Journal of Theoretical and Applied Information Technology
15th June 2024. Vol.102. No. 11

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

A HYBRID SARIMA-LSTM APPROACH FOR IMPROVED TIME SERIES PREDICTION OF AEROSOL OPTICAL DEPTH ACROSS DELHI,INDIA

NAUMI KRISHNA K. PANICKER[1], J. VALARMATHI[2]

[1]Research Scholar, School of Electronics Engineering, Vellore Institute of Technology, Vellore, India

[2]Professor, School of Electronics Engineering, Vellore Institute of Technology, Vellore, India

E-mail: [1]naumikrishna@gmail.com, [2]jvalarmathi@vit.ac.in

ABSTRACT

Atmospheric aerosols are one of the indispensable particles in understanding atmospheric dynamics and are essential for accurate environmental forecasting and policy development. The literature on AOD time series forecasting usually uses either statistical methods, which handle linear patterns but struggle with non-linearities, or machine learning (ML) and deep learning (DL) methods, which capture non-linearities but can be limited in accurately processing the linear components present in the data. This study introduces a hybrid model that combines statistical methods and ML techniques to effectively address both the linear and non-linear components present in AOD time series data. The primary goal of this work was to understand the potential of the hybrid SARIMA-LSTM (seasonal autoregressive integrated moving average—long short-term memory) model to enhance the forecasting capacity of AOD time series data. The proposed model was compared to its baseline models, SARIMA and LSTM, by utilizing monthly data from the Moderate Resolution Imaging Spectroradiometer (MODIS) satellite across the Delhi region of India from 2001 to 2019. The performance of these models was evaluated based on root mean square error (RMSE), coefficient of determination (R^2), and mean absolute percentage error (MAPE) during both training and testing phases. The proposed model outperformed the baseline models in all three metrics. The findings of this study advocate hybrid modeling as a promising tool for improving the accuracy of time series prediction of AOD because it can handle both linear and non-linear aspects present in the data.

Keywords: *SARIMA, LSTM, Hybrid, Time Series Prediction, Forecasting, AOD, MODIS*

1. INTRODUCTION

Atmospheric aerosols are multifaceted, tiny particles that have profound effects on the climate, air quality, and human health. These particles affect the climate through direct and indirect interactions. They directly affect the Earth's radiation budget by scattering or absorbing the solar radiation, thereby resulting in cooling or warming effects [1]. Atmospheric aerosols have an indirect impact on the climate by modifying the dynamics of cloud formation, as they serve as cloud condensation nuclei (CCN). The substantial amount of CCN can lead to clouds with more droplets that are more reflective and last longer, potentially leading to a cooling effect. This is referred to as the Twomey effect [2]. The atmospheric aerosols can affect the size of cloud droplets as well, influencing precipitation patterns known as the Albrecht effect [3]. Being tiny in size, these particles can penetrate deep into the lungs, causing respiratory and cardiovascular health issues. The scattering and absorption of atmospheric aerosols also contribute to visibility issues, a particularly profound problem in urban cities with high pollution levels [4]. These impacts and the complex behavior of atmospheric aerosols highlight the need for continuous understanding, monitoring, and prediction of their patterns, thereby aiding in the development of effective mitigation strategies.

AOD is the most widely used parameter to quantify the total aerosol concentration in a region. It is a dimensionless parameter with usual values between 0 and 1. Higher values of AOD indicate a higher aerosol concentration in the atmosphere, and vice versa. An increase in aerosol concentration can lead to a higher amount of scattering or absorption, thereby affecting the Earth's radiation budget and climate [5]. AOD measurements are thus significant for assessing the various impacts of aerosols on climate and air quality, and are crucial for the development of plans for air quality management.

Understanding AOD dynamics can lead to more accurate climate estimates. However, the limited amount of observational AOD data still remains a major challenge[6]. AOD is measured using satellite remote sensing and ground-based instruments. While ground-based instruments provide data at specific locations, they have limited spatial coverage and can also suffer from maintenance and calibration issues. Satellite remote sensing provides greater spatial coverage, but it also faces challenges such as cloud cover and irregular terrain effects. As a result, predictive modeling has become useful. Predictive modeling can be primarily classified as global models and time series forecasting. Global models such as Chemistry Transport Models (CTM) and Deterministic Weather Prediction Models (DM) can predict the AOD, but their accuracy is restricted by the AOD's inherent variability and the complexity of atmospheric processes [7]. To develop the simulations for understanding the variations of aerosols, physical, chemical, and dynamic processes need to be considered. Under these circumstances, time series forecasting started gaining momentum as it used historical data to predict future AOD values.

By considering the literature focusing on the time series forecasting of AOD, it can be classified primarily into two categories: statistical modeling and machine learning, and deep learning methodologies. The most widely used statistical models for AOD forecasting are ARIMA and its extended version, SARIMA, as they depend upon the concept of serial correlation, where previous data influences future values [8] . In 2013, B. Abish and K. Mohankumar, to overcome the difficulties of using complex deterministic models, developed a simple and efficient model for predicting the future values of AOD across the northern Indian region by utilizing the multiangle imaging spectroradiometer (MERRA) data from the years 2000 to 2010. This model showed promising results for accurately forecasting AOD data in the region but underestimated the high values of AOD that would have resulted from extreme events [9]. Developing on the work proposed by Abish and Mohankumar (2013), Soni et al., (2014) applied the ARIMA model to predict the MODIS-AOD across eight selected sites in the northern Indian and Himalayan regions. The model performed well in low-AOD sites and performed satisfactorily whenever there were high values of AOD. The ARIMA model was successful in simulating seasonality [10]. Following this, Soni et al., (2015) again utilized ARIMA for the modeling of MODIS-derived AOD time series data across 11 coal mines across India, covering the period from March 2000 to December 2012. The study successfully demonstrated the effectiveness of ARIMA in achieving satisfactory predictions at all the selected coal mines [11]. Taneja et al., (2016) considered New Delhi and employed the ARIMA model by utilizing monthly average AOD data from Terra MODIS over a period of ten years (2004–2014) to analyze future values of AOD. The results showed the feasibility of such a simple model for simulating future values of AOD, although the model encountered issues when dealing with extreme values [12]. Soni et al., (2016) analyzed AOD data from MODIS and AERONET over the Indo-Gangetic Plains (IGP) from 2001–2012 using the ARIMA model. They assessed the model's accuracy with statistical metrics and found a significant correlation between the two datasets, showing MODIS data could be derived from AERONET data by adding a specific value. They also predicted an increasing trend in AOD for 2013–2017, demonstrating the model's effectiveness in forecasting [13]. Kumar et al., (2018) conducted a comprehensive analysis of AOD and applied ARIMA across the entire region of IGP and nine specific stations located in the upper, central, and lower IGP regions using the Terra MODIS Collection 6 enhanced Deep Blue (DB) AOD retrieval algorithm. The study showed that while the ARIMA model provided reliable predictions in general, the non-stationary behavior of aerosol loading and heterogeneity in aerosol properties limited its performance, which was witnessed in areas such as Lahore, Kolkata, Karachi, and Multan. The authors suggested the potential use of more advanced models like adaptive neuro-fuzzy inference systems and artificial neural networks to better handle the complexities of AOD behavior [14]. Li et al., (2019) conducted a comparative analysis of AOD in the U.S. (United States) and China from 2003 to 2015, and employed SARIMA to reveal high AOD values in the eastern regions and distinct seasonal peaks in the summer. The study highlighted that SARIMA effectively captures temporal variations and trends, providing reliable short-term predictions. However, it was also noted that the model's reliance on past data might limit its accuracy in predicting extreme events [15]. The study by Abuelgasim et al., (2021) on AOD variability over the United Arab Emirates (UAE) for the period of 2003 to 2018 highlighted the effectiveness of SARIMA modeling for AOD forecasting. SARIMA outperformed the Exponential Smoothing State Space Model (ETS-error, trend, and seasonal) model. The findings of this study confirmed SARIMA's reliability for monthly AOD forecasts, making it suitable for capturing seasonal

variations and trends in the UAE's AOD data [16]. Singh et al., (2022) employed ARIMA for the time series forecasting of monthly mean AOD and angstrom exponent (AE) obtained from MODIS across three different locations in the IGP, namely, Jaipur, Kanpur, and Ballia, from the period 2003 to 2018. The performance of the ARIMA model was compared to three statistical models: average, naive, and drift methods. The ARIMA model outperformed all three models and emerged as the best-fit model to predict both AOD and AE in all three regions [17]. Dutta et al., (2022) applied the ARIMA model to predict future AOD levels across various Indian states. The results indicated a significant shift in aerosol pollution levels, with southern Indian states expected to become more vulnerable by 2023. This study emphasized the necessity of ARIMA for projecting future scenarios based on historical trends and seasonal patterns [18]. However, it is important to note that various studies have highlighted a key limitation of ARIMA and SARIMA models related to their inability to effectively deal with non-linearity in the data.

Recently, ML and DL models started gaining popularity due to their ability to deal with complex data. Nabavi et al., (2018) compared deterministic weather prediction models (DMs) and machine learning algorithms (MLAs) to address the gaps in forecasting the monthly mean AOD in West Asia. MLAs outperformed DMs in terms of prediction error and accuracy [19]. Eltahan and Moharm (2020) developed a simple and efficient long-short-term memory (LSTM) model to estimate and predict the temporal trend of AOD for air quality in five Egyptian cities [20]. Naumi et al., (2021) developed the Prophet model for forecasting aerosol optical depth (AOD) across major urban areas in India, such as Delhi, Mumbai, Kolkata, and Trivandrum. The Prophet model outperformed the traditional SARIMA model in terms of accuracy and efficiency, providing better forecasts with lower errors and faster execution times [21]. Daoud et al., (2021) tested the accuracy of predicting AOD in four areas within the global dust belt: the Eastern Libyan Desert, the Saudi Arabian Peninsula, the Indian Subcontinent, and China. To accomplish this, the authors used three models, namely LSTM, CNN-LSTM (Convolutional Neural Networks-Long-Short Term Memory), and ConvLSTM (Convolutional Long-Short Term Memory), which use algorithms to detect patterns in data as well as a traditional Fast Fourier Transform (FFT) algorithm. The ConvLSTM algorithms performed the best, with RMSE within ±10% [22]. Zaheer et al., (2023) developed the SVR model by utilizing pre-processing techniques like forward feature selection. (FFS) and the grey wolf optimizer optimization algorithm for the prediction of AOD across Pakistan using satellite data. The SVR-GWO model was compared to SVR and multiple linear regression (MLR) models and was able to achieve better performance [23]. Nevertheless, nonlinear models may not be able to cope with linear and nonlinear patterns equally effectively due to their inability to identify the underlying patterns.

The primary requirement of any time series forecasting scenario is the development of more efficient and accurate forecasting models. Previous studies on AOD time series forecasting have shown that both statistical and ML/DL models can successfully forecast AOD. However, these methods often struggle with the linearities and non-linearities present in AOD time series data. Statistical models, while simple and flexible, face issues when dealing with non-linearities. On the other hand, non-linear models may not efficiently capture linear trends. Motivated by the limitations observed in both approaches, our work aims to enhance the accuracy of AOD time series forecasting by combining the strengths of both linear and non-linear models into hybrid systems.

Hybrid systems based on residual forecasting have shown promising results in various applications [24],[25],[26],[27],[28] . In these models, the residual, which is the difference between the predicted and actual values of the first model, serves as the input for the second model. In our study, we use the statistical model SARIMA as the first model and the DL model LSTM as the second. Such hybrid models, which use residuals, can correct biased forecasts that may occur due to overfitting, underfitting, or model mis-specification, thereby improving overall forecasting accuracy [29].

The Delhi region holds significant importance from an aerosol perspective due to its high levels of air pollution, particularly from anthropogenic sources. Delhi consistently ranks among the most polluted cities globally, with particulate matter (PM2.5 and PM10) and AOD levels frequently exceeding safe limits [30]. This pollution arises from various sources, including vehicular emissions, industrial activities, construction dust, and biomass burning [31]. The high aerosol concentrations in Delhi have severe public health implications, including respiratory and cardiovascular diseases, and are linked to increased mortality rates. Accurate forecasting of AOD in Delhi is significant for overcoming the health risks, informing policy decisions, and comprehending the broad-ranging environmental impacts of aerosols.

By knowing the underlying patterns and the futuristic patterns, it can aid in creating awareness about aerosol concentrations and public health.

The main objective of this study is to develop and evaluate a hybrid SARIMA-LSTM model for improved time series forecasting of AOD across the Delhi region. The specific objectives are:

1. To develop a hybrid model that is a combination of statistical and deep learning models based on residuals to improve the AOD prediction accuracy.
2. To compare the performance of the hybrid SARIMA-LSTM model to its individual SARIMA and LSTM models through performance metrics.
3. To validate the performance of the proposed hybrid model through additional analyses, including autocorrelation function (ACF) plots and residual density distribution of training residuals.

Hence, in this paper, we aim to develop an accurate AOD time series prediction model with an efficient structure that adapts to real-world situations and gives better prediction capability. In order to deal with the real-world problem, which may contain linear and non-linear components, we are proposing the hybrid SARIMA-LSTM.

To the best of our knowledge, hybrid systems that perform residuals modeling combining the SARIMA and LSTM were not proposed or assessed for AOD time series prediction across the Delhi region.

The primary contributions of this study are:

1. The development of a hybrid methodology that merges the DL and statistical linear approaches aimed at enhancing the accuracy of AOD time series prediction.

2. A comprehensive evaluation of the hybrid model using the AOD dataset from the MODIS satellite for the Delhi region and a comparative analysis with its constituent models, SARIMA and LSTM, to verify the superiority of hybrid modeling via residual forecasting in improving AOD time series prediction accuracy.

Section 2 details the study area, dataset, and mathematical background of the models used in this study. Section 3 describes the proposed methodology for development. Section 4 presents the results and discussions, focusing on the comparative analysis of model performances and their practical implications. Finally, Section 5 concludes the study with a summary of findings, limitations, and potential directions for future research.

2. MATERIALS AND METHODS

2.1 Study Region

In this study, Delhi (located from latitudes of 28º 0.21’ to 28º 0.53’ North and longitudes of 76º 0.20’ to 77º 0.37’ East, with elevations ranging from 213.3 to 305.4 meters above sea level) is considered as shown in figure 1. A densely populated national capital, it is currently ranked as the third most polluted city in the world, according to the IQAir 2023 report [32]. Surrounded by the Himalayas to the north, the Thar Desert to the west, the Vindhyan Ranges to the south, and the Gangetic Plains to the east, Delhi's geographical location, high population and high urbanization level necessitates a focus on environmental and air quality studies.

The primary sources for the alarming rise in air pollution in this region are anthropogenic and natural aerosol sources. Vehicular emissions, industrial discharges, and biomass burnings from the surrounding regions are the main factors in the anthropogenic aerosol formations [33]. Natural sources, such as dust from the Great Indian Desert and neighboring regions, contribute to these sources. These aerosol sources lead to high AOD levels and thereby influence not only local but also regional climatic patterns. Therefore, the understanding of these patterns and time series modeling of atmospheric aerosols in Delhi is significant.

2.2 Data Description

The AOD data at 550 nm wavelength collected by the MODIS sensor on the Terra satellite developed by NASA is utilized in this study. The Terra satellite was launched in 1999 and is part of NASA's Earth Observing System (EOS). It orbits the Earth from pole to pole, crossing the equator in the morning at approximately 10.30 a.m. IST [34], [35]. MODIS captures data in 36 spectral bands ranging from visible to thermal infrared at varying spatial resolutions. This sensor plays a critical role in observing and measuring large-scale global

Journal of Theoretical and Applied Information Technology
15th June 2024. Vol.102. No. 11

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

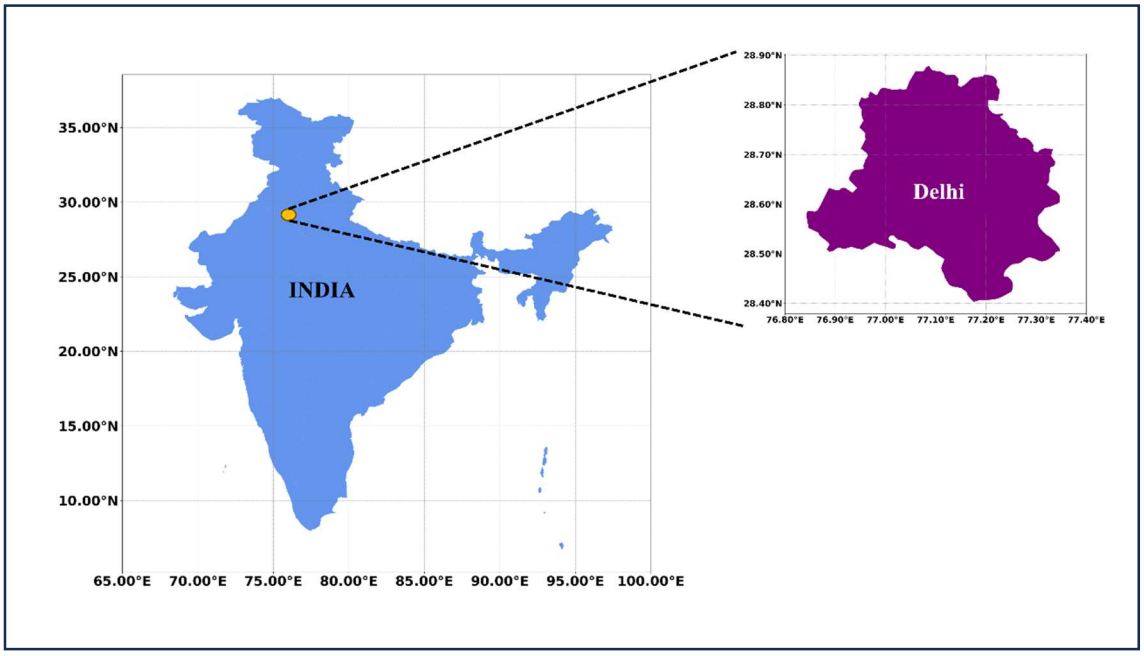

Figure 1. The representation of location of the study region.

dynamics such as changes in the Earth's cloud cover, radiation budget, and processes occurring in the oceans, on land, and in the lower atmosphere. MODIS data are widely used for studying Earth's surface and atmosphere, providing insights into environmental conditions and trends.

In this study, the monthly mean area-averaged AOD data at 550 nm over the Delhi region from 2001 to 2019 were collected from the Terra satellite via Giovanni. Giovanni is an online portal developed by NASA GES DISC, a data center that archives and distributes Earth science data available in http://disc.sci.gsfc.nasa.gov/giovanni. Version 6.1 (MOD08_M3 V 6.1) is utilized in this study. The spatial resolution is 1º x 1º, and the temporal resolution is monthly. The dataset is processed using the Dark Target (DT) algorithm.

2.3 SARIMA

SARIMA is one of the most widely used statistical models and is a derivative of ARIMA. This model effectively handles the seasonal variations, an aspect that the conventional ARIMA model does not adequately address[36] .The general representation of SARIMA is (p, d, q) (P, D, Q, s). The general representation of SARIMA is $(p,d,q)(P,D,Q,s)$, where the first set of parameters (p,d,q) defines the non-seasonal components, and the second set (P,D,Q) together with s captures the seasonal elements of the model. Here p and P are the number of autoregressive terms in non-seasonal and seasonal terms, respectively. The terms d and D stand for the number of differences and seasonal differences, respectively, undergone to make the data stationary. The q and Q represent the number of non-seasonal and seasonal moving average terms, respectively.

The simplified mathematical representation of SARIMA is given below in **Error! Reference source not found.** :

$$\varphi(L)\phi(L^s)\nabla^d\nabla_s{}^D = \theta(L)\Theta(L^s)\varepsilon_t \qquad (1)$$

where

- $\varphi(L) = 1 - \sum_{i=1}^{p} \varphi_i L^i$: the non-seasonal AR terms.
- $\phi(L^s) = 1 - \sum_{i=1}^{p} \phi_{is} L^{is}$: the seasonal AR terms.
- $\theta(L) = 1 + \sum_{i=1}^{q} \theta_i L^i$: non-seasonal MA terms.
- $\Theta(L^s) = 1 + \sum_{i=1}^{q} \Theta_i L^{is}$: seasonal MA terms.

From equation (1), it is clear that the SARIMA model captures the non-seasonal and seasonal patterns present in the data, thereby effectively modeling seasonal data.

2.4 LSTM

LSTM is an advanced version of recurrent neural network (RNN) introduced by Hochreiter and Schmidhuber in order to mainly overcome the gradient vanishing and gradient exploding issues [37]. LSTMS are extensively used in time series applications as they have the ability to learn the long-term dependencies existing in the time series due to presence of memory cell. The existence of non-linear activation functions benefits in dealing with the non-linear behaviors of the data.

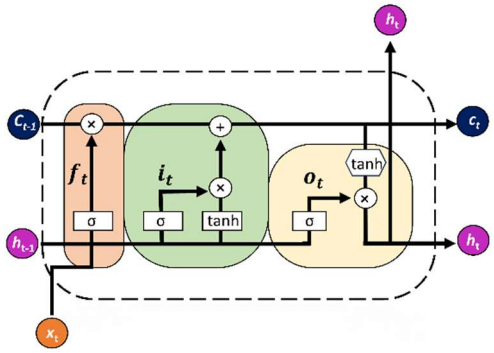

Figure 2: The structure of LSTM.

Figure 2 depicts the structure of the LSTM unit. It has mainly three gates, which act as regulators. They decided on the flow of information. The three gates are the forget gate, the input gate, and the output gate. The gates are basically sigmoid functions where the output values lie between 0 and 1.

The initial layer of the LSTM cell is the forget gate, which is assigned the task of determining which information is to be kept or discarded from the cel ıg the hidden state (h_{t-1}) and current input (x_t), results in either 0 or 1 to either discard or maintain the information. Any value closer to 0 implies to neglect more information, and a closer value to 1 indicates to store more information. The mathematical equation (2) describing the forget gate is given below:

$$f_t = \sigma\left(W_f \cdot [h_{t-1}, x_t] + b_f\right) \quad (2)$$

where W_f and b_f are the weight and bias parameters for the forget gate, respectively.

The input gate layer is the next layer, determining the new information to be added to the cell state. It first applies a sigmoid function on h_{t-1} and x_t to decide which parts of the new input are important, using the previous hidden state and the current input. Simultaneously, a tanh layer produces candidate values for updating the cell state as shown in equation (3) and (4):

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (3)$$

$$\widetilde{C_t} = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C) \quad (4)$$

Here, $W_i, W_C, b_i,$ and b_C represent the weights and biases related to the input gate and the generation of candidate values.

The cell state C_t is updated by merging the old cell state C_{t-1} , modulated by the forget gate, with the new candidate values, influenced by the input gate as given below in equation (5):

$$C_t = f_t * C_{t-1} + i_t * \widetilde{C_t} \quad (5)$$

This equation highlights the selective updating process, incorporating new, relevant information while discarding outdated data.

The final layer is the output gate, which determines the aspects of the cell state to be used in the output. It starts with a sigmoid function evaluating the inputs to generate a gating vector. The cell state is then processed through a tanh function to normalize its values:

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (6)$$

$$h_t = o_t * \tanh(C_t) \quad (7)$$

This results in the final output h_t of the LSTM cell, which is ready for subsequent processing or as input for the next timestep in sequence-based tasks.

By making use of gates and cell states, the LSTM cell can control the flow of information and what is to be remembered and discarded, thereby overcoming the vanishing gradient problem.

2.5 Hybrid SARIMA-LSTM

The MODIS AOD data contains both linear and non-linear components. The SARIMA model can be of use to deal with the linear part of the data and LSTM can deal the non-linear component, The proposed hybrid SARIMA-LSTM can thus handle both the linear and non-linear components as it combines the strength of both SARIMA and LSTM.

Suppose the MODIS AOD data at time t is represented as A_t and it's a combination of linear and non-linear components as given in equation (8):

$$A_t = L_t + N_t \quad (8)$$

Here, L_t and N_t are the linear and non-linear components, respectively. The SARIMA model is utilized to take care of the seasonality and trend parts of the MODIS AOD data and generate the SARIMA-based predicted linear component. The residual is obtained by differencing the actual MODIS data from the predicted linear component by SARIMA, as indicated in equation (9):

$$R_t = A_t - \hat{L}_t \quad (9)$$

The residual from SARIMA is fed as the input to LSTM for predicting the non-linear part and it predicts $\hat{N}_t$. The final prediction is the combination of the predicted linear and non-linear components respectively as shown in equation (10):

$$\hat{A}_t = \hat{L}_t + \hat{N}_t \quad (10)$$

The proposed model has the potential to incorporate both the linear and non-linear dynamics of the data, thereby enhancing the prediction accuracy [38]. The workflow of the proposed model is depicted in figure 3.

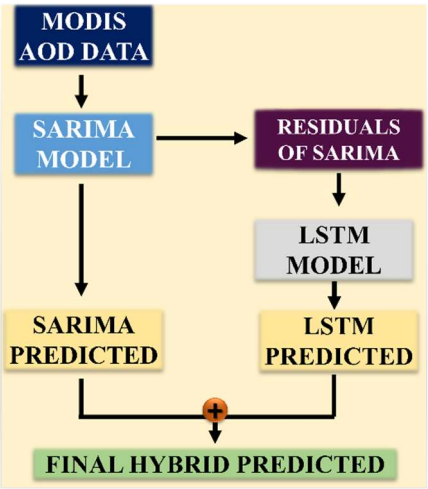

Figure 3. The workflow of hybrid SARIMA-LSTM.

3 SYSTEM DESCRIPTION

In this section, the various steps involved in the proposed framework are discussed. The proposed framework is illustrated in figure 4.

The experimental set-up for the current study was performed on a platform operated on a 64-bit Windows 11 Home Single Language OS with an x64-based processor architecture. The experiments were performed in the Jupyter Notebook environment provided by Anaconda version 3. The system utilized was equipped with an AMD Ryzen 5 5600 H processor with Radeon graphics, capable of operating at 3.30 GHz with 24 GB of RAM.

3.1 Data Collection

The monthly mean AOD data for a period of 19 years (Jan 2001 to Dec 2019) across Delhi derived from MODIS were accessible at the Giovanni online system at http://disc.sci.gsfc.nasa.gov/giovanni. The Giovanni online system was developed by the National Aeronautics and Space Administration's (NASA) Goddard Earth Sciences Data and Information Services Center (NASA GES DISC). The Giovanni platform serves as a web-based application providing Earth Science scientific data for the purpose of climate research and environmental monitoring applications.

3.2 Data Pre-Processing

Data pre-processing is a significant step as it aids in improving the accuracy and effectiveness of the models by cleaning and standardizing the data. There was no missing data present in the data. Prior to LSTM modeling, the data underwent min-max normalization as shown in the equation (11). This transformation results in the values lying between 0's and 1's, thereby maintaining the original distribution and making the data suitable for further analyses.

$$X_{norm} = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (11)$$

Here X is the original dataset, X_{max} and X_{min} are the maximum and minimum AOD values, and X_{norm} is the final normalized output.

3.3 Data Splitting

In order to check the predictive potential of the models, the data was split into training and testing with a 0.894 ratio, so that the monthly mean AOD data from January 2001 to December 2017 becomes the training data and the remaining data from January 2018 to December 2019 becomes the

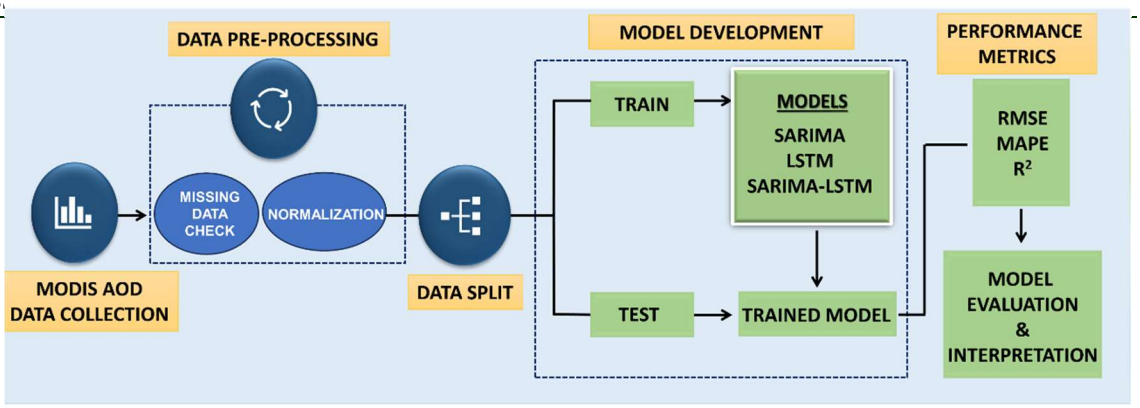

Figure 4. The framework of the proposed methodology

test data. During the training process, the models learn how to predict patterns, and their predictive capabilities are analyzed during the testing phase, depending on how they respond to the unseen data (test data).

3.4 Model Development

For SARIMA modeling, four steps are involved. The first step involves the identification of parameter values. We use seasonal differences in the data to identify the most suitable D values. Using ACF and partial autocorrelation function (PACF), the maximum parameter ranges for p and q are established. In the parallel grid search procedure, the Akaike information criterion (AIC) is used as an error measure in conjunction with the specified parameter range to choose the optimal model. The best-achieved model parameters are then used for estimation purposes. For the p, P, q, and Q components, the ranges were fixed within 0 to 2, and for the differencing parameters d and D, the ranges were varied from 0 to 1. The seasonality component, S, was set to 12 to reflect the monthly seasonality.

LSTM modeling use the grid search approach to determine the best hyperparameters. The grid search technique determines parameters such as batch size, maximum number of epochs, and optimizer type. The initial learning rate, activation function for state, and gate function are set to 0.001, hyperbolic tangent (tanh), and sigmoid, respectively. For AOD time series prediction, the best-identified model is used. Batch sizes range from 1 to 100. The maximum number of epochs varies between 10 and 300. The optimizers are set to Adam or RMSprop. The loss function is mean square error (MSE).

For hybrid modeling, the same approaches are followed. The residual from SARIMA is fed as the input for LSTM modeling.

3.5 Performance Metrics

To validate the performance of the models under consideration, the accuracy of the models has to be evaluated with respect to accuracy. For this purpose, the most prevalent three evaluation metrics—RMSE, MAPE, and R^2 which has been used extensively in the literature for AOD time series prediction are employed. These performance measures are determined as given in the equations (12), (13) and (14) as follows:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(d_i - \hat{d}_i)^2} \quad (12)$$

$$MAPE = \left(\frac{100\%}{n}\right)\sum_{i=1}^{n}\left|\frac{d_i - \hat{d}_i}{d_i}\right| \quad (13)$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(d_i - \hat{d}_i)^2}{\sum_{i=1}^{n}(d_i - \bar{d})^2} \quad (14)$$

where, n is the total number of observations, d_i is the actual AOD data at i^{th} sample, $\hat{d}_i$ is the predicted AOD data and $\bar{d}$ is the mean of the actual AOD data d.

4. RESULTS AND DISCUSSION

In this section, the data analysis and predictive capacities of the proposed model are compared to those of single models.

Table 1 provides the descriptive statistics of the monthly mean MODIS AOD_{550nm} considered in this study.

Table 1: Descriptive Statistics of Monthly Mean MODIS AOD_{550nm} Over Delhi (2001–2019).

Statistic	AOD Values
Count	228
Mean	0.775
Standard Deviation	0.266
Minimum	0.420
25th Percentile (Q1)	0.585
Median (50th Percentile)	0.716
75th percentile (Q3)	0.902
Maximum	1.856

There was no missing data. The typical level of AOD across the Delhi region is approximately 0.775, which is the average AOD value. The standard deviation of 0.266 indicates a moderate variability in AOD values from month to month. The AOD values vary from a minimum of 0.420 to 1.856, indicating variability due to seasonality and the presence of extreme events. The three-percentile values aid in understanding the spread of the AOD values and whether there is a presence of skewness. Since the median value is 0.716, it implies that half of the observation values are below this value. The Q1 and Q3 values of 0.585 and 0.902 indicate that the data is probably right-skewed.

Figure 5 illustrates the time series variation of monthly mean MODIS AOD_{550nm} data over the Delhi region from January 2001 to December 2019. The most noticeable observation from this plot is the peak value of AOD observed during the alternate years during the month of July. A similar observation was reported by Taneja et. al. (2016) over the Delhi region, and the possible reason for this phenomenon is the presence of sizeable dynamics in the lower atmosphere. An increasing trend in AOD values over the years can be noted. Every year in the winter (December–January–February), there is a decrease, and this has been linked to the higher rate of fine-mode particles during winters [12] .

In order to understand the underlying patterns, the ACF and PACF plots, as shown in figure 6, were considered. From figure 6 (a), it can be seen that there are some significant spikes at early lags, but the autocorrelations fall off quickly within the confidence interval (blue shaded region). Since the ACF is tailing off gradually, it is an indication of a potential autoregressive (AR) process. The lag repeats at every 12th position indicating the presence of seasonality. The PACF plot depicted in figure 6(b) shows no significant spikes outside the confidence interval, indicating higher lags aren't required in the AR part of the model.

To further understand whether there is a presence of trend and strong seasonality the decomposition of time series data is utilized. The monthly mean MODIS AOD_{550nm} is decomposed into three components, the trend, seasonality and residual. There is an increasing trend with some periods of slight decline. There exists a clear, regular oscillations indicating the seasonal fluctuations present in the data. The periodicity is consistent and strong, indicating a significant presence of seasonality within the AOD data. The residual or random plot is the plot obtained after the removal of seasonal and trend components. The residuals do not seem to have any pattern and thus the seasonal and trend components have been removed.

The monthly mean MODIS AOD_{550nm} across the Delhi region from 2001 to 2019 thus exhibits an increasing trend with strong seasonality. It demonstrates non-stationary behavior, implying that it has complex and changing patterns.

4.1 Comparison of the Proposed Model to SARIMA and LSTM Models

To validate the performance of the proposed model (SARIMA-LSTM), it was compared to its baseline models, SARIMA and LSTM. Initially, a plot comparing the actual values versus the predicted values for both the training and testing phases was performed. The performances were quantitatively analyzed using the performance metrics RMSE, MAPE, and R^2 during training and testing. Furthermore, the ACF of the residual obtained during the training process as well as the residual density distribution were also used to assess each of the model's performance.

In this study, the SARIMA model with configuration (2,1,2) and (2,0,2,12) was used as an outcome of having the lowest AIC using the parallel grid search method. For the standalone LSTM model, using grid search, the best hyperparameters were chosen. The model employed an extensive architecture with a total of 10 LSTM layers. The maximum number of epochs was 100, with a batch size of 32. The optimizer chosen was the Adam method with a learning rate of 0.001. For the LSTM model designed to predict the residuals from the

SARIMA outputs, the epochs were run for 100 with a batch size of 12, incorporating 5 layers to sufficiently model the complexity of the residuals. The Adam optimizer was chosen for its efficiency in handling sparse gradients on noisy problems, with a learning rate of 0.001 to ensure gradual and steady learning.

From figure 8, it can be observed that during both training and testing, the SARIMA model was able to effectively capture the seasonality and trend present in the data. The LSTM model was able to detect the pattern and follow the spiky nature of the data. However, the model is not consistent in predicting the peaks and troughs effectively. During both training and testing, the hybrid SARIMA-LSTM model was able to get a good fit with the actual data. This shows its potential to achieve more accurate predictions of AOD time series data.

From table 2, in the training phase, the hybrid SARIMA-LSTM model did better than the individual SARIMA and LSTM models, with an RMSE of 0.1508 compared to 0.1733 for SARIMA and 0.2162 for LSTM. In terms of MAPE, the hybrid model achieved a 13.5325% error rate, which is lower than SARIMA's 15.5007% and LSTM's 20.5125%. This indicates that the hybrid model was able to predict the training data with approximately 12.68% and 34.05% better accuracy than SARIMA and LSTM, respectively. The R^2 value for the hybrid model stands at 0.6328, which is higher than the 0.5798 of SARIMA and significantly higher than the 0.3449 of LSTM. This represents a 9.14% improvement in the coefficient of determination over SARIMA and an 83.51% improvement over LSTM during training, suggesting that the hybrid model explains the variability of the response data more effectively.

Model	RMSE	MAPE (%)	R^2
Training Performance			
SARIMA	0.1733	15.5007	0.5798
LSTM	0.2162	20.5125	0.3449
Proposed model	0.1508	13.5325	0.6328
Testing Performance			
SARIMA	0.1768	17.1807	0.6098
LSTM	0.2484	18.4296	0.2473
Proposed model	0.1471	13.0726	0.6867

Table 2: Training and testing performances of Monthly Mean MODIS AOD_{550nm} Over Delhi (2001–2019).

The testing phase results were consistent with the training performance, with the hybrid SARIMA-LSTM model outperforming both SARIMA and LSTM models. The RMSE for the hybrid model during testing was 0.1471, which is lower than SARIMA's 0.1768 and LSTM's 0.2484, indicating 16.82% and 40.79% improvement, respectively. In terms of MAPE, the hybrid model had a lower error percentage of 13.0726% compared to SARIMA's 17.1807% and LSTM's 18.4296%. This represents a 23.91% and 29.07% improvement in predictive accuracy over SARIMA and LSTM, respectively, during the testing phase.

The R^2 value of 0.6867 for the hybrid model indicates that the model explains approximately 12.62% more variability of the test data than SARIMA's 0.6098 and a substantial 177.88% improvement over LSTM's 0.2473. This suggests that the hybrid model not only predicts more accurately but also generalizes better to unseen data. These improvements validate the hybrid model's robustness and its enhanced ability to generalize to new data, surpassing SARIMA's and LSTM's performance under varying conditions.

As illustrated in table 2, the hybrid SARIMA-LSTM model performed better in the training phase than the SARIMA and LSTM models, as it had an RMSE of 0.1508 and 13.5325% MAPE. On the other hand, those for SARIMA and LSTM were 0.1733 and 15.5007% and 0.2162 and 20.5125%, respectively. The MAPE values indicate that the hybrid model performed with approximately 12.68% and 34.05% better accuracy in predicting the training data than SARIMA and LSTM. In addition, the corresponding R^2 value for the hybrid model in the training phase was 0.6328, which is higher than the 0.5798 for SARIMA and significantly higher than 0.3449 for LSTM. These values reveal an approximately 9.14% and 83.51% improvement in the coefficient of determination by the hybrid model compared to SARIMA and LSTM. In other words, the hybrid model explains the variability of the response data in the training phase better than the individual models.

Similarly, in the testing phase, the hybrid SARIMA-LSTM model outperformed the SARIMA and LSTM models as illustrated by the comparatively low RMSE of 0.1471 and MAPE of 13.0726%. SARIMA and LSTM models RMSE were, respectively, recorded as 0.1768 and 0.2484; MAPE of 17.1807% and 18.4296%. Therefore, approximately 16.82% and 40.79% improvement in

the accuracy of prediction are realized for the hybrid model over SARIMA and LSTM. Furthermore, the R^2 value of 0.6867 for the hybrid model shows that it explains about 12.62% more variability of the test data than SARIMA and a significant 177.88% improvement over LSTM. Therefore, the hybrid model not only predicts accurately but also generalizes better. Consequently, the hybrid model shows that it is more robust and generalizes better to new data.

The above improvements of performance metrics emphasize the robustness of the hybrid model and its improved generalization capabilities across varying conditions. Indeed, it outperformed not only SARIMA and LSTM.

Figure 9 depicts the ACF of residuals and their corresponding density distributions during the training process for SARIMA, LSTM, and SARIMA-LSTM models. These plots aids in further understanding and validating the performances of all the three models. A model is said to be a good fit if it has all of its autocorrelation coefficients within the confidence limits (the blue-shaded region in the ACF plots), which implies the residuals are white noise. Both the SARIMA and proposed models have autocorrelation coefficients within the confidence limits, thereby indicating the desirable property of handling time series prediction through the ability to capture temporal dependencies. The LSTM model's ACF plot shows stronger spikes outside the confidence interval. This suggests that the model's residuals still have some autocorrelation, which means that the LSTM might not be capturing all of the data's temporal dependencies.

When considering the residual density distribution plots in figure 9, both the SARIMA and hybrid SARIMA-LSTM models exhibit a normal distribution with a slight deviation. The hybrid model's residuals are more closely aligned with the normal distribution, as seen by the tighter fit of the histogram to the Kernel density estimation (KDE) curve, indicating a more accurate and consistent prediction of the training data. The LSTM residuals exhibit a more pronounced deviation from normality with heavier tails, potentially due to their inability to learn the temporal patterns effectively.

The proposed method of the hybrid SARIMA-LSTM model has a better overall performance as it can accommodate both linear and non-linear patterns of the time series data of AOD. The linear and seasonal patterns are effectively modeled by the SARIMA component and the residual containing the non-linear components is processed by the LSTM model. Finally, it combines both of these and gets a better and efficient model, compared to using SARIMA or LSTM alone. LSTM model when used alone fails in efficiently forecasting the AOD as it is unable to detect the underlying patterns and this might be attributed to the small amount of dataset available [39].

The significant improvement in RMSE, MAPE, and R^2 metrics during both training and testing phases implies the effectiveness of the proposed hybrid models. The higher R^2 obtained by the proposed hybrid model indicates the greater predictive potential. The absence of autocorrelation coefficients during the training and testing phases indicates that the model has successfully captured the underlying pattern and is providing a good fit.

5. CONCLUSION

The main objective of this study was to develop a hybrid model based on residual that can enhance the prediction accuracy of AOD time series data in the Delhi region. This study has demonstrated that the hybrid model outperforms the individual SARIMA and LSTM models, which have traditionally been employed within the literature to predict AOD, often failed to detect the underlying patterns efficiently. The results obtained reveals that the proposed model was able to enhance the accuracy of forecasting the AOD time series data considered. The hybrid model was able to achieve the lowest RMSE, MAPE and highest R^2 values, indicating the least error, highest accuracy, and best fit. The proposed hybrid model effectively captures both linear and non-linear patterns in AOD data, resulting in superior predictive capability. The individual SARIMA model, while better at linear dependencies, and the LSTM model, which handles non-linear patterns, both fall short when compared to the hybrid methodology. The stable performance of the hybrid model across training and testing phases highlights its strong capabilities and reliability for time series forecasting of AOD. The ACF and residual plots also further validates the efficiency of the proposed model. Thus, by combining the strength of linear and non-linear models, the proposed method was able to enhance the prediction accuracy. This enhanced modeling ability is essential for effective environmental management and health risk assessment in urban areas like Delhi, where aerosol concentrations can

have a significant impact on air quality and public health. The hybrid SARIMA-LSTM approach improves prediction accuracy while also providing deeper insights into the dynamics of aerosol distributions, allowing for more informed decision-making.

The proposed model as well as the standalone models faced issues in dealing with the random extreme events present in the data. Future studies can focus on developing effective approaches to handle these extreme events. Future works can incorporate more features, such as meteorological variables, to further increase forecasting accuracy. The study can also be extended to different geographical locations to provide a more detailed understanding of aerosol behavior and its implications and assess the performance of the proposed model.

This study advances the application of an effective and potent hybrid methodology for the efficient prediction of atmospheric aerosols, which can aid in a better understanding of climatic dynamics and decision-making policies for improving air quality and climate.

ACKNOWLEDGMENT:

The authors also acknowledge the Giovanni online data system (http://disc.sci.gsfc.nasa.gov/giovanni) for providing MODIS data utilized for the current study.

REFERENCES:

[1] R. J. Charlson *et al.*, 'Climate forcing by anthropogenic aerosols', *Science (1979)*, vol. 255, no. 5043, pp. 423–430, 1992.

[2] S. Twomey, 'Pollution and the planetary albedo', *Atmospheric Environment (1967)*, vol. 8, no. 12, pp. 1251–1256, 1974.

[3] B. A. Albrecht, 'Aerosols, cloud microphysics, and fractional cloudiness', *Science (1979)*, vol. 245, no. 4923, pp. 1227–1230, 1989.

[4] D. Chen, Y. Zhao, J. Zhang, H. Yu, and X. Yu, 'Characterization and source apportionment of aerosol light scattering in a typical polluted city in the Yangtze River Delta, China', *Atmos Chem Phys*, vol. 20, no. 17, pp. 10193–10210, 2020, doi: 10.5194/acp-20-10193-2020.

[5] A. K. Ranjan, A. K. Patra, and A. K. Gorai, 'A Review on Estimation of Particulate Matter from Satellite-Based Aerosol Optical Depth: Data, Methods, and Challenges', *Asia Pac J Atmos Sci*, vol. 57, no. 3, pp. 679–699, 2021, doi: 10.1007/s13143-020-00215-0.

[6] P. Stier *et al.*, 'The aerosol-climate model ECHAM5-HAM', *Atmos Chem Phys*, vol. 5, no. 4, pp. 1125–1156, 2005.

[7] M. Chin *et al.*, 'Light absorption by pollution, dust, and biomass burning aerosols: a global model study and evaluation with AERONET measurements', in *Annales Geophysicae*, Copernicus GmbH, 2009, pp. 3439–3464.

[8] K. Yürekli, H. Simsek, B. Cemek, and S. Karaman, 'Simulating climatic variables by using stochastic approach', *Build Environ*, vol. 42, no. 10, pp. 3493–3499, 2007, doi: https://doi.org/10.1016/j.buildenv.2006.10.046.

[9] B. Abish and K. Mohanakumar, 'A stochastic model for predicting aerosol optical depth over the north Indian region', *Int J Remote Sens*, vol. 34, no. 4, pp. 1449–1458, 2013, doi: 10.1080/01431161.2012.723149.

[10] K. Soni, S. Kapoor, K. S. Parmar, and D. G. Kaskaoutis, 'Statistical analysis of aerosols over the Gangetic-Himalayan region using ARIMA model based on long-term MODIS observations', *Atmos Res*, vol. 149, pp. 174–192, 2014, doi: 10.1016/j.atmosres.2014.05.025.

[11] K. Soni, K. S. Parmar, and S. Kapoor, 'Time series model prediction and trend variability of aerosol optical depth over coal mines in India', *Environmental Science and Pollution Research*, vol. 22, no. 5, pp. 3652–3671, 2015, doi: 10.1007/s11356-014-3561-9.

[12] A. S. Taneja K Ahmad K Attri SD, 'Time series analysis of aerosol optical depth over New Delhi using Box–Jenkins ARIMA modeling approach', *Atmos Pollut Res*, vol. 7, no. 4, pp. 585–596, 2016.

[13] K. Soni, K. S. Parmar, S. Kapoor, and N. Kumar, 'Statistical variability comparison in MODIS and AERONET derived aerosol optical depth over Indo-Gangetic Plains using time series modeling', *Science of the Total Environment*, vol. 553, pp. 258–265, 2016, doi: 10.1016/j.scitotenv.2016.02.075.

[14] M. Kumar *et al.*, 'Long-term aerosol climatology over Indo-Gangetic Plain: Trend, prediction and potential source fields', *Atmos Environ*, vol. 180, pp. 37–50, 2018, doi: 10.1016/j.atmosenv.2018.02.027.

[15] X. Li, C. Zhang, B. Zhang, and K. Liu, 'A comparative time series analysis and

modeling of aerosols in the contiguous United States and China', *Science of the Total Environment*, vol. 690, pp. 799–811, 2019.

[16] A. Abuelgasim, M. Bilal, and I. A. Alfaki, 'Spatiotemporal variations and long term trends analysis of aerosol optical depth over the United Arab Emirates', *Remote Sens Appl*, vol. 23, p. 100532, 2021, doi: https://doi-org.egateway.vit.ac.in/10.1016/j.rsase.2021.100532.

[17] A. Singh, S. Singh, A. K. Srivastava, S. Payra, V. Pathak, and A. K. Shukla, 'Climatology and model prediction of aerosol optical properties over the Indo-Gangetic Basin in north India', *Environ Monit Assess*, vol. 194, no. 11, 2022, doi: 10.1007/s10661-022-10440-x.

[18] M. Dutta and A. Chatterjee, 'A deep insight into state-level aerosol pollution in India: Long-term (2005–2019) characteristics, source apportionment, and future projection (2023)', *Atmos Environ*, vol. 289, 2022, doi: 10.1016/j.atmosenv.2022.119312.

[19] S. O. Nabavi, L. Haimberger, R. Abbasi, and C. Samimi, 'Prediction of aerosol optical depth in West Asia using deterministic models and machine learning algorithms', *Aeolian Res*, vol. 35, pp. 69–84, 2018, doi: 10.1016/j.aeolia.2018.10.002.

[20] M. Eltahan and K. Moharm, 'Atmospheric Aerosol Prediction over Egypt with LSTM-RNN using NASA's MERRA-2', in *2nd Novel Intelligent and Leading Emerging Sciences Conference, NILES 2020*, 2020, pp. 93–98. doi: 10.1109/NILES50944.2020.9257885.

[21] N. K. K. Panicker and J. Valarmathi, 'AOD Forecasting using Prophet Model across Four Major Urban Areas in India', in *2021 International Conference on Wireless Communications, Signal Processing and Networking, WiSPNET 2021*, 2021, pp. 400–405. doi: 10.1109/WiSPNET51692.2021.9419423.

[22] N. Daoud, M. Eltahan, and A. Elhennawi, 'Aerosol optical depth forecast over global dust belt based on LSTM, CNN-LSTM, CONV-LSTM and FFT algorithms', in *EUROCON 2021 - 19th IEEE International Conference on Smart Technologies, Proceedings*, 2021, pp. 186–191. doi: 10.1109/EUROCON52738.2021.9535571.

[23] K. Zaheer, S. Saeed, and S. Tariq, 'Prediction of aerosol optical depth over Pakistan using novel hybrid machine learning model', *Acta Geophysica*, vol. 71, no. 4, pp. 2009–2029, 2023, doi: 10.1007/s11600-023-01072-x.

[24] G. P. Zhang, 'Time series forecasting using a hybrid ARIMA and neural network model', *Neurocomputing*, vol. 50, pp. 159–175, 2003.

[25] M. Khashei and M. Bijari, 'A novel hybridization of artificial neural networks and ARIMA models for time series forecasting', *Appl Soft Comput*, vol. 11, no. 2, pp. 2664–2675, 2011.

[26] D. S. de O. S. Júnior, J. F. L. de Oliveira, and P. S. G. de Mattos Neto, 'An intelligent hybridization of ARIMA with machine learning models for time series forecasting', *Knowl Based Syst*, vol. 175, pp. 72–86, 2019.

[27] J. F. L. de Oliveira and T. B. Ludermir, 'A distributed PSO-ARIMA-SVR hybrid system for time series forecasting', in *2014 IEEE international conference on systems, man, and cybernetics (SMC)*, IEEE, 2014, pp. 3867–3872.

[28] J. F. L. de Oliveira, E. G. Silva, and P. S. G. de M. Neto, 'A Hybrid System Based on Dynamic Selection for Time Series Forecasting', *IEEE Trans Neural Netw Learn Syst*, vol. 33, no. 8, pp. 3251–3263, 2022, doi: 10.1109/TNNLS.2021.3051384.

[29] P. S. G. de Mattos Neto, G. D. C. Cavalcanti, D. S. de O. Santos Júnior, and E. G. Silva, 'Hybrid systems using residual modeling for sea surface temperature forecasting', *Sci Rep*, vol. 12, no. 1, p. 487, 2022, doi: 10.1038/s41598-021-04238-z.

[30] V. Singh, S. Singh, and A. Biswal, 'Exceedances and trends of particulate matter (PM2.5) in five Indian megacities', *Science of The Total Environment*, vol. 750, p. 141461, 2021, doi: https://doi-org.egateway.vit.ac.in/10.1016/j.scitotenv.2020.141461.

[31] K. Patel *et al.*, 'Sources and dynamics of submicron aerosol during the autumn onset of the air pollution season in Delhi, India', *ACS Earth Space Chem*, vol. 5, no. 1, pp. 118–128, 2021.

[32] '2 | 2023 World Air Quality Report'.

[33] T. Singh *et al.*, 'Climatological trends in satellite-derived aerosol optical depth over North India and its relationship with crop

residue burning: Rural-urban contrast', *Science of the Total Environment*, vol. 748, 2020, doi: 10.1016/j.scitotenv.2020.140963.

[34] M. D. King *et al.*, 'Cloud and aerosol properties, precipitable water, and profiles of temperature and water vapor from MODIS', *IEEE Transactions on Geoscience and Remote Sensing*, vol. 41, no. 2 PART 1, pp. 442–456, 2003, doi: 10.1109/TGRS.2002.808226.

[35] L. A. Remer *et al.*, 'Global aerosol climatology from the MODIS satellite sensors', *Journal of Geophysical Research: Atmospheres*, vol. 113, no. D14, 2008.

[36] G. E. P. Box, G. M. Jenkins, and G. C. Reinsel, *Time series analysis: Forecasting and control: Fourth edition*. 2013. doi: 10.1002/9781118619193.

[37] S. Hochreiter and J. Schmidhuber, 'Long short-term memory', *Neural Comput*, vol. 9, no. 8, pp. 1735–1780, 1997.

[38] J. F. L. De Oliveira and T. B. Ludermir, 'A hybrid evolutionary decomposition system for time series forecasting', *Neurocomputing*, vol. 180, pp. 27–34, 2016.

[39] P. S. Muhuri, P. Chatterjee, X. Yuan, K. Roy, and A. Esterline, 'Using a long short-term memory recurrent neural network (LSTM-RNN) to classify network attacks', *Information (Switzerland)*, vol. 11, no. 5, 2020, doi: 10.3390/INFO11050243.

Journal of Theoretical and Applied Information Technology
15th June 2024. Vol.102. No. 11

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

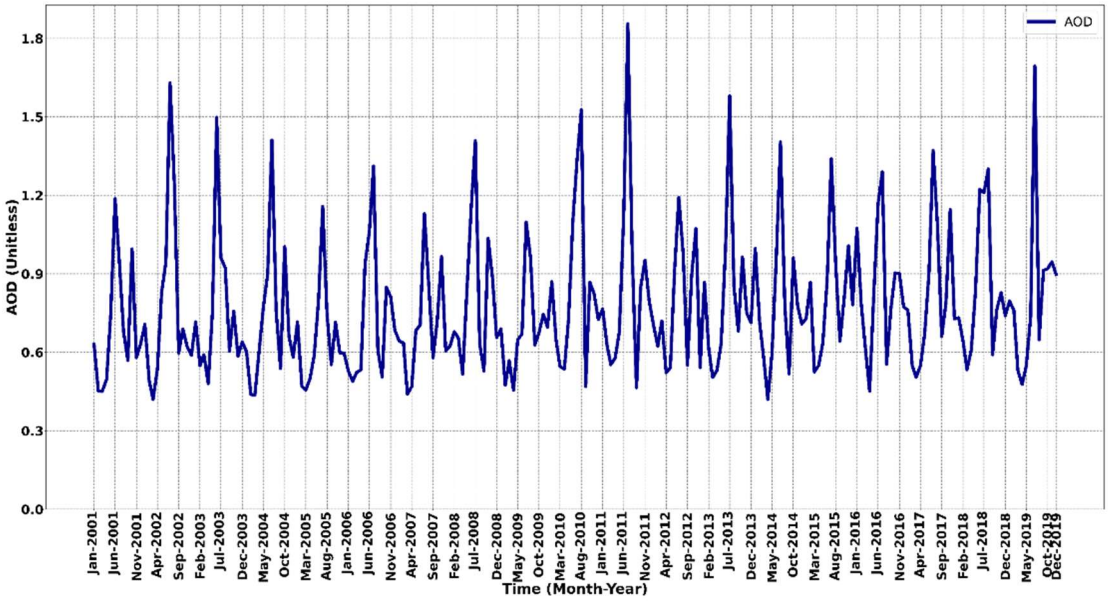

Figure 5. Monthly temporal variations in MODIS AOD_{550nm} Over Delhi (Jan 2001- Dec 2019).

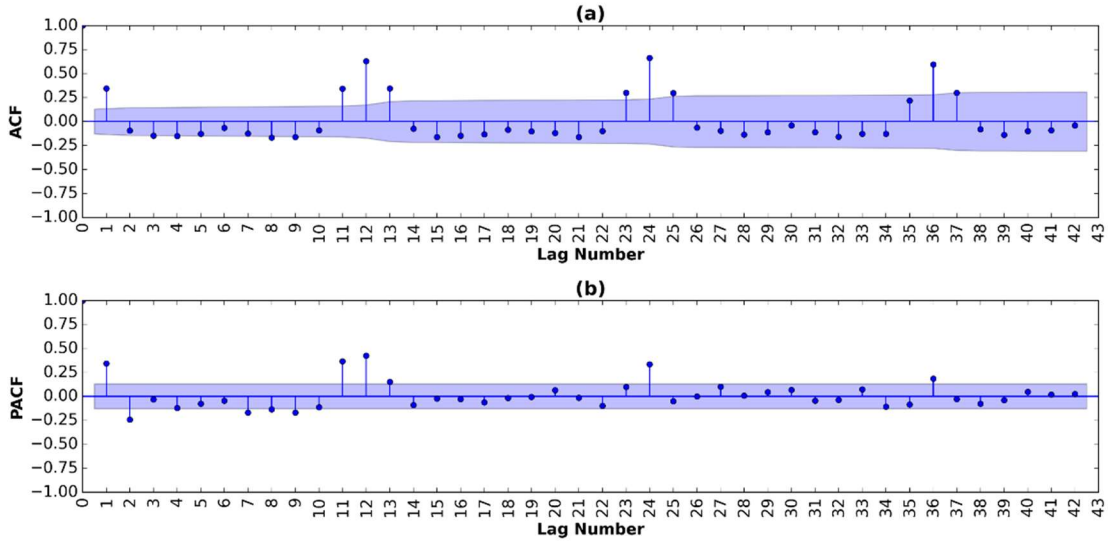

Figure 6: The ACF and PACF analysis of monthly MODIS AOD_{550nm} over Delhi from 2001 to 2019.

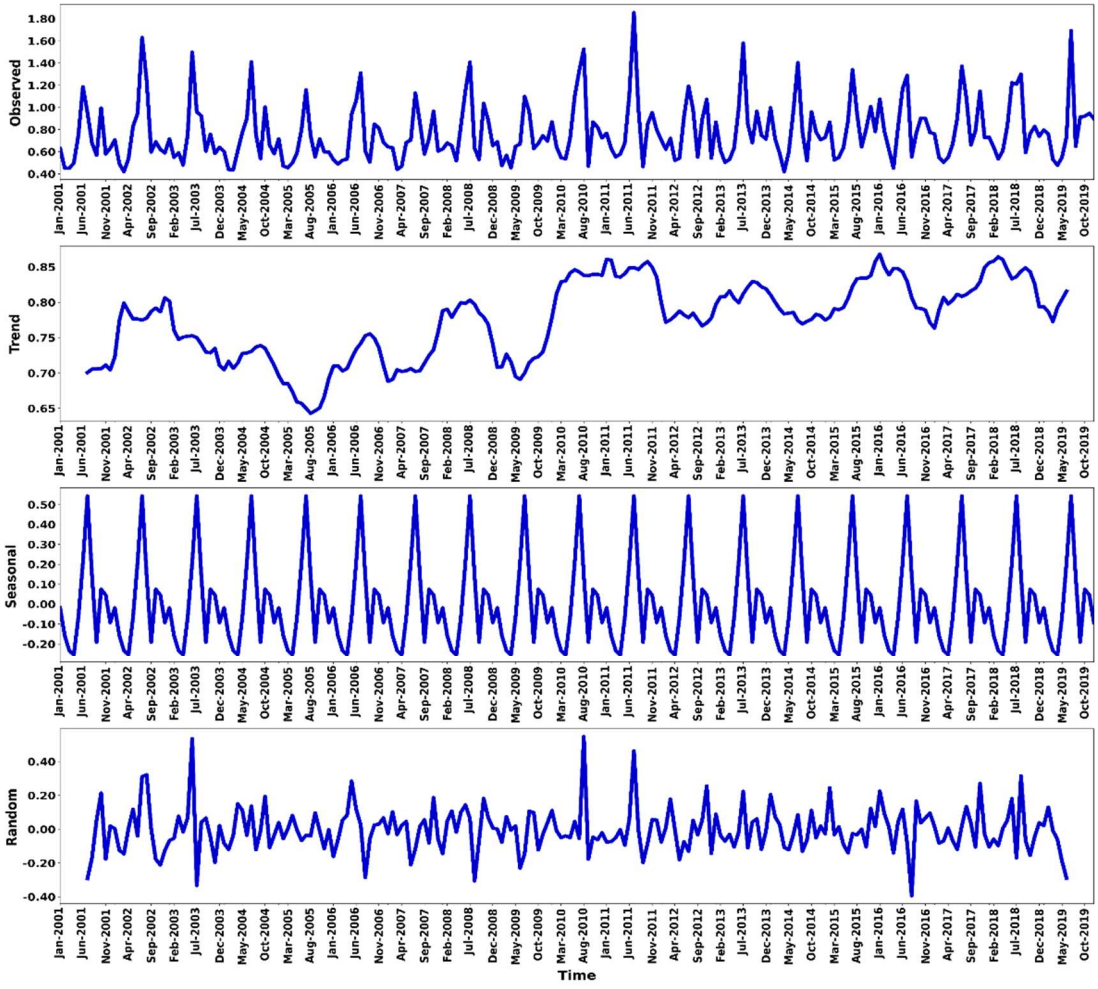

Figure 7: Decomposed time series analysis of monthly MODIS AOD_{550nm} over Delhi from 2001 to 2019.

Journal of Theoretical and Applied Information Technology
15th June 2024. Vol.102. No. 11

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

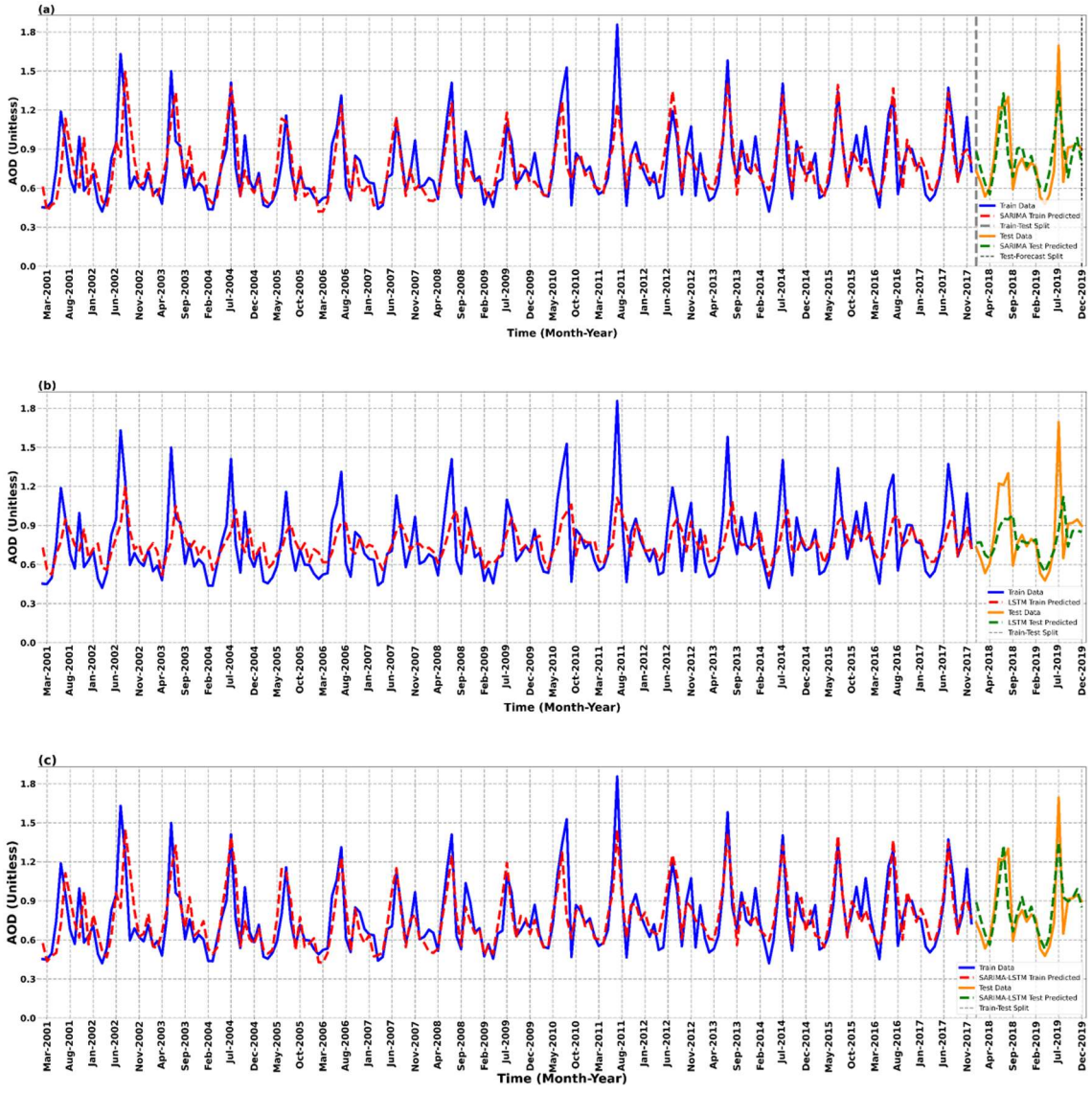

Figure. 8: The plot showing the actual, and predicted values of MODIS-AOD$_{550nm}$ during training (Jan 2001- Dec 2017) and testing (Jan 2018-Dec 2019) processes over the Delhi region using (a) SARIMA, (b) LSTM, and (c) SARIMA-LSTM. The grey dashed line indicates the Train-Test split.

Journal of Theoretical and Applied Information Technology
15th June 2024. Vol.102. No. 11

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

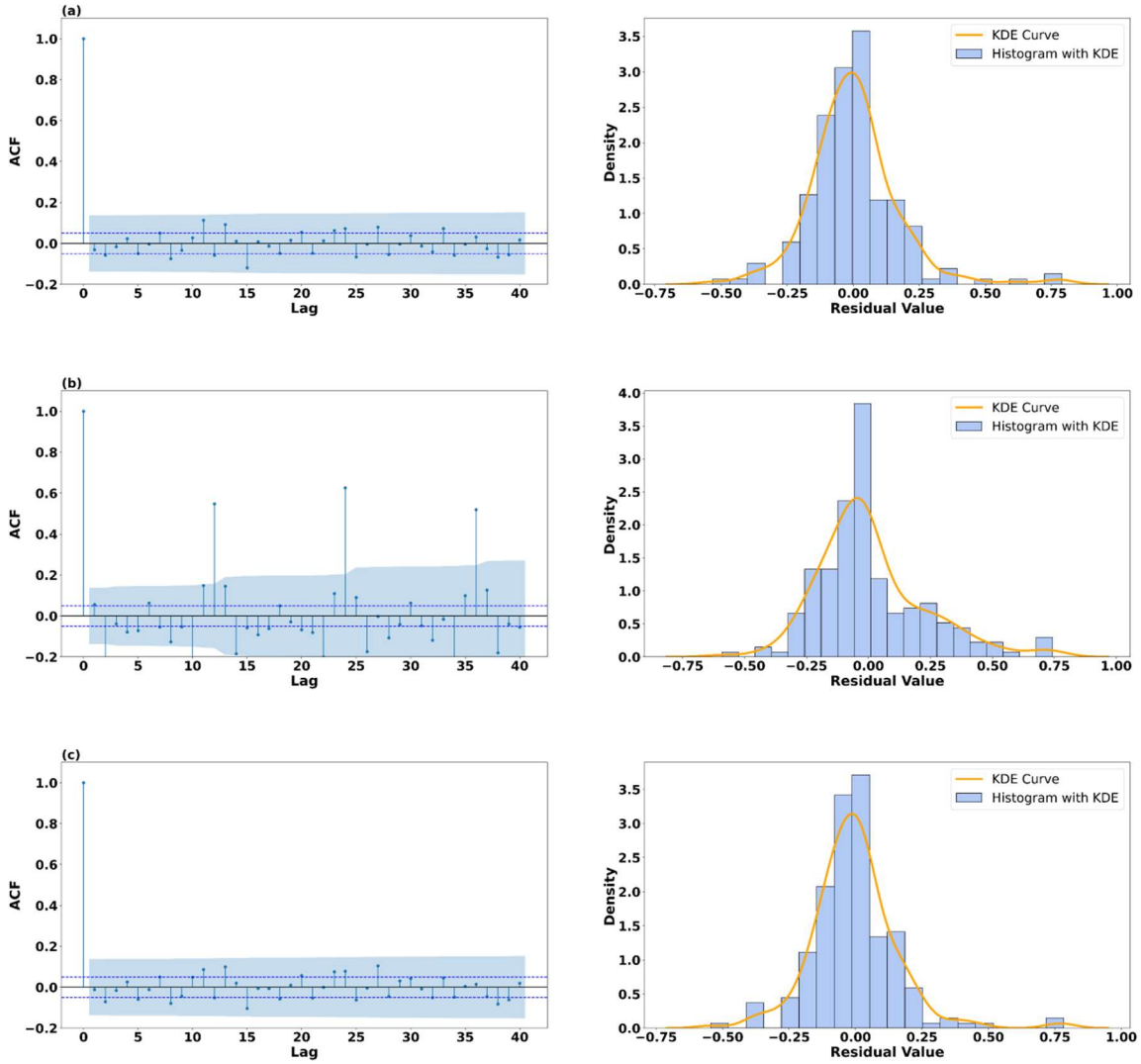

Figure. 9: The residual plot obtained for (a) SARIMA, (b) LSTM, and (c) SARIMA-LSTM during training (Jan 2001-Dec 2017) processes over the Delhi region.