

Article

A New Method for Calculating Water Quality Parameters by Integrating Space-Ground Hyperspectral Data and Spectral-In Situ Assay Data

Donghui Zhang ^{1,2}, Lifu Zhang ^{1,2,3,*}, Xuejian Sun ^{1,2}, Yu Gao ^{2,4}, Ziyue Lan ², Yining Wang ², Haoran Zhai ², Jingru Li ², Wei Wang ², Maming Chen ², Xusheng Li ⁵, Liang Hou ⁶ and Hongliang Li ⁷

¹ State Key Laboratory of Remote Sensing Science, Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100094, China; zhangdonghui@aircas.ac.cn

² Progoo Research Institute, Tianjin Progoo Information Technology Co., Ltd, Tianjin 300380, China; zhanglf@radi.ac.cn

³ Key Laboratory of Oasis Eco-agriculture, Xinjiang Production and Construction Corps, Shihezi University, Shihezi 832003, China;

⁴ School of Earth Sciences, Chengdu University of Technology, Chengdu 610059, China; gao7819@foxmail.com

⁵ National Key Laboratory of Remote Sensing Information and Imagery Analyzing Technology, Beijing Research Institute of Uranium Geology, Beijing 100029, China; saintlxs@foxmail.com

⁶ Institute of Agricultural Information and Economy, Hebei Academy of Agriculture and Forestry Sciences, Shijiazhuang 050051, China; giantark@163.com

⁷ Tianjin Institute of Metrological Supervision and Testing, Tianjin 300192, China; caoshangfei-666@163.com

* Correspondence: zhanglf@radi.ac.cn; Tel.: +86-1371-6974-736.

Abstract: The effective integration of aerial remote sensing data and ground multi-source data has always been one of the difficulties of quantitative remote sensing. A new monitoring mode is designed which installs the hyperspectral imager on the UAV and places a buoy spectrometer on the river. Water samples are collected simultaneously to obtain in situ assay data of total phosphorus, total nitrogen, COD, turbidity and chlorophyll during data collection. The cross correlogram spectral matching (CCSM) algorithm is used to match the data of the buoy spectrometer with the UAV spectral data to reduce the UAV data noise significantly. An absorption characteristics recognition algorithm (ACR) is designed to realize a new method for comparing UAV data with laboratory data. This method takes into account the spectral characteristics and the correlation characteristics of test data synchronously. It is concluded that the most accurate water quality parameters can be calculated by using the regression method under five scales after the regression tests of multiple linear regression method (MLR), support vector machine method (SVM) and neural network (NN) method. This new working mode of integrating spectral imager data with point spectrometer data will become a trend in water quality monitoring.

Keywords: hyperspectral imager; UAV remote sensing; water quality monitoring; space-ground data; buoy spectrometer; water eutrophication; absorption characteristics

0. Introduction

With the agricultural, industrial and commercial utilization of water resources, a large amount of sewage is produced. The premise of controlling water pollution is to monitor the changes of water quality. It can be divided into contact technology and non-contact technology from the instrument principle. The former includes water probe method, assay method and biological method; The latter includes remote sensing spectroscopy, laser method and transmission method. Each method has its scope of application and shortcomings [1]. For example, the water inlet probe needs to wipe the sensor regularly, the chemical method will produce secondary pollution, the biological method basically

has no quantitative ability, the processing process of remote sensing spectroscopy is complex, the laser method lacks mechanism basis, and the transmission method can only play a better effect indoors.

It can be divided into satellite, airborne, UAV and water surface from the perspective of data acquisition platform. Common satellite data include Sentinel [2], Landsat8 [3], Hyperion [4], MODIS [5], IKONOS [6], MERIS [7], AHSI [8], PRISMA [9]; Airborne and UAV airborne data include HyMap [10], HIS [11], Spectral Evolution [12], VNIR [13], Hyperspectral Imager [14], Ocean Optics [15], Headwall [16], Gaia Sky-mini [17]; There are a large number of micro sensors represented by ASD [16] for water surface data. The data of different platforms have obvious advantages and disadvantages. It is necessary to carry out integrated application in order to further improve the accuracy of water quality monitoring.

The collaborative work of data has always been a potential breakthrough in the accuracy of water quality calculation. Scholars have gradually developed from hardware integration and platform integration to data integration. The core idea is to realize the mutual calibration of the original data and the mutual verification of the result data. The initial steps begin with sensor hardware integration. For example, build a set of autonomous robots to realize long-distance data transmission [18]; a variety of satellite data are analyzed synchronously to remove noise [3, 19]; a set of multi parameter monitoring stations are built to comprehensively monitor seismicity, geomagnetic field change, water temperature, pressure, salinity, chemistry, ocean current and gas generation [20]; a set of software and hardware is designed, which can be applied to the construction of monitoring stations in marine and continental waters [21]. Further, high-precision monitoring of water composition can be realized by cooperating with ground and satellite data with the help of the integrated application of sensors on different platforms [22]. The temperature field information of AVHRR data is studied by taking buoy and ship measurements as reference [23]; the band ratio algorithm of in-situ data and sentinel-2 image is scientifically integrated [24]; the joint calculation of satellite data and monitoring station data is realized [25]; the model of UAV hyperspectral and ground measured data is used to realize water quality monitoring [17]; a set of collaborative image processing flow is designed [26].

Nowadays, data integration, which can comprehensively consider the effects of time and space, is becoming a research hotspot. A general idea of spatio-temporal fusion of multi-source remote sensing data is designed [27,28]. Space stations, air stations, field or ground hyperspectral systems have been built one after another [29]. A working mode combining small UAV system and small sensors is designed. It is considered that this new observation mode will become a common tool for water resources management in the future [15]. If the data with different spectral resolution, spatial resolution and temporal resolution can be analyzed uniformly, the reliability of the results will be further improved [30]. The collaborative working mode of MODIS, Landsat, hydrological data and DEM ScienceDirect is studied and discussed [31]. In short, the opening and application of data is the key to water environment monitoring [32].

Great requirements are put forward for the algorithm efficiency with the increase of the data volume. It is obviously the simplest and most efficient method to analyze the correlation between some ground test data and sensor data. The designed semi empirical algorithm has a good extraction effect for chlorophyll-a and phycocyanin concentrations [33]. Hyperspectral technology can play a role in the identification of crop nitrogen stress and water stress, and establish the relationship between crop state indexes and spectral data [34]. The characteristic band of chlorophyll a in hyperspectral imager of coastal ocean (HICO) image is extracted by neural network [14]. The prediction ability of partial least squares regression for hyperspectral remote sensing and in-situ chlorophyll a concentration is tested [35]. With the support of 52 sampling point data, the machine learning method is selected, and the technology of combining in-situ assay data with hyperspectral data is studied for the urban inland water, and it is considered that this technology can overcome the limitations of traditional band selection methods [36].

A large number of more difficult and better algorithms have been designed on the basis of mastering the transmission mechanism of water quality spectral data [37]. Artificial neural network (ANN) and wavelet neural network (WNN) models are used to calculate the daily and hourly values of salinity, temperature and dissolved oxygen in the bay water [38]. The accuracy of pixel based a pixel-based deep neural network regression (pixel_DNNR) model and a patch-based deep neural network regression (patch_DNNR) model are compared, and the content information difficult to be extracted by conventional methods such as permanganate index, total nitrogen, total phosphorus, ammonia nitrogen and heavy metals is obtained by using aerial VNIR hyperspectral data [13]. The R2 coefficient of the designed neural network algorithm can exceed 0.9 through the calculation of phosphorus, nitrogen, biochemical oxygen demand, chemical oxygen demand and chlorophyll a [17].

Some new ideas have also been recognized based on the gradual maturity of the above methods. A hybrid bayesian back-propagation neural network approach to multivariate modelling [39] and a three-step semi analytical algorithm [40] to calculate the inherent optical properties of ocean, coastal and inland waters. These studies are a good attempt. Multi algorithm index and look-up table (MAIN-LUT) technology can avoid the defect of the algorithm falling into the local optimal solution, which has been verified in the calculation of chlorophyll a [41]. A matching pixel by pixel algorithm is designed to establish the linear regression model of chlorophyll a, depth and turbidity [42]. Convolution recurrent neural network (convRNN) and other depth learning methods have achieved good results in crop information extraction [43]. Chlorophyll a content in inland and coastal waters is calculated by fluorescence analysis technique [44].

Analysis of existing water quality parameters, mainly focusing on chlorophyll a [4,14,22,28,45–48]; Suspended particulate matter [9,22,47,49]; dissolved organic matter [22,49]; transparency [50], total phosphorus [6] total nitrogen [51], ammonia nitrogen [13], biochemical oxygen demand [52], water color, colored dissolved organic matter (CDOM), dissolved organic carbon [24], transparency [17], pH [10], turbidity [53], water depth [42]. In addition, some regional parameters also have preliminary conclusions, including salinity, temperature and dissolved oxygen [38], water temperature, pressure, salinity, chemistry, ocean current [20] and heavy metals [13]. It can be concluded that the above indicators gradually change from small watershed research to regional research, and finally to global scale exploration. This will undoubtedly lead the expansion of technology boundary.

In this paper, we explored the core technology of spectral collaborative processing by deploying buoy spectrometer, UAV hyperspectral image data acquisition and river in-situ sampling and testing on a river that has attracted much attention from the local government. The research contents include the matching method of spectral data, the selection technology of water quality characteristic bands, and the calculation accuracy of water quality parameters at different scales. A new algorithm (Absorbance Characteristics Recognition, ACR) is designed, which can take into account the advantages of supervised method and unsupervised method. The relatively optimal calculation models for total phosphorus, total nitrogen, COD, turbidity and chlorophyll are established respectively by comparing various regression methods. The results provide a scientific basis for local analysis of water pollution sources and environmental treatment.

1. Study Area and Data Collection

1.1. Study area

Foshan City is located in Guangdong province, which is the southeast part of China. Lingnan Avenue River, the main sewage river in the city center, is selected as the study area. The river is located from 113.12°E to 113.14°E longitude and from 22.98°N to 23.03°N latitude (Figure 1). The river length is 7.78km and is one of the important drainage channels in the urban area. Located at the intersection of Tanzhou waterway and Pingzhou

waterway, this area is the most densely populated area in the city. The river is mainly polluted by the domestic drainage of residents and the drainage of commercial places. At the same time, some small processing plants are distributed on both banks to discharge industrial sewage.

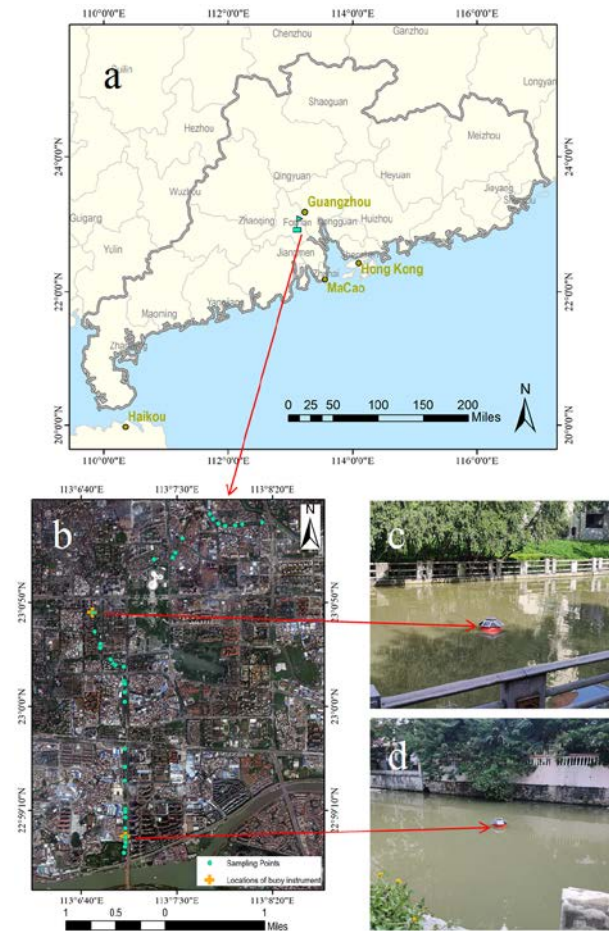


Figure 1. Geographic location of the study area and the selected sampling positions. (a) Map showing the location of the study area, Foshan, Guangdong province, China. (b) 36 water quality sampling points are distributed along the river. During the acquisition of hyperspectral data by UAV, two buoy hyperspectral sensors were set up simultaneously in the middle and downstream of the river. (c) The first buoy hyperspectral sensor, No. A. There is some shadow interference in this position (d) The second buoy hyperspectral sensor, No. B. There are no shadows in this position.

The hyperspectral data of UAV with a total area of 0.92 km² were obtained, and the laboratory data of 36 points were collected simultaneously on the river. The collection of water sample points and storage of samples in accordance with the Chinese Environmental Quality Standards for Surface Water (GB3838-2002). The turbidity, total phosphorus, total nitrogen, COD and chlorophyll contents of each sampling point were obtained within 12 hours. The local government and residents are very concerned about the water quality of this river. As a test water system for controlling river pollution, they believe that the water quality of such an important river directly reflects the basic situation of the local environment.

1.2. Data collection

1.2.1. Hyperspectral image acquisition

Nano hyperspec, a visible and near-infrared spectrometer developed in the United States was used for the hyperspectral image acquisition. The wavelength range is

398.7nm-1000.46nm, the number of bands is 272, and the spatial bandwidth is 640. The data is stored on the built-in SSD disk with a maximum frame rate of 350Hz. The sensor is mounted on the DJI M600PRO, which can work continuously for 35 minutes with a load of 6kg and a flight speed of 18 m/s.

The spectrometer is calibrated by integrating sphere to ensure that its wavelength position is accurate before flight. According to the field survey of the study area, there are many buildings with a height of nearly 100m on both banks of the river. Therefore, in order to ensure safety, the design navigation height is 120m. The acquisition dates were August 16, 2021, and August 17, 2021, and 10 strips data with a spatial resolution of 0.075m were generated in total. Geometric correction is completed according to UAV attitude and navigation POS data. POS data has 7 parameters, including longitude, latitude, altitude, rolling, pitching, heading and time. Atmospheric correction is achieved by laying calibration cloth with reflectivity of 11%, 32% and 56% simultaneously during UAV operation, and by linear fitting according to the actual reflectivity of the calibration cloth (Figure 2). The flight direction is along the river and the reflectivity uncertainty caused by water flow can be ignored due to the slow velocity of the river.

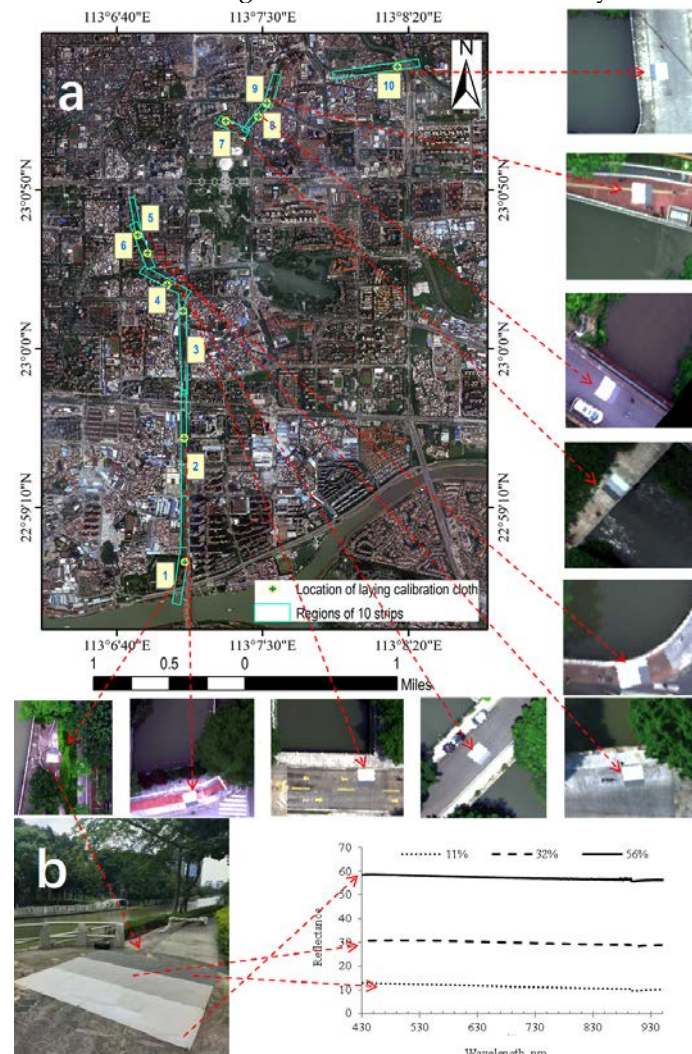


Figure 2. Distribution of 10 strips and the information of radiometric calibration cloth. (a) The radiation calibration cloth is laid for each strip, and 3 reflectivity calibration cloth are laid respectively. The cloth is laid in a flat and unobstructed place with an area of 3×3m. (b) This is the standard reflectance of the calibration cloth. They are 11%, 32% and 56% respectively. In the later calibration, they are selectively used according to the field illumination.

The calibration cloth can radiometrically calibrate the UAV image and convert the DN value into water reflectance [54], which can be expressed as:

$$\rho_{water} = \frac{DN_{water} - DN_{cloth}}{DN_{calibrationplate} - DN_{cloth}} (\rho_{calibrationplate} - \rho_{cloth}) + \rho_{cloth} \quad (1)$$

where ρ_{water} and DN_{water} are the water reflectance and DN value; ρ_{cloth} and DN_{cloth} are the calibration cloth reflectance measured under the same solar illumination and DN value; $\rho_{calibrationplate}$ and $DN_{calibrationplate}$ are the calibrated reference board under the same solar illumination and DN value respectively.

1.2.2. Water surface hyperspectral data acquisition

The author's team has developed a buoy spectrometer water quality detection system that can be applied to rivers, lakes, ponds and other waters. The system consists of Hyscan micro intelligent spectrometer, fixed buoy and water quality data cloud service platform. The instrument control and data return are completed in the cloud. The spectral range is 400 nm to 1000 nm and the instrument weighs 20kg. The power supply of the instrument is solar energy plus rechargeable battery pack, which can automatically collect a group (10 spectra) of spectral data in 30 minutes. It can work continuously for more than 3 months in good daylighting conditions places (Figure 3 a). It can automatically retrieve a variety of water quality parameters, realize real-time data transmission, and support cloud data storage, real-time display and statistical analysis. The data can be transmitted to the screen, iPad and mobile terminal in real time, and the water quality can be viewed anytime and anywhere (Figure 3 b). The buoy spectrometer collects spectral data while the UAV is flying. A total of 200 water spectral data were obtained in two days (Qu et al., 2008). The significance of these data is that, on the one hand, they can calibrate UAV data to reduce the data uncertainty caused by atmosphere, shadow, light intensity, etc. On the other hand, collecting water samples around the buoy spectrometer can directly build the relationship between various water quality parameters and spectra, find out the characteristic bands, and help establish a more accurate model of hyperspectral images.



Figure 3. The system is composed of intelligent water quality spectrometer and data analysis cloud service platform. (a) The water quality spectrometer is fixed on the water surface, collects spectral data regularly, and transmits it to the cloud service in real time through 4G / 5G network. (b) The system supports cloud data storage, statistical analysis, and supports real-time viewing on the user's client.

1.2.3. Water parameter sampling and measurement

Collect a bottle of 500 mL sample at each sampling site and keep it in a box with ice bag, and complete the chemical testing within 12 hours (Table 1). The contents of total phosphorus, total nitrogen and cod were obtained by the assay instrument DR6000 [55]. Among them, (1) the content of total phosphorus is obtained by adding 5ml potassium dihydrogen phosphate to the water sample and heating to digestion at 150 °C for 30 minutes, with an precision of 0.01mg/L; (2) the content of total nitrogen is obtained by

adding 2ml potassium nitrate to the water sample and heating it to 105 °C for digestion for 30 minutes, with an precision of 0.1mg/ L; (3) the reagent added for COD is potassium hydrogen phthalate. After adding 2ml, it is heated and digested at 150 °C for 2 hours to obtain the test value with an precision of 0.1mg/L; (4) The turbidity test value is obtained by TSS portable instrument. Place the probe in the water sample for 2 hours to obtain a continuous set of values. After averaging, the test value with an accuracy of 0.1mg/l is obtained; (5) Using a similar measurement method, with the support of HQ40d instrument, the chlorophyll value can be obtained, and the precision can reach 0.01 µg/L.

Table 1. Statistical values of water quality parameters of different strips consisting of 45 sampling test data.

Strips	Total phosphorus (mg/L)		Total nitrogen (mg/L)		COD (mg/L)		Turbidity (mg/L)		Chlorophyll (mg/L)	
	Range	Mean	Range	Mean	Range	Mean	Range	Mean	Range	Mean
1	0.7-1.0	0.82	7.0-12.0	9.55	5.0-22.0	15.73	24.10-42.30	29.21	3.59-6.04	5.15
2	0.9-1.1	0.98	6.0-8.0	7.00	11.0-16.0	13.00	29.90-34.90	31.60	4.55-5.20	4.91
3	1.0-1.2	1.08	9.0-13.0	11.25	9.0-13.0	11.50	35.40-40.10	37.33	4.15-4.41	4.25
4	0.8-1.8	1.30	11.0-15.0	13.00	12.0-21.0	16.50	34.00-49.50	41.75	4.29-4.49	4.39
5	1.0-1.2	1.10	12.0-13.0	12.50	12.0-13.0	12.50	31.10-31.70	31.40	5.01-5.51	5.26
6	1.1-1.2	1.15	11.0-12.0	11.50	13.0-17.0	15.00	48.30-48.60	48.45	3.92-4.26	4.09
7	1.1-1.2	1.15	13.0-14.0	13.50	11.0-11.0	11.00	49.00-50.10	49.55	3.46-3.59	3.53
8	1.1-1.1	1.10	13.0-13.0	13.00	16.0-18.0	17.00	45.40-47.30	46.35	4.34-4.55	4.45
9	1.4-1.4	1.40	12.0-12.0	12.00	14.0-18.0	16.20	42.90-51.70	48.36	3.36-4.52	3.97
10	1.0-1.2	1.10	20.0-20.0	20.0	12.0-13.0	12.50	26.50-26.50	26.50	3.84-4.19	4.02
Buoy sensor A	1.0-1.1	1.05	13.0-18.0	15.25	13.0-19.0	15.25	36.50-40.30	38.48	4.18-4.44	4.34
Buoy sensor B	0.8-1.1	0.90	6.0-8.0	7.20	5.0-9.0	7.00	29.90-30.70	30.20	4.13-4.40	4.29

2. Methodology

2.1. Workflow

A set of technical process of water quality parameter extraction is designed for buoy spectrometer, UAV hyperspectral image data and test data at sampling points (Figure 4). The cross correlogram spectral matching (CCSM) algorithm can effectively match space and ground data (Section 2.2) and further improve the accuracy of UAV data (Section 3.1). A new absorbance characteristics recognition algorithm (ACR) (Section 2.3) is designed to compare the ground test data with the UAV data. This method can combine the advantages of supervised method and unsupervised method to select the overlapping band as the potential effective band for modeling (Section3.2). Four scale amplification tests (Section 2.4) are carried out at the sampling points in addition to the in-situ scale in order to verify the scale effect, and the sensitive bands of water quality parameters at different scales are further studied. Using two band cluster analysis (Section 2.4) and three regression algorithms (refer to Section 2.5 for the algorithm and Section 3.3 for the result), the accuracy evaluation results of five types of water quality parameters are obtained (refer to Section 2.6 for the algorithm and Section 3.4 for the result). The prediction results of five water quality parameters at modeling points are drawn, and the distribution law of water quality parameters in the upstream, midstream and downstream of Lingnan Avenue River (Section 3.5) are analyzed based on these.

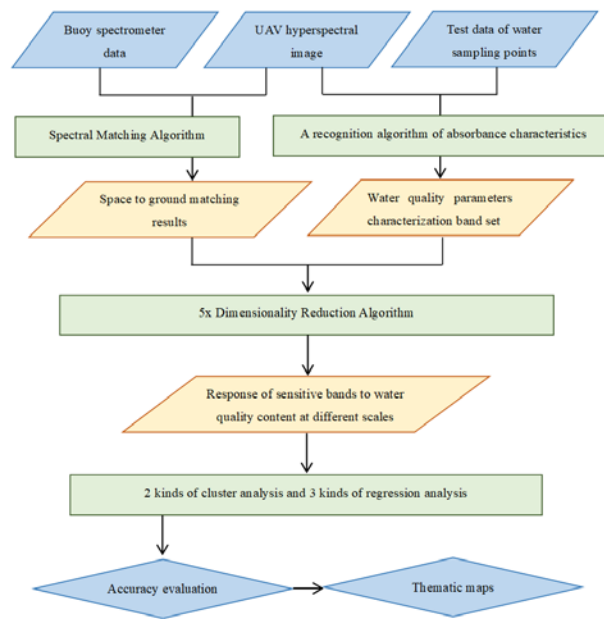


Figure 4. Workflow of the new method for calculating water quality parameters by integrating space-ground hyperspectral image data and spectral-in situ assay data.

2.2. Spectral matching algorithm for UAV and buoy data

Since the sensor of the buoy spectrometer is only 10cm away from the water surface and the spectral energy source is a stable halogen lamp, in addition, the water surface is in a shaded dark environment, which can be recognized as the true reflectance of the water. Although the UAV spectrum has been corrected by calibration cloth, there are still some errors due to interference such as shadow occlusion and light intensity change. The cross correlogram spectral matching (CCSM) algorithm [56] is used to calculate the linear correlation coefficient between buoy spectral data and UAV spectral data through the relative translation of spectral axis, and draw the cross-correlation coefficient diagram to remove these errors. It is considered that if the cross-correlation coefficient of the two bands reaches the maximum, it is a similar band. The secondary calibration of UAV spectral data is realized by this method.

This algorithm determines the similarity of the spectrum, which depends on the spectral shape rather than the reflectance, and can overcome the spectral error caused by atmospheric and sensor noise. It is especially sensitive to spectral shape error caused by water surface structure. The calculation formula of cross correlation is:

$$r_m = \frac{n \sum R_r R_t - \sum R_r \sum R_t}{\sqrt{[n \sum R_r^2 - (\sum R_r)^2][n \sum R_t^2 - (\sum R_t)^2]}} \quad (2)$$

Where r_m is the cross-correlation coefficient; m is the band matching position. Generally, the value is - 20 to 20. When the value is 0, the band does not move; n is the number of bands where the two spectral curves coincide; R_r is the spectral collected by buoy; R_t is the UAV pixel spectral.

A continuous curve can be draw by the cross-correlation coefficients of all matching bands position [57]. The calibration is realized by expressing and comparing the difference between the reference spectral and the actual spectral. The calculation formula of difference degree is:

$$RMS = \sqrt{\frac{\sum_{-m}^m (R_m - r_m)^2}{k}} \quad (3)$$

Where RMS is the root mean square difference of cross-correlation coefficient; R_m is the cross-correlation coefficient curve of the buoy spectrum itself; r_m is the cross-correlation coefficient curve of buoy spectral and UAV pixel spectral; k is the calculation coefficient, and the value is $2m + 1$.

2.3. Absorbance characteristics recognition algorithm (ACR)

Spectral feature selection can be divided into unsupervised band selection method and supervised band selection method according to whether there is chemical tests data. The basic idea of unsupervised band selection method is statistical spectral indicators, such as variance, information entropy, signal-to-noise ratio and optimal index factor method. Estimate the importance of each band or between bands to the component content according to the differences between indicators. Generally, the reason why this method makes it difficult to improve the accuracy to a certain extent is the lack of specific purpose. The supervised band selection method achieves relatively better calculation accuracy on the basis of certain training samples. Methods include regression analysis, principal component analysis, partial least squares, support vector machine and neural network. The core purpose is to select a subset of bands with a number of D ($d < D$) from all wavelengths D of hyperspectral images by some search method, so as to maximize the evaluation criterion function, no matter which method is adopted.

An unsupervised band selection method for extracting water material content is designed. Absorbance reflects the sensitivity of each wavelength to water substances. The reflectance is converted to absorbance, and the proportional parameter of absorbance is set to 100 for the spectral data of each sampling point. The formula is:

$$A_i = |LOG_{10}^{R_i/100}| \quad (4)$$

Where, A_i is the absorbance value of band i ; R_i is the reflectance value of band i . The absorbance calculation results are brought into a new absorbance characteristic extraction algorithm for feature band extraction. The formula is:

$$S_i = [(A_i - SD_{A_i}) \times (A_i + AVG_{A_i})] / [(A_i + SD_{A_i}) \times (A_i - AVG_{A_i})] \quad (5)$$

Where, S_i is the calculated value of absorbance characteristics; A_i is the absorbance value of band i ; SD_{A_i} the absorbance standard deviation of the band i ; AVG_{A_i} is the average absorbance of the band i . It is considered that the first 30 bands with the highest absorbance contain the information of main pollutants in water quality according to the principle of unsupervised feature extraction. These bands are selected for calculating the content of water pollutants as potential independent variables.

The multiple linear regression technique in the supervised band selection method is used to establish the correlation between each band and content of the spectral at the sampling point. The formula is:

$$y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, X = \begin{bmatrix} 1 & x_{11} & \dots & x_{1i} \\ 1 & x_{21} & \dots & x_{2i} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_{n1} & \dots & x_{ni} \end{bmatrix}, \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_n \end{bmatrix}, \epsilon = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix} \quad (6)$$

Where, y_i is the chemical tests data of each sampling point; X is the spectral reflectance value of the corresponding test point, β is the band coefficient value and ϵ is the intercept value. The correlation coefficients are sorted, and the first 30 bands are also selected as the result of another characteristic band.

Comparing the results of unsupervised and supervised band selection methods, the overlapping bands are selected. These overlapping bands have an indicative relationship with the main indicators of water quality (Figure 5).

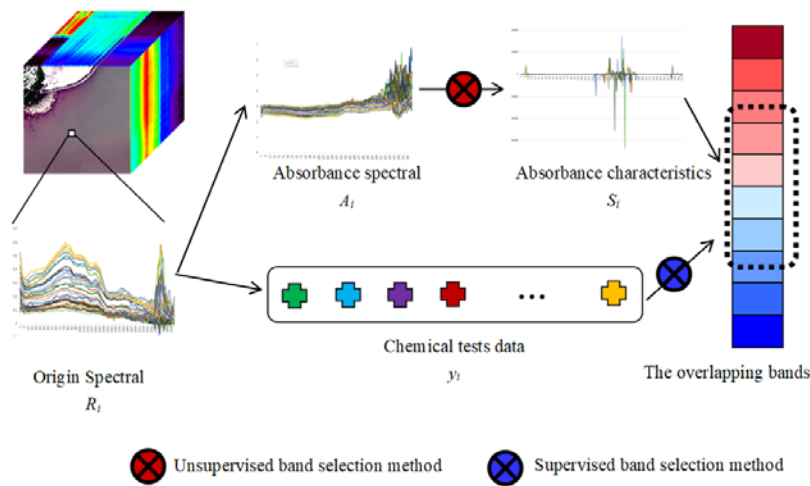


Figure 5. The flow of a recognition algorithm of absorbance characteristics. The characteristic bands selected by supervised method and unsupervised method are obtained through direct and indirect methods, and the overlapping bands are used as the effective bands.

2.4. 5x Dimensionality Reduction Algorithm

The uncertainty of information extraction caused by scale effect and the scale dependence of extraction accuracy must be considered in the calculation of surface parameters using hyperspectral remote sensing [58].

There are three main methods to obtain different scales remote sensing data. (1) The Sampling method, which expands the original image into a series of images with different resolutions through scale; (2) The multi-sensor method, which obtains the data of sensors with different resolutions in the same area, such as IKONOS pan 1m, SPOT pan 20m, TM 30m and MODIS 250m; (3) The variable altitude method, which obtains different resolution data of the same sensor by adjusting the flight altitude. The three methods have advantages and disadvantages. For example, the sampling method will lead to the unreliability of the subsequent conclusions; Due to the different spectral response functions of sensors in multi-sensor method, the work of unified standard will also cause computational complexity in evaluating scale effect; The variable altitude sensor is certain, so the data obtained with different resolutions have good comparability, but it is difficult to obtain. The improved sampling method is used to expand the spectral data from point data to five different levels of polygon data in this paper. Four adjacent pixels around the sampling point are taken as four scale levels. The number of pixels involved in the calculation is 1, 8, 16, 24 and 32 respectively (Figure 6). Take the spectral mean as the spectral value of each level.

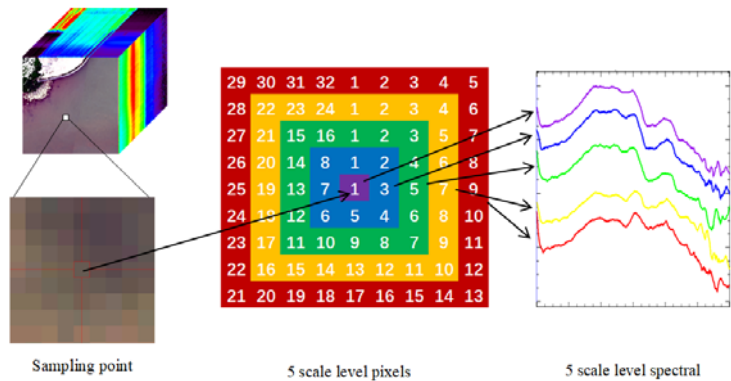


Figure 6. Spectral data sampling method at five scale levels.

There is a high correlation between adjacent bands of hyperspectral data [59]. A method integrating the advantages of hierarchical clustering and fuzzy clustering is designed to realize the rapid bands selection. The filtered band modeling can not only significantly improve the stability and prediction accuracy of the model, but also improve the extraction efficiency. Hierarchical clustering and fuzzy clustering algorithms are selected for feature band selection.

The steps of hierarchical clustering method are: (1) Calculate the distance between bands and combine the nearest bands into the same class; (2) Calculate the distance between classes and merge the nearest classes; (3) Repeat this process until all bands are merged into one class. The distance here is the Pearson correlation between bands. The greater the correlation, the smaller the distance, and merge. The steps of fuzzy clustering method are as follows: (1) The similarity matrix of the model is established according to the similarity coefficient method, and the value is between - 1 and 1; (2) The transitive closure is established, and different level cut sets are obtained by transforming the fuzzy equivalent matrix; (3) The fuzzy similarity matrix satisfying transitivity is clustered by setting different confidence levels. The corresponding clustering bands are combined to complete the evaluation of characteristic bands after the two kinds of clustering are realized.

2.5. Regression Models

Multiple linear regression method (MLR), support vector machine method (SVM) and neural network (NN) method are selected to establish the regression model between water quality parameters and characteristic bands in this paper. Generally, there is a linear correlation between water quality parameters and reflectance of characteristic band, which is suitable for modeling with multivariate linear model. It is necessary to introduce hyperplane to establish the regression relationship when the linear separability of the characteristic band decreases, the support vector can play a powerful role to further improve the regression accuracy. Neural network model is needed to participate in the calculation of a large amount of data, because support vector machine is only suitable for the task of small batch samples.

2.6. Model evaluation

R^2 (coefficient of determination), which reflects the accuracy of model fitting data and represents the proportion of variance explained by the model. The range is 0 to 1. The closer to 1, the stronger the explanatory ability of the variables of the equation to y , and the better the model fits the data. The closer to 0, the worse the model fits. For example, $R^2=0.6$ means that the model explains 60% of the uncertainty, and the model is acceptable. The R^2 coefficient calculation formula is as follows:

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (7)$$

Where, n is the sample size; y_i is the assay value of content of point i ; \hat{y}_i is the content prediction value of spectral method of point i ; \bar{y} is the mean of the assay value of the samples.

RMSE is the root mean square error in the same unit as the true value, which values range from 0 to infinity. For example, RMSE = 1 indicates that the average difference between the predicted value and the real value is 1. When the predicted value is completely consistent with the real value, it is equal to 0, that is, the perfect model; The greater the error, the greater the RMSE value, and the worse the model. The calculation formula is as follows:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (8)$$

Where, n is the sample size; y_i is the assay value of content of point i ; \hat{y}_i is the content prediction value of spectral method of point i .

3. Results and discussion

3.1. Space to ground matching results

Comparing the average reflectance of 10 UAV strips with 2 buoy spectrometers, it is concluded that UAV spectra have more burrs, and the reflectance is more affected by illumination change than buoy spectrometers. The two buoy spectrometers have good similarity and consistent spectral patterns (Figure 7). The reflectivity is mainly affected by the weak level of liquid level (such as waves). UAV data has great mutation in the first 5 bands and the last 30 bands, indicating that it should not be selected as the characteristic band in the subsequent modeling. The secondary calibration coefficient of each band spectrum is obtained according to the cross-correlation coefficient.

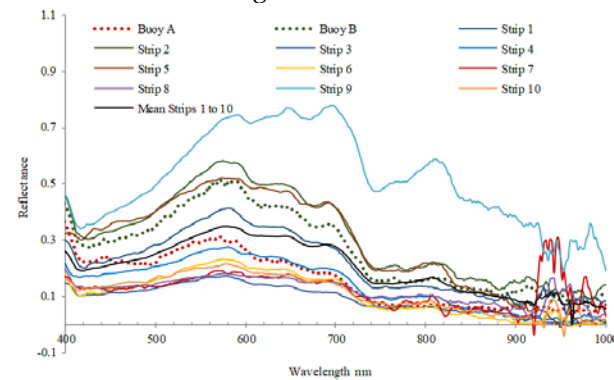
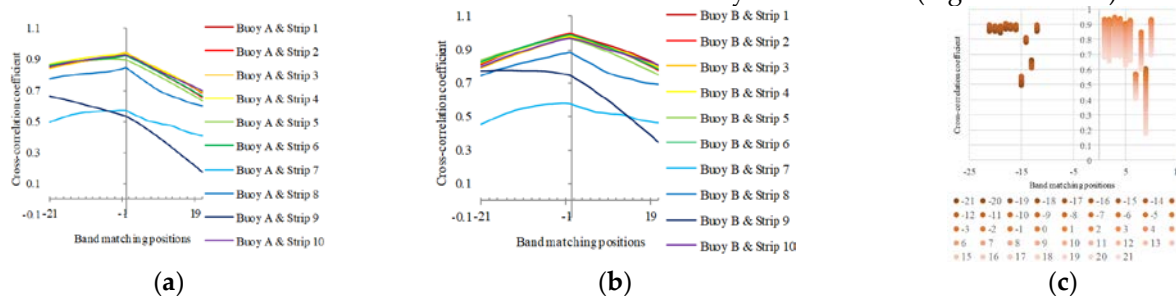


Figure 7. Comparison of mean reflectance between the data of two buoy spectrometers and 10 strips of UAV.

Draw the cross-correlation coefficient between the average reflectance of 10 UAV strips and buoys A & B, and the Fig. 7 reflects the change of correlation coefficient when the spectral of the two devices moves ± 21 . It can be concluded that, (1) The positions of reflectance peaks and valleys of UAV spectral and buoy spectral are highly consistent. The correlation shows a downward trend in both positive and negative directions (Figure 8 a and b); (2) The high correlation coefficients are -15 to -20, 0-6 and 10-13. When the cross-correlation coefficient moves in the positive (long wave) direction, the point value is more (Figure 8 c and d); (3) The cross-correlation coefficients of bands strips 7, strips 8 and strips 9 vary greatly, which shows that the spectral characteristics of these three bands are sensitive. When establishing the water quality calculation model, the characteristic bands selected on these three bands may not be robust (Figure 8 e and f).



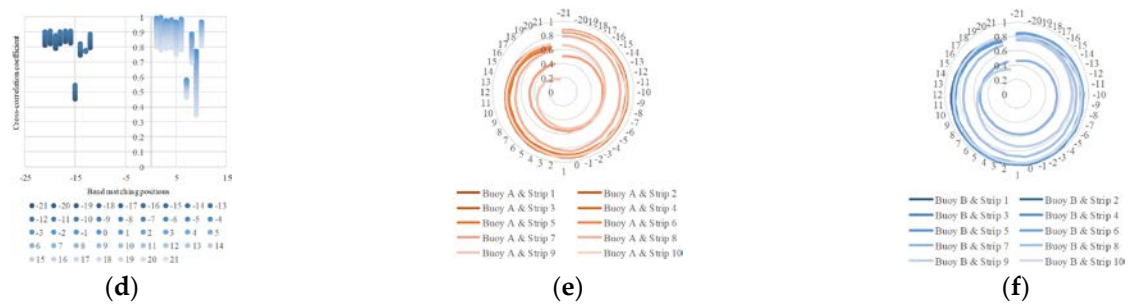


Figure 8. Positive and negative moving diagram, scatter diagram and radar diagram for cross-correlation coefficient between buoy spectrometer a and UAV spectral data. Figure (a) and figure (b) show that the central wavelength positions of the two sensors are basically the same, because the correlation coefficient shows a downward trend with the left and right shift of the wavelength. The important conclusion is that the characteristic band found by the buoy sensor on the water surface can be extended to UAV data. Figure (c) and figure (d) show that the correlation coefficients of the two sensors are largely distributed in several obvious intervals. When the central wavelength is negative, the correlation coefficient is more concentrated and the correlation degree is high. In the positive direction, the correlation coefficients are quite different. Figure (e) and figure (f) show the correlation coefficients of individual strips jump to a large extent with the movement of the central wavelength, which is likely due to the sudden change of light or shadow. These bands need to be eliminated during modeling, otherwise it may cause over fitting or under fitting.

3.2. Water quality parameters characterization band set

The reflectance data of 272 bands at each position are collected according to the longitude and latitude of the sampling point. On the hyperspectral images of strips 1 to 10, there are 11, 4, 4, 2, 2, 2, 2, 2, 5 and 2 valid data respectively. Buoy a and buoy B have 4 and 5 valid data respectively. So, a total of 45 groups of valid data are formed (Figure 9 a). It is concluded that the spectral data of the same strips have great similarity, indicating that the water quality with similar distance is also similar. The spectral of sampling points of different strips are significantly different, which is a very favorable phenomenon for subsequent modeling. The sensor has obvious noise at both ends, including 400 nm to 410 nm and 920nm to 1000 nm.

It is considered that as long as the wavelength of light is fixed, the absorption coefficient of the same substance will remain unchanged according to the principle that the absorption coefficient is related to the wavelength of incident light and the substance passed by light [60]. This phenomenon is very suitable to be used as the material content calculation. Take 10 as the base and 100 as the parameter to convert the absorbance of the spectrum to obtain the ratio of incident light to transmitted light on the water surface (Figure 9 b). It is concluded that the absorbance increases significantly with the increase of wavelength. The longer the wavelength, the more energy the water absorbs. If this trend is not maintained, it is caused by the material composition of the water body. The corresponding band can be selected to retrieve its material content.

Calculate the absorbance characteristic bands (Figure 9 c) according to formula (5), and sort the characteristic bands of each sampling point after calculating the absolute value. The spectral data corresponding to 45 sampling points have 45 sorting possibilities. The first 30 bands are selected as the final result of unsupervised characteristic band selection according to the principle of maximum simple addition value (Figure 9 d). It can be seen that there is no participation of any chemical test data in the whole process. The result of the calculation is: 785 nm, 747 nm, 727 nm, 781 nm, 774 nm, 787 nm, 725 nm, 776 nm, 783 nm, 803 nm, 805 nm, 809 nm, 754 nm, 778 nm, 678 nm, 730 nm, 794 nm, 758 nm, 772 nm, 798 nm, 745 nm, 736 nm, 790 nm, 750 nm, 421 nm, 743 nm, 741 nm, 674 nm, 767 nm, 416 nm are selected.

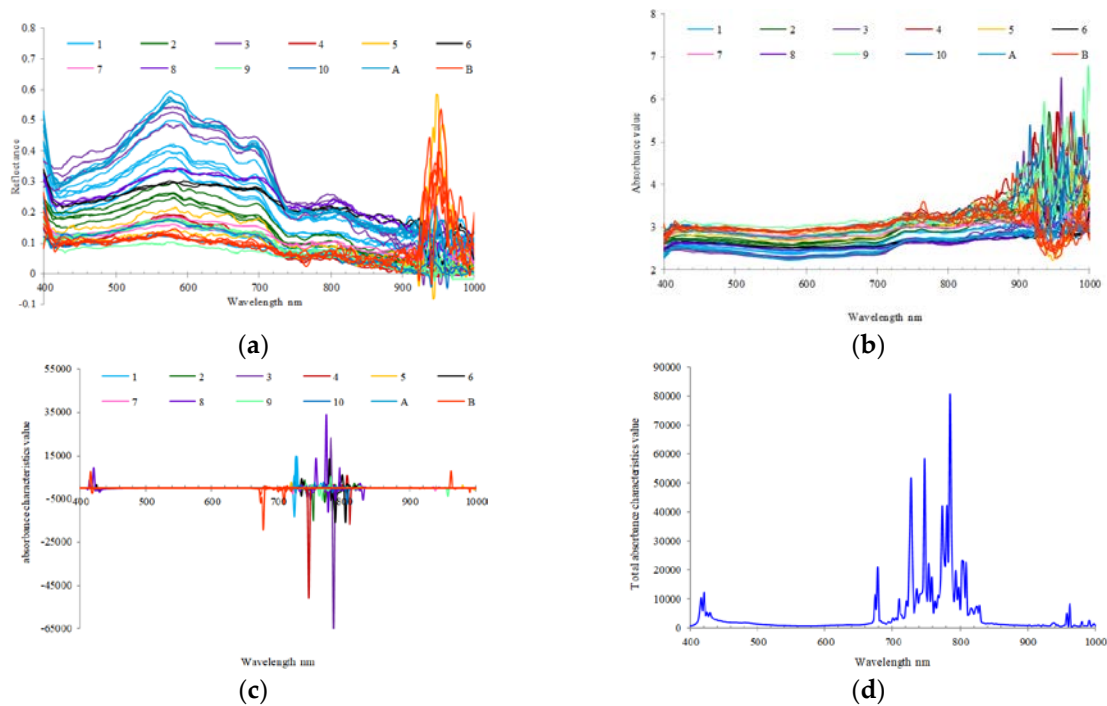


Figure 9. The spectral data, the absorbance data, the absorbance characteristics data and the total absorbance characteristics data of sampling points on each of the 10 strips and the spectral data of buoy A, B spectrometers. (a) The spectral data of sampling points; (b) The absorbance data of sampling points; (c) The absorbance characteristics data of sampling points; (d) The numerical ranking of 272 bands after passing the recognition algorithm of absorbance characteristics.

Analyze the correlation between the contents of five water quality parameters and the full wavelength to obtain the band number of the top 30 in positive correlation and negative correlation (Figure 10 a). Remove the first 10 bands (400nm to 420nm) and the last 30 bands (920nm to 1000nm) when selecting the characteristic band due to the interference of instrument noise. In general, the correlation coefficient of COD and chlorophyll is high, which reflects that the extraction accuracy may be higher. (1) There was a negative correlation between total phosphorus and all bands, and the correlation coefficient ranged from -0.116 to -0.460; (2) There was a negative correlation between total nitrogen and all bands, and the correlation coefficient ranged from -0.116 to -0.460; (3) COD showed a positive correlation with all bands, and the correlation coefficient ranged from 0.303 to 0.416; (4) Turbidity has a negative correlation with 420 nm to 700 nm, and a positive correlation with subsequent bands, with correlation coefficients ranging from -0.282 to 0.094; (5) Chlorophyll showed a positive correlation with all bands, and the correlation coefficient ranged from 0.078 to 0.384.

Overlay the characteristic bands selected by the correlation coefficient method with the characteristic bands selected by the unsupervised method (Figure 10 b). It is considered that the overlapping wavelength region can improve the calculation accuracy of water quality parameters to the greatest extent because it is selected by both supervised and unsupervised methods. The characteristic band sets of total phosphorus are 425nm to 434nm, with a total of 5 bands; The characteristic band sets of total nitrogen are 671nm-682nm and 694nm-711nm, with a total of 15 bands; The characteristic band set of COD is: 700nm, 722nm-736nm, 765nm-771nm, with a total of 12 bands; The characteristic band set of turbidity is: 427nm-434nm, 773nm-778nm, with a total of 7 bands; The characteristic bands of chlorophyll are 425nm-434nm, with a total of 3 bands.

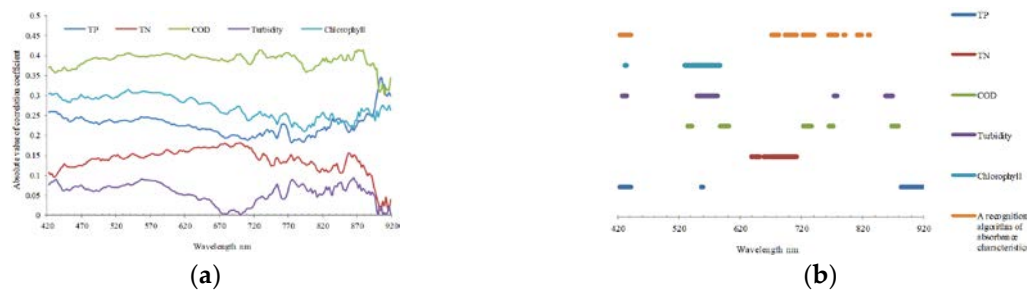


Figure 10. The water quality parameters characterization band set. (a) The absolute value of correlation coefficient between five water quality parameters and all bands; (b) The comparison chart of characteristic bands selected by supervised method and unsupervised method.

3.3. Response of sensitive bands to water quality content at different scales

The effect intensity of scale effect is preliminarily judged by cluster calculation. The clustering results of 272 bands in 5 scales are obtained according to the two algorithms Section 2.5. The results show that the category identification positions are 521 nm, 656 nm, 721 nm, 829 nm, 929 nm and 963 nm respectively (Figure 11). The results of clustering under different scales have great similarity, except fuzzy clustering at 16 scale. In addition, the similarity is also reflected in the merging of short wave and long wave with the change of wavelength at all scales. Although the red and blue band ranges in Fig. 11 are discontinuous, they can be aggregated into one kind of spectral. These phenomena imply that it has little effect on the extraction accuracy of water quality parameters under the current five scale divisions. The underlying reason why the scale effect can be ignored is that the spatial resolution of UAV hyperspectral is very high and the river channel is relatively narrow.

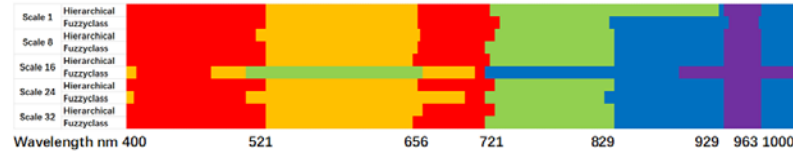


Figure 11. Clustering results of hierarchical clustering method and fuzzy clustering method at 5 scales. The same color in the figure indicates that the cluster is the same class and there are 5 categories in total.

The relatively best regression methods of different water quality indicators appear on different scales (Table 2). (1) ACR method has the highest R^2 value (0.6142) only in the calculation of total phosphorus although ACR method combines the characteristic bands selected by supervised and unsupervised methods. The RMSE value of ACR method is the smallest in chlorophyll calculation, but considering that R^2 is only 0.1431, it cannot be selected as the final calculation model. (2) Surprisingly, the MLR, SVM and NN methods did not reach the highest R^2 and lowest RMSE when calculating all water quality indicators at scale 1 after comparing the regression results of all five scales. On the one hand, it shows that only one pixel is selected in the quantitative calculation of hyperspectral data, which cannot represent the real situation of water environment; On the other hand, it is impossible to calculate the accurate water quality index because the selected pixel is not necessarily the point of collecting water samples due to the inherent error of GPS positioning (0.5-1m). (3) Scale 8 is a relatively balanced amount of data relative to the other four scales. The highest R^2 is reached in the calculation of total nitrogen, COD and turbidity, which is 0.7949, 0.6249 and 0.7105 respectively, and RMSE is also the lowest in all results, which shows a good calculation effect under this scale. (4) The calculation results of scale 16 and scale 24 are similar to that of scale 1. There are no higher R^2 and lower RMSE in the calculation results of the other three methods except that the RMSE of total phosphorus in scale 24 is 0.1741 (ranking first, but R^2 is only 0.3845) and the R^2 of total nitrogen in scale 16 is 0.7868 (ranking second). However, the reason for this phenomenon

is significantly different from scale 1. It is more because the typical characteristic position of reflectance is not significant, which is caused by excessive spectral averaging. (5) The R^2 of chlorophyll reached 0.6289, which was significantly higher than that of ACR and other four scales with the scale enlarged to 32. In addition, the R^2 of TN is also as high as 0.7662 (ranking third). The reason for this phenomenon is that chlorophyll is evenly dispersed and fully mixed in the water body. Similarly, TN is the collection of various nitrogen elements such as ammonia nitrogen, nitrogen and nitrogen oxide in water. With the scale enlargement, it can also extract more accurate results.

Table 2. Regression results for the water quality parameters.

Scale	Method	Accuracy	TP	TN	COD	Turbidity	Chlorophyll
1*	ACR	RMSE	0.2113	3.4244	3.9972	7.0520	0.0062
		R^2	0.6142	0.3201	0.1673	0.3054	0.1431
1	MLR	RMSE	0.1799	2.5217	3.7454	5.5209	0.6104
		R^2	0.3698	0.4276	0.2688	0.5742	0.0900
	SVM	RMSE	0.1858	2.9075	3.3585	7.5495	0.5614
		R^2	0.3684	0.2532	0.4274	0.2139	0.2638
	NN	RMSE	0.2024	2.9117	4.0375	6.7541	0.4725
		R^2	0.2026	0.2369	0.1502	0.3628	0.4546
8	MLR	RMSE	0.1820	1.0607	3.7279	3.9585	0.4787
		R^2	0.3277	0.7949	0.2866	0.7105	0.4431
	SVM	RMSE	0.1762	2.0915	3.5193	5.6504	0.4778
		R^2	0.4400	0.2078	0.3837	0.4796	0.4648
	NN	RMSE	0.2223	3.2793	2.6825	8.4609	0.5512
		R^2	0.0381	0.0320	0.6249	0.1279	0.2578
16	MLR	RMSE	0.1767	1.0815	3.8211	6.1119	0.4391
		R^2	0.3657	0.7868	0.2504	0.3100	0.5313
	SVM	RMSE	0.1867	1.9938	3.6679	6.3852	0.4730
		R^2	0.3241	0.2857	0.3376	0.3467	0.4798
	NN	RMSE	0.2218	3.2981	3.9968	8.4518	0.6190
		R^2	0.0422	0.0208	0.1673	0.0022	0.0640
24	MLR	RMSE	0.1741	2.3017	3.7519	5.3045	0.4488
		R^2	0.3845	0.0341	0.2774	0.4802	0.5104
	SVM	RMSE	0.1912	1.9701	3.6856	6.3527	0.4832
		R^2	0.2775	0.3201	0.3299	0.3429	0.4607
	NN	RMSE	0.2055	3.3055	4.0253	7.1815	0.5828
		R^2	0.1776	0.0165	0.1554	0.2796	0.1705
32	MLR	RMSE	0.1772	1.1327	3.8439	5.1770	0.3907
		R^2	0.3624	0.7662	0.2415	0.5049	0.6289
	SVM	RMSE	0.1866	1.9808	3.6631	6.4209	0.4725
		R^2	0.3215	0.2941	0.3408	0.3434	0.4813
	NN	RMSE	0.2189	3.3162	4.0816	8.4297	0.6094
		R^2	0.0673	0.0101	0.1316	0.0074	0.0929

* The ACR method has only scale 1 data.

Comparing ACR, MLR, SVM and NN4 calculation methods, the conclusions are as follows: (1) The ACR method of total phosphorus and the MLR method of total nitrogen, turbidity and chlorophyll reached the highest value of R^2 on the corresponding scale (Figure 12 a). The ACR method of chlorophyll and the MLR method of total phosphorus, total nitrogen, turbidity reached the minimum value of RMSE on the corresponding scale respectively (Figure 12 b); (2) SVM method does not reach the relative maximum of R^2 (Figure 12 c) and the relative minimum of RMSE (Figure 12 d) on all scales, which shows the shortcomings of this method; (3) The COD regression coefficient R^2 of NN method reaches the relative maximum (Figure 12 e) and the RMSE of COD calculated by NN method reaches a relative minimum (Figure 12 f) at scale 8, which indicates the best method and scale of COD.

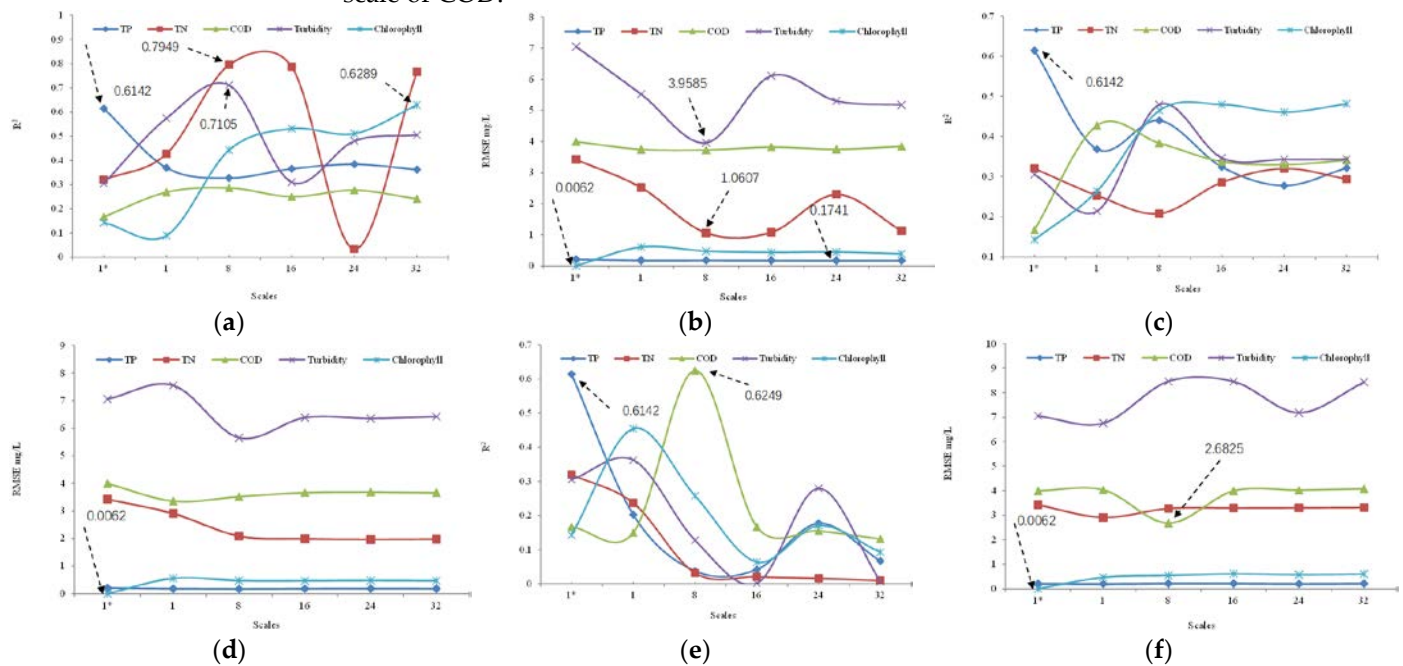


Figure 12. Comparison of regression results between ACR and MLR, SVM and NN methods. (a) R^2 values regressed by ACR method and MLR method at different scales; (b) RMSE values regressed by ACR method and MLR method at different scales; (c) R^2 values regressed by ACR method and SVM method at different scales; (d) RMSE values regressed by ACR method and SVM method at different scales; (e) R^2 values regressed by ACR method and NN method at different scales; (f) RMSE values regressed by ACR method and NN method at different scales.

3.4. Accuracy evaluation

According to the response of sensitive bands to water quality content at different scales (Section 3.3), the scale 1 data of ACR method is selected to calculate the total phosphorus content, the scale 8 data of MLR method is selected to calculate the total nitrogen and turbidity, the scale 8 data of NN method is selected to calculate the COD, and the scale 32 data of MLR method is selected to calculate the chlorophyll.

The accuracy of data is limited in terms of sampling points. The reason is that the number of sampling points is generally small, and there are individual extreme values. In addition, there is a certain inherent deviation in the spectral data because the hyperspectral data of UAV is obtained in 2 days. The accuracy of COD (Figure 13 a) and turbidity (Figure 13 b) is low comparing the calculation results of five water quality parameters. COD data generally needs to be obtained by testing for several consecutive days. The test data only includes single time data, which cannot reflect the actual situation of water quality COD. Turbidity should reflect the comprehensive situation within a certain water depth and thickness, which is difficult to calculate for hyperspectral data. The comparison

accuracy of total phosphorus, total nitrogen and chlorophyll are 0.6925 (Figure 13 c), 0.7291 (Figure 13 d) and 0.7658 (Figure 13 e) respectively, which is acceptable.

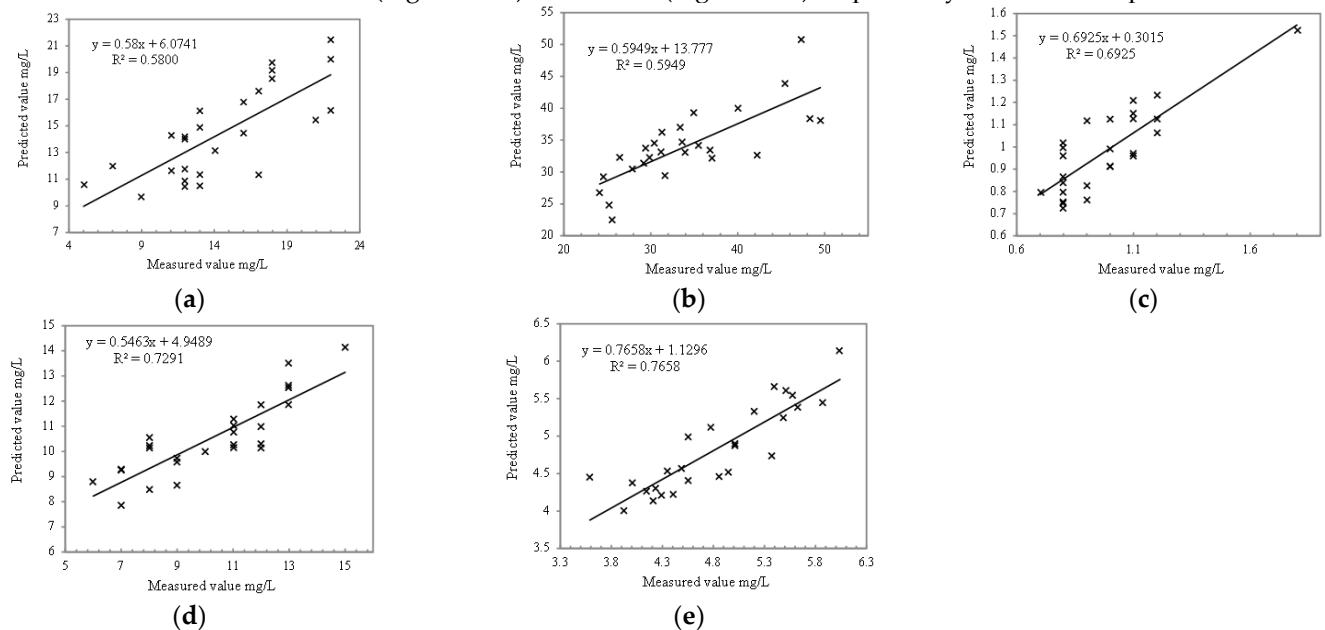


Figure 13. Comparison between the measured and predicted values of each water quality parameter in the modeling dataset. (a) Comparison between predicted and measured values of total phosphorus; (b) Comparison between predicted and measured values of total nitrogen; (c) Comparison between predicted value and measured value of COD; (d) Comparison between predicted value and measured value of Turbidity; (e) Comparison between predicted value and measured value of Chlorophyll.

3.5. Mapping and water quality evaluation

The river in the study area flows slowly from north to south, and the velocity is lower than 0.1m/s under normal conditions. Some river sections have weak backflow, and the overall hydrological situation is similar to that of inland lakes, which is conducive to the hyperspectral work. The results showed that the content of total phosphorus changed gently, ranging from 0.4061 mg/L to 2.0605 mg/L (Figure 14 a); The content of total nitrogen changed sharply, ranging from 0.1323 mg/L to 109.834 mg/L; The content of COD changes violently, ranging from 0.0251 mg/L to 48.327 mg/L; The content of turbidity changes very sharply, ranging from 1.8461 to 3248.68; The content of chlorophyll also changed sharply, ranging from 0.0878 mg/L to 338.2971 mg/L by calculating five water quality parameters of the river. The pollutant content of the whole river shows great difference. The reasons are as follows: on the one hand, the river channel is narrow (the narrowest part is less than 5m) and the flow velocity is slow, and many piers lead to the accumulation of pollutants; On the other hand, there are many urban commercial and domestic sewage outlets, and all kinds of pollutants show a sharp increase near the sewage outlets.

Four typical areas are selected, which are the starting point (No. 1), catchment (No. 2), direct flow (No. 3) and end point of the river (No. 4). Different areas show different laws (Figure 14 b). (1) The river presents the state of pollutant accumulation on the north bank due to the inflow of the upstream mainstream river at the starting point of the river. The other four pollutants increase significantly except that the law of total nitrogen is not significant. This phenomenon reflects that a large part of the pollutants in the river come from the upstream mainstream river; (2) The river channel leaks out of the ground again, and all kinds of pollutants show explosive growth under the combined action of chemistry and physics at the catchment. Moreover, the river here is narrow, which makes the water present the characteristics of typical black odor water body; (3) The river enters a down-

stream state of hundreds of meters, and the concentration of pollutants decreases significantly at the direct current. A pollutant strip appears in the west of the center of the river due to the action of water flow. Moreover, two circular high-value areas of pollutants can be seen, and it can be inferred that there are underwater sewage outlets at these two locations. It is speculated that there are two underwater sewage outlets, because two circular high-value areas of pollutants can be seen; (4) Various pollutants are fully diluted and reduced at the end of the river. On the one hand, there is a large area of open water in the downstream, which has a significant scouring effect. At the same time, the relative concentration of pollutants is significantly reduced after a certain distance of flow due to the river's degradation ability.

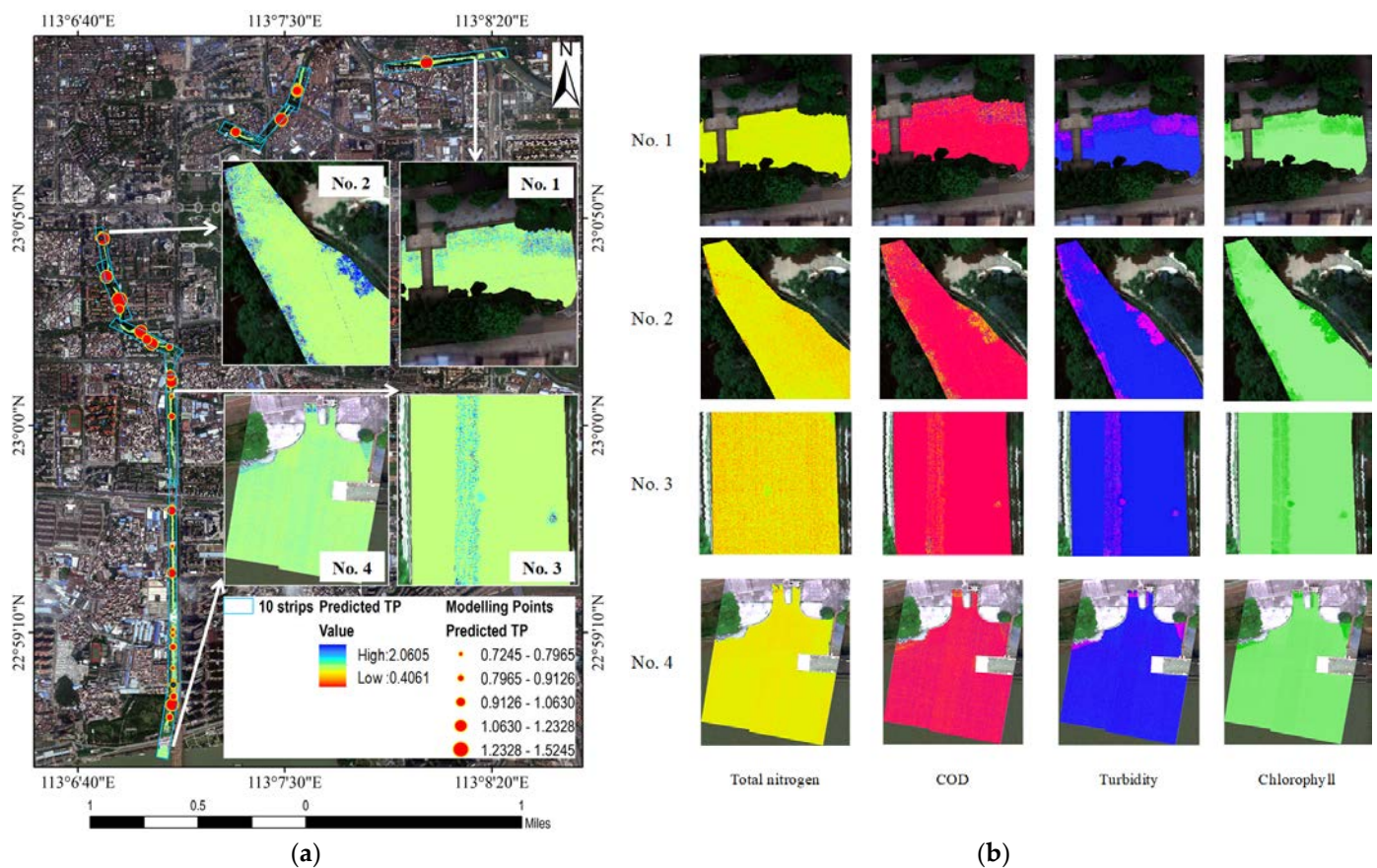


Figure 14. Calculation results of water quality parameters in the whole river and spatial distribution of five parameters in typical areas. (a) Calculation results of total phosphorus and content in four typical areas; (b) Contents of total nitrogen, COD, turbidity and chlorophyll in four typical areas.

The river hyperspectral image data are divided into downstream, midstream and upstream according to the distribution of 10 bands (Figure 15). The calculation shows that the content of total phosphorus in the upstream and midstream is low, ranging from 0.4061 mg/L to 1.6528 mg/L, and there is a high value in the upstream, reaching 2.0605 mg/L (Figure 16 a). The distribution of total nitrogen in the three river sections is close (Figure 16 b). The minimum value is 0.1323 mg/L in the downstream and the maximum value is 109.834 mg/L in the midstream. The COD content in the downstream reaches is significantly higher than that in the upstream and midstream, up to 48.327mg/l (Figure 16 c). The three river sections show a trend of gradual reduction of COD, which is in line with the objective law of COD. The turbidity in the midstream is significantly higher than that in the upstream and downstream, with a peak of 3248.68 (Figure 16 d). This river section combines all kinds of pollutants from the upstream. At the same time, the purifi-

cation capacity of the river has not played a significant role, resulting in such high turbidity. There is no significant watershed difference in the distribution of chlorophyll, but it has a great correlation with the content of total phosphorus and total nitrogen, reflecting the promotion effect on aquatic algae due to water eutrophication (Figure 16 e).

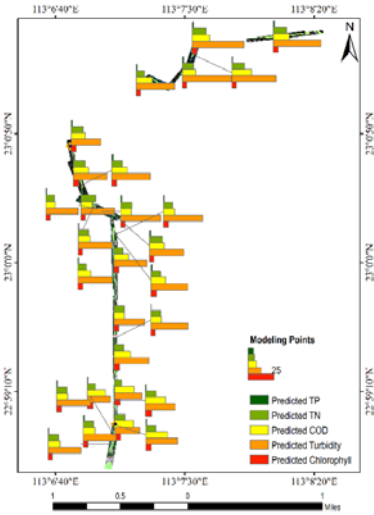


Figure 15. The prediction results of five water quality parameters at modeling points.

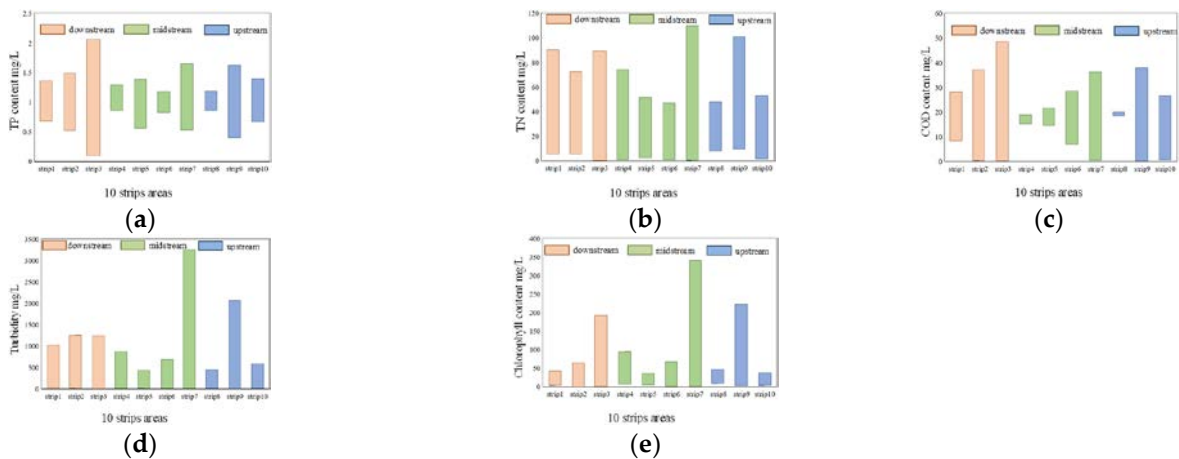


Figure 16. Distribution law of water quality parameters in the upstream, midstream and downstream of Lingnan Avenue River.

4. Conclusion

The future water environment monitoring work will show the characteristics of high data fusion of multiple platforms. In this paper, a new remote sensing monitoring mode of water quality is designed and implemented, that is, a buoy spectrometer with continuous working ability on the water surface and a flight platform for large-area synchronous monitoring in the air. The conclusions are as follows: (1) the data of the flight platform is limited by atmospheric interference, shadow and pixel resolution, which needs the calibration of the water surface spectrometer. The airborne spectral data will be more real through simple coefficient conversion, which is the fundamental guarantee for the calculation accuracy of water quality; (2) The traditional characteristic band selection method is based on the correlation between reflectivity and content. Although a large number of algorithm tests have been carried out, the applicability has been questioned due to the inherent limitations of water optical model. A band selection algorithm (ACR algorithm) with reflectivity related to content and strong absorbance characteristics is proposed, which improves the accuracy of calculation results to a certain extent, especially in the

extraction of total phosphorus and chlorophyll; (3) Spatial-spectral differences should be fully considered when comparing test data for hyperspectral data combination of spectroscopy and optical imaging. The reason is the best results of different water quality parameters appear on different scales. This scale effect has a certain relationship with the algorithm, which leads to the relatively complex problem. This paper makes a preliminary exploration. The research results not only have scientific reference significance for the processing and analysis of point and polygon hyperspectral data, but also provide a complete solution for the monitoring and treatment of small watershed rivers in urban areas.

Author Contributions: Donghui Zhang: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data Curation, Writing - Original Draft, Writing - Review & Editing. Lifu Zhang: Conceptualization, Methodology, Validation, Formal analysis, Investigation, Data Curation, Writing - Original Draft, Writing - Review, Editing, Project administration & Funding acquisition. Xuejian Sun: Software, Validation, Formal analysis, Investigation & Resource. Yu Gao, Ziyue Lan, Yining Wang and Haoran Zhai: Methodology, Software, Resource & Editing. Jingru Li, Wei Wang and Maming Chen: Software, Investigation & Data Curation. Xusheng Li, Liang Hou and Hongliang Li: Supervision, Formal analysis, Investigation & Data Curation.

Funding: Please add: This research was funded by the National Natural Science Foundation of China, grant number 41830108 and the National Natural Science Foundation of China, grant number 41977154.

Data Availability Statement: Not applicable.

Conflicts of Interest: The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. Zhang, L. F.; Zhang, L. S.; Sun, X. J.; Chen, J.; Wang, S.; Zhang, H.M. & Tong, Q. X. Spectral monitoring online system for water quality assessment based on satellite-ground data integration. *Journal of Global Change Data & Discovery*, **2021**, 5(1), 1-10. <https://doi.org/10.3974/geodp.2021.01.01>.
2. Arvor, D.; Betbeder, J.; Daher, F.; Blossier, T. & Junior, C. Towards user-adaptive remote sensing: knowledge-driven automatic classification of sentinel-2 time series. *Remote Sensing of Environment*, **2021**, 264(17), 112615. <https://doi.org/10.1016/j.rse.2021.112615>.
3. Brezonik, P. L.; Olmanson, L. G.; Finlay, J. C. & Bauer, M. E. Factors affecting the measurement of cdom by remote sensing of optically complex inland waters. *Remote Sensing of Environment*, **2015**, 157(Sp. Iss. SI), 199-215. <https://doi.org/10.1016/j.rse.2014.04.033>.
4. Flores, A.; Griffin, R.; Dix, M.; Romero-Oliva, C. S. & Barreno, F. Hyperspectral satellite remote sensing of water quality in lake atitlán, guatemala. *Frontiers in Environmental Science*, **2020**, 8. <https://doi.org/10.3389/fenvs.2020.00007>.
5. Yang, M. M.; Ishizaka, J.; Goes, J. I.; Gomes, H. D. R.; Maúre, Elígio de Raús & Hayashi, M. et al. Improved modis-aqua chlorophyll-a retrievals in the turbid semi-enclosed ariake bay, japan. *Remote Sensing*, **2018**, 10(9). <https://doi.org/10.3390/rs10091335>.
6. Jiaming, L.; Yanjun, Z.; Di, Y. & Xingyuan, S. Empirical estimation of total nitrogen and total phosphorus concentration of urban water bodies in china using high resolution ikonos multispectral imagery. *Water*, **2015**, 7(11), 6551-6573. <https://doi.org/10.3390/w7116551>.
7. Lavigne, H.; Zande, D.; Ruddick, K.; Santos, J. & Kratzer, S. Quality-control tests for oc4, oc5 and nir-red satellite chlorophyll-a algorithms applied to coastal waters. *Remote Sensing of Environment*, **2021**, 255(1-2), 112237. <https://doi.org/10.1016/j.rse.2020.112237>.
8. Liu, Y.; & Xiao, C. C. Water extraction on the hyperspectral images of gaofen-5 satellite using spectral indices. **2020**. <https://doi.org/10.5194/isprs-archives-XLIII-B3-2020-441-2020>.
9. Niroumandjadidi, M.; Bovolo, F. & Bruzzone, L. Water quality retrieval from prisma hyperspectral images: first experience in a turbid lake and comparison with sentinel-2. *Remote Sensing*. **2020**. <https://doi.org/10.3390/rs12233984>.
10. Rianza, A.; Buzzi, J.; Garcia-Melendez, E.; Carrere, V.; Sarmiento, A. & Mueller, A. Monitoring acidic water in a polluted river with hyperspectral remote sensing (hymap). *International Association of Scientific Hydrology Bulletin*, **2015**, 60(5-6), 1064-1077. <https://doi.org/10.1080/02626667.2014.899704>.
11. J Suomalainen; Oliveira, R. A.; Hakala, T.; Koivumki, N.; Markelin, L. & Nsi, R., et al. Direct reflectance transformation methodology for drone-based hyperspectral imaging. *Remote Sensing of Environment*, **2021**, 266. <https://doi.org/10.1016/j.rse.2021.112691>.
12. Guimarães Tainá, Veronez Mauricio; Emilie, K.; Luiz, G.; Fabiane, B. & Leonardo, I. An alternative method of spatial autocorrelation for chlorophyll detection in water bodies using remote sensing. *Sustainability*, **2017**, 9(3), 416. <https://doi.org/10.3390/su9030416>.

13. Niu, C.; Tan, K.; Jia, X. & Wang, X. Deep learning based regression for optically inactive inland water quality parameter estimation using airborne hyperspectral imagery. *Environmental Pollution*, **2021**, 117534. <http://doi.org/10.1016/j.envpol.2021.117534>.
14. Pahlevan, N.; Smith, B.; Binding, C.; Gurlin, D. & Giardino, C. Hyperspectral retrievals of phytoplankton absorption and chlorophyll-a in inland and nearshore coastal waters. *Remote Sensing of Environment*, **2020**, <https://doi.org/10.1016/j.rse.2020.112200>.
15. Rhb, A.; Ms, B.; Dd, A.; Rs, B.; Kqa, C. & Kb, B. Unmanned aerial system based spectroradiometer for monitoring harmful algal blooms: a new paradigm in water quality monitoring - sciencedirect. *Journal of Great Lakes Research*, **2019**, 45(3), 444-453. <http://creativecommons.org/licenses/by-nc-nd/4.0/>.
16. Wei; Huang; Wang; Zhou & Cao. Monitoring of urban black-odor water based on nemerow index and gradient boosting decision tree regression using uav-borne hyperspectral imagery. *Remote Sensing*, **2019**, 11(20), 2402. <https://doi.org/10.3390/rs11202402>.
17. Zhang, Y.; Wu, L.; Ren, H.; Liu, Y. & Dong, J. Mapping water quality parameters in urban rivers from hyperspectral images using a new self-adapting selection of multiple artificial neural networks. *Remote Sensing*, **2020**, 12(2), 336. <http://doi.org/10.3390/rs12020336>.
18. Aguzzi, J.; Albiez, J.; Flgel, S.; God, O. R. & Zhang, G. A flexible autonomous robotic observatory infrastructure for benthopelagic monitoring. *Sensors*, **2020**, 20(6), 1614. <https://doi.org/10.3390/s20061614>.
19. Warren, M. A.; Simis, S. & Selmes, N. Complementary water quality observations from high and medium resolution sentinel sensors by aligning chlorophyll- a and turbidity algorithms. *Remote Sensing of Environment*, **2021**, 265. <https://doi.org/10.1016/j.rse.2021.112651>.
20. Favali, P.; Beranzoli, L.; D 'Anna, G.; Gasparoni, F. & Finch, E. A fleet of multiparameter observatories for geophysical and environmental monitoring at seafloor. *Annals of geophysics*, **2006**, 49(2-3), 659-680. <https://doi.org/10.4401/ag-3126>.
21. J González; Herrera, J. L. & Varela, R. A. A design proposal of real-time monitoring stations: implementation and performance in contrasting environmental conditions. *Scientia Marina*, **2012**, 76S1(S1), 235-248. <https://doi.org/10.3989/scimar.03620.191>.
22. Arabi, B.; Salama, M. S.; Pitarch, J. & Verhoef, W. Integration of in-situ and multi-sensor satellite observations for long-term water quality monitoring in coastal areas. *Remote Sensing of Environment*, **2020**, 239, 111632-. <https://doi.org/10.1016/j.rse.2020.111632>.
23. Banzon, V.; Smith, T. M.; Chin, T. M.; Liu, C. & Hankins, W. A long-term record of blended satellite and in situ sea-surface temperature for climate monitoring, modeling and environmental studies. *Earth System Science Data*, **2016**, 8,1(2016-04-28), 8(1), 165-176. <https://doi.org/10.5194/essd-8-165-2016>.
24. Kaire, T.; Tiit, K.; Alo, L.; Margot, S.; Birgot, P. & Tiina, N. First experiences in mapping lake water quality parameters with sentinel-2 msi imagery. *Remote Sensing*, **2016**, 8(8), 640-. <https://doi.org/10.3390/rs8080640>.
25. Vassiliki, M.; Dionissios, K.; George, P. & Elias, D. An appraisal of the potential of landsat 8 in estimating chlorophyll-a, ammonium concentrations and other water quality indicators. *Remote Sensing*, **2018**, 10(7), 1018. <http://doi.org/10.3390/rs10071018>.
26. Page, B. P.; Olmanson, L. G. & Mishra, D. R. A harmonized image processing workflow using sentinel-2/msi and landsat-8/oli for mapping water clarity in optically variable lake systems. *Remote Sensing of Environment*, **2019**, 231, 111284-. <https://doi.org/10.1016/j.rse.2019.111284>.
27. Zhu, X.; Cai, F.; Tian, J. & Williams, T. Spatiotemporal fusion of multisource remote sensing data: literature survey, taxonomy, principles, applications, and future directions. *Remote Sensing*, **2018**, 10(4), 527. <https://doi.org/10.3390/rs10040527>.
28. Chunmei, Cheng; Yuchun, Wei; Guonian, & Ning. Remote sensing estimation of chlorophyll-a concentration in taihu lake considering spatial and temporal variations. *Environmental monitoring and assessment*. **2019**. <https://doi.org/10.1007/s10661-018-7106-4>.
29. Hongbin, Liu; Dan, Jia-Huan; Sun & Da-Wen. Applications of imaging spectrometry in inland water quality monitoring-a review of recent developments. *Water, air and soil pollution*. **2017**. <https://doi.org/10.1007/s11270-017-3294-8>.
30. Suel, E.; Bhatt, S.; Brauer, M.; Flaxman, S. & Ezzati, M. Multimodal deep learning from satellite and street-level imagery for measuring income, overcrowding, and environmental deprivation in urban areas. *Remote Sensing of Environment*, **2021**, 257, 112339. <https://doi.org/10.1016/j.rse.2021.112339>.
31. Xl, A.; Feng, L. A.; Gmf, B.; Dsb, B.; Lai, J. C. & Yz, A. Monitoring high spatiotemporal water dynamics by fusing modis, landsat, water occurrence data and dem - sciencedirect. *Remote Sensing of Environment*, **2021**, 265. <https://doi.org/10.1016/j.rse.2021.112680>.
32. Hestir, E. L.; Brando, V. E.; Bresciani, M.; Giardino, C.; Matta, E. & Villa, P. Measuring freshwater aquatic ecosystems: the need for a hyperspectral global mapping satellite mission. *Remote Sensing of Environment*, **2015**, 167, 181-195. <https://doi.org/10.1016/j.rse.2015.05.023>.
33. Jongcheol, P.; Yakov, P.; Sang-Soo, B.; Yongseong, K.; Minjeong, K. & Hyuk, L. Optimizing semi-analytical algorithms for estimating chlorophyll-a and phycocyanin concentrations in inland waters in korea. *Remote Sensing*, **2017**, 9(6), 542. <https://doi.org/10.3390/rs9060542>.
34. Karimi, Y.; Prasher, S. O.; Mcnairn, H.; Bonnell, R. B.; Dutilleul, P. & Goel, P. K. Discriminant analysis of hyperspectral data for assessing water and nitrogen stresses in corn. *Transactions of the Asae*, **2005**, 48(2), 805-813. <https://doi.org/10.13031/2013.18303>.
35. Ryan, K. & Ali, K. Application of a partial least-squares regression model to retrieve chlorophyll-a concentrations in coastal waters using hyper-spectral data. *Ocean Science Journal*, **2016**, 51(2), 209-221. <https://doi.org/10.1007/s12601-016-0018-8>.
36. Sarigai, Yang; J.; Zhou, A.; Han, L. & Xie, Y. Monitoring urban black-odorous water by using hyperspectral data and machine learning. *Environmental Pollution*, **2021**, 269(10), 116166. <https://doi.org/10.1016/j.envpol.2020.116166>.

37. Dekker, A. G.; Hoogenboom, H. J.; Goddijn, L. M. & Malthus, T. J. M. The relation between inherent optical properties and reflectance spectra in turbid inland waters. *Remote Sensing Reviews*, **1997**, 15(1-4), 59-74. <https://doi.org/10.1080/02757259709532331>.
38. Alizadeh, M. J. & Kavianpour, M. R. Development of wavelet-ann models to predict water quality parameters in hilo bay, pacific ocean. *Marine Pollution Bulletin*, **2015**, 98(1-2), 171-178. <https://doi.org/10.1016/j.marpolbul.2015.06.052>.
39. Chua, C. G. & Goh, A. T. C. A hybrid Bayesian back-propagation neural network approach to multivariate modelling. *International Journal for Numerical and Analytical Methods in Geomechanics*, **2003**, 27(8), 651-667. <https://doi.org/10.1002/nag.291>.
40. Dsfj, A.; Hl, A.; Cj, A.; Dd, B.; Jd, C. & Ab, D. A three-step semi analytical algorithm (3saa) for estimating inherent optical properties over oceanic, coastal, and inland waters from remote sensing reflectance - sciencedirect. *Remote Sensing of Environment*, **2021**, 263. <https://doi.org/10.1016/j.rse.2021.112537>.
41. Salem, S. I.; Higa, H.; Kim, H.; Kazuhiro, K. & Oki, T. Multi-algorithm indices and look-up table for chlorophyll-a retrieval in highly turbid water bodies using multispectral data. *Remote Sensing*, **2017**, 9(6), 556. <https://doi.org/10.3390/rs9060556>.
42. Tung-Ching. A study of a matching pixel by pixel (mpp) algorithm to establish an empirical model of water quality mapping, as based on unmanned aerial vehicle (uav) images. *International journal of applied earth observation and geoinformation*, **2017**, 58, 213-224. <https://doi.org/10.1016/j.jag.2017.02.011>.
43. Turkoglu, M. O.; D'Aronco, S.; Perich, G.; Liebisch, F.; Streit, C. & Schindler, K. Crop mapping from image time series: deep learning with multi-scale label hierarchies. **2021**. <https://doi.org/10.1016/j.rse.2021.112603>.
44. Jing, Z.; Hui, W. B.; Yw, B.; Qin, Z. B. & Yla, B. Deep network based on up and down blocks using wavelet transform and successive multi-scale spatial attention for cloud detection. *Remote Sensing of Environment*, **2021**, 261. <https://doi.org/10.1016/j.rse.2021.112483>.
45. Chen, F.; Xiao, D. & Li, Z. Developing water quality retrieval models with in situ hyperspectral data in poyang lake, china. *Geo-Spatial Information Science*, **2016**, 19(4), 255-266. <https://doi.org/10.1080/10095020.2016.1258201>.
46. Gurlin, D.; Gitelson, A. A. & Moses, W. J. Remote estimation of chl-a concentration in turbid productive waters — return to a simple two-band nir-red model?. *Remote Sensing of Environment*, **2011**, 115(12), 3479-3490. <https://doi.org/10.1016/j.rse.2011.08.011>.
47. James, B. & Tsai, S. Optimization of a semi-analytical algorithm for multi-temporal water quality monitoring in inland waters with wide natural variability. *Remote Sensing*, **2015**, 7(12), 16623-16646. <https://doi.org/10.3390/rs71215845>.
48. Pyo, J. C.; Yong, S. K.; Min, J. H.; Nam, G. & Park, Y. Effect of hyperspectral image-based initial conditions on improving short-term algal simulation of hydrodynamic and water quality models. *Journal of Environmental Management*, **2021**, 294(3), 112988. <https://doi.org/10.1016/j.jenvman.2021.112988>.
49. GITELSON. The peak near 700 nm on radiance spectra of algae and water : relationships of its magnitude and position with chlorophyll concentration. *International Journal of Remote Sensing*, **1992**, 13(17), 3367-3373. <https://doi.org/10.1080/01431169208904125>.
50. Cui, T.; Jie, Z.; Jing, L.; Lim, B. & Roslinah, S. Hyperspectral water quality retrieval model: taking malaysia inshore sea area as an example. *International Society for Optics and Photonics*. **2007**. <https://doi.org/10.1117/12.750915>.
51. Mbuh & Mbongowo, J. Optimization of airborne real-time cueing hyperspectral enhanced reconnaissance (archer) imagery, in situ data with chemometrics to evaluate nutrients in the shenandoah river, virginia. *Geocarto International*, **2017**, 1-24. <http://dx.doi.org/10.1080/10106049.2017.1343395>.
52. Jouann Ea U, S.; Reroutes, L.; Durand, M. J.; Boukabache, A.; Picot, V. & Primault, Y. Methods for assessing biochemical oxygen demand (bod): a review. *Water Research*, **2014**, 49(feb.1), 62-82. <https://doi.org/10.1016/j.watres.2013.10.066>.
53. Song, K.; Li, L.; Li, S.; Tedesco, L.; Hall, B. & Li, L. Hyperspectral remote sensing of total phosphorus (tp) in three central indiana water supply reservoirs. *Water Air & Soil Pollution*, **2012**, 223(4), 1481-1502. <https://doi.org/10.1007/s11270-011-0959-6>.
54. Soppa, M. A.; Silva, B.; Steinmetz, F.; Keith, D. & Bracher, A. Assessment of polymer atmospheric correction algorithm for hyperspectral remote sensing imagery over coastal waters. *Sensors*, **2021**, 21(12), 4125. <https://doi.org/10.3390/s21124125>.
55. Ramoelo, A.; Skidmore, A. K.; Schlerf, M.; Mathieu, R. & Heitkonig, I. M. A. Water-removed spectra increase the retrieval accuracy when estimating savanna grass nitrogen and phosphorus concentrations. *Isprs Journal of Photogrammetry & Remote Sensing*, **2011**, 66(4), 408-417. <https://doi.org/10.1016/j.isprsjprs.2011.01.008>.
56. F. Van der Meer. Spectral Curve Shape Matching with a Continuum Removed CCSM Algorithm. *International Journal of Remote Sensing*, **2000**, 21(16), 3179-3185. <http://doi.org/10.1080/01431160050145063>.
57. Yu, X.; Yi, H.; Liu, X.; Wang, Y.; Liu, X. & Zhang, H. Remote-sensing estimation of dissolved inorganic nitrogen concentration in the bohai sea using band combinations derived from modis data. *International Journal of Remote Sensing*, **2016**, 37(2), 327-340. <http://dx.doi.org/10.1080/01431161.2015.1125555>.
58. Xiong, J.; Lin, C.; Ma, R. & Cao, Z. Remote sensing estimation of lake total phosphorus concentration based on modis: a case study of lake hongze. *Remote Sensing*, **2019**, 11(17), 2068. <https://doi.org/10.3390/rs11172068>.
59. Cannistra, A. F.; Shean, D. E. & Cristea, N. C. High-resolution cubesat imagery and machine learning for detailed snow-covered area. *Remote Sensing of Environment*, **2021**, 258, 112399. <https://doi.org/10.1016/j.rse.2021.112399>.
60. Rahman, H. A.; Harun, S. W.; Yasin, M. & Ahmad, H. Fiber optic salinity sensor using beam-through technique. *Optik - International Journal for Light and Electron Optics*, **2013**, 124(8), 679-681. <https://doi.org/10.1016/j.jileo.2012.01.020>.