Analysis of Seasonality of Air Quality Index in Muzaffarnagar Using the Link Relative Method

Vanshika Saini[1], Richa[2]

[1]Student, MSc. Statistics, School of Statistics, Maa Shakumbhari University, Saharanpur, Uttar Pradesh, India.
[2]Assistant Professor, School of Statistics, Maa Shakumbhari University, Saharanpur, Uttar Pradesh, India.

Abstract

Air pollution has emerged as a serious environmental and public health issue in India, especially in industrial and agricultural regions like Muzaffarnagar, Uttar Pradesh. To understand its seasonal behavior, this study analyses the Air Quality Index (AQI) data from January 2022 to December 2024 using the Link Relative Method. Daily AQI values were first aggregated into monthly averages to reduce daily fluctuations and capture broader seasonal trends. The link relative method was then applied to capture the continuous monthly fluctuations, facilitating the derivation of seasonal indices that reflect recurring patterns in air quality. These indices reveal clear seasonal patterns as high AQI levels during winter months like November and December, and improved air quality during the monsoon months such as July and August. The analysis highlights how climatic factors and anthropogenic activities jointly influence pollution levels throughout the year. The findings of this study provide valuable insights into temporal pollution dynamics, which can help plan localized mitigation strategies and inform timely public health responses for better air quality management in Muzaffarnagar and similar regions.

Keywords: Air Quality, Air Quality Index (AQI), Air Pollution, Air Pollution Patterns, Monthly Averages, Link Relative Method, Chain Relatives, Seasonal Indices, Seasonal Variation, Time Series Analysis, Muzaffarnagar.

1. Introduction

Air quality is a critical determinant of both public health and environmental sustainability. Clean air is essential for human survival and plays a vital role in maintaining ecological balance [10, 21]. However, with increasing industrialization, urban expansion and intense agricultural activities, air pollution has emerged as a significant environmental challenge especially in rapidly developing countries like India. To assess and communicate pollution levels effectively, the Air Quality Index (AQI) has been developed as a standardized indicator. It translates the concentrations of multiple atmospheric pollutants into a single numerical value, providing an accessible and comprehensive measure of ambient air quality. According to the United States Environmental Protection Agency (EPA), the AQI includes six major pollutants: $PM_{2.5}$, PM_{10}, nitrogen dioxide (NO_2), sulfur dioxide (SO_2), carbon monoxide (CO), and ozone (O_3) [18, 20]. By converting their concentrations into sub-indices and determining an overall index based on the most severe pollutant, the AQI allows for easy interpretation of the severity of pollution and its associated health risks. The AQI is widely used by governments, researchers, and the public to monitor pollution levels and guide preventive measures. In the Indian context, many cities and towns experience persistent air quality issues due to both anthropogenic and natural factors. Muzaffarnagar, a key industrial and agricultural hub in Uttar Pradesh, frequently records elevated AQI values, particularly in the winter months. Sources such as vehicular emissions, industrial activities, road dust, and agricultural stubble burning contribute significantly to air pollution. These factors often push AQI readings into the 'Poor' (201–300), 'Very Poor' (301–400), or even 'Severe' (401–500) categories, posing serious public health risks. However, during the monsoon season, rainfall often helps clear airborne pollutants, resulting in temporary improvements in air quality.

To explore how AQI fluctuates throughout the year in response to such influences, it is important to adopt a statistical approach that can isolate seasonal effects. The Link Relative Method is one such technique that is widely used in time series analysis to identify and quantify seasonal variations [2]. In this method, monthly AQI values are compared using relative links from one month to the next, and average link relatives are calculated to develop seasonal indices. These indices help determine how specific months consistently affect air quality in comparison to others. The Link Relative Method is particularly effective in highlighting persistent seasonal effects that might be masked in raw data. For example, months like January and November may consistently show higher-than-average AQI values due to winter inversion and crop residue burning, while July and August may reveal lower values owing to the cleansing effect of monsoon rains.

In this study, monthly average AQI data from January 2022 to December 2024 for Muzaffarnagar has been analyzed using the Link Relative Method. The seasonal indices derived through this method provide clear insights into how air quality behaves in different months, allowing for a better understanding of both natural cycles and human-induced influences. By identifying these patterns, the study not only brings forth the cyclic nature of pollution but also contributes to more effective environmental planning, awareness campaigns, and public health strategies. Such analyses are vital for shaping data-driven, seasonally targeted policies that aim to reduce pollution and promote a healthier living environment in urban centers like Muzaffarnagar.

2. Review of Literature

Air pollution has become a serious issue in recent decades due to rapid urbanization, industrial growth, and vehicular emissions. To address the deteriorating air quality and its adverse effects on public health, numerous researchers have undertaken the task of analysing and forecasting air pollution levels using a range of statistical, machine learning, and hybrid approaches. These studies vary in scope, methodology, and geographical focus, but together they contribute to a broader understanding of air pollution dynamics and the applicability of different analytical models. Das et al. (2022) conducted a comparative evaluation of ten models to analyse and forecast short-term $PM_{2.5}$ concentrations across five different monitoring stations. They tested three forecasting horizons—daily, weekly, and monthly—and employed statistical metrics like RMSE, MAE, and MAPE. Their findings suggest that deep learning models, particularly LSTM, are effective for short-term forecasting, while k-Nearest Neighbors (kNN) performs well for longer-term projections. Traditional models such as ARIMA showed limited capacity to capture long-term variations, underscoring the adaptability of modern machine learning models to dynamic pollution patterns. In a study from Sofia, Bulgaria, Marinov et al. (2022) used time series methods to analyse air quality data between 2015 and 2019. They focused on pollutants such as CO, NO_2, O_3, and $PM_{2.5}$, applying ARIMA models at varying time granularities—ranging from 3 to 24 hours. The study incorporated imputation techniques for missing data and emphasized identifying whether pollution levels exceeded WHO thresholds. Tyagi et al. (2022) emphasized AQI analysis during the COVID-19 pandemic in Delhi, where they applied time series modelling techniques to study the behaviour of pollutants such as PM_{10}, $PM_{2.5}$, CO, SO_2, NO_2, NH_3, and O_3. Data from CPCB revealed notable improvements in air quality during the pandemic-related lockdown. With high R^2 values (e.g., 0.95 for PM_{10}), the models showed good fit, highlighting the relevance of such analyses for understanding air quality dynamics during periods of reduced anthropogenic activity. While the study was limited to one modelling approach, it stressed the importance of continuous AQI monitoring for public health. Bhatti et al. (2021) focused on Lahore, Pakistan, analysing $PM_{2.5}$ and PM_{10} levels and correlating them with O_3, NO, and SO_2. Using the HYSPLIT model, they traced pollution sources, identifying cross-border transport from Afghanistan as a key contributor. Their SARIMA-based analysis indicated a likely increase in particulate matter beyond 100 $\mu g/m^3$ in the coming year, raising concerns about non-compliance with Pakistan's NEQS. The study called attention to the necessity for bilateral cooperation and localized mitigation strategies. In another study, Nath et al. (2021) explored long-term pollutant trends in Kolkata using both traditional statistical and deep learning models. They compared AR, SARIMA, and Holt-Winters models with stacked LSTM, bidirectional LSTM, and convolutional LSTM models. Interestingly, statistical models outperformed deep learning ones, likely due to the limited size of the dataset. This emphasizes that while deep learning holds promise, its performance is highly data-dependent. The study also suggests that more comprehensive datasets may yield better results with deep learning. Gourav et al. (2019) applied ARIMA models to forecast SO_2 and NO_2 levels in Delhi on a seasonal and monthly basis. The study demonstrated the capability of ARIMA in modelling both stationary and non-stationary series and evaluated results using MSE, MAE, and RMSE. This research highlighted the potential of time series analysis in environmental monitoring and emphasized the need for early warnings to combat pollution surges, especially in densely populated urban areas. Samal et al. (2019) conducted a comparative study of SARIMA and Prophet models for analysing air pollutant concentrations in Bhubaneswar, specifically focusing on RSPM, SO_2, NO_2, and SPM. Their findings highlighted the limitations of traditional linear regression in capturing time-dependent structures, whereas SARIMA and Prophet models successfully identified seasonal trends and provided reliable forecasts. The study offers valuable insights for improving regional strategies in air quality assessment and planning. Lee et al. (2018) proposed a hybrid SARIMA–SVM approach to effectively model the combined linear and nonlinear patterns present in air pollution datasets from South Korea. Their model achieved significantly higher accuracy compared to traditional models, with improvements of approximately 20.81% for particulate matter (PM) and 43.77% for ozone (O_3) forecasts. The results emphasize the increasing importance of hybrid techniques in environmental modelling, especially when dealing with pollutants influenced by both straightforward trends and intricate interactions. Gocheva-Ilieva et al. (2014) focused their analysis on Blagoevgrad, Bulgaria, using PCA-based factor analysis and the Box–Jenkins methodology. The study grouped pollutants (NO, NO_2, NO_x, PM_{10}, SO_2, O_3) into three distinct sources and applied SARIMA models for short-term forecasting. Their results, which included variance-stabilizing transformations and BIC-based model selection, showed accurate predictions for up to 72 hours, offering a simple yet effective framework for smaller cities with limited resources. Lee et al. (2012) used SARIMA to analyse monthly and seasonal API trends in Johor, Malaysia. Their Box-Jenkins modelling approach revealed Pasir Gudang to be the most polluted, primarily due to industrial activity. The research emphasized the role of seasonal variability in pollution patterns and the importance of targeted interventions. It further illustrated the feasibility of using SARIMA for continuous environmental monitoring. Tsakiri and Zurbenko (2011) conducted an in-depth analysis of ozone concentration trends in Albany, New York. They decomposed the ozone time series into long-term, seasonal, and short-term components and used a vector autoregressive model along with the Kalman filter for short-term analysis. The study highlighted solar radiation and temperature as dominant predictors, demonstrating high R^2 values when these variables were considered. This multidimensional approach reinforces the need to include meteorological factors in air quality studies. Kumar and Jain (2010) used ARMA and ARIMA models to examine daily concentrations of CO, NO, NO_2, and O_3 at a traffic-heavy urban site in Delhi. To enhance model selection, they applied multiple criteria (AIC, BIC, FPE) and used ACF and PACF for parameter estimation. With MAPE values for one-day-ahead predictions falling between 12% and 24%, the forecasting accuracy was considered satisfactory. The study confirmed that with proper pre-processing and model selection, ARIMA models can provide reasonably accurate short-term forecasts.

These studies, when considered together, reveal that there is an increasing reliance on both time series and hybrid models for effective analysis and forecasting of air pollution levels in different contexts. While traditional statistical methods like ARIMA and SARIMA remain robust for structured data and shorter horizons, machine learning and deep learning techniques are increasingly effective, especially with larger datasets and non-linear pollutant behaviours. The integration of meteorological factors, seasonal and spatial data has proven to be critical in generating meaningful and implementable insights for urban planning and public health policy.

3. Research Gap

While several studies have examined air quality patterns in major metropolitan cities using advanced statistical and machine learning models, limited research has focused on smaller industrial and semi-urban regions like Muzaffarnagar. Moreover, much of the existing literature emphasizes forecasting rather than understanding the underlying seasonal behaviours that influence AQI fluctuations. This leaves a gap in identifying consistent monthly trends that could support preventive action before pollution spikes occur. Additionally, many time series analyses focus on short-term data or daily fluctuations, often overlooking the value of multi-year seasonal decomposition. In the context of this study, the application of the Link Relative Method on three years of AQI data offers a novel perspective, as it enables the clear identification of recurring seasonal patterns specific to this region.

4. Methodology

Research Design:

This study adopts a descriptive and exploratory research design to analyse the Air Quality Index (AQI) patterns in Muzaffarnagar over a specified period. The primary objective is to examine seasonal variations in AQI through time series analysis. The research focuses on understanding the temporal dynamics of air quality, offering insights into how pollution behave across different months and seasons.

Research Approach:

A quantitative research approach has been employed, utilizing classical time series analysis techniques. Specifically, the Link Relative Method has been applied to monthly average AQI data to extract seasonal indices. The Link Relative Method was chosen for its simplicity and effectiveness in isolating seasonal patterns from monthly AQI data, making it ideal for a three-year dataset with clear cyclic trends. Unlike complex models like SARIMA, it requires minimal computational resources and is well-suited for exploratory analysis in smaller regions like Muzaffarnagar. This method helps in identifying recurring seasonal patterns by calculating average link relatives and determining the extent to which each month contributes to fluctuations in air quality. The entire analysis was performed in Microsoft Excel, ensuring transparency in computation and ease in graphical representation of trends and seasonal behaviors.

Data Source and Collection:

The study is based on secondary data obtained from the Central Pollution Control Board (CPCB) [19]. Daily AQI values for Muzaffarnagar were collected for the period spanning January 1, 2022, to December 31, 2024. These daily figures were aggregated into monthly averages, which were then used for the application of the Link Relative Method to analyze seasonal patterns and determine monthly deviations in AQI. Any missing values in the dataset were carefully handled to ensure consistency and reliability in the time series analysis.

5. Statistical Analysis

To explore the seasonal patterns of air quality in Muzaffarnagar from 2022 to 2024, the study employed the Link Relative Method on monthly average AQI values. The daily AQI readings for each of the three years were first aggregated to calculate monthly averages, with each month's value representing the average AQI for that specific month and year. This smoothed out daily fluctuations and provided a clear representation of seasonal and temporal trends in air quality.

Next, to quantify the month-to-month changes, link relatives (LRs) were calculated using the formula:

$$\text{Link Relative} = \left(\frac{\text{AQI of Current Month}}{\text{AQI of Previos Month}}\right) \times 100$$

This formula was applied to each year’s monthly data individually to obtain the percentage change in AQI from one month to the next. Since January has no preceding month, link relatives were calculated only from February onward in each year. The resulting link relatives for corresponding months across the three years were then averaged to compute the Average Link Relatives for each calendar month.

Following this, chain relatives (CRs) were computed. The chain relative for January was set as 100 (base), and subsequent months were calculated using:

$$\text{Chain Relative}_{\text{Month}} = \left(\frac{\text{Average LR of Month} \times \text{CR of Previous Month}}{100}\right)$$

These chain relatives represented the cumulative influence of seasonal variations across successive months. However, since chain relatives can introduce upward or downward drifts, they were adjusted to reflect only seasonal variation.

To correct for this and to ensure that the seasonal indices reflect only cyclical behavior, the January chain relative was recalculated as follows:

$$\text{C} = \left(\frac{\text{LR of January} \times \text{CR of December}}{100}\right)$$

This new value C often differs from the assumed base of 100. A correction factor (d) was therefore introduced to eliminate this drift:

$$\text{d} = \frac{\text{C} - 100}{12}$$

Each subsequent chain relative was then adjusted using this correction factor:

$$\text{Adjusted CR for } \text{r}^{\text{th}} \text{ month} = \text{Original CR} - (\text{r} - 1) \times \text{d}$$

These adjusted CRs represent the pure seasonal component for each month, isolated from trend and irregular fluctuations.

Finally, to convert these adjusted CRs into seasonal indices, each adjusted CR was expressed as a percentage of the overall average of all adjusted CRs, ensuring the average of all indices equals 100.

$$\text{Seasonal Index} = \left(\frac{\text{Chain Relative}}{\text{Average of All Chain Relatives}}\right) \times 100$$

The final seasonal indices thus obtained represent the typical monthly deviations from the overall trend, offering insight into how air quality typically behaves in a given month.

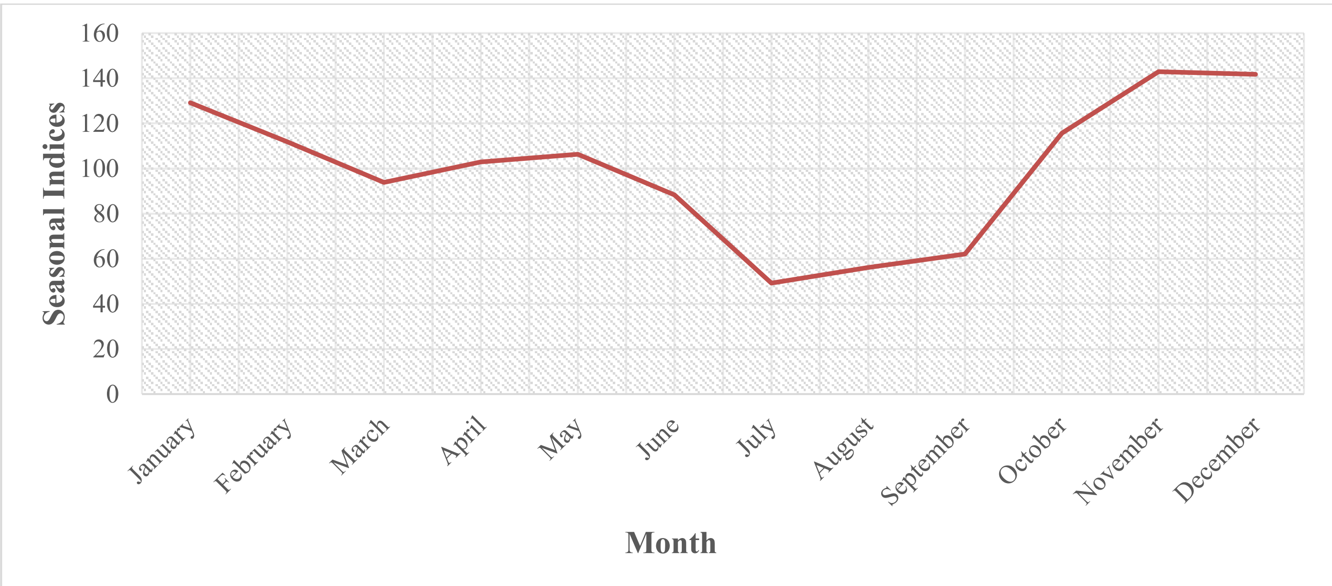
Table 1: Seasonal Indices (Link Relative Method)

Months	Seasonal Indices
January	129.1059
February	111.8077
March	93.8406
April	102.9772
May	106.3035
June	88.3066
July	49.1919
August	56.0680
September	62.0178
October	115.7280
November	142.9263
December	141.7265

*Calculated by author using AQI data from CPCB, 2022–2024

The above data is pictured in the following graph.

Figure 1: Monthly Seasonal Indices of AQI (2022–2024) calculated using Link Relative Method



The seasonal indices in Table 1 revealed a consistent pattern: higher values were observed during the winter months such as January and November, indicating worsening air quality likely due to low wind movement and increased emissions. In contrast, lower indices were observed during the monsoon season—particularly July and August—suggesting an improvement in air quality due to rainfall-induced pollutant dispersion.
As seen in Figure 1, the line graph representing the monthly seasonal indices of AQI (calculated using the Link Relative Method) clearly reveals a recurring and interpretable seasonal pattern. AQI values are significantly higher during the winter months—particularly in November, December, and January—indicating severe air pollution during this period. In contrast, the indices sharply decline during the monsoon months, with the lowest values observed in July, August, and September, reflecting improved air quality due to rainfall. The transition months, such as April, May, and October, show moderate seasonal effects. This pattern confirms the seasonal nature of air pollution in Muzaffarnagar, with winter months contributing most significantly to elevated AQI levels.

6. Results

The analysis of monthly Air Quality Index (AQI) values for Muzaffarnagar from January 2022 to December 2024, using the Link Relative Method, provided a clearer and more interpretable view of seasonal variations in air quality. The method revealed distinct monthly patterns that aligned more closely with environmental expectations. The seasonal indices derived from the link relatives indicated significant fluctuation across the months, reflecting the recurring influence of climatic and anthropogenic factors. The indices showed that air quality tends to deteriorate sharply during the winter months. November and December recorded the highest seasonal indices—142.9263 and 141.7265, respectively—clearly indicating peak pollution levels, likely driven by factors such as atmospheric inversion, stubble burning, and increased use of domestic heating. January also showed a high index of 129.1059, confirming the continuation of winter pollution. On the other hand, the most notable improvement in air quality was observed during the monsoon months—July, August and September—when the seasonal index fell to 49.1919, 56.0680 and 62.0178, respectively. This drop can be attributed to the cleansing effect of rainfall, which helps wash away suspended pollutants from the atmosphere. The transitional months of April, May, and October showed moderate indices ranging between 102.9772 and 115.7280, pointing to relatively average pollution levels. These variations indicate that air quality in Muzaffarnagar follows a strong seasonal cycle, with its peak clearly observed in winter and its lowest during the monsoon. These results validate the seasonal nature of air pollution in the region and highlight specific months where intervention or public awareness campaigns could be most impactful. The sharp contrast between seasons underscores the importance of incorporating seasonality into both monitoring and policy design for more effective air quality management.

7. Conclusion

This study provides a comprehensive understanding of the seasonal behaviour of air quality in Muzaffarnagar over a three-year period (2022–2024) using the Link Relative Method. By transforming daily AQI data into monthly averages and analysing the resulting seasonal indices, the research successfully captured the recurring patterns in pollution levels throughout the year. The findings clearly indicate that AQI levels rise significantly during the winter months—especially in November, December, and January—while the monsoon season consistently records the lowest levels of air pollution, likely due to natural atmospheric cleansing. The application of the Link Relative Method proved effective in isolating the seasonal component from the overall trend, offering a more accurate and interpretable picture of air quality fluctuations. These insights not only validate commonly observed pollution cycles but also reinforce the need for season-specific strategies in environmental management. Targeted measures during high-risk months, such as stricter emission control and public awareness campaigns, could significantly mitigate the health impacts of poor air quality.

Overall, this analysis highlights the importance of statistical approaches in environmental research and emphasizes how time series methods can inform more responsive and data-driven policy interventions aimed at promoting cleaner and healthier living conditions.

8. Futuristic Approach

To enhance the understanding and forecasting of air quality, future research should integrate meteorological parameters such as temperature, humidity, wind speed, and rainfall, as these factors significantly influence AQI levels. Incorporating advanced statistical and machine learning models like SARIMA, LSTM, or hybrid approaches could yield more accurate and dynamic forecasts. Additionally, the use of satellite data and remote sensing techniques can provide spatially detailed insights into pollution spread, especially in urban and semi-urban regions. Another promising direction is linking AQI data with public health indicators, hospital records, or respiratory illness trends, which can help in designing targeted awareness campaigns and timely intervention strategies. Such multidimensional analysis would support more informed environmental policymaking and promote sustainable urban living.

References

1. Saini, V., Richa, “Prediction of Air Quality Index Using Time Series Modelling: A Review Study”, International Journel For Multidisciplinary Research, Volume 7, March 2025.
2. Elhance, D.N., Elhance, R., Elhance V., “Fundamentals of Statistics”, 61th Edition, Kitab Mahal, 2024, 13.1-13.74.
3. Leon, C., “Environmental Studies: An Integrated Approach”, Callisto Reference, USA, 2023, 135-139.
4. Das, R., Middya, A. I., Roy, S., “High granular and short term time series forecasting of pm 2.5 air pollutant-a comparative review”, Artificial Intelligence Review, 2022,55(2), 1253-1287.
5. Marinov, E., Petrova-Antonova, D., Malinov, S., “Time series forecasting of air quality: a case study of Sofia City”, Atmosphere, 2022, 13(5), 788.
6. Tyagi, A., Gaur, L., Singh, G., Kumar, A., “Air Quality Index (AQI) Using Time Series Modelling During COVID Pandemic”, In International Conference on Artificial Intelligence and Sustainable Engineering: Select Proceedings of AISE 2020, April 2022, Volume 2 (pp. 441-452). Singapore: Springer Singapore.
7. Bhatti, U. A., Yan, Y., Zhou, M., Ali, S., Hussain, A., Qingsong, H., Yuan, L., “Time series analysis and forecasting of air pollution particulate matter (PM 2.5): an SARIMA and factor analysis approach”, 2021, Ieee Access, 9, 41019-41031.
8. Liao, K., Huang, X., Dang, H., Ren, Y., Zuo, S., Duan, C., “Statistical approaches for forecasting primary air pollutants: a review”, Atmosphere, 2021, 12(6), 686.

9. Nath, P., Saha, P., Middya, A. I., Roy, S., “Long-term time-series pollution forecast using statistical and deep learning methods. Neural Computing and Applications”, 2021, 1-20.
10. Patel, L., Friedman, E., Johannes, S. A., Lee, S. S., O'Brien, H. G., Schear, S. E., “Air pollution as a social and structural determinant of health”, *The Journal of Climate Change and Health*, 2021,*3*, 100035.
11. Gourav, Rekhi, J. K., Nagrath, P., Jain, R., “Forecasting Air Quality of Delhi Using ARIMA Model”, Advances in Data Sciences, Security and Applications: Proceedings of ICDSSA 2019, 2019, 612, 315.
12. Samal, K. K. R., Babu, K. S., Das, S. K., Acharaya, A., “Time series based air pollution forecasting using SARIMA and prophet model”, In proceedings of the 2019 international conference on information technology and computer communications, August 2019, 80-85.
13. Lee, N. U., Shim, J. S., Ju, Y. W., Park, S. C., “Design and implementation of the SARIMA–SVM time series analysis algorithm for the improvement of atmospheric environment forecast accuracy”, Soft Computing, 2018, 22, 4275-4281.
14. Gocheva-Ilieva, S. G., Ivanov, A. V., Voynikova, D. S., Boyadzhiev, D. T., “Time series analysis and forecasting for air pollution in small urban area: an SARIMA and factor analysis approach”, Stochastic environmental research and risk assessment, 2014, 28, 1045-1060.
15. Lee, M. H., Abd Rahman, N. H., Latif, M. T., Nor, M. E., Kamisan, N. A. B., “Seasonal ARIMA for forecasting air pollution index: A case study. American Journal of Applied Sciences”, 2012, 9(4), 570-578.
16. Tsakiri, K. G., Zurbenko, I. G., “Prediction of ozone concentrations using atmospheric variables”, Air Quality, Atmosphere & Health, 2011, 4, 111-120.
17. Kumar, U., Jain, V. K., “ARIMA forecasting of ambient air pollutants (O 3, NO, NO2 and CO)”, Stochastic Environmental Research and Risk Assessment, 2010, 24, 751-760.
18. EPA (2005). Six common air pollutants. U. S. environmental protection agency. https://www.epa.gov/criteria-air-pollutants
19. Central Pollution Control Board (CPCB), Government of India. https://cpcb.nic.in/
20. https://www.nasa.gov/general/what-is-air-quality/
21. https://soe.epa.sa.gov.au/environmental-themes/air/importance-of-clean-air
22. https://muzaffarnagar.nic.in/