

# Correlation analysis of atmospheric pollutants and meteorological factors using statistical tools in Pune, Maharashtra

Sneha Khedekar<sup>1,\*</sup>, Sunil Thakare<sup>2</sup>

<sup>1</sup> Research Scholar, Dr.D Y Patil Institute of technology, Pimpri, Pune, Maharashtra, India.

<sup>2</sup>Research Guide, Dr. D Y Patil Institute of Technology, Pimpri, Pune, Maharashtra, India.

**Abstract.** Air pollution has gotten worse due to the speeding up of urbanisation and industry, and the outlook for pollution control is not promising. A significant worldwide challenge that humanity is currently facing is climate change. India has suggested carbon neutrality and a carbon peak as ways to combat climate change. The intricate link and association between atmospheric contaminants and climatic variables that affect air quality, however, must be further elucidated. This work uses Pune's 2017–2021 high-resolution air pollution reanalysis open data set in conjunction with statistical techniques of the Pearson Correlation Coefficient (PCC) to compute and illustrate the design and analysis of environmental monitoring big data. The PCC is easy to use, immediately showed how contaminants and meteorological conditions relate to one another in time and space, and made environmental management agencies' jobs easier. The experimental results show that all contaminants are positively associated, with the exception of ozone, which is adversely connected. Pollutants are more influenced by meteorological factors than by temperature, which are all positively associated. Due to its strong negative relationship with the five pollutants, wind speed has a greater effect on the dispersion of pollutants.

**Keywords**-urbanisation, pollutants, air quality, Pearson Correlation Coefficient, meteorological.

## 1 Introduction

A crucial aspect of protecting the environment is reducing urban air pollution. Urban environmental protection has become more crucial to overall national environmental protection as the country's urbanisation process has accelerated. Air pollution control and prevention are critical for human survival, but they are also an essential first step in the development of the economy. Over the past few decades, air pollution and its mitigation have been ongoing scientific issues. However, they continue to be major global issues. The National Air Quality Index was introduced in 2015 by IIT Kanpur and the Indian government. Governments in populous and developing nations view the regulation of air as a crucial task.

---

\* Corresponding author: [snehakhedekar1@gmail.com](mailto:snehakhedekar1@gmail.com)

Cities with high populations and industrial activity, like Pune, are particularly susceptible to pollution increases brought on by rising aspirations and energy use. The study area considered is Pune city which takes into account 6 stations namely, Bhumkar Chowk, Hadapsar, Lohegaon, Katraj, Pashan and Shivajinagar. There is a strong association between the distribution of air pollution and surface meteorological parameters, and there is a definite relationship with seasonal fluctuations, according to research by Tan Lina, Fang Xiaoting, and others. The complex and mutable interactions between atmospheric pollutants and meteorological variables influence air quality. It is necessary to clarify the correlation and complex relationship between each other. A statistical computation method called the Pearson Correlation Coefficient (PCC) can be used to evaluate the correlation between variables. Numerous academics have recently employed correlation to analyse the relationship between air pollution indicators, making it one of the widely used techniques in the field of environmental governance.

## 2 Studies and Methodologies

Francis Chizoruo Ibe, et.al. [5] used statistical methods to examine the amounts of various air contaminants in some areas of Imo State, southern Nigeria. The existence, quantities, and potential sources of emission of PM<sub>10</sub>, NO<sub>2</sub>, SO<sub>2</sub>, CO, VOC, and H<sub>2</sub>S in specific areas of Imo State, Southeast Nigeria, were examined. This study also attempted to determine whether there might be a connection between the regional variation and dispersion of pollutants and anthropogenic and industrial activity in the study area. The air pollutant concentrations in the research area varied noticeably, as seen by the ANOVA and Box and Whisker graphs. Additionally, whereas Principal Component Analysis (PCA) further grouped the pollutants into a single coherent component indicating a comparable source of emission, Hierarchical Cluster Analysis (HCA) had already classified the pollutants into one large category. The study's use of Box and Whisker plots helped to clarify how the pollutant concentrations varied according to time of day. The observed positive connection between NO<sub>2</sub>, SO<sub>2</sub>, CO, and PM<sub>10</sub> also suggests that their respective emission sources are comparable. Planning for the environment and developing measures to reduce air pollution may benefit from this. The link and origin of the air pollutants under investigation were also shown by the PCA and HCA used in the current study. Knowing where the air pollutants come from is crucial for controlling and preventing atmospheric emissions, especially those from point sources. The wind rose diagrams were used to identify the predominant wind direction and speed that were in charge of the dispersion of air contaminants in the study area. These results highlight the importance of ongoing air quality monitoring and atmospheric analysis in cities and towns. Therefore, air quality monitoring is a proactive and crucial activity that is essential for managing, controlling, and monitoring the atmosphere. Therefore, it is imperative to regularly and continuously monitor the levels of criteria air pollutants as CO, NO<sub>2</sub>, SO<sub>2</sub>, and PM<sub>10</sub>.

Pornpun Watcharavitoon, et.al. [14] examined Bangkok's terrible air pollution. The Pollution Control Department of the Ministry of Natural Resources and Environment, Thailand, gathered the weather and air quality data utilised in this study from 1996 to 2009. The hourly air quality and meteorological data were measured at ten residential and seven roadway locations. According to Pearson's chi-square cross tabulation statistics, the 24-hour mean PM<sub>10</sub> concentrations at both roadside and residential locations were found to be significantly higher than the Thai National Ambient Air Quality Standards (NAAQS) and World Health Organisation (WHO) recommendations. The daily maximum O<sub>3</sub> (O<sub>3</sub>-1hr) concentration at both sites was higher than the Thai NAAQS. The CO<sub>8hr</sub> values at both sites were greater

than the 8-hour time-weighted average of CO (CO8hr), but lower than the Thai NAAQS. The daily 1-hour mean concentration of NO<sub>2</sub> and the 24-hour average SO<sub>2</sub> concentration were both greater than the WHO recommendations, although both sites met the Thai NAAQS. A stepwise multiple linear regression model was used to analyse the important variables affecting PM<sub>10</sub>, CO<sub>8hr</sub>, O<sub>3</sub>-1hr, NO<sub>2</sub>, and SO<sub>2</sub> levels at both locations. The results showed a declining correlation with meteorological parameters and a growing association with the region and seasons. O<sub>3</sub>-1hr levels, on the other hand, showed a diminished connection with the region under investigation. This study found that traffic emissions are the main element causing the geographic disparity in airborne pollutants in BKK, Thailand, even though meteorological circumstances may be the main factors driving the temporal changes. Our study demonstrates the various spatiotemporal facets of air pollution in BKK and offers practical solutions to address this problem.

David Nuñez-Alonso, et.al. [4] presents data on the pollution distribution in the province and city of Madrid from 22 monitoring stations between 2010 and 2017. Data on air pollution were interpreted and modelled using statistical methods. The data are yearly average concentrations of nitrogen oxides, ozone, and particulate matter (PM<sub>10</sub>) that were gathered in Madrid and its surrounding areas, one of the biggest cities in Europe with little research on the quality of its air. In order to show how these pollutants, relate to one another and to the local population, a map of their distribution was created. An association between several contaminants was established by the multivariate analysis using correlation analysis, PCA, CA. The findings allowed for the classification of different monitoring stations based to each of the four pollutants, exposing information about their mechanisms and average annual limits, illustrating their geographic distribution, and tracking their levels. The multivariate analysis's assessment states that NO<sub>2</sub> levels in the most parts of the Madrid province was more than the yearly limit was also corroborated by the development of contour maps using the geostatistical technique known as ordinary kriging.

Chen Chao, et.al. [3] calculated Using China's 2013–2018 high-resolution air pollution reanalysis open data set in conjunction with statistical techniques of the Pearson Correlation Coefficient (PCC), it was possible to visualise the design and analysis of environmental monitoring big data. This approach is simple to apply, clearly illustrates the relationship between climatic variables and contaminants in the temporal and spatial sequence, and makes it simple for environmental management departments to u The experimental results show that all contaminants are positively associated, with the exception of ozone, which is adversely connected. Pollutants are more affected by meteorological factors than by temperature, air pressure, and humidity, which are all positively associated.

## **3 Data Collected**

### **3.1 Sources of Data Collection**

Data collection is done from various sources as below:

- a) Pune Municipal Corporation.
- b) Maharashtra Pollution Control Board website (<https://mpcb.gov.in/node>).
- c) SAFAR-India-System of Air Quality and Weather Forecasting And Research website (<http://safar.tropmet.res.in/>).
- d) Indian Institute of Tropical Meteorology website (<https://www.tropmet.res.in/>).

### **3.2 Data Collected**

Following data is collected:

1. Air pollutant data and meteorological data from 6 stations for Year 2016, 2017, 2018, 2019, 2020 & 2021.
  - a) Bhumkar Chowk
  - b) Hadapsar
  - c) Lohegaon
  - d) Katraj
  - e) Pashan
  - f) Shivajinagar
2. Pollutants measured - PM2.5, PM10, O3, NO2, CO.
3. Meteorological parameters considered – temperature, wind, rainfall, uv-index.

### **4 Pearson's Correlation Co-efficient**

The test statistic that assesses the statistical association, or relationship, between two continuous variables is called Pearson's correlation coefficient. Because it is based on the method of covariance, it is regarded as the best method for determining the association between variables of interest. It provides details on the size of the association or correlation as well as the relationship's slant. The range of coefficient values ranges from +1 to -1, with +1 denoting a perfect positive relationship, -1 denoting a perfect negative relationship, and 0 denoting the absence of any link.

Degree of correlation:

1. Perfect: If the value is close to 1, the correlation is considered to be perfect, meaning that if one variable rises, the other tends to rise as well (if it's positive), or to fall (if it's negative).
2. High: A correlation is considered strong if the coefficient value is between 0.50 and 1.
3. Moderate: A correlation is considered to be of a medium degree if its value falls between 0.30 and 0.49.
4. Low: A correlation is considered to be of low degree when its value is less than +.29.
5. No correlation: When the value is zero, there is no association.

## 5 Correlation Analysis

### 1.1 Inter-correlation between pollutants

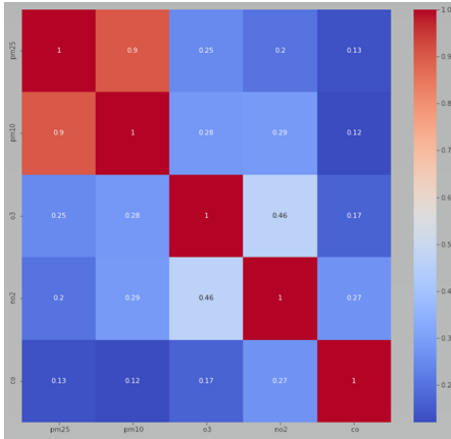


Fig. 1 Pearson Correlation Coefficient for Bhumkar Chowk

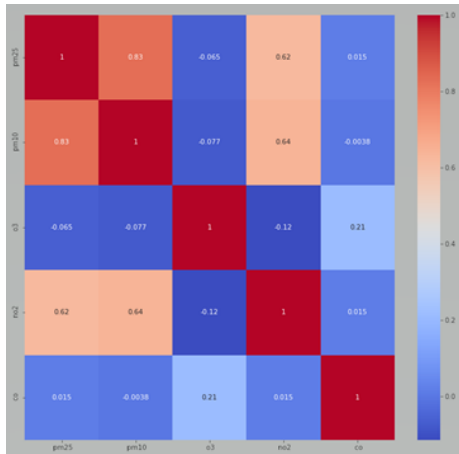


Fig. 2 Pearson Correlation Coefficient for Hadapsar

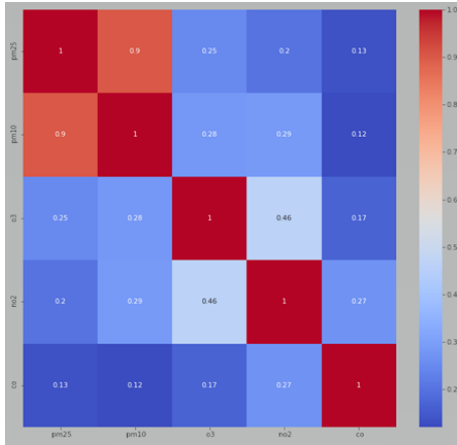


Fig. 3 Pearson Correlation Coefficient for Karvenagar

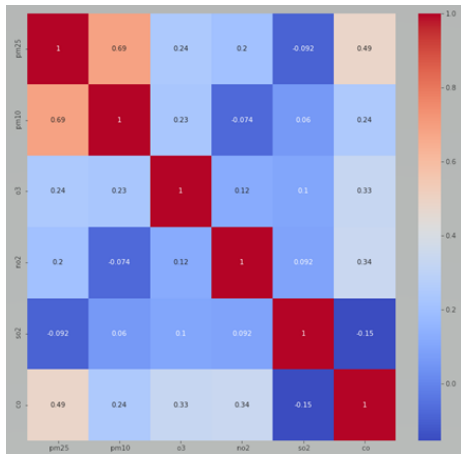


Fig. 4 Pearson Correlation Coefficient for Katraj

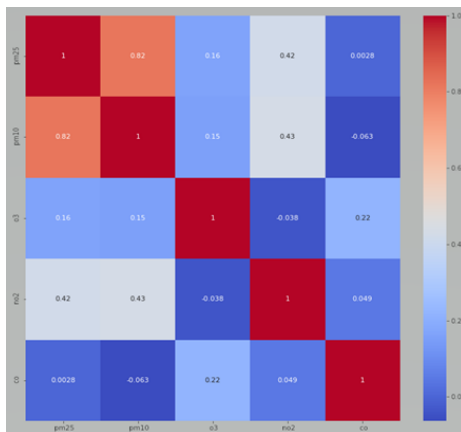


Fig. 5 Pearson Correlation Coefficient for Shivajinagar

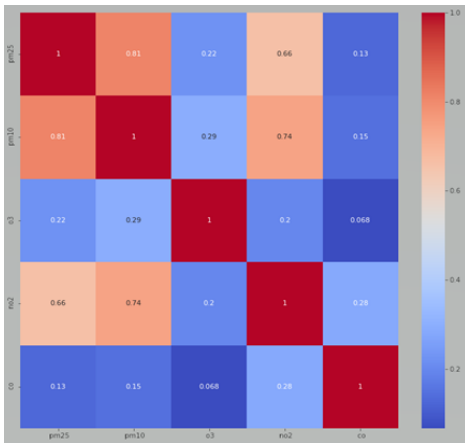


Fig. 6 Pearson Correlation Coefficient for Pashan

The above figures 1, 2, 3, 4, 5,6 shows the Pearson’s correlation matrix for Bhumkar Chowk, Hadapsar, Karvenagar, Katraj, Pashan and Shivajinagar area to show the relationship between the pollutants.

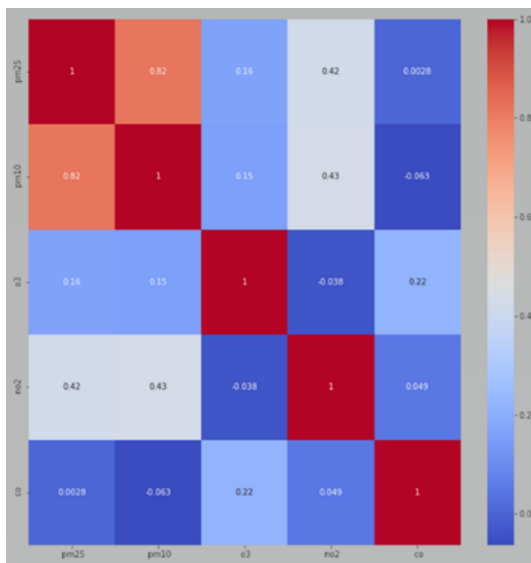


Fig. 7 Pearson Correlation Coefficient for all the stations

The above matrix (Fig. 7) shows the relationship of all the 6 stations considered together for the years 2017-2021. It can be observed from the matrix PM 2.5 and PM10 are very highly correlated with co-efficient 0.82 but PM2.5 and PM10 both are very less correlated with O3, moderately correlated with NO2 and almost no correlation with CO.

## 5.2 MONTHLY CORRELATION

To find the effect of time and weather month-wise correlation for the year 2017-2021 for all the 6 stations has been done using Pearson’s Correlation coefficient. The result is shown in the table below. In this we can observe that in the months of July to September the pollutants come down.

Table 1. Monthwise Correlation of Pollutants

Month	PM <sub>2.5</sub> & PM <sub>10</sub>	PM <sub>2.5</sub> and NO <sub>2</sub>	PM <sub>10</sub> and NO <sub>2</sub>	O <sub>3</sub> and NO <sub>2</sub>	O <sub>3</sub> and CO	NO <sub>2</sub> and CO
Jan	Very high	Low	Very less	Low	Low	Moderate
Feb	High	Low	Low	Low	Low	Moderate
March	High	Low	Low	Moderate	Low	Low
April	High	Moderate	Moderate	Low	Low	Low
May	High	Moderate	Low	Moderate	Very Low	--
June	Moderate	Low	Low	--	--	--
July	Moderate	Low	Very Low	--	Low	Low
Aug	Moderate	Low	--	--	Low	Low
Sept	High	Low	Very Less	--	Low	--
Oct	High	Low	Low	--	Low	--
Nov	High	Moderate	Low	--	Low	--
Dec	High	Low	Very Less	--	Very Less	Very Less

## 5.3 Pearsons Correlation for Pollutants & Meteorological Parameters

To show the relationship between the various meteorological parameters like temperature, rainfall, windspeed, uv-index and the pollutants considered, the following graph of Pearson’s correlation coefficient is given. The figure is only for all 6 region considering all the values for pollutants as well as for meteorological parameters considering years 2017-2021.



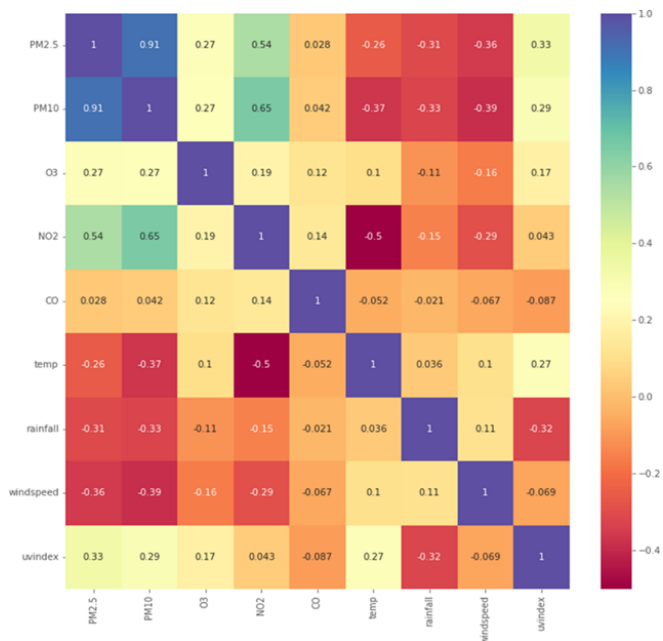


Fig. 8 Pearsons correlation for pollutants & meteorological parameters

## 6 Conclusion

Using the 6-year plenty of-level data of PUNE city according to the monthly granularity, the mean per day data between the six pollutants (PM2.5, PM10, NO2, CO, O3) is calculated. The meteorological factors (TEMP, RAIN, UV, WIND) and the Pearson correlation coefficient of pollutants are examined statistically. The experiment's findings show a positive correlation between the pollutants, with the exception of ozone, which has a negative correlation; a greater influence of meteorological variables on the pollutants; and a relationship between the pollutants and temperature. The six contaminants have a strong negative association with wind speed, which has a bigger effect on the dispersion of pollutants.

## References

1. C. Fahe., Environment and Sustainable Development, **vol. 45**, no. 3, pp. 5-15, (2020), doi: 10.19758/j.cnki.issn1673-288x.202003005
2. C. Linjun, W. Shuai, G. Zhengyu, Chinese Environmental Science, **vol. 37**, no.11, pp. 4003-4012, 2017, doi: 10.3969/j.issn.1000-6923.2017.11.001.(2019)
3. C. Chao, B. Won Min, International Journal of Contents **Vol.18**, No.1, Mar.(2022).
4. D. Nuñez-Alonso, L.V. Pe´rez-Arribas, S. Manzoor ,and J. O. Ca´ceres, Journal of Analytical Methods in Chemistry, (2019), Article ID 9753927, 9 pages, <https://doi.org/10.1155/2019/9753927>.
5. F. C. Ibe, A. I. Opar, C. E. Durua, I. B. Obinna, M. C. Enegho, Scientific African, (2019).

6. F. Xiaoting, D. Huabo, H. Mingwei, *Environmental Pollution and Control*, **vol. 41**, no. 5, pp. 541-546, (2019), doi: 10.15985/j.cnki.1001-3865.2019.05.009.
7. J. Qiqing, C. Wencong, X. Bingye, *China Environmental Monitoring*, **vol. 36**, no. 05, pp. 88-95, (2020), doi: 10.19316/j.issn.1002-6002.2020.05.12.
8. K. Lei, T. Xiao, *Earth Syst. Sci. Data*, **vol. 13**, pp. 529-570, (2021) doi: <https://doi.org/10.5194/essd-13-529-2021>.
9. L. Hui, X. Dunsheng, C. Hong, , " *Environmental Science Research*, **vol. 32**, no. 06, pp. 993-1000, 2019, doi: 10.13198/j.issn.1001-6929.2019.01.08.( 2017).
10. L. Jian and L. Yinkun. *Journal of North China Institute of Science and Technology*, **vol. 16**, no. 04, pp. 93-97, doi: 10.3969/j.issn.1672-7169.2019.04.016.(2019)
11. L. Qing, L. Dian, W. Liwei, *Environmental Science and Management*, **vol. 46**, no.04, pp. 136-140,doi: 10.3969/j.issn.1673-1212.2021.04.034, (2021).
12. L. Zihao, H. Jianwu, K. Deya, *Environmental Science and Management*, **vol. 44**, no. 02, pp. 43-48, doi: 10.3969/j.issn.1673-1212.2019.02.010, (2019).
13. L. Zuoyun and L. Henglin, *Hunan Journal of Ecological Sciences*, **vol. 8**, no. 03, pp. 69-75, doi: 10.3969/j.issn.2095-7300.2021.03.011. (2021)
14. P. Watcharaviton, C.-Pin Chio, C. Chan, *Aerosol and Air Quality Research*, pp. 1741-1754, doi: 10.4209/aaqr.2012.11.0305 (2013).
15. T. Jiayang, F. Chuanbo, Y. Renyong,, *Environmental Pollution and Control*, **vol. 40**, no. 4, pp. 445-449, doi: 10.15985/j.cnki.1001-3865.2018.04.016, (2018).
16. T. Lina, *Environmental Science and Management*, **vol. 46**, no. 3, pp. 64-68, doi: 10.3969/j.issn.1673-1212.2021.03.016, (2021).
17. W. Geng, L. Yijiang, W. Lijing, *Journal of Liaoning Normal University (Natural Science Edition)*, **vol. 41**, no.04, pp. 516-522, doi: 10.11679/lxsblk2018040516. (2018)
18. X. Kai, R. Xuechang, and C. Renhua. *Environmental Engineering*, **vol. 10**, no. 06, pp. 101-112, doi: 10.13205/j.hjgc.202109014.(2021)
19. Y. Liu, W. Shigong, and Z. Ying, *Journal of Lanzhou University (Natural Science Edition)*, **vol. 54**, no. 6, pp. 731-738, 2018, doi: 10.13885/j.issn.0455-2059.2018.06.003.
20. Z. Jieqiong, W. Yaqian, G. Shuang, *China Environmental Science*, vol. 38, no. 10, pp. 64-74, doi: 10.3969/j.issn.1000-6923.2018.10.008. (2018)