Gray Prediction of Air Pollutants in Beijing Based on Improved Residual GM (1, 1) Model

Yong Shao[1, a], Yang Zhang [2, b] and Changshun Yan [2, c]

[1]School of software, Beijing University of Technology, Beijing 100124, China;

[2]School of software, Beijing University of Technology, Beijing 100124, China.

[a]shaoyong@bjut.edu.cn, [b]15600520022@163.com, [c]yuewuxing@bjut.edu.cn

Keywords: air quality, Gray prediction, improved residual model, residual correction.

Abstract. Based on the gray system theory, the original gray predicting model, the residual correction model and the residual model have been established for PM_{10}, SO_2 and NO_2, which are the major atmospheric pollution factors in Beijing. After comparing those three models, this paper established a new improved residual model, and the results show that: the improved residual model has reached a better accuracy in forecasting. This model can be used to predict the concentration of common pollutants in air in Beijing for the next few years as well as to forecast the development of air pollutants.

1. Introduction

Air quality has been a hot issue concerned by people, especially in cities like Beijing, where traffic is crowded, economy is growing fast, urbanization and industrial development are rapid and energy consumption is high. With the rapid development, air pollution became increasingly serious. Although the gray GM (1, 1) model [1,2] and the residual correction model [3, 4] have been used by many scholars to predict the air quality, the precision of these two models is not easy to meet the requirements in most cases. Therefore, for the air quality prediction in Beijing from 2016 to 2020, there are some problems such as large spatial scale, long time span, discretization of original data and other issues. Based on the research of GM (1, 1) model, this paper compared three forecasting models and chooses the improved the residual model which has the best prediction precision to simulate prediction of the air pollutant dominating factor in Beijing from 2016 to 2020.

2. GM (1, 1) modeling theory

The gray system theory mainly applies on those uncertain systems with "small sample" and "poor information", information in those systems are partial known and partial unknown. The gray system theory analyzes those "partial known" information in a specific system, and digs out important information through those data, in this way, it will reach a correct description and understand of the real world [5]. Gray prediction GM (1, 1) model is the core content of gray system theory, which can be used to forecast some unknown data through some known data, conceptual quantization by quantitative information abstraction, quantitative conceptual modeling, and finally model optimization [6].

2.1 The Establishment of Gray GM (1, 1) Model

When using GM (1, 1) modeling, the original data series must meet or basically accorded with the exponential change rule, and the change speed cannot be too fast, in addition, whether to meet the quasi-smoothness condition is also an important threshold to establish the gray system model of a sequence. Therefore, generally, before modeling, data need to be test for its quasi-smoothness and quasi-exponential law.

First step: Quasi-smoothness and quasi-exponential rule test [5].

Second step: GM (1, 1) modeling.

Set the time series $x^{(0)} = (x^{(0)}(1), x^{(0)}(2), \cdots, x^{(0)}(n))$ for nonnegative original sequences, the original sequence subtracts to generate a new sequence $x^{(1)} = (x^{(1)}(1), x^{(1)}(2), \cdots, x^{(1)}(n))$, among them $x^{(1)}(k) = \sum_{i=1}^{k} x^{(0)}(i), k = 1, 2, \cdots, n$. The first-order univariate differential equation was used to get the whitened differential equation:

$$\frac{dx^{(1)}}{dt} + ax^{(1)} = b, \text{ a is the development coefficient and b is the gray effect.} \tag{1}$$

The parameter vector $\hat{a} = [a, b]^T$ can be determined by using the least squares estimation formula:

$$\hat{a} = (B^T B)^{-1} B^T Y \tag{2}$$

Where Y, B are:

$$Y = \begin{bmatrix} x^{(0)}(2) \\ x^{(0)}(3) \\ \vdots \\ x^{(0)}(n) \end{bmatrix} \quad B = \begin{bmatrix} -x^{(1)}(2) & 1 \\ -x^{(1)}(3) & 1 \\ \vdots & \vdots \\ -x^{(1)}(n) & 1 \end{bmatrix} \tag{3}$$

The general solution of the above differential equation is: $x^{(1)}(k) = Ce^{-ak} + \frac{b}{a}$, C is a constant, the solution is discretized, and according to $k = 0$, $X_{k+1}^{(1)} = X_{(1)}^{(1)}$, so the differential equation of the discrete numerical solution:

$$\hat{x}^{(1)}(k+1) = \left(x^{(0)}(1) - \frac{b}{a} \right) e^{-ak} + \frac{b}{a} \quad , k = 1, 2, \cdots, n-1 \tag{4}$$

Furthermore, the gray prediction model of the original series $X^{(0)}$ can be obtained by accumulating and reducing:

$$\hat{x}^{(0)}(k+1) = \hat{x}^{(1)}(k+1) - \hat{x}^{(1)}(k) = (1 - e^a)[x^{(0)}(1) - \frac{b}{a}]e^{-ak}, \quad k = 1, 2, \cdots, n \tag{5}$$

2.2 Establishment of GM (1, 1) Residual Error Correction Model

When the randomness of data fluctuates greatly or the regularity is not strong, the prediction precision of GM (1, 1) model may not meet the requirement. At this point, residual correction can be applied to the model to improve the predicting accuracy. The model can be modified according to the corresponding principle, until the accuracy is reached [4].Its principles and steps are as follows:

(1) Set up the residual data sequence:

$$\varepsilon_k^{(0)} = x_k^{(0)} - \hat{x}_k^{(0)} \quad , k = 1, 2, 3 \ldots, n. \tag{6}$$

(2)According to the actual situation, we can take some sub series of $\varepsilon_k^{(0)}$ to build the model. If the sequence has both positive and negative numbers, non-negative processing on those numbers is required. The method of non-negative processing of residual data sequence is: each data added the absolute value of 2 times of the minimum negative. After the processing, the non-negative sequence is deemed as initial sequence, and the GM (1, 1) model was applied again to that sequence.

$$\mu^{(0)} = \varepsilon_k^{(0)} + 2 | \varepsilon_{min}^{(0)} | \tag{7}$$

(3) Residual sequence prediction: Firstly, the first accumulation sequence $\mu^{(1)}$ of the residual sequence $\mu^{(0)}$ is obtained, and then $\mu^{(1)}$ is accumulated and subtracted to obtain its prediction value $\hat{\mu}^{(0)}$.

(4) Restore residual sequence prediction: with the application of equation (8), the residual sequence predicted value after non-negative treatment will be restored to the original sequence's prediction sequence $\hat{\varepsilon}_k^{(0)}$.

$$\hat{\varepsilon}_k^{(0)} = \hat{\mu}_k^{(0)} - 2\,|\,\varepsilon_{\min}^{(0)}\,| \tag{8}$$

(5) A residual correction model is added to the original model to obtain the predicted value $\hat{x}_{1k}^{(0)}$:

$$\hat{x}_{1k}^{(0)} = \hat{x}_k^{(0)} + \hat{\varepsilon}_k^{(0)} \tag{9}$$

2.3 Establishment of Improved Gray Residual Forecasting Model

The key to the establishment of the improved gray prediction model is to establish the GM (1, 1) model with the absolute value of all the values in the residual sequence as the original sequence. Markov process is used to decide the sign of the residual prediction value at k> n, to correct the gray prediction value [7].

1)Firstly, the GM (1,1) model is established for the residual absolute value sequence $\left|\varepsilon_k^{(0)}\right|$ (k = 1,2,3 ..., n),and the time response function is obtained:

$$\left|\hat{e}^{(0)}(k+1)\right| = (1-e^{a1})[\left|e^{(0)}(1)\right| - \frac{b1}{a1}]e^{-a1k}, k=1,2,\cdots\cdots n-1 \tag{10}$$

2) The time response function of the improved residual gray prediction model is obtained by the superposition of the formula (10) to the absolute value and the formula (5):

$$\chi^{(0)}(k+1) = (1-e^{a})[\chi^{(0)}(1) - \frac{b}{a}]e^{-ak} + \text{sgn}\,(k+1)\,(1-e^{a1}) \tag{11}$$

In sgn(k+1)= $[\left|e^{(0)}(1)\right| - \frac{b1}{a1}]e^{-a1k}, k=1,2,\cdots\cdots n-1$,if $\varepsilon_k^{(0)} \geq$ 0,then sgn(k)=1,if $\varepsilon_k^{(0)} \leq$ 0,then sgn(k)=-1.When $1 \leq k \leq n$, the value of sgn (k) is determined by the sign of the original residual. When k> n, the value of sgn (k) needs to be predicted by Markov process.

3. Model accuracy test

The accuracy of the model is usually tested by residual test, posterior test, correlation test, small error probability test[5].The average relative error α is smaller, the result is better, the greater the correlation degree ε is, the result is better, the average variance ratio C and the probability of error p are smaller, the result is better .For the GM (1,1) prediction model, we need to test these four aspects.

Refer to the table of accuracy check level, if p, C are within the allowable range to calculate the predicted value, otherwise it is necessary to analyze the residual series of corrections before forecasting. Table 1 shows the values of various indicators:

Table 1 Posterior Differential Inspection Level Reference

Critical value / Accuracy class	α	C	ε	p
1 (good)	<0.01	<0.35	>0.90	>0.95
2(qualified)	<0.05	<0.50	>0.80	>0.80
3(barely)	<0.10	<0.65	>0.70	>0.70
4(unqualified)	$\geq$0.10	$\geq$0.65	$\leq$0.70	$\leq$0.70

4. Forecast of Air Pollutants in Beijing

There are three main pollutants in the air in China: PM_{10}, NO_2, and SO_2. The followed shown annual average emission concentration of three pollutants in Beijing from 2009 to 2015 are from Beijing Yearbook and the Beijing State of the Environment Bulletin. As shown in Table 2:

Table 2 Data of Atmospheric Pollutants in Beijing   mg/m³

Year/years	SO_2	NO_2	PM_{10}
2009	0.034	0.053	0.121
2010	0.032	0.057	0.121
2011	0.028	0.055	0.114
2012	0.028	0.052	0.109
2013	0.0265	0.056	0.1081
2014	0.0218	0.0567	0.1158
2015	0.0135	0.05	0.1015

4.1 Establishing GM (1, 1) prediction model for main impact factor of air pollutants

In order to apply the theory of gray modeling, the original sequence has been verified to perfectly satisfy the smoothness and quasi index law. In accordance of   the modeling process of gray prediction model, taking the annual average concentration of PM_{10}, SO_2 and NO_2 in Beijing from 2009 to 2014(when the improved residual model was built, the data collected in 2015 was the last set of data. Due to the uncertainty of the situation in 2015, the residual symbol was ignored) as original sample, the original model of gray prediction (As shown in Table 3), Improved Residual Model were built. A residual correction model was established for those original models whose precision was not satisfied.

Table 3 Gray Prediction Models of PM_{10}, SO_2, and NO_2

Pollutants	Gray model
PM_{10}	$\hat{x}^{(1)}(k+1) = -7.9601e^{-0.0148k} + 8.0811$
SO_2	$\hat{x}^{(1)}(k+1) = 0.4156e^{-0.0794k} + 0.4496$
NO_2	$\hat{x}^{(1)}(k+1) = 74.6476e^{0.00074k} - 74.5946$

4.2 Comparison of raw data fitting and accuracy test

The GM (1, 1) gray prediction model, the residual correction model and the improved residual model were used to fit the original data. The results of the three gray prediction models are shown in Table4, 5, 6. (note: the monitored value is denoted by M, the predicted value is denoted by P, the residual is denoted by R, the residual symbol is denoted as RS, the relative error is denoted by RR, and the absolute value of prediction residual is denoted as APR).

Table 4 The Result of PM_{10}Fitting in Beijing from 2009 to 2014(except for residual error and error)

	PM_{10}											
	The original model					Residual Correction Model				Improved Residual Model		
Years	M	P	R	RS	RR %	P	R	RR %	APR	P	R	RR%
2009	0.121	0.121	0		0	0.121	0	0	0	0.121	0	0
2010	0.121	0.117	0.0040	+	3.34	0.117	0.0040	3.34	0.0026	0.1196	0.0014	1.157
2011	0.114	0.1152	-0.0012	-	1.09	0.114	0	0	0.0031	0.1121	-0.0019	1.667
2012	0.109	0.1136	-0.0046	-	4.18	0.1094	-0.0004	0.367	0.0037	0.1099	-0.0009	0.826
2013	0.1081	0.1119	-0.0038	-	3.5	0.1072	0.0009	0.833	0.0044	0.1075	0.0006	0.555
2014	0.1158	0.1102	0.0056	+	4.8	0.1153	0.0005	0.432	0.0052	0.1154	0.0004	0.345
2015	0.1015	0.1086	-0.0071	-	6.99	0.1031	-0.0016	1.576	0.0061	0.1025	-0.0010	0.985

Table 5 The Result of SO$_2$Fitting in Beijing from 2009 to 2014(except for residual error and error)

SO$_2$									
	The original model					Improved Residual Model			
Years	M	P	R	RS	RR %	APR	P	R	RR%
2009	0.034	0.034	0		0	0	0.034	0	0
2010	0.032	0.0317	0.0003	+	0.81	0.0007	0.0324	-0.0004	1.25
2011	0.028	0.0293	-0.0013	-	4.71	0.0009	0.0284	-0.0004	1.4286
2012	0.028	0.0271	0.0009	+	3.28	0.001	0.0281	-0.0001	0.3571
2013	0.0265	0.0250	0.0015	+	5.6	0.0012	0.0262	0.0003	1.132
2014	0.0218	0.0231	-0.0013	-	5.99	0.0015	0.0216	0.0002	0.9174
2015	0.0135	0.0213	-0.0078	-	57.8	0.018	0.0231	-0.0096	71.1

Table 6 The Result of NO$_2$ Fitting in Beijing from 2009 to 2014(except for residual error and error)

NO$_2$												
	The original model					Residual Correction Model			Improved Residual Model			
Years	M	P	R	RS	RR %	P	R	RR%	APR	P	R	RR%
2009	0.053	0．053	0		0	0.053	0	0	0	0.053	0	0
2010	0.057	0.0553	0.0017	+	0.298	0.0553	0.0017	0.298	0.0015	0.0568	0.0002	0.3508
2011	0.055	0.0553	-0.0003	-	0.54	0.055	0	0	0.0015	0.0538	0.0012	2.182
2012	0.052	0.0553	-0.0033	-	6.42	0.0529	-0.0009	1.73	0.0014	0.0539	-0.0019	3.654
2013	0.056	0.0554	0.0006	+	1.11	0.0561	-0.0001	0.1786	0.0014	0.0568	0.0008	1.482
2014	0.0567	0.0554	0.0013	+	2.25	0.0571	-0.0004	0.705	0.0014	0.0568	-0.0001	0.1764
2015	0.05	0.0555	-0.0055	-	11	0.0578	-0.0078	15.6	0.0013	0.0568	-0.0068	13.6

The test results of GM (1, 1) gray prediction original model, residual correction model and improved residual model are shown in Table 4. In comparison with Table 1, the original model of PM$_{10}$ and NO$_2$ is unqualified. After the residual correction, the accuracy grade only improved a bit, while the improved residual model accuracy of PM$_{10}$ and SO$_2$ can achieve excellent accuracy. The results of this study show that:

(1) For those original sequences who have better smoothness and more obvious exponential characteristics and only obey one single changing rule, the original GM (1, 1) model can achieve good prediction result, and it is not necessary to carry out residual correction; however, if the improved residual model is applied, the prediction accuracy will be higher.

(2) For those original sequences who have obvious fluctuation, the predicting result of GM (1, 1) model is not satisfying. In this case, a more accurate prediction model could be obtained by establishing the residual correction model. However, it should be noted that it is necessary to select a reasonable residual tail modeling according to the actual situation, because the chosen model would have a certain impact on the model accuracy. For example, the original model of NO$_2$ for the residual error correction is selected from the third residual start of the residual sequence, making the revised model greatly improved accuracy, and even higher than the accuracy of the improved residual model. While the establishment of improved residual model does not need to pay attention to this problem, because the improved residual model combines the gray prediction model with the Markov process, it can overcome the influence of random volatility data on prediction accuracy, and can improve the prediction effect, the improved residual model can be used directly to obtain better prediction results.

In summary, by the use of improved residual models for air pollutants to predict the main factors, the accuracy level can be achieved. Comparing with original prediction model and residual correction model, improved residual models can better achieve predicting accuracy. In one word, improved residual model has a certain practicability in air pollutant prediction, and can contribute to the scientific predicting.

5. Forecast results

According to the established gray prediction model, draw the fitting and forecasting trend curve of Fig. 1 to Fig. 3, can be seen from the figure, improve the residual model relative to the other two models, the predicted value and the original data fitting better. Therefore, this paper selects the improved residual model to predict the annual mean concentration of air pollutants in Beijing from 2016 to 2020, and the prediction results are shown in Table 7. From the prediction results, by 2020, Beijing SO_2 annual concentration can reach the national standard; PM_{10}, NO_2 still cannot meet the national secondary standard.

Table 7 Beijing major pollution factor 2016-2020 predictive value

years	PM_{10} (mg/m³)	SO_2(mg/m³)	NO_2(mg/m³)
2016	0.0997	0.0176	0.0568
2017	0.0969	0.0207	0.0567
2018	0.0918	0.0198	0.0568
2019	0.0903	0.0191	0.0568
2020	0.0866	0.0188	0.0568
National level standards	0.04	0.02	0.04
National secondary standard	0.07	0.06	0.04

Note: The national air quality standards from GB3095-2012 "ambient air quality standards," the standard since January 1, 2016 formally implemented.
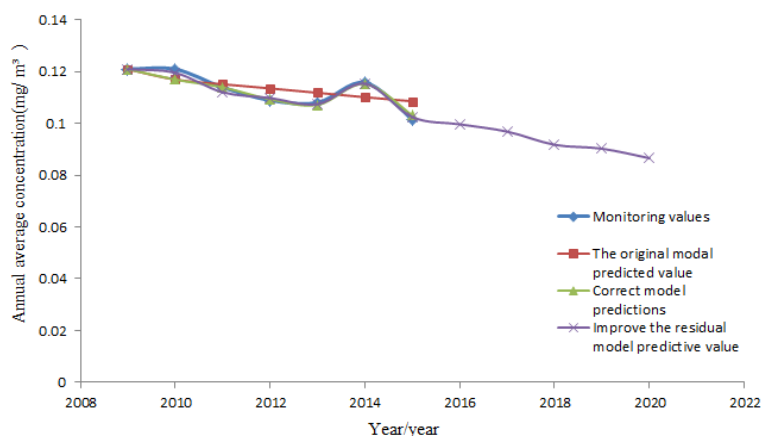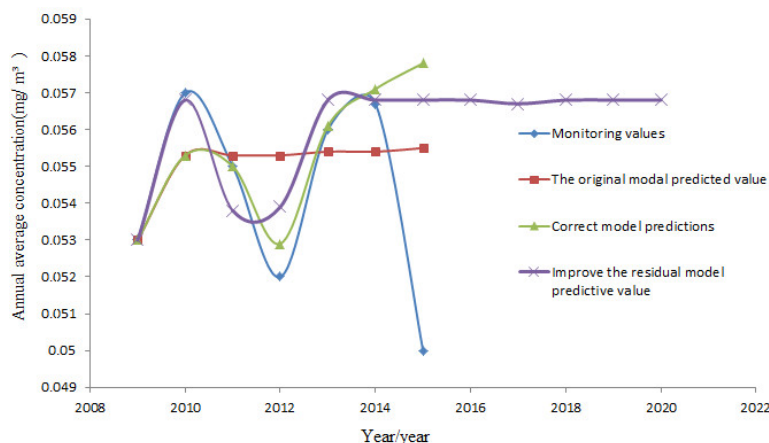

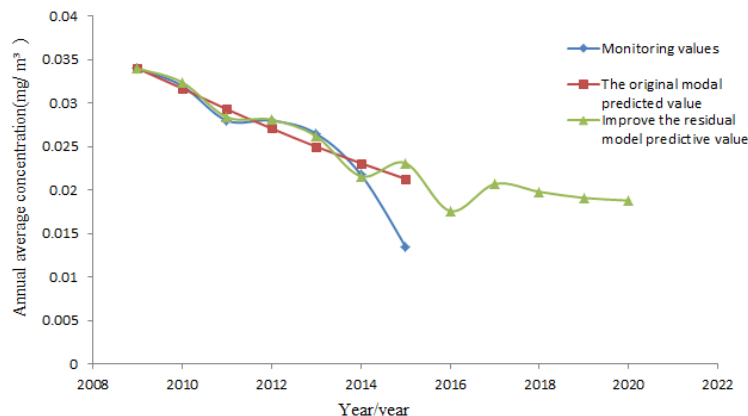Fig.1 PM_{10} Curve fitting

Fig.2 SO_2 Curve fitting

Fig.3 NO$_2$ Curve fitting

6. Conclusion and discussion

The predicted value obtained by using the improved residual model is close to the actual value, and the error is small, and good results can be obtained. By 2020, the air pollutant SO$_2$'s concentration in Beijing can be reduced to reach the standard putting forward clearly by the "13th Five-Year Plan", the plan was that keep developing a green society, strive to improve the ecological environment, decrease the emission of air pollutant, alleviate Beijing "big city disease" and other prominent issues, effectively improve the quality of ecological environment. The prediction results of NO$_2$ and PM$_{10}$ are not very satisfactory, mainly because the control of NO$_2$ and PM$_{10}$ in Beijing has yet to be strengthened, the continuous increase of the number of motor vehicles and a variety of industrial processes continue to be affected by the impact of Tianjin and Hebei Province, making NO$_2$ and PM$_{10}$ concentrations of the trend of instability. Therefore, in order to achieve the goal putting forward on March 2016 national two sessions which is to improve the ecological environment and build a clear environment for people, Beijing needs to enhance the effort to take air pollution into control and to take effective measures to fundamentally improve and govern air pollution.

Acknowledgments

References

[1] Xuexin Liu, An Xue, Gray prediction of air quality in Beijing in 2008 [J] ,Environmental Engineering 2006, 24 (2): 69-71.

[2] Wei Dai, Dong Qu, Prediction of Air Pollutant SO$_2$ in Qinhuangdao City Based on GM (1,1) Model [J], Northwest Agriculture and Forestry University 2011,21 (1): 55-57.

[3] Ling Zhao, Minkang Tang, Prediction of Fire Accident Based on Residual Modified GM (1,1) Model, Jiangxi University of Science and Technology 2012,42 (18): 82-85.

[4] Bin Li, Weimin Zhang, Application of GM_1_1 Residual Correction Model in Prediction of Hepatitis B Incidence,   Chinese Journal of Health Statistics,2014,31 (2).

[5] Sifeng Liu, Naiming Xie, Gray system theory and its application, sixth ed., Science Press, pp.98-100,154-155.

[6] Hualong Yang, Jinxia Liu, Wu Zheng, Improvement and Application of Gray Forecast GM (1,1) Model ,Mathematics in Practice and Theory , Dalian Maritime University 2011,41 (23): 39.

[7] Peifeng Cheng, Wan Zheng, Study on Forecasting Pavement Performance Based on Improved Gray Model [J],Journal of China & Foreign Highway, 2014 (6).