Air Quality Trends in Indian Cities: A Time Series Analysis

Abhi Thumar (202101123),* Charith Kutikuppala (202101128),† and Harsh Mungara (202101130)‡
Dhirubhai Ambani Institute of Information & Communication Technology,
Gandhinagar, Gujarat 382007, India
SC475, Time Series Analysis

Abstract

Air pollution is a major concern in India, affecting the health of millions. This project examines the air quality in 435 Indian cities from 2010 to 2023. We started by gathering and cleaning the air quality data to make sure it was ready for analysis. We then used graphs to show how air pollution levels have changed over the years. Through exploratory data analysis, we identify patterns and outliers in the data. Time series decomposition allows us to separate these trends into seasonal variations and underlying trends, enhancing our understanding of the cyclic nature of air pollution. Additionally, we looked at how much each city contributes to pollution in its state, giving us insight into where pollution is coming from. We also conduct stationarity checks and autocorrelation analyses to prepare the data for forecasting. We aim to understand how air quality has changed over time and what might be causing these changes. Additionally, we apply a statistical method called SARIMAX to predict future air quality levels based on past data. The goal of this study is to better understand the factors influencing air quality in Indian cities and suggest ways to improve it, making our cities healthier places to live.

I. INTRODUCTION

Air pollution in India remains a critical environmental and health issue. In this project, we delve into the air quality across 435 cities from 2010 to 2023, examining pollutants like PM10, PM2.5, and others that significantly impact health. Our study begins with gathering and cleaning extensive air quality data, ensuring it is ready for thorough analysis. We use graphical methods to illustrate how air pollution levels have shifted over the years and deploy statistical tools to uncover underlying trends and seasonal patterns. The goal is to figure out how air quality has changed over time and what might be causing these changes. We also aim to predict future air pollution levels, which could help in planning better urban environments and public health strategies. This predictive insight is crucial for policymakers and health professionals aiming to reduce the effects of air pollution and enhance public health.

A. Proposed Approach

- **Data Extraction and Preprocessing:** We started by collecting air quality data from Kaggle[1], cleaning it up to make sure it was usable for our analysis. This included fixing any errors in the data, filling in missing values, and making sure everything was consistent.
- **Exploratory Data Analysis (EDA):** We looked at the data closely to understand the levels of different pollutants and how they vary. This helped us get a basic understanding of the air quality and spot any unusual patterns.
- **Selection of Cities of Interest:** We calculated how much pollution each city contributes to its state and focused on those that contribute the most. This helped us target our analysis on the most polluted areas.
- **Time Series Decomposition:** We broke down the historical air quality data to identify ongoing trends and seasonal patterns. This step helped us see how air quality has been changing over the years and at different times of the year.
- **Trend and Seasonality Identification:** We focus on identifying and quantifying the trends and seasonal variations in air quality, which are crucial for predicting future pollution levels.
- **Stationarity Check:** We checked if our data was stationary, meaning its statistical properties like mean and variance don't change over time. This is important because many forecasting methods assume the data they are using is stationary.
- **Autocorrelation Analysis:** We analyzed if and how current air quality levels are related to previous levels. This helps in understanding how past conditions might predict future conditions.
- **Detrending and Model Fitting:** We removed trends from our data to focus on more detailed fea-

*Electronic address: 202101123@daiict.ac.in
†Electronic address: 202101128@daiict.ac.in
‡Electronic address: 202101130@daiict.ac.in

tures like seasonal changes. Then, we used a SARIMAX model, which is good for data that shows both seasonal trends and non-seasonal shifts, to forecast future air quality.

- **Forecasting:** Using our model, we tried to predict future levels of air pollution. These predictions can help in making plans to improve air quality and public health.

II. MODEL

In this project, we aim to forecast air quality by analyzing time-series data on air pollutants collected from various cities across India. Given the sequential nature of our data, with clear trends and potential seasonal patterns due to varying environmental conditions throughout the year, it is crucial to employ robust forecasting models that can effectively handle such characteristics. The ARIMA model was chosen due to its flexibility in modeling a wide range of time series data by addressing non-stationarities through differencing, while capturing the dynamics via autoregressive and moving average components. Additionally, to incorporate the influence of seasonal variations observed in air quality metrics and potential external factors like weather or industrial activity, the SARIMAX model was selected. SARIMAX extends ARIMA by integrating seasonal decomposition and exogenous variables, making it highly suitable for our data's complexity and enhancing the forecast accuracy.[2]

A. ARIMA Model

The **ARIMA model**, short for *AutoRegressive Integrated Moving Average*, is a popular forecasting technique that combines autoregressive features with moving averages. It is particularly useful for time series data that shows patterns of autocorrelation.

1. *Model Components*

- **AR (Autoregressive) part:** This component predicts the variable of interest using a linear combination of previous values.
- **I (Integrated) part:** This involves differencing the time series to make it stationary, i.e., constant mean and variance over time.
- **MA (Moving Average) part:** This uses the dependency between an observation and a residual error from a moving average model applied to lagged observations.

2. *Mathematical Representation*

The ARIMA model is expressed as follows:

$$Y_t = \alpha + \sum_{i=1}^{p} \phi_i Y_{t-i} + \sum_{i=1}^{q} \theta_i \epsilon_{t-i} + \epsilon_t \tag{1}$$

Where:

- ϕ_i are the parameters of the AR part.
- θ_i are the parameters of the MA part.
- p is the order of the AR part.
- q is the order of the MA part.
- ϵ_t is the error term.
- α is a constant.

B. SARIMAX Model

The **SARIMAX model** (Seasonal AutoRegressive Integrated Moving Average with eXogenous variables model) extends ARIMA by accounting for seasonality and incorporating external variables. It is ideal for datasets with seasonal trends and where external factors influence the predictions.

1. *Model Enhancements*

- **Seasonal Elements:** It includes seasonal differencing and seasonal terms in the AR and MA parts, which handle patterns that repeat at fixed intervals.
- **Exogenous Variables:** It integrates variables external to the dataset to improve the accuracy of forecasts.

2. *Mathematical Representation*

The general form of the SARIMAX model can be written as:

$$\begin{aligned} Y_t = \alpha + \sum_{i=1}^{p} \phi_i Y_{t-i} + \sum_{i=1}^{q} \theta_i \varepsilon_{t-i} + \sum_{i=1}^{P} \Phi_i Y_{t-si} \\ + \sum_{i=1}^{Q} \Theta_i \varepsilon_{t-si} + \beta X_t + \varepsilon_t \end{aligned} \tag{2}$$

Where:

- ϕ and Φ are the non-seasonal and seasonal AR coefficients, respectively.

- θ and Θ are the non-seasonal and seasonal MA coefficients, respectively.
- p and P are the orders of the non-seasonal and seasonal AR terms.
- q and Q are the orders of the non-seasonal and seasonal MA terms.
- s is the length of the seasonal cycle.
- β is the coefficient for the exogenous variables X_t.
- ϵ_t is the error term.

III. RESULTS AND DISCUSSION

Across various states in India, the pollution impact due to PM2.5, CO, NO2, and SO2 shows significant variation, highlighting a mixture of urban, industrial, and regional factors contributing to air quality degradation. For instance, large cities like Chennai, Kanpur, and Kolkata consistently appear with high pollution levels across different pollutants, indicating the cumulative effect of dense population, heavy traffic, and industrial activities. In contrast, smaller or less urbanized cities generally show lower percentages of pollutants, though this could also be attributed to less comprehensive monitoring.

Particularly in North India, Uttar Pradesh and Bihar exhibit high levels of all pollutants, with cities like Kanpur, Ghaziabad, Patna, and Muzaffarpur standing out. This suggests a strong influence of industrial activities alongside vehicular emissions. Moving towards South India, cities like Coimbatore and Chennai in Tamil Nadu show alarmingly high concentrations of pollutants, particularly SO2 and PM2.5, which are likely linked to their industrial bases and urban traffic.

A. Trend, Seasonality and Residual Plots

1. *Ghaziabad*

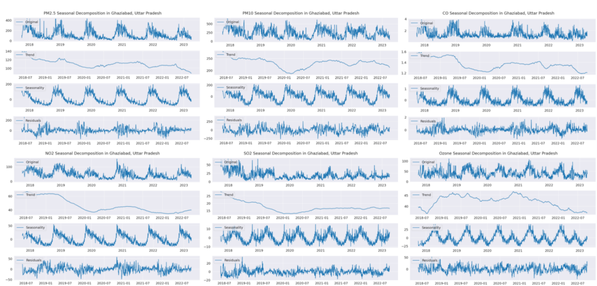

FIG. 1: Seasonal decomposition of air quality parameters (PM2.5, PM10, CO, NO2, SO2, Ozone) measured in Ghaziabad, Uttar Pradesh, from 2018 to 2023.

The graphs depict the seasonal decomposition of various air quality parameters measured in Ghaziabad, Uttar Pradesh, from 2018 to 2023. Here are some inferences and observations from these graphs:

a. General Observations Across All Pollutants:

1. **Seasonality**: All pollutants exhibit clear seasonal patterns, indicating a recurring cycle throughout the year. This phenomenon is common in air quality data due to changes in weather, temperature, and human activities like heating and agricultural practices.
2. **Trend**: Most pollutants show a decreasing trend over the years, suggesting a gradual improvement in air quality or the effectiveness of pollution control measures.
3. **Residuals**: The residuals generally show random fluctuations, indicating that the model has accounted for most of the systematic information in the original data. However, some spikes suggest occasional unmodeled influences, such as weather events, industrial activity, or burning.

b. Specific Observations and Possible Anomalies:

1. **PM2.5 and PM10**: Both particulate matters show significant spikes, especially during certain times of the year, likely corresponding to winter months when heating increases and atmospheric conditions trap pollutants closer to the ground.
2. **CO**: The carbon monoxide graph shows less pronounced seasonal variation compared to PM2.5 and PM10 but maintains a slight seasonal cycle. The trend is relatively flat from 2020 onwards.
3. **NO2**: Nitrogen dioxide shows sharp peaks, which might be associated with increased vehicular traffic or industrial activity. A decreasing trend suggests improvements, possibly due to stricter emission regulations or changes in traffic patterns.
4. **SO2**: Sulphur dioxide exhibits more variability in its residuals, which could be due to irregular industrial activities or changes in fuel types used in industries and vehicles.
5. **Ozone**: Ozone is unique because it shows less of a declining trend compared to other pollutants and has more pronounced seasonality, which could be related to its photochemical formation in the presence of sunlight, hence more prevalent in summer months.

c. Abnormalities and Their Possible Causes:

- **Sharp Peaks**: Spikes in PM2.5 and PM10 around the same time each year could be due to Diwali when fireworks significantly contribute to air pollution.

- **Abnormal Dips in Data**: Any sudden drops in air pollutant levels might be due to lockdowns (such as those during the COVID-19 pandemic), which drastically reduced human activity and thus emissions.

2. Varanasi

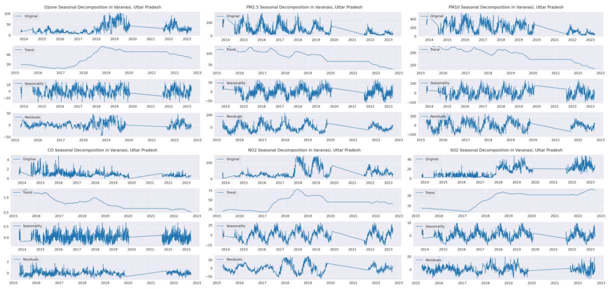

FIG. 2: Seasonal decomposition of air quality parameters (PM2.5, PM10, CO, NO2, SO2, Ozone) measured in Varanasi, Uttar Pradesh, from 2014 to 2023.

The provided graphs illustrate the seasonal decomposition of various air quality parameters measured in Varanasi, Uttar Pradesh, over the period from 2014 to 2023. Here are the detailed observations and possible interpretations:

a. General Observations Across All Pollutants:

1. **Seasonality**: All parameters exhibit clear seasonal patterns, indicative of typical variations in air quality related to climatic conditions, agricultural burning, festivals, and changes in traffic volume throughout the year.

2. **Trend**: There is a mix of increasing and decreasing trends among different pollutants, reflecting various influences like policy changes, economic activities, and environmental regulations.

3. **Residuals**: The residuals generally show random fluctuations around zero, suggesting that the seasonal and trend components have been reasonably captured, though some parameters show large spikes indicating unmodeled or irregular events.

b. Specific Observations and Possible Anomalies:

1. **PM2.5**: Shows a general decline in trend until around 2020, then stabilizes. The significant spikes in seasonal components often occur in colder months when heating and agricultural burning are more prevalent.

2. **PM10**: Similar to PM2.5, PM10 levels show a declining trend with prominent seasonal peaks. The slight rebound in recent years could be influenced by local construction activities or reduced effectiveness of pollution control measures.

3. **CO**: The carbon monoxide levels show a steady decline over the years, reflecting possibly improved fuel standards and better vehicle emissions controls. Seasonal peaks could be related to increased heating during winter.

4. **NO2**: Shows a notable increase in trend from around 2017 onwards, potentially linked to increased vehicular traffic or industrial emissions. The seasonal pattern remains relatively consistent over the years.

5. **SO2**: A steady increase in trend is observed, which is somewhat unusual compared to other pollutants and could be attributed to specific local industrial activities or changes in fuel usage that increase sulfur emissions.

6. **Ozone**: Displays a somewhat stable trend with slight decreases, which could be related to overall reductions in precursors like NOx and volatile organic compounds. Ozone peaks during warmer months due to its photochemical formation processes.

c. Abnormalities and Their Possible Causes:

- **Sudden Spikes**: Commonly observed in the PM2.5 and PM10 data during winter months, possibly due to the burning of crop residue and increased use of fireworks during festivals like Diwali.

- **Anomalous Dips**: Particularly visible around 2020, which could correspond to reduced economic and social activities during the COVID-19 lockdowns, leading to temporarily improved air quality.

3. Noida

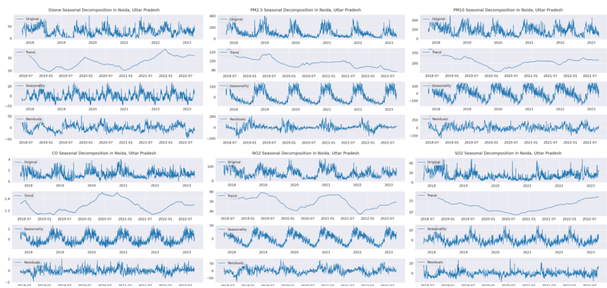

FIG. 3: Seasonal decomposition of air quality parameters (PM2.5, PM10, CO, NO2, SO2, Ozone) measured in Noida, Uttar Pradesh, from 2018 to 2023.

The provided graphs illustrate the seasonal decomposition of several air quality parameters measured in Noida, Uttar Pradesh, from 2018 to 2023. Let's discuss the key observations and possible abnormalities:

a. General Observations Across All Pollutants:

1. **Seasonality**: All pollutants exhibit clear seasonal variations, which could be associated with changes in weather, industrial activity, and other seasonal factors such as agricultural burning and fireworks during festivals.

2. **Trend**: Most pollutants show a declining trend over the years, indicating possible improvements in air quality control measures or shifts in industrial and traffic patterns.

3. **Residuals**: The residuals are mostly centered around zero with occasional spikes, indicating that the seasonal and trend components have captured most of the systematic variation in the data. Spikes in residuals could suggest external events or anomalies not captured by the model.

b. Specific Observations and Possible Anomalies:

1. **PM2.5 and PM10**: These graphs show pronounced seasonal peaks, often during the winter months, which can be attributed to increased heating emissions and atmospheric inversions that trap pollutants near the ground. A declining trend suggests improved regulations and changes in emission sources.

2. **CO**: The carbon monoxide graph exhibits a fluctuating trend with a noticeable drop around 2020, possibly due to reduced vehicular traffic during COVID-19 lockdowns. The rise in seasonality around winter is typical due to increased heating and vehicular use in colder temperatures.

3. **NO2**: There's a significant decline in the trend, which is promising as NO2 is primarily emitted from vehicles and industrial activity. The reduction could be linked to better emissions technology and stricter environmental policies.

4. **SO2**: Shows a more consistent seasonal pattern with smaller peaks. The gradual decrease in trend may indicate a reduction in the use of sulfur-containing fuels and stricter industrial emission standards.

5. **Ozone**: Interestingly, ozone shows an upward trend initially followed by a decline. This could be due to changes in precursors like NO2 and VOCs (Volatile Organic Compounds). Ozone's seasonality, peaking in warmer months, aligns with its photochemical formation processes.

c. Abnormalities and Their Possible Causes:

- **COVID-19 Impact**: Noticeable declines in some pollutants during 2020 could be directly linked to the global pandemic's impact on reducing traffic, industrial activities, and overall human mobility.

- **Sudden Spikes**: Occasional spikes in PM2.5 and PM10 may occur due to unregulated or unusually high burning activities, possibly from agricultural stubble burning or festival-related fireworks.

- **Residual Spikes**: These might indicate local events like construction activities, unregulated industrial operations, or temporary lapses in pollution control measures.

4. Chandrapur

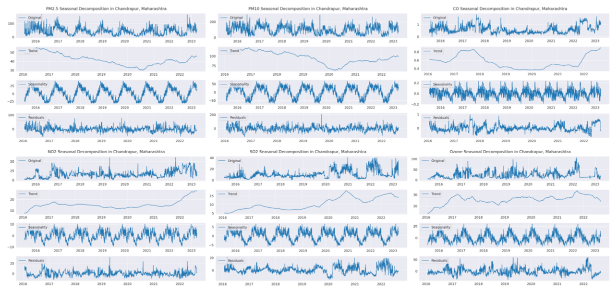

FIG. 4: Seasonal decomposition of air quality parameters (PM2.5, PM10, CO, NO2, SO2, Ozone) measured in Chandrapur, Maharashtra, from 2016 to 2023.

The seasonal decomposition graphs for Chandrapur, Maharashtra, covering the period from 2016 to 2023 for PM2.5, PM10, CO, NO2, SO2, and Ozone, provide a detailed analysis of the patterns in air pollutant concentrations over time. Here are the inferences and observations from these graphs:

a. General Observations Across All Pollutants:

1. **Seasonality**: All pollutants exhibit clear seasonal variations, which is typical due to the influence of weather conditions, heating practices, agricultural activities, and other seasonal factors.

2. **Trend**: Most pollutants show a declining or stable trend, indicating potential effectiveness of air quality management strategies or changes in emission sources.

3. **Residuals**: Fluctuations in the residuals indicate deviations from the seasonal and trend patterns, which might be caused by unusual events or changes in local activities.

b. Specific Observations and Possible Anomalies:

1. **PM2.5 and PM10**: Both exhibit similar seasonal peaks, typically seen during winter months when colder temperatures and lower wind speeds can trap pollutants. The trend for PM2.5 is gradually decreasing, while PM10 remains relatively stable, suggesting some success in managing finer particulate sources.

2. **CO**: Displays a decreasing trend, which could be linked to improvements in vehicle emissions and fuel quality. Seasonal peaks are less pronounced, suggesting that CO is less affected by seasonal factors compared to particulate matter.

3. **NO2**: Shows a relatively stable trend with a slight decrease, reflecting possibly improved controls over emissions from vehicles and industrial activities. The seasonal pattern suggests higher winter concentrations.

4. **SO2**: This pollutant has a slight downward trend and a seasonal pattern that might correspond with industrial activity cycles, particularly if linked to industries that have seasonal production patterns or fluctuating sulfur content in fuels.

5. **Ozone**: Displays a relatively stable trend but with noticeable seasonality, peaking during warmer months due to its formation from precursors like NOx and VOCs in the presence of sunlight.

c. Abnormalities and Their Possible Causes:

- **Sudden Spikes**: Notable spikes in PM2.5, PM10, and Ozone during certain years could be related to specific events such as wildfires, festivals (e.g., Diwali), or sudden changes in industrial activity.

5. Pune

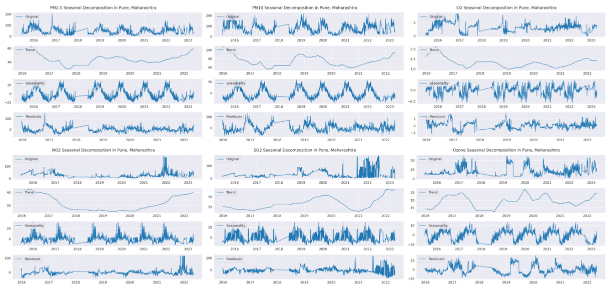

FIG. 5: Seasonal decomposition of air quality parameters (PM2.5, PM10, CO, NO2, SO2, Ozone) measured in Pune, Maharashtra, from 2016 to 2023.

The seasonal decomposition graphs for Pune, Maharashtra, provide insights into the patterns of various air pollutants from 2016 to 2023. Given Pune's status as an industrial hub, these patterns are particularly relevant in understanding the impact of industrial activities along with other urban factors on air quality.

a. General Observations Across All Pollutants:

1. **Seasonality**: Each pollutant shows distinct seasonal patterns, typically with higher levels in colder months due to factors such as increased heating and lower atmospheric mixing heights, and during specific local events or festivals that involve burning or fireworks.

2. **Trend**: There are varying trends across pollutants. Some show a decreasing trend, possibly due to improved regulations and technologies, while others show stability or increases, indicating ongoing challenges in air quality management.

3. **Residuals**: The residuals, which represent fluctuations not explained by seasonality or trend, vary significantly, suggesting the influence of unpredictable factors such as unregulated emissions or changes in meteorological conditions.

b. Specific Observations and Possible Anomalies:

1. **PM2.5**: There's an initially decreasing trend until around 2019, followed by a gradual increase, likely due to changes in local industrial activities or increased vehicular traffic. Seasonal peaks are pronounced during winter months, aligning with agricultural burning practices and cooler, calmer conditions. The dip observed in 2020 reflects the impact of COVID-19 lockdown measures, which temporarily reduced emissions.

2. **PM10**: PM10 levels have shown a slightly increasing trend in recent years, which could be linked to ongoing construction activities and industrial outputs, common in a growing industrial city. Seasonality is similar to PM2.5 with peaks in colder months.

3. **CO**: CO levels exhibit a decreasing trend, likely reflecting improvements in fuel quality and vehicle emission standards. Seasonality is less pronounced compared to particulate matter, indicating sources such as vehicles and industrial combustion are more consistent year-round.

4. **NO2**: NO2 levels have been increasing since 2020, possibly due to a recovery in industrial activities post-lockdown. Seasonal spikes are often observed during winter due to increased heating and stable atmospheric conditions that inhibit pollutant dispersion.

5. **SO2**: SO2 levels show a gradual increase since 2020, potentially due to increased use of fossil fuels in local industries. Anomalies, such as spikes in 2021 and 2022, could be related to specific industrial events or regulatory changes affecting sulfur emissions.

6. **Ozone**: Ozone levels have remained relatively stable but with an upward tick in recent years, hinting at increasing precursors like VOCs and NOx. Seasonality is evident, with peaks during warmer months due to photochemical reactions facilitated by sunlight.

c. Possible Causes for Anomalies:

- **Economic Growth and Industrial Activity**: Fluctuations in pollution levels often align with economic activities. For instance, periods of industrial growth in Pune could explain rising trends in some pollutants despite overall improvements in technology and regulations.
- **Meteorological Conditions**: Changes in weather patterns, such as unseasonably warm or cold weather, can significantly affect pollution levels, as observed in seasonal peaks.
- **Regulatory Changes**: Implementation of stricter or relaxed environmental regulations can lead to noticeable changes in air quality data.

Considering Pune's industrial backdrop, the observed trends and seasonal variations are consistent with the combined effects of urbanization, industrial emissions, vehicular traffic, and regulatory measures. These insights are critical for formulating targeted air quality improvement strategies that address both seasonal patterns and long-term trends in pollutant levels.

6. Patna

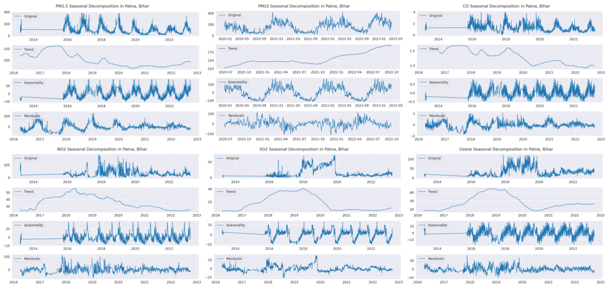

FIG. 6: FIG. 5: Seasonal decomposition of air quality parameters (PM2.5, PM10, CO, NO2, SO2, Ozone) measured in Patna, Bihar, from 2014 to 2023.

a. General Observations Across All Pollutants:

- **Seasonality:** All pollutants show clear seasonal patterns, where certain months repeatedly show higher or lower levels of pollutants.
- **Trend:** There's a general decline in the levels of most pollutants, especially visible in the trend lines for PM2.5, PM10, and NO2, suggesting improved air quality over the years.
- **Residuals:** The residuals (the differences between the observed and predicted values based on trend and seasonality) for most pollutants are quite noisy, indicating other factors influencing pollution levels not captured by just seasonal and trend components.

b. Specific Observations and Possible Anomalies:

- **PM2.5 and PM10:** Both particulate matter measurements peak around the same times each year, suggesting a common source or event causing these spikes.
- **CO and Ozone:** Carbon monoxide and ozone levels don't show as strong a declining trend as particulate matter, with CO showing more fluctuation in recent years.
- **NO2:** There's a sharp decline in levels after 2020, which is an interesting anomaly.
- **SO2:** Sulfur dioxide shows a relatively stable pattern with less pronounced seasonal spikes compared to the other pollutants.

c. Possible Causes for Anomalies:

- **Seasonal Variations:** These are likely linked to climatic conditions such as temperature and wind patterns, which can affect how pollutants are dispersed or formed.
- **Festival and Crop Burning:** High pollution peaks, especially visible in particulate matter graphs, may correlate with local festivals and agricultural practices like crop burning.
- **Lockdown Effects:** The notable decrease in NO2 levels after 2020 could be due to reduced vehicular and industrial activity during COVID-19 lockdowns, as NO2 is primarily emitted from combustion engines and industrial processes.
- **Regulatory Changes:** Gradual improvements in air quality could also be due to stricter environmental regulations and better emission controls in the region.

7. Howrah

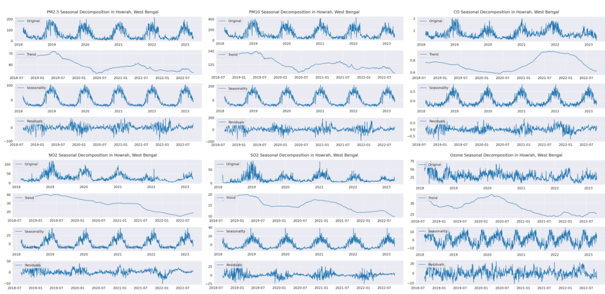

FIG. 7: Seasonal Decomposition of Air Pollutants in Howrah, West Bengal

a. General Observations Across All Pollutants:

- **Trend of Improvement:** Most pollutants, including PM2.5, PM10, CO, NO2, and SO2, exhibit a downward trend over time. This suggests effective regulatory measures and possibly changes in local practices that have contributed to improving air quality.

- **Clear Seasonality:** There is a consistent seasonal pattern across pollutants, with spikes generally occurring during cooler months and other specific periods, influenced by both natural and human factors.

- **Residuals Mostly Stable:** The residuals for most pollutants are fairly stable, indicating that the seasonal and trend components of the models are capturing the majority of the variability in pollutant levels.

b. Specific Observations and Possible Anomalies:

- **PM2.5 and PM10:** Both exhibit similar seasonal peaks during winter, reflecting increased particulate pollution possibly due to heating and stagnant atmospheric conditions.

- **CO:** The biannual peaks in carbon monoxide could be related to changes in temperature, which can affect both the rate of emissions (from heating) and chemical reactions in the atmosphere.

- **NO2:** Shows less seasonality compared to particulate matters, hinting at its sources being less influenced by seasonal climatic changes and more by ongoing human activities such as traffic and industrial output.

- **SO2:** The significant peaks could be strongly linked to specific events like local festivals and industrial operations, which might increase the burning of sulfur-rich materials.

- **Ozone:** Unlike other pollutants, ozone peaks occur mainly in warmer months due to its photochemical formation from precursors like NOx and VOCs under sunlight.

c. Possible Causes for Anomalies:

- **Seasonal Activities and Festivals:** Spikes in PM, SO2, and even CO during festivals such as Diwali and Kali Puja could be due to the extensive use of fireworks and increased vehicular traffic as people participate in festivities.

- **Industrial Emissions:** Given Howrah's industrial base, fluctuations in industrial activity, including changes in production schedules or fuel use, can directly influence levels of SO2, NO2, and particulate matter.

- **Meteorological Conditions:** The increased pollution during winter months can be attributed to lower temperatures and lower wind speeds, which hinder the dispersion of pollutants.

- **Pandemic Lockdowns:** The noticeable dips in pollutant levels during 2020 align with the reduced industrial activity and traffic during the COVID-19 lockdowns, which drastically cut down emissions temporarily.

8. Singrauli

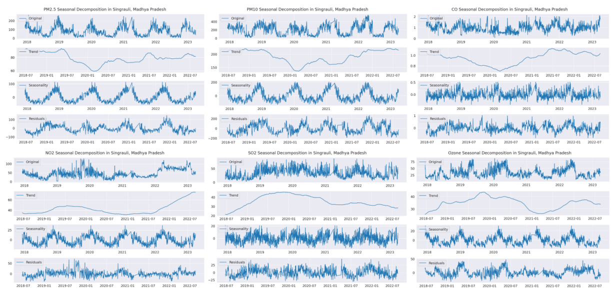

FIG. 8: Seasonal decomposition of air quality data in Singrauli, Madhya Pradesh

a. General Observations Across All Pollutants:

- **Trend:** There's a noticeable downward trend in PM10 and PM2.5, particularly strong in PM10. This could be indicative of effective pollution control measures or changes in industrial activity.

- **Seasonality:** Both PM2.5 and PM10 exhibit clear seasonal patterns, peaking typically around the colder months. This could be due to increased heating activities, and possibly due to crop burning practices in nearby agricultural areas.

- **Anomalies/Peaks:** Notable spikes in PM levels can correspond to specific events such as festivals (like Diwali), where firework usage significantly increases air pollution.

b. Specific Observations and Possible Anomalies:

- **PM2.5 and PM10:** The levels of PM2.5 and PM10 are showing a noticeable downward trend, particularly pronounced in PM10, which could signify effective pollution control measures or changes in industrial activities. These pollutants exhibit clear seasonal patterns, with peaks typically observed during colder months, likely due to increased heating activities and agricultural practices such as crop burning. Additionally, notable spikes in PM levels, especially during events like Diwali, indicate instances of significant air pollution caused by activities such as firework usage.

- **CO:** Carbon monoxide levels are displaying a decreasing trend over the years, potentially linked to advancements in emission standards for vehicles and industries. The fluctuations in CO levels throughout the year may be associated with changes in vehicle usage and heating demand during colder months.

- **NO2:** Nitrogen dioxide levels exhibit a generally stable trend with slight increases in recent years, suggesting a possible rise in vehicular traffic or industrial emissions. Peaks in NO2 levels, particularly during late 2020 and early 2021, may be attributed to variations in traffic and industrial activities, potentially influenced by the resurgence in industrial operations post-initial COVID-19 lockdowns.

- **SO2:** Sulfur dioxide levels demonstrate a downward trend, possibly indicating the adoption of cleaner technologies in power plants and the enforcement of stricter industrial regulations. Similar to other pollutants, SO2 levels exhibit higher concentrations during colder months, primarily due to increased coal consumption for heating and power generation.

- **Ozone:** Ozone levels show a slight decline over the observed period. It typically rises during warmer months due to the presence of sunlight, facilitating photochemical reactions necessary for ozone formation.

c. Possible Causes for Anomalies:

- **Festival Effects:** Notable spikes in PM levels during festivals like Diwali.

- **Seasonal Influences:** Seasonal patterns in PM levels possibly influenced by heating activities and crop burning.

- **Industrial Activity:** Trends in CO, NO2, and SO2 may be influenced by changes in industrial output and emission standards.

- **Pandemic Impact:** Fluctuations during 2020 and slightly beyond likely due to COVID-19 lockdowns reducing industrial activity and traffic.

- **Local Events and Practices:** Fire usage during festivals, agricultural burning practices, and changes in industrial output can create visible spikes or drops in the data trends.

9. Chennai

a. General Observations Across All Pollutants :

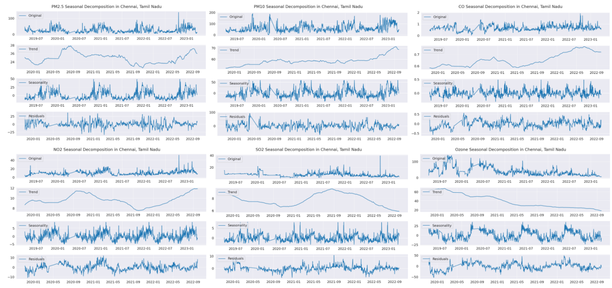

FIG. 9: Seasonal decomposition of air quality parameters (PM2.5, PM10, CO, NO2, SO2, Ozone) measured in Chennai, Tamil Nadu, from 2016 to 2023.

- **Trend:** Most pollutants show a general downward trend over the observed period, suggesting improvements in air quality management and possibly effective pollution control measures being implemented.

- **Seasonality:** There are clear seasonal variations for each pollutant, with peaks generally occurring at specific times of the year. This indicates that certain seasonal factors strongly influence pollutant levels.

- **Residuals**: The residuals, which represent the fluctuations not explained by the seasonal and trend components, are relatively stable across most pollutants, indicating that the models used are capturing most of the variability effectively.

b. Specific Observations and Possible Anomalies:

- **PM2.5 and PM10**: Both show similar seasonal patterns with peaks typically occurring during specific months. The trends indicate a gradual decrease over time, which is a positive sign of air quality improvement.

- **CO (Carbon Monoxide)**: The trend for CO shows a very slow decline and has less pronounced seasonal peaks compared to particulate matter, which might indicate different emission sources or control measures affecting it.

- **NO2 (Nitrogen Dioxide)**: Shows a noticeable downward trend, particularly after early 2020, which might be linked to reduced vehicular traffic or industrial activities.

- **SO2 (Sulfur Dioxide)**: The levels of SO2 are relatively low and show minor seasonal fluctuations, suggesting effective control and possibly the switch to cleaner fuel sources in the region.

- **Ozone**: Ozone levels display a notable summer peak, which is typical due to the photochemical reactions driven by increased sunlight during these months.

c. Possible Causes for Anomalies:

- **Meteorological Conditions:** Seasonal peaks, particularly for PM2.5, PM10, and ozone, can be influenced by weather conditions such as temperature, sunlight, and humidity, which affect the formation and dispersion of pollutants.

- **Festivals and Public Events:** Events that involve burning of fireworks or increased vehicular traffic can cause temporary spikes in pollutants like PM10 and SO2.

- **Regulatory Changes:** Implementation of stricter pollution control norms can lead to the observed declines in NO2 and CO levels.

- **COVID-19 Pandemic:** The noticeable dips in NO2 and other pollutants' levels in early 2020 can be directly linked to the lockdowns imposed during the COVID-19 pandemic, which drastically reduced traffic and industrial activities.

10. Jodhpur

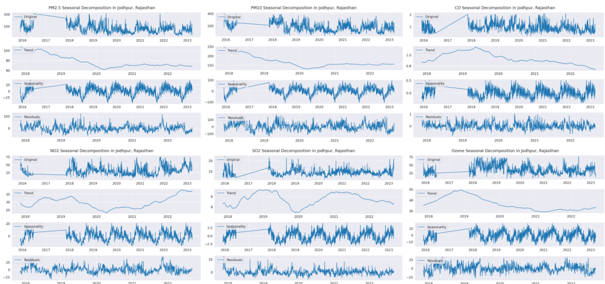

FIG. 10: Seasonal Decomposition of Air Pollutants in Jodhpur, Rajasthan

a. General Observations Across All Pollutants in Jodhpur, Rajasthan:

- **Trend:** Most pollutants, including PM2.5, PM10, and SO2, show a downward trend over the observed period, suggesting improvements in air quality or effective pollution control measures.

- **Seasonality:** Clear seasonal patterns are observed in pollutants like CO, NO2, and Ozone, indicating that their levels are influenced by seasonal factors such as temperature changes, festival periods, or industrial activities.

- **Residuals:** The residuals, which represent fluctuations not explained by the seasonal and trend components, are relatively stable across most pollutants, indicating effective model performance.

b. Specific Observations and Possible Anomalies:

- **PM2.5 and PM10:** Both particulate matters show similar seasonal peaks, possibly correlated with specific weather conditions like low wind during winter months that inhibit the dispersion of particulates. The declining trend in PM2.5 and stability in PM10 post-2018 suggest effective air quality control measures or changes in local industrial practices.

- **CO (Carbon Monoxide):** Consistent seasonal patterns may be due to increased combustion from heating during cooler months, with the downward trend suggesting improvements in combustion efficiency or emissions regulations.

- **NO2 (Nitrogen Dioxide):** An increasing trend from around 2020 could be due to a rise in industrial activity or traffic, with consistent seasonal variations indicating steady sources of NO2 emissions.

- **SO2 (Sulfur Dioxide):** A decreasing trend might reflect successful implementation of SO2 control technologies or a shift towards cleaner industrial processes. Seasonal spikes could be linked to specific, seasonal industrial activities.

- **Ozone:** A significant decreasing trend could be due to reductions in precursor emissions like NOx and VOCs, or climatic conditions affecting ozone formation. Seasonal patterns may reflect variations in sunlight and temperature, critical for ozone generation.

c. Possible Causes for Anomalies:

- **Geographic and Climatic Factors:** Jodhpur's arid climate and occasional dust storms could significantly impact levels of particulate matter.

- **Cultural Practices and Festivals:** Events like Diwali may lead to temporary spikes in particulate matter and SO2 due to the burning of firecrackers.

- **Regulatory Changes:** Ongoing implementation of stricter air quality regulations might be contributing to the overall downward trend in pollutants.

- **Pandemic Impact:** Notable reductions in pollutant levels during 2020 could be tied to the COVID-19 lockdowns, which saw reduced traffic and industrial activity, particularly affecting NO2 and CO levels.

11. Bengaluru

a. General Observations Across All Pollutants in Bengaluru, Karnataka:

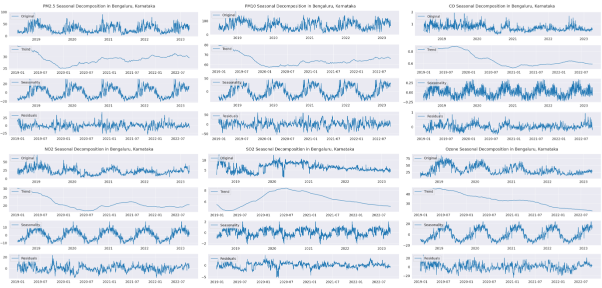

FIG. 11: Seasonal Decomposition of Air Pollutants in Bengaluru

- **Overall Trend of Decline:** Most pollutants, including PM2.5, PM10, CO, and SO2, show a general downward trend over the observed period.
- **Seasonal Variations:** Clear seasonal patterns are evident across all pollutants, with peaks often occurring around the middle of the year.
- **Stable Residuals Indicating Effective Model Fit:** The residuals for most pollutants do not show significant patterns, suggesting that the seasonal and trend components of the decomposition models are capturing the majority of variations in pollution levels effectively.

b. Specific Observations and Possible Anomalies:

- **PM2.5 and PM10:** Both pollutants show declining trends with mid-year seasonal peaks. The sharper seasonal variation in PM10 might indicate it is more affected by local sources like construction dust, which can vary more dramatically across the year compared to PM2.5. Being an IT hub, Bengaluru's industrial emissions might be lower, but urban construction related to IT infrastructure growth could contribute to PM10 levels.
- **CO (Carbon Monoxide):** The gradual decline in CO could be due to better management of sources like traffic emissions, reflecting improvements in vehicle technology and possibly increased use of public transport or remote working practices encouraged by IT companies.
- **NO2 (Nitrogen Dioxide):** The slight downward trend, stabilizing around 2021, might reflect effective emissions control associated with a modern vehicle fleet and reduced industrial emissions, possibly due to the dominance of the IT sector over more polluting industries.
- **SO2 (Sulfur Dioxide):** The long-term decline in SO2 is likely due to stricter emissions controls and the transition to cleaner technologies, which are often adopted faster in technology-driven cities like Bengaluru.
- **Ozone:** The slight decreasing trend might relate to reductions in NOx and VOCs, crucial precursors for ozone formation.

c. Possible Causes for Anomalies:

- **Meteorological Conditions:** Seasonal weather patterns, particularly temperature and wind, can greatly influence pollution levels and their dispersion in this region.
- **Urban Development and IT Sector Expansion:** As Bengaluru continues to expand its IT infrastructure, construction and increased vehicular traffic in certain areas could influence pollution levels, particularly PM10 and NO2.
- **Local Festivals and Public Events:** Events that involve large gatherings and the use of vehicles or fireworks can lead to short-term spikes in pollutants like PM10 and NO2.
- **Global and Local Events:** The COVID-19 pandemic, resulting in lockdowns and a significant reduction in activities, likely contributed to the dips observed in 2020 across most pollutants. The IT sector's quick shift to remote working may have also played a role in reducing traffic and emissions during this period.

12. Ahmedabad

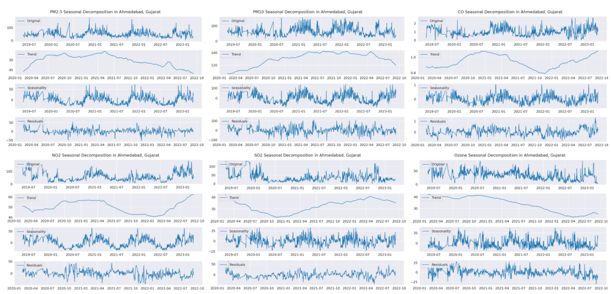

FIG. 12: Seasonal Decomposition of Air Pollutants in Ahmedabad

a. General Observations Across All Pollutants:

- **Overall Decline in Pollution Levels:** Most pollutants, including PM2.5, PM10, CO, and SO2, display a general downward trend. This suggests improvements in air quality, potentially due to successful pollution control measures or changes in urban activities, possibly influenced by the pandemic lockdowns.
- **Seasonal Variations:** Clear seasonal patterns are evident across all pollutants, with peaks often occurring during the winter months. This trend may

be exacerbated by meteorological conditions that inhibit pollutant dispersion, such as temperature inversions common in northern Indian cities during winter.

- **Stable Residuals Indicating Model Fit:** The residuals for most pollutants do not show significant patterns, suggesting that the seasonal and trend components of the decomposition models are capturing the majority of variations in pollution levels effectively.

b. Specific Observations and Possible Anomalies:

- **PM2.5 and PM10:** Both show declining trends with sharper seasonal peaks during the winter, possibly due to increased heating requirements and vehicular use. Being a metro city away from the coast, Ahmedabad might experience less dispersion of these particles due to the lack of sea breezes, which can help clear air pollutants in coastal cities.
- **CO:** The slight decline until early 2021 and subsequent stabilization could be linked to changes in traffic patterns and industrial activities, especially during COVID-19 lockdowns. The annual spikes in CO could be associated with the increased combustion from heating during colder months.
- **NO2:** There is an increasing trend from mid-2020, possibly indicating a rebound in industrial and vehicular emissions after initial pandemic restrictions. The peaks during winter could be due to increased heating and reduced atmospheric mixing.
- **SO2:** Shows a decreasing trend, suggesting effective control measures or changes in operations of sulfur-emitting industries. The less pronounced seasonality compared to particulate matters indicates different emission sources or more effective control mechanisms.
- **Ozone:** Exhibits a slightly decreasing trend but remains relatively stable. The peaks during warmer months are due to its formation from sunlight and precursor chemicals from traffic and industry.

c. Possible Causes for Anomalies:

- **Geographic and Climatic Factors:** Being a metro city away from the coastal region, Ahmedabad's high temperatures and abundant sunlight contribute to the formation of ozone. The lack of sea breezes leads to poorer dispersion of air pollutants.
- **Cultural Festivals:** Events like Diwali can cause short-term spikes in PM levels due to the extensive use of fireworks.
- **Pandemic Impact:** The notable declines in pollutant levels at the onset of the COVID-19 pandemic in early 2020 highlight the significant impact of reduced human and industrial activities.

13. Rajamahendravaram

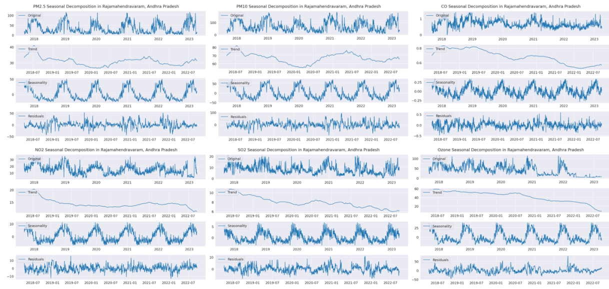

FIG. 13: Seasonal Decomposition of Air Pollutants in Rajamahendravaram

a. General Observations Across All Pollutants:

- **Trend:** Both particulate matter (PM2.5 and PM10) and carbon monoxide (CO) display a general declining trend over the observed period. This suggests effective air quality improvements, potentially due to enhanced regulatory measures, shifts in industrial activity, or improvements in traffic management and vehicle emissions controls.
- **Seasonality:** There are clear seasonal patterns for all pollutants, with peaks often occurring during specific months. These variations can be influenced by weather conditions such as temperature and sunlight, which affect both the dispersal and chemical reactions of pollutants in the atmosphere.
- **Residuals:** For most pollutants, the residuals do not show significant patterns, suggesting that the seasonal and trend components of the decomposition models are effectively capturing the majority of variations in pollution levels.

b. Specific Observations and Possible Anomalies:

- **PM2.5 and PM10:** These particulate matters show declining trends but with seasonal peaks, which might be exacerbated by local events such as festivals or agricultural practices like crop burning.
- **CO (Carbon Monoxide):** The gradual decline could be attributed to better fuel quality and stricter emissions controls in urban vehicles, reflecting ongoing improvements in local transportation policies.
- **NO2 (Nitrogen Dioxide):** The increasing trend from 2021 could suggest a rise in industrial activities or vehicular traffic, possibly linked to local economic growth or development projects.
- **Ozone (O3):** A decrease in ozone levels might indicate effective control of precursor emissions (NOx and VOCs) or changes in meteorological conditions that are less favorable for ozone formation.

c. Possible Causes for Anomalies:

- **Meteorological Influences:** Seasonal weather changes significantly impact the formation and dispersal of pollutants, with cooler temperatures and lower wind speeds in winter potentially trapping pollutants closer to the ground.
- **Local and Seasonal Activities:** Festivals like Diwali, which involve the burning of fireworks, and agricultural practices such as crop burning can lead to temporary spikes in PM levels.
- **Economic and Infrastructural Developments:** Increases in NO2 levels might be linked to infrastructural developments that boost vehicular and industrial activities.
- **Pandemic Impact:** The significant reductions in pollutant levels during the early stages of the COVID-19 pandemic demonstrate the impact of reduced human activity, including industrial operations and traffic, on air quality.

B. Mean, Variance and ACF plots

a. Mean and Variance

PM2.5 and PM10 (Particulate Matter)

- **Decreasing Mean:** Indicates improvements in air quality, likely due to effective pollution reduction measures.
- **Increasing Mean:** Suggests worsening air quality, possibly due to increased emissions.
- **Decreasing Variance:** Shows that the levels of particulates are becoming more uniform, indicating stable environmental conditions or consistent control measures.
- **Increasing Variance:** Means there are greater fluctuations in particulate levels, possibly due to variable emission sources or changing environmental conditions.

CO (Carbon Monoxide)

- **Decreasing Mean:** Suggests reductions in emissions, possibly from improved technology or regulatory compliance in traffic and industrial sources.
- **Increasing Mean:** Indicates more emissions, potentially from increased vehicle usage or industrial activity.
- **Decreasing Variance:** Points to consistent CO levels, possibly due to uniform traffic flow or steady industrial operations.
- **Increasing Variance:** Indicates fluctuating CO levels, which could be caused by varying traffic patterns or industrial outputs.

NO2 (Nitrogen Dioxide)

- **Decreasing Mean:** Generally indicates improving air quality, often due to decreased emissions from vehicles and industry.
- **Increasing Mean:** May reflect increased emissions from industrial activities or higher vehicle traffic.
- **Decreasing Variance:** Suggests more stable emissions, likely due to effective emission controls.
- **Increasing Variance:** Means more erratic NO2 levels, potentially due to inconsistent industrial activities or traffic conditions.

SO2 (Sulfur Dioxide)

- **Decreasing Mean:** Typically reflects improved air quality, often due to reduced emissions from industrial sources like power plants.
- **Increasing Mean:** Suggests heightened emissions, potentially leading to poorer air quality.
- **Decreasing Variance:** Indicates that SO2 emissions are becoming more consistent, likely due to steady industrial operations.
- **Increasing Variance:** Points to more irregular SO2 levels, which may result from fluctuating industrial outputs.

Ozone

- **Decreasing Mean:** Could be due to less intense sunlight or lower levels of ozone precursors (like VOCs and NOx).
- **Increasing Mean:** Suggests more intense sunlight or an increase in the chemicals that form ozone.
- **Decreasing Variance:** Means that the conditions for ozone formation are becoming more stable.
- **Increasing Variance:** Indicates more variable conditions affecting ozone formation, such as changes in weather or emissions of precursor chemicals.

b. Auto Correlation Plot

- **High Initial Autocorrelation:** Most plots show a very high initial autocorrelation at lag 0 (which is always 1, as it's the correlation of the series with itself), with a rapid drop-off as the lag increases. This pattern suggests that the values of pollutants are highly dependent on their immediate past values, indicating persistence in air quality levels over short intervals.

- **Gradual Decline:** The autocorrelation values generally exhibit a gradual decline rather than an abrupt change, suggesting that the air quality data possesses a smooth transition over time without sudden shifts in levels, which is typical for environmental data.

- **Seasonal Patterns:** The consistent peaks at regular intervals in some plots (such as ozone and PM10) could indicate seasonal patterns, where air quality parameters are influenced by systematic seasonal factors. For instance, higher levels of ozone are often found during warmer months due to increased sunlight driving photochemical reactions.

- **Negative Autocorrelation:** Some plots show negative autocorrelation at higher lags, suggesting possible over-differencing or perhaps cyclic variations that aren't captured by simpler models.

- **Long-term Dependence:** The slow decay of the autocorrelation function in many pollutants suggests long-term dependence, meaning past values have a lingering effect on future values over an extended period. This might be due to sustained pollution sources or slow-changing meteorological conditions.

C. Heat Map

Particulate Matter:

- **PM2.5 and PM10:** The PM10 levels are particularly high across all cities, with Ghaziabad showing the highest concentration at 235.7 $\mu g/m^3$. This is followed by Muzaffarnagar and Patna with levels above 150 $\mu g/m^3$. For PM2.5, Ghaziabad again leads, indicating a serious concern regarding airborne particulate pollution in these areas.

Gaseous Pollutants:

- **Carbon Monoxide (CO):** Levels are comparatively lower across the board, with most cities averaging around 1 $\mu g/m^3$. The highest recorded in this category is relatively low, highlighting lesser concern about CO in these regions compared to particulate matter.

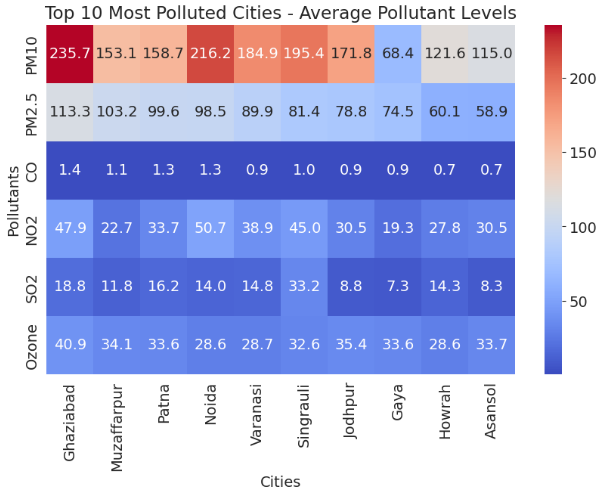

FIG. 14: Heat Map of Top 10 Most Polluted Cities

- **Nitrogen Dioxide (NO2):** Shows significant variation, with Singrauli exhibiting notably high levels at 50.7 $\mu g/m^3$. Other cities like Ghaziabad, Varanasi, and Jodhpur also show elevated NO2 levels, suggesting traffic and industrial activities as probable contributors.

- **Sulphur Dioxide (SO2):** In Singrauli stands out with the highest levels at 33.2 $\mu g/m^3$, which could be attributed to industrial emissions, particularly from power plants. Other cities generally show moderate levels.

- **Ozone (O3):** The concentration is relatively high in all cities but varies less dramatically between them, with Ghaziabad and Asansol showing the highest readings, over 33 $\mu g/m^3$. This pollutant is a secondary air pollutant formed by the reaction of sunlight with other air pollutants like NOx and volatile organic compounds.

D. Final Result

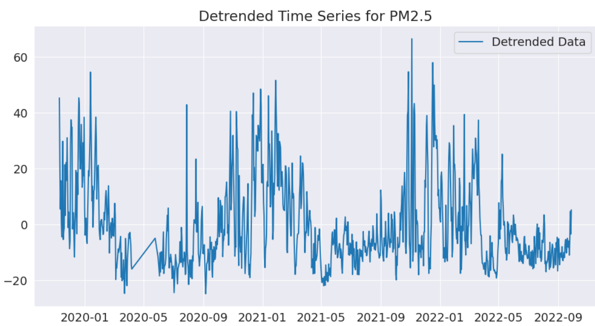

FIG. 15: Detrended Time series for PM2.5

In this project, we aimed to model and forecast different pollutant levels using a time series analysis approach. The key steps in our analysis included detrending the data, fitting a SARIMA model, and forecasting future values.

Detrending the Time Series Data Initially, we focused on detrending the data to remove long-term trends and make the time series more stationary. This was achieved by subtracting a rolling mean with a 365-day window from the original data. The resulting detrended data fluctuated around zero, indicating that the trend component had been effectively removed. This step is crucial as it simplifies the underlying data structure, making it easier for our model to capture the true patterns.

Fitting the SARIMA Model We then fitted a Seasonal ARIMA (SARIMA) model with the parameters $(1, 1, 1) \times (1, 1, 1, 12)$. This model configuration suggests that both seasonal and non-seasonal differences were necessary to stabilize the series. The statistical significance of the model's coefficients, indicated by very low p-values, confirms that the chosen parameters were meaningful. However, a warning about potential non-stationarity implies that some aspects of the data may not have been fully accounted for, which could affect the model's accuracy.

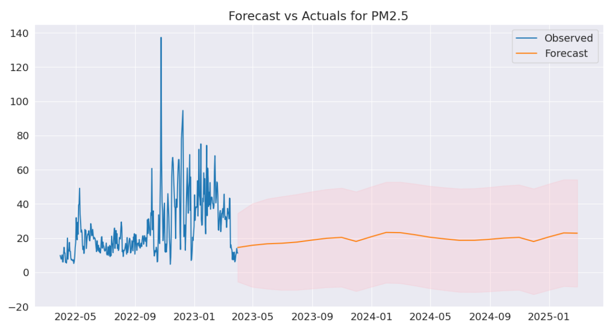

FIG. 16: Forecast VS Actual for PM2.5

Forecasting Future Values The forecasting plot revealed how the model predicts future PM2.5 levels. The forecasts seemed to align well with the recent trends observed in the data, although the confidence intervals widened over time. This widening indicates increasing uncertainty in the predictions as we project further into the future.

IV. CONCLUSION

This project analyzed air quality trends from 2010 to 2023 across different regions in India using ARIMA and SARIMAX models. Our study highlighted a significant difference in air pollution between heavily populated cities and rural areas. Cities like Delhi and Kanpur experienced higher pollution levels mainly due to dense traffic and industrial activities, while rural areas generally had better air quality except during specific events like crop burning.

Understanding these differences helps us pinpoint where and when to implement targeted environmental policies. For instance, North Indian cities suffer more in winter due to climatic conditions that trap pollutants close to the ground, suggesting a need for seasonal pollution management strategies. In contrast, cities in South India benefit from milder weather and often manage better because of their climate and stricter pollution control policies.

During the COVID-19 lockdowns, we saw a clear drop in pollution, showing how reducing human activities can quickly improve air quality. This demonstrates the potential benefits of changes in policies and public behavior on environmental health. However, our models sometimes did not predict sudden pollution spikes caused by unexpected events such as festivals or sudden industrial activity. These limitations showed that while our forecasting models are useful, they need to be adjusted to better handle unexpected changes.

In conclusion, using ARIMA and SARIMAX models helped us understand and predict general trends in air quality. These models were good at showing long-term trends and seasonal changes. But, we found that they need to be better at predicting sudden changes in air quality caused by unusual events. Future work could look into using more complex models or adding more data to improve accuracy. This project has shown us the importance of using time series analysis to help make decisions about environmental policies and protect public health.

V. ADDITIONAL RESOURCES

Below are additional resources that provide more insight into our analysis:

- **Video Presentation:** For a detailed explanation and walkthrough of our analysis, view our video on YouTube at Video Presentation.
- **Google Colab Notebook:** The code used in this analysis is available on Google Colab. Access it here: Google Colab Notebook.

[1] Abhishek-S-Jha. n.d. "Time Series Air Quality Data of India (2010-2023)." Kaggle. Accessed May 9, 2024. https://www.kaggle.com/datasets/abhisheksjha/time-series-air-quality-data-of-india-2010-2023.

[2] Artley, Brendan. 2022. "Time Series Forecasting with ARIMA , SARIMA and SARIMAX." Towards Data Science, June 27, 2022. https://towardsdatascience.com/time-series-forecasting-with-arima-sarima-and-sarimax-ee61099e78f6.