Using grey Holt-Winters model to predict the air quality index for cities in China

Lifeng Wu[a], Xiaohui Gao[a], Yanli Xiao*[a], Sifeng Liu[b], Yingjie Yang[b]

[a] *College of Economics and Management, Hebei University of Engineering, Handan 056038, China*

[b] *Centre for Computational Intelligence, De Montfort University, Leicester, LE1 9BH, UK*

Abstract:

The randomness, non-stationarity and irregularity of air quality index series bring the difficulty of air quality index forecasting. To enhance forecast accuracy, a novel model combining grey accumulated generating technique and Holt-Winters method is developed for air quality index forecasting in this paper. The grey accumulated generating technique is utilized to handle non-stationarity of random and irregular data series and Holt-Winters method is employed to deal with the seasonal effects. To verify and validate the proposed model, two monthly air quality index series from January in 2014 to December in 2016 collected from Shijiazhuang and Handan in China are taken as the test cases. The experimental results show that the proposed model is remarkably superior to conventional Holt-Winters method for its higher forecast accuracy.

Keywords: Air quality index forecasting; Holt-Winters method; grey accumulated generating technique; Handan; Shijiazhuang

1. Introduction

Air pollution is a serious problem in many parts of the world. In China, many cities have been adversely affected by hazy weather since the beginning of 2013. Air quality index (AQI) is used to measure the air quality and the greater value is, the more serious pollution it is. AQI is based on the concentrations of six airborne pollutants-PM2.5 and PM10, ozone, sulfur dioxide,

[0]*Corresponding Author: E-mail address: xylfwb@163.com.

nitrogen dioxide and carbon monoxide. The AQI continues rising. Repeatedly, moderate, severe and extremely severe hazy weather appeared in the North China, Hebei Province. Measurements in Shijiazhuang (the capital of Hebei Province) and Handan (a city of Hebei Province) often exceed 500 and hit hazardous levels. Around 300 is considered hazardous. Thus, how to forecast the AQI precisely plays an important role in both controlling air pollution and promoting the regional sustainable development [1]. In view of existing situation, the prediction methods of AQI can be divided into the following five categories: principal component regression technique [2], ordinal time series model [3], grey model [4], adaptive fuzzy model [5] and hybrid models [1, 6, 7].

However, the deficiencies still exist in the methods mentioned above. These models often cannot thoroughly handle non-stationarity of random and irregular data series, and did not take seasons variation this into consideration, although different seasons and regions all come with their own weather conditions [8]. Thus, the grey accumulated generating technique is introduced to conduct the non-stationary characteristics associated with AQI series in order to improve the forecast accuracy, and the Holt-Winters method is developed to deal with the seasonal effects in this paper.

The rest of this paper is organized as follows. A novel grey Holt-Winters model is put forward in Section 2. The AQI of Handan and Shijiazhuang are predicted respectively in Section 3. The conclusion is presented in Section 4.

2. Grey Holt-Winters model

For the original time series $X^{(0)} = \{x^{(0)}(1), x^{(0)}(2), \cdots, x^{(0)}(n)\}$, the classical Holt-Winters method follows the equations:

$$S_t = \alpha \frac{x^{(0)}(t)}{I_{t-L}} + (1-\alpha)(S_{t-1} + b_{t-1}), 0 < \alpha < 1$$

$$b_t = \gamma(S_t - S_{t-1}) + (1-\gamma)b_{t-1}, 0 < \gamma < 1$$

$$I_t = \beta \frac{x^{(0)}(t)}{S_t} + (1-\beta) I_{t-L}, 0 < \beta < 1$$

Where L means the length of seasonality, such as the months of one year,the days of one week. I_t is the correction coefficient of seasonality and b_t denotes the trend. α, γ and β are weighting coefficients, which varies from 0 to 1. Then, the initial values are given by:

$$I_L = \frac{\overline{x^{(0)}(L)}}{\overline{x^{(0)}(t)}}, S_{L+1} = x^{(0)}(L+1),$$

$$b_{L+1} = \frac{x^{(0)}(L+1) - x^{(0)}(1) + x^{(0)}(L+2) - x^{(0)}(2) + x^{(0)}(L+3) - x^{(0)}(3)}{3L}.$$

Where $\overline{x^{(0)}(L)}$ is the average value of the same quarter in different years and $\overline{x^{(0)}(t)}$ is the average of overall actual value.

Generally speaking, the Holt-Winters smoothing method can to deal with univariate time sequence that includes both trend and seasonal factors [9, 10, 11]. The method obtain its popularity since its simple model formulation and accurate predicting results. However, the classical additive and multiplicative Holt-Winters methods can become unreliable if the noise (non-stationarity of random and irregular) dominates the trend and seasonal components of the data [12]. To smooth the randomness of original data, the following definition is given.

Definition 1 [13] For the original time series $X^{(0)} = \{x^{(0)}(1), x^{(0)}(2), \cdots, x^{(0)}(n)\}$, the $r(r \in R_+)$-order accumulated generating sequence $X^{(r)} = \{x^{(r)}(1), x^{(r)}(2), \cdots, x^{(r)}(n)\}$ can be generated by r-order accumulated generating operator (r-AGO) as follows:

$$x^{(r)}(k) = \sum_{i=1}^{k} \binom{k-i+r-1}{k-i} x^{(0)}(i); k = 1, 2, \cdots, n.$$

Set $\binom{r-1}{0} = 1, \binom{k-1}{k} = 0, \binom{k-i+r-1}{k-i} = \frac{(r+k-i-1)(r+k-i-2)\cdots(r+1)r}{(k-i)!}$. r-order inverse accumulated generating operator (IAGO) of $X^{(r)}$ is expressed as follows:

$$X^{(-r)} =^{(\lceil r \rceil)} X^{(\lceil r \rceil - r)} = \{^{(\lceil r \rceil)} x^{(\lceil r \rceil - r)}(1), ^{(\lceil r \rceil)} x^{(\lceil r \rceil - r)}(2), \cdots, ^{(\lceil r \rceil)} x^{(\lceil r \rceil - r)}(n)\}$$

where $\lceil r \rceil = \min\{n \in Z | r \leq n\}$,$^{(\lceil r \rceil)} x^{(\lceil r \rceil - r)}(k) =^{(\lceil r-1 \rceil)} x^{(\lceil r \rceil - r)}(k) -^{(\lceil r-1 \rceil)} x^{(\lceil r \rceil - r)}(k-1)$. The IAGO is the inverse operation of AGO. In general, when $0 < r < 1$, $x^{(1)}(k) = \sum\limits_{i=1}^{k} x^{(0)}(i)(k =$

$1, 2, \cdots, n$). The r-IAGO of $X^{(r)}$ is computed as follows:

$$X^{(-r)} = {}^{(1)}X^{(1-r)} = \{{}^{(1)}x^{(1-r)}(1), {}^{(1)}x^{(1-r)}(2), \cdots, {}^{(1)}x^{(1-r)}(n)\}$$

where ${}^{(1)}x^{(1-r)}(k) = x^{(1-r)}(k+1) - x^{(1-r)}(k)$.

AGO is widely used in grey models for its ability to smooth the randomness of original data [13-15]. By means of IAGO, the prediction value can be transformed back to the original sequence. Through AGO, the disorderly data may be converted into regular trend form. Then the forecasting of the sequence with trend and seasonality can use Holt-Winters smoothing method. Thus we give the following definition:

Definition 2 For the original time series $X^{(0)} = \{x^{(0)}(1), x^{(0)}(2), \cdots, x^{(0)}(n)\}$, r-AGO is given in Definition 1. Grey Holt-Winters model (GHW) follows the equations

$$S_t = \alpha \frac{x^{(r)}(t)}{I_{t-L}} + (1-\alpha)(S_{t-1} + b_{t-1}), 0 < \alpha < 1$$

$$b_t = \gamma(S_t - S_{t-1}) + (1-\gamma)b_{t-1}, 0 < \gamma < 1$$

$$I_t = \beta \frac{x^{(r)}(t)}{S_t} + (1-\beta)I_{t-L}, 0 < \beta < 1$$

Its initial values are given by:

$$I_L = \frac{\overline{x^{(r)}(L)}}{\overline{x^{(r)}(t)}}, S_{L+1} = x^{(r)}(L+1),$$

$$b_{L+1} = \frac{x^{(r)}(L+1) - x^{(r)}(1) + x^{(r)}(L+2) - x^{(r)}(2) + x^{(r)}(L+3) - x^{(r)}(3)}{3L}.$$

Where $\overline{x^{(r)}(L)}$ is the average value of the same quarter in different years and $\overline{x^{(r)}(t)}$ is the average of overall actual value. The forecasting form is $F^{(r)}(t+m) = (S_t + mb_t)I_{t+m-L}$.

If $r = 0$, GHW is the conventional Holt-Winters model. The flow chart of GHW model is given in Fig.1. The process of calculating GHW is as follows:

[Fig.1. Flow chart of GHW model]

Step 1: Set the order number r and obtain the r-AGO sequence of $X^{(0)}$ according to Definition 1;

Step 2: Compute the parameters (S_t, b_t, I_t) by using Definition 2;

Step 3: Obtain the predictive value by using the equation $F^{(r)}(t+m) = (S_t + mb_t)I_{t+m-L}$;

Step 4: Transform the prediction value $F^{(r)}(t+m)$ back to the original sequence $\hat{X}^{(0)}$ by means of IAGO.

In this paper, the predictive values are calculated by using different α, γ, β. The α, γ, β that produce a small mean square error for the fitted values and shows an expected future growth are chosen.

3. Experiments and analysis

To evaluate the performance of GHW prediction, two cities of China are employed, that is, Shijiazhuang and Handan. In 2015, Handan is ranked as the top 7th polluted city and Shijiazhuang is ranked as the top 8th polluted city in China. The simple map of the study area is shown in Fig.2. The AQI data was from Ministry of Environment Protection of the People's Republic of China for 2014-2016. which can be found on the internet at: http://www.zhb.gov.cn/. The monthly AQI data series (the average of every month) from January in 2014 to November in 2016 in Handan are collected. Mean absolute percentage error (MAPE $= \frac{1}{n}\sum\limits_{k=1}^{n}|\frac{x^{(0)}(k)-\hat{x}^{(0)}(k)}{x^{(0)}(k)}| \times 100\%$) compares the forecasted values with the actual values to evaluate the precision. All the considered forecasting models are performed in Excel2003.

[Insert Fig.2 about here]

Case in Handan

Handan is located in the intersection area of Hebei, Shanxi, Henan, and Shandong. Handan, the third largest city of Hebei province with an area of 1.2×10^5 km^2 and population over 10 million, lies in the east of Taihang Mountain. The iron and steel production plays a key role in the local economy. Handan has a typical monsoon climate with dry and windy spring, moist and hot summers, and dry and cold winters. From February to March, sandstorms are common in

Handan. From April to October, the atmospheric dusts are mainly derived from building dusts. From September to December, the pollution from directly burning of straw in countryside have a much greater impact on the AQI. During winter and spring, Handan needs a huge quantity of coal to provide energy for heating, consequently, producing a large amount of SO_2, CO, NOx and other air pollutants pumped into the atmosphere. Therefore, the hazy weather has obvious seasonal characteristic. GHW model is suitable for the AQI forecasting in Handan.

In order to examine the current situation of AQI and to predict the future trend, we use the monthly AQI data series from January in 2014 to September in 2016 as training data. The AQI values from October to November in 2016 are taken as the testing samples and are kept to verify the prediction accuracy. The results are reported in Table 1 and Fig.3.

[Insert Table 1 about here]

[Insert Fig.3 about here]

Table 1 reflects that the fitness values of GHW approximate actual value more than that of conventional method. Therefore, GHW yields the lower MAPE compared with the conventional method. So it is suitable to forecast the AQI with distinct seasonality features. The GHW forecasting values of next four months are reported in Table 2.

[Insert Table 2 about here]

Both in-sample and out-of-sample prediction performance results show that the GHW enjoys higher accuracy. The short term forecasting results are considered reasonable.

Case in Shijiazhuang

Shijiazhuang is located at the transition zone of the east slope of the Taihang Mountain and Heibei plain and lies at southwest of Beijing. It is the capital of Hebei province. Because of the special geographical climate conditions and rapid development of this city, Shijiazhuang is a serious polluted city. On December 3 and 4, 2016, Shijiazhuang saw the AQI reach the most hazardous level. In Shijiazhuang on December 19, 2016, the concentration of PM2.5 exceeded 1000 micrograms per cubic meter at 1 pm and by 6 pm hovered around 900 at some measuring stations, according to the National Environmental Monitoring Center website. As the most polluted areas, it is extremely urgent to solve the haze of air pollution. Therefore, AQI forecasting is significance so that some appropriate measures can be taken to reduce the impact of haze on lives of citizens. Shijiazhuang has a typical temperate and monsoonal climate with four clearly distinct seasons. It is very hot in summer and there is little rain, while the winter is frigid and dry. GHW model is also suitable for the AQI forecasting in Shijiazhuang.

To examine the current situation of AQI and to predict the future trend, we use the monthly AQI data series from January in 2014 to November in 2016 as training data. The AQI value of December in 2016 is taken as the testing sample and is kept to verify the prediction accuracy. The results are reported in Table 3 and Fig.4.

[Insert Table 3 about here]

[Insert Fig.4 about here]

Table 3 reflects that GHW yields the lower MAPE compared with the conventional method. So it is suitable to forecast the AQI with distinct seasonality features. The GHW forecasting values of next four months are reported in Table 4.

[Insert Table 4 about here]

Both in-sample and out-of-sample prediction performance results show that the GHW enjoys higher accuracy. The short term forecasting results are also acceptable. It can be seen from Table 4 that the air pollution will inevitably return to hazardous levels in Shijiazhuang. The government should also step up more efforts to improve air quality.

4. Conclusion

It is necessary and important to seek an effective and reliable method for AQI forecasting. In this paper, the experimental results indicate that the proposed method can consider the seasonal effects and obtain more accurate forecasting results. This study found that by using the fraction order accumulation can eliminate the impact of irregular data and improve AQI forecasts. Two real cases was seen to reach a good order and the coefficients. However, the order and the coefficients may be not the most optimal. More computational experiments will be necessary to make conclusions on the performance of different orders and the coefficients. Future research will discuss the method which can obtain ideal order number and the coefficients. GHW can be applied to the AQI forecasting in other region in the future, to further confirm its effectiveness.

Acknowledgements

The relevant researches done in this paper are supported by the National Natural Science Foundation of China (No.71401051), the Humanistic and Social Science Foundation of Education Ministry (No.15YJA630017), the science and technology project of science and technology department in Henan province (No.172102210257), and the project of high-level talent in Hebei province.

References

[1] Deyun Wang, ShuaiWei, Hongyuan Luo, et al. A novel hybrid model for air quality forecasting based on two-phase decomposition technique and modified extreme learning machine. Science of the Total Environment 2017, 580:719-733.

[2] Anikender Kumar, Pramila Goyal. Forecasting of air quality in Delhi using principal component regression technique. Atmospheric Pollution Research 2011(2):436-444.

[3] Sung Eun Kim. Ordinal time series model for forecasting air quality index for ozone in Southern California. Environmental Modeling & Assessment 2017, 22(2): 175-182.

[4] Cheng-Chung Lee, Terng-Jou Wan, Chao-Yin Kuo, et al. Modified grey model for estimating traffic tunnel air quality. Environ Monit Assess (2007) 132:351-364.

[5] Yiliao Song, Shanshan Qin, Jiansheng Qu, et al. The forecasting research of early warning systems for atmospheric pollutants: a case in Yangtze River Delta region. Atmospheric Environment 118 (2015):58-69.

[6] ANIKENDER KUMAR, P. GOYAL. Forecasting of air quality index in Delhi using neural network based on principal component analysis. Pure Appl. Geophys. 2013(170):711-722.

[7] PauloS. G. deMattosNeto, Francisco Madeiro, Tiago A.E.Ferreira, et al. Hybrid intelligent system for air quality forecasting using phase adjustment. Engineering Applications of Artificial Intelligence 32(2014): 185-191.

[8] Maria Ikram, Zhijun Yan, Yan Liu, et al. Seasonal effects of temperature flctuations on air quality and respiratory disease: a study in Beijing. Nat Hazards 79(2015):833-853.

[9] G. Sudheer, A. Suseelatha. Short term load forecasting using wavelet transform combined with Holt-Winters and weighted nearest neighbor models. Electrical Power and Energy Systems 64 (2015):340-346.

[10] Tiago Mendes Dantas, Fernando Luiz Cyrino Oliveira, Hugo Miguel Varela Repolho. Air transportation demand forecast through Bagging Holt Winters methods, Journal of Air Transport Management, 2017(59):116-123.

[11] Marina Theodosiou. Forecasting monthly and quarterly time series using STL decomposition. International Journal of Forecasting, 2011, 27(4):1178-1195.

[12] Liljana Ferbar Tratar, Blaz Mojskerc, Ales Toman. Demand forecasting with four-parameter exponential smoothing. International Journal of Production Economics, 2016,181:162-173.

[13] Lifeng Wu, Sifeng Liu, Yang Yingjie. Grey double exponential smoothing model and its application on pig price forecasting in China. Applied Soft Computing, 2016, 39:117-123.

[14] Guo-Dong Li, Shiro Masuda, Masatake Nagai. The prediction for Japan's domestic and overseas automobile production. Technological Forecasting and Social Change 2014,87:224-231.

[15] Che-Jung Chang, Der-Chiang Li, Chien-Chih Chen, et al. A forecasting model for small non-equigap data sets considering data weights and occurrence possibilities. Computers & Industrial Engineering 2014,67:139-145.