

*Original Research*

# Savitzky–Golay Denoising and Chla Concentration Inversion Based on ZY-1 02D Images: a Case Study of Nansi Lake, China

Fei Meng<sup>1</sup>, Jiawei Zhang<sup>2</sup>, Pingjie Fu<sup>1\*</sup>, Xinyue Yang<sup>1</sup>, Jianfei Feng<sup>1</sup>, Yu Cui<sup>1</sup>

<sup>1</sup>Shandong Jianzhu University School of Surveying and Geo-Informatics, Jinan 250101, China

<sup>2</sup>College of Geodesy and Geomatics, Shandong University of Science and Technology, Qingdao 266590, China

*Received: 17 April 2024*

*Accepted: 6 August 2024*

## Abstract

The importance of hyperspectral remote sensing technology in inland water quality monitoring research has achieved fruitful results. This research used the hyperspectral satellite images of ZY-1 02D and considered Nansi Lake, Shandong Province, China as the main research area. First, the Savitzky–Golay (SG) filtering method was used to denoise ZY-1 02D images. Meanwhile, combined with the XGBoost model, the denoised and original images were applied to retrieve the Chlorophyll-a (Chla) concentration in the water. We found that compared with the original image, the signal-to-noise ratio (SNR) of 7–5D and 9–5D filtered images has been improved in varying degrees. Based on the Chla concentration in the water, the three-band parameters of 7–5D, 9–5D, and the original (OD) image were extracted. The SNR of the characteristic bands obtained from the 7–5D image was significantly higher than other OD images, and it had the highest accuracy for Chla concentration inversion (coefficient of determination  $R^2=0.8737$ , root-mean-square error  $RMSE = 4.2259 \mu\text{g}\cdot\text{L}^{-1}$ ). This study innovatively utilized the SG filtering method to denoise ZY-1 02D hyperspectral satellite images and the XGBoost model applied to the images was established to invert the Chla concentration of water bodies, which realized large-scale visualization and high-precision monitoring of Chla concentration in the Nansi Lake, and provided a new idea for improving the accuracy of remote sensing methods for monitoring the water quality of inland water.

**Keywords:** hyperspectral remote sensing, ZY-1 02D, Savitzky-Golay filtering, chlorophyll-a concentration inversion, Nansi Lake, China

## Introduction

In recent years, substantial advancements have been achieved in the estimation of Chlorophyll-a (Chla)

concentration in inland water using optical remote sensing [1–5]. Satellite remote sensing provides significant benefits over traditional on-site sampling monitoring methods, especially for large-scale and long-term monitoring [6]. The utilization of satellite remote sensing images to monitor the spatial distribution of Chlorophyll-a (Chla) concentration in large inland lakes has become a research

\*e-mail: fupjrs@126.com

focus [7–9]. The current remote sensing inversion of Chla concentration in water mainly relies on medium-to-low resolution remote sensing data, and there is a need for improvement in its accuracy. As remote sensing technology advances, there is an increasing need for hyperspectral data products. Hyperspectral images, which possess higher spectral resolution and more bands compared to multispectral remote sensing images, offer greater potential for quantitative research on water quality parameters. Therefore, satellites equipped with hyperspectral sensors have gradually gained attention in various countries. In this study, we utilized the remote sensing images of the first civilian hyperspectral operational satellite (ZY-1 02D) of China to invert the Chla concentration, by leveraging its high spectral resolution and signal-to-noise ratio (SNR), we aimed to improve the accuracy of Chla estimation.

Hyperspectral remote sensing images provide valuable sources of spectral and spatial information for ground objects. However, the acquisition and transmission of such images are susceptible to various sources of noise, such as sensor interference, atmospheric conditions, natural light illumination, and challenges related to ground terrain and mixed pixel effects. The influence of noise on hyperspectral images manifested in two aspects: spatial and spectral-domain noises. In the field of remote sensing digital image processing, noise removal is an important task [10]. Noise in hyperspectral remote sensing images can have a significant detrimental effect on their quality, leading to negative impacts on downstream tasks, such as classification, recognition, and unmixing [11].

Therefore, hyperspectral image denoising holds academic significance and practical value [12]. Numerous scholars have conducted in-depth studies on denoising techniques for spectral signals and 2D images, resulting in remarkable achievements [13–17]. Despite the advantages of hyperspectral remote sensing images, they also present some unique challenges. One of these challenges is spectral integration, which refers to the high correlation between adjacent bands in the images. Traditional methods that do not consider this characteristic into account may not effectively denoise spectral signals or spatial images with high dimensions. As a result, the valuable spectral information contained in hyperspectral images is often underutilized due to the presence of noise [18]. In response, researchers have been actively exploring the development of denoising models and optimization methods that consider the specific characteristics of hyperspectral remote sensing images [19–21]. To effectively denoise hyperspectral images, Savitzky and Golay proposed the Savitzky-Golay (SG) filtering method [22], which uses a polynomial curve fitting method based on least squares to smooth the spectral curve. Madden [23] proposed a solution method of an arbitrary order SG filter based on a linear equation. The hyperspectral curve is affected by noise, resulting in jagged distributions. These jagged distributions represent spectral domain noise in the hyperspectral image, which significantly impacts its high spectral resolution [24, 25]. The hyperspectral curve processed by SG filtering can smooth some subtle sawtooth noise and not affect

the overall spectral characteristics of the entire curve to better reflect the absorption characteristics of the spectrum; thus, it is widely used in hyperspectral image denoising studies [26–30].

In this study, we employed the SG filtering method to denoise hyperspectral remote sensing images captured by the ZY-1 02D satellite. Utilizing the measured Chla concentration data in Nansihu Lake, we extracted spectra from the denoised images and derived characteristic band parameters using the three-band method. Subsequently, we developed an Extreme Gradient Boosting (XGBoost) [31] model for Chla concentration inversion. This approach enabled large-scale monitoring of Chla concentration in Nansihu Lake.

## Materials and Methods

### Study Area

Nansi Lake, a significant water body in Shandong Province, China, is made up of four interconnected lakes: Weishan Lake, Zhaoyang Lake, Dushan Lake, and Nanyang Lake, collectively referred to as Nansi Lake. This lake complex spans 1,266 Km<sup>2</sup>, the area measures 126 km in the north-south direction and 5–25 km in the east-west direction. Its total storage capacity amounts to 4.7 billion cubic meters, making it the second-largest freshwater lake in the Huaihe River Basin of China. Nansi Lake plays a vital role as a water source and storage lake for the eastern route of the South-to-North Water Diversion Project of China, underscoring the importance of maintaining its water quality. This study focused on Dushan Lake, the uppermost lake in the Nansi Lake system, as the primary research area. In particular, we conducted hyperspectral remote sensing inversion to estimate the concentration of Chla of Nansi Lake. Fig. 1 displays the spatial distribution of the remote sensing images and in-suit points in the study area.

### Satellite Images and Measured Data

#### *Measured Data*

On September 12, 2021, 38 sampling points were set up in the upper part of Nansi Lake (Dushan Lake) to sample the surface water of Dushan Lake. To save sampling time and increase the number of samples, the samples were transported back to the indoor dark place to determine the Chla concentration within 5 h. During sampling and transportation, the samples were always kept in a dark and low-temperature environment. The Chla concentration was measured using an American Arno brand ChloroTech 121A Chla analyzer, and GPS was used to locate the sampling points.

#### *ZY-1 02D Satellite Image*

The ZY-1 02D satellite was launched by China's Long March 4B carrier rocket at China Taiyuan Satellite Launch

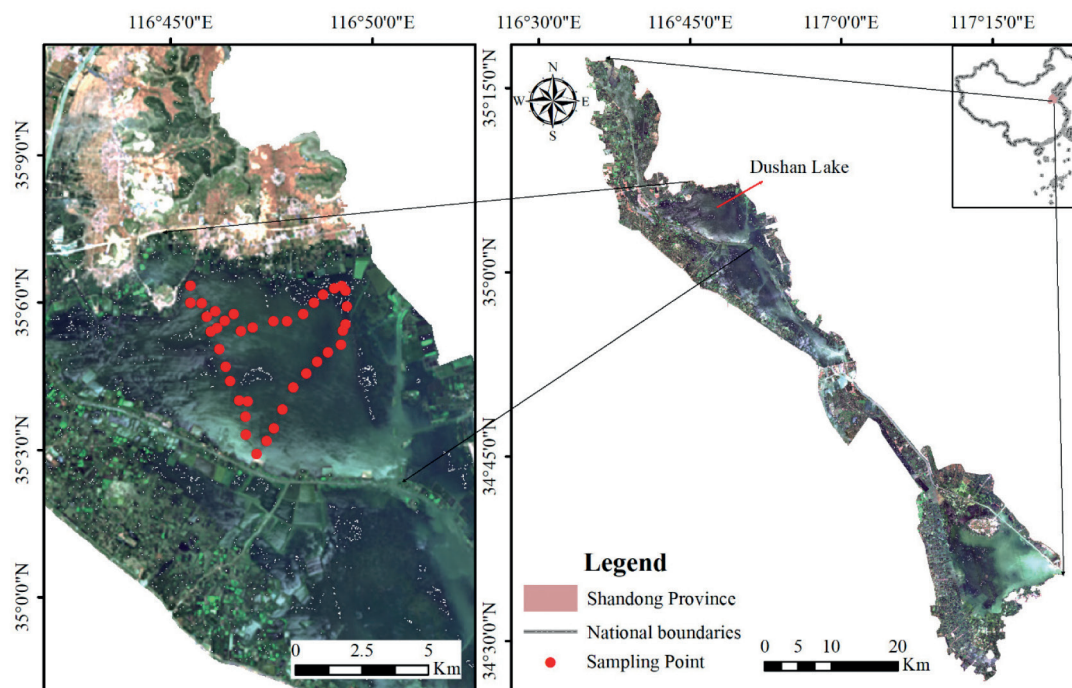


Fig. 1. Study area and sampling points.

Center at 11:26 on September 12, 2019. The satellite was developed by the China Academy of Space Technology and has two payloads: a visible near-infrared and a hyperspectral camera. Table 1 displays the main parameters of the ZY-1 02D satellite's hyperspectral camera payload. This satellite was located in the medium resolution, wide range observation, and quantitative remote sensing mission, which can provide rich spectral information on ground objects. Its hyperspectral camera subsystem was divided into 166 spectral channels in the spectral range of 400–2500 nm, with a visible near-infrared spectral resolution of 10 nm, a shortwave infrared spectral resolution of 20 nm, and the image covers an area of 60 km in width, with a spatial resolution of 30 m.

Considering the high spatial resolution, spectral resolution, and SNR, the comprehensive performance was at the leading level in the world. Its advantages could significantly improve the inversion effect and accuracy of water Chla. The image used in this study was the ZY-1 02D satellite image of the upper lake part of Nansi Lake, which was nearly synchronized with the field measurement time.

### Technical Route

In this study, we employed a combination of satellite images and measured Chla concentration to estimate the Chla concentration in water. Firstly, we performed preprocessing on the acquired ZY-1 02D hyperspectral image, followed by spectral SG filtering to eliminate spectral noise. We

compared the signal-to-noise ratio (SNR) of the image before and after noise reduction to assess the effectiveness of the denoising process. Based on the Chla measured concentration, the three-band method was applied to extract the characteristic bands of the original and denoised images. Subsequently, the XGBoost algorithm was utilized to model 70% of the dataset and then validate the entire dataset. A machine learning model was established to estimate Chla concentration. The accuracy of the image inversion before and after denoising was compared, and the technical route is shown in Fig. 2.

### Savitzky–Golay

In the plane coordinate system, a curve is used to fit a set of data, assuming that this curve is

$$y = a_0 + a_1x + a_2x^2 + a_3x^3 + a_4x^4$$

here, the curve equation is used to substitute the abscissa of each point, and the objective is to minimize the sum of squared differences between the obtained value and the corresponding ordinate of the point. The curve has the highest fit, so that all coefficients can be determined  $a_p$  ( $p=0, 1, 2, 3, 4$ ). The SG filtering method is a classic least-squares smoothing algorithm that uses a simplified least-squares fitting convolution to smooth the curve. Here,  $y_i$  represents the value of the  $i^{\text{th}}$  point before smoothing,  $\hat{y}_i$  represents the value after smoothing, the smoothing

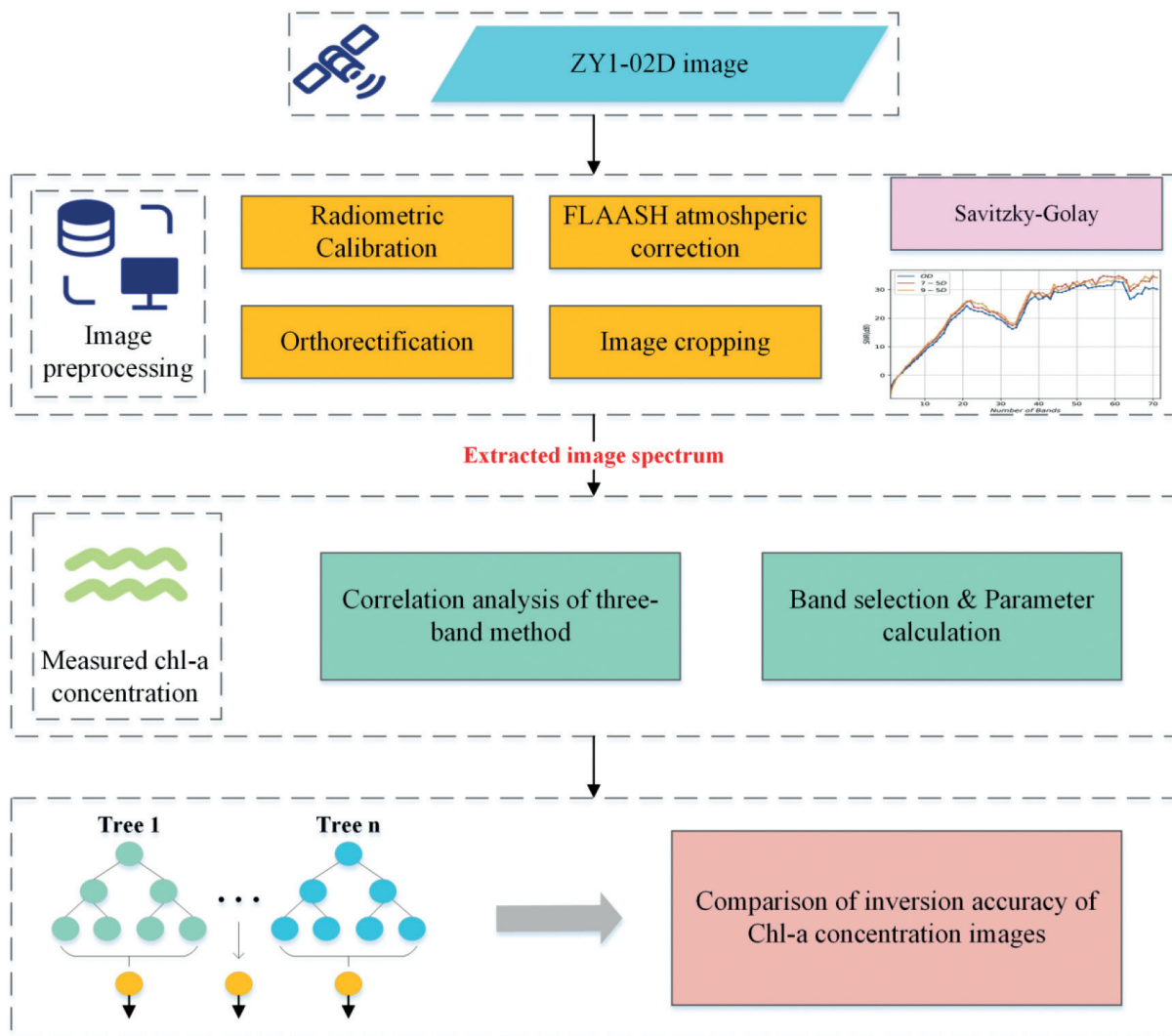


Fig. 2. Flowchart depicting the technical route.

window size is  $2m+1$ , and the fitting curve is a  $p$ -order polynomial:

$$\hat{y}_i = \sum_{k=0}^p a_k i^k \quad (1)$$

The error of this polynomial fit curve is:

$$\varepsilon = \sum_{j=i-m}^{i+m} (\hat{y}_i - y_i)^2 \quad (2)$$

To minimize the error, the partial derivatives of  $\varepsilon$  are taken, and the partial derivatives of  $\varepsilon$  are zero, and the following linear equations are obtained:

$$\begin{cases} \frac{\partial \varepsilon}{\partial a_0} = 2 \sum_{j=i-m}^{i+m} (a_0 + a_1 j + \dots + a_p j^p - y_j) = 0 \\ \frac{\partial \varepsilon}{\partial a_1} = 2 \sum_{j=i-m}^{i+m} j(a_0 + a_1 j + \dots + a_p j^p - y_j) = 0 \\ \dots \\ \frac{\partial \varepsilon}{\partial a_p} = 2 \sum_{j=i-m}^{i+m} j^p (a_0 + a_1 j + \dots + a_p j^p - y_j) = 0 \end{cases} \quad (3)$$

By solving the linear equation to obtain  $a_0, a_1, \dots, a_p$ , a fitted polynomial curve can be obtained.



### Three-Band Combined Parameter Construction of Chla Concentration

For the inversion of Chla concentration in water, three main methods are used at home and abroad: empirical, bio-optical, and semi-empirical/semi-analytical methods. Each of the three methods has advantages and limitations. Among the three methods, the three-band semi-analysis (TBS) method, based on the semi-analytical method proposed by Dall’Olmo [32] is the most widely used and the most common method for obtaining the Chla concentration [33, 34].

Gitelson et al. [35] suggested a method that utilizes a three-band sensitive spectral combination extraction technique to estimate the chlorophyll content in terrestrial plants, the form of which is:

$$C_{chl-a} \propto [R^{-1}(\lambda_1) - R^{-1}(\lambda_2)] \cdot R(\lambda_3) \quad (4)$$

Giorgio [36] selected 660–690 nm, 700–750 nm, and  $\geq 730$  nm as the value ranges of  $\lambda_1$ ,  $\lambda_2$ , and  $\lambda_3$ , respectively, and the parameter model based on a three-band method was developed for the inversion of Chla concentration in water. The model’s mathematical expression is as follows:

$$C_{chl-a} = A \cdot [R^{-1}(\lambda_1) - R^{-1}(\lambda_2)] \cdot R(\lambda_3) + B \quad (5)$$

where  $C_{chl-a}$  is the concentration of Chla;  $R^{-1}(\lambda_1)$  and  $R^{-1}(\lambda_2)$  are the reciprocals of the reflectance at wavelengths  $\lambda_1$  and  $\lambda_2$ , respectively;  $R(\lambda_3)$  is the reflectance at wavelength  $\lambda_3$ ;  $A$  and  $B$  are indeterminate coefficients.

Based on the three-band method, this study used the three-dimensional spectral index construction method to identify the position of the feature bands. This index construction method selects characteristic bands by collectively examining three bands through traversal. It traverses the combined correlations between all bands through iterative operations, which fully considers the synergy between bands and minimizes the influence of irrelevant wavelengths [37]. The correlation between the image spectrum and measured Chla concentration was determined using the three-dimensional spectral index construction method, which resulted in different characteristic band positions compared to those selected by the traditional semi-empirical/semi-analytical three-band method.

### XGBoost Model

According to the data characteristics of this study, we selected the XGBoost algorithm to perform multivariate regression and fit the dependent variables. XGBoost was designed by Chen Tianqi. When using CART as the base classifier, a regularization parameter is employed to regulate the model’s complexity [38]. By effectively reducing overfitting of the data, the XGBoost model improves the generalization ability of the data. The model’s mathematical expression is as follows:

$$\hat{y}_i = \varphi(x_i) = \sum_{k=1}^K f_k(x_i), f_k \in F \quad (6)$$

where  $\hat{y}_i$  is the predicted value of the XGBoost model of Chla concentration, the parameter value  $x_i$  represents the  $i^{\text{th}}$  three-band combination,  $K$  is the number of trees,  $f_k$  is the  $k^{\text{th}}$  tree model.

### Accuracy Verification

In this study, the Pearson correlation coefficient was used as the basis for selecting the correlation between the characteristic band parameters and measured Chla concentration. The Pearson correlation coefficient, also known as the Pearson product-moment correlation coefficient, was employed to assess the linear correlation between  $X$  and  $Y$  in the two datasets. Its value ranges between  $-1$  and  $1$ . The calculation formula is shown in (7).

$$\begin{aligned} \rho_{X,Y} &= \frac{\text{cov}(X,Y)}{\sigma_X \sigma_Y} = \\ &= \frac{E(XY) - E(X)E(Y)}{\sqrt{E(X^2) - (E(X))^2} \sqrt{E(Y^2) - (E(Y))^2}} = \\ &= \frac{n \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{\sqrt{n \sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2} \sqrt{n \sum_{i=1}^n y_i^2 - (\sum_{i=1}^n y_i)^2}} \end{aligned} \quad (7)$$

where  $E$  is the mathematical expectation;  $\text{cov}(X,Y)$  represents the covariance of  $X$  and  $Y$ ;  $\sigma_X$  and  $\sigma_Y$  are the standard deviations of  $X$  and  $Y$ , respectively. The closer the value of  $\rho_{X,Y}$  is to one, the higher the correlation between the two datasets.

In this study, the model fitting accuracy was evaluated using the coefficient of determination ( $R^2$ ), which compares the predicted values with the actual measured values. The  $R^2$  is obtained by dividing the explained variation by the total variation, which represents the quality of the regression line fit. Its value ranges between 0 and 1. The degree of accuracy of the regression equation is higher when  $R^2$  is closer to 1 and worse when it approaches 0. The accuracy of the model was evaluated in this study using the root-mean-square error (RMSE), which quantifies the extent of the difference between the predicted and actual values. A lower value of RMSE implies a higher level of accuracy, as there is less discrepancy between the predicted and measured values. Formula (8) was used to calculate  $R^2$ , while (9) was used to calculate RMSE.

$$R^2 = \frac{\sum (y_i - \bar{y})(y_j - \bar{y})}{\sqrt{\sum (y_i - \bar{y})^2 \sum (y_j - \bar{y})^2}} \quad (8)$$

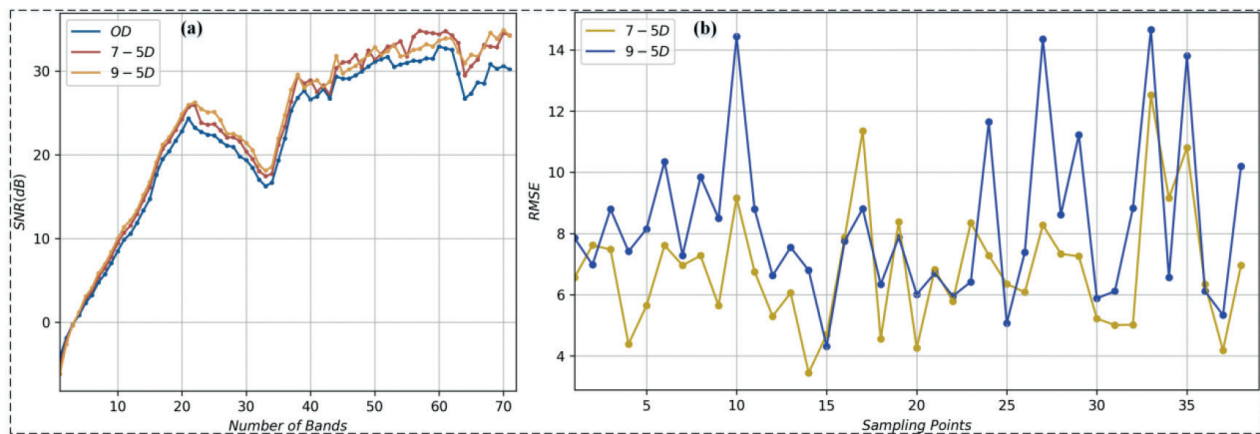


Fig. 3. SNR and RMSE. (a) Curves of the calculation results for different image SNR. (b) RMSE of the denoised image relative to the original image.

$$R_{MSE} = \sqrt{\frac{\sum_{i=1}^n (y_i - \bar{y}_j)^2}{n}} \quad (9)$$

where  $n$  represent the number of samples,  $y_i$  and  $y_j$  are measured and predicted values of the Chla concentration, respectively;  $\bar{y}_i$  and  $\bar{y}_j$  are the mean values of the measured and predicted Chla concentrations, respectively.

## Results and Discussion

### ZY-1 02D Image Denoising

The denoising of the preprocessed ZY-1 02D image using the SG filtering method is discussed. Filter window coefficients  $m$  and  $n$  can be selected over a relatively wide range. Generally, multiple iterations are required for different data to find the optimal kernel parameter range. The spectral curve used in this experiment was relatively complex. To preserve the original spectral curve characteristics to the greatest extent and achieve the best image denoising effect and inversion result, extensive iterative tests were conducted. Based on the experimental results, the SG filter window coefficients  $m = 7$ ,  $n = 5$  (7-5D), and  $m = 9$ ,  $n = 5$  (9-5D) were retained as the image results after noise reduction. Subsequently, they were used to compare the inversion accuracy of Chla concentration with the data obtained from the original image (OD).

To verify the denoising effect of the test image, the SNR was objectively measured. The higher the SNR, the better the denoising effect. The three image datasets were imported into the ENVI5.3 IDL console, and the SNR was calculated using the local standard deviation method based on edge elimination through code. The SNR results are shown in Fig. 3(a). The SNR of 7-5D and 9-5D were

slightly improved compared with OD, but the denoising effect varied in different bands.

To further examine the similarity between the spectrum of the denoised image and the original image, their RMSEs in each band of the sampling point spectrum were also calculated. The smaller the RMSE, the closer the denoised image spectrum is to the original image spectrum. The RMSE of the denoised image relative to that of the original image is shown in Fig. 3(b). The RMSE between 7-5D and OD was smaller compared with 9-5D, indicating that 7-5D had less image information loss. The image quality has a certain influence on the inversion of the Chla concentration in water. Then, the characteristic bands of the OD, 7-5D, and 9-5D data were extracted to build the optimal concentration estimation model.

Fig. 4 shows the reflectance curves of sampling points OD, 7-5D, and 9-5D. It can be observed that the spectral curve of the denoised image had a significant improvement in the fine sawtooth compared with the original image.

### Characteristic Band Extraction

Specifically, the extraction of feature bands is divided into four steps: (1) Extract spectral data from the original image and the denoised image (7-5D and 9-5D) according to the actual measurement point positions. (2) Calculate the correlation coefficient between a single band and the measured Chla concentration. (3) Follow the method in Technical Route Section, iteratively cycle through all possible band combinations, and calculate the correlation coefficient between all combinations and the measured Chla concentration. (4) Sort all correlation coefficients and select the five largest characteristic bands as input to the model. The correlation coefficient statistics of the first five groups of band combinations between the three groups of spectral data and the Chla concentration are shown in Table 1. Simultaneously, the SNR of the sampling points

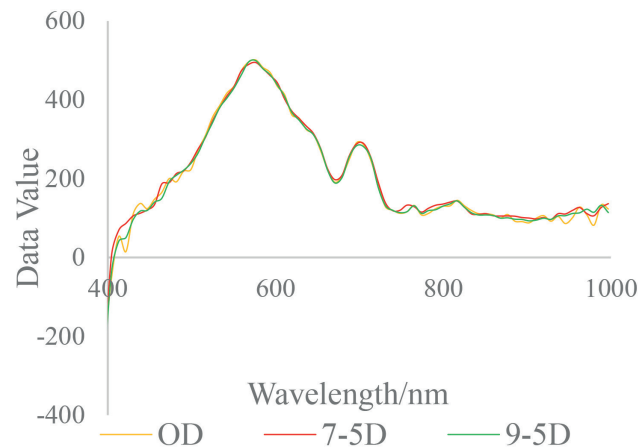


Fig. 4. The reflectivity curve of a sampling point.

Table 1. Three-band combination and band SNR statistics.

	OD			7-5D			9-5D		
	$\lambda_1/\text{nm}$ SNR/dB	$\lambda_2/\text{nm}$ SNR/dB	$\lambda_3/\text{nm}$ SNR/dB	$\lambda_1/\text{nm}$ SNR/dB	$\lambda_2/\text{nm}$ SNR/dB	$\lambda_3/\text{nm}$ SNR/dB	$\lambda_1/\text{nm}$ SNR/dB	$\lambda_2/\text{nm}$ SNR/dB	$\lambda_3/\text{nm}$ SNR/dB
1	713.7949 26.797	533.5523 19.466	559.2343 22.805	713.7949 29.311	533.5523 20.706	559.2343 24.195	696.5913 24.709	559.2343 24.838	645.2094 21.404
2	713.7949 26.797	533.5523 19.466	602.2487 22.307	713.7949 29.311	533.5523 20.706	602.2487 23.653	713.7949 29.535	533.5523 21.192	559.2343 24.838
3	713.7949 26.797	533.5523 19.466	567.8867 24.326	713.7949 29.311	533.5523 20.706	550.5819 22.945	713.7949 29.535	533.5523 21.192	567.8867 25.926
4	713.7949 26.797	533.5523 19.466	576.4771 23.211	713.7949 29.311	533.5523 20.706	567.8867 25.700	696.5913 24.709	550.5819 23.281	619.4298 22.476
5	713.7949 26.797	533.5523 19.466	585.0676 22.710	713.7949 29.311	533.5523 20.706	585.0676 23.796	696.5913 24.709	559.2343 24.838	619.4298 22.476

at the corresponding characteristic wavelengths in OD, 7–5D, and 9–5D were extracted. It can be seen from Table 1 that the SNR of the characteristic bands extracted at 9–Dd and 7–5D was slightly improved.

### Hyperspectral Image Retrieval of Chla Concentration in Nansi Lake

The method used in this study was the modified normalized difference water index (MNDWI) proposed by Xu [39]. Water bodies were extracted from OD, 7–5D, and 9–5D images, respectively. Based on the combination of the five bands with the best correlation, the three-band formula was used to calculate the inversion parameters of the three sets of images. Using the XGBoost model, 70% of the sampling points were selected for modeling, and the inversion accuracy of the Chla concentration was verified at all sampling points. Fig. 5 shows the image inversion results. The high and low values

of Chla concentration in the original image were relatively concentrated, and the high-value areas accounted for a large proportion. With the adjustment of the image noise reduction parameters, the distribution of high and low values of Chla became more dispersed. Fig. 6 shows a fitting diagram of the image inversion accuracy. The inversion accuracy of the Chla concentration in the image after noise reduction was significantly improved compared to the original image. Among them, using 7–5D image data to invert Chla concentration had the highest accuracy, with an  $R^2$  of 0.8737 and RMSE of  $7.45 \mu\text{g}\cdot\text{L}^{-1}$ . In addition, the Chla concentration in Dushan Lake is generally high. During the field investigation, it was found that the lake had undergone multiple large-scale drainage just before the satellite transit and actual measurement. The direction was from south to north, causing the water flow to concentrate towards Dushan Lake, causing pollution. In addition, there are many rice and terrace fields around Dushan Lake, and the soaking and rotting of straw after returning to the fields, pesticide, and fertilizer residues, etc., will also

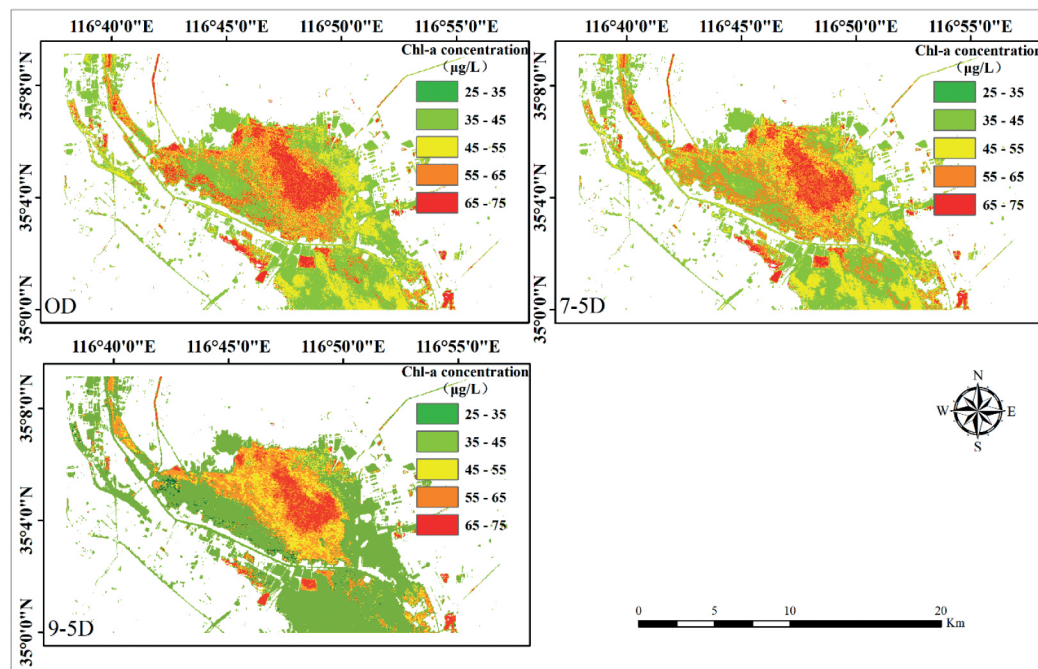


Fig. 5. Chla concentration image inversion results.

affect the water quality of the lake, so the concentration of Chla in Dushan Lake will be higher.

#### Selection of Feature Bands in the Process of Chlorophyll-a Inversion

Regarding the selection of characteristic bands, in the study of Liu et al [40], the authors refer to the previous research experience and, through the traditional band positioning method, determined the central wavelengths of  $\lambda_1$ ,  $\lambda_2$ , and  $\lambda_3$  from the range of band selection of the three-band semi-empirical model to be 671 nm, 705 nm, and 731 nm respectively. This is different from the three-band parameter model constructed by the three-dimensional spectral index method based on the three-band method model formula in this study to invert the concentration of Chla. The selected three-band positions were quite different because the three-dimensional spectral index method adopted in this study did not consider the selection range of each band in the three-band semi-empirical model, but the overall loop iterative traversal of all the bands to achieve the positioning of the band positions.

#### Necessity of Hyperspectral Denoising and Selection of Inversion Model

The inversion accuracy of the original image was low, which may be caused by the deviation between the image

acquisition time and sampling time. In addition, the direct error between the measured point and the pixel position of the extracted image spectrum, radiometric calibration of the image, atmospheric correction, and other errors affect the inversion results of the image. Therefore, in the study of water Chla inversion using hyperspectral images, it is necessary to optimize further the image preprocessing method to reduce the influence of various errors on inversion accuracy. In terms of the inversion model, XGBoost model was selected to invert the Chla concentration of images to alleviate the overfitting phenomenon caused by the machine learning algorithm due to the small number of sampling points used in this study. This is because previous studies have found that the XGBoost model can be trained on sparse data using the sparse-aware split-finding method. The XGBoost model will always have better results in water quality inversion, which is similar to the results of previous studies, the application of this method holds significant importance in remote sensing image processing, particularly in addressing the challenge of low sampling rates in field data.

#### Conclusions

In this study, a hyperspectral image of China's domestic satellite (ZY-1 02D) was used as the research object, the Savitzky-Golay method was used to denoise the ZY-1 02D image, and the denoised image was used to retrieve the Chla concentration of Nansi Lake. The conclusions are as follows:



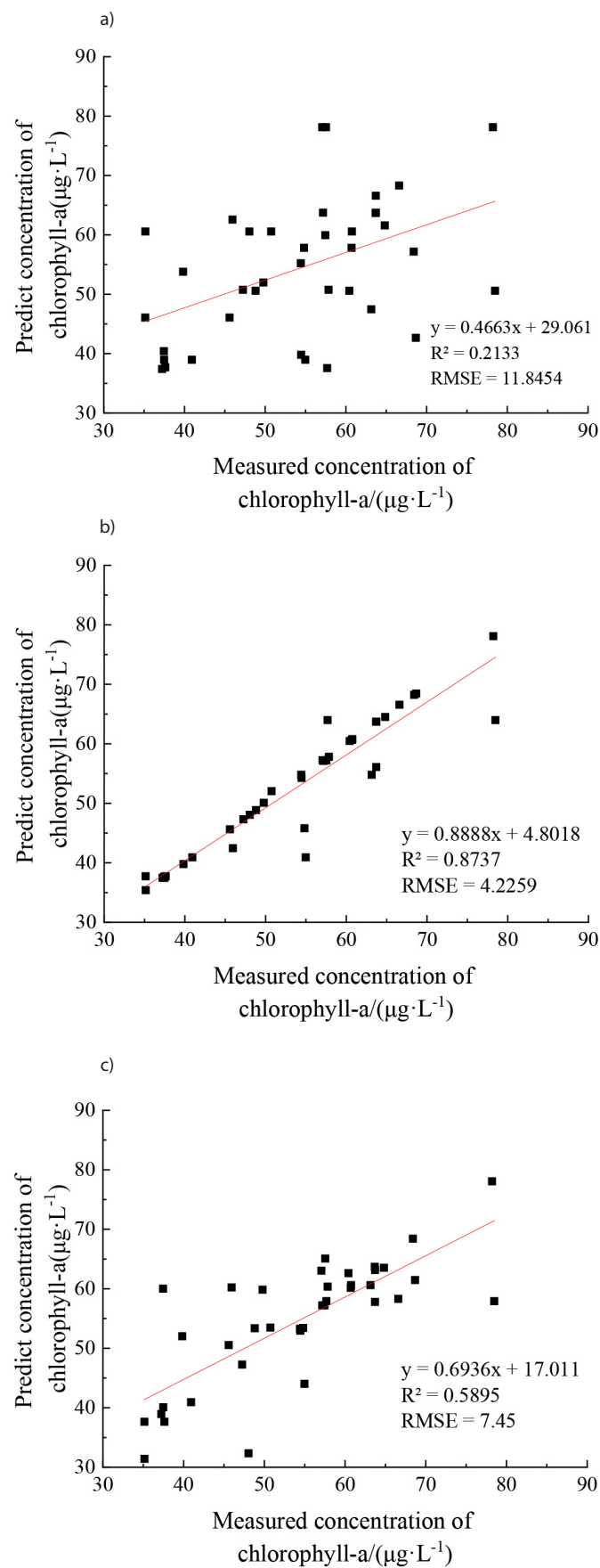


Fig. 6. Fitting diagram of the inversion accuracy of Chla concentration in different image data (a) OD, (b) 7-5D, (c) 9-5D.

1. The SNR of the 7–5D and 9–5D denoised images was significantly improved. Therefore, SG filtering can effectively denoise the hyperspectral ZY-1 02D image, and different parameter settings can produce varying image-denoising effects. The ZY-1 02D satellite exhibits excellent performance, and the hyperspectral images acquired by it have undergone noise reduction processing, which can provide better and more convenient monitoring methods for water quality remote sensing monitoring.

2. In the inversion of Chla concentration in Nansihu Lake performed using the XGBoost machine learning model, the OD image inversion accuracy was the lowest, with an  $R^2$  of 0.2133 and RMSE of  $11.8454 \mu\text{g}\cdot\text{L}^{-1}$ ; the inversion accuracy of 9–5D images ranked second, with an  $R^2$  of 0.5895 and RMSE of  $7.45 \mu\text{g}\cdot\text{L}^{-1}$ , and the accuracy of the 7–5D image was the best with an  $R^2$  of 0.8737 and RMSE of  $4.2259 \mu\text{g}\cdot\text{L}^{-1}$ . Experiments showed that noise reduction processing on hyperspectral images (with appropriate parameter adjustment) could significantly improve the inversion accuracy of Chla concentration.

3. By performing Savitzky–Golay denoising on hyperspectral images, we effectively improved the accuracy of retrieval of Chla concentration in water. This result is of great significance for water environment monitoring and management. The concentration of Chla in the northeastern area of Dushan Lake is relatively high. Environmental management of surrounding villages and other areas with human activities should be strengthened to prevent the deterioration of water quality.

### Acknowledgments

This research was supported in part by the National Natural Science Foundation of China under Grant 42101388, and Shandong Natural Science Foundation (No. ZR2022MD070), Shandong Top Talent Special Foundation (No. 0031504), and Jinan City and University Integration Development Project (JNSX2023065). Thanks to the editors and anonymous reviewers for their detailed and constructive comments. Thanks to all anonymous reviewers.

### Conflict of Interest

The authors declare no conflict of interest.

### References

1. NEIL C., SPYRAKOS 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 Sensing of Environment*. **229**, 159, **2019**.
2. FENG L., DAI Y., HOU X., XU Y., LIU J., ZHENG C. Concerns about phytoplankton bloom trends in global lakes. *Nature*. **590** (7846), E35, **2021**.
3. CAO Z., MA R., MELACK J.M., DUAN H., LIU M., KUTSER T., XUE K., SHEN M., QI T., YUAN H. Landsat

- observations of chlorophyll-a variations in Lake Taihu from 1984 to 2019. *International Journal of Applied Earth Observation and Geoinformation*. **106**, **2022**.
4. GUAN Q., FENG L., HOU X.J., 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 Sensing of Environment*. **246**, 111890, **2020**.
5. HE Y., WANG X., XU F. How reliable is chlorophyll-a as algae proxy in lake environments? New insights from the perspective of n-alkanes. *Science of the Total Environment*. **836**, 155700, **2022**.
6. SHI K., ZHANG Y., XU H., ZHU G., QIN B., HUANG C., LIU X., ZHOU Y., LV H. Long-Term Satellite Observations of Microcystin Concentrations in Lake Taihu during Cyanobacterial Bloom Periods. *Environmental Science & Technology*. **49** (11), 6448, **2015**.
7. LI J., GAO M., FENG L., ZHAO H., SHEN Q., ZHANG F., WANG S., ZHANG B. Estimation of Chlorophyll-a Concentrations in a Highly Turbid Eutrophic Lake Using a Classification-Based MODIS Land-Band Algorithm. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*. **12** (10), 3769, **2019**.
8. NIU C., TAN K., JIA X., WANG X. Deep learning based regression for optically inactive inland water quality parameter estimation using airborne hyperspectral imagery. *Environ Pollut*. **286**, 117534, **2021**.
9. PAHLEVAN N., SMITH B., SCHALLES J., BINDING C., CAO Z., MA R., ALIKAS K., KANGRO K., GURLIN D., HÀ N., MATSUSHITA B., MOSES W., GREB S., LEHMANN M. K., ONDRUSEK M., OPPELT N., STUMPF R. Seamless retrievals of chlorophyll-a from Sentinel-2 (MSI) and Sentinel-3 (OLCI) in inland and coastal waters: A machine-learning approach. *Remote Sensing of Environment*. **240**, **2020**.
10. LU M., HU L., YUE T., CHEN Z., CHEN B., LU X., XU B. Penalized Linear Discriminant Analysis of Hyperspectral Imagery for Noise Removal. *IEEE Geoscience and Remote Sensing Letters*. **14** (3), 359, **2017**.
11. RASTI B., GHAMISI P., BENEDIKTSSON J.A. Hyperspectral Mixed Gaussian and Sparse Noise Reduction. *IEEE Geoscience and Remote Sensing Letters*. **17** (3), 474, **2020**.
12. HUANG S., LIU Z., WANG Y., WANG R. Wide-stripe noise removal method of hyperspectral image based on fusion of wavelet transform and local interpolation. *Optical Review*. **24** (2), 177, **2017**.
13. RASTI B., SCHEUNDERS P., GHAMISI P., LICCIARDI G., CHANUSSOT J. Noise Reduction in Hyperspectral Imagery: Overview and Application. *Remote Sensing*. **10** (3), 482, **2018**.
14. ZENG Z., HUANG T.-Z., CHEN Y., ZHAO X.-L. Nonlocal Block-Term Decomposition for Hyperspectral Image Mixed Noise Removal. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*. **14**, 5406, **2021**.
15. BOURENNANE S., FOSSATI C., LIN T. Noise Removal Based on Tensor Modelling for Hyperspectral Image Classification. *Remote Sensing*. **10** (9), 1330, **2018**.
16. JI L., WANG L., GENG X. An Automatic Bad Band Pre-Removal Method for Hyperspectral Imagery. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*. **12** (12), 4985, **2019**.
17. MA T.-H., XU Z., MENG D. Remote Sensing Image Denoising via Low-Rank Tensor Approximation and Robust Noise Modeling. *Remote Sensing*. **12** (8), **2020**.

18. DAO P.D., MANTRIPRAGADA K., HE Y., QURESHI F.Z. Improving hyperspectral image segmentation by applying inverse noise weighting and outlier removal for optimal scale selection. *ISPRS Journal of Photogrammetry and Remote Sensing*. **171**, 348, **2021**.
19. MAFFEI A., HAUT J. M., PAOLETTI M. E., PLAZA J., BRUZZONE L., PLAZA A. A Single Model CNN for Hyperspectral Image Denoising. *IEEE Transactions on Geoscience and Remote Sensing*. **58** (4), 2516, **2020**.
20. WANG M., WANG Q., CHANUSSOT J., LI D. Hyperspectral Image Mixed Noise Removal Based on Multidirectional Low-Rank Modeling and Spatial–Spectral Total Variation. *IEEE Transactions on Geoscience and Remote Sensing*. **59** (1), 488, **2021**.
21. GOLILARZ N.A., GAO H., PIRASTEH S., YAZDI M., ZHOU J., FU Y. Satellite Multispectral and Hyperspectral Image De-Noiseing with Enhanced Adaptive Generalized Gaussian Distribution Threshold in the Wavelet Domain. *Remote Sensing*. **13** (1), 101, **2020**.
22. SCHETTINO B.M., DUQUE C.A., SILVEIRA P.M. Current-Transformer Saturation Detection Using Savitzky–Golay Filter. *IEEE Transactions on Power Delivery*. **31** (3), 1400, **2016**.
23. HH M. Comments on the Savitzky-Golay Convolution Method for Least-Squares Fit Smoothing and Differentiation of Digital Data. *Analytical Chemistry*. **50** (9), **1978**.
24. MIGLANI A., RAY S.S., VASHISHTA D.P., PARIHAR J.S. Comparison of Two Data Smoothing Techniques for Vegetation Spectra Derived From EO-1 Hyperion. *Journal of the Indian Society of Remote Sensing*. **39** (4), 443, **2011**.
25. PANG L., XIAO J., MA J., YAN L. Hyperspectral imaging technology to detect the vigor of thermal-damaged *Quercus variabilis* seeds. *Journal of Forestry Research*. **32** (2), 461, **2020**.
26. YANG R., TIAN H., KAN J. Classification of Sugar Beets Based on Hyperspectral and Extreme Learning Machine Methods. *Applied Engineering in Agriculture*. **34** (6), 891, **2018**.
27. ZHU S., CHAO M., ZHANG J., XU X., SONG P., ZHANG J., HUANG Z. Identification of Soybean Seed Varieties Based on Hyperspectral Imaging Technology. *Sensors (Basel)*. **19** (23), **2019**.
28. ZHOU Q., HUANG W., FAN S., ZHAO F., LIANG D., TIAN X. Non-destructive discrimination of the variety of sweet maize seeds based on hyperspectral image coupled with wavelength selection algorithm. *Infrared Physics & Technology*. **109** (10), 103418, **2020**.
29. YANG G., WANG Q., LIU C., WANG X., FAN S., HUANG W. Rapid and visual detection of the main chemical compositions in maize seeds based on Raman hyperspectral imaging. *Spectrochim Acta A Mol Biomol Spectrosc*. **200**, 186, **2018**.
30. XIN Z., JUN S., YAN T., QUANSHENG C., XIAOHONG W., YINGYING H. A deep learning based regression method on hyperspectral data for rapid prediction of cadmium residue in lettuce leaves. *Chemometrics and Intelligent Laboratory Systems*. **200**, **2020**.
31. CHEN T., GUESTIN C. XGBoost: A Scalable Tree Boosting System. *Proceedings of the 22<sup>nd</sup> ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, San Francisco, California. **2016**.
32. DALL'OLMO G., GITELSON A.A., RUNDQUIST D.C. Towards a unified approach for remote estimation of chlorophyll-a in both terrestrial vegetation and turbid productive waters. *Geophysical Research Letters*. **30** (18), **2003**.
33. CUI Y., MENG F., FU P., YANG X., ZHANG Y., LIU P. Application of hyperspectral analysis of chlorophyll a concentration inversion in Nansi Lake. *Ecological Informatics*. **64**, 101360, **2021**.
34. ZHANG J., FU P., MENG F., YANG X., XU J., CUI Y. Estimation algorithm for chlorophyll-a concentrations in water from hyperspectral images based on feature derivation and ensemble learning. *Ecological Informatics*. **71**, 101783, **2022**.
35. GITELSON A.A., GRITZ Y., MERZLYAK M.N. Relationships between leaf chlorophyll content and spectral reflectance and algorithms for non-destructive chlorophyll assessment in higher plant leaves. *The Journal of Plant Physiology*. **160** (3), 271, **2003**.
36. DALL'OLMO G., GITELSON A.A. Effect of bio-optical parameter variability on the remote estimation of chlorophyll-a concentration in turbid productive waters: experimental results. *Applied Optics*. **44** (3), 412, **2005**.
37. FU P., YANG K., MENG F., ZHANG W., CUI Y., FENG F., YAO G. A new three-band spectral and metal element index for estimating soil arsenic content around the mining area. *Process Safety and Environmental Protection*. **157**, 27, **2022**.
38. ZHANG J., MENG F., FU P., JING T., XU J., YANG X. Tracking changes in chlorophyll-a concentration and turbidity in Nansi Lake using Sentinel-2 imagery: A novel machine learning approach. *Ecological Informatics*. **81**, 102597, **2024**.
39. XU H. A Study of Information Extraction of Water Body with the Modified Normalized Difference Water Index(MNDWI). *Journal of Remote Sensing*. **9**, 589, **2005**.
40. LIU Y., LI J., XIAO C., ZHANG F., WANG S. Inland water chlorophyll-a retrieval based on ZY-1 02D satellite hyperspectral observations. *National Remote Sensing Bulletin*. **26** (1), 168, **2022**. [In Chinese]