ISSN: 2688-8645

AQI Prediction Based on GMC(1,n) Model

Lanxi Zhang[1, a, *]

[1]School of Science, Southwest University of Science and Technology, Mianyang, 621000, China

[a]Corresponding author Email: zhanglanxi1203@163.com

Abstract

Air quality affects people's daily life and physical and mental health. As a quantitative description of air quality, AQI is an important basis for measuring air quality. It can accurately reflect the actual air quality and the evaluation results, so the prediction of AQI is of great significance. Therefore, this paper takes AQI as the dependent variable sequence and PM2.5 as the independent variable sequence, applies the $GMC(1,n)$ gray prediction model to predict the AQI of Chengdu and Beijing, and compares with the three gray models. The comparative analysis of the average absolute percentage error and relative error of each model shows that the $GMC(1,n)$ model has good performance and can be used as an effective method for AQI prediction.

Keywords

AQI; GMC (1,n).

1. Introduction

AQI describes how clean or polluted the air is and how it affects health. The Environmental Protection Agency calculated the air quality index to pass five major pollution standards: ground-level ozone, particulate pollution, carbon monoxide, sulfur dioxide, and nitrogen dioxide. For these pollutants, the Environmental Protection Agency has established a national ambient air quality standard to protect public health. The two pollutants, ground-level ozone and airborne particles, constitute the country's greatest threat to human health. In this context, the establishment of a scientific and reasonable AQI prediction model will play a positive role in the treatment of air pollution.

Grey system theory is founded in 1982 by the Chinese scholar Professor Deng Julong [1]. It is a method specially used to study uncertain system problems where "some information is known and some information is unknown". The gray prediction of the change of a certain system's characteristic sequence is bound to be affected by a variety of influencing factors, and this mechanism of action may include the essential characteristics of uncertainty and incomplete information. In order to solve the trend prediction problem of this type of system behavior sequence, Professor Deng Julong proposes a multivariate GM (1, n) model on the basis of the univariate model [2].Many scholars have also improved the multivariate gray model on this basis. Guo et al. [3] combined the advantages of the self-memory principle and proposed a self-memory multivariate prediction model. Aiming at the problem of unsatisfactory simulation and prediction accuracy of traditional models, Zeng et al. [4] introduced linear correction terms and gray action terms to overcome the structural defects of the traditional model have improved the model's ability to describe the law of data changes and interdependence between variables. Tien [4] based on the convolution integration technology, proposed the GMC (1, n) model, and proposed a series of improved models [5, 6] to improve the processing ability of multiple feature sequences.

The above methods have effectively improved the modeling effect of the GM (1, n) model. The calculation of AQI is affected by ground-level ozone, particulate pollution, carbon monoxide, sulfur dioxide, and nitrogen dioxide. Therefore, this paper adopts multivariate gray prediction to improve the prediction accuracy. This paper will apply the GMC(1,n) model proposed by Tien with AQI as the dependent variable sequence and PM2.5 as the independent variable sequence to predict the AQI in Beijing and Chengdu, and compare it with three traditional gray prediction models. Contrast, test the practicability of the model, and obtain a practical method for AQI prediction.

2. AQI Prediction based on GMC(1,n) Model

2.1. GMC(1,n)

Tien proposed a new GM(1,n) model with convolution solution in 2005, referred to as GMC(1,n) model for short. The GMC(1,n) model is of great significance to the development of grey forecasting theory, and it greatly expands the accuracy and application scope of the multivariate grey forecasting model. This article first discusses the GMC(1,n) model. The basic theory is reviewed.

Let $X_1^{(0)}$ be the dependent variable sequence:

$$X_1^{(0)} = (X_1^{(0)}(1), X_1^{(0)}(2), \cdots, X_1^{(0)}(m)) \tag{1}$$

Sequence $X_i^{(0)}(i = 2,3,\cdots,n)$ is a sequence of independent variables that is highly correlated with sequence $X_1^{(0)}$ and is called a correlation factor sequence:

$$X_i^{(0)} = (X_i^{(0)}(1), X_i^{(0)}(2), \cdots, X_i^{(0)}(m)) \tag{2}$$

$X_i^{(1)}$ is the first-order cumulative generating sequence of $X_i^{(0)}(i = 2,3,\cdots,n)$,abbreviated as $1 - AGO$ sequence:

$$X_i^{(1)} = (X_i^{(1)}(1), X_i^{(1)}(2), \cdots, X_i^{(1)}(m)) \tag{3}$$

Where

$$X_i^{(1)}(k) = \sum_{j=1}^{k} X_i^{(0)}(j) \quad j = 1,2,\cdots m \tag{4}$$

Then the whitening equation of model $GMC(1,n)$ is called

$$\frac{dX_1^{(1)}(rp+t)}{dt} + aX_1^{(1)}(rp+t) = \sum_{i=2}^{n} b_i X_i^{(1)}(t) + u \tag{5}$$

Where

$$X_1^{(1)}(rp+t) = \sum_{k=rp+1}^{rp+t} X_1^{(0)}(k)\,, t = 1,2,\ldots,r+rf \tag{6}$$

And

$$X_1^{(1)}(t) = \sum_{k=1}^{t} X_i^{(0)}(k)\,, t = 1,2,\ldots,r+rf \tag{7}$$

$1-AGO$ sequence of $X_1^{(0)}(t)$ and $X_i^{(0)}(t)(i=2,3,\dots,n)$ respectively.r is the number of data points used for modeling, rp is the time delay coefficient, and rf is the number of points used for prediction. a and $b_2, b_{2,\dots,}b_n$ are parameters to be estimated.

The background values of $X_1^{(0)}(t)$ and $X_i^{(0)}(t)(i=2,3,\dots,n)$ are defined as

$$Z_1^{(1)}(rp+t)=\frac{1}{2}[X_1^{(1)}(rp+t)+X_1^{(1)}(rp+t-1)] \tag{8}$$

And

$$Z_i^{(1)}(t)=\frac{1}{2}[X_i^{(1)}(t)+X_i^{(1)}(t-1)] \tag{9}$$

Thus the whitening equation can be approximately written as

$$X_1^{(0)}(rp+t)+aZ_1^{(1)}(rp+t)=\sum_{i=2}^{n} b_i Z_i^{(1)}(t)+u \tag{10}$$

Using the above formula, the least square estimation formula of $GMC(1,n)$ model can be obtained

$$[a,b_2,b_{2,\dots,}b_n,u]^T=B^T(B^TB)^{-1}Y_R \tag{11}$$

Where

$$B=\begin{bmatrix}-Z^{(1)}(rp+2) & Z_2^{(1)}(2) & \cdots & Z_n^{(1)}(2) & 1\\ -Z^{(1)}(rp+3) & Z_2^{(1)}(3) & \cdots & Z_n^{(1)}(3) & 1\\ \vdots & & \vdots & & \vdots\\ -Z^{(1)}(rp+r) & Z_2^{(1)}(r) & \cdots & Z_n^{(1)}(\mathrm{r}) & 1\end{bmatrix} \tag{12}$$

$$Y_R=[X^{(0)}(2)\quad X^{(0)}(3)\quad \cdots \quad X^{(0)}(r)]^T \qquad \{13\}$$

The solution of the whitening equation is

$$\widehat{X_1^{(1)}}(rp+t)=X_1^{(0)}(rp+t)e^{-a(t-1)}+\int_1^t e^{-a(t-\tau)}f(\tau)dt \tag{14}$$

Where

$$f(t)=\sum_{i=2}^{n} b_i X_i^{(1)}(t)+u \tag{15}$$

The discrete solution of the $GMC(1,n)$ model obtained by using the two-point trapezoidal formula to discretize the solution of the whitening equation is

$$\widehat{X_1^{(1)}}(rp+t)=X_1^{(0)}(rp+t)e^{-a(t-1)}+u(t-2)\times\sum_{\tau=2}^{t}\frac{1}{2}\{e^{-a(t-\tau)}f(\tau)+e^{-a(t-\tau+1)}f(\tau-1)\} \tag{16}$$

Where $u(t-2)$ is the unit step function, defined as

$$u(t-2)=\begin{cases}0, t<2\\1, t\geq 2\end{cases} \tag{17}$$

Using any discretization, the predicted value $\widehat{X_1^{(1)}}$ of the sequence $1-AGO$ of the characteristic sequence can be calculated, and the predicted value $\widehat{X_1^{(0)}}$ of the characteristic sequence can be obtained by the following cumulative reduction equation

$$\widehat{X_1^{(0)}}(rp+t)=\widehat{X_1^{(1)}}(rp+t)-\widehat{X_1^{(1)}}(rp+t-1) \tag{18}$$

2.2. Case Analysis

In simple terms, AQI is data that can quantitatively describe air quality, which is mainly affected by daily power generation, industrial production, and automobile exhaust emissions. As cities with rapid economic development in my country, Beijing and Chengdu should pay more attention to air quality. PM2.5 has become one of the most important factors affecting people's health in our country, and it is also the main indicator for calculating AQI. Therefore, this article takes AQI as the dependent variable sequence and PM2.5 as the independent variable sequence. Therefore, the AQI and PM2.5 data of Beijing and Chengdu are collected, and $GMC(1,n)$ is applied to the prediction research of urban AQI. The first 8 data are used as the basic data to build the model, and the latter two data are used as the prediction. And compared with the traditional gray models $GM(1,1)$, $DMC(1,1)$ and $GM(1,n)$.

2.2.1 Forecast Data

In this paper, the AQI values of Beijing and Chengdu for ten consecutive months are randomly selected as the basic data, and PM2.5 is used as the sequence of related factors. The specific data are shown in Table 1.

Table 1. Monthly content of AQI and PM2.5

Time	Chengdu		Beijing	
	AQI	PM2.5	AQI	PM2.5
2018.01	108	80	64	34
2018.02	87	64	80	50
2018.03	84	58	115	82
2018.04	84	45	98	59
2018.05	78	33	91	45
2018.06	69	25	111	43
2018.07	66	20	89	44
2018.08	87	31	83	31
2018.09	54	22	60	28
2018.10	65	42	74	42

2.2.2 Forecast Result

The data of Beijing and Chengdu with the data node of 10 are divided into two parts. The first part is the data set of the first 8 time-nodes, which is used as the training set of the model; the second part is the data set of the last 2 time-nodes, using $GMC(1,n)$ model makes predictions.

In order to reflect the generalization ability of the model used in this article, we compare the other three gray models, using $GM(1,1)$, $DMC(1,1)$ and $GM(1,n)$. MAPE is used as a comparison model The quantitative indicators of prediction ability, the prediction results of each model and the MAPE value are shown in Table 2.

Table 2. Forecast error of Chengdu AQI monthly data

Time	Chengdu				
	Relative value	GMC(1,n)	GM(1,n)	GM(1,1)	DGM(1,1)
2018.01	108	0.	0.	0.	0.
2018.02	87	-0.06728772	-9.54616499	0.30619435	0.32511393
2018.03	84	1.64052534	33.46152687	0.54012266	0.55037874
2018.04	84	-1.06288767	19.00557844	-2.13831319	-2.13618988
2018.05	78	0.56185758	-1.68189761	1.2681103	1.26260658
2018.06	69	4.15023071	-13.53381283	7.75670458	7.74405569
2018.07	66	1.39604788	-24.07592881	8.32486627	8.30553117
2018.08	87	-24.04685791	-28.10097823	-15.02992552	-15.05550981
2018.09	54	5.13040328	-12.61151516	15.68988821	15.65847075
2018.10	65	-8.15171123	11.60059908	2.48194376	2.44508904
MAPE		6.18279800	20.51843318	15.83417573	15.81243019

Table 3. Forecast error of Beijing AQI monthly data

Time	Beijing				
	Relative value	GMC(1,n)	GM(1,n)	GM(1,1)	DGM(1,1)
2018.01	64	0.	0.	0.	0.
2018.02	80	41.47064318	-19.28769461	13.81005796	14.72647023
2018.03	115	22.39190836	32.00779689	-19.74443668	-19.42397357
2018.04	98	24.65794801	11.41117073	-1.27665775	-1.56679812
2018.05	91	5.69248209	-10.72732379	7.21373798	6.29806493
2018.06	111	-26.35650402	-36.34616004	-11.272901	-12.82931548
2018.07	89	-6.47952035	-13.31442006	12.26377917	10.05113022
2018.08	83	-15.94141104	-29.79163891	19.82413782	16.93947221
2018.09	60	-3.48969477	-11.99090405	44.40853981	40.83578132
2018.10	74	-4.293303	-2.01227122	32.01735562	27.74012899
MAPE		16.45755711	18.16767820	41.75245428	38.52497385

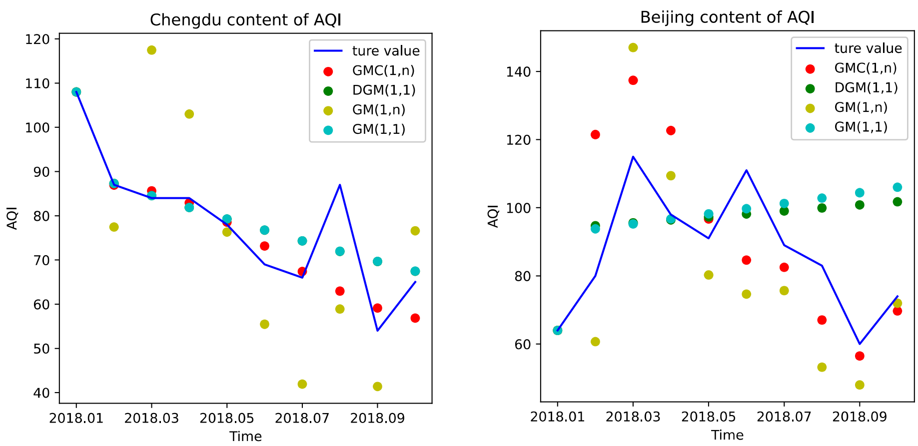

Figure 1. AQI

Table 2 shows the prediction error of AQI content in Chengdu from January to October 2018. From the table, it can be seen that compared with the other three gray models, $GMC(1,n)$ has better prediction accuracy. Table 3 shows the prediction error of AQI content in Beijing from January to October 2018. From the table, it can be seen directly that $GMC(1,n)$ has a smaller MAPE. In order to more intuitively reflect the predictive ability of the model, this article compares four The prediction results of this model are displayed, as shown in Figure 1. It can be clearly shown from the figure that the forecast result of $GMC(1,n)$ is closer to the original time series. Therefore, the $GMC(1,n)$ model presents excellent generalization performance in the prediction of AQI in Chengdu and Beijing, and can be used as an effective tool for AQI prediction.

3. Conclusion

The calculation of AQI is affected by multiple indicators. If only univariate gray prediction is performed on AQI, larger errors may occur. Therefore, this article considers the multivariate $GMC(1,n)$ gray model, which is further improved on the basis of multivariate. Forecast accuracy to reduce errors and get more accurate results. Judging from the AQI prediction results in Chengdu and Beijing, the $GMC(1,n)$ model is significantly better than the other three gray models, with higher prediction accuracy and higher reliability, and can be used as a quantitative prediction method for AQI in actual practice in the application.

Acknowledgments

This work is supported by Southwest University of Science and Technology College Students' 2021 Innovation Fund Project Accurately Funded Special Project. (JZ21-054).

References

[1] On the Accumulation Generation in Grey System Theory [J]. Systems Engineering Theory and Practice, 1982(05):10-15.

[2] Deng J L. The control problem of grey systems[J]. System Control Letter, 1982, 1(5): 288-294.

[3] Guo X J, Liu S F, Wu L F, et al. A multi-variable grey model with a self-memory component and its application on engineering prediction[J]. Engineering Applications of Artificial Intelligence, 2015, 42: 82-93.

[4] Zeng B, Luo C M, Liu S F,et al. Development of an optimization method for the GM(1, N) model[J]. Engineering Applications of Artificial Intelligence, 2016, 55: 353-362.

[5] Tien T L. The indirect measurement of tensile strength of material by the grey prediction model GMC(1, N)[J]. Measurement Science Technology, 2005, 16(6): 1322– 1328.

[6] Tien T L. Forecasting CO2 output from gas furnace by grey model IGMC(1, N)[J]. J of the Chinese Society of Mechanical Engineers, 2010, 31(1): 55–65.

[7] Tien T L. The indirect measurement of tensile strength by the new model FGMC(1, N)[J]. Measurement, 2011, 44(10): 1884–1897.