5-2019

PM2.5 Data Reliability and Air Quality Improvement Trends in Beijing

Huimin Li
The University of Texas Rio Grande Valley

$PM_{2.5}$ DATA RELIABILITY AND AIR QUALITY IMPROVEMENT TRENDS IN BEIJING

A Thesis

by

HUIMIN LI

Submitted to the Graduate School of
The University of Texas Rio Grande Valley
In partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

May 2019

Major Subject: Mathematics

PM$_{2.5}$ DATA RELIABILITY AND AIR QUALITY IMPROVEMENT TRENDS IN BEIJING

A Thesis
by
HUIMIN LI

COMMITTEE MEMBERS

Dr. Xiaohui Wang
Chair of Committee

Dr. Santanu Chakraborty
Committee Member

Dr. Demba Fofana
Committee Member

Dr. Chu-Lin Cheng
Committee Member

May 2019

ABSTRACT

Li, Huimin, $PM_{2.5}$ Data Reliability and Air Quality Improvement Trends in Beijing. Master of Science (MS), May, 2019, 68 pp., 15 tables, 14 figures, 60 references, 37 titles.

$PM_{2.5}$ has been a main environmental concern due to its adverse effects on human health and society. We used data from two sources: monitoring station of the U.S. Embassy in Beijing, and several nearby monitoring stations of the Chinese Ministry of Environmental Protection. This study includes investigating (1) $PM_{2.5}$ historical data reliability, (2) $PM_{2.5}$ real-time data reliability, and (3) air quality improvement trends in Beijing over the past decade. We used graphical methods, descriptive statistics, correlation analysis, and inferential analyses including paired samples t-test, ANOVA, and Kruskal-Wallis test. We reported effect sizes to aid study on practical significance. Inferential procedures' assumptions were checked. Results showed that $PM_{2.5}$ historical and real-time data were highly consistent between two data sources. Air quality improvement has become significant since 2015 in Beijing but the annual average $PM_{2.5}$ concentration in 2016 with 72.9 $\mu g.m^{-3}$ was still over two times the WHO's recommended level 35 $\mu g.m^{-3}$.

Keywords: $PM_{2.5}$ historical data, $PM_{2.5}$ real-time data, reliability, trends

DEDICATION

The completion of my master's studies would not have been possible without the full support and encouragement of my family. My mother, my father and my husband have wholeheartedly inspired, motivated and supported me by all means to accomplish this degree. Thank you for your love and patience.

ACKNOWLEDGMENTS

Firstly, I would like to express my greatly sincere gratitude to my advisor Dr. Xiaohui Wang, for her infinite guidance, patience and immense knowledge throughout my graduate studies. She provided me continuous support for my master's study, related research and great opportunity to take part in joint researches. Her guidance helped me in all the research and writing of this thesis. The most valuable lesson that I learned from her was how to become an independent learner.

In addition to my advisor, my sincere thanks also go to my thesis committee members: Dr. Santanu Chakraborty, Dr. Demba Fofana, and Dr. Chu-Lin Cheng for their time, expertise, insightful comments, and encouragement which helped improve the quality of my thesis.

TABLE OF CONTENTS

LIST OF TABLES

Page

LIST OF FIGURES

CHAPTER I

INTRODUCTION

Beijing, the capital of China, is the second largest city in the country with about 21.71 million of the permanent population in 2017 with a 2.6% average annual growth rate in the recent ten years (Figure 1) (Beijing Municipal Bureau of Statistics, 2018). In the recent decades, Beijing has experienced rapid economic growth, the gross domestic product (GDP) increased from 1139.2 billion Renminbi (RMB) in 2008 to 2801.49 billion RMB in 2017 with a 10.8% average annual growth rate (Figure 1) (Beijing Municipal Bureau of Statistics, 2018). Coupled with the rapid economic growth is the sharp increase in energy consumption. Data published officially by Beijing Statistical Yearbook (2018) indicates that Beijing's primary energy consumption increased from 57.86 million tons of standard coal equivalent in 2008 to 71.33 million tons of standard coal equivalent in 2017 with an average annual growth rate of 2.2%. Nevertheless, the high energy consumption along with stagnant atmospheric conditions has led to severe and chronic air pollution in Beijing (Liang *et al*., 2016; Tian *et al*., 2007).

Fine particulate matter ($PM_{2.5}$) is the key contributors to air pollution with aerodynamic diameters less than 2.5 micrometers (Guo, *et al*., 2014; Liang *et al*., 2016). $PM_{2.5}$ can come from various sources. Zhang *et al* (2013) and Zhang *et al* (2014) found the percentages of source factors that contributed to $PM_{2.5}$ concentrations in Beijing as follows: soil dust 15%, coal combustion 18%, biomass burning 12%, traffic and waste incineration emission 4%, industrial pollution 25%, and secondary inorganic aerosols 26%. It means that $PM_{2.5}$ primarily come from industrial pollution and secondary inorganic aerosols, while rarely from traffic and waste incineration emission.

$PM_{2.5}$ has attracted worldwide attention during the past years due to its adverse effects on human health. $PM_{2.5}$ pollution has become a worldwide environmental issue, in addition to China, many developing countries such as India have experienced severe $PM_{2.5}$ pollution in recent decades (World Bank, 2016). Owing to their minute size, $PM_{2.5}$ can bypass the nose and throat and penetrate deep into the lungs and some may even enter the circulatory system resulting in impairing human health. Schwartz *et al* (1996) revealed that $PM_{2.5}$ was one of the causative factors of daily mortality (non-accidental death) and showed the strongest correlation with it (r = 1.5%, 95% CI: 1.1%–1.9%). Samoli *et al*. (2004) confirmed the linear relationship between mortality and $PM_{2.5}$. Cohen *et al*. (2005) pointed out that about 0.8 million premature deaths were estimated per year because of outdoor air pollution in the world. Zhang *et al*. (2008) found 281,361 deaths were caused by PM pollution in 2004 in the 111 Chinese cities covering most large and medium-sized cities in China and accounted for more than 70% of the national GDP in 2004. Scientists in the study (Pope *et al*., 2002), where data collected from about 500,000 adults in metropolitan areas throughout the United States, estimated that for every 10 micrograms per cubic meter increase in fine particulate air pollution, there is an associated 4%, 6%, and 8% increased risk of all-cause, cardiopulmonary and lung cancer mortality, respectively. Pope *et al*. (2009) found that a decrease of 10 $\mu g.m^{-3}$ of fine particulate concentration was associated with an estimated increase in life expectancy equal to 0.77 (SE = 0.17) year.

Besides the harmful impacts on human health, $PM_{2.5}$ is also known to negatively affect agriculture (Burney *et al*., 2014), reduce visibility (Deng *et al*., 2008; Tao *et al*., 2009; Zhao *et al*., 2013), and affect climate changes (Tai *et al*., 2010).

U.S. Embassy in Beijing is operated by the U.S. Department of State which started reporting hourly $PM_{2.5}$ concentrations in April 2018 followed by the US consulates in Guangzhou (November 2011), Shanghai (December 2011), Chengdu (June 2012), and Shenyang (April 2013). China's Ministry of Environmental Protection (MEP, superseded by Ministry of Ecology and Environment of the People's Republic of China in March 2018) added $PM_{2.5}$ as air pollutants in January 2013 (GB 3095-2012, 2012) and began to publish hourly $PM_{2.5}$ readings from 496 national monitoring stations in 74 major cities since January 2013.

Report of pollutants' readings helps public have a certain understanding of air quality, but public raised suspicions on the data released by MEP sites compared to those reported by the U.S. Embassy in Beijing (Grammaticas, 2011; Wong 2013) due to the traditional phenomena "numbers make leaders" and "leaders make numbers" in China (Liu *et al*, 2009). Discrepancies of $PM_{2.5}$ have been reported between the MEP sites and the U.S. Embassy before 2013 (Spegele, 2012). Stoerk (2016) compared $PM_{2.5}$ released by two sources: the U.S. Embassy and Beijing Municipal Environmental Protection Bureau (BMEPB) from 2008 to 2013. They found misreporting of the official air quality data prior to 2013. Data reliability is a prior condition of data analyses, and confirmation of the reliability of $PM_{2.5}$ data could provide more required confidence in the air quality measurements in China. Hence, the requirement to check $PM_{2.5}$ data reliability is profound.

We investigated the $PM_{2.5}$ data reliability in Beijing through historical and real-time data. The reliability of $PM_{2.5}$ historical data was explored by comparing hourly $PM_{2.5}$ historical concentrations and daily $PM_{2.5}$ percentages between the U.S. Embassy and four nearby MEP monitoring stations based on three years' data from 2014 to 2016. $PM_{2.5}$ percentage is a measurement to assess air quality proposed in this study. $PM_{2.5}$ real-time data reliability was investigated by comparing the hourly $PM_{2.5}$ real-time data released by the U.S. Embassy and three MEP monitoring stations from December 2017 to March 2018.

As China's air pollution has become severe, in order to reduce heavy pollution and improve air quality, "Air Pollution Prevention and Control Action Plan" (hereinafter referred to as Action Plan) was issued by China's State Council on 10^{th} September 2013 which included ten air pollution prevention and control measures. The specific indicators of the Action Plan are that by 2017, PM_{10} concentrations shall decrease by 10% compared to 2012 in Chinese urban areas. The annual number of days with fairly good air quality will gradually increase. $PM_{2.5}$ concentrations will fall by around 25%, 20% and 15% in Beijing-Tianjin-Hebei, Yangtze River Delta and Pearl River Delta region respectively. Besides the requirement of regional air quality improvement, the Action Plan proposed a clear requirement for Beijing's air quality improvement that by 2017, annual average $PM_{2.5}$ concentration should be controlled below 60 $\mu g.m^{-3}$.

Our assessment on improvement trends of air quality measured by both $PM_{2.5}$ concentrations and $PM_{2.5}$ percentages consists of three analyses (1) exploring air quality improvement trends from 2008 to 2017 by calendar view; (2) investigating annual trends of air quality improvement between 2012 and 2016; (3) checking seasonal trends of air quality improvement between 2012 and 2016;

The remainder of the thesis is organized as follows. Chapter II investigates the $PM_{2.5}$ historical data reliability. Chapter III explores the $PM_{2.5}$ real-time data reliability. Chapter IV assesses air quality improvement trends from 2008 to 2017. Chapter V concludes our findings with discussions. Chapter VI provides the limitations of our study and outlines our future work. Statistical software R (version 3.5.1) conducted all statistical analyses.

CHAPTER II

$PM_{2.5}$ HISTORICAL DATA RELIABILITY

In this chapter, we explore the consistency of $PM_{2.5}$ historical data released by the U.S. Embassy and the MEP sites in Beijing. Section 2.1 concludes previous studies about the reliability and consistency of $PM_{2.5}$ data. Data collection information and measurements are provided in Section 2.2. Section 2.3 outlines the statistical methods used in this chapter. Finally, section 2.4 summarizes the results.

2.1 Previous Studies

Researchers (Liang *et al*., 2016) explored the reliability and consistency of hourly $PM_{2.5}$ data by cross-validating data from the U.S. diplomatic posts with those of the nearby MEP sites in the five major Chinese cities including Beijing based on three-year data since January 2013. Their exploration of $PM_{2.5}$ data reliability consisted of three analyses. The first analysis was to compare the three $PM_{2.5}$ concentration ranges duration (in hours), and only 5 (1 in Guangzhou, 3 in Chengdu and 1 in Shenyang) of 33 significant statistical tests supported that the statistics were generally agreeable between the two data sources. Then, they studied the air quality assessment to compare the weather-adjusted monthly mean, median and 90th percentile of $PM_{2.5}$ data as well as yearly changes between U.S. posts and the averages of MEP sites. The high correlation in the statistics and yearly changes verified the high consistency of $PM_{2.5}$ data between the two data sources. The last analysis was to test the effect of winter heating in Beijing and Shenyang, the highly consistent heating effect between two data sources in Beijing provided another support on the $PM_{2.5}$ data reliability and consistency between U.S. posts and MEP sites.

Air Quality Assessment Report (2015) studied the correlation coefficients of hourly $PM_{2.5}$ data from May 1, 2014, to December 31, 2014, between the U.S. Embassy and six nearby MEP sites: Nongzhanguan, Dongsihuan, Qianmen E St, South Ring Road, Olympics, Haidian Wanliu. It found the two closet sites Nongzhanguan and Dongsihuan had the highest correlation coefficients 0.96 and 0.95 respectively, and the farthest sites Haidian and Wanliu still had a correlation coefficient of 0.77. The findings verified the $PM_{2.5}$ data consistency between the U.S. Embassy and the MEP sites.

2.2 Measurements

2.2.1 Data Collection

Beijing Municipal Environmental Monitoring Center (BMEMC) and other official air quality publishing platforms in China report hourly real-time data of pollutants, $PM_{2.5}$ historical data is not accessible to the public. Fortunately, third parties created by civic efforts such as $PM_{2.5}$ historical data (https://www.aqistudy.cn/historydata/), AQISTUDY.cn (https://www.aqistudy.cn/), and BEIJINGAIR.SINAAPP.com (http://beijingair.sinaapp.com/) have been collecting $PM_{2.5}$ historical data since late 2013 (Liu *et al*., 2016). Hourly $PM_{2.5}$ historical data were collected between 2014 and 2016 from the U.S. Embassy in Beijing (http://www.stateair.net/web/historical/1/1.html) and BEIJING.SINAAPP.com. We compared the $PM_{2.5}$ historical data between the U.S. Embassy and 4 nearby MEP monitoring stations Nongzhanguan, Dongsihuan, Dongsi, and Olympics which are geographically close to the U.S. Embassy (Figure 2) to explore the reliability of $PM_{2.5}$ historical data in Beijing.

2.2.2 Air Quality Standards

Table 1 displays air quality standards based on $PM_{2.5}$ concentrations (μg.m^{-3}) between China (GB 3095-2012, 2012) and the US (U.S. Environmental Protection Agency, 2016). This table shows that the same $PM_{2.5}$ concentrations in China and the US are possibly classified into different air quality index categories due to their different air quality standards. Big differences are found in the first four air quality index categories 0 – 50, 51 – 100, 101 -150 and 151 – 200 between China and US, so air quality could be reported differently between the MEP sites and the U.S. Embassy due to this difference even with same $PM_{2.5}$ concentrations.

In order to improve ambient air quality, the Ambient Air Quality Standard (AAQS) was originally issued in China in 1982 (Zhang *et al*., 2016), the latest amendment was made in 2012 and added the $PM_{2.5}$ as one of the pollutants, while the US formulated the National Ambient Air Quality Standards (NAAQS) in 1971 and added the standards of $PM_{2.5}$ in 1997, the new NAAQS was amended in 2012. Table 2 shows the $PM_{2.5}$ guidelines in China, the World Health Organization (WHO) and the United States (US). As can be observed, Chinese grade 1 annual and 24-average $PM_{2.5}$ concentrations are 15 μg.m^{-3} and 35 μg.m^{-3} respectively while the grade 2 values are 35 μg.m^{-3} and 75 μg.m^{-3} respectively. $PM_{2.5}$ grade 2 values are consistent with the WHO Interim Target-1 values while $PM_{2.5}$ grade 1 standard in China is comparable with the US secondary standard.

2.2.3 $PM_{2.5}$ Data Validity

Missing hourly $PM_{2.5}$ concentrations existed in all data sources. To obtain representative data, averages were only computed when periods had at least 75% non-missing values in our study (Ma *et al*., 2011; Li *et al*., 2013; Hu *et al*., 2015; Fontes *et al*., 2017), namely daily average

$PM_{2.5}$ concentrations were only counted for days with at least 75% of hourly concentrations (18 hourly concentrations). $PM_{2.5}$ data validity requirements for monthly, seasonal and annual average $PM_{2.5}$ concentrations are considered similarly and listed in Table 3.

2.2.4 $PM_{2.5}$ Percentages Definition

Liang *et al* (2016) compared the duration (in hours) of three $PM_{2.5}$ concentrations ranges ($PM_{2.5} \leq 35$ μg.m^{-3}; $PM_{2.5} > 35$ μg.m^{-3}; $PM_{2.5} > 150$ μg.m^{-3}) between the U.S. Embassy and nearby MEP sites in five cities to study the $PM_{2.5}$ data reliability. Inspired by this study, we proposed a new measurement to assess air quality and to study the $PM_{2.5}$ historical data reliability. We defined the daily $PM_{2.5}$ percentage as the percent of hours with $PM_{2.5}$ greater than 35 μg.m^{-3} during the valid daily hours. $PM_{2.5}$ data validity criterion is also used to compute the daily, monthly, seasonal and annual average $PM_{2.5}$ percentages. Hence, valid daily hours are between 18 and 24. We adopted 35 μg.m^{-3} as the threshold since it's the standards for China's grade 1 24-hour average $PM_{2.5}$ concentration and daily $PM_{2.5}$ concentration for the US, as well as the interim target-1 value of the annual average concentration of WHO. As 35 μg.m^{-3} is the critical point of China's excellent air quality, $PM_{2.5}$ percentage can be regarded as the percent of unhealthy air quality within a day. Higher $PM_{2.5}$ percentage means worse air quality. In comparison to durations of $PM_{2.5}$ ranges, $PM_{2.5}$ percentages are computed daily and simplify the variable dimensions. In addition, $PM_{2.5}$ percentage makes it feasible to compare the air quality on the daily bases. In Chapter IV, $PM_{2.5}$ percentage will be used to assess air quality improvement trends.

2.3 Methods

The two-fold objective in this section includes comparing (1) hourly $PM_{2.5}$ concentrations from 2014 to 2016 between two data sources and (2) daily percentages in 2015 and 2016 between two data sources. To further explore the $PM_{2.5}$ historical data reliability, we compared the hourly $PM_{2.5}$ concentrations and daily $PM_{2.5}$ percentages between the MEP sites as well. We excluded the comparisons of daily $PM_{2.5}$ percentages between two data sources in 2014 due to the sample sizes of MEP sites in certain months violating the data validity requirement (Table 4). Study of reliability of $PM_{2.5}$ historical data by year can inform us what extent $PM_{2.5}$ data reliability changes over the years.

Correlations of $PM_{2.5}$ data between sites were explored by Pearson's correlation coefficient test. Paired samples t-test was used to test the difference of annual average $PM_{2.5}$ concentrations and annual average $PM_{2.5}$ percentages between sites. In a paired samples t-test, each assumption refers to the differences between the pairs, not the original data values. As a parametric procedure, the paired samples t-test has the assumption of normality. Histogram overlaid with normal curve and boxplot are the effective graphical technique to test the normality assumption and present the distribution of data. Since outliers can bias results, in comparisons of hourly $PM_{2.5}$ concentrations between sites, we removed the extreme outliers in observations that the absolute differences of $PM_{2.5}$ concentrations between sites were greater than or equaled to 350 $\mu g.m^{-3}$. Threshold 350 $\mu g.m^{-3}$ was the difference between the upper bound of $PM_{2.5}$ concentrations in the air quality categories severely polluted and moderately polluted respectively. All statistical analyses were conducted after removing outliers.

Since we have 10 paired samples t-tests performed to compare $PM_{2.5}$ data each year between sites simultaneously, to reduce the risk of rejecting the null hypotheses which are true, we applied the Bonferroni correction to adjust the p-values. Bonferroni correction is to divide the significance level (0.05) by the number of comparisons being made (Dunn 1961). In this chapter, the significance level would be 0.5% and used 99.5% confidence intervals for differences. Since 3 statistical tests to compare the daily $PM_{2.5}$ percentages between MEP sites in 2016 were excluded due to violation of data validity requirement, 7 paired samples t-tests were performed, so the significance level was 0.7% and 99.3% confidence intervals of differences were conducted. *p-value in this Chapter means that the p-value was less than the significance level.

Statistical significance (p-value) can reveal whether the difference between groups exists but cannot show the magnitude of the difference. Large sample size will greatly increase the probability of significant result, hence, with adequately large sample size, the statistical test will almost always show a significant difference (Sullivan *et al*., 2012). In contrast, practical significance can reveal the practical results in the real world (Kirk, 1996). Effect size is an important index used to quantify the degree of practical significance of study results that is to measure the magnitude of the difference. Effect size has two advantages over statistical significance testing that it is independent of sample size and scale-free index. Hence, both statistical significance (p-value) and effect size are essential results to be reported. We performed Cohen's d effect size to measure the magnitude of the differences in annual average $PM_{2.5}$ concentrations and percentages. Cohen's d effect size for paired samples t-test is calculated by the following formula (Gibbons *et al*., 1993):

$$\text{Cohen's d} = \frac{\bar{X}_D}{s_D}$$

where the numerator and denominator are the mean and standard deviation of differences respectively. Cohen (1988) defined effect size as “small: d = 0.20”, “medium: d = 0.50” and “large: d = 0.80”.

2.4 Results

2.4.1 Comparing Hourly $PM_{2.5}$ Concentrations by Year

Table 5 provides the geographical location of monitoring stations and the distance between monitoring stations. Nongzhanguan is closest to U.S. Embassy (2.07 km), followed by Donsihuan (2.24 km), Dongsi (5.15 km) and Olympics (6.68km). The two nearest MEP monitoring stations are Nongzhanguan and Dongsihuan (1.89 km) while the two farthest MEP monitoring stations are Dongsihuan and Olympics (8.75 km).

Table 6 shows that the hourly $PM_{2.5}$ concentrations in four MEP sites were positively highly correlated to that in the U.S. Embassy with Pearson's correlation coefficients above 0.93 in three years. The closest site Nongzhanguan had the highest correlation coefficients 0.949, 0.955 and 0.956 with the U.S. Embassy in 2014, 2015 and 2016 respectively while the lowest correlation coefficients 0.94 and 0.941 were found in 2015 and 2016 between the U.S. Embassy and the farthest site Olympics.

We plotted histograms and boxplots (Figure 3) to present the distributions of differences of hourly $PM_{2.5}$ concentrations between sites. The distributions were symmetric, though two-side tails found, t-tests are powerful with large sample size (>7000) in our study. We performed 12 paired samples t-tests for multiple null hypotheses that the MEP sites had the same annual average $PM_{2.5}$ concentrations with the U.S. Embassy. Detailed and summary of comparison results were listed in Table 6 and Table 8 respectively. 10 of 12 statistical tests were significant with p-values less than 0.005, but effect sizes were either small or negligible. 18 paired samples t-tests were conducted to compare the annual average $PM_{2.5}$ concentrations between MEP sites in three years. We found 16 of 18 all statistical tests were significant, same as comparisons between the U.S. Embassy and the MEP sites, effect sizes were either small or negligible.

Paired samples t-tests showed that comparisons between the MEP sites had a higher percentage of significant differences compared to the comparisons between two data sources. Given the above fact and only small or negligible effect sizes found in comparisons between two data sources, we conclude that the $PM_{2.5}$ historical data were consistent between the two data sources.

2.4.2 Comparing Daily $PM_{2.5}$ Percentage by Year

In addition to comparing the hourly $PM_{2.5}$ concentrations, we explored the consistency of $PM_{2.5}$ historical data between two data sources by comparing daily $PM_{2.5}$ percentages. Comparisons of daily $PM_{2.5}$ percentage in 2014 and 3 comparisons of daily $PM_{2.5}$ percentages in 2016 were excluded in our study due to violation of the data validity requirements (Table 7).

Table 7 shows the Pearson's correlation coefficients between the U.S. Embassy and the MEP sites were extremely high (all above 0.97). $PM_{2.5}$ percentage differences between sites were approximately normally distributed (Figure 4). We found that there were no significant differences in annual average $PM_{2.5}$ percentage between the U.S. Embassy and two MEP sites: Nongzhanguan and Olympics, while the annual average $PM_{2.5}$ percentage of Dongsihuan and Dongsi were significantly greater than that of the U.S. Embassy in 2015 and 2016 with small effect sizes (Table 7). The 99.3% confidence intervals of $PM_{2.5}$ percentage differences showed the maximum absolute $PM_{2.5}$ percentage difference was 4.4% in the comparison between the U.S. Embassy and Dongsihuan in 2016, that means differences of hours with $PM_{2.5}$ greater than 35 $\mu g.m^{-3}$ between two data sources were within 1 hour. Total 9 paired sample t-tests were conducted to compare annual average $PM_{2.5}$ percentages between MEP sites, and 6 of 9 statistical tests were significant with small effect size.

Though 50% of paired sample t-tests were significant for comparing annual average $PM_{2.5}$ percentages between the U.S. Embassy and the MEP sites, the significant differences were small. Furthermore, a higher percentage of significant results and small effect sizes were found in the comparisons between MEP sites. All results support the high agreement of $PM_{2.5}$ data between the U.S. Embassy and the MEP sites.

CHAPTER III

$PM_{2.5}$ REAL-TIME DATA RELIABILITY

3.1 Introduction

In the previous Chapter, our results confirmed the reliability of $PM_{2.5}$ historical data released by two data sources. But as Liang *et al* (2016) mentioned in their study that $PM_{2.5}$ has become a performance measure for local government officials since January 2015, hence continued check of $PM_{2.5}$ data reliability is of great importance. Besides, different from historical data, real-time data is firsthand data. Therefore, study the $PM_{2.5}$ data reliability through $PM_{2.5}$ real-time data is very meaningful and a good way to check the wholeness of $PM_{2.5}$ data. Our study appears to be the first that assesses the $PM_{2.5}$ data reliability using real-time data.

$PM_{2.5}$ real-time data were collected from two data sources: the U.S. Embassy in Beijing and the China National Environmental Monitoring Centre (CNEMC). CNEMC is the directly affiliated institution of Chinese Ministry of Ecology and Environment (MEP) and releases the pollutants' real-time data in the China National Urban Air Quality Real-Time Publishing Platform. The National Urban Air Quality Real-Time Publishing Platform is developed by the Institute of Advanced Technology and the School of Engineering at Sun Yat-sen. According to the requirements of latest Ambient Air Quality Standard (GB3095-2012), the platform reports six monitoring indicators: $PM_{2.5}$, PM_{10}, Sulphur Dioxide (SO_2), Nitrogen Dioxide (NO_2), Ozone (O_3) and Carbon Monoxide (CO) with real-time hourly concentrations, daily average concentrations, AQI and the past 24-hour historical values (Figure 5).

3.2 Measurements

China National Environmental Monitoring Centre (CNEMC) reports hourly real-time $PM_{2.5}$ values for 12 monitoring stations in Beijing: Huairou, Dingling, Changping, Shunyi, Olympics, Wanliu, Nongzhanguan, Dongsi, Guanyuan, Gucheng, Tiantan, and Wanshou. To conduct real-time data reliability analysis, we collected and compared hourly $PM_{2.5}$ real-time data from the U.S. Embassy and its three nearby sites in Beijing: Nongzhanguan, Dongsi, and Olympics.

Hourly real-time $PM_{2.5}$ concentrations were randomly collected in 76 days during December 2017 and March 2018 in sites. 27 days collected 1 hourly $PM_{2.5}$ real-time value, 21, 17, and 10 days collected 2, 3, and 4 hourly $PM_{2.5}$ real-time values respectively, 1 day collected 5 hourly $PM_{2.5}$ real-time values (Figure 6). 165 time points were collected but due to missing values, the U.S. Embassy has 164 valid samples, Nongzhanguan, Dongsi and Olympics has 134, 137, and 138 valid samples respectively. Hourly $PM_{2.5}$ real-time values were mostly randomly collected during Beijing time from 12:00 am to 12:00 pm and 10:00 pm to 11:00 pm (Figure 6).

3.3 Methods

In this part, we used the same statistical methods to check the normality assumptions and examined the correlations of $PM_{2.5}$ real-time concentrations between sites as Chapter II. Since we have 6 paired t-tests performed to compare average $PM_{2.5}$ real-time concentrations between two data sources as well as between the sites, the significance level would be 0.8%, and 99.2% confidence intervals of differences were performed consequently. *p-value in this Chapter means that the p-value was less than the significance level. Finally, Cohen's d effect size was performed to test the magnitude of the differences.

3.4 Results

Table 9 shows the best correlation was found between the U.S. Embassy and the nearest site Nongzhanguan with correlation coefficient 0.994, followed by 0.984, 0.981 with Dongsi and Olympics respectively. The closer distance led to higher correlation coefficients between the U.S. Embassy and the MEP sites. Figure 7 presents the trends of $PM_{2.5}$ real-time concentrations of the U.S. Embassy and the MEP sites matched well. Paired samples t-test showed that the U.S. Embassy had lower average $PM_{2.5}$ real-time concentration than that of Nongzhanguan with a small difference (p-value = <0.001*, 99.2% CI: (-4, -0.5), Cohen's d = 0.307), but had no significant differences with that of Dongsi (p-value = 0.03) and Olympics (p-value = 0.727). In the comparisons of $PM_{2.5}$ real-time concentrations between MEP sites, no significant differences were found.

Though we found a significant difference in average $PM_{2.5}$ real-time data between the U.S. Embassy and Nongzhanguan, the difference was small, and a higher percent of non-significant results were observed. Based on the results, we would say that $PM_{2.5}$ real-time data between two data sources had great consistency.

CHAPTER IV

AIR QUALITY IMPROVEMENT TRENDS

In this Chapter, our exploration on the air quality improvement trends from 2008 to 2017 includes three aspects (1) exploring air quality improvement trends from 2008 to 2017 by calendar view; (2) investigating annual trends of air quality improvement between 2012 and 2016; (3) studying seasonal trends of air quality improvement between 2012 and 2016.

4.1 Previous Studies

Greenstone *et al* (2018) compared the government $PM_{2.5}$ data in 2013 and 2017 from more than 200 monitors throughout China and found air quality improving dramatically since 2013. Specifically, $PM_{2.5}$ concentrations in 2017 fell by 36%, 27% and 34% for the region of Beijing-Tianjin-Hebei, Pearl River Delta and Yangtze River Delta respectively compared to 2013. And the annual average $PM_{2.5}$ concentration in Beijing decreased from 90.6 $\mu g.m^{-3}$ of 2013 to 58.8 $\mu g.m^{-3}$ of 2017.

Liu *et al* (2018) plotted a calendar view of the daily concentrations of $PM_{2.5}$ in Beijing from 2008 to 2017 obtained from the Beijing Environmental Protection Monitoring Centre (BEPMC) which showed that the pollution in Beijing intensified from 2008 to 2011 and then began to alleviate in 2012. The reduction has become more significant since 2015. But the air quality still fell below the international standard for healthy air.

Liang *et al* (2016) assessed the air quality between 2013 to 2015 in five cities in China: Beijing, Shanghai, Guangzhou, Chengdu, and Shenyang, and found a decrease on the $PM_{2.5}$ concentrations in all cities, the decline became significant in 2015. But the annual $PM_{2.5}$ concentration of 80 $\mu g.m^{-3}$ in Beijing in 2015 was more than double the WHO's first interim standard of 35 $\mu g.m^{-3}$ and eight times the WHO's annual air quality guideline of 10 $\mu g.m^{-3}$.

Fontes *et al* (2017) studied the trends of $PM_{2.5}$ concentrations in Chinese five cities: Beijing, Shanghai, Guangzhou, Chengdu, and Shenyang. They collected hourly $PM_{2.5}$ concentrations from the Embassy and Consulates of the U.S. in China from 2008 to 2015. Annual and seasonal average $PM_{2.5}$ concentrations were the average of all hourly concentrations in the specific year and season respectively. In the study of trends in Beijing, the lowest annual average $PM_{2.5}$ concentration was found in 2012 of 90.5 ± 81.7 $\mu g.m^{-3}$, while the highest value was found in 2015 with 104.3 ± 97.7 $\mu g.m^{-3}$. The seasonal evaluation indicated that the $PM_{2.5}$ concentrations in warm seasons decreased 30% - 35% between 2009 and 2015 but almost no variations were observed during cold seasons.

4.2 Measurements

Ten-year hourly $PM_{2.5}$ historical concentrations ($\mu g.m^{-3}$) from 2008 to 2017 were obtained from the U.S. Embassy in Beijing. Daily $PM_{2.5}$ percentages and daily average $PM_{2.5}$ concentrations were calculated based on data validity. Both $PM_{2.5}$ concentrations and $PM_{2.5}$ percentages were assessed to explore air quality improvement trends. According to the data validity, the number of valid days in ten years was listed and due to data validity requirements violated in 2008, 2009, 2010, 2011 and 2017 (Table 10), annual trends and seasonal trends analyses were only applicable from 2012 to 2016.

4.3 Methods

Firstly, calendar view with colors representing the daily $PM_{2.5}$ concentrations and daily $PM_{2.5}$ percentages were plotted. Days with missing values were filled with gray. The calendar view is more informative than other plots (point plot, line plot, etc.) since it can present trends on multiple temporal scales simultaneously and provide intuitive insight (Liu *et al*., 2018).

We explored the annual and seasonal trends by $PM_{2.5}$ concentrations and $PM_{2.5}$ percentages from 2012 to 2016. One-way analysis of variance (ANOVA) is a common inferential procedure to compare means of more than two groups and has two basic assumptions: normality of data in each group and homogeneity of variance. Normal Q-Q plot in diagnostic plots is the intuitive ways to test normality (Abraham *et al*., 2006). Homogeneity of variance assumption was assessed by the Levene test. Assumptions of normality and homogeneous variance of $PM_{2.5}$ concentrations were met after log transformation. One-way ANOVA and its corresponding post-hoc test Tukey's HSD test were performed to test the differences in the annual and seasonal average $PM_{2.5}$ concentrations from 2012 to 2016. The η^2 effect size was conducted to evaluate the magnitude of the differences. Cohen (1988) indicate that in η^2 effect size, 0.01, 0.059 and 0.138 stand for small, medium and large effect size respectively.

Since $PM_{2.5}$ percentages data violated the normality and homogeneous variance assumptions, we conducted non-parametric Kruskal–Wallis test to investigate the annual and seasonal median $PM_{2.5}$ percentages between 2012 and 2016. Dunn test was used for multiple comparisons of groups. The ε^2 effect size was performed. ε^2 effect sizes 0.01, 0.08 and 0.26 indicate for small, medium and large effect size respectively (Kruskal–Wallis Test Handbook). In this Chapter, *p-value, **p-value, and ***p-value indicate that p-values were less than 0.05 (significance level), 0.01, and 0.001 respectively.

4.4 Results

4.4.1 Calendar View

$PM_{2.5}$ concentrations. The graph (Figure 9) shows the air quality was severely polluted between 2008 and 2011, then it got alleviated slightly in 2012 but intensified in 2013, a great improvement has found since 2015. We noticed that higher $PM_{2.5}$ concentrations mainly concentrated in cold seasons: autumn (October and November), winter (December, January, and February) and one month in Spring (March), while lower concentrations occurred in warm seasons (April to September) especially after 2013. Significant improvement was found in warm seasons since 2015 where most days had excellent (green) and good (yellow) air quality. By contrast, during the cold seasons, significant improvement was observed in January and February since 2016. There was almost no significant improvement observed from October to December. Based on the air quality index categories, we defined three $PM_{2.5}$ concentrations ranges "$PM_{2.5} \leq 35\ \mu g.m^{-3}$", "$35\ \mu g.m^{-3} < PM_{2.5} \leq 150\ \mu g.m^{-3}$" and "$PM_{2.5} > 150\ \mu g.m^{-3}$" to stand for the states of excellent air, polluted air, and severely polluted air respectively. We found that 2010 and 2011 reached the highest percentages of days with severe pollution with 22.3% and 22.1% respectively, followed by 2013 with a percentage of 19.9%. Severe pollution days reduced to 17.8%, 14.4%, 10.7%, and 8.9% respectively between 2014 and 2017. The percentages of days with excellent air quality increased from 14.7% in 2009 to 25.1% in 2012, but decreased to 17.5% in 2013, then kept a gradual increase trend to 21.4%, 27.2%, 32.8%, 36.1% in 2014, 2015, 2016 and 2017 (Table 11).

$PM_{2.5}$ percentages. The calendar view of daily $PM_{2.5}$ percentages was plotted in Figure 10 with colors representing daily $PM_{2.5}$ percentages. Figure 10 shows that $PM_{2.5}$ percentages decreased slightly compared to $PM_{2.5}$ concentrations over the years. Contrast to $PM_{2.5}$ concentrations, the higher $PM_{2.5}$ percentages mainly focused in warm seasons (April to August) while lower $PM_{2.5}$ percentages happened in cold seasons (January to March and September to December). To further study the air quality improvement trends by $PM_{2.5}$ percentages, we classified the $PM_{2.5}$ percentages into three ranges "$PM_{2.5}$ percentage = 0", "0 < $PM_{2.5}$ percentage < 100" and "$PM_{2.5}$ percentage = 100" to indicate the states of healthy air, sub-healthy air, and unhealthy air respectively. "$PM_{2.5}$ percentage = 0" means that all hourly $PM_{2.5}$ concentrations within one day were within 35 $\mu g.m^{-3}$, while "$PM_{2.5}$ percentage = 100" means that all hourly $PM_{2.5}$ concentrations within one day were greater than 35 $\mu g.m^{-3}$.

We found the days with healthy air had the lowest percentages in the last decade, while the days with sub-healthy air experienced the highest percentages except 2009 (Figure 11). From the perspective of $PM_{2.5}$ percentages, days with unhealthy air reached the highest percentages in 2008, 2009, and 2010 with 47.6%, 49.5%, and 47.9% respectively, followed by 2013 and 2014 with percentages 42.4% and 44.2% (Table 11), which means above 40% days in these five years had all hourly $PM_{2.5}$ concentration greater than 35 $\mu g.m^{-3}$. Meanwhile, 2008, 2009 and 2013 had lowest percentages of days with healthy air with 1.6%, 3.3%, and 3.3% respectively, which means only 1.6% or 3.3% days in these years had all hourly $PM_{2.5}$ concentration within 35 $\mu g.m^{-3}$. Decrease of $PM_{2.5}$ percentage has become significantly since 2015. The percentages of days with unhealthy air reduced to 35.6%, 32.0%, and 24.4% while the percentage of days with healthy air increased to 9.7%, 10.7%, and 11.7% in 2015, 2016, and 2017 respectively.

4.4.2 Annual Trends of Air Quality Improvement

$PM_{2.5}$ concentrations. Table 12 represents the summary statistics of $PM_{2.5}$ concentrations and $PM_{2.5}$ percentages measured from 2012 to 2016. We found the highest annual average $PM_{2.5}$ concentrations in 2013 with 102.2 ± 83.0 μg.m^{-3}, $PM_{2.5}$ concentrations had a clear decrease from 2013 to 2016 and the annual average concentration reduced from 102.0 μg.m^{-3} in 2013 to 72.9 μg.m^{-3} in 2016.

The result (p-value<.001***) indicates the $PM_{2.5}$ concentrations significantly changed over the years, but the effect size was small ($\eta^2 = 0.03$). As we observed in the calendar view of $PM_{2.5}$ concentrations, post hoc test shows that a significant decrease was found since 2015. Average $PM_{2.5}$ concentration in 2016 decreased by 18.9%, 28.5% and 25.3% compared to that in 2012, 2013 and 2014 respectively. And average $PM_{2.5}$ concentration in 2015 decreased by 19% and 15.4% compared to that in 2013 and 2014 respectively (Table 13).

$PM_{2.5}$ percentages. Result (Table 12) shows 2013 had the highest annual median $PM_{2.5}$ percentage with 90.9%, namely, days in 2013 had 90.9% hours with $PM_{2.5}$ concentrations greater than 35 μg.m^{-3} on average. Similar to $PM_{2.5}$ concentrations, $PM_{2.5}$ reduced continuously from 2013 to 2016 but the median $PM_{2.5}$ percentage with 75% in 2016 was still high. Kruskal–Wallis test (p-value < 0.001***) shows there was a significant difference in median percentages among years, but the difference was small ($\varepsilon^2 = 0.015$). The decreases in annual median $PM_{2.5}$ percentages were 22.1% and 19.1% for 2015 with comparisons to that in 2013 and 2014 respectively, and the decreases in annual median $PM_{2.5}$ percentages were 17% and 14.3% for 2016 with comparisons to that in 2013 and 2014.

4.4.3 Season Trends of Air Quality Improvement

$PM_{2.5}$ concentrations. Table 14 shows the summary statistics of $PM_{2.5}$ concentrations and $PM_{2.5}$ percentages measured by seasons from 2012 to 2016. Combined with Figure 13, the $PM_{2.5}$ concentrations present a significant seasonal trend, higher concentrations mainly happened in winter, followed by autumn and spring, while the lower concentrations occurred during summer. As we can observe, the average $PM_{2.5}$ concentrations in spring and autumn decreased slightly over the years while summer had a large decline in 2015 and 2016 compared to previous years. In winter 2013 experienced the highest average $PM_{2.5}$ concentration, in contrast, $PM_{2.5}$ concentrations alleviated greatly in 2016.

We found seasonal average $PM_{2.5}$ concentrations were significantly different from 2012 to 2016, the medium difference in average $PM_{2.5}$ concentrations was found in summer while small differences were found in the other three seasons (Table 15). Pairwise comparisons show that in spring, average $PM_{2.5}$ concentration in 2016 was lower than that in 2014 with a decrease of 22.7%. Average $PM_{2.5}$ concentrations decreased 40.8%, 36.4% and 27.5% in 2015 compared to that in 2012, 2013 and 2014 respectively, and it decreased 39.4% and 34.9% in 2016 compared to that in 2012 and 2013 respectively in summer. Average $PM_{2.5}$ concentration in 2015 was lower than that in 2013 and 2014 in autumn with decreases of 15.6% and 23.4% respectively, and average $PM_{2.5}$ concentration in 2016 decreased 37.6%, 27.9%, and 28.2% compared to that in 2013, 2014, and 2015 respectively in winter.

$PM_{2.5}$ percentages. $PM_{2.5}$ percentages show a different seasonal pattern from $PM_{2.5}$ concentrations. The higher median $PM_{2.5}$ percentages were found in spring while lower median $PM_{2.5}$ percentages were observed in (Table 14). $PM_{2.5}$ percentages in winter have reduced since

2014 and had a sharp decrease since 2015. $PM_{2.5}$ percentages decreased greatly in 2015 and 2016 in summer compared to 2012 and 2013, while spring and autumn did not show continuous downward trends.

Results (Table 15) show that the median $PM_{2.5}$ percentages were significantly different among years in summer (p-value < 0.001***), autumn (p-value = 0.039*) and winter (p-value = 0.014***), but all differences were small. In further analyses, we found in summer, the median $PM_{2.5}$ percentage in 2015 was less than that in both 2012 and 2013 with decreases of 37.5% and 34.8% respectively, and 2016 had lower median $PM_{2.5}$ percentages than that in 2012 with the decrease of 20.8%. Though we found a significant difference in median $PM_{2.5}$ percentages in autumn, the post-hoc test shows no significant differences between any two years, and we will discuss this contradictory result in the Discussion chapter. In winter, only 2016 had lower median $PM_{2.5}$ percentages than that in 2013 (p-value = 0.011*) with a decrease of 46.4%.

CHAPTER V

CONCLUSIONS AND DISCUSSIONS

We compared hourly $PM_{2.5}$ concentrations and daily $PM_{2.5}$ percentages between the U.S. Embassy and four nearby MEP sites as well as between the MEP sites based on three years' data since 2014 to study the reliability of $PM_{2.5}$ historical data between two data sources. We found almost all paired samples t-tests were significant for comparing annual average $PM_{2.5}$ concentrations between sites from 2014 to 2016. But all significant differences were either small or negligible, which meant the differences were not practically significant. In comparisons of daily $PM_{2.5}$ percentages, 50% paired samples t-tests showed small differences on annual average $PM_{2.5}$ percentages between two data sources, while higher percent (66.7%) of paired samples t-tests were statistically significant in comparisons of annual average $PM_{2.5}$ percentages between the MEP sites with small effect sizes. Our findings supported that $PM_{2.5}$ historical data between two data sources were highly consistent. Generally, the closer distance between the U.S. Embassy and the MEP sites led to higher correlation coefficients, and the correlations of $PM_{2.5}$ concentrations between the U.S. Embassy and the MEP sites has increased over the years.

To further explore the $PM_{2.5}$ data reliability, investigation on $PM_{2.5}$ real-time data reliability was conducted as well. Analyses results showed that $PM_{2.5}$ real-time data in the U.S. Embassy had extremely high correlation coefficients 0.994, 0.984 and 0.981 with that in Nongzhanguan, Dongsi and Olympics respectively, and closer distance between the U.S. Embassy led to higher correlation coefficients. After Bonferroni correction, only 1 of 3 comparisons of average $PM_{2.5}$ real-time concentrations between two data sources was significant with small effect size, 99.2% confidence interval showed that the absolute differences of $PM_{2.5}$ real-time concentrations in the significant comparison group were within 4 $\mu g.m^{-3}$. Given these facts, we would say that $PM_{2.5}$ real-time data between two data sources had a great agreement.

Official statistics indicate that population, gross domestic product (GDP) and energy consumption had consistent increase trends from 2008 to 2017 with average annual growth rates of 2.6%, 10.8%, and 2.2% respectively. We studied the air quality improvement trends in Beijing from 2008 to 2017 by three analyses that were calendar view of daily $PM_{2.5}$ concentrations and daily $PM_{2.5}$ percentages from 2008 to 2017, the study of annual and seasonal air quality improvement trends between 2012 and 2016.

Our assessment on the air quality improvement trends found that air quality has significantly improved since 2015. Highest average $PM_{2.5}$ concentration was found in 2013 of 102 ± 83 $\mu g.m^{-3}$, and highest median $PM_{2.5}$ percentage was detected in 2013 of 90.9%. The annual average $PM_{2.5}$ concentrations reduced from 102.0 $\mu g.m^{-3}$ in 2013 to 72.9 $\mu g.m^{-3}$ in 2016. The decreases in annual average $PM_{2.5}$ concentrations were 19.0% and 28.5% in 2015 and 2016 respectively compared to that in 2013. The annual median $PM_{2.5}$ percentages decreased from 90.9% in 2013 to 75% in 2016. And the annual median $PM_{2.5}$ percentages in 2015 and 2016 decreased by 22.1% and 17.5% with comparisons to that in 2013. Though air quality had a significant improvement since 2015, the annual average $PM_{2.5}$ concentration in 2016 of 72.9 $\mu g.m^{-3}$ was still over two times the first interim target of annual $PM_{2.5}$ concentration 35.0 $\mu g.m^{-3}$ of WHO and Chinese grade 2 annual average concentration 35.0 $\mu g.m^{-3}$. And the median $PM_{2.5}$ percentage of 75% in 2016 means that there were 75% hours in days in 2016 with unhealthy air ($PM_{2.5}$ concentration > 35 $\mu g.m^{-3}$). Due to the lack of data in the second half of 2017, we did not analyze the annual and seasonal trends in 2017, the average $PM_{2.5}$ concentration of 69.2 $\mu g.m^{-3}$ in the first half of 2017 was lower than that in 2016 but was still above the specific indicator of below 60 $\mu g.m^{-3}$ in Beijing by 2017 set in the Action Plan.

Percentages of days with excellent air quality ($PM_{2.5}$ concentration $\leq$ 35 $\mu g.m^{-3}$) increased from 17.5% in 2013 to 27.2%, 32.8% and 36.1% in 2015, 2016 and 2017 respectively, and the percentages of days with severe air pollution ($PM_{2.5}$ concentration > 150 $\mu g.m^{-3}$) decreased from 19.9% to 14.4%, 10.7% and 8.9% in 2015, 2016 and 2017 respectively. These facts indicate the annual number of days with excellent air quality has gradually increased over the years.

Study of seasonal trends has detected that winter had higher $PM_{2.5}$ concentrations than the other three seasons while the lower concentrations occurred during summer. Air quality has improved greatly in summer since 2015, average $PM_{2.5}$ concentrations decreased by 40.8% and 39.4% in 2015 and 2016 compared to that in 2012 respectively which had highest average $PM_{2.5}$ concentration of 86.5 $\mu g.m^{-3}$ in summer. Small improvement has occurred in autumn since 2015, and small improvement happened in spring and winter beginning in 2016.

In term of $PM_{2.5}$ percentages, percent of unhealthy air has decreased slightly in summer since 2015 with decreases of 37.5% and 20.8% in median $PM_{2.5}$ percentages in 2015 and 2016 compared to that in 2012 respectively. Median $PM_{2.5}$ percentage decreased by 46.6% in 2016 compared to that in 2013 in winter. But the percent of unhealthy air had no variations in spring and autumn.

Though China made big progress in the battle with air pollution, there is still a long way to achieve the target set by China's Ambient Air Quality Standards (GB 3095-2012) and by the Action Plan. The fight against air pollution will continue.

In the study of $PM_{2.5}$ historical data reliability, we noticed that the U.S. Embassy had higher annual average $PM_{2.5}$ concentration than the MEP sites in 2014, but then it had a lower concentration than the MEP sites since 2015.

In the study of seasonal air quality improvement trends by $PM_{2.5}$ percentages, contradictory results were found between Kruskal–Wallis test and multiple pairwise comparisons. Kruskal–Wallis test was significant (p-value = 0.039*) to compare the median $PM_{2.5}$ percentages in autumn but the post-hoc tests showed no significant differences between any two years. We found some explanations for the contradictory results. (1) A lack of statistical power. For example, when groups have small sample sizes, pairwise comparison tests are not statistically powerful, it is less likely to detect significant differences. (2) A high number of factor levels. The more the pairwise comparisons, the more the p-values get penalized to reduce the risk of rejecting null hypotheses while they are true. (3) A weakly significant effect (p-value of inferential procedure equals to or closes to the significance level). (4) A conservative multiple comparisons test. The more conservative the test, the more likely to fail to reject the false null hypothesis. In our study, the sample size (over 90 in five years) was not the reason to create contradictory results. But factor levels (five years created 10 pairwise comparisons) and the weakly significant effect (p-value = 0.039*) could be the explanations to the contradictory results.

CHAPTER VI

LIMITATIONS AND FUTURE WORK

This study has three potential limitations. Firstly, in the study of $PM_{2.5}$ historical data reliability by comparing hourly $PM_{2.5}$ concentrations between sites, the sample sizes were too large, which led to nearly all paired samples t-tests were significant. Secondly, in the $PM_{2.5}$ real-time data reliability assessment, $PM_{2.5}$ real-time data were missing in some days during the data collection period. Thirdly, $PM_{2.5}$ concentrations are known to be influenced by pollutants emissions, meteorological conditions and their interaction (Sun *et al*., 2014). Weather variables, e.g., wind direction, humidity, and temperature can influence $PM_{2.5}$ concentrations. Those factors were not considered in the study of air quality improvement trends but can be the potential direction for future studies.

For future work, we will perform simple random sampling in hourly $PM_{2.5}$ concentrations data for all sites to obtain proper sample sizes and then conduct paired samples t-tests. We will consider repeating this process in a certain number of times (e.g. 1000 times) and then check the results. Secondly, $PM_{2.5}$ data reliability assessment will be extended from one city to five cities Beijing, Shanghai, Guangzhou, Chengdu, and Shenyang. $PM_{2.5}$ real-time data will be collected in the winter-heating period for Beijing and Shenyang, and in winter for Shanghai, Guangzhou, and Chengdu. More efforts will be adopted to collect 2 to 4 hourly $PM_{2.5}$ real-time concentrations daily to prevent missing values. Thirdly, we will consider adopting statistical models to perform a meteorological adjustment to minimize the meteorological effects on $PM_{2.5}$ concentrations.

REFERENCES

Abraham, B., & Ledolter, J. (2006). Introduction to regression modeling. Belmont, CA: Thomson Brooks/Cole.

Altman DG, Bland JM. Statistics notes: the normal distribution. BMJ. 1995;310 (6975):298.

Beijing Municipal Bureau of Statistics. (2018). Retrieved from http://www.bjstats.gov.cn/tjsj/.

Beijing Municipal Environmental Protection Bureau (BMEPB). (2012). http://www.bjepb.gov.cn/bjhrb/xxgk/fgwj/qtwj/zcjd/608633/index.html.

Beijing Statistical Yearbook. (2018). Retrieved from http://tjj.beijing.gov.cn/nj/main/2018-tjnj/zk/indexeh.htm.

Burney, J., & Ramanathan, V. (2014). Recent climate and air pollution impact on Indian agriculture. Proceedings of the National Academy of Sciences, 111(46), 16319-16324. doi:10.1073/pnas.1317275111.

Center for Statistical Science at Peking University, Guanghua School of Management at Peking University. (2015). Air Quality Assessment Report: Study on $PM_{2.5}$ Pollution Status in Beijing Urban District from 2010 to 2014. https://wenku.baidu.com/view/9c45b7031711cc7931b716c8.html.

CNEMC. (2012). Retrieved from http://www.cnemc.cn/showSearchContent.do?contentId=33399&siteId=2002.

Cohen, A. J., Anderson, H. R., Ostro, B., Pandey, K. D., Krzyzanowski, M., Künzli, N., . . . Smith, K. (2005). The Global Burden of Disease Due to Outdoor Air Pollution. Journal of Toxicology and Environmental Health, Part A, 68(13-14), 1301-1307. doi:10.1080/15287390590936166.

Cohen, J. (1988). Statistical power analysis for the behavioral sciences (2nd ed., p. 283). Hillsdale, NJ: Erlbaum.

Deng, X., Tie, X., Wu, D., Zhou, X., Bi, X., Tan, H., . . . Jiang, C. (2008). Long-term trend of visibility and its characterizations in the Pearl River Delta (PRD) region, China. Atmospheric Environment, 42(7), 1424-1435. doi: 10.1016/j.atmosenv.2007.11.025.

Dunn, O. J. (1961). Multiple Comparisons among Means. Journal of the American Statistical Association, 56:293, 52-64, DOI: 10.1080/01621459.1961.10482090.

Fontes, T., Li, P., Barros, N., & Zhao, P. (2017). Trends of $PM_{2.5}$ concentrations in China: A long term approach. Journal of Environmental Management, 196, 719-732. doi: 10.1016/j.jenvman.2017.03.074.

GB 3095-2012. (2012). Ambient Air Quality Standards. Retrieved from http://210.72.1.216:8080/gzaqi/Document/gjzlbz.pdf.

Ghasemi, A., & Zahediasl, S. (2012). Normality Tests for Statistical Analysis: A Guide for Non-Statisticians. International Journal of Endocrinology and Metabolism, 10(2), 486-489. doi:10.5812/ijem.3505.

Gibbons, R. D., Hedeker, D. R., & Davis, J. M. (1993). Estimation of effect size from a series of experiments involving paired comparisons. Journal of Educational Statistics, 271-279.

Glass, G. V., McGaw, B., & Smith, M. L. (1981). Meta-analysis in social research. Beverly Hills: Sage Publications.

Grammaticas, D. (2011, December 07). China's pollution, a toxic issue - BBC News. Retrieved from https://www.bbc.co.uk/news/world-asia-16068926.

Greenstone, M., & Schwarz, P. (2018). Is China Winning its War on Pollution? The Air Quality Life Index™. The Energy Policy Institute at the University of Chicago. Retrieved from https://aqli.epic.uchicago.edu/.

Guo, S., Hu, M., Zamora, M. L., Peng, J., Shang, D., Zheng, J., . . . Zhang, R. (2014). Elucidating severe urban haze formation in China. Proceedings of the National Academy of Sciences, 111(49), 17373-17378. doi:10.1073/pnas.1419604111.

Guo, S., Hu, M., Guo, Q., Zhang, X., Schauer, J. J., & Zhang, R. (2013). Quantitative evaluation of emission controls on primary and secondary organic aerosol sources during Beijing 2008 Olympics. Atmospheric Chemistry and Physics, 13(16), 8303-8314. doi:10.5194/acp-13-8303-2013.

HJ 633-2012. (2012). Technical Regulation on Ambient Air Quality Index. Retrieved from http://210.72.1.216:8080/gzaqi/Document/aqijsgd.pdf.

Hojat, M., & Xu, G. (2004). A Visitors Guide to Effect Sizes – Statistical Significance Versus Practical (Clinical) Importance of Research Findings. Advances in Health Sciences Education, 9(3), 241-249. doi:10.1023/b: ahse.0000038173. 00909.f6.

How to interpret contradictory results b... (n.d.). Retrieved from https://help.xlstat.com/customer/es/portal/articles/2852820-how-to-interpret-contradictory-results-between-anova-and-multiple-pairwise-comparisons-?b_id=9283.

Hu, H., Ha, S., Henderson, B. H., Warner, T. D., Roth, J., Kan, H., & Xu, X. (2015). Association of Atmospheric Particulate Matter and Ozone with Gestational Diabetes Mellitus. Environmental Health Perspectives. doi:10.1289/ehp.1408456.

Kirk, R. E. (1996). Practical Significance: A Concept Whose Time Has Come. Educational and Psychological Measurement, 56(5), 746–759. https://doi.org/10.1177/0013164496056005002.

Kruskal–Wallis Test. (n.d.). Retrieved from http://rcompanion.org/handbook/F_08.html.

Li, P., Xin, J., Wang, Y., Wang, S., Shang, K., Liu, Z., . . . Wang, M. (2013). Time-series analysis of mortality effects from airborne particulate matter size fractions in Beijing. Atmospheric Environment, 81, 253-262. doi: 10.1016/j.atmosenv.2013.09.004.

Liang, X., Li, S., Zhang, S., Huang, H., & Chen, S. X. (2016). $PM_{2.5}$ data reliability, consistency, and air quality assessment in five Chinese cities. Journal of Geophysical Research: Atmospheres, 121(17). doi:10.1002/2016jd024877.

Liang, X., Zou, T., Guo, B., Li, S., Zhang, H., Zhang, S., . . . Chen, S. X. (2015). Assessing Beijing's $PM_{2.5}$ pollution: Severity, weather impact, APEC and winter heating. Proceedings of the Royal Society A: Mathematical, Physical and Engineering Science, 471(2182), 20150257. doi:10.1098/rspa.2015.0257.

Liu, J., Li, J., & Li, W. (2016). Temporal Patterns in Fine Particulate Matter Time Series in Beijing: A Calendar View. Scientific Reports, 6(1). doi:10.1038/srep3222.

Liu, J., & Yang, H. (2009). China Fights Against Statistical Corruption. Science, 325(5941), 675-676. doi:10.1126/science.325_675.

Liu, L., Silva, E. A., & Liu, J. (2018). A decade of battle against $PM_{2.5}$ in Beijing. Environment and Planning A: Economy and Space. doi:10.1177/0308518x18766633.

Ma, Y., Chen, R., Pan, G., Xu, X., Song, W., Chen, B., & Kan, H. (2011). Fine particulate air pollution and daily mortality in Shenyang, China. Science of The Total Environment, 409(13), 2473-2477. doi: 10.1016/j.scitotenv.2011.03.017.

Malhotra, N. K., & Birks, D. F. (2007). Marketing research: An applied approach. Harlow: Prentice Hall.

Martini, F. M., Hasenkopf, C. A., & Roberts, D. C. (2015). Statistical analysis of $PM_{2.5}$ observations from diplomatic facilities in China. Atmospheric Environment, 110, 174-185. doi: 10.1016/j.atmosenv.2015.03.060.

McCune, B. and Grace, J.B. 2002 Analysis of Ecological Communities. MJM Software Design, Gleneden Beach, Oregon, USA.

Pope, C. A., Burnett, R. T., Thun, M. J., Calle, E. E., Krewski, D., Ito, K., & Thurston, G. D. (2002). Lung Cancer, Cardiopulmonary Mortality, and Long-term Exposure to Fine Particulate Air Pollution. JAMA: The Journal of the American Medical Association, 287(9), 1132–1141.

Pope, C. A., Ezzati, M., & Dockery, D. W. (2009). Fine Particulate Air Pollution and US County Life Expectancies. N Engl J Med. 2009 January 22; 360(4): 376–386. doi:10.1056/NEJMsa0805646.

Pope, C. A., III, M. Ezzati, and D.W. Dockery. (2009). Fine-particulate air pollution and life expectancy in the United States, N. Engl. J. Med., 360, 376–386.

Prajapati, S. K. (2012). Ecological effect of airborne particulate matter on plants, Environ. Skep. Crit., 1, 12–22.

Report on the State of the Environment in China. (2017). Retrieved from http://www.mep.gov.cn/hjzl/zghjzkgb/lnzghjzkgb/201805/P020180531534645032372.pdf.

Samoli, E., Analitis, A., Touloumi, G., Schwartz, J., Anderson, H. R., Sunyer, J., . . . Katsouyanni, K. (2004). Estimating the Exposure–Response Relationships between Particulate Matter and Mortality within the APHEA Multicity Project. Environmental Health Perspectives, 113(1), 88-95. doi:10.1289/ehp.7387.

Schwartz, J., Dockery, D. W., & Neas, L. M. (1996). Is Daily Mortality Associated Specifically with Fine Particles? Journal of the Air & Waste Management Association, 46(10), 927-939. doi:10.1080/10473289.1996.10467528.

Spegele, B. (2012, January 23). Comparing Pollution Data: Beijing vs. U.S. Embassy on $PM_{2.5}$. Retrieved from https://blogs.wsj.com/chinarealtime/2012/01/23/comparing-pollution-data-beijing-vs-u-s-embassy-on-pm2-5/.

Stoerk, T. (2016). Statistical corruption in Beijing's air quality data has likely ended in 2012. Atmospheric Environment, 127, 365-371. doi: 10.1016/j.atmosenv.2015.12.055.

Sullivan, G. M., & Feinn, R. (2012). Using Effect Size—or Why the P-Value Is Not Enough. Journal of Graduate Medical Education, 4(3), 279-282. doi:10.4300/jgme-d-12-00156.1.

Table of Historical Particulate Matter (PM) National Ambient Air Quality Standards (NAAQS). (2018, March 15). Retrieved from https://www.epa.gov/pm-pollution/table-historical-particulate-matter-pm-national-ambient-air-quality-standards-naaqs.

Tai, A. P., Mickley, L. J., & Jacob, D. J. (2010). Correlations between fine particulate matter ($PM_{2.5}$) and meteorological variables in the United States: Implications for the sensitivity of $PM_{2.5}$ to climate change. Atmospheric Environment, 44(32), 3976-3984. doi: 10.1016/j.atmosenv.2010.06.060.

Tao, J., Ho, K., Chen, L., Zhu, L., Han, J., & Xu, Z. (2009). Effect of chemical composition of $PM_{2.5}$ on visibility in Guangzhou, China, 2007 spring. Particuology, 7(1), 68-75. doi: 10.1016/j.partic.2008.11.002.

Tian, H., Hao, J., Hu, M., & Nie, Y. (2007). Recent Trends of Energy Consumption and Air Pollution in China. Journal of Energy Engineering, 133(1), 4-12. doi:10.1061/(asce)0733-9402(2007)133:1(4).

U.S. Environmental Protection Agency. (2016). Technical assistance document for the reporting of daily air quality: The air quality index (AQI).

Wong, E. (2013, January 12). On Scale of 0 to 500, Beijing's Air Quality Tops 'Crazy Bad' at 755-New York Times. Retrieved from www.nytimes.com/2013/01/13/science/earth/beijing-air-pollution-off-the-charts.html?_r=0).

World Bank. (2016). The cost of air pollution: strengthening the economic case for action (English). Washington, D.C.: World Bank Group. http://documents.worldbank.org/curated/en/781521473177013155/The-cost-of-air-pollution-strengthening-the-economic-case-for-action.

World Health Organization. (2005). WHO Air quality guidelines for particulate matter, ozone, nitrogen dioxide and sulfur dioxide Global update 2005 -Summary of risk assessment.

Zhang, H., Wang, S., Hao, J., Wang, X., Wang, S., Chai, F., & Li, M. (2016). Air pollution and control action in Beijing. Journal of Cleaner Production, 112, 1519-1527. doi: 10.1016/j.jclepro.2015.04.092.

Zhang, M., Song, Y., Cai, X., & Zhou, J. (2008). Economic assessment of the health effects related to particulate matter pollution in 111 Chinese cities by using economic burden of disease analysis. Journal of Environmental Management, 88(4), 947-954. doi: 10.1016/j.jenvman.2007.04.019.

Zhang, R., Jing, J., Tao, J., Hsu, S., Wang, G., Cao, J., . . . Zhao, Y. (2013). Chemical characterization and source apportionment of $PM_{2.5}$ in Beijing: Seasonal perspective. Atmospheric Chemistry and Physics Discussions, 13(4), 9953-10007. doi:10.5194/acpd-13-9953-2013.

Zhang, R., Jing, J., Tao, J., Hsu, S., Wang, G., Cao, J., . . . Shen, Z. (2014). Corrigendum to "Chemical characterization and source apportionment of $PM_{2.5}$ in Beijing: Seasonal perspective" published in Atmos. Chem. Phys., 13, 7053–7074, 2013. Atmospheric Chemistry and Physics, 14(1), 175-175. doi:10.5194/acp-14-175-2014.

Zhao, H., Che, H., Zhang, X., Ma, Y., Wang, Y., Wang, H., & Wang, Y. (2013). Characteristics of visibility and particulate matter (PM) in an urban area of Northeast China. Atmospheric Pollution Research, 4(4), 427-434. doi:10.5094/apr.2013.049.

APPENDIX A

APPENDIX A

TABLES

Table 1. Air quality standards based on $PM_{2.5}$ concentrations ($\mu g.m^{-3}$)

	China		US	
Air Quality Index (AQI)	AQI Categories	24-hour Average $PM_{2.5}$ Concentrations ($\mu g.m^{-3}$)	AQI Categories	24-hour Average $PM_{2.5}$ Concentrations ($\mu g.m^{-3}$)
0 - 50	Excellent	0 - 35	Good	0.0 - 12.0
51 - 100	Good	35 - 75	Moderate	12.1 - 35.4
101 - 150	Light polluted	75 - 115	Unhealthy for sensitive groups	35.5 - 55.4
151 - 200	Moderately polluted	115 - 150	Unhealthy	55.5 - 150.4
201 - 300	Heavily polluted	150 - 250	Very unhealthy	150.5 - 250.4
301 - 500	Severely polluted	250 - 500	Hazardous	250.5 - 500.4

Table 2. $PM_{2.5}$ guidelines in China, the World Health Organization (WHO) and the US

	China (2012) [a]		**WHO (2005)** [b]				**US (2012)** [c]	
$PM_{2.5}$	Grade 1	Grade 2	Interim Target-1	Interim Target-2	Interim Target-3	AQG [d]	Primary	Secondary
Annual average	15	35	35	25	15	10	12	15
24-hour average	35	75	75	50	37.5	25	35 [e]	35 [e]

[a] GB 3095-2012, 2012

[b] World Health Organization, 2005

[c] National Ambient Air Quality Standards (NAAQS)

[d] Air quality guideline (AQG)

[e] instead of 24-hour average, these are 98th percentile.

Table 3. $PM_{2.5}$ data validity requirement

$PM_{2.5}$ Concentrations and $PM_{2.5}$ Percentages	Data Validity Requirement
Annual average	All months in the year have at least 75% of daily $PM_{2.5}$ values or the year has at least 75% of hourly concentrations (if the annual average concentration is calculated by the average of all hourly concentrations within the year)
Seasonal average	All months in the season have at least 75% of daily $PM_{2.5}$ values
Month average	The month has at least 75% of daily $PM_{2.5}$ values, for months (January, March, May, July, August, October, and December) with 31 days, have at least 24 daily $PM_{2.5}$ values; for months (April, June, September, and November) with 30 days, have at least 23 daily $PM_{2.5}$ values; for February with 28 days, has at least 21 daily $PM_{2.5}$ values; for February with 29 days, has at least 22 daily $PM_{2.5}$ values.
Daily average [a]	The day has at least 75% of hourly $PM_{2.5}$ concentrations (18 hourly $PM_{2.5}$ concentrations)

[a] daily average $PM_{2.5}$ concentrations or daily percentages.

$PM_{2.5}$ values: $PM_{2.5}$ concentrations or $PM_{2.5}$ percentages.

Table 4. Number of valid days by month and by year of the U.S. Embassy and the MEP sites

	Jan.	Feb.	Mar.	Apr.	May	June	July	Aug.	Sep.	Oct.	Nov.	Dec.	Total
2014													
U.S. Embassy	31	28	31	30	31	29	31	30	30	31	29	29	360
Nongzhanguan	31	27	30	**15**	29	30	28	31	29	29	29	31	339
Dongsihuan	30	26	30	**16**	31	30	28	31	30	**20**	30	31	333
Dongsi	31	26	30	**17**	29	30	30	31	30	30	30	29	343
Olympics	31	26	30	**16**	31	30	30	31	30	28	30	**16**	329
2015													
U.S. Embassy	31	28	29	29	31	29	31	31	30	31	30	30	360
Nongzhanguan	31	25	29	28	30	30	31	31	27	28	28	31	349
Dongsihuan	29	25	29	27	27	30	24	31	29	31	30	31	343
Dongsi	31	25	30	30	29	30	31	31	30	31	30	31	359
Olympics	31	25	30	28	31	28	31	28	30	31	29	31	353
2016													
U.S. Embassy	31	29	31	30	31	30	31	31	30	31	30	31	366
Nongzhanguan	31	29	31	29	31	30	30	29	29	31	30	26	356
Dongsihuan	31	29	31	28	29	29	24	29	29	31	30	25	345
Dongsi	31	29	30	29	31	27	24	29	28	27	30	27	342
Olympics	31	29	28	28	31	30	29	29	29	31	30	26	351

The bold is to illustrate that number of daily $PM_{2.5}$ concentrations ($\mu g.m^{-3}$) and daily $PM_{2.5}$ percentages within the month are less than the data validity requirements, annual, seasonal and monthly average $PM_{2.5}$ concentrations and $PM_{2.5}$ percentages cannot be computed due to this violation. Hence, comparisons of daily $PM_{2.5}$ percentages between sites were excluded in 2014.

Table 5. The geographical location of the U.S. Embassy and the MEP sites.

Monitoring Station	Location [b]	U.S. Embassy	Nongzhanguan	Dongsihuan	Dongsi	Olympics
U.S. Embassy	39.955°N, 116.467°E [a]	-	2.07 km	2.24 km	5.15 km	6.68 km
Nongzhanguan	39.937°N, 116.461°E	-	-	1.89 km	3.86 km	7.4 km
Dongshihuan	39.939°N, 116.483°E	-	-	-	5.74 km	8.75 km
Dongsi	39.929°N, 116.417°E	-	-	-	-	6.13 km
Olympics	39.982°N, 116.397°E	-	-	-	-	-

[a] Prior to February 16, 2009, the Beijing air quality monitor was located at No. 3 Xiu Shui Bei Jie, Chaoyang District, Beijing (39.31°N, 116.44°E). Beginning on February 16, 2009, at 23:00 BJT, the air monitor began operating at No. 55 An Jia Lou Rd., Chaoyang District, Beijing (Fontes et al., 2017).

[b] Retrieved from http://beijingair.sinaapp.com/.

Distance between two monitoring stations was calculated in https://gps-coordinates.org/distance-between-coordinates.php.

Table 6. Comparison results of hourly $PM_{2.5}$ concentrations between the U.S. Embassy and the MEP sites.

		N	Mean ± SD	99.5% CI of the Difference	Correlation Coefficient	Paired t-test [a]	Cohen's d Effect Size
2014							
Test 1	U.S. Embassy Nongzhanguan	8118	6.1 ± 29	(5.2, 7)	0.949	<0.001*	0.211 (small)
Test 2	U.S. Embassy Dongsihuan	7988	1 ± 32.5	(-0.1, 2)	0.935	0.009	0.029 (negligible)
Test 3	U.S. Embassy Dongsi	8225	9.4 ± 32.7	(8.4, 10.4)	0.936	<0.001*	0.288 (small)
Test 4	U.S. Embassy Olympics	7862	6.6 ± 32.6	(5.5, 7.6)	0.936	<0.001*	0.202 (small)
Test 5	Nongzhanguan Dongsihuan	7906	-5.4 ± 20.5	(-6.1, -4.8)	0.973	<0.001*	0.264 (small)
Test 6	Nongzhanguan Dongsi	8129	3.4 ± 24.9	(2.7, 4.2)	0.96	<0.001*	0.138 (negligible)
Test 7	Nongzhanguan Olympics	7765	0.6 ± 24.8	(-0.2, 1.4)	0.96	0.026	0.025 (negligible)
Test 8	Dongsihuan Dongsi	8004	8.4 ± 32.1	(7.4, 9.4)	0.934	<0.001*	0.261 (small)
Test 9	Dongsihuan Olympics	7642	5.4 ± 28.8	(4.4, 6.3)	0.946	<0.001*	0.186 (negligible)
Test 10	Dongsi Olympics	7882	-2.3 ± 21.7	(-3, -1.6)	0.967	<0.001*	0.106 (negligible)
2015							
Test 1	U.S. Embassy Nongzhanguan	8333	-2.3 ± 27.2	(-3.1, -1.5)	0.955	<0.001*	0.085 (negligible)
Test 2	U.S. Embassy Dongsihuan	8216	-7.9 ± 30.5	(-8.7, -7)	0.95	<0.001*	0.26 (small)
Test 3	U.S. Embassy Dongsi	8461	-4.6 ± 27.5	(-5.5, -3.8)	0.955	<0.001*	0.169 (negligible)
Test 4	U.S. Embassy Olympics	8364	0.63 ± 30.4	(-0.3, 1.6)	0.94	0.06	0.021 (negligible)
Test 5	Nongzhanguan Dongsihuan	8069	-5.9 ± 20.4	(-6.5, -5.2)	0.979	<0.001*	0.287 (small)
Test 6	Nongzhanguan Dongsi	8324	-2.5 ± 20.4	(-3.1, -1.8)	0.976	<0.001*	0.12 (negligible)
Test 7	Nongzhanguan Olympics	8227	3 ± 25.4	(2.5, 3.6)	0.961	<0.001*	0.12 (negligible)
Test 8	Dongsihuan Dongsi	8207	3.3 ± 27.5	(2.4, 4.2)	0.959	<0.001*	0.12 (negligible)

Test 9	Dongsihuan Olympics	8118	9 ± 32.8	(8, 10)	0.944	<0.001*	0.275 (small)
Test 10	Dongsi Olympics	8365	5.3 ± 24.7	(4.5, 6)	0.965	<0.001*	0.214 (small)
2016							
Test 1	U.S. Embassy Nongzhanguan	8456	-3.3 ± 22.5	(-4, -2.7)	0.956	<0.001*	0.149 (negligible)
Test 2	U.S. Embassy Dongsihuan	8288	-7.4 ± 24.9	(-8.1, -6.6)	0.951	<0.001*	0.295 (small)
Test 3	U.S. Embassy Dongsi	8232	-6.8 ± 25.5	(-7.6, -6)	0.949	<0.001*	0.267 (negligible)
Test 4	U.S. Embassy Olympics	8402	-1 ± 25.7	(-1.8, -0.2)	0.941	<0.001*	0.038 (negligible)
Test 5	Nongzhanguan Dongsihuan	8235	-4.1 ± 14.8	(-4.6, -3.7)	0.983	<0.001*	0.279 (small)
Test 6	Nongzhanguan Dongsi	8193	-3.6 ± 17.4	(-4.2, -3.1)	0.977	<0.001*	0.209 (small)
Test 7	Nongzhanguan Olympics	8358	2.4 ± 19.8	(1.8, 3)	0.966	<0.001*	0.12 (negligible)
Test 8	Dongsihuan Dongsi	8027	0.6 ± 22.3	(-0.1, 1.3)	0.962	0.024	0.025 (negligible)
Test 9	Dongsihuan Olympics	8186	6.6 ± 23.8	(5.8, 7.3)	0.955	<0.001*	0.276 (small)
Test 10	Dongsi Olympics	8129	5.9 ± 19.7	(5.3, 6.6)	0.97	<0.001*	0.302 (small)

N: number of pairs.
SD: standard deviation.
[a] Alternative hypothesis was two-sided.
*p-value<0.005 (Bonferroni correction).

Table 7. Comparison results of daily $PM_{2.5}$ percentages between the U.S. Embassy and the MEP sites.

		N	Mean ± SD	99.5% CI of the Difference	Correlation Coefficient	Paired t-test [a]	Cohen's d Effect Size
2015							
Test 1	U.S. Embassy Nongzhanguan	345	0.3 ± 7.6	(-0.9, 1.4)	0.979	0.5	0.036 (negligible)
Test 2	U.S. Embassy Dongsihuan	338	-1.9 ± 8	(-3.1, -0.7)	0.977	<0.001*	0.236 (small)
Test 3	U.S. Embassy Dongsi	354	-1.6 ± 7.7	(-2.8, -0.4)	0.978	<0.001*	0.207 (small)
Test 4	U.S. Embassy Olympics	348	-0.02 ± 7.2	(-1.1, 1.1)	0.981	0.966	0.002 (negligible)
Test 5	Nongzhanguan Dongsihuan	331	-2.2 ± 9.6	(-3.6, -0.7)	0.967	<0.001*	0.225 (small)
Test 6	Nongzhanguan Dongsi	347	-2 ± 8.8	(-3.3, -0.6)	0.971	<0.001*	0.225 (small)
Test 7	Nongzhanguan Olympics	341	-0.2 ± 8.7	(-1.6, 1.1)	0.972	0.623	0.027 (negligible)
Test 8	Dongsihuan Dongsi	341	0.3 ± 10.3	(-1.3, 1.9)	0.962	0.603	0.028 (negligible)
Test 9	Dongsihuan Olympics	335	2 ± 9.5	(0.5, 3.8)	0.968	<0.001*	0.212 (small)
Test 10	Dongsi Olympics	351	1.6 ± 6.8	(0.5, 2.6)	0.983	<0.001*	0.228 (small)
2016							
Test 1	U.S. Embassy Nongzhanguan	356	-0.3 ± 8.6	(-1.5, 1)	0.976	0.589	0.029 (negligible)
Test 2	U.S. Embassy Dongsihuan	345	-3 ± 9.4	(-4.4, -1.6)	0.971	<0.001*	0.322 (small)
Test 3	U.S. Embassy Dongsi	341	-2.5 ± 8.8	(-3.8, -1.2)	0.974	<0.001*	0.28 (small)
Test 4	U.S. Embassy Olympics	351	0.8 ± 8.8	(-0.5, 2.1)	0.974	0.096	0.089 (negligible)
Test 5	Nongzhanguan Dongsihuan	343	-2.6 ± 6.6	(-3.5, -1.6)	0.986	<0.001*	0.395 (small)
Test 6	Nongzhanguan Dongsi	341	-2.3 ± 8.1	(-3.5, -1.1)	0.979	<0.001*	0.285 (small)
Test 7	Nongzhanguan Olympics	350	1.1 ± 8.1	(-0.1, 2.2)	0.979	0.014	0.132 (negligible)

N: number of pairs. SD: standard deviation. [a] Alternative hypothesis was two-sided. Tests in 2014 and Test 8, 9 and 10 in 2016 were excluded due to violation of the data validity requirement. Bonferroni correction: *p-value<0.005 in 2015; *p-value<0.007 in 2016.

Table 8. Summary of comparison results of hourly $PM_{2.5}$ concentrations and daily $PM_{2.5}$ percentages between the U.S. Embassy and the MEP sites.

	Number of Tests	Number of Tests with a Significant Result	Number of Tests with a Non-significant Result
Comparing hourly $PM_{2.5}$ concentrations			
Between U.S. Embassy and 4 MEP sites	12	10 (83.3%)	2 (16.7%)
Between 4 MEP sites	18	16 (88.9%)	2 (11.1%)
Comparing daily $PM_{2.5}$ percentages			
Between U.S. Embassy and 4 MEP sites	8 [a]	4 (50%)	4 (50%)
Between 4 MEP sites	9 [b]	6 (66.7%)	3 (33.3%)

Effect sizes in all statistical tests were either small or negligible.

[a] 4 statistical tests were excluded in 2014 due to violation of data validity requirement.

[b] 6 statistical tests in 2014 and 3 statistical tests in 2016 were excluded due to violation of data validity requirement.

Table 9. Comparison results of $PM_{2.5}$ real-time concentrations between the U.S. Embassy and the MEP sites.

		N	Mean ± SD	99.2% CI of the Difference	Correlation Coefficient	Paired t-test [a]	Cohen's d Effect Size
Test 1	U.S. Embassy Nongzhanguan	133	-2.3 ± 7.3	(-4, -0.5)	0.994	<0.001*	0.307 (small)
Test 2	U.S. Embassy Dongsi	136	-2.2 ± 11.4	(-4.8, 0.5)	0.984	0.03	0.189 (negligible)
Test 3	U.S. Embassy Olympics	137	0.4 ± 12.7	(-2.5, 3.3)	0.981	0.727	0.03 (negligible)
Test 4	Nongzhanguan Dongsi	123	-0.2 ± 10.6	(-2.8, 2.4)	0.987	0.866	0.015 (negligible)
Test 5	Nongzhanguan Olympics	124	2.4 ± 13.2	(-0.8, 5.5)	0.98	0.049	0.179 (negligible)
Test 6	Dongsi Olympics	128	2.5 ± 11	(-0.1,5.1)	0.986	0.012	0.226 (small)

N: number of pairs

SD: standard deviation

[a] Alternative hypothesis was two-sided.

*p-value<0.008 (Bonferroni correction)

Table 10. Number of valid days by month and by year in 2008 and 2017

Year	Jan.	Feb.	Mar.	Apr.	May	June	July	Aug.	Sep.	Oct.	Nov.	Dec.	Total
2008	**0**	**0**	**0**	**22**	31	30	**19**	26	27	31	**5**	**0**	191
2009	**0**	**9**	28	24	**21**	23	29	29	30	29	27	24	273
2010	26	28	29	30	31	**21**	31	28	**19**	31	27	31	332
2011	27	28	26	**18**	29	30	31	**23**	30	**23**	30	31	326
2012	27	29	31	30	28	30	28	24	30	31	29	25	342
2013	31	28	31	30	29	30	31	31	30	31	30	29	361
2014	31	28	31	30	31	29	31	30	30	31	29	29	360
2015	31	28	29	29	31	29	31	31	30	31	30	30	360
2016	31	29	31	30	31	30	31	31	30	31	30	31	366
2017	30	28	31	30	31	30	**0**	**0**	**0**	**0**	**0**	**0**	180

The bold is to illustrate that the number of valid days within the month is less than the data validity requirements.

Table 11. Summary statistics of ranges of $PM_{2.5}$ concentrations and $PM_{2.5}$ percentages in 2008 - 2017

	$PM_{2.5}$ Concentration Ranges N (%)			**$PM_{2.5}$ Percentage Ranges N (%)**		
Year	[0, 35]	(35, 150]	>150	0	(0, 100)	100
2008	34 (17.8%)	133 (69.6%)	24 (12.6%)	3 (1.6%)	97 (50.8%)	91 (47.6%)
2009	40 (14.7%)	186 (68.1%)	47 (17.2%)	9 (3.3%)	129 (47.3%)	135 (49.5%)
2010	55 (16.6%)	203 (61.1%)	74 (22.3%)	13 (3.9%)	160 (48.2%)	159 (47.9%)
2011	78 (23.9%)	176 (54.0%)	72 (22.1%)	25 (7.7%)	166 (50.9%)	135 (41.4%)
2012	86 (25.1%)	195 (57.0%)	61 (17.8%)	23 (6.7%)	180 (52.6%)	139 (40.6%)
2013	63 (17.5%)	226 (62.6%)	72 (19.9%)	12 (3.3%)	196 (54.3%)	153 (42.4%)
2014	77 (21.4%)	219 (60.8%)	64 (17.8%)	17 (4.7%)	184 (51.1%)	159 (44.2%)
2015	98 (27.2%)	210 (58.3%)	52 (14.4%)	35 (9.7%)	197 (54.7%)	128 (35.6%)
2016	120 (32.8%)	207 (56.6%)	39 (10.7%)	39 (10.7%)	210 (57.4%)	117 (32.0%)
2017	65 (36.1%)	99 (55.0%)	16 (8.9%)	21 (11.7%)	115 (63.9%)	44 (24.4%)

N: number of ranges
%: percentage of ranges

Table 12. Summary statistics of $PM_{2.5}$ concentrations ($\mu g.m^{-3}$) and $PM_{2.5}$ percentage (%) in 2012 - 2016

Year	N	$PM_{2.5}$ Concentrations ($\mu g.m^{-3}$) Mean ± SD (Median)	$PM_{2.5}$ Percentages (%) Mean ± SD (Median)
2012	342	89.9 ± 67.0 (75.4)	67.7 ± 36.2 (83.3)
2013	361	102.0 ± 83.0 (76.5)	72.5 ± 32.7 (90.9)
2014	360	97.6 ± 80.9 (74.1)	70.8 ± 34.2 (87.5)
2015	360	82.6 ± 76.7 (58.2)	62.6 ± 37.3 (70.8)
2016	366	72.9 ± 65.4 (56.6)	60.6 ± 38.7 (75.0)

N: Number of valid days

SD: standard deviation

Table 13. Comparison results of $PM_{2.5}$ concentrations (µg.m-3) and $PM_{2.5}$ percentage (%) in 2012 – 2016

				Tukey's (HSD) Test		
	Levene Test	One-way ANOVA	η^2 Effect Size	Comparison	Adjust p-value	Decrease [a]
$PM_{2.5}$ concentrations	0.079	<0.001***	0.03 (small)	2016 - 2012	<0.001***	18.9%
				2015 -2013	<0.001***	19%
				2016 - 2013	<0.001***	28.5%
				2015 -2014	0.004**	15.4%
				2016 -2014	<0.001***	25.3%
				Dunn Test		
	Levene Test	Kruskal–Wallis Test	ε^2 Effect Size	Comparison	Adjust p-value	Decrease [b]
$PM_{2.5}$ percentages	<0.001***	<0.001***	0.015 (small)	2015 -2013	0.008**	22.1%
				2016 - 2013	<0.001***	17.5%
				2015 -2014	0.02*	19.1%
				2016 -2014	<0.001***	14.3%

[a] Decrease in average $PM_{2.5}$ concentrations ($\mu g.m^{-3}$).
[b] Decrease in median $PM_{2.5}$ percentages (%).
*p-value<0.05, **p-value<0.01, ***p-value<0.001

Table 14. Summary statistics of $PM_{2.5}$ concentrations ($\mu g.m^{-3}$) and $PM_{2.5}$ percentages (%) by season in 2012 – 2016

	Spring		Summer		Autumn		Winter	
	N	Mean ± SD (Median)	N	Mean ± SD (Median)	N	Mean ± SD (Median)	N	Mean ± SD (Median)
$PM_{2.5}$ concentrations ($\mu g.m^{-3}$)								
2012	89	91.6 ± 62.7 (79.9)	82	86.5 ± 52.9 (78.5)	90	80.4 ± 63.1 (59.5)	81	102.2 ± 85.2 (76.0)
2013	90	91.8 ± 69.5 (70.3)	92	80.5 ± 47.1 (70.1)	91	96.2 ± 72.4 (71.5)	88	140.8 ± 116.9 (109.6)
2014	92	92.7 ± 60.8 (82.1)	90	70.6 ± 45.7 (60.1)	90	106 ± 90.6 (75.9)	88	121.9 ± 105.7 (97.8)
2015	89	76.3 ± 48.9 (63.5)	91	51.2 ± 30.7 (45.6)	91	81.2 ± 83.3 (51.1)	89	122.5 ± 104.2 (90.9)
2016	92	71.7 ± 60.0 (62.5)	92	52.4 ± 29.1 (50.1)	91	79.8 ± 62.0 (68.2)	91	87.9 ± 91.0 (46.5)
$PM_{2.5}$ percentages (%)								
2012	89	72.2 ± 33.9 (91.7)	82	76.0 ± 34.4 (100)	90	59.0 ± 36.6 (62.5)	81	64.0 ± 37.9 (75.0)
2013	90	74.2 ± 32.9 (91.7)	92	76.4 ± 30.4 (95.8)	91	69.1 ± 34.2 (83.3)	88	70.3 ± 33.4 (85.4)
2014	92	77.4 ± 31.0 (100)	90	67.9 ± 35.0 (84.5)	90	70.9 ± 34.5 (91.7)	88	66.6 ± 35.8 (83.3)
2015	89	71.2 ± 33.0 (87.5)	91	57.8 ± 37.5 (62.5)	91	56.8 ± 41.1 (70.8)	89	64.7 ± 35.6 (69.6)
2016	92	64.0 ± 38.5 (83.3)	92	63.7 ± 36.7 (79.2)	91	63.9 ± 39.2 (83.3)	91	50.7 ± 39.2 (45.8)

Spring has 92 days; Summer has 92; Autumn has 91 days; Winter has 91 days for the years 2008, 2012 and 2016, and 90 days for other years.

N: Number of valid days

SD: standard deviation

Table 15. Comparisons results of $PM_{2.5}$ concentrations ($\mu g.m^{-3}$) and $PM_{2.5}$ percentage (%) by season in 2012 - 2016

Season	Levene Test	One-way ANOVA	η^2 Effect Size	Tukey's (HSD) Test Comparison	Adjust p-value	Decrease [a]
$PM_{2.5}$ concentrations ($\mu g.m^{-3}$)						
Spring	0.446	0.01*	0.029 (small)	2016 - 2014	0.011*	22.7%
Summer	0.725	< 0.001***	0.089 (medium)	2015 - 2012	<0.001***	40.8%
				2016 - 2012	<0.001***	39.4%
				2015 - 2013	<0.001***	36.4%
				2016 - 2013	<0.001***	34.9%
				2015 - 2014	0.022*	27.5%
Autumn	0.112	0.01*	0.029 (small)	2015 - 2013	0.036*	15.6%
				2015 - 2014	0.024*	23.4%
Winter	0.446	< 0.001***	0.043 (small)	2016 - 2013	<0.001***	37.6%
				2016 - 2014	0.032*	27.9%
				2016 - 2015	0.02*	28.2%

Season	Levene Test	Kruskal–Wallis Test	ε^2Effect Size	Dunn Test Comparison	Adjust p-value	Decrease [b]
$PM_{2.5}$ percentages (%)						
Spring	0.158	0.093	0.017 (small)			
Summer	0.032*	<0.001***	0.044 (small)	2015 -2012	0.003**	37.5%
				2016 -2012	0.036*	20.8%
				2015 - 2013	0.007*	34.8%
Autumn	0.076	0.039*	0.022 (small)			
Winter	0.24	0.014**	0.028 (small)	2016 -2013	0.011*	46.4%

[a] Decrease in average $PM_{2.5}$ concentrations ($\mu g.m^{-3}$).

[b] Decrease in median $PM_{2.5}$ percentages (%).

*p-value<0.05, **p-value<0.01, ***p-value<0.001

APPENDIX B

APPENDIX B

FIGURES

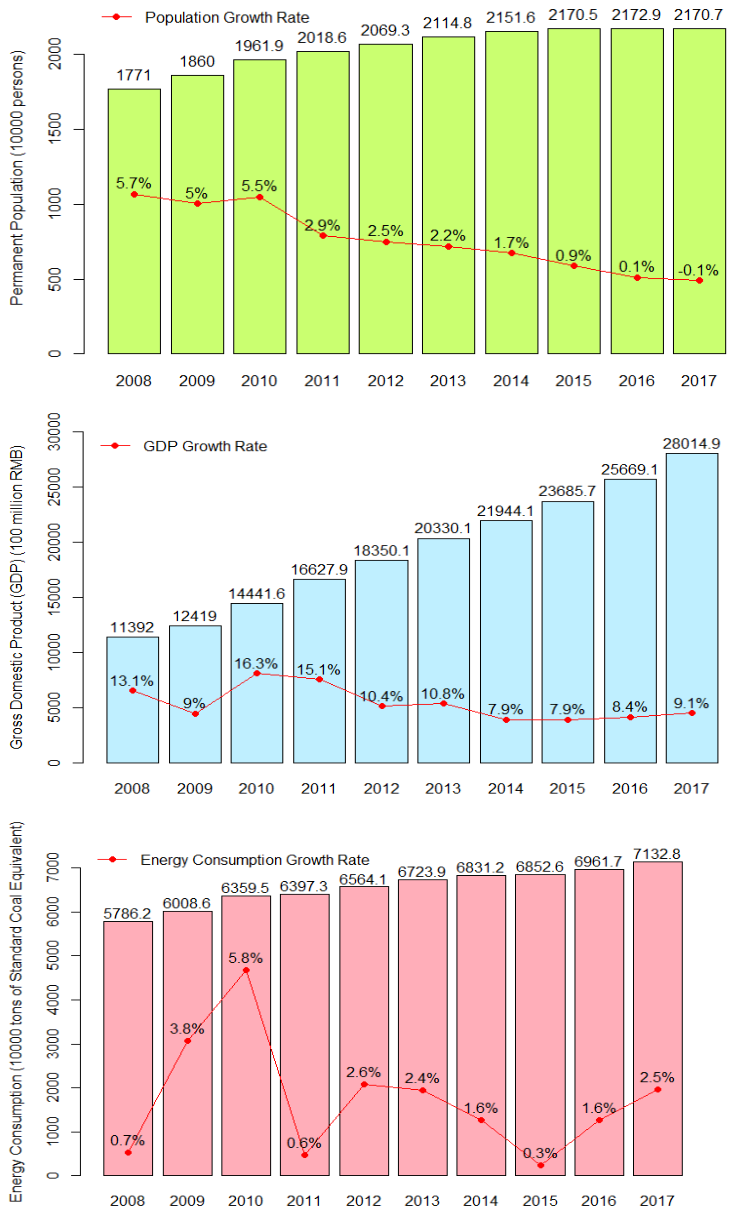


Figure 1. Trends of population, GDP and energy consumption in Beijing in 2008 - 2017

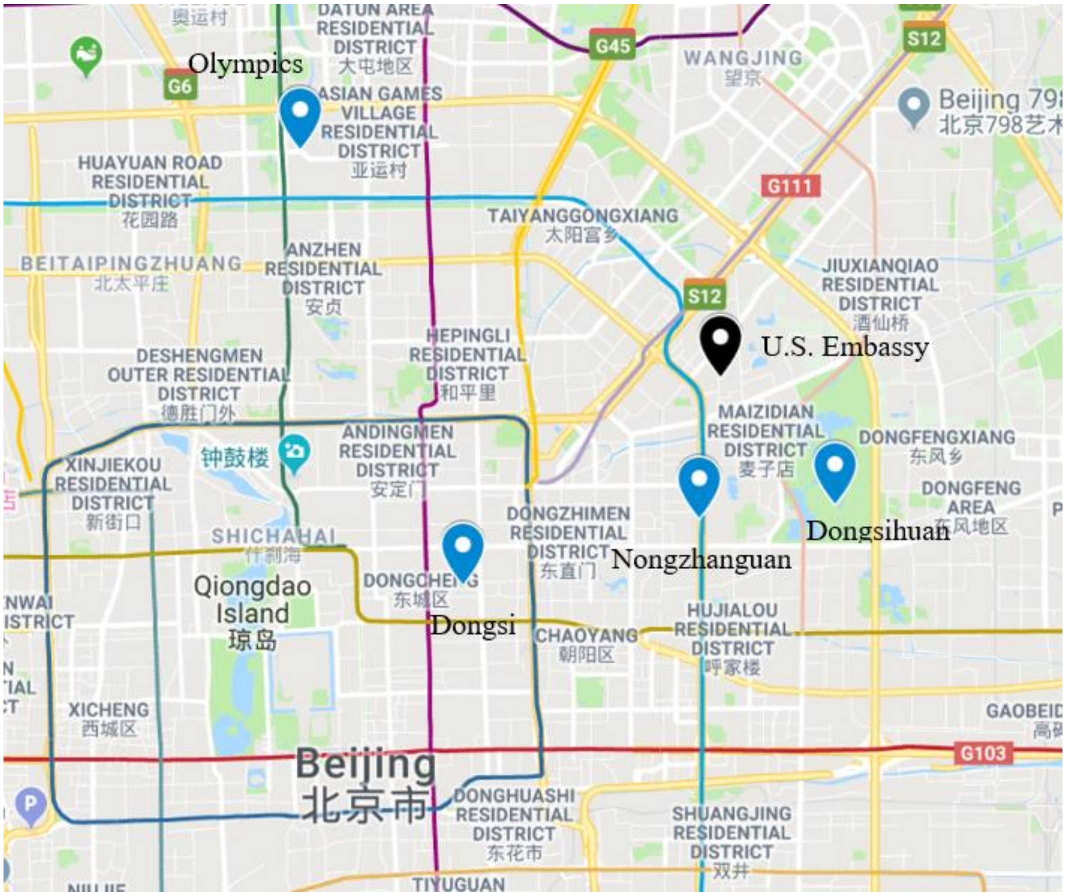
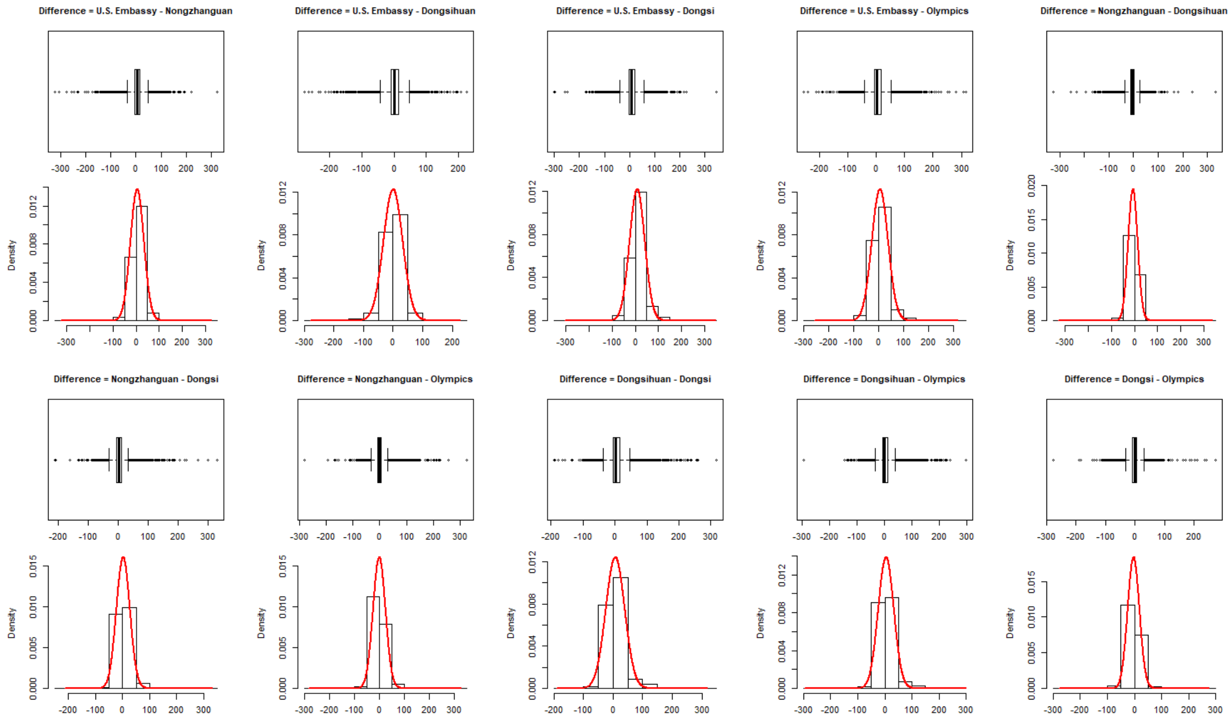
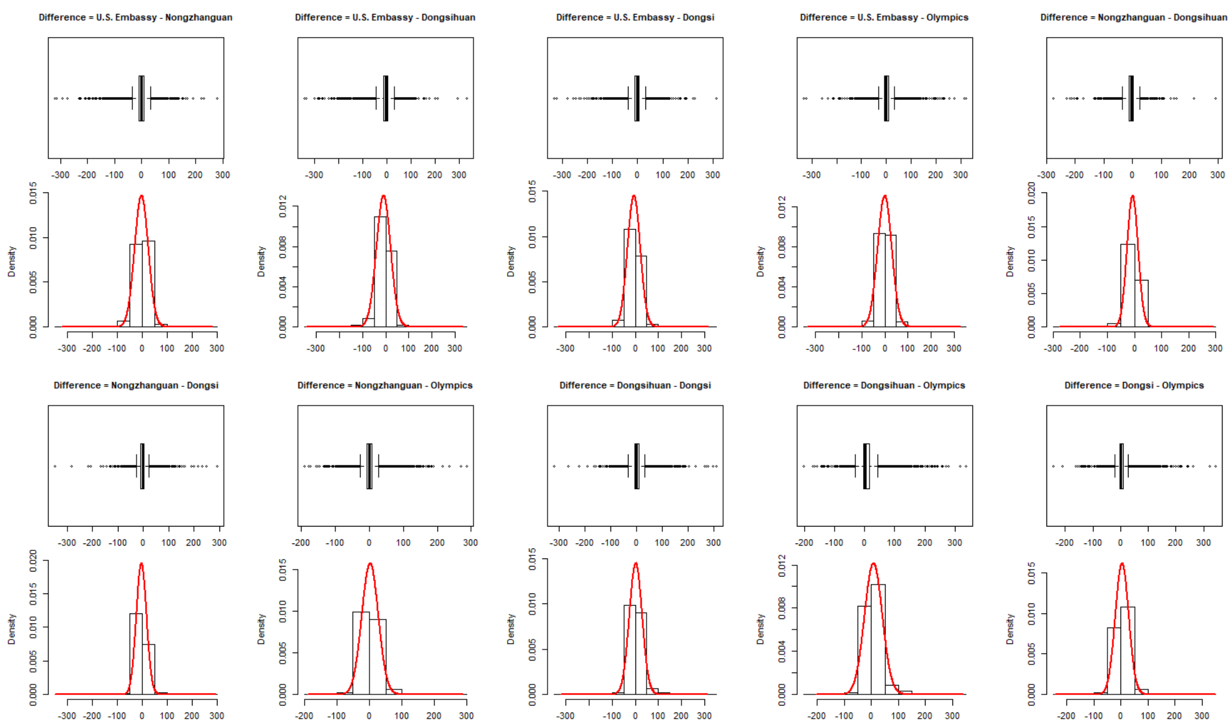


Figure 2. The geographical location of the U.S. Embassy and the MEP sites.

2014


Difference = U.S. Embassy - Nongzhanguan
Difference = U.S. Embassy - Dongsihuan
Difference = U.S. Embassy - Dongsi
Difference = U.S. Embassy - Olympics
Difference = Nongzhanguan - Dongsihuan
Difference = Nongzhanguan - Dongsi
Difference = Nongzhanguan - Olympics
Difference = Dongsihuan - Dongsi
Difference = Dongsihuan - Olympics
Difference = Dongsi - Olympics

2015


Difference = U.S. Embassy - Nongzhanguan
Difference = U.S. Embassy - Dongsihuan
Difference = U.S. Embassy - Dongsi
Difference = U.S. Embassy - Olympics
Difference = Nongzhanguan - Dongsihuan
Difference = Nongzhanguan - Dongsi
Difference = Nongzhanguan - Olympics
Difference = Dongsihuan - Dongsi
Difference = Dongsihuan - Olympics
Difference = Dongsi - Olympics

2016

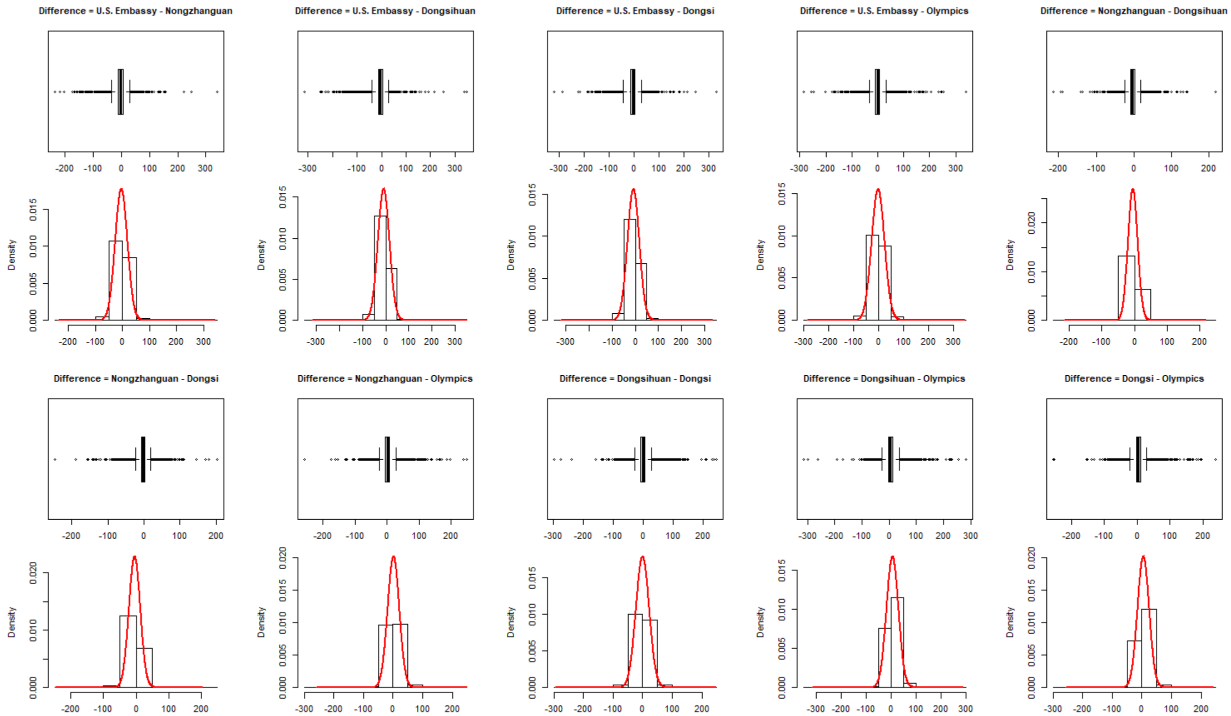
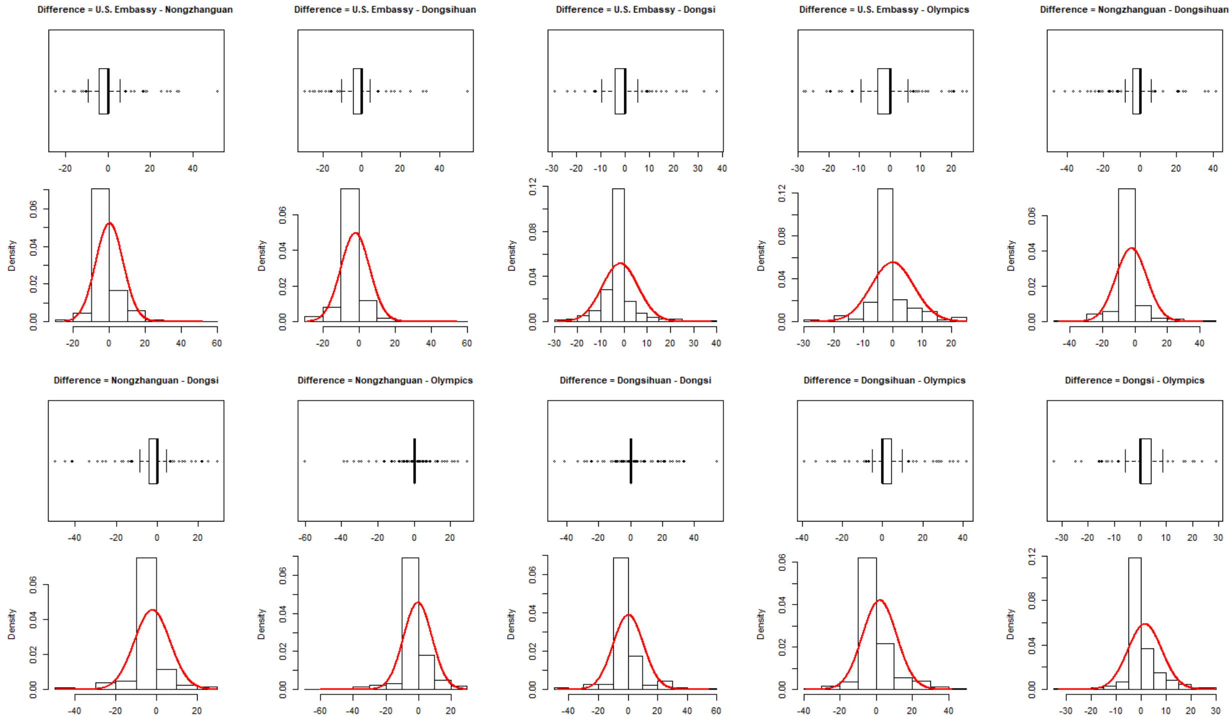


Figure 3. Distribution of difference of hourly $PM_{2.5}$ concentrations between sites in 2014 - 2016

2015



2016

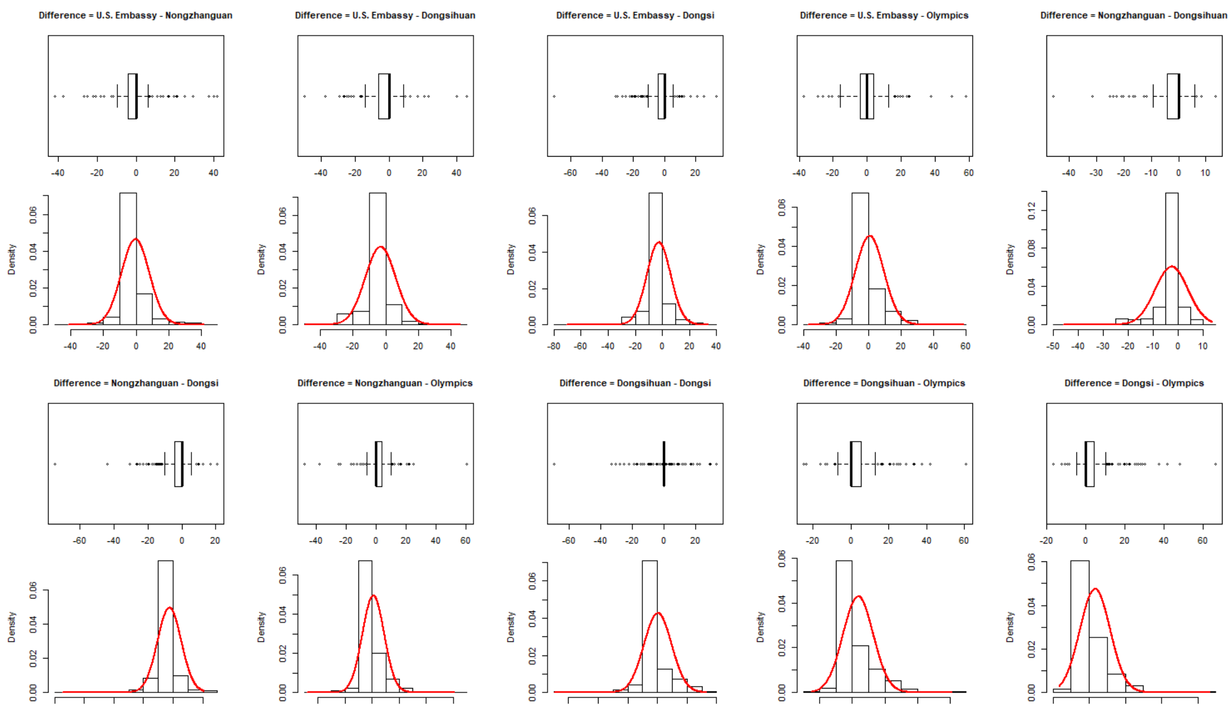


Figure 4. Distribution of difference of daily $PM_{2.5}$ percentages between sites in 2015 - 2016

Figure 5. National Urban Air Quality Real-Time Publishing Platform

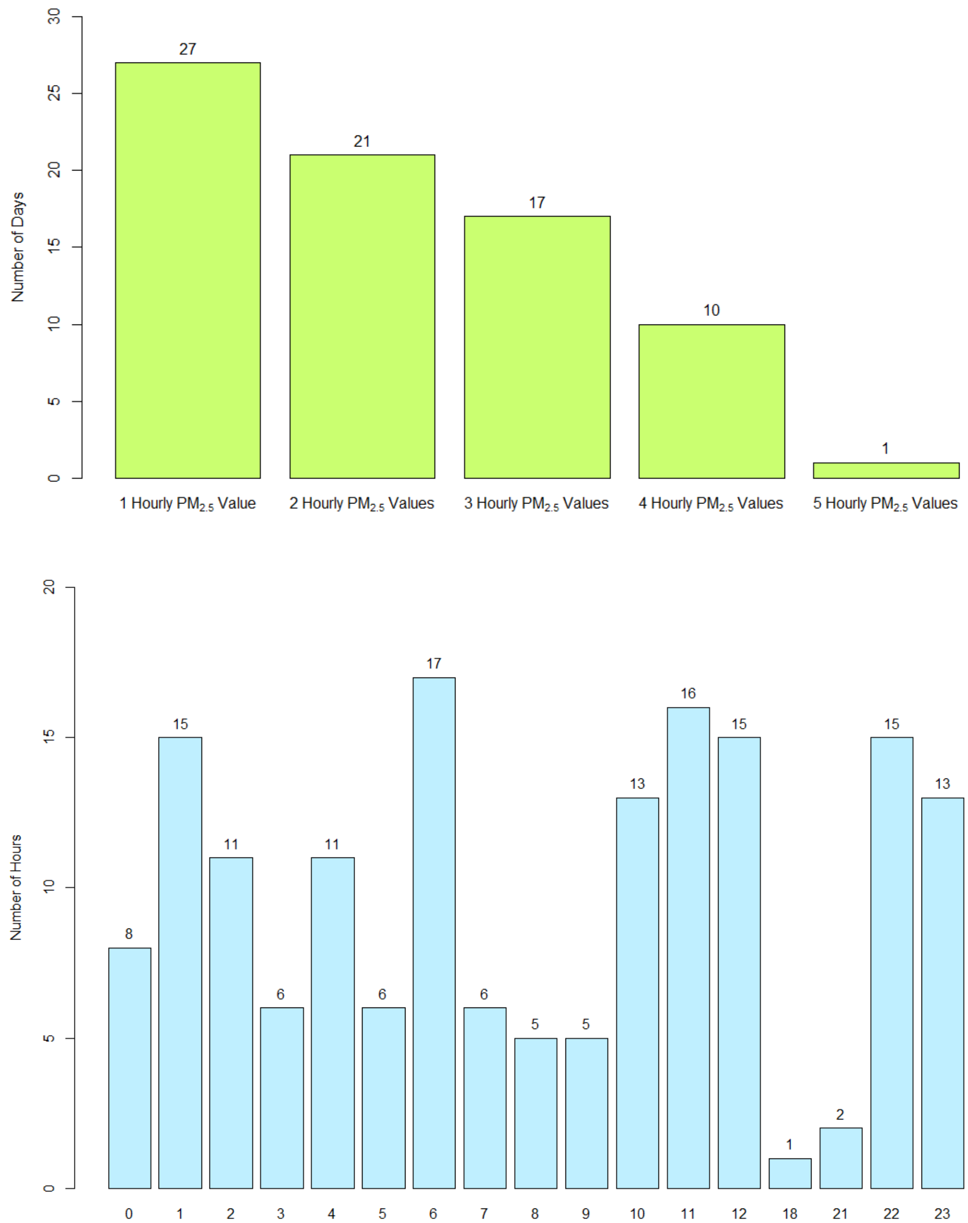


Figure 6. $PM_{2.5}$ real-time concentrations collection description

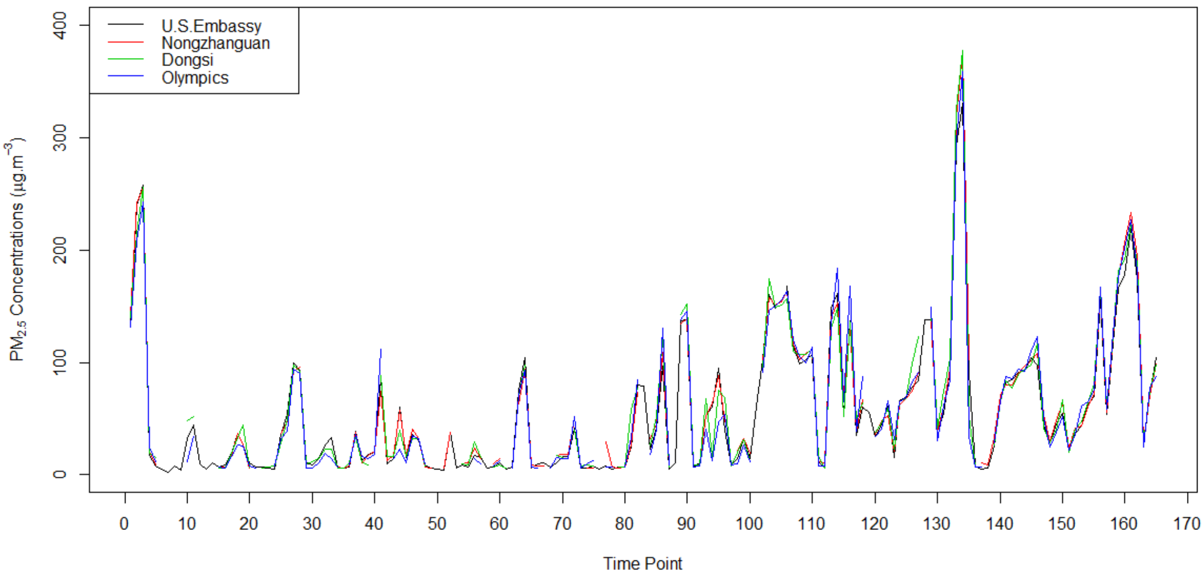


Figure 7. Trends of $PM_{2.5}$ real-time concentrations of the U.S. Embassy and the MEP sites

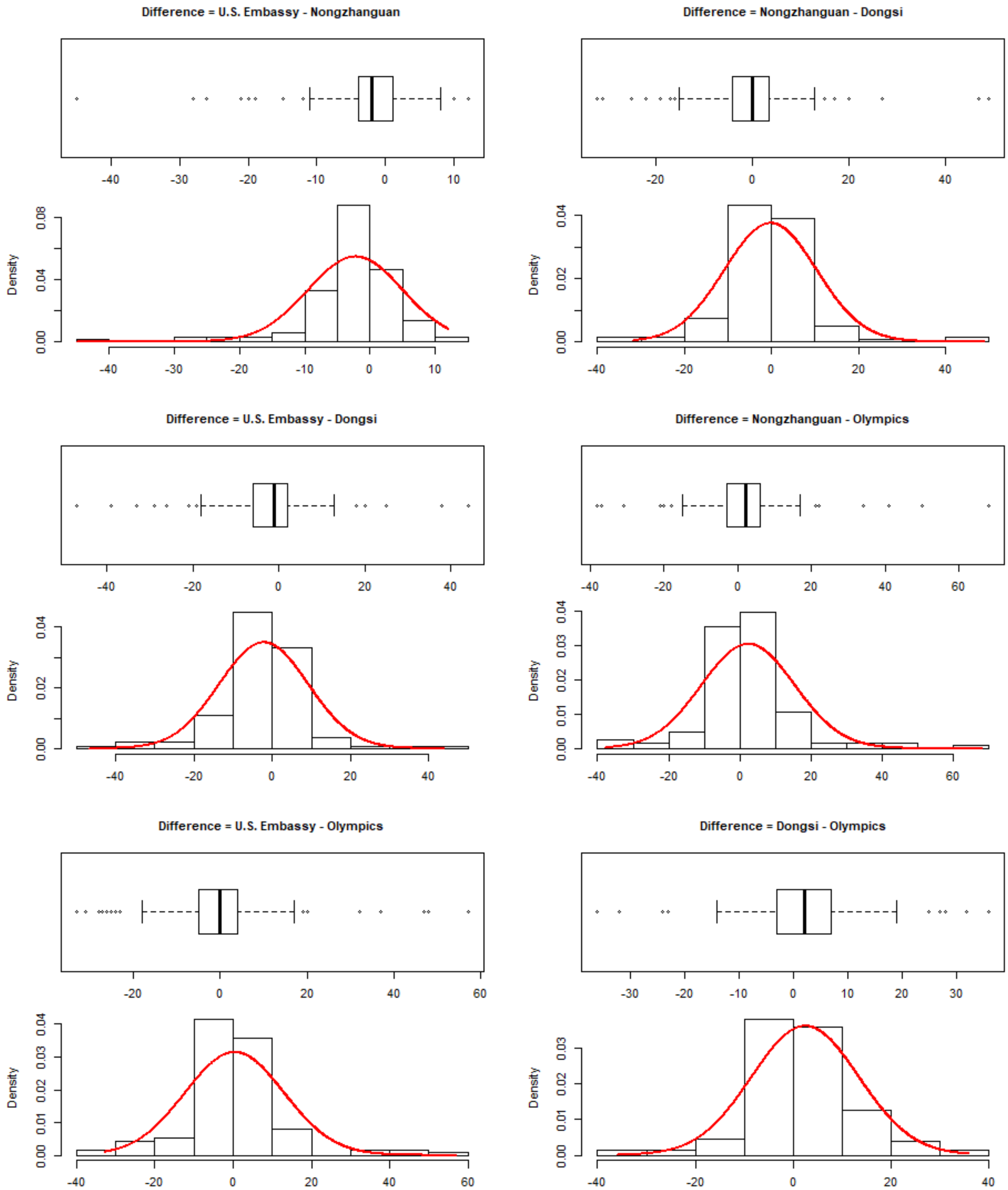


Figure 8. Distribution of difference of the $PM_{2.5}$ real-time concentrations between sites

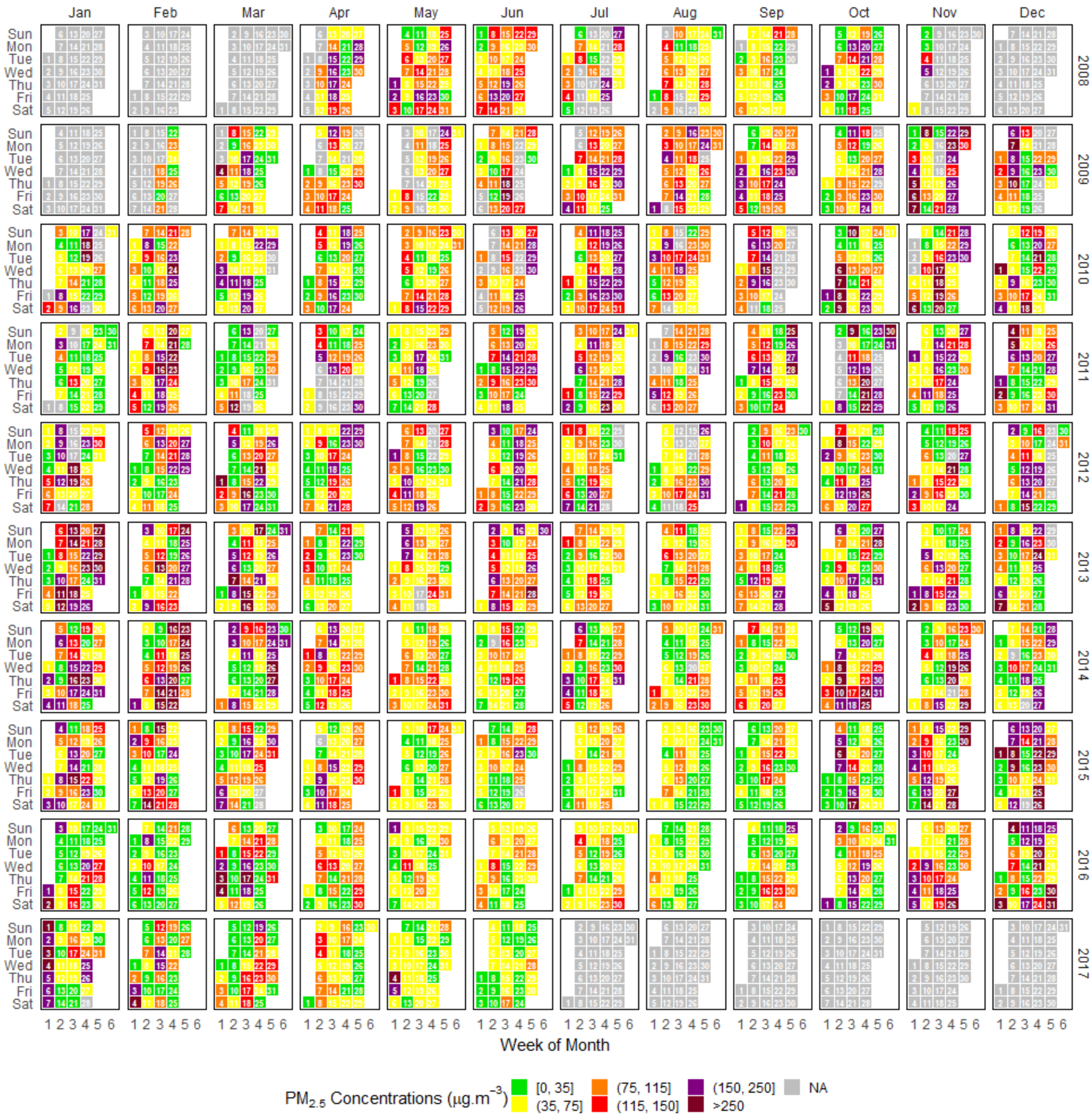


Figure 9. Calendar view of $PM_{2.5}$ concentrations in 2008 - 2017

Figure 10. Calendar view of $PM_{2.5}$ percentages in 2008 - 2017

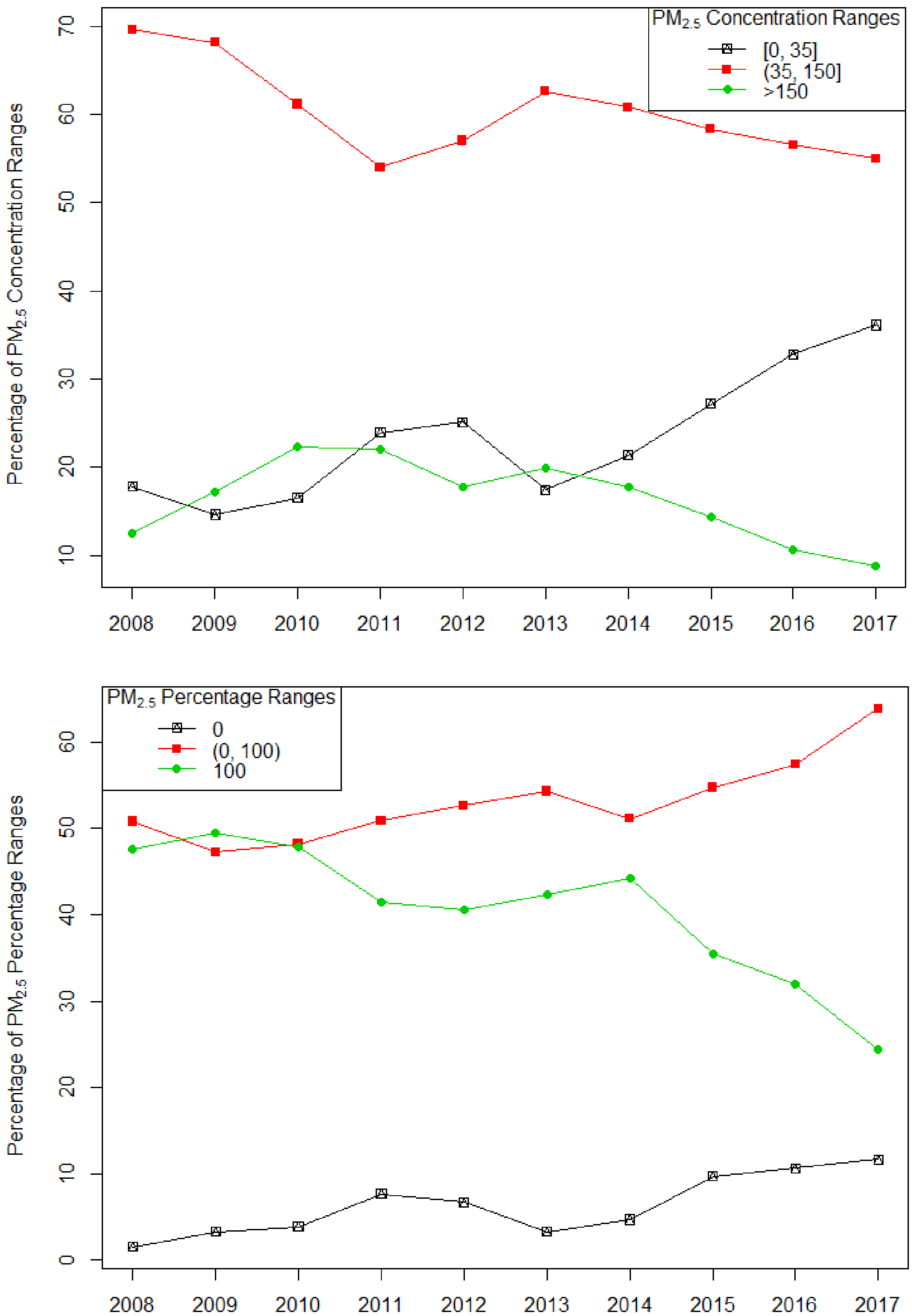
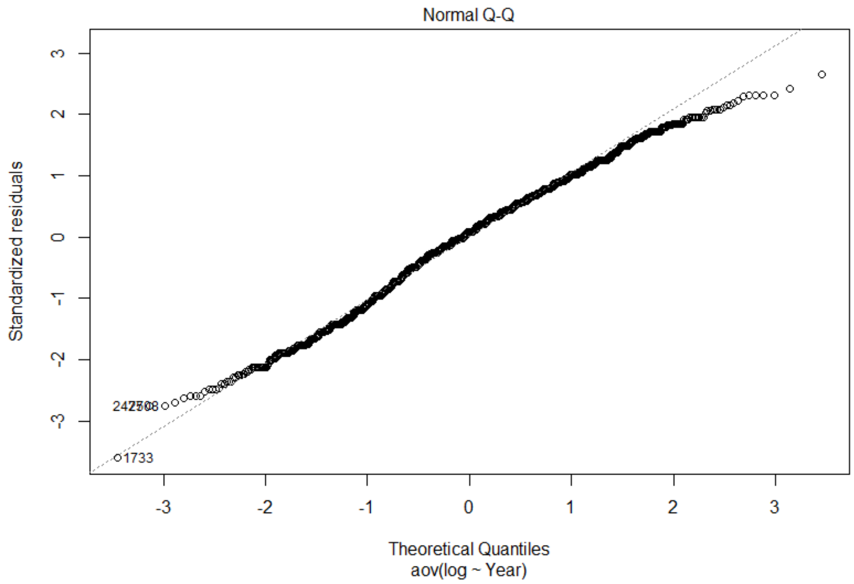


Figure 11. Annual trends of ranges of $PM_{2.5}$ concentration and $PM_{2.5}$ percentage in 2008 - 2017



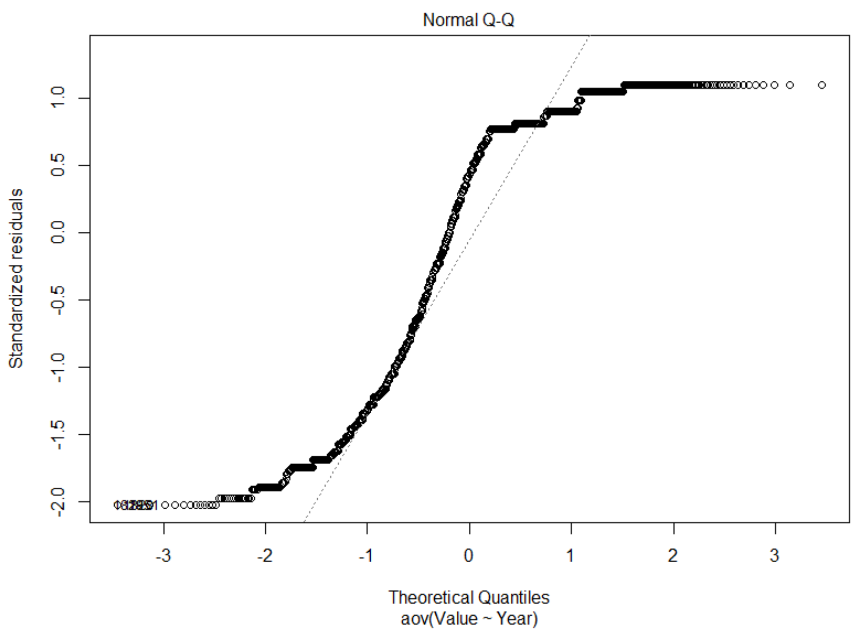


Figure 12. Normality check of $PM_{2.5}$ concentrations (top) and $PM_{2.5}$ percentages (bottom)

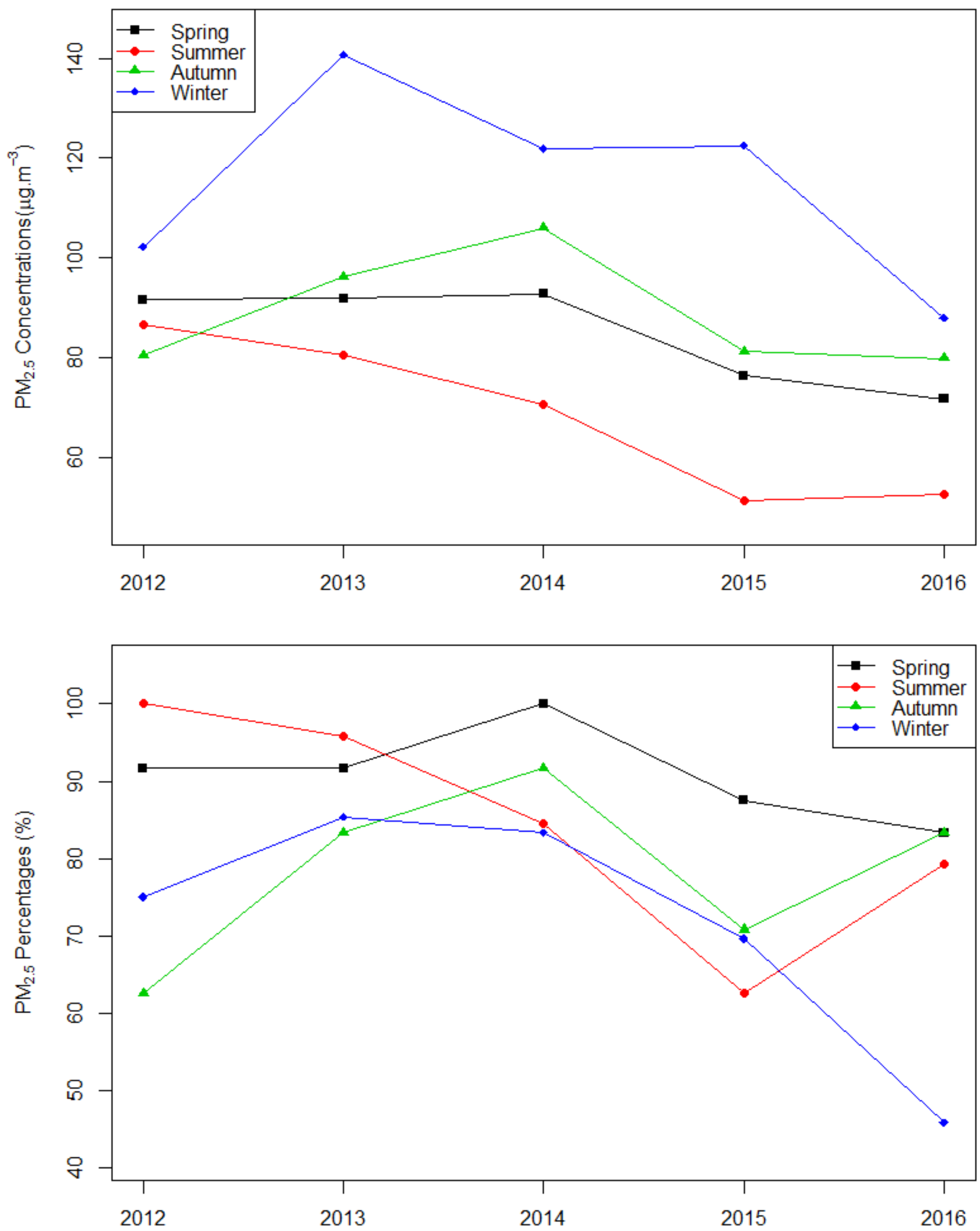


Figure 13. Seasonal trends of $PM_{2.5}$ concentrations and $PM_{2.5}$ percentages in 2012 - 2016

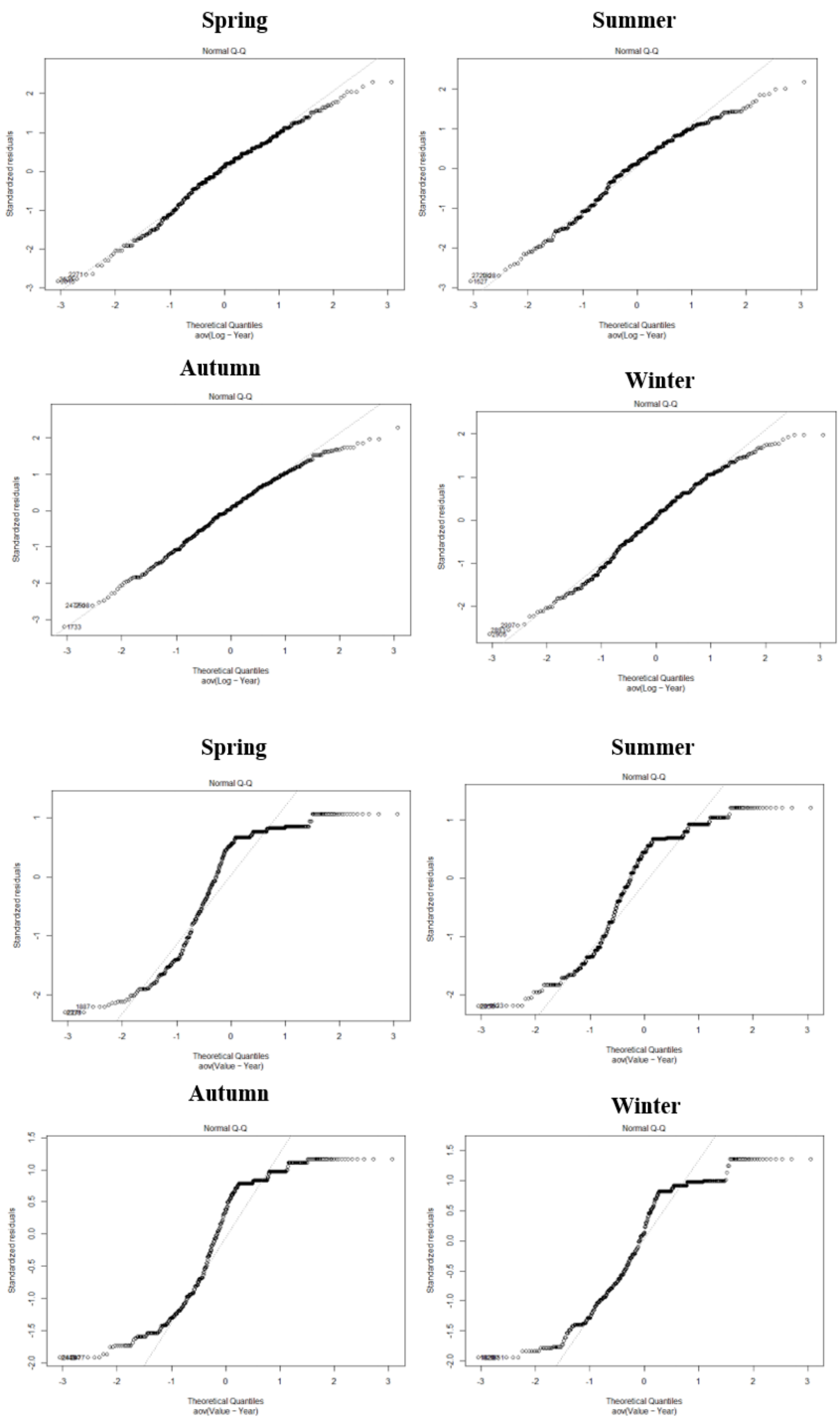


Figure 14. Normality check of $PM_{2.5}$ concentrations (top) and $PM_{2.5}$ percentages (bottom) by season

BIOGRAPHICAL SKETCH

Huimin Li was born in Hunan, China on September 25, 1990. She completed her master with a major in Statistics in May 2019 from the University of Texas Rio Grande Valley. She graduated from Xiangtan University in June 2013 with a Bachelor of Business Administration.

During her graduate education, she was the evaluation coordinator in the South Texas Early Prevention Study (STEPS) Pre-K Project from December 2017 to May 2019. She worked along with Co-PI Dr. Xiaohui Wang, oversaw all field activities related to data management and data evaluation. The duties included designing experiment, survey, website, and database; setting logistics of data collection; creating data collection schedules for all data collectors; conducting collected data quality check; performing preliminary statistical analysis using R for reports and publications; leading one Undergraduate Research Assistant. She also assisted her advisor to perform data cleaning, data analyses and write the preliminary report on several projects through her graduate study period.

Huimin concentrated on applied statistics and performed a lot of data evaluation by statistical software and statistical methods. She also made several presentations on the area of statistics including applied statistics, experimental design and statistical computing. With the graduate degree, she plans to pursue a Ph.D. degree in statistics and become more professional in this interesting field. She can be contacted via email huimin.li01@utrgv.edu.