# Remotely observed variations of reservoir low concentration chromophoric dissolved organic matter and its response to upstream hydrological and meteorological conditions using Sentinel-2 imagery and Gradient Boosting Regression Tree

Zeliang Zhang, Weining Zhu, Jiang Chen and Qian Cheng

## ABSTRACT

Freshwater lakes are facing increasingly serious water quality problems. Remote sensing techniques are effective tools for monitoring spatiotemporal information of chromophoric dissolved organic matter (CDOM), a biochemical indicator for water quality. In this study, the Gradient Boosting Regression Tree (GBRT) model and Sentinel-2A/B imagery were combined to estimate low CDOM concentrations ($0.003\ m^{-1} < a_{CDOM}(440) < 1.787\ m^{-1}$) in Xin'anjiang Reservoir, an important drinking water resource in Zhejiang Province, China, providing the CDOM distributions and dynamics with high spatial (10 m) and temporal (5 day) resolutions. The possible environmental factors that may affect CDOM spatiotemporal patterns and dynamics were analyzed using Sentinel-2 image-observed data in 2018. Results showed that CDOM in the reservoir exhibited a clear increased gradient from its transition and lacustrine zones to the riverine zones, indicating that the rivers carried a substantial load of organic matter to the lake. The precipitation may increase CDOM concentrations but it has a delayed effect, while it may also shortly decrease CDOM concentrations due to the rainwater dilution. We also found that the correlations between CDOM and water temperature, air pressure, and wind speed were very low, indicating that these factors may not have significant impacts on CDOM variations in the reservoir. This study demonstrated that the GBRT model and Sentinel-2 imagery have the potential to accurately monitor CDOM spatiotemporal variations in reservoirs with low CDOM concentrations, which advances our understanding on the relations between the dissolved organic matter and its coupling environmental factors in river-reservoir systems.

**Key words** | CDOM, GBRT, remote sensing, reservoir, Sentinel-2

**Zeliang Zhang**
Department of Marine Science, Ocean College,
Zhejiang University,
Zhoushan,
China

**Weining Zhu** (corresponding author)
Key Laboratory of Ocean Observation-Imaging Testbed of Zhejiang Province, Ocean College,
Zhejiang University,
Zhoushan,
China
and
Institute of Marine Sensing and Network, Department of Marine Information Science, Ocean College,
Zhejiang University,
Zhoushan,
China
E-mail: *zhuwn@zju.edu.cn*; *zhuweining@gmail.com*

**Jiang Chen**
School of Remote Sensing and Information Engineering,
Wuhan University,
Wuhan,
China

**Qian Cheng**
School of Tourism and Urban-Rural Planning,
Zhejiang Gongshang University,
Hangzhou,
China

## HIGHLIGHTS

- Low-concentration reservoir CDOM (chromophoric dissolved organic matter) can be estimated by using Sentinel-2 images and machine learning.
- The GBRT (Gradient Boosting Regression Tree) method performed much better than other traditional and machine learning methods.
- The satellite observed CDOM variations demonstrated some correlations to the upstream hydrological and meteorological conditions of the reservoir.



doi: 10.2166/ws.2020.342

## INTRODUCTION

Freshwater lakes and reservoirs in high population density regions are facing increasingly serious water-quality problems, which are caused by human activities and climate change (Li *et al.* 2017; Mushtaq & Lala 2017). As one of the water quality indicators, chromophoric dissolved organic matter (CDOM) is the biologically active component of dissolved organic matter (DOM). CDOM absorbs up to 90% of the underwater solar radiation in 400–500 nm (Belanger *et al.* 2008), protecting underwater ecosystems from harmful UV radiation exposure (Stedmon *et al.* 2000). However, CDOM also has a few negative effects on the processes of drinking water treatment (Zhang *et al.* 2011), such as reducing the effectiveness of oxidants and disinfectants and producing undesirable disinfection by-products during oxidation processes (Baghoth *et al.* 2011). Understanding the sources, concentration, and cycling of CDOM in freshwaters is important for managing aquatic resources and predicting the outcomes of environmental change (Olmanson *et al.* 2020). As a result, monitoring CDOM in reservoirs and studying its spatiotemporal distribution are very important for effective water quality evaluation and drinking water conservation (Zhang *et al.* 2011).

Compared with the traditional methods of collecting field samples and measuring CDOM's absorption coefficients in the lab using spectrophotometers, remote sensing techniques can monitor CDOM variations over a large scale. The traditional remote sensing models for estimating CDOM include empirical, semi-analytical, spectral matching, matrix inversion, and so on (Kutser *et al.* 2005a; Brezonik *et al.* 2015; Ruescas *et al.* 2018). However, these models do not perform well for complex waters, particularly for those inland waters with low-concentration CDOM (Chen *et al.* 2019). Zhu *et al.* compared 15 CDOM retrieval methods using the data in Lake Huron, and they found that most methods did not provide reliable estimates of CDOM levels, and even those that provided the best estimates tended to yield underestimates at high CDOM levels and overestimates at low CDOM concentrations (Zhu *et al.* 2014). Artificial neural network (ANN) has been also used for CDOM inversion since the last century (Heddam 2014), and it shows potential to deal with complex inland water. Recently, machine learning-based methods have also been used for CDOM inversion and obtained accurate results (Ruescas *et al.* 2018). So far, there are many available machine learning-based methods, such as RFR (Random Forest Regression), Gradient Boosting Regression Tree (GBRT), SVR (Support Vector Machine Regression), and GPR (Gaussian Process Regression) (Ruescas *et al.* 2018; Shah *et al.* 2019), but which one is more suitable for low-concentration CDOM in freshwater lakes/reservoirs remains unknown, and of course we want to determine the best one. We have noticed that recently the GBRT was proposed for machine learning (Deng *et al.* 2019), and it has been proved to be more robust and preferred than the other machine learning methods, such as the Random Forest and Support Vector Machine (Yang *et al.* 2020). GBRT has the advantages that it can fit complex nonlinear relationships but does not need prior data transformation or elimination of the outlier scans (Elith *et al.* 2008). Therefore, in this study, we tested the GBRT model and compared it with the other several machine learning methods, and our results did show that GBRT was the best among the tested models.

Some land-oriented satellite sensors, such as the Landsat series, have fine spatial resolutions (30 m in Landsat 8) and have been demonstrated with accepted accuracy for inland CDOM estimation. Some Landsat 8-based band-ratio models have been proposed for monitoring low CDOM concentrations (the absorption coefficients of CDOM at 440 nm ranges 0.066–1.242 $m^{-1}$) (Chen *et al.* 2019). However, the temporal resolution (16 days) of Landsat series satellites is too low for complex inland waters, which are often highly varied in short periods (Toming *et al.* 2016). In the case of cloud cover on the day of the satellite passing, it is difficult to obtain high-quality Landsat images in a region within several months. Therefore, researchers have paid more attention to Sentinel-2 sensors, which have higher spatial (10 m) and temporal resolutions (5 days) (Chen *et al.* 2017b). Sentinel-2 sensors were widely applied to monitor CDOM concentrations (Toming *et al.* 2016; Xu *et al.* 2018; Al-Kharusi *et al.* 2020). Thus, in this study, Sentinel-2 was selected as image data to estimate CDOM concentrations. The bands and their spatial resolutions of Sentinel-2A/B are shown in Table 1.

**Table 1** | Spectral bands for the Sentinel-2A/B sensors

| | Sentinel-2A | | Sentinel-2B | | |
|---|---|---|---|---|---|
| Band number | Central wavelength (nm) | Bandwidth (nm) | Central wavelength (nm) | Bandwidth (nm) | Spatial resolution (m) |
| 1 | 443.9 | 27 | 442.3 | 45 | 60 |
| 2 | 496.6 | 98 | 492.1 | 98 | 10 |
| 3 | 560.0 | 45 | 559 | 46 | 10 |
| 4 | 664.5 | 38 | 665 | 39 | 10 |
| 5 | 703.9 | 19 | 703.8 | 20 | 20 |
| 6 | 740.2 | 18 | 739.1 | 18 | 20 |
| 7 | 782.5 | 28 | 779.7 | 28 | 20 |
| 8 | 835.1 | 145 | 833 | 133 | 10 |
| 8a | 864.8 | 33 | 864 | 32 | 20 |
| 9 | 945.0 | 26 | 943.2 | 27 | 60 |
| 10 | 1,373.7 | 75 | 1,376.9 | 76 | 60 |
| 11 | 1,613.7 | 143 | 1,610.4 | 141 | 20 |
| 12 | 2,202.4 | 242 | 2,185.7 | 238 | 20 |

The goals of this study are: (1) exploring the ability of the GBRT model and Sentinel-2 images to retrieve low-concentration CDOM in Xin'anjiang Reservoir; (2) identifying the possible environmental factors that may affect CDOM spatiotemporal patterns of the reservoir. To achieve the goals, field measurements were conducted in Xin'anjiang Reservoir from April to October 2018. The Sentinel-2 image-observed results demonstrated the CDOM spatiotemporal variations, and hydrological and meteorological data of the reservoir in 2018 were used to study the couplings between CDOM and water temperature, river flow, rainfall, and other environmental factors.

## DATA AND METHODS

### Study site

Xin'anjiang Reservoir (also known as Lake Qiandaohu) is in the western part of Zhejiang Province and the southern part of Anhui Province, China (Figure 1). It is the third major river system in Anhui province and its discharge makes the largest river, Xin'anjiang River (or Qiantang River), in Zhejiang province (Wang *et al.* 2012). This reservoir has a water area of 580 $km^2$, 1,078 islands with each area larger than 2,500 $m^2$, a maximal water volume of $178.4 \times 10^8$ $m^3$, and a basin area of 10,480 $km^2$ (Wu *et al.* 2015). The recorded mean annual air temperature is 17 °C, and the mean annual precipitation is 1,637 mm (1961–2014). Due to its high water clarity and good water quality, Xin'anjiang Reservoir is a key drinking water source for China's Yangtze River Delta Region, serving a surrounding population of at least ten million (Xin *et al.* 2007). However, short-term algal blooms have appeared in the lake since the 1990s (Zhang *et al.* 2014), making its water quality worse periodically.

According to its geographical and ecological function, the Xin'anjiang Reservoir consists of a riverine zone, a transition zone and a lacustrine zone (Liu *et al.* 2014). The concentration ranges of chlorophyll, total suspended matter and total phosphorus in Xin'anjiang Reservoir in October 2018 were 2.402–13.165 μg/L, 0.476–3.430 mg/L, and 0.009–0.112 mg/L, respectively (Zeng *et al.* 2020). In previous studies, the reservoir was divided into five sub-regions to reveal nutrient status in different aquatic environments (Li *et al.* 2017). In this study, to better analyze temporal and spatial variations of CDOM, we also divided the reservoir into the northeastern (NE, the riverine zone), northwestern (NW, the riverine zone), southwestern (SW, the riverine zone), central (C, the transition zone) and southeastern areas (SE, the lacustrine zone), see Figure 1.

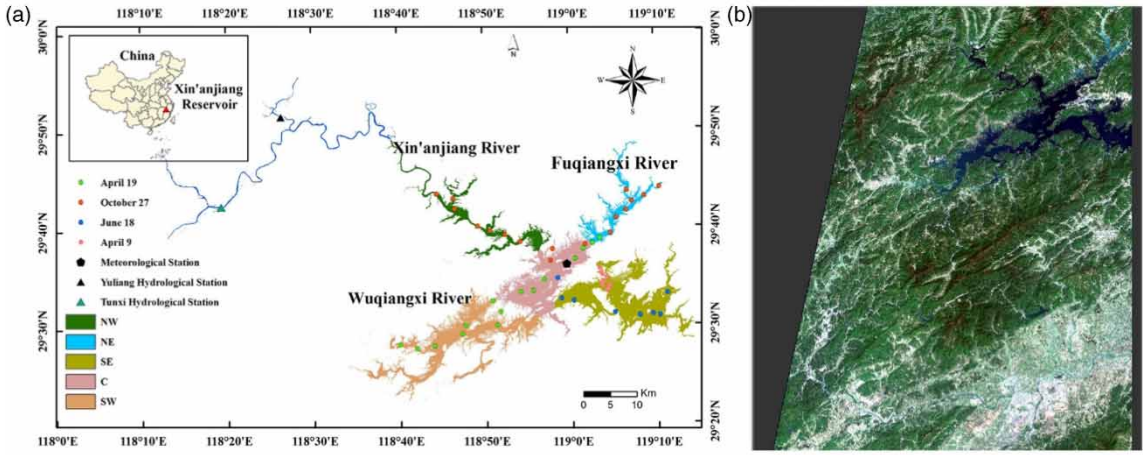


**Figure 1** | (a) Study site map showing the field sampling points, locations of meteorological stations and hydrological stations, three main inflow rivers, and five sub-regions (C, NE, NW, SE, and SW) in Xin'anjiang Reservoir. (b) A Sentinel-2 raw image acquired on April 9 in 2018, showing the land cover in the watershed of the Xin'anjiang Reservoir.

## Field measurement

Four sampling field trips were conducted in Xin'anjiang Reservoir from April to October in 2018, covering the seasons of spring (twice in April), summer (June), and autumn (October) in 2018 (see Figure 1), and in total 50 samples were obtained (see Table 1). Water samples were collected at the surface (about 10 cm below) using a standard water-fetching bottle, and then preserved in amber bottles (polypropylene 250 mL) at ambient water temperatures before being delivered to the laboratory for analysis within 24 h, measuring some water quality parameters such as the concentrations of CDOM, COD, TP, suspended substances, and chlorophyll-a. At each sampling location, the above-surface water spectra were measured using an ASD (Analytical Spectral Devices, Inc.) FieldSpec® spectroradiometer (wavelength range 325–1,075 nm with 1 nm interval) (ASD User Guide, https://www.malvernpanalytical.com). Three in-situ optical properties were measured: the above-surface radiance ($L_t$), sky radiance ($L_i$), and downwelling irradiance ($E_d$) (Zhu *et al.* 2014). To minimize uncertainties, ten spectra were measured and the median one was selected for calculating the remote sensing reflectance ($R_{rs}$ ($\lambda$), $sr^{-1}$):

$$R_{rs}(\lambda) = \frac{L_t - \rho L_i}{E_d} \tag{1}$$

where the surface reflectance factor $\rho = 0.028$, according to Mobley 1999.

In the laboratory, water samples were filtered through a GF/F glass microfiber membrane (0.45 $\mu$m) under low pressure (<5 atm). The filtered sample was used to measure CDOM absorption coefficients $a_{\mathrm{CDOM}}(\lambda)$ with a Cary-100 spectrophotometer (Cary-100 User Guide, https://www.agilent.com/cs/library/usermanuals/public/1972_7000.pdf) with a Milli-Q baseline correction. $a_{\mathrm{CDOM}}(\lambda)$ can be calculated by the following equation:

$$a_{CDOM}(\lambda) = A(\lambda)\frac{ln10}{Pathlength} = A(\lambda) \times 230.3 \tag{2}$$

where $A(\lambda)$ is the Cary-measured CDOM absorbance and the pathlength is the path length (1 cm) of the used cuvette. In this study, CDOM concentrations are parameterized by its absorption coefficient at 440 nm (Zhu *et al.* 2014).

In order to use Sentinel-2A/B data to derive $a_{\mathrm{CDOM}}(440)$, the field spectra of $R_{rs}$ were resampled to Sentinel-2 bands using its spectral response function (Chen *et al.* 2017b).

$$R_{rs}(B_k) = \frac{\int_{\lambda_m}^{\lambda_n} RSR(\lambda) \times R_{rs\_field}(\lambda) d(\lambda)}{\int_{\lambda_m}^{\lambda_n} RSR(\lambda) d(\lambda)} \tag{3}$$

where $R_{rs}(B_k)$ is the simulated $R_{rs}$ for the $k$-th ($k = 2, 3, 4$) band of Sentinel-2A/B, which is summed from $\lambda_m$ to $\lambda_n$ for the $k$-th band, and $R_{\mathrm{rs\_field}}$ are the field measured spectra.

### Hydrological and meteorological data

The monthly flow data of Xin'anjiang River in 2018 was collected from two hydrological stations called Tunxi and Yuliang (Figure 1), from the Taihu basin hydrological information service system (http://data.cma.cn/). Meteorological data, including rainfall, air temperature, wind speed, and humidity were obtained at the Chun'an meteorological station (Figure 1) and downloaded from the China Meteorological Data-Sharing Service System.

### Sentinel-2A/B acquisition and preprocessing

The combined Sentinel-2A and Sentinel-2B can provide global coverage of the Earth's land surface for every 5 days with 10 m, 20 m, and 60 m spatial resolutions. A total of 13 cloud-free Sentinel-2 images were collected in 2018 to develop the CDOM retrieval algorithm and analyze its spatiotemporal variations in Xin'anjiang Reservoir. Since cloudless images cannot be obtained for September 2018, the image acquired at the closest time (October 1) was used instead. Two quasi-synchronous Sentinel-2 images with two field measurements (on 9 April and 18 June 2018) were used to validate the GBRT model. The FLAASH (Atmospheric Analysis of Spectral Hypercubes) module in ENVI 5.5 (Sentinel-2 atmospheric correction function was added in the latest version) was used for atmospheric correction, which gives the earth surface reflectance ($R_t$). Then, a $5\times5$ mean filtering was used to reduce the image uncertainty. At last, the remote sensing reflectance ($R_{rs}$, $sr^{-1}$) was calculated by

$$R_{rs}(\theta, \varphi) = \frac{R_t}{\pi} - \frac{L_r(\theta, \varphi)}{E_d} \tag{4}$$

where $L_r(\theta, \varphi)$ is the surface radiance at zenith angle $\theta$ and azimuth angle $\varphi$; $E_d$ is the downwelling irradiance. The two variables were estimated using the Hydrolight (Ver. 5.0), a well-known radiative transfer model developed by Curtis Mobley.

### Gradient Boosting Regression Tree

As a non-linear model, GBRT can fit complex nonlinear relationships and does not need prior data transformation or elimination of the outlier scans (Elith *et al.* 2008). The GBRT model uses the CART tree as the weak classifier and requires multiple iterations. The newly generated regression tree would fit the error of the previous tree in each iteration, which is the biggest difference from the Random Forest. In Random Forest, multiple regression trees are independent, and the training results of different trees are not further optimized. The gradient descent method is used to move toward the negative gradient of loss function in each iteration, which makes the loss function decline. In general, each iteration of the GBRT model produces a weak classifier, whose accuracy is not high, but integration of weak classifiers can achieve higher accuracy (Natekin & Knoll 2013; Rokach 2016; Kuang *et al.* 2018). The weighted sum of the prediction results of each regression tree is the predicted value, see Figure 2.

The GBRT model is expressed as follows:

$$F_m(\mathrm{x}) = \sum_{i=1}^{n} \theta_i f_i(x) \tag{5}$$

where $n$ is the number of weak learners, $\theta$ is the coefficient (reducing the over fitting), $f$ is the weak learner, and $F$ is the final general model. For more details, please refer to Zhao *et al.* 2019.

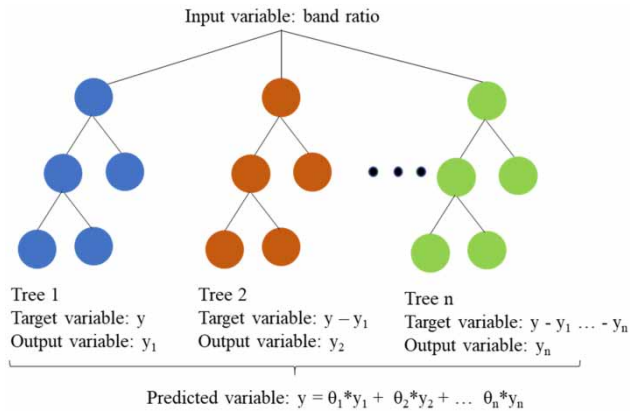


**Figure 2** | The construction process of GBRT model.

There are several important parameters of the GBRT model: (a) the max depth of each weak learner (generally no more than five), (b) The maximum quantity of weak learners, and (c) the learning rate (higher learning rate means a stronger correction and makes the model more complex). Thus, we need to choose appropriate parameters for improving the performance of the model.

### Other alternative machine learning algorithms

Machine-learning methods have been used to simulate the nonlinear relationships between CDOM concentration and remote sensing reflectance. They are mainly classified into tree-based algorithms, neural networks, and kernel methods. Many previous studies have confirmed that neural networks can predict CDOM absorption from ocean color (Kishino *et al.* 2005; Chen *et al.* 2017a), but it requires large training data and possibly produces the risk of over-fitting and being trapped in a local minimum (Zhan *et al.* 2001). In this work, four nonlinear regression machine-learning algorithms were tested and compared: RFR (Random Forest Regression), GBRT, SVR (Support Vector Machine Regression), and GPR (Gaussian Process Regression). RFR and GBRT are representatives of decision families, while SVR and GPR are the representatives of kernel methods.

## ALGORITHM DEVELOPMENT

### Algorithm architecture

The band-ratio algorithms involving $R_{rs}$ in blue, green, and red domains have been widely used for CDOM remote sensing in freshwater lakes (Kutser *et al.* 2005b; Mannino *et al.* 2014; Joshi *et al.* 2017). In addition, the band-ratio models can reduce more uncertainty of atmospheric correction than using single-band reflectance (Stramska & Stramski 2005; Cherukuru *et al.* 2016). In this study, all Sentinel-2 band-ratios combined among B2($R_{rs}$490), B3($R_{rs}$560), and B4($R_{rs}$665) were tested as the inputs of the GBRT model and we determined the optimal combinations in terms of four indicators: coefficient of determination ($R^2$), root mean squared error (RMSE), Bias and mean absolute percentage error (MAPE).

The field-measured 50 samples were randomly divided into two independent datasets with a sample proportion of approximately 3:1, a common ratio used in machine learning (Shah *et al.* 2019). Therefore, the GBRT's training and validation datasets contain 37 and 13 field measured samples, respectively. We also used the 16 samples collected in April 19 and June 18 for image validation because the Sentinel-2 also acquired the images in the two days. The image-derived spectra were input into the GBRT model and then the model outputs were validated by the 16 field-measured CDOM concentrations. Note that during the two field measurements we collected 23 samples, but seven were not used due to the cloud cover in the images. The optimal number of base learners is 50. The learning rate is 0.1. The max depth of each learner is 5 and other parameters are configured by the model default.

### Comparison to other CDOM retrieval algorithms

In order to test the advantages of GBRT for estimating low-concentration CDOM in a freshwater environment, two previous traditional models were also compared. A logarithmic model (see Equation (6)), was developed based on Landsat 8 in lakes and rivers in Minnesota, where $a_{\mathrm{CDOM}}(440)$ ranges from $0.51\,\mathrm{m}^{-1}$ to $25.1\,\mathrm{m}^{-1}$ (Brezonik *et al.* 2015). Another model was a unary quadratic polynomial Landsat-8 model (see Equation (7)), in a Brazil reservoir, where $a_{\mathrm{CDOM}}(440)$ ranges from $0.644\,\mathrm{m}^{-1}$ to $1.413\,\mathrm{m}^{-1}$ (Alcantara *et al.* 2016). The second model is more comparable with our model in terms of water type and CDOM range.

$$a_{\mathrm{CDOM}}(440) = A\ln\left(\frac{B3}{B4}\right) + B \tag{6}$$

$$a_{\mathrm{CDOM}}(440) = A\left(\frac{B4}{B2}\right)^2 + B\left(\frac{B4}{B2}\right) + C \tag{7}$$

where $B2$, $B3$, and $B4$ are the bands of Landsat 8. The parameters $A$, $B$, and $C$ need to be re-calibrated for different study sites.

**Table 2** | Descriptive statistics of the field measured $a_{CDOM}(440)$ in Xin'anjiang Reservoir, 2018

| Date | Min | Max | Mean | SD | Samples |
|---|---|---|---|---|---|
| Apr. 9 | 0.248 | 0.994 | 0.609 | 0.291 | 10 |
| Apr. 19 | 0.103 | 0.506 | 0.252 | 0.116 | 15 |
| June 18 | 0.597 | 1.162 | 0.796 | 0.196 | 8 |
| Oct. 27 | 0.710 | 1.788 | 1.142 | 0.339 | 17 |
| All | 0.103 | 1.788 | 0.713 | 0.442 | 50 |

## RESULTS AND DISCUSSION

### Field measured CDOM and above-surface spectra

In many previous study sites, CDOM concentrations in inland waters were often found in wide ranges (Griffin *et al.* 2011; Kutser 2012), for example, $a_{CDOM}(440)$ ranged from 0.6 to 19.4 $m^{-1}$ in 15 Minnesota lakes (Brezonik *et al.* 2005). Comparing with these CDOM highly-varied lakes, CDOM concentrations in Xin'anjiang Reservoir were relatively low in 2018. The field measured $a_{CDOM}(440)$ ranged from 0.1 to 1.78 $m^{-1}$ with mean 0.7 $m^{-1}$. Our field measured data showed that there were also seasonal variations: the measured $a_{CDOM}(440)$ in April was within 0.1–0.99 $m^{-1}$ (mean 0.39 $m^{-1}$), but in October the range changed to 0.7–1.8 $m^{-1}$ (mean 1.14 $m^{-1}$), demonstrating that CDOM concentrations were higher in autumn than in the early spring. The measured CDOM concentrations during our four field trips are shown in Table 2.

The measured remote sensing reflectance of the Xin'anjiang Reservoir is shown in Figure 3, showing the typical spectral signatures of complex inland freshwater (Toming *et al.* 2016; Xu *et al.* 2018). The low absorptions and scatterings of chlorophyll and carotene in phytoplankton produced the peaks at 570 nm (Gurlin *et al.* 2011; Liu *et al.* 2011). The small reflection valleys at 670 nm may be caused by maximum absorptions of chlorophyll-a (Beatriz Juarez *et al.* 2008). The chlorophyll fluorescence at around 700 nm, caused by the minimum absorptions of algal pigments and pure water, were not remarkable in Xin'anjiang Reservoir (Chen *et al.* 2020).

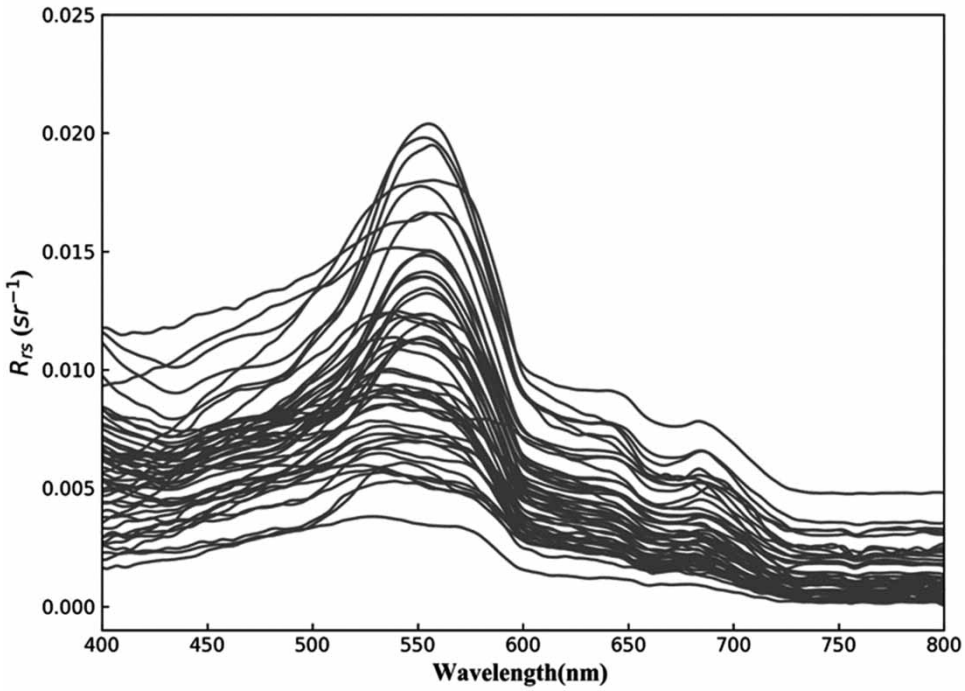


**Figure 3** | The field measured above-surface spectra ($R_{rs}$) in Xin'anjiang Reservoir.

## Model assessment and comparison

To determine the optimal inputs of the GBRT model, all the possible band-ratio combinations were examined and the statistical results are shown in Table 3. Over-fitting occurred when B2/B3 and B4/B3 were used as inputs, which led the model to perform poorly on the validation dataset. In terms of four indicators ($R^2$, *RMSE*, *Bias*, and *MAPE*), B2/B3 and B4/B2 were selected to retrieve CDOM concentrations. The best GBRT model yielded $R^2 = 0.92$ and 0.95, $MAPE = 0.3$ and 18%, $RMSE = 0.16$ and $0.13\ m^{-1}$, and $Bias = -0.0015$ and $-0.01\ m^{-1}$ for the training and validation datasets, respectively. The GBRT-model-retrieved CDOM were compared with the field measured CDOM (see results in Figure 4(a) and 4(b)). Sixteen samples out of the training and validation datasets were further compared with the derived $a_{CDOM}(440)$ from two quasi-synchronous (within 3 hours) Sentinel-2 images (see Figure 3(c)). The B2/B3 and B4/B2 algorithms applied on Sentinel-2 images produced acceptable accuracy ($R^2 = 0.77$, $RMSE = 0.1\ m^{-1}$, $Bias = 0.062\ m^{-1}$ and $MAPE = 19.0\%$), indicating that the Sentinel-2 GBRT model had acceptable accuracy for retrieving CDOM concentrations in Xin'anjiang Reservoir. When the CDOM concentrations at lower levels

**Table 3** | Modeling $a_{CDOM}(440)$ using different Sentinel-2 band combinations

| Input | Dataset | $R^2$ | MAPE | RMSE | Bias |
|---|---|---|---|---|---|
| B2/B3, B4/B2 | training | 0.92 | 0.3% | 0.16 | −0.0015 |
| | validation | 0.95 | 18% | 0.13 | −0.0121 |
| B2/B3, B4/B3 | training | 0.97 | 0.01% | 0.01 | 0.0013 |
| | validation | 0.06 | 72% | 0.48 | −0.1026 |
| B4/B3, B4/B2 | training | 0.78 | 0.59% | 0.28 | −0.0047 |
| | validation | 0.61 | 46% | 0.29 | −0.0172 |
| B2/B3, B4/B2, B4/B3 | training | 0.910 | 0.44% | 0.21 | −0.0034 |
| | validation | 0.71 | 36% | 0.24 | −0.0058 |

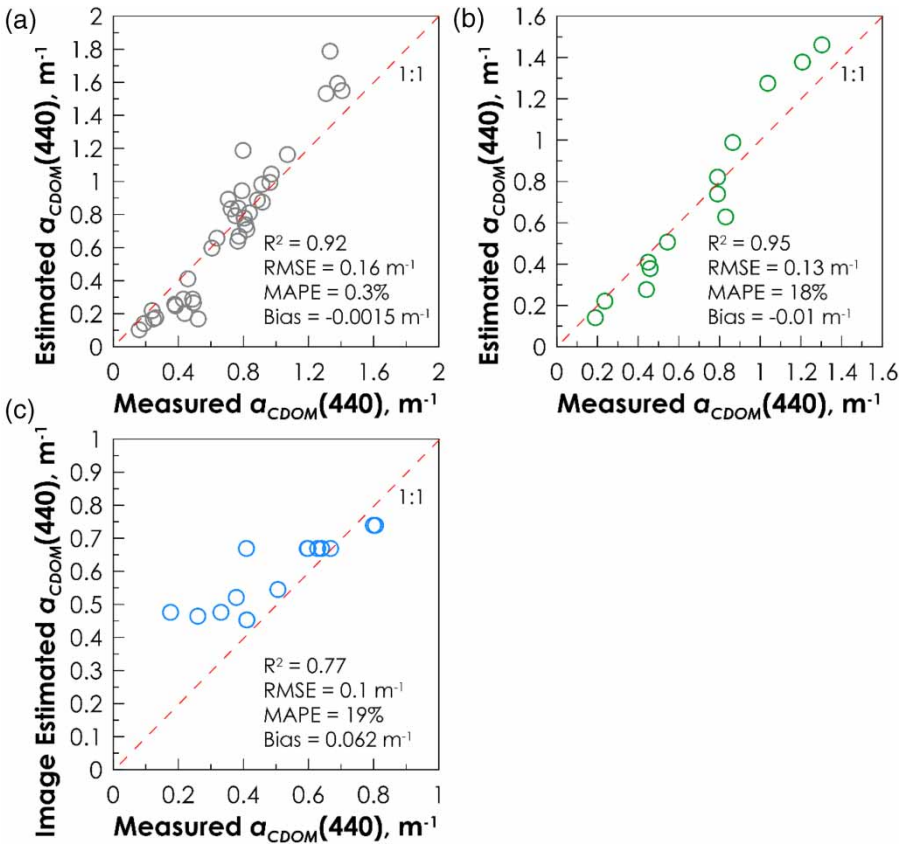


**Figure 4** | Estimating CDOM based on a GBRT model and Sentinel-2 band ratios B2/B3 and B4/B2: the field-measured versus the model-estimated CDOM for (a) the training dataset and (b) the validation dataset, and (c) validating the model using 16 field samples, which were collected at the same time (within 3 hours) of two Sentinel-2 images.

($<0.37\ m^{-1}$), there is still a certain overestimation. The extremely low CDOM concentrations will become weaker to the spectral response, which means that CDOM absorbance at low wavelengths may be overridden by Chl-a and SPM absorptions. At the same time, the weaker CDOM absorbance signal transmitted to the satellite image may be seriously affected by the noise.

In this study, two models developed from previous studies were tested with our situ-data in Xin'anjiang Reservoir. We found that the logarithmic model ($R^2 = 0.0009$, $RMSE = 0.4578\ m^{-1}$, see Figure 5(a)) and the polynomial model ($R^2 = 0.1125$, $RMSE = 0.4389\ m^{-1}$, see Figure 5(b)) are not able to retrieve low CDOM concentrations in Xin'anjiang Reservoir. The relation between low CDOM concentrations and $R_{rs}$ in Xin'anjiang Reservoir may be more complex. Simple logarithmic and polynomial models did not achieve a good fit in Xin'anjiang Reservoir. Whether the GBRT model can be generalized to other low-concentration reservoirs requires more data in support. We also compared the GBRT with the Random Forest (RF), Support Vector Machine (SVM), and Gaussian Process (GP) regression models. Band-ratios B2/B3 and B4/B3 were selected as inputs in all models. Results from the statistical analysis are shown in Table 4. Of the kernel methods, SVM did not perform well for either the training ($R^2 = 0.36$, MAPE = 70.1%) or the validation datasets ($R^2 = 0.25$, MAPE = 65.7%), and GP produced over-fitting for the validation dataset ($R^2 = 0$, MAPE = 82.0%). Of the decision families, compared to the RF model, GBRT's $R^2$ increased 25% while its MAPE decreased 62.1% for the validation dataset.

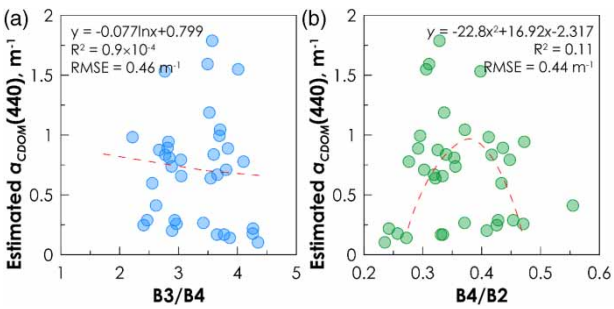


**Figure 5** | Testing two previous empirical models to retrieve the low CDOM concentrations in Xin'anjiang Reservoir.

**Table 4** | Comparison of four representative machine learning algorithms

| Algorithm | Dataset | $R^2$ | MAPE | RMSE | Bias |
|---|---|---|---|---|---|
| SVR | Training | 0.36 | 70.1% | 0.37 | −0.031 |
| | Validation | 0.25 | 65.7% | 0.39 | −0.047 |
| GRR | Training | 0.99 | 0.10% | 0.01 | 0.002 |
| | Validation | 0.02 | 82.0% | 0.77 | −0.51 |
| RFR | Training | 0.50 | 47.3% | 0.22 | 0.006 |
| | Validation | 0.76 | 47.5% | 0.31 | 0.002 |
| GBRT | Training | 0.92 | 0.3% | 0.16 | −0.002 |
| | Validation | 0.95 | 18% | 0.13 | −0.01 |

## Spatiotemporal variations of CDOM

The trained GBRT model was applied to 12 Sentinel-2 images in 2018 to estimate CDOM concentrations in Xin'anjiang Reservoir (see results in Figure 6). To further analyze CDOM's spatiotemporal variations, monthly mean $a_{CDOM}$(440) in five sub-regions were estimated from the valid pixels in each region (Figure 7). The results showed that estimated monthly mean CDOM concentrations of the entire reservoir in 2018 were low, with the range 0.07–1.90 $m^{-1}$ and mean 0.62 $m^{-1}$. Estimated monthly mean CDOM concentrations in different sub-regions demonstrated similar variation trends (see Figure 8(a)). They fluctuated from January to April (0.51–0.61 $m^{-1}$), increased from May (0.57 $m^{-1}$) to October (0.76 $m^{-1}$), and became lower in November (0.55 $m^{-1}$) and December (0.54 $m^{-1}$). The satellite observed CDOM's seasonal variations in Xin'anjiang Reservoir were consistent with some previous observations in Mälaren boreal lakes (Kutser 2012). In Xin'anjiang Reservoir, CDOM concentrations dropped significantly in August (0.59 $m^{-1}$), which may be caused by the heavy rainfall that occurred shortly before the image acquisition time. This phenomenon can be interpreted as being the consequence of dilution due to a large amount of freshwater brought by the rainfall (Shi *et al.* 2018; Lyu *et al.* 2020). After the algal bloom, it was found that CDOM concentrations increased sharply, indicating that phytoplankton degradation was likely to be a major source of CDOM under bloom conditions (Zhao *et al.* 2009). In summer, there is a pressing need to monitor and control the water quality of Xin'anjiang Reservoir to prevent algae outbreaks. The image-observed highest CDOM concentrations in autumn 2018 (see the September and October images in Figure 5) were consistent

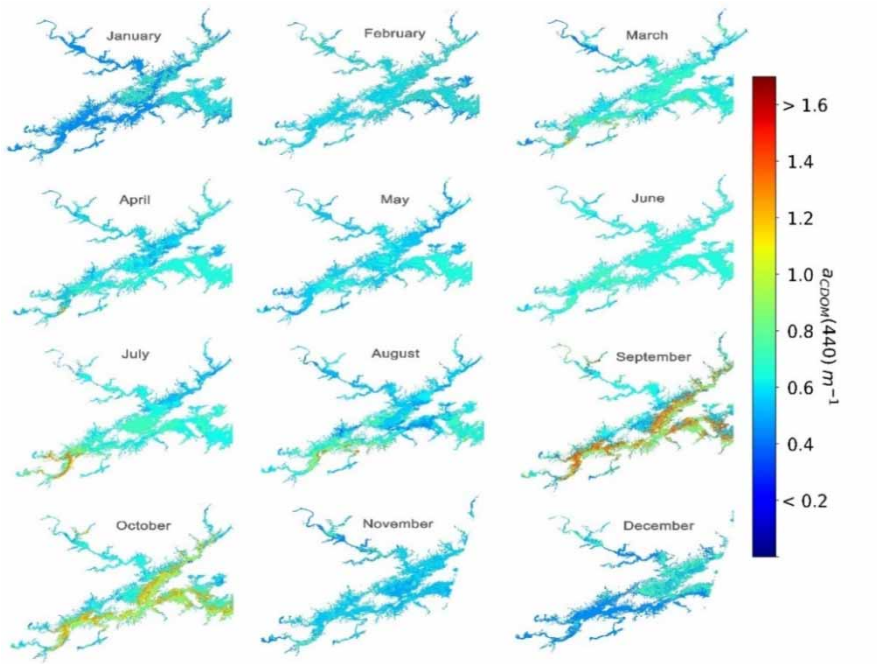


**Figure 6** | Monthly estimated CDOM concentrations observed from Sentinel-2 images in Xin'anjiang Reservoir in 2018.

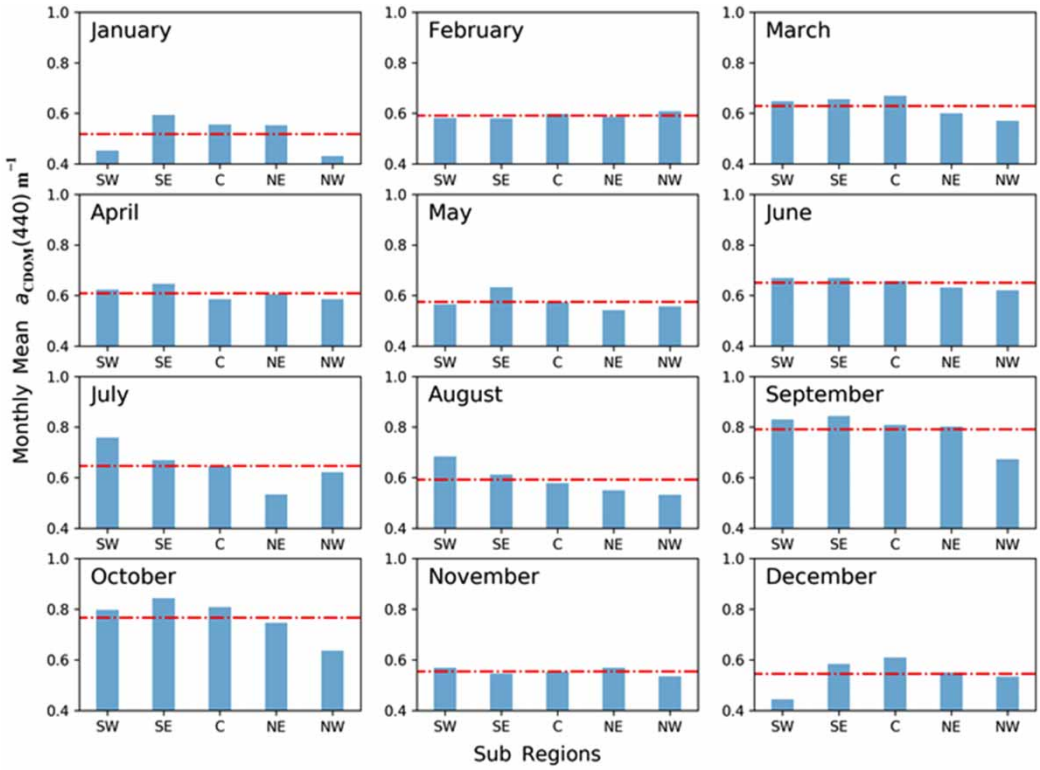


**Figure 7** | Estimated monthly mean $a_{CDOM}(440)$ derived from Sentinel-2 images in five sub-regions of Xin'anjiang Reservoir. The red lines represent the mean values of the entire reservoir. All subfigures have the same labels and axes as the subfigure of January.

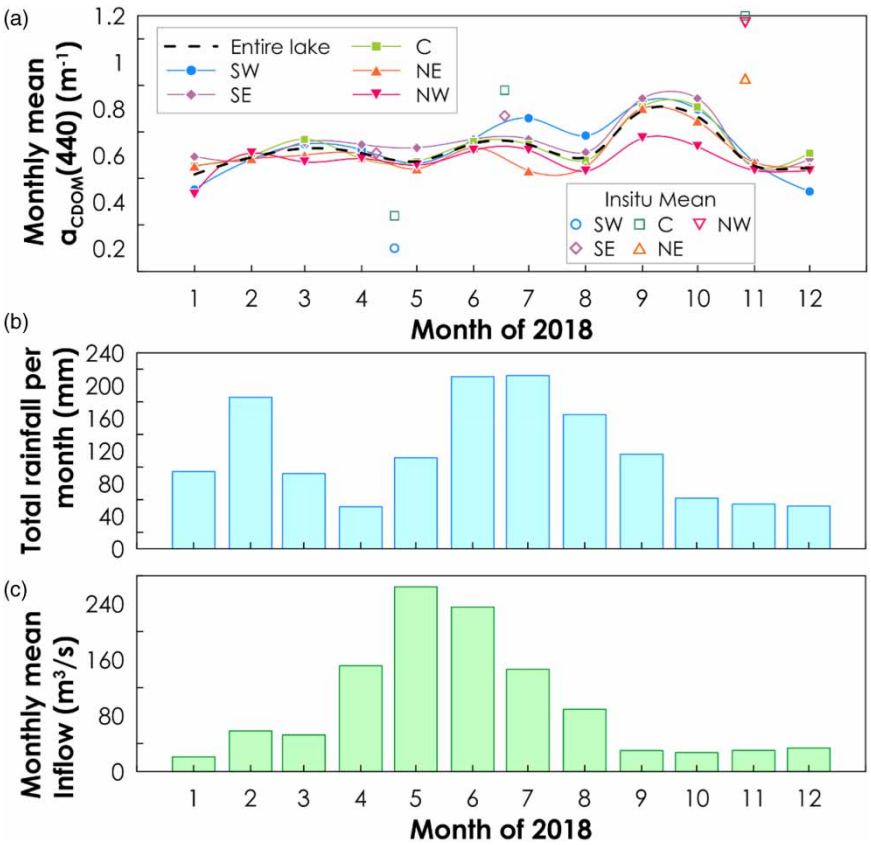


**Figure 8** | (a) The image-estimated and field measured CDOM temporal variations (smoothed) in five sub-regions as well as the entire Xin'anjiang Reservoir, (b) monthly total precipitation in Xin'anjiang Reservoir in 2018, and (c) monthly mean inflow rate from the Xin'anjiang River in 2018.

with the field measured results (see Table 2). The CDOM-rich water presented during the autumn was possibly caused by the occurrence of the algal bloom (Zhang *et al.* 2014).

Based on the image-observed CDOM distribution in Xin'anjiang Reservoir, its monthly estimated average $a_{CDOM}(440)$ in five sub-regions decreased in the following order: SE $0.656 \pm 0.132\ m^{-1}$, Central $0.636 \pm 0.124\ m^{-1}$, SW $0.635 \pm 0.156\ m^{-1}$, NE $0.605 \pm 0.151\ m^{-1}$, and NW $0.575 \pm 0.113\ m^{-1}$. CDOM concentrations in SE (lacustrine zone) and C (transition zone) regions were typically higher than those in the NW, NE, and SW regions (riverine zone). It is known that the watershed runoff and rainfall would bring aquatic CDOM and terrestrial humic-like substances into the reservoir (Zhang *et al.* 2011; Zhou *et al.* 2016), causing a large amount of CDOM to be carried to and pooled in lacustrine and transition zones through three rivers (Xin'anjiang River, Fuqiangxi River, and Wuqiangxi River), which then resulted in relatively high CDOM concentrations in lacustrine and transition zones.

## Relations between CDOM and hydrological and meteorological factors

Some previous studies have indicated that the increased monthly mean precipitation and inflow rate may increase lacustrine CDOM concentrations (Zhou *et al.* 2016). However, in this study, CDOM concentrations in Xin'anjiang Reservoir peaked in September ($0.79\ m^{-1}$) and October ($0.76\ m^{-1}$) 2018, not in the typical rainy months from May to July (see Figure 8(a)–8(c)). It is known that photochemical degradation and photobleaching are important CDOM removal mechanisms (Moran *et al.* 2000; Tzortziou *et al.* 2007; Helms *et al.* 2008; Yoshida *et al.* 2018). For example, a 12-day experiment conducted in Lake Taihu found that CDOM absorptions $a_{CDOM}(355)$ and $a_{CDOM}(280)$ decreased

by 29.8% and 20.8% when CDOM was exposed to natural solar radiation (Zhang *et al.* 2009). In 2018, Xin'anjiang Reservoir experienced high-intensity UV-B radiation during the hot and wet season (from June to August) that would significantly increase photobleaching of the surface water and thereby partly remove CDOM introduced by the precipitation and inflow. In addition, the precipitation also has a short-term effect on CDOM, in that a large amount of rainwater may dilute the CDOM already in lake water (Zhou *et al.* 2015). Therefore, CDOM concentrations estimated in Xin'anjing Reservoir were not too high during the rainy season in 2018. In comparison, the higher CDOM concentrations in September (0.79 $m^{-1}$) and October (0.76 $m^{-1}$) may be caused by several reasons: (a) after the high-precipitation months, surface runoff and groundwater with abundant soil organic matter still gradually released CDOM into the water, (b) the solar radiation became weaker and hence slowed down the photobleaching of CDOM, and (c) during the autumn, more and more deciduous leaves decayed and directly increased the organic matter content in soil and water. Therefore, the Xin'anjiang CDOM peaks in September and October were actually the lag effect of the rainfall during the wet seasons. The delay effect of the rainfall can be also observed in the image-derived CDOM concentrations in March and April, which may be caused by the relatively large rainfall in February.

The correlations between the satellite-derived CDOM concentrations and hydrological and meteorological factors (temperature, air pressure, and wind) were analyzed. The $R^2$ between CDOM concentrations and water temperature, air pressure, and wind speed were 0.25, 0.001 and 0.008 respectively, indicating that these factors may not have significant impacts on CDOM variations in the Xin'anjiang Reservoir.

### Model uncertainty

Due to the very large area of the lake and limited sampling cost, it was difficult to complete a full-coverage sampling of the entire lake within a couple of days and then repeat such full sampling at different seasons, so the full-coverage data were aggregated from the different sub-regions and times. The inversion model based on such aggregated data may work well for the entire lake over a year-long period, but may cause uncertainty for a specific sub-region at a specific time. The aggregated model tends to highlight the global status but eliminate the regional details. Figure 8(a) shows the image-estimated and the field-measured mean CDOM concentrations over the five sub-regions in different months. For the entire lake and some sub-regions (SE and C on 9 April and 18 June), the estimated and measured mean values matched very well within 0.4–0.8 $m^{-1}$; however, in some regions (SW, C, NE, and NW on 19 April and 27 October), some relatively small (0.2–0.4 $m^{-1}$) and large (0.8–1.2 $m^{-1}$) observations were 'averaged' by the aggregated model, hence introducing uncertainties. We suggest that for estimating the regional details, region-oriented specific models should be separately made.

## CONCLUSIONS

A remote sensing model was developed to estimate low-concentration CDOM in Xin'anjiang Reservoir, China, using a GBRT algorithm and Sentinel-2 images (two-band ratios, blue/green and red/blue as the input). By the image validation, the model performance was acceptable with accuracy $RMSE = 0.1\ m^{-1}$ and $MAPE = 19.0\%$. The GBRT model was then applied to observe CDOM variations in Xin'anjiang Reservoir in 2018. CDOM exhibited a clear increased gradient from the transition and lacustrine zones to the riverine zones, indicating that the rivers carried a substantial load of organic matter to the reservoir. In 2018, the highest CDOM concentrations were observed in September (0.79 $m^{-1}$) and October (0.76 $m^{-1}$) rather than in high precipitation and inflow rate months. We also found that the correlations between CDOM concentrations and water temperature, air pressure, and wind speed were very low ($R^2 = 0.25$, 0.001 and 0.008 respectively), indicating that these factors may not have significant impacts on CDOM variations in Xin'anjiang Reservoir.

Since the field measurement and observations were only conducted within one year, more environmental factors and events may largely or occasionally change CDOM and water quality scenarios, and hence the GBRT model is still subject to be further improved in future. Nevertheless, this study demonstrated that the GBRT model and Sentinel-2 imagery have the potential to accurately monitor CDOM spatiotemporal variations in reservoirs of low concentrations, and

hence advancing our understanding on the relations between the dissolved organic matter and its coupling environmental factors in river, lake, and reservoir systems.

## ACKNOWLEDGEMENTS

This study was supported by the National Key R&D Program of China (2017YFB0503902) and the National Natural Science Foundation of China (No. 41971373, 41876031).

## DATA AVAILABILITY STATEMENT

Data cannot be made publicly available; readers should contact the corresponding author for details.

## REFERENCES

Alcantara, E., Bernardo, N., Watanabe, F., Rodrigues, T., Rotta, L., Carmo, A., Shimabukuro, M., Goncalves, S. & Imai, N. 2016 Estimating the CDOM absorption coefficient in tropical inland waters using OLI/Landsat-8 images. *Remote Sensing Letters* **7**, 661–670.

Al-Kharusi, E. S., Tenenbaum, D. E., Abdi, A. M., Kutser, T., Karlsson, J., Bergstrom, A. K. & Berggren, M. 2020 Large-scale retrieval of coloured dissolved organic matter in northern lakes using Sentinel-2 data. *Remote Sensing* **12**, 16.

Baghoth, S. A., Sharma, S. K. & Amy, G. L. 2011 Tracking natural organic matter (NOM) in a drinking water treatment plant using fluorescence excitation-emission matrices and PARAFAC. *Water Research* **45**, 797–809.

Beatriz Juarez, A., Barsanti, L., Passarelli, V., Evangelista, V., Vesentini, N., Conforti, V. & Gualtieri, P. 2008 In vivo microspectroscopy monitoring of chromium effects on the photosynthetic and photoreceptive apparatus of *Eudorina unicocca* and *Chlorella kessleri*. *Journal of Environmental Monitoring* **10**, 1313–1318.

Belanger, S., Babin, M. & Larouche, P. 2008 An empirical ocean color algorithm for estimating the contribution of chromophoric dissolved organic matter to total light absorption in optically complex waters. *Journal of Geophysical Research-Oceans* **113**, 14.

Brezonik, P., Menken, K. D. & Bauer, M. 2005 Landsat-based remote sensing of lake water quality characteristics, including chlorophyll and colored dissolved organic matter (CDOM). *Lake and Reservoir Management* **21**, 373–382.

Brezonik, P. L., Olmanson, L. G., Finlay, J. C. & Bauer, M. E. 2015 Factors affecting the measurement of CDOM by remote sensing of optically complex inland waters. *Remote Sensing of Environment* **157**, 199–215.

Chen, J., He, X., Zhou, B. & Pan, D. 2017a Deriving colored dissolved organic matter absorption coefficient from ocean color with a neural quasi-analytical algorithm. *Journal of Geophysical Research-Oceans* **122**, 8543–8556.

Chen, J., Zhu, W., Tian, Y. Q., Yu, Q., Zheng, Y. & Huang, L. 2017b Remote estimation of colored dissolved organic matter and chlorophyll-a in Lake Huron using Sentinel-2 measurements. *Journal of Applied Remote Sensing* **11**, 036007-1-16.

Chen, J., Zhu, W., Pang, S. & Cheng, Q. 2019 Applicability evaluation of Landsat-8 for estimating low concentration colored dissolved organic matter in inland water. *Geocarto International*. https://dx.doi.org/10.1080/10106049.2019.1704071

Chen, J., Zhu, W. N., Tian, Y. Q. & Yu, Q. 2020 Monitoring dissolved organic carbon by combining Landsat-8 and Sentinel-2 satellites: case study in Saginaw River estuary, Lake Huron. *Science of the Total Environment* **718**, 11.

Cherukuru, N., Ford, P. W., Matear, R. J., Oubelkheir, K., Clementson, L. A., Suber, K. & Steven, A. D. L. 2016 Estimating dissolved organic carbon concentration in turbid coastal waters using optical remote sensing observations. *International Journal of Applied Earth Observation and Geoinformation* **52**, 149–154.

Deng, L., Yang, W. Y. & Liu, H. 2019 PredPRBA: prediction of protein-RNA binding affinity using gradient boosted regression trees. *Frontiers in Genetics* **10**, 11.

Elith, J., Leathwick, J. R. & Hastie, T. 2008 A working guide to boosted regression trees. *Journal of Animal Ecology* **77**, 802–813.

Griffin, C. G., Frey, K. E., Rogan, J. & Holmes, R. M. 2011 Spatial and interannual variability of dissolved organic matter in the Kolyma River, East Siberia, observed using satellite imagery. *Journal of Geophysical Research-Biogeosciences* **116**, G03018-1-13.

Gurlin, D., Gitelson, A. A. & Moses, W. J. 2011 Remote estimation of chl-a concentration in turbid productive waters – return to a simple two-band NIR-red model? *Remote Sensing of Environment* **115**, 3479–3490.

Heddam, S. 2014 Generalized regression neural network (GRNN)-based approach for colored dissolved organic matter (CDOM) retrieval: case study of Connecticut River at Middle Haddam Station, USA. *Environmental Monitoring and Assessment* **186**, 7837–7848.

Helms, J. R., Stubbins, A., Ritchie, J. D., Minor, E. C., Kieber, D. J. & Mopper, K. 2008 Absorption spectral slopes and slope ratios as indicators of molecular weight, source, and photobleaching of chromophoric dissolved organic matter. *Limnology and Oceanography* **53**, 955–969.

Joshi, I. D., D'Sa, E. J., Osburn, C. L., Bianchi, T. S., Ko, D. S., Oviedo-Vargas, D., Arellano, A. R. & Ward, N. D. 2017 Assessing chromophoric dissolved organic matter (CDOM) distribution, stocks, and fluxes in Apalachicola Bay using combined field, VIIRS ocean color, and model observations. *Remote Sensing of Environment* **191**, 359–372.

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 Sensing of Environment* **99**, 66–74.

Kuang, L., Yu, L., Huang, L., Wang, Y., Ma, P. J., Li, C. B. & Zhu, Y. J. 2018 A personalized QoS prediction approach for CPS service recommendation based on reputation and location-aware collaborative filtering. *Sensors* **18**, 32.

Kutser, T. 2012 The possibility of using the Landsat image archive for monitoring long time trends in coloured dissolved organic matter concentration in lake waters. *Remote Sensing of Environment* **123**, 334–338.

Kutser, T., Pierson, D., Tranvik, L., Reinart, A., Sobek, S. & Kallio, K. 2005a Using satellite remote sensing to estimate the colored dissolved organic matter absorption coefficient in lakes. *Ecosystems* **8**, 709–720.

Kutser, T., Pierson, D. C., Kallio, K. Y., Reinart, A. & Sobek, S. 2005b Mapping lake CDOM by satellite remote sensing. *Remote Sensing of Environment* **94**, 535–540.

Li, Y., Zhang, Y., Shi, K., Zhu, G., Zhou, Y., Zhang, Y. & Guo, Y. 2017 Monitoring spatiotemporal variations in nutrients in a large drinking water reservoir and their relationships with hydrological and meteorological conditions based on Landsat 8 imagery. *Science of the Total Environment* **599**, 1705–1717.

Liu, Z., Li, Y., Lue, H., Xu, Y., Xu, X., Huang, J., Tan, J. & Guo, Y. 2011 Inversion of suspended matter concentration in Lake Chaohu based on Partial Leastsquares Regression. *Hupo Kexue* **23**, 357–365.

Liu, X. H., Zhang, Y. L., Shi, K., Zhu, G. W., Xu, H. & Zhu, M. Y. 2014 Absorption and fluorescence properties of chromophoric dissolved organic matter: implications for the monitoring of water quality in a large subtropical reservoir. *Environmental Science and Pollution Research* **21**, 14078–14090.

Lyu, L. L., Wen, Z. D., Jacinthe, P. A., Shang, Y. X., Zhang, N., Liu, G., Fang, C., Hou, J. B. & Song, K. S. 2020 Absorption characteristics of CDOM in treated and non-treated urban lakes in Changchun, China. *Environmental Research* **182**, 13.

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 Sensing of Environment* **152**, 576–602.

Mobley, C. D. 1999 Estimation of the remote-sensing reflectance from above-surface measurements. *Applied Optics* **38**, 7442–7455.

Moran, M. A., Sheldon, W. M. & Zepp, R. G. 2000 Carbon loss and optical property changes during long-term photochemical and biological degradation of estuarine dissolved organic matter. *Limnology and Oceanography* **45**, 1254–1264.

Mushtaq, F. & Lala, M. G. N. 2017 Remote estimation of water quality parameters of Himalayan lake (Kashmir) using Landsat 8 OLI imagery. *Geocarto International* **32**, 274–285.

Natekin, A. & Knoll, A. 2013 Gradient boosting machines, a tutorial. *Frontiers in Neurorobotics* **7**, 21.

Olmanson, L. G., Page, B. P., Finlay, J. C., Brezonik, P. L., Bauer, M. E., Griffin, C. G. & Hozalski, R. M. 2020 Regional measurements and spatial/temporal analysis of CDOM in 10,000+ optically variable Minnesota lakes using Landsat 8 imagery. *The Science of the Total Environment* **724**, 138141.

Rokach, L. 2016 Decision forest: twenty years of research. *Information Fusion* **27**, 111–125.

Ruescas, A. B., Hieronymi, M., Mateo-Garcia, G., Koponen, S., Kallio, K. & Camps-Valls, G. 2018 Machine learning regression approaches for colored dissolved organic matter (CDOM) retrieval with S2-MSI and S3-OLCI simulated data. *Remote Sensing* **10**, 25.

Shah, S. H., Angel, Y., Houborg, R., Ali, S. & McCabe, M. F. 2019 A random forest machine learning approach for the retrieval of leaf chlorophyll content in wheat. *Remote Sensing* **11**, 26.

Shi, L., Mao, Z., Liu, M. & Zhang, Y. 2018 Effects of rainstorm on the spectral absorption properties of chromophoric dissolved organic matter and particles in Lake Qiandao. *Hupo Kexue* **30**, 358–374.

Stedmon, C. A., Markager, S. & Kaas, H. 2000 Optical properties and signatures of chromophoric dissolved organic matter (CDOM) in Danish coastal waters. *Estuarine Coastal and Shelf Science* **51**, 267–278.

Stramska, M. & Stramski, D. 2005 Variability of particulate organic carbon concentration in the north polar Atlantic based on ocean color observations with Sea-viewing Wide Field-of-view Sensor (SeaWiFS). *Journal of Geophysical Research-Oceans* **110**, 16.

Toming, K., Kutser, T., Laas, A., Sepp, M., Paavel, B. & Noges, T. 2016 First experiences in mapping lake water quality parameters with Sentinel-2 MSI imagery. *Remote Sensing* **8**, 14.

Tzortziou, M., Osburn, C. L. & Neale, P. J. 2007 Photobleaching of dissolved organic material from a tidal marsh-estuarine system of the Chesapeake Bay. *Photochemistry and Photobiology* **83**, 782–792.

Wang, X. L., Wang, Q., Wu, C. Q., Liang, T., Zheng, D. H. & Wei, X. F. 2012 A method coupled with remote sensing data to evaluate non-point source pollution in the Xin'anjiang catchment of China. *Science of the Total Environment* **430**, 132–143.

Wu, Z. X., Zhang, Y. L., Zhou, Y. Q., Liu, M. L., Shi, K. & Yu, Z. M. 2015 Seasonal-spatial distribution and long-term variation of transparency in Xin'anjiang reservoir: implications for reservoir management. *International Journal of Environmental Research and Public Health* **12**, 9492–9507.

Xin, Z., GaoFu, X. U., DongWei, S., YongJie, G. U., Hui, G. A. O., XiaoHua, L. U. O. & XiaoYong, C. 2007 Maintenance and natural regeneration of *Castanopsis sclerophylla* populations on islands of Qiandao Lake Region. *Acta Ecologica Sinica* **27**, 424–431.

Xu, J., Fang, C. Y., Gao, D., Zhang, H. S., Gao, C., Xu, Z. C. & Wang, Y. Q. 2018 Optical models for remote sensing of chromophoric dissolved organic matter (CDOM) absorption

in Poyang Lake. *Isprs Journal of Photogrammetry and Remote Sensing* **142**, 124–136.

Yang, J. C., Chuang, H. C. & Kuan, C. M. 2020 Double machine learning with gradient boosting and its application to the Big N audit quality effect. *Journal of Econometrics* **216**, 268–283.

Yoshida, K., Endo, H., Lawrenz, E., Isada, T., Hooker, S. B., Prasil, O. & Suzuki, K. 2018 Community composition and photophysiology of phytoplankton assemblages in coastal Oyashio waters of the western North Pacific during early spring. *Estuarine Coastal and Shelf Science* **212**, 80–94.

Zeng, S., Li, Y., Lyu, H., Xu, J., Dong, X., Wang, R., Yang, Z. & Li, J. 2020 Mapping spatio-temporal dynamics of main water parameters and understanding their relationships with driving factors using GF-1 images in a clear reservoir. *Environmental Science and Pollution Research* **27**, 33929–33950.

Zhan, H. G., Shi, P. & Chen, C. Q. 2001 Inversion of oceanic chlorophyll concentrations by neural networks. *Chinese Science Bulletin* **46**, 158–161.

Zhang, Y. L., van Dijk, M. A., Liu, M. L., Zhu, G. W. & Qin, B. Q. 2009 The contribution of phytoplankton degradation to chromophoric dissolved organic matter (CDOM) in eutrophic shallow lakes: field and experimental evidence. *Water Research* **43**, 4685–4697.

Zhang, Y. L., Yin, Y., Feng, L. Q., Zhu, G. W., Shi, Z. Q., Liu, X. H. & Zhang, Y. Z. 2011 Characterizing chromophoric dissolved organic matter in lake Tianmuhu and its catchment basin using excitation-emission matrix fluorescence and parallel factor analysis. *Water Research* **45**, 5110–5122.

Zhang, Y. L., Wu, Z. X., Liu, M. L., He, J. B., Shi, K., Wang, M. Z. & Yu, Z. M. 2014 Thermal structure and response to long-term climatic changes in Lake Qiandaohu, a deep subtropical reservoir in China. *Limnology and Oceanography* **59**, 1193–1202.

Zhao, J., Cao, W., Wang, G., Yang, D., Yang, Y., Sun, Z., Zhou, W. & Liang, S. 2009 The variations in optical properties of CDOM throughout an algal bloom event. *Estuarine Coastal and Shelf Science* **82**, 225–232.

Zhao, W., Li, J., Zhao, J., Zhao, D. & Zhu, X. 2019 PDD_GBR: research on evaporation duct height prediction based on gradient boosting regression algorithm. *Radio Science* **54**, 949–962.

Zhou, Y. Q., Zhang, Y. L., Shi, K., Liu, X. H. & Niu, C. 2015 Dynamics of chromophoric dissolved organic matter influenced by hydrological conditions in a large, shallow, and eutrophic lake in China. *Environmental Science and Pollution Research* **22**, 12992–13003.

Zhou, Y. Q., Zhang, Y. L., Jeppesen, E., Murphy, K. R., Shi, K., Liu, M. L., Liu, X. H. & Zhu, G. W. 2016 Inflow rate-driven changes in the composition and dynamics of chromophoric dissolved organic matter in a large drinking water lake. *Water Research* **100**, 211–221.

Zhu, W. N., Yu, Q., Tian, Y. Q., Becker, B. L., Zheng, T. & Carrick, H. J. 2014 An assessment of remote sensing algorithms for colored dissolved organic matter in complex freshwater environments. *Remote Sensing of Environment* **140**, 766–778.

First received 8 June 2020; accepted in revised form 16 November 2020. Available online 30 November 2020