Analysis of Current Status of Water Pollution in the Yellow River Based on Principal Component

Chang Liu¹ *, Longjiang Fan², Shipeng Qin³

¹College of Information Science and Technology (College of Cyber Security Oxford Brookes College), Chengdu University of Technology, Chengdu, Sichuan, 610059
²The College of Environment and Civil Engineering, Chengdu University of Technology, Chengdu, Sichuan, 610059
³The College of Nuclear Technology and Automation Engineering, Chengdu University of Technology, Chengdu, Sichuan, 610059

*Corresponding author: chang_liu2020@cdut.edu.cn

Abstract. The Yellow River is the "mother river" of the Chinese nation. The Yellow River is vital to the economic, social development and ecological environment of the surrounding areas. In recent years, with economic development and the acceleration of urbanization, the water environment and water ecology in the Ningxia section of the Yellow River have been polluted to varying degrees. Therefore, it is urgent to investigate the pollution laws of the water environment of the Yellow River and carry out corresponding treatments. First, we extracted six major pollutant indicators from three aspects, and then we conducted a comprehensive evaluation of the water environment of a certain section of the Yellow River in the past 8 months: The entropy method is used to determine the weight of various pollutants' impact on the water environment, and the comprehensive evaluation index Q of water quality is obtained. Then the Q value of 8 months is compared with the actual data, and then the expected water environment (Type III water) Define the water environment comprehensive evaluation standard index Q_s, combine Q and Q_s to define the degree coefficient h reflecting the degree of water environment quality, and then divide the degree coefficient into intervals according to the principle of equal division, as the index of water environment classification. Finally calculated h=0.0088, according to the water quality grading index, it is concluded that the comprehensive water quality of a certain section of the Yellow River in the past 8 months has been classified as "poor", and it belongs to the very poor level in the "poor" category.

1. Introduction

The Yellow River is the mother river of the Chinese nation and an important water intake for industrial and agricultural production and domestic water in our country. The flowing water of the Ningxia section of the Yellow River is of great significance to the socio-economic development and the construction of ecological civilization in the Ningxia area of the Yellow River.

With the rapid economic growth in recent years and the acceleration of urbanization, the water environment and water ecology of the Ningxia section of the Yellow River have been polluted to varying...
degrees. Water pollution refers to the decline in the use value of water caused by various harmful substances. For example, excessive arsenic and excessive mercury will cause water pollution and reduce water quality. In recent years, the water pollution problem of the Yellow River has become increasingly serious. In July 2016, when General Secretary Xi Jinping inspected Ningxia, he clarified the positioning of the "ecological security barrier", gave Ningxia the mission of "strengthening the protection of the Yellow River", and required Ningxia to undertake the important mission of maintaining ecological security in the northwest and the whole country.

Doing a good job in the protection and management of the Yellow River is conducive to the simultaneous development of the construction of ecological civilization and social and economic development in Ningxia, and is also of important strategic significance to the protection of water resources throughout the country. Achievements in the contemporary era, benefits in the future.

This article analyzes the current state of water environmental pollution in a certain section of the Ningxia section of the Yellow River based on the current state of water environmental pollution in the river section or based on data recorded in relevant data.

2. Problem analysis
First, we consider extracting the main pollutant indicators from different aspects, and then conduct a comprehensive evaluation of the water environment of a certain section of the Yellow River in the past eight months. We intend to use the entropy method to determine the weight of various pollutants' impact on the water environment, obtain a comprehensive evaluation index of water quality, and then use the results of the comparison of the 8-month comprehensive evaluation index value of water quality with actual data, and then pass the expected water environment. The standard value defines the degree coefficient that reflects the degree of water environment quality, and then divides the degree coefficient into intervals as an indicator of water environment classification. Finally, the comprehensive water quality of a certain section of the Yellow River in the past 8 months is obtained.

3. Model establishment and solution

3.1. Selection of water quality indicators and data preprocessing

3.1.1. Selection of water quality indicators. The evaluation of water quality should consider selecting appropriate indicators from the aspects of pollution degree, self-purification capacity, and biological carrying capacity. Since total phosphorus, fluoride, chemical oxygen demand, and hypemanganate index are indicators of inorganic and organic pollution; ammonia nitrogen is the main oxygen consumption in water bodies. The higher the ammonia nitrogen content, the more likely it is to cause damage to fish and other aquatic organisms; The content of dissolved oxygen represents the self-purification capacity of the water body. In summary, we selected six indicators of total phosphorus, fluoride, chemical oxygen demand, permanganate index, ammonia nitrogen, and dissolved oxygen to study the current state of water pollution.

3.1.2. Data preprocessing. Since the data given in Annex 1 is missing in Section A in the 7th and August, and this data is necessary to evaluate the water environment pollution of the section, we use interpolation to fill in the missing data and then build the model.

3.2. Water environment pollution assessment model based on entropy method ([11])
Firstly, a decision matrix is formed with the monitoring values of 6 pollutants on 5 sections. Since each type of pollutant has a different degree of impact on the water environment, we need to assign weights to these 6 types of indicators. Here is an analogy to the concept of entropy in information theory.

In information theory, entropy is a measure of uncertainty. The greater the amount of information, the smaller the uncertainty and the smaller the entropy; the smaller the amount of information, the greater the uncertainty and the greater the entropy. According to the characteristics of entropy, we can
judge the randomness and disorder degree of an event by calculating the entropy value, or use the entropy value to judge the degree of dispersion of an index. The greater the degree of dispersion of the index, the better the index has for the comprehensive evaluation. The greater the impact. Therefore, the entropy method is used to assign weights to avoid deviations caused by human subjective factors, so that the water environmental quality of a certain section of the Yellow River can be more objectively evaluated from the existing data.

3.2.1. Entropy method weighting steps. Step1: Build the data decision matrix $x_{ij}$.

$$(x_{ij})_{m \times n} = \begin{pmatrix} X_{11} & \cdots & X_{1n} \\ \vdots & \ddots & \vdots \\ X_{m1} & \cdots & X_{mn} \end{pmatrix}$$

Where $X_{ij}$ is the value of the $j$th pollutant of the $i$-th section, $m$ represents the number of sections, $n$ represents the number of pollutant indicators, $m=5$, $n=6$.

Step2: Normalization of indicators

Since the dimensions of the various indicators are different, they need to be standardized before using them to calculate the comprehensive indicators. Moreover, because the meanings of the positive and negative indicators are different (the higher the positive indicator, the better, the lower the negative indicator, the better), we use different algorithms for data standardization for high and low indicators. The specific method is as follows:

For positive indicators:

$$x'_{ij} = \frac{x_{ij} - \min x_{ij}}{\max x_{ij} - \min x_{ij}} \tag{2}$$

For negative indicators:

$$x'_{ij} = \frac{\max x_{ij} - x_{ij}}{\max x_{ij} - \min x_{ij}} \tag{3}$$

Step3: Calculate the proportion of the $i$-th type cross section in the index under the $j$-th type index

$$P_{ij} = \frac{x_{ij}}{\sum_{i=1}^{n} x_{ij}} \quad (i = 1 \ldots m ; j = 1 \ldots n) \tag{4}$$

Step4: Calculate the entropy value of the $j$-th index

$$e_{j} = -k \sum_{i=1}^{m} P_{ij} \ln P_{ij} \tag{5}$$

Step5: Calculate information entropy redundancy

$$d_{j} = 1 - e_{j} \tag{6}$$

Step6: Calculate the weight of each indicator

$$w_{j} = \frac{d_{j}}{\sum_{j=1}^{n} d_{j}} \tag{7}$$

So far, we have completed the weighting of these 6 types of indicators.
3.2.2. Establishment of Water Environment Pollution Evaluation Model Based on Entropy Method. The standard value $I_3$ (target water quality value) is defined as follows:

Because Class I, II, and III water meet the drinking water standards, and the target water quality required in Annex I is based on the standard value of Class III water as the measurement limit. Therefore, we define the standard deviation $d_{ij}$:

\begin{align*}
    d_{ij} &= x_{ij} - I_3 \\
    d_{ij} &= I_3 - x_{ij}
\end{align*}

(8)

In the above formula: the first formula represents the benefit index (dissolved oxygen, etc.). The larger the value of such index, the better the water quality.

The second formula represents cost-based indicators (permanganate index, etc.). The smaller the value of such indicators, the better the water quality.

It can be seen from the above formula that the larger the $d_{ij}$, the smaller the pollution degree of the j-th pollutant of the i-th section to the water environment. The following is the sum of the standard deviations of each type of pollutant:

\[ D_j = \sum_{i=1}^{5} d_{ij} \]

(9)

In the above formula, $D_j$ represents the sum of the standard deviations of the jth pollutant in each section, and the size of $D_j$ can reflect the influence of the jth pollutant on the water environment. The following defines the comprehensive water environment assessment index $Q$:

\[ Q = \sum_{j=1}^{6} D_j w_j \]

(10)

The meaning of $Q$ in the above formula is an evaluation index obtained by weighting the sum of the standard deviations of various pollutants. The value of $Q$ can reflect the degree of monthly water environmental pollution in a certain section of the Yellow River. The larger the $Q$, the higher the water environmental quality. The better, the worse. The following defines the comprehensive evaluation standard index $Q_s$ of the water environment:

\begin{align*}
    Q_s &= \left( I_{III} - \frac{I_I + I_{II} + 3I_{III}}{3} \right) * n * \bar{w}_j \\
    Q_s &= \left( \frac{I_I + I_{II} + 3I_{III}}{3} - I_{III} \right) * n * \bar{w}_j
\end{align*}

(11)

In the above formula: The first formula represents a cost-based index (permanganate index, etc.). The smaller the value of this type of index, the better the water quality. The second formula represents the benefit index (dissolved oxygen, etc.). The larger the value of such index, the better the water quality.

Among them, $I_j$, $II_j$, $III_j$ represent the national standard value of the jth pollutant belonging to the I, II, and III water, $n$ is the number of sections of a certain section of the Yellow River, $w_{-j}$ refers to each type of pollution the average value of the weight of each month in the past 8 months. Therefore, after calculation, $Q_s$ is a fixed value, which serves as a standard to measure the quality of the water environment.

Analysis of the rationality of the definition of $Q_s$: The target water quality expected by the subject is the III water quality, so we find a balance point between the I, II, and III water standard values. This balance point uses the average of the three types of water standard values. Instead, the corresponding algebraic calculation of the target water quality and the water quality at the equilibrium point can finally get the comprehensive water environment evaluation standard index $Q_s$. By comparing $Q_s$ and $Q$, the quality of the water environment of a certain section of the Yellow River can be roughly judged. If $Q$ is
much greater than \( Q_s \), it is indicated the quality of the water environment is very good, and vice versa. Therefore, the definition is reasonable.

In order to comprehensively evaluate the water environmental quality of a certain section of the Yellow River, we need to use \( Q_s \) and \( Q \) to define a degree coefficient \( h \) that reflects the degree of water quality:

\[
h = \frac{\sum_{i=1}^{8}(Q_i - Q_s)}{\sum_{i=1}^{8}|Q_i - Q_s|} \tag{12}
\]

In the above formula, \( Q_i \) represents the comprehensive evaluation index of the water environment every month, and the ratio \( h \) represents the degree of water environment pollution in a certain section of the Yellow River in the past 8 months.

The meaning of the numerator: The numerator refers to the sum of the difference between the comprehensive water environment evaluation index \( Q \) and the comprehensive water environment evaluation standard index \( Q_s \) for each month. The difference is positive or negative, indicating that the water environment is better or worse than the standard value. If the final summation result is positive, it means that the water environment quality of a certain section of the Yellow River has been good during this period. If the summation result is negative, it means that the water environment quality of a certain section of the Yellow River has been poor during this period.

The meaning of the denominator: The denominator refers to the sum of the absolute value of the difference between the comprehensive water environment evaluation index \( Q \) and the comprehensive water environment evaluation standard index \( Q_s \) for each month, indicating the total deviation of the water environment quality and the standard value each month.

The meaning of the ratio \( h \): According to the above analysis, it can be seen that the value range of \( h \) is between \([-1, 1]\), with 0 as the limit. When the value of the ratio \( h \) is less than 0, it indicates that a certain section of the Yellow River has been polluted during this period of time. Seriously, it did not meet the target water quality standard. The ratio \( h \) greater than 0 indicates that the water quality of a certain section of the Yellow River has reached the basic standard of the target water quality during this period, but the water quality is not clear. Therefore, we uphold the principle of uniform distribution to define the following water environment water quality standards:

\[
\begin{align*}
h & \in [-1, 0] \\
0 & \leq h < \frac{1}{4} \\
\frac{1}{4} & \leq h < \frac{1}{2} \\
\frac{1}{2} & \leq h < \frac{3}{4} \\
\frac{3}{4} & \leq h \leq 1
\end{align*}
\tag{13}
\]

Analysis of the rationality of defining water environment water quality standards: For \( h \in [-1, 0] \), we define water quality as "very poor". From the above analysis, we know that when the value of the ratio \( h \) is less than 0, it indicates that a certain river in the Yellow River has been the water environment in Duan is seriously polluted and does not meet the target water quality standard, so the definition is reasonable. For the range of \( h \in (0,1] \), we use the equal division method to divide it into four levels: "excellent", "good", "medium" and "poor". The distance between the levels is the same, which is more reasonable.

3.3. Solving the model

By formula (1-3), using MATLAB to solve, the \( Q \) value of each month is obtained as shown in Table 1, and the change curve of \( Q \) value over the past 8 months is obtained.
Table 1. Q value per month.

| Month | 1    | 2     | 3     | 4     | 5     | 6     | 7     | 8     |
|-------|------|-------|-------|-------|-------|-------|-------|-------|
| Q     | 7.73 | 10.63 | 8.77  | 4.21  | -0.74 | 3.14  | 2.80  | 7.43  |

Through formula (1-5), use MATLAB to solve the problem and get h=0.0088

3.4. Problem one result analysis

According to the standard of formula (1-6), we can know that the h ∈ (0,1/4) we solved, so we can conclude that the water environment quality of a certain section of the Yellow River in the past 8 months is "poor", because h=0.0088 Close to 0, it can even be said that the water environment of a certain section of the Yellow River has been seriously polluted during this period. If no corresponding measures are taken, the water quality of a certain section of the Yellow River will suffer more serious pollution, and the survival of aquatic organisms will face serious threats, or even Lead to biological extinction.

The analysis in Table 1 shows that in the past eight months, only the first, second, third, and August water environment quality has reached the standard, and the water environmental quality in the remaining months has not reached the standard. Especially in the fifth month, the water environment quality was far below the expected target water quality. Starting from about the second month, the water environment quality decreased month by month. From the fifth month, the water environment quality began to improve slightly. Therefore, it can be inferred that in the fifth month, the government department has taken care of the water environment of a certain section of the Yellow River and gradually gained Effectiveness.

4. Evaluation of the model

4.1. Advantages of the model

(1): The entropy method is used to calculate the weight of pollutants, and the weight assignment is objective.

(2): Defined the comprehensive evaluation index Q for evaluating the water environment, and defined the comprehensive evaluation standard index Q_s of the water environment, which makes the evaluation of the water environment quality more reasonable.

(3): We have studied the changes of water environment pollution in a certain section of the Yellow River from two perspectives, making the evaluation more objective and comprehensive.

(4): We use principal component analysis to extract the main components and give each principal component a new meaning.

4.2. Disadvantages of the model

(1): We ignore the changes in the water environment pollution caused by human factors.

(2): We have neglected the possibility of catalysis between pollutant elements, which will affect the degradation rate of pollutants and affect the quality of water environment.

References

[1] Gang Shiting, Jia Tao, Deng Yinger, Liu Weipo, Peng Xin. Application of set pair analysis model based on entropy weight method in groundwater quality evaluation of Tomatong Basin[J]. Journal of Yangtze River Scientific Research Institute, 2018, 35 (09):23-27.

[2] Li Zhiping, He Yujiang, Zhu Zhongdao. Application of Entropy Method in the Evaluation of Rural Safe Drinking Water Quality[J]. People's Yellow River, 2007(05): 35-36.

[3] Chen Zhaoming, Wang Wei, Zhao Ying, Xu Zeyu. Improved principal component analysis and multiple regression integration of Hanfeng Lake water quality assessment and prediction[J/OL]. Environmental monitoring management and technology: 1-8 [2020-08-18].https://doi.org/10.19501/j.cnki.1006-2009.20200713.001.

[4] Liang Yukai, Lin Guangyu. Groundwater quality evaluation of Tianjin Binhai New Area based on principal component analysis and water quality labeling index[J]. Groundwater, 2020,
[5] Huang Faming, Li Youjun, Liu Dalin, Guo Jianfei, Chen Jiawu, Wang Jianwei. Lake water quality evaluation based on improved TOPSIS method[J]. Hydropower Energy Science, 2020, 38(04): 51-54.

[6] Gu Xuezhi, Xiao Changlai, Jiang Xingming, Liang Xiujuan, Wang Yuehang, Lou Yang. Application of two improved TOPSIS models in groundwater quality evaluation[J]. Water-saving irrigation, 2016(05): 60-64.

[7] Men Baohui, Liang Chuan. Water quality evaluation attribute recognition model based on the weight of coefficient of variation [J]. Journal of Harbin Institute of Technology, 2005(10): 69-71.

[8] Jiang Qiyuan. Mathematical Model (Third Edition) [M]. Beijing: Higher Education Press, 2003.

[9] Si Shoukui, Sun Xijing. Mathematical modeling algorithm and application [M]. Beijing: National Defense Industry Press, 2016.