

## Two GM (2,1) Improved Model to Predict Air Quality Index

Qianwen An

School of electrical and electronic engineering, North China Electric Power  
University, Baoding 071000, China.

765456234@qq.com

**Keywords:** Least square method to improve    Dimension number replacement    Changes in air quality

**Abstract.** At present, air pollution becomes coordinated development of Beijing-Tianjin-Hebei area major issues. This paper studies the Beijing-Tianjin-Hebei regional air quality issues. In view of the Beijing-Tianjin-Hebei 13 city air quality data, the data properly filtered, by finding and analyzing data, eventually selecting the air quality index (AQI) as a description of the air quality index. AQI complex and irregular, a day was used as observation units, AQI fluctuate in a month, but AQI slow to change in a day, so in hours on the AQI forecast. First, the GM based on least squares (2,1) model to predict the AQI, Data inspection, we find that the result of long-term forecasting model of relative error is larger. For further optimization models, creative introduction of such number of replacement method in this paper to improve the Least square method GM (2,1) model. Model shows, such as grey replacement model can improve the AQI forecast accuracy for a long time.

### 1. Introduction

Beijing-Tianjin-Hebei coordinated development is the major national strategy, has become a hot topic. Promote the development of Beijing-Tianjin-Hebei synergy, is a major strategic decision made by the Party Central Committee and the State Council. Air pollution become the major bottleneck of regional development. Promote the development of Beijing-Tianjin-Hebei synergy, to control air pollution and improve the ecological environment as an important breakthrough in coordinated development of Beijing, Tianjin and Hebei, take the lead in making progress on collaborative prevention and control of atmospheric pollution. Establishing mathematical model of the air quality in Beijing, Tianjin and Hebei region, predicted air quality, convenient pedestrian travel arrangements in advance.

Urban air quality in Beijing, Tianjin and Hebei 13 large amount of data and contains more indicators, with all indicators to analyze the changes in air quality, complex and cumbersome, so the data should be a reasonable screening. Based on the AQI (The air quality index) value definition of AQI description air quality indicator, AQI description index, the greater the pollution the more serious, greater harm to human health. After all the discussion, will use the AQI reflect the air quality situation.

In order to get the air quality in Beijing, Tianjin and Hebei province 13 variations take AQI average day, drawing AQI values in Beijing, Tianjin and Hebei 13 January change in the image(Figure 1).

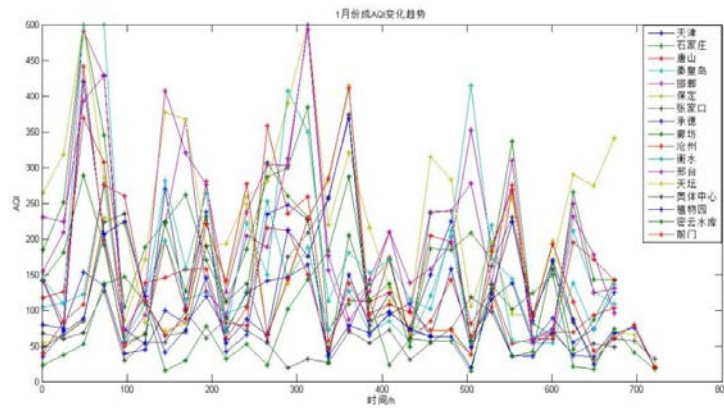


Figure 1: changes in January Beijing, Tianjin and Hebei province 13 AQI line chart

According to the image can be seen within one month did not rule any AQI values every day, up and down, it cannot be in days the AQI forecast. In order to further study the changes of AQI, rendering the AQI in the day in Beijing, Tianjin and Hebei province(Figure 2).

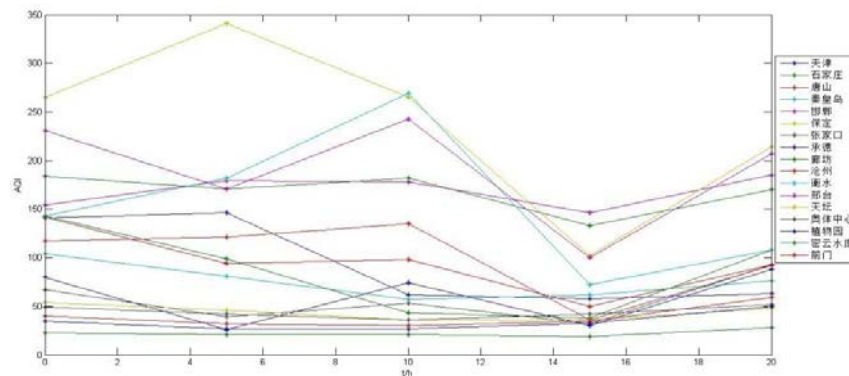


Figure 2: a change of days AQI

Can be seen from figure II AQI change is slow in one day, there is a certain range, is predictable, so gray value of AQI forecasts to predict the future.

Grey forecasting to predict the evolution of random data, but applies only to short-term forecasts, long-term forecasting errors. Hourly data is given and the AQI value, so we measured in hours, based on the nature of the least square method to improve GM (2,1) 【1】 model to describe the change of AQI.

## 2. Assumptions

1. using the data is accurate and true, ignore the effects of other pollutants in the air.
2. assuming that within each region without natural factors (dust storms, heavy rains).
3. the soft factors that affect air quality in the region (such as population and industrial development) remained stable.
4. assuming that the city on behalf of the entire city.

## 3. Predict changes in air quality——Least squares modified GM (2,1) forecast model

### 3.1 Gray forecast model introduced

Grey theory holds that the behavior of the system is a hazy, the data is complex, but it is, after all, in order, is an overall function. Grey number formation, is the law of finding out from the clutter. Meanwhile, is generating data models by grey, instead of the original data model, therefore, gray forecast data is generated by the GM (2,1) models obtained by this method of inverse consequences.

Least square method to improve GM (2,1) model to determine the general solution  $X^{(1)}(t) = C_1 e^{\lambda_1 t} + C_2 e^{\lambda_2 t} + b/a_2$  undetermined coefficients C1 and C2 in more accurate error is smaller.

### 3.2 Modeling

Improved GM established the least squares (2,1) model.

First, given a raw sequence is:

$$x^{(0)} = (x^{(0)}(1), x^{(0)}(2), \dots, x^{(0)}(n)),$$

A progressive generation of sequence as follows:

$$x^{(1)} = (x^{(1)}(1), x^{(1)}(2), \dots, x^{(1)}(n)),$$

A regressive sequence that is generated as follows:

$$\alpha^{(1)} x^{(0)} = (\alpha^{(1)} x^{(0)}(2), \dots, \alpha^{(1)} x^{(0)}(n)),$$

X (1) generates the sequence is:

$$z^{(1)} = (z^{(1)}(2), z^{(1)}(3), \dots, x^{(1)}(n)),$$

From the above equations are GM (2,1) albino of the model equation:

$$\frac{d^2 x^{(1)}}{dt^2} + a_1 \frac{dx^{(1)}}{dt} + a_2 x^{(1)} = b \quad (1).$$

### 3.3 Model solution

The equation (1), when the characteristic equation

$$\lambda^2 + a_1 \lambda + a_2 = 0$$

When the discriminant  $\Delta = a_1^2 - 4a_2 > 0$ , you can get two real roots

$$\lambda_1 = \frac{-a_1 + \sqrt{\Delta}}{2}, \lambda_2 = \frac{-a_1 - \sqrt{\Delta}}{2}$$

Heat equation (1) General solutions for:

$$X^{(1)}(t) = C_1 e^{\lambda_1 t} + C_2 e^{\lambda_2 t} + \frac{b}{a_2} \quad (2)$$

Where C1, C2 is the undetermined coefficients.

Determine the values of C1, C2, we used [1]:

$$\varepsilon(t) = X^{(1)}(t) - \widehat{X^{(1)}(t)} = C_1 e^{\lambda_1 t} + C_2 e^{\lambda_2 t} + \frac{b}{a_2} - \widehat{X^{(1)}(t)} \quad (3)$$

Ask C<sub>1</sub>, C<sub>2</sub>, order::

$$f(C_1, C_2) = \sum_{i=1}^n \varepsilon(t_i)^2 = \sum_{i=1}^n \left[ C_1 e^{\lambda_1 t_i} + C_2 e^{\lambda_2 t_i} + \frac{b}{a_2} - \widehat{X^{(1)}(t)} \right]^2$$

Reaches its minimum value.

Order:

$$\frac{\partial f}{\partial C_1} = 0, \frac{\partial f}{\partial C_2} = 0$$

Calculated values of C<sub>1</sub>, C<sub>2</sub>.

In the  $C_1$  and  $C_2$  into  $X^{(1)}(t)$ , determine the expression of  $X^{(1)}(t)$  General solutions. By the time response of  $X^{(1)}(t)$  get  $X^{(1)}$ , then by the time response down to get  $X^{(0)}(t)$ .

### 3.4 Analysis of model results

When choosing Beijing Dong SI 2015-1-2 1-7 data as input, through the MATLAB computing obtained predicted values of  $X^{(0)}$ , as shown in table 1.

Time (1—2)	$X^{(0)}(t)$	$\widehat{X^{(0)}}(t)$	Relative error%
1:00 a.m	61	60.3186	1.100
2:00 a.m	63	62.8000	0.317
3:00 a.m	66	64.9976	1.510
4:00 a.m	67	66.8196	0.260
11:00 a.m	67	62.5373	6.667

Table 1: least square method to improve GM (2,1) model  $X(0)$  predictive value

Grey forecast model, given its relative error calculated as follows:

$$\frac{x_0(k) - \widehat{x_0(k)}}{x_0(k)} \times 100\% \quad (4)$$

Table a shows that least squares modified GM (2,1) model to more accurately predict AQI values, the average relative error of 1.9708%.Relative error of the predicted value of 1-4 in 2%, 11 o'clock is 6.667%, you can see that with GM (2,1) model prediction error for a long time. In order to get an accurate long-term predictions, we improved the model.

## 4. Improved model——Same Dimension Gray Recurrence Dynamic Model[2]

### 4.1 Model theory

This model is a model of an improvement, with given column GM (2,1) grey model, the first prediction of the model (gray), added after the given column, and remove the first known data, maintaining data series such as. Then again build GM (2,1) grey model and forecast next values, so advance, one forecast, which they were elected until the completion of task or reach a certain accuracy requirements.

### 4.2 Modeling

(1) To determine the optimal number of dimensions.

General GM (2,1) grey model takes only 4 data sequences to modeling and forecasting, select variable-length data sequences for modeling, prediction and accuracy will be different. In order to improve the predictive accuracy of the model, we chose the original data sequences of different length in different dimensions of GM (1,1) model(Table 2)l, in order to achieve the objective of selecting optimal prediction model. Original data selection for Beijing April or January first data.

Serial number	Dimension	Data series	Model
Model 1	4	1—4 时	$x^{(1)}(t) = 1420.4e^{0.0403t} + 0.000015e^{2.1790t} - 1420.6$
Model 2	5	1—5 时	$x^{(1)}(t) = 1441.3e^{0.0397t} - 0.001847e^{1.2824t} - 1441.3$
Model 3	6	1—6 时	$x^{(1)}(t) = 1219.7e^{0.0467t} - 0.542562e^{0.5363t} - 1218.5$
Model 4	7	1—7 时	$x^{(1)}(t) = (350.07 - 150.69i)e^{(0.099-0.045i)t} + (350.07 + 150.69i)e^{(0.099+0.045i)t} - 698.8$
Model 5	8	1—8 时	$x^{(1)}(t) = (-275.77 - 328.94i)e^{(-0.003+0.082i)t} + (-275.77 + 328.94i)e^{(-0.003-0.082i)t} + 554.0$
Model 6	9	1—9 时	$x^{(1)}(t) = (-878.08 + 206.14i)e^{(-0.044+0.055i)t} + (-878.08 - 206.14i)e^{(-0.044-0.055i)t} + 1759.1$

Table 2: different dimensions of GM (2,1) grey model

In order to determine the optimal number of dimensions, we are going to Beijing in April or January data as test data to verify the accuracy of the model. Test results are shown in table 3.

Serial number	2nd-2 a.m	2nd-3 a.m	2nd-4 a.m	2nd-5 a.m	The average relative error%
Actual value	61	63	66	67	
Model 1 (4)	60.742	63.245	65.913	69.247	0.86
Model 2 (5)	60.680	63.092	65.485	67.557	0.46
Model 3 (6)	60.501	62.961	65.229	67.077	0.43
Model 4 (7)	60.319	62.800	64.998	66.820	0.64
Model 5 (8)	60.164	62.681	64.856	66.631	0.83
Model 6 (9)	60.016	62.616	64.784	66.506	0.96

Table 3: different dimensions of GM (2,1) comparison of model results

Table 3 results show, the six model could accurately predict the AQI value, the average relative error of less than 1%. In contrast, models, six-dimensional grey forecasting model of relative error smaller, higher accuracy, so we choose six GM (2,1) grey replacement model for AQI value prediction.

(2) Building dynamic GM (2,1) model.

First known columns six GM (2,1) grey model method and least square method to improve GM (2,1) model is very similar to, not detailed here. After getting the first predicted value will be added

after the given column and remove the given column the first known data, maintaining data series such as, and then re-establish GM (2,1) model, and so on, until you reach the desired objectives or precision.

#### 4.3 Analysis of model results

In order to observe the improved GM (2,1) grey prediction model can more accurately predict air quality through comparison of the results of the two models to the MATLAB figure(Figure 3).

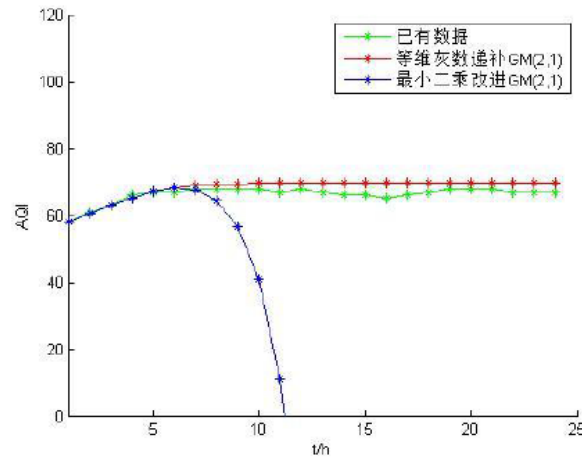


Figure 3: prediction results

Figure III shows the number of dimensions, such as ash filled the improved model predictive accuracy was higher than that of the original model, and the forecast time horizon is much larger than the original model, the improved GM (2,1) models to achieve the desired effect.

#### 5. Model analysis

Using error analysis method of grey prediction model of error. Error is mainly grey prediction model predictions and the deviation between the original data. Whitening of the grey forecast due out equations and finally abstract function is formed by the original fitting of the data, so the data cannot be absolutely precise forecast and raw data one by one. Dimension number replacement improved model of grey number range and accuracy than the original model, the improved GM (2,1) models to achieve the desired effect, you can predict the future air quality conditions in a day.

#### 6. Model advantages

1 Will be necessary to omit extraneous data that can reduce a model of computation, but also ensure the accuracy.

2 AQI values without the law, based on the laws of nature change GM (2,1) forecast that AQI values, error is smaller.

3 Dimension number replacement grey model of innovation and improvement of least squares GM (2,1) model, improve the range of forecasts, reducing the relative error.

#### References

- [1] Shen Jihong, Zhao Xiren. Use of least square method in improving GM (2, 1) model [j]. Journal of Harbin Engineering University, 2001, 22 (4): 64-66.
- [2] Zhang Jingjing, Ma Dongdong. dimensional grey replacement model in the prediction of population – case study in Zhejiang Province [j]. Wenzhou University Journal of natural sciences, 2011, 32(4): 37-42.
- [3] Si Shoukui, Sun Xi cyanine. mathematical modeling algorithms and applications. defense industry publishing house, 2014.

[4] Zhang Junliang. grey control system model for forecasting analysis [j]. computers and networks.2012(11).

[5]AQI, <http://baike.so.com/doc/5411492-5649600.html>.