|  |
| --- |
| Air Pollution in Chiang Mai, Thailand |
|  |
| **June 7**  Authored by: Worasom Kundhikanjana |

# Key Messages

In this report, we studied air pollution in Chiang Mai, Thailand. The major pollutants are PM2.5 and PM10. The pollutions are high during the winter season (October – April). Factors contributed to the pollutions level, such as weather patterns, the effect of global warming, and agricultural burning are considered.

The first section of this report provides a geographical background of the Chiang Mai. In section 2, we first understand the behaviors of different pollutants and their relationships. In section 3-4, the effect of weather patterns, fire hotspots, and human activities are studied. These sections will serve as a robustness check for our machine learning model.

The last section provides detail about the definition of AQI, data sources, and how the data are put together for modeling.

Contents

[Air Pollution in Chiang Mai, Thailand 1](file:///C:\Users\Benny\Documents\Fern\aqi_thailand2\reports\chiang_mai_report.docx#_Toc42434236)

[Key Messages 2](#_Toc42434237)

[1 About Chiang Mai 4](#_Toc42434238)

[1.1 Geography 4](#_Toc42434239)

[1.2 Pollution Monitoring Stations 5](#_Toc42434240)

[2 Air Pollutions Data 6](#_Toc42434241)

[2.1 Air Pollution Overtime 6](#_Toc42434242)

[2.2 Seasonal Pattern ~~Winter Season~~ 8](#_Toc42434243)

[2.3 Air Pollution Overtime 9](#_Toc42434244)

[2.4 Correlation Between Pollutants 10](#_Toc42434245)

[3 Effect of Weather Pattern 12](#_Toc42434246)

[3.1 Effect of Climate Change 12](#_Toc42434247)

[3.2 Effect of Wind Speed 13](#_Toc42434248)

[4 Agricultural Burning 13](#_Toc42434249)

[4.1 Strong Correlation between Particle Pollution Levels and Agricultural Burning. 13](#_Toc42434250)

[4.2 Fire Zone 15](#_Toc42434251)

[5 Predicting PM2.5 Level 16](#_Toc42434252)

[6 Technical Details 18](#_Toc42434253)

[6.1 Definitions of AQI 18](#_Toc42434254)

[6.2 Weather Data 19](#_Toc42434255)

[6.3 Hotspot Data 20](#_Toc42434256)

[6.4 Other Data 20](#_Toc42434257)

# About Chiang Mai

## Geography

Chiang Mai is one of the biggest cities in the northern part of Thailand. It also shares a border with Myanmar in the northwestern direction. Laos is on the northeast side . China is in the north with Yunan as the nearest big city. These neighboring countries still practice agricultural burning, which is suspected as a contributing factor to air pollution.

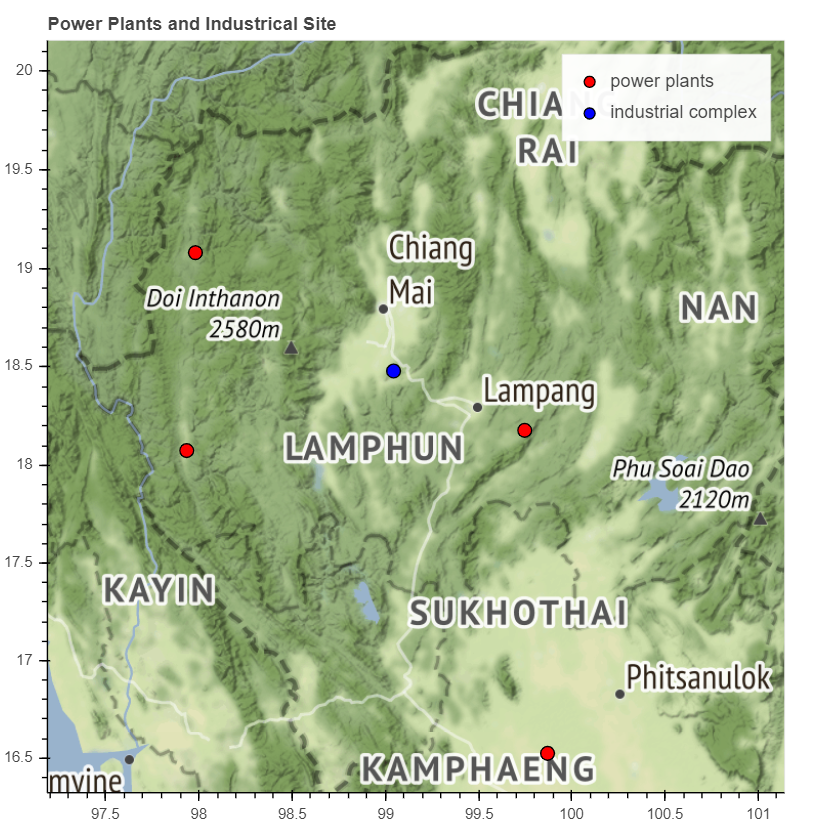


Figure 1: map of Chiang Mai with nearby industrial complex and power plants, which could potentially generate pollutions. The dashed line indicates the Thailand-Myanmar border.

Figure 1 shows a map of Chiang Mai and the surrounding province. Notice that Chiang Mai is a valley enclosed by high mountain ranges; thus, stagnant air and extreme weather patterns could have a multiplication effect on the existing air pollution problems.

There are four power plants in the neighboring provinces of Chiang Mai as shown in Figure 1, as red dots. There is a controversial coal power plant is in Mae Moh, Lampang, which is about 50 km southeast of Chiang Mai. This power plant used to cause air pollution problems especially sulfur dioxide pollutants. Currently, the Electricity Generating Authority of Thailand states that this problem has been resolved by installing an air filtration system within the plant in 1995. The other three power plants are oil and natural gas power plants. These power plants often affect two types of pollutants the SO2 and CO level. The former is unique to power plants, but the later could be generated by other burning activities. Since the power plants are located on the east and west sides of Chiang Mai, there would be a relationship between the pollutants and the wind direction. As we will see later in the report that this is not the case.

The closest industrial complex from downtown Chiang Mai is 25 km away in the southern direction, shown as a blue dot in Figure 1. Many industries are concentrated in this complex, including food, sugar, jewelry, and electronics. Again, if the industrial complex are the source of air pollution, the wind direction from the south would increase the air pollution, which is not the case here.

## Pollution Monitoring Stations

Figure 2 is a map plotted with the pollution monitoring stations. In Thailand, the Pollution Control Department (PCD) is in charge of monitoring environmental pollutions in air, waters and also in charge of environmental policy. PCD has two monitoring stations both in the downtown area shown as a pair of magenta-colored dots in Figure 2. The station in the north is near City Hall, and the station in the south is by Yupparaj Wittayalai School. These stations started operating since June of1995, measuring carbon monoxide gas (CO), ozone (O3), nitrogen dioxide (NO2), sulfur dioxide (SO2), and coarse particle(PM10). After a while, small particulate matterPM2.5) attracted public attention as a major contributor of air pollution. Accordingly, the Yupparaj Wittanyalai School station and the station at City Hall started monitoring small particle pollution in mid-2005 and mid-2016 respectively. This report will focus on analyzing the data from Yupparaj Wittayalai School station.

Since mid-2018, Chiang Mai University started setting up monitoring stations across Chiang Mai shown as yellow dots in Figure 2. These stations concentrate on measuring PM2.5 and PM10. So far, there are about 40 monitoring stations across and around the city. Although, the data from these stations are not sufficient for a machine learning more, they special information can potentially be used to form a spatial map, which can be correlated with the map from the satellite or studying effect of the local traffic. These will be the subject of further study.

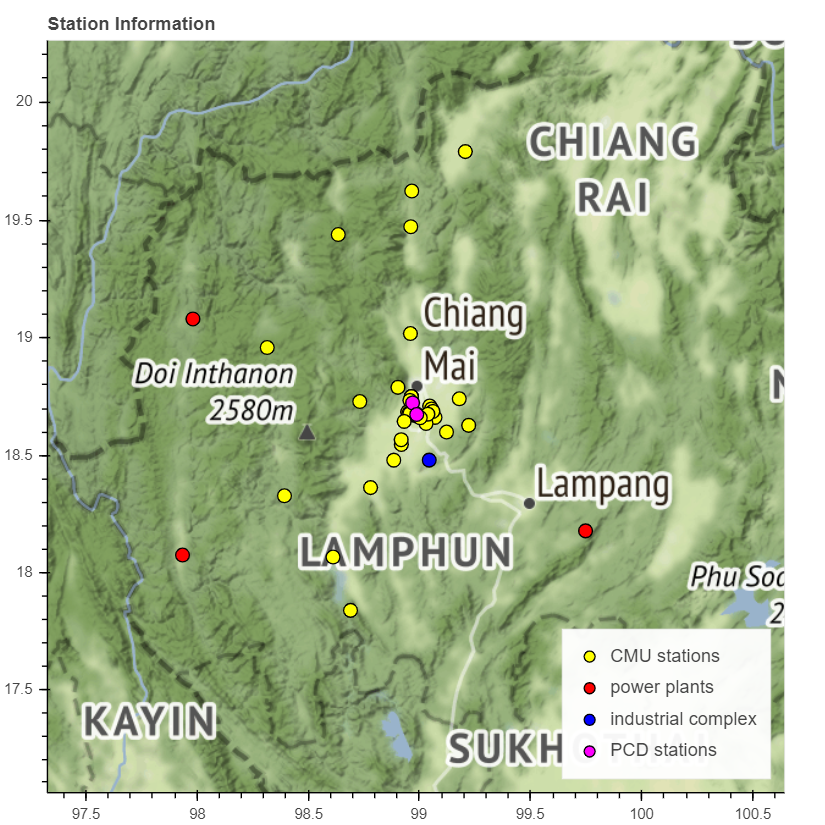


Figure 2: a map of pollution monitor stations in Chiang Mai, nearby industrial complex and power plants

# Air Pollutions Data

## Air Pollution Overtime

Figure 3 and Figure 4 show hourly pollution levels detected from Yupparaj Wittayalai School station since 1996. The pollutants with the most AQIs are in Figure 3, which are PM2.5, PM10, and O3. These values often exceed the unhealthy level (AQI=150) and sometimes reach a very unhealthy level (AQI>200) during the winter. On average the pollutants with highest AQI are PM2.5, PM10 and O3 respectively. For the other pollutants, CO, NO2, and SO2, the pollution levels are well within the satisfactory-moderate limit.

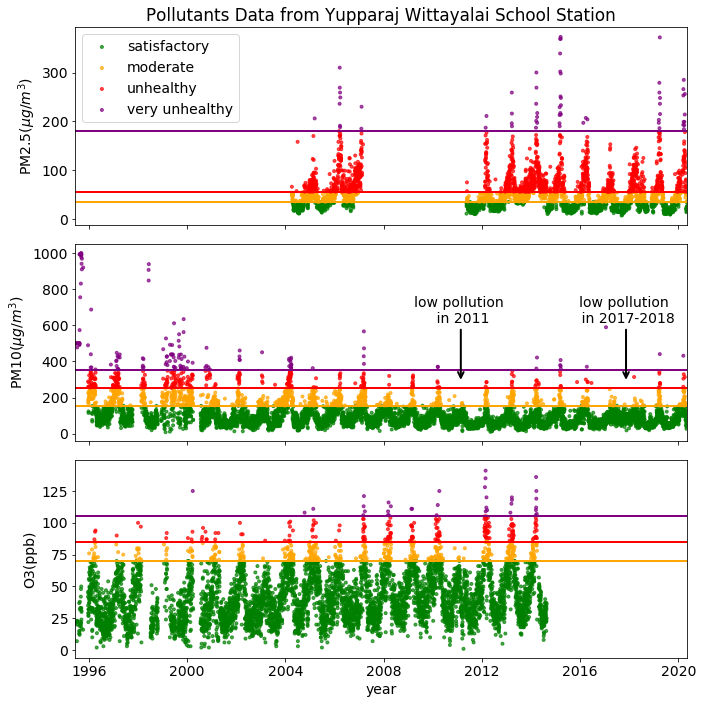


Figure 3: Daily maximum pollution level for PM2.5, PM10, and O3 from Yupparaj Wittayalai School station. The pollution levels have a seasonal pattern with high values during winter. In 2011, 2016 and 2017, the pollution levels are relatively low. The horizontal yellow, red and purple lines indicate the transition to moderate, unhealthy, and very unhealthy levels accordingly.

Note that the data for PM2.5 is missing between 2007 and 2011. Including the PM2.5 before 2007 could affect the accuracy of the model; thus, only the values after 2011 should be used for modeling. Alternatively, one could build a model for PM10 and infer PM2.5 from the PM10 value. Interestingly, in 2011, 2017, and 2018 the data points of PM2.5, PM10, and O3 are noticeably lower. These three years could serve as case studies to identify major pollution sources.

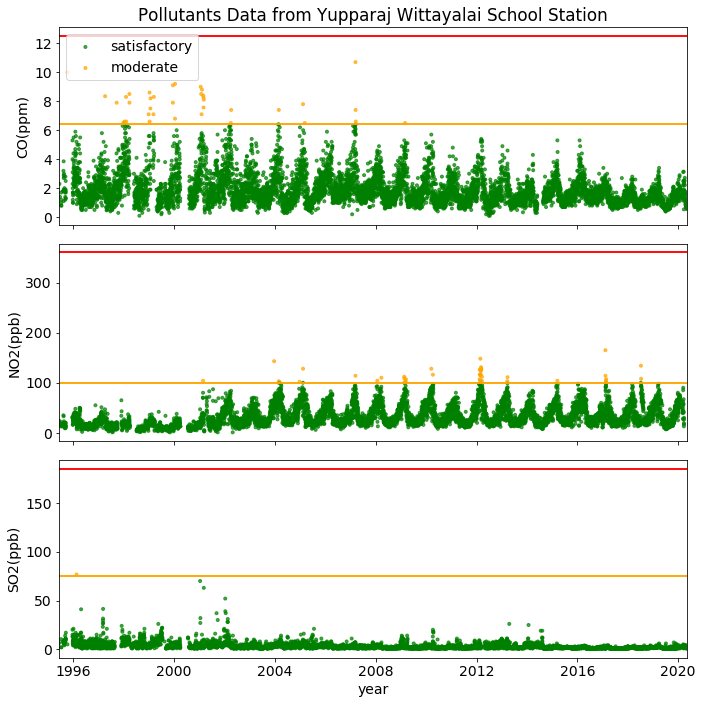


Figure 4: Daily maximum pollution levels for CO, NO2, and SO2 from Yupparaj Wittayalai School station. The horizontal yellow and red lines indicate the transition to moderate and unhealthy levels respectively.

## Seasonal Pattern ~~Winter Season~~

The official winter season starts on a different day every year and the historical records of those date are hard to trace. In this report, we define our winter months from the seasonal pattern of PM2.5 level. This is to be between the 1st of December and the 30th of April. This season will be used to calculate the yearly trends (called winter year) in the later part of this report, for example in Figure 6 the report as well. Figure 5 plots the daily average of PM2.5 levels over 14 years. The solid blue line shows the mean This average value exceeds the healthy limit, marked across the graph in red, between December and April. This seasonal pattern in Figure 5 will be compared with the wind and hotspots pattern in Figure 8, Figure 10, and Figure 12.

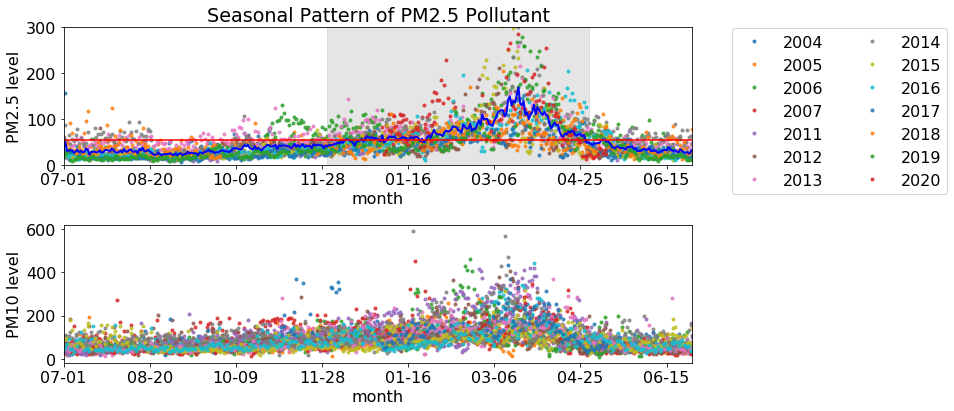


Figure 5: Daily average of PM2.5 level in a different year. The solid blue line is the average pollution level for the same day of year. The average crosses the healthy limit (horizontal red line) between December – April.

To understand the seasonal behavior of the air pollutions, we look at this average pollution levels on the same calendar date from different year in Figure 5. Different colors indicate the data from different year. The solid blue line is the average value. The PM2.5 level significantly increases in winter season. The average values exceed the health limit (horizontal red line) between the 1st of December and the 30th of April. We call this period *an air pollution season*.

The pollution season behavior is important for studying the yearly trend. Typically, the yearly average values are lower than the average of the values from just the air pollution season. Therefore, it is more informative to studying change in the average from the air pollution season. Moreover, the seasonal pattern Figure 5 is also useful when studying relationship with pollutions sources, which may exhibit a similar seasonal pattern. These relationship are discussed in Figure 8, Figure 10, and Figure 12.

## Air Pollution Overtime

Figure 6 is plotted with the average PM10 levels during the pollution season over the years. The average is well kept between 45 and 80, which is a good AQI level, since 2004 up to now. No major points stand out, interpreting no major improvements or worsening of air pollution. However, it is necessary to point out that low yearly average values could be misleading. They are different from daily or hourly average values, which are often much higher than the yearly values. When broken down into days or hours, PM10 levels can exceed the limits.

Note that the years with low pollution levels (2011,2017, 2018) in Figure 3 is shifted to 2010, 2016 and 2017 in Figure 6. This is because the low number is from the pollution season of 2010, which also covers the early months of 2011.

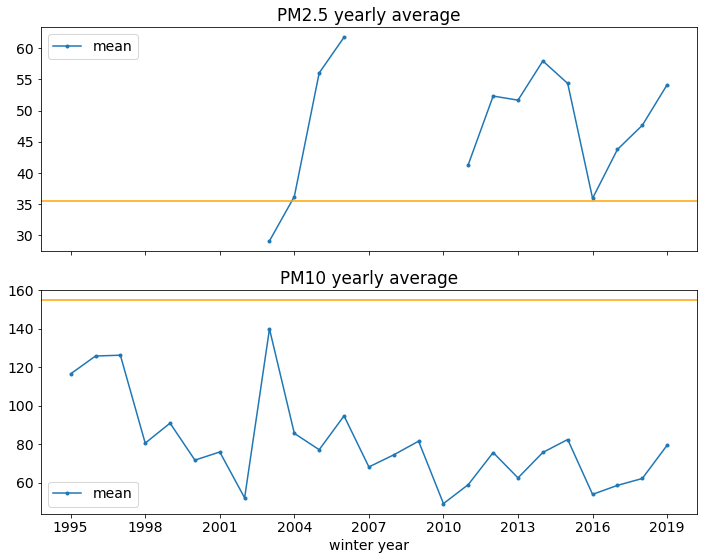


Figure 6: a yearly average of PM10 value for different pollution years. An average of pollution year only accounts for the data between the December 1st and April 30th of the next year. For example, the pollution year 2016 means the data between December, 2016 and April 2017. The yellow line indicates the transition from a good to a moderate limit. The low pollution seasons are the ends of 2010, 2016, and 2017.

## Correlation Between Pollutants

Except for SO2, the pollutants show a seasonal pattern with high values during the winter (December – April), and their levels are highly correlated. This correlation is shown by a plot of correlation coefficients among the pollutants in Figure 7. PM2.5, PM10, NO2, and O3 have correlation coefficients above 0.5 indicating high correlation. The same seasonal patterns and the high correlation coefficients indicate that similar chemical reactions generate these pollutants, which in turn suggests that air pollution may be generated from the same reoccurring sources.

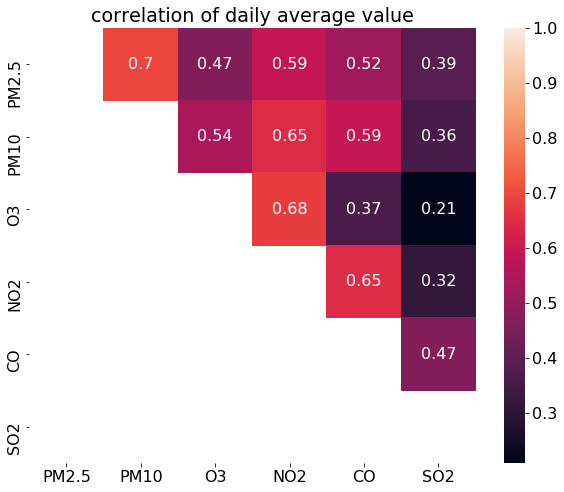


Figure 7: Spearman correlation coefficient among pollutants. Red color indicates high correlation coefficients, which are the cases for PM2.5, PM10, NO2, and O3.

The high correlation behavior among the pollutants also helps infer the missing data. As mentioned earlier, PM10 can be used to infer PM2.5. Besides, since the NO2 concentrations can be seen from the satellites, such a map can also be used to understand the spatial distribution of the particle pollutants.

However, a closer look at the peak value of each pollutant indicates slightly different rising and falling patterns. Figure 8 shows the average rising and falling behaviors of different pollutants over a calendar date. Each line is an average pollution level of the same calendar date from different years, similar to Figure 5. The y-axis values are the pollution level normalized by the maximum value for each pollutant. The normalization process allows a comparison between different pollutants. All pollutants except for SO2 have a clear and rather synchronous seasonal pattern with a peak in March. The PM2.5 level have a steep angle from early January, while the other pollutants show a gradual change from early October. All pollutant levels drop sharply in April.

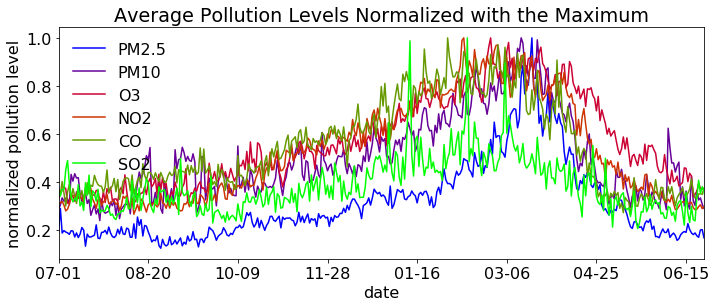


Figure 8: Day of year averages show rising and falling behaviors for different pollutants in the winter season. SO2­ does not have a significant seasonal pattern. The level is normalized by the maximum average maximum value.

# Effect of Weather Pattern

## Effect of Climate Change

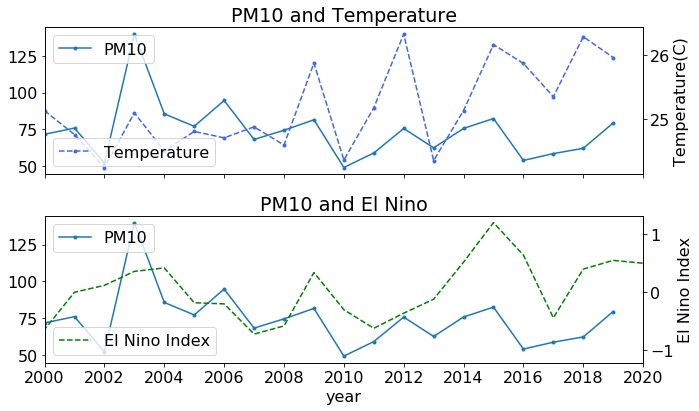


Figure 9: The effect of rising average temperature, and El Nino-La Nina effects on PM10 level in Chiang Mai since 2000. Both effects have low correlation with the average PM10 levels, correlation coefficient=0.07 and 0.26, respectively.

Climate change contributes to extreme weathers and increase a number of seasons wild fire. This is not a direct cause for air pollution in Chiang Mai. Figure 9 compares the yearly trend of temperature, El Nino Index, and PM10 levels. A rising trend can be seen in the average temperature (dash blue line). Since the average pollution levels do not show either increasing or decreasing trend, the temperature and El Nino index have a low correlation with the pollution level with correlation coefficient=0.07 and 0.26, respectively.

## Effect of Wind Speed

Air circulation is considered very important to reduce the pollution level in an area. Chiang Mai is surrounded by high mountains. During the winter(December – April), the average wind speed is about 5 km/hr, and in the monsoon season, the wind speed increase to about 15 km/hr causes this to speed up Figure 10 compares the average PM2.5 seasonal patterns with the patterns of the average wind speed. The wind speed decreases drastically in November and increases again in April. The month with most stagnant airflow is between December and January, however the pollution levels peaks in March. Therefore, the low wind speed is not a major contribution factor for high pollution level. Furthermore, since the wind speed has small effect on the pollutions level, we can also rule out the industrial complex and power plants from the sources of air pollution.



Figure 10: Comparison between the PM2.5 (solid line) and wind speed (dashed line) pattern in the winter season. The wind speed pattern is negatively correlated with the PM2.5 level (Spearman correlation coefficient = - 0.26)

# Agricultural Burning

## Strong Correlation between Particle Pollution Levels and Agricultural Burning.

Many countries in South East Asia still practice agricultural burning, and this is mostly done during the winter season.. Figure 11 shows satellite detected hotspots around Chiang Mai. These hotspots represent fire activities, very likely to be from agricultural burning or wildfires. There are significant burning activities around the Chiang Mai during the winter. There are dense burning activities in Myanmar and Laos, less activity in Thailand, and light activities in China. We will see that these activities are major contributing factors to high particle pollution in Chiang Mai.

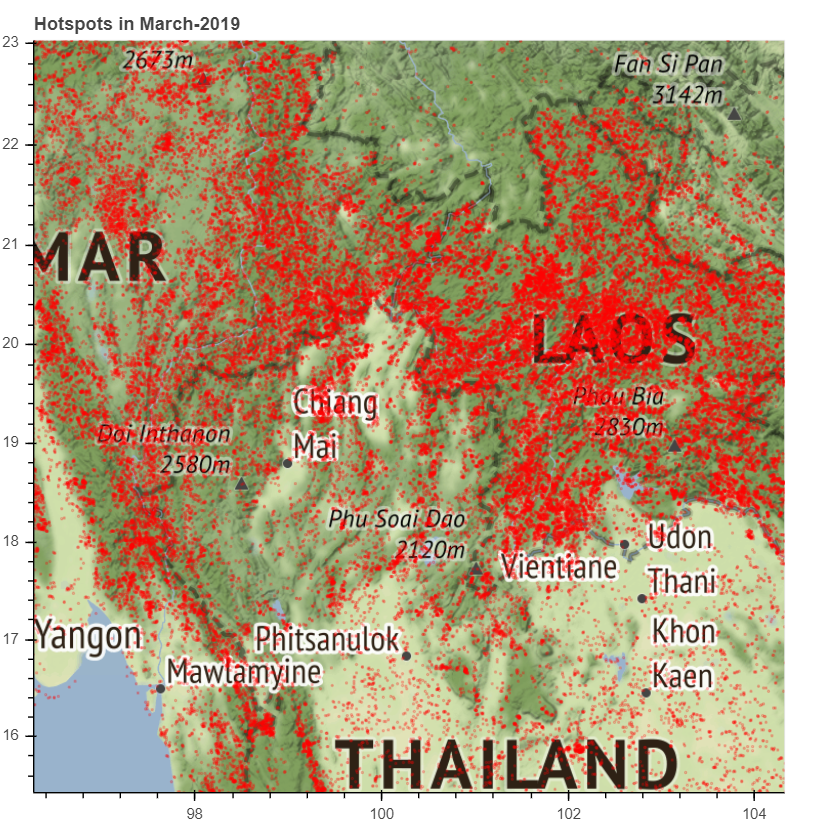


Figure 11: Satellite detected hotspots in March 2019

Figure 12 compares the seasonal pattern of fire hotspots and those of the particle pollutions. The pattern highly correlates with the PM2.5 level (Spearman correlation coefficient = 0.9). Following the highest burning activities in March, both blue and purple lines have reached its peaks. A similar pattern is shown with PM10with correlation coefficient = 0.7.

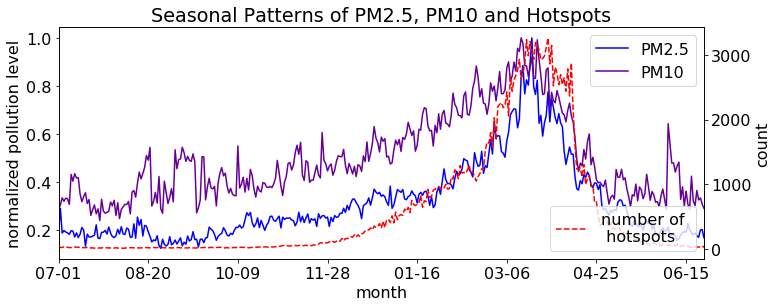


Figure 12: Compare the seasonal pattern of the number of hotspots within 1000 km from Chiang Mai(red) and the particle pollutions (blue and purple). The number of hotspots has a similar pattern as the PM2.5’s with Spearman correlation coefficient = 0.9, while PM10 and the number of hotspots has correlation coefficient = 0.7.

The total number of hotspots in each year also determines the average pollution levels. Figure 13 compares the average particle pollution levels for the pollution season and the number of hotspots. In the pollution year of 2010, 2016, and 2017 with relatively lower particle pollution levels, the number of hotspots is also lower.

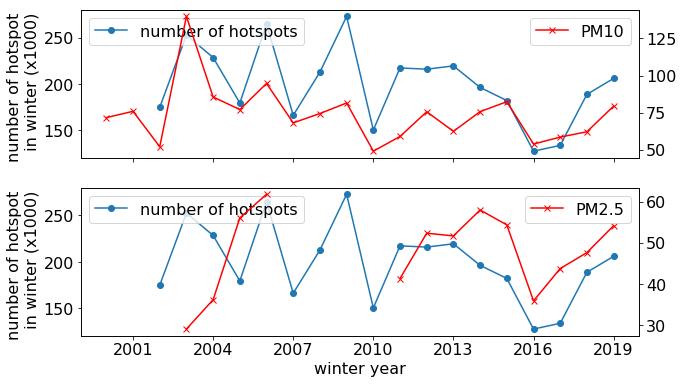


Figure 13: High correlation between the seasonal average PM10 and PM2.5 levels and the total number of fire hotspots in that season. In winter 2010, 2016, and 2017, the fire activities are significantly lower, thus lower pollution levels.

## Fire Zone

Since concentration of hotspots differ at various distance from Chiang Mai, it is important to divide the burning activity into distance-based zones. Figure 14 shows four fire zone. The first zone is with in 100 km from Chiang Mai within Thailand boarder. The second zone is between 100 and 400 km from Chiang Mai. A lot of fire activities in this zone are in Myanmar and Laos. The third zone is between 400 and 700 km from Chiang Mai. Most fire activities here are in Laos. The last zone is between 700 and 1000 km. Most fire activities in this zone are in Cambodia, which is much further away and the burning activities concentrate in December, which is not the month with the highest pollution in Chiang Mai.

Note that the fire activities as far as 400 – 1000 km are included in the model because there are evidence that the particulate pollution can travel as far as China and Taiwan, which is more than 1000 km away. In addition, the fire zones also distinguish between the fires from within and outside Thailand. Since solving the burning activities in these two zones imply different type of policy. The former is an internal affair, while the latter are international collaboration.

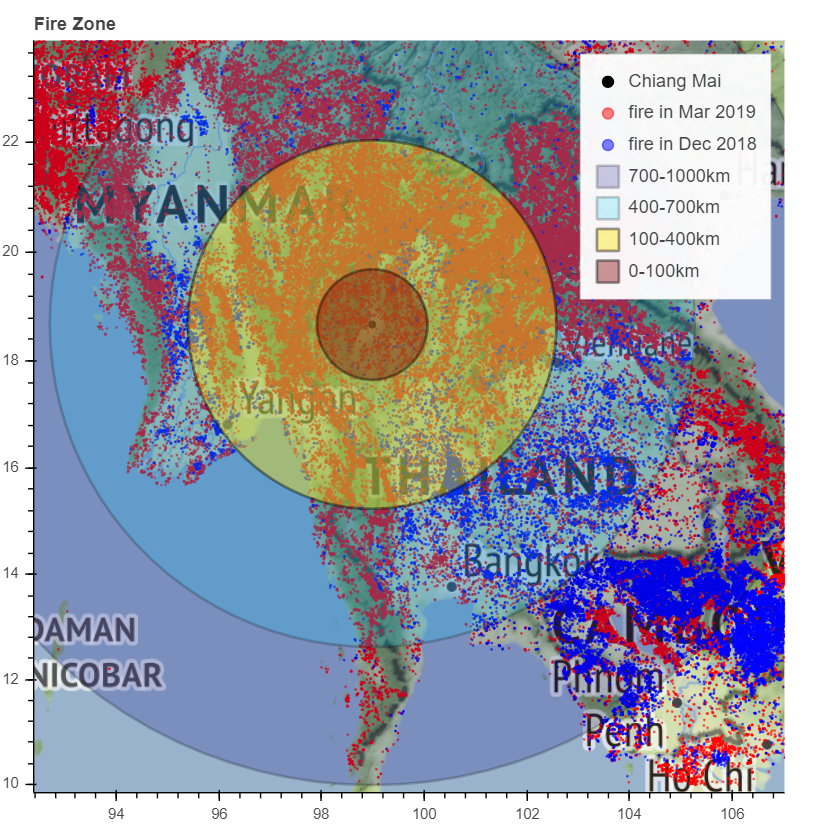


Figure : Four fire zone and hotspot concentration in December 2018(blue) and March 2019 (red).

# Predicting PM2.5 Level

In this section, we describe a machine learning model built to predict hourly PM2.5 level. The weather data (temperature, wind direction, wind speed, and humidity), and fire hotspot are used as the model inputs. The model served two purposes. First, identify the major factor contributed to the pollution. Second, model can be used to predict pollution level if certain policies are implemented.

Figure 15 show the model performance on the data between November 2018, and April 2019. The model was not allowed to see this data during training. The model has pretty good performance with R2 score = 0.94 (perfect prediction has R2 score = 1)

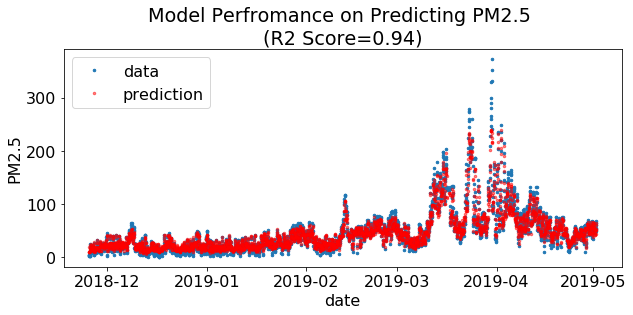


Figure 15: Model performance on the data between November 2018, and April 2019. The actual data (blue) and the prediction(red) are in agreement. The model has R2 score = 0.94(the best possible value =1)

Figure 16 shows the model input ranked by order of importance. The order of importance index is obtained by measuring the decreased in model accuracy if shuffling an input column. The order of importance value is in arbitrary unit. As expect, the fire features at different distance from Chiang Mai are most important features. The weather pattern such as temperature, and wind speed only have weak effect on the pollution level. Note that one feature ‘time\_of\_day’ refer to the hour of the day. It is included in the model to help improve the accuracy.

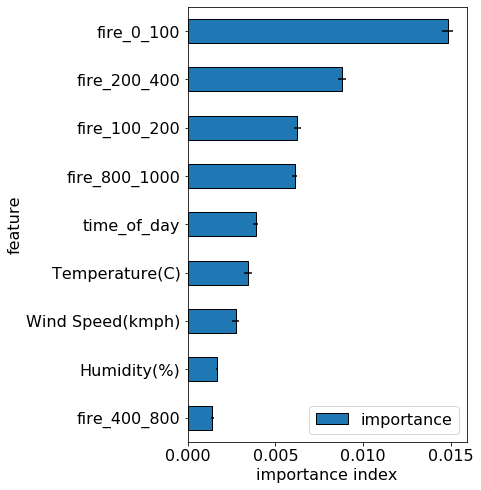


Figure 16: A rank of input’s order of importance. The fire feature within 400 km from Chiang Mai are most important.

# Technical Details

## Definitions of AQI

This report uses US AQI for pollution standards. Sometimes the local government pollution standard is lower compared to those in developed countries Therefore, it is crucial to compare the US and Thailand AQI standards. Figure 17compares the AQI standard for the US and Thailand for different pollutants. The pollution levels are on the horizontal axis, while the hazard levels are color-coded. For each pollutant, the corresponding AQI values do not scale linearly across different hazardous levels, but AQI below 50 means very good air quality, and above 200 AQI means unhealthy.

Thailand and the US have a slightly different cut-off value. However, the values considered to be hazard are in agreement for most pollutants, except for SO2. For SO2, 200 ppb is still considered healthy in Thailand, while this exceeds the healthy limit for the US standard.

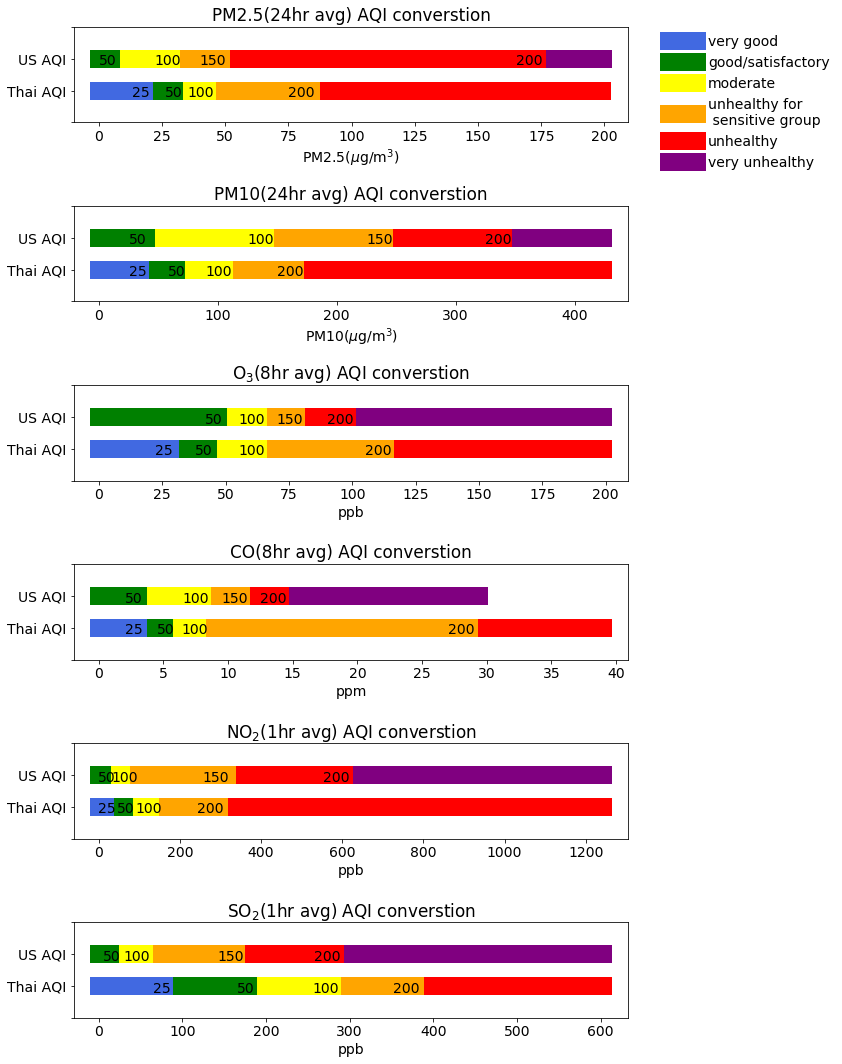


Figure 17: Comparison between US and Thailand AQI standards. The text on the plot indicates the maximum AQI for that range.

## Weather Data

Data preprocessing

## Hotspot Data

<https://earthdata.nasa.gov/earth-observation-data/near-real-time/firms/c6-mcd14dl>

## Other Data

US AQI standard <https://airnow.gov/index.cfm?action=airnow.calculator>

Thailand AQI standard <http://air4thai.pcd.go.th/webV2/aqi_info.php>

Holiday information in Thailand are from <https://www.timeanddate.com/holidays/thailand/>

El Niño index is from <https://origin.cpc.ncep.noaa.gov/products/analysis_monitoring/ensostuff/ONI_v5.php>