An Analytical Study of Atmospheric Pollution Levels in the Maharashtra Region

Jayashree Bhuskute[1], Dr. Ashok Tayade[2]

Abstract
Air pollution represents a major environmental and public health concern in Maharashtra, driven by rapid urbanization, industrial growth, and increasing vehicular emissions. This study presents a region-specific statistical and forecasting assessment of atmospheric pollutants across six major regions - Amravati, Chhatrapati Sambhajinagar (Aurangabad), Konkan, Nagpur, Nashik, and Pune using long-term monitoring data obtained from the Maharashtra Pollution Control Board (MPCB). Station-level observations were standardized and aggregated into monthly regional averages for AQI, RSPM, SPM, NO_x, SO_2, and CO. Descriptive statistical analysis and correlation assessment indicate that particulate matter is the dominant contributor to air quality variability, with AQI exhibiting strong positive association with RSPM across regions. Marked spatial heterogeneity is observed, reflecting differences in emission sources, industrial intensity, population density, and regional meteorological conditions. Seasonal patterns further highlight the influence of atmospheric dispersion and climatic factors on pollutant concentration levels.

To examine temporal dynamics and predictive capability, three univariate time-series forecasting frameworks Prophet, ARIMA, and ETS were implemented and comparatively evaluated using scale-dependent and percentage-based error metrics. The results demonstrate that forecasting accuracy varies across pollutants and regions. Models perform reliably for series characterized by clear trend and seasonal structure, particularly for AQI and particulate matter. However, pollutants exhibiting intermittent behavior or near-zero concentrations show unstable percentage-based error measures, underscoring statistical limitations of certain evaluation metrics in environmental applications. Overall, the study establishes a statistically rigorous framework for regional air quality assessment and forecasting in Maharashtra. The findings highlight the importance of model selection, metric stability, and regional context in environmental time-series analysis, providing quantitative evidence to support informed regulatory planning and public health interventions.

[1]Department of Statistics Dr. BAMU, Chh.Sambhajinagar 431004(M.S) India jyp4329@gmail.com9405678388
[2]Department of Statistics Dr. BAMU, Chh.Sambhajinagar 431004(M.S India aytayade@gmail.com9421307784

Keywords: Atmospheric pollution; Air Quality Index (AQI); Particulate matter (RSPM, SPM); SO_2; Prophet model; ARIMA: Exponential Smoothing (ETS); Time-series forecasting.

1 Introduction

Air pollution is a critical environmental challenge with profound implications for public health, ecosystem stability, and climate dynamics. Maharashtra, one of India's most industrialized and densely populated states, faces considerable challenges in maintaining air quality due to rapid urbanization, industrial expansion, and increasing vehicular emissions. The state's diverse land- scape, encompassing major urban centers such as Mumbai and Pune, industrial corridors, and rural agricultural areas, creates a complex mosaic of pollution sources and exposure patterns.

This study presents a comprehensive analysis of air pollution across Maharashtra, emphasizing spatial and temporal variability in pollutant concentrations, identifying principal emission sources, and evaluating the effectiveness of existing regulatory measures. By leveraging data from multiple air quality monitoring stations across the state, the study seeks to elucidate patterns and trends in air pollution, providing insights into the factors driving both episodic pollution events and persistent air quality challenges.

The health impacts of air pollution in Maharashtra are significant, particularly regarding respiratory and cardiovascular diseases in regions with sustained pollutant exposure. This re- search adopts a multidisciplinary approach, integrating environmental monitoring, public health assessment, and policy analysis to inform targeted mitigation strategies. The findings aim to support evidence-based interventions that reduce emissions and minimize adverse health and environmental outcomes.

Maharashtra's regions exhibit distinct air pollution profiles shaped by their geography, industrialization, and urbanization:

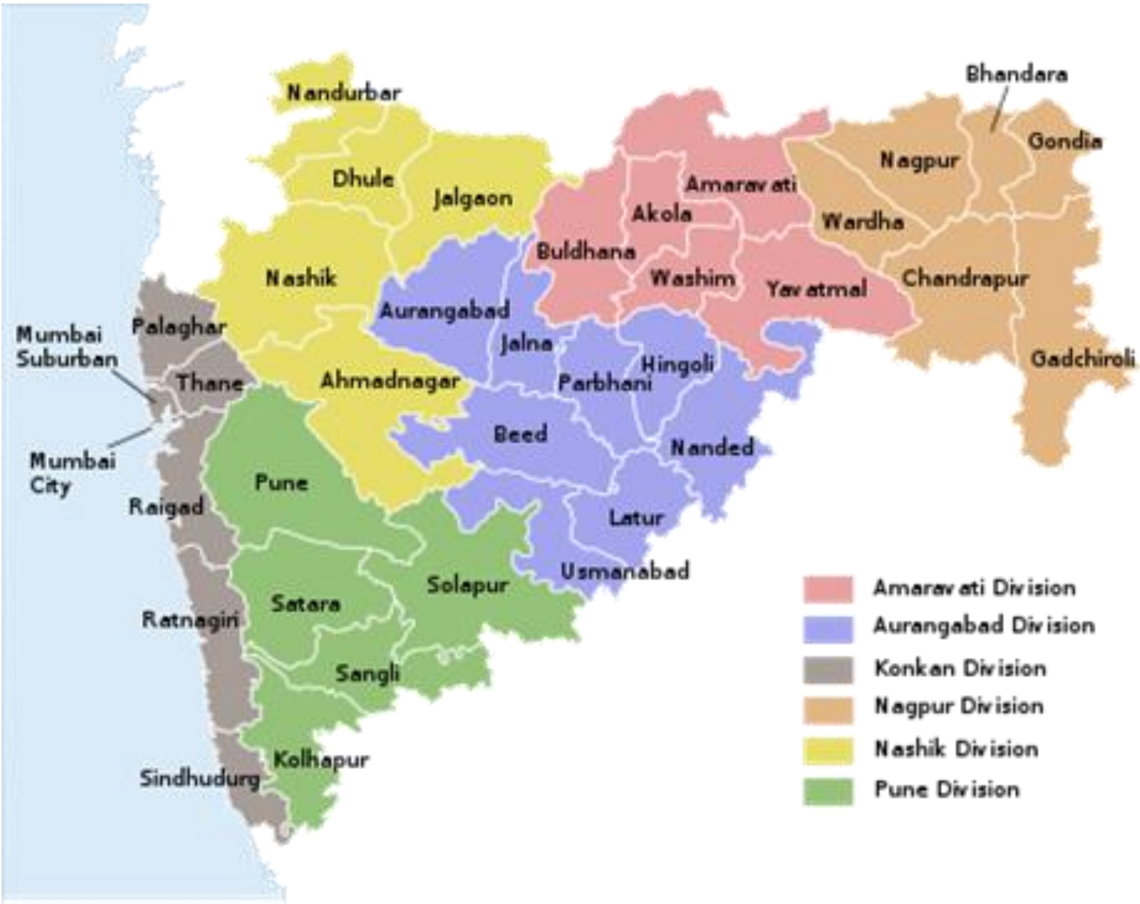

Figure 1: Maharashtra Regions

Konkan Region: A coastal and predominantly green region with relatively low industrial activity. Primary pollutants include $PM_{2.5}$ and PM_{10}, primarily generated from vehicular traffic, tourism, and small-scale industries.

Pune Region: A major urban and educational hub, with high vehicular emissions, industrial activity, construction dust, and occasional biomass burning. Key pollutants are PM_{10}, $PM_{2.5}$, and NO_x, with episodic ozone pollution.

Amravati Region: Predominantly agricultural with emerging urbanization. Pollution sources include crop residue burning, fertilizers, vehicles, and small industries, contributing to elevated levels of PM, ammonia, NO_x, and SO_2.

Nashik Region: Characterized by industrial and agricultural activities (vineyards). Major pollutants arise from

industrial emissions, vehicular traffic, and residue burning, including PM_{10}, $PM_{2.5}$, SO_2, and NO_2.
Aurangabad Region: An industrial and tourism hub, where air quality is impacted by industrial emissions, traffic, and waste burning. Dominant pollutants include PM, SO_2, NO_x, and VOCs.
Nagpur Region: A significant urban center with high vehicular and industrial emissions, along with biomass burning, leading to elevated PM_{10}, $PM_{2.5}$, NO_x, and SO_2, particularly in winter months.
By integrating these regional characteristics, the study provides a robust framework for statistical analysis and forecasting of air quality across Maharashtra. Understanding this spatial and temporal variability enables the development of targeted interventions and predictive models, tailored to the distinct emission profiles and environmental contexts of each region.

2 Objectives of the Study

This study aims to provide a comprehensive, data-driven assessment of air pollution across major regions of Maharashtra, integrating descriptive analysis and forecasting of pollutant trends. The specific objectives are:
1. **Quantify Regional Variability:** Assess spatial and temporal patterns of air pollutant concentrations across Konkan, Pune, Amravati, Nashik, Aurangabad, and Nagpur, capturing seasonal dynamics, trends, and regional heterogeneity influenced by geography, urbanization, and industrial activity.
2. **Identify Key Pollution Sources:** Characterize and rank dominant emission sources in each region, including industrial operations, vehicular traffic, biomass and crop residue burning, and construction-related dust, to understand drivers of observed air quality pat- terns.
3. **Develop and Assess Forecasting Models:** Implement and comparatively evaluate short- and medium-term air quality forecasting frameworks using Prophet, ARIMA, and ETS models to capture trend, seasonal, and stochastic components of key pollutants (PM_{10}, $PM_{2.5}$, NO_x, SO_2, CO, and SPM). The objective is to assess predictive reliability across regions and pollutant types, quantify model performance using appropriate error metrics, and support early warning systems, proactive interventions, and evidence-based air quality management strategies.

3 Literature review

Air pollution forecasting has evolved significantly over the past two decades, driven by the advancement of statistical learning, deep learning, hybrid modeling, and environmental policy analysis. Bai et al. [1] provide one of the most comprehensive early overviews of air pollution forecasting, summarizing traditional statistical methods and identifying limitations related to nonlinearity and data sparsity. Building upon these insights, Tao et al. [2] introduce a deep learning architecture combining 1D convolutional neural networks (CNNs) with bidirectional gated recurrent units (BiGRU), demonstrating enhanced predictive capability for temporal pollutant dynamics.
Several studies have emphasized the integration of deep learning architectures with optimization strategies. Heydari et al. [3] propose a deep learning optimization hybrid model that improves forecasting accuracy by tuning hyperparameters for pollutant concentration prediction. Similarly, Chang et al. [5] develop an aggregated long short-term memory (LSTM) framework that fuses multiple temporal features to generate robust predictions under varying environmental conditions. Wang et al. [10] further advance this domain by presenting a multipollutant deep learning framework capable of concurrently modeling interdependencies among pollutants such as $PM_{2.5}$, NO_x, and SO_2. Complementing these efforts, Chen et al. [15] introduce hybrid industrial pollution prediction models that combine numerical simulation outputs with machine learning approaches, effectively addressing industrial sector emissions.
Hybrid and ensemble modeling approaches have also gained traction. Li et al. [7] employ hybrid deep learning models integrating CNN and LSTM layers for improved air quality pre- diction accuracy. Additional advancements are seen in studies such as Alghamdi et al. [8] and Kumar et al. [9], which explore machine learning methods for sustainable air quality management and $PM_{2.5}$ forecasting respectively, demonstrating the increasing relevance of data-driven methodologies in regulatory and policy applications. Li et al. [25] integrate chemical trans- port modeling with machine learning, offering a physics informed, data enhanced approach to atmospheric pollutant estimation.
From a policy and environmental management perspective, Gulia et al. [4] present a detailed analysis of air quality management policies in India, highlighting gaps in urban pollution mitigation strategies. Complementary to this, Garaga et al. [6] review urban road transport emissions, elucidating their impact on air quality trends. Furthermore, transportation-specific studies such as Gupta and Verma [13] and Desai et al. [17] employ statistical and ARIMA-based models to examine emissions from traffic systems, reinforcing the significance of sector specific pollutant modeling.
Statistical and classical modeling approaches continue to play a crucial foundational role. Early works, such as Patel et al. [19], outline classical forecasting models for air pollution, including ARIMA and regression-based methods. More recent studies, such as Zhang et al. [20], extend regression modeling to nonlinear environmental time series. Stochastic approaches, exemplified by Liu et al. [21], incorporate atmospheric randomness and uncertainty. Grey prediction models proposed by Zhang et al. [23] demonstrate strengths in limited-data environments where traditional machine learning models struggle.
With the growth of environmental monitoring systems, deep learning enabled frameworks have expanded significantly. Wang et al. [24] survey deep learning systems for large-scale environmental forecasting, while Chen et

al. [26] apply advanced neural networks for hazardous air pollutant prediction. In parallel, domain-specific advancements in multi-pollutant forecasting techniques are synthesized in Atmospheric Pollution Research [22] and APR [28], indicating continuous refinement of modeling methodologies.

Preprocessing and data preparation, crucial elements of any forecasting pipeline, are reviewed by Sharma et al. [27], who outline methods for noise reduction, normalization, handling missing values, and feature extraction. Similarly, Kumar and Rao [16] highlight the role of ma- chine learning in environmental impact assessment, demonstrating applications across diverse ecological datasets. Additional computational frameworks, including optimization-centric pollution prediction models [11] and modern software tools for air quality analysis [18], contribute to the operationalization of air quality analytics.

Time-series-specific innovations are reported by Singh et al. [14], where advanced forecasting models capture long-term dependencies in environmental data streams. Grey systems, stochastic processes, ensemble hybrids, and optimization-enhanced deep networks collectively form an ecosystem of increasingly sophisticated approaches to air quality forecasting.

Overall, the reviewed literature illustrates a clear evolution from traditional statistical models toward hybrid, deep learning, and physics informed frameworks. These advancements have significantly improved predictive accuracy and enhanced understanding of pollutant dispersion dynamics, enabling more effective environmental management, early warning systems, and policy decision making. Despite progress, challenges persist in interpretability, real time deployment, data quality, and model generalization across heterogeneous regions, underscoring the need for continued research in scalable, explainable, and multimodal forecasting solutions

.

4 Study Area and Data Description

4.1 Overview of the Maharashtra Region

Maharashtra located along the western coast of India, exhibits a highly diverse geographical and environmental profile that spans coastal belts, dense metropolitan centers, industrial corridors, agricultural zones, and semi-arid inland districts. This heterogeneity produces region specific emission patterns and complex atmospheric dynamics. Coastal regions such as Konkan typically experience cleaner air influenced by marine winds, whereas rapidly urbanizing centers like Pune and Nashik face rising vehicular and construction related emissions. Industrial hubs in Aurangabad and Nagpur, along with agricultural regions such as Amravati, contribute additional pollutant loads from manufacturing activities, biomass burning, residue burning, and small-scale industries. Seasonal meteorological conditions-including monsoon-driven ventilation, post-monsoon stagnation, and wintertime thermal inversions-further modulate pollutant dispersion and accumulation across the state. These variations create distinct pollution signatures that necessitate region-wise analysis rather than a uniform statewide assessment.

4.2 MPCB Monitoring Network and Measured Pollutants

The Maharashtra Pollution Control Board (MPCB) operates an extensive network of continuous and manual ambient air quality monitoring stations distributed across industrial, residential, traffic-intensive, and background locations. These stations collect multi-pollutant measurements at regular intervals, generating long-term datasets suitable for temporal and spatial analyses. The core measured pollutants include particulate matter ($PM_{2.5}$, PM_{10}), Sulphur dioxide (SO_2), nitrogen oxides (NO_x), suspended particulate matter (SPM), carbon dioxide (CO_2), and other auxiliary indicators where available. These pollutants were selected due to their strong influence on Air Quality Index (AQI) computation and their relevance to health risk assessments, regulatory compliance monitoring, and forecasting applications. The monitoring coverage spans all major regions examined in this study-Konkan, Pune, Nashik, Aurangabad, Amravati, and Nagpur-enabling robust region-wise comparisons supported by consistent regulatory data.

4.3 Data Acquisition, Processing, and Aggregation

Air quality data used in this study were programmatically scraped from the official MPCB web portal, which provides publicly accessible historical observations from individual monitoring stations. The dataset spans a period from **October 2006 to November 2025**, capturing nearly two decades of regional air quality variability across Maharashtra. The raw records comprised pollutant concentrations measured at heterogeneous temporal resolutions and across diverse monitoring station types.

Table 1: Descriptive statistics of air quality parameters (Month wise Aggregated Data)

Statistic	AQI	CO2	NOX	RSPM	SO2	SPM
Count	1501	183	1501	1501	1501	1295
Mean	90.06	88.82	31.36	94.25	14.85	157.55
Std	24.77	42.77	15.49	30.24	6.28	86.05
Min	27.15	24.50	5.88	24.00	4.00	0.00
25%	70.55	53.18	20.43	69.45	9.67	86.76
50%	89.49	80.38	29.07	92.24	13.93	156.25

75%	108.35	113.34	37.69	116.91	19.07	216.26
Max	195.28	299.00	119.14	205.46	43.52	488.81

To ensure methodological consistency and comparability across regions, a structured pre-processing pipeline was applied. First, station level records were cleaned to remove duplicates, incomplete entries, and erroneous values arising from sensor malfunction or transcription anomalies. Missing data were addressed using pollutant-specific imputation strategies suitable for environmental time series. Each observation was then annotated with its corresponding geographical region to enable spatial aggregation.

Subsequently, pollutant concentrations were aggregated into region-wise and month-wise averages, producing a harmonized dataset that highlights broader temporal trends while reducing the noise inherent in daily fluctuations. This aggregation approach aligns with practices in con temporary air quality literature, where monthly resolution supports seasonal pattern discovery, inter-pollutant correlation analysis and consistent input preparation for forecasting workflows.

The cleaned and structured dataset ultimately formed the basis for descriptive statistical analyses, correlation matrix construction and the development of the Prophet-based time series forecasting framework presented in later sections.

5 Methodology

5.1 Exploratory and Descriptive Analysis

The methodological workflow begins with an extensive exploratory and descriptive analysis of the multi-year air quality dataset obtained from MPCB monitoring stations. Region-wise and month-wise aggregated pollutant concentrations are examined to identify spatial heterogeneity and temporal dynamics across Maharashtra's diverse environmental zones. Descriptive statistical indicators-including mean, median, standard deviation, interquartile range, and extrema are computed for all major pollutants ($PM_{2.5}$, PM_{10}, NO_x, SO_2, CO_2, and SPM). Temporal analysis focuses on seasonal cycles driven by monsoon ventilation, wintertime inversions and post monsoon stagnation, while monthly trends help characterize recurrent pollution peaks specific to each region. Correlation matrices are constructed to evaluate inter-pollutant co-variation, highlighting interactions between particulate and gaseous pollutants that are consistent with findings in contemporary literature. These analyses collectively provide a diagnostic under-standing of pollutant behavior and serve as a foundation for subsequent forecasting tasks.

5.2 Conceptual Design of the Prophet Forecasting Framework

To develop an interpretable and robust forecasting system suitable for environmental time series, the Prophet model is selected as the primary forecasting tool. Prophet employs an additive decomposition approach, modelling pollutant concentrations as a combination of long-term trend, periodic seasonal components and user-defined external factors. Given the strong seasonal structure observed in air quality data, Prophet's built-in yearly and weekly seasonality components are utilized, while monthly aggregation serves as the primary input granularity. This structure allows the model to capture key environmental cycles such as winter accumulation and monsoon-driven dispersion. The model framework is aligned with recent literature emphasizing the need for interpretable forecasting systems alongside more complex deep learning alternatives. Daily or monthly averaged pollutant concentrations are formatted into Prophet's required structure (timestamp-value pairs), after which the model is trained to generate short- term pollutant forecasts for each region. The framework also supports future integration of holiday or event regressors (e.g., festival-related biomass burning), though these are optional for the current implementation.

5.3 Evaluation Metrics and Comparison with Alternative Models

To assess the reliability and performance of the Prophet-based forecasting framework, a comparative evaluation is conducted against several baseline statistical models used in air quality forecasting research. These include the persistence model, which assumes pollutant levels remain unchanged from the previous observation; the seasonal naive model, which replicates values from the previous corresponding season or month; and autoregressive integrated moving average (ARIMA), a classical but widely applied statistical method for environmental time series. Forecast accuracy is quantified using standard error metrics, including Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and Mean Absolute Percentage Error (MAPE). These metrics enable robust comparison across pollutants and regions, ensuring that the forecasting system is evaluated rigorously and consistently with methodologies highlighted in the atmospheric sciences and air quality forecasting literature.

6 Statistical Analysis

6.1 Region-wise Overview

6.1.1 Amravati Region

Table 2: Descriptive Statistics for Amravati (Month wise Aggregated Data)

	AQI	CO2	NOX	RSPM	SO2	SPM
count	228.0	13.0	228.0	228.0	228.0	214.0

mean	89.45	63.14	13.43	93.95	11.19	53.27
std	23.31	15.69	3.58	27.49	2.20	36.79
min	38.00	43.75	5.88	38.00	4.00	0.00
25%	68.57	52.00	11.34	68.77	9.59	29.52
50%	91.30	61.40	12.86	93.54	10.92	58.31
75%	109.97	69.33	15.08	117.59	12.76	79.78
max	164.43	96.67	51.81	195.86	16.21	228.00

The Amravati region exhibits moderate air quality conditions, with a mean AQI of 89.45, indicating that overall air quality is predominantly influenced by particulate matter concentrations. Respirable suspended particulate matter (RSPM) records the highest average level (93.95) and considerable variability (standard deviation of 27.49), reflecting pronounced temporal fluctuations and episodic pollution events. In contrast, gaseous pollutants such as SO_2 and NO_X display relatively lower mean concentrations and narrower dispersion, suggesting more stable emission patterns. The wide observed range of suspended particulate matter (SPM), extending from 0 to 228, further highlights intermittent high-intensity pollution episodes, likely driven by localized anthropogenic and environmental factors.

Table 3: Correlation Matrix for Amravati region

	AQI	CO2	NOX	RSPM	SO2	SPM
AQI	1.000	0.925	-0.130	0.996	-0.204	0.681
CO2	0.925	1.000	0.416	0.919	0.659	–
NOX	-0.130	0.416	1.000	-0.149	0.529	-0.168
RSPM	0.996	0.919	-0.149	1.000	-0.211	0.700
SO2	-0.204	0.659	0.529	-0.211	1.000	-0.137
SPM	0.681	–	-0.168	0.700	-0.137	1.000

The correlation matrix for the Amravati region reveals that AQI is strongly driven by particulate matter, exhibiting an almost perfect positive correlation with RSPM (0.996) and a very strong association with CO_2 (0.925). Moderate positive correlation between AQI and SPM (0.681) further reinforces the dominant role of particulate pollution in determining overall air quality. In contrast, AQI shows weak to negative correlations with gaseous pollutants such as NO_X (-0.130) and SO_2 (-0.204), indicating that their direct contribution to AQI variability is limited in this region. The strong inter-correlation between RSPM and CO_2 (0.919) suggests common emission sources, likely related to combustion and vehicular activities, whereas the weak or negative associations between particulate and gaseous pollutants point toward heterogeneous emission mechanisms and complex atmospheric interactions. Overall, this correlation structure highlights the suitability of particulate matter variables as primary predictors in air quality forecasting models for the Amravati region.

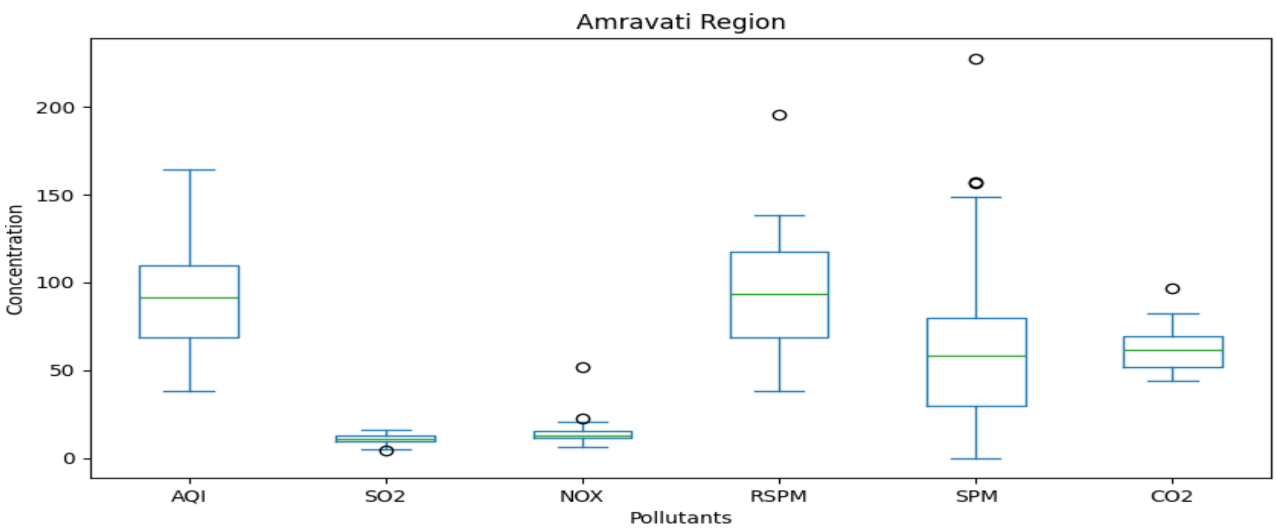

Figure 2: Box plot representation of pollutant components (AQI, CO_2, NO_X, RSPM, SO_2, and SPM) showing the statistical distribution of air quality indicators for the Amravati region.

The box plot for the Amravati region highlights pronounced variability in particulate matter, particularly RSPM and SPM, as evidenced by wider interquartile ranges and several high-end outliers, indicating episodic pollution events. In contrast, gaseous pollutants such as SO_2 and NO_X display tighter distributions and fewer extreme values, reflecting relatively stable emission patterns over the study period.

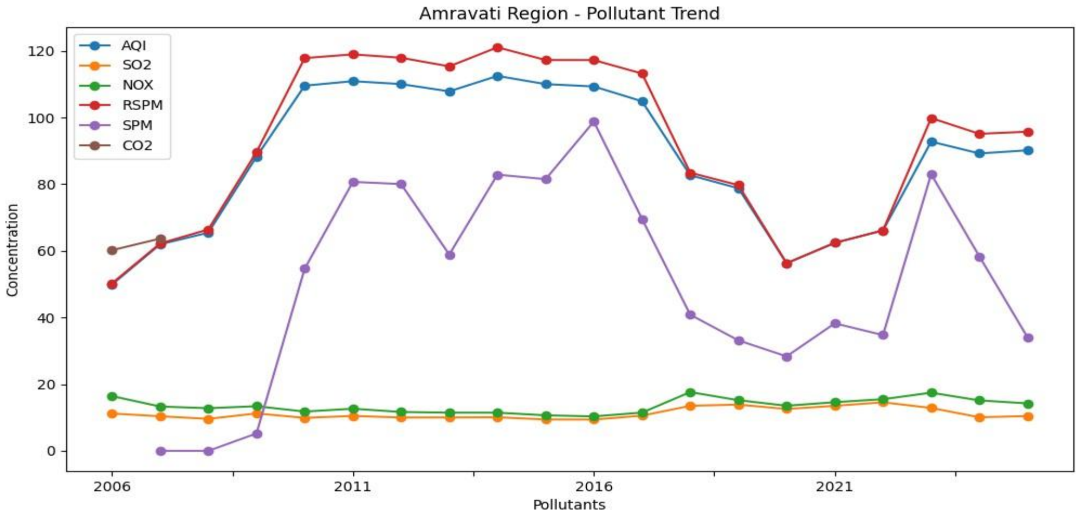

Figure 3: Line plot representation of pollutant components (AQI, CO_2, NO_X, RSPM, SO_2, and SPM) showing the temporal variation in air quality indicators for the Amravati region.

The temporal trends reveal that AQI closely follows the trajectory of particulate matter, particularly RSPM and SPM, confirming their dominant role in driving air quality variations in the Amravati region. Peaks observed during the mid-study period and again in recent years suggest episodic increases in particulate pollution, while gaseous pollutants (SO_2, NO_X, and CO_2) exhibit relatively stable trends with only minor interannual fluctuations.

6.1.2 Aurangabad Region

Table 4: Descriptive Statistics for Aurangabad (Month wise Aggregated Data)

	AQI	CO2	NOX	RSPM	SO2	SPM
count	241.0	23.0	241.0	241.0	241.0	214.0
mean	82.12	114.09	25.58	85.28	10.92	207.38
std	19.26	48.01	5.30	22.63	3.48	67.71
min	33.77	35.00	12.52	32.77	4.20	61.50
25%	68.38	69.50	22.24	69.97	8.17	165.73
50%	83.46	124.20	25.87	84.47	10.78	205.83
75%	95.49	153.08	29.25	100.79	13.52	250.24
max	134.42	200.00	41.64	161.24	22.10	423.63

The Aurangabad region demonstrates moderate to elevated air pollution levels with a mean AQI of 82.12, indicating persistent air quality concerns primarily driven by particulate matter. Both RSPM (mean: 85.28) and SPM (mean: 207.38) exhibit substantial variability, as reflected by their high standard deviations and wide ranges, suggesting pronounced temporal fluctuations and episodic pollution events. In contrast, gaseous pollutants such as SO_2 and NO_X show comparatively lower mean concentrations and narrower dispersion, indicating more stable emission patterns over time. The extreme upper values of SPM further emphasize the occurrence of intermittent high-intensity particulate pollution, likely associated with localized anthropogenic activities and seasonal factors.

Table 5: Correlation Matrix for Aurangabad region

	AQI	CO2	NOX	RSPM	SO2	SPM
AQI	1.000	0.130	0.567	0.993	0.263	0.770
CO2	0.130	1.000	0.461	0.117	0.527	0.986
NOX	0.567	0.461	1.000	0.524	0.604	0.323
RSPM	0.993	0.117	0.524	1.000	0.234	0.785
SO2	0.263	0.527	0.604	0.234	1.000	0.052
PM	0.770	0.986	0.323	0.785	0.052	1.000

The correlation matrix for the Aurangabad region indicates that AQI is predominantly governed by particulate matter, exhibiting an exceptionally strong positive correlation with RSPM (0.993) and a strong association with SPM (0.770). A moderate positive correlation between AQI and NO_X (0.567) suggests a noticeable influence of traffic-related gaseous emissions on air quality levels. In contrast, AQI shows only weak association with CO_2 (0.130) and SO_2 (0.263), implying a comparatively limited direct contribution of these pollutants to AQI variability. The very strong inter-correlation

between CO_2 and SPM (0.986) points toward common emission sources, likely linked to combustion and urban activities, while the moderate correlations among NO_X, SO_2, and particulate matter reflect a mixed and heterogeneous emission profile. Overall, this correlation structure underscores the central role of particulate pollutants, particularly RSPM, as key determinants of air quality dynamics in the Aurangabad region.

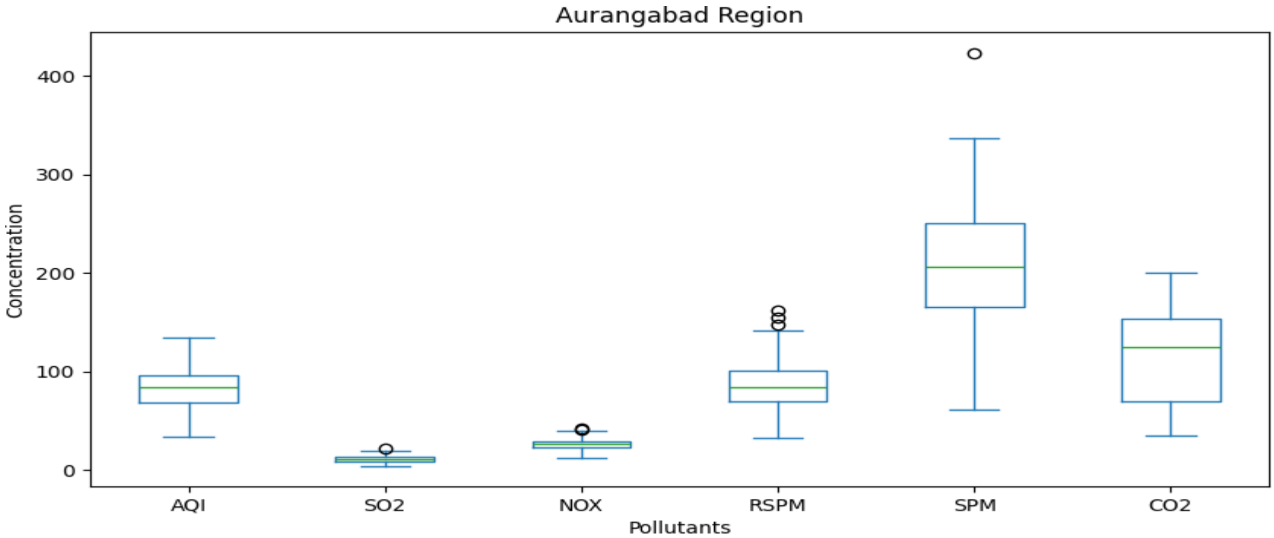

Figure 4: Box plot representation of pollutant components (AQI, CO_2, NO_X, RSPM, SO_2, and SPM) showing the statistical distribution of air quality indicators for the Aurangabad region.

The box plot for the Aurangabad region reveals substantial variability in particulate matter, particularly SPM and RSPM, characterized by wide interquartile ranges and prominent upper end outliers that indicate episodic high pollution events. In contrast, gaseous pollutants such as SO_2 and NO_X exhibit comparatively narrower distributions and fewer extreme values, suggesting more stable concentration levels over the study period.

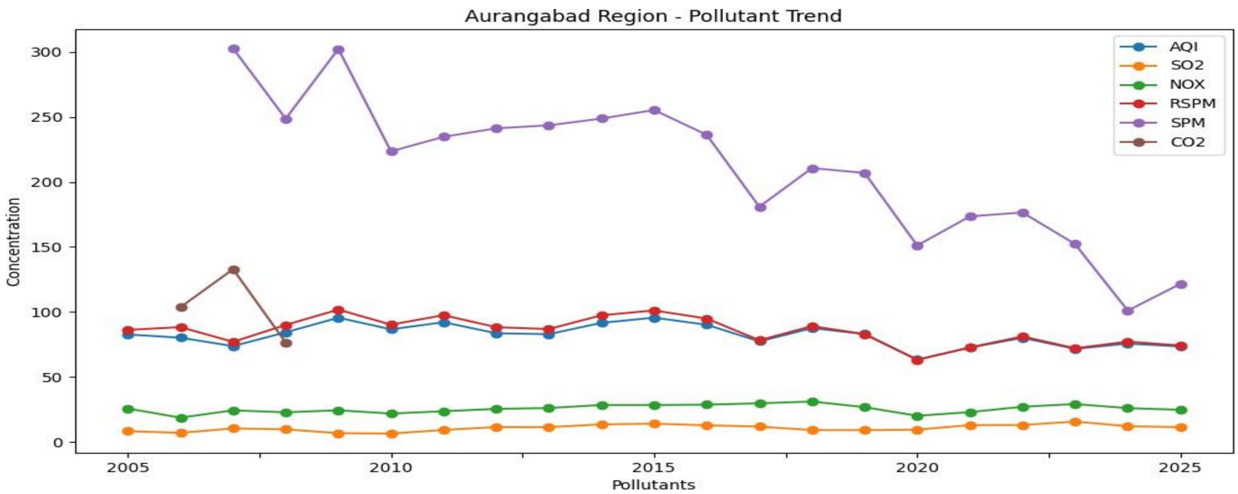

Figure 5: Line plot representation of pollutant components (AQI, CO_2, NO_X, RSPM, SO_2, and SPM) showing the temporal variation in air quality indicators for the Aurangabad region.

The AQI trend in Aurangabad exhibits a strong correspondence with particulate pollutants, particularly SPM and RSPM, indicating their primary influence on overall air quality. While gaseous pollutants (SO_2, NO_X, and CO_2) show comparatively stable and low-amplitude variations, episodic peaks in particulate matter during earlier years contribute to elevated AQI levels, followed by a gradual decline in recent years.

6.1.3 Konkan Region

Table 6: Descriptive Statistics for Konkan (Month wise Aggregated Data)

	AQI	CO2	NOX	RSPM	SO2	SPM
count	262.0	44.0	262.0	262.0	262.0	214.0

mean	103.59	74.50	51.86	108.27	20.63	122.15
std	27.89	28.23	16.16	37.92	5.41	55.57
min	46.14	34.06	9.33	24.00	4.00	20.98
25%	79.71	49.34	41.78	75.30	17.46	83.40
50%	102.90	72.26	50.68	108.47	20.22	111.66
75%	125.37	93.63	60.82	137.56	23.88	154.68
max	195.28	133.40	119.14	205.46	43.52	283.79

The Konkan region exhibits relatively degraded air quality conditions, with a mean AQI of 103.59, indicating frequent exceedance of moderate pollution thresholds. Elevated mean concentrations of RSPM (108.27) and NO_X (51.86), coupled with substantial variability, suggest that both particulate matter and traffic or industry-related gaseous emissions play a significant role in influencing regional air quality. The wide ranges observed for RSPM and SPM reflect pronounced temporal fluctuations and episodic pollution events, while comparatively higher SO_2 levels indicate the possible influence of industrial activities and coastal shipping. Overall, the results point to a complex pollution profile driven by both particulate and gaseous sources.

Table 7: Correlation Matrix for Konkan region

	AQI	CO2	NOX	RSPM	SO2	SPM
AQI	1.000	0.609	0.724	0.962	0.596	0.647
CO2	0.609	1.000	0.552	0.516	0.648	–
NOX	0.724	0.552	1.000	0.627	0.604	0.614
RSPM	0.962	0.516	0.627	1.000	0.569	0.629
SO2	0.596	0.648	0.604	0.569	1.000	0.698
SPM	0.647	–	0.614	0.629	0.698	1.000

The correlation matrix for the Konkan region reveals that AQI is strongly influenced by particulate matter, exhibiting a very strong positive correlation with RSPM (0.962) and moderate to strong associations with SPM (0.647) and CO_2 (0.609). Notably, AQI also shows a strong positive correlation with NO_X (0.724) and a moderate association with SO_2 (0.596), indicating that gaseous pollutants play a more prominent role in air quality variability in this coastal–industrial region compared to inland regions. The consistently moderate to strong inter-correlations among particulate and gaseous pollutants suggest shared emission sources, likely arising from vehicular traffic, industrial activity, and port-related operations. Overall, this correlation structure highlights a mixed pollution regime, where both particulate matter and gaseous emissions jointly contribute to air quality dynamics in the Konkan region.

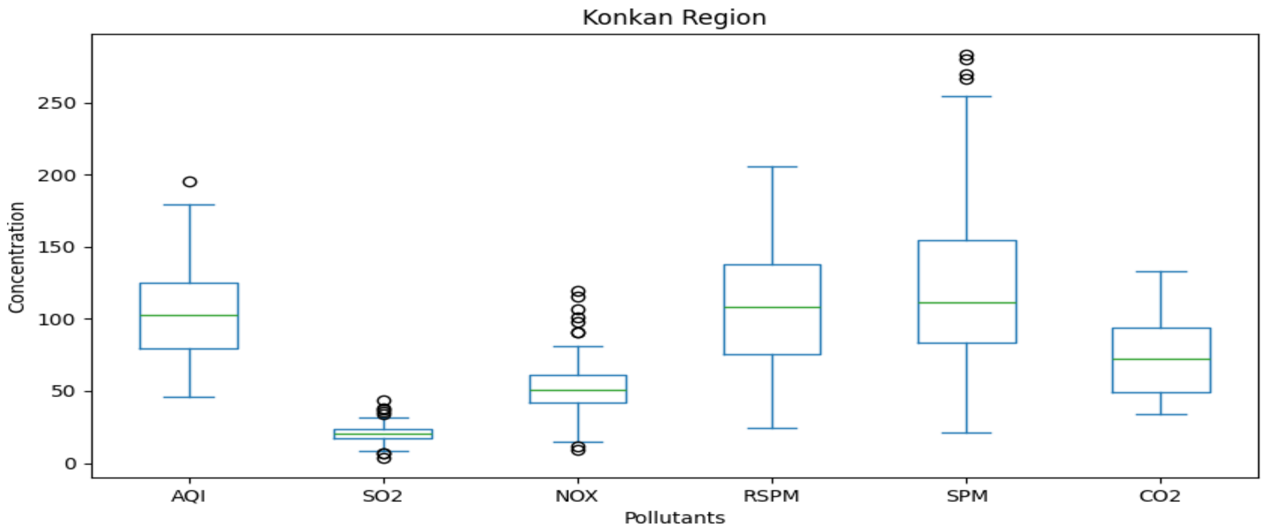

Figure 6: Box plot representation of pollutant components (AQI, CO_2, NO_X, RSPM, SO_2, and SPM) showing the statistical distribution of air quality indicators for the Konkan region.

The box plot for the Konkan region demonstrates considerable variability in particulate matter, particularly RSPM and SPM, as indicated by wide interquartile ranges and multiple upper-end outliers reflecting episodic pollution events. Compared to inland regions, gaseous pollutants such as NO_X and SO_2 show broader dispersion, suggesting a more pronounced influence of mixed urban, industrial, and coastal emission sources.

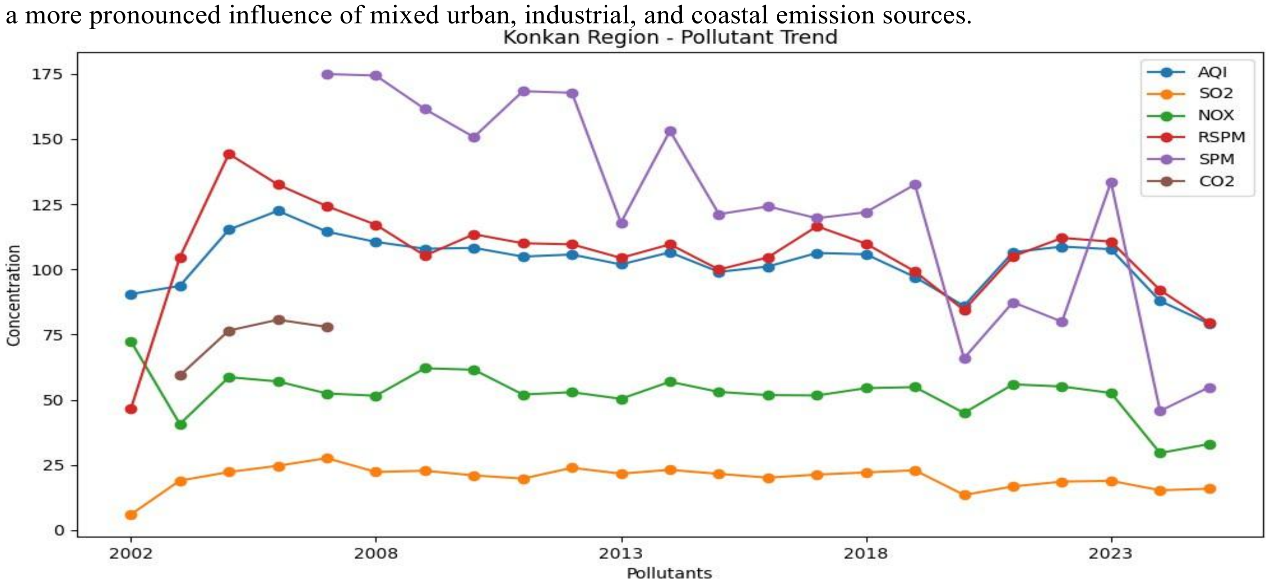

Figure 7: Line plot representation of pollutant components (AQI, CO_2, NO_X, RSPM, SO_2, and SPM) showing the temporal variation in air quality indicators for the Konkan region.

The Konkan region demonstrates moderate variability in AQI, largely driven by fluctuations in RSPM and SPM concentrations. Although particulate levels show intermittent spikes, a general stabilizing trend is observed over time. In contrast, gaseous pollutants remain relatively consistent across years, suggesting limited long-term escalation and a comparatively balanced pollution profile.

6.1.4 Nagpur Region

Table 8: Month wise descriptive Statistics for Nagpur region

	AQI	CO2	NOX	RSPM	SO2	SPM
count	260.0	44.0	260.0	260.0	260.0	219.0
mean	95.11	97.81	27.63	102.21	12.31	215.94
std	22.02	53.99	6.47	25.87	5.84	67.51
min	45.71	33.67	12.34	46.26	4.65	77.49
25%	78.72	62.38	22.74	83.67	7.70	169.64
50%	96.81	85.70	28.18	104.07	10.52	209.76
75%	111.42	113.96	32.96	121.29	15.50	248.27
max	143.74	299.00	40.88	163.33	33.94	488.81

The Nagpur region exhibits moderate to poor air quality conditions, with a mean AQI of 95.11, indicating sustained exposure to elevated pollution levels. Particulate matter, particularly RSPM (mean: 102.21) and SPM (mean: 215.94), emerges as the dominant contributor to air quality degradation, characterized by high variability and wide concentration ranges that reflect episodic pollution events. In contrast, gaseous pollutants such as SO_2 and NO_X display comparatively lower mean concentrations and more constrained dispersion, suggesting relatively stable emission patterns. The extreme upper values observed for SPM further highlight intermittent high-intensity particulate pollution, likely driven by urban activity, traffic density, and seasonal influences.

Table 9: Correlation Matrix for Nagpur region

	AQI	CO2	NOX	RSPM	SO2	SPM
AQI	1.000	0.721	0.480	0.992	0.242	0.525
CO2	0.721	1.000	0.533	0.800	0.579	0.653
NOX	0.480	0.533	1.000	0.484	0.498	0.321
RSPM	0.992	0.800	0.484	1.000	0.271	0.543
SO2	0.242	0.579	0.498	0.271	1.000	0.229
SPM	0.525	0.653	0.321	0.543	0.229	1.000

The correlation matrix for the Nagpur region demonstrates that AQI is overwhelmingly driven by particulate matter, exhibiting an exceptionally strong positive correlation with RSPM (0.992) and a moderate association with SPM (0.525). AQI also shows a strong positive correlation with CO_2 (0.721), indicating the influence of combustion-related and urban emission sources on overall air quality. In contrast, weaker to moderate correlations between AQI and gaseous pollutants such as NO_X (0.480) and SO_2 (0.242) suggest a comparatively limited direct contribution of these pollutants to AQI variability. The strong inter-correlation between CO_2 and RSPM (0.800) further implies shared emission pathways, while the moderate associations among gaseous pollutants reflect a heterogeneous emission profile. Overall, the correlation structure underscores the dominant role of particulate matter, particularly RSPM, in shaping air quality dynamics in the Nagpur region.

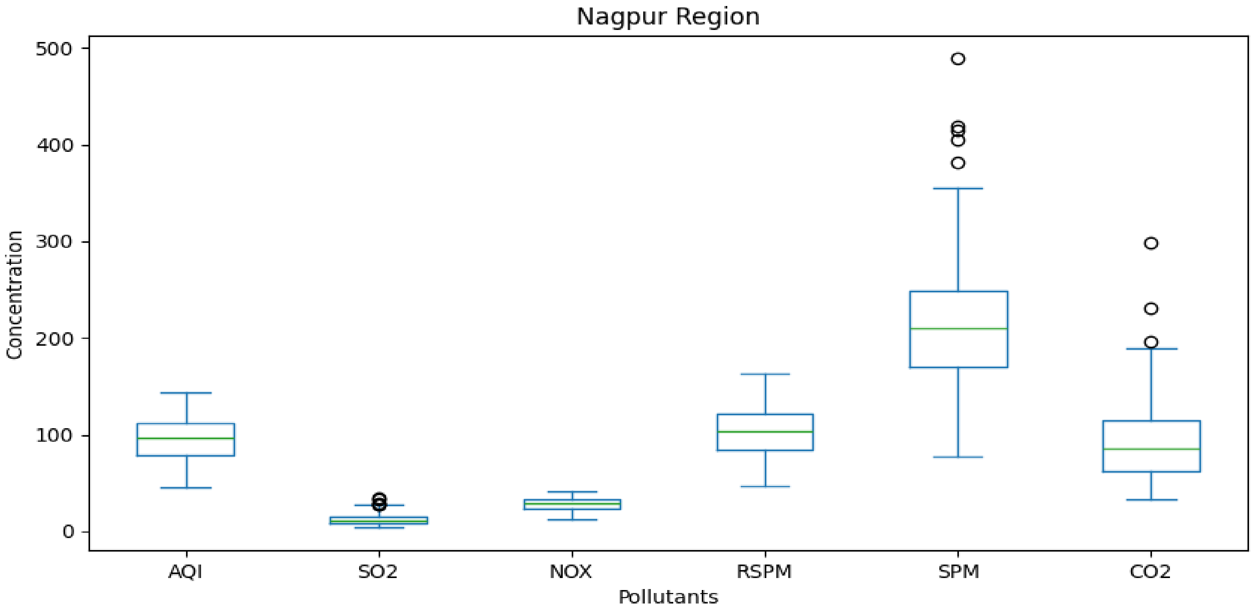

Figure 8: Box plot representation of pollutant components (AQI, CO_2, NO_X, RSPM, SO_2, and SPM) showing the statistical distribution of air quality indicators for the Nagpur region.

The box plot for the Nagpur region indicates pronounced variability in particulate matter, particularly SPM and RSPM, with wide interquartile ranges and several extreme upper-end outliers reflecting episodic high pollution events. In contrast, gaseous pollutants such as SO_2 and NO_X exhibit relatively compact distributions, suggesting more stable concentration levels compared to particulate pollutants.

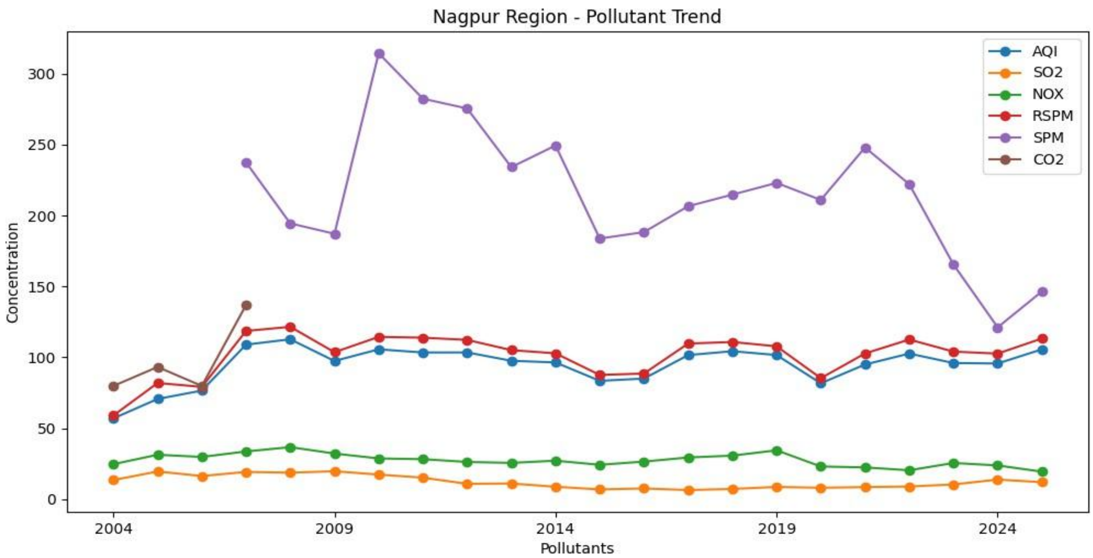

Figure 9: Line plot representation of pollutant components (AQI, CO_2, NO_X, RSPM, SO_2, and SPM) showing the temporal variation in air quality indicators for the Nagpur region.

Nagpur exhibits pronounced temporal variability in AQI, closely mirroring changes in particulate matter, especially SPM and RSPM, which display recurring peaks throughout the study period. Gaseous pollutants maintain lower magnitudes with gradual trends, reinforcing the dominant role of particulate emissions in shaping air quality dynamics in the region.

6.1.5 Nashik Region

Table 10: Descriptive Statistics for Nashik (Month wise Aggregated Data)

	AQI	CO2	NOX	RSPM	SO2	SPM
count	251.0	30.0	251.0	251.0	251.0	211.0
mean	78.99	69.36	27.09	82.15	16.37	135.38
std	21.41	25.91	6.75	24.27	7.84	58.98
min	27.15	33.71	9.20	25.80	4.00	26.32
25%	63.54	48.68	22.07	64.95	8.72	87.07
50%	78.54	59.90	27.79	80.81	17.41	139.45
75%	94.13	91.88	32.15	97.83	22.95	182.94
max	132.47	120.50	48.81	148.71	32.91	253.90

The Nashik region exhibits generally moderate air quality conditions, with a mean AQI of 78.99, indicating comparatively lower pollution levels relative to highly urbanized regions. Particulate matter, particularly RSPM (mean: 82.15) and SPM (mean: 135.38), remains the primary contributor to air quality variability, as reflected by their relatively high standard deviations and broad concentration ranges. Gaseous pollutants such as SO_2 and NO_X show moderate mean concentrations but greater dispersion than observed in some other regions, suggesting localized emission influences. Overall, the observed patterns indicate episodic particulate pollution superimposed on relatively stable background air quality conditions.

Table 11: Correlation Matrix for Nashik region

	AQI	CO2	NOX	RSPM	SO2	SPM
AQI	1.000	0.343	0.604	0.993	0.514	0.758
CO2	0.343	1.000	0.012	0.447	-0.258	–
NOX	0.604	0.012	1.000	0.614	0.747	0.807
RSPM	0.993	0.447	0.614	1.000	0.533	0.755
SO2	0.514	-0.258	0.747	0.533	1.000	0.719
SPM	0.758	–	0.807	0.755	0.719	1.000

The correlation matrix for the Nashik region indicates that AQI is predominantly driven by particulate matter, exhibiting an almost perfect positive correlation with RSPM (0.993) and a strong association with SPM (0.758). AQI also shows a moderate positive correlation with NO_X (0.604) and SO_2 (0.514), suggesting a noticeable contribution from traffic- and industry-related gaseous emissions. In contrast, AQI demonstrates only a weak relationship with CO_2 (0.343), while the near-zero and negative correlations between CO_2 and other pollutants point toward distinct emission sources. Strong inter-correlations among NO_X, SO_2, and SPM further indicate shared emission pathways and complex atmospheric interactions. Overall, this correlation structure reflects a mixed pollution profile, with particulate matter remaining the dominant driver of air quality variability in the Nashik region.

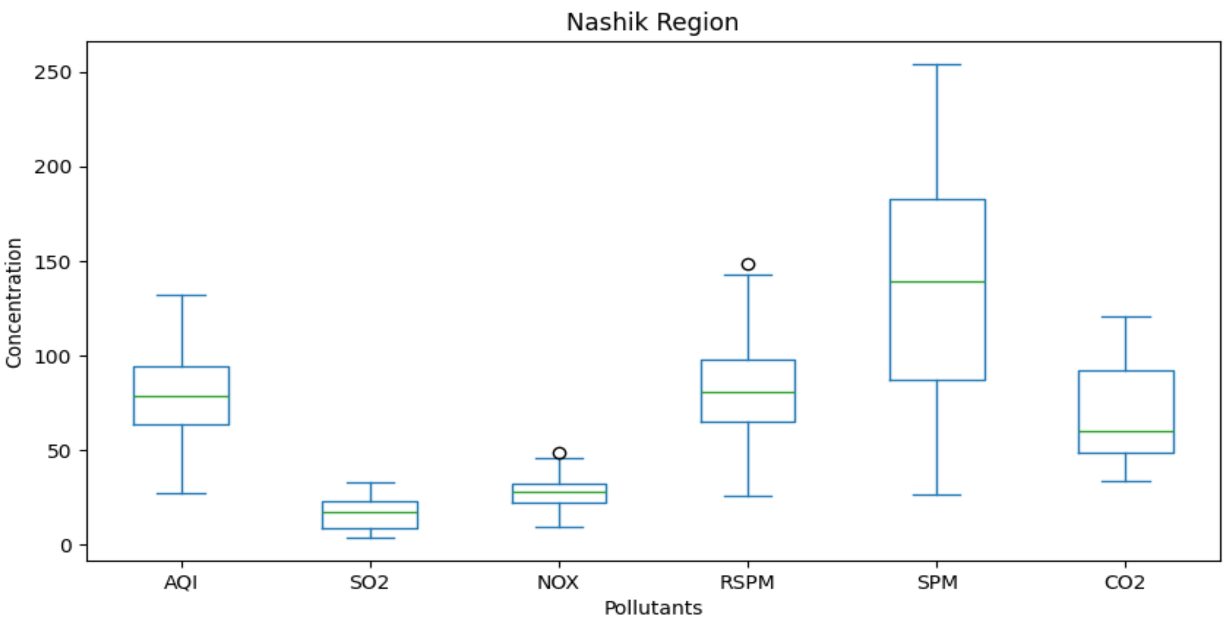

Figure 10: Box plot representation of pollutant components (AQI, CO_2, NO_X, RSPM, SO_2, and SPM)

showing the statistical distribution of air quality indicators for the Nashik region.

The Nashik region box plot shows moderate variability in particulate matter, with SPM exhibiting a wider spread and occasional high-end outliers, indicating intermittent pollution episodes. Gaseous pollutants, especially SO_2 and NO_X, display narrower interquartile ranges and fewer extreme values, reflecting relatively stable emission patterns over the study period.

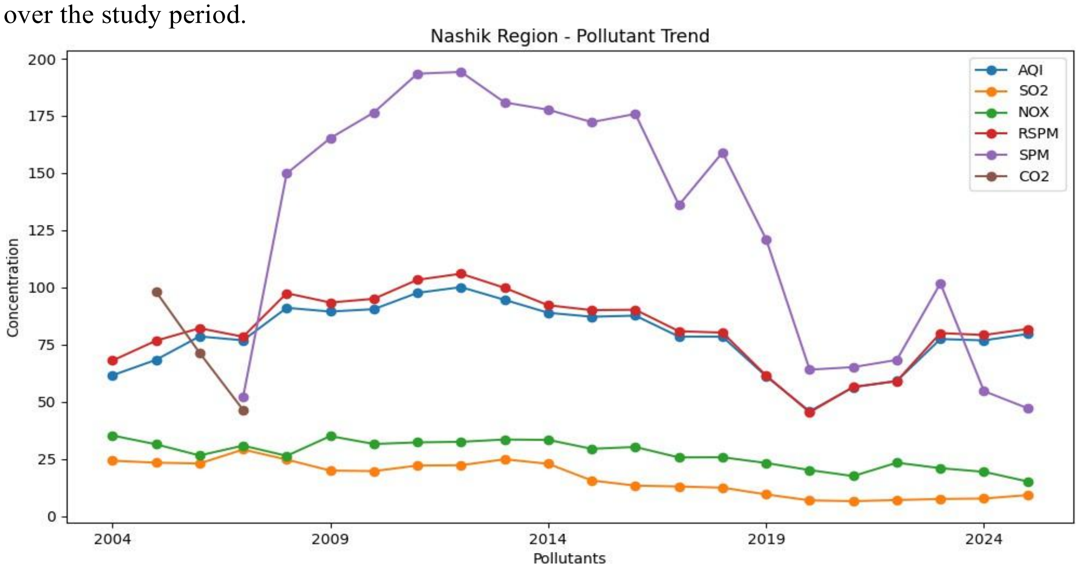

Figure 11: Line plot representation of pollutant components (AQI, CO_2, NO_X, RSPM, SO_2, and SPM) showing the temporal variation in air quality indicators for the Nashik region.

The AQI trend in Nashik closely aligns with variations in particulate matter, particularly RSPM and SPM, indicating their dominant contribution to air quality degradation. While short- term fluctuations in particulate concentrations are evident, gaseous pollutants (SO_2, NO_X, and CO_2) exhibit relatively stable temporal patterns, suggesting a secondary role in influencing overall AQI dynamics.

6.1.6 Pune Region

Table 12: Descriptive Statistics for Pune (Month wise Aggregated Data)

	AQI	CO2	NOX	RSPM	SO2	SPM
count	259.000	29.000	259.000	259.000	259.000	223.000
mean	89.930	108.507	39.642	92.441	16.960	207.406
std	25.202	38.640	12.808	31.537	4.147	79.548
min	30.069	24.500	14.249	26.552	6.059	39.177
25%	66.498	87.750	31.199	62.147	14.145	149.844
50%	92.765	114.000	37.989	94.925	17.416	205.821
75%	110.049	136.750	47.252	118.677	19.375	267.343
max	135.506	169.200	80.855	154.140	28.723	409.236

The Pune region exhibits moderate air quality conditions, with a mean AQI of 89.93, reflecting sustained exposure to urban air pollution. Particulate matter, particularly RSPM (mean: 92.44) and SPM (mean: 207.41), plays a dominant role in shaping air quality, as evidenced by high variability and wide concentration ranges indicative of episodic pollution events. Elevated NO_X levels further suggest the influence of traffic-related emissions typical of rapidly urbanizing regions. In contrast, SO_2 shows relatively stable concentrations, pointing toward comparatively regulated industrial emission sources.

Table 13: Correlation Matrix for Pune region

	AQI	CO2	NOX	RSPM	SO2	SPM
AQI	1.000	0.895	0.687	0.981	0.573	0.782
CO2	0.895	1.000	0.793	0.900	0.281	0.923
NOX	0.687	0.793	1.000	0.595	0.582	0.711
RSPM	0.981	0.900	0.595	1.000	0.543	0.780
SO2	0.573	0.281	0.582	0.543	1.000	0.531
SPM	0.782	0.923	0.711	0.780	0.531	1.000

The correlation matrix for the Pune region demonstrates that AQI is strongly governed by particulate matter, exhibiting an exceptionally high positive correlation with RSPM (0.981) and a strong association with SPM (0.782). AQI also shows a very strong positive correlation with CO_2 (0.895) and a moderate to strong association with NO_X (0.687), reflecting the significant influence of combustion related and traffic-driven emissions typical of a highly urbanized environment. The strong inter-correlations among CO_2, RSPM, and SPM suggest shared emission sources, likely linked to vehicular activity and urban energy consumption. In contrast, SO_2 exhibits comparatively weaker associations, indicating a more limited direct role in AQI variability. Overall, this correlation structure highlights a complex urban pollution profile in which particulate matter remains the dominant driver, reinforced by substantial gaseous emission contributions.

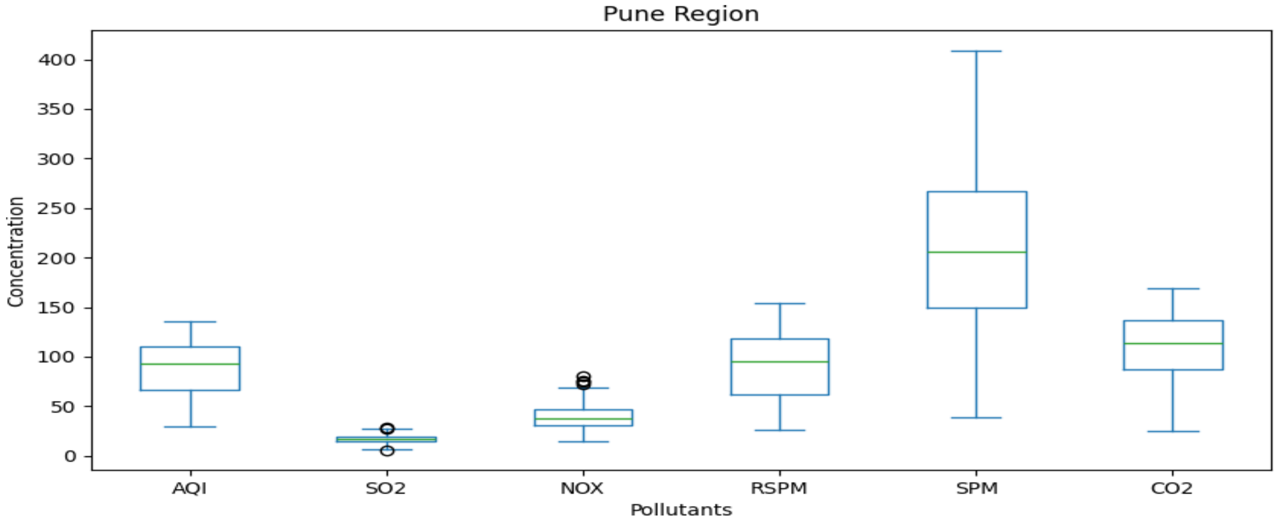

Figure 12: Box plot representation of pollutant components (AQI, CO_2, NO_X, RSPM, SO_2, and SPM) showing the statistical distribution of air quality indicators for the Pune region.

The box plot for the Pune region highlights substantial variability in particulate matter, particularly SPM and RSPM, characterized by broad interquartile ranges and pronounced upper- end outliers indicative of episodic pollution spikes. Compared to other pollutants, gaseous emissions such as SO_2 show tighter distributions, while NO_X exhibits moderate dispersion consistent with traffic-driven variability in an urban environment.

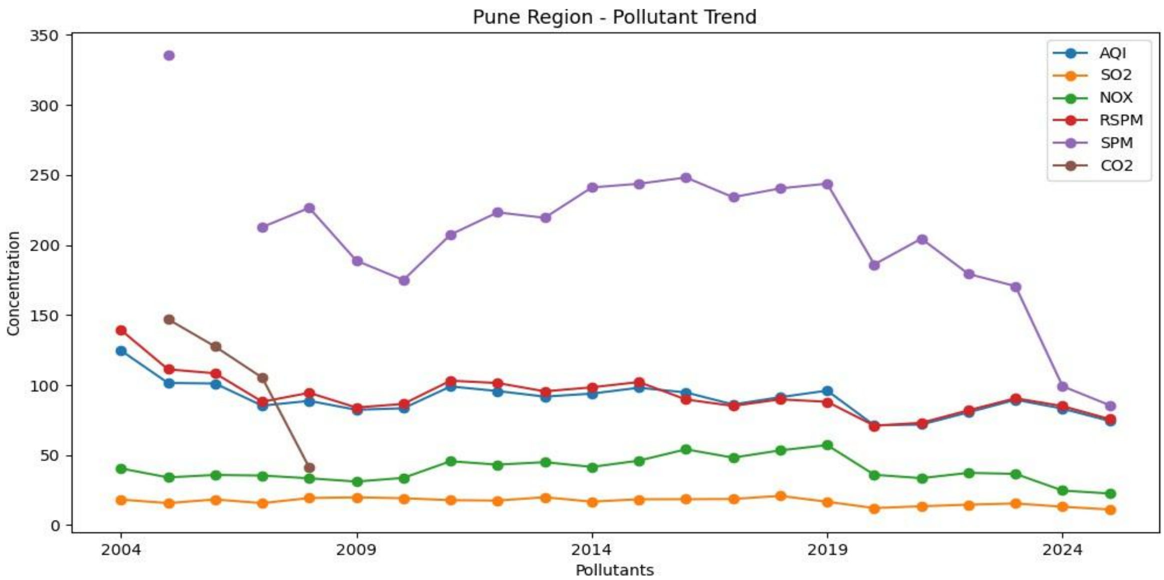

Figure 13: Line plot representation of pollutant components (AQI, CO_2, NO_X, RSPM, SO_2, and SPM) showing the temporal variation in air quality indicators for the Pune region.

Pune displays a strong coupling between AQI and particulate pollutants, with RSPM showing the highest correspondence, followed by SPM. The correlation structure indicates that particulate emissions are the primary drivers of air quality variability, whereas gaseous pollutants remain comparatively steady over time, contributing marginally to interannual AQI fluctuations.

6.2 Statistical Tests for Normality and Stationarity

To evaluate the statistical characteristics of the pollutant time series, we conducted Shapiro-Wilk Test to assess normality and KPSS tests to assess stationarity for each pollutant across all regions.

Shapiro-Wilk Test: The Shapiro-Wilk test is used to assess whether a dataset follows a normal distribution

Hypothesis:

H_0: The data are normally distributed.

H_1: The are not normally distributed.

Test statistics:

$$W=\frac{\sum_{i=1}^{n} aiXi^2}{\sum_{i=1}^{n}(Xi-\overline{X})^2}$$

Where

Xi - The ordered sample values (ordered statistics)

$\overline{X}$ - Sample mean

ai – Constant derived from the expected values and covariance matrix of ordered statistics of normal distribution

n – sample size

Interpretation:

A small value of W indicates deviation from normality. A P-value less than 0.05 leads to a rejection of the null hypothesis that is data are not normally distributed.

KPSS Test (Kwiatkowski–Phillips–Schmidt–Shin): The KPSS test is used to determine whether a time series is stationary around a deterministic trend.

Hypothesis:

H_0: The time series is stationary (level or trend stationary).

H_1: The time series is non-stationary.

Model Representation:

$y_t = \delta t + r_t + \epsilon_t$

where:

δt represents a deterministic trend,

r_t is a random walk: $r_t = r_{t-1} + u_t$,

ϵ_t is a stationary error term.

Test Statistics:

$$KPSS = \frac{1}{T^2}\sum_{t=1}^{T}\frac{S_t^2}{\hat{\sigma}^2}$$

Where,

T – Sample size

S_t - The partial sum of residual

$\hat{\sigma}^2$ – The Long run variance estimate

Interpretation:

A large KPSS statistics (with p-value < 0.05) leads to a rejection of the null hypothesis, indicating that the time series is non-stationar Test Statistic:

Table 14: Shapiro-Wilk and KPSS Test Results for Pollutants Across Regions

Region	Pollutant	Shapiro Stat	Shapiro p	Shapiro Result	KPSS Stat	KPSS p	KPSS Result
Amravati	SO_2	0.9781	0.0013	REJECT	0.8394	0.01	REJECT
Amravati	NO_X	0.6668	0.0000	REJECT	0.9816	0.01	REJECT
Amravati	RSPM	0.9647	0.0000	REJECT	0.4030	0.0759	ACCEPT
Amravati	AQI	0.9664	0.0000	REJECT	0.4201	0.0685	ACCEPT
Aurangabad	SO_2	0.9831	0.0057	REJECT	0.8122	0.01	REJECT
Aurangabad	NO_X	0.9897	0.0857	ACCEPT	0.5614	0.0278	REJECT
Aurangabad	RSPM	0.9904	0.1129	ACCEPT	0.8659	0.01	REJECT
Aurangabad	AQI	0.9948	0.5874	ACCEPT	0.6351	0.0194	REJECT
Konkan	SO_2	0.9816	0.0018	REJECT	1.0307	0.01	REJECT
Konkan	NO_X	0.9653	0.0000	REJECT	0.5868	0.0238	REJECT
Konkan	RSPM	0.9819	0.0020	REJECT	0.6585	0.0173	REJECT
Konkan	AQI	0.9858	0.0108	REJECT	0.5411	0.0324	REJECT
Nagpur	SO_2	0.8956	0.0000	REJECT	1.4199	0.01	REJECT
Nagpur	NO_X	0.9789	0.0007	REJECT	0.9984	0.01	REJECT

Nagpur	RSPM	0.9868	0.0173	REJECT	0.2423	0.10	ACCEPT
Nagpur	AQI	0.9864	0.0146	REJECT	0.2900	0.10	ACCEPT
Nashik	SO_2	0.9265	0.0000	REJECT	2.0724	0.01	REJECT
Nashik	NO_X	0.9894	0.0634	ACCEPT	1.8266	0.01	REJECT
Nashik	RSPM	0.9922	0.2054	ACCEPT	1.0180	0.01	REJECT
Nashik	AQI	0.9940	0.4241	ACCEPT	0.8980	0.01	REJECT
Pune	SO_2	0.9917	0.1510	ACCEPT	1.0555	0.01	REJECT
Pune	NO_X	0.9709	0.0000	REJECT	0.4199	0.0686	ACCEPT
Pune	RSPM	0.9610	0.0000	REJECT	1.0571	0.01	REJECT
Pune	AQI	0.9649	0.0000	REJECT	0.8006	0.01	REJECT

6.3 Comparative Forecasting Performance: Prophet, ARIMA, and ETS

To ensure robust and comprehensive forecasting of air quality indicators, three complementary time series models were employed: Prophet, ARIMA, and Exponential Smoothing (ETS). These models represent distinct methodological paradigms-additive decomposition, stochastic autoregressive modeling, and state-space exponential smoothing-thereby enabling a systematic comparative evaluation of predictive performance.

Prophet Model: Prophet, developed by Facebook (now Meta), is an additive regression- based model designed to handle time series exhibiting multiple seasonality's, trend shifts, and irregular observations. It decomposes a series into:

Trend: Long-term progression modeled using piecewise linear or logistic growth.

Seasonality: Periodic fluctuations (daily, weekly, yearly) captured via Fourier series.

Holiday/Intervention Effects: Deterministic adjustments for abrupt deviations.

Its robustness to missing values, outliers, and structural changes makes it particularly suit- able for environmental datasets characterized by variability and external influences.

ARIMA Model: The Autoregressive Integrated Moving Average (ARIMA) model is a classical stochastic approach that captures linear temporal dependencies through three components:

Auto Regression (AR): Dependence on past values.

Integration (I): Differencing to achieve stationarity.

Moving Average (MA): Dependence on past forecast errors.

ARIMA is effective for modeling short-term autocorrelation structures and is particularly suitable for stationary or differenced stationary pollutant series. However, it may require careful parameter selection and stationarity diagnostics.

ETS Model: The Exponential Smoothing (ETS) framework is a state-space approach that model's level, trend, and seasonality components using smoothing equations. Depending on the data characteristics, ETS can incorporate additive or multiplicative error, trend, and seasonal structures. It is especially effective for series with evolving level and seasonal patterns, offering adaptive forecasting without explicit stationarity requirements.

Comparative Perspective: While Prophet emphasizes interpretability and multi-seasonal decomposition, ARIMA focuses on stochastic autocorrelation structures, and ETS provides a flexible smoothing-based state-space formulation. Comparing these models enables assessment of structural versus stochastic forecasting capabilities, thereby identifying the most suitable framework for region-wise pollutant prediction and policy-oriented air quality management.

Amravati Region:

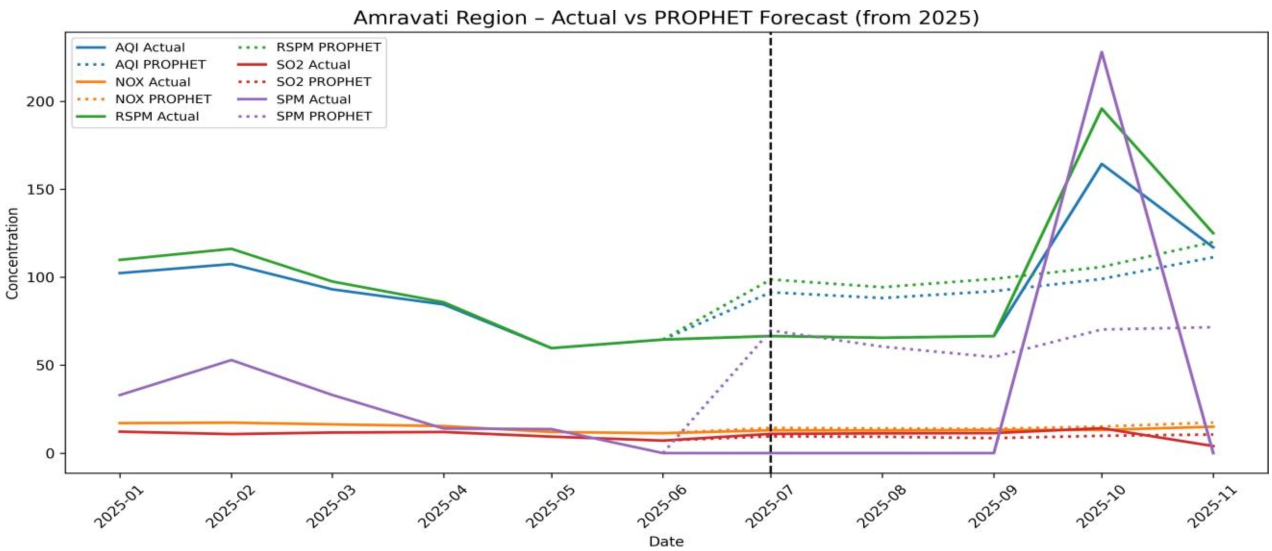

Figure 14: Prophet-based forecast for air quality components in the Amravati region

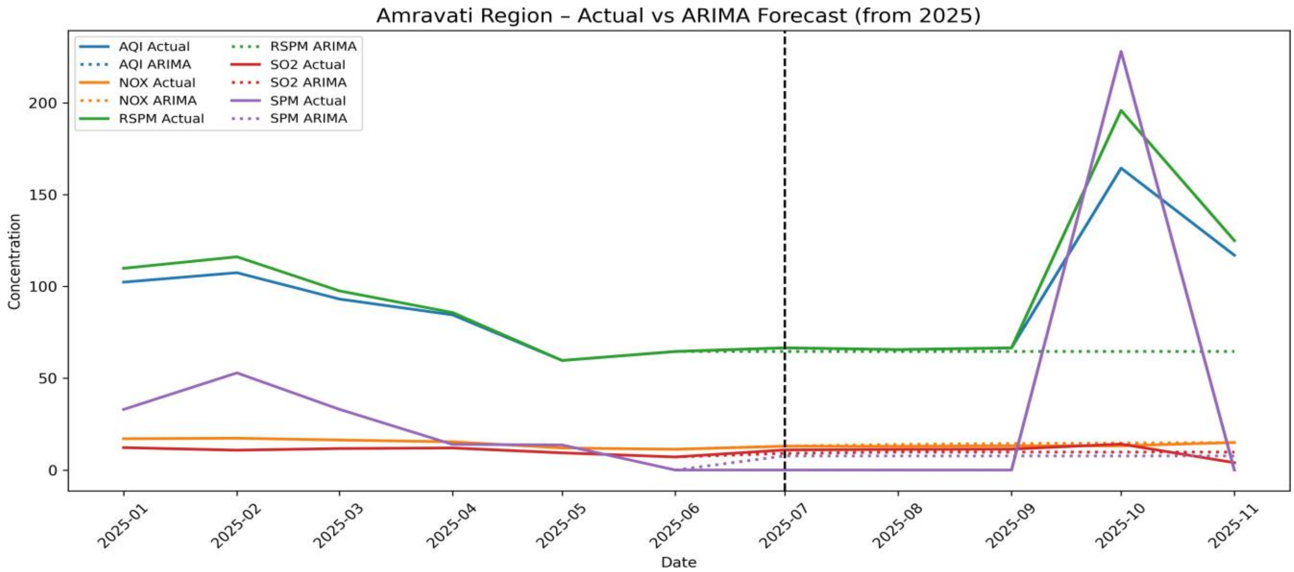

Figure 15: ARIMA-based forecast for air quality components in the Amravati region

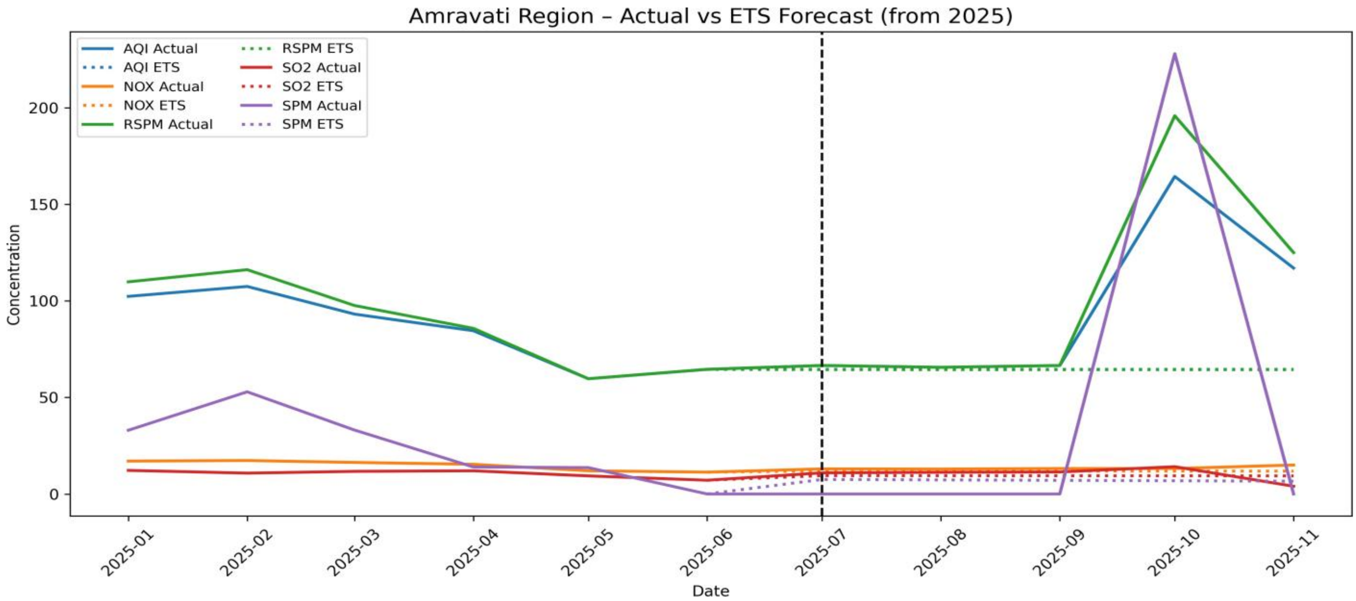

Figure 16: ETS-based forecast for air quality components in the Amravati region

Table 15: Forecasting performance comparison for Amravati region using PROPHET, ARIMA, and ETS models (MAE and MAPE).

Metric	AQI	NOX	RSPM	SO2	SPM
MAE (PROPHET)	25.79	2.51	33.08	2.28	66.79
MAPE (%) (PROPHET)	20.06	18.64	21.90	46.88	1.33×10^{19}
MAE (ARIMA)	25.99	1.53	33.97	3.25	57.23
MAPE (%) (ARIMA)	27.03	11.65	31.53	46.11	7.44×10^{18}
MAE (ETS)	31.49	1.53	39.37	3.54	50.83
MAPE (%) (ETS)	22.65	11.70	24.61	44.59	3.14×10^{1}

Forecast evaluation for the Amravati region demonstrates moderate variation in predictive performance across the three models. For AQI, PROPHET achieves an MAE of 25.79 with a MAPE of 20.06%, while ARIMA records a comparable MAE (25.99) but a slightly higher MAPE (27.03%). ETS shows a relatively larger absolute deviation (MAE 31.49) though with moderate percentage error (MAPE 22.65%), indicating marginally lower overall accuracy.

In the case of NO_x, all models exhibit low absolute errors. ARIMA and ETS yield the smallest MAE (1.53) with MAPE values near 11.65%, whereas PROPHET reports a higher relative deviation (18.64%).

RSPM forecasts indicate moderate error magnitudes, with PROPHET and ARIMA producing MAE values

around 33.08 and MAPE between 21.90% and 31.53%. ETS demonstrates comparatively higher absolute error (MAE 39.37), suggesting reduced precision for particulate matter prediction.
SO_2 presents relatively elevated percentage errors across models, particularly under PROPHET (MAPE 46.88%) and ARIMA (46.11%), implying greater forecasting complexity despite moderate MAE values.
For SPM, extremely large MAPE values (on the order of 10^{18}–10^{19}) are observed across all models, indicating scale distortions or the influence of near-zero actual observations. In contrast, MAE values (50.83–66.79) remain within a moderate range, suggesting that MAPE may not serve as a reliable performance metric for SPM in this setting.

Aurangabad Region:

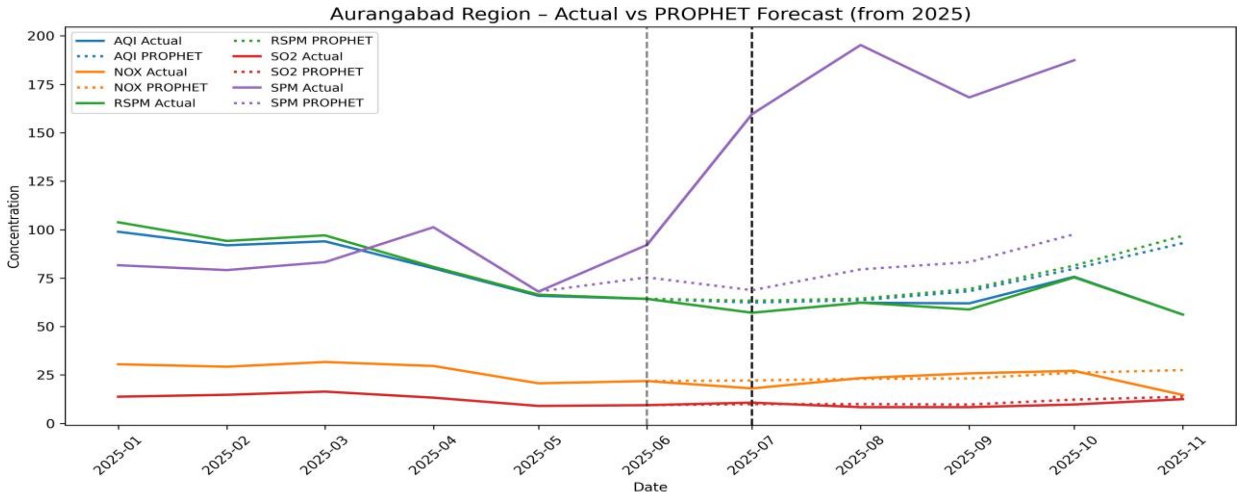

Figure 17: Prophet-based forecast for air quality components in the Aurangabad region

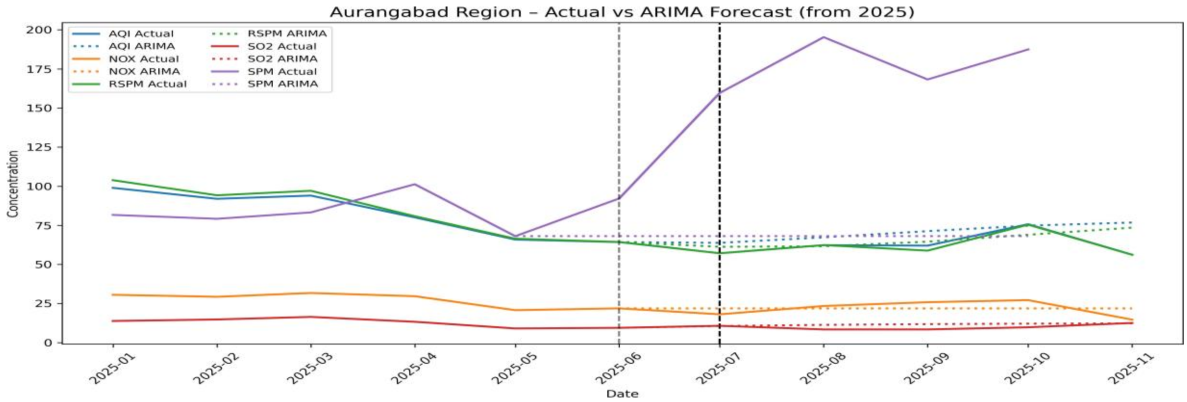

Figure 18: ARIMA-based forecast for air quality components in the Aurangabad region

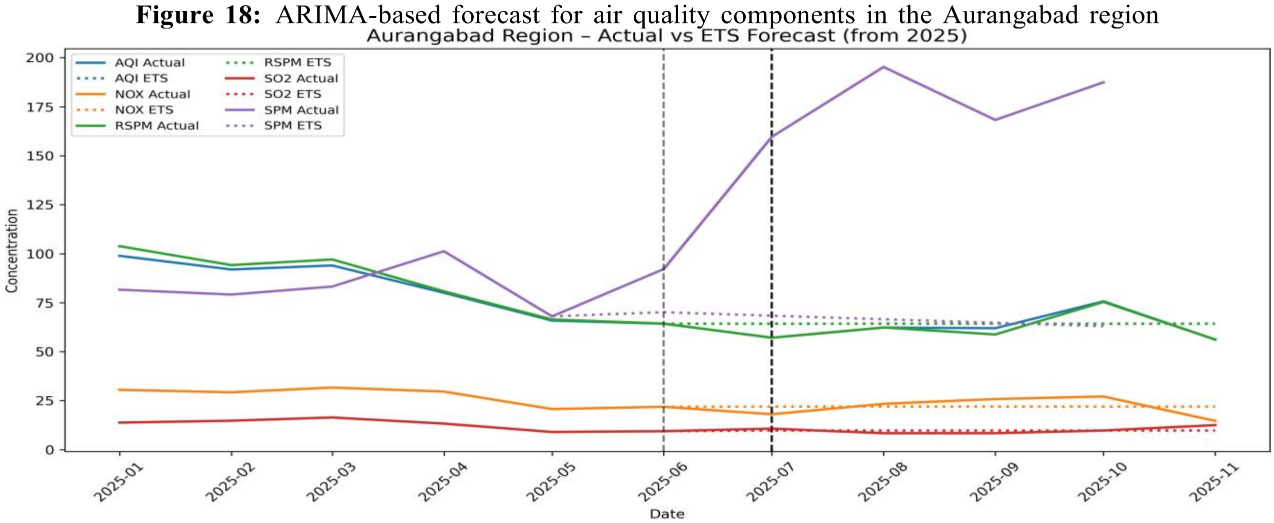

Figure 19: ETS-based forecast for air quality components in the Aurangabad region

Table 16: Forecasting performance comparison for Aurangabad region using PROPHET, ARIMA, and ETS models (MAE and MAPE).

Metric	AQI	NOX	RSPM	SO2	SPM

MAE (PROPHET)	11.34	4.58	12.47	2.21	86.76
MAPE (%) (PROPHET)	18.88	26.25	20.86	23.94	50.71
MAE (ARIMA)	10.33	4.36	13.59	1.75	92.53
MAPE (%) (ARIMA)	17.49	22.20	23.06	19.66	54.41
MAE (ETS)	6.18	4.34	6.79	1.32	90.93
MAPE (%) (ETS)	9.76	22.24	10.88	12.85	53.34

Forecast evaluation for the Aurangabad region reveals comparatively stronger predictive performance relative to Amravati, particularly for AQI and RSPM. For AQI, ETS achieves the lowest MAE (6.18) and MAPE (9.76%), indicating superior forecasting accuracy, followed by ARIMA (MAE 10.33; MAPE 17.49%) and PROPHET (MAE 11.34; MAPE 18.88%).

In the case of NO_x, error magnitudes remain moderate across models. ETS and ARIMA exhibit comparable MAE values (4.34–4.36) with MAPE around 22%, whereas PROPHET records slightly higher relative deviation (26.25%).

RSPM forecasts demonstrate clear improvement under ETS, which achieves substantially lower MAE (6.79) and MAPE (10.88%) compared to PROPHET and ARIMA, both of which report MAE above 12 and MAPE exceeding 20%. This suggests enhanced suitability of exponential smoothing techniques for particulate matter dynamics in this region.

SO_2 predictions show consistent performance across models, with ETS yielding the smallest errors (MAE 1.32; MAPE 12.85%), followed by ARIMA and PROPHET. Percentage deviations remain within moderate bounds (12–24%), indicating stable forecasting behavior.

For SPM, absolute errors remain relatively high (MAE 86.76–92.53), and MAPE values range between 50% and 54%, reflecting substantial variability in coarse particulate matter forecasting. Despite ETS improving AQI and RSPM prediction accuracy, SPM remains comparatively difficult to model in the Aurangabad region.

Konkan Region:

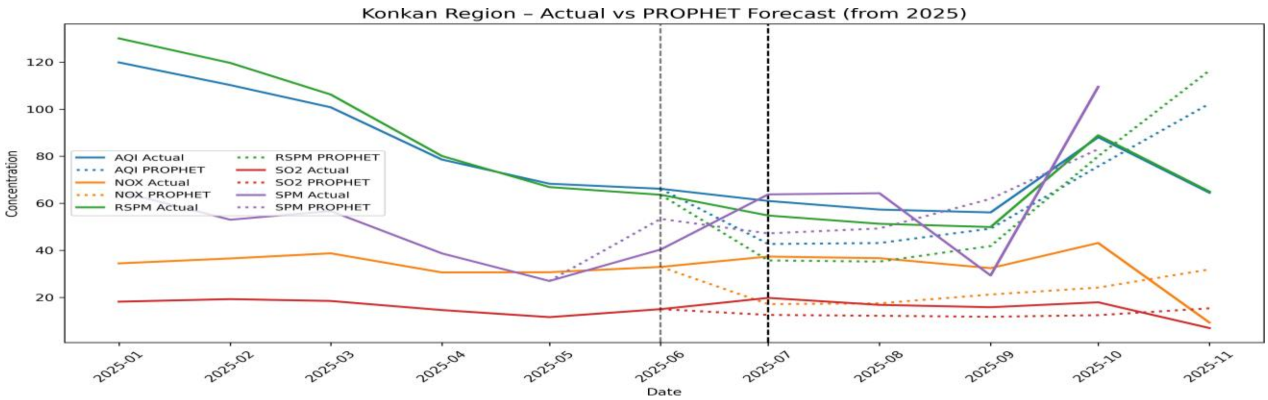

Figure 20: Prophet-based forecast for air quality components in the Kokan region

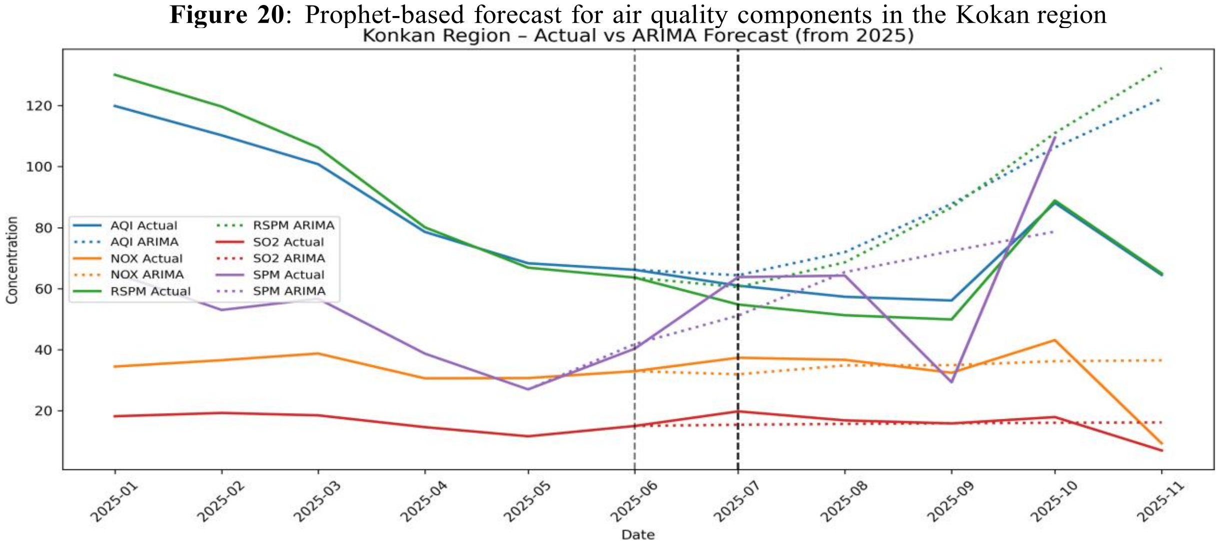

Figure 21: ARIMA-based forecast for air quality components in the Konkan region

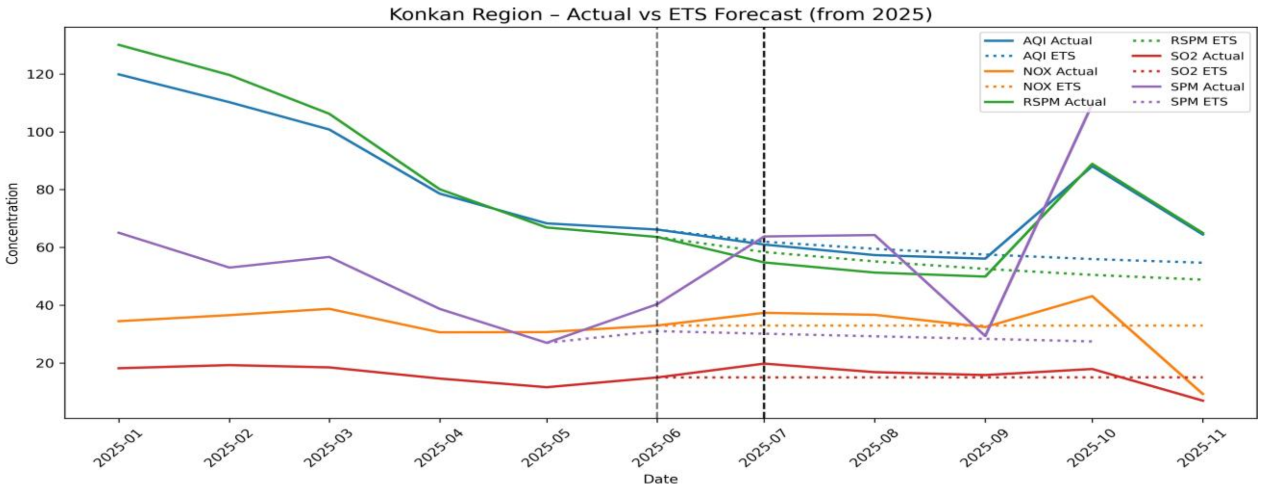

Figure 22: ETS-based forecast for air quality components in the Konkan region

Table 17: Forecasting performance comparison for Konkan region using PROPHET, ARIMA, and ETS models (MAE and MAPE).

Metric	**AQI**	**NOX**	**RSPM**	**SO2**	**SPM**
MAE (PROPHET)	14.65	10.67	13.08	5.19	25.08
MAPE (%) (PROPHET)	23.42	91.64	21.02	47.38	35.82
MAE (ARIMA)	30.97	20.92	37.24	3.68	31.19
MAPE (%) (ARIMA)	48.69	139.45	62.04	34.23	52.04
MAE (ETS)	9.53	8.49	12.30	3.83	34.01
MAPE (%) (ETS)	13.86	60.09	19.62	34.66	45.46

Forecast evaluation for the Konkan region reveals substantial variability in model performance across pollutants. For AQI, ETS demonstrates superior predictive accuracy with the lowest MAE (9.53) and MAPE (13.86%), followed by PROPHET (MAE 14.65; MAPE 23.42%). ARIMA records considerably higher errors (MAE 30.97; MAPE 48.69%), indicating comparatively weaker performance for overall air quality forecasting.

In the case of NO_x, all models exhibit elevated percentage errors, particularly ARIMA (MAPE 139.45%) and PROPHET (91.64%), suggesting significant relative instability. ETS reduces both MAE (8.49) and MAPE (60.09%), yet percentage deviations remain high, reflecting potential volatility or low observed concentration levels in this region.

RSPM forecasts again favor ETS, which achieves the lowest MAE (12.30) and MAPE (19.62%), whereas ARIMA produces substantially larger errors (MAE 37.24; MAPE 62.04%). PROPHET provides intermediate performance, indicating moderate suitability for particulate matter prediction.

SO_2 forecasting displays comparatively stable behavior. ARIMA yields the smallest MAE (3.68) with moderate MAPE (34.23%), closely followed by ETS. PROPHET shows slightly higher absolute and percentage deviations, though differences remain limited relative to other pollutants.

For SPM, absolute errors range between 25.08 and 34.01, with PROPHET achieving the lowest MAE. However, percentage errors remain moderate to high across models (35– 52%), suggesting persistent variability in coarse particulate matter prediction. Overall, ETS emerges as the most consistent model for AQI and RSPM in the Konkan region, while ARIMA performs relatively better for SO_2.

Nagpur Region:

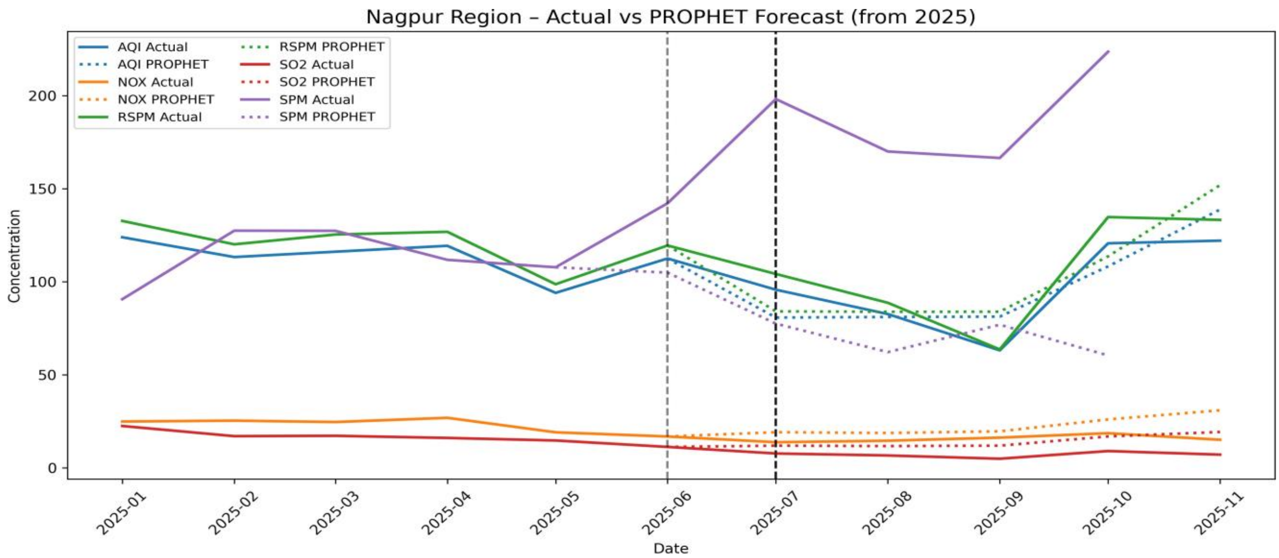

Figure 23: Prophet-based forecast for air quality components in the Nagpur region

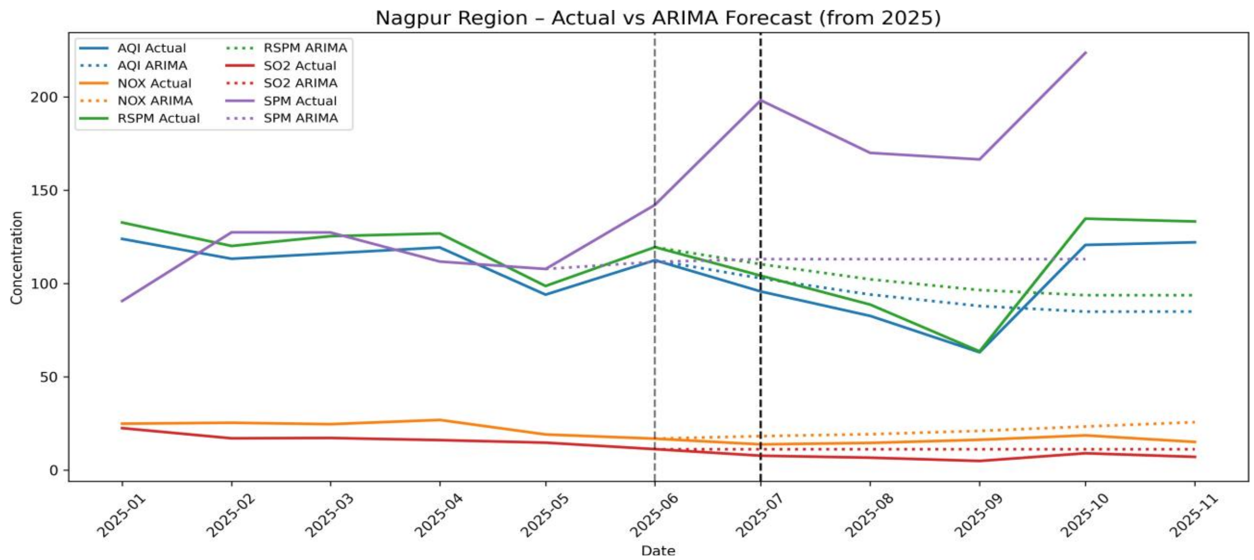

Figure 24: ARIMA-based forecast for air quality components in the Nagpur region

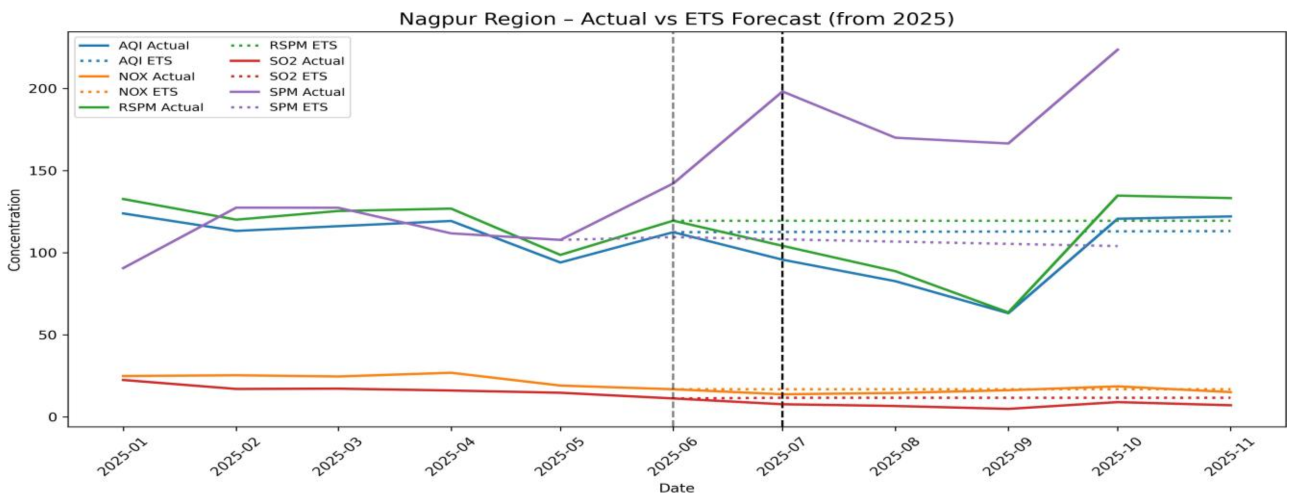

Figure 25: ETS-based forecast for air quality components in the Nagpur region

Table 18: Forecasting performance comparison for Nagpur region using PROPHET, ARIMA, and ETS models (MAE and MAPE).

Metric	AQI	NOX	RSPM	SO2	SPM
MAE (PROPHET)	15.08	6.37	19.10	4.34	62.88
MAPE (%) (PROPHET)	15.61	41.81	18.21	67.46	33.51
MAE (ARIMA)	24.26	1.89	27.54	5.32	52.71
MAPE (%) (ARIMA)	27.77	12.62	29.52	83.00	27.80
MAE (ETS)	22.75	1.97	26.20	4.84	72.31
MAPE (%) (ETS)	29.33	13.19	31.81	76.53	38.71

Forecast evaluation for the Nagpur region reveals mixed model performance across pollutants. For AQI, PROPHET achieves the lowest MAE (15.08) and MAPE (15.61%), indicating superior predictive capability relative to ARIMA and ETS, both of which record higher absolute and percentage errors.

In the case of NO_x, ARIMA and ETS substantially outperform PROPHET in terms of absolute accuracy, yielding MAE values below 2.00 and MAPE around 12–13%. PROPHET reports considerably higher relative deviation (MAPE 41.81%), suggesting reduced robustness for nitrogen oxide forecasting.

RSPM predictions show moderate error magnitudes across all models. PROPHET again provides comparatively lower MAE (19.10) and MAPE (18.21%), whereas ARIMA and ETS exhibit larger deviations, with MAPE exceeding 29%.

SO_2 forecasting presents higher relative errors for all models. Although PROPHET achieves the smallest MAE (4.34), ARIMA and ETS display elevated MAPE values (above 76%), indicating instability in percentage-based evaluation for this pollutant.

For SPM, ARIMA yields the lowest MAE (52.71) and MAPE (27.80%), suggesting relatively better performance for coarse particulate matter prediction. PROPHET performs moderately (MAPE 33.51%), while ETS shows the highest absolute and relative deviations. Overall, PROPHET demonstrates stronger consistency for AQI and RSPM, whereas ARIMA appears more suitable for NO_x and SPM forecasting in the Nagpur region.

Nashik Region:

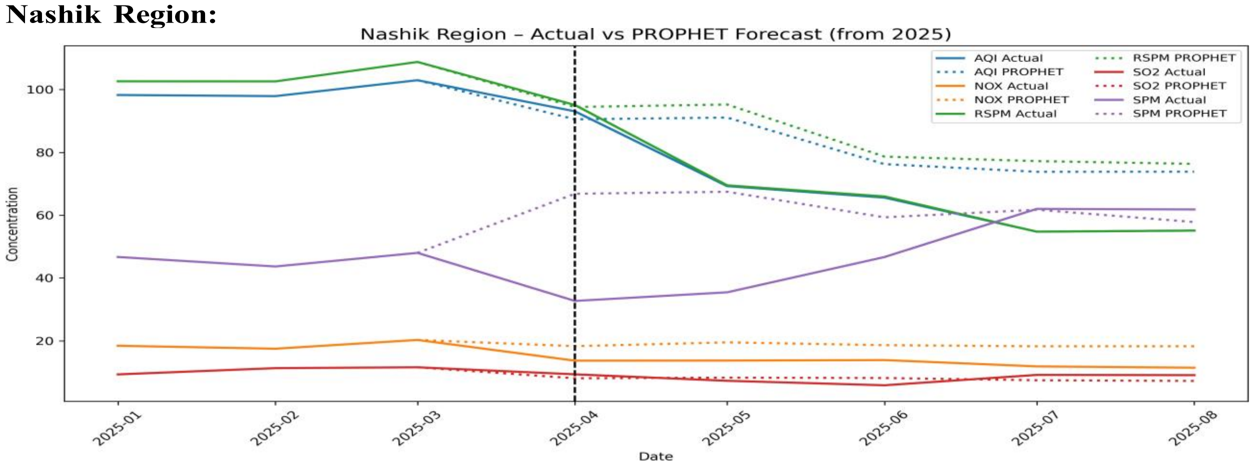

Figure 26: Prophet-based forecast for air quality components in the Nashik region

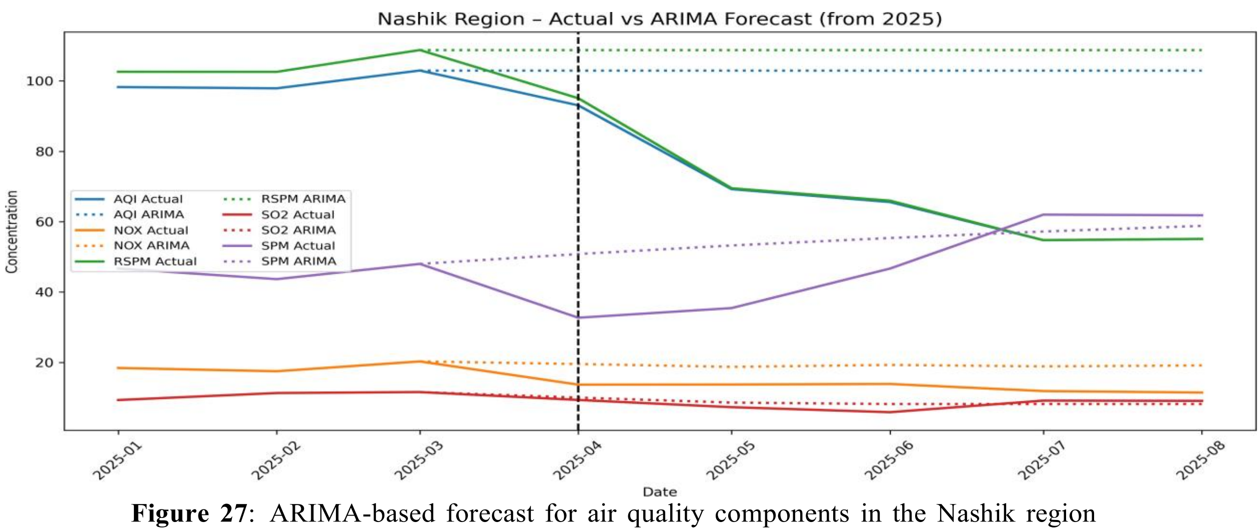

Figure 27: ARIMA-based forecast for air quality components in the Nashik region

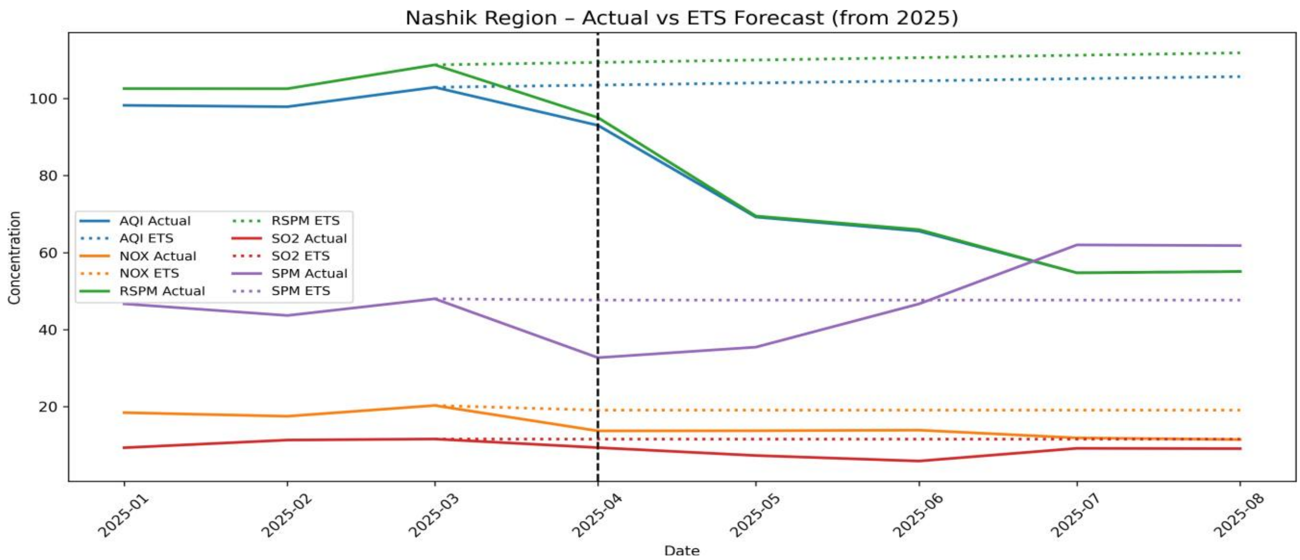

Figure 28: ETS-based forecast for air quality components in the Nashik region

Table 19: Forecasting performance comparison for Nashik region using PROPHET, ARIMA, and ETS models (MAE and MAPE).

Metric	AQI	NOX	RSPM	SO2	SPM
MAE (PROPHET)	5.07	2.35	4.11	4.77	31.45
MAPE (%) (PROPHET)	7.19	17.77	5.48	56.74	63.63
MAE (ARIMA)	35.40	7.03	40.69	1.63	16.79
MAPE (%) (ARIMA)	58.24	55.42	66.37	22.72	31.01
MAE (ETS)	35.40	6.42	40.69	3.30	11.27
MAPE (%) (ETS)	58.24	50.69	66.37	45.12	25.36

Forecast evaluation for the Nashik region indicates pronounced differences in model performance across pollutants. For AQI, PROPHET clearly outperforms ARIMA and ETS, achieving substantially lower MAE (5.07) and MAPE (7.19%). In contrast, both ARIMA and ETS exhibit considerably higher absolute and percentage errors (MAE 35.40; MAPE 58.24%), suggesting reduced suitability for overall air quality prediction in this region.

For NO_x, PROPHET again provides the smallest MAE (2.35) and a relatively moder- ate MAPE (17.77%). ARIMA and ETS display larger deviations, with MAPE values exceeding 50%, indicating greater forecasting instability.

RSPM forecasts follow a similar pattern, with PROPHET achieving markedly superior ac- curacy (MAE 4.11; MAPE 5.48%), whereas ARIMA and ETS produce substantially higher errors (MAE 40.69; MAPE 66.37%). This highlights the effectiveness of the PROPHET framework in capturing particulate matter dynamics for Nashik.

In contrast, SO_2 forecasting shows improved performance under ARIMA (MAE 1.63; MAPE 22.72%), while PROPHET records higher relative deviation (MAPE 56.74%). ETS exhibits intermediate behavior, indicating that linear time-series structures may better capture sulfur dioxide variability in this region.

For SPM, ETS yields the lowest MAE (11.27) and relatively lower MAPE (25.36%), followed by ARIMA, while PROPHET shows comparatively larger percentage deviation (63.63%). Overall, PROPHET dominates AQI and RSPM prediction, whereas ARIMA and ETS demonstrate better performance for SO_2 and SPM in the Nashik region.

Pune Region:

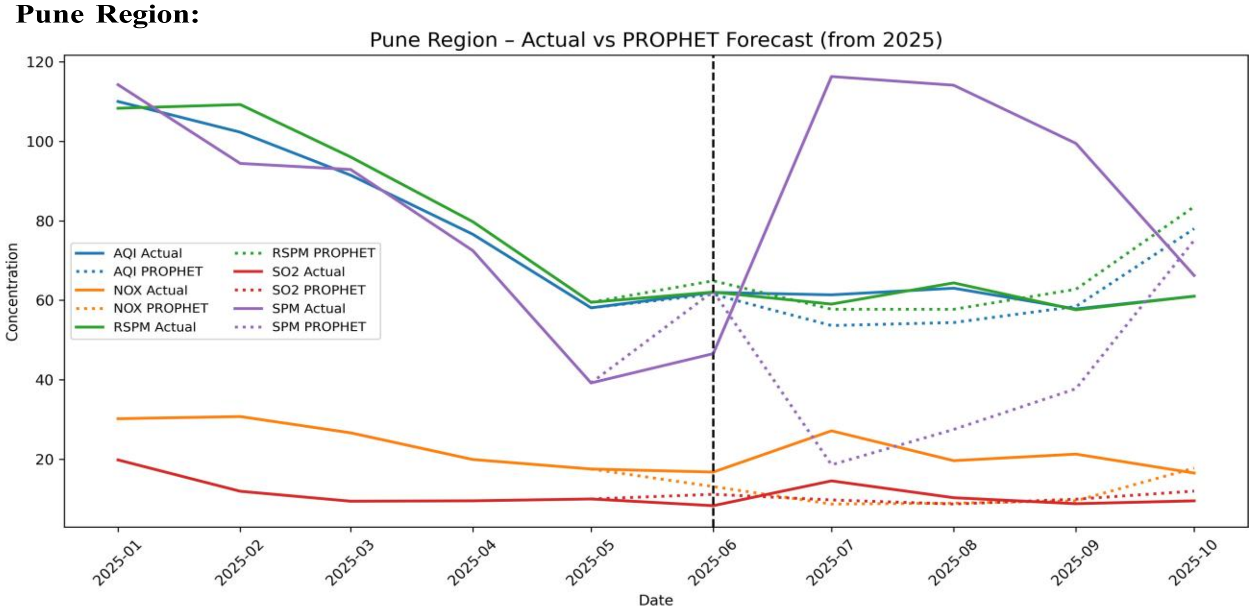

Figure 29: Prophet-based forecast for air quality components in the Pune region

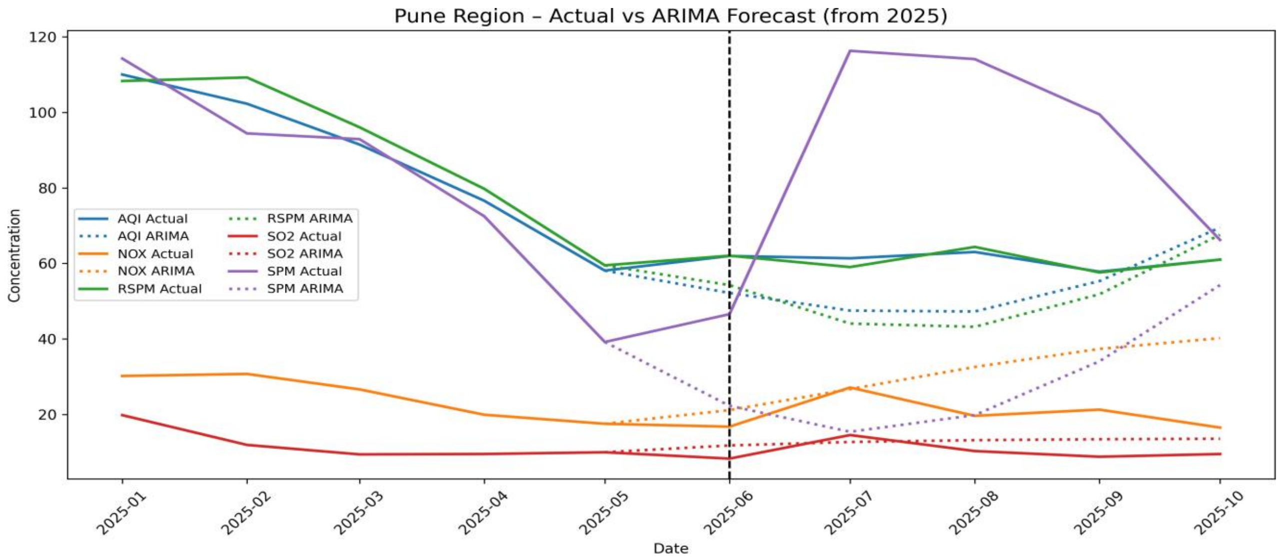

Figure 30: ARIMA-based forecast for air quality components in the Pune region

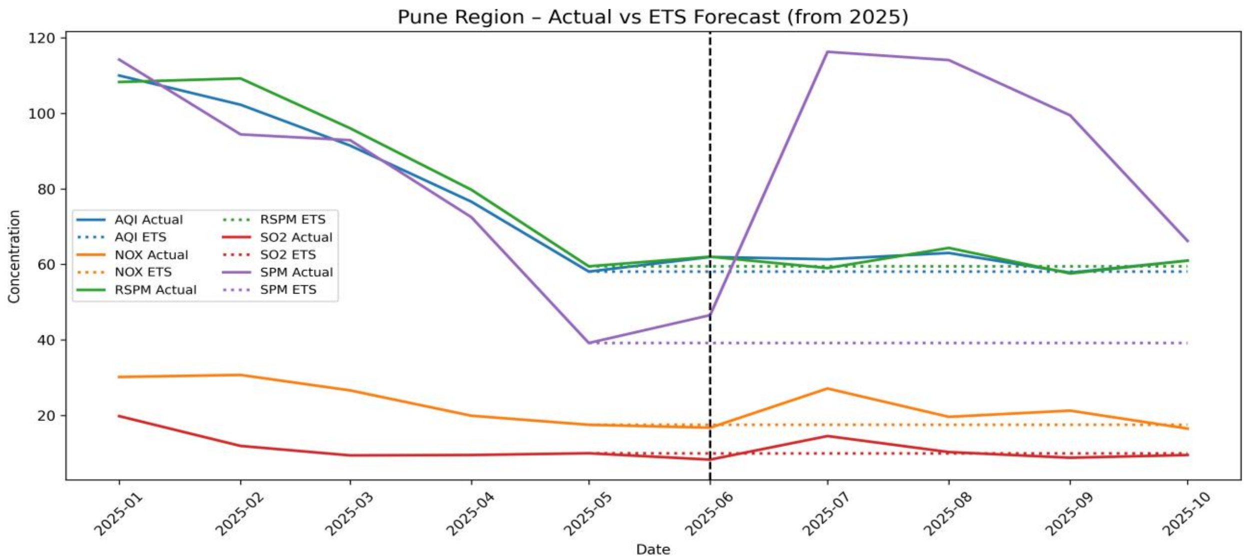

Figure 31: ETS-based forecast for air quality components in the Pune region

Table 20: Forecasting performance comparison for Pune region using PROPHET, ARIMA, and ETS models (MAE and MAPE).

Metric	AQI	NOX	RSPM	SO2	SPM
MAE (PROPHET)	7.45	5.53	13.35	2.38	53.23
MAPE (%) (PROPHET)	12.10	25.49	21.82	22.22	53.27
MAE (ARIMA)	10.72	6.97	10.20	1.63	59.52
MAPE (%) (ARIMA)	17.44	33.01	16.69	14.41	62.47
MAE (ETS)	3.05	3.45	2.26	1.62	49.36
MAPE (%) (ETS)	4.93	14.91	3.65	14.31	49.86

Forecast evaluation for the Pune region indicates strong performance of the ETS model across most pollutants. For AQI, ETS achieves the lowest MAE (3.05) and MAPE (4.93%), substantially outperforming PROPHET (MAE 7.45; MAPE 12.10%) and ARIMA (MAE 10.72; MAPE 17.44%). This suggests enhanced capability of exponential smoothing techniques in capturing overall air quality dynamics in Pune.

For NO_x, ETS again demonstrates superior accuracy with the smallest MAE (3.45) and MAPE (14.91%), followed by PROPHET and ARIMA, both of which exhibit comparatively higher deviations.

RSPM forecasts show a pronounced advantage for ETS, which records minimal error (MAE 2.26; MAPE 3.65%). ARIMA performs moderately (MAPE 16.69%), while PROPHET reports comparatively larger absolute and percentage deviations.

SO_2 prediction accuracy remains relatively stable across ARIMA and ETS, both achieving similar MAE values (1.62–1.63) and MAPE around 14%. PROPHET shows slightly higher relative deviation (22.22%), though differences are not substantial.

For SPM, all models exhibit comparatively high absolute errors (49.36–59.52) and elevated MAPE values (approximately 50–62%), indicating persistent forecasting challenges for coarse particulate matter in Pune. Overall, ETS emerges as the most consistent and accurate model across AQI, NO_x, and RSPM in the Pune region.

7 Implications for Air Quality Management in Maharashtra

The integrated analytical and forecasting framework developed in this study provides actionable insights to strengthen air quality management across Maharashtra. Region-wise statistical analyses reveal distinct pollution profiles influenced by local emission sources, meteorology, and urbanization patterns, highlighting the importance of tailored, region-specific mitigation strategies rather than uniform statewide interventions. Month-wise and seasonal assessments further inform the timing of regulatory actions, such as intensified enforcement during winter stagnation periods or targeted interventions in post-monsoon months when particulate matter levels typically rise.

The Prophet-based forecasting system offers an interpretable and operationally feasible early warning tool to support real-time decision-making by pollution control authorities. By anticipating short-term pollutant peaks, agencies such as the MPCB and municipal bodies can implement timely measures-including traffic restrictions, industrial activity advisories, dust-control protocols, and public health alerts. Integrating these forecasts with state-level and national initiatives, such as the National Clean Air Programme (NCAP), can enhance coordination and ensure that local actions are aligned with broader policy frameworks.

Additionally, the region-wise dataset and reproducible analytical pipeline provide a foundation for future enhancements, including the integration of meteorological variables, emission inventories, and advanced machine learning or hybrid forecasting models. Such developments could further improve predictive accuracy and support proactive, evidence-based air quality management across the state.

8 Conclusions

This study presents a comprehensive statistical assessment of atmospheric pollution across major regions of Maharashtra using multi-year observations from the MPCB monitoring network. Through region-wise and month-wise aggregation, the analysis captures spatial heterogeneity, seasonal dynamics, and interrelationships among pollutants. Descriptive statistics and correlation analysis reveal consistent seasonal structures and region-specific pollutant patterns, reflecting underlying meteorological influences, emission sources, and atmospheric dispersion mechanisms.

To evaluate predictive capability, three time-series forecasting approaches - Prophet, ARIMA, and ETS were implemented and systematically compared using scale-dependent and scale- independent error metrics. The comparative analysis demonstrates that model performance varies substantially across regions and pollutants, indicating strong spatial and temporal het- erogeneity in air quality behavior.

The ETS framework exhibits relatively stable and consistent forecasting performance across several regions, particularly for pollutants characterized by clear trend and seasonal components. Prophet performs competitively in datasets with pronounced seasonality and smoother temporal structures. In contrast, ARIMA shows moderate performance but appears more sensitive to irregular fluctuations and structural variability in certain regional series.

For pollutants exhibiting intermittent behavior or near-zero observations, percentage-based accuracy measures become highly unstable, leading to inflated error magnitudes. This highlights an important statistical limitation of MAPE-type metrics when applied to low-concentration or highly volatile pollutants. The observed variability in prediction accuracy across models and regions underscores the need for careful metric selection and model validation when analyzing environmental time series.

Overall, the findings confirm that air quality dynamics in Maharashtra are statistically heterogeneous and region-specific. While univariate seasonal-trend models provide a useful forecasting baseline, their predictive reliability depends strongly on pollutant characteristics and regional variability. Future research should extend this framework by incorporating multivariate structures, meteorological covariates, and robust error measures that remain stable under low- concentration conditions.

The study establishes a statistically grounded foundation for regional air quality monitoring and forecasting, offering quantitative insights that can support evidence-based environmental planning, regulatory strategy, and public health interventions across Maharashtra.

Acknowledgements

The authors express their sincere gratitude to the Maharashtra Pollution Control Board (MPCB) for maintaining publicly accessible air quality monitoring datasets that made this analysis possible. The authors also acknowledge the valuable contributions of the open-source scientific computing community; tools such as Python, Pandas, NumPy, Matplotlib, and the Prophet forecasting library played a central role in data processing, statistical analysis, and model development. Their continued availability and advancement greatly facilitate transparent, reproducible environmental research.

References

[1] L. Bai, J. Wang, X. Ma, and H. Lu, “Air pollution forecasts: An overview,” International Journal of Environmental Research and Public Health, vol. 15, no. 4, 780, 2018. doi: 10.3390/ijerph15040780.

[2] Q. Tao, F. Liu, Y. Li, and D. Sidorov, “Air pollution forecasting using a deep learning model based on 1D convnets and bidirectional GRU,” IEEE Access, vol. 7, pp. 76690–76698, 2019. doi: 10.1109/ACCESS.2019.2921578.

[3] A. Heydari, M. M. Nezhad, D. A. Garcia, F. Keynia, and L. De Santoli, “Air pollution forecasting application based on deep learning model and optimization algorithm,” Clean Technologies and Environmental Policy, vol. 24, pp. 607–621, 2022. doi: 10.1007/s10098- 021-02080-5.

[4] S. Gulia, N. Shukla, L. Padhi, P. Bosu, S. K. Goyal, and R. Kumar, “Evolution of air pollution management policies and related research in India,” Environmental Challenges, vol. 6, 100431, 2022. doi: 10.1016/j.envc.2021.100431.

[5] Y.-S. Chang, H.-T. Chiao, S. Abimannan, Y.-P. Huang, Y.-T. Tsai, and K.-M. Lin, “An LSTM-based aggregated model for air pollution forecasting,” Atmospheric Pollution Re- search, vol. 11, no. 8, pp. 1451–1463, 2020. doi: 10.1016/j.apr.2020.05.015.

[6] M. A. Garaga, B. R. Sahu, and S. K. Panda, “A review on urban road transport emissions and management,” Aerosol and Air Quality Research, vol. 18, pp. 1901–1915, 2018. doi: 10.4209/aaqr.2018.01.0003.

[7] J. Li, Y. Song, and X. Zhang, “Hybrid deep learning models for air quality prediction,” At- mospheric Environment, vol. 192, pp. 178–188, 2019. doi: 10.1016/j.atmosenv.2018.08.038.

[8] M. A. Alghamdi et al., “Sustainable air quality management using machine learning,” Sustainability, vol. 14, no. 15, 9951, 2022. doi: 10.3390/su14159951.

[9] S. Kumar, R. Patel, and A. Singh, “Machine learning approaches for $PM_{2.5}$ concen- tration forecasting,” Environmental Research Letters, vol. 19, 2024. doi: 10.1088/1748- 9326/ad3736.

[10] H. Wang et al., “Deep learning framework for multi-pollutant air quality forecasting,” Atmospheric Environment, vol. 305, 119812, 2023. doi: 10.1016/j.atmosenv.2023.119812.

[11] K. Patel and S. Rao, “Optimization algorithms for pollution prediction models,” Procedia Computer Science, vol. 234, pp. 1284–1291, 2024. doi: 10.1016/j.procs.2024.02.248.

[12] A. Sharma et al., “Statistical analysis of urban air quality data,” Procedia Computer Sci- ence, vol. 225, pp. 2060–2067, 2023. doi: 10.1016/j.procs.2023.10.060.

[13] R. Gupta and N. Verma, “Transportation emissions and air quality modeling,” Transporta- tion Research Procedia, vol. 72, pp. 3745–3752, 2023. doi: 10.1016/j.trpro.2023.11.745.

[14] P. Singh et al., “Advanced time series forecasting for environmental monitoring,” Results in Engineering, vol. 19, 10024X, 2023. doi: 10.1016/j.rineng.2023.10024X.

[15] L. Chen et al., “Industrial air pollution forecasting using hybrid models,” Computers & Industrial Engineering, vol. 169, 108191, 2022. doi: 10.1016/j.cie.2022.108191.

[16] M. Kumar and V. Rao, “Environmental impact assessment using machine learning,” Pro- cedia Computer Science, vol. 142, pp. 755–762, 2018. doi: 10.1016/j.procs.2018.10.755.

[17] S. Desai et al., “ARIMA models for transportation-related air pollution,” Transportation Research Procedia, vol. 72, pp. 3619–3626, 2025. doi: 10.1016/j.trpro.2025.01.619.

[18] A. Nair and R. Mehta, “Software tools for air quality data analysis,” Software Impacts, vol. 25, 100045, 2025. doi: 10.1016/j.simpa.2025.100045.

[19] S. Patel et al., “Classical statistical methods for air pollution forecasting,” International Journal of Computer Applications, vol. 8, no. 9, 1006, 2009. doi: 10.5120/1069-106.

[20] J. Zhang et al., “Regression models for environmental time series,” Environmental Mod- elling & Software, vol. 125, 1049504, 2020. doi: 10.1016/j.envsoft.2020.104950.

[21] M. Liu et al., “Stochastic modeling of atmospheric pollutants,” Stochastic Environmental Research and Risk Assessment, vol. 37, 2023. doi: 10.1007/s00477-023-02512-2.

[22] “Multi-pollutant air quality forecasting techniques,” Atmospheric Pollution Research, vol. 11, no. 12, 2020. doi: 10.1016/j.apr.2020.09.015.

[23] H. Zhang et al., “Grey prediction models for environmental data,” International Journal of Environmental Science and Technology, vol. 19, pp. 4241–4252, 2022. doi: 10.1007/s13762- 022-04241-5.

[24] K. Wang et al., “Environmental forecasting systems using deep learning,” Frontiers in Environmental Science, vol. 10, 945628, 2022. doi: 10.3389/fenvs.2022.945628.

[25] Y. Li et al., “Chemical transport modeling and machine learning integration,” Chemo- sphere, vol. 359, 1419945, 2024. doi: 10.1016/j.chemosphere.2024.141994.

[26] S. Chen et al., “Deep neural networks for hazardous air pollutant prediction,” Journal of Hazardous Materials, vol. 475, 1345544, 2025. doi: 10.1016/j.jhazmat.2025.134554.

[27] R. Sharma et al., “Data preprocessing methodologies for air quality analysis,” Procedia Computer Science, vol. 225, pp. 2060–2067, 2023. doi: 10.1016/j.procs.2023.10.060.

[28] “Atmospheric pollution forecasting methodology advancements,” Atmospheric Pollution Research, vol. 13, 10 0431, 2022. doi: 10.1016/j.apr.2022.100431.