

Water quality assessment of the New Chaobai River based on PCA-AHP reduced-dimensional combined subjective two-layer empowerment evaluation model

SUN Yuan-tao¹, DENG Ken², and WANG Hong-yang³

¹*School of Chemical Engineering, Tianjin University - 309207005*

²*School of Chemical Engineering, Tianjin University - 3019207073*

³*College of Intelligence and Computing, Tianjin University - 3019244233*

Abstract

The comprehensive evaluation of river water quality is important for regional environmental protection and ecological balance of water resources, as well as river ecological planning and management. Most of the river water quality evaluation methods are mainly based on the evaluation objectives and water quality standards to achieve the analysis of a single river, it is difficult to achieve the quantification of river water quality pollution indicators within the watershed and the comparison of different sampling sites and even different river water quality status. With the continuous development of the testing technology and analytical instruments, there is an increasing variety of water quality indicators. A rapid, accurate and comprehensive water quality evaluation model is of great practical importance for the quantitative evaluation of the overall water quality of the river basin. In order to quantitatively evaluate the existing sampling points of the New Chaobai River and form an objective and comprehensive evaluation of all indicators to amend the existing evaluation methods, we combined the principal component analysis (PCA) and hierarchical analysis (AHP), to build a PCA-AHP double-weighted water quality comprehensive quantitative evaluation model. PCA can quickly filter out water quality parameters, reduce dimensions and calculate the objective weights; AHP can perform category analysis on the water quality parameters after dimensionality reduction and calculate the subjective weights. We again applied hierarchical analysis to the subjective and objective weights, and combined with the analysis and observation of the measured data and the indicators and sampling points, we obtained a subjective weight influenced by these factors, and then assigned weights to the two weights again.

The main research work and results are as follows:

- 1) We used grey predictions to fit the missing data, analyzed the data in both temporal and spatial dimensions, and gained a clear initial idea of the data characteristics of the current data.
- 2) In order to comprehensively, accurately conduct quantitative evaluation of water quality indicators data from several sampling points of the New Chaobai River, we propose to build a PCA-AHP two-layer weighting water quality comprehensive evaluation model, use SPSS software for PCA analysis, and program the two-layer AHP;

- 3) Through the model we build, we modify the existing evaluation methods and analyze the strength and weakness of the two models in combination with the data, so that they can reach a complementary state. We also analyze the sampling points, in terms of model evaluation methods, sampling points and the actual situation reflected by the data. The sampling points are also analysed, and suggestions for improvement are made at three levels: the model evaluation method, the choice of sampling points and the development of the actual situation reflected in the data.

[Key Words]: Water quality evaluation, PCA-AHP, New ChaoBai River, Policy Suggestions, Dimensionality reduction

Contents

ABSTRACT	I
1 Introduction	1
1.1 Background	1
1.2 Our Works	2
2 General Assumption and Symbol Explanation	2
2.1 General Assumption	2
2.2 Symbol Explanation	3
3 Main Factors and Data Selection	4
3.1 Main Factors	4
3.1.1 PH	4
3.1.2 Dissolved Oxygen	4
3.1.3 Ammonia Nitrogen	4
3.1.4 Total Phosphorus	4
3.1.5 Total Nitrogen	5
3.1.6 Chemical Oxygen Demand	5
3.1.7 BOD	5
3.1.8 Potassium Permanganate	5
3.1.9 Fluorine	5
3.2 Data Selection	6
3.3 Data Pretreatment	6
3.3.1 Time Dimension	6
3.3.2 Spatial Dimension	11
4 APPROACH	12
4.1 Principal Component Analysis	13
4.1.1 Basic principles	13
4.1.2 Calculation steps	13
4.2 AHP	16
4.2.1 Build a hierarchical model	17
4.3 PCA-AHP two-layer modeling	18
4.3.1 PCA-AHP comprehensive empowerment	18
5 Analysis of sampling points	21
5.1 Introduction to sampling sites	21
5.2 Sampling point reasonableness analysis and coordinate supplement	21
5.2.1 Analysis of the premise	21
5.2.2 Rational analysis	21
6 TREATMENT PLAN AND ADVICE	22
6.1 Suggestions for improvements in evaluation methods and measures for improvement	22
6.2 Recommendations for sampling point selection	22
6.3 Recommendations for specific governance directions	23
REFERENCE	24

1 Introduction

1.1 Background

The New Chaobai River is a 458-kilometer-long river that passes through Beijing, Hebei, and Tianjin. It belongs to the Haihe River System (called the New Chaobai River in Tianjin) and is an important part in the Haihe River System Management Project.

Water quality evaluation is a key work of water environmental management. By making a reasonable assessment of the water quality, a scientific water environment remediation plan can be formulated, which can effectively promote the improvement of the water environment and the sustainable utilization of water resources.

序号	标准值 项目	分类	I 类	II 类	III 类	IV 类	V 类
1	水温 (°C)	人为造成的环境水温变化应限制在： 周平均最大温升≤1 周平均最大温降≤2					
2	pH 值 (无量纲)	6~9					
3	溶解氧	≥ 饱和率 90% (或 7.5)	6	5	3	2	
4	高锰酸盐指数	≤	2	4	6	10	15
5	化学需氧量 (COD)	≤	15	15	20	30	40
6	五日生化需氧量 (BOD ₅)	≤	3	3	4	6	10
7	氨氮 (NH ₃ -N)	≤	0.15	0.5	1.0	1.5	2.0
8	总磷 (以 P 计)	≤ (湖、库 0.01)	0.02 (湖、库 0.025)	0.1 (湖、库 0.05)	0.2 (湖、库 0.1)	0.3 (湖、库 0.2)	0.4
9	总氮 (湖、库, 以 N 计)	≤	0.2	0.5	1.0	1.5	2.0
10	铜	≤	0.01	1.0	1.0	1.0	1.0
11	锌	≤	0.05	1.0	1.0	2.0	2.0
12	氟化物 (以 F ⁻ 计)	≤	1.0	1.0	1.0	1.5	1.5
13	硒	≤	0.01	0.01	0.01	0.02	0.02
14	砷	≤	0.05	0.05	0.05	0.1	0.1
15	汞	≤	0.0005	0.0005	0.0001	0.001	0.001
16	镉	≤	0.001	0.005	0.005	0.005	0.01
17	铬 (六价)	≤	0.01	0.05	0.05	0.05	0.1
18	铅	≤	0.01	0.01	0.05	0.05	0.1
19	氰化物	≤	0.005	0.05	0.2	0.2	0.2
20	挥发酚	≤	0.002	0.002	0.005	0.01	0.1
21	石油类	≤	0.05	0.05	0.05	0.5	1.0
22	阴离子表面活性剂	≤	0.2	0.2	0.2	0.3	0.3
23	硫化物	≤	0.05	0.1	0.2	0.5	1.0
24	粪大肠菌群 (个/L)	≤	200	2000	10000	20000	40000

Figure 1. Surface water environmental quality standard item limit value

The picture is commonly used in a water quality evaluation. The water quality grade is divided by the degree of some indicators. This method has the advantage of being easy to classify, but the contradiction caused by the mismatch of the corresponding categories of different indicator values makes it difficult to classify the water. It is still tough to solve the problem. For example, the water quality with a permanganate index of 2 and a chemical oxygen demand of 20 is too one-sided in its decisive indicator. Such a water quality evaluation method is suitable for the evaluation of a single water quality and regulation

and requirements for river water quality water, but hard to compare different water quality in the same category. In addition, much present analyses only reflect the characteristics of the data , but do not take into account the actual situation, such as the specificity and importance of certain indicators. As a result, it needs to be improved by new empowerment method.

1.2 Our Works

Currently, single-factor evaluation method, comprehensive pollution index method, Nemerow pollution index method, neural network method, etc, are commonly used for water quality evaluation. These methods have a positive effect on general evaluation, but they also have few shortcomings. For example, one of the disadvantage needed to be improved is the determination of the weight value of each index in the calculation process of different water quality evaluation methods. Common index weight assignment methods include analytic hierarchy process, expert scoring method and other subjective weighting methods, as well as entropy weight method, coefficient of variation method and other objective weighting methods, as well as their combined weighting methods. These weighting methods and the evaluation methods have different disadvantages, so we attempt to filter the data through a **new modified weighting method** to establish a new water quality evaluation model.

We have completed the following works:

- We selected **9 test indexes** of Huangbai bridge, Yujialing bridge, Ningchegu floodgate and Xiaogaotuo in even months in the data sheet as our analysis object, and analyze these data in dimensions of time and space.
- We use **principal component analysis (PCA)** to screen indicators, and combine with **analytic hierarchy process (AHP)** to carry out a **comprehensive weight** assignment to get our evaluation model, and evaluate the existing data, compare the results with the water quality evaluation method mentioned above to analyze the rationality of the model, and modify it, and objectively analyze the advantages and disadvantages of the two methods.
- We analyze the distribution of sampling points, and give suggestions to supple sampling points through the research on the field situation.
- Combined with the model and analysis results, we give scientific suggestions and improvement measures to the water management department to improve the river water quality.

2 General Assumption and Symbol Explanation

2.1 General Assumption

In order to ensure that our model can accurately evaluate the water quality, analyze the existing data and modify the existing water quality evaluation methods, we make the following assumptions:

- The time point we choose, the even-numbered months from 2014 to 2018, is the time variable obtained after ignoring the date, and the data of Huangbai Bridge in March 31st, 2014 is assumed to be the data of April, 2014. We also assume that the time variable we selected can exactly reflect

the average value of all indicators in the last two months.

- To a certain extent, the values obtained by weighting **all indicators** can accurately **evaluate** the water quality in the model system we have established
- Our prediction of future data is based on the assumption that the current external environment remains unchanged, that is, external environmental factors such as environmental governance and requirements for factory emission reductions remain on the last day of the data we hold in 2018 and there is no other change during our forecast time.
- When we use the analytic hierarchy process, our judgment on the importance and impact of the index is **basically rational and logical**.
- When we encounter an inaccurate value less than a certain number, we take it upward as that number. For example, when an index number is < 0.2 , we take this data as 0.2 for drawing and calculation, and assume that the value is accurate and has no impact on the correctness of the result.

2.2 Symbol Explanation

Symbol	Basic description
x_{ij}	Sampling point data, where i is the sampling point, j is the index
L_{ij}	Principal component feature vector
α	Percentage of variance of principal components
W_{ij}	Weights obtained by principal component analysis
Q_{ij}	Weights obtained from hierarchical analysis
β	Secondary weights
F	Corrected combined weights
WQ	Water quality index, where $WQ = F \times x_j$

3 Main Factors and Data Selection

3.1 Main Factors

3.1.1 PH

pH, called the hydrogen ion concentration index, acid-base value, is a scale of the hydrogen ion activity in a solution, that is, a measure of the acid-base degree of the solution generally. PH value equal to 7 means that the water environment is neutral. PH value greater than 7 means that the water is alkaline, and the larger the value, the stronger the alkaline water. On the contrary, the PH value is less than 7, it means that the water is acidic, and the smaller the value, the more acidic the water is.

The pH of the water environment changes with the seasons, day and night. In the aquarium feeding process, the long-term change of pH value is usually lower. This is mainly due to the decomposition of the organic matter in the water and the lack of oxygen. When the dissolved oxygen in the water environment is insufficient, the oxidation process is inhibited, and residual bait, organic fertilizers, and fish feces will gradually accumulate. If these things are not sufficiently decomposed, they will produce and accumulate various organic acids, resulting in a gradual decrease in pH and a change in the pH of the water.

3.1.2 Dissolved Oxygen

Dissolved in the air in the molecular state of oxygen in water is called dissolved oxygen, the content of dissolved oxygen in water and the partial pressure of oxygen in the air, the temperature of the water are closely related. In the natural situation, the oxygen content in the air does not change much, so the water temperature is the main factor, the lower the water temperature, the higher the content of dissolved oxygen in water. The molecular state of oxygen dissolved in water is called dissolved oxygen, usually recorded as DO, expressed in milligrams of oxygen per liter of water. The amount of dissolved oxygen in the water is a measure of the self-purification ability of the water body.

3.1.3 Ammonia Nitrogen

The presence of chemical nitrogen in the form of free ammonia (NH_3) and ammonium ions (NH_4^+) is called ammonia nitrogen. Ammonia nitrogen is a nutrient in the water body, can lead to water eutrophication phenomenon, is the main oxygen-consuming pollutants in the water body, toxic to fish and some aquatic organisms.

3.1.4 Total Phosphorus

Total phosphorus is the result of water samples after the elimination of various forms of phosphorus into orthophosphate measured in milligrams of phosphorus per liter of water sample. Phosphorus in water can be in the form of elemental phosphorus, orthophosphate, condensed phosphate, pyrophosphate, metaphosphate and organic group-bound phosphate. Its main sources are domestic sewage, fertilizers, organophosphorus pesticides and phosphate detergents used in recent detergents. Phosphate can interfere with the coagulation process in water plants. Phosphorus in water bodies is a key element needed for algae growth, and excess phosphorus is the main cause of water fouling and odor, eutrophication in lakes and red tides in bays. China's surface water environmental quality standards (GB3838-2002) provide the following allowable values for total phosphorus.

Table 1.Total phosphorus tolerance value

	Class I	Class II	Class III	Class IV	Class V
TP (mg/L)	≤ 0.02	≤ 0.1	≤ 0.2	≤ 0.3	≤ 0.4
	≤ 0.01 (lakes)	≤ 0.025 (lakes)	≤ 0.050 (lakes)	≤ 0.1 (lakes)	≤ 0.2 (lakes)

3.1.5 Total Nitrogen

Total Nitrogen, abbreviated as TN, the total nitrogen content in water is one of the important indicators of water quality. Total nitrogen is defined as the total amount of inorganic and organic nitrogen in various forms in water. It includes inorganic nitrogen such as NO_3^- , NO_2^- and NH_4^+ and organic nitrogen such as proteins, amino acids and organic amines, and is calculated as milligrams of nitrogen per liter of water. It is often used to indicate the extent to which a water body is contaminated with nutrients.

3.1.6 Chemical Oxygen Demand

Chemical oxygen demand is a chemical measurement of the amount of reducing substances that need to be oxidized in a water sample. The oxygen equivalent of substances (generally organic matter) that can be oxidized by strong oxidizing agents in wastewater, wastewater treatment plant effluent and contaminated water. It is an important and quickly measurable organic pollution parameter in the study of river pollution and the nature of industrial wastewater, as well as in the operation and management of wastewater treatment plants, and is often expressed by the symbol COD.

3.1.7 BOD

Biochemical oxygen demand is the amount of dissolved oxygen consumed during the biochemical reactions carried out by microorganisms to decompose the biochemically degradable organic matter present in water under certain conditions. It is expressed in milligrams per liter or percent, ppm. It is a comprehensive indicator of the content of organic pollutants in water. If the time for biological oxidation is five days it is called five-day biochemical oxygen demand (BOD_5), and accordingly there are BOD_{10} , BOD_{20} .

3.1.8 Potassium Permanganate

Under certain conditions, the amount of oxygen consumed by potassium permanganate to oxidize organic matter and some reduced inorganic matter in water. It is expressed as CODMn. It can indirectly reflect the total amount of organic and partially reduced inorganic substances in water. In the discipline of water treatment, it is mainly used as an indicator of the content of organic matter. The larger the value, the higher the content of organic matter in water, the worse the water quality

3.1.9 Fluorine

Fluorine is the most active non-metallic element, which can be combined with almost all metals and non-metallic elements at room temperature, and can react violently with water, and exists in nature in the form of stable fluoride.

The sources of fluorine in water bodies can be divided into two categories: natural sources and anthropogenic pollution. Since fluorine exists widely in nature and fluoride has high solubility, rainwater,

surface water and groundwater flowing through fluoride-containing soil and rocks can leach some fluorine from them, so all kinds of water bodies contain certain amount of fluoride. The fluorine content in rainwater is relatively low, about 0.05 0.1mg/L, surface water is generally 0.2 0.5mg/L, groundwater contains fluorine from trace to more than 10mg/L, groundwater flowing through fluoride-bearing mineral layer can reach 2 5mg/L, hot spring water contains high fluorine, usually 10mg/L, some can be as high as 300mg/L.

3.2 Data Selection

In order to make the selected data more valid and fit the model, in the time dimension, we chose even numbered months from **2014-2018** as the time variables, as described in the assumptions (General Assumption section).

In terms of sampling points, the given data of 2015-2017 are **only four sampling points**, namely Huangbai Bridge, Ningchegu Gate, Yujialing Bridge and Xiaogao Tuo, but for these four sampling points, all indicators of Xiaogao Tuo are missing in 2014, while the other sampling points of Jinwei Highway Bridge, Jinyu Highway Bridge, Xiaoguan Village and Wucun Gate appear in the 14-year data, except for Wucun Gate which has corresponding indicator data in 18 years. The general fitting method is impossible to fit three years and more data from one year data, and the fitted data obtained by using Markov chain Monte Carlo method is too random, so we choose to abandon these four sampling points.

According to the completeness of the data and the degree of influence, we choose the above nine indicators as the initial indicators. It should be noted that the reason for not choosing water temperature indicators is that dissolved oxygen is very obviously affected by water temperature, and the changes of these two groups of indicators are basically the **same**, and the dissolved oxygen content basically decreases in the same proportion with the increase of water temperature. As for the indicators such as iron and manganese, if the two-year data are fitted through the three-year data, such data are not acceptable. In the data processing of this model, the maximum degree of fit we can accept is to fit the data of one year based on the data of four years, so we abandoned the indicators of iron, manganese, selenium, chloride, nitrate, anion selector, to fit the data of the defective years of the selected indicators to supplement.

3.3 Data Pretreatment

After data filtering, we finally determined that the data needed were the data of PH, dissolved oxygen, ammonia nitrogen, total phosphorus, total nitrogen, potassium permanganate index, chemical oxygen demand, biochemical oxygen demand and fluoride at four sampling points of Huangbai Bridge(Excluding Yujialing Bridge February 2014 data), Xiaogao Tuo, Ningchegu Gate and Yujialing Bridge in even-numbered months from 2014 to 2018. The absolute error of all the predicted data **percentages does not exceed 5%**, which is within the range we can accept and use. We also performed **gray prediction** of fluoride data for each sampling site for each even-numbered month in 2014 by using the fluoride data for all sampling sites from 2015 to 2018. At this point, all the data we need have been obtained. We analyzed the data in two dimensions.

3.3.1 Time Dimension

We used time as a variable from 2014 to 2018 for all even-numbered months of each indicator data for a curve fitting with a five-year horizontal axis span, from which we obtained a time change curve for each indicator, a total of 36 time change curves for the four sampling points, so as to have a clear understanding

of the change of all indicators over time for each sampling point.

The curves and analysis are as follows.

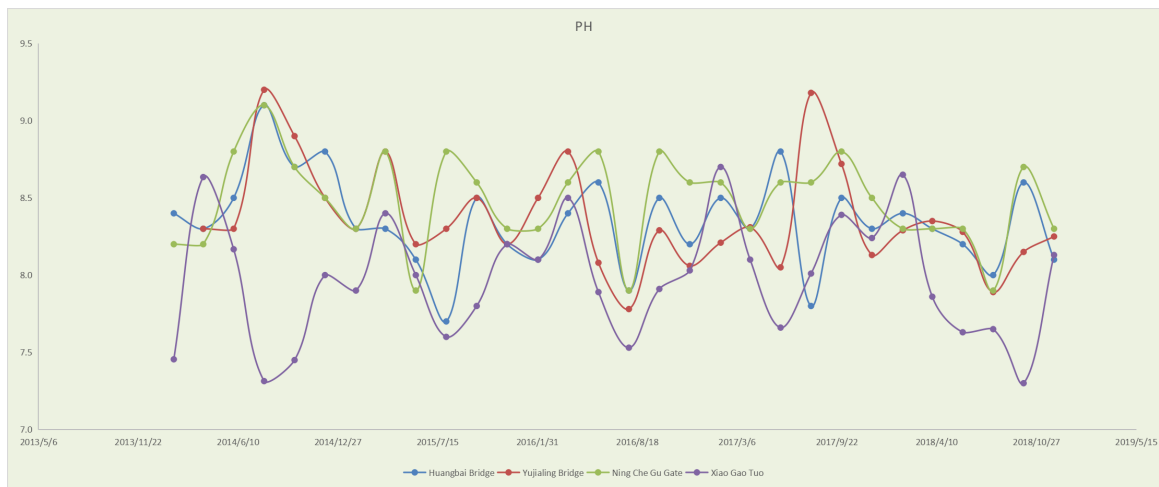


Figure 1: PH change curve of each sampling point in the New Chaobai River from 2014 to 2018

It can be seen that the fluctuation of PH influenced by month is more significant, while for different years, there is no particularly large fluctuation, basically fluctuating in a horizontal straight line, which also reflects that the PH of these four sampling points is more stable Over a span of time measured in years.

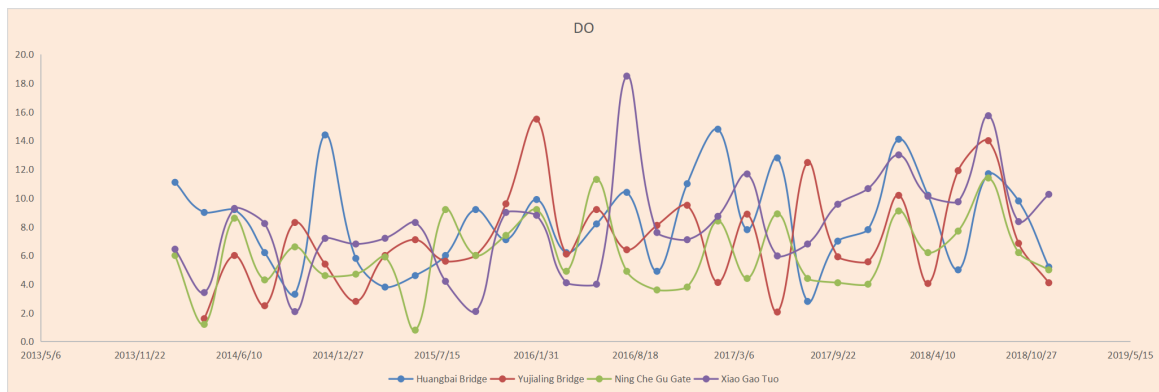


Figure 2: Dissolved oxygen change curve of each sampling point in the New Chaobai River from 2014 to 2018

Dissolved oxygen was significantly affected by water temperature, with higher water temperatures resulting in lower dissolved oxygen values. Basically, the values are higher in autumn and winter and lower in spring and summer seasons. Some of the curves that do not fit this conclusion are due to biases in the data processing and fitting. There is a slight **upward trend** in the dissolved oxygen values over the entire five-year time span, with differences in the time when the curve peaks at the four sampling sites. We can attribute the cause of this phenomenon to the influence of the location of the sampling sites and the external environment.

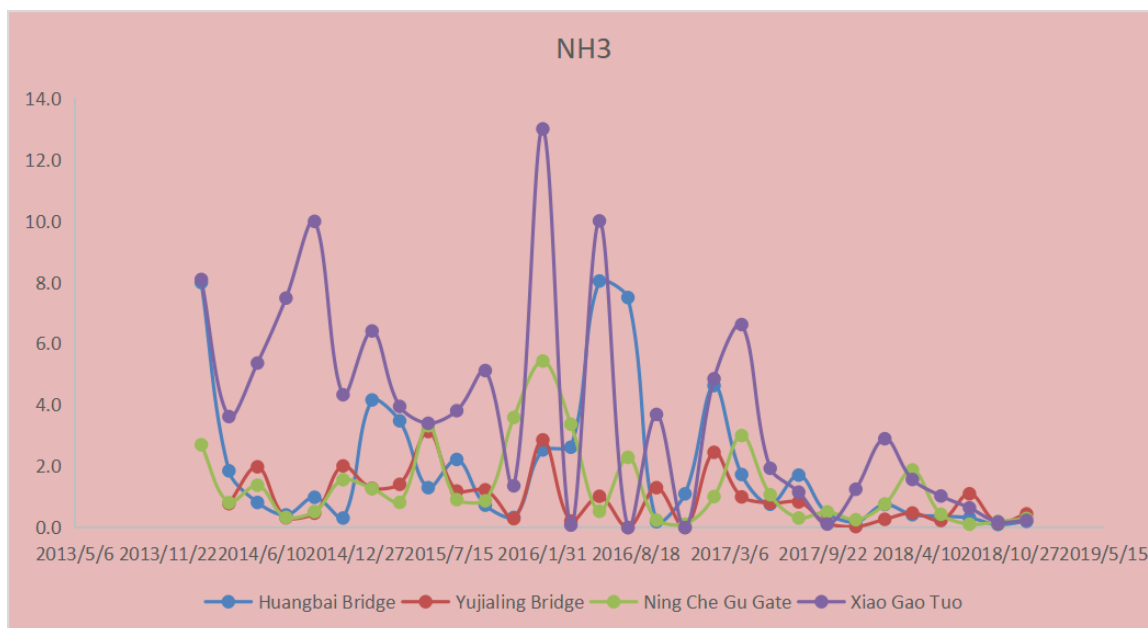


Figure 3: Ammonia nitrogen change curve of each sampling point in the New Chaobai River from 2014 to 2018

The values of ammonia nitrogen show a trend of increasing and then decreasing throughout the time span. The values of ammonia nitrogen in 2016 were very large at all sampling sites, and the level of this indicator varied very much from one sampling site to another, which we will analyze later in the spatial dimension with annual averages. In addition, the values for 2018 also show a slight decrease compared to the values prior to 2016, which is also attributed to the treatment of the New Chaobai River.

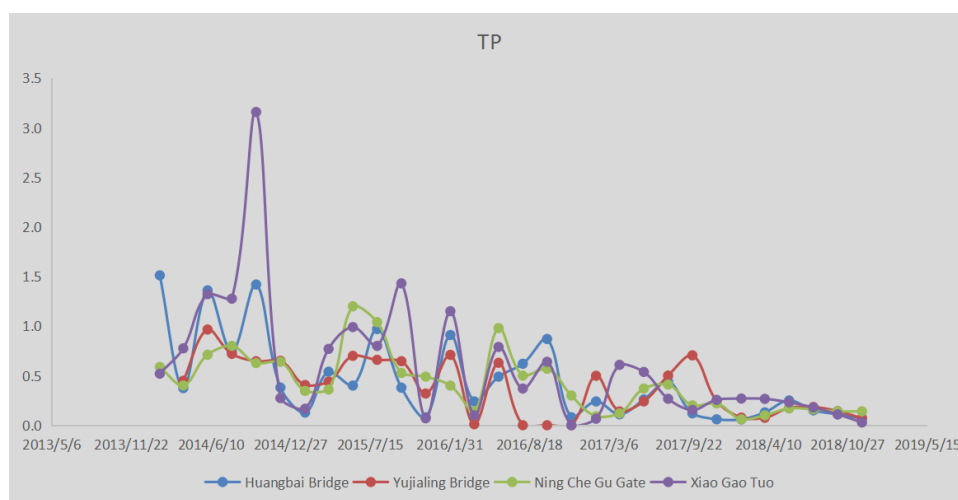


Figure 4: Total phosphorus change curve of each sampling point in the New Chaobai River from 2014 to 2018

The fluctuation of total phosphorus in various periods gradually diminished from 2014 until it reached stability in 2016, showing an overall **decreasing trend**. Obviously, this change is inseparably related to the management of river pollution.

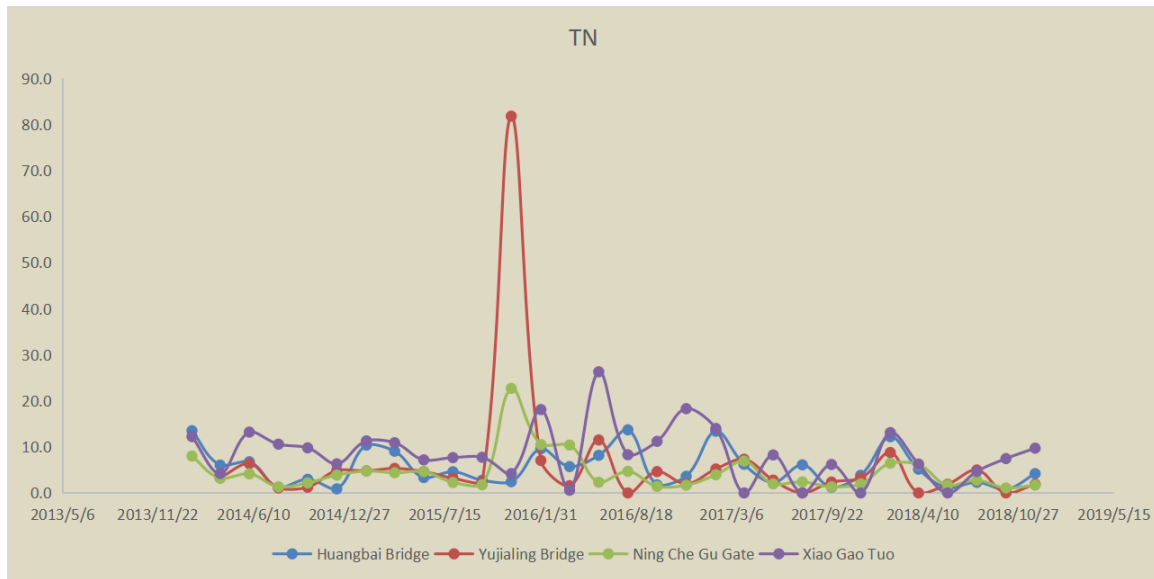


Figure 5: Total Nitrogen curve of each sampling point in the New Chaobai River from 2014 to 2018

Total nitrogen fluctuates relatively steadily throughout the time span of the data sampling, and the fluctuation is smaller than that of PH. The trend is consistent across the sampling sites, with the exception of a large bump at Yujialing Bridge at the end of 2015. If we rule out the data as the cause, we can only speculate that there was some **unexpected situation** at this sampling site at this time.

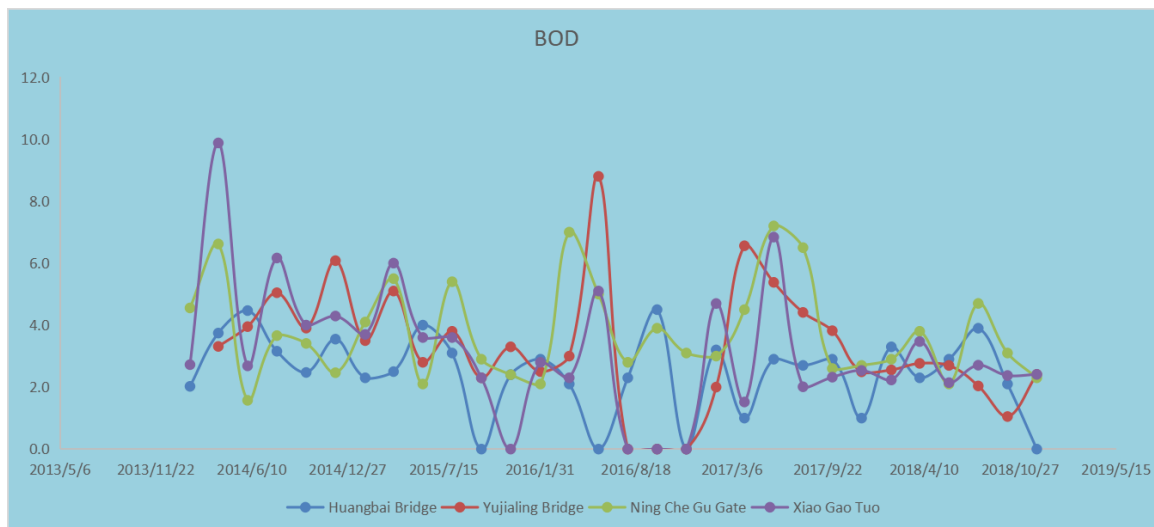


Figure 6: COD curve of each sampling point in the New Chaobai River from 2014 to 2018

The variation of COD is very **irregular**, and the curve shows that it is less affected by time, and the curves of the four sampling sites show surprisingly consistent trends in this indicator. There is a strong correlation between COD and factory discharges during the same period, so we speculate that these erratic changes with consistent trends are due to changes in industrial domestic wastewater discharges near the New Chaobai River over time.

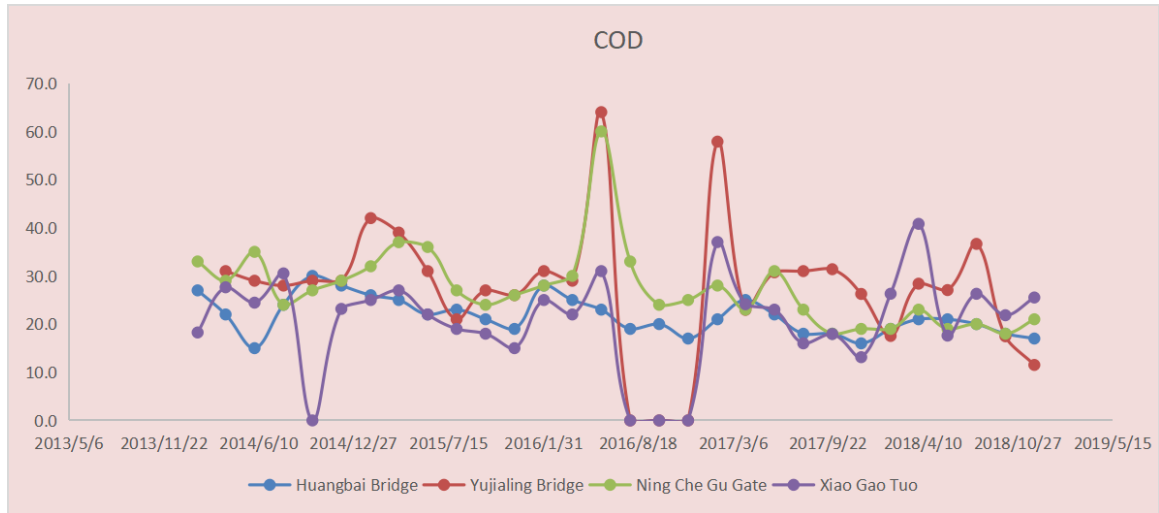


Figure 7: BOD curve of each sampling point in the New Chaobai River from 2014 to 2018

The variation of BOD is also very irregular and has a high consistency of variation with COD, and both reach a trough in August 2016. The definitions of the two indicators, and our later analysis of the indicators, lead to conclusions that are largely consistent with these conclusions we obtained from the images, i.e., there is a **high degree of consistency** between BOD and COD.

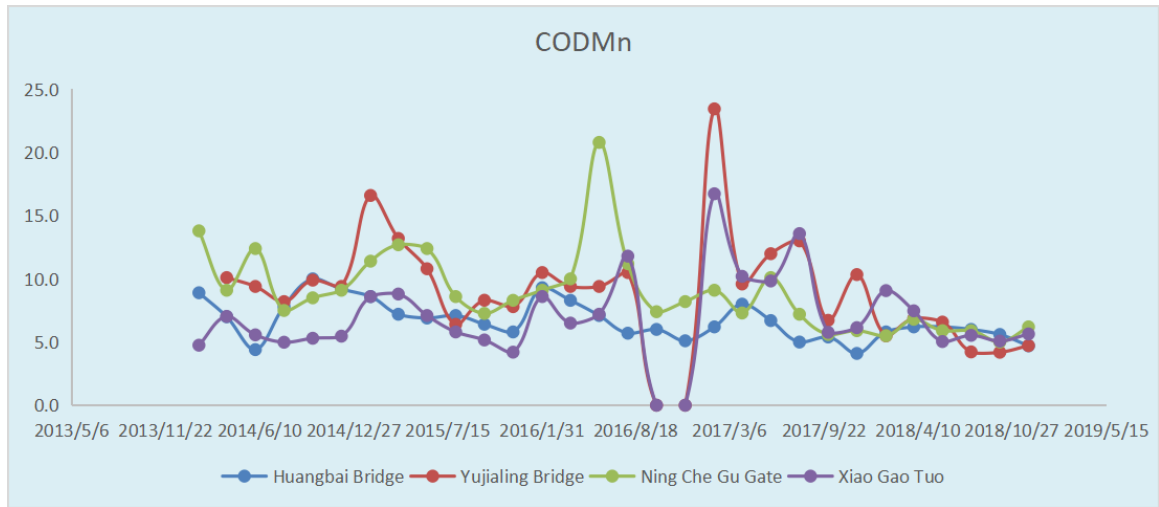


Figure 8: The Permanganate Index curve of each sampling point in the New Chaobai River from 2014 to 2018

The Permanganate Index also has almost the same characteristics as the COD and BOD curves, which is the accordance for the approximation of the weights of these three indicators in the later AHP weighting process.

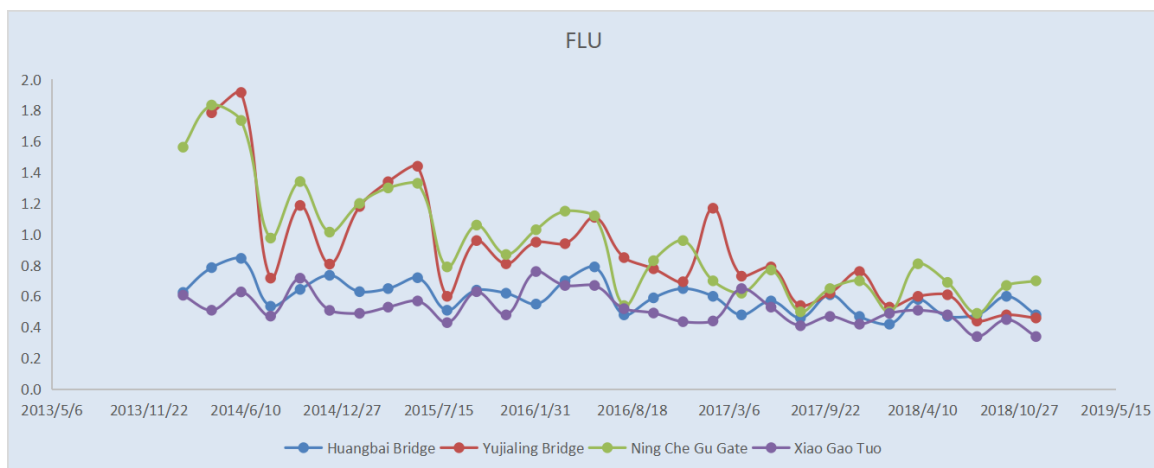


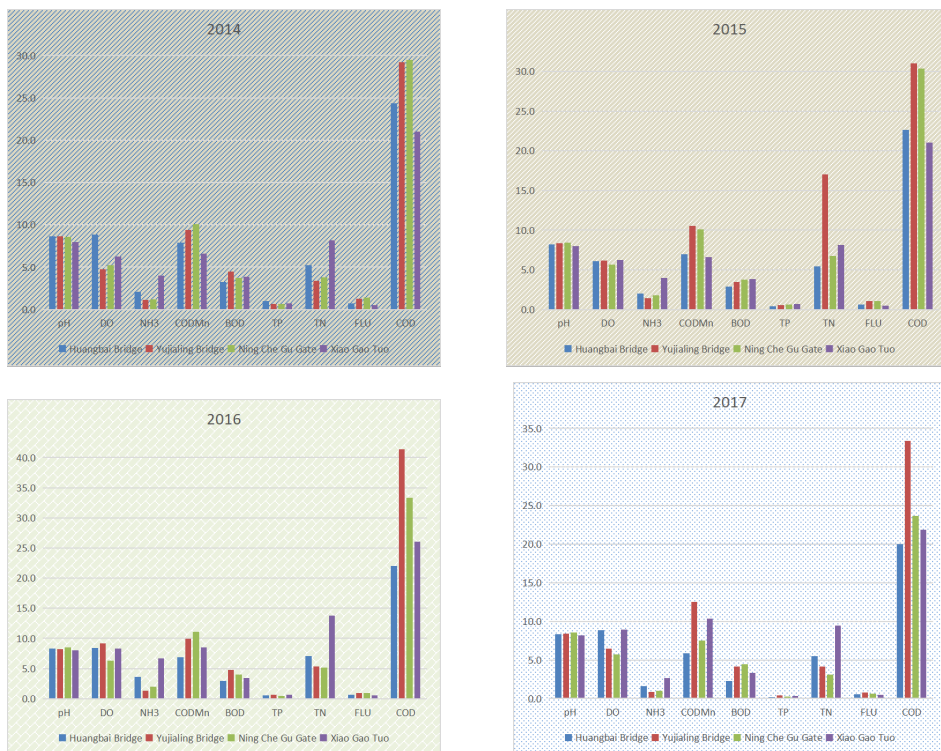
Figure 9: Fluoride curve of each sampling point in the New Chaobai River from 2014 to 2018

The fluoride content is very **stable** with a slight downward trend, and the fluoride values for the entire time span meet the fluoride requirements for Class I in the existing evaluation system. This indicator has been at an excellent level.

3.3.2 Spatial Dimension

We compare the data of the same indicator at four sampling points **at the same time** (here we take the annual average) to get the spatial difference of the same indicator in the same year, i.e. the sampling point difference, and make a bar chart for observation and analysis.

The histogram is as follows.



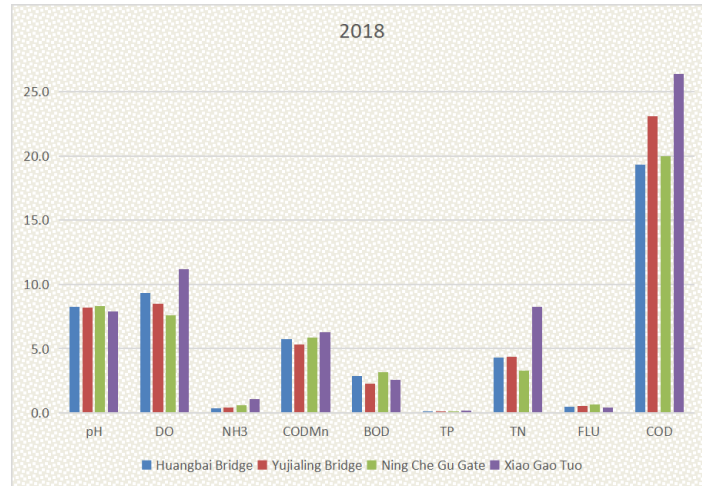


Figure 10: Sampling sites from 2014 to 2018

We analyze these histograms by comparing the values of each indicator at different sampling sites at the same time.

1. PH remained essentially the same at each sampling site for each year, and did not vary by sampling sites.

2. Dissolved oxygen had a slight **decrease** in value between 2014 and 2015, with the most obvious decrease at the Huangbai Bridge sampling site. Thereafter, the dissolved oxygen values increased year by year, with the greatest increase from 2017 to 2018, and this rising change was most obvious at the Xiaogao Tuo sampling site.

3. The value of ammonia nitrogen had a significant increase between 2015 and 2016, and then decreased year by year in the following two years. The value of ammonia nitrogen at the sampling point of Xiaogao Tuo has been the highest, and Huangbai Bridge also dropped to the lowest in 2018 from the second highest before 2017. In 2018, the value of ammonia nitrogen at the remaining three sampling points was basically the same except for Xiaogao Tuo.

4. By analyzing the various indicators at different sampling points, we can derive from the graph the ranking of the same indicator at different sampling points in the same period. By deduction we found that the New Chaobai River was effectively governed in 2016 and the water quality of Yujialing Bridge began to slowly improve, while in terms of chemical oxygen demand Xiaogao Tuo still had a significant increase from 2017 to 2018 years.

4 APPROACH

In order to avoid the disadvantage that the objective weighting method is too dependent on the data itself, which causes the weighting not to fit the actual use well, and the lack of objectivity of the subjective weighting method, which causes the control of the modeling results by the emotion of each person. We use **both PCA and AHP models** for simultaneous modeling and finally reassign the weights of each indicator from both models through data analysis and the influence of each indicator on the evaluation results to obtain a weighting method that combines the advantages of both assignment methods and fits the data and the actual situation^[1].

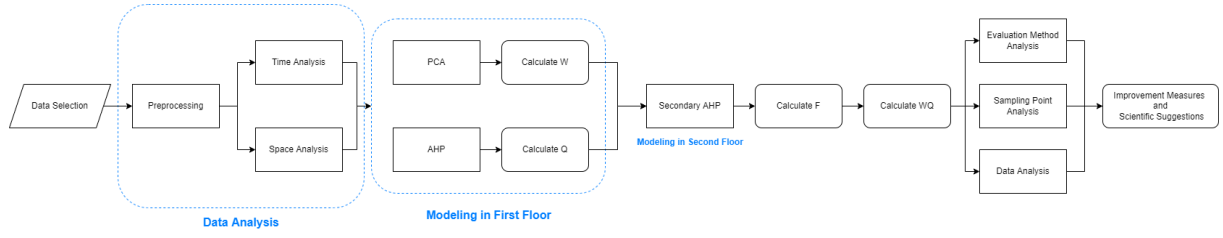


Figure 11: PCA-AHP flow chart

4.1 Principal Component Analysis

Principal Component Analysis (PCA), also known as principal component analysis, is a multivariate statistical analysis method in which multiple parameter variables are computed by linear transformation to select a smaller number of significant variables.

4.1.1 Basic principles

With n samples, each involving m variables, denoted by $X_1, X_2, X_3 \dots X_m$, forming a data matrix of order $m \times n$.

$$\begin{bmatrix} X_{11} & X_{12} & \dots & X_{1m} \\ X_{21} & X_{22} & \dots & X_{2m} \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ X_{n1} & X_{n2} & \dots & X_{nm} \end{bmatrix} \quad (1)$$

When m is large, it is complicated and tedious to examine the problem in m -dimensional space. And dimensionality reduction can easily overcome this difficulty.

Let their combined indicators be represented by $Z_1, Z_2 \dots Z_m$. The new composite indicator can be represented by a linear combination of the original indicators.

$$\begin{aligned} z_1 &= l_{11}x_1 + l_{12}x_2 + \dots + l_{1m}x_m \\ z_2 &= l_{21}x_1 + l_{22}x_2 + \dots + l_{2m}x_m \\ &\dots \\ z_m &= l_{m1}x_1 + l_{m2}x_2 + \dots + l_{mm}x_m \end{aligned} \quad (2)$$

i.e.

$$Z = LX \quad (3)$$

The required is L

4.1.2 Calculation steps

1. The bias correlation coefficients between the scalars met the requirements by the **Kaiser-Meyer-Olkin (KMO)** test and could be used for principal component analysis. Then, we imported the data into SPSS software.

Table 1: Raw data

PH	DO	NH ₃	CODMn	BOD	TP	TN	FLU	COD
8.35	8.31	1.94	6.69	2.85	0.45	5.51	0.60	21.67
8.35	8.31	1.04	9.55	3.81	0.48	6.86	0.90	31.60
8.48	7.02	1.04	8.94	3.80	0.43	4.42	0.95	27.37
7.79	8.17	4.19	7.47	3.61	0.61	9.81	0.52	23.19

2. **Standardize** the original 9 indicators to eliminate the effect of variables in level and magnitude

Table 2: Standardised data

ZPH	ZDO	ZNH ₃	ZCODMn	ZBOD	ZTP	ZTN	ZFLU	ZCOD
0.23041	0.87448	-0.12895	-1.12019	-1.47073	-0.50589	-0.48954	-0.67722	-0.95958
0.38953	-0.36238	-0.76053	1.05447	0.64188	-0.14594	0.08836	0.74731	1.26330
0.83086	-1.24924	-0.55760	0.59461	0.61623	-0.79443	-0.95528	0.96006	0.31567
-1.45081	0.73713	1.44708	-0.52888	0.21262	1.44626	1.35646	-1.03014	-0.61939

3. Based on the normalized data matrix, the correlation coefficient matrix is found.

Table 3: Correlation

	Z _{score} (PH)	Z _{score} (DO)	Z _{score} (NH ₃)	Z _{score} (CODMn)	Z _{score} (BOD)	Z _{score} (TP)	Z _{score} (TN)	Z _{score} (FLU)	Z _{score} (COD)
Z _{score} (PH)	1.000	-.682	-.963	.471	.038	-.977	.477	.809	-.947
Z _{score} (DO)	-.682	1.000	.642	-.831	-.711	.556	-.716	-.941	-.578
Z _{score} (NH ₃)	-.963	.642	1.000	-.585	-.111	-.904	-.636	-.836	.830
Z _{score} (CODMn)	.471	-.831	-.585	1.000	.859	-.275	.974	.887	-.215
Z _{score} (BOD)	.038	-.711	-.111	.859	1.000	.156	.762	.616	.159
Z _{score} (TP)	-.977	.556	.904	-.275	.156	1.000	-.282	.673	.985
Z _{score} (TN)	.477	-.716	-.636	.974	.762	-.282	1.000	.845	-.187
Z _{score} (FLU)	.809	-.941	-.836	.887	.616	-.673	.845	1.000	-.639
Z _{score} (COD)	-.947	.578	.830	-.215	.159	.985	-.187	-.639	1.000

From the above figure, it can be seen that there is a **strong** correlation between the variables, which is suitable for principal component analysis.

4. By using factor analysis, the principal components are identified and the information contained in each principal component is interpreted appropriately.

(1) From the scree plot, we can see that the slope turns flat after the third point, so we initially determine that there are two principal components.

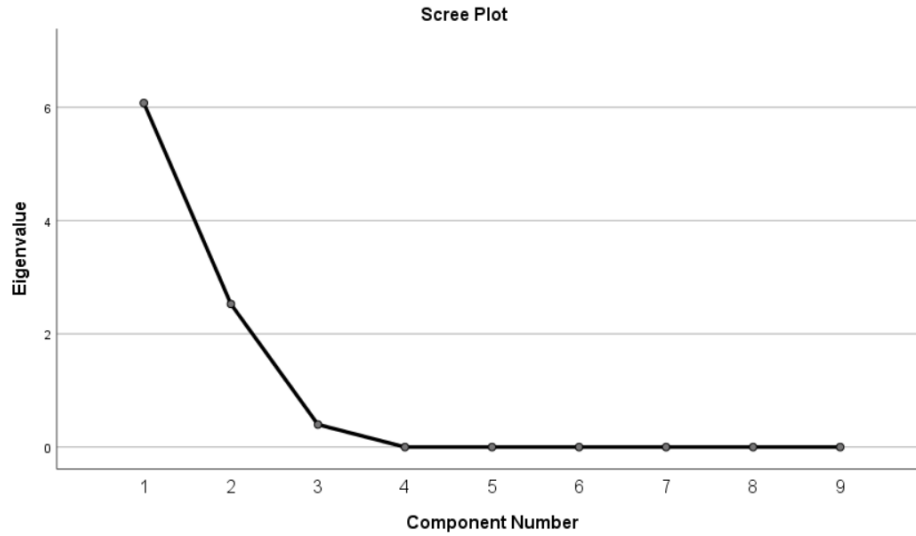


Figure 12: Scree Plot

(2) The first few principal components with a cumulative variance contribution of 85% or more can be selected as the final principal component.

Table 4: Total Variance Explained

	Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
		Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
Raw	1	6.076	67.513	67.513	6.076	67.513	67.513
	2	2.256	28.069	95.582	2.526	28.069	95.582
	3	.398	4.418	100.000			
	4	7.206E-16	8.007E-15	100.000			
	5	4.980E-16	5.534E-15	100.000			
	6	5.101E-17	5.668E-16	100.000			
	7	-9.496E-17	-1.055E-15	100.000			
	8	-1.522E-16	-1.691E-15	100.000			
	9	-9.788E-16	-1.088E-14	100.000			
Rescaled	1	6.076	67.513	67.513	6.076	67.513	67.513
	2	2.526	28.069	95.582	2.526	28.069	95.582
	3	.398	4.418	100.000			
	4	7.206E-16	8.007E-15	100.000			
	5	4.980E-16	5.534E-15	100.000			
	6	5.101E-17	5.668E-16	100.000			
	7	-9.496E-17	-1.055E-15	100.000			
	8	-1.522E-16	-1.691E-15	100.000			
	9	-9.788E-16	-1.088E-14	100.000			

From this figure, it can be further determined that the principal components are Z_1 , Z_2 .

(3) Using the component matrix, the linear correlation coefficients of the principal components with each indicator can be observed. The **larger** the absolute value of the coefficients, the more representative the principal components are of the variable. PH, DO, NH_3 , FLU are more representative of principal component 1. BOD is more representative of principal component 2.

Table 5: Component Matrix

	Raw Component		Rescaled Component	
	1	2	1	2
$Z_{score}(PH)$	-.892	-.451	.892	-.451
$Z_{score}(DO)$	-.899	-.242	-.899	-.242
$Z_{score}(NH_3)$	-.909	.320	-.909	.320
$Z_{score}(CODMn)$.814	.572	.814	.572
$Z_{score}(BOD)$.481	.855	.481	.855
$Z_{score}(TP)$	-.780	.626	-.780	.626
$Z_{score}(COD)$.788	.522	.788	.522
$Z_{score}(FLU)$.987	.156	.987	.156
$Z_{score}(TN)$	-.741	.645	-.741	.645

5. Since we are using factor analysis with SPSS software, we transform the component matrix into the eigenvectors V of the principal components by calculation.
i.e.

$$Z = \begin{bmatrix} 0.36, & -0.36, & -0.37, & 0.33, & 0.20, & -0.32, & -0.30, & 0.40, & 0.32 \\ -0.28, & -0.15, & 0.20, & 0.36, & 0.54, & 0.39, & 0.41, & 0.10, & 0.33 \end{bmatrix} \times \begin{bmatrix} PH \\ DO \\ NH_3 \\ CODMn \\ BOD \\ TP \\ TN \\ FLU \\ COD \end{bmatrix} \quad (4)$$

6.Finally, by using the eigenvector formula for the principal components:

$$W_j = \sum_{i=0}^2 \frac{L_{ij}a_j}{\sum_{i=0}^2 L_{ij}^2} \quad (5)$$

The principal component analysis weights are:

$$W = [0.110584, 0.097362, 0.104728, 0.111878, 0.100534, 0.112648, 0.11012, 0.1014, 0.106546] \quad (6)$$

4.2 AHP

The main function of principal component analysis in water quality evaluation is to reduce the dimensions of the water quality parameters of the original data, and to screen out the main factors affecting water quality, which can quickly evaluate the main environmental problems of the evaluation object.

The **disadvantage** is that the weights obtained are only based on the measured data, and the impact of the physical and chemical properties of the water quality parameters on the environmental pollution is **ignored**. Because the principal component analysis method relies on the weight calculated by the relationship between the various index parameters of the original data itself, (the original data is determined and not affected by the subjective factors of the calculator), so the calculated weight is called the **objective weight** of the original data.

The analytic hierarchy process discusses the weight of each pollution factor for a certain environmental problem. The basis for calculating the weight comes from the nature of the water quality parameters related to the environmental problem. The calculated weights can well reflect the ranking of the impacts of various water quality parameters on environmental issues. The **disadvantage** is that it is often difficult to take the measured data into account, and the calculated weight cannot represent the actual pollution of the research object. At the same time, when there are too many water quality parameters, the constructed hierarchical structure model is too complicated as well as the calculation.

In order to be able to solve the weakness of subjective weighting and objective weighting, we can **combine** principal component analysis and hierarchical analysis to build a PCA-AHP dimensionality reduction combination weighting evaluation model [2,3,4]. Based on the actual raw measurement data, we use PCA to reduce the dimensionality of the data, filter out the main water quality pollution parameters, and then can evaluate the main environmental problems of the study object. Meanwhile, we use AHP to discriminate the filtered water quality parameters. The advantages of this model are that both the original data and the water quality pollution parameters of their own properties are considered, and the final combination of weights calculated could take into account the actual pollution status of water quality parameters and their properties respectively, so that the evaluation results are more fit the reality.

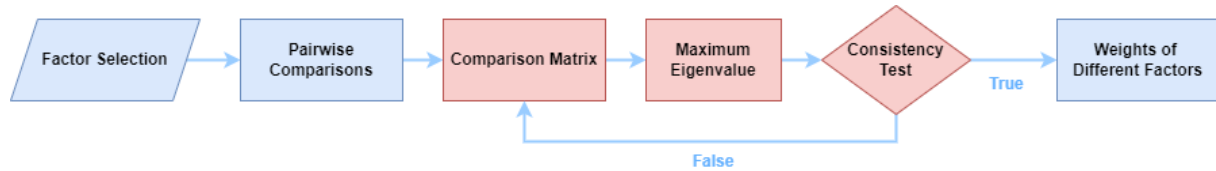


Figure 13: AHP

4.2.1 Build a hierarchical model

The highest level refers to the target level, denoted by O, which represents the water quality status obtained by calculating each indicator after weighting in the analytic hierarchy model, while the lowest level refers to the 9 indicators we selected, denoted by $C_n(n=1, 2, \dots, 9)$.

Index comparison Through the data processing analysis and the understanding of the importance of each index in the existing water quality evaluation system, we believe that the importance of the index is ranked as: $NH_3 > TP \approx TN > FLU \approx BOD \approx COD \approx CODMn > DO \approx PH$

Calculation of the comparison matrix According to our judgment on the importance of the index, we subjectively get the comparison matrix:

$$C = \begin{bmatrix} 1 & 5 & 5 & 7 & 7 & 7 & 7 & 9 & 9 \\ 1/5 & 1 & 3 & 5 & 5 & 5 & 5 & 7 & 7 \\ 1/5 & 1/3 & 1 & 3 & 3 & 3 & 3 & 5 & 5 \\ 1/7 & 1/5 & 1/3 & 1 & 4 & 3 & 3 & 5 & 5 \\ 1/7 & 1/5 & 1/3 & 1/3 & 1 & 1 & 1 & 3 & 3 \\ 1/7 & 1/5 & 1/3 & 1/3 & 1 & 1 & 1 & 3 & 3 \\ 1/7 & 1/5 & 1/3 & 1/3 & 1 & 1 & 1 & 3 & 3 \\ 1/9 & 1/7 & 1/5 & 1/5 & 1/3 & 1/3 & 1/3 & 1 & 1 \\ 1/9 & 1/7 & 1/5 & 1/5 & 1/3 & 1/3 & 1/3 & 1 & 1 \end{bmatrix} \quad (7)$$

Consistency test We find the eigenvector Q_{max} corresponding to the largest eigenvalue λ_{max} of each level of judgment matrix. The elements of Q_j obtained by normalization (simplified as a unit vector) are the ranking weights of the relative importance of the factors of the same level to the factors of the previous level.

The eigenvector Q_{max} corresponding to the maximum eigenvalue λ_{max} is used as the degree of inconsistency of the evaluation weight vector. The greater the degree of inconsistency, the greater the judgment error caused. The magnitude of the value of $\lambda - n$ can be used to measure the degree of inconsistency of the judgment matrix.

Define the consistency index as:

$$CI = \frac{\lambda - n}{n - 1} \quad (8)$$

If $CI=0$, there is complete consistency; if CI is close to 0, there is satisfactory consistency; the larger the CI , the more serious the inconsistency.

To measure the size of CI , we introduce the random consistency index RI :

$$RI = \frac{CI_1 + CI_2 + \dots CI_n}{n} \quad (9)$$

Among them, the value of the random consistency index RI is only related to the order of the judgment matrix, and the corresponding relationship is as follows:

Table 6: Standard value of average random consistency index RI

N(Matrix dimension)	1	2	3	4	5	6	7	8	9
RI	0	0	0.58	0.90	1.12	1.24	1.32	1.41	1.45

Considering that the deviation of consistency may be caused by random reasons, it is also necessary to compare CI and random consistency index RI to obtain the test coefficient CR . The formula is as follows:

$$CR = \frac{CI}{RI} \quad (10)$$

Generally, if $CR < 0.1$, the judgment matrix is considered to pass the consistency test, otherwise it does not have satisfactory consistency.

Results of the calculation Bring it in the comparison matrix we set, **the consistency check is passed**, and the eigenvector corresponding to the largest eigenvalue obtained is normalized to get the weight.

$$Q = [0.392, 0.211, 0.122, 0.087, 0.048, 0.048, 0.048, 0.022, 0.022] \quad (11)$$

4.3 PCA-AHP two-layer modeling

4.3.1 PCA-AHP comprehensive empowerment

Building the model We have a subjective judgment of the importance of each index in the previous data pre-processing, and apply this judgment to the AHP model, combined with the PCA model based only on the measured data, the weights W and Q in the two models obtained again using hierarchical analysis to re-assign the weights, so as to correct the over-reliance of PCA on the measured data in a subjective way and ignore the actual condition factors, while **correcting** the shortcomings of AHP too

subjective to link it to the measured data ^[1,5].

We use the same steps of hierarchical analysis as above, and the consistency test is passed, we normalize the obtained feature vectors to get the weights.

$$\beta = [0.75, 0.25] \quad (12)$$

And then, by the formula

$$F = \beta^T [Q, W] \quad (13)$$

We obtain the modified composite weights

$$F = [0.181, 0.126, 0.109, 0.106, 0.087, 0.096, 0.094, 0.081, 0.085] \quad (14)$$

We weighed all index data at the same time to get the water quality index of the current month (WQ), and to get the model's result for water quality evaluation by WQ value.

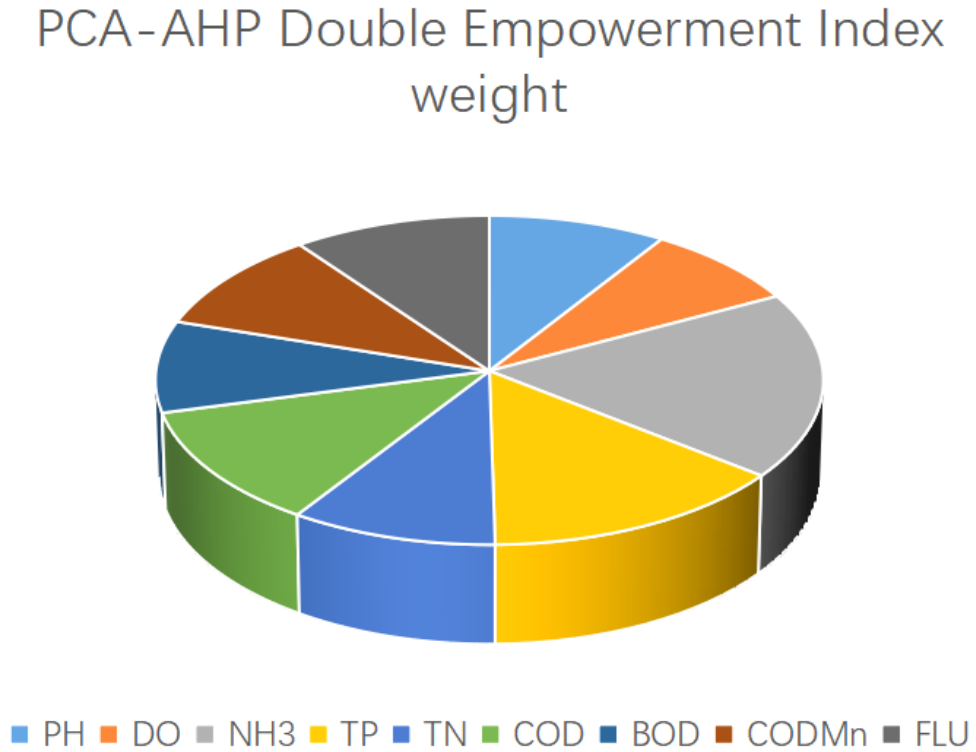


Figure 14: Weighting pie chart

Calculation of results We first normalize all values of each indicator for each sampling point indicator over the whole time span to counterbalance the unfair effects on the calculation results caused by the difference in order of magnitude after assigning weights, **so that all indicators are calculated at the same order of magnitude and comparability is obtained.**

In the normalization process, the impact of dissolved oxygen trend and the rest of the indicators of the opposite trend, that is, the higher the dissolved oxygen water quality evaluation results are better, the larger the value of other indicators assessment results are worse. Therefore, when we weighted the calculation, **we corrected the sign of the dissolved oxygen data:** when it is used in the weighting operation,

a negative sign is added to ensure that the results are consistent with the impact of the indicators.

We plotted a representation of the final WQ values to obtain a curve of the variation of WQ over a five-year time span at four sampling points.

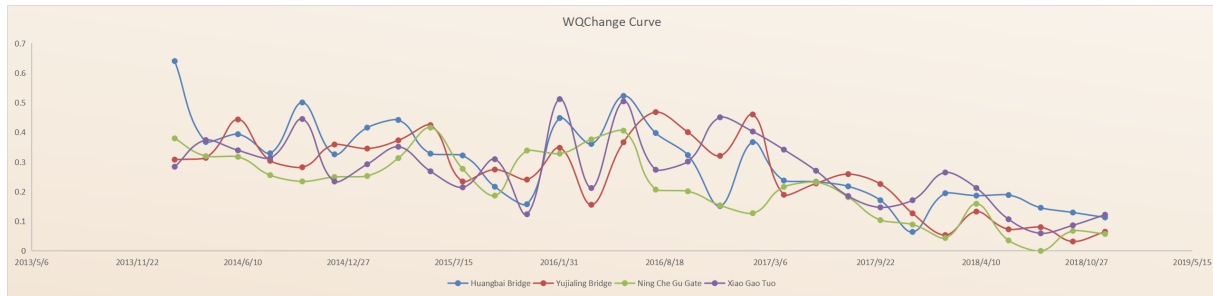


Figure 15: WQ

As can be seen from the figure, the overall WQ curve trend contains many of the water quality characteristics obtained during our previous data pre-processing and is also highest in 2016, with an overall trend showing a **downward trend**, which indicates that our model fits the characteristics of the available data and further illustrates the practical applicability of our model for the analysis of this problem. (Note: Higher WQ values represent more severe pollution levels).

By comparing the results obtained with the existing evaluation methods, we conducted a standard WQ value calculation (by calculating the average WQ value of the existing data of each type) for the existing V, IV and III categories, and obtained the following correspondence, **so that the water quality WQ value for the three-level classification.**

Table 7: Comparison of outcome values with existing evaluations

Existing evaluation level	Average of improved evaluation methods	Maximum value of improved evaluation methods
Class V	0.2790	0.6402
Class IV	0.1427	0.3739
Class III	0.0879	0.1296

Thus, in the comprehensive evaluation of existing data, our model can better meet the requirements of the comprehensive comparison between different data, but also the evaluation of the water quality of a place has a quantitative index can be queried, so that the water quality level is more specific.

Comparison with existing evaluation methods and rationality analysis The existing evaluation methods can better reflect the impact on water quality of a single indicator, and is more suitable for water quality for water use assessment and characteristics of the indicators used when limited, but the rest of the indicators are vulnerable to an extreme indicator masking leading to completely ignore the role of the remaining indicators on water quality evaluation. Our PCA-AHP double-layer weighting model solves the problem of commonly used weighting methods or subjective or objective, to achieve an organic unification of objective and subjective evaluation, through the weighting of all indicators to obtain a final determinant of the method, **the water quality status of such a common category of general expression of the characteristics of numerical, specific, and perfectly fit the existing model and the significant characteristics of the data.** Therefore, we believe that the use of such a model for water quality evaluation is very reasonable and feasible.

Improvements to existing evaluation methods For example, for the data from Xiao Gao Tuo in 2014, we calculated a WQ value of 0.3739, which exceeds the average value of Class IV water, so we classified it as Class V. Under more balanced and reasonable weighting conditions, the overall water quality status of this sampling site for this month is relatively poor.

In contrast, for the August 2018 data from Ning Che Gu Lock, we calculated a WQ value of 0.00013, which is a very small value, and its rating as Class V water in the existing evaluation method is influenced by the high TN value, which overshadows the excellent levels of other indicators, and therefore the water quality condition at this sampling site for this month is relatively good in our model.

Therefore for us in the original evaluation method for individual indicators of the special requirements of the premise, increase our WQ value, the above three average value as a cut-off point for water quality level evaluation, **not only to retain the original method of evaluation of practicality, but also to increase the evaluation method of comparative performance, as well as for the overall water quality situation of the numerical and specific function.**

5 Analysis of sampling points

5.1 Introduction to sampling sites

Table 8: Existing sampling sites

Sampling sites	Longitude and Latitude	Basic Information
Yujialing Bridge	(117.645837, 39.217584)	Located in the south of Tianjin, the second most southerly of all the sites sampled
Ning Che Gu Gate	(117.664711, 39.159724)	Located in the south of Tianjin, the most southerly of all the sites sampled
Xiao Gao Tuo	(117.135913, 39.744498)	Located in Langfang, Hebei, the most northern of all sampling sites
Huangbai Bridge	(117.374036, 39.402133)	Located on the west bank of the Chaobai River, Erwangzhuang Reservoir
Jinwei Highway Bridge	(117.202148, 39.382593)	Located in Yongding New River in Southern Tianjin
Jingyu Highway Bridge	(117.160419, 39.264747)	Located on the Yongding Xinhe River in southern Tianjin
Xiaoguanzhuang	(117.759673, 39.327731)	The Ji Canal in southeastern Tianjin, which has a tributary that flows into the New Chaobai River

5.2 Sampling point reasonableness analysis and coordinate supplement

5.2.1 Analysis of the premise

- The Yongding New River joins the New Chaobai River to the south of the Ningchegu Gate
- A tributary of the Ji Canal flows into the New Chaobai River
- The Chaobai River is now known as the Chaobai Xinhe in Tianjin

5.2.2 Rational analysis

(1) Sampling point analysis

The three sampling sites selected from the original data, Jinwei Highway Bridge, Jinyu Highway Bridge and Xiaoguanzhuang, which I screened out because of data deficiencies, are not directly on the New Chaobai River, so the water quality data from these three sites are not reasonable for the assessment of the water quality of the Chaobai River due to the temporal differences caused by spatial mobility. For the four sampling sites with complete data, they are located in the north and south of the Chaobai River, which can largely reflect the water quality of the north and south ends of the river, but this distribution of sampling sites leads to the lack of sampling sites in the middle of the river.

(2) Sampling point requirements

After discussion and study of the relevant dissertations, we understand that sampling points should be set

up mainly at the main inlets and outlets of rivers, at the confluence of tributaries and at major outfalls, and distributed as evenly as possible in space so that the water quality of the whole river can be consistently reflected.

(3) Sampling point supplement

(117.666942, 39.157771) located at the confluence of the Chaobai and Yongding rivers, it reflects the water quality of the two rivers at the confluence

(117.583867, 39.28545) located in the Qilihai Reservoir on the east bank of the New Chaobai River, the west bank is home to the Beihuaidian Locks and the Tianjin Ancient Coast and Wetlands, which are of great value for water quality assessment.

(117.491631, 39.445053) located in the residential area of the New Chaobai River Basin, it can reflect the water quality condition of the residential area.

(117.478695, 39.480052) is located at the confluence of the New Chaobai River and the Qinglongwan Minus River, which is an important water quality monitoring site.

(117.420506, 39.567221) on the north side of the Chaori Bridge, close to the centre of Baodi district.

(117.285544, 39.692033) located in the Chaobai River National Wetland Park, in the heart of a major industrial park, this is a sampling site worth measuring [6,7]

(4) Sampling Point Group Analysis after Supplementing Sampling Points

The additional sampling sites are evenly distributed along the entire New Chaobai River, reflecting water quality changes in each section of the New Chaobai River and allowing for spatial comparisons, with sampling environments such as **wetland parks, river confluences, living areas and industrial discharge areas**, fully satisfying the richness of sampling site characteristics and allowing for more comprehensive and diverse sampling data.

6 TREATMENT PLAN AND ADVICE

6.1 Suggestions for improvements in evaluation methods and measures for improvement

Because of the respective strengths and weaknesses, as well as capacity deficiencies of this existing water quality evaluation method and our PCA-AHP two-layer assignment model, it is suggested that a combination of the two evaluation methods should be adopted, i.e. the improvements of the existing evaluation methods mentioned above, could achieve more functionality in water quality evaluation. See the end of Part IV for a more detailed analysis

6.2 Recommendations for sampling point selection

Through our analysis, the existing sampling points have uneven spatial distribution, too few sampling points, lack of representative sampling points for different features, etc. New sampling points should be found according to the analysis in Part V. The specific coordinates are given in Part V. For the problem of missing sampling point data, the efficiency of data collection and collation should be improved to enhance the completeness and validity of the data.

6.3 Recommendations for specific governance directions

After our calculation of the model ammonia, total phosphorus, total nitrogen and fluoride have a greater impact on the water quality status, but the rest of the indicators are second to none. However through the pre-processing of the data and the analysis of the measured data, we have concluded that the **fluoride** content at each sampling point is very stable and has excellent values over the entire time span, so it only needs to be maintained and does not require much to invest in this area for treatment. However, for the three indicators of **total phosphorus, total nitrogen and oxygen and ammonian**, they need to be treated vigorously. Many of the WQ indicators that are excellent in the improved model are rated as Class V water in the existing evaluation system, and are classified as poor V in the "Information List of Water Quality Control Units to be improved during the 13th Five-Year Plan Period", so these indicators still need to be treated with. Although there is no obvious trend in data processing, PH is a very important indicator for water quality for its own sake and should not be neglected in treatment. Some of the hazardous indicators in the Xiaogao Tuo, such as TN and COD, have been on the rise in recent years and need to be treated more intensively. On the whole, the water quality of the New Chaobai River has improved significantly compared with the previous situation, but the water quality is still not optimistic, and the data is still basically at the V and IV level, so it is necessary to raise the degree of attention, and take corresponding measures to treat the symptoms, to precisely implement the actual situation in each area, to take different degrees of treatment for different indicators, and to maximize the treatment efforts while making a reasonable allocation of resources. The maximum efficiency of governance and the maximum benefit to the people.

References

- [1] Zhang Qiang. Comprehensive evaluation of river water quality based on PCA-AHP dimension reduction combined weighting model[D].hebei university, 2020.DOI:10.27103/d.cnki.ghebu.2020.001912.
- [2] Zheng Rongwei, Xu Yanxue, Cheng Jing. Construction of river chief system evaluation system based on AHP [J]. Journal of Zhejiang University of Water Resources and Hydropower, 2020, 32(04):25-29.
- [3] Hu Xiaolei, Wang Yizhen. Application of Analytic Hierarchy Process (AHP) in Evaluation of Demonstration Rivers and Lakes [J]. Water Resources Development and Management, 2020(09):77-81.DOI:10.16616/j.cnki.10-1326/TV.2020.09.15.
- [4] MA Kai, LUO Yi, TIAN Yiyuan, YANG Wei, FU Xianting, LI Shuhan. Water Quality Evaluation of Baiyangdian River Based on Principal Component Analysis [C] // 2021 Annual Conference on Science and Technology of Chinese Society of Environmental Sciences - Environmental Engineering Technology Innovation and Application Branch Field Papers (□), 2021:20-25+32.DOI:10.26914/c.cnkihy.2021.022029.
- [5] Zhu Yongjun, Wu Qiong, Zhan Zhongyu. Water quality assessment based on principal component analysis and artificial neural network coupling model [J]. Jiangsu water conservancy, 2021(08):48-54.DOI:10.16310/j.cnki.jssl.2021.08.014.
- [6] Jiang Hao, An Yanling, Wu Qixin, Luo Jin, Peng Wenbo. Study on spatial distribution of water quality in Chishui River Basin during dry season [C] // Abstract of papers of the 14 th Annual Conference of the Chinese Society of Mineral and Rock Geochemistry., 2013:436.
- [7] Research on Intelligent Analysis and Application of Online Monitoring Data of Water Ecological Environment in Three Gorges Reservoir Area [D]. University of Chinese Academy of Sciences (Chongqing Institute of Green Intelligent Technology, Chinese Academy of Sciences), 2018.