

## Estimation of total phosphorus concentration using a water classification method in inland water



Chenggong Du<sup>a,b</sup>, Qiao Wang<sup>c</sup>, Yunmei Li<sup>a,b,\*</sup>, Heng Lyu<sup>a,b</sup>, Li Zhu<sup>c</sup>, Zhubin Zheng<sup>a,b</sup>, Shuang Wen<sup>a,b</sup>, Ge Liu<sup>d</sup>, Yulong Guo<sup>e</sup>

<sup>a</sup> Key Laboratory of Virtual Geographic Environment, Ministry of Education, College of Geographic Science, Nanjing Normal University, Nanjing, 210023, China

<sup>b</sup> Jiangsu Center for Collaborative Innovation in Geographical Information Resource Development and Application, Nanjing, 210023, China

<sup>c</sup> Satellite Environment Application Center, Ministry of Environmental Protection, Beijing, 100029, China

<sup>d</sup> Northeast Institute of Geography and Agricultural Ecology, Chinese Academy of Sciences, Changchun, 130000, China

<sup>e</sup> College of Resources and Environmental Science, Henan Agricultural University, Zhengzhou, 450002, China

### ARTICLE INFO

#### Keywords:

Total phosphorus  
Remote sensing  
Classification  
Inversion models  
Sentinel-3

### ABSTRACT

Total phosphorus (TP) is an important factor of eutrophication. The accurate monitoring of TP concentration is conducive to the management of water environments. To improve the accuracy of TP remote sensing inversion, this paper improved and studied the following three aspects. First, a new optical classification method was constructed to classify water into two different types. Compared with other classification methods, the proposed method can be more effective for improving the accuracy of TP inversion. Second, data regression analysis and fitting (DRF), back propagation neural network (BP) and random forest (RF) were used to build TP inversion models based on the different water type. The validation results indicated that the DRF performed best, with the highest precision for all water types, and was suitable for TP inversion. Finally, the performance of the classification and inversion algorithms on Sentinel-3 data was evaluated and one image was acquired for mapping the TP concentration in Taihu Lake. All in all, this study improves the accuracy of TP inversion by combining the new classification algorithm with DRF. This is significant for the development of class-based water quality parameter inversion algorithms for water color remote sensing, and this approach can be applied in the effective management and control of the eutrophication of lake environments.

### 1. Introduction

Eutrophication is a serious phenomenon among environmental pollution, and it severely threatens the sustainable development of society and the survival of mankind (Le et al., 2010; Qin et al., 2012). With the rapid development of the industrial and agricultural economy, a large amount of industrial and agricultural wastewater and municipal sewage is discharged into lake water, resulting in lake water eutrophication (Liu and Qiu, 2007; Savage et al., 2010). Phosphorus is an essential element of algae growth and an important factor in eutrophication. The root cause of eutrophication is the increase in nutrients, which is generally considered to be mainly phosphorus, followed by nitrogen (Vollenweider et al., 1980). According to the OECD (World Economic Cooperation and Development Organization) study, 80% of water eutrophication is attributable to phosphorus, 10% of the eutrophication of the water body is directly related to nitrogen and phosphorus, and the remaining 10% of the water body eutrophication is due to nitrogen and

other factors. Therefore, the accurate monitoring of TP in lake water is an important link in the effective management and control of eutrophication in the lake environment. Nevertheless, conventional water quality monitoring has failed to characterize TP dynamics because of the limitations in spatial sampling. Remote sensing has the characteristics of timeliness and wide range and has been widely used in various types of lake pollution research (Akbar et al., 2010; Guo et al., 2016; Shang et al., 2016; Zheng et al., 2015; Zhu et al., 2013); thus, it might be the ideal tool for inverse TP content.

There are several research studies concerning the inversion model for TP, and they can be divided into two categories: indirect and direct inversion models. Indirect models generally need to find the relationship between TP and optically active substances (chlorophyll-a (Chl-a), total suspended matter (TSM) and color dissolved organic matter (CDOM)), and then invert TP concentration based on the inversion of optically active substances (Song et al., 2014). Thus, the accuracy of this method seriously depends on the inversion of the optically active

\* Corresponding author at: Key Laboratory of Virtual Geographic Environment, Ministry of Education, College of Geographic Science, Nanjing Normal University, Nanjing, 210023, China.

E-mail address: [648666077@qq.com](mailto:648666077@qq.com) (Y. Li).

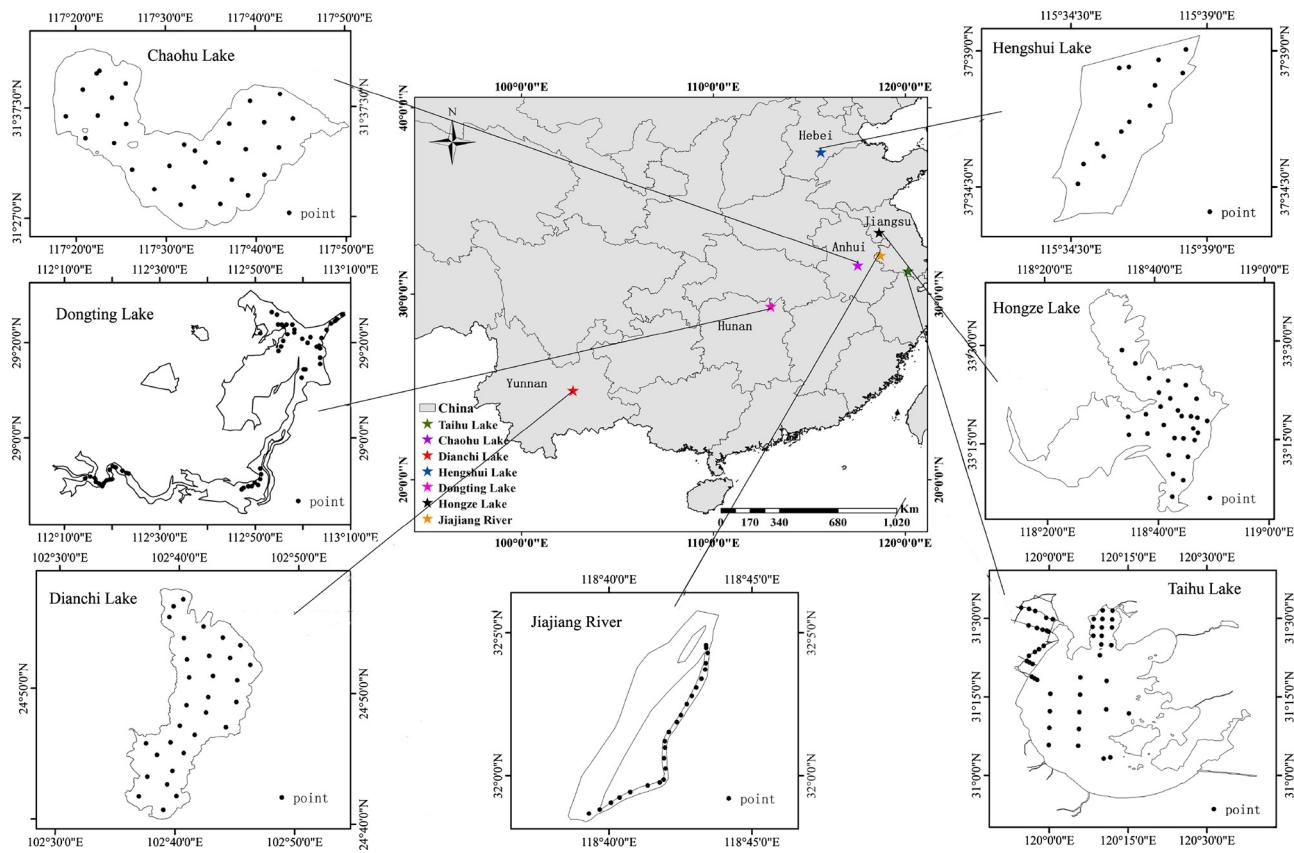


Fig. 1. Spatial distribution of study area.

substances and the relationship between the two parameters; compared with indirect models, the direct model constructs a TP concentration estimation model from remote sensing reflectance ( $R_{rs}$ ) based on the relationship between  $R_{rs}$  and TP directly. But for constructing the direct inversion models, a variety of mathematical methods were tested and used without considering the mechanism. This would greatly affect the applicability of the model because it is limited by study area and data (Chang and Xuan, 2013; Gao et al., 2015; Isenstein and Park, 2014; Sun et al., 2014b; Wu et al., 2010). To sum up, both the indirect and the direct inversion models were developed in their respective study areas based on the strong correlation between TP and Chl-a, TSM or  $R_{rs}$ . Therefore, the applicability of the models need to be validated in other water bodies. From the premier study, oceanic waters and inland clean waters are the major inhabitants, and only fewer studies were conducted on inland eutrophication lakes. Since the complex optical character, the inversion algorithm of TP in inland lakes is not mature enough, while the deteriorating water environment is in urgent need of a universal application of a TP inversion algorithm. Accordingly, this paper attempts to establish a generalized TP inversion algorithm for different water bodies.

The construction of a universal phosphor-inversion model needs to consider different water body composition and optical characteristics. Previous studies have shown that the sources and the influencing factors of TP are different in different types of water bodies. When the non-algal particles in the water is dominant, TP is mainly attached to ISM and has a high correlation with ISM when the main source is land-based input. When the phytoplankton is the main constituent, TP is highly correlated with Chl-a and is affected by the absorption and degradation of algae (Domagalski et al., 2006; Pacciaroni and Crisp, 2007; Shi et al., 2017). Correspondingly, the  $R_{rs}$  of a water body is influenced by the water components. Therefore, it is necessary to classify the water bodies and then construct TP remote sensing estimation model based on the study of different water body characteristics. Some water

classification methods based on  $R_{rs}$  were proposed to improve the accuracy of water color parameter inversion (Lyu et al., 2015; Shi et al., 2013; Sun et al., 2014a). Common classification methods mainly include cluster analysis, the maximum chlorophyll index (MCI) and the groove depth or baseline height in the  $R_{rs}$ . However, these classification methods are mainly aimed at chlorophyll or TSM inversion, not considering the characteristics of TP in different bodies of water. Then, it is necessary to construct a specific method to construct TP remote sensing models for various water properties and improve the accuracy of TP content inversion based on the different water type.

Among water color sensors, the Sentinel 3 is the third satellite in the six satellites of the Copenhagen mission, carrying the most advanced sensor, which is the Ocean and Land Color Imager (OLCI) instrument. It will conduct frequent, near real-time global monitoring of the ocean and its data will be freely available globally. And it provides new data sources for water color remote sensing.

Consequently, in this study, the objectives are to (1) propose a simple and effective remote sensing-based classification method for TP concentration inversion by classifying water bodies into non-algae particles or phytoplankton dominated waters; (2) confirm the most appropriate TP retrieval algorithm for estimating TP concentrations across different trophic states in inland waters and (3) verify the applicability of the algorithms on Sentinel-3 data.

## 2. Study area and in situ data

### 2.1. Study area

The data used in this study were collected in Taihu Lake, Hongze Lake, Dongting Lake, Dianchi Lake, Chaohu Lake, the Jiajiang River and Hengshui Lake from 2009 to 2016. These lakes contain eutrophic inland freshwater lakes, then the algorithms constructed from the data would be more versatile. Among these locations, Taihu Lake, Hongze Lake,

**Table 1**  
Sampling points information.

Study area	Numbers	Sampling time
Taihu Lake	119	August 9, 2010 (23); August 1(13) and 3(17), 2013; October 23, 2014 (35); July 22, 2016 (31)
Chaohu Lake	16	June 14(7) and 15(9), 2009
Dianchi Lake	18	September 19(9) and 20(9), 2009
Hengshui Lake	12	August 16, 2016
Dongting Lake	16	August 9, 2013
Hongze Lake	48	June 3(11), 4(15) and 6(22), 2015
Jiajiang River	19	October 15, 2014

Dongting Lake and Chaohu Lake belong to China's five major freshwater lakes. Taihu Lake, located in Jiangsu province, is a typical eutrophication water body, and its water composition is the most complex among the above water bodies (Du et al., 2016; Ma et al., 2006; Shi et al., 2015). Dianchi Lake, Chaohu Lake and Hengshui Lake are phytoplankton-dominated water bodies located in Yunnan province, Anhui province and Hebei province, respectively (Chen et al., 2013; Huang et al., 2014). In contrast, Hongze Lake, Dongting Lake and the Jiajiang River, which have lower Chl-a concentrations, are non-algae particles dominated water bodies (Zheng et al., 2015). Hongze Lake and the Jiajing River are both located in Jiangsu province, while Dongting Lake is located in Hunan province. The spatial distribution of these areas is shown in Fig. 1.

## 2.2. In situ data

The number of in situ data is 248 in total, there are 119 points in Taihu Lake, 16 in Chaohu Lake, 18 in Dianchi Lake, 12 in Hengshui Lake, 16 in Hongze Lake, 48 in Dongting Lake and 19 in Jiajiang River. And the specific point's information and sampling time is shown in Table 1.

Each point was measured for reflectance by an Analytical Spectral Device called a Field Spec spectroradiometer (350–1050 nm range, with a 1.5-nm sampling interval and an FWHM of ~3 nm).  $R_{rs}(sr^{-1})$  was calculated via the following equation.

$$R_{rs}(\lambda) = (L_t - r^* L_{sky}) / (L_p * \pi / \rho_p) \quad (1)$$

where  $L_t$  is the total radiance of the water surface;  $r$  is skylight reflectance at the air-water surface;  $L_{sky}$  is the measured radiance from the sky;  $L_p$  is the measured radiance of the gray reference panel; and  $\rho_p$  is the reflectance of the gray diffuse panel. The data in the range of 400–900 nm, which is generally used for water color remote sensing, were used to calibrate and validate the performance of the TP concentration estimation algorithms.

Water samples were also collected from the surface and taken back to the laboratory in a refrigerated environment for water quality (TP, Chl-a, Inorganic suspended matter (ISM), organic suspended matter (OSM) and TSM), and water component absorption coefficient ( $a_p$  and  $a_{ph}$ ) analysed in less than 12 h. Water samples were filtered on Whatman GF/F glass fiber filters. TP was measured using an ultraviolet-visible spectrophotometer (Shimadzu, UV-3600) (Huang et al., 2015). Chl-a, TSM, ISM and OSM were determined spectrophotometrically using the method described by Lorenzen and Chen et al (Chen, 2006; Lorenzen, 1967).

## 3. Model development

### 3.1. Framework for TP estimation algorithms

Seen from Fig. 2, the overall framework of TP inversion consists of two steps, which are constructions of optical classification method and TP estimation model. The datasets used for classification should ideally include a complete representation of all water types. In the first step,

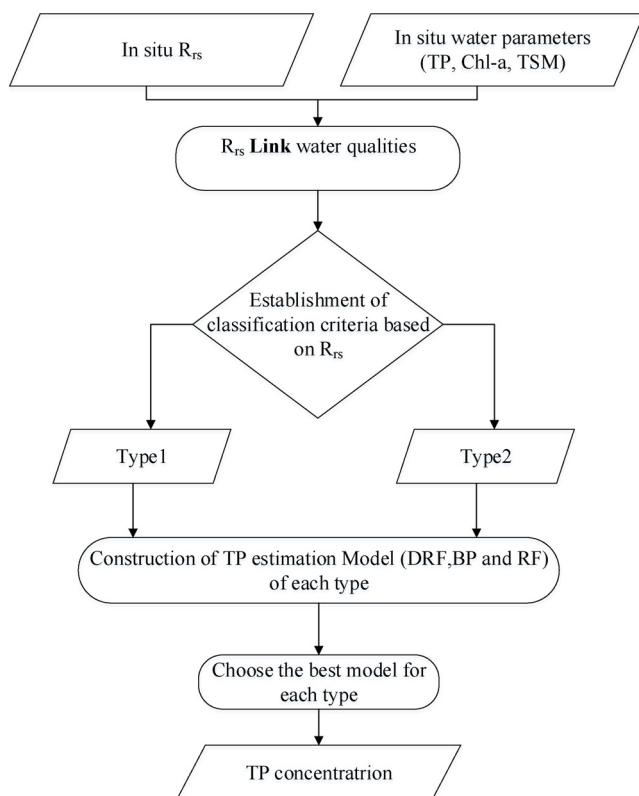


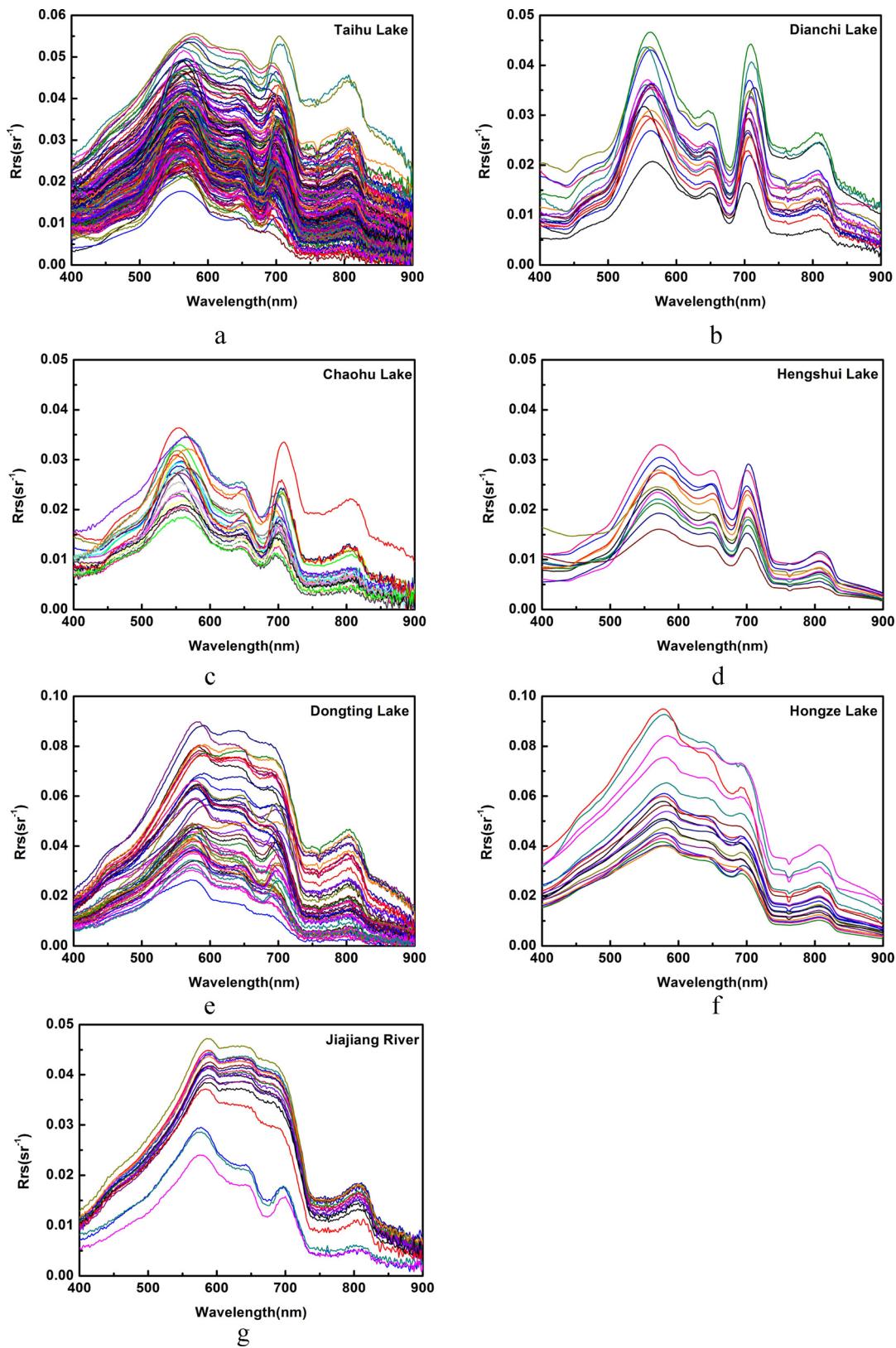
Fig. 2. Framework for TP estimation algorithms based on optical classification.

through the analysis of  $R_{rs}$  and water qualities, the relationship between  $R_{rs}$  and water qualities concentration was determined. According to the difference of water quality in different water bodies, the classification criterion based on  $R_{rs}$  was established. The next step is to apply the modeling methods (DRF, BP and RF) to different types of water bodies for their respective characteristics. Then the most appropriate estimation models for TP were chosen for the two water bodies. Finally, a hybrid algorithm inversion algorithm for TP concentration based on optical classification with high accuracy was established.

### 3.2. Criteria for water classification based on remote sensing reflectance

$R_{rs}$  is a prerequisite for inverting water quality parameters among the apparent optic properties (AOP). Different concentrations of water quality parameters cause the different features in  $R_{rs}$ . The  $R_{rs}$  and the mean content of the water parameters from the study areas are shown in Fig. 3 and Table 2. The Chl-a concentrations in Taihu Lake, Dianchi Lake, Chaohu Lake and Hengshui Lake are higher than other lakes, and they are 56.52 µg/L, 97.15 µg/L, 38.17 µg/L and 51.93 µg/L, respectively. The spectral curves of these lakes show significant differences in the peaks and valleys of the visible light range and the near-infrared range, and they are similar to the studies conducted by Huang et al. (2014), Sun et al. (2011), Huang et al. (2014) and Sun et al. (2011). In contrast, the ISM levels in Dongting Lake, Hongze Lake and the Jiajiang River are 49.52 mg/L, 35.59 mg/L and 59.11 mg/L, which are higher than Chl-a concentration. The spectral curves of  $R_{rs}$  in these lakes are flatter and higher than the other lakes due to higher ISM and lower Chl-a contents. These findings are similar to Zheng et al. (2015), Cao et al. (2017), Cao et al. (2017) and Zheng et al. (2015), who reported that the non-algae particles dominates in Dongting Lake and Hongze Lake. Thus, it can be seen that there is a distinct difference of  $R_{rs}$  between phytoplankton or non-algae particles dominated water bodies.

In previous studies, the ratio of Chl-a/TSM indicates the relative quantity of Chl-a and TSM, it is commonly used to judge which water



**Fig. 3.** Measured remote sensing reflectance. a-g represents Taihu Lake, Dianchi Lake, Chaohu Lake, Hengshui Lake, Dongting Lake, Hongze Lake and the Jiajiang River, respectively.

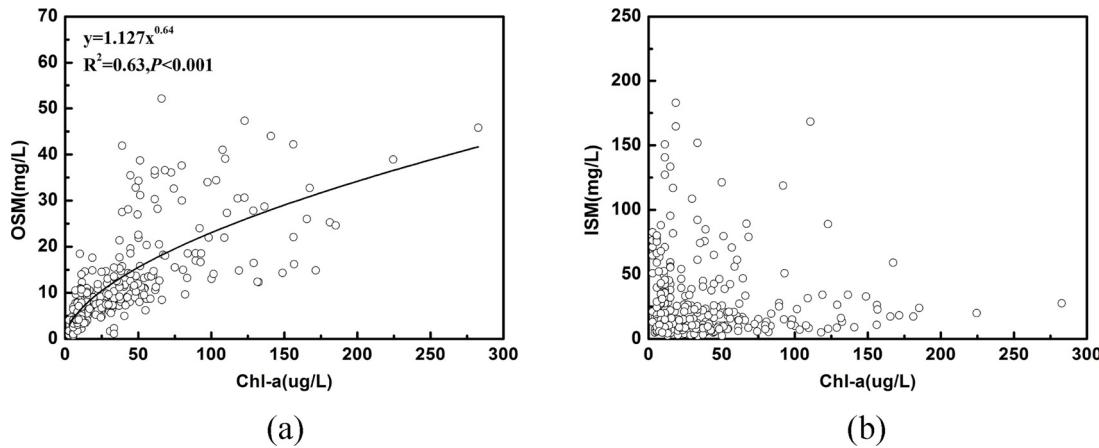
component is predominant in the water optics (Lyu et al., 2015). However, TSM is the sum of ISM and OSM, while OSM may contain algal particles that contain Chl- $\alpha$  as a main component. From Fig. 4, it can be observed that OSM has a higher correlation with Chl- $\alpha$  than ISM.

Thus, to eliminate the impact of the Chl- $\alpha$ , the ratio of Chl- $\alpha$ /ISM is used as the index to judge the type of water. By calculating the ratio of Chl- $\alpha$  and ISM in Table 2, it is found that the ratios of Dongting Lake, Hongze Lake and the Jiajiang River, which are non-algae particles-dominated

**Table 2**

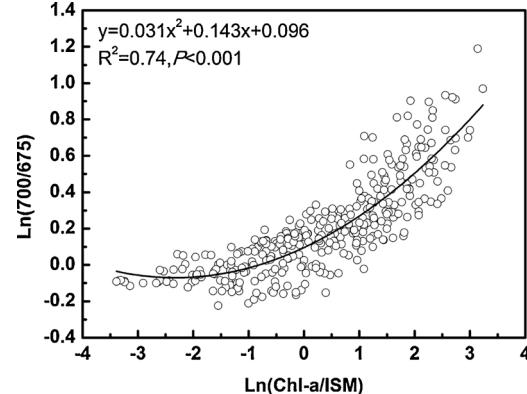
Mean values of water quality parameters in different lakes (Bold values indicate higher concentration).

Study area	Mean Chl-a ( $\mu\text{g/L}$ )	Mean TSM (mg/L)	Mean ISM (mg/L)	Mean OSM (mg/L)	Mean Chl-a/ISM
Taihu Lake	<b>56.52 ± 34.33</b>	42.24 ± 38.18	34.09 ± 34.70	12.69 ± 6.08	2.53 ± 2.67
Dianchi Lake	<b>97.15 ± 35.45</b>	46.53 ± 7.80	9.93 ± 3.91	36.60 ± 7.52	9.78 ± 7.10
Chaochu Lake	<b>38.17 ± 23.04</b>	18.64 ± 10.64	7.49 ± 4.65	11.15 ± 7.32	5.09 ± 3.75
Hengshui Lake	<b>51.93 ± 11.11</b>	40.51 ± 11.92	11.78 ± 7.03	28.72 ± 6.51	6.84 ± 4.93
Dongting Lake	16.90 ± 12.83	56.81 ± 49.41	<b>49.52 ± 47.07</b>	7.29 ± 3.89	0.40 ± 0.78
Hongze Lake	15.48 ± 8.18	42.93 ± 16.96	<b>35.59 ± 15.67</b>	7.34 ± 1.97	0.44 ± 0.33
Jiajiang River	5.96 ± 9.46	60.62 ± 24.85	<b>59.11 ± 24.63</b>	1.52 ± 0.53	0.15 ± 0.48

**Fig. 4.** (a) Power function relationship between OSM and Chl-a; (b) Relationship between ISM and Chl-a.

waters, were 0.40, 0.44 and 0.15, respectively. In addition, the ratios of Taihu Lake, Dianchi Lake, Chaohu Lake and Hengshui Lake, which are phytoplankton-dominated waters, were 2.53, 9.78, 6.84 and 4.40, respectively. Thus, according to the previous study, the water is classified as non-algae particles-dominated water or phytoplankton-dominated water by Chl-a/ISM ratios less than or greater than 1, respectively (Shen et al., 2015; Sun et al., 2012; Tan et al., 2016).

To address the classification by  $R_{rs}$ , the relationship between the ratio of Chl-a/ISM and  $R_{rs}$  was analyzed to determine the specific bands and classification criteria. Using the above analysis of the spectral curves of  $R_{rs}$  in different types water, the significant differences mainly appear in the visible light range and the near-infrared range, especially at 675 and 700 nm. In the  $R_{rs}$  vicinity of 675 nm, almost all curves have a reflection valley, and the curve performance is particularly evident in Taihu Lake, Dianchi Lake, Chaohu Lake and Hengshui Lake, which are phytoplankton-dominated water bodies. Moreover, in the  $R_{rs}$  vicinity of 700 nm, the reflection peak of the curves in the four lakes is very obvious. These characteristics are caused by changes in the dominant factors of the water optics. The 675-nm reflection valley is mainly due to the maximum absorption of Chl-a in the red band, and the 700-nm reflection peak is mainly caused by particle backscatter and the absorption and scatter of water. Comparing the standard deviation of  $R_{rs}$  in each wavelength,  $R_{rs}$  at 675 and 700 nm ( $R_{rs,675}$  and  $R_{rs,700}$ ) have significant differences in different water bodies. In addition, the valley of 675 nm and the peak of 700 nm are strongly related to the relative quantity of Chl-a and ISM, i.e., when the Chl-a content increases or the ISM content decreases, the  $R_{rs}$  value tends to be bigger at 700 nm and smaller at 675 nm, while the value of  $R_{rs}(700)/R_{rs}(675)$  becomes bigger. The cooperative change of  $R_{rs}(700)/R_{rs}(675)$  and Chl-a/ISM indicate the probability of classifying the water into phytoplankton- or non-algae particles-dominated optic water. In Fig. 5, the good correlation between  $\ln(R_{rs}700/R_{rs}675)$  and  $\ln(\text{Chl-a}/\text{ISM})$ , with an  $R^2 = 0.74$ , can be observed. This infers that the  $\ln(R_{rs}700/R_{rs}675)$  can indicate the relative ratio between Chl-a and ISM in water. Corresponding to value 1 of Chl-a/ISM, the  $\ln(R_{rs}700/R_{rs}675)$  is 0.096

**Fig. 5.** Regression between  $\ln(R_{rs}700/R_{rs}675)$  and  $\ln(\text{Chl-a}/\text{ISM})$ .

according to Fig. 5. Thus, two optical types were proposed to discriminate the turbid inland waters. The types are (1) Type 1 (non-algae particles dominated):  $\ln(R_{rs}700/R_{rs}675) \leq 0.096$  and (2) Type 2 (phytoplankton dominated):  $0.096 < \ln(R_{rs}700/R_{rs}675)$ .

### 3.3. Development of hybrid algorithms for operationally estimating TP

In phytoplankton-dominated and non-algae particles-dominated water bodies, the relationship between TP and water quality parameters is different. TP has a better correlation with ISM in non-algae particles-dominated water, whereas it has a better correlation with Chl-a in phytoplankton-dominated water (Chen et al., 2015; Ferris and Tyler, 1985; Hui and Yao, 2016; Wu et al., 2010). Due to the complexity of case II water, TP inversion models might be different for different types of water. Based on the above classification, several TP retrieval algorithms were constructed to analyze the applicability for TP inversion, including data regression analysis and fitting (DRF), back-propagation neural networks (BP) and random forest (RF).

### 3.3.1. Data regression analysis and fitting models

DRF has been widely used in the inversion of water quality parameters (i.e., Chl-a, TSM, CDOM and so on) (Cao et al., 2017; Li et al., 2012; Mannino et al., 2014; Oliveira et al., 2016; Olmanson et al., 2016; Tian et al., 2016), including the linear model, exponential model, power function model and quadratic function model. Furthermore, stepwise regression analysis and enter regression analysis were also carried out to establish the multiple regression analysis model considering the common action and influence of multivariable to TP concentration. In the modeling process, the exact selection of the characteristic band is the basis for achieving higher inversion accuracy. Therefore, the correlation between TP and  $R_{rs}$  was calculated. The four bands with the highest correlation and their band combinations, such as ratio, difference and the NDVI pattern, were used as parameters for modeling.

### 3.3.2. Machine learning methods

Machine learning methods are algorithms that can obtain the rule from the data automatically and forecast the unknown data with the law. Among the numerous machine algorithms, the development of BP is relatively mature and easy to implement (Keiner and Yan, 1998; Schiller and Doerffer, 1999; Zhao et al., 2009). The network, with a hyperbolic tangent function hidden layer and a linear output layer, has been proven capable of characterizing any static nonlinear relationship (Liu et al., 2015). Therefore, BP has been increasingly adopted in remote sensing (Kishino et al., 2005; Song et al., 2012; Zhang et al., 2003). The RF algorithm is a non-parametric statistical technique that is capable of synthesizing regression or classification functions based on discrete or continuous datasets. The RF learning process is faster, and it also has a capability to address complex relationships between predictors due to the noise and large amounts of data (Mutanga et al., 2012). It has been found in previous research that the RF can increase the accuracy of remote sensing image classification (Cortijo and Blanca, 1999; Ghimire et al., 2010; Rodriguez-Galiano et al., 2012). Thus, the two methods were used in this research for TP inversion. The most important part of the machine model is the selection of variables. First of all, the correlation between  $R_{rs}$  and TP was analyzed. The four bands with the highest correlation and the combinations (difference, ratio and NDVI) of these bands were chosen as input variables, the number of independent variables can reach 22. The amount was determined based on our research, firstly, more variables do not necessarily improve model accuracy. When multiple regression analysis is used, the input of too many independent variables could saturate the output of the results. Secondly, if the number is too small, the independent variables are insufficient to characterize TP content. Therefore, after many adjustments, the number of 4 was chosen.

In this study, BP contained three layers: an input layer, a hyperbolic tangent function hidden layer and a linear output layer. The  $R_{rs}$  of the four bands that had higher correlation with TP concentration and their band combinations were chosen as the input layer, and their corresponding TP concentrations were the output data. The BP neural network uses an S-type hyperbolic tangent function as the hidden layer transfer function and the linear function as the output layer function to quickly train the model. With the maximum number of trainings set for 1000 times, the learning rate is 0.001. The most important step in BP neural network design is the choice of the number of nodes in the

hidden layer, which directly affects the mapping ability of the network to complex problems. If the number of hidden nodes is too small, then the network convergence speed is too slow to achieve the accuracy of the request; however, if there are too many nodes, the amount of calculation is increased, thus leading to over-fitting and reducing the generalization of the network capacity. Therefore, this study determined the optimal number of nodes in the hidden layer using several experiments.

The random forest model is constructed by the random forest data packet in R language to invert the TP content. The model input argument is the  $R_{rs}$  of the four bands that had higher correlation with TP concentration and band combinations, and the dependent variable is TP content. In the RF regression algorithm, the number of decision trees (ntree) and the number of independent variables (mtry) are required to create the branch. To find the ntree and mtry values that can best predict the TP content, the two parameters were optimized based on the root mean square error of calibration (RMSEC). The ntree values were tested from 500 to 9500 with a 1000 interval, whereas mtry was tested from 1 to 21 using a single interval (Mutanga et al., 2012).

### 3.4. Assessment of the accuracy of the algorithms

The accuracy of the inverse model was measured by the mean absolute percent error (MAPE) and Root Mean Square Error (RMSE). In addition, the two parameters are calculated as:

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - y'_i}{y_i} \right| * 100\% \quad (2)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - y'_i)^2} \quad (3)$$

where  $y_i$  and  $y'_i$  are the retrieved and measured TP respectively.

## 4. Results and discussion

### 4.1. The different water types and their properties

The results of true classification was divided based on the ratio of Chl-a and ISM concentration. When Chl-a/ISM is less than 1, water is type1, otherwise it is type2. The true and model classification are shown in Table 3. In Dianchi Lake, Chaohu Lake, Hengshui Lake and Jiajiang Lake, the accuracy of the proposed classified method could reach 100%. There are two points misclassified in Dongting Lake and three points misclassified in Hongze Lake. Seven points were misclassified in Taihu Lake because its water component is the most complex. The overall classification accuracy for the all lakes is high and acceptable, and it is concluded that this method can effectively distinguish non-algae particles- and phytoplankton-dominated water.

#### 4.1.1. Water quality parameters of different water types

Based on the water classification method developed above, the sampling data was divided into two categories: type 1 with 103 points and type 2 with 145 points. As seen in Table 4, the minimum, maximum, mean and standard deviation (SD) of the parameters (TP, Chl-a, TSM, ISM and OSM) of different types were calculated. The range of TP

**Table 3**

Comparison between the results of true classification and model classification.

Classification	Taihu	Dianchi	Chaohu	Hengshui	Dongting	Hongze	Jiajiang
True classificaiton	Type1	40	0	0	0	36	13
	Type2	79	18	16	12	12	3
Model classificaiton	Type1	33	0	0	0	38	16
	Type2	86	18	16	12	10	3

**Table 4**

Statistics for the in situ data sets for Type 1 and 2. Min, Max, Mean, and SD represent the minimum, maximum, average and standard deviation, respectively.

Water types	Parameters	Minimum	Maximum	Mean	SD
Type 1	TP	0.023	0.25	0.096	0.051
	Chl-a	2.79	122.76	13.866	14.232
	TSM	7.812	200.533	51.961	40.304
	ISM	4.4	182.933	45.246	38.252
	OSM	0.667	30.667	6.715	4.765
	Chl-a/ISM	0.034	2.959	0.543	0.570
Type 2	TP	0.014	0.396	0.133	0.082
	Chl-a	11.532	324.198	63.439	40.863
	TSM	8.025	195.563	36.871	28.282
	ISM	2.4	168.25	20.439	25.296
	OSM	1.133	52.1	16.433	9.779
	Chl-a/ISM	0.413	23.168	4.919	4.111

content and TSM of the two types are similar. Chl-a generally shows an increasing trend from type 1 ( $2.79 \mu\text{g/L}$ – $122.76 \mu\text{g/L}$ , mean =  $13.866 \mu\text{g/L}$ ) to type 2 ( $11.532 \mu\text{g/L}$ – $324.198 \mu\text{g/L}$ , mean =  $63.439 \mu\text{g/L}$ ). In contrast, ISM shows a decreasing trend from type 1 ( $4.4 \text{ mg/L}$ – $182.933 \text{ mg/L}$ , mean =  $45.246 \text{ mg/L}$ ) to type 2 ( $2.4 \text{ mg/L}$ – $168.25 \text{ mg/L}$ , mean =  $20.439 \text{ mg/L}$ ). Table 4 also shows the ratio of Chl-a/ISM. The mean value of Chl-a/ISM in type 1 of 0.543 is much lower than the value of 4.919 in type 2, indicating that the two types of water bodies are distinguished effectively.

Through the above analysis, there is a large difference in the concentration of water components in different categories. Due to phosphorus's physical properties and chemical form, phosphorus in water is mainly divided into particulate phosphorus and dissolved phosphorus. Particulate phosphorus is often attached to suspended particles, and the correlation with the concentration of ISM is relatively high. On the other hand, aquatic plant and phytoplankton growth are affected by dissolved phosphorus, and the correlation with Chl-a concentration is higher. Then in the two type's water bodies, the influencing factors of phosphorus may be different. As seen in Fig. 6(a), in type 1, which is dominated by non-algae particles, TP and ISM have a good correlation, with an  $R^2$  of 0.731. Thus, ISM could explain 73.1% of the changes in TP content. And there is no relationship between TP and Chl-a. Apparently, TP is mainly affected by non-algae particles because P is adsorbed on it. This phenomenon often appears in non-algae particles dominated lakes, and the estuaries which are the intersections of rivers and lakes, where the water parameters are affected mainly by land-based runoff. In type 2, TP has a positive correlation with Chl-a ( $R^2$  is 0.65), and the Chl-a could explain 65% of the changes in TP. While the relationship between TP and ISM is weak. In this kind of water, phytoplankton is predominant, and the growth and decay process of algae has an important effect on the content of phosphorus. The concentration of TP increases with the increase of Chl-a content. For whole data, TP has weak relationship between Chl-a and ISM.

According to the above results, this classification can reflect the relationship between phosphorus and water components in different waters. After classification, the correlation between TP and  $R_{rs}$  was enhanced by highlighting the main influencing factors of TP in inland water. Thus, the classified water results could provide a better data set for construction of the TP inversion model.

#### 4.1.2. Optical properties of different water types

Particulate absorption ( $a_p$ ) is an important inherent optical quantity of water bodies that significantly affects the underwater light field and water color variations, especially for highly turbid inland waters (Sun et al., 2012), and it is composed of phytoplankton ( $a_{ph}$ ) and non-algal particle absorption ( $a_{nap}$ ). Fig. 7(a) and (b) show the  $a_{ph}$  and  $a_{nap}$  spectra of the two types of water bodies. The value of  $a_{ph}$  for type 2 is higher than the other type. At 440 and 675 nm, there is a significant

characteristic peak, mainly due to Chl-a being the main pigment source for phytoplankton and the strong absorption of Chl-a in the blue and red bands (Dekker, 1993; Haardt and Maske, 1987). The type 1 water body is dominated by non-algae particles, and the characteristics are not obvious. In contrast, in Fig. 7(b), the value of  $a_{nap}$  is the largest. The overall trend presents that exponential decay in the blue band absorption is strong; with the increase in wavelength, the absorption coefficient decreased rapidly. As for the contributions of phytoplankton and non-algal particles to total particulate absorption, large differences were found among the two optical types of waters. The type 1 water presents the least phytoplankton contribution to the total particulate absorption; the overall contribution rate is less than 50%, and it gradually reaches the highest at the red wave bands. The type 2 water exhibits the largest phytoplankton contribution, and most wavelengths have the ratios larger than 50%.

The characteristics of  $R_{rs}$  in Fig. 7(d) are obviously different in type 1 and type 2. The value of  $R_{rs}$  in type 1 is higher, in the range of 400 nm to approximately 700 nm, which is similar to the shape of Dongting Lake, Hongze Lake and Jiajiang Lake. Based on the above analysis, the shape of the spectral type 1 curve is mainly caused by the higher concentration of ISM with little algae. The shape of the spectral type 2 curve is similar in Dianchi and Chaohu Lake, showing the characteristics of severe eutrophication of the spectral curve, of which algae dominate. At 550 nm, there is a reflection peak because of the weak absorption of Chl-a and carotene and the scattering of cells. Similarly, the reflection of the peak near 700 nm has become more obvious, and the position of the reflection peak has a significant redshift. The difference between the values of the peak near 700 and the valley near 675 is larger. These features are both caused by the higher concentration of Chl-a. Based on the above analysis, the two types classified by the proposed classification algorithm have significant differences. Spectral characteristics correspond well to the water quality parameter concentration.

#### 4.1.3. Comparing of the classification algorithms

Many scholars have established a variety of water classification methods for their respective research objectives. Among the proposed algorithms, three common classification algorithms (i.e., cluster analysis, depths of  $R_{rs}$  near 680 nm (TD680) and the maximum chlorophyll index (MCI)) were compared in this study. Cluster analysis is a process of classifying data into different classes or clusters by maximizing intra-group similarity and inter-group variability. In this study, the water was classified into 3 types by the clustering method, which had the best performance when selected from some algorithms calculated by SPSS. TD680 was established by Sun et al. (2011) to refine the optical classification of turbid waters in order to improve the performance of water color algorithms, mainly based on the strong correlations between the trough depths of  $R_{rs}$  near 680 nm and the ratios of ISM to TSM concentrations for water classification (Sun et al., 2011). According to the original classification criteria, the data could be divided into three categories. The MCI was defined by (Gower et al., 2005). MCI is a measure of the reflectance peak at 709 nm in water-leaving reflectance. Through the analysis on the relation between MCI and Chl-a content, the water was classified for inversion of Chl-a. In published research, Chl-a content had a good correlation with MCI (Binding et al., 2010; Matsushita et al., 2015; Matthews, 2011; Matthews et al., 2012). However, in our actual sampled data, as shown in Fig. 8(b), there is no correlation between MCI and Chl-a. The MCI method is based on oceanic water bodies for establishment and use. Inland lake water is more complex, so the model does not apply.

Based on the results classified by the new classification method (i.e., two types: Type 1 (103 points) and Type 2 (145 points)), cluster analysis (i.e., three types: CA1 (68 points), CA2 (146 points) and CA3 (34 points)) and TD680 (i.e., three types: TD1 (38 points), TD2 (191 points) and TD3 (19 points)), the correlation between TP and  $R_{rs}$  for different types was calculated to test the applicability of the classification

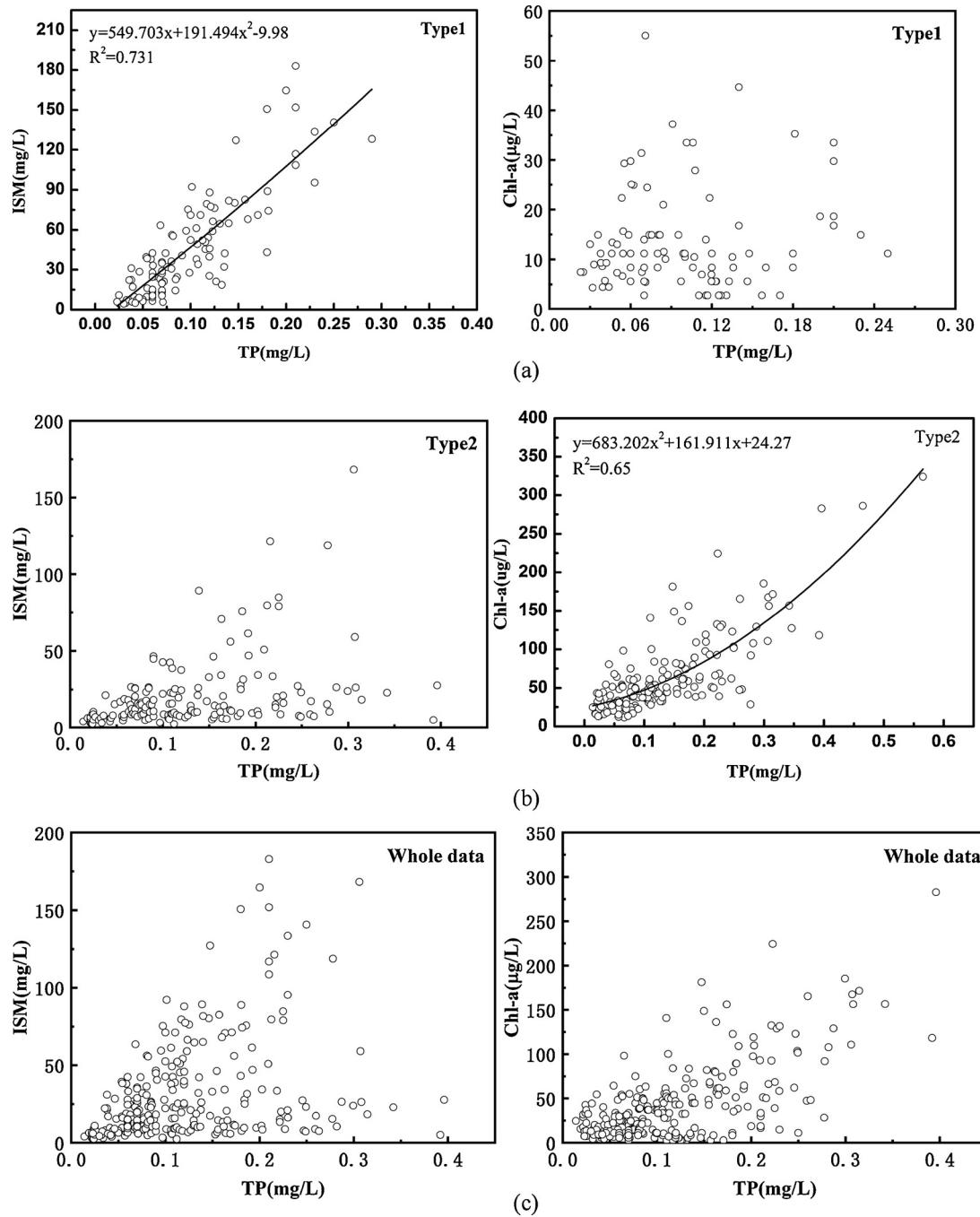


Fig. 6. Statistical relationships between TP concentration and other water parameters. (a) type 1; (b) type 2 and (c) whole data.

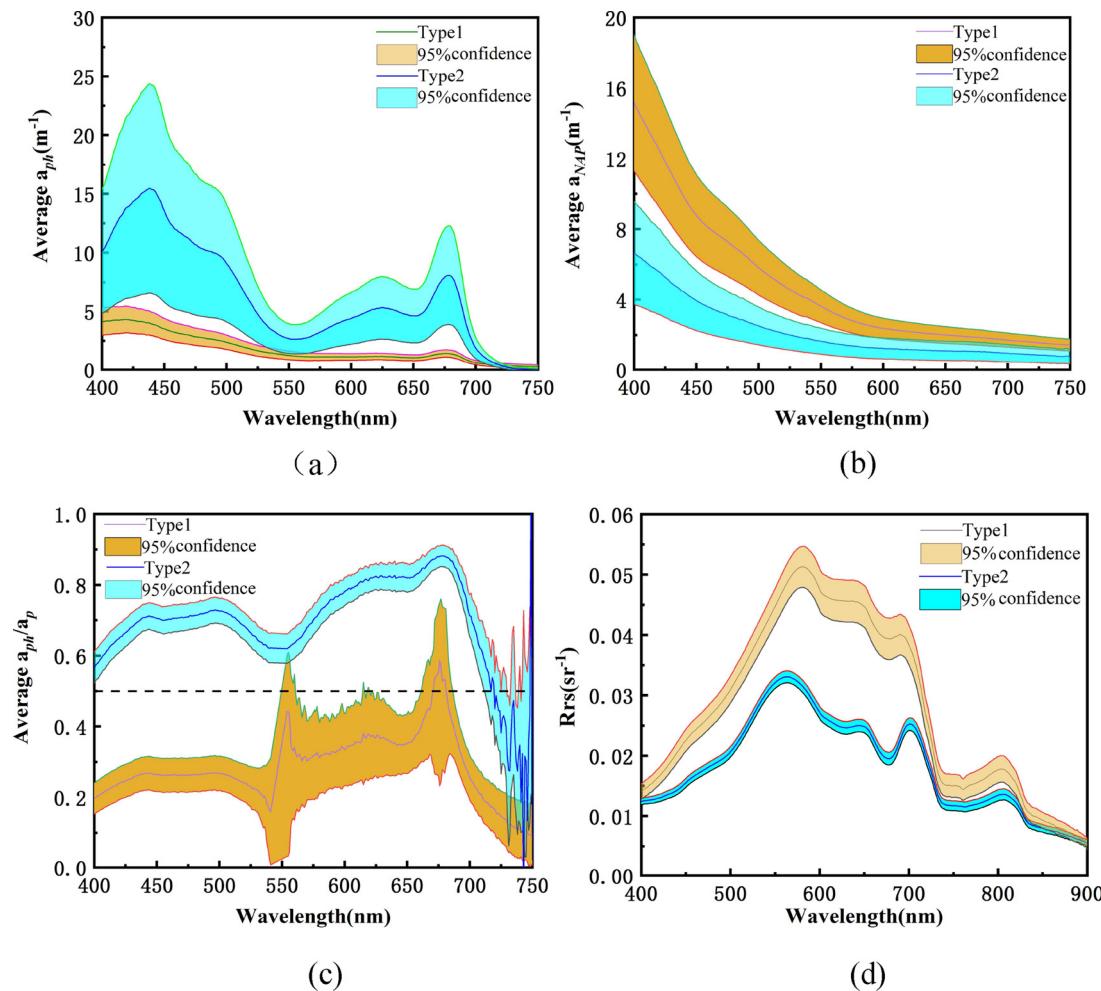
method for TP inversion. As seen in Fig. 8(a), type 1 and type 2 both have higher correlation than other types; the largest values for the two types were 0.82 and 0.73. For types CA1, CA2 and CA3, the values vary greatly in all bands, and the largest are –0.64, 0.5 and 0.66. The largest values of TD1, TD2 and TD3 are 0.68, 0.48 and 0.4, respectively. By comparing the results, it is obvious that the types classified by the new method had the maximum correlation values. Therefore, it is proven that the classification algorithm proposed in this paper is helpful for improving TP retrieval accuracy.

The reasons for the poor effect of the two methods were analyzed as follows. For cluster analysis, the curve trend and the peak position of the three types of spectral curves are basically the same in Fig. 8(c); the only difference was the overall value. It can be said that this classification algorithm does not reflect the optical trend characteristics of

different water bodies. In addition, it is based on the size of the height of the classification, so this algorithm is prone to misclassification. For TD680, four typical spectral curves were selected in Fig. 8(d). Case 2 and case 3 have distinctly different spectral characteristics, and the peaks on both sides of the groove are the opposite. However, the calculated TD680 is the same. Therefore, this method will be mistakenly classified. In conclusion, these existing methods cannot effectively divide the water into non-algae particles dominated and predominantly phytoplankton water bodies.

#### 4.2. TP estimation algorithms for different water types

Among all of the sample points, 76 were used to construct the TP inversion model, whereas 27 points were used to verify the precision in



**Fig. 7.** Average absorption spectra of the two optical types of water: (a) phytoplankton; (b) non-algal particles; (c) contribution of phytoplankton to total particulate absorption; (d) Means of Remote sensing reflectance of the two types.

type 1 water. In type 2 water, 106 points were used for modeling, and 39 points were used for verification. The accuracy of each algorithm for each group was re-evaluated by comparing the estimated TP values with the measured TP values. The modeling results and detailed information about the optimal model for each type are shown in Table 5 and Fig. 9, (1) for type 1, the DRF gives the best accuracy, and the combination of  $R_{rs}$  798 and  $R_{rs}$  803 was used to construct a polynomial equation. The RMSE and MAPE are 0.034 mg/L and 20.8%, respectively. These values are smaller than the other two algorithms; (2) for type 2, the DRF has higher accuracy, and the logarithmic transform of TP and  $R_{rs}$  730 were used to construct a linear model. The RMSE and MAPE are 0.057 mg/L and 22.55%, respectively. The comparison with the other two models also shows that DRF is the best choice; (3) for the overall data, the RMSE and MAPE of the DRF is also lowest, the RMSE and MAPE are 0.058 mg/L and 37.51%.

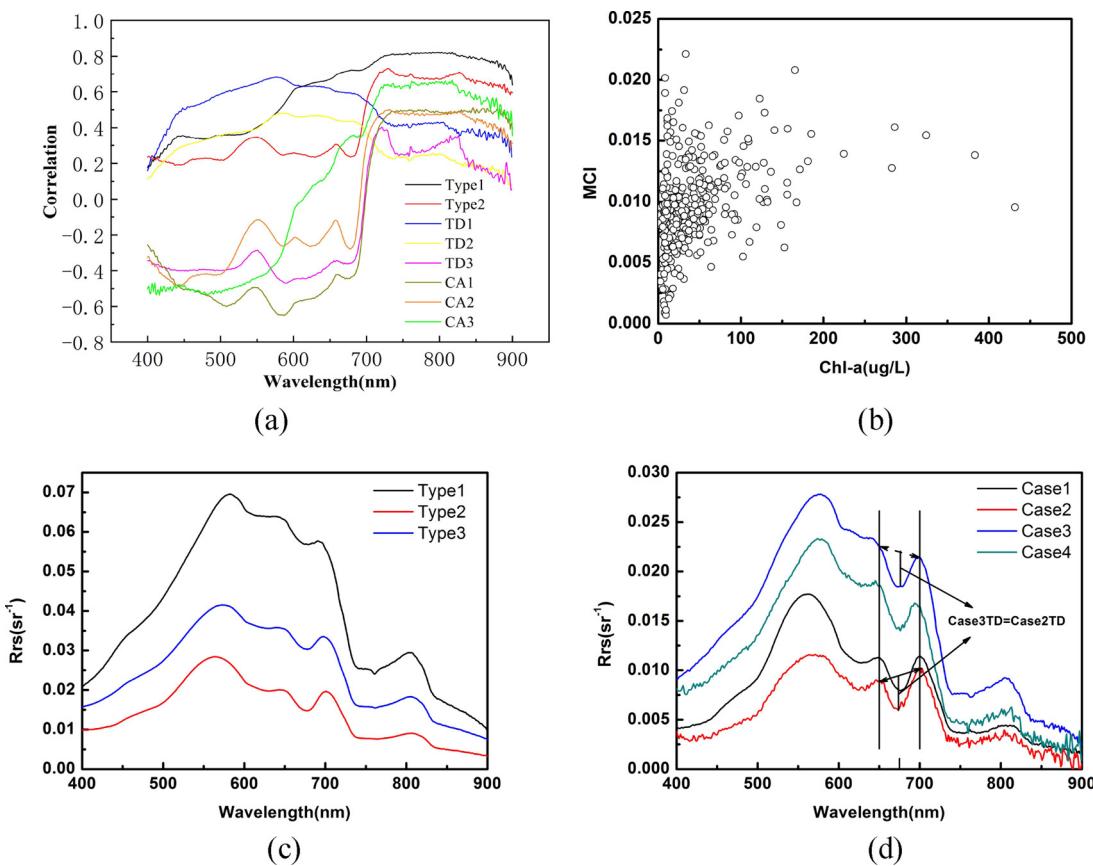
It can be concluded that the DRF model displays good accuracy in all three datasets. In addition, after classification, for BP and RF, the accuracy of type 1 and type 2 are also better than the whole data. For BP, the RMSEs of type 1, type 2 and whole data are 0.057 mg/L, 0.056 mg/L and 0.062 mg/L, and the MAPEs are 38.83%, 39.44% and 47.93%. For RF, the RMSEs of type 1, type 2 and whole data are 0.06 mg/L, 0.053 mg/L and 0.061 mg/L, and the MAPEs are 37.71%, 36.70% and 42.10%. That means the inversion results are improved after type 1 and 2 classification. In the low value area, the verification points are both well distributed near the line 1:1. In contrast, there is an underestimating phenomenon in the high value area. It is deduced that this may be caused by the dissolved phosphorus in the water. It cannot

be fully monitored by the model when the TP content is high.

Based on the results above, it is feasible to use the remote sensing technique to invert TP, and the accuracy of the three inversion methods has been improved by classification. DRF method inversion results have higher accuracy in all datasets compared with BP and RF. DRF, as a simple and direct and effective method, has been widely used in remote sensing inversion, and its application is also applicable to the inversion of TP concentrations in water. All the validation results have the phenomenon of underestimation in the high value area. To further improve the accuracy of the model, the relationship between dissolved phosphorus and algae and suspended matter needs to be further explained, which will also be the focus of the next step.

In order to analyze the influence of Chl-a and ISM concentration on the established model, the relationship between the relative errors and the ratio of TP to Chl-a (TP: Chl-a) in type1 and the ratio of TP to ISM (TP: ISM) was studied, shown as Fig. 10. It can be seen that the relative error of the model has no relationship with TP:Chl-a in type1. And the model relative error and the TP: ISM value is basically not correlated in type2. It can be concluded that Chl-a has little influence on the inversion model in type1 and ISM also has little influence on the inversion model in type2.

There are several existed TP estimation algorithms, the comparison between the TP inversion models proposed and the existed algorithms was shown in Table 6 (Gao et al., 2015; Huang et al., 2015; Isenstein and Park, 2014; Wu et al., 2010). The validation data consisted the validation data of type1 and type2. Seen from Table 6, MAPEs of the algorithms proposed by the previous studies are all more than 50%,



**Fig. 8.** (a) Correlation between TP and  $R_{rs}$  for different types (the new classification methods: Type1 and Type2; depths of  $R_{rs}$  near 680 nm (TD680); TD1, TD2 and TD3; cluster analysis: CA1, CA2 and CA3); (b) the relationship between MCI and Chl-a. (c) Classification results of  $R_{rs}$  from cluster analysis. (d) Examples of depths of  $R_{rs}$  near 680 nm (TD680).

**Table 5**  
Performance of the TP estimation algorithms (Bold values indicate higher accuracy).

Classification	DRF	BP	RF	The optimal algorithm	Model parameter
Type1 (n = 103)					
RMSE(mg/L)	<b>0.034</b>	0.057	0.06	$y = 2.203x^2 + 1.903x + 0.035$	y: TP; x: $R_{rs} 798 + R_{rs} 803$
MAPE (%)	<b>20.38</b>	38.83	37.71		
$R^2$	0.80	0.28	0.24		
Type2 (n = 145)					
RMSE(mg/L)	<b>0.057</b>	0.056	0.053	$y = 0.99x + 2.199$	y: ln(TP); x: ln( $R_{rs} 730$ )
MAPE (%)	<b>22.55</b>	39.44	36.70		
$R^2$	0.66	0.52	0.61		
Whole data (n = 248)					
RMSE(mg/L)	<b>0.058</b>	0.062	0.061	$y = 2.514x^{0.679}$	y: TP; x: $R_{rs} 827$
MAPE (%)	<b>37.51</b>	47.93	42.10		
$R^2$	0.45	0.34	0.43		

while the MAPE of the method in this paper is only 21.66%. And the RMSEs of the early algorithms are high up to greater than 0.06 mg/L, while the RMSE of the proposed method is only 0.049 mg/L. The estimation accuracy of the new methods was markedly improved.

#### 4.3. Applicability analysis of the TP inversion algorithm for Sentinel-3 data

To ensure the continuation of the data, on the basis of MERIS, the OLCI has been improved in some performance. The main improvements are as follows: the number of bands increased from 15 to 21; the signal-to-noise ratio is improved; the accuracy of the positioning is improved; the return cycle is shortened, the global ocean return cycle is less than 4 days and MERIS is approximately 15 days; data release more frequently, and every 3 h L1B and L2 products are released ([https://sentinels.copernicus.eu/documents/247904/685236/Sentinel-3\\_User\\_Handbook](https://sentinels.copernicus.eu/documents/247904/685236/Sentinel-3_User_Handbook)).

It seems that the OLCI sensor will further promote the development of water color remote sensing, which is of great significance.

The measured  $R_{rs}$  data were simulated to the OLCI sensor, and then the proposed classification method was used to classify. Bands 9 and 11 were chosen for spectral classification, whose center wavelengths were similar to the hyperspectral data. The classification results were as follows: Type 1 with 148 points and type 2 with 100 points. Compared with the true classification result in Table 3, there are some misclassifications in type 1 and type 2. In addition, as seen in Fig. 11, although the difference in  $R_{rs}$  is not as significant as the hyperspectral spectrum due to the wide band of Sentinel 3, the spectral difference in the typical band was still present.

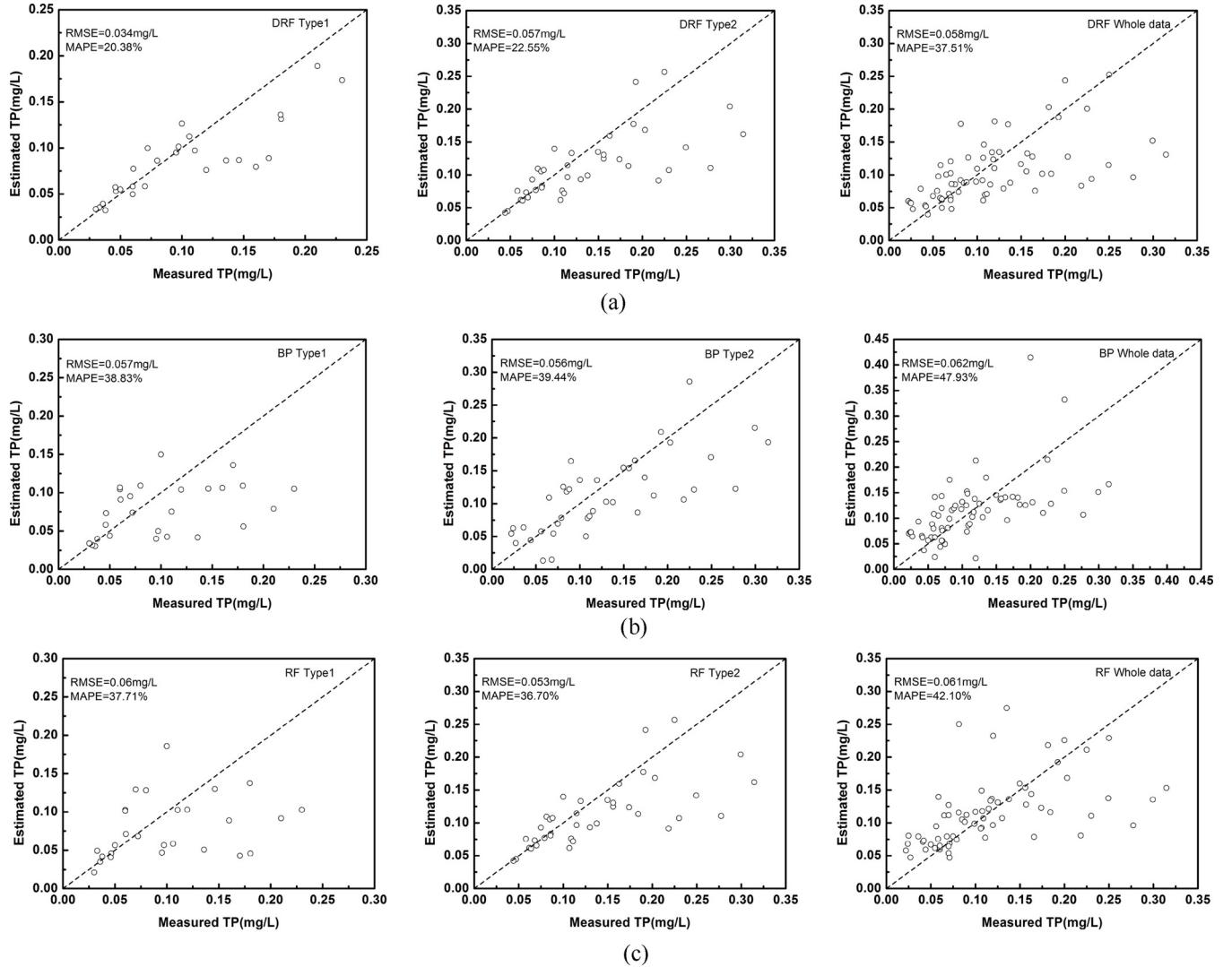


Fig. 9. Comparison of the measured and estimated TP values for the different types based on DRF, BP and RF. (a): DRF; (b): BP; (c): RF.

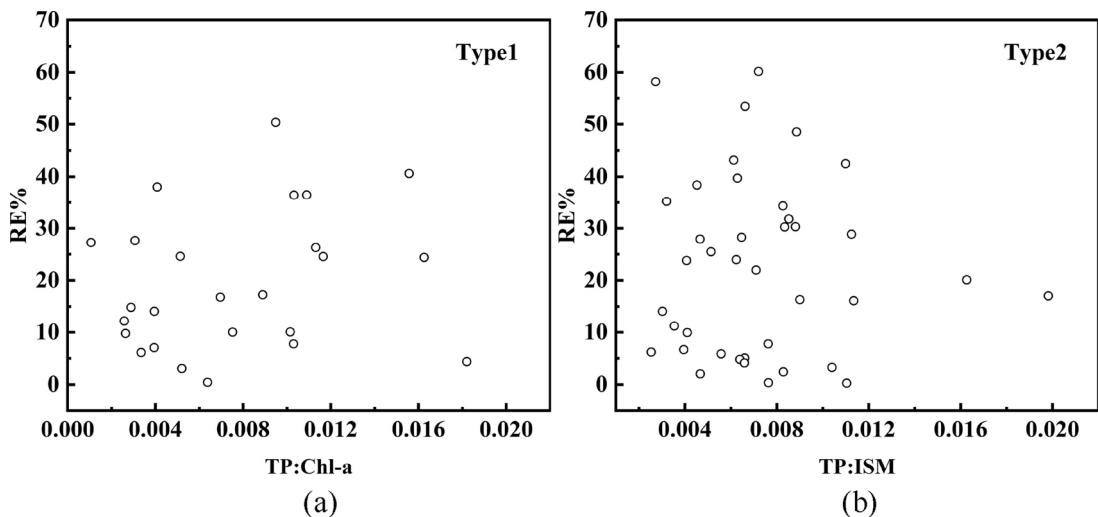


Fig. 10. Analysis of influencing factors on the inversion model (a):Type1; (b) Type2.

**Table 6**

The MAPE and RMSE of TP estimation algorithms.

Model comparison	MAPE (%)	RMSE (mg/L)
The new classification method and inversion models	<b>21.66</b>	<b>0.049</b>
Wu et al. (2010)	55.12	0.062
Isenstein and Park (2014)	63.33	0.084
Huang et al. (2015)	50.59	0.060
Gao et al. (2015)	75.64	0.098

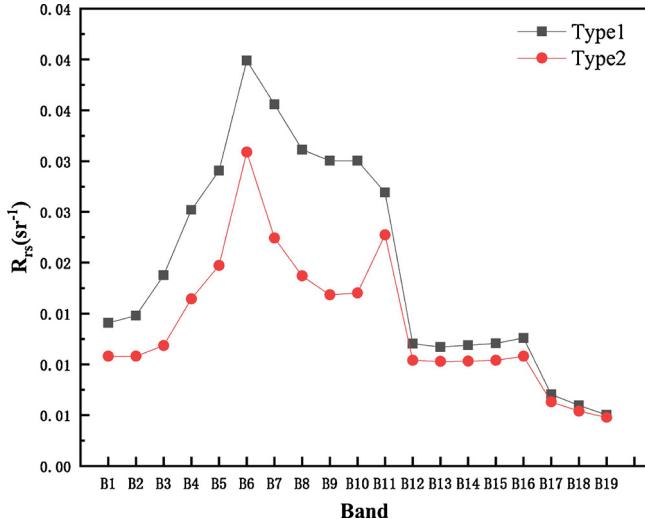


Fig. 11. Means of remote sensing reflectance of the two types.

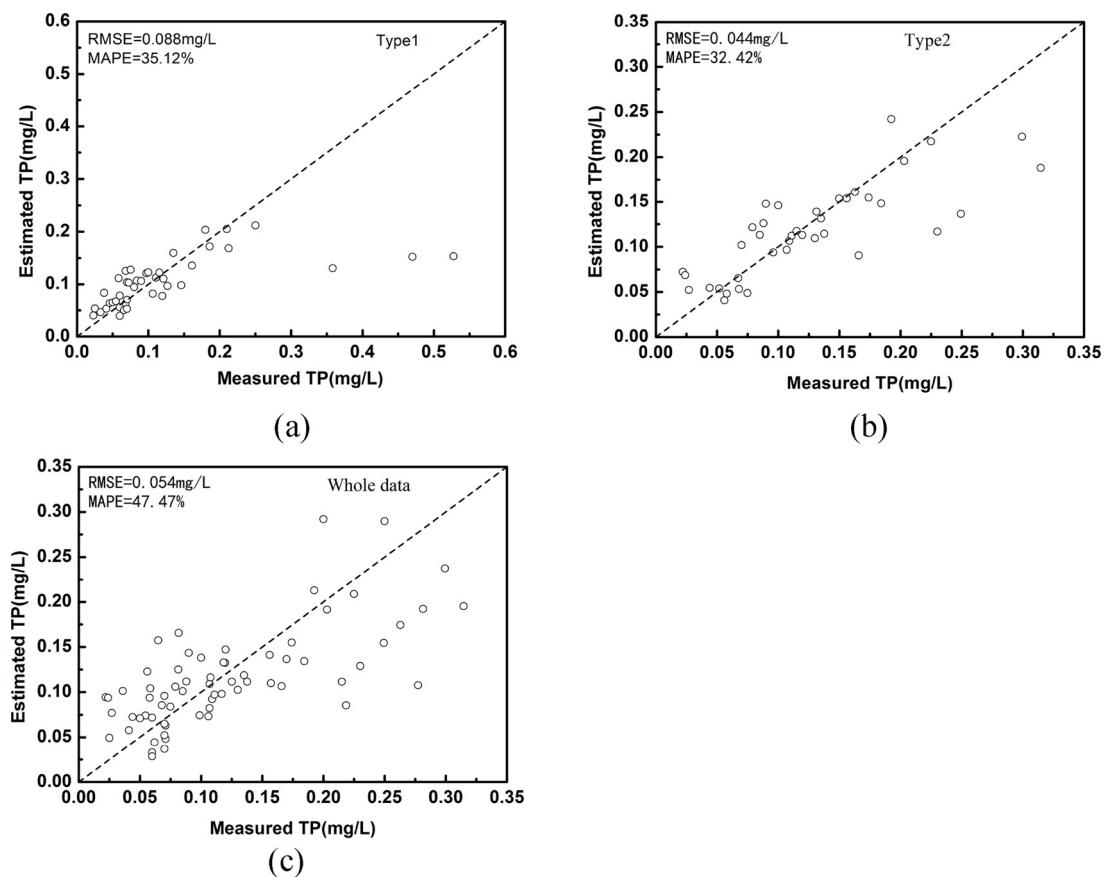
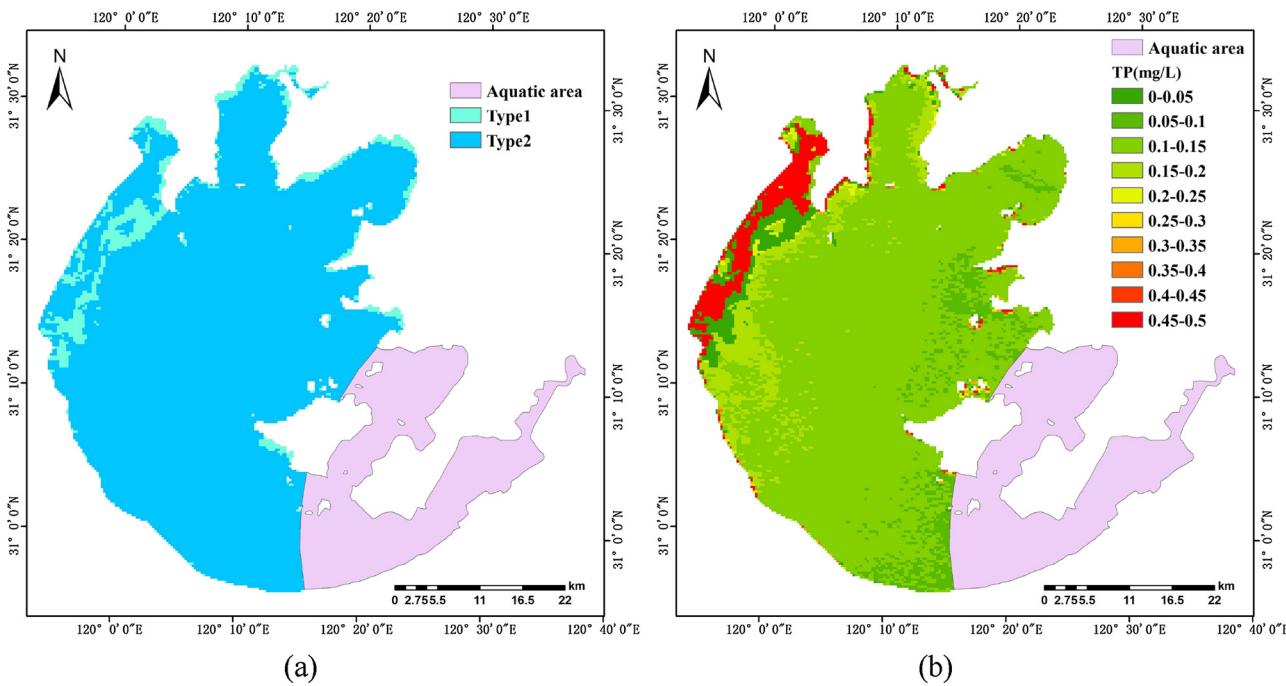


Fig. 12. Comparison of the measured and estimated TP values for the different types. (a) Type 1; (b) Type 2 and (c) whole data.



**Fig. 13.** (a) Water distribution in Taihu Lake on July 24th, 2017; (b) spatial distribution of the TP concentration in Taihu Lake on July 24th, 2017.

inflow rivers are primarily located in the northwestern region, which brings a large amount of sediment and pollutants into the lake (Du et al., 2017; Qin et al., 2007). These river inputs cause the northwestern area dominated by non-algae particles and the TP concentration is higher than other areas.

## 5. Conclusions

To retrieve the TP concentration from remote sensing data, a new water classification method was proposed in this study, and three kinds of models were compared to build a TP inversion algorithm.

Although the phosphorus is not the optic activity matter, its quantity can be reflected by suspended matter or phytoplankton. In fact, some of the phosphorus presenting in granular form should be adsorbed on inorganic suspended solids and algae particles. In addition, the symbiotic relationship between the growth of algae and phosphorus indicates that the Chl-a might reflect the concentration of TP. In this research, a new classification method was established, by which the complex water conditions could be divided into two optical types. They are (1) Type 1:  $\ln(R_{rs700}/R_{rs675}) \leq 0.096$  and (2) Type 2:  $0.096 < \ln(R_{rs700}/R_{rs675})$ . Type 1 belongs to non-algae particles dominated water, whereas type 2 is dominated by Chl-a. It was found that ISM strongly correlated with TP in type 1 water, whereas Chl-a had a good positive correlation with TP in type 2 water. Moreover, the inherent optical parameters  $a_{nap}$  and  $a_{ph}$  presented obviously different features in type 1 and 2 waters. Meanwhile, in the type1 water, the overall contribution rate of phytoplankton to the total particulate absorption is less than 50%, whereas in type 2 water, it is greater than 50% at most wavelengths. The different optical properties make it necessary to construct different TP inversion models based on the water type.

Three kinds of models, DRF, BP and RF, were constructed to examine the performance of TP inversion. The results show that the accuracy of the three methods is improved after water classification compared with the original data. In addition, in the three methods, DRF performs best, with the highest precision in different types of waters. The phenomenon of underestimation exists in the validation results due to the dissolved phosphorus in water.

The classification method and DRF were applied to Sentinel-3 data

to validate the suitability of the method for a water color sensor. The simulated Sentinel-3 bands yielded good accuracy for TP estimation. These findings indicate that the DRF described in this study could be suitable for current and future quantitative monitoring of TP using the Sentinel-3 OLCI in inland common eutrophic water such as Taihu Lake and Dongting Lake.

## Acknowledgements

This research was financially supported by the National Key Research and Development Program of China [2017YFB0503902], Major Science and Technology Program for Water Pollution Control and Treatment [2017ZX07302-003], the National Natural Science Foundation of China [41671340], [41501374], [41571324], [41701412], [41701422], [41701423] and the Natural Science Foundation of Jiangxi Province [2017BAB213024]. The authors would like to thank, Meng Mu, Shaohua Lei, Jie Xu, Xiaolei Ding and Shun Bi for their participation in field sample collection and experimental analysis.

## References

- Akbar, T.A., Hassan, Q.K., Achari, G., 2010. A remote sensing based framework for predicting water quality of different source waters. *Int. Arch. Photogramm.* 38.
- Binding, C.E., Greenberg, T.A., Jerome, J.H., Bukata, R.P., Letourneau, G., 2010. An assessment of MERIS algal products during an intense bloom in Lake of the Woods. *J. Plankton Res.* 33, 793–806.
- Cao, Z., Duan, H., Feng, L., Ma, R., Xue, K., 2017. Climate- and human-induced changes in suspended particulate matter over Lake Hongze on short and long timescales. *Remote Sens. Environ.* 192, 98–113.
- Chang, N.B., Xuan, Z.M., 2013. Monitoring nutrient concentrations in Tampa Bay with MODIS images and machine learning models. *IEEE Int. Conf. Netw. Sens.* 702–707.
- Chen, Y., 2006. Discussion on possible error for phytoplankton chlorophyll-a concentration analysis using hot-ethanol extraction method. *J. Lake Sci.* 18, 550–552.
- Chen, S., Han, L., Chen, X., Li, D., Sun, L., Li, Y., 2015. Estimating wide range total suspended solids concentrations from MODIS 250-m imagers: an improved method. *ISPRS J. Photogramm.* 99, 58–69.
- Chen, X., Yang, X., Dong, X., Liu, E., 2013. Environmental changes in Chaohu Lake (southeast, China) since the mid 20th century: the interactive impacts of nutrients, hydrology and climate. *Limnol.—Ecol. Manage. Inland Waters* 43, 10–17.
- Cortijo, F.J., Blanca, N.P.D.L., 1999. The performance of regularized discriminant analysis versus non-parametric classifiers applied to high-dimensional image classification. *Int. J. Remote Sens.* 20, 3345–3365.
- Dekker, A.G., 1993. Detection of Optical Water Quality Parameters for Eutrophic Waters

- by High Resolution Remote Sensing. Amsterdam Vrije Universiteit.
- Domagalski, J., Lin, C., Luo, Y., Kang, J., Wang, S., Brown, L.R., Munn, M.D., 2006. Eutrophication study at the Panjiakou-Daheiting Reservoir system, northern Hebei Province, People's Republic of China: Chlorophyll-a model and sources of phosphorus and nitrogen. *Acta Geochim.* 25, 136–137.
- Du, C., Li, Y., Wang, Q., Liu, G., Zheng, Z., Mu, M., Li, Y., 2017. Tempo-spatial dynamics of water quality and its response to river flow in estuary of Taihu Lake based on GOCI imagery. *Environ. Sci. Pollut. Res. Int.* 24, 28079–28101.
- Du, C., Li, Y., Wang, Q., Zhu, L., Lu, H., 2016. Inversion model and daily variation of total phosphorus concentrations in Taihu Lake based on GOCI data. *huan jing ke xue = huanjingxue/[bian ji, Zhongguo ke xue yuan huan jing ke xue wei yuan hui "huan jing ke xue" bian ji wei yuan hui.]* 37, 862.
- Ferris, J., Tyler, P., 1985. Chlorophyll-total phosphorus relationships in Lake Burragorang, New South Wales, and some other Southern Hemisphere lakes. *Mar. Freshw. Res.* 36, 157–168.
- Gao, Y., Gao, J., Yin, H., Liu, C., Xia, T., Wang, J., Huang, Q., 2015. Remote sensing estimation of the total phosphorus concentration in a large lake using band combinations and regional multivariate statistical modeling techniques. *J. Environ. Manage.* 151, 33–43.
- Ghimire, B., Rogan, J., Miller, J., 2010. Contextual land-cover classification: incorporating spatial dependence in land-cover classification models using random forests and the Getis statistic. *Remote Sens. Lett.* 1, 45–54.
- Gower, J., King, S., Borstad, G., Brown, L., 2005. Detection of intense plankton blooms using the 709 nm band of the MERIS imaging spectrometer. *Int. J. Remote Sens.* 26, 2005–2012.
- Guo, Q.Z., Wu, X.X., Bing, Q.X., Pan, Y.Y., Wang, Z.H., Fu, Y., Wang, D.C., Liu, J.N., 2016. Study on retrieval of chlorophyll-a concentration based on landsat OLI imagery in the Haihe River, China. *Sustainability (Basel)* 8.
- Haardt, H., Maske, H., 1987. Specific *in vivo* absorption coefficient of chlorophyll a at 675 nm. *Limnol. Oceanogr.* 32, 608–619.
- Huang, C., Guo, Y., Yang, H., Li, Y., Zou, J., Zhang, M., Lyu, H., Zhu, A., Huang, T., 2015. Using remote sensing to track variation in phosphorus and its interaction with chlorophyll-a and suspended sediment. *IEEE J-Stars* 8, 4171–4180.
- Huang, C., Wang, X., Yang, H., Li, Y., Wang, Y., Chen, X., Xu, L., 2014. Satellite data regarding the eutrophication response to human activities in the plateau lake Dianchi in China from 1974 to 2009. *Sci. Total Environ.* 485–486, 1–11.
- Hui, J., Yao, L., 2016. Analysis and inversion of the nutritional Status of China's Poyang Lake Using MODIS Data. *J. Indian Soc. Remote* 44, 837–842.
- Isenstein, E.M., Park, M.-H., 2014. Assessment of nutrient distributions in Lake Champlain using satellite remote sensing. *J. Environ. Sci.* 26, 1831–1836.
- Keiner, L.E., Yan, X.H., 1998. A neural network model for estimating sea surface chlorophyll and sediments from thematic mapper imagery. *Remote Sens. Environ.* 66, 153–165.
- Kishino, M., Tanaka, A., Ishizaka, J., 2005. Retrieval of Chlorophyll a, suspended solids, and colored dissolved organic matter in Tokyo Bay using ASTER data. *Remote Sens. Environ.* 99, 66–74.
- Le, C., Zha, Y., Li, Y., Sun, D., Lu, H., Yin, B., 2010. Eutrophication of lake waters in China: cost, causes, and control. *Environ. Manage.* 45, 662–668.
- Li, Y., Wang, Q., Wu, C., Zhao, S., Xu, X., Wang, Y., Huang, C., 2012. Estimation of chlorophyll a concentration using NIR/red bands of MERIS and classification procedure in inland turbid Water. *IEEE Trans. Geosci. Remote* 50, 988–997.
- Liu, W., Qiu, R., 2007. Water eutrophication in China and the combating strategies. *J. Chem. Technol. Biotechnol.* 82, 781–786.
- Liu, J.M., Zhang, Y.J., Yuan, D., Song, X.Y., 2015. Empirical estimation of total nitrogen and total phosphorus concentration of urban water bodies in China using high resolution IKONOS multispectral imagery. *Water* 7, 6551–6573.
- Lorenzen, C.J., 1967. Determination of chlorophyll and pheo-pigments: spectro-photometric equations. *Limnol. Oceanogr.* 12, 343–346.
- Lyu, H., Li, X., Wang, Y., Jin, Q., Cao, K., Wang, Q., Li, Y., 2015. Evaluation of chlorophyll-a retrieval algorithms based on MERIS bands for optically varying eutrophic inland lakes. *Sci. Total Environ.* 530–531, 373–382.
- Ma, R., Tang, J., Dai, J., 2006. Bio-optical model with optimal parameter suitable for Taihu Lake in water colour remote sensing. *Int. J. Remote Sens.* 27, 4305–4328.
- Mannino, A., Novak, M.G., Hooker, S.B., Hyde, K., Aurin, D., 2014. Algorithm development and validation of CDOM properties for estuarine and continental shelf waters along the northeastern U.S. coast. *Remote Sens. Environ.* 152, 576–602.
- Matsushita, B., Yang, W., Yu, G., Oyama, Y., Yoshimura, K., Fukushima, T., 2015. A hybrid algorithm for estimating the chlorophyll-a concentration across different trophic states in Asian inland waters. *ISPRS J. Photogramm.* 102, 28–37.
- Matthews, M.W., 2011. A current review of empirical procedures of remote sensing in inland and near-coastal transitional waters. *Int. J. Remote Sens.* 32, 6855–6899.
- Matthews, M.W., Bernard, S., Robertson, L., 2012. An algorithm for detecting trophic status (chlorophyll-a), cyanobacterial-dominance, surface scums and floating vegetation in inland and coastal waters. *Remote Sens. Environ.* 124, 637–652.
- Mutanga, O., Adam, E., Cho, M.A., 2012. High density biomass estimation for wetland vegetation using world view-2 imagery and random forest regression algorithm. *Int. J. Appl. Earth Obs.* 18, 399–406.
- Oliveira, E.N., Fernandes, A.M., Kampel, M., Cordeiro, R.C., Brandini, N., Vinzon, S.B., Grassi, R.M., Pinto, F.N., Fillipo, A.M., Paranhos, R., 2016. Assessment of remotely sensed chlorophyll-a concentration in Guanabara Bay, Brazil. *J. Appl. Remote Sens.* 10, 026003.
- Olmanson, L.G., Brezonik, P.L., Finlay, J.C., Bauer, M.E., 2016. Comparison of Landsat 8 and Landsat 7 for regional measurements of CDOM and water clarity in lakes. *Remote Sens. Environ.* 185, 119–128.
- Pacciaroni, M., Crisp, G., 2007. Chlorophyll signatures and nutrient cycles in the Mediterranean Sea: a model sensitivity study to nitrogen and phosphorus atmospheric inputs. *Biogeosci. Discuss.* 4, S1033–S1044.
- Qin, B., Gao, G., Zhu, G., Zhang, Y., Song, Y., Tang, X., Xu, H., Deng, J., 2012. Lake eutrophication and its ecosystem response. *Chin. Sci. Bull.* 58, 961–970.
- Qin, B., Xu, P., Wu, Q., Luo, L., Zhang, Y., 2007. Environmental issues of Lake Taihu, China. *Hydrobiologia* 581, 3–14.
- Rodriguez-Galiano, V.F., Ghimire, B., Rogan, J., Chica-Olmo, M., Rigol-Sánchez, J.P., 2012. An assessment of the effectiveness of a random forest classifier for land-cover classification. *ISPRS J. Photogramm.* 67, 93–104.
- Savage, C., Leavitt, P.R., Elmgren, R., 2010. Effects of land use, urbanization, and climate variability on coastal eutrophication in the Baltic Sea. *Limnol. Oceanogr.* 55, 1033–1046.
- Schiller, H., Doerffer, R., 1999. Neural network for emulation of an inverse model operational derivation of case II water properties from MERIS data. *Int. J. Remote Sens.* 20 (1), 735–746.
- Shang, S., Lee, Z., Shi, L., Lin, G., Wei, G., Li, X., 2016. Changes in water clarity of the Bohai Sea: observations from MODIS. *Remote Sens. Environ.* 186, 22–31.
- Shen, Q., Li, J., Zhang, F., Sun, X., Li, J., Li, W., Zhang, B., 2015. Classification of several optically complex waters in China using *in situ* remote sensing reflectance. *Remote Sens. Basel* 7, 14731–14756.
- Shi, K., Li, Y., Li, L., Lu, H., Song, K., Liu, Z., Xu, Y., Li, Z., 2013. Remote chlorophyll-a estimates for inland waters based on a cluster-based classification. *Sci. Total Environ.* 444, 1–15.
- Shi, K., Zhang, Y., Zhou, Y., Liu, X., Zhu, G., Qin, B., Gao, G., 2017. Long-term MODIS observations of cyanobacterial dynamics in Lake Taihu: responses to nutrient enrichment and meteorological factors. *Sci. Rep.* 7, 40326.
- Shi, K., Zhang, Y., Zhu, G., Liu, X., Zhou, Y., Xu, H., Qin, B., Liu, G., Li, Y., 2015. Long-term remote monitoring of total suspended matter concentration in Lake Taihu using 250 m MODIS-Aqua data. *Remote Sens. Environ.* 164, 43–56.
- Song, K., Li, L., Wang, Z., Liu, D., Zhang, B., Xu, J., Du, J., Li, L., Li, S., Wang, Y., 2012. Retrieval of total suspended matter (TSM) and chlorophyll-a (Chl-a) concentration from remote-sensing data for drinking water resources. *Environ. Monit. Assess.* 184, 1449.
- Song, K.S., Li, L., Tedesco, L., Li, S., Shi, K., Hall, B., 2014. Remote estimation of nutrients for a drinking water source through adaptive modeling. *Water Resour. Manage.* 28, 2563–2581.
- Sun, D., Li, Y., Wang, Q., Le, C., Huang, C., Shi, K., 2011. Development of optical criteria to discriminate various types of highly turbid lake waters. *Hydrobiologia* 669, 83–104.
- Sun, D., Li, Y., Wang, Q., Le, C., Lv, H., Huang, C., Gong, S., 2012. Specific inherent optical quantities of complex turbid inland waters, from the perspective of water classification. *Photochem. Photobiol. Sci.* 11, 1299–1312.
- Sun, D., Hu, C., Qiu, Z., Cannizzaro, J.P., Barnes, B.B., 2014a. Influence of a red band-based water classification approach on chlorophyll algorithms for optically complex estuaries. *Remote Sens. Environ.* 155, 289–302.
- Sun, D., Qiu, Z., Li, Y., Shi, K., Gong, S., 2014b. Detection of total phosphorus concentrations of turbid inland waters using a remote sensing method. *Water Air Soil Pollut.* 225, 1953.
- Tan, J., Cherkauer, K., Chaubey, I., 2016. Developing a comprehensive spectral-biogeochemical database of midwestern rivers for water quality retrieval using remote sensing data: a case study of the Wabash River and its tributary, Indiana. *Remote Sens.*—Basel 8, 517.
- Tian, L., Wai, O.W.H., Chen, X., Li, W., Li, J., Li, W., Zhang, H., 2016. Retrieval of total suspended matter concentration from Gaofen-1 Wide Field Imager (WFI) multi-spectral imagery with the assistance of Terra MODIS in turbid water—case in Deep Bay. *Int. J. Remote Sens.* 37, 3400–3413.
- Vollenweider, R., Rast, W., Kerekes, J., 1980. Phosphorus loading concept and Great Lakes eutrophication. Annual Cornell University Conference, vol. 11, 207–234.
- Wu, C., Wu, J., Qi, J., Zhang, L., Huang, H., Lou, L., Chen, Y., 2010. Empirical estimation of total phosphorus concentration in the mainstream of the Qiantang River in China using Landsat TM data. *Int. J. Remote Sens.* 31, 2309–2324.
- Zhang, T., Fell, F., Liu, Z.S., Preusker, R., Fischer, J., He, M.X., 2003. Evaluating the performance of artificial neural network techniques for pigment retrieval from ocean color in case I waters. *J. Geophys. Res. Atmos.* 108, 307–336.
- Zhao, Y.Q., Xi-Li, W., Jiang, S., 2009. Study on neural network model for Weihe River Water quality retrieving using remote-sensing image. *Remote Sens. Technol. Appl.* 24, 63–67.
- Zheng, Z., Li, Y., Guo, Y., Xu, Y., Liu, G., Du, C., 2015. Landsat-based long-term monitoring of total suspended matter concentration pattern change in the wet season for Dongting Lake, China. *Remote Sens.*—Basel 7, 13975–13999.
- Zheng, Z., Ren, J., Li, Y., Huang, C., Liu, G., Du, C., Lyu, H., 2016. Remote sensing of diffuse attenuation coefficient patterns from Landsat 8 OLI imagery of turbid inland waters: a case study of Dongting Lake. *Sci. Total Environ.* 573, 39–54.
- Zhu, W., Tian, Y.Q., Yu, Q., Becker, B.L., 2013. Using hyperion imagery to monitor the spatial and temporal distribution of colored dissolved organic matter in estuarine and coastal regions. *Remote Sens. Environ.* 134, 342–354.