ISPRS Journal of Photogrammetry and Remote Sensing 183 (2022) 451–469

Contents lists available at ScienceDirect

ISPRS Journal of Photogrammetry and Remote Sensing

journal homepage: www.elsevier.com/locate/isprsjprs

Performance of deep learning in mapping water quality of Lake Simcoe with long-term Landsat archive

Hongwei Guo, Shang Tian, Jinhui Jeanne Huang *, Xiaotong Zhu, Bo Wang, Zijie Zhang

College of Environmental Science and Engineering/Sino-Canada Joint R&D Centre for Water and Environmental Safety, Nankai University, Tianjin 300457, China

ARTICLE INFO

Keywords:
Deep learning
Remote sensing
Water quality
Chlorophyll-*a*
Total phosphorous
Total nitrogen

ABSTRACT

Remote sensing provides full-coverage and dynamic water quality monitoring with high efficiency and low consumption. Deep learning (DL) has been progressively used in water quality retrieval because it efficiently captures the potential relationship between target variables and imagery. In this study, the multimodal deep learning (MDL) models were developed and rigorously validated using atmospherically corrected Landsat remote sensing reflectance data and synchronous water quality measurements for estimating long-term Chlorophyll-*a* (Chl-*a*), total phosphorus (TP), and total nitrogen (TN) in Lake Simcoe, Canada. Since TP and TN are non-optically active, their retrievals were based on the fact that they are closely related to the optically active constituents (OACs) such as Chl-*a*. We trained the MDL models with one in-situ measured data set (for Chl-*a*, $N = 315$, for TP and TN, $N = 303$), validated the models with two independent data sets ($N = 147$), and compared the model performances with several DL, machine learning, and empirical algorithms. The results indicated that the MDL models adequately estimated Chl-*a* (mean absolute error (MAE) = 32.57%, Bias = 10.61%), TP (MAE = 42.58%, Bias = −2.82%), and TN (MAE = 35.05%, Bias = 13.66%), and outperformed several other candidate algorithms, namely the progressively decreasing deep neural network (DNN), a DNN with trainable parameters similar to MDL but without splitting input features, the eXtreme Gradient Boosting, the support vector regression, the NASA Ocean Color two-band and three-band ratio algorithms, and another empirical algorithm of Landsat data in clear lakes. Using the MDL models, we reconstructed the historical spatiotemporal patterns of Chl-*a*, TP, and TN in Lake Simcoe since 1984, and investigated the effects of two water quality improvement programs. In addition, the physical mechanism and interpretability of the MDL models were explored by quantifying the contribution of each feature to the model outputs. The framework proposed in this study provides a practical method for long-term Chl-*a*, TP, and TN estimation at the regional scale.

1. Introduction

Water quality monitoring is key to maintaining water security. Traditional water quality monitoring methods mainly rely on in-situ measurements, sample collection, and lab analyses. These methods are time-consuming and labor-intensive (Bierman et al., 2011). Besides, measurements at the sampling point scale are also considered insufficient to represent the spatiotemporal characteristics of water quality in the whole water surface (Chawla et al., 2020). Remote sensing has long been recognized as a potential supplement to the traditional methods (Palmer et al., 2015) because it supports quick-access and cost-effective water quality monitoring at extensive spatiotemporal scales. As early as the 1970s, remote sensing has been used to estimate the concentration of suspended solids (SS) in inland waters (Holyer, 1978; Ritchie et al., 1976). Over the past 40 years, the developments of sensors, image correction technology, and retrieval algorithms have contributed to substantial progress in remote sensing of water quality (Matthews, 2011; Sagan et al., 2020). A large amount of research towards making better use of images, more diverse water quality parameter (WQP) retrieval, and more robust and generalized algorithm development was reported progressively.

The Landsat series of satellites provide the longest open-source earth observations, and are one of the most focused data sources in water quality retrieval (Zhang et al., 2021). Landsat 4 and 5 with Thematic

* Corresponding author at: B406, College of Environmental Science and Engineering, 38 Tongyan Rd., Haihe Education Park, Jinnan District, Tianjin 300350, China.

E-mail addresses: guohw@mail.nankai.edu.cn (H. Guo), tianshang@mail.nankai.edu.cn (S. Tian), huangj@nankai.edu.cn (J. Jeanne Huang), zhuxiaotong16@mails.ucas.ac.cn (X. Zhu), curry010519@163.com (B. Wang), zijiezhang_lhz@163.com (Z. Zhang).

https://doi.org/10.1016/j.isprsjprs.2021.11.023
Received 20 June 2021; Received in revised form 26 October 2021; Accepted 26 November 2021
Available online 6 December 2021
0924-2716/© 2021 International Society for Photogrammetry and Remote Sensing, Inc. (ISPRS). Published by Elsevier B.V. All rights reserved.

Mapper (TM) and Landsat 7 with Enhanced Thematic Mapper Plus (ETM+) were launched in the 1980s and 1990s successively. Since then Landsat has provided multispectral data with six spectral bands (450–2350 nm) with a spatial resolution of 30 m and a 16-day repeat cycle. A series of studies on water quality retrieval using TM and ETM+ data were conducted (Odermatt et al., 2012; Xiong et al., 2020). Griffin et al., (2018) developed empirical models using TM and ETM+ surface reflectance (SR) and chromophoric dissolved organic matter (CDOM) data derived from discharge-constituent regression-based models for CDOM and dissolved organic carbon estimation in major Arctic rivers. Setiawan et al., (2019) estimated the Secchi disk depth (SDD) of Lake Maninjau (Indonesia) using an empirical model developed with TM and ETM+ data between 1987 and 2018. The Landsat 8 launched in 2013 is equipped with a 12-bit push-broom sensor Operational Land Imager (OLI). Compared to the previous 8-bit whisk-broom TM/ETM+, OLI adds Band 1 (435–451 nm) at deep blue and violet wavelengths and significantly improves the signal to noise ratio (SNR) (Mancino et al., 2020), which makes OLI appropriate for application to inland waters (Pahlevan et al., 2017). It is worth to mention that the Landsat 9 launched in September 2021 and continues the legacy of long-term earth observation. The 14-bit OLI-2 onboard Landsat 9 is expected to provide improved SNR for darker targets such as coastal and inland waters. The combined Landsat 8/9 constellation is able to reduce the repeat cycle to eight days (Masek et al., 2020). To date, extensive research on water quality retrieval by OLI was conducted (Page et al., 2019; Pahlevan et al., 2017, 2014; Sagan et al., 2020). However, there exist substantial differences in design between OLI and TM/ETM+ (Mancino et al., 2020). These differences challenge the generation of spatiotemporal continuous water quality retrieval. To produce internally consistent Landsat remote sensing reflectance (R_{rs}) for aquatic application, Pahlevan et al., (2018) implemented a physics-based atmospheric correction method of TM and ETM+ data. Although this method dramatically improves the consistency between the sensors with the application of system calibration gains and provides a practical way to the long-term water quality retrieval using Landsat archive, large variability in the performance was still found (Vanhellemont, 2019). As being data-driven, machine learning (ML) models were expected to be able to learn the sensor differences with data covering these differences, which might be an alternative solution.

Chlorophyll-*a* (Chl-*a*) is one of the most relevant parameters in water quality retrieval because it quantitatively indicates the state of eutrophication and aids the evaluation of health risks of aquatic ecosystems (Hunter et al., 2009). At present, Chl-*a* estimation employs two broad categories of methods, i.e., analytical and empirical methods (Sagan et al., 2020). Analytical methods derive inherent optical properties (IOPs) such as absorption and backscattering coefficients from R_{rs} and then estimate Chl-*a* by the IOPs (Pahlevan et al., 2020). The fundamental light-water interactions enhance model generalization of the analytical methods, but also require detailed measurements and complex analyses of spectral information. Empirical methods estimate Chl-*a* directly from R_{rs} based on: 1) the blue-green (O'Reilly et al., 1998) or near-infrared (NIR)-red (Gons, 1999) band ratios; 2) the spectral shape delineated by slope or peak reflectance/absorption differences (Matthews and Odermatt, 2015); 3) the optical water types classified by clustering spectra (Spyrakos et al., 2018) or linking spectra to concentrations of water quality constituents (Uudeberg et al., 2019). These methods have been successfully applied to Chl-*a* estimation in the open ocean and inland and coastal waters at the regional or even global scale (Liu et al., 2020; Neil et al., 2019). In recent years, ML is more and more used for Chl-*a* retrieval, because it not only improves the retrieval accuracy but also reduces the uncertainty due to complex physical processes (Cao et al., 2020; Kwiatkowska and Fargion, 2003; O'Reilly and Werdell, 2019; Pahlevan et al., 2020). Nitrogen (N) and phosphorus (P) commonly co-limit primary productivity in lakes. Therefore, total phosphorous (TP) and total nitrogen (TN) are important factors in the initiation, propagation, and maintenance of harmful algal blooms (Paerl and Otten, 2013). However, TP and TN are judged as non-optically active, and hence research on their retrievals is limited (Sagan et al., 2020; Xiong et al., 2020). In the existing research, TP and TN are mainly retrieved indirectly by their correlations with optically active constituents (OACs) (e.g., Chl-*a* and SDD) (Lu et al., 2020; Song et al., 2012). The application of ML enhances the ability to capture the potential relationship among OACs, non-OACs, and R_{rs}, and is expected to promote non-OAC retrieval by saving the efforts on prior OAC retrieval. Chang et al., (2013) mapped the spatiotemporal distribution of TP in Tampa Bay with R_{rs} derived from Moderate Resolution Imaging Spectroradiometer (MODIS) data and a suite of genetic programming models. Wang et al., (2018) developed a back-propagation neural network with both satellite-derived sea surface salinity and R_{rs} generated from the Geostationary Ocean Color Imager data and estimated TP and TN of the coastal regions of the East China Sea. Guo et al., (2021a) compared the performances of several ML algorithms for TP and TN retrievals and estimated TP and TN of a small urban lake in northern China using the optimal models. In summary, ML is not only progressively applied to Chl-*a* retrieval, but also contributes to TP and TN retrievals. Meanwhile, Chl-*a*, TP, and TN are usually highly cross-correlated, but there are few studies on the simultaneous retrieval of them and attempting to effectively separate their estimates.

As part and an improvement of traditional ML, deep learning (DL) focuses on large-size and deep neural networks (DNN), which has attracted wide attention in recent years (Ma et al., 2019; Reichstein et al., 2019). Based on multi-layer learning (Urban et al., 2016), DL was proved to outperform traditional ML models with substantial improvement in capturing the potential association between environmental variables and remote sensing images (Yuan et al., 2020; Zhu et al., 2017). Peterson et al., (2020) presented the robustness of the progressively decreasing DNN (pDNN) for estimating multiple WQPs with harmonized Landsat 8 and Sentinel-2 data in both fluvial and reservoir waters. Cao et al., (2019) evaluated the bandwidth effects of common high spatial resolution satellite sensors on optical properties of inland waters using DNN and several other methods. Pu et al., (2019) configured a convolutional neural network (CNN) termed AlexNet to fit the relationship between Landsat 8 images and water quality levels, and found that CNN improved both the accuracy and coverage of inland lake monitoring. These studies indicate the strengths of DL for remote sensing retrieval of WQPs (especially non-OACs) in inland waters. Inspired by the fact that humans learn and process complex information through multiple senses (e.g., visual, auditory, and kinesthetic), Ngiam et al., (2011) proposed the multimodal DL (MDL). MDL first uses multiple independent models to learn different kinds of features, then fuses the learning results and inputs them into a new independent model for the final prediction (Ramachandram and Taylor, 2017). MDL has been widely used in human activity (e.g. action, gesture, and emotion) recognition, medical applications (e.g. medical image interpretation and computer-aided diagnosis), and autonomous driving (Ramachandram and Taylor, 2017). Recently, MDL has also attracted great interests in remote sensing applications, including imagery classification (Hong et al., 2021), crop yield prediction (Maimaitijiang et al., 2020), semantic labeling (Audebert et al., 2018), and image segmentation (Suel et al., 2021). However, MDL is rarely applied to remote sensing retrieval of WQPs.

In this study, we intended to develop and validate MDL models for long-term Chl-*a*, TP, and TN estimation with atmospherically corrected Landsat R_{rs} data. More specifically, we first corrected Landsat raw images to R_{rs} data set since 1984 and extracted four visible and NIR bands and 20 derived remote sensing indices (RSIs) as model input features. Then, the MDL models were trained and rigorously validated using the matchups of the pre-screened features and Chl-*a*, TP, and TN measurements of Lake Simcoe, Canada. Strengths and limitations of the MDL models in mapping long-term water quality patterns were also investigated. Finally, we quantified and discussed the contribution of each feature to the estimation of Chl-*a*, TP, and TN, in an attempt to explore

the physical mechanism and interpretability of the MDL models.

2. Materials and methods

2.1. Study site

Lake Simcoe was selected as the study site because it provides long-term and systematic water quality monitoring data. Lake Simcoe is located in southern Ontario, Canada, bordering on Simcoe County, Durham Region, and York Region. Lake Simcoe covers 722 km^2 and is fed by 35 major streams and rivers (Fig. 1). Lake Simcoe supplies drinking water, agricultural water, and diverse recreation opportunities to more than 450,000 people in the watershed and Ontario, including parts of the Oak Ridges Moraine and the Greenbelt. Over the years, climate change and anthropogenic activities like urban development, agriculture, and recreation have significantly affected the water quality of Lake Simcoe (Crossman et al., 2016). To rectify these problems, the Canadian government and related organizations implemented a series of environmental improvement programs, for example, the Lake Simcoe Environmental Management Strategy (LSEMS) program implemented in 1990 and the Lake Simcoe Protection Plan (LSPP) implemented in 2008. The implementation of these programs significantly improved the water quality of Lake Simcoe.

2.2. Satellite data and processing

To reconstruct the long-term water quality patterns in Lake Simcoe with high spatial resolution, Landsat raw images from 1982 to 2020 were selected as the satellite data source in this study. The image screening followed several criteria: 1) cloud cover was constrained to $\leq$ 70%; 2) cloud-covered area visually estimated from Landsat look natural color images was $<$ 30% of the lake area; 3) images in winter (December to February) were excluded because the lake was frozen; 4) images of frozen lake in other months were further excluded by visual examination. After applying these criteria, a total of 225 qualified Level-1 images were downloaded from the United States Geological Survey portal (https://earthexplorer.usgs.gov). Note that except for the observation gap of Landsat 5 and 8 between 5 May 2012 and 11 April 2013, 74 Landsat 7 images after 31 May 2003 (the scan line corrector failed) were only used for training and validating the models, but not for water quality mapping to avoid the impacts of data gaps. No eligible Landsat 4 TM image was screened out. The Level-2 surface reflectance (SR) of TM, ETM+, and OLI is atmospherically corrected by the Landsat Ecosystem Disturbance Adaptive Processing System (LEDAPS) (Schmidt et al., 2013) and the Landsat 8 Collection 1 Land Surface Reflectance Code (LaSRC) (Vermote et al., 2016). Since both methods are developed for land surface applications, the SR data may have uncertainties when applied over water (Ilori et al., 2019). Therefore, the Level-1 top-of-atmosphere (TOA) data was used rather than the Level-2 SR data in this study.

The TOA data was atmospherically corrected to R_{rs} using the dark spectrum fitting (DSF) and exponential extrapolation (EXP) algorithms integrated into ACOLITE (version 20210114.0) (Vanhellemont, 2019). During the atmospheric correction, the Rayleigh corrected reflectance (ρ_{rc}) was also exported. Given the low SNR of TM and ETM+ short-wave infrared (SWIR) bands, the NIR and SWIR bands rather than the default of two SWIR bands were used for the EXP algorithm. We compared the percentages of invalid pixels (negative values) produced by EXP and DSF algorithms at visible and NIR bands (Fig. A.1). The results showed that DSF had a lower invalid pixel percentage than EXP. In addition, we trained and validated the MDL models with the outputs of the three atmospheric correction methods and compared the model performances (Fig. A.2–A.4). The results also indicated that the models performed better with DSF outputs. Therefore, DSF was selected as the atmospheric correction method in this study. The closest SWIR bands to 1,600 nm were used for removing cloud and other non-water pixels with a threshold of 0.0215. To avoid the severe land adjacency effect, a 300 m buffer inward the water boundary was removed for water body (Wang et al., 2020).

RSIs are proved to provide better sensitivity than individual spectral bands (Gao et al., 2020). To enable the models to learn more useful information and achieve more accurate water quality retrieval, we

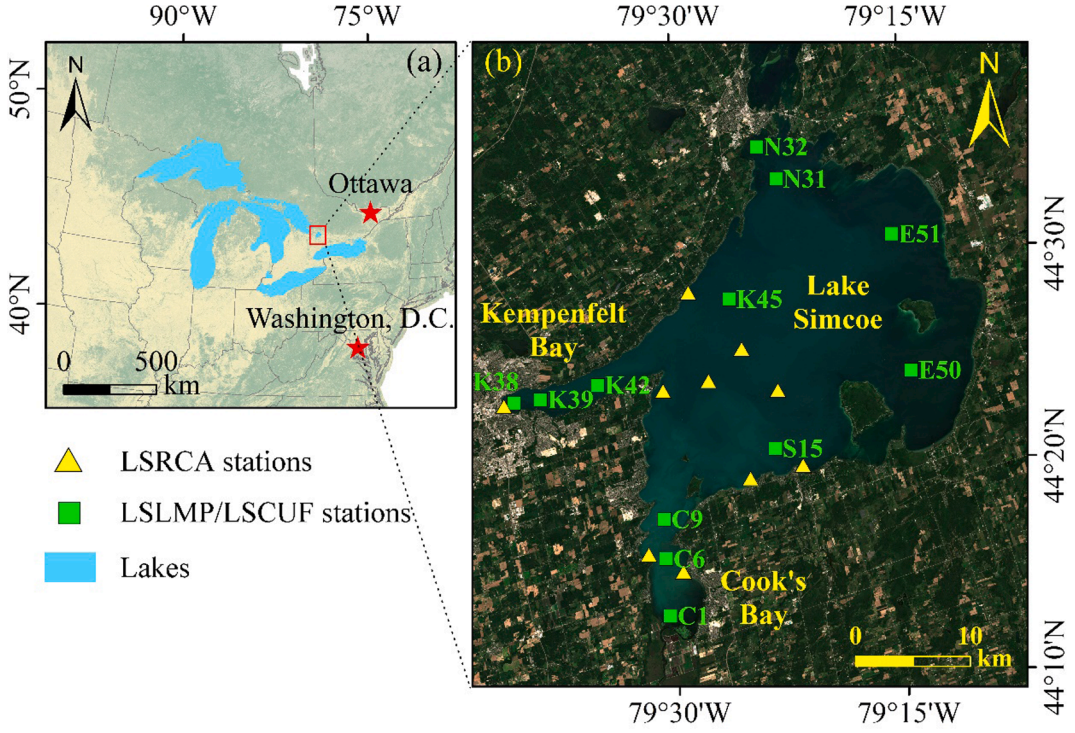

Fig. 1. Locations of Lake Simcoe (a) and the water quality monitoring stations (b). The base map used in (b) was derived from the true color composite image (Band 4, 3, and 2) ingested by Sentinel-2 on 14 August 2020.

calculated 20 RSIs (Table A.1) based on the R_{rs} of visible and NIR bands to supplement the four individual spectral bands. The selection of RSIs referred to the spectral characteristics of Lake Simcoe presented by the matched-up data set used in this study (Fig. 2). In Lake Simcoe, the reflectance peaks appeared at the green band and close reflectance appeared at the blue band. The red and NIR bands showed relatively strong absorption. We speculated that the effects of phytoplankton, CDOM, and non-algal particles on non-water absorption were balanced in Lake Simcoe (Spyrakos et al., 2018).

All features were split into three categories based on the number of bands participating in the feature calculation, namely $F_{1\text{-band}}$, $F_{2\text{-band}}$, and $F_{3\text{-band}}$. Then, we calculated the Pearson correlation coefficients (r) between 24 features and each WQP, respectively. Some features belonged to vegetation indices, they might confuse the models due to their weak correlations with WQPs. Moreover, the more features, the more samples were needed for model training. Therefore, we further screened the features with high contributions to the prediction by setting a threshold for $|r|$. The optimal threshold was determined by the model performances with different thresholds within [0.1, 0.9]. The features with $|r|$ being beyond the optimal threshold were excluded. The remaining features (Table 1) were used to match up with the water quality measurements for model development and validation.

2.3. Water quality measurements

We used three different ground truth data sets of Chl-*a*, TP, and TN to train and validate the MDL models, i.e., in-situ measurements obtained from the Lake Simcoe lake monitoring program (LSLMP) (for Chl-*a*, $N = 315$, for TP and TN, $N = 303$), the Lake Simcoe Clean-up Fund (LSCUF) ($N = 89$), and the Lake Simcoe Region Conservation Authority (LSRCA) ($N = 58$).

LSLMP is conducted by the Ministry of the Environment, Conservation and Parks, Ontario, Canada, and provides biweekly measurements of chemical and physical WQPs in Lake Simcoe. The LSLMP data from 2013 to 2019 was downloaded from https://data.ontario.ca. The LSLMP data and satellite data were matched-up to develop MDL models for estimating Chl-*a*, TP, and TN. To ensure the high quality of the in-situ and concurrent satellite data, Chl-*a*, TP, and TN measurements were selected with the time interval between the field measurement and satellite sensing being constrained to < 6 h (Cao et al., 2020). The coefficient of variation (CV) was used to detect the spatial variability of matched pixels in each spectral band (Zibordi et al., 2009). If the CV of the 3×3 window around the matched pixel was greater than 20%, the pixel was excluded.

The LSCUF and LSRCA measurements were used as independent data sets to further validate the model generalization. LSCUF is part of the Government of Canada's Comprehensive Approach to Clean Water and is managed by the Environment Canada. LSCUF shares the same sampling locations with LSLMP and provides Chl-*a*, TP, and TN measurements from 2007 to 2012 (https://open.canada.ca). The LSRCA is continually involved in various kinds of monitoring activities including water quality monitoring at 10 lake water quality stations in Lake Simcoe. The Chl-*a*, TP, and TN measurements of LSRCA from 2010 to 2019 were obtained from https://data.lsrca.on.ca. The distributions of the three ground truth data sets and corresponding R_{rs} data used in this study were shown in Fig. A.5 and Fig. A.6. Note that the negative R_{rs} was masked as it might confuse the models.

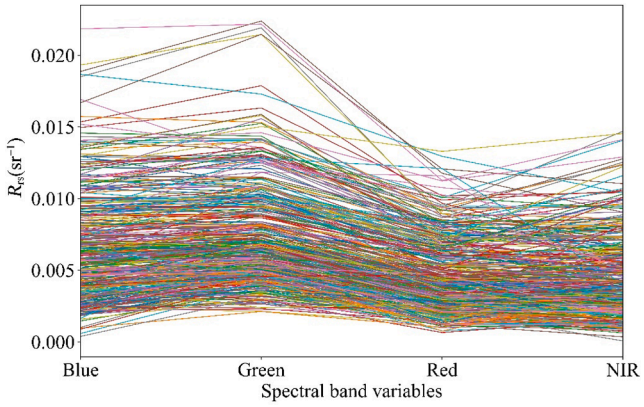

Fig. 2. The spectral characteristics of Lake Simcoe presented by the matched-up data set used in this study. The spectra were obtained from the DSF algorithm integrated into ACOLITE (version 20210114.0).

Using the in-situ measurements of LSLMP, LSCUF, and LSRCA, we analyzed the correlations of OACs Chl-*a* and SDD with non-OACs TP and TN (Fig. 3). The results indicated that Chl-*a* and SDD showed good positive and negative correlations with TP and TN, which provided a basis for indirect estimation of TP and TN by remote sensing images in Lake Simcoe.

2.4. Model development and evaluation

In this study, we attempted to train a DL model that could adequately fit the relationship between the features generated from satellite data and the target WQP, so as to implement water quality retrieval. Here, we employed a simple neural network (Fig. 4a) to illustrate how this model worked. For each sample, the outputs y (i.e., $[y_1 \quad y_2 \quad y_3]$) of the input layer could be computed by Eq. (1):

$$y = \sigma\left(w_1^{\mathrm{T}} x + b_1\right) \tag{1}$$

where x is the input feature vectors $[x_1 \quad x_2]$, representing the R_{rs} of the visible and NIR bands and screened RSIs. w_1 and b_1 are the weight vector $\begin{bmatrix} w_{11}^{(1)} & w_{21}^{(1)} \\ w_{12}^{(1)} & w_{22}^{(1)} \\ w_{13}^{(1)} & w_{23}^{(1)} \end{bmatrix}$ and bias vector $\left[b^{(1)} \quad b^{(1)} \quad b^{(1)}\right]$ applied to the interconnections between the neurons in the input layer and hidden layer. σ is the activation function, which converts the linear transformation between the two layers into a nonlinear transformation and thus allows nonlinear fitting.

The output $\hat{y}$ of the hidden layer, representing the estimated WQP, could be computed by Eq. (2):

$$\hat{y} = w_2^{\mathrm{T}} y + b_2 \tag{2}$$

where w_2 and b_2 are the weight vector $\left[w_{11}^{(2)} \quad w_{21}^{(2)} \quad w_{31}^{(2)}\right]$ and bias vector $\left[b^{(2)}\right]$ applied to the interconnections between the neurons in the hidden layer and output layer. The activation function was not applied in the final WQP prediction.

The model can be determined by solving for w and b. To better fit the relationship between the input features and WQPs, we first designed and constructed more complex model structures, and then optimized w and b by defining the loss function and using the gradient descent. The detailed steps of model construction and hyperparameter optimization were as follows:

For each WQP, we constructed an MDL model integrated by four sub neural networks (sNN) (Fig. 4b). In nature, $F_{1\text{-band}}$, $F_{2\text{-band}}$, and $F_{3\text{-band}}$ belonged to spectral information, but the number of sources providing information in each feature category was different, resulting in different feature sensitivity and noises. If all three categories of features were input into the fully connected neural network at the same time, the interaction of noises might either increase the difficulty of model training (underfitting) or make the model learn excessive noises (overfitting). Therefore, we allocated an sNN for the $F_{1\text{-band}}$ (sNN1), $F_{2\text{-band}}$ (sNN2), and $F_{3\text{-band}}$ (sNN3), respectively, considering the three sNNs could effectively balance the useful information against noises during

Table 1
List of the indices, formulas, and references of the input features.

Indices		Features	Formulas	References
$F_{1\text{-band}}$	F1–4	Individual spectral bands	B, G, R, NIR	—
$F_{2\text{-band}}$	F6	Ratio of red to blue	R/B	(Hewson et al., 2001)
	F8	Ratio of red to green	R/G	(Gamon and Surfus, 1999)
	F9	Ratio of NIR to green	NIR/G	(Gitelson et al., 2002)
	F12	Normalized difference green and red index	(G – R)/(G + R)	(Ahamed et al., 2011)
	F14	Blue NDVI	(NIR – B)/(NIR + B)	(Hancock and Dougherty, 2007)
	F15	Normalized difference green and NIR index	(G – NIR)/(G + NIR)	(Le Maire et al., 2004)
$F_{3\text{-band}}$	F16	Chlorophyll vegetation index	NIR × R/G	(Hunt et al., 2011)
	F17	Green-blue NDVI	(NIR – G + B)/(NIR + G + B)	(Wang et al., 2007)
	F19	Red-blue NDVI	(NIR – B + R)/(NIR + B + R)	(Wang et al., 2007)
	F21	Shape Index	(2 × R – G – B)/(G – B)	(Escadafal. 1994)
	F24	Visible atmospherically resistant index	(G – R)/(G + R + B)	(Gitelson et al., 2003)

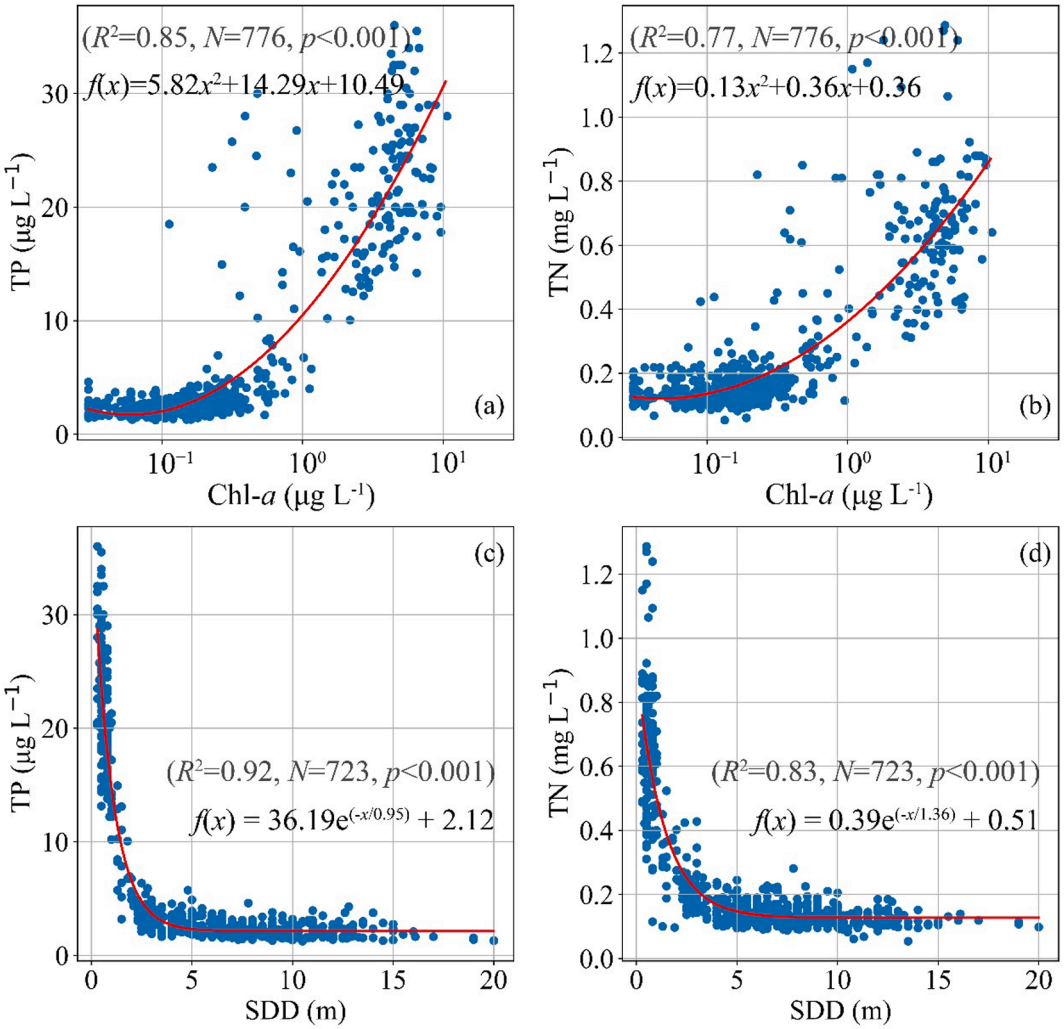

Fig. 3. Correlations of OACs Chl-*a* and SDD with non-OACs TP and TN in Lake Simcoe. The determination coefficient (R^2) and *p*-value (*p*) were used to evaluate the goodness of fit and the statistical significance.

learning. The outputs of the sNNs were essentially nonlinear mapping of the input features after weighted summation by hidden layers. Then these outputs were concatenated and input into the fourth sNN (sNN4) for the final WQP prediction. sNN1 and sNN3 were set with two hidden layers, while sNN2 and sNN4 were set with three hidden layers. The numbers of neurons in the input layers of sNN1, sNN2, and sNN3 were determined by the feature number of each feature category. The numbers of neurons in the output layer of the sNNs needed to be specified according to model performance. In this study, they were set to the same as the input layers for convenience. Since the input of sNN4 was the outputs of the other sNNs and the output of sNN4 was the target WQP, the numbers of neurons in the input layer and output layer of sNN4 were set to ten and one, respectively. The numbers of neurons in the hidden layer of the four sNNs were different and were specified according to the model performance during training. The numbers of neurons were set as few as possible to avoid overfitting.

The Exponential Linear Unit (ELU) was used as the activation function. In model optimization, the mean squared error (MSE) was used as

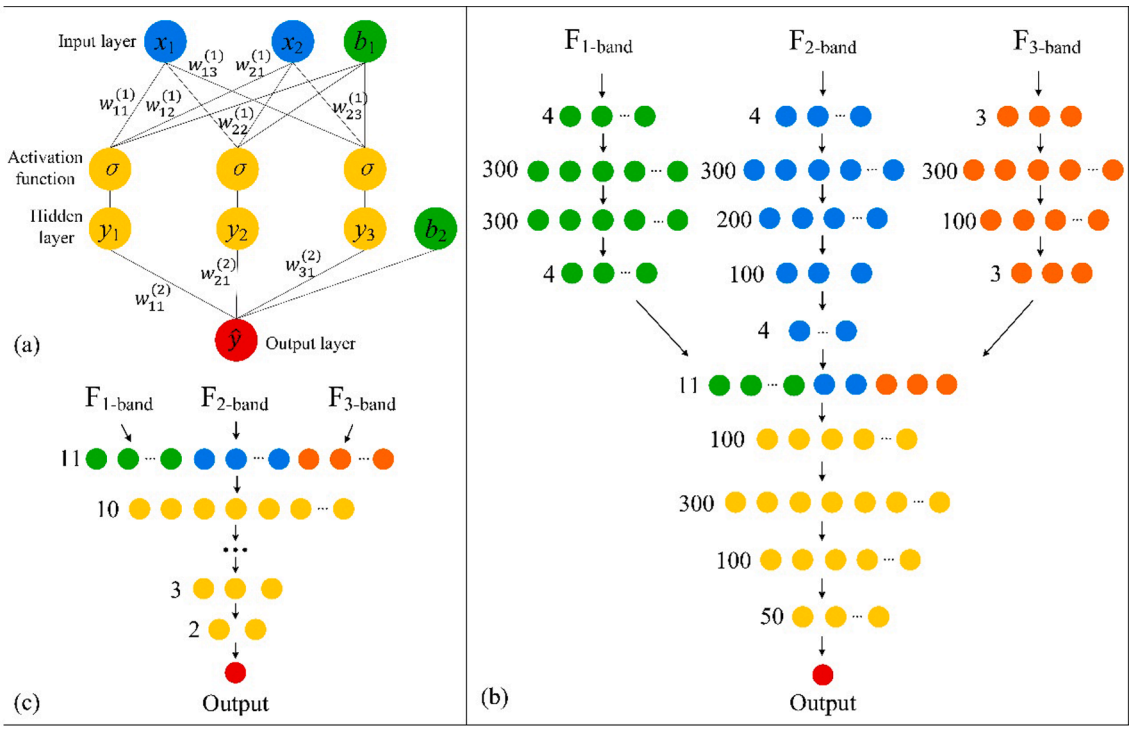

Fig. 4. Structures of a simple neural network (a) illustrating how the input features (i.e., R_{rs} of the visible and NIR bands and screened RSIs) relate to the WQPs, the MDL model (b), and the pDNN model (c). In (b–c), green, blue, orange, and yellow denote the sNN1, sNN2, sNN3, and sNN4 of the MDL model, respectively. Note that we labeled the neuron numbers of the Chl-*a* model to display the model structure. The neuron numbers of the Chl-*a*, TP, and TN models were slightly different. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

the loss function:

$$L(w,b) = \frac{1}{N}\sum_{i=1}^{N}(\hat{y}_i - y_i)^2 \tag{3}$$

where y_i and $\hat{y}_i$ are the measured and estimated WQP, N is the number of samples.

The stochastic gradient descent (SGD) was used for hyperparameter update in model training because SGD only uses one random sample to calculate the gradient in each update instead of all samples. This mode not only speeds up the computation, but the instability of the gradient also helps the model escape the saddle point to find a better solution (Bottou, 2012). The w and b updates using SGD can be formulated as follows:

$$w = w - \eta\frac{\partial L(w,b)}{\partial w} \tag{4}$$

$$b = b - \eta\frac{\partial L(w,b)}{\partial b} \tag{5}$$

where η is the learning rate, determining the step size at each iteration while moving towards the minimum of the $L(w,b)$.

In addition to MDL, we also constructed four other candidate DL and ML models, including the pDNN (Section 1, Fig. 4c), a DNN with a similar number of trainable parameters but no sNN (sDNN), the eXtreme Gradient Boosting (XGBoost) (Chen and Guestrin, 2016), and the support vector regression (SVR) (Cortes and Vapnik, 1995) for comparison purpose. XGBoost and SVR, representing two different kinds of ML algorithms, have been frequently used in water quality retrieval over years (Cao et al., 2020; Guo et al., 2021b). In this study, MDL, pDNN, and sDNN models were implemented by PyTorch (version 1.7.0) (Paszke et al., 2019) in Python. XGBoost and SVR models were implemented using the xgboost (version 1.1.1) and scikit-learn (version 0.20.4) libraries in Python, respectively. To enable the models to learn enough information and avoid overfitting, the "train_test_split" function in scikit-learn was used to randomly split the LSLMP data set into 70% training set and 30% test set before model training. In the training of MDL, pDNN, and sDNN models, dropout was used to reduce overfitting (Srivastava et al., 2014). In addition, the losses on the training set and the test set were strictly monitored to ensure a balanced model performance. In the training of XGBoost and SVR models, the grid-search technique was used to optimize the hyperparameters. To further avoid overfitting, a 5-fold cross-validation was used to ensure that the training set was randomly distributed in different segments (Fan et al., 2017). The hyperparameters and the main codes (taking Chl-*a* as an example) of the DL and ML models developed in this study were listed in Table A.2 and Table A.3.

We also included the NASA Ocean Color two-band (OC2) and three-band (OC3) ratio algorithms (O'Reilly and Werdell, 2019) and another empirical algorithm of Landsat 5 TM in clear lakes (Giardino et al., 2001) to verify the performance of the MDL model for Chl-*a* estimation. Both OC2 and OC3 were based on the ratio of blue to green R_{rs}:

$$\text{Chl-}a = 10^{a+bX+cX^2+dX^3+eX^4} \tag{6}$$

In OC2,

$$X = \log_{10}\left(\frac{R_{rs}(490)}{R_{rs}(560)}\right) \tag{7}$$

In OC3,

$$X = \log_{10}\left(\frac{R_{rs}(443).>R_{rs}(489)}{R_{rs}(560)}\right) \tag{8}$$

where the greater of $R_{rs}(443)$ and $R_{rs}(489)$ is used. Note that the coastal aerosol band (433–453 nm) of OLI was only used for OC3 and not used for other model training because of its absence on TM and ETM+ .

The algorithm proposed by Giardino et al., (2001) (hereafter termed as MLR for brevity) estimates Chl-*a* by relating R_{rs} at 490 nm and 560 nm to in-situ measured Chl-*a* with multiple linear regression:

$$\text{Chl-}a = a \times R_{rs}(490) - b \times R_{rs}(560) - c \quad (9)$$

The OC2, OC3, and MLR models were trained and validated using the same data sets as other ML and DL models. To avoid possible effects of the data set splitting, model training was repeated 100 times for each algorithm, re-splitting the data set in every run (Tavares et al., 2021). The bands used in the original algorithms were converted to the closest comparable TM, ETM+, and OLI bands.

Two widely used log-transformed metrics (Seegers et al., 2018; Smith et al., 2021) were calculated for evaluating the model performances on the LSLMP test set and two independent data sets (i.e., LSCUF and LSRCA), namely the mean absolute error (MAE) and Bias. They can be formulated as follows:

$$\text{MAE} = 100 \times \left\{ 10^{\left[\frac{\sum_{i=1}^{N} |\log_{10}(\hat{y}_i/y_i)|}{N}\right]} - 1 \right\} \quad (10)$$

$$\text{Bias} = 100 \times \left\{ 10^{\left[\frac{\sum_{i=1}^{N} \log_{10}(\hat{y}_i/y_i)}{N}\right]} - 1 \right\} \quad (11)$$

where N is the number of the matched-up pairs, $\hat{y}_i$ and y_i are the estimated and measured WQP, respectively. For both metrics, values closer to zero indicate better model performance.

3. Results

3.1. Preliminary feature screening

Before model training, we calculated r between the WQPs and the features, and preliminarily screened the features using a threshold determined by model performances, to avoid excessive redundant noises being input into the models and improve the model efficiency (Fig. 5, Table A.4).

$F_{1\text{-band}}$ (F1–4) showed positive correlations with the three WQPs, with $r \geq 0.4$. Chl-*a* had the strongest correlation with F2 (green band), with r reaching 0.61. The correlation between Chl-*a* and F1 (blue band) was close to that between Chl-*a* and F2 ($r = 0.52$). The correlations between Chl-*a* and F4 (NIR band) and F3 (red band) were close, with r of 0.49 and 0.40, respectively. Like Chl-*a*, TP also had the strongest correlation with F2, with r reaching 0.55. The following was the correlation between TP and F3 ($r = 0.50$). The correlations between TP and F1 and F4 were the same, with both r reaching 0.47. TN had the strongest correlation with F3, with r of 0.57. The correlations between TN and F4 and F2 were close, with r reaching 0.55 and 0.52, respectively. The correlation between TN and F1 was slightly weaker than those between TN and F2, F3, and F4 ($r = 0.46$). These results indicated the potential of using the spectral bands of atmospherically corrected Landsat R_{rs} to estimate Chl-*a*, TP, and TN.

In addition to the individual spectral bands, we also calculated 20 common RSIs (Table A.1) so that the models could learn more effective information. $F_{2\text{-band}}$ (F5–F15) showed relatively weak correlations with Chl-*a*, TP, and TN, with r fluctuating between –0.31–0.35, –0.46–0.47, and –0.43–0.42, respectively. Chl-*a* had the strongest correlation with F14 ($r = 0.35$). The correlation between Chl-*a* and F8 was the weakest, with r of –0.17. TP and TN showed strong positive correlations with F6, F14, and F9, but strong negative correlations with F8, F12, and F15. For $F_{3\text{-band}}$ (F16–24), three WQPs showed negative correlations with F22 and F24, and positive correlations with the other features. Chl-*a*, TP, and TN had the strongest correlations with F16, with r of 0.51, 0.56, and 0.56, respectively, while had the weakest correlations with F23, F18, and F20, with r of 0.05, 0.17, and 0.13, respectively. Except for F16, r of Chl-*a* and other features fluctuated in a small range of –0.29–0.35. Except for F16, F17, F21, and F24, r of TP and TN and other features fluctuated in a small range of –0.24–0.28.

Fig. 6 showed that the MDL models performed best with the threshold of 0.3. The MAE of the Chl-*a*, TP, and TN models was 33.14%,

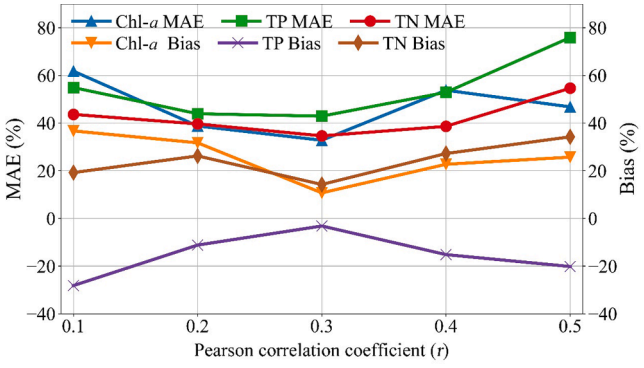

Fig. 6. The performances of the MDL models with features screened by different Pearson correlation coefficient (r) thresholds.

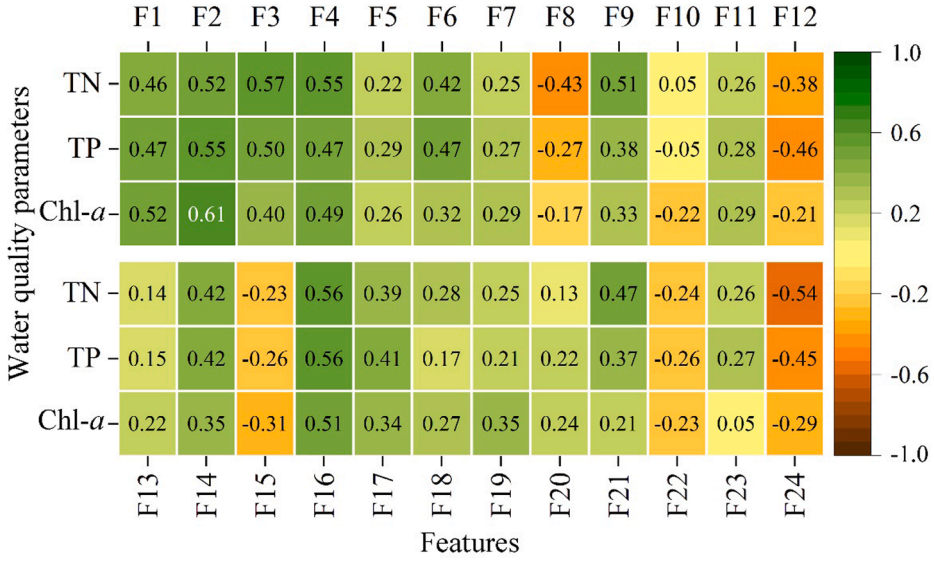

Fig. 5. The r between the WQPs and the features. F1–4, F5–15, and F16–24 are the $F_{1\text{-band}}$, $F_{2\text{-band}}$, and $F_{3\text{-band}}$, respectively.

43.26%, and 35.40%, respectively. The Bias of the Chl-*a*, TP, and TN models was 10.85%, −3.20%, and 14.37%, respectively. Therefore, 0.3 was set as the threshold for preliminary screening of the features.

In summary, although the correlations of $F_{2\text{-band}}$ and $F_{3\text{-band}}$ with WQPs were not as strong as those of $F_{1\text{-band}}$ with WQPs, we anticipated that $F_{2\text{-band}}$ and $F_{3\text{-band}}$ could supplement the individual spectral bands. We also observed differences in the correlations of three WQPs with some features and in the screened features with the same threshold. These differences might make it possible to effectively separate the estimates of three WQPs. The features with $|r| < 0.3$ were excluded before model development.

3.2. Model performances

To better explore the model training and test processes, especially the overfitting, here, we separately presented the model performances on the training set and the test set (Fig. 7). In general, the models of three WQPs performed well on both training set and test set. We noted that the performances of the Chl-*a*, TP, and TN models on the test set decreased in different degrees compared to those on the training set, but the overall performances maintained a good balance. The results suggested that slight overfitting existed in the models, but it was controlled at a good level. In this study, the model evaluation scores on the test set were chosen to represent the model performances in application. Chl-*a* belongs to OAC, thus the Chl-*a* model was expected to perform better than the models of non-OACs (i.e., TP and TN). However, the results showed that the model performances of TP (MAE = 42.58%, Bias = −2.82%) and TN (MAE = 35.05%, Bias = 13.66%) were quite close to that of Chl-*a* (MAE = 32.57%, Bias = 10.61%). Although there existed slight overfitting, the models developed in this study adequately estimated Chl-*a*, TP, and TN.

In addition to the LSLMP test set, we further used the LSCUF and LSRCA measurements as independent data sets to validate the model generalization (Fig. 8). The model performances on the independent data sets were maintained with high accuracy, although different degrees of accuracy loss was observed on the models of all WQPs. The validation MAE and Bias of the Chl-*a* model were 27.87% and 23.73%, respectively. We speculated that the narrower range of Chl-*a* concentration in the LSCUF and LSRCA data sets led to the MAE improvement. Although TP model showed a clear accuracy loss and the underestimation on the LSLMP test set even changed to an overestimation on the LSCUF and LSRCA data sets, the overall accuracy remained at 52.10% and 22.24%, respectively. The validation MAE and Bias of the TN model remained at 47.83% and 18.61%, with decreases of 36.46% and 36.24%, respectively. Through the validation of the LSLMP test set and two independent data sets, the MDL models developed in this study were demonstrated to be capable to adequately estimate long-term Chl-*a*, TP, and TN of Lake Simcoe with the Landsat R_{rs} data.

Moreover, we trained and validated a set of other candidate models using the LSLMP data set to further investigate the strengths of the MDL models. Fig. 9 showed the performances of the pDNN, sDNN, XGBoost, SVR, OC2, OC3, and MLR models on the LSLMP test set. As expected, the evaluation results indicated that the MDL models outperformed all other models for all three WQPs. For Chl-*a*, the MAE and Bias were improved by 23.88% and 53.60% by the MDL model compared to those of the optimal candidate model (pDNN). The performance of the pDNN (MAE = 42.79%, Bias = 22.87%) for Chl-*a* estimation was proved to be better than the sDNN (MAE = 53.76%, Bias = 35.12%) and other two traditional ML models, which was consistent with the research results in (Peterson et al., 2020). Although OC2, OC3, and MLR were developed for Chl-*a* estimation in clear Case-I waters and considered to be applicable for Lake Simcoe (with Chl-*a* < 10 μg L^{-1}), the results showed that

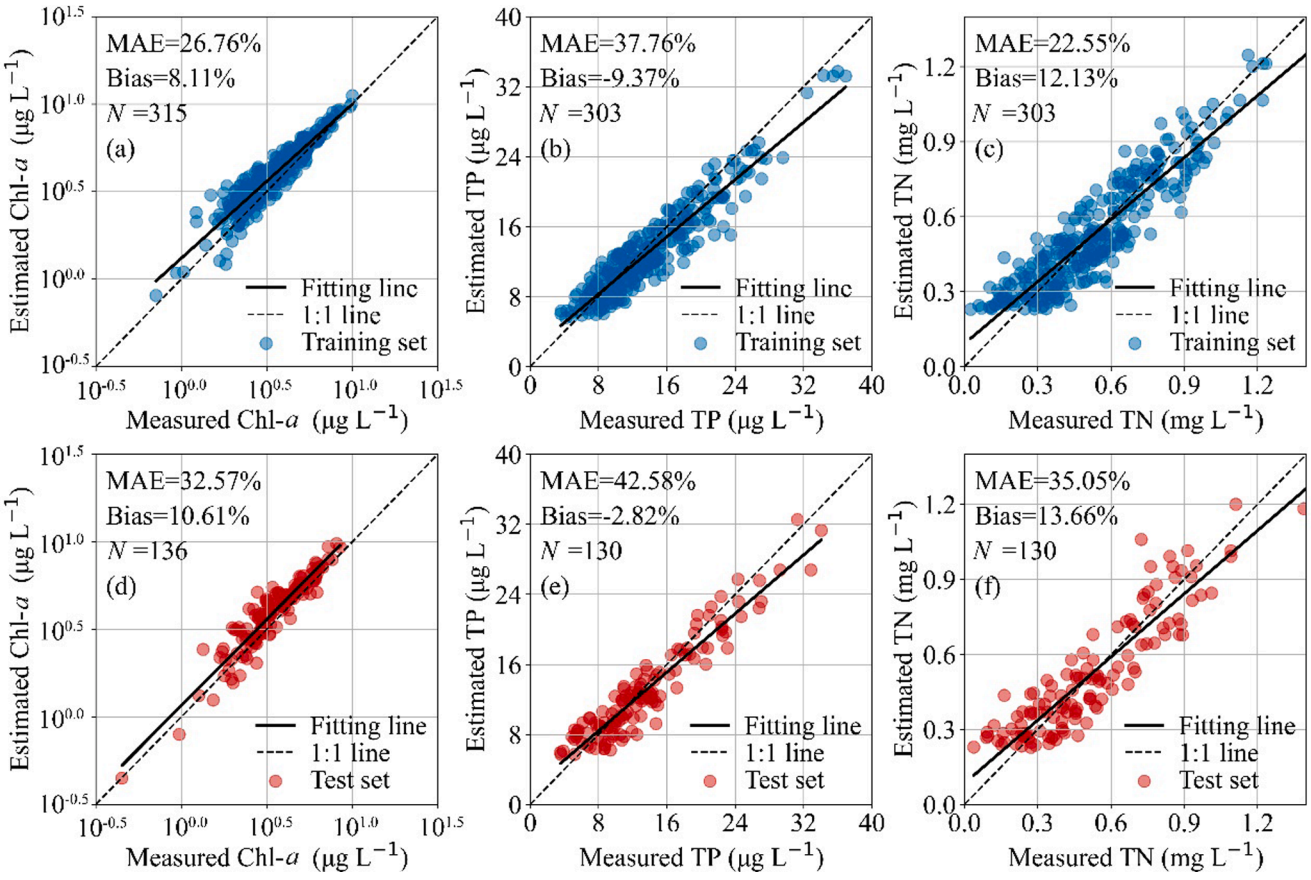

Fig. 7. Performances of the MDL models on both training set (a–c) and test set (d–f).

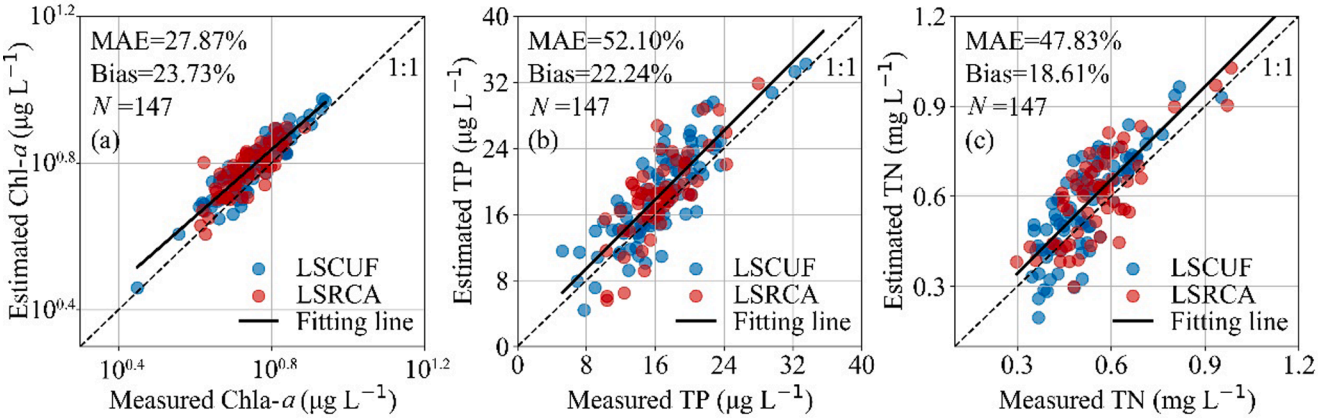

Fig. 8. The performances of the Chl-*a* (a), TP (b), and TN (c) models on the two independent data sets, namely, the LSCUF (blue) and LSRCA (red). Note that the MAE and Bias were the averages of the two data sets. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

the MDL model performed better than all of them. More specifically, OC3 performed slightly worse than pDNN (MAE = 42.79%, Bias = 22.87%), with MAE and Bias of 48.86% and 31.89%, respectively. The performance of OC2 was inferior to sDNN (MAE = 53.76%, Bias = 35.12%) and XGBoost (MAE = 60.59%, Bias = 39.36%), with MAE and Bias of 89.61% and 47.82%, respectively. The following was the MLR algorithm, of which a high scatter was also observed, with MAE and Bias of 96.98% and 54.44%, respectively. SVR performed the worst among all candidate algorithms for Chl-*a* estimation, with MAE and Bias of 117.89% and 58.99%, respectively. The MAE and Bias of the MDL model for TP estimation were 51.80% and 46.79% better than the second-best model (sDNN). The performance of TN estimation was also improved by the MDL model, with MAE and Bias being increased at least 55.40% and 49.87%, respectively. These results suggested that the MDL models had strengths in both OAC and non-OAC retrieval. We noted that XGBoost also showed good potential, especially for the estimation of TP (MAE = 115.58%, Bias = 19.81%) and TN (MAE = 78.58%, Bias = 27.25%). This might benefit from that XGBoost not only adds a regularization to the objective function to control the overfitting but also uses both the first and second derivatives for more accurate and efficient loss calculation (Cao et al., 2020).

3.3. Spatiotemporal variation of water quality

The Canadian government began to implement LSEMS in 1990 and LSPP in 2009, aiming to improve the water quality and protect the ecological health of Lake Simcoe. We reconstructed the spatiotemporal distributions of water quality in Lake Simcoe since 1984 with the Landsat R_{rs} data and the MDL models developed in this study. Here, to analyze the water quality improvement effects of LSEMS and LSPP in Lake Simcoe, we mapped the average spatial distributions of Chl-*a*, TP, and TN of Lake Simcoe during 1984–1989, 1990–2008, and 2009–2020, respectively (Fig. 10). The annual water quality in Lake Simcoe and its different areas from 1984 to 2020 was also calculated. In addition, the spatial averages of three WQPs from 1984 to 2020 were mapped to analyze the long-term spatial distribution characteristics of water quality in Lake Simcoe. Moreover, the spatial distributions of water quality on two typical dates were compared with true-color composite (TCC) images to investigate the spatial distribution characteristics of three WQPs.

3.3.1. Spatial variation

We compared the water quality variation of Lake Simcoe and its five most concerned populated and estuarine intensive areas (i.e. the Cook's Bay (CB), the Kempenfelt Bay (KB), offshore areas in the east, north, and west of the lake (hereafter termed as east, north, and west for brevity)). The water quality of CB was quantified by the average estimates of C1, C6, and C9. The water quality of KB was quantified by the average estimates of K38, K39, and K42. The water quality of the north was quantified by the average estimates of N31 and N32. The water quality of the east and west was represented by the estimates of E50 and K45, respectively. The water quality before and after the LSEMS implementation and after the LSPP implementation were quantified by the average estimates of 1984–1989, 1990–2008, and 2009–2020, respectively.

After the LSEMS and LSPP implementation, Chl-*a*, TP, and TN in Lake Simcoe showed decreasing trends, with percentages of 4.27%, 32.85%, and 29.52%, respectively. More specifically, after the LSEMS implementation, Chl-*a*, TP, and TN decreased by 8.36%, 28.79%, and 25.31%, respectively. After the LSPP implementation, Chl-*a*, TP, and TN further decreased by 2.28%, 5.58%, and 4.65%, respectively.

Chl-*a* decreased significantly after the LSEMS implementation. Especially in the CB, east, north, and KB, Chl-*a* decreased by 18.92%, 7.54%, 4.13%, and 2.52%, respectively. Chl-*a* showed no significant change after the LSPP implementation, with an average concentration of 2.26 ± 0.59 μg L^{-1}. In some areas of the east, Chl-*a* even rebounded, while in the center of the lake, Chl-*a* decreased to a certain extent. TP decreased significantly after LSEMS and LSPP implementation. Especially in the KB, west, CB, and north, TP decreased by 21.67%, 15.59%, 12.54%, and 9.72%, respectively. A significant TP decrease in the area extending from CB to the center of the lake was also observed. TP decreased slightly in the east, with a percentage of 0.78%. After LSPP implementation, TP in the east, KB, CB, and west further decreased by 17.79%, 14.17%, 11.33%, and 9.59%, respectively. TP also decreased to some extent in the center of the lake. TP in the north showed a slight rebound, with a percentage of 1.23%. After the LSEMS implementation, TN in the west, KB, and north showed significant decreases, with percentages of 19.54%, 15.76%, and 12.17%, respectively. Meanwhile, TN in the area extending from CB to the center of the lake also showed a significant decrease. TN did not decrease significantly in the east and even increased to a certain extent in some areas. A certain TN increase was also observed in the center of the lake. After the LSPP implementation, TN showed a significant decrease in the east, with a percentage of 16.43%. In the west, KB, and CB, TN further decreased by 11.33%, 8.09%, and 7.28%, respectively. TN in the north showed a slight rebound, with a percentage of 0.73%. In addition, a certain increase of TN was observed in the center of the lake.

According to the average spatial distributions of water quality from 1984 to 2020, Chl-*a*, TP, and TN in the east, CB, west, north, and KB were significantly higher than those in other areas. Three populated cities (i.e., Keswick, Orillia, and Barrie) are distributed near CB, north, and KB. Although the east and west are more sparsely populated, the

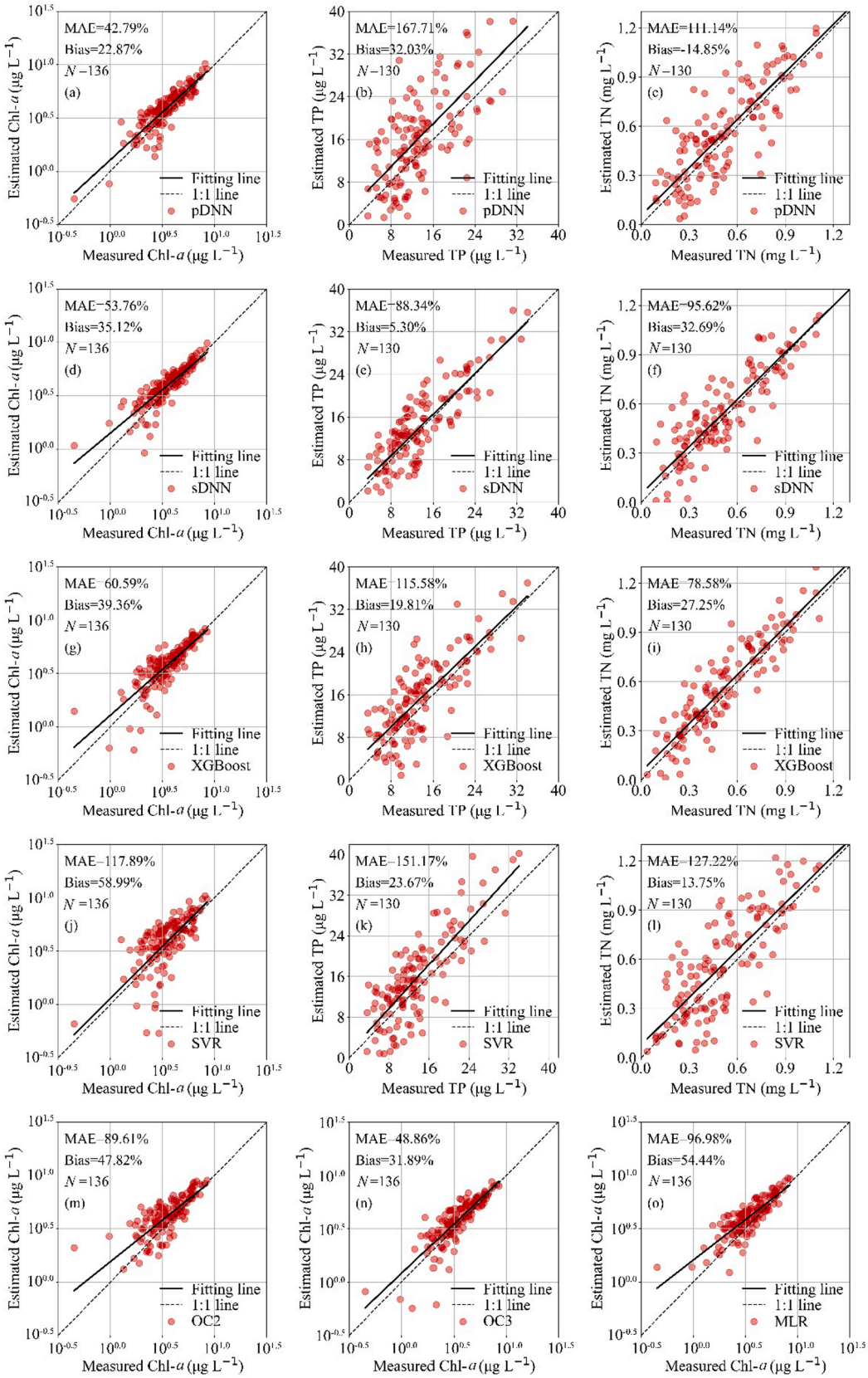

Fig. 9. The performances of the pDNN (a–c), sDNN (d–f), XGBoost (g–i), SVR (j–l), OC2 (m), OC3 (n), and MLR (o) models on the LSLMP test set.

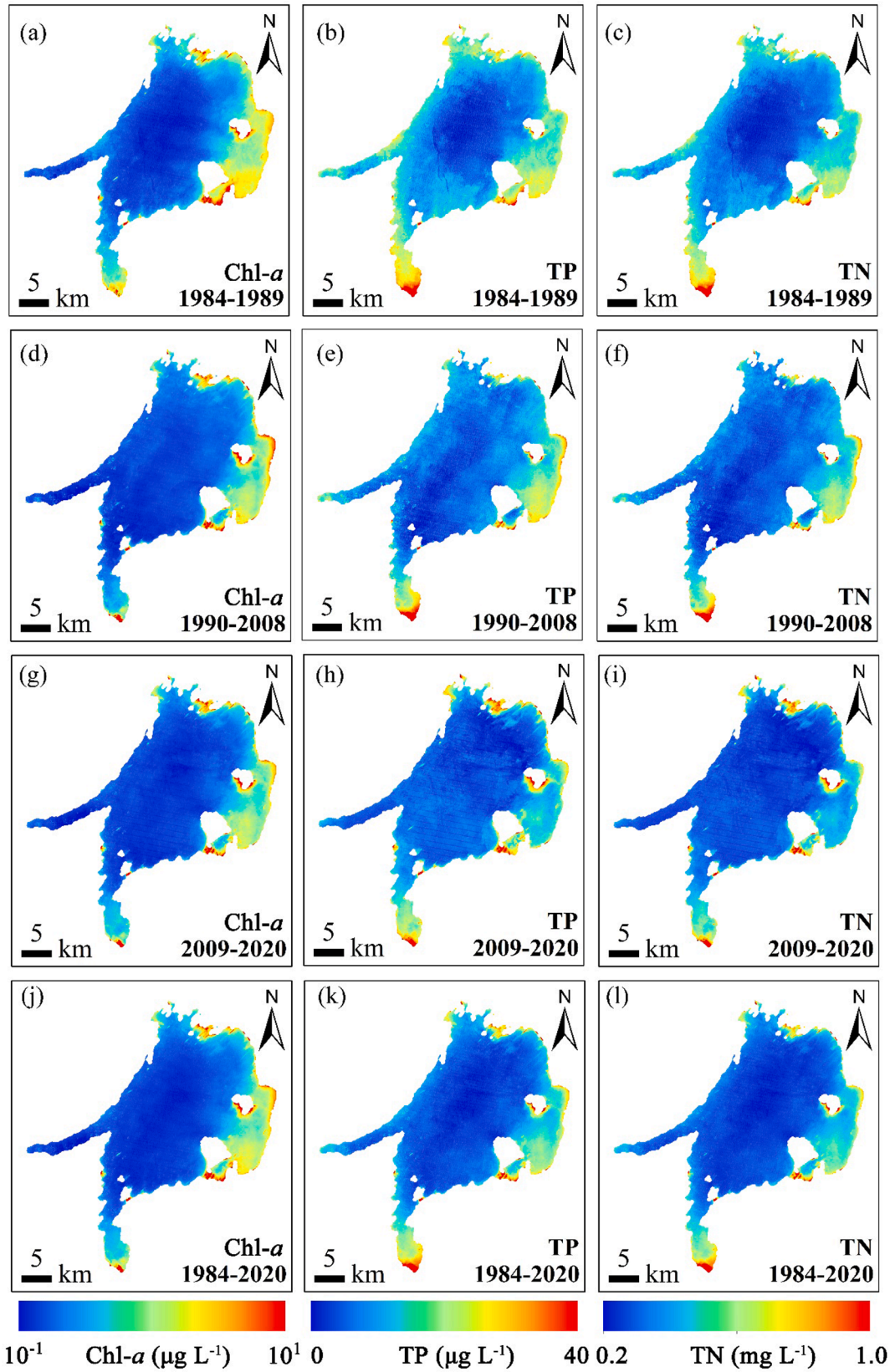

Fig. 10. Mapping of the average spatial distributions of Chl-*a*, TP, and TN in Lake Simcoe during 1984–1989 (a–c), 1990–2008 (d–f), 2009–2020 (g–i), and 1984–2020 (j–l).

areas are widely covered with farmland. In addition, extensive estuaries are distributed in these areas. Therefore, we inferred that these areas received a mass of nutrients from urbanization and intensive agricultural activities, which led to algae growing vigorously and even algal blooms.

Fig. 11 showed the TCC images and the corresponding Chl-*a*, TP, and TN estimates on 19 September 1984 and 22 October 2014. Compared with the TCC images, we considered that the models adequately estimated Chl-*a*. Meanwhile, we observed that the spatial distributions of the Chl-*a*, TP, and TN showed high consistency. P and N are the main limiting nutrients for algal growth, and P and N are highly homologous (Paerl and Otten, 2013), we hence speculated that TP and TN indirectly affected the spectral characteristics of water mainly by affecting the concentration of Chl-*a* in Lake Simcoe. Adding RSIs in addition to the individual spectral bands to expand the model input features might make the models better learn the correlations between TP, TN, and Chl-*a* and the spectral features, and in turn contribute to the retrieval of the non-OACs TP and TN. However, the cross-correlation between Chl-*a*, TP, and TN challenged the accurate estimation of any WQP without the impacts of other WQPs. Therefore, we further calculated the contribution of each feature to the model outputs to explain the feasibility for effective separation of Chl-*a*, TP, and TN estimates by the MDL models. This content was discussed in detail in Section 4.3.

3.3.2. Temporal variation

Fig. 12 illustrated the annual model estimated water quality variation of Lake Simcoe and its five most concerned areas from 1984 to 2020. The water quality of the whole lake was the spatial average.

From 1984 to 2020, Chl-*a*, TP, and TN of Lake Simcoe showed fluctuating downward trends, with percentages of 64.79%, 60.63%, and

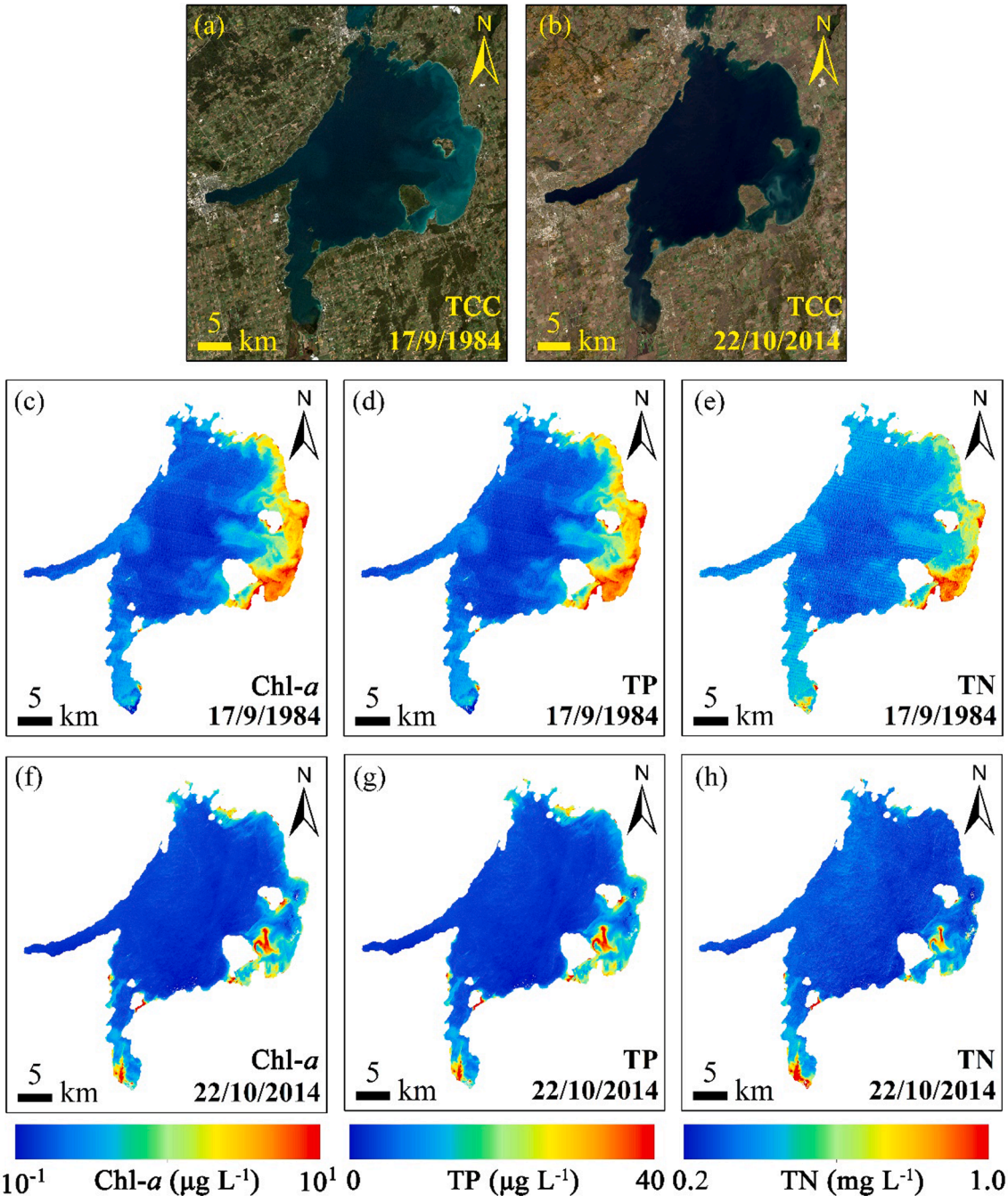

Fig. 11. TCC images (a–b) and the corresponding Chl-*a*, TP, and TN estimates on 17 September 1984 (c–e) and 22 October 2014 (f–h).

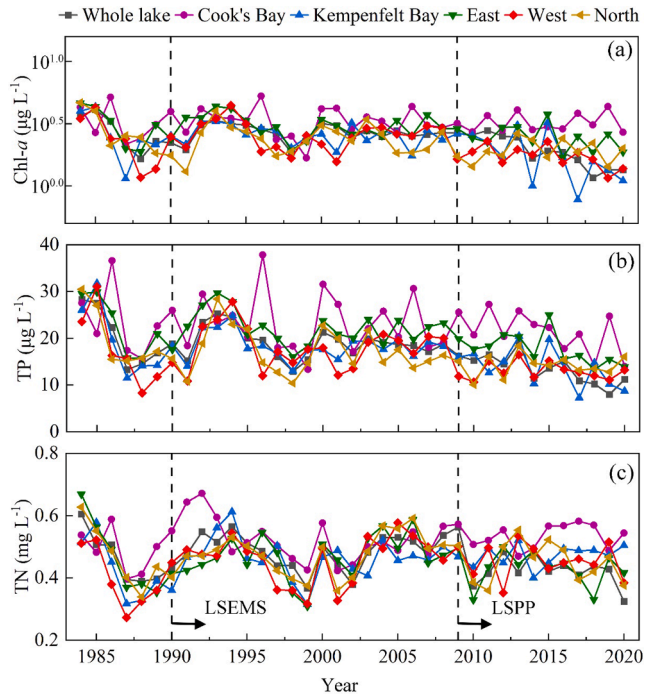

Fig. 12. The estimated Chl-*a* (a), TP (b), and TN (c) variation of Lake Simcoe and its five most concerned areas from 1984 to 2020.

36.92%, respectively. Chl-*a*, TP, and TN decreased significantly from 1984 to 1988, with percentages of 57.32%, 49.75%, and 36.47%, respectively. From 1988, Chl-*a*, TP, and TN gradually increased and reached local peaks of 3.36 μg L^{-1}, 18.88 μg L^{-1}, and 0.57 mg L^{-1} three to four years after the LSEMS implementation. After that, the implementation effect of the LSEMS appeared. Three WQPs decreased significantly and reached local troughs of 1.87 μg L^{-1}, 12.93 μg L^{-1}, and 0.37 mg L^{-1} in 1998 (Chl-*a*, TP) and 1999 (TN), respectively. Chl-*a* and TP increased continuously in 1998 and 1999 and gradually decreased from 2000 to 2020, with percentages of 58.74% and 48.59%, respectively. TN showed a significant increasing trend from 1999 to 2008, with a percentage of 43.75%. After the LSPP implementation, Chl-*a*, TP, and TN showed significant decreasing trends again, with percentages of 52.65%, 35.43%, and 28.56%, respectively.

The water quality variation in the five most concerned areas from 1984 to 2020 tended to be consistent. Chl-*a*, TP, and TN of the CB were higher than those of other areas in most years. In 1992 after the LSEMS implementation, Chl-*a*, TP, and TN decreased significantly and reached the local troughs of 1.68 μg L^{-1}, 13.35 μg L^{-1}, and 0.43 mg L^{-1} in 1999. From 2000 to 2020, Chl-*a* and TN fluctuated in ranges of 2.70–4.36 μg L^{-1} and 0.42–0.58 mg L^{-1}, respectively. TN even showed a slight increasing trend, and the LSPP implementation also failed to change this trend. Since 2000, TP maintained a decreasing trend, especially after the LSPP implementation, with a percentage of 47.93%.

The Chl-*a* of KB decreased significantly from 1992 to 2001 after the LSEMS implementation, with a percentage of 40.56%. TP and TN decreased significantly from 1994 to 2001 after the LSEMS implementation, with percentages of 37.53% and 18.23%, respectively. From 2002 to the LSPP implementation, Chl-*a*, TP, and TN fluctuated in ranges of 1.74–3.22 μg L^{-1}, 16.01–19.77 μg L^{-1}, and 0.41–0.52 mg L^{-1}, respectively. After the LSPP implementation, Chl-*a* and TP began to decrease significantly, with percentages of 55.56% and 39.94%, respectively, while TN increased to a certain extent, with a percentage of 4.43%.

Chl-*a* and TP in the east were higher than those in other areas except CB, while TN in the east was close to those in other areas except CB. From 1993 to 1999 after the LSEMS implementation, Chl-*a* and TP decreased rapidly, with percentages of 46.28% and 34.26%, respectively. From 2000 to the LSPP implementation, Chl-*a* and TP fluctuated in ranges of 2.54–3.73 μg L^{-1} and 18.73–24.11 μg L^{-1}, respectively. After the LSPP implementation, Chl-*a* and TP decreased significantly, with percentages of 36.76% and 29.54%, respectively. TN maintained an increase of 24.39% in six years after the LSEMS implementation. From 1996, TN began to decrease. But when dropped to 0.31 mg L^{-1} in 1999, TN began to rise to 0.53 mg L^{-1} in 2003. Since 2003, TN fluctuated in the range of 0.33–0.59 mg L^{-1}. Until the LSPP implementation, TN showed a slight decrease.

In the west, although Chl-*a* and TP increased significantly in the first four years after the LSEMS implementation, with percentages of 76.73% and 89.26%, respectively, they decreased continuously after the LSEMS and LSPP implementation, with percentages of 56.70% and 43.65%, respectively. After the LSEMS implementation, TN decreased from 1994 to 1999, with a percentage of 41.92%. From then to 2005, TN maintained an increasing trend with a percentage of 14.77%. Since 2005, including the LSPP implementation, TN showed a slight decreasing trend, with a percentage of 25.39%.

The variation of Chl-*a*, TP, and TN in the north was similar to that in the west from 1984 to 2020. After the LSEMS implementation, Chl-*a* decreased significantly from 1993 to 1997, with a percentage of 56.27%. From 1998 to 2003, Chl-*a* increased significantly, with a percentage of 83.87%. TP decreased significantly from 1993 to 1998 after the LSEMS implementation, with a percentage of 61.43%. From 1999 to 2003, TP increased significantly by 55.02%. Since 2003, including the implementation of LSPP, Chl-*a* and TP showed a decreasing trend, with percentages of 39.34% and 29.03%, respectively. After the LSEMS implementation, TN decreased significantly from 1994 to 1999, with a percentage of 37.25%. From 2000 to 2006, TN increased by 17.56%. Since 2006, including the LSPP implementation, TN showed a slight decreasing trend, with a percentage of 10.68%.

4. Discussion

4.1. Model strengths

The MDL models proposed in this study were demonstrated to perform well on the LSLMP test set and two independent data sets (LSCUF and LSRCA) (Section 3.2) for Chl-*a*, TP, and TN estimation, and outperformed several other candidate algorithms i.e., pDNN, sDNN, XGBoost, SVR, OC2, OC3, and MLR. As part of ML, DL has been proved to perform better than traditional ML in many fields (Zhu et al., 2017). In traditional ML, feature engineering before model training is necessary, because extracting features from a large amount of complex data using domain knowledge is considered to make the model more efficient. More specifically, to better explore the potential of ML for water quality retrieval, we usually manually seek spectral bands and their simple or complex derivative features (such as band ratio) most related to WQPs before model development. The performance of most traditional ML models depends on the accuracy of feature extraction in the feature engineering. DL is usually considered to have the ability to conduct feature engineering by itself. By combining and transforming the input low-level features, the more effective high-level features can be extracted for model prediction (Reichstein et al., 2019; Urban et al., 2016). In this study, to supplement the individual spectral bands, a set of simple RSIs were added as the model input features. The features could not be ensured to be highly correlated with the WQPs, and we only made a simple screening based on the Pearson correlation analyses (Section 3.1). Faced with such a large number of features containing redundant information, the DL models showed higher win rates than the traditional ML models.

In addition, we used the idea of multimodal deep learning to split all features into three categories ($F_{1\text{-band}}$, $F_{2\text{-band}}$, and $F_{3\text{-band}}$) according to the number of the source spectral bands. Subsequently, we constructed an sNN for each category of features. Each sNN mapped the input

features to a higher dimension and eventually reduced them to the input dimension to try to learn more useful information (Fig. 4b). We hypothesized that the three sNNs fully learned all features. Finally, the outputs of the three sNNs were concatenated and input into a new sNN for further information mining and the final prediction of the WQPs. According to the result that the MDL models outperformed the pDNN and even sDNN (with a similar number of trainable parameters) models, we inferred that learning each category of features independently through the sNNs could make the model efficiently control noises when learning more useful information. We also compared the model performances with several other feature splitting methods (Table A.5). The results showed that the MDL models performed better with the feature splitting of the $F_{1\text{-band}}$, $F_{2\text{-band}}$, and $F_{3\text{-band}}$. The model performances of most feature splitting methods were better than those of the sDNN. These results proved the contribution of the sNNs.

We also evaluated the comparability of the MDL models-derived Chl-*a*, TP, and TN by using the overlapping areas of the adjacent TM, ETM+, and OLI images (Zhang et al., 2021). For the comparison between TM and ETM+, the TM image on 31 July 2007 and ETM+ image on 1 August 2007 were selected. For the comparison between ETM+ and OLI, the ETM+ image on 15 September 2015 and OLI image on 16 September 2015 were selected (Fig. 13, Table A.6). Previous studies have shown that the time windows within several days are oftentimes used for the matching-up of images and water quality measurements, and are generally considered to produce reliable results (Chen et al., 2020). Recently, Olmanson et al., (2020) demonstrated that a more relaxed time window of up to ± 30 days could be used to accommodate the coarse temporal resolution of OLI in Minnesota lake environments.

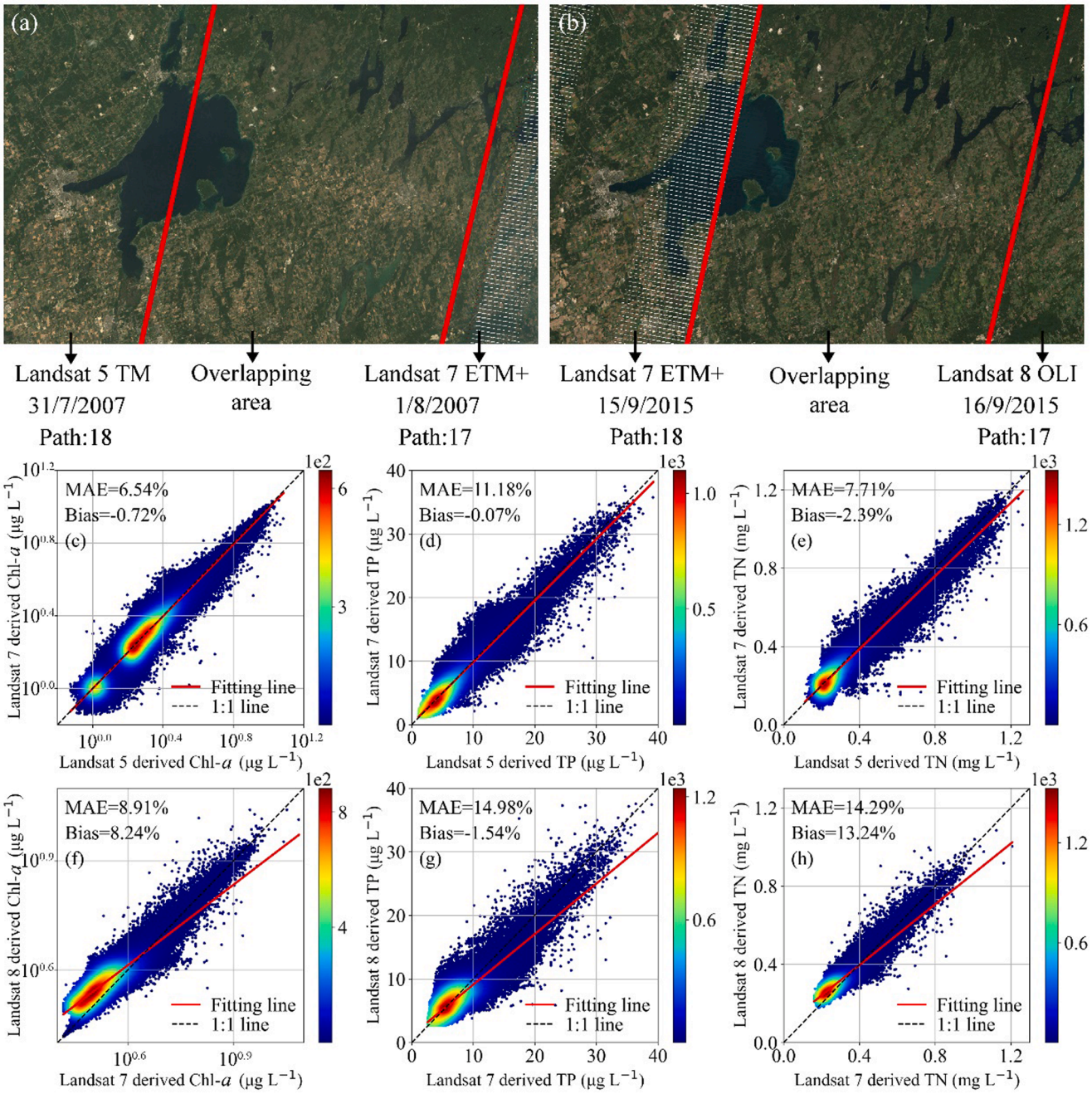

Fig. 13. Overlapping areas of the TM and ETM+ images (a) and the ETM+ and OLI images (b); comparisons of the retrieved Chl-*a* (c, f), TP (d, g), and TN (e, h) between TM (31 July 2007) and ETM+ (1 August 2007) and ETM+ (15 September 2015) and OLI (16 September 2015).

Therefore, here we ignored the impacts of the 1-day time windows of the two groups of images on the comparisons. The results indicated that the MDL models derived continuous Chl-*a*, TP, and TN estimates from TM, ETM+, and OLI. Generally, TM and ETM+ are considered to be comparable (Mancino et al., 2020), thus the Chl-*a* (MAE = 6.54%, Bias = −0.72%), TP (MAE = 11.18%, Bias = −0.07%), and TN (MAE = 7.71%, Bias = −2.39%) derived from TM and ETM+ showed good consistency. Although there exist substantial differences in design between OLI and TM/ETM+, and the consistency of the Chl-*a* (MAE = 8.91%, Bias = 8.24%), TP (MAE = 14.98%, Bias = −1.54%), and TN (MAE = 14.29%, Bias = 13.24%) estimates showed slight declines, we observed that the consistency was also better than that of the four spectral bands (i.e., blue, green, red, and NIR bands) (Table A.6). These results proved that the MDL models had the potential to independently learn the correlations between TM, ETM+, and OLI R_{rs} and WQPs and generate continuous time series water quality data sets using Landsat data.

4.2. Model limitations

The MDL models proposed in this study also had some limitations. First, as data-driven models, the model performance was highly dependent on data. All ground truth data for model training and validation was from the in-situ measurements of Lake Simcoe. Although these data sets covered sufficient water quality conditions of Lake Simcoe in time and space, the lack of training and validation data of other water bodies would lead to weak model generalization. For example, since the Chl-*a* data was only ranging in 0–11 μg L^{-1} (Fig. A.5), applying the models directly to lakes with more serious eutrophication might be unfeasible. Therefore, the models proposed in this study were judged to be effective at the regional scale, and it was necessary to use new data to further calibrate the models before applying them to other water bodies. Second, the sample size in this study was not large enough for ML, although we strictly controlled the number of features being input into the models by the Pearson correlation analyses to reduce the required sample size. Generating data sets of larger size by including in-situ measurements from more water bodies was expected in future research. Third, we observed that the MDL model almost failed to estimate TN in the low concentration region (< 3 mg L^{-1}), although the sample size in this region was not small (Fig. A.5). We inferred that the low concentration of non-OAC TN could not significantly change the concentration of the OACs in water, resulting in that the models not fully learned the correlation between the low TN concentration and the spectral characteristics. The correlations of TN with Chl-*a* and SDD (Fig. 3) also showed that TN hardly changed with Chl-*a* and SDD in the concentration range of 0–0.3 mg L^{-1}. However, the "black box" property of DL limited further analysis of the causes of the TN estimation failure in the low concentration region. In future research, biochemical experiments might be used to further clarify the correlation between low TN concentrations and OACs to address this issue. Fourth, compared with traditional ML models, DL models have more complex structures, more hyperparameters, and higher error susceptibility. These characteristics make the DL models more prone to overfitting. Therefore, during the model development, carefully designing the model structure, tuning each hyperparameter as well as strictly monitoring the model losses on both training set and test set are essential. Besides, the high requirement of computational power and long training time also challenge the model development.

4.3. MDL effectively separated Chl-a, TP, and TN estimates

P and N are the main limiting nutrients for the growth of aquatic plants and are highly homologous, hence the concentrations of Chl-*a*, TP, and TN are cross-correlated. This challenges the accurate retrieval of any WQP without the impacts of other WQPs. To analyze the feasibility of the effective separation of Chl-*a*, TP, and TN estimates by the MDL models and improve the model interpretability to a certain extent, we used the integrated gradients algorithm in Captum (version 0.3.1) (Kokhlikyan et al., 2020) to calculate the contribution of each feature to the model outputs (Fig. 14). Captum is a model interpretability library for PyTorch. Using the state-of-the-art algorithms integrated into Captum, the contributions of features, layers, and neurons to the DL models can be quantified as a score within [–1, 1]. Zero means no contribution whereas positives indicate positive contributions and negatives the opposite. For convenient comparison, we used zero-valued baselines for all features in the contribution calculation. Different baseline settings lead to different contribution scores, even the reversal of positive and negative. Therefore, the results were relative contribution scores when the baselines of all features were zero.

Fig. 14 indicated that the contribution score of each feature to the MDL models for Chl-*a*, TP, and TN estimation was different. For example, in the $F_{1\text{-band}}$, F2 (green band) had the highest positive contribution to Chl-*a* and TP estimates, with scores of 0.75 and 0.67, while F3 (red band) had the highest positive contribution to TN, with a score of 0.62. This result matched the Pearson correlation analysis (Fig. 5). It was worth mentioning that the contribution scores were not always aligned with *r*. F4 (NIR band) had positive contributions to TP and TN, but a negative contribution to Chl-*a*. Although the input $F_{2\text{-band}}$ and $F_{3\text{-band}}$ of each WQP model were different after screening by the threshold, the same results were also observed. F9 and F16 had positive contributions to Chl-*a* and TN, but negative contributions to TP. F17 had negative contributions to TP and TN, but a positive contribution to Chl-*a*. Besides, we also observed that some features did not contribute to the models, for example, the contributions of F3 to Chl-*a* and TP were almost zero. Since the baselines of all features were set to zero, these contribution scores were only relative results. However, we speculated that it was the features that contributed to the models and the differences in their contributions that provided the basis for the MDL models to effectively separate Chl-*a*, TP, and TN estimates.

4.4. Feasibility of long-term water quality retrieval

To generate long-term Chl-*a*, TP, and TN patterns of Lake Simcoe from Landsat raw images, we first used the DSF algorithm integrated into ACOLITE to atmospherically correct the TM, ETM+, and OLI raw images to R_{rs} data. Then, the ground truth water quality data was obtained from the long-term (2007–2019) in-situ measurements of 22 automatic monitoring stations widely distributed in Lake Simcoe. We considered that the water quality data could cover sufficient water quality conditions of Lake Simcoe. Subsequently, MDL models were constructed to estimate Chl-*a*, TP, and TN of Lake Simcoe using the R_{rs} data. We rigorously validated the models, and the results indicated that the MDL models performed well on the LSLMP test set and two independent data sets (i.e., LSCUF and LSRCA) and outperformed four candidate DL and ML models and three other empirical algorithms. Based on the above analysis, we considered that the MDL models proposed in this study could be used to reconstruct the long-term spatial distributions of Chl-*a*, TP, and TN in Lake Simcoe with the Landsat archive.

However, the coarse temporal resolution of the Landsat images and noises (e.g., cloud) resulted in quite a small number of utilizable images in one year, which inevitably challenged the annual average accuracy of water quality. We calculated the absolute differences (ADs) of the annual average in-situ measured and satellite estimated water quality of Lake Simcoe from 2007 to 2019, and quantitively analyzed the relationship between the ADs and the number of utilizable images (Fig. 15). The sample size of the in-situ measurements was more than 110 every year, covering spring, summer, and autumn, which was considered to objectively represent the annual average water quality of Lake Simcoe. The results showed that the average ADs of Chl-*a*, TP, and TN were 1.12 μg L^{-1}, 5.14 μg L^{-1}, and 0.42 mg L^{-1}, respectively, and the ADs increased with the decrease of the utilizable image number. In 2014, the ADs of Chl-*a*, TP, and TN reached 1.21 μg L^{-1}, 8.57 μg L^{-1}, and 0.67 mg

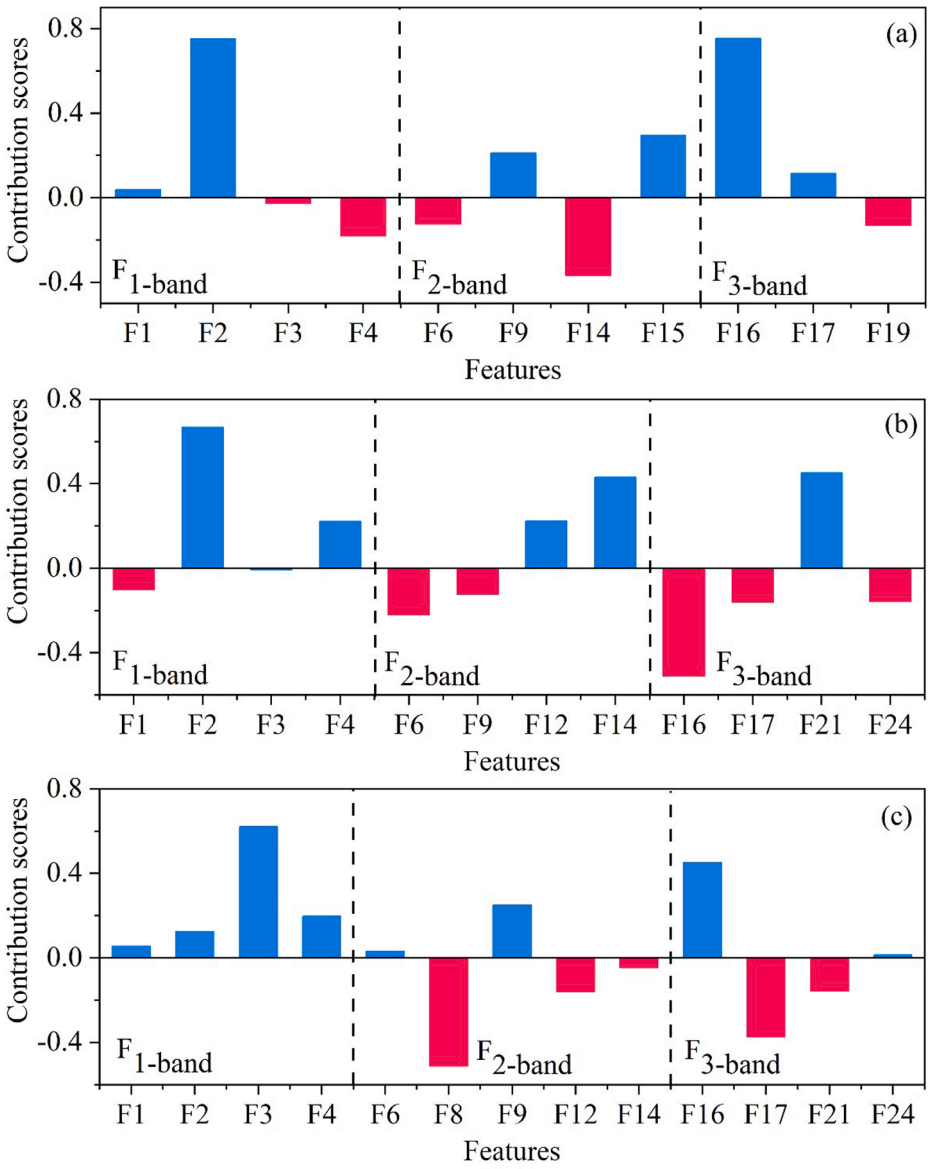

Fig. 14. The contribution scores of each feature to the Chl-*a* (a), TP (b), and TN (c) estimates. Note that the results were relative contribution scores when the baselines of all features were zero.

L^{-1}, respectively, due to that the water quality was estimated with only one utilizable image. While when the number of the utilizable images increased to four, ADs of Chl-*a*, TP, and TN significantly decreased to 1.16 μg L^{-1}, 5.41 μg L^{-1}, and 0.46 mg L^{-1}, with percentages of 6.76%, 36.87%, and 31.34%, respectively. When the number of the utilizable images was greater than six in 2015 and 2016, the ADs of Chl-*a*, TP, and TN were 1.06 μg L^{-1}, 2.79 μg L^{-1}, and 0.20 mg L^{-1} respectively. These results indicated that the annual average water quality generated in this study had a certain deviation compared with the in-situ measurements owing to the limited utilizable image number. Therefore, we considered that these annual averages were only applicable to long-term water quality change trend analysis of Lake Simcoe, but not to the quantitative delineation year by year.

5. Conclusions

In this study, MDL models were developed and validated with R_{rs} data derived from Landsat raw images and synchronous water quality measurements in Lake Simcoe for long-term Chl-*a*, TP, and TN estimation. The MDL models were trained with 70% LSLMP data (for Chl-*a*, $N = 315$, for TP and TN, $N = 303$) and validated with 30% LSLMP data (for Chl-*a*, $N = 136$, for TP and TN, $N = 130$) as well as two independent data sets, namely LSCUF ($N = 89$) and LSRCA ($N = 58$). The performances of the MDL models were also compared with other DL, ML, and empirical models (i.e., pDNN, sDNN, XGBoost, SVR, OC2, OC3, and MLR). The results indicated that the MDL models adequately estimated Chl-*a* (MAE = 32.57%, Bias = 10.61%), TP (MAE = 42.58%, Bias = −2.82%), and TN (MAE = 35.05%, Bias = 13.66%) of Lake Simcoe, and outperformed the candidate models. Although TP and TN are non-OACs, the MDL models effectively captured the potential correlations among OACs, TP and TN, and R_{rs}, promoting direct non-OAC retrieval without prior OAC retrieval. Using the MDL models, the spatial distributions of Chl-*a*, TP, and TN of Lake Simcoe since 1984 were mapped, and in sequence, the water quality improvement effects of the LSEMS implemented in 1990 and the LSPP implemented in 2009 were quantitatively analyzed. The MDL models were demonstrated to be capable to reconstruct the long-term water quality spatial distribution and contribute to efficient water quality monitoring in Lake Simcoe using Landsat R_{rs} data and in-situ measurements covering sufficient water quality conditions. However, due to the limitation of the training data, MDL models were

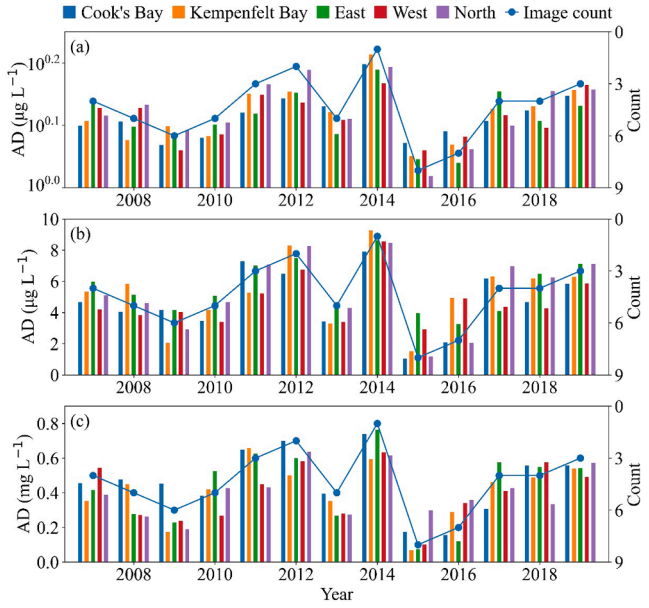

Fig. 15. The number of utilizable images and the ADs of the annual average in-situ measured and satellite estimated Chl-*a* (a), TP (b), and TN (c) from 2007 to 2019.

considered effective at the regional scale. It was necessary to recalibrate the models with new data if the models needed to be applied in other water bodies. In addition, the cross-correlation among Chl-*a*, TP, and TN challenges the accurate quantification of any WQP without the impacts of other WQPs. This study quantified the contribution of each feature to the models, thus improving the interpretability of the MDL models for effective separation of Chl-*a*, TP, and TN estimates to a certain extent. The MDL models provide a practical regional-scale method for long-term Chl-*a*, TP, and TN estimation using the Landsat archive.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was supported by the National Key Research and Development Program of China [No. 2016YFC0400709]. We appreciate the careful and thoughtful comments of two anonymous reviewers, which helped improve this paper substantially. We also acknowledge Mike Walter, Steve Auger, Ben Longstaff, and Brian Gann at Lake Simcoe Region Conservation Authority and James Yiuwah Li at Ryerson University for providing valuable comments on this study.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.isprsjprs.2021.11.023.

References

Ahamed, T., Tian, L., Zhang, Y., Ting, K.C., 2011. A review of remote sensing methods for biomass feedstock production. Biomass Bioenergy 35 (7), 2455–2469. https://doi.org/10.1016/j.biombioe.2011.02.028.

Audebert, N., Le Saux, B., Lefèvre, S., 2018. Beyond RGB: Very high resolution urban remote sensing with multimodal deep networks. ISPRS J. Photogramm. Remote Sens. 140, 20–32. https://doi.org/10.1016/j.isprsjprs.2017.11.011.

Bierman, P., Lewis, M., Ostendorf, B., Tanner, J., 2011. A review of methods for analysing spatial and temporal patterns in coastal Water quality. Ecol. Indic. 11 (1), 103–114. https://doi.org/10.1016/j.ecolind.2009.11.001.

Bottou, L., 2012. Stochastic gradient descent tricks. Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics) 7700 LECTU, 421–436. https://doi.org/10.1007/978-3-642-35289-8_25.

Cao, Z., Ma, R., Duan, H., Pahlevan, N., Melack, J., Shen, M., Xue, K., 2020. A machine learning approach to estimate chlorophyll-a from Landsat-8 measurements in inland lakes. Remote Sens. Environ. 248, 111974. https://doi.org/10.1016/j.rse.2020.111974.

Cao, Z., Ma, R., Duan, H., Xue, K., 2019. Effects of broad bandwidth on the remote sensing of inland waters: Implications for high spatial resolution satellite data applications. ISPRS J. Photogramm. Remote Sens. 153, 110–122. https://doi.org/10.1016/j.isprsjprs.2019.05.001.

Chang, N.B., Xuan, Z., Yang, Y.J., 2013. Exploring spatiotemporal patterns of phosphorus concentrations in a coastal bay with MODIS images and machine learning models. Remote Sens. Environ. 134, 100–110. https://doi.org/10.1016/j.rse.2013.03.002.

Chawla, I., Karthikeyan, L., Mishra, A.K., 2020. A review of remote sensing applications for water security: Quantity, quality, and extremes. J. Hydrol. 585, 124826. https://doi.org/10.1016/j.jhydrol.2020.124826.

Chen, J., Zhu, W., Tian, Y.Q., Yu, Q., 2020. Monitoring dissolved organic carbon by combining Landsat-8 and Sentinel-2 satellites: Case study in Saginaw River estuary. Lake Huron. Sci. Total Environ. 718, 137374. https://doi.org/10.1016/j.scitotenv.2020.137374.

Chen, T., Guestrin, C., 2016. XGBoost: A scalable tree boosting system, in. In: Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 785–794. https://doi.org/10.1145/2939672.2939785.

Cortes, C., Vapnik, V., 1995. Support-Vector Networks. Mach. Learn. 20, 273–297. https://doi.org/10.1023/A:1022627411411.

Crossman, J., Futter, M.N., Palmer, M., Whitehead, P.G., Baulch, H.M., Woods, D., Jin, L., Oni, S.K., Dillon, P.J., 2016. The effectiveness and resilience of phosphorus management practices in the Lake Simcoe watershed, Ontario. Canada. J. Geophys. Res. G Biogeosciences 121 (9), 2390–2409. https://doi.org/10.1002/jgrg.v121.910.1002/2015JG003253.

Escadafal. R, B.A. and B.M.H., 1994. Indices spectraux pour la dégradation des milieux naturels en Tunisie aride. Proc. 6e'me Symp. Int. Mes. Phys. Signatures en Télédétection, ISPRS-CNES, Val d'lsère, Fr. 253–259.

Fan, Y., Li, W., Gatebe, C.K., Jamet, C., Zibordi, G., Schroeder, T., Stamnes, K., 2017. Atmospheric correction over coastal waters using multilayer neural networks. Remote Sens. Environ. 199, 218–240. https://doi.org/10.1016/j.rse.2017.07.016.

GAMON, J.A., SURFUS, J.S., 1999. Assessing leaf pigment content and activity with a reflectometer. New Phytol. 143 (1), 105–117. https://doi.org/10.1046/j.1469-8137.1999.00424.x.

Gao, L., Wang, X., Johnson, B.A., Tian, Q., Wang, Y., Verrelst, J., Mu, X., Gu, X., 2020. Remote sensing algorithms for estimation of fractional vegetation cover using pure vegetation index values: A review. ISPRS J. Photogramm. Remote Sens. 159, 364–377. https://doi.org/10.1016/j.isprsjprs.2019.11.018.

Giardino, C., Pepe, M., Brivio, P.A., Ghezzi, P., Zilioli, E., 2001. Detecting chlorophyll, Secchi disk depth and surface temperature in a sub-alpine lake using Landsat imagery. Sci. Total Environ. 268 (1-3), 19–29. https://doi.org/10.1016/S0048-9697(00)00692-6.

Gitelson, A.A., Kaufman, Y.J., Stark, R., Rundquist, D., 2002. Novel algorithms for remote estimation of vegetation fraction. Remote Sens. Environ. 80 (1), 76–87. https://doi.org/10.1016/S0034-4257(01)00289-9.

Gitelson, A.A., Viña, A., Arkebauer, T.J., Rundquist, D.C., Keydan, G., Leavitt, B., 2003. Remote estimation of leaf area index and green leaf biomass in maize canopies. Geophys. Res. Lett. 30 (5), n/a–n/a. https://doi.org/10.1029/2002GL016450.

Gons, H.J., 1999. Optical teledetection of chlorophyll a in turbid inland waters. Environ. Sci. Technol. 33 (7), 1127–1132. https://doi.org/10.1021/es9809657.

Griffin, C.G., McClelland, J.W., Frey, K.E., Fiske, G., Holmes, R.M., 2018. Quantifying CDOM and DOC in major Arctic rivers during ice-free conditions using Landsat TM and ETM+ data. Remote Sens. Environ. 209, 395–409. https://doi.org/10.1016/j.rse.2018.02.060.

Guo, H., Huang, J.J., Chen, B., Guo, X., Singh, V.P., 2021a. A machine learning-based strategy for estimating non-optically active water quality parameters using Sentinel-2 imagery. Int. J. Remote Sens. 42 (5), 1841–1866. https://doi.org/10.1080/01431161.2020.1846222.

Guo, H., Huang, J.J., Zhu, X., Wang, B.o., Tian, S., Xu, W., Mai, Y., 2021b. A generalized machine learning approach for dissolved oxygen estimation at multiple spatiotemporal scales using remote sensing. Environ. Pollut. 288, 117734. https://doi.org/10.1016/j.envpol.2021.117734.

Hancock, D.W., Dougherty, C.T., 2007. Relationships between blue- and red-based vegetation indices and leaf area and yield of alfalfa. Crop Sci. 47 (6), 2547–2556. https://doi.org/10.2135/cropsci2007.01.0031.

Hewson, R.D., Cudahy, T.J., Huntington, J.F., 2001. Geologic and alteration mapping at Mt fitton, South Australia, using ASTER satellite-borne data. In: International Geoscience and Remote Sensing Symposium (IGARSS), pp. 724–726, 10.1109/igarss.2001.976615.

Holyer, R.J., 1978. Toward universal multispectral suspended sediment algorithms. Remote Sens. Environ. 7 (4), 323–338. https://doi.org/10.1016/0034-4257(78)90023-8.

Hong, D., Gao, L., Yokoya, N., Yao, J., Chanussot, J., Du, Q., Zhang, B., 2021. More Diverse Means Better: Multimodal Deep Learning Meets Remote-Sensing Imagery Classification. IEEE Trans. Geosci. Remote Sens. 59 (5), 4340–4354. https://doi.org/10.1109/TGRS.2020.3016820.

Hunter, P.D., Tyler, A.N., Gilvear, D.J., Willby, N.J., 2009. Using remote sensing to aid the assessment of human health risks from blooms of potentially toxic cyanobacteria. Environ. Sci. Technol. 43 (7), 2627–2633. https://doi.org/10.1021/es802977u.

Ilori, C.O., Pahlevan, N., Knudby, A., 2019. Analyzing performances of different atmospheric correction techniques for Landsat 8: Application for coastal remote sensing. Remote Sens. 11, 469. https://doi.org/10.3390/rs11040469.

Kokhlikyan, N., Miglani, V., Martin, M., Wang, E., Alsallakh, B., Reynolds, J., ... & Reblitz-Richardson, O. (2020). Captum: A unified and generic model interpretability library for pytorch. arXiv preprint arXiv:2009.07896.

Kwiatkowska, E.J., Fargion, G.S., 2003. Application of machine-learning techniques toward the creation of a consistent and calibrated global chlorophyll concentration baseline dataset using remotely sensed ocean color data. IEEE Trans. Geosci. Remote Sens. 41 (12), 2844–2860. https://doi.org/10.1109/TGRS.2003.818016.

le Maire, G., François, C., Dufrêne, E., 2004. Towards universal broad leaf chlorophyll indices using PROSPECT simulated database and hyperspectral reflectance measurements. Remote Sens. Environ. 89 (1), 1–28. https://doi.org/10.1016/j.rse.2003.09.004.

Liu, G.e., Li, L., Song, K., Li, Y., Lyu, H., Wen, Z., Fang, C., Bi, S., Sun, X., Wang, Z., Cao, Z., Shang, Y., Yu, G., Zheng, Z., Huang, C., Xu, Y., Shi, K., 2020. An OLCI-based algorithm for semi-empirically partitioning absorption coefficient and estimating chlorophyll a concentration in various turbid case-2 waters. Remote Sens. Environ. 239, 111648. https://doi.org/10.1016/j.rse.2020.111648.

Lu, S., Deng, R., Liang, Y., Xiong, L., Ai, X., Qin, Y., 2020. Remote sensing retrieval of total phosphorus in the pearl river channels based on the GF-1 remote sensing data. Remote Sens. 12, 1420. https://doi.org/10.3390/RS12091420.

Ma, L., Liu, Y., Zhang, X., Ye, Y., Yin, G., Johnson, B.A., 2019. Deep learning in remote sensing applications: A meta-analysis and review. ISPRS J. Photogramm. Remote Sens. 152, 166–177. https://doi.org/10.1016/j.isprsjprs.2019.04.015.

Maimaitijiang, M., Sagan, V., Sidike, P., Hartling, S., Esposito, F., Fritschi, F.B., 2020. Soybean yield prediction from UAV using multimodal data fusion and deep learning. Remote Sens. Environ. 237, 111599. https://doi.org/10.1016/j.rse.2019.111599.

Mancino, G., Ferrara, A., Padula, A., Nolè, A., 2020. Cross-comparison between Landsat 8 (OLI) and Landsat 7 (ETM+) derived vegetation indices in a Mediterranean Environment. Remote Sens. 12, 291. https://doi.org/10.3390/rs12020291.

Masek, J.G., Wulder, M.A., Markham, B., McCorkel, J., Crawford, C.J., Storey, J., Jenstrom, D.T., 2020. Landsat 9: Empowering open science and applications through continuity. Remote Sens. Environ. 248, 111968. https://doi.org/10.1016/j.rse.2020.111968.

Matthews, M.W., 2011. A current review of empirical procedures of remote sensing in Inland and near-coastal transitional waters. Int. J. Remote Sens. 32 (21), 6855–6899. https://doi.org/10.1080/01431161.2010.512947.

Matthews, M.W., Odermatt, D., 2015. Improved algorithm for routine monitoring of cyanobacteria and eutrophication in inland and near-coastal waters. Remote Sens. Environ. 156, 374–382. https://doi.org/10.1016/j.rse.2014.10.010.

Neil, C., Spyrakos, E., Hunter, P.D., Tyler, A.N., 2019. A global approach for chlorophyll-a retrieval across optically complex inland waters based on optical water types. Remote Sens. Environ. 229, 159–178. https://doi.org/10.1016/j.rse.2019.04.027.

Ngiam, J., Khosla, A., Kim, M., Nam, J., Lee, H., Ng, A.Y., 2011. Multimodal deep learning, in: Proceedings of the 28th International Conference on Machine Learning, ICML 2011. pp. 689–696.

O'Reilly, J.E., Maritorena, S., Mitchell, B.G., Siegel, D.A., Carder, K.L., Garver, S.A., Kahru, M., McClain, C., 1998. Ocean color chlorophyll algorithms for SeaWiFS. J. Geophys. Res. Ocean. 103 (C11), 24937–24953. https://doi.org/10.1029/98JC02160.

O'Reilly, J.E., Werdell, P.J., 2019. Chlorophyll algorithms for ocean color sensors - OC4, OC5 & OC6. Remote Sens. Environ. 229, 32–47. https://doi.org/10.1016/j.rse.2019.04.021.

Odermatt, D., Gitelson, A., Brando, V.E., Schaepman, M., 2012. Review of constituent retrieval in optically deep and complex waters from satellite imagery. Remote Sens. Environ. 118, 116–126. https://doi.org/10.1016/j.rse.2011.11.013.

Olmanson, L.G., Page, B.P., Finlay, J.C., Brezonik, P.L., Bauer, M.E., Griffin, C.G., Hozalski, R.M., 2020. Regional measurements and spatial/temporal analysis of CDOM in 10,000+ optically variable Minnesota lakes using Landsat 8 imagery. Sci. Total Environ. 724, 138141. https://doi.org/10.1016/j.scitotenv.2020.138141.

Paerl, H.W., Otten, T.G., 2013. Harmful Cyanobacterial Blooms: Causes, Consequences, and Controls. Microb. Ecol. 65 (4), 995–1010. https://doi.org/10.1007/s00248-012-0159-y.

Page, B.P., Olmanson, L.G., Mishra, D.R., 2019. A harmonized image processing workflow using Sentinel-2/MSI and Landsat-8/OLI for mapping water clarity in optically variable lake systems. Remote Sens. Environ. 231, 111284. https://doi.org/10.1016/j.rse.2019.111284.

Pahlevan, N., Balasubramanian, S.V., Sarkar, S., Franz, B.A., 2018. Toward long-term aquatic science products from heritage Landsat missions. Remote Sens. 10, 1337. https://doi.org/10.3390/rs10091337.

Pahlevan, N., Lee, Z., Wei, J., Schaaf, C.B., Schott, J.R., Berk, A., 2014. On-orbit radiometric characterization of OLI (Landsat-8) for applications in aquatic remote sensing. Remote Sens. Environ. 154, 272–284. https://doi.org/10.1016/j.rse.2014.08.001.

Pahlevan, N., Schott, J.R., Franz, B.A., Zibordi, G., Markham, B., Bailey, S., Schaaf, C.B., Ondrusek, M., Greb, S., Strait, C.M., 2017. Landsat 8 remote sensing reflectance (Rrs) products: Evaluations, intercomparisons, and enhancements. Remote Sens. Environ. 190, 289–301. https://doi.org/10.1016/j.rse.2016.12.030.

Pahlevan, N., Smith, B., Schalles, J., Binding, C., Cao, Z., Ma, R., Alikas, K., Kangro, K., Gurlin, D., Hà, N., Matsushita, B., Moses, W., Greb, S., Lehmann, M.K., Ondrusek, M., Oppelt, N., Stumpf, R., 2020. Seamless retrievals of chlorophyll-a from Sentinel-2 (MSI) and Sentinel-3 (OLCI) in inland and coastal waters: A machine-learning approach. Remote Sens. Environ. 240, 111604. https://doi.org/10.1016/j.rse.2019.111604.

Palmer, S.C.J., Kutser, T., Hunter, P.D., 2015. Remote sensing of inland waters: Challenges, progress and future directions. Remote Sens. Environ. 157, 1–8. https://doi.org/10.1016/j.rse.2014.09.021.

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Köpf, A., Yang, E., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., Bai, J., Chintala, S., 2019. PyTorch: An Imperative Style, High-Performance Deep Learning Library. Adv. Neural Inform. Process. Syst., 32 Curran Associates, Inc., pp. 8024–8035

Peterson, K.T., Sagan, V., Sloan, J.J., 2020. Deep learning-based water quality estimation and anomaly detection using Landsat-8/Sentinel-2 virtual constellation and cloud computing. GIScience Remote Sens. 57 (4), 510–525. https://doi.org/10.1080/15481603.2020.1738061.

Pu, F., Ding, C., Chao, Z., Yu, Y., Xu, X., 2019. Water-quality classification of inland lakes using Landsat8 images by convolutional neural networks. Remote Sens. 11, 1674. https://doi.org/10.3390/rs11141674.

Ramachandram, D., Taylor, G.W., 2017. Deep multimodal learning: A survey on recent advances and trends. IEEE Signal Process. Mag. 34 (6), 96–108. https://doi.org/10.1109/MSP.2017.2738401.

Hunt, E.R., Daughtry, C.S.T., Eitel, J.U.H., Long, D.S., 2011. Remote sensing leaf chlorophyll content using a visible band index. Agron. J. 103 (4), 1090–1099. https://doi.org/10.2134/agronj2010.0395.

Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N., Prabhat, 2019. Deep learning and process understanding for data-driven Earth system science. Nature 566, 195–204. https://doi.org/10.1038/s41586-019-0912-1.

Ritchie, J.C., Schiebe, F.R., McHenry, J.R., 1976. Remote sensing of suspended sediments in surface waters. Photogramm. Remote Sens. 42, 1539–1545.

Sagan, Vasit, Peterson, Kyle T., Maimaitijiang, Maitiniyazi, Sidike, Paheding, Sloan, John, Greeling, Benjamin A., Maalouf, Samar, Adams, Craig, 2020. Monitoring inland water quality using remote sensing: potential and limitations of spectral indices, bio-optical simulations, machine learning, and cloud computing. Earth-Science Rev. 205, 103187. https://doi.org/10.1016/j.earscirev.2020.103187.

Schmidt, G., Jenkerson, C., Masek, J., Vermote, E.F., Gao, F., 2013. Landsat Ecosystem Disturbance Adaptive Processing System (LEDAPS) Algorithm Description. Open-file Rep. 2013–1057, 1–27.

Seegers, B.N., Stumpf, R.P., Schaeffer, B.A., Loftin, K.A., Werdell, P.J., 2018. Performance metrics for the assessment of satellite data products: an ocean color case study. Opt. Express 26 (6), 7404. https://doi.org/10.1364/OE.26.007404.

Setiawan, F., Matsushita, B., Hamzah, R., Jiang, D., Fukushima, T., 2019. Long-term change of the secchi disk depth in Lake Maninjau, Indonesia shown by landsat TM and ETM+ data. Remote Sens. 11, 2875. https://doi.org/10.3390/rs11232875.

Smith, B., Pahlevan, N., Schalles, J., Ruberg, S., Errera, R., Ma, R., Giardino, C., Bresciani, M., Barbosa, C., Moore, T., Fernandez, V., Alikas, K., Kangro, K., 2021. A Chlorophyll-a Algorithm for Landsat-8 Based on Mixture Density Networks. Front. Remote Sens. 1, 5. https://doi.org/10.3389/frsen.2020.623678.

Song, Kaishan, Li, Lin, Li, Shuai, Tedesco, Lenore, Hall, Bob, Li, Linhai, 2012. Hyperspectral remote sensing of total phosphorus (TP) in three central Indiana water supply reservoirs. Water. Air. Soil Pollut. 223 (4), 1481–1502. https://doi.org/10.1007/s11270-011-0959-6.

Spyrakos, E., O'Donnell, R., Hunter, P.D., Miller, C., Scott, M., Simis, S.G.H., Neil, C., Barbosa, C.C.F., Binding, C.E., Bradt, S., Bresciani, M., Dall'Olmo, G., Giardino, C., Gitelson, A.A., Kutser, T., Li, L., Matsushita, B., Martinez-Vicente, V., Matthews, M.W., Ogashawara, I., Ruiz-Verdú, A., Schalles, J.F., Tebbs, E., Zhang, Y., Tyler, A.N., 2018. Optical types of inland and coastal waters. Limnol. Oceanogr. 63, 846–870. https://doi.org/10.1002/lno.10674.

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., Salakhutdinov, R., 2014. Dropout: A simple way to prevent neural networks from overfitting. J. Mach. Learn. Res. 15, 1929–1958.

Suel, Esra, Bhatt, Samir, Brauer, Michael, Flaxman, Seth, Ezzati, Majid, 2021. Multimodal deep learning from satellite and street-level imagery for measuring income, overcrowding, and environmental deprivation in urban areas. Remote Sens. Environ. 257, 112339. https://doi.org/10.1016/j.rse.2021.112339.

Tavares, M.H., Lins, R.C., Harmel, T., Fragoso, C.R., Martínez, J.M., Motta-Marques, D., 2021. Atmospheric and sunglint correction for retrieving chlorophyll-a in a productive tropical estuarine-lagoon system using Sentinel-2 MSI imagery. ISPRS J. Photogramm. Remote Sens. 174, 215–236. https://doi.org/10.1016/j.isprsjprs.2021.01.021.

Urban, G., Geras, K. J., Kahou, S. E., Aslan, O., Wang, S., Caruana, R., ... & Richardson, M. (2016). Do deep convolutional nets really need to be deep and convolutional?. arXiv preprint arXiv:1603.05691.

Uudeberg, K., Ansko, I., Põru, G., Ansper, A., Reinart, A., 2019. Using opticalwater types to monitor changes in optically complex inland and coastalwaters. Remote Sens. 11, 2297. https://doi.org/10.3390/rs11192297.

Vanhellemont, Q., 2019. Adaptation of the dark spectrum fitting atmospheric correction for aquatic applications of the Landsat and Sentinel-2 archives. Remote Sens. Environ. 225, 175–192. https://doi.org/10.1016/j.rse.2019.03.010.

Vermote, E., Justice, C., Claverie, M., Franch, B., 2016. Preliminary analysis of the performance of the Landsat 8/OLI land surface reflectance product. Remote Sens. Environ. 185, 46–56. https://doi.org/10.1016/j.rse.2016.04.008.

Wang, D., Cui, Q., Gong, F., Wang, L., He, X., Bai, Y., 2018. Satellite retrieval of surface water nutrients in the coastal regions of the East China Sea. Remote Sens. 10, 1896. https://doi.org/10.3390/rs10121896.

WANG, Fu-min, HUANG, Jing-feng, TANG, Yan-lin, WANG, Xiu-zhen, 2007. New Vegetation Index and Its Application in Estimating Leaf Area Index of Rice. Rice Sci. 14 (3), 195–203. https://doi.org/10.1016/S1672-6308(07)60027-4.

Wang, Shenglei, Li, Junsheng, Zhang, Bing, Lee, Zhongping, Spyrakos, Evangelos, Feng, Lian, Liu, Chong, Zhao, Hongli, Wu, Yanhong, Zhu, Liping, Jia, Liming, Wan, Wei, Zhang, Fangfang, Shen, Qian, Tyler, Andrew N., Zhang, Xianfeng, 2020. Changes of water clarity in large lakes and reservoirs across China observed from long-term MODIS. Remote Sens. Environ. 247, 111949. https://doi.org/10.1016/j.rse.2020.111949.

Xiong, Yujiu, Ran, Yili, Zhao, Shaohua, Zhao, Huan, Tian, Qixuan, 2020. Remotely assessing and monitoring coastal and inland water quality in China: Progress, challenges and outlook. Crit. Rev. Environ. Sci. Technol. 50 (12), 1266–1302. https://doi.org/10.1080/10643389.2019.1656511.

Yuan, Qiangqiang, Shen, Huanfeng, Li, Tongwen, Li, Zhiwei, Li, Shuwen, Jiang, Yun, Xu, Hongzhang, Tan, Weiwei, Yang, Qianqian, Wang, Jiwen, Gao, Jianhao, Zhang, Liangpei, 2020. Deep learning in environmental remote sensing: Achievements and challenges. Remote Sens. Environ. 241, 111716. https://doi.org/10.1016/j.rse.2020.111716.

Zhang, Yibo, Zhang, Yunlin, Shi, Kun, Zhou, Yongqiang, Li, Na, 2021. Remote sensing estimation of water clarity for various lakes in China. Water Res. 192, 116844. https://doi.org/10.1016/j.watres.2021.116844.

Zhu, Xiao Xiang, Tuia, Devis, Mou, Lichao, Xia, Gui-Song, Zhang, Liangpei, Xu, Feng, Fraundorfer, Friedrich, 2017. Deep Learning in Remote Sensing: A Comprehensive Review and List of Resources. IEEE Geosci. Remote Sens. Mag. 5 (4), 8–36. https://doi.org/10.1109/MGRS.2017.2762307.

Zibordi, Giuseppe, Berthon, Jean-François, Mélin, Frédéric, D'Alimonte, Davide, Kaitala, Seppo, 2009. Validation of satellite ocean color primary products at optically complex coastal sites: Northern Adriatic Sea, Northern Baltic Proper and Gulf of Finland. Remote Sens. Environ. 113 (12), 2574–2591. https://doi.org/10.1016/j.rse.2009.07.013.