

Investigation of the relationship between vehicle number in Changzhou and air pollution

Le Yang^{1,a,*}, You Jiajie^{2,b}

¹Nanjing University of Posts and Telecommunications, Nanjing, China

² China Construction Bank Jiangsu Branch Banking Department, Nanjing, China

^aleyang@njupt.edu.cn, ^b66405034@qq.com

Keywords: ARIMA model, AQI, vehicle pollution, emission reduction

Abstract: Along with economic development, the number of vehicles in China grows rapidly. Vehicle emission has become a major source of pollution in urban settings. Based on data on the number of vehicles from 2004 to 2014 in Changzhou, the ARIMA model is built to predict the number of vehicles from 2015 to 2020. A comparison of the AQI in Changzhou between 2014 and 2015 shows that both the improvement of standards for vehicle fuel and the dissemination of new technology in clean-energy cars can help improve air quality.

Introduction

In 2013, continuous large-scale haze pollution broke out in China, an incident in which 17 provinces or cities and autonomous regions were involved. The extent and duration of the haze were unprecedented. According to the “Annual Report for Vehicle Pollution Prevention in China in 2013”^[1], the number of vehicles over the whole of China in 2012 reached 2.24 hundred million, and the total amount of pollutant emission from vehicles throughout the country reached 46,121,000 tons. Thus, the establishment of a vehicle emission inventory and research on its control strategy are important parts in the process of controlling vehicle pollution. In this paper, the ARIMA model is built based on vehicle number data from 2004 to 2014, and it is used to predict the number of vehicles from 2015 to 2020. Meanwhile, a comparison between the air quality index (AQI) in 2014 and that in 2015 in Changzhou is carried out to find out the relationship between the number of vehicles and the AQI, which is of important practical significance for the reduction of vehicle emissions in an urban environment.

Prediction of number of vehicles

The growth of the number of vehicles is influenced by factors such as the GDP, the population and city policies. The autoregressive integrated moving average model, or ARIMA, combines the advantages of an autoregressive model (AR model) and a moving average model (MA model) effectively, and can predict and take into account the synthetic influence of all factors. The number of vehicles in Changzhou changes with the imbalances in time, thus it can be treated as an imbalanced time series. By establishing a time series model for some relevant values and estimating parameters based on historical data, simulation and prediction can be realized.

Principle of ARIMA model

The model, which is built based on differential data series, is called the ARIMA model^[2], and denoted as $\{X_t\} \square ARIMA(p,d,q)$, where $\{X_t\}$ is the data series, p and q are orders of the model,

and d is the differential number^[4]. When $d=0$, the ARIMA model coincides with an ARMA model, which is defined as:

$$X_t = \sum_{i=1}^p a_i X_{t-i} + e_t + \sum_{j=1}^q b_j e_{t-j} \quad (1)$$

wherein a_i and b_j are the parameters we need to estimate; $e_t \sim N(0, \sigma_x^2)$ is a white noise series, and σ_x^2 is the variance of the white noise. When p and q are equal to 0, respectively, the model coincides with a MA model and an AR model.

Algorithm steps of ARIMA model

After the time series data are balanced, by using the ARIMA model for prediction, the autocorrelation functions (ACF) and partial autocorrelation functions (PACF) of the original data should be first calculated to determine the corresponding model and its appropriate order correctly.

As for the balanced time series $\{X_t\}$, the autocorrelation coefficient of the sample should be determined according to

$$\hat{g}_k = \frac{\sum_{j=1}^{N-k} (X_k - \bar{X})(X_{t+k} - \bar{X})}{\sum_{j=1}^N (X_k - \bar{X})^2}, \forall 0 < k < n \quad (2)$$

wherein, \hat{g}_k is the autocorrelation coefficient and \bar{X} is the average of the data series.

The Yule-Walker equation is solved via the Cramer principle^[3], and the partial correlation coefficient of the sample is

$$\hat{r}_k = \frac{\hat{D}_k}{\hat{D}}, \forall 0 < k < n \quad (3)$$

where, $\hat{D} = \begin{vmatrix} 1 & \hat{g}_1 & \mathbf{L} & \hat{g}_{k-1} \\ \hat{g}_1 & 1 & \mathbf{L} & \hat{g}_{k-2} \\ \mathbf{M} & \mathbf{M} & & \mathbf{M} \\ \hat{g}_{k-1} & \hat{g}_{k-2} & \mathbf{L} & 1 \end{vmatrix}$, and $\hat{D}_k = \begin{vmatrix} 1 & \hat{g}_1 & \mathbf{L} & \hat{g}_{k-2} & \hat{g}_1 \\ \hat{g}_1 & 1 & \mathbf{L} & \hat{g}_{k-3} & \hat{g}_2 \\ \mathbf{M} & \mathbf{M} & & \mathbf{M} & \mathbf{M} \\ \hat{g}_{k-2} & \hat{g}_{k-3} & \mathbf{L} & 1 & \hat{g}_{k-1} \\ \hat{g}_{k-1} & \hat{g}_{k-2} & \mathbf{L} & \hat{g}_1 & \hat{g}_k \end{vmatrix}$, is obtained by replacing

the k -th column vector with $[\hat{g}_1 \ \hat{g}_2 \ \mathbf{L} \ \hat{g}_k]^T$, while \hat{r}_k is the partial correlation coefficient.

The model order is determined preliminarily based on the truncation property of \hat{g}_k and \hat{r}_k .

The judging criteria are shown in Table 1.

Table 1. ARMA model identification criteria for balanced time series

| Category | Model identification | | |
|---------------------------------|----------------------|------------|-------------|
| | $AR(p)$ | $MA(q)$ | $ARMA(p,q)$ |
| Autocorrelation coefficient | Trailing | Truncation | Trailing |
| Partial correlation coefficient | Truncation | Trailing | Trailing |

After the model is determined, the AIC function is employed to determine the model order accurately^[4],

$$AIC(k) = \ln \hat{S}_k^2 + \frac{2k}{N}, \quad AIC(k) = \ln \hat{S}_k^2 + \frac{2k}{N} \quad k = 0, 1, \mathbf{L}, p_0 \quad (4)$$

wherein, k is the model order, \hat{S}_k^2 is variance of the residual, and N is the length of the handled data series.

After the model is identified, the parameters in the model should be estimated. In this paper, least square estimation is used to estimate parameters $a_1, a_2, \mathbf{L}, a_p, b_1, b_2, \mathbf{L}, b_q$ so that the sum of squared residuals

$$\sum_{t=1}^N e_t^2 = \sum_{t=1}^N (b_q^{-1}(B) a_p(B) X_t)^2 \quad (5)$$

reaches a minimum, where B is corresponding delay operator, and $\{X_t\}$ is the stable series after processing.

Prediction of number of vehicles

After a survey of current published yearbooks, data on the number of vehicles from 2004 to 2014 have been obtained in order to build the ARIMA model, and then the number of vehicles from 2015 to 2020 was predicted. According to the algorithm presented above, the estimated result is expressed as:

$$y = -0.0214 \times x^3 + 0.919 \times x^2 - 1.5928 \times x + 7.874 \quad (6)$$

Figure 1 shows the time-dependent trend of vehicle number from 2004 to 2014. Based on the prediction, the vehicle number in 2015 should be 841,200, while the result given in the yearbook is 822,000. The error between the predicted and actual values is below 2.3%. It is thus clear that the prediction can effectively reflect growth of vehicle number in Changzhou at present. The prediction also shows that in 2020 the number of vehicles in Changzhou will exceed 1,400,000. Such an excessive vehicle number will definitely lead to a huge pressure for the reduction of vehicle pollution.

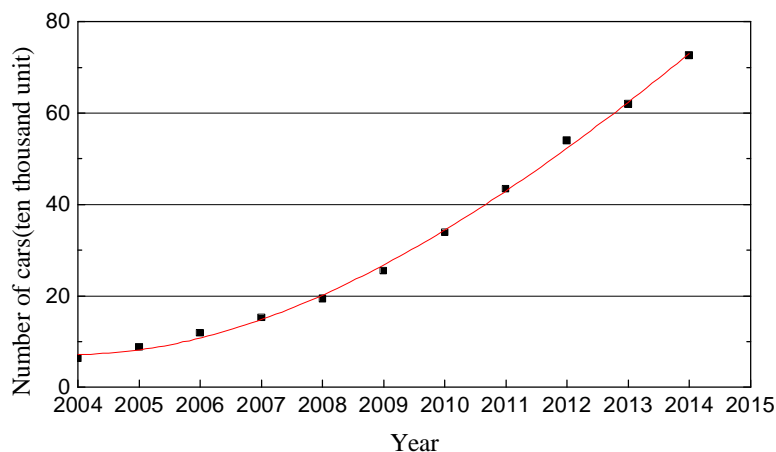


Figure 1. Time-dependent trend of vehicle number

Analysis of pollutants

Pollutant emission of vehicles is influenced by factors such as vehicle emission standards and actual service life. It is known from [7] that the emission of CO and PM in mini buses is very striking, while heavy diesel cars (large buses, heavy trucks, medium trucks) are important sources of NO_x and PM .

The comparison of Changzhou's AQI between 2014 and 2015 was carried out on a daily and monthly basis, as shown in Figures 2 to 5. It is known from relevant data that in Changzhou at the end of 2015, the number of civilian cars reached 968,000, and increased by 96,000 compared with the end of previous year, which is a 10.9% increase. Among them, the number of private cars reached 822,000, and increased by 95,200 compared with the end of the previous year, which is a 13.1% increase. It is also known from Figures 2 to 5 that, although number of vehicles in 2015 increased 10.9% compared to 2014, the values of two indexes i.e. PM2.5 and PM10 in 2015 were still lower than those in 2014. The concentration index of CO in 2015 is basically the same as that of 2014. As for the concentration index of NO_2 , the relevant value in the first half of 2015 is obviously better than that of 2014; however, the value of the index reached the limit on some days in the second half of 2015, which leads to a slightly higher overall average than in 2014.

By analyzing the above data, the change of the AQI in 2014 and 2015 may be due to the following reasons:

- 1) Although National Standard IV came into effect officially from July 1, 2013 in Changzhou, like many areas, some vehicles satisfying National Standard III were registered by agencies in advance. Therefore, those cars were still being digested in 2014. However, in 2015, vehicles in the market or ready to register already satisfied National Standard IV. It follows that the improvement of standards for vehicle fuel can help to improve air quality.
- 2) According to statistics at the end of 2014, the number of public vehicles was 2978, and there are almost 900 pure electric vehicles and gas vehicles among them. However, according to 2015 data, the number of public vehicles in operation was 3135, which is similar to that of the previous year, but there are 2250 pure electric vehicles and gas vehicles among them. Public vehicles using clean energy form 71.8% of the total. Taking into account that public vehicles play a major role in traffic network, we can greatly reduce the pollution from these vehicles by replacing them with clean energy vehicles.

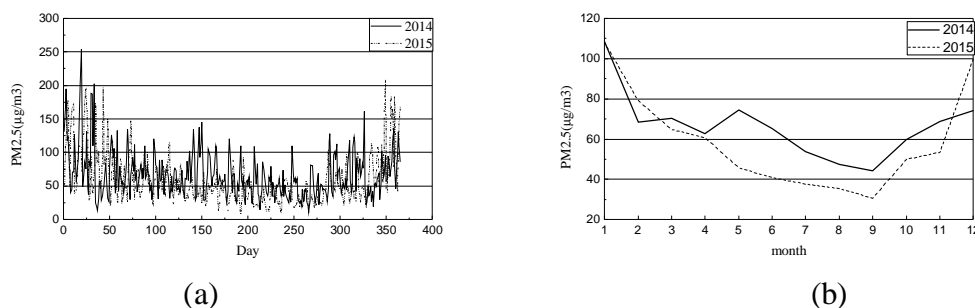


Figure 2. Comparison of PM2.5 concentration in different years

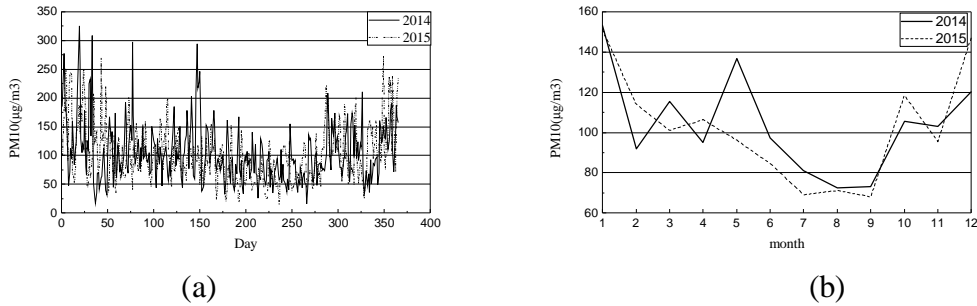


Figure 3. Comparison of PM10 concentration in different years

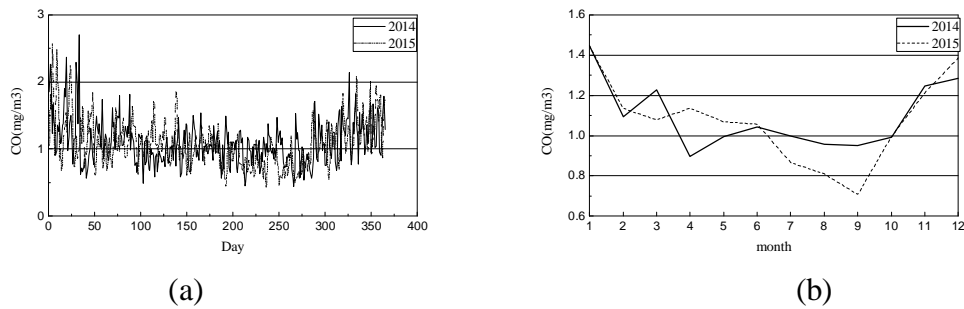


Figure 4 Comparison of CO concentration in different years

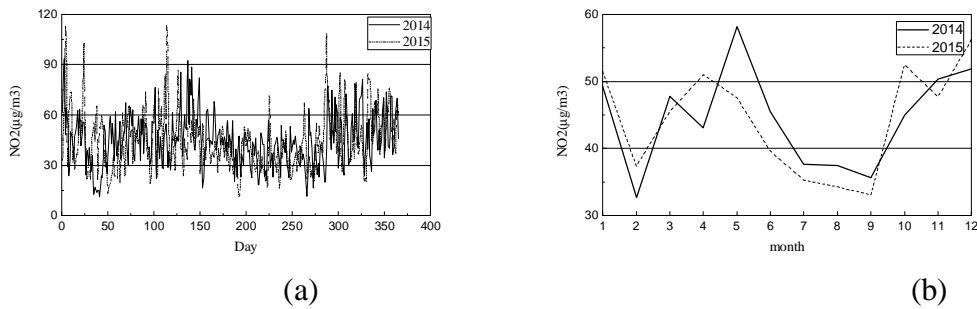


Figure 5. Comparison of NO_2 concentration in different years

Conclusion

In this paper, data on the number of vehicles from 2004 to 2014 in Changzhou were obtained from published yearbooks. The ARIMA model was built to predict the vehicle number in Changzhou from 2015 to 2020, and its accuracy was verified, as results show that the error between the predicted and actual values is below 2.3%. It is thus clear that the prediction can effectively reflect the growth of the number of vehicles in Changzhou at present. A comparison of Changzhou's AQI between 2014 and 2015 was also carried out on a daily and on a monthly basis, and the results indicate that the improvement of standards for vehicle fuel and the dissemination of new technology for clean-energy cars can help improve air quality.

Acknowledgements

This research is supported by National Natural Science Foundation of China (No. 41401480), A project funded by the NASG Key Laboratory of Land Environment and Disaster Monitoring (No. LEDM2014B03), Science foundation for talent bringing of Nanjing University of Posts and Telecommunications (No. NY212036), Science Foundation of Nanjing University of Posts and

Telecommunications (No. NY214110, No. NY215181).

Reference

[1] Ministry of Environmental Protection of the People's Republic of China. Annual report of motor vehicle pollution prevention and control in China (2013) [EB].

<http://www.zhb.gov.cn/gkml/hbb/qt/201401/W020140126591490573172.pdf>

[2] Liu Jiye, Chen Xihong, Xue Lunsheng and Liu Qiang. A combined ARIMA and ANN method for prediction of satellite clock error[J]. Journal of Geodesy and Geodynamics, 2013, 33(4): 107-111

[3] Zhao Liang, Lan Xiaoqi, Sheng Jianyue. Application of ARIMA model in satellite clock error forecasting[J]. Journal of Water Resources and Architectural Engineering, 2012, 10(1): 135-137

[4] Peng Sijun, Shen Jiachao, Zhu Xue. Forecast of PM2.5 based on the ARIMA model[J]. Safety and Environmental Engineering. 2014, 21(6): 125-128