Research Article

The Adaptive-Clustering and Error-Correction Method for Forecasting Cyanobacteria Blooms in Lakes and Reservoirs

Xiao-zhe Bai, Hui-yan Zhang, Xiao-yi Wang, Li Wang, Ji-ping Xu, and Jia-bin Yu

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

Correspondence should be addressed to Hui-yan Zhang; zhanghuiyan369@126.com

Globally, cyanobacteria blooms frequently occur, and effective prediction of cyanobacteria blooms in lakes and reservoirs could constitute an essential proactive strategy for water-resource protection. However, cyanobacteria blooms are very complicated because of the internal stochastic nature of the system evolution and the external uncertainty of the observation data. In this study, an adaptive-clustering algorithm is introduced to obtain some typical operating intervals. In addition, the number of nearest neighbors used for modeling was optimized by particle swarm optimization. Finally, a fuzzy linear regression method based on error-correction was used to revise the model dynamically near the operating point. We found that the combined method can characterize the evolutionary track of cyanobacteria blooms in lakes and reservoirs. The model constructed in this paper is compared to other cyanobacteria-bloom forecasting methods (e.g., phase space reconstruction and traditional-clustering linear regression), and, then, the average relative error and average absolute error are used to compare the accuracies of these models. The results suggest that the proposed model is superior. As such, the newly developed approach achieves more precise predictions, which can be used to prevent the further deterioration of the water environment.

1. Introduction

Cyanobacteria blooms are the consequence of eutrophication, which can no longer be ignored. It clearly interferes with lake ecosystem functioning (severe loss of aquatic biodiversity) and services such as the provisioning of drinking water and recreation [1]. The formation of cyanobacteria blooms is a multifactor coupled multidimensional coordination complex dynamical system with an intrinsic strong nonlinear dissipative structure [2]. Precise prediction of cyanobacteria blooms can effectively mitigate eutrophication and protect the lake ecosystem. The cyanobacteria blooms have always exhibited high variability and instability, which makes them very difficult to monitor and predict [3].

Cyanobacteria-bloom data are a typical and complex type of chaotic time-series data based on long-term field observation and simulation experiments [4]. Traditionally, global and local prediction algorithms have been used for the classical modeling and prediction of chaotic time series [5, 6]. The global method attempts to approximate the entire time series on all attractors and seeks a function that is valid at every point. Then, we can predict the future values from knowledge of all previous elements of the time series. However, the parameters may be changed when new information is added to the model [7]. The local method only uses the nearest information to the local attractors, and the dynamics of the system are described locally, step by step, in the phase space [8].

Recently, researchers found that the prediction precision can be improved by using some combined techniques. These include artificial neural networks combined with fuzzy theory [9, 10], the unbiased composite kernel LSSVR [11], and K-nearest neighbors with phase space reconstruction [12]. Generally, these methods can obtain better results than traditional individual models, but they are complex, and the parameter selection is affected by personal experience.

The variation of cyanobacteria blooms is of a dynamic and sensitive nature, as ecological environments show different characteristics at different intervals. Local linearization is a useful method for modeling and prediction in this situation, and the choice of the dimension of the regression model is also an important issue. In this paper, we propose

a validity index to cluster some typical operating intervals on research objects for dynamic state identification. The typical operating intervals are subsets of samples, which are of unequal sizes. The subset based on the division of similarity also has considerable influence on the operating-point range of the modeling process. The optimal number of neighbors participating in the modeling is set by the particle swarm optimization (PSO) algorithm. By using an error-correction method to adjust the function coefficients, the model can approximate the local attractor and improve the prediction accuracy.

In this paper, the following contributions are made:

(i) The clustering algorithm based on the validity index proposed in this paper is used to split the time-series data into unequal intervals and to narrow the field of prediction by finding partial neighbor information of current points in each subset. Using this approach, the prediction speed can be improved. In addition, this validity index is easy to understand and convenient for popularization and application.

(ii) Prediction deviation is used to adjust model parameters online, which means the idea of error-correction is introduced, and the proposed prediction model has the ability to resist external disturbances.

The rest of the paper is organized as follows. In Section 2, the prediction model is introduced. In Section 3, experiments and comparisons of different models are presented. Conclusions are given in Section 4.

2. Model-Based Prediction of Cyanobacteria Blooms in the Short Term

The objective is to predict the general trend of cyanobacteria blooms over a short period. The proposed prediction model involves the following three steps: first, using the proposed concise validity index, the time-series dataset is split into typical subsets. The number of clusters is determined by data similarity characteristics, and the maximum value of the validity index corresponds to the optimal clustering number. Then, the error-correction method, PSO, and fuzzy regression algorithm are used to optimize the data-driven model, which is used to predict the cyanobacteria blooms for the next time. The values of multiple nearest neighbors of the subset are used to express the evolutionary trends in the time series, and the error between the predicted value and actual value can be used as feedback to correct the model coefficients dynamically based on fuzzy theory. In this way, the prediction of cyanobacteria blooms can be performed precisely without generating cumulative errors.

2.1. Optimal Number of Clusters Based on the Validity Index.
The objective of the current work is to predict the fluctuation of cyanobacteria blooms over a consecutive future time. The most basic step of the proposed prediction method is splitting the time-series modeling data into a set of typical subsets with nearest-neighbor similarity via the adaptive-clustering principle, according to which the number of

clusters is determined by the similarity characteristics of the data. Clustering is an unsupervised learning process. The optimal number of clusters and the partition are determined according to a validity index that rewards higher similarity within each cluster and lower similarity between different clusters. Research results show that there is no validity index that is applicable to every case. At present, many validity indices have been used in various studies. The commonly used clustering validity indices include Calinski-Harabasz (CH) [13], Hartigan (Ht) [14], Homogeneity-Separation (HS) [15], and Krzanowski-Lai (KL) [16].

Based on the preliminary research on chlorophyll a concentration, this paper proposes an improved cluster validity index according to its characteristics. The natural properties of the clustering results are used to evaluate intraclass tightness and interclass separability. We define the intraclass distance and interclass distance to compute the cluster validity index, which is named the clustering comprehensive quality. The formula for the intraclass distance is as follows:

$$\mathrm{Dwc}\,(c)$$
$$= \frac{1}{c} \sum_{w=1}^{c} \left(\left(\frac{n_w \left(n_w - 1 \right)}{2} \right)^{-1} \sum_{i=1, i \neq j}^{n_w} \left| x_{n_w \cdot i} - x_{n_w \cdot j} \right| \right). \tag{1}$$

Here, c is the number of clusters; n_w is the data size of each cluster, where $w = 1, 2, \ldots, c$; and $x_{n_w \cdot i}, x_{n_w \cdot j}$ denote the ith and jth data of the subset, which has n_w samples, where $i, j = 1, 2, \ldots, n_w$.

The formula for the interclass distance is as follows:

$$\mathrm{Dbc}\,(c) = \left(\frac{c \left(c - 1 \right)}{2} \right)^{-1} \sum_{u=1, u \neq v}^{c} \left| c_u - c_v \right|. \tag{2}$$

Here, c_u, c_v denote the uth and vth cluster centers, where $u, v = 1, 2, \ldots, c$.

The formula for the proposed cluster validity index is as follows:

$$\mathrm{Coq}\,(c)$$
$$= 1 - \left(\left(\varepsilon \times \mathrm{Dwc}\,(c) \right)^p + \left((1 - \varepsilon) \times \frac{1}{\mathrm{Dbc}\,(c)} \right)^q \right). \tag{3}$$

Here, ε, p, and q are the weights for balancing the intraclass distance and interclass distance. For the general case, they are all positive real numbers, and $p, q > 0$, $0 < \varepsilon < 1$. In application, when the intraclass distance is too small or the interclass distance is too large, the three parameters can be adjusted to avoid weakening the performance indicator of clustering comprehensive quality. Obviously, the greater the clustering comprehensive quality is, the better the effect of clustering is.

In this paper, $p = 1$, $q = 1$, and $\varepsilon = 0.7$ are set based on the characteristics of the distribution of the chlorophyll a concentration data.

2.2. K-Nearest Neighbors Based on the Particle Swarm Optimization Algorithm.
Splitting into different clusters is equivalent to the local division of the motion law. In the K-nearest

method, each sample can be represented by its closest K neighbors. After determining the best cluster partition, it is critical to construct a model to predict the value of the evolutionary track using local similarity information. It is a difficult task to select the appropriate number of similar sample points. The particle swarm optimization algorithm is a new global-optimization evolutionary algorithm based on the foraging behavior of birds, which was advanced by Dr. Eberhart and Dr. Kennedy [17]. In this paper, the particle swarm optimization algorithm is used to determine the optimal number of most similar neighbors among a cluster for a sample point.

2.3. Fuzzy Linear Regression Model.

Consider the classical linear regression equation

$$y = A_0 + \sum_{m=1}^{f} A_m x_m. \qquad (4)$$

Here, A_m is the regression coefficient, where $m = 1, 2, \ldots, f$, and x_m is the independent variable of the model.

In a real application, the data often have fuzzy uncertainties due to various factors. In this paper, fuzzy linear regression is introduced.

The fuzzy linear regression equation is represented as follows:

$$\widetilde{y}_d = \widetilde{A}_0 + \widetilde{A}_1 x_{1d} + \cdots + \widetilde{A}_f x_{fd} = \widetilde{A}_0 + \sum_{l=1}^{f} \widetilde{A}_l x_{ld}. \qquad (5)$$

Here, $\widetilde{A}_l$ is the fuzzy regression coefficient, where $l = 1, 2, \ldots, f$; x_{ld} is the fuzzy independent variable, where $d = 1, 2, \ldots, h$; and $\widetilde{y}_d$ is the fuzzy dependent variable.

To simplify the use of the model, the triangular fuzzy representation was adopted as $(\widetilde{A}_l(\alpha_l, c_l))$ and $(\widetilde{y}_d(y_d, c_{yd}))$, where α_l and y_d are central values of triangular fuzzy numbers and c_l and c_{yd} are the fuzzy amplitudes.

2.4. Model Improvement and Solving.

Optimization of the regression coefficients should be subject to two constraint conditions: one ensures that the model fits the data well, and the other ensures the minimal fuzziness of the regression function. By scaling the degree of the fit using the closeness degree, the problem is transformed into a problem that minimizes the fuzzy amplitude under the condition $h(\widetilde{A}_0 + \sum_{l=1}^{f} \widetilde{A}_l x_{ld}, \widetilde{y}_d) \geq h_0$. The transformation is given by

$$s\left(\widetilde{A}_0 + \sum_{l=1}^{f} \widetilde{A}_l x_{ld}\right) \longrightarrow \min$$

$$h\left(\widetilde{A}_0 + \sum_{l=1}^{f} \widetilde{A}_l x_{ld}, \widetilde{y}_d\right) \geq h_0. \qquad (6)$$

Here, h_0 is the given closeness-degree standard, and $s(\widetilde{A}_0 + \sum_{l=1}^{f} \widetilde{A}_l x_{ld})$ is the fuzzy amplitude of $\widetilde{y}_d$.

The formulation of the closeness degree is represented as follows:

$$h\left(\widetilde{A}_0 + \sum_{l=1}^{f} \widetilde{A}_l x_{ld}, \widetilde{y}_d\right) = 1 - \frac{\left|y_d - \sum_{l=1}^{f} \alpha_l x_{ld}\right|}{\sum_{l=1}^{f} c_l |x_{ld}|}. \qquad (7)$$

Given a set of weights $\omega = \{\omega_0, \omega_1, \ldots, \omega_f\}$, the fuzzy amplitude of the system under the weights is represented as follows:

$$s = \sum_{l=0}^{f} \omega_l c_l. \qquad (8)$$

Obviously, the fuzzy linear regression is essentially an optimization problem, which is solved by minimizing the objective function subject to the following constraints:

$$\sum_{l=1}^{f} c_l \omega_l \longrightarrow \min$$

$$\sum_{l=1}^{f} c_l x_{ld} \longrightarrow \min$$

$$\sum_{l=1}^{f} \alpha_l x_{ld} - (1 - h_0) \sum_{l=1}^{f} c_l |x_{ld}| \leq y_d$$

$$\sum_{l=1}^{f} \alpha_l x_{ld} + (1 - h_0) \sum_{l=1}^{f} c_l |x_{ld}| \geq y_d. \qquad (9)$$

Here, ω_l is the weight of the fuzzy amplitude in the objective function, and, generally, its value is 1.

In a linear programming problem, the weight of each fuzzy amplitude in the objective function should be different because the influence of each regression variable on the dependent variable is different. Since the correlation coefficient between the regression variable and dependent variable can roughly reflect the degree of influence of each regression variable on the dependent variable, the weights are corrected according to correlation coefficients. The formulation is represented as follows:

$$\sum_{l=1}^{f} \omega_l = 1,$$

$$\omega_l = \frac{|r_l|}{\sum_{l=1}^{f} |r_l|} \quad l = 1, 2, \ldots, f. \qquad (10)$$

Here, r_l is the correlation coefficient between the lth regression variable and the dependent variable.

2.5. Model Evaluation.

After the model is solved, a certain standard can be selected to evaluate the fitting performance of the model.

(1) *Closeness degree.* The fitting accuracy is generally considered to be high when the closeness degree is greater than 0.5.

(2) The ratio Φ_1 of the relative error between the predicted value and actual value to the actual value and the ratio Φ_2 of the fuzzy amplitude to the actual value are

$$\Phi_1 = \frac{\left| y_d - \sum_{l=1}^{f} \alpha_l x_{ld} \right|}{y_d},$$

$$\Phi_2 = \frac{\sum_{l=1}^{f} c_l \left| x_{ld} \right|}{y_d}. \tag{11}$$

When Φ_1 and Φ_2 are less than 20%, the fitting accuracy is better.

2.6. Prediction Model. In view of the characteristics of the random and nonperiodic chaotic behavior in the evolution process of cyanobacteria blooms, a fuzzy linear regression prediction model is constructed. According to the theory of the local linearization of nonlinear systems, a simple and novel validity index is used to split the sample data into several subsets with the adaptive-clustering algorithm. Then, the particle swarm optimization algorithm is used to select the optimal number of nearest neighbors in clusters of modeling data based on the Euclidean distance. Finally, with the fuzzy characteristics of the monitoring data, the fuzzy linear regression model is constructed.

First, the adaptive-clustering algorithm is used to split the time-series data into unequal intervals and narrow the field of prediction. The most important aspect is the choice of the number of clusters. To obtain the best clustering partition, a simple and effective validity index is proposed to evaluate the quality of clustering. Second, the subset of current monitoring points is selected, and the particle swarm optimization algorithm is used to determine the optimal number of most similar neighbors from among the subset for a sample point. Third, with the selected nearest-neighbor data, a fuzzy linear regression method based on error-correction is adopted for the local small-range prediction. This model effectively utilizes the historical information of the nearest neighbors of the subset to make the predictive value closer to the actual value and revise the prediction model dynamically based on the deviation.

3. Experiments and Comparison

This section presents experiments on the application of the prediction model to cyanobacteria-bloom forecasting. In this paper, we used chlorophyll a concentration as the characterization index to characterize cyanobacterial bloom formation. The proposed adaptive-clustering and error-correction method was able to capture the dynamic characteristics of high nonlinearity and uncertainty.

It has been confirmed that the evolution of cyanobacteria blooms has a chaotic property, and the sensitivity characteristics of the chaotic system can cause great differences in evolutionary trajectories due to slight disturbances. Therefore, the chaotic system cannot be predicted over the long term. A set of chlorophyll a concentrations with 2000 data values
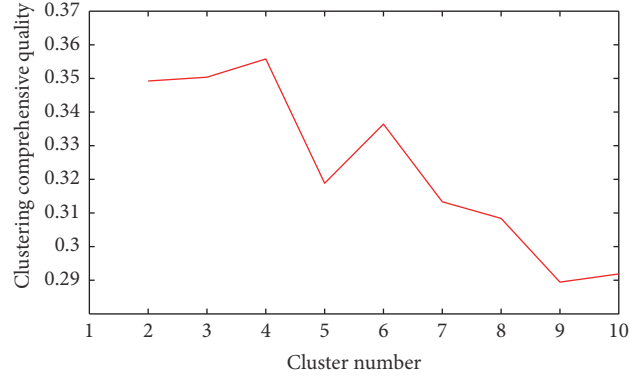

FIGURE 1: Clustering comprehensive quality.

collected every 4 hours from January 1, 2011, to December 31, 2011, in the Jin Shu monitoring site of Taihu was used to evaluate the prediction of chlorophyll a concentration from January 1, 2012, to January 7, 2012.

3.1. Evaluation Indicator. To evaluate the accuracy of these predictions, the average absolute error and the average relative error between predicted values and actual values can be calculated as follows:

$$e_{av} = \frac{1}{g} \sum_{z=1}^{g} e_z = \frac{1}{g} \sum_{z=1}^{g} \left| Acv_z - Pcv_z \right|,$$

$$e_{rz} = \frac{Acv_z - Pcv_z}{Acv_z},$$

$$e_{rav} = \frac{1}{g} \sum_{z=1}^{g} e_{rz} = \frac{1}{g} \sum_{z=1}^{g} \left| \frac{Acv_z - Pcv_z}{Acv_z} \right|. \tag{12}$$

Here, e_{av} is the average absolute error, e_z is the absolute error of the chlorophyll a concentration, Acv_z is the actual value, Pcv_z is the predicted value, g is the prediction length, e_{rz} is the relative error, and e_{rav} is the average relative error.

3.2. Optimal Cluster Number. The training data were optimized based on the validity index proposed in this paper, and the curve of clustering comprehensive quality with different numbers of clusters is shown in Figure 1. The optimal number of clusters was computed according to formula (3); obviously, $c = 4$.

The same training data are split according to the above-mentioned four kinds of clustering validity indices: the CH index, Ht index, HS index, and KL index. The curves of the indices are shown in Figure 2.

For the CH index, KL index, and HS index functions, the number of clusters at which the function attains its maximum is taken as the optimal number of clusters. For the Ht index function, the minimum number of clusters with index-function value less than or equal to 10 is taken as the optimal number of clusters. As shown in Figure 2, the optimal numbers of clusters obtained by the CH index, KL index, HS index, and Ht index are $c = 7$, $c = 6$, $c = 4$, and $c = 1$, respectively.
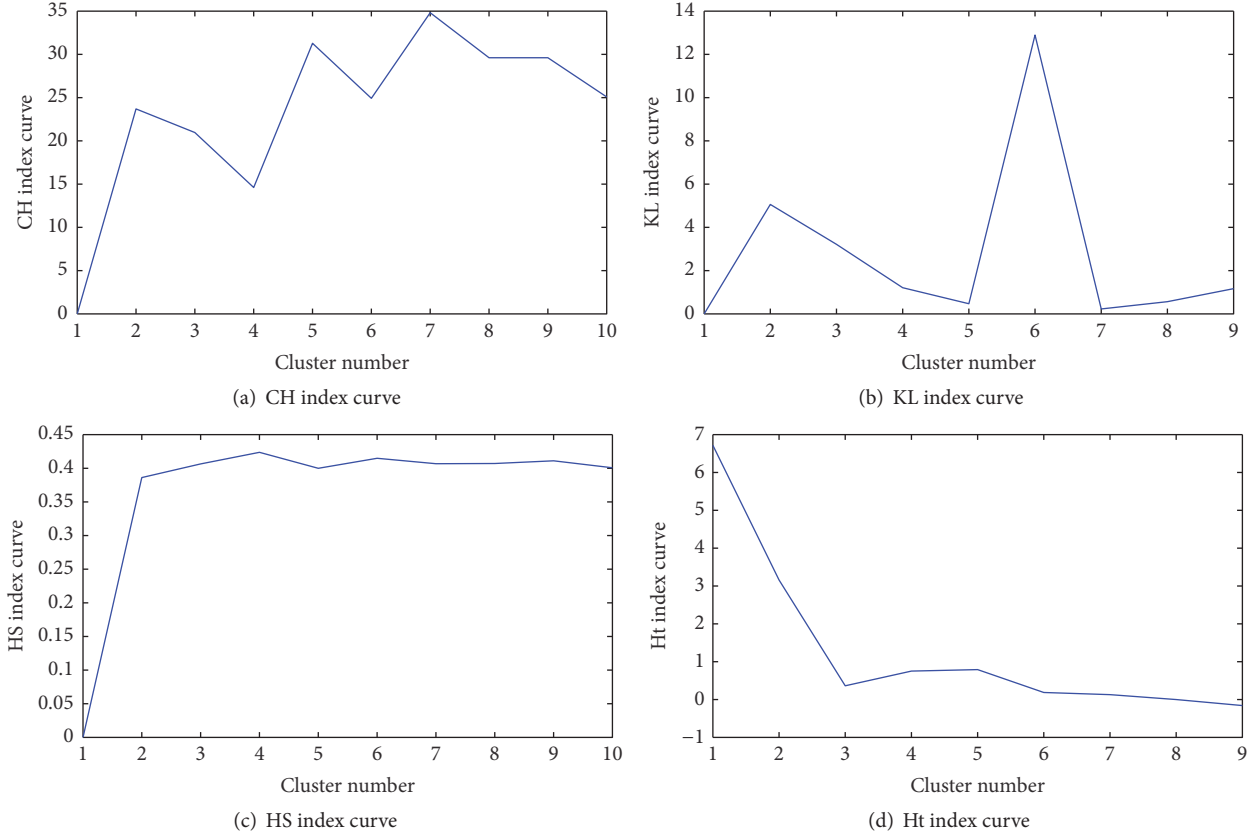
(a) CH index curve

(b) KL index curve

(c) HS index curve

(d) Ht index curve

FIGURE 2: Clustering result curves based on common validity indices.

TABLE 1: Fuzzy coefficients.

Fuzzy coefficient $\widetilde{A}_l$	Central value α_l	Fuzzy amplitude c_l
$\widetilde{A}_1$	0.0000	0.0000
$\widetilde{A}_2$	0.0103	0.0506
$\widetilde{A}_3$	0.0838	0.0000
$\widetilde{A}_4$	0.0000	0.0000
$\widetilde{A}_5$	0.0000	0.1267
$\widetilde{A}_6$	0.1378	0.0000
$\widetilde{A}_7$	0.0000	0.0000
$\widetilde{A}_8$	0.4427	0.0556
$\widetilde{A}_9$	0.0000	0.0000
$\widetilde{A}_{10}$	0.3180	0.0000

3.3. Model Construction and Evaluation. The optimal number of neighbors was determined using the particle swarm optimization algorithm for each subset. The optimal number of neighbors was found to be 10 by optimization.

Therefore, 10 nearest-neighbor samples were included in the fuzzy linear regression prediction model in each subset. H was set to 0.8, and Lingo software was used to compute the fuzzy coefficients. The fuzzy coefficients are listed in Table 1.

To evaluate the fitting accuracy of the model, relevant data were substituted into the model. The predicted values and actual values are listed in Table 2.

As shown in Table 2, most of the closeness degrees exceed 0.5; the values of Φ_1 and Φ_2 are less than 20%. This means that the fitting accuracy of the model is high.

3.4. Experimental Results. The values obtained using the fuzzy linear regression model for 20-step prediction and the actual values are shown in Figure 3.

The values predicted by the fuzzy linear regression model based on adaptive-clustering and error-correction are consistent with the actual values of chlorophyll a concentration, and the model has good prediction precision.

To assess the effectiveness of the proposed method, traditional-clustering linear regression and adaptive-clustering linear regression based on the four kinds of clustering validity indices described above were used to validate the experimental results. In addition, phase space reconstruction was used. The fuzzy linear regression model based on adaptive-clustering and error-correction was calculated with

TABLE 2: Predicted values and actual values.

Actual value/(mg/L)	Fit values/(mg/L)			Evaluation index		
	Central value	Fuzzy amplitude	Interval range	h_0	Φ_1/%	Φ_2/%
1.3000	1.4234	0.1674	[1.2560, 1.5908]	0.26	9.49	12.88
1.5250	1.4376	0.1897	[1.2479, 1.6273]	0.54	5.73	12.44
3.1168	3.2198	0.3391	[2.8807, 3.5589]	0.70	3.30	10.88
2.9000	2.9297	0.3004	[2.6293, 3.2301]	0.90	1.02	10.36
3.1000	3.0101	0.3292	[2.6809, 3.3393]	0.73	2.90	10.62
3.7000	3.5301	0.4127	[3.1174, 3.9428]	0.59	4.59	11.15
3.7000	3.6085	0.3099	[3.2986, 3.9184]	0.70	2.47	8.38
2.5000	2.7006	0.3308	[2.3698, 3.0314]	0.40	8.02	13.23
1.8000	1.9335	0.2579	[1.6756, 2.1914]	0.48	7.42	14.33
3.2000	3.0964	0.2286	[2.5678, 3.3250]	0.55	3.24	7.14

TABLE 3: Comparison of results.

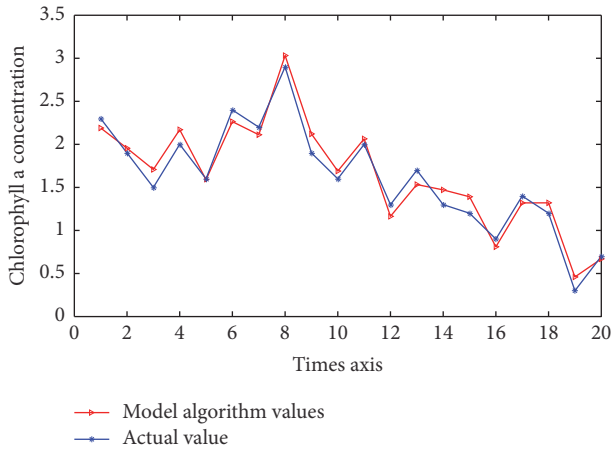| Prediction method	Evaluation index	
	Average absolute error	Average relative error
Phase space reconstruction	0.4440	0.3365
Traditional-clustering linear regression	0.1887	0.1368
Adaptive-clustering linear regression (Ht index)	0.1818	0.1443
Adaptive-clustering linear regression (KL index)	0.1380	0.1204
Adaptive-clustering linear regression (CH index)	0.1456	0.1231
Adaptive-clustering linear regression (clustering comprehensive quality)	0.1553	0.1171
Adaptive-clustering fuzzy linear regression (clustering comprehensive quality)	0.1276	0.0994

FIGURE 3: Prediction result of chlorophyll a concentration based on the fuzzy linear regression model.

the center value of the predicted value. The results of other prediction models are shown in Figure 4, and the errors are listed in Table 3.

According to Figure 4 and Table 3, the prediction precision of the model constructed in this paper is relatively high. In this way, the effectiveness of the proposed method and the feasibility of the model are verified. Although the optimal clustering number obtained by the HS index is the same as that obtained by the cluster validity index we used in this paper, the calculation method for the HS index is difficult to understand, and its generality is not strong. It is clear that the validity index proposed in this paper is convenient and easy to understand and has a strong practicality.
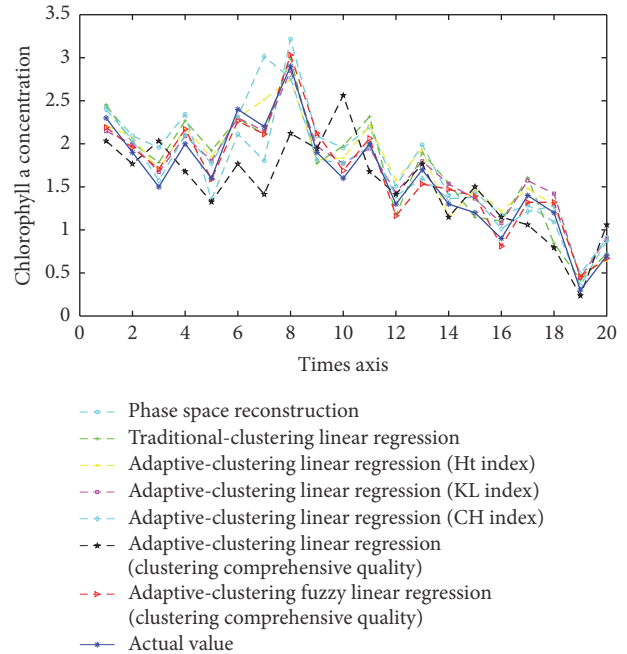

FIGURE 4: Comparison of results of prediction models.

4. Conclusion and Future Work

In this paper, the adaptive-clustering and error-correction methods were introduced to predict the chlorophyll a concentration used to characterize cyanobacterial bloom formation. Considering the chaotic and uncertain characteristics of chlorophyll a concentration, the adaptive-clustering

algorithm was introduced here to obtain some typical subsets. In addition, the optimal number of nearest neighbors for the model was optimized by particle swarm optimization. Finally, the fuzzy linear regression method based on error-correction was used to revise the model dynamically near the operating point. We found that the combined method can characterize the evolutionary track of cyanobacteria blooms in lakes and reservoirs. It can be concluded that the model constructed in this paper can capture complex dynamics, such as the trends of chlorophyll a concentration in cyanobacteria blooms.

This model is intended for short-term forecasting. The complexities of data-driven modeling can greatly increase the difficulty of prediction. Future work should explore additional learning methods and even combine different learning methods to predict cyanobacteria blooms.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

References

[1] B. W. Ibelings, M. Bormans, J. Fastner, and P. M. Visser, "CYANOCOST special issue on cyanobacterial blooms: synopsis—a critical review of the management options for their prevention, control and mitigation," *Aquatic Ecology*, vol. 50, pp. 595–605, 2016.

[2] L. S. Chen, *Nonlinear Biodynamic System*, Science Press, Beijing, China, 1993.

[3] B. Q. Qin, G. J. Yang, J. R. Ma et al., "Dynamics of variability and mechanism of harmful cyanobacteria bloom in Lake Taihu, China," *Kexue Tongbao/Chinese*, vol. 61, no. 7, pp. 759–770, 2016.

[4] X. Y. Wang, "Chaotic time series prediction method of algal bloom in urban lake and reservoir based on complex network," Chinese Patent, 201510128961.5P, 2015.

[5] P. R. L. Alves, L. G. S. Duarte, and L. A. C. P. D. Mota, "Improvement in global forecast for chaotic time series," *Computer Physics Communications*, vol. 207, pp. 325–340, 2016.

[6] Q.-F. Meng, Y.-H. Peng, and J. Sun, "The improved local linear prediction of chaotic time series," *Chinese Physics*, vol. 16, no. 11, pp. 3220–3225, 2007.

[7] L. Su and C. Li, "Local prediction of chaotic time series based on polynomial coefficient autoregressive model," *Mathematical Problems in Engineering*, vol. 2015, Article ID 901807, 14 pages, 2015.

[8] M. N. Islam and B. Sivakumar, "Characterization and prediction of runoff dynamics: a nonlinear dynamical view," *Advances in Water Resources*, vol. 25, no. 2, pp. 179–190, 2002.

[9] G. Heydari, M. Vali, and A. A. Gharaveisi, "Chaotic time series prediction via artificial neural square fuzzy inference system," *Expert Systems with Applications*, vol. 55, pp. 461–468, 2016.

[10] P. L. Gentili, H. Gotoda, M. Dolnik, and I. R. Epstein, "Analysis and prediction of aperiodic hydrodynamic oscillatory time series by feed-forward neural networks, fuzzy logic, and a local nonlinear predictor," *Chaos. An Interdisciplinary Journal of Nonlinear Science*, vol. 25, no. 1, Article ID 013104, 2015.

[11] H. Wang, F. Sun, Y. Cai, and Z. Zhao, "Online chaotic time series prediction using unbiased composite kernel machine via Cholesky factorization," *Soft Computing*, vol. 14, no. 9, pp. 931–944, 2010.

[12] H. Tongal and R. Berndtsson, "Phase-space reconstruction and self-exciting threshold modeling approach to forecast lake water levels," *Stochastic Environmental Research & Risk Assessment*, vol. 28, no. 4, pp. 955–971, 2014.

[13] S. Kermani, N. Samadzadehaghdam, and M. EtehadTavakol, "Automatic color segmentation of breast infrared images using a Gaussian mixture model," *Optik*, vol. 126, no. 21, pp. 3288–3294, 2015.

[14] R. C. de Amorim and C. Hennig, "Recovering the number of clusters in data sets with noise features using feature rescaling factors," *Information Sciences*, vol. 324, pp. 126–145, 2015.

[15] G. Chen, S. A. Jaradat, N. Banerjee, T. S. Tanaka, M. S. Ko, and M. Q. Zhang, "Evaluation and comparison of clustering algorithms in analyzing ES cell gene expression data," *Statistica Sinica*, vol. 12, no. 1, pp. 241–262, 2002.

[16] H. Guo, X. Liu, and Z. Sun, "Multivariate time series prediction using a hybridization of VARMA models and Bayesian networks," *Journal of Applied Statistics*, vol. 43, no. 16, pp. 2897–2909, 2016.

[17] Y. Liu, X. J. Zhang, X. M. Guan, and D. Delahaye, "Potential odor intensity grid based uav path planning algorithm with particle swarm optimization approach," *Mathematical Problems in Engineering*, vol. 2016, no. 2, Article ID 7802798, 16 pages, 2016.