ICIC Express Letters
Part B: Applications
Volume **7**, Number **1**, January **2016**

ICIC International ©2016 ISSN 2185-2766
pp. **97–103**

AN IMPROVED GREY MODEL FOR URBAN AIR QUALITY FORECAST

Yuan Huang, Guoyan Huang and Jiadong Ren

College of Information Science and Engineering
Yanshan University
No. 438, Hebei Avenue, Qinhuangdao 066004, P. R. China
757918272@qq.com

Received June 2015; accepted September 2015

Abstract. *The forecast of air quality is critical to enable proper precautions to be taken before and during certain events. Urban air quality varies non-linearly and depends on multiple factors, which will lead to the low forecast accuracy. In order to improve the forecast accuracy, an improved $GM(1,1)$ model called $IOBGM(1,1)$ model is proposed to forecast urban air quality in this paper. The original sequence and the background value of the $GM(1,1)$ model are simultaneously optimized to improve the forecast accuracy based on the analysis of the factors that can influence the forecast accuracy. The improved $GM(1,1)$ model is employed to forecast the concentrations of PM_{10} and SO_2 of Handan before 2020, one of the most polluted cities in China. Experimental results show that the improved $GM(1,1)$ model has higher forecast accuracy, the forecast results of which are much more reliable.*

Keywords: Air quality, Grey model, Optimizing, Forecast

1. **Introduction.** The air quality information, such as the concentration of particulate matter (PM_{10}), nitrogen dioxide (NO_2) and sulfur dioxide (SO_2), is of great importance to support air pollution control and protect human health from damage by air pollution. Most major cities in the northeast of China, especially Beijing, Tianjin and Hebei province (Jing-Jin-Ji area) have experienced severe short-term pollution events. Accurate forecast for the changing trend of urban air quality is helpful for providing the public and governments with important information, according to which effective measures are timely taken to prevent and control emissions of air pollutants.

The forecast approaches of urban air quality fundamentally branch into two main streams: deterministic approaches and statistical approaches [1]. Deterministic approaches can be performed without a large quantity of historical data, but it demands sufficient knowledge of pollutant sources, the real-time emission quantity and temporal physical processes under the planetary boundary layer. Statistical forecasting requires relatively less detailed data and it is inexpensive and easy to operate, and, is relatively accurate. The commonly used statistical forecast approaches include regression analysis method, artificial neural network, etc. [2, 3, 4]. Zheng et al. [6] combined artificial neural network with linear regression to infer the real-time and fine-grained air quality. The experiments show the mean forecast accuracy is higher than that of the regression tree method. Wang et al. [5] utilized a hybrid artificial neural network to enhance the forecast accuracy by revising the error term of the traditional method. However, the artificial neural network models need a large number of measured data to track the change of the data for obtaining better forecast results. However, the acquisition of a large amount of data is often difficult.

The grey system theory, originally presented by Deng [7], focuses on model uncertainty and information insufficiency in analyzing systems via researches on forecasting and decision making. Limited time and space impose restriction on the data acquisition of air quality forecast which always contains incomplete information. So the air quality forecast system belongs to a typical grey system.

The $GM(1,1)$ model of grey theory can use fewer data for modeling and forecasting, thereby making up for the inadequacy of the artificial neural network method. Pai et al. [9] utilized seven types of $GM(1,1)$ model to forecast hourly PM_{10} and SO_2 concentrations in Banciao city of Taiwan. Pan et al. [8] employed grey dynamic model group to forecast the air quality changing trend of Tianjin in China. The experimental results show that $GM(1,1)$ model is an efficiently early warning tool for providing particulate matter information for the public. In order to improve the forecast accuracy, an improved $GM(1,1)$ model called $IOBGM(1,1)$ model is proposed in this paper, the original sequence and the background value of which are optimized at the same time. Experimental results show that the $IOBGM(1,1)$ model has higher forecast accuracy, the forecast results of which have the important reference for the assessment, management and decision making of urban air quality.

The rest of paper is organized as follows. Section 2 introduces the traditional grey forecast model of air quality and the impact factors of forecast accuracy of model. An improved air quality forecast model called $IOBGM(1,1)$ model is given in Section 3. Experimental results and analysis are given in Section 4. Section 5 concludes the paper.

2. The Grey Forecast Model of Air Quality.

2.1. The traditional $GM(1,1)$ forecast model.

The traditional $GM(1,1)$ model is shown as follows.

Step 1 Denote the original sequence as:

$$x^{(0)} = \left\{x^{(0)}(1), x^{(0)}(2), \cdots, x^{(0)}(N)\right\} \tag{1}$$

where $N = 1, 2, 3, \cdots$.

The sequence $x^{(0)}$ for 1-AGO (Accumulated Generating Operation) can be:

$$x^{(1)} = \left\{x^{(1)}(1), x^{(1)}(2), \cdots, x^{(1)}(N)\right\} \tag{2}$$

where $x^{(1)}(k) = \sum_{i=1}^{k} x^{(0)}(i)$, $k = 1, 2, \cdots, N$.

Step 2 Establish the background value:

$z^{(1)}$ is the mean generating sequence for $x^{(1)}$. The sequence of the background value is defined as:

$$z^{(1)} = \left\{z^{(1)}(2), z^{(1)}(3), \cdots, z^{(1)}(N)\right\} \tag{3}$$

where

$$z^{(1)}(k+1) = \frac{1}{2}\left(x^{(1)}(k) + x^{(1)}(k+1)\right) \tag{4}$$

Step 3 Build up constant differential equation:

$$\frac{dx^{(1)}}{dt} + ax^{(1)} = u \tag{5}$$

where a is called the developing coefficient, and u is called the control variable.

The Ordinary Least Square (OLS) method is utilized to calculate a and u:

$$\hat{U} = [a \quad u]^T = (B^TB)^{-1}B^TY \tag{6}$$

where,

$$B = \begin{bmatrix} -z_1^{(1)}(2) & 1 \\ -z_1^{(1)}(3) & 1 \\ \vdots & \vdots \\ -z_1^{(1)}(N) & 1 \end{bmatrix} Y = \left[x_1^{(0)}(2)\ x_1^{(0)}(3)\cdots x_1^{(0)}(n)\right]^T \tag{7}$$

Step 4 Obtain the discrete form of first-order grey differential equation. The solution of $x^{(1)}$ is:

$$\hat{x}^{(1)}(k+1) = \left[x^{(1)}(1) - \frac{u}{a}\right] e^{-ak} + \frac{u}{a} \tag{8}$$

where a is called the developing coefficient, $k = 1, 2, \cdots, N$.

Thus, the forecast values can be derived by one inverse accumulated generating operation $(1 - IAGO)$:

$$\hat{x}^{(0)}(k+1) = \hat{x}^{(1)}(k+1) - \hat{x}^{(1)}(k) = (1 - e^a)\left(x^{(0)}(1) - \frac{u}{a}\right) e^{-ak} \tag{9}$$

2.2. **The impact factors of forecast accuracy of $GM(1,1)$ model.** The analysis of the traditional $GM(1,1)$ model shows the impact factors of forecast accuracy of $GM(1,1)$ model mainly include the original sequence factor and the background value factor.

The original sequence factor. The data of air quality forecast system is constantly distorted due to the interference of external factors such as human activities. The original sequence of $GM(1,1)$ model cannot correctly reflect the essential change laws of the forecast system. In order to improve the forecast accuracy of air quality, we need reduce the interference effect of disturbance factors. Therefore, generated by the original sequence, the buffer operator is proposed [10]. The buffer operator can weaken or eliminate the influence of interference in the system, modify the original data, and improve the smoothness of discrete data and the forecast accuracy of air quality.

The background value factor. According to Equation (8), the fitting and forecast accuracy of $GM(1,1)$ model depend on parameters a and u. And the values of a and u depend on the construction of the background value, so the background value sequence $z^{(1)}$ becomes the key to directly affecting the forecast accuracy of $GM(1,1)$ model. Equation (4) shows the background value $z^{(1)}(k+1)$ is the average value of $x^{(1)}(k)$ and $x^{(1)}(k+1)$ in the range $[k, k+1]$. The background value $z^{(1)}(k+1)$ is the abscissa midpoint of trapezoid, as shown in Figure 1. However, the forecast curve of $GM(1,1)$ model is exponential curve, and the midpoint of the actual curve (L) is always less than the abscissa midpoint of trapezoid $(L + \triangle L)$ in the range $[k, k+1]$. When the sequence data changes gently, $x^{(1)}(k)$ is close to $x^{(1)}(k+1)$, the difference between L and $L + \triangle L$ is very small and the deviation of the model also is small. However, when the change of sequence data is

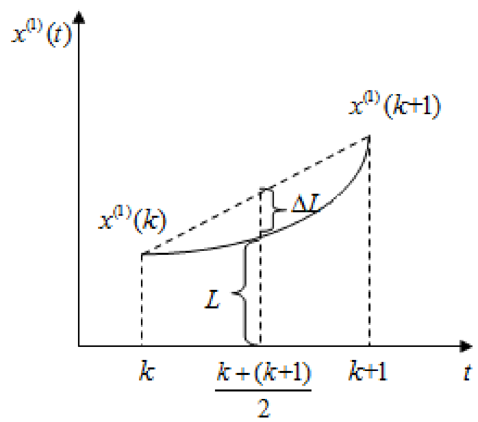

FIGURE 1. Principle diagram of the background value

intensive, the difference between $x^{(1)}(k)$ and $x^{(1)}(k+1)$ becomes larger, and the difference between L and $L+\triangle L$ also increases, thus leading to the great reduce of forecast accuracy.

The improvement to either of the impact factors is limited for improving the forecast accuracy of grey model. So the improved $GM(1,1)$ model called $IOBGM(1,1)$ model, the original sequence and the background value of which are optimized at the same time, is used to forecast the air quality and improve the forecast accuracy.

3. The Improved $GM(1,1)$ Forecast Model.

3.1. The improvement of the original sequence.

Definition 3.1. *Assume that the original sequence* $x^{(0)} = \{x^{(0)}(1), x^{(0)}(2), \cdots, x^{(0)}(N)\}$ *is the behavior data sequence of system, that* d *is an operator worked on* $x^{(0)}$*, and that the sequence, obtained by having* d *worked on* $x^{(0)}$*, is denoted as:*

$$x^{(0)}d = \{x^{(0)}(1)d, x^{(0)}(2)d, \cdots, x^{(0)}(N)d\} \tag{10}$$

where d *is called a sequence operator and* $x^{(0)}d$ *is called the first-order operator sequence.*

Theorem 3.1. ***Theorem of Fixed Points*** [10] *Assuming that* $x^{(0)}$ *is the behavior data sequence of system and that* d *is a sequence operator,* d *must satisfy* $x^{(0)}(N)d = x^{(0)}(N)$.

Theorem 3.2. ***Theorem of Sufficient Usage of Information*** [10] *Each data* $x^{(0)}(k)$ $(k = 1, 2, \cdots, N)$ *in the behavior data sequence of system participates fully in the whole process in which the operator works on* $x^{(0)}$.

Theorem 3.3. ***Theorem of Analytic Representations*** [10] *For any* $x^{(0)}(k)d$ $(k = 1, 2, \cdots, N)$ *can be described with a uniform and elementary analytic representation in* $x^{(0)}(1), x^{(0)}(2), \cdots, x^{(0)}(N)$.

Definition 3.2. *All sequence operators, satisfying the above three theorems, are called buffer operators.*

Definition 3.3. *A novel buffer operator:*

$$x^{(0)}d_1 = \frac{2x^{(0)}(k)x^{(0)}(N)}{x^{(0)}(k) + x^{(0)}(N)} \tag{11}$$

where $k = 1, 2, \cdots, N$*,* d_1 *is a buffer operator, and* $x^{(0)}d_1 = \{x^{(0)}(1)d_1, x^{(0)}(2)d_1, \cdots, x^{(0)}(N)d_1\}$ *is a new sequence obtained by the buffer operator* d_1 *working on the original sequence.*

3.2. The improvement of the background value.

According to the analysis of the impact factors of background values, to improve model accuracy, a novel background value is given as Equation (12) to improve forecast accuracy at the same time with the novel buffer operator.

Definition 3.4. *A novel background value:*

$$z^{(1)}(k+1) = \frac{N/2+1}{N}x^{(1)}(k) + \frac{N/2-1}{N}x^{(1)}(k+1) \tag{12}$$

where N *is the number of elements in the original sequence,* $N = 1, 2, 3, \cdots$.

3.3. The forecast process of the $IOBGM(1,1)$ model.

Step 1 The annual concentration of an air pollutant is selected as the original sequence. The first time is represented by $x^{(0)}(1)$, the second time is represented by $x^{(0)}(2)$ and the N time is represented by $x^{(0)}(N)$. The novel buffer operator is used to work on the original sequence, which generates a new sequence as input of $IOBGM(1,1)$ model. The form of new sequence is shown in Equation (10).

Step 2 The background value is optimized by using Equation (12) to replace Equation (4). The sequence of novel background values is defined as:

$$z^{(1)'} = \left\{z^{(1)'}(2), z^{(1)'}(3), \cdots, z^{(1)'}(N)\right\} \tag{13}$$

where $z^{(1)'}(k+1) = \frac{N/2+1}{N}x^{(1)}(k) + \frac{N/2-1}{N}x^{(1)}(k+1)$.

Step 3 According to Step 3 and Step 4 of the traditional $GM(1,1)$ model, we can calculate the forecast values of the air pollutant concentration.

Thus, the original sequence and the background value are optimized at the same time in the $IOBGM(1,1)$ model to improve the forecast accuracy.

4. Experiments.

4.1. Datasets. According to the public announcement of the Ministry of Environmental Protection of China issued in January 2015, Handan of Hebei province ranked the 4th in cities where the air quality is relatively poor in China. And relevant researches show that with the action of southwest wind, the air pollutants of Handan easily impact the air quality of Beijing. If atmospheric pollution of Handan could be mitigated, it would help to ease the current situation of air pollution in Beijing. The air pollutants impacting air quality in Handan are mainly PM_{10} and SO_2, so we will forecast the two kinds of the air pollutant concentration based on the annual concentrations of PM_{10} and SO_2 of the environment quality bulletin from 2008 to 2014 issued by Handan Municipal Environmental Protection Bureau.

Table 1. Annual concentrations of PM_{10} and SO_2 of Handan (2008-2014)

	2008	2009	2010	2011	2012	2013	2014
PM_{10} (ug/m^3)	101	102	91	82	92	88	83
SO_2 (ug/m^3)	53	46	45	44	39	40	39

4.2. The forecast results and analysis. The annual concentrations of PM_{10} and SO_2 of Handan from 2008 to 2012 are used as respective the original sequences to respectively establish the $IOBGM(1,1)$ model for forecasting the annual concentrations of PM_{10} and SO_2 of Handan from 2013 to 2014. The real values of PM_{10} and SO_2 concentrations in 2013 and 2014 are used as the accuracy test values of $IOBGM(1,1)$ model. The forecast results of $IOBGM(1,1)$ model are respectively compared with those of the traditional $GM(1,1)$ model and $BGM(1,1)$ model, $GM(1,1)$ model whose the background value is only improved. The forecast results and relative errors of three kinds of air quality forecast models respectively are shown in Table 2 and Table 3.

Table 2 and Table 3 present that when the background value of $GM(1,1)$ model is only improved, the forecast accuracy is improved to limited extent. And simultaneous optimization of the original sequence and the background value significantly improves the forecast accuracy. For PM_{10}, the average relative error of the traditional $GM(1,1)$ model is 6.9% and that of $IOBGM(1,1)$ model is only 1.9%. The forecast accuracy improves by 5.0%, compared with that of the traditional $GM(1,1)$ model. For SO_2, the average relative error of the traditional $GM(1,1)$ model reaches up to 6.0% and that of

TABLE 2. Forecast results and relative error of PM_{10} (ug/m^3)

Year	Real value	Traditional $GM(1,1)$ Forecast value	Traditional $GM(1,1)$ Relative error	$BGM(1,1)$ Forecast value	$BGM(1,1)$ Relative error	$IOBGM(1,1)$ Forecast value	$IOBGM(1,1)$ Relative error
2013	88	81.5	7.4%	82.8	5.9%	86.7	1.5%
2014	83	77.8	6.3%	78.4	5.5%	84.9	2.3%

TABLE 3. Forecast results and relative error of SO_2 (ug/m^3)

Year	Real value	Traditional $GM(1,1)$ Forecast value	Traditional $GM(1,1)$ Relative error	$BGM(1,1)$ Forecast value	$BGM(1,1)$ Relative error	$IOBGM(1,1)$ Forecast value	$IOBGM(1,1)$ Relative error
2013	40	38.0	5.0%	38.2	4.5%	38.5	3.7%
2014	39	36.3	6.9%	36.5	6.4%	37.6	3.6%

$IOBGM(1,1)$ model is 3.7%. So using the $IOBGM(1,1)$ model to forecast the air quality can significantly improve the forecast accuracy.

The $IOBGM(1,1)$ model is applied to forecasting the annual concentrations of PM_{10} and SO_2 of Handan from 2015 to 2020. The annual concentrations of PM_{10} and SO_2 from 2008 to 2014 are used as the original sequence respectively. The new sequences are generated respectively by the novel buffer operator working on the original sequence as follows:

$$x^{(0)}d_{1PM_{10}} = \{91 \quad 92 \quad 87 \quad 82 \quad 87 \quad 85 \quad 83\} \tag{14}$$

$$x^{(0)}d_{1SO_2} = \{45 \quad 42 \quad 42 \quad 41 \quad 39 \quad 39 \quad 39\} \tag{15}$$

The new sequences $x^{(0)}d_{1PM_{10}}$ and $x^{(0)}d_{1SO_2}$ respectively used as input of $IOBGM(1,1)$ model to forecast the annual concentrations of PM_{10} and SO_2 of Handan before 2020.

Figure 2 and Figure 3 show the mean error between fitting values and real values of PM_{10} from 2008 to 2014 is only 1.2% and that the mean error between fitting values and real values of SO_2 is only 2.2%. The forecast results show that the annual concentrations of PM_{10} and SO_2 present the gradual decline in the next few years. In 2020, the concentration of PM_{10} is 75ug/m^3 and the concentration of SO_2 is 34ug/m^3. In comparison with the concentrations of PM_{10} and SO_2 in 2014, the concentrations of PM_{10} and SO_2 in 2020 decline with respective descent rate of 9.6% and 12.8%. However, the concentration of PM_{10} exceeds the secondary annual concentration thresholds of PM_{10} of the ambient

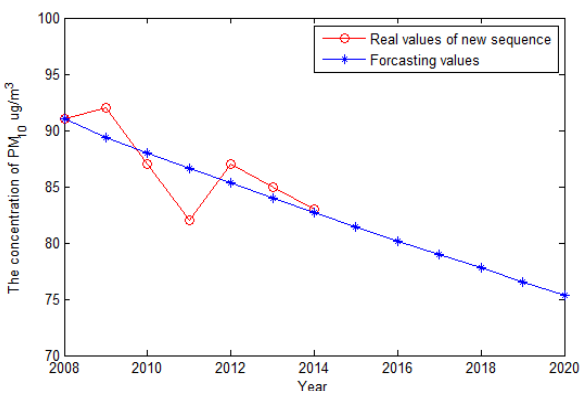

FIGURE 2. Forecast results of PM_{10} concentration

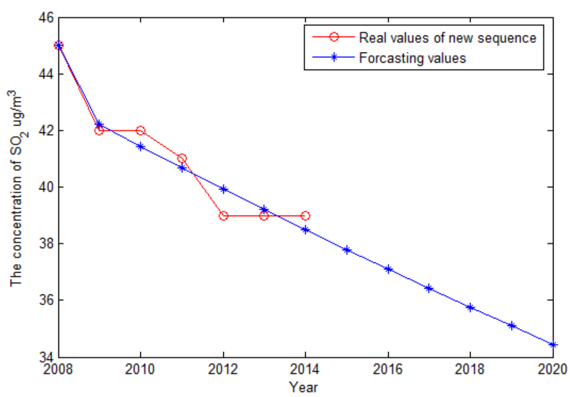

FIGURE 3. Forecast results of SO_2 concentration

air quality standard of China (70ug/m^3), which suggests that although the air quality of Handan has been obviously improved and the prevention and control of atmospheric pollution treatment measures has achieved a certain effect, more governance is required for PM_{10} pollution.

5. **Conclusions.** Based on the simultaneous optimizations of the original sequence and the background value, an improved $GM(1,1)$ model called $IOBGM(1,1)$ model is proposed to forecast urban air quality, and takes Handan as an example for experiments and analysis. Through the comparison of the forecast results of PM_{10} and SO_2 concentrations, the forecast accuracy of $IOBGM(1,1)$ model is much higher than that of traditional $GM(1,1)$ model. With ideal forecast effect, $IOBGM(1,1)$ model meets the requirements for air quality forecast. The air quality of Handan before 2020 is forecasted by $IOBGM(1,1)$ model. The forecast results show that the concentrations of PM_{10} and SO_2 obviously decrease and that the air quality of Handan has been improved. The forecast and analysis provide a scientific basis for the prevention and control of air pollution in Handan and even other cities of China, and have a certain practical significance. In the future, we would like to apply the improved $GM(1,1)$ model to more cities, further improve the forecast accuracy and study the root causes of air pollution.

Acknowledgment. This work is supported by the National Natural Science Foundation of China under Grant No. 61170190 and No. 61472341, and the Natural Science Foundation of Hebei Province P. R. China under Grant No. F2012203062, No. F2013203324 and No. F2014203152. Authors also gratefully acknowledge the helpful comments and suggestions of reviewers, which have improved the presentation.

REFERENCES

[1] Y. Zhang, M. Bocquet, V. Mallet et al., Real-time air quality forecasting, part I: History, techniques, and current status, *Atmospheric Environment*, vol.60, pp.632-655, 2012.
[2] J. Shen, L. Zhong, F. He and D. Chen, Development of air quality forecast model based on clustering and multiple regression, *Environmental Science and Technology*, vol.38, pp.63-66, 2015.
[3] M. S. Baawain and A. S. Al-Serihi, Systematic approach for the prediction of ground-level air pollution using an artificial neural network, *Aerosol and Air Quality Research*, vol.14, pp.124-134, 2014.
[4] X. Hua, H. Tang, H. Zhang and L. Zhang, Application of spatial interpolation method in air quality monitoring based on RBF neural network, *Transducer and Microsystem Technologies*, vol.34, pp.157-160, 2015.
[5] P. Wang, Y. Liu, Z. Qin and G. Zhang, A novel hybrid forecasting model for PM_{10} and SO_2 daily concentrations, *Science of the Total Environment*, vol.505, pp.1202-1212, 2015.
[6] Y. Zheng, X. Yi, M. Li et al., Forecasting fine-grained air quality based on big data, *Proc. of the 21st ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Sydney, Australia, 2015.
[7] J. L. Deng, Control problems of grey systems, *Systems and Control Letters*, vol.1, pp.288-294, 1982.
[8] L. Pan, B. Sun and W. Wang, City air quality forecasting and impact factors analysis based on grey model, *Proc. of the 2011 SREE Conference on Engineering Modeling and Simulation*, Hong Kong, China, 2011.
[9] T.-Y. Pai, C.-L. Ho, S.-W. Chen et al., Using seven types of $GM(1,1)$ model to forecast hourly particulate matter concentration in banciao city of Taiwan, *Water, Air, and Soil Pollution*, vol.217, pp.25-33, 2011.
[10] S. Liu, The three axioms of buffer operator and their application, *Journal of Grey System*, vol.3, pp.39-48, 1991.