International Journal of Engineering and Technical Research (IJETR)
ISSN: 2321-0869 (O) 2454-4698 (P), Volume-3, Issue-9, September 2015

PM2.5 concentration regression analysis based on Meteorological comprehensive index

Chen Zhao

***Abstract*— Adopting international exchange station ground comprehensive meteorological index data, the PM2.5 statistical analysis model is established by using principal component analysis theory and dimensionless reduction technology, the numerical results are obtained by employing Eviews software.**

***Index Terms*—meteorological index; principal component regression; time series modeling; PM2.5.**

I. Introduction

On January 20, 2013, in order to improve air quality, Wuhan City carried out artificial rainfall, which was the first time that used this method to alleviate the smog. According to the official data, artificial rain a few days before and after the air quality index showed a rising trend after the first, the 18th air quality index of 257 20th starting at 169, 22nd returned to 261. Among them, the PM2.5 as primary pollutants, visible under appropriate conditions, to carry out artificial precipitation can effectively alleviate the haze. Artificial rainfall washed the formation of haze of aerosol ionic solids, so as to alleviate the haze. Stakeholders also suggested that artificial rain mist and haze become normalized, as a system of fixed, funded and staffed [1]. However to carry out artificial precipitation need harsh weather conditions, and because of this, limited time to carry out artificial precipitation, blind artificial modification of rain should not be significantly reduced smog, can also cause a waste of money and resources. Studies have shown that when sources are relatively stable, weather conditions are important factors affecting PM2.5 [2], conditions conducive to the dispersion and settling of particles PM2.5 concentrations can be reduced, and not conducive to particle dispersion and settling of the meteorological conditions can lead to increased concentration of PM2.5 [3]. So the study of the relationship between PM2.5 and meteorological conditions, finding the best time to carry out artificial precipitation become a matter of urgency. In this paper we aiming at the characteristics of PM2.5, and studying the relationship between PM2.5 and meteorological conditions, according to the results we make observations and recommendations about timing problem of removing fog by artificial rain.

We will analysis international exchange station ground comprehensive meteorological index data below.

II. Data source

The data of PM2.5 concentration of Wuhan in 2014 was come from Wuhan Environmental Monitoring Center [4], the daily meteorological data of Wuhan in 2014 was come from China Meteorological data sharing service system, including 20-20 precipitation, large evaporation and maximum wind speed, the direction of maximum wind speed, the average air pressure, average wind speed, average temperature, average vapor pressure, average relative humidity, sunshine hours, small evaporation etc. [5].

III. Mathematical model

A. Principal component analysis of meteorological index

In order to make the analysis question be more comprehensive[6], we selected a large number of meteorological indices, but not every index is effective, even there is high correlation between some indices, which added the complexity of the question analysis and work intensity. Principal component analysis can transform the original multiple variables into a small number of linear composite index, which plays a role in reducing the dimension. Through the dimensionality reduction the variables which are related to each other was assembled, in order to reduce the number of variables to be analyzed, and reduce the difficulty of the analysis. Using principal component analysis, according to the intrinsic link between the data, it can be divided into several main components, the main components of different components are mutually orthogonal variables, each principal component is not related to each other, these principal components can provide a complete characterization of the whole index system.

First use PM2.5 and many indices of regression analysis. Then screened seven indicators (20-20 precipitation, large evaporation, average pressure, average wind speed, average temperature, average vapor pressure, average relative humidity), to principal component analysis. Before the principal component analysis, the data were tested by KMO and Bartlett. The test of data was mainly used to test the validity and reliability of the data, to realize whether the data can be analyzed by principal component analysis, data is shown in table 1.

From the test results, the KMO test coefficient is 0.655, the significance level is less than 0.05, so the data can be considered as the main component analysis, and numerical data is shown in table 2

Table 1 KMO test and Bartlett test

Sample sufficient Kaiser-Meyer-Olkin Metric		655
Bartlett Spherical degree test	Approximate chi square	1681
	df	19
	Sig.	.00021

Table 2 Component matrix of principal component analysis

	1	2	3
20-20 precipitation (0.1mm)	.002	.749	.345
large evaporation	-.602	.568	-.267

Chen Zhao, School of computer science, Wuhan University of Technology, 430070, Wuhan, China.

average pressure (0.1hundred Pa)	-.906	-.110	-.079
average wind speed（0.1m/s)	-.236	.0045	.928
average temperature（0.1C)	-.9743	.021	-.001
average vapor pressure	.953	.179	-.031
average relative humidity（Percentage)	.120	-.868	.104

According to the experimental results, the first three principal components explained 85.499% of the total variation of the original variables, so that the PM2.5 related meteorological index can be well reflected by the first three principal components. Using the component matrix coefficient, the simple linear combination of the original variables is constructed, and the linear expressions of the three synthetic indexes are as follows

$$X = -0.979z_5 + 0.959z_6 - 0.930z_3 - 0.629z_2 \quad (1)$$

$$Y = -0.880z_7 + 0.739z_1 \quad (2)$$

$$Z = 0.920z_4 \quad (3)$$

The average pressure, average temperature, average vapor pressure reflect the overall situation of atmospheric pressure and temperature, and 20-21 rainfall, large evaporating capacity, average relative humidity reflect the overall situation of the vapor in the air, and the last Principal component reflects the average wind speed.

.

B. Time series of principal components

After obtaining the comprehensive index of the above structure, first the data of each comprehensive index is standardized, and then the time series analysis and modeling of the comprehensive index are carried out. Test each synthetic index of the time series for stationary and pure randomness. The test results are divided into different types according to the results of the test, further take different analysis methods.

3.2.1 Stationary test

There are two kinds of methods for the stationary test, the graph test method and the structural test statistic. The method of graph test has some subjectivity, so this paper uses the method of constructing test statistics, and uses the Eviews unit root test to test the stability of the sequence. From Figure 3 we can see that the significance level of the first comprehensive index unit root test is less than 0.05, so we can determine it as a stationary sequence. In the same way, the first order difference sequence of second and third index is stable, the calculative data is shown in table 3 .

Table 3 T statistics test for first order differential

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-12.38755	0.0000
Test critical values:	1% level	-3.452066	
	5% level	-2.870996	
	10% level	-2.571880	

3.2.2 Pure randomness test [7]

Pure randomness test is mainly to test whether the sequence is white noise, if the sequence is white noise sequence, then it is pure random, has no research value, only the non pure random time sequence can be further studied with ARMA model to fit and forecast. Using Eviews to calculate the values of the three sequences of Q, see Table 4.

Table 4 Q statistic of three synthetic indexes

Sequence	x	y	z
P value of Q statistic	0.0001	0.0001	0.0001

Q statistic is used in large sample cases, the P values of the three sequences are less than 0.05, so it is considered that these sequences belong to the non white noise series, which has the analysis value and can be further used for model selection.

3.2.3 Model establishment and diagnosis

Using Eviews to get the graph of first order difference of the three comprehensive index of partial autocorrelation, according to its characteristics, select the model parameters. Multiple parameters of the model, and use the ACI information value as the evaluation criteria, and finally get the ARMA model of the three comprehensive indicators are as follows:

$$DX_t = 0.86DX_{t-1} - 0.34DX_{t-2} + d\varepsilon_t - 0.76d\varepsilon_t \quad (4)$$

$$DY_t = 0.01 + 0.22DY_{t-1} + d\varepsilon_t - 1.03d\varepsilon_{t-1} \quad (5)$$

$$Z_t = Z_{t-1} + \varepsilon_t - 0.83\varepsilon_{t-1} \quad (6)$$

Use the "static" method to estimate the second comprehensive index of first order differential, the solid line in the figure represents the predictive value of Dy, two dotted lines provides 2 times the standard deviation of the confidence interval. It can be seen from the picture, the "Static" method has a big fluctuation of the forecast value. At the same time, the variance ratio of the model has good simulation of the actual series of fluctuations, Theil unequal coefficient is 0.488, the covariance ratio is 0.488, which indicates that the model has a good prediction. The other two are similar to the situation of the general index.

C. Regression analysis of PM2.5 and meteorological index

Regression analysis is to study the causal relationship between the number of two or more variables, based on the correlation analysis, based on the correlation between the number of two or more variables to determine a suitable mathematical model, so that to use one or a number of variables to describe and predict the other or a number of variables.

The principal component analysis was used to obtain three comprehensive indices, and the time series model of the three indexes was carried out, and the regression analysis was carried out using these three indices and PM2.5 concentration index. The significance level of the F statistic of variance analysis of regression analysis is less than 0.01. So, it can be considered that the regression equation can be fitted by the 99% confidence level. and

$$k = -0.391zx - 0.217zy - 0.098zz \quad (7)$$

k represents the PM2.5 concentration index, zx,zy,zz are the first two and three comprehensive index of the standard after the standardization. According to the results of principal component analysis and regression analysis, the positive and negative correlation between the variables is summarized :

IV. CONCLUSION

Principal component and regression analysis showed that under the same conditions, the greater the rainfall, the lower the PM2.5 concentration. , which indicated that wet deposition can effectively reduce the particulate matter in the atmosphere, and when the wind speed increased, the concentration of PM2.5 increased, but when the relative humidity increased, the concentration of PM2.5 increased. Generally speaking, pollutants can spread in two main weather conditions ——precipitation and wind, continuous heavy rainfall can scour atmospheric haze particles, so once the rain is over air quality will change for the better. In addition, heavy precipitation often accompanied by strong winds, strong winds have conducive to the diffusion and dilution of pollutants. But if rainfall duration time is short, the rainfall is too small, it is very difficult to achieve the effect of purifying air, this is because in the short weak rainfall conditions, not only can ‘t scour air haze. It will also increase the relative humidity in the air, moist air adds the particulate matter suspended in the air with a layer of water film coat ", which is more likely to cause accumulation of pollutants and increase the concentration of PM2.5 [8].

The results also show that the higher the evaporation rate is, the higher the concentration of PM2.5, because of the large amount of evaporation, the particulate matter forms because of air cooling and condensation, and thus the concentration of PM2.5 is higher.

In summary, although the rainfall can rely on the wet deposition to reduce the PM2.5, but when the rainfall is small, the wind speed is relatively low and the humidity is relatively large, it can’ t reduce the concentration of PM2.5 effectively, and may even be due to the evaporation of a large number of ground water, aerosol is difficult to spread and lead to an increase in PM2.5 concentration. So the artificial rainfall relief work, should be considered implementing when the wind speed is relatively large and relatively humidity is low.

REFERENCES

[1] He Tao. The artificial rainfall to ease haze in Wuhan, contribute but not go on for long Guangzhou daily [N],2013-1-28（1）．

[2] Mei Pengyu. The influence of stable weather conditions on the quality of the environment in Tianjin [J]. urban environment and urban ecology,2006，19(4)：37—49．

[3] Xu Wei, Gan Qinghui,Tangqiang, the climate characteristics and the influencing factors of Shantou haze in 1951-2006, [J]. Meteorological and environmental newspapers.2008，24(3)：42—45．

[4] Wuhan environmental monitoring center. Environmental quality bulletin [DB/OL].http://www.whemc.cn/2015-01-01

[5] China Meteorological science data sharing service network, the daily data of climate of Chinese international exchange station [DB/OL]. http://cdc.nmic.cn/dataSetLogger.do?changeFlag=dataLogger2014-12-20

[6] I.T.Jolliffe.Principal Component Analysis[J/OL].http://wenku.baidu.com/link?url=Exy5k2cOT-EMx5-y6V5AIX6SwU3J4ItPiNFQWh7KgIVtBg9DM2wiXfB4dYh-VcImCZbUsNKMIif-jMzq_QGX3UbLXKBkN3lfB_3dngie91a,2014-12-03

[7] [Johannes Mayer.Adaptive Random Testing by Bisection with Restriction[M]. Germany:Springer Berlin Heidelberg,2005:231-256

[8] Liu Shanshan , the cloud is not be able to rain. The Shijiazhuang air quality in the first "severe" in June [N]. Yanzhao Metropolis net.2014-06-17（3）.