*IOSR Journal of Environmental Science, Toxicology and Food Technology (IOSR-JESTFT)*
*e-ISSN: 2319-2402, p- ISSN: 2319-2399.Volume 20, Issue 9 Ser. 1 (September 2026), PP 16-22*
*www.iosrjournals.org*

# Time-Series Analysis and Forecasting of Air Quality Index in Selected Indian Cities: A Comparative Study of SARIMA and Holt-Winters Models

Tirupathi Rao Padi[1] ·Sarode Shirisha[1*] ·Ede Naga Anilkumar[2] ·Godide Meghana[3]

*Department of Statistics, Pondicherry University (A Central University), Puducherry, India*

***Abstract***
*Air pollution is a major environmental and public health concern in India, particularly in rapidly urbanizing cities. This study examines the temporal patterns and forecasts the Air Quality Index (AQI) for Patna, Kanpur, and Delhi using statistical time-series methods. Daily AQI observations from 2015-2020 were analyzed using exploratory analysis, time-series decomposition, the Augmented Dickey-Fuller (ADF) test, SARIMA, and Holt-Winters exponential smoothing. The results indicate persistent poor air quality and pronounced annual seasonality across the three cities, with the ADF test indicating stationarity in all three monthly AQI series. Overall, Holt-Winters outperformed the specified SARIMA model, particularly for Kanpur, while AQI-category analysis showed a predominance of Poor and Very Poor conditions. These findings highlight the usefulness of seasonal time-series forecasting for air-quality monitoring, environmental management, and public-health planning.*

***Keywords:*** *Air Quality Index; air pollution; time-series analysis; SARIMA; Holt-Winters; forecasting; stationarity; India*

---

Date of Submission: 03-09-2026 Date of Acceptance: 14-09-2026

---

## I. INTRODUCTION

Air pollution has become a major environmental and public-health concern in rapidly urbanising cities across India. The growth of industrialisation, transportation, urban development, and population has contributed significantly to the deterioration of ambient air quality. Major sources include industrial emissions, vehicular exhaust, construction dust, biomass burning, and other human activities. Meteorological factors such as temperature, humidity, wind speed, rainfall, and atmospheric conditions also influence the dispersion and accumulation of pollutants. Prolonged exposure to high levels of air pollution can adversely affect human health and increase the risk of respiratory and cardiovascular diseases. Therefore, continuous monitoring and accurate forecasting of air quality are essential for effective environmental management and timely intervention. The Air Quality Index (AQI) provides a simplified measure of overall air pollution based on major pollutants. Higher AQI values generally indicate poorer air quality and greater potential health risks. Analysing AQI as a time series helps identify long-term trends, seasonal variations, and irregular fluctuations in pollution levels. In the present study, AQI data from 2015 to 2020 are analysed for Patna, Kanpur, and Delhi to compare their temporal air-quality patterns. Descriptive statistical methods are used to summarise the characteristics and distribution of AQI observations. Time-series plots and decomposition techniques are applied to examine the trend, seasonal, and irregular components of the series. Monthly AQI data are further analysed to identify recurring annual seasonal patterns. For forecasting, the Seasonal Autoregressive Integrated Moving Average (SARIMA) model is employed to capture temporal dependence and annual seasonality. The Holt-Winters exponential smoothing method is also applied as a comparative forecasting approach. The performance of both models is assessed using appropriate forecasting accuracy measures and diagnostic tests. The comparison helps identify a suitable forecasting method for the selected cities. Overall, the study provides a systematic assessment of AQI patterns and forecasting performance, contributing to improved air-quality monitoring, environmental planning, and timely public-health interventions.

## II. LITERATURE REVIEW

Air pollution is a major environmental and public-health concern in India, particularly in rapidly urbanising cities. Industrial emissions, vehicular activities, particulate matter, biomass burning, and meteorological conditions contribute to deteriorating air quality. The adverse health effects and associated disease burden of air pollution in India have been widely documented [5,6], highlighting the importance of effective air-quality monitoring and forecasting.

The Air Quality Index (AQI) provides a simplified measure of overall air quality and is widely used for communicating pollution levels. Kumar and Goyal [1] demonstrated the applicability of statistical time-series methods for forecasting AQI in Delhi. Their subsequent studies examined principal component regression and neural network approaches, demonstrating the potential of statistical and nonlinear methods for air-quality prediction [2,3]. These studies established the importance of historical pollutant and AQI information for forecasting future air-quality conditions.

Time-series models such as ARIMA and SARIMA are particularly useful for air-quality data because they can account for temporal dependence and seasonal variation. Sharma et al. [8] demonstrated the applicability of ARIMA for forecasting $PM_{2.5}$ concentrations in the Delhi region. SARIMA extends the ARIMA framework by incorporating seasonal dependence and is therefore appropriate for AQI series exhibiting recurring annual patterns.

Holt-Winters exponential smoothing provides an alternative approach for modelling level, trend, and seasonal components. Hyndman et al. [12] established a state-space framework for exponential-smoothing methods, demonstrating their usefulness for seasonal forecasting. Wu et al. [11] further applied a grey Holt–Winters approach to AQI forecasting and reported improved forecasting performance, supporting the applicability of exponential-smoothing methods to air-quality data.

Recent studies have increasingly incorporated machine-learning and hybrid techniques for AQI prediction [3,9,15]. Although these approaches can capture complex nonlinear relationships, classical statistical models remain attractive because of their interpretability, relatively simple implementation, and transparent diagnostics. Stationarity is also an important consideration in ARIMA-family modelling, and the Augmented Dickey-Fuller test provides a standard method for assessing unit-root behaviour [13].

Despite considerable research on air-quality forecasting, comparatively fewer studies have provided a direct comparison of classical seasonal forecasting methods across multiple Indian cities using a common framework. Therefore, the present study analyses monthly AQI data for Patna, Kanpur, and Delhi during 2015–2020 using time-series decomposition, the ADF test, SARIMA, and Holt-Winters exponential smoothing. Forecasting performance is compared using RMSE, while AQI categories, inter-city correlations, and regression relationships are examined to provide a comprehensive statistical assessment of AQI behaviour across the selected cities.

## III. OBJECTIVES

The study aims to examine the trends and temporal patterns of the Air Quality Index (AQI) in selected Indian cities during the period 2015-2020. It seeks to identify and characterize the trend, seasonal, and irregular components of city-wise AQI and to assess the stationarity of the monthly AQI series using the Augmented Dickey-Fuller (ADF) test. The study further aims to develop AQI forecasting models based on Seasonal Autoregressive Integrated Moving Average (SARIMA) and compare their forecasting performance with Holt-Winters exponential smoothing. The accuracy of the forecasting models will be evaluated using the Root Mean Square Error (RMSE). In addition, the study examines AQI-category distributions, inter-city correlations, and simple regression relationships to understand similarities and differences in air-quality patterns across the selected cities. Finally, the study aims to provide statistical evidence that can support air-quality management, environmental monitoring, and urban planning decisions.

## IV. MATERIALS AND METHODS

4.1 Data Source

The study uses the "Air Quality Data in India" Kaggle dataset compiled by Rohan Rao. The data contain air-quality observations collected from monitoring stations and include city, date, AQI, AQI category, and pollutant measurements such as PM2.5, PM10, NO, $NO_2$, NOx, $NH_3$, CO, $SO_2$, $O_3$, benzene, toluene, and xylene.

**Table 1:** Principal dataset characteristics

| Characteristic | Reported information |
|---|---|
| Number of rows | 29,531 |
| Number of variables | 16 |
| Study period | 2015-2020 |
| Non-missing AQI | 24,850 |
| Non-missing xylene | 11,422 |
| Main outcome | AQI |

4.2 City Selection
The data contains an initial ranking in which Ahmedabad, Delhi, and Patna have the highest average AQI over the full dataset. After filtering to 2015-2020 and recalculating the ranking, the detailed time-series analysis uses Patna, Kanpur, and Delhi. The present article follows the latter set because these are the cities used in the decomposition, forecasting, and comparative analyses.

4.3 Data Preparation
The date variable was converted to date-time format, and the data were restricted to the 2015-2020 study period. Observations with missing AQI were removed. The reported preprocessing workflow used median imputation for selected pollutant variables. For time-series modelling, observations were aggregated to monthly mean AQI for each selected city.

4.4 Exploratory Data Analysis

Exploratory analysis included AQI distribution assessment, city-wise comparison, monthly trends, and major air pollutants.

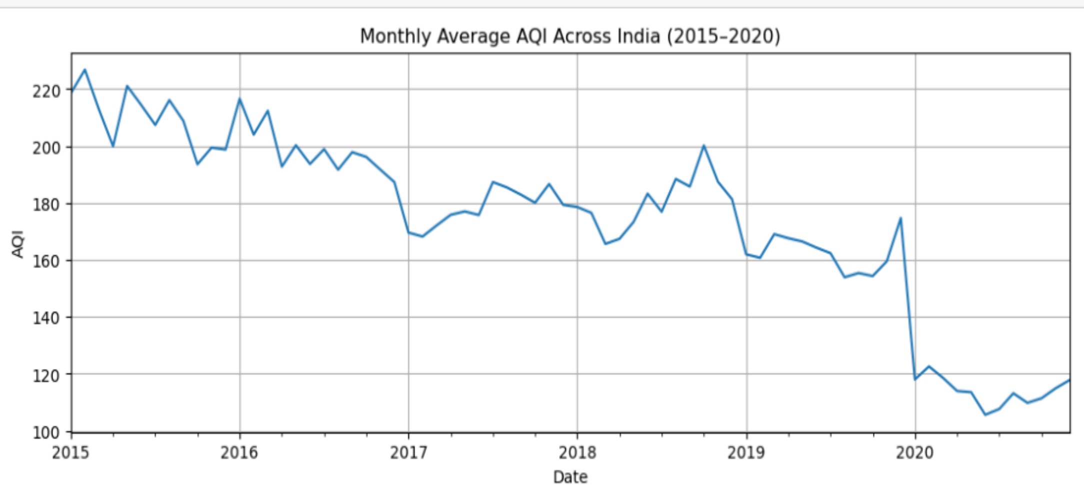


**Figure 1:** Monthly Average AQI Across India (2015–2020).

From Figure 1, A decreasing trend in AQI is observed over the study period, with a sharp decline in early 2020, likely associated with reduced emissions during the COVID-19 lockdown restrictions. The series also exhibits clear seasonal fluctuations, characterized by recurring periodic spikes and dips in AQI.

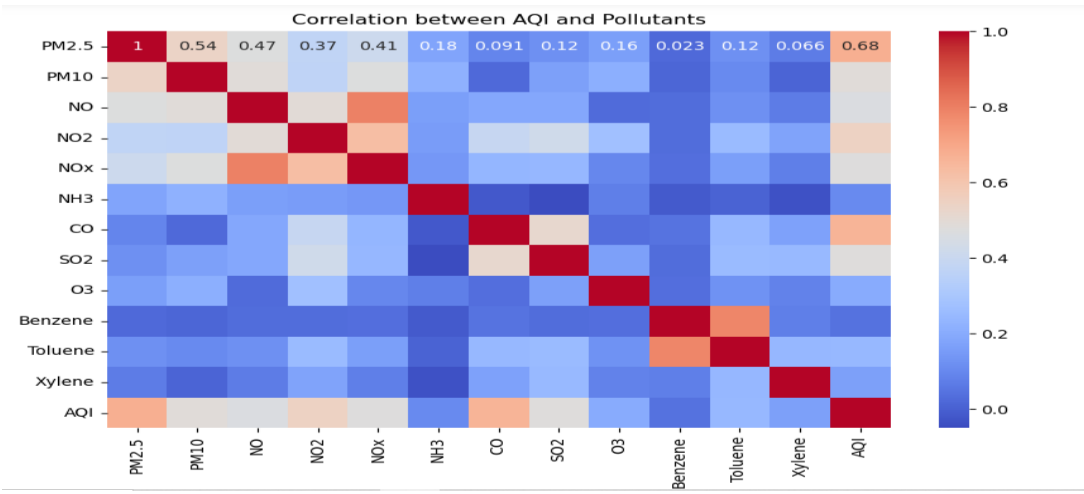


**Figure 2:** Pearson correlation matrix showing the relationships between AQI and major air pollutants

Figure 2 identifies $PM_{2.5}$ as the strongest reported positive correlate of AQI (r = 0.68), followed by $PM_{10}$ (r = 0.54). $PM_{10}$ also shows a moderate correlation (0.54). Gaseous pollutants such as CO, $SO_2$ and $O_3$ had weak correlations with AQI.

4.5 Time-Series Decomposition

Monthly city-level AQI series were decomposed into trend, seasonal, and residual components. Patna was reported to show a generally increasing trend with a consistent annual cycle. Kanpur exhibited a rise until around 2018, followed by a dip and more erratic seasonal behaviour. Delhi showed high AQI through approximately 2019, followed by a decline around 2020, together with a strong annual seasonal cycle and substantial residual variability.

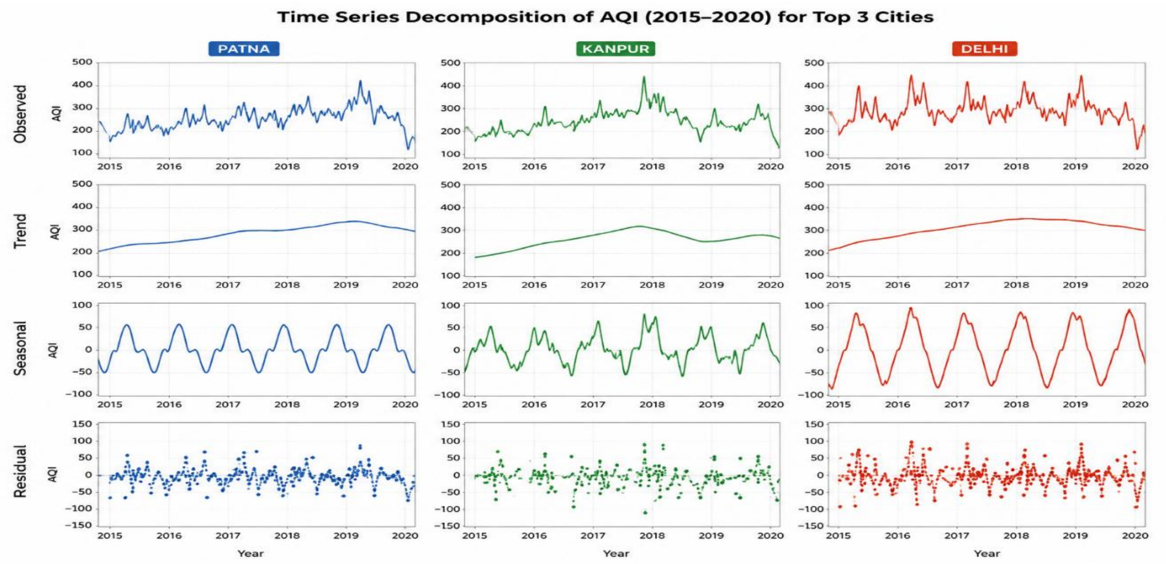


**Figure 3:** Time-series decomposition of monthly AQI (2015-2020) for Patna, Kanpur, and Delhi.

The decomposition demonstrates that seasonality is an important feature of AQI behaviour in all three cities, although its strength and regularity differ among them. Patna exhibits a comparatively stable seasonal pattern, Kanpur shows greater irregularity, and Delhi displays the strongest seasonal and residual variability. These differences indicate that AQI dynamics are influenced by both recurring seasonal factors

4.6 Stationarity Test

Stationarity was assessed with the Augmented Dickey-Fuller test. The null hypothesis states that the series has a unit root and is non-stationary, whereas the alternative hypothesis states that the series is stationary. The reported p-values are less than 0.001 for all three cities, leading to rejection of the null hypothesis.

**Table 2:** ADF stationarity results

| City | ADF statistic | p-value | Decision |
|---|---|---|---|
| Patna | −7.1573 | <0.001 | Reject $H_0$ |
| Kanpur | −7.4644 | <0.001 | Reject $H_0$ |
| Delhi | −9.4955 | <0.001 | Reject $H_0$ |

4.7 SARIMA Model

The Seasonal Autoregressive Integrated Moving Average model is represented as SARIMA (p,d,q) (P,D,Q,s), where p, d and q are the non-seasonal autoregressive, differencing and moving-average orders, P, D and Q are their seasonal counterparts, and s is the seasonal period. The reported model uses SARIMA (1,1,1) (1,1,1,12), with a 12-month seasonal cycle.

4.8 Holt-Winters and Forecast Evaluation

An additive Holt-Winters exponential-smoothing model with a 12-month seasonal period was used as a benchmark. Forecast performance was evaluated using root mean squared error (RMSE) lower RMSE indicates smaller forecast error. The reported comparison uses a 12-month forecasting horizon.

## V. Results and Discussion

5.0 Results

5.1 Forecasting Results

The SARIMA forecasts remain at relatively high AQI levels. Delhi is forecast to fluctuate approximately between 250 and 300, Kanpur begins above 300 and moves toward approximately 250-300, and Patna begins around 280 and remains broadly within the 250-300 range. Forecast uncertainty increases with the forecast horizon.
Holt-Winters forecasts retain recurring seasonal peaks and troughs. Patna remains approximately within 250-300, Kanpur shows a mild decline toward the end of the forecast window, and Delhi cycles approximately between 250 and 320. These patterns indicate that seasonal structure is an important component of AQI dynamics.

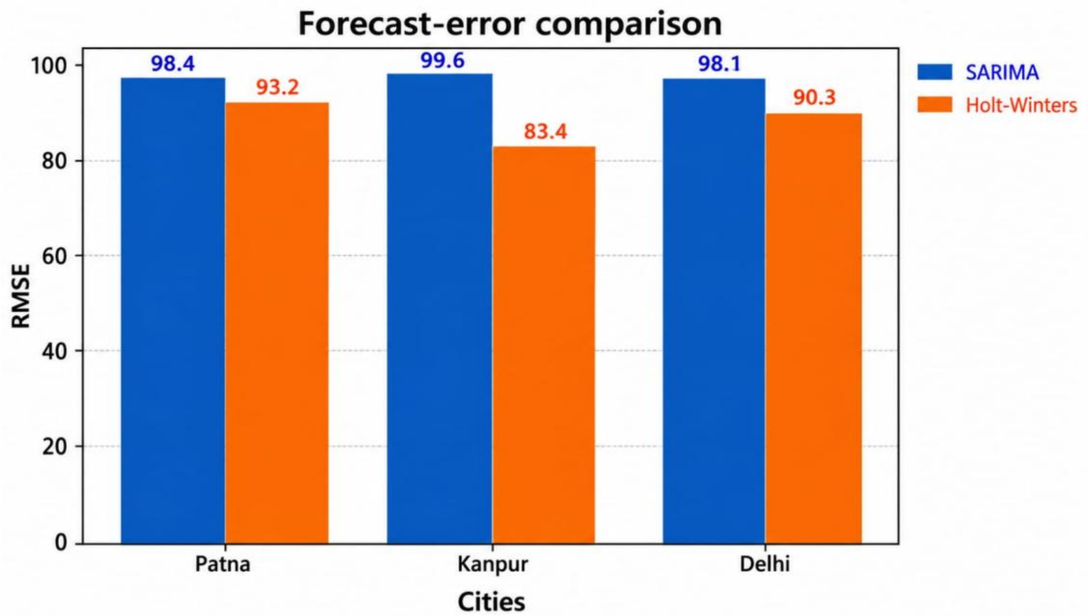


**Figure 4:** Reported RMSE comparison between SARIMA and Holt-Winters models.

Since lower RMSE indicates better forecasting accuracy, the Holt-Winters model is better than the SARIMA model in all three cities. The forecast-error comparison indicates that the Holt-Winters model outperforms the SARIMA model across Patna, Kanpur, and Delhi, as evidenced by consistently lower RMSE values. Therefore, based on RMSE, the Holt-Winters model provides relatively more accurate forecasts for the three cities.

**Table 3:** Forecast accuracy

| City | SARIMA RMSE | Holt–Winters RMSE | Better reported model |
|---|---|---|---|
| Patna | 99.09 | 93.44 | Holt-Winters |
| Kanpur | 98.95 | 83.31 | Holt-Winters |
| Delhi | 98.99 | 90.80 | Holt-Winters |

5.2 AQI Category Distribution

**Table 4:** Reported distribution of AQI categories

| City | Moderate (%) | Poor (%) | Very Poor (%) |
|---|---|---|---|
| Patna | 25 | 33 | 42 |
| Kanpur | 21 | 38 | 42 |
| Delhi | 19 | 42 | 39 |

From Table 4, Patna has 42% Very Poor, 33% Poor, and 25% Moderate observations. Kanpur has 42% Very Poor, 38% Poor, and 21% Moderate observations. Delhi has 39% Very Poor, 42% Poor, and 19% Moderate observations. Thus, unhealthy AQI categories dominate across all three cities.

5.3 Inter-City Correlation

**Table 5:** Reported Pearson correlations between city-level AQI series

| City pair | Correlation | Interpretation |
|---|---|---|
| Delhi - Kanpur | 0.20 | Low positive |

| | | |
|---|---|---|
| Delhi - Patna | 0.30 | Low positive |
| Kanpur - Patna | 0.15 | Very weak |

From Table 5, the highest reported correlation is 0.30 between Delhi and Patna, while Delhi-Kanpur and Kanpur-Patna correlations are lower. These weak relationships indicate that AQI dynamics are not strongly synchronized across the three cities.

5.4 Regression Analysis

**Table 6**: Reported OLS regression results

| City | $R^2$ | Coefficient | p-value | Interpretation |
|---|---|---|---|---|
| Patna | 0.001 | 0.0813 | 0.859 | Not significant |
| Kanpur | 0.003 | 0.1891 | 0.655 | Not significant |
| Delhi | 0.001 | −0.0388 | 0.927 | Not significant |

We observe from Table 6, that all three regression specifications have near-zero $R^2$ and statistically non-significant coefficients. The source analysis therefore concludes that the regression models do not provide useful predictive information in their current form.

5.5 Discussion

The analysis demonstrates a persistent pattern of elevated pollution levels, pronounced annual seasonality, and irregular temporal fluctuations across the selected cities. Delhi exhibits particularly strong seasonal and residual variability, reflecting substantial fluctuations in air quality throughout the study period. In contrast, Patna displays a relatively stable and consistent annual seasonal cycle, whereas Kanpur is characterized by greater irregularity and comparatively higher forecast uncertainty.

A key finding of the study is the comparative forecasting performance of the two approaches. Although the SARIMA model explicitly accounts for seasonal autocorrelation and is well suited to structured seasonal time-series data, the specified Holt-Winters exponential smoothing model produced lower reported RMSE values for all three cities. The most notable improvement was observed for Kanpur, where the RMSE decreased from 98.95 under SARIMA to 83.31 under Holt-Winters. This result suggests that the level-trend-seasonal structure captured by the Holt-Winters model was more closely aligned with the observed monthly AQI patterns over the evaluated forecast period.

Furthermore, the weak inter-city correlations and statistically non-significant regression relationships indicate that AQI dynamics are largely city-specific rather than governed by a common temporal pattern. Differences in local emission sources, meteorological conditions, geographical characteristics, and short-term environmental events may contribute to these distinct patterns. Therefore, the findings emphasize the importance of adopting city-specific approaches to AQI monitoring, forecasting, pollution control, and environmental management.

## VI. CONCLUSION

This study provides a statistical time-series assessment of Air Quality Index (AQI) patterns in Patna, Kanpur, and Delhi using monthly data from 2015 to 2020. The findings reveal consistently elevated levels of air pollution, distinct annual seasonal patterns, and substantial differences in AQI behaviour among the three cities. The Augmented Dickey-Fuller (ADF) test rejected the unit-root hypothesis for all three monthly AQI series, indicating stationarity and supporting their suitability for time-series analysis.

The forecasting results show that the specified Holt-Winters model achieved lower RMSE values than the selected SARIMA model for all three cities, with the most notable improvement observed for Kanpur. The AQI-category analysis further indicates that Poor and Very Poor air-quality conditions constitute a large proportion of the observations, highlighting the persistence of unhealthy air-quality conditions. In addition, the relatively weak correlations between the cities suggest limited cross-city predictability and indicate that local factors may play an important role in determining AQI variations.

Overall, the findings emphasise the importance of city-specific air-quality monitoring and forecasting for effective environmental management and public health planning. The study demonstrates the practical usefulness of classical time-series methods for identifying seasonal patterns and forecasting AQI. Future research should evaluate a broader range of SARIMA specifications, apply rolling-origin validation, incorporate relevant meteorological variables, and reproduce the final forecasting metrics using the complete original observed dataset. These improvements can strengthen the reliability of future AQI forecasts and provide more robust support for air-quality management and timely intervention.

**Declarations**

**Funding:** No specific funding information is reported in the supplied research paper.
**Conflict of interest:** No conflict of interest
**Data availability:** The study uses the "Air Quality Data in India" Kaggle dataset compiled by Rohan Rao.
**Code availability:** The supplied research paper used Python code for data preparation, visualization, forecasting, RMSE evaluation, correlation, and regression.
**Ethics approval:** Not applicable; the study uses environmental monitoring data and does not involve human participants.

## References

[1]. Kumar, A., & Goyal, P. (2011). Forecasting of daily air quality index in Delhi. Science of the Total Environment, 409(24), 5517–5523. https://doi.org/10.1016/j.scitotenv.2011.08.069.
[2]. Kumar, A., & Goyal, P. (2011). Forecasting of air quality in Delhi using principal component regression technique. Atmospheric Pollution Research, 2(4), 436–444. https://doi.org/10.5094/APR.2011.050.
[3]. Kumar, A., & Goyal, P. (2013). Forecasting of air quality index in Delhi using a neural network based on principal component analysis. Pure and Applied Geophysics, 170(4), 711–722. https://doi.org/10.1007/s00024-012-0583-4.
[4]. Goyal, P., Kumar, A., & Yadav, V. K. (2012). Forecasting of air pollutants in Delhi using different statistical techniques. Indian Journal of Air Pollution Control, 12(2), 57–66.
[5]. Khilnani, G. C., & Tiwari, P. (2018). Air pollution in India and related adverse respiratory health effects: Past, present, and future directions. Current Opinion in Pulmonary Medicine, 24(2), 108–116. https://doi.org/10.1097/MCP.0000000000000463.
[6]. India State-Level Disease Burden Initiative Air Pollution Collaborators. (2019). The impact of air pollution on deaths, disease burden, and life expectancy across the states of India: The Global Burden of Disease Study 2017. The Lancet Planetary Health, 3(1), e26–e39. https://doi.org/10.1016/S2542-5196(18)30261-4.
[7]. India State-Level Disease Burden Initiative Air Pollution Collaborators. (2021). Health and economic impact of air pollution in the states of India: The Global Burden of Disease Study 2019. The Lancet Planetary Health, 5(1), e25–e38. https://doi.org/10.1016/S2542-5196(20)30298-9.
[8]. Sharma, V., Ghosh, S., Mishra, V. N., & Kumar, P. (2024). Spatio-temporal variations and forecast of $PM_{2.5}$ concentration around selected satellite cities of Delhi, India using ARIMA model. Physics and Chemistry of the Earth, 134, 103849. https://doi.org/10.1016/j.pce.2024.103849.
[9]. Dhanalakshmi, S., & Poongothai, M. (2024). Design and synthesis of time series forecasting and deep learning prediction model for Air Quality Index prediction in Indian cities. Indian Journal of Engineering and Materials Sciences, 31(4), 589–600. https://doi.org/10.56042/ijems.v31i4.2132.
[10]. Agarwal, S., et al. (2020). Air quality forecasting using artificial neural networks with real-time dynamic error correction in highly polluted regions. Atmospheric Pollution Research.
[11]. Wu, L., Gao, X., Xiao, Y., Liu, S., & Yang, Y. (2017). Using grey Holt–Winters model to predict the air quality index for cities in China. Natural Hazards, 88(2), 1003–1012. https://doi.org/10.1007/s11069-017-2901-8.
[12]. Hyndman, R. J., Koehler, A. B., Snyder, R. D., & Grose, S. (2002). A state space framework for automatic forecasting using exponential smoothing methods. International Journal of Forecasting, 18(3), 439–454. https://doi.org/10.1016/S0169-2070(01)00110-8.
[13]. Dickey, D. A., & Fuller, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root. Journal of the American Statistical Association, 74(366a), 427–431. https://doi.org/10.1080/01621459.1979.10482531.
[14]. Bahadure, N. B., Gonge, S., Parashar, D., Shah, B., Patil, P. D., Renugadevi, M., & Raju, N. (2024). SVR machine learning and SARIMA-based air quality index classification and forecasting system. [Journal article].
[15]. Yenkikar, A., et al. (2025). Explainable forecasting of air quality index using a hybrid random forest and ARIMA model. MethodsX, 15, 103517. https://doi.org/10.1016/j.mex.2025.103517.
[16]. Bahadur, F. T., Shah, S. R., & Nidamanuri, R. R. (2026). Geospatial time series forecasting of air quality across Delhi using ARIMA and LSTM models. Journal of Atmospheric and Solar-Terrestrial Physics, 279, 106723. https://doi.org/10.1016/j.jastp.2026.106723.
[17]. Ranjan, H., Verma, P. K., Yadava, M. K., & Kumar, S. (2026). Multi-horizon forecasting of $PM_{2.5}$ concentrations in Delhi using a CNN-LSTM hybrid deep learning model. Advances in Space Research. https://doi.org/10.1016/j.asr.2026.07.106.