Distributional and Developmental Analysis of PM2.5 in Beijing, China

Alexander Guo
July 2016

Table of Contents

ABSTRACT

Introduction. Substandard air quality in Beijing, China has recently been brought to the attention of the Chinese government, and due to public pressure, emphasized as an issue of extreme importance. The Ministry of Environmental Protection of the People's Republic of China has since pursued an increasingly aggressive plan to curtail air pollution. Specifically, particulate matter with a diameter of less than 2.5 micrometers, which pose a large threat to people's health and the environment, is an issue of large concern. In addition, although the United States and China differ in the specific method to calculate air quality index (AQI) based on PM2.5 concentration, both the United States Embassy in Beijing and the government of China have increased monitoring of PM2.5 and concentrations of other pollutants in recent years, and have made real-time data available to the public. This report utilized hourly historical data from April 8th, 2008 to May 31st, 2016 from the U.S. Embassy in Chaoyang District, Beijing for two main objectives. First was to attempt to fit probability distributions to the data to better understand the data as a whole, and second was to uncover any yearly, seasonal, monthly, daily, and hourly patterns and trends that may arise. Tentative reasons and suggestions for improvements in air quality were also noted.

Methodology. In these data, 66,650 hours and 2687 days provided valid data. Lognormal, gamma, and Weibull distributions were fit to the data as a whole, as well as data for each year through an estimation of parameters. The Chi-squared test was employed to compare the actual data with the fitted distributions, and with the chi-squared distribution, p-values were computed and compared for each of the distributions within a dataset. The data were also used to uncover trends, patterns, and improvements in PM2.5 concentration over the period of time with valid data.

Results and Discussion. The data show a clear indication that Beijing's air quality is unhealthy, with an average of 94.07μg/m3 across all 66,650 hours with valid data. It was found that no distribution fit the entire dataset of all 2687 days well, but each of the three distribution types was optimal in at least one of the yearly data sets, with the lognormal distribution found to fit recent years better. An improvement in air quality beginning in 2014 was discovered, with the first five months of 2016 reporting an average PM2.5 concentration that is 23.8% lower than the average of the same period in all years. This trend may be due to the various measures the Chinese government has put into place to curtail pollution. It was also found that the winter and fall months contained more days in both good and extremely polluted categories, leading to a higher average but a comparable median in these months. Additionally, the evening hours, especially in the winter, reported much higher PM2.5 concentrations than the afternoon hours, possibly due to the prohibition of trucks in the city in the daytime and the increased use of coal for heating in the colder months when residents are home in the evening. Lastly, through analysis of special intervals that attracted media attention for either unnaturally good or bad air quality, the government's temporary pollution control measures, such as more intensive road-space rationing and factory closures, are shown to be effective

Conclusion. From this study, it is possible to conclude that air quality in Beijing is improving steadily and do follow standard probability distributions to an extent, but remains unhealthy for those living in the city. Seasonal, monthly, and daily patterns were also discovered, contributing to a complex situation of air quality. However, despite improvements and patterns, it remains clear that more needs to be done to ensure a healthy Beijing.

1 INTRODUCTION

1.1 AIR QUALITY INDEX

In the United States, air quality has been measured since 1968 when the National Air Pollution Control Administration initiated the development of an air quality index. This index, however, proved to be inconsistent due to the lack of data available, lack of agreement on weighting factors, and inconsistent air quality standards across the nation.

Air quality indexes such as abovementioned were developed following the first iteration of the U.S. Clean Air Act of 1955. Amended in 1963, 1970, 1977, and 1990, the Clean Air Act represented milestones in environmental protection within the United States. In the 1970 edition of the legislation, six pollutants were identified as criteria for calculating air quality: carbon monoxide, nitrogen dioxide, sulfur dioxide, particulate matter, hydrocarbons, and photochemical oxidants[1]. Although revised to some degree in subsequent legislation, these six pollutants have been used in air quality index (AQI) calculation in the United States and around the world since. The most relevant of these pollutants to Chinese cities is particulate matter – in particular, particulate matter with a diameter smaller than 2.5μm, emitted from factories, vehicles, and other sources[2]. PM2.5, as it is known, can enter human lungs or even the bloodstream, which can cause premature death in people with heart or lung disease, heart attacks, irregular heartbeat, aggravated asthma, decreased lung function, and increased respiratory symptoms. With regards to the environment, PM2.5 can acidify lakes and streams, pollute soil, damage sensitive forests and farm crops, affect ecosystems, and contribute to acid rain. Particulate matter also causes haze and reduced visibility.

All air quality indexes are scales in which values are calculated based on various pollutants. Standards, formulas, and warnings assigned to air quality index ranges, however, differ between nations. The United States Environmental Protection Agency (EPA), formed in 1968, was tasked to create the air quality index in the United States, and to review the National Ambient Air Quality Standards every five years to ensure safe air quality for U.S. citizens. As such, the EPA frequently updates its standards on air quality and methodology for air quality index calculation.

The following chart represents the current (2013) EPA breakpoints for AQI values[4].

Table 1.1.1 EPA Breakpoints for Pollutants[4]

Breakpoints							AQI	Category
O3 (ppm) 8-hr	O3 (ppm) 1-hr	PM10 (μg/m3) 24-hour	PM2.5 (μg/m3) 24-hr	CO (ppm) 8-hour	SO2 (ppb) 1-hour	NO2 (ppb) 1-hour	AQI	
0-0.059	-	0 - 54	0.0 – 12.0	0.0 - 4.4	0 - 35	0 - 53	0 - 50	Good
0.06 - 0.075	-	55 - 154	12.1 - 35.4	4.5 - 9.4	36 - 75	54 - 100	51 - 100	Moderate
0.076 - 0.095	0.125 - 0.164	155 - 254	35.5 - 55.4	9.5 - 12.4	76 - 185	101 - 360	101 - 150	Unhealthy for Sensitive Groups
0.096 - 0.115	0.165 - 0.204	255 - 354	55.5 - 150.4	12.5 - 15.4	(186 - 304)	361 - 649	151 - 200	Unhealthy
0.116 - 0.374	0.205 - 0.404	355 - 424	150.5 - 250.4	15.5 - 30.4	(305 - 604)	650 - 1249	201 - 300	Very unhealthy
	0.405 - 0.504	425 - 504	250.5 - 350.4	30.5 - 40.4	(605 - 804)	1250 - 1649	301 - 400	Hazardous
	0.505 - 0.604	505 - 604	350.5 - 500.4	40.5 - 50.4	(805 - 1004)	1650 - 2049	401 - 500	Hazardous

The equivalent government agency in the People's Republic of China is the Ministry of Environmental Protection (MEP), directly controlled by the State Council of the People's Republic of China. Founded in 2008 from the State Environmental Protection Agency, the MEP's mission is "to improve environmental quality and build a beautiful China which enjoys blue sky, green land and clean water.[5]" In being "responsible for establishing a sound basic system for environmental protection,[6]" the MEP's current (2012) air quality index calculation standards are delineated in Table 1.1.2)[7]

Table 1.1.2 Chinese MEP Breakpoint[7]

AQI	Pollutant Concentration Standards									
	SO_2 (24hr avg, μg/m^3)	SO_2 (1hr avg, μg/m^3)*	NO_2 (24hr avg, μg/m^3)	NO_2 (1hr avg, μg/m^3)*	PM_{10} (24hr avg, μg/m^3)	CO (24hr avg, μg/m^3)	CO (1hr avg, μg/m^3)*	O_3 (24hr avg, μg/m^3)	O_3 (8hr avg, μg/m^3)	$PM_{2.5}$ (24hr avg, μg/m^3)
0	0	0	0	0	0	0	0	0	0	0
50	50	150	40	100	50	2	5	160	100	35
100	150	500	80	200	150	4	10	200	160	75
150	475	500	180	700	250	14	35	300	215	115
200	800	650	280	1200	350	24	60	400	265	150
300	1600	800	565	2340	420	36	90	800	800	250
400	2100	**	750	3090	500	48	120	1000	***	350
500	2620	**	940	3840	600	60	150	1200	***	500

Note: *The 1-hour average concentrations for SO2, NO2, and CO should only be extracted from real-time hourly data; for daily readings, use the 24-hour average concentration.
**For 1-hour average SO2 readings greater than 800μg/m3, AQI shall not be calculated with this pollutant standard. Instead, use the 24-hour average concentration.
***For 8-hour average O3 readings greater than 800μg/m3, AQI shall not be calculated with this pollutant standard. Instead, use the 1-hour average concentration.

Even with different breakpoints in PM2.5, both the EPA and MEP present the same AQI scale and use the same formula to determine the air quality index of a PM2.5 concentration (μg/m3) that is not a breakpoint.

$$I_p = \frac{I_{Hi} - I_{Lo}}{BP_{Hi} - BP_{Lo}}\left(C_p - BP_{Lo}\right) + I_{Lo}$$

In this piecewise linear scale[7] [4], I_p is the index for pollutant p, C_p is the rounded concentration of pollutant p, BP_{Hi} is the breakpoint that is greater than or equal to C_p, BP_{Lo} is the breakpoint that is less than or equal to C_p, I_{Hi} is the AQI value corresponding to BP_{Hi}, and I_{Lo} is the AQI value corresponding to BP_{Lo}. This is computed for all relevant pollutants p, and the air quality index is computed by finding the maximum value of all such I_p. Because PM2.5 is the major pollutant in Chinese cities such as Beijing, only AQI values based on PM2.5 concentrations is relevant in this study. PM2.5 breakpoints, corresponding AQI breakpoints and both MEP and EPA categories can be summarized in Table 1.1.4.

Figure 1.1.3

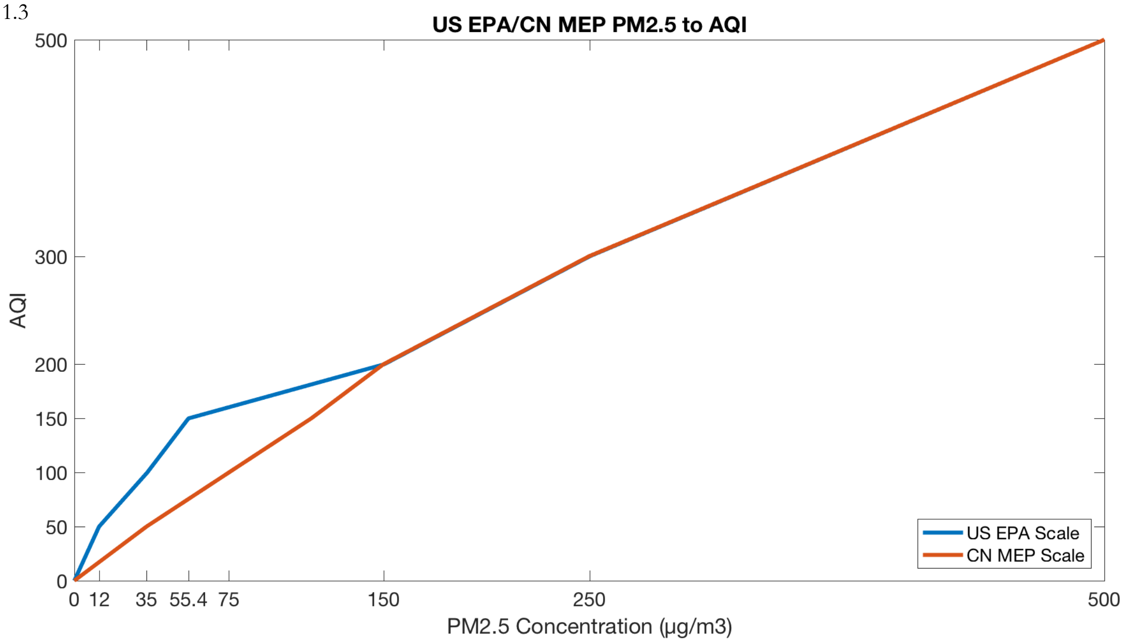


Table 1.1.4 Corresponding Breakpoints and Category Labels

EPA PM$_{2.5}$	**MEP PM$_{2.5}$**	**AQI**	**EPA Category[4]**	**MEP Category[7]**
0-12.0	0-35	0-50	Good	Excellent
12.1-35.4	35-75	51-100	Moderate	Good
35.5-55.4	75-115	101-150	Unhealthy for Sensitive Groups	Lightly Polluted
55.5-150.4	115-150	151-200	Unhealthy	Moderately Polluted
150.5-250.4	150-250	201-300	Very Unhealthy	Heavily Polluted
250.5-500	250-500	301-500	Hazardous	Severely Polluted

For comparison, the World Health Organization (WHO) Air Quality Guidelines (AQG) maintains its PM2.5 standard at 25 µg/m3 (24-hour mean)[8]. Figure 1.1.3 compares EPA and MEP PM2.5 breakpoints through conversion to AQI.

For PM2.5 concentrations under 150 µg/m3, it is evident that the same PM2.5 concentration is converted to a higher AQI value on the US scale when compared to computing with the Chinese MEP breakpoints. It should also be noted that the Chinese MEP PM2.5 to AQI relationship is more linear. In this study, all data and charts will be presented with both MEP and EPA PM2.5 breakpoints.

1.2 AIR QUALITY IN BEIJING AND CHINA: AN OVERVIEW

Due to rapid industrialization at the expense of the environment, substandard air quality has plagued Beijing for years and has attracted global media attention continuously. However, before the late 2000s, air pollution problems were widely unheard of within the Chinese population. The U.S. Department of State, in an effort to protect the health of U.S. citizens abroad, began publishing hourly readings of PM2.5 concentrations online in 2008, recorded at its embassy in Beijing. The issue of substandard air quality, however, was considered a state secret by the government of the People's Republic of China until

2012 and 2013, when officials finally admitted to the catastrophic problem of air pollution that is at hand[9].

In 2012, due to public pressure Chinese officials finally began to report PM2.5 concentrations on an hourly basis in Beijing, but additional controversies prevailed. On June 5th, 2012, for example, U.S. embassy reported a PM2.5 concentration of 47 µg/m3, categorizing this as "unhealthy." The Beijing Environmental Protection Bureau measured, with its 27 monitoring stations across Beijing, PM2.5 concentrations from 51 to 79 µg/m3, yet categorizing the air quality situation as "good." Subsequently, the Chinese government told the U.S. embassy to stop publishing their own data of air quality in China, calling US embassy readings as un-scientific, and unfair due to the fact that China is a developing country and needs less stringent air quality standards (as evident from previous section)[10].

As of 2016, China still takes issue with the PM2.5 reports of U.S. embassy and its consulates, but the government readily admits that air pollution is an enormous problem in China that must be solved. China's government has repeatedly "declared war" on pollution, promising legislation reform and stricter enforcement. The 2013 Action Plan for the Prevention and Control of Air Pollution was the first major plan implemented by the central government regarding the prevention and control of air pollution. In the 35-point plan, China acknowledged that environmental protection is "related to the great rejuvenation of the Chinese national dream," and that the current situation is "grim" and is "harming people's health and affecting social harmony and stability."[11] The plan called for a reduction of PM2.5 levels in the Beijing/Tianjin/Hebei (BTH, one of the three key regions of China, joined by the Pearl River Delta (PRD) region and the Yangtze River Delta (YRD) region) region by 25% by 2017, the control of coal consumption by prohibiting new coal-powered plants in certain regions and replacing coal with other sources of energy, a cut in iron-making and steel-making capacity by 15 million tons in 2015, the control if the number of automobiles on city roads, and an increase of non-fossil fuel energy usage, among other goals.[12]

With China's new Ministry of Environmental Protection under the direct jurisdiction of the State Council of the People's Republic of China, the MEP has become a cabinet-level ministry in the executive branch of the Chinese government with greater power than the State Environmental Protection Administration that it replaced. Its current minister is Chen Jining, a graduate from the Department of Environmental Engineering at Tsinghua University who served as the Director of the Environmental Engineering Department, as well as the president of Tsinghua University. Chen Jining is frequently praised by the Chinese public as having a background in environmental engineering, known to the people as president of a prestigious university, and his hard-liner stance on environmental protection. China's current president Xi Jinping, too, expressed his thoughts on the issue, stating that he will "punish, with an iron hand, any violators who destroy ecology or environment, with no exceptions."[13]

The Ministry of Environmental Protection (english.mep.gov.cn) has since released additional plans and legislation to combat air pollution, such as its 2014 Environmental Protection Law of the People's Republic of China (entered into force January 1, 2015). This law decrees that polluting companies can face fines of any amount, non-governmental organizations (NGOs) are welcome to sue companies, and local governments will be held accountable for implementing and enforcing policies. The MEP also provides monthly updates on air quality readings and comparisons to previous months and years[14]. Beijing's own government has additionally emphasized the need for cleaner air, demonstrated in its 2013 84-point plan[15] to tackle air quality related issues. Additionally, since the 2008 Olympics, Beijing has implemented a road space rationing policy, prohibiting 20% of all private cars from entering the city on each weekday. An increasing amount of regulation has also been implemented surrounding emissions of trucks, and such trucks are prohibited from accessing urban Beijing in daytime hours. Cars from outside Beijing are also prohibited from entering the city during daily rush hours, with certain exceptions. On days when smog alerts are issued in accordance to the system implemented in 2013, half of private cars are prohibited from the roads of Beijing each day (an odd-even system). Such a rationing system and related regulations, when coupled with the immensely difficult lottery-based license plate registration system, has decreased vehicle emissions tremendously.

Amidst such sweeping reform regarding environmental protection and air quality legislation and enforcement, critics have often expressed problems in such legislation and enforcement that continue to impair China's plan to improve its air quality. A famous documentary, *Under the Dome,* created by Chai Jing[13], one of China's most famous journalists, exposed China's air quality problem and was censored from Chinese websites following heated discussion on social media. In it, Chai Jing points to China's outdated energy structure, monopolies by state-owned natural resources companies, and lack of individual effort and foresight as issues hindering improvements in air quality. She sees large institutional obstacles and oversight issues, stemming from decentralization and inefficient power allocation. Although the central government passes strict legislation, it is, for the most part, up to local governments to enforce these standards and policies. Local governments, however, are frequently self-driven and corrupt, and as such it is difficult for change to really take place.[13] Hongjun Zhan, a former Chinese legislator, claims that legislation has improved tremendously in the past years, but enforcement is still weak and needs improvement.[9]

1.3 OBJECTIVES

Regardless of position on the issue, it is clear that China's, and specifically Beijing's, road to clean air will be long and arduous. The Chinese government, although acknowledging this fact, consistently reports that PM2.5 concentration is decreasing and the air quality is steadily improving. The ultimate objective of this report is to provide a comprehensive analysis of the air quality in Beijing from April 2008 to May 2016 and any patterns and trends in these data, to allow for the general public to quantitatively comprehend the conclusions that follow, provide preliminary analysis of causes and factors to this pollution, and give future policy and enforcement suggestions.

There are two specific objectives in this study. The first goal is to fit a probabilistic distribution of daily PM2.5 data in Beijing and to estimate reductions in PM2.5 needed to meet Chinese and U.S. AQI standards through such a distribution. It is known that various distributions, such as the lognormal, Weibull, Gamma, log-logistic, type V Pearson, and extreme-value distributions may fit data on pollutant concentration[16]. Although the lognormal distribution has been used widely to determine statistical distributions of air pollution data, other distributions have also been prominent in fitting such data. Lu (2012), for example, showed that lognormal, type V Pearson, and Weibull distributions are each appropriate in measuring different pollutants in various cities in Taiwan[17]. In Santiago, Chile, various pollutants including PM2.5 were fitted with the type-V Pearson probabilistic distribution (Morel, et. al, 1998)[18]. Choosing a correct distribution with the least difference from observed data will allow for predictions of frequency of concentrations that exceed the ambient air quality standard[17]. To the best of my knowledge, no group has analyzed recent PM2.5 data from Beijing, one of the China's most polluted cities.

The second objective of this report is to discover and analyze any trends or patterns that may arise from the data, such as seasonal, monthly, day-of-week, time-of-day, and any improvements that may be the result of stricter pollution control laws.

2 METHODOLOGY

Following the U.S. Embassy in Beijing's initiation of PM2.5 monitoring in 2008, the embassy has released hourly data as well as large amounts of historical data beginning in 2014. As of July 2016, the U.S. Department of State, through its embassy in Beijing, has published hourly data of PM2.5 concentration from April 8th, 2008, to May 31st, 2016, in pre-formatted .csv files at http://www.stateair.net/web/historical/1/1.html, available for public use and analysis. The Department of State states, in its Data Use Statement, that "State Air observational data are not fully verified or validated; these data are subject to change, error, and correction. The data and information are in no way official… If observational data are used for analyses, displayed on web pages, or used for other programs or products, the analysis results, displays, or products must indicate that these data are not fully verified or validated.[19]" As such, although not fully verified, the data are still capable of providing valuable analysis and insight into the air quality situation in Beijing.

Although it is possible that air quality data such as these are entirely random and follows no distinct statistical distribution, fitting a statistical distribution is useful for simplified analysis and future predictions. In this study, lognormal, Weibull, and Gamma distributions are used to fit the data, and the distribution with least deviation from the observed data is selected based on the chi-squared goodness-of-fit (GOF) test.

2.1 PROBABILITY DISTRIBUTIONS

The three probability distributions that will be fitted in this study, as mentioned earlier, are the lognormal, Weibull, and Gamma distributions. Each of these probability density functions (pdfs) are discussed in more detail below.

Lognormal Distribution. The pdf of the lognormal distribution, $p_l(x \mid \mu, \sigma)$, is

$$p_l(x \mid \mu, \sigma) = \frac{1}{\sqrt{2\pi} x \ln(\sigma_g)} \exp\left[-\frac{(\ln x - \ln\mu_g)^2}{2\ln(\sigma_g)^2}\right]$$

where x is the pollutant concentration of a specific pollutant species, and $\ln(\mu_g)$ and $\ln(\sigma_g)$ are the location and shape parameters that are equal to the logarithms of the geometric mean and standard geometric deviation, respectively[17].

Weibull Distribution. The pdf of the Weibull distribution, $p_w(x_i)$, is

$$p_w(x \mid a, b) = \frac{b}{a}\left(\frac{x}{a}\right)^b \exp\left[-\left(\frac{x}{a}\right)^b\right]$$

for positive values of x. b and a are the shape and scale parameters of the distribution function, respectively[20].

Gamma Distribution. The pdf of the gamma distribution $p_g(x_i \mid \alpha, \beta)$, is

$$p_g(x_i \mid \alpha, \beta), = \frac{x^{\alpha-1} e^{-x\beta}}{\beta^\alpha \Gamma(\alpha)}$$

for positive values of x. β and α are positive values that are the scale and shape parameters of the distribution function, respectively[16].

2.2 PARAMETER ESTIMATIONS

For each pdf, its parameters were estimated based on various methods that allowed for the best fit of the data. Gamma and Weibull distribution parameters were estimated with the maximum likelihood estimation (MLE) method, and σ for the lognormal distribution was estimated as the square root of the unbiased estimate of the variance of the log of the data[21]. These computations are executed with built-in libraries in MATLAB.

2.3 GOODNESS OF FIT TESTS

Various goodness-of-fit tests are employed to test the deviation of the null hypothesis from the observed data, but only a handful can be applied to continuous data in which parameters are estimated, not specified. In this study, the chi-squared goodness-of-fit test is used to determine whether to reject the null hypothesis, in the process selecting the most appropriate probability distribution to fit the data. The chi-squared test is calculated as:

$$\mathrm{X}^2 = \sum_{i=1}^{r} \frac{(O_i - E_i)^2}{E_i}$$

where O_i is the observed frequency of bin *i* and E_i is the expected frequency of bin *i* based on the null hypothesis. The number of PM2.5 concentration bins selected was based on the the following formula[22]:

$$n = \lfloor 1.88 \times N^{0.4} \rfloor$$

where N is the number of days with valid data (sample size), and *n* is the number of bins used for grouping. The number of degrees of freedom *k* is selected with the following formula:

$$k = n - 1 - p$$

where *n* is the number of bins and *p* is the number of parameters estimated when fitting the distribution.

The computed statistical value from the chi-squared test is used to calculate the *p*-value, the probability of observing a test statistic at least as extreme as the observed value under the null hypothesis, based on the chi-squared distribution for a specific degree of freedom. The smaller the value of *p,* the more likely the null hypothesis is rejected. If the *p*-value is smaller than the significance level α, which is set at 0.05, a common value, then the null hypothesis is rejected. The distribution that provided the largest p-value was selected as the optimal distribution to fit the data. It should be noted that although a large p-value provides less evidence to reject the null hypothesis, the null hypothesis should also not be accepted definitively.

Upon determination of the optimal distribution, conclusions and predictions may be drawn from the data.

2.4 ANALAYSIS OF PATTERNS AND TRENDS

The data downloaded from the U.S. Embassy database of hourly-recorded data are utilized in a variety of methods to discover trends and improvements in PM2.5 concentrations, in addition to seasonal, monthly, daily, and hourly patterns.

Trends in Data. Trends are analyzed with yearly, seasonal, and monthly data. For a yearly analysis, data are presented in tables and charts that show basic statistics of each year, improvements in averages, percentage of days in each year within breakpoints defined by the EPA and MEP, and any changes in these respective bins. Seasonal and monthly data are analyzed through comparing averages of individual seasons (3-month periods) and months throughout the period of data collection.

Patterns in Data. Seasonal, monthly, daily, and hourly patterns in the data were analyzed. Basic statistics were calculated for each of the four seasons by compiling data from each year's respective season, and for each of the twelve months by compiling data from each year's respective month. Percentages of days in each season and month category were also calculated and visualized. Hourly averages were also computed and visualized.

Following an analysis of trends and patterns, special intervals were selected and subsequently analyzed. These intervals are subsets of the data available, and chosen based on relative media coverage, either due to unnaturally good air quality or unnaturally substandard air quality. Basic statistics and percentages of days in each bin were also computed.

3 RESULTS AND DISCUSSION

3.1 PROBABILITY DISTRIBUTIONS

Table 3.1.1 is a list of parameters for each distribution for data within each year and for the entire data upon the conclusion of parameter estimation and the fitting of the respective distributions. Table 3.1.2 presents results of applying the chi-squared goodness of fit test on each of the distributions, and the computed p-value based on the chi-squared distribution for the degree of freedom specified. Lower χ values represent smaller degrees of error, and translates into larger p-values. A p-value smaller than 0.05 does not provide enough evidence to accept the null hypothesis, so the null hypothesis is rejected in these cases. Table 3.1.2 also highlights p-values greater than 0.05, signifying failure of null hypothesis rejection. It should be noted that the only dataset in which all null hypotheses were rejected is that of all data.

Table 3.1.1 Estimated Parameters

	Lognormal		Gamma		Weibull	
	μ_g	σ_g	a	b	α	β
2008*	4.2532	0.6943	2.60266	33.1455	97.1385	1.77777
2009*	4.38524	0.719089	2.30877	43.8359	113.168	1.56202
2010	4.37345	0.778579	1.99547	52.1223	115.306	1.45095
2011	4.25352	0.86937	1.61113	61.4334	107.529	1.29183
2012	4.19693	0.832727	1.81453	49.4661	98.887	1.4056
2013	4.32504	0.799853	1.81751	56.1049	111.886	1.34373
2014	4.25793	0.847135	1.68899	57.8674	106.525	1.30558
2015	4.03169	0.914754	1.451	56.9417	88.0822	1.18887
2016*	3.81149	0.908346	1.43771	46.2883	70.8962	1.18585
All Data	4.22989	0.836676	1.74468	53.8655	102.895	1.34141

*Note: 2008 and 2016 have incomplete data (less than 70% of total days in year). 2009 has a large amount of missing data in January and February, which may have an effect on the year's distribution.

Table 3.1.2 Chi-Squared Statistic and p-value

			Lognormal		Gamma		Weibull	
	Valid Days	df	X	p-value	X	p-value	X	p-value
2008*	188	12	20.732	0.054442	11.61	0.4775	11.579	0.48007**
2009*	271	14	18.841	0.1711	18.295	0.19366**	44.27	5.35e-05
2010	332	16	28.26	0.029435	23.408	0.10325**	29.123	0.023121
2011	325	16	21.043	0.17686	9.3884	0.89656	8.0591	0.94709**
2012	340	16	35.592	0.0032954	15.35	0.49919**	16.087	0.44689
2013	360	16	7.8714	0.9526**	21.675	0.15401	32.577	0.0084033
2014	359	16	24.32	0.082754**	24.825	0.072967	28.998	0.023949
2015	360	16	12.945	0.67678**	27.633	0.034959	32.318	0.0090884
2016*	152	11	18.526	0.07016**	21.633	0.027377	20.596	0.037805
All Data	2687	41	145.13665	1.4747e-13	84.69617	7.12617e-5**	138.01978	1.97795e-12

*Note: 2008 and 2016 have incomplete data (less than 70% of total days in year). 2009 has a large amount of missing data in January and February, which may have an effect on the year's distribution.
**The largest p-value among the three distributions for that year. Shaded boxes represent p-values higher than the significance level, 0.05.

Table 3.1.3 summarizes these results, noting the optimal distribution for each dataset, and whether the hypothesis to accept the distribution was rejected. It should be noted that there seems to be an increasing trend of lognormally distributed data in recent years. In Figure 3.1.4, which shows a comparison of the three distribution types on each dataset, it should be noted that the lognormal distribution frequently overestimates data in the lower bins, characteristic of more days with better air quality. This is one indication of an improving trend in Beijing's air quality. Figure 3.1.5 is similar to Figure 3.1.4; however, Figure 3.1.5 shows data and distributions of all data. It should also be noted that the lognormal distribution is also known to be perhaps the distribution that best fits air quality data, based on previous studies.

It should be noted again that all distributions for the entire dataset led to null hypothesis rejection. It is possible that the reason for this is that there exists unnatural PM2.5 concentration trends throughout the years that undermine any specific distribution. These trends are analyzed concretely and in more detail in 3.2.

Table 3.1.3 Optimal Distributions

	Optimal Distribution	Optimal p-value
2008*	Weibull	0.48007
2009*	Gamma	0.19366
2010	Gamma	0.10325
2011	Weibull	0.94709
2012	Gamma	0.49919
2013	Lognormal	0.9526
2014	Lognormal	0.082754
2015	Lognormal	0.67678
2016*	Lognormal	0.07016
All Data	Gamma	7.12617e-5

*Note: 2008 and 2016 have incomplete data (less than 70% of total days in year). 2009 has a large amount of missing data in January and February, which may have an effect on the year's distribution.
** Shaded boxes represent p-values higher than the significance level, 0.05.

It is now possible to estimate the probability in which the PM2.5 concentration exceeds standards. Table 3.1.6 compares actual percentage of days exceeding a certain standard and the estimated percentage of days

Figure 3.1.4

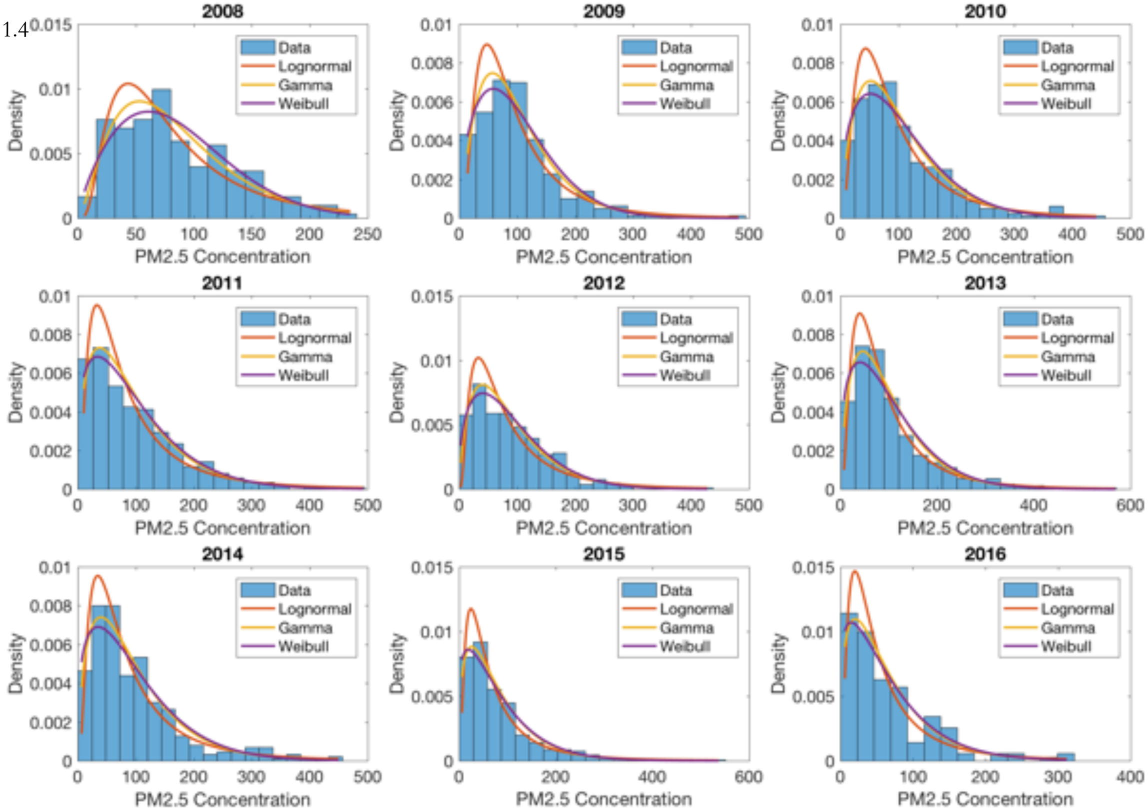
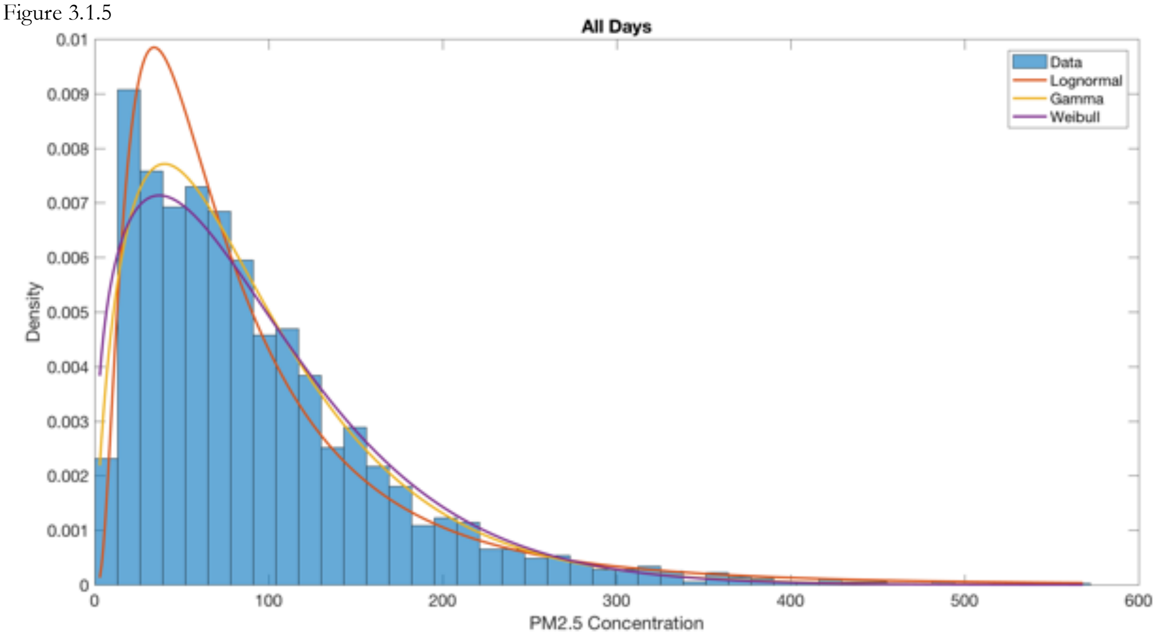


Figure 3.1.5



exceeding the respective standard for each year's optimal distribution. It is clear that the distributions are accurate to a certain extent, with differences amounting to, on average, a few percentage points. Thus, this study has demonstrated the capability to fit distributions on yearly PM2.5 data in Beijing, and has found the Weibull, Gamma, and Lognormal distributions to somewhat accurately fit the data, with lognormal distributions more frequently overestimating data at smaller concentrations but remaining the optimal distribution in recent years.

Table 3.1.6 Estimated Percent Of Days Exceeding AQI Standard

	Act. # d.>12 μg m³	Actual %>12 μg/m³	Est. %>12 μg/m³	Act. # d.>25 μg m³	Actual %>25 μg/m³	Est. %>25 μg/m³	Act. # d.>35 μg m³	Actual %>35 μg/m³	Est. %>35 μg/m³	Act. # d.> 150	Actual %> 150	Est. %> 150
2008*	184	97.9%	94.8%	173	92.0%	87.6%	155	82.4%	80.6%	24	12.8%	9.3%
2009*	271	100.0%	92.8%	246	90.8%	84.8%	231	85.2%	77.6%	47	17.3%	15.7%
2010	329	99.1%	93.6%	298	89.8%	85.9%	277	83.4%	79.1%	74	22.3%	18.9%
2011	323	99.4%	88.3%	270	83.1%	79.4%	247	76.0%	72.6%	72	22.2%	18.9%
2012	335	98.5%	94.9%	289	85.0%	86.0%	255	75.0%	78.1%	60	17.6%	15.4%
2013	355	98.6%	95.5%	323	89.7%	85.8%	297	82.5%	76.8%	72	20.0%	17.6%
2014	349	97.2%	93.5%	313	87.2%	82.4%	282	78.6%	73.0%	64	17.8%	16.1%
2015	337	93.6%	89.7%	292	81.1%	74.7%	262	72.8%	63.8%	52	14.4%	12.7%
2016*	140	92.1%	84.1%	109	71.7%	65.6%	92	60.5%	53.6%	13	8.6%	7.1%
All Data	2623	97.6%	95.0%	2313	86.1%	86.3%	2098	78.1%	78.8%	478	17.8%	17.5%

*Note: 2008 and 2016 have incomplete data (less than 70% of total days in year). 2009 has a large amount of missing data in January and February, which may have an effect on the year's distribution.

** 12 μg/m3, 25μg/m3, and 35μg/m3 are the standards for the US EPA, WHO, and CN MEP, respectively. 150μg/m3 converts to around AQI 200 on both EPA and MEP scales.

3.2 STATISTICS, PATTERNS AND TRENDS

GENERAL ANALYSIS

The mean of all recorded values, data from 65,660 hours following 4,419 hours removed due to missing data, is 94.07 (AQI: EPA 171; MEP 124) with a median of 69 (AQI: EPA 158; MEP 92.5) and a large standard deviation of 88.46, as seen from the distribution in 3.1.

Similarly, when each the arithmetic mean of the hours of each day are computed, 296 days are removed due to insufficient data (under the circumstances that at least 6 hours out of 24 reported missing values). Thus, from 2983 days, 2687 days remain with valid data. With this metric, the arithmetic mean of daily PM2.5 concentrations is 93.98 (AQI: EPA 171; MEP 124) with a median of 74 (AQI: EPA 161; MEP 99) and a large standard deviation of 74.97. This standard deviation, however, is smaller than the standard deviation of hourly data, possibly because of outliers removed through averaging each day's values. As such, all values of PM2.5 will be tabulated and calculated from daily PM2.5 average concentrations henceforth unless otherwise specified.

An analysis of consecutive sequence of days in which the average PM2.5 concentration remained between certain values is shown in Table 3.2.1. Such sequences show that by US EPA AQI standards, a daily average of "Good" (AQI 0-50) has never been achieved for a period of time longer than three consecutive days, whereas the longest string of "Unhealthy" or worse days (PM2.5 54.4+, US EPA AQI 150+) by U.S. EPA standards is 25 days, and 10 days by CN MEP standards (PM2.5 75+, MEP AQI 150+). It should be noted that on average, a sequence of consecutive days in which the PM2.5 concentration remained above 25 μg/m3, the World Health Organization upper limit, was 9.84 days, with 2313 total days. The average number of days in a sequence in which the PM2.5 concentration was below the WHO Air Quality Standard (AQS), in comparison, was only 1.59 days, with 373 total days. Such data show that substandard air quality, by WHO standards, plagued Beijing in 86% of days between 2008-4-1 and 2016-5-31. The issue of pollution in Beijing, then, is clearly extremely serious and needing of efficient remedy. To illustrate some extreme examples of all data and daily averages, Table 3.2.2 is a list of the twenty best and worst such data points.

Table 3.2.3 is a description of bin assignments and color codes that will be used extensively throughout the remainder of this report.

Table 3.2.1: Analysis of Continuous Sequences**

PM2.5 Range	Significance	Longest-First Day	Longest -Last Day	Longest Duration	Average Duration	% of Ind. Seq. >1 d.	Total # of days
0-12	EPA AQI 0-50	2015-09-10	2015-09-12	3 days	1.21 days	15.1%	64
0-35/35.4*	MEP AQI 0-50 and US EPA AQI 0-100	2015-08-20	2015-09-03	15 days	1.86 days	46.9%	594
0-75	MEP AQI 0-100	2015-08-14	2015-09-14	32 days	3.11 days	68.8%	1354
55.4+	EPA AQI 150+	2009-07-26	2009-08-19	25 days	4.19 days	76.2%	1723
75+	MEP AQI 150+	2009-11-21	2009-11-30	10 days	2.23 days	58.5%	775
150/150.4+*	MEP/EPA AQI 200+	2011-02-15	2011-02-23	9 days	1.97 days	50.0%	478
250/250.4+*	MEP/EPA AQI 300+	2014-02-20	2014-02-26	7 days	1.67 days	32.9%	122
0-25	Below WHO AQS	2015-8-21	2015-8-28	8 days	1.59 days	37.2%	373
25+	Above WHO AQS	2009-07-01	2009-09-05	68 days	9.84 days	91.1%	2313

* Note: MEP/EPA breakpoints differ as such, but data for these breakpoints are very similar, and thus they are included together.

** If data from a day following a consecutive sequence are missing, the sequence is not terminated if the subsequent day's PM2.5 concentration lies between the concentrations required for a day to be within a sequence.

Table 3.2.2a Extreme Hours

Year	Month	Day	Hour	$PM_{2.5}$
2012	9	28	9	0
2012	9	28	14	0
2015	3	21	16	0
2015	10	1	8	0
2009	10	3	3	1
2010	9	21	2	1
2012	6	10	17	1
2012	9	28	15	1
2012	9	28	16	1
2012	9	30	2	1
				…
2013	1	12	22	805
2013	1	12	16	810
2013	1	12	18	824
2013	1	12	15	845
2013	1	12	20	852
2013	1	12	21	858
2013	1	12	19	886
2012	1	23	1	972
2010	2	14	0	980
2012	1	23	0	994

Table 3.2.2b Extreme Days

Year	Month	Day	$PM_{2.5}$
2012	9	28	2.92
2008	10	23	5.83
2015	12	15	6.08
2016	4	2	6.13
2015	10	10	6.26
2014	11	12	6.54
2015	6	12	6.63
2015	12	2	6.75
2014	10	12	6.79
2015	10	11	7.04
			…
2012	1	19	428.54
2013	1	29	429.78
2010	12	21	441.50
2014	1	16	448.42
2014	2	25	449.75
2015	12	1	464.38
2009	11	7	482.25
2011	2	21	492.75
2015	12	25	537.25
2013	1	12	568.57

Table 3.2.3 Scale Breakpoint Ranges, Bin Assignments, and Colors

Bin Number	CN MEP PM2.5 Range (μg/m3)	US EPA PM2.5 Range (μg/m3)	AQI (CN MEP & US EPA)	Color (R, G, B) [4]
1	0-35	0-12	0-50	Green (0,228,0)
2	35-75	12.1-35.4	51-100	Yellow (255,255,0)
3	75-115	35.5-55.4	101-150	Orange (255,126,0)
4	115-150	55.5-150.4	151-200	Red (255,0,0)
5	150-250	150.5-250.4	201-300	Purple (143,63,151)
6	250-500	250.5-500	301-500	Maroon (126,0,35)
7	500+	500+	500+	Black (0,0,0)

YEARLY PATTERNS AND TRENDS

Figure 3.2.4

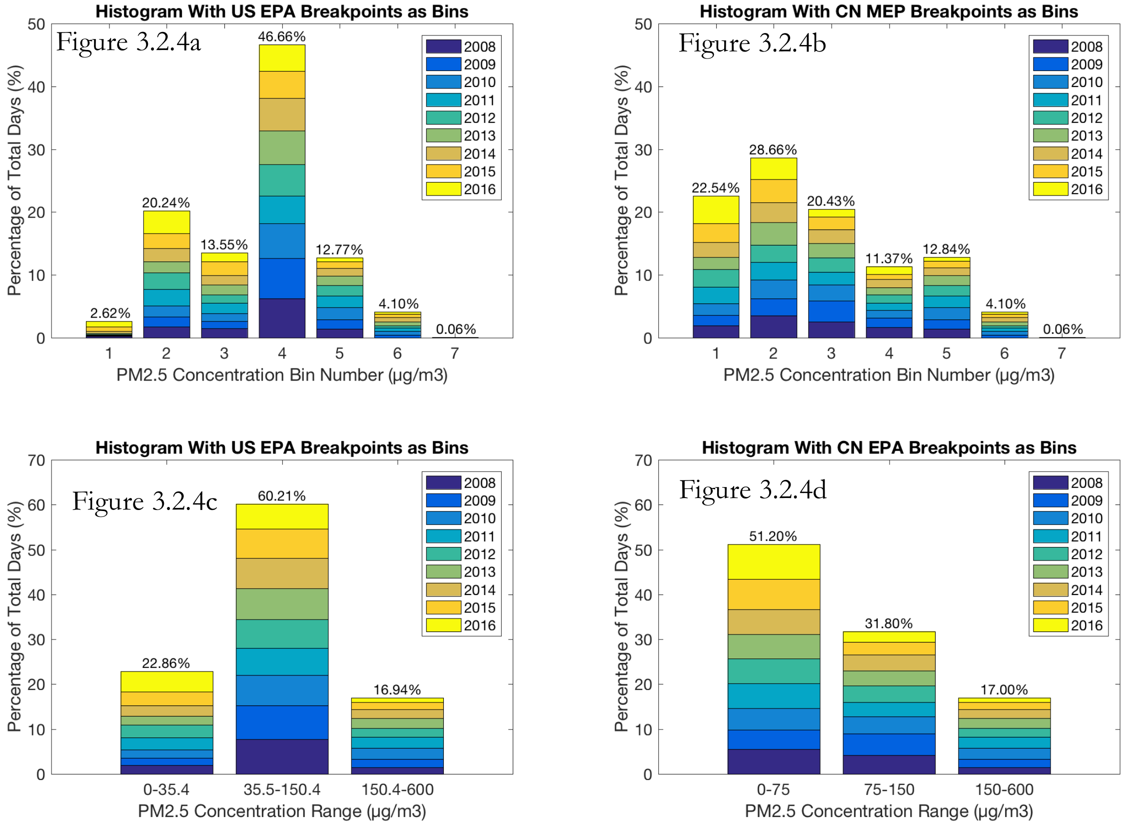


Table 3.2.5a Yearly Histogram Data for US EPA Breakpoints

Year	Tot. # val. d	Bin 1 #d	Bin 1 %d	Bin 2 #d	Bin 2 %d	Bin 3 #d	Bin 3 %d	Bin 4 #d	Bin 4 %d	Bin 5 #d	Bin 5 %d	Bin 6 #d	Bin 6 %d	Bin 7 #d	Bin 7 %d
2008	188	4	2.1	29	15.4	25	13.3	106	56.4	24	12.8	0	0	0	0
2009	271	0	0	40	14.8	28	10.3	156	57.6	36	13.3	11	4.1	0	0
2010	332	3	0.9	52	15.7	38	11.4	165	49.7	57	17.2	17	5.1	0	0
2011	325	2	0.6	76	23.4	48	14.8	128	39.4	54	16.6	17	5.2	0	0
2012	340	5	1.5	82	24.1	39	11.5	154	45.3	51	15	9	2.6	0	0
2013	360	5	1.4	58	16.1	52	14.4	173	48.1	50	13.9	21	5.8	1	0.3
2014	359	10	2.8	67	18.7	50	13.9	168	46.8	40	11.1	24	6.7	0	0
2015	360	23	6.4	76	21.1	71	19.7	139	38.6	33	9.1	17	4.7	1	0.3
2016	152	12	7.9	50	32.9	19	12.5	58	38.2	9	5.9	4	2.6	0	0
Total	2687	64	2.6	530	20.2	370	13.6	1247	46.7	354	12.8	120	4.1	2	0.1

*Note: 2008. 2009 and 2016 have incomplete data (less than 75% of total days in year), which may have an effect on the year's distribution, and thus should not be compared directly to other years (see table 3.2.6)

Table 3.2.5b Yearly Histogram Data for CN MEP Breakpoints

Year	Tot. # val. d	Bin 1 #d	Bin 1 %d	Bin 2 #d	Bin 2 %d	Bin 3 #d	Bin 3 %d	Bin 4 #d	Bin 4 %d	Bin 5 #d	Bin 5 %d	Bin 6 #d	Bin 6 %d	Bin 7 #d	Bin 7 %d
2008	188	33	17.6	60	31.9	43	22.9	28	14.9	24	12.8	0	0	0	0
2009	271	40	14.8	66	24.4	81	29.9	37	13.7	36	13.3	11	4.1	0	0
2010	332	55	16.6	89	26.8	78	23.5	36	10.8	57	17.2	17	5.1	0	0
2011	325	78	24	82	25.2	59	18.2	34	10.5	55	16.9	17	5.2	0	0
2012	340	85	25	84	24.7	70	20.6	41	12.1	51	15	9	2.6	0	0
2013	360	63	17.5	115	31.9	74	20.6	36	10	50	13.9	21	5.8	1	0.3
2014	359	77	21.5	103	28.7	71	19.8	44	12.3	40	11.1	24	6.7	0	0
2015	360	98	27.2	120	33.3	65	18.1	25	6.9	34	9.4	17	4.7	1	0.3
2016	152	59	38.8	47	30.9	16	10.5	17	11.2	9	5.9	4	2.6	0	0
Total	2687	588	22.5	766	28.7	557	20.4	298	11.4	356	12.8	120	4.1	2	0.1

*Note: 2008. 2009 and 2016 have incomplete data (less than 75% of total days in year), which may have an effect on the year's distribution, and thus should not be compared directly to other years (see table 3.2.6)

Figure 3.2.6

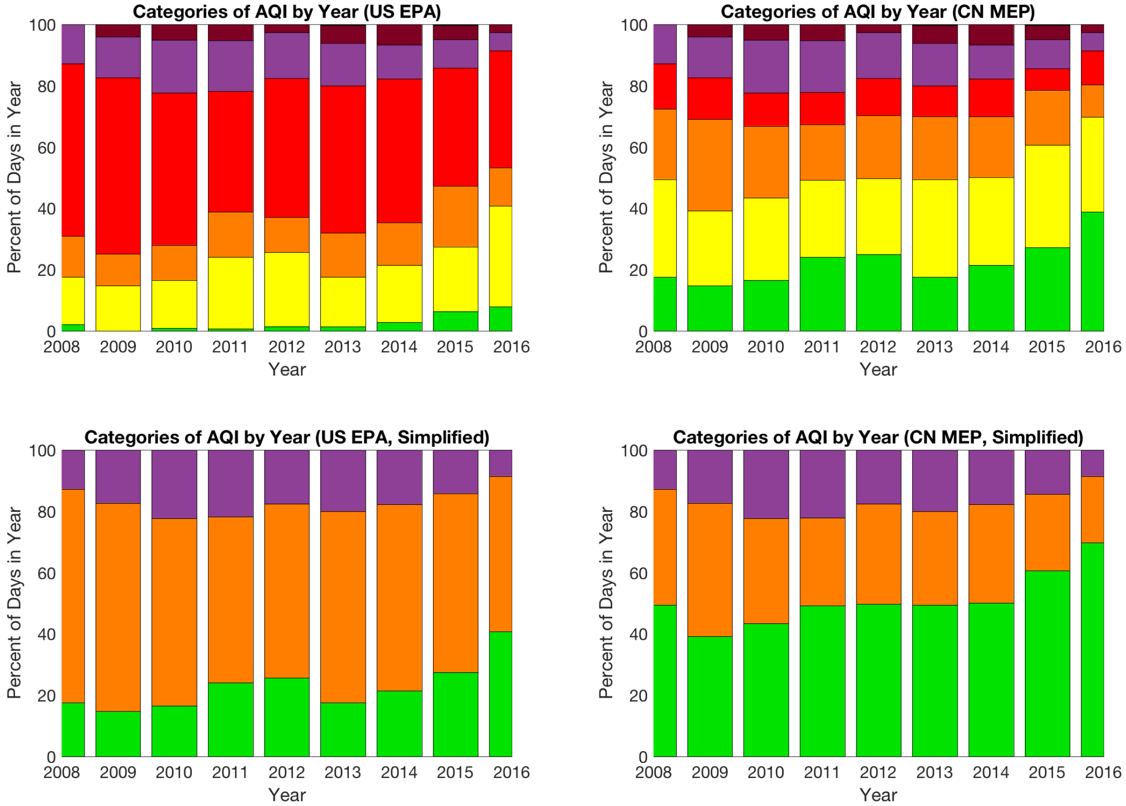


Table 3.2.7 Basic Yearly Statistical Analysis

Year	Min	25th	Median (Rank)	75th	Max	Mean (Rank)	IQR	STD	Avg YoY %Change
2008*	5.8	46.6	75.2 (4)	121.6	235.4	86.3 (3)	75.0	50.7	
2009	14.6	55.3	88.3 (9)	125.2	482.3	101.2 (7)	69.9	69.5	+17.3%
2010	9.9	49.7	84.4 (8)	138.5	441.5	104.0 (9)	88.8	76.4	+2.8%
2011	10.4	36.6	77.5 (7)	141.4	492.8	99.0 (6)	104.7	80.8	-4.8%
2012	2.9	35.0	75.4 (5)	125.3	428.5	89.8 (4)	90.3	66.6	-9.3%
2013	7.4	47.4	76.4 (6)	130.4	568.6	102.0 (8)	83.0	83.1	+13.6%
2014	6.5	41.8	75.0 (3)	124.5	450.0	97.7 (5)	82.8	90.0	-4.2%
2015	6.1	30.9	58.2 (2)	106.0	537.2	82.6 (2)	75.1	76.7	-15.5%
2016*	6.1	21.9	47.0 (1)	85.1	312.6	66.5 (1)	63.3	61.0	-19.5%
All Days	2.9	39.6	74.2	123.8	568.6	94.0	84.3	75.0	

*Note: 2008 and 2016 have incomplete data (less than 70% of total days in year). 2009 has a large amount of missing data in January and February, which may have an effect on the year's distribution.

Figure 3.2.4a is a histogram broken into bins defined by the US EPA PM2.5 breakpoints, and Figure 3.2.4b is the corresponding histogram for CN MEP breakpoints. Figure 3.2.4c and 3.2.4d represent the same data as Figures 3.2.4a and Figure 2.2.4b, respectively, but with some bins combined. Bins 1-7 in Figures 3.2.4a and 3.2.4b correspond to the respective breakpoints in Tables 1.1.1 and 1.1.2, and color and breakpoint references can be found in Table 3.2.3. For example, bin 4 represents an AQI of 150-200 ("Unhealthy" with US EPA breakpoints and "Moderately Polluted" with CN MEP breakpoints). Figures 3.2.4c and 3.2.4d combine bins 1-2, 3-4, and 5-7 into three bins for more convenient visualization. It can be noted that there is a great difference between these two standard systems, partly due to the large range of PM2.5 concentrations (55.4-150) categorized in bin 4 (AQI 150-200) by US EPA standards. It should also be noted that although there were only 188 valid days in 2008, 271 valid days in 2009, and 152 valid days in 2016, all were incorporated into the data in an effort to ensure comprehensiveness. However, due to such incomplete data only covering certain parts of the year (seasons have a significant effect on PM2.5, as will be shown later), they should not be directly compared with data from other years, which are more comprehensive.

From the histograms in Figure 3.2.4, it is apparent that there seems to be an increasing percentage of days in the bins with small PM2.5 concentrations, and a decreasing percentage of days in the bins with large PM2.5 concentrations. Table 3.2.5 shows this in tabulated data form, and Figure 3.2.6 displays this trend in a bar graph. In Table 3.2.5, each cell in each percentages column is colored a shade of its representative bin color; with a shade closer to the representative bin color representing a higher value, and a shade closer to white representing a lower value. It should be noted that for bins 1 and 2 with both EPA and MEP standards, higher percentages occurred more recently, while in bins 4, 5, and 6, lower percentages occurred more recently.

These gains can be visualized most easily in Figure 3.2.9 (Figure 3.2.9a is with US EPA standards; Figure 3.2.9b is with CN MEP standards), nine graphs in which six represent trends in each of the six AQI categories, and three represent trends in combined bins of AQI categories. As mentioned earlier, data from 2008, 2009, and 2016 should not be compared directly to those of other years due to a lack of data. Thus, this section of the report will focus on comparing 2015 all-year data to the average in addition to comparing the first five months of 2016 to the same periods of previous years.

Especially notable is the increase of "Good" days by US EPA standards in 2015: 6.4% of days in 2015 were in bin 1 (PM2.5 0-12, AQI 0-50), 3.8 percentage points (146%) higher than the average of 2.6% and 5.8 percentage points higher than the low of 0.6 (967%, 2011) in valid years. There were 23 such days in 2015, almost four times the average. In the first 5 months of 2016, there have already been more of such days than the entirety of 2014, and at least double that of every other year except for 2014 and 2015.

By Chinese MEP standards, Beijing's air quality has also made significant gains. Days in bin 1 (MEP 0-35, AQI 0-50, "Excellent") have reached 27.2% in 2015, 9.7 percentage points (55.4%) higher than that of 2013,

Table 3.2.8a Yearly Histogram Data for US EPA Breakpoints (January 1st to May 31st)

Year	Tot. # val. d	Bin 1 #d	Bin 1 %d	Bin 2 #d	Bin 2 %d	Bin 3 #d	Bin 3 %d	Bin 4 #d	Bin 4 %d	Bin 5 #d	Bin 5 %d	Bin 6 #d	Bin 6 %d	Bin 7 #d	Bin 7 %d
2008	53	0	0	8	15.1	2	3.8	32	60.4	11	20.8	0	0	0	0
2009	81	0	0	15	18.5	9	11.1	52	64.2	4	4.9	1	1.2	0	0
2010	144	0	0	29	20.1	19	13.2	72	50	21	14.6	3	2.1	0	0
2011	128	1	0.8	42	32.8	24	18.8	44	34.4	11	8.6	6	4.7	0	0
2012	144	1	0.7	33	22.9	16	11.1	66	45.8	23	16	5	3.5	0	0
2013	149	3	2	20	13.4	18	12.1	69	46.3	23	15.4	15	10.1	1	0.7
2014	151	3	2	23	15.2	14	9.3	74	49	23	15.2	14	9.3	0	0
2015	148	5	3.4	23	15.5	28	18.9	71	48	18	12.2	3	2	0	0
2016	152	12	7.9	50	32.9	19	12.5	58	38.2	9	5.9	4	2.6	0	0
Total	1150	25	2.2	243	21.1	149	13	538	46.8	143	12.4	51	4.4	1	0

*Note: 2008 and 2009 do not have enough valid days to meet the standard (75% of total days in first five months, or about 113 days), and thus should not be directly compared to other years.

when the first major pollution legislation was enacted. Comparing 2016 data, it is also clear that air quality is improving, as “Excellent” days (MEP PM2.5 0-35, AQI 0-50) in the first five months of 2016 have already almost surpassed those in all of 2013.

Table 3.2.7 shows some basic statistics of each year. It should be noted that there is a large range, interquartile range, and standard deviation in the data of each year, showing the wide distribution of data. 2016 has been, according to the data, much cleaner than the average (only 70.8% of the average), but, like mentioned earlier, this comparison is not completely fair.

Tables 3.2.8a and 3.2.8b show a comparison between the first five months of each year, such that 2016 data can be directly compared to previous years. Especially notable is the increase in MEP “Excellent” days and EPA “Good” and “Moderate” days (MEP PM2.5 0-35, AQI 0-50; EPA PM2.5 0-35.4, AQI 0-100). 40.8% of days in the first five months of 2016 are in bins 1 and 2 for US breakpoints. This is an increase of 17.5 percentage points (75%) from the average of 23.3%,

Table 3.2.8b Yearly Histogram Data for CN MEP Breakpoints (January 1st to May 31st)

Year	Tot. # val. d	Bin 1 #d	Bin 1 %d	Bin 2 #d	Bin 2 %d	Bin 3 #d	Bin 3 %d	Bin 4 #d	Bin 4 %d	Bin 5 #d	Bin 5 %d	Bin 6 #d	Bin 6 %d	Bin 7 #d	Bin 7 %d
2008	53	8	15.1	12	22.6	14	26.4	8	15.1	11	20.8	0	0	0	0
2009	81	15	18.5	26	32.1	26	32.1	9	11.1	4	4.9	1	1.2	0	0
2010	144	29	20.1	43	30	35	24.3	13	9	21	14.6	3	2.1	0	0
2011	128	42	33.6	41	32	17	13.3	10	7.8	11	8.6	6	4.7	0	0
2012	144	34	23.6	37	25.7	27	18.9	18	12.5	23	16	5	3.5	0	0
2013	149	23	15.4	43	28.9	29	19.5	15	10.1	23	15.4	15	10.1	1	0.7
2014	151	26	17.2	36	23.8	33	21.9	19	12.6	23	15.2	14	9.3	0	0
2015	148	28	18.9	52	35.1	30	20.3	16	10.8	18	12.2	3	2	0	0
2016	152	59	38.8	47	30.9	16	10.5	17	11.2	9	5.9	4	2.6	0	0
Total	1150	234	20.3	337	29.3	227	19.7	125	10.9	143	12.4	51	4.4	1	0

*Note: 2008 and 2009 do not have enough valid days to meet the standard (75% of total days in first five months, or about 113 days), and thus should not be directly compared to other years.

and an increase of 25.4 percentage points from such days in 2013. As a purely subjective addition, such days in bins 1 and 2 by US breakpoints and bin 1 by Chinese breakpoints correspond to the largest percentage of “blue skies” days, and is thus particularly important to analyzing the trend of days with satisfactory air quality in Beijing. In addition, 7.9% of days in the first five months of 2016 are “Good” (US EPA Bin 1), a rise of 5.7 percentage points from the average of 2.2%. There have already been 12 such days in 2016, almost half of total such days from 2008 to 2016 in which data are available.

In addition to an increase in days with air quality better than the standard, there has been a substantial decrease of days with “Unhealthy” air quality by some standards. 38.2% of days in 2016 have had an average PM2.5 concentration between 55.4 and 150.4 (US EPA bin 4, AQI 150-200), down 8.6 percentage points (18.4%) from the average of 46.8%. It should be noted, however, that such a decrease was not as substantial as the increase in satisfactory days, as analyzed earlier. By Chinese MEP standards, days in bin 4 have actually occurred more frequently in 2016 than the 9-year average, diminishing an extremely optimistic outlook based on previous analysis. However, it should be apparent that a trend of decreasing bad air quality days is still forming, as evident in the 9.5 percentage point (61%) drop in “Very Unhealthy” (EPA) or “Heavily Polluted” (MEP) (bin 5) from the 15.4% in 2013, and a 7.5 percentage point (74%) drop in “Hazardous” (EPA) or “Severely Polluted” (MEP) (bin 6) from the 10.1% in 2013. Although such drops are substantial, it should also be noted that such days in 2013 are among the highest of all years. This, along with data from Table 3.2.5, leads to a conclusion that the air decreased in quality from 2008 to 2013, but has improved steadily since then in almost all measures. A possible reason for this trend is the initial focus on the economy by the Chinese government before 2013, followed by an increased emphasis on the environment in subsequent years.

In all of Figures 3.2.9a and 3.2.9b, two linear trend lines are plotted: one from 2008 to 2016, and one from 2013 to 2016, the period in which various pollution laws were in force. Regardless of the fact that only the first five months of 2016 are represented, it is clear that a positive trend is developing, especially with regards to data since 2013, leading Beijing to a future with cleaner air.

It can also be tentatively concluded from comparing Jan-May data to full-year data that the distribution within the first five months is similar to the distribution of the entire year based on percentage of days in the various bins. There as some discrepancies, however, such as the fact that only 5 days in the first 5 months of 2015 had an average PM2.5 concentration of under 12, but that there were 23 total such days in the year of 2015. However, such an analysis is extremely tentative and distribution within a year is analyzed further in a later section.

Figure 3.2.9a

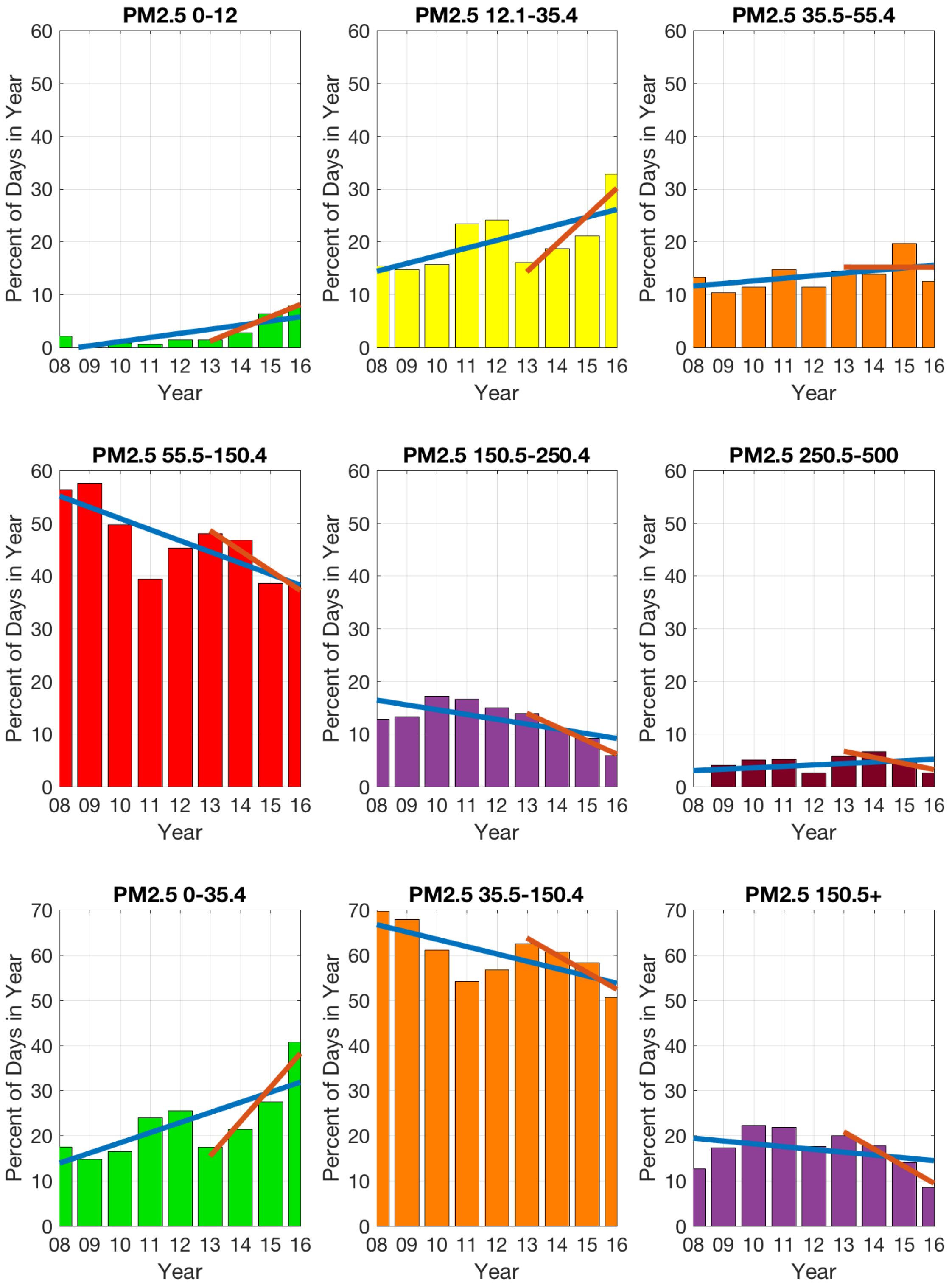

PM2.5 0-12
PM2.5 12.1-35.4
PM2.5 35.5-55.4
PM2.5 55.5-150.4
PM2.5 150.5-250.4
PM2.5 250.5-500
PM2.5 0-35.4
PM2.5 35.5-150.4
PM2.5 150.5+
Percent of Days in Year
Year
0
10
20
30
40
50
60
70
08
09
10
11
12
13
14
15
16

Figure 3.2.9b

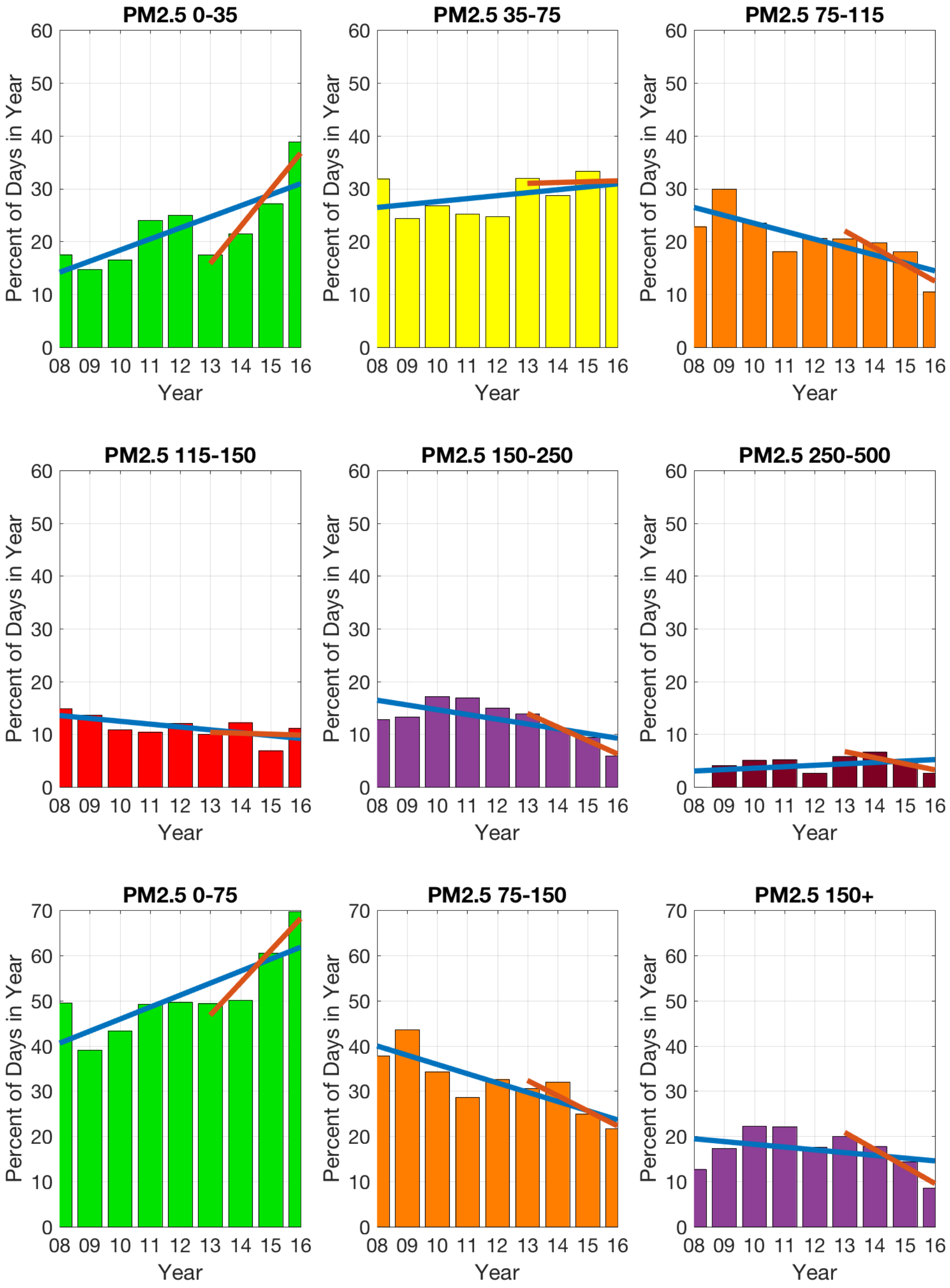

PM2.5 0-35
PM2.5 35-75
PM2.5 75-115
PM2.5 115-150
PM2.5 150-250
PM2.5 250-500
PM2.5 0-75
PM2.5 75-150
PM2.5 150+
Percent of Days in Year
Year
0
10
20
30
40
50
60
70
08
09
10
11
12
13
14
15
16

Figure 3.2.10

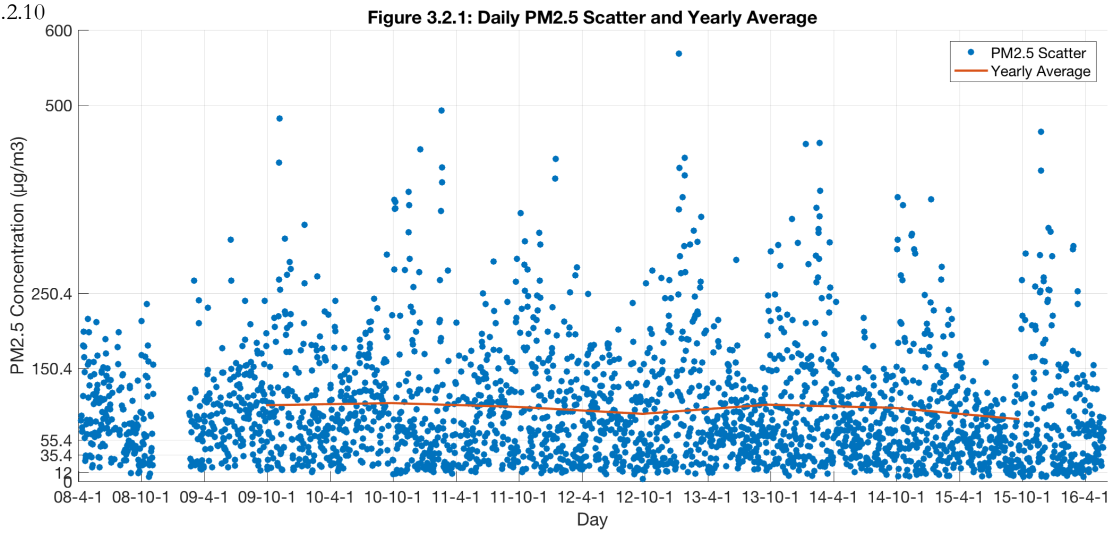
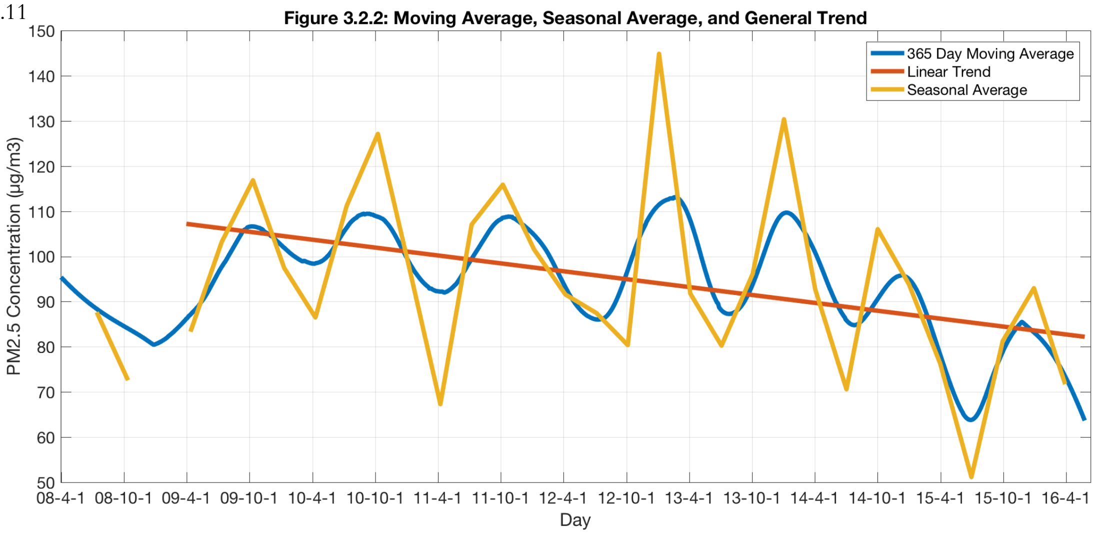


Figure 3.2.11



SEASONAL PATTERNS AND TRENDS

Figure 3.2.10 shows a basic scatter plot of daily average PM2.5 concentrations in the time period between April 1st, 2008, and May 31st, 2016, the extent of available data. The y-axis is labeled according to US EPA breakpoints for contextualization and simplicity of visualization, and a yearly PM2.5 concentration average 2D line is plotted through the data. The x-axis is labeled in a yy-mm-dd format. From this graph, a pattern begins to appear yearly: most of the days with average PM2.5 concentration above 250.4 µg/m3 appear to be in the fall and winter months (between 10/1 and 4/1 of the following year). Such an observation is echoed in Figure 3.2.11, a 365-day moving average, average by season (3-month period), and overall trend line. This pattern is, without doubt, evident in Figure 3.2.11, with the 365-day moving average resembling a sine or cosine wave with a period of about a year and a large amplitude. The seasonal average seems to be more extreme, with winter months resulting in a much higher average PM2.5 concentration than spring months. This is shown more clearly in Table 3.2.12.

Table 3.2.12 Basic Seasonal Statistical Analysis

Year	Min	25th	Median (Rank)	75th	Max	Mean (Rank)	IQR	STD	Mean %of All Avg
Spring	6.1	40.5	69.6 (1)	113.7	351.3	84.2 (1)	73.2	58.8	89.6%
Summer	6.6	45.0	78.6 (4)	117.1	321.2	86.5 (2)	72.1	52.5	92.0%
Fall	2.9	37.7	72.5 (2)	138.2	482.3	99.8 (3)	100.5	83.5	106.2%
Winter	6.1	34.4	76.9 (3)	149.4	568.6	107.7 (4)	114.9	97.7	114.6%
All Days	2.9	39.6	74.2	123.8	568.6	94.0	84.3	75.0	100%

Upon analyzing such data in more detail, however, it appears that the trend that emerges is not as simple as when a simple mean is computed for each season. It appears that fall and winter days, with lower 25th percentiles but higher 75th percentiles than spring and summer days, have larger IQRs and STDs. From Table 3.2.13a and b, there have been more "Good" (bin 1, US EPA PM2.5 0-12 AQI 0-50) days in the fall and winter, contrary to the conclusions made by comparing averages. In addition, spring and summer have less "Very Unhealthy/Heavily Polluted" (bin 5, EPA/MEP PM2.5 150.5/150-250.4/250, AQI 200-300) and "Hazardous/Severely Polluted" (bin 6, EPA/MEP PM2.5 250.4/250-500, AQI 300-500) days, but more "Unhealthy" (bin 4, EPA PM2.5 55.4-150.4 AQI 100-150) days (bin 4 in MEP are similar across the four seasons). However, it is clear that there are many more days in bin 6 in the fall and winter than in the spring and summer, and that the maximum days in the fall and winter are much more polluted than those of the spring and summer. Days above AQI 150 by US EPA standards are 13.24%, 12.1%, 21.8%, and 24.8% in the spring, summer, fall, and winter, respectively, showing that such days occur around twice as much in the fall and winter than in the spring and summer. Days that have more satisfactory air quality seem to exist significantly less in the summer, however (17.5% in bin 1 of MEP), than in the spring, fall and winter (21.8%, 23.2% and 25.5%, respectively). As such, it is clear that more heavily polluted days occur in the fall and winter, whereas satisfactory days fall more frequently in the spring, fall, and winter. The average of PM2.5 concentration in the winter and fall may be higher due to these extreme values; the medians, as shown in Table 3.2.12, are very similar. A visualization of this pattern is clearly shown in Figure 3.2.14 with the larger middle sections in the spring and summer and larger extreme sections for the fall and winter. A reason for a higher frequency of polluted days in the fall and winter may be the use of coal for

Table 3.2.13a Seasonal Histogram Data for US EPA Breakpoints

Year	Tot. # val. d	Bin 1 #d	Bin 1 %d	Bin 2 #d	Bin 2 %d	Bin 3 #d	Bin 3 %d	Bin 4 #d	Bin 4 %d	Bin 5 #d	Bin 5 %d	Bin 6 #d	Bin 6 %d	Bin 7 #d	Bin 7 %d
Spring	740	12	1.6	149	20.1	105	14.2	376	50.8	84	11.4	14	1.9	0	0
Summer	670	5	0.7	114	17	102	15.2	368	54.9	77	11.5	4	0.6	0	0
Fall	669	26	3.9	130	19.4	98	14.6	269	40.2	99	14.8	47	7	0	0
Winter	608	21	3.5	137	22.5	65	10.7	234	38.5	94	15.5	55	9	2	0.3
Total	2687	64	2.4	530	19.7	370	13.8	1247	46.4	354	13.2	120	4.5	2	0.1

Table 3.2.13b Seasonal Histogram Data for CN MEP Breakpoints

Year	Tot. # val. d	Bin 1 #d	Bin 1 %d	Bin 2 #d	Bin 2 %d	Bin 3 #d	Bin 3 %d	Bin 4 #d	Bin 4 %d	Bin 5 #d	Bin 5 %d	Bin 6 #d	Bin 6 %d	Bin 7 #d	Bin 7 %d
Spring	740	161	21.8	238	32.3	162	21.9	80	10.8	85	11.5	14	1.9	0	0
Summer	670	117	17.5	197	29.4	185	27.6	89	13.3	78	11.6	4	0.6	0	0
Fall	669	155	23.2	188	28.1	111	16.6	69	10.3	99	14.8	47	7	0	0
Winter	608	155	25.5	143	23.5	99	16.3	60	9.9	94	15.5	55	9	2	0.3
Total	2687	588	21.9	766	28.5	557	20.7	298	11.1	356	13.2	120	4.5	2	0.1

heating for many living in Beijing.

Figure 3.2.14

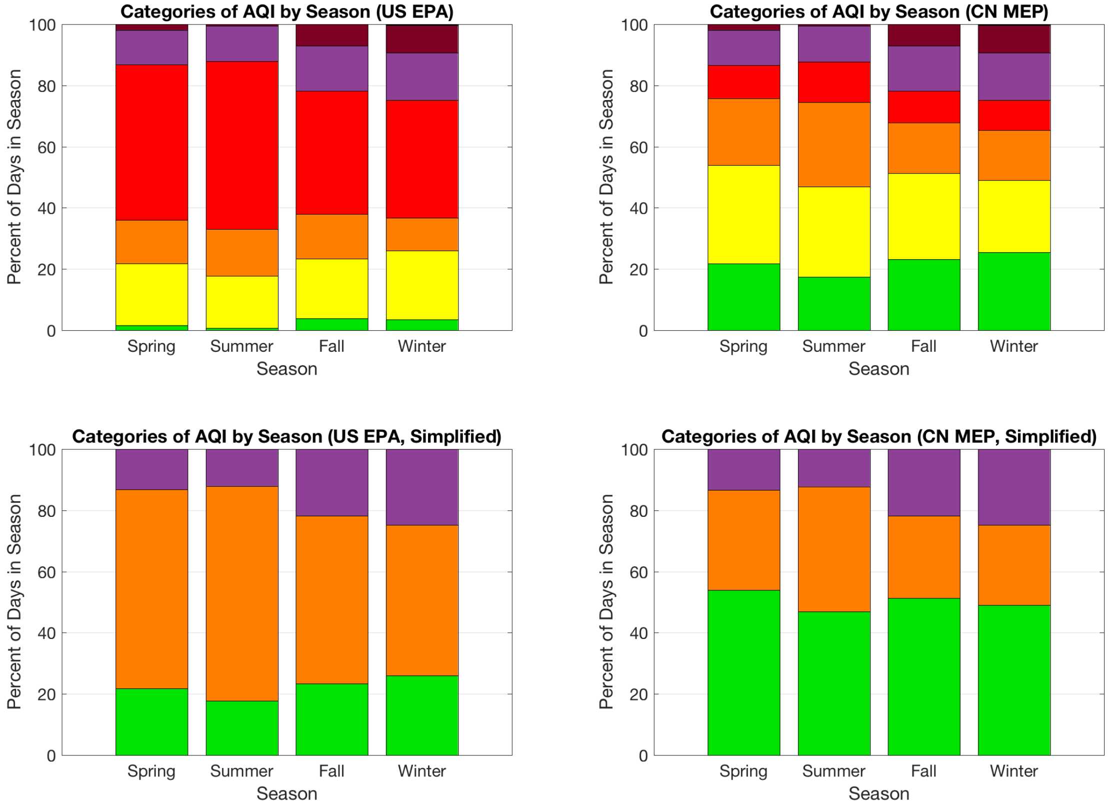


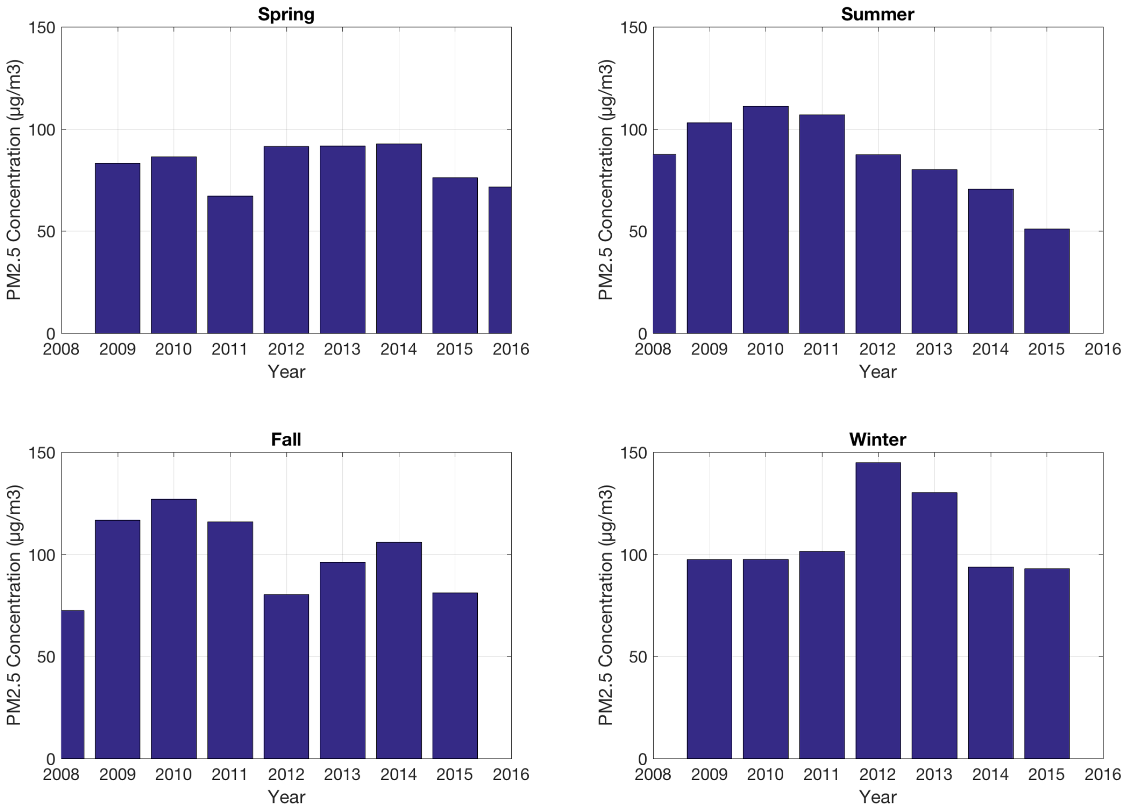
Table 3.2.15 Seasonal and Yearly Data***								
Year	Spring	YoY % Change	Summer	YoY % Change	Fall	YoY % Change	Winter**	YoY % Change
2008	-*		87.6		72.6		-*	
2009	83.3		103.2	+17.8%	116.9	+61.0%	97.4	
2010	86.5	+3.8%	111.2	+7.8%	127.1	+8.7%	97.6	+0.2%
2011	67.3	-22.2%	107.0	-3.8%	115.9	-8.9%	101.5	+4.0%
2012	91.6	+36.1%	87.4	-18.3%	80.4	-30.6%	144.9	+42.8%
2013	91.8	+0.2%	80.2	-8.2%	96.2	+19.7%	130.4	-10.0%
2014	92.7	+1.0%	70.6	-12.0%	106.0	+10.2%	93.8	-28.1%
2015	76.3	-17.7%	51.2	-27.5%	81.2	-23.4%	93.0	-0.9%
2016	71.7	-6.0%	-*		-*		-*	
Average	84.2		86.5		99.8		107.7	

*Insufficient valid data

** Winter starts 12/1 of the respective year and continues to 2/28 or 2/29 of the next. Other seasons exist solely in the respective year (spring: 3/1-5/31, summer: 6/1-8/31, fall: 9/1-11/30).

***All data are average PM2.5 concentration for that season in μg/m3. Data are shaded according to its value, with red being largest and green being smallest.

Figure 3.2.16



It should be noted, however, that this analysis may not be fully representative of the data because of a concurrent trend in PM2.5 concentration throughout the nine years.

In analyzing seasonal patterns, one may also discover trends in air quality over the years with regards to a specific season. Shown in Table 3.2.15, 2011 had the best spring in terms of daily PM2.5 average (67.3), 2015 had the best summer (51.2), 2008 had the best fall (72.6), and 2015 had the best winter (93.0). A trend begins to emerge, similar to previous analyses, that the PM2.5 concentration increased until 2013, in which it then decreased. The data in Table 3.2.15 are visually shown in Figure 3.2.16, and the clearest trend is data from the summer: average PM2.5 concentrations have been decreasing every year since 2010 (Average has decreased 54% since 2010). When summer 2016 data becomes available, such a trend will be put to the test.

It should also be noted that despite large improvements in air quality in 2015 for the spring, summer, and fall seasons (-17.7%, -27.5%, -23.4% respectively), the spring of 2011 and the fall of 2008 had better air quality than the respective seasons of 2015. The reason for such good quality in the spring of 2011 and the fall of 2008 is unknown, but the massive cleanup efforts for the Beijing 2008 Olympics may be a reason for good air quality in the fall of 2008.

Table 3.2.17 Basic Monthly Statistical Analysis

Month	Min	25th	Median (Rank)	75th	Max	Mean (Rank)	IQR	STD	Mean % of Avg
Jan	10.3	33.9	72.4 (5)	150.9	568.6	107.3 (9)	117.0	101.5	114.2
Feb	7.1	35.0	79.8 (8)	146.4	492.8	107.4 (10)	111.4	95.9	114.3
Mar	7.5	29.8	70.1 (3)	127.8	351.3	93.6 (7)	98.0	76.8	99.6
Apr	6.1	39.5	73.4 (6)	116.5	249.4	83.2 (4)	77.0	52.4	88.5
May	12.5	45.3	67.2 (1)	103.9	214.3	76.8 (1)	58.7	42.4	81.8
Jun	6.6	49.7	82.6 (12)	127.2	321.1	91.9 (6)	77.4	55.2	97.8
Jul	8.4	46.1	82.5 (11)	121.6	292.4	90.1 (5)	75.5	54.1	95.8
Aug	10.8	39.1	71.7 (4)	105.9	243.0	77.3 (2)	66.8	46.9	82.2
Sep	2.9	35.7	69.2 (2)	110.3	301.8	79.9 (3)	74.6	57.2	85.0
Oct	5.8	34.8	73.4 (6)	156.5	378.0	105.8 (8)	121.8	90.9	112.5
Nov	6.5	43.9	81.3 (10)	165.0	482.3	114.6 (12)	121.1	94.3	121.9
Dec	6.1	34.3	79.8 (8)	160.0	537.3	108.4 (11)	125.7	95.9	115.4
All Days	2.9	39.6	74.2	123.8	568.6	94.0	84.3	75.0	100%

MONTHLY PATTERNS AND TRENDS

An analysis of monthly patterns and trends is essentially a similar but more detailed analysis of seasons. As such, a similar pattern is seen in the monthly data as in seasonal data. The conclusion that fall and winter months have both more satisfactory days and more heavily polluted days than the spring and summer is echoed in a monthly analysis. Table 3.2.17 shows basic statistics with PM2.5 concentrations grouped by month. It should be noted that the spring, summer, and early fall months have much smaller IQRs and STDs and much lower maximums than other months. May has the lowest median and average, whereas June is 6th by average but 12th by median. Such an observation may be indicative of the less frequent heavily polluted days in June, but also of the absence of satisfactory days.

Table 3.2.18a and b show this pattern in more detail. It is clear that although May had the best average and median, there were 0 days throughout the nine months of May (261 days) analyzed with a PM2.5 average concentration of "Good" by US EPA standards (bin 1, PM2.5 0-12, AQI 0-50). May additionally only had 16.5% of days in the "Excellent" category by CN EPA breakpoints (bin 1, PM2.5 0-35, AQI 0-50), the subjective standard of "blue skies" days. This is 5.4 percentage points (24.7%) lower than the average. However, May also only had 6.5% of days with PM2.5 concentration exceeding 150 (AQI >200), 11.2 percentage points (63.2%) lower than the average, and 22 percentage points (77%) lower than the high of 28.5% (November). Instead, May ranked second highest in percentage of days in US EPA bin 4 ("Unhealthy," PM2.5 55.5-150.4, AQI 150-200), with 56.3% in this category, 9.9 percentage points (21.3%) higher than the average of 46.4% and 21.8 percentage points (37.8%) higher than the low (October). January, however, had 26% of days in MEP bin 1, ranked highest of all months, but also 25% of days in bins 5, 6, and 7 (PM2.5 >150.4). As such, 51% of days in January are at the two extremes, whereas in May, only 23% are at these extremes (16.5% at low end (PM2.5 <35) and 6.5% at high (PM2.5>150).

Months in the late fall, winter, and early spring have the highest frequency of days with PM2.5 above 150, but such months also have the highest frequency of days with PM2.5 below 35 (subjective "blue skies" standard). As such, months in the late spring and summer have more days concentrated in the PM2.5 range of 35 to 150 (US EPA: 100-200, CN MEP: 50-200), with less variance in air quality. On the other hand, months in the late fall, winter, and early spring have a higher average and a wider variation of air quality. This pattern is clearly shown in Figure 3.2.19, where the portion of red (bin 4) in Figure 3.2.19a is largest in the summer and smallest in the winter, whereas days in green and yellow are more frequent in the winter. An unexpected spike of days in bin 5 occurs in June and July, however. It is thus clear that a

Table 3.2.18a Monthly Histogram Data for US EPA Breakpoints

Year	Tot. # val. d	Bin 1 #d	Bin 1 %d	Bin 2 #d	Bin 2 %d	Bin 3 #d	Bin 3 %d	Bin 4 #d	Bin 4 %d	Bin 5 #d	Bin 5 %d	Bin 6 #d	Bin 6 %d	Bin 7 #d	Bin 7 %d
Jan	204	5	2.5	49	24.0	25	12.3	74	36.3	33	16.2	17	8.3	1	0.5
Feb	206	8	3.9	45	21.8	19	9.2	88	42.7	26	12.6	20	9.7	0	0.0
Mar	236	9	3.8	55	23.3	29	12.3	91	38.6	38	16.1	14	5.9	0	0.0
Apr	243	3	1.2	51	21.0	22	9.1	138	56.8	29	11.9	0	0.0	0	0.0
May	261	0	0.0	43	16.5	54	20.7	147	56.3	17	6.5	0	0.0	0	0.0
Jun	221	2	0.9	34	15.4	29	13.1	124	56.1	29	13.1	3	1.4	0	0.0
Jul	229	1	0.4	37	16.2	34	14.8	122	53.3	34	14.8	1	0.4	0	0.0
Aug	220	2	0.9	43	19.5	39	17.7	122	55.5	14	6.4	0	0.0	0	0.0
Sep	225	9	4.0	46	20.4	37	16.4	108	48.0	22	9.8	3	1.3	0	0.0
Oct	237	12	5.1	48	20.3	32	13.5	83	35.0	40	16.9	22	9.3	0	0.0
Nov	207	5	2.4	36	17.4	29	14.0	78	37.7	37	17.9	22	10.6	0	0.0
Dec	198	8	4.0	43	21.7	21	10.6	72	36.4	35	17.7	18	9.1	1	0.5
Total	2687	64	2.4	530	19.7	370	13.8	1247	46.4	354	13.2	120	4.5	2	0.1

simple monthly average is not representative of monthly patterns.

As in seasonal analysis, monthly analysis also provides insight on yearly trends. As such, it is pertinent to further compare and analyze yearly improvements through the observation of PM2.5 data for every specific month.

Table 3.2.18b Monthly Histogram Data for CN MEP Breakpoints

Year	Tot. # val. d	Bin 1 #d	Bin 1 %d	Bin 2 #d	Bin 2 %d	Bin 3 #d	Bin 3 %d	Bin 4 #d	Bin 4 %d	Bin 5 #d	Bin 5 %d	Bin 6 #d	Bin 6 %d	Bin 7 #d	Bin 7 %d
Jan	204	53	26.0	54	26.5	26	12.7	20	9.8	33	16.2	17	8.3	1	0.5
Feb	206	51	24.8	45	21.8	39	18.9	25	12.1	26	12.6	20	9.7	0	0.0
Mar	236	64	27.1	61	25.8	42	17.8	17	7.2	38	16.1	14	5.9	0	0.0
Apr	243	54	22.2	72	29.6	55	22.6	32	13.2	30	12.3	0	0.0	0	0.0
May	261	43	16.5	105	40.2	65	24.9	31	11.9	17	6.5	0	0.0	0	0.0
Jun	221	35	15.8	58	26.2	62	28.1	34	15.4	29	13.1	3	1.4	0	0.0
Jul	229	37	16.2	71	31.0	57	24.9	29	12.7	34	14.8	1	0.4	0	0.0
Aug	220	45	20.5	68	30.9	66	30.0	26	11.8	15	6.8	0	0.0	0	0.0
Sep	225	55	24.4	69	30.7	48	21.3	28	12.4	22	9.8	3	1.3	0	0.0
Oct	237	59	24.9	61	25.7	38	16.0	17	7.2	40	16.9	22	9.3	0	0.0
Nov	207	41	19.8	58	28.0	25	12.1	24	11.6	37	17.9	22	10.6	0	0.0
Dec	198	51	25.8	44	22.2	34	17.2	15	7.6	35	17.7	18	9.1	1	0.5
Total	2687	588	21.9	766	28.5	557	20.7	298	11.1	356	13.2	120	4.5	2	0.1

Figure 3.2.19

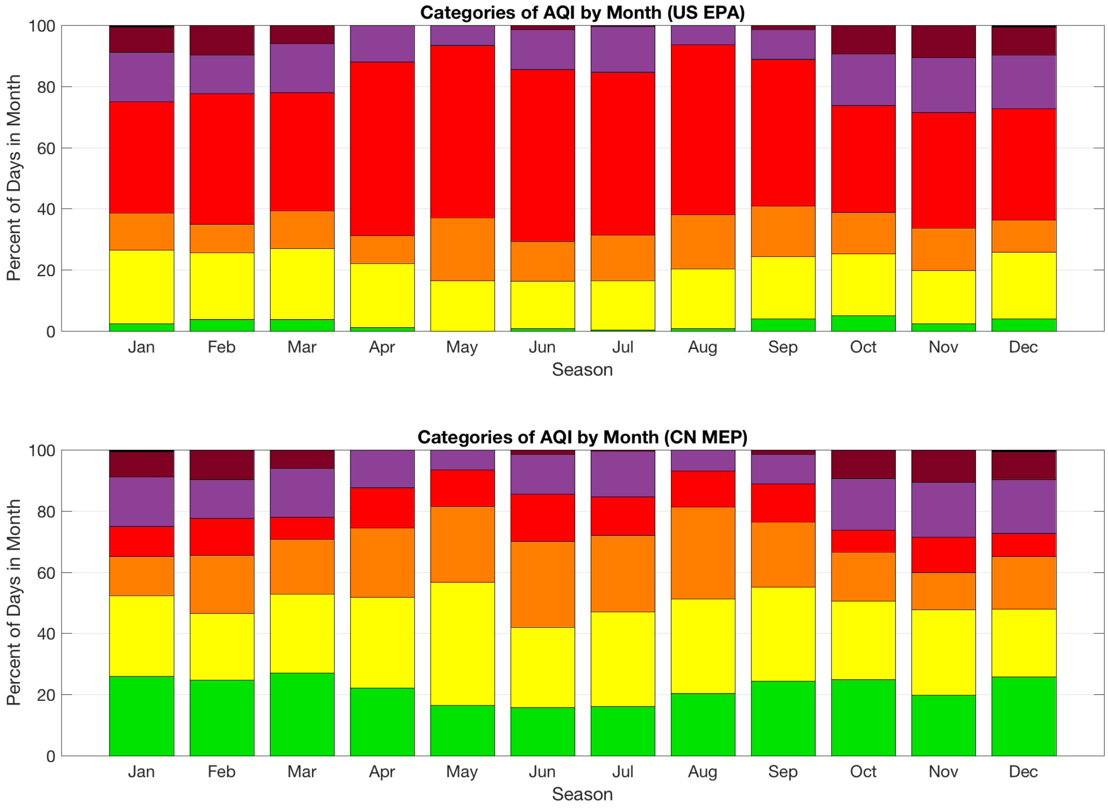


Table 3.9.20 is such a list, detailing the PM2.5 average concentration of every month, and how such a month compares against the same month the previous year. Table 3.9.21 ranks each year by the respective month, with each individual cell color-coded to match its year. Although, like mentioned earlier, monthly averages may not be the most representative measure of each month's air quality as a whole, this graph clearly shows that most months of 2016 so far, as well as the mid-late months of 2015, are among the months with the best air quality. Specifically, June 2015 to October 2015 had the best air quality with regards to the respective month of the seven or eight years in which data is valid. Although January-April, November, and December of 2015 were substandard, an apparent trend is already forming. From Table 3.9.20, it is clear that 9 of the past 12 months reported PM2.5 averages below the same month in the previous year, and 10 out of the 12 months were below the nine-year average for that month. November and December 2015 were anomalies, despite the numerous "red alerts" for pollution issued by the government. Additionally, although March 2016 seemed to also be an anomaly, the other months of 2016 so far are ranked either 1st or 2nd in their respective month categories.

From Table 3.9.22, it can be observed that six of the past 12 months have PM2.5 averages that place them in the best 7 months (86 valid months total). However, only these 7 months are below the 55.4 US EPA threshold, with averages below this threshold representing days that are "Unhealthy for Sensitive Groups." 74 of the 79 other months have averages in US EPA bin 4, "Unhealthy," and the remaining 5 are in US EPA bin 5, "Very Unhealthy." (BY CN MEP standards, 23 months were in bin 2, "Good," 49 months were in bin 3, "Slightly Polluted," 8 months were in bin 4, "Moderately Polluted," and 6 months were in bin 5, "Heavily Polluted." Such statistics, at least according to US EPA standards, are a reminder that despite improvements in air quality over the past year, Beijing's air is still chronically unsafe.

Table 3.9.20 Monthly PM2.5 Average Concentrations and YoY Changes

	Jan	Jan %YoY	Feb	Feb % YoY	Mar	Mar % YoY	Apr	Apr % YoY	May	May % YoY	Jun	Jun %YoY
2008	-*	-	-*	-	-*	-	-*	-	98.3	-	99.8	-
2009	-*	-	-*	-	78.7	-	89.3	-	-*	-	96.9	-2.9%
2010	89.8	-	97.2	-	92.7	17.8%	79.9	-10.5%	87.0	-	-*	-
2011	43.6	-51.4%	150.3	54.6%	56.1	-39.5%	-*	-	63.8	-26.7%	108.2	-
2012	115.2	164.1%	80.6	-46.4%	96.3	71.6%	87.8	-	90.4	41.7%	99.4	-8.1%
2013	194.3	68.7%	123.6	53.4%	123.6	28.3%	65.8	-25.0%	84.7	-6.3%	111.4	12.1%
2014	118.3	-39.1%	174.8	41.4%	110.4	-10.7%	95.5	45.1%	72.3	-14.7%	59.6	-46.5%
2015	108.0	-8.7%	96.8	-44.6%	89.9	-18.6%	80.1	-16.1%	60.0	-17.0%	54.0	-9.5%
2016	72.1	-33.2%	44.2	-54.3%	93.2	3.7%	66.5	-17.0%	55.3	-7.8%	-*	-
Avg	107.3		107.4		93.6		83.2		76.8		91.9	

*Insufficient valid data (less than 75% of days in month had valid daily averages).

	Jul	Jul % YoY	Aug	Aug % YoY	Sep	Sep % YoY	Oct	Oct % YoY	Nov	Nov% YoY	Dec	Dec % YoY
2008	-*	-	69.2	-	57.7	-	84.3	-	-*	-	-*	-
2009	105.3	-	106.1	53.2%	108.4	87.9%	88.0	4.4%	156.2	-	106.0	-
2010	123.4	17.2%	98.2	-7.4%	-*	-	118.6	34.7%	139.5	-10.7%	97.0	-8.5%
2011	107.2	-13.2%	-*	-	95.1	-	150.7	27.0%	110.0	-21.2%	108.6	12.0%
2012	80.3	-25.1%	81.3	-	59.6	-37.4%	94.9	-37.0%	86.3	-21.5%	107.4	-1.1%
2013	68.0	-15.3%	61.9	-23.9%	90.9	52.6%	106.5	12.2%	90.8	5.2%	100.4	-6.6%
2014	88.9	30.8%	62.2	0.5%	70.2	-22.8%	140.8	32.2%	105.9	16.6%	75.2	-25.1%
2015	55.0	-38.1%	44.6	-28.2%	47.1	-32.9%	72.1	-48.8%	124.8	17.8%	161.6	115.0%
2016	-*	-	-*	-	-*	-	-*	-	-*	-	-*	-
Avg	90.1		77.3		79.9		105.8		114.6		108.4	

*Insufficient valid data (less than 75% of days in month had valid daily averages).

Table 3.9.21 Sorted Months by PM2.5 Concentration

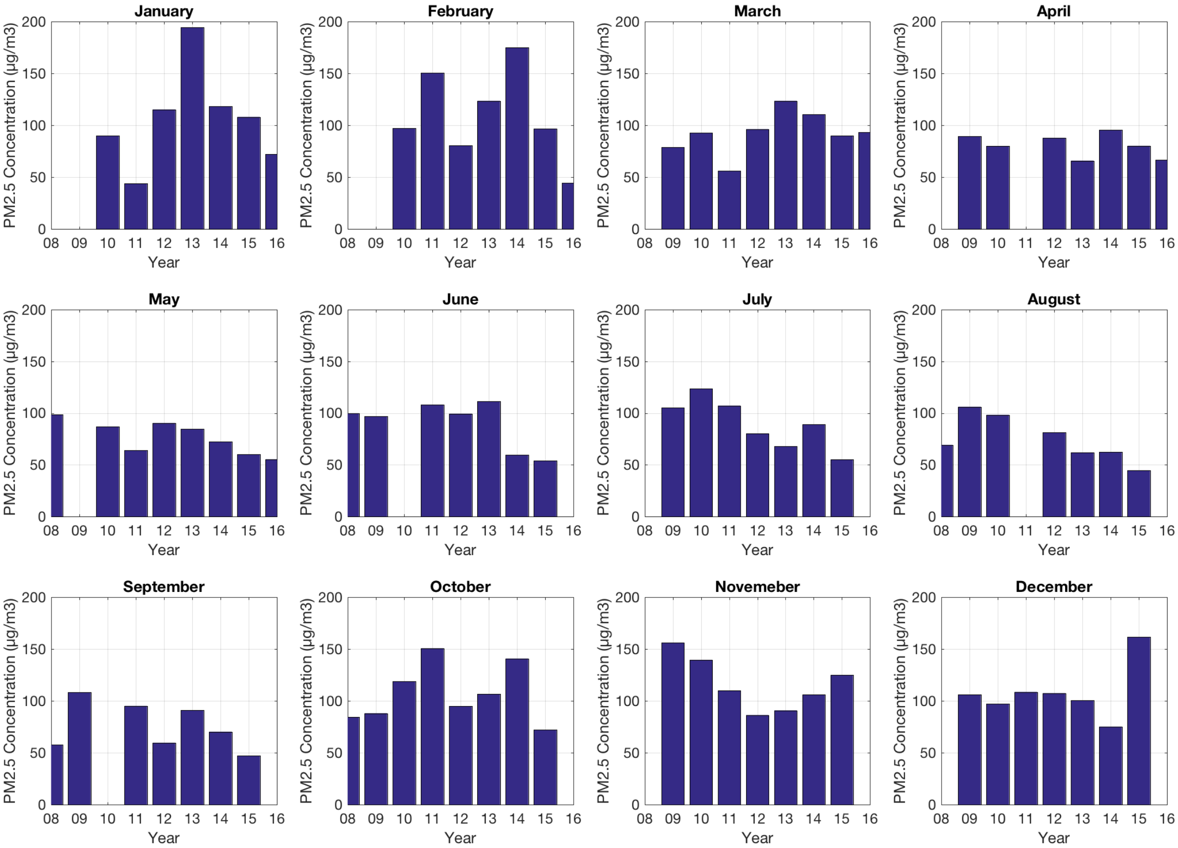
	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
1	2011	2016	2011	2013	2016	2015	2015	2015	2015	2015	2012	2014
2	2016	2012	2009	2016	2015	2014	2013	2013	2008	2008	2013	2010
3	2010	2015	2015	2010	2011	2009	2012	2014	2012	2009	2014	2013
4	2015	2010	2010	2015	2014	2012	2014	2008	2014	2012	2011	2009
5	2012	2013	2016	2012	2013	2008	2009	2012	2013	2013	2015	2012
6	2014	2011	2012	2009	2010	2011	2011	2010	2011	2010	2010	2011
7	2013	2014	2014	2014	2012	2013	2010	2009	2009	2014	2009	2015
8	*	*	2013	*	2008	*	*	*	*	2011	*	*

Note: Colors do not have significance beyond comparison reasons. There is no relation to bin colors.

Table 3.9.22 Best 10 Months and Worst 10 Months (Average PM2.5 Concentration)

	Best 10 Months			Worst 10 Months		
	Year	Month	PM2.5	Year	Month	PM2.5
1	2011	1	43.6	2013	1	194.3
2	2016	2	44.2	2014	2	174.8
3	2015	8	44.6	2015	12	161.6
4	2015	9	47.1	2009	11	156.2
5	2015	6	54.0	2011	10	150.7
6	2015	7	55.0	2011	2	150.3
7	2016	5	55.3	2014	10	140.8
8	2011	3	56.1	2010	11	139.5
9	2008	9	57.7	2015	11	124.8
10	2012	9	59.6	2013	2	123.6

Figure 3.2.23



DAILY PATTERNS AND TRENDS

Figure 3.2.25

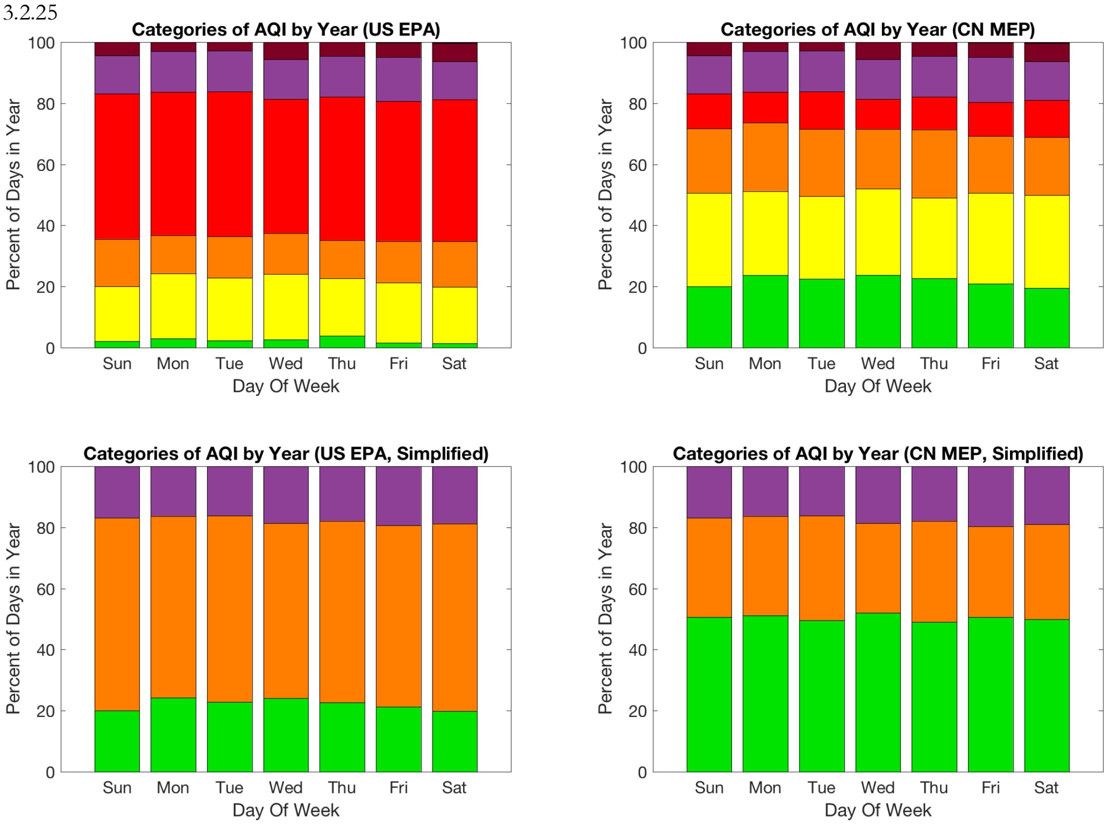
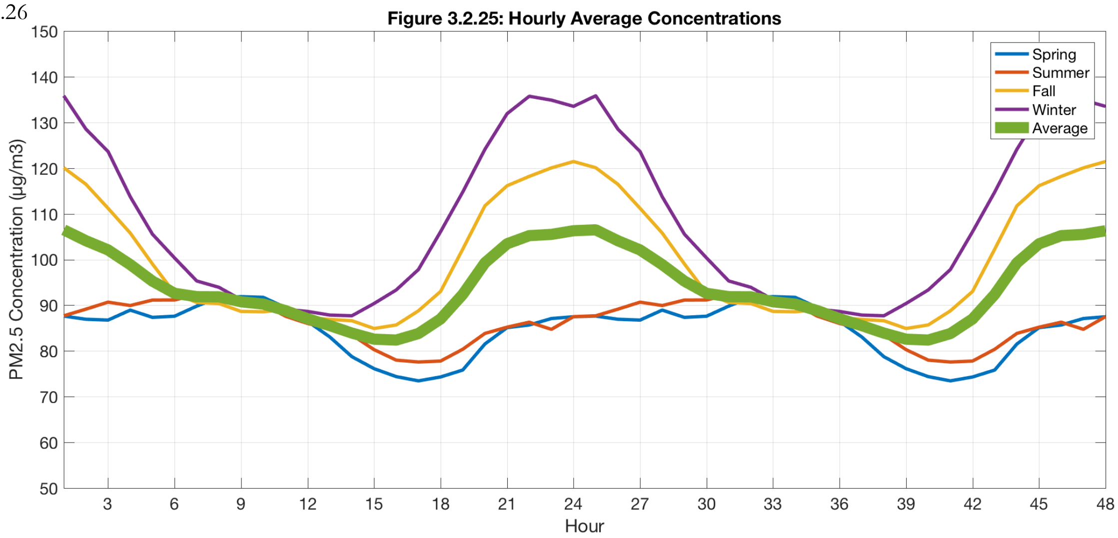


Table 3.9.24 Basic Statistics By Day of Week

Day of Week	Min	25th	Median (Rank)	75th	Max	Mean (Rank)	IQR	STD	Mean % of Avg
Sun	6.8	43.3	73.9 (3)	119.8	416.3	92.8 (3)	76.5	71.1	98.7
Mon	7.1	36.1	71.3 (1)	122.2	492.8	90.1 (1)	86.1	71.1	95.9
Tues	6.1	38.3	75.6 (5)	124.8	464.4	92.4 (2)	86.6	73.7	98.4
Wed	6.5	36.8	71.4 (2)	122.7	417.5	94.2 (5)	86	77.5	100.2
Thu	5.8	37	75.7 (6)	123	448.4	93.4 (4)	86	75.1	99.4
Fri	2.9	41.7	74.5 (4)	128.2	537.3	96.4 (6)	86.5	76.2	102.6
Sat	6.1	41.9	76.7 (7)	137.9	568.6	98.5 (7)	95.9	80	104.8
All Days	2.9	39.6	74.2	123.8	568.6	94	84.3	75	100%

analyzing PM2.5 concentration averages and distributions for each day of the week, no distinct pattern results. Table 3.9.24 shows that every median is within 2.8 points (3.8%) of the total median, and every average is within 4.5 points (4.8%) of the total average. Figure 3.9.25 further shows that each day of the week exhibits similar distributions in each scale. There does seem to be a slight trend that latter days of the week (most notably Friday and Saturday) exhibit worse air quality (medians are among the highest and means are the two highest), but this analysis is tentative at best and any pattern that may be observed from this data may simply be the result of anomalies.

Figure 3.2.26



HOURLY PATTERNS AND TRENDS

Not only do the data show yearly, seasonal and monthly patterns and trends, specifically pertinent is the pattern of average PM2.5 concentrations by the hour. As shown in Figure 3.2.26 (48 hours plotted), on average, PM2.5 hourly concentrations are lowest in the afternoon (lowest average: 82.4 at 3:00PM), and highest at midnight (highest average: 106.5 at 12:00AM). This large, 24.1 point difference, as shown in the graph, is larger in the winter and fall and smaller in the spring and summer. Specifically, winter PM2.5 average concentrations are highest at 12:00AM (135.8), and lowest at 1PM (87.7), almost a 50 point difference. Such a large difference may be the result of the use of coal for heating in many homes during the night, in addition to the allowance of large freight trucks into the city during the evening hours (in contrast to daytime hours). Weather may also be a factor.

On the other hand, summer PM2.5 concentrations show the least discrepancies between hours, with a low of 77.6 at 4PM and a high of 92.3 at 6AM, a difference of only 14.7 µg.m3. However, despite differences between seasons, averages seem to converge in the 9AM-12PM period (four hour averages: spring: 87.7, summer: 87.2, fall: 88.0, winter: 89.0). These averages, none more than 2 µg/m3 away from each other, contrast the 23.5 µg/m3 range between entire seasonal averages. The reason for this convergence of concentrations is unclear.

It is also unclear if the increase of cars during morning and evening rush hours contribute to such PM2.5 trends. On average, PM2.5 concentrations do not increase through the morning rush hour, but concentrations increase dramatically (especially in the fall and winter) following the evening rush hour. This increase may or may not be an effect of rush hour traffic.

SPECIAL INTERVALS

In the past nine years, numerous periods of time have attracted media attention due to Beijing's air quality. In January 2013, Beijing's "airpocalypse" was subject to international concern, while Beijing's efforts to clean up the air in preparation for the 2008 Olympic games, the 2014 Asia-Pacific Economic Conference, and the 2015 Victory Day parade also attracted media attention. These four periods are briefly analyzed in this study.

Specifically these four periods were as follows:

- 2008 Olympic Games (8/8/2008 to 8/24/2008)
- January 2013 "Airpocalypse" (1/1/2013 to 1/31/2013)
- "APEC blue" (11/1/2014 to 11/14/2014)
- "Parade blue" preparations and parade (8/20/2015 to 9/3/2015).

Table 3.2.27 is a table detailing analysis of these four periods, and Figure 3.2.28 are bar graphs, displaying data for each day in each period.

Table 3.2.27a Select Special Interval Statistics, US EPA Bins

	# Valid Days	Mean PM2.5	Median PM2.5	% in Bin 1	% in Bin 2	% in Bin 3	% in Bin 4	% in Bin 5	% in Bin 6	% in Bin 7	Avg. in sm. per.*
Olym.	17	58.6	45.1	0	11.8	41.2	29.4	0	0	0	79.4
Jan-13	31	194.3	147.8	0	12.9	0	38.7	16.3	29.0	3.2	91.8
APEC	14	49.9	44.9	14.3	21.4	28.6	35.7	0	0	0	102.1
Parade	15	18	15.4	20.0	80.0	0	0	0	0	0	83.2

*average PM2.5 concentration in same period in other years.

Table 3.2.27b Select Special Interval Statistics, CN MEP Bins

	# Valid Days	Mean PM2.5	Median PM2.5	% in Bin 1	% in Bin 2	% in Bin 3	% in Bin 4	% in Bin 5	% in Bin 6	% in Bin 7	Avg. in sm. per.*
Olym.	17	58.6	45.1	11.8	47.1	17.6	5.9	0	0	0	79.4
Jan-13	31	194.3	147.8	12.9	9.7	16.1	12.9	16.3	29.0	3.2	91.8
APEC	14	49.9	44.9	35.7	42.9	14.3	7.1	0	0	0	102.1
Parade	15	18	15.4	100.0	0	0	0	0	0	0	83.2

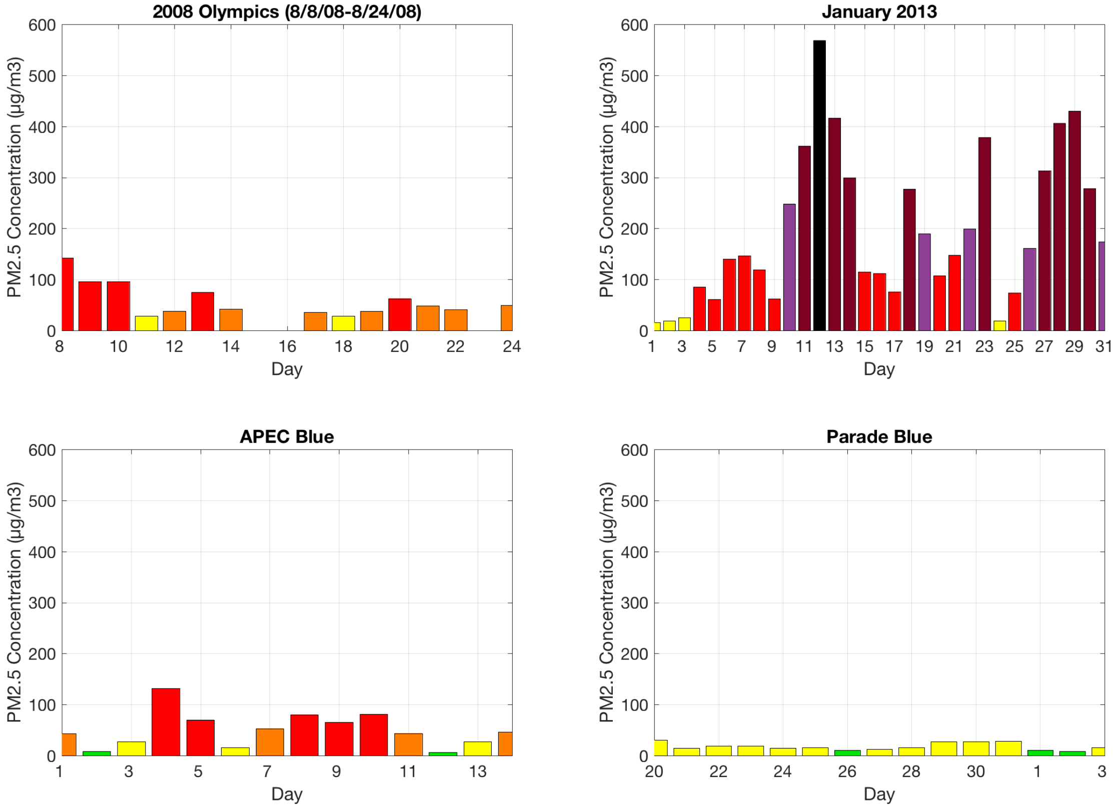
*average PM2.5 concentration in same period in other years.

It is clear that the January 2013 “airpocalypse” contained extremely polluted days, and it is thus unsurprising that January 2013 was the month with the highest PM2.5 average concentration (194.3). With 48.5% of days beyond bins 1-4 (PM2.5 >150/150.4, AQI>200, Categories: US EPA: “Very Unhealthy,” “Hazardous,” or beyond index; CN MEP: “Heavily Polluted,” “Severely Polluted”, beyond index). This is 212% of the January average excluding 2013. On the other hand, none of the days in the Olympics period, “APEC blue,” or “parade blue” had PM2.5 concentration averages in these bins.

Additionally, the average of days within the preparation for the 2015 Victory Day Parade was only 18 µg/m3, with 100% days occurring in US EPA bins 1 or 2 and CN MEP bin 1. The maximum concentration in this period was only 28.2 µg/m3, compared with the maximum of 568.6 µg/m3 concentration achieved in the 2013 January period (568.8 was also the highest concentration of any day in the entire nine years analyzed). The “parade blue” period was also the longest consecutive sequence of days, in all days analyzed, in which the PM2.5 concentration was below 35 µg/m3 (US EPA: AQI 100; CN MEP: AQI 50). The average of 18 µg/m3 is 65.2 µg/m3 (78%) lower than the average concentration in the same period in other years, and meets the WHO AQS of 25 µg/m3. Note that a subset of this period, the days between 8/21/2015 and 8/28/2015, inclusive, was the longest period of consecutive days in which the air quality did not exceed the WHO AQS.

However, it is worthy to note that the Olympics period and the APEC 2014 period, with mean concentrations of 58.6 and 49.9 µg/m3, respectively, were not among the best periods among all days analyzed. A 58.6 µg/m3 concentration is considered “Unhealthy” with US EPA standards, and “Good” by CN MEP standards. Similarly, a 49.9 µg/m3 concentration is considered “Unhealthy for Sensitive Groups” with US EPA standards, and “Good” by CN MEP standards. Even if these periods were entire months (as opposed to 17 and 14 days, respectively), they would only be ranked 10th and 5th of all months surveyed. However, these periods maintained average concentrations 26.2% (34.3 µg/m3) and 51.1% (57.2 µg/m3) below the averages of 79.4 and 83.2 µg/m3 in the same period of other years, and is a large deviation from the average of all days in the 2687 days surveyed of 94.0 µg/m3.

Figure 3.2.28



Possible reasons for an increased percentage of days with air quality meeting the standard are the Chinese government's extensive efforts to temporarily mitigate the effects of large polluting sources. Half of private cars are prohibited from driving in the urban area of the city every day, dozens of factories are brought to a close, and people are urged not to stage barbeques or set of firecrackers. Numerous other causes of air pollution, such as geography and weather situations in the period, were not considered in this study but also may have an effect on air quality.

4 CONCLUSION

Hourly PM2.5 concentrations from 2687 days across nine years, reported from the U.S. Embassy in Chaoyang District, Beijing, were analyzed in this study. Multiple conclusions may be drawn pertaining to the ultimate objective of discovering trends and patterns in the data and providing reference for the Chinese government, as specified in the introduction.

It is first important to note that air quality in Beijing is unhealthy, irrespective of any analysis conducted in this study. An average of 94.0 μg/m3 across the 2687 days with valid days is “Unhealthy” by US standards, “Slightly Polluted” by Chinese standards, and 276% the World Health Organization Air Quality Standard of 25 μg/m3.

Three different probability distributions are known to fit air quality data, namely the Weibull, gamma, and lognormal distributions. Results from the optimal selection through *p*-value selection indicated that each of the three distributions fit at least one year-long dataset. Lognormal distributions fit datasets from more recent years best, and all of the three distributions utilized provided sufficient evidence to reject the null hypothesis when analyzing data from the entire nine-year period. This may be an indication of trends developing throughout the years within the data that causes deviations from each of the three distributions analyzed. Such a primitive sign of a trend was further analyzed in the study.

In analyzing a trend throughout the years, it was discovered that an increasing percentage of days in recent years have averaged at more satisfactory levels of air quality. Results show that the first five months of 2016 reported a 23.8% improvement in average PM2.5 daily concentration over the average of the first five months of all years with available data, with 38.8% days below the Chinese AQI standard, compared to the mean of 20.3%, and 7.9% days below the US AQI standard, compared to the mean of 2.4%. Six of the 7 months with the lowest average PM2.5 concentrations occurred in the last 12 months of available data (June 2015-May 2016), and 7 of the last 12 months with available data reported average PM2.5 concentrations that were the lowest averages for that respective month in any of the years analyzed. Reasons for such an improvement may perhaps be attributed to the Chinese government’s massive efforts on curtailing pollution, from road rationing to better regulation in vehicle emissions and factory waste. However, as mentioned earlier, although these improvements are clear, according to the WHO, the air quality remains at an unhealthy level.

Both seasonal and monthly patterns are evident in the data analyzed, with fall and winter months reporting higher average PM2.5 concentrations, but with medians similar to the spring and summer months. A further inspection reveals a greater percentage of days in both excellent and polluted categories in the winter and fall months, with most of the extremely polluted days overall occurring in these months. The widespread use of coal for heating near the winter months may be a contributing cause for such days with extremely substandard air quality.

While each day of the week reported similar distributions and averages of PM2.5 concentrations, hourly differences in average PM2.5 concentrations are clear. Average concentrations in the early hours of a day (esp. 12am) were much higher than average concentrations in the afternoon with a 34.1 μg/m3 range; this pattern is exacerbated in the winter, with a range of over 50 μg/m3. It is possible that the hours of the night may report higher PM2.5 concentrations because trucks are prohibited from the urban areas of Beijing in daytime hours, in addition to increased use of coal in the winter for heating in the evening and extremely early morning hours, when residents are home.

Lastly, it was found that numerous temporary restrictions implemented by the Chinese government preluding and during important events were effective to a large extent. The 2008 Beijing Olympics, the 2014 Asia Pacific Economic Conference, and the day of and days leading up to the 2015 Victory Day Parade all reported PM2.5 concentrations significantly lower than the average of the same period in other years.

It is important to note that there were several limitations that inhibited all sections of this study. First, PM2.5 is not the sole pollutant that affects air quality, and although frequently described as most prominent in cities such as Beijing, it is not indicative of the entire air quality situation. Second, geographical and meteorological conditions were not considered in this study, with various factors such as wind speed, wind direction, and rainfall known to affect air quality as well. Third, data was recorded at only one monitoring site in Beijing (US Embassy in Chaoyang District), and thus, 1) its readings may not be indicative of the air quality of Beijing as a whole, and 2) its readings may include inaccuracies due to measurement errors that can carry through to this analysis. Thus, data by the Chinese government, which has dozens of monitoring stations throughout Beijing, may be more accurate. However, historical data from these sources are not readily accessible to the public.

For the section on probability distributions, the choice of the number of bins for the chi-squared goodness of fit test remains a nontrivial task that may be a limitation, for there is no guarantee of the optimal choice of bins and degrees of freedom with any formula. More bins/degrees of freedom presents more power in the chi-squared goodness of fit test calculation, but with too many bins inaccuracies start to materialize as samples in each bin is reduced. In addition, differences between the distributions used may not be significant enough to declare one as optimal, and distributions other than Weibull, lognormal, and gamma may be more representative of the data.

In conclusion, probability distributions were successfully fit on multiple year-long datasets, and statistically significant trends and patterns were discovered in the data. It is also evident that, although improvements in air quality, possibly stemming from to the government's extensive pollution control measures in recent years, exist, more needs to done to create a healthy city to live in.

5 REFERENCES

1. U.S. EPA. "Clean Air Act Overview: Evolution of the Clean Air Act." U.S. Environmental Protection Agency. Accessed July 22, 2016. https://www.epa.gov/clean-air-act-overview/evolution-clean-air-act.

2. U.S. EPA. "Particulate Matter (PM) Pollution: Particulate Matter (PM) Basics." U.S. Environmental Protection Agency. Accessed July 22, 2016. https://www.epa.gov/pm-pollution/particulate-matter-pm-basics#PM.

3. U.S. EPA. "Particulate Matter (PM) Pollution: Health and Environmental Effects of Particulate Matter (PM)." U.S. Environmental Protection Agency. Accessed July 22, 2016. https://www.epa.gov/pm-pollution/health-and-environmental-effects-particulate-matter-pm.

4. U.S. Environmental Protection Agency Office of Air Quality Planning and Standards. *Technical Assistance Document for the Reporting of Daily Air Quality -- the Air Quality Index*. By United States Environmental Protection Agency. Report no. EPA-454/B-13-001. December 2013. Accessed July 22, 2016. https://www3.epa.gov/airnow/aqi-technical-assistance-document-dec2013.pdf.

5. Ministry of Environmental Protection. "Mission." Ministry of Environmental Protection of the People's Republic of China. Accessed July 22, 2016. http://english.sepa.gov.cn/About_SEPA/Mission/.

6. Ministry of Environmental Protection. "Mandates." Ministry of Environmental Protection of the People's Republic of China. Accessed July 22, 2016. http://english.sepa.gov.cn/About_SEPA/Mandates/.

7. Ministry of Environmental Protection of the People's Republic of China People's Republic of China National Environmental Protection Standards. *Technical Regulation on Ambient Air Quality Index (on trial)*. By Ministry of Environmental Protection. Report no. HJ633-2012. February 29, 2012. Accessed July 22, 2016. http://kjs.mep.gov.cn/hjbhbz/bzwb/dqhjbh/jcgfffbz/201203/W020120410332725219541.pdf.

8. WHO Media Center. "Ambient (Outdoor) Air Quality and Health." World Health Organization. Last modified March 2014. Accessed July 22, 2016. http://www.who.int/mediacentre/factsheets/fs313/en/.

9. Zhang, Jim. "What is China Doing to Tackle its Air Pollution?" BBC News. Last modified January 20, 2016. Accessed July 22, 2016. http://www.bbc.com/news/world-asia-china-35351597.

10. AP. "China Tells U.S. to Stop Tweets on Beijing's Bad Air." USA Today News. Last modified June 5, 2012. Accessed July 22, 2016. http://usatoday30.usatoday.com/news/world/story/2012-06-05/china-beijing-air-quality-tweets/55403444/1.

11. State Council of the People's Republic of China. *Action Plan for the Prevention and Control of Air Pollution*. Report no. 2013 No. 37. September 10, 2013. Accessed July 22, 2016. http://www.gov.cn/zwgk/2013-09/12/content_2486773.htm.

12. Finamore, Barbara. "China Pledges to Tackle Air Pollution with New Plan." Natural Resources Defense Council. Last modified September 13, 2013. Accessed July 22, 2016. https://www.nrdc.org/experts/barbara-finamore/china-pledges-tackle-air-pollution-new-plan.

13. Ruan, Lotus Yang. "Why China Can't Fix Its Environment." The Diplomat. Last modified March 16, 2015. Accessed July 22, 2016.

http://thediplomat.com/2015/03/why-china-cant-fix-its-environment/.

14. "News Release." Ministry of Environmental Protection of the People's Republic of China. Accessed July 22, 2016. http://english.sepa.gov.cn/News_service/news_release/.

15. Beijing Municipal People's Government. *Beijing Municipal People's Government on the Issuance of Beijing 2013-2017 Clean Air Action Plan Focused on Task Decomposition*. Report no. [2013]49. September 2, 2013. Accessed July 22, 2016. http://zhengwu.beijing.gov.cn/gzdt/gggs/t1322955.htm.

16. Wang, Xi, Ren Jie Chen, Bing Heng Chen, and Hai Dong Kan. *Application of Statistical Distribution of PM10 Concentration in Air Quality Management in 5 Representatives Cities of China*. March 30, 2013. Accessed July 25, 2016. doi:10.3967/0895-3988.2013.08.002.

17. Lu, Hsin-Chung. *Comparisons of Statistical Characteristic of Air Pollutants in Taiwan by Frequency Distribution*. February 22, 2012. Accessed July 25, 2016. DOI:10.1080/10473289.2003.10466194.

18. Morel, Benoit, Sonia Yeh, and Luis Cifuentes. *Statistical Distribution for Air Pollution Applied to the Study of the Particulate Problem in Santiago*. Research report no. 33 (1999) 2575}2585. June 29, 1998. Accessed July 25, 2016. DOI:10.1016/S1352-2310(98)00380-X.

19. U.S. Department of State. "Beijing - Historical Data." Mission China. Accessed July 23, 2016. http://www.stateair.net/web/historical/1/1.html.

20. MathWorks. "Weibull Distribution." MathWorks Documentation. Accessed July 26, 2016. http://www.mathworks.com/help/stats/weibull-distribution.html.

21. MathWorks. "Fitdist: Fit Probability Distribution Object to Data." MathWorks Documentation. Accessed July 26, 2016. http://www.mathworks.com/help/stats/fitdist.html?searchHighlight=fitdist.

22. "4.10. Number of Bins in Distribution Fitting." Palisade. Accessed August 15, 2016. http://kb.palisade.com/index.php?pg=kb.page&id=57.