ISSN 2616-5775 Vol. 4, Issue 3: 88-93, DOI: 10.25236/AJCIS.2021.040314

Annual Water Consumption Forecast of Hefei Based on ARIMA Model

Zhu Bo[1,3,a,*], Liu Yezheng[1,b], Zheng Feifei[2,c]

[1]*School of Management, Hefei University of Technology, Hefei, Anhui Province, China*
[2]*College of Civil Engineering and Architecture, Zhejiang University, Hangzhou Zhejiang Province, China*
[3]*Hefei Water Supply Group Co Ltd,, Hefei, Anhui Province, China*
[a]*nationsteed@outlook.com,* [b]*liuyezheng@hfut.edu.cn,* [c]*feifeizheng@zju.edu.cn*
**Corresponding Author*

Abstract: *The ARIMA entity model was used in Hefei's current water demand coding sequence from 1990 to 2018. According to the ADF (Unit Root Test), it is clear that the number of differences d in the entity model is 2, and the main parameters in the entity model p, q are basic identification based on time series analysis related graphs, and based on the Akaike Information Content Rule (AIC) and other methods It is clear that the optimal entity model is ARIMA (1,2,1). ARIMA (1,2,1) is used to predict and analyze Hefei's water demand in the next two years. The results show that the reason for the deviation between the estimated value and the specific value is relatively small, indicating that the actual effect of the actual model predictive analysis is excellent.*

Keywords: *ARIMA entity model of annual water consumption; predictive analysis; AIC*

1. Introduction

The development of a city cannot do without water. Water is necessary for both production and living. The water demand of a city is the total water demand of a city in a period of time. It is the embodiment of living and production demand of water city. All aspects of social life are mainly composed of the following ten parts: agriculture, forestry, animal husbandry, fishery and water conservancy; Industry; Construction; Transportation, warehousing, postal services; Digital economy, industrial Internet and information service industry; Industry and commerce, accommodation and catering industry; Financial, real estate, commercial and residential services ;Public goods supply and public management; Water, electricity, gas and other public utilities. In modern society, all walks of life use water, urban water and power supply has become a key indicator of national economic development. Urban water demand index is not only concerned by water supply and drainage related science and technology, but also an important reference for economists to study macro and micro economic operation.

In recent years, many scholars at home and abroad have deeply studied the problem of water supply demand forecasting. Sun Xiaoting et al. (2017) [1] used chaos theory to build a model to predict the overall trend of series data, and then modified the prediction results based on training neural network with small sample; Wang pan et al. (2014) [2] established a water demand forecasting model by using random forest algorithm. Guo Guancheng et al. (2018) [3] used bidirectional long-term and short-term neural networks to construct short-term water supply prediction; Shabani s et al. (2017) [4] reconstructed the phase space of chaotic time series, combined with SVR to build a water prediction model.Bai y et al. (2014) [5] used the multi-scale correlation vector regression method to build the urban water demand forecasting day model; He Bo et al. (2019) [6] established a decision-making model of urban daily water supply based on multi granularity characteristics and xgboost model.

However, the research methods of scholars in the above literature need to collect a large number of index data related to urban water supply demand;At the same time, because there is no unified standard for the scale of indicators, the selection of indicators depends too much on the subjective factors of scholars. ARIMA model does not need exogenous variables, only endogenous variables. At the same time, the historical data needed by the model is easy to collect, and relatively less. A large number of experiments show that the prediction accuracy of ARIMA model is better. ARIMA model has developed rapidly in all aspects of time series data prediction in recent years, and has a large number of application scenarios. Ding shouluan et al. (2003) [7], Shan Liang (2015) [8], Wang Jianshu et al. (2018) [9] applied

ISSN 2616-5775 Vol. 4, Issue 3: 88-93, DOI: 10.25236/AJCIS.2021.040314

ARIMA model to medical disease prediction, Liu Zhenwei (2009) [10], Guo Xiaofeng (2012) [11], sun Zhaohui (2015) [12], Ao Xiqin (2017) [13] used ARIMA model to forecast macro economy, and GE Ling et al. (2018) [14] used ARIMA model to forecast railway passenger volume.Wang Yuefen et al. (2014) [15], pan Yurong et al. (2018) [16] applied ARIMA model to study the power consumption and short-term electricity price of the whole society, and Rafael Ortega huedo et al. (2019) [17] used ARIMA model to study the social change trend.The antibiotic resistance rate of European hospitals, Gu Jianwei et al. (2018) [18] on oil well production, Cai Chengzhi et al. (2019) [19] using ARIMA model to predict world wheat production, the above research has achieved good results.In view of this, this paper intends to use ARIMA model to predict the annual water demand of Hefei city.

2. ARIMA model

1970In the early s, box and Jenkin proposed a new prediction method based on time series, namely Arima, which is also called box Jenkins model or box Jenkins method. In ARIMA (P, D, q) model, the autoregressive term is AR and the autoregressive term is p;The number of Ma terms is Q, and D is the difference when the time series becomes stationary. The first step of ARIMA model is to stabilize the practice sequence, and the second step is to regress the lag value of dependent variable and the current and lag value of random error.Arima includes MA (moving average), AR (autoregressive), ARMA (autoregressive moving average) and ARIMA （Modified here)

AR (P) model is established by the weighted sum of P-term and a random disturbance term:

$$X_t = c + \sum_{i=1}^{p} \varphi_i X_{t-i} + \varepsilon_t \quad (1)$$

Among them, X_t Is the observed value of time t, P is the number of autoregressive terms, ε_t It is a white noise sequence with a mean value of 0 and constant variance. The MA (q) model has the current value of random interference term and the weighted sum of lag Q term to construct the model. Its form is

$$X_t = c + \sum_{t=1}^{q} \theta_i \varepsilon_{t-i} \quad (2)$$

ARMA (P, q) model uses the weighting of the past value, the current value and the lagged random disturbance term of time series to build the model

$$X_t = c + \sum_{i=1}^{p} \varphi_i X_{t-i} + \sum_{i=1}^{q} \theta_i \varepsilon_{t-i} \quad (3)$$

When q = 0, ARMA (P, q) model changes to AR (P) model; When p = 0, ARMA (P, q) model changes to MA (q) model. For convenience and security, ARMA (P, q) model is usually used to model time series directly. However, ARMA (P, q) model needs a fixed time series, and in the real environment, it is mostly non-stationary time series. Time series are usually smoothed by difference operation. In fact, ARMA (P, q) model usually differentiates D times and then becomes a stationary sequence. ARMA (P, q) model is called ARIMA (P, D, q) model.

The key to building ARIMA (P, D, q) model is to identify the three parameters of P, D and Q in the model. Parameter D is the number of differences needed to smooth the time series. The stationarity of identification data can be determined by time series, and the data stationarity needs to be determined by unit root test. The parameters p and Q can be determined in advance through the partial autocorrelation coefficient function diagram and the autocorrelation function diagram of fixed time series. The autocorrelation function graph and partial autocorrelation coefficient function graph gradually decay to 0 in P and Q periods, and both of them will appear tailing phenomenon. Then, according to Akaike information criterion, Bayesian information criterion (AIC, SC minimum principle) and adjusted R2 maximum principle, the optimal P and Q are determined, so as to determine the optimal model parameters of ARIMA (P, D, q).

ISSN 2616-5775 Vol. 4, Issue 3: 88-93, DOI: 10.25236/AJCIS.2021.040314

3. Case study

The water demand data of this paper comes from the statistical annual report of Hefei water supply group. The water demand data of Hefei from 1990 to 2018 are collected. The water demand of Hefei from 1990 to 2017 is used to establish the water demand, and the water demand of Hefei in 2018 is used to test the prediction effect of ARIMA model. Record Xi as the water demand data series of Hefei from 1990 to 2017, and the water demand sequence diagram (Figure 1). From Figure 1, we can see that the trend of water demand of Hefei is obviously rising, which shows that the data series is not stable. In order to change the non-stationary sequence into a stationary sequence, the first step is to take the natural logarithm Yi = lnXi for the data Xi, the second step is to make a difference for the data, which is recorded as dYi, and at the same time, plot the scattered points of the sequence dYi (Figure 2). From the trend of the fitting curve in Figure 2, we can make a preliminary judgment that the time series dYi is stationary.

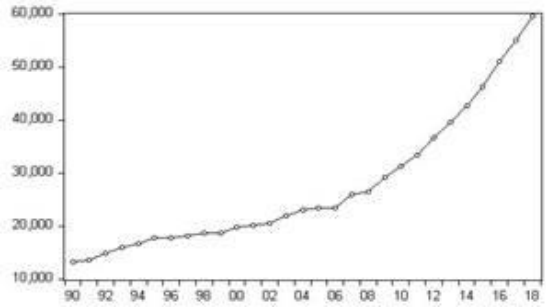

Figure 1 Scatter plot of time series Xi

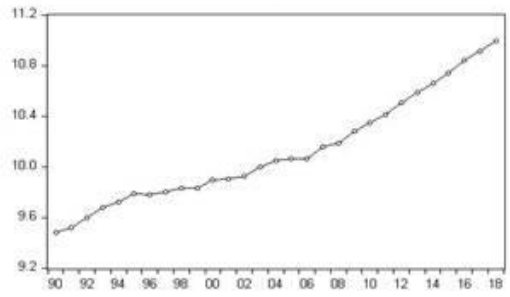

Figure 2 Scatter plot of Xi natural logarithm Yi

The third step is to observe the stationarity by drawing the trend diagram of dYi. Figure 3 shows that dYi is obviously unstable, so the second difference for Yi is recorded as, and Figure 4 is the trend diagram of d2yi. From the diagram, d2yi is a stationary sequence, and the stationarity of d2Yi will be determined by the unit root test (Eviews software is used in this paper). If it is not stable enough, the third difference will be carried out until it is stable. The unit root test results of d2Yi are shown in Table 1.From table 1, the T statistical value of ADF test is -6.740550, which is less than the critical point and probability value of 1% of the test level of -3.724070. P is close to 0, which is much smaller than the significance level of 0.05. Therefore , the null hypothesis that the time series d2Yi has a unit root is rejected. The time series d2Yi has no unit root, so the time series d2Yi is fixed [20].

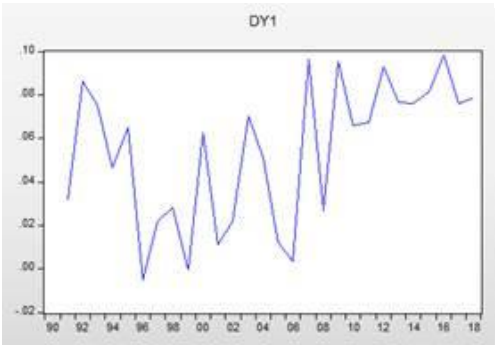

Figure 3 Differential sequence dYi trend chart

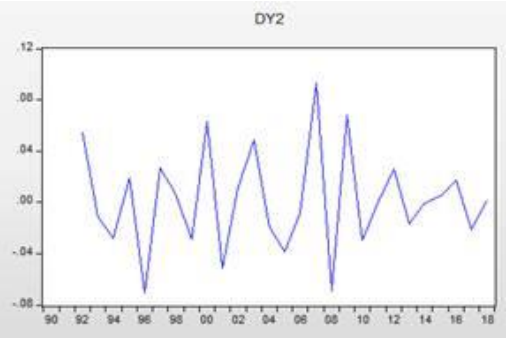

Figure 4 Trend chart of differential sequence d2Yi

Table 1 ADF unit root test of d2Yi

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-6.74055	0.0000
Test critical values:	1% level	-3.72407	
	5% level	-2.986225	
	10% level	-2.632604	

Since the time series d2Yi is stationary, it can be determined that the parameter D in ARIMA (P, D, q) is 2.Then we continue to determine P and Q in ARIMA (P, D, q).After determining the three parameters, we will draw the autocorrelation graph and local autocorrelation graph of time series d2Yi (see Figure

ISSN 2616-5775 Vol. 4, Issue 3: 88-93, DOI: 10.25236/AJCIS.2021.040314

5).The partial correlation function diagram in Figure 5 can lag behind the two steps of 95% confidence interval, and the subsequent steps are all within the confidence interval, and there is tailing. Combined with the principle of simplicity, it is more appropriate that P is 1 or 2; Looking at the autocorrelation function diagram in Figure 5, we can see that Q is 1. Therefore, ARIMA (2,2,1) and ARIMA (1,2,1) are the primary models.

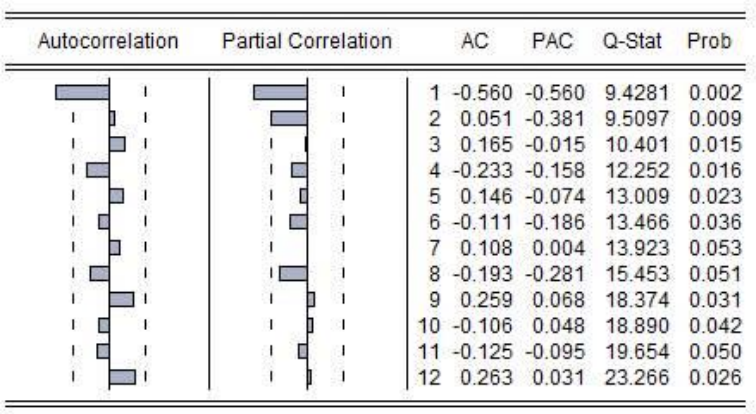

Figure 5 Autocorrelation and partial autocorrelation of time series d2Yi

The following is the most ARIMA model determined by Akaike information criterion, adjusted R2 and Bayesian criterion. The comparison of AIC, SC and adjusted R2 of the two models is shown in Table 2. It can be seen from table 2 that the AIC and SC of ARIMA (1,2,1) are the smallest, and the adjusted R2 of the two models are very close. Therefore, the optimal model to be considered is ARIMA (1,2,1).

Table 2 Comparison of the results of the two models

Model	Adjusted R2	AC	SC
ARIMA(1,2,1)	0.4706780	-3.839092	-3.647116
ARIMA(2,2,1)	0.4719402	-3.834762	-3.594792

The coefficient estimation table of ARIMA (1,2,1) is obtained by using the least square method in Eviews software and the estimate equation command, as shown in Table 3

Table 3 Main estimation results of ARIMA (1,2,1)

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.001834	0.001960	0.935939	0.359
AR(1)	0.092834	0.260414	0.356485	0.7247
MA(1)	-1.000000	5991.793	-0.000167	0.9999
SIGMASQ	0.000833	0.110812	0.00752	0.9941
R-squared	0.470678	Mean dependent var		0.001726
Adjusted R-squared	0.401636	S.D. dependent var		0.040434
S.E. of regression	0.031278	Akaike info criterion		-3.839092
Sum squared resid	0.022501	Schwarz criterion		-3.647116
Log likelihood	55.82774	Hannan-Quinn criter		-3.782007
F-statistic	6.817279	Durbin-Watson stat		1.992762
Prob（F-statistic）	0.00188			
Inverted AR Roots	0.09			
Inverted MA Roots	1			

It can be seen from table 3 that the ARIMA (1,2,1) model equation established by the water consumption of Hefei from 1990 to 2018 is as follows:

$$\Delta^2(x_t) = 0.001834 + 0.09834\Delta^2(x_{t-1}) + \varepsilon_t - \varepsilon_{t-1} \quad (4)$$

ISSN 2616-5775 Vol. 4, Issue 3: 88-93, DOI: 10.25236/AJCIS.2021.040314

The F statistic of ARIMA (1,2,1) model is 6.817279, and its corresponding p value is 0.001880, which is far less than the test level of 0.05. Therefore, ARIMA (1,2,1) model is significant on the whole, which indicates that ARIMA (1,2,1) model has fully extracted the water consumption data information of Hefei from 1990 to 2017.

Then we test the residuals of ARIMA (1,2,1) model, and the test results are shown in Figure 6.

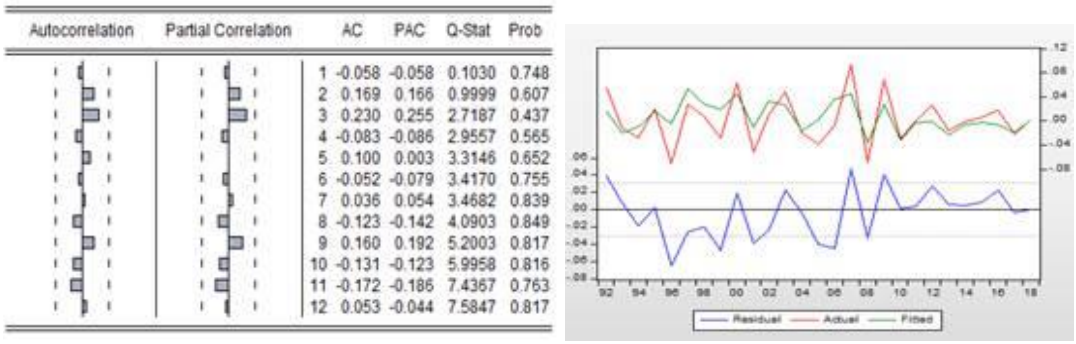

Figure 6 Residual white noise test of ARIMA (1,2,1)

It can be seen from Figure 6 that the prob (residual autocorrelation coefficient) of ARIMA (1,2,1) model is within the 95% confidence interval, and the minimum probability of the value of the lag correlation function is 0.437, which is far greater than the test level of 0.05, so it cannot be rejected.The initial assumption is that ARIMA (1,2,1) model does not have autocorrelation. The conclusion fully shows that ARIMA (1,2,1) model is appropriate.

Finally, the ARIMA (1,2,1) model is used to predict the water demand of Hefei in the next four years, that is, from 2019 to 2022. The prediction results of urban water demand of Hefei are shown in Table 4.

Table 4 Predicted values of ARIMA (1,2,1)

particular year	2019	2020	2021	2022
Estimate	63598.55	71507.85	77752.98	83144.09
				Unit:10000m3

It can be seen from table 4 that the predicted annual water consumption of Hefei in 2019 is 635.9855 million tons, while the actual water consumption of Hefei in 2019 is 618.3613 million tons (the statistical interval is from December 26, 2018 to December 26, 2019). The absolute error between the two is 17.6242 million tons, and the relative error is only 2.85%, which fully shows that the prediction accuracy of ARIMA (1,2,1) is appropriate.

4. Results and Discussion

The medium and long-term prediction of urban water consumption is an important basis for urban water supply development planning. At the same time, the data is also an important reference for urban economic operation, so the prediction of urban water demand is of great significance. ARIMA model needs less historical water consumption data, and also considers the correlation and stability of the data itself. This paper uses ARIMA model to predict the annual water consumption of Hefei city. Taking the water demand data of Hefei from 1990 to 2018 as the data source, the unit root test, autocorrelation and partial autocorrelation graph are used to test and analyze the stability of time series and identify the parameters. Finally, the optimal model is determined by Akaike criterion and other discriminant methods. The ARIMA (1,2,1) prediction model is established by taking the logarithm of water demand data of Hefei, The model is used to predict the water demand of Hefei from 2019 to 2022. Through comparison, it is found that the predicted value in 2019 is very close to the real value, and the effectiveness of the model is identified in practice. The results show that the ARIMA (1,2,1) model can be used to predict the annual water consumption of Hefei City, which has a certain reference value for the annual planning of water supply enterprises, long-term construction planning and the decision-making of government economic management departments.

References

[1] Sun Xiao-ting, Liu Niandong, Du Kun, Zhou Ming, Ren ganghong(2017). Prediction of water supply

ISSN 2616-5775 Vol. 4, Issue 3: 88-93, DOI: 10.25236/AJCIS.2021.040314

based on chaos local area method and neural network [J]. Civil architecture and environmental engineering, vol.39 ,no.05,pp.135-139

[2] Wang pan, Lu Baohong, Zhang Hanwen, Zhang Wei, sun Yinfeng, Ji Yu(2014). Water demand prediction model based on Stochastic Forest Model and its application [J]. Water resources protection, vol.30 ,no.01,pp.34-37 + 89

[3] Guo Guancheng, Liu Shuming, Li Junyu, Zhou Ren, Zhu Xiaoyun(2018). Study on water quantity prediction method based on bidirectional long and short time neural network [J]. Water supply and drainage, vol.54 no.03,pp.123-126.

[4] Shabani S, Yousefi P, Naser G(2017). Support Vector Machines in Urban Water Demand Forecasting Using Phase Space =Reconstruction[J]. Procedia Engineering,no.186,pp.537-543.

[5] Bai Y, Wang P, Li C, et al(2014). A multi-scale relevance vector regression approach for daily urban water demand forecasting[J]. Journal of Hydrology,vol.517,no.2,pp.236-245.

[6] He Bo, Ma Jing, Gao Heyu. Prediction of urban daily water supply based on multi granularity characteristics and xgboost model [J / OL]. Journal of Yangtze River academy of Sciences http://kns.cnki.net/kcms/detail/42.1171.TV.20190509.1102.004.html.

[7] Shan Liang(2015). Prediction and analysis of hand foot mouth disease incidence trend in Jianye District of Nanjing City Based on time series model [J]. Practical preventive medicine, vol.22 no.9,pp.1143-1147.

[8] Ding Shouluan, Kang Jiaqi, Wang Jiezhen(2003).ARIMA model in the prediction of incidence rate [J]. China Hospital Statistics.no.01.

[9] Wang Jianshu, Liu Qiang, hang Hui, Qin Jiangchun, Yang Haibing(2018). Application of ARIMA product seasonal model in prediction of other infectious diarrhea in Suzhou [J]. Occupational and health. No.06.

[10] Guo Xiaofeng(2012). Prediction and analysis of China's CPI trend based on ARIMA model [J]. Statistics and decision making.no.11.

[11] Liu Zhenwei(2009). Empirical analysis on the future trend of China's CPI based on ARMA model [J]. North China finance. No.09.

[12] Ao Xiqin, Gong Yujie, Wang Jinting, et al(2017). Prediction of CPI in Anhui Province Based on SARIMA model [J]. Journal of Bengbu University, vol.6,no.3,pp. 83-86.

[13] Sun Zhaohui(2015). Application of time series model in GDP forecast of Anhui Province [J]. Journal of Bengbu University, vol.4 ,no.2,pp. 95-99.

[14] Ge Ling, Zhang Jie(2018). Prediction of railway passenger traffic volume based on product season model [J]. Journal of Chongqing business and Technology University (NATURAL SCIENCE EDITION), vol.35,no.3,pp.18-25.

[15] Wang Yuefen, Wang Lu(2014). Empirical analysis of total domestic electricity consumption in Ningbo based on ARIMA model [J]. Journal of Zhejiang University of Technology (SOCIAL SCIENCE EDITION), vol.32,no.3,pp.204-206.

[16] Pan Yurong, Jia Chaoyong(2018). Short term electricity price forecasting based on seasonal ARIMA model [J].Journal of Baicheng Normal University,vol.32 ,no.6,pp. 18-24.

[17] Rafael Ortega-Huedo, Marina Cuesta, Andreas Hoefer, Bruno Gonzalez-Zorn(2019). Econometric ARIMA methodology to elucidate the evolution of trends in nosocomial antimicrobial resistance rates in the European Union[J]. International Journal of Antimicrobial Agents.

[18] Gu Jianwei, Sui Gulei, Li Zhitao, Liu Wei, Wang Yike, Zhang Yigen, Cui Wenfu(2018). Oil well production prediction based on ARIMA Kalman filter data mining model [J]. Journal of Shenzhen University (Science and Engineering Edition). no.06.

[19] Cai Chengzhi, Li Yingying, Zheng Weiwei, Jiang Xingzi, Yu Fei, Zuo Jin, Zeng Xiaoshan(2019). Prediction and analysis of world wheat yield per unit area from 2018 to 2022 based on ARIMA model [J]. Hebei Agricultural Sciences, vol.23,no.04,pp. 89-92.

[20] Wang Yanmei, Chen Xizhen, Dong Naiming(2012). The development trend of per capita GDP in Shanxi Province -- Based on time series, least square regression and quantile regression[J]. Science, technology and engineering, vol.12 ,no.18,pp. 4575-4578.