Bay Area Air Quality in 20 days after the Camp Fire

CEE 263D Course Project

Shuyi Yin*

syin3@stanford.edu

Abstract

Although the Bay Area did not witness any calsualties, it did suffer from severe air quality circumstances for two weeks during the Camp Fire period (Nov. 7th to Nov. 25th). It is therefore natural to revisit the air quality in the Bay Area during that period and to see what air pollutants specifically caused the smog that shut down businesses and public services. In this paper, we crawled and examined time series of air pollutant concentrations monitored by Bay Area Air Quality Management District (BAAQMD), calculated ozone concentrations by assuming it only came from NO and NO_2, studied and discussed an interesting mini-case spotted in the data for plateaued overnight ozone concentrationand quantified the relationship between logarithms of reaction rate of NO_2 photolysis and $PM_{2.5}$ concentrations.

1   Introduction

Three-hour drive north of San Francisco downtown, Pulga, California witnessed the first sparks of Camp Fire, the deadliest and most destructive wildfire in California history to date. Although it made a narrow escape from the fire, the Bay Area could not save itself from two weeks of heavy smog. To understand what air pollutants were present in the smog and therefore adverse to public health, and more importantly, how they interacted with each other, we studied several pollutants recorded in the BAAQMD data set. Specifically, they are:

- Ozone: O_3 in ppb;

- Nitric oxides (NO_x): NO and NO_2, in ppb;

- $PM_{2.5}$: in μg/m^3;

- Carbon Monoxide: CO, pptm (parts per ten million).

---

*This project can be found at my GitHub repo: *Bay-Area-air-pollution-after-Camp-Fire*.

We examine the time series of five pollutants at San Francisco downtown and at Redwood City, to compare the smogs in north and south Bay. The reason why our analysis in this paper is limited to these two locations is that stations across BAAQMD are not consistent in air pollutant monitoring, e.g. at some stations Black Carbon (BC) is also monitored and logged, while Ozone (O_3) is vacant. After analysis, we plot the time series and spot interesting events in the data and provide discussions.

The rest of the paper is structured as follows. In **section 2**, we briefly introduce the data set we used in the project, with examples. In **section 3**, we describe the chemical reactions informing Ozone (O_3) near surface and derive mathematical expressions in calculating reaction coefficients. In **section 4**, we present our calculations and discuss the plots. Also, we will talk about an interesting mini-case spotted in the data. Lastly, we conclude by a summary of discoveries (see **section 5**), limitations in our analysis and future directions in **section 6**.

2 Dataset

Data are crawled from BAAQMD website[2]. Specifically, historical air monitoring data is presented at each station on each date available. An example is air monitoring data for San Francisco - Arkansas St station, on November 8th, 2018. On the webpage, we can select (1) date, (2) measurement for pollutants, and (3) station. In our paper, we select all pollutants, stations of **San Francisco - Arkansas St., ID 5011** and **Redwood City, ID 6004**, and days from **November 1st, 2018** to **November 30th, 2018**. A Python script was written to automate the process and time series were saved to Pandas dataframe for later process [1]. Each of the two Pandas dataframe objects has 6 columns for 5 pollutants and NO_x and 720 rows for 30 days with 24 records per day.

3 Methods

In this section, we present the three most important building blocks of the analysis in this paper. They range from the web crawler I built myself, chemical reaction and coefficients for O_3 production with NO and NO_x, and the mathematical derivations for O_3 concentration calculation:

- Web crawler, written in Python, which is used to download data from BAAQMD;
- NO_x related cheimcal reactions involved in O_3 formation, with reaction coefficients;
- Mathematical derivation of O_3 concentration by two methods: PSSA and Photostationary State Relation equation.

3.1 Web crawling

The data display page of BAAQMD contains multiple tables besides the actual data table, so the crawler is more complicated than typical BeautifulSoup[2] scripts. As can be seen in the function defined below, the data table in the original page source is defined by **\<tr>\</tr>** and class of **dataSep**

---

[2]Link is also accessible: http://www.baaqmd.gov/.

in HTML. We only retrieve the six measures from the tables and treat missing values in the orginal table as **NaN**. Units are manually copied.

Listing 1: Crawling BAAQMD webpage in Python

```python
import re
import requests
import pandas as pd
from bs4 import BeautifulSoup
def download(url):
    res = requests.get(url)
    soup = BeautifulSoup(res.content,'lxml')
    tables = soup.find_all(lambda tag: tag.name == 'tr' and tag.get('class') == ['dataSep'])

    # create dataframe
    dic = {}
    pollutant = ["PM2.5", "NO2", "NOx", "NO", "O3", "CO"]
    for i in pollutant:
        dic[i] = []
    for table in tables:
        cont = table.contents
        count = 0
        for item in cont:
            if item != '\n':
                if count == 0:
                    for j in pollutant:
                        if j in item.contents[0]:
                            break
                elif len(item.contents) == 1:
                    dic[j].append(pd.np.nan)
                else:
                    thing = item.contents[0]
                    dic[j].append(thing.contents[0])
                count += 1
    df = pd.DataFrame.from_dict(dic).astype(float)
    return df
```

3.2 Chemical reactions

Book [3] presents the photolysis of nitrogen dioxide in N_2 and O_2 in the following table 1. From these reactions, we can calculate O_3 concentrations by assuming the dominant source is NO_x related chemicals. That is to say, we are ignoring O_3 formation from hydrocarbons (e.g. CH_4, C_2H_6), alkenes (e.g. C_2H_4), aromatics (e.g. C_6H_6) and aldehydes (e.g. CH_3CHO, C_2H_5CHO), due to lack of data. Meanwhile, although we are collecting CO concentrations, we are not calculating the dynamics of CO-related O_3 emissions either, due to lack of dynamics coefficients. The other critical assumption we are making is the reaction rate constants at $298K = 76F$ also hold in our cases, although temperatures in most of our days are actually lower at levels of $60K \sim 65K$.

Reactions are initiated by the photolysis of nitrogen dioxide into nitrogen oxide and oxygen atom (*Reaction 1*). The rate of this decomposition depends on light intensity: stronger the sunlight,

3

faster the decomposition. The oxygen atom then combines with oxygen molecule to form an ozone molecule, in the presence of a third molecule (denoted by M here, $[M] = 10^6$ppm for any third molecule). The ozone molecule and the oxygen atom then react with the oxides of nitrogen in *Reactions 3-11*. Therefore, it is clearly seen that O_3 molecule is involved only in *Reactions 2,3,11*.

Table 1: Reactions and rate constants for the photolysis of nitrogen dioxide in N_2 and O_2.

No.	Reaction	Rate constant at 298K	Reference
1	$NO_2 + h\nu \longrightarrow NO + O$	Depends on light intensity	
2	$O + O_2 + M \longrightarrow O_3 + M$	2.33×10^{-5}ppm^{-2}min^{-1}	Johnston (1968)
3	$O_3 + NO \longrightarrow NO_2 + O_2$	2.95×10^{1}ppm^{-1}min^{-1}	Johnston abd Crosby (1954)
4	$O + NO_2 \longrightarrow NO + O_2$	1.38×10^{4}ppm^{-1}min^{-1}	Schuck et al. (1966)
5	$O + NO_2 + M \longrightarrow NO_3 + M$	4.50×10^{-3}ppm^{-2}min^{-1}	Schuck et al. (1966)
6	$NO_3 + NO \longrightarrow 2\,NO_2$	1.48×10^{4}ppm^{-1}min^{-1}	Schott and Davidson (1958)
7	$O + NO + M \longrightarrow NO_2 + M$	2.34×10^{-3}ppm^{-2}min^{-1}	Kaufman (1958)
8	$2\,NO + O_2 \longrightarrow 2\,NO_2$	7.62×10^{-10}ppm^{-1}min^{-1}	Glasson and Tuesday (1963)
9	$NO_3 + NO_2 \longrightarrow N_2O_5$	4.43×10^{3}ppm^{-1}min^{-1}	Schott and Davidson (1958)
10	$N_2O_5 \longrightarrow NO_3 + NO_2$	1.38×10^{1}min^{-1}	Mills and Johnston (1951)
11	$NO_2 + O_3 \longrightarrow NO_3 + O_2$	0.46×10^{-1}ppm^{-1}min^{-1}	Ghormley et al. (1973)

3.3 Mathematical derivations

3.3.1 PSSA

From the section above in 3.2, we know the consumption and production of O_3 are only in *Reaction 2,3,11*. We can therefore write the rate of change in O_3 concentration as the following differential equation:

$$\frac{d[O_3]}{dt} = R_2 - R_3 - R_{11},$$

where R_2, R_3, R_{11} are the rates of the three reactions, respectively.

Further, if assume $[O_3]$ stays very close to steady state, we have

$$\frac{d[O_3]}{dt} = R_2 - R_3 - R_{11}$$
$$= k_2[O][O_2][M] - k_3[O_3][NO] - k_{11}[NO_2][O_3] = 0$$
$$\therefore [O_3] = \frac{k_2[O][O_2][M]}{k_3[NO] + k_{11}[NO_2]}$$

To figure out $[O_3]$, we need to find $[O]$ first. Therefore, similarly, we know O is only involved in *Reaction 1,2,4,5,7*, and we can write:

4

$$\frac{\mathrm{d[O]}}{\mathrm{dt}} = k_1[\mathrm{NO_2}] - k_2[\mathrm{O}][\mathrm{O_2}][\mathrm{M}] - k_4[\mathrm{O}][\mathrm{NO_2}] - k_5[\mathrm{O}][\mathrm{NO_2}][\mathrm{M}] - k_7[\mathrm{O}][\mathrm{NO}][\mathrm{M}]$$

In PSSA, we assume $\frac{\mathrm{d[O]}}{\mathrm{dt}} = 0$. Hence

$$[\mathrm{O}] = \frac{k_1[\mathrm{NO_2}]}{k_2[\mathrm{O_2}][\mathrm{M}] + k_4[\mathrm{NO_2}] + k_5[\mathrm{NO_2}][\mathrm{M}] + k_7[\mathrm{NO}][\mathrm{M}]}$$

Therefore, we can perform calculations in both directions as follows. The latter tells us how strong sunlight is and hence how strong the photolysis of $\mathrm{NO_2}$ is:

- Given k_1, calculate $[\mathrm{O_3}]$;

- Given $[\mathrm{O_3}]$, find out k_1 and thus strength of sunlight.

3.3.2 Photostationary State Relation

Using only **Photostationary State Relation equation**, we can also write the expression for $\mathrm{O_3}$ concentration, which later will be compared with PSSA results.

$$[\mathrm{O_3}] = \frac{k_1[\mathrm{NO_2}]}{k_3[\mathrm{NO}]}$$

4   Results & Discussion

In this section, we present our analysis of air quality in San Francisco downtown and Redwood City and make comparisons.

4.1   General

Fig. 1(a) shows $\mathrm{PM_{2.5}}$ concentrations during polllution was an order of magnitude higher than normal levels, the highest concentration $241.0\mu g/m^3$ for San Francisco and $181.0\mu g/m^3$ for Redwood City both occurring on Nov. 16th. The smog arrived in the Bay Area on late night Nov. 8th and left the region on Nov. 21st. In general, $\mathrm{PM_{2.5}}$ concentrations were higher in downtown San Francisco, but the two locations roughly see the same trend over time. In Redwood City, the concentraion of particulate matters vary more than that in San Francisco does, with more spikes in the blue curve. In Fig. 1 (b), we present the concentrations of $\mathrm{O_3}$, NO and $\mathrm{NO_2}$ during the week of Nov. 19th to Nov. 26th. Notice the unusual behavior of three pollutants on the 23rd, where $\mathrm{O_3}$ level did not drop overnight but remained high. We will discuss this case later specifically in **section 4.2**.

4.2   Zoom-in on the 19th to 26th period

Following the discussion in **section 4.1**, we zoom in on the time axis and examine the k_1 values for the week Nov. 19th to Nov. 26th (see Fig. 2). The rate of reaction decreased first during the day of the Nov. 21st, then returned to normal at noon on the Nov. 22nd. Suddenly, it did peaked again on the
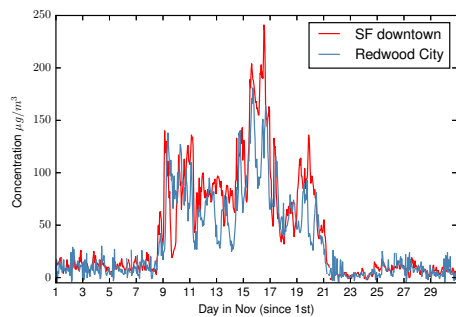
(a) PM$_{2.5}$ concentration at San Francisco and Redwood City.
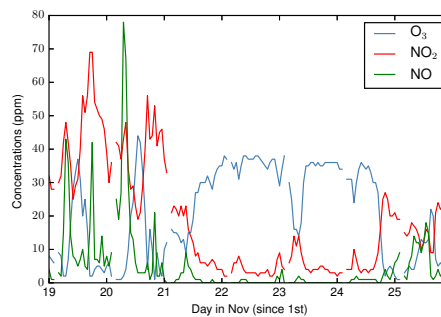
(b) Zoom-in view on pollutant concentrations at downtown San Francisco in 18th to 25th period.

Figure 1: Pollutants arrived at San Francisco and Redwood city roughly at the same time, in the evening of the 8th and left the region on the 21st. The concentration of PM$_{2.5}$ is way higher than normal during this two-week period. Pollutants in San Francisco behaved strangely for 21st to 23rd.

same day at midnight. The normal noon peak on the Nov. 23rd was rather low and things returned normal on the following days. Looking at data in Redwood City, we could not find a similar behavior: daily peaks occur at noon and was rather consistent in peak values, except the peak on Nov. 24th.

Clearly, it is the plateaued concentration of O$_3$ at San Francisco that peaked k$_1$ at midnight. It is then natural to ask what plateaud [O$_3$]. Checking historical weather data in San Francisco[3], we noticed that it rained from Nov. 21st to the Nov. 23rd. While cleaning up the weakening smog (the fire is being controlled at the same time), the rain and cloud created a favorable environment for O$_3$ formation. That's why we are seeing such a rapid increase in O$_3$ concentration on the 21st, which also benefits from stronger light intensity after the smog. A reasonable hypothesis is the rain cleaned up the smog and reduced human mobility in the city for the days. It also clouded the sky and created some layer favorable of O$_3$ formation. At the same time, some sources other than the NO$_x$ pathway contributed to O$_3$ and kept the concentration high. It is of low probability that human activity introduced extra O$_3$, because any burning of the O$_3$-related hydrocarbons would have introduced NO and NO$_2$ in the process, unless people were not burning gas but pure hydrocarbons, and concentrations for NO and NO$_2$ remained low for the period.
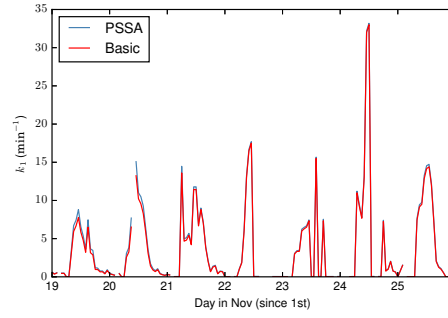
4.3 PM$_{2.5}$'s influence on k$_1$

In Fig. 3, the two colors compare the rates of coefficient computed by PSSA and Basic methods. The steelblue colored curve is the result of PSSA and the red curve denotes the curve for the Photostationary State Relation output. Under our assumption that O$_3$ is only produced by NO$_x$ related reactions, the methods do not produce significantly different k$_1$ values, hence *Reaction 11* is not important to include in our calculation.

---

[3]https://www.timeanddate.com/weather/usa/san-francisco/historic?month=11&year=2018

6
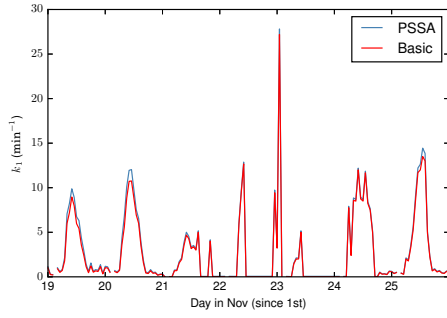
(a) Zoom-in view of k_1 values at San Francisco.   (b) Zoom-in view of k_1 values at Redwood City.

Figure 2: K_1 value was extremely high at midnight Nov. 23rd in San Francisco and disappeared during daytime Nov. 23rd, while a great peak occurred at Redwood City at noon Nov. 24rd.

Another trend exhibited in Fig. 3 is also a good topic to discuss. Notice during the two-week smog from the Nov. 8th to the Nov. 21st, the k_1 values were significantly lower than normal, before and after the smog. This demonstrates smog full of $PM_{2.5}$ reduced visibility for human and also scattered greater portion than usual of the sunlight reaching the near-surface. Therefore, the photolysis of nitrogen dioxide was slowed down and hence the production of ozone near surface. So can we establish a solid mathematical relationship describing this?

Several steps were taken before the fitting of linear model:

- Feature selection: $PM_{2.5}$ and k_1 curves in general are noisy throughout the 720 hours of the month, so we cannot use the whole dataset: some summarizing is needed. We choose the daily maximum of $PM_{2.5}$ concentration and daily maximum k_1 values for the 30 days to represent the data. An evident relationship between daily maximum concentrations and rate of reactions is enough in demonstrating the true existance of such a relationship.

- Scale of data: Plotting the data we can see the $[PM_{2.5}]$ values are concentrated near the origin on x-axis. Taking a logarithm scale on the $[PM_{2.5}]$ values is therefore necessary. Again plotting the $\mathbf{k_1} \sim \log [\mathbf{PM_{2.5}}]$, there is a clear non-linear relationship and the k_1 values seem to be condensed together. So a second logarithm was taken on k_1.

The final model is $\mathbf{\log k_1} \sim \mathbf{\log [PM_{2.5}]}$. Fig.4 presents the result of linear fitting. With R^2 values of 0.47 and 0.45, we say there are strong negative linear relationship between $\mathbf{\log k_1}$ and $\mathbf{\log [PM_{2.5}]}$. Specifically, we can write out the log-log linear relationship:
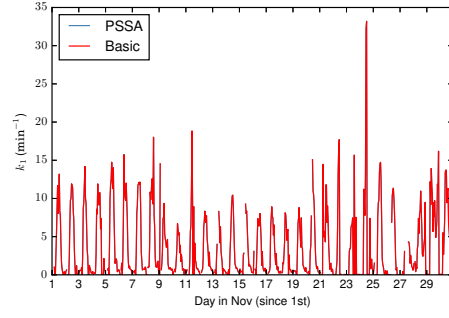
$$\text{San Francisco: } \log k_1 = -0.24 \times \log [PM_{2.5}] + 1.39$$

$$\text{Redwood City: } \log k_1 = -0.23 \times \log [PM_{2.5}] + 1.42$$

The fact that coefficients in two linear relationships are close tells that other factors affecting photolysis of NO_2 are roughly equivalent in San Francisco and Redwood City. If we can extend our result to other cities, maybe we can conclude some useful dynamics out of it.
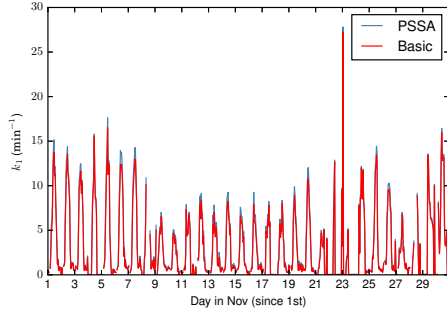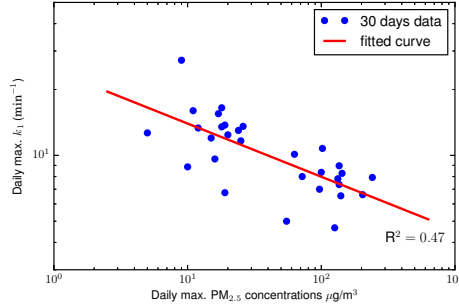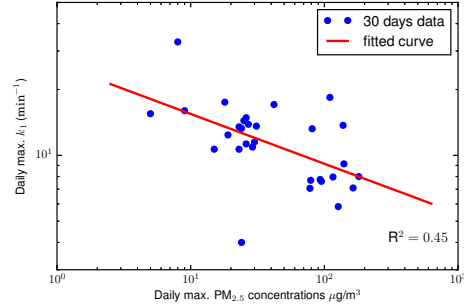
(a) K$_1$ values at San Francisco.

(b) K$_1$ values at Redwood City.

Figure 3: Smog blocked sunlight, hindered photolysis of NO$_2$ and thus reduced the value of k$_1$ for the period.

(a) Linear regression on San Francisco daily maximum PM$_{2.5}$ and k$_1$ values.

(b) Linear regression on Redwood City daily maximum PM$_{2.5}$ and k$_1$ values.

Figure 4: Regression analysis showed great fit between daily maximum $\log$ PM$_{2.5}$ and daily maximum $\log$ k$_1$ at both locations, further proving the effect of PM$_{2.5}$ induced smog on light intensity reaching the ground.

5 Conclusion

In this paper, we studied Bay Area air quality in the 3 weeks after the Camp Fire in November 2018. We discussed our data set from BAAQMD and methods in crawling, analyzing and modeling the data. Moreover, we presented our results and compared air quality in San Francisco and Redwood City. Specifically, our findings are:

- PM$_{2.5}$ concentrations at both locations increased by 10-fold during the pollution, highest records reaching 250μg/m^3 and 200μg/m^3 in San Francisco and Redwood City, respectively. Air pollution in San Francisco is little worse than Redwood City.

- Calculations by two methods PSSA and Photostationary State Relation equation produce very close results on rate of reaction for *Reaction NO$_2$ + hν $\longrightarrow$ NO + O* (this rate is denoted by k$_1$ throughout the paper). San Francisco and Redwood City witnessed very similar trends in the study period for k$_1$ values.

- A closer look at k_1 values of the week 19th to 26th revealed an interesting case where O_3 concentration remained high and NO and NO_2 stayed low overnight in San Francisco. A hypothesis was suggested that some other sources beyond the scope of this paper might have played an important role in the accumulation of O_3 concentration. Also, the favorable condition introduced by the rainy weather of the week could have contributed.

- A negative linear relationship was found between logarithm of daily maximum k_1 values and logarithm of daily maximum $PM_{2.5}$ concentrations at both locations. Confidence for both fittings are high and the coefficients were close. We used daily maximum to avoid noise in per hour data.

6   Future Work

In this paper, we made several assumptions before we carried out the analysis. First of all, we assume O_3 is only produced by NO and NO_2 in the cities of San Francisco and Redwood City. In reality, O_3 comes from a variety of sources, anthropogenic and natural. Including the neglected reactions will establish a more solid basis for the analysis. Secondly, we assumed rate of coefficients in the chemical reaction table is transferable to our temperature. which is typically 10F lower than 298K. In our case, the real rates should be smaller than the assumed values. So there is an intrinsic inaccuray in our calculation. Thirdly, we are limited in the quality and quantity of data. For some days in our study period, $PM_{2.5}$ concentrations were measured negative. Although very rare, this reduces our confidence in the data source. We did not examine data sources other than the BAAQMD cited in this paper. If we can cross-reference, then we will have a better idea in data quality. Also, this paper is limited in the measurements available. We did not use the CO concentrations, which are available at both locations. We could not either extrapolate concentrations only available at nearby sites, such as Black Carbon at San Jose. If we could, then our analysis would achieve more.

References

[1] Wes McKinney. Data structures for statistical computing in python. In Stéfan van der Walt and Jarrod Millman, editors, *Proceedings of the 9th Python in Science Conference*, pages 51 – 56, 2010.

[2] Vineeth G. Nair. *Getting Started with Beautiful Soup*. Packt Publishing, 2014.

[3] John H Seinfeld. *Air pollution: physical and chemical fundamentals*. McGraw-Hill Companies, 1975.