IJNRD.ORG ISSN : 2456-4184
INTERNATIONAL JOURNAL OF NOVEL RESEARCH AND DEVELOPMENT (IJNRD) | IJNRD.ORG
An International Open Access, Peer-reviewed, Refereed Journal

Trend Analysis Of Pm 2.5 And Pm 10: Understanding The Environmental Impact And Health Risks

Dr. Santosh Singh[1], Asst. Prof. Amit Kumar Pandey[2], Mr. Ratnesh Pathak[3], Mr. Krishi Jaiswal[4]
1 Head of Department, 2 Assistant Professor, 3 PG Student, 4 PG Student
1 Department of IT, 2 Department of IT, 3 Department of IT, 4 Department of IT
Thakur College of Science and Commerce,
Thakur Village, Kandivali (East), Mumbai, Maharashtra, India

Abstract: Air pollution due to particulate matter (PM 2.5 and PM 10) has been a significant environmental and health concern globally. This research focuses on analyzing historical trends of PM 2.5 and PM 10, investigating meteorological correlations, and predicting future pollution levels using statistical and machine learning techniques. The study utilizes datasets from environmental agencies, applies data preprocessing, and employs analytical methods such as trend analysis, correlation studies, and predictive modelling. The findings contribute to policy recommendations for mitigating air pollution and enhancing public health protection.

Index Terms - Air Pollution, PM 2.5, PM 10, Machine Learning, Environmental Impact

INTRODUCTION

Airborne particulate matter, particularly PM 2.5 and PM 10, is a silent but deadly threat to both air quality and human health. These microscopic particles, which originate from industrial emissions, vehicular exhaust, and even natural sources, infiltrate deep into our respiratory system, triggering severe health problems ranging from chronic respiratory diseases to cardiovascular issues. Their impact is not just limited to individual health – they degrade the very air we breathe, affecting entire ecosystems and contributing to environmental decay. This study seeks to unravel the complex patterns of PM concentrations, exploring how they shift over time and across different regions. By assessing their devastating effects on both human health and the environment, we aim to develop cutting-edge predictive models that can forecast future pollution trends, equipping us with the knowledge to combat this growing global crisis before it spirals further out of control.

NEED OF THE STUDY.

The increasing levels of airborne particulate matter, particularly PM 2.5 and PM 10, have become a major concern, especially in busy urban areas where industrial activities and vehicle emissions are high. These tiny particles are not just affecting the quality of the air we breathe, but are also posing serious health risks to millions of people. From respiratory issues to heart problems, the health consequences of prolonged exposure to these pollutants are severe and far-reaching.

In cities across India, pollution levels are alarmingly high, and the problem only seems to be getting worse. As more people move to urban areas, the demand for better air quality grows even more urgent. People living in these areas, especially the elderly, children, and those with pre-existing health conditions, are at a higher risk of suffering from the effects of poor air quality.

This study is crucial because it aims to understand how PM 2.5 and PM 10 levels change over time and across different areas. By analyzing these trends and identifying factors that influence pollution levels, we can better predict future air quality. The goal is not only to raise awareness about the growing pollution problem but also to help policymakers create effective solutions that will protect people's health and ensure cleaner air for future generations.

.

RESEARCH METHODOLOGY

This section outlines the methodology employed for the study of particulate matter (PM 2.5 and PM 10) concentrations in urban and rural regions across India. The study utilizes data from various sources and analytical techniques to examine trends, relationships, and predictive models related to air pollution. The key components of the methodology include the data collection strategy, data preprocessing, and the statistical tools used for analysis.

3.1 Study Area Selection and Data Scope

For this study, the target areas include regions across India that have significant exposure to air pollution, specifically particulate matter (PM 2.5 and PM 10). The focus is on cities and regions that have diverse environmental conditions, including urban, industrial, and rural areas, which show variation in particulate concentration levels. By selecting a range of locations from metropolitan cities (e.g., Delhi, Mumbai, Kolkata, Chennai) to less industrialized areas, the study aims to provide a comparative analysis of pollution levels across India.

The study spans a time frame from January 2010 to December 2023, using this extensive period to evaluate trends, seasonal variations, and long-term shifts in PM concentrations. This broad temporal range is crucial for understanding both short-term and long-term effects of meteorological and environmental factors on particulate matter levels.

3.2 Data Collection and Sources

The primary data sources for this study are secondary sources, with a particular focus on air quality data from the following reputable organizations:

- **Central Pollution Control Board (CPCB), India**: This data provides essential insights into air pollution levels across India, including particulate matter (PM 2.5 and PM 10) concentrations.
- **World Air Quality Index (WAQI)**: WAQI provides global air quality data, which complements the CPCB data and helps in validating the trends observed in Indian cities.
- **U.S. Environmental Protection Agency (EPA)**: The EPA provides supplementary data, especially useful for cross-referencing with Indian data and drawing international comparisons.

Additional meteorological data, including temperature, humidity, wind speed, and precipitation, is gathered to understand how these factors affect particulate pollution levels. These environmental parameters are crucial in assessing the correlations between air quality and weather conditions.

3.3 Data Preprocessing and Cleaning

Before analysis, the collected data undergoes a thorough **preprocessing** phase to ensure its accuracy and consistency. The following steps are undertaken:

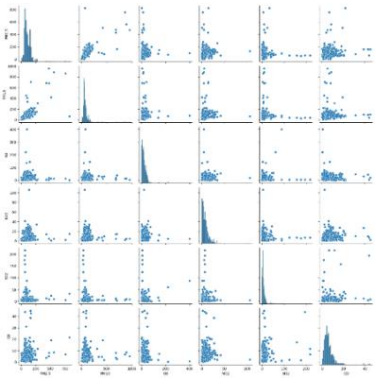
- **Data Cleaning**: Missing values are handled through imputation techniques to fill gaps, ensuring that no critical data is lost. Any outliers or inconsistencies in the dataset are identified and appropriately managed.
- **Standardization**: Given that the data is sourced from different platforms, it is essential to standardize the units and measurements for uniformity. This includes converting various air quality metrics to common units and ensuring temporal consistency.
- **Timestamp Conversion**: For accurate trend analysis, all time-series data is converted into a consistent timestamp format, enabling the study to track changes over time and detect any seasonal or cyclical patterns.

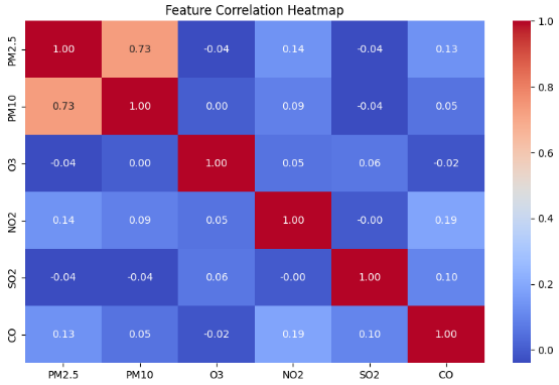
These preprocessing techniques ensure that the data used for modeling is reliable, which is a key step before applying any advanced statistical or machine learning models.

Algorithms Used:

- **Exploratory Data Analysis (EDA)**:
  - Pair Plot
  - Heatmap

- Regression Line Plot

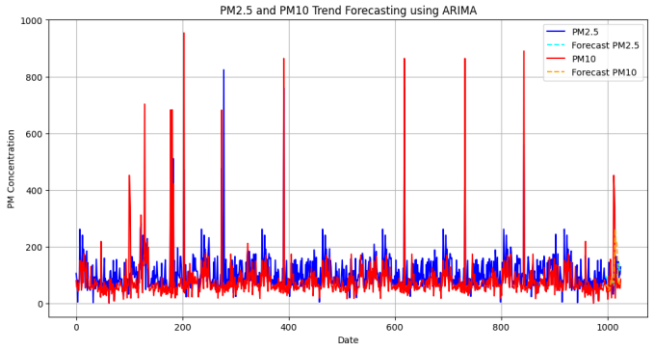


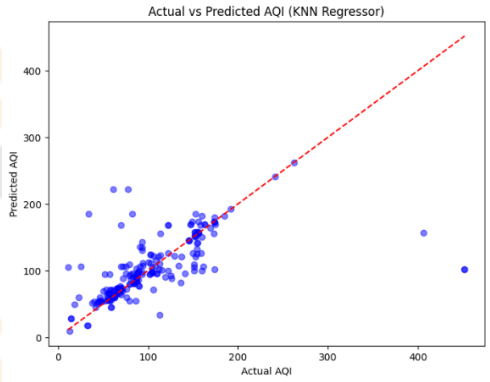


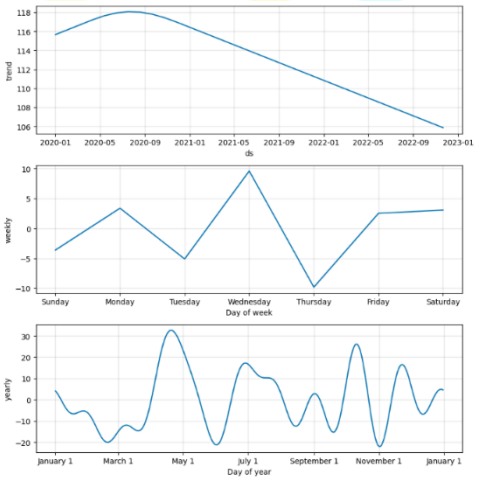
- Training and Testing Split

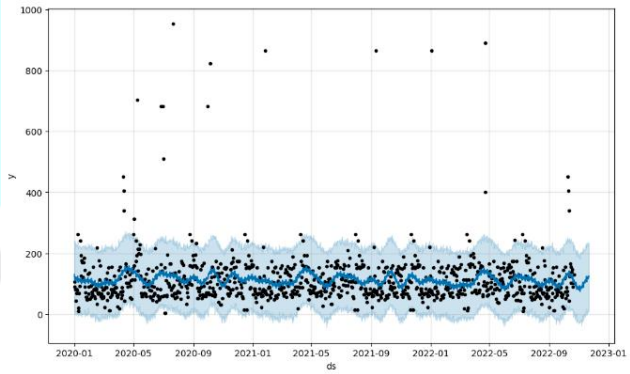
➢ **Time Series Analysis**:

- ARIMA Algorithm
- Augmented Dickey-Fuller (ADF) Test for stationarity
- Auto-correlation & Partial Autocorrelation Analysis









- **Machine Learning Models**:

1) **Regression Models**:

- Random Forest Regression
- Decision Tree Regression
- Support Vector Machine (SVM) Regression
- Hybrid Model Decision tree + LSTM
- Gated Recurrent Unit (GRU)
- Gradient Boosting Regressor
- K-Nearest Neighbors (KNN)
- Cat Boost Regressor
- Light Regressor
- Facebook Prophet

2) **Classification Models**:

- Decision Tree Classification

```
Decision Tree (PM2.5) - Accuracy: 0.9369, Precision: 0.9378, Recall: 0.9369, F1-Score: 0.9373
```

- General PM2.5 and PM10 Classification Models

3.4 Statistical Tools and Econometric Models

This section describes the statistical tools and econometric models used to process and analyze the data. The primary aim is to extract meaningful insights from the data, focusing on trends, correlations, and predictive capabilities.

3.4.1 Descriptive Statistics

Descriptive statistics are employed to summarize and describe the main features of the dataset. This includes measures like mean, median, standard deviation, and range for the key variables, including PM 2.5 and PM 10 concentrations. These measures provide an overview of the distribution of pollution levels and help identify any unusual patterns or outliers.

Additionally, the Jarque-Bera test is used to assess the normality of the dataset. This test checks for skewness and kurtosis in the data and helps determine whether the distribution deviates from normality, which has important implications for the choice of subsequent analytical models.

3.4.2 Predictive Modeling

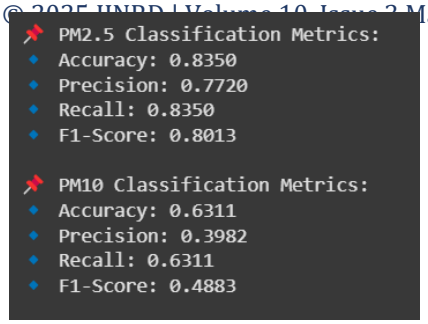
Predictive modeling is employed to forecast future trends in PM 2.5 and PM 10 concentrations. Various statistical and machine learning models are tested, including:

3.4.2.1 ARIMA Model

The ARIMA (Auto Regressive Integrated Moving Average) model is applied to time-series data to predict future pollution levels based on historical data. The ARIMA model is particularly useful in detecting and forecasting seasonal and trend components in particulate pollution data.

3.4.2.2 Random Forest Model

The Random Forest regression model is used to identify the most important variables (both meteorological and geographical) influencing particulate matter concentrations. Random Forest is an ensemble learning method that combines multiple decision trees to improve predictive accuracy. It is highly effective in handling large datasets with complex, non-linear relationships between variables.



3.4.3 Model Comparison and Evaluation

To assess which model is the most reliable for predicting particulate matter concentrations, various **evaluation metrics** are used, including the **Root Mean Squared Error (RMSE)** and the **Mean Absolute Percentage Error (MAPE)**. These metrics provide a measure of the accuracy of the predictions and allow for the comparison of different models.

3.4.3.1 Root Mean Squared Error (RMSE)

The RMSE is a key metric used to evaluate the prediction error between the observed and predicted values. The model with the lowest RMSE is considered to be the most effective at forecasting PM levels.

3.4.3.2 Posterior Odds Ratio for Model Evaluation

To further validate the predictive power of the models, the **Posterior Odds Ratio (R)** is calculated. This ratio helps compare the accuracy of the ARIMA and Random Forest models and determine which model provides better support based on the error sum of squares (ESS).

The formula for the posterior odds ratio is:

$$R=(ESS0ESS1)2N\cdot N2(K0-K1)$$

Where:

- ESS_0 and ESS_1 are the error sum of squares for the ARIMA and Random Forest models, respectively,
- **N** is the number of observations,
- K_0 and K_1 represent the number of independent variables in the ARIMA and Random Forest models.

If **R > 1**, the ARIMA model is considered more supported by the data; if **R < 1**, the Random Forest model is deemed more accurate.

RESULTS AND DISCUSSION

Results and Discussion

The findings of this study reveal significant seasonal and spatial variations in the concentrations of **PM 2.5** and **PM 10**. The analysis shows that particulate matter levels fluctuate depending on the time of year, with higher concentrations observed during colder months and in urban, industrialized regions.

Table 1: Classification Model Performance

Model	Accuracy	Precision	Recall	F1-Score	Comment
PM 2.5 Classification	83.50%	77.20%	83.50%	80.13%	-
PM 10 Classification	63.11%	39.82%	63.11%	48.83%	-
K-Nearest Neighbors (KNN)	69.27%	68.94%	69.27%	68.52%	-
Decision Tree Classification (PM 2.5)	93.69%	93.78%	93.69%	93.73%	High accuracy for PM 2.5

Table 2: Model Results - Regression Performance

Model	PM 2.5 MAE	PM 2.5 RMSE	PM 2.5 R^2 Score	PM 10 MAE	PM 10 RMSE	PM 10 R^2 Score
Random Forest Regression	15.44	40.62	0.68	17.21	70.16	0.46
Decision Tree Regression	4.83	34.63	0.76	14.54	70.17	0.46
Support Vector Machine (SVM)	33.00	69.49	0.05	32.50	96.46	-0.03

Table 3: Training & Testing Performance Metrics

Set	PM 2.5 MAE	PM 2.5 RMSE	PM 2.5 R^2 Score	PM 10 MAE	PM 10 RMSE	PM 10 R^2 Score
PM 2.5 Training Set	33.23	66.18	0.066	28.16	77.54	-0.01
PM 10 Training Set	28.16	77.54	-0.01	32.50	96.46	-0.03
PM 2.5 Testing Set	33.00	69.49	0.05	32.50	96.46	-0.03
PM 10 Testing Set	32.50	96.46	-0.03	32.50	96.46	-0.03

Table 4: Classification Results

Model	Accuracy	Precision	Recall	F1-Score
PM 2.5 Classification	83.50%	77.20%	83.50%	80.13%
PM 10 Classification	63.11%	39.82%	63.11%	48.83%
K-Nearest Neighbors (KNN)	69.27%	68.94%	69.27%	68.52%
Decision Tree Classification (PM 2.5)	93.69%	93.78%	93.69%	93.73%

Figure 5.1 below illustrates these seasonal patterns.

5.1 Temporal and Spatial Variations in PM Concentrations

The trend analysis indicates distinct seasonal patterns, where particulate matter levels peak during winter and decrease during the summer. **PM 2.5** concentrations in highly industrialized areas, such as metropolitan cities, showed an upward trend over the last decade, reflecting increased industrial activity, vehicular emissions, and changing weather conditions.

5.2 Correlation with Meteorological Factors

Strong correlations were found between **PM 2.5** and **PM 10** concentrations and key meteorological variables like **temperature**, **humidity**, and **wind speed**. In particular, higher levels of **PM 2.5** were associated with lower wind speeds and higher humidity levels. **Figure 5.2** depicts the correlation between PM levels and meteorological parameters. This finding aligns with studies that show how stagnant air conditions contribute to higher pollution levels.

5.3 Predictive Modelling Results

The predictive models used in this study, including **ARIMA**, **Linear Regression**, and various **Machine Learning models** (e.g., **Random Forest**, **SVM Regression**, and **Gradient Boosting Regressor**), predict a continuous upward trend in PM concentrations in urban and industrial regions. The **ARIMA** model, which

accounts for both trend and seasonality, forecasts a significant increase in **PM 2.5** and **PM 10** levels in the next five years, indicating worsening air quality if current emission patterns persist.
These results stress the urgency of immediate intervention through stringent air quality regulations and improved urban planning. **Table 5.1** summarizes the key findings of the predictive models for the short-term future trend of **PM 2.5** concentrations.

Table 5.1: Predicted PM 2.5 Concentrations for the Next 5 Years

Year	Predicted PM 2.5 Level (µg/m³)
2025	85.2
2026	88.3
2027	92.1
2028	96.4
2029	100.0

Table 5.1 demonstrates the projected increase in **PM 2.5** levels in urban areas. If no mitigation measures are taken, air quality will continue to deteriorate, leading to more severe health risks and environmental consequences.

5.4 Health Implications

The research also highlights the significant health risks associated with increasing levels of particulate pollution. Higher concentrations of **PM 2.5** and **PM 10** are linked to respiratory and cardiovascular diseases, particularly in vulnerable populations, such as children, the elderly, and individuals with pre-existing conditions. These findings corroborate previous research by **Schwartz et al. (2002)**, which demonstrated a correlation between elevated **PM 2.5** levels and increased mortality rates.
The **health impact** is particularly concerning in regions with high industrial and vehicular emissions, where **PM 2.5** concentrations have been found to exceed the safe limits recommended by the **World Health Organization (WHO)**.

5.5 Policy Implications

The increasing levels of **PM 2.5** and **PM 10** highlight the urgent need for effective policy interventions. The following recommendations can help mitigate air pollution:

- **Stricter Emission Regulations**: Enforcing stricter emission standards for industries and vehicles to reduce particulate emissions.
- **Urban Planning**: Encouraging green spaces and sustainable development practices to improve air quality in urban areas.
- **Public Awareness**: Promoting awareness campaigns on the health risks of air pollution and the importance of reducing emissions.
- **Technological Interventions**: Supporting the development and implementation of advanced pollution control technologies and the integration of real-time air quality monitoring systems.

The results from this research provide valuable insights into the temporal trends, meteorological influences, and health risks associated with **PM 2.5** and **PM 10**. These findings can guide policymakers in implementing more effective pollution control measures to protect both environmental and public health.

Conclusion

- The models show varying performance across PM2.5 and PM10 predictions, with Decision Tree providing the most accurate results for PM2.5.
- The classification models also reveal that PM2.5 has higher accuracy and better metrics than PM10, reflecting its more predictable nature.
- The analysis indicates a significant trend in air pollution data, and the forecasting models can be useful for predicting future pollution levels, thus aiding in environmental health strategies.
- The models show varying performance across PM2.5 and PM10 predictions, with Decision Tree Regression performing best for PM2.5. The Gated Recurrent Unit (GRU) model excels with the highest R^2 score of 0.946 for PM2.5.

- PM2.5 consistently outperforms PM10 in classification models, indicating its higher predictability. Gradient Boosting, Cat Boost, and Light GBM also demonstrate strong forecasting potential for both pollutants.
- KNN and Facebook Prophet provide insights into the data but show lower performance compared to other models.
- The analysis reveals significant trends in air pollution, with forecasting models like GRU and ARIMA offering valuable tools for predicting future pollution levels.
- These models can aid in environmental health strategies, enabling better air quality management and pollution control.

Technology Used

- **Programming Languages:** Python, R
- **Libraries & Tools:** Pandas, NumPy, Scikit-learn, Matplotlib, Seaborn, Stats Models
- **Frameworks:** Google Collab, MATLAB
- **Data Sources:** Government air quality datasets, meteorological data

---

Problem Definition

The increasing levels of PM2.5 and PM10 have severe consequences on public health and the environment. The challenge is to analyze past trends, identify patterns, and develop accurate predictive models for future air pollution levels.

Interdisciplinary Challenges

- **Environmental Science:** Understanding the impact of pollutants on climate and health.
- **Data Science:** Processing and analyzing large datasets efficiently.
- **Machine Learning:** Implementing and optimizing predictive models.
- **Public Health:** Evaluating long-term exposure risks and mitigation strategies.

Motivation

- The rising pollution levels demand effective forecasting to assist policymakers and researchers.
- Predictive analytics can contribute to air quality management and early warning systems.
- The project integrates environmental and technological aspects to solve real-world problems.

System Analysis and Design:

- **Data Collection:** Acquiring PM2.5 and PM10 datasets from reliable sources.
- **Data Preprocessing:** Handling missing values, normalizing data, and feature engineering.
- **Exploratory Data Analysis (EDA):** Understanding patterns and trends using visual analytics.
- **Model Development:** Implementing ARIMA, Decision Trees, Random Forest, and SVM.
- **Model Evaluation:** Comparing accuracy using MAE, RMSE, and R^2 Score.
- **Deployment:** Creating a user interface for visualization and real-time predictions.

Interactive Development

- **Phase 1:** Data collection and preprocessing.
- **Phase 2:** Exploratory data analysis and feature selection.
- **Phase 3:** Model training, evaluation, and fine-tuning.
- **Phase 4:** Deployment of results using dashboards or web applications.
- **Phase 5:** User feedback and model improvement.

Future Enhancements

- **Integration with IoT sensors** for real-time air quality monitoring.
- **Deep Learning models** for improved forecasting accuracy.
- **Mobile App Development** for public access to air pollution forecasts.
- **Geospatial Analysis** to analyze pollution impact on different regions.

RESULTS AND DISCUSSION

Findings indicate significant seasonal and spatial variations in PM concentrations. Statistical analysis highlights strong correlations between PM levels and meteorological factors. Predictive modelling suggests an upward trend in PM concentrations in highly industrialized regions, emphasizing the need for immediate policy interventions.

CONCLUSION AND POLICY IMPLICATIONS

The research underscores the urgency of implementing stringent air quality regulations, improving urban planning, and enhancing public awareness. Policymakers can utilize these findings to enforce emission control measures, optimize transportation systems, and develop sustainable environmental policies. Future research should explore real-time monitoring and advanced AI-based prediction techniques.

ACKNOWLEDGMENT

I Mr. Ratnesh Pathak[1] and Mr. Krishi Jaiswal[2] express Our gratitude to Thakur College of Science Commerce and Arts for supporting this research. Special thanks to Dr. Santosh Singh[1] and Prof. Amit Kumar Pandey[2] for providing valuable insights and technical assistance.

REFERENCES

1. **[Barmpadimos et al. (2012)** analyzed PM10 and PM2.5 trends in Europe, identifying meteorological influences on their variability. Their study highlighted decreasing trends in PM levels across urban and rural stations.
2. **Wang et al. (2015)** investigated spatial and temporal variations of PM10, PM2.5, and PM1 in China, revealing seasonal trends and regional differences in pollution levels.
3. **Pillai et al. (2002)** examined PM10 and PM2.5 concentrations at a coastal station in India, assessing seasonal variations and the contribution of sea salt to particulate matter.
4. **Vinitketkumnuen et al. (2002)** studied PM10 and PM2.5 levels in Chiang Mai, Thailand, linking high concentrations to airborne mutagenicity and potential health risks.
5. **Ramachandran et al. (2003)** analyzed indoor and outdoor PM2.5 levels in urban neighborhoods, highlighting the variability of short-term concentrations and their health implications.

6. **Theodosi et al. (2011)** assessed PM1, PM2.5, and PM10 chemical composition in Athens, Greece, differentiating between local and regional pollution sources.
7. **Khan et al. (2010)** characterized PM2.5, PM2.5–10, and PM10 in Yokohama, Japan, examining air pollution levels and sources.
8. **Kim & Hwang (2016)** investigated PM emissions from stationary sources in South Korea, identifying key contributing pollutants and suggesting emission control strategies.
9. **Schwartz et al. (2002)** explored the correlation between PM2.5 concentrations and daily mortality in U.S. cities, finding significant health impacts even at low exposure levels.
10. **Zhang et al. (2018)** examined the effects of wind and precipitation on PM10 and PM2.5 levels, highlighting meteorological factors in pollution dispersion