Desalination and Water Treatment
www.deswater.com

doi: 10.5004/dwt.2023.29473

291 (2023) 20–31
April

Inversion of water quality by remote-sensing monitoring based on machine learning in complex freshwater environments

Shuangshuang Zhou[a,b,c], Min Xiao[a,b,c], Jiaxuan Zhang[a,b,c], Yingping Huang[a,b,d,*], Zhuan Jin[a,b,c], Biao Xiong[a,b,c,*]

[a]*Hubei Engineering Technology Research Center for Farmland Environment Monitoring, China Three Gorges University, Yichang 443002, Hubei, China, emails: chem_ctgu@126.com (Y. Huang), biaoxiong1122@126.com/1173052866@qq.com (B. Xiong), 3517545010@qq.com (S. Zhou), xiaomin2013@163.com (M. Xiao), 859159273@qq.com (J. Zhang), 1533602154@qq.com (Z. Jin)*
[b]*Engineering Research Center of Eco-Environment in Three Gorges Reservoir Region of Ministry of Education, China Three Gorges University, Yichang, 443002, Hubei, China*
[c]*College of Hydraulic and Environmental Engineering, China Three Gorges University, Yichang 443002, Hubei, China*
[d]*College of Computer and Information Technology, China Three Gorges University, Yichang 443002, Hubei, China*

Received 26 September 2022; Accepted 14 March 2023

ABSTRACT

Small-area urban water areas play an important role in the human ecological environment and as a source of drinking water. Most such urban water bodies are long and narrow, so ordinary remote-sensing data sources are not suitable for monitoring. To address this problem, we use the four water-quality metrics of total nitrogen, total phosphorus, turbidity, and algal density as detected from unmanned aerial vehicle (UAV) remote-sensing images. We apply an improved least squares model to evaluate and invert the water quality in the complex freshwater environments of Longhu Lake, which is a tributary of the Yangtze River. These environments combine the polymorphism characteristics of urban rivers. The results show that (1) the band-ratio model is appropriate to estimate the water-quality parameters, and inversion by band combination is more accurate than inversion with a single band. (2) Comparison of the unary linear function, polynomial function, and exponential function with the least squares model shows that the latter produces the best inversion results and the smallest error. In addition, data from different periods are summed to verify the applicability of the model. The inversion results indicate that the elements responsible for polluting city rivers must be further explored by comparing point source pollution with non-point source pollution. The results show that the UAV multispectral estimation model based on the least squares method is accurate and stable and can provide strong support for water-quality monitoring in small areas. This method has important practical significance for improving intelligent and automated water-monitoring technology.

Keywords: Water-quality inversion; Machine learning; Unmanned aerial vehicle remote sensing

---

1. Introduction

Water is one of the most important factors in the natural environment and an indispensable resource for human civilization [1]. The increasingly serious problem of water pollution necessitates a timely, accurate, and comprehensive understanding of the water pollution status of specific water bodies, which is vital to solve the water-pollution problem [2]. The main factors affecting water quality include suspended solids (turbidity), algae, and chemicals

---

* Corresponding authors.

such as nutrients, pesticides, metals, and dissolved organic substances. Water-quality monitoring is the main basis for water-quality evaluation and prevention of water pollution [3]. Conventional water quality monitoring usually takes the form of fixed-point monitoring, which is easily affected by many external factors such as manpower, material resources, and weather, resulting in many restrictions on the time, frequency, and quantity of water sampling that can be done [4,5]. Based on these monitoring data, it is difficult to recreate the actual situation of the entire water body [6]. However, in the 1960s, remote-sensing technology was developed, bringing with it the advantages of low monitoring cost and a large monitoring range, thereby overcoming the limitations of conventional water-quality monitoring. Applying remote-sensing technology to monitor water quality involves detecting spectral signals and then linking these signals to water-quality parameters through experience or by applying a model. Quantitative remote sensing of water quality can be implemented if the data correlate with pollutant concentration. A previous study reported that neither total nitrogen nor total phosphorus directly affect the spectral reflectance, whereas the spectral reflectance correlates statistically with chlorophyll a and suspended matter [7]. At present, the main methods for monitoring water-quality parameters using remote-sensing data include analytical methods, empirical and semi-empirical methods, and water-quality inversion by machine-learning methods. The data source required by the analysis method is difficult to obtain, so the quantitative remote-sensing inversion based on this method is difficult to implement [8]. Thus, the empirical and semi-empirical methods may not be the optimal solution [9]. In contrast, the machine-learning algorithm does not depend on a specific model and improves the complex relationship between independent and dependent variables through the algorithm itself, providing an effective method for inverting water-quality parameters [10].

In addition, machine-learning methods can be used for rapid inversion of water-quality factors [11]. In previous work, an autoregressive neural network was used to predict the water level of Lake Jablanica; the results demonstrated that a single water level datum can be used to accurately predict the water level [12]. Machine learning (such as support vector machines, random forests, and backpropagation neural networks) was applied to construct water-quality inversion models [13–15]. This method has been used in various water quality parameter-inversion models. However, these inversion methods are suitable for single, large-area water bodies. Some researchers have also tried to reverse water quality in micro water areas [16].

In this paper, we study the water-quality parameters and unmanned aerial vehicle (UAV) multispectral images of surface-water samples collected from Longhu Lake, Yidu City, in the middle and lower reaches of the Yangtze River. We use the least squares method and construct an inversion model based on four water-quality parameters: total nitrogen (TN), total phosphorus (TP), turbidity (TUB), and algae density (ACD). The results of the model map the water-quality distribution of the study area.

The main contributions of this paper are as follows: (1) Because the study area involves the middle and lower reaches of the Yangtze River, which finally flow into the Yangtze River, the results provide a theoretical basis for water-quality inversion of the Yangtze River. (2) The combination of UAV remote-sensing data and the machine-learning algorithm provide a basis for evaluating the water quality of small water bodies. (3) Finally, the results provide a reference for selecting the sensitive bands for estimating the water-quality parameters from small areas of water without involving significant optical characteristics (e.g., total nitrogen and total phosphorus).

2. Data and technique

2.1. Study area

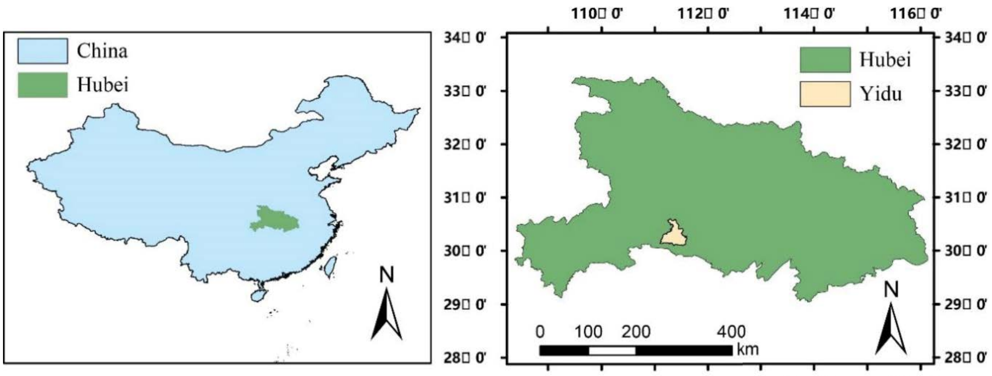
The study area is Longhu (30°23′N,111°30′E), which is in Yidu City, Hubei Province, China. In 2019, this area was included in the pilot project for ecological protection and restoration of the Three Gorges area of the middle and lower reaches of the Yangtze River. Two typical areas of Longhu Lake labeled A and B (Fig. 1c and d, respectively (serve as the research areas). The environments of these two areas differ: Area A consists of two continuous water areas with a narrow water surface surrounded by aquatic plants, mostly reeds. Area B consists of seven continuous ponds, with a wide water surface and no aquatic plants but with surrounding orchards. The entire area is under a strong anthropomorphic influence (Fig. 1).

2.2. Data treatment

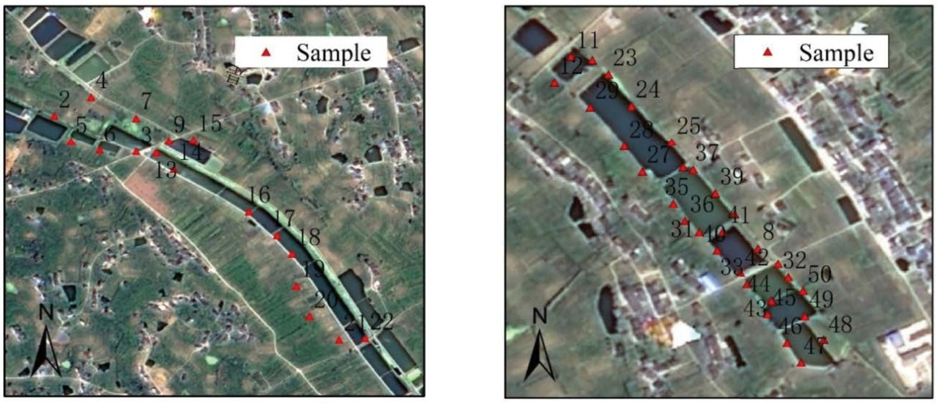
2.2.1. Acquisition and pre-processing of UAV images

The study used a multirotor UAV Ely 4RTK produced by the Dajiang Company (Shenzhen, China) and equipped with six semiconductor sensors (FC6310R) [17]: one visible-light sensor for true-color imaging and five monochromatic sensors for multispectral imaging. These sensors captured five bands, namely, the blue band, green band, red band, red-edge band, and near-infrared band. Table 1 lists the parameters of the multispectral sensors. The bandwidth of each band was 32, 32, 32, 32, and 52 nm, respectively, and was mainly determined by the geometric size of the incident slit of the spectral instrument. The multispectral light intensity sensor on the top of the fuselage detected the real-time incident light intensity in the five bands, thus allowing the imaging to be compensated and obtaining more accurate spectral information [18]. The water areas studied were approximately 6 km long. Due to limitations imposed by battery life, a single mission could fly 1 km. Multiple flight missions were required to obtain complete UAV images of the study area. UAV images were collected from 10:00 to 15:00 from an altitude of 200 m under a clear sky and with low wind speeds [19]. The sensor was set to continuous shooting mode and the data were saved to a SD card. The exposure time was set according to the light intensity.

We used DJI GS Pro ground station software for trajectory planning. By tapping the interactive design on the screen and planning the route mission, the UAV could fly automatically to the waypoint according to the flight mission. The UAV usually flew 1.1 km on each flight while photographing at equal intervals. The flight speed was 14.5 m/s, the flight altitude was 200 m [20], and the route overlap rate



(a)Hubei Province (b) Yidu City



(c) Area A (d) Area B

Fig. 1. Study area.

Table 1
Parameters of multispectral sensors

Unmanned aerial vehicle band	Wavelength
B (blue)	450 ± 16 nm
G (green)	560 ± 16 nm
R (red)	650 ± 16 nm
RE (red edge)	730 ± 16 nm
NIR (near infrared)	840 ± 26 nm

and side overlap were both 75% [17,21,22]. The main course angle was usually obtained automatically by the UAV's perception of the local sunlight and wind speed. After completing the flight mission, the UAV automatically returned to the take-off point.

2.2.2. Collection and detection of water-quality parameters

On-site sampling was implemented with the UAV in flight. The sampling sites were evenly distributed 100 to 150 m apart on both sides of the river. The sampling points in the two study areas totaled 48; Fig. 1 shows the distribution of the sampling points. During sampling, we used the mobile phone software Jiyin Footprint to record the position of the sampling sites. Following the technical guidelines for water-quality sampling, the water samples were collected 0.5 m below the water surface [23]. Once the water samples were collected, their water-quality parameters were obtained in the laboratory within 3 d.

Four water-quality parameters were collected in this experiment: total nitrogen, total phosphorus, turbidity, and algae density. The total phosphorus was determined as per the Water Quality Determination of Total Phosphorus Ammonium Molybdate Spectrophotometry (CB 11893-1989). First, under neutral conditions, the sample was digested by potassium persulfate, and all the phosphorus contained was oxidized to orthophosphate. Next, in an acidic medium, orthophosphate was reacted with ammonium molybdate to form phosphomolybdate heteropoly acid in the presence of antimony salt, which was immediately reduced by adding ascorbic acid to form a blue complex. At a specific wavelength, combined with color reaction, we determined the absorbance and made a calibration curve to obtain the phosphorus content [24]. The determination of total nitrogen was done following the Determination of Total Nitrogen in Water Quality by Alkaline Potassium Persulfate Digestion Ultraviolet Spectrophotometry (HJ 636-2012). The nitrogen-containing substances contained in the water sample were converted at the temperature of the experimental system, which was between 120°C and 124°C. After adding alkaline potassium persulfate solution, it converted into nitrate substances, and the converted test solution was placed in the ultraviolet spectrophotometer for determination. The measurement was made at 220 and 275 nm and judged according to the relative fraction of total nitrogen in the water body and the corrected absorbance, which was measured by using the standard curve method [25]. Turbidity was measured directly by a turbidimeter according to standard HJ 1075-2019 [26], and algae density was measured by using spectrophotometry [27].

2.2.3. Multispectral data processing

Compared with satellite remote-sensing images, the images obtained by multispectral sensors are relatively simple to process, notably because the complex and tedious atmospheric correction process is avoided. The software exported images from the multispectral sensors and performed band operations. The spectral reflectance data are available in [28,29].

We first used Pix4DCapture to splice and crop the multispectral images (Fig. 2), then imported the cropped images into ENVI5.3. Finally, the sampling points on the water surface were found from the latitude and longitude coordinates, allowing us to construct 5 × 5 matrices centered on the sampling points as the region of interest. Three 5 × 5 regions of interest were drawn for each sampling point, and we took the average spectral reflectance of all points in the region as the spectral reflectance of that point.

2.3. Selection and modeling of spectral parameters

To reduce the interference of background information, effective spectral information was extracted. After obtaining from each band the spectral information at the sampling point, we tried a variety of combinations, and then applied a correlation analysis with the measured water body parameters to find the band or band combinations that met the requirements. Table 2 lists the formulas used to combine the bands.

The inversion model of water-quality elements was constructed by combining spectral data and water-quality parameters. Four functional models were established by using MATLAB software: a linear regression model, a power function model, a polynomial model, and a least squares model. The spectral data finally determined after processing and selection were fit with the measurement data, and the corresponding models were generated in each case. The optimal prediction model was selected based on the evaluation criteria of the model. Finally, the optimal model was

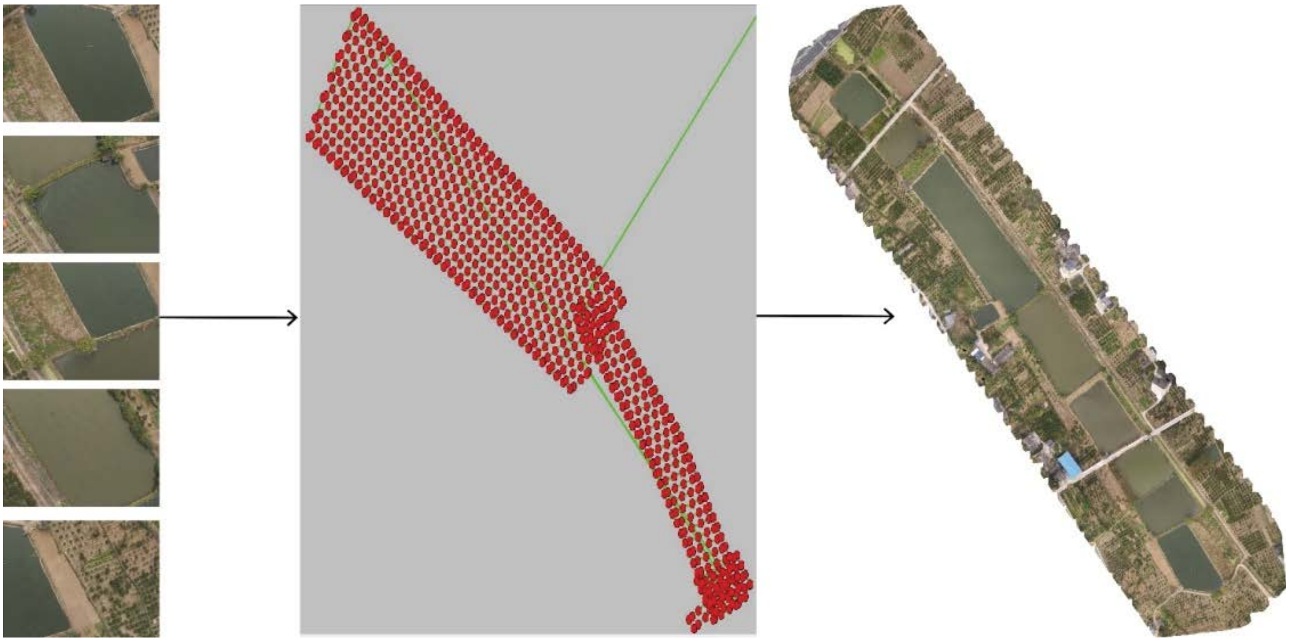


Fig. 2. Stitching unmanned aerial vehicle images.

Table 2
Formulas for combining multispectral bands

Water-quality parameters	Band combination
Total phosphorus	G/RE, R/RE, R/NIR, RE/NIR, (B+G)/NIR, (B+G)/RE, (B+R)/NIR, (B+RE)/G
Total nitrogen	NIR/R, B/R, (B+G)/R, G/R, RE/R, (B+RE)/R, (B+G+RE)/(R+NIR)
Turbidity	B, B+RE, B+NIR, B+RE+NIR, B+R+NIR, G/R, RE/R, NIR/R, (RE+NIR)/(B+G+R)
Algae density	RE/R, B/R, G/R, (B+G)/R, R/B, (B+RE)/R, (B+G+RE)/(R+NIR)

used to invert TN, TP, TUB, and ACD in ENVI5.3 using the acquired multispectral images, and spatial distribution maps were made.

3. Least squares method

3.1. Least squares method

The least squares method is a mathematical optimization technique that finds the best fit to data by minimizing the sum of the squares of the errors [30]. Given n sets of observations (X_1, Y_1), (X_2, Y_2), …, (X_n, Y_n) for n points, the sample regression function fits this set of values as best as possible. The criteria for selecting the best-fit curve minimizes the total fitting error (i.e., the total residual). Sample regression is done as follows:

$$Y_i = \hat{\beta}_0 + \hat{\beta}_1 X_i + e_i \Rightarrow e_i = Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i \tag{1}$$

The residual sum of squares is

$$Q = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right)^2 = \sum_{i=1}^{n}\left(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i\right)^2 \tag{2}$$

A linear fit is determined by minimizing Q. The quantities β_0 and β_1 are variables and are considered to be functions of Q, which creates a problem for finding the extreme value, which is obtained by differentiating. Next, we calculate the partial derivatives of Q with respect to the two variables:

$$\begin{cases} \dfrac{\partial Q}{\partial \hat{\beta}_0} = 2\sum_{i=1}^{n}\left(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i\right)(-1) = 0 \\ \dfrac{\partial Q}{\partial \hat{\beta}_1} = 2\sum_{i=1}^{n}\left(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i\right)(-X_i) = 0 \end{cases} \tag{3}$$

The objective-function solution is:

$$\hat{\beta}_0 = \frac{\sum X_i^2 \sum Y_i - \sum X_i \sum X_i Y_i}{n\sum X_i^2 - \left(\sum X_i\right)^2} \tag{4}$$

$$\hat{\beta}_1 = \frac{n\sum X_i Y_i - \sum X_i \sum Y_i}{n\sum X_i^2 - \left(\sum X_i\right)^2} \tag{5}$$

3.2. Dataset construction and accuracy evaluation

3.2.1. Basic properties of water from two sampling points

Table 3 lists the specific water-quality parameters. We measured and counted the maximum, mean, normal deviation, and coefficient of deviation of the water-quality parameters. The mean is a measure of the central tendency of a dataset, and the standard deviation is estimated based on the sample and reflects the degree of dispersion in the values relative to the mean. The coefficient of variation (CV) is the dimensionless fraction of the normal deviation of the original information with respect to the mean of the original data. The three metrics are calculated as follows:

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} \tag{6}$$

$$\text{Stdev} = \sqrt{\frac{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2}{n-1}} \tag{7}$$

$$\text{CV} = \left(\text{Stdev} \div \bar{x}\right) \times 100\% \tag{8}$$

3.2.2. Model evaluation criteria

The experiment involved 48 water-quality samples, with 75% chosen at random as the training set and the remaining 25% as the validation set. The training set was used to find the best fit and the validation set was used to check the accuracy of the model.

The root mean square error (RMSE), mean absolute error (MAE), and mean absolute percentage error (MAPE) were used to determine the accuracy of the most commonly used water-quality inversion models. The range of values of the three rating indicators of the regression model is [0, +∞), and the RMSE indicates the variation between the predicted value and the quantified value [31]. The smaller the value, the higher the accuracy of the model. The smaller the MAE, the better the model. MAPE is commonly used

Table 3
Basic water properties at two sampling points

Sampling point	Parameter	Max.	Min.	Average	Std. Dev.	Coefficient of variation
Area A	Total nitrogen (mg/L)	0.179	0.031	0.061	0.035	0.564
	Total phosphorus (mg/L)	0.12	0.02	0.055	0.029	0.527
	Turbidity (NTU)	29.2	16	21.666	3.338	0.154
	Algae density (×10^6 cells/L)	1.071	0.236	0.508	0.262	0.516
Area B	Total nitrogen (mg/L)	0.149.	0.04	0.065	0.024	0.372
	Total phosphorus (mg/L)	0.093	0.016	0.058	0.018	0.316
	Turbidity (NTU)	20.4	9.9	15.123	2.485	0.164
	Algae density (×10^6 cells/L)	2.182	0.325	0.838	0.415	0.495

to measure prediction accuracy: the smaller the MAPE, the better the model (greater than 100% indicates a poor model). The three evaluation indicators are calculated as follows:

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}\left|\hat{y}_i - y_i\right| \tag{9}$$

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2} \tag{10}$$

$$\text{MAPE} = \frac{100\%}{n}\sum_{i=1}^{n}\left|\frac{\hat{y}_i - y_i}{y_i}\right| \tag{11}$$

In addition, the association coefficient between forecast and actual value is used to access the forecast results. The closer the correlation coefficient is to unity, the better the correlation between the two data sets. The coefficient of determination R^2 is calculated as follows:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}\left(\hat{y}_i - y_i\right)^2}{\sum_{i=1}^{n}\left(\bar{y}_i - y_i\right)^2} \tag{12}$$

where n is the number of samples, y_i is the actual value, $\hat{y}_i$ is the predicted value, and $\bar{y}$ is the average of the actual values.

4. Results and analysis

4.1. Spectral characteristics of multispectral image regions

Fig. 3a and b show cross-sectional spectra from the UAV multispectral image of areas A and B, respectively. The left sides of Fig. 3a and b show the five channels captured by the UAV. The curves marked by the dotted box on the right are the spectra corresponding to the section of water marked by the dotted box on the left. The multispectral images show that the spectral reflectance of the water body differs significantly from the reflectance of other types of bottom objects; that is, the water body has a low and relatively constant spectral reflectance, which corresponds to the general law of spectral characteristics of bottom objects [16]. The magnitude of spectral reflectance of the cross section in decreasing order is B, G, R, RE, and NIR, so the spectral reflectance gradually decreases with increasing wavelength.

4.2. Data correlation analysis

4.2.1. Analysis of water-quality parameters at sampling points

Table 3 lists the water quality parameters, which show that the normal deviation of the turbidity is largest in study areas A and B, the relative deviation of TN is largest in area A, and the relative deviation of ACD is largest in area B. The coefficient of variation is between 0.164 and 0.564, and the overall variation is moderate. Of the four water-quality

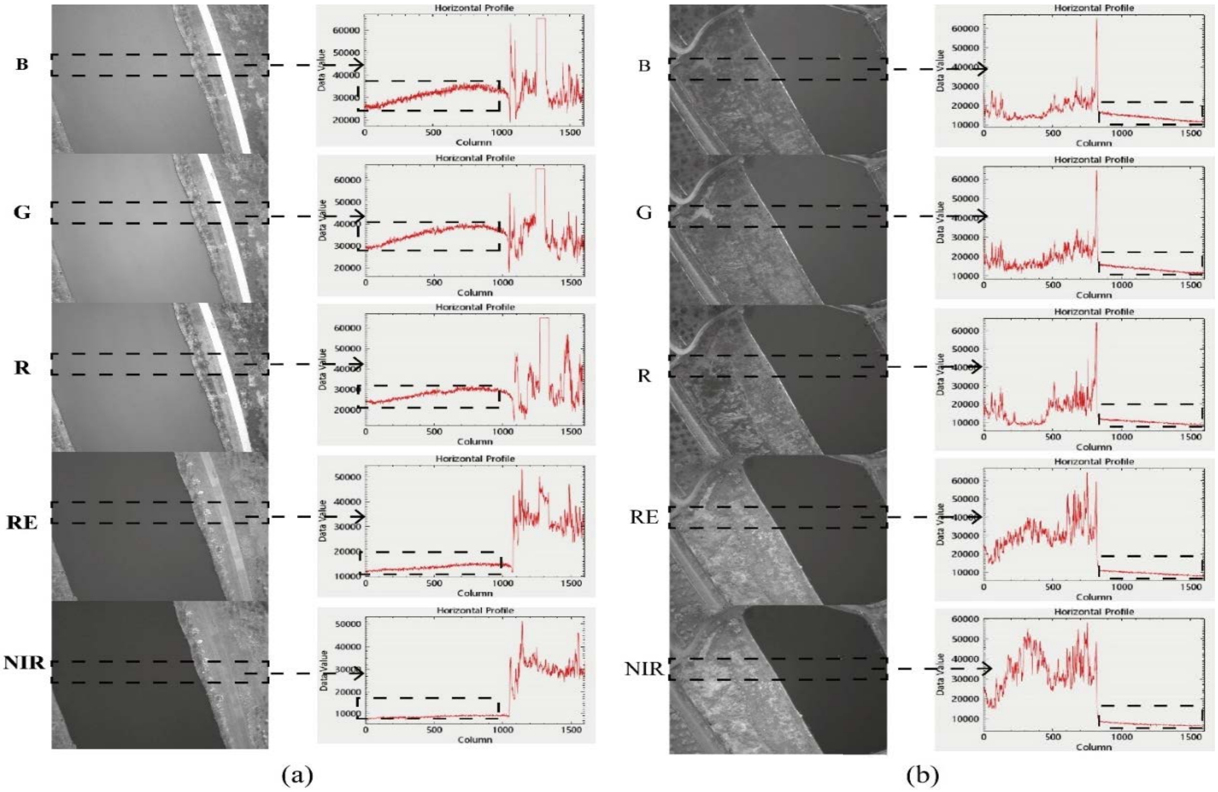


Fig. 3. Multispectral image of water and associated spectra.

parameters, the coefficient of variation of TN is largest in area A, and the degree of variation is high, which may be related to the uneven distribution of water bodies in area A and the uneven distribution of near-shore water bodies. Areas with fallen leaves have a higher TN content.

4.2.2. Data correlation analysis

The independent variable is the band combination, the dependent variable is the corresponding water-quality parameter, SPSS is used for correlation analysis, and several spectral parameters corresponding to each water-quality element index with $p < 0.05$ and meeting the statistical requirements are selected [31], as shown in Table 4. The results of the correlation analysis of areas A and B are given in Table 4. Area A has 3, 6, 6, and 3 sets of spectral parameters that meet the requirements of TN, TP, TUB, and ACD, respectively, and 5, 6, 5, and 6 sets of spectral parameters in area B meet the requirements of TN, TP, TUB, and ACD, respectively. We select the band combination with the highest correlation for fitting.

4.2.3. Construction of inversion model

Taking the optimal spectral parameters as independent variables, we established the linear function, polynomial function, power function model, and least squares model between each water-quality parameter and band factor. Tables 5 and 6 list the accuracy of the water-quality parameter-inversion models for areas A and B, respectively. To more intuitively express the accuracy evaluation outcomes of each model, the coefficient of determination and the error of the model are plotted as line graphs (Figs. S1 and S2).

Table 5 shows that the determination coefficients of the water quality parameter inversion model in Area A are all greater than 0.5. The polynomial model and the least squares model have the same coefficient of determination. However, the RMSEs of the least squares method are small. This result shows that the least squares model is more accurate and produces better fits. Table 6 shows that the R^2 of the inversion model of ACD and TP for area B are both about 0.7, in the inversion model of TN, the R^2 of the linear function is less than 0.5, and in the inversion model of TUB, the R^2 of the linear function model and the power function model are both less than 0.5. The determination coefficients of the polynomial model and least squares model are similar, but the RMSEs of the least squares model are smaller. Over-fitting is better controlled, as illustrated more intuitively in Figs. S1 and S2. The least squares model produces the best results in inverting each water-quality parameter with the RMSE as the standard. This method represents a significant improvement over the traditional gradient algorithm [32].

Table 4
Correlation coefficients between spectral parameters and water-quality parameters

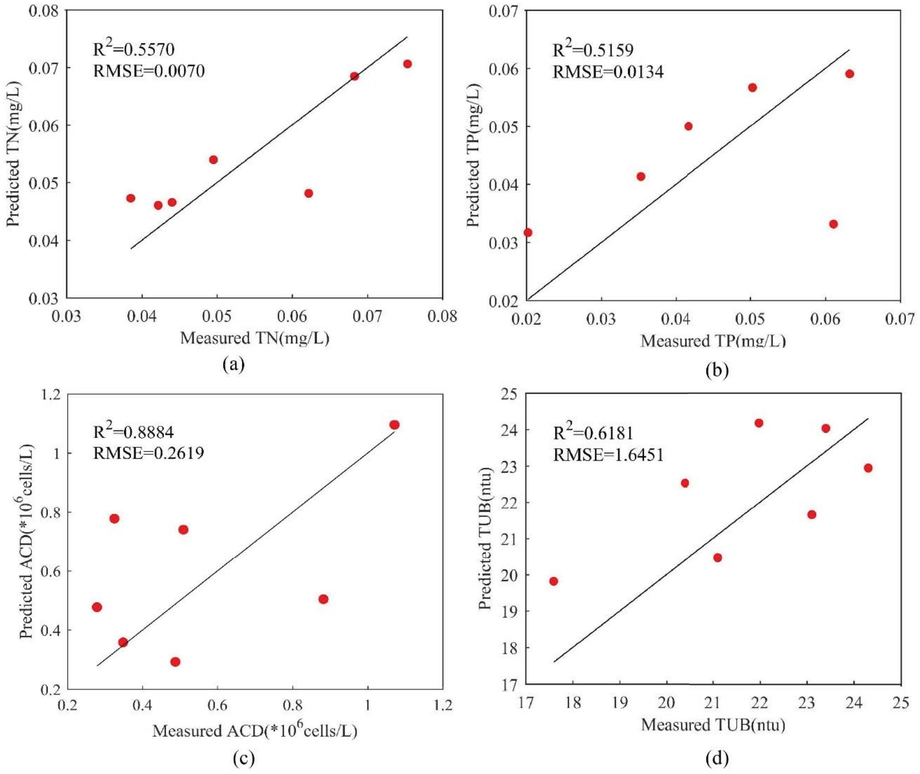
Water-quality parameters	Sensitive band	R^2	p	Sensitive band	R^2	p
	A			B		
Total nitrogen	B/R	0.841	0.004	(B+G)/R	0.599	0.003
	B/G	0.678	0.007	G/R	0.697	0.000
	(B+G)/R	0.562	0.021	NIR/R	0.637	0.001
				(B+RE)/R	0.654	0.001
				(B+G+RE)/(R+NIR)	0.544	0.009
Total phosphorus	RE/R	0.789	0.012	G/R	0.768	0.000
	B/R	0.782	0.013	RE/R	0.791	0.000
	R/NIR	0.765	0.016	(B+RE)/R	0.674	0.001
	RE/NIR	0.765	0.016	RE/(B+R)	0.611	0.003
	(B+G)/NIR	0.689	0.04	(B+NIR)/(B+R)	0.671	0.001
	(B+R)/NIR	0.673	0.047			
Turbidity	B	0.78	0.008	G/R	0.584	0.004
	B+RE	0.749	0.013	RE/R	0.661	0.001
	B+NIR	0.824	0.003	NIR/R	0.596	0.003
	RE+NIR	0.709	0.022	RE/(B+R)	0.58	0.005
	B+RE+NIR	0.772	0.009	(RE+NIR)/(B+G+R)	0.655	0.001
	(B+RE)/R	0.664	0036			
Algae density	B/R	0.71	0.021	B/R	0.656	0.001
	R/B	−0.638	0.047	G/R	0.821	0.000
	(B+G)/R	0.747	0.013	(B+G)/R	0.756	0.000
				(B+RE)/G	0.553	0.008
				(B+RE)/R	0.818	0.000
				(B+G+RE)/(R+NIR)	0.68	0.001

Table 5
Inversion model evaluation of water quality parameters for area A

Water-quality parameters	Fitting model	R^2	MAE	MAPE	RMSE
Total nitrogen	Linear fit	0.704	0.019	0.341	0.026
	Polynomial	0.910	0.011	0.218	0.016
	Power	0.81	0.014	0.253	0.021
	Least squares	0.910	0.011	0.218	0.013
Total phosphorus	Linear fit	0.588	0.017	0.314	0.024
	Polynomial	0.618	0.016	0.307	0.025
	Power	0.607	0.016	0.307	0.023
	Least squares	0.618	0.016	0.307	0.020
Turbidity	Linear fit	0.597	1.820	0.084	2.729
	Polynomial	0.600	1.828	0.086	2.904
	Power	0.591	1.807	0.083	2.748
	Least squares	0.600	1.828	0.085	2.430
Algae density	Linear fit	0.503	0.130	0.284	0.179
	Polynomial	0.674	0.116	0.262	0.155
	Power	0.572	0.129	0.292	0.167
	Least squares	0.674	0.116	0.262	0.13

Table 6
Inversion model evaluation of water-quality parameters for area B

Water-quality parameters	Fitting model	R^2	MAE	MAPE	RMSE
Total nitrogen	Linear fit	0.490	0.014	0.219	0.020
	Polynomial	0.672	0.010	0.158	0.016
	Power	0.528	0.014	0.219	0.019
	Least squares	0.672	0.010	0.156	0.015
Total phosphorus	Linear fit	0.625	0.010	0.236	0.012
	Polynomial	0.701	0.008	0.202	0.011
	Power	0.633	0.009	0.224	0.012
	Least squares	0.701	0.008	0.201	0.010
Turbidity	Linear fit	0.234	1.529	0.103	2.169
	Polynomial	0.640	1.026	0.067	1.524
	Power	0.308	1.441	0.097	2.059
	Least squares	0.640	1.026	0.067	1.416
Algae density	Linear fit	0.670	0.205	0.298	0.269
	Polynomial	0.741	0.181	0.240	0.244
	Power	0.715	0.614	0.667	0.250
	Least squares	0.741	0.181	0.240	0.227



(a) (b) (c) (d)

Fig. 4. Linear fits to plots of predicted value vs. true value for area A.

4.3. Model validation

Linear fitting is applied between the measured value of the validation dataset and the estimated value produced by the corresponding model. The results are shown in Figs. 4 and 5. The figure panels (a)–(d) are TN, fitted plots of TP, algal density, and turbidity, respectively. The predicted value is a red dot, and the black line is the 1:1 line. The closer the predicted value is to the 1:1 line, the more accurate is the estimate [33].

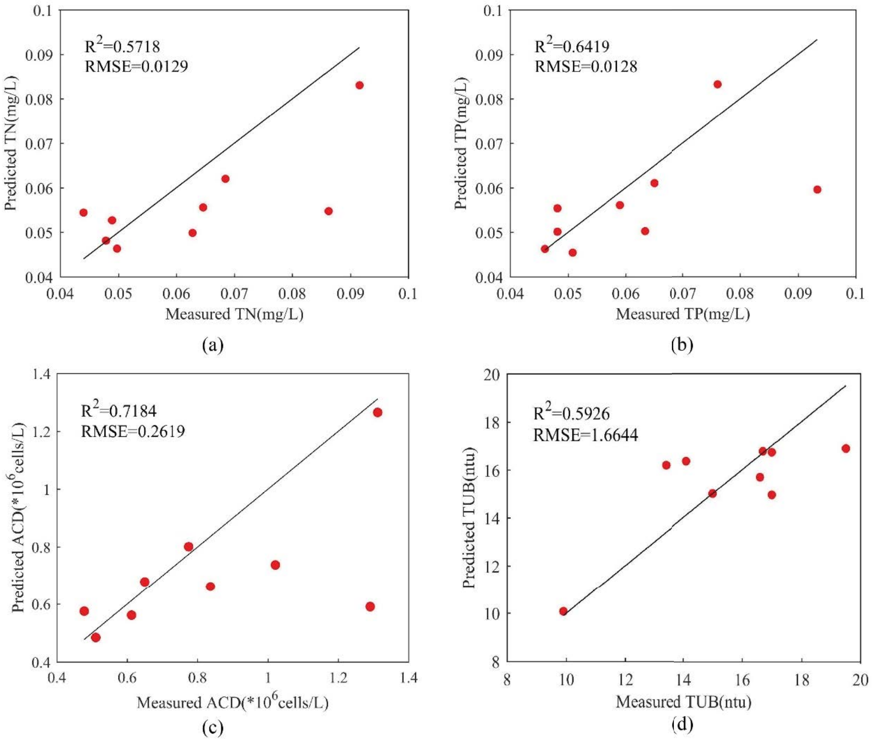


Fig. 5. Linear fits to plots of predicted value vs. true value for area B.

Fig. 4 shows that the model accuracy for ACD (TP) in area A is the highest (lowest), and the model accuracy of the ACD (TN) in area B is the highest (lowest). Except for a few points that are far from the 1:1 line, TN, TP, TUB, and ACD are all close to the 1:1 line, which is consistent with their determination coefficients ranging between 0.5 and 0.9 ($p < 0.01$), indicating that each inversion model is stable. The inversion accuracy of the least squares model is close to that of traditional satellite remote-sensing models for the four water-quality parameters used herein [34]. We can therefore estimate the spatial distribution of the water-quality parameters in the Longhu area [35].

4.4. Multispectral image inversion of water-quality parameters

The least squares model estimates each pixel in the multispectral image, and the estimation results for TN, TP, TUB, and ACD in the East Lake area are inverted [36,37]. First, the water body is extracted from the multispectral image by drawing the region of interest; next, the optimal inversion model corresponding to the four established parameters is inserted into the band-calculation tool to calculate the water quality corresponding to each pixel in the area. Finally, the inversion results are presented in stages for different concentrations, and the distribution of water-quality parameters in the study area is visualized. Different colors represent different values. The closer the color is red, the larger the value; the closer the color is to blue, the smaller the value [38] (Fig. 6, area A and Fig. 7, area B). The inversion results show that the concentration of TN in the study area is between 0.02 and 0.2 mg/L, the concentration of TP is

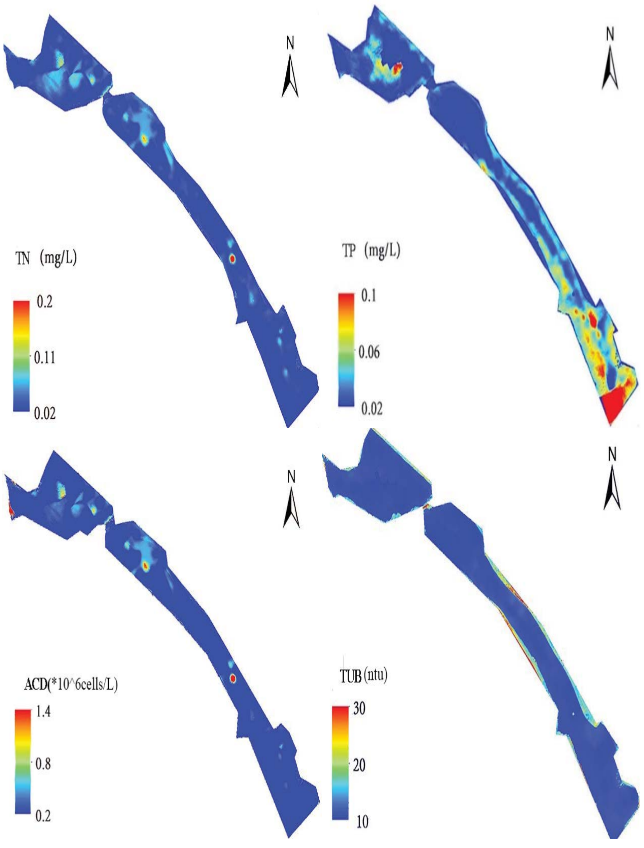


Fig. 6. Inversion results of region A.

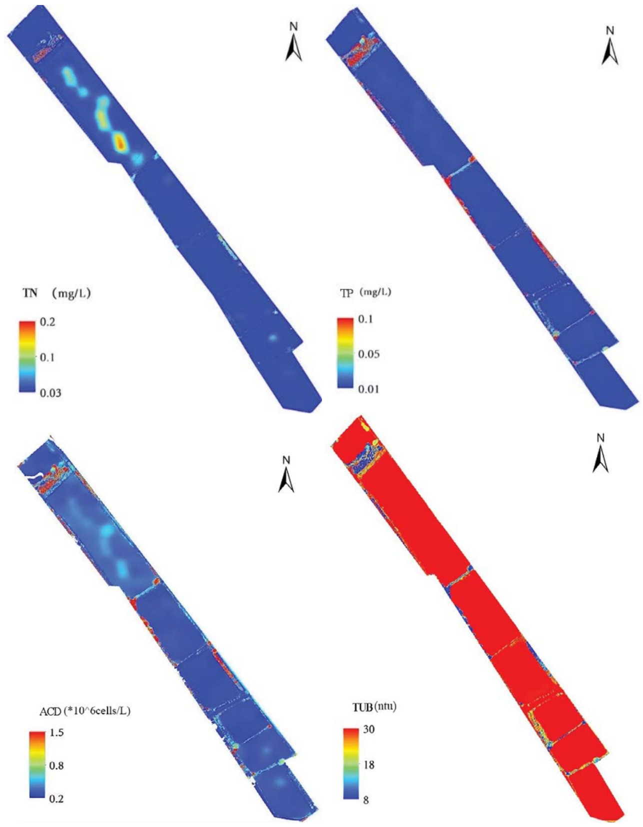


Fig. 7. Inversion results of region B.

Table 7
Basic item limits for environmental quality standards of surface water

Water-quality parameter	I	II	III	IV	V
TN≤	0.2	0.5	1.0	1.5	2.0
TP≤	0.02	0.1	0.2	0.3	0.4

0.01–0.1 mg/L, the TUB is 8–30 NTU, and the ACD is 0.2 × 10^6 to 1.5 × 10^6 cells/L. Among them, the content of TP in part of study area A is high, which may be due to the influence of fallen leaves in the coastal area. The turbidity of most of study area B is close to 30 NTU, and no value is abnormally high. This result may be related to the enhancement of animal and plant diversity, concentrations of suspended solids, and human activity involving ecological restoration [39].

4.5. Water-quality classification

Based on various inversion results for water-quality parameters and in accordance with GB 3838-2002 (environmental quality standards for surface water), the two water-quality indicators TP and TN are assessed. Table 7 lists the standard limits of surface water. The results show that the TN (TP) concentration in the study area is basically at the Class I (II) water-body level.

5. Conclusion

Based on UAV multispectral images and measured water-quality data, this paper quantitatively inverts the four water-quality parameters TN, TP, TUB, and ACD of a typical Longhu river. The results show that the appropriate mathematical transformation of reflectance improves the correlation between some water-quality parameters. The band-ratio model can be used to estimate water-quality parameters. The least squares inversion model of water-quality parameters improves the model accuracy and controls overfitting. Although the number of water samples is small in this work, some specific band combinations correlate strongly with water-quality parameters and have high inversion accuracy and good stability, which allows the water-quality parameters of small areas to be predicted. This work provides a preliminary verification of the reliability and flexibility of multispectral remote-sensing monitoring of water-quality parameters of small water bodies, and the study of different types of water bodies should increase our understanding of river polymorphism and provide a practical starting point for future UAV water monitoring. Although the water-quality inversion model applied in one area is not necessarily suitable for other areas, water-quality inversion is consistent, which provides a technical route for applying this technology in other areas. In addition, the inversion model should improve in follow-up work, allowing a model to be established by considering the season and the water body and thereby establishing a more universal inversion model.

Acknowledgments

This work was financially supported by the National Natural Science Foundation of China (Grant No. 22136003), The Foundation of Key Laboratory of Yangtze River Water Environment, Ministry of Education (Tongji University), China (Grant No. YRWEF202103), the 111 Project (D20015), and Hubei Province Support of Introducing Foreign Talents and Intelligence (Grant No. 2019BJH004).

References

[1] V. Sagan, K.T. Peterson, M. Maimaitijiang, P. Sidike, J. Sloan, B.A. Greeling, S. Maalouf, C. Adams, Monitoring inland water quality using remote sensing: potential and limitations of spectral indices, bio-optical simulations, machine learning, and cloud computing, Earth-Sci. Rev., 205 (2020) 103187, doi: 10.1016/j.earscirev.2020.103187.

[2] Z. Yang, X. Lu, Y. Wu, P. Miao, J. Zhou, Retrieval and model construction of water quantity parameters for UAV hyperspectral remote sensing, Sci. Survey. Mapp., 45 (2020) 60–64.

[3] B. Bansod, R. Singh, R. Thakur, Analysis of water quality parameters by hyperspectral imaging in Ganges River, Spat. Inf. Res., 26 (2018) 203–211.

[4] Y. Tian, H. Huang, G. Zhou, Q. Zhang, J. Tao, Y. Zhang, J. Lin, Aboveground mangrove biomass estimation in Beibu Gulf using machine learning and UAV remote sensing, Sci. Total Environ., 781 (2021) 1–18.

[5] M. Gholizadeh, A. Melesse, L. Reddi, A comprehensive review on water quality parameters estimation using remote sensing techniques, Sensors-Basel, 16 (2016) 1–43.

[6] Z. Hu, Y. Zhou, Research on urban water quality monitoring method based on low-altitude multi-spectral remote sensing, Geo-spatial Inf., 18 (2020) 4–8.

[7] K. Dörnhöfer, N. Oppelt, Remote sensing for lake research and monitoring – recent advances, Ecol. Indic., 64 (2016) 105–122.

[8] L. Wang, H. Bai, Research review on retrieval of water quality parameters about lake based on remote sensing techniques, GNSS World China, 38 (2013) 57–61.

[9] X. Jin, C. Wang, Z. Yuan, Research on reservoir water quality monitoring method based on remote sensing, Henan Sci. Technol., 40 (2021) 43–46.

[10] A. Ruescas, M. Hieronymi, G. Mateo-Garcia, S. Koponen, K. Kallio, G. Camps-Valls, Machine learning regression approaches for colored dissolved organic matter (CDOM) retrieval with S2-MSI and S3-OLCI simulated data, Remote Sens., 10 (2018) 786, doi: 10.3390/rs10050786.

[11] A.P. Piotrowski, M. Osuch, M.J. Napiorkowski, P.M. Rowinski, J.J. Napiorkowski, Comparing large number of metaheuristics for artificial neural networks training to predict water temperature in a natural river, Comput. Geosci.-UK, 64 (2014) 136–151.

[12] A. Hamzic, Z. Avdagic, S. Omanovic, A Sequential Approach for Short-Term Water Level Prediction Using Nonlinear Autoregressive Neural Networks, IEEE, Sarajevo, Bosnia and Herzegovina, 2016, pp. 1–7.

[13] M. Mamun, J.-J. Kim, Md. A. Alam, K.-G. An, Prediction of algal Chlorophyll-a and water clarity in monsoon-region reservoir using machine learning approaches, Water, 12 (2020) 30, doi: 10.3390/w12010030.

[14] F. Ma, Q. Jiang, L. Xu, Y. Liang, R. Wang, S. Su, Retrieval of water quality parameters based on BP neural network algorithm in Miyun Reservoir, Ecol. Environ. Sci., 29 (2020) 569–579.

[15] Y. Chen, L. Liu, M. Chen, Comparative analysis of water quality inversion models based on UAV multispectral data, China Water Transport, 22 (2022) 29–31.

[16] Y. Liu, K. Xia, H. Feng, Y. Fang, Inversion of water quality elements in small and microsize water region using multispectral image by UAV, Acta Sci. Circum., 39 (2019) 1241–1249.

[17] R. Kong, Analysis of the Effect of DJI Genie 4 RTK Parameter Settings on the Efficiency of Aerial Surveying and Mapping, Pearl River Water Transport, (2020) 53–54.

[18] L. Chang-hou, Study on Relationship Between the Spectrum Band Width and the Absorbance Error, Analysis and Technology and Instruments, (2004) 65–67.

[19] D. Xiao, Y. Pan, J. Feng, J. Yin, Y. Liu, L. He, Remote sensing detection algorithm for apple fire blight based on UAV multispectral image, Comput. Electron. Agric., 199 (2022) 1–12.

[20] H. Zhu, Y. Huang, Y. Li, F. Yu, G. Zhang, L. Fan, J. Zhou, Z. Li, M. Yuan, Predicting plant diversity in beach wetland downstream of Xiaolangdi reservoir with UAV and satellite multispectral images, Sci. Total Environ., 819 (2022) 1–16.

[21] X. Tao, Y. Li, Q. Luan, J. Jiang, Estimation of anthocyanin content in *Pinus elliottii* based on UAV remote sensing, Acta Agric. Univ. Jiangxiensis, 43 (2021) 1065–1077.

[22] J. Li, H. Huang, J. Xiu, B. Li, H. Zhang, Effect and compensation of overlap influenced by flight parameter of oblique aerial camera, Opt. Precis. Eng., 28 (2020) 1254–1264.

[23] B. Chen, X. Mu, P. Chen, B. Wang, J. Choi, H. Park, S. Xu, Y. Wu, H. Yang, Machine learning-based inversion of water quality parameters in typical reach of the urban river by UAV multispectral data, Ecol. Indic., 133 (2021) 1–18.

[24] J. Xin, Study on main influencing factors of determination of total phosphorus in water by ammonium molybdate spectrophotometry, China Resour. Compr. Util., 40 (2022) 23–25.

[25] D. Xie, Study on determination of total nitrogen in water by alkaline potassium persulfate method, Leather Technol., 3 (2022) 25–26.

[26] C. Chen, Q. Lu, Research and suggestion on the applicability of turbidity meter method to measure turbidity in water, Chem. Eng. Des. Commun., 47 (2021) 61–62.

[27] X. Zhou, S. Ma, Z. Shang, Y. Wang, L. Guo, C. Lin, Determination of cell density of *Microcystis aeruginosa* by spectrophotometry, Water Conserv. Technol. Supervision, 24 (2016) 50–51.

[28] X. Zhu, L. Liu, Z. Ye, UAV remote sensing monitoring method for water quality, China Water Transport, (2021) 157–159.

[29] Q. Shao, X. Guo, Y. Li, Y. Wang, D. Wang, J. Liu, J. Fan, F. Yang, Using UAV remote sensing to analyze the population and distribution of large wild herbivores, J. Remote Sens., 22 (2018) 497–507.

[30] J. Wei, F. Jia-yuan, The least square method and its application, J. Commun. Univ. China (Sci. Technol.), 27 (2020).

[31] H. Zhang, B. Yao, S. Wang, G. Wang, Remote sensing estimation of the concentration and sources of coloured dissolved organic matter based on MODIS: a case study of Erhai lake, Ecol. Indic., 131 (2021) 1–12.

[32] S. Zhu, Z. Chen, Y. Zhang, Cotton seeding emergence information extraction based on UAV digital image, Mod. Electron. Tech., 45 (2022) 61–64.

[33] J. Lu, J.U.H. Eitel, M. Engels, J. Zhu, Y. Ma, F. Liao, H. Zheng, X. Wang, X. Yao, T. Cheng, Y. Zhu, W. Cao, Y. Tian, Improving unmanned aerial vehicle (UAV) remote sensing of rice plant potassium accumulation by fusing spectral and textural information, Int. J. Appl. Earth Obs. Geoinf., 104 (2021) 1–15.

[34] Y. Zhang, Y. Zhang, Y. Zha, K. Shi, Y. Zhou, M. Liu, Estimation of diffuse attenuation coefficient of photosynthetically active radiation in Xin'anjiang reservoir based on Landsat 8 data, Environ. Sci., 36 (2015) 4420–4429.

[35] W. Zhou, H. Yang, L. Xie, H. Li, L. Huang, Y. Zhao, T. Yue, Hyperspectral inversion of soil heavy metals in three-river source region based on random forest model, Catena, 202 (2021) 1–10.

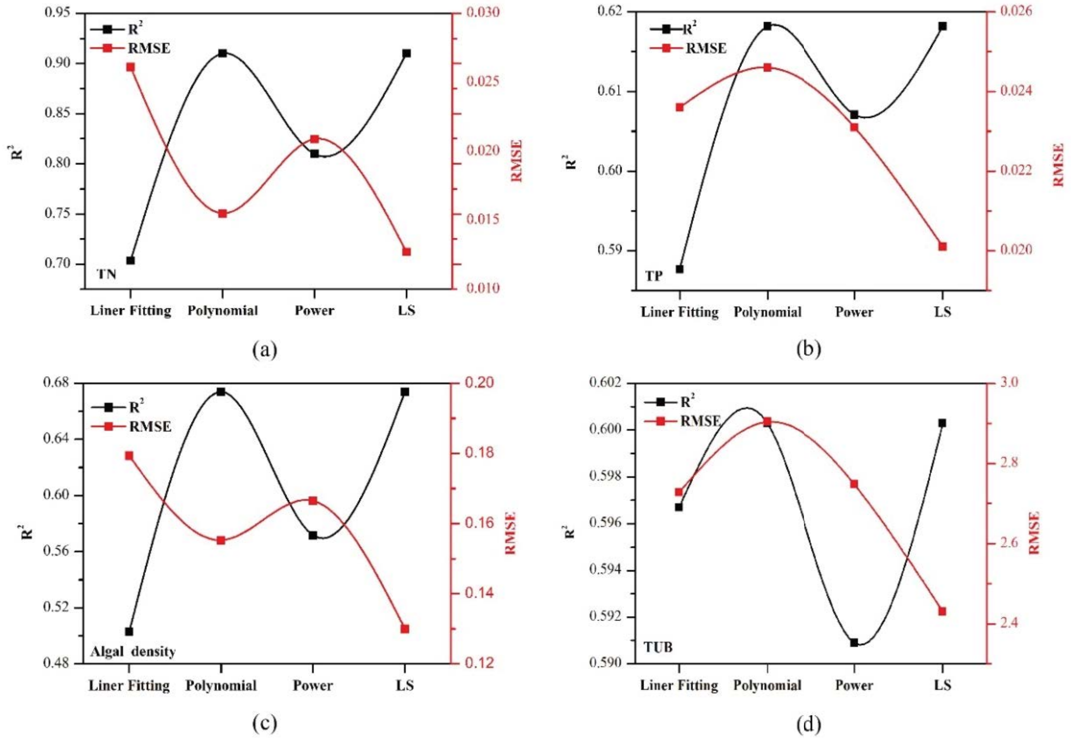
[36] X. Sòria-Perpinyà, E. Vicente, P. Urrego, M. Pereira-Sandoval, A. Ruíz-Verdú, J. Delegido, J.M. Soria, J. Moreno, Remote sensing of cyanobacterial blooms in a hypertrophic lagoon (Albufera of València, Eastern Iberian Peninsula) using multi-temporal Sentinel-2 images, Sci. Total Environ., 698 (2020) 1–10.

[37] K. Matsui, H. Shirai, Y. Kageyama, H. Yokoyama, Improving the resolution of UAV-based remote sensing data of water quality of Lake Hachiroko, Japan by neural networks, Ecol. Inf., 62 (2021) 1–13.

[38] F. Wang, C. Zhao, H. Yang, H. Jiang, L. Li, G. Yang, Non-destructive and in-site estimation of apple quality and maturity by hyperspectral imaging, Comput. Electron. Agric., 195 (2022) 1–9.

[39] Z. Yuan, The Monitoring and Analysis of Chlorophyll-a and Turbidity by Remote Sensing in MinJiang River, Fuzhou University, 2016, p. 74.

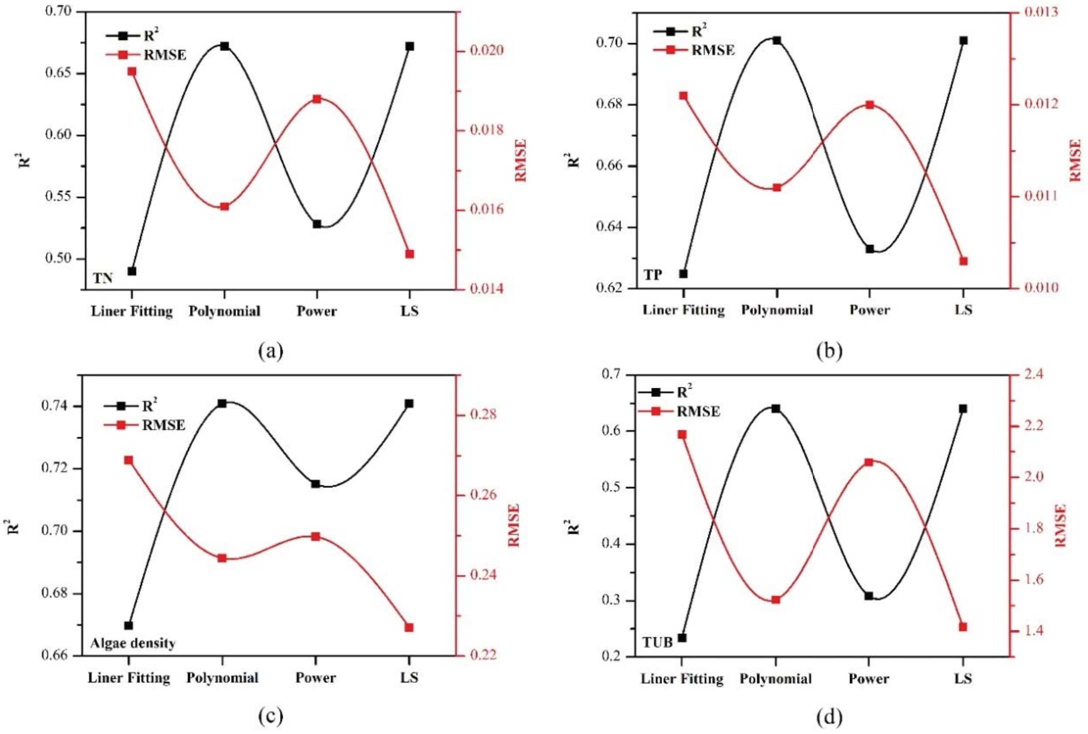
Supporting information



(a) (b)

(c) (d)

Fig. S1. The accuracy comparison of each model in area A.



(a) (b)

(c) (d)

Fig. S2. The accuracy comparison of each model in area B.