Running Head: Beijing PM2.5 Haze Pollution Problem

Title: PM2.5 Data Exploration and How the Environment Affects PM2.5 in Beijing

Yinglu Deng

Abstract

Beijing (the Capital of China) faces severe haze pollution problems mainly due to coal-based energy consumption which increases the PM2.5 concentrations. PM2.5 is a 2.5 micrometer or less than 2.5 particle pollutant exposure in the air. Because of the poor air quality, it increases hospital admissions for lung or heart causes, chronic or acute bronchitis. Here we discovered the data of PM2.5 and meteorological variations starting from the year of 2010 to 2015. The PM2.5 time series analysis shows average PM2.5 concentrations in Winter is slightly higher than the other seasons. By using seasonal ARIMA, the PM2.5 concentrations can be predicted for 2015 and the trend is decreasing year by year. Also, the null hypothesis of the environment has no effect on the PM2.5 (like Temperature, wind, snow, rain) is proven as false and MARSS model proves that the environment shows a significant effect on PM2.5. We end with a discussion of the possible results by which the dynamic factors may affect PM2.5 in some way. Predicting the future trend and giving out some plausible suggestions for communities and government to solve for this PM2.5 pollution problem.

Keywords: Beijing, PM2.5, haze pollution, factor, AICc, seasonal, null hypothesis, time series, MARSS, ARIMA.

Introduction

Since late 1970s, Mainland China has gradually reformed and developed its economy. With the rapid development in industrial production and demanding of residents' life, people are lack of the awareness of environmental protection, regulations and laws which leads to poor pollution problem. Especially in the east of China, it causes widespread haze due to more and more wood and coal are consumed for heating in cold winter.

Haze has a greater impact on climate, environment, and human health. There is greater impact for the range of PM1 to PM2.5 in haze. Dust and sand will appear within the range of PM2.5 to PM10. Those extreme weather events may be caused by the increase in PM2.5 concentration. Also, PM2.5 breaks the balance of atmospheric radiation so that the ground becomes colder and the atmosphere gets hotter, which seriously affects the climate change. And it may aggravate regional atmospheric heating effects. Moreover, fine particle pollution has become one of the most crucial environmental problems in the world. Additionally, PM2.5 concentration will cause the phenomenon of urban acid rain and photochemical haze, resulting in a decrease in the visibility of atmosphere and obstructing land and air transportation. Finally, haze contains relatively high humidity which can directly infect viruses and bacteria. The PM2.5 is also known as lung-entry particles which can directly enter the human's alveoli and the blood system, and it will cause cardiovascular diseases easily. The prevalence of lung cancer has increased rapidly year by year in Beijing. Based on the report for 2012, average 104 lung cancer patients were diagnosed every day.

Haze is mainly caused by economic and social activities, weather conditions, and atmospheric circulation in Beijing. Approximately 90% of the haze comes from anthropogenic emissions which are directly from human social and economic activities. For Beijing's annual average PM2.5 emissions, 26% of them accounts for coal, 19% comes from motor vehicles and 10% for industrial. Taking 2016 as an example, the annual average PM2.5 concentration reached 73 ug/m^3 in Beijing which is 4.8 times the annual average concentration in Los Angeles where is the place of the most polluted city in the United States, and it is 7.2 times the recommended standard of the WHO. (World Health Organization, 2016)

Additionally, the haze is seasonal in Beijing. In the high temperature season of summer and autumn, the vertical movement of gas is so active that the industrial waste gas is easily dissipated. But the momentary sunshine is just an illusion. It does not mean that the haze is gone, and governance has been effective. When it comes to severely cold winter, the air sinks and the haze stays on the ground. Beijing is back to the situation that the haze locks the city again. China's rapid economic development has greatly improved people's life quality in the past four decades, but it has also brought many new problems at the same time. Haze is a problem we urgently need to solve now. (**Figure 1**)

In this paper, I will explore three general questions and one hypothesis to deeply understand the cause and trend of PM2.5 concentration. First, what is the day of average highest pm2.5 in each month? Secondly, which season has severer pm2.5 problem in a year? Last, does the problem of pm2.5 become more serious year by year and what is the trend for it? And we will also test if the null hypothesis of the environment has no effect on the PM2.5 is true.

Methods

From the 2010-2014 PM2.5 time series plot, the time series is roughly stable (**Figure 2**). Also, it is clearly that this time series has seasonal features that PM2.5 is higher in Autumn and Winter and lower in Spring and Summer, so we need to do more data visualization for exploring other seasonal characteristics.

The original data has been divided into different seasons. Among them, March, April and May are designated as Spring; June, July and August are designated as Summer; September, October, and November are classified as Autumn; and December, January and February are classified as Winter. In the average PM2.5 box plot, the upper quartile of Winter's PM2.5 is higher than the other seasons, and there are also more outliers of higher concentrations in Winter. (**Figure 3a**) Moreover, a PM2.5 density plot on different seasons (**Figure 3b**) is helpful for us to identified that Autumn and Winter curve are righter shewed than Spring and Summer.

For the prediction model, we have established the ACF diagram (**Figure 4a**) and the PACF diagram (**Figure 4b**). Both the autocorrelation coefficient and the partial correlation coefficient are tailed. So, the sequence is not an AR model or an MA model. Furthermore, we try to use the auto.arima function which combines the waves of different frequencies and different periods. It can support the fitting of time series with seasonal trends, so we fit two models including the non-seasonal ARIMA model and the seasonal ARIMA model. (**Table 1**). The fitted AICc values of ARIMA (0,0,0) and ARIMA (0,0,0)(0,0,1)[12] are 571.64 and 566.34 respectively. The seasonal ARIMA model fits better than the non-seasonal ARIMA model due to the lower AICc value. Then we predict trend of PM2.5 concentration for 2015 by using seasonal ARIMA model. (**Figure 4c)** It is obvious that the average PM2.5 concentration was decreasing since the year of 2013. In 2013, the average PM2.5 concentration reached a peak at 190 ug/m^3, then it decreased to 175 ug/m^3 in 2014. Besides, there was a small peak at 140 ug/m^3 in the end of 2014. As shown on the line graph, there are two highest peaks at 130 ug/m^3 and 115 ug/m^3 respectively the predicted line (blue line) for 2015. Therefore, the average PM2.5 concentration is decreasing year by year.

Finally, we use MARSS model to test null hypothesis of the environment (like Temperature, wind, snow, rain) has no effect on the PM2.5. The target is PM2.5 concentration (ug/m^3). And there are four covariates including TEMP_mean (mean Temperature (â„ƒ)), Iws_mean (mean Cumulated wind speed (m/s)), Is_mean (mean Cumulated hours of snow), Ir_mean (mean Cumulated hours of rain). First, we do data transformation for our covariate data by z-score, it allows us to calculate the probability of a score occurring within a standard normal distribution. Then we can start to build the MARSS model with each covariate separately and bootstrap them.

For these four models, they all contain 0 within the lower CI and upper CI which means it is not significant that the null hypothesis is true, and we need to reject the null hypothesis, and the environment has effect on the PM2.5 concentration. Based on the U matrix and C matrix, they show wind and snow exist long-term trends and affect PM2.5 due to the matrix does not include 0. Besides, Q matrix (the process error variance-covariance matrix) tells us that all covariances show stochastic variation. (**Figure 5**)

Results

First, we build a lower AICc seasonal ARIMA model to predict the average PM2.5 concentration for next year. The result can be concluded that the predicted average PM2.5 is lower in 2015 compared to the previous year and the average PM2.5 concentration was decreasing year by year especially starting from 2013. But there are still concerns about the model since the seasonal ARIMA model can only predict a small part of the average PM2.5 value. Because the latest observations have a greater impact on the predicted value, but the data from few years ago will have smaller impacts on the predicted year. When the prediction order continues to increase, the prediction result will return at the average level, the model is related to the lower order of the model and the shorter memory of the sequence. Since the time span of the data is only 5 years, the sample size is too small for establishing a accurate model, so we still need to collect more data to fit our model.

Lastly, we need to reject the null hypothesis and the environment has effect on the PM2.5 concentration by the MARSS model. All four covariances (temperature, wind, snow, rain) have impact on average PM2.5 concentration. The PM2.5 shows stochastic variation. For the U matrix (long-term trend) and C matrix, it shows the trend and covariate effects exist on wind and snow. For Wind, the maximum likelihood estimation of U matrix and C matrix are -0.0159 and -0.4623 respectively, which means there is a negative relationship between wind and PM2.5. Higher wind speed will blow away the pollutant and lower down the amount of PM2.5 concentration. For Snow, the maximum likelihood estimation of U matrix and C matrix are 0.0277 and 0.6246 respectively. It contains a positive relationship between snow and PM2.5, more cumulated hours of snow can lead to a higher amount of PM2.5.

Discussion

In summary, although the concentration of air pollution emissions has shown a downward trend in Beijing, the damage to human health is still serious. In 2016, haze pollution resulted in 5,900 premature deaths, 361,500 hospitalizations for respiratory diseases, 129,900 hospitalizations for cardiovascular diseases, 344.1 million asthma visits in Beijing. Among them, PM2.5 is the biggest and mainly threat to residents' health. Second, the air pollution causes the huge economic loss to Beijing's residents. The total economic loss for health problem was 67.925-billion-yuan which accounts for 3% of Beijing's GDP in 2016. Third, from the perspective of time trends from 2010 to 2015, Beijing's economic loss in health first increased in a fluctuating manner and then decreased year by year.

To better solve this question, we have the following measures. First, the government needs to strengthen the emissions control of pollutant including inhalable particulate matter and nitrogen oxides. Although the air pollution health loss has shown a decreasing trend in various districts of Beijing from 2015 to 2016, this does not mean haze pollution control can be slack. In 2016, Beijing's air pollution health loss was 1.1 times that of 2009, and the haze pollution situation is still not optimistic. The government should effectively start to work from the city's motor vehicles, coal combustion, industrial activities, dust and regional joint prevention and control, and adopt several effective measures to control the haze pollution in Beijing, especially the strict control of motor vehicle exhaust emissions.

Secondly, the air pollution prevention and control countermeasures in each district need to be adjusted according to the actual situation. There are significant differences in air pollution health losses and historical trends in various districts of Beijing, which are mainly caused by factors such as geographic location and permanent population. As a whole, each district of Beijing needs to focus on different areas in the efforts to reduce air pollution and health losses in accordance with their respective development missions and local conditions.

Finally, we should vigorously develop relatively clean energy sources in order to reduce sulfur dioxide emissions such as hydropower, nuclear power, wind power, and solar power, and reduce our dependence on coal. However, China is a big country, energy transformation takes a long time and process. Even if we start to act now, China will still need to use a lot of coal in ten to twenty years. In this regard, small coal-burning enterprises should be merged and closed, and large enterprises must be further technologically deepened to improve the application of desulfurization technology to increase the thermal efficiency of coal combustion. In rural areas, the state should vigorously promote the application of coal “decoupling combustion technology” stoves and provide subsidies to farmers who actively use new stoves. In large and medium-sized cities across the country, the number of fuel vehicles should be gradually limited, the issuance of fuel vehicle licenses should also be reduced, and electric vehicles should be promoted.

Acknowledgements

This work was supported by Professor Albert Ruhi and Charlotte Kwongs for statistical advice that greatly improved this manuscript. The Beijing PM2.5 dataset was provided by UC Irvine of Machine Learning Repository.

Literature Cited

Chen, S.M. and L. Y. He (2014), "Welfare Loss of China's Air Pollution: How to Make Personal Vehicle Transportation Policy", China Economic Review, 31, pp. 106 – 118

Chen, Y. Y., A. Ebenstein and M. Greenstone, et al. (2013) "Evidence on the Impact of Sustained Exposure to Air Pollution on Life Expectancy from China's Huai River Policy", PNAS, 110(32), pp. 12936 – 12941

Chen SY, Lin YL, Chang WT, Lee CT, Chan CC. Increasing emergency room visits for stroke by elevated levels of fine particulate constituents. Sci Total Environ. 2014;473-474:446–50.

Kim KH, Kabir E, Kabie S. A review on the human health impact of airborne particulate matter. Environ Int. 2015;74:136–43.

P. Jiang, Q. Dong, and P. Li, "A novel hybrid strategy for $PM_{2.5}$ concentration analysis and prediction," *Journal of Environmental Management*, vol. 196, pp. 443–457, 2017.

Ruckerl R, Schneider A, Breitner S, *et al*. Health effects of particulate air pollution: A review of epidemiological evidence. Inhal Toxicol 2011;**23**:555–92.

W. Sun and J. Sun, “Daily $PM_{2.5}$ concentration prediction based on principal component analysis and LSSVM optimized by cuckoo search algorithm,” *Journal of Environmental Management*, vol. 188, pp. 144–152, 2017.

World Health Organization Ambient (outdoor) air pollution in cities database 2014. https://www.who.int/ (2016). Accessed 26th Nov 2021.

Tables

```
ARIMA(0,0,0) with non-zero mean

Coefficients:
         mean
      98.8295
s.e.   3.5341

sigma^2 estimated as 762.1:  log likelihood=-283.71
AIC=571.43   AICc=571.64   BIC=575.62
Series: pm_q3
ARIMA(0,0,0)(0,0,1)[12] with non-zero mean

Coefficients:
        sma1     mean
      0.5751  98.6264
s.e.  0.2242   4.7123

sigma^2 estimated as 631.4:  log likelihood=-279.96
AIC=565.92   AICc=566.34   BIC=572.2
```

Table 1. Summary of non-seasonal ARIMA model and seasonal ARIMA model.

```
##
## MARSS fit is
## Estimation method: kem
## Convergence test: conv.test.slope.tol = 0.5, abstol = 0.001
## Algorithm ran 15 (=minit) iterations and convergence was reached.
## Log-likelihood: -99.96075
## AIC: 207.9215   AICc: 208.6488
##
##          ML.Est Std.Err low.CI up.CI Est.Bias Unbias.Est
## U.1    -0.00743   0.177  -0.36 0.327 0.000744   -0.00668
## Q.Q     1.63910   0.289   1.07 2.192 0.088755    1.72786
## x0.x0 -0.28551   1.334  -3.05 2.332 0.024585   -0.26093
## C.1     0.00661   0.177  -0.36 0.327 0.000744    0.00735
## Initial states (x0) defined at t=0
##
## CIs calculated at alpha = 0.05 via method=parametric
## Bias calculated via  bootstrapping with  bootstraps.
```

Table 2a. Marss model for TEMP_mean (mean Temperature (â,,ƒ)) and average PM2.5 concentration.

```
##
## MARSS fit is
## Estimation method: kem
## Convergence test: conv.test.slope.tol = 0.5, abstol = 0.001
## Algorithm ran 15 (=minit) iterations and convergence was reached.
## Log-likelihood: -95.937
## AIC: 199.874   AICc: 200.6013
##
##          ML.Est Std.Err low.CI  up.CI Est.Bias Unbias.Est
## U.1    -0.0159   0.155 -0.772 -0.167  0.00848   -0.00738
## Q.Q     1.4334   0.253  0.949  1.880  0.06174    1.49510
## x0.x0  0.2206   1.206 -2.135  2.411  0.04273    0.26333
## C.1    -0.4623   0.155 -0.772 -0.167  0.00848   -0.45382
## Initial states (x0) defined at t=0
##
## CIs calculated at alpha = 0.05 via method=parametric
## Bias calculated via  bootstrapping with  bootstraps.
```

Table 2b. Marss model for Iws_mean (mean Cumulated wind speed (m/s)) and average PM2.5 concentration.

```
##
## MARSS fit is
## Estimation method: kem
## Convergence test: conv.test.slope.tol = 0.5, abstol = 0.001
## Algorithm ran 15 (=minit) iterations and convergence was reached.
## Log-likelihood: -93.63661
## AIC: 195.2732   AICc: 196.0005
##
##        ML.Est Std.Err low.CI up.CI Est.Bias Unbias.Est
## U.1    0.0277   0.170  0.292 0.963 -0.00126     0.0264
## Q.Q    1.3276   0.236  0.826 1.807  0.07256     1.4001
## x0.x0 -2.3937   1.284 -4.902 0.155 -0.01804    -2.4118
## C.1    0.6246   0.170  0.292 0.963 -0.00126     0.6234
## Initial states (x0) defined at t=0
##
## CIs calculated at alpha = 0.05 via method=parametric
## Bias calculated via  bootstrapping with  bootstraps.
```

Table 2c. Marss model for Is_mean (mean Cumulated hours of snow) and average PM2.5 concentration.

```
##
## MARSS fit is
## Estimation method: kem
## Convergence test: conv.test.slope.tol = 0.5, abstol = 0.001
## Algorithm ran 15 (=minit) iterations and convergence was reached.
## Log-likelihood: -99.9613
## AIC: 207.9226   AICc: 208.6499
##
##         ML.Est Std.Err low.CI up.CI Est.Bias Unbias.Est
## U.1   -0.00720   0.169 -0.337  0.32  0.00124   -0.00596
## Q.Q    1.63913   0.296  1.017  2.16  0.08030    1.71943
## x0.x0 -0.29934   1.349 -2.913  2.33 -0.01287   -0.31221
## C.1   -0.00335   0.169 -0.337  0.32  0.00124   -0.00211
## Initial states (x0) defined at t=0
##
## CIs calculated at alpha = 0.05 via method=parametric
## Bias calculated via  bootstrapping with  bootstraps.
```

Table 2d. Marss model for Ir_mean (mean Cumulated hours of rain) and average PM2.5 concentration.

Figures

Figure 1. A pedestrian walks amid heavy smog in Beijing, Dec 29, 2015. [Photo/Xinhua]

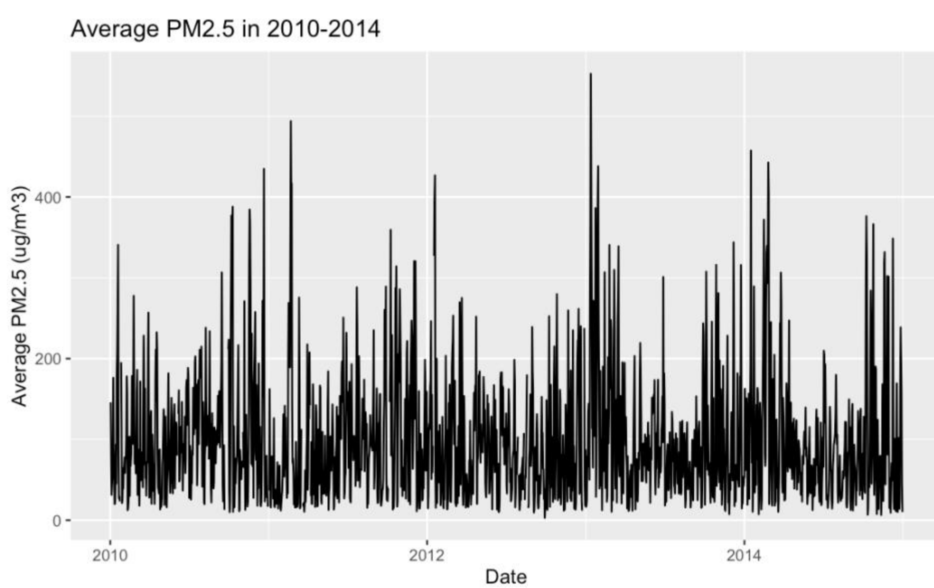

Figure 2. Average PM2.5 time series plot in the year of 2010-2014.

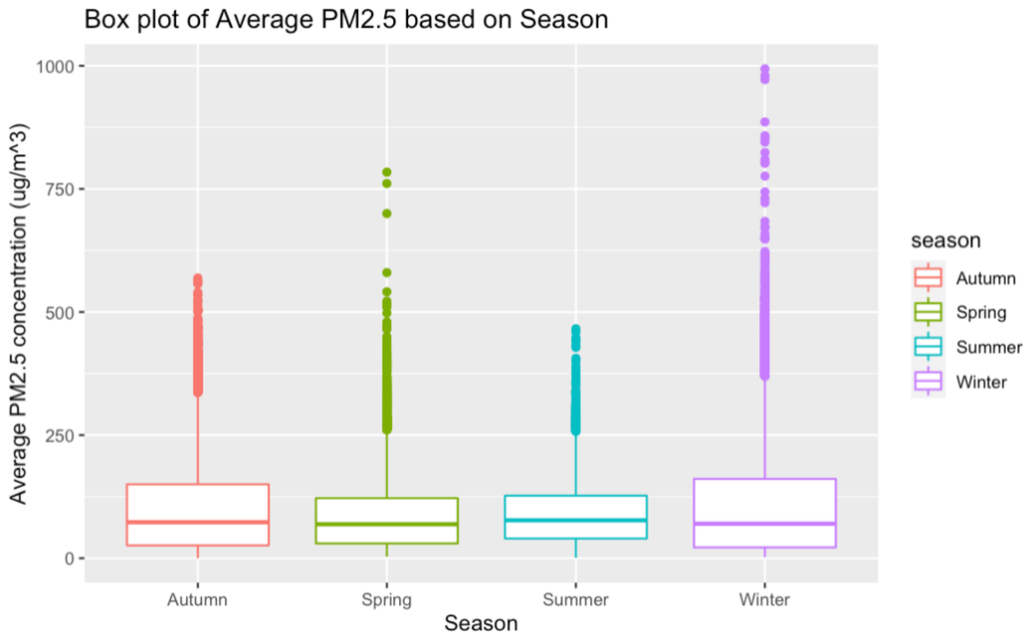

Figure 3a. The box plot for average PM2.5 based on season.

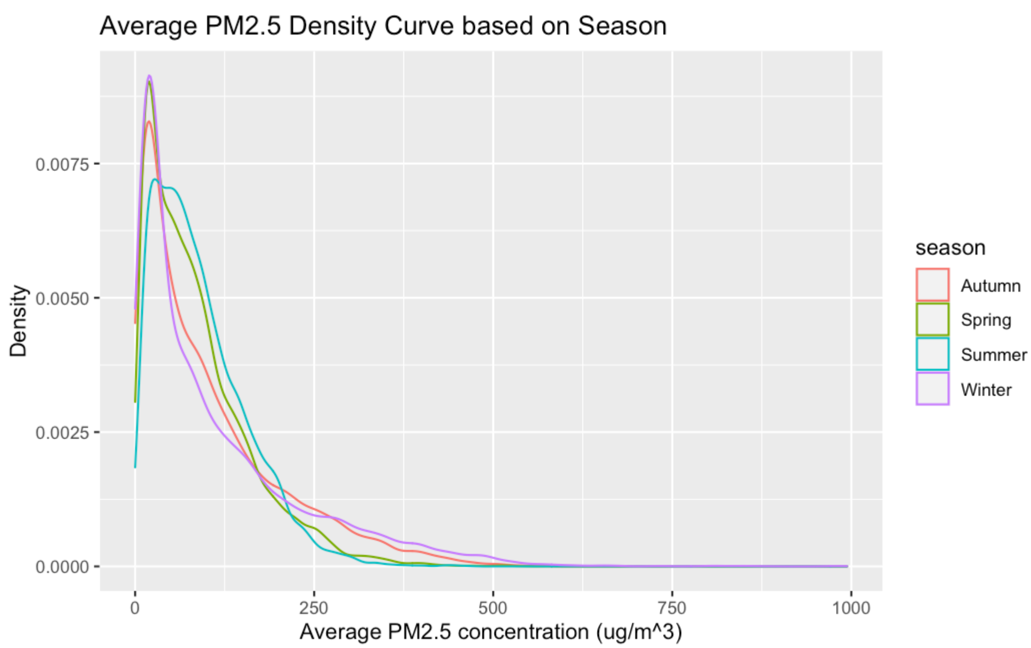

Figure 3b. Average PM2.5 concentration density curve based on season.

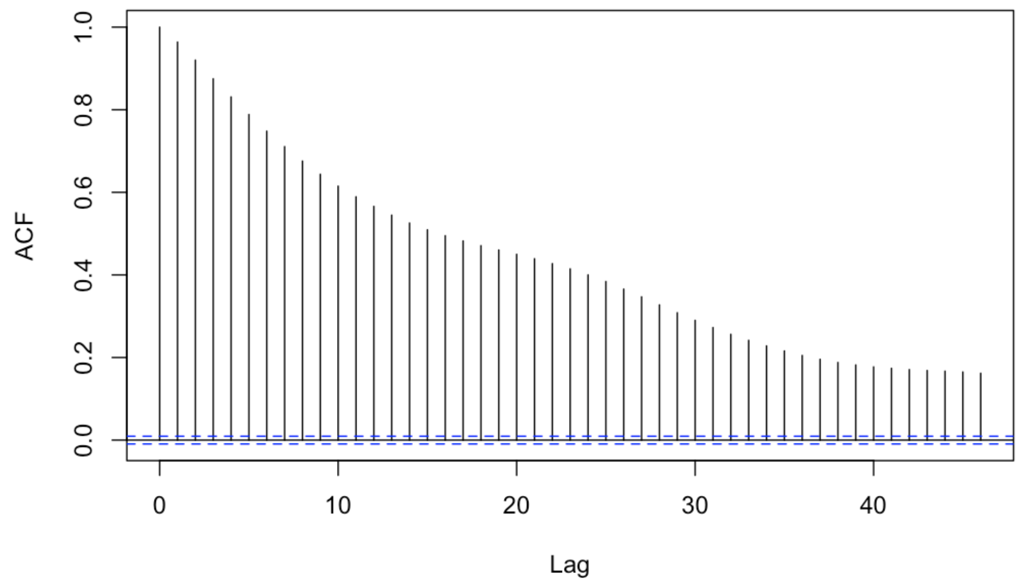

Figure 4a. The autocorrelation function (ACF) graph.

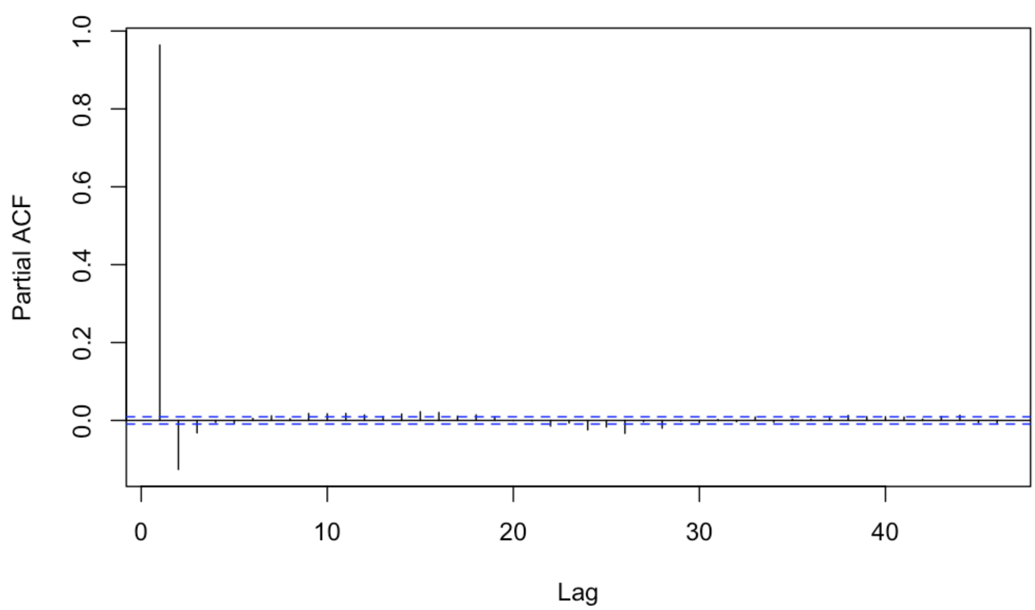

Figure 4b. The partial autocorrelation function (PACF) graph.

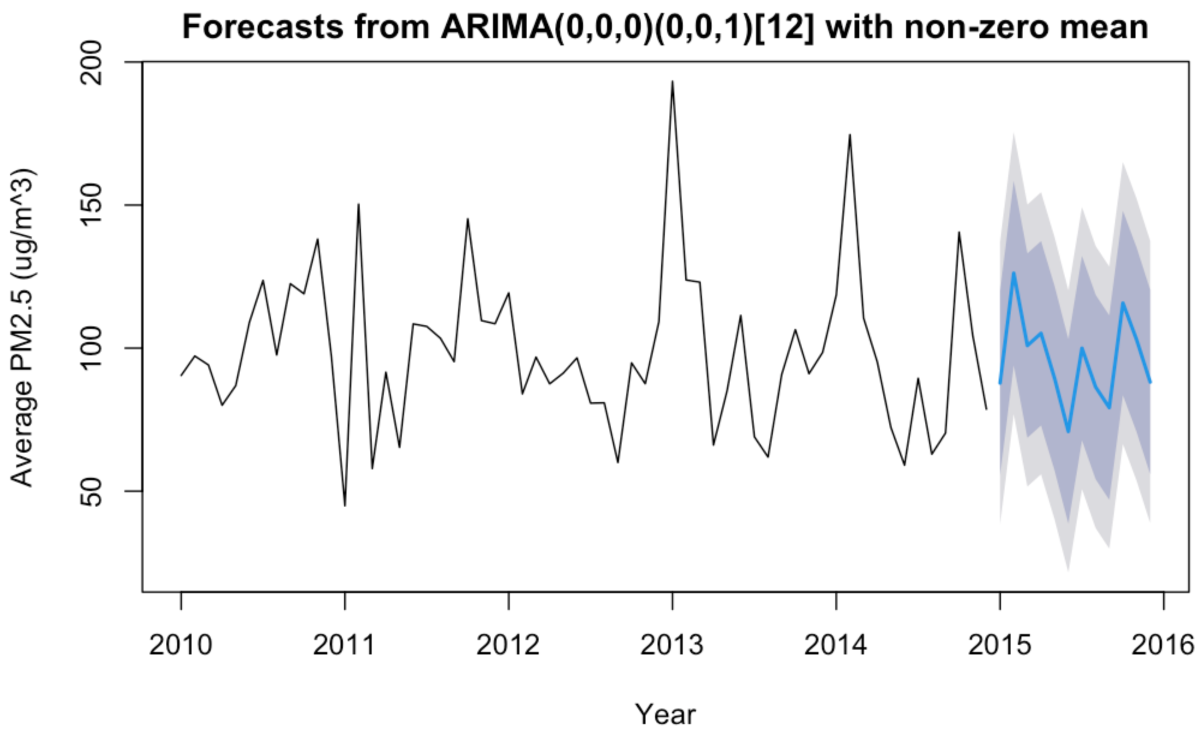

Figure 4c. Seasonal ARIMA forecast average PM2.5 plot for the year of 2015.