

## Research Article

# Prediction Research of Red Tide Based on Improved FCM

Xiaomei Hu,<sup>1</sup> Dong Wang,<sup>1</sup> Hewei Qu,<sup>1</sup> and Xinran Shi<sup>2</sup>

<sup>1</sup>The Key Laboratory of Intelligent Manufacturing and Robotics, School of Mechatronic Engineering and Automation, Shanghai University, Mailbox 232, No. 149 Yanchang Road, Shanghai 200072, China

<sup>2</sup>Department of Mechanical Engineering, College of Engineering, University of Michigan, Ann Arbor, MI 48105, USA

Correspondence should be addressed to Xiaomei Hu; [sufeimasohxm@163.com](mailto:sufeimasohxm@163.com)

Received 21 September 2015; Revised 15 December 2015; Accepted 17 December 2015

Academic Editor: Daniela Boso

Copyright © 2016 Xiaomei Hu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Red tides are caused by the combination effects of many marine elements. The complexity of the marine ecosystem makes it hard to find the relationship between marine elements and red tides. The algorithm of fuzzy *c*-means (FCM) can get clear classification of things and expresses the fuzzy state among different things. Therefore, a prediction algorithm of red tide based on improved FCM is proposed. In order to overcome the defect of FCM which is overdependent on the initial cluster centers and the objective function, this paper gains the initial cluster centers through the principle of regional minimum data density and the minimum mean distance. The feature weighted cluster center is added to the objective function. Finally, the improved FCM algorithm is applied in the prediction research of red tide, and the results show that the improved FCM algorithm has good denoising ability and high accuracy in the prediction of red tides.

## 1. Introduction

Ocean as the cradle of human beings provides humans with abundant biological resources and mineral resources. Along with the rapid growth of population and economic society in the 21st century, we march into the sea to alleviate the shortage of resources, followed by the marine ecological environment pollution and destruction. Because of a large amount of untreated waste water directly discharged into the ocean and global climate change, harmful red tide species increased dramatically. The cause of the red tides is more complex. Although the occurrence mechanism of red tide has not yet been determined, most scholars believe that red tide occurrence is closely related to water eutrophication [1, 2].

There have been some studies on the red tide prediction. Using numerical method, Gibson et al. established NPZ ecological dynamic model and analyzed five kinds of seston feeding functions [3]. Zhang et al. combined the multifactor including the meteorological and hydrological data to forecast the red tide [4]. Wang et al. established a multivariate adaptive spline regression model to forecast the red tide [5]. However, the accuracy of the traditional red tide prediction is flawed. With the rapid rise of artificial intelligence, artificial

neural network (ANN) has been applied to the red tide prediction and has made a lot of achievements [6, 7]. Artificial neural network also has many defects, such as low learning efficiency and unstable network learning and memory.

In this paper, a fuzzy *c*-means clustering algorithm is proposed for the red tide prediction. The algorithm is optimized from the initial cluster center selection and the improvement of the objective function. As the experiment shows, the prediction of red tide based on improved FCM algorithm has better clustering results, converging speed and robustness to noise and outliers.

## 2. Related Research

*2.1. Research on the Prediction Method of Red Tide.* Red tide is caused by the comprehensive action of multiple factors, such as the sudden proliferation and accumulation of some plankton [8]. It has the characteristics of common natural disasters, namely, inhomogeneity, diversity, difference, burst, reproducibility, disorder, randomness, and predictability in time and space [9]. As one of the most serious marine disasters, the scale of the red tide and the loss of the economy are increasing year by year. Red tide also has a negative

effect on the marine economy sustainable development in our country; therefore the prediction research of red tide has an important meaning in environmental protection and reducing the economic loss [10–12].

According to the research on the prediction of red tide, there are already many prediction methods, such as empirical prediction method, statistical prediction method, numerical model prediction method, and artificial neural network [13–15].

(1) *Empirical Prediction Method.* There is a certain regularity between the occurrence of red tide and the changes of environmental factors, such as meteorological conditions, oceanographic processes, and ecological factors. Empirical prediction method is based on the certain regularity.

(2) *Statistical Prediction Method.* The empirical prediction method only depends on the change of a certain environmental factor. However, red tide in fact is caused by many factors. The statistical prediction method can have comprehensive analysis on the factors; therefore it has stronger predictive ability. Its main method includes principal component analysis, discriminant analysis, and stepwise regression method.

(3) *Numerical Model Prediction Method.* Due to the lack of detailed understanding of the occurrence mechanism of red tide, statistical prediction method is to a certain degree subjective and blind. Numerical model prediction method uses a variety of mathematical tools to analyze, solve, and simulate the model based on various physical and chemical biological coupling factors.

(4) *Artificial Neural Network.* With the rapid development of computer technology, artificial intelligence technology, and biotechnology, artificial neural network has been applied to the prediction of the red tide. However, the research of this method has been a certain constraint because it has the slow convergence speed and it is easy to fall into local minimum values.

**2.2. Research on the Fuzzy  $c$ -Means Clustering Algorithm.** Clustering is one of the most basic activities of human understanding of the world. The purpose of clustering is to make the same class of things as similar as possible and different categories of things as different as possible. According to the different values of membership degree, the clustering method can be divided into hard clustering method and fuzzy clustering method. As for the hard clustering method, 0 means that the sample must not fall into this category and 1 means that the sample must belong to this category. Fuzzy clustering method is a combination of fuzzy theory and clustering analysis. The fuzzy clustering algorithm proposed by Dunn and extended by Bezdek is the most well-known and the most frequently used method [16]. The major algorithms include transitive closure method based on fuzzy equivalence relation, the method based on similarity relation and fuzzy relation, and most tree method based on fuzzy graph theory. However, these methods have the high

complexity of calculation and are not suitable for large data; therefore, they have been gradually reduced in research. The fuzzy clustering algorithm based on objective function is widely studied. Within-Groups Sum of Squared Error is used to construct the objective function. This algorithm is simple and effective, also supported by classical theory. Fuzzy  $c$ -means clustering algorithm is the most widely used fuzzy clustering algorithm based on objective function [17].

Let  $X = \{x_1, x_2, \dots, x_n\}$  be sample dataset with  $n$  samples, and each sample  $x_k$  has  $s$  properties; namely,  $x_k = \{x_{k1}, x_{k2}, \dots, x_{ks}\}$ ; then the matrix of the sample dataset can be expressed as follows:

$$X = (x_{ns}), \quad n = 1, 2, \dots, n; \quad s = 1, 2, \dots, s. \quad (1)$$

The objective function of FCM clustering algorithm  $J_m(U, V)$  is defined as [18–20]

$$J_m(U, V) = \sum_{k=1}^n \sum_{i=1}^c (\mu_{ik})^m d_{ik}^2(x_k, v_i), \quad (2)$$

where  $c$  is clustering category,  $2 \leq c \leq n$ ;  $U$  is matrix of membership degree,  $U = \{\mu_{ik}\}$ ,  $i = 1, 2, \dots, c$ ,  $k = 1, 2, \dots, n$ ;  $V$  is matrix of cluster center,  $V = \{v_i\}$ ,  $i = 1, 2, \dots, c$ ;  $\mu_{ik}$  is membership degree of the  $k$ th element in the  $i$ th clustering category;  $v_i$  is cluster center of the  $i$ th clustering category,  $v_i = \{v_{i1}, v_{i2}, \dots, v_{is}\}$ ,  $i = 1, 2, \dots, c$ ;  $m$  is weighting exponent and empirical value is  $1.5 \leq m \leq 2.5$ ;  $d_{ik}(x_k, v_i)$  is the Euclidean distance of target data between  $x_k$  and  $v_i$ .

The basic idea of FCM algorithm is to find fuzzy matrix of membership degree and cluster center, making the objective function minimum.

Membership degree function satisfies the following equations:

$$\begin{aligned} \sum_{i=1}^c (\mu_{ik})^m &= 1, \quad \forall k, \\ \sum_{k=1}^n (\mu_{ik})^m &\geq 0, \quad \forall i, \end{aligned} \quad (3)$$

$$\mu_{ik} \in [0, 1], \quad \forall i, k.$$

According to the Lagrange multiplier optimization algorithm, the update formulas of membership degree and cluster center are expressed as

$$\mu_{ik} = \frac{1}{\sum_{l=1}^c [d_{ik}(x_k, v_i)/d_{lk}(x_k, v_l)]^{2/(m-1)}}, \quad (4)$$

$$i = 1, 2, \dots, c, \quad k = 1, 2, \dots, n,$$

$$v_i = \frac{\sum_{k=1}^n (\mu_{ik})^m x_k}{\sum_{k=1}^n (\mu_{ik})^m}, \quad i = 1, 2, \dots, c. \quad (5)$$

FCM algorithm is described as follows.

*Step 1.* Give the number of clustering categories  $c$  ( $2 \leq c \leq n$ ). Set the number of iterations  $k = 1$ , weighting exponent  $m$ , and stop parameters  $\varepsilon$ .

Step 2. Randomly select cluster centers  $v_i(k)$ ,  $i = 1, 2, \dots, c$ .

Step 3. According to formula (4), update the subjection function.

Step 4. According to formula (5), update the cluster center and set the updated cluster center  $v_i(k + 1)$ .

Step 5. If  $\|v_i(k + 1) - v_i(k)\| > \epsilon$ , then turn to Step 3,  $k = k + 1$ ; otherwise, quit the loop and get the clustering results.

Although FCM clustering algorithm as a classical algorithm is widely applied in a variety of key areas [21–23], it has its own shortcomings and deficiencies. On one hand, the initial cluster centers are randomly selected, which leads to a possible local convergence rather than the global optimal solution; on the other hand, the objective function of FCM algorithm considers only the distance between sample data and cluster centers, but it ignores the effect of the distance between cluster centers on the solution. To solve the problem above, a variety of different improved algorithms are proposed [24–26]. In this paper, an improved FCM algorithm is proposed and applied to make a red tide prediction.

### 3. An Improved Fuzzy $c$ -Means Clustering Algorithm

3.1. Selection of the Initial Cluster Centers. The selection principle of the initial cluster center is to make the initial cluster center within a certain threshold contain more data. This not only ensures that the clustering algorithm finds the cluster centers in a number of feasible regions, but also effectively reduces the impact of the noise and outliers on the objective function.

Let  $X = \{x_1, x_2, \dots, x_n\}$  be the sample dataset, the regional minimum threshold  $a$ , and regional minimum data density  $b$ ; according to the principle above, the selection steps of initial cluster centers are as follows.

Step 1. Calculate Euclidean distance between any two samples in the dataset to generate distance matrix  $A$ . According to the distance matrix  $A$ , choose the nearest two samples.

Step 2. With the center of the two samples as center and the regional threshold  $a$  as radius, plan  $a$  circular region. If the data density in the region is equal or greater than minimum data density  $b$  in this region, this center is chosen as a kind of initial cluster center; otherwise the region is removed.

Step 3. Choose the nearest two samples in the remaining samples outside the region. Repeat Step 2 until  $c$  classes are found. If the selected classes are less than  $c$ , the criteria of  $a$  and  $b$  will be relaxed.

Figure 1 shows the selection process of the initial cluster center. According to the distance matrix, the two red samples have the nearest distance among all the samples. However the data density in its region is small, so the red samples are removed from the sample data. In the rest of the sample data,

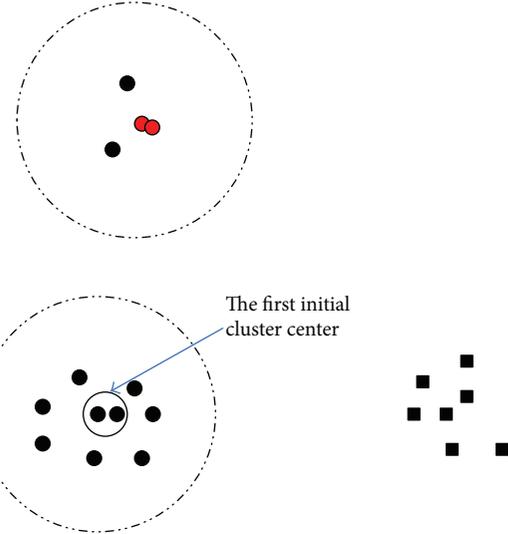


FIGURE 1: The selection of the initial cluster centers through the principle of regional minimum data density.

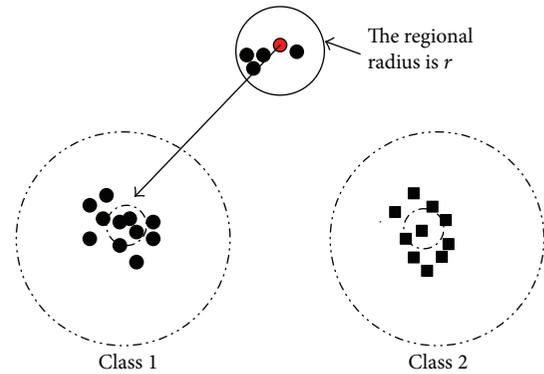


FIGURE 2: The selection of the initial cluster centers through the principle of the minimum mean distance.

the initial cluster centers are continued to look for until  $c$  classes are found.

According to formula (4), the membership degree of the element  $x_i$  in the  $i$ th clustering category is determined by the relative ratio of  $d_{ij}$  and  $d_{cj}$  ( $c \neq i$ ), but the real distance  $d_{ij}$  is not reflected in the solution of subjection function.

As shown in Figure 2, the red data point is located on the bisecting line of two datasets below it. It is obviously difficult to distinguish the class of the red data point using the traditional FCM subjection function. The principle of the mean distance minimum is proposed to classify the boundary data: firstly, with the boundary red data point as center,  $r$  as radius, plan  $a$  circular region; secondly, respectively, calculate the mean value of the distance between the points in the circular region and the two cluster centers; finally, the boundary red data point is assigned to the class where the mean value of the distance is smaller.

3.2. Improvement of the Objective Function. As the most common analysis method of FCM, fuzzy clustering algorithm

based on dissimilar objective function only considers weighting Euclidean distance between sample data and cluster centers, without taking into account the distance between each cluster center [27]. Therefore, a weighting fuzzy  $c$ -means based on dissimilar objective function (DWFCM) is proposed and weighted distance between cluster centers is added to dissimilar objective function according to the effect of the distance between cluster centers on clustering results.

Dissimilar objective function with distance between cluster centers is defined as [28]

$$J_m(U, V, \eta) = \sum_{k=1}^n \sum_{i=1}^c (\mu_{ik})^m d_{ik}^2(x_k, v_i) - \frac{1}{c(c-1)} \sum_{i=1}^c \sum_{j=1}^c (\eta_{ij})^m d_{ij}^2(v_i, v_j), \quad (6)$$

where  $\eta_{ij} = \beta(\min\{d_{ik}(v_i, x_k), d_{jk}(v_j, x_k)\} / \max\{d_{ab}(v_a, v_b)\})$ ,  $0 \leq \beta \leq 1$ ,  $a, b \in c$ .  $d_{ij}(v_i, v_j)$  is the Euclidean distance between cluster centers  $v_i$  and  $v_j$ .

According to the Lagrange multiplier optimization algorithm under the constraint conditions of formula (3), the update formulas of cluster center and membership degree are expressed as

$$\frac{\partial J_m}{\partial v_i} = 0 \implies v_i = \frac{\sum_{k=1}^n (\mu_{ik})^m x_k - (1/c(c-1)) \sum_{j=1}^c (\eta_{ij})^m v_j}{\sum_{k=1}^n (\mu_{ik})^m - (1/c(c-1)) \sum_{j=1}^c (\eta_{ij})^m}, \quad (7)$$

$$\mu_{ik} = \begin{cases} \frac{(d_{ik}^2(x_k, v_i))^{-1/(m-1)}}{\sum_{j=1}^c (d_{jk}^2(x_k, v_j))^{-1/(m-1)}} & d_{ik}^2(x_k, v_i) > 0 \quad (1 \leq j \leq c) \\ 1 & d_{ik}^2(x_k, v_i) = 0 \quad (1 \leq j \leq c) \\ 0 & j \neq i, d_{jk}^2(x_k, v_j) = 0. \end{cases} \quad (8)$$

Formula (6) adds the effect of the cluster centers on clustering results. This optimization obtains the minimum weighting Euclidean distance between sample data and cluster centers and the maximum Euclidean distance between cluster centers. However, this optimization does not consider the effect of the distance between cluster centers on clustering results. Therefore, DWFCM algorithm adds feature weight to Euclidean distance among cluster centers.

In order to obtain the feature weight, the similarity coefficient  $r_{ij}$  between the cluster centers should be calculated. There are several common methods as follows [29].

(1) The correlation coefficient method is defined as

$$r_{ij} = \frac{\sum_{p=1}^s |v_{ip} - \bar{v}_i| |v_{jp} - \bar{v}_j|}{\left[ \sum_{p=1}^s (v_{ip} - \bar{v}_i)^2 \sum_{p=1}^s (v_{jp} - \bar{v}_j)^2 \right]^{1/2}}, \quad (9)$$

where  $\bar{v}_i = (1/s) \sum_{p=1}^s v_{ip}$ ,  $\bar{v}_j = (1/s) \sum_{p=1}^s v_{jp}$ ,  $1 \leq i, j \leq c$ .  $v_{ip}$ ,  $v_{jp}$  are the  $p$ th attribute of the  $i$ th,  $j$ th cluster center.

(2) The least arithmetic average method is defined as

$$r_{ij} = \frac{2 \sum_{p=1}^s \min(v_{ip}, v_{jp})}{\sum_{p=1}^s (v_{ip} + v_{jp})}. \quad (10)$$

(3) The angle cosine method is defined as

$$r_{ij} = \frac{\sum_{p=1}^s v_{ip} v_{jp}}{\sqrt{\sum_{p=1}^s v_{ip}^2} \sqrt{\sum_{p=1}^s v_{jp}^2}}. \quad (11)$$

The correlation coefficient method is used to calculate the similarity coefficient  $r_{ij}$  in the DWFCM algorithm.

The feature weight  $L_{ij}$  is defined as

$$L_j = \sum_{p=1}^c r_{pj}, \quad (12)$$

$$L_{ij} = \frac{r_{ij}}{L_j}.$$

The objective function of DWFCM is defined as

$$J_m(U, V, \eta) = \sum_{k=1}^n \sum_{i=1}^c (\mu_{ik})^m d_{ik}^2(x_k, v_i) - \frac{1}{c(c-1)} \sum_{i=1}^c \sum_{j=1}^c L_{ij} (\eta_{ij})^m d_{ij}^2(v_i, v_j). \quad (13)$$

According to the Lagrange multiplier optimization algorithm under the constraint conditions of formula (3), formula (7) is modified as follows:

$$v_i = \frac{\sum_{k=1}^n (\mu_{ik})^m x_k - (1/c(c-1)) \sum_{j=1}^c L_{ij} (\eta_{ij})^m v_j}{\sum_{k=1}^n (\mu_{ik})^m - (1/c(c-1)) \sum_{j=1}^c L_{ij} (\eta_{ij})^m}. \quad (14)$$

Considering the unit of sample data is not unified and the data is not complete, the data are normalized before the sample is classified. The normalized functions are as follows:

$$x'_{ns} = \frac{x_{ns} - \bar{x}_n}{s_n}, \quad (15)$$

where  $\bar{x}_n = (1/s) \sum_{j=1}^s x_{ns}$ ,  $s_n = \sqrt{(1/(s-1)) \sum_{j=1}^s (x_{ns} - \bar{x}_n)^2}$ .

**3.3. The Description of Improved FCM Algorithm.** The steps of improved FCM algorithm are as follows.

*Step 1.* Set the regional threshold  $a$ , regional density  $b$ , regional radius  $r$ , category number  $c$ , iteration  $k = 1$ , weighting exponent  $m$ , and stop parameters  $\varepsilon$ .

*Step 2.* According to the number of regional thresholds, regional density, and regional radius, accomplish the selection of  $c$  initial cluster centers.

*Step 3.* Calculate the correlation coefficient  $r_{ij}$  and feature weight  $L_{ij}$  among cluster centers.

*Step 4.* According to formulas (8) and (14), update the subjection function and the cluster center and set the updated cluster center  $v_i(k+1)$ . Normalize the original sample.

*Step 5.* If  $\|v_i(k+1) - v_i(k)\| > \varepsilon$ , then turn to Step 3,  $k = k + 1$ ; otherwise, quit the loop and get the subjection function  $\mu_{ik}$ .

*Step 6.* The samples are classified according to the following methods:

$$d_{ik}(v_i, x_k) \leq d_{jk}(v_j, x_k), \quad (16)$$

$$k = 1, 2, \dots, n, \quad i, j = 1, 2, \dots, c.$$

In this case, the sample  $x_k$  is classified into the  $i$ th class.

#### 4. Red Tide Prediction Analysis Based on Improved FCM Algorithm

*Example 1.* Some changes of the marine environment factors accompany the process of the red tide from happening to extinction. The original sample set is composed of marine environment factors which have a great influence on the red tide. FCM, possibilistic fuzzy  $c$ -mean algorithm (PCM), and DWFCM are used to classify the original sample set. Considering that the increase of nitrogen concentration is the key factor of eutrophication of the seawater, meanwhile, the eutrophication of the seawater is the primary condition of the occurrence of red tide, and the change of phytoplankton density is also an important indicator to measure the occurrence of red tide; nitrogen concentration ( $\mu\text{mol/L}$ ) and phytoplankton density ( $10^4/\text{cubic meter}$ ) are chosen as the research factors. 21 original samples from the State Oceanic Administration are shown in Table 1 where the last sample is used for forecasting and other samples are used for cluster analysis.

Before clustering, the normalized processing of original samples is completed. If the horizontal axis represents nitrogen concentration and the vertical axis represents the phytoplankton density, the simulation results of three different algorithms are shown in Figures 3, 4, and 5, and the comparison of clustering results is shown in Table 2.

In Figures 3, 4, and 5, the samples where red tide occurs are marked as the inverted triangle. The samples where red tide will occur are marked as the asterisk. The samples where red tide does not occur are marked as rhombic. The red dots represent cluster centers of samples. In Table 2, the error score is the number of the misclassified data pieces, and the error rate is the percentage of the error score in the total data. From Figures 3, 4, and 5 and Table 1, DWFCM has the best performance among three algorithms: the selection of clustering center is the most reasonable; the number of iterations is also the least; the accuracy is also the best.

*Example 2.* In order to sufficiently demonstrate superiority of the proposed optimization model in this paper, this example chooses another original sample set which also have a great influence on the red tide. Considering that the water temperature is the key factor of plankton growth speed and the transparency can be used to evaluate the density

TABLE 1: The original sample set with 21 elements.

Sample	Nitrogen concentration ( $\mu\text{mol/L}$ )	Phytoplankton density ( $10^4/\text{cubic meter}$ )
1	0.4	148
2	0.54	109
3	0.1	71.5
4	0.07	31.8
5	0.54	109
6	0.31	83
7	0.42	113.1
8	0.21	93.3
9	0.18	31.1
10	0.21	97.2
11	0.4	74.5
12	0.36	125
13	0.56	187
14	0.39	146
15	0.07	32
16	0.47	205
17	0.43	136
18	0.43	147
19	0.26	160
20	0.32	105.4
21	0.37	103.7

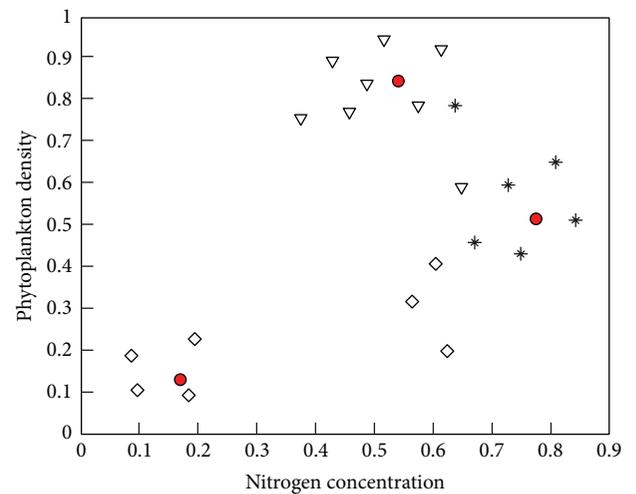


FIGURE 3: The clustering results of FCM algorithm in Example 1.

of plankton, water temperature ( $^{\circ}\text{C}$ ) and transparency (m) are chosen as the research factors. 32 original samples are shown in Table 3 where the last sample is used for forecasting and other samples are used for cluster analysis. The specific clustering results and the analysis of clustering results are shown in Tables 3 and 4 and Figures 6, 7, and 8.

As to the FCM algorithm, the sum of the membership degree of the same sample belonging to all categories is 1,

TABLE 2: The comparison of clustering results in Example 1.

	FCM	PCM	DWFCM
Clustering center	(0.55, 0.84) (0.78, 0.50) (0.18, 0.12)	(0.56, 0.81) (0.85, 0.43) (0.16, 0.19)	(0.54, 0.86) (0.81, 0.55) (0.17, 0.18)
Iteration number	11	13	8
Error score	4	3	1
Error rate	20%	15%	5%

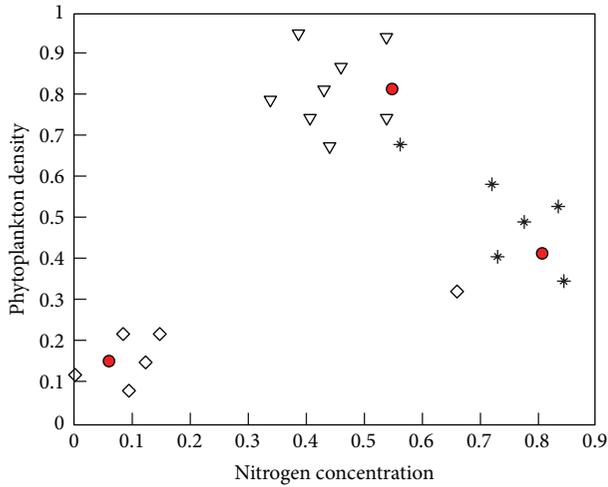


FIGURE 4: The clustering results of PCM algorithm in Example 1.

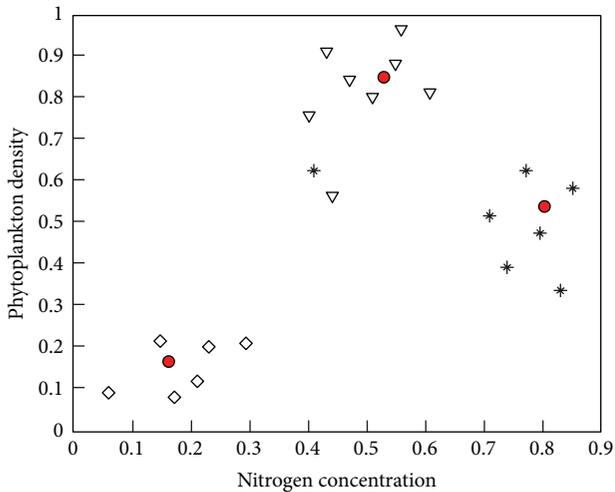


FIGURE 5: The clustering results of DWFCM algorithm in Example 1.

which makes FCM algorithm sensitive to noise and outliers [30]. Figure 6 shows that FCM algorithm makes the wrong clustering division for some boundary samples. In PCM algorithm, the value of the membership degree reflects the real Euclidean distance between sample data and cluster centers, which makes PCM algorithm have good robustness to noise and outliers. Figure 7 shows that PCM algorithm makes the right clustering division for boundary samples, but there are still a large number of overlapping clusters.

TABLE 3: The original sample set with 32 elements.

Sample	Water temperature (°C)	Transparency (m)
1	26.47	1.5
2	26.22	1.2
3	24.47	1.2
4	24.2	3.5
5	25.15	1.2
6	24.2	2.5
7	23.7	2.2
8	26.6	0.9
9	26.5	0.4
10	26.5	0.5
11	26.1	0.6
12	26.2	1.5
13	26.5	0.9
14	24.9	1.8
15	25.0	2.5
16	25.0	2.5
17	25.8	1.1
18	26.1	1.8
19	26.2	1.5
20	25.4	1.2
21	27.4	5.0
22	27.6	2.1
23	27.0	2.0
24	26.8	4.0
25	26.8	5.0
26	26.8	4.5
27	26.6	0.9
28	26.5	0.4
29	26.5	0.5
30	26.1	0.6
31	26.2	1.5
32	27.5	3.9

TABLE 4: The comparison of clustering results in Example 2.

	FCM	PCM	DWFCM
Clustering center	(0.64, 0.10) (0.67, 0.48) (0.66, 0.84)	(0.65, 0.11) (0.68, 0.50) (0.69, 0.80)	(0.74, 0.12) (0.69, 0.39) (0.70, 0.82)
Iteration number	18	19	13
Error score	7	5	2
Error rate	22%	16%	6%

Because DWFCM algorithm uses the principle of regional minimum data density and the minimum mean distance to select the fixed initial cluster centers, which can effectively avoid the influence of noise and outliers, it also considered the influence of the weighted cluster center on the objective function, which makes the clustering results more accurate. After the last set is added to the clustering sample, the last set is clustered to the red tides cluster using the DWFCM

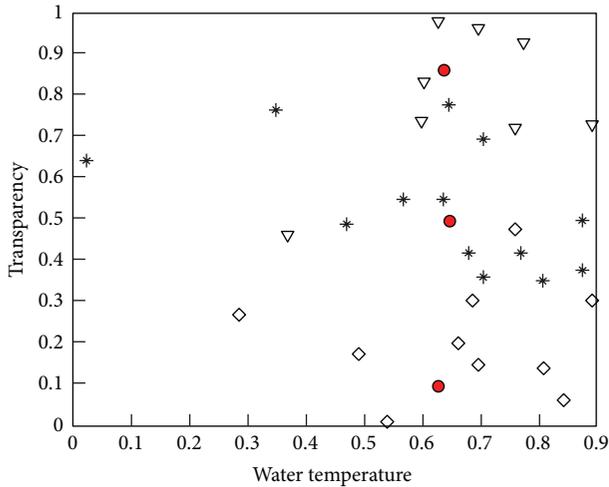


FIGURE 6: The clustering results of FCM algorithm in Example 2.

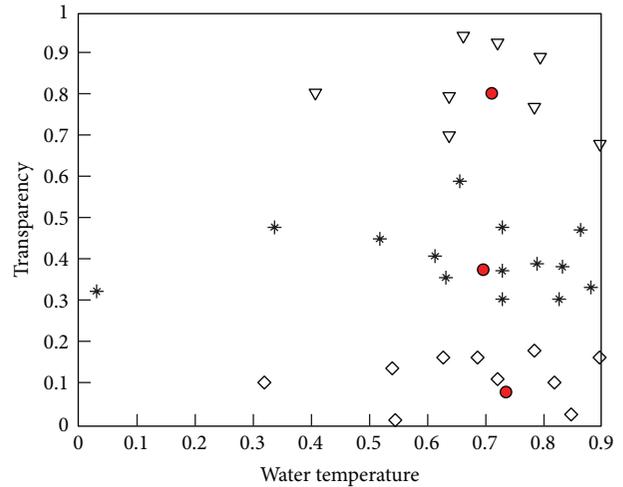


FIGURE 8: The clustering results of DWFCM algorithm in Example 2.

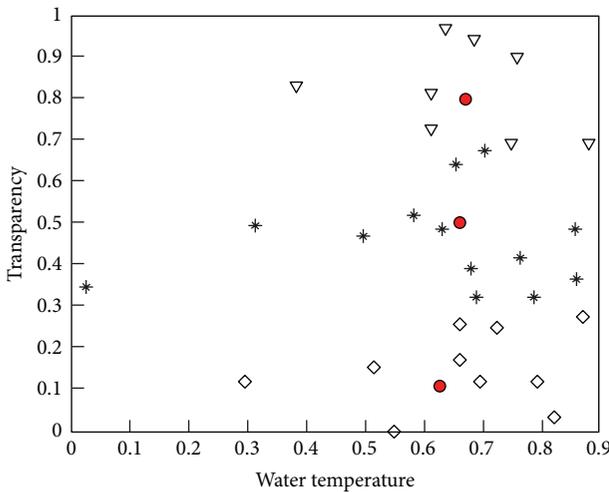


FIGURE 7: The clustering results of PCM algorithm in Example 2.

algorithm and the fact is that the last set is indeed the sample when red tide occurs. The simulation results show that DWFCM algorithm can effectively achieve the prediction of red tide disaster.

From Figures 6, 7, and 8 and Table 4, FCM algorithm and PCM algorithm randomly select the initial cluster centers, which make the number of iterations and the clustering results very different from each other. However, DWFCM algorithm has selected the fixed initial cluster centers before the iteration, so the number of iterations is significantly reduced and the clustering results are unified.

### 5. Conclusion

In view of the complexity of the red tide disaster and the shortage of the previous prediction algorithm, a DWFCM algorithm is proposed to predict red tide. The initial cluster centers are chosen by the principle of regional minimum data density and the minimum mean distance, and the objective

function is optimized by using the weighted cluster center. The simulation results show that DWFCM algorithm has better denoising ability and can optimize the prediction model of red tide disaster and get more accurate predictive results. However, DWFCM algorithm introduces many parameters. In order to get accurate parameter values, a lot of experiments have to be done. Therefore, the study of combining DWFCM algorithm with other algorithms to overcome the defects in the DWFCM algorithm is the focus of study in the future.

### Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

### Acknowledgment

This work was supported by the Grand Science & Technology Program Shanghai China (no. 14DZ1100700).

### References

- [1] W. Hong-Li and F. Jian-li, *Ecological Dynamics and Prediction of Red Tide*, Tianjin University Press, Tianjin, China, 2006.
- [2] S.-Z. Feng, F.-Q. Li, and S.-Q. Li, *Introduction to Marine Science*, Higher Education Press, Beijing, China, 1999.
- [3] G. A. Gibson, D. L. Musgrave, and S. Hinckley, "Non-linear dynamics of a pelagic ecosystem model with multiple predator and prey types," *Journal of Plankton Research*, vol. 27, no. 5, pp. 427–447, 2005.
- [4] J. F. Zhang, Y. Bai, J. Yu et al., "Forecast of red tide in the South China Sea by using the variation trend of hydrological and meteorological factors," *Marine Science Bulletin*, vol. 8, no. 2, pp. 60–74, 2006.
- [5] H.-L. Wang, J.-P. Xiang, and G. Ge, "Multivariate adaptive spline regression prediction of total phytoplankton," *Marine Technology*, vol. 25, no. 3, pp. 7–9, 2006.

- [6] R. Zhang, H. Yan, and L. P. Du, "Research on prediction of red tide based on fuzzy neural network," *Bulletin of Marine Science*, vol. 8, no. 1, pp. 83–91, 2006.
- [7] H. Wilson and F. Recknagel, "Towards a generic artificial neural network model for dynamic predictions of algal abundance in freshwater lakes," *Ecological Modelling*, vol. 146, no. 1–3, pp. 69–84, 2001.
- [8] W.-L. Huang and D.-W. Ding, *Prediction Mechanism and Technology of Red Tide Disaster*, Ocean Press, Beijing, China, 2004.
- [9] F. Jin-Qing, "Attention to the complexity of science and research," *Nature Magazine*, vol. 24, no. 1, pp. 7–14, 2002.
- [10] F. Jian-Feng, *Study on the Nonlinear Dynamics of the Planktonic Ecosystem and the Prediction of Red Tide*, Engineering Mechanics of Tianjin University, 2005.
- [11] D.-M. Guan and X.-W. Zhan, "Red tide disasters in the coastal waters in China and its countermeasures," *Marine Environmental Science*, no. 2, pp. 60–63, 2003.
- [12] C.-S. Wang, S.-M. Tang, and P.-Q. Song, "Economic loss assessment of red tide disaster in China," *Marine Environmental Science*, no. 3, pp. 428–431, 2011.
- [13] X.-L. Wang, P.-Y. Sun, and Z.-H. Gao, "The status and progress of the prediction of red tide in China," *Progress in Marine Science*, vol. 21, no. 1, pp. 93–98, 2003.
- [14] H. L. Wang, J. F. Feng, S. P. Li, and F. Shen, "Statistical analysis and prediction of the concentration of harmful algae in bohai bay," *Transactions of Tianjin University*, vol. 11, no. 4, pp. 308–312, 2005.
- [15] X. Wei-Yi and Z. De-Di, "Numerical simulation of the process of the red tide in the real sea area," *Oceanologia*, vol. 32, no. 6, pp. 598–604, 2001.
- [16] K.-L. Wu and M.-S. Yang, "Alternative c-means clustering algorithms," *Pattern Recognition*, vol. 35, no. 10, pp. 2267–2278, 2002.
- [17] N. Yong, *The application of data mining to marine environment online monitoring and HVB prediction system [M.S. thesis]*, Shandong University, Jinan, China, 2008.
- [18] M. N. Ahmed, S. M. Yamany, N. Mohamed, A. A. Farag, and T. Moriarty, "A modified fuzzy C-means algorithm for bias field estimation and segmentation of MRI data," *IEEE Transactions on Medical Imaging*, vol. 21, no. 3, pp. 193–199, 2002.
- [19] N. R. Pal, K. Pal, J. M. Keller, and J. C. Bezdek, "A possibilistic fuzzy c-means clustering algorithm," *IEEE Transactions on Fuzzy Systems*, vol. 13, no. 4, pp. 517–530, 2005.
- [20] H. Timm and R. Kruse, "A modification to improve possibilistic fuzzy cluster analysis," in *Proceedings of the IEEE International Conference on Fuzzy Systems (FUZZ-IEEE '02)*, vol. 2, pp. 1460–1465, IEEE, Honolulu, Hawaii, USA, May 2002.
- [21] X. Li, X. Lu, J. Tian, P. Gao, H. W. Kong, and G. W. Xu, "Application of fuzzy c-means clustering in data analysis of metabolomics," *Analytical Chemistry*, vol. 81, no. 11, pp. 4468–4475, 2009.
- [22] H.-K. Xu, J.-H. Chuai, Z.-H. Zhang, and H.-W. Fan, "Data mining of traffic flow in road tunnel," *Journal of Chang'an University*, vol. 25, no. 4, pp. 66–69, 2005.
- [23] F. Yu, H. Xu, L. Wang, and X. Zhou, "An improved automatic FCM clustering algorithm," in *Proceedings of the 2nd International Workshop on Database Technology and Applications (DBTA '10)*, pp. 1–4, IEEE, Wuhan, China, November 2010.
- [24] J. Yu and H. Huang, "A new weighting fuzzy c-means algorithm," in *Proceedings of the IEEE International Conference on Fuzzy System (FUZZ '03)*, pp. 896–901, May 2003.
- [25] J. M. Leski, "Generalized weighted conditional fuzzy clustering," *IEEE Transactions on Fuzzy Systems*, vol. 11, no. 6, pp. 709–715, 2003.
- [26] J. Li, X.-B. Gao, and L.-C. Jiao, "New fuzzy clustering algorithm based on feature weighting," *Electronic Journal*, vol. 34, no. 1, pp. 89–92, 2006.
- [27] W. Zebing and C. Baozhen, "The study of an improved FCM clustering algorithm," in *Proceedings of the 2nd International Conference on Signal Processing Systems (ICSPS '10)*, vol. 2, pp. V2-96–V2-100, IEEE, Dalian, China, July 2010.
- [28] J. F. Zhang, Y. Bai, J. Yu et al., "Forecast of red tide in the South China sea by using the variation trend of hydrological and meteorological factors," *Marine Science Bulletin*, vol. 8, no. 2, pp. 60–74, 2006.
- [29] W. Jin-Pei and S. De-Shan, *Modern Data Analysis*, Machinery Industry Press, Beijing, China, 2006.
- [30] N. R. Pal, K. Pal, J. M. Keller, and J. C. Bezdek, "A possibilistic fuzzy c-means clustering algorithm," *IEEE Transactions on Fuzzy Systems*, vol. 13, no. 4, pp. 517–530, 2005.



# Hindawi

Submit your manuscripts at  
<http://www.hindawi.com>

