Vol.3.Issue.4.2015 (Oct-Dec.)

BULLETIN OF MATHEMATICS AND STATISTICS RESEARCH

A Peer Reviewed International Research Journal

http://www.bomsr.com

ISSN:2348-0580

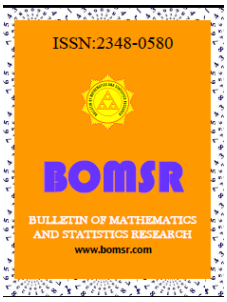
RESEARCH ARTICLE

THE APPLICATION OF HOLT EXPONENTIAL SMOOTHING MODEL ON 2014 AIR QUALITY INDEX IN MIYUN COUNTY, BEIJING, CHINA

RENHAO JIN[1], TAO LIU[1], SHA WANG[1]

[1]School of Information, Beijing Wuzi University, Beijing, China

Email :Renhao.jin@outlook.com (Renhao Jin)



ABSTRACT

In order to study the changes of air quality index (AQI) in Miyun County, Beijing, China and predict the trend of AQI value, this paper constructed a time-series analysis. The non-stationary trend is found in the AQI data and it is difficult to match the trends in the data to any model of linear, polynomial and exponential models, so the Holt exponential smoothing model is chosen. The modeling results shows the Holt model sufficiently model the data, thus a short trend of AQI value can be predicted using the established model.

Keywords : Air Quality Index (AQI); prediction; Holt Exponential Smoothing Model.

1. INTRODUCTION

Beijing is the capital of China and one of the most populous cities in the world. Its population in 2013 was 21.15 million. The city proper is the 3rd largest in the world. The metropolis, located in northern China, is governed as a direct-controlled municipality under the national government, with 14 urban and suburban districts and two rural counties. It is home to the headquarters of most of China's largest state-owned companies and many large multinational companies, and is a major hub for the national highway, expressway, railway, and high-speed rail networks. As China economic is boosting over 20 years, Beijing is always an attraction in the world. However, in recent 2-3 years Beijing is air pollution problem is often in the headlines of many news articles. China government has noticed this problem and done a lot of measures to control the air pollution in Beijing. In this paper, the air quality index (AQI) is used as a comprehensive figure to measure the air quality. As the AQI increases, an increasingly large percentage of the population is likely to experience increasingly severe adverse health effects [1]. Different countries have their own air quality indices, corresponding to different national air quality standards. This paper only concerns the AQI defined by China government [2].The reasonable analysis and forecast of AQI can help the government make and check their air control police and let the hospitals to prepare their daily patient service.

China's Ministry of Environmental Protection (MEP) is responsible for measuring the level of air pollution in China. The AQI level is based on the level of 6 atmospheric pollutants, namely sulfur dioxide (SO2), nitrogen dioxide (NO2), suspended particulates smaller than 10 μm in aerodynamic diameter (PM10), suspended particulates smaller than 2.5 μm in aerodynamic diameter (PM2.5), carbon monoxide (CO), and ozone (O3) measured at the monitoring stations in China [2]. Table 1 displays the AQI value and its corresponding level and health implications. As shown in Table 1, when AQI value is less than 100, the air is no effect for daily life, but when AQI is larger than 200, it can may case heavy adverse health effects.

In this paper, the study area is in Miyun County of Beijing, which is situated in northeast Beijing and has an area of 2,227 square kilometers and a population of half million. The Miyun County is famous tourism place is Beijing with SimaTai Great wall and large Miyun reservoir supplying water for the whole Beijing. It is chosen to be study area as its tourism industry is highly determined by its air quality. As shown in Table 1, health implication of AQI is mainly related to outdoor activities. There is one air quality monitor to examine the air pollution and it publish the AQI value every day. The data is extracted from their everyday report from Jan. 1st 2014 to Dec. 29th 2014.

A lot of methods have been used to analysis and forecast of time series data, such as autoregressive model, autoregressive moving average model, autoregressive conditional heteroscedasticity model, autoregressive integrated moving average model (ARIMA),Holt Exponential Smoothing Model and so on [3]. In the balance of predict and explanation, Holt exponential smoothing model is a wildly used model. In this paper, the AQI data in Miyun County of 2014 is fitted by the Holt exponential smoothing model [4]. All computations are done by using SAS software (SAS® 9.4, SAS Institute Inc., Cary, N.C.) [5].

Table 1. AQI and Health Implications by China's Ministry of Environmental Protection.

AQI	**Air Pollution Level**	**Health Implications**
0–50	Excellent	No health implications
51–100	Good	Few hypersensitive individuals should reduce outdoor exercise
101–150	Lightly Polluted	Slight irritations may occur, individuals with breathing or heart problems should reduce outdoor exercise
151–200	Moderately Polluted	
201–300	Heavily Polluted	Healthy people will be noticeably affected. People with breathing or heart problems will experience reduced endurance in activities. These individuals and elders should remain indoors and restrict activities
300+	Severely Polluted	Healthy people will experience reduced endurance in activities. There may be strong irritations and symptoms and may trigger other illnesses. Elders and the sick should remain indoors and avoid exercise. Healthy individuals should avoid outdoor activities

2. Modeling

2.1. Description of the data

A timing diagram is firstly plot using all the AQI data of 2014 in Miyun County, Beijing. As shown in Figure 1, the AQI values range from 31 to 427 with the annual mean value 118. AQI values peak at spring and winter season, and for the other period of 2014 the AQI seems stationary. It is reasonable to have large AQI values in spring and winter months, as the temperature is relative low in Beijing at that time, ranging from -10℃ to 5℃ and it often leads tofog and haze weather in low temperature. The number of days for every AQI Pollution level in Miyun County, Beijing in 2014 are shown in Table 2, and 52.89% of days in 2014 are in Good or Excellent Air level. However, 12.39% days of 2014 in Miyun are in Heavily Polluted or Severely Polluted. So, in general the air condition in Miyun County is acceptable and suitable for the tourism industry.

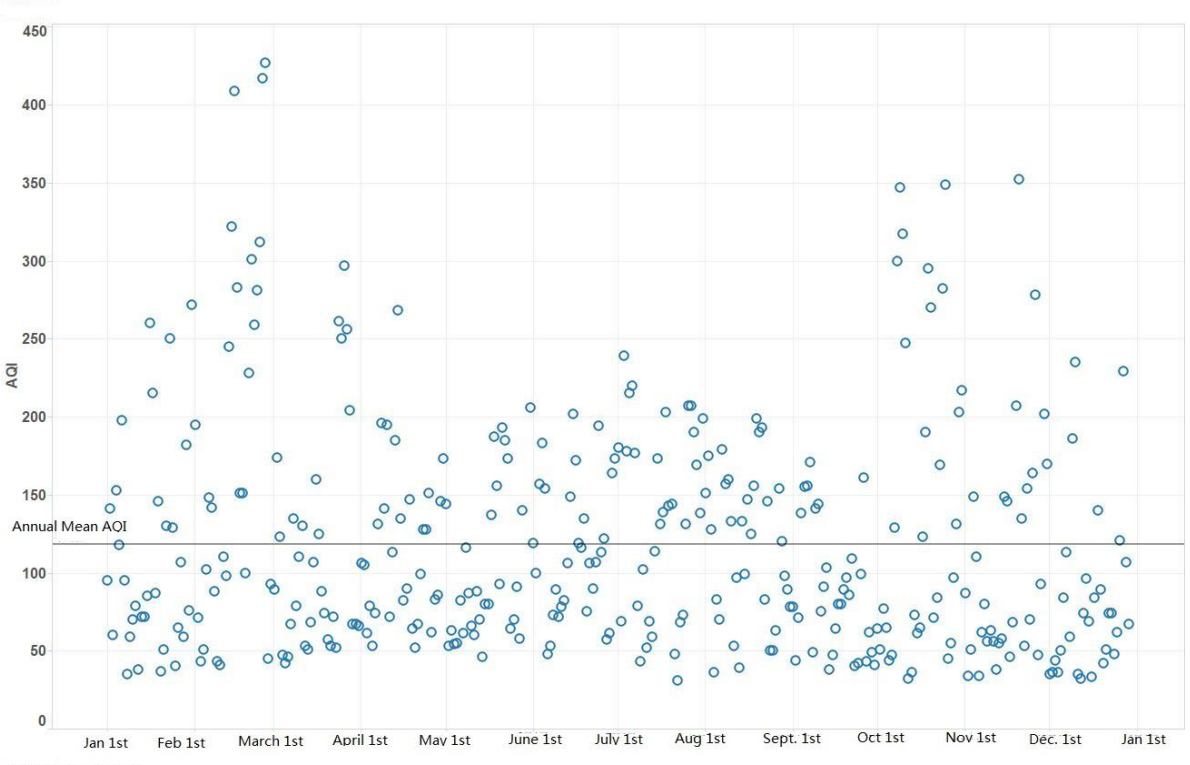


Figure 1. Thetiming diagram plot of AQI value of 2014 in Miyun County, Beijing.

Table 2. The number of days for every AQI Pollution level in Miyun County, Beijing in 2014.

Air Pollution Level	# Days	Percentage	Cumulative Percentage
Excellent	48	13.22	13.22
Good	144	39.67	52.89
Lightly Polluted	74	20.39	73.28
Moderately Polluted	52	14.33	87.60
Heavily Polluted	35	9.64	97.25
Severely Polluted	10	2.75	100.00

As shown in Figure 1, the two peaks on the two sides of the plot break the hypotheses of weaker stationary. It also has no linear trend in diagram and very difficult to match the trend in the Figure to any curve model such as polynomial models and exponential models, thus an exponential smoothing model could be fitted to the data [6, 7]. Because the data is only in one year, i.e., from

Jan. 1st 2014 to Dec. 29th 2014, it cannot be fitted with seasonal model. In the exponential smoothing models, Holt exponential smoothing model is a prefer model which is generally used [8].

2.2. Holt exponential smoothing modeling

Simple exponential smoothing does not do well when there is a trend in the data, which is inconvenient. In such situations, several methods were devised under the name "double exponential smoothing" or "second-order exponential smoothing", which is the recursive application of an exponential filter twice, thus being termed "double exponential smoothing". This nomenclature is similar to quadruple exponential smoothing, which also references its recursion depth. The basic idea behind double exponential smoothing is to introduce a term to take into account the possibility of a series exhibiting some form of trend. This slope component is itself updated via exponential smoothing. Holt exponential smoothing method is the most popular double exponential smoothing method, proposed by Holt (1957) with extending simple exponential smoothing to allow forecasting of data with a trend [4, 9].

Holt method is to concentrate on the series of increments $X_t - X_{t-1}$, and then estimate the slop parameter to a linear trend by exponential smoothing of these differences. In a series with a linear trend, this should equal the slope of the trend with some added noise specific for the situation at the time index t. The trend slop, which is allowed to be time varying, is denoted b_t. The idea is basically to update the true level using the present observation X_t from the previous level $\tilde{X}_{t-1}$ to $\tilde{X}_t$ by an adjustment to the previous slope element b_{t-1} using exponential smoothing. Moreover, the basic formula for exponential smoothing is applied to update from the estimate of b_{t-1} to an estimate of actual b_t as anaverage of last slop element b_{t-1} and the present observed increment$\tilde{X}_t - \tilde{X}_{t-1}$ of the estimated true level. Expressed as formulas, these two updating equations then become

$$\tilde{X}_t = \alpha X_t + (1-\alpha)(\tilde{X}_{t-1} + b_{t-1})$$

and

$$b_t = \gamma\left(\tilde{X}_t - \tilde{X}_{t-1}\right) + (1-\gamma)b_{t-1}.$$

The formula for prediction is

$\hat{X}_{T+i} = \hat{X}_T + \mathrm{i}b_T$.

In this formulation, two weighting parameters (α and γ) are used for the two updating equations. The least squared method is often used to estimate parameters of Holt method [4, 10].

In the model fitting of the Holt exponential smoothing method, the initial parameters of $\tilde{X}_0$ and b_0 need to be determined. In this study, the initial value of the smoothing series $\tilde{X}_0$ is set to be X_1, i.e.,$\tilde{X}_0 = X_1$. The initial value of the trend series b_0 can be defined by many ways, and a simple method is defined that for an arbitrarilyn,

$b_0 = (\mathrm{X}_{\mathrm{n}+1} - X_1)/\mathrm{n}$ [4, 11].

The smooting curved line of Holt method is shown in Figure 2. The two parameters of Holt modeling are estimated as $\alpha = 0.11$ and$\gamma = 0.25$.It can be seen from the plot that the fitted line well matches the AQI changing scatterplot and the fluctuation in the data is expressed by the fitted model. Thus, it is reasonable to use the fitted Holt smoothing method to predict future values, with the equation$\hat{X}_{T+i} = \hat{X}_T + \mathrm{i}b_T$. It is noted that Holt smooting model only can be used to predict future short steps, just like all the other time series methods [12].

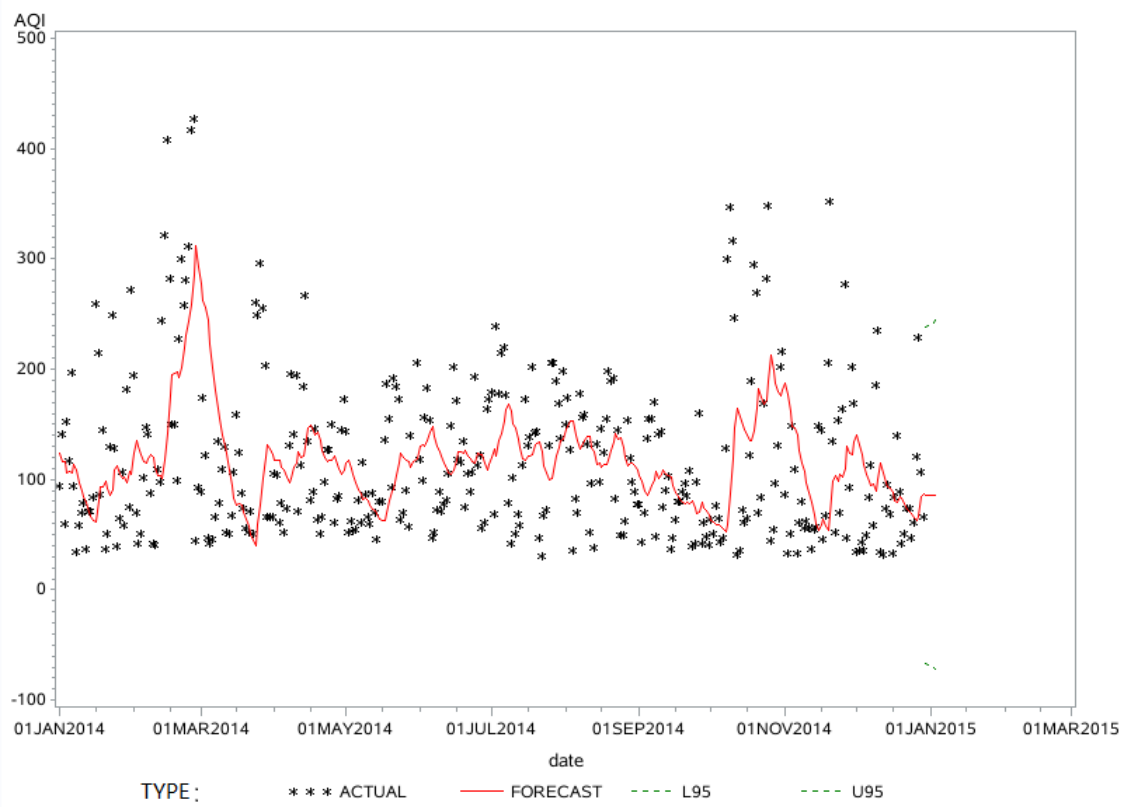


Figure 2. The fitted Holt exponential smoothing model and timing diagram plot of AQI value of 2014 in Miyun County, Beijing.

3. Conclusions

This paper does a study on 2014 the air quality index (AQI) in Miyun County, Beijing, China. In the process of model building, the original AQI data is found to be non-stationary. Because it is difficult to match the trends in the data to any model of linear, polynomial and exponential models, the Holt exponential smoothing model is chosen as it can model complex trends comparing with other smoothing models. In this study, the initial value of the smoothing series $\tilde{X}_0$ is set to be X_1, and the initial value of the trend series b_0 is defined that for an arbitrarilyn,$b_0 = \mathrm{X}_{\mathrm{n}+1} - X_1/\mathrm{n}$. The Holt model fitting result is adequately to capture the trends in the original AQI data, and it can be used to predict AQI value of future short days. The predict model can be used to help government and other authorities to take advanced measures to the coming air condition.

Acknowledgements

This paper is funded by the project of National Natural Science Fund, Logistics distribution of artificial order picking random process model analysis and research (Project number: 71371033); and funded by intelligent logistics system Beijing Key Laboratory (No.BZ0211); and funded by scientific-research bases---Science & Technology Innovation Platform---Modern logistics information and control technology research (Project number:PXM2015_014214_000001); University Cultivation Fund Project of 2014-Research on Congestion Model and algorithm of picking system in distribution center(0541502703).

REFERENCES

[1]. Garcia, Javier; Colosio, Joëlle (2002). Air-quality indices: elaboration, uses and international comparisons. Presses des MINES.

[2]. "People's Republic of China Ministry of Environmental Protection Standard: Technical Regulation on Ambient Air Quality Index".

Access:http://kjs.mep.gov.cn/hjbhbz/bzwb/dqhjbh/jcgfffbz/201203/W020120410332725219541.pdf

[3]. Box, G.E.P., Jenkins, G.M., and Reinsel, G.C.(1994), Time Series Analysis: Forecasting and Control, 3rd edition, Prentice Hall: Englewood Cliffs, New Jersey.

[4]. Anders Milhoj (2013). Practical Time Series Analysis Using SAS. NC: SAS Institute Inc, Cary.

[5]. SAS Institute Inc, (2014). SAS/STAT® 9.4 User's Guide: The ESM Procedure (Book Excerpt). NC: SAS Institute Inc, Cary.

[6]. Bollerslev T. Generalized autoregressive conditionalheteroskedasticity [J]. Journal of Econometrics, 1986, 31 (3):309-317.

[7]. Engle R.F. Autoregressive conditional heteroskedasticity withestimates of the variance of United Kingdom inflation [J].Econometric, 1982, 50 (4): 989-1004.

[8]. Engle R.F., Kroner F.K. Multivariate Simultaneous Generalized ARCH [J].Econometric Theory, 1995, 11:135-149.

[9]. Tsay, R.S., and Tiao, G.C. (1984), "Consistent Estimates of Auto-regressive Parameters and Extended Sample Auto-correlation Function for Stationary and Non-stationary ARMA models," Journal of American Statistical Association, 79, 84-96.

[10]. Cox, D. R., & Wermuth, N. (1991). A simple approximation for bivariate and trivariate normal integrals. International Statistical Review/Revue Internationale de Statistique, 59(2), 263-269.

[11]. Engle Robert F. Dynamic Conditional Correlation: A SimpleClass of Multivariate GARCH Models [J]. Journal of Business and Economic Statistics, 2002, 20 (3):341-347.

[12]. Engle R.F., Lilien D.M., Robins R.P. Estimating time-varyingrisk Premia in the term structure: The ARCH-M model [J].Econometrica, 1987, 55: 395-406.