

Application of chaotic prediction model based on wavelet transform on water quality prediction

L Zhang¹, Z H Zou^{1,2} and Y F Zhao¹

¹School of Economics and Management, Beihang University, Beijing 100191, China

E-mail: zouzhihong@buaa.edu.cn

Abstract. Dissolved oxygen (DO) is closely related to water self-purification capacity. In order to better forecast its concentration, the chaotic prediction model, based on the wavelet transform, is proposed and applied to a certain monitoring section of the Mentougou area of the Haihe River Basin. The result is compared with the simple application of the chaotic prediction model. The study indicates that the new model aligns better with the real data and has a higher accuracy. Therefore, it will provide significant decision support for water protection and water environment treatment.

1. Introduction

The first paragraph after a heading is not indented (Bodytext style). Along with economic development and improvements in people's standard of living, bodies of water are facing more and more serious pollution and destruction. However, polluted water has certain self-purification abilities. It can reduce the concentration of pollutants to normal levels by physical, chemical and biological functions. Dissolved oxygen (DO) is the basis for studying water bodies' self-purification abilities. Consequently, research concerning trends of DO in water and predicting its concentration has substantial theoretical and realistic significance. Because of the complexity of the water environment and the limitation of single methods, combined forecasting methods have been widely used, such as the radial basis function (RBF) water quality prediction model based on particle swarm optimization, the BP neural network prediction model based on ant-colony optimization and the gray and neural network combination model [1-3]. These methods led to accurate predictions of water quality trends and concentrations of pollutants.

Water environment systems are quite complex. The DO value could be affected by plenty of factors. As a signal processing technology, the wavelet transform has been widely used recently. The decomposition and reconstruction of original signals can lead to better extraction of the feature information of the original signal [4]. The chaotic prediction model based on the wavelet transform is established by combining the wavelet transform and the chaotic prediction model, which has simple calculations and high precision. Comparing it with the chaotic prediction model proves the validity of the new model.

2. The chaotic prediction model based on wavelet transform

2.1. Preparing the new file with

The first paragraph after a heading is not indented (Bodytext style). Chaos comes from nonlinear dynamic systems. It is a kind of pseudo randomness showed by deterministic systems. Thus,



seemingly random phenomena produced in chaotic systems are predictable. Chaos is a kind of resilience of periodic motion produced by deterministic systems. This motion is very sensitive to initial conditions, which makes it suitable for short-term forecasting [5]. Wavelet analysis has developed in the past 20 years. As an analytic method of wavelet analysis, the wavelet transform makes up for the defects of the discrete Fourier transform with a fixed window and no discrete orthogonal basis. The multi-scale discrete wavelet transform can decompose the original water quality time series on different temporal resolutions [6] and gain high frequency and low frequency sequences with more single frequencies. This can strengthen the chaotic features of sequences and make them more predictable.

The steps of the chaotic prediction model based on the wavelet transform are as follows:

- Perform wavelet decomposition of the original time series data based on the à Troun algorithm.

Many algorithms of the wavelet transform require down sampling of the original sequences. The resulted approximation and detail sequences contain only half of the points of the original sequences, which decreases prediction accuracy. With its easy calculations and translational invariance, the à Troun algorithm is conducive for extracting the feature information of the original sequences. The choice of the wavelet function is an important factor affecting the accuracy of the prediction. After repeated trial calculations, db2 of the Daubechies wavelet group was picked for this model. The algorithm decomposition process is as follows:

Let the original sequence be $c_0(t)$, $t=0,1,2, \dots, n-1$. N is the length of the sequence. Then

$$c_j(t) = \sum_l h(l) c_{j-1}(t+2^{j-1}l), j=1,2,\dots,k \quad (1)$$

$$w_j(t) = c_{j-1}(t) - c_j(t), j=1,2,\dots,k \quad (2)$$

where k is the decomposition scale, $h(l)$ is the low-pass filter and $c_j(t)$ and $w_j(t)$ are the scale coefficients and wavelet coefficients under scale j , respectively.

- Refactor the decomposed scale coefficients and wavelet coefficients.

The scale coefficients and wavelet coefficients can be refactored to get original time sequences, namely:

$$c_0(t) = \sum_{j=1}^k w_j(t) + c_k(t) \quad (3)$$

where ck is an approximate sequence, and $\sum w_j$ is a detail sequence, denoted by dk . With a single frequency component and simpler sequence structure, these two sequences are more predictable.

- Test the chaos characteristics for ck and dk . Establish the chaotic model to predict sequences with chaos characteristics. Use the neural network method to predict sequences without chaos characteristics. Add the two predicted outcomes together to get the final prediction result.

The chaos characteristics of the chaotic time sequences have to be tested before prediction, which mainly includes phase-space reconstruction and the calculation of the maximum Lyapunov exponent [7]. The chaotic forecasting model can predict the time series with chaos characteristics. The chaos weighted first-order local prediction method believes that the prediction center in the phase space has approximately a first-order linear relationship with its next point. This relationship is approximately the same as the relationship between point F , which is nearest to the prediction center by Euclidean distance and its next point, namely: $Y(q+1)=a+bY(q)$. The specific steps are as follows:

a. Reconstruct the phase space and determine the embedding delay time[8] and embedding dimension [9].

b. Calculate the F points nearest the final point and their weights.

$$p_j = \frac{\exp(-(d_j - d_{\min}))}{\sum_{j=1}^n \exp(-(d_j - d_{\min}))} \quad (4)$$

where $j=1,2,\dots,n$. d_j is the Euclidean distance between the first j point and the final point. d_{min} is the minimum among d_i .

c. Use the least squares method to fit the first-order linear relationship between point F and its next point, and then obtain the corresponding coefficient fitting coefficient, namely:

$$y(N_j + 1) = a_j e + b_j y(N_j) \quad (5)$$

where $y(N_j+1)$ is the next point of $y(N_j)$ and a_j , b_j are fitting coefficients. $eT=(1,1,\dots,1)F$.

d. To predict the point, weight the gained fitting coefficients and substitute to the prediction center and points waiting for prediction.

e. Separate the prediction of the original sequence from the point to be predicted.

3. Case study

3.1. Date source

The research object is the average weekly monitoring value of DO for 258 weeks between 2010 and 2014 of a monitoring section in the Mentougou section of the Haihe river Basin. All data comes from China's environmental monitoring network.

3.2. Establishment of prediction model

The model is established based on the DO concentrations of the former 256 weeks to predict concentrations in the last two weeks. The DO concentration curve of the former 256 weeks in this district is shown in Figure 1.

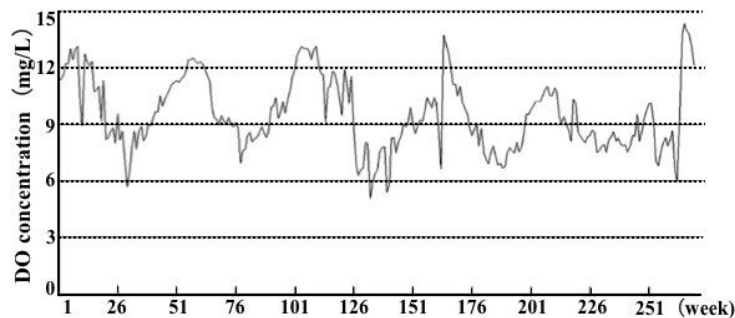


Figure 1. DO concentration for 256 weeks of a monitoring section of Haihe.

When using the auto-correlation method to calculate the embedding time delay τ of DO concentration sequence, we calculate the auto-correlation coefficient under different time delays. When the auto-correlation coefficient becomes negative for the first time, the corresponding time delay is the result that we desire. According to Table 1, τ equals 13.

Table 1. Auto-correlation coefficient changing with the embedding time delay.

T	1	2	3	4	5	6	7	8	9	10	11	12	13
R	0.85	0.76	0.71	0.66	0.62	0.60	0.66	0.47	0.36	0.26	0.17	0.09	-0.02

The cao algorithm is applied to calculate changes of $E1(m)$ and $E2(m)$ along with the embedding dimensions (see Figure 2) and $m \geq 3$.

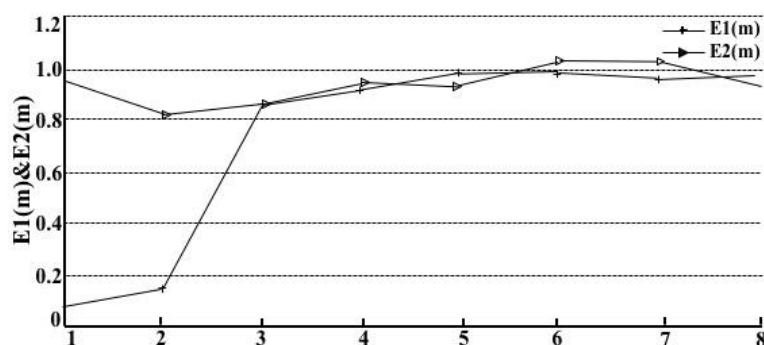


Figure 2. Changes of $E1(m)$ and $E2(m)$ along with embedding dimensions.

The DO concentration sequence has an obvious annually periodic feature, namely the average orbital period $\tau p \approx 52$ weeks, which equals the result of FFT. This provides important additional information for reasonably determining time delay and dimensions. Embedding dimensions separately determined can hardly consider the combination of both. Given the principle of $\tau p \approx \tau w = \tau(m-1)$, we chose $\tau=13$, $m=5$, so $\tau w = 52$. This was plugged into the minimum data account method to calculate the maximum Lyapunov exponent of the DO concentration sequence. The result was 0.0025; the sequence was a chaotic sequence. Meantime, $E2(m)$ changed along with m , which also shows the chaotic characters of the sequence.

The appropriate decomposition scale should be selected for wavelet decomposition. The standard for choosing the decomposition level is minimizing the prediction error. Normally, the decomposition level is 3~5. To be prudent, a decomposition level of 1~5 was calculated, and the results showed that 3 is the best choice.

Next, we refactored the decomposed scale coefficients and the wavelet coefficients to get an approximating sequence and a detail sequence (see Figure 3).

Then, we selected parameters for the approximating sequence and the detail sequence and calculated the maximum Lyapunov exponent (see Table 2). Results showed that the maximum Lyapunov exponent for both sequences were positive, which indicates chaotic features. Thus, the chaotic prediction model can be applied to these two sequences.

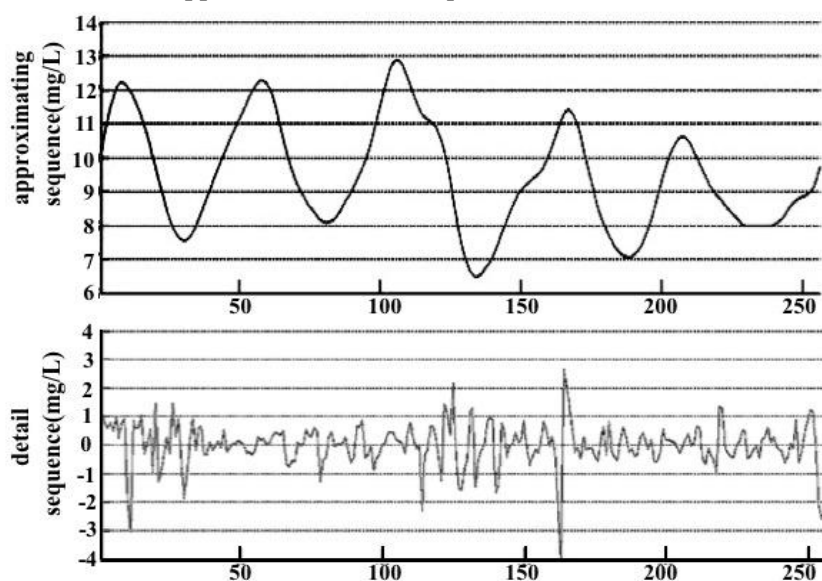


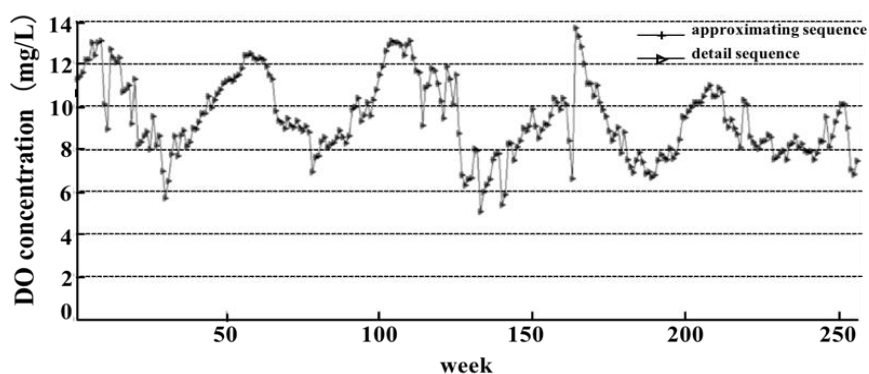
Figure 3. Approximate sequence and detail sequence after wavelet decomposition and refactoring.

Table 2. Parameters of approximating and detail sequences based on maximum Lyapunov exponent.

Research Object	τ	m	Maximum Lyapunov exponent
Approximating Sequence ck	13	5	0.0030
Detail Sequence dk	2	4	0.0017

3.3. Model fitting and prediction

3.3.1. Model fitting. The combined prediction model was applied to fit the data of the previous 256 weeks and the predicted DO concentration of the upcoming 2 weeks. The two results were added together to get the fitting result of the combined model (see Figure 4). According to the figure, the chaotic prediction model based on the wavelet transform was the best fit to the trend.

**Figure 4.** Model fitting figure.

3.3.2. Model prediction. The nonlinear system is very sensitive to initial conditions, which makes it suitable for short-term forecasting. The chaotic prediction model (model 1), the artificial neural network prediction model (model 2) and the chaotic prediction model based on the wavelet transform (model 3) were applied to forecast DO concentration in the next two weeks. The MRE (mean relative error) and RMSE (root mean square error) were used to evaluate the models (see Table 3). The RMSE is very sensitive to the maximum and minimum error, which enables it effectively reflect the accuracy of the prediction results.

Table 3. Comparison of two prediction models.

Model	Model 1 (Week 257)	Model 1 (Week 258)	Model 2 (Week257)	Model 2 (Week258)	Model 3 (Week 257)	Model 3 (Week 258)
Predictor	7.670	7.684	8.196	8.873	8.049	8.829
Real value	8.02	8.27	8.02	8.27	8.02	8.27
Errors	0.349	0.585	0.176	0.603	0.029	0.559
MRE	0.043	0.071	0.022	0.073	0.004	0.068
Average MRE	0.0572		0.0474		0.0356	
RMSE	0.4824		0.0444		0.3956	

According to Table 3, the average MRE and RMSE of the combined prediction model is lower than Model 1 and Model 2, implying that the combined model has better accuracy. With the same data amount, approximating and detail sequence can easily find points near to the center by Euclidean distance, which makes them better predictors. A more accurate prediction of DO concentration can provide better reference for water self-purification abilities and water conservation work in this district.

Conclusion

The water environment is very complicated, and water indicators could show different sequence features. The chaotic prediction model based on the wavelet transform can be applied to the short-term prediction of the fluctuation index with large-scale samples. Based on the volatility of the DO concentration and making use of the good stripping effect of the wavelet transform, the low frequency and high frequency data sequences were gained. The chaotic prediction model based on the wavelet transform was established by combining the wavelet transform and the chaotic prediction model, which involves simple calculations and high precision. Through the comparison of average MRE, the combined model has higher accuracy and can provide a more solid basis for water protection and water environment treatment.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (Nos. 51478025).

References

- [1] Bahaa M K, Ayman G A, Hussein K and Ashraf E.-S 2012 Application of artificial neural networks for the prediction of water quality variables in the Nile delta *J. Water Resour. Protection*
- [2] Cao L Y 1997 Practical method for determining the minimum embedding dimension of a scalar time series *Physical D* **110**: 43-50
- [3] Chen J J, Mao H X, Li P C, et al. 2014 A wavelet autoregressive model and its application to water quality forecast *Hydro-Science and Engineering* **2**: 95-99
- [4] Chen M and Ye X Z 2008 Research on the chaotic time series decision method *Information Technology* **2008**(6): 23-26
- [5] Cheng S F, Cheng X H, Yang L 2014 Application of wavelet neural network with improved particle swarm optimization algorithm in power transformer fault diagnosis *Power. Syst. Prot. Control* **42**
- [6] Ding J, Wang W S and Zhao Y L 2003 Characteristics of daily flow variation in the Yangtze River ,1, optimum determination of delay time for reconstruction of a phase space *Advances In Water Science* **14**(4): 407-11
- [7] Ding J, Wang W S and Zhao Y L 2003 Characteristics of daily flow variation in the Yangtze River ,2, optimum determination of embedding dimension for reconstruction of a phase space *Advances In Water Science* **14**(4):412-16
- [8] Dogan E, Sengorur B and Koklu R 2009 Modeling biological oxygen demand of the Melen River in Turkey using an artificial neural network technique *J. Environ. Manage.* **90** 1229–35
- [9] Gazzaz N M, Yusoff M K, Aris A Z, Juahir H and Ramli M F 2012 Artificial neural network modeling of the water quality index for Kinta River (Malaysia) using water quality variables as predictors *Mar. Pollut. Bull.* **64** 2409–20
- [10] Govindaraju R S 2000 Artificial neural networks in hydrology. II: hydrologic applications *J. Hydrol. Eng.* **5** 124–37
- [11] Liu D J and Zou Z H 2011 Applications of Gray Forecast Model combined with Artificial Neural Networks Model to water quality forecast *Systems Engineering* **29**(9): 105–09
- [12] Mohanmad J A and Mohanmad R K 2015 Development of wavelet-ANN models to predict water quality parameters in Hilo Bay, Pacific Ocean *Mar. Pollut. Bull.* **98**: 171-78

- [13] Musavi S and Golabi M 2008 Application of artificial neural networks in the river water quality modeling: Karoon River, Iran *J. Appl. Sci.* **8(12)** : 2324–28
- [14] Nian H, Gu S M and Li X 2014 Back propagation neural network based on ant colony optimization algorithm for the prediction of water quality *J. Minnan Normal University(Nat. Sci.)* **83**: 62-67
- [15] Okkan U 2012 Wavelet neural network model for reservoir inflow prediction *Sci. Iranica* **19**: 1445–55
- [16] Palani S, Tklich P, Balasubramanian R and Palanichamy J 2011 ANN application for prediction of atmospheric nitrogen deposition to aquatic ecosystems *Mar. Pollut. Bull.* **62**: 1198–206
- [17] Qi J X and Chen J 2014 Radial basis function (RBF) water quality prediction model based on particle swarm optimization *Gansu Water Resour. Hydropower Technol.* **50(5)**: 1-4
- [18] Xu L Q and Liu S Y 2013 Study of short-term water quality prediction model based on wavelet neural network *Math. Comput. Model.* **58(3)**: 807-13
- [19] Yao S P, Hu C Z and Zheng L 2006 Chaotic time sequence prediction based on wavelet decomposition and refactoring *J. Projectiles, Rockets, Missiles and Guidance* **26(2)**: 1133-35
- [20] Zhang F, Dai H and Tang D 2014 A conjunction method of wavelet transform- particle swarm optimization-support vector machine for stream flow forecasting *J. Appl. Math.* **10**: 1-10
- [21] Zou Z H and Feng W 2008 Add-weighted one-rank local-region method based on phase space reconstruction and application to dissolved oxygen prediction *Math. Practice Theory* **38(13)**: 70-75