Analysis of Air Quality Data of Nanning in 2019

1 Background

With the development of society, the problem of air pollution has become increasingly serious. Population health risks caused by it have attracted wide attention from the society. Typical areas with rapid economic development, represented by China's first-tier and second-tier cities, are densely populated, greatly affected by anthropogenic emission sources, and have a large exposure risk of air pollution. So, I collect the air pollution data of Nanning in 2019, and analyzes the temporal air quality characteristics of Nanning city according to the hourly data. The first step is to analyze the time series and obtain the temporal distribution characteristics of the air quality index of Nanning.

Figure 1 the position of Nanning city

(source: https://www.sohu.com/a/117148270_394146)

In the atmosphere, a variety of pollutants can interact with each other in a complex way to produce secondary pollutants, which are transmitted over long distances with meteorological conditions, causing regional pollution [1]. According to the survey, in 2015, the annual mean of PM2.5 and PM10 in 299 cities in China reached the second-level standard of Ambient Air Quality Standard (GB 3095-2012), which was 24% and 38% respectively [2]. In 2017, nearly 75% of the 74 major cities in China still failed to reach the annual average PM2.5 concentration limit (GB 3095-2012), and in 2018, the annual average PM2.5 concentration in Beijing-Tianjin-Hebei and its surrounding areas and the Fenhe and Weihe River Plain exceeded 71% and 66% of the annual average PM2.5 concentration limit (GB 3095-2012), respectively [3]. Although China has taken strict measures to control air pollution emissions, it is still one of the countries with the most serious air pollution in the world at present [4]. Atmospheric pollutants mainly enters the body through the respiratory system and breathing, blood circulation and immune to the human body caused serious harm, exposed to the people at the same time is affected by all kinds of pollutants in the air, has a large number of reports that mortality increased air pollution not only can directly lead to population, with a variety of diseases such as lung cancer, cardiovascular disease and higher incidence of respiratory disease on [5-7]. Air pollution has caused great losses to the safety of life and property of Chinese residents. Studies have shown that in 2006, residents of 113 cities in China suffered great health losses due to PM10 pollution, with a health economic loss of RMB 341.403 billion [8]. About 1.1 million deaths in China in 2015 can be attributed to long-term exposure to PM 2.5 pollution [9].

2 Data Used

Air pollution data: The air pollution data in this paper were collected from the air quality online monitoring platform (https://www.aqistudy.cn), the data source of which was the air quality online real-time publishing system platform of China Environmental Monitoring Station.

3 Data Analysis

This data set has been processed in the early stage, and the pre-processing rules of the data will not be detailed. Here, we will discuss the graphical display method. First, draw the time series diagram to make intuitive judgment:

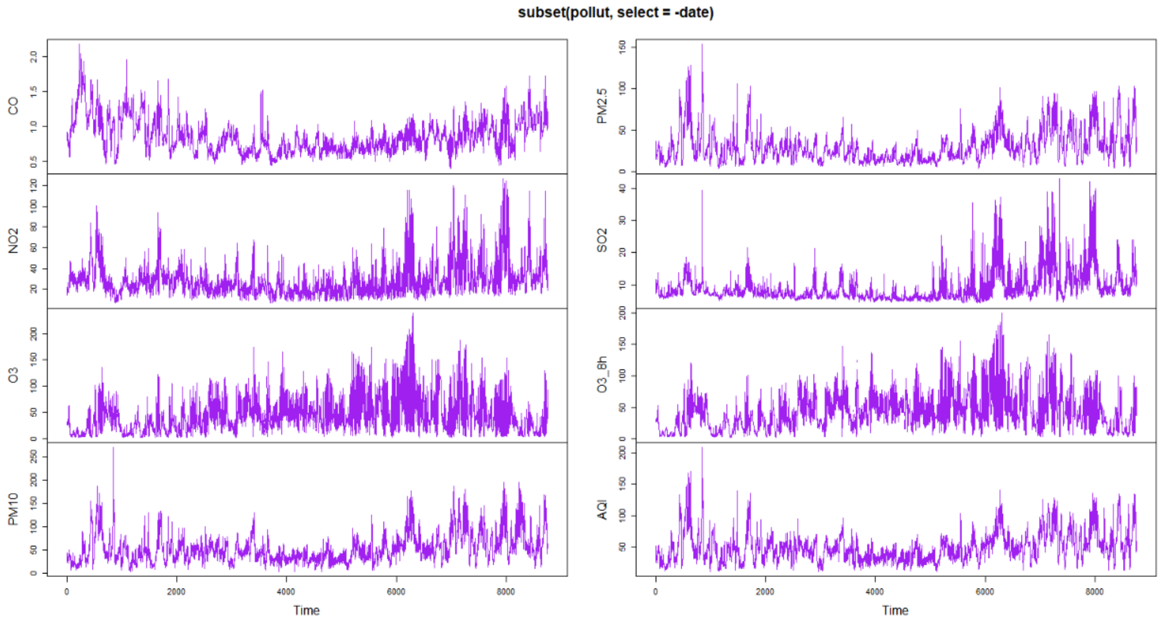


Figure 2 Full review of the data

We can also quickly summarize the data as a whole using the SummaryPlot () function, which graphically displays time series variations, statistical indicators, and frequency plots:

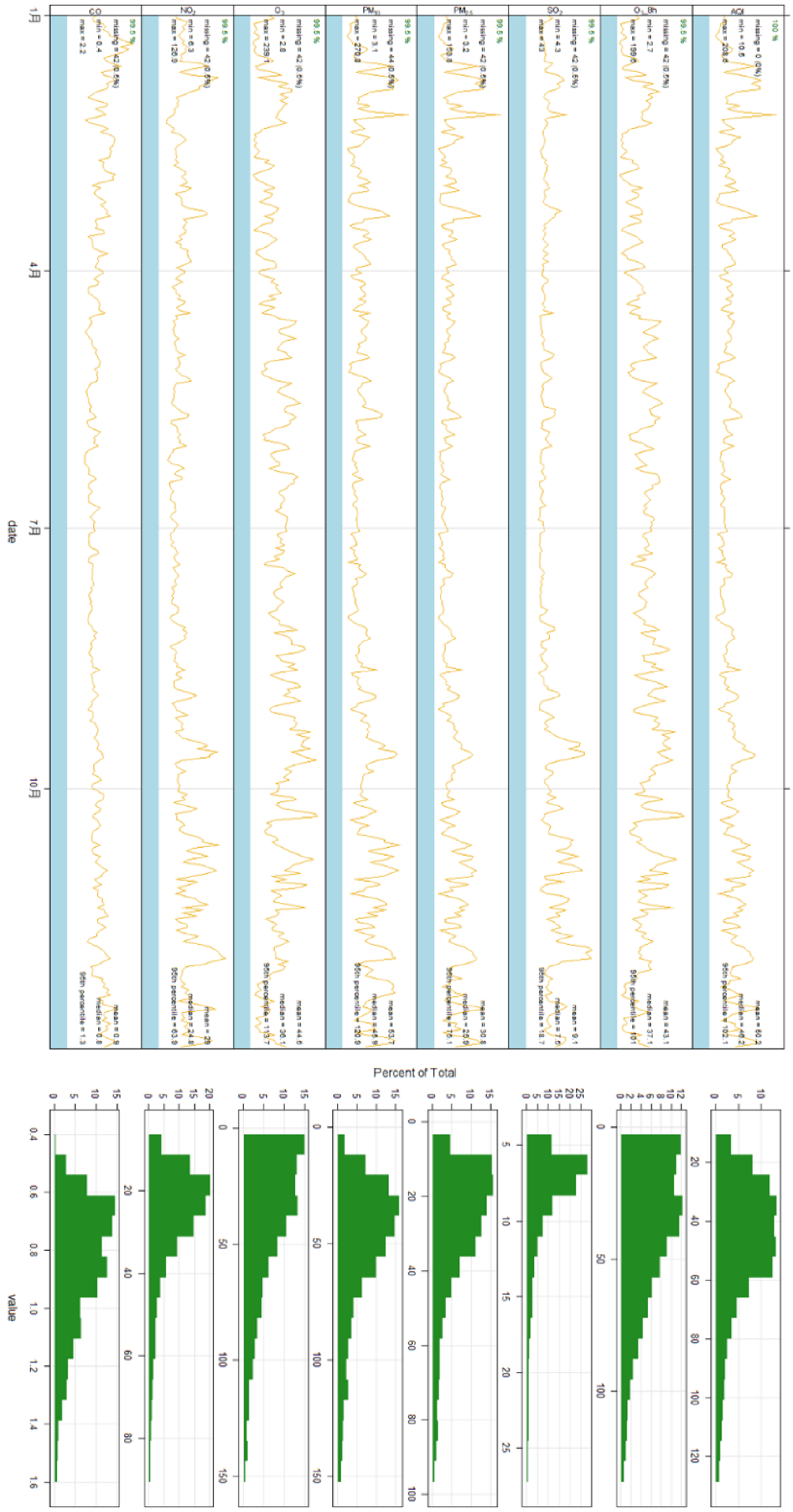


Figure 3 Time series, frequency distribution and statistical analysis (mean, median /95% interstitial value, minimum/maximum value, data amount and proportion of missing value)

Compile pollutant calendar map according to the PM2.5 concentration limit corresponding to the air quality sub-index in the Ambient Air Quality Index (AQI) Technical Regulations (Trial) (HJ633-2012) issued by the Ministry of Ecology and Environmental Protection, the daily mean PM2.5 concentration was divided into six levels and the numerical range and color definition were given respectively:

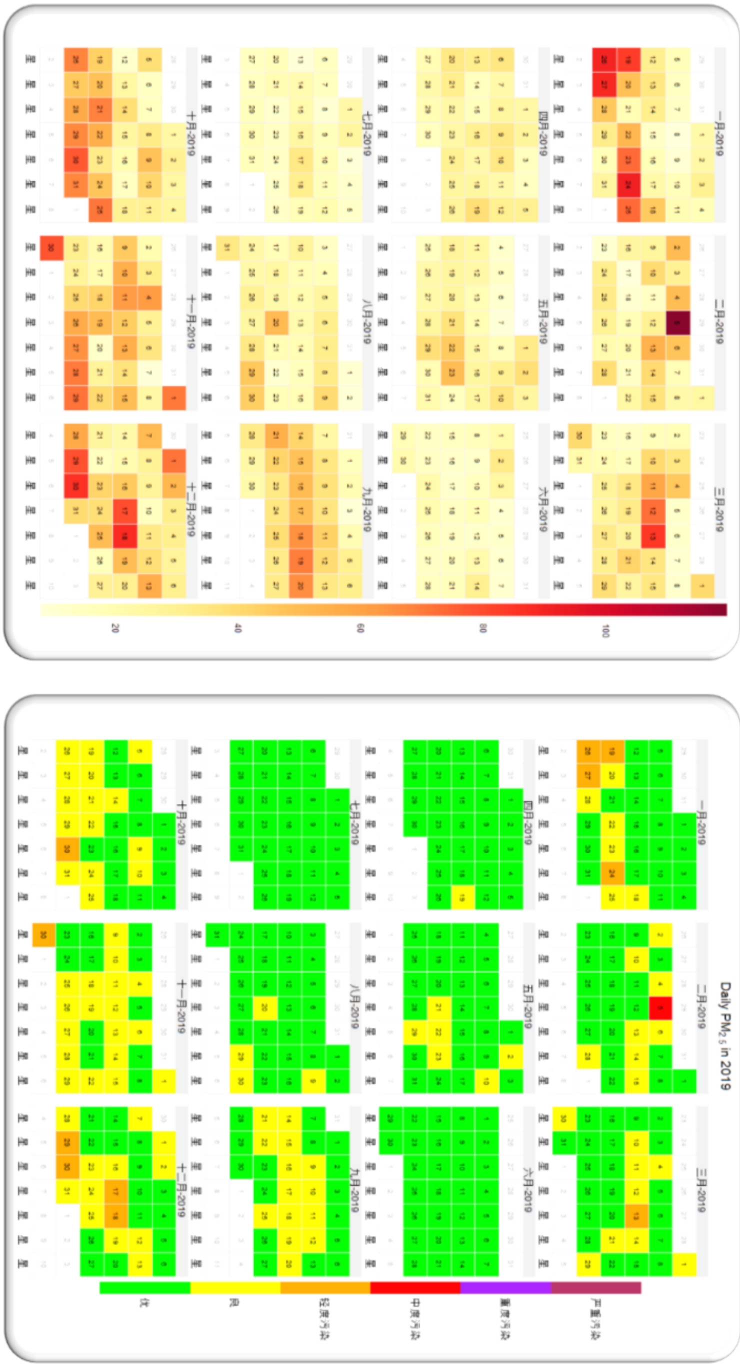


Figure 4 Daily concentration of PM2.5

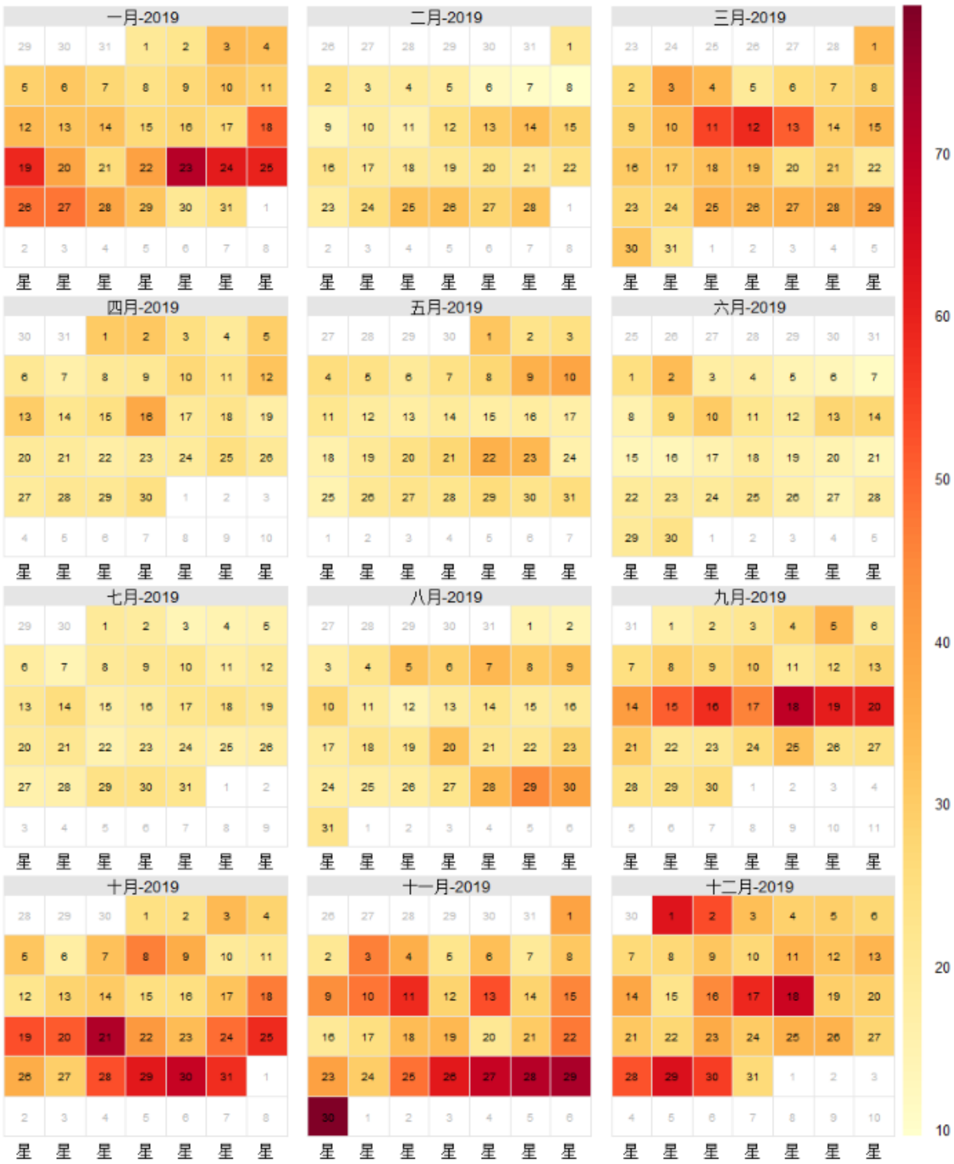


Figure 5 Daily concentration of NO2

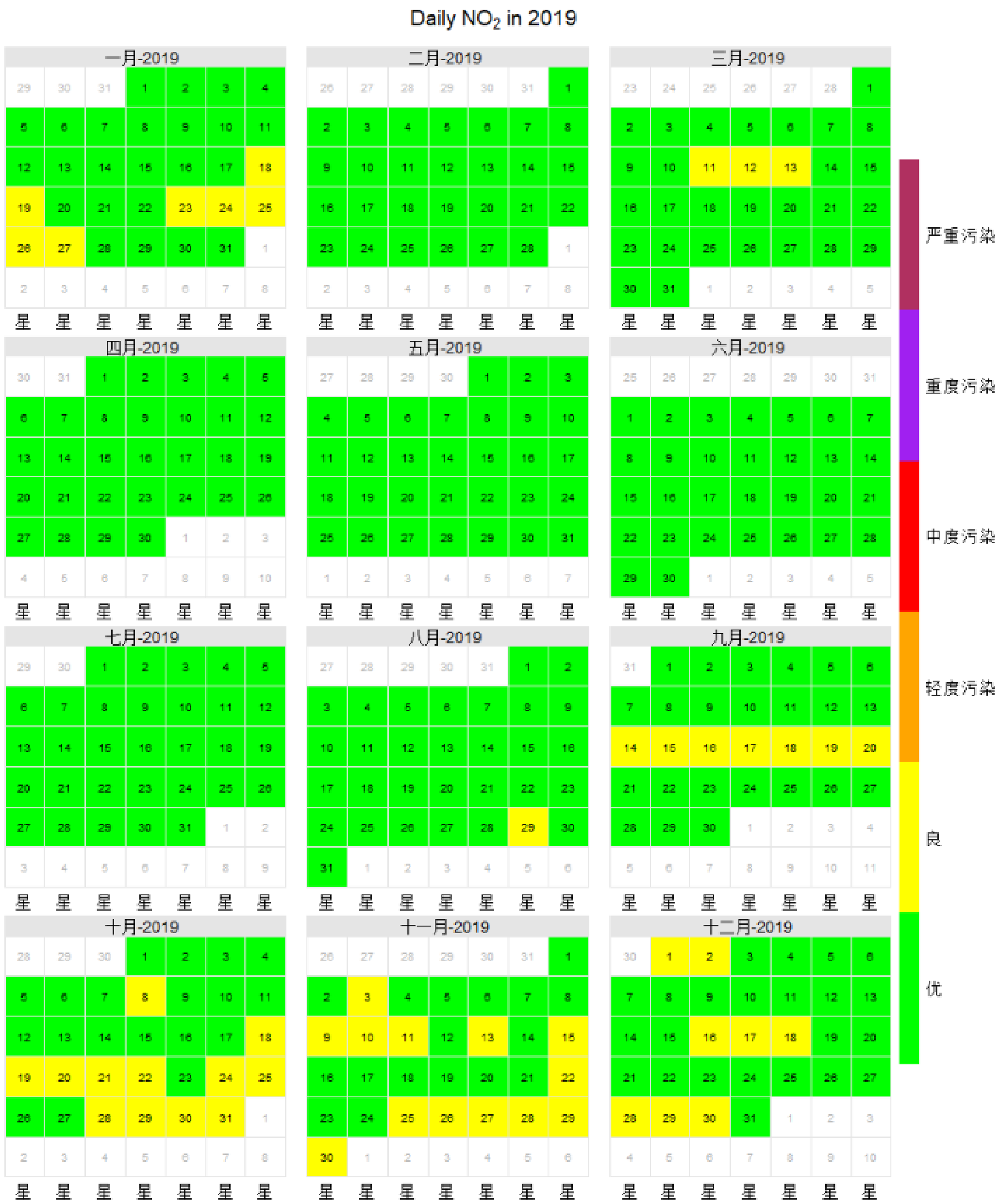
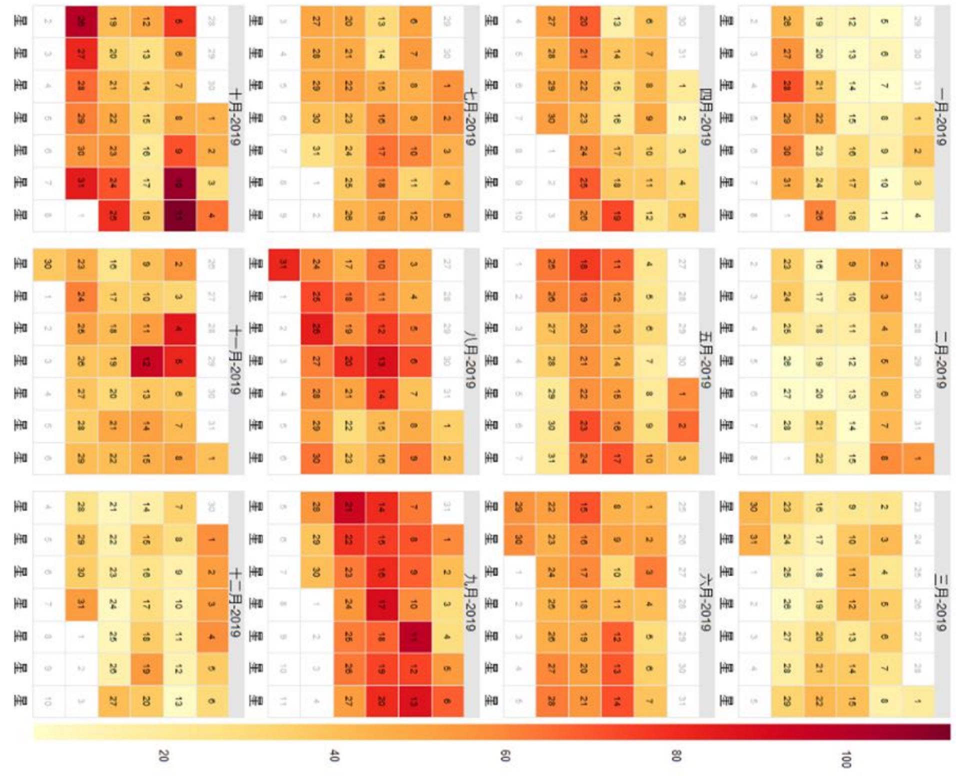


Figure 6 Daily concentration of NO2 after classified



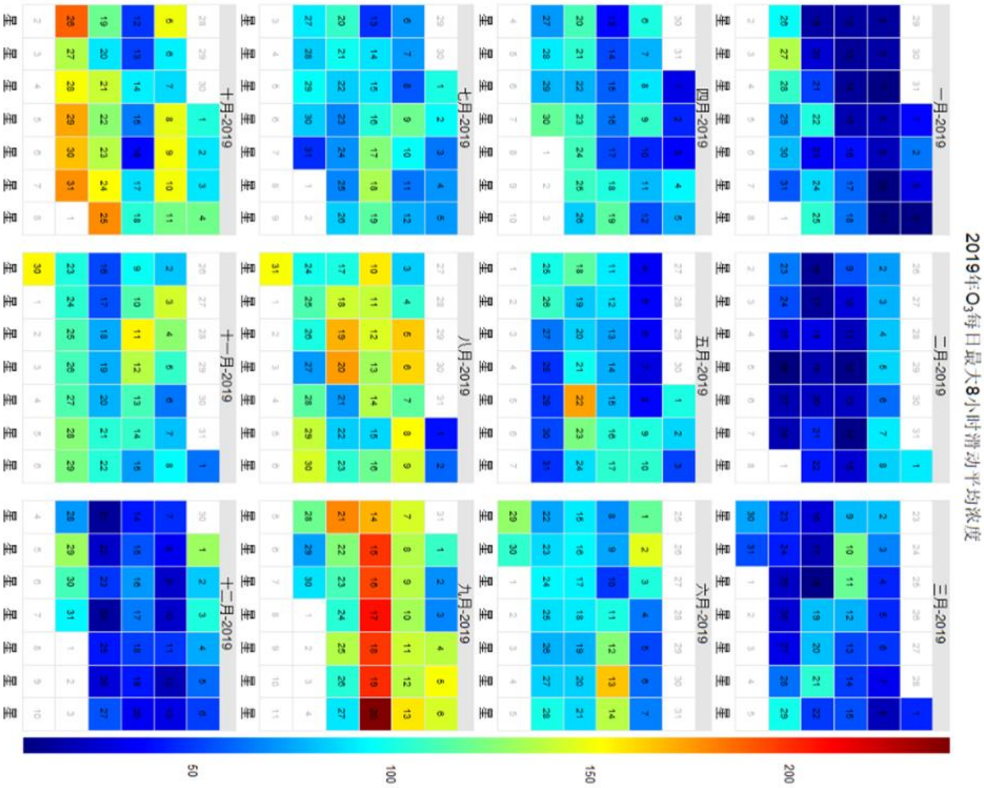


Figure 7 Daily concentration of O3

We can also use Trendlevel( ) to plot the time variation trend of pollutants:

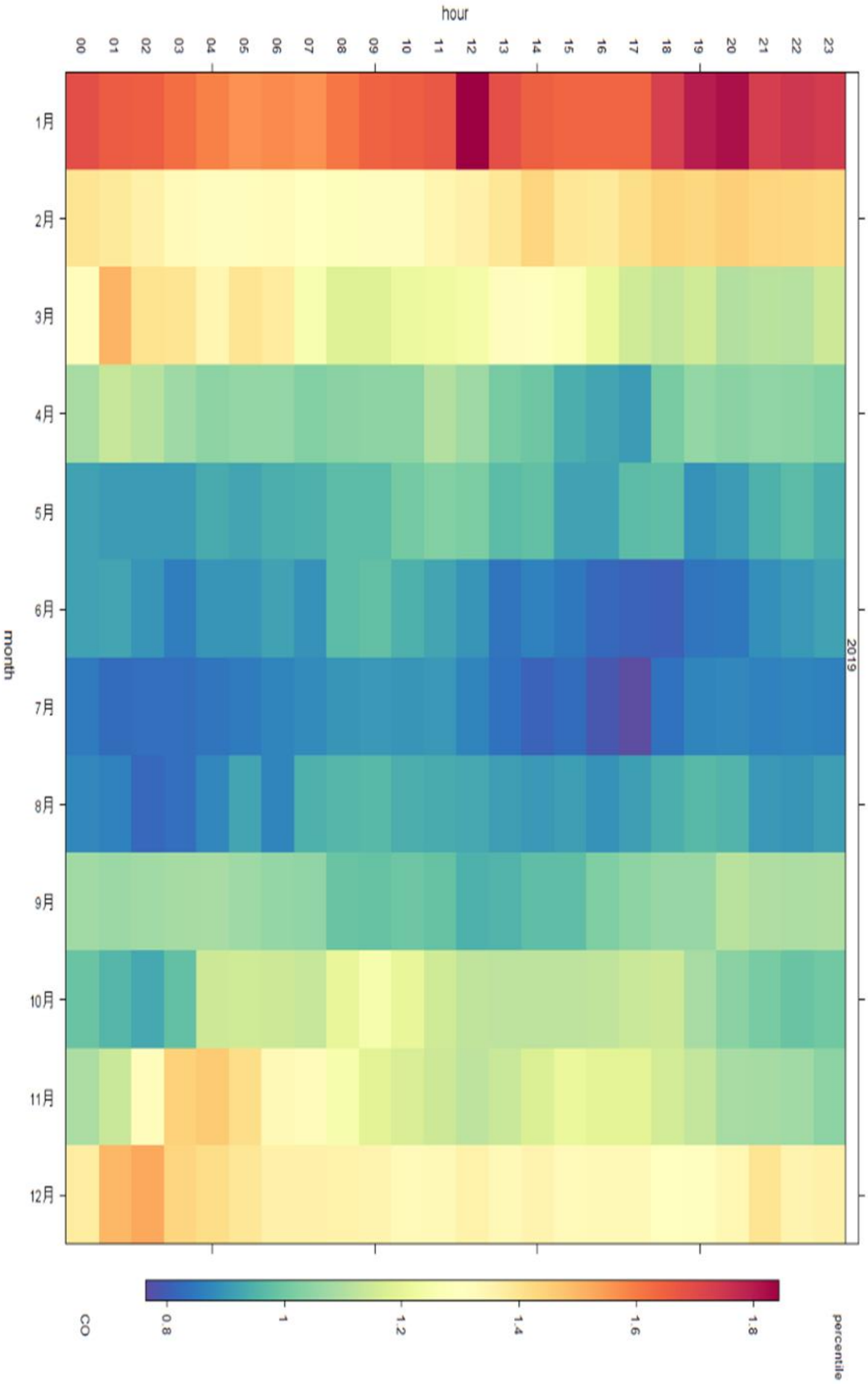


Figure 8 PM2.5_Daily_Monthly variable trend

Time series analysis is a commonly used analysis method of atmospheric monitoring data, including the analysis of seasonal, monthly, week-day, hourly changes and other characteristics, in order to reveal the temporal change rule of pollutants and predict the change trend.

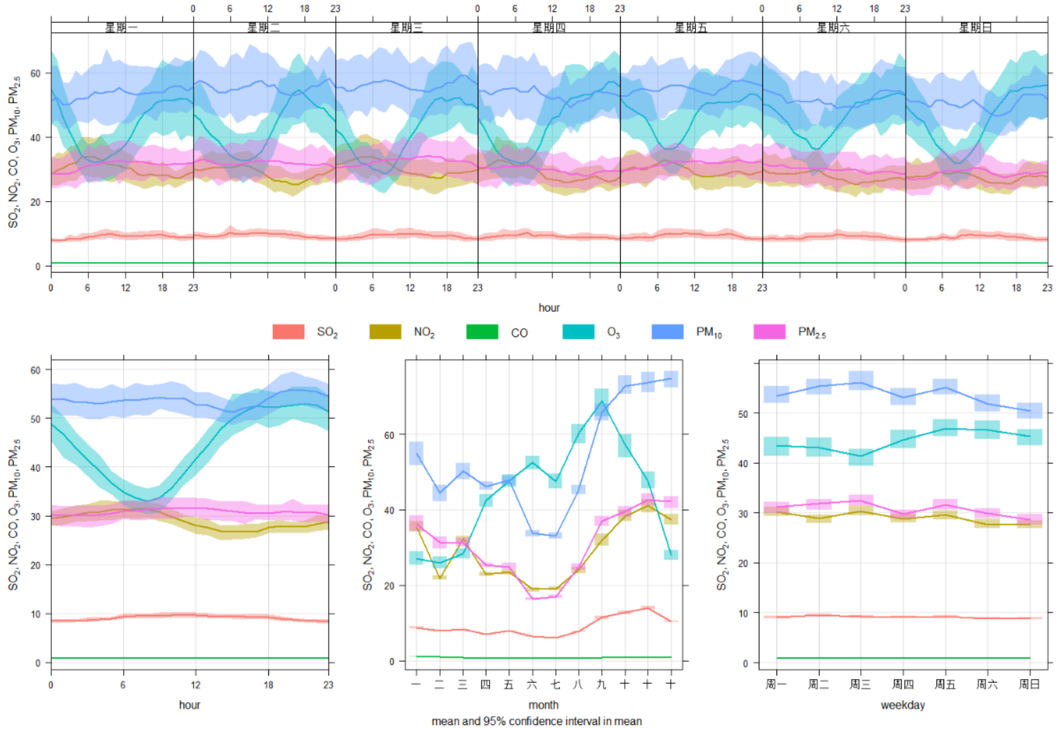


Figure9 (a) Daily_Weekly_Monthly

It is easy to cover up the change rule of low-concentration pollutants (such as SO2) in the same graph with multiple pollutants of different orders of magnitude. TimeVariation function provides a method to standardize data of different orders of magnitude (divide by their mean values to achieve comparison in the same numerical range). Normalise = True (FALSE by default) is selected:

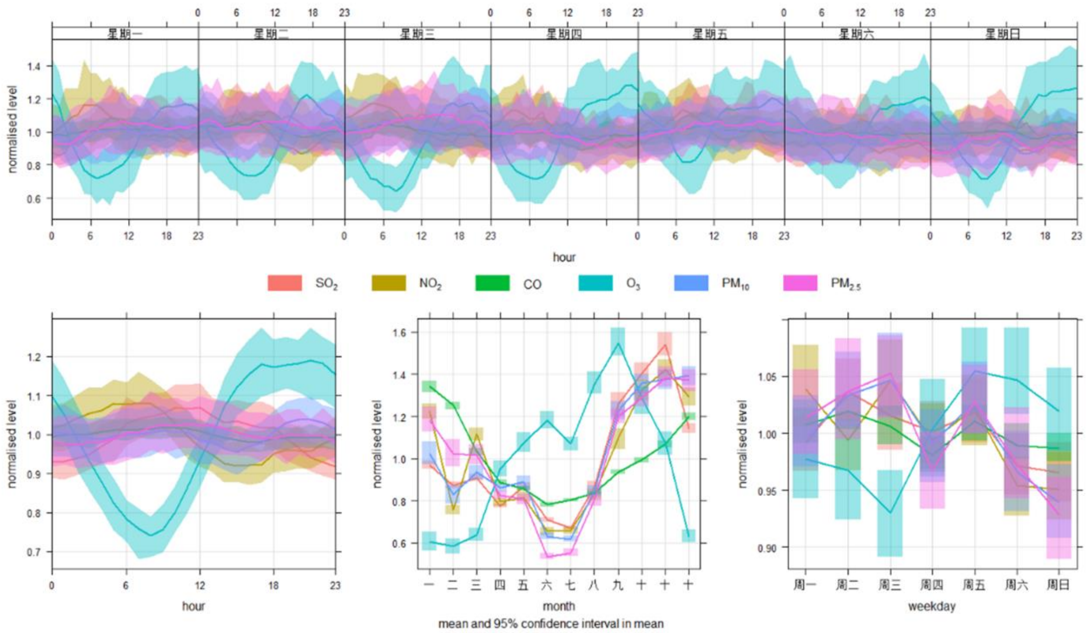


Figure 9 (b) Standardized time variations of six pollutants (mean and 95% confidence interval)

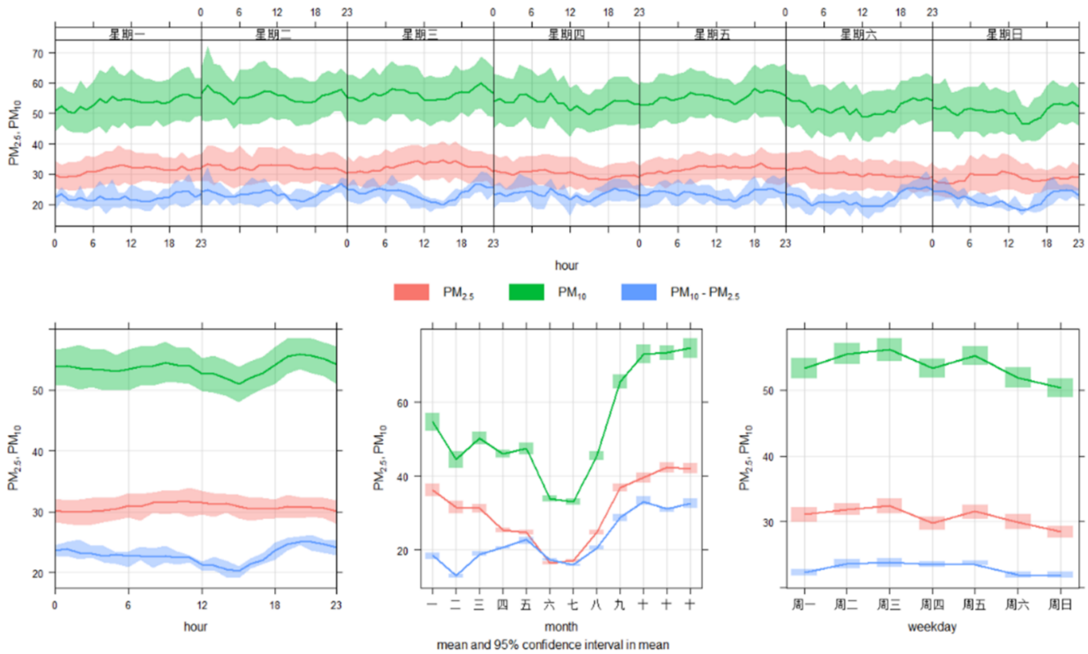


Figure 10 PM2.5 PM10 Difference

This function also provides the function of calculating the difference value. We use the parameter difference to calculate the concentration of coarse particulate matter and show its time change. It can reflect the change characteristics of coarse particulate matter: the monthly change shows that the concentration of December and November is higher, indicating that the dust pollution is heavier in autumn; The diurnal variation showed that the concentration was higher in the morning and evening travel periods, indicating that this point was affected by road dust sources. It can be clearly seen that the concentration of coarse particulate matter peaks in the morning and evening travel peak hours of the four seasons, indicating that road traffic dust is an important source of coarse particulate matter at this point. It can also be clearly seen that the peak time in winter is significantly delayed compared with other seasons, reflecting the seasonal changes of the dawn and dusk line and the length of day and night.

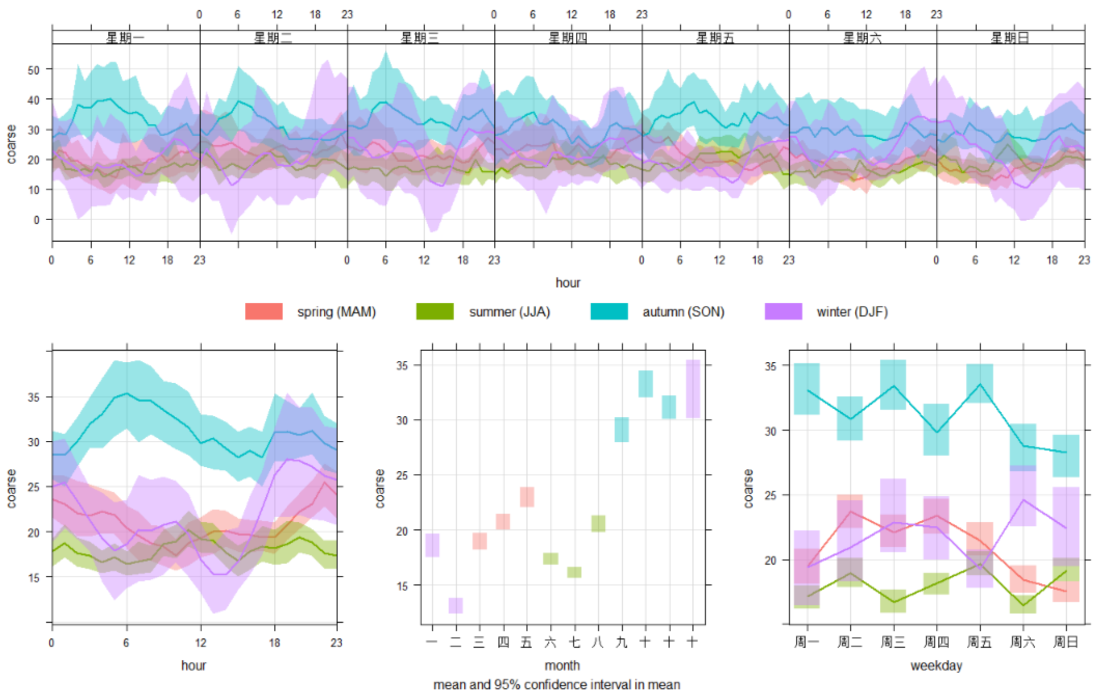


Figure 11 Time series of the coarse particles（PM2.5、PM10）

In addition, the functional parameters of the GROUP( ) can realize further combined processing of data, such as drawing data of different seasons and point locations at the same time. We use the parameters of the GROUP( ) to draw seasonal time changes of the coarse variable:

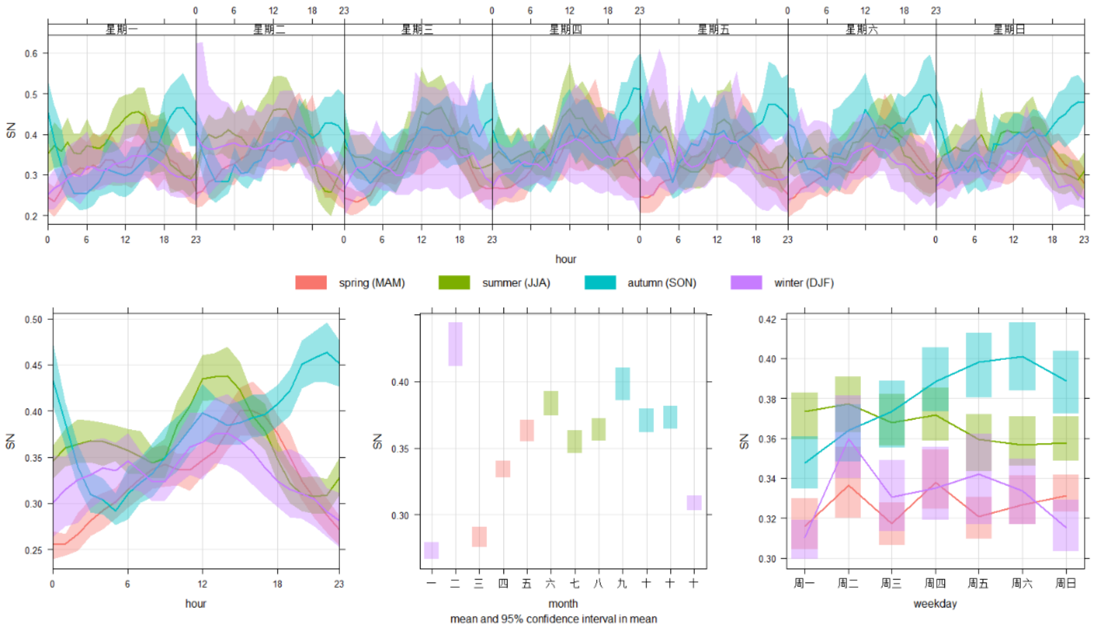


Figure 12 Time series of SN（SO2/NO2)

The partitioning function of ggplot2 was used to display the time change sequence of each pollutant:

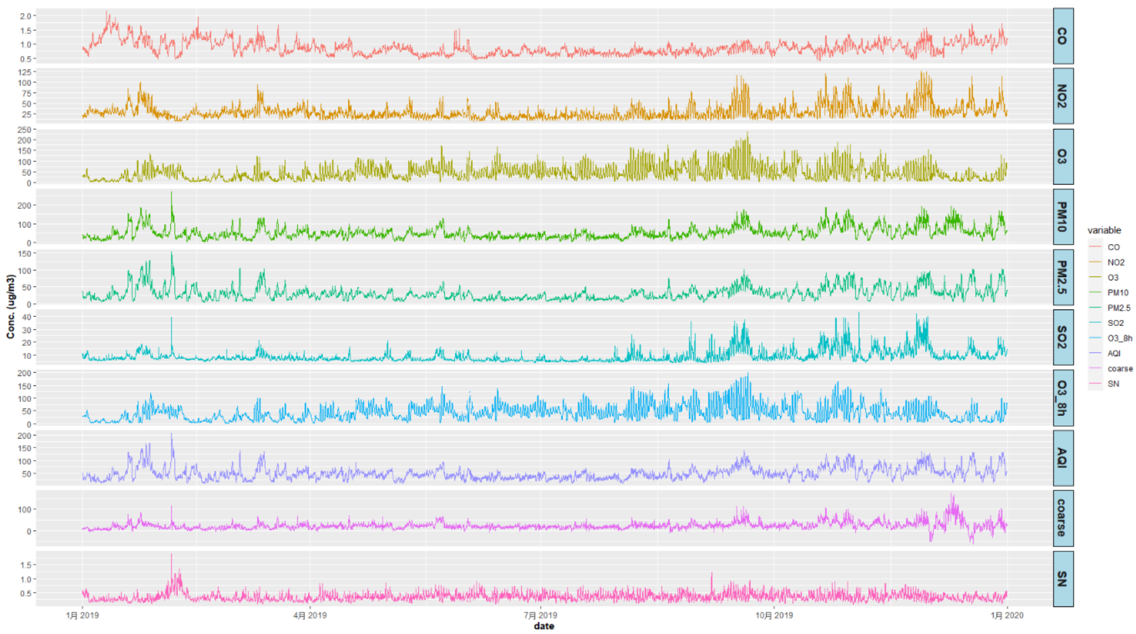


Figure 13 Time Series of kinds of Pollutants respectively

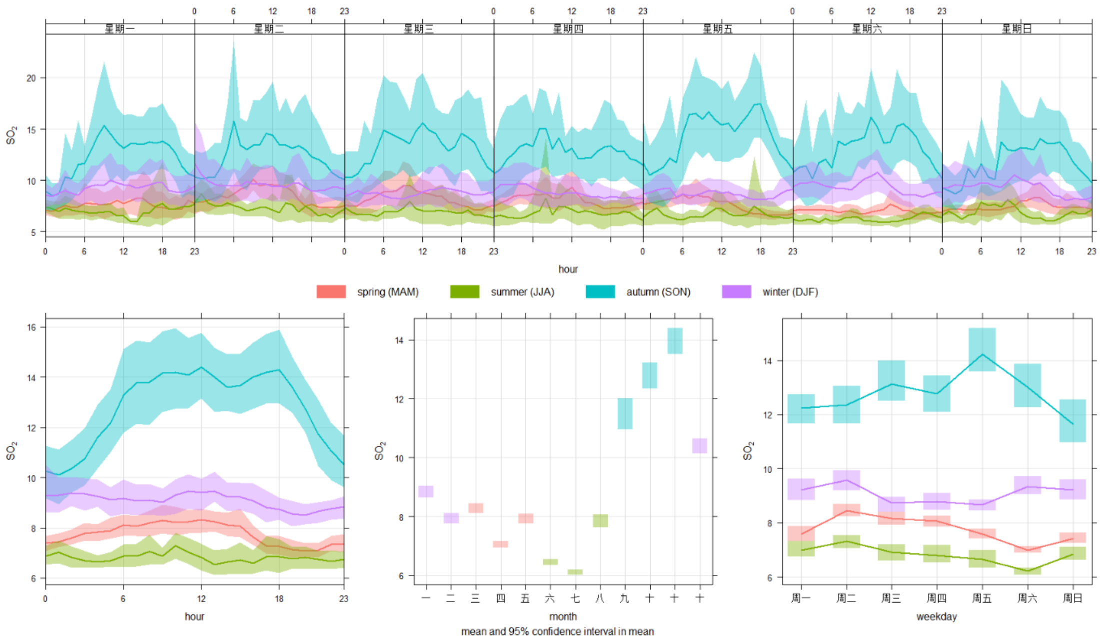


Figure 14 Time series of SO2

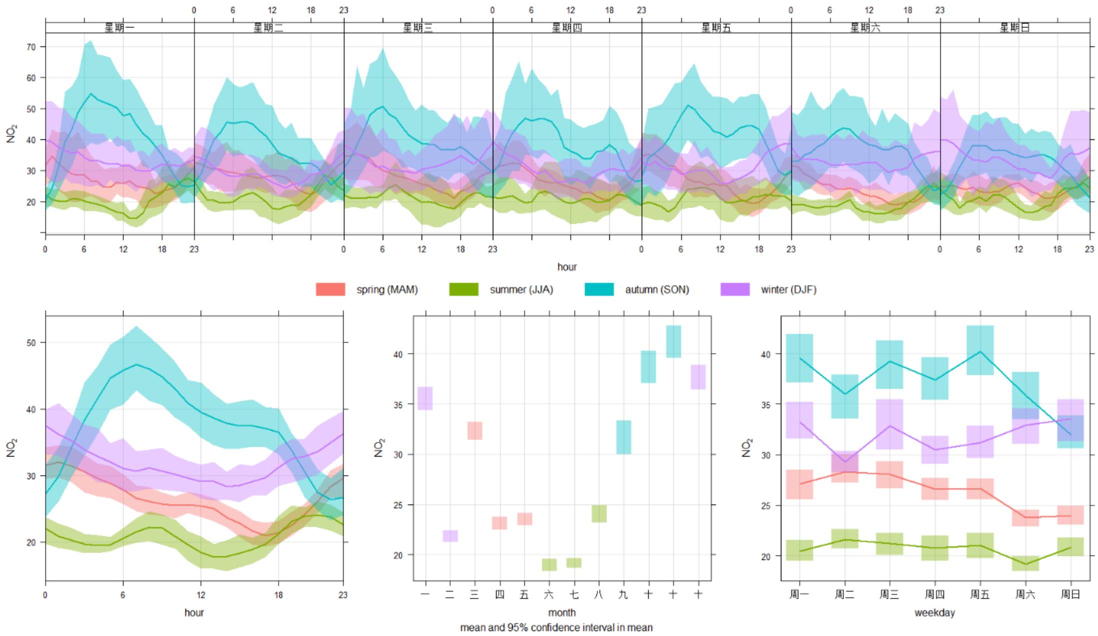


Figure 15 Time series of NO2

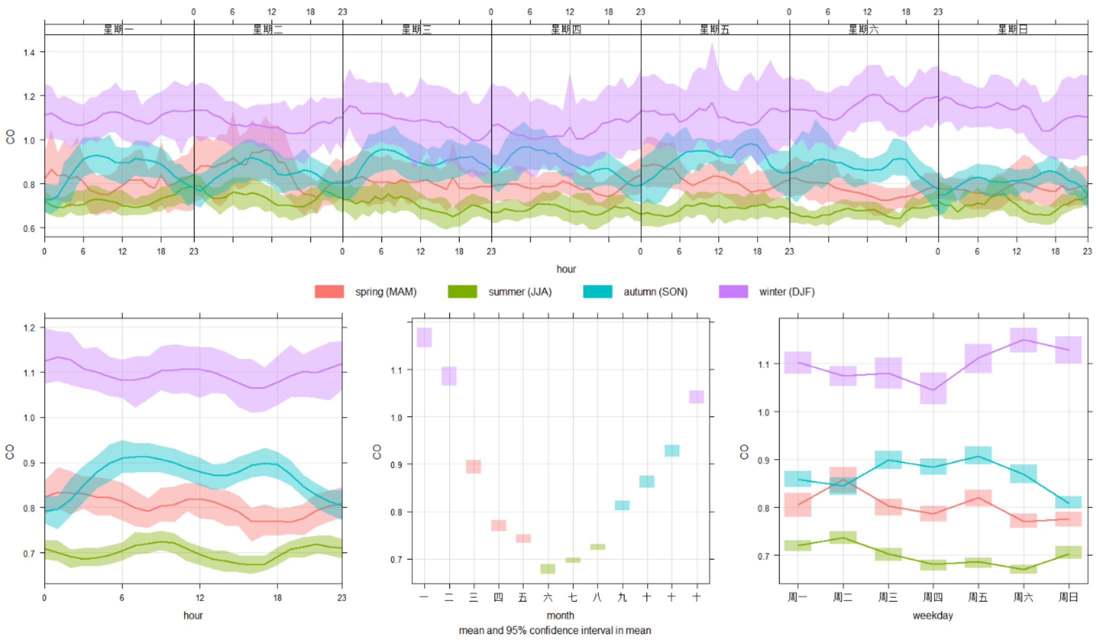


Figure 16 Time series of CO

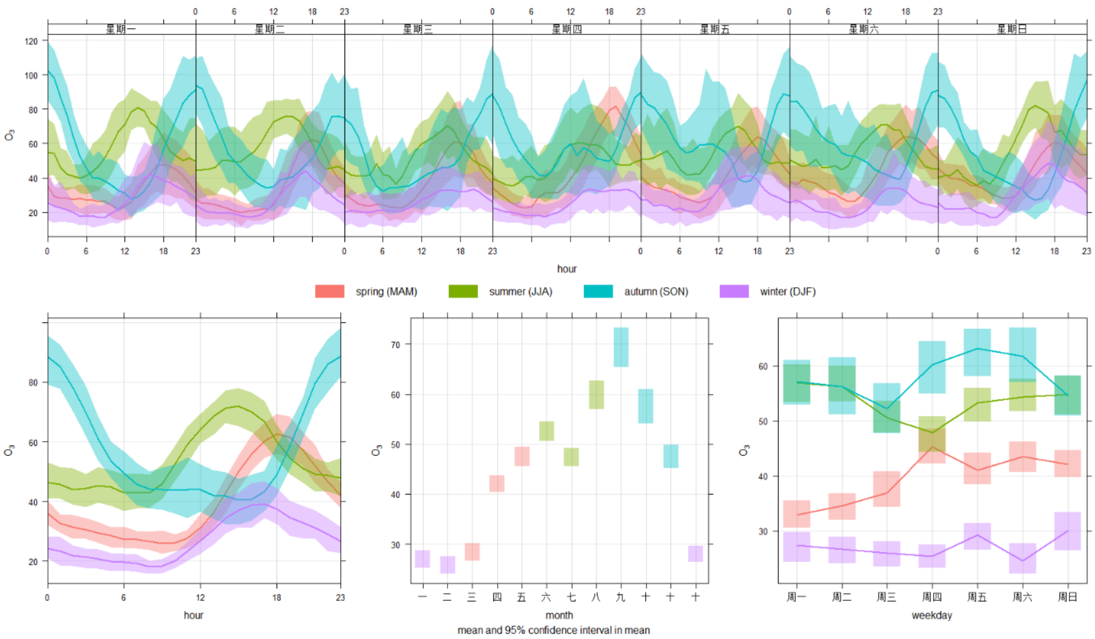


Figure 17 Time series of O3

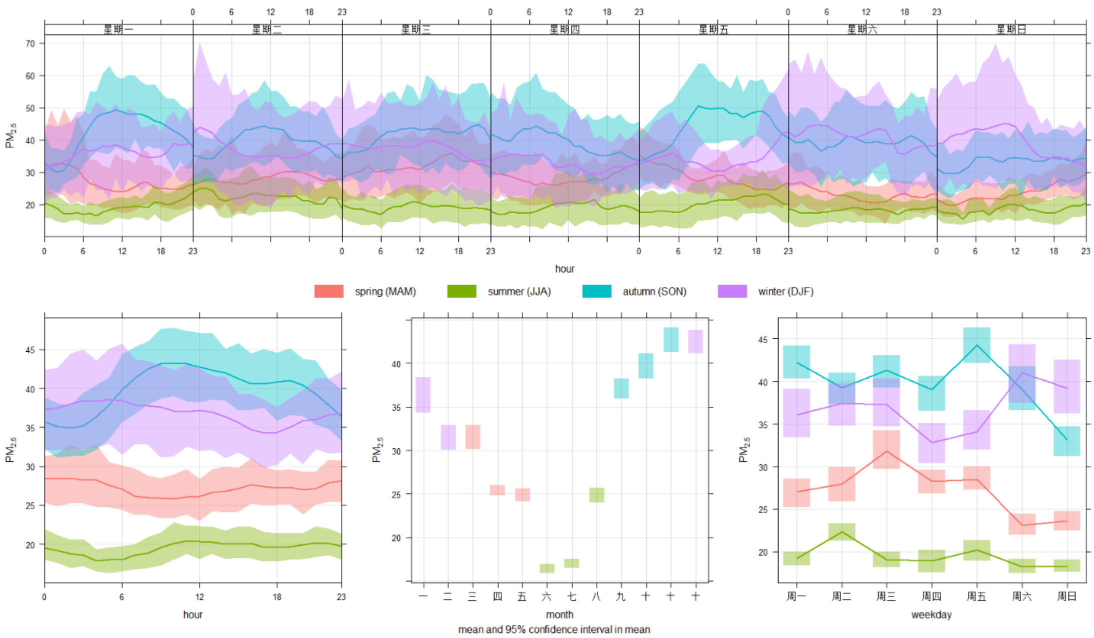


Figure 18 Time series of PM2.5

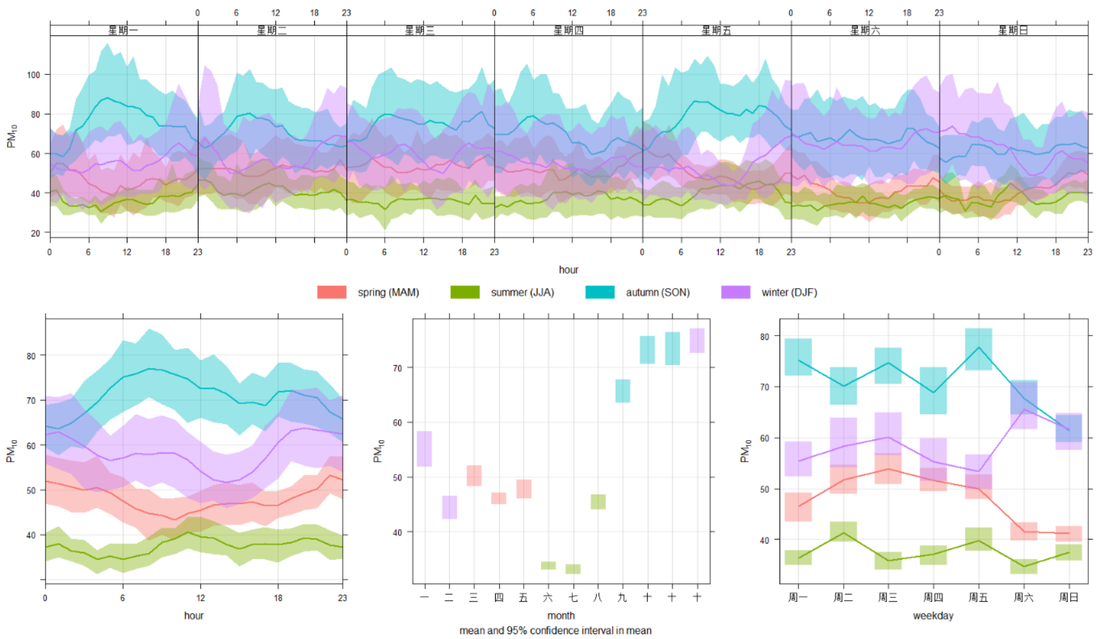


Figure 19 Time series of PM10

Using the partitioning function of GGPLOT2 (graphic parameters of GGPLOT2 will not be introduced here but in R), the average diurnal variation of each pollutant in different seasons is plotted:

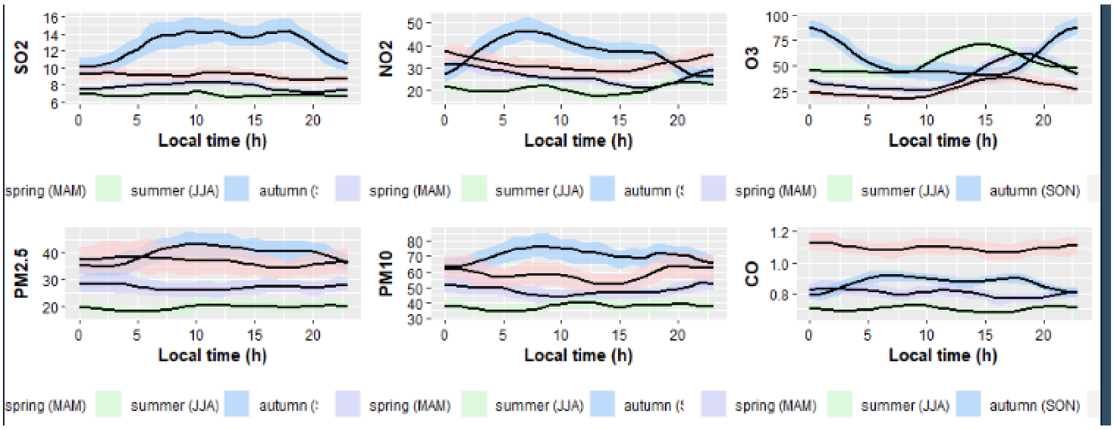


Figure 20 Daily change of all pollutants in different seasons

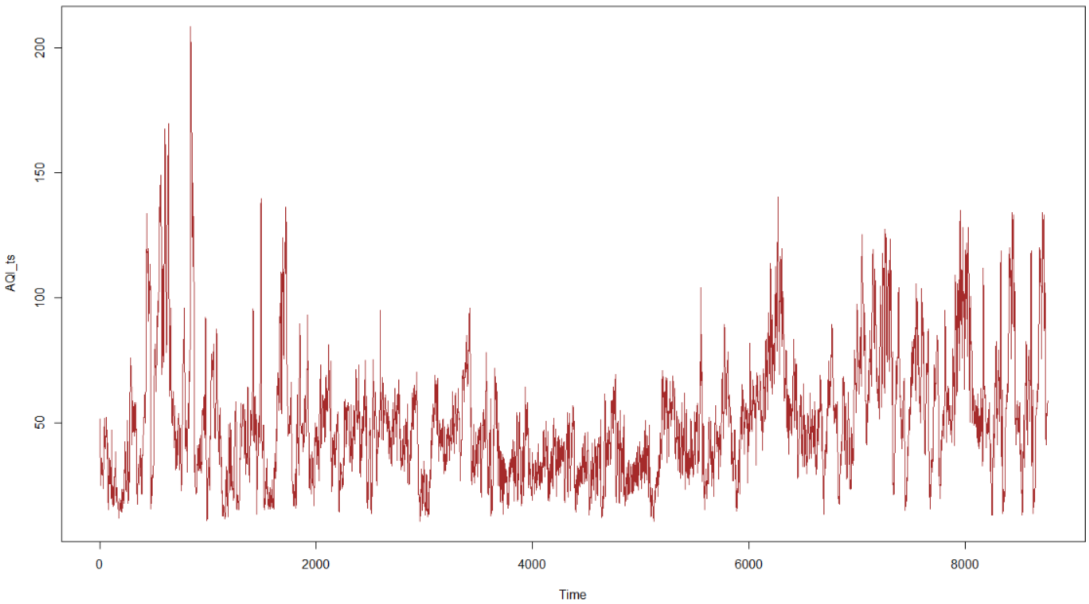


Figure 21 Full preview of AQI

There is a logical relationship between pollutants. For example, when the local environment is dominated by fine particulate matter sources, the ratio of PM2.5/PM10 is higher and the correlation is better. The ratio of PM2.5/PM10 is low and the correlation between PM2.5 and PM10 is poor when dust pollution is the main factor.Below, we assume that there is a linear relationship between pollutants to calculate the time change rule of the ratio between pollutants, and the LinearRelation function is used to show the time change characteristics of the slope between pollutants:

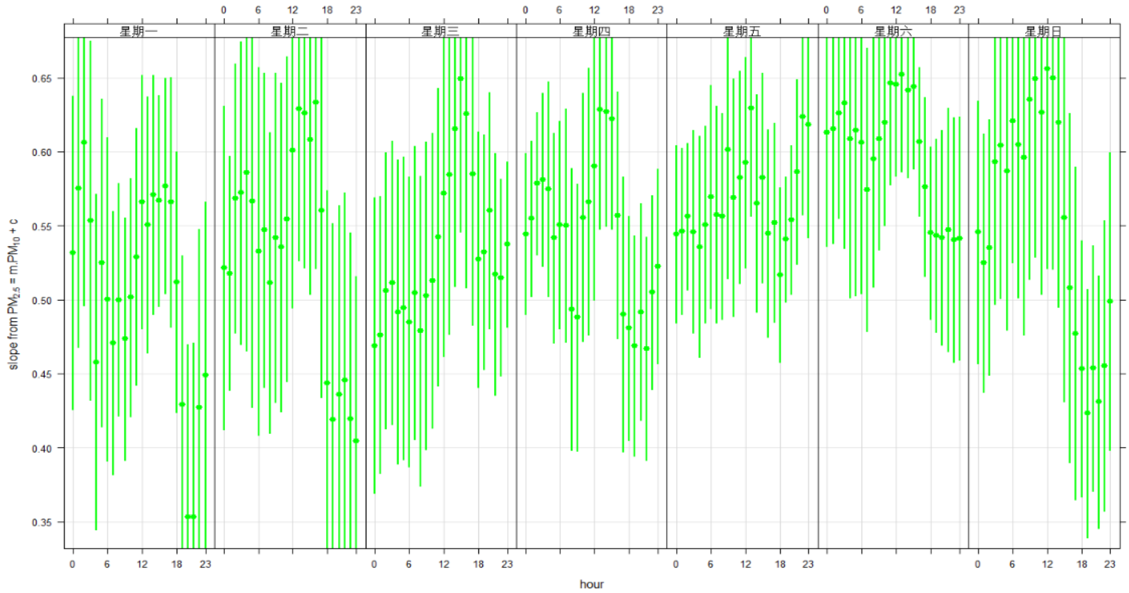


Figure 22 Linear Relation of PM2.5_PM10

It can be known from the above analysis that the time series data of atmospheric pollutants have certain periodicity (including obvious seasonal fluctuation, weekly and daily variation, etc.), so the autocorrelation and partial correlation of the data are briefly analyzed. I want to use the ARMA model to forecast and analyze the trend of time change. However, the data set currently crawled is only one year old, which cannot meet the decomposition requirements of analyzing seasonal factors. When the amount of data accumulated in the later period is enough, I will plan to launch a detailed time change prediction modeling.

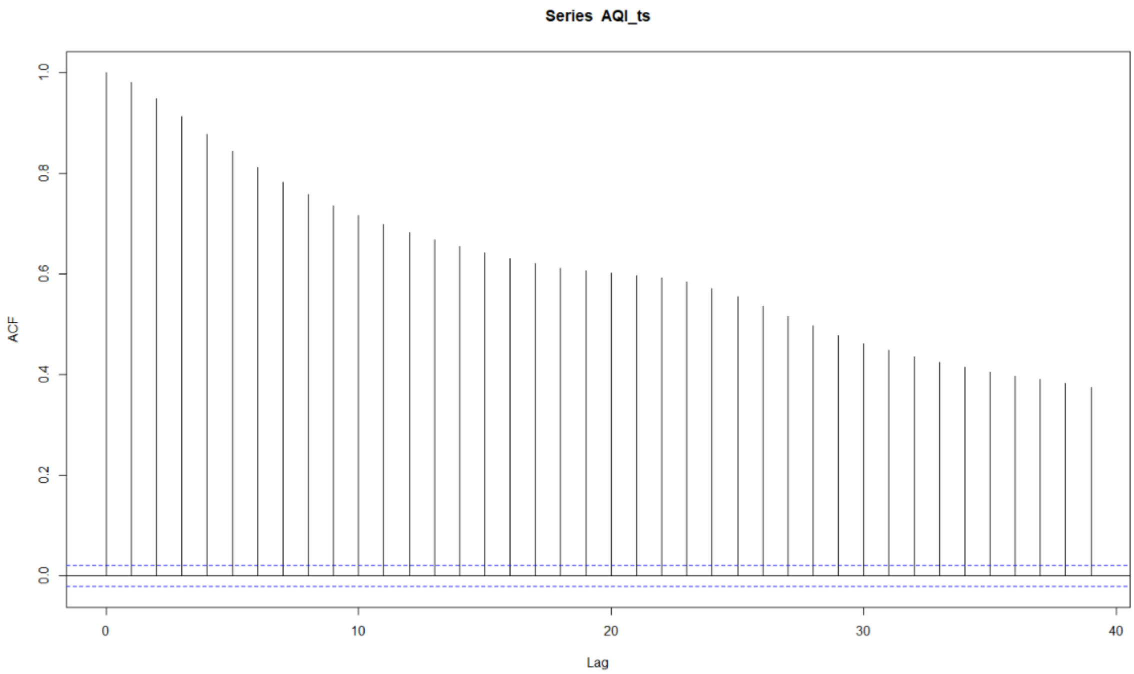


Figure 23 ACF _AQI

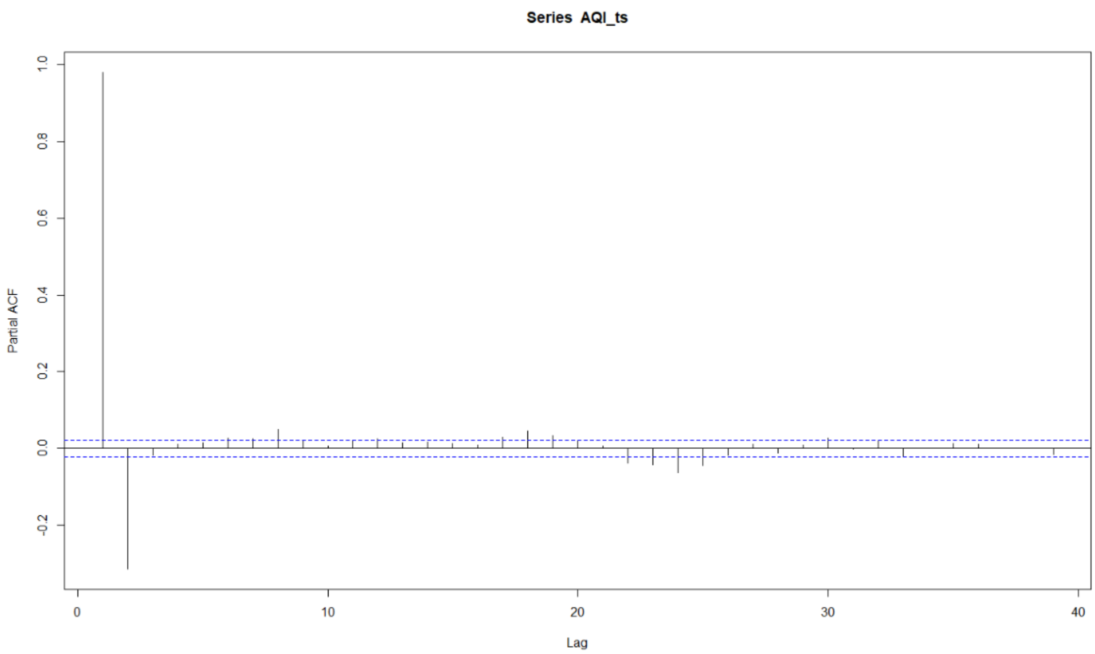


Figure 24 ACF _AQI

This is the distribution of AQI values for different stations:

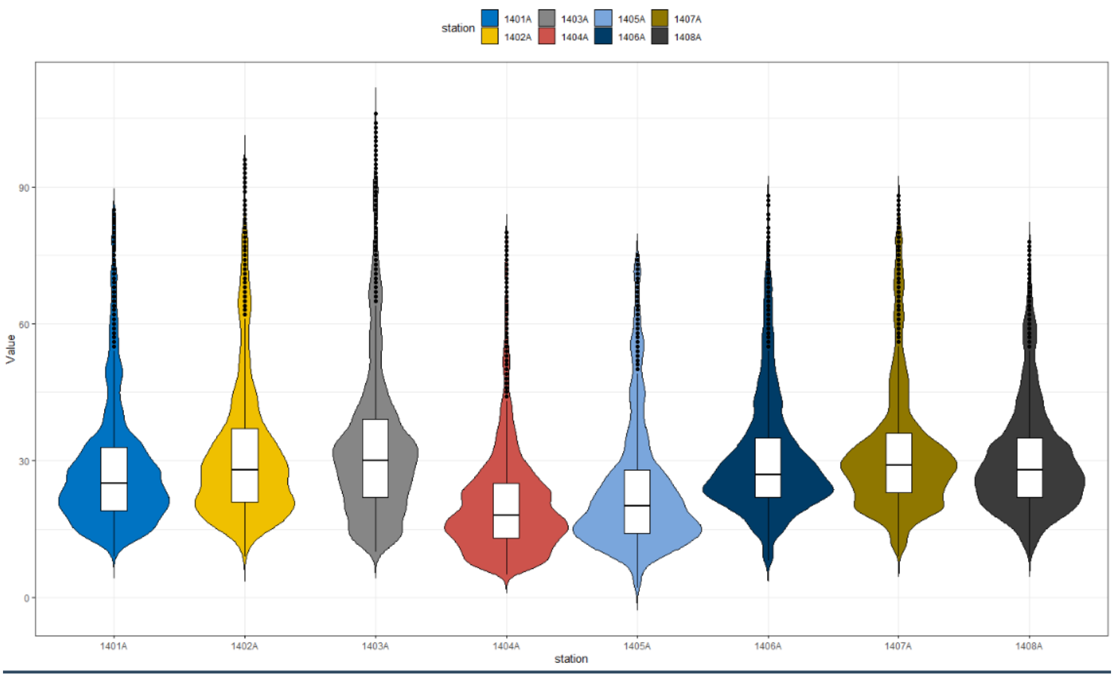


Figure 25 Spatial Interpolation

4 Conclusions

In terms of seasonal difference, Nanning's air quality is significantly affected by the climate type, showing the characteristics of good air quality in spring and summer and relatively poor air quality in winter and autumn. Except for O3, the four-seasons concentration trends of SO2, NO2, CO, PM2.5 and PM10 and the seasonal trend in AQI all shows the phenomenon of autumn> winter> spring> summer.

5 Discussion

The air pollution data collected in this paper are only from 8 national control points, which leads to insufficient representativeness of the model. In the future research, more data from municipal control points can be collected or monitoring points can be set manually to increase the number of samples. The air quality of the site may be related to the environmental factors of the site, such as building area, vegetation area, water area, population density and roads. I have also done some other parts in the R studio. I drew some scatterplots which reveals the spatial relationship between the variables and the spatial distribution of the variables about some monitoring stations. I haven't finished the part due to the incompleteness of data. Therefore, I mainly focused on the front part this time and I may accomplish the unfinished part when I get more data.

Reference

[1] 吴丹, 张世秋. 中国大气污染控制策略与改进方向评析[J]. 北京大学学报(自然科学版), 2011,47(06): 1143-1150.

[2] 黄晓虎, 韩秀秀, 李帅东, 等. 城市主要大气污染物时空分布特征及其相关性[J]. 环境科学研究, 2017,30(07): 1001-1011.

[3] 王韵杰, 张少君, 郝吉明. 中国大气污染治理:进展·挑战·路径[J]. 环境科学研究, 2019,32(10): 1755-1762.

[4] ZHANG Z, ZHENG N, ZHANG D, et al. Rayleigh based concept to track NOx emission sources in urban areas of China[J]. Science of The Total Environment, 2020,704: 135362.

[5] KIM K, KABIR E, KABIR S. A review on the human health impact of airborne particulate matter[J]. Environment International, 2015,74: 136-143.

[6] BRUNEKREEF B. Epidemiological evidence of effects of coarse airborne particles on health[J]. European Respiratory Journal, 2005,26(2): 309-318.

[7] BROOK R D, RAJAGOPALAN S, POPE C A, et al. Particulate Matter Air Pollution and Cardiovascular Disease[J]. Circulation, 2010,121(21): 2331-2378.

[8] 陈仁杰, 陈秉衡, 阚海东. 我国 113 个城市大气颗粒物污染的健康经济学评价[J]. 中国环境科学, 2010,30(03): 410-415.

[9] COHEN A J, BRAUER M, BURNETT R, et al. Estimates and 25-year trends of the global burden of disease attributable to ambient air pollution: an analysis of data from the Global Burden of Diseases Study 2015[J]. The Lancet, 2017,389(10082): 1907-1918.