

## Article

# Research on Inversion Mechanism of Chlorophyll—A Concentration in Water Bodies Using a Convolutional Neural Network Model

Yun Xue <sup>1,2</sup>, Lei Zhu <sup>1,3</sup>, Bin Zou <sup>2,\*</sup>, Yi-min Wen <sup>4</sup>, Yue-hong Long <sup>1</sup> and Song-lin Zhou <sup>1</sup>

<sup>1</sup> School of Municipal and Surveying Engineering, Design Institute Co., Ltd., Hunan City University, Yiyang 413000, China; yunxue1209@hncu.edu.cn (Y.X.); weizhang@hncu.edu.cn (L.Z.); yhlong@hncu.edu.cn (Y.-h.L.); liufl@hncu.edu.cn (S.-l.Z.)

<sup>2</sup> Key Laboratory of Metallogenic Prediction of Nonferrous Metals and Geological Environment Monitoring, School of Geosciences and Info-Physics, Central South University, Changsha 410083, China

<sup>3</sup> School of Civil and Surveying Engineering, Jiangxi University of Science and Technology, Ganzhou 341000, China

<sup>4</sup> School of Computer Science and Information Security, Guilin University of Electronic Technology, Guilin 541004, China; ymw@guet.edu.cn

\* Correspondence: 210010@csu.edu.cn

**Abstract:** For Case-II water bodies with relatively complex water qualities, it is challenging to establish a chlorophyll-a concentration (Chl-a concentration) inversion model with strong applicability and high accuracy. Convolutional Neural Network (CNN) shows excellent performance in image target recognition and natural language processing. However, there little research exists on the inversion of Chl-a concentration in water using convolutional neural networks. Taking China's Dongting Lake as an example, 90 water samples and their spectra were collected in this study. Using eight combinations as independent variables and Chl-a concentration as the dependent variable, a CNN model was constructed to invert Chl-a concentration. The results showed that: (1) The CNN model of the original spectrum has a worse inversion effect than the CNN model of the preprocessed spectrum. The determination coefficient ( $R_p^2$ ) of the predicted sample is increased from 0.79 to 0.88, and the root mean square error ( $RMSEP$ ) of the predicted sample is reduced from 0.61 to 0.49, indicating that preprocessing can significantly improve the inversion effect of the model.; (2) among the combined models, the CNN model with Baseline1\_SC (strong correlation factor of 500–750 nm baseline) has the best effect, with  $R_p^2$  reaching 0.90 and  $RMSEP$  only 0.45. The average inversion effect of the eight CNN models is better. The average  $R_p^2$  reaches 0.86 and the  $RMSEP$  is only 0.52, indicating the feasibility of applying CNN to Chl-a concentration inversion modeling; (3) the performance of the CNN model (Baseline1\_SC ( $R_p^2 = 0.90$ ,  $RMSEP = 0.45$ )) was far better than the traditional model of the same combination, i.e., the linear regression model ( $R_p^2 = 0.61$ ,  $RMSEP = 0.72$ ) and partial least squares regression model (Baseline1\_SC ( $R_p^2 = 0.58$ ,  $RMSEP = 0.95$ )), indicating the superiority of the convolutional neural network inversion modeling of water body Chl-a concentration.

**Keywords:** convolutional neural network; chlorophyll-a; Dongting Lake



**Citation:** Xue, Y.; Zhu, L.; Zou, B.; Wen, Y.-m.; Long, Y.-h.; Zhou, S.-l. Research on Inversion Mechanism of Chlorophyll—A Concentration in Water Bodies Using a Convolutional Neural Network Model. *Water* **2021**, *13*, 664. <https://doi.org/10.3390/w13050664>

Academic Editor:  
Alessandro Bergamasco

Received: 31 December 2020  
Accepted: 27 February 2021  
Published: 28 February 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

The problem of lake eutrophication is one of the most important water environmental problems facing human beings [1]. Although the current degree of eutrophication in Dongting Lake's water body is relatively low, due to factors such as the social and economic development of the basin, the water quality deteriorated and the eutrophication trend intensified [2,3]. Chlorophyll-a (Chl-a) is an important component of aquatic algae, with its concentration often used to estimate the biomass and productivity of phytoplankton, and it is also an important parameter reflecting the degree of eutrophication of water bodies. Carrying out Chl-a monitoring is of great significance in understanding the status and evolution trend of Dongting Lake eutrophication [4,5]. At present, water quality monitoring

methods can accurately determine the various indicators of the water quality at certain locations, but due to the large labor and material resources required, it is difficult to meet the needs of real-time, rapid, large-scale water environment monitoring. Hyperspectral remote sensing technology are rapid, macroscopic, and high spectral resolution, which provides the possibility for rapid monitoring of Chl-a concentration [6–8].

The material composition of Case-II water bodies is relatively complex, and the optical properties are mainly affected by phytoplankton pigments, suspended particles, and yellow substances. There is a complex coupling mechanism between the spectra of different components, which brings great uncertainty to the inversion of Chl-a concentration. A simple empirical model can hardly meet the accuracy requirements of Chl-a concentration inversion. How to accurately simulate the relationship between spectral information and Chl-a concentration and to establish a Chl-a concentration inversion model with strong applicability and high accuracy has always been a difficult point [9,10]. Scholars at home and abroad have done a lot of research on improving the accuracy of Chl-a concentration inversion. Band ratio [11,12], three-band [13–17], and four-band [18–20] algorithms are commonly used band selection algorithms for Chl-a concentration inversion and have many successful applications. Yang et al. [7] used vegetation index and Chl-a concentration to establish a regression model and used Landsat images to study the temporal and spatial distribution characteristics of Chl-a concentration in East Lake. Cheng et al. [18] used the band ratio, three-band, four-band model, and Chl-a concentration to establish a regression model, and inverted the Chl-a concentration in Taihu Lake. However, the multi-band characteristics of hyperspectral are not fully utilized, and the Chl-a information carried by other spectral bands in the visible light range is ignored. Secondly, in high turbidity lakes, the influence of suspended solids concentration on the spectrum limits the application of the above model [6,19,21]. Baseline correction is to subtract the baseline value from the spectral reflectance [22–25], which can be used to weaken the non-Chl-a spectral information in the in-situ spectrum. Wei et al. [26] found that baseline correction can weaken the spectral contribution of suspended sediment to Chl-a. Thirdly, there are a lot of uncertain factors in the relationship between the water reflection spectrum and the water composition. Using traditional empirical and semi-empirical models to invert the nonlinear relationship between Chl-a concentration and the reflectance spectrum, good prediction results cannot be obtained [11,27,28]. Convolutional Neural Networks (CNN) is a type of feature extractor composed of convolutional layers and pooling layers and a network model composed of feedforward neural networks. CNN's weight-sharing network method makes it resemble a biological neural network structure, thereby reducing the complexity of the model and the amount of weight, and better extracting data features. Through the activation function, CNN can realize multilevel, nonlinear transformation. CNN was successfully applied in many fields, such as image target recognition and natural language processing [29–32]. However, for Case-II water bodies with relatively complex water qualities, it is still challenging to establish a Chl-a concentration inversion model with strong applicability and high accuracy, with few studies on the CNN model of Chl-a concentration in water bodies [33–35]. Syariz et al. [36] used Sentinel-3 and in-situ Chl-a concentration to achieve the Chl-a concentration inversion using CNN and found that the inversion accuracy is higher than that of the neural network model. However, for turbid water bodies, based on in-situ spectroscopy, Chl-a's CNN inversion model has not been reported. To sum up, (1) For the filtering of bands, it is easy to cause information loss and affect the accuracy of Chl-a concentration inversion. This study uses preprocessed in-situ spectra as the input of the model. (2) Aiming at the problem of the spectral contribution of suspended solids to Chl-a in high turbidity lakes, the baseline correction method is used to eliminate its influence. (3) Whether the CNN model is feasible to retrieve the Chl-a concentration, what is the best mode for the CNN model to retrieve the Chl-a concentration, and whether it is superior to the traditional Chl-a concentration inversion model is still an open question.

Taking China's Dongting Lake as an example, this study collected 90 water samples and their spectra. Taking different inversion factors (strong correlation factors and principal component factors) of different spectrum types (original spectrum, 500–750 baseline spectrum, 750 baseline spectrum, and envelope removal spectrum) as input independent variables and Chl-a concentration as the dependent variable, eight Chl-a concentration inversion models based on one-dimensional CNNs were constructed and their performances were compared. The aim was to study a new method of Chl-a concentration inversion and to explore the best mode of the CNN model applied to Chl-a concentration inversion. It is expected to provide a scientific selection basis for the hyperspectral inversion of Chl-a concentration and rapid diagnosis of eutrophication in Dongting Lake.

## 2. Materials and Methods

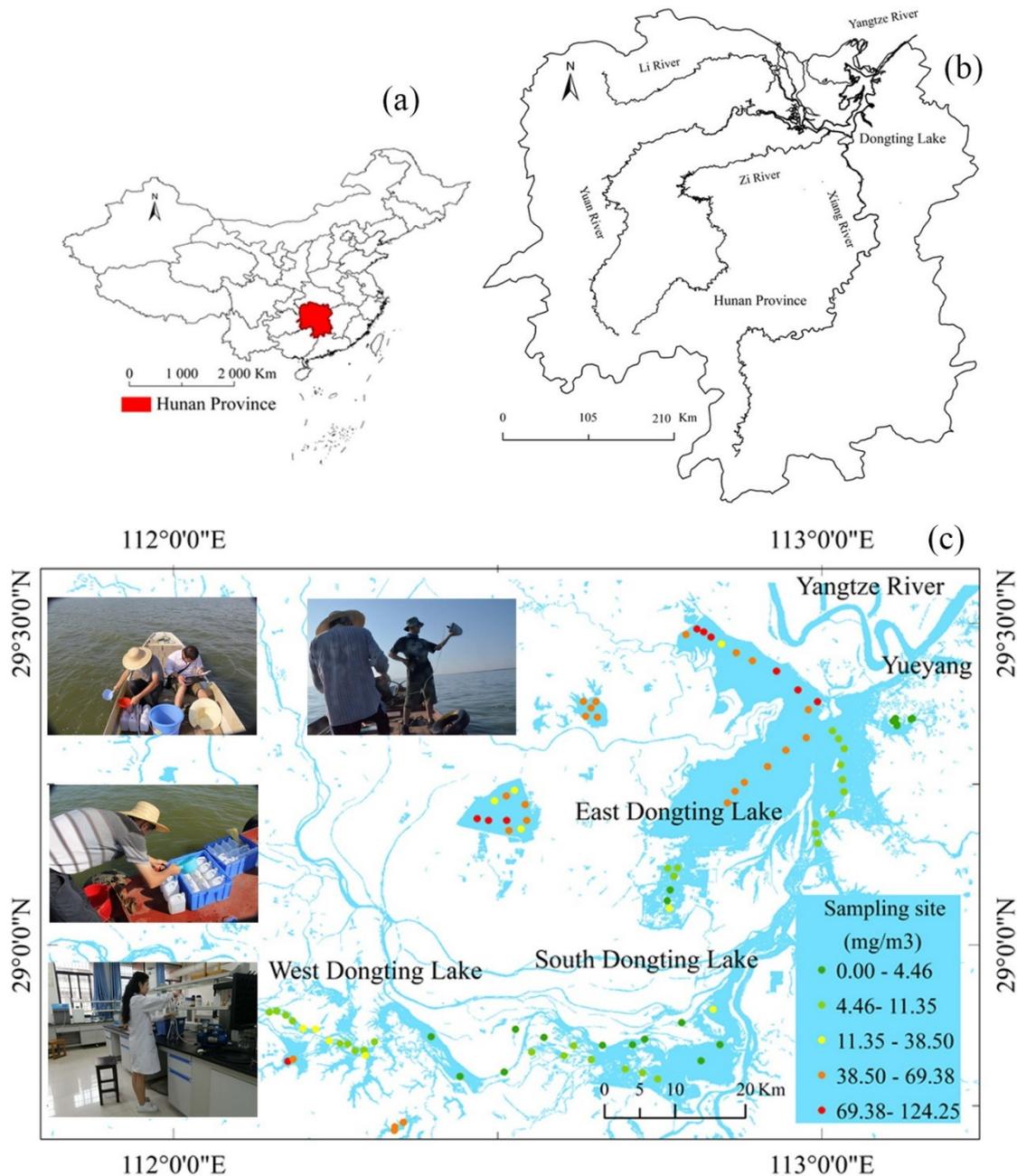
### 2.1. Study Area

Dongting Lake is the second largest freshwater lake in China, located in the north of Hunan Province and south of the Yangtze River. Its geographic location is between  $28^{\circ}30'–29^{\circ}31'$  N and  $111^{\circ}40'–113^{\circ}10'$  E. The four rivers of Xiang, Zi, Yuan, and Li merge into Dongting Lake from south to north. The north is connected to the Yangtze River. It is a typical throughput lake with a total drainage area of 257,200 km<sup>2</sup>, a lake area of 2579 km<sup>2</sup>, a lake length of 143 km, an average lake width of 17 km, a maximum water depth of 23.5 m, an average water depth of 6.39 m, and a water cycle period of about 18.2 days. At present, it is divided into East Dongting Lake, South Dongting Lake, West Dongting Lake, and Datong Lake (Figure 1). Dongting Lake not only has very rich biodiversity resources, but also has the function of regulating the flood and runoff of the Yangtze River. In recent years, with the impact of economic development and human activities, a large number of point source and nonpoint source pollutants were discharged into the basin, and the water quality of Dongting Lake is not currently optimistic. The overall lake is at a mesotrophic level. The water quality is showing a downward trend. The eutrophication of the water body is gradually increasing, and Dongting Lake is beginning to show a eutrophication trend [2]. Carrying out monitoring of Chl-a concentration and understanding the distribution characteristics of Chl-a concentration is of great significance in understanding the status and evolution trend of Dongting Lake eutrophication.

### 2.2. Sampling Plan Design

According to the "Technical Guidelines for Sampling of Lakes and Reservoirs" (GB/T14581-93), the layout of sampling points should consider the hydrodynamic conditions of the lake water body, the area of the lake body, the shape of the lake basin, the replenishment conditions, the water outlet and intake, the location and scale of the sewage facilities, and other conditions. At the same time, considering the particularity of remote sensing, a large area of relatively uniform water area should be selected to arrange sampling points to maximize the correspondence between the "point data" of the ground sampling points and the "surface data" of the remote-sensing data. Considering the inheritance of the data, the sampling points were arranged in combination with the data accumulation of Dongting Lake for many years to facilitate the study of the time evolution of Dongting Lake. The location of the water quality sampling points of Dongting Lake is shown in Figure 1. Due to the large area of Dongting Lake, and it has been divided into multiple lakes, according to the specific conditions, take the speedboat, fishing boat and other means of transportation, divide it into two groups, and carry out ground experiments. To have similar lighting conditions as the remote sensing data, a day with clear weather and a little cloud was selected, and the ground experiment was completed on 9–12 August 2013, which lasted four days. The water spectrum data in the field were measured and the water samples were collected and brought back to the laboratory for analysis to obtain water quality data. A total of 90 samples were collected. During the sampling process, the sampling point number, location, weather conditions, sampling time, sampling point characteristics, and other information were simultaneously recorded. In August 2013, the

monthly average values of environmental factors at the Yangtze River exit of Dongting Lake were: water temperature 32 °C, dissolved oxygen 5.6 mg/L, total nitrogen 1.54 mg/L, total phosphorus 0.048 mg/L, transparency 0.38 m, monthly precipitation 132.2 mm, the maximum daily precipitation is 51.4 mm, the number of days with daily precipitation greater than 0.1 mm is 9 days, the number of sunshine hours is 245.2 h, and the average wind speed is 3.5 m/s.



**Figure 1.** Location of sampling points. (a) Map of China. (b) Location map of Dongting Lake in Hunan (c) The location of sampling points and the distribution of chlorophyll concentration in Dongting Lake.

### 2.3. Spectral Collection and Chl-a Concentration Determination

The water surface spectrum measurement method was used to measure the water surface spectrum. In the process of spectrum collection, to avoid the influence of water specular reflection and ship shadow on the measurement results as much as possible, and to better extract the water reflectance, a certain observation angle is used to observe the water spectrum. The angle of observation azimuth was approximately  $135^\circ$  between the instrument observation plane and the sun incident plane. The angle of observation zenith was approximately  $40^\circ$  between the instrument and water surface normal. These angles avoid most of the reflectance from direct solar radiation, while also reducing the ship-shadow effect. We measured the standard gray plate, water, and sky on the open ship deck 30 cm above the water surface several times and calculated the remote sensing reflectance ( $R_{rs}$ ) as the average of the observation station spectra [37,38]. For each sample point, ten water surface spectra were collected using an ASD Field Spec portable field spectrometer (spectral range 350–1050 nm, spectral resolution 3 nm). Abnormal values were eliminated and an averaging processing was carried out to obtain the measured spectrum value of the sample point. The transparency of the water body was measured using a Saishi plate. The concentration of Chl-a was measured using the hot ethanol method. A certain volume (500–1000 mL) of water sample was taken using a 47 mm diameter GF/C filter membrane and the filter membrane was placed in a refrigerator for more than 48 h. The samples were extracted with 90% hot ethanol, then 90% ethanol was used as the reference solution to measure the absorbance at 665 and 750 nm on a spectrophotometer. One drop of 1% dilute hydrochloric acid was added to acidify and then the Chl-a concentration was calculated.

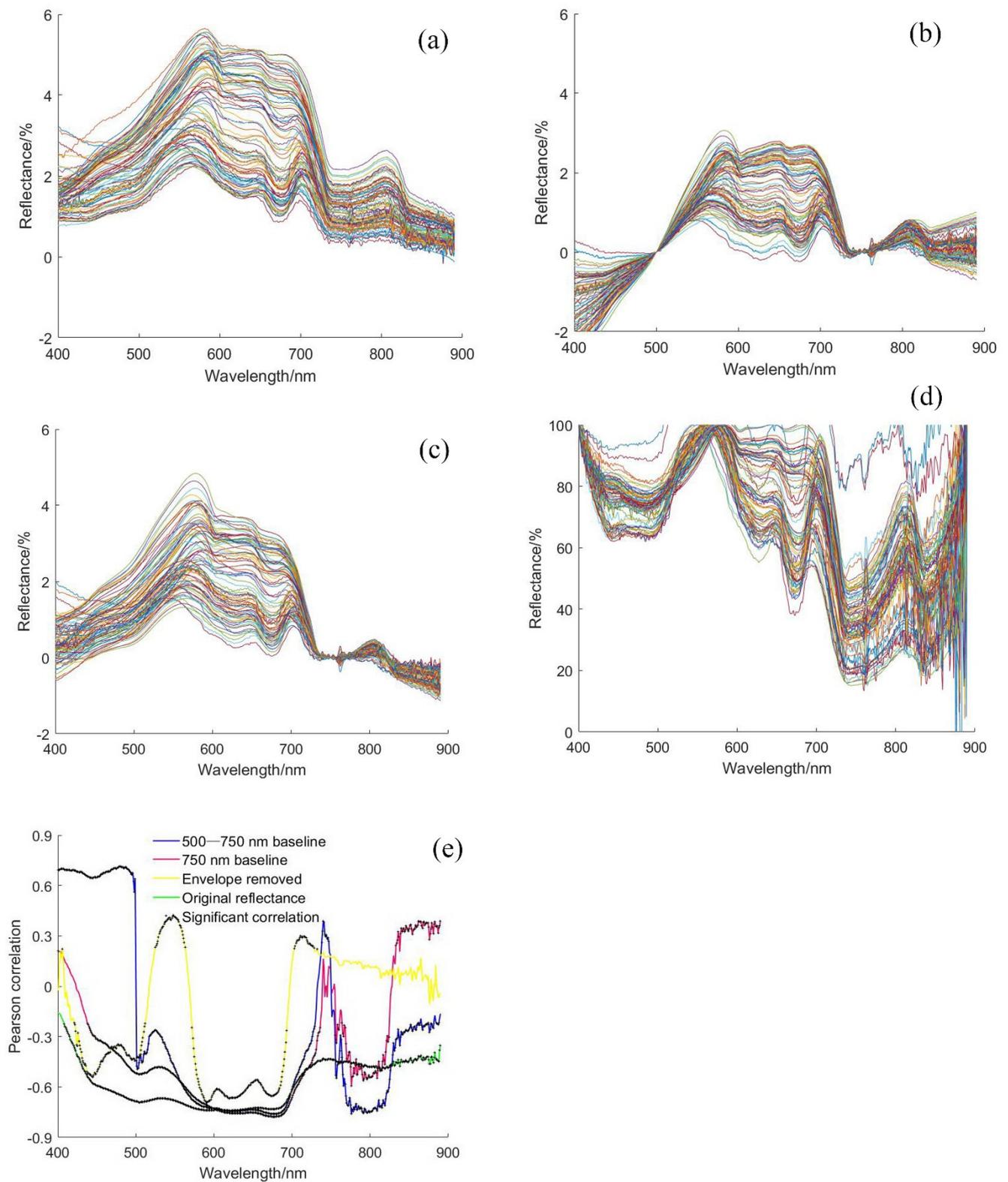
### 2.4. Spectra Pretreatment

#### 2.4.1. Spectral Denoising and Resampling

Spectral measurement is easily affected by many factors, such as observation angle and illumination, which makes the signal-to-noise ratio of spectral data low. Preprocessing of spectral data can reduce its impact. In this study, the low signal-to-noise ratio bands (350–399 nm and 891–1050 nm) were removed. The Savitzky-Golay algorithm was used to smooth the spectrum [39]. The effective bands of each spectrum were resampled at 10 nm intervals (Figure 2a) to reduce the correlation between bands and improve data processing efficiency.

#### 2.4.2. Baseline Correction

The baseline correction involves subtract the baseline value from the spectral reflectance. In the inversion of Chl-a concentration in inland water bodies, it is mainly used to weaken the nonChl-a spectrum information in the measured spectrum. In water color remote sensing research, the 750 nm spectral reflectance of indoor water is approximately zero. The baseline correction with 750 nm as the baseline enhanced the fluorescence peak of chlorophyll by about 700 nm and the absorption peak by about 680 nm in water. In addition, the backscattering coefficient of suspended matter in water at 500–750 nm decreased linearly with the increase in wavelength. The straight line connecting the reflectance values of 500 nm and 750 nm was taken as the baseline and a 500–750 nm baseline correction was performed on all spectral data, highlighting the water body Chl-a spectral information. This study uses 500–750 baseline spectra and 750 baseline spectra as the two preprocessing spectra of the CNN model (Figure 2b,c).



**Figure 2.** Spectral curve characteristics. (a) Original spectrum. (b) Baseline spectrum at 500–750 nm. (c) Baseline spectrum at 750 nm. (d) Envelope removal spectrum. (e) Correlation analysis.

### 2.4.3. Envelope Removal

The envelope is the line connecting the start and end points of the absorption valley of the spectral curve. The envelope removal method is a spectral analysis method that

effectively enhances the absorption characteristics of interest. It effectively highlights the absorption and reflection characteristics of the spectral curve and normalizes them to a consistent background, which is conducive to the comparison of characteristic values with other spectral curves. This study used the envelope removal spectrum as a preprocessing spectrum of the CNN model (Figure 2d).

### 2.5. Inversion Factor Selection

Correlation analysis reflects the strength of the linear relationship between two variables. The Pearson correlation coefficient between the reflectance and Chl-a concentration was used in this work to indicate the strong correlation band of the difference in the chlorophyll concentration of water bodies. The calculation formula is:

$$r = \frac{\text{Cov}(x_i, y)}{\sigma_{x_i} \sigma_y} \quad (1)$$

In the formula,  $r$  represents the Pearson correlation coefficient,  $x_i$  represents the reflectance of the  $i$  band,  $y$  represents the Chl-a concentration,  $\text{Cov}$  represents the covariance, and  $\sigma$  represents the standard deviation. A band with a significance level of 0.05 during the t-test of the correlation coefficient was considered to be a strong correlation band.

The strong correlation band is used as the strong correlation factor (SC) for Chl-a concentration inversion (Figure 2e).

Through principal component transformation, this study transformed the preprocessed spectrum into linearly independent principal component variables. The calculation formula is:

$$C_j = v_{1j}X_1 + v_{2j}X_2 + \dots + v_{ij}X_i + \dots + v_{mj}X_m \quad (2)$$

where  $X_i$  is the reflectance of the  $i$ -band;  $C_j$  is the  $j$ -th new component after the main component transformation;  $m$  is the number of spectral variables before the main component transformation.  $V_{ij}$  is the eigenvector corresponding to the  $i$ -th eigenvalue in the correlation coefficient matrix of the original variables. To avoid the false abandonment of the effective weak signal in the modeling process, all the principal components extracted in this study are retained as the independent variables of the model. These independent variables are used as the principal component factors (PC) of the Chl-a concentration inversion.

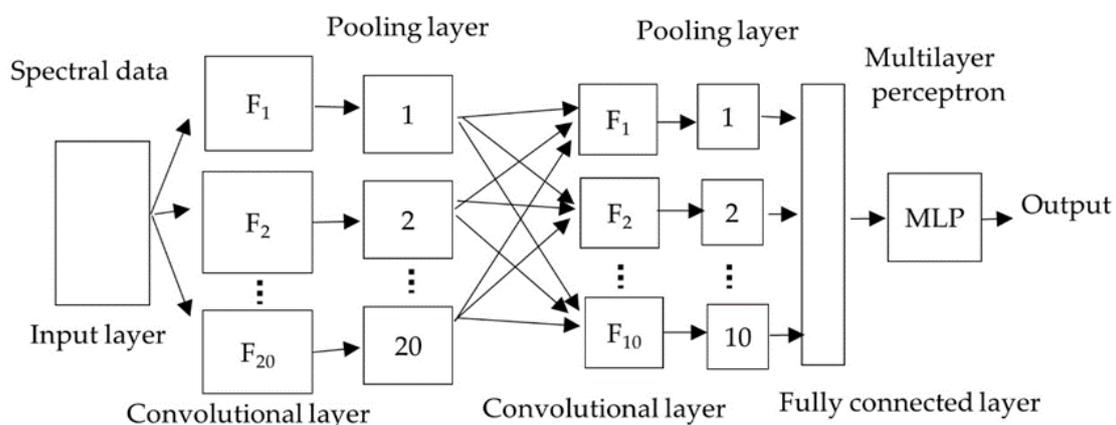
### 2.6. Convolutional Neural Network Model

Convolutional neural network is a feedforward neural network with excellent performance in image target recognition and natural language processing. CNN is mainly composed of three parts, namely, the convolutional layer, the pooling layer, and the fully connected layer. The convolutional layer is composed of multiple feature planes. The neurons in the feature plane are locally connected to the previous feature plane through the convolution kernel. The convolution kernel slides on the feature plane according to a certain step length to achieve weight-sharing. The function of the pooling layer is to downsample the local features extracted by the convolutional layer, reduce network free parameters, and improve the robustness of feature data. Generally, the average pooling or maximum pooling methods are adopted. The fully connected layer flattens the feature map output by the pooling layer and is fully connected with the multilayer perceptron.

The convolution layer performs feature extraction on the convolution operation of the input signal through the convolution kernel. The size and number of convolution kernels directly determine the overall performance of the network. Generally, the smaller the convolution kernel, the more detailed the extracted features, but the overall structure information of the input signal will be lost, which reduces the generalization of the network. Conversely, the larger the convolution kernel, although the overall structure information is preserved, the details of the signal will be missing. The number of convolution kernels is different, which means that the feature extraction methods of the input signal are different. Generally, the number of convolution kernels is positively correlated with the overall

performance of the network, but after increasing to a certain upper limit, not only will the improvement of the network performance become very limited, but also the network calculation scale will increase. The pooling layer down-samples the feature maps extracted by the convolutional layer. The sampling ratio should not be too large. Although an increase in the sampling ratio will improve the generalization ability of the network and reduce the network scale, an excessively large down-sampling ratio will lose a large amount of characteristic information and cause a decrease in network performance. Therefore, the size and number of convolution kernels, the size of downsampling, the number of convolutional layers, and the number of pooling layers all increase the uncertainty of the CNN inversion model of the Chl-a concentration [18,19].

Through repeated tests, this study designed a 7-layer CNN model for Chl-a concentration inversion, including 1 input layer, 2 convolutional layers, 2 pooling layers, 1 fully connected layer, and 1 output layer (Figure 3). The input layer inputted the original spectrum and the convolutional layer was behind the input layer. The size of the convolution kernel of the first convolutional layer was 5, the number was 20, and the linear rectification function was used as the activation function. The size of the convolution kernel of the second convolutional layer was 3, the number was 10, and the linear rectification function was used as the activation function. The pooling layer adopted the mean pooling method and the downsampling scale was 2. The output layer was the dense layer, using a linear function for activation for real value regression. The cross-validation method was used to select the best parameters and the mean square error was used as the cost function of the model. The experimental environment was the Windows 10 system using Python language and Tensorflow 2.0 as the backend to build a CNN model of Chl-a concentration inversion based on the Keras deep learning framework.



**Figure 3.** The structure of the Convolutional Neural Network (CNN) prediction model.

### 2.7. Model Evaluation Criteria

The coefficient of determination ( $R^2$ ), the root mean square error ( $RMSE$ ), the relative percent deviation ( $RPD$ ), and the relative error ( $RE$ ) were selected as the criteria for judging the predictive ability of the model. Among them,  $R^2$  included the determination coefficient of training samples ( $R_T^2$ ) and the determination coefficient of prediction samples ( $R_P^2$ ), and  $RMSE$  included the root mean square error of training samples ( $RMSE_T$ ) and the root mean square error of prediction samples ( $RMSE_P$ ), as follows:

$$\begin{aligned}
 R^2 &= \frac{\sum_1^n (\hat{y}_i - \bar{y})^2}{\sum_1^n (y_i - \bar{y})^2} \\
 RMSE &= \sqrt{\frac{\sum_1^n (\hat{y}_i - y_i)^2}{n}} \\
 RPD &= SD / RMSE \\
 RE &= \sum_{i=1}^n \left| \frac{\hat{y}_i - y_i}{y_i} \right| / n
 \end{aligned} \tag{3}$$

where  $\hat{y}$  is the predicted value of the sample,  $\bar{y}$  is the mean of the measured sample,  $y$  is the measured sample value,  $n$  is the number of samples denoted by  $i = 1, 2 \dots, n$ , and  $SD$  is the standard deviation. Generally, the larger the  $R_p^2$ ,  $R_T^2$ , and  $RPD$ , the smaller the  $RMSE_P$ ,  $RMSE_T$ , and  $RE$ , the higher the model accuracy, and the smaller the deviation.

### 3. Results

#### 3.1. Descriptive Statistics of Chl-a Concentration

Dongting Lake is a typical water-through lake. The water body with fast water exchange has a low Chl-a concentration, while the relatively closed water body has a high Chl-a concentration (maximum 124.25 mg/m<sup>3</sup>). The concentration difference between different sample points was very large (standard deviation up to 30.00 mg/m<sup>3</sup>) (Table 1). The chi-square test showed that the Chl-a concentration data did not obey the normal distribution at the 0.05 significance level. The original data were transformed normally by the Box–Cox method [40]. Divided into two groups, which lasted four days, a total of 90 samples of spectrum and water quality data were collected. Of the samples, 70% were randomly selected as training samples and 30% were selected as test samples.

**Table 1.** Descriptive statistics of Chl-a concentration (mg/m<sup>3</sup>).

Minimum	Maximum	Average Value	Standard Deviation	Skewness	Kurtosis
1.86	124.25	31.35	30.00	0.87	2.82

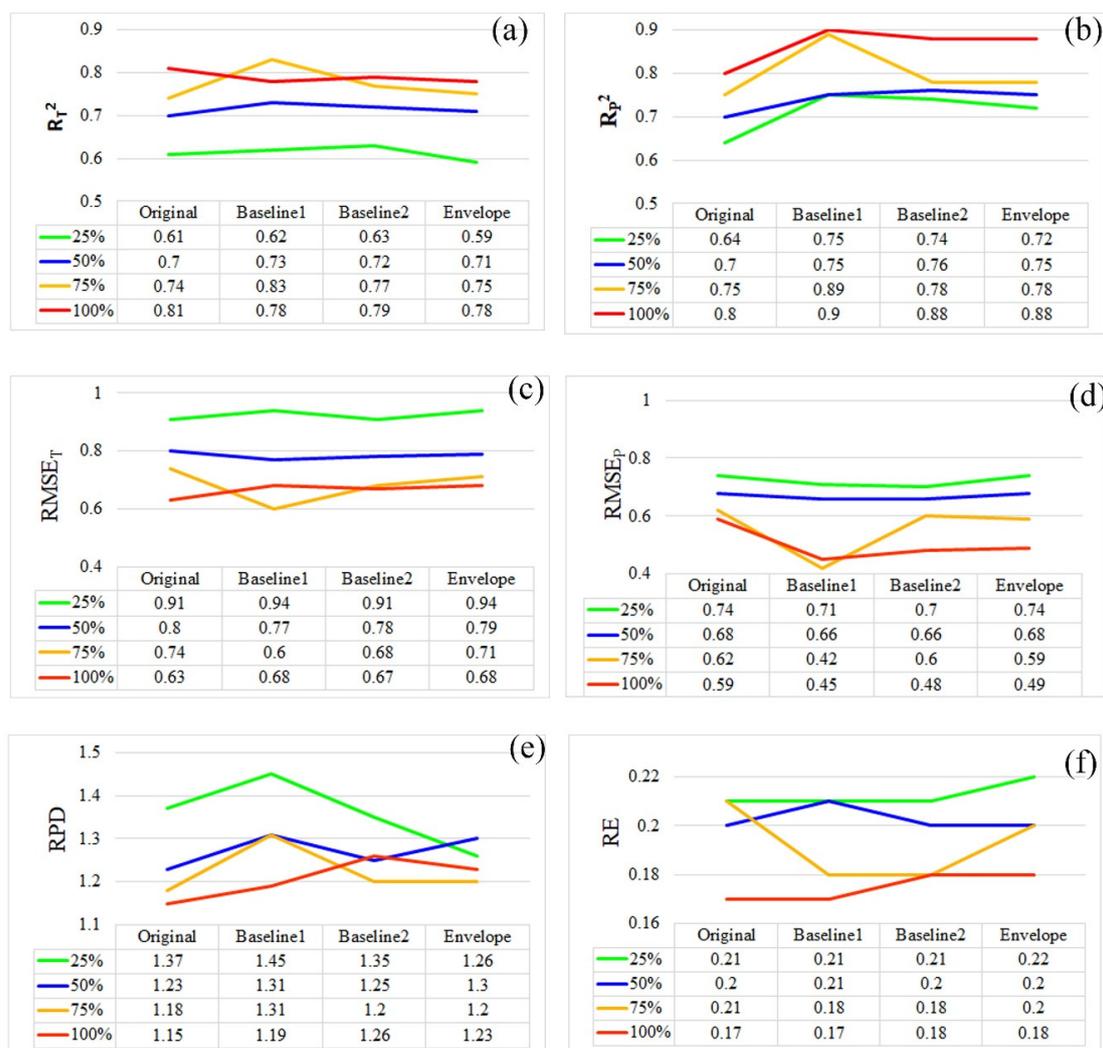
#### 3.2. Spectral Curve Characteristics and Characteristic Band Recognition

As seen in Figure 2a, before 580 nm, the spectral reflectance rose rapidly as the wavelength increased, reaching a peak near 585 nm, then revealing tiny reflection peaks near 650 nm, 700 nm, and 800 nm, with absorption valleys at 630 nm, 680 nm, and 740 nm. On the whole, the waveband after 600 nm, especially the waveband between 600 and 800 nm, showed larger fluctuations in the spectral curve, with larger reflection peaks at 700 nm and larger absorption valleys at 740–780 nm. The Pearson correlation coefficient between the Chl-a concentration and the spectrum was calculated (Figure 2e). The results showed many significant correlation bands between the four spectra and the water Chl-a concentration, and the strong correlation bands with a correlation coefficient greater than 0.5 were mainly concentrated between 578 and 695 nm. The inverse principal component transformation of regression coefficients is an effective way to reconstruct the spectrum, obtain the contribution intensity of each band, and identify the characteristic band. The principal component regression curves of the four spectra are shown in Figure 2f.

#### 3.3. Robustness Analysis of Chl-a Concentration CNN Model

Of the training samples, 25%, 50%, 75%, and 100% (i.e., 16, 32, 47, and 63 samples) were selected in the model as new training samples to train the model. The test samples were all 27 samples for model verification. Taking the strong correlation factors of the four preprocessing spectrum types (original spectrum, 500–750 baseline spectrum, 750 baseline spectrum, and envelope removal spectrum) as input independent variables, the concentration of Chl-a was the dependent variable and the concentration of Chl-a was

retrieved using the CNN model. The performance of the model is shown in Figure 4. It can be seen from the results that: (1) When the number of samples was 16, the overall performance of the CNN model of the 500–750 baseline spectrum was the best ( $R_T^2 = 0.62$ ,  $RMSE_T = 0.94$ ,  $R_P^2 = 0.75$ ,  $RMSE_P = 0.71$ ,  $RPD = 1.45$ , and  $RE = 0.21$ ), and the generalization performance of the CNN model of the original spectrum was the worst ( $R_T^2 = 0.61$ ,  $RMSE_T = 0.91$ ,  $R_P^2 = 0.64$ ,  $RMSE_P = 0.74$ ,  $RPD = 1.37$ , and  $RE = 0.21$ ), whereas when the number of samples was 63, the overall performance of the CNN model of the same 500–750 baseline spectrum was the best ( $R_T^2 = 0.78$ ,  $RMSE_T = 0.68$ ,  $R_P^2 = 0.90$ ,  $RMSE_P = 0.45$ ,  $RPD = 1.19$ , and  $RE = 0.17$ ), and the generalization performance of the CNN model of the original spectrum was the worst ( $R_T^2 = 0.81$ ,  $RMSE_T = 0.63$ ,  $R_P^2 = 0.80$ ,  $RMSE_P = 0.59$ ,  $RPD = 1.15$ , and  $RE = 0.17$ ); (2) generally speaking, the more samples there were, the better the overall performance of the CNN models of the four spectra were improved. Under different sample sizes, the average performances of the CNN models of the four spectra were ranked from good to bad: 500–750 baseline spectrum ( $R_T^2 = 0.74$ ,  $RMSE_T = 0.75$ ,  $R_P^2 = 0.82$ ,  $RMSE_P = 0.56$ ,  $RPD = 1.32$ , and  $RE = 0.19$ ), 750 baseline spectrum ( $R_T^2 = 0.73$ ,  $RMSE_T = 0.76$ ,  $R_P^2 = 0.79$ ,  $RMSE_P = 0.61$ ,  $RPD = 1.27$ , and  $RE = 0.19$ ), envelope removal spectrum ( $R_T^2 = 0.71$ ,  $RMSE_T = 0.78$ ,  $R_P^2 = 0.78$ ,  $RMSE_P = 0.63$ ,  $RPD = 1.25$ , and  $RE = 0.20$ ) and original spectrum ( $R_T^2 = 0.72$ ,  $RMSE_T = 0.77$ ,  $R_P^2 = 0.72$ ,  $RMSE_P = 0.66$ ,  $RPD = 1.23$ , and  $RE = 0.20$ ).



**Figure 4.** Model robustness evaluation analysis. (a)  $R_T^2$ , (b)  $R_P^2$ , (c)  $RMSE_T$ , (d)  $RMSE_P$ , (e) RPD and (f) RE. In the figure, Baseline1 is a 500–750 nm baseline, and Baseline2 is a 750 nm baseline.

### 3.4. Accuracy Analysis of Chl-a Concentration Inversion

Tables 2 and 3, and Figure 5 show the inversion results of the CNN model after eight combinations of two inversion factors (strong correlation factor and principal component factor) of four spectral types (original spectrum, 500–750 baseline spectrum, 750 baseline spectrum, and envelope removal spectrum). The inversion effect of the CNN model of the preprocessed spectrum was good (500–750 baseline spectrum, 750 baseline spectrum, and envelope removal spectrum), and the CNN model of the original spectrum showed the worst inversion effect (Tables 2 and 3, and Figure 4). The results showed that preprocessing significantly improved the model inversion effect. The inversion effect of the CNN model with strong correlation factor ( $R_T^2 = 0.79$ ,  $RMSE_T = 0.67$ ,  $R_P^2 = 0.87$ ,  $RMSE_P = 0.50$ ,  $RPD = 1.21$ , and  $RE = 0.18$ ) and the CNN model with principal component factor ( $R_T^2 = 0.76$ ,  $RMSE_T = 0.71$ ,  $R_P^2 = 0.85$ ,  $RMSE_P = 0.53$ ,  $RPD = 1.24$ , and  $RE = 0.18$ ) were not much different (Table 2 and Figure 5), showing that the selection of inversion factors had little effect on the performance of the model. Among the combined models, the CNN model of Baseline1\_SC ( $R_T^2 = 0.78$ ,  $RMSE_T = 0.68$ ,  $R_P^2 = 0.90$ ,  $RMSE_P = 0.45$ ,  $RPD = 1.19$ , and  $RE = 0.17$ ) showed the best effect and the average inversion effect of 8 CNN models is also better ( $R_T^2 = 0.78$ ,  $RMSE_T = 0.69$ ,  $R_P^2 = 0.86$ ,  $RMSE_P = 0.52$ ,  $RPD = 1.23$  and  $RE = 0.18$ ) (Table 2 and Figure 5). All showed higher  $R_P^2$  and lower  $RMSE_P$ , indicating the feasibility of applying CNN to the modeling of Chl-a concentration hyperspectral inversion. To verify the superiority of CNN model applied to Chl-a concentration inversion modeling, the performances of the CNN, linear regression, and partial least square regression models for Chl-a concentration inversion were compared. The performance of the CNN model (Baseline1\_SC ( $R_P^2 = 0.90$ ,  $RMSE_P = 0.45$ )) was much better than the same combination of traditional models, i.e., the linear regression model (LR) (Baseline1\_SC ( $R_P^2 = 0.61$ ,  $RMSE_P = 0.72$ )) and the partial least squares regression model (PLSR) (Baseline1\_SC ( $R_P^2 = 0.58$ ,  $RMSE_P = 0.95$ )). This showed the superiority of CNN applied to the inversion modeling of Chl-a concentration.

**Table 2.** Inversion results of the combined model.

Pretreatment Method	Inversion Factor	Alias	$R_T^2$	$RMSE_T$	$R_P^2$	$RMSE_P$	RPD	RE
Original	SC	Original_SC	0.81	0.63	0.80	0.59	1.15	0.17
	PC	Original_PC	0.65	0.85	0.78	0.62	1.33	0.20
Baseline 500–750 nm	SC	Baseline1_SC	0.78	0.68	0.90	0.45	1.19	0.17
	PC	Baseline1_PC	0.81	0.65	0.85	0.56	1.20	0.17
Baseline 750 nm	SC	Baseline2_SC	0.79	0.67	0.88	0.48	1.26	0.18
	PC	Baseline2_PC	0.78	0.68	0.87	0.48	1.26	0.18
Envelope removed	SC	Envelope_SC	0.78	0.68	0.88	0.49	1.23	0.18
	PC	Envelope_PC	0.80	0.65	0.88	0.47	1.18	0.17

**Table 3.** Means of various combination models.

Pretreatment Method	$R_T^2$	$RMSE_T$	$R_P^2$	$RMSE_P$	RPD	RE
Original	0.73	0.74	0.79	0.61	1.24	0.19
Baseline 1	0.80	0.67	0.88	0.51	1.20	0.17
Baseline 2	0.79	0.68	0.88	0.48	1.26	0.18
Envelope	0.79	0.67	0.88	0.48	1.21	0.18
All models	0.78	0.69	0.86	0.52	1.23	0.18

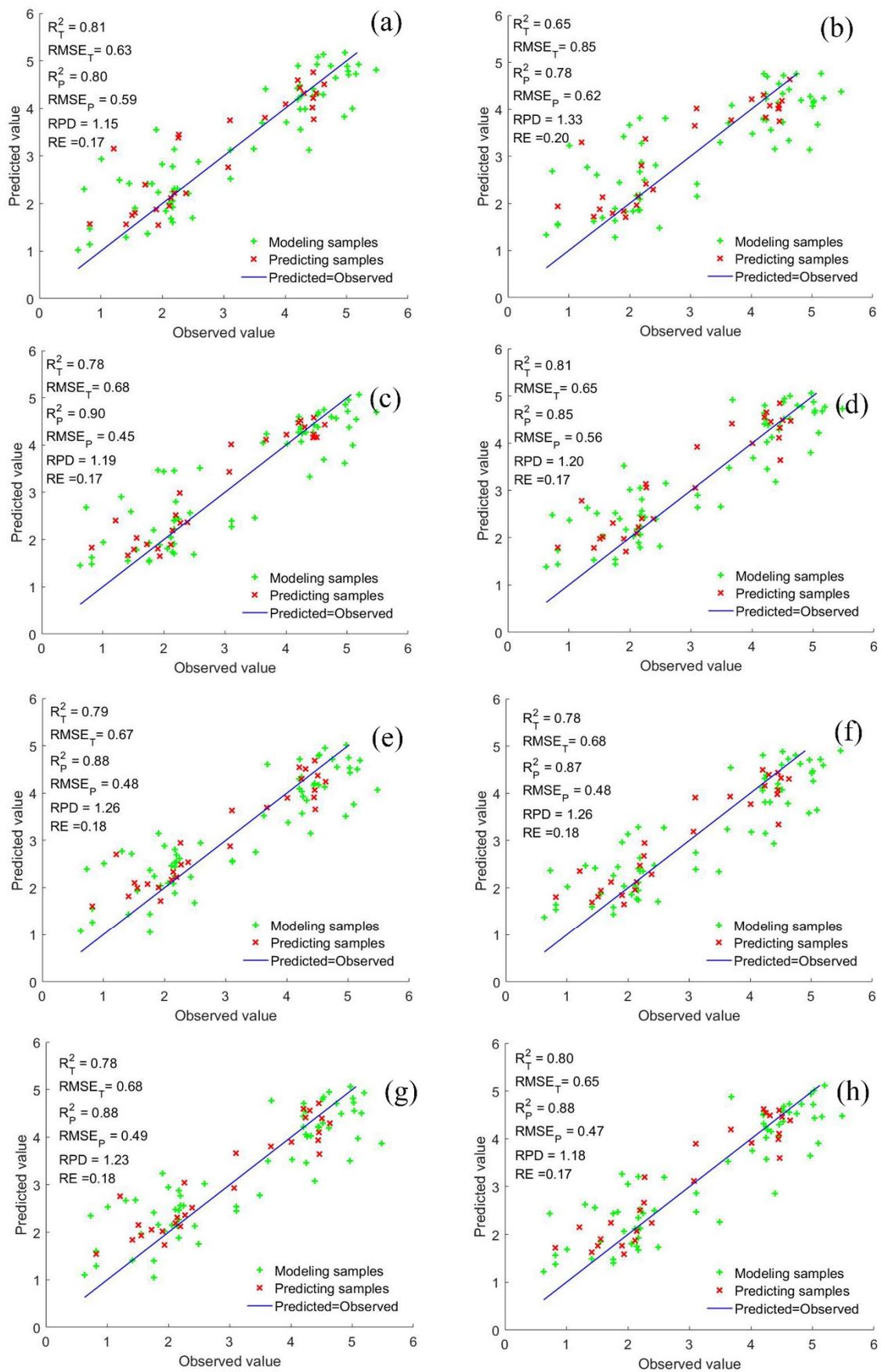
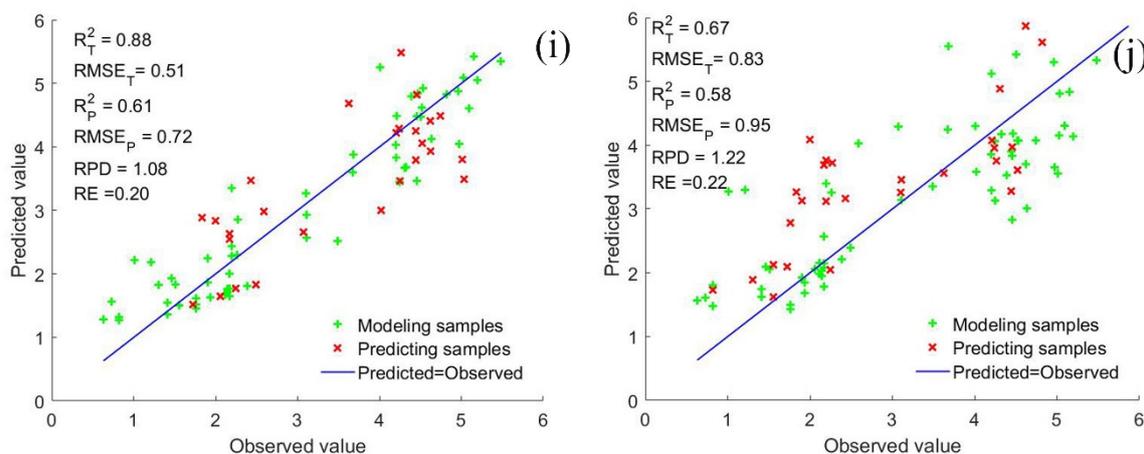


Figure 5. Cont.



**Figure 5.** Inversion results. (a) Original\_SC, (b) Original\_PC, (c) Baseline1\_SC, (d) Baseline1\_PC, (e) Baseline2\_SC, (f) Baseline2\_PC, (g) Envelope\_SC, (h) Envelope\_PC, (i) LR, and (j) PLSR.

#### 4. Discussion

As a high-performance deep learning model, CNN has stronger feature extraction and model expression capabilities than shallow learning methods. However, for Case-II water bodies with relatively complex water qualities, it is still challenging to establish a highly applicable and accurate Chl-a concentration inversion model. Few research works exist on remote sensing inversion modeling of Chl-a concentration using CNN [36]. This study explored the possibility of quantitatively inverting the Chl-a concentration based on measured spectra using the CNN model. Its novelty lies in the inversion of Chl-a concentration using a CNN model with four types of spectra (original spectrum, 500–750 baseline spectrum, 750 baseline spectrum, and envelope removal spectrum) and two inversion factors (strong correlation factor and principal component factor).

Among the combined models of CNN, the CNN model of Baseline1\_SC showed the best effect, and the average inversion effect of the eight combined CNN models was better, indicating the feasibility of the CNN inversion modeling of the Chl-a concentration. The performance of the CNN model was far better than the same combination of traditional models, i.e., the linear regression model and partial least squares regression model, indicating the superiority of the CNN inversion modeling of Chl-a concentration. This was similar to previous studies [41–46]. Xu Y et al. found in a comparative study of machine learning models used to retrieve Chl-a concentration in Lake Taihu that deep learning methods had better inversion accuracy and better model stability [42]. Wang C et al. found in a study of near-infrared spectroscopy used to predict soil moisture content that the prediction accuracy of the CNN model was better than traditional models, such as PLSR [46]. However, previous studies did not provide the optimal scheme for applying the CNN model with different preprocessing methods and different inversion factor combinations to the inversion modeling of Chl-a concentration in order to provide a reference for the inversion modeling of Chl-a concentration.

This study randomly selected 25%, 50%, 75%, and 100% of the samples to train the model to test the robustness of the CNN model. The results showed that as the number of training samples increased, the overall performances of the CNN models for the four spectra were improved. The number of samples had an impact on the performance of the CNN model. Enough samples improved the stability of the CNN model, which was consistent with the conclusions of other studies [36,47–49]. However, the cost of obtaining a large number of in situ samples is high. Considering the integration of laboratory samples and in situ samples and increasing the number of samples is an issue that needs to be considered to improve the robustness of CNN. This study only considered the influence of sample size on the robustness of the model, but the Chl-a concentration is also affected

by water quality parameters such as water temperature, dissolved oxygen, dissolved phosphorus, and total nitrogen. The influence of water quality parameters of inland water bodies, coastal water bodies, and ocean water bodies carries certain differences [29]. A Chl-a concentration inversion model that takes into account the influence of different regions and different types of water quality parameters is a problem that needs to be considered to improve the spatial robustness of the model. The robustness of the model is also affected by time. Although this study was an experiment carried out by simultaneous sampling of water quality parameters and spectrum, the sampling lacked time continuity. The dominant substances in different water bodies were often different, resulting in different optical properties and different Chl-a concentrations in water bodies. A Chl-a concentration inversion model taking into account the time duration is another issue that needs to be considered to improve the time robustness of the model.

This research establishes a semi-empirical model, and the inversion result of the model is affected by sampling location, time, and weather conditions. To eliminate the influence of these factors, a physical model based on the theory of radiation transmission needs to be established. However, due to the complexity of the physical model and too many parameters, the application of this model is limited. The characteristics of water spectrum are complex, regional, seasonal, etc., resulting in the same algorithm can only guarantee its accuracy in a certain area. Limited experimental data led to temporal and spatial limitations in the built model. If more sample data are used, the resulting impact could be eliminated or weakened so the validity and applicable boundaries of the model can be further analyzed and discussed.

## 5. Conclusions

Taking China's Dongting Lake as an example, 90 water samples and their spectra were collected in this study. Using eight combinations of two inversion factors (strong correlation factor and principal component factor) of four spectral types (original spectra, 500–750 baseline spectra, 750 baseline spectra, and envelope removal spectra) as independent variables and Chl-a concentration as the dependent variable, a CNN model was constructed to invert Chl-a concentration. The results showed that: (1) As the number of training samples increased, the overall performances of the CNN model for the four spectra were improved. Under different sample sizes, the average performances of the CNN models of the four spectra were ranked from good to bad, namely, 500–750 baseline spectrum, 750 baseline spectrum, envelope removal spectrum, and original spectrum; (2) the CNN model of the preprocessing spectrum showed a good inversion effect, and the CNN model of the original spectrum showed the worst inversion effect, indicating that the preprocessing significantly improved the inversion effect of the model; (3) the inversion effects of the CNN model with the strong correlation factor and the CNN model with the principal component factor were not much different, indicating that the choice of inversion factor had little effect on the performance of the model; (4) among the combined models, the CNN model of Baseline1\_SC showed the best effect and the average inversion effects of the eight combinations of CNN models were better, indicating the feasibility of applying CNN to the inversion modeling of water body Chl-a concentration; (5) the performance of the CNN model was far better than the traditional model of the same combination, i.e., the linear regression and partial least squares regression models, indicating the superiority of the CNN inversion modeling of water body Chl-a concentration.

**Author Contributions:** Y.X. provided fund support. B.Z. provided technical guidance. Y.X. and L.Z. conceived the methodology. Y.-h.L., Y.-m.W. and S.-l.Z. participated in the revision. Y.X. made the data processing, the analysis of them and the writing of the paper. All authors have read and agreed to the published version of the manuscript.

**Funding:** This project is funded by the Open Fund of the Key Laboratory of Non-ferrous Metal Mineralization Prediction and Geological Environment Monitoring (Central South University) of the Ministry of Education (2020YSJS12) and the Hunan Provincial Science and Technology Department (2019SK2112, 2019TP2079, 2018TP1042, 2017TP2006).

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The data used to support the findings of this study are available from the corresponding author upon request.

**Conflicts of Interest:** The authors declare no competing interests.

## References

1. Guan, Q.; Feng, L.; Hou, X.; Schurgers, G.; Zheng, Y.; Tang, J. Eutrophication changes in fifty large lakes on the Yangtze Plain of China derived from MERIS and OLCI observations. *Remote Sens. Environ.* **2020**, *246*, 111890. [[CrossRef](#)]
2. Xue, Y.; Liu, F.; Liu, J.; Sun, Y. Annual variation characteristics of eutrophication in Dongting Lake, China. *Proc. Inst. Civ. Eng. Water Manag.* **2020**, *173*, 208–216. [[CrossRef](#)]
3. Dong, S.R.; Jiang, J.; Wang, Y.J.; Li, C.L.; Shi, Y.; Yang, Y.; Yang, Y.; Li, L.H.; Cai, B.; You, J.B.; et al. Impact of water body environments on the microbial community of *Oncomelania hupensis* snails in marshlands around the eastern Dongting Lake. *Zhongguo Xue Xi Chong Bing Fang Zhi Za Zhi = Chin. J. Schistosomiasis Control* **2020**, *32*, 132–139.
4. Huang, Z.; Shang, X.H.; Li, Z.M.; Yin, X.C.; Hu, Y.M. Investigation on the content of persistent organic pollutants in fish from Dongting Lake in China. *Zhonghua Yu Fang Yi Xue Za Zhi [Chin. J. Prev. Med.]* **2019**, *53*, 1236–1241.
5. Xue, J.; Xia, S.; Zhang, L.; Abe, E.M.; Zhou, J.; Li, Y.; Hao, Y.; Wang, Q.; Xu, J.; Li, S.; et al. High-resolution remote sensing-based spatial modeling for the prediction of potential risk areas of schistosomiasis in the Dongting Lake area, China. *Acta Trop.* **2019**, *199*, 105077. [[CrossRef](#)]
6. Gai, Y.; Yu, D.; Zhou, Y.; Yang, L.; Chen, C.; Chen, J. An Improved Model for Chlorophyll-a Concentration Retrieval in Coastal Waters Based on UAV-Borne Hyperspectral Imagery: A Case Study in Qingdao, China. *Water* **2020**, *12*, 2769. [[CrossRef](#)]
7. Yang, X.; Jiang, Y.; Deng, X.; Zheng, Y.; Yue, Z. Temporal and Spatial Variations of Chlorophyll a Concentration and Eutrophication Assessment (1987–2018) of Donghu Lake in Wuhan Using Landsat Images. *Water* **2020**, *12*, 2192. [[CrossRef](#)]
8. Shin, Y.; Kim, T.; Hong, S.; Lee, S.; Lee, E.; Hong, S.; Lee, C.; Kim, T.; Park, M.S.; Park, J.; et al. Prediction of Chlorophyll-a Concentrations in the Nakdong River Using Machine Learning Methods. *Water* **2020**, *12*, 1822. [[CrossRef](#)]
9. Kim, K.B.; Jung, M.; Tsang, Y.F.; Kwon, H. Stochastic modeling of chlorophyll-a for probabilistic assessment and monitoring of algae blooms in the Lower Nakdong River, South Korea. *J. Hazard. Mater.* **2020**, *400*, 123066. [[CrossRef](#)] [[PubMed](#)]
10. Papefus, M.; Schaeffer, B.; Pollard, A.I.; Loftin, K. Exploring the potential value of satellite remote sensing to monitor chlorophyll-a for US lakes and reservoirs. *Environ. Monit. Assess.* **2020**, *192*, 1–22. [[CrossRef](#)]
11. Cao, Z.; Ma, R.; Duan, H.; Pahlevan, N.; Melack, J.; Shen, M.; Xue, K. A machine learning approach to estimate chlorophyll-a from Landsat-8 measurements in inland lakes. *Remote Sens. Environ.* **2020**, *248*, 111974. [[CrossRef](#)]
12. Yu, Y.; Chen, S.; Qin, W.; Lu, T.; Li, J.; Cao, Y. A Semi-Empirical Chlorophyll-a Retrieval Algorithm Considering the Effects of Sun Glint, Bottom Reflectance, and Non-Algal Particles in the Optically Shallow Water Zones of Sanya Bay Using SPOT6 Data. *Remote Sens.* **2020**, *12*, 2765. [[CrossRef](#)]
13. Neil, C.; Spyarakos, E.; Hunter, P.D.; Tyler, A.N. A global approach for chlorophyll-a retrieval across optically complex inland waters based on optical water types. *Remote Sens. Environ.* **2019**, *229*, 159–178. [[CrossRef](#)]
14. Buma, W.G.; Lee, S. Evaluation of Sentinel-2 and Landsat 8 Images for Estimating Chlorophyll-a Concentrations in Lake Chad, Africa. *Remote Sens.* **2020**, *12*, 2437. [[CrossRef](#)]
15. Xie, T.; Chen, Y.; Lu, W. Retrieval of Chlorophyll-a in Lower Reaches of the Minjiang River Via Three-Band Bio-Optical Model. *Laser Optoelectron. Prog.* **2020**, *57*, 071701. [[CrossRef](#)]
16. Lu, X.; Situ, C.; Ma, W.; Dai, X.; He, C.; Li, L.; Zhou, L. Comparative analysis of four semi-analytical models for estimating chlorophyll-a concentration in case-2 waters using field hyperspectral reflectance. *Int. J. Remote Sens.* **2020**, *41*, 584–594. [[CrossRef](#)]
17. Xu, M.; Liu, H.; Beck, R.; Lekki, J.; Yang, B.; Shu, S.; Kang, E.L.; Anderson, R.; Johansen, R.; Emery, E.; et al. A spectral space partition guided ensemble method for retrieving chlorophyll-a concentration in inland waters from Sentinel-2A satellite imagery. *J. Great Lakes Res.* **2019**, *45*, 454–465. [[CrossRef](#)]
18. Cheng, C.; Wei, Y.; Lv, G.; Xu, N. Remote sensing estimation of chlorophyll-a concentration in Taihu Lake considering spatial and temporal variations. *Environ. Monit. Assess.* **2019**, *191*, 84. [[CrossRef](#)] [[PubMed](#)]
19. Menon, H.B.; Adhikari, A. Remote Sensing of Chlorophyll-A in Case II Waters: A Novel Approach With Improved Accuracy Over Widely Implemented Turbid Water Indices. *J. Geophys. Res. Oceans* **2018**, *123*, 8138–8158. [[CrossRef](#)]
20. Kwon, Y.S.; Baek, S.H.; Lim, Y.K.; Pyo, J.; Ligaray, M.; Park, Y.; Cho, K.H. Monitoring Coastal Chlorophyll-a Concentrations in Coastal Areas Using Machine Learning Models. *Water* **2018**, *10*, 1020. [[CrossRef](#)]
21. Malahlela, O.E.; Oliphant, T.; Tsoeleng, L.T.; Mhangara, P. Mapping chlorophyll-a concentrations in a cyanobacteria- and algae-impacted Vaal Dam using Landsat 8 OLI data. *S. Afr. J. Sci.* **2018**, *114*, 64–72. [[CrossRef](#)]
22. Gao, B.; Li, R. Improving Water Leaving Reflectance Retrievals from ABI and AHI Data Acquired Over Case 2 Waters from Present Geostationary Weather Satellite Platforms. *Remote Sens.* **2020**, *12*, 3257. [[CrossRef](#)]
23. Renosh, P.R.; Doxaran, D.; De Keukelaere, L.; Ignacio Gossn, J. Evaluation of Atmospheric Correction Algorithms for Sentinel-2-MSI and Sentinel-3-OLCI in Highly Turbid Estuarine Waters. *Remote Sens.* **2020**, *12*, 1285. [[CrossRef](#)]

24. Xue, K.; Boss, E.; Ma, R.; Shen, M. Algorithm to derive inherent optics properties from remote sensing reflectance in turbid and eutrophic lakes. *Appl. Opt.* **2019**, *58*, 8549–8564. [[CrossRef](#)] [[PubMed](#)]
25. Ignacio Gossn, J.; Ruddick, K.G.; Ines Dogliotti, A. Atmospheric Correction of OLCI Imagery over Extremely Turbid Waters Based on the Red, NIR and 1016 nm Bands and a New Baseline Residual Technique. *Remote Sens.* **2019**, *11*, 220. [[CrossRef](#)]
26. Wei, Y.; Wang, G.; Cheng, C.; Zhang, J.; Sun, X. Baseline Correction of Spectrum for the Inversion of Chlorophyll-a Concentration in the Turbidity Water. *Spectrosc. Spect. Anal.* **2012**, *32*, 2546–2550.
27. Rodriguez-Lopez, L.; Duran-Llacer, I.; Gonzalez-Rodriguez, L.; Abarca-del-Rio, R.; Cardenas, R.; Parra, O.; Martinez-Retureta, R.; Urrutia, R. Spectral analysis using LANDSAT images to monitor the chlorophyll-a concentration in Lake Laja in Chile. *Ecol. Inform.* **2020**, *60*, 101183. [[CrossRef](#)]
28. Pirasteh, S.; Mollae, S.; Fatholahi, S.N.; Li, J. Estimation of Phytoplankton Chlorophyll-a Concentrations in the Western Basin of Lake Erie Using Sentinel-2 and Sentinel-3 Data. *Can. J. Remote Sens.* **2020**, *46*, 585–602. [[CrossRef](#)]
29. Liu, Y.; Yao, L.; Xia, Z.; Gao, Y.; Gong, Z. Geographical discrimination and adulteration analysis for edible oils using two-dimensional correlation spectroscopy and convolutional neural networks (CNNs). *Spectrochim. Acta Part A* **2021**, *246*, 118973. [[CrossRef](#)] [[PubMed](#)]
30. Zeng, Y.; Dai, H. CT Image Segmentation of Liver Tumor with Deep Convolutional Neural Network. *J. Med. Imag. Health* **2021**, *11*, 337–344. [[CrossRef](#)]
31. Chen, T.; Sun, Y.; Li, T. A semi-parametric estimation method for the quantile spectrum with an application to earthquake classification using convolutional neural network. *Comput. Stat. Data Anal.* **2021**, *154*, 107069. [[CrossRef](#)]
32. Lv, Z.; Ding, H.; Wang, L.; Zou, Q. A Convolutional Neural Network Using Dinucleotide One-hot Encoder for identifying DNA N6-Methyladenine Sites in the Rice Genome. *Neurocomputing* **2021**, *422*, 214–221. [[CrossRef](#)]
33. Yu, B.; Xu, L.; Peng, J.; Hu, Z.; Wong, A. Global chlorophyll-a concentration estimation from moderate resolution imaging spectroradiometer using convolutional neural networks. *J. Appl. Remote Sens.* **2020**, *14*, 034520. [[CrossRef](#)]
34. Han, Z.; He, Y.; Liu, G.; Perrie, W. Application of DINCAE to Reconstruct the Gaps in Chlorophyll-a Satellite Observations in the South China Sea and West Philippine Sea. *Remote Sens.* **2020**, *12*, 480. [[CrossRef](#)]
35. Wang, X.; Li, Z.; Wang, W.; Wang, J. Chlorophyll content for millet leaf using hyperspectral imaging and an attention-convolutional neural network. *Cienc. Rural* **2020**, *50*, e20190731. [[CrossRef](#)]
36. Syariz, M.A.; Lin, C.; Nguyen, M.V.; Jaelani, L.M.; Blanco, A.C. WaterNet: A Convolutional Neural Network for Chlorophyll-a Concentration Retrieval. *Remote Sens.* **2020**, *12*, 1966. [[CrossRef](#)]
37. Xing, Q.; Lou, M.; Chen, C.; Shi, P. Using in situ and Satellite Hyperspectral Data to Estimate the Surface Suspended Sediments Concentrations in the Pearl River Estuary. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2013**, *6*, 731–738. [[CrossRef](#)]
38. Zhang, F.; Li, J.; Shen, Q.; Zhang, B.; Wu, C.; Wu, Y.; Wang, G.; Wang, S.; Lu, Z. Algorithms and Schemes for Chlorophyll a Estimation by Remote Sensing and Optical Classification for Turbid Lake Taihu, China. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2015**, *8*, 350–364. [[CrossRef](#)]
39. Sun, X.; Ma, R.; Chen, Y.; Zhen, H.; Ma, C. Experimental Study on Determination of Trisodium Phosphate Concentration in Water Based on Spectral Technique. *Spectrosc. Spect. Anal.* **2020**, *40*, 3205–3210.
40. Cheddad, A. On Box-Cox Transformation for Image Normality and Pattern Classification. *IEEE Access* **2020**, *8*, 154975–154983. [[CrossRef](#)]
41. Wang, Y.; Li, M.; Ji, R.; Wang, M.; Zheng, L. Comparison of Soil Total Nitrogen Content Prediction Models Based on Vis-NIR Spectroscopy. *Sensors* **2020**, *20*, 7078.
42. Caroppo, A.; Leone, A.; Siciliano, P. Comparison Between Deep Learning Models and Traditional Machine Learning Approaches for Facial Expression Recognition in Ageing Adults. *J. Comput. Sci. Technol.* **2020**, *35*, 1127–1146. [[CrossRef](#)]
43. Annala, L.; Ayramo, S.; Polonen, I. Comparison of Machine Learning Methods in Stochastic Skin Optical Model Inversion. *Appl. Sci.* **2020**, *10*, 7097. [[CrossRef](#)]
44. Zhang, W.; Liu, H.; Wu, W.; Zhan, L.; Wei, J. Mapping Rice Paddy Based on Machine Learning with Sentinel-2 Multi-Temporal Data: Model Comparison and Transferability. *Remote Sens.* **2020**, *12*, 1620. [[CrossRef](#)]
45. Xu, Y.; Dong, X.; Wang, J. Use of Remote Multispectral Imaging to Monitor Chlorophyll-a in Taihu Lake: A Comparison of Four Machine Learning Models. *J. Hydroecol.* **2019**, *40*, 48–57.
46. Wang, C.; Wu, X.; Li, L.; Wang, Y.; Li, Z. Convolutional Neural Network Application in Prediction of Soil Moisture Content. *Spectrosc. Spect. Anal.* **2018**, *38*, 36–41.
47. Moghadas, D. One-dimensional deep learning inversion of electromagnetic induction data using convolutional neural network. *Geophys. J. Int.* **2020**, *222*, 247–259. [[CrossRef](#)]
48. Benhaddi, M.; Ouarzazi, J. Multivariate Time Series Forecasting with Dilated Residual Convolutional Neural Networks for Urban Air Quality Prediction. *Arab. J. Sci. Eng.* **2021**. [[CrossRef](#)]
49. Yang, J.; Wang, X.; Wang, R.; Wang, H. Combination of Convolutional Neural Networks and Recurrent Neural Networks for predicting soil properties using Vis-NIR spectroscopy. *Geoderma* **2020**, *380*, 114616. [[CrossRef](#)]