CHEMICAL ENGINEERING TRANSACTIONS

VOL. 66, 2018

A publication of AIDIC

The Italian Association of Chemical Engineering
Online at www.aidic.it/cet

Guest Editors: Songying Zhao, Yougang Sun, Ye Zhou

ISBN 978-88-95608-63-1; **ISSN** 2283-9216

DOI: 10.3303/CET1866103

Water Pollution Evaluation in Lakes Based on Factor Analysis-Fuzzy Neural Network

Cheng Yang[a], Suiju Lv[a], Feng Gao[b]*

[a]College of Civil Engineering, North Minzu University, Yinchuan 750021, China
[b]School of General Education, North Minzu University, Yinchuan 750021, China
gaofeng0502@163.com

The water quality assessment factors directly determine whether the results are grounded in reality. There is a traditional way, i.e. the fuzzy neural network, which not only has great subjectivity in the selection of input samples but also lacks scientific rationality. This paper uses the factor analysis to ensure the availability of input samples. With the Yuehai Lake as a study case, a water quality assessment is conducted using a factor analysis-fuzzy neural network model. First, the factor analysis performs dimension reduction process on input samples to identify the variance contributions of evaluation and other factors. Next, the linspace function of MATLAB interpolates across different levels of evaluation criteria at an equal interval to generate the sets of samples, and assess the water quality using the factor analysis-fuzzy neural network model available after training and testing. The findings show that the water quality of Yuehai Lake seems good, as Class I-II. This method can also bear out when the more sever pollution occurs in the case of Class II water. It is proved by the instance of the Yuehai Lake water quality identification that the factor analysis-fuzzy neural network model is feasible and easy to operate, and even more, it can derive more practical results.

1. Introduction

The water quality assessment is qualitative and provides scientific clues to water pollution control and management. There are a lot of non-linear, non-stationary and uncertain factors in the water environment. The traditional water quality assessment models such as gray system approach, water quality identification index method, fuzzy comprehensive index method, and analytic hierarchy process (Karmakar and Mujumdar, 2006; Xu et al., 2010; Zhu and Chen, 2011; Pang et al., 2008) all fail to accurately depict the nonlinear evolution process of the water environment system, while the fuzzy neural network model can organically integrate the fuzzy technology and the neural network to build a neural network or adaptive fuzzy system that allows "automatic" treatment for fuzzy information, so that many of practical problems can be settled by using this model. Chen and Li (2005) incorporated the artificial neural network and fuzzy recognition theory to construct a fuzzy artificial neural network recognition model and applied it to the comprehensive assessment on the water quality of the Yangtze River tributary Tuojiang River during the drought period. The results show that this model has the objectivity and practicality. Yang and Wei, (2007) blended the fuzzy system with the neural network and proposed a water quality assessment model which could enable a clear inference process, strong generalization. Zhou (2007) focused on how well two artificial intelligence methods, i.e. the fuzzy systems and neural networks, could seem and integrated them organically. Then the fuzzy neural network based on T-S model was applied to the water quality assessment with good effect.

As described above, when the fuzzy rules are determined, the performance of the fuzzy control system depends on the membership of each subset of the fuzzy variables. This is a multi-parameter optimization problem. In general, it is difficult to obtain a global optimum; the choice of water quality assessment factors directly affects whether the assessment results are grounded in reality. However, most scholars choose the water quality pollution assessment factors only according to the local pollution situation, that is, take some representative indicators as the assessment factors, lacking the scientific basis.

In view of the above issues, this study uses the factor analysis to determine what the weight of each factor is, and chooses the water quality indicator with the accumulated weight of 85% or above as the assessment

Please cite this article as: Yang C., Lv S., Gao F., 2018, Water pollution evaluation in lakes based on factor analysis-fuzzy neural network, Chemical Engineering Transactions, 66, 613-618 DOI:10.3303/CET1866103

factor (Gao and Feng, 2014). Then, a factor analysis-fuzzy neural network is used to assess the water quality with the Haihai Lake as an example to demonstrate the feasibility of this method.

2. Theory and method

The fuzzy modeling technology excessively depends on the veracity of the membership function. Fuzzy neural network as a model with strong self-adaptability not only realizes the automatic update of the fuzzy model, but also can timely amend the membership function of each fuzzy subset, making the fuzzy modeling more reasonable. The fuzzy neural network is a high-order feed-forward type. Unlike the BP neural network, it uses the multiplication neuron instead of the addition neuron in the output layer. It features that the connection weight of the hidden layer with the output layer takes l, no need to change it in the learning and training processes, so that the fuzzy neural network has less training parameters and a faster convergence speed. The fuzzy neural network model based on Takagi-Sugeno fuzzy system is shown in Figure 1.

In this network, there are four input neurons. S, P, and "•" respectively represent the addition, the multiplication, and logical operations (Nie and Deng, 2008; Zhou and Wang, 2010; Kosko, 1992; Li and Chen, 1996; Chen and Wu, 1997).

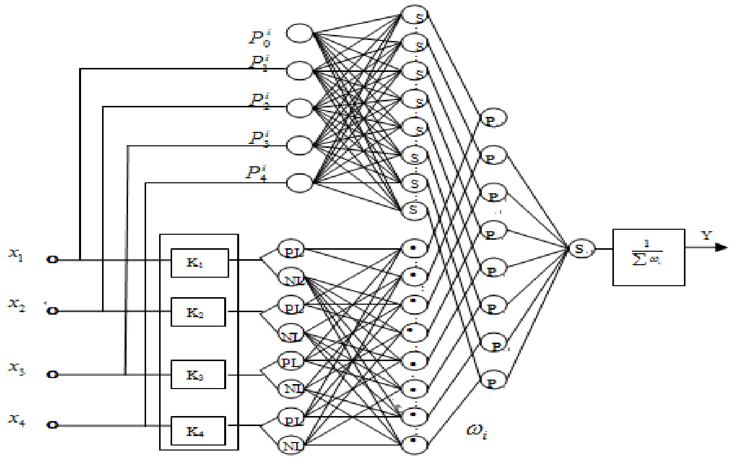

Figure 1: Pi‑Sigma fuzzy neural network with four inputs

2.1 The network output is given as follows:

$$y_n = \frac{\sum_{i=1}^{m} \omega^i y^i}{\sum_{i=1}^{m} \omega^i} = \frac{\sum_{i=1}^{m} \left[\mu_{A_1^i}(x_1) \cdot \mu_{A_2^i}(x_2) \cdot \mu_{A_3^i}(x_3) \cdot \mu_{A_4^i}(x_4) \left(p_0^i + p_1^i x_1 + p_2^i x_2 + p_3^i x_3 + p_4^i x_4 \right) \right]}{\sum_{i=1}^{m} \left[\mu_{A_1^i}(x_1) \cdot \mu_{A_2^i}(x_2) \cdot \mu_{A_3^i}(x_3) \cdot \mu_{A_4^i}(x_4) \right]} \tag{1}$$

2.2 Fuzzy neural network learning algorithm:

a. error calculation

$$e = \frac{1}{2}(y_d - y_c)^2 \tag{2}$$

b. coefficient correction

$$p_j^i(k) = p_j^i(k-1) - \alpha \frac{\partial e}{\partial p_j^i} \tag{3}$$

$$\frac{\partial e}{\partial p_j^i} = (y_d - y_c)\omega^i / \sum_{i=1}^{m} \omega^i \cdot x_j \tag{4}$$

Where, p_j^i is the neural network coefficient; α is the network learning rate; x_j is the network input parameter; ω^i is the input parameter membership degree continued product.
c. parameter correction

$$c_j^i(k) = c_j^i(k-1) - \beta \frac{\partial e}{\partial c_j^i} \tag{5}$$

$$b_j^i(k) = b_j^i(k-1) - \beta \frac{\partial e}{\partial b_j^i} \tag{6}$$

Where, c_j^i, b_j^i are the center and width of the membership function, respectively.

2.3 Building the water quality assessment model based on fuzzy neural network

The process of water quality assessment of fuzzy neural network based on factor analysis is shown in Figure 2.

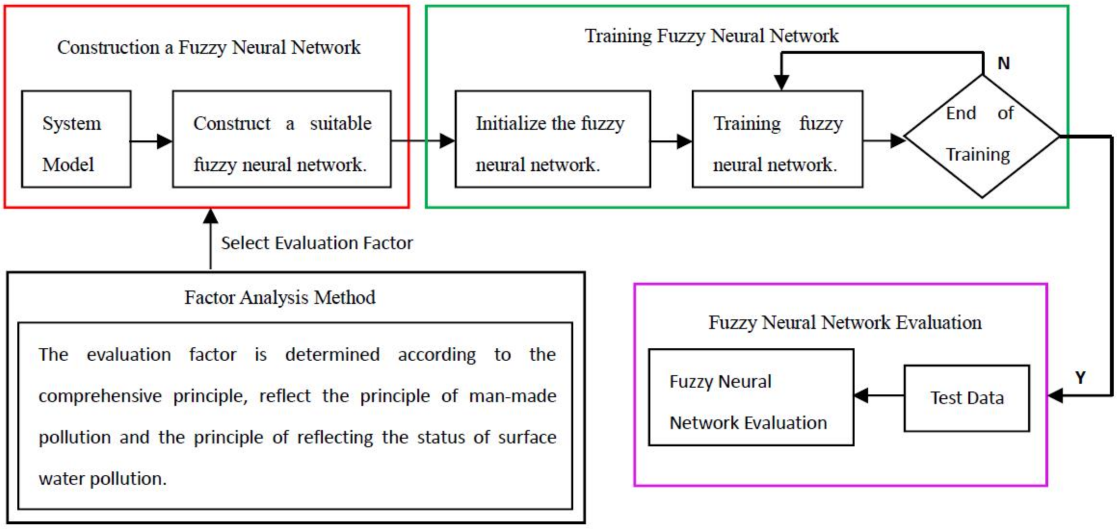

Figure 2: Algorithm flow of water quality assessment based on Factor analysis - fuzzy neural network

Where, the fuzzy neural network determines the numbers of fuzzy neural network input and output nodes and fuzzy membership functions according to the training samples. The fuzzy membership function has a center and width randomly obtained.

3. Application of factor analysis-fuzzy neural network in water quality assessment

Yuehai Wetland Park is located in Jinfeng District, Yinchuan, Ningxia, with a total area of 2013hm^2 and a core planning area of approximately 22km^2. The non-point source pollution caused by return water of upstream farmland irrigation and the lack of ecological revetment as necessary along the banks converges in the lake, leading to the pollution of the water body of the Haihai Lake. This source is monitoring data from the section of Haihu Lake in 2012 ~2016, combined with the actual situation, this study traces the following assessment indicators: PH, DO, ammonia nitrogen, potassium permanganate, COD, BOD_5, etc.

3.1 Selection of assessment factors

The matlab is used herein to carry out factor analysis on original data, first normalize the original indicator data to obtain the standardization matrix Z, conduct KMO and Bartlett test on it, and verify whether it is suitable for factor analysis (Lu et al., 2012; Lei et al., 2009), the result is shown as Table 1.

Table 1: Test results of KMO and Bartlett

Kaiser-Meyer-Olkin Measure of Sampling	Bartlett's Test of Sphericity		
	Approx	df	Sig.
0.512	189.992	73	0.00

The KMO value is 0.582>0.5, Bartlett test value is 189.992, and the freedom degree is 73. Its significance level Sig is 0.00<0.05, achieving the standard. The results show that it is suitable for factor analysis.
We learn from the factor analysis that when the common factors are 6, the cumulative contribution rate of its variance reaches 86.594%, which can represent most of information of the original assessment indicator. The variance contribution rate and factor score coefficient are shown in Tables 2 and 3.

Table 2: Variance contribution rates

Common factor	Contribution rate (%)	Cumulative cont. rate (%)
F1	35.647	35.647
F2	25.124	60.771
F3	15.228	75.999
F4	10.595	86.594

The weight of each indicator is calculated by the formula:

$$\omega_{\mathrm{i}} = \frac{\sum_{j=1}^{m} \beta_{ji} e_i}{\sum_{i=1}^{p} \sum_{j=1}^{m} \beta_{ji} e_i} \tag{7}$$

Where: ω is the weight; e is the variance contribution rate; β is the factor score coefficient

Table 3: Factor score coefficient

Indicator	F1	F2	F3	F4
DO	0.124	0.121	-0.155	-0.250
KMnO4	0.092	0.149	0.881	-0.122
BOD5	-0.121	0.432	0.048	0.141
NH4	0.029	0.269	0.120	-0.110
TP	-0.112	-0.013	-0.035	0.952
TN	0.198	0.240	0.085	0.184
F−	0.285	-0.009	0.066	-0.112
sulfate	-0.113	0.144	-0.002	0.069
Chlorides	-0.131	0.096	0.143	0.090

According to Tables 2, 3 and formula (7), the weight of each indicator can be available, as shown in Table 4.

Table 4: Weight of each indictor

DO	KMnO4	BOD5	NH4	TP	TN	F−	sulfate	Chlorides
0.034	0.269	0.123	0.119	0.073	0.229	0.137	0.004	0.012

The weights of the indicators in Table 4 are ranked in descending order. The cumulative weight of the first four indicators reaches 86.594%. Therefore, four indicators that affect water quality are chosen as assessment factors, i.e. they are in turn $KMnO_4$, TN, F^-, BOD_5.

3.2 Model application

In this paper, the four assessment indicators as chosen in Section 3.1 are used as the inputs of the factor analysis-fuzzy neural network, the water quality level as the output item, and the uniform interpolation water quality assessment standard as the training sample.

3.2.1 Generation of training samples

In fact, there are few data on water quality assessment. If the surface water quality grading standard is used as a training sample, the sample size is too small, inevitably resulting in a lack of network fit that affects the assessment results (Yang and Wu, 2004; Dai, 1999), hence to adopt MATLAB herein. The *linspace* function performs uniformly interval interpolation on grading standard data from surface water quality to generate a set of training samples, where there are 200 sets of data respectively generated from those less than Class I, between Class I ~ Class II, Class II ~ Class III, Class III ~ Class IV, Class IV ~ Class V, forming a total of 1,000 sets of data. In order to avoid loss of generality, 900 sets of data out of them are taken randomly as training samples and the remaining 100 sets of data are used as test samples.

3.2.2 Network training

This study uses the samples obtained from the environmental quality standard of surface water to train the model. First, the coefficients are randomly initialized. Based on a set of input values, the network output value is obtained according to formula (1), compare the output values and expected values, adjust the coefficient values, repeat the above steps 900 times, and the network is well trained, as shown in Figure 3. The remaining 100 sets of data are used as test samples to check the network with the results shown in Figure 4.
In order to verify the generalization performance and assessment precision of the network, 5 sets of classification standard values in the *National Standard for Surface Water Environment Quality* are used as test samples. After pretreatment, they are input into a trained network for testing. The results are shown in the Table 5.

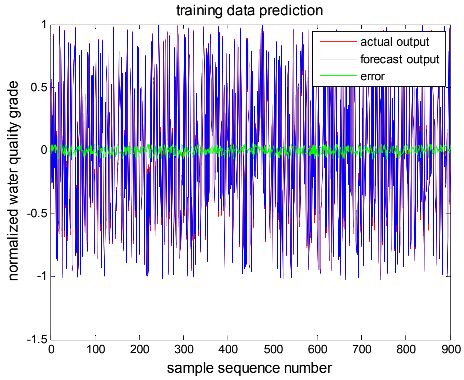

Figure 3: Training process of fuzzy neural network

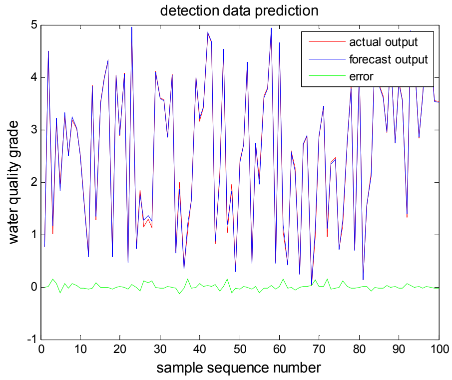

Figure 4: Fitting process of test sample

Table 5: Comparison of test results of standard samples

standard sample evaluation results	1	2	3	4	5
model classification results	1	2	3	4	5

As shown in Table 5, the assessment results of the standard samples coincide with the classification results after training. It is proved that the factor analysis-fuzzy neural network model features precise assessment and good generalization.

3.2.3 Water quality assessment

The trained model is used to assess the water quality of the monitoring section of the Haihai Lake in 2012-2016, and the results are shown in Figure 5:

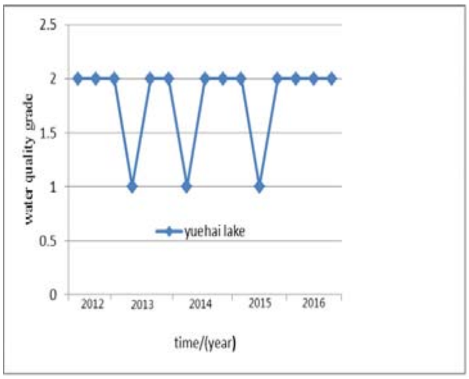

Figure 5: Water quality evaluation results of yuehai lake

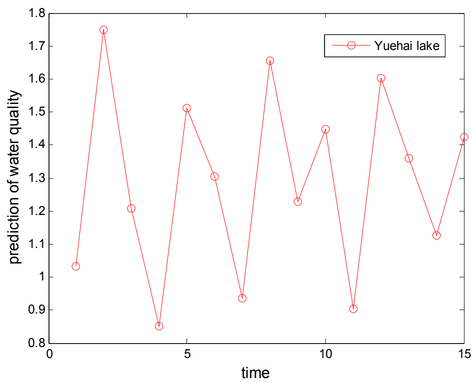

Figure 6: Actual value of evaluation result of yuehai lake

As shown in Figure 5, the assessment results of the Lake Haihai Lake water quality are I-II, and the water quality is good, but there is also a difference in whether the pollution is severe or not in the Class II water. Therefore, the grounded results are shown in Figure 6. It is clearly known in which time frame the water quality is better. This model specially applies to heavily polluted rivers or lakes.

4. Conclusions

(1) Aiming at the fact that multiple indicators spoil the effect of the overall water quality assessment, the factor analysis helps determine the variance contribution to scientifically choose the assessment factors and propose a factor analysis-fuzzy neural network model for this purpose herein. The results show that the Yuehai Lake water quality is in good conditions, regarded as Class I-II. The method can also determine when the pollution is worse in the case of Class II water.
(2) The instance of the Yuehai Lake water quality identification bears out that the factor analysis-fuzzy neural network identification model is feasible, easy to operate, and more of that, it can derives practical results, especially apply to those rivers or lakes polluted severely.

Acknowledgment

This research is supported by Key Scientific Research Projects in 2017 at North Minzu University (Item Number: 2017KJ39)

Reference

Chen S.Y., Li Y.W., 2005, Water quality evaluation based on fuzzy artificial neural network, Advances in Water Science, 16(1), 88-91.
Chen T.P., Wu X., 1997, Characteristics ofactivationfunctionin sigma-pi neural networks, Journal of Fudan University, 1997, 36(6), 639-644.
Dai W.Z., 1999, A Method of Multiobjective Synthetic Evaluation Based on Artificial Neural Networks and Its Applications, Systems Engineering-Theory & Practice, (5), 29-34,40.
Gao F., Feng M.Q., 2014, Ater Quality Evaluation Based on Modified VQ Neural Network, International Journal of Earth Sciences and Engineering, 27(5), 1721-1726,
Karmakar S., Mujumdar P.P., 2006, Grey fuzzy optimization model for water quality management of a river system, Adv Water Resour, 29(7), 1088-1105, DOI: 10.1016/j.advwatres.2006.04.003
Kosko B., 1992, Neural Networks and Fuzzy Systems: A Dyhamical Systems Approach to Machine Intelligence, NJ: Prentice Hall Inc, 28(8), 956-957.
Lei L.N., Shi W.R., Fan M., 2009, Water quality evaluation analysis based on improved SOM neural network, Chinese Journal of Scientific Instrument, 30(11), 2380-2383.
Li C.M., Chen T.P., 1996, Approximation problem in sigma-pi mural networks, Chenese Science Bulletin, 41(13), 1073-1074．
Lu W.X., Chu H.B., Wang X.H., Gong L., 2012 Application of Hopfield Neural Network Based on Factor Analysis to Water Quality Evaluation, Bulletin of Soil and Water Conservation, 32(1), 197-200.
Nie Y., Deng W., 2008, Hybrid learning algorithm for Pi-sigma neural network and analysis of its convergence, Computer Engineering and Applications, 44(35), 56-58, DOI: 10.3778/j.issn.1002-8331.2008.35.017
Pang Z.L., Chang H.J., Li Y.Y., Zhang N.Q., Du R.Q., 2008, Analytical Hierarchy Process (AHP) Evaluation of Water Quality in Danjiangkou Reservoir-source of the Middle Line Project to Transfer Water from South to North, China, Acta Ecologica Sinica, 28(4), 1810-1819.
Xu M.D., Lu J.J., Li C.S., 2010, Assessment on Tributary Water Quality of Fenhe River in Taiyuan City, China Water & Waste Water, 26(2), 105-108.
Yang H.F., Wei Y., 2007, Research on Water Quality Assessment Model Based on Fuzzy Neural Network, Journal of Yunnan Nationalities University (Natural Sciences Edition), 16(3), 255-258.
Yang J., Wu Y.M., 2004, ANN method for comprehensive evaluation of water quality in Hanjiang River, Engineering Journal of Wuhan University, 7(1), 51-59.
Zhou Y., Wang L.A., 2010, Application of Fuzzy Neural Network in Water Quality Assessment of Chongqing Drinking Water Sources, Environment and Ecology in the Three Gorges, 3(1), 33-35.
Zhou Z.X., 2007, The application of Fuzzy Neural Network based on T-S model in water qulity evaluation, Nanjing: Hohai University.
Zhu L., Chen W., 2011, Fuzzy Complex Index in Water Quality Assessment of Municipal ities, Journal of Wuhan University of Technology, 23(8), 61-65, DOI: 10.3321/j.issn:1671-4431.2001.08.019