Journal of Computing and Electronic Information Management

ISSN: 2413-1660

Grey-Multivariate Mixed Time Series Analysis Method in Application of Cyanobacteria Growth Modeling and Water Bloom Prediction

Li Wang*, Tianrui Zhang, Xiaoyi Wang, Xuebo Jin, Jiping Xu, Huiyan Zhang, Jiabin Yu, Qian Sun, Tingli Su

School of Computer and Information Engineering, Beijing Technology and Business University, Beijing 100048, China

Abstract: In the work of water bloom control, water bloom prediction has been a problem. Because the mechanism of algal bloom formation is different in each stage, it is difficult to predict the whole formation process of water bloom by using a certain model. Therefore, algae growth stage of the longest duration and obvious change law is taken as the research object in the paper and the multivariate time series analysis technique is adopted to conduct time model analysis for characteristic factors in the growth stage of algae. Characteristic factors of affecting the algae growth are used as the input of model and characteristic factors of representation are used as the output of model, which is combined with the trend, periodicity and random variation of algae growth stage to establish multivariate mixed time series model. At the same time, the grey theory is added to the model to further improve the prediction algorithm of model. Finally, water bloom multi-factor prediction in algae growth stages based on multidimensional time series analysis and grey theory is proposed. Feature factors can be comprehensively analyzed and predicted and the prediction accuracy of the model is improved by the method. The method of this paper and the traditional method are used to build model with the data water bloom characteristics in Taihu Lake, Jiangsu. The conclusion is that the prediction method is more consistent with the measured results and absolute value of the average prediction error is smaller.

Keywords: water bloom, model, grey theory, characteristics, prediction

1 Introduction

Algal bloom is a kind of phenomenon that algal reproduction and accumulation reaches a certain concentration in eutrophic water. The harm of water bloom is very big. It not

only destroys the structure and function of aquatic ecosystem, but also the sustainable development and utilization of water resources are threatened.

Water bloom prediction has been a difficult point in the prevention and control of water bloom. Because of the complexity of ecological processes, the impact of human activities ,water quality information measurement and reporting, receiving and so on, it is difficult to establish the ecological model of water bloom directly [1-6].With the further research of water bloom prediction, some scholars have made some progress by trying soft sensor [7-9] to establish a precise mathematical model to predict the formation of water bloom [10-15].However, the prediction method of water bloom still has the problem of low precision. From the point of view of mathematical analysis, the time series of the characteristic factors of water bloom is regarded as the multiple period stable sequence and the multivariate time series analysis method [16-20] is used to predict in the literature [21], which improves the prediction accuracy and expands the model of water bloom prediction

The whole production of algal bloom is a complex ecological processes. It consists of 4 stages: germination, growth, outbreak and death. However, the change mechanism of each stage is not exactly the same. It is not reasonable to predict the whole formation process of water bloom if only a model is used. Therefore, algal growth stage of the longest duration and obvious change law is used to establish the multivariate time series model by the characteristic factors of water bloom. The algal growth phase is described and predicted by this model to solve the following problems.

(1) During the formation of the water bloom, not only some characterization factors such as the change of chlorophyll concentration and algae density can directly represent the formation process of water bloom but also the change of influencing factor also has the reference value. Therefore, the prediction method of all characteristic factors, including the influence factors, needs to be studied.

(2) In the algae growth stage, the multivariate time series of the characteristic factors of water bloom not only have random and periodic changes but also each characteristic factor also has the varying degree of monotonous change tendency. When the tendency of this monotonous is accumulated to a certain degree, it will enter the stage of water bloom. Therefore, when the time sequence characteristic factors are described, a feature of random and periodic changes need to be considered and the trend change and interactive effects of multiple characteristic factors must be taken into account.

(3) The trend, periodicity and randomness of data can be fully excavated by time series analysis method. However, the accuracy of the method depends on the limitation of the data length. In the time series model, it is difficult to guarantee the accuracy of the prediction if the original data is directly used to predict the future. This method

reduces the credibility of the results of water bloom prediction. Therefore, a method needs be studied, which applies to the combination of low data prediction technology and time series model.

2. Multivariate Mixed Time Series Modeling of Characteristic Factors

2.1 Determine the Timing Structure of the Characteristic Factors

In the stage of algal growth, the changes of trend, random and periodic are presented by the multivariate time series of characteristic factors of water bloom in nature. However, the algae growth mechanism does not change with time, so the characteristic factors of water algae growth stage are multivariate variance stationary time series. The characteristic factor vector (Y_t) is decomposed into trend term (F_t), periodic term (C_t) and random term (R_t).

$$\boldsymbol{Y}_t = \boldsymbol{F}_t + \boldsymbol{C}_t + \boldsymbol{R}_t$$

$$\boldsymbol{Y}_t = \begin{pmatrix} y_{1t} \\ y_{2t} \\ \vdots \\ y_{nt} \end{pmatrix}, \boldsymbol{F}_t = \begin{pmatrix} f_{1t} \\ f_{2t} \\ \vdots \\ f_{nt} \end{pmatrix}, \boldsymbol{C}_t = \begin{pmatrix} c_{1t} \\ c_{2t} \\ \vdots \\ c_{nt} \end{pmatrix}, \boldsymbol{R}_t = \begin{pmatrix} r_{1t} \\ r_{2t} \\ \vdots \\ r_{nt} \end{pmatrix} \tag{1}$$

fit is the trend of the first I characteristic factor. cit is a periodic term of first characteristic factors. rit is the random term of the 'i'characteristic factor. i=1,2,…,n

2.2 Establish the Characteristic Factor Time Series Trend Item Model

Each feature factors will have monotonicity change trend in different degree in the algae growth process. So the trend is n-dimensional multiple regression model by taking time as the independent variable. Its expression is as follows.

$$\boldsymbol{F}_t = \boldsymbol{F}(t) = \begin{pmatrix} F_{1t} \\ F_{2t} \\ \vdots \\ F_{nt} \end{pmatrix} = \begin{pmatrix} b_1 g(t) + y_{01} \\ b_2 g(t) + y_{02} \\ \vdots \\ b_n g(t) + y_{0n} \end{pmatrix} = \begin{pmatrix} y_{01} & b_1 \\ y_{02} & b_2 \\ \vdots & \vdots \\ y_{0n} & b_n \end{pmatrix} \begin{pmatrix} 1 \\ g(t) \end{pmatrix} \tag{2}$$

F (t) is an n-dimensional monotone multiple regression function. bi is the change rate of the characteristic factor. g(t) is a monotonic function. y0i is the initial value of the

characteristic factor. i=1,2,... N

2.3 Set Up the Characteristic Factor Time Series Period Item Model

There is interaction between the periodic variations of the multiple characteristic factors, so the interaction effect cannot be described by the periodic term of one element cycle model. Multiple latent period model, which applies to mining the potential periodicity of the data and reflects the interaction of multi-week period is proposed in the paper. The model is as follows.

$$\boldsymbol{C}_t = \boldsymbol{C}(t) = \begin{pmatrix} c_{1t} \\ c_{2t} \\ \vdots \\ c_{nt} \end{pmatrix} = \begin{pmatrix} \sum_{j=1}^{q} a_{1j}\cos(\omega_j t + \varphi_{1j}) \\ \sum_{j=1}^{q} a_{2j}\cos(\omega_j t + \varphi_{2j}) \\ \vdots \\ \sum_{j=1}^{q} a_{nj}\cos(\omega_j t + \varphi_{nj}) \end{pmatrix} \tag{3}$$

C (T) is a multi-potential periodic function of multiple latent period model. Q is the number of potential cycle angle frequency. ωj is the jth angular frequency. A_{ij} is the amplitude of ith characteristic factor of the multiple latent period model and jth angular frequency. Φ_{ij} phase is the amplitude of ith characteristic factor of the multiple latent period model and jth angular frequency.i=1,2,…,n.

2.4 Establishe a Time Series Stochastic Term Model of Characteristic Factors

After subtracting the F_t and C_t from Yt, the stationary random part (R_t) of the stochastic term (Y_t) is described by a multivariate autoregressive model, which is widely used and is suitable for prediction. Expressions are as follows.

$$\boldsymbol{R}_t = \sum_{j=1}^{p} \boldsymbol{H}_j \boldsymbol{R}_{t-j} + \boldsymbol{E}_t$$

$$\boldsymbol{H}_j = \begin{pmatrix} \eta_{11j} & \eta_{12j} & \cdots & \eta_{1nj} \\ \eta_{21j} & \eta_{22j} & \cdots & \eta_{2nj} \\ \vdots & \vdots & \ddots & \vdots \\ \eta_{n1j} & \eta_{n2j} & \cdots & \eta_{nnj} \end{pmatrix}, \boldsymbol{E}_t = \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \\ \vdots \\ \varepsilon_{nt} \end{pmatrix} \tag{4}$$

P is a multivariate autoregressive order. Hj is multivariate auto regression coefficient matrix (n*n). Rt-j is a random item at the t-J time.. E_t is a n-dimensional white noise vector of Mutual independence and obeying N [0, Q].Q is a n-dimensional white noise

variance matrix. η_{ikj} is the multiple regression coefficient, which is ith characteristic factor to kth characteristic factor. e_{it} is the white noise of the ith characteristic factor .i,k=1,2,…,n.

2.5 Establish Multivariate Mixed Time Series Model

Multiple time series model is proposed by combining Multiple regression models (equation (2))of trend items(F_t),Multiple latent period models(equation (3)) of periodic terms(C_t) and Multivariate autoregressive model (equation(4))of random term(R_t).The model is as follows.

$$\boldsymbol{Y}_t = \boldsymbol{F}_t + \boldsymbol{C}_t + \boldsymbol{R}_t = \boldsymbol{F}(t) + \boldsymbol{C}(t) + \sum_{j=1}^{p} \boldsymbol{H}_j \boldsymbol{R}_{t-j} + \boldsymbol{E}_t \tag{5}$$

Formula (5) is a multivariate mixed time series model of characteristic factors.

3. Multi Factor Prediction of Water Bloom

The trend, periodicity and randomness of data can be fully excavated by the time series analysis method. However, the accuracy of the method depends on the limitation of the data length. In the time series model, it is difficult to guarantee the accuracy of the prediction if the original data is directly used to predict the future. This method reduces the credibility of the results of water bloom prediction. Grey theory mainly studies the uncertainty of a few data. The advantage of time series model is preserved, the limitations of the prediction are overcome and the prediction accuracy is improved by using grey theory and time series model.

The Nesting system grey prediction method is used to obtain the predictive value of each behavior variable by embedding the GM (1,1) model into the GM (1, N) model for system grey prediction model of a certain structure in Grey Theory .GM (1,1) is a gray prediction model.GM (1, N) is composed of 1 dependent variables and N-1 independent variables. Raw data is added to each 1 step prediction and the first raw data are deleted. Then fitting the time series model and model is modified all the time. Nesting system grey prediction method is used to make the multivariate mixed time series prediction. The flow chart is shown in Figure 1.Specific steps are as follows.

(1) Setting the number of prediction steps (H) and making u=1,2……, H is represented as the predicted step number.

(2) Setting $\{Y_{ut}\}$ (t=1,2,3……).N are represented as a Multivariate time series prediction When the uth step prediction. Multiple regression models are established as follows.

$$\boldsymbol{F}_{ut}=\boldsymbol{F}_u(t)=\begin{pmatrix}F_{u1t}\\F_{u2t}\\\vdots\\F_{unt}\end{pmatrix}=\begin{pmatrix}y_{u01}\ b_{u1}\\y_{u02}\ b_{u2}\\\vdots\ \ \vdots\\y_{u0n}\ b_{un}\end{pmatrix}\begin{pmatrix}1\\g_u(t)\end{pmatrix}$$

(3) The trend term is subtracted from 1 and the remaining data are established as a multiple latent period model. The model is as follows.

$$\boldsymbol{C}_{ut}=\boldsymbol{C}_u(t)=\begin{pmatrix}\sum_{j=1}^{q_u}a_{u1j}\cos(\omega_{uj}t+\varphi_{u1j})\\\sum_{j=1}^{q_u}a_{u2j}\cos(\omega_{uj}t+\varphi_{u2j})\\\vdots\\\sum_{j=1}^{q_u}a_{unj}\cos(\omega_{uj}t+\varphi_{unj})\end{pmatrix}$$

In the model, $q_u,a_{uij},\omega_{uj},\varphi_{uij},i=1,2,\cdots,n$ are represented as multiple latent period function of periodic term and multiple latent period model as well as the corresponding model parameters.

(4) Period items ($\boldsymbol{C}_{ut}$) are subtracted from the $\boldsymbol{Y}_{ut}$ and the remaining data is established as a multivariate autoregressive model.The model is as follows.

$$\boldsymbol{R}_{ut}=\sum_{j=1}^{p_u}\boldsymbol{H}_{uj}\boldsymbol{R}_{u(t-j)}+\boldsymbol{E}_{ut}$$

(5) Multivariate mixed time series model is established.

$$\boldsymbol{Y}_{ut}=\boldsymbol{F}_u(t)+\boldsymbol{C}_u(t)+\sum_{j=1}^{p_u}\boldsymbol{H}_{uj}\boldsymbol{R}_{u(t-j)}+\boldsymbol{E}_{ut}$$

(6) The minimum mean square error prediction formula of multivariate mixed time series model is adopted to carry forward 1 steps prediction.

$$\begin{aligned}&\boldsymbol{Y}_{u(N+1)}=\boldsymbol{Y}_{N+u}\\&=\boldsymbol{F}_u(N+1)+\boldsymbol{C}_u(N+1)+\sum_{j=1}^{p_u}\boldsymbol{H}_{uj}\boldsymbol{R}_{u(N+1-j)}\end{aligned}$$

$\boldsymbol{Y}_{N+u}$ is characteristic factor vector of the N+u time. At this point, the one step prediction of multivariate mixed time series of characteristic factors is completed.

(7) The first data of sequential (Y_{u1}) is removed in the group and one step predictive value of time series ($Y_{u(N+1)}$) is added. A new sets of time series data $\{Y_{ut}\}$ are obtained(t=2,3……N, N+1).This time series data is used to predict the time series data of the(u+1)th step $\{Y_{(u+1)t}\}$(u=1,2,……N).

(8) Making u=u+1 and Repeating steps (2) ~ (6).Multivariate time series of characteristic factors can be predicted by successive.

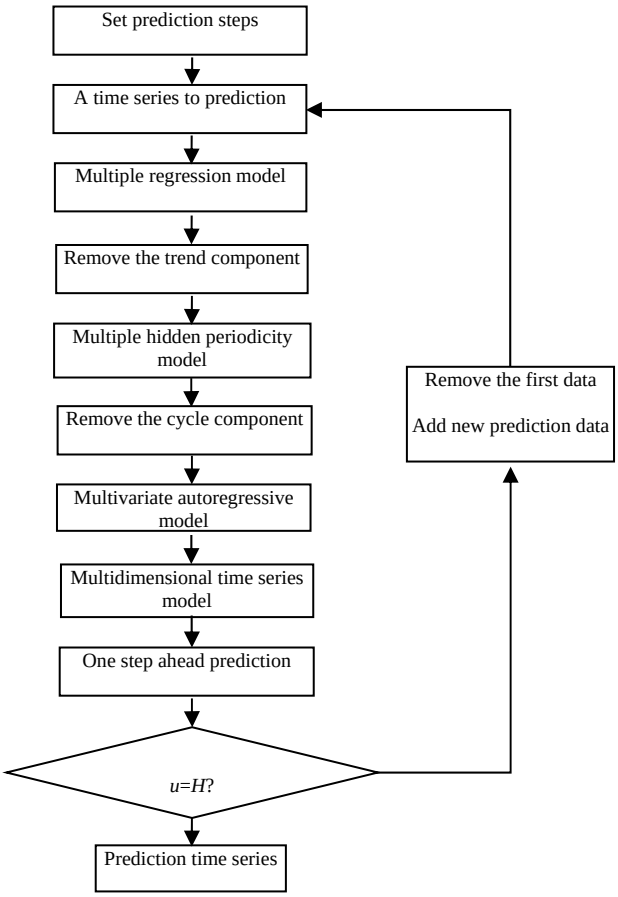

Fig.1 Flow chart of prediction procedure.

4. Instance Verification

The essential nutrients for aquatic plant growth are 1 important factors such as nitrogen. In addition, the growth of algae is also affected by the environmental factors such as water temperature. At the same time, a large number of diffuse algae will be counterproductive in the water environment such as the change of the pH, oxygen consumption and so on.

Chlorophyll concentration is not only an important reference index, which measures primary productivity and eutrophication status of water body but also judges the final

indicator of algae biomass and algal bloom. Total nitrogen, pH v, temperature and oxygen consumption are four influencing factors. Chlorophyll and algal density are two characterization factors. The 6 factors are related to the formation of water bloom. The water bloom characteristics of Taihu in Jiangsu province from June 2009 to June 2012 are monitored to verify the prediction model proposed in this paper. Water bloom characteristic factor data were recorded for 1102 days by testing equipment. Two characterization factors are predicted from the 1075 days to the 1102 days. The predicted results are shown in Figure 2 to 7.The blue curve in the figure is the actual value and the red curve is the predicted value in the figure.

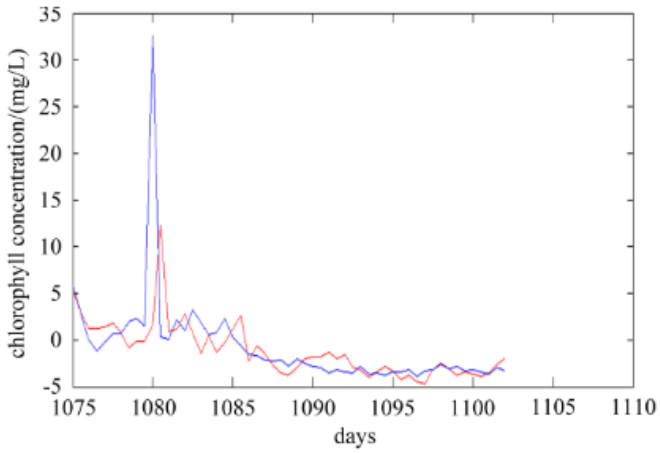

Fig.2 Prediction of chlorophyll concentration.

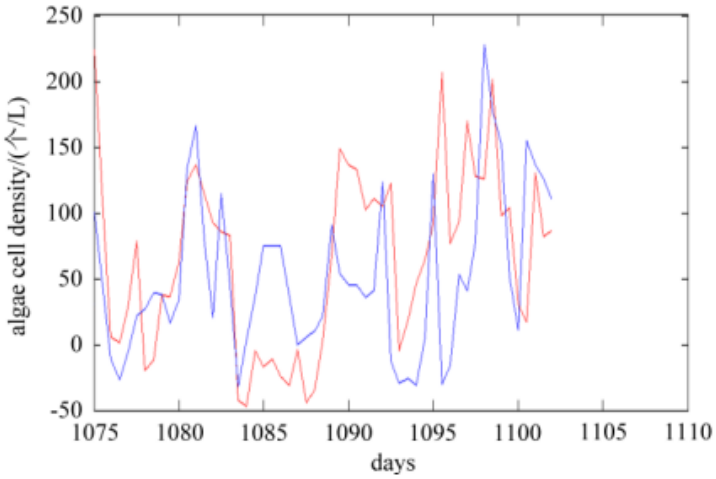

Fig.3 Prediction of algae cell density.

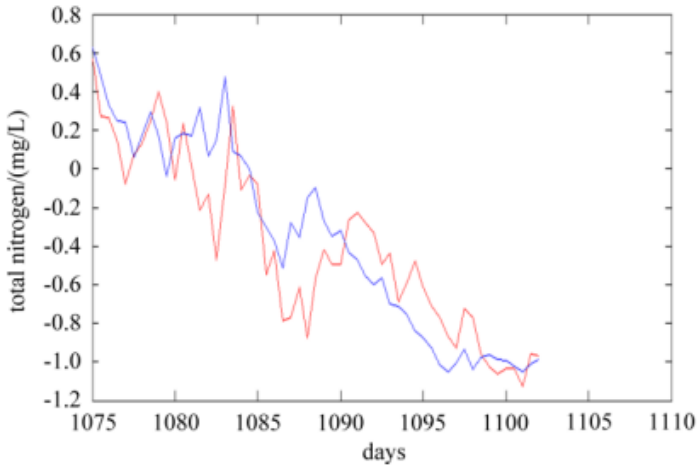

Fig.4 Prediction of total nitrogen.

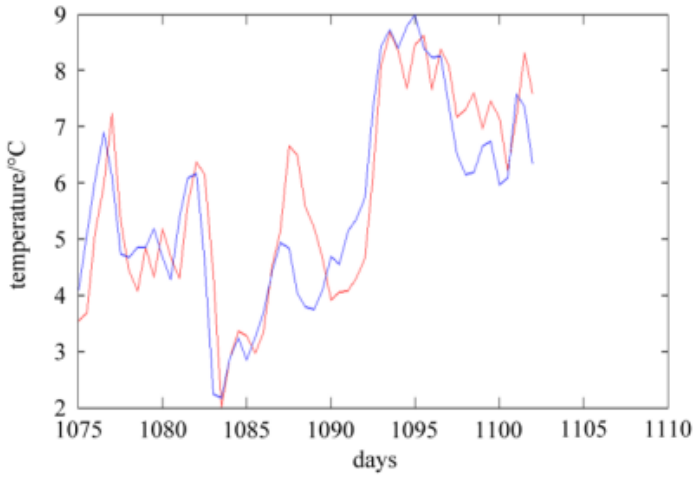

Fig.5 Prediction of temperature.

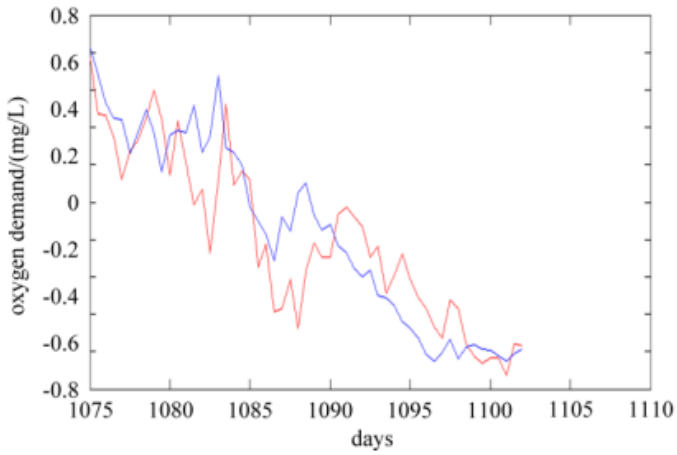

Fig.6 Prediction of oxygen demand.

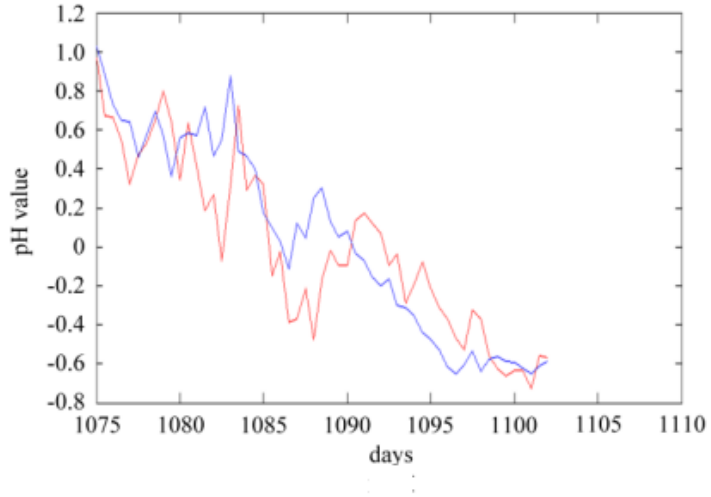

Fig.7 Prediction of pH

5. Results and Discussion

In addition, in order to explain the advantages of the multiple mixed time series analysis and grey theory to the prediction of water bloom. The prediction results of a hybrid timing analysis, grey theory and a hybrid timing analysis as well as Multiple hybrid timing analysis are given and compared with the method proposed in this paper.

Four methods are used to predict multivariate time series of bloom characteristic factor the 1075 days to the1102 days. The absolute value of the predicted mean error of the 6 characteristic factors is shown in table 1.

Table 1 Comparison of prediction errors of different methods

method	One element time series	One element time series and grey theory	Multivariate time series	Multivariate time series and grey theory
chlorophyll	0.1883	0.1789	0.1814	0.1743
Algae density	1.1703	1.0024	0.7528	0.6984
pH	0.0480	0.0578	0.2092	0.0676
water temperature	0.1198	0.1107	0.3055	0.1380
total nitrogen	0.6421	0.5214	0.6315	0.5785
Oxygen consumption	1.8125	1.4667	1.3808	1.0088

The results of multivariate time series prediction of water bloom characteristics is more consistent with the experimental results and the average prediction error absolute value of small by comparing to the other three methods.

Acknowledgments

This work was financially supported by Innovation ability promotion project of Beijing municipal colleges and universities (PXM2014_014213_000033), Major Project of Beijing Municipal Education Commission science and technology development plans (KZ201510011011), and General Project of Beijing Municipal Education Commission science and technology development plans (SQKM201610011009). Those supports are gratefully acknowledged.

References

[1] David A Carona, Marie-Eve Garneaua and Erica Seuberta. Harmful algae and their potential impacts on desalination operations off southern California. Water Research, 2010, 44(2):385-416.

[2] Wang Xiaoyi, Tang Lina and Liu Zaiwen. Formation mechanism of cyanobacteria bloom in urban lake reservoir. Journal of Chemical Industry and Engineering (China), 2012, 63(5): 1492-1497.

[3] Wang Xiaoyi, Tang Lina and Liu Zaiwen. Research on the fuzzy petri net optimization modeling of water bloom formation process. Acta Electronica Sinica, 2013, 41(1):68-71.

[4] Hao Qiwen, Wang Xiaoyi and Xu Jiping. Information system for water quality monitoring of lakes and reservoirs and water-bloom early-warning. Computer Engineering, 2013, 39(1):287-293.

[5] Dong Shuoqi, Liu Zaiwen and Wang Xiaoyi. Intelligent agent modeling and simulating of algal bloom formation mechanism. Journal of Jiangnan University (Natural Science Edition), 2012, 11(4):412-417.

[6] Wang Xiaoyi, Zhao Xiaoping and Liu Zaiwen. Comprehen-sive mechanism modeling on city lake cyanobacteria bloom formation. Acta Scientiae Circumstantiae, 2012, 32(7):1677-1683.

[7] Du Wenli, Guan Zhenqiang and Qian Feng. Dynamic soft sensor series error modeling based on time compensation. Journal of Chemical Industry and Engineering (China), 2010, 61(2):439-443.

[8] Ma Yong, Huang Dexian and Jin Yihui. Discussion about dynamic soft-sensing modeling. Journal of Chemical Industry and Engineering (China), 2005, 56(8):1516-1519.

[9] Liu Zaiwen, Wang Xiaoyi and Cui Lifeng. Computing method for unmeasurable process parameters. Journal of Tsinghua University, 2007, 47(S2):1742-1746.

[10] Shuhaibar B N and Riffat R. A process for harmful algal bloom location prediction using GIS and trend analysis for the terrestrial waters of kuwait. Journal of Environmental Informatics, 2008, 12(2):160-173.

[11] Liu Zaiwen, Yang Bin and HuangZhenfang. Water-bloom short-time predicting system of Beijing based on neural network. Computer Engineering and Applications, 2007, 43(28): 243-245.

[12] Liu Zaiwen, Li Mengxun and Wang Xiaoyi. The method of mid-term and short-term prediction for water bloom based on LSSVM and RBFNN. Computers and Applied Chemistry, 2012, 29(10):1189-1194.

[13] Zhu Shiping, Liu Zaiwen and Wang Xiaoyi. Gray theory and neural network prediction for water bloom. Computer Engineering and Applications, 2011, 47(13):231-233.

[14] Chen Yunfeng, Yin Fucai and Lu Genfa. A catastrophe model for water bloom prediction: A case study of China's Lake Chaohu. Human and Ecological Risk Assessment, 2007, 13(4):914-921.

[15] Liu Zaiwen, Wu Qiaomei and Wang Xiaoyi. Algae growth modeling based on optimization theory and application to water-bloom prediction. Journal of Chemical Industry and Engineering (China), 2008, 59(7):1869-1873.

[16] Victor Chan and William Q Meeker. Time series modeling of degradation due to outdoor weathering. Communications in Statistics-Theory and Methods, 2008, 37(3):408-424.

[17] Wang Li, Li Xiaoyang and Jiang Tongmin. Life prediction of product based on degradation amount distribution time series analysis. Journal of Beijing University of Aeronautics and Astronautics, 2011, 37(4):492-498.

[18] Wang Li, Zhang Huiyan and Xue Hong. Fault prognostics and reliability estimation of DC motor using time series analysis based on degradation data. Journal of Theoretical and Applied Information Technology, 2012, 45(2):568-572.

[19] McGovern Amy, Rosendahl Derek H and Brown Rodger A. Identifying predictive multi-dimensional time series motifs: an application to severe weather prediction. Data Mining and Knowledge Discovery, 2011, 22(1-2):232-258.

[20] Yang Shuzi, Wu Ya and Xuan Jianping. Time series analysis in engineering application. Wuhan: Huazhong University of Science & Technology Press, 2007.

[21] Wang Li, Liu Zaiwen, Wu Chengrui, Hua Wei and Zhang Xue. Water bloom prediction and factor analysis based on multidimensional time series analysis. Journal of Chemical Industry and Engineering (China), 2013, 64(12):4643-4649.

[22] Wang Li, Wang Xiaoyi, et al. Time-varying nonlinear modeling and analysis of algal bloom dynamics. Nonlinear Dynamics, 2016, 84(1): 371-378.