Statistical Prediction of AQI of Amravati City and Parametric Analysis using PCA and Z - Score Variability

Nidhi Tiwari

Associate Professor, Department of Chemistry, Shri Shankaracharya Professional University, Bhilai
Corresponding Author Email: *nidhitiwari0308[at]gmail.com*

Abstract: ***In India, the rate of urban air pollution has increased to a concerning level. The majority of cities struggle with inadequate air quality, which falls short of healthy air standards. Governmental organizations utilize the air quality index (AQI), which is a numerical representation, to inform the public about the current and projected levels of air pollution. A significant portion of the populace is anticipated to suffer from progressively severe unfavourable health impacts as the AQI rises. In this paper we have used Statistical analysis model, ARIMA (Autoregressive Integrated Moving Average) and PCA (Principal component analysis) and Z - Score variability to understand the dataset and for sequential prediction of AQI data of Amravati region of Maharashtra.***

Keywords: Statistical prediction, ARIMA, PCA, Air quality Index, Z - Score, Air Pollution, AQI

1. Introduction

Air is the only thing that keeps humans alive. For our own well - being, its quality needs to be seen and comprehended [1]. Millions of individuals worldwide have respiratory deaths and physiological problems as a result of air pollution. The biggest environmental risk, as determined by science, is air pollution. The rate of population growth has been significantly accelerated by the releases of hazardous gases brought on by fast industrialization [2]. Hazardous compounds are contaminating the air, which is seriously harming our health. The unregulated pollution has caused a dramatic decrease in air quality. Air pollution levels are measured and communicated using the Air Quality Index (AQI).

The AQI is determined by 12 parameters, or air pollutants, which include nitrogen dioxide (NO_2), sulfur dioxide (SO_2), carbon monoxide (CO), ozone (O_3), PM10 (particulate matter at a diameter of 10 microns or less), PM2.5 (particulate matter at a diameter of 2.5 microns or less), NH_3 (ammonia), and benzene. The air quality index (AQI) is computed in different applications using the six pollutants PM_{10}, $PM_{2.5}$, SO_2, NO_2, CO, and O_3. Serious health risks can result from extremely dirty air, as indicated by a high AQI rating. The AQI allows for the real - time monitoring of air quality.

2. Air Quality Index: Need and Standards

A new set of Indian National Air Quality Standards (INAQS) for 12 parameters—carbon monoxide (CO), nitrogen dioxide (NO2), sulfur dioxide (SO2), particulate matter (PM) of less than 2.5 microns size (PM2.5), PM of less than 10 microns size (PM10), Ozone (O3), Lead (Pb), Ammonia (NH3), Benzo (a) Pyrene (BaP), Benzene (C6H6), Arsenic (As), and Nickel (Ni)]—have been notified by CPCB (http: //www.cpcb. nic. in). With the exception of CO and O3, the first eight parameters (Table01) have both yearly and short - term (1/8/24 hours) standards; the other four parameters only have annual standards.

Table 1: Indian National Air Quality Standards (units: μg/m3 unless mentioned otherwise) [3]

Pollutant	SO_2	NO_2	PM2.5	PM10	O_3		CO (mg/m^3)		Pb	NH_3
Averaging Time (hr.)	24	24	24	24	1	8	1	8	24	24
Standard	80	80	60	100	180	100	4	2	1	400

BaP, C6H6, As, and Ni have annual standards

Using the AQI as a benchmark, which ranges from 0 to 500. Air pollution levels and associated health risks increase with increasing AQI values. An AQI score of 50, for instance, indicates good air quality with minimal potential to harm the public's health, but an AQI value of more than 300 indicates hazardous air quality. The national air quality standard for the pollutant, the threshold that the EPA [**4**] as established to safeguard human health, generally correlates to an AQI value of 100. Thus, it is commonly accepted that AQI readings of less than 100 are adequate. When AQI values rise beyond 100, air quality is considered hazardous, first for some susceptible populations and then for everyone. The purpose of the AQI is to educate us about the health implications of local air quality. The EPA [4]split the AQI scale into the following six categories to make it easier to understand as shown below:

1) **Good: -** A community's air quality is deemed acceptable and air pollution is either not a concern or not very dangerous if its AQI value is between 0 and 50 [4].
2) **Moderate: -** Few pollutants may affect very less population with moderate health issues, AQI for such a community is in between 51 and 100, air quality is acceptable. For example, Respiratory problems might arise in individuals who are exceptionally sensitive to ozone [4].
3) **Unhealthy for Sensitive Groups**: - Some individuals are more vulnerable than others to the negative effects of specific air pollutants. This implies that their impact will

probably be smaller than that of the broader population. People with respiratory conditions [10] and children and adults who engage in outdoor activities, for instance, are more vulnerable to ozone exposure, whereas those who have heart problems are more susceptible to carbon monoxide exposure. Certain individuals could be susceptible to many pollutants. There may be health implications for members of sensitive groups when AQI readings are between 101 and 150. When the AQI is in this range, the general public is probably not going to be impacted [4].

4) **Unhealthy**: - If the AQI values are in between 151 and 200 indicates, each and every individual may experience very serious health effects. Members of sensitive groups may experience more serious health problems [4].
5) **Very Unhealthy**: - If the AQI values are in between 201 and 300, it triggers a health alert, meaning everyone may experience more serious health effects [4].
6) **Hazardous**: - If the AQI values over 300, it elicits health warnings of emergency conditions. The entire population is more likely to be affected. Each AQI category has been assigned a colour by Environmental Protection Agency (EPA) so that a general citizen can easily and quickly comprehend the importance of the air pollution levels in his community. For instance the colour orange indicates that environment is "unhealthy for sensitive groups"; the colour red indicates that environment may be "unhealthy for all", and so on [4].

General AQI Table 02 [5] used in India shows Indian air quality index criteria and associated health communication: -

Table 2: AQI category and range

AQI category	AQI Range	Health communication
Good	0 - 50	Air quality is satisfactory, with little or no health risk.
Satisfactorily	51 - 100	Air quality is acceptable; however, there may be some health concerns for sensitive individuals
Moderate	101 - 200	Air quality may pose health risks, especially for sensitive groups and those with respiratory issues
Poor	201 - 300	Health warnings may be issued, particularly for vulnerable groups such as children, elderly, and individuals with existing health conditions
Very poor	301 - 400	Health warnings of emergency conditions. The entire population may experience health effects.
Severe	401 - 500	Health alert: everyone may experience more serious health effects. Emergency actions may be necessary

By providing information, the AQI informs the public, policymakers, and healthcare professionals about outdoor activities, health precautions and environmental regulations. This helps individuals to make informed decisions. It helps individuals understand the potential risks associated [6] with breathing polluted air and take appropriate actions to protect their health, such as reducing outdoor activities during high pollution days or using air filtration systems indoors.

Overall, the AQI serves as a vital tool in raising awareness about air pollution and its impacts on public health and the environment [7], as well as guiding efforts to improve air quality through regulatory measures and pollution control strategies.

3. Dataset Source

The data set consists of 09 various parameters $PM_{2.5}$, PM_{10}, NO, NO_2, NO_x, NH_3O_3, Benzene, Toluene and calculated AQI values of 265 days of City – Amravati, Maharashtra, India. This data set was download from the website www.keggle. com.

4. Results And Discussion

4.1 Prediction of AQI using ARIMA

Prediction of Air Quality Index was done on the basis of the historical information downloaded using Auto Regressive Integrated Moving Average (ARIMA) model. This model is purely statistical model and provides the predicted values. The calculation was done using MS Excel v.2013. A comparative graph between the predicted and true / calculated AQI is illustrated in **Figure 01**. It can be observed from the said figure that true AQI is significantly overlapped with the predicted AQI. In order to find the quality of prediction few parameter like R^2, Root Mean Square Error (RMSE), Mean Absolute Error (MAE), RMSE Standard Ratio (RSR) and Cosine Symmetry (CS) were calculated and found as 0.9452152, 15.0690007, 10.706857, 0.2340616, and 0.9832339 respectively. The values of R2 and CS are much closed to 01 which is an indicative of good quality prediction.

Paper ID: SR24501103453 DOI: https://dx.doi.org/10.21275/SR24501103453

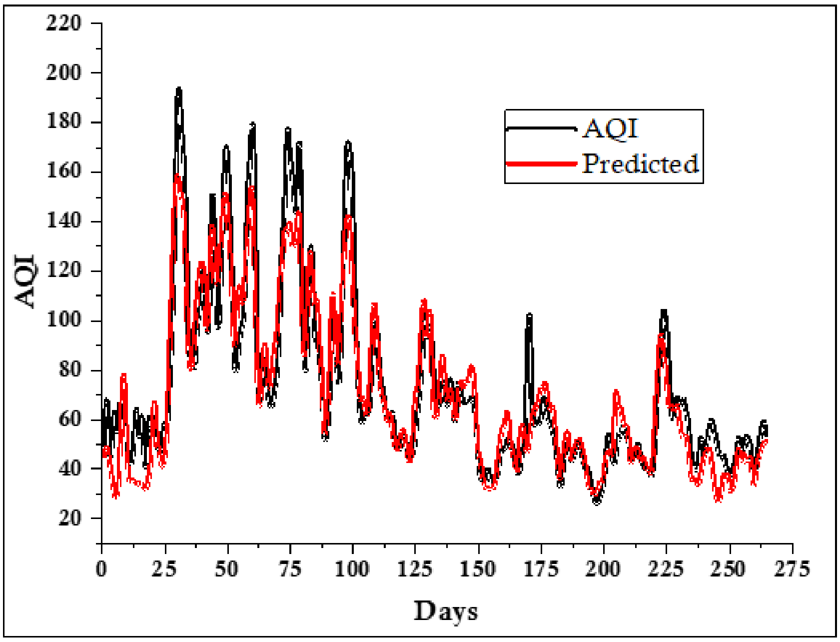

Figure 1: Comparison between True and Predicted AQI

Many researchers have found a relationship between air pollution exposure and adverse health impacts on humans. Data mining techniques are one of the approaches to forecast AQI and analyze it [1]. Data analysis by ARIMA [8] provides an accuracy and an insightful approach towards understanding AQI and its mitigation if required.

The Figure (b) is the scree plot obtained by PCA analysis. This is the graph between the component and the corresponding Eigen values. In this case, 10 components were considered therefore 10 various points can be seen. From this graph, it may be concluded that the Eigen value corresponding to component 1 to component 2 changes rapidly similarly, component 2 to component 3, hence component 1 and component 2 were chosen to draw the loading plot, illustrated in **Figure 02**.

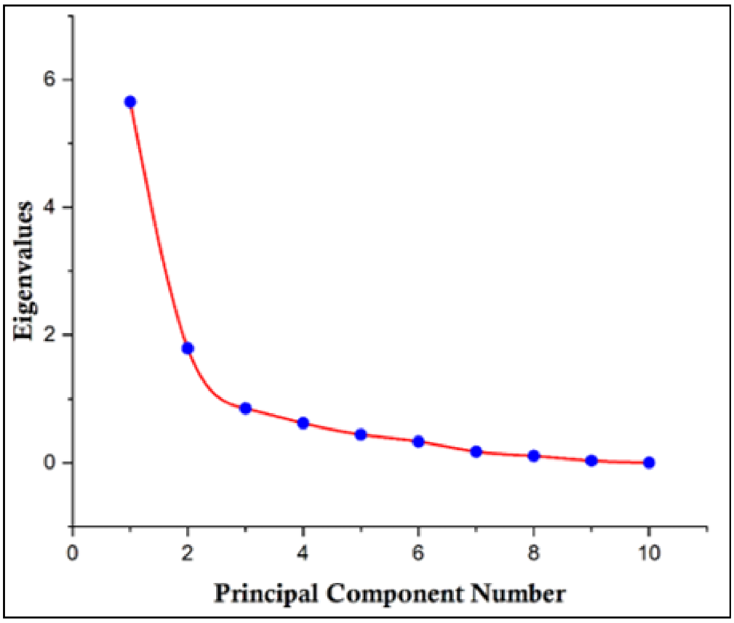

Figure 2: Scree Plot

ISSN: 2319-7064
SJIF (2022): 7.942

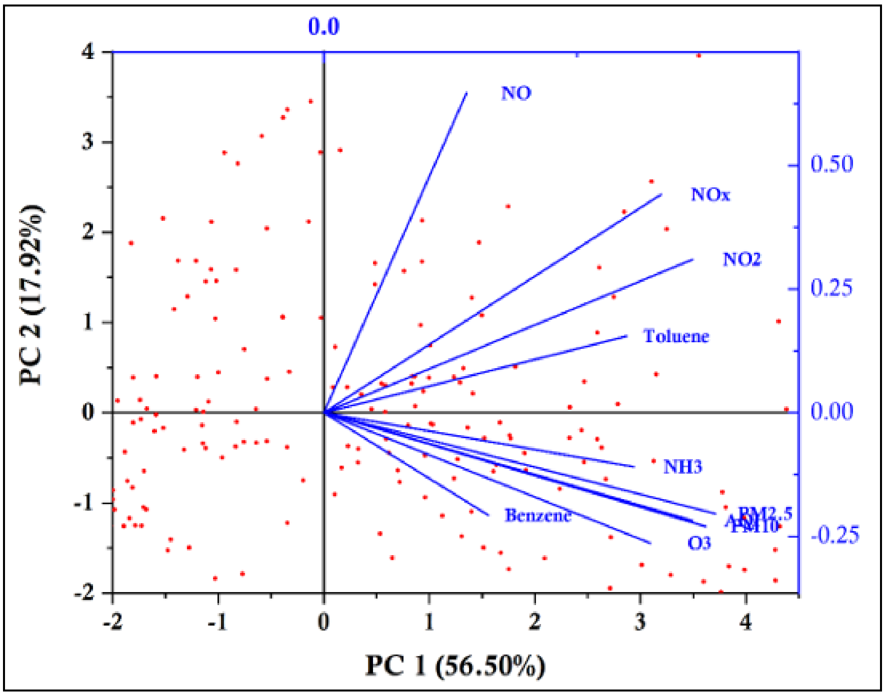

Figure 3: Loading Plot

The PCA loading plot of the AQI parameters is shown in Figure 03. The correlation between any two parameters can be determined using the said plot. All the parameters including AQI values are expressed using the blue line. Acute angle between the two parameters indicates that there is a positive correlation between the parameter considered, Right angle is the indicative of no correlation and obtuse angle expresses the existence of negative correlation. If the angle between the two blue lines is 0^o the correlation is +1 whereas when the angle is 180^o the correlation is - 1. In order to understand the dependency of AQI on other parameters, the angle created by the corresponding lines with the line assigned to AQI is observed. It may be seen that the angle between the lines assigned to AQI and PM_{10} is almost 0^o which indicates the highest dependency of PM_{10} on AQI value. There are few more parameter which highly affect the value of AQI are $PM_{2.5}$, O_3, NH_3 and Benzene. Therefore, the ways to decrease the counts of these parameters need to be researched in order to reduce the dangerous looking AQI. Table 02 provides the correlation between the various parameters.

Table 3: Correlation between the parameters

Correlation	PM2.5	PM10	NO	NO2	NOx	NH3	O3	Benzene	Toluene	AQI
PM2.5	1.00	0.96	0.08	0.66	0.56	0.67	0.75	0.31	0.57	0.91
PM10	0.96	1.00	0.05	0.64	0.52	0.60	0.74	0.35	0.55	0.90
NO	0.08	0.05	1.00	0.57	0.76	0.09	- 0.01	0.00	0.32	0.07
NO2	0.66	0.64	0.57	1.00	0.97	0.60	0.51	0.19	0.65	0.60
NOx	0.56	0.52	0.76	0.97	1.00	0.51	0.41	0.15	0.62	0.50
NH3	0.67	0.60	0.09	0.60	0.51	1.00	0.56	0.32	0.35	0.55
O3	0.75	0.74	- 0.01	0.51	0.41	0.56	1.00	0.30	0.42	0.73
Benzene	0.31	0.35	0.00	0.19	0.15	0.32	0.30	1.00	0.25	0.31
Toluene	0.57	0.55	0.32	0.65	0.62	0.35	0.42	0.25	1.00	0.53
AQI	0.91	0.90	0.07	0.60	0.50	0.55	0.73	0.31	0.53	1.00

4.2 Z - Score Analysis

In this section Standard Normalization of all the parameters along with AQI for 14 various stations were taken for study. The variability in the z - score values of the data points can be seen as box and whisker plots in the Figure 04. The filled circle inside each box shows the median of the normalized values for that parameter. The vertical lines which in most cases lie beyond the boxes represent the maximum and minimum values. For example in case of Alkalinity the median z - score value is 0 whereas, the minimum and maximum z - score values are app. - 0.65 and +1.0 respectively. The right and left edges of the box indicate the (25 percentile) and (75 percentile), respectively in each of the boxes. The whiskers extend to the most extreme data points that are not considered outliers, and the outliers are plotted individually using the '*' symbol outside the boxes. Z - Scores are estimated in order to obtain a set of linearly transformed scores **[11]**.

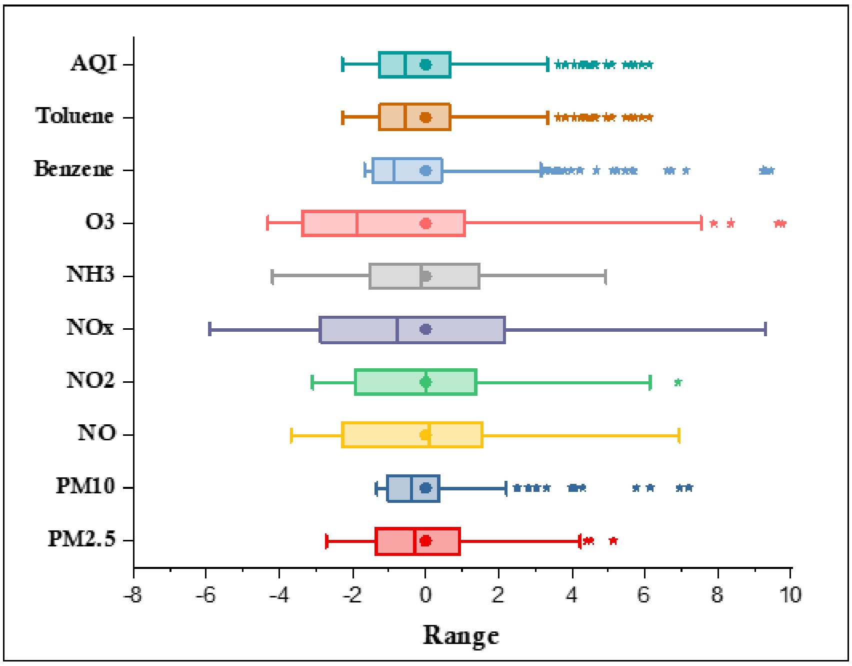

Figure 4: Z - Score variability

5. Conclusion

Arima modelling is used for future forecast of air - quality trends. ARIMA model demonstrated good predictive performance with low error rates, it could be concluded that the model provides valuable insights into future air quality fluctuations, aiding in proactive decision - making. Scree plot analysis, often used in principal component analysis (PCA), helps identify the number of significant components in the data. It could be concluded that two components PC1 and PC2 capture the majority of variability in the AQI data, simplifying interpretation in explaining air quality and potentially guiding targeted pollution control measures. Parameters affecting the value of AQI are PM_{10}, $PM_{2.5}$, O_3, NH_3 and benzene out of which the most affecting parameter is PM_{10}. The extent of air quality anomalies detected through Z - score analysis. In nut shell, the combined findings from ARIMA modeling, scree plot analysis, and Z - score analysis would provide a comprehensive understanding of air quality dynamics, enabling stakeholders to make informed decisions aimed at improving air quality and public health.

References

[1] N. S. Gupta, Y. Mohta, K. Heda, R. Armaan, B. Valarmathi, G. Arulkumaran, Prediction of air quality index using Machine learning Techniques: A comparative studies; Hindawi; Journal of Environmental and public health: 2023, 26 pages (2023).

[2] K. Aroh, Globilization, Industralization and population growth the fundamental drivers of pollution in a given Society, (India); SSRN electronic Journal, DOI: 10.2139/ssrn.3453740m, (2018), https: //doi. org/10.1155/2023/4196267.

[3] National air quality Index, central pollution control board; FINAL - REPORT_AQI_, page no 17 (2023)

[4] Air Quality Index: A guide to air quality and Health. USEPA, EPA - 454/R.00 - 005, http: //ww. epa. gov. page no.3 (2000)

[5] General AQI Table, source: Chat GPT AI.

[6] Kanchan, A. K. Gorai, P. Goyal, A Review on Air Quality Indexing System; Asian Journal of Atmospheric Environment, 9 - 2: 101 - 113 (2015).

[7] H. Raj, S. Vijaykumar, A study on air quality Index; International Journal of computer Science and Engineering, 7 (3): 961 - 966 (2019).

[8] H. N. Mahendra, S. Mallikarjunaswamy, D. Mahesh Kumar, S. Kumari, S. Kashyap, S. Fulwani, A. Chatterjee, Assessment and prediction of Air quality level using Arima Model: a case study of Surat city, Gujrat State, India; Nature Environment and pollution technology, 22 (1), 100 - 210 (2023).

[9] Kumar, P. Goyal, Forecasting of Air Quality Index in Delhi using Neural network based on Principal Component analysis; Pure and Applied Geophysics, 170 (4): 711 - 722 (2013) DOI: 10.1007/s00024 - 012 - 0583 - 4

[10] LeongW., KalaniR., Ahmad. Z., Prediction of Air pollution Index, API using support vector machines (SVM); J. Environ. Chem. Eng., 8 (3): Article 103208 (2020).

[11] M. Tripathi, S. K. Singal, Use of Principal Component Analysis for parameter selection for

[12] Development of a novel Water Quality Index: A case study of river Ganga, India, Ecological Indicators 96: 430–436 (2019).