Journal of
Applied Remote Sensing

RemoteSensing.SPIEDigitalLibrary.org

Remote estimation of colored dissolved organic matter and chlorophyll-a in Lake Huron using Sentinel-2 measurements

Jiang Chen
Weining Zhu
Yong Q. Tian
Qian Yu
Yuhan Zheng
Litong Huang

SPIE.

Jiang Chen, Weining Zhu, Yong Q. Tian, Qian Yu, Yuhan Zheng, Litong Huang, "Remote estimation of colored dissolved organic matter and chlorophyll-a in Lake Huron using Sentinel-2 measurements," *J. Appl. Remote Sens.* **11**(3), 036007 (2017), doi: 10.1117/1.JRS.11.036007.

Remote estimation of colored dissolved organic matter and chlorophyll-a in Lake Huron using Sentinel-2 measurements

Jiang Chen,[a] Weining Zhu,[a,*] Yong Q. Tian,[b] Qian Yu,[c] Yuhan Zheng,[a] and Litong Huang[a]

[a]Zhejiang University, Ocean College, Zhejiang, China
[b]Central Michigan University, Institute for Great Lakes Research, Department of Geography, Mount Pleasant, Michigan, United States
[c]University of Massachusetts, Department of Geosciences, Amherst, Massachusetts, United States

Abstract. Colored dissolved organic matter (CDOM) and chlorophyll-a (Chla) are important water quality parameters and play crucial roles in aquatic environment. Remote sensing of CDOM and Chla concentrations for inland lakes is often limited by low spatial resolution. The newly launched Sentinel-2 satellite is equipped with high spatial resolution (10, 20, and 60 m). Empirical band ratio models were developed to derive CDOM and Chla concentrations in Lake Huron. The leave-one-out cross-validation method was used for model calibration and validation. The best CDOM retrieval algorithm is a B3/B5 model with accuracy coefficient of determination $(R^2) = 0.884$, root-mean-squared error $(\mathrm{RMSE}) = 0.731\ \mathrm{m}^{-1}$, relative root-mean-squared error $(\mathrm{RRMSE}) = 28.02\%$, and bias $= -0.1\ \mathrm{m}^{-1}$. The best Chla retrieval algorithm is a B5/B4 model with accuracy $R^2 = 0.49$, $\mathrm{RMSE} = 9.972\ \mathrm{mg/m}^3$, $\mathrm{RRMSE} = 48.47\%$, and bias $= -0.116\ \mathrm{mg/m}^3$. Neural network models were further implemented to improve inversion accuracy. The applications of the two best band ratio models to Sentinel-2 imagery with $10\ \mathrm{m} \times 10\ \mathrm{m}$ pixel size presented the high potential of the sensor for monitoring water quality of inland lakes. © *2017 Society of Photo-Optical Instrumentation Engineers (SPIE)* [DOI: 10.1117/1.JRS.11.036007]

Keywords: colored dissolved organic matter; chlorophyll-a; remote sensing; Sentinel-2; Lake Huron.

Paper 170457 received May 27, 2017; accepted for publication Jul. 18, 2017; published online Aug. 16, 2017.

1 Introduction

Colored dissolved organic matter (CDOM) is usually the proxy of dissolved organic carbon (DOC) abundance.[1] It consists of aliphatic and aromatic polymers and typically has strong optical absorption for ultraviolet and short visible light.[2,3] CDOM is an important water quality parameter because it directly relates to aquatic ecosystem processes,[4] drinking water safety,[5] and contaminant transport.[6] There are two major sources of CDOM in aquatic ecosystems, allochthonous inputs of terrestrial materials[7] and autochthonous production by phytoplankton, benthic algae, and aquatic macrophytes.[8,9] The major sinks of CDOM are usually photobleaching and microbial decomposition.[10,11] CDOM decomposes and releases compounds that can be utilized for growth of aquatic biota. CDOM levels in regional and global scales are indicators to understand carbon cycling and climate change.[12] Chlorophyll-a (Chla) is a general measure of phytoplankton biomass,[2,13] and phytoplankton is an important base biota in the primary production, carbon cycling modeling, and monitoring the eutrophication of inland waters,[14] which are habitats of a large variety of flora and fauna based on the photosynthesis of Chla.[15] In complex

*Address all correspondence to: Weining Zhu, E-mail: zhuwn@zju.edu.cn

1931-3195/2017/$25.00 © 2017 SPIE

case II waters, biological attributes of freshwater are a mixture of relatively high concentration of Chla, nonalgal particles (NAP), and CDOM,[16] and the three water quality parameters are independent of each other, which then leads to degrade the performance of Chla retrieval. Currently, many lakes all over the world are suffering from frequent algae bloom. Accordingly, Chla concentration retrieval in synoptic scale is essential for water quality assessment and management of these productive waters.

Inland freshwater ecosystems, such as those in lakes, rivers, and reservoirs, are important quantitative components in carbon cycling.[17] Using high-resolution satellite imagery, researchers have estimated global lake abundance and found there are ~117 million lakes whose individual area is >0.002 km^2.[18] These massive lakes are pools of dissolved organic matter and particulate organic matter[19] and hence with great significance for carbon cycling.

Chla and CDOM both interact with solar radiation and have considerable impacts on the underwater light field. Their amount in surface water can be quantified by remote sensing, which overcomes the spatial and temporal limitations as well as the high costs of the conventional field-based water quality measurement.[20] For water quality assessments of middle and small lakes, we usually use the Landsat designed originally for terrestrial remote sensing applications.[21–26] Landsat series instruments are in a 30-m spatial resolution and with a 16-day revisit period. Their temporal resolution is not good for the high dynamic of water quality mapping. The newly launched Sentinel-2 satellite combines and improves the spectral and spatial characteristics of the Landsat TM/ETM+ and SPOT XS series, respectively.[27] Sentinel-2 carries an improved instrument radiometric digitization (12 bits) that is helpful for low-radiance aquatic environment applications. Its spatial resolution is 10 m in wavelength range 490 to 665 nm and the revisit period is 5 days when the Sentinel-2 operation system is in working order. These improved satellite parameters motivate us to evaluate the potential of Sentinel-2 for inland water quality estimation, because so far there are few studies using Sentinel-2 data to do the same work in lakes.[28,29] Using Sentinel-2 images, our study's objective is to (1) develop empirical models for CDOM and Chla retrievals in the case study site, Lake Huron, one of the Great Lakes, (2) improve models using a neural network model, and (3) map CDOM and Chla from Sentinel-2 with 10 m × 10 m pixel size.

2 Materials and Methods

2.1 *Study Site*

The Saginaw River and Kawkawlin River plume regions of Lake Huron (Fig. 1) were selected as the study site for water quality monitoring because of their water's complex bio-optical properties. The two rivers are the major tributaries flowing into Saginaw Bay, Lake Huron. The Kawkawlin River is relatively smaller with length 28.2 km and discharge area of 647 km^2. 40.2% of its watershed is deciduous forests (Fig. 1, red area indicates vegetation dominated). The Saginaw River is the largest river entering Lake Huron, with length 36 km and discharge area of 22,260 km^2. The headwaters of the Saginaw River cover massive vegetation and forests occupied about 30% of the overall discharge area. The Midland, Bay City, and Saginaw region is under the status of highly industrialization and urbanization as well as population growth. The surrounding environment from nature to human may have increasing adverse impact on water quality in Lake Huron, such as increased CDOM and Chla concentration and reduced water clarity.[30]

2.2 *Field Data Collection and Laboratory Measurements*

Field measurements of water quality were carried out on May 10 and October 18, 2012, and May 7, 2013. During these three cruises, 41 surface water samples were collected to determine CDOM and Chla concentrations. Water above-surface spectra were measured using the Hyperspectral Surface Acquisition System (HyperSAS) and Hyperspectral Ocean Color Radiometer (HyperOCR). Viewing direction and water surface reflectance factor ρ followed the recommended values by Mobley.[31] The two sensors in HyperSAS measured L_t (water-leaving

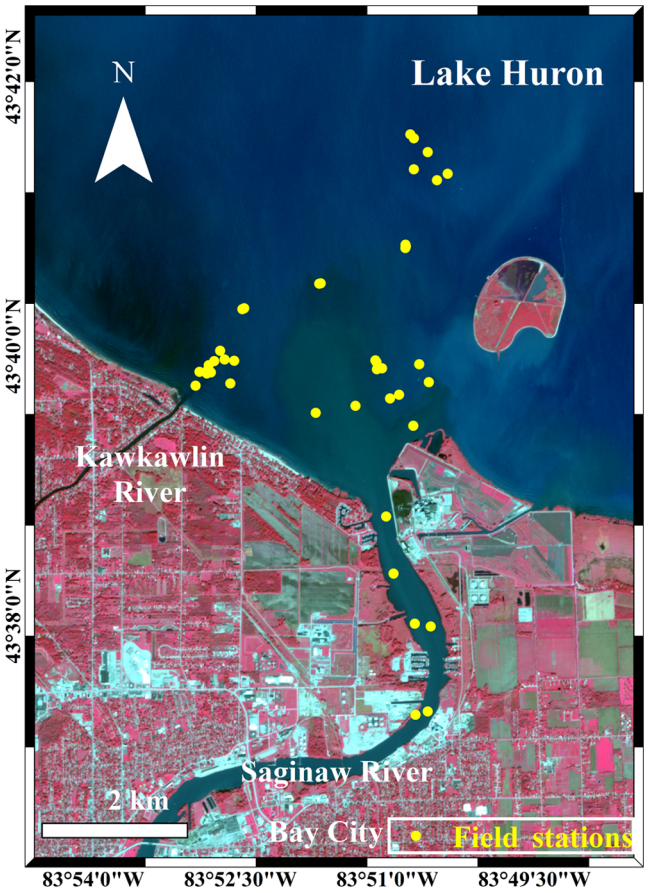

Fig. 1 Location and environment of the Saginaw River and Kawkawlin River plume regions in Lake Huron, and the sampling locations during three cruises in 2012 and 2013. The base map is a Sentinel-2 multispectral instrument (MSI) image collected on October 7, 2016.

radiance) and L_i (sky radiance), respectively, while one sensor in HyperOCR measured E_d (above-surface downwelling irradiance) simultaneously. The R_{rs} (remote sensing reflectance) at each sampling point was calculated as

$$R_{\mathrm{rs}} = \frac{L_t - \rho L_i}{E_d}. \tag{1}$$

More field measurement details are shown by Zhu et al.[32]

Table 1 gives the center wavelengths, bandwidths, and spatial resolution of the first seven bands of Sentinel-2. To use Sentinel-2 data to derive $a_{\mathrm{CDOM}}(440)$ (absorption coefficient of CDOM at 440 nm) and Chla concentrations, the first step is using field measured R_{rs} (Fig. 2) to simulate the first seven Sentinel-2 bands using its relative spectral response (RSR) functions[22]

$$R_{\mathrm{rs}}(B_i) = \frac{\int_{\lambda_m}^{\lambda_n} \mathrm{RSR}(\lambda) * R_{\mathrm{rs}}\_\mathrm{measured}(\lambda)\mathrm{d}\lambda}{\int_{\lambda_m}^{\lambda_n} \mathrm{RSR}(\lambda)\mathrm{d}\lambda}, \tag{2}$$

where R_{rs}_measured is the field measured spectra and $R_{\mathrm{rs}}(B_i)$ is the simulated R_{rs} for the i'th band of Sentinel-2, which is computed from λ_m to λ_n for the i'th band.

In laboratory, $a_{\mathrm{CDOM}}(440)$ and Chla concentrations of each sample were determined by the quantitative filter technique.[32] The CDOM absorbance was measured using a Cray-60 spectroradiometer with a 1-cm path-length cuvette and Milli-Q baseline correction. Chla was extracted with a 1:1 mixture of 90% acetone and dimethylsulfoxide and analyzed by a 10-AU Turner Fluorometer. The measured water quality results show wide variations (Table 2), which imply the complexity of inland aquatic environment. $a_{\mathrm{CDOM}}(440)$ ranges from 0.11 to 8.46 m^{-1}

Table 1 Center wavelengths, bandwidths, and spatial resolution of the first seven bands for Sentinel-2.

Band	Center wavelength (nm)	Bandwidth (nm)	Spatial resolution (m)
B1	443	20	60
B2	490	65	10
B3	560	35	10
B4	665	30	10
B5	705	15	20
B6	740	15	20
B7	783	20	20

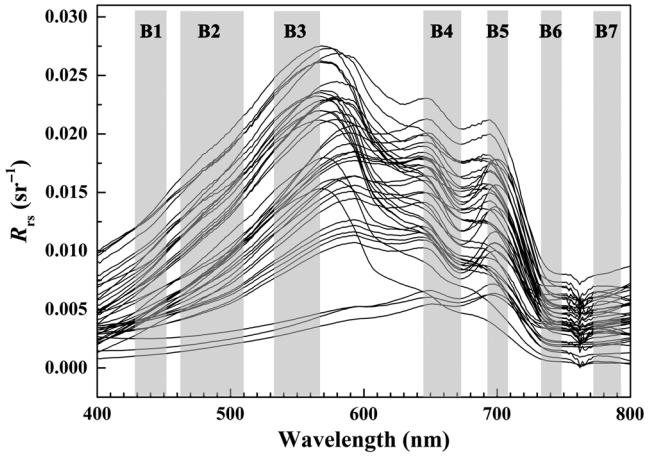

Fig. 2 The measured R_{rs} (λ) spectra. The gray-shaded bars show the ranges of Sentinel-2 MSI bands.

Table 2 Descriptive statistics of the measured water parameters, including $a_{CDOM}(440)$, Chla, and DOC.

Statistics	Min	Max	Mean	Median	SD	CV (100%)
$a_{CDOM}(440)(m^{-1})$	0.11	8.46	2.61	2.06	2.12	81.2
Chla (mg/m^3)	1.62	51.68	20.46	13.02	14.13	69.06
DOC (mg/L)	3.29	17.86	6.94	6.62	3.08	44.35

Note: SD is the standard deviation and CV is the coefficient of variation in percent.

with a mean value of 2.61 m^{-1}. Chla shows a range of 1.62 to 51.68 mg/m^3 with a mean value of 20.46 mg/m^3. DOC ranges from 3.29 to 17.86 mg/L with a mean value of 6.94 mg/L. In addition, $a_{CDOM}(440)$ and Chla show large variations with coefficient of variation (CV) values 81.2% and 69.06%, respectively. More detailed information can be referred to in the work by Zhu et al.[32,33]

2.3 *Sentinel-2 Acquisition and Preprocessing*

Sentinel-2A and Sentinel-2B were launched on June 23, 2015, and March 7, 2017, respectively, which provide a global coverage of the Earth land surface for every 5 days. However, the

Sentinel-2 operation system is in the status of instrument testing, and we cannot obtain imagery with the rigorous 5-day interval for the same area. Furthermore, frequent cloud cover reduces the availability of available images in our study area. As a result, few high-quality images could be achieved from the Sentinels Scientific Data Hub.[34] Finally, one high-quality Sentinel-2A image of Lake Huron on October 10, 2016, was downloaded for mapping CDOM and Chla concentrations with 10 m × 10 m pixel size. The image was preprocessed as the following two steps.

1. Atmospheric correction: It was implemented using the Sentinel-2 Toolbox in Sentinel Application Platform to derive a Bottom of Atmosphere reflectance image,[28] namely, irradiance reflectance (R_t), which is defined as the ratio of upwelling irradiance (E_u) to downwelling irradiance (E_d).[35]
2. Water surface reflectance correction: Remote sensing reflectance (R_{rs}) is usually the standard input variable for ocean color remote sensing algorithms. We obtained R_{rs} from Sentinel-2 imagery using the following equation:

$$R_{\mathrm{rs}} = \frac{R_t}{\pi} - \frac{L_r}{E_d}, \tag{3}$$

where L_r is the surface radiance reflected upward by water surface and E_d is the downwelling irradiance. The two variables were derived by the Hydrolight, a well-known model for simulating radiative transfer in aquatic environment. To estimate the two variables, some parameters, including image acquisition locations and dates, solar zenith angles, cloud cover, and wind speed, were input into Hydrolight. These parameters could be acquired from either image metadata or the National Climatic Data Center. The other input parameters, such as water quality components and water depth, were set by their defaults because these parameters are not related to L_r and E_d.

2.4 *Model Calibration and Validation*

To minimize random factor effects, we used a reliable scheme, which is an uncertainty assessment method known as leave-one-out cross-validation (LOOCV), to implement model calibration and validation. The subset (40 samples) was randomly selected from the total 41 field measured samples for model calibration, and the rest of the subset (1 sample) was used for model validation. Thus, the above process would be repeated for 41 times. We evaluated algorithm performance by four indicators: coefficient of determination (R^2), root-mean-squared error (RMSE), relative root-mean-squared error (RRMSE), and bias

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{N} (x_i^{\mathrm{estimated}} - x_i^{\mathrm{measured}})^2}{N}}, \tag{4}$$

$$\mathrm{RRMSE} = \frac{\mathrm{RMSE}}{\sum_{i=1}^{N} x_i^{\mathrm{measured}}/N} \times 100\%, \tag{5}$$

$$\mathrm{Bias} = \frac{\sum_{i=1}^{n} (x_i^{\mathrm{estimated}} - x_i^{\mathrm{measured}})}{N}, \tag{6}$$

where $x_i^{\mathrm{estimated}}$ and x_i^{measured} represent the estimated and measured $a_{\mathrm{CDOM}}(440)$ or Chla concentrations for the No. i sample, respectively.

The linear, logarithmic, power, and exponential band ratio models were tested, respectively. Then, the R^2, RMSE, and RRMSE values of all model validations were compared with each other to determine the best band ratio model for CDOM monitoring. Numerous remote sensing algorithms have been developed to estimate Chla concentrations for complex waters, in which two-band/three-band/four-band (2-B/3-B/4-B) ratio models are widely used following the below equations:[15,33,36,37]

$$\text{Chla} \propto R_{\text{rs}}(\text{B1})/R_{\text{rs}}(\text{B2}), \tag{7}$$

$$\text{Chla} \propto [1/R_{\text{rs}}(\text{B1}) - 1/R_{\text{rs}}(\text{B2})]R_{\text{rs}}(\text{B3}), \tag{8}$$

$$\text{Chla} \propto [1/R_{\text{rs}}(\text{B1}) - 1/R_{\text{rs}}(\text{B2})]/[1/R_{\text{rs}}(\text{B4}) - 1/R_{\text{rs}}(\text{B3})]. \tag{9}$$

Note that B1 is located at red band. B2, B3, and B4 are located at near-infrared (NIR) band and B4 > B3 > B2.

To create the best band ratio model for Chla estimation, 2-B/3-B/4-B models incorporating all possible combinations of B1 to B7 of Sentinel-2 were tested using linear, logarithmic, power, and exponential functions. In addition, we used the LOOCV method to implement model development and validation. Finally, we chose the best model with higher R^2 and lower RMSE, RRMSE, and bias value.

3 Results and Discussion

3.1 *Colored Dissolved Organic Matter Retrieval*

In remote sensing, the adjacent bands are possibly correlated to each other. The correlation coefficients between two Sentinel-2 bands are shown in Table 3. The results indicated that B1 was highly correlated with both B2 and B3. There are strong statistical relationships between B4 and B5, and B6 and B7. Absorption is considered only as to CDOM optical property, while backscattering is ignored. Thus, Pearson's correlative coefficients between R_{rs} and $a_{\text{CDOM}}(440)$ present negative values (Fig. 3). This is in agreement with previous studies in which $a_{\text{CDOM}}(440)$ was inversely correlated with remote sensing reflectance.[38] The relative high R values were mainly occurred at the green bands within 550 to 575 nm. The high correlations are due to the ratio of CDOM absorption to water's total absorption at the green band for complex inland waters.[39] In addition, we examined the Pearson's correlative coefficients between R_{rs} at Sentinel-2 band and $a_{\text{CDOM}}(440)$. The outcomes show that B3 (green band) generates the highest R value −0.897 with $a_{\text{CDOM}}(440)$, which is greater than B1 and B2, −0.741 and −0.890, respectively. B4 to B7, by contrast, present a weak correlation (R value is no more than −0.251).

Band ratio models tend to remove the certain impacts of atmospheric correction errors.[40] Furthermore, previous study proved that including wavelengths >600 nm in the band ratio model improves CDOM retrieval accuracy notably, particularly for optical complex inland waters.[32] In this study, a band ratio algorithm assembled in the form that short bands (B1 to B3) highly correlated with $a_{\text{CDOM}}(440)$ acted as the numerator and longer bands (B4 to B7) as the denominator had higher accuracy to derive $a_{\text{CDOM}}(440)$. The results indicated that the exponential model is more reliable to build a relationship between $a_{\text{CDOM}}(440)$ and variables,

Table 3 R^2 values between first seven bands of Sentinel-2.

Band No.	B1	B2	B3	B4	B5	B6	B7
B1	—	—	—	—	—	—	—
B2	**0.968**	—	—	—	—	—	—
B3	**0.866**	**0.954**	—	—	—	—	—
B4	0.114	0.165	0.292	—	—	—	—
B5	0.012	0.022	0.078	**0.854**	—	—	—
B6	0.196	0.167	0.198	0.575	0.692	—	—
B7	0.153	0.120	0.139	0.491	0.648	**0.986**	—

Note: Bold values exhibit correlations which were highly significant ($p < 0.05$).

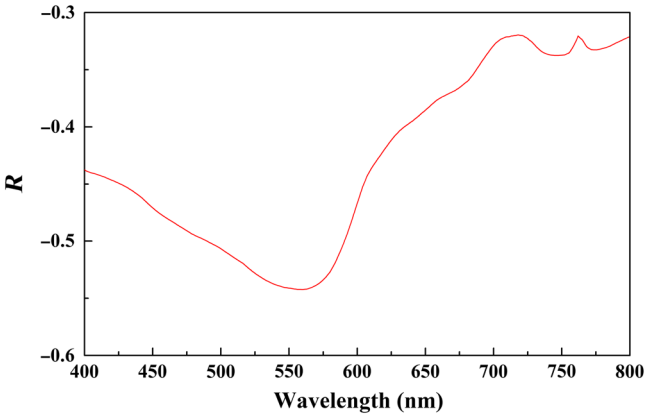

Fig. 3 Pearson's correlative coefficients (R) between R_{rs} and $a_{CDOM}(440)$.

and the function is such that $y = ae^{bx}$, where y is the $a_{CDOM}(440)$ and x is the remote sensing reflectance of band ratio.

Based on the LOOCV analytic method, the statistical values of model validation results can be seen in Table 4. The better four models are highlighted in bold with higher R^2 and lower RMSE, RRMSE, and bias. In addition, Fig. 4 shows the measured $a_{CDOM}(440)$ against those from Sentinel-2 estimated values based on the four better band ratio models. B3/B5 band ratio model with the highest performance ($R^2 = 0.884$, RMSE $= 0.731$ m^{-1}, RRMSE $= 28.02\%$, and bias $= -0.1$ m^{-1}) among all different algorithms was as follows:

$$a_{CDOM}(440) = 22.283e^{-1.724x}, X = R_{rs}(B3)/R_{rs}(B5). \quad (10)$$

As expected, band ratio models that incorporated bands >600 nm obtained much better performance for CDOM estimations.[32] The B5 band (>600 nm) is used here to remove the effects of particulate matters, as well as for normalizing purposes. a_{CDOM} at short bands (B1 and B2) are much larger than longer bands (B3, green band), considering the exponential decay of a_{CDOM}

Table 4 LOOCV model validation results of CDOM model performance corresponding to different band ratios.

Variables	a	b	R^2	RMSE ($m-1$)	RRMSE (100%)	Bias ($m-1$)
B1/B4	12.171	−3.23	0.581	1.357	51.98	0.02
B1/B5	13.283	−3.347	0.696	1.167	44.73	−0.1
B1/B6	16.349	−1.415	0.685	1.186	45.45	−0.09
B1/B7	14.027	−1.181	0.676	1.205	46.17	−0.1
B2/B4	**20.899**	**−2.952**	**0.779**	**0.987**	**37.83**	**−0.05**
B2/B5	**18.618**	**−2.709**	**0.849**	**0.836**	**32.04**	**−0.14**
B2/B6	18.204	−0.977	0.729	1.106	42.37	−0.11
B2/B7	14.712	−0.791	0.693	1.179	45.18	−0.12
B3/B4	**28.966**	**−2.015**	**0.832**	**0.859**	**32.9**	**−0.03**
B3/B5	**22.283**	**−1.724**	**0.884**	**0.731**	**28.02**	**−0.1**
B3/B6	20.61	−0.619	0.655	1.252	47.98	−0.12
B3/B7	16.425	−0.506	0.617	1.323	50.7	−0.15

Note: The bold rows indicate models with better performance.

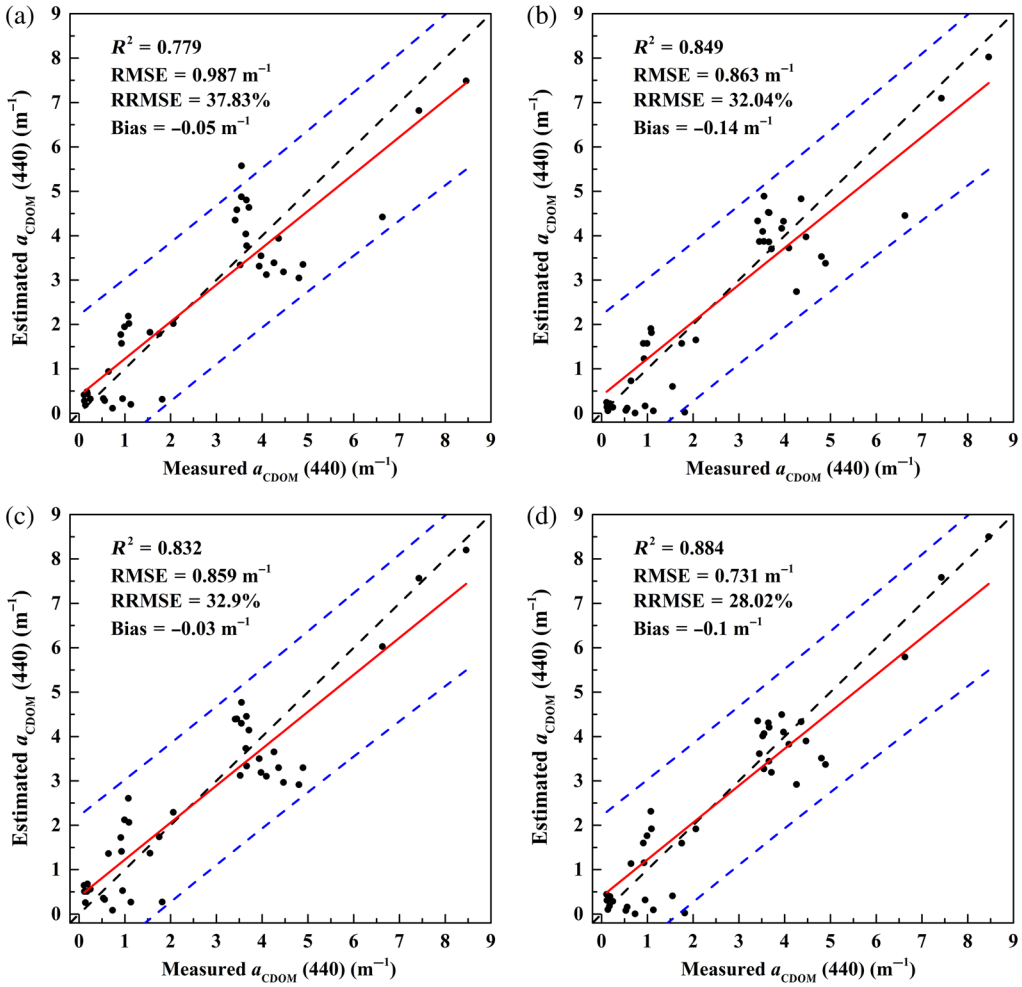

Fig. 4 Scatter plots showing the measured $a_{\mathrm{CDOM}}(440)$ versus Sentinel-2 estimations using four better band ratio models: (a) B2/B4, (b) B2/B5, (c) B3/B4, and (d) B3/B5. The red solid line indicates the regression line between estimated and measured $a_{\mathrm{CDOM}}(440)$ values. The double blue dashed lines are the 95% prediction bands, and the black dashed line is the $y = x$ line.

with the increasing of wavelengths. However, Chla and NAP also absorb light strongly at the short blue band. Finally, water-leaving signals from short blue bands are very weak, causing a_{CDOM} at those bands to take small proportions of the water's total absorptions, while such a case is opposite for the green band. The green band is thus more sensitive to derive CDOM, and models using the green band achieve better accuracy for complex inland waters.[39]

3.2 *Chlorophyll-a Retrieval*

We found the best Chla retrieval model among all testing: $\mathrm{Chla} = 25.985x^{3.117}$, $x = R_{\mathrm{rs}}(B5)/R_{\mathrm{rs}}(B4)$, with $R^2 = 0.49$, $\mathrm{RMSE} = 9.972\ \mathrm{mg/m^3}$, $\mathrm{RRMSE} = 48.47\%$, and $\mathrm{bias} = -0.116\ \mathrm{mg/m^3}$ (Fig. 5). All samples were nearly located inside of the 95% prediction bands, and the best model performance was acceptable. The best model was not the 3-B/4-B algorithm but the 2-B model. The reason is that the 2-B model, an NIR–red model, is just simple and sufficient without involving more uncertainties carried by the additional bands in the 3-B/4-B models.[33] Gurlin et al.[37] compared several NIR–red models with different levels of complexity for case II waters and found that the simple applicable NIR–red 2-B algorithm holds a high potential for Chla retrieval. The best results of the 2-B/3-B/4-B models in our study site are generally excellent and acceptable. We also tested normalized difference chlorophyll index and fluorescence line height algorithms for Chla estimation and found the same results that band ratio empirical models gave the best performance.[41]

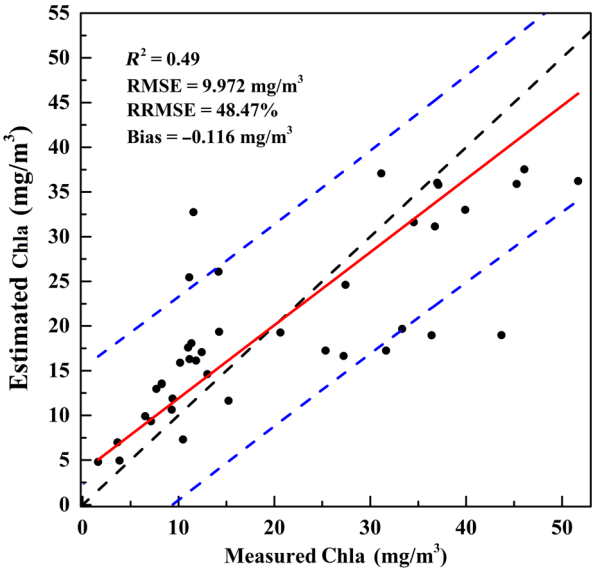

Fig. 5 Scatter plots showing the measured Chla versus Sentinel-2 estimations using B5/B4 power model. The red solid line, blue, and black dashed lines refer to the same as in Fig. 4.

3.3 *Neural Network Model*

To improve the accuracy of water quality retrieval models, an artificial neural network model was considered, which provides an alternative intelligence algorithm compared to traditional empirical statistical models. Beale and Jackson[42] introduced that any mathematical function could be depicted via neural network (NN), which is well-known as the Kolmogorov representation theorem. In addition, 80% to 90% of applied cases utilize the backpropagation (BP) pattern to develop network architecture. A typical NN model is comprised of one input layer, one or more hidden layers, and one output layer. One hidden layer is highly recommended as a first choice along with increasing the number of neurons. To avoid overfitting and obtain good accuracy, we set the rule that the optimal number of neurons in the hidden layer was determined by selecting the satisfying model with higher accuracy and fewer neurons. Therefore, in this study, about 1 to 20 neurons were tested to determine the best architecture that combined a minimal MSE value with the least neurons.[43] Note that it is the preliminary result using NN models, and training of NN will need a larger dataset to guarantee the robustness of the neural network for model prediction in future work. The NN model was conducted by MATLAB®, and the detailed implementation process may be found in studies.[38,43] During the NN model development, 26 samples randomly chosen from the total 41 samples were used to establish a stable network, and the remaining 15 samples were used to assess model accuracy.

Based on the above empirical model analysis, four variables, including $R_{rs}(B2)/R_{rs}(B4)$, $R_{rs}(B2)/R_{rs}(B5)$, $R_{rs}(B3)/R_{rs}(B4)$, and $R_{rs}(B3)/R_{rs}(B5)$, acted as the four input nodes of the input layer for the CDOM-NN model. Eventually, the CDOM-NN model comprises the following: an input layer with four nodes and a hidden layer with eights neurons. The CDOM-NN model presented the results with $R^2 = 0.913$, RMSE $= 0.601$ m^{-1}, RRMSE $= 23.3\%$, and bias $= -0.09$ m^{-1}.

Sun et al.[22] used multivariate regression analysis to estimate phycocyanin pigment concentration in inland waters from Landsat imagery and obtained acceptable performance. In this study, a multivariate model was proposed to estimate Chla

$$\begin{aligned}\text{Chla} = {} & 3.134 + 5636.315R_{rs}(B1) - 10816.193R_{rs}(B2) + 6671.622R_{rs}(B3) - 11010.068R_{rs}(B4) \\ & + 10878.736R_{rs}(B5) + 4838.647R_{rs}(B6) - 6863.896R_{rs}(B7). \end{aligned} \quad (11)$$

In contrast, the Chla estimations accuracy was improved using the multivariate regression model, and validation results reveal $R^2 = 0.733$, RMSE $= 7.268$ mg/m^3, RRMSE $= 35.5\%$, and bias $= -0.292$ mg/m^3 (Fig. 6).

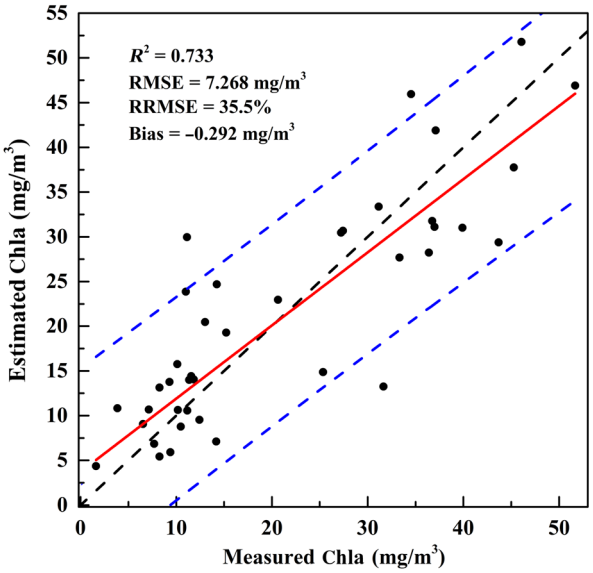

Fig. 6 Validation results of Chla retrievals. Scatter plots showing the measured Chla against those from Sentinel-2 estimations based on the multivariate regression model.

Equation (11) implies that R_{rs} at B1 to B7 bands may be a function of Chla, but the special equation may be not highly accurate to account for Chla. Similarly, the Chla-NN model, which comprises the following: an input layer with seven nodes and a hidden layer with nine neurons, is employed to determine Chla concentrations. The Chla-NN model presented the results with $R^2 = 0.95$, RMSE $= 3.127$ mg/m^3, RRMSE $= 10.4\%$, and bias $= 0.77$ mg/m^3. Figure 7 shows the validation results from the two NN models. Apparently, those models were more accurate than the empirical models. Simple band ratio models somehow remove atmospheric correction error when deriving water quality parameters from satellite imagery.[40] In general, the short (blue) waveband poses greater atmospheric correction errors,[44] and the analysis of the relationship within bands in Sec. 3.1 provides a basis for the optimal band selecting.

3.4 *Application to Sentinel-2 Imagery*

The Sentinel-2 operation system is now in the status of instrument testing, and we cannot obtain imagery with the rigorous 5-day interval for the same area. The exact match between *in situ* measurement and satellite overpass is usually difficult to obtain, particularly if the satellite

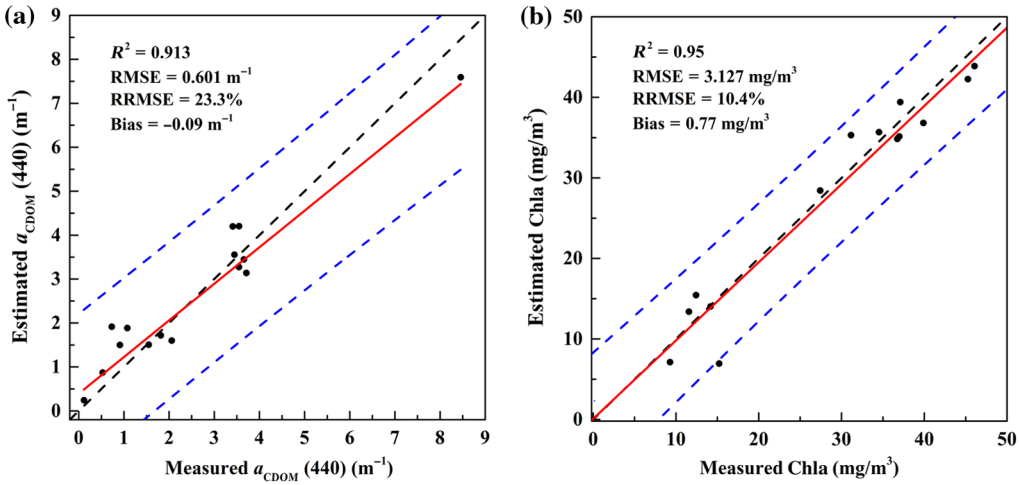

Fig. 7 Validation results of (a) $a_{CDOM}(440)$ retrievals and (b) Chla retrievals, based on NN model.

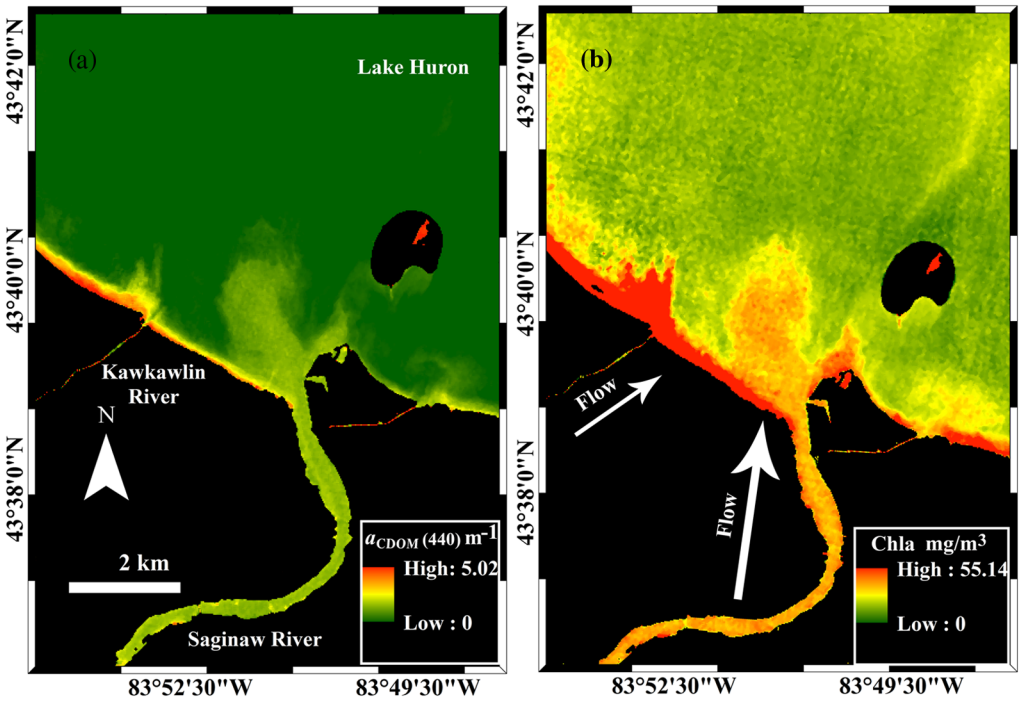

Fig. 8 CDOM and Chla estimations using a Sentinel-2 image on October 7, 2016. Results of atmospheric correction using the 10-m pixel size from Sen2cor output.

temporal resolution is relatively long now. So far, a few high-quality images could be achieved from the Sentinels Scientific Data Hub.[34] Band ratio models tend to remove the certain impacts of atmospheric correction errors.[40] NN models are more accurate, while they also involve multivariable inputs. Direct atmospheric correction assessment should be conducted with *in situ* observations in the future to use the NN model on Sentinel-2 imagery.

Atmospheric correction is a critical issue for image-based water color remote sensing. We would like to emphasize that the atmospheric correction module, Sen2cor, used in this study is particularly designed for atmospheric correction of Sentinel-2 imagery. The Sen2cor toolkit has also been tested for atmospheric correction over waters and obtained good results in previous studies.[28,29] When the toolkit was used in our study, some module input parameters were adjusted according to the field or regional realities in Lake Huron. Given the results of the Sen2cor toolkit being reliable, then water surface reflectance effects could be removed using the HydroLight, a well-known radiative transfer numerical model.[45] As a result, the remote sensing reflectance (R_{rs}), a standard input of ocean color algorithm, can be obtained from Sentinel-2 imagery. For more validation, we also compared the atmospheric correction results of Sentinel-2 with the help of a moderate-resolution imaging spectrometer (MODIS) R_{rs} product (freely downloaded from Ref. 46). We found that R_{rs} of Sentinel-2 kept good consistency with R_{rs} of MODIS in our study site.

Figure 8 shows the CDOM and Chla retrievals with 10 m × 10 m pixel size from Sentinel-2 imagery on October 7, 2016. The Saginaw River and Kawkawlin River plume regions of Lake Huron show spatial distributions of CDOM and Chla with ranges of 0 to 5.02 m^{-1} and 0 to 55.14 mg/m^3, respectively. The Kawkawlin River water was much more turbid than the Saginaw River which contains less CDOM and Chla. In estuarine and plume regions, CDOM and Chla concentrations were higher than those in the open water of Lake Huron. These image-derived distribution patterns were consistent with *in situ* observations and field measurements during the same month in another year. The results provide the evidence that Sentinel-2 offers the potential to assess water quality for complex inland waters.

Numerous remote sensing models have been proposed to derive the CDOM and Chla concentrations in productive waters, including empirical, semiempirical, semianalytical, and analytical models.[32,47] Owing to its simplicity and availability among those algorithms, studies

based on empirical algorithms are easier to get and much more common. However, empirical algorithms are often area dependent and time dependent due to no rigorous theoretical basis to support them, which indicates that the models here could not be directly used for other waters.[48] Note that the band ratio setting might be suitable for those waters, where the optical properties are similar to Lake Huron. Then, empirical models should reparameterize the equations.

4 Conclusions

This study proved that Sentinel-2 imagery can continuously monitor water color parameters, particularly, to assess water quality dynamic for many lakes in different scales. Results show that the best band ratio for deriving $a_{\mathrm{CDOM}}(440)$ is B3/B5, and it is a reasonable and applicable model in Lake Huron. For Chla, the best band combination is the B5/B4 2-B model rather than 3-B or 4-B models. In addition, a multivariate regression model was proposed to estimate Chla and we obtained good results. CDOM-NN and Chla-NN models have been trained based on a BP neural network, using models with a hidden layer of eight and nine neurons, respectively, and they were selected as the best model architecture. Validation results showed that CDOM estimation accuracy is $R^2 = 0.913$, RMSE $= 0.601$ m^{-1}, and RRMSE $= 23.3\%$, and the Chla model also performed well with $R^2 = 0.95$, RMSE $= 3.127$ mg/m^3, and RRMSE $= 10.4\%$. The estimated CDOM and Chla from a Sentinel-2 image on October 10, 2016, show a general spatial distribution pattern that is consistent with our field measurements in the same season. Although the proposed algorithms have performed well in Lake Huron, our study is still preliminary, and more studies should be conducted for further model improvement and validation.

Acknowledgments

This research was supported by the National Natural Science Foundation of China (No. 41471346), the Natural Science Foundation of Zhejiang Province (No. LY17D010005), an internal grant from Central Michigan University, and two collaborative grants from the National Science Foundation (Grant Nos. 1025547 and 1230261).

References

1. D. A. Hansell and C. A. Carlson, *Biogeochemistry of Marine Dissolved Organic Matter*, Academic Press, San Diego, California (2002).
2. J. T. Kirk, *Light and Photosynthesis in Aquatic Ecosystems*, Cambridge University Press, Cambridge, United Kingdom (1994).
3. C. A. Stedmon, S. Markager, and R. Bro, "Tracing dissolved organic matter in aquatic environments using a new approach to fluorescence spectroscopy," *Mar. Chem.* **82**(3), 239–254 (2003).
4. J. N. Houser, "Water color affects the stratification, surface temperature, heat content, and mean epilimnetic irradiance of small lakes," *Can. J. Fish. Aquat. Sci.* **63**(11), 2447–2455 (2006).
5. P. Herzsprung et al., "Variations of DOM quality in inflows of a drinking water reservoir: linking of van Krevelen diagrams with EEMF spectra by rank correlation," *Environ. Sci. Technol.* **46**(10), 5511–5518 (2012).
6. B. A. Bergamaschi et al., "Methyl mercury dynamics in a tidal wetland quantified using in situ optical measurements," *Limnol. Oceanogr.* **56**(4), 1355–1371 (2011).
7. J. Aitkenhead-Peterson et al., "Sources, production, and regulation of allochthonous dissolved organic matter inputs to surface waters," in *Aquatic Ecosystems: Interactivity of Dissolved Organic Matter*, S. E. G. Findlay and and R. L. Sinsabaugh, Eds., pp. 26–70, Academic Press, San Diego, California (2003).
8. S. Bertilsson and J. Jones, "Supply of dissolved organic matter to aquatic ecosystems: autochthonous sources," in *Aquatic Ecosystems: Interactivity of Dissolved Organic Matter*, S. E. G. Findlay and and R. L. Sinsabaugh, Eds., pp. 3–24, Academic Press, San Diego, California (2003).

9. D. L. Bade et al., "Sources and fates of dissolved organic carbon in lakes as determined by whole-lake carbon isotope additions," *Biogeochemistry* **84**(2), 115–129 (2007).
10. P. G. Coble, "Marine optical biogeochemistry: the chemistry of ocean color," *Chem. Rev.* **107**(2), 402–418 (2007).
11. M. A. Moran and R. G. Zepp, "Role of photoreactions in the formation of biologically labile compounds from dissolved organic matter," *Limnol. Oceanogr.* **42**(6), 1307–1316 (1997).
12. K. Mopper, "Photochemical source of biological substrates in sea water: implications for carbon cycling," *Nature* **341**, 19 (1989).
13. Y. Huot et al., "Relationship between photosynthetic parameters and different proxies of phytoplankton biomass in the subtropical ocean," *Biogeosciences* **4**(5), 853–868 (2007).
14. T. Kutser, "Quantitative detection of chlorophyll in cyanobacterial blooms by satellite remote sensing," *Limnol. Oceanogr.* **49**(6), 2179–2189 (2004).
15. W. J. Moses et al., "Satellite estimation of chlorophyll-concentration using the red and NIR bands of MERIS—the Azov Sea case study," *IEEE Geosci. Remote Sens. Lett.* **6**(4), 845–849 (2009).
16. A. Morel and L. Prieur, "Analysis of variations in ocean color," *Limnol. Oceanogr.* **22**(4), 709–722 (1977).
17. J. J. Cole et al., "Plumbing the global carbon cycle: integrating inland waters into the terrestrial carbon budget," *Ecosystems* **10**(1), 172–185 (2007).
18. C. Verpoorter et al., "A global inventory of lakes based on high-resolution satellite imagery," *Geophys. Res. Lett.* **41**(18), 6396–6402 (2014).
19. G. M. Wilkinson, M. L. Pace, and J. J. Cole, "Terrestrial dominance of organic matter in north temperate lakes," *Global Biogeochem. Cycles* **27**(1), 43–51 (2013).
20. N. Torbick et al., "Mapping chlorophyll-a concentrations in West Lake, China using Landsat 7 ETM+," *J. Great Lakes Res.* **34**(3), 559–565 (2008).
21. F. Mushtaq and M. G. Nee Lala, "Remote estimation of water quality parameters of Himalayan lake (Kashmir) using Landsat 8 OLI imagery," *Geocarto Int.* **32**, 1–12 (2016).
22. D. Sun et al., "Estimating phycocyanin pigment concentration in productive inland waters using Landsat measurements: a case study in Lake Dianchi," *Opt. Express* **23**(3), 3055–3074 (2015).
23. M. J. Butt and M. Nazeer, "Landsat ETM+ Secchi Disc Transparency (SDT) retrievals for Rawal Lake, Pakistan," *Space Res.* **56**(7), 1428–1440 (2015).
24. M. G. Allan et al., "Empirical and semi-analytical chlorophyll a algorithms for multi-temporal monitoring of New Zealand lakes using Landsat," *Environ. Monit. Assess.* **187**(6), 364 (2015).
25. K. Singh et al., "Blue–red–NIR model for chlorophyll-a retrieval in hypersaline–alkaline water using Landsat ETM+ sensor," *IEEE J. Sel. Top. Appl. Earth Obs.* **7**(8), 3553–3559 (2014).
26. N. Torbick et al., "Mapping inland lake water quality across the Lower Peninsula of Michigan using Landsat TM imagery," *Int. J. Remote Sens.* **34**(21), 7607–7624 (2013).
27. J. Hedley et al., "Capability of the Sentinel 2 mission for tropical coral reef mapping and coral bleaching detection," *Remote Sens. Environ.* **120**, 145–155 (2012).
28. K. Toming et al., "First experiences in mapping lake water quality parameters with Sentinel-2 MSI imagery," *Remote Sens.* **8**(8), 640 (2016).
29. K. Dörnhöfer et al., "Water constituents and water depth retrieval from Sentinel-2A—a first evaluation in an oligotrophic lake," *Remote Sens.* **8**(11), 941 (2016).
30. C. Hoard et al., "Comparison of streamflow and water-quality data collection techniques for the Saginaw River, Michigan," US Geological Survey (2012).
31. C. D. Mobley, "Estimation of the remote-sensing reflectance from above-surface measurements," *Appl. Opt.* **38**(36), 7442 (1999).
32. W. Zhu et al., "An assessment of remote sensing algorithms for colored dissolved organic matter in complex freshwater environments," *Remote Sens. Environ.* **140**, 766–778 (2014).
33. W. Zhu et al., "Issues and potential improvement of multiband models for remotely estimating chlorophyll-a in complex inland waters," *IEEE J. Sel. Top. Appl. Earth Obs.* **8**(2), 562–575 (2015).
34. https://scihub.copernicus.eu/dhus/#/home

35. C. D. Mobley, *Light and Water: Radiative Transfer in Natural Waters*, Academic Press, San Diego, California (1994).
36. P. V. Zimba and A. Gitelson, "Remote estimation of chlorophyll concentration in hyper-eutrophic aquatic systems: model tuning and accuracy optimization," *Aquaculture* **256**(1–4), 272–286 (2006).
37. D. Gurlin, A. A. Gitelson, and W. J. Moses, "Remote estimation of chl-a concentration in turbid productive waters—return to a simple two-band NIR–red model?" *Remote Sens. Environ.* **115**(12), 3479–3490 (2011).
38. D. Y. Sun et al., "A neural-network model to retrieve CDOM absorption from in situ measured hyperspectral data in an optically complex lake: Lake Taihu case study," *Int. J. Remote Sens.* **32**(14), 4005–4022 (2011).
39. J. Chen et al., "Estimation of colored dissolved organic matter from Landsat-8 imagery for complex inland water: case study of Lake Huron," *IEEE Trans. Geosci. Remote Sens.* **55**, 2201–2212 (2017).
40. N. Cherukuru et al., "Estimating dissolved organic carbon concentration in turbid coastal waters using optical remote sensing observations," *Int. J. Appl. Earth Obs. Geoinf.* **52**, 149–154 (2016).
41. R. Beck et al., "Comparison of satellite reflectance algorithms for estimating chlorophyll-a in a temperate reservoir using coincident hyperspectral aircraft imagery and dense coincident surface observations," *Remote Sens. Environ.* **178**, 15–30 (2016).
42. R. Beale and T. Jackson, *Neural Computing—An Introduction*, CRC Press, Boca Raton, Florida (1990).
43. J. Chen et al., "Remote sensing of absorption and scattering coefficient using neural network model: development, validation, and application," *Remote Sens. Environ.* **149**, 213–226 (2014).
44. A. Morel and B. Gentili, "A simple band ratio technique to quantify the colored dissolved and detrital organic material from ocean color remotely sensed data," *Remote Sens. Environ.* **113**(5), 998–1011 (2009).
45. W. Zhu and Q. Yu, "Inversion of chromophoric dissolved organic matter from EO-1 Hyperion imagery for turbid estuarine and coastal waters," *IEEE Trans. Geosci. Remote Sens.* **51**(6), 3286–3298 (2013).
46. https://oceancolor.gsfc.nasa.gov/
47. B. M. Lesht, R. P. Barbiero, and G. J. Warren, "Satellite ocean color algorithms: a review of applications to the Great Lakes," *J. Great Lakes Res.* **38**(1), 49–60 (2012).
48. G. Wu et al., "Statistical model development and estimation of suspended particulate matter concentrations with Landsat 8 OLI images of Dongting Lake, China," *Int. J. Remote Sens.* **36**(1), 343–360 (2015).

Jiang Chen received his BS degree in geographic information science from Huazhong Agricultural University, Wuhan, China, in 2015. He is currently pursuing his master's degree in the Department of Ocean Science, Ocean College, Zhejiang University, Zhejiang, China. His research interests include water bio-optical properties and ocean color remote sensing.

Weining Zhu received his PhD in geosciences from the University of Massachusetts, Amherst, Massachusetts, USA, in 2012. He is currently an associate professor in the Department of Ocean Science, Ocean College, Zhejiang University, Zhejiang, China. His research interests include ocean color, hyperspectral remote sensing, and geographic information system spatial analysis.

Yong Q. Tian received his PhD in computer science from the University of Waikato, Hamilton, New Zealand, in 1995. Since 2011, he has been a professor and a cohort of faculty with specializations relevant to Great Lakes Research at Central Michigan University, Mount Pleasant, Michigan, USA. His current research interests include geographic information system and remote sensing, as well as coastal ecological modeling and analysis.

Qian Yu received her BS and MS degrees in geography from Nanjing University, Nanjing, China, in 1999 and 2001, respectively, and her PhD from the University of California, Berkeley, California, USA, in 2005. Since 2006, she has been a faculty member in the Department of Geosciences, University of Massachusetts, Amherst, Massachusetts, USA. Her current research

interests include hyperspectral and high-resolution remote sensing of vegetation and water biogeochemistry.

Yuhan Zheng received her BE degree in water resources and hydropower engineering from Wuhan University, Wuhan, China, in 2015. She is currently pursuing her master's degree in the Department of Ocean Science, Ocean College, Zhejiang University, Zhejiang, China. Her research interests include remote sensing classification with high-resolution imagery and hydrological remote sensing.

Litong Huang is currently pursuing his master's degree in the Department of Ocean Science, Ocean College, Zhejiang University, Zhejiang, China. His research interests include ocean color and remote sensing applications.