

A Mathematical And Statistical Approach To Environmental Pollution Analysis A Case Study of Air Pollution Indicators in the Peenya Industrial Area

Kunal B. Rathod

Assistant Professor, Department of Mathematics, Gopikabai Sitaram Gawande Mahavidyalaya, Umardhed
kunalbrathod9@gmail.com

Abstract

Air pollution is often discussed through large numbers and technical reports, but the real value of those numbers appears only when they are examined carefully. Mathematics provides a simple way to turn environmental observations into measurable patterns. This paper presents a mathematical and statistical analysis of annual average air-pollution indicators recorded at the Peenya Industrial Area. The study considers sulphur dioxide (SO_2), nitrogen dioxide (NO_2) and particulate matter (PM_{10}) for 2015–16, 2016–17 and 2017–18. The data reported by the Central Pollution Control Board (CPCB) are analysed using arithmetic mean, sample standard deviation, coefficient of variation, percentage change and a simple linear trend model. The mean concentrations during the study period were $2.63 \mu g/m^3$ for SO_2 , $27.57 \mu g/m^3$ for NO_2 and $104.03 \mu g/m^3$ for PM_{10} . From the first to the last observation, SO_2 decreased by 48.72% and PM_{10} decreased by 40.08%, while NO_2 increased by 26.24%. The fitted trend slopes were $-0.95 \mu g/m^3/year$ for SO_2 , $+2.65 \mu g/m^3/year$ for NO_2 and $-25.45 \mu g/m^3/year$ for PM_{10} . The PM_{10} series also shows the strongest linear fit among the three pollutants, although the short three-year record limits any long-term conclusion. The results demonstrate that different pollutants can behave differently within the same industrial environment. The paper argues that a basic mathematical framework, when applied to reliable monitoring data, can support environmental interpretation, comparison and sustainable planning. The study is intentionally descriptive rather than causal; the calculated trends should not be interpreted as proof that a particular source caused a change.

Keywords: Air pollution; mathematical modelling; environmental statistics; trend analysis; SO_2 ; NO_2 ; PM_{10} ; sustainable development; statistical analysis.

1. Introduction

Environmental problems are usually visible in everyday life, but their scale is understood through measurements. Air quality is a good example. A person may notice dust, smoke or traffic emissions, but a monitoring station produces numerical observations that can be compared over time. Once the observations are available, mathematics becomes useful because it helps us answer simple but important questions: What is the average level? How much does it vary? Is there a rising or falling tendency? Is one pollutant changing differently from another?

Air pollution is also an interdisciplinary problem. It involves chemistry, physics, environmental science, public health, engineering and data analysis. The World Health Organization's 2021 Global Air Quality Guidelines provide health-based guideline levels for major pollutants including $PM_{2.5}$, PM_{10} , NO_2 , SO_2 , ozone and carbon monoxide. These guidelines are intended as evidence-informed references and are not legally binding national standards. This distinction is important when environmental observations are interpreted.

In India, the Central Pollution Control Board and State Pollution Control Boards maintain air-quality monitoring programmes and publish data for different locations. The CPCB material for Peenya Industrial Area provides annual average concentrations for several pollutants. For the three years selected in this study, SO_2 , NO_2 and PM_{10} have consistent observations, making them suitable for a small mathematical case study. CPCB reports that PM_{10} exceeded the applicable National Ambient Air Quality Standard in the measured years in this Peenya dataset, while SO_2 and NO_2 were within the stated standards.

The purpose of this paper is not to make a broad claim about all industrial areas. Instead, it demonstrates how a small environmental dataset can be analysed systematically. The approach is deliberately understandable: descriptive statistics are followed by percentage change and a simple trend equation. Such a framework is useful for teaching, preliminary research and environmental data interpretation, and it can later be extended to larger datasets and more advanced models.

Copyright: © by the Author(s)
This is an open access article distributed under the terms of the Creative Commons Attribution License and permitted unrestricted use, distribution, and reproduction in any medium.

2. Objectives of the Study

- To analyse selected air-pollution indicators using elementary mathematical and statistical methods.
- To calculate the mean, sample standard deviation and coefficient of variation for SO₂, NO₂ and PM₁₀.
- To measure the percentage change between the first and last observations.
- To estimate the direction and strength of a simple linear trend.
- To compare the observed PM₁₀ levels with the applicable CPCB standard reported with the dataset.
- To discuss how mathematical analysis can support environmental monitoring and sustainable planning.

3. Review of Related Work

Environmental data analysis has a long history of using statistical summaries, regression, time-series methods and mathematical models. In environmental applications, the quality of the conclusion depends strongly on the quality of the observations and on whether the selected method matches the data. A simple trend line can be useful for describing direction, but it should not be confused with a causal model.

Air-quality research increasingly combines statistical analysis with modelling and forecasting. Recent work on Indian cities has used statistical forecasting approaches such as autoregression, moving averages and ARIMA to study PM_{2.5} predictability across different urban settings. Other recent studies have combined statistical analysis with artificial intelligence to model aerosol concentrations and meteorological relationships. These studies show the value of data-driven approaches, but they also highlight the importance of long time series and appropriate explanatory variables when stronger conclusions are required.

The present work takes a more modest approach. It uses a short, verified dataset and focuses on transparent calculations that can be reproduced by another researcher. This is useful because environmental statistics should not become unnecessarily complicated when the available dataset is small. The mathematical method should be strong enough to answer the research question but simple enough for the reader to follow.

4. Data Source and Study Area

The data used in this paper are annual average air-pollutant concentrations reported for the Urban Eco Park / Peenya Industrial Area in Bengaluru. The CPCB source gives the values for 2015–16, 2016–17 and 2017–18. The three selected pollutants are SO₂, NO₂ and PM₁₀. The same CPCB material also reports PM_{2.5} for 2017–18 but not for the earlier two years in the table, so PM_{2.5} is not used for the three-year trend calculation.

Peenya is an industrial area and therefore provides a meaningful setting for studying the interaction between industrial activity, transport, road dust and measured air quality. CPCB documentation also notes that the monitoring station at Peenya was operational from 2008 and that continuous ambient air-quality data generation began from 2011 in the regional monitoring programme. citetum1search17

5. Dataset Used in the Study

Year	SO ₂ (µg/m ³)	NO ₂ (µg/m ³)	PM ₁₀ (µg/m ³)
2015–16	3.9	20.2	127.0
2016–17	2.0	37.0	109.0
2017–18	2.0	25.5	76.1

Source: Central Pollution Control Board, Peenya Industrial Area air-quality data.

6. Methodology

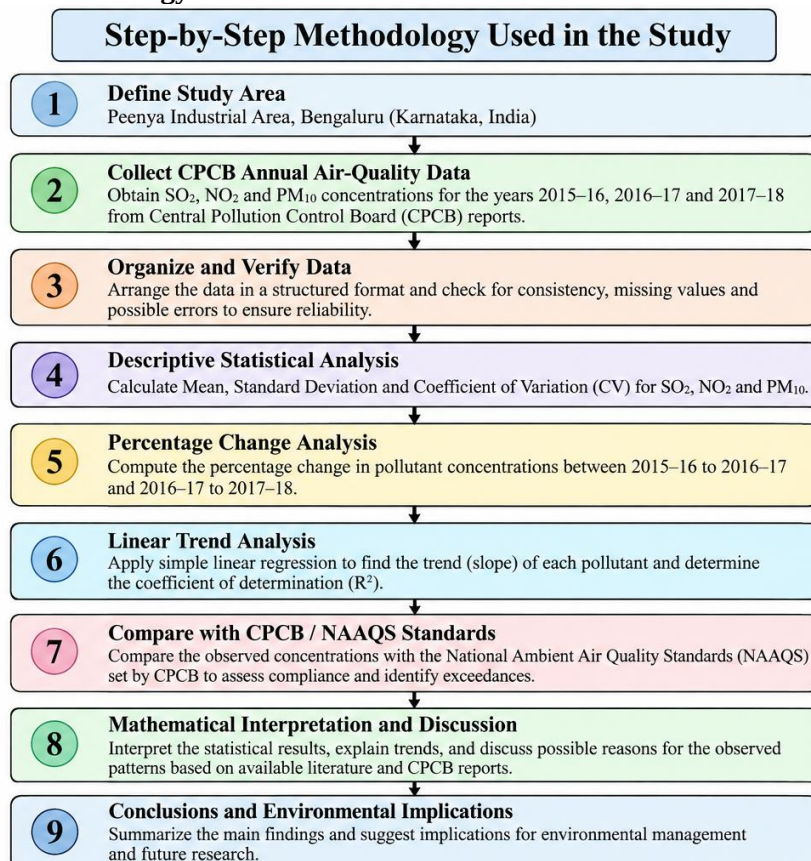


Figure 2. Step-by-step methodology used in the study.

6.1 Arithmetic Mean

For n observations, the arithmetic mean is calculated as:

$$\bar{x} = (x_1 + x_2 + \dots + x_n) / n$$

6.2 Sample Standard Deviation

The sample standard deviation is used to describe year-to-year variation:

$$s = \sqrt{[\sum (x_i - \bar{x})^2 / (n - 1)]}$$

6.3 Coefficient of Variation

$$CV = (s / \bar{x}) \times 100$$

The coefficient of variation is useful when the amount of variation needs to be considered relative to the mean. It is particularly helpful when different pollutants have very different numerical scales.

6.4 Percentage Change

$$\% \text{ Change} = [(Final \text{ value} - Initial \text{ value}) / Initial \text{ value}] \times 100$$

6.5 Linear Trend

A simple linear model is written as $y = a + bx$, where y is the pollutant concentration, x is the coded time variable, a is the intercept and b is the slope. A negative slope indicates a decreasing fitted trend, while a positive slope indicates an increasing fitted trend. Because only three annual observations are available, the trend is treated as descriptive evidence rather than a forecasting model.

7. Results and Discussion

7.1 Descriptive Statistics

Pollutant	Mean	Sample SD	CV	First-to-last change
SO ₂	2.63	1.10	41.66%	-48.72%
NO ₂	27.57	8.59	31.16%	+26.24%
PM ₁₀	104.03	25.81	24.81%	-40.08%

The mean concentration of PM₁₀ is the highest of the three pollutants in numerical terms, at 104.03 $\mu\text{g}/\text{m}^3$. Its sample standard deviation is 25.81 $\mu\text{g}/\text{m}^3$. SO₂ has the smallest mean

concentration at $2.63 \mu\text{g}/\text{m}^3$, while NO_2 has a mean of $27.57 \mu\text{g}/\text{m}^3$. These averages describe the dataset; they should not be interpreted as a direct comparison of health risk because different pollutants have different properties and health effects.

The coefficient of variation is highest for SO_2 (41.66%), followed by NO_2 (31.16%) and PM_{10} (24.81%). The higher relative variation in SO_2 is partly a consequence of its low mean value. For that reason, both absolute variation and relative variation are reported rather than relying on one statistic alone.

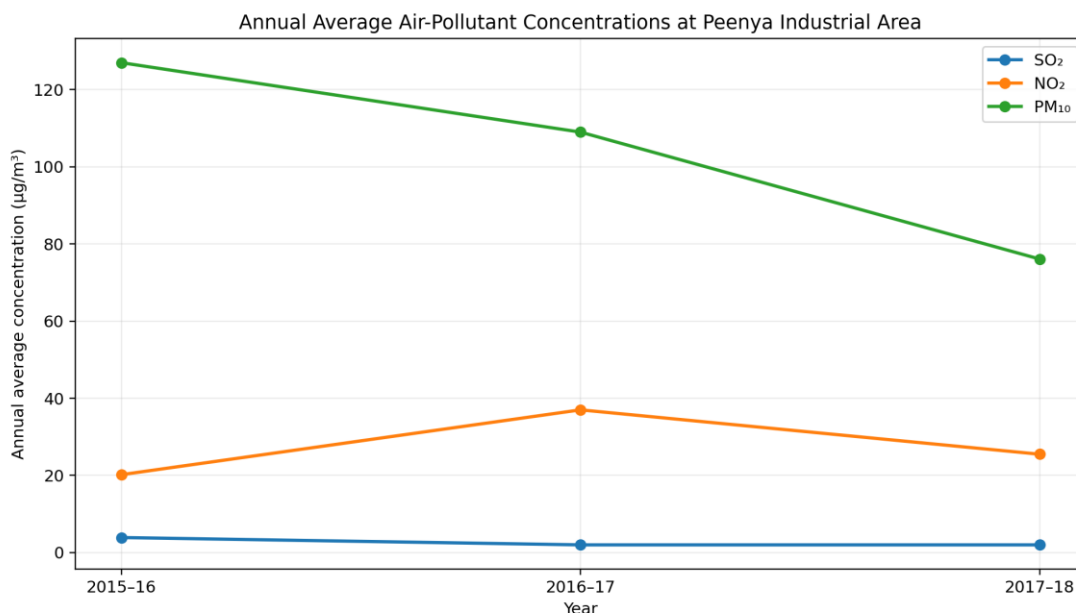


Figure 3. Observed annual average concentrations of SO_2 , NO_2 and PM_{10} .

7.2 Percentage Change

The first-to-last comparison gives a clear picture of the direction of change. SO_2 fell from $3.9 \mu\text{g}/\text{m}^3$ in 2015–16 to $2.0 \mu\text{g}/\text{m}^3$ in 2017–18, a decrease of 48.72%. PM_{10} decreased from 127.0 to $76.1 \mu\text{g}/\text{m}^3$, a decrease of 40.08%. NO_2 moved in the opposite direction: it increased from 20.2 to $25.5 \mu\text{g}/\text{m}^3$, an increase of 26.24%.

The middle year is also important. NO_2 reached $37.0 \mu\text{g}/\text{m}^3$ in 2016–17 and then fell to $25.5 \mu\text{g}/\text{m}^3$ in 2017–18. Therefore, the final value alone does not tell the full story. This is one reason why even a small time series should be viewed as a sequence rather than as two isolated numbers.

7.3 Linear Trend Analysis

Pollutant	Slope ($\mu\text{g}/\text{m}^3/\text{year}$)	R^2	Interpretation
SO_2	-0.95	0.750	Downward fitted trend
NO_2	+2.65	0.095	Upward fitted trend
PM_{10}	-25.45	0.972	Strong downward fitted trend

The estimated slopes are $-0.95 \mu\text{g}/\text{m}^3/\text{year}$ for SO_2 , $+2.65 \mu\text{g}/\text{m}^3/\text{year}$ for NO_2 and $-25.45 \mu\text{g}/\text{m}^3/\text{year}$ for PM_{10} . The corresponding R^2 values are approximately 0.750, 0.095 and 0.972. The high R^2 for PM_{10} means that a straight line describes the three observed PM_{10} points relatively closely. It does not mean that the same rate will continue in future years. The low R^2 for NO_2 reflects the rise followed by a decline, which is not well represented by a straight line.

This is an important mathematical point. A trend line is a summary of observed data, not an explanation of why the data changed. Factors such as traffic, road dust, industrial activity, meteorology and changes in monitoring conditions may influence measured concentrations. The present dataset does not contain enough explanatory variables to separate those effects.

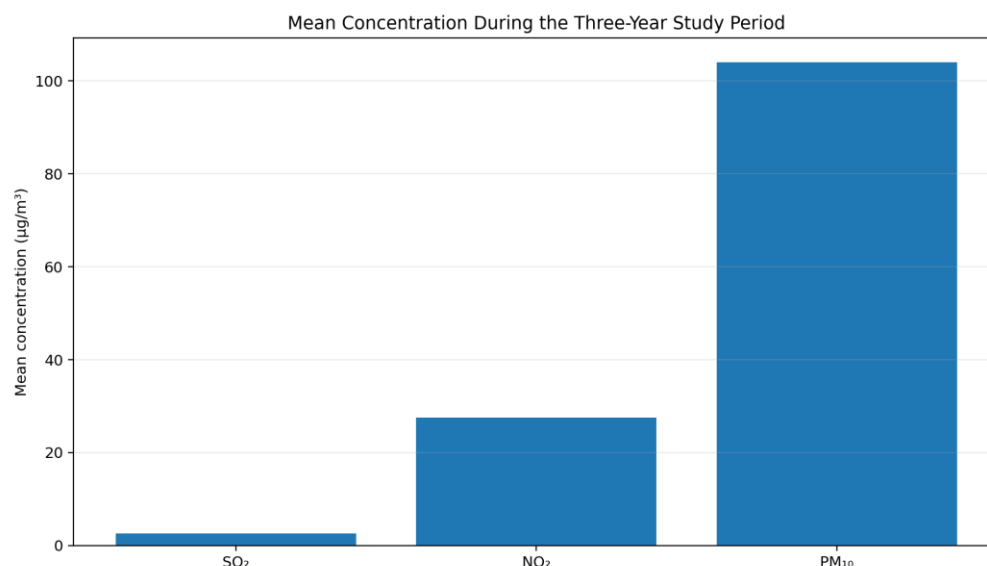


Figure 4. Mean concentration of the three pollutants over the study period.

8. Comparison with the Reported CPCB Standard

The CPCB source associated with the Peenya data reports annual standards of 50 µg/m³ for SO₂, 40 µg/m³ for NO₂ and 60 µg/m³ for PM₁₀ in the comparison table. On that basis, the PM₁₀ observations of 127.0, 109.0 and 76.1 µg/m³ are above the reported annual standard in all three years. The SO₂ and NO₂ values remain below their respective reported standards in the same dataset. CPCB attributes the PM₁₀ exceedance in this material to factors including vehicular movement and road dust. This is a source-reported interpretation; the present paper does not independently establish causality.

WHO guideline values should not be mixed with CPCB legal standards. The WHO 2021 annual guideline for PM₁₀ is 15 µg/m³ and for NO₂ is 10 µg/m³; the WHO guideline framework is health-based and is not itself a national legal standard.

9. Mathematical Interpretation

The analysis shows why environmental mathematics is more than calculation. The mean gives a typical level, the standard deviation describes variation, the coefficient of variation puts that variation into context, percentage change gives an easy-to-understand comparison, and the trend equation gives a compact description of direction. Together, these measures provide a more balanced view than a single number.

For example, saying that PM₁₀ was 104.03 µg/m³ on average is useful, but the statement becomes more informative when we add that it fell from 127.0 to 76.1 µg/m³ during the selected period. Similarly, saying that NO₂ averaged 27.57 µg/m³ does not reveal its temporary rise to 37.0 µg/m³ in 2016–17. Mathematics helps preserve these details while keeping the interpretation concise.

10. Relevance to Sustainable Development

Reliable environmental data can support better planning. If a pollutant repeatedly exceeds a standard, authorities can examine possible sources and consider targeted measures. If the pollutant decreases, the trend can be monitored to see whether the improvement is sustained. In both cases, mathematics provides a common language for comparing observations across time and locations.

The study also fits the conference theme of Mathematics for a Greener Planet and Environmental Data and Statistics. The conference is interdisciplinary and invites work that connects scientific and technological approaches with environmental protection.

A future version of this work could use monthly or daily observations and add meteorological variables such as rainfall, temperature, humidity and wind speed. Correlation, multiple regression, seasonal analysis and time-series models could then be used. A larger dataset would also make it possible to examine whether changes are persistent and whether environmental factors explain part of the observed variation.

11. Limitations

- The main analysis contains only three annual observations, so the trend estimates should not be treated as long-term forecasts.
- The study focuses on one industrial area and should not be generalized to all industrial areas or all cities.
- Only pollutants with consistent observations for all three years are used in the main trend calculations.
- The dataset does not include enough explanatory variables to establish causal relationships.
- The CPCB source's explanation of possible PM₁₀ sources is reported as an attributed interpretation rather than independently tested in this paper.

12. Future Scope

The next stage of this research should use a longer time series. Ten or more years of annual data, or preferably monthly/daily data, would provide a much stronger basis for statistical modelling. The analysis could be expanded to PM_{2.5}, PM₁₀, NO₂, SO₂, CO and O₃. A multivariate model could examine the relationship between pollution and meteorological variables. Time-series models such as ARIMA or other forecasting methods could be compared using out-of-sample validation rather than simply fitting a line to the observed data. Recent Indian research demonstrates the potential of statistical forecasting and AI-based models for air-quality prediction, but the choice of model should always follow the structure and quality of the available data.

A further extension would be a local study using verified air-quality observations from Maharashtra. Such a study would make the research more geographically relevant while retaining the same mathematical framework.

13. Conclusion

This paper presented a simple mathematical and statistical approach for analysing air-pollution observations from the Peenya Industrial Area. The calculations showed that SO₂ decreased by 48.72% and PM₁₀ decreased by 40.08% between the first and last observations, while NO₂ increased by 26.24%. The PM₁₀ series had the strongest linear fit, whereas the NO₂ series showed a much weaker linear pattern because its middle-year value was higher than both the first and final observations.

The main lesson is that environmental data should be read as a pattern rather than as a collection of isolated numbers. Mean, variation, percentage change and trend analysis are simple tools, but when they are applied carefully they can reveal useful information for environmental monitoring. At the same time, the limitations of a short dataset must be stated clearly. A responsible mathematical study should distinguish between what the data show, what a model suggests and what remains unknown.

The approach can be extended to longer datasets, additional pollutants and more advanced statistical models. In this way, mathematics can contribute directly to environmental understanding and to the broader goal of sustainable development.

References

1. Central Pollution Control Board. (2018). Annual report 2017–18. Ministry of Environment, Forest and Climate Change, Government of India.
2. Central Pollution Control Board. (n.d.). Action plan for severely polluted Peenya industrial area. Government of India.
3. Central Pollution Control Board, Regional Directorate (South), Bengaluru. (2018). Annual report 2017–18.
4. World Health Organization. (2021). WHO global air quality guidelines: Particulate matter (PM_{2.5} and PM₁₀), ozone, nitrogen dioxide, sulfur dioxide and carbon monoxide. World Health Organization.
5. World Health Organization. (2021). WHO global air quality guidelines: Executive summary. World Health Organization.
6. World Health Organization. (2026). Air quality standards database (3rd ed.). World Health Organization.
7. World Health Organization. (2024). Ambient (outdoor) air pollution. World Health Organization.

8. Chen, S. (2017). Beijing multi-site air quality data. UCI Machine Learning Repository. <https://doi.org/10.24432/C5RK5G>
9. De Vito, S., Massera, E., Piga, M., Martinotto, L., & Di Francia, G. (2008). On field calibration of an electronic nose for benzene estimation in an urban pollution monitoring scenario. *Sensors and Actuators B: Chemical*, 129(2), 750–757.
10. Box, G. E. P., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). *Time series analysis: Forecasting and control* (5th ed.). Wiley.
11. Wilks, D. S. (2011). *Statistical methods in the atmospheric sciences* (3rd ed.). Academic Press.
12. Montgomery, D. C., Peck, E. A., & Vining, G. G. (2021). *Introduction to linear regression analysis* (6th ed.). Wiley.
13. Kumar, N., & Sahu, M. (2026). Understanding PM_{2.5} predictability across diverse urban environments: A comparative evaluation of statistical forecasting models in four Indian cities. *Intelligent Climate and Eco-Environment*, 1, 100015.
14. Sustainable data-driven insights: Statistical analysis and artificial intelligence-driven modelling of aerosol concentrations in Hyderabad district, India. (2024). *Environmental Advances*, 16, 100539.