

The Regression Analysis between the Meteorological Synthetic Index Sequence and PM2.5 Concentration

Weijuan Liang, Zhaogan Zhang, Jing Gao, Wanyu Li, Xiaofan Liu, Liyuan Bai, Yufeng Gui

College of Science, Wuhan University of Technology, Wuhan, China

Email: 1361195693@qq.com

Received 14 September 2015; accepted 26 October 2015; published 29 October 2015

Copyright © 2015 by authors and Scientific Research Publishing Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

Adapting daily meteorological data provided by China International Exchange Station, and using principal component analysis of meteorological index for dimension reduction comprehensive, the regression analysis model between PM2.5 and comprehensive index is established, by making use of Eviews time series modeling of the comprehensive principal component, finally puts forward opinions and suggestions aim at the regression analysis results of using artificial rainfall to ease haze.

Keywords

Meteorological Index, Principal Component Analysis, Time Series Modeling, PM2.5, Haze

1. Introduction

Significant pollutant, obviously in the appropriate conditions, the implementation of artificial rainfall can effectively ease the haze. The rain of artificial rainfall scours the solid formation of haze like aerosol ion etc., and then the haze is eased. Relevant personnel are also advised to make the artificial rainfall to ease haze to become normalized, confirmed as institutions, to implement the fund and staff [1]. However, the implementation of artificial rainfall also requires harsh weather conditions, because of this, the opportunity implementation of artificial rainfall is limited, implementing the artificial rainfall to ease the haze blindly would not be effectively, but also causes the waste of money and resources. The study shows that when the pollution sources are relatively stable, the meteorological conditions are important factors to affecting the PM2.5, which can reduce the PM2.5 concentration, meanwhile when the meteorological condition is not conducive to the diffusion and the deposition of particles will lead to an increase in PM2.5 concentration [2] [3]. Therefore, it is urgent to find out the re-

relationship between PM2.5 and the meteorological conditions to get the best time to implement artificial rainfall. In this paper, we aim at the diversification characteristics of the PM2.5, study the relationship between PM2.5 and meteorological conditions, and the suggestions are put forward according to the research results.

2. Data Whence

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].

3. Data Analysis

3.1. Principal Component Analysis of Meteorological Indices

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 (see Table 1).

First use PM2.5 and many indices of regression analysis, see Table 2. 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.

From the test results, the KMO test coefficient is 0.669, the significance level is less than 0.05, so the data can be considered as the main component analysis.

Table 1. Kaiser-Meyer-Olkin test and Bartlett test.

	Sample sufficient Kaiser-Meyer-Olkin Metric	0.669
	Approximate chi square	1771.072
Bartlett Spherical degree test	df	21
	Sig.	0.000

Table 2. Component matrix of principal component analysis.

	1	2	3
20 - 20 precipitation (0.1 mm)	0.002	0.744	0.350
Large evaporation	-0.632	0.579	-0.279
Average pressure (0.1 hundred Pa)	-0.927	-0.105	-0.081
Average wind speed (0.1 m/s)	-0.238	0.005	0.917
Average temperature (0.1 °C)	-0.983	0.010	-0.001
Average vapor pressure	0.960	0.181	-0.023
Average relative humidity (Percentage)	0.115	-0.879	0.113

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

$$X = -0.983z_5 + 0.96z_6 - 0.927z_3 - 0.632z_2 \tag{1}$$

$$Y = -0.879z_7 + 0.744z_1 \tag{2}$$

$$z = 0.917z_4 \tag{3}$$

The average pressure, average temperature, average vapor pressure reflect the overall situation of atmospheric pressure and temperature, and 20 - 20 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.

3.2. 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 **Table 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.

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**.

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

Table 3. First comprehensive index T statistics test for first order differential.

Augmented Dickey-Fuller test statistic		t-statistic	Prob.
		-12.388	0.0000
Test critical values:	1% level	-3.452	
	5% level	-2.871	
	10% level	-2.572	

Table 4. P value of time series Q statistic of three synthetic indexes.

Sequence	x	y	z
P value of Q statistic	0.000	0.000	0.000

ACI information value as the evaluation criteria, and finally get the ARMA model of the three comprehensive indicators are as follows:

$$DX_t = 0.86DX_{t-1} - 0.34DX_{t-2} + d\varepsilon_t - 0.76d\varepsilon_{t-1} \tag{4}$$

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

$$Z_t = Z_{t-1} + \varepsilon_t - 0.83\varepsilon_{t-1} \tag{6}$$

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.

3.3. 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

$$k = -0.391zx - 0.217zy - 0.098zz \tag{7}$$

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 as **Figure 1** and **Table 5**:

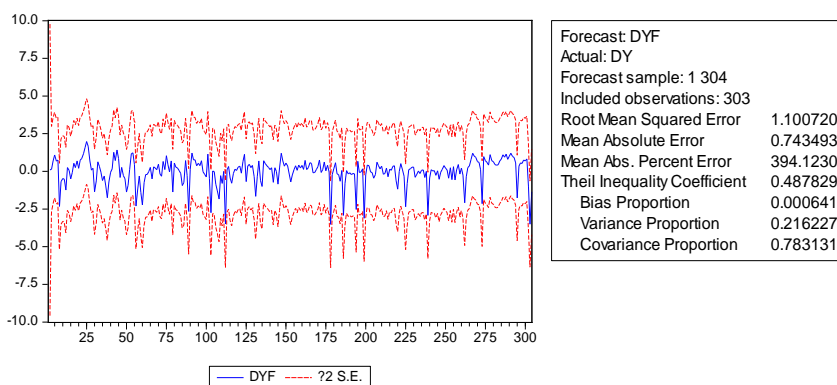


Figure 1. Second comprehensive index of first order differential sequence prediction.

Table 5. The positive and negative linear correlation between the index variables.

Index	Rainfall	Evaporating capacity	Air pressure	Wind speed	Temperature	Vapor pressure	Relate humidity	PM2.5
X			-		-	+		-
Y	-	-					-	-
Z				+				-
PM2.5	-	-	+	-	+	-	+	

Note: The symbol indicates the linear correlation between the two variables, + means positive correlation, - means negative correlation.

4. Results Analysis and Suggestions

Principal component and regression analysis showed that under the same conditions, the greater the rainfall, the lower the PM_{2.5} concentration, which indicated that wet deposition can effectively reduce the particulate matter in the atmosphere, and when the wind speed increased, the concentration of PM_{2.5} increased, but when the relative humidity increased, the concentration of PM_{2.5} 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 is often accompanied by strong winds which are 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, the air haze can't be scour, and the relative humidity in the air will also increase, 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 PM_{2.5} [8].

The results also show that the higher the evaporation rate is, the higher the concentration of PM_{2.5}, because of the large amount of evaporation, the particulate matter forms because of air cooling and condensation, and thus the concentration of PM_{2.5} is higher.

In summary, although the rainfall can rely on the wet deposition to reduce the PM_{2.5}, 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 PM_{2.5} effectively, and may even be due to the evaporation of a large number of ground water, aerosol is difficult to spread and leads to an increase in PM_{2.5} concentration. So the artificial rainfall relief work should be considered implementing when the wind speed is relatively large and relatively humidity is low.

Acknowledgements

The paper is financially supported by students' innovation and entrepreneurship training program, Wuhan University of technology, China (No. 146814003).

References

- [1] He, T. (2013) The Artificial Rainfall to Ease Haze in Wuhan, Contribute but Not Go on for Long Guangzhou Daily.
- [2] Mei, P.Y. (2006) The Influence of Stable Weather Conditions on the Quality of the Environment in Tianjin. *Urban Environment and Urban Ecology*, **19**, 37-49.
- [3] Xu, W., Gan, Q.H. and Tang, Q. (2008) The Climate Characteristics and the Influencing Factors of Shantou Haze in 1951-2006. *Meteorological and Environmental Newspapers*, **24**, 42-45.
- [4] Wuhan Environmental Monitoring Center (2015) Environmental Quality Bulletin. <http://www.whemc.cn/>
- [5] (2014) China Meteorological Science Data Sharing Service Network, the Daily Data of Climate of Chinese International Exchange Station. <http://cdc.nmic.cn/dataSetLogger.do?changeFlag=dataLogger>
- [6] Jolliffe, I.T. (2014) Principal Component Analysis. http://wenku.baidu.com/link?url=Exy5k2cOT-EMx5-y6V5AIX6SwU3J4ItPiNFQWh7KgIVtBg9DM2wiXfB4dYh-VcImCZbUsNKMIif-jmZq_OGX3UbLXKBkN3lfB_3dngie91a
- [7] Mayer, J. (2005) Adaptive Random Testing by Bisection with Restriction. Springer, Berlin, 231-256. http://dx.doi.org/10.1007/11576280_18
- [8] Liu, S.S. (2014) The Cloud Is Not Be Able to Rain. The Shijiazhuang Air Quality in the First "Severe" in June. *Yanzhao Metropolis Net*.