

Water environment risk assessment model based on information entropy and stochastic fuzzy theory

K Tong^{1,3} and C Yang²

¹Key Laboratory of Watershed Geographic Science, Nanjing Institute of Geography and Limnology, Chinese Academy of Sciences, Nanjing, 210008, China

²Nanjing Guohuan Science and Technology Co Ltd, Nanjing, 210042, China

E-mail: tongkun0502@hotmail.com

Abstract: At present, under the background of the river length system, the concern for water quality becomes more and more urgent. For there are many influencing factors of water environment change and it is difficult to describe. This paper proposes a fuzzy random risk rate calculation model for water environment water quality exceeding the standard aiming at the randomness and fuzziness of water environment system. At the same time, information entropy was introduced to solve the calculation. Use an example for calculations, the results show that the rate of the DO pollution is 4.3%, COD is 0.16%, and NH₄-N is 0.00% .The model calculation results are as good as observed.

1. Introduction

Nowadays, with the advancement of the “river length system”, China's water environment security has received increasing attention. The quality of the water environment is directly related to water security. With the development of industrialization and urbanization, the pressure on the water environment is increasing. Scholars at home and abroad research on water environment safety are mainly focused on water quality and aquatic ecology safety. For example, Ahmadi believes that water environment safety is water quality safety [1]. Larry believes that the shortage of water resources and the deterioration of the ecological environment are the reasons that restrict the social and economic development of Botswana in South Africa [2]. Then many methods that used to evaluate the water quality were developed such as artificial neural network [3], multivariate statistical techniques [4], grey evaluation method [5] and so on. Methods mentioned above are all deterministic methods [6]. But as we know, the factors that affect the risk of water environment system are numerous, complex and unpredictable. The uncertainties caused by such factors are hard to be quantified. At the meantime, it is necessary to attribute this type of uncertainty to randomness, using stochastic theory to calculate the risk rate that exceeds the standard. However, these uncertainties are not random in many aspects and sometimes they come from someone's subjective knowledge. Using traditional risk analysis theory to assess the risk of water environment systems always cannot fully reflect the actual situation. The actual water environment assessment usually is generally vague. The river is an open and large system. The diffusion, migration, degradation and conversion of pollutants into rivers are complicate. They are affected by temperature, flow conditions, river structure, microorganisms, degradation capacity, and location of pollution sources. The influence makes the concentration of pollutants in the water present a certain randomness and ambiguity, and it became necessary to introduce fuzzy mathematics method into the river water quality risk analysis [7,8]. This paper take the water quality system as a fuzzy



event, proposes a fuzzy evaluate model with water quality risk, and solves it through information entropy.

According to large number of studies, the state of water quality is random, the water quality state function Z can be described as follows [7].

$$Z = C_R - (C + C_I) \quad (1)$$

Among them, C_R is the source water quality control standard, C is the contribution of the upstream pollutants to the water source, and C_I is the initial value of the source water quality. $Z < 0$ means that the water quality is below the control index and the pollution accident occurs; when $Z = 0$, the water quality is in a critical state; when $Z > 0$, the water quality is lower than the control standard, and the water quality is good. Therefore, the risk of water pollution can be expressed as:

$$P = P[Z(X_1, X_2, X_3, \dots) < 0] \quad (2)$$

Among them, P is the fuzzy risk rate of water quality. X_1, X_2, X_3, \dots represent random variables that determine the state of the water quality. Based on the distribution of these random variables, the distribution of the state function $Z(X_1, X_2, X_3, \dots)$ of the water quality can be calculated. Combined with the fuzzy membership functions, it is possible to obtain the water pollution risk by integrating the continuous random variable probability calculation formula.

2. Methodology

2.1. Fuzzy model

The fuzzy model was used to calculate the risk of the water quality. When the risk of water quality exceeds the standard is calculated based on fuzzy-stochastic theory, the key issue is the probability distribution of membership function and water quality state random variables. There are many types of membership functions that describe fuzzy events, such as normal distribution, trapezoidal distribution, Cauchy distribution, and diamond distribution (triangular distribution). In this paper, trapezoidal distribution is used to establish a general fuzzy model. The risk rate membership function $\mu_c(x)$ can be described as follows [7]. Its graphical representation is shown in figure 1.

$$\mu_c(z) = \begin{cases} 1, & z < a \\ (b-z)(b-a)^{-1}, & a \leq z \leq b \\ 0, & z > b \end{cases} \quad (3)$$

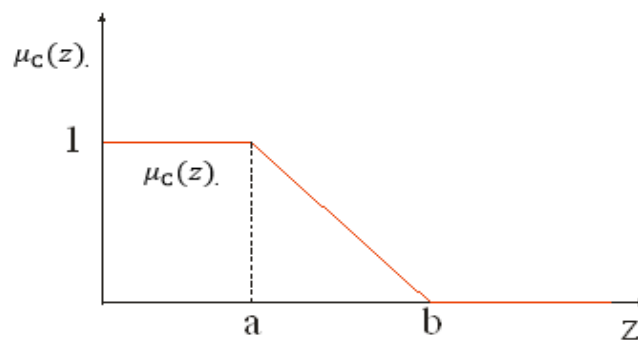


Figure 1. Risk rate membership function trapezoidal distribution map [9].

In the short term, the water quality status value Z presents a symmetry distribution due to various uncertainties [10]. However, in the long-term, due to the uncertainty such as river basin (region) socio-economic development, industrial and agricultural water discharges increasing, the initial value Z shows a definite growth trend, which can be determined based on long-term observational data. The study of the transfer between Long-term incremental deterministic trends and short symmetry random distribution is still rare. This paper studies the risk of water environment with short-term stochastic, so considering the water quality status value Z obeys the symmetrical distribution. In order to predict the long-term risk rate, deterministic trends can be used to predict the water quality, and then divide the time step into shorter time step. Symmetrically distribute were considered each shorter time step, then the water quality risk rate can be calculated from the probability distribution function that determined by the interval symmetry distribution prediction result. Its probability density is:

$$f(z) = \frac{1}{\sqrt{2\pi}\delta} e^{-\frac{(z-\mu)^2}{2\delta^2}} \quad (4)$$

Where μ is the expectation of a random variable and δ is the variance of the random variable.

Then formula (2) can be updated as follows [7]:

$$P_c = P(C) = \int_{-\infty}^{+\infty} \mu_c(z) f(z) dz = \begin{cases} \int_{-\infty}^0 \frac{1}{\sqrt{2\pi}\delta} \exp\left(-\frac{(z-\mu)^2}{2\delta^2}\right) dz, a > 0, b > 0 \\ \int_{-\infty}^a \frac{1}{\sqrt{2\pi}\delta} \exp\left(-\frac{(z-\mu)^2}{2\delta^2}\right) dz + \int_a^0 \frac{(b-z)}{(b-a)} \frac{1}{\sqrt{2\pi}\delta} \exp\left(-\frac{(z-\mu)^2}{2\delta^2}\right) dz, a < 0, b > 0 \\ \int_{-\infty}^a \frac{1}{\sqrt{2\pi}\delta} \exp\left(-\frac{(z-\mu)^2}{2\delta^2}\right) dz + \int_a^b \frac{(b-z)}{(b-a)} \frac{1}{\sqrt{2\pi}\delta} \exp\left(-\frac{(z-\mu)^2}{2\delta^2}\right) dz, a < 0, b < 0 \end{cases} \quad (5)$$

2.2. Information entropy

The information entropy was used to calculate the parameters a and b . When integrating the above functions, it can be found that the determination of feature parameters a, b is critical, and the feature function is related to the statistical characteristics of random variables, and $a=A(\mu, \delta)$, $b=B(\mu, \delta)$. The eigenvalues of membership function are generally determined by subjective empirical methods and amplification coefficient method. So sometimes the result is very subjective. Due to the limitations of various conditions, it is often difficult to obtain reasonable and credible result. In order to make the Eigen function determination more theoretical, the entropy function was introduced to determine the characteristic parameters.

The concept of entropy was derived from thermodynamics. In 1986, the German scientist Clausius proposed a new state function, entropy, which was used to quantify the second law of thermodynamics. The definition of entropy in thermodynamics is the logarithm of the number of possible states of the system, usually it was called thermal entropy. The thermal entropy can only increase but cannot be reduced at any known evolution process of orphaned physical system. Therefore, when the system is out of order, the entropy is maximized; when the system is in an ordered, the entropy is minimum. In 1948, Shannon, the founder of information theory, proposed the concept of information volume, information entropy can measure the overall uncertainty of information. Opposite to the thermal entropy, the information entropy can only be reduced but cannot be increased. So thermal entropy and information entropy are negative. And it was proved that any system if want to obtain information must increase the thermal entropy, both they are quantitatively related. In information theory, the source output is a random quantity, so its uncertainty can be measured by probability distribution. Information entropy is a physical quantity that represents the chaotic and random degree of the system,

that means, the greater the information entropy of the system, the higher the randomness of the system, the more random features of the water environment can be reflected in water quality risk studies. The water quality risk entropy can be expressed as follows [11].

$$H_c(\mathcal{C}) = - \int_{-\infty}^{+\infty} \mu_c(z) f(x) \log(\mu_c(x) f(x)) dx \quad (6)$$

Since the Eigen function is related to the statistical characteristics of random variables, let $a = k \times \mu$, $b = m \times \delta$, where k and m are random coefficients, $k, m = -1, -0.5, 0, 0.1, 0.15 \dots 1$, using the trial algorithm to calculate the entropy. In addition, the entropy function has maximum value. Therefore, the pollution probability of the water quality can be calculated when find out the characteristic parameters a and b corresponding to the maximum value.

3. Application

Taking a river in Zhejiang Province Taizhou City as an example, it is located in drinking water protection area, and the water quality is Class II according to the water quality standard. The function of river is mainly irrigation, considering flood control and water supply. The water quality of the river is good. Dissolved oxygen (mg/L), permanganate (mg/L), and ammonia nitrogen (mg/L) were be analyzed this time. According to Surface Water Environment Quality Standard (GB3838-2002), the standards for dissolved oxygen (mg/L), permanganate (mg/L), and ammonia nitrogen (mg/L) for Type II are 6, 4, and 0.5 mg/L. The fuzzy random theory was used to determine the risk rate of water quality that exceeding the standard. Considering the lack of long observational information, it is necessary to use the fuzzy theory to calculate the water environment parameters [12].

3.1. Example

The risks of water quality pollutants were analyzed by fuzzy stochastic risk method. The changes of water quality state variables (difference between water quality standard and observation data) are shown in figure 2.

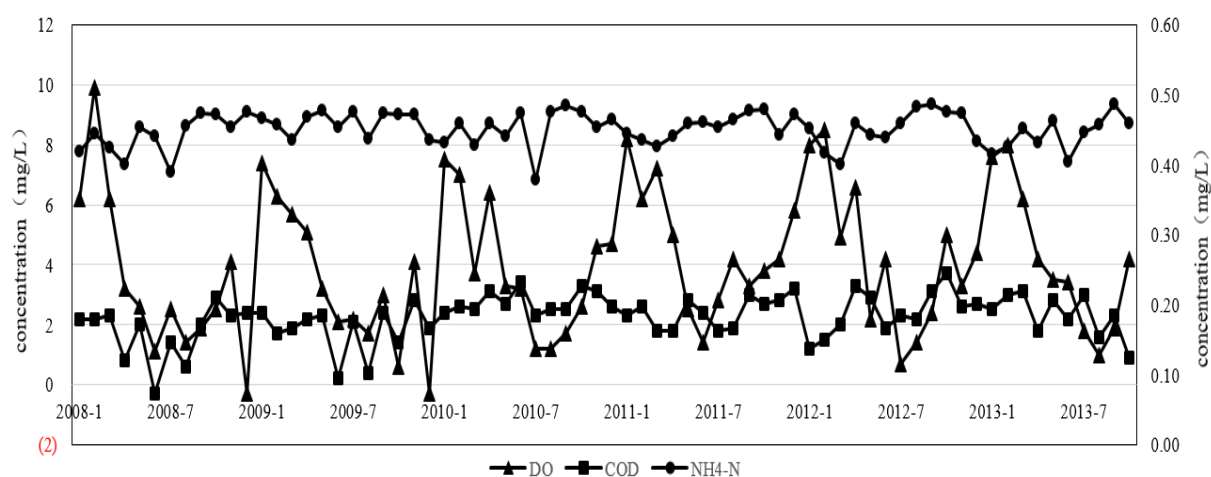


Figure 2. Water Quality Status Variables.

The results of the mean, variance, and distribution function of each variable are calculated as follows (table 1):

Table 1. Variable distribution feature value calculation.

Item	Mean μ	Variance σ	Distribution function
DO	3.948	2.310	$f(x) = \frac{1}{5.790} e^{\left(-\frac{(x-3.948)^2}{10.672}\right)}$
COD	2.242	0.760	$f(x) = \frac{1}{1.905} e^{\left(-\frac{(x-2.242)^2}{1.155}\right)}$
NH ₄ -N	0.450	0.020	$f(x) = \frac{1}{0.050} e^{\left(-\frac{(x-0.45)^2}{0.001}\right)}$

3.2. Results

Information entropy is a physical quantity that represents the chaotic and random degree of the system, that means, the greater the information entropy of the system, the higher the randomness of the system. When the distribution density function of Z is determined, there are two variables a and b need to be calculated. As the characteristic function is related to the statistical characteristics of the random variable, set $a = k \times \mu$, $b = m \times \delta$, where k and m are random coefficients, k, m = -1, -0.5, 0, 0.1, 0.15 ... 1, and the entropy is calculated using a trial algorithm. Taking dissolved oxygen as an example, as shown in table 2, when a=2.29 and b=2.31, the maximum value of H_C is 1.15, and the risk of water quality is 4.3%. Calculating DO, COD, and NH₄-N three indicators, the probability of decibel of pollution is 4.3%, 0.16%, and 0.3%, and the probability of pollution is small, which is basically consist with the observation data.

Table 2. Water pollution risks of different indicators.

Item	a	b	k	m	H _C	P of model	P of observations
DO	2.29	2.31	0.58	1	1.15	0.0430	0.0286
COD	0.74	0.76	0.33	1	1.24	0.0016	0.0142
NH ₄ -N	-0.45	-0.02	-1	-1	0.78	0.0030	0.0000

4. Conclusion

In this paper, a model for calculating the risk of excessive water quality based on the fuzzy theory is adopted. The randomness and fuzziness of water environment are comprehensively considered, and information entropy is used to help solving the model. Then choose one river in Zhenjiang Province, the results show that the rate of the DO pollution is 4.3%, COD is 0.16%, and NH₄-N is 0.00% .The model calculation results are as good as observed. Therefore the model can be tried to calculate other risk rate.

5. Discussion

In November 2016, the central government issued river length system for protecting the river and lake, requested that the river length system must be established in every country before the end of 2018, and clarified the high requirements for the water ecology of “hechang, Shuiqing, anlv and Jingmei (in Chinese)”. Therefore, it is very important to accurately calculate the water quality risk of rivers, reservoirs and lakes in the context of random and fuzzy water quality, then some accident can be avoided as soon as possible. This paper calculates the risk of water quality on one river by using fuzzy theory and information entropy, and effectively avoids the characteristics of strong subjective willingness in the previous calculation process. Although the result shows very well, this paper only choose three index to analyze, so next more index should be used to further inspected the model. In addition, we cannot consider all factors in the process of calculation, especially the unobservable

factors. This is what we will do next. How to think about the impact of unobserved factors on water quality in the model?

Acknowledgments

This work was supported by “One-Three-Five” Strategic Planning of Nanjing Institute of Geography and Limnology, Chinese Academy of Sciences (Grant No. NIGLAS2017GH07).

References

- [1] Ahmadi A 2012 Integrated planning of land use and water allocation on a watershed scale considering social and water quality issues *J. Water Resour. Plan. Manage.* **138** 671-81
- [2] Swatuk L A and Dianne R 2004 Integrating policy, destine-grating practice: water resources management in Botswana *Phys. Chem. Earth* **29** 1357-64
- [3] Yan H, Zou Z and Wang H 2010 Adaptive neuro fuzzy inference system for classification of water quality status *J. Environ. Sci.* **22** 1891-6
- [4] Huang F, Wang X Q and Lou L P 2010 Spatial variation and source apportionment of water pollution in Qiantang River (China) using statistical techniques *Water Res.* **44** 1562-72
- [5] Shi H and Deng Y 2012 A grey model for evaluation of information systems security *J. Comput.* **7** 284-291.
- [6] Li R, Zou Z and Yan A 2016 Water quality assessment in Qu River based on fuzzy water pollution index method *J. Environ. Sci.* **50** 87-92
- [7] Li L W and Shi Z 2007 Water risk rate calculation method based on fuzzy event probability theory *J. Hydraul. Eng.* **4** 417-21
- [8] Finney B A and Bowles D S 1982 Windham M P. Random differential e2 quations in river water quality modeling *Water Resour. Re.* **18** 122-34
- [9] Yang J 2014 Application of improved fuzzy comprehensive evaluation method in water quality evaluation (Chongqing: Chongqing University)
- [10] Padgett W J and Rao A 1979 Estimation of BOD and DO probability distribution *J. Environ. Eng. Div.* **105** 525-33
- [11] Li LW and Shi Z 2007 Water quality risk rate calculation method based on fuzzy event probability theory *J. Hydraul. Eng.* **38** 417-21
- [12] Kaufmann A 1975 *Introduction to the Fuzzy Subsets* (New York: New York Academic Press)