included two moving average terms, reflecting the influence of past errors on current predictions.

The fitted ARIMA (1,1,2)(0,0,0)[365] model (Eq. [10]) satisfied the criteria of residual independence, as assessed by the Ljung-Box test, and stationarity, verified through the augmented Dickey-Fuller (ADF) test (Table VII).

Table V. Results of Granger causality tests.

Comparison	F-value	p-value
PM ₁₀ vs. PP	144	< 2.2e - 16
PM ₁₀ vs. WS2M	242.08	< 2.2e - 16
PM ₁₀ vs. WS10M	169.14	< 2.2e - 16
PM ₁₀ vs. PS	371.59	< 2.2e - 16

PP: precipitation; WS2M (WS10M): wind at 2 (10) m above the ground; PS: atmospheric pressure; PM₁₀: particles smaller than 10 μm.

Figures in bold indicate significance.

$$PM_{10,t} = 1.2787 PM_{10,t-1} - 0.2787PM_{10,t-2} - 0.7196a_{t-1} - 0.2392a_{t-2} + a_t \quad (10)$$

Furthermore, when comparing the MAE obtained in this study with the results reported by Hernández et al. (2021), it is evident that the univariate ARIMA model exhibits a significantly lower MAE (1.8664 μg m⁻³) than the one obtained using the evolutionary algorithm (12.5 μg m⁻³). This difference might be attributed to the evolutionary algorithm being applied by Hernández-Vega et al. (2021), whose study covered only seven months with hourly frequency and was applied at a nearby, but not identical, monitoring station.

It is worth noting that an evolutionary algorithm is an optimization method inspired by the principles of natural selection and biological evolution, in which a population of candidate solutions evolves through multiple iterations using operators such as mutation, crossover, and selection, to progressively find optimal or near-optimal solutions for a given problem.

However, the mean absolute percentage error (MAPE) of the ARIMA model is considerably higher due to noise that is not fully explained by the model,

Table VII. Selection of ARIMA models and accuracy metrics for PM₁₀.

Accuracy metrics	
AIC	9044.21
AICc	9044.22
BIC	9069.18
Ljung-box (p-value)	0.9658
Augmented Dickey-Fuller (p-value)	> 0.01
RMSE	1.1328
MSE	1.2831
MAE	0.9208
MAPE	611.64

AIC: Akaike information criterion; AICc: Akaike information criterion corrected version; BIC: Bayesian information criterion; RMSE: root mean squared error; MSE: mean squared error; MAE: mean absolute error; MAPE: mean absolute percentage error.

distorting the calculations by inflating errors. This results in a higher percentage error compared to other metrics, reflecting unexplained variability (noise) within the series.

3.6.2 Residual analysis of univariate models

The residuals of the ARIMA models exhibited stationarity and independence, indicating that their mean and variance remain constant over time, and no evidence of significant autocorrelation between lags was found. This behavior further supports the overall validity of the model.

In the ACF plots (Fig. 7), showing residual correlations up to lag 40, most values fall within the confidence limits. This indicates that the fitted model accurately captures the dynamics of the time series and does not leave significant autocorrelations unaccounted for. However, two significant peaks are

Table VI. Coefficient test for the univariate ARIMA (1,1,2)(0,0,0)[365] model.

Coefficient	Estimate	Standard error	z value	Pr (> z)
ϕ_1	0.278737	0.035123	7.9359	2.089e⁻¹⁵
θ_1	0.719585	0.034952	20.5878	< 2.2e⁻¹⁶
θ_2		0.030916	7.7386	1.005e⁻¹⁴

Figures in bold indicate significance.

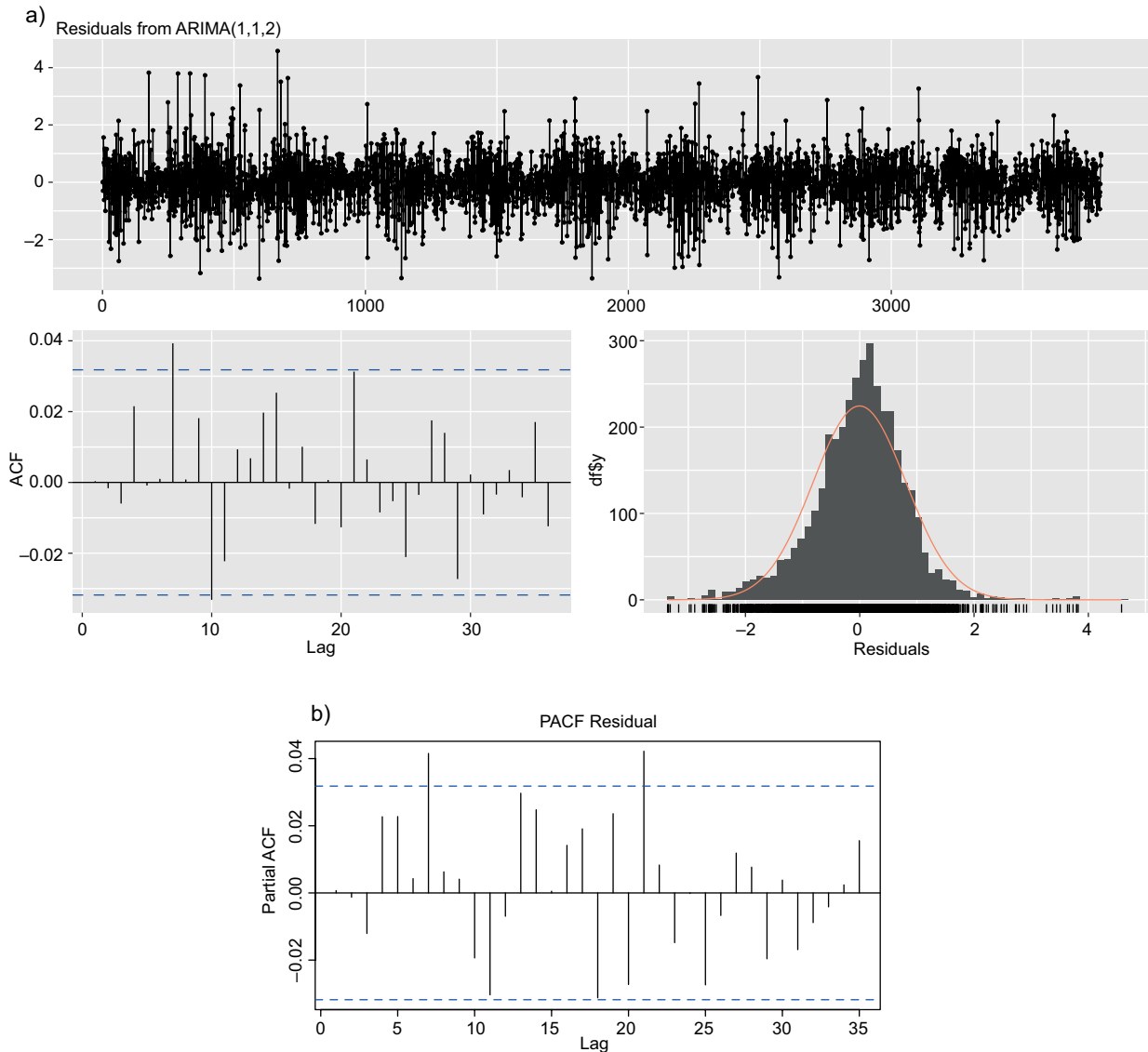


Fig. 7. Autocorrelation function (ACF) and partial autocorrelation function (PACF) of residuals from the ARIMA (1,1,2)(0,0,0)[365] model.

identified at lags 7 and 21, which could be associated with sporadic events.

The skewness observed in the residual histogram reflects a moderate asymmetry, while the kurtosis suggests a slight deviation from normality. Despite these deviations, residuals predominantly remain within acceptable statistical limits, supporting the validity of the model fit for the PM_{10} time series. Additionally, the PACF plot (Fig. 7) illustrates correlations between a value and its specific lag while

controlling for intermediate lag effects. In our case, it indicates that most correlations remain within confidence bounds for residuals, which implies that the fitted model adequately captures the dynamics of the PM_{10} pollutant time series.

3.7 Transfer function ARIMA model for PM_{10}

A general transfer function model incorporates AR and MA terms to capture the internal dynamics of the output series, as well as transfer terms that represent

the dynamic relationship between input variables and the output variable. In this case, the fitted model for the PM₁₀ series is an ARIMA (1,1,1)(0,0,0)[365], which includes transfer function terms for the meteorological variables (precipitation, wind speed at 2 and 10 m above ground, and atmospheric pressure) (Eq. [11]).

$$PM_{10t} = -0.0726PP_t - 1.471WSM2_t + 1.4729WSM10_t - 0.2699PS + \frac{(1 - 0.9662B)}{(1 - 0.4367B) \nabla^{d=1}(B)} a_t \quad (11)$$

The estimated coefficients for the transfer function model were significant (Table VIII), indicating that each term significantly contributes to the model's fit, similar to the study by (Analitis et al., 2020), where meteorological variables were used to analyze their relationship with particulate matter

The autoregressive coefficient (ϕ_1) represents the dependence of Y_t on its own lagged value Y_{t-1} . A positive value indicates that if the difference (change) in the previous period was high and/or positive, the current period's difference tends to follow the same pattern. A value close to one for a_{t-1} suggests that the error from the previous period strongly influences the current error, representing a "rebound" effect. This occurs when a large error in one period tends to be compensated in the following one, which is common in phenomena with rapid fluctuations. In this case, it indicates that exogenous variables cause abrupt

Table VIII. Estimated coefficients of the ARIMA PM₁₀ model with transfer function terms.

Coefficient	Estimate	Standard error	p-values
ϕ_1	0.4367	0.0185	< 2.2e ⁻¹⁶
θ_1	0.9662	0.0069	< 2.2e ⁻¹⁶
ω_0	0.0726	0.015	1.402e ⁻⁶
ω_1	1.471	0.0973	< 2.2e ⁻¹⁶
ω_2	1.4729	0.0948	< 2.2e ⁻¹⁶
ω_3	0.2699	0.0169	< 2.2e ⁻¹⁶

ω_0 : precipitation; ω_1 : wind speed at 2 m above ground; ω_2 : wind speed at 10 m above ground; ω_3 : atmospheric pressure.

Figures in bold indicate significance.

adjustments in the model's response, leading to sudden changes in the response to external shocks.

The model coefficients were transformed back to the original scale (Eqs. [12] to [15]). The suffix "tr" indicates that the coefficient is estimated on the original scale of the PM₁₀ data, measured in $\mu\text{g m}^{-3}$.

$$\omega_{0 \text{ tr}} = \left[(-4.999963)(-0.0726) + 1 \right]^{\frac{1}{-4.999963}} - 1 = -0.06 \quad (12)$$

$$\omega_{1 \text{ tr}} = \left[(-0.1796091)(-1.471) + 1 \right]^{\frac{1}{-0.1796091}} - 1 = -0.729 \quad (13)$$

$$\omega_{2 \text{ tr}} = 1 - \left[(2 + 0.1654505)(1.4729) + 1 \right]^{\frac{1}{2 + 0.1654505}} = -0.938 \quad (14)$$

$$\omega_{3 \text{ tr}} = \left[(-2.847771)(-0.2699) + 1 \right]^{\frac{1}{-2.847771}} - 1 = -0.18 \quad (15)$$

All the estimates for the effects of the variables are negative, suggesting that an increase in the values of these variables is associated with a reduction in PM₁₀ concentrations. It is estimated that each 1-unit increase in precipitation reduces, on average, the concentration of PM₁₀ by 0.064 $\mu\text{g m}^{-3}$, keeping all other variables constant.

A stronger wind near the surface (2 m, $\omega_{1 \text{ tr}}$) favors the dispersion of pollutants, reducing the concentration of PM₁₀. Each increase in wind speed at 2 m is associated with a decrease of PM₁₀ of 0.729 $\mu\text{g m}^{-3}$.

On the other hand, wind at 10 m ($\omega_{2 \text{ tr}}$) suggests that an increase in wind speed at this level correlates with a decrease of 0.938 $\mu\text{g m}^{-3}$ in PM₁₀, which could reflect a more complex dynamic of vertical transport or advection, similar to what was reported by Contreras-Arreola and González (1999), where it was shown that wind direction and persistence influence PM₁₀ concentrations, with more pronounced seasonal differences in summer than in winter.

Finally, the effect of atmospheric pressure ($\omega_{3 \text{ tr}}$) indicates that for each increase in atmospheric pressure, the concentration of PM₁₀ decreases by 0.18 $\mu\text{g m}^{-3}$, which may be related to meteorological conditions

that promote the dispersion or removal of particles in the atmosphere.

A detailed analysis of the model residuals is presented in Figure 8. The residuals do not exhibit clear patterns and remain randomly distributed, indicating that the model adequately captures the main dynamics of the series. The residual histogram suggests a degree of normality, while the absence of autocorrelation in the ACF plots confirms the independence of the errors, thereby validating the fitted model.

This analysis demonstrates that the fitted ARIMA model is robust and suitable for capturing temporal dynamics and interactions with meteorological variables (Cao, 2024), providing an effective tool for the analysis and prediction of PM_{10} .

The transfer function model (Eq. [10]) produced results in terms of confidence intervals comparable to those reported by Alvarado et al. (2010), as they capture a significant portion of the actual values, as shown in Figure 9. Although extreme events are not adequately captured, the model successfully describes the average behavior of the data.

A key advantage of this approach is its simplicity compared to complex methods like MARS, which

segment data and fit polynomials to model nonlinear relationships. The performance metrics show that the MSE was $1.0259 \mu\text{g m}^{-3}$, reflecting moderate variability in the model's errors, while the RMSE indicates that the predictions deviate, on average, $1.0129 \mu\text{g m}^{-3}$ from the actual values. The MAE, which measures the average absolute errors without being influenced by extreme values, is estimated at $0.8083 \mu\text{g m}^{-3}$. This metric provides a more stable and reliable interpretation of the model's average error magnitude, making it particularly useful in contexts where extreme values do not dominate the analysis (Table IX).

Regarding the model, it is suitable for capturing general trends and making average projections but requires additional adjustments to improve its performance in predicting extreme events, especially in contexts where these peaks have a significant impact (Fig. 10).

3.8 Outliers in the residuals of the models

The detection of outliers is based on the decomposition of the time series into trend, seasonality, and residual components, following Hyndman and

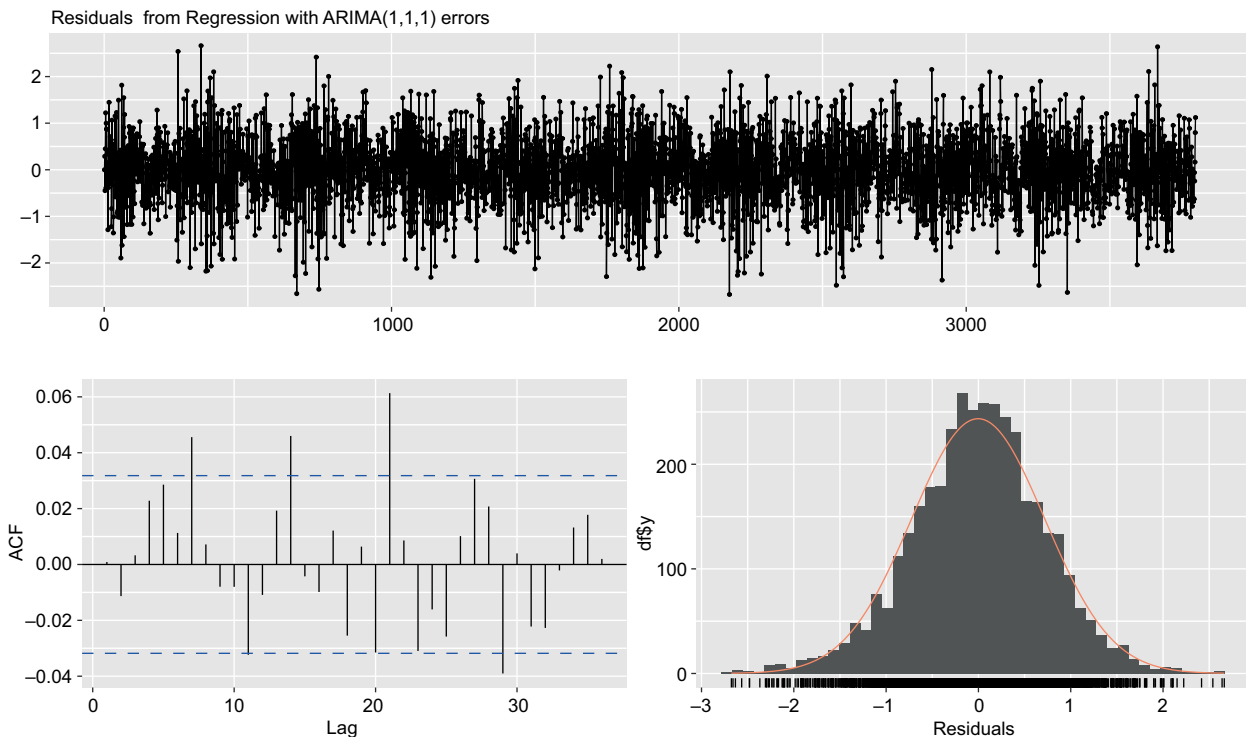


Fig. 8. Residual analysis of the ARIMA (1,1,1) (0,0,0) [365] model with transfer functions.

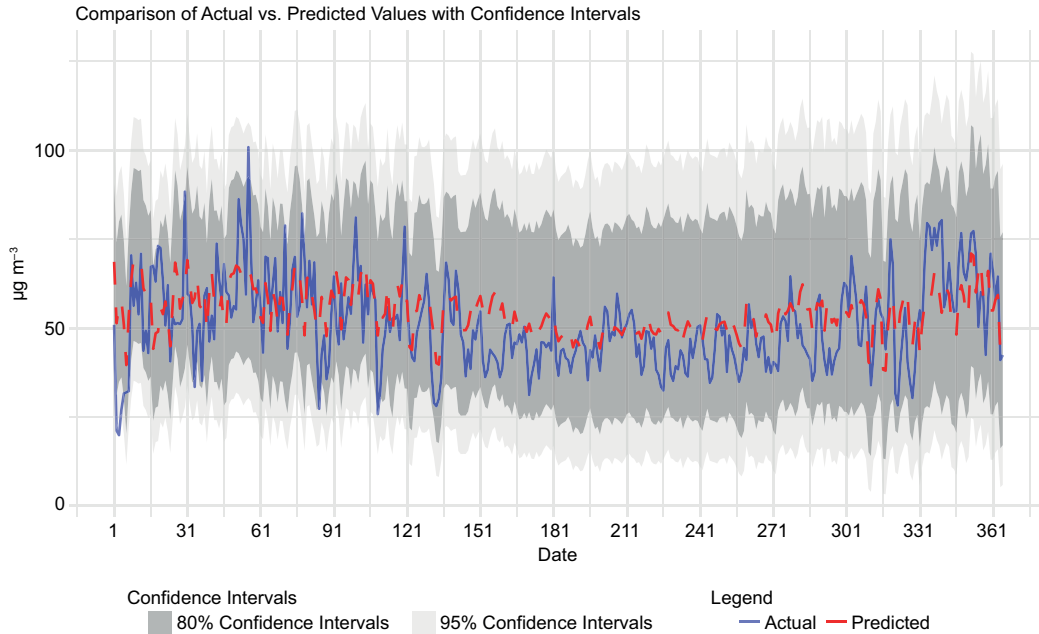


Fig. 9. Prediction on the test set with confidence intervals for PM₁₀ for the average data of the test portion.

Table IX. Performance metrics of the transfer function model.

Accuracy metrics	
AIC	8320.02
AICc	8320.05
BIC	8363.71
Ljung-box (p-value)	0.957
Augmented Dickey-Fuller (p-value)	> 0.01
RMSE	1.0129
MSE	1.0259
MAE	0.8083
MAPE	992.54

AIC: Akaike information criterion; AICc: Akaike information criterion corrected version; BIC: Bayesian information criterion; RMSE: root mean squared error; MSE: mean squared error; MAE: mean absolute error; MAPE: mean absolute percentage error.

Athanasopoulos (2021). The residuals of both the univariate model and the transfer function model show a low proportion of outliers, with 0.89 and 0.5%, respectively. However, the transfer function model exhibits fewer additive outliers (AO) and temporary

changes (TC), whereas the univariate model captures more level shifts (LS) (Table X). This suggests that the meteorological variables in the transfer function model influence the propagation of changes over time, while the univariate model better explains the individual behavior of the PM₁₀ pollutant (Fig. 11).

3.9 GARCH models

Two GARCH models were applied to the residuals of the ARIMA models: one univariate and one with transfer functions. In both cases, the skewed generalized error distribution (sGED) was used. The two models share similar characteristics in terms of asymmetry (0.9399 and 0.8640, respectively) and kurtosis (1.7341 and 1.7108, respectively), which are typical properties of the sGED distribution.

For the conditional variance of the residuals from the univariate ARIMA model, a GARCH (2,2) process was applied (Fig. 12). This means that two lags were used for both the conditional variance and the shock term, providing a more detailed fit to the fluctuations in the residuals' volatility, as described in Eq. (16).

$$\sigma_t^2 = \omega + \alpha_1 \varepsilon_{t-1}^2 + \alpha_2 \varepsilon_{t-2}^2 + \beta_1 \sigma_{t-1}^2 + \beta_2 \sigma_{t-2}^2 \quad (16)$$

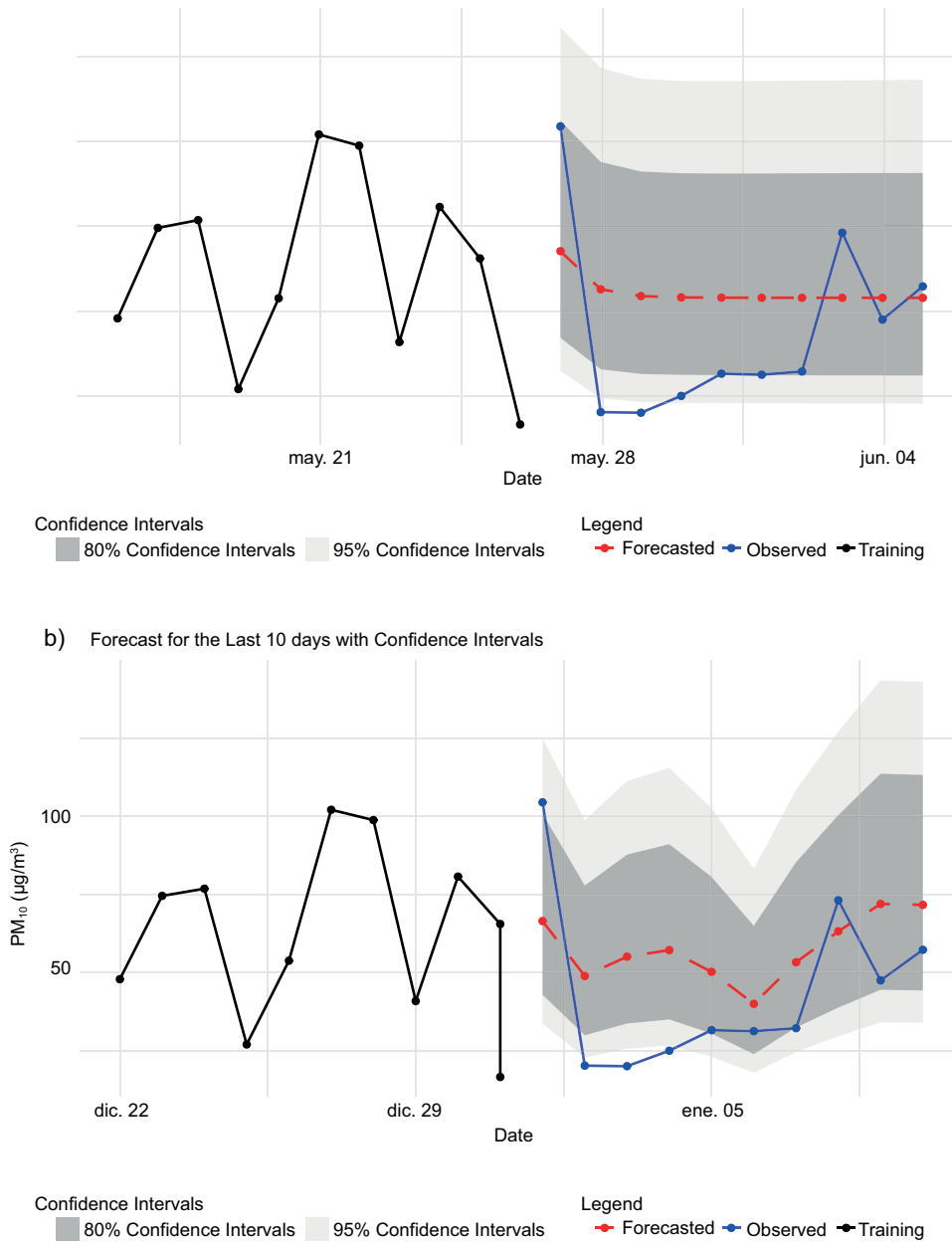


Fig. 10. Prediction for the last seven days in the test set with confidence intervals for PM₁₀: (a) univariate ARIMA; (b) ARIMA with transfer functions.

Table X. Types of outliers by model.

Model	AO	LS	TC
Univariate	17	10	7
Transfer	14	0	5

AO: additive outlier; LS: level shift; TC: temporary change.

For the transfer function model, the conditional variance follows a GARCH (1,2) process (Fig. 13). This means that one lag was used for the conditional variance and two lags for the shock terms, allowing for an efficient capture of fluctuations in the residuals' volatility, as described in equation (17).

$$\sigma_t^2 = \omega + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2 + \beta_2 \sigma_{t-2}^2 \quad (17)$$

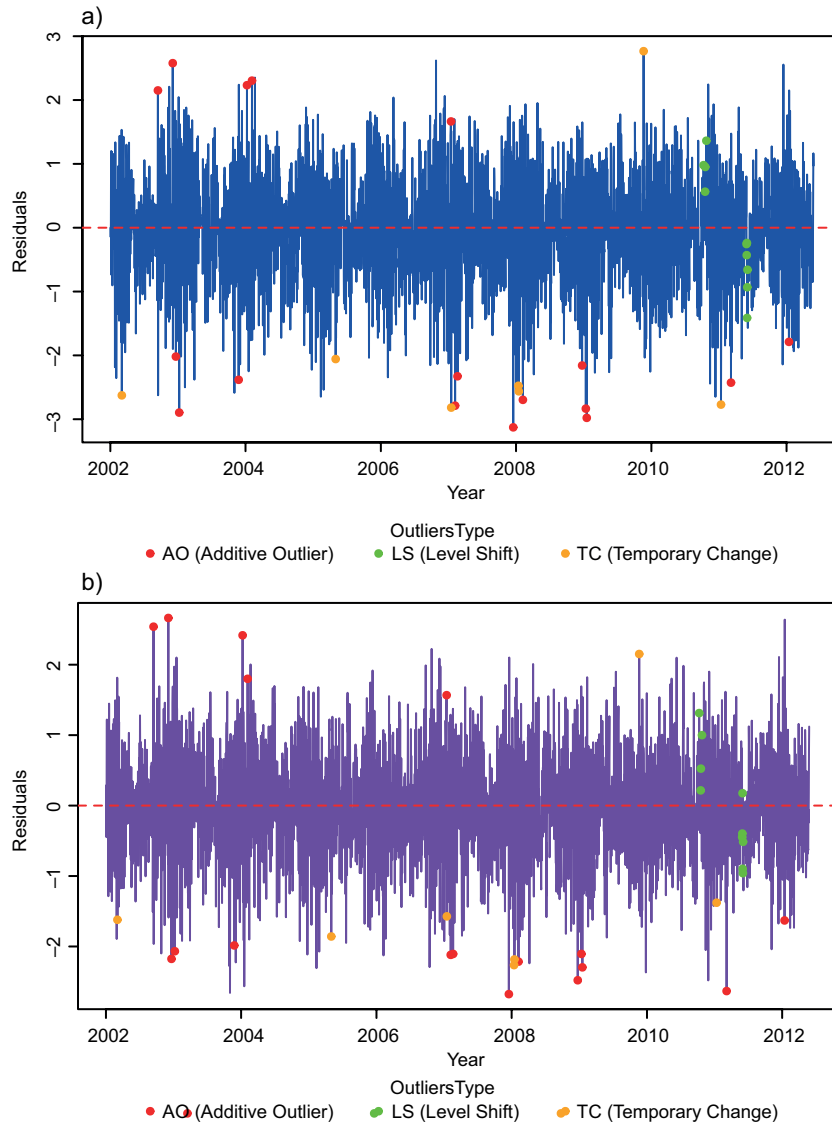


Fig. 11. Outliers in the residuals of the ARIMA models: (a) univariate ARIMA model, and (b) ARIMA model with transfer functions.

The estimated parameters for the GARCH (2,2) model are: $\alpha_1 \approx 0.037$, $\alpha_2 \approx 0.045$ and $\beta_1 + \beta_2 \approx 0.20 + 0.69 = 0.89$, which results in a total sum of $\alpha_1 + \alpha_2 + \beta_1 + \beta_2 \approx 0.97$. This value, close to 1, indicates very high persistence in volatility, suggesting that the effects of past shocks on the conditional variance diminish slowly over time.

Similarly, the GARCH (1,2) model associated with the transfer function model shows the following parameter estimates: $\alpha_1 \approx 0.058$ and $\beta_1 + \beta_2 \approx 0.44 + 0.47 = 0.91$, which results in a total persistence of α_1

+ $\beta_1 + \beta_2 \approx 0.97$. This high persistence implies that future volatility strongly depends on past volatility, maintaining its effect over extended periods.

The joint interpretation of both models indicates a high persistence in volatility over time, with past volatilities (β) having a more significant influence compared to recent shocks (α), which reflects a behavior where episodes of high or low volatility tend to persist, suggesting that changes in conditional variance are smoother and more sustained over time, rather than responding abruptly to new shocks. Such

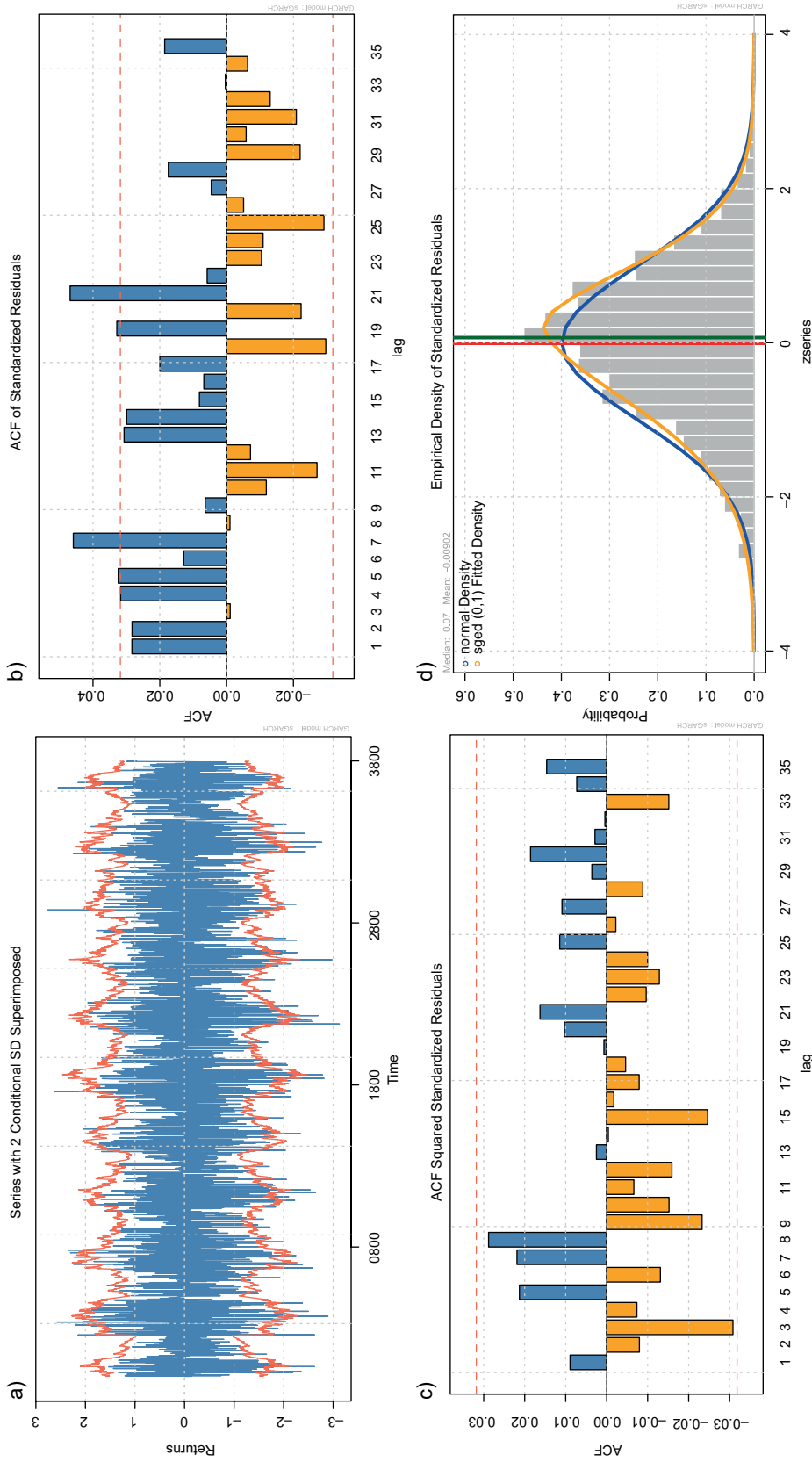


Fig. 12. GARCH (2,2): (a) series with two conditional standard deviations overlaid; (b) autocorrelation function of standardized residuals; (c) autocorrelation function of squared standardized residuals, and (d) empirical density of standardized residuals.

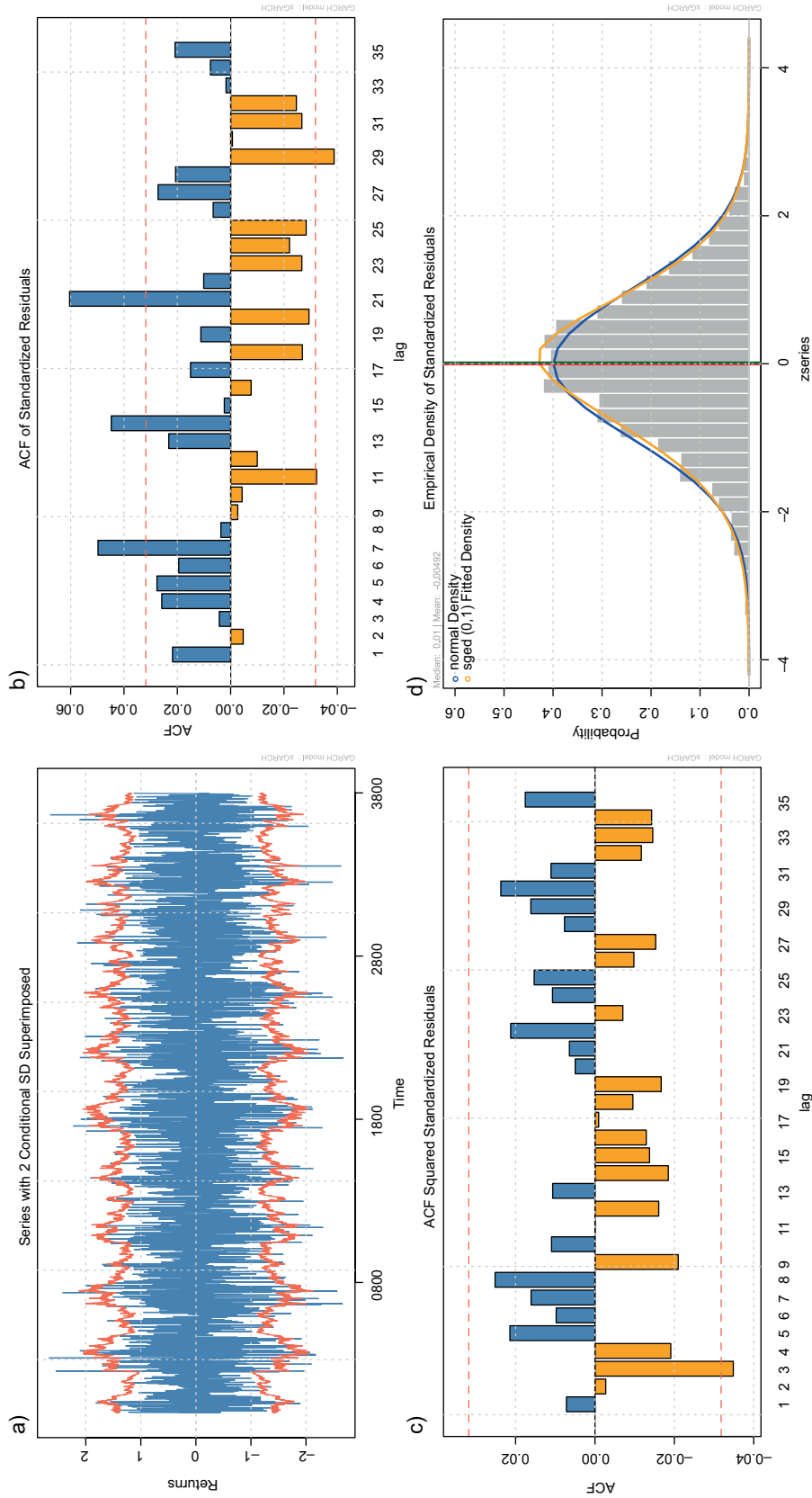


Fig. 13. GARCH (1,2): (a) series with two conditional standard deviations overlaid; (b) autocorrelation function of standardized residuals; (c) autocorrelation function of squared standardized residuals, and (d) empirical density of standardized residuals.

persistent fluctuations in volatility indicate changes in the conditional variance that may be driven by extreme environmental events as demonstrated in the work of Alexis et al. (2022), such as sudden storms or abrupt shifts in wind patterns, which can affect PM_{10} concentrations.

Another significant difference between the two models is that the residuals of the univariate ARIMA model depend more on the volatility associated with the magnitude of squared errors from two periods ago, reflecting a longer memory regarding past shocks. In contrast, the residuals of the transfer function model exhibit a relatively shorter memory concerning the impact of past shocks, indicating that the effects of recent errors dissipate more quickly. This behavior can be explained by the fact that, in the transfer function model, the mean effect has been captured more efficiently through meteorological variables, which better reflect the deterministic and exogenous dynamics of the system. As a result, the number of unexplained residuals is reduced, allowing for a more precise fit and decreasing the need for a complex conditional variance model.

On the contrary, in the GARCH(2,2) model associated with the univariate ARIMA, there is a higher proportion of unexplained residuals in the mean. This necessitates a more flexible and complex conditional variance model to capture the remaining fluctuations. The need for greater flexibility is reflected in the GARCH(2,2) structure, which allows for a more effective representation of prolonged fluctuations and the effects of past shocks on volatility. In both cases, the residuals of the ARIMA + transfer + GARCH model achieved normal residuals (Figs. 12 and 13, respectively).

The GARCH component captures conditional heteroscedasticity, periods of high or low volatility, associated with abrupt changes in external variables, modeling how the residual variance evolves over time in response to climatic or environmental phenomena. By adding transfer functions to the ARIMA + GARCH framework, exogenous variables are explicitly included in the equation, which not only explains PM_{10} dynamics based on its own history but also directly links volatility to factors such as precipitation or wind speed, attributing a portion of the observed volatility to these external drivers. In contrast, a purely ARFIMA + GARCH (or SARFIMA

+ GARCH) model captures long-memory behavior and fractional seasonality in the PM_{10} series, modeling volatility through GARCH (Reisen et al., 2014); however, without exogenous variables, it cannot distinguish how much of the variability in the mean or volatility is caused by specific external events like a precipitation spike or a strong wind episode.

4. Conclusion

The ARIMA (1,1,1)(0,0,0)[365] model is more effective in explaining the behavior of the PM_{10} variable rather than forecasting it. Additionally, the selected meteorological variables indeed have a significant impact on PM_{10} levels, confirming their influence on the dynamics of pollutant concentrations. While all meteorological variables included in the model were statistically significant, precipitation stood out due to its pronounced variability, reflected in recurrent episodes of intense rainfall and extended droughts characteristic of this region. Consequently, changes in precipitation behavior notably influence PM_{10} concentrations, highlighting its critical role compared to other statistically significant meteorological variables. Pollutant data show moderate noise due to their nature, with a few outliers that do not significantly affect the analysis. The high persistence in GARCH models suggests that volatility changes are driven more by past trends than recent events, indicating that pollutant dynamics are shaped by structural factors and long-term patterns, reinforcing the stability and predictability of the time series.

It is recommended to use the model for short-term forecasts with a maximum horizon of seven days, as no significant peaks in PM_{10} concentrations are observed within this period. However, caution should be exercised when interpreting the results, considering that the use of confidence intervals allows for defining a more precise range in which PM_{10} concentrations may fluctuate, providing a more robust assessment of the forecasts.

As a proposal for future work, it would be beneficial to increase the number of monitoring stations that record additional meteorological variables, such as temperature, relative humidity, and traffic flow, along with information on industrial activities. Expanding the set of variables would allow for a more precise validation of the impact of anthropogenic activities on PM_{10} concentrations.

Additionally, it is recommended to use more detailed monitoring frequencies and to address outlier treatment more rigorously at the hourly frequency. This approach could enhance the accuracy and reliability of predictive models.

Acknowledgments

The data were obtained from the NASA Langley Research Center (LaRC) POWER Project, funded through the NASA Earth Science/Applied Science Program. The first author thanks CONAHCyT for the financial support.

References

- Aguirre-López MA, Rodríguez-González MÁ, Soto-Villalobos R, Gómez-Sánchez LE, Benavides-Ríos ÁG, Benavides-Bravo FG, Walle-García O, Pámanes-Aguilar MG. 2022. Statistical analysis of PM₁₀ concentration in the Monterrey Metropolitan Area, Mexico (2010-2018). *Atmosphere* 13: 297. <https://doi.org/10.3390/atmos13020297>
- Alexis E, Plocoste T, Nuiro SP. 2022. Analysis of particulate matter (PM₁₀) behavior in the Caribbean area using a coupled SARIMA-GARCH model. *Atmosphere* 13: 862. <https://doi.org/10.3390/atmos13060862>
- Alvarado SA, Silva CS, Cáceres DD. 2010. Modelación de episodios críticos de contaminación por material particulado (PM₁₀) en Santiago de Chile. Comparación de la eficiencia predictiva de los modelos paramétricos y no paramétricos. *Gaceta Sanitaria* 24: 466-472. <https://doi.org/10.1016/j.gaceta.2010.07.008>
- Analitis A, Barratt B, Green D, Beddows A, Samoli E, Schwartz J, Katsouyanni K. 2020. Prediction of PM_{2.5} concentrations at the locations of monitoring sites measuring PM₁₀ and NO_x, using generalized additive models and machine learning methods: A case study in London. *Atmospheric Environment* 240: 117757. <https://doi.org/10.1016/j.atmosenv.2020.117757>
- Badran G, Ledoux F, Verdin A, Abbas I, Roumie M, Genevray P, Landkocz Y, Lo Guidice J-M, Garçon G, Courcot D. 2020. Toxicity of fine and quasi-ultrafine particles: Focus on the effects of organic extractable and non-extractable matter fractions. *Chemosphere* 243: 125440. <https://doi.org/10.1016/J.CHEMOSPHERE.2019.125440>
- Bai J, Perron P. 1998. Estimating and testing linear models with multiple structural changes. *Econometrica* 66: 47-78. <https://doi.org/10.2307/2998540>
- Bai J, Perron P. 2003. Computation and analysis of multiple structural change models. *Journal of Applied Econometrics* 18: 1-22. <https://doi.org/10.1002/jae.659>
- Box GEP, Jenkins GM, Reinsel GC, Ljung GM. 2015. *Time series analysis: Forecasting and control*. John Wiley & Sons.
- Buoli M, Grassi S, Caldiroli A, Carnevali GS, Mucci F, Iodice S, Cantone L, Pergoli L, Bollati V. 2018. Is there a link between air pollution and mental disorders? *Environment International* 118: 154-168. <https://doi.org/10.1016/j.envint.2018.05.044>
- Cao C. 2024. How to better predict the effect of urban traffic and weather on air pollution? Norwegian evidence from machine learning approaches. *Journal of Economic Behavior & Organization* 221: 544-569. <https://doi.org/10.1016/j.jebo.2024.03.018>
- Cárdenas-Moreno PR, Moreno-Torres LR, Lovallo M, Telesca L, Ramírez-Rojas A. 2021. Spectral, multifractal and informational analysis of PM₁₀ time series measured in Mexico City Metropolitan Area. *Physica A: Statistical Mechanics and its Applications* 565: 125545. <https://doi.org/10.1016/j.physa.2020.125545>
- Cong L, Chen R, Sera F, Vicedo-Cabrera AM, Guo Y, Tong S, Coelho MSZS, Saldiva PHN, Lavigne E, Matus P, Valdes Ortega N, Osorio Garcia S, Pascal M, Stafoggia M, Scortichini M, Hashizume M, Honda Y, Hurtado-Díaz M, Cruz J, Nunes B, Teixeira JP, Kim H, Tobias A, Íñiguez C, Forsberg B, Åström C, Ragettli MS, Guo Y-L, Chen B-Y, Bell ML, Wright CY, Scovronick N, Garland RM, Milojevic A, Kyselý J, Urban A, Orru H, Indermitte E, Jaakkola JJK, Rytí NRI, Katsouyanni K, Analitis A, Zanobetti A, Schwartz J, Chen J, Wu T, Cohen A, Gasparrini A, Kan H. 2019. Ambient particulate air pollution and daily mortality in 652 cities. *The New England Journal of Medicine* 381: 705-715. <https://doi.org/10.1056/NEJMoa1817364>
- Contreras-Arreola JL, González G. 1999. Análisis espectral del viento y de partículas menores de 10 micrómetros (PM₁₀) en el área metropolitana de Monterrey, México. *Revista Internacional de Contaminación Ambiental* 15: 95-102.
- Gasca-Sánchez FM, Santuario-Facio SK, Ortiz-López R, Rojas-Martínez A, Mejía-Velázquez GM, Garza-Pérez EM, Hernández-Hernández JA, López-Sánchez RC, Cardona-Huerta S, Santos-Guzmán J. 2021. Spatial

- interaction between breast cancer and environmental pollution in the Monterrey Metropolitan Area. *Heliyon* 7: e07915. <https://doi.org/10.1016/j.heliyon.2021.e07915>
- Gladkova E, Saychenko L. 2022. Applying machine learning techniques in air quality prediction. *Transportation Research Procedia* 63: 1999-2006. <https://doi.org/10.1016/j.trpro.2022.06.222>
- Granger CWJ. 1969. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica* 37: 424-438. <https://doi.org/10.2307/1912791>
- Hernández-Vega JI, González-Rodríguez M, Reyes-Varela E, Reynoso-Guajardo LA, Palomares-Gorham DG. 2021. Adaptación de algoritmo evolutivo mediante el método de evolución diferencial aplicado para problemas de la calidad del aire. *Revista ELECTRO* 43: 1-6.
- Huang J, Liu Q, Guo X. 2019. Short-term effects of particulate air pollution on human health. In: *Encyclopedia of environmental health* (Nriagu JO, Ed.). Elsevier, 655-662. <https://doi.org/10.1016/B978-0-12-409548-9.10991-1>
- Hyndman RJ, Athanasopoulos G. 2018. *Forecasting: Principles and practice*. 2nd ed. OTexts, Melbourne.
- Hyndman RJ, Athanasopoulos G. 2021. *Forecasting: Principles and practice*. 3rd ed. OTexts, Melbourne.
- INECC. 2019. *Calidad del aire. Equipos de medición*. Instituto Nacional de Ecología y Cambio Climático, Mexico. Available at: https://datos.gob.mx/dataset/calidad_aire_equipos_medicion (accessed on August 17, 2025).
- Leal-Iga J. 2019. Efectos físicos de la contaminación atmosférica percibidos de manera inconsciente por la ciudadanía, en el área metropolitana de la ciudad de Monterrey, Nuevo León, México. *Revista de Salud Pública* 21: 423-429. <https://doi.org/10.15446/rsap.V21n4.74959>
- Moritz S, Bartz-Beielstein T. 2017. imputeTS: Time series missing value imputation in R. *The R Journal* 9: 207-218. <https://doi.org/10.32614/RJ-2017-009>
- Myridakis A, Stephanou EG. 2025. 1.8 - Aerosols PM_{2.5} and PM₁₀. In: *Comprehensive sampling and sample preparation*, 2nd ed. (Soylak M, Ed.). Elsevier, 178-188. <https://doi.org/10.1016/B978-0-443-15978-7.00016-3>
- NASA. 2025. *Prediction of Worldwide Energy Resource (POWER)*. NASA POWER | Data Access Viewer (DAV). Available at: <https://power.larc.nasa.gov/data-access-viewer/> (accessed on August 17, 2025).
- Ordóñez-Iriarte JM. 2020. Salud mental y salud ambiental. Una visión prospectiva. Informe SESPAS 2020. *Gaceta Sanitaria* 34: 68-75. <https://doi.org/10.1016/j.gaceta.2020.05.007>
- Peterson RA. 2023. Yeo-Johnson normalization. In: *Package 'bestNormalize': Normalizing transformation functions 1.9.0*. 28-30. Comprehensive R Archive Network, CRAN Repository. Available at: <https://cran.r-project.org/web/packages/bestNormalize/bestNormalize.pdf> (accessed on May 14, 2024).
- Quintana-Belmares R, Hernández-Pérez G, Montiel-Dávalos A, Gustafsson Å, Miranda J, Rosas-Pérez I, López-Marure R, Alfaro-Moreno E. 2018. Urban particulate matter induces the expression of receptors for early and late adhesion molecules on human monocytes. *Environmental Research* 167: 283-291. <https://doi.org/10.1016/J.ENVRRES.2018.07.033>
- Raga GB, Baumgardner D, Castro T, Martínez-Arroyo A, Navarro-González R. 2001. Mexico City air quality: A qualitative review of gas and aerosol measurements (1960-2000). *Atmospheric Environment* 35: 4041-4058. [https://doi.org/10.1016/S1352-2310\(01\)00157-1](https://doi.org/10.1016/S1352-2310(01)00157-1)
- Paital B, Agrawal PK. 2022. Role of environmental factors in transmission of COVID-19. In: *COVID-19 in the environment: Impact, concerns, and management of coronavirus* (Rawtani D, Hussain CM, Khatri N, Eds.). Elsevier, 35-72. <https://doi.org/10.1016/B978-0-323-90272-4.00017-8>
- R Core Team. 2023. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna.
- Reisen VA, Sarnaglia AJQ, Reis NC, Lévy-Leduc C, Santos JM. 2014. Modeling and forecasting daily average PM₁₀ concentrations by a seasonal long-memory model with volatility. *Environmental Modelling & Software* 51: 286-295. <https://doi.org/10.1016/J.ENVSOF.2013.09.027>
- Renjie C, Peng Y, Xia M, Lijun W, Cong L, Yue N, Yunning L, Jiangmei L, Jinlei Q, Jinling Y, Haidong K, Maigen Z. 2025. Associations between coarse particulate matter air pollution and cause-specific mortality: A nationwide analysis in 272 Chinese cities. *Environmental Health Perspectives* 127: 017008. <https://doi.org/10.1289/EHP2711>
- Rodríguez Portal JA, Javier González-Barcalá F, Jorda RM, Martínez González C. 2011. El aire es nuestro:

- la importancia de mantener su calidad. *Archivos de Bronconeumología* 47: 23-26. [https://doi.org/10.1016/S0300-2896\(11\)70007-2](https://doi.org/10.1016/S0300-2896(11)70007-2)
- Scapini V, Torres S, Rubilar-Torrealba R. 2023. Meteorological, PM_{2.5} and PM₁₀ factors on SARS-COV-2 transmission: The case of southern regions in Chile. *Environmental Pollution* 322: 120961. <https://doi.org/10.1016/j.envpol.2022.120961>
- SSA. 2014. NORMA Oficial Mexicana NOM-025-SSA1-201. Salud ambiental. Valores límite permisibles para la concentración de partículas suspendidas PM₁₀ y PM_{2.5} en el aire ambiente y criterios para su evaluación. Secretaría de Salud, Mexico. *Diario Oficial de la Federación*, August 20.
- Silva C, Firinguetti L, Trier A. 1994. Contaminación ambiental por partículas en suspensión: modelamiento estadístico. *Actas XXI Jornadas Nacionales de Estadística*, Concepción, Chile.
- Ukaogo PO, Ewuzie U, Onwuka CV. 2020. Environmental pollution: Causes, effects, and the remedies. In: *Microorganisms for sustainable environment and health* (Chowdhary P, Raj A, Verma D, Akhter Y, Eds.). Elsevier, 419-429. <https://doi.org/10.1016/B978-0-12-819001-2.00021-8>
- Vallabani NVS, Gruziova O, Elihn K, Juárez-Facio AT, Steimer SS, Kuhn J, Silvergren S, Portugal J, Piña B, Olofsson U, Johansson C, Karlsson HL. 2023. Toxicity and health effects of ultrafine particles: Towards an understanding of the relative impacts of different transport modes. *Environmental Research* 231: 116186. <https://doi.org/10.1016/J.ENVRES.2023.116186>
- Villarreal-Macés SG, Díaz-Viera MA. 2018. Estimación geoestadística de la distribución espacial de la precipitación media mensual y anual en Nuevo León, México (1930-2014). *Tecnología y Ciencias del Agua* 9: 106-130. <https://doi.org/10.24850/j-tyca-2018-05-05>
- Villasenor R, Ortiz E, Watson J, Chow J. 2000. Spatial and temporal variations in ambient PM_{2.5} and PM₁₀ in Mexico City. *Journal of Aerosol Science* 31: 901-902. [https://doi.org/10.1016/S0021-8502\(00\)90911-X](https://doi.org/10.1016/S0021-8502(00)90911-X)
- Xuan VN. 2024. Determinants of environmental pollution: Evidence from Indonesia. *Journal of Open Innovation: Technology, Market, and Complexity* 10: 100386. <https://doi.org/10.1016/j.joitmc.2024.100386>
- Zeileis A, Kleiber C, Krämer W, Hornik K. 2003. Testing and dating of structural changes in practice. *Computational Statistics & Data Analysis* 44: 109-123. [https://doi.org/10.1016/S0167-9473\(03\)00030-6](https://doi.org/10.1016/S0167-9473(03)00030-6)