

Water quality prediction analysis of Qingyi River based on time series

Yulan Luo¹, Qingsong Chen², Ying Liu^{1*}, Xiaohui Xie¹, and Qianying Du¹

¹Departments of Geosciences and Environmental Engineering, Southwest Jiaotong University, 610000, Chengdu, China

²Sichuan Province Environmental Protection Science & Technology Engineering Co., Ltd, 610041, Chengdu, China

Abstract. According to the current situation of water quality in drainage basin, the key to improve the prediction accuracy is to select the appropriate prediction model of water quality. The time series method excellently reflected the continuity of the future data in the case of emphasizing historical data. What's more, the time series method has the higher short-term prediction accuracy and simple modeling process. So, the time series method was used to establish the Auto-Regressive and Moving Average (ARMA) model for the time series of the concentration of dissolved oxygen (DO), biochemical oxygen demand (BOD5), chemical oxygen demand (CODCr), ammonia nitrogen (NH₃-N) and total nitrogen (TN) at the Guidu fu section of Qingyi River from January 2011 to December 2015. Then, the concentrations of the five water quality indicators from January to June 2016 were predicted, which were verified and analyzed with the measured values. The results show that the model has fine fitting effect and higher prediction accuracy, which can accurately reflect the current and future change trends of the water quality.

1 Introduction

The water environment problem in drainage basin is one of the most important problems faced by environmental management in China in recent years. Water quality prediction can evaluate water quality changes as early as possible, which has great significance for water environment protection. According to the difference between the theoretical basis and the solution method when establishing the water quality prediction model, the water quality prediction model is divided into mechanism water quality model and non-mechanism water quality model. Because the mechanism water quality prediction model uses the governing equation to describe the changing trend of water quality, it is necessary to identify the parameters with practical physical significance in the modeling process, which leads to the modeling process more complex [1]. Therefore, non-mechanism water quality prediction models are widely used in water quality prediction at present. The commonly used non-mechanism water quality prediction models include artificial neural network method [2], grey system theory method [3-6], time series method [7-8] and so on. But in practical application, the mathematical theory of artificial neural network method is not perfect, the data training speed is too slow, and sometimes it is impossible to obtain results [9]. The grey system theory method has weak anti-interference ability and gray deviation, which leads to the model prediction accuracy not meeting the requirements [10]. The time series method has a relatively complete mathematical theory foundation, which can make full use of historical data to make quantitative predictions for

future water quality, and the short-term prediction accuracy is well [11].

Time series method is a mature data processing method. It analyzes and studies the corresponding time series mathematical model established by the dynamic water quality parameter data, and excavates the periodic information of the water quality change to make an accurate prediction of the data change trend. What's more, the method has been widely used in domestic and foreign economic, life science, physics, computer and other fields [12-17]. In the field of water quality prediction, the time series method is mainly applied to the prediction of water quality indicators, and most of the prediction ranges are more than 12 steps [18]. However, the practice shows that the time series method has higher accuracy in short-term prediction, and the longer the prediction time, the greater the error [19]. Because of the uncertainties of external factors on the water quality of Qingyi River and the insufficient information of relevant data, the prediction accuracy of long-term time scale cannot meet the requirements. Therefore, this study used six-step prediction in the prediction time, in order to ensure the higher accuracy. Qingyi River is a tributary of Minjiang River. Due to the rapid development of social economy along the river in recent years, it inevitably has a negative impact on the water quality. Hongya section of Qingyi River, as the drinking water source of Hongya County in Meishan City, undertakes the drinking water safety guarantee of nearly 400,000 people. This work established an Auto-Regressive and Moving Average (ARMA) model to fit the water quality data of the Hongya section of the Qingyi River from January 2011 to

* Corresponding author: 642823770@qq.com

December 2015, and predicted its evolution trend. In order to provide a scientific basis for local water environment management and protection.

2 Water quality prediction method based on time series

The time series refers to a series of measured data arranged in chronological order. Water quality prediction model of time series is to use the historical measured data of water body to predict the current value of water quality at a certain time in the future.

If a linear combination of the past interference value and the current interference value of time series $\{X_t\}$ is represented by ε_t ; $\{\varepsilon_t\}$ is white noise and is recorded as $WN(0, \sigma^2)$; $a_0, a_1, a_2, \dots, a_p$ ($a_p \neq 0$) are the auto-regressive coefficient; $b_0, b_1, b_2, \dots, b_p$ ($b_p \neq 0$) are the moving average coefficient. Both the auto-regressive coefficient and the moving average coefficient are real numbers and are determined by the least squares parameter estimation method, X_t represents the value of random phenomena at time t . Then, the Auto-Regressive and Moving Average model as follows:

$$X_t = a_1X_{t-1} + a_2X_{t-2} + \dots + a_pX_{t-p} + \varepsilon_t - b_1\varepsilon_{t-1} - b_2\varepsilon_{t-2} - \dots - b_q\varepsilon_{t-q}, t \in Z \quad (1)$$

Which is recorded as ARMA (p, q) model, and the time series that satisfies the model are called the ARMA (p, q) series.

In the equation (1), X_{t-p} represents the measured value of water quality indicator at time $t-p$, and ε_t represents the interference value of the current time, that is, the data in the series have non-extractable information for establishing the model. Therefore, X_t is a linear combination function of the measured value of water quality indicator in the previous time p , the past interference value and the current interference value.

Generally, the modeling process of ARMA (p, q) model are divided into four steps: stationarity test and white noise test of the measured value of water quality, coefficient estimation and model test of water quality prediction model.

3 Establishment of ARMA model for Qingyi River

In this work, five water quality indicators data such as DO, BOD₅, COD_{Cr}, NH₃-N and TN were selected from the Guidu fu section of Qingyi River from January 2011 to June 2016. The ARMA(p, q) model was established with the measured concentrations of the five water quality indicators as sample data from January 2011 to December 2015, and the significance test of the model was carried out. Then, the concentration values of five water quality indicators from January to June 2016 were predicted and verified. The results show that the model is effective and the fitting effect is well. This section takes the concentration of COD_{Cr} indicator as an example to describe the modeling process in detail.

3.1. Stationarity test for time series of water quality measured data

The stationary series is the basis of time series analysis. Many stationarity testing methods in time series analysis assume that data samples are from stationary and ergodic stochastic processes, that is, their expectations, variances and auto-covariance functions do not change with time, and can replace the overall average with time average. In this paper, the stationarity of time series $\{X_t\}$ of COD_{Cr} concentration in Qingyi River were tested by Augmented Dickey Fuller (ADF) Test [20] and Phillips-Perron (PP) test [21].

Table 1. ADF and PP test results of time series $\{X_t\}$

| Test methods | Critical value | T statistics | Probability |
|--------------|----------------|--------------|-------------|
| ADF test | ADF value | -3.636710 | 0.0289 |
| | 1% level | -3.546099 | |
| | 5% level | -2.911730 | |
| PP test | 10% level | -2.593551 | |
| | PP value | -3.735216 | 0.0372 |
| | 1% level | -3.546099 | |
| | 5% level | -2.911730 | |
| | 10% level | -2.593551 | |

As shown in Table 1, the statistics of ADF and PP methods of COD_{Cr} concentration in the time series are -3.636710 and -3.735216, respectively, which are less than the critical values of 1% level, 5% level and 10% level. Therefore, the original hypothesis is rejected and the series passed the test at 99% confidence level, that is, the series are stationary series.

3.2. White noise test for time series of water quality measured data

The white noise test for time series of water quality measured data is a necessary step in the modeling of water quality prediction model of Qingyi River. White noise series means that there is no correlation between the series itself, and the series composed of historical water quality measured data has no significance for future water quality prediction. That is to say, the white noise series without any information cannot be used to establish time series model. Only time series without white noise test can be used for modeling. This work used the Q statistic to test whether the time series $\{X_t\}$ is a white noise series. Given the significance level α , when the P values (probability) of the test statistic is less than the significance level α , the time series of the water quality indicator is considered to be a non-white noise series. The test results as shown in Fig.1.

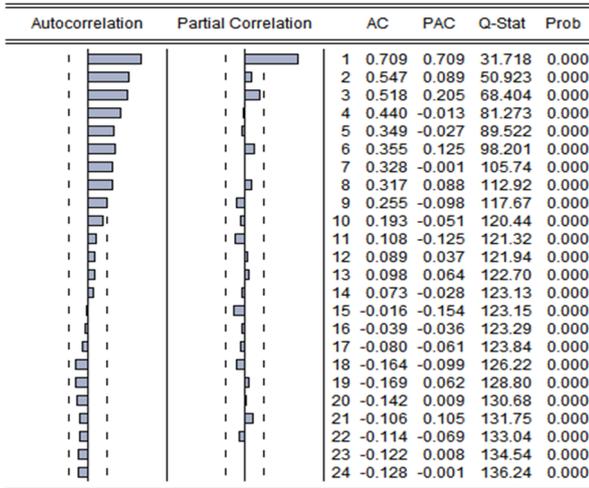


Fig.1. The white noise test results of COD_{Cr} time series {X_t}

As can be seen from Fig.1, the *P* values of COD_{Cr} white noise test is less than the significant level ($\alpha=0.05$), so the original assumption that the series {X_t} is white noise is rejected, that is, the series {X_t} is not white noise series.

3.3. Coefficient estimation of water quality prediction model

The order *p* and *q* of the ARMA (*p*, *q*) model can be determined by observing the autocorrelation coefficient and partial correlation coefficient of the water quality measured data time series {X_t} after stationarity test and determination of the non-white noise series.

As can be seen from Fig.1, both autocorrelation and partial correlation coefficient diagrams of series are trailing. Therefore, ARMA (*p*, *q*) model should be selected to determine the optimal model order after determining multiple sets of *p* and *q* values and validity test of the model and significance test of the parameters through AIC criterion or BIC criterion [22]. The best model is ARMA (2, 1) by calculating on Eviews 8.0 software many times, and the estimated coefficients as shown in Table 2. Then, by introducing the fitted model coefficients into equation (1), the final expression of the model can be obtained as follows:

$$X_t = 1.578X_{t-1} - 0.575X_{t-2} + \varepsilon_t - 0.975\varepsilon_{t-1} \quad (2)$$

Table 2. The result of coefficient estimates

| Variables | Coefficients | T statistics | Probability |
|-----------|--------------|--------------|-------------|
| AR(1) | 1.578 | 14.612 | 0.0000 |
| AR(2) | -0.575 | -5.272 | 0.0000 |
| MA(1) | -0.975 | -50.815 | 0.0000 |

3.4. Testing of water quality prediction model

Before the model can be used for prediction, the time series of water quality measured data must be tested, including the white noise test of residual. Only the residual series is white noise to prove that the model has extracted all information of the original time series {X_t}. The white noise of the residual can be obtained by the

above white noise test method. Fig.2 shows the autocorrelation and partial correlation of residuals.

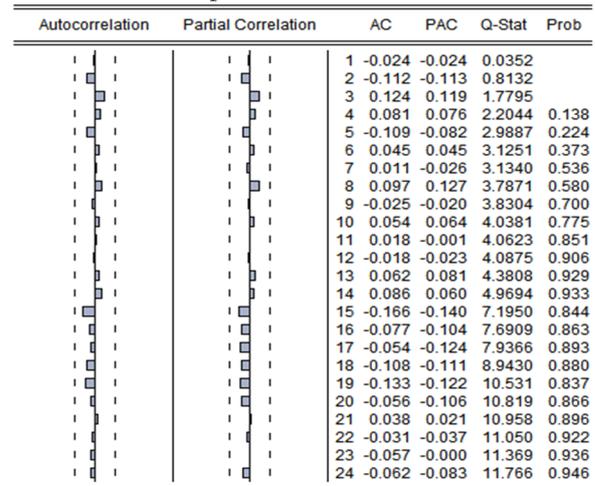


Fig.2. The autocorrelation and partial correlation of residual

As can be seen from Fig.2, the autocorrelation coefficients and partial correlation of the residuals fall within the range of two standard deviations, so it was considered that the residuals are white noise series. Combined with Q statistics and corresponding *P* values, it can be found that *P* values are greater than 0.05, that is to say, the residual is a white noise series, and the model is effective. In the coefficient estimation of Table 2, the *P* values corresponding to all t-test results is less than 0.05, so the parameter is significant. In conclusion, the model ARMA (2, 1) is effective.

The actual COD_{Cr} concentration data, the fitting result of the model and the residual of Guidu fu section from January 2011 to December 2015 are shown in Fig.3, in which the horizontal axis is time (2011-2015) and the vertical axis is COD_{Cr} concentration. It can be seen that the fluctuations between the fitted and measured values of the model are consistent and the residuals are small, which shows that the fitting effect of the model is well.

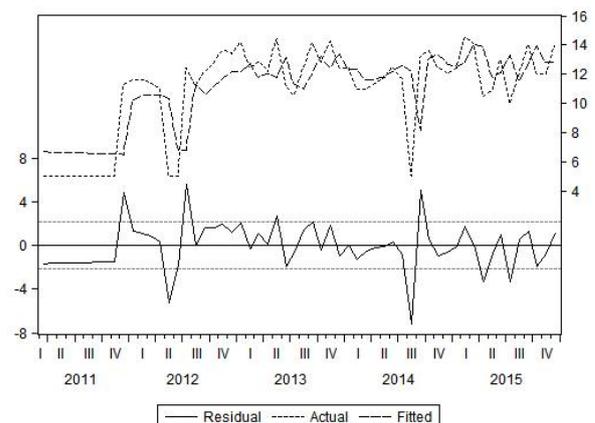


Fig.3 The fitting effect of time series {X_t}

4 Verification and analysis of water quality prediction results

Based on the above methods, 300 data of five water quality indicators of DO, BOD₅, COD_{Cr}, NH₃-N and TN for Qingyi River from January 2011 to December 2015 were used to establish the corresponding water quality prediction model and predicted the concentrations and the change trend of five water quality indicators from January to June 2016. The prediction results as shown in Fig. 4-8.

4.1. DO prediction results

Dissolved oxygen (DO) means the amount of molecular oxygen dissolved in water, expressed in milligrams of oxygen per liter of water. It is an important indicator for measuring water pollution and an important condition for water body to achieve self-purification. The higher the degree of contamination of the water, the less DO concentration. The predicted results of DO are shown in Fig. 4.

As shown in Fig. 4, the DO concentration of Qingyi River from 2011 to 2015 shows a relatively stable characteristic, which is above 6mg/L and meets the requirement of the environment quality standard for surface water class II and above. However, in 2016, DO concentration began to decrease, and the concentration are lower than 6 mg/L from April. Water quality category reduced from class II to class III. According to the data of water quality survey, the overall water quality of the the Guidu fu section of Qingyi River decreased in the first half of 2016 compared with the same period in 2015. From January to June, the water temperature gradually increased, which also led to the decrease of DO concentration in water. The predictive value are compared with the measured value in 2016, and the overall trend is consistent, showing the accuracy of ARMA prediction model.

4.2. BOD₅ prediction results

Biochemical oxygen demand (BOD₅) refers to the amount of oxygen consumed by microorganisms when they decompose organic compounds in water. It is a comprehensive indicator of the concentration of aerobic pollutants such as organic compounds that can be decomposed by microorganisms in water. The prediction results are shown in Fig. 5.

As can be seen from Fig. 5, the concentration of BOD₅ in Qingyi River is within the standard of Class I from 2011 to 2016. However, it has been rising slowly since December 2015, which indicates that the water quality has a downward trend.

4.3. COD_{Cr} prediction results

In general, chemical oxygen demand (COD_{Cr}) is used to represent the total amount of organic matter in water. The predicted results are shown in Figure 6.

As shown in Fig. 6, the concentration of COD_{Cr} in the water body from 2011 to 2015 conforms to the standard of Class II. and it has an obvious upward trend since December 2015. In June 2016, the concentration of COD_{Cr} is more than 15mg/L, which belongs to the

standard of Class III. As an important indicator of organic pollutant content in water, the concentration of COD_{Cr} increased, which indicated that the influence of organic pollutants on Qingyi River water increased. From the Fig. 6, it can be seen that the fitting effect between the predictive value and the measured value is well, which can better reflect the actual change trend of chemical oxygen demand.

4.4. NH₃-N prediction results

Ammonia nitrogen (NH₃-N) is a product of organic matter degradation by microorganisms in water, and it is a kind of inorganic nutrients. If the concentration of NH₃-N in water body is too high, it can lead to eutrophication of water body, which is harmful to aquatic organisms. Therefore, the prediction of NH₃-N concentration in water is particularly important. The prediction results are shown in Fig. 7.

As shown in Fig. 7, the concentration of NH₃-N fluctuated greatly from 2011 to 2014 in the Guidu fu section of Qingyi River. From 2011 to 2015, the change of concentration was in line with the standard of Class III, and there was a trend of decreasing year by year. The concentration of NH₃-N from January to June 2016 also showed a downward trend, which began to be lower than 0.15 mg/L in June and reaching the standard of class I. Except for some errors in April, the results of the overall simulation are in good agreement with the measured values, which can reflect the change trend of short-term water quality.

4.5. TN prediction results

Total nitrogen (TN) refers to the total amount of various forms of nitrogen in water, which is one of the important indicators used to measure water quality. The prediction of TN concentration is helpful to understand the status of water contaminated by nutrients. The predicted results are shown in Fig. 8.

It can be seen from Fig. 8 that the concentration of TN in Qingyi River maintained basically stable trend from 2011 to 2014, which belonged to the water quality standard of Class III. But from June 2015, the concentration has generally increased, and it belongs to water quality standard of Class IV. Since 2016, the concentration of TN continues to rise, which may increase further, indicating that the pollution of water by nutrients will increase and should be further controlled.

4.6. Result analysis

From Fig. 4-8, it can be seen that the ARMA model can efficiently simulate the change of water quality in Qingyi River from 2011 to 2015, which is consistent with the actual change trend. It can predict the concentration changes of DO, BOD₅, COD_{Cr}, NH₃-N and TN indicators. In comparison with the measured values in 2016, all the indicators can be reasonably simulated and predicted. And especially, the short-term prediction is well. Because the abnormal operation of sewage pipe network and

centralized treatment facilities in some industrial parks along the Qingyi River in the first half of 2016, and the direct discharge of sewage into the river, which led to less deviations in the prediction results of the indicators. Therefore, the influence of external factors should also be

taken into account in practical application in order to improve the accuracy of prediction. Generally, the prediction results are reasonable and can provide a scientific reference for water quality control in drainage basin.

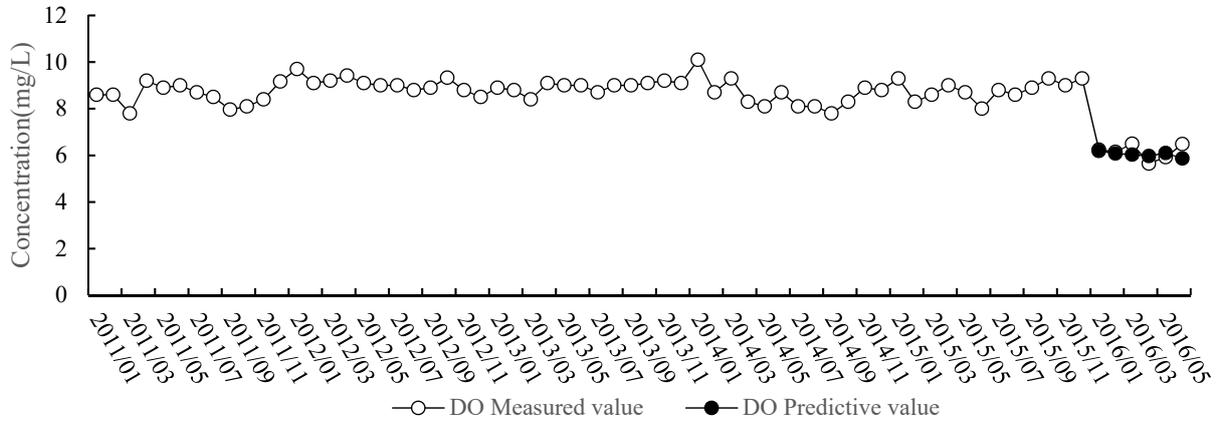


Fig.4. Changing tendency and predictive value of DO concentration from 2011 to 2016

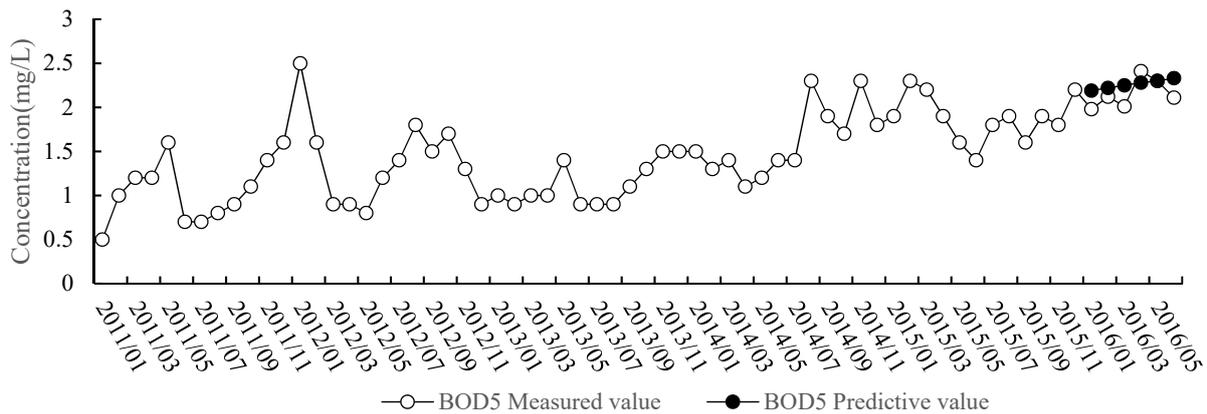


Fig.5. Changing tendency and predictive value of BOD₅ concentration from 2011 to 2016

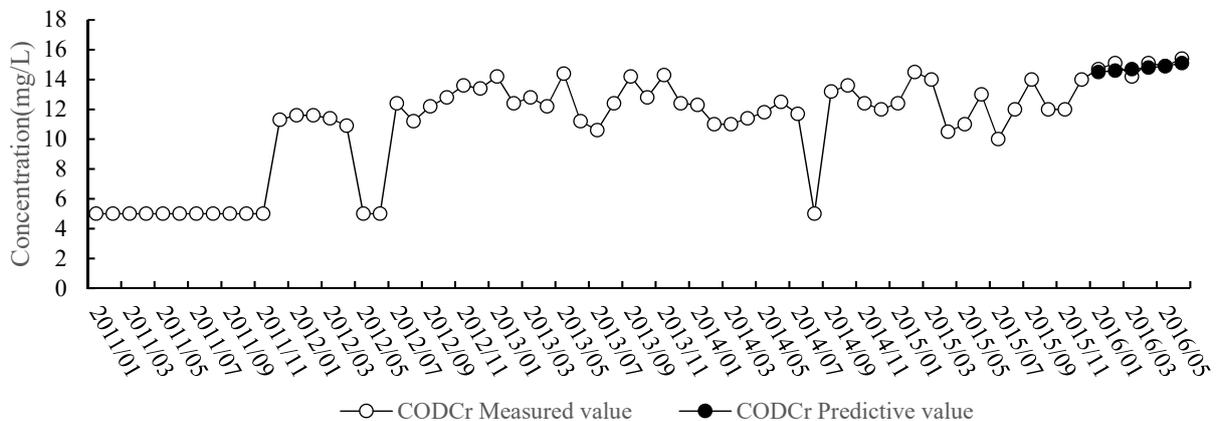


Fig.6. Changing tendency and predictive value of COD_{Cr} concentration from 2011 to 2016

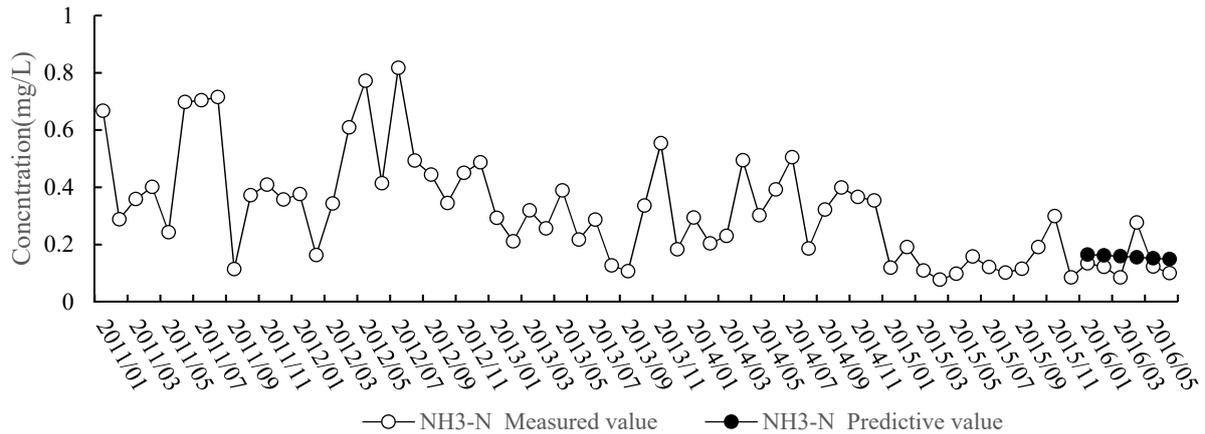


Fig.7 Changing tendency and predictive value of NH₃-N concentration from 2011 to 2016

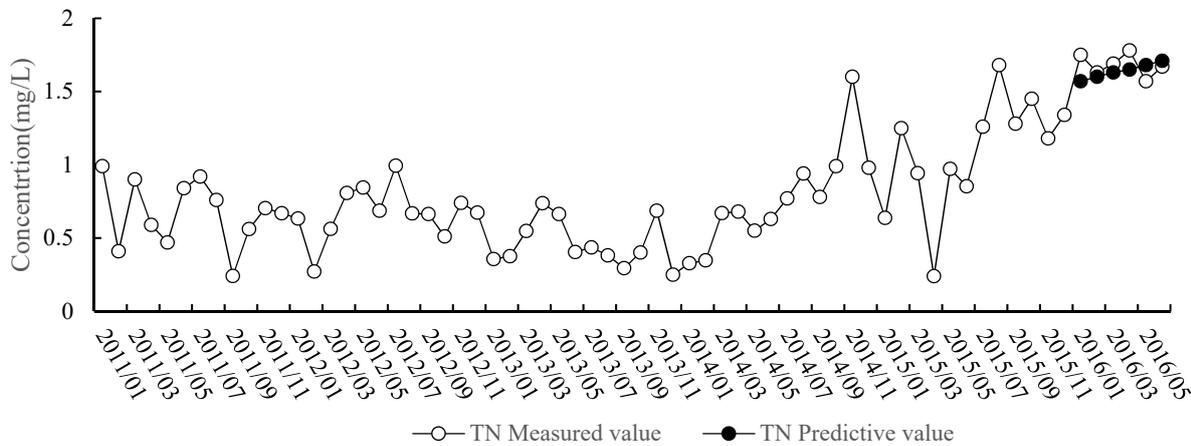


Fig.8. Changing tendency and predictive value of TN concentration from 2011 to 2016

5 Conclusions

The ARMA model based on the time series of water quality indicators can be used to simulate and predict water quality indicators accurately. The DO, BOD₅, COD_{Cr}, NH₃-N and TN indicators have achieved good results in the prediction of 2016, showing the accuracy of the model predictions. It can grasp the short-term trend of pollutants and provide reliable basis for water environment planning and management in drainage basin.

The prediction and analysis of five indicators of Guidu fu section in Qingyi River shows that the concentrations of BOD₅, COD_{Cr} and TN indicators will increase in the next months. It result the reduction of DO concentration, which indicates that controlling the concentrations of BOD₅, COD_{Cr} and TN is still an important task of water quality management and control in Qingyi River.

Acknowledgement

This study was supported by the National Natural Science Foundation of China (Grant No.51779211, Grant No.51209178), and the Sichuan Science and Technology Program (Grant No. 2019YJ0233, Grant No.18ZDYF1829). We deeply appreciate the editors' and anonymous reviewers' significant comments and suggestions for adding the quality of the study.

References

1. C. Zhang, L. Y. Wang, Y. Gao, X. P. Gao, J. Saf. Environ, **11**, 149-152 (2011)
2. Q. C. Guo, Z. F. He, Inf. Technol, **12**, 65- 67 (2012)
3. N. Li, L. C. Wang, G. Xie, Y. N. Wei, H. Neng, Environ. Sci. Technol, **35**, 195-199 (2012)
4. Z. Lu, Y. Li, X. Ma, D.Y. Zhang, B. L. Huang, M. Zhang, L. Zhu, C. Liu, Environ. Sci. Technol, **35**, 174-178 (2012)
5. Y. Z. Hu, Chongqing University, Chongqing (2010)
6. R. Zhang, B. S. Sun, Y. L. Wang, Z. J. Si, X. F. Li, T. W. Yan, Chin. J. Environ. Eng, **7**, 3089-3093 (2013)
7. Z. Y. Liang, Y. Liu, H. Sheng, Y. H. Yong, H. C. Guo, L. Zhao, B. He, J. Environ. Sci, **34**, 754-762

(2014)

8. Y. H. Hu, Z. f. Wu, D.Q. Li, M. N. Zhuo, *Ecol. Environ*, **15**, 94-98 (2006)
9. H. Yu, B. S. Sun, Y. N. Li, Y. Zhang, G. S. Qi, *China. Environ. Sci*, **34**, 810-816 (2014)
10. Q. C. Guo, X. R. Zhao, *Comput. Digit. Eng*, **41**, 683-685 (2013)
11. B. Jiang, L. Qiu, Z. H. Zheng, *Agr. Technol*, **36**, 68-70 (2016)
12. D. Olivier, C. Amelie, *Cliom*, **5**, 79-100 (2011)
13. L. Kevin, P. Vinhthuy, *BMC Bioinf*, **15**, 24-25 (2014)
14. S. Shervin, K. Mohammad, M. Salem, M. Goranneviss, P. Khorshid, *J. Fus. Energy*, **29**, 371-375 (2010)
15. G. Valeriy, S. Olga, D. Kristina, *Sign. Imag. Anal. Biomed. Life. Sci*, **823**, 107-126 (2015)
16. C. B. Huang, Y. Y. Dian, Z. X. Zhou, D. Wang, R. D. Chen, *J. Remot. Sens*, **19**, 657-668 (2015)
17. X. B. Huang, H. B. Li, Y. C. Zhu, Y. X. Wang, X. X. Zheng, Y. G. Wang, *High. Volt. Eng*, **43**, 1943-1949 (2017)
18. D. Li, C. Zhou, P. F. Yuan, M. Wang, S. Jiang, J. Zhang, *Environ. Sci. Technol*, **35**, 184-188 (2012)
19. L. Ai, East China Normal University, Shanghai (2010)
20. Z. Chen. *J. Guangzhou Univ (Nat. Sci. Ed)*, **7**, 5-10 (2008)
21. L. Fang, W. X. Zou, *J. Quant. Tech. Econ*, **1**, 151-160 (2007)
22. Y. D. Zhou, H. Q. Wang, W. Y. Lu, *Time Series Analysis and Application* (2015)