NAIS Data Analysis

Celestine Atieno Oliewo

November 10, 2024

1. Data

The Neutral cluster and Air Ion Spectrometer (NAIS) was set up in the rooftop container at ISAC-CNR Bologna on the 16th of May 2024. The data has been collected continuously with regular cleaning.

1.1 NAIS Processor Python Package

The nais.processor module converts the NAIS data to NetCDF files. This was applied to the 'block' records and converted to daily NetCDF files at a 5-minute resolution. The data was further resampled to a 1-hour time resolution while combining the daily files into one file. It allows;

- inlet loss correction following the Gormley and Kennedy (1948) equations is significant for small particles. These are calculated as functions of particle size, flow rate, and geometry of the inlet. The inlet length, which is an input was measured as 85cm.
- ion mode correction, to account for biases in measuring the different modes (positive, negative, and neutral particles). The method by Wagner et al. (2016) is applied.
- conversion to standard conditions for temperature (293.15 K) and pressure (101325.0 Pa), unless otherwise specified.
- removes corona ions, thus removing background noise from the measurements.

The nais.checker module is applied to identify bad data and draw bounding boxes around bad data in the size distributions. The corresponding flags identify the bad data for each variable.
Using the nais.utils module, we set the bad data to NaN, and the resulting file is used for further analysis.

1.2 Complementary Data

Data for further analysis was downloaded from the ISAC server. All the data is in hourly resolution (UTC zone). These included some common pollutants (NO, NO2, NOx, black carbon, and ozone) and meteorological variables (temperature, Relative Humidity, total Rain, Wind Speed, Wind Direction, and air pressure).
The data preprocessing included removing any possible bad data, like negative values where they were improbable (replacing with NaN, or 0s).
Data for solar radiation was obtained as Hourly mean visible global irradiance - top component, from the environmental agency (ARPAE) site, for urban Bologna.

2. Analysis

2.1 NAIS Data Overview

The data used for this analysis spans from May 17th to October 29th, 2024. Below is the spectral plot of the variables measured by the NAIS, the particles data is the sum of the positively and negatively charged neutral particles.

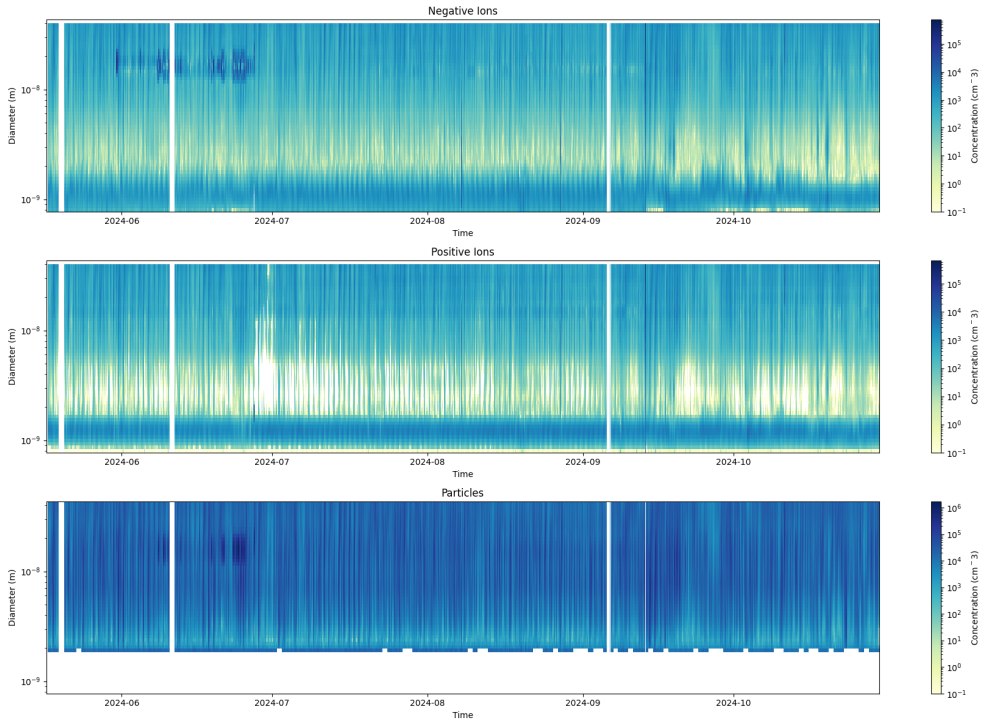
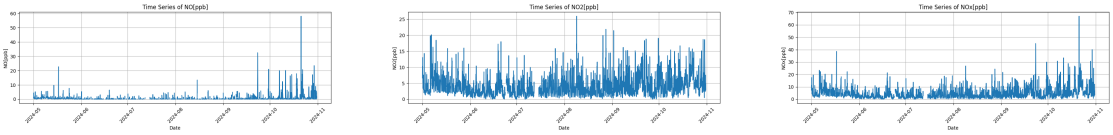


Figure 1: Spectral Plot

2.2 Complementary Data Overview

Time series for the common pollutants, running from May 1st to 30th Oct, while black carbon data runs from January.



(a) NO plot (b) NO2 plot (c) NOx plot

Figure 2: Nitrogen Oxides

(a) Ozone plot (b) Black Carbon plot

Figure 3: O3 and BC

3. Nanoparticle Ranking Analysis

The analysis in this section follows the method applied by Aliaga et. al (2023). This provides an alternative method of identifying new particle formation (NPF) events as opposed to the method proposed by Dal Maso et. al (2005) which may be tedious in the case of a larger dataset and also introduces human biases. The process involves several steps as below;

3.1 Extract data for $N_{2.5-5}$

A subset of the data for the diameter range 2.5nm to 5nm was extracted and used for further analysis. The concentration values were averaged over the diameter range and below is a plot of the variability.

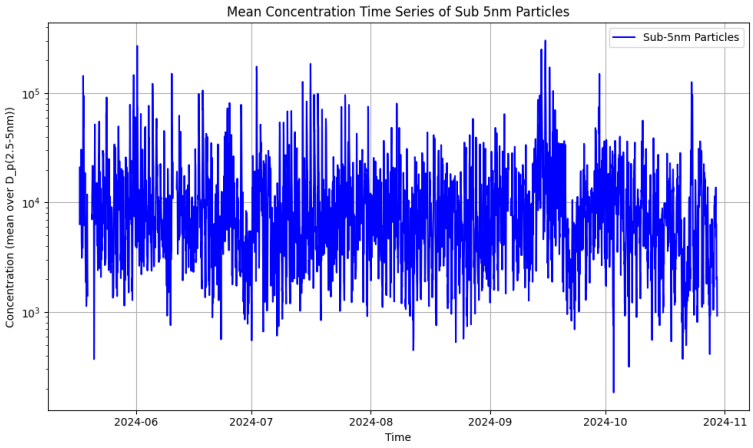


Figure 4: Time series for concentration Values

3.2 Smooth out the time series

Rolling median over 2-hour intervals was applied to the data to smooth the concentration values.

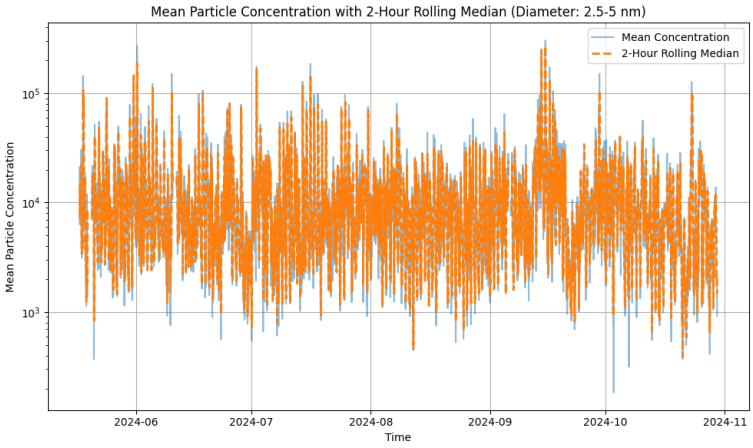


Figure 5: Time series for smoothed concentration Values

3.3 Identify diurnal background and active regions

The dataset was divided into months and the diurnal behavior was plotted as below;

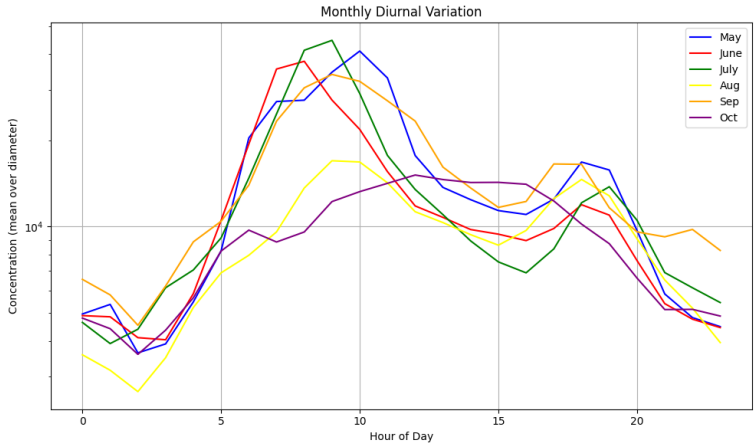


Figure 6: Diurnal Variation of the sub-5nm particles concentration by month

There is a peak in the mid-day hours between hours 6 and 12 for all the months with the exception of October and a smaller peak in the later evening at around 1800 hours. August records a notably lower concentration level in the mid-morning peak, which is almost the same level as the evening peak.
In October, the concentration levels rise slowly in the morning hours and stay rather constant through the day, before slowly declining in the late evening. Across the months, the background period appears around hours 1 and 3. All the hours referred to are UTC zone.

3.4 Find the background number concentration for each day $N_{B;2.5-5}$

The background concentration corresponding to a given day is determined based on the median value of $N_{2.5-5}$ in the so-called background region (0100 hours - 0300 hours).

3.5 Find the active peak daytime number concentration for each day $N_{A;2.5-5}$

This was calculated as the max value of $N_{2.5-5}$ in the so-called active region (0600 hours - 1200 hours).

3.6 Determine the daily change in number concentration $\Delta N_{2.5-5}$

The daily change for each day was calculated as the max value from the active region minus the median value from the background region.

$$\Delta N_{2.5-5} = N_{A;2.5-5} - N_{B;2.5-5}$$

Below is a time series plot of the daily change in concentrations for the selected sub-5nm diameter range.

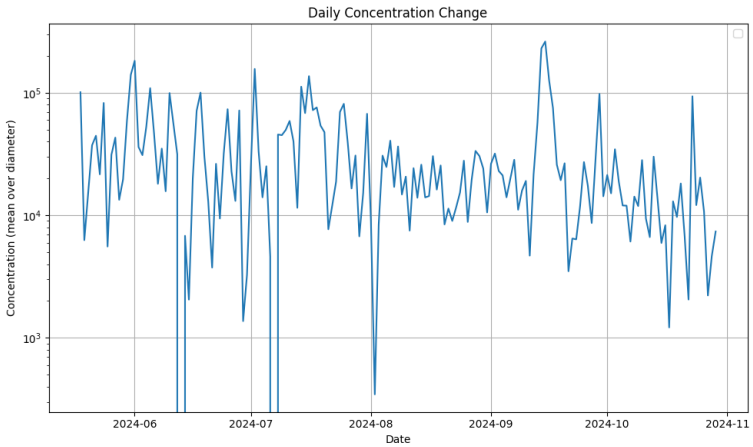


Figure 7: Daily Change Trend

The day with the maximum daily change was 15th September which looks like a Class I NPF event. Below is a plot of the variation in concentration for the sub-5nm range.

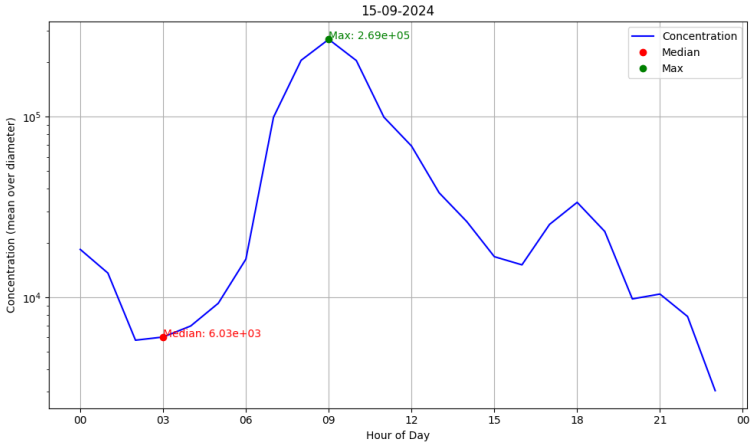


Figure 8: 15th September 2024

3.7 Rank and group the days

The daily change values were grouped into percentiles of 10%.
The days that recorded the median value in each percentile were plotted 9 to see the intensity of NPF events variation with the ranking process.

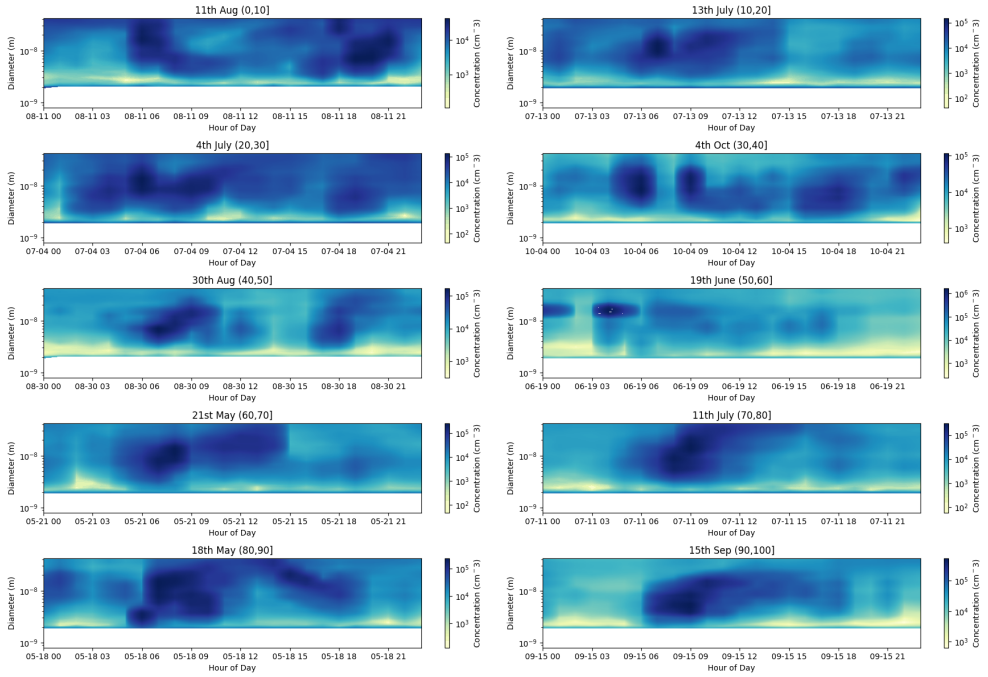
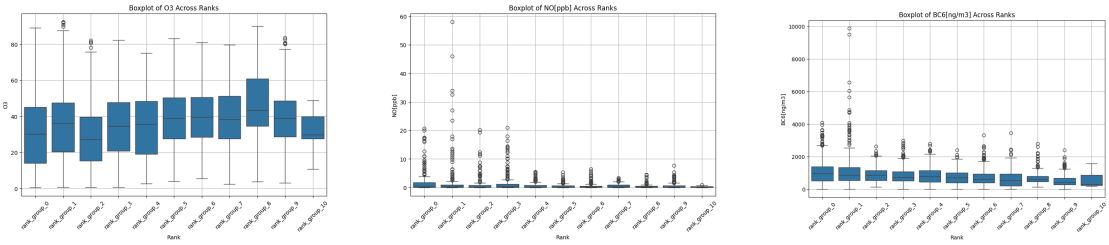


Figure 9: Intensity of NPF Events Variation with Rank

The plots show the days ranked from the 60th percentile can be categorized as NPF event days.

3.8 Complementary Data Ranking

To identify some factors that contribute to or inhibit the NPF events, we plotted the variations for some of the complementary data according to the ranks for the daily change values as described in the previous step.

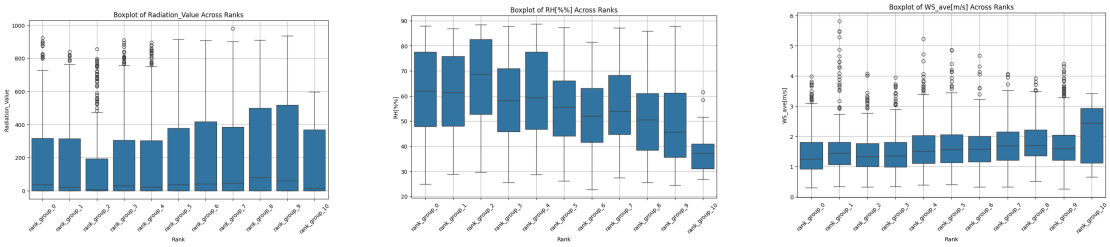


(a) Ozone Variation (b) NO Variation (c) Black Carbon Variation

Figure 10: Common Pollutants

There is no clear trend in the ozone variation with rank 10a. The concentration of NO is decreasing with the ranks 10b, while also the outliers reducing with rank. This supports

the fact that NO inhibits NPF events. The concentrations of black carbon also seems to reduce with rank, considering the median value from the boxplot 10c.



(a) Solar Radiation Variation (b) RH Variation (c) Wind Speed Variation

Figure 11: Some Meteorological Variables

The values for solar radiation appear to have a greater variation, the higher the rank while having fewer outliers, thus higher radiation values encourage NPF events.
Relative humidity on the other hand tends to reduce up the ranks, looking at the median values.
The values for wind speed do not seem to have a significant impact on the ranking of NPF events intensity.

4. NPF Mode Fitting

An alternative way to divide the days based on the NPF event intensity proposed by Aliaga et. al (2023) is depicting the log ($\Delta N_{2.5-5}$) distribution and the number of Gaussian curves needed to describe the distribution determined for the data subsets.
To determine the optimal number of Gaussian curves to fit to the data, we analyzed the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) curves to find a balance between model simplicity and data fitting.
The BIC curve penalizes model complexity more strictly, while the AIC curve focus more on fitting the data and may allow for more complex models.

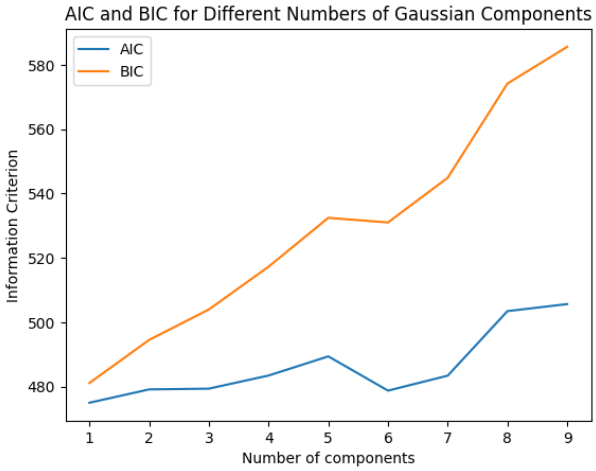


Figure 12: Optimal Number of Gaussian Components

The AIC curve has the minimum value at 6 components, while the BIC curve increases with components with no clear minimum. Based on these findings, we fit the data with 6 Gaussian curves 13, also indicating the number of days under each curve.

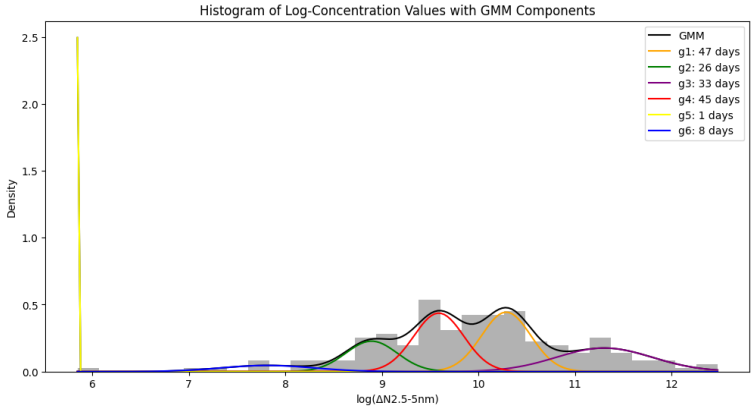


Figure 13: Log ($\Delta N_{2.5-5}$) fit to 6 Gaussian Curves

Based on the 6 groups, below is the spectral plot for the days with median values, based on the daily changes calculated earlier.

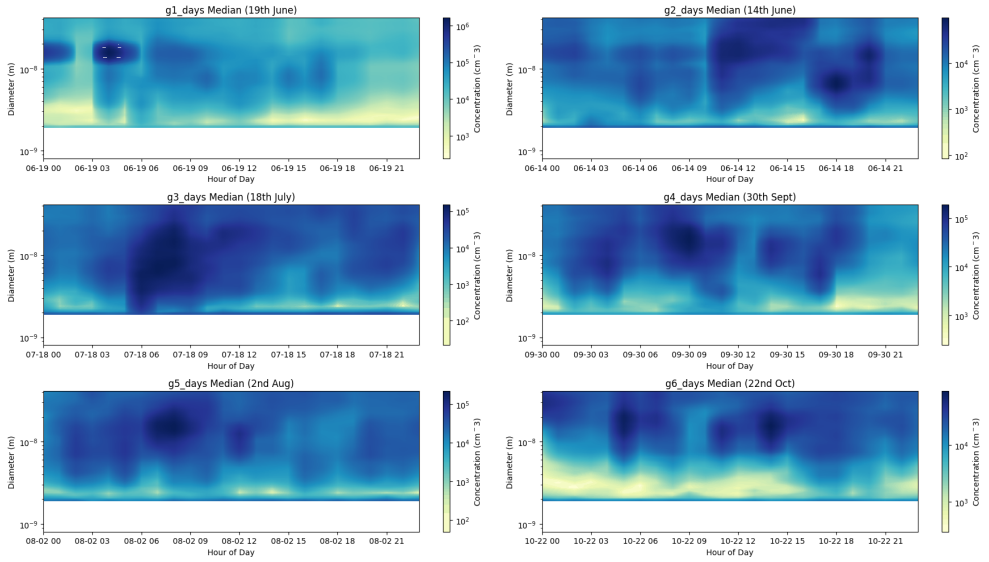


Figure 14: Gaussian groups

From the spectral plots, it is clear that the 33 days in group 3 are those to be classified as NPF event days.
The diurnal variation plots for each group 15 further shows the clear peak (growth) in the early morning hours for group 3, which we will now classify as NPF event days.

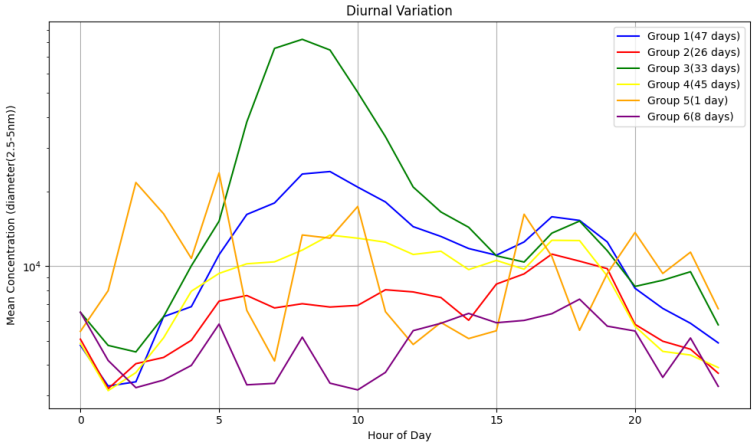
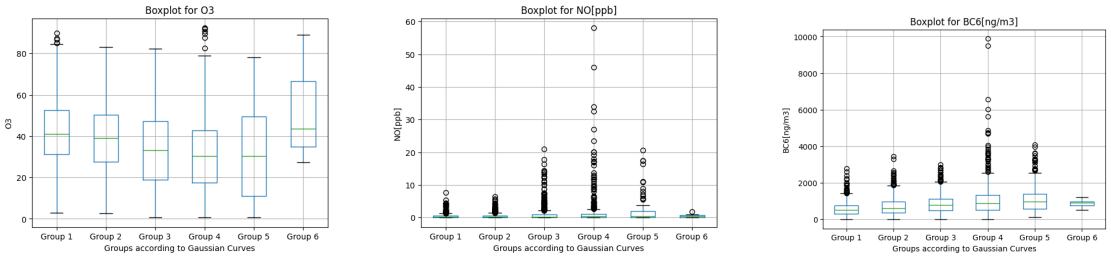


Figure 15: Diurnal Variations for the groups of 6 Gaussian Curves

The 47 days in group 1 could represent weaker NPF event days since they display a small peak in the mid-morning as well.
We could classify the group 3 days as local NPF event days and the group 1 days as regional NPF event days.

4.1 Variation for Complementary Data

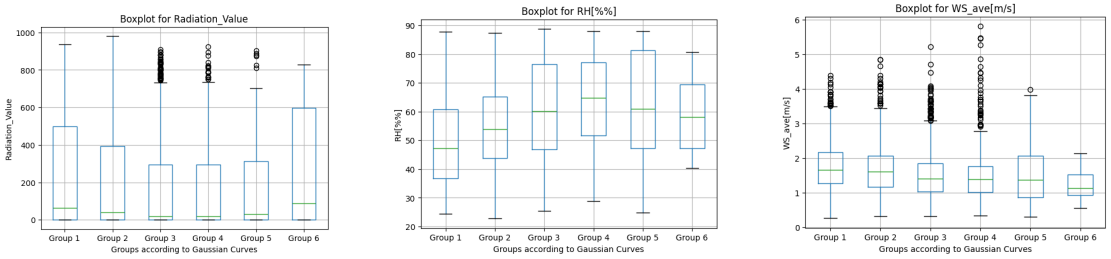
From plot 13, we divided the complementary data into 6 groups, based on the log values. That is; g3 as group 1 (highest values from the histogram), g1 as group 2, and so on. With this grouping, we plotted the boxplots for the variables to see how they vary for each group.



(a) Ozone Variation (b) NO Variation (c) Black Carbon Variation

Figure 16: Common Pollutants

Unlike in the ranking method above where there was no clear trend with ozone 10a, using the mode fitting method, there is a clear downward trend in ozone concentrations, being higher for stronger NPF event days.
Black carbon and NO show an upward trend, up the groups, to imply they inhibit NPF events thus lower concentrations are recorded on NPF event days.
The high concentrations recorded in group 4 could represent increased combustion or traffic, which is also recorded for black carbon in the same group.
Some interesting meteorological variables were also plotted below.



(a) Solar Radiation Variation (b) RH Variation (c) Wind Speed Variation

Figure 17: Some Meteorological Variables

The days with stronger NPF events record higher values of solar radiation, considering the reducing size of the boxplots from groups 1 to 4. This implies the higher radiation values contribute to the NPF events and their intensity.
The values of relative humidity on the other hand are lower on NPF event days. This implies high radiation and lower humidity favor photochemical reactions leading to NPF

events.
The median values of wind speed are quite low for all groups, but the outliers on the positive side increase with groups implying NPF events are favored by more stable air masses.

From the above analysis applying the ranking method and mode fitting, the mode fitting method seems to better categorize the data according to NPF event intensity.