Sensitivity Analysis of River Chlorophyll-a to Environmental Factors Based on Support Vector Machine Regression Algorithm in Northern China

Li Xu (xuli0031@163.com)
Hebei University of Architecture
Guizhen Hao
Hebei University of Architecture
Simin Li
Hebei University of Engineering
Peiran Guo
Hebei University of Engineering
Fengzhi Song
Linyi University
Yong Zhao
Hebei University of Architecture

Research Article

Keywords: Water environmental factor, Eutrophication, Chlorophyll-a, Sensitivity analysis

Posted Date: May 4th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1554031/v1

Abstract

The outbreak of planktonic algae seriously affects the water environmental quality of the artificially reformed rivers in the north of China, and it is difficult to control. In this study, taking a typical small river dam area in the north of China—Qingshui River dam area in Zhangjiakou City as an example, the spatiotemporal variation characteristics of environmental factors were analyzed firstly. On this basis, the Chlorophyll-a (Chl-a) prediction model was established by using the Support Vector Regression (SVR). Further, the sensitivity of Chl-a was analyzed, moreover, we confirmed the main limiting factors of Chl-a. The results showed that in 2018, the average content of Chl-a was 126.25ug/L, while the content of DO and pH was also totally high. The maximum content of TN was 16.68mg/L and high in the whole year; the maximum value of CODmn was 50.00mg/L, which was slightly higher in summer; the average value of NH_4^+-N and TP was only 0.78mg/L and 0.18mg/L. The eutrophication evaluation showed that the proportion of areas with hyper eutrophication is 73%, and the most severe eutrophication was in summer. We used the RBF kernel function SVR model and 10-fold cross-validation method to optimize the parameters. The penalty parameter *c* is 1.4142, the kernel function parameter *g* is 1, and the training and verification errors were only 0.032 and 0.067. The application effect was good. Based on the sensitivity analysis of SVR prediction model, the sensitivity coefficients of Chl-a were TP, WT, DO, pH, TN, NH+ 4-N in turn. According to the analysis of the present water pollution situation, TP is the limiting factor of Chl-a in Qingshui River with 0.571 sensitive coefficient and 33% contribution rate, and it is also the main prevention and control factor of planktonic algae outbreak.

Highlights

1. The water quality of qingshui river is poor and the eutrophication is serious.
2. Using the SVR to establish a prediction model of Chlorophyll a is effectivt.
3. Sensitivity analysis and water pollution showed the limiting factor of Chl-a is TP.
4. Sensitivity coefficient of TP is 0.751,and its contribution rate is 33%.

1. Introduction

Most cities in northern part of China are built by rivers. Rivers are often an important part of people's life and entertainment. Therefore, there are more and more phenomena that urban inland rivers are artificially reformed into landscape water bodies (Li Xu, 2019). However, due to the unreasonable reconstruction of many rivers and increasing human interferences, the self-purification capacity is reduced with the enhancement of river closure and poor hydraulic condition. Some common human interferences include: the hardening of urban inland rivers, the construction of bank protection and the setting of rubber dams etc. (Hammock et al., 2019) Finally, the planktonic algae in the rubber dam area grew rapidly, the water quality deteriorated seriously, and even local algal blooms occurred (B. Wu et al., 2022). In recent years, the related departments of phytoplankton control have also invested a lot of manpower and material resources to carry out the prediction and sensitivity analysis of Chl-a in artificially reformed rivers in northern part of China and to explore the limiting factors of planktonic algae growth scientifically. It has a great significance for water pollution prevention and control and early warning of local algae outbreak (Bonsdorff, 2021; Kahru et al., 2020; Zaragüeta and Acebes, 2017; Zou et al., 2020).

In this study, a typical northern man-made urban inland river—Qingshui River in Zhangjiakou City was taken as the research object. The relationship between environmental factors and the indicative index (Chl-a) of planktonic algae was analyzed through sampling and monitoring for one year. On this basis, some prevention and control factors of planktonic algae were explored by a prediction model of Chl-a, which was established by the SVR algorithm, in order to provide a theoretical basis for scientifically and effectively reducing the risk of algal blooms.

Zhangjiakou City is located in the core area of water conservation of the capital in the northwest of Hebei Province. It belongs to a semiarid zone with 330–400 mm annual precipitation and a serious shortage of water resources. The average annual total water resources are 1.799 billion m3, of which the surface water resources are only 1.162 billion m^3 (Hongwei P et al., 2020). Qingshui River is a tributary of Yang River which belongs to Yongding River System. It rises from the southern foot of Huapi Mountain in Chongli County. It is an important sub-basin of Yongding River Basin and the primary source of Yongding River Basin. It is a multi-functional water body such as drinking and entertainment, and it is also an important watershed in the core area of water conservation in the capital (Lei W et al., 2019). 31 rubber dams have been built in Qingshui River in urban area. The total length of the rubber dams is 2,315km. The maximum annual upstream supply is only 50,000 m^3 per day. It is a typical northern artificially reformed river, and it is also an ideal research object of Chl-a prediction and sensitivity analysis.

2. Material And Methods

2.1 Layout of sampling sites and Index Detection

The environmental factors and Chl-a in rubber dam area of Qingshui River were continuously monitored from March to November 2018 (no sampling in winter because of freeze). 11 monitoring points were set up, sampling twice a month (once in the middle of the month and once at the end of the month). The specific location of the sampling point is shown in Fig. 1 The stainless-steel flip river sampler was used to collect subsurface water (0.5 m underwater). 1000 ml polyethylene bottle was used for water sample storage. The clean sampling bottle were able to sample after three times of rinsing. The whole process of transportation and preservation was in a 4 °C light-proof incubator, and the collected water samples were tested within 24 hours. Among them, water temperature (WT), PH, dissolved oxygen (DO) was measured by American HACH portable water quality meter. Total nitrogen (TN), ammonia nitrogen (NH_4^+-N) and total phosphorus (TP) were determined using the spectrophotometry according to environmental quality standards of China for surface water (GB3838-2002). Permanganate index (CODmn) was determined through the permanganate method according to national standard, and Chl-a was analyzed by repeated freeze-thaw ethanol extraction. The correlation coefficients “r” of calibration curves of environmental factors experimental monitoring were all greater than or equal to 0.999, the accuracy of detection data met the requirements of national standards, and the chemicals used in the tests were all guaranteed reagent.

2.2 Data analysis and statistics

Excel 2010 was used to collate and count the monitoring data of environmental factors. Origin8.0 was applied to draw the spatiotemporal variation curve of environmental factors. The Pearson correlation between Chl-a and environmental factors (selecting input factors of SVR, merging verbosity factors and eliminating noise factors) were analyzed using SPSS18.0.

The universal index equation by a logarithmic power function was the way to assess the current situation of regional eutrophication. Compared with other evaluation methods, this way can better consider the comprehensive pollution of components in water with more applicable and objective.(Li Z et al., 2010) Calculation formula:

$$\mathrm{EI} = 10.77 \times \sum_{j=1}^{n} W_j \times \left(\ln X_j\right)^{1.1826}$$

$$X_j = \begin{cases} \left(C_{jo}/C_j\right)^2 C_j \leq C_{jo}(indexDO) \\ 1 C_j > C_{jo}(indexDO) \\ C_{jo}/C_j C_j \geq C_{jo}(exceptDO) \\ 1 C_j < C_{jo}(exceptDO) \end{cases}$$

In the formula, W_j is the normalized weight value of environmental factor j. X_j represents the "standard value" of environmental factor j. C_{jo} is the extremely poor reference value of environmental factor j (value refer to Table 1-a). C_j describes the measured value of environmental factor j. According to the size of EI comprehensive index, water eutrophication is divided into five categories: oligotrophic, mesotrophic, eutrophic, hyper eutrophic, extreme eutrophic, as shown in Table 1-b.

Table 1. (a) reference values of eutrophication indicators; (b) Classification standard of eutrophication of water body

(a)

Project	DO (mg/L)	NH_4^+-N (mg/L)	TN (mg/L)	TP (ug/L)	COD_{mn} (mg/L)	Chl-a (ug/L)
C_{jo}	40	0.01	0.02	1.00	0.12	0.4

(b)

Nutrition degree	EI				
	EI≤20	20<EI≤39.42	39.42<EI≤61.29	61.29<EI≤76.28	76.28<EI
Grading standard	oligotrophic	mesotrophic	eutrophic	hyper eutrophic	extreme eutrophic

2.3 SVR and Sensitivity Analysis

In this study, the SVR, which is widely used in other fields, was the approach to build a Chl-a prediction model, moreover, the sensitive factors of Chl-a in the model were analyzed. It further expanded the application of traditional SVM in the analysis of Chl-a concentration prediction model (Azimi-Pour et al., 2020; B. Wu et al., 2022; C.-C. Wu et al., 2022). The basic research approaches are as follows: first, the measured data are normalized, and the input variables are determined by linear correlation. Second, the kernel function is confirmed as the RBF kernel function, and the optimal parameters C and g are selected by 10-fold cross-verification method. Finally, 80% of the measured data are used for model training, 20% of the measured data are participated in model verification, and finally the optimal regression function of Chl-a is solved.

The sensitivity analysis of Chl-a was based on the partial derivative method, and this model was used to reflect the influence of the change of environmental factors on Chl-a. The influence degree was measured by the sensitivity coefficient. Expression of sensitivity analysis:

$$K_t^i = [f\left(x_1, \cdots, x_i + \Delta x, x_n\right) - f\left(x_1, \cdots, x_n\right)] / \Delta x$$

$$K^i = \sqrt{\sum_{t=1}^{t} \left(K_t^i\right)^2 / t}$$

In the formula: K_t^i is the sensitivity coefficient of environmental factor *i* to Chl-a in t-th sample. $f\left(x_1, \cdots, x_i + \Delta x, x_n\right)$ is the predicted value after the addition of the n-th environmental factor. $f\left(x_1, \cdots, x_n\right)$ is the output value of t sample. K^i is the whole sample sensitivity coefficient of environmental factor *i*.

3. Result And Discussion Coefficient

3.1 Current status of water quality in the study area

The present situation of water environmental factors in Qingshui River dam area of Zhangjiakou City is shown in Table 2. The pH values are between 7.33 and 10.35, and the average value is 8.76. It shows that the Qingshui River dam area is an alkaline water body which is related to the proliferation of planktonic algae in the study area and the consumption of CO_2 by photosynthesis and the production of O_2 and OH^- (Ding et al., 2020). The content of DO and pH are overall at a high level which is between 5.31~22.00mg/L, with an average value of 13.11mg/L. Most water bodies exceed the natural DO content (8 ~ 9mg/L), therefore, it belongs to supersaturated dissolved oxygen water bodies. It also related to the growth state of planktonic algae in the study area. NH+ 4-N and TN are the main indicative indexes of nitrogen content in water. NH+ 4-N concentration in Qingshui River dam area is 0.01~3.89mg/L; the TN concentrations are 1.34~16.68mg/L. The average concentration is 0.78mg/L (NH+ 4-N) and 5.87mg /L(TN). TN concentration was much higher than NH+ 4, indicating that the contents of nitrate nitrogen and organic nitrogen in water are relatively high. The concentrations of TP range from 0.01 to 0.46mg/L with an average value of 0.18mg/L. The content of phosphorus in water varies greatly. COD_{mn} content is the amount of oxidant which consumed by reducing substances, which is an important and rapid measured organic pollution parameter. COD_{mn} in Qingshui River water sample are 0.10~50.00mg/L with an average value of 9.26mg/L which indicates that the content of organic matter in the water was high and the water quality was poor. Chl-a is an indicative index of planktonic algae, which directly reflects the abundance of algae in the water. The maximum value of concentration of Chl-a is 746.12ug/L with an average value of 126.25ug/L. The overall content is obviously at a high level, and the abundance value of planktonic algae is slightly large, mainly due to the poor hydrodynamic conditions and strong sealing of artificially reformed rivers in the north of China. Meanwhile, it also reflects that the external physical interference of planktonic algae growth in artificially reformed rivers in the north of China is weak, and it is strongly affected by the environmental factors of the water body itself.

According to the reference value of Class III water in the Chinese Surface Water Quality Standard (China GB3838-2002), the overproof rate of pH was 30%, in particular, it was 50% in summer. The overproof rate of NH+ 4-N and TP is 20% and 11%. And the maximum value exceeds the standard by 3.89 times (NH+ 4-N) and 1.15 times (TP), which is acceptable. The overproof rate of CODmn is 67%. In spring and summer, this rate is more than 70%, and the highest multiples of the over standard content is more than 8 times. The TN content of all tested samples exceed the standard. The average multiple of the over standard content is more than 5 times. From the overproof situation of the single environmental factor, the overall water quality of Qingshui River was somewhat poor.

Table 2. present situation of water quality in Qingshui River Dam area

Statistic quantity	Period of time	PH	DO (mg/L)	NH_4^+ (mg/L)	TN (mg/L)	TP (mg/L)	COD_{mn} (mg/L)	Chl-a (ug/L)
Average value	Spring	7.59	12.60	0.78	4.67	0.12	6.17	75.23
	Summer	9.08	13.59	0.80	4.39	0.12	11.27	231.82
	Autumn	8.38	11.65	0.66	7.32	0.09	8.84	64.19
	Synthesis	8.76	13.11	0.78	5.87	0.18	9.26	126.25
MAX	Spring	**10.35**	**22.00**	1.92	11.15	**0.46**	13.00	215.04
	Summer	10.34	**22.00**	2.82	10.09	0.36	**50.00**	**746.12**
	Autumn	9.79	17.26	**3.89**	**16.68**	0.23	38.29	218.71
MIN	Spring	8.08	7.74	0.46	1.49	0.05	1.40	3.49
	Summer	7.96	7.04	0.09	1.34	0.03	0.80	3.83
	Autumn	7.33	5.31	0.01	2.01	0.01	0.10	0.82
Over-standard rate %	Spring	42	0	30	100	27	72	—
	Summer	50	0	18	100	9	75	—
	Autumn	13	0	15	100	2	42	—
	Synthesis	32	0	20	100	11	67	—
Quality standard of surface water		6~9	≥5	≤1.0	≤1.0	≤0.2	≤6	—

The concentration values of environmental factors in different locations during the same period are able to reflect the change law of spatial scale, as shown in Fig.2.. In different seasons, the spatial scale change of NH+ 4-N concentration is obvious which increases gradually along the water flow direction with an average increase of 2 to 3 times. The spatial scale change of TN concentration is also evident, and the trend is decreasing along the water flow direction, with an average decrease of 2 to 3 times. On this basis, it can be indicated, that the process of the nitrogen migration with the water flow is slow, and the nitrogen have a degradation trend. A part of organic nitrogen and nitrate nitrogen had a trend of transformation to ammonia nitrogen. The change of TP and COD_{mn} content in water body had a uniform law to a certain extent, but the change along the water flow direction was not obvious. TP had a slightly decreasing trend. And the relative change trend of COD_{mn} has no significant law.

3.2 Analysis on the present situation of eutrophication in the study area

The universal index equation by a logarithmic power function is based on the eutrophication classification standard of rivers, lakes and reservoirs in China. And it is optimized by particle swarm optimization algorithm after standardized treatment. Therefore, the equation can be widely applied to the eutrophication evaluation in lakes, reservoirs, surface water bodies and any type of water body in China (Li Z et al., 2010; Xia Z. and Chen Y., 2020). Eutrophication evaluation is able to comprehensively reflect the current situation of water pollution (Cunha et al., 2021). According to the monitoring data of environmental factors in Qingshui River dam area in 2018, the evaluation results are shown in Table 3. The eutrophication of water body in Qingshui River dam area of Zhangjiakou city is serious. The whole monitoring area is at the eutrophic level and hyper eutrophic level all the time, and the proportion of areas at the hyper eutrophic level is 73%. Among them, the grade of eutrophication is the most serious in summer, only S1 sampling site is at the eutrophic level, and all the other sampling sites are at the hyper eutrophic level. The proportion of hyper eutrophic regions in spring and autumn was 72% and 55%. In the eutrophication level of the sampling area, S2 and S8 were at the hyper eutrophic level during the whole period of time. The other sampling areas showed the hyper eutrophic level in two seasons. The grade of eutrophication of Qingshui River in Zhangjiakou City was overall serious, the problem was widely distributed, and there may be a risk of further deterioration. We consider the reasons are as follows: (1) this section of water body belongs to man-made modified water body, which is strongly closed and strongly affected by human activities, moreover, the water pollution is more serious (Tedengren, 2021); (2) there is a recharge of renewable water body in the upstream of this section of water body, and the nutrient content is higher than natural water body, resulting in the rapid accumulation of nutrients in small closed water bodies downstream, which can easily lead to the aggravation of water eutrophication (Hartshorn et al., 2016). Thus, it can be seen that, establish a suitable prediction model according to the strong closure of man-made rivers in the north of China. Locate accurately the sensitive factors which affects eutrophication, furthermore, reduce the exogenous input and endogenous release of sensitive factors. It is helpful to scientifically solve the serious problem of eutrophication of man-made rivers in the north of China.

Table 3. results of eutrophication assessment

Project	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11
Spring	62.23	64.17	63.13	62.03	61.09	61.23	60.42	62.74	65.17	64.57	63.34
Summer	58.98	63.33	62.34	64.39	63.73	67.31	66.99	64.62	69.25	67.75	66.09
Autumn	63.53	61.42	60.93	60.34	62.94	64.00	63.09	62.16	59.23	61.28	55.51
Standard	oligotrophic		mesotrophic		eutrophic		hyper eutrophic		extreme eutrophic		

3.3 Establishment of Chl-a prediction model and sensitivity analysis of Chl-a

Chl-a is an indicator of planktonic algae outbreak, which is an important discriminant factor of water eutrophication (Chen et al., 2020). The complex environmental factors in the water body are the main factors that affect the concentration of Chl-a. Using environmental factors to predict Chl-a accurately, the linear system is difficult to realize perfectly. In contrast, it is more reasonable to use nonlinear system to predict the concentration of Chl-a in water. In this study, we used the more mature SVR in other fields to build the Chl-a prediction model, and analyzed the sensitive factors of Chl-a in the model(Na et al., 2022).

3.3.1 Determining the input variables of the prediction model by correlation

In the process of building the SVR prediction model, use environmental factors as input variables to predict Chl-a. How to select reasonable environmental factors as input variables is particularly important to the accuracy of the SVR prediction model (Soranno et al., 2020). By using the correlation analysis among environmental factors, the noise factor (the absolute value of correlation with Chl-a is less than 0.1) was eliminated, and the verbosity factor was merged (the absolute value of correlation between environmental factors was greater than 0.8). And then reduced the influence of dependent input variables on the prediction model.

The correlation of environmental factors in water in Qingshui River rubber dam area of Zhangjiakou City is shown in Table 4. There is a significant correlation between Chl-a and WT, DO, NH+ 4-N, TP. The correlation coefficients are 0.483, 0.460, 0.335 and 0.316, respectively, the absolute values are all greater than 0.1. It shows that the four environmental factors have a strong correlation with Chl-a and meet the requirements of model input. Chl-a is significantly correlated with pH and TN, the correlation coefficients are 0.640 and-0.551, respectively, and the absolute values were all greater than 0.1. It indicates that pH and TN had a certain correlation to Chl-a, which reached the requirements of model input. However, the linear correlation shows no significance of pH, TN and Chl-a, which may be due to the relatively complex factors affecting pH value in water and the high content of TN. There is a weak correlation between Chl-a and COD_{mn}, and the correlation coefficient is only 0.009, which belongs to the noise factor input by the model. It may be due to the fact that the contribution of COD_{mn} in water only comes from algae organic compounds with Chl-a content. And the linear correlation between COD_{mn} and Chl-a content is weak, so COD_{mn} cannot be used as the input variable of the model. It can also be seen from the table that the correlation coefficients between environmental factors are not greater than 0.8. It indicates that there is no strong overlap between environmental factors as input variables of the model. According to the linear correlation, it could be judged that the input variables of the SVR prediction model were WT, pH, DO, NH+ 4-N, TN and TP.

Table 4. correlation coefficient and significance between Chl-a content and environmental factors in Qingshui River rubber dam area

	WT	pH	DO	NH+ 4-N	TN	TP	COD	Chl-a
WT	1	0.511**	0.03	0.076	-0.482**	0.035	0.339**	0.483**
pH		1	0.458**	0.138	-0.687**	0.253**	0.337**	0.640*
DO			1	0.054	-0.228**	0.224**	0.065	0.460**
NH+ 4-N				1	-0.231	0.118	-0.028	0.335**
TN					1	-0.26**	-0.354**	-0.551*
TP						1	0.32	0.316**
COD_{mn}							1	0.009
Chl-a								1

Note：*P<0.05, Significant correlation；**P<0.01, Extremely significant correlation（Two-sided test significance）

3.3.2 Model parameter optimization

There are two very important parameters c and g (gamma) in the SVR model, in which the penalty parameter c is the tolerance to the error and affects the fitting end point of the model. The kernel function parameter g is a parameter in the RBF kernel function formula, which determines the number of support vectors after the data is mapped to the new feature space. Therefore, to establish the SVR prediction model with the best fitting effect, it is necessary to adjust the parameters to reach the optimal value.

K-fold cross-validation is an important method for generalization ability estimation in machine learning, and it has unbiasedness (Jian S et al., 2019). Finally, it is obtained that the k-fold cross-verification error is the average value of k times generalization errors, and the prediction model selects the parameters c and g by 10-fold cross-validation comparison method. The optimization results of specific parameters are shown in Fig.3(a), from which the optimal parameters can be seen: the penalty parameter c is 1.4142; the kernel function parameter g is 1; and the mean square error is only 0.01975, which meets the allowable error of model parameter optimization.

3.3.3 Model building and application

In the process of modeling, firstly, the data set was randomly divided into training samples (80%) and verification samples (20%). The training samples were trained and studied through SVR, and then a model with good fitting effect was established. The remaining verification sample data were able to use for model verification. It can be seen from Fig.3(b.c) that the relative error between the real value and the predicted value of the model training sample is only 0.032. The relative error between the real value of the verification sample and the predicted value is 0.067. It indicates: (1) the fitting effect of the prediction model of SVR is ideal;(2) the input variables of the model are reasonable;(3) the selection of kernel function parameters and penalty parameters meet the requirements of the prediction model;(4) the method of using SVR to predict Chl-a of man-made rivers in the north of China is feasible and effective.

3.3.4 Chl-a sensitivity analysis

The prediction model of Chl-a is established to analyze the degree of change of Chl-a which is indicative index of planktonic algae density, caused by the change of environmental factors. The method based on partial derivative was able to use to analyze the sensitivity of a single sample, however, the particularity of the environmental factor of a single sample was strong, so it was difficult to reflect the sensitivity of Chl-a in entire Qingshui River. Therefore, this study made a comprehensive calculation of all the tested samples to determine the sensitivity coefficient of Chl-a to each environmental factor, and further analyzed the sensitivity factor of Chl-a in Qingshui River. The percentage of the total sample sensitivity coefficient of a single environmental factor to the sum of all sensitivity coefficient is the contribution of environmental factors to the growth rate of Chl-a. And this contribution rate can accurately reflect the relative contribution of a certain environmental factor to the growth of Chl-a.

Fig.4. Sensitivity coefficient and contribution rate of environmental factors to Chl-a

As can be seen from Fig.4, the sensitivity coefficient of Chl-a to TP and WT is 0.571 and 0.394, respectively, the contribution rate is 33% and 22%. Followed by DO and pH, the sensitivity coefficient is 0.281 and 0.243, and the contribution rate is 16% and 14%. The sensitivity to TN and NH+ 4-N is the lowest, the sensitivity coefficient is only 0.104 and 0.166, and the contribution rate is 6% and 9%. Planktonic algae photosynthesis produces a large amount of organic matter, absorbs CO_2 and releases a large amount of O_2, which increased the content of DO in water, and then Chl-a is sensitive to DO. The pH value also affects the metabolic activity of planktonic algae, but the sensitivity is common and the contribution rate is at a low level. WT as a factor

that changes with the environmental temperature, the increase of temperature directly affects he activity of algae enzymes, so it is highly sensitive. However, because WT is an external environmental factor, which is difficult for human intervention. When the growth of WT at a certain node causes a large change in Chl-a, the content of limiting factors should be reduced to control the reproduction rate of planktonic algae. TN and NH+ 4-N are necessary nutrients for the growth and reproduction of planktonic algae; however, their sensitivity and contribution are weak. According to the analysis of the current situation of water quality in Qingshui River, the content of TN and NH+ 4-N in Qingshui River were both high, therefore, it belongs to nitrogen content rich type. It is very difficult for nitrogen fluctuation to affect the growth of planktonic algae in the water body which contains lots of nitrogen. Thus, TN and NH+ 4-N are the influential contributing factors of eutrophication in the river.

TP, as another necessary nutrient for the growth and reproduction of planktonic algae, has the greatest contribution to the growth rate of Chl-a, indicating that TP is the strongest sensitive factor of Chl-a. According to the analysis of the current situation of water quality in Qingshui River, the concentration proportion of TP was slightly lower, which contributed greatly to the growth rate of Chl-a. It indicates that the fluctuation of TP content in Qingshui River had the greatest influence on the indicative index of planktonic algae, i.e., Chl-a. And Chl-a restrict the growth and reproduction of planktonic algae directly. When a large amount of phosphorus enters the river water body, it may lead to the proliferation of planktonic algae and the rapid deterioration of eutrophication-condition. Therefore, through the SVR prediction model combined with the analysis of the present situation of water quality, TP is the sensitive factor and limiting factor of Chl-a in Qingshui River, which is the prevention and control factor of planktonic algae outbreak.

4. Conclusion

(1) First, the water quality of Qingshui River in Zhangjiakou City was poor, and the average content of Chl-a was 126.25mg/L, which reflected the abundance of planktonic algae and affected the overall content of DO and pH. The content of TN exceeded the standard in the whole year. The maximum value was 16.68mg/L, and attenuated obviously along the water flow direction. The content in autumn was higher than in the other seasons. Second, the annual over-standard rate of CODmn was 67%, and the maximum value was 50.00mg/L, which in summer was slightly higher than the others, and the spatial change was not obvious. Third, compared with NH + 4-N and TP, the average value was only 0.78mg/L, 0.18mg/L. The NH + 4-N content was higher in spring and increased significantly along the water flow direction. The seasonal change of TP was not obvious, therefore, decreased lightly along the water flow direction.

(2) The universal index formula by logarithmic power function was used to evaluate the eutrophication of Qingshui River. The results showed that the eutrophication of Qingshui River reached extremely high levels, and the whole monitoring area was a at the level of eutrophication and hyper eutrophication all the time. The proportion of areas at the level of hyper eutrophication was 73%, among which the degree of eutrophication in summer was the most serious.

(3) The prediction model of Chl-a of artificially reconstructed water bodies in the north of China was established with the method of SVR. After normalization at first, there was a weak correlation between Chl-a and COD_{mn} through correlation analysis. COD_{mn} belonged to noise factor and was not used as the input variable of the model. Then the parameter optimization of 10-fold cross-verification method was as the way to determine that the penalty parameter c was 1.4142 and the kernel function parameter g was 1.4142. Finally, the training samples and verification samples were studied by RBF kernel function SVR. The results showed that the error was small, therefore, the training and verification errors were only 0.032 and 0.067 respectively.

(4) The sensitivity of Chl-a to environmental factors was analyzed based on SVR prediction model. The results showed that Chl-a was the most sensitive to TP and WT, followed by DO and pH, and the lowest to TN and NH + 4-N. According to the present situation of water pollution, the content of TN and NH + 4-N in Qingshui River was high, which belonged to the type of extremely rich in nitrogen. In the water with high nitrogen content, nitrogen fluctuation was difficult to affect the growth of planktonic algae. TP was the

sensitive factor and limiting factor of Chl-a in Qingshui River, furthermore, it was also the prevention and control factor of planktonic algae outbreak.

Declarations

Supplementary Information Not applicable.

Data Availability All data generated or analyzed during this study are included in this manuscript. The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

Author Contributions Li Xu: Methodology, Sample collection, Formal analysis, Writing - original draft. **Guizhen Hao:** Data curation, Resources. **Simin Li:** Writing - review & editing. **Peiran Guo**: Software, Draft Editing, Formal analysis. **Fengzhi Song**: Review and Editing. **Yong zhao**: Sample collection.

Funding This study is funded by the National Nature Science Foundation of China(No.51508149) and the Basic scientific research business fee of Hebei Provincial Department of Education(ZD2021316 2021QNJS01).The authors also acknowledge the management of Hebei key Laboratory of Water Quality Enginerring and Comprehensive Utilization of Water Resources for the facility provided during the study.

Ethics approval Not applicable**.**

Consent to participate Not applicable.

Consent for publication All the authors reviewed and approved the manuscript for publication.

Competing interests The authors declare no competing interests.

References

1. Azimi-Pour M, Eskandari-Naddaf H, Pakzad A (2020) Linear and non-linear SVM prediction for fresh properties and compressive strength of high volume fly ash self-compacting concrete. Constr Build Mater 230. https://doi.org/10.1016/j.conbuildmat.2019.117021
2. Bonsdorff E, Eutrophication (2021) : Early warning signals, ecosystem-level and societal responses, and ways forward.Ambio50,753–758. https://doi.org/10.1007/s13280-020-01432-7
3. Chen B, Liu H, Xiao W, Wang L, Huang B (2020) A machine-learning approach to modeling picophytoplankton abundances in the South China Sea. Prog Oceanogr 189:102456. https://doi.org/10.1016/j.pocean.2020.102456
4. Cunha DGF, Finkler NR, Lamparelli MC, Calijuri MdoC, Dodds WK, Carlson RE (2021) Characterizing Trophic State in Tropical/Subtropical Reservoirs: Deviations among Indexes in the Lower Latitudes. Environ Manage 68:491–504. https://doi.org/10.1007/s00267-021-01521-7

5. Ding X, Guo X, Zhang C, Yao X, Liu S, Shi J, Luo C, Yu X, Yu Y, Gao H (2020) Water conservancy project on the Yellow River modifies the seasonal variation of Chlorophyll-a in the Bohai Sea. Chemosphere 254:126846. https://doi.org/10.1016/j.chemosphere.2020.126846
6. Hammock BG, Moose SP, Solis SS, Goharian E, Teh SJ (2019) Hydrodynamic Modeling Coupled with Long-term Field Data Provide Evidence for Suppression of Phytoplankton by Invasive Clams and Freshwater Exports in the San Francisco Estuary. Environ Manage 63:703–717. https://doi.org/10.1007/s00267-019-01159-6
7. Hartshorn N, Marimon Z, Xuan Z, Cormier J, Chang N-B, Wanielista M (2016) Complex interactions among nutrients, chlorophyll-a, and microcystins in three stormwater wet detention basins with floating treatment wetlands. Chemosphere 144:408–419. https://doi.org/10.1016/j.chemosphere.2015.08.023
8. Hongwei P, Jia Y, Hongjuan z (2020) Characteristics and driving forces of runoff evolution in Qingshui River Basin under changing environment. South-to-North Water Transfers and Wate 18:1–13. https://doi.org/10.13476/j.cnki.nsbdqk.2020.0022
9. Jian S, Qi L, Mingqiang C, Long Ren (2019) Optimization of Madel for Identification of Oil/Gas and Water Layers While Drilling Based on Machine Learning. J Xi'an Shiyou University(Natural Sci Edition) 34:79–85
10. Kahru M, Elmgren R, Kaiser J, Wasmund N, Savchuk O (2020) Cyanobacterial blooms in the Baltic Sea: Correlations with environmental factors. Harmful Algae 92. https://doi.org/10.1016/j.hal.2019.101739
11. Lei W, Ting L, Jianzhi X (2019) Study on the Effect of Different Land Use Scenarios on Runoff in Qingshuihe Basin of Zhangjiakou Based on SWAT Model. Res Soil Water Conserv 26:245–251. https://doi.org/10.13869/j.cnki.rswc.2019.04.037
12. Li, Xu (2019) Prediction and Sensitivity Analysis of Chlorophyll a in Dam Area of Small and Medium-sized Riversin North China. Dissertation, Hebei University of Architecture
13. Li Z, Wang J, Guo C (2010) A universal index formula for eutrophic evaluation using a logarithmic power function. Acta Sci Circum 30:664–672. https://doi.org/10.13671/j.hjkxxb.2010.03.025
14. Na L, Shaoyang C, Zhenyan C, Xing W, Yun X, Li X, Yanwei G, Tingting W, Xuefeng Z, Siqi L (2022) Long-term prediction of sea surface chlorophyll- A concentration based on the combination of spatio-temporal features. Water Res 211. https://doi.org/10.1016/j.watres.2022.118040
15. Soranno PA, Cheruvelil KS, Liu B, Wang Q, Tan P, Zhou J, King KBS, McCullough IM, Stachelek J, Bartley M, Filstrup CT, Hanks EM, Lapierre J, Lottig NR, Schliep EM, Wagner T, Webster KE (2020) Ecological prediction at macroscales using big data: Does sampling design matter? Ecological Applications 30. https://doi.org/10.1002/eap.2123
16. Tedengren M (2021) Eutrophication and the disrupted nitrogen cycle. Ambio 50:733–738. https://doi.org/10.1007/s13280-020-01466-x
17. Wu B, Dai S, Wen X, Qian C, Luo F, Xu J, Wang X, Li Y, Xi Y (2022) Chlorophyll-nutrient relationship changes with lake type, season and small-bodied zooplankton in a set of subtropical shallow lakes.

Ecol Ind 135:108571. https://doi.org/10.1016/j.ecolind.2022.108571

18. Wu C-C, Chiu C-C, Fu S-Y (2022) A Risk Classification System for Elderly Falls Using Support Vector Machine. Biomedical Engineering: Applications, Basis and Communications. https://doi.org/10.4015/S101623722250017X
19. Xia Z, Chen Y (2020) Three water quality evaluation methods applied to water quality analysis and comparison of Jiang'an River. Environ Dev 32:15–18. https://doi.org/10.16647/j.cnki.cn15-1369/X.2020.04.007
20. Zaragüeta M, Acebes P (2017) Controlling Eutrophication in A Mediterranean Shallow Reservoir by Phosphorus Loading Reduction: The Need for an Integrated Management Approach. Environ Manage 59:635–651. https://doi.org/10.1007/s00267-016-0815-y
21. Zou W, Zhu G, Cai Y, Vilmi A, Xu H, Zhu M, Gong Z, Zhang Y, Qin B (2020) Relationships between nutrient, chlorophyll a and Secchi depth in lakes of the Chinese Eastern Plains ecoregion: Implications for eutrophication management. J Environ Manage 260:109923. https://doi.org/10.1016/j.jenvman.2019.109923

Figures

Figure 1

Layout of sampling sites

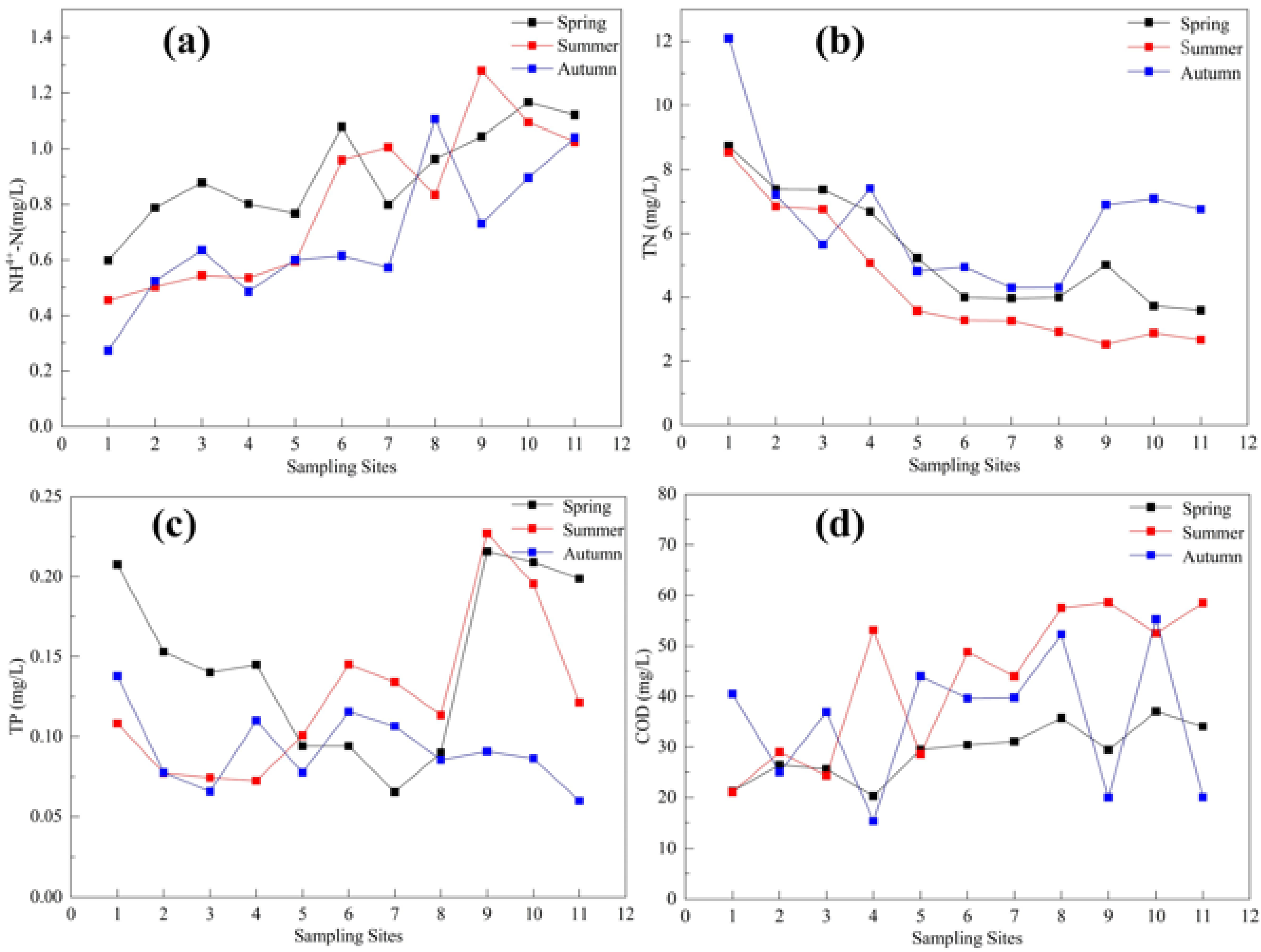

Figure 2

Spatial variation of NH+ 4-N(a), TN(b), TP(c) and COD(d)

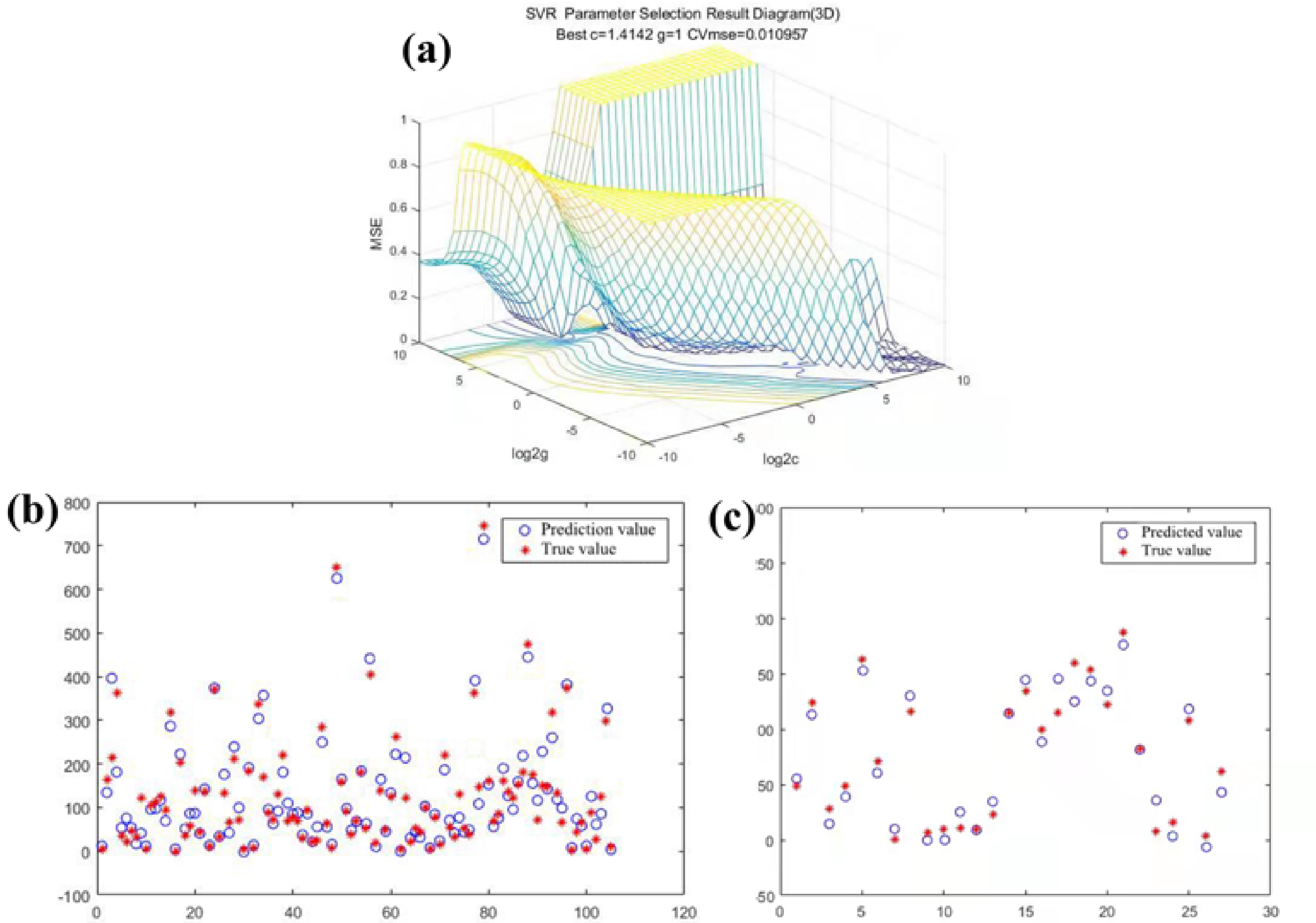

Figure 3

(a) 3D view of SVR parameter optimization result; (b) Error between the real value and the predicted value of the model training sample; (c)Error between the real value of the verification sample and the predicted value

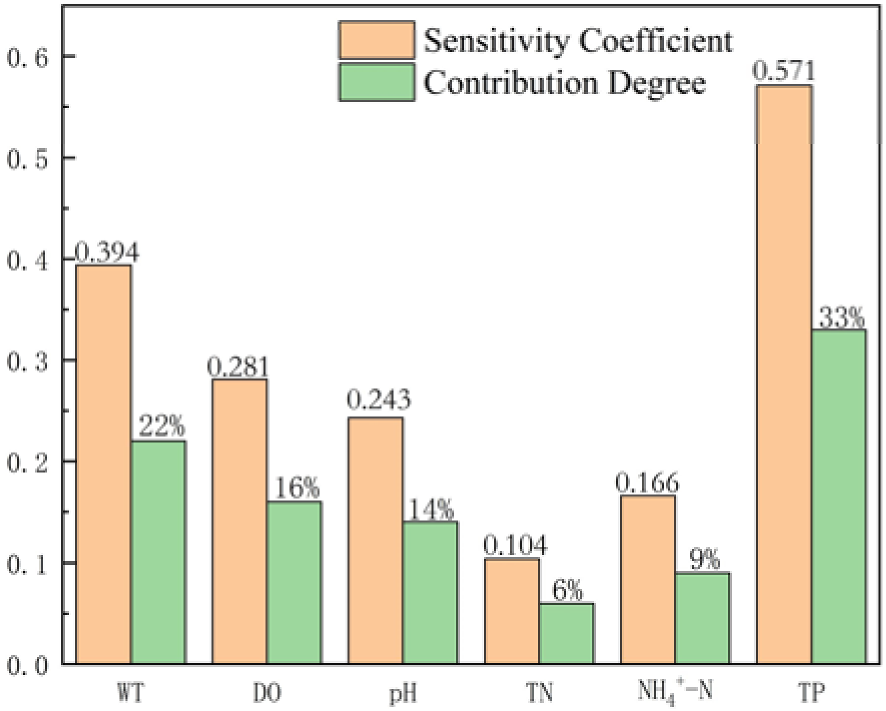

Figure 4

Sensitivity coefficient and contribution rate of environmental factors to Chl-a

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Graphicalabstract.png