Time Series Analysis of Fine Particulates Matter (PM2.5) in Chaoyang District of Beijing

A thesis submitted in partial fulfillment of the requirements for the degree of Master of Science in Environmental Science and Health

by

Yang Peng

Dr. Wei Yang/Thesis Advisor

August, 2013
We recommend that the thesis prepared under our supervision by

Yang Peng

etitled

Time Series Analysis of Fine Particulates Matter (PM2.5) in Chaoyang District of Beijing

be accepted in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Wei Yang Ph.D., Advisor

Vera Samburova Ph.D., Committee Member

Julie Smith-Gagen Ph.D., Graduate School Representative

Marsha H. Read, Ph. D., Dean, Graduate School

August, 2013
ABSTRACT

Air pollution in Beijing is becoming hard to ignore, however, there are no long-term studies examining the particle pollution such as PM 2.5 concentration. This study describes current and predicts future PM 2.5 pollution for a three-year period, 1st January 2010 to 31st December 2012, from the U.S. Embassy monitoring site in Chaoyang, Beijing. We calculated the 24-hour and annual PM2.5 concentrations under EPA method and using the Box-Jenkins method to build a SARIMA model (Seasonal Autoregressive Integrated Moving Average). The results showed the PM2.5 concentration decreased from 2010 to 2012. The 24-hour PM2.5 concentration was 294μg/m³, ranged from 265.2μg/m³ to 318.5μg/m³. And the annual PM2.5 concentration was 94.5μg/m³ which yearly range from 89.5μg/m³ to 99.0μg/m³ and the quarterly value from 73.1μg/m³ to 108.6μg/m³. The 24-hours and annual concentration is about 7 to 8 times higher than the US National Ambient Air Quality Standard (NAAQS). There is an urgent need for long-term studies to guide policy makers to improve the air quality in Beijing.

Keywords: Beijing; PM2.5; Air pollution;
ACKNOWLEDGMENTS

I would like to give my thanks to all those that have given their support, understanding, and help during the creation of this work. First, I would like to thank my family. Secondly, thanks to Mitsuo Yoshida for proving technical help on the twitter historical data search engine and website for data capture, David G. Streets from the Argonne National Laboratory whom provided helpful suggestions. Lastly, thanks are given to my committee, Dr. Wei Yang, Dr. Julie Gagen and Dr. Vera Samburova for their unconditional help not only on this thesis but also on my graduate school study.
TABLE OF CONTENTS

ABSTRACT .................................................................................................................. i
ACKNOWLEDGMENTS .............................................................................................. ii
LIST OF FIGURES ..................................................................................................... iv
LIST OF TABLES ....................................................................................................... v

Introduction ................................................................................................................... 1

Materials and Methods ............................................................................................... 8

  Materials: Data and Monitor ....................................................................................... 8
  Methods PM2.5 Calculation Methods ......................................................................... 9
  Method :SARIMA ....................................................................................................... 10
    1. Data Preparation .................................................................................................. 10
    2. Model Estimation ................................................................................................ 11
    3. Forecast .............................................................................................................. 12

Results .......................................................................................................................... 13

  SARIMA modelling .................................................................................................... 14
    1. Data Preparation .................................................................................................. 14
    2. Model Identification and Estimation .................................................................... 20
    3. Prediction ............................................................................................................ 22

Conclusion .................................................................................................................... 27

Appendix ...................................................................................................................... 29

References .................................................................................................................... 32
LIST OF FIGURES

Figure 1. locations of U.S. Embassy and BJMEMC air monitoring sites ........................... 5
Figure 2. the study area located in Chaoyang district of Beijing, China ........................... 5
Figure 3. PM2.5 daily averages (of hourly averages per day) US Embassy, Beijing 2010-2012 ......................................................................................................... 14
Figure 4. PM2.5 daily averages (of hourly averages per day) US Embassy, Beijing 2010 ................................................................................................................... 15
Figure 5. PM2.5 daily averages (of hourly averages per day) US Embassy, Beijing 2011 ................................................................................................................... 15
Figure 6. PM2.5 daily averages (of hourly averages per day) US Embassy, Beijing 2012. ..................................................................................................................... 16
Figure 7. Time series plots of PM2.5 monthly and quarterly average US Embassy, Beijing 2012 .............................................................................................................. 16
Figure 8. PM2.5 Seasonal trends analysis by Linear Model 2010-2012 US Embassy, Beijing 2010-2012 ................................................................................................. 17
Figure 9. PM2.5 Trends and Correlation Analysis 2010-2012 US Embassy, Beijing 2012 ..................................................................................................................... 18
Figure 10. Trend and correlation analysis for PM2.5 after log transformation .......... 19
Figure 11. PM 2.5 plots after seasonal differencing, US Embassy, Beijing 2012........ 20
Figure 12. Prediction using SARIMA (1,0,2)(1,1,1)12 Model ......................................... 22
Figure 13. 2013 Beijing PM2.5 concentration (1st Jan 2010-31st Mar2013) ............ 22
Figure 14. PM2.5 hourly concentration comparison, data from BJMEMC and U.S. Embassy monitoring sites, February 2012 ......................................................... 24
LIST OF TABLES

Table 1. Historical studies of PM2.5 in China .......................................................... 3
Table 2. History of the NAAQS for Particulate Matter During the Period 1997-2012 by EPA .......................................................... 6
Table 3. History of the AQQS for PM2.5 during the Period 1997-2012 by MEP, China .......................................................... 7
Table 4. 24-hours PM2.5 Standard Design Value (Based on three year average of yearly 98th percentile 24-hr values), US Embassy, Beijing 2010-2012.............. 13
Table 5. Annual Average Design Values (μg/m³) US Embassy, Beijing 2010-2012... 13
Table 6. Results of Unit root tests ............................................................................. 18
Table 7. Parameter Estimates for SARIMA (1,0,2)(1,1,1)12 Model......................... 21
Table 8. Autocorrelation Check of Residuals: SARIMA (1,0,2)(1,1,1)12 Model....... 21
Table 9. Different monitor detection and calculation method between BJMEMC and U.S. Embassy ................................................................. 25
Table 10. Air Quality Index ...................................................................................... 29
Table 11. NAAQS of PM2.5 ................................................................................... 30
Introduction

In China, high population density, rapid urbanization and industrialization have led to an increase in emissions\textsuperscript{1}. The emissions have exacerbated outdoor air pollution, PM 2.5 specifically, in large and medium size cities resulting in reduced visibility\textsuperscript{2}. Air pollution has been linked with respiratory and cardiovascular diseases\textsuperscript{3,4}. Recently, the National Development and Reform Commission (NDRC) reported that haze shrouded nearly one-quarter of China and about 600 million people were affected\textsuperscript{3}. Compared to the southern Chinese cities, northern cities are colder and have greater heavy industry. A study from Proceedings of the National Academy of Sciences indicates (PNAS) reported on average, southern Chinese have about 5.5 years longer life expectancy than northern Chinese because northern Chinese use coal for winter heating\textsuperscript{2}. PM 2.5 air pollution is one of the major pollutants in northern China, especially in Beijing. The highest PM2.5 hourly concentration in Beijing was on January 12th, 2013 at 886\textmu/m\textsuperscript{3}, 25 times higher than the EPA’s 24-hour standard. During the first month of 2013, only 6, out of 31 days, met the EPA’s standard for PM2.5. Beijing has more than 20 million people, making this a significant area of concern.

Most of the previous studies on PM 2.5 in Beijing focused on the chemical composition, source apportionment, and seasonal characteristics. The first comprehensive PM2.5 measurement study was done in 1990. Later, with the economic growth and expansion of transportation system, the overall PM2.5 concentrations begin to increase. In 2008, the Beijing authority introduced pollution control measures designed to reduce local emissions for the Olympic Games. Recent studies show PM2.5 concentrations have dropped to levels lower than in 1989. Urban and suburban or rural PM2.5
concentration comparisons shown in Table 1 indicate that urban PM2.5 concentrations are lower than suburban and rural area of Beijing.

Furthermore, Table 1 also indicates that PM2.5 concentrations can be high during the summer and winter. The air quality in Beijing depends on the meteorological conditions and topography.

Surrounding hills obstruct pollutant movement and unfavorable meteorological conditions such as high temperatures, high humidity and low wind speed boost the photochemical formation of ozone, exacerbate the heat island effect and directly influence the PM2.5 level.

This study describes PM2.5 by using three years hourly data to calculate the 24-hour and annual PM2.5 concentrations by using EPA methods. Moreover, we predict future PM2.5 concentrations in Beijing.

<table>
<thead>
<tr>
<th>Measurement Intervals</th>
<th>PM2.5 concentration(μ/m³)</th>
<th>Monitoring site(s)</th>
<th>Time period of measurements</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hourly</td>
<td>89 (2.9-300)</td>
<td>IP (urban)</td>
<td>8/2003</td>
<td>Chan et al.(2005)</td>
</tr>
<tr>
<td>Daily</td>
<td>91 (30-159)</td>
<td>CM (urban)</td>
<td>6/1999</td>
<td>Bergin et al. (2001)</td>
</tr>
<tr>
<td></td>
<td>114 (84-163)</td>
<td>SB (suburban)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>169 (19-368)</td>
<td>IP, CM, SB (urban/suburban)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>136</td>
<td>PU (suburban)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Weekly</td>
<td>37-357</td>
<td>TU, CG (urban)</td>
<td>7/1999-2000</td>
<td>He et al. (2001)</td>
</tr>
<tr>
<td></td>
<td>33</td>
<td>BN (urban)</td>
<td>7/2002</td>
<td>Xu et al. (2005)</td>
</tr>
<tr>
<td>Seasonal (Summer)</td>
<td>93</td>
<td>BN, CS, TH, MY, PG(urb/sub)</td>
<td>Summer 2001-2003</td>
<td>Wang et al. (2005)</td>
</tr>
<tr>
<td></td>
<td>76</td>
<td>CG (urban)</td>
<td>Summer 2000</td>
<td>He et al. (2001)</td>
</tr>
<tr>
<td></td>
<td>122</td>
<td>PG (rural)</td>
<td>6/2002-7/2002</td>
<td></td>
</tr>
<tr>
<td></td>
<td>91</td>
<td>BN, CS, YH</td>
<td>6/2002-7/2002</td>
<td></td>
</tr>
</tbody>
</table>
### Table 1. Historical studies of PM2.5 in China.

<table>
<thead>
<tr>
<th>Season</th>
<th>Concentration (μg/m³)</th>
<th>Location</th>
<th>Year</th>
<th>Authors</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Seasonal</td>
<td>54.2</td>
<td>TSI(urban)</td>
<td>Spring1989</td>
<td>Chen et al(1994)</td>
</tr>
<tr>
<td></td>
<td>75.5</td>
<td></td>
<td>Summer</td>
<td></td>
</tr>
<tr>
<td></td>
<td>60.9</td>
<td></td>
<td>Fall</td>
<td></td>
</tr>
<tr>
<td></td>
<td>72.6</td>
<td></td>
<td>Winter</td>
<td></td>
</tr>
<tr>
<td></td>
<td>76.4</td>
<td>CZ(urban)</td>
<td>Spring2002</td>
<td>Duan et al. (2006)</td>
</tr>
<tr>
<td></td>
<td>89.0</td>
<td></td>
<td>Summer</td>
<td></td>
</tr>
<tr>
<td></td>
<td>79.8</td>
<td></td>
<td>Fall</td>
<td></td>
</tr>
<tr>
<td></td>
<td>122.1</td>
<td></td>
<td>Winter</td>
<td></td>
</tr>
<tr>
<td></td>
<td>54.3(10.1-191.6)</td>
<td>BN(urban)</td>
<td>Spring2010</td>
<td>Lingda et.al(2010)</td>
</tr>
<tr>
<td></td>
<td>59.1(15.8-203.7)</td>
<td></td>
<td>Summer</td>
<td></td>
</tr>
<tr>
<td></td>
<td>51.2(9.8-112.5)</td>
<td></td>
<td>Fall</td>
<td></td>
</tr>
<tr>
<td></td>
<td>57.5(11.6-219)</td>
<td></td>
<td>Winter</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Annual</th>
<th>Concentration (μg/m³)</th>
<th>Location</th>
<th>Year</th>
<th>Authors</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>64.78</td>
<td>USE(urban)</td>
<td>December 7,</td>
<td>Jin-feng et al(2013)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2010–December 6,2011</td>
<td></td>
</tr>
</tbody>
</table>

IP =Institute of Atmospheric Physics, CM =Chinese Academy of Meteorological Sciences, SB=S Southern Observational Base, PU=Peking University, TU=Tsinghua University, CG=Chegongzhuang, DS=Dong Si Environmental Protection Bureau, AP=airport, YL=Yong Le Dian, MT=Ming Tombs, CS=Capital Steel Company, YH=Yihai Garden, PG=Pinggu, MY=Miyun, BJ=Beijing (unspecified), BN=Beijing Normal University, USE=U.S. Embassy

* (Spring: March–May; Summer: June–August; Fall: September–November; Winter: January, February and December).

Many projects to reduce air pollution were implemented in Beijing during the last 10 years. A coal-to-electricity heat supply project started in 2003 and aims to supply electric heating to the entire Beijing metropolitan area by 2015. Since 2008, Beijing authorities reported that projects aimed to reduce coal burning, traffic congestion, relocate factories away from the city center, plant trees and ban barbecues decreased sulfur dioxide emissions. However, the effects are still unknown due to limited scientific research.
During a haze-air pollution event on October 22nd, 2011, a Chinese news blogger drew public attention to PM2.5 levels in Beijing reported by the U.S. Embassy hourly via Twitter since 2008. After that, the public started to pay close attention to the PM2.5 monitoring data released by U.S. Embassy located in diplomatic area in Chaoyang District, Beijing. Chaoyang is the largest district located at the center of the Beijing metropolitan area (Figure 2). It occupies 475 km² and it has a total population of 3,545,000. Chaoyang has the highest population density of 7,463/km² in Beijing and a rapidly growing population rate. Public outcry regarding the availability of PM2.5 data from the US Embassy has impelled the Beijing government to start releasing PM 2.5 monitoring data of their own. The Beijing Municipal Environmental Monitoring Center (BJMEMC) first released PM2.5 data in February 2012 from the Chegongzhuang monitoring site. The geographic location of the US Embassy and Chegongzhuang monitoring site are shown in Figure 1.
The U.S. Environmental Protection Agency (EPA) revised the National Ambient Air Quality Standards (NAAQS) for particulate matter three times since 1997 (Table 2). In 2006, the EPA revoked the 24-hour PM10 NAAQS and changed the 24-hour NAAQS from the arithmetic mean of 50μg/m$^3$ to ‘not to be exceeded the standard more than once per year on average over a 3-year period’. In 2006, the EPA reduced the PM2.5 NAAQS 24-hour standard from 65μg/m$^3$ to 35μg/m$^3$ and in 2012 the EPA established a primary PM2.5 annual level as 12.0μg/m$^3$ to protect public health and a secondary standard of 15.0μg/m$^3$ to protect public welfare. We should notice that all the PM2.5 NAAQS are a 3 years average.

| NAAQS by EPA, U.S. |  |
|-------------------|---|---|
| Final Rule and Averaging Time | PM10(μg/m$^3$) | PM2.5(μg/m$^3$) |
| 1997 | 24h | 150 | 65 |
China also has ambient air quality standards (AAQS) for particulate matters; however, there was no standard for PM2.5 until February, 2012. However, researchers have measured PM2.5 over the years in Beijing (summarized in Table 1). In previous studies, daily, weekly monthly and seasonal PM2.5 concentrations are much higher than the NAAQS. However, the AAQS are not strict as NAAQS, instead of thee-year averages; AAQS use the one-year arithmetic mean. Since calculations and methods are different, AAQS and NAAQS are not comparable. Furthermore, AAQS are less extensive than NAAQS, and standards for PM2.5 are very recent.

<table>
<thead>
<tr>
<th>Year</th>
<th>Averaging Period</th>
<th>PM10(μg/m3)</th>
<th>PM2.5(μg/m3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1996</td>
<td>24h</td>
<td>50(i)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>150(ii)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>250(iii)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Annual</td>
<td>40(i)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>100(ii)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>150(iii)</td>
<td></td>
</tr>
<tr>
<td>2012</td>
<td>24h</td>
<td>50</td>
<td>35</td>
</tr>
<tr>
<td></td>
<td></td>
<td>15</td>
<td>75</td>
</tr>
<tr>
<td></td>
<td>Annual</td>
<td>40</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td></td>
<td>70</td>
<td>35</td>
</tr>
</tbody>
</table>

i: for urban area; ii: for pastoral region; iii: for agriculture and forestry area
Table 3. History of the AQQS for PM2.5 during the Period 1997-2012 by MEP, China

The purpose of the study is to expand previous research by examining data from a longer time range (3 years), use more precise calculation methods (hourly PM 2.5 and annual PM 2.5), use highly valid data (more than 90% scheduled sampling days data are valid), and compare these results to the EPA’s more conservative NAAQS standards. Furthermore, we will forecast the PM 2.5 values for the near future using time series analysis.
Materials and Methods

Materials: Data and Monitor

A doctoral student in the Computer Science Department at the University of Tsukuba built a website to capture the hourly data published (twittered) by U.S. embassy\textsuperscript{9}. Thirty-five thousand hourly PM2.5 measurements were retained based on the following criterion: 1) the measurement was kept if the measurements were available for at least 75\% (18 or more) of the hours during a 24-hours period. 2) If less than 75\% of the 24-hour period measurements were available, impute zero for each missing value and average over all 24-hours. The measurement was kept if it exceeds the 24-hour standard (35μg/m\textsuperscript{3}). 3) The measurement was kept if at least 75\% of the scheduled sampling days per quarter were available. 4) If three years of continuous data from single monitoring site were available.

The US Embassy uses an ambient particulate mass monitor (Model: Met One BAM 1020) to measure PM2.5 particulates as an indicator of the air quality on the Embassy compound located in Chaoyang district, Beijing. This real-time data monitor of ambient particle mass is designed to work under the Federal Equivalent Method (FEM) for continuous PM2.5 particulate monitoring, meets standards set forth in 40 CFR 58, Appendix D to Part 58\textsuperscript{10}, and are suitable for comparison with the NAAQS.

The US Embassy began twittering hourly PM2.5 data 25th August 2008 and data collection ended 30th April, 2013. This project used data between 1st January 2010 to 31st December 2012, which covers a 1010-day period, with 92.2\% valid sampling data. Other time-frames of data, specifically 2008, 2009
and 2013, were missing 50% or more than 50% of scheduled sampling days per quarter. In other words, data from those years have not achieved the minimum requirement to meet the EPA standards.

**PM2.5 Calculation Methods**

We plan to compare our results to the National Ambient Air Quality Standards (NAAQS), all of the daily must meet the standards of particulates matter in 40CFR Part 50 by U.S. EPA\textsuperscript{11}. For data to meet the current NAAQS annual PM2.5 standard, the 3-year average of the spatially averaged annual mean PM2.5 concentrations must be less than or equal to 12.0 \( \mu \text{g/m}^3 \) (for the secondary standard related to welfare, less than or equal to 15.0 \( \mu \text{g/m}^3 \)). To meet the current 24-hour PM2.5 standard, the 3-year average of the annual 98th percentile values for PM2.5 at each monitoring site must be less than or equal to 35 \( \mu \text{g/m}^3 \).

Calculation for 24-hour PM2.5 value (Standard Design Value, based on three year average of yearly 98th percentile 24-hr values)

1. Determine 98th percentile value for each year over three year period. First, we sorted all data values from lowest to highest by year, and then applied sign ranks, then identified the 98th percentile value of each year by multiplying the number of samples taken in the year by 0.98.

2. Average 98th percentile values over three years.

3. Round to the result nearest 1.0\( \mu \text{g/m}^3 \).

4. Compare the result to the NAAQ 24-hour standard of 35\( \mu \text{g/m}^3 \).

Calculation for Annual PM2.5 (The Annual Average Design Value)

1. Calculate the average of each quarter of each year over a three-year period.
2. Average four quarters in a calendar year to determine the average for each year.

3. Average the three annual values.

4. Round the resulting value to the nearest 0.1μg/m³.

5. Compare the result to the NAAQS PM2.5 annual primary standard of 12μg/m³ (or the secondary standard of 15μg/m³).11

Method: SARIMA

Time series analysis is widely used in statistics, econometrics, weather furcating, earthquake predictions and many other areas.12 In brief, time series analysis extracts meaningful characteristics from data and uses this information to predict future values. In time series analysis, the Box-Jenkins method commonly applies to ARIMA models to find the best-fit model in order to make predictions. There are three steps to construct ARIMA model by the Box-Jenkins method: 1) data preparation, 2) model identification and estimation, 3) furcating, which uses the model to predict future situations.

1. Data Preparation

After getting background information and carefully defining the variable, the most important and the first step in any time series analysis is data preparation. In this part, we determine if the series is stationary, an important mathematical and statistical property of time series analysis. A stationary time series indicates the mean, variance, autocorrelation, etc. of the series are constant over time.13 Probability theory of time series analysis relies on the assumption that the series is stationary. In order to examine if the series is stationary, visual inspection of a plot of the observations against time should not have a significant upward or downward trend. Second, the ACF plot can also indicate changes over
time. The Dickey Fuller test tests unit root in a time series. Third, the ACF test is an augmented version of Dickey-Fuller test for larger and more complicated sets of time series. Finally, the ADF test tests whether a time series variable is non-stationary by autoregressive model that use the existence of unit root as a null hypothesis.\(^{14}\)

If results from these tests show the series are non-stationary, differencing or log-transformation may be able to transfer the non-stationary series to stationary series. After we confirm the series is stationary, next test is the white noise test. White noise is defined as a purely random series which has a constant mean (usually zero) and constant variance.\(^{15}\) A white noise series is linearly unpredictable.\(^{16}\)

2. Model Estimation

Previous studies provide evidence that particulate matter concentrations are seasonal.\(^{17,18}\) Thus, a simple ARIIMA model is not sufficient. A multiplicative seasonal autoregressive integrated moving average model (SARIMA) was selected to as forecast model. The seasonal autoregressive integrated moving average model of Box and Jenkins (1970) is denoted as an SARIMA \((p;d;q)^*(P; D; Q)s\)

- \(p\) is the number of autoregressive terms,
- \(d\) is the number of non-seasonal differences (integrated part)
- \(q\) is the number of lagged forecast errors in the prediction equation (moving average term).
- \(P\) is AR order of seasonal part
- \(D\) order of seasonal difference
- \(Q\) is MA order of seasonal part
- \(S\) is period of the seasonal component
Same as ARIMA model estimation, the autocorrelation (ACF) plot provides information to choose the value of ‘q’ and ‘Q’. The partial autocorrelation (PACF) function plots are also useful aids in identifying appropriate ‘p’ and ‘Q’ values\(^9\).

3. Forecast

1) The last step of Box-Jenkins SARIMA modeling is to use the best-fit model to forecast the future situation, and graph forecasts against actual values.

Statistical software SAS version 9.3 was used for all analyses.
Results

PM2.5

The following tables (Table 4. and table 5.) show the 24-hour and annual values for PM2.5 concentration by year.

Table 4. 24-hours PM2.5 Standard Design Value (Based on three year average of yearly 98th percentile 24-hr values), US Embassy, Beijing 2010-2012

<table>
<thead>
<tr>
<th>Year</th>
<th>Quarter</th>
<th>PM2.5 value (μg/m³)</th>
<th>Yearly value (μg/m³)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2010</td>
<td>1</td>
<td>94.5</td>
<td>94.5</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>87.7</td>
<td>94.5</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>105.3</td>
<td>94.5</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>108.6</td>
<td>94.5</td>
</tr>
<tr>
<td>2011</td>
<td>1</td>
<td>82.1</td>
<td>94.925</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>85.5</td>
<td>94.925</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>104.8</td>
<td>94.925</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>107.3</td>
<td>94.925</td>
</tr>
<tr>
<td>2012</td>
<td>1</td>
<td>97.4</td>
<td>89.525</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>93.5</td>
<td>89.525</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>73.1</td>
<td>89.525</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>94.1</td>
<td>89.525</td>
</tr>
</tbody>
</table>

|       | 2010-2012 | 94.5            |

Table 5. Annual Average Design Values (μg/m³) US Embassy, Beijing 2010-2012
SARIMA modelling

1. Data Preparation

Figure 3 shows the PM2.5 daily average from 2010 to 2012 and the following Tables (Figure 4-6) are the PM2.5 daily average by each year. Visually they appear smooth, without upward or downward trends.

Figure 3. PM2.5 daily averages (of hourly averages per day) US Embassy, Beijing 2010-2012.
Figure 4. PM2.5 daily averages (of hourly averages per day) US Embassy, Beijing 2010

Figure 5. PM2.5 daily averages (of hourly averages per day) US Embassy, Beijing 2011
Fine particles often have seasonal patterns. Both the daily values and quarterly average of PM2.5 also reveals patterns based on the time of year. As shown in Table 8, PM2.5 values are typically higher in the fourth quarter (October through December).

The linear model fitted against time is an estimator of trend. Table 9 shows the downward trend over
the entire time span. The series appears to decline slowly. The horizontal black line drawn at about 100 indicates the mean of the series. This indicates the minimal trend.

Figure 8. PM2.5 Seasonal trends analysis by Linear Model 2010-2012 US Embassy, Beijing 2010-2012.

**Stationary or non-stationary**

Visual inspection is the first step to judge whether the series is stationary or not. Table 9, the trend analysis graphic, indicates minimal trend, providing evidence that this is stationary series. The second step, to the autocorrelation (ACF) function plot, shows how values of the series are correlated with past values of the series, and Table 10 provided evidence that the ‘value’ is a stationary time series since the ACF plot decreases rapidly and it was to zero after Lag2 (Figure 9). Moreover, the larger spike at lag 11 in ACF plot provides evidence that the PM2.5 daily value series has a seasonal autoregressive component. However, there is no peak at later stages indicating that the series is stationary at seasonal levels.
Table 6 provides results of the ADF test and show that all of the p-values are small enough to reject the null hypothesis that the series has a unit root. In other words, it is a stationary series.

<table>
<thead>
<tr>
<th>Augmented Dickey-Fuller Unit Root Tests</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Type</strong></td>
</tr>
<tr>
<td>Zero Mean</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Single Mean</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Trend</td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

Table 6. Results of Unit root tests

**White noise test**

The white noise test shows that the p-value for the first six autocorrelations is <0.0001, thus, the white noise hypothesis is strongly rejected, the series is not a white noise series.
**Log Transformation**

As we mentioned above at data preparation part, the series fluctuate wildly. In order to reduce the variation caused by extreme values, log-transformation was used to smooth the series. Table 11 showed the series after log transformation. Visually it appears smoother and more stable. The ACF plot decreases quickly after Lag 2 and provides evidence that the series is stationary after log-transformation.

![Trend and Correlation Analysis for Ivalue](image)

**Figure 10. Trend and correlation analysis for PM2.5 after log transformation**

**Seasonal Differencing**

The previous study indicated that PM2.5 in Beijing exhibit seasonality\(^2\). Although ACF and PACF are relatively stable, raw data plots (Table 4-7) exhibited seasonality. It is well-known that PM concentrations are the highest in winter, because particles are not easily dispersed into the air during the cold weather under unfavourable meteorological condition. It is necessary to eliminate seasonal fluctuations in order to understand and forecast the trends. A seasonal differencing method was used
before we fit the model. Table 12 plots the series after seasonal adjustments and the ACF plot decreases quickly after Lag2 providing additional evidence that this is a stationary series.

Figure 11. PM 2.5 plots after seasonal differencing, US Embassy, Beijing 2012.

2. Model Identification and Estimation

As mentioned before, since PM2.5 concentrations are seasonal, a SARIMA model was selected. By looking at the plots of Table 3.4 and previous seasonal differencing methods, we may define the potential model SARIMA \((p;d;q)^s(P; D; Q)s\) = SARIMA \((1,0,2)(1,1,1)_{12}\).

Model Check

The Table 8 lists four parameters in the model; the t-value provides a significance test for the parameter estimates and indicates whether some terms in the model may be unnecessary. In this series, the t-value for all autoregressive parameter is highly significant. The t-value for ‘MU’ parameter indicates the mean term adds little to the model.
Table 7. Parameter Estimates for SARIMA (1,0,2)(1,1,1)12 Model

Table 8. Autocorrelation Check of Residuals: SARIMA (1,0,2)(1,1,1)12 Model

The Table 9 autocorrelation checks of residuals that have same form as the autocorrelation check for white noise. The X² tests (P>0.05) indicate that we cannot reject the hypothesis that the residuals are uncorrelated, that means the model SARIMA (1,0,2)(1,1,1)12 is adequate for the series, and there is no need to try more complex models.
3. Prediction

The last step of the time series analysis is prediction. SARIMA model is a good method to forecast the future. As the Table 13 shows 90 days prediction started from ‘2013-01-01’.

![Prediction using SARIMA (1,0,2)(1,1,1)12 Model](image)

**Figure 12.** Prediction using SARIMA (1,0,2)(1,1,1)12 Model

![2013 Beijing PM2.5 concentration (1st Jan 2010-31st Mar2013)](image)

**Figure 13.** 2013 Beijing PM2.5 concentration (1st Jan 2010-31st Mar2013)
Discussion

Based on our data, the 24-hour PM2.5 standard design value was 294μg/m³ (Table 4). Compared to the 24-hour NAAQS, 35μg/m³, our values are about 7.4 times higher. The Annual Average Design Value (based on three years’ data) is 94.5μg/m³ (Table 5). Compared to the annual NAAQS 12μg/m³, our values are about 6.9 times higher than the NAAQS.

The overall PM2.5 concentrations are decreasing. However, the actual values were much higher than the predicted values. This may be due to the 2013 winter; 9 out of 30 days were fog-haze days, compared to the same time in 2010 and 2011. Secondly, extremely high PM2.5 concentrations in the first quarter of 2013 caused the actual PM2.5 value to increase. Compared to the quarterly concentrations from 2010 to 2012, the 1st quarter PM2.5 concentration in 2013 is about 50% higher than the other three-year’s concentration. Thirdly, industrial activities in neighboring provinces may increase the actual PM2.5 concentration.

Wang et al. (2013) indicated U.S. Embassy PM2.5 observations exhibited the same trends as citywide PM2.5 values, but the embassy’s concentration values were higher. Forty-three hourly sample data were collected both from the BJMEMC and the USE monitoring site between 2nd February to 4th February, 2012. Figure 12 shows the data; although the trends of 2 set of series are similar, data from the USE reported higher concentrations.
A potential reason for the differences between the USE and BJMEM PM2.5 values are the monitoring devices. The BJMEMC used a ‘Thermo Fisher Scientific’ monitor, which uses a tapered element oscillating microbalance (TEOM) detection method. The USE site used BAM 1020 from ‘Met One’ company which uses a β-ray decay detection method. Both of the TEOM and Beta-ray detection method are widely used in PM2.5 mass measurement. However, there are concerns about the accuracy of TEOM including lose semi-volatile material and values measured using the TEOM method are, on average, 18.3% lower than the Gravimetric method across Europe. For PM10 detection, a study found the Beta-ray decay method was 15.8% higher than the TEOM method, and the deviation between TEOM and Beta-decay method was 7.5% (cite). On the other hand, a U.S. study used Gravimetric samplers as a reference method to compare other PM measurements and reported that the TEOM method was more accurate than Beta method.
25

<table>
<thead>
<tr>
<th>Monitor</th>
<th>BJMEMC</th>
<th>U.S. Embassy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Detection Method</td>
<td>Thermo Fisher Scientific</td>
<td>Met One BAM 1020</td>
</tr>
<tr>
<td>Calculation</td>
<td>Tapered Element Oscillating Microbalance (TEOM)</td>
<td>β-ray decay</td>
</tr>
<tr>
<td></td>
<td>24-hour and Annual PM2.5 concentration: arithmetic mean</td>
<td>24-hour and Annual PM2.5 concentration: based on 3 year period</td>
</tr>
</tbody>
</table>

Table 9. Different monitor detection and calculation method between BJMEMC and U.S. Embassy PM2.5 monitoring sites.

Another reason is that PM2.5 is high during the winter months in Chaoyang Beijing, because fine particles are hard to dissipate during the winter in Beijing. As I mentioned before the air quality in Beijing depends on the meteorological conditions and topography, meteorological conditions are unfavorable for pollutants dispersion during the cold season. Moreover, charcoal heating is widely used during the winter season and fine particle are more readily formed in cooler weather. The sources of PM2.5 also varied by seasons and that help to explain the seasonality of the series.

Limitations

This study used data on one site thus it is limited in its ability to describe the PM2.5 pollution situation for whole district or city. However there is no other PM2.5 data available. Existing Beijing Authority PM2.5 monitoring stations are currently not working but they are planning to set more than 50 new air monitoring sites in the near future thanks to a new legislative mandate that aims to implement multiple monitoring sites in 2016. These data will provide a historical baseline to compare future spatial data from multi-monitoring sites.
Future Directions

As air pollution in Beijing draws greater public concern, more legislative environmental projects are implemented every year. This research provides a basis for future studies on the effects of these projects. Furthermore, I found that the USE site recorded higher levels of PM2.5 than the BJMEMC site for multiple reasons. Further research should include interdisciplinary collaboration and include data from Beijing neighboring provinces. Moreover, ozone and PM10 studies also important for research on air pollution.
Conclusion

Beijing is an important international city that is losing its fame and talents due poor air quality. This thesis has described 3 years of PM2.5 data in Chaoyang District of Beijing. Moreover, 24-hour and annual PM2.5 concentrations at Chaoyang District of Beijing exceeded NAAQS standards. However, concentrations are decreasing over time.

The SARIMA model was built by using Box-Jenkins method. However, the differences between predicted and observed PM2.5 concentrations may be due to the extreme high PM2.5 concentrations in January 2013. The record-breaking values were caused by local, regional pollutions and meteorological conditions.

We compared the PM2.5 levels to other Asian cities, like Ahwaz, Iran (annual: 372μg/m³) and Ludhiana, India (annual: 251μg/m³). The PM2.5 concentration in Beijing was not one of the worst, not even close. But we still have a long way to go to attain the standards. For future studies, we recommend more data from spatial monitoring site measuring both PM2.5, PM10 and PM2.5/PM10 ratio studies, as well as new interdisciplinary studies to further investigate potential sources of the pollution. Last, but not least, studies should not only focus on the local pollution but also regional air quality from neighboring provinces.
Appendix

What is particle pollution?

Particle pollution is a mixture of microscopic solids and liquid droplets suspended in air. This pollution, also known as particulate matter, is made up of a number of components, including acids, such as nitrates and sulfates, organic chemicals, metals, soil or dust particles, and allergens, such as fragments of pollen or mold spores.

The size of particles is directly linked to their potential for causing health problems. Small particles less than 10 micrometers in diameter pose the greatest problems because they can get deep into your lungs and some may even get into your bloodstream. Exposure to such particles can affect both your lungs and your heart. Larger particles are of less concern although they can irritate your eyes, nose and throat.

Small particles of concern include “fine particles”, such as those found in smoke and haze, which are 2.5 micrometers in diameters or less; and “coarse particles”, such as those found in wind-blown dust, which have diameters between 2.5 and 10 micrometers.

What is PM2.5?

PM2.5 refers to particulate matter that is 2.5 micrometers or smaller in size. 2.5 micrometers is approximately 1/30 the size of a human hair; so small that several thousand of them could fit on the
PM2.5 can be emitted directly from both human activities and natural sources. “Secondary” PM2.5 can form from gases, such as oxides of nitrogen (NO$_x$) or sulfur dioxide (SO$_2$), reacting in the atmosphere. Other secondary particles include organic carbon particles, which can be formed when certain volatile organic compounds react with other gases in the atmosphere. Sources of particles include burning activities, motor vehicles emissions, and other combustion activities. The U.S Environmental Protection Agency (EPA) set the first PM2.5 National Ambient Air Quality (NAAQS) in 1997 at 15μg/m$^3$ for the annual standard and at 65μg/m$^3$ for daily standard. In 2006, EPA strengthened the daily standard by lowering it to 35μg/m$^3$. In 2012, EPA strengthened the annual standard, set first (12μg/m$^3$) and secondary (15μg/m$^3$) annual standard.

### Air Quality Index

<table>
<thead>
<tr>
<th>Air Quality Index</th>
<th>Air Quality</th>
<th>Health Advisory</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-50</td>
<td>Good</td>
<td>None.</td>
</tr>
<tr>
<td>51-100</td>
<td>Moderate</td>
<td>Unusually sensitive people should consider reducing prolonged or heavy exertion.</td>
</tr>
<tr>
<td>101-150</td>
<td>Unhealthy for Sensitive Groups</td>
<td>People with heart or lung disease, older adults, and children should reduce prolonged or heavy exertion.</td>
</tr>
<tr>
<td>151-200</td>
<td>Unhealthy</td>
<td>People with heart or lung disease, older adults, and children should avoid prolonged or heavy exertion. Everyone else should reduce prolonged or heavy exertion.</td>
</tr>
<tr>
<td>201-300</td>
<td>Very Unhealthy</td>
<td>People with heart or lung disease, older adults, and children should avoid all physical activity outdoors. Everyone else should avoid prolonged or heavy exertion.</td>
</tr>
</tbody>
</table>

Table 10. Air Quality Index
How does PM2.5 affect the environment?

The fine particles lead to health effects and also reduce the visibility. It is estimated that in certain parts of the U.S. the visual range has been reduced by 70% of natural conditions. Because these particles are so small, they can travel great distances affecting areas in other states or even regions. It is believed that one-third of the haze seen over the Grand Canyon comes from Southern California29.

The fine particles also contribute to acid rain that affects all biological or man-made and by thus affecting the environment, can have repercussions to human health34.

What are the National Ambient Air Quality Standards (NAAQS) for PM2.5?

<table>
<thead>
<tr>
<th>PM2.5</th>
<th>primary</th>
<th>Annual</th>
<th>12 μg/m³</th>
<th>annual mean, averaged over 3 years</th>
</tr>
</thead>
<tbody>
<tr>
<td>secondary</td>
<td>Annual</td>
<td>15 μg/m³</td>
<td>annual mean, averaged over 3 years</td>
<td></td>
</tr>
<tr>
<td>primary and secondary</td>
<td>24-hour</td>
<td>35 μg/m³</td>
<td>98th percentile, averaged over 3 years</td>
<td></td>
</tr>
</tbody>
</table>

Table 11. NAAQS of PM2.5

Primary standards provide public health protection, including protecting asthmatics, children, and the elderly population.
Secondary standard provide public welfare protection, including protection against decreased visibility and damage to animals, crops, vegetation, and buildings.

**What are Design Values?**

Design values are metrics for assessing air quality, 24-hour design values, and annual values. Design value calculations are based on three-year date that followed by EPA protocols for sample complement, missing data substitutions, sample validity, routing, sample frequency, and averaging calculation.
References


& Waste Management Association, 47(6), 682-689.