J. Ind. Geophys. Union, 30(2) (2026), 122-135
https://doi.org/10.71122/JIGU.30(2)2026.0013

A comparative study of stochastic nature of anthropogenic aerosol using SARIMA and LSTM methods

J. Arul Asir[1*], H. Johnson Jeyakumar[1] and C. P. Anil Kumar[2]
[1]P.G and Research Department of Physics, Pope's College, Manonmaniam Sundaranar University, Abishekapatti, Tirunelveli – 627012, Tamil Nadu, India.
[2]Equatorial Geophysical Research Laboratory, Indian Institute of Geomagnetism, Tirunelveli- 627011, Tamil Nadu, India.
[*]Corresponding author: arulasirphy@popescollege.edu.in

ABSTRACT

The troposphere is a dynamic region characterized by intricate interactions among natural and anthropogenic processes. Among these, aerosol loading, particularly fine particulate matter ($PM_{2.5}$), poses significant challenges in understanding the greenhouse effect and complicates air quality assessment due to its spatiotemporal variability. This study presents a comparative stochastic modelling framework for forecasting anthropogenic aerosol concentrations using both parametric and non-parametric approaches over Alandur (13.0° N, 80.10° E), Tamil Nadu, India. Specifically, a Seasonal Autoregressive Integrated Moving Average (SARIMA) model and a Long Short-Term Memory (LSTM) neural network, were employed to forecast $PM_{2.5}$ levels using observational data from 2023 and 2024. SARIMA, a classical statistical model, effectively captured regular patterns and seasonal trends. In contrast, the LSTM model, a deep learning approach, demonstrated enhanced performance in modelling non-linear dependencies and temporal dynamics. Model evaluations based on Root Mean Square Error (RMSE), indicated that LSTM consistently outperformed SARIMA, underlining its effectiveness in capturing the stochastic and complex behaviour of aerosol concentrations. These findings underscore the potential of integrating advanced machine learning techniques with traditional time series models to enhance the accuracy of air quality forecasting and environmental risk mitigation strategies.

Keywords: Aerosol, $PM_{2.5}$ levels, Alandur (Tamil Nadu), Health impact, Seasonal ARIMA, Machine learning, Neural network.

INTRODUCTION

Particulate matter (PM) in the atmosphere, originating from both natural and anthropogenic sources, consists of fine solid or liquid particles suspended in air. These aerosols are either directly emitted or formed via secondary chemical reactions of precursor gases. As a result, PM is a heterogeneous mixture with varied chemical compositions and size distributions, often exhibiting a core-shell morphology where secondary constituents condense around a primary core. Based on aerodynamic diameter, PM is commonly classified into fine particles ($PM_{2.5}$; ≤2.5 μm), ultrafine particles ($PM_{0.1}$; ≤0.1 μm), and nanoparticles ($PM_{0.05}$; ≤0.05 μm or ≤50 nm). This spatial-temporal distribution of PM concentration and composition is driven by complex interactions between emission sources, atmospheric transport mechanisms, in-situ chemical transformations, and removal processes such as dry and wet deposition. In addition to primary and secondary sources, background particulate matter representing PM levels in the absence of local anthropological influence, also contributes to the overall spatiotemporal variability (Yang et al., 2011: Celo and Dabek-Zlotorzynska, 2011; Seinfeld and Pandis, 2016; Guevara, 2016; Donahue et al., 2016). Alongside emission sources and transformation mechanisms, atmospheric conditions play a crucial role in modulating PM behaviour. Meteorological parameters such as wind speed, temperature, and relative humidity, significantly influence the dispersion, accumulation, and chemical transformation of aerosols (Zhang et al., 2018). Recent studies (Bhattarai et al., 2024; Pan et al., 2024; Yim et al., 2024), highlight the growing concern over fine particulate matter ($PM_{2.5}$) due to its combined impact on air quality degradation and climate alteration. With an atmospheric lifetime ranging from days to weeks, $PM_{2.5}$ is capable of long-range transport, leading to trans-boundary pollution and contributing to both regional and global environmental challenges. Meteorological influences and public health consequences are linked to PM exposure.

Andreae (2009) established a global correlation between CCN (Cloud Condensation Nuclei) concentrations and Aerosol Optical Thickness (AOT), showing that polluted marine and continental regions have nearly ten times higher CCN concentrations and about five times greater AOT than remote areas. This highlights the crucial role of aerosol optical properties in influencing cloud microphysics. Complementing these findings, Zhang et al. (2012) investigated CCN activation dynamics in Tianjin, China, using a thermal-gradient diffusion chamber (Droplet Measurement Technologies - CCNC). alongside aerosol size and composition analysis. Their findings confirmed that larger or more hygroscopic particles activate at lower supersaturation, while smaller or less hygroscopic particles, require higher supersaturation. Additionally, meteorological factors significantly influence $PM_{2.5}$ concentrations. Zhao et al. (2017), using the WRF-Chem model, found that increased anthropogenic aerosol loading in China, significantly enhanced cloud droplet formation and the liquefaction process in cumulus clouds. Following these results, Cheung et al. (2019) studied $PM_{2.5}$ over northern Taiwan from April 2017 to March 2018. In this study, continental air masses exhibited higher hygroscopicity (K-values), whereas local air masses showed higher Condensation Nuclei and CCN concentrations but lower K-values. These differences were reflected in the $PM_{2.5}$ chemical composition. The study also found that new particle formation enhanced

CCN levels through coagulation with smaller particles. Gao et al. (2023) further expanded on this by quantifying Aerosol Radiation Interactions (ARIs) and Aerosol Cloud Interactions (ACIs) over an eight-year period (2013–2021). Their results showed that ARI's contributed more substantially to $PM_{2.5}$ enhancement (5.59 $\mu g\, m^{-3}$) than ACI's (3.96 $\mu g\, m^{-3}$) during peak pollution episodes in eastern China. While aerosol composition determines particle activation and growth, meteorological parameters such as temperature and humidity, govern their dispersion and transformation.

Besides, Megaritis et al. (2014), employing a three-dimensional chemical transport model, quantified the seasonal sensitivity of $PM_{2.5}$ across Europe to variations in meteorological parameters, identifying temperature as the predominant influence year-round. Similarly, Tai et al. (2012) observed strong positive correlations between $PM_{2.5}$ and temperature across most of the United States, except for nitrate components in the southeast, which exhibited an inverse relationship. Principal component analysis further revealed that key meteorological modes explained 20–40% of the daily variability in $PM_{2.5}$ concentrations. Similarly, long term environmental changes, such as climate variability and land cover alterations, also influence $PM_{2.5}$ levels. Fu et al. (2016) used global chemical transport models to quantify these impacts over East Asia, finding that climate change can cause significant seasonal shifts in $PM_{2.5}$ concentrations, with reductions in winter and increases in summer. Although land cover and land use changes contributed to regional variability, anthropogenic emissions remained the primary driver of air quality deterioration. Farahat and Abuelgasim (2022) analyzed PM_{10} and $PM_{2.5}$ variations during cloud seeding missions in the United Arab Emirates. Their findings indicated that $PM_{2.5}$ levels were most elevated in industrial areas and were significantly influenced by cloud seeding, whereas urban PM_{10} concentrations remained higher but less sensitive to such interventions. In a similar manner**,** Singh et al. (2021) analyzed $PM_{2.5}$ trends across five Indian megacities Delhi, Kolkata, Mumbai, Hyderabad, and Chennai alongside the U.S. data over a six-year period (2014–2019). Delhi emerged as the most polluted city, followed by Kolkata and Mumbai. Expanding on the urban pollution patterns, Gu et al. (2024) also evaluated the health impacts of $PM_{2.5}$ across Southeast Asia using air quality modeling. They estimated approximately 900,000 premature deaths annually, with 87% linked to $PM_{2.5}$ exposure. Of these, 77% were due to local emissions primarily from industrial (45%) and residential (17%) sectors, while 23% were attributed to transboundary pollution. Time series forecasting can play a vital role in environmental monitoring by enabling predictive insights into complex and temporally structured phenomena.

Owing to the inherent stochasticity of atmospheric processes, model selection must align with the data's structural characteristics and forecasting objectives (Chatfield, 1975; Gautam and Singh, 2020). Previously, the power spectral analysis using a Blackman Tukey window. revealed dominant biennial and seasonal periodicities in $PM_{2.5}$ concentrations over Alandur's industrial corridor during pre- and post- pandemic periods. Further, wavelet-based phase coherence, confirmed the persistence of these periodic components, highlighting strong meteorological influences (Asir et al., 2024). In this study, we compare two forecasting frameworks Seasonal Autoregressive Integrated Moving Average (SARIMA) and Long Short-Term Memory (LSTM) networks, representing parametric and non-parametric approaches, respectively. SARIMA, grounded in linear stochtic modeling, is well-suited for seasonal stationarity, while LSTM networks, with their memory-cell architecture, effectively capture nonlinear and dependencies.

OBSERVATIONAL SITE AND AQUIRED DATA

The observational site selected for this study is Alandur (13.0° N, 80.10° E), located in the southern part of Chennai, Tamil Nadu (India). Known for its dense population, heavy vehicular traffic, and industrial zones, Alandur presents a complex urban environment encompassing residential, commercial, and transportation sectors. This diverse setting makes it a suitable candidate for evaluating aerosol loading and forecasting in mixed-use urban areas. $PM_{2.5}$ concentration data were sourced from the Central Pollution Control Board's (CPCB) Continuous Ambient Air Quality Monitoring Stations (CAAQMS) via the portal (https://airquality.cpcb.gov.in), covering the years 2023 and 2024, with 730 daily records in total. Due to routine maintenance, the 2023 dataset had 45 missing values, addressed using a 21-day rolling mean imputation. For the 12 missing values in 2024, a 7-day moving average was applied. Each year was treated as a separate timeline, while this approach helped preserve temporal continuity, it has smoothed out certain seasonal components in 2023; this limitation was considered during SARIMA model fitting. Outliers were detected using the Inter quartile Range (IQR) method, identifying values outside the Q_1–Q_3 range. The outlier rates were 1.37% for 2023 and 2.46% for 2024. Instead of removal, these values were normalized using the Z-score method to retain extreme event information while minimizing their impact on model training**.** A sensitivity analysis was also performed to ensure that the results were not biased by the chosen window sizes and their results are given in Table 1.

Table 1. Sensitivity analysis: Actual vs imputed $PM_{2.5}$ data (2023 and 2024)

Year	Imputation method	Statistic	Actual ($PM_{2.5}$)	Imputed ($PM_{2.5}$)	% Change (actual vs imputed)
2023	21-Day window	Mean	45.244	44.947	−0.65
		Median	40.660	40.070	−1.45
		Mode	28.145	26.161	−7.05
		Std dev	24.016	24.027	0.05
		Julian days	365	365	0.83
2024	7-Day window	Mean	45.114	45.096	−0.04
		Median	29.010	29.010	0.00
		Mode	13.620	13.620	0.00
		Std dev	35.396	35.355	−0.12
		Julian Days	366	366	0.00

METHODS

In time series analysis, data collected over time have temporal dependencies and non-stationarity commonly observed in real world (Shumway and Stoffer, 2011). The Autocorrelation Function (ACF) quantifies correlations between observations at different lags, aiding pattern detection and model development. Robust time series models integrate deterministic trends with stochastic components to capture both systematic and random variations (Kirchgässner and Wolters, 2007). These methods are essential in environmental and physical sciences, where complex temporal dynamics are common**.** A crucial prerequisite in this analysis is determining the stationarity of the time series, as it ensures consistent statistical properties over time and underpins the development of accurate models and reliable forecasts. In many practical systems, stationarity is maintained through regular control actions or continuous component maintenance, which stabilize underlying processes and prevent structural changes**.** A time series is called stationary, if the joint probability distribution of any n observations, $\{z_{t+1}, z_{t+2}, \ldots, z_{t+n}\}$ (i.e., realizations of the time series at times t + 1, t + 2, . . . , t + n) of this series remain the same as another set of n observations shifted by k time units, that is, $\{z_{t+1+k}, z_{t+2+k}, \ldots, z_{t+n+k}\}$ for practical purposes (Bisgaard and kulahci, 2011).

Extending the concept of stationarity, the autocorrelation function (ACF) and the Partial Autocorrelation Function (PACF), provide critical insights into the underlying structure of a time series. When a series exhibits a trend, the ACF typically shows large positive autocorrelations at small lags, reflecting the similarity of adjacent observations and resulting in a slow decay over increasing lags. Seasonality introduces distinct periodic peaks in the ACF at multiples of the seasonal frequency, signalling regular cyclical patterns. The coexistence of trend and seasonality, produces an ACF characterized by both gradual decline and periodic spikes. Conversely, a lack of significant autocorrelation across all lags indicates a purely random process, commonly termed white noise (Hyndman and Athanasopoulos, 2014), as provided in Eq.1.

Autocorrelation function is expressed as $r_k = \frac{\sum_{t=k+1}^{T}(y_t-\bar{y})(y_{t-k}-\bar{y})}{\sum_{t=1}^{T}(y_t-\bar{y})^2}$ (1)

where, r_k is the autocorrelation at lag k, y_t is the value of the time series at time (t), y_{t-k} Lagged value of the time series k steps before time (t), $\bar{y}$ is the mean of the time series value

Like the ACF, PACF depends solely on the second order properties of the process. The partial autocorrelation at lag k, denoted as α_K, represents the correlation between X_1 and X_{k+1}, adjusted for the influence of intervening observations $X_2,\ldots,X_K$ (Brockwell and Davis, 2016). PACF quantifies the direct relationship between observations at different lags, by removing the effects of intermediate lags. This makes the PACF a valuable tool for identifying which lagged terms have a meaningful and independent impact on the current observation. Negative partial correlation is often referred to as the reflection coefficient which lie within the range of –1 and 1 (Broerson, 2006; Hyndman and Athanasopoulos, 2014) as in Eq. 2.

The partial autocorrelation function is expressed as $\phi_{n,n} = \frac{\rho_n - \sum_{k=1}^{n-1}\phi_{n-1,k}\rho(n-k)}{1-\sum_{k=1}^{n-1}\phi_{n-1,k}\rho(k)}$ (2)

where, $\phi_{n,n}$ is the partial autocorrelation at lag n, ρ_n is the autocorrelation at lag n, ϕ_{n-1} Partial autocorrelation at lag k from the previous step (lag n−1), $k\rho(n-k)$ autocorrelation at lag n−k

However, visual inspection of the ACF and PACF alone is insufficient to conclusively determine stationarity. Unit root tests are applied to statistically assess the need for differencing. Among these, the Augmented Dickey-Fuller (ADF) test is widely used, extending the Dickey-Fuller test by including lagged differences to account for autocorrelation. The ADF models the first-order difference of the series as a regression on lagged values and lagged differences up to m lags, providing a formal test for non-stationarity (Cromwell et al., 1994; Pal and Prakash, 2017). The ADF Test is given by Eq. 3

$$\Delta x(t) = \alpha_1 x(t-1) + \Sigma_J \beta_j \Delta x(t-j) + e(t) \quad (3)$$

where, $\Delta x(t)$ represents the change in the variable x at time t. $\alpha_1 x(t-1)$ represents the current change in $x(t)$ which is influenced by its previous value $x(t-1)$, with $\alpha_1$as the coefficient, $\beta_j \Delta x(t-j)$ represents the influence of past changes in x(t) on the current change, β_j are the coefficients for each lag j, and $e(t)$ is the error term or noise

Parametric and non-parametric methods

Due to limited observational data, multiple stochastic processes can explain the same time series. Model selection begins by defining candidates and applying estimation or inference. Parametric models, like ARIMA and SARIMA, have fixed parameters, while nonparametric models, such as LSTM networks or power spectral analysis, involve flexible or infinite parameters (Fan and Yao, 2005). Although the distinction can be blur, it guides estimation methods. Empirical evidence is mixed: parametric models offer efficiency and interpretability, while nonparametric methods provide flexibility for complex patterns (Gautam and Singh, 2020). Model choice should align with data characteristics and forecasting goals.

ARIMA and SARIMA

The Autoregressive (AR) model was introduced by Yule (1927) and expanded by Walker (1931), while Slutzky (1937) developed the Moving Average (MA) model. Wold (1938) laid the theoretical groundwork for the combined ARMA framework. Box and Jenkins (1976) integrated these components into the Autoregressive Integrated Moving Average (ARIMA) model, designed for effective modeling and forecasting of both stationary and non-stationary time series. The ARIMA model comprises three parts: autoregressive (ar), which relates observations to their past values; integrated (I), which applies differencing to achieve stationarity; and moving average (ma), which models dependence on past forecast errors.

For seasonal data, SARIMA extends ARIMA by incorporating seasonal differencing and seasonal AR and MA terms. These models are widely applied in environmental and risk analysis for capturing complex temporal dynamics (Yule, 1927; Walker, 1931; Slutzky, 1937; Wold, 1938; Box and Jenkins, 1976; Kumar and Jain, 2010; Hyndman and Athanasopoulos, 2014; Pal and Prakash, 2017; Naveen and Anu, 2017; Kaur et al., 2023). ARIMA (p, d, q) model is expressed as Eq. 4.

$$Y_t = C + \sum_{i=1}^{p} \phi_t y_{t-i} + \sum_{J=1}^{q} \theta_j \varepsilon_{t-j} + \varepsilon_t \quad (4)$$

where, c - is the constant term, ϕ_t - is the coefficient of AR, θ_j - is the coefficient of MA

The SARIMA model is typically denoted as SARIMA (p, d, q) (P, D, Q) $_s$, where S is the length of the seasonal cycle (for example, 12 for monthly data with yearly seasonality). The general form of the SARIMA (p, d, q) (P, D, Q)$_s$ is expressed as Eq. 5:

$$Y_t = C + \sum_{i=1}^{p} \phi_t y_{t-i} + \sum_{I=1}^{P} \boldsymbol{\phi_t} Y_{t-IS} + \sum_{J=1}^{q} \theta_j \varepsilon_{t-j} + \sum_{J=1}^{Q} \boldsymbol{\theta_J} \varepsilon_{t-JS} + \varepsilon_t \quad (5)$$

where, c - is the constant term, ϕ_t - is the coefficient of AR, θ_j - is the coefficient of ma, $\boldsymbol{\phi_t}$ - is the coefficient of seasonal AR, $\boldsymbol{\theta_j}$ - is the coefficient of ma. The differencing equation is represented as Integration I_D (Differencing equation) $y'_t = Y_t - Y_{t-s}$

In SARIMA modeling, the seasonal order P, D, Q, and s reflect recurring patterns in time series data. However, this seasonal structure can vary from year to year due to external factors such as climate variations or changes in emission sources. These factors can alter the strength, timing, or frequency of seasonal peaks and valleys. SARIMA addresses this by iteratively adjusting seasonal parameters during model fitting, enabling it to capture both stable seasonal patterns and annual fluctuations. Hence, the seasonal order varies from year to year, and SARIMA models capture this by adjusting the seasonal parameters, ensuring the model adapts to evolving patterns while maintaining predictive accuracy.

Long Short – Term Memory (LSTM) Neural Network Analysis

The increasing availability of computational power and large-scale data has greatly advanced the effectiveness of modeling techniques in uncovering complex and hidden patterns in time series data. Among these, neural networks have emerged as powerful tools due to their ability to learn intrinsic relationships directly from raw inputs. Recurrent Neural Networks (RNNs) are particularly well-suited for modeling sequential data, as they use the current input and previous hidden state to capture temporal dependencies. However, traditional RNNs often suffer from vanishing or exploding gradient problems, which limit their ability to learn long-term dependencies. To address this, Long Short-Term Memory (LSTM) networks a specialized form of RNNs were developed (Figure 1). LSTMs introduce a cell state that carries information across time steps and use a gating mechanism to regulate the flow of information. Each LSTM cell contains three gates: the forget gate, which removes irrelevant past information; the input gate, which decides what new information to store; and the output gate, which determines the information to pass forward. These gates, governed by sigmoid activation functions and element-wise operations, allow

LSTMs to selectively retain relevant information and discard noise. This architecture enables LSTMs to model long-term, non-linear dependencies more effectively than standard RNNs, reducing overfitting and computational complexity. Due to these advantages, LSTMs are widely applied in fields such as weather forecasting, anomaly detection, and environmental time series modeling, where long-term memory and adaptability are crucial (Rebala et al., 2019; Yu et al., 2021; Luo and Gong, 2023).

Model evaluation metrics MSE and RMSE

MSE involves squaring the errors, its value is always non-negative, ranging from zero to infinity. Furthermore, MSE increases exponentially as the magnitude of errors increases. Therefore, a model exhibiting a lower MSE value, ideally approaching zero, is considered to have superior predictive accuracy**.** The RMSE is the square root of the mean squared error (MSE). Taking the root does not affect the relative ranks of models, but it yields a metric with the same units X (Hudson, 2022). The basic equation can be expressed as,

$$\text{MSE} = {}^{1}/_{n} \sum_{I=1}^{n} (y_i - \hat{y}_i)^2 \qquad (6)$$

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{I=1}^{n} (y_i - \hat{y}_i)^2} \qquad (7)$$

where, y_i is the actual value, $\hat{y}_i$ is the predicted value, n is the number of observations.

RESULTS AND DISCUSSION

Figure 2(a) and (b) illustrate the time series of daily $PM_{2.5}$ concentrations for 2023 and 2024, with the x-axis representing months and the y-axis indicating concentration intensity. In 2023, the dry winter months (January–February) exhibited a sharp increase in $PM_{2.5}$ levels, peaking in early January. In contrast, 2024 showed milder winter fluctuations, with peaks occurring later in January. Summer and post-monsoon months (March–June), revealed moderate variations, with generally higher concentrations in 2023 compared to 2024. During the southwest monsoon (July–August), 2024 recorded abrupt fluctuations, while 2023 displayed more stable trends. These seasonal variations are consistent with findings by Chen et al. (2020), who reported similar patterns across Delhi, Hyderabad, Chennai, and Mumbai (2015–2018), with peak $PM_{2.5}$ concentrations in winter and minimum levels during the monsoon due to effective wet scavenging. Additionally, intercity differences were attributed to variations in wind patterns and $PM_{2.5}$ transport dynamics.

The SARIMA model was implemented in Python - Jupyter Lab environment to forecast and analyse the stochastic behaviour of $PM_{2.5}$ concentrations. Effective time series forecasting requires modelling future values as functions of past observations, with particular emphasis on capturing trend and seasonality, two fundamental components shaping temporal patterns (Peixeiro, 2022). Accurate identification of these components is essential for informed model selection and reliable forecasting. To analyse the temporal dependence in the $PM_{2.5}$ series, the autocorrelation function (ACF) and partial autocorrelation function (PACF) were applied to identify significant lag structures and guide model selection.

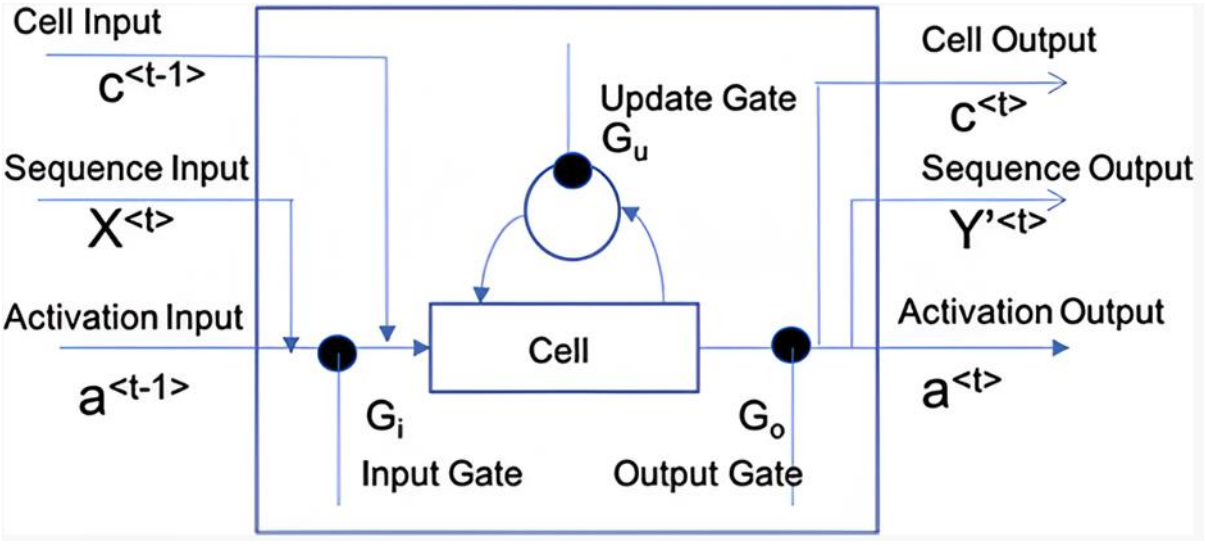


Figure 1. LSTM architecture showing how an LSTM uses its Input Gate (G_i), Update/Forget Gate (G_u), and Output Gate (G_o) to control what information is added, retained, or passed forward at each time step. By combining the current input with the previous cell and hidden states, the LSTM produces an updated cell state and activation output.

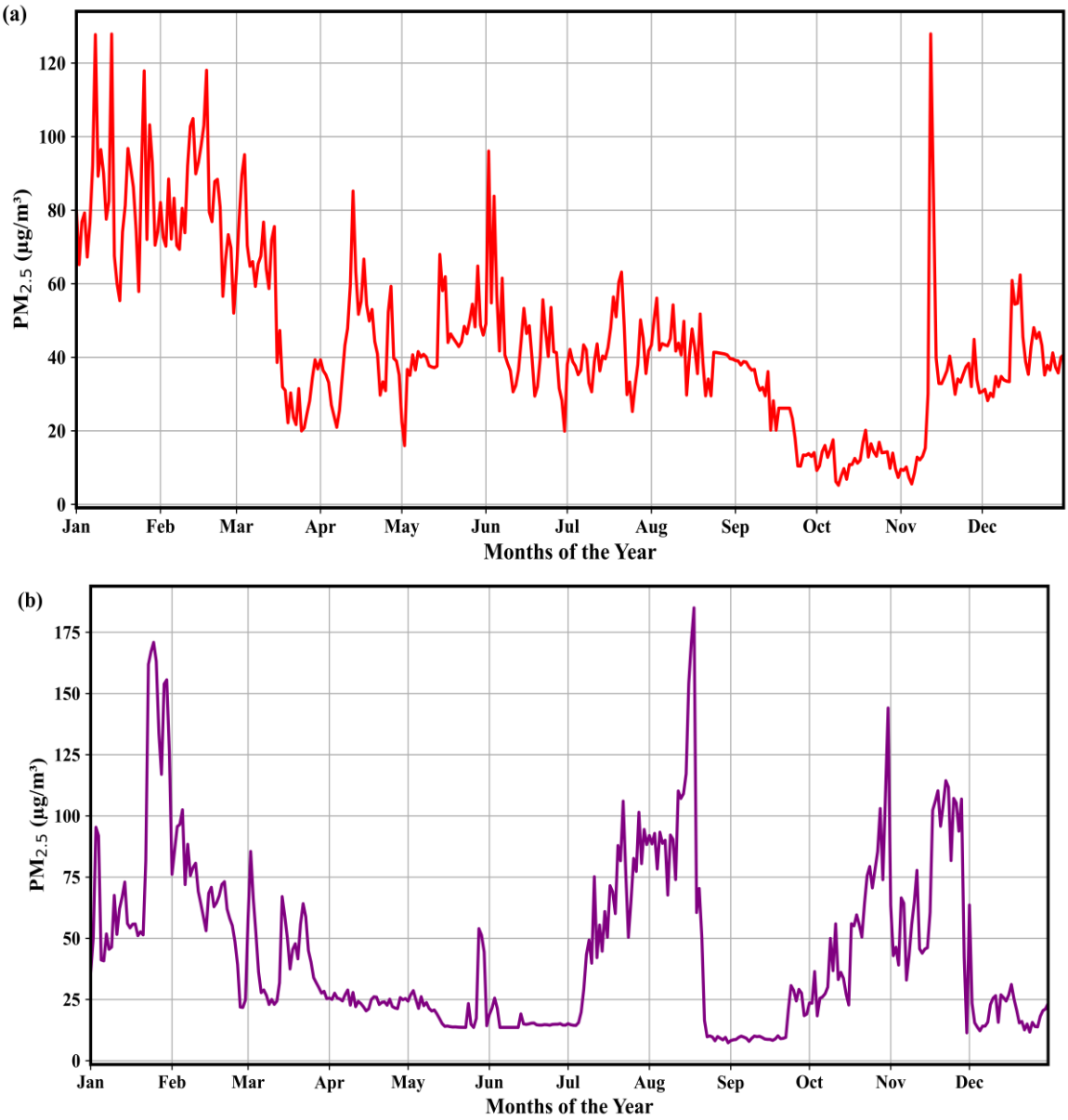


Figure 2. Time series analyses of particulate matter ($PM_{2.5}$) concentrations for the years, **(a)** 2023 and **(b)** 2024

The ACF captures both direct and indirect correlations of a time series with its lagged values, thereby offering insights into persistence and stationarity. A slow decay of the ACF beyond the confidence bounds, as observed for the 2023 dataset (Figure 3a), is indicative of a pronounced trend and potential seasonal effects, consistent with non-stationary dynamics (Kumar and Jain, 2010). In contrast, the PACF designed to isolate the direct effect of each lag by controlling for the intervening lags, revealed significant initial spikes followed by minor fluctuations (Figure 3b), suggesting an autoregressive structure embedded within a non-stationary process (Hyndman and Athanasopoulos, 2014). For 2024, the ACF and the PACF a comparable structure with marginal variations was observed (not presented here due to brevity). To supplement the visual diagnostics, the Augmented Dickey-Fuller (ADF) test was applied to statistically assess the stationarity of both time series. The results returned p-values exceeding the 0.05 threshold, indicating the presence of a unit root in both cases, and thereby confirming non-stationarity at a 95% confidence level. These findings substantiate the necessity for appropriate differencing and seasonal adjustment in model specification to accommodate the intrinsic stochasticity and temporal dependencies characteristic of urban $PM_{2.5}$ concentrations.

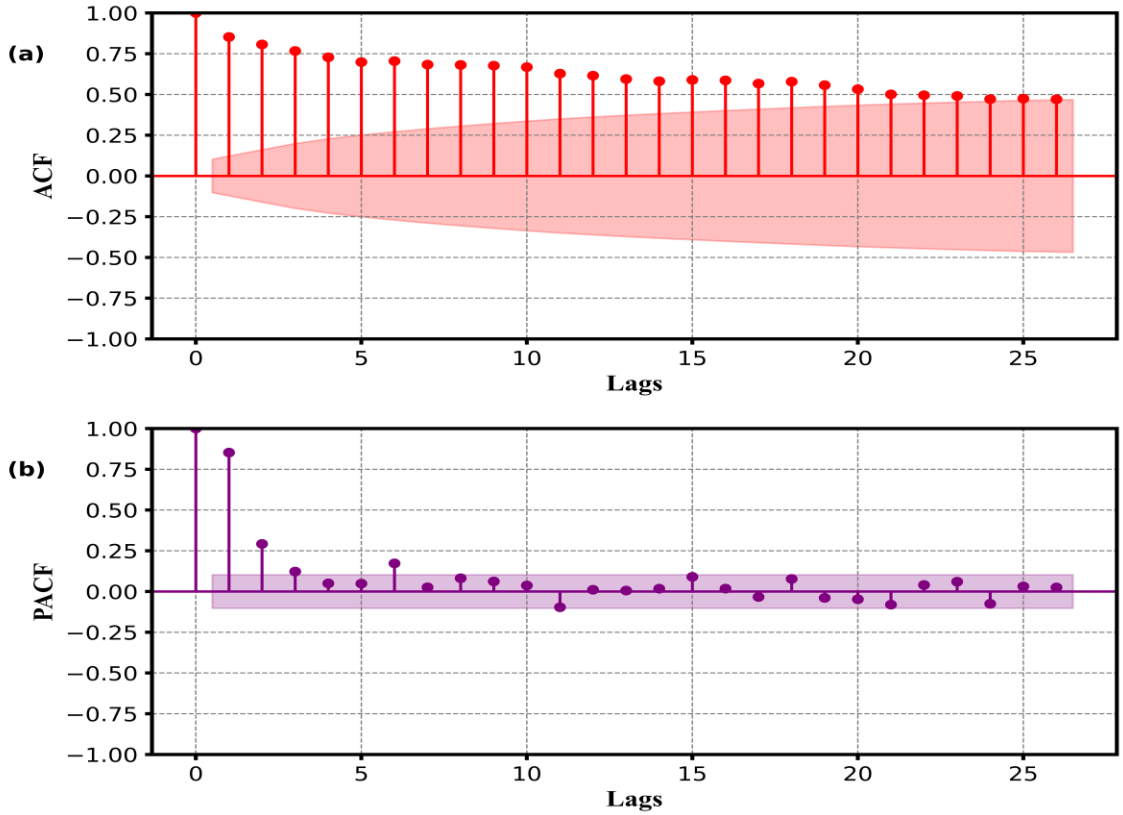


Figure 3. **(a)** ACF and **(b)** PACF plot for the year 2023

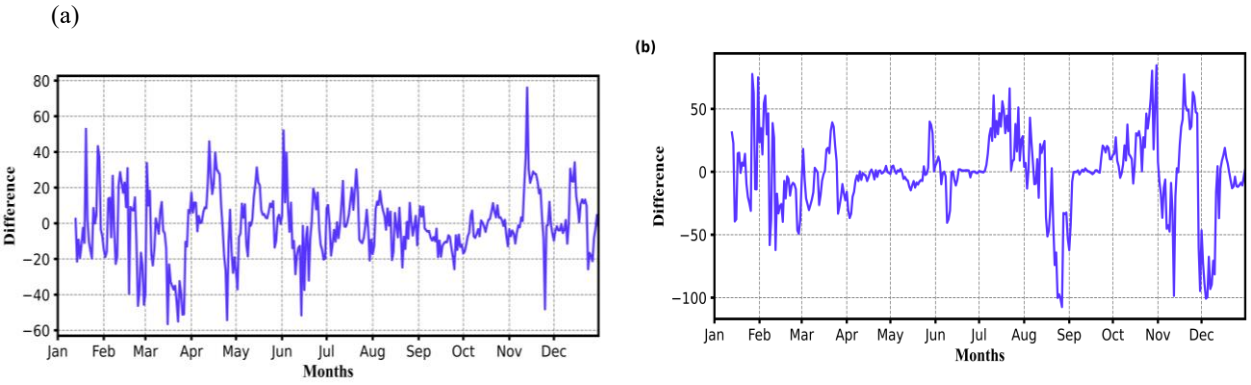


Figure 4. Seasonal first difference of $PM_{2.5}$ plots for the year **(a)** 2023, and **(b)** 2024

Figure 4 (a) and 4 (b**)** illustrate the seasonally differenced $PM_{2.5}$ time series for 2023 and 2024, respectively. A seasonal differencing lag of 12 was applied to capture annual periodicity, resulting in the exclusion of the first 12 observations in each series. Post-differencing, the Augmented Dickey-Fuller (ADF) test yielded statistically significant p-values of 0.0029 (2023) and 0.021 (2024), allowing for rejection of the null hypothesis of a unit root and confirming that both differenced series are stationary.

Following stationarity confirmation, the autocorrelation (ACF) and partial autocorrelation (PACF) plots of the differenced series were examined to identify model parameters, as shown in Figure 5 (a) and 5 (b). For 2023 (Figure 5a), the ACF dropped below the 95% confidence interval by lag 5 and approached zero by lag 10. A slight negative autocorrelation was observed at lag 11. The corresponding PACF (Figure 5b) showed a peak at lag 2 near the confidence limit, decaying to zero by lag 3, with a negative spike at lag 11.

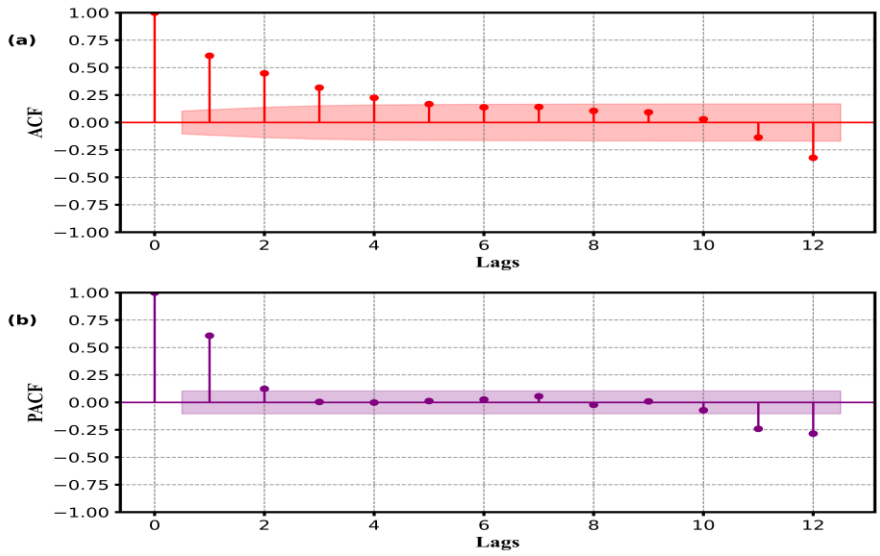


Figure 5. Seasonal differenced plot of **(a)** ACF, and **(b)** PACF for the year 2023

Table 2. AIC, BIC, and HQIC values for various fitted models for the year 2023 and 2024

SARIMA MODEL order 2023	AIC	BIC	HQIC
$(0,0,0)\ (2,1,4)_{11}$	2594.63	2620.768	2605.08
$(1,0,1)\ (2,1,4)_{11}$	2290.35	2323.92	2303.77
$(1,1,1)\ (2,1,4)_{11}$	2280.65	2314.10	2294.06
$(1,1,1)\ (2,1,5)_{11}$	2196.54	2233.44	2211.31
SARIMA MODEL order 2024	AIC	BIC	HQIC
$(0,0,0)\ (3,1,7)_{12}$	2530.98	2570.52	2546.86
$(1,1,1)\ (3,1,6)_{12}$	2294.83	2338.40	2312.31
$(1,0,1)\ (3,1,6)_{12}$	2293.51	2337.12	2311.00

Model order selection for 2023 and 2024

The lag at which autocorrelation (ACF) and partial autocorrelation (PACF) values fall within the 95% confidence interval indicating statistical insignificance, is commonly used to estimate the orders of the moving average MA(q) and autoregressive AR(p) components, respectively. In the presence of pronounced seasonality, the ACF of the seasonally differenced series often shows significant spikes at seasonal lags (e.g., lag 12), implying a stronger seasonal MA component. Conversely, the PACF generally exhibits weaker or fewer seasonal spikes, suggesting a lower seasonal AR order (P), compared to the seasonal MA order (Q). Appropriate model order selection is critical to balance complexity and performance. Underfitting (model under-specification) can miss essential patterns, while overfitting (model over-specification), may capture noise instead of the underlying signal. Thus, ACF and PACF diagnostics serve as key initial tools in guiding SARIMA model specification. To refine model selection, the Akaike Information Criterion (AIC) is employed to evaluate multiple candidate models, with preference given to the model exhibiting the lowest AIC value. This indicates an optimal trade-off between goodness-of-fit and model parsimony. Table 2 presents the Akaike Information Criterion, Bayesian Information Criterion, and HQIC values for several models, from which the optimal configuration was chosen based on minimum information criteria values. The SARIMA $(1, 1, 1)\ (2, 1, 5)_{11}$ configuration yielded the lowest Akaike Information Criterion (AIC), indicating a better balance between model fit and complexity for 2023. Further the seasonal period of 11 is chosen as the data had a large number of missing values filled by statistical methods on different months. This reflected in a trade off in performance of the model when it was modelled for a period of 12 months. Hence, a seasonal period of 11 months was chosen for and optimal Fit. For the year 2024, the SARIMA $(1, 0, 1)\ (3, 1, 6)_{12}$. configuration yielded the lowest Akaike Information Criterion (AIC), indicating a better balance between model fit and complexity. Table 2 gives their model summaries. Minor variations are seen among the all the models.

Tables 3 and **4** present the estimated coefficients for the optimal SARIMA model of order $(1,1,1)\ (2,1,5)_{11}$, $(1,0,1)\ (3,1,6)_{12}$ applied to the $PM_{2.5}$ time series for the years 2023 and 2024, respectively. The Coefficient column reports the estimated parameter values for both seasonal and non-seasonal

components. The standard error reflects the uncertainty associated with each estimate. The z-statistic represents the ratio of the coefficient to its standard error, providing a measure of statistical significance, while P>|z| indicates the two-tailed p-value for the null hypothesis that the coefficient equals zero. For both years, the non-seasonal parameters were statistically significant ($p < 0.05$), suggesting a strong contribution to model performance. However, the seasonal components, while retained for structural completeness, showed comparatively weaker significance especially in 2024 implying less pronounced seasonal effects in that year. This observation supports the visual trends observed in the time series plots. The Ljung–Box Q statistic ($L_1 = 0.67$ and $L_I = 0.57$ for both models), indicates that the residuals approximate white noise, confirming that the temporal structure of the series was adequately captured and validating the assumption of residual independence critical for forecast reliability.

Table 3. Summary of the SARIMA model order $(1,1,1)\ (2,1,5)_{11}$

Model order for 2023		$(1,1,1)\ (2,1,5)_{11}$		
L Jung box (Q)		0.67		
Order	Co-efficient value	Standard error	Z- statistics	$\rho > \|z\|$
AR (1)	0.5160	0.096	5.355	0
MA (1)	-0.8272	0.073	-11.361	0
AR. S (1)	-1.0798	0.90	-11.968	0
AR. S (2)	-0.5290	0.075	-7.071	0
MA. S (1)	0.2163	0.120	1.800	0.072
MA. S (2)	-0.5602	0.088	-6.343	0
MA. S (3)	-0.4551	0.106	-4.2187	0
MA. S (4)	0.1424	0.079	1.801	0.072
MA. S (5)	0.0200	0.083	0.242	0

Table 4. Summary of the SARIMA model order $(1,0,1)\ (3,1,6)_{12}$

Model order for 2024		$(1,0,1)\ (3,1,6)_{12}$		
Ljung box (Q)		0.57		
Order	Co-efficient value	Standard error	Z- Statistics	$\rho > \|z\|$
AR (1)	0.9255	0.020	47.33	0
MA (1)	-0.1938	0.040	-4.808	0
AR. S (1)	0.2591	0.436	0.594	0.533
AR. S (2)	-0.3742	0.457	-0.818	0.413
AR. S (3)	0.0013	0.211	0.006	0.995
MA. S (1)	-1.2103	0.445	-2.720	0.007
MA. S (2)	0.5239	0.808	0.648	0.517
MA. S (3)	-0.3209	0.548	-0.586	0.558
MA. S (4)	0.0744	0.211	0.352	0.725
MA. S (5)	-0.0218	0.170	-0.128	0.898
MA. S (6)	0.0390	0.129	0.302	0.736

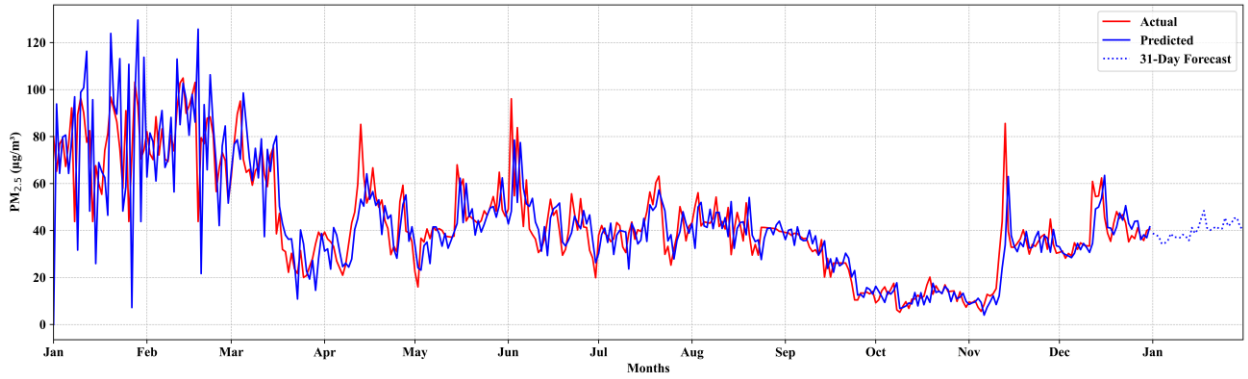


Figure 6. Forecast generated using the SARIMA model of the order $(1,1,1)\ (2,1,5)_{11}$ for the year 2023. The red line represents the actual values. The blue line represents the modelled forecast value. The blue dotted line represents the 31 days out of sample forecast.

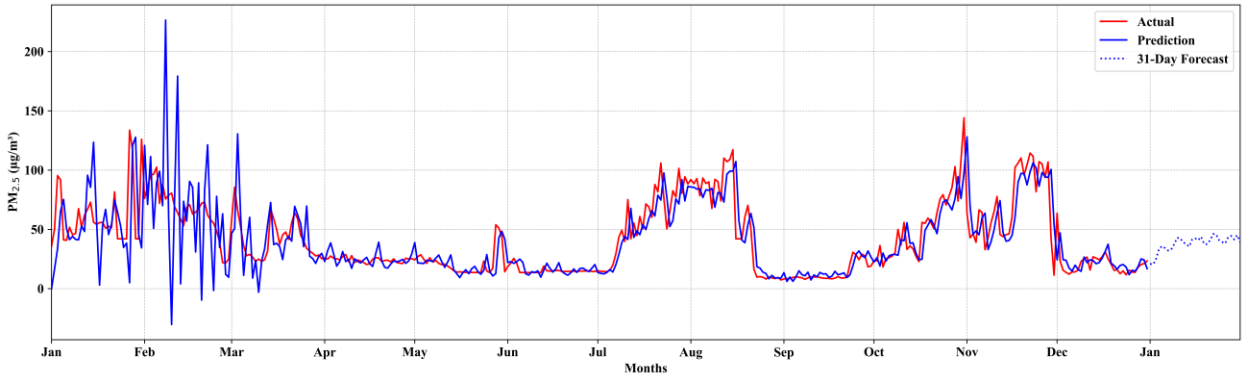


Figure 7. Forecast generated using the SARIMA model of the order (1,0,1) $(3,1,6)_{12}$ for the year 2024. The red line represents the actual values. The blue line represents the modelled forecast value. The blue dotted line represents the 31 days out of sample forecast.

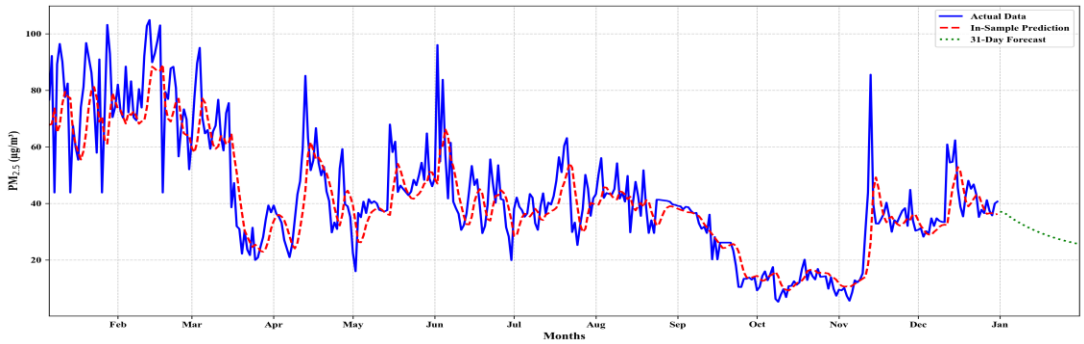


Figure 8. LSTM model forecast for the year 2023. The blue line represents the actual values. The red dashed line represents the modelled forecast value. The green line represents the 31 days out of sample forecast

Figure 6 and 7 present the SARIMA model forecasts for the years 2023 and 2024. The models yielded a Root Mean Square Error (RMSE) of 14.55 µg m^{-3} for the year 2023 and 22.86 µg m^{-3} for 2024, indicating a notable decline in predictive accuracy across the two periods. The increase in RMSE for 2024 suggests greater variability or reduced seasonality in the data, which may have affected the model's ability to capture the underlying temporal patterns. This further underscores the limitations of SARIMA in adapting to evolving or nonlinear dynamics, particularly under conditions of reduced seasonal regularity or abrupt shifts in pollution levels. Several studies have reported RMSE values comparable to the present findings, confirming the robustness of SARIMA models in air quality forecasting. Khamala et al. (2022) used an SARIMA $(1,0,0)(2,1,2)_{12}$ model for aerosol data over East Africa, capturing seasonal patterns effectively and suggesting applicability in broader environmental studies. Bunnang (2024) also applied SARIMA $(3,1,3)$ $(1,1,2)_{12}$ to forecast PM_{10}, achieving an RMSE of 20.839 µg m^{-3}. A comparison with SARIMA-GARCH showed no significant improvement, favoring SARIMA. Similarly, Khanal et al. (2023) evaluated forecasting models for $PM_{2.5}$ in Kathmandu. SARIMA yielded the lowest RMSE of 16.991 µg m^{-3}, outperforming models like Prophet. Other similar studies carried out at different places include, Jain and Mandowara (2019), Mangayarkarasi et al. (2021), Mancini et al. (2022), Atoui et al. (2022), Sakib et al. (2023), and Khan et al. (2024).

LSTM Results

The LSTM model was implemented in Jupyter Lab using Tensor Flow and Keras. It is a stacked neural network designed to predict future $PM_{2.5}$ values based on past observations. The LSTM hyperparameters were selected through a validation based manual tunning process. The model architecture includes three stacked LSTM layers with 50 units each. The first two layers were designed to capture short- and medium-term temporal dependencies, the third layer served as a final temporal encoder before prediction. Dropout layers (rate: 0.3) were added after each LSTM layer to prevent overfitting. A dense layer with one neuron provides the final output. The models were programmed with three layers. The data were scaled to a 0-1 range using MinMaxScaler for efficient training and the train, validation, in 2023 and 2024, the data was divided into three sets. Training (Jan 1- Sep 12), validation (Sep 13-Nov 5), and testing (Nov 6 -Dec 31). A sliding window of size 5 was used to generate input sequences, where each sequence

predicts the next value. The model was compiled using the Adam optimizer and mean squared error loss, with RMSE as an evaluation metric. Early stopping was used to halt training after 10 epochs without validation improvement. Predictions were inverse-transformed and compared with actual values. **Figure 8** and **9** present the LSTM model forecasts for the years 2023 and 2024. The convergence of the model during training was evaluated by plotting the training and validation loss curves for 2023 and 2024, not shown.

The LSTM model for each timeline was run over 10 iterations to avert a particular weight initialization and they are reported in Table 5. It achieved RMSE values of 11.13 μg m^{-3} for 2023 and 17.44 μg m^{-3} for 2024, demonstrating strong predictive accuracy across both years. In comparison, the SARIMA model recorded higher RMSEs of 14.55 μg m^{-3} (2023) and 22.86 μg m^{-3} (2024), indicating reduced performance. This comparative evaluation in this case emphasizes that the LSTM outperformed SARIMA in modelling the nonlinear, stochastic behaviour of $PM_{2.5}$ concentrations, especially under conditions of temporal variability and seasonal shifts. The significant reduction in RMSE underscores LSTM's effectiveness in capturing long-range dependencies and adapting to complex atmospheric dynamics that SARIMA fails to represent adequately. Several studies have reported RMSE values comparable to the present findings, confirming the robustness of LSTM models. Seng et al. (2021) applied an LSTM-based MMSL model for $PM_{2.5}$ prediction in Beijing, achieving RMSEs of 11.15, 10.94, 15.62, and 12.22 μg m^{-3} at four representative stations. The model outperformed traditional methods like ARMA (RMSE: 13.96). In comparison, Tsokov et al. (2022) proposed a hybrid CNN–LSTM spatiotemporal model for hourly $PM_{2.5}$ prediction in Beijing, achieving RMSEs of 30.21 μg m^{-3} (Dongsi), 24.71 μg m^{-3} (Wanliu), and 27.39 μg m^{-3} (Changping) across three stations. The model incorporated optimized architectures via a genetic algorithm and utilized spatial data from surrounding stations. The ensemble of the top-performing models further improved results, with Wanliu achieving the lowest RMSE of 20.74 μg m^{-3}. Other similar studies carried out at different places include, Kim et al. (2019), Xayasouk et al. (2020) and Kristiani et al. (2022).

Model performance comparison

Figure 10 shows the results of the SARIMA and LSTM models' performance analysis. In 2023, the LSTM model demonstrated greater prediction power and flexibility to temporal fluctuations with a significantly lower root mean square error (RMSE) of 11.31 compared to the SARIMA model's 14.55 RMSE. Similarly, with RMSE values of 17.68 and 22.86 in 2024, respectively, the LSTM model continued to outperform SARIMA, despite an increase in error size. The LSTM beats SARIMA by 23.52% in 2023 and 23.71% in 2024, respectively, in percentage terms. The observed increase in RMSE between 2023 and 2024 for the SARIMA model indicates that its predictive power is declining over time. This variation suggests that the data's underlying temporal and seasonal dynamics have changed over the course of the two years. The inherent linearity and assumption of stationarity of the SARIMA model limit its ability to capture nonlinear and time-varying interactions in real-world data. SARIMA is unable to accurately represent abrupt or nonlinear fluctuations caused by other variables present in complex atmospheric and environmental systems, but it performs well for periodic and steady series. Incorporating exogenous variables. The model's performance did not significantly improve, according to preliminary testing. This implies that the short-term fluctuations in the data were not significantly impacted by these factors. By including such features, SARIMAX or hybrid deep learning models perform better when modeling long-term variation. The LSTM model not only captured the seasonal signatures of the unseen future, but also provided a clearer indication of the underlying trend, demonstrating its strength in modeling both short-term fluctuations and long-term temporal dynamics.

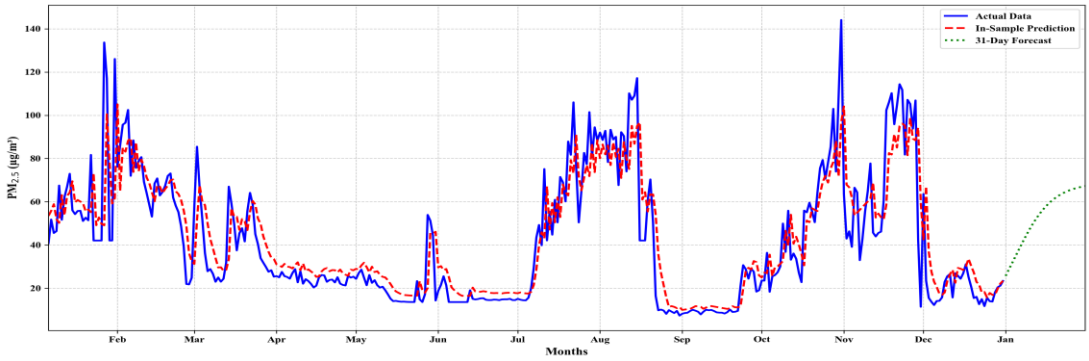


Figure 9. LSTM model forecast for the year 2024. The blue line represents the actual values. The red dashed line represents the modelled forecast value. The green line represents the 31 days out of sample forecast

Table 5. LSTM Models mean RMSE values obtained over 10 Iterations

Year	Number of runs	Mean RMSE	Standard deviation	Range (Min–Max)
2023	10	11.13	0.444	10.348 – 11.732
2024	10	17.44	0.182	17.180 – 17.688

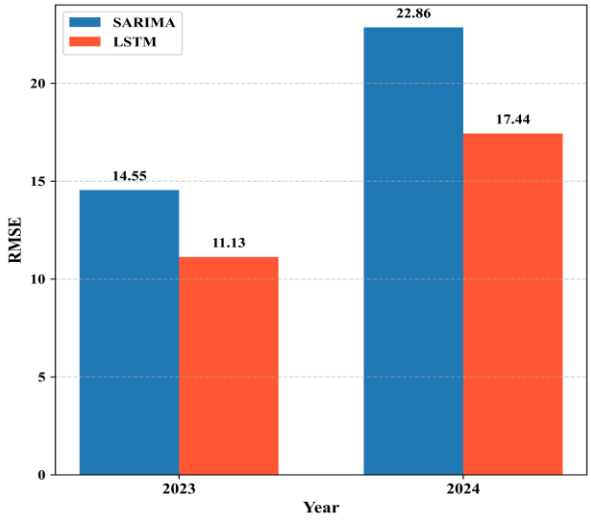


Figure 10. Bar graph comparison of RMSE values for the year 2023 and 2024. The blue bar represents SARIMA model the red bar represents the LSTM model. The RMSE values obtained from SARIMA model is higher the LSTM model in both timelines

CONCLUSIONS

The forecasting performance of parametric (SARIMA) and non-parametric (LSTM) models for $PM_{2.5}$ concentrations over Alandur in 2023 and 2024 was assessed in this study. SARIMA demonstrated greater uncertainty when extrapolating to unknown data, especially in 2024, despite its ability to capture seasonal linear patterns. However, the LSTM model performed better than SARIMA, particularly when it came to identifying non-linear patterns, as evidenced by lower RMSE values (11.13 and 17.44 for 2023 and 2024, respectively) than SARIMA (14.55 and 22.86). Both models were evaluated after imputing the 41-day missing period, the LSTM model achieved superior forecasting performance compared to the SARIMA model, demonstrating its greater capability to capture temporal dependencies even with imputed data. These findings demonstrate how LSTM can be used as an alternative for air quality forecasting in dynamic environmental conditions by non-linear dependencies in atmospheric time series. To enhance environmental risk assessment and early warning systems, the integration of deep learning with traditional statistical models serves as a promising approach.

Acknowledgements

The authors are grateful to the Director, Indian Institute of Geomagnetism, Navi Mumbai for constant support and extending the same to the Ministry of Science and Technology, Government of India. The authors are grateful to the referees for their valuable comments which enhanced the scientific value of this manuscript to several fold.

Author Credit Statement

J. Arul Asir: Programming, methodology, manuscript preparation; H. Johnson Jeyakumar: Supervision, manuscript review; C. P. Anil Kumar: Conceptualization, methodology, supervision, manuscript corrections, revision.

Data availability

Data available with Public domain bins of CPCB

Compliance with Ethical Standards

The authors declare that they have no conflict of interest and adhere to copyright norms.

References

Andreae, M. O., 2009. Correlation between cloud condensation nuclei concentration and aerosol optical thickness in remote and polluted regions. In: Atmos. Chem. Phys., 9, pp 542-556.

Asir, J. A., Jeyakumar, H. J. and Kumar, C. P. A., 2024. Wavelet Based Semblance Analysis of Anthropogenic Aerosol Control on Air Quality During the COVID-19 Pandemic Period. Aerosol Science and Engineering., 9, pp 140-151.

Atoui, A., Slim, K., Andaloussi, S. A., Moilleron, R. and Khraibani, Z., 2022. Time Series Analysis and Forecasting of the Air Quality Index of Atmospheric Air Pollutants in Zahleh, Lebanon. Atmosph.Climate Sci., 12(04), pp 728-749.

Bhattarai, H., Tai, A. P. K., Val Martin, M.V. and Yung, D. H. Y., 2024. Responses of fine particulate matter ($PM_{2.5}$) air quality to future climate, land use, and emission changes: Insights from modeling across shared socioeconomic pathways. Science of the Total Environment., 948, No. 174611.

Bisgaard, S. and Kulahci, M., 2011. Time Series Analysis and Forecasting by Example. Wiley. Wiley Series in Probability and Statistics., pp 48-50.

Box, G. E.P. and Jenkins, G. M. 1976, Time series analysis: forecasting and control, Holden-Day publications, San Fransico.

Brockwell, P. J. and Davis, R. A., 2016. Time Series: Theory and Methods, 3rd Ed., p 83.

Broersen, P. M. T., 2006. Automatic Autocorrelation and Spectral Analysis, 1st Ed., pp 90-95.

Bunnag, T., 2024. Forecasting PM_{10} Caused by Bangkok's Leading Greenhouse Gas Emission Using the SARIMA and SARIMA-GARCH Model. Int. J. Energy Economics and Policy., 14(1), pp 418-426.

Celo, V. and Dabek-Zlotorzynska, E., 2011. Concentration and Source Origin of Trace Metals in $PM_{2.5}$ Collected at Selected Canadian Sites within the Canadian National Air Pollution Surveillance Program. In Environmental Science and Engineering. Springer Science and Business Media, Deutschland GmbH., pp 19–38.

Chatfield, C., 1975. The Analysis of Time Series: Theory and Practice, 1st Ed, Springer New York, NY., p 60.

Chen, Y., Wild, O., Conibear, L., Ran, L., He, J., Wang, L. and Wang, Y., 2020. Local characteristics of and exposure to fine particulate matter ($PM_{2.5}$) in four Indian megacities. Atmosph. Environment: X., 5, No. 100052.

Cheung, H. C., Chou, C. C.-K., Lee, C. S. L., Kuo, W.-C. and Chang, S.-C., 2019. Hygroscopic properties and CCN activity of atmospheric aerosols under the influences of Asian continental outflow and new particle formation at a coastal site in East Asia., 20(10), 5911-5922.

Cromwell, J. B., Hannan, M. J., Labys, W. C. and Terraza, M., 1994. Multivariate tests for time series models. Sage Publications., Inc, pp 21-23.

Donahue, L. N., Posner, L. N., Westervelt, D. M., Li, Z., Shrivastava, M., Presto, A. A., Sulivan, R. C., Adams, P. J., Pandis, S. N. and Robinson, A. L., 2016. Where Did This Particle Come From? Sources of Particle Number and Mass for Human Exposure Estimates in Airborne Particulate Matter Sources, Atmospheric Processes and Health, ed. R. M. Harrison, R. E. Hester, and X. Querol, The Royal Society of Chemistry., 2016, pp 35-71.

Fan, J. and Yao, Q., 2005. Nonlinear time series. Nonparametric and parametric methods., p 9.

Farahat, A. and Abuelgasim, A., 2022. Effect of cloud seeding on aerosol properties and particulate matter variability in the United Arab Emirates. Int. J. Environmental Sci. Technol., 19(2), 951-968.

Fu, Y., Tai, A. P. K. and Liao, H., 2016. Impacts of historical climate and land cover changes on fine particulate matter ($PM_{2.5}$) air quality in East Asia between 1980 and 2010. In: Atmospheric Chemistry and Physic., Vol. 16(16), pp 10369-10383.

Gao, D., Zhao, B., Wang, S., Wang, Y., Gaudet, B., Zhu, Y., Wang, X., Shen, J., Li, S., He, Y., Yin, D. and Dong, Z., 2023. Increased importance of aerosol-cloud interactions for surface $PM_{2.5}$ pollution relative to aerosol-radiation interactions in China with the anthropogenic emission reductions. Atmosph. Chem. Phys., 23(22), 14359-14373.

Gautam, A. and Singh, V., 2020. Parametric versus non-parametric time series forecasting methods: A review. J. Engineering Sci. Technol. Rev., 13(3), 165-171.

Gu, Y., Fang, T. and Yim, S. H. L., 2024. Source emission contributions to particulate matter and ozone, and their health impacts in Southeast Asia. Environment Int., 186, 108578.

Guevara, M., 2016 Emissions of Primary Particulate Matter. In: Airborne Particulate Matter Sources, Atmospheric Processes and Health, ed. R. M. Harrison, R. E. Hester, and X. Querol, The Royal Society of Chemistry, pp.1-34.

Hodson, T. O., 2022. Root-mean-square error (RMSE) or mean absolute error (MAE): when to use them or not. Geoscientific Model Development., 15(14), 5481–5487.

Hyndman, R.J. and Athanasopoulos, G.,2014. Forecasting: principles and practice, 2nd Ed., pp 27-46, 157-180, 223-276.

Jain, S. and Mandowara, V. L., 2019. Study on Particulate Matter Pollution in Jaipur City. Int. J. Appl. Engineering Res., 14(3), 637-645.

Kaur, J., Parmar, K. S. and Singh, S., 2023. Autoregressive models in environmental forecasting time series: a theoretical and application review. Environmental Sci. Poll. Res., 30(8), 19617-19641.

Khamala, G. W., Makokha, J. W. and Boiyo, R., 2022. Statistical Analysis of Aerosols Characteristics from Satellite Measurements over East Africa Using Autoregressive Moving Average (ARIMA). OALib., 09(11), 1-14.

Khan, D. R., Patankar, A. B. and Khan, A., 2024. An experimental comparison of classic statistical techniques on univariate time series forecasting. Proced. Comput. Sci., 235, 2730-2740.

Khanal, I., Prakash Dhakal, O., Kumar Guragain, M., Karki, P. and Pandey, B., 2023. Proc. 14th IOE Graduate Conference Comparative Analysis of Time Series Forecasting Models for Predicting $PM_{2.5}$ level in Kathmandu : SARIMA, Prophet and XG-Boost., 14, pp 1182-1189.

Kim, H. S., Park, I., Song, C. H., Lee, K., Yun, J. W., Kim, H. K., Jeon, M., Lee, J. and Han, K. M., 2019. Development of a daily PM_{10} and $PM_{2.5}$ prediction system using a deep long short-term memory neural network model. Atmosph. Chem. Phys., 19(20), 12935-12951.

Kirchgässner, G. and Wolters, J., 2007. Introduction to Modern Time Series Analysis., pp15-19.

Kristiani, E., Lin, H., Lin, J. R., Chuang, Y. H., Huang, C. Y. and Yang, C. T., 2022. Short-Term Prediction of $PM_{2.5}$ Using LSTM Deep Learning Methods. Sustainability (Switzerland)., 14(4), No. 2068.

Kumar, U. and Jain, V. K., 2010. ARIMA forecasting of ambient air pollutants (O_3, NO, NO_2 and CO). Stochastic Environmental Res. Risk Assess., 24(5), 751-760.

Luo, J. and Gong, Y., 2023. Air pollutant prediction based on ARIMA-WOA-LSTM model. Atmosph. Poll. Res., 14 (6), No.101761.

Mancini, S., Francavilla, A., Graziuso, G. and Guarnaccia, C., 2022. An Application of ARIMA modelling to air pollution concentrations during covid pandemic in Italy. J. Phys. Conf. Ser., 2162(1), No. 012009.

Mangayarkarasi, R., Vanmathi, C., Khan, M. Z., Noorwali, A., Jain, R. and Agarwal, P., 2021. COVID19: Forecasting Air Quality Index and Particulate Matter ($PM_{2.5}$). Comput. Materials and Contin., 67(3), 3363-3380.

Megaritis, A. G., Fountoukis, C., Charalampidis, P. E., Denier Van Der Gon, H. A. C., Pilinis, C. and Pandis, S. N., 2014. Linking climate and air quality over Europe: Effects of meteorology on $PM_{2.5}$ concentrations. Atmosph, Chem. Phys., 14(18), 10283-10298.

Naveen, V. and Anu, N., 2017. Time Series Analysis to Forecast Air Quality Indices in Thiruvananthapuram District, Kerala, India. Int. J. Engineering Res. Appl., 07(06), 66-84.

Pal, A. and Prakash, P.K.S., 2017. Practical Time Series Analysis: Master Time Series Data Processing, Visualization, and Modelling Using Python. Packt Publishing Ltd., pp 47-83, 113, 193.

Pan, R., Song, J., Yi, W., Liu, J., Song, R., Li, X., Liu, L., Yuan, J., Wei, N., Cheng, J., Huang, Y., Zhang, X. and Su, H., 2024. Heatwave characteristics complicate the association between $PM_{2.5}$ components and schizophrenia hospitalizations in a changing climate: Leveraging of the individual residential environment. Ecotoxicology and Environmental Safety, 271, No. 115973.

Peixeiro, M., 2022. Time Series Forecasting in Python. Manning, Shelter Island., p 31.

Rebala, G., Ravi, A. and Churiwala, S., 2019. An Introduction to Machine Learning., pp 134-136.

Sakib, S. R., Sara, K. N., Rasel Tahmid, M. T. H., Rasel, H., Masudul, M., Asif, M. I. A., Aynul, M., Nahid, A. H., Rahman, S., Mridha, M. F. and Islam, A., 2023. Time Series Analysis and Forecasting of Air Quality Index of Dhaka City of Bangladesh., pp 67-76.

Seinfeld, J.H. and Pandis, S.N., 2016. Atmospheric chemistry and physics: from air pollution to climate change, 3rd Ed., p 350.

Seng, D., Zhang, Q., Zhang, X., Chen, G. and Chen, X., 2021. Spatiotemporal prediction of air quality based on LSTM neural network. Alexandria Engineering J., 60(2), 2021–2032.

Shumway, R. H. and Stoffer, D. S., 2011. Time Series Analysis and Its Applications, 4th Ed, Springer New York., p 1.

Singh, V., Singh, S. and Biswal, A., 2021. Exceedances and trends of particulate matter ($PM_{2.5}$) in five Indian megacities. Sci. Total Environment., 750, No.141461.

Slutzky, E., 1937. Summation of random causes as the source of cyclic processes, Econometrica., 5, 105–146.

Tai, A. P. K., Mickley, L. J., Jacob, D. J., Leibensperger, E. M., Zhang, L., Fisher, J. A. and Pye, H. O. T., 2012. Meteorological modes of variability for fine particulate matter ($PM_{2.5}$) air quality in the United States: Implications for $PM_{2.5}$ sensitivity to climate change. Atmosph. Chem. Phys., 12(6), 3131–3145.

Tsokov, S., Lazarova, M. and Aleksieva-Petrova, A., 2022. A Hybrid Spatiotemporal Deep Model Based on CNN and LSTM for Air Pollution Prediction. Sustainability Switzerland., 14(9), No 5104.

Walker, G., 1931 On periodicity in series of related terms. Proc. R. Soc. Lond. Ser., A 131(818), 518-532.

Wold, H., 1938. A Study in the Analysis of Stationary Time Series-Stockholm, Almqvist and Wiksel., pp 214-234.

Xayasouk, T., Lee, H. M. and Lee, G., 2020. Air pollution prediction using long short-term memory (LSTM) and deep autoencoder (DAE) models. Sustainability (Switzerland)., 12(6), No. 2570.

Yang, F., Tan, J., Zhao, Q., Du, Z., He, K., Ma, Y., Duan, F., Chen, G. and Zhao, Q., 2011. Characteristics of $PM_{2.5}$ speciation in representative megacities and across China. Atmosph. Chem. Phys., 11(11), 5207-5219.

Yim, S. H. L., Li, Y., Huang, T., Lim, J. T., Lee, H. F., Chotirmall, S. H., Dong, G. H., Abisheganaden, J., Wedzicha, J. A., Schuster, S. C., Horton, B. P. and Sung, J. J. Y., 2024. Global health impacts of ambient fine particulate pollution associated with climate variability. Environment Int., 186, No. 108537.

Yu, M., Xu, F., Hu, W., Sun, J. and Cervone, G., 2021. Using Long Short-Term Memory (LSTM) and Internet of Things (IoT) for Localized Surface Temperature Forecasting in an Urban Environment. IEEE Access., 9, 137406-137418.

Yule, G. U., 1927. On a Method of Investigating Periodicities in Disturbed Series, with Special Reference to Wolfer's Sunspot Numbers Phil. Trans. R. Soc. Lond., A, 226, 267-298

Zhang, B., Jiao, L., Xu, G., Zhao, S., Tang, X., Zhou, Y. and Gong, C., 2018. Influences of wind and precipitation on different-sized particulate matter concentrations ($PM_{2.5}$, PM_{10}, $PM_{2.5–10}$). Meteorol. Atmosph, Phys., 130(3), 383-392.

Zhang, Q., Meng, J., Quan, J., Gao, Y., Zhao, D., Chen, P. and He, H., 2012. Impact of aerosol composition on cloud condensation nuclei activity. Atmosph. Chem. Phys., 12(8), 3783–3790.

Zhao, B., Liou, K. N., Gu, Y., Li, Q., Jiang, J. H., Su, H., He, C., Tseng, H. L. R., Wang, S., Liu, R., Qi, L., Lee, W. L. and Hao, J., 2017. Enhanced PM2.5 pollution in China due to aerosol-cloud interactions. Scientific Rep., 7(1), No. 4453.

Received on: 30-6-2025; Revised on 16-12-2025; Accepted on 23-12-2025