Desalination and Water Treatment
www.deswater.com
doi: 10.5004/dwt.2018.22085

106 (2018) 285–291
February

Water quality forecasting of Haihe River based on improved fuzzy time series model

Ranran Li*

School of Business Administration, Shandong Technology and Business University, Yantai 264005, China, email: liranran1101@163.com

Received 24 April 2017; Accepted 19 February 2018

ABSTRACT

Fuzzy time series (FTS) forecasting model has both advantages of fuzzy theory and time series, and can overcome the problem that historical data is inaccurate or incomplete. It has been applied in many fields and has achieved lots of good results, but few studies have paid attention to the application in water quality forecasting. This study introduced the method to short-term forecasting of water quality. Some improvements on the calculation process of FTS forecasting was made and the improved fuzzy time series (IFTS) model was proposed. Potassium permanganate index (COD_{Mn}) and dissolved oxygen (DO) concentrations of Sanchakou Station of Haihe River were predicted by this proposed method. FTS and modified GM (1,1) methods were also used for prediction. Through a series of contrast and analysis, it is concluded that this proposed IFTS method has obvious advantages than modified GM (1,1) method in prediction of data with no apparently monotonous trend, and better than the FTS method. Therefore, it is a good water quality forecasting method and can be widely used in short-term water quality forecasting. The prediction results provided an important theoretical basis for Haihe River water quality management.

Keywords: Water quality forecasting; Fuzzy time series; Fuzzy c-means; Grey prediction

1. Introduction

Surface water quality analysis and forecasting is quite important for current water quality management. Accurate prediction of water quality is the basis of water quality control, plans setting, and water quality incidents management. Therefore, research on various methods of prediction of water quality has important theoretical values and practical significances. There are already numerous water quality forecasting methods. Water quality simulation method [1,2], regression analysis [3], time series [4], neural network method [5,6] and grey theory [7], are commonly used with positive results in sewage treatment, drinking water management, and various other fields. However, there are some disadvantages with each method. On account of the variety of type and quantity of pollutants in water, as well as complexity of water dilution and self-purification, water quality simulation method is most based on the random water quality model. When the amount of considering random variable is large, the calculation accuracy is poor. Regression analysis, time series and neural network method usually need a lot of historical time series, which are often difficult to be obtained. Grey theory prediction usually adopts GM (1,1) model. GM (1,1) model requires less data, and is a widely used prediction method. Many scholars improved this method from initial condition, background value and other aspects, consequently further improved the precision. Even so, the grey prediction is monotonic; its performance is not satisfactory for data with no apparently monotonic trend. Water quality concentrations have no obviously monotonous trend in most cases, so more reasonable prediction method is required.

In recent years, many researchers have combined fuzzy theory and time sequence, and gained fuzzy time series (FTS) forecasting model. This model has both advantages of fuzzy theory and time series, and can overcome the problem that historical data is inaccurate or incomplete. FTS forecasting has been applied in economic, finance, science and other fields, and has achieved lots of good results [8–17]. Tsaur [18] used adaptive fuzzy time series model

*Corresponding author.

1944-3994 / 1944-3986

to forecast tourism demand. Wang [19] adopted two-factor high-order fuzzy time series model for temperature prediction. Cheng [20] applied weighted fuzzy time series model to forecast innovation diffusion of products. Besides, there have been many applications in stock price prediction [21–23], exchange rates prediction [24,25], export amount prediction [26] and dry bulk shipping index prediction [27]. However, few studies have paid attention to the application of fuzzy time series in water quality forecasting. As water environmental system has a lot of uncertainty, and other fuzzy theory has achieved pleasurable results in water quality assessment and other aspects [28–34], here considering applying fuzzy time series to short-term prediction of water quality. In this article, fuzzy time series was applied in water quality forecasting, and the realization of fuzzy time series forecasting method was improved. Modified GM (1,1) method was also used. By a series of contrast and analysis, it is concluded that this improved method has a higher prediction precision.

2. Materials and methods

2.1. Study area and data source

Haihe River is one of the seven major rivers in China. The basin area is 2.65×10^6 km^2. It is the largest water system in north China. The main stream of Haihe River is the access to the sea for Ziya, Daqing, Nanyun, Beiyun and Yongding River, at the same time it has functions of waterlogging drainage, water storage, water supply, shipping, tourism and environmental protection [35,36]. The water quality management of Haihe River has gained national and local government's great attention. Haihe River is in Beijing-Tianjin-Tangshan industrial zone, where heavy industry is developed. It is also located in the place of severe water shortages, with dry climate, less precipitation, more evaporation, and less river runoff. So Haihe River is a serious polluted river. According to the past investigation and research, the main pollution factors in the water are potassium permanganate index (COD_{Mn}), dissolved oxygen (DO) and ammonia nitrogen (NH_3-N). To understand the evolution trend of Haihe River environmental quality, and thus to discover the cause of the deterioration of water quality in time and take the corresponding control measures, the concentrations of the main pollution factors in Haihe River should be predicted accurately. To verify the effectiveness of the proposed method, concentrations of COD_{Mn} and DO are predicted in this study. The prediction is based on the monitoring data of Sanchakou Station of Haihe River. Sanchakou Station is located at Ping'an Street in Hebei District in Tianjin, and on the east of Haihe River. Its coordinates are 117°11′21.9″E, 39°08′22.5″N. The specific location is shown in Fig. 1. The data were from Ministry of Environmental Protection of the People's Republic of China.

Fig. 1. The map of Haihe River Basin and location of Sanchakou Station.

2.2. Fuzzy time series model

Let subset $Y(t) \subset R(t = 0,1,\ldots)$ be a given domain, $f_i(t)$ $(i = 1,2,\ldots)$ are fuzzy sets in the domain, $F(t)$ is a set composed of $f_i(t)(i = 1,2,\ldots)$, then $F(t)$ is called fuzzy time series defined on $f_i(t)(i = 1,2,\ldots)$. Other related concepts about fuzzy are not elaborated in detail here, which can be seen in many researches [37–39]. According to Refs. [8–10], the forecasting framework consists of four steps: 1) definition and division of domains; 2) definition of fuzzy sets and fuzzy data; 3) establishment of fuzzy relationship; 4) defuzzification and forecasting.

2.2.1. Definition and division of domains

Let the time series be $X(t) = \{x_1, x_2, \ldots, x_n\}$, x_t is any data point in the sample data. Let U be the domain, $U = [D_{min} = \alpha_1, D_{max} + \alpha_2]$, where D_{min} and D_{max} are the minimum and maximum values of the sample series, respectively, α_1 and α_2 are appropriate constants. The domain was divided into several equal intervals, which are $u_1, u_2, \ldots, u_k$.

2.2.2. Definition of fuzzy sets and fuzzy data

The domain U was divided into k subspaces, there is $U = [u_1, u_2, \ldots, u_k]$. Then $A_t = (u_1(x_t), u_2(x_t), \ldots, u_k(x_t))$ is fuzzy set in domain U. $u_1(x_t), u_2(x_t), \ldots, u_k(x_t)$ are the membership degrees of corresponding interval $u_1, u_2, \ldots, u_k$ for x_t. If x_t belongs to u_i, $1 \le i \le k$, $u_j(x_t)$, is the membership degree of corresponding interval u_j for x_t, there is

$$u_j(x_t) = \begin{cases} 0, & j < i-1 \\ 0.5, & j = i-1 \\ 1, & j = i \\ 0.5, & j = i+1 \\ 0, & j > i+1 \end{cases} \tag{1}$$

The membership degree of x_t to u_j on both sides decreases gradually.

2.2.3. Establishment of fuzzy relationship

Next w weeks' data are used to establish the fuzzy relationship, so fuzzy relations of w-order model can be got.

Assuming that A_i^{t-w}, A_j^{t-w+1}, …, A_l^{t-1} are the corresponding fuzzy sets of $F(t-w)$, $F(t-w+1)$, …, $F(t-1)$, respectively [40]. If there is

$$F(t-w), F(t-w+1), \ldots, F(t-1) \rightarrow F(t), \tag{2}$$

then

$$A_i^{t-w}, A_j^{t-w+1}, \cdots A_l^{t-1} \rightarrow A_m^t \left(i,j,l,m = 1,2,\cdots,k\right). \tag{3}$$

Define $C(t) = f(t-1) = [C_1, C_2, \ldots, C_k]$ be the standard vector of fuzzy time series $F(t)$, $f(t-1)$ is the fuzzified variation of the first factor $F(t)$ between times $t-1$ and $t-2$; k is the number of intervals in the domain. The operation matrix of $F(t)$ is

$$O^w(t) = \begin{bmatrix} f(t-2) \\ f(t-3) \\ \vdots \\ f(t-w) \end{bmatrix} = \begin{bmatrix} O_{11} & O_{12} & \cdots & O_{1k} \\ O_{21} & O_{22} & \cdots & O_{2k} \\ \vdots & \vdots & \vdots & \vdots \\ O_{(w-1)1} & O_{(w-1)2} & \cdots & O_{(w-1)k} \end{bmatrix} \tag{4}$$

The fuzzy relations $R(t)$ is

$$R(t) = O^w(t) \otimes C(t) = \begin{bmatrix} R_{11} & R_{12} & \cdots & R_{1k} \\ R_{21} & R_{22} & \cdots & R_{2k} \\ \vdots & \vdots & \vdots & \vdots \\ R_{(w-1)1} & R_{(w-1)2} & \cdots & R_{(w-1)k} \end{bmatrix} \tag{5}$$

where C_j, $O_{ij} \in [0,1]$, $i \in [1,w-1]$, $j \in [1,k]$, $R_{ij} = O_{ij} \times C_j$. According to $R(t)$, the forecasting membership of time t is

$$f(t) = [\max(R_{11}, \ldots, R_{(w-1)1}), \ldots, \max(R_{1k}, \ldots, R_{(w-1)k})] = [f_{t1}, \ldots, f_{tk}] \tag{6}$$

2.2.4. Defuzzification and forecasting

The defuzzification of the result uses gravity method:

$$F(t) = F_t = \sum_{i=1}^{k} c_i f_{ti} \Big/ \sum_{i=1}^{k} f_{ti} \tag{7}$$

where c_i is the center of each interval, f_{ti} is the membership, F_t is predicted value, k is the number of the intervals.

2.3. Improved fuzzy time series model

The distribution of data is usually not uniform in practice, so it is more reasonable to divide domain into several unequal intervals than equal ones. Fuzzy c-means (FCM) is an excellent method in unequal classification. It has been applied in many fields [41,42]. But in this method, how to determine the optimal clustering number has not got the consistent conclusion. For short-term forecasting, under the condition of small amount of data, it can be clustered for different given clustering number, respectively. And then compare forecasting results through the root mean square error (RMSE). Select the clustering number when RMSE is the smallest as the optimal clustering number, and choose this forecasting result as the final result [43].

As the distribution of sample points is usually not uniform, when the difference between sample data is not too large, some sample may be in the same range. Then according to the previous fuzzification method, different sample data may be fuzzed into the same fuzzy set, which cannot make full use of the sensitivity of the data, and result in low forecasting accuracy. Therefore, reasonable definition of fuzzy sets and fuzzy data is particularly important. Here a new data fuzzification method based on distance [44] is proposed.

2.3.1. Definition and division method of domains based on FCM

Let the time series be $X(t) = \{x_1, x_2, \ldots, x_n\}$. Cluster the sample data according to the FCM method. For each clustering number $k_i \in N^+$, there are k_i clustering centers, which are $c_1, c_2, \ldots, c_{k_i}$. Choose the halfway point of the adjacent two clustering centers as a boundary point of domain division. The boundary points are dented as $d_2, d_3, \ldots, d_{k_i}$. Divide the domain U into k_i intervals, which are $u_1, u_2, \ldots, u_{k_i}$, then there is $u_1 = [D_{\min} + \sigma_1, d_2]$, $u_2 = [d_2, d_3]$, , $u_{k_i} = [d_{k_i}, D_{\max} + \sigma_2]$, where $D_{\min}$, $D_{\max}$ are the minimum and maximum values of the sample series, respectively, σ_1, σ_2 are appropriate constants.

Let k_m be the maximum number of clustering sample sequence. $k_m = [\sqrt{n}]$, where n is the total number of sample, [•] denotes INTPART. For different clustering numbers 2, 3…, k_m, different forecasting results can be obtained according to the previous steps respectively. And then each corresponding RMSE can be calculated:

$$RMSE = \sqrt{\sum_{t=1}^{n} (X(t) - F(t))^2 / n} \tag{8}$$

where $X(t)$, $F(t)$ are the actual and forecasted values of sample data at time t. Then choose the set with smallest RMSE as the last forecasting results.

2.3.2. Definition of fuzzy sets and fuzzy data based on distance

Let $d_1 = D_{min} + \sigma_1$, $d_{k_i+1} = D_{max} + \sigma_2$, then the boundary points of divided domains are $d_1, d_2, \ldots, d_{k_i+1}$. If x_t belongs to u_i, there is $1 \le i \le k_i$, let $u_i(x_t) = 1$. Otherwise, there is $1 \le j \le k_i$ and $j \ne i$, $1 \le t \le n$, let $d_{\min} = \min_{m=1,2,\ ,k_i} \{d_{m+1} - d_m\}$, let

$$u_j(x_k) = \frac{d_{\min}}{|d_{j+1} - x_k| + |x_k - d_j|} \tag{9}$$

Fuzz all the sample data, the corresponding fuzzy time series can be obtained:

$$\begin{cases} A_1 = (u_1(x_1), u_2(x_1), \cdots, u_{k_i}(x_1)) \\ A_2 = (u_1(x_2), u_2(x_2), \cdots, u_{k_i}(x_2)) \\ \vdots \\ A_n = (u_1(x_n), u_2(x_n), \cdots, u_{k_i}(x_n)) \end{cases} \quad (10)$$

The following steps are same as the FTS method.

2.4. Modified GM (1,1) model

Modified GM (1,1) method is also used for comparison in the following section. Traditional GM (1,1) method can be seen in Ref. [45]. The modified GM (1,1) method is given below, which improved both initial value and weight μ of traditional GM (1,1) [46]. Next is the calculation process of modified GM (1,1). The time response formula of GM (1,1) is

$$\widehat{X}^{(1)}(k+1) = ce^{-ak} + \frac{\mu}{a}, k = 0,1,\cdots,n-1 \quad (11)$$

In order to get the constant c, an initial value should be supposed first. Suppose $\hat{X}^{(1)}(1) = X^{(0)}(1)$, then there is

$$\widehat{X}^{(1)}(1) = c + \frac{\mu}{a} = X^{(0)}(1) \quad (12)$$

$$c = X^{(0)}(1) - \frac{\mu}{a} \quad (13)$$

Then

$$\widehat{X}^{(1)}(k+1) = [X^{(0)}(1) - \frac{\mu}{a}]e^{-ak} + \frac{\mu}{a}, k = 0,1,\cdots,n-1 \quad (14)$$

After inverse accumulated generating operation for formula (10), the prediction formula of original sequence can be obtained

$$\widehat{X}^{(0)}(k+1) = \widehat{X}^{(1)}(k+1) - \widehat{X}^{(1)}(k) = c\cdot(1-e^{a})\cdot e^{-ak}, k = 1,2,\cdots,n-1 \quad (15)$$

Let $C = c\ (1 - e^{a})$, according to formula (10) and (13), there is

$$\widehat{X}^{(1)}(k+1) = C(1-e^{a})^{-1}e^{-ak} + \frac{\mu}{a}, k = 0,1,\cdots,n-1 \quad (16)$$

$$\widehat{X}^{(0)}(k+1) = \widehat{X}^{(1)}(k+1) - \widehat{X}^{(1)}(k) = C\cdot e^{-ak}, k = 1,2,\cdots,n-1 \quad (17)$$

The improved initial value is

$$C = \frac{[X^{(0)}(1) - \frac{\mu}{a}](1-e^{a})^{-1} + \sum_{k=2}^{n} X^{(0)}(k)e^{-a(k-1)}}{(1-e^{a})^{-2} + \sum_{k=2}^{n} e^{-2a(k-1)}} \quad (18)$$

Then by formula (16), the predicted values of modified model are got when $\mu = 0$. Next calculate sum of deviation square under $\mu = 0$ according to $s = \sum_{k=1}^{n} [\widehat{X}^{(0)}(k) - X^{(0)}(k)]^2$. On the basis of this, add a tiny amount $\Delta\mu > 0$, that is $\mu \Leftarrow \mu + \Delta\mu$, repeat the above until $\mu = 1$. During this process sum of deviation squares between predicted values and actual values under different weights can be compared, and select the weight as the optimal weight when sum of deviation square is the smallest. The prediction model established.

2.5. Deviation indexes of forecasting accuracy

Mean Square Error (MSE), Mean Absolute Error (MAE), Mean Absolute Percent Error (MAPE) are calculated to compare accuracy of methods. The calculation formulas are as follows.

$$MSE = \frac{\sum_{t=1}^{n} (X_{actual}(t) - X_{forecasted}(t))^2}{n} \quad (19)$$

$$MAE = \frac{\sum_{t=1}^{n} |(X_{actual}(t) - X_{forecasted}(t))|}{n} \quad (20)$$

$$MAPE = \frac{\sum_{t=1}^{n} |(X_{actual}(t) - X_{forecasted}(t)) / X_{actual}(t) * 100\%|}{n} \quad (21)$$

All these three deviation indexes reflect system error, and indicate the discrete degree between predicted data and the actual sequence. The accuracy of the forecasting method is the higher when these indexes are the smaller. MSE and MAE are both absolute indexes. They are affected by the dimension. MAPE is a relative index, which is not affected by the dimension.

3. Results

All calculations in this study were realized by matlab. First, the proposed prediction model was constructed based on the monitoring concentrations of COD_{Mn} and DO from the 1st week in 2015 to the 52nd week in 2016, respectively. Three weeks' data is missing among them; therefore, the model was constructed based on 101 weeks' data. Select the number of weeks' data for fuzzy relationship establishment according to the actual situation. Take $w = 5$ for example here. The data was clustered according to the preceding steps of the IFTS method. Calculate the RMSEs under different clustering numbers of COD_{Mn} and DO. Find the cluster numbers when the RMSEs are the minimum for COD_{Mn} and

DO respectively. Then it can be got that the optimal number of clusters for COD_{Mn} and DO are 11 and 9. Therefore, this two group results are taken as the final forecasting results for COD_{Mn} and DO, respectively. Then the model was used to fit the data of the last 96 weeks and forecast data from the 1^{st} to the 10^{th} week in 2017.

The fitting results of COD_{Mn} and DO by IFTS and corresponding monitoring data are shown in Fig. 2. In order to prove the effectiveness of the proposed method, FTS method and modified GM (1,1) method, a commonly used prediction method of water quality, are also adopted for the concentrations fitting of COD_{Mn} and DO. Fitting results of the last 96 weeks by all methods can be obtained. To compare fitting accuracy of these three methods more intuitively, the previous three deviation indexes (MSE, MAE and MAPE) are calculated. All these deviation indexes were shown in Tables 1 and 2 for COD_{Mn} and DO, respectively.

Next this method was used to forecast the concentrations of COD_{Mn} and DO from the 1^{st} to the 10^{th} week in 2017. FTS and modified GM (1,1) method are also adopted for the forecasting. Results from all methods and monitoring data are shown in Tables 3 and 4. The previous three deviation indexes were calculated and shown in Tables 5 and 6 for COD_{Mn} and DO, respectively.

4. Discussion

From Fig. 2, Tables 1 and 2, the data fitting of IFTS model is good. According to Tables 1 and 2, deviation indexes values of IFTS model is the smallest in the three methods both for COD_{Mn} and DO. This suggests that IFTS model has a higher fitting accuracy than both FTS and modified GM (1,1) model.

The fitting results of DO is worse than COD_{Mn} in all these three methods. That's because the concentrations of COD_{Mn} are relatively stable, while the concentrations of DO have a larger fluctuation. From previous studies, the concentrations of COD_{Mn} are stable in a whole year; the concentrations of DO are higher in spring and winter, but lower in summer and autumn. These also can be seen from the weekly data in Fig. 1.

The fitting errors of DO are higher. Given the fluctuation of data, smoothing of data is considered before the prediction of DO. Here three-point smoothing method [45] is used, and then make fitting as before. The new deviation

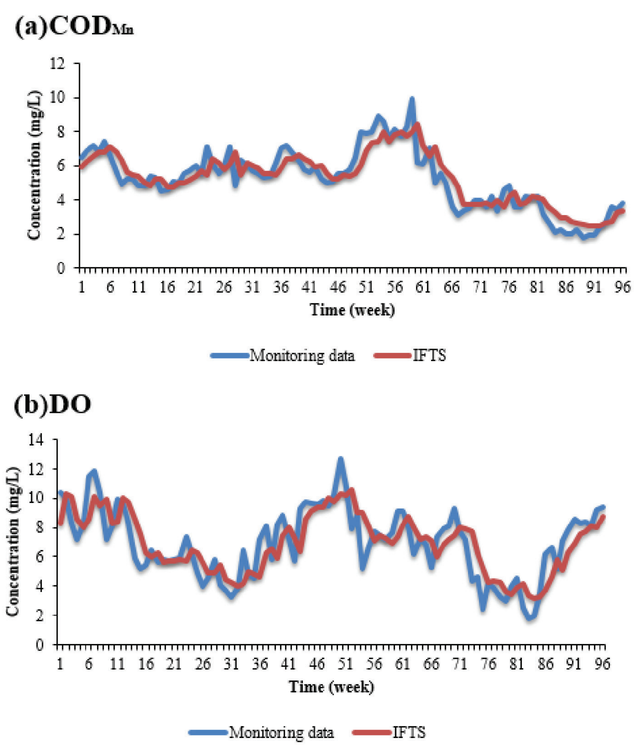

Fig. 2. Fitting results by IFTS and monitoring data of the last 96 weeks.

Table 1
Deviation indexes values of different methods' fitting results for COD_{Mn}

Method	MSE	MAE	MAPE
IFTS	0.65	0.64	13.42%
FTS	0.70	0.64	13.79%
Modified GM (1,1)	2.27	1.14	26.46%

Table 2
Deviation indexes values of different methods' fitting results for DO

Method	MSE	MAE	MAPE
IFTS	2.10	1.15	18.56%
FTS	2.33	1.17	18.82%
Modified GM (1,1)	5.28	1.98	37.29%

Table 3
Forecasting results from all methods and monitoring data for COD_{Mn}

Time	Monitoring data	IFTS	FTS	Modified GM (1,1)
1	3.50	3.60	3.36	3.57
2	3.50	3.50	3.36	3.76
3	4.00	3.61	3.50	3.95
4	4.30	3.84	3.50	4.16
5	4.60	4.05	4.00	4.38
6	4.90	4.28	4.36	4.61
7	5.20	4.54	4.36	4.85
8	4.70	4.86	5.00	5.10
9	6.30	4.71	4.64	5.37
10	5.60	5.42	6.00	5.65

Table 4
Forecasting results from all methods and monitoring data for DO

Time	Monitoring data	IFTS	FTS	Modified GM (1,1)
1	7.89	8.84	9.00	8.21
2	7.88	8.12	9.17	8.17
3	6.91	8.07	8.00	8.13
4	7.49	7.84	7.64	8.09
5	7.85	7.81	7.00	8.06
6	7.77	7.80	7.36	8.02
7	7.58	7.78	7.36	7.98
8	7.96	7.71	7.36	7.95
9	7.36	7.84	7.36	7.91
10	9.55	7.67	7.50	7.87

Table 5
Deviation indexes values of different methods' forecasting results for COD_{Mn}

Method	MSE	MAE	MAPE
IFTS	0.37	0.47	9.27%
FTS	0.53	0.59	11.94%
Modified GM (1,1)	0.14	0.28	5.55%

Table 6
Deviation indexes values of different methods' forecasting results for DO

Method	MSE	MAE	MAPE
IFTS	0.43	0.55	6.92%
FTS	0.96	0.78	9.64%
Modified GM (1,1)	0.54	0.55	6.98%

indexes values of IFTS model are calculated and they are 1.13, 0.93, and 14.66%. By contrast, the prediction accuracy of IFTS model after data smoothing has increased significantly. It is visible that data fluctuation has considerable impact on model's accuracy.

It can be seen from the index values in Table 5 that the precision of IFTS is higher than that of FTS method, but lower than that of modified GM (1,1). This verified the advantage of IFTS method over FTS method. Further through the data analysis in Table 3, we can see the significant monotonous trend of this small data series. Therefore, to forecast data with obviously monotonous trend, gray prediction method is the first choice. Based on the analysis of the data in Table 4 and the index values in Table 6, IFTS has higher accuracy than both the FTS and modified GM (1,1) methods for sequence predictions with no noticeable trend.

5. Conclusions

FTS model has both advantages of fuzzy and time series, and can overcome the problem that historical data is inaccurate or incomplete. The calculation process of FTS was improved, and the IFTS method was proposed. As the distribution of data is usually not uniform in practice, it is more reasonable to divide domain into several unequal intervals than equal ones. FCM is an excellent method in unequal classification. As how to determine the optimal cluster number for the FCM algorithm was still an unsolved problem, the optimal clustering number was to be determined based RMSE, which avoided the defects of subjective definition. Besides, the fuzzy sets and data fuzzification method were defined based on the distance. This made the fuzzy sets change with the change of data, and enhanced the sensitivity and interpretability of the data. The two improvements both improved the model precision.

This study introduced IFTS method to water quality forecasting. IFTS method was applied in concentration prediction of COD_{Mn} and DO based on data from Sanchakou Station of Haihe River. Modified GM (1,1) methods were also used for prediction. By a series of contrast and analysis, it is concluded that IFTS method has higher accuracy than FTS method and possesses obvious advantages than modified GM (1,1) method in forecasting of data with no fixed trend.

FTS model provides a new thought for water quality forecasting. But prediction accuracy and calculated amount of this method remains to be improved.

Acknowledgment

This work was supported by Shandong Technology and Business University Ph.D. Research Foundation (No. 5002-014-306332).

References

[1] R.V. Thomann, The future 'Golden Age' of predictive models for surface water quality and ecosystem management, J. Environ. Eng., 124 (1998) 94–103.
[2] C.H. Tang, Y.J. Yi, Z.F. Yang, J. Sun, Risk forecasting of pollution accidents based on an integrated Bayesian Network and water quality model for the South to North Water Transfer Project, Ecol. Eng., 96 (2015) 109–116.
[3] X. Wang, J.J. Zhang, V. Babovic, Improving real-time forecasting of water quality indicators with combination of process-based models and data assimilation technique, Ecol. Indic., 66 (2016) 428–439.
[4] W.H. Deng, G.Y. Wang, X.R. Zhang, A novel hybrid water quality time series prediction method based on cloud model and fuzzy forecasting, Chemometr. Intell. Lab. Syst., 149 (2015) 39–49.
[5] H.M. Nagy, M. Watanabe, M. Hirano, Prediction of sediment load concentration in rivers using artificial neural network model, J. Hydraul. Eng., 128 (2002) 588–595.
[6] S.E. Kim, I.W. Seo, Artificial neural network ensemble modeling with conjunctive data clustering for water quality prediction in rivers, J. Hydro-environ. Res., 9 (2015) 325–339.

[7] A. Yan, Z.H. Zou, Y.F. Zhao, Forecasting of dissolved oxygen in the Guanting reservoir using an optimized NGBM (1, 1) model, J. Environ. Sci., 29 (2015) 158–164.
[8] Q. Song, B.S. Chissom, Fuzzy time series and its models, Fuzzy Set. Syst., 54 (1993) 269–277.
[9] Q. Song, B.S. Chissom, Forecasting enrollments with fuzzy time series: Part I, Fuzzy Set. Syst., 54 (1993) 1–10.
[10] Q. Song, B.S. Chissom, Forecasting enrollments with fuzzy time series: Part II, Fuzzy Set. Syst., 62 (1994) 1–8.
[11] J. Sullivan, W.H. Woodall, A comparison of fuzzy forecasting and Markov modeling, Fuzzy Set. Syst., 64 (1994) 279–293.
[12] S.M. Chen, Forecasting enrollments based on high-order fuzzy time series, Cybernet. Syst., 33 (2002) 1–16.
[13] K. Huarng, Heuristic models of fuzzy time series for forecasting, Fuzzy Set. Syst., 123 (2001) 369–386.
[14] L.W. Lee, H. Wang, S.M. Chen, Y.H. Leu, Handling forecasting problems based on two-factor high-order fuzzy time series, IEEE T. Fuzzy Syst., 14 (2006) 468–477.
[15] L.W. Lee, L.H. Wang, S.M. Chen, Temperature prediction and TAIFEX forecasting based on fuzzy logical relationships and genetic algorithm, Expert Syst. Appl., 33 (2007) 539–550.
[16] S. Askari, N. Montazerin, M.H.F. Zarandi, A clustering based forecasting algorithm for multivariable fuzzy time series using linear combinations of independent variables, Appl. Soft Comput., 35 (2015) 151–160.
[17] F.M. Talarposhti, H.J. Sadaei, R. Enayatifar, M. Mahmud, T. Eslami, Stock market forecasting by using a hybrid model of exponential fuzzy time series, Int. J. Approx. Reason., 70 (2015) 79–98.
[18] R.C. Tsaur, T.C. Kuo, The adaptive fuzzy time series model with an application to Taiwan's tourism demand, Expert Syst. Appl., 38 (2011) 9164–9171.
[19] N.Y. Wang, S.M. Chen, Temperature prediction and TAIFEX forecasting based on automatic clustering techniques and two-factor high-order fuzzy time series, Expert Syst. Appl., 36 (2009) 2143–2154.
[20] C.H. Cheng, Y.S. Chen, Y.L. Wu, Forecasting innovation diffusion of products using trend-weighted fuzzy time-series model, Expert Syst. Appl., 36 (2009) 1826–1832.
[21] C.H. Cheng, T.L. Chen, L.Y. Wei, A hybrid model based on rough sets theory and genetic algorithms for stock price forecasting, Inform. Sciences 180 (2010) 1610–1629.
[22] M.H.F. Zarandi, A. Molladavoudi, A. Hemmati, Fuzzy time series based on defining interval length with Imperialist Competitive Algorithm, Fuzzy Information Processing Society (NAFIPS), 2010 Annual Meeting of the North American, (2010) 1–7.
[23] J.W. Liu, T.L. Chen, C.H. Cheng, Y.H. Chen, Adaptive-expectation based multi-attribute FTS model for forecasting TAIEX, Comput. Math. Appl., 59 (2010) 795–802.
[24] Y. Leu, C.P. Lee, Y.Z. Jou, A distance-based fuzzy time series model for exchange rates forecasting, Expert Syst. Appl., 36 (2009) 8107–8114.
[25] Y.Y. Hsu, S.M. Tse, B. Wu, A new approach of bivariate fuzzy time series analysis to the forecasting of a stock index. International Journal of Uncertainty, Fuzziness Knowledge-Based Syst., 11 (2011) 671–690.
[26] H.L. Wong, Y.H. Tu, C.C. Wang, Application of fuzzy time series models for forecasting the amount of Taiwan export, Expert Syst. Appl., 37 (2010) 1465–1470.
[27] M. Versaci, F.C. Morabito, Fuzzy time series approach for disruption prediction in Tokamak reactors, IEEE T. Magn, 39 (2003) 1503–1506.
[28] L.A. Zadeh, Fuzzy Sets, Inform. Control, 8 (1965) 338–353.
[29] N.B. Chang, H.W. Chen, S.K. Ning, Identification of river water quality using the fuzzy synthetic evaluation approach, J. Environ. Manage., 63 (2001) 293–305.
[30] H. Gharibi, A.H. Mahvi, R. Nabizadeh, H. Arabalibeik, M. Yunesian, M.H. Sowlat, A novel approach in water quality assessment based on fuzzy logic, J. Environ. Manage., 112 (2012) 87–95.
[31] J.S.R. Jang, C.T. Sun, Neuro-fuzzy modeling and control, P. IEEE 83 (1995) 378–406.
[32] S. Karmakar, P.P. Mujumdar, Grey fuzzy optimization model for water quality management of a river system, Adv. Water Resour., 29 (2006) 1088–1105.
[33] C.L. Karr, E.J. Gentry, Fuzzy control of pH using genetic algorithms, IEEE T. Fuzzy Syst., 1 (1993) 46–53.
[34] A. Lermontov, L. Yokoyama, M. Lermontov, M.A.S. Machado, River quality analysis using fuzzy water quality index: Ribeira do Iguape river watershed, Brazil. Ecol. Indic., 9 (2009) 1188–1197.
[35] X. Ke, C.Y. Wang, D.B. Jing, Y. Zhang, H.J. Zhang, Assessing water quality by ratio of the number of dominant bacterium species between surface/subsurface sediments in Haihe River Basin, Mar. Pollut. Bull., 98 (2015) 267–273.
[36] Y. Qin, D.W. Yang, H.M. Lei, K. Xu, X.Y. Xu, Comparative analysis of drought based on precipitation and soil moisture indices in Haihe basin of North China during the period of 1960–2010, J. Hydrol., 526 (2015) 55–67.
[37] Y. Icaga, Fuzzy evaluation of water quality classification, Ecol. Indic., 7 (2007) 710–718.
[38] W. Ocampo-Duque, N. Ferre-Huguet, J.L. Domingo, M. Schuhmacher, Assessing water quality in rivers with fuzzy inference systems: a case study, Environ. Int., 32 (2006) 733–742.
[39] Z.H. Zou, Y. Yun, J.N. Sun, Entropy method for determination of weight of evaluating indicators in fuzzy synthetic evaluation for water quality assessment, J. Environ. Sci., 18 (2006) 1020–1023.
[40] S.H. Cheng, S.M. Chen, W.S. Jian, Fuzzy time series forecasting based on fuzzy logical relationships and similarity measures, Inform. Sciences, 327 (2016) 272–287.
[41] H. Verma, R.K. Agrawal, A. Sharan, An improved intuitionistic fuzzy c-means clustering algorithm incorporating local information for brain image segmentation, Appl. Soft Comput., 46 (2016) 543–557.
[42] L.Q. Li, W.X. Xie, Z.X. Liu, A novel quadrature particle filtering based on fuzzy c-means clustering, Knowledge-Based Syst., 106 (2016) 105–115.
[43] G. Chen, H.W. Qu, A new forecasting method of fuzzy time series model, Control Decis., 28 (2013) 105–114.
[44] G.H. Wang, J. Yao, Forecasting model of fuzzy time series based on K-means algorithm. Acta Anal. Funct. Appl., 17 (2015) 58–65.
[45] X.M. Wang, The grey system analysis and practical application, Huazhong University of Science and Technology Press, Wuhan, China, (2001).
[46] X. Yang, The improved grey prediction GM (1,1) model in MATLAB, Jiangsu Sci. Tech. Inform., 4 (2014) 69–70.