

Research Article

Study on Water Quality Inversion Model of Dianchi Lake Based on Landsat 8 Data

Jiaju Cao, Xingping Wen , Dayou Luo, and Yinlong Tan

Faculty of Land Resource Engineering, Kunming University of Science and Technology, Kunming 650093, China

Correspondence should be addressed to Xingping Wen; wfxyp@qq.com

Received 29 May 2022; Revised 29 July 2022; Accepted 1 August 2022; Published 16 September 2022

Academic Editor: Nives Galić

Copyright © 2022 Jiaju Cao et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Efficient, comprehensive, continuous, and accurate monitoring of organic pollution in lakes can provide a reliable basis for water quality assessment and water pollution prevention. This paper takes Dianchi Lake as the research object, aiming at the four important water quality indexes of permanganate index (COD), dissolved oxygen (DO), hydrogen ion (pH), and ammonia nitrogen (NH₃-N); based on the correlation analysis of Landsat 8 data and measured water quality data, an inversion model is constructed to obtain the spatial distribution of the four indexes. The results show that the relative errors of permanganate index (COD) in neural network and multiple regression are 9.68% and 17.48%, respectively; 3.81% and 3.36% in dissolved oxygen (DO); 1.25% and 1.58% in hydrogen ion (pH); in ammonia nitrogen (NH₃-N), it is 15.39% and 24.97%, respectively. The lowest COD in the study area is 6.2 mg/L and the highest is 9.8 mg/L; in 2018, the DO is 5.81 mg/L at the lowest and 9.05 mg/L at the highest; the lowest pH is 5.9 mg/L, the highest is 8.54 mg/L, and the lowest NH₃-N is 0.22 mg/L, the highest is 0.41 mg/L. The inversion results of the overall pollutant concentration in the study area are consistent with the actual situation, with only some slight deviations in some areas. The two inversion models can effectively monitor the water quality and spatial distribution of Dianchi Lake. The remote sensing inversion model of water quality has the value of in-depth research and promotion.

1. Introduction

At present, industrial production, livestock and poultry breeding, and human daily life have produced a large number of organic pollutants, which have seriously polluted the surrounding water bodies and endangered aquatic organisms. With the gradual amplification of the food chain, it has gradually affected human health [1–4]. Therefore, it is of great significance for water quality evaluation and inland water pollution prevention to grasp the pollution degree of organic pollutants in time and accurately.

The traditional water resources monitoring is mainly carried out by manual sampling and testing, which is mainly based on fixed-point and profile sampling. The method has a large workload, a high cost and is time-consuming and laborious. With the development of earth observation technology, remote sensing inversion has gradually become an important way of water resources and water environment monitoring. The basic principle of remote sensing inversion

of water quality is to use appropriate remote sensing image bands to build an inversion model for qualitative or quantitative evaluation of water quality. Due to the large coverage of remote sensing, this method is mostly used for water quality inversion monitoring in large-area waters.

With the development of remote sensing technology, the work of using remote sensing to monitor water quality information is gradually carried out. The types of pollutants that can be retrieved from satellite images are also greatly increased, and the retrieval accuracy is continuously improved. The retrieval methods of water quality parameters mainly include an empirical method [5], a semiempirical method [6], and an analytical method [7]. At present, many algorithms have been applied to remotely sense the inversion of water quality parameters [8, 9]. Based on the research on the inversion of surface water environmental parameters from remote sensing data and measured data, the inversion models of water quality parameters such as statistical regression and neural network are established [10]. Deshpande

et al. took the Jui dam of Jalna District, Maharashtra as the study area and considered pH, total soluble solids, total hardness, total alkalinity, calcium, magnesium, sodium, potassium, chlorine, sulfate, nitrate, and fluoride. The weighted arithmetic water quality index method is used to find the surface water quality index [11]. Artificial neural network [12–16] has the ability of Distributed Association, self-learning, and self-organization. It is often used to mine data relationships and build prediction models. In recent years, it has been widely used in the field of water quality parameter inversion and water quality evaluation. Scholars at home and abroad have made some progress in the application of neural network to water environment: Li et al. combined particle swarm optimization (PSO), chaos theory, adaptive strategy, and back propagation artificial neural network (BP ANN), and proposed a new Weihe River water quality evaluation model [17]. Woo Kim Young et al. used the machine learning (ML) algorithm to extract Chla from MSI. The model interpretation and spatial change of Chla within and between lakes confirmed the effectiveness of lgbm in retrieving the derived Chla from MSI from lakes and estuaries [18]. Chen Jinyue et al. used the machine learning method (GA ANN) combining genetic algorithm and artificial neural network to retrieve Chla concentration [19]. Chen Zhu et al. made correlation analysis between MODIS remote sensing data and measured chlorophyll a concentration [20]. Based on multitemporal remote sensing images and field observation data, Yanhu et al. used back propagation (BP) neural network to establish an inversion model for water quality parameters of inland reservoirs, [21]. Huang Jingjing et al. used machine learning algorithm to establish a remote sensing model for Shenzhen Bay [22].

At present, the research on chlorophyll a concentration, water temperature, suspended solids, salt content, etc. is relatively mature, while the research on other water quality indicators is relatively few [23–25]. Chemical oxygen demand (COD) is the amount of oxidant required to oxidize the reducing substances in water samples when measuring water samples by chemical methods [26]. In the research and management of river pollution caused by domestic wastewater and industrial wastewater, chemical oxygen demand (COD) is one of the important indicators that can be measured quickly. Ammonia nitrogen is a nutrient in the water and one of the main oxygen-consuming pollutants in the river. Excessive ammonia nitrogen content will lead to eutrophication of the water body, which is not conducive to the healthy growth of aquatic organisms. Nonaqueous color water quality parameters, such as ammonia nitrogen, hydrogen ion, permanganate index, and dissolved oxygen, are also important indicators to measure water quality. The high concentration of nitrogen and phosphorus in water will directly lead to water eutrophication, which is the main problem in most of the water environment at present.

Multiple regression model and artificial neural network method were used to establish remote sensing inversion models of four water quality parameters: permanganate index (COD), dissolved oxygen (DO), hydrogen ion (pH), and ammonia nitrogen ($\text{NH}_3\text{-N}$). Through the comparison and analysis of the inversion accuracy between the multiple

regression model and the neural network model with parameters such as relative error and correlation coefficient, the inversion of water quality using remote sensing technology has the characteristics of strong information synthesis ability, fast acquisition speed, time saving and labor saving, and can better reflect the spatial distribution characteristics of water quality. Artificial neural network is a large-scale parallel nonlinear dynamic system. Further in-depth analysis of the inversion of water quality parameters, and establishing a rigorous, scientific and efficient water quality monitoring system belonging to Dianchi Lake are required as soon as possible, so as to provide a guarantee for the solution of increasingly frequent water resource shortages, water environment pollution, and flood disasters.

2. Study Area and Data

2.1. Overview of the Study Area. Kunming, Yunnan Province is located at $102^{\circ} 10' - 103^{\circ} 40'$ E and $24^{\circ} 23' - 26^{\circ} 22'$ N. Located in the southwest border of China and the middle of Yunnan Guizhou Plateau, the city center is 1895 meters above sea level, surrounded by mountains on three sides and Dianchi Lake on the south. The Dianchi Lake Reserve in the study area is located in the southwest of Kunming City, Yunnan Province (Figure 1). It is the largest freshwater lake in Yunnan Province, located in the plain area of the central Yunnan basin, with 34 rivers flowing into it. The average elevation of the lake surface is about 1886 m, the average water depth of the whole lake is 5.3 m, the length of the lake shoreline is about 163 km, the Lake area is 309.5 km^2 (when the water level is 1887.4 m), and the water storage capacity is 1.56 billion cubic meters. In the north of the lake, there is a ridge across the east and west, which divides the lake into open sea (298.7 km^2) and Caohai (10.8 km^2). Dianchi Lake is a lake of earthquake fault subsidence type. With low latitude, high altitude, constant temperature, and long water change cycle, the rainfall in Dianchi Lake Basin is divided into rainy season and dry season, where the rainy season is from May to October and the dry season is from November to April. The physical and chemical environment of Dianchi Lake water body is constantly changing with the development of Dianchi Lake pollution, and the ecological environment quality of Dianchi Lake Basin is only at the medium level.

2.2. Data Collection and Processing. On the premise of ensuring data quality, this paper selects Landsat 8 images and Dianchi Lake measured points as the research data. Remote sensing images are from the United States Geological Survey (USGS) (<https://glovis.usgs.gov/>) The measured data are from China national environmental monitoring station. Ten Landsat 8 images with little or no cloud from 2013–2018 were selected.

In order to objectively and truly reflect the surface information and reduce the errors caused by data factors, the remote sensing image data are preprocessed such as radiometric calibration, atmospheric correction, image clipping, etc. (1) Radiometric calibration: the process of converting the gray value of the remote sensing image

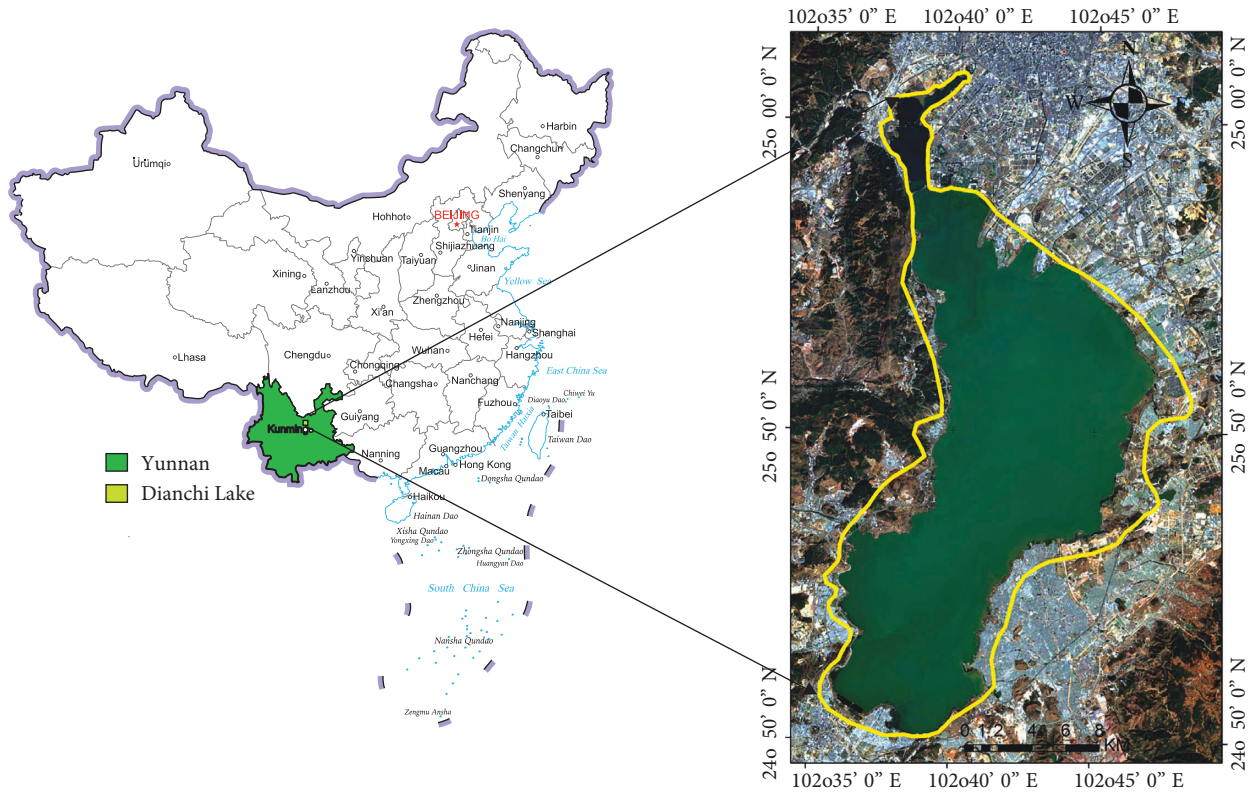


FIGURE 1: Study area.

recorded by the sensor into the radiance of the entrance pupil of the sensor. (2) Atmospheric correction: using the hyperspectral image header file to obtain various correction parameters required, obtaining the reflectivity through the atmospheric radiative transfer equation, and then using the spectral smoothing technology to remove the noise left in the atmospheric correction process. (3) Crop: removing areas outside the study. When acquiring the emissivity data of the sampling point, due to force majeure factors such as weather and climate, the data will change a little and will need to be calculated so that the data can be used normally.

The measured data used in the experiment come from the observation data of buoy stations along the coast of Dianchi Lake and the results of subsequent processing in the laboratory. The distribution of buoy stations is shown in Figure 1. The stations are evenly distributed in the Dianchi Lake area, thus ensuring the integrity and scientificity of water environment monitoring in the Dianchi Lake Basin. There are 36 groups of measured data (Table 1). After removing some abnormal points, 24 groups of data are selected to establish the inversion model, and the other 8 groups are used to test the accuracy of the model. The inversion models of permanganate (COD), dissolved oxygen (DO), hydrogen ion (pH), and ammonia nitrogen ($\text{NH}_3\text{-N}$) were constructed by multiple regression and neural network. The accuracy of the two remote sensing inversion models is tested and compared by using the test sample points.

The purpose of water quality remote sensing monitoring is to reverse study the relationship between water quality concentration information contained in water reflection

TABLE 1: Collection time of sample points.

Date	Number of sample points
April 20, 2013	2
February 2, 2014	2
January 4, 2015	4
January 9, 2017	4
February 13, 2018	4
April 2, 2018	4
May 22, 2013	2
April 23, 2014	2
November 22, 2016	4
January 28, 2018	4
March 1, 2018	4

information and remote sensing total radiation information obtained by satellite sensors. If we can make full use of the water body information captured by remote sensing satellites, it must be helpful to improve the accuracy of water quality inversion.

3. Establishment of Remote Sensing Inversion Model for Water Quality Parameters

3.1. Multiple Regression Model. Multiple regression model is a statistical analysis method to study the linear or nonlinear relationship between a dependent variable and one or more independent variables. The regression model determines the causal relationship between variables by specifying dependent variables and independent variables, establishes the regression model, estimates the parameters of the model

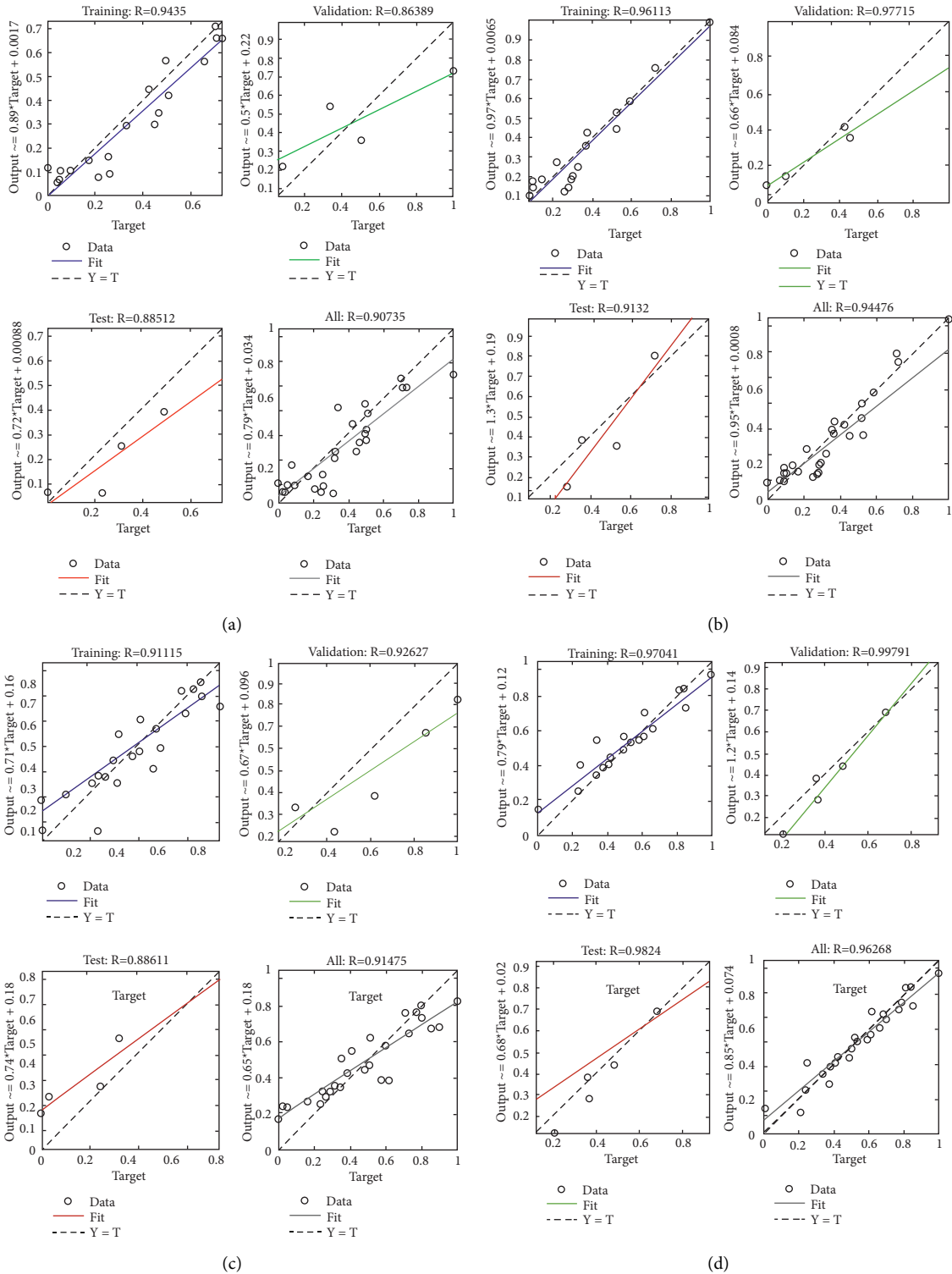


FIGURE 2: Schematic diagram of network regression analysis of various water quality parameters.

according to the measured data, and then evaluates whether the regression model can fit the measured data well. Further prediction can be made according to independent variables.

The study of multiple linear regression model belongs to a completely accidental correlation, and there is no physical relationship between Landsat satellite bands and variables. Therefore, although the program is scalable, the results still have limitations, which can be applied to Landsat satellites and lakes in the study.

3.2. Artificial Neural Network Model. The full name of neural network is artificial neural network (ANN). It uses physically realizable devices or computers to simulate some structures and functions of neural networks in biology, and is applied to engineering fields. The focus of neural network is not to use physical devices to completely replicate the neural cell network in the organism, but to extract the available parts to overcome the problems that cannot be solved by the current computer or other systems such as learning, control, recognition, and expert system.

Using the measured water quality parameters and Landsat 8 remote sensing image data, a three-layer neural network model including input layer, hidden layer, and output layer is constructed. The input layer is the single band and band combination data of remote sensing reflectance extracted from remote sensing images, and the output layer is the water quality parameter concentration data. Through experimental comparison and analysis, the number of neurons in the hidden layer is set. The number of neurons in different hidden layers will directly affect the accuracy of the model. After training, the training results of each group tend to be stable and meet the error requirements, and the error performance change, training situation, and linear regression fitting are better. The fitting of each part of the sample is shown in Figure 2. Several groups of highly correlated data are used as the input layer of the neural network, and the measured water quality parameters are used as the output layer. The neural network model with appropriate number of neurons in the hidden layer is selected.

4. Results and Analysis

4.1. Multiple Regression Model Analysis. Analyze the correlation between measured data and single band and band combination of remote sensing images [27], and select the data group with good correlation to build a statistical regression model [28]. Compare and analyze the fitting effects of linear, quadratic polynomial, cubic polynomial, logarithm, and other algorithms. List all the functional relations with large R^2 in the models with different factors, and select the fitting model with single band or band combination with high R^2 . These models are tested with 8 sample data used for testing. Finally, the best multiple regression model is selected according to R^2 and error size.

Analyze the correlation between permanganate index (COD), dissolved oxygen (DO), hydrogen ion (pH), and ammonia nitrogen ($\text{NH}_3\text{-N}$). The bands with high correlation of COD concentration in water are B5, B2, and B4,

TABLE 2: Correlation index between COD and band.

Factor	R^2
B5	0.818
B2	0.739
B2 + B5	0.828
B4 + B5	0.666
B5/B2	0.679
$(B2 - B5)/(B2 + B5)$	0.686
$B5/(B2 + B5)$	0.626
B2 - B5	0.453

TABLE 3: Correlation index between DO and band.

Factor	R^2
B7	0.562
B1	0.501
B1 + B7	0.580
B1 - B7	0.631
B1/B7	0.819
$B1/(B1 + B7)$	0.819
$(B1 - B7)/(B1 + B7)$	0.820
B1*B7	0.400

TABLE 4: Correlation index between PH and band.

Factor	R^2
B3	0.647
B4	0.557
B3 + B4	0.654
B3*B4	0.615
B3 - B4	0.401
$B3/(B3 + B4)$	0.415

TABLE 5: Correlation index between $\text{NH}_3\text{-N}$ and band.

Factor	R^2
B6	0.529
B7	0.541
B6 + B7	0.672
B6*B7	0.421

and the band combination is B2 + B5, B4 + B5, B5/B2, $(B2 - B5)/(B2 + B5)$, $B5/(B2 + B5)$ (Table 2); The bands with high correlation of DO concentration in water are B7 and B1, and the band combination is B1/B7, $B1/(B1 + B7)$, $(B1 - B7)/(B1 + B7)$, B1 - B7, B1 + B7, B1*B7 (Table 3); the bands with high correlation of pH concentration in water are B3 and B4, and the band combination is B3 + B4, B3 - B4, B3*B4, $B3/(B3 + B4)$ (Table 4); the bands with high correlation of nh_3n concentration in water are B6 and B7, and the band combination is B6 + B7 and B6*B7 (Table 5). The single band and band combination of these remote sensing images with high correlation with water quality parameters can be used as factors for retrieving water quality parameters and establishing multiple regression models.

According to the band combination factor selected as the model, a multiple regression model between the band combination factor and water quality parameters is

TABLE 6: Fitting model of COD concentration.

Factor	Model	R^2	Functional relation
B5	Linear logarithm	0.818	$y = 0.036x + 1.9692$
B2	Linear logarithm	0.739	$y = 0.0368x - 5.8195$
B2 + B5	Linear logarithm	0.828	$y = 0.0194x - 2.5423$
B4 + B5	Linear logarithm	0.666	$y = 0.0179x - 1.8273$
B5/B2	Cubic	0.679	$y = 124.86x^3 - 143.9x^2 + 69.475x - 5.4311$
$(B2 - B5)/(B2 + B5)$	Cubic	0.680	$y = -235.2x^3 + 322.04x^2 - 158.7x + 34.705$

TABLE 7: Fitting model of DO concentration.

Factor	Model	R^2	Functional relation
B1/B7	Cubic	0.819	$y = 0.0005x^3 - 0.0307x^2 + 0.2222x + 9.299$
$B1/(B1 + B7)$	Cubic	0.819	$y = -13122x^3 + 34968x^2 - 31065x + 9209.4$
$(B1 - B7)/(B1 + B7)$	Cubic	0.820	$y = -1640.2x^3 + 3821.3x^2 - 2969.2x + 778.67$
B1 - B7	Linear logarithm	0.631	$y = -0.0139x + 12.958$
B1 + B7	Linear logarithm	0.580	$y = -0.0143x + 13.99$

established. B2 + B5 is used as the factor for retrieving COD concentration, and the R^2 of the linear model is the best, reaching 0.828. The best model of each factor is tested with the data of eight test stations. Finally, according to R , R^2 , and error, the linear model with B2 + B5 as the factor is selected as the statistical regression model (Table 6). $(B1 - B7)/(B1 + B7)$ is used as the factor for inversion of DO concentration, and the R^2 of the cubic polynomial model is the best, reaching 0.820. After comparison and analysis, the linear model with $(B1 - B7)/(B1 + B7)$ as the factor is selected as the statistical regression model (Table 7). B3 + B4 is used as the factor for inversion of pH concentration, and the R^2 of the linear model is the best, reaching 0.654. After comparison and analysis, the linear model with B3 + B4 as the factor is selected as the multiple regression model (Table 8). B6 + B7 is used as the factor for inversion of $\text{NH}_3\text{-N}$ concentration, and the R^2 of the cubic polynomial model is the best, reaching 0.672. After comparison and analysis, the cubic polynomial model with B6 + B7 as the factor is selected as the multiple regression model (Table 9).

4.2. Analysis of Artificial Neural Network Mode. By analyzing the correlation between single band and band combination and measured water quality parameters, it is concluded that B5, B2, B2 + B5, B4 + B5, B5/B2, $(B2 - B5)/(B2 + B5)$, and $B5/(B2 + B5)$ have high correlation with measured COD concentration, and the model fitting effect is the best when the number of neurons in the hidden layer is 7. The fitting degree of each group of samples is above 0.86, and the model fitting degree is very good (Figure 2(a)). The data of B7, B1, B1/B7, $B1/(B1 + B7)$, $(B1 - B7)/(B1 + B7)$, B1 - B7, B1 + B7, B1 * B7, and 9 groups are highly correlated with the measured DO concentration, and the model fitting degree is the best when the number of neurons is 8. The fitting degree of samples in each group is above 0.91, and the model fitting degree is good (Figure 2(b)). B3, B4, B3 + B4, B3 - B4, B3 * B4, B3/(B3 + B4), and 6 groups of data have high correlation with the

TABLE 8: Fitting model of pH concentration.

Factor	Model	R^2	Functional relation
B3	Linear logarithm	0.647	$y = 0.0026x + 6.8589$
B4	Linear logarithm	0.557	$y = 0.0043x + 6.8119$
B3 + B4	Linear logarithm	0.654	$y = 0.0017x + 6.7299$
B3 * B4	Logarithm	0.615	$y = 0.7254 \ln(x) - 0.4599$

measured pH concentration, and the model fitting degree is the best when the number of neurons is 7. The fitting degree of each group of samples is above 0.88, and the model fitting degree is very good (Figure 2(c)). The data of B6, B7, B6 + B7, B6 * B7, and 4 groups have high correlation with the measured $\text{NH}_3\text{-N}$ concentration, and the model fitting degree is the best when the number of neurons is 6. The fitting degree of samples in each group is above 0.96, and the model fitting degree is very good (Figure 2(d)). Overall, the fitting effect of water quality parameters is good.

4.3. Model Comparison. For better verification, the same test sample for difference and error comparison is selected (Figure 3 and Table 10). In COD (Figure 3(a)), the overall trend of the predicted value of neural network is closer to the measured value than the predicted value of multiple regression. However, the difference of individual points is large. In the average relative error, the neural network is 9.68%, while the multiple regression is 17.48%. The effect of the neural network model is better. In DO (Figure 3(b)), the predicted value of neural network and the predicted value of multiple regression are close to the measured value. In the average relative error, the neural network is 3.31%, while the multiple regression is 3.86%. The effects of neural network and multiple regression models are good. In pH (Figure 3(c)), the predicted value of neural network and multiple regression are close to the measured value. In the average relative error, the neural network is 1.25%, while the

TABLE 9: Fitting model of NH₃-N concentration.

Factor	Model	R ²	Functional relation
B6 + B7	Linear logarithm	0.672	$y = -0.0083x + 0.9465$
B6	Linear logarithm	0.529	$y = -0.0113x + 0.821$
B7	Cubic	0.541	$y = -0.0001x^3 + 0.0108x^2 - 0.3386x + 3.943$

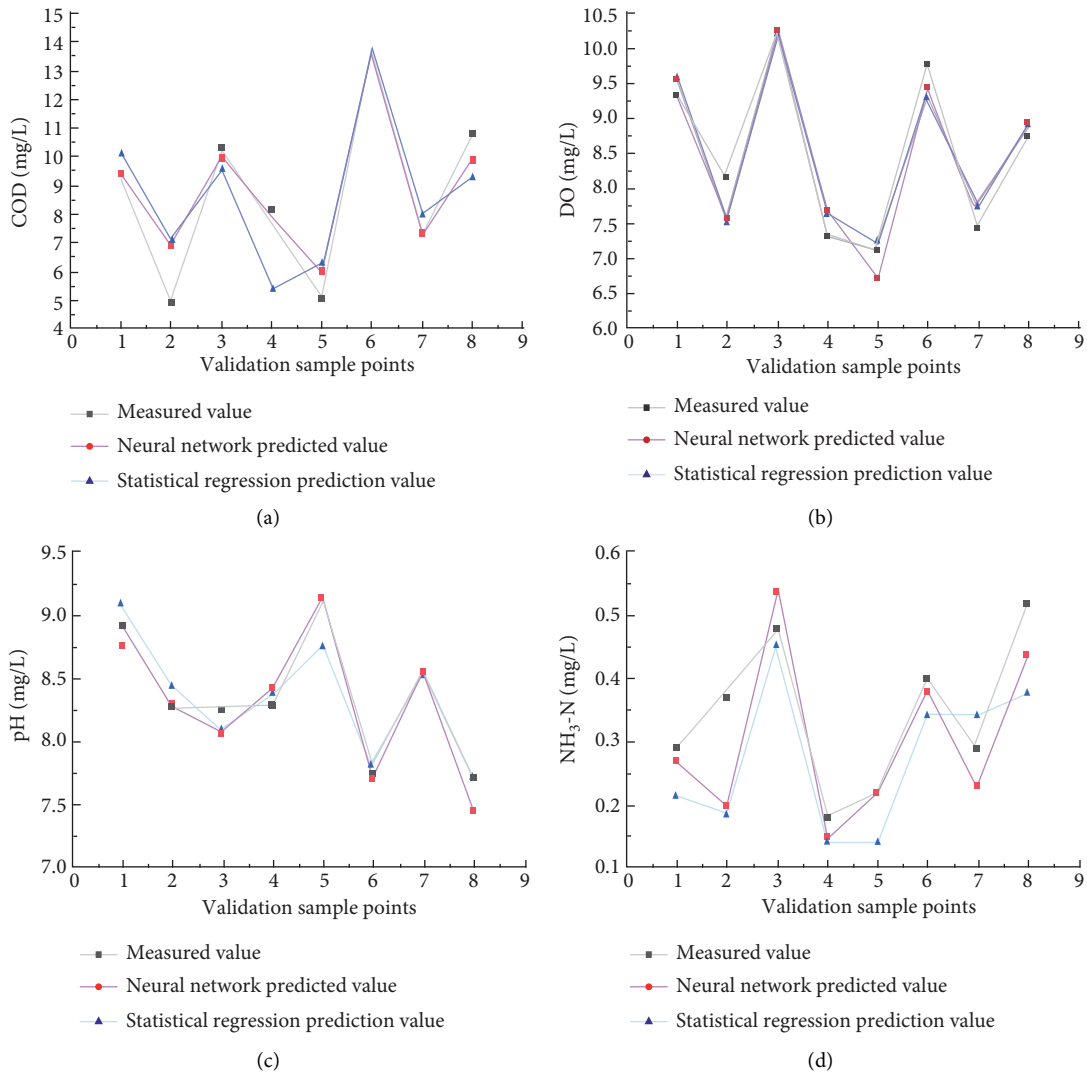


FIGURE 3: Comparison of measured values, neural network predicted values, and multiple regression predicted values.

multiple regression is 1.58%. The effects of neural network and multiple regression models are good. In NH₃-N (Figure 3(d)), the difference between the predicted value of neural network and the predicted value of multiple regression is smaller than the measured value, but the difference between the predicted value and the measured value is larger. The average relative error of neural network is 15.39%, while that of multiple regression is 24.97%. The

effect of neural network model and multiple regression model is general.

The analysis results show that neural network has obvious advantages in monitoring the concentration of pollutants in Dianchi Lake. The COD, DO, and pH concentrations retrieved by this model are in good agreement with the measured data, and the ammonia nitrogen concentration is relatively low. Therefore, the applicability of

TABLE 10: Relative errors of different models.

Water quality parameters	COD		DO		pH		NH ₃ -N	
	Relative error of neural network (%)	Relative error of neural network (%)	Relative error of neural network (%)	Relative error of multiple regression (%)	Relative error of neural network (%)	Relative error of multiple regression (%)	Relative error of neural network (%)	Relative error of multiple regression (%)
	4.25	7.77	2.54	2.83	1.79	1.97	6.90	26.07
	40.41	44.4	7.12	7.64	0.36	2.00	45.95	49.78
	3.01	7.13	0.29	0.13	2.30	2.06	12.5	5.33
	2.72	33.77	4.92	4.32	1.57	1.09	16.67	22.78
	19.8	25.26	6.18	1.40	0	4.25	0	36.41
	1.17	0.18	3.28	4.81	0.39	0.90	5	14.23
	1.62	7.51	4.31	4.01	0.23	0.23	20.69	18.31
	8.33	13.75	1.83	1.71	3.37	0.13	15.38	27.33
Average relative error (%)	9.68	17.48	3.81	3.36	1.25	1.58	15.39	24.97

this model to the retrieval of NH₃-N concentration in Dianchi Lake is limited.

4.4. Distribution Map of Water Quality Parameters. Based on the neural network inversion model with high accuracy and the data information of each band of Landsat 8 image, the concentrations of four water pollutants were inverted to obtain the spatial distribution of COD, DO, pH, and NH₃-N concentrations in the study area in 2013, 2014, and 2018 (Figure 4). The monitoring results show that the COD concentration index shows a downward trend, the small decrease in DO is consistent with the actual situation, the pH and NH₃-N index change little, and may be affected by time, water temperature, etc.

Taking the inversion results of water quality in 2018 as an example. The monitoring results show that the lowest COD of each water body in the study area is 6.2 mg/L and the highest is 9.8 mg/L; DO is 5.81 mg/L at the lowest and 9.05 mg/L at the highest; The lowest pH is 5.9 mg/L and the highest is 8.54 mg/L, and the lowest NH₃-N is 0.22 mg/L and the highest is 0.41 mg/L. The inversion results of the overall pollutant concentration in the study area are consistent with the actual situation, with only some slight deviations in some areas. The concentrations of COD and pH are relatively high. Generally speaking, the water quality in this section is poor, and there are some problems.

The experimental results show that the water quality of Dianchi Lake is moderately polluted, belonging to the first-class surface water source protection zone. However, when the midstream flows through urban and rural residential areas, the discharge of human daily domestic sewage and industrial sewage will increase accordingly, resulting in an increase in COD content, and then the degree of water pollution. In addition, with the application of nitrogen fertilizer in farmland runoff, the content of NH₃-N in the water body will also increase. DO content in water is affected by many factors, such as water temperature, dissolved ions, microorganisms, etc; in

eutrophic water, it is mainly controlled by biological processes. Therefore, when the number of algae rises to a certain order of magnitude, the number of algae and the vigorous degree of life activities will inevitably play a leading role in the change of DO in the water body. The pH in the water body is mainly affected by the CO₂ content. The rapid consumption of CO₂ by algae under photosynthesis increases the pH. It is verified that the actual NH₃-N content in the water body in this area is more than 0.3 mg/L, but the spatial distribution of the monitoring results of the pollutant concentration in Dianchi Lake is not consistent with the actual situation: the concentration is low compared with the actual situation. The possible reasons for the monitoring error of pollutant content in some water bodies are as follows: (1) Although the shooting time of remote sensing images and the sampling time of water quality data are the same day, the time points cannot be accurately matched, and there will always be some differences between the time in the morning and the time in the evening. Objectively, there is a gap in the concentration of pollutants in water bodies flowing through residential areas during this period; (2) Errors caused by atmospheric correction; (3) The error of the measured water quality concentration at the sampling point. For the water quality samples at the sampling point, the error generated when measuring the water quality concentration in the laboratory, as a result, the water quality factor content at the measured sampling point is relatively small; (4) As some rivers are relatively narrow, and the images used in this study are Landsat 8 remote sensing images with a resolution of 30m, the number of pixels in the narrow river reach is small, and the distinction between the river and the bank surface features is not obvious, resulting in mixed pixels, which will affect the DN value of the image, resulting in low monitoring results of water pollution concentration. The evaluation results are similar to the water quality status in the water resources bulletin. The neural network model can monitor the concentration of water pollutants with high accuracy.

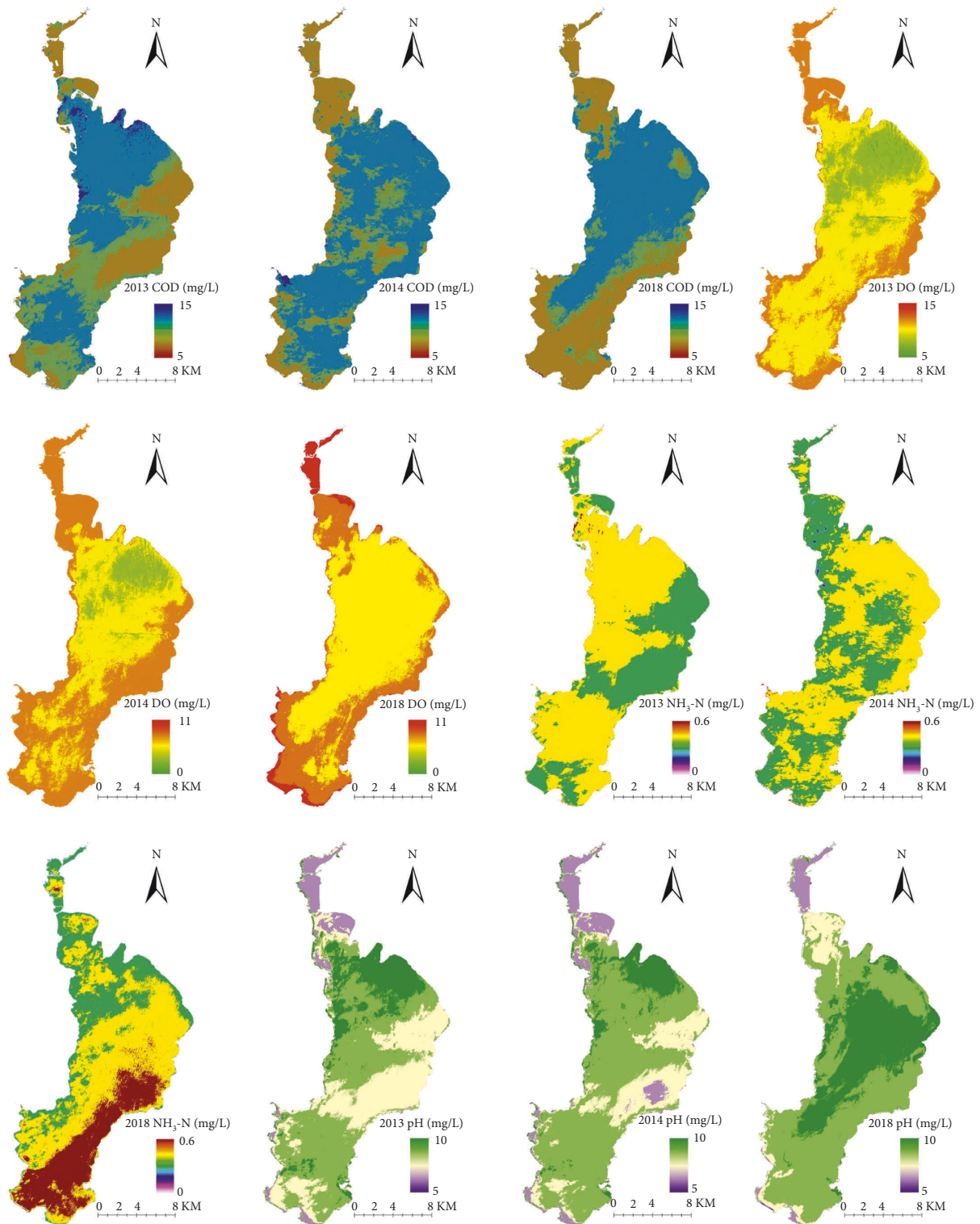


FIGURE 4: Water quality inversion distribution.

5. Discussion and Conclusion

Based on the measured water quality data and synchronous image data, the correlation analysis results show that there is an internal relationship between the spectral reflectance and the concentration of organic pollutants, the spectral

characteristic bands of four important water quality parameters in Dianchi Lake area are found by using the correlation analysis method. The Landsat 8 data of Landsat series are used for band operation. The single band or band combination image data are correlated with the measured water quality factor data. The band or band combination

with high correlation is selected to establish a neural network model for inversion of water quality factors, and the inversion results are compared with the measured data. The results show that it is feasible to use Landsat 8 remote sensing image data to monitor the concentration of water pollutants.

Statistical regression model cannot deal with the nonlinear relationship between data well; Artificial neural network has the ability to deal with complex nonlinear relationships between various interacting variables, and shows good estimation performance. Therefore, the development of artificial neural network model will become a research hotspot in the field of pollution in recent years. In addition, the combination of various neural network models and statistical regression models has become an important means of estimating concentration. The addition of neural network models can effectively improve the accuracy of concentration estimation.

The construction of water quality parameter inversion model needs to consider many factors and inaccurate information processing. In view of this situation, the powerful nonlinear fitting ability of neural network can greatly improve the accuracy of remote sensing inversion model of water quality parameters. At present, neural network algorithm has been widely used in the inversion of water quality parameters. Based on the measured data, combined with the multiple regression model and the inversion model of permanganate index (COD), dissolved oxygen (DO), hydrogen ion (pH) and ammonia nitrogen (NH₃-N) and suspended solids concentration developed by artificial neural network, the temporal and spatial analysis of the inversion results is briefly analyzed. In addition, for the components of inland water bodies, seasonal and regional differences are large. Therefore, it is still necessary to build local relevant models based on the actual situation to provide reasonable data support for local water pollution prevention and control.

Data Availability

Data used to support the findings of this study are included within the article and additional data are available from the author upon request.

Conflicts of Interest

The authors declare that there are no conflicts of interest.

Acknowledgments

The authors gratefully acknowledge the USGS and the Hydrology and Environmental Monitoring of China for providing data used in this study.

References

- [1] Y. Kown, S. Baek, Y. Lim et al., "Monitoring coastal chlorophyll-a concentrations in coastal areas using machine learning models," *Water*, vol. 10, no. 8, p. 1020, 2018.
- [2] K. Dörnhöfer and N. Oppelt, "Remote sensing for lake research and monitoring—recent advances," *Ecological Indicators*, vol. 64, pp. 105–122, 2016.
- [3] T. Yuqiang and Z. Ruihan, "Occurrence and distribution of persistent organic pollutants in water of Chinese lakes and reservoirs," *Journal of Lake Sciences*, vol. 32, no. 2, pp. 309–324, 2020.
- [4] Y. Wang, H. Xia, J. Fu, and G. Sheng, "Water quality change in reservoirs of Shenzhen, China: detection using LANDSAT/TM data," *Science of the Total Environment*, vol. 328, no. 1, pp. 195–206, 2004.
- [5] Y. Park, K. H. Cho, J. Park, S. M. Cha, and J. H. Kim, "Development of early-warning protocol for predicting chlorophyll-a concentration using machine learning models in freshwater and estuarine reservoirs, Korea," *Science of The Total Environment*, vol. 502, pp. 31–41, 2015.
- [6] J. Wang, Y. Zhang, F. Yang et al., "Spatial and temporal variations of chlorophyll-a concentration from 2009 to 2012 in poyang lake, China," *Environmental Earth Sciences*, vol. 73, no. 8, pp. 4063–4075, 2015.
- [7] M. Zoljoodi, M. Moradi, and N. Moradi, "Seasonal and interannual cycles of total phytoplankton phenology metrics in the Persian Gulf using ocean color remote sensing," *Continental Shelf Research*, vol. 237, Article ID 104685, 2022.
- [8] T. W. Cui, J. Zhang, K. Wang et al., "Remote sensing of chlorophyll a concentration in turbid coastal waters based on a global optical water classification system," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 163, pp. 187–201, 2020.
- [9] J. E. Tavora, E. Boss, D. Doxaran, and P. Hill, "An algorithm to estimate suspended particulate matter concentrations and associated uncertainties from remote sensing reflectance in coastal environments," *Remote Sensing*, vol. 12, no. 13, p. 2172, 2020.
- [10] H. Hossen, W. E. Mahmood, A. Negm, and T. Nakamura, "Assessing water quality parameters in burullus lake using sentinel-2 satellite images," *Water Resources*, vol. 49, no. 2, pp. 321–331, 2022.
- [11] S. M. Deshpande, U. S. Bhagwat, and K. R. Aher, "Mathematical computation of weighted arithmetic water quality index of Jui dam of jalna District, Maharashtra," *Bulletin of Pure and Applied Sciences- Geology*, vol. 40, no. 2, pp. 219–226, 2021.
- [12] B. Chen, X. Mu, P. Chen et al., "Machine learning-based inversion of water quality parameters in typical reach of the urban river by UAV multispectral data," *Ecological Indicators*, vol. 133, Article ID 108434, 2021.
- [13] S. Bansal and G. Ganesan, "Advanced evaluation methodology for water quality assessment using artificial neural network approach," *Water Resources Management*, vol. 33, no. 9, pp. 3127–3141, 2019.
- [14] S. Y. Liu, H. J. Tai, Q. S. Ding, D. Li, L. Xu, and Y. Wei, "A hybrid approach of support vector regression with genetic algorithm optimization for aquaculture water quality prediction," *Mathematical and Computer Modelling*, vol. 58, no. 3–4, pp. 458–465, 2013.
- [15] M. Majumder and R. Roy, "A quick prediction of hardness from water quality parameters by artificial neural network," *International Journal of Environment and Sustainable Development*, vol. 17, no. 2/3, pp. 247–257, 2018.
- [16] I. W. Seo, S. H. Yun, and S. Y. Choi, "Forecasting water quality parameters by ANN model using pre-processing technique at the downstream of cheongpyeong dam," *Procedia Engineering*, vol. 154, pp. 1110–1115, 2016.

- [17] M. S. Li, W. Wu, B. S. Chen, L. Guan, and Y. Wu, "Water quality evaluation using back propagation artificial neural network based on self-adaptive particle swarm optimization algorithm and chaos theory," *Computational Water, Energy, and Environmental Engineering*, vol. 06, no. 03, pp. 229–242, 2017.
- [18] Y. Woo Kim, T. H. Kim, J. Shin et al., "Validity evaluation of a machine-learning model for chlorophyll a retrieval using sentinel-2 from inland and coastal waters," *Ecological Indicators*, vol. 137, Article ID 108737, 2022.
- [19] J. Chen, S. Chen, R. Fu et al., "Remote sensing estimation of chlorophyll-A in case-II waters of coastal areas: three-band model versus genetic algorithm-artificial neural networks model," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 14, pp. 3640–3658, 2021.
- [20] Z. Chen, M. Dou, R. Xia, G. Li, and L. Shen, "Spatiotemporal evolution of chlorophyll-a concentration from MODIS data inversion in the middle and lower reaches of the Hanjiang River, China," *Environmental Science and Pollution Research*, vol. 29, no. 25, pp. 38143–38160, 2022.
- [21] Y. He, Z. Gong, Y. Zheng, and Y. Zhang, "Inland reservoir water quality inversion and eutrophication evaluation using BP neural network and remote sensing imagery: a case study of dashahe reservoir," *Water*, vol. 13, no. 20, p. 2844, 2021.
- [22] J. Huang, D. Wang, F. Gong, Y. Bai, and X. He, "Changes in nutrient concentrations in shenzhen bay detected using Landsat imagery between 1988 and 2020," *Remote Sensing*, vol. 13, no. 17, p. 3469, 2021.
- [23] X. Wu, "Study on hydrogeological characteristics of Xinyang city," *Groundwater*, vol. 31, no. 4, 2009.
- [24] J. Yuan, Z. Niu, and X. Wang, "Atmospheric correction of Hyperion hyperspectral images based on FLAASH," *Spectroscopy and Spectral Analysis*, vol. 29, no. 5, 2009.
- [25] J. Ma, *Estimation of Chlorophyll Content of Vegetation Based on Model Simulation and Oli Remote Sensing Images*, Hebei Normal University, Hebei, China, 2014.
- [26] L. Wang, X. Shi, B. Sun et al., "Study on remote sensing inversion of COD concentration in hulun lake based on MODIS data," *Environmental Engineering*, vol. 12, no. 6, pp. 103–108, 2014.
- [27] G. L. Feyisa, H. Meilby, R. Fensholt, and S. R. Proud, "Automated Water Extraction Index: a new technique for surface water mapping using Landsat imagery," *Remote Sensing of Environment*, vol. 140, pp. 23–35, 2014.
- [28] J. E. Escoto, A. C. Blanco, R. J. Argamosa, and J. M. Medina, "Estimation of water quality in Pasig River using empirical ordinary least squares regression model from Sentinel-2 satellite images," the International Archives of the Photogrammetry, *Remote Sensing and Spatial Information Sciences*, vol. XLVI-4/W6, pp. 161–168, 2021.