HANOI UNIVERSITY OF SCIENCE AND TECHNOLOGY
SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

PROJECT REPORT

ANALYSIS OF AIR POLLUTION LEVELS AND THE IMPACT OF METEOROLOGICAL FACTORS IN MAJOR VIETNAMESE CITIES BASED ON ONLINE DATA

Course: Project 2

Supervisor: PhD. Tran Nguyen Ngoc

Author	Student ID
Pham Nguyen Hai Nhi	20225549

July, 2025

Contents

List of Figures

List of Tables

1 Abstract

In the context of escalating air pollution, a growing concern in Vietnam's major cities, understanding the distribution characteristics and influencing factors of air quality is crucial for effective environmental management and public health protection.

This study assesses air pollution levels in large urban areas of Vietnam by analyzing key indicators such as AQI, $PM_{2.5}$, PM_{10}, CO, NO_2, O_3, and SO_2. The research aims to identify spatio-temporal distribution trends of air pollution and the relationship between pollutants and meteorological factors including temperature, humidity, atmospheric pressure, and wind speed. Data were collected from the Ministry of Natural Resources and Environment's online platform and integrated with corresponding meteorological data. The methodology encompasses descriptive statistics, spatio-temporal trend analysis, and comparative assessments across geographical regions.

Results indicate a strong correlation between air pollution and traffic as well as industrial activities. Notably, the Northern region consistently recorded higher average pollution levels than the South and exhibited distinct seasonal patterns. Furthermore, the study reveals that meteorological conditions conducive to temperature inversions significantly contribute to increased pollutant concentrations at specific times.

These findings provide essential data and insights, serving as a foundation for developing and implementing effective air pollution control policies in Vietnam.

2 Introduction

Air pollution has become one of the most urgent environmental and public health challenges globally, especially in developing countries experiencing rapid urbanization and industrialization. The World Health Organization (WHO) warns that air pollution causes millions of premature deaths annually and contributes to numerous serious respiratory, cardiovascular, and cancerous diseases (WHO, 2021). In Vietnam, with continuous economic growth and increasing urban populations, air pollution has become progressively severe, particularly in major cities and key industrial zones (UNEP, 2019).

To address this issue, a clear understanding of air pollution's characteristics, causes, and dynamics is essential. Previous studies indicate that air quality is influenced by complex interactions between anthropogenic emission sources (like traffic, industry, construction) and meteorological factors (such as temperature, humidity, pressure, and wind speed), which govern the formation, dispersion, and accumulation of pollutants in the atmosphere (Seinfeld Pandis, 2016). However, a comprehensive and detailed analysis of the air pollution landscape across Vietnam's regions, particularly clarifying the role of specific meteorological factors like temperature inversions, remains limited.

This study aims to provide a more holistic and in-depth view of air pollution in Vietnam. We focus on analyzing time series data of Air Quality Index (AQI) and key pollutants ($PM_{2.5}$, PM_{10}, CO, NO_2, O_3, and SO_2) from the Ministry of Natural Resources and Environment's data, while integrating relevant meteorological factors. The primary objectives of this research are to:

- Analyze the trends and fluctuations of AQI across space (between regions and monitoring stations) and time (by hour, day of the week, and season).
- Assess the correlation between AQI and major pollutant concentrations, as well as the influence of meteorological factors.
- Identify and analyze the role of temperature inversions in increasing pollutant concentrations, particularly in sensitive areas.

The findings from this study are expected to provide crucial scientific evidence and data, contributing to the development of more effective air quality management strategies, supporting sustainable urban planning, and enhancing public awareness regarding the impacts of air pollution in Vietnam.

3 Background Knowledge

Research on air quality is an increasingly critical field globally, especially in rapidly urbanizing and industrializing developing countries like Vietnam. High levels of air pollution directly impact public health and negatively affect the environment, economy, and overall quality of life. To better understand the background knowledge of this study, we need to examine concepts, key pollutants, emission sources, meteorological factors, and the air pollution situation in Vietnam.

3.1 Concepts and Air Quality Index (AQI)

Air quality is a measure of how clean or polluted the air is, assessed based on the concentration of various pollutants in the atmosphere.

The Air Quality Index (AQI) is a standard tool widely used to convey air quality information to the public. AQI converts the concentrations of different air pollutants (such as $PM_{2.5}$, PM_{10}, CO, NO_2, O_3, and SO_2) into a single, easily understandable number, accompanied by color-coded categories and health recommendations. A higher AQI value indicates greater air pollution and higher health risks. In Vietnam, AQI calculation often adheres to national standards, typically referencing guidelines from the U.S. Environmental Protection Agency (US EPA) or the World Health Organization (WHO).

AQI Calculation Formula:

$$\mathrm{AQI}_i = \frac{C_i - C_{\mathrm{low}}}{C_{\mathrm{high}} - C_{\mathrm{low}}} \times (\mathrm{AQI}_{\mathrm{high}} - \mathrm{AQI}_{\mathrm{low}}) + \mathrm{AQI}_{\mathrm{low}}$$

Where:

- AQI_i: AQI value for pollutant i
- C_i: Measured concentration of pollutant i
- C_{low}: Lower concentration breakpoint of the range containing C_i
- C_{high}: Upper concentration breakpoint of the range containing C_i
- $\mathrm{AQI}_{\mathrm{low}}$: AQI corresponding to C_{low}
- $\mathrm{AQI}_{\mathrm{high}}$: AQI corresponding to C_{high}

The overall Air Quality Index (AQI) at a given time is determined as the maximum AQI among the measured pollutants:

$$\mathrm{AQI}_{\mathrm{overall}} = \max\left(\mathrm{AQI}_{\mathrm{PM}_{2.5}}, \mathrm{AQI}_{\mathrm{PM}_{10}}, \mathrm{AQI}_{\mathrm{O}_3}, \mathrm{AQI}_{\mathrm{CO}}, \mathrm{AQI}_{\mathrm{NO}_2}, \mathrm{AQI}_{\mathrm{SO}_2}\right)$$

3.2 Key Air Pollutants

This study focuses on the following key pollutants, each with distinct origins and impacts:

- **Particulate Matter ($PM_{2.5}$ and PM_{10}):**

  - **$PM_{2.5}$:** Particles with diameters $\leq 2.5\,\mu$m. These can penetrate deep into the lungs and enter the bloodstream, causing respiratory and cardiovascular diseases, and even cancer.
  - **PM_{10}:** Particles with diameters $\leq 10\,\mu$m. Though larger than $PM_{2.5}$, they can still affect the upper respiratory tract and cause health problems.
  - *Sources:* Combustion of fuels (traffic, industry, biomass burning), construction, industrial processes, road and soil dust.

- **Carbon Monoxide (CO):**
  - A colorless, odorless, and tasteless gas resulting from incomplete combustion of fossil fuels.
  - *Sources:* Mainly from motor vehicles, open burning, and wildfires.
  - *Impacts:* Reduces blood's oxygen-carrying capacity, causing headaches, dizziness, and at high levels, can be fatal.

- **Nitrogen Dioxide (NO_2):**
  - A reddish-brown gas with a strong odor, produced from high-temperature combustion.
  - *Sources:* Traffic exhaust, power plants, and industrial activities.
  - *Impacts:* Irritates the respiratory system and aggravates conditions like asthma.

- **Tropospheric Ozone (O_3):**
  - A secondary pollutant formed by reactions between NO_x and VOCs under sunlight.
  - *Sources:* Not directly emitted; formed from precursor pollutants.
  - *Impacts:* Affects lung function and especially harmful to sensitive groups such as children, the elderly, and those with respiratory diseases.

- **Sulfur Dioxide (SO_2):**
  - A colorless gas with a pungent smell, produced by burning sulfur-rich fuels.
  - *Sources:* Coal power plants, industry, oil and gas production.
  - *Impacts:* Causes respiratory irritation and contributes to acid rain.

3.3 Major Air Pollution Emission Sources in Vietnamese Cities

Air pollution emission sources in Vietnam, especially in large urban areas, can be categorized as follows:

- **Mobile Sources:** Primarily road vehicles such as motorbikes, cars, and buses. With an extremely high density of motorbikes, this is a major source of pollutants like $PM_{2.5}$, PM_{10}, CO, NO_2, and volatile organic compounds (VOCs).

- **Stationary Sources:** Include industrial factories (e.g., cement production, thermal power plants, chemical and steel industries), craft villages, and residential buildings. Emissions are often caused by burning fossil fuels such as coal or wood for industrial or domestic use.

- **Area Sources:** Comprise smaller-scale but widespread activities. Include household waste burning, agricultural biomass burning (e.g., rice straw), and dust from construction sites.

3.4 Influence of Meteorological Factors on Air Quality

Meteorological factors play a pivotal role in the dispersion, accumulation, or transformation of pollutants:

- **Temperature:** High temperatures generally promote convection, allowing pollutants to disperse upward and dilute. However, elevated temperatures can also accelerate chemical reactions that generate secondary pollutants such as ozone (O_3).
- **Humidity:** High humidity can cause fine particulate matter (e.g., $PM_{2.5}$) to absorb water, increasing their size and deposition rate. On the other hand, it may enhance surface chemical reactions, potentially increasing pollutant concentrations. Fog and mist, often associated with high humidity and stagnant air, contribute to pollution build-up.
- **Atmospheric Pressure:** High-pressure systems are associated with stable weather, low wind speeds, and downward air movement (subsidence). This results in atmospheric stability that inhibits vertical mixing, trapping pollutants near the surface.
- **Wind Speed and Direction:** Strong winds help disperse pollutants away from emission sources. Wind direction determines the transport path of pollutants. Low or stagnant wind conditions cause pollutant accumulation locally.
- **Temperature Inversion Phenomenon:** In a temperature inversion, air temperature increases with altitude in a certain atmospheric layer. This inversion layer acts as a "lid," preventing the usual upward movement of cooler air and pollutants. Pollutants get trapped near the surface, leading to prolonged high pollution levels. Common during nighttime and early mornings, especially in winter and dry seasons, or in valleys and basins surrounded by mountains.

3.5 Air Pollution Situation in Vietnam

Vietnam has been and continues to face increasing air pollution challenges, particularly in major cities like Hanoi and Ho Chi Minh City. Reports from the Ministry of Natural Resources and Environment (MONRE) and international organizations (WHO, UNEP) consistently indicate that $PM_{2.5}$ concentrations frequently exceed permissible limits, significantly impacting public health.

- **Seasonal Characteristics:** Pollution is typically more severe during the winter and dry seasons (from around October to April of the following year) due to unfavorable meteorological conditions (less rain, low winds, high likelihood of temperature inversions). Summer and rainy seasons (from May to September) generally see better air quality thanks to rain washout and stronger winds.
- **Regional Characteristics:** Northern Vietnam, especially the Red River Delta, often experiences higher pollution levels than the South due to the influence of the Northeast monsoon, cold and dry climatic conditions, and the concentration of

industrial activities, traffic, and biomass burning. Southern Vietnam, with its hot and humid tropical monsoon climate, generally has better dispersion conditions, although it still faces pressure from traffic and industry.

- **Hourly Characteristics:** Pollutant concentrations tend to increase during peak traffic hours (early morning and late afternoon/evening) and can remain high throughout the day in areas affected by industrial and construction activities.

This study builds upon this solid foundation of knowledge to provide a detailed analysis, presenting evidence of the influence of meteorological and anthropogenic factors on air quality, specifically clarifying the role of temperature inversions. From this, it proposes practical applications and future research directions suitable for the Vietnamese context.

4 Materials and Methods

4.1 Data Collection for Air Quality and Weather

To analyze air quality in major urban areas of Vietnam, the research team developed an automated data collection tool (crawler) using Python. This tool integrates the requests library to access data from Application Programming Interfaces (APIs) and the pandas library for processing, formatting, and storing data in tabular form (as .csv files).

4.1.1 Data Sources

Data was collected from two main sources:

- **Air Quality Data**: Data was retrieved from the Vietnam Electronic Environmental Information Portal, under the Ministry of Natural Resources and Environment, via Envisoft's API system at the address:

  ```
  https://envisoft.gov.vn/eos/services/call/json/qi_detail_for_eip?
     station_id=<station_id>
  ```

  The data includes indices for: `AQI`, $\mathtt{PM}_{2.5}$, $\mathtt{PM}_{10}$, `CO`, $\mathtt{NO}_2$, $\mathtt{O}_3$, and $\mathtt{SO}_2$.

- **Weather Data**: Was retrieved from the OpenWeatherMap platform via API in the format:

  ```
  https://api.openweathermap.org/data/2.5/weather?id=<city_id>&appid=<
     API_KEY>.
  ```

  The data includes: `temperature`, `humidity`, `atmospheric pressure`, `wind speed`, and geographical location (`longitude`, `latitude`).

4.1.2 Monitoring Station List

The system is set up to collect data from 12 air quality monitoring stations, with a predefined list of `station_id` codes. Correspondingly, weather data is gathered from 12 associated locations using their respective `city_ids`. These locations are primarily in major urban centers like Hanoi and Ho Chi Minh City, along with several other key industrial hubs.

4.1.3 Data Collection and Storage Method

The data collection process is automated and follows a three-step cycle:

1. **Retrieve Air Quality Data**
   For each `station_id`, the program sends a request to the `Envisoft` API. It then extracts specific data fields: `Timestamp`, $\mathtt{PM}_{2.5}$, $\mathtt{PM}_{10}$, `CO`, $\mathtt{NO}_2$, $\mathtt{O}_3$, $\mathtt{SO}_2$, and the station `Name`.

2. **Retrieve Weather Data**
   For each corresponding `city_id`, the program fetches current weather data from OpenWeatherMap. It collects the following parameters: `Temperature`, `Humidity`, `Pressure`, `Wind Speed`, `Latitude`, and `Longitude`.

3. **Merge and Store Data**
   Data from both sources are combined into a unified record and saved into a CSV file named `air_quality_weather_data.csv`. Before saving, the system checks for duplicates by comparing the `Name` and `Timestamp` to prevent overwriting or creating redundant records.

4.1.4 Automated Collection and Frequency

The program is set to execute periodically once every hour via GitHub Actions. There's a 3-second pause between each data retrieval request to reduce the load on the API servers and avoid violating access policies. This process helps build a continuous and near real-time data stream.

4.1.5 Error Handling and Robustness

The tool includes an integrated exception handling mechanism to ensure stability when encountering connection errors, response errors, or malformed data. If weather data can't be retrieved, the system still records the AQI data, leaving the weather fields blank. All errors during the collection process are logged into the system journal for review and troubleshooting.

4.1.6 Output Data Structure

The output `.csv` data file includes the following fields:

Data Field	Meaning
Timestamp	Time of measurement
Name	Name of the monitoring station
Latitude	Geographical latitude of the station
Longitude	Geographical longitude of the station
AQI	Air Quality Index
$PM_{2.5}$	Fine particulate matter index
PM_{10}	Coarse particulate matter index
CO	Carbon Monoxide
NO_2	Nitrogen Dioxide
SO_2	Sulfur Dioxide
O_3	Ozone
Temperature	Air temperature (°C)
Humidity	Relative humidity (%)
Pressure	Atmospheric pressure (hPa)
Wind Speed	Wind speed (m/s)

4.2 Data Collection

Due to the inherent limitations of real-time data collection, where the crawler must send hourly API requests for each station, gathering long-term historical data proved challenging. Therefore, the data scope for this study is limited to April 8, 2025, to July 5, 2025.

Although this timeframe is relatively short, it coincides with the transitional period from spring to summer. This period is characterized by significant weather and meteorological fluctuations, which notably impact the dispersion and diffusion of air pollutants. Consequently, the data collected during this phase remains representative, allowing for a generalized assessment of air pollution levels in key regions of Vietnam. This is particularly relevant given the rapid urbanization and highly variable tropical monsoon climate.

4.2.1 Spatial Scope

This study focuses on analyzing data from two major cities representing Vietnam's Northern and Southern regions:

- **Hanoi**: The capital city and a political and cultural center, characterized by high population density and heavy traffic, is considered a high-risk pollution area in the North.

- **Ho Chi Minh City**: The country's most dynamic economic hub, representing the Southern region with very high population density, construction activities, and traffic volume.

  In addition to these two main cities, data was also collected from several neighboring provinces in the Northern and Southern regions to broaden the analysis by geographical area. Comparing data between these regions allows for assessing differences in air pollution levels between the two parts of the country at the same time, while also supporting the development of spatial pollution zoning models.

4.2.2 Overview of Collected Data

Here's an overview of the data collected for this study:

- **Number of AQI Stations**: 12 stations across various provinces/cities.
- **Number of Weather Points**: 12 corresponding locations.
- **Time Period**: From April 8, 2025, to July 5, 2025.
- **Sampling Frequency**: Hourly.

The data is stored in CSV table format, with each record containing the measurement `timestamp`, `station name`, air pollution indices ($PM_{2.5}$, PM_{10}, CO, NO_2, O_3, SO_2), and accompanying meteorological information (`temperature`, `humidity`, `pressure`, `wind speed`, `geographical coordinates`).

4.3 Data Preprocessing and Analysis Methods

4.3.1 Data Preprocessing

After collecting data from the two primary sources, the Envisoft system (Ministry of Natural Resources and Environment) and OpenWeatherMap, the research team performed preprocessing steps to ensure the data's quality, consistency, and reliability before analysis.

4.3.1.1 Handling Missing Data

Thanks to the crawler system's hourly scheduling and its duplicate-checking mechanism based on `Name` and `Timestamp`, most of the collected data is complete and free from redundancy. However, some records still have missing information due to connection errors or unresponsive APIs.

Here's how missing data was handled:

- **AQI records were retained** even if weather data was missing, to ensure no air quality information was lost.
- **Records were removed** if `AQI` or `Timestamp` information was entirely absent or invalid.
- **Missing quantitative values (NaNs) were kept** as is, allowing for flexible handling during later analysis phases, rather than being immediately imputed or dropped.

4.3.1.2 Time Series Validation and Anomaly Detection

To assess data quality from the monitoring stations, we visualized the daily and hourly AQI time series for each station. By plotting these time series, we easily spotted several clear visual anomalies, such as:

- **Fixed or nearly constant `AQI` values** over extended periods (e.g., always fluctuating around `75 ± 1` for several days).
- **Extremely small fluctuations** in index series compared to other stations under similar `meteorological conditions`.
- **"Step-like" trends** in some stations, where `AQI` values suddenly jump up or drop down, then remain constant at the new level for a long time.

These anomalies often indicate issues with sensor calibration, measurement errors, or technical limitations at certain stations.

4.3.1.3 Anomaly Detection Using Statistical and Machine Learning Algorithms

In addition to visually inspecting time series for unusual data segments, we applied three anomaly detection methods to flag suspicious data points. This combination of methods was chosen to ensure comprehensive coverage of various types of anomalies that may appear in environmental data:

- **`IQR (Interquartile Range)`**
  This method uses quantile statistics to define reasonable value thresholds. Values falling outside the range of `[Q1{1.5×IQR, Q3+1.5×IQR]` are marked as anomalies. `IQR` is suitable for detecting simple amplitude-based outliers but doesn't consider the relationship between nearby data points.
- **`Isolation Forest`**
  This machine learning model uses random binary trees to `"isolate"` data points. Anomalous points are typically isolated after only a few splits of the data. This method performs well with high-dimensional data and can detect global anomalies without requiring assumptions about data distribution.
- **`Local Outlier Factor (LOF)`**
  `LOF` assesses how `"local"` an outlier is by comparing the density of a data point to its neighbors in the feature space. Points with a lower neighboring density compared to their surrounding area are considered anomalous. This method is particularly useful for detecting `local anomalies`—points that aren't necessarily far from the mean but are still distinct from their immediate vicinity.

4.3.1.4 Rationale for Combining All Three Methods

Each method has its own strengths and weaknesses, and each is sensitive to different types of anomalies:

- **`IQR`** is simple and effective for large amplitude outliers.

- **`Isolation Forest`** is powerful at handling non-linear and multidimensional anomalies.

- **`LOF`** excels at detecting local anomalies or those within data clusters.

Combining these methods helps increase the reliability of detection and avoids missing unusual cases that have different underlying natures.

All points flagged as anomalous aren't removed from the dataset. Instead, we added an `is_anomaly` column to record the status of each point, which helps with optional future analyses. This decision aims to avoid mistakenly excluding values that might reflect sudden pollution events such as:

- Environmental incidents (wildfires, chemical accidents, etc.)

- Localized traffic fluctuations

- Extreme meteorological events

- Temporary sensor malfunctions

Only with verified information from independent sources (e.g., reports from monitoring stations, satellite data, news) can the root cause be confirmed and deeper processing decisions be made. Therefore, this approach maintains neutrality and flexibility for subsequent analysis.

4.3.1.5 Supplementing Data Source Identification

To support future scalability and integration with other data sources (e.g., data from AQICN, IQAir, etc.), we've added two new fields:

- **`source`**: Identifies the data origin (e.g., `"gov"` for data provided by the Ministry of Natural Resources and Environment via Envisoft).

- **`source_id`**: The station's identifier within its respective source system.

This allows for better data organization by station system, and also supports merging, comparing, or cross-calibrating data from various collection sources.

4.3.2 Data Analysis

After completing the data collection and preprocessing steps, the data was analyzed using three main approaches to comprehensively assess air quality in the study areas. These analytical methods include: temporal analysis, spatial analysis, and multivariate correlation analysis.

4.3.2.1 Temporal Analysis

Temporal analysis was conducted to clarify the trends and fluctuations of air pollution indicators over time. This approach aims to identify patterns linked to social activities and meteorological conditions.

1. **Descriptive Statistics and Time Series Visualization**
   In the initial phase of the analysis, the research team calculated basic descriptive statistics. This provided an overview of air quality across the studied areas. Specifically, these statistics included:

   - **Mean value**: Reflects the background pollution level for each indicator.
   - **Standard deviation**: Measures the degree of fluctuation, showing the stability or variability of indicators over time.
   - **Frequency of exceeding pollution thresholds (`AQI > 100` ratio)**: Calculated based on the World Health Organization (WHO)'s "unhealthy" classification, this assesses how often residents are exposed to hazardous air environments.

   The goal of this analysis step was to identify prominent characteristics of air pollution, determine the severity in each area, and lay the groundwork for deeper analyses regarding time, space, and connections with meteorological factors.

   Simultaneously, the team also developed time series plots for each monitoring station and key parameter, including: `AQI`, $\texttt{PM}_{2.5}$, $\texttt{PM}_{10}$, `CO`, $\texttt{NO}_2$, $\texttt{SO}_2$, $\texttt{O}_3$, `Temperature`, `Humidity`, `Pressure`, and `Wind Speed`.

   Visualizing data as time series clarified the fluctuation trends of each indicator. This helped detect periods of sudden change or long-term trends, thereby supporting the assessment of meteorological conditions and urban activities on air quality.

2. **Comparison by Weekday and Weekend**
   To assess the impact of societal activities—particularly traffic and industrial operations—on urban air quality, the research team analyzed the distribution of the AQI by weekday (Monday to Friday) and weekend (Saturday and Sunday). This grouping reflects the differing intensity of residential and economic activities in major urban areas.

   Method:

   - AQI data from each monitoring station was categorized by day type (weekday vs. weekend).
   - The mean, standard deviation, and probability distribution (histogram or Kernel Density Estimate - KDE) of AQI were calculated for each group.
   - Statistical tests (such as an independent t-test) were used to determine the significance of differences between the two groups.

   This analysis provides clear insight into the relationship between socioeconomic activities and air pollution levels, thereby supporting the proposal of flexible traffic and production management policies over time to mitigate impacts on public health.

3. **Analysis by Hour of Day: Aggregated Mean and Standard Deviation by Station**
   This analysis dives into AQI data across all stations and observation days, broken down by each hour of the day, creating hourly charts for individual days. This helps pinpoint periods of escalating pollution, which often align with peak traffic

hours (early morning and late afternoon), reflecting the impact of daily activities and production on air quality.

Next, to outline the overall trends and stability of air quality throughout the day, the research team calculated the mean and standard deviation of AQI for each hour, aggregated across all observation days at each station. This analysis provides insights into:

- Average pollution trends by hour of day in each area.
- The extent of AQI variability (fluctuation) over time, shown through the standard deviation, which assesses air quality stability.

4.3.2.2 Spatial Analysis

Spatial analysis was conducted to assess the distribution of air quality across different geographical areas. The focus of this analysis was to compare pollution levels between two major urban centers—Hanoi and Ho Chi Minh City—and extend the examination to their neighboring provinces. This allowed for an investigation into regional pollution differences between the North and the South of Vietnam.

1. **Comparison of Time Series by Region**
   To identify spatial differences in pollution fluctuations, time series plots were constructed at three analytical levels:

   - **Intra-urban**: Comparing stations within the same city (Hanoi or Ho Chi Minh City) to identify local hotspots and assess the level of urban homogeneity.
   - **Intra-regional**: Comparing stations within the same region (Northern or Southern Vietnam), reflecting general pollution trends and the dispersion levels of the indicators.
   - **Between Urban Hubs and Regions**: Conducting an overall comparison between Hanoi and Ho Chi Minh City, and between the Northern and Southern regions, to evaluate the influence of climatic conditions, urban density, and socioeconomic activities on pollution levels.

2. **Regional Descriptive Statistical Analysis**
   Data was grouped by geographical area (city and region), and descriptive statistics were compiled, including:

   - **Mean AQI value**: Reflects the background pollution level.
   - **Standard deviation**: Indicates the stability of air quality.
   - **Frequency of exceeding pollution thresholds (`AQI > 100`)**: Assesses the level of public health risk in each region.

   These results help quantify air quality differences between areas and serve to prioritize environmental management efforts.

3. **Spatial Visualization**
   To enhance data exploration capabilities, geographical spatial visualization methods were applied:

Interactive Map: Implemented using Plotly, this allows users to directly interact with the map (zoom in, zoom out, hover to view detailed information). This effectively supports visualization and aids in communication or decision-making processes.

4.3.2.3 Multivariate Correlation Analysis

To explore the relationships between the Air Quality Index (AQI) and surrounding environmental factors, a series of correlation and regression analyses were performed. These analyses primarily focused on evaluating AQI's dependency on its constituent pollutants and meteorological conditions.

1. **Correlation Analysis by Station and by Hour**

   (a) **Calculating Correlations by Station**
   For each station, the following relevant variables were used to calculate Pearson correlation coefficients, indicating the degree of linear association between pairs of variables:

   - **Pollutants**: $PM_{2.5}$, PM_{10}, CO, NO_2, SO_2, O_3
   - **Meteorological Factors**: Temperature, Humidity, Pressure, Wind Speed
   - **Overall Air Quality Index**: AQI

   For each station with sufficient data, a correlation matrix was computed and visualized as a heatmap. These heatmaps highlight notable relationships.

   (b) **Hourly Correlation Analysis**
   Beyond location, the time of day is also a crucial factor influencing air quality. Therefore, data was grouped by hour of the day, and similarly, Pearson correlation coefficients were calculated for each hourly interval.

   This allows for identifying diurnal periods when meteorological factors most strongly influence pollution levels. The visualization of hourly correlation matrices was also performed using heatmaps, aiding in the comparison of fluctuation trends throughout the day.

2. **Visualization of AQI – Other Factors Correlation by Hour and Space**
   From the detailed correlation matrices, a combined heatmap showing AQI – parameter correlations by hour was generated for each station. This allows for tracking changes in the relationship between AQI and meteorological factors over the daily cycle. This approach helps identify correlations that only appear at specific times (e.g., early morning, late evening, or peak hours).

3. **Detection of Temperature Inversion Conditions**
   In addition to analyzing the correlation between meteorological factors and air quality, this study also focused on the detection of temperature inversion conditions—a specific meteorological phenomenon that can lead to severe accumulation of air pollutants, particularly impacting the Air Quality Index (AQI).

   (a) **Climate Context of Northern Vietnam**
   Northern Vietnam features a tropical monsoon climate with four distinct seasons. A temperature inversion occurs when a warmer layer of air lies above a

colder layer near the ground, preventing air from rising and circulating normally. Consequently, pollutants from traffic, industry, and daily activities cannot disperse, leading to their accumulation near the surface and a severe increase in the AQI.

Although temperature inversions are generally known to be more prevalent in the dry winter months (November to March) due to low temperatures, overcast skies, and calm winds, our data indicates that similar conditions can still occur and cause pollution during other periods.

(b) **Criteria for Detecting Temperature Inversion Conditions**
Instead of building a complex function, we used a set of combined criteria to identify periods with a high probability of temperature inversions. These criteria are applied directly to each data record (i.e., each specific time and monitoring station), based on key meteorological factors influencing atmospheric stability and pollutant accumulation:

- **High Atmospheric Pressure (`Pressure > 75th` Percentile)**: Rather than a fixed threshold, we defined "high pressure" based on the 75th percentile (P75) of actual pressure data from Northern Vietnam. This approach provides a relative and more flexible definition of high pressure, suitable for the region's pressure distribution. High pressure typically indicates stable, less mobile air, which allows pollutants to be trapped.
- **Low Wind Speed (`Wind Speed < 2 m/s`)**: Wind plays a critical role in dispersing pollutants. Low wind speeds signify minimal air movement, allowing pollutants to accumulate.
- **High Humidity (`Humidity > 85%`)**: High humidity often correlates with stable atmospheric conditions and can lead to the formation of fog or haze, increasing the potential for pollutant accumulation.
- **Early Morning Hours (Hour from 0 to 6)**: The period from midnight to 6 AM is when the ground cools most rapidly due to radiative heat loss, often leading to radiative inversions where cold air is trapped near the surface.
- **High Air Quality Index (`AQI > 100`)**: This condition was included to ensure that our focus remains on instances where the combination of meteorological factors genuinely results in poor air quality, indicating significant pollutant accumulation.

A moment in time is flagged as having potential inversion conditions if all these criteria are simultaneously met.

4.4 Tools and Software Utilized

The entire data analysis process in this study was primarily conducted using the Python programming language, combined with several powerful data science libraries and visualization tools. This approach ensured computational efficiency, result reliability, and clear visualization capabilities.

4.4.1 Programming Environment and Language

- **Python 3.10**: The main programming language used for all stages of data preprocessing, statistical analysis, and data visualization.

4.4.2 Key Libraries and Analytical Tools

Here are the primary libraries and analytical tools used in this study:

- **`pandas`**: Used for tabular data manipulation, grouping, calculating descriptive statistics, and preparing data for analysis.
- **`NumPy`**: Supported efficient array operations and numerical computations with large datasets.
- **`SciPy`**: Utilized for conducting statistical tests, such as the `t-test` for comparing weekdays and weekends.
- **`Matplotlib`** and **`Seaborn`**: Employed for visualizing data in various chart formats, including line plots, distribution plots, bar charts, and notably, heatmaps to represent correlation matrices.
- **`Plotly`**: Used to create interactive maps and charts, enhancing data presentation and exploration.
- **`Scikit-learn`**: Supported data standardization, correlation matrix computation, and the implementation of multivariate analysis methods (should the study expand to regression or clustering).

4.4.3 Infrastructure and Platform

Jupyter Notebook was the primary development environment. It allowed for the seamless integration of source code with descriptive text, analytical results, and illustrative figures within a single document.

4.4.4 Automation and Result Storage

- **`Git and GitHub`**: These tools were used for managing source code and analysis results, facilitating change tracking and efficient collaboration.
- **`CSV`** and **`Parquet`**: These formats were chosen for storing intermediate data and analysis outputs, ensuring high retrieval performance and compatibility with data processing tools.

The selection of the tools mentioned was carefully considered to optimize for the specific requirements of each analysis step. This approach also ensures transparency, reproducibility, and the scalability of the research findings for future work.

5 Results

***Note**: For the analysis results section, only two representative stations were selected from each region for detailed illustration:*

- *Northern Region: Hanoi – 556 Nguyen Van Cu (KK)*
- *Southern Region: Ho Chi Minh City – Ministry of Natural Resources and Environment Inter-agency Complex*

*This selection ensures conciseness in presentation, and both stations had complete and stable data throughout the survey period. Comprehensive results for all stations are presented in the **Appendix**.*

5.1 Data Overview

5.1.1 Data Scope

The dataset comprises over **20,000 samples**, collected from **April 8, 2025**, to **July 5, 2025**. This data covers **12 monitoring stations** located across **9 provinces and cities**, including `Hanoi`, `Ho Chi Minh City`, and other provinces in both the **Northern** and **Southern** regions.

List of Monitoring Stations and Coordinates:

- **Ha Noi:** DHBK – cong Parabol duong Giai Phong (21.0052, 105.8418)
- **Ha Noi:** 556 Nguyen Van Cu (21.0491, 105.8831)
- **Ha Noi:** Cong vien Nhan Chinh – Khuat Duy Tien (21.0031, 105.7947)
- **TP.HCM:** Khu Lien co quan Bo TN&MT – 20 Ly Chinh Thang (10.7823, 106.6834)
- **TP.HCM:** Le Huu Kieu – P. Binh Trung Tay, Quan 2 (10.7823, 106.7528)
- **Da Nang:** DH Su pham – Khuon vien truong (16.0622, 108.1594)
- **Thai Nguyen:** Duong Hung Vuong – TP. Thai Nguyen (21.59315, 105.8431)
- **Phu Tho:** Duong Hung Vuong – TP. Viet Tri (21.33847, 105.3633)
- **Bac Giang:** Khu lien co quan tinh – P. Ngo Quyen (21.3015, 106.22603)
- **Ha Nam:** Cong vien Nam Cao – P. Quang Trung (20.536, 105.9165)
- **Long An:** UBND TP. Tan An – 76 Hung Vuong – P.2 (10.5391, 106.4045)
- **Binh Duong:** 593 Dai lo Binh Duong – P. Hiep Thanh (10.9923, 106.6577)

The collected indices include: $\texttt{AQI}$, $\texttt{PM}_{2.5}$, $\texttt{PM}_{10}$, $\texttt{CO}$, $\texttt{NO}_2$, $\texttt{SO}_2$, $\texttt{O}_3$, `Temperature`, `Humidity`, `Pressure`, and `Wind Speed`.

5.1.2 Basic Descriptive Statistics

Indicator	Mean	Std	Min	Max	Exceedance Rate (%)
AQI	54.73	27.86	6.00	172.00	7.27
$PM_{2.5}$	51.65	29.14	1.28	172.20	–
PM_{10}	43.87	22.84	1.49	167.59	–
CO	12.14	7.89	0.03	59.04	–
NO_2	8.10	9.83	0.04	76.23	–
O_3	15.98	17.01	0.43	177.59	–
SO_2	5.60	6.22	0.40	54.84	–
Temperature (°C)	27.81	3.26	17.73	39.01	–
Humidity (%)	78.64	13.06	24.00	100.00	–
Pressure (hPa)	1006.66	3.64	995.00	1021.00	–
Wind Speed (m/s)	2.88	1.62	0.00	13.96	–

Table 1: Descriptive statistics of collected air quality and weather indicators

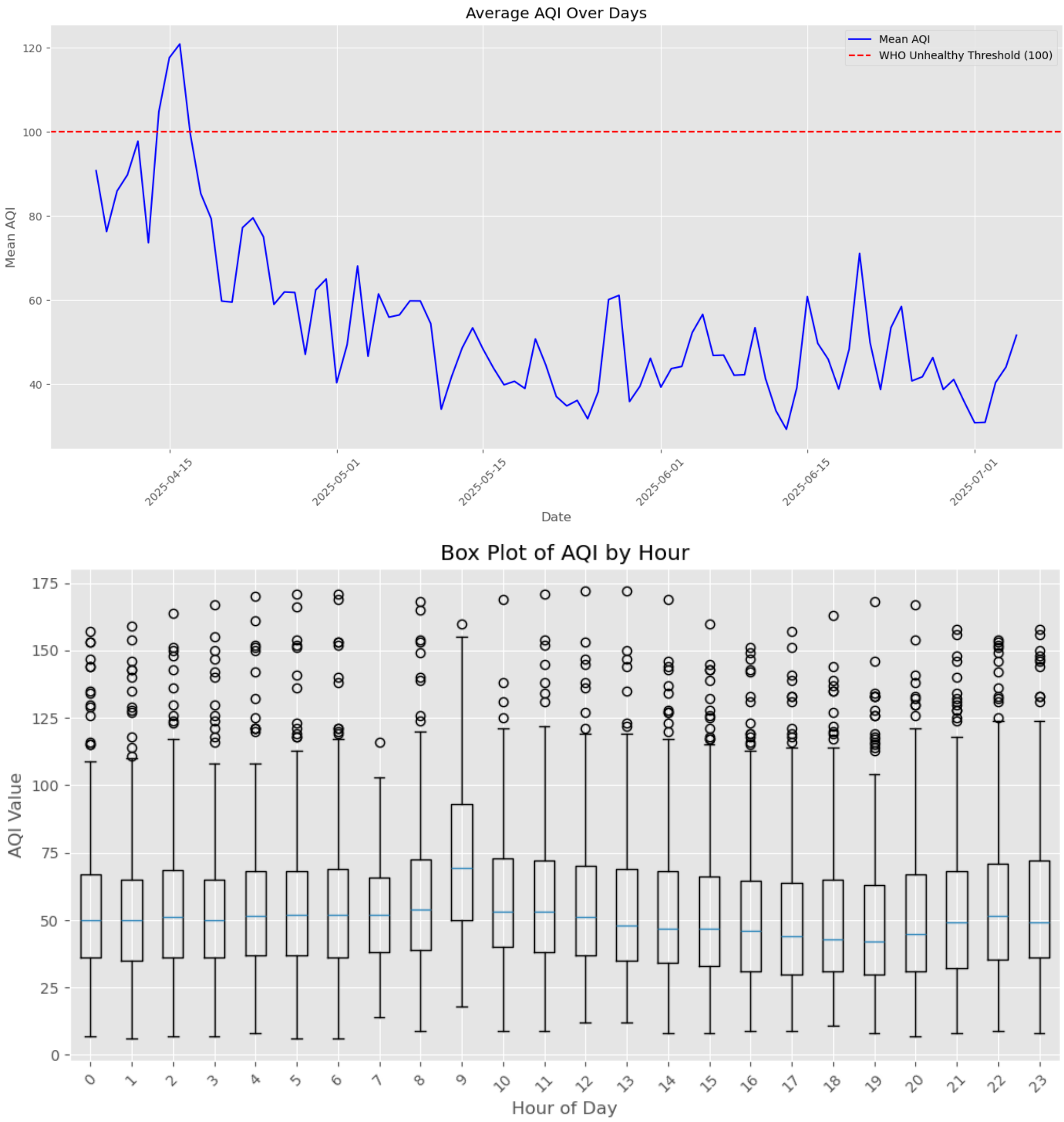

Figure 1: Daily and hourly averages of AQI from the observational dataset.

5.2 Temporal Analysis

5.2.1 Time Series Fluctuations

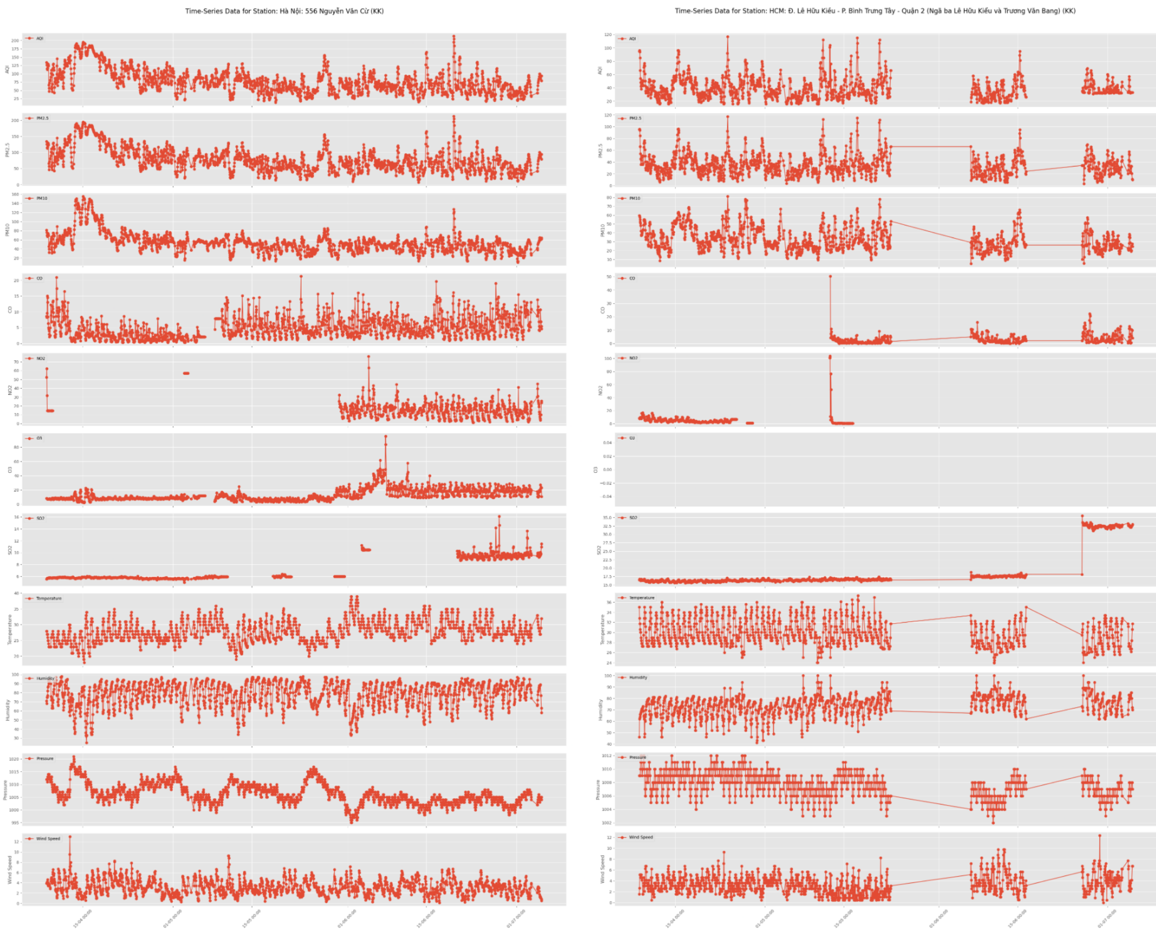

Figure 2: Hourly variations of air quality and meteorological data during the survey period at each station.

5.2.2 AQI Comparison by Day of the Week

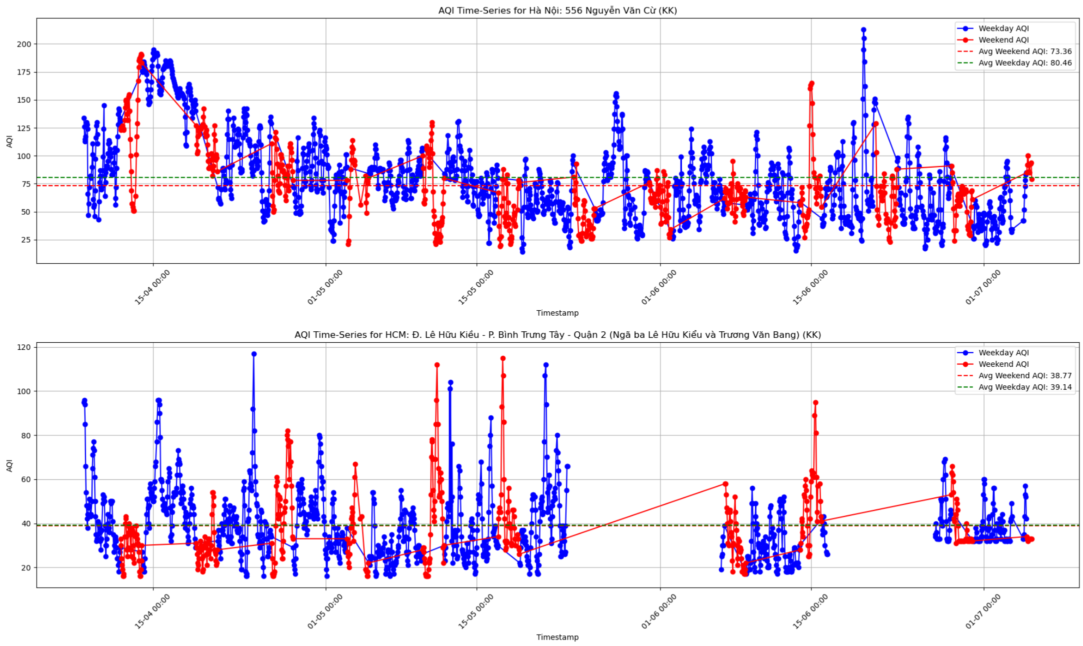

Figure 3: Time series of daily AQI, categorized by weekdays (Mon–Fri) and weekends (Sat–Sun).

Here's a breakdown of the Air Quality Index (AQI) analysis, comparing weekday and weekend averages for two representative stations in Vietnam:

Metric	**Hanoi: 556 Nguyen Van Cu (KK)**	**HCM: Le Huu Kieu - Truong Van Bang (KK)**
Average AQI on Weekends	73.36	38.77
Average AQI on Weekdays	80.46	39.14
Difference (Weekend - Weekday)	-7.11	-0.36
Weekend Days Analyzed	503	376
Weekday Days Analyzed	1356	972

Table 2: Weekday vs. Weekend AQI Comparison in Hanoi and Ho Chi Minh City

5.2.3 AQI Analysis by Hour of Day

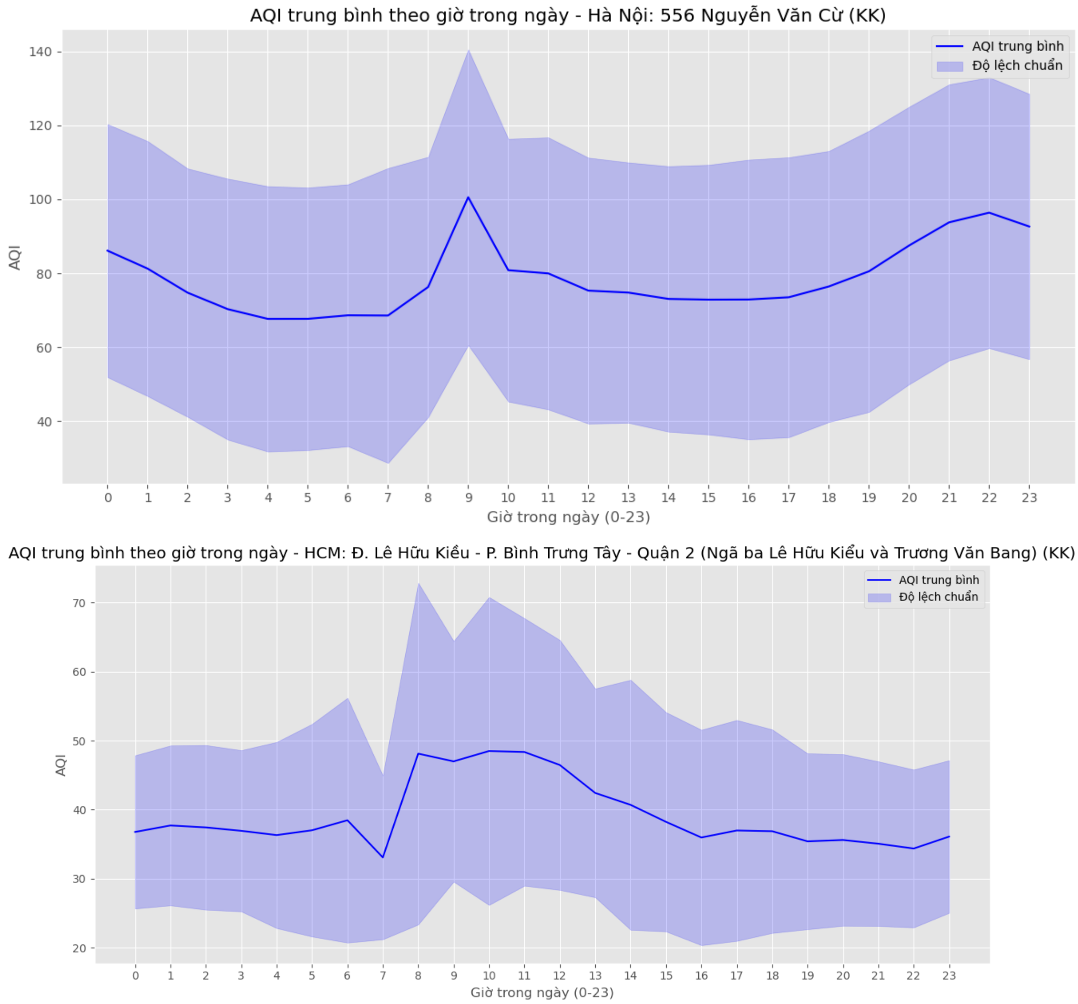

Figure 4: Hourly average and standard deviation of AQI at monitoring stations.

5.3 Spatial Analysis

5.3.1 Intra-City Comparison

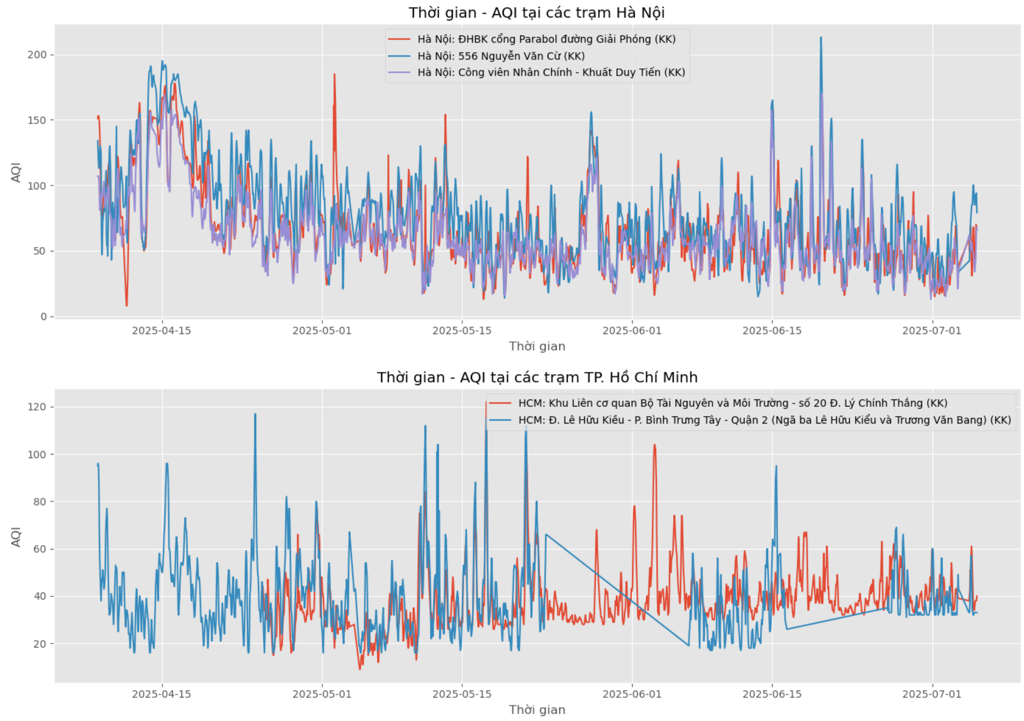

Figure 5: AQI time series at monitoring stations in Hanoi and Ho Chi Minh City.

5.3.2 Regional Descriptive Statistical Analysis

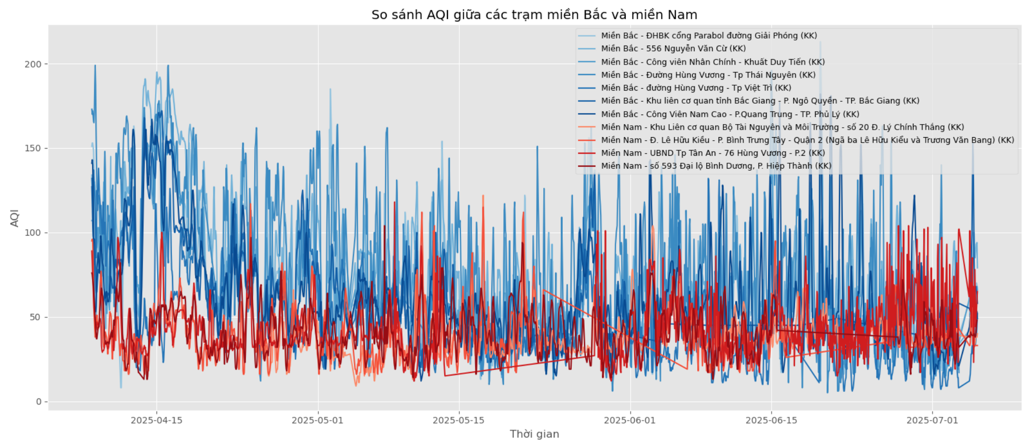

Figure 6: Time series of AQI at monitoring stations in Northern and Southern Vietnam

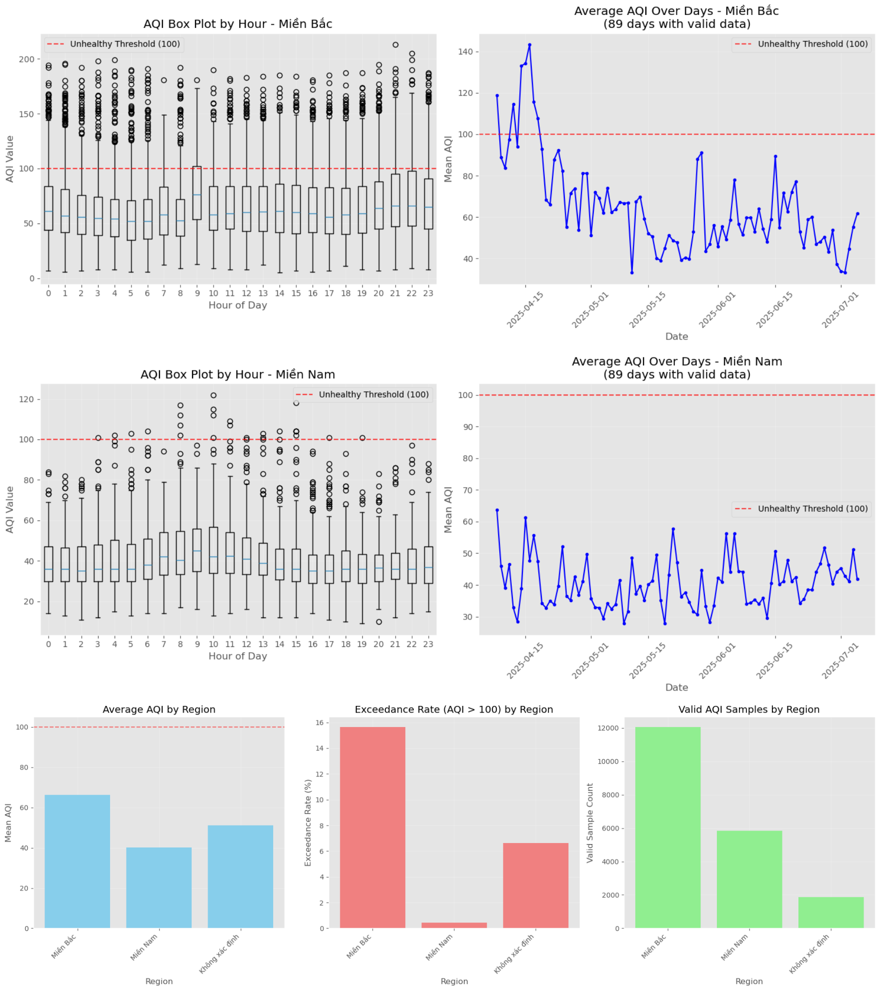

Figure 7: Comparison Chart of Average AQI Between Northern and Southern Monitoring Stations

Region	Total Samples	Valid Samples	% Missing	Mean	Std	Exc. Thresh. (%)
Northern	12,726	12,053	5.3%	66.3	34.3	15.6%
Southern	5,840	5,836	0.1%	40.2	14.8	0.4%
Undefined	1,868	1,868	0.0%	51.2	26.5	6.6%

Table 3: Comparing AQI Across Regions

5.3.3 Spatial Visualization

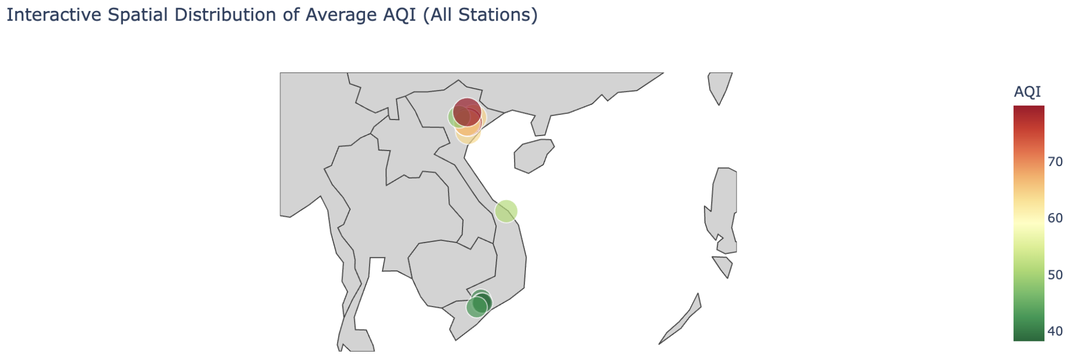

Figure 8: Interactive map showing the geographical locations of air quality monitoring stations across Vietnam.

5.4 Multivariate Correlation Analysis

5.4.1 Correlation by Station

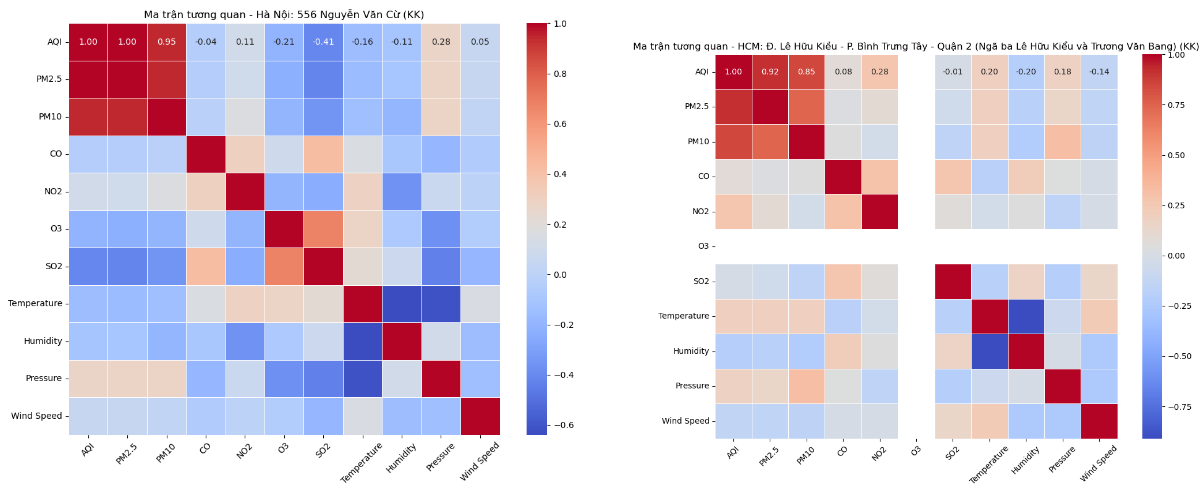

Figure 9: Correlation matrices of various indicators at each monitoring station

5.4.2 Hourly Correlation

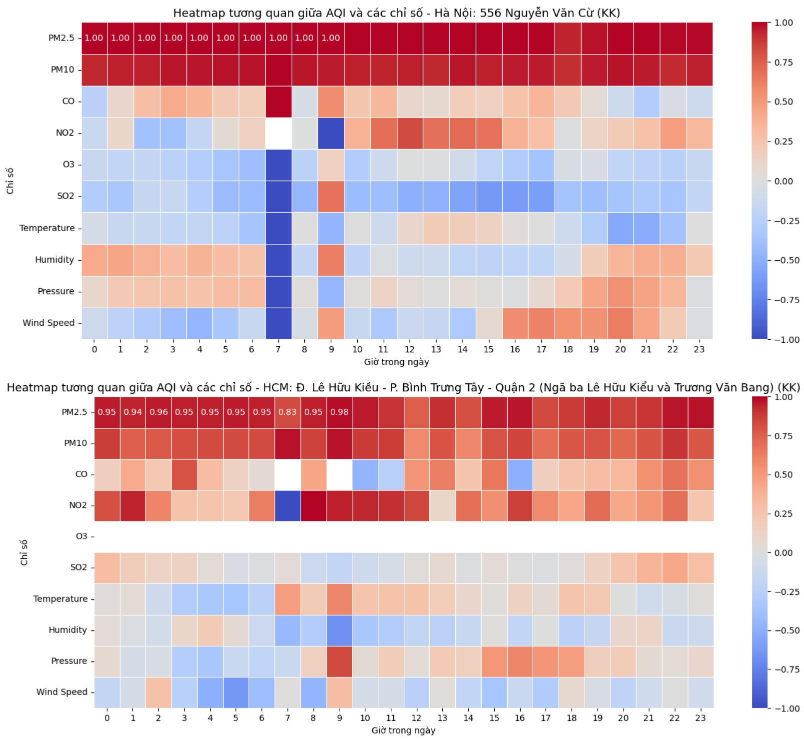

Figure 10: Hourly heatmaps of AQI correlations with other indicators at monitoring stations.

5.5 Temperature Inversion Detection

The pressure threshold (75th percentile) was calculated to be:**1009.00 hPa**.

Temperature Inversion Summary

Table 4: Diurnal Frequency

Day	Hours
2025-04-15	12
2025-04-16	2
2025-05-01	3
2025-05-02	3
2025-05-13	1
2025-05-28	6

Inversions on 6 of 89 days (6.7%).

Table 5: Hourly Frequency

Hour	Count
0	5
1	4
2	3
3	2
4	2
5	2
6	9

Table 6: AQI Comparison

Condition	Avg AQI
With Inversion	147.89
Without Inversion	63.20

T-test: $t = 15.78$, $p = 0.0000$

6 Discussion

6.1 Key Findings Summary

Our analysis of AQI time series data from numerous monitoring stations nationwide reveals several key conclusions. *Please note that, due to computational limitations and to focus on primary characteristics, the correlation analyses of individual pollutants (CO, NO_2, O_3, SO_2) and the influence of meteorological factors are summarized specifically from stations in Hanoi.*

6.1.1 Relationship Between AQI and Pollutants

Analysis of the AQI's relationship with various pollutants reveals some clear patterns:

- **Particulate Matter ($PM_{2.5}$ and PM_{10}) are the Dominant Factors:** AQI shows a very high correlation with $PM_{2.5}$ and PM_{10} across most hours. This confirms the primary role of fine and coarse particulate matter in overall air quality, primarily originating from traffic, construction activities, and fuel combustion.

- **CO and NO_2 Play a Significant Role During Morning Peak Hours:** Pollutants like CO and NO_2 are crucial during morning peak hours. These gases are mainly products of fuel combustion in internal combustion engines (cars, motorbikes), which are major emission sources in Vietnamese cities.

- **O_3 and SO_2 Have Less Direct Impact on AQI:** Ozone and sulfur dioxide generally show weak or negative correlations with AQI and particulate matter concentrations. While NO_2 and O_3 can be linked through photochemical reactions that produce ozone under sunlight, SO_2 (primarily from industrial coal combustion) contributes less to overall AQI compared to particulate matter and CO/NO_2 in this dataset.

6.1.2 Regional and Temporal AQI Fluctuations

Here's an overview of how AQI varies across regions and over time:

- **Distinct Regional Differences:** We observed a significant difference in AQI between the Northern and Southern regions, with the North generally showing higher AQI values. Notably, AQI in the Northern region displayed a clear seasonality, with high average levels recorded in early April (within the analyzed period).

- **Intra-Regional Variation:** Even within the same city—such as Hanoi—monitoring stations recorded considerable AQI discrepancies. This highlights the influence of geographical location, traffic density, and residential or industrial activities in specific urban areas.

- **Diurnal (Hourly) Fluctuations:**
  - AQI varies by hour, typically averaging higher during active hours from 6 AM to 9 PM.
  - However, no pronounced spikes were observed during typical rush hours (7–9 AM and 5–7 PM), unlike patterns seen in many other urban centers.

  - Interestingly, stations in the Northern region did not exhibit clear AQI trends related to daily activity patterns, as observed in Southern stations. This suggests possible influences from geographical, meteorological, or urban-structural differences.

- **Weekend Effect:** On weekends, average AQI levels were lower than on weekdays, reflecting a likely reduction in emission-generating activities, such as industrial production and road traffic.

6.1.3 Influence of Meteorological Factors

Meteorological factors play a crucial role in either dispersing or accumulating pollutants:

- **Wind Speed:** Shows a significant **positive correlation with AQI** in the evening. This suggests that dust and pollutants may be transported from external urban sources (e.g., traffic, construction, or agricultural activities) into the monitoring areas during these hours.

- **Temperature and Humidity:**
  - *Temperature:* High temperatures generally show a **negative correlation with AQI**, as they promote atmospheric convection and pollutant dispersion.
  - *Humidity:* High humidity, particularly in the early morning, may **increase fine particulate matter** concentrations. Moisture-laden conditions help particles absorb pollutants, leading to higher AQI readings.

- **Atmospheric Pressure:** High pressure conditions in the evening and early morning are associated with **higher AQI levels**. This is indicative of atmospheric inversions or stagnant conditions, where pollutants are trapped near the ground and fail to disperse.

6.1.4 Evidence and Hypotheses on Temperature Inversions in Northern Vietnam

Based on a deeper analysis of meteorological factors (pressure, wind speed, humidity, and time of day) combined with high AQI, this study found strong evidence for the existence of conditions favorable for temperature inversions in Northern Vietnam:

- **Inversion-like events**, defined by the following conditions:
  - `Pressure > 1009 hPa`,
  - `Wind Speed < 2 m/s`,
  - `Humidity > 85%`,
  - `Time between 0:00-6:00 AM`,
  - `AQI > 100` ,

  were primarily observed in **April and May**, with a clear peak around **6:00 AM**.

- The **average AQI during these inversion-like conditions** was **147.89**, significantly higher than the **63.20** observed when such conditions were absent. A t-test confirmed this difference to be **highly statistically significant** (`p-value < 0.0001`, supporting the hypothesis that these conditions contribute to severe pollutant accumulation.
- Additional meteorological factors, such as large day–night temperature differentials, further support the **likelihood of radiative inversions** occurring during nighttime and early morning hours.

6.2 Comparison with Previous Research

The findings of this study align with many widely published discoveries in the field of meteorological pollution and air quality:

- **Role of PM2.5 and PM10:** The dominance of fine particulate matter in shaping the AQI is a consistent conclusion globally (*WHO*, *EPA*) and in major Asian cities (e.g., *Beijing, Delhi, Hanoi, Ho Chi Minh City*).
- **CO/NO_2 Link to Traffic:** The connection between CO and NO_2 concentrations and peak hour traffic activity is a common characteristic of urban air pollution.
- **Meteorological Influence:** The principles concerning the influence of wind (dispersion), temperature (convection), and humidity (particle absorption) on air quality are all confirmed in various studies.
- **Temperature Inversion Phenomenon:** Empirical evidence of increased AQI under conditions of high pressure, low wind, high humidity, and early morning hours strongly reinforces research on radiative inversions as a primary cause of severe pollution episodes worldwide, particularly in regions with basin topography or similar climatic conditions.
- **Seasonality:** The trend of higher pollution during the dry/cold season in monsoon climates has also been well-documented.

However, this study also highlights some points that warrant further discussion:

- **High AQI in Early April in Northern Vietnam:** While winter is typically the peak pollution season in Northern Vietnam, recording high average AQI levels in early April might reflect transitional season characteristics or the influence of seasonal emission sources that haven't been analyzed in depth.
- **Lack of Distinct AQI Peaks During Traffic Rush Hours:** This could be due to the specific characteristics of the monitoring stations, high background pollution levels, or more complex meteorological factors affecting local dispersion during these hours.

6.3 Explaining the Results: Underlying Reasons

Analysis results reflect the combined impact of various natural and anthropogenic factors, specifically:

6.3.1 Geographical and Climatic Characteristics of Northern Vietnam

Northern Vietnam, especially the Hanoi area and the Red River Delta provinces, possesses geographical and climatic features that make it particularly susceptible to temperature inversions, especially during the winter-spring season. During this time, cold air masses move in, causing temperature inversions where a layer of cold air lies close to the ground, trapped by a warmer air layer above. This phenomenon impedes the vertical diffusion of pollutants, leading to the accumulation of fine particulate matter and harmful gases near the ground, resulting in severe air pollution, particularly in the early mornings and at night.

From a geographical and topographical perspective, the Northern Delta region has a flat and sheltered terrain, especially large urban areas like Hanoi. The capital is surrounded by low mountain ranges to the North and West, creating a semi-enclosed spatial structure that further restricts air convection. Furthermore, Hanoi is one of the cities with the highest population and vehicle density in the country. Coupled with suboptimal urban planning regarding ventilation, high building density, and uneven tree distribution, these factors increase the potential for localized pollution accumulation, especially under adverse meteorological conditions such as low wind, high humidity, and rising pressure.

The prevalent conditions of low rainfall, high humidity, and low wind speeds in the early part of the year further reduce the atmosphere's ability to disperse and wash away fine particulate matter. This causes the Air Quality Index (AQI) in this region to tend to rise and remain at alarming levels for several consecutive days.

6.3.2 High Population Density and Dense Traffic

In major urban centers like Hanoi, Ho Chi Minh City, and many centrally-governed cities, high population density and dense traffic are significant contributors to air pollution. During the active hours from 6 AM to 9 PM, continuous traffic flow from a large number of motorbikes, cars, buses, etc., contributes to the emission of pollutants such as NO_2 and CO, along with fine particulate matter ($PM_{2.5}$, PM_{10}) generated from fuel combustion and mechanical friction between tires, brakes, and road surfaces.

These emissions are continuously added to the atmosphere throughout the day. When meteorological conditions are unfavorable for dispersion, the AQI tends to remain elevated for extended periods throughout the active hours. This phenomenon is particularly pronounced in city centers, main roads, or busy intersections.

6.3.3 Industrial and Construction Activities

Areas like Hanoi, Bac Ninh, Thai Nguyen, and Binh Duong are major centers for industrial development and construction, with many factories, industrial zones, and large-scale construction projects. Here, emission sources such as combustion furnaces, industrial production processes, material transport vehicles, and especially dust from construction activities are primary contributors to high concentrations of fine particulate matter ($PM_{2.5}$, PM_{10}).

Unlike traffic pollution, which tends to concentrate during peak hours, emissions from industrial and construction activities often occur continuously throughout the day. This causes the AQI to remain elevated for extended periods from morning until evening. Furthermore, if meteorological conditions are unfavorable (low wind, high humidity, temperature inversions, etc.), pollutants from these activities tend to accumulate longer in the air, exacerbating pollution levels.

6.3.4 Seasonal Influence

During the dry season (January to April), Northern Vietnam and many other provinces often experience a lack of rainfall, low air humidity, and limited plant biomass. These conditions significantly reduce the natural air-cleansing capacity. The absence of processes like the washout of fine particulate matter by rain and the absorption of gaseous pollutants by vegetation leads to the accumulation of pollutants such as $PM_{2.5}$, PM_{10}, NO_2, and CO in the atmosphere.

Conversely, during the summer and rainy season (June to September), frequent heavy rains help to wash away and disperse pollutants. Simultaneously, increased wind speeds and atmospheric convection create favorable conditions for the dilution of polluted air, significantly contributing to a reduction in AQI and fine particulate matter concentrations.

6.3.5 Potential Errors Due to Measurement Equipment and Monitoring Station Distribution

A portion of the data inaccuracies may stem from differences among measurement devices at various monitoring stations. These include variations in sensitivity, calibration frequency, maintenance conditions, and distance from major emission sources. Such factors can impact the accuracy and consistency of the recorded AQI values.

Furthermore, the uneven distribution of monitoring stations—being densely concentrated in urban areas like Hanoi or Ho Chi Minh City, but sparse in rural or suburban regions—can lead to data bias. This means that the analysis results primarily reflect air quality in densely populated areas, rather than providing a comprehensive overview at a regional or national level.

6.4 Limitations of the Study

Although this study provides an overview of air pollution in several representative areas of Vietnam, several limitations should be noted:

- **Limited Data Coverage**
  The monitoring station network is primarily concentrated in urban and delta regions, while many mountainous or rural areas still lack comprehensive AQI monitoring systems. This creates an imbalance in data distribution and can introduce bias in regional comparisons.

- **Errors from Equipment and Data Sources**
  Technical errors from sensors, along with differences in measurement standards and methods across data sources, can affect the consistency and accuracy of the analysis results.

- **Incomplete Coverage of Meteorological Factors**
  The study primarily analyzed basic meteorological factors such as temperature, humidity, pressure, and wind speed. However, other crucial factors like solar radiation, UV index, cloud cover, or photochemical decomposition rates in the atmosphere were not integrated into the model. These factors significantly influence the formation, dispersion, and transformation of pollutants, especially O and NO.

  Furthermore, due to data limitations and processing capacity, the study could not fully and accurately analyze the correlations between environmental indicators at all stations across every hour of the day. Hourly and regional variations, as well as non-linear interactions between meteorological factors and pollutants, remain underexplored. This highlights the need for expanded research with more complex multivariate models in the future.

- **Limited Survey Period**
  The analyzed data focuses on a short period, therefore not fully reflecting long-term pollution trends or seasonal cyclical patterns throughout the year.

- **Lack of Socio-Behavioral Factor Analysis**
  Important contributing factors such as the frequency of personal vehicle use, agricultural burning practices (e.g., rice straw burning), or the level of environmental policy compliance were not included in the quantitative analysis, despite their direct impact on air quality in many areas.

6.5 Proposed Practical Applications and Future Research Directions

6.5.1 Practical Applications

- **Real-time Air Pollution Warning and Monitoring:**
  Based on the strong correlation between AQI and parameters like $PM_{2.5}$, PM_{10}, pressure, and humidity, a real-time hourly air pollution early warning system can be developed for urban areas, particularly Hanoi and Northern provinces – areas heavily influenced by weather conditions and traffic activities.

- **Urban Planning and Environmental Policy Support:**
  The results indicating high AQI in spring and early summer, as well as clear disparities between areas within the same city, suggest a need to re-evaluate urban, residential, and industrial planning. This also points to the necessity of increasing green spaces or environmental buffer zones at pollution hotspots.

- **Optimizing Outdoor Activities and Community Life:**
  By identifying high AQI hours (6 AM–9 PM) and lower pollution levels on weekends, residents and authorities can proactively adjust school schedules, work hours, or outdoor events to reduce exposure to pollution.

6.5.2 Future Research Directions

- **Expanding Research Scope (Spatial and Temporal):**
  Supplementing data from monitoring stations in mountainous, rural, and Central regions will provide a comprehensive national overview. Furthermore, the survey

period should be extended to cover all four seasons to analyze annual climatic influences.

- **Integrating Satellite Data and Machine Learning Models:**
  Utilizing remote sensing data from `Sentinel-5P`, `MODIS`, or `Himawari-8`, combined with deep learning models, to forecast AQI levels or detect pollution hotspots spatially.

- **In-depth Analysis of Temperature Inversions and Pollutant Accumulation Mechanisms:**
  Continue to deeply investigate the phenomenon of temperature inversions in Northern Vietnam – a factor that can drastically increase AQI under specific meteorological conditions.

- **Incorporating Socio-Behavioral Factors:**
  Further research into the influence of population density, traffic volume, and manual fuel burning activities to build cause-and-effect models, thereby supporting data-driven policy decisions.

7 Conclusion

This study analyzed Air Quality Index (AQI) data from various monitoring stations across Vietnam, incorporating meteorological factors such as temperature, humidity, pressure, and time. Our findings clearly show that AQI fluctuations are significantly linked to weather conditions, time of day, day of the week, and specific local areas. Pollution levels tend to be higher during peak traffic hours, on weekdays, and during the spring–early summer, especially in urban areas with high vehicle and population densities.

These discoveries not only deepen our understanding of air pollution characteristics in Vietnam but also highlight the potential for practical applications. This includes developing early warning systems, supporting green urban planning, and formulating environmental policies tailored to specific regional characteristics.

For future research, we recommend expanding the study's scope both spatially (to cover the entire nation, including rural and urban areas) and temporally (through annual and seasonal analyses). Integrating remote sensing data and applying machine learning models will be effective approaches to enhance the accuracy and proactive nature of air pollution monitoring and forecasting.

References

- **World Health Organization (WHO).** (2021). *WHO global air quality guidelines: Particulate matter (PM2.5 and PM10), ozone, nitrogen dioxide, sulfur dioxide and carbon monoxide.* Geneva: World Health Organization.
  Note: This is a key foundational document for air quality standards and health impacts.

- **United Nations Environment Programme (UNEP).** (2019). *Air Pollution in Asia and the Pacific: Towards a Regional Response.* Bangkok, Thailand.
  Note: This provides a broad overview of air pollution sources and challenges in the Asian region, relevant to the context of Vietnam.

- **Seinfeld, J. H., & Pandis, S. N.** (2016). *Atmospheric Chemistry and Physics: From Air Pollution to Climate Change* (3rd ed.). John Wiley & Sons.
  Note: A fundamental textbook covering chemistry, physics, dispersion, and temperature inversions.

- **Trinh, T. T., Trinh, T. T., Le, T. T., Nguyen, T. D. H., & Tu, B. M.** (2018). Temperature inversion and air pollution relationship, and its effects on human health in Hanoi City, Vietnam. *Environmental Science and Pollution Research*, 25(34), 34698–34707.
  Note: Empirical study on inversion effects in Hanoi.

- **Bo Tai nguyen va Moi truong Viet Nam (MONRE).**

Appendix

A Time Series Fluctuations

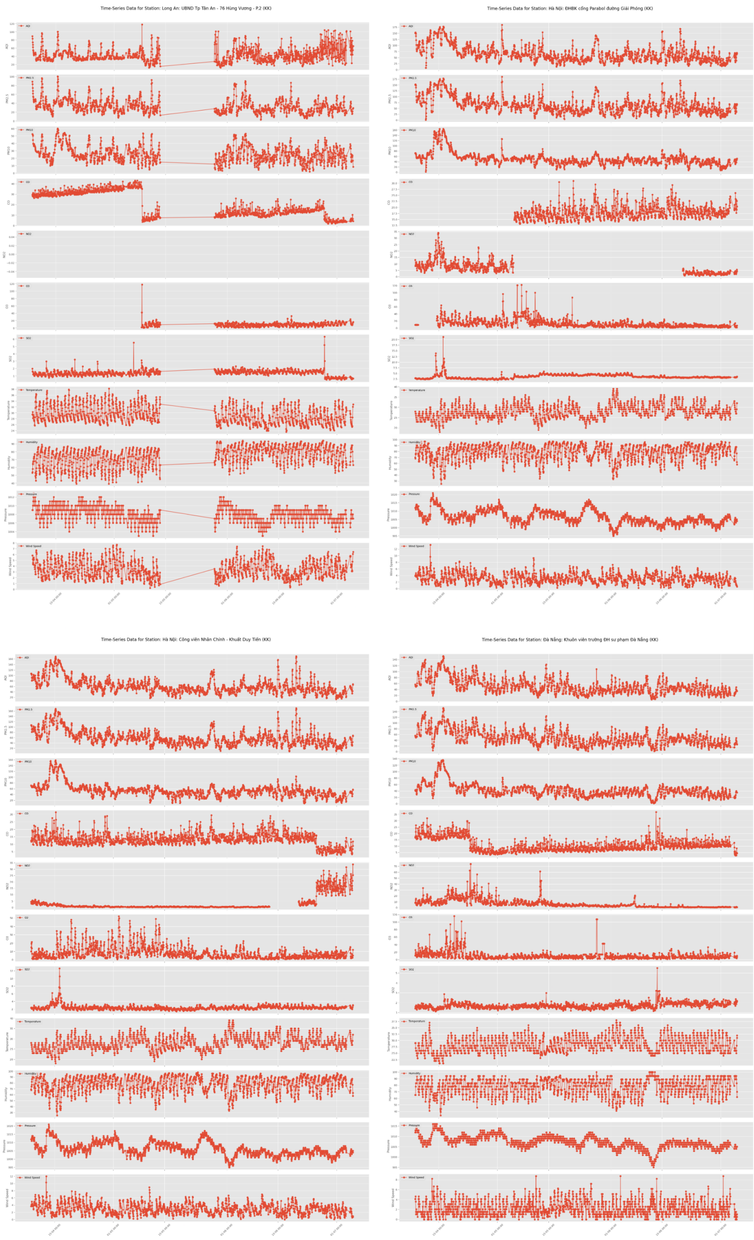

Figure A1: Time series of all major air quality indicators at Long An, HN - Giai Phong, HN - Khuat Duy Tien, Da Nang

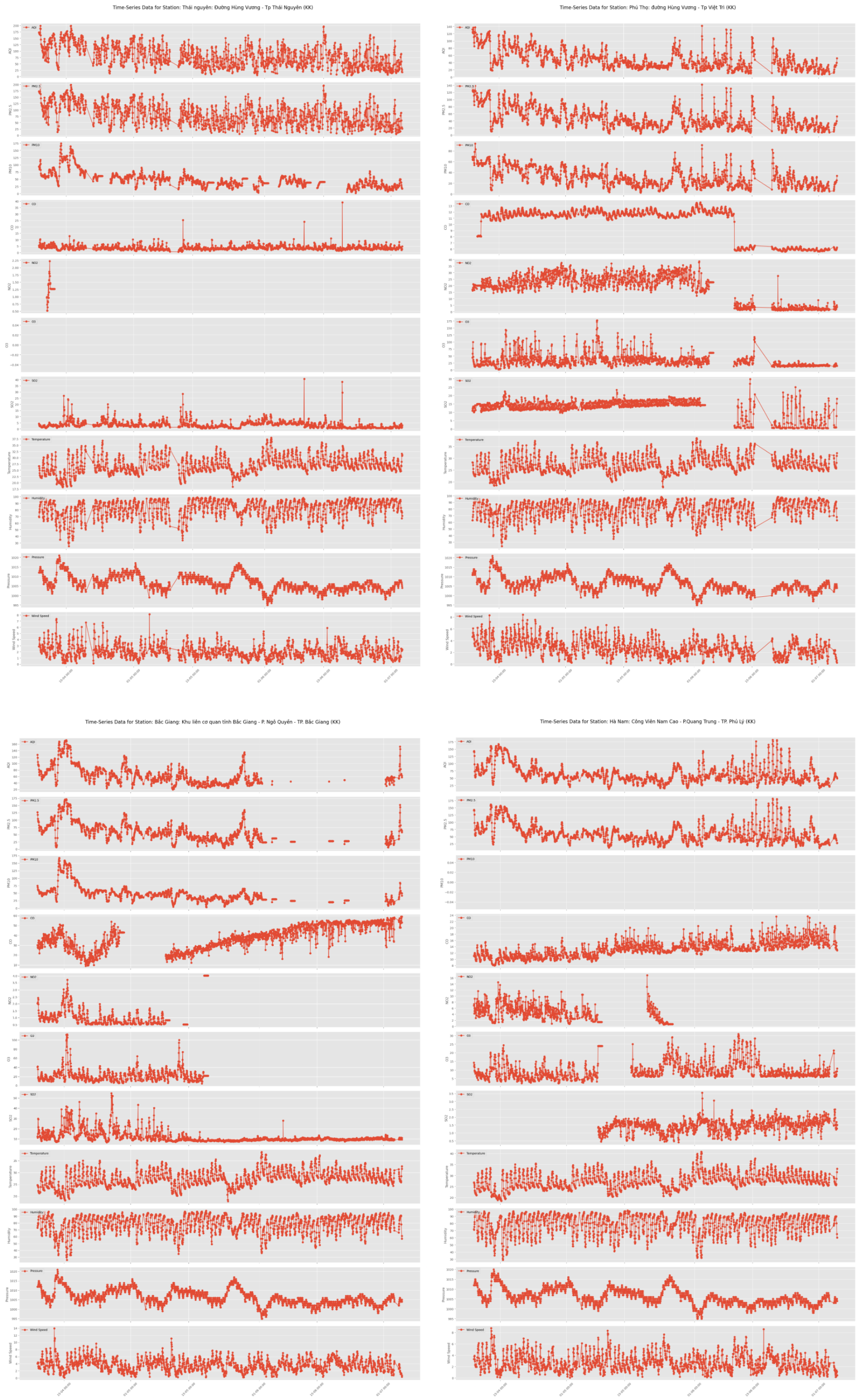

Figure A2: Time series of all major air quality indicators at Thai Nguyen, Phu To, Bac Giang, Ha Nam

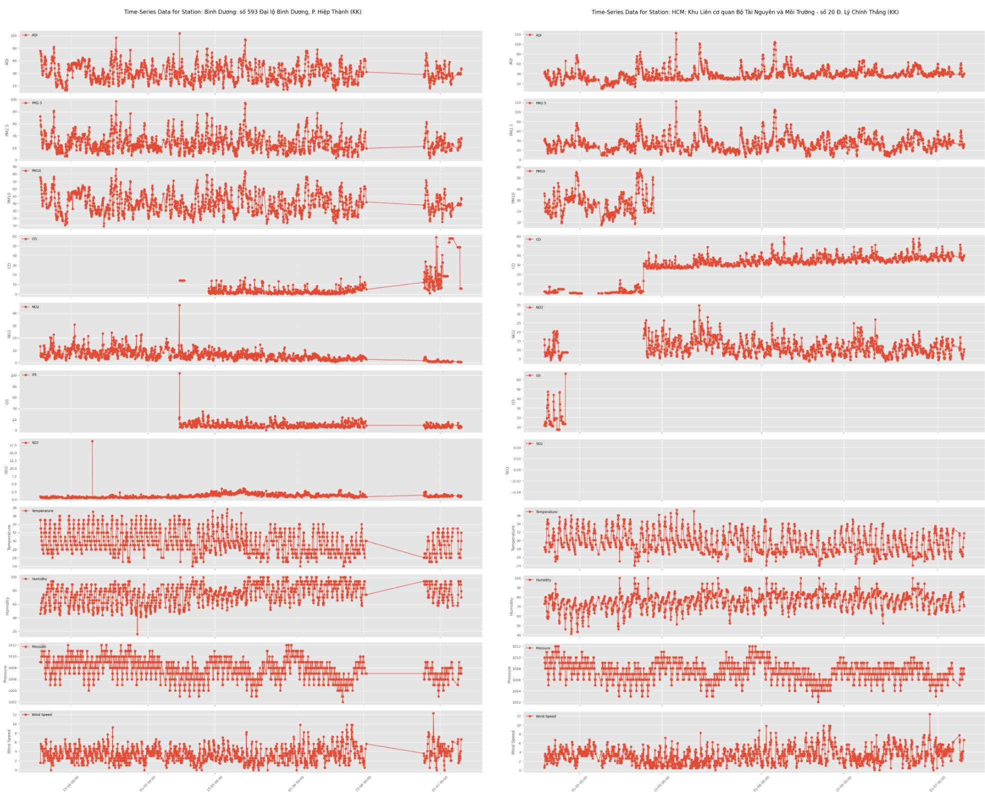

Figure A3: Time series of all major air quality indicators at Binh Duong, HCM - Bo Tai nguyen va Moi truong

B AQI Comparison by Day of the Week

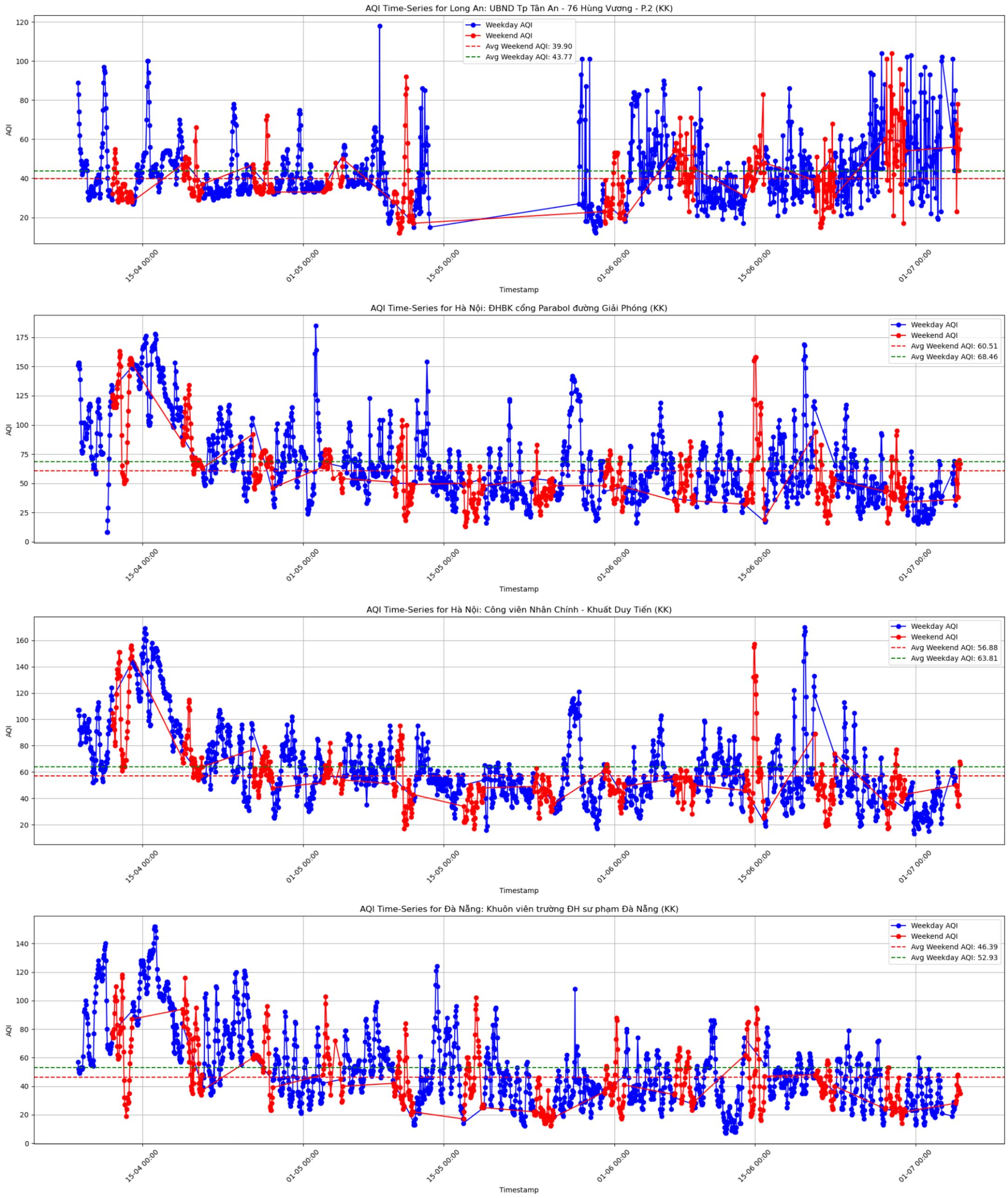

Figure B1: Comparison of AQI by day of the week at Long An, HN - Giai Phong, HN - Khuat Duy Tien, Da Nang

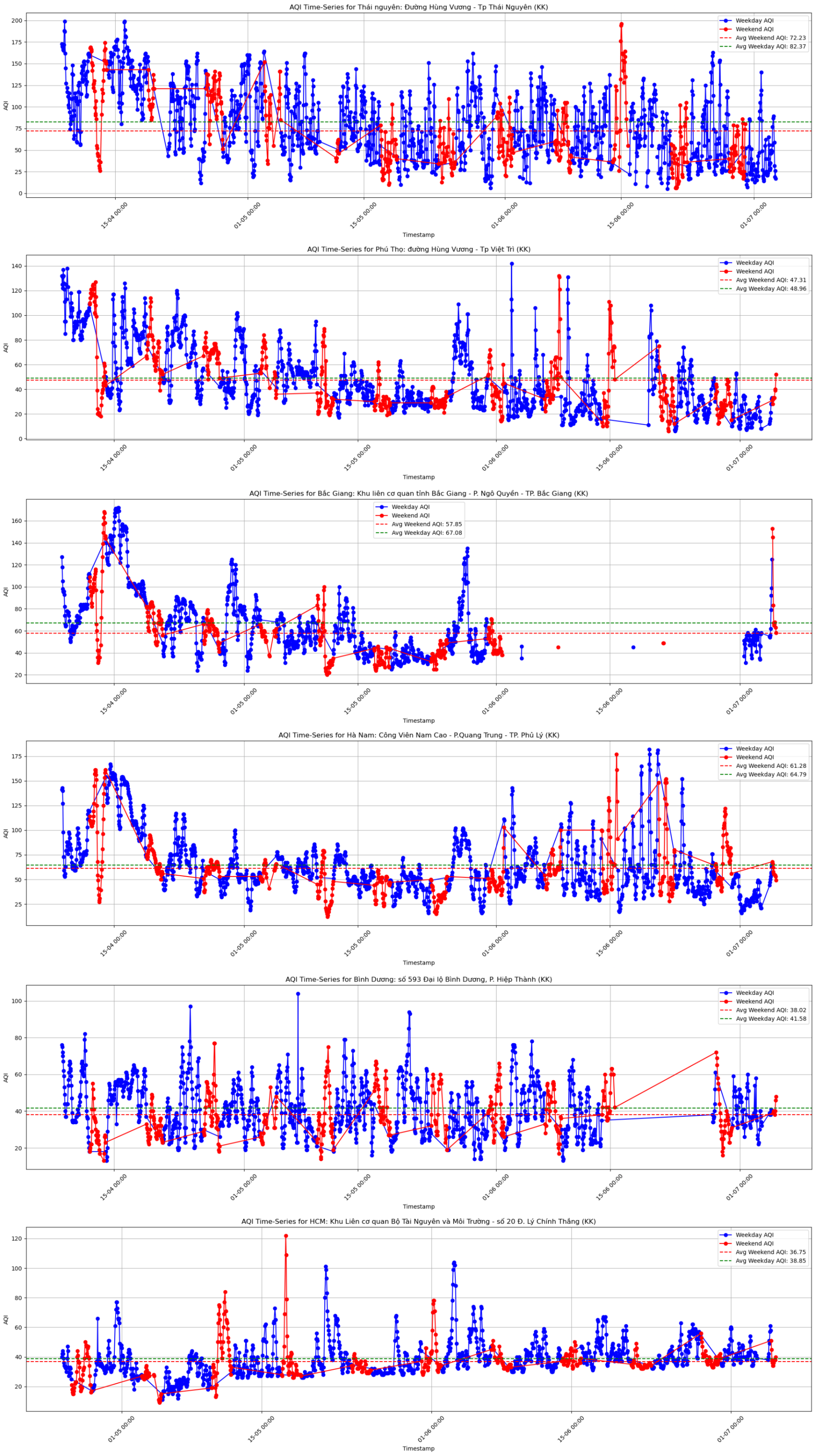

Figure B2: Comparison of AQI by day of the week at Thai Nguyen, Phu To, Bac Giang, Ha Nam, Binh Duong, HCM - Bo Tai nguyen va Moi truong

Monitoring Station	**Weekend AQI**	**Weekday AQI**	**Difference**	**Weekend Days**	**Weekday Days**
Long An: UBND Tp Tan An – 76 Hung Vuong – P.2	39.90	43.77	-3.87	407	1094
Ha Noi: DHBK – cong Parabol – Giai Phong	60.51	68.46	-7.96	498	1344
Ha Noi: Cong vien Nhan Chinh – Khuat Duy Tien	56.88	63.81	-6.93	502	1358
Da Nang: DH Su pham – Khuon vien truong	46.39	52.93	-6.54	502	1366
Thai Nguyen: Duong Hung Vuong – TP Thai Nguyen	72.23	82.37	-10.14	419	1320
Phu Tho: Duong Hung Vuong – TP Viet Tri	47.31	48.96	-1.65	492	1224
Bac Giang: Khu lien co quan tinh – P. Ngo Quyen	57.85	67.08	-9.23	498	1357
Ha Nam: Cong vien Nam Cao – P. Quang Trung	61.28	64.79	-3.50	501	1354
Binh Duong: 593 Dai lo Binh Duong – P. Hiep Thanh	38.02	41.58	-3.56	399	1095
HCM: Khu lien co quan – 20 D. Ly Chinh Thang	36.75	38.85	-2.09	413	1084

Table B1: Weekday vs. Weekend AQI Comparison

C AQI by Hour of Day

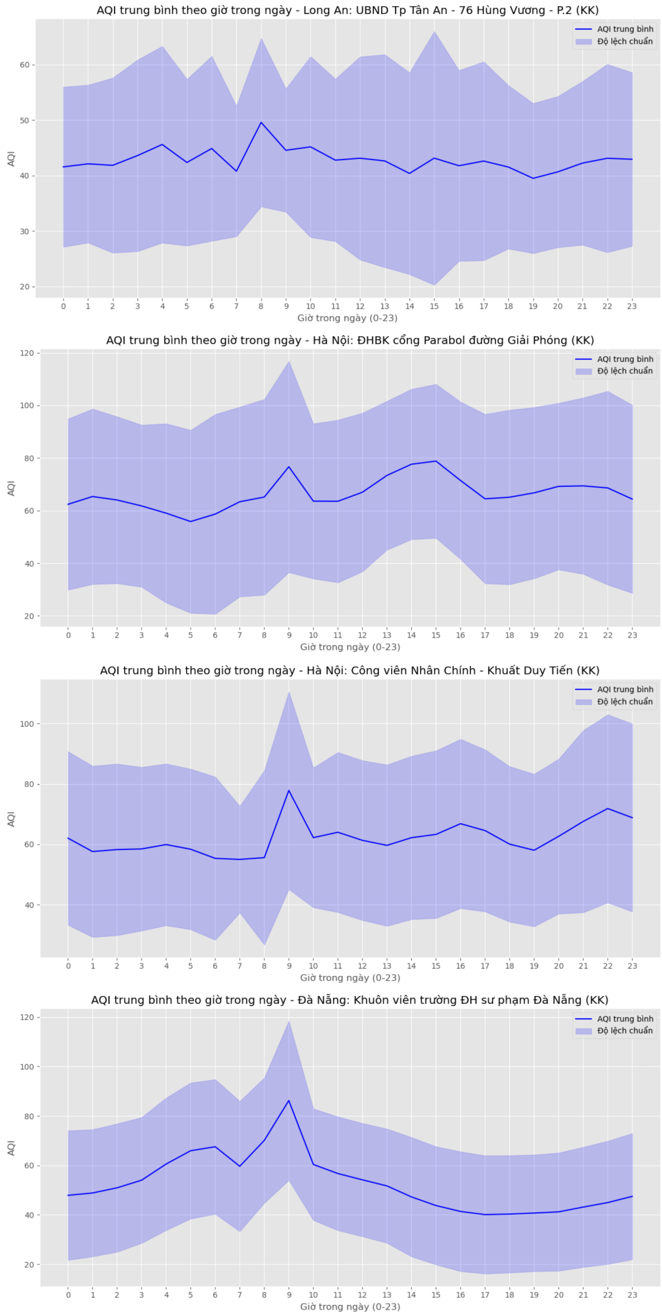

Figure C1: Hourly variation of AQI at Long An, HN - Giai Phong, HN - Khuat Duy Tien, Da Nang

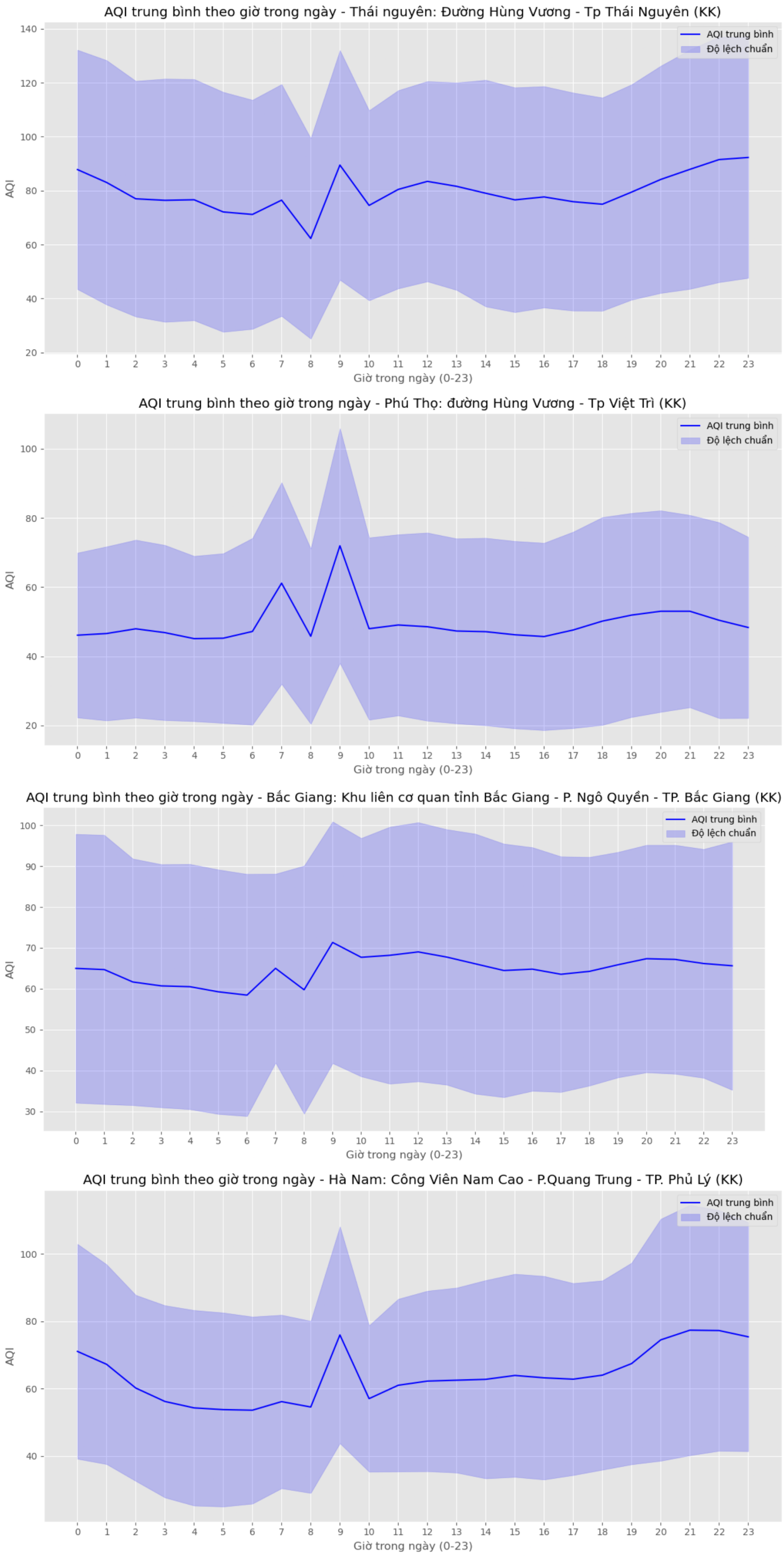

Figure C2: Hourly variation of AQI at Thai Nguyen, Phu To, Bac Giang, Ha Nam

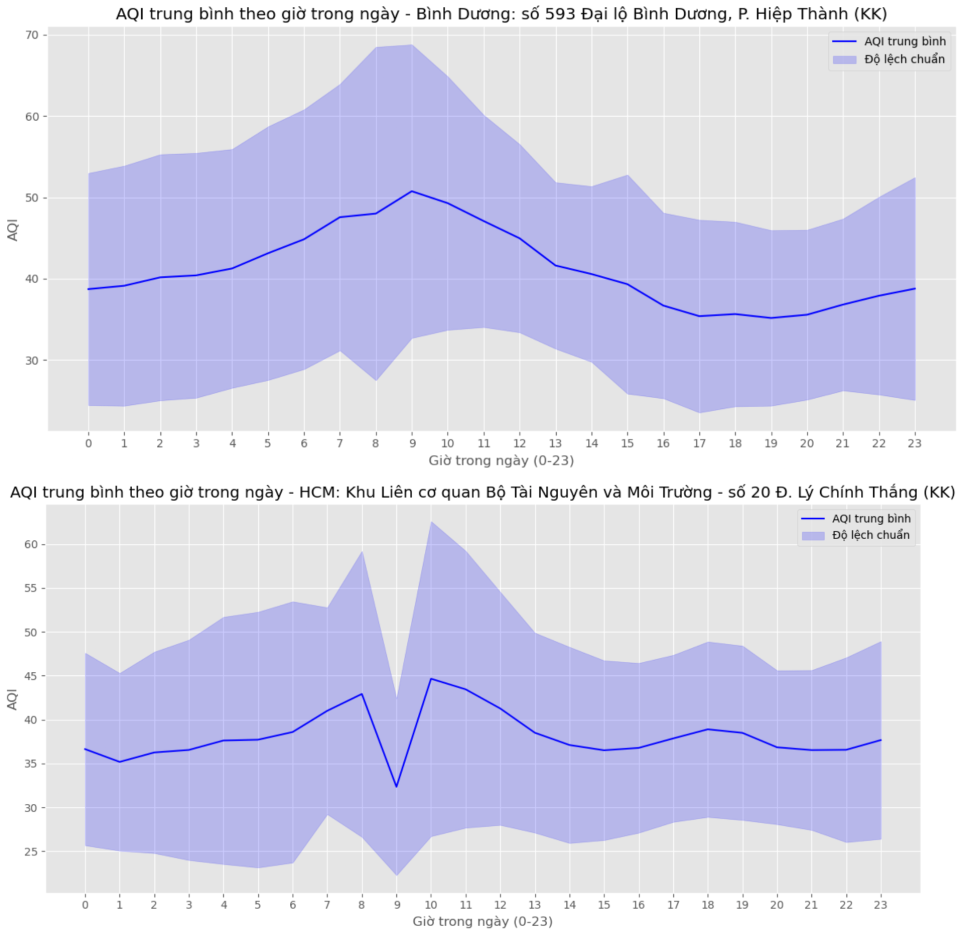

Figure C3: Hourly variation of AQI at Binh Duong, HCM - Bo Tai nguyen va Moi truong

D Correlation by Station

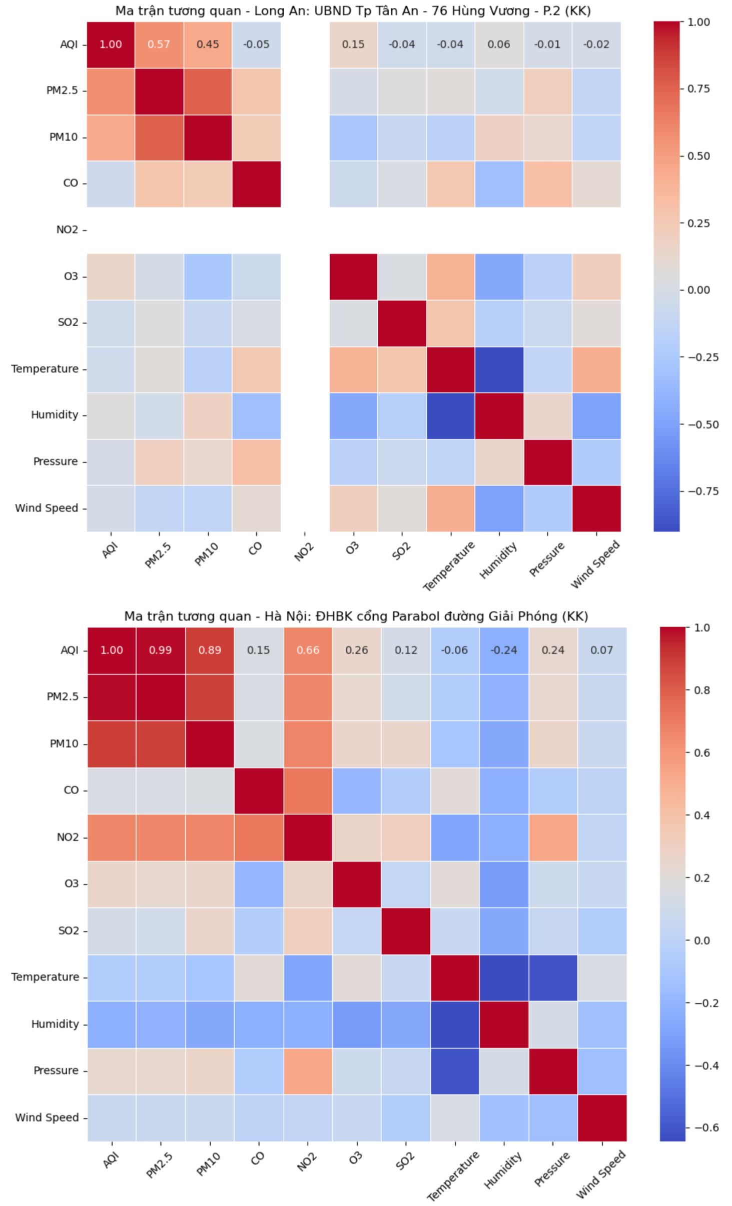

Figure D1: Correlation of all major air quality indicators at Long An, HN - Giai Phong

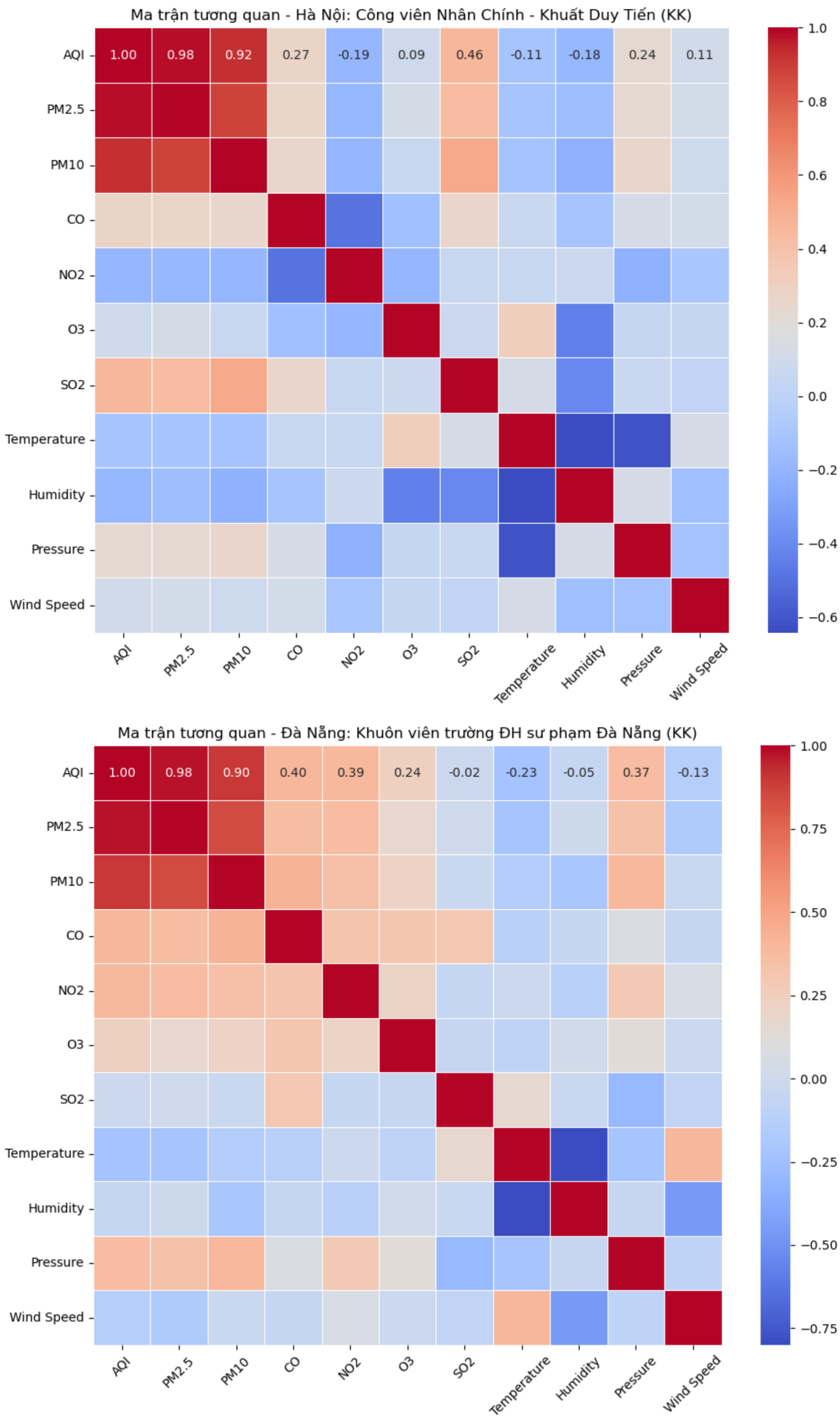

Figure D2: Correlation of all major air quality indicators at HN - Khuat Duy Tien, Da Nang

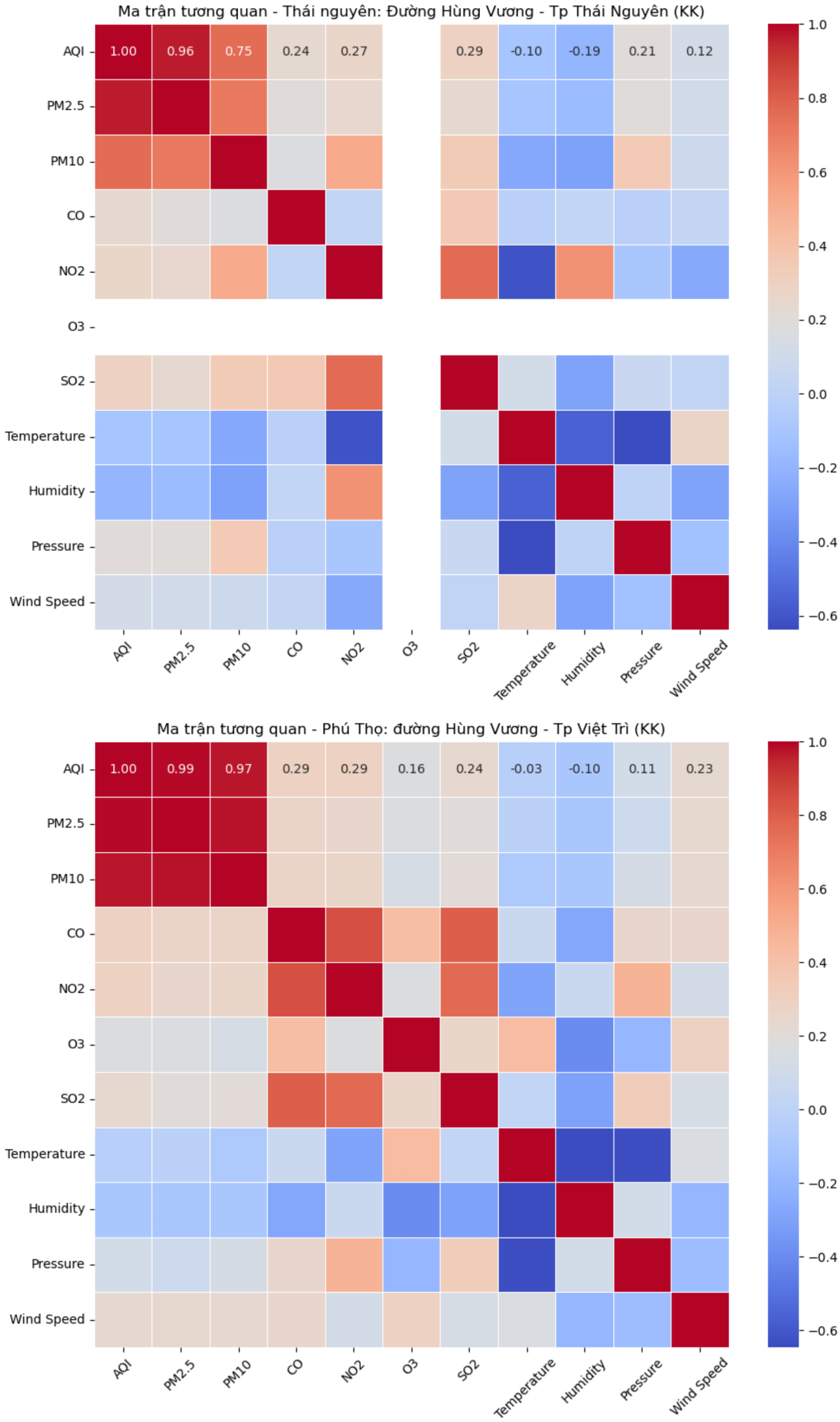

Figure D3: Correlation of all major air quality indicators at Thai Nguyen, Phu Tho

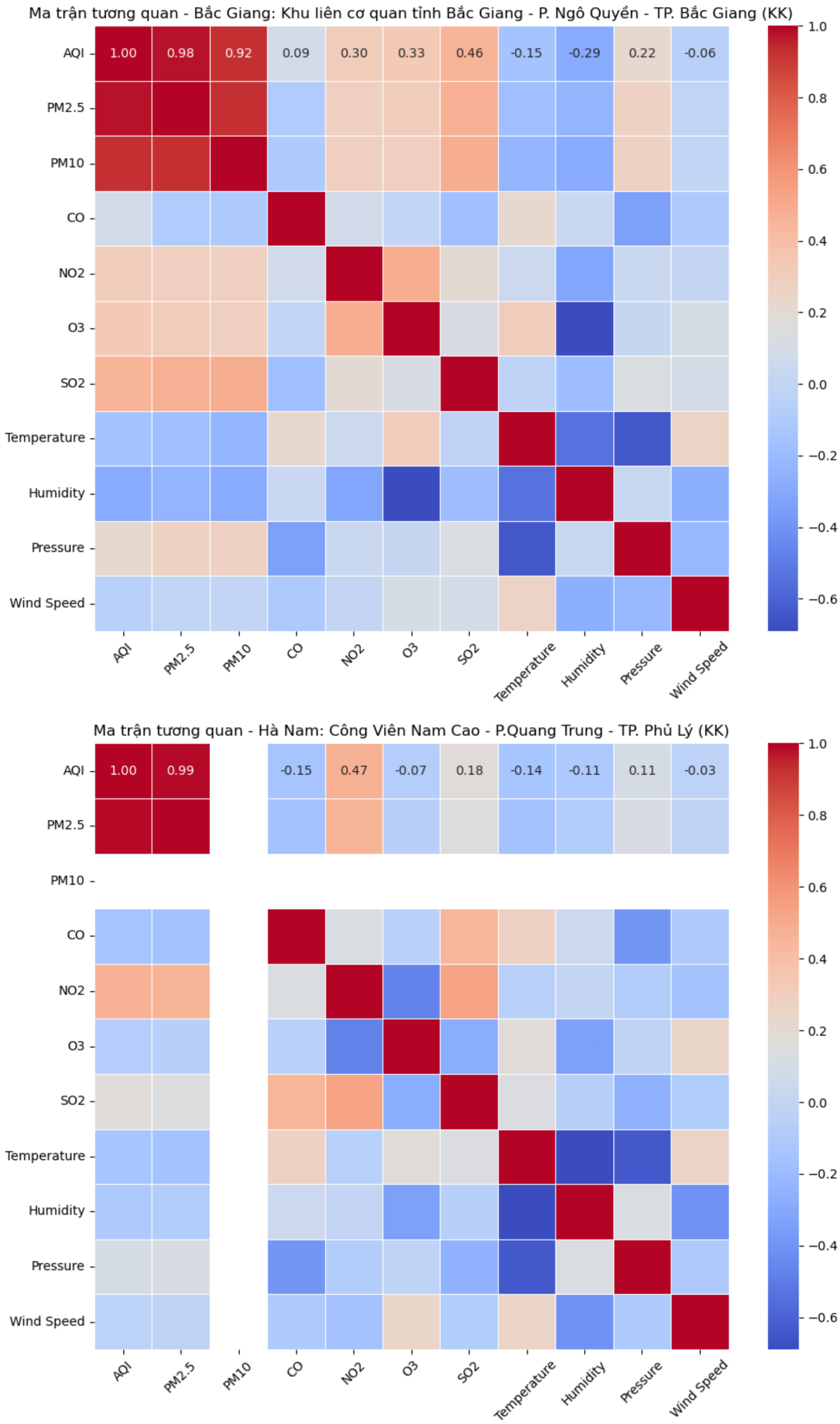

Figure D4: Correlation of all major air quality indicators at Bac Giang, Ha Nam

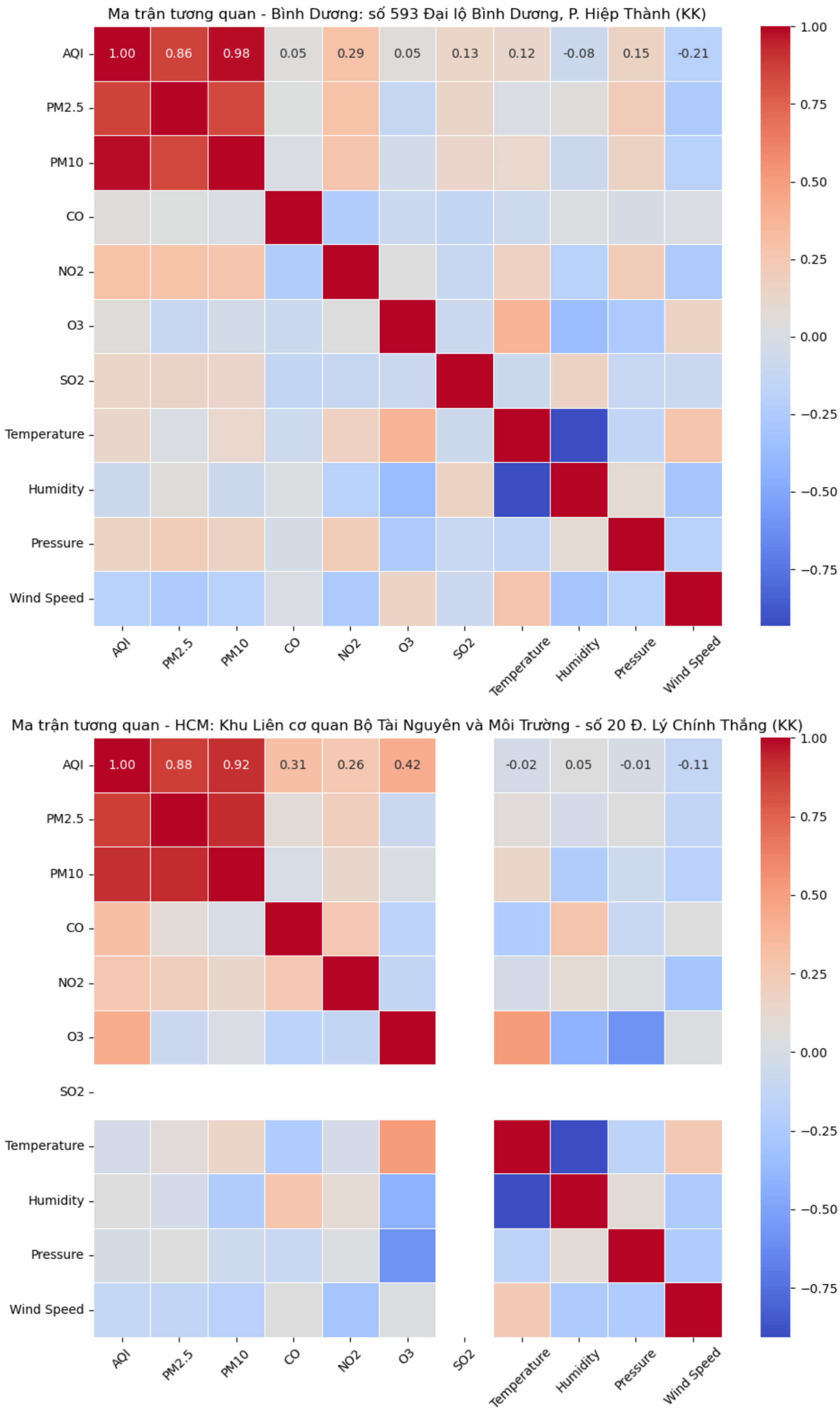

Figure D5: Correlation of all major air quality indicators at Binh Duong, HCM - Bo Tai nguyen va Moi truong

E Hourly Correlation

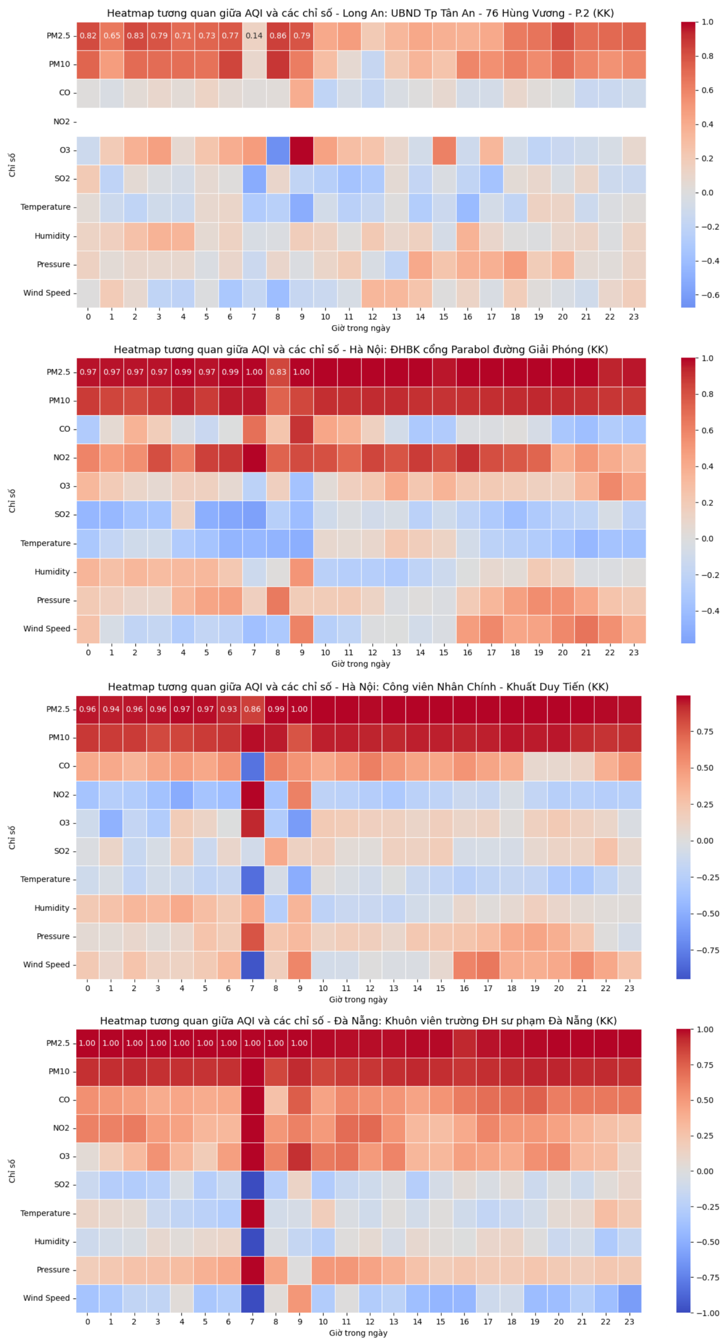

Figure E1: Hourly correlation of all major air quality indicators at Long An, HN - Giai Phong, HN - Khuat Duy Tien, Da Nang

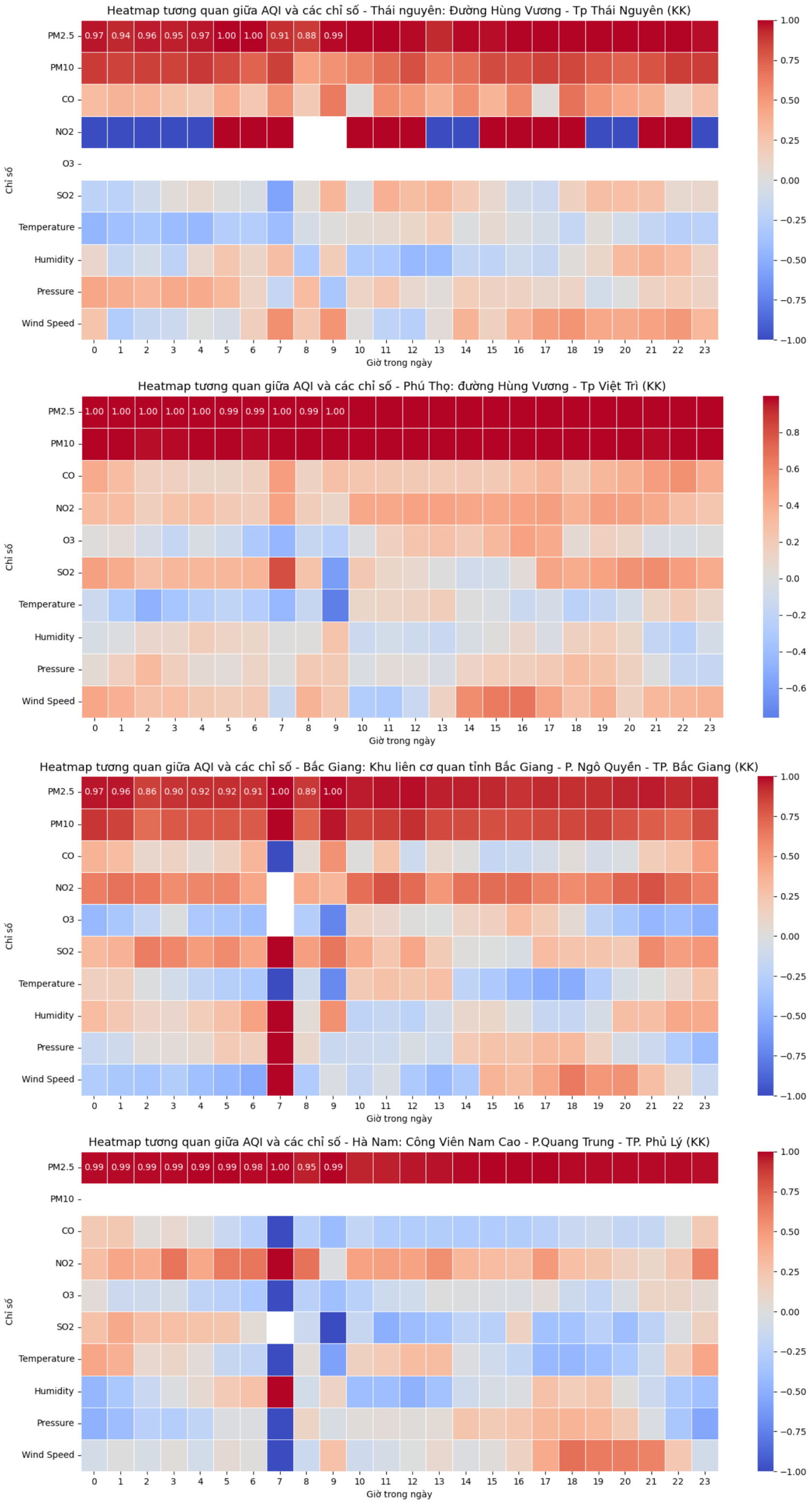

Figure E2: Hourly correlation of all major air quality indicators at Thai Nguyen, Phu To, Bac Giang, Ha Nam

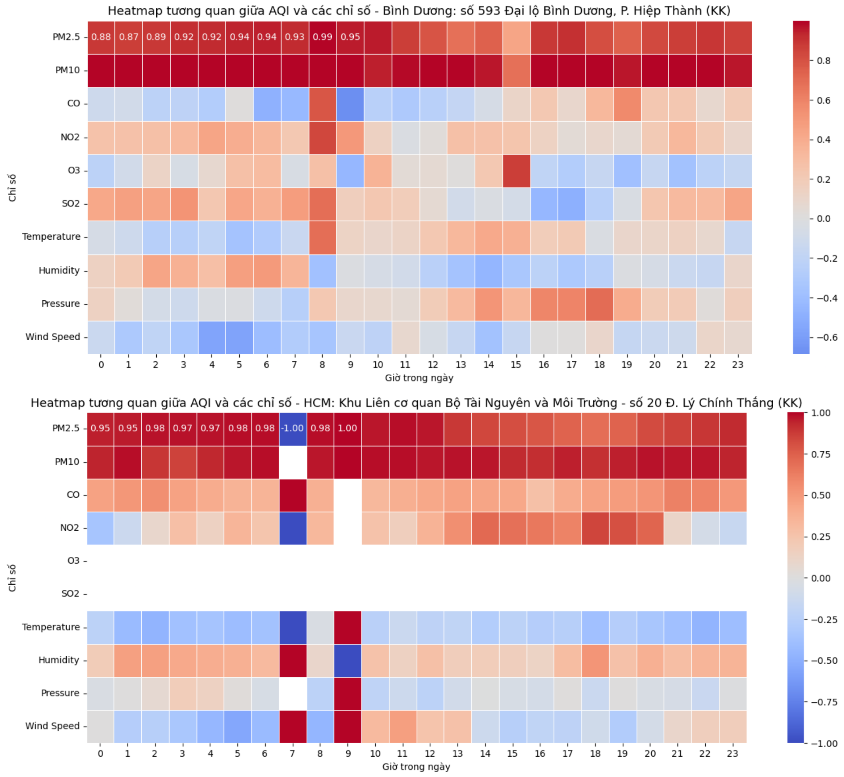

Figure E3: Hourly correlation of all major air quality indicators at Binh Duong, HCM - Bo Tai nguyen va Moi truong