Using a fractional order grey seasonal model to predict the dissolved oxygen and pH in the Huaihe River

Kai Zhang and Lifeng Wu

ABSTRACT

To accurately forecast the seasonal fluctuations of dissolved oxygen (DO) and pH in Huaihe River, a grey seasonal model with fractional order accumulation is proposed, optimized by particle swarm optimization (PSO-FGSM(1,1)). We use this new model to carry out an empirical analysis based on the DO and pH data from 2014 to 2018 from Huaibin, Bengbu, Chuzhou monitoring points. The comparison results show that the PSO-FGSM(1,1) model accuracy is significantly higher than the Holt-Winters model with grey wolf optimization (GWO-Holt-Winters). The prediction results indicated that the pollution of the Huaihe River has regional characteristics. The Huaibin and Chuzhou sections of the Huaihe River are slightly polluted, and the Bengbu section is seriously polluted.

Key words | dissolved oxygen, GWO-Holt-Winters model, pH, prediction, PSO-FGSM(1,1)

Kai Zhang (corresponding author)
Lifeng Wu
College of Management Engineering and Business,
Hebei University of Engineering,
Handan 056038,
China
E-mail: *276423643@qq.com*

HIGHLIGHTS

- Fractional order grey model was combined with seasonal factors.
- The prediction effect of the new combination model is better than that of the traditional seasonal model.
- The proposed model was first used to predict DO and pH.
- Data from three monitoring sites on the Huaihe River were used to model the analysis.
- It provides a new method to study the seasonal fluctuation sequence of a small number of samples.

INTRODUCTION

River pollution has always been a hotspot of public concern, and therefore the subject of extensive and in-depth research (Schwarzenbach *et al.* 2010). Water pollution control needs to be systematic, requiring prevention, effective identification of pollution and control of pollutants. In addition to traditional industrial pollution, life and agricultural pollution have also become the main causes of pollution (Moss 2007). It is aquatic life that is directly affected by river pollution.

The prediction of water pollution is very important for river pollution control. Models for predicting water pollution are used in many places. For example, an artificial neural network was used to predict water pollution sources in Turkey and the Johor River in Malaysia (Najah *et al.* 2009). To ensure the water quality, it was predicted in the main canal of the South-to-North Water Transfer Project (Tang *et al.* 2014).

Dissolved oxygen (DO) is an important indicator for evaluating water quality and aquatic environment. There are two major methods in DO prediction. One is the traditional statistical model. For example, a new linear regression in fuzzy mathematics proved to have higher Nash-Sutcliffe efficiency in DO prediction (Khan & Valeo 2015). An optimized nonlinear grey Bernoulli model using a particle swarm optimization algorithm was proposed for forecasting dissolved oxygen in the Guanting reservoir (An *et al.* 2015). Nonlinear mathematical modeling approaches, namely the modified response surface method, showed the best performance in comparison with the multilayer perceptron neural network (Keshtegar & Heddam 2017). Nonlinear models gave better results than linear models in DO prediction (Basant *et al.* 2010). The other is intelligent computing

doi: 10.2166/wst.2020.596

models. For example, a machine learning model was used to assess the predictability of column minimum DO (Ross & Stock 2019). The multivariate adaptive regression splines have similar accuracy to least square support vector machine in water pollution prediction (Kisi & Parmar 2016). Attention-based recurrent neural networks can achieve more accurate DO prediction in both short-term and long-term prediction (Liu *et al.* 2019). Nature-inspired algorithms and optimally pruned extreme learning machines are only suitable for short-term forecasting (Heddam 2016). Moreover, support vector machines and artificial neural networks are also widely used in DO prediction (Liu *et al.* 2013; Ji *et al.* 2017), including improved versions. Hybrid model combine traditional and intelligent models make the prediction more accurate (Huan *et al.* 2018; Li *et al.* 2018a). However, these models need a large number of samples to ensure accuracy, but data loss due to monitoring equipment failure is inevitable.

The grey model has high accuracy on a small number of samples (Wu *et al.* 2013). Since the grey system theory was proposed (Deng 1989), the grey prediction model has been studied extensively. It was first applied to the prediction of water supply in 1993 (Wang 1993). Since then, grey prediction models have been widely used and the prediction accuracy has been improved. The grey model combined with fuzzy mathematics was applied to the Tunga–Bhadra river system in India (Karmakar & Mujumdar 2006). An artificial neural network whose parameters are represented by grey numbers has good accuracy for river stage forecasting (Alvisi & Franchini 2012). A new grey water-forecasting model is used to forecast water demand (Wu *et al.* 2017). In order to improve the accuracy of the grey prediction model, a fractional order grey prediction model has been proposed (Wu *et al.* 2015), and the optimal fractional order grey model has been applied to predict the amount of waste-sewage water discharged into the Yangtze River basin (Li *et al.* 2018b). To enhance the performance of the grey model in processing seasonal data, a lot of research has been carried out. The grey prediction model with the ratio-to-moving-average deseasonalization method has been proposed to predict seasonal time series (Tseng *et al.* 2001). An improved seasonal rolling grey forecasting model and grey model with improved seasonal indexes have been applied to seasonal data (Cheng *et al.* 2017; Xiao *et al.* 2017; Wang *et al.* 2018). These show that grey prediction models with seasonal indexes have better forecast results.

According to the United Nations' World Water Development Report 2018, about 3.6 billion people, nearly half the world's population, lack water, and that number could rise to 5.7 billion by the middle of the century. The United Nations is calling on countries to develop environmental policies to improve water quality. Water shortage is already serious in China. More than a third of the land is very short of water. To alleviate water pressure, the Chinese government has issued the opinions of the state council on implementing the strictest water resources management system. Controlling water pollution is an important measure to realize water resources management. In the past decades of extensive development in China, water resources have been seriously polluted. Rivers have gradually lost their capacity to absorb pollution and purify themselves. The continuous accumulation of pollutants in water has become a major environmental safety hazard. Water pollution has expanded from being river pollution to being river basin pollution.

DO is one of the main criteria for detecting water pollution. The DO content directly affects the growing environment of fish in rivers. Rivers with good water quality often have an environment suitable for fish. The microbial population is kept within reasonable limits. Such rivers often have a good capacity for self-purification, and can cope with a certain amount of pollution. However, to facilitate the discharge of sewage, a large number of enterprises and factories have been built on the river banks. Raw sewage is discharged directly into rivers. In water pollution, organic pollution accounts for a large proportion, with the main pollutants being ammonia nitrogen, chemical oxygen demand, etc. A lot of organic matter is broken down by a lot of microbes. A large amount of DO is consumed in the decomposition process. Fish die from lack of oxygen. The decline of fisheries has caused people to pay attention to sewage treatment. Using the Huaihe River basin as an example, in the national water resources protection plan (2016–2030), the target for limiting the discharge of major pollutants into Huaihe River basin in 2030 is 266,000 tons per year. The prediction of DO concentration can provide a reference for pollution control and enhance the confidence of aquaculture industry.

To treat river pollution, a large number of sewage treatment plants were built and put into use. The pH value of sewage exceeds the normal range and can cause damage to the sewage treatment equipment. At the same time, the pH value is the most convenient way to check the water quality. It is widely used by fishers to detect the living conditions of fish, as it does not need expensive detection equipment. The pH value of 6.5 to 8.3 is the normal range for fish growth. Outside these pH values, the growth of fish will be inhibited and the growth cycle will be prolonged, thus missing the best time for fish to come to market. In order to protect the fishery in the Huaihe River basin, the fishery ecological protection research center in the Huaihe River basin was established. To support national policies and optimize the future aquaculture environment, the DO concentration and pH value are predicted in this paper.

Because the Huaihe River sewage detection equipment was put in late, not much data has been obtained. The grey prediction model has high precision in prediction involving a small number of samples. Because of the seasonal variation of rivers, an improved grey seasonal model with fractional order accumulation was proposed. The optimal order of the proposed model was obtained using the particle swarm optimization algorithm (PSO-FGSM(1,1)). The DO concentration and pH value of the Huaihe River were predicted by the PSO-FGSM(1,1) model, and compared with the prediction results of the traditional seasonal prediction Holt-Winters model.

METHODS

The Holt-Winters model

The Holt-Winters model is a traditional statistical model, and has been widely used since it was proposed by Winters (1960). It is used to process data with seasonal characteristics. According to the characteristics of the data sequence, seasonal, trend and level items are extracted separately. After being calculated by exponential smoothing, the three parts extrapolate the predicted value. A prediction model is established based on the Holt-Winters model. There are two kinds of calculation methods for the Holt-Winters model, namely addition and multiplication. The multiplication Holt-Winters model works better for solving seasonal time series prediction problems. Therefore, the multiplication Holt-Winters model is used for comparison in this paper. The establishment process of the model is as follows.

For the time series $X^{(0)} = \{x^{(0)}(1), x^{(0)}(2), \ldots, x^{(0)}(n),\}$, the multiplication Holt-Winters model formulation is as follows:

$$S_t = \alpha \frac{x^{(0)}(t)}{I_{t-L}} + (1-\alpha)(S_{t-1} + b_{t-1}), \quad 0 < \alpha < 1 \tag{1}$$

$$b_t = \gamma(S_t - S_{t-1}) + (1-\gamma)b_{t-1}, \quad 0 < \gamma < 1 \tag{2}$$

$$I_t = \beta \frac{x^{(0)}(t)}{S_t} + (1-\beta)I_{t-L}, \quad 0 < \beta < 1 \tag{3}$$

where L is the seasonal length, such as four quarters, 12 months etc; I is a seasonal index; b represents the tendency. The original calculation formulation is as follows:

$$I_t = \frac{\overline{x^{(0)}(L)}}{\overline{x^{(0)}(t)}}, \quad S_{L+1} = x^{(0)}(L+1); \tag{4}$$

$$b_{L+1} = \frac{x^{(0)}(L+1) - x^{(0)}(1) + x^{(0)}(L+2) - x^{(0)}(2) + x^{(0)}(L+3) - x^{(0)}(3)}{3L}; \tag{5}$$

where $\overline{x^{(0)}(L)}$ represents the average of the same month in different years, and $\overline{x^{(0)}(t)}$ is the general average.

The grey seasonal model with fractional order accumulation

The data varied substantially among different seasons but the traditional grey prediction model is only suitable for time series that exhibit an exponential trend, so it is not capable of effectively predicting data with large fluctuations. Thus, in order to eliminate the fluctuations in the data and enhance the forecasting performance, the seasonal data is summarized into annual data in this study. The annual data is predicted by the grey model (1,1) (GM(1,1)) with fractional order accumulation (Wu *et al.* 2013). The monthly data is then restored using the seasonal indexes.

The original monthly data is $X^{(0)} = \{x^{(0)}(1), x^{(0)}(2), \cdots, x^{(0)}(n)\}$ with a cycle of L, and the modeling process for the FGSM(1,1) model is as follows.

Step 1. The yearly data is:

$$Y^{(0)} = \{y^{(0)}(1), y^{(0)}(2), \cdots y^{(0)}(m)\}, m = \left\lceil \frac{n}{L} \right\rceil$$

where $y^{(0)}(1) = \sum_{i=1}^{L} x^{(0)}(i), y^{(0)}(2) = \sum_{i=L+1}^{2L} x^{(0)}(i), \cdots, y^{(0)}(m) = \sum_{i=(m-1)L+1}^{mL} x^{(0)}(i)$.

Step 2. By using $y^{(r)}(k) = \sum_{j=1}^{k} C_{k-j+r-1}^{k-j} y^{(0)}(j)$, the r-order accumulation sequence is:

$$Y^{(r)} = \{y^{(r)}(1), y^{(r)}(2), \cdots, y^{(r)}(m)\} \tag{6}$$

where $C_{r-1}^{0} = 1, C_{k}^{k+1} = 0, C_{k-j+r-1}^{k-j} = \frac{(k-j+r-1)(k-j+r-2)\cdots(r+1)r}{(k-j)!}$.

Step 3. For the r-order accumulation sequence $Y^{(r)}$, the first-order differential equation with one variable (i.e., the

FGM(1,1) model) can be defined as:

$$\frac{dy^{(r)}}{dt} + ay^{(r)} = b \tag{7}$$

where a is the development coefficient and b is the grey action quantity. The solution to Equation (7) is:

$$y^{(r)}(t+1) = \left(y^{(0)}(1) - \frac{b}{a}\right)e^{-at} + \frac{b}{a} \tag{8}$$

The least squares estimate minimizes the sum of the squared residuals, so the parameters are obtained by using the least squares method. The unknown parameters $\hat{a}, \hat{b}$ can be solved by using the following formulae:

$$\begin{bmatrix} \hat{a} \\ \hat{b} \end{bmatrix} = (B^{\mathrm{T}}B)^{-1}B^{\mathrm{T}}Y \tag{9}$$

where

$$B = \begin{bmatrix} -0.5(y^{(r)}(1) + y^{(r)}(2)) & 1 \\ -0.5(y^{(r)}(2) + y^{(r)}(3)) & 1 \\ \vdots & \vdots \\ -0.5(y^{(r)}(m-1) + y^{(r)}(m)) & 1 \end{bmatrix}, \quad Y = \begin{bmatrix} y^{(r)}(2) \\ y^{(r)}(3) \\ \vdots \\ y^{(r)}(m) \end{bmatrix}$$

Step 4. Input $\hat{a}$ and $\hat{b}$ into the time response function:

$$\hat{y}^{(r)}(k+1) = (y^{(0)}(1) - \frac{\hat{b}}{\hat{a}})e^{-\hat{a}k} + \frac{\hat{b}}{\hat{a}} \tag{10}$$

and $\hat{y}^{(r)}(k+1)$ is the fitting value at time $k+1$.

Step 5. For $\hat{Y}^{(r)} = \{\hat{y}^{(r)}(1), \hat{y}^{(r)}(2), \cdots, \hat{y}^{(r)}(m), \cdots\}$, the predictive sequence is:

$$\begin{aligned} \alpha^{(r)}\hat{Y}^{(r)} = \{&\alpha^{(1)}\hat{y}^{(r)(1-r)}(1), \alpha^{(1)}\hat{y}^{(r)(1-r)}(2), \cdots, \\ &\alpha^{(1)}\hat{y}^{(r)(1-r)}(m), \alpha^{(1)}\hat{y}^{(r)(1-r)}(m+1), \cdots\} \end{aligned} \tag{11}$$

where $\alpha^{(1)}\hat{y}^{(r)(1-r)}(k) = \hat{y}^{(r)(1-r)}(k) - \hat{y}^{(r)(1-r)}(k-1)$. Then, the forecasting value is:

$$\hat{y}^{(0)}(1), \hat{y}^{(0)}(2), \cdots, \hat{y}^{(0)}(m), \hat{y}^{(0)}(m+1), \hat{y}^{(0)}(m+2), \cdots.$$

Step 6. The seasonal index $S_i (i = 1, 2, \cdots, L)$ is calculated by using the corresponding period average method:

$$S_i = \frac{\bar{x}_i^{(0)}(i)}{\bar{x}^{(0)}} \tag{12}$$

where $\bar{x}_i^{(0)}(i)$ indicates the average of month i (at time point i) and $\bar{x}^{(0)}$ represents the average of all months.

Step 7. For:

$$\hat{X}^{(0)} = \{\hat{x}^{(0)}(1), \hat{x}^{(0)}(2), \cdots, \hat{x}^{(0)}(n), \hat{x}^{(0)}(n+1), \cdots\} \tag{13}$$

where $\hat{x}^{(0)}(1) = \hat{y}^{(0)}(1)\frac{S_1}{L}$, $\hat{x}^{(0)}(2) = \hat{y}^{(0)}(1)\frac{S_2}{L}$, $\cdots$, $\hat{x}^{(0)}(L) = \hat{y}^{(0)}(1)\frac{S_L}{L}$, $\hat{x}^{(0)}(L+1) = \hat{y}^{(0)}(2)\frac{S_{L+1}}{L}$, $\cdots$, $\hat{x}^{(0)}(n) = \hat{y}^{(0)}(m)\frac{S_n}{L}$, $\hat{x}^{(0)}(n+1) = \hat{y}^{(0)}(m+1)\frac{S_{n+1}}{L}$, $\cdots$.

Step 8. The mean absolute percentage error (MAPE) is used to test model accuracy, and the calculation process is as follows:

$$\text{MAPE} = \frac{1}{n}\sum_{k=1}^{n}\left|\frac{\hat{x}^{(0)}(k) - x^{(0)}(k)}{x^{(0)}(k)}\right| \times 100\% \tag{14}$$

The steps involved in the modeling process can be illustrated using a flow chart (Figure 1).

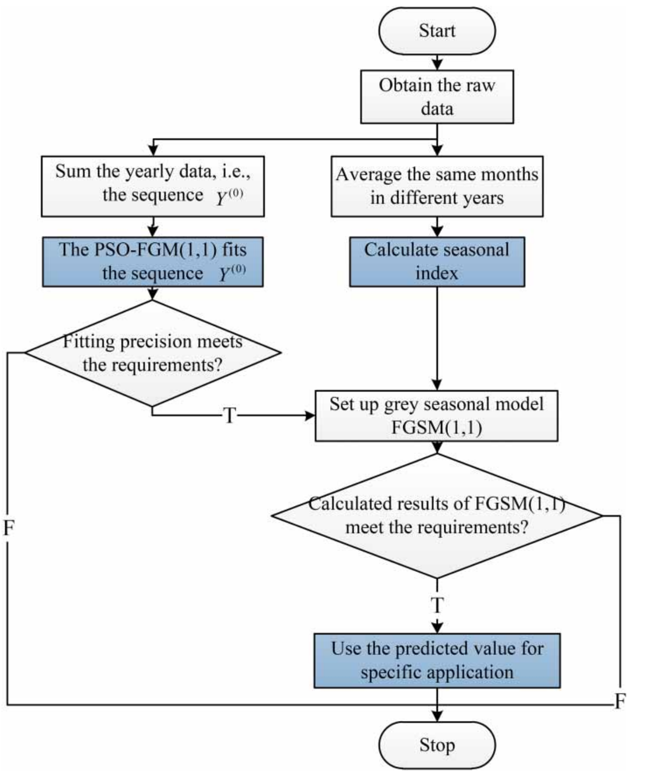

Figure 1 | The steps in the prediction process using the proposed FGSM(1,1) model.

DATA AND CASE ANALYSIS

Study area status

The importance of Huaihe River lies in its geography, population, agriculture, industry and transportation. In terms of geography, the Huaihe River is the dividing line between south and north China, between the Yangtze River and the Yellow River. The main branch of the Huaihe River starts in Henan and passes through Anhui and Jiangsu provinces. The Huaihe River basin has a total population of 165 million, the highest population density among all the basins in China. In terms of agriculture, the Huaihe River basin is an important grain-producing area in China, with grain output accounting for 17.3% of the country's total grain output. In terms of industry, the Huaihe River basin is dominated by coal, power, food and light industries. A number of large national coal production bases have been built. The annual coal production accounts for one eighth of the national coal production. In terms of transportation, the Beijing–Shanghai, Beijing–Kowloon and Beijing–Guangzhou railway arteries pass through the Huaihe River basin.

In this paper, three monitoring points of the Huaihe River are investigated, namely, Huaibin, Bengbu and Chuzhou. The Huaibin monitoring station is located at the junction of Henan and Anhui provinces. The Huaihe River crosses Bengbu city, and Bengbu sluice is a large water conservancy project in the middle reaches of the Huaihe River. The Bengbu monitoring point is located on Bengbu sluice. Chuzhou is a city in Anhui province, bordering Jiangsu province. The locations of the Huaibin, Bengbu, and Chuzhou monitoring points are shown in Figure 2.

In order to verify the effectiveness of the FGSM(1,1) model, we established models for DO and pH from 2014 to 2018. The data is from the official website of ministry of ecology and environment of the People's Republic of China (http://datacenter.mee.gov.cn/websjzx/queryIndex.vm). The current water quality data relates to a standard (GB 3838–2002) published in 2002. The data loss caused by instrument failure was calculated by the anterior-posterior average method. Weekly data were summed into monthly data to facilitate the calculation of the monthly seasonal indexes.

The calculation and prediction of DO

The calculation results of the PSO-FGSM(1,1) model

Monthly data from 2014 to 2018 are summed. The grey seasonal model with fractional order accumulation was used to

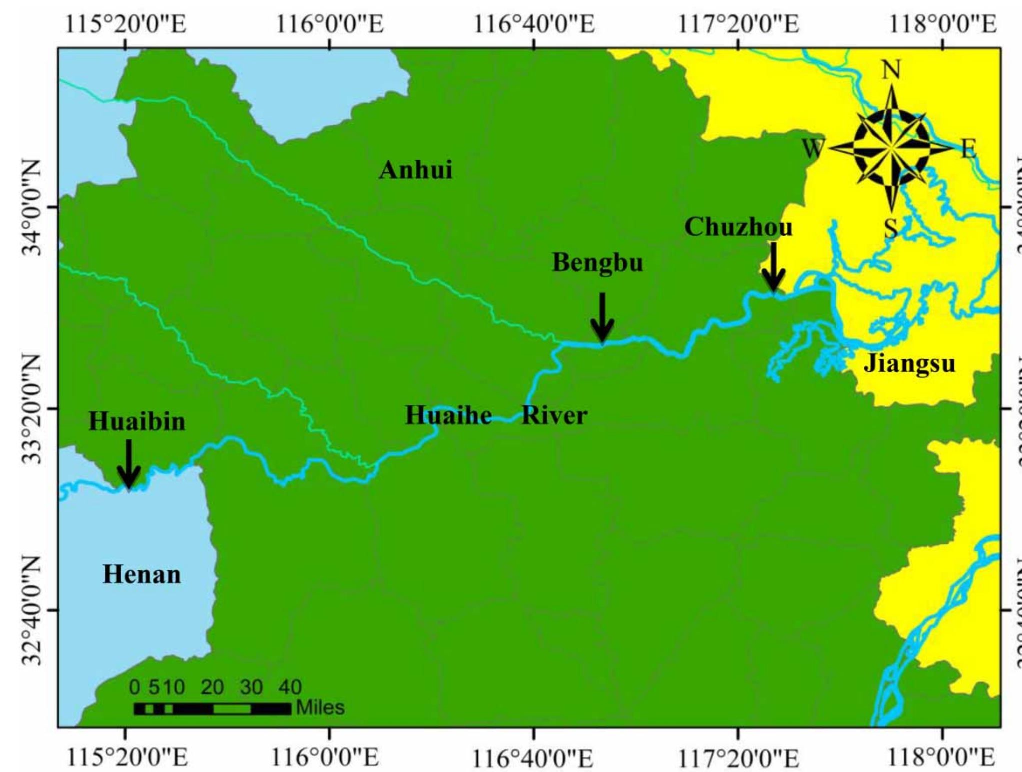

Figure 2 | The locations of the three monitoring points.

Table 1 | Fitting results of the three monitoring points using the PSO-FGSM(1,1) model

	Huaibin		Bengbu		Chuzhou	
Year	Actual value	Fitting value	Actual value	Fitting value	Actual value	Fitting value
2014	116.65	116.65	81.15	81.15	103.57	103.57
2015	112.68	109.78	87.31	87.31	103.76	106.81
2016	105.61	103.90	85.85	82.61	110.97	102.35
2017	90.64	98.86	70.23	76.51	90.14	98.07
2018	95.02	94.54	73.42	70.19	96.38	93.97
MAPE		3.44%		4.29%		5.50%

Table 2 | Seasonal indexes of the three monitoring points

Month	Huaibin	Bengbu	Chuzhou
1	1.30	1.41	1.39
2	1.38	1.33	1.32
3	1.23	1.17	1.21
4	1.03	0.92	1.00
5	0.83	0.86	0.88
6	0.81	0.74	0.83
7	0.83	0.65	0.68
8	0.80	0.71	0.71
9	0.86	0.81	0.85
10	0.84	0.89	0.86
11	0.94	1.12	1.04
12	1.15	1.39	1.23

fit the data, with the optimal order being calculated by particle swarm algorithm (PSO-FGSM(1,1)). The error (MAPE) value of the three monitoring points was 3.44%, 4.29% and 5.50%. The model has better fitting accuracy. The fitting results are shown in Table 1.

The seasonal indexes represent the seasonal trend of the data. Seasonal indexes are only meaningful when each data set has a similar variation trend. Through the data calculation of Huaibin, Bengbu and Chouzhou, the respective seasonal indexes were obtained and are shown in Table 2. It can be seen that the seasonal indexes of three monitoring points have the same variation trend, and the seasonal indexes are higher in winter than in summer. One reason is that dissolved oxygen is inversely proportional to temperature: the higher the temperature, the lower the concentration of dissolved oxygen. On the other hand, pollution is worse in summer than in winter.

Using Equation (13), the monthly fitting values of DO concentrations at the three monitoring points from 2014 to 2018 were calculated. The formula for calculating MAPE is shown in Equation (14). The results are listed in Tables 3 and 4.

The calculation results of the Holt-Winters model

As a comparison, the Holt-Winters model was used to fit data from the three monitoring points from 2014 to 2018. The grey wolf algorithm was used to optimize the three parameters in the model (GWO-Holt-Winters). Since the data of 2014 was the first cycle, it was not used in the fitting and its MAPE

Table 3 | Fitting results of the grey seasonal model at Huaibin monitoring point

	2014		2015		2016		2017		2018	
	Actual value	Fitting value	Actual value	Fitting value	Actual value	Fitting value	Actual value	Fitting value	Actual value	Fitting value
1	13.58	12.68	13.13	11.93	11.93	11.29	8.13	10.74	9.81	10.27
2	13.25	13.38	13.85	12.60	13.53	11.92	7.66	11.34	11.45	10.85
3	13.48	12.00	12.35	11.29	10.17	10.69	8.40	10.17	9.17	9.73
4	10.00	10.05	9.48	9.46	8.96	8.95	8.70	8.52	7.70	8.14
5	7.37	8.09	8.61	7.61	6.47	7.21	8.06	6.86	5.60	6.56
6	6.29	7.91	7.56	7.44	7.60	7.04	7.35	6.70	6.50	6.41
7	7.94	8.03	6.96	7.55	6.86	7.15	7.72	6.80	6.34	6.50
8	8.05	7.74	6.07	7.28	7.69	6.89	6.55	6.56	6.19	6.27
9	8.38	8.37	6.91	7.87	8.42	7.45	6.55	7.09	7.08	6.78
10	7.50	8.16	8.72	7.68	7.20	7.26	5.38	6.91	7.60	6.61
11	8.98	9.11	8.41	8.58	7.75	8.12	7.41	7.72	8.12	7.39
12	11.84	11.14	10.64	10.49	9.05	9.93	8.74	9.44	9.47	9.03

Table 4 | Fitting results of the grey seasonal model at Bengbu and Chuzhou monitoring points

Month	2014 fitting value	2015 fitting value	2016 fitting value	2017 fitting value	2018 fitting value
Bengbu					
1	9.51	10.23	9.68	8.96	8.22
2	9.02	9.71	9.19	8.51	7.80
3	7.91	8.51	8.05	7.45	6.84
4	6.25	6.72	6.36	5.89	5.40
5	5.80	6.24	5.90	5.46	5.01
6	5.03	5.41	5.12	4.74	4.35
7	4.36	4.69	4.44	4.11	3.77
8	4.80	5.17	4.89	4.53	4.16
9	5.47	5.89	5.57	5.16	4.73
10	6.04	6.49	6.14	5.69	5.22
11	7.56	8.13	7.69	7.13	6.54
12	9.41	10.12	9.58	8.87	8.14
Chuzhou					
1	12.02	12.39	11.87	11.38	10.90
2	11.36	11.72	11.23	10.76	10.31
3	10.47	10.80	10.35	9.91	9.50
4	8.67	8.94	8.57	8.21	7.87
5	7.56	7.80	7.47	7.16	6.86
6	7.17	7.39	7.08	6.79	6.50
7	5.85	6.03	5.78	5.54	5.31
8	6.16	6.35	6.09	5.83	5.59
9	7.33	7.56	7.24	6.94	6.65
10	7.42	7.65	7.33	7.03	6.73
11	8.98	9.26	8.88	8.51	8.15
12	10.58	10.91	10.46	10.02	9.60

value is 0, therefore it is omitted. The fitting results of the GWO-Holt-Winters model are shown in Tables 5 and 6.

The error analysis

The error of the two models was averaged. The results are given in Figure 3 and Table 7. It can be seen that the MAPE of the PSO-FGSM(1,1) model is lower than that of the GWO-Holt-Winters model. It proves that the PSO-FGSM(1,1) model is more suitable for processing data with seasonal fluctuations. Seasonality is a major feature of rivers and water pollution is affected by the characteristics of the river, presenting a cyclical change, but the total concentration of pollutants keeps changing steadily

Table 5 | Fitting results of the GWO-Holt-Winters model at Huaibin monitoring points

Month	2015 fitting value	2016 fitting value	2017 fitting value	2018 fitting value
1	13.13	11.93	10.30	9.81
2	18.04	12.63	9.08	10.39
3	13.16	11.95	7.10	10.06
4	10.45	8.79	6.78	7.81
5	7.77	7.17	6.67	6.20
6	8.22	6.44	7.58	5.57
7	7.79	7.44	7.47	6.37
8	6.86	6.71	7.38	6.10
9	6.71	8.08	7.24	6.65
10	6.67	8.13	6.50	6.79
11	9.24	8.21	6.23	8.28
12	10.46	9.57	8.74	9.94

from year to year. According to these variation characteristics of river pollutants, the years and months data were calculated by the PSO-FGSM(1,1) model. For the yearly data, the PSO-FGM(1,1) model is used to predict the total annual value. For a small number of samples of data that do not fluctuate significantly, the PSO-FGM(1,1) model has a better processing effect. The optimization algorithm was used to find the optimal order, making the results more accurate. For the monthly data, different years of data for the same month are processed, and the seasonal indexes are calculated. The fluctuation characteristics of the data are reflected in the seasonal indexes. The combination of data for years and months results in more and more accurate predictions, overcoming the shortcoming of the traditional grey model applied to a small number of samples, and the prediction accuracy of the seasonal model is improved.

DO prediction using the PSO-FGSM(1,1) model

Through error analysis, the proposed model is proven to have a better effect. The model is applied to DO prediction of rivers, which can provide reference points for short- and long-term river pollution control. The principle of protection by stages for river pollution control has been widely adopted. One is pollution control characterized by seasons, and the other is by years. The PSO-FGSM(1,1) model responds to this principle, giving an annual and monthly forecast of pollutant indexes. The DO predictions for the next 2 years at the three monitoring sites are shown in Figure 4.

Table 6 | Fitting results of the GWO-Holt-Winters model at Bengbu and Huaibin monitoring points

	2015		2016		2017		2018	
Month	**Bengbu**	**Chuzhou**	**Bengbu**	**Chuzhou**	**Bengbu**	**Chuzhou**	**Bengbu**	**Chuzhou**
1	9.89	13.43	10.08	11.78	8.27	12.01	8.61	10.49
2	13.35	17.75	8.93	11.46	8.05	9.46	9.29	10.87
3	8.84	11.70	8.32	10.73	7.45	8.73	7.85	10.33
4	7.36	9.25	6.46	9.11	5.94	8.24	5.20	7.23
5	6.39	7.48	5.91	8.14	5.72	7.22	5.07	6.78
6	5.23	7.43	5.33	8.04	5.03	6.84	4.00	6.03
7	4.87	6.01	4.85	6.47	4.35	5.66	2.88	4.80
8	3.74	5.22	6.72	5.26	3.95	6.59	3.69	5.75
9	6.19	5.76	6.59	8.24	4.11	6.92	4.04	6.38
10	6.90	8.29	7.71	8.07	3.99	5.93	4.85	6.87
11	8.92	8.72	9.13	11.54	4.59	6.21	7.41	9.56
12	10.76	10.33	9.32	13.19	7.81	7.86	9.44	10.99
MAPE	13.95	14.00	10.30	11.00	9.31	11.00	11.47	9.00

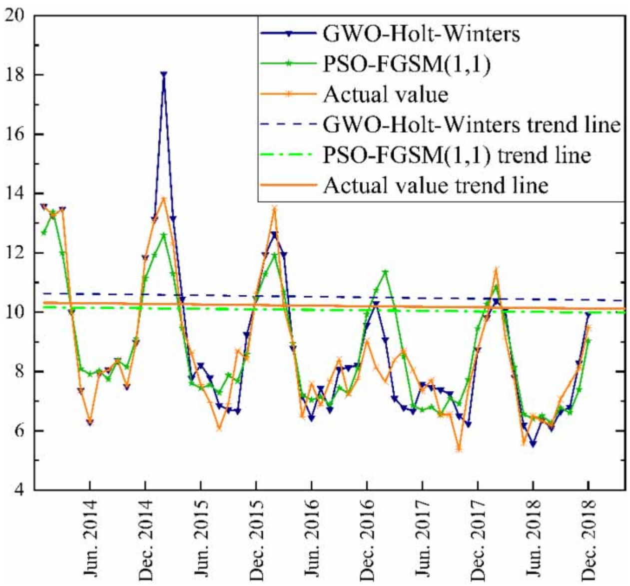

Figure 3 | The fitting results associated with the two different models in Huaibin.

Table 7 | The MAPE of the two models

Model	**Huaibin**	**Bengbu**	**Chuzhou**
PSO-FGSM(1,1)	8.66%	10.73%	9.74%
GWO-Holt-Winters	10.29%	13.20%	11.62%

The prediction results show that the seasonal variation characteristics of the three monitoring points are obvious and similar. Summer is still the peak season for pollution, and pollution is relatively light in winter. The pollution in Bengbu is more serious than in Huaibin and Chuzhou, so the Bengbu government should formulate policies to control pollution according to its seasonal characteristics. Measures for water pollution control and remediation should be implemented. In summer, the emphasis should be on governance, and strict water control measures should be implemented. The management of industrial and urban sewage discharge should be strengthened to make it meet standards before discharge. In winter, ecological restoration is emphasized. According to the annual characteristics of water pollution, short- and long-term water pollution control measures need to be developed. Many measures need to be taken to curb the trend of water pollution.

The calculation and prediction of pH using the PSO-FGSM (1,1) model

The normal pH range is 6.5–8.3, and pollution changes the pH. There are so many different kinds of pollutants in river that it is difficult to detect them fully. Measuring pH can quickly represent the general situation of river pollution, providing a direction for targeted detection.

After summing the data of the same year, the results fitted by the PSO-FGSM(1,1) model are shown in Table 8. The average errors are 1.74%, 0.42%, and 2.04%, so the fitting accuracy is high. The calculated results of seasonal indexes are shown in Figure 5. The normal standard for

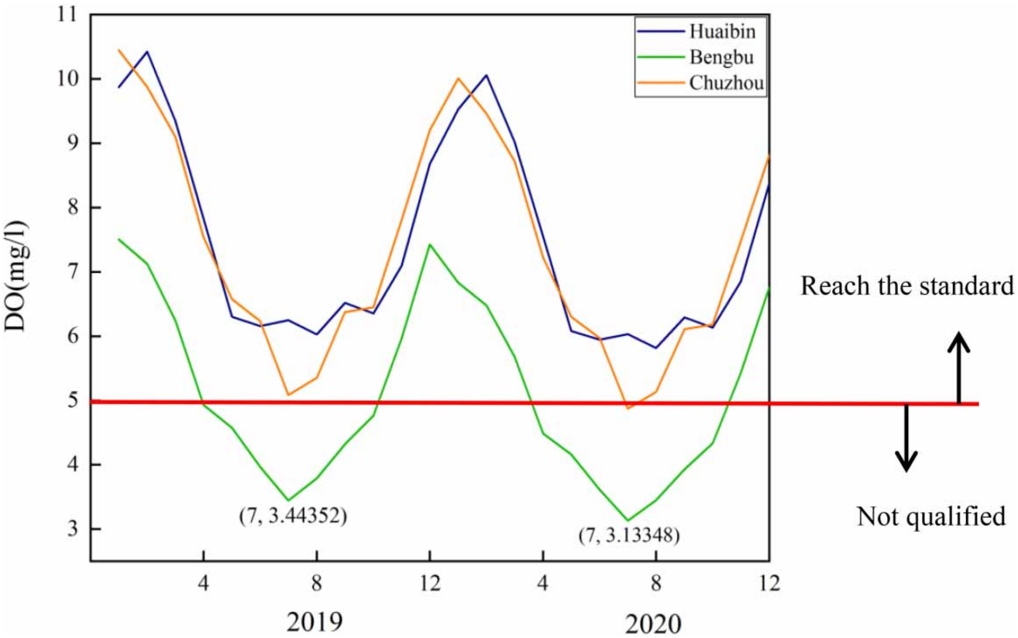

Figure 4 | Prediction results of DO at the three different monitoring sites.

pH is a range. Once outside this range, it is considered abnormal. In addition, pH value has no dimensions. Therefore, it makes more sense to analyze the predicted results. Error was averaged over the same year, and the results are shown in Table 9. It can be seen that the fitting effect on the pH values of the three monitoring points is good. The accuracy of future prediction is guaranteed.

The predicted results for 2019 and 2020 are shown in Figure 6. From these, the conclusion is consistent with the prediction results of DO. The pollution in the Bengbu section of the Huaihe River is serious and the water quality is alkaline, which will worsen. It is therefore urgent to control the Huaihe River in Bengbu section. Huaibin is upstream of Bengbu, while Chuzhou is downstream. The river water quality in both cities is clearly better than in Bengbu. In the pH forecasts for Huaibin and Chuzhou, there were few months of pollution and overall water quality improved. River pollution control is a long-term process, and it is difficult to restore the river's ecosystem. The government departments of Huaibin and Chuzhou should formulate relevant policies on the ecology of the river basin, and prevent the recurrence of river pollution.

Table 8 | Fitting results of the PSO-FGSM(1,1) model at the three different monitoring points

	Huaibin		**Bengbu**		**Chuzhou**	
	Actual value	**Fitting value**	**Actual value**	**Fitting value**	**Actual value**	**Fitting value**
2014	92.81	92.81	89.27	89.27	94.81	94.81
2015	92.75	89.89	88.67	88.17	94.88	94.05
2016	88.78	88.80	87.08	87.85	97.73	96.82
2017	85.42	88.70	88.16	88.34	93.44	97.41
2018	89.87	89.88	91.11	91.14	98.73	96.66
MAPE		1.74		0.42		2.04

As for Bengbu, river pollution control remains a priority. The cause of pollution in the Bengbu section of the Huaihe River is complicated. Bengbu is a general industrial base in Anhui province and an important old industrial city. It is the largest central city on the main branch of the Huaihe River, and listed in the top 100 wealthy Chinese cities. Industry accounts for 44.5% of local gross domestic product. The extensive development has seriously damaged the ecological environment for a long time. The specific reasons are as follows.

Firstly, the construction of municipal sewage treatment plants and supporting pipe networks is old-fashioned. Since the reform and opening up, the process of urbanization has been developing rapidly, and the population has been moving to urban areas. The city's sewage capacity has not kept pace with economic development. As a result, a large amount of substandard sewage is directly discharged into the Huaihe River, polluting the environment.

Secondly, Bengbu is located on a plain and has abundant water resources. It is an important grain-producing area in

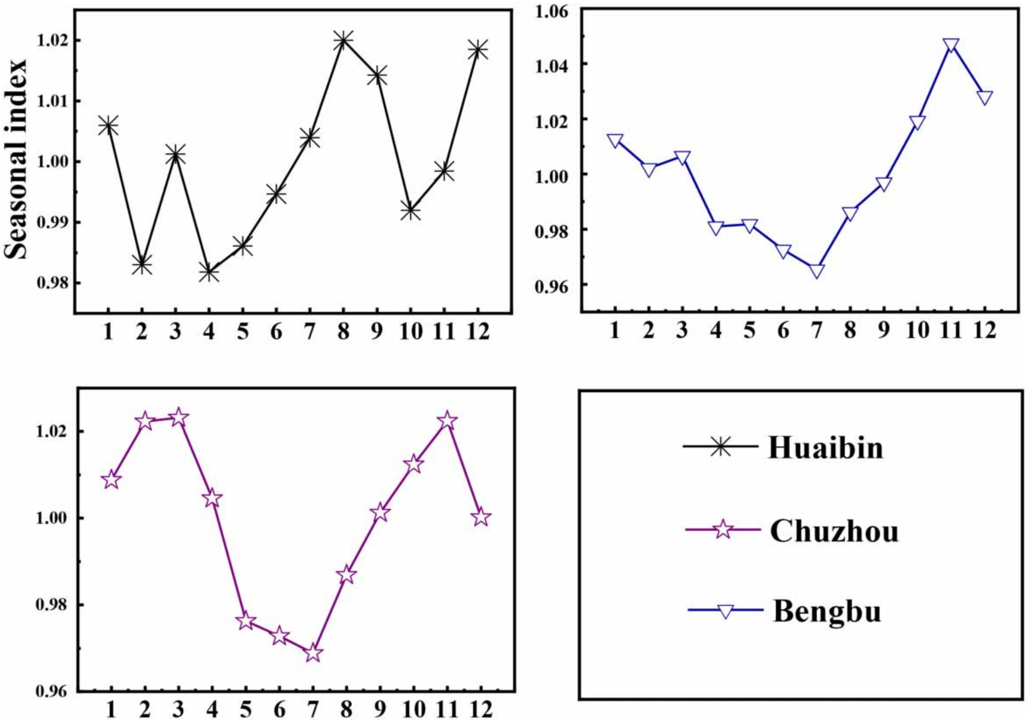

Figure 5 | Seasonal indexes of pH.

Table 9 | Error analysis of pH

MAPE	Huaibin	Bengbu	Chuzhou
2014	2.72%	2.24%	2.97%
2015	3.50%	2.05%	2.16%
2016	3.13%	2.01%	1.44%
2017	3.88%	2.11%	4.31%
2018	1.81%	1.77%	2.29%

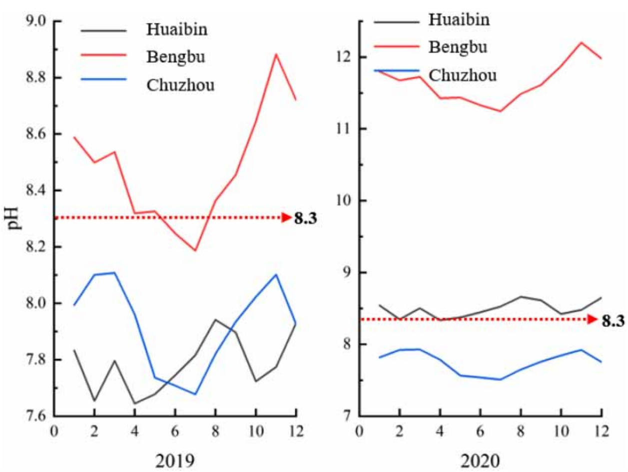

Figure 6 | Prediction results of pH at the three different monitoring sites.

China. Large areas of arable land have increased the use of pesticides and chemical fertilizers. The heavy use of pesticides and fertilizers not only pollutes rivers, but also soils and groundwater. The self-purification capacity of the agricultural ecosystems has been breached, resulting in pollutants being discharged directly into the nearby Huaihe River.

Thirdly, abundant water and vast plains create conditions for aquaculture. As a nationally important food industry, Bengbu's agricultural and associated food processing industry have developed rapidly. Aquaculture has polluted the environment.

Fourthly, there are many industries in Bengbu, including for chemicals, building materials and non-ferrous metals. The existence of highly polluting industries worsens the local environment. To cut costs, high-polluting companies discharge untreated sewage into the Huaihe River.

These not only reduce the output of aquatic products, but also harm people's health. The deteriorating ecological environment in Bengbu has aroused wide concern. The government has improved environment governance laws and regulations, such as the 2019 work plan for the prevention and control of water pollution in Bengbu city. The treatment of river pollution must be sustained. On the one hand, the government will deepen supply-side reform and develop clean energy. On the other hand, enterprises with high energy consumption, water consumption and pollution have been restricted. Lastly, existing enterprises must meet standards of sewage discharge. Moreover, in order to control the pollution of the Huaihe River, Bengbu is listed as a comprehensive pollution control city. Under the government's strict sewage treatment measures, the trend of water pollution in Bengbu section of Huaihe River has been restrained. The government should actively continue to promote the transformation of energy structure, and coordinate economic development and water pollution control.

CONCLUSIONS

In this study, we have highlighted the seasonal changes occurring in the DO and pH of the Huaihe River. The seasonal indexes were multiplied by the predicted results of the PSO-FGM(1,1) model. We obtained the following conclusions.

The new model was empirically compared with the Holt-Winters model. The results showed that the PSO-FGSM(1,1) model has a better prediction effect in processing seasonal data, improving the accuracy of the traditional seasonal model.

Under the strict control of government departments, the overall pollution trend in the Huaihe River has been curbed, but the forms of pollution have expanded regionally. The prediction results for Huaibin and Chuzhou showed that the local water quality is good, and the Bengbu prediction results showed poor water quality. The Bengbu section of the Huaihe River basin is the focus of water pollution control.

In this paper, DO and pH, two pollution indexes that affect fish growth, were researched. Other pollution indicators can be further studied in the future, including chemical oxygen demand, ammonia nitrogen, potassium permanganate index and heavy metals. A more comprehensive reference for treating water pollution will be provided.

ACKNOWLEDGEMENTS

The relevant researches are supported by the National Natural Science Foundation of China (7187 1084), the Excellent Young Scientist Foundation of Hebei Education Department (SLRC2019 001) and the Natural Science Foundation of Hebei Province (E2020402074).

DATA AVAILABILITY STATEMENT

All relevant data are included in the paper or its Supplementary Information.

REFERENCES

Alvisi, S. & Franchini, M. 2012 Grey neural networks for river stage forecasting with uncertainty. *Physics and Chemistry of the Earth, Parts A/B/C* **42**, 108–118. https://doi.org/10.1016/j.pce.2011.04.002.

An, Y., Zou, Z. & Zhao, Y. 2015 Forecasting of dissolved oxygen in the Guanting reservoir using an optimized NGBM (1, 1) model. *Journal of Environmental Sciences* **29**, 158–164. https://doi.org/10.1016/j.jes.2014.10.005.

Basant, N., Gupta, S., Malik, A. & Singh, K. P. 2010 Linear and nonlinear modeling for simultaneous prediction of dissolved oxygen and biochemical oxygen demand of the surface water-a case study. *Chemometrics and Intelligent Laboratory Systems* **104** (2), 172–180. https://doi.org/10.1016/j.chemolab.2010.08.005.

Cheng, C., Miao, S., Luo, B. & Sun, Y. 2017 Forecasting monthly energy production of small hydropower plants in ungauged basins using grey model and improved seasonal index. *Journal of Hydroinformatics* **19** (6), 993–1008. https://doi.org/10.2166/hydro.2017.062.

Deng, J. 1989 Introduction to grey system theory. *The Journal of Grey System* **1** (1), 1–24. https://doi.org/10.1007/978-3-642-16158-2_1.

Heddam, S. 2016 Use of optimally pruned extreme learning machine (OP-ELM) in forecasting dissolved oxygen concentration (DO) several hours in advance: a case study from the Klamath River, Oregon, USA. *Environmental Processes* **3** (4), 909–937. https://doi.org/10.1007/s40710-016-0172-0.

Huan, J., Cao, W. & Qin, Y. 2018 Prediction of dissolved oxygen in aquaculture based on EEMD and LSSVM optimized by the Bayesian evidence framework. *Computers and Electronics in Agriculture* **150**, 257–265. https://doi.org/10.1016/j.compag.2018.04.022.

Ji, X., Shang, X., Dahlgren, R. A. & Zhang, M. 2017 Prediction of dissolved oxygen concentration in hypoxic river systems using support vector machine: a case study of Wen-Rui Tang River, China. *Environmental Science and Pollution Research* **24** (19), 16062–16076. https://doi.org/10.1007/s11356-017-9243-7.

Karmakar, S. & Mujumdar, P. P. 2006 Grey fuzzy optimization model for water quality management of a river system. *Advances in Water Resources* **29** (7), 1088–1105.

Keshtegar, B. & Heddam, S. 2017 Modeling daily dissolved oxygen concentration using modified response surface method and artificial neural network: a comparative study. *Neural Computing and Applications* **30** (10), 2995–3006. https://doi.org/10.1007/s00521-017-2917-8.

Khan, U. T. & Valeo, C. 2015 A new fuzzy linear regression approach for dissolved oxygen prediction. *Hydrological Sciences Journal* **60** (6), 1096–1119. https://doi.org/10.1080/02626667.2014.900558.

Kisi, O. & Parmar, K. S. 2016 Application of least square support vector machine and multivariate adaptive regression spline models in long term prediction of river water pollution. *Journal of Hydrology* **534**, 104–112. https://doi.org/10.1016/j.jhydrol.2015.12.014.

Li, C., Li, Z., Wu, J., Zhu, L. & Yue, J. 2018a A hybrid model for dissolved oxygen prediction in aquaculture based on multi-scale features. *Information Processing in Agriculture* **5** (1), 11–20. https://doi.org/10.1016/j.inpa.2017.11.002.

Li, S., Meng, W. & Xie, Y. 2018b Forecasting the amount of waste-sewage water discharged into the Yangtze River basin based on the optimal fractional order grey model. *International Journal of Environmental Research and Public Health* **15** (1), 20. https://doi.org/10.3390/ijerph15010020.

Liu, S., Xu, L., Li, D., Li, Q., Jiang, Y., Tai, H. & Zeng, L. 2013 Prediction of dissolved oxygen content in river crab culture based on least squares support vector regression optimized by improved particle swarm optimization. *Computers and Electronics in Agriculture* **95**, 82–91. https://doi.org/10.1016/j.compag.2013.03.009.

Liu, Y., Zhang, Q., Song, L. & Chen, Y. 2019 Attention-based recurrent neural networks for accurate short-term and long-term dissolved oxygen prediction. *Computers and Electronics in Agriculture* **165**, 104964. https://doi.org/10.1016/j.compag.2019.104964.

Moss, B. 2007 Water pollution by agriculture. *Philosophical Transactions of the Royal Society B: Biological Sciences* **363** (1491), 659–666. http:// doi.org/10.1098/rstb.2007.2176.

Najah, A., Elshafie, A., Karim, O. A. & Jaffar, O. 2009 Prediction of Johor River water quality parameters using artificial neural networks. *European Journal of Scientific Research* **28** (3), 422–435. http://dx.doi.org/.

Ross, A. C. & Stock, C. A. 2019 An assessment of the predictability of column minimum dissolved oxygen concentrations in Chesapeake Bay using a machine learning model. *Estuarine, Coastal and Shelf Science* **221**, 53–65. https://doi.org/10.1016/j.ecss.2019.03.007.

Schwarzenbach, R. P., Egli, T., Hofstetter, T. B., Von Gunten, U. & Wehrli, B. 2010 Global water pollution and human health. *Annual Review of Environment and Resources* **35**, 109–136. https:// doi.org/10.1146/annurev-environ-100809-125342.

Tang, C., Yi, Y., Yang, Z. & Cheng, X. 2014 Water pollution risk simulation and prediction in the main canal of the South-to-North Water Transfer Project. *Journal of Hydrology* **519**, 2111–2120.

Tseng, F., Yu, H. & Tzeng, G. 2001 Applied hybrid grey model to forecast seasonal time series. *Technological Forecasting and Social Change* **67** (2–3), 291–302. https:// doi.org/10.1016/S0040-1625(99)00098-0.

Wang, H. 1993 The grey algebraical curve model for municipal water supply forecasting. *Systems Engineering-Theory & Practice* **1993**, 3.

Wang, Z., Li, Q. & Pei, L. 2018 A seasonal GM(1,1) model for forecasting the electricity consumption of the primary economic sectors. *Energy* **154**, 522–534. https://doi.org/10.1016/j.energy.2018.04.155.

Winters, P. R. 1960 Forecasting sales by exponentially weighted moving averages. *Management Science* **6** (3), 324–342. https:// doi.org/10.1278/mnsc.6.3.324.

Wu, L., Liu, S., Yao, L., Yan, S. & Liu, D. 2013 Grey system model with the fractional order accumulation. *Communications in Nonlinear Science and Numerical Simulation* **18** (7), 1775–1785. https://doi.org/10.1016/j.cnsns.2012.11.017.

Wu, L., Liu, S., Fang, Z. & Xu, H. 2015 Properties of the GM(1,1) with fractional order accumulation. *Applied Mathematics and Computation* **252**, 287–293. https://doi:10.1016/j.amc.2014.12.014.

Wu, H. A., Zeng, B. & Zhou, M. 2017 Forecasting the water demand in Chongqing, China using a grey prediction model and recommendations for the sustainable development of urban water consumption. *International Journal of Environmental Research and Public Health* **14** (11), 1386. https:// doi.org/10.3390/ijerph14111386.

Xiao, X., Yang, J., Mao, S. & Wen, J. 2017 An improved seasonal rolling grey forecasting model using a cycle truncation accumulated generating operation for traffic flow. *Applied Mathematical Modelling* **51**, 386–404. https://doi.org/10.1016/j.apm.2017.07.010.

First received 6 September 2020; accepted in revised form 3 December 2020. Available online 15 December 2020