water

MDPI

Article

Study on Parameter Inversion Model Construction and Evaluation Method of UAV Hyperspectral Urban Inland Water Pollution Dynamic Monitoring

Jiaqi Chen [1,2], Jinyu Wang [1,2], Shulong Feng [1], Zitong Zhao [1,2], Mingjia Wang [1], Ci Sun [1], Nan Song [1] and Jin Yang [1,*]

[1] Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences, Changchun 130033, China
[2] University of Chinese Academy of Sciences, Beijing 100049, China
* Correspondence: yang_jin1988@163.com

Abstract: The problem of environmental water pollution is becoming increasingly important. Inland rivers and lakes form interconnected water networks with fragile water ecosystems, and urban water pollution problems occur frequently. Chemical oxygen demand (COD), dissolved oxygen (DO), total phosphorus (TP), total nitrogen (TN), and ammonia nitrogen (NH3-N) in inland rivers are important indicators to evaluate water health quality. Timely and accurate reflection of dynamic changes to the key indices of urban river health status are of vital practical significance to adjust water treatment policy and ensure the stability of the aquatic environment and people's health. This study used COD, DO, TP, TN and NH3-N as typical water quality parameters for a reservoir in Guangxi Province, China and established a set of standardized processes covering UAV hyperspectral sampling and ground spectral correction, spectral data preprocessing, and modeling. In combination with machine learning and statistical analysis, an inversion method for measuring urban inland water pollution from UAV hyperspectral imaging with different dynamic monitoring parameters was proposed. And we compared the different combinations of preprocessing algorithm-regression algorithm and dimensionality reduction algorithm to get a unified model for quantitative estimation of water quality parameter concentration. We evaluated the performance of the proposed model according to root mean square error (RMSE), mean absolute error (MAE), mean absolute percentage error (MAPE), and coefficient of determination (R^2). The experimental results showed that our model was superior to other algorithms in RMSE, MAE, MAPE, and R^2. The MAPE of this model ranged from 0.01 to 0.12 and R^2 ranged from 0.84 to 0.98 in all water quality parameters. In general, this study provides an effective tool for decision-makers to investigate the source and physical mechanism of water pollution and establish a graded water quality evaluation model.

Keywords: hyperspectral imaging; UAV; inland river; water quality monitoring; model building

Citation: Chen, J.; Wang, J.; Feng, S.; Zhao, Z.; Wang, M.; Sun, C.; Song, N.; Yang, J. Study on Parameter Inversion Model Construction and Evaluation Method of UAV Hyperspectral Urban Inland Water Pollution Dynamic Monitoring. *Water* **2023**, *15*, 4131. https://doi.org/10.3390/w15234131

Academic Editors: Yung-Tse Hung and Laura Bulgariu

Received: 6 October 2023
Revised: 26 October 2023
Accepted: 17 November 2023
Published: 29 November 2023

1. Introduction

Owing to the influence of natural factors and human activities, the water quality of many inland rivers has been severely degraded, which leads to the need for water quality monitoring technology. Traditional water quality monitoring uses manual sampling and testing methods, which are low in monitoring efficiency and consume substantial human and material resources. In addition, isolated monitoring points cannot accurately describe the water quality of the whole water area, nor can they obtain the spatial distribution of water quality. These limitations make it difficult to meet the monitoring needs [1] of modern urban governance systems. Remote sensing of water quality, as an auxiliary to the existing monitoring system, can obtain multi-scale and multi-temporal information on rivers and lakes and help track the location of possible pollution sources. Effective monitoring of water quality changes is important for the protection of urban rivers. In recent years,

water quality monitoring models using remote sensing technology have been widely used in the quantitative estimation [2,3] of a variety of water quality parameters, including chemical oxygen demand (COD), dissolved oxygen (DO), total phosphorus (TP), total nitrogen (TN) and ammonia nitrogen (NH3-N). All abbreviations in the text are shown in Abbreviations Index. The accumulation of these elements leads to water eutrophication, changes in turbidity, and increases in total suspended matter content. They directly lead to the rapid accumulation of chlorophyll, which substantially [4] disturbs urban river ecosystems. Therefore, real-time and accurate monitoring of water quality changes, as well as determining pollution sources and mechanisms that cause water quality changes, are crucial to solving the problem of urban inland water pollution and are the basis and priority for protecting the water environment. In recent years, with the rapid development of remote sensing technology, inversion of water quality parameter changes using remote sensing has become a common means of water quality monitoring. Satellite remote sensing technology can effectively monitor large urban rivers, but most urban rivers are less than 100 m in width and belong to the category of small and medium-sized rivers. Village and town rivers are usually less than 10 m in width. There is a dense, widely distributed network of these small and medium-sized rivers, which often have very silted channels with slow flow rates, and illegal and unplanned discharge from the surrounding environment. Therefore, the water quality is prone to deterioration, and there is a high incidence of river pollution. These conditions increase the need for high spatial and temporal resolution remote sensing data. The existing satellite-borne remote sensing monitoring methods often cannot meet the monitoring needs of modern urban river and lake water environments. In terms of analysis, a variety of methods have been used to monitor water quality, including deep learning methods such as empirical methods, convolutional neural networks (CNN), deep convolutional generation and adversarial networks, and Bayesian neural networks. Most of these grade [5–8] water quality according to the concentration of water quality parameters and achieve relatively good performance. However, the spatial resolution of the above studies is relatively poor, and the water quality cannot be monitored in a small range. Therefore, it cannot be applied to the monitoring of water pollution in small inland water bodies in cities. Some researchers, such as Bonansea et al., have studied satellite data to monitor the change in water quality parameters through polynomial regression modeling, using the relationship between spectra. Similarly, Gu et al. used two GF-5 hyperspectral images to estimate chlorophyll a concentration and suspended sediment concentration in the Yangtze River Estuary and [9–11] obtained a prediction accuracy of 0.815. However, this kind of method requires many data sets as training data to obtain relatively good prediction results, which results in a long data collection period and limited accuracy. By improving the traditional CNN research for the prediction of COD concentration using hyperspectral remote sensing and a gated recurrent neural network (GRNN), WANG et al. achieved good modeling performance [12,13]. However, their data were simulated in the laboratory which may not have practical significance. In recent years, researchers have studied the use of low-altitude remote sensing technology from unmanned aerial vehicles (UAV), and have used UAV to monitor chlorophyll a. Jung Min Ahn achieved good results in predicting cyanobacteria blooms in controlled rivers using hyperspectral images combined with machine learning [14–19]. In summary, there are few studies on inversion methods for UAV hyperspectral dynamic monitoring parameters for urban inland water pollution in China and elsewhere. And the analysis algorithm is inefficient. In view of this, remote sensing inversion models of spectral features of water quality parameters were constructed by taking COD, DO, TP, TN and NH3-N for five typical water quality parameters in Beihai City, Guangxi Province as research objects through different pretreatment and regression methods. In this study, the water pollution index in a reservoir was quantitatively and dynamically monitored by mining the spectral information in the wavelength range of 400–1000 nm obtained by the UAV hyperspectral imaging system. And we compared different combinations of pre-processing algorithm, regression algorithm and dimensionality reduction algorithm, and finally establishes a

unified model of quantification of the concentration of the water quality parameters to be used for the effective quantitative estimation of the concentration of water quality parameters, so as to achieve the quantitative and dynamic monitoring of water pollution indexes in the water reservoirs.The specific studies were to: (1) obtain the spectral data of the hyperspectral image in the range of 400–1000 nm; (2) eliminate the interference from the instrument itself and the surrounding environment by selecting suitable spectral preprocessing; (3) find different combinations of preprocessing and regression algorithms for different monitoring indicators, and study ways of evaluating the performance of the models in the process; (4) through three methods—principal component analysis (PCA), continuous projection algorithm (SPA) and simulated annealing algorithm (SAA)—to select the best wavelength combination to optimize the inversion model and improve inversion efficiency. This approach can achieve the rapid monitoring of water pollution indicators for COD, DO, TP, TN and NH3-N, to establish a comprehensive analysis model.

2. Materials and Methods

2.1. Study Area Overview

In this study, a reservoir in Guangxi Province was selected as the research area. The reservoir is a centralized drinking water source at the Beihai level, located on the Sanhe River, with a normal reservoir capacity of 11.53 million m^3 (Figure 1). Analysis of the reservoir is helpful to understand the ecological environment status of the Beijiao water plant and provide for the safety of the urban water supply. Hyperspectral images were taken at multiple points in five different water sections in the reservoir, including the village on the lower slope, the Poxinling, the Caohualing, the head of the dam, and the reservoir (Table 1). At the same time, ground spectral data were collected by a micro-optical fiber spectrometer in the same water area. Water samples were collected using a cup-type fixed- depth water quality sampler in the experimental area. There were 20 sampling points in a single experiment, and 500 mL water samples were collected at each point. The samples were sent to professional institutions for testing according to the national standard, and the measured data of the detected content of COD, DO, TP, TN and NH3-N were used as the standard for subsequent spectral modeling.

Figure 1. Map of distribution of sampled waters.

Table 1. Water quality monitoring at the center point of the reservoir basin.

Serial Number	Section Name	Latitude (N)	Longitude (E)
1	In the library	21.596800900°	109.232536650°
2	Downhill Village	21.607201300°	109.237450290°
3	Caohualing Village	21.60265254°	109.240883460°
4	Slope Heart Ridge	21.612006920°	109.230904750°
5	Dam Head	21.589200229°	109.229440652°

2.2. UAV Hyperspectral Data Acquisition and Preprocessing

In this study, we used an unmanned airborne hyperspectral imaging system integrated by Hangzhou Hyperspectral Imaging Technology Co., Ltd. (Hangzhou, China) as shown in Figure 2b. The core spectral module of the system was independently developed by CIOMP.

(**a**) (**b**)

Figure 2. Experimental instrument: (**a**) UAV hyperspectral imaging system, and (**b**) miniature fiber optic spectrometer.

The spectral range covered 400–1000 nm, the spectral resolution was better than 3 nm, and the focal length of the lens (kowa) was 35 mm. The spectral range of the miniature optical fiber spectrometer covered 200–1100 nm, and the spectral resolution was better than 1.5 nm. The hyperspectral image of the waters of Niuweiling, Guangxi was collected in reflection mode, and the spectral curve of the water body was obtained, as shown in Figure 3. The processing process is shown in Figure 4.

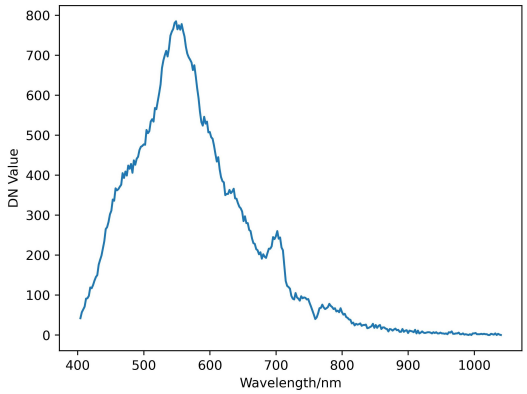


Figure 3. Spectral reflectance curve of the water body.

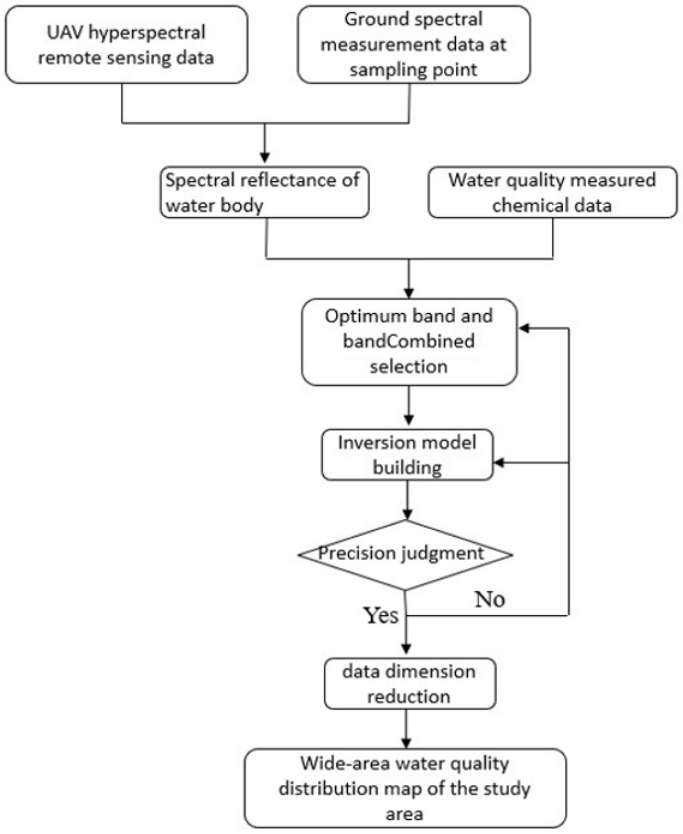


Figure 4. Water sample data processing flow.

2.3. ROI Selection and Pretreatment Method

The experiment time was between 10:00 and 14:00 in the day. During the experiment, a high reflectivity image was obtained by scanning the gradient reflectivity whiteboard, and the hyperspectral image data of five different water areas in the reservoir were collected. The reflectivity of the reservoir water was corrected using Formula (1):

$$I_R = (I_O - I_B)/(I_W - I_B) \tag{1}$$

where, I_R represents the corrected reflectivity hyperspectral image data, I_O represents the original water body hyperspectral image data, I_W represents the reflectivity whiteboard hyperspectral image data taken under the same experimental environment, I_B represents the reflectivity dark plate hyperspectral image data taken in the same experimental environment.

The micro-optical fiber spectrometer self-developed by CIOMP was used to correct the hyperspectral image data according to Formula (2).

$$R = \begin{pmatrix} r_{1,1} & r_{1,2} & r_{1,3} & \cdots & r_{1,n} \\ r_{2,1} & r_{2,2} & r_{2,3} & \cdots & r_{2,n} \\ r_{3,1} & r_{3,2} & r_{3,3} & \cdots & r_{3,n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ r_{m,1} & r_{m,2} & r_{m,3} & \cdots & r_{m,4} \end{pmatrix} \rightarrow \begin{pmatrix} r'_{1,1} & r'_{1,2} & r'_{1,3} & \cdots & r'_{1,n} \\ r'_{2,1} & r'_{2,2} & r'_{2,3} & \cdots & r'_{2,n} \\ r'_{3,1} & r'_{3,2} & r'_{3,3} & \cdots & r'_{3,n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ r'_{m,1} & r'_{m,2} & r'_{m,3} & \cdots & r'_{m,4} \end{pmatrix} = R' \tag{2}$$

A modified model of UAV reflectance and ground reflectance was established, where n is the number of bands, m is the number of sample points, and r is the original hyperspectral reflectance. The processed hyperspectral image is shown in Figure 5.

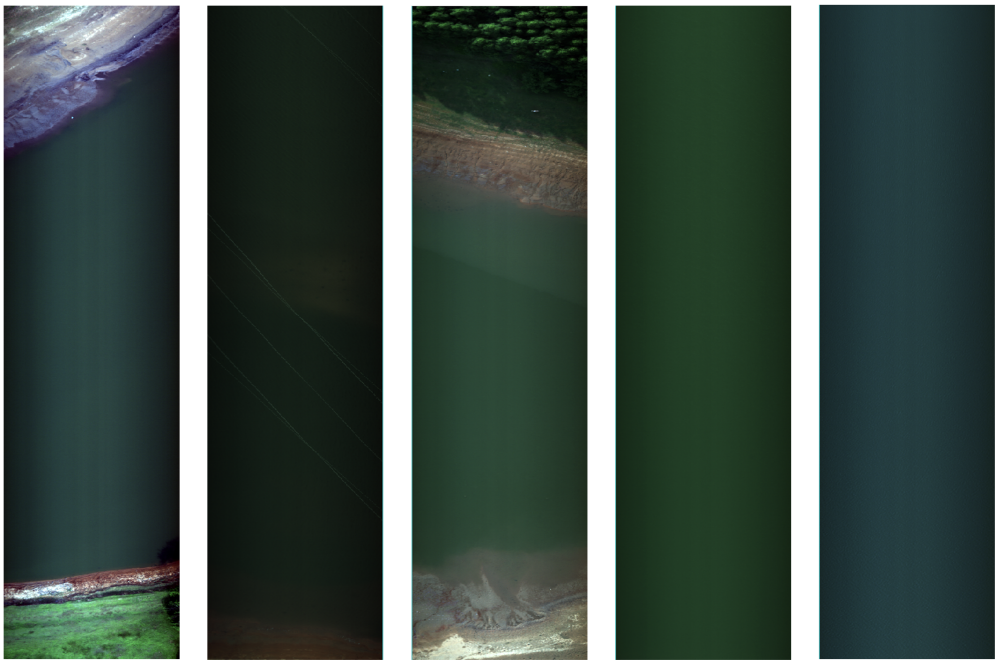
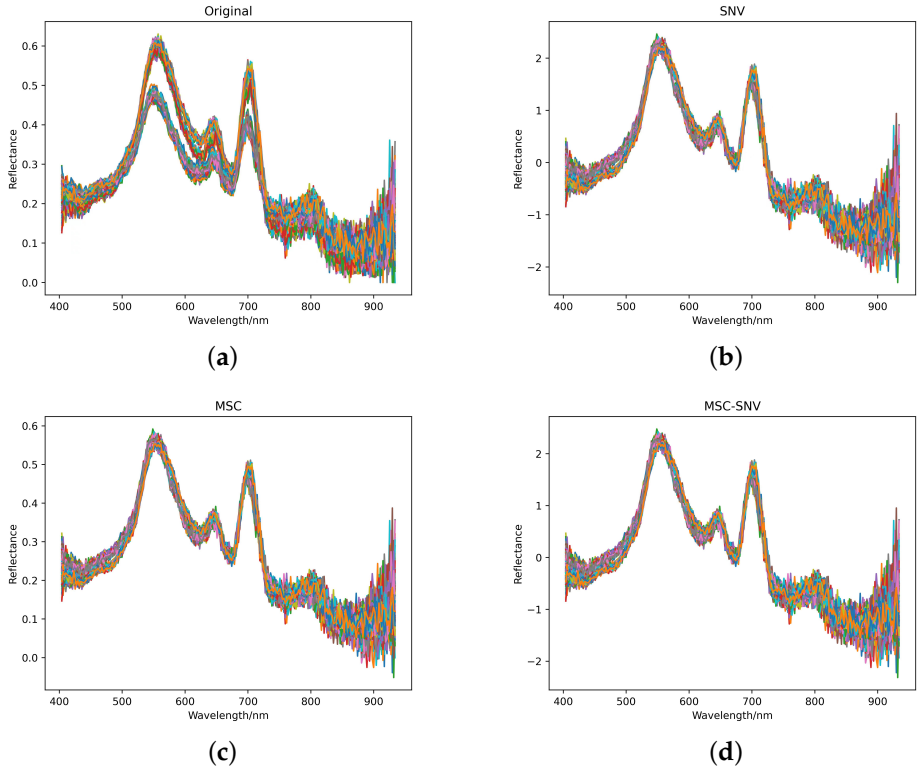


Figure 5. Hyperspectral image of the studied water area.

In this study, 50 points in each of the above five regions were selected for water sampling. Of the data points, 70% were randomly selected as the training set and the remaining 30% were the prediction set. Before analyzing the ROI , it was necessary to preprocess the data to eliminate the system noise and the influence of the surrounding environment. In this study, SNV, MSC, WAVE, MMS and cascade were used for preprocessing. SNV and MSC can reduce the influence of the scattering on the spectrum owing to the uneven particle distribution and different particle size. The effects of some of the different pretreatment methods are shown in Figure 6.



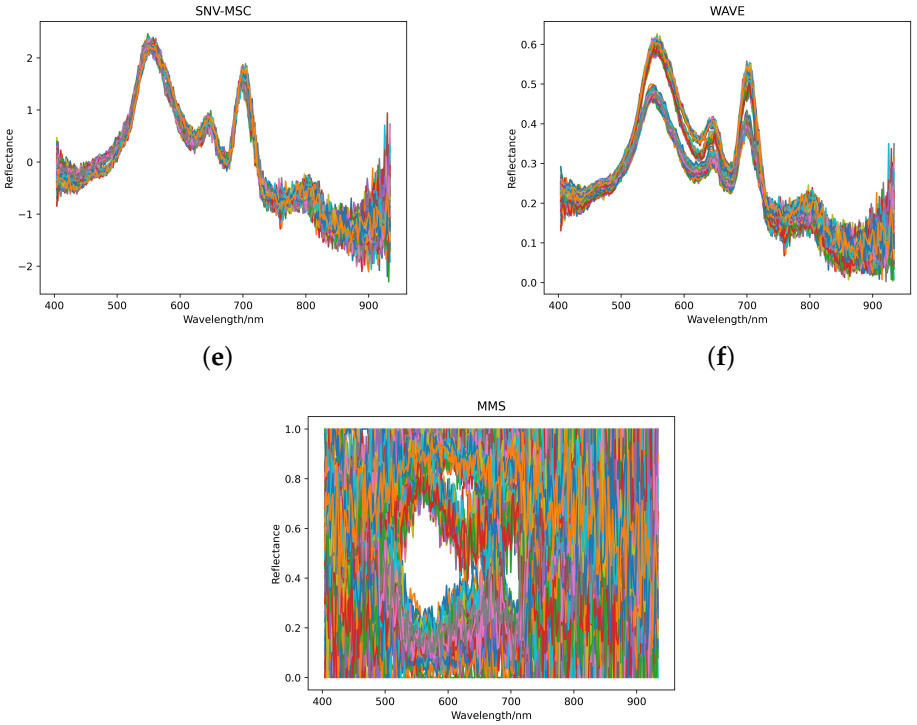


Figure 6. Spectral curve of water data: (**a**) Spectral curve of original data, (**b**) spectral curve of SNV pretreatment. (**c**) Spectral curve after MSC pretreatment, (**d**) spectral curve after SNV+MSC pretreatment, (**e**) spectral curve after MSC+SNV pretreatment, (**f**) spectral curve after WAVE pretreatment, (**g**) spectral curve after MMS pretreatment.

3. Analytical Model and Evaluation Criteria

3.1. Modeling Methods for the Full Spectrum Regression Model

In this study, LR, SVR, PLSR and RFR regression algorithms were used to model and analyze the COD, DO TN, TP, and NH3-N content and spectral reflectance data in a reservoir water body. LR is used where the relationship between variables is simple. SVR is carried out by finding an optimal curve or hyperplane in the data set, that is, by minimizing the training error and maximizing the effective edge [20,21]. SVR regression uses a kernel function to deal with nonlinear problems, which makes it easier to separate input samples. It has strong nonlinear fitting ability and robustness, and only calculates the support vector in the prediction. This greatly reduces the computation process. For the training data set, $D = (x, y)$, where x_i is the n-dimensional and y_i the scalar predicted value, the relaxation variable sum is introduced to minimize the objective function as follows: ξ_i ξ_i^*

$$F = \min \frac{1}{2}\|\omega\|^2 + C\sum_{i=1}^{n}(\xi_i + \xi_i^*) \tag{3}$$

where: C is the penalty coefficient; w and b are the coefficient and intercept of the model respectively. Kernel functions commonly introduced include Gaussian kernel, polynomial kernel and sigmoid kernel function. Here, the Gaussian kernel function was selected for modeling in the study, as shown in Equation (4):

$$K(x_i, x_j) = \exp(-\|x_i - x_j\|^2)/(2\sigma^2) \tag{4}$$

where: σ is a hyperparameter. PLSR is a regression modeling method for multi-dependent variables to multi-independent variables. The final model contained all the original independent variables, and had good model interpretability [22,23]. RFR is composed of multiple regression trees, and there is no correlation between each decision tree in the forest.

The final output of the model is determined by all decision trees in the forest, which has strong anti-interference ability and anti-overfitting ability [24–30].

In this study, the root-mean-square error(RMSE), mean absolute error (MAE) and determination coefficient (R^2) were used to evaluate the effect of the regression analysis model. In general, the smaller the RMSE and MAE error and the closer the R^2 is to 1, the better the prediction performance of the model. The specific parameters are calculated as follows:

$$RMSE = \left(\frac{1}{m} \sum_{i=1}^{m} (y_i - \hat{y}_l)^2 \right)^{1/2} \quad (5)$$

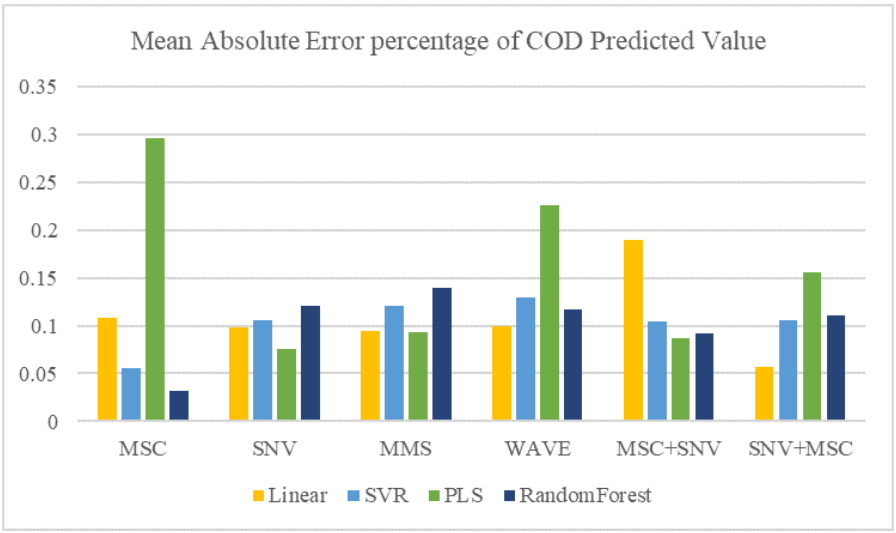
$$MAE = \frac{1}{m} \sum_{i=1}^{m} |y_i - \hat{y}_i| \quad (6)$$

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

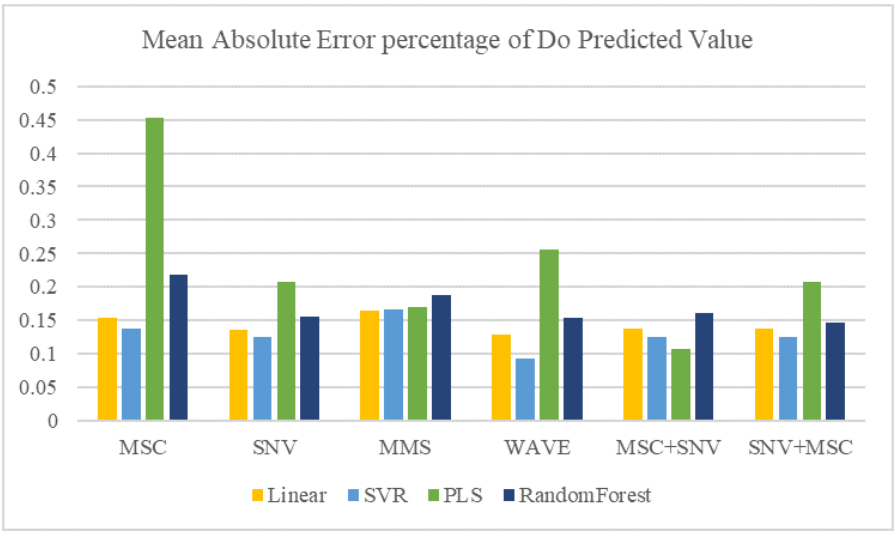
where, RMSE represents the root mean square error, MAE represents the average absolute value error, R^2 represents the determination coefficient, m represents the number of data points, y_i represents the true value of the key indicator content in the water body data point, $\hat{y}_i$ represents the predicted value of the key data indicators COD, DO, TN, TP, and NH3-N, content, $\bar{y}_i$ represents the average value of the true content of key indicators COD, DO, TN, TP, and NH3-N in each data point set.

The expected values of the COD, DO, TN, TP, and NH3-N contents of different key indicators obtained by full-spectrum modeling are shown in Tables 2–6, and the average absolute error percentage of predicted values of water quality parameters is shown in Figure 7.

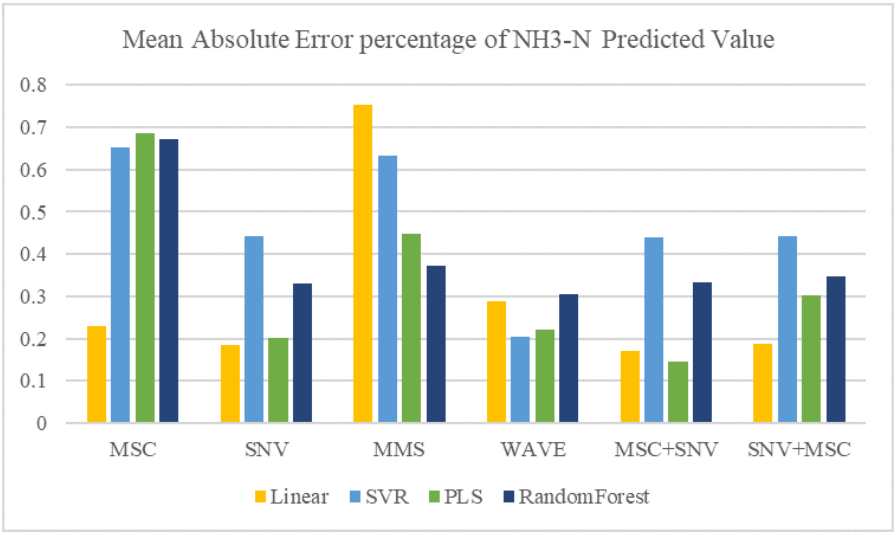
A smaller MAPE indicates a better prediction by the model. On the basis of the above tables and graphs and combining the RMSE and MAE data, we found that the best prediction model of COD was established by the combination of MSC preprocessing and the RandomForest regression algorithm, while the best prediction model of DO was established by the combination of WAVE preprocessing and the SVR regression algorithm. The best prediction model of NH3-N was established by the combination of MSC+SNV pretreatment and PLS regression algorithm, while the best prediction model of TP was established by the combination of MSC+SNV pretreatment and the RandomForest regression algorithm. The best prediction model of TN was established by the combination of MMS pre-processing and the PLS regression algorithm. All these factors provided data input for the following research.



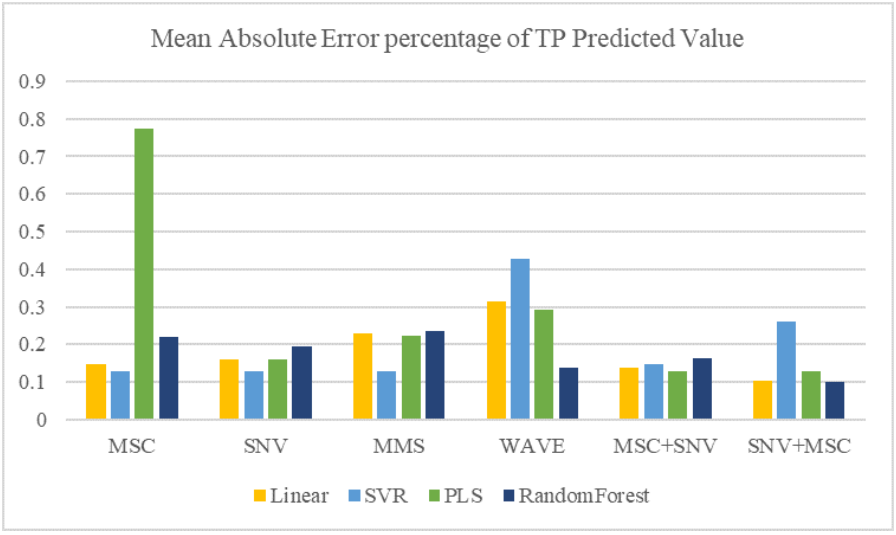
(**a**)


Mean Absolute Error percentage of Do Predicted Value
0.5
0.45
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
MSC
SNV
MMS
WAVE
MSC+SNV
SNV+MSC
Linear
SVR
PLS
RandomForest

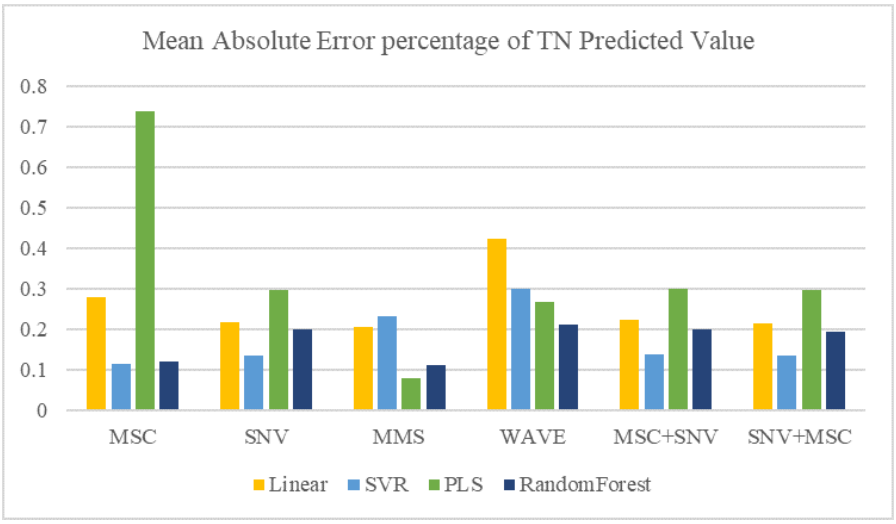
(**b**)


Mean Absolute Error percentage of NH3-N Predicted Value
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
MSC
SNV
MMS
WAVE
MSC+SNV
SNV+MSC
Linear
SVR
PLS
RandomForest

(**c**)


Mean Absolute Error percentage of TP Predicted Value
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
MSC
SNV
MMS
WAVE
MSC+SNV
SNV+MSC
Linear
SVR
PLS
RandomForest

(**d**)



(**e**)

Figure 7. Histogram of prediction and evaluation of water quality parameters: (**a**) MAPE cluster histogram of predicted COD, (**b**) MAPE cluster histogram of predicted DO, (**c**) MAPE cluster histogram of predicted NH3-N, (**d**) MAPE cluster histogram of predicted TP, (**e**) MAPE cluster histogram of predicted TN.

Table 2. Prediction results of COD with full spectrum.

Model Preprocessing	Regression Model	RMSE	MAE	R^2	Predicted
MSC	Linear	0.0877	0.7071	0.8227	0.8921
	SVR	0.0980	0.0829	0.7773	0.9448
	PLS	0.0742	0.0565	0.8726	0.7037
	RandomForest	0.0678	0.0202	0.8930	0.9686
SNV	Linear	0.0849	0.0647	0.8381	0.9022
	SVR	0.0980	0.0825	0.7788	0.8944
	PLS	0.0735	0.0561	0.8769	0.9244
	RandomForest	0.0894	0.0285	0.8154	0.8795
MMS	Linear	0.0574	0.0496	0.9232	0.9051
	SVR	0.0975	0.0971	0.7816	0.8796
	PLS	0.0728	0.0489	0.8781	0.9063
	RandomForest	0.0077	0.0038	0.9985	0.8611
WAVE	Linear	0.0656	0.0514	0.9005	0.9007
	SVR	0.0742	0.0643	0.8730	0.8711
	PLS	0.0721	0.0484	0.8806	0.7740
	RandomForest	0.0100	0.0039	0.9976	0.8825
MSC+SNV	Linear	0.0860	0.0689	0.8292	0.8100
	SVR	0.0980	0.0828	0.7771	0.8953
	PLS	0.0742	0.0566	0.8727	0.9132
	RandomForest	0.0686	0.0199	0.8905	0.9085
SNV+MSC	Linear	0.0825	0.0658	0.8441	0.9430
	SVR	0.0980	0.0825	0.7788	0.8944
	PLS	0.0735	0.056	0.8769	0.8444
	RandomForest	0.0922	0.0288	0.8045	0.9895

Table 3. DO full spectrum prediction results.

Model Preprocessing	Regression Model	RMSE	MAE	R^2	Predicted
MSC	Linear	0.2789	0.2308	0.8951	0.8455
	SVR	0.4589	0.3604	0.7159	0.8630
	PLS	0.2059	0.1611	0.9428	0.5471
	RandomForest	0.2462	0.1311	0.9182	0.7813
SNV	Linear	0.2693	0.2221	0.9022	0.8644
	SVR	0.4593	0.3612	0.7051	0.8744
	PLS	0.2020	0.1571	0.9441	0.7923
	RandomForest	0.1931	0.1027	0.9496	0.8443
MMS	Linear	0.1288	0.1071	0.9775	0.8359
	SVR	0.1136	0.1021	0.9825	0.8332
	PLS	0.1483	0.1072	0.9703	0.8307
	RandomForest	0.1778	0.0825	0.9571	0.8128
WAVE	Linear	0.1411	0.1152	0.9731	0.8703
	SVR	0.1175	0.1020	0.9813	0.9064
	PLS	0.1428	0.1101	0.9724	0.7448
	RandomForest	0.1682	0.0748	0.9617	0.8467
MSC+SNV	Linear	0.2766	0.2281	0.8968	0.8622
	SVR	0.4587	0.3604	0.7160	0.8741
	PLS	0.2059	0.1612	0.9428	0.8934
	RandomForest	0.2966	0.1251	0.8812	0.8386
SNV+MSC	Linear	0.2661	0.2199	0.9044	0.8630
	SVR	0.4593	0.3612	0.7152	0.8744
	PLS	0.2020	0.1571	0.9449	0.7923
	RandomForest	0.2078	0.1011	0.9421	0.8538

Table 4. NH3-N full spectrum prediction results.

Model Preprocessing	Regression Model	RMSE	MAE	R^2	Predicted
MSC	Linear	0.0648	0.0525	0.78	0.7718
	SVR	0.0800	0.0687	0.663	0.3485
	PLS	0.0539	0.0413	0.8459	0.3131
	RandomForest	0.0548	0.0148	0.8529	0.3269
SNV	Linear	0.0632	0.0511	0.7918	0.8148
	SVR	0.0775	0.0685	0.663	0.5574
	PLS	0.0548	0.0409	0.85125	0.7984
	RandomForest	0.0632	0.0197	0.7837	0.6705
MMS	Linear	0.0447	0.0356	0.9079	0.2477
	SVR	0.0949	0.097	0.5081	0.3672
	PLS	0.0548	0.035	0.859	0.5531
	RandomForest	0.0063	0.0035	0.9976	0.6270
WAVE	Linear	0.0447	0.0369	0.8842	0.7128
	SVR	0.0707	0.06169	0.7331	0.7967
	PLS	0.0548	0.035	0.8604	0.7792
	RandomForest	0.0100	0.002	0.9971	0.6951
MSC+SNV	Linear	0.0632	0.052	0.786	0.8300
	SVR	0.0775	0.0686	0.6635	0.5616
	PLS	0.0548	0.04	0.8461	0.8542
	RandomForest	0.0529	0.0135	0.8566	0.6679
SNV+MSC	Linear	0.0632	0.0503	0.7969	0.8128
	SVR	0.0775	0.0685	0.6634	0.5574
	PLS	0.0539	0.04097	0.8511	0.6984
	RandomForest	0.0632	0.021	0.7699	0.6541

Table 5. TP full spectrum prediction results.

Model Preprocessing	Regression Model	RMSE	MAE	R^2	Predicted
MSC	Linear	0.0055	0.0042	0.824	0.8513
	SVR	0.0141	0.0151	0.436	0.8718
	PLS	0.0045	0.0033	0.8769	0.2256
	RandomForest	0.0548	0.0012	0.9008	0.7795
SNV	Linear	0.0055	0.0041	0.8343	0.8385
	SVR	0.0045	0.0152	0.4361	0.8718
	PLS	0.0055	0.0034	0.8768	0.8410
	RandomForest	0.0632	0.0018	0.8051	0.8049
MMS	Linear	0.0032	0.0029	0.9216	0.7713
	SVR	0.0141	0.0152	0.4361	0.8718
	PLS	0.0045	0.0029	0.8782	0.7754
	RandomForest	0.0055	0.0003	0.9981	0.7641
WAVE	Linear	0.0045	0.0321	0.9012	0.6872
	SVR	0.0141	0.0151	0.4361	0.5718
	PLS	0.0042	0.0029	0.8828	0.7067
	RandomForest	0.0041	0.0002	0.9938	0.8631
MSC+SNV	Linear	0.0051	0.0041	0.8291	0.8533
	SVR	0.0141	0.0151	0.4362	0.8718
	PLS	0.0044	0.0039	0.8727	0.8369
	RandomForest	0.0041	0.0012	0.8871	0.8949
SNV+MSC	Linear	0.0050	0.0041	0.8392	0.7405
	SVR	0.0141	0.0152	0.4362	0.8718
	PLS	0.0044	0.0034	0.8769	0.8410
	RandomForest	0.0057	0.0017	0.7964	0.7749

Table 6. TN full spectrum prediction results.

Model Preprocessing	Regression Model	RMSE	MAE	R^2	Predicted
MSC	Linear	0.2755	0.0698	0.8721	0.7200
	SVR	0.0990	0.0838	0.8323	0.8838
	PLS	0.0755	0.0571	0.9021	0.2605
	RandomForest	0.0671	0.0224	0.9228	0.8771
SNV	Linear	0.0849	0.0682	0.8766	0.7819
	SVR	0.0990	0.0834	0.8332	0.8648
	PLS	0.0742	0.0561	0.9055	0.7024
	RandomForest	0.0949	0.0285	0.8241	0.7990
MMS	Linear	0.0600	0.0518	0.9371	0.7924
	SVR	0.0949	0.0951	0.8432	0.7671
	PLS	0.0742	0.0496	0.9057	0.9210
	RandomForest	0.0084	0.0053	0.9981	0.8871
WAVE	Linear	0.0686	0.0541	0.9192	0.5748
	SVR	0.0735	0.0638	0.9071	0.7005
	PLS	0.0707	0.0564	0.9123	0.7305
	RandomForest	0.0063	0.0033	0.9938	0.7862
MSC+SNV	Linear	0.0837	0.0681	0.8731	0.7747
	SVR	0.0990	0.0837	0.8324	0.8620
	PLS	0.0755	0.0574	0.9024	0.6989
	RandomForest	0.0735	0.0185	0.9074	0.7995
SNV+MSC	Linear	0.0837	0.0671	0.8801	0.7833
	SVR	0.0985	0.0832	0.8332	0.8638
	PLS	0.0742	0.0569	0.9055	0.7024
	RandomForest	0.0872	0.0284	0.8709	0.8050

3.2. Spectrum Dimensionality Reduction

The spectral dimension reduction methods used in this study included SPA, PCA, and SAA. SPA is one of the commonly used wavelength selection methods, which can select the shortest collinear wavelength combination. For COD parameter prediction, in this study, the MSC-RFR analysis model was established with each wavelength group, and the wavelength combination with the largest determination coefficient was taken to be the best combination. In this study, a total of 20 characteristic wavelengths were selected by SPA. The wavelength combination was selected by PCA by extracting the principal component which contributed the most to the original information and combining the corresponding principal component load coefficient. In this study, the PCA algorithm was used to select 15 characteristic wavelengths. SAA is a random combinatorial optimization algorithm, which considers not only the optimal solution, but also the deteriorating solution within a certain limit, finds the local optimal solution, and then obtains the global optimal solution. In this study, SAA was used to select 20 characteristic wavelengths, and the optimal wavelength combination selected by SPA, PCA, and SAA was used to simplify the MSC-RFR model. The MSC-SPA-RFR model, MSC-PCA-RFR model, and MSC-SAA-RFR model were established. The other four water quality parameters were calculated in the same way, and their specific performance is shown in Table 7.

Table 7. Spectral dimensionality reduction analysis results.

	Spectral Dimensionality Reduction Method	Predicted	MAPE
COD	MSC-SAA-RFR	0.9871	0.0129
	MSC-SPA-RFR	0.9412	0.0588
	MSC-SPA-RFR	0.9132	0.0868
DO	WAVE-SAA-SVR	0.9291	0.0709
	WAVE-SPA-SVR	0.7647	0.2353
	WAVE-PCA-SVR	0.9096	0.0904
NH3-N	MSC+SNV-SAA-PLS	0.8746	0.1254
	MSC+SNV-SPA-PLS	0.5897	0.4103
	MSC+SNV-PCA-PLS	0.5951	0.4049
TP	MSC+SNV-SAA-RFR	0.7230	0.277
	MSC+SNV-SPA-RFR	0.8941	0.1059
	MSC+SNV-PCA-RFR	0.8000	0.2
TN	MMS-SAA-PLS	0.8871	0.1129
	MMS-SPA-PLS	0.8752	0.1248
	MMS-PCA-PLS	0.9210	0.079

4. Results and Discussion

Part 3 shows that we were able to establish corresponding preprocessing and full-spectrum regression modeling methods for the indicators COD, TN, TP, NH3-N, and DO, as can be seen from Tables 2–6 above. As can be seen from Table 7, for COD, the MSC method was used to pretreat the hyperspectral reflectance data of reservoir water, and the MSC-RFR analysis model was established. RMSE, MAE, R^2 and MAPE were used to evaluate the full spectrum analysis model. Then, three different dimensionality reduction methods were used to reduce the spectral dimension of the reservoir water hyperspectral reflectance data to establish the MSC-SAA-RFR, MSC-SPA-RFR, and MSC-SPA-RFR analysis models. Comparison of the experimental results showed that the MSC-SAA-RFR model had a higher R^2 value and a lower MAPE value and could better predict the COD content in water. For DO, the WAVE-SVR full spectrum analysis model was first established, and then three different dimensionality reduction methods were used to reduce the spectral dimension of reservoir water hyperspectral reflectance data to establish the WAVE-SAA-SVR, WAVE-SPA-SVR, and WAVE-PCA-SVR analysis models. The comparative experimental results showed that the WAVE-SAA-SVR model had a higher R^2 value and lower MAPE

value, which could better predict the DO content in water. For NH3-N, the full spectrum analysis model of MSC-PLS was first established, and then three different dimensionality reduction methods were used to reduce the spectral dimension of the reservoir water hyperspectral reflectance data to establish the MSC+SNV-SAA-PLS, MSC+SNV-SPA-PLS, MSC+ SNV-PACC-PLS analysis models. The comparison of the three experimental results showed that the MSC+SNV-SAA- PLS model had a higher R^2 value and lower MAPE value, and could better predict the NH3-N content in water. For TP, the full spectrum analysis model of MSC-RFR was first established, and then spectral dimension reduction was performed on the hyperspectral reflectance data of reservoir water by three different dimensionality reduction methods to establish the MSC+SNV-SAA-RFR, MSC+SNV-SPA-RFR, and MSC+SNV-PCA-RFR analysis models. Comparing the three experimental results showed that the MSC+SNV-SPA-RFR model had a higher R^2 value and lower MAPE value, which could better predict the TP content in water. For TN, the full spectrum analysis model of MMS-PLS was first established, and then three different dimensionality reduction methods were used to reduce the spectral dimension of the reservoir water hyperspectral reflectance data to establish the MMS-SAA-PLS, MMS-SPA-PLS, and MMS-PCA-PLS analysis models. The comparative experimental results of the three showed that the MMS-PCA-PLS model had a higher R^2 value and lower MAPE value, which could better predict the TP content in the water body. In addition, we believe that the quality of the evaluation model was not determined by a single level of R^2 and RMSE. Only through the comprehensive evaluation of R^2, RMSE, and MAPE could a more robust water quality parameter analysis model be obtained.

5. Conclusions

Because the detection indicators of urban river sewage cannot be monitored over a large area quickly and dynamically, this study used a UAV to collect hyperspectral image data, which was effective and flexible, with a higher spatial resolution and less medium interference than satellite collection. Combining machine learning and statistical analysis, we examined the parameter inversion method for dynamic monitoring of urban inland water pollution on the basis of UAV hyperspectral imaging technology. At the same time, we established a set of standardized processes from UAV hyperspectral sampling, ground spectrum correction, spectral data preprocessing, and modeling. We combined the above-mentioned models into a unified end-to-end structure. In general, from the perspective of theoretical significance, this study perfected a set of systematic data collection, physical and chemical analysis of water samples, spectral preprocessing, and data modeling of the key indicators of water quality parameters, providing a guideline for subsequent research. From the perspective of application significance, this method can monitor the change in urban river water quality in real time which is conducive to the tracking of pollution sources, and provide a decision-making basis for urban river water environment management by establishing an early warning system. This will minimize the harm caused by sewage to people's lives.

However, this research has some limitations. On the one hand, the data for our study were collected between 10:00 a.m. and 2:00 p.m. on sunny days, which did not include monitoring during some extreme bad weather. In the follow-up work, we plan to add weather and other parameter factors into the model to correct this, further improve the robustness of the analysis and prepare for all-weather monitoring. On the other hand, we will conduct experiments in China and other parts of the world to further verify the effectiveness of the quantitative method of water quality parameter concentration in this study, and accelerate the construction of and intelligent monitoring system of water environment.

Author Contributions: Conceptualization, J.C. and J.W.; Methodology, C.S.; Software, M.W.; Validation, J.C., J.W. and Z.Z.; Formal analysis, Z.Z.; Investigation, J.W.; Resources, J.Y.; Data curation, C.S.; Writing—original draft preparation, J.C.; Writing—review and editing, J.C.; Visualization, S.F.; Supervision, J.Y.; Project administration, N.S.; Funding acquisition, J.C. All authors have read and agreed to the published version of the manuscript.

Funding: Jilin Province and the Chinese Academy of Sciences Science and Technology Cooperation in High-tech Industrialisation Special Funds Project (2022SYHZ0025); Jilin Province Science & Technology Development Program Project in China (20200403009SF).

Data Availability Statement: Data underlying the results presented in this paper are not publicly available at this time but may be obtained from the authors upon reasonable request.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

COD	Chemical Oxygen Demand
DO	Dissolved Oxygen
TP	Total Phosphorus
TN	Total Nitrogen
NH3-N	Ammonia Nitrogen
UAV	Unmanned Aerial Vehicle
RMSE	Root Mean Square Error
MAPE	Mean Absolute Percentage Error
MAE	Mean Absolute Error
R^2	Coefficient of Determination
CNN	Convolutional Neural Network
DCGAN	Deep Convolutional Generation and Adversarial Network
BNN	Bayesian Neural Network
SSC	Suspended Sediment Concentration
GRNN	Gated Recurrent Neural Network
Chl-a	ChlorophylL-A
PCA	Principal Component Analysis
SPA	Successive Projections Algorithm
SAA	Simulated Annealing Algorithm
CIOMP	Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences
ROI	Region of Interest
SNV	Standard Normal Variate Correction
MSC	Multiplicative Scatter Correction
MMS	Min-Max Standardization
WAVE	Wavelet Transform
LR	LinearRegression
SVR	Support Vector Regression
PLS	Partial Least Squares Regression
RFR	Random Forest Regression

References

1. Palmer, S.C.; Kutser, T.; Hunter, P.D. Remote sensing of inland waters: Challenges, progress and future directions. *Remote Sens. Environ.* **2015**, *157*, 1–8.
2. Harish, B.; Manjulavani, K.; Shantosh, M.; Supriya, V.M. Change detection of land use and land cover using remote sensing techniques. In Proceedings of the 2017 IEEE International Conference on Power, Control, Signals and Instrumentation Engineering (ICPCSI), Chennai, India, 21–22 September 2017; pp. 2806–2810.
3. HaRa, J.; Atique, U.; An, K.G. Multiyear links between water chemistry, algal chlorophyll, drought-flood regime, and nutrient enrichment in a morphologically complex reservoir. *Int. J. Environ. Res. Public Health* **2020**, *17*, 3139.
4. Lewis, W.M., Jr.; Wurtsbaugh, W.A.; Paerl, H.W. Rationale for control of anthropogenic nitrogen and phosphorus to reduce eutrophication of inland waters. *Environ. Sci. Technol.* **2011**, *45*, 10300–10305.
5. Adamiak, M.; Będkowski, K.; Majchrowska, A. Aerial imagery feature engineering using bidirectional generative adversarial networks: A case study of the pilica river region, Poland. *Remote Sens.* **2021**, *13*, 306.

6. Doña, C.; Chang, N.B.; Caselles, V.; Sánchez, J.M.; Camacho, A.; Delegido, J.; Vannah, B.W. Integrated satellite data fusion and mining for monitoring lake water quality status of the Albufera de Valencia in Spain. *J. Environ. Manag.* **2015**, *151*, 416–426.
7. Ghaderi, D.; Rahbani, M. Tracing suspended matter in Tiab estuary applying ANN and Remote sensing. *Reg. Stud. Mar. Sci.* **2021**, *44*, 101788. [CrossRef]
8. Dona, C.; Sanchez, J.M.; Caselles, V.; Domínguez, J.A.; Camacho, A. Empirical relationships for monitoring water quality of lakes and reservoirs through multispectral images. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2014**, *7*, 1632–1641. [CrossRef]
9. Bonansea, M.; Ledesma, M.; Rodriguez, C.; Pinotti, L. Using new remote sensing satellites for assessing water quality in a reservoir. *Hydrol. Sci. J.* **2019**, *64*, 34–44. [CrossRef]
10. Bao, Y. Diurnal variation of chlorophyll a concentration in Taihu Lake based on GOCI image classification. *Spectrosc. Spectr. Anal.* **2016**, *36*, 2562–2567.
11. Gu, Q.; Li, Q.; Zhou, M. Water Quality Monitoring of the Yangtze Estuary by Using GF-5 Hyperspectral Image. In Proceedings of the 2019 12th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI), Suzhou, China, 19–21 October 2019; pp. 1–5. [CrossRef]
12. Jichang, T.; Xueqin, Y.; Chaobo, C.; Song, G.; Jingcheng, W.; Cheng, S. Water quality prediction model based on GRU hybrid network. In Proceedings of the 2019 Chinese Automation Congress (CAC), Hangzhou, China, 22–24 November 2019; pp. 1893–1898. [CrossRef]
13. Du, C.; Wang, Q.; Li, Y.; Lyu, H.; Zhu, L.; Zheng, Z.; Wen, S.; Liu, G.; Guo, Y. Estimation of total phosphorus concentration using a water classification method in inland water. *Int. J. Appl. Earth Obs. Geoinf.* **2018**, *71*, 29–42. [CrossRef]
14. Ahn, J.M.; Kim, B.; Jong, J.; Nam, G.; Park, L.J.; Park, S.; Kang, T.; Lee, J.K.; Kim, J. Predicting Cyanobacterial Blooms Using Hyperspectral Images in a Regulated River. *Sensors* **2021**, *21*, 530.
15. Flynn, K.F.; Chapra, S.C. Remote Sensing of Submerged Aquatic Vegetation in a Shallow Non-Turbid River Using an Unmanned Aerial Vehicle. *Remote Sens.* **2014**, *6*, 12815–12836.
16. Pajares, G. Overview and Current Status of Remote Sensing Applications Based on Unmanned Aerial Vehicles (UAVs). *Photogramm. Eng. Remote Sens.* **2015**, *81*, 281–329. [CrossRef]
17. Su, T.C.; Chou, H.T. Application of Multispectral Sensors Carried on Unmanned Aerial Vehicle (UAV) to Trophic State Mapping of Small Reservoirs: A Case Study of Tain-Pu Reservoir in Kinmen, Taiwan. *Remote Sens.* **2015**, *7*, 10078–10097.
18. Zaman, B.; Jensen, A.M.; Clemens, S.; McKee, M. Retrieval of Spectral Reflectance of High Resolution Multispectral Imagery Acquired with an Autonomous Unmanned Aerial Vehicle: AggieAir™. *Photogramm. Eng. Remote Sens.* **2014**, *80*, 1139–1150.
19. Su, T.C. 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. *Int. J. Appl. Earth Obs. Geoinf.* **2017**, *58*, 213–224. [CrossRef]
20. Pour, S.H.; Shahid, S.; Chung, E.S.; Wang, X.J. Model output statistics downscaling using support vector machine for the projection of spatial and temporal changes in rainfall of Bangladesh. *Atmos. Res.* **2018**, *213*, 149–162. [CrossRef]
21. Zhang, X.; Liu, G.; Wang, H.; Li, X. Application of a Hybrid Interpolation Method Based on Support Vector Machine in the Precipitation Spatial Interpolation of Basins. *Water* **2017**, *9*, 760. [CrossRef]
22. Leone, A.P.; Viscarra-Rossel, R.A.; Amenta, P.; Buondonno, A. Prediction of soil properties with PLSR and vis-NIR spectroscopy: Application to mediterranean soils from Southern Italy. *Curr. Anal. Chem.* **2012**, *8*, 283–299. [CrossRef]
23. Hou, L.; Li, X.; Li, F. Hyperspectral-based Inversion of Heavy Metal Content in the Soil of Coal Mining Areas. *J. Environ. Qual.* **2019**, *1*, 57–63. [CrossRef]
24. Shi, C.; Liao, D.; Zhang, T.; Wang, L. Hyperspectral Image Classification Based on Expansion Convolution Network. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 1–16. [CrossRef]
25. Xi, J.; Ersoy, O.K.; Cong, M.; Zhao, C.; Qu, W.; Wu, T. Wide and Deep Fourier Neural Network for Hyperspectral Remote Sensing Image Classification. *Remote Sens.* **2022**, *14*, 2931. [CrossRef]
26. Vaiphasa, C.; Skidmore, A.K.; de Boer, W.F.; Vaiphasa, T. A hyperspectral band selector for plant species discrimination. *ISPRS J. Photogramm. Remote Sens.* **2007**, *62*, 225–235. [CrossRef]
27. Xu, Y.; Du, Q.; Younan, N. Particle swarm optimization-based band selection for hyperspectral target detection. In Proceedings of the 2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Beijing, China, 10–15 July 2016; pp. 5872–5875. [CrossRef]
28. Zhao, W.; Wang, L.; Zhang, Z. A novel atom search optimization for dispersion coefficient estimation in groundwater. *Future Gener. Comput. Syst.* **2019**, *91*, 601–610. [CrossRef]
29. Zhong, K.; Luo, Q.; Zhou, Y.; Jiang, M. TLMPA: Teaching-learning-based Marine Predators algorithm. *AIMS Math.* **2021**, *6*, 1395–1442. [CrossRef]
30. Hennessy, A.; Clarke, K.; Lewis, M. Hyperspectral Classification of Plants: A Review of Waveband Selection Generalisability. *Remote Sens.* **2020**, *12*, 113. [CrossRef]