

PAPER • OPEN ACCESS

Prediction of water quality based on artificial neural network with grey theory

To cite this article: W Zhai *et al* 2019 *IOP Conf. Ser.: Earth Environ. Sci.* **295** 042009

View the [article online](#) for updates and enhancements.

Prediction of water quality based on artificial neural network with grey theory

W Zhai¹, X Zhou^{2,3}, J Man⁴, Q Xu¹, Q Jiang¹, Z Yang¹, L Jiang¹, Z Gao¹, Y Yuan¹, W Gao^{1,5}

¹ School of Safety Engineering, Ningbo University of Technology, Ningbo, Zhejiang, 315211, China

² School of Safety Science and Emergency Management, Wuhan University of Technology, Wuhan 430070, Hubei, China

³ School of Resources and Environmental Engineering, Wuhan University of Technology, Wuhan 430070, Hubei, China

⁴ Ningbo Liwah Pharmaceutical Co., Ltd. Zhenhai Branch, Ningbo, Zhejiang, 315204, China

⁵ Author to whom any correspondence should be addressed.

Abstract. In this paper, the grey theory, three type of artificial neural network (back-propagation neural network, radial basis function neural network, and generalized regression neural network) and their combination were used to predict the pH values in the evaluation of water quality. Based on the measured data from the Xielugang in Jiaxin with the post-hoc analysis for the c and p values of the prediction, the results showed that the prediction by using the generalized regression neural network has the averaged relative error 0.61%, and $c < 0.65$, $p > 0.7$.

1. Introduction

With the rapid development of Chinese economy, the life level of Chinese people has been dramatically improved. However, the environment in some places in China have been damaged, which included the water quality. In order to protect the environment, it is important to protect the water quality for the rivers and lakes [1]. It can provide a solid support if the water quality can be predicted. Because the factors affecting the water quality are various, and the change in water quality has the nonlinear characteristics and time-varying parameters, current sensors cannot be reacted effectively [2]. However, the modern mathematics and artificial intelligence technology can provide a path to build the model to predict water quality.

The grey theory and artificial neural network (ANN) can be used to predict the water quality. The grey theory can analyze the discrete data and find the relationship between those discrete data, which can reflect the nonlinear relationship among the various factors and resulting in the dynamic and real-time predicting of the water quality [3]. ANN is also a common method for the prediction of water quality. ANN can transform the nonlinear problems in low-dimension into a separable problem in high-dimension and use it to predict the water quality. Currently, the most common ANN for predicting the water quality is the back-propagation neural network [4]. The combination of the grey theory and ANN showed the advantages of both methods for predicting the water quality [5].



2. Method

2.1. Grey theory

The grey theory is suitable for the research which has the insufficient data. The internal relationship of the insufficient data can be found by using the data mining. Based on the grey theory, the raw data is transformed by the technique of accumulative generation operation, so that the volatility in the raw data are smoothed, and a new sequence shows the strong regularity.

The steps for building the grey GM (1,1) model are as follows:

Set the raw data sequence:

$$X^{(0)} = \{x^{(0)}(1), x^{(0)}(2), \dots, x^{(0)}(n)\}$$

Generate the new sequence:

$$X^{(1)} = \{x^{(1)}(1), x^{(1)}(2), \dots, x^{(1)}(n)\}$$

$$\text{where } x^{(1)}(k) = \sum_{i=1}^k x^{(0)}(i) \quad (i)k = 1, 2, \dots, n$$

Then, generate the mean value:

$$Z^{(1)} = \{z^{(1)}(1), z^{(1)}(2), \dots, z^{(1)}(n)\}$$

$$z^{(1)}(k) = 1/2 [x^{(1)}(k) + x^{(1)}(k-1)] \quad k=2, 3, \dots, n$$

In order to establish the GM(1,1) model, the differential equation based on the first-order grey module is as follows:

$$\begin{cases} \frac{dx^{(1)}(t)}{dt} + ax^{(1)}(t) = u \\ x^{(1)}(1) = x^{(0)}(1) \end{cases}$$

The solutions of the above differential equation is:

$$x^{(1)}(t) = [x^{(0)}(1) - u/a]e^{-at} + u/a$$

a and u values are:

$$[\hat{a}, u]^T = (A^T A)^{-1} A^T Y$$

where A is:

$$A = \begin{bmatrix} -\frac{1}{2}[x^{(1)}(1) + x^{(1)}(2)] & 1 \\ -\frac{1}{2}[x^{(1)}(2) + x^{(1)}(3)] & 1 \\ \vdots & \vdots \\ -\frac{1}{2}[x^{(1)}(n-1) + x^{(1)}(n)] & 1 \end{bmatrix}$$

$$Y = \begin{bmatrix} x^{(0)}(2) \\ x^{(0)}(3) \\ \vdots \\ x^{(0)}(n) \end{bmatrix}$$

Substituting the least square method solution of a and u, the approximate solution is:

$$\hat{x}^{(1)}(t) = \left[x^{(0)}(1) - \frac{\hat{u}}{\hat{a}} \right] e^{-\hat{a}(t-1)} + \frac{\hat{u}}{\hat{a}}$$

The GM (1,1) dynamic prediction model is:

$$\hat{x}^{(0)}(k+1) = (1 - e^{\hat{a}}) \left[x^{(0)}(1) - \frac{\hat{u}}{\hat{a}} \right] e^{-\hat{a}k}$$

2.2. Artificial neural networks

ANNs are the hot research in the field of artificial intelligence. The most common ANN is the back-propagation neural network (BPNN). In this paper, radial basis function neural network (RBFNN), and generalized regression neural network (GRNN) were also utilized.

2.3. Predictive index

The accuracy of the prediction value is determined by the error test formula. The test formula is shown in **Table 1**. And the evaluation level is shown in **Table 2**.

Table 1. Error test calculation formula

	Residual	Raw data
Mean difference	$\bar{\varepsilon} = \frac{1}{n} \sum_{i=1}^n \varepsilon_i^{(0)}$	$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i^{(0)}$
variance	$S_1^2 = \frac{1}{n} \sum_{i=1}^n (\varepsilon_i^{(0)} - \bar{\varepsilon})^2$	$S_2^2 = \frac{1}{n} \sum_{i=1}^n (X_i^{(0)} - \bar{X})^2$
Posterior difference test ratio	$c = \frac{S_1}{S_2}$	
Error probability	$p = P\{ \varepsilon_i^{(0)} - \bar{\varepsilon} < 0.6745 S_2\}$	

Table 2. Evaluation level

Prediction evaluation level	P	c
good (Level 1)	>0.95	<0.35
qualified (Level 2)	>0.8	<0.5
reluctant (Level 3)	>0.7	<0.65
failed (Level 4)	≤ 0.7	≥ 0.65

2.4. Experimental data

pH is an important indicator to measure the water quality. Therefore, in this paper pH was the one which was predicted. The data in this article are from the weekly report of the cross section of Jiaying Xielu Port in the Taihu Lake, which is published by the Ministry of Environmental Protection of the People's Republic of China. The selected data included 28 groups from 2 weeks to 29 weeks in 2017. In addition to the pH, the data also included DO, COD_{mn}, and NH₃-N.

3. Result

By using the grey theory, the pH value of water quality was predicted. The results are shown in **Table 3**. We use the 2-21 weeks' data as the input to training ANNs, and then predict the pH value in 22-29 week. The c and p values then can be calculated by the error test formula. The results are shown in **Table 4**.

Table 3. Predictions of pH values by using the grey theory and artificial neural network

Week	Measured value	Grey theory		Artificial neural networks					
				BPNN		RBFNN		GRNN	
		Predictive value	Relative error	Predictive value	Relative error	Predictive value	Relative error	Predictive value	Relative error
22	7.02	6.97	-0.67%	7.44	6.01%	7.08	0.89%	7.05	0.38%
23	7.12	6.97	-2.15%	7.50	5.33%	7.06	-0.83%	7.06	-0.89%
24	7.07	6.96	-1.55%	7.77	9.95%	7.06	-0.15%	7.02	-0.68%
25	7.01	6.95	-0.79%	7.84	11.8%	7.12	1.55%	7.03	0.29%
26	6.91	6.95	0.56%	7.55	9.29%	7.11	2.90%	7.03	1.67%
27	7.00	6.94	-0.82%	7.42	6.06%	7.11	1.57%	7.04	0.54%
28	7.04	6.94	-1.47%	7.35	4.45%	7.10	0.82%	7.04	-0.01%
29	7.01	6.93	-1.13%	7.30	4.11%	7.08	0.96%	7.04	0.45%

Table 4. Predictive index by using grey theory and artificial neural networks

	Average relative error	c	P
Grey	1.14%	0.65	0.75
BPNN	7.13%	3.29	0.00
RBFNN	1.21%	0.93	0.75
GRNN	0.61%	0.57	0.75

Based on the **Table 3**, the averaged relative error by using the grey theory is 1.14%; and the averaged relative error by using BPNN, RBFNN and GRNN are 7.13%, 1.21% and 0.61% respectively. It demonstrated that the deviation between the predicted and measured values by using BPNN is larger than that of grey theory, RBFNN and GRNN. The relative error that was predicted by RBFNN method is comparable to that of grey theory. The relative error by using GRNN is smaller than that of grey theory. From **Table 4**, it showed that the precision of prediction by using GRNN is higher than other three methods. The GRNN method can achieve the third level and it is very close to the second level. Therefore, by using GRNN, pH value can be predicted.

4. Conclusion

In this paper, the grey theory and artificial neural networks have been used to predict water quality. The results showed that the GRNN have the best performance in prediction. The precision of the prediction by using GRNN is in the third level and it is very close to the second level.

Acknowledgments

This paper is supported by Ningbo University of Technology College Student Science and Technology Innovation Fund (2016041), Zhejiang Science and Technology Innovation Plan Fund (2018R428011) and Ningbo University of Technology Starup Fund (D2016009).

References

- [1]Sun Z, Wang B and Ji H 2011 Water quality prediction based on probability combination *China Environmental Science* **31** 1657-1662.
- [2]Wang X, Zhao X and Liu Z 2011 Research on lake eutrophication forecasting methods based on grey theory *Computer Simulation* **28** 17-19.
- [3]Zhou X, Tang J and Shi B 2016 Analysis and forecast of influential factors of gas content in deep coal seam on the basis of the grey entropy *Coal Geology & Exploration* **44** 19-23.
- [4]Zhao X, Lei X, Shen Z and Liu Z 2007 Application of artificial neural network to Xinjiang Muguahu reservoir water quality evaluation *Journal of Shihezi University (Natural Science)* **25** 236-239.
- [5]Zhou L, Mu Z, Xing W and Kong G 2011 Study of service life prediction of organic coatings based on grey neural network *Equipment Environmental Engineering* **8** 62-65.