

Application of GM (1, 1) model in PM2.5 content prediction

Hongfu Ai^{1, a}, Ying Shi^{1, b}

¹Jilin Agricultural University, Changchun, 130117, China

^aemail: aixin1115@163.com, ^bemail:7833949@qq.com

Keywords: PM2.5; GM (1, 1) Model; Prediction

Abstract. In recent years, the frequent occurrence of fog and haze weather phenomenon, and the governance of the haze are more difficult. In order to reduce the harm caused by the PM2.5 content of the main components of its prediction research has gradually become a hot spot. In this paper, taking Changchun city as an example, using the grey theory GM (1, 1) to establish the prediction model, the PM2.5 content in the next 2 days in Changchun city was predicted. Experimental results show that the forecast effect is good, can be used as an effective method of fog and haze weather forecast.

Introduction

Haze weather has become a problem more difficult to manage; the effect of a variety of control methods for the haze is not very satisfactory. So in this case, the prediction of the haze is particularly important. Haze on people's lives brought great harm, if they can more accurately predict the occurrence of haze becomes more significant meaning [1]. The main source of pollution of the haze is PM2.5, in general the value of the index as a haze of judgment standards, so for the prediction of PM2.5 content is particularly important. A large amount of data has been confirmed, PM2.5 content value is a non-linear data, and volatile, dynamic and random.

Among the prediction method, GM gray system theory (1, 1) model for a single column of data and information are scarce data, forecast the effect is more ideal. So in this paper with the GM (1,1) of gray theory as a model to predict PM2.5 in Changchun City , experimental results show that the prediction method is better, it can be applied to Changchun fog haze scientific and rational prediction, has research value.

GM (1, 1) model

Grey system theory proposed by Deng Ju-long [2], after years of development, gray theory is applied to various fields, and to prove its effectiveness. Among the gray models GM (1, 1) model in predicting is the most widely used. The model is found by the content data itself (time series data) of internal rules, dynamic forecast. For different subjects, this model also has a corresponding improvement. Common improved models include: (1) dynamic GM (1, 1) model; (2) the metabolic GM (1, 1) model; (3) Residual GM (1, 1) model; (4) GM (1 1) cluster model.

GM (1, 1) model is widely used because a small amount of sample data, relatively scarce information can be the object of study [3]. Based on a small sample of the information obtained it find the law to extract useful information to achieve effective control and prediction of the research objectives of the system [4].

In this study, the dynamic grey model GM (1, 1) was used to predict the content of PM2.5. The main operation rules are listed below:

Set through the normalization process and weaken the data series for the operator after the operation:

$$x^0(t) = x^0(1), x^0(2), \dots, x^0(n) \quad (1)$$

The data accumulated in the above formula to do, the next time the data is generated sequence

$$x^1(t) = x^1(1), x^1(2), \dots, x^1(n) \quad (2)$$

While x^1 do mean that the data was obtained in a 1-AVG

The first order differential equation of PM2.5 content prediction model is constructed:

$$\frac{dx^1(t)}{dt} + ax^1(t) = u \quad (3)$$

The parameters of a and u in the formula are determined by the least square method in the present study. Through the above treatment, the final prediction model is shown below:

$$\hat{x}^1(t+1) = \left(x^0(1) - \frac{b}{a}\right) e^{-at} + \frac{u}{a} \quad (4)$$

Where b is the least square method, the coefficient t matrix is calculated.

In this way, after the grey dynamic GM (1, 1) model to predict the PM2.5 content of Changchun City, the trend of the data has been obtained, and there is a certain error.

Forecasting process

According to the characteristics of the data used in this research, combined with the practical problems, the flow chart for predicting the content of PM2.5 is established, which is shown in Figure 1:

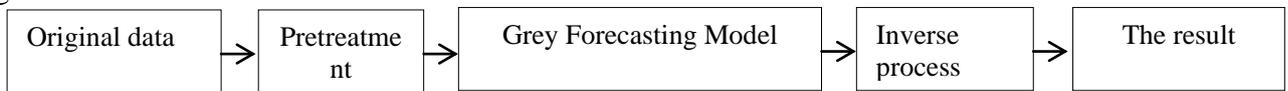


Fig. 1 Prediction of flow chart

For the working principle of the forecast model is described as follows:

(1) For raw data acquisition, data in this study was applied by 10 air quality monitoring stations in real-time monitoring data Changchun City has been the daily average PM2.5 concentration values, true and reliable.

(2) Since the original data changes even more, in order to make better prediction of the original data preprocessing, including normalization, to eliminate differences in dimension on the data itself, and in order to improve the smoothness of the data itself, applied to the data processing weakening operator, it aims to improve the prediction model can be obtained by better predict and reduce the prediction error.

(3) Binding studies the characteristics of the sample data, based on the classic gray GM (1, 1) model, a prediction model for this study, and specific algorithms can refer to formula (1) to formula (4).

(4) Will give the corresponding prediction by the prediction model, since the data prior to prediction of pre-processing operations do so after obtaining prediction, the prediction data for the corresponding inverse processing operations.

(5) Obtained by the inverse process the final prediction, the prediction result is the same level with the original data, the final prediction data directly understandable.

Simulations

Data Processing

For the data to support this study, we collected Changchun value for PM2.5 monitoring data as a sample, the sample total of 20 sets of data (December 1, 2015 to December 20), true and reliable data. Raw data obtained are shown in Table 1: PM2.5 Content Unit: ($\mu\text{g} / \text{m}^3$).

Tab.1 Initial data

Date	PM _{2.5}	Date	PM _{2.5}
01	62.1	11	86.3
02	74.3	12	150.3
03	54.0	13	239.9
04	30.0	14	209.0
05	29.8	15	58.8
06	53.8	16	35.0
07	125.4	17	47.0
08	108.5	18	71.1
09	104.3	19	54.0
10	71.6	20	50.3

In order to verify the effect of the prediction model, and in order to improve the accuracy of forecasting, reduce the computational model, the paper raw data were normalized, rule processing with reference to formula (5), so that the original data were between 0.1 to 0.9, eliminating the difference data on the dimension itself, the processed data as shown in table 2.

$$x' = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \times 0.8 + 0.1 \quad (5)$$

Where, x is the raw data obtained, the minimum measured for the sample x_{\min} , x_{\max} measured sample for the maximum. x' is the normalized value of a post.

Tab.2 After normalization of data

Date	PM _{2.5}	Date	PM _{2.5}
01	0.2230	11	0.3151
02	0.2694	12	0.5588
03	0.1921	13	0.9000
04	0.1008	14	0.7823
05	0.1000	15	0.2104
06	0.1914	16	0.1198
07	0.4640	17	0.1655
08	0.3997	18	0.2573
09	0.3837	19	0.1921
10	0.2592	20	0.1781

In order to make more intuitive predictive value for the predicted values by the corresponding prediction model was further anti-normalization process to obtain the final prediction.

Application of gray GM (1,1) prediction model can be introduced when the relevant operator of the raw data processing, common operator mainly strengthening and weakening operator, which will weaken the operator changes the speed of the sub-data sequence slowdown and the operator will strengthen even more dramatic changes in the original data[5]. But whether it is strengthening or weakening operator, introduce the operator is intended to increase the smoothness of the original data. According to the content of PM2.5 Changchun data values characteristic (volatile), in this study we introduce weakening operator to increase the accuracy of predictions. Specific operator described as follows:

The raw sequence data

$$x = (x_1, x_2, \dots, x_n) \quad (6)$$

The m is set to weaken operator, the weakening of the data sequence is processed

$$xm = (x_1 m, x_2 m, \dots, x_n m) \quad (7)$$

Among them:

$$x_k m = \frac{1}{n-k+1} (x_k + x_{k+1} + \dots + x_{kn}) \quad k = 1, 2, \dots, n \quad (8)$$

Simulation Process and Results

Flowchart entire prediction model shown in Figure 1, in this model, for the gray GM (1,1) part, a dynamic prediction mode equal dimensions, from 20 sets of data obtained, each selected data 6 The

gray dynamical prediction values [6]. Information obtained by rolling forecasts 7th to 20th prediction data values, combined with the prediction GM (1,1) Model two days of data values, that is, Changchun December 21 and 22 final prediction value for PM2.5 (to predict the results of the anti-normalization). The results are shown in Table 3. Also lists the prediction error value and the actual value in the table. Where in the PM2.5 concentration values in units of: ($\mu\text{g} / \text{m}^3$).

Tab.3 Comparison between predicted and actual values

Date	Actual value	Prediction	Error(%)
December 21st	145.6	123.8	14.97
December 22nd	188.6	172.9	8.32

Analyses

Through the above table it can be seen that the predictive model (GM (1,1)) for predicting PM2.5 concentration values of Changchun there are some errors, and the predicted first data error is larger, while the second data error small. Observation data shows that the content of the sample prior to December 21 of a smaller value. According to forecasts the characteristics and principles of the model so that the prediction error is too large, and the predictive value of December 22 and the actual value of the error is small, reflecting the effectiveness of the prediction model, although there are some errors, but predicted more ideal to verify the validity of the model.

Conclusions

In this paper, through the VS2010 platform C++ language to achieve, the PM2.5 content of Changchun city as the original data to predict the original data. After the raw data preprocessing, as a dynamic prediction data GM (1,1) model, this study collected the CCP 20 sets of data, and then the next 2 data to predict, and lists the predicted value and the actual value between errors. The results show that the error is small; it can be effective predictive model Changchun PM2.5 content.

Acknowledgement

In this paper, the research was sponsored by the Education Department of Jilin province "Twelve five " planning scientific research: (Project No. Ji Zi [2015] No. 183) and 2015 Institute of higher education and scientific research of Jilin Province: (Project No. JGJX2015D34).

References

- [1] AI Hong-fu, SHI Ying. Study on Prediction of Haze Based on BP Neural Network [J] Computer Simulation. 2015.1
- [2] ZHANG Chuan-ke. Application of Grey Systematic Theory in the Comprehensive Assenment of markenty Variety [J]. Journal of Anhui Agri Sci. 2005.33 (7):1214, 1311.
- [3] Luo Kaiyuan, Tian Tianxiong. Application of Greg Control System Forecasting Forest-stand Growth [J]. Journal of Guizhou University (Natural Science). 1998.15(1);16~20.
- [4] ZHANG Zhenghu, YUAN Mengke, DENG Jianhui, XUE Shouning. Displacement Prediction of slope Based on imprived Gery-Time series time-varying model [J]. Chinese Journal of Rock Mechanics and Engineering. 2014.33(2):3791~3796.
- [5] Sun Xiaoyun, Zhang Tao, Wang Zhendong, Wang Mingming. On the Prediction Method of Ultimate Resistance Capacity of Rock Bolt Based on the Improved Grtay GM(1,1) Model [J]. Journal of Hebei Normal University/ Natural Science Edition: 2014.38(6):589~595.
- [6] WANG Xi-ning, NIE Ting-yang, CHEN Lu, LI Shi-min, LIANG Jie. The Forcast Method of the Citizens' FoodBase on the Improved Gray System Model [J]. Journal of System Science. 2014.22(4):81~84.