Improving remote sensing estimation of Secchi disk depth for global lakes and reservoirs using machine learning methods

Yibo Zhang\textsuperscript{a,b}, Kun Shi\textsuperscript{a,b,c}, Xiao Sun\textsuperscript{a,b}, Yunlin Zhang\textsuperscript{a,b}, Na Li\textsuperscript{a,b}, Weijia Wang\textsuperscript{a,b}, Yongqiang Zhou\textsuperscript{a,b}, Wei Zhi\textsuperscript{a}, Mingliang Liu\textsuperscript{d}, Yuan Li\textsuperscript{e}, Guangwei Zhu\textsuperscript{a,b}, Boqiang Qin\textsuperscript{a,b}, Erik Jeppesen\textsuperscript{g}, Jian Zhou\textsuperscript{a,b} and Huiyun Li\textsuperscript{a,b}

\textsuperscript{a}State Key Laboratory of Lake Science and Environment, Nanjing Institute of Geography and Limnology, Chinese Academy of Sciences, Nanjing, China; \textsuperscript{b}University of the Chinese Academy of Sciences, Beijing, China; \textsuperscript{c}CAS Center for Excellence in Tibetan Plateau Earth Sciences, Beijing, China; \textsuperscript{d}Department of Civil and Environmental Engineering, The Pennsylvania State University, State College, PA, USA; \textsuperscript{e}Institute of Environmental Protection Science, Hangzhou, China; \textsuperscript{f}School of Tourism and Urban & Rural Planning, Zhejiang Gongshang University, Hangzhou, China; \textsuperscript{g}Department of Bioscience and Arctic Research Centre, Aarhus University, Silkeborg, Denmark

ABSTRACT

Secchi disk depth (SDD) is a simple but particularly important indicator for characterizing the overall water quality status and assessing the long-term dynamics of water quality for diverse global waters. For this reason, countless efforts have been made to collect SDD data from the field and through remote sensing systems. Many empirical and semianalytical algorithms have been proposed to estimate SDD from different satellite images for a specific or regional water. However, the construction of a robust global SDD estimation model is still challenging due to the nonlinear response of SDD to optical properties and the complex physical and biogeochemical processes of different waters. Therefore, machine learning methods to better interpret nonlinear processes were used to improve remote sensing estimations of SDD for global lakes and reservoirs based on a global matchup dataset from Landsat TM (N = 4099), ETM+ (N = 2420), and OLI (N = 1249) covering in situ SDD from 0.01 m to over 18 m. Overall, extreme gradient boosting (XGBoost) and random forest (RF) had better SDD retrievals than back propagation neural network, support vector regression, empirical and quasi-analytical models showing high precision with mean relative error of approximately 30% and good agreements with the long-term in situ SDD in different waters with various optical properties. Our results can support long-term global-level water quality evaluation and thus making informed decisions about development policy.

ARTICLE HISTORY
Received 22 April 2022
Accepted 15 August 2022

KEYWORDS
Global lakes; reservoirs; water clarity; XGBoost; random forest

1. Introduction

Water quality is being paid increasingly attention by the public as complete water environment information is an important basis for making informed decisions regarding water resource use and development policies (Downing et al. 2021; Grant et al. 2021; Keeler et al. 2012). Secchi disk depth (SDD), which refers to the penetration ability of visible light into underwater, is a simple but particularly important indicator for characterizing overall water quality status and assessing the long-term dynamics of water quality for diverse global waters (Capuzzo et al. 2015; Lisi and Hein 2019; Song et al. 2020, 2022). SDD is influenced by the abundance of various dissolved and particulate matter in the water column (Liu et al. 2021a; Maciel et al. 2021; Wang et al. 2020); these water constituents play vital roles in regulating many chemical, physical and biological processes including dissolved oxygen and thermal stratification, primary productivity, and biological vertical migration (Jane et al. 2021; Karlsson et al. 2009; Zhang et al. 2015). In addition, good water quality has been proven to have great recreational value for activities such as boating, swimming, fishing, and sightseeing (Keeler et al. 2012); improved water SDD indicates good water quality and leads to increased visits to lakes and reservoirs. Keeler et al. (2015) stated that visitors were willing to pay an additional US$22 for every one-meter increase in water SDD.

SDD has records for more than 150 years back to 1865 because of its simplicity in measurement (Maciel et al. 2021). Although traditional SDD ship-based observations are accessible and highly accurate, insufficient spatial and temporal resolutions hinder global long-term assessments of SDD dynamics (Liu et al. 2021a; Luís et al. 2019). In contrast, satellite remote sensing, with large-scale observations and higher
spatiotemporal resolutions, has become an important alternative tool for synoptic estimations of SDD (Chawla, Karthikeyan, and Mishra 2020; Page, Olmanson, and Mishra 2019; Zhang et al. 2020).

To date, both empirical regressions (Kabiri 2022a, 2022b; Kabiri and Moradi 2016; Liu et al. 2020; Ren et al. 2018; Song et al. 2020, 2022; Zhang et al. 2021c) and semianalytical models (Jiang et al. 2019; Lee et al. 2016; Msusa, Jiang, and Matsushita 2022; Wang, Son, and Harding 2009) have been developed to estimate SDD through satellite data. Due to the nonlinear response of SDD to optical properties and the complex physical and biogeochemical processes of different waters, the construction of a robust model for different global waters is still challenging. For different clear and turbid waters, the sensitive band varies with the optical properties of the water environment. Doron et al. (2011) and Alikas and Kratzer (2017) suggested that clear ocean water absorbs less at short visible wavelengths than at longer wavelengths; thus, the short visible bands are more suitable for clear ocean water SDD remote sensing estimation. However, a previous study indicated that for turbid waters the reference band used for SDD estimation is suggested to move to longer wavelengths (Shi et al. 2015). In other words, at present, there is no unified empirical or semianalytical model to accommodate differing optical properties for remote estimation of the surface SDD for global lakes and reservoirs.

Recently, based on satellite images, many data-driven methods including artificial neural networks (ANNs) (Li et al. 2021b; Liu et al. 2021b), support vector regression (SVR) (Xiao et al. 2018), generalized additive models (Li et al. 2021a), decision trees (Song et al. 2021), random forests (RF) (Belgiu and Dragut 2016; Li et al. 2021a; Mudiyanselage et al. 2022) and deep learning (Guo et al. 2022; He et al. 2022; Zhang et al. 2021a) have been applied to different research fields such as classification and differentiation, object and change detection, and parameter inversion. With the accumulation of sample size and image data, massive amounts of data support the application of machine learning methods in water color remote sensing. For instance, Balasubramanian et al. (2020) trained an ANNs for backscattering coefficient retrievals and thus estimated total suspended matter (TSM) in different waters. Shen et al. (2020) and Yu et al. (2021) constructed an SDD model based on random forest regression and a deep neural network algorithm, respectively for turbid and eutrophic waters in eastern China and plateau lakes in Yunnan Province. Giannini et al. (2021) retrieved chromophoric dissolved organic matter, chlorophyll-a, and TSM concentrations in the Northeast Pacific coastal waters through a neural net algorithm. The state-of-the-art algorithms in the above studies suggest that machine learning methods can capture nonlinear relationships between input features and biophysical parameters or inherent optical properties and generally show improved performance over simple regression methods.

Our hypothesis is that machine learning models have improved performance over published empirical and semianalytical models for global lakes and reservoirs based on a large and extensive SDD dataset. Therefore, we compile global in situ SDD datasets and develop and validate machine learning methods with general applicability to estimate the SDD from satellites through multiperspective comparison. We aim to develop a universal SDD remote sensing estimation model widely used to diverse lakes and reservoirs with various optical properties.

2. Materials and methods

2.1 In situ data

The in situ SDD dataset was collected from five independent sources (Table S1): the Chinese lake dataset (Zhang et al. 2021c); the European lake dataset (Mantzouki et al. 2018) (https://portal.edirepository.org); the American National Aquatic Resource Surveys (ANARS) dataset (https://www.epa.gov); the Global Freshwater Quality dataset (https://gemstat.bafg.de/); and the routine water monitoring dataset collected from Lake Taihu (China), Xin’anjiang Reservoir (China), Lake Kasumigaura (Japan), Lake Mjosa (Norway), Fellows Reservoir (USA), Shayne Reservoir (USA) and Stockton Reservoir (USA) (Fig. S1). The last dataset was not a part of any other datasets. Both Lake Taihu and Lake Kasumigaura represent shallow lakes with turbid waters (Shi et al. 2015; Terrel et al. 2012). All four reservoirs and Lake Mjosa represented clear or moderately turbid waters. Datasets 1–4 were used for model development in Section 3.2, and dataset 5 was used for model testing in Section 3.3.
2.2 Satellite data acquisition and application

The surface reflectance collection 1 products of Landsat 5 TM, 7 ETM+, and 8 OLI were collected from the United States Geological Survey (USGS) website (https://earthexplorer.usgs.gov). These products have been geometrically and atmospherically corrected by USGS (Masek et al. 2006; Vermote et al. 2016). Landsat 5 TM and Landsat 7 ETM+ surface reflectance were generated using the Landsat ecosystem disturbance adaptative processing system and 6S radiative transfer model (Masek et al. 2006). Landsat 8 OLI surface reflectance was generated using the land surface reflectance code and 6S radiative transfer model (Vermote et al. 2016). Each image was quality controlled by removing the shadow, cloud, and snow pixels through the Fmask algorithm (Zhu and Woodcock 2012). The abnormal reflectance on water surfaces from the Fmask algorithm and atmospheric correction were excluded with thresholds (i.e. surface reflectance at green band of <0.02 or surface reflectance at any visible bands of >0.5) from Hou et al. (2022). The images with sun glint were manually excluded to ensure accurate SDD estimation. The surface reflectance was divided by a constant π (3.14) to obtain the remote sensing reflectance (\(R_{rs}\)) (Zhang et al. 2021c).

The spatial and temporal variations of SDD were acquired according to the following steps. First, the normalized difference water index (NDWI) was used to delineate water and land interfaces. The valley of the two peaks from the NDWI histogram was considered the threshold to extract the water area (Hou et al. 2017). Second, the extracted images were normalized before being entered into the RF and extreme gradient boosting (XGBoost) algorithms in Python. Third, outputs from the RF and XGBoost algorithms were denormalized and transformed into SDD values for further application.

2.3 Matching rules

The matchups between satellite data and in situ SDD were based on the following rules: 1) the time interval was set to within 3 days between an in situ SDD measurement and Landsat observation (Olmanson, Bauer, and Brezonik 2008); 2) the reflectance with the coefficient of variation>10% in a 3 × 3 pixel window was excluded (Cao et al. 2020); and 3) the average value in a 3 × 3 pixel window centered at the sampling location was employed.

2.4 Machine learning methods

Four machine learning methods were used to retrieve the SDD: back propagation neural network (BP), SVR, RF, and XGBoost. The BP and SVR are single models, and RF and XGBoost are ensemble learning models.

In BP modeling, neurons in each layer are connected to those in the next layer by connection weights (\(W^{layer}\)). Random weights were assigned at the start of the learning. The error and output are then calculated to update the \(W^{layer}\) until the criterion is reached. The weighted output is then summed to produce the final prediction by using a nonlinear transfer function.

The SVR works by mapping the input data nonlinearly to a feature space with more dimensions. It is a supervised method for completing optimal regression tasks. In the SVR training process, the input layers are referenced by coordinates to form the support vectors. Once these support vectors have been created, the SVR can then automatically predict the new observations (Li and Simske 2010).

The XGBoost is a scalable gradient boosting machine model (Chen et al. 2015). Comparing to the traditional gradient boosting machine model, the XGBoost technique is robust to noisy predictors and generally avoids overfitting owing to its regularized boosting processing (Chen et al. 2015). In addition, algorithmic optimizations and parallel and distributed computing increase the rate of XGBoost learning, which enables faster model exploration than existing popular solutions (Chen et al. 2015). The XGBoost algorithm adds trees continuously to split features, until the stopping criterion is reached, and the importance score in each tree is summed to achieve the final prediction.

The RF is another ensemble learning model (Breiman 2001). In RF processes, the whole training set is divided into several sub-training sets based on the bootstrap method, and each sub-training set is randomly selected to construct a decision tree for independent training. Each model prediction is integrated by averaging the model result to obtain the final prediction.

The training and validation processes of the above models were implemented using the scikit-learn and
XGBoost libraries coded in Python language. In this study, the RF model was established with a mean absolute error (MAE) criterion and 1000 iterations. The SVR model was established with a nonlinear radial basis function kernel. The XGBoost model was established with the objective of “reg:linear” and the loss function of root mean square error (RMSE). The designed BP had five layers, including one input layer, three hidden layers with tansig activation function and one output layer with a linear activation function. We established independent models with different hyperparameters groups. The optimal hyperparameters group was determined by the performance of the corresponding model that reported the best accuracy for validation dataset.

2.5 Model development and implementation

We incorporated the visible, near-infrared (NIR) and short-wave infrared (SWIR) bands of Landsat 5/7/8 data and four band ratios that achieved better predictive performance (blue/green, blue/red, green/red and red/NIR) into the model for global SDD estimation of lakes and reservoirs. The importance calculation is based on out-of-bag (oob) error (Breiman 2001). If the feature is important, its change will greatly affect the oob error. For instance, the error E1 can be obtained by using oob data for a tree, and then randomly change the feature F1 in oob data and keep other features unchanged to obtain the error E2. E1-E2 can be used to describe the importance of feature F1. The relative importance of each feature, defined as the ratio of each feature to the sum of importance scores, was calculated through the RF model. The details of the variable importance scores for the major predictors are shown in Fig. S2. The overall framework of SDD mode development, testing, comparison and robustness analysis is briefly summarized as follows (Figure 1):

(1) The training dataset (70%) was randomly selected from the whole dataset, and the remaining data (30%) were used to validate the model performance. Four machine learning

![Figure 1. The overall framework of the study. The content of this study included SDD mode development, testing, comparation and robustness analysis.](image)
methods (i.e. SVR, BP, RF and XGBoost) were used to develop the SDD models. The training and validation datasets were normalized before being entered into the machine learning methods.

(2) The two ensemble machine learning methods were applied to the long-term testing dataset in some typical lakes and reservoirs.

(3) For model comparison, the frequently-used empirical band ratio (EBR) model (Olmanson, Bauer, and Brezonik 2008), empirical single band (ESB) model (Allee and Johnson 1999) and quasi-analytical algorithm (QAA) (Lee et al. 2016) were also evaluated. Specifically, the EBR, ESB and QAA models were defined as follows:

\[ SDD_{EBR} = a \times (R_{RS}(Blue)/R_{RS}(Red)) + b \]  
(1)

\[ SDD_{ESB} = a \times R_{RS}(Red)^3 + b \times R_{RS}(Red)^2 + c \times R_{RS}(Red) + d \]  
(2)

\[ SDD_{QAA} = \frac{1}{2.5\text{Min}(K_d^o)} \ln \left( \frac{0.14 - R_{RS}^{\text{det}}}{0.013} \right) \]  
(3)

where \( K_d^o \) in equation (3) is the diffuse attenuation coefficient at five wavelengths \( (K_d(443), K_d(481), K_d(530), K_d(554), \text{and } K_d(656)) \) which could be estimated from Lee et al. (2016), with \( R_{RS}^{\text{det}} \), the remote sensing reflectance corresponding to the wavelength of minimum \( K_d^o \).

(4) The robustness of the machine learning models was tested by analyzing the error evolution of the validation dataset when altering the number of in situ measurements in the model training or adding noise of 1–100% for input parameters.

2.6 Statistical analyses

Statistical indicators (e.g. maximum, minimum, average, and standard deviation values) were calculated in Excel software. To evaluate the model performance, we used MRE, MAE, RMSE, normalized root mean square error (NRMSE), log transformed bias and the coefficient of determination \( (R^2) \) in this study:

\[ \text{MRE} = \frac{100\%}{M} \sum_{i=1}^{M} \left| \frac{\text{SDD}_{\text{est}_i} - \text{SDD}_{\text{meas}_i}}{\text{SDD}_{\text{meas}_i}} \right| \]  
(4)

\[ \text{MAE} = \frac{1}{M} \sum_{i=1}^{M} |\text{SDD}_{\text{est}_i} - \text{SDD}_{\text{meas}_i}| \]  
(5)

\[ \text{RMSE} = \sqrt{\frac{1}{M} \sum_{i=1}^{M} (\text{SDD}_{\text{est}_i} - \text{SDD}_{\text{meas}_i})^2} \]  
(6)

\[ \text{NRMSE} = \frac{100\% \times \sqrt{\frac{1}{M} \sum_{i=1}^{M} (\text{SDD}_{\text{est}_i} - \text{SDD}_{\text{meas}_i})^2}}{\frac{1}{M} \sum_{i=1}^{M} \text{SDD}_{\text{meas}_i}} \]  
(7)

\[ \text{bias} = 10 \times \left( \sum_{i=1}^{M} \log_{10} \left( \frac{\text{SDD}_{\text{meas}_i}}{\text{SDD}_{\text{est}_i}} \right) \right) / M \]  
(8)

where SDD_{meas} and SDD_{est} are the measured and estimated SDDs, respectively, and M is the number of values.

3 Result

3.1 Global SDD distribution from in situ observations

The in situ SDD observations were primarily located in America, Europe and Asia with a wide range of 0.01–36.71 m and an average of 1.60 m (Figure 2a). The statistical information of the SDD for the five continents is summarized graphically in box plots (Figure 2b). North America has an obviously wider SDD range than the other four continents. Oceania \((N = 17)\) has a relatively high SDD, with an average SDD of 6.35 m. However, the limited data can not completely represent SDD dynamics of the Oceania. South America \((N = 376)\) and North America \((N = 30,465)\) have moderate SDD values, with average SDDs of 3.34 m and 2.77 m, respectively. Europe \((N = 9955)\), Africa \((N = 386)\) and Asia \((N = 5047)\) have relatively low SDD, with average SDDs of 1.53 m, 2.01 m and 2.07 m, respectively (Figure 2b). Specifically, lakes with high SDDs were mainly located in eastern North America at longitudes of 80°-90° W and Tibet at longitudes of 80°-95° E, and few high SDDs were located in South America. In contrast, the
lakes in the middle region of North America at longitudes of 98°-110° W and the Yangtze River basin of China at longitudes of 105°-125° E generally had low SDDs (Figure 2).

3.2 Model development

The concurrent in situ SDD observations with Landsat 5 TM, Landsat 7 ETM+ and Landsat 8 OLI were 4099, 2420 and 1249, respectively (Table S2). The spatial distribution of the matchups for Landsat 5/7/8 images and a histogram of the SDDs are shown in Fig. S3. The SDD dataset covered a wide range of 0.01-18.25 m with a median value of 1.10 m (Table S3).

Scatterplots of the measured vs. estimated Ln-transformed SDDs (Figures 3, 4, Fig. S4 and Fig. S5) show the model performances for Landsat 5/7/8 data. The statistical information of the model performances is shown in Table S4. The two ensemble learning models (i.e. XGBoost and RF) showed higher accuracies for the training dataset with an average MAE = 0.21 m, MRE = 14%, RMSE = 0.44 m and $R^2 = 0.96$, than the two single models (i.e. SVR and BP), with an average MAE = 0.66 m, MRE = 52%, RMSE = 1.24 and $R^2 = 0.66$. In addition, the validation results of the two ensemble learning models also showed outperformance, with an average MAE = 0.66 m, MRE = 65% and RMSE = 1.25 m over the two single models (i.e. SVR and BP), with an average MAE = 0.79 m, MRE = 67% and RMSE = 1.63 m. Regarding the satellite platforms, the performance of Landsat 8, with an average MAE = 0.43 m, MRE = 34% and RMSE = 0.71 m, was obviously better than that of Landsat 5/7, with an average MAE = 0.66 m, MRE = 57% and RMSE = 1.36 m, for both the training and validation datasets.

3.3 Model test with a long-term SDD dataset in typical lakes and reservoirs

We therefore applied the two ensemble machine learning methods to the long-term SDD dataset in some typical lakes and reservoirs (Fig. S1) to test their performance and stability. Figures 5 and 6 show a time-series comparison of Landsat estimated and in situ measured SDDs in Lake Taihu, Xin’anjiang Reservoir, Lake Kasumigaura, Lake Mjosa and three USA reservoirs based on XGBoost and RF models. In situ SDDs with high frequencies were linearly plotted to demonstrate both annual and inter-annual variation (Figures 5 and 6a-6c). The in situ SDD during special seasons was dotty plotted to demonstrate interannual variations (Figure 6d-6h).

A clear seasonal variation in SDD, lower in winter and higher in summer, can be seen from satellite-estimated SDD values, which was consistent with the in situ observed pattern in Lake Taihu and
Xin’anjiang Reservoir (Figure 5). From the compared information of long-term in situ and estimated SDD in Table S5, we found that the average observation frequency of Landsat series data (594) was obviously higher than that of field surveys (322) during the period of 1985–2020. In addition, the satellite data recorded abundant water conditions with a wider SDD range of 0.08–23.71 m than that of field surveys with an SDD range of 0.1–15.3 m (Table S5).

The average SDD of each lake/reservoir estimated from the satellite was very close to the in situ SDD values (Table S5). Regression analysis showed that the slopes between in situ SDDs and XGBoost-estimated values (1.03) or RF-estimated values (0.94) were close to 1, the $R^2$ values were more than 0.99, and the MRE values were less than 10% (Figure 7). All these results suggested that the XGBoost and RF models could result in reasonable and nearly identical SDD values for different waters.

3.4 Model performances comparison and robustness analyses

For Landsat 5/7, the performance of the four machine learning models and two widely used empirical models are summarized in Figure 8a and 8b. Although the
red band and blue-red ratio achieve higher relative importance than other single bands and band ratios (Fig. S2), the empirical algorithms based on these two features performed much worse than machine learning methods in the final SDD prediction. In addition, the two ensemble models (RF and XGBoost) performed better with MREs of approximately 30% than the other two single machine learning models (BP and SVR) with MREs of approximately 50%.

For Landsat 8, the performance of the four machine learning models, two empirical models and the QAA model are summarized in Figure 8c. Overall, the empirical algorithms performed much worse than the machine learning methods. The two ensemble models (RF and XGBoost) performed well for all datasets with MREs below 30%, while the other two machine learning models (BP and SVR) achieved a performance comparable to that of the QAA models with MREs of approximately 55%.

The residuals plots between Landsat 5/7/8 estimation and in situ SDD further confirmed the different performances for various models (Figs. S6-S8). The two single machine learning models (BP and SVR), two empirical modes (EBR and ESB) and the QAA model showed overestimation at low SDD levels and underestimation at high SDD levels. In contrast, the
two ensemble models (RF and XGBoost) provided stable performance across the full dynamic range of SDD.

The robustness of the four machine learning models was tested by altering the number of in situ measurements in the model training in this context (Fig. S9). The NRMSE of the test dataset was averaged over 100 trials with random sample selections. Extremely abnormal data, such as NRMSE beyond 5 times the standard deviation, were excluded. The four machine learning models showed gradual improvement in terms of accuracy as the training sample ratio increased. In addition, RF and XGBoost outperformed SVR and BP in a wide range of training sample ratios. The BP and SVR models have stricter sample size requirements than the RF and XGBoost models. The performance of the BP model decreased rapidly with decreasing sample size; in particular, it decreased significantly when working with the reduced datasets (training sample ratio < 90%) from Landsat 8 (Fig. S9c). At least 870 in situ measurements can ensure an acceptable accuracy of the BP model with NRMSE below 100% for a test dataset. In contrast, the performances of RF and XGBoost are less sensitive to the varying sample size, and a training sample ratio of 10% can ensure that the NRMSEs of the RF and

Figure 5. The time-series in situ and estimated SDD based on the XGBoost and RF algorithms from different waters in China. a-c Lake Taihu. d-f Xin’anjiang Reservoir.
Figure 6. The time-series in situ and estimated SDD based on the XGBoost and RF algorithms from different waters. a-c Lake Kasumigaura. d-e Lake Mjosa. f Fellows Reservoir. g Shayne Reservoir. h Stockton Reservoir.
Figure 7. Scatterplots of the in situ measured vs. estimated SDD in typical lakes and reservoirs. Each point (error bar) represents the average (standard deviation) value of each lake/reservoir from the long-term observation in Figures 5 and 6.

Figure 8. Model performances of SVR, BP, XGBoost, RF, EBR, ESB and QAA in Taylor plots for different Landsat sensors. a Landsat 5. b Landsat 7 (b). c Landsat 8. The closer model points to the coordinate origin indicate higher accuracy of the models.
XGBoost models are below 100% for a validation dataset.

To determine how the input parameters influence the model performance, random errors of 1%–100% in the input parameters were artificially introduced during model training, and the SDD was calculated with each of the established models. The NRMSE of the validation dataset was also averaged over 100 different trials with different random sample selections. The NRMSE for the four machine learning models gradually increased with increasing introduced error. In addition, the NRMSEs of the SVR and BP models were 100 times larger than those of the RF and XGBoost models over a wide range of scenarios (amount of introduced error). When the introduced error increased from 1% to 100%, the XGBoost and RF models maintained good performance, with NRMSEs below 100%. In contrast, the SVR and BP models showed large variations in their NRMSEs of more than 100 fold (Fig. S10).

4. Discussion

4.1 Algorithm uncertainties along with lake trophic status

We determined the algorithm uncertainties along with the water optical types. The water SDDs were first transformed into a lake trophic index using an empirical relationship [TSI_{SDD} = 60–14.41 \times \text{Ln}(SDD)] proposed by Carlson (1977). The TSI_{SDD} was thus divided into three trophic levels (e.g. oligotrophic: 0–40 (SDD > 4.0 m); mesotrophic: 40–50 (2.0 m < SDD ≤ 4.0 m), eutrophic: 50–100 (SDD ≤ 2.0 m)) according to the taxonomy proposed by Olmanson, Bauer, and Brezonik (2008).

Bias was calculated for the full dataset as well as the above three trophic levels (Table 1). It should be noted that bias values closer to unity indicate less biased results and values less than one indicate negative biases (Seegers et al. 2018). For the full dataset, XGBoost and RF reported less biased results than the other models, with bias values closer to unity (Landsat 5 TM: 1.07, Landsat 7 ETM+: 1.08, and Landsat 8 OLI: 1.05).

Algorithm performance varied for each trophic level (Table 1). The two single machine learning models (BP and SVR) and two empirical modes (EBR and ESB) reported underestimated SDD results with bias<1 in eutrophic water and overestimated SDD results with bias>1 in oligotrophic water. The two empirical modes (EBR and ESB) reported more biased results with bias values far from unity. Collective consideration of bias designates XGBoost and RF as two good performers for the full dataset as well as for each trophic level.

4.2 Advantages and further optimization

The relationships between the SDD and optically active parameters are complex and variable (Jiang et al. 2019). Each band in the empirical models is sensitive to different water constituents, which leads to underestimation or overestimation of the SDD for different optical waters (Balasubramanian et al. 2020). The accuracy of the QAA algorithm was low for global lakes and reservoir datasets primarily because the empirical parameters from ocean water in the semi-analytical algorithm are unsuitable for the optically complex waters of global lakes and reservoirs. Therefore, it has proven difficult to retrieve SDDs at a global scale using the most widely employed empirical (EBR and ESB) and QAA models (Figure 8).

The machine learning models in this research were based on a broad water condition, with SDD values ranging from approximately 0.01 m to over 18 m. Machine learning methods can capture nonlinear relationships between multiple input features (e.g. remote sensing from single band, band ratio, and band combination) and SDD values and generally produce better performance than empirical and semi-analytical methods (Figure 8). This study also found that the ensemble models of XGBoost and RF were more robust than the other two single models (BP and SVR) (Fig. S9 and Fig. S10). The uncertainties from
added noise in XGBoost and RF modeling might be corrected through boosting and cross-validation processing (Shen et al. 2020). In addition, the XGBoost and RF models could effectively avoid overfitting and interference from outliers and noise during the training process (Shen et al. 2020). Therefore, the two ensemble models (RF and XGBoost) performed reasonably better than the single models (BP and SVR). Considering the good performance of the XGBoost and RF methods, these two ensemble models have great potential for water monitoring applications in global lakes and reservoirs. Since the band settings of earth resource satellite data (such as Sentinel 2 A/B, Gao-Fen series, CBERS-2, ZY series and HJ A/B) and the newly launched Landsat 9 are similar to those of Landsat 5/7/8, the experience gained from this study could be used for current earth resource satellite data to derive spatiotemporal SDD distributions in global lakes and reservoirs. Moreover, these earth resource satellite data were characterized by numerous improvements (such as high spatial resolution, and shorter revisit time), which make them appropriate for various inland waters, especially for those with small water surfaces.

As data-driven methods, the training sample size greatly impact the accuracy of machine learning models (Fig. S9). Although a robust analysis indicated that our dataset was sufficient for model construction, our training and validation samples were not evenly distributed (i.e. too many samples had SDD values around the median, while few samples had SDDs at the two extreme ends of the distribution). Some uncertainties observed for extremely clear and turbid waters (Figs. S6-S8) were caused by their scarcity in the training samples. For this reason, more in situ SDD measurements in extremely clear and turbid waters might improve SDD estimation using machine learning methods. Therefore, we welcome more in situ SDDs from scientific researchers and environmental monitoring departments to contribute to and enrich our SDD dataset, which will further optimize and improve our model accuracies.

4.3 Model implication for water environmental management
Accurate long-term water quality monitoring on a national or global scale plays an increasingly important role in designing effective water environmental management. The construction and validation of SDD remote sensing estimation model for global lakes and reservoirs in this research demonstrated that the XGBoost and RF methods with Landsat images can obtain comprehensive spatial-temporal coverage of water quality. The Landsat satellites provide global coverage and the longest record of earth observations, which are the result of data from documented historical Landsat 5 TM and Landsat 7 ETM+, the current Landsat 8 OLI and the newly launched Landsat 9 OLI (Che, Zhang, and Liu 2021; Masek et al. 2020). The results of a long-term Landsat water clarity database will help reveal the historical evolution process and monitor the current situation of water quality. The proposed estimation models will play a crucial role in some remote lakes and reservoirs where available field data are scarce. For instance, the United Nations Sustainable Development Goal (http://www.sdg6data.org/indicator/6.3.2) and European Water Framework Directive (https://www.iksr.org/en/eu-directives/european-water-framework-directive) focus on sustainable water resource protection and freshwater quality evaluation. Harsh environmental conditions within the studied regions and the costs of sampling and laboratory analysis mean that only a few countries or territories are able to provide sufficient data. Our results are promising for solving these problems quickly and economically.

Providing future forecast analysis of water quality due to the effects of climate change and localized anthropogenic activities is also a basic need for lake and reservoir management (Deeds et al. 2021) and has recently been given increasing attention in various research fields such as limnology and environmental science (Cusack et al. 2016; Mao, Lee, and Choi 2009; Saber, James, and Hayes 2019). Climate influences on water quality are multifaceted. For instance, wind affects water quality through resuspension of lake sediment in shallow lakes (Cao et al. 2017; Hou et al. 2017). Temperature affects the growth of algae and regulates thermal structure (Liu, Feng, and Wang 2019; Qin et al. 2019). Rainfall affects water quality through dilution and leaching actions (Bonansea et al. 2015; Hou et al. 2017). Watershed land use and vegetation cover, generally related to localized anthropogenic activities, also play important roles in changing water quality by altering watershed erosion or nutrient loading (Olmanson, Breznik, and Bauer 2014; Zhang et al. 2021b). The results of the Landsat water
clarity database hopefully disentangle the effects of climate change and localized anthropogenic activities on water quality and can predict future trends and temporal shifts in water quality by simulating future climate with a climatic model; these results may help inform decisions for development policy.

5. Conclusion

This research compiled global in situ SDD datasets and developed and validated machine learning methods with general applicability to estimate the SDD for global lakes and reservoirs from satellites through multiperspective comparisons. Overall, machine learning methods generally produce better performance (MRE decrease more than 25%) than empirical and semianalytical methods. The ensemble models of XGBoost and RF were more robust than the single models (BP and SVR) throughout diverse statistical tests and assessments. In addition, the XGBoost and RF models could result in reasonable and nearly identical patterns for different waters, which could be used for long-term global-level SDD estimation and water quality assessment. More studies are needed to map the temporal-spatial distribution of the SDD, elucidate the driving mechanisms of SDD dynamics, and predict future SDD change scenarios for global lakes and reservoirs based on our machine learning models. In addition, more in situ SDDs, especially in extremely clear and turbid waters, are encouraged to improve SDD estimation precision and model applicability.

Acknowledgements

We would like to thank the USGS for providing all Landsat data. We would also like to thank Breiman, Leo and TianQi Chen for developing and sharing the scikit-learn and XGBoost libraries. We would like to express our gratitude to the executive editor and four anonymous reviewers for their critical comments and constructive suggestions.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by the National Natural Science Foundation of China [42007160]; National Natural Science Foundation of China [41922005]; National Natural Science Foundation of China [41771472].

Author contributions

Yibo Zhang and Yunlin Zhang designed the study. Yibo Zhang, Kun Shi, Xiaon Sun, Na Li, Weijia Wang conducted the experiments. Yibo Zhang, Yongqiang Zhou, Wei Zhi, Mingliang Liu, Yuan Li, Jian Zhou and Huiyun Li collected the data. Yibo Zhang, Kun Shi and Xiao Sun analyzed the data. Yibo Zhang, Yunlin Zhang, Guangwei Zhu, Boqiang Qin and Erik Jeppesen wrote and edited the manuscript.

Data availability

The in situ SDD in this study can be found in Github repositories with the identifier (https://github.com/YiboNiglas/Machine_learning_SDD.git).

References

Alikas, K., and S. Kratzer. 2017. “Improved Retrieval of Secchi Depth for optically-complex Waters Using Remote Sensing Data.” Ecological Indicators 77: 218–227. doi:10.1016/j.ecolind.2017.02.007.

Allee, R. J., and J. E. Johnson. 1999. “Use of Satellite Imagery to Estimate Surface Chlorophyll a and Secchi Disc Depth of Bull Shoals Reservoir, Arkansas, USA.” International Journal of Remote Sensing 20 (6): 1057–1072. doi:10.1080/014311699212849.

Balasubramanian, S. V., N. Pahlevan, B. Smith, C. Binding, J. Schalles, H. Loisel, D. Gurlin, et al. 2020. “Robust Algorithm for Estimating Total Suspended Solids (TSS) in Inland and Nearshore Coastal Waters.” Remote Sensing of Environment 246:111768. doi:10.1016/j.rse.2020.111768.

Belgiu, M., and L. Dragut. 2016. “Random Forest in Remote Sensing: A Review of Applications and Future Directions.” ISPRS Journal of Photogrammetry and Remote Sensing 114: 24–31. doi:10.1016/j.isprsjprs.2016.01.011.

Bonasea, M., M. C. Rodriguez, L. Pinotti, and S. Ferrero. 2015. “Using multi-temporal Landsat Imagery and Linear Mixed Models for Assessing Water Quality Parameters in Rio Tercero Reservoir (Argentina).” Remote Sensing of Environment 158: 28–41. doi:10.1016/j.rse.2014.10.032.

Breiman, L. 2001. “Random Forests.” Machine Learning 45 (1): 5–32. doi:10.1023/A:1010933404324.

Cao, Z. G., H. T. Duan, L. Feng, R. H. Ma, and K. Xue. 2017. “Climate- and human-induced Changes in Suspended Particulate Matter over Lake Hongze on Short and Long Timescales.” Remote Sensing of Environment 192: 98–113. doi:10.1016/j.rse.2017.02.007.
Cao, Z. G., R. H. Ma, H. T. Duan, N. Pahlevan, J. Melack, M. Shen, and K. Xue. 2020. “A Machine Learning Approach to Estimate chlorophyll-A from Landsat-8 Measurements in Inland Lakes.” Remote Sensing of Environment 248: 111974. doi:10.1016/j.rse.2020.111974.

Capuzzo, E., D. Stephens, T. Silva, J. Barry, and R. M. Forster. 2015. “Decrease in Water Clarity of the Southern and Central North Sea during the 20th Century.” Global Change Biology 21 (6): 2206–2214. doi:10.1111/gcb.12584.

Carlson, R. E. 1977. “A Trophic State Index for Lakes.” Limnology and Oceanography 22 (2): 361–369. doi:10.4319/lo.1977.22.0361.

Chawla, I., L. Karthikeyan, and A. K. Mishra. 2020. “A Review of Remote Sensing Applications for Water Security: Quantity, Quality, and Extremes.” Journal of Hydrology 585. doi:10.1016/j.jhydrol.2020.124826.

Che, X. H., H. K. Zhang, and J. P. Liu. 2021. “Making Landsat 5, 7 and 8 Reflectance Consistent Using MODIS nadir-BrDF Adjusted Reflectance as Reference.” Remote Sensing of Environment 262. doi:10.1016/j.rse.2021.112517.

Chen, T. Q., T. He, M. Benesty, V. Khotilovich, Y. Tang, and H. Cho. 2015. “Xgboost: Extreme Gradient Boosting.” R Package Version 0.4-2 1 (4): 1–4.

Cusack, C., T. Dabrowski, K. Lyons, A. Berry, G. Westbrook, R. Salas, C. Duffy, G. Nolan, and J. Silke. 2016. “Harmful Algal Bloom Forecast System for SW Ireland. Part II: Are Operational Oceanographic Models Useful in a HAB Warning System.” Harmful Algae 53: 86–101. doi:10.1016/j.hal.2015.11.013.

Deeds, J., A. Amirbahman, S. A. Norton, L. C. Bacon, and R. A. Hoveland. 2021. “Shifting Baselines and Cross-scale Drivers of Lake Water Clarity: Applications for Lake Assessment.” Limnology and Oceanography 999: 1–14. doi:10.1002/lio.11873.

Doron, M., M. Babin, O. Hembise, A. Mangin, and P. Garnesson. 2011. “Ocean Transparency from Space: Validation of Algorithms Estimating Secchi Depth Using MERIS, MODIS and SeaWiFS Data.” Remote Sensing of Environment 115 (12): 2986–3001. doi:10.1016/j.rse.2011.05.019.

Downing, J. A., S. Polasky, S. O. Mollmeast, and S. C. Newbold. 2021. “Protecting Local Water Quality Has Global Benefits.” Nature Communications 12 (1): 2709. doi:10.1038/s41467-021-22863-3.

Giannini, F., B. P. V. Hunt, D. Jacoby, and M. Costa. 2021. “Performance of OLCI Sentinel-3A Satellite in the Northeast Pacific Coastal Waters.” Remote Sensing of Environment 256. doi:10.1016/j.rse.2021.112317.

Grant, L., L. Vanderkelen, L. Gudmundsson, Z. Tan, M. Perroud, V. M. Stepanenko, A. V. Debolskiy, et al. 2021. “Attribution of Global Lake Systems Change to Anthropogenic Forcing.” Nature Geoscience 14 (11): 849–854. doi:10.1038/s41561-021-00833-x.

Guo, H., S. Tian, J. Jeanne Huang, X. Zhu, B. Wang, and Z. Zhang. 2022. “Performance of Deep Learning in Mapping Water Quality of Lake Simcoe with long-term Landsat Archive.” ISPRS Journal of Photogrammetry and Remote Sensing 183: 451–469. doi:10.1016/j.isprsjprs.2021.11.023.

He, Y., Z. Lu, W. J. Wang, D. Zhang, Y. L. Zhang, B. Q. Qin, K. Shi, and X. F. Yang. 2022. “Water Clarity Mapping of Global Lakes Using a Novel Hybrid deep-learning-based Recurrent Model with Landsat OLI Images.” Water Research 215: 118241. doi:10.1016/j.watres.2022.118241.

Hou, X., L. Feng, Y. Dai, C. Hu, L. Gibson, J. Tang, Z. Lee, et al. 2022. “Global Mapping Reveals Increase in Lacustrine Algal Blooms over the Past Decade.” Nature Geoscience 15 (2): 130–134. doi:10.1038/s41561-021-00887-x.

Hou, X. J., L. Feng, H. T. Duan, X. L. Chen, D. Y. Sun, and K. Shi. 2017. “Fifteen-year Monitoring of the Turbidity Dynamics in Large Lakes and Reservoirs in the Middle and Lower Basin of the Yangtze River, China.” Remote Sensing of Environment 190: 107–121. doi:10.1016/j.rse.2016.12.006.

Jane, S. F., G. J. A. Hansen, B. M. Kraemer, P. R. Leavitt, J. L. Mincer, R. L. North, R. M. Pilla, et al. 2021. “Widespread Deoxygenation of Temperate Lakes.” Nature 594 (7861): 66–70. doi:10.1038/s41586-021-03550-y.

Jiang, D. L., B. Matsushita, F. Setiawan, and A. Vundo. 2019. “An Improved Algorithm for Estimating the Secchi Disk Depth from Remote Sensing Data Based on the New Underwater Visibility Theory.” ISPRS Journal of Photogrammetry and Remote Sensing 152 (1): 13–23. doi:10.1016/j.isprsjprs.2019.04.002.

Kabiri, K. 2022a. “Estimation of the Secchi Disk Depth from the NASA MODIS-Aqua Diffuse Attenuation Coefficient Data in the Northern Persian Gulf and the Gulf of Oman: A Spatiotemporal Assessment.” Regional Studies in Marine Science 52: 102359. doi:10.1016/j.rsma.2022.102359.

Kabiri, K. 2022b. “Remote Sensing of Water Clarity in the near-shore Zone Using a cross-sensor-based Method: Feasibility Study: Kish Island, Persian Gulf.” Journal of Coastal Conservation 26 (4): 26. doi:10.1007/s11852-022-00875-2.

Kabiri, K., and M. Moradi. 2016. “Landsat-8 Imagery to Estimate Clarity in near-shore Coastal Waters: Feasibility Study - Chahabar Bay, Iran.” Continental Shelf Research 125: 44–53. doi:10.1016/j.csr.2016.06.016.

Karlsson, J., P. Byström, J. Ask, P. Ask, L. Persson, and M. Jansson. 2009. “Light Limitation of nutrient-poor Lake Ecosystems.” Nature 460 (7254): 506–510. doi:10.1038/nature08179.

Keeler, B. L., S. Polasky, K. A. Brauman, K. A. Johnson, J. C. Finlay, A. O’Neill, K. Kovacs, and B. Dalzell. 2012. “Linking Water Quality and well-being for Improved Assessment and Valuation of Ecosystem Services.” Proceedings of the National Academy of Sciences of the United States of America 109 (45): 18619–18624. doi:10.1073/pnas.1215991109.

Keeler, B. L., S. A. Wood, S. Polasky, C. Kling, C. T. Filstrup, and J. A. Downing. 2015. “Recreational Demand for Clean Water: Evidence from Geotagged Photographs by Visitors to Lakes.” Frontiers in Ecology and the Environment 13 (2): 76–81. doi:10.1890/140124.

Lee, Z. P., S. L. Shang, L. Qi, J. Yan, and G. Lin. 2016. “A semi-analytical Scheme to Estimate Secchi-disk Depth from
Landsat-8 Measurements.” Remote Sensing of Environment 177: 101–106. doi:10.1016/j.rse.2016.02.033.
Li, D., and S. Simske. 2010. “Example Based single-frame Image super-resolution by Support Vector Regression.” Journal of Pattern Recognition Research 1 (1): 104–118. doi:10.13176/11.253.
Li, J., E. Garshick, J. E. Hart, L. X. Li, L. H. Shi, A. Al-Hemoud, S. D. Huang, and P. Koutrakis. 2021a. “Estimation of Ambient PM2.5 in Iraq and Kuwait from 2001 to 2018 Using Machine Learning and Remote Sensing.” Environment International 151: 106445. doi:10.1016/j.envint.2021.106445.
Li, X. H., D. K. Yang, J. S. Yang, G. Zheng, G. Q. Han, Y. Nan, and W. Q. Li. 2021b. “Analysis of Coastal Wind Speed Retrieval from CYGNSS Mission Using Artificial Neural Network.” Remote Sensing of Environment 260: 112454. doi:10.1016/j.rse.2021.112454.
Lisi, P. J., and C. L. Hein. 2019. “Eutrophication Drives Divergent Water Clarity Responses to Decadal Variation in Lake Level.” Limnology and Oceanography 64 (S1): S49–S59. doi:10.1002/ino.11095.
Liu, C., L. P. Zhu, J. S. Li, J. B. Wang, J. T. Ju, B. J. Qiao, Q. F. Ma, and S. L. Wang. 2021a. “The Increasing Water Clarity of Tibetan Lakes over Last 20 Years according to MODIS Data.” Remote Sensing of Environment 253: 112199. doi:10.1016/j.rse.2020.112199.
Liu, D., H. T. Duan, S. Loiselle, C. M. Hu, G. Q. Zhang, J. L. Li, H. Yang, et al. 2020. “Observations of Water Transparency in China’s Lakes from Space.” International Journal of Applied Earth Observation and Geoinformation 92: 102187. doi:10.1016/j.jag.2020.102187.
Liu, H., B. He, Y. Zhou, X. Yang, X. Zhang, F. Xiao, Q. Feng, S. Liang, X. Zhou, and C. Fu. 2021b. “Eutrophication Monitoring of Lakes in Wuhan Based on Sentinel-2 Data.” GIScience & Remote Sensing 58 (5): 776–798. doi:10.1080/15481603.2021.1940738.
Liu, X., J. Feng, and Y. Wang. 2019. “Chlorophyll a Predictability and Relative Importance of Factors Governing Lake Phytoplankton at Different Timescales.” Science of the Total Environment 648: 472–480. doi:10.1016/j.scitotenv.2018.08.146.
Luis, K. M. A., J. E. Rheuban, M. T. Kavanaugh, D. M. Glover, J. Wei, Z. Lee, and S. C. Doney. 2019. “Capturing Coastal Water Clarity Variability with Landsat 8.” Marine Pollution Bulletin 145: 96–104. doi:10.1016/j.marpolbul.2019.04.078.
Maciel, D. A., C. C. F. Barbosa, E. M. L. D. M. Novo, R. Flores Júnior, and F. N. Begliomini. 2021. “Water Clarity in Brazilian Water Assessed Using Sentinel-2 and Machine Learning Methods.” ISPRS Journal of Photogrammetry and Remote Sensing 182: 134–152. doi:10.1016/j.isprsjprs.2021.10.009.
Mantzouki, E., J. Campbell, E. van Loon, P. Visser, I. Konstantinou, M. Antoniou, G. Giuliani, et al. 2018. “A European Multi Lake Survey Dataset of Environmental Variables, Phytoplankton Pigments and Cyanotoxins.” Scientific Data 5 (1): 180226. doi:10.1038/sdata.2018.226.
Mao, J. Q., J. H. Lee, and K. W. Choi. 2009. “The Extended Kalman Filter for Forecast of Algal Bloom Dynamics.” Water Research 43 (17): 4214–4224. doi:10.1016/j.watres.2009.06.012.
Masek, J. G., E. F. Vermote, N. E. Saleous, R. Wolfe, F. G. Hall, K. F. Huemmrich, F. Gao, J. Kutler, and T. K. Lim. 2006. “A Landsat Surface Reflectance Dataset for North America, 1990–2000.” IEEE Geoscience and Remote Sensing Letters 3 (1): 68–72. doi:10.1109/LGRS.2005.857030.
Masek, J. G., M. A. Wulder, B. Markham, J. McCorkel, C. J. Crawford, J. Storey, and D. T. Jenstrom. 2020. “Landsat 9: Empowering Open Science and Applications through Continuity.” Remote Sensing of Environment 248: 111968. doi:10.1016/j.rse.2020.111968.
Msusa, A. D., D. Jiang, and B. Matsushita. 2022. “A Semi-analytical Algorithm for Estimating Water Transparency in Different Optical Water Types from MERIS Data.” Remote Sensing 14 (4): 868. doi:10.3390/rs14040868.
Mudiyanselage, S. S. J. D., A. Abd-Elrahman, B. Wilkinson, and V. Elocurs. 2022. “Satellite-derived Bathymetry Using Machine Learning and Optimal Sentinel-2 Imagery in South-West Florida Coastal Waters.” GIScience & Remote Sensing 59 (1): 1143–1158. doi:10.1080/15481603.2022.2100597.
Olmanson, L. G., M. E. Bauer, and P. L. Brezonik. 2008. “A 20-year Landsat Water Clarity Census of Minnesota’s 10,000 Lakes.” Remote Sensing of Environment 112 (11): 4086–4097. doi:10.1016/j.rse.2007.12.013.
Olmanson, L. G., P. L. Brezonik, and M. E. Bauer. 2014. “Geospatial and Temporal Analysis of a 20-year Record of Landsat-based Water Clarity in Minnesota’s 10,000 Lakes.” Journal of the American Water Resources Association 50 (3): 748–761. doi:10.1111/jawr.12138.
Page, B. P., L. G. Olmanson, and D. R. Mishra. 2019. “A Harmonized Image Processing Workflow Using Sentinel-2/MSI and Landsat-8/OLI for Mapping Water Clarity in Optically Variable Lake Systems.” Remote Sensing of Environment 231: 111284. doi:10.1016/j.rse.2019.111284.
Qin, B. Q., H. W. Paelr, J. D. Brookes, J. G. Liu, E. Jeppesen, G. W. Zhu, Y. L. Zhang, H. Xu, K. Shi, and J. M. Deng. 2019. “Why Lake Taihu Continues to Be Plagued with Cyanobacterial Blooms through 10 Years (2007–2017) Efforts.” Science Bulletin 64 (6): 354–356. doi:10.1016/j.scib.2019.02.008.
Ren, J. L., Z. B. Zheng, Y. M. Li, G. N. Lv, Q. Wang, H. Lyu, C. C. Huang, et al. 2018. “Remote Observation of Water Clarity Patterns in Three Gorges Reservoir and Dongting Lake of China and Their Probable Linkage to the Three Gorges Dam Based on Landsat 8 Imagery.” Science of the Total Environment 625:1554–1566. doi:10.1016/j.scitotenv.2018.01.036.
Saber, A., D. E. James, and D. F. Hayes. 2019. “Long-term Forecast of Water Temperature and Dissolved Oxygen Profiles in Deep Lakes Using Artificial Neural Networks Conjugated with Wavelet Transform.” Limnology and Oceanography 65 (6): 1297–1317. doi:10.1002/lno.11390.
Seegers, B. N., R. P. Stumpf, B. A. Schaeffer, K. A. Loftin, and P. J. Werdell. 2018. “Performance Metrics for the Assessment
of Satellite Data Products: An Ocean Color Case Study.” Opt Express 26 (6): 7404–7422. doi:10.1364/OE.26.007404.

Shen, M., H. T. Duan, Z. G. Cao, K. Xue, T. C. Qi, J. G. Ma, D. Liu, K. S. Song, C. L. Huang, and X. Y. Song. 2020. “Sentinel-3 OLCI Observations of Water Clarity in Large Lakes in Eastern China: Implications for SDG 6.3.2 Evaluation.” Remote Sensing of Environment 247: 111950. doi:10.1016/j.rse.2020.111950.

Shi, K., Y. L. Zhang, G. W. Zhu, X. H. Liu, Y. Q. Zhou, H. Xu, B. Q. Qin, G. Liu, and Y. M. Li. 2015. “Long-term Remote Monitoring of Total Suspended Concentration in Lake Taihu Using 250m MODIS-Aqua Data.” Remote Sensing of Environment 164: 43–56. doi:10.1016/j.rse.2015.02.029.

Song, K. S., G. Liu, Q. Wang, Z. D. Wen, L. L. Lyu, Y. X. Du, L. W. Sha, and C. Fang. 2020. “Quantification of Lake Clarity in China Using Landsat OLI Imagery Data.” Remote Sensing of Environment 243: 111800. doi:10.1016/j.rse.2020.111800.

Song, K. S., Q. Wang, G. Liu, P. A. Jacinthe, S. J. Li, H. Tao, Y. X. Du, et al. 2022. “A Unified Model for High Resolution Mapping of Global Lake (> 1 Ha) Clarity Using Landsat Imagery Data.” Science of the Total Environment 810:151188. doi:10.1016/j.scitotenv.2021.151188.

Song, X. P., W. L. Huang, M. C. Hansen, and P. Potapov. 2021. “An Evaluation of Landsat, Sentinel-2, Sentinel-1 and MODIS Data for Crop Type Mapping.” Science of Remote Sensing 3: 100018. doi:10.1016/j.srs.2021.100018.

Terrel, M. M., T. Fukushima, B. Matsuhashi, K. Yoshimura, and A. Imai. 2012. “Long-term Light Environment Variability in Lake Biwa and Lake Kasumigaura, Japan: Modeling Approach.” Limnology 13 (2): 237–252. doi:10.1007/s10201-012-0372-x.

Vermote, E., C. Justice, M. Claverie, and B. Franch. 2016. “Preliminary Analysis of the Performance of the Landsat 8/ OLI Land Surface Reflectance Product.” Remote Sensing of Environment 185: 46–56. doi:10.1016/j.rse.2016.04.008.

Wang, M. H., S. Son, and L. W. Harding. 2009. “Retrieval of Diffuse Attenuation Coefficient in the Chesapeake Bay and Turbid Ocean Regions for Satellite Ocean Color Applications.” Journal of Geophysical Research Oceans 114 (C10): C10011. doi:10.1029/2009JC005286.

Wang, S. L., J. S. Li, B. Zhang, Z. P. Lee, E. Spyrokos, L. Feng, C. Liu, et al. 2020. “Changes of Water Clarity in Large Lakes and Reservoirs across China Observed from long-term MODIS.” Remote Sensing of Environment 247:111949. doi:10.1016/j.rse.2020.111949.

Xiao, X. X., T. J. Zhang, X. Y. Zhong, W. W. Shao, and X. D. Li. 2018. “Support Vector Regression snow-depth Retrieval Algorithm Using Passive Microwave Remote Sensing Data.” Remote Sensing of Environment 210: 48–64. doi:10.1016/j.rse.2018.03.008.

Yu, Z. Y., K. Yang, Y. Luo, and Y. L. Yang. 2021. “Secchi Depth Inversion and Its Temporal and Spatial Variation analysis—A Case Study of Nine Plateau Lakes in Yunnan Province of China.” International Journal of Applied Earth Observation and Geoinformation 100: 102344. doi:10.1016/j.jag.2021.102344.

Zhang, E. Z., L. Liu, L. C. Huang, and K. S. Ng. 2021a. “An Automated, Generalized, deep-learning-based Method for Delineating the Calving Fronts of Greenland Glaciers from multi-sensor Remote Sensing Imagery.” Remote Sensing of Environment 254: 112265. doi:10.1016/j.rse.2021.112265.

Zhang, Y. B., K. Shi, Y. L. Zhang, M. J. Moreno-Madrinán, X. Xu, Y. Q. Zhou, B. Q. Qin, G. W. Zhu, and E. Jeppesen. 2021b. “Water Clarity Response to Climate Warming and Wetting of the Inner Mongolia-Xinjiang Plateau: A Remote Sensing Approach.” Science of the Total Environment 796: 148916. doi:10.1016/j.scitotenv.2021.148916.

Zhang, Y. B., Y. L. Zhang, K. Shi, Y. Q. Zhou, and N. Li. 2021c. “Remote Sensing Estimation of Water Clarity for Various Lakes in China.” Water Research 192: 116844. doi:10.1016/j.watres.2021.116844.

Zhang, Y. L., B. Q. Qin, K. Shi, Y. B. Zhang, J. M. Deng, M. Wild, L. Li, et al. 2020. “Radiation Dimming and Decreasing Water Clarity Fuel Underwater Darkening in Lakes.” Science Bulletin 65 (19): 1675–1684. doi:10.1016/j.scib.2020.06.016.

Zhang, Y. L., Z. X. Wu, M. L. Liu, J. B. He, K. Shi, M. Z. Wang, and Z. M. Yu. 2015. “Thermal Structure and Response to Long-term Climatic Changes in Lake Qiandaohu, a Deep Subtropical Reservoir in China.” Limnology and Oceanography 59 (4): 1193–1202. doi:10.4319/lo.2014.59.4.1193.

Zhu, Z., and C. E. Woodcock. 2012. “Object-based Cloud and Cloud Shadow Detection in Landsat Imagery.” Remote Sensing of Environment 118: 83–94. doi:10.1016/j.rse.2011.10.028.