



Water Environment Quality Analysis Based on Information Diffusion Theory and Fuzzy Neural Network

Hai-tao Chen, Ke-ke Xie and Wen-chuan Wang[†]

School of Water Resources, North China University of Water Resources and Electric Power, Zhengzhou City, Henan Province, 450045, PR China

[†]Corresponding author: Wen-chuan Wang; zzchenhaitao@126.com (Hai-tao Chen)

Nat. Env. & Poll. Tech.
Website: www.neptjournal.com

Received: 26-12-2019

Revised: 16-01-2020

Accepted: 30-03-2020

Key Words:

Information diffusion theory
Fuzzy neural network
Water quality analysis
Mathematical modelling

ABSTRACT

Reservoirs play a key role in many infrastructure functions for people like flood control, irrigation, and water supply. In this work, we focused on the water quality evaluation model for Shimen Reservoir. Based on the monthly changes of factors such as pH, nitrate, ammonia nitrogen (NH₃-N) and total nitrogen (TN) in 2013 and 2014, the information diffusion theory and fuzzy neural network technology were utilized to evaluate the water quality comprehensively. The probability distribution of these four factors in the reservoir was analysed and the water quality of the reservoir evaluated. The results show its reliability and these two methods can provide a basis for water quality control of Shimen Reservoir. Furthermore, the methods can be universally applied to the analysis and research of water quality in other regions..

INTRODUCTION

The Shimen Reservoir is a large-scale water conservancy project with flood control, irrigation, and urban water supply as its main function, taking into account the functions of power generation and fish farming. Therefore, it is important to ensure the basic stability of the water quality of the reservoir. At present, there are many methods for water quality evaluation, and more methods such as single factor evaluation method, comprehensive evaluation method and Nemerow index method are used. The single-factor index method is simple to calculate, but it can only reflect the change of a single factor, so many scholars combine single-factor evaluation with other methods to get better results. Song (2018) used a single factor evaluation method combined with an improved grey correlation to evaluate water quality. Wu (2019) evaluates the groundwater environment in Pinghai Bay, Putian City based on a single factor and fuzzy mathematics comprehensive evaluation method.

Information diffusion theory can dig out more information when the amount of data is insufficient and the samples are insufficient. It can be well combined with artificial neural network models and applied to many fields. Currently, mainly involved in meteorological disasters, signal processing, geological disasters, surveying and mapping, etc. Liu et al. (2019) applied entropy information diffusion theory to the risk assessment of agricultural drought and flood disasters in

the middle and lower reaches of the Yangtze River. Zhong et al. (2019) evaluation of flash flood risk was based on information diffusion method. Wang et al. (2016) used information diffusion theory to study flood and drought disaster risk characteristics in southern China. Lu et al. (2014) applied information diffusion technology in the probability analysis of grassland biological disaster risk. Besides, it is involved in some aspects such as thunderstorms, pests, typhoons and crop yields.

Regarding water environment issues, some scholars have applied the information diffusion theory. Li (2007) used the information diffusion technology to study the river health risk estimation model under the condition of incomplete information. The inspection of reservoir water quality is very complicated. Therefore, grasping the changes in the content of each element in the reservoir can better analyse the water quality and maintain the normal use of the reservoir. This paper analyses the risk probability of each element's pollution index through a combination of single factor index and information diffusion theory. It is of great significance to prevent the pollution index from exceeding the standard and the water quality to be stable.

The T-S fuzzy neural network model is an organic combination of fuzzy logic and neural networks. It inherits the advantages of both fuzzy logic and neural networks. It can represent highly nonlinear complex systems with

fewer fuzzy rules, which is very suitable for water quality evaluation. Mo et al. (2017) evaluated the water quality of the main rivers in Qinzhou based on the T-S fuzzy network model. Zhang et al. (2018) combined fuzzy neural network with LSSVM-MC to predict comprehensive water quality. Zhao (2018) uses neural networks to study early warning of aquaculture and transportation environment. This paper uses a fuzzy neural network to evaluate the water quality of the Shimen Reservoir.

MATERIALS AND METHODS

Single Factor Index

The single factor evaluation method is to determine the category of comprehensive water quality of the water body by the category of the single index with the worst water quality. This method is simple to calculate and can directly reflect the pollution of a single factor. With reference to the “Environmental Quality Standard for Surface Water” (GB3838-2002) Category V water standard(2002), the formula is:

$$P_i = C_i/S_i \quad \dots(1)$$

Where, C_i is the measured value of type i pollutant, and S_i is the evaluation standard of type i pollutant.

When $P_i \leq 1$, it means that the water body is not polluted; when $P_i > 1$, it means that the water body is polluted.

The standard index for PH value is:

$$PH_j \leq 7.0$$

$$SPH_j = |7.0 - PH_j| / (7.0 - PH_{sd}) \quad \dots(2)$$

$$PH_j > 7.0$$

$$SPH_j = |PH_j - 7.0| / (PH_{su} - 7.0) \quad \dots(3)$$

Where, PH_{sd} is the lower limit of the evaluation standard; PH_{su} is the upper limit of the evaluation standard; and PH_j is the measured value of PH.

According to the standard limit value of the supplementary project for centralized drinking water and surface water sources, the standard value of nitrate (in N) is 10 mg/L.

Information Diffusion Theory

Information diffusion is a kind of fuzzy mathematical processing method for set-valued samples. In order to make up

for the lack of information, it is considered to preferentially use the fuzzy information of samples, so as to set-valued samples.

The information diffusion method can turn a sample with observations into a fuzzy set, that is, a single-valued sample into a set-valued sample. The most commonly used model is the normal diffusion model.

Suppose that the set of actual observation samples is X , the sample series is $X = \{x_1, x_2, \dots, x_n\}$, and the discourse domain is $U = \{u_1, u_2, \dots, u_m\}$. A single-valued observation sample can carry the information spread to all points in U .

$$f_i(u_j) = \frac{1}{h\sqrt{2\pi}} \exp\left[-\frac{(x_i - u_j)^2}{2h^2}\right] \quad \dots(4)$$

Where, h is the information diffusion coefficient, which is determined according to the maximum value b , the minimum

$$\text{value } a, \text{ and the number of samples } n \text{ in the sample set. } \square = \begin{cases} 0.8146 \times (b - a), n = 5 \\ 0.5690 \times (b - a), n = 6 \\ 0.4560 \times (b - a), n = 7 \\ 0.3860 \times (b - a), n = 8 \\ 0.3362 \times (b - a), n = 9 \\ 0.2986 \times (b - a), n = 10 \\ 2.6851 \times \frac{(b-a)}{n-1}, n \geq 11 \end{cases} \quad \dots(5)$$

Set

$$C_i = \sum_{j=1}^m f_i(u_j), i = 1, 2, \dots, m \quad \dots(6)$$

Where, C_i is the normal diffusion information sum of the observation sample x_i .

The membership function of the corresponding fuzzy subset is:

$$g_i = \frac{f_i(u_j)}{C_i}, i = 1, 2, \dots, n; j = 1, 2, \dots, m \quad \dots(7)$$

Assume that after performing the above processing on all the samples, the number of samples with an observed value of u_j is inferred to be $q(u_j)$.

Let

$$q(u_j) = \sum_{i=1}^n g_i(u_j) \quad \dots(8)$$

$$Q = \sum_{j=1}^m q(u_j) \quad \dots(9)$$

Q is the sum of the number of sample points at each u_j point.

Table 1: Standard value of basic items for environmental quality standard for surface water (unit in mg/L).

Classification standard value/items	Class I	Class II	Class III	Class IV	Class V
pH (dimensionless)	6~9				
NH ₃ -N	0.15	0.5	1.0	1.5	2.0
TN (N for lakes and reservoirs)	0.2	0.5	1.0	1.5	2.0

$$p(u_j) = \frac{q(u_j)}{Q} \quad \dots(10)$$

Where $p(u_j)$ is the frequency value at which the sample point falls at u_j , which can be used as an estimate of the probability. Its transcendence probability is:

$$P(u \geq u_j) = \sum_{k=1}^m p(u_k) \quad \dots(11)$$

Where, $P(u \geq u_j)$ is the probability of exceeding .

T-S Fuzzy Neural Network

The T-S fuzzy system can be defined by the ‘‘if-then’’ rule form. In the case of the rule R_i , the fuzzy reasoning is as follows:

$$R_i: \text{if } x_1^i \text{ is } A_1^i, x_2^i \text{ is } A_2^i, \dots, x_k^i \text{ is } A_k^i, \text{ then } y_i = p_0^i + p_1^i x_1 + \dots + p_k^i x_k \quad \dots(12)$$

Where A_j^i is the fuzzy set of the fuzzy system; p_j^i is the fuzzy system parameter; y_i is the output obtained according to the fuzzy rules, the input part (the if part) is fuzzy, and the output part (the then part) is determined. The fuzzy inference output is a linear combination of the inputs.

T-S fuzzy system is a kind of fuzzy system with strong adaptive ability. The model can not only update automatically but also continuously modify the membership function of the fuzzy subset. It consists of an antecedent network and an after ware network.

Antecedent network

The first layer is the input layer. Assume that the input quantity $x = [x_1, x_2, \dots, x_k]^T$, and the number of nodes in this layer is k .

The second layer is the fuzzification layer. The membership degree of each input variable x_j is calculated according to the fuzzy rules.

$$u A_j^i = \exp \left(- \frac{(x_j - c_j^i)^2}{b_j^i} \right) \quad \dots(13)$$

Where, c_j^i and b_j^i are the center and width of the membership function respectively; k is the input parameter; n is the number of fuzzy subsets.

The third layer is the fuzzy rule layer. The fuzzy operator is used as the multiplication operator to perform fuzzy calculations for each membership degree.

$$\omega_i = u A_1^i(x_1) * u A_2^i(x_2) * \dots * u A_k^i(x_k) \quad \dots(14)$$

The fourth layer is used to normalize the applicability of each rule. The calculation expression is:

$$\bar{\omega}_i = \frac{\omega_i}{\sum_{i=1}^n \omega_i} \quad \dots(15)$$

After ware network

The first layer is the input layer, which is used to provide the constant term of the fuzzy rule follower.

The role of the second layer is to calculate each postscript of the rule.

$$y_i = p_0^i + p_1^i x_1 + \dots + p_k^i x_k \quad \dots(16)$$

The third layer calculates the output value of the fuzzy model based on the fuzzy calculation results to achieve clear calculations.

$$y_c = \sum_{i=1}^n (p_0^i + p_1^i x_1 + \dots + p_k^i x_k) \bar{\omega}_i \quad \dots(17)$$

Algorithm learning

The model’s error analysis, coefficient correction, and parameter correction methods are as follows:

Error calculation:

$$e = \frac{1}{2} (y_d - y_c) \quad \dots(18)$$

Where y_d is the expected output of the network; y_c is the actual output of the network; e is the error between the expected output and the actual output.

Coefficient correction:

$$p_j^i(k) = p_j^i(k-1) - \alpha \frac{\partial e}{\partial p_j} \quad \dots(19)$$

$$\frac{\partial e}{\partial p_j} = (y_d - y_c) \omega_i / \sum_{i=1}^n \omega_i x_j \quad \dots(20)$$

Where, p_j^i is the neural network coefficient, x_j is the learning efficiency of the network, x_j is the network input parameter, and ω_i is the membership product of input parameters.

Parameter correction:

$$c_j^i(k) = c_j^i(k-1) - \beta \frac{\partial e}{\partial c_j} \quad \dots(21)$$

$$b_j^i(k) = b_j^i(k-1) - \beta \frac{\partial e}{\partial b_j} \quad \dots(22)$$

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

Empirical formula for the number of nodes in the hidden layer:

$$M = \sqrt{I + O} + a \quad \dots(23)$$

Table 2: T-S fuzzy neural network construction parameter table.

I	O	M	Recommended M value
2	1~3	4~8	8
3	1~4	5~9	8
4	1~5	6~10	10
5	1~5	8~12	10
6	1~3	9~14	12

Where, M is the number of nodes in the hidden layer; I is the number of nodes in the input layer; O is the number of nodes in the output layer

Survey of Research Area

Shimen Reservoir is located at 122°45'00" east longitude and 40°22'00" north latitude. It is a large-scale water conservancy project with flood control, irrigation, and urban water supply as its main function while taking into account power generation and fish farming functions. The maximum dam height is 47 meters, the dam is 350 meters long, the annual runoff is 117 million cubic meters, and the total storage capacity is 102.2 million cubic meters. The reservoir provides an average of 25 million cubic meters of industrial and urban domestic water and 36 million cubic meters of agricultural water to Yingkou and Gaizhou. Reservoir operation has played a huge benefit and harm removal benefits

RESULTS AND ANALYSIS

In this paper, the monthly pH, nitrate, ammonia nitrogen, and total nitrogen contents of Shimen Reservoir in 2013 and 2014 were selected to ensure the stability of the water environment of the reservoir. The feasibility of the information diffusion theory and T-S neural network in the analysis of reservoir water quality was verified. The main research

Table 3: Probability of exceeding the pH single factor index.

Single factor index	Probability
0	1.000000
0.1	0.831928
0.2	0.502622
0.3	0.287776
0.4	0.215550
0.5	0.161780
0.6	0.065910
0.7	0.005141
0.8	0.000040
0.9	0.000000
1.0	0.000000

contents are as follows:

- (1) After determining the single factor index of each element, use information diffusion technology to process each single factor index to analyse the occurrence probability of pH, nitrate, ammonia nitrogen and total nitrogen in Shimen Reservoir.
- (2) The T-S fuzzy neural network was used to evaluate the water quality of the Shimen Reservoir in 2013 and 2014.

Water Quality Analysis of Shimen Reservoir Based on Information Diffusion Theory and Single Factor Index

The single factor index corresponding to the four factors was calculated from Equations 1-3. According to the information diffusion theory, each single factor index is used as the information diffusion sample for diffusion. X_1 , X_2 , X_3 and X_4 are information diffusion samples of pH, nitrate, ammonia nitrogen, and total nitrogen, respectively. With reference to the "Environmental Quality Standard for Surface Water" (GB3838-2002), the set [0,1] of one-dimensional space can be used to take points at equal intervals as the discourse of each factor. Set the discrete domain of PH as: $U_1 = \{0,0.1,0.2,\dots,1\}$; the discrete domain of nitrate as: $U_2 = \{0,0.1,0.2,\dots,1.4\}$; the discrete domain of $\text{NH}_3\text{-N}$ as: $U_3 = \{0,0.025,0.05,\dots,0.325\}$ and the discrete domain of TN as: $U_4 = \{1.0,1.6,1.7,\dots,2.5\}$. Then the control points of PH, nitrate, $\text{NH}_3\text{-N}$ and TN are $m_1=11$, $m_2=15$, $m_3=14$, $m_4=16$. The number of samples is $n = 24$. According to formula 4-11, the diffusion coefficients of each factor can be obtained as $h_1 = 0.064209$, $h_2 = 0.075883$, $h_3 = 0.019496$, $h_4 = 0.093395$. Surpassing probability of each factor is given in Tables 3-6, and the probability density and surpass probability of the four elements in Figs. 1-4.

It can be seen from the Figs. 1-4 that the probability of the corresponding single factor index of four elements in Shimen Reservoir appears. The maximum probability of pH is $\text{SPH}_j = 0.1$, which is 32.93%. Nitrate has the highest probability at 46.13%; $\text{NH}_3\text{-N}$ has the highest probability at 18.99% and TN has the highest probability at 23.78%.

It can be seen from Fig. 1 and Table 3 that the transcendental probability curve of pH is on the side less than 1. Therefore, the pH value of the Shimen Reservoir is relatively stable.

It can be known from Table 5 and Fig. 2 that the single factor exponential distribution of nitrate is distributed on the side of $P_i \leq 1$.

Table 4: Probability of exceeding the nitrate single factor index.

Single factor index	Probability
0	1.000000
0.1	0.999637
0.2	0.981154
0.3	0.783971
0.4	0.322698
0.5	0.074134
0.6	0.042579
0.7	0.041663
0.8	0.040967
0.9	0.031620
1.0	0.009881
1.1	0.000688
1.2	0.000009
1.3	0.000000
1.4	0.000000

Table 5: Probability of exceeding the NH₃-N single factor index.

Single factor index	Probability
0	1.000000
0.025	0.985991
0.05	0.911267
0.075	0.761674
0.1	0.571740
0.125	0.398739
0.15	0.258229
0.175	0.164238
0.2	0.078687
0.225	0.015765
0.25	0.000952
0.275	0.000014
0.3	0.000000
0.325	0.000000

That is, the nitrate content of the water body meets the standard.

From Table 6 and Fig. 3, when $P_i \leq 0.075$, the content of NH₃-N conforms to type I water, and the probability of occurrence is $P \geq 76.17\%$; when $0.075 < P_i \leq 0.25$, the NH₃-N content accords with the water of class II, and the probability of occurrence at this time is $76.17\% \leq P \leq 0.09\%$. That is, most of the ammonia nitrogen content in the water body meets the standards of Class I and II water.

It is clear from Table 7 and Fig. 3 that the single factor index of total nitrogen in the water body is greater than 1. Therefore, the biggest cause of water pollution may be excessive nitrogen content.

Water Quality Evaluation of Shimen Reservoir Based on T-S Fuzzy Neural Network

Table 6: Probability of exceeding the TN single factor index.

Single factor index	Probability
1.0	1.000000
1.1	1.000000
1.2	0.999999
1.3	0.999912
1.4	0.997685
1.5	0.978861
1.6	0.904393
1.7	0.813638
1.8	0.690715
1.9	0.498245
2.0	0.260474
2.1	0.103946
2.2	0.043997
2.3	0.013959
2.4	0.003121
2.5	0.000000

The basic steps and evaluation results of Shimen reservoir water quality evaluation based on T-S fuzzy neural network are as follows:

- (1) Selected network structure: The construction of the fuzzy neural network determines the number of input and output points of the fuzzy neural network according to the dimensions of the training samples. Selected four indicators of pH, nitrate, NH₃-N and TN, so the number of input nodes is 4. Water quality levels I-V are represented by numbers 1-5, and the number of output nodes is 1. It can be known from Table 2 that the number of nodes in the hidden layer is 10, so a 4-10-1 network structure is formed.
- (2) Generate training samples: In this paper, 400 sets of training samples are generated by interpolating water quality index standard data "Environmental Quality Standard for Surface Water" (GB3838-2002) with the evenly spaced distribution. Obtain water quality level indicators based on network predictions. When the predicted value is less than 1.5, the water quality level is Class I; when the predicted value is 1.5 to 2.5, the water quality level is Class II; when the predicted value is 2.5 to 3.5, the water quality level is Class III; when the predicted value is 3.5 to 4.5 When the predicted value is greater than 4.5, the water quality level is Category V.
- (3) Network training and testing: The training sample is used to train the fuzzy neural network 100 times. The training result is shown in Fig. 5. Fifty sets of data were drawn from a random sample to verify the accuracy of the model. The verification results are shown in Fig. 6.

It can be seen from Fig. 6 that the error between the actual water quality level and the model output water quality level is small, and it can play an accurate prediction role.

- (4) Water quality evaluation: Select the content of pH, nitrate, $\text{NH}_3\text{-N}$ and TN from month to month from 2013 to 2014. The trained T-S fuzzy neural network was used to evaluate the Shimen reservoir. The evaluation results are shown in Fig. 7. It can be seen from the figure that the water quality level of Shimen Reservoir is basically

maintained at Class III and IV, and the water environment is relatively stable.

CONCLUSIONS

- (1) 24 sample data are selected in this paper. The nitrate, $\text{NH}_3\text{-N}$, TN and pH in Shimen Reservoir were evaluated based on single factor index and information diffusion theory. The results show that the pH, nitrate and ammonia nitrogen in Shimen Reservoir are stable and meet

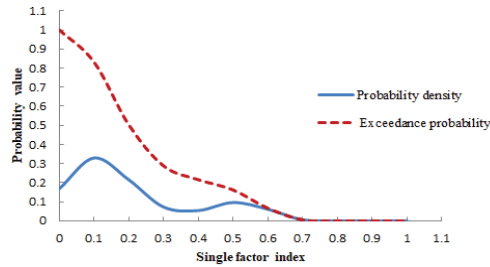


Fig.1: Probability distribution of the pH single factor index.

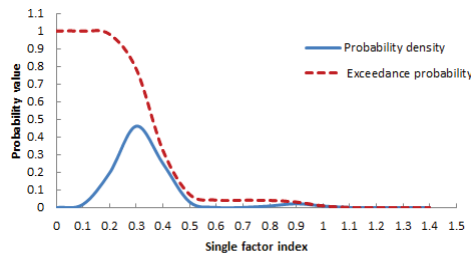


Fig. 2: Probability distribution of the nitrate single factor index.

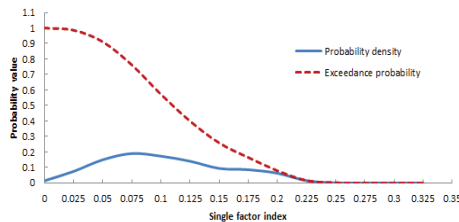


Fig.3: Probability distribution of the $\text{NH}_3\text{-N}$ single factor index.

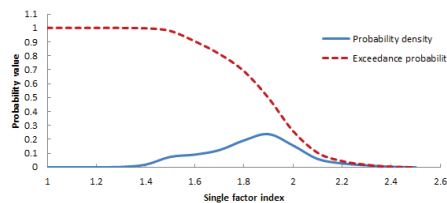


Fig.4: Probability distribution of the TN single factor index.

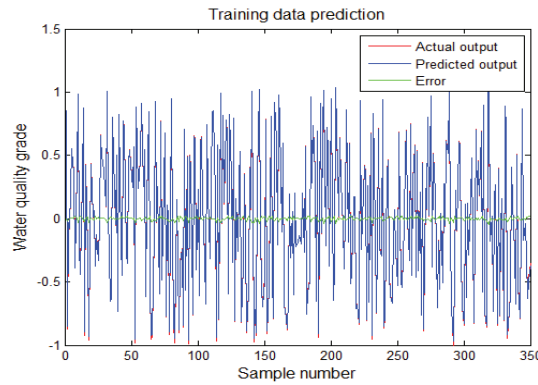


Fig. 5: Prediction simulation results of training data of fuzzy neural network model.

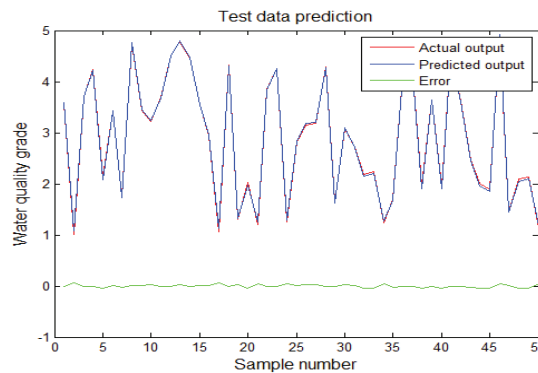


Fig. 6: Prediction simulation results of fuzzy neural network test data.

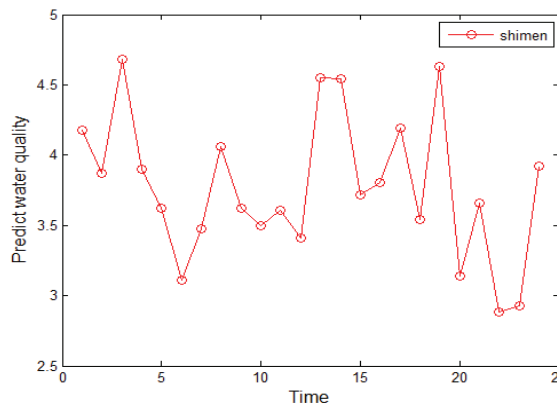


Fig. 7: Evaluation of water quality in Shimen Reservoir by the fuzzy neural network.

- national standards, but the total nitrogen contents are significantly higher than the standard.
- (2) Analysis of water quality of Shimen Reservoir from 2013 to 2014 based on T-S fuzzy neural network. The results show that the comprehensive water quality of Shimen Reservoir is good and stable.
 - (3) It is feasible to use information diffusion technology combined with single factor index to evaluate water quality. The analysis results are clear and have guiding significance for the water quality control of Shimen Reservoir. T-S fuzzy neural network fuzzy takes single factor prediction as input and comprehensive evaluation

as output. It has effectively evaluated the water environment quality of the Shimen Reservoir and obtained good evaluation results.

REFERENCES

- GB3832-2002 Environmental Quality Standard for Surface Water [S]. National Environmental Protection Agen. Standards Press of China, Beijing.
- Ji, Q. 2017. Evaluation of *Matsucoccus sinensis* Hazarb based on information diffusion theory. Hubei Agricultural Sciences. 56(02): 273-275.
- Jia, Z.H. and Li, X.G. 2017. Analysis on variation characteristics of Dagu river based on T-S fuzzy neural network. Environmental Science and Management. 42(01): 50-54.
- Jin, B., Du, Z.X. and Xue, S.C. 2018. Operation reliability assessment of straddle monorail vehicle based on T-S fuzzy neural network. Urban Mass Transit, 21(08): 77-81+85.
- Li, M. 2007. The Study of risk estimation model of river health by incomplete information. Xi'an University of Technology.
- Liu, Y., You, M., Zhu, J., Wang, F. and Ran, R. 2019. Integrated risk assessment for agricultural drought and flood disasters based on entropy information diffusion theory in the middle and lower reaches of the Yangtze River, China. International Journal of Disaster Risk Reduction, 38: 101194.
- Lu, H., Yang, L.Z. and Gao, J.M. 2014. The application of information diffusion technique in probabilistic analysis to grassland biological disasters risk. Ecological Modelling, 272: 264-270.
- Luo, B.L., Zhang, C. and Huang, W.H. 2009. Flooding Risk Assessment of Rice Production Based on Information Diffusion in Hunan Province. Chinese Journal of Agrometeorology.30(3): 458-462.
- Mo, C.X., Ruan, Y.L., Mo, G.Y. et al. 2017. Water quality assessment of main rivers in Qingzhou city based on T-S fuzzy neural network. Pearl River, 38(8): 80-83.
- Song, R.W. 2018. Application of single factor evaluation method and improved grey correlation degree in water quality evaluation. Shaanxi Water Resources. 3: 103-107.
- Wang, Y., Zhang, Q. and Han, L.Y. 2016. Risk characteristics of flood and drought disasters in southern China based on information diffusion theory. Journal of Arid Meteorology, 34(06): 919-926.
- Wu, M. 2019. Assessment of groundwater environment for Pinghai Bay, Putian City based on single factor evaluation method and fuzzy comprehensive evaluation method. Heilongjiang Environmental Journal, 43(03): 49-54.
- Zhang, Z.Y., Zhang, X.P., Zhang, L. et al. 2018. Integrated water quality prediction based on LSSVM-MC and fuzzy neural network. Jiangxi Science, 36(02): 331-334+339.
- Zhao, Y.T. 2018. Research on environment early-warning of aquaculture and transportation based on fuzzy neural network. Shanghai Ocean University.
- Zhong, M., Wang, J., Gao, L. et al. 2019. Fuzzy risk assessment of flash floods using a cloud-based information diffusion approach. Water Resources Management, 33(7): 2537-2553.