Statistical Analysis of Changes PM2.5 under Local Policies among Five Cities in China

Report prepared for Feifang Hu

By

Jian Sun
Menghan Wang
Mian Wang
Xinran Xu

GWU Statistical Consulting Class
Professor Feifang Hu

Dec 06, 2016

STATISTICAL CONSULTING CLASS

CONSULTANTS: Jian Sun
Menghan Wang
Mian Wang
Xinran Xu

DATE: Dec 06, 2016
CLIENT: Feifang Hu
PROJECT: Policies of controlling PM2.5 for Five cities in China

EXECUTIVE SUMMARY

Currently, the high concentration of PM 2.5 have became hit news in China, since it would cause horrible disease for human beings. To avoid worse consequence, Chinese government is taking measures to control the situation and keep it from getting worse. Therefore, the PM2.5 data until September, 2016 of five main Chinese cities, BeiJing, ShangHai, GuangZhou, ShenYang and ChengDu, is chosen to analyze. According to the various cities' different policies, this paper did t-test, ANOVA and CCF to test if the government policies worked on controlling the air quality. What's more, time series forecasting is done to judge that if the government target will be fulfilled on time. The attached plots directly reveal that the changes of PM 2.5 for recent years. In general, the whole crisis is gradually relieved. The folks can still trust their government.

1 Introduction

The aim of this report is to compare the changes of air qualities under the local government's policies for Beijing, Shanghai, Chengdu, Shenyang and Guangzhou for controlling concentration of PM2.5. And we also want to investigate how does the concentration of PM2.5 in Beijing affect that of Shanghai.

1.1 Study Design:

In this investigation, hourly concentration of PM 2.5 (ug/m3) is used to evaluate the situation of air condition in Shanghai, Beijing, Chengdu, Shenyang and Guangzhou these five main cities in China. We need to investigate if the policies launched by the local governments on controlling PM2.5 improved their air conditions. The raw data of PM2.5 in the five cities are from the website of U.S Department of State Mission China. The data for Shanghai is from 2011-2016; for Beijing is from 2008-2016; for Chengdu is from 2012-2016; for Shenyang is from 2013-2016 and for Guangzhou is from 2011-2016. We did some cleaning of the data and the imputation of the missing value. We transferred the hourly data to daily average. After getting the final data set, we use t-test, ANOVA and ACF methods to measure the change of concentration of PM2.5 (ug/m3) in these five cities under their related local policies.

1.2 Policies and Laws:

Due to a rapid economic development, industrial expansion and urbanization during the last few decades, there is an increasingly occurrence of haze or smog and the air pollution issue became a national-scale problem in China, especially in the most developed and high-populated city, including Beijing, Shanghai, Chengdu, Guangzhou and Shenyang.

High occurrence of extreme haze episodes and high concentration of PM2.5 in recent years not only leads to a global concern due to its adverse health effects, but also triggers the Chinese government to tackle the serious air quality problem especially PM2.5 pollution. From February 29th, 2012, the Chinese Ministry of Environmental Protection (MEP) published the third revision of the "the national ambient air quality standards" (NAAQS), in which PM2.5 is included into the NAAQS for the first time. PM2.5 monitoring has not yet been introduced in the national network in China before the new NAAQS, although PM2.5 levels have been reported in research studies in some developed cities such as Beijing, Shanghai, and Guangzhou (Zhang 2015).

In recent years, the national air quality monitoring network is continuously operated and maintained by Department of the Environment for each city. The network comprises of 496 stations in 74 cities since 2012, which is later extended to 946 monitoring stations in 190 cities from 2013. At each monitoring site, the real-time mass concentration of PM2.5 is measured by commercial instruments (Rodolfo 2015).

The 12th Five-Year Plan (FYP) of 2011-2015, which sets out both economic and political guidelines, represents a change emphasizing on environmental pollution as a problem related to the entire country. More recently, premier Li Keqiang has already indicated in his speech that the government will limit factory emissions of tiny harmful particulate matter (PM2.5) , a major cause of air pollution, down by 25% in 13th Five-Year Plan (FYP). This is the first time in the China's history that a specific PM2.5 target has been included in a FYP. Other main national policies and laws directly concerning air pollution are: National Air Quality Policy (2011-2015), the Clean Air Action Plan (2013-2017), the Environmental Protection Law (2015), the Air Pollution Prevention and Control Law (2015), the Law on Promoting Clean Production (2012), the Energy Conservation Law (2007), and the Environmental Impact Assessment Law (2003) (Rodolfo 2015).

1.2.1 Clean Air Action Plan

In order to improve the air quality, the municipal government of Beijing, Shanghai, and Chengdu promised to reduce the particle density by 25% or more on the PM 2.5 scale until 2017. Urban PM2.5 originates mainly from sources such as traffic-related emissions, biomass burning, as well as regional constructions dust. To archive this aim the local municipal government released a set of policies to cut vehicle emissions and industrial pollution in its five-year Clean Air Action Plan (2013-2017). This plan included targets for pollution prevention in sex sectors, which are energy, industry, transportation, construction, agriculture and social life. In this study, we use monthly data from January 2013 to September 2016, which is the most recent PM2.5 data we can get from Mission China, to check the performance of five-year Clean Air Action Plan in Beijing, Shanghai, and Chengdu. From the boxplots in Table 1 we can see the PM2.5 decreases year by year in all three cities. Until the end of September 2016, Beijing reduces its PM2.5 by 41%, Shanghai by 24% and Chengdu by 33%. All three cities reduced its

PM2.5 by 20%, among which Beijing and Chengdu reach the goal of reducing PM2.5 by 25% in advance. Therefore, we can conclude that the policy works well.

1.2.2 Driving Restriction Policy

A driving restriction policy, as a control-and-command rationing measure, is a politically acceptable policy tool to address traffic congestion and air pollution in some countries and cities. Beijing was the first city in China to implement this policy.

As the capital city of China, Beijing has experienced explosive growth during the last three decades. Market reforms, increased population and rising wealth have fuelled strong demand for mobility, which in turn has reshaped the lifestyle of residents. The total number of automobiles in Beijing jumped sharply from 2.583 million in 2005 to 4.809 million in 2010, of which about 3.207 million were owned by individual households (Beijing Transportation Research Center, 2011a). However, this rapid increase in mobility comes at a high price. Beijing is already one of the world's most polluted cities, and vehicular emissions contribute heavily to local air pollution.

The Beijing Municipal Government has adopted a series of policy instruments in an attempt to reduce the transportation emissions in the city. The policy instruments by municipal government include investing in public transportation (building new subways and reducing bus fare), controlling automobile ownership (lottery policy), restricting car use (driving restriction policies), or increasing the cost of private transport (such as parking fees and fuel taxes) to push residents to switch from private automobiles to public transit. Nevertheless, among all of these transportation policies, the driving restriction policy stands out as the most prominent instrument favored by the Beijing municipal government.

In preparation for the 2008 Olympics, Beijing imposed rules barring cars from the central city from 6:00 a.m. until midnight by restriction cars depending on whether the last number of the license plate was even or odd, based on license plate number. From July 1st to September 20, 2008 (80 days), the driving restriction rules were used to control traffic and air pollution in support of the Olympics. With this program, it was calculated that the PM2.5 in Beijing was reduced by 28.87%, from 100.55 ug/m^3 to 71.52 ug/m^3, during the 80 days restriction period. To compare the concentration of PM2.5 before this restriction policy with that under the restriction policy, we draw a boxplot using 80 days' data before the program and 80 days' data under the program. It is clearly to see the PM2.5 level was reduced significantly from Table 2 .

These temporary restrictions were followed by similar restrictions that remained in place after the Olympics were over. On September 28, 2008, Beijing officials announced a new driving restriction measure to be enforced from October 11, 2008 to April 10, 2009. This new measure put last digit numbers of the license plates into five groups: 1 and 6, 2 and 7, 3 and 8, 4 and 9, and 5 and 0. Private automobiles with the last digit numbers in each group were forbidden from using the road inside, but excluding, the 5th ring road on a designated weekday from 6:00 a.m. to 9:00 p.m (Wang, Xu, Zheng, and Qin 2013).

Beijing was the first city in China to implement a driving restriction policy in order to control traffic volume and air pollution. More recently, several Chinese cities from north to south, including Changchun, Lanzhou, Hangzhou, Guiyang, and Chengdu, began to rely on driving restrictions to alleviate traffic congestion.

Starting from Jan 1st, 2016, we noticed that Guangzhou launched a rule to bans all out of town vehicles depending on whether the last number of the license plate was even or odd in the next nine months as a temporary driving restriction. In order to see if there is any difference in the concentration of PM2.5 during the same nine months in 2015 and 2016, we applied two samples T-test procedure. As shown in Table 3, the P-value is much smaller than 0.05, which indicates that there is a statistical significant difference between the mean of PM2.5 in 2015 and the mean of PM2.5 in 2016, though the boxplot in Table 3 does not clearly show the difference. Two sample T-test reveals that the PM2.5 was reduced significantly in Guangzhou after the driving restriction plan. However, the PM2.5 before the restriction is quite low comparing to other four cities, so we would rather say that the aim of restriction is to reduce the traffic pressure than reduce PM2.5.

In the end, we compare the change of PM2.5 of five cities from 2013 to 2016. The result shows that Beijing and Guangzhou both reduce PM2.5 over 40%, while the level of PM2.5 in Guangzhou is the lowest and Chengdu is the highest in 2016 (as shown in Table 4).

In summary, all the policies and laws are targeted to reduce emissions from industry, transport, outdoor agricultural waste and indoor biomass waste. To control the emissions from industry, the government need to control emissions from coal burning industries, boost more efficient use of coal and renewable energy, and promote the use of electricity and natural gas. To control the emissions from transport, the government need to control the number of vehicles, increase the cost of driving a car, take management of vehicles from other cities, set emission standard and promote clean energy vehicles. To control the emissions from agricultural waste, the government need to ban on straw burning inside cities and encourage the use of waste straw as a resource. To control emissions from indoor biomass, the government need to control emissions from cooking and heating, and promote clean cooking fuels and electrification. The last and most important thing concerning the policies is to put the policies and laws really into force, not just list them on the blueprint.

2 Methodology

2.1 Time series forecasting

Time Series Analysis, comprises methods for analyzing time series data in order to extract meaningful statistics and other characteristics of the data. Time Series forecasting, is the use of a model to predict future values based on previously observed values. In this project, time series method is used to predict the concentration of PM2.5 for the next 12 months in monthly average value, in order to test if the local policies efficient to each city.

2.2 Arima Model

An autoregressive integrated moving average (ARIMA) model is a generalization of an autoregressive moving average (ARMA) model. Both of these models are fitted to time series data either to better understand the data or to predict future points in the series. ARIMA models are applied in some cases where data show evidence of non-stationarity, where an initial differencing step can be applied to reduce the non-stationarity. In this project the Arima model is used to build up the time series model with decomposed data which has removed the seasonal trend and effect.

2.3 Cross Correlation

In signal processing, cross-correlation is a measure of similarity of two series as a function of the displacement of one relative to the other. This is also known as a sliding dot product or sliding inner-product. It is commonly used for searching a long signal for a shorter, known feature. In this project the cross correlation method is used to analyze the inter effect of the concentration of PM2.5 between Beijing and Shanghai.

3 Data Analysis

3.1 Time Series Analysis

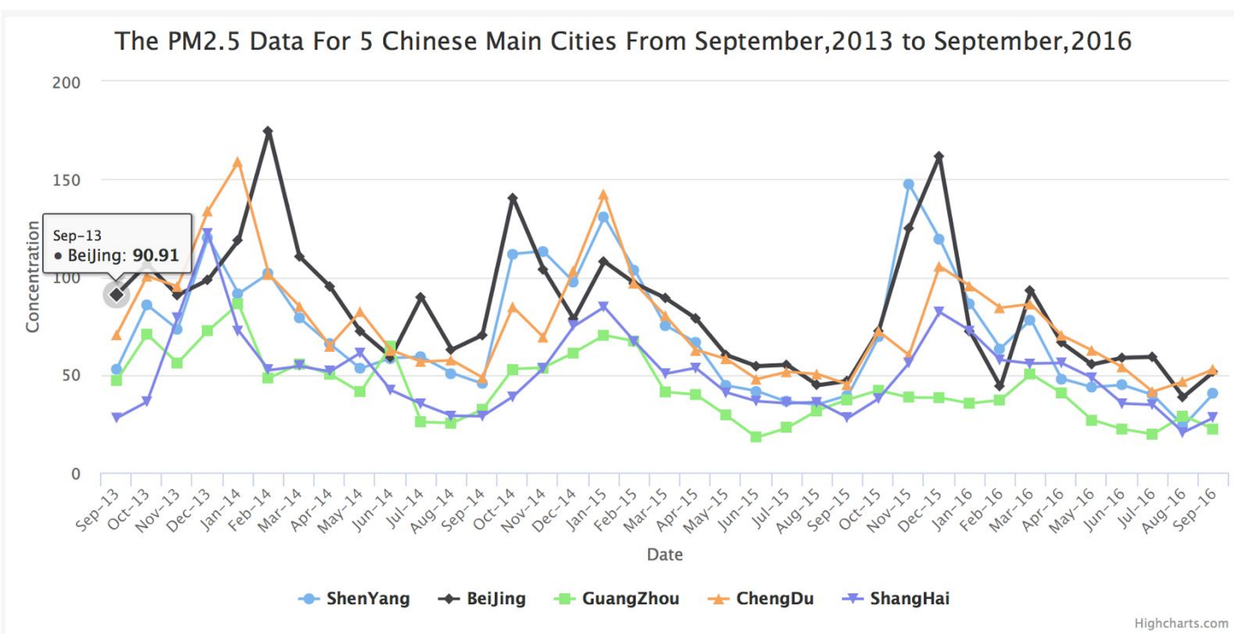
Definition Explanation

Time Series Analysis, comprises methods for analyzing time series data in order to extract meaningful statistics and other characteristics of the data.

Time Series forecasting, is the use of a model to predict future values based on previously observed values.

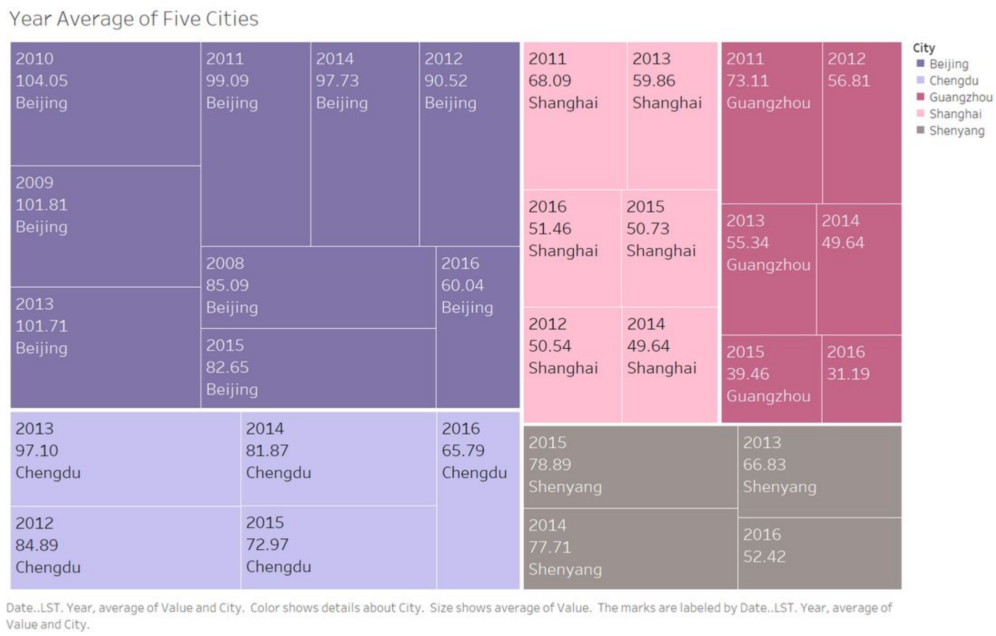
We will follow five steps here to realize the prediction for the future analysis. These five steps are: visualize the time series; stationarize the series; plot ACF/PACF charts and find optimal parameters; build the ARIMA model; make predictions. Since we are interested in knowing the cross relationships among all the cities, we have a cross-correlation analysis later after doing the prediction.

3.1.1 Visualize The Time Series



Plot 1. Line Chart of PM 2.5 of 5 cities along the years

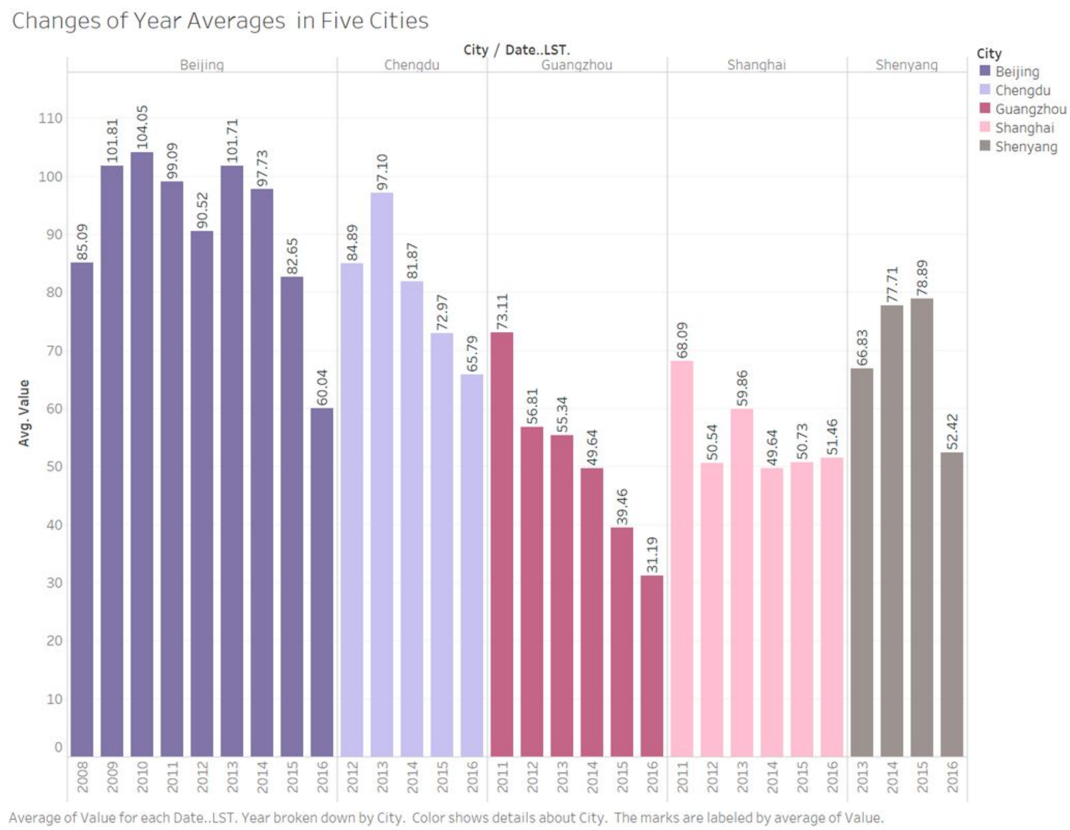
To show the five cities' concentration changes along the years since 2013, we used highcharts.com created this line chart with each dot representing the 13th day's pm2.5 concentration in that month, and each line represents different cities. Since each city have different time frame regarding the dataset, we only have all of them started in 2013. Through this plot, we could find how are these cities perform differently at different time in this month.



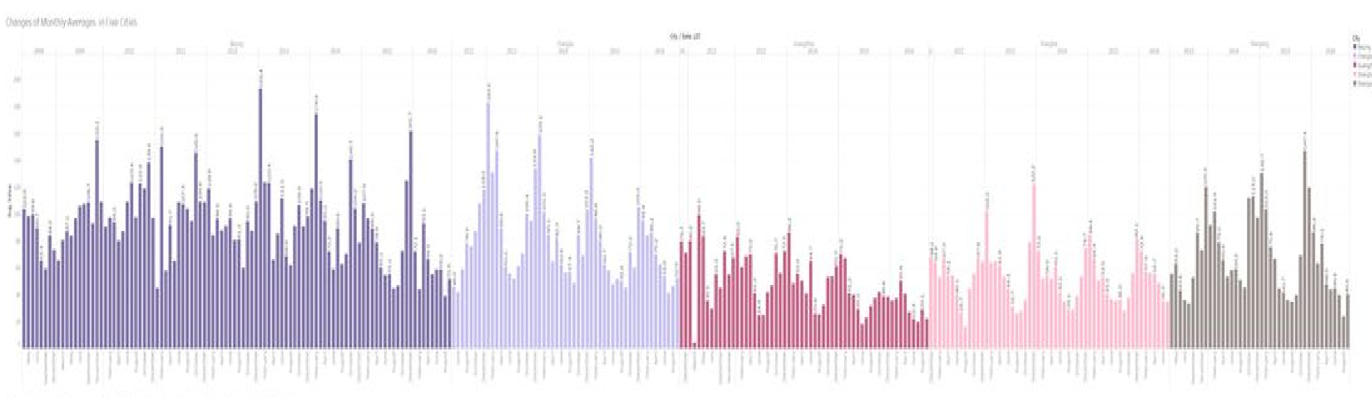
Plot 2. Tree Map of Yearly Average of 5 Cities

To compare the yearly changes among each cities. We made this tree map to show how each city is performing among its time frame. The position and size of the rectangle represents the order of the sample. Each color represents different cities while the most left top corner is the highest value and the most right bottom corner is the lowest value. Since we used yearly averages in this plot, we can see that Beijing has the biggest dataset as well the highest yearly averages. We also got the information that 2010 in Beijing was the worst year with a 104.05ug/m3 yearly averages. One thing to pay attention to is that, we only have partial dataset in 2016 since the dataset is still updating, we should consider the fact that winter time's concentration has not been considered yet in this comparison. Once the winter values been added, the averages of 2016 would increase a lot.

More information is showed in the bar chart below. Still having the yearly averages representing each bar, we could easily noticed that Guangzhou has the best decreasing trend and lowest yearly averages in general. Shanghai appears to be the second best among the five cities but it keeps the same and seems to be a little bit increasing after the highest year of 2013. Still, we need to consider the fact that averages would increase of 2016 due to the lack of winter data. We only have four years' dataset for Shenyang and its yearly averages varies a lot. It is hard to see the trend for the following years but it has a relatively low start at the year of 2013. Chengdu has a relatively high start at the year of 2012 and reaches its highest value at 2013. Then Chengdu is performing well on the trend of decreasing. Beijing appears to be the worst city among these five cities. It has an averages of over 90 ug/m3 while 2008 and 2015 has a year average below 90 ug/m3. It is also very hard to predict the future performance based on the existing dataset. We have also made a monthly averages plot so that we can see the seasonal effect directly.

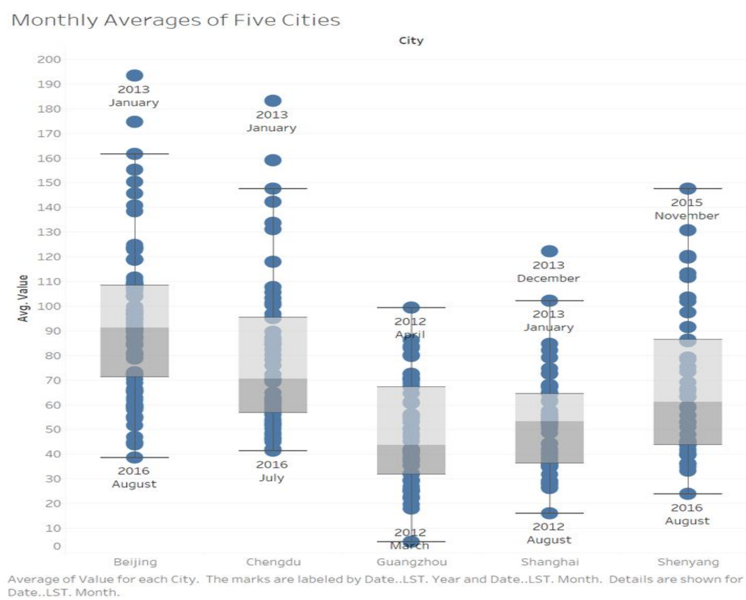


Plot 3. Bar Chart of Yearly Average of 5 Cities



Plot 4. Bar Chart of Monthly Average of 5 Cities

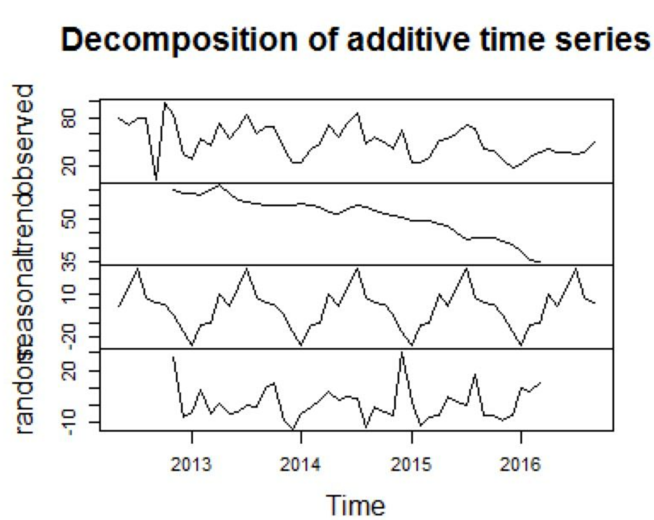
With monthly averages of five cities, we have also made this boxplot to show the ranges of each city. Still, we can see that Guangzhou has the lowest averages while Beijing has the highest averages and the biggest ranges. The highest and lowest values were all printed in this plot and we noticed that while all the other cities have their highest value in winter time around January and lowest value in summer time around August, Guangzhou had its highest value in April 2012 and its lowest value just one month ago. This was an interesting point that we could dig further for future research.



Plot 5. Boxplot of Monthly Averages of Five Cities

3.1.2 Stationarize the Series

As we noticed from above plots, all cities have a strong seasonal trend among the years. Thus, when we are building the time series model, we should eliminate the seasonal effect and trend, use the random value to build the model and then add on the seasonal effect and trend to predict for future values. We could use R to realize the decomposition. We are putting a decomposition result of Guangzhou here as an example. Please see Appendix for the other cities.



Plot 6. Decomposition of Observed Data in GuangZhou

In this plot, from top to bottom are: the observed data, the trend, the seasonal effect, and the random. We can see that trend and seasonal has a strong effect on the observed data.
After we have acquired the random series, we are testing for its stationary. When it is stationary in time series model, its properties do not depend on the time at which the series is observed. Thus, stationary is the common assumption of time series models.
We used Augmented Dickey-Fuller test to test the random values and they all turn out to be stationary at 95% confidence. Tests results are attached in the Appendix. Since we have proved the stationary, all cities will be discussed one by one as following.

3.2 Plot ACF/PACF charts and Build ARIMA model and Prediction

All the ACF and PACF plots are attached in the Appendix Table 5.

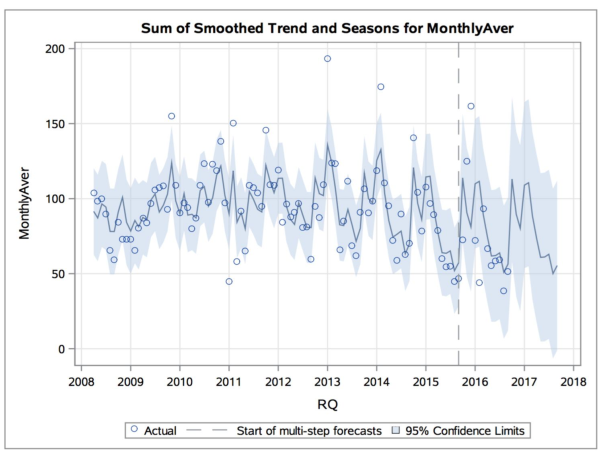
3.2.1 Beijing

By using decomposed data, we built ARIMA model in SAS, then the value p+d and q is decided. p+d=1, q=1.
Then the model is decided,

Autoregressive Factors: 1-0.51881B**(1)

MovingAverageFactors: 1-0.16749B**(1) .

Based on this model, we added the trend and seasonal effect back to do forecasting. The result is shown in the following. The until 9, 2017, the value of PM 2.5 in BeiJing will change around 75 ug/m3.



Plot 7. The forecasting plot for BeiJing

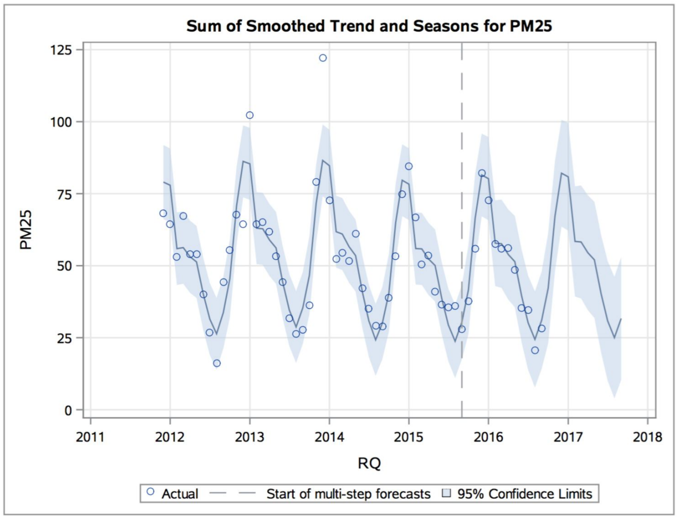
3.2.2 Shanghai

By using decomposed data, we built ARIMA model in SAS, then the value p+d and q is decided. p+d=3, q=1.
Then the model is decided,

Autoregressive Factors: 1-1.43102B**(1)+0.66524B**(2)+0.05689B**(3)

MovingAverageFactors: 1-0.82483B**(1) .

Based on this model, we added the trend and seasonal effect back to do forecasting. The result is shown in the following. The until 9, 2017, the value of PM 2.5 in ShangHai will change around 50 ug/m3.



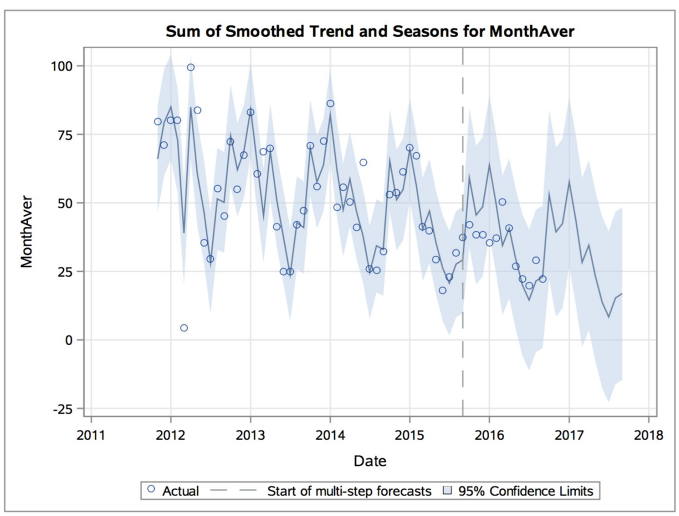
Plot 8. The forecasting plot for ShangHai

3.2.3 Guangzhou

By using decomposed data, we built ARIMA model in SAS, then the value p+d and q is decided. p+d=1, q=0. Then the model is decided,

Autoregressive Factors: 1-0.42309B**(1) .

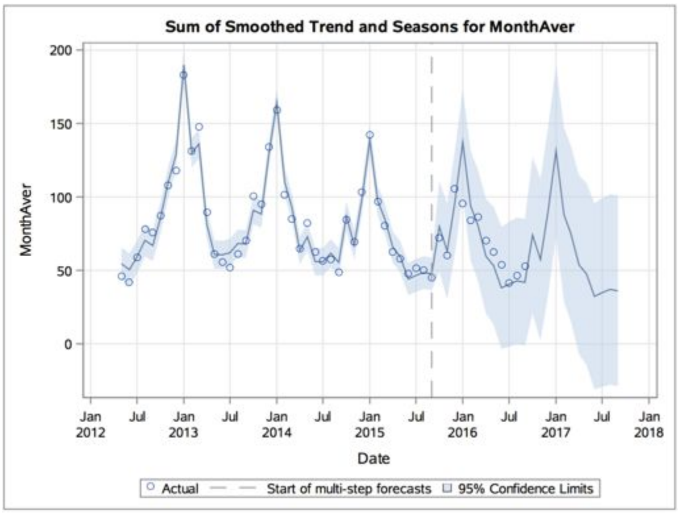
Based on this model, we added the trend and seasonal effect back to do forecasting. The result is shown in the following. The until 9, 2017, the value of PM 2.5 in GuangZhou will change around 30 ug/m3.



Plot 9. The forecasting plot for GuangZhou

3.2.4 Chengdu

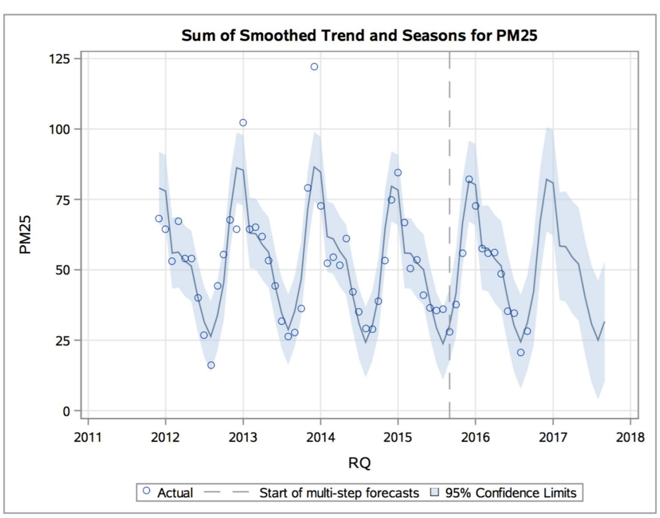
By using decomposed data, we built ARIMA model in SAS, then the value p+d and q is decided. p+d=5, q=6. Then the model is decided,
Autoregressive Factors: 1-0.28043B**(1)-0.21716B**(2)-0.01084B**(3)-0.03999B**(4)+0.6348B**(5)
MovingAverageFactors:
1+0.35646B**(1)+0.23737B**(2)-0.01443B**(3)-0.28273B**(4)+0.57764B**(5)+0.10453B**(6) .
Based on this model, we added the trend and seasonal effect back to do forecasting. The result is shown in the following. The until 9, 2017, the value of PM 2.5 in ChengDu will change around 75 ug/m3.



Plot 10. The forecasting plot for ChengDu

3.2.5 Shenyang

By using decomposed data, we built ARIMA model in SAS, then the value p+d and q is decided. p+d=6, q=4. Then the model is decided,
Autoregressive Factors:
1-0.89092B**(1)+0.64806B**(2)-0.24033B**(3)-0.53984B**(4)+0.2873B**(5)+0.03028B**(6)
MovingAverageFactors: 1-0.04285B**(1)+0.60583B**(2)+0.28228B**(3)-0.34614B**(4) .
Based on this model, we added the trend and seasonal effect back to do forecasting. The result is shown in the following. The until 9, 2017, the value of PM 2.5 in ShenYang will change around 60 ug/m3.



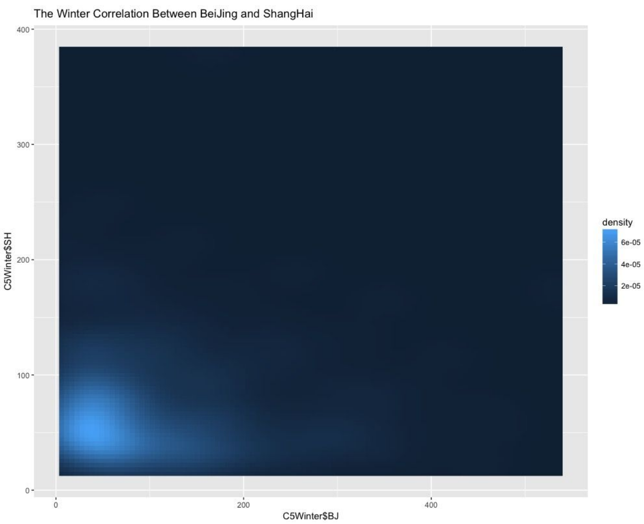
Plot 11. The forecasting plot for ShenYang

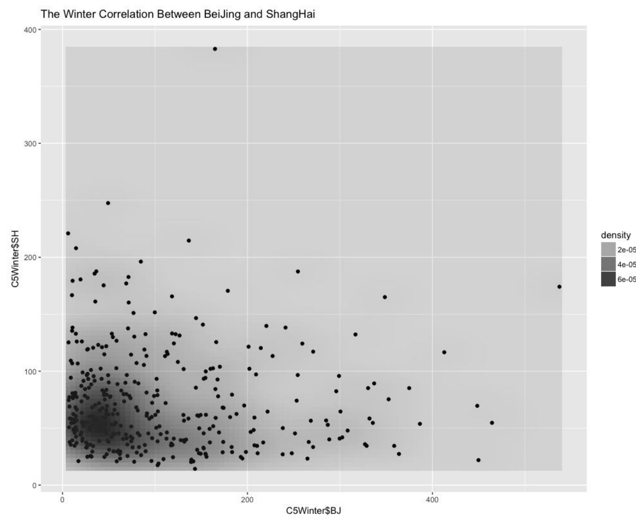
Generally speaking, among the five cities, the air quality in GuangZhou is the best one from current to future.

3.3 Cross-Correlation

3.3.1 BeiJing to ShangHai

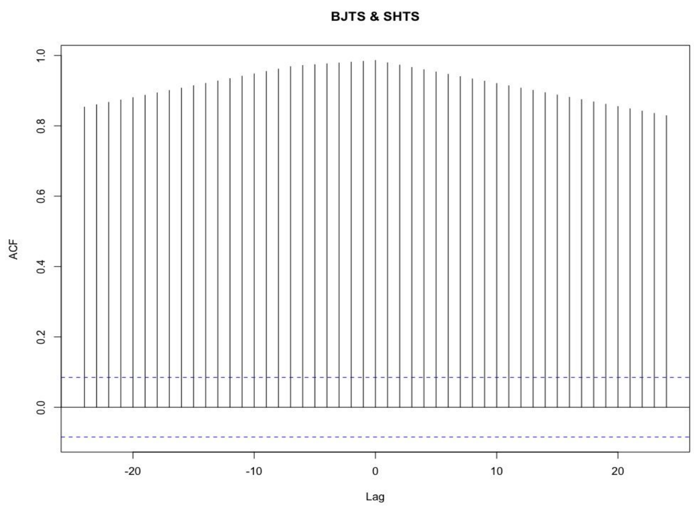
In this part, the client required us to test the cross-correlation between Beijing and Shanghai during Winter Time. According to the total plot for BeiJing to ShangHai from winter, 2012 to winter, 2016, the higher the PM 2.5 value of BeiJing is, the higher that of ShangHai is, when the PM 2.5 value is less than 100 ug/m3.





Plot 12. The Density plot for BeiJing and ShangHai at Winter time ; Plot 13. The Density plot for BeiJing and ShangHai at Winter time

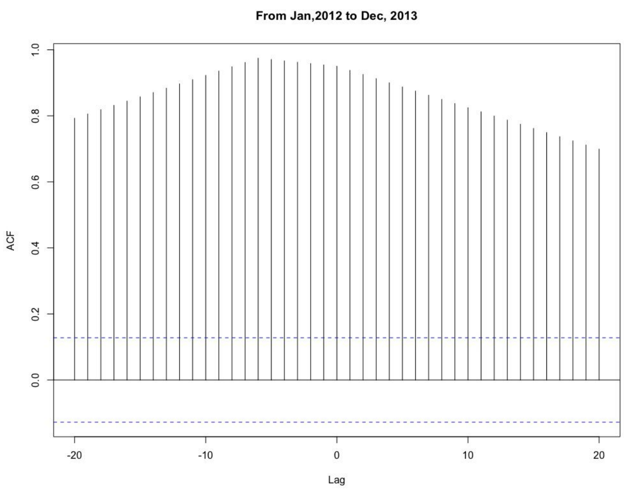
And for both cities, most of winter days have a concentration value within 100ug/m3, the density is 6*10^5. The following is the CCF plot, the details values are attached in the Appendix.



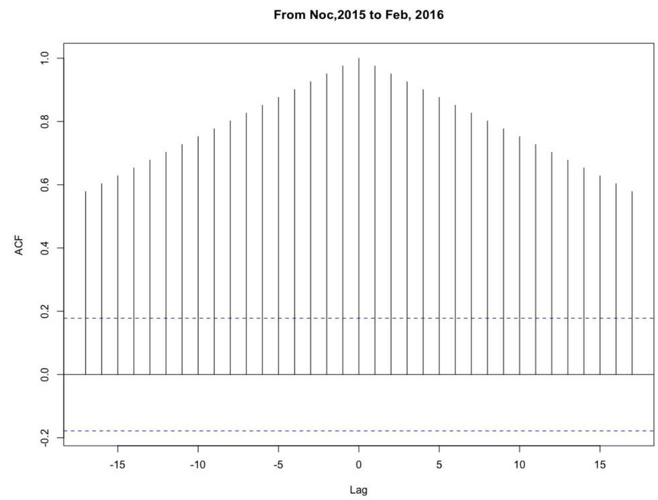
Plot 14. The CCF of BeiJing and ShangHai

When Lay=-1, the coefficient got max value, 0.921.
That means, the PM 2.5 of Shanghai will be affected by that of Beijing after 1 day. Actually, for the next 20 days, the coefficient was higher than 0.85, a very high value.
Hence, Shang Hai nearly gets more than 85% influence of PM 2.5 from Beijing for the next 20 days. The PM 2.5 in Beijing is too gargantuan to reduce much after long journey. The PM2.5 of BeiJing has huge influence on that of ShangHai in winter. Hence, it will be good for the crowd in ShangHai to walk with breathing mask during the winter.

However, this conclusion will cover the effect of government policy, it's not objective. Therefore, the CCF among 2 time periods are tested separately.
The first is from Jan, 2012 to Dec, 2013; the second is from Nov, 2015 to Feb, 2016.
For the first time period, ShangHai got largest PM2.5 effect, 0.974, from BeiJing after 6 days. What's more for the next 13 days, the coefficient is higher than 0.80.
For the second time period, ShangHai got largest PM2.5 effect, 0.975, from BeiJing after 1 day. However, after 8 days, the coefficient is less than 0.80.
Hence, we believe that the policy works well to control the PM 2.5 concentration for ShangHai in Winter. Additionally, the situation will be better if the relative policy is lasting for the future.



Plot 15. The CCF plot from Jan, 2012 to Dec, 2013;



Plot 16. The CCF plot from Nov, 2012 to Feb, 2013

4 Conclusion

Based on the data analysis, the air quality in Beijing appears to be the worst city among all five cities. The PM2.5 in Beijing has an yearly averages of 101.7 ug/m3 in 2013 and has a yearly average of 59.9 ug/m3 in 2016. The five year Clean Air Action Plan and driving restriction policy stand out among lots of related air quality control policies in Beijing. Those policies significantly reduced the PM2.5 by over 40% from 2013 to 2016. The driving restriction policy in its first 80 days reduced the PM2.5 dramatically by 28%. Based on the dataset and our statistical analysis, the performance of policies in Beijing is impressive, however, it is very hard to predict the future performance based on the existing dataset. Conversely, air quality in Guangzhou is the best among five cities, with the lowest PM2.5 of 32 ug/m3 in 2016. Since the PM2.5 in Guangzhou before the driving restriction policy is quite low comparing to the other four cities, we would rather say that the aim of restriction is to reduce the traffic pressure than reduce PM2.5. Until the end of September 2016, Beijing reduced its PM2.5 by 41%, Shanghai by 24% and Chengdu by 33%. All three cities reduced its PM2.5 by 20%, among which Beijing and Chengdu reach the goal of reducing PM2.5 by 25% in advance. Therefore, we can confidently conclude that the Clean Air Action Plan and other policies jointly play an important role in reducing the PM2.5.

Meanwhile, given that the data from five cities are stationary, it's convenient for us to launch time series analysis. According to time series analysis, we know that:
until 9, 2017, the value of PM 2.5 in BeiJing will change around 75 ug/m3;
the value of PM 2.5 in ShangHai will change around 50 ug/m3;
the value of PM 2.5 in GuangZhou will change around 30 ug/m3;
the value of PM 2.5 in ChengDu will change around 75 ug/m3;
the value of PM 2.5 in ShenYang will change around 60 ug/m3.

For the next 12 months, the air quality in GuangZhou may be still the best one among five cities. About the cross-correlation from BeiJing to ShangHai, the situation is bad along the whole time. But if we separated the time and did the test, we found that from Jan, 2012 to Dec, 2013, ShangHai got largest PM2.5 effect, 0.974, from BeiJing after 6 days. What's more, for the next 13 days, the coefficient is higher than 0.80.

From Nov, 2015 to Feb, 2016, ShangHai got largest PM2.5 effect, 0.975, from BeiJing after 1 day. However, after 8 days, the coefficient is less than 0.80. According to the Clean Air Action Plan, until the end of September 2016, Beijing reduced its PM2.5 by 41%, people have reason to believe that BeiJing's controlling policy is effective for ShangHai on reducing the lasting days of PM 2.5 in air at Winter. What's more, the reducing percentage of ShangHai is 24%, which is very close to the target percent, 25%. Therefore, a hypothesis is drawn that during ShangHai's Winter, from Nov, 2015 to Feb, 2016, it took 8 days to reduce the coefficient gotten from BeiJing lower than 0.8, because of the combined action of both BeiJing and ShangHai on Clean Air Action Plan.
Hence, we are confident to say that the policy works well to control the PM 2.5 concentration for ShangHai in Winter. However, the coefficient, 0.8, is still very large, the 8 days is still so long. The government still has a long journey to experience on controlling PM 2.5.
Even so, we still believe that the situation will be better if the relative policy is lasting for the future.

References

1. Zhang, Y.-L. and Cao, F. Fine particulate matter ($PM_{2.5}$) in China at a city level. *Sci. Rep.* 5, 14884; doi: 10.1038/srep14884 (2015).
2. Rodolfo Andres Hernandez, Prevention and Control of Air Pollution in China: A Research Agenda for Science and Technology Studies, *S.A.P.I.EN.S* [Online], 8.1 | 2015, Online since 15 December 2015, connection on 14 December 2016. URL: http://sapiens.revues.org/1734.
3. Lanlan Wang, Jintao Xu, Xinye Zheng, and Ping Qin "Will a Driving Restriction Policy Reduce Car Trips? A Case Study of Beijing, China". Environment for Development Discussion Paper Series. September 2013 EfD DP 13-11.

Appendix

Table 1

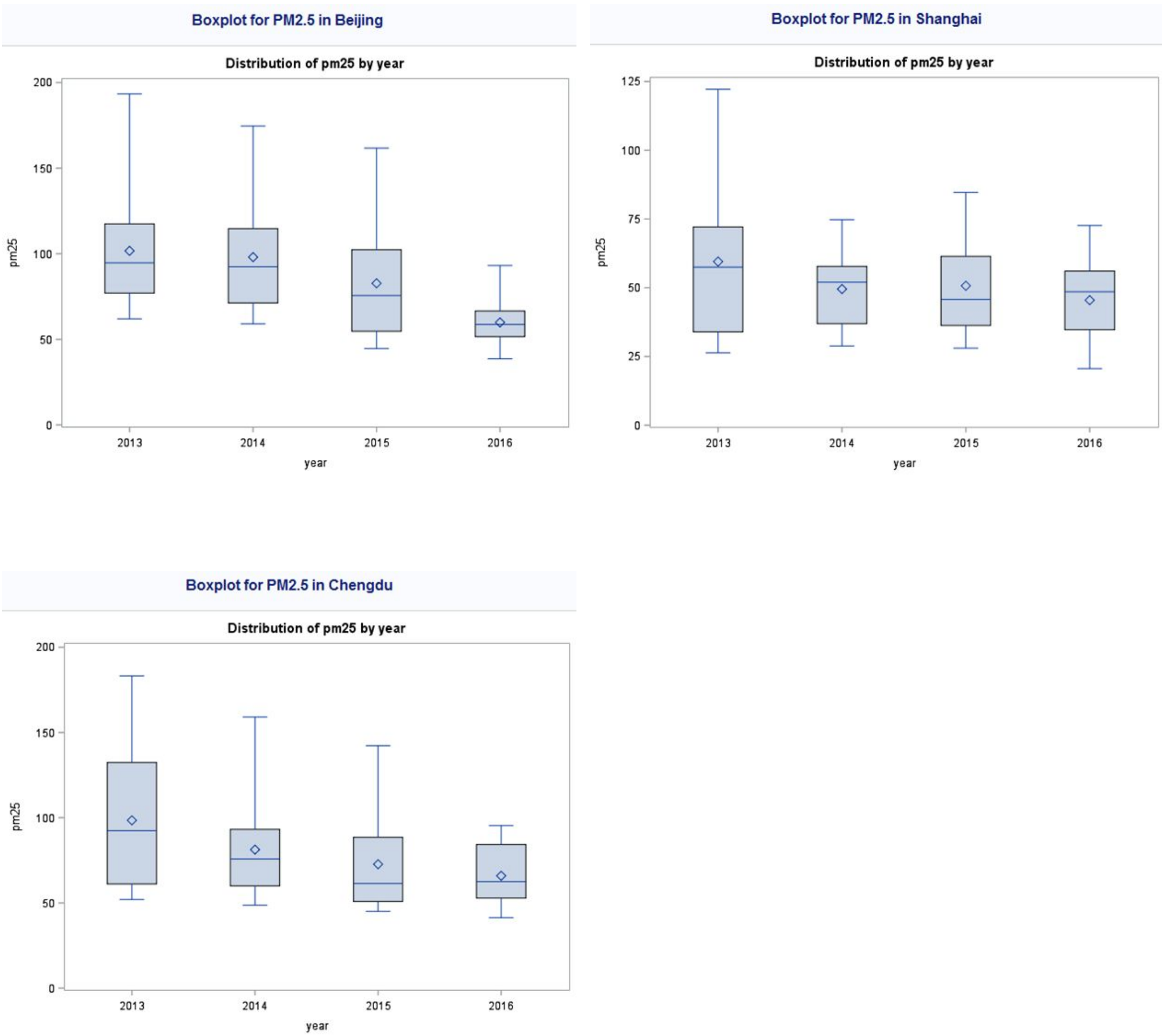


Table 2

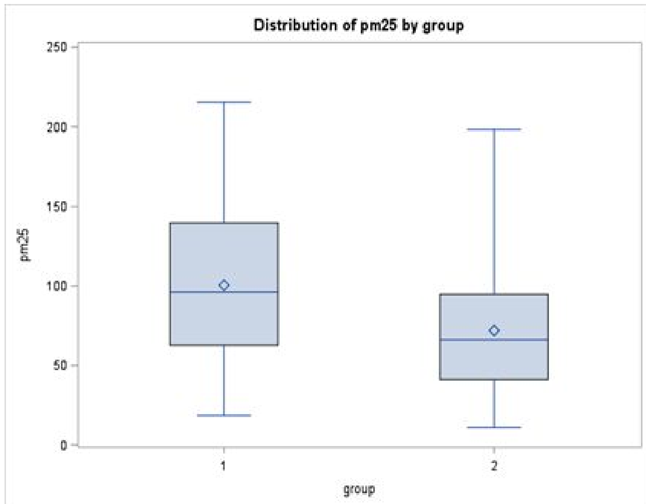
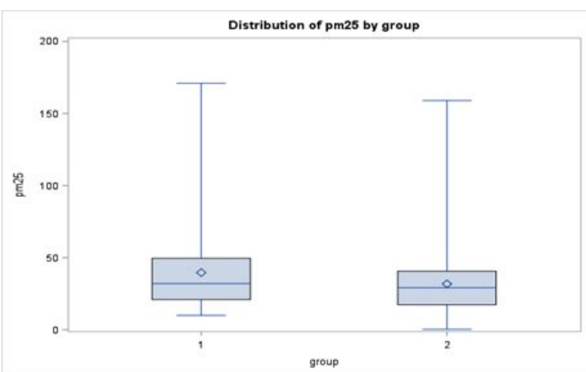


Table 3



Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	540	3.84	0.0001
Satterthwaite	Unequal	510.06	3.84	0.0001

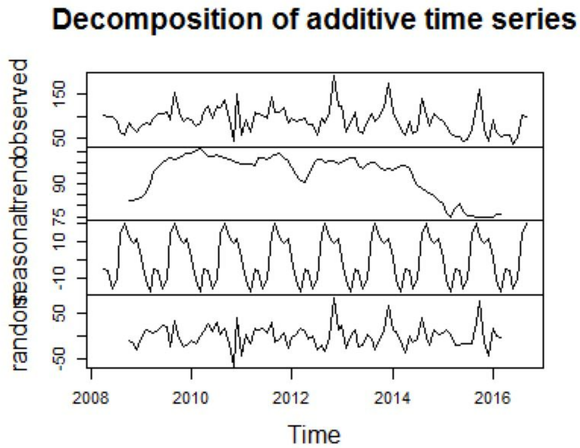
Equality of Variances				
Method	Num DF	Den DF	F Value	Pr > F
Folded F	272	268	1.69	<.0001

Table 4

City	PM2.5 2013	PM2.5 2016	Change
Beijing	101.70	59.90	41.10%
Shanghai	59.51	45.47	23.59%
Chengdu	98.46	65.91	33.06%
Guangzhou	55.39	31.82	42.55%
Shenyang	66.69	53.33	20.03%

Table 5

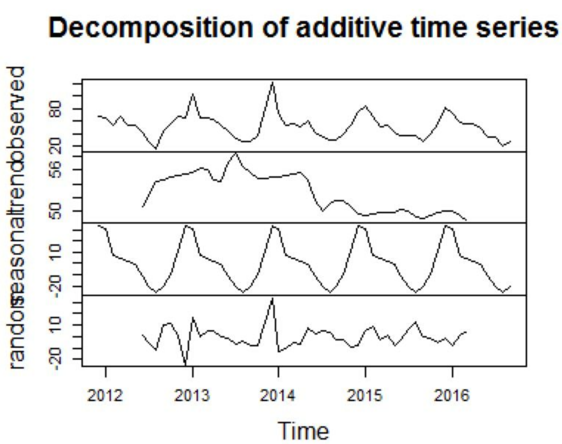
Table 5.1 BeiJing




```
        Augmented Dickey-Fuller Test

data:  na.omit(randombj)
Dickey-Fuller = -5.9391, Lag order = 4, p-value = 0.01
alternative hypothesis: stationary
```

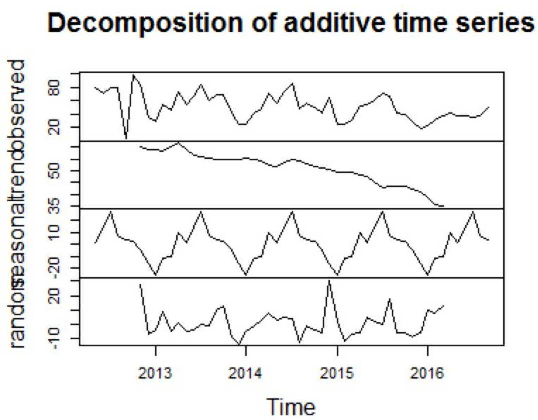
Table 5.2 ShangHai




```
        Augmented Dickey-Fuller Test

data:  na.omit(randomsh)
Dickey-Fuller = -5.3401, Lag order = 3, p-value = 0.01
alternative hypothesis: stationary
```

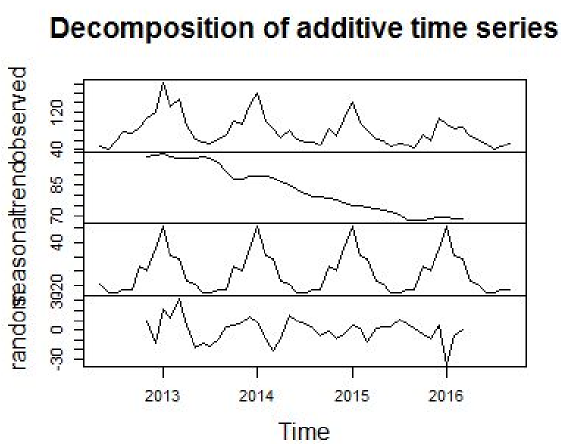
Table 5.3 GuangZhou




```
Augmented Dickey-Fuller Test

data:  na.omit(randomgz)
Dickey-Fuller = -6.5932, Lag order = 3, p-value = 0.01
alternative hypothesis: stationary
```

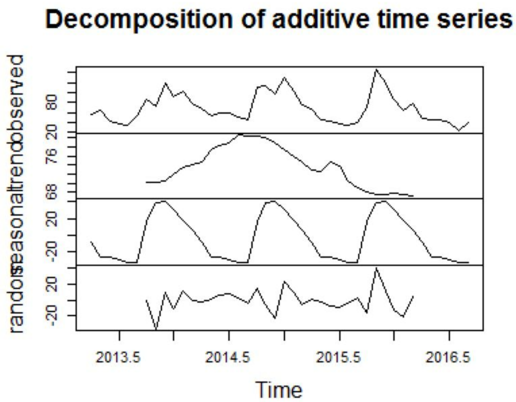
Table 5.4 ChengDu




```
Augmented Dickey-Fuller Test

data:  na.omit(randomcd)
Dickey-Fuller = -4.4844, Lag order = 3, p-value = 0.01
alternative hypothesis: stationary
```

Table 5.5 ShenYang



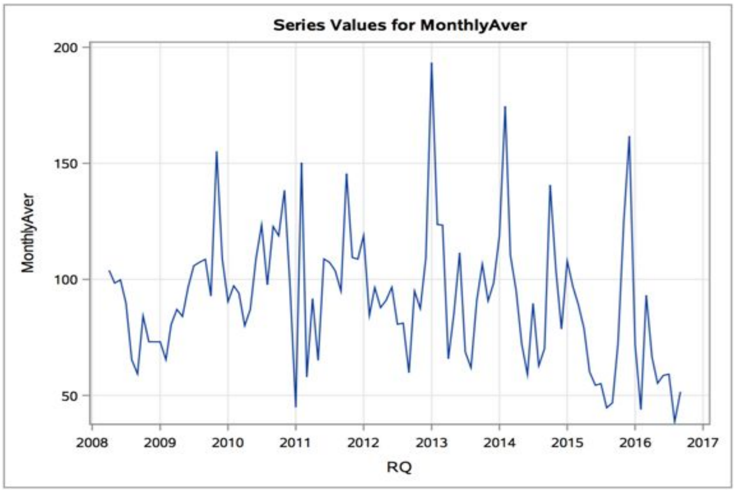

```
        Augmented Dickey-Fuller Test

data:  na.omit(randomsy)
Dickey-Fuller = -3.489, Lag order = 3, p-value = 0.06343
alternative hypothesis: stationary
```

Table 6

Table 6.1 BeiJing

Overall time plot and Deciding p+d and q



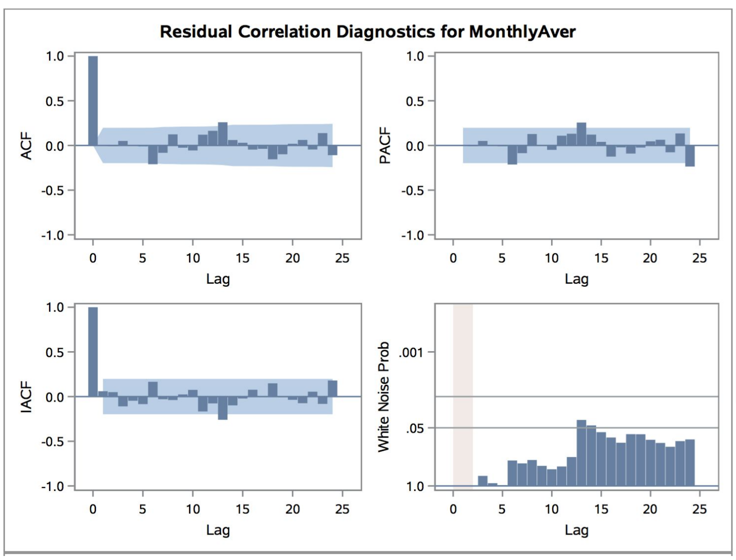
Error series model: AR(12)

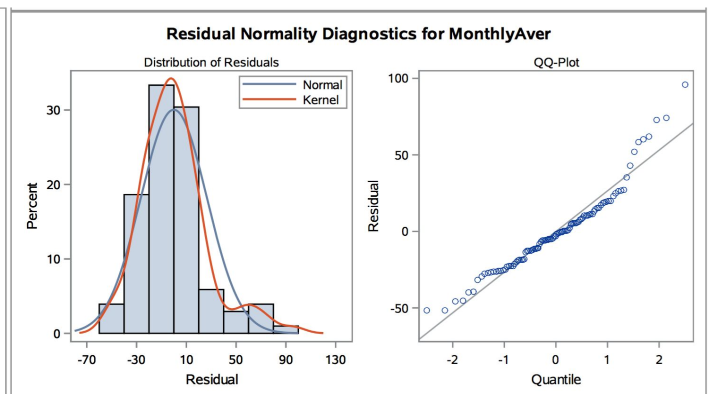
Minimum Table Value: BIC(1,1) = 6.561196

ARMA(p+d,q) Tentative Order Selection Tests		
SCAN		
p+d	q	BIC
0	1	6.653008
6	0	6.72992

(5% Significance Level)

ACF, PACF and Residual Plot



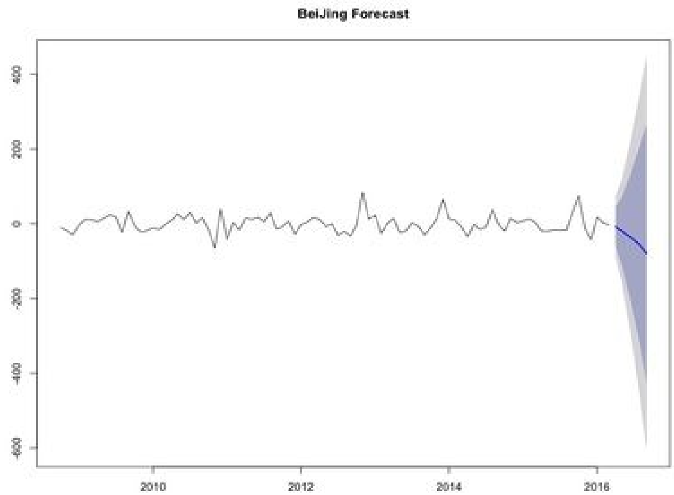


The ARIMA Model and Decomposed Data

The ARIMA Procedure

Autoregressive Factors	
Factor 1:	1 - 0.51881 B**(1)

Moving Average Factors	
Factor 1:	1 - 0.16749 B**(1)



The forecast plot by original data and the forecast plot with decomposed data.

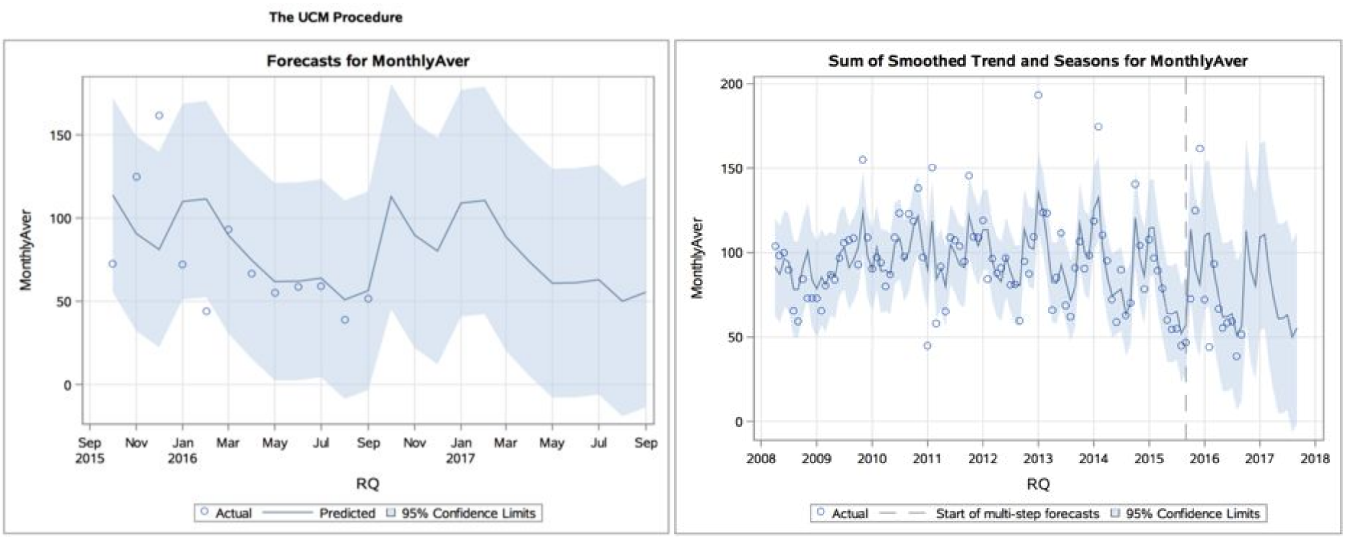
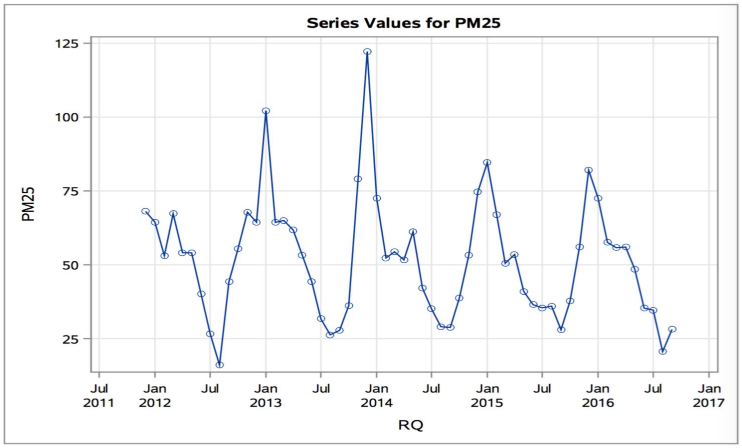


Table 6.2 ShangHai

Overall time plot and Deciding p+d and q



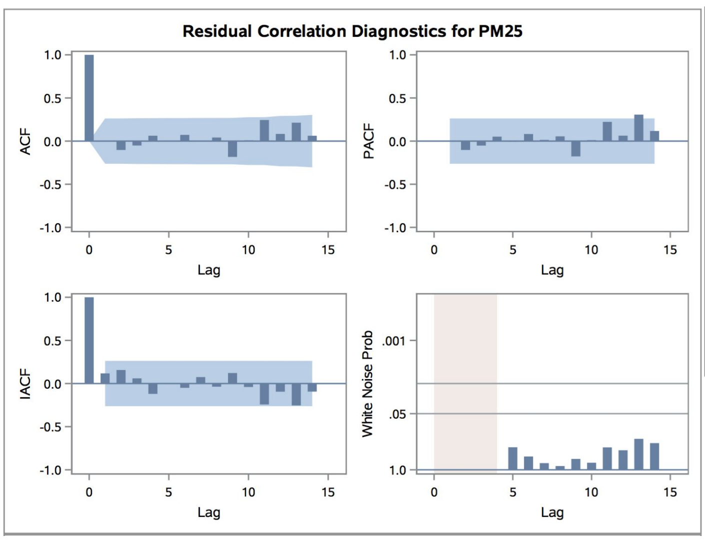
Error series model: AR(11)

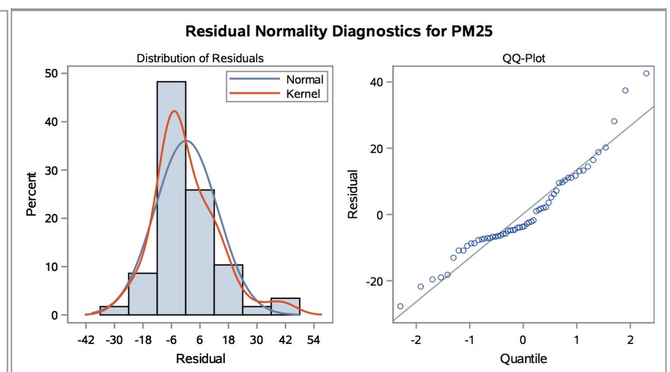
Minimum Table Value: BIC(3,1) = 4.883017

ARMA(p+d,q) Tentative Order Selection Tests		
SCAN		
p+d	q	BIC
2	1	5.099721
1	2	5.335244

(5% Significance Level)

ACF, PACF and Residual Plot



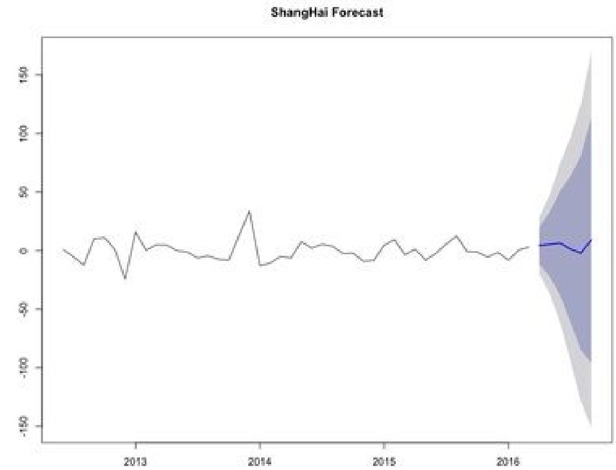


The ARIMA Model and Decomposed Data

The ARIMA Procedure

Autoregressive Factors	
Factor 1:	1 - 1.43102 B**(1) + 0.66524 B**(2) + 0.05689 B**(3)

Moving Average Factors	
Factor 1:	1 - 0.82483 B**(1)



The forecast plot by original data and the forecast plot with decomposed data

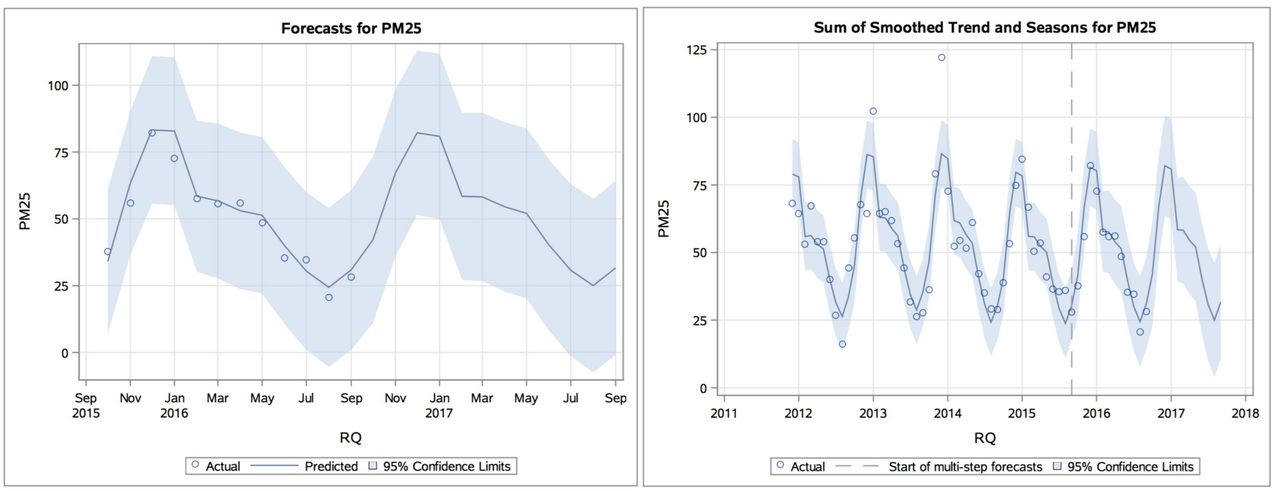
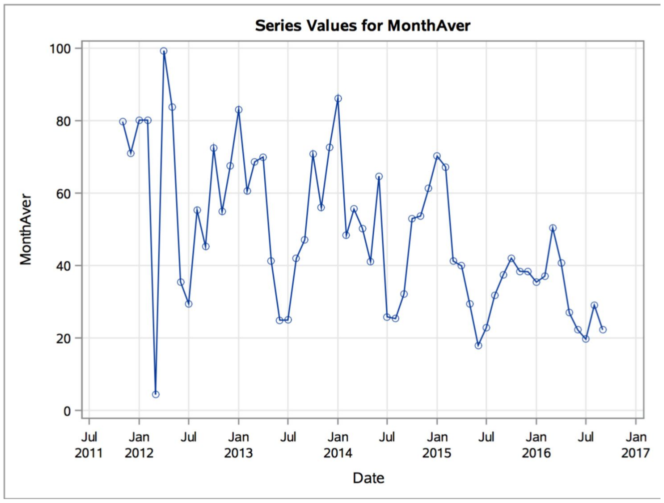


Table 6.3 GuangZhou

Overall time plot and Deciding p+d and q



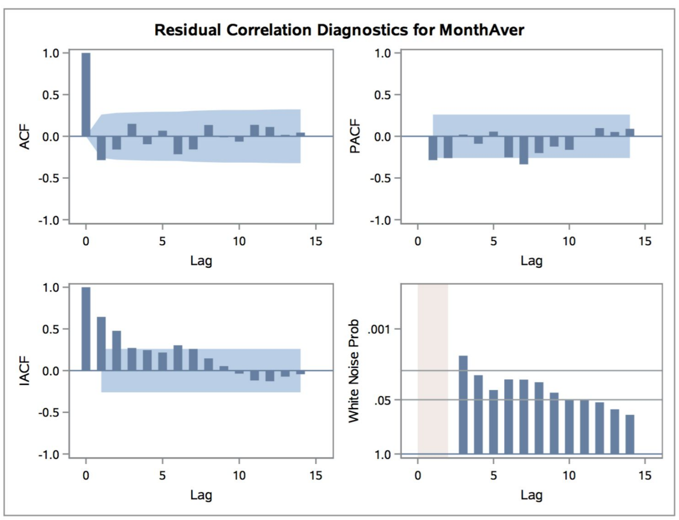
Error series model: AR(12)

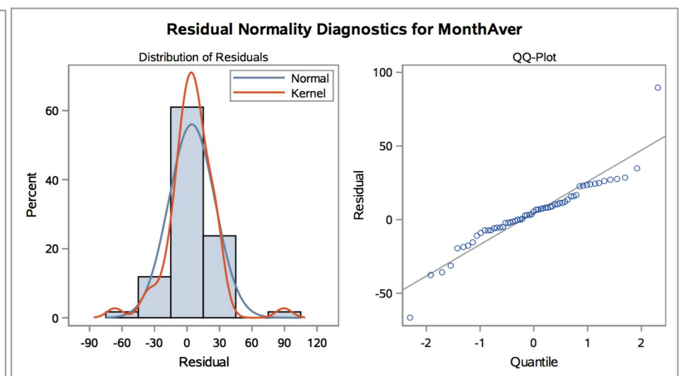
Minimum Table Value: BIC(1,0) = 5.010649

ARMA(p+d,q) Tentative Order Selection Tests		
SCAN		
p+d	q	BIC
1	0	5.010649
0	1	5.34866

(5% Significance Level)

ACF, PACF and Residual Plot



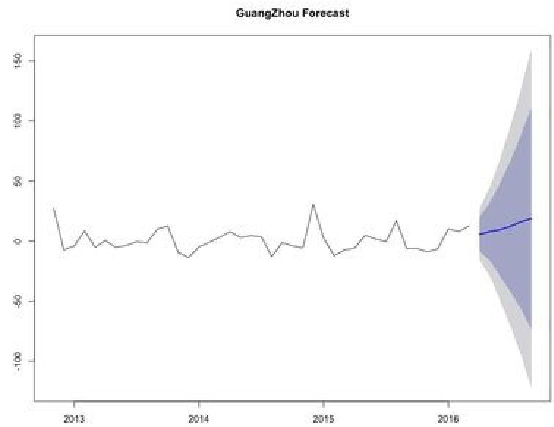


The ARIMA Model and Decomposed Data

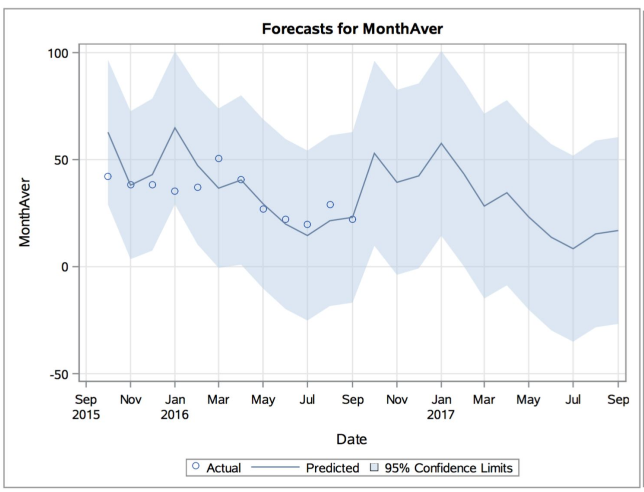
Model for variable MonthAver

No mean term in this model.

Autoregressive Factors	
Factor 1:	1 - 0.8686 B**(1)
Factor 2:	1 - 0.23018 B**(6)



The forecast plot by original data and the forecast plot with decomposed data.



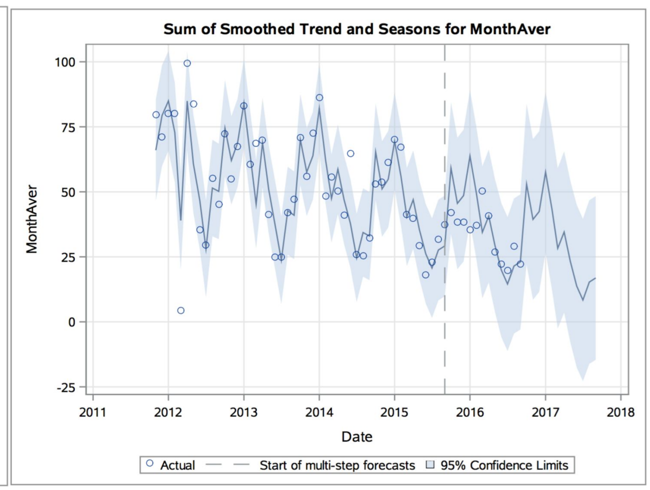
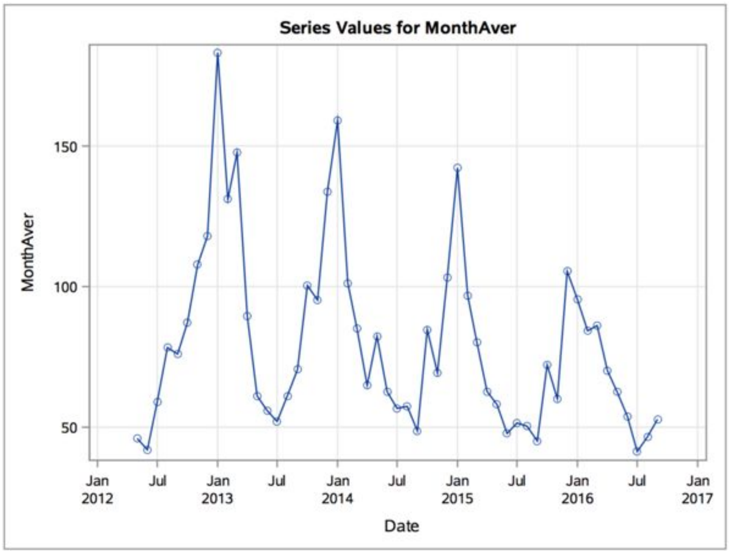


Table 6.4 ChengDu

Overall time plot and Deciding p+d and q



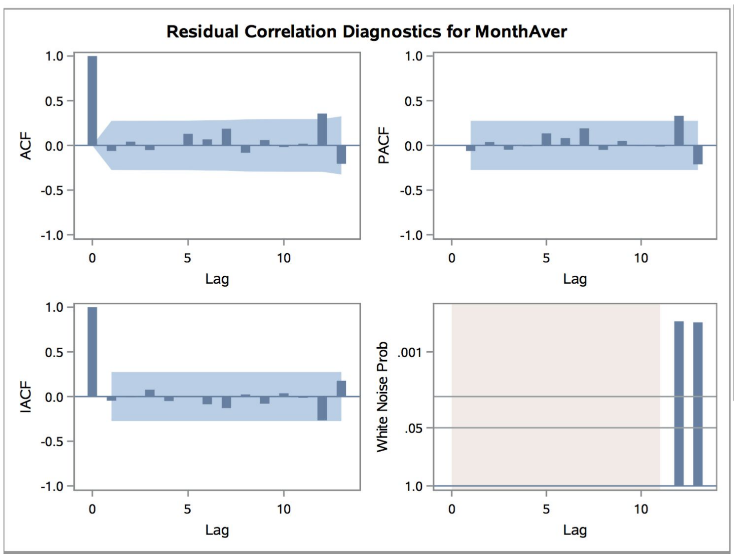
Error series model: AR(12)

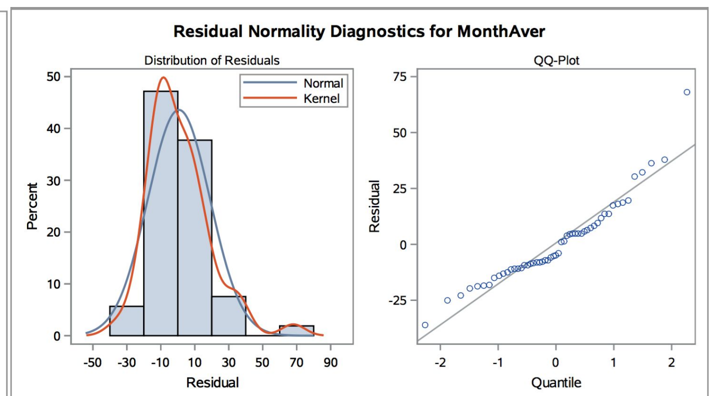
Minimum Table Value: BIC(5,6) = 5.665551

ARMA(p+d,q) Tentative Order Selection Tests		
SCAN		
p+d	q	BIC
2	1	5.871579
4	0	5.815365
0	6	6.460728

(5% Significance Level)

ACF, PACF and Residual Plot

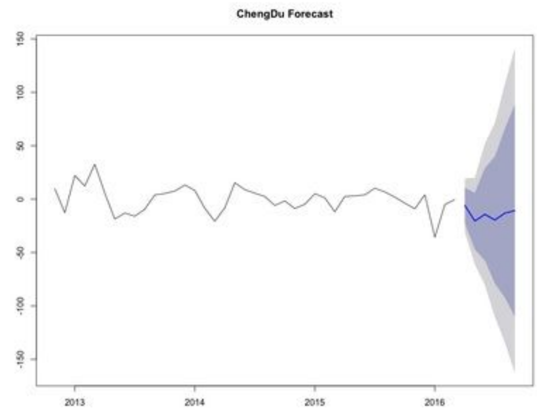




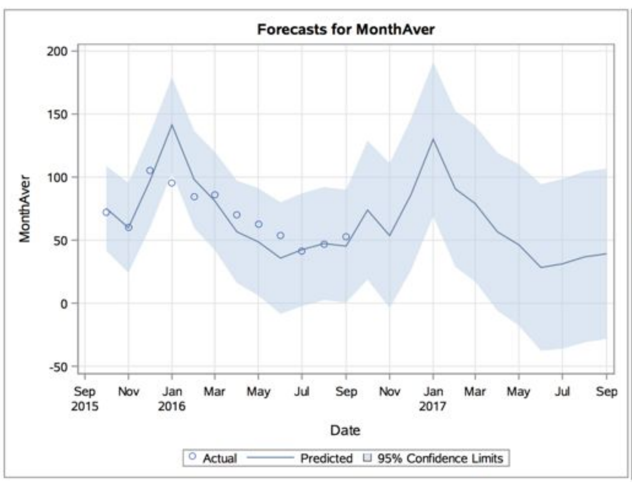
The ARIMA Model and Decomposed Data

Autoregressive Factors	
Factor 1:	1 - 0.28043 B**(1) - 0.21716 B**(2) - 0.01084 B**(3) - 0.03999 B**(4) + 0.6348 B**(5)

Moving Average Factors	
Factor 1:	1 + 0.35646 B**(1) + 0.23737 B**(2) - 0.01443 B**(3) - 0.28273 B**(4) + 0.57764 B**(5) + 0.10453 B**(6)



The forecast plot by original data and the forecast plot with decomposed data.



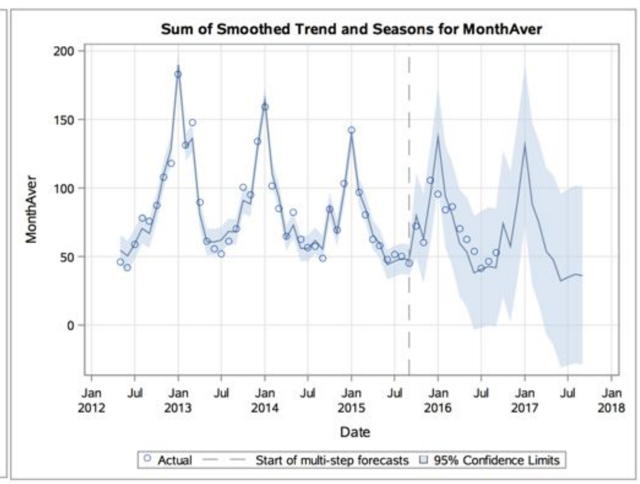
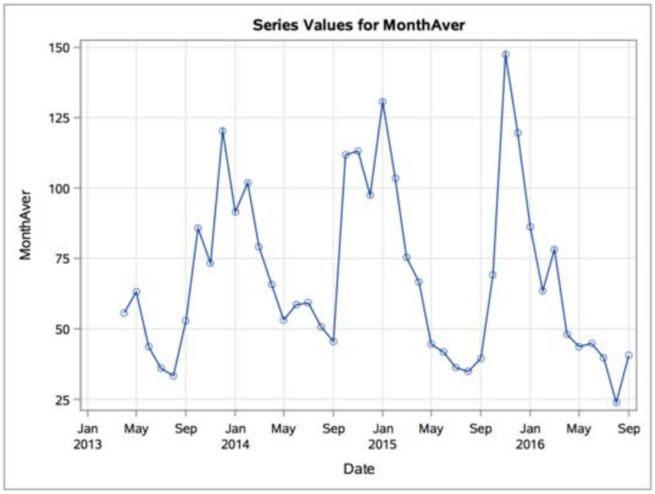


Table 6.5 ShenYang

Overall time plot and Deciding p+d and q

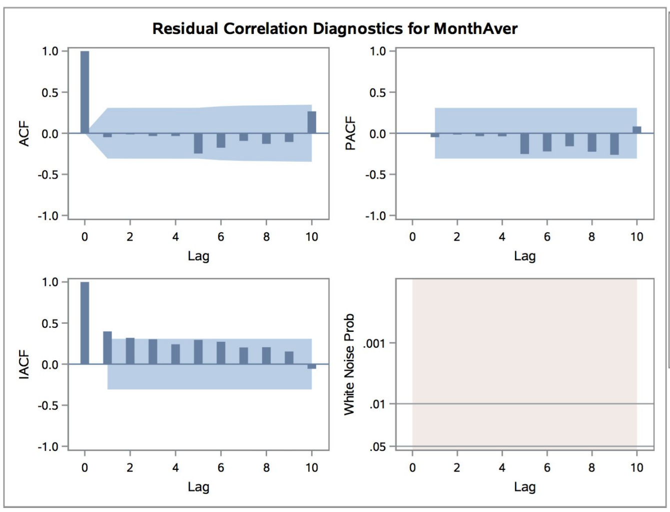


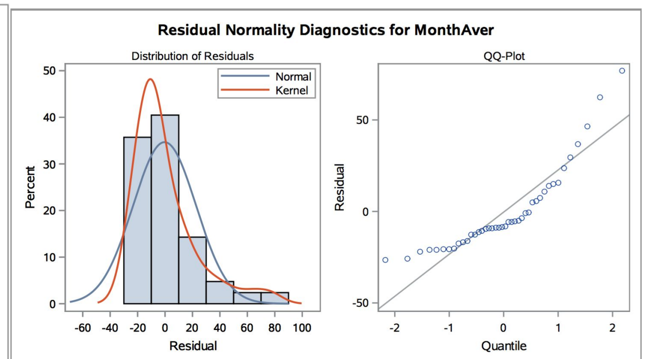
Minimum Table Value: BIC(6,4) = 5.805647

ARMA(p+d,q) Tentative Order Selection Tests		
SCAN		
p+d	q	BIC
1	1	6.156153
5	0	6.109166
0	6	6.824426

(5% Significance Level)

ACF, PACF and Residual Plot



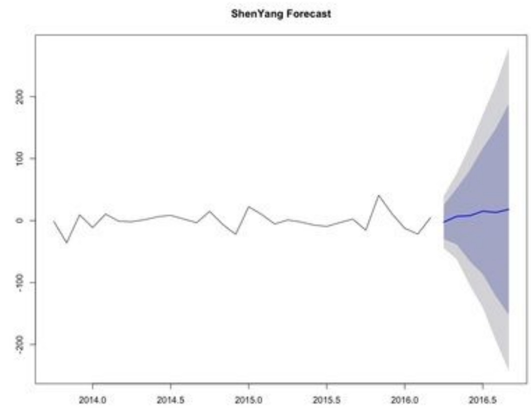


The ARIMA Model and Decomposed Data

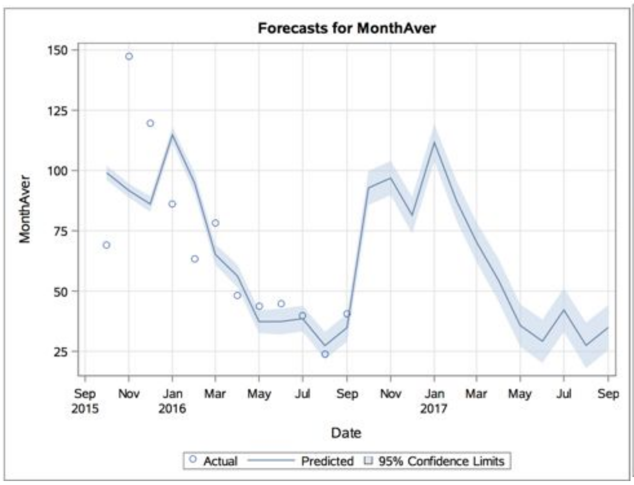
The ARIMA Procedure

Autoregressive Factors	
Factor 1:	1 - 0.89092 B**(1) + 0.64806 B**(2) - 0.24033 B**(3) - 0.53984 B**(4) + 0.2873 B**(5) + 0.03028 B**(6)

Moving Average Factors	
Factor 1:	1 - 0.04285 B**(1) + 0.60583 B**(2) + 0.28228 B**(3) - 0.34614 B**(4)



The forecast plot by original data and the forecast plot with decomposed data.



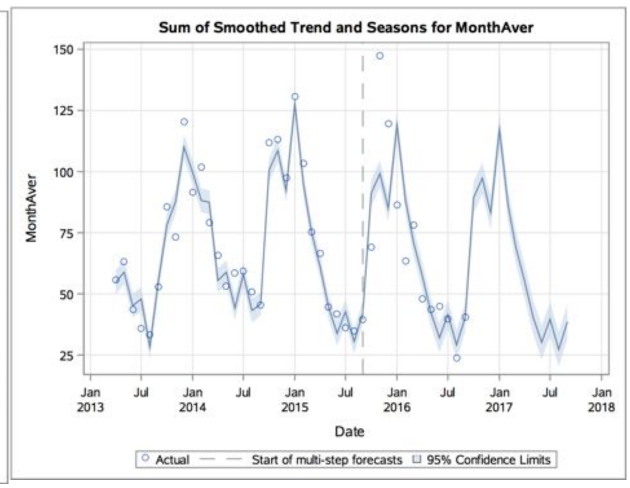
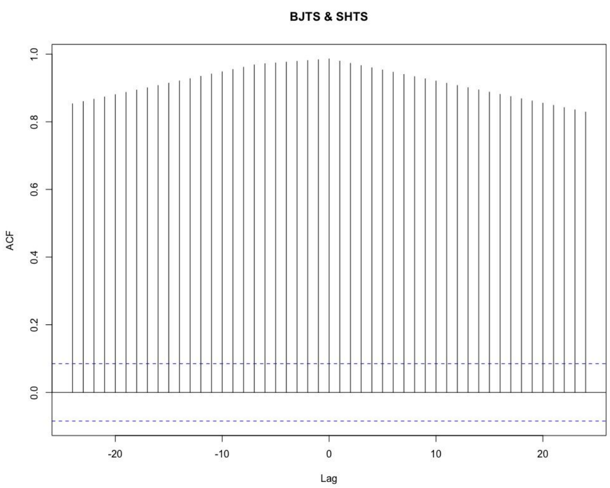


Table 7

Table 7.1 Effect of BeiJing to ShangHai from Jan, 2012 to Sep, 2016

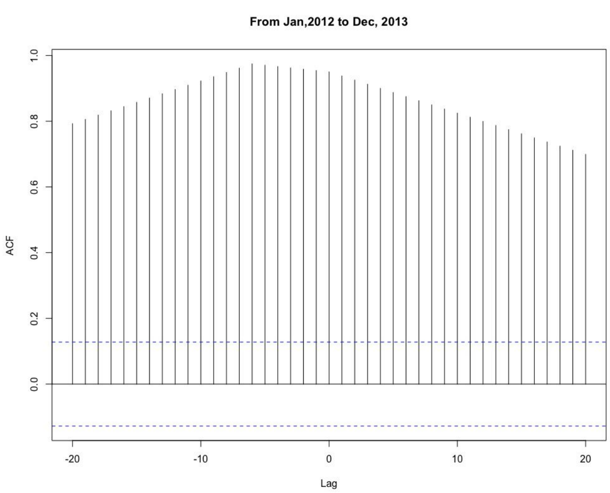



```
> BJtoSH

Autocorrelations of series ‘X’, by lag

  -24   -23   -22   -21   -20   -19   -18   -17   -16   -15   -14   -13   -12
0.853 0.860 0.867 0.874 0.880 0.887 0.894 0.901 0.907 0.914 0.921 0.928 0.934
  -11   -10    -9    -8    -7    -6    -5    -4    -3    -2    -1     0     1
0.941 0.948 0.955 0.962 0.968 0.972 0.974 0.976 0.979 0.981 0.984 0.986 0.979
    2     3     4     5     6     7     8     9    10    11    12    13    14
0.973 0.966 0.960 0.953 0.947 0.940 0.934 0.927 0.921 0.914 0.907 0.901 0.894
   15    16    17    18    19    20    21    22    23    24
0.888 0.881 0.875 0.868 0.862 0.855 0.849 0.842 0.835 0.829
```

Table 7.2 Effect of BeiJing to ShangHai from Jan, 2012 to Dec, 2013

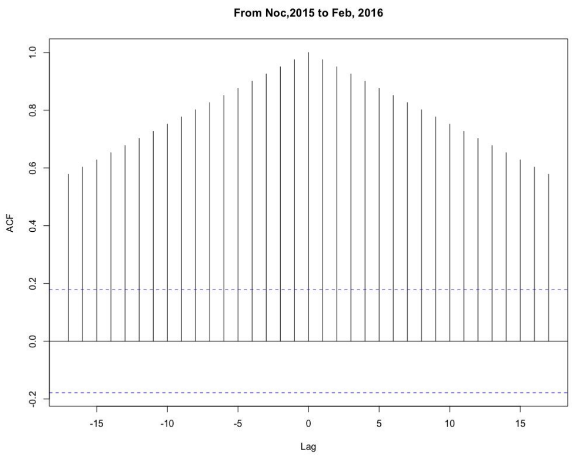



```
> BJtoSH2=ccf(BJTS2, SHTS2, type = "correlation", main="From Jan,2012 to Dec, 2013")
> BJtoSH2

Autocorrelations of series ‘X’, by lag

  -20   -19   -18   -17   -16   -15   -14   -13   -12   -11   -10    -9    -8    -7
0.793 0.806 0.819 0.832 0.845 0.858 0.871 0.884 0.897 0.910 0.923 0.935 0.948 0.961
   -6    -5    -4    -3    -2    -1     0     1     2     3     4     5     6     7
0.974 0.970 0.966 0.962 0.958 0.954 0.950 0.938 0.925 0.913 0.900 0.887 0.875 0.862
    8     9    10    11    12    13    14    15    16    17    18    19    20
0.850 0.837 0.825 0.812 0.800 0.787 0.774 0.762 0.749 0.737 0.724 0.712 0.699
```

Table 7.3 Effect of BeiJing to ShangHai from Nov, 2015 to Feb, 2016




```
> BJtoSH3=ccf(BJTS3, SHTS3, type = "correlation", main="From Noc,2015 to Feb, 2016")
> BJtoSH3

Autocorrelations of series ‘X’, by lag

  -17   -16   -15   -14   -13   -12   -11   -10    -9    -8    -7    -6    -5    -4
0.578 0.603 0.628 0.653 0.678 0.702 0.727 0.752 0.777 0.802 0.826 0.851 0.876 0.901
   -3    -2    -1     0     1     2     3     4     5     6     7     8     9    10
0.926 0.950 0.975 1.000 0.975 0.950 0.926 0.901 0.876 0.851 0.826 0.802 0.777 0.752
   11    12    13    14    15    16    17
0.727 0.702 0.678 0.653 0.628 0.603 0.578
```

SAS CODE

```
FILENAME REFFILE '/folders/myfolders/6289/Project2/BJmonthAver.xlsx';

PROC IMPORT DATAFILE=REFFILE
        DBMS=XLSX
        OUT=BJi;
        GETNAMES=YES;
RUN;

PROC print DATA=BJi; RUN;


%web_open_table(WORK.IMPORT1);

proc timeseries data=BJi plot=series
   out=series
   outtrend=trend
   outseason=season print=seasons;;
   id RQ interval=Month;
   var MonthlyAver;
run;

proc arima data=BJi;
identify var=MonthlyAver minic p=(0:6) q=(0:6) scan;
run;

proc arima data=BJi;
```

```
identify var=MonthlyAver;
estimate plot p=1 q=1;
run;

proc arima data=BJi;
identify var=MonthlyAver;
estimate p=1 q=1 method=ml;
run;

proc arima data=BJi;
  identify var=MonthlyAver;
  estimate p=(1)(6) q=(1)(6) noint method=ml;
run;

proc ucm data = BJi;
  id RQ interval = Month;
  model MonthlyAver;
  irregular p=1 q=1;
  level checkbreak;
  slope var = 0 noest;
  season length = 12 type=trig plot=smooth;
  estimate back=24 plot=panel;
  forecast lead=24 back=12  print=forecasts plot=decomp;
run;

FILENAME REFFILE '/folders/myfolders/6289/Project2/CDmonthAver.xlsx';
PROC IMPORT DATAFILE=REFFILE
        DBMS=XLSX
        OUT=CDD;
        GETNAMES=YES;
RUN;
PROC print DATA=CDD; RUN;
proc timeseries data=CD plot=series;
  id Date interval=Month;
  var MonthAver;
run;
proc arima data=CD;
identify var=MonthAver minic p=(0:6) q=(0:6) scan;
run;
proc arima data=CD;
identify var=MonthAver;
estimate plot p=5 q=6;
```

```
run;
proc arima data=CD;
identify var=MonthAver;
estimate p=5 q=6 method=ml;
run;
proc arima data=CD;
  identify var=MonthAver;
  estimate p=(1)(6) q=(1)(6) noint method=ml;
run;
proc ucm data = CDD;
  id Date interval = Month;
  model MonthAver;
  irregular p=5 q=6;
  level checkbreak;
  slope var = 0 noest;
  season length = 12 type=trig plot=smooth;
  estimate back=24 plot=panel;
  forecast lead=24 back=12  print=forecasts plot=decomp;
run;

FILENAME REFFILE '/folders/myfolders/6289/Project2/GZmonthAver.xlsx';
PROC IMPORT DATAFILE=REFFILE
        DBMS=XLSX
        OUT=GZh;
        GETNAMES=YES;
RUN;
PROC print DATA=GZh; RUN;
proc timeseries data=GZh plot=series;
  id Date interval=Month;
  var MonthAver;
run;
proc arima data=GZh;
identify var=MonthAver minic p=(0:6) q=(0:6) scan;
run;
proc arima data=GZh;
  identify var=MonthAver;
  estimate p=(1)(6) noint method=ml;
run;
proc ucm data = GZh;
  id Date interval = Month;
  model MonthAver;
  irregular p=1 q=6;
```

```
   level checkbreak;
   slope var = 0 noest;
   season length = 12 type=trig plot=smooth;
   estimate back=24 plot=panel;
   forecast lead=24 back=12  print=forecasts plot=decomp;
run;

FILENAME REFFILE '/folders/myfolders/6289/Project2/SHmonthAver.xlsx';
PROC IMPORT DATAFILE=REFFILE
	DBMS=XLSX
	OUT=SHN;
	GETNAMES=YES;
RUN;
PROC print DATA=SHN; RUN;
proc timeseries data=SHN plot=series;
   id RQ interval=Month;
   var PM25;
run;
proc arima data=SHN;
identify var=PM25 minic  p=(0:6) q=(0:6) scan;
run;
proc arima data=SHN;
identify var=PM25;
estimate plot p=3 q=1;
run;
proc arima data=SHN;
identify var=PM25;
estimate p=3 q=1 method=ml;
run;
proc arima data=SHN;
   identify var=PM25;
   estimate p=(1)(6) q=(1)(6) noint method=ml;
run;
proc ucm data = SHN;
   id RQ interval = Month;
   model PM25;
   irregular p=3 q=1;
   level checkbreak;
   slope var = 0 noest;
   season length = 12 type=trig plot=smooth;
   estimate back=24 plot=panel;
   forecast lead=24 back=12  print=forecasts plot=decomp;
```

```
run;

FILENAME REFFILE '/folders/myfolders/6289/Project2/SYmonthAver.xlsx';
PROC IMPORT DATAFILE=REFFILE
        DBMS=XLSX
        OUT=SY;
        GETNAMES=YES;
RUN;
PROC print DATA=SY; RUN;
proc timeseries data=SY plot=series;
  id Date interval=Month;
  var MonthAver;
run;
proc arima data=SY;
identify var=MonthAver minic p=(0:6) q=(0:6) scan;
run;
proc arima data=SY;
identify var=MonthAver;
estimate plot p=6 q=4;
run;
proc arima data=SY;
identify var=MonthAver;
estimate p=6 q=4 method=ml;
run;
proc arima data=SY;
  identify var=MonthAver;
  estimate p=(1)(6) q=(1)(6) noint method=ml;
run;
proc ucm data = SY;
  id Date interval = Month;
  model MonthAver;
  irregular p=6 q=4;
  level checkbreak;
  slope var = 0 noest;
  season length = 12 type=trig plot=smooth;
  estimate back=24 plot=panel;
  forecast lead=24 back=12  print=forecasts plot=decomp;
run;
```

R CODE

```
BJDT=read.xlsx("/Users/LoveChina/Documents/6289/Project2/cleaneddata/BJDaymean.xlsx",header=TRUE,sheetName = 'BJWinter')
SHDT=read.xlsx("/Users/LoveChina/Documents/6289/Project2/cleaneddata/SHdayandmonth.xlsx",header=TRUE,sheetName = 'SHWinter')

BJTS=ts(BJDT[,3])
SHTS=ts(SHDT[,3])
BJTS2=ts(BJDT[1:235,3])
SHTS2=ts(SHDT[1:235,3])
BJTS3=ts(BJDT[415:535,3])
SHTS3=ts(SHDT[422:542,3])

BJtoSH1=ccf(BJTS, SHTS, type = "correlation")
BJtoSH1
BJtoSH2=ccf(BJTS2, SHTS2, type = "correlation", main="From Jan,2012 to Dec, 2013")
BJtoSH2
BJtoSH3=ccf(BJTS3, SHTS3, type = "correlation", main="From Noc,2015 to Feb, 2016")
BJtoSH3

###plot relative plot
C5=read.xlsx("/Users/LoveChina/Documents/6289/Project 2/ccf.xlsx",header=TRUE,sheetName = 'CITY5')
C5Winter=read.xlsx("/Users/LoveChina/Documents/6289/Project 2/ccf.xlsx",header=TRUE,sheetName = 'ForWinter')

library(corrgram)
colorset=function(ncol){
  colorRampPalette(c("darkgoldenrod4","burlywood1","darkkhaki","darkgreen","red"))(ncol)
}
corrgram(C5[,4:8], order=TRUE, lower.panel=panel.shade, upper.panel=panel.pie,
        diag.panel=panel.minmax,
        main="A Correlation Among 5 Cities")

corrgram(C5Winter[,4:8], order=TRUE, lower.panel=panel.shade, upper.panel=panel.pie,
        diag.panel=panel.minmax,
        main="A Correlation Among 5 Cities in Winter")

###density plot
library(ggplot2)
p <- ggplot(C5Winter, aes(x=C5Winter$BJ, y=C5Winter$SH))
p + stat_density2d(aes(fill=..density..), geom="raster", contour=FALSE)+
  ggtitle("The Winter Correlation Between BeiJing and ShangHai")
```

```
p + geom_point() +
  stat_density2d(aes(alpha=..density..), geom="tile", contour=FALSE)+
  ggtitle("The Winter Correlation Between BeiJing and ShangHai")
```