ICIC Express Letters
Part B: Applications
Volume **7**, Number **8**, August **2016**

ICIC International ©2016 ISSN 2185-2766
pp. 1739–1745

AN IMPROVED GREY-MARKOV MODEL FOR URBAN AIR QUALITY FORECAST

Yuan Huang, Guoyan Huang, Bing Zhang and Wanchang Jiang

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

Received December 2015; accepted March 2016

Abstract. *Urban air quality forecast plays an important role in the regulatory plans aimed at the control and reduction of air pollutants. In view of the nonlinearity, data insufficiency and data fluctuation of air quality forecast, an improved Grey-Markov model called IGMM is proposed to improve the forecast accuracy of urban air quality. The metabolic GM(1,1) model is utilized to replace traditional GM(1,1) model and simultaneously the original sequence is optimized to reduce data disturbance. The experiment results show that IGMM can be employed in air quality forecast with its less relative error and higher reliability, compared with grey model and traditional Grey-Markov model.*

Keywords: Air quality, GM(1,1) model, Markov chain, Optimizing, Forecast

1. **Introduction.** Air quality forecast helps the public to effectively avoid physical damage caused by air pollution and the environmental protection department to strengthen the supervision of pollution sources. Besides, it has important significance on the improvement of emergency capability in heavy pollution days and of the atmospheric environment.

Air quality forecast widely uses statistical forecast methods and deterministic forecast methods [1]. Although a preference toward deterministic chemical models for air quality forecast is perceived, the forecast accuracy of air pollutant concentrations with them does not seem greater than what can be obtained with statistical models. The trend in recent years has been to use more statistical methods instead of traditional deterministic models for statistical models are much easier, quicker and economical to implement [2]. The statistical forecast models include artificial neural network models and grey models and so on. Perez [3] combined the artificial neural networks and a nearest neighbor method for PM_{10} forecast in Santiago. Feng et al. [4] presented a hybrid model combining air mass trajectory analysis and wavelet transformation to improve the artificial neural network forecast accuracy of air pollution concentrations. Pan et al. [5] employed grey dynamic model group to forecast the air quality changing trend of Tianjin in China.

Limited time and space impose restriction on the data acquisition of air quality forecast which always contains incomplete information. The grey model can solve these problems as a solution to insufficient data, poor information and uncertainty. However, the data sequence fluctuation is often encountered in the air quality forecast. The forecast accuracy of GM(1,1) model of grey theory for data sequence with large random fluctuations is low. If it is used to forecast long-term data, the forecast accuracy will be greatly reduced.

Markov chain forecast model can be used to forecast a system with randomly varying time series. Therefore, Markov chain model can improve the forecast accuracy of GM(1,1) model especially when data fluctuation is large, thereby making up for the deficiency of the grey method. However, Markov forecast model requires time series to have the characteristics of stationary process. However, the urban air quality forecast is a kind of non-stationary random process. We adopt GM(1,1) model to fit time series of urban air quality, to find out the change rule and to compensate for the limitations of Markov process. So the grey model and the Markov model are combined, which can increase the reliability and stability of the forecast results and avoid the limitations of single forecast model. Yang and Sun [6] established a Grey-Markov model to forecast the air pollution concentrations of Pingdingshan in China. However, the combination of grey model with Markov model is rarely used to forecast urban air quality.

The traditional Grey-Markov model (TGMM) can be used to forecast air quality based on the combination of the advantages of the grey model and the Markov model. However, the addition of new disturbance factors in the forecast process is ignored. And the accuracy of the Grey-Markov model is still influenced by the random fluctuations of the data in the modeling process of grey model. The grey metabolic model can overcome the disadvantage of the new disturbance factors into the system, which can reduce the accuracy of forecast. And the weakening buffer operator can weaken the influence of interference in the system. In order to improve the forecast accuracy, an improved Grey-Markov model called IGMM is proposed in this paper, with using the metabolic GM$(1,1)$ model instead of traditional GM$(1,1)$ model and the original sequence of GM$(1,1)$ is optimized to reduce data disturbance simultaneously. The experimental results show that the improved Grey-Markov model has higher forecast accuracy.

The rest of paper is organized as follows. Section 2 introduces the improved grey model. An improved Grey-Markov model is given in Section 3. The experimental results and analysis are given in Section 4. Section 5 concludes the paper.

2. The Improved Grey Model.

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

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

Step 1 $X^{(1)} = x^{(1)}(1), x^{(1)}(2), \cdots, x^{(1)}(N)$ is obtained by 1-AGO (one time accumulated generating operation) based on the original sequence $X^{(0)} = x^{(0)}(1), x^{(0)}(2), \cdots,$ $x^{(0)}(N)$ where $x^{(1)}(k) = \sum_{i=1}^{k} x^{(0)}(i)$, $k = 1, 2, \cdots, N$.

Step 2 Build up constant differential equation:

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

where a and u are respectively called the developing coefficient and the control variable.

Step 3 Let $\hat{U}$ be the parameters vector, and the Ordinary Least Square (OLS) method is utilized to calculate a and u:

$$\hat{U} = [a \quad u]^T = (B^TB)^{-1}B^TY \tag{2}$$

$$B = \begin{bmatrix} -\frac{1}{2}\left(x^{(1)}(2) + x^{(1)}(3)\right) & 1 \\ -\frac{1}{2}\left(x^{(1)}(3) + x^{(1)}(4)\right) & 1 \\ \vdots & \vdots \\ -\frac{1}{2}\left(x^{(1)}(k) + x^{(1)}(k+1)\right) & 1 \end{bmatrix} Y = \left[x_1^{(0)}(2)\ x_1^{(0)}(3) \cdots x_1^{(0)}(N)\right]^T \tag{3}$$

where, $k = 1, 2, \cdots, N-1$.

Step 4 Obtain the discrete form of first-order grey differential equation.

$$\hat{x}^{(1)}(k+1) = \left[x^{(0)}(1) - \frac{u}{a}\right] e^{-ak} + \frac{u}{a} \tag{4}$$

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

Step 5 Apply the Inverse Accumulated Generating Operation (IAGO), and then we have

$$\hat{x}^{(0)}(k+1) = \hat{x}^{(1)}(k+1) - \hat{x}^{(1)}(k) = (1 - e^{a})\left(x^{(0)}(1) - \frac{u}{a}\right) e^{-ak} \tag{5}$$

2.2. **The improvement of GM(1, 1) model.**

2.2.1. *The metabolic GM*$(1,1)$ *model.* The data before time $t = N$ are used as the data sequence in the modeling process of traditional $\mathrm{GM}(1,1)$ model. However, as time goes on, some new random disturbance factors are added into the system, which affect the system. Therefore, the longer the forecast lasts, the lower the forecast accuracy is. However, the metabolic $\mathrm{GM}(1,1)$ model can avoid these defects. Wang et al. [7] established a metabolic $\mathrm{GM}(1,1)$ predicting model of water demand. Moreover the model is applied to predict the water demand of a certain city that is lacking in water in the next 10 years in North China. Zhu et al. [8] used the metabolic $\mathrm{GM}(1,1)$ model to predict the river quality. The precision of forecasting is higher than the traditional $\mathrm{GM}(1,1)$ model. In order to reflect the influence of random disturbance in the future, and to improve the forecast accuracy, the metabolic $\mathrm{GM}(1,1)$ model is used to replace the traditional $\mathrm{GM}(1,1)$ model in this paper.

The latest forecast data $x^{(0)}(N+1)$ is obtained by one time grey forecast. The data $x^{(0)}(N+1)$ is added to the original data sequence, while the oldest data $x^{(0)}(1)$ is removed. The new sequence $\left\{x^{(0)}(2), x^{(0)}(3), \cdots, x^{(0)}(N+1)\right\}$ is used as the original sequence in accordance with the modeling steps of traditional $\mathrm{GM}(1,1)$ model to modeling. Repeat the above data replacement process until the forecast target is completed.

2.2.2. *The improvement of the original sequence.* The data of air quality forecast system is constantly distorted due to the interference of external factors, such as weather and motor vehicle exhaust. The original sequence of $\mathrm{GM}(1,1)$ model will directly affect the accuracy of the forecast system. It is necessary to reduce the interference effect of disturbance factors in order to improve the forecast accuracy of air quality. Therefore, the buffer operator is proposed based on the calculation of the original sequence [9]. The weakening buffer operator can weaken or eliminate the influence of interference in the system and improve the forecast accuracy of air quality.

Definition 2.1. *Assume that the original sequence* $X^{(0)} = \left\{x^{(0)}(1), x^{(0)}(2), \cdots, x^{(0)}(N)\right\}$ *is the behavior data sequence of system, that d is an operator working on* $\{X^{(0)}\}$*, and that the sequence, obtained by having d work on* $\{X^{(0)}\}$*, is denoted as:*

$$X^{(0)}d = \left\{x^{(0)}(1)d, x^{(0)}(2)d, \cdots, x^{(0)}(N)d\right\} \tag{6}$$

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

Theorem 2.1. ***Theorem of Fixed Points*** [9] *Assume 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 2.2. ***Theorem of Sufficient Usage of Information*** [9] *Each datum* $x^{(0)}(k)$ $(k = 1, 2, \cdots, N)$ *in the behavior data sequence of system participates fully in the whole process in which the operator works on* $x^{(0)}$.

Theorem 2.3. ***Theorem of Analytic Representations*** [9] *For any* $x^{(0)}(k)d$ $(k = 1, 2, \cdots, N)$ *can be described with a uniform and elementary analytic representation in* $x^{(0)}(1), x^{(0)}(2), \cdots, x^{(0)}(N)$.

Above axioms are jointly called three axioms of buffer operators. All sequence operators, satisfying the above three axioms, are called buffer operators. And $\left\{X^{(0)}d\right\}$ is called the buffer sequence.

A novel weakening buffer operator is defined as follows:

$$x^{(0)}d_1 = \sqrt{\frac{x^{(0)}(k)x^{(0)}(N) + x^{(0)}(N)^2}{2}} \tag{7}$$

where $k = 1, 2, \cdots, N$, d_1 is a buffer operator.

3. The Improved Grey-Markov Model.

The modeling steps of improved Grey-Markov model are shown as follows.

Step 1 From Section 2, a forecast trend curve equation can be built:

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

Step 2 Divide state intervals. The relative errors are calculated based on the forecast curve and the real data sequence. Relative error range is obtained by subtracting the minimum relative error value from the maximum relative error value. We can equally divide the relative error range into m states according to the relative errors. Any state can be denoted as:

$$E_j = [A_j, B_j], \quad j = 1, 2, \cdots, m \tag{8}$$

where A_j and B_j respectively denote the upper and lower bounds of the j-th state.

Step 3 Calculate the transition probability. For Markov-chain series, the transition probability from state E_i to E_j can be established by using an equation as follows:

$$P_{ij}(k) = \frac{n_{ij}(k)}{n_i}, \quad i, j = 1, 2, \cdots, m \tag{9}$$

where $P_{ij}(k)$ is the transition probability of state E_j transferring from state E_i through k steps, k is the number of transition steps each time, n_i is the number of data in state E_i, $n_{ij}(k)$ is the number of original data of state E_j transferring from state E_i through k steps, and its transition probability matrix can be expressed as follows:

$$P_k = \begin{bmatrix} P_{11}(k) & P_{12}(k) & \cdots & P_{1m}(k) \\ P_{21}(k) & P_{22}(k) & \cdots & P_{2m}(k) \\ \vdots & \vdots & \vdots & \vdots \\ P_{m1}(k) & P_{m2}(k) & \cdots & P_{mm}(k) \end{bmatrix} \tag{10}$$

The transition probability matrix reflects the transition rules of the states in a system, which is the foundation of the Grey-Markov forecast model, and the future trend of the system can be predicted by studying the transition probability matrix.

Step 4 Calculate the forecast data. After the determination of the future state transition of a system, the relative residual error zone $[A_j, B_j]$ is obtained. The forecast value $\hat{y}(j)$ is obtained by the following equation.

$$\hat{y}(j) = \hat{x}^{(0)}(j)\left(1 + \frac{A_j + B_j}{2}\right) \tag{11}$$

The forecast process of the IGMM is presented in Algorithm 1.

Algorithm 1: Improved Grey-Markov Model

Input: The annual concentration of an air pollutant,
$X^{(0)} = \left\{x^{(0)}(1), x^{(0)}(2), \cdots, x^{(0)}(N)\right\}$

Output: The forecast values of the air pollutant concentration,
$Y^{(0)} = \left\{y^{(0)}(1), y^{(0)}(2), \cdots, y^{(0)}(M)\right\}$

1. Do
2. Build the improved GM$(1,1)$ model and generate prediction function
3. Calculate relative errors and divide states
4. Calculate transition probability matrix
5. Calculate the forecast values
6. Return $Y^{(0)} = \left\{y^{(0)}(1), y^{(0)}(2), \cdots, y^{(0)}(M)\right\}$

4. Experiments.

4.1. **Datasets.** According to the statistics from Ministry of Environmental Protection of the People's Republic of China, in the API evaluation system, PM_{10} acts as the primary pollutant in over 99% of pollution days and SO_2 as the primary pollutant in the rest days. The air pollutants impacting air quality in Beijing are mainly PM_{10} and SO_2. Therefore, we respectively forecast the annual concentrations of PM_{10} and SO_2 of Beijing based on the environment quality bulletin from 2005 to 2014 issued by Beijing Municipal Environmental Protection Bureau. The datasets are detailed in Table 1.

Table 1. Annual concentrations of PM_{10} and SO_2 of Beijing (ug/m^3)

	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014
PM_{10}	142	151	148	122	121	121	114	109	108	116
SO_2	50	53	47	36	34	32	28	28	27	22

4.2. **The establishment of IGMM model.** We take PM_{10} as an example to introduce the modeling process of the improved Grey-Markov model, and the modeling process of SO_2 is the same as that of PM_{10}.

Step 1 The annual concentrations of PM_{10} from 2005 to 2012 are used as the original sequence to establish the improved GM$(1,1)$ model. The annual concentrations of PM_{10} from 2013 to 2014 are forecast by the improved Grey-Markov model.

Step 2 The new sequence is generated with the novel weakening buffer operator working on the original sequence as follows: $X^{(0)}d_1 = \{117\ 119\ 118\ 112\ 112\ 112\ 110\ 109\}$.

Step 3 The metabolic GM$(1,1)$ model is used to forecast the new sequence $\{X^{(0)}d_1\}$, and the relative errors between the original values and the forecast values are calculated. Then state intervals are divided based on the relative errors. According to the relative errors of the original values and the forecast values of PM_{10}, three states are divided as follows: $E_1 = (-0.027, -0.012]$, $E_2 = (-0.012, 0.003]$, and $E_3 = (0.003, 0.018]$. The forecast results and the relative errors lying in the states are listed in Table 2.

Table 2. Forecast results and relative errors

Year	Original value	Forecast value	Relative error	State
2005	117	117	0	2
2006	119	118	0.8%	3
2007	118	116	1.7%	3
2008	112	115	−2.6%	1
2009	112	113	−0.9%	2
2010	112	111	0.9%	3
2011	110	110	0	3
2012	109	108	0.9%	3

Step 4 Transition probability matrix can be calculated according to the state partition in Table 2:

$$P = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1/4 & 0 & 3/4 \end{bmatrix}$$

Because 2012 lies in state 3, the initial vector is $V_0 = (0\ 0\ 1)$ and $V_1 = V_0 \cdot P = (1/4\ 0\ 3/4)$. Therefore, the 2013 lies in state 3. The forecast value of 2013 obtained by improved GM(1,1) is 106 ug/m^3, so the forecast value obtained by the improved Grey-Markov is 108 ug/m^3, that is $106 + (1 + 0.5 \times (0.003 + 0.018)) = 108$.

Step 5 As the new element, the forecast concentration of PM_{10} in 2013 is added in the original sequence to replace the concentration of PM_{10} in 2005. That is, the annual concentrations of PM_{10} from 2006 to 2013 are used as the original sequence to establish the improved GM$(1,1)$ model to forecast the concentration of PM_{10} in 2014. Repeating Step 2, Step 3 and Step 4, we can get the concentration of PM_{10} in 2014 is 109 ug/m^3.

4.3. **The forecast results and analysis.** The forecast results of Improved Grey-Markov model (IGMM) are respectively compared with those of Traditional Grey-Markov model (TGMM) and GM$(1,1)$ model. The forecast results and relative errors of the three kinds of air quality forecast models are respectively shown in Table 3 and Table 4.

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

Year	PM_{10}						
	Real value	GM$(1,1)$		TGMM		IGMM	
		Forecast value	Relative error	Forecast value	Relative error	Forecast value	Relative error
2013	108	102	5.6%	105	2.8%	108	0%
2014	116	97	16.4%	99	14.7%	110	5.2%

TABLE 4. Forecast results and relative error of SO_2 (ug/m^3)

Year	SO_2						
	Real value	GM$(1,1)$		TGMM		IGMM	
		Forecast value	Relative error	Forecast value	Relative error	Forecast value	Relative error
2013	27	22	18.5%	24	11.1%	27	0%
2014	22	20	9.1%	21	4.5%	24	9.1%

Table 3 and Table 4 show that although the relative error of the IGMM is slightly higher than that of the TGMM as for the forecast concentration of SO_2 in 2014. The forecast accuracy of the Improved Grey-Markov model (IGMM) as the whole is significantly improved compared with that of GM$(1,1)$ model and the Traditional Grey-Markov model (TGMM). For PM_{10}, the average relative error of the TGMM model is 8.8% and that of IGMM model is only 2.6%. For SO_2, the average relative error of the TGMM model is 7.8% and that of IGMM model is 4.6%. The forecast concentrations of the two pollutants in 2013 are the same as the real values, and the relative error is 0%.

So using the IGMM model to forecast the air quality can significantly improve the forecast accuracy. The improved Grey-Markov model has higher prediction accuracy, which makes it suitable for the urban air quality forecast. The IGMM is adopted to forecast the annual concentrations of PM_{10} and SO_2 of Beijing from 2015 to 2019.

Figure 1 shows the annual concentrations of PM_{10} and SO_2 are gradually decreasing in the next few years. In 2019, the concentration of PM_{10} is 109 ug/m^3 and that of SO_2 is 19 ug/m^3. In comparison with the concentrations of PM_{10} and SO_2 in 2014, the concentrations of PM_{10} and SO_2 in 2019 decline with respective descent rate of 6.4% and 13.6%.

Although Beijing has gained some achievements on the air pollution control and the concentration of pollutants has declined, the governance of the main pollutant PM_{10} is still far from the standard of environmental protection. In 2019, the concentration of PM_{10} is still 1.6 times more than the secondary annual concentration thresholds of PM_{10} of the ambient air quality standard of China (70 ug/m^3), which suggests that more governance is required for PM_{10} pollution.

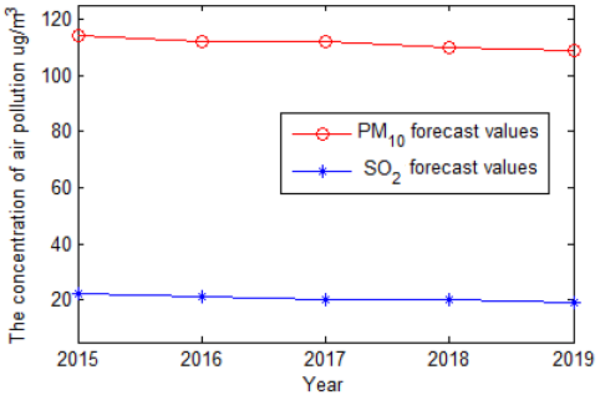

FIGURE 1. Forecast results of the air pollutant concentrations

5. **Conclusions.** An improved Grey-Markov model is proposed to forecast urban air quality in this paper. The improved Grey-Markov model avoids the limitations of single forecast model and effectively improves the forecast accuracy. We take Beijing 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 IGMM is much higher than that of GM$(1,1)$ model and traditional Grey-Markov model. With ideal forecast effect, the improved Grey-Markov model meets the requirements for air quality forecast. The forecast and analysis provide a scientific basis for the prevention and control of air pollution in Beijing and even other cities in China, and have a certain practical significance. In the future, we would like to apply the improved Grey-Markov 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] H. J. S. Fernando, M. C. Mammarella, C. Grandoni et al., Forecasting PM_{10} in metropolitan areas: Efficacy of neural networks, *Environmental Pollution*, vol.163, pp.62-67, 2012.
[3] P. Perez, Combined model for PM_{10} forecasting in a large city, *Atmospheric Environment*, vol.60, pp.271-276, 2012.
[4] X. Feng, Q. Li, Y. Zhu et al., Artificial neural networks forecasting of PM2.5 pollution using air mass trajectory based geographic model and wavelet transformation, *Atmospheric Environment*, vol.107, pp.118-128, 2015.
[5] L. Pan, B. Sun and W. Wang, City air quality forecasting and impact factors analysis based on grey model, *Proc. of the SREE Conference on Engineering Modeling and Simulation*, Hongkong, China, 2011.
[6] J. Yang and B. Sun, The prediction of air pollutant concentration in Pingdingshan based on grey Markov model, *Mathematics in Practice and Theory*, vol.44, pp.64-70, 2014.
[7] H. Wang, F. Ma, K. Yang and B. Lu, A study of information renewal GM$(1,1)$ model for predicting urban medium and long-term water demand, *Engineering Journal of Wuhan University*, vol.37, pp.32-35, 2004.
[8] C. Zhu, J. Lu, Z. Hao et al., Application of information renewal GM$(1,1)$ model in the prediction of water quality in the river, *Journal of Hebei University of Engineering (Natural Science Edition)*, vol.24, pp.45-47, 2007.
[9] S. Liu, The three axioms of buffer operator and their application, *Journal of Grey System*, vol.3, pp.39-48, 1991.