# A Comparitive Study of Air Quality Index with Pollution in Bengaluru


Athulya Shanty
(*1740225*)
Department of Statistics
CHRIST (*Deemed to be University*)

Angel Mary Abraham
(*1740228*)
Department of Statistics
CHRIST (*Deemed to be University*)

Namitha Antony
(*1740250*)
Department of Statistics
CHRIST (*Deemed to be University*)



***Abstract*— The present environmental conditions in the city demands an initiative to overcome the issue of depleting nature. This research is an attempted study to analyze the effect of vehicular pollution and various weather conditions on the Air Quality Index (AQI) of Bengaluru city. The prediction of air quality by measuring AQI is noted as a major tool to identify the present living conditions in the city. The increasing rate of AQI can outturn into uncertain living circumstances and health concerns for the population in the city. The outcomes of the research depicted that it is an essential requirement to conduct periodic testing of vehicular emissions and weather changes in order to resist the plight of hazardous living conditions in the city.**



**Keywords— Air Quality Index (AQI), Environmental factors, Health concerns, Multiple Linear Regression model**


## I. INTRODUCTION

Air pollution has always been a global cause of worry. The city of Bengaluru is in the grip of rising air pollution. It is ranked second among the highest polluted cities in India after Delhi. The Garden city has almost 14% of school going children suffering due to air pollution. Air pollution has emerged as a major concern for Bangalore, with more than 75% of its areas where air quality monitoring was carried out violated the prescribed standards. The extremely populated areas of Bengaluru such as Rail Wheel factory, Yelahanka, Export promotional park ITPL, Yeshwanthpura and Central Silk Board depict severe impacts of air pollution. Due to the presence of numerous vehicles and industries such densely populated areas have experienced serious consequences of pollution. The city has the largest concentration of IT parks, electronic and bio-tech industries and is popularly known as the 'Silicon Valley of India. The rapid development and changing landscape of the city over the past few years has led to a rise in the population approximately to 8.4 million inhabitants and number of vehicles were increased by approximately 4 million. This has impacted on the infrastructure of the city with tremendous challenges and pressure with air pollution being the most crucial one. A significant increase in air pollutants and environmental deterioration has been witnessed by the city. The air pollutants have also impacted of public health causing breathing disorders, lung cancer, pulmonary cancer, COPD and several other illnesses. All such hazardous effects of air pollutants reflects the need of testing and estimating the quality of air periodically.

The key factors affecting the air quality of the city included factors such as large number of vehicles, raise in temperature, humidity, rainfall and wind along with many other climate changes.

The most commonly used technique for estimating the pollutants in the air is Air Quality Index (AQI). This process involves detection of poisonous pollutants such as $PM_{10}$, $PM_{2.5}$, $SO_2$, $NO_2$, CO, $O_3$ and $NH_3$ in the air. The government agencies use AQI method to communicate with public regarding the health concerns due to air pollutants. The risk on public health increases with the rise in AQI. There is an air quality reference index corresponding to certain air quality standards. This reference criteria can be used to predict the impact and severity of pollutants on public health. The value of AQI ranging from 0-50 represents a low AQI with no health implications, 50-100 range of AQI represents a moderate level of AQI with health concerns for the sensitive and elderly group of population. Whereas, if the AQI exceeds the range of 100 it depicts that air quality of the area under consideration is extremely hazardous for the health of general public. The paper reflects thoroughly upon the major air pollutants sources of air pollution and their causes, by considering a common technique of determining the quality of air in city of Bengaluru. It also provides the general impact of this pollutants on public health and the environment.

## II. LITERATURE REVIEW

Trodd et al [1] states that air pollution is progressively turning into a worldwide concern and is accepted to be among the main sources of life threats on the planet. The rapid development of economies and retraining the amount of contaminated air depicts that Bangalore is one of India's developing cities and, despite the fact that the development is profiting monetarily, it also has a deteriorating environment. This paper gives a basic examination of the air pollution rate in the city over the period 2005-2011 at 6 explicit areas where estimations have been reliably recorded.

Harish et al [2] has made an endeavor to consider on urban air pollution in Bangalore city caused by the discharge of pollutants from vehicles. The research endeavors the impacts of increasing vehicles in the city. In view of the realities and information acquired, the situations in regards to future vehicle growth and their impact for travel is discussed to overcome emissions problems. The outcomes presented by significant issues on vehicles and facts of existing circumstance will be utilized for the suggestions.

Nigam et al [3] differentiates air quality Index as an instrument for recognize the present situation of air quality. Six unique strategies for evaluating Air quality Index (AQI) in view of four toxins synergistic impact viz., $PM_{10}$, $PM_{2.5}$, $SO_2$ and $NO_2$ were utilized to compare air quality in the study. The daily AQI estimation depicted that air quality status in the investigation region under different references ranging from good, moderate, satisfactory and unacceptable class for computation of different AQI.

Mamta et al [4] conducts a study and reports the investigation of the surrounding air in Delhi city utilizing air quality list (AQI). The determined AQIs values of SPM and RSPM fall under all the four classifications with varying rates. The general AQI was found to fall under the classification 'poor' and 'extremely poor' inferable from the effect of pollutants like RSPM and SPM. Therefore it is seen that SPM causes primary contamination of air at the chosen study destinations in Delhi.

Kanchan, et al [5] states that air quality index (AQI) or air pollution index (API) is usually used to report the level of danger that air can cause to the public. Most of the AQI or API indices can be classified as single pollutant index or multi-pollutant index with different aggregation method. This paper attempt to analyze and review all the major air quality indices developed worldwide.

Zhai et al [6] surveys a FS-GA-BPNN model predicting the everyday normal Air Quality Index (AQI) is proposed. The models BPNN, GA-BPNN and FS-GA-BPNN were set up to meet the expectations exactness, speculation, capacity and dependability. The outcome demonstrates that the model FS-GA-BPNN performs better than normal BPNN, recommending the need to explore broad information and highlight extraction for effective AQI.

Lin et al [7] surveyed the Chinese government and advanced a progression of forceful control measures to handle ecological issues. The outcomes depicts a complex relationship between air quality and visibility. The results demonstrate that the improvement in visibility in both Beijing and Guangzhou was mostly due to the decreased $PM_{10}$ fixation. In Guangzhou, improved atmospheric visibility. The lower wind speed, together with potential changes could clarify the reason of no improvement in visibility pattern found in Shanghai or Chengdu.

Ghosh et al [8] identified New Delhi as the most polluted city. It is undoubtedly a disturbing issue for the wellbeing of our future ages in India. We are aware about the dangerous impacts of pollutants on the wellbeing of humans and environment. Some sudden measures and severe laws ought to be made to resist the air contamination in urban areas of India. Fatal illnesses like cancer and asthma is increasing amongst the Indian population. Pollution is in fact in fact responsible for increasing diseases.

Greenstone et al [9] states that more than 660 million Indians breathe air that fails India's National Air Quality Standards. This Research proposes that fulfilling the guidelines that would expand future in India by one year. Despite these huge advantages, the effectively executing programs convey that clean air is difficult to achieve. Three exercises were conducted for planning viable change like ensuring that regulatory data is reliable and unbiased, efficient framing regulations, culture of piloting and evaluating new policy.

Rizwan, et al [10] states that Air pollution is in charge of numerous medical issues in the urban territories. This paper gives a proof based understanding into the status of air pollution in Delhi and its consequences upon human wellbeing and control measures that should be initiated. The Vehicular emissions and modern exercises were observed to be related with indoor just as open air pollution in Delhi. Delhi has found a way to decrease the level of air pollution in the city during the most recent 10 years. However, further precautions must be taken to maintain low degrees of pollution.

## III. METHODOLOGY

According to the assembled data and parameters, it is observed that the technique of Multiple Linear Regression model is best suitable for the research analysis. The chosen categories of factors affecting the air quality along with AQI depicts the characteristics of variables required for analysis using Multiple Regression. It is noted from the gathered data that the quality of air which is predicted using AQI shows a variation due to extraneous factors such as number of vehicles, temperature, humidity, wind speed and rainfall. Such a variation in AQI is due to the dependency of air quality on this external factors. By involving the concept of Multiple Linear Regression in the prediction of air quality it is observed that the monthly AQI calculated using the average composition of major air pollutants is the Dependent Variable and all other factors involving vehicles, temperature, humidity, wind, rainfall and rainy days are the Independent Variables. Thus, the chosen data reflects one Dependent Variable and six Independent Variable for the research analysis using Multiple Regression.

### A. *Materials*

The materials and framework used for the research analysis includes the monthly data collected for the time span of 30 months. The considered time period is from the month of April 2016 to September 2018. The recordings of this chosen 30 months is utilized in the research to predict the air quality of Bengaluru city. The data and statistics assembled for the analyzing the air quality of Bengaluru city is recorded as follows:

- The AQI for all the months from April 2016 to September 2018 is calculated. The AQI is predicted based upon the average composition of harmful pollutants such as PM10, PM2.5, SO2, NO2, CO, O3 and NH3 in the air.

- The number of vehicles in Bengaluru city during the months of April 2016 to September 2018. It particularly involves the total number transport and non- transport vehicles along with 2 wheelers, 3 wheelers, 4 wheelers, other transports and LMV goods.

- The Climate change during the chosen time span of 30 months is recorded. The change in climate of Bengaluru involves variations in factors such as amount of rainfall, wind speed, temperature and humidity.

- The data for temperature changes from April 2016 to September 2018 is recorded in Celsius.

- The average rainfall (in mm) and the number of rainy days during the chosen span is obtained.

- The average speed of wind from April 2016 to September 2018 is calculated in km/hr.

- The humidity changes in the climate during the chosen months is also recorded.

### B. *Multiple Linear Regression Model*

The concept of Multiple Regression describes the relationship between multiple independent variables and one dependent variable. The independent and dependent variables are also known as predictor and criterion variables respectively. A dependent variable is represented as a function of several independent variables with corresponding coefficients, along with the constant term. Multiple

Regression is possible only with the involvement of at least two or more predictor variables. Generally, the Dependent Variable is symbolized using Y and Independent Variables are represented using $X_1$, $X_2$, $X_3$, … $X_k$. The model for multiple regression is given by:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots.. + \beta_k X_k$$

Here $\beta_0$ is the intercept and $\beta_1$, $\beta_2$, $\beta_3, \ldots.$ $B_k$ are analogous to the slope in linear regression equation and are also called regression coefficients. The appropriateness of the multiple regression model is tested by the F-test in the ANOVA table. A significant F indicates a linear relationship between Y and at least one of the X's.

The constructed Multiple Regression model can be used to check predictive ability by examining the coefficient of determination ($R^2$). R-Square, also known as the Coefficient of determination is a commonly used statistic to evaluate model fit. R-square is 1 minus the ratio of residual variability. When the variability of the residual values around the regression line relative to the overall variability is small, the predictions from the regression equation are good. For example, if there is no relationship between the X and Y variables, then the ratio of the residual variability of the Y variable to the original variance is equal to 1.0. Then R-square would be 0. If X and Y are perfectly related then there is no residual variance and the ratio of variance would be 0, making R-square = 1. In most cases, the ratio and R-square will fall somewhere between these extremes, that is, between 0 and 1.0. This ratio value is immediately interpretable in the following manner. If the value of R-square is 0.4 then, the variability of the Y values around the regression line is 1- 0.4 times the original variance; in other words it is explained 40% of the original variability, and are left with 60% residual variability. Ideally, it should explain most if not all of the original variability. The R-square value is an indicator of how well the model fits the data. An R-square close to 1.0 indicates that it has been accounted for almost all of the variability with the variables specified in the model.

Multiple regression technique does not test whether data are linear. It proceeds by assuming that the relationship between the Y and each of Xi's is linear. Hence as a rule, it is prudent to always look at the scatter plots of (Y, Xi), i= 1, 2… k. If any plot suggests non linearity, one may use a suitable transformation to attain linearity. Another important assumption is non-existence of multicollinearity that is the independent variables are not related among themselves). At a very basic level, this can be tested by computing the correlation coefficient between each pair of independent variables. Other assumptions include those of homoscedasticity and normality.

Homoscedasticity describes a situation in which the error term that is, the “noise” or random disturbance in the relationship between the independent variables and the dependent variable is the same across all values of the independent variables and Multivariate Normality assumes that residuals are normally distributed.

Multiple regression analysis is used when the interest is to predict a continuous dependent variable from a number of independent variables. If dependent variable is dichotomous, then logistic regression should be used.

- The beta value is used in measuring how effectively the predictor variable influences the criterion variable, it is measured in terms of standard deviation.

- R is the measure of association between the observed value and the predicted value of the criterion variable.

- R Square or $R^2$, is the square of the measure of association which indicates the percent of overlap between the predictor variables and the criterion variable.

## IV. ANALYSIS AND RESULTS

The obtained dataset containing the AQI levels and the variations in impacting factors from April 2016 to September 2018 was aggregated and analyzed for the process of research work. Initially, the basic frame of descriptive statistics is acquired for the dataset. It indicates that the mean and standard deviation statistics for all the parameters in the research dataset. For time span of chosen 30 months, it suggests that the average AQI level is 83.6985 with a standard deviation of 15.46399. It depicts that the average air quality is acceptable. The average number of vehicles are 6941584.6667 suggesting a huge impact of vehicular exposure on the air quality. The average climatic changes including Temperature, Rainy days, Rainfall, Wind, and Humidity were attained as 24.4000 C, 15.5667 days, 56.0580mm, 13.0833km/hr and 62.7000 C respectively. This mean quantities depicts a huge impact of weather related factors on air quality. The values of standard deviations for the dependent and independent variables are highly deviated from the mean suggesting that the data points are spread over a wider range of values. The maximum observed level of AQI during the study period is 125.583 and the minimum value is 58.2. In case of the Independent Variables, the maximum number of registered vehicles in Bengaluru is 6078289 and minimum number was 7683753. The maximum recorded temperature is 29 C and the minimum is 20 C. The number of rainy days ranges from 0 to 29 for various months. The maximum amount of rainfall determined is 144.36mm, while the minimum value is 0.03mm. The highest observes speed of wind was found to be 22.8km/hr and the minimum speed is 8.4km/hr. The humidity at 80% is the maximum and 20% is minimum during the study period of 30 months. The data also verifies that AQI, Temperature, Wind and Rainfall are positively skewed whereas, rainy days, vehicles and humidity are negatively skewed. The kurtosis obtained for all the variables are less than 3, which signifies that the data is platykurtic. It means the distribution produces fewer and less extreme outliers than does the normal distribution.

TABLE I. DESCRIPTIVE STATISTICS

| | Mean | Std. Deviation | N |
|---|---|---|---|
| AQI | 83.6985 | 15.46399 | 30 |
| Vehicles | 6941584.6667 | 449519.24744 | 30 |
| Temperature | 24.4000 | 2.54070 | 30 |
| Rainy Days | 15.5667 | 8.82662 | 30 |
| Avgrainfall | 56.0580 | 46.95246 | 30 |
| Avgwind | 13.0833 | 4.14122 | 30 |
| Humidity | 62.7000 | 12.33204 | 30 |

The Correlation statistics directing the mutual relationship or interdependency between the factors affecting AQI is calculated. It reflects that AQI is negatively associated with extrinsic causes such as Rainy Days, Average Rainfall, Average Wind and Humidity. Whereas, it is positively associated with factors such as Temperature and Vehicles. The negative correlation suggests an inverse relationship between the two variables, higher values of one variable tend to be associated with lower values of the other. This leads to the interpretation that the increasing climatic factors such as rainy days, rainfall, wind speed at moderate level and humidity in the environment will decrease the amount of AQI. The lowering the values of AQI depicts a satisfactory air quality and possess no risk and harm to the surroundings or public health. On the other hand, the AQI has a positive correlation with factors such as Vehicles and Average Temperature. The statistical association suggests a linear relationship between this variables. It interprets that the raising temperature and vehicles can lead to extreme levels of AQI causing a severely unhealthy living conditions even for the general public. The AQI levels beyond the range of 100 outturns into acute illnesses and has tremendous threat on the surrounding environment.

The Multiple Linear Regression Model for the dataset is constructed and interpreted with one dependent variable that is AQI and six predictor variables viz., vehicles, temperature, humidity, wind, rainfall and rainy days. The value of $R^2$ obtained by summarizing the Multiple Regression model was attained as 0.757. The coefficient of determination, denoted by $R^2$, is the proportion of the variance in the dependent variable that is predictable from the independent variables. Therefore, the data states 75.7% variation in the levels of AQI during the time period of April 2016 to September 2018 due to vehicles, temperature, humidity, wind, rainfall and rainy days. The remaining 24.3% variation in AQI is due to other extrinsic factors. The Adjusted R-squared that adjusts the statistic based on the number of independent variables in the model is obtained as 0.694.

TABLE II. MODEL SUMMARY

| Model | R | R Square | Adjusted R Square | Std Error of the Estimate | Change Statistics | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | R Square Change | F Change | df1 | df2 | Sig. F Change |
| 1 | .870[a] | 0.757 | 0.594 | 8.55899 | 0.757 | 11.944 | 6 | 23 | 0 |

a. **Predictors:** Humidity, Vehicles, Temperature, Avgwind, Avgrainfall, Rainy Days

b. **Dependent Variable:** AQI

The Analysis of Variance (ANOVA) is calculated, that depicts the Regression and Residual Sum of Squares of Air Quality dataset. The Regression model is Significant at 1% Level of Significance with a Sum of Squares 5250.020 and F-value of 11.944. This Explained Sum of Squares or the Model Sum of Squares describes that the acquired data and variables are well-fitted into the Multiple Regression Model. The Residual sum of squares (RSS) is obtained as 1684.896. It indicates the amount of variance in a data set that is not explained by the regression model. It gives the measure of error remaining between the regression function and the dataset.

The Coefficients table is obtained for the Independent factors viz., Temperature, Vehicles, Rainy Days, Average Rainfall, Average Wind and Humidity. By increase in one vehicle the AQI level is decreased by -1.288E-5 that is almost a negligible value. As temperature increases by 1 C the AQI level is decreased by -4.301. With the increase in one Rainy day AQI is increased by 1.717. With the increase in Rainfall the AQI is reduced by -.130. The increasing wind causes the AQI to be reduced by 2.290. Also, the increasing humidity decreases the AQI by -1.208.

TABLE III. ANOVA SUMMARY

| Model | Sum of Squares | df | Mean Square | F | Sig. |
|---|---|---|---|---|---|
| Regression | 5250.02 | 6 | 875.003 | 11.944 | .000[b] |
| Residual | 1684.9 | 23 | 73.256 | | |
| Total | 6934.92 | 29 | | | |

A visual representation for the normality of the Dependent Variable AQI is procured that illustrates the Normal Distribution of AQI in terms of its Frequency and Regression Standard Residuals.

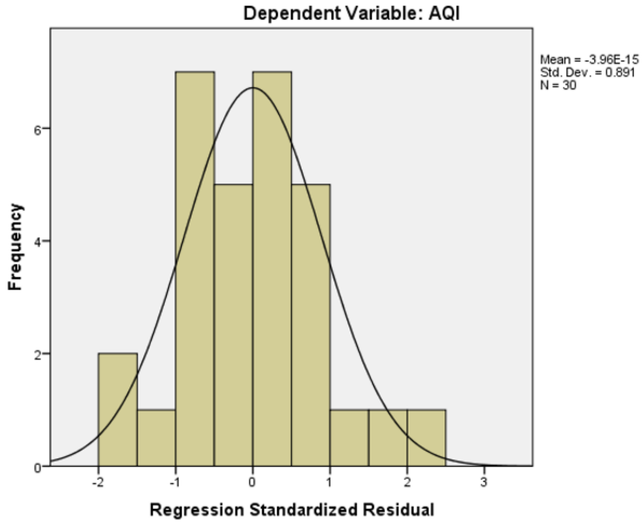


Fig. 1. Normality of AQI

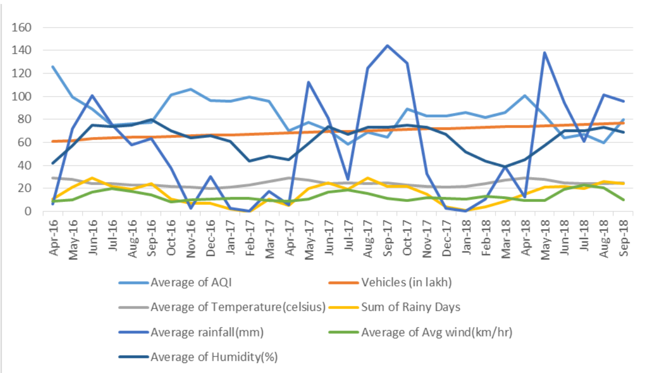


Fig. 2. Dependency Chart

## V. FINDINGS

1. Air Quality Index (AQI) of Bangalore during the time period of April 2016 to September 2018 was affected widely by factors such as number of vehicles, temperature, rainy days, rainfall, wind and humidity.

2. The comparison of obtained AQI values with the air quality reference range depicts moderate level of AQI for most of the months during study period.

3. The analytical study of the dataset verified the correlation and interdependency amongst the variables.

4. AQI is negatively associated with extrinsic factors Rainy Days, Average Rainfall, Average Wind and Humidity. This states that increasing climatic factors such as rainy days, rainfall, and wind speed at moderate level and humidity in the environment will decrease the amount of AQI.

5. The increasing Temperature and Vehicles can raise the AQI levels by causing serious harm to environment and public health.

## VI. CONCLUSIONS

Bengaluru experienced varying levels of air pollution from April 2016 to September 2018. Many of the densely populated regions the city experienced either high or critical levels of pollution as analyzed during the study. The Air Quality Index (AQI) determined by the composition several gases such as $PM_{10}$, $PM_{2.5}$, $SO_2$, $NO_2$, CO, $O_3$ and $NH_3$ depicts that the Reduction in $SO_2$ and $NO_2$ levels in the city may be an outcome of a combing of various interventions such as implementing stringent emission norms for vehicles and improvement in the quality of fuel. Also, most of the vehicles are switching to LPG for automobile fuel. Meanwhile, the levels of $PM_{10}$ and $PM_{2.5}$ remains high or critical during time period of study. Uncontrolled exposures to elevated concentrations of $PM_{10}$ and $PM_{2.5}$ can trigger health risks and can reduce the life expectancy. The critical levels of PM are likely to have a damaging effect on the health of the citizens in Bangalore that may result in a tremendous burden on the public health system. Also if the pollutants affect skilled young human resources, it would pose a threat to the city's economy that is growing exponentially. It is the need of the hour to address the public about the impacts of air pollution and adhering to stringent measures of pollution control. The AQI levels in the city must be estimated periodically and variations in the air quality levels should be communicated with public. Also, the awareness about raising number of vehicles and high temperature causing air pollution must be alerted in the city. The time period chosen for the study reflects 'Moderate' levels of AQI for most of the months. But even the Moderate level of AQI can cause serious health concerns among the sensitive group. The KSPCB has taken many initiatives to improve the quality of air, such as improving fuel quality, introducing alternate fuels for auto rickshaws, improved traffic management, implementation of emission norms, promoting use of green fuel by industries for DG sets, etc. In a nutshell, there's is a need for wider research and publications in the area of Bengaluru air pollution and its effects on public health.

## ACKNOWLEDGMENT

We give praise and thanks to the Almighty God. This paper is ostensibly supported by the Christ (Deemed to be University), Bangalore. We would also like to express our sincere gratitude towards the Head of the Department of Statistics, Prof. Joy Paulose for giving us such an opportunity. We would also like to thank Dr Nagaraja M S for his constant support throughout this study. Our team members have made valuable comment suggestions on this project which gave us an inspiration to improve our project. We would like to thank all those who have been instrumental in our work. page.

## REFERENCES

[1] N. Trodd and G. N. Santos, "Air Pollution in Bangalore , India : A Six-Year Trend and Health Implication Analysis," no. August 2015, 2014.

[2] M. Harish, "a Study on Air Pollution By Automobiles," vol. 4, no. 3, pp. 25–36, 2012.

[3] S. Nigam, B. P. S. Rao, N. Kumar, and V. A. Mhaisalkar, "Air Quality Index-A Comparative Study for Assessing the Status of Air Quality," *Res. J. Eng. Technol.*, vol. 6, no. 2, p. 267, 2015.

[4] P. Mamta and J. Bassin, "Analysis of ambient air quality using air quality index-A case study," *Int. J. Adv. Eng. Technol.*, vol. 1, no. 2, pp. 106–14, 2010.

[5] Kanchan, A. K. Gorai, and P. Goyal, "A review on air quality indexing system," *Asian J. Atmos. Environ.*, vol. 9, no. 2, pp. 101–113, 2015.

[6] B. Zhai and J. Chen, "Research on the forecasting of air quality index (AQI) based on FS-GABPNN: A case study of Beijing, China," *Proc. Int. ISCRAM Conf.*, vol. 2017-May, no. May, pp. 307–321, 2017.

[7] M. Lin *et al.*, "Regression analyses between recent air quality and visibility changes in megacities at four haze regions in china," *Aerosol Air Qual. Res.*, vol. 12, no. 6, pp. 1049–1061, 2012.

[8] D. Ghosh and P. parida, "AIR POLLUTION AND INDIA : CURRENT SCENARIO," no. May 2014, pp. 5–7, 2015.

[9] M. Greenstone, S. Harish, R. Pande, and A. Sudarshan, "The Solvable Challenge of Air Pollution in India," p. 40, 2017.

[10] Rizwan, B. Nongkynrih, and S. K. Gupta, "Air pollution in Delhi: Its Magnitude and Effects on Health," *Med. Teach.*, vol. 5, no. 4, pp. 137–143, 1983.