Monitoring Air Pollution Impacts of COVID-19 in India

Mimansha AGRAWAL[1]*, Mamta AGRAWAL[2], Zafer ASLAN[3],

İlknur DÖNMEZ[4], Ali GÜNEŞ[5]

Abstract

Objective: Air quality is an important issue that is affecting human health, flora, fauna, and ecosystems. Vehicular pollution, demolition, construction, and population burst are the main reasons for degrading air quality. Hence, the main aim of the study is to monitor, analyse, as well as predict the air pollution component i.e., particulate matter before and after the lockdown of COVID-19 in various Indian cities of India.

Methods: The data collected is from Central Pollution Control Board (CPCB) of India from 2015-2020 for the pollutants such as PM2.5 and PM10. Pre-Processing techniques such as removal of null values using MICE imputer on multivariate data, merging the columns, and minmax feature scaling is applied. The cities are classified and further subdivided based on Terrain (coastal, mountainous and plains) and based on Population (highly populated and less populated). The study predicts the values for Particulate matter by using the Seasonal Autoregressive Integrated Moving Average (SARIMA) model using optimal hyperparameters for better prediction.

Results: The presented system for Air Quality Index (AQI) values includes 2020 data, and the forecasted results were highly optimistic. Further, this scenario is pure since 2020 is an outlier. A better prediction may be possible by skipping 2020. The performance of the forecasting model is calculated by the Root Mean Squared Error (RMSE) error which may help control the degraded air quality. The predicted AQI values for the year 2021 are predicted through SARIMA and produced 18.301 RMSE value for Lucknow and 46.198 for Delhi.

[1] *VIT Bhopal University, School of Computer Science and Engineering, Bhopal, India*
[2] *VIT Bhopal University, School of Advance Sciences (Mathematics), Bhopal, India*
[3] *Istanbul Aydin University, Computer Engineering Department, Istanbul, Turkey*
[4] *Istanbul Arel University, Computer Science Department, Istanbul, Turkey*
[5] *Istanbul Aydin University, Computer Engineering Department, Istanbul, Turkey*

* Corresponding author
E-mail: mimanshaagrawal1998@gmail.com; ORCID: 0000-0001-7545-390X

Mamta Agrawal: mamta.agrawal@vitbhopal.ac.in; ORCID: 0000-0001-5738-3450,
Zafer Aslan: zaferaslan@aydin.edu.tr; ORCID: 0000-0001-7707-7370,
İlknur Dönmez: ilknurdonmez@aydin.edu.tr; ORCID: 0000-0002-8344-1180,
Ali Güneş: aligunes@aydin.edu.tr; ORCID: 0000-0003-0220-5475

Conclusion: COVID-19 restricted human activities in 2020, and the air quality level significantly improved. This paper presents a model to forecast AQI values in various Indian cities for the coming year by considering surge reduction in various pollutant levels. In Future work, a hybrid approach of models can be followed to make predictions.

Keywords: COVID-19, Air Quality, Particulate Matter, Data Analysis, Environmental Air Pollutants

Introduction

COVID-19 first made contact in Wuhan, China, in December 2019. SARS-CoV-2 (Severe Acute Respiratory Syndrome coronavirus 2) was declared a global pandemic by the World Health Organization (WHO), and the world entered a state of lockdown. In India, the first case reported was on 30th January 2020, followed by a countrywide lockdown from 25th March 2020 (1). Total mortalities due to coronavirus has already surpassed three million. The virus has reportedly mutated, making it even more dangerous. Multiple variants of coronavirus are active throughout the globe. Every variant goes through gradual mutation depending on the regional conditions and show slightly different properties (2). Infection and mortality rates vary depending on the variant as different variants react differently to certain regional conditions. Transmission and mortality rates have proven to be influenced by several factors. These factors include air temperature, wind speed, pollution, and other geographical conditions (3). PM2.5 is the main culprit in cases of air pollution affecting the pandemic. Even in the lockdown period, the levels of PM2.5 emission are still alarming. The main reason for this is the crop burning and forest fires, which have plagued multiple countries in 2020. Particulate matter in the air act as a carrier for the virus, increasing transmission rates (4). The damage done to the immune system increases the mortality rate due to coronavirus, as the host becomes more susceptible to it. Previous studies show that the severity of air pollution correlates to the severity of COVID cases.

Although the coronavirus has forced us to enter our homes for about half a year, pollution levels in the country have not diminished (5). In fact, a recent Greenpeace report suggests that nearly 2 lakhs people were killed by air pollution in serval metro cities such as Delhi, Mumbai etc. despite severe closures (6). An estimated one million deaths each year and an estimated 9.8 lakhs preterm births, equivalent to an annual economic loss of 10.7 lakh crores, are caused by air pollution from mineral oil in India, according to the report. India is estimated to bear losses amounting to 5.4 percent of India's GDP annually, the third highest cost from global air pollution (7). The current study was conducted to determine the difference between the pollution screening (PM10 and PM2.5) before and after the closing time (March 25 to June 15, 2020) and the comparable period

from 5 years ago, i.e., 2015 to 2019 to assess the impact of air quality closures in major cities of India (8).

Data Collection and Methodology

The database is taken from the World Air Quality Index historical data platform. The filtered database contains 29,531 cases for 23 Indian cities. In addition, the data includes polluting air concentrations from various cities in India arranged by day. Recorded samples were collected from January 2015 to July 2020. The reason why these cities are being considered is because these cities are centres of air pollution. Details are analysed how air pollution varies before and after the closing time. The study will determine the results of which city was most polluted and before and after the closure.

In 2016, the Government of India, under the 'Swachh Bharat Mission', introduced the 'National Air Quality Index' (NAQI) which is a published regularly by the Central Pollution Control Board (CPCB). The AQI monitoring network consists of monitoring eight major parameters for calculating the amount of air index, while the AQI personal monitoring network focuses mainly on PM10 and PM2.5 pollution. The database contains air quality data and AQI (Air Quality Index) for cities at the daily level of various stations in many Indian cities from 2015-2020.

Data Pre-Processing

In the pre-processing step, unwanted rows and null values are deleted from the dataset. This Dataset has many columns with a large number of missing values. This missing data does not affect predictions because the average value of a particular column is used to estimate Indian AQI values between 29,531 cases.

Manage and Remove Null values

Through the Data analysis, it reveals that there are many missing values. In particular, a few cities have large numbers of missing values. In the AQI column, full details can only be obtained for Delhi. So, by calculating the percentage of missing values in each column, it helps in null values removal. (Multivariate imputation with chain-binding measurements) is used in the removal of null values in multivariate data. Minmax feature scaling is performed if needed.

Air Pollutants Pattern Visualization

Air pollution level in India is analysed using Python 3.6 data viewing tools to gain insight into patterns. Figure 1 provides a better overview of pollution levels (PM10 and PM2.5) in terms of months and years. Monthly data displayed in Figure 1 shows a significant decrease in air pollution in July and August. Such an effect is likely due to the monsoon season, which occurs during these months. A significant decrease in air pollution is observed in about June, when there is a

slight rise to very high levels during the winter. This decline, height, and altitude can be attributed to the burning of crops in northern parts of India.
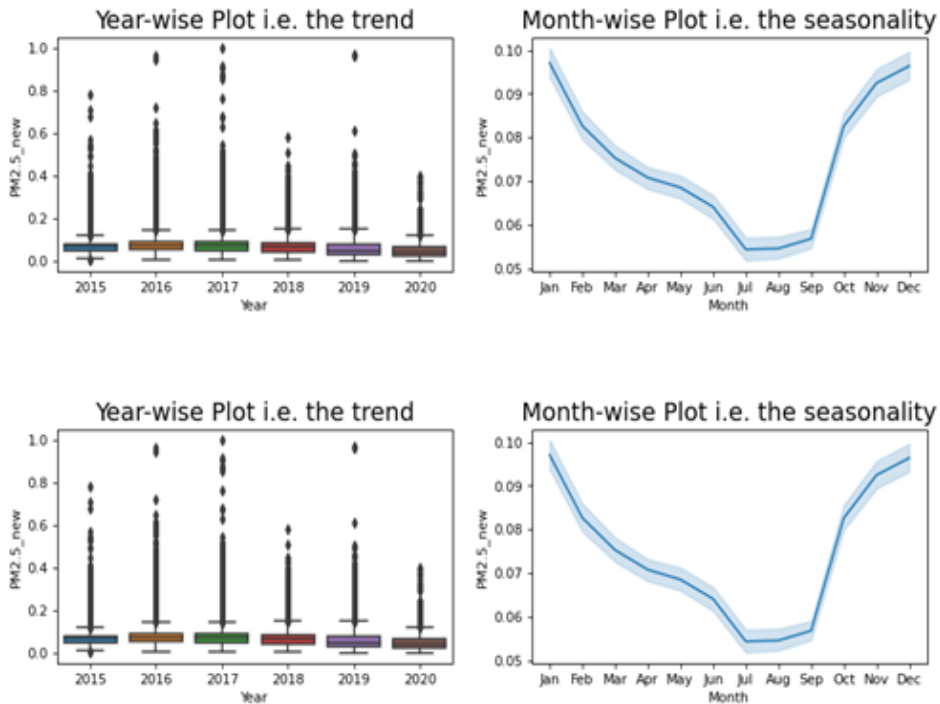

Figure 1. *Concentration level measured in terms of months in India for PM2.5 and PM10*

Analysis of the Pollutants

India remains one of the most polluted countries in the world and this has crippled the world war against coronavirus. The issue of particulate matter in air pollution is reported to act as a coronavirus carrier and increase transmission rates. Investigators have focused on PM2.5 million air pollution which is responsible for the estimated 4.2 million premature deaths each year worldwide. (9) This includes the deaths of more than a million people in China, more than half a million in India, nearly 200,000 in Europe, and more than 50,000 in the United States. Humans can also indirectly create PM10 by a combination of atmospheric chemicals. Some gases can withstand chemical reactions and form particulate matter, such as sulphur dioxide forming sulphates. While other coarse particles are subjected to such a process, these "second" particles are more common in PM2.5. As shown in Figure 2, Particulate Matter (PM2.5, PM10, etc.) has a

modest reduction all that time is still the cause of deaths related to air pollution.
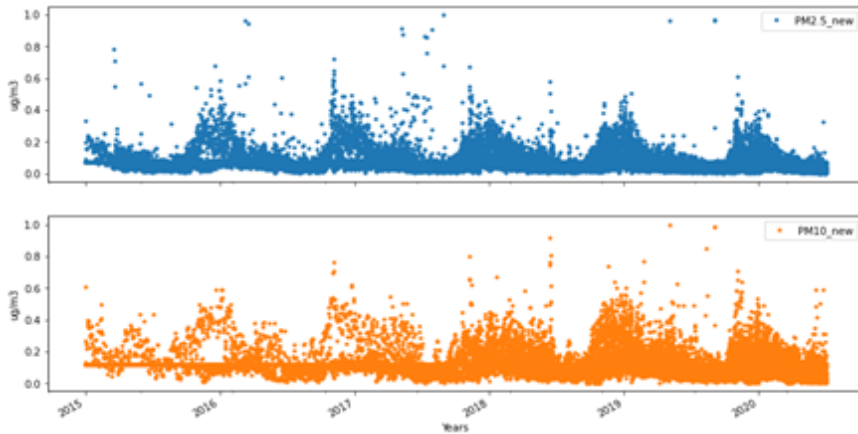

Figure 2. *Analysis of the pollutants in various years*

Proposed Classification

Various cities can be classified on the basis of terrain and on the basis of population. This classification is done so that the cities with high AQI and most polluted may be better visualized before and after the lockdown. The classification in Figure 3 is based on type of terrain is sub divided as coastal city, Mountain or hilly city and plains, with the cities as Mumbai, Shillong and Lucknow respectively. On the other hand, for the classification based on population we have highly populated (Delhi) and least populated (Visakhapatnam).
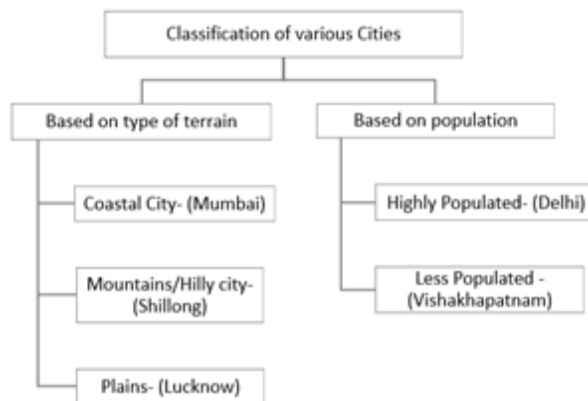

Figure 3. *Proposed classification of cities*

Effect of Lockdown on AQI

Lockdown has significantly affected AQI levels in India's leading cities. In this regard we consider the data from 2015-2020. (10) Air quality indicators (AQI) are often used by government agencies to communicate with the public how much air is currently polluted or how much it is expected to be. There are six categories of AQI level indication, namely Good, Satisfactory, Moderate, Poor, Very Poor, and Severe. The proposed AQI will consider various contaminants such as (PM10, PM2.5) for a period of 24 hours on average as determined by the National Ambient Air Quality Standards. It is interesting to note that the level of pollution in India generally decreases as summer approaches. This can also be proved by the Figure 4. Below which shows the AQI of last 5 years in India. However, the reduction of the 2020 march is very clear compared to March 2019. This could be because of lockdown. Most of the air pollution by 2020 is based on the use of domestic fuel, thermal power plants and plant heat. 2020 also shows high temperatures that have led to an unusual number of forest fires and other heat-related natural disasters, which cause air pollution. (11) Pollution levels from these sources have shown consistency and an increase in some cases. In some places, pollution levels have dropped so low that people can breathe fresh air, or see the blue sky and clear water. These improvements in air quality in some cities have allowed us to save lives. This does not mean that epidemics are good for life, but having better air quality allows us to save lives from chronic airborne diseases.
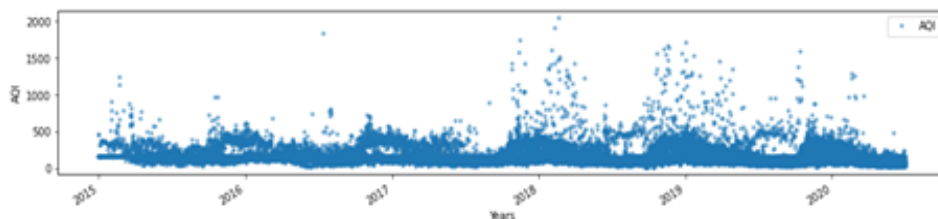

Figure 4. AQI of India in last 5 years (2015-2020)

Model Formation

A model to predict particulate matter values for Indian cities is proposed using Seasonal Autoregressive Integrated Moving Average (SARIMA). To forecast univariate time series data, ARIMA (Auto Regression Integrated Moving Average Models) is most preferred. The reason behind this is its inherent ability to handle trends in time series data, so it is applied in various domains. However, ARIMA cannot support time-series data with a seasonal component. An extension to ARIMA (parameters mentioned in Eq. (1)) that supports the direct modelling of the seasonal component of a series is called SARIMA (Seasonal Auto Regression

Integrated Moving Average Models). ARIMA (p, d, q) (P, D, Q) (1) where (p, d, q) are non-seasonal part of the model, (P, D, Q) is a seasonal part of the model.

Results and Discussions

The forecasting models presented in this paper are implemented using Python3.6. The optimal parameters for the customized SARIMAX model are (1, 0, 0) × (1, 0, 1, 12). Here, the numeric value 12 is used to indicate the usage of monthly data. The model does not include an external variable, so it is SARIMA. The model selection criterion is the Akaike's Information Criterion (AIC) by default. To forecast AQI values, the dataset is divided into training and testing data. The proposed SARIMA model considers AQI values from 2015 to 2018 (till June) as the training dataset, and July 2018 to June 2019 as the test dataset. The reason for excluding 2020 is that the year is an outlier because of COVID-19; therefore, including data from 2020 may deviate from the actual prediction.
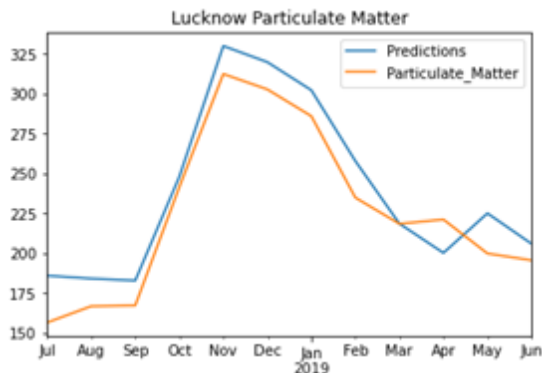

Figure 5. *Test and train prediction data of particulate matter for classification based on terrain.*
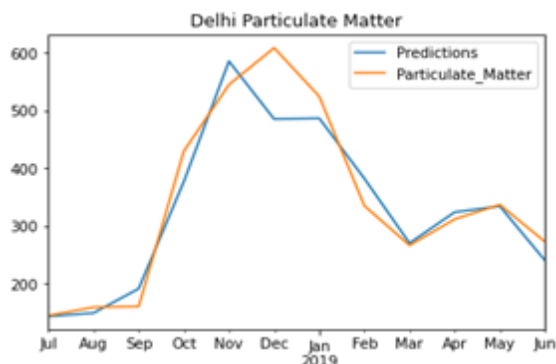

Figure 6. *Test and train prediction data of particulate matter for classification based on population.*

Figure 5 and 6 shows the performance of SARIMA in predicting the AQI values for 2020 and 2021 for classification based on terrain and population respectively. The predicted values are fairly close to our actual values obtained using SARIMA in both the figures. The error value between the existing versus forecasted results is summarized in Table 1. Figure 5 shows a comparison between the SARIMA values and the actual values in the month-wise corpus for 2019. It is quite fascinating how looking at previous values gives us so much insight into future air pollution trends. However, there is a discrepancy at the peak of the graph, where our model is unable to make predictions with high accuracy. The SARIMA model shows AQI values for 2018-2019 (Jul-Jun). Figure 6 shows the month-wise predicted results of AQI through SARIMA for 2020. The resultant data reveals a gap between the actual versus predicted AQI values.
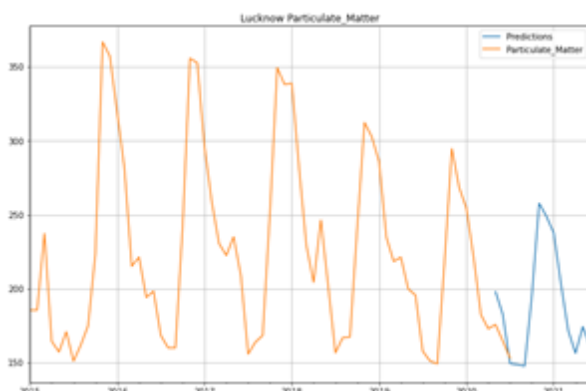

Figure 7. *Forecasted particulate matter values through SARIMA for the year 2020–2021 of Lucknow.*
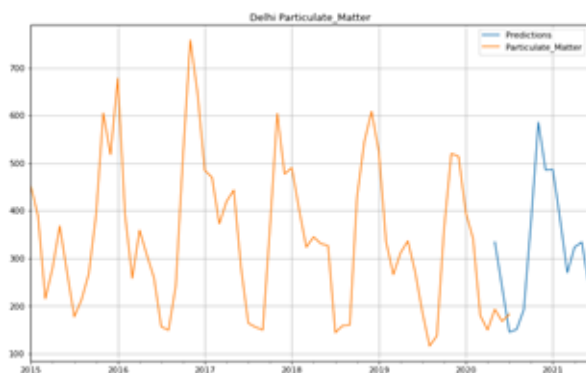

Figure 8. *Forecasted particulate matter values through SARIMA for the year 2020–2021 of Delhi.*

Mimansha AGRAWAL, Mamta AGRAWAL, Zafer ASLAN, İlknur DÖNMEZ, Ali GÜNEŞ

Table 1. Comparing the errors through SARIMA of various classifications

No.	Classification	City	Root Mean Squared Error
1	Based on Terrain	Lucknow	18.301
2	Based on Population	Delhi	46.198

The forecasting of AQI values through SARIMA is carried out and plotted in Figure 7 and 8. Predicting the levels of AQI for the upcoming year (2020–2021) is however difficult for analysts and researchers. If a predictive model considers 2020 data, there may be a chance of inaccurate prediction by the model for next year because 2020 is an outlier and also, not enough data is available for all the cities for prediction. So, to avoid the wrong predictions, the forecasting for other cities is not done. (12) However, not including the 2020 data in our dataset might lead to wrong predictions, and COVID-19 could have lasting effects that could lead to poor predictions. Consequently, the presented forecasting system for AQI values includes 2020 data, and the forecasted results are highly optimistic. Further, this scenario is pure since 2020 is an outlier. A better prediction may be possible by skipping 2020. Figure 7 shows the AQI values forecasted of Delhi using SARIMA and Figure 8 shows the AQI values of Lucknow for 2020–2021. Prediction for 2021 AQI through SARIMA produces 18.301 RMSE value for Lucknow and 46.198 for Delhi.

Conclusion

The AQI level classification models are designed using a machine learning and trained with average monthly Particulate matter (PM10 and PM2.5) values recorded in various cities in India using SARIMAX model. Many statistical tools and semi-automated tools help researchers predict air quality by considering several pollutants and seasonal parameters. However, an automated machine learning model to forecast and monitor air quality is required, especially in urban areas. COVID-19 restricted human activities in 2020, and the air quality level significantly improved. This paper presents a model to forecast AQI values in various Indian cities for the coming year by considering surge reduction in various pollutant levels. However, the main limitation to this study is because of the lack of samples in the dataset, the proposed Particulate matter model predicts the results for only 2 cities. Future work in this domain can include the hybrid approach of models to make predictions.

References

1.WHO (World Health Organization). World Health Organization Emergency Dashboard. Available at: https://covid19.who.int/region/searo/country/in. Accessed 8.07.2020.

2.Chen LWA, Chien LC, Li Y, Lin G. Nonuniform impacts of COVID-19 lockdown on air quality over the United States. Sci. Total Environ. 2020. Available at: https://doi.org/10.1016/j.scitotenv.2020.141105.

3.Coker ES, Cavalli L, Fabrizi E, Guastella G, Lippo E, Parisi ML, Pontarollo N, Rizzati M, Varacca A, Vergalli S. The Effects of Air Pollution on COVID-19 Related Mortality in Northern Italy. Environ Resour Econ (Dordr). 2020:4;1-24.

4.Coronavirus (COVID-19) related mortality rates and the effects of air pollution in England. Available at: https://www.ons.gov.uk/economy/environmentalaccounts/methodologies/

5.Ali N, Islam F. The Effects of Air Pollution on COVID-19 Infection and Mortality-A Review on Recent Evidence. Front Public Health. 2020;8:580057.

6.Wu X, Nethery RC, Sabath MB, Braun D, Dominici F. Air pollution and COVID-19 mortality in the United States: Strengths and limitations of an ecological regression analysis. Sci Adv. 2020;6(45):eabd4049.

7.Nie D, Shen F, Wang J, Ma X, Li Z, Ge P, Ou Y, Jiang Y, Chen M, Chen M, Wang T, Ge X. Changes of air quality and its associated health and economic burden in 31 provincial capital cities in China during COVID-19 pandemic. Atmos Res. 2021;249:105328.

8.MoEFC (Ministry of Environment, Forest and Climate change) NCAP National Clean Air Programme. In: Sundaray SNK, Bharadwaj DSR (eds). New Delhi, 2019.

9.Le Tertre A, Medina S, Samoli E, Forsberg B, Michelozzi P, Boumghar A, Vonk JM, Bellini A, Atkinson R, Ayres JG, Sunyer J, Schwartz J, Katsouyanni K. Short-term effects of particulate air pollution on cardiovascular diseases in eight European cities. J Epidemiol Community Health. 2002;56(10):773-9.

10.Pozzer A, Dominici F, Haines A, Witt C, Münzel T, Lelieveld J. Regional and global contributions of air pollution to risk of death from COVID-19. Cardiovascular Research. 2020;116(14):2247–2253.

11.Thorpe A, Harrison RM. Sources and properties of non-exhaust particulate matter from road traffic: a review. Sci Total Environ. 2008;400(1-3):270-82.

12.Ye Yao, Jinhua Pan, Zhixi Liu, Xia Meng, Weidong Wang, Haidong Kan, Weibing Wang, Ambient nitrogen dioxide pollution and spreadability of COVID-19 in Chinese cities. Ecotoxicology and Environmental Safety. 2021;208:111421.