Remote Sensing of Environment 253 (2021) 112200

Contents lists available at ScienceDirect

Remote Sensing of Environment

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

Hyperspectral retrievals of phytoplankton absorption and chlorophyll-*a* in inland and nearshore coastal waters

Nima Pahlevan [a,b,*], Brandon Smith [a,b], Caren Binding [c], Daniela Gurlin [d], Lin Li [e], Mariano Bresciani [f], Claudia Giardino [f]

[a] NASA Goddard Space Flight Center, Greenbelt, MD, USA
[b] Science Systems and Applications, Inc. (SSAI), Lanham, MD, USA
[c] Environment and Climate Change Canada, Burlington, ON, Canada
[d] Wisconsin Department of Natural Resources, Madison, WI, USA
[e] Purdue School of Science, Indiana University-Purdue University, IN, USA
[f] National Research Council of Italy, IREA, Milan, Italy

ARTICLE INFO

Keywords:
Hyperspectral
Inland and coastal waters
HICO
Phytoplankton absorption
Chlorophyll-*a*
Machine learning
Algorithm development

ABSTRACT

Following more than two decades of research and developments made possible through various proof-of-concept hyperspectral remote sensing missions, it has been anticipated that hyperspectral imaging would enhance the accuracy of remotely sensed in-water products. This study investigates such expected improvements and demonstrates the utility of hyperspectral radiometric measurements for the retrieval of near-surface phytoplankton properties[1], i.e., phytoplankton absorption spectra (a_{ph}) and biomass evaluated through examining the concentration of chlorophyll-*a* (Chl*a*). Using hyperspectral data (409–800 nm at ~5 nm resolution) and a class of neural networks known as Mixture Density Networks (MDN) (Pahlevan et al., 2020), we show that the median error in a_{ph} retrievals is reduced two-to-three times ($N = 722$) compared to that from heritage ocean color algorithms. The median error associated with our a_{ph} retrieval across all the visible bands varies between 20 and 30%. Similarly, Chl*a* retrievals exhibit significant improvements (i.e., more than two times; $N = 1902$), with respect to existing algorithms that rely on select spectral bands. Using an independent matchup dataset acquired near-concurrently with the acquisition of the Hyperspectral Imager for the Coastal Ocean (HICO) images, the models are found to perform well, but at reduced levels due to uncertainties in the atmospheric correction. The mapped spatial distribution of Chl*a* maps and a_{ph} spectra for selected HICO swaths further solidify MDNs as promising machine-learning models that have the potential to generate highly accurate aquatic remote sensing products in inland and coastal waters. For a_{ph} retrieval to improve further, two immediate research avenues are recommended: a) the network architecture requires additional optimization to enable a simultaneous retrieval of multiple in-water parameters (e.g., a_{ph}, Chl*a*, absorption by colored dissolved organic matter), and b) the training dataset should be extended to enhance model generalizability. This feasibility analysis using MDNs provides strong evidence that high-quality, global hyperspectral data will open new pathways toward a better understanding of biodiversity in aquatic ecosystems.

1. Introduction

Phytoplankton are primary producers in aquatic environments, providing a food source for other organisms, such as zooplankton (Falkowski et al., 2003). Their distribution in space and time as well as their composition and community structure serve as indicators for measuring the biodiversity and health of aquatic ecosystems. Understanding phytoplankton dynamics elucidates how aquatic ecosystems respond to climate variability (e.g., extended periods of dry and/or wet seasons) or to landcover and/or landuse changes (Sinha et al., 2017). Of prime concern in inland (e.g., lakes, rivers, reservoirs, especially water supplies) and coastal (e.g., estuaries, bays) waters is the presence of harmful algal blooms (HABs) triggered typically by rising temperatures and enhanced nutrient levels (Anderson, 2004; Brooks et al., 2016).

* Corresponding author at: NASA Goddard Space Flight Center, Greenbelt, MD, USA.
E-mail address: nima.pahlevan@nasa.gov (N. Pahlevan).

[1] Codes are accessible via https://github.com/STREAM-RS/STREAM-RS

https://doi.org/10.1016/j.rse.2020.112200
Received 5 April 2020; Received in revised form 12 November 2020; Accepted 16 November 2020
Available online 27 November 2020

Quantifying phytoplankton biomass, their community composition and cell-size distribution is critical in documenting the onset, evolution, and dissipation of HABs (Schofield et al., 1999) and their associated public health and ecosystem impacts.

Aquatic remote sensing has long proven an efficient means for detecting HABs and addressing phytoplankton biodiversity (IOCCG, 2014). While HAB detection has primarily been based on examining band-arithmetic indices (Bresciani et al., 2011; Gower et al., 2008; Tomlinson et al., 2004), identifying phytoplankton community structure in the open ocean has relied upon phytoplankton biomass as inferred through near-surface concentrations of chlorophyll-*a* (Chl*a*), the primary pigment in all phytoplankton types and/or species (Alvain et al., 2005; Brewin et al., 2011; Uitz et al., 2010; Uitz et al., 2015). A complete characterization of phytoplankton biodiversity, however, requires high-fidelity retrievals of phytoplankton absorption coefficients (a_{ph}) (Brewin et al., 2011; Bricaud et al., 2007; Hoepffner and Sathyendranath, 1993; Uitz et al., 2015) that fall under a broader suite of products termed the Inherent Optical Properties (IOPs). For a full attribution of phytoplankton community structure via a_{ph}, numerous studies have suggested that a fairly fine spectral sampling is necessary to ensure subtle spectral features are captured (Roelke et al., 1999). Assuming remotely sensed observations over inland and coastal waters are accurately compensated for the atmospheric effects (Frouin et al., 2019), the remaining hurdle is to robustly retrieve $a_{ph}(400 < \lambda < 700$ nm) in areas where the absorption budget is dominated by non-algal particles (a_{nap}) and/or colored dissolved organic matter (a_{cdom}) (Bukata et al., 1995; Feng et al., 2005; IOCCG, 2000, 2006; Strömbeck and Pierson, 2001). Further, estimating Chl*a* from hyperspectral remote sensing across a wide array of trophic conditions (e.g., $0.1 <$ Chl*a* < 940 mg m^{-3}; (Eleveld et al., 2017)) remains a daunting task and no single functional algorithm has been offered to-date (Kutser, 2004; Neil et al., 2019; Spyrakos et al., 2018; Tilstone et al., 2017).

Recently, Pahlevan et al. (2020) demonstrated the utility of Mixture Density Networks (MDN) in the context of Chl*a* retrieval for Sentinel-2 and Sentinel-3 multispectral observations. The algorithm harnesses all available spectral bands and, through training, automatically learns most suitable band combinations for various water types. In this study, we extend this algorithm to the inversion problem of hyperspectral $a_{ph}(\lambda)$ (λ is dropped hereafter) and Chl*a* from hyperspectral data using two separate sets of *in situ* data (Sections 3 and 4). Following a rigorous evaluation of model performances (Section 5), the developed models are applied to atmospherically corrected images of the Hyperspectral Imager for Coastal Ocean (HICO) to assess image-derived a_{ph} spectra and Chl*a* maps in highly eutrophic and turbid waters. A list of critical notations and acronyms used throughout this research is provided in Table 1.

Table 1
List of critical notations, acronyms, and symbols.

Symbol	Description
R_{rs}	*In situ* remote sensing reflectance
a_{ph}	*In situ* phytoplankton absorption
a^*_{ph}	Phytoplankton specific absorption
$\widehat{R}_{rs}$	Retrieved (image-derived) R_{rs}
$\widehat{a}_{ph}$	Estimated a_{ph} from R_{rs}
$\widehat{a}^r_{ph}$	Retrieved (image-derived) a_{ph}
$\widehat{a}^*_{ph}$	Retrieved (image-derived) phytoplankton specific absorption
Chla^e	Estimated Chl*a* from R_{rs}
Chla^r	Retrieved (image-derived) Chl*a*
MDN-I	MDN applied to R_{rs}
MDN-S	MDN applied to $\widehat{R}_{rs}$

Note that *in situ* phytoplankton absorption measurements are *in vivo*.

2. Background

2.1. Phytoplankton absorption (a_{ph})

A thorough review of the existing IOP algorithms has been provided in Werdell et al. (2018). For completeness, a brief overview of fundamentals of bio-optical inversion methods is presented. In general, radiometric quantities like R_{rs}, defined as the ratio of water-leaving radiance to the total downwelling irradiance just above the water (Mobley, 1999), is assumed to be a function of specific IOPs (e.g., spectral shapes of absorption and backscattering, volume scattering function) and concentrations of water constituents like Chl*a*, Suspended Particulate Matter (SPM), and phycocyanin (PC), i.e., $R_{rs} = F$ [IOP, Chl*a*, PC, SPM] (Mobley, 1994). This function (F) is unique as each permutation of IOPs and concentrations yields a single R_{rs} spectrum. Given a satellite-derived R_{rs} spectrum, the goal in bio-optical retrieval schemes is to approximate an inverse function (F^{-1}), which is not unique as a single R_{rs} spectrum may correspond to varied combinations of IOPs and concentrations (Defoin-Platel and Chami, 2007; Sydor et al., 2004; Yang et al., 2011). Identifying a solution is more complex when spectral information is available only for a limited number of bands. Further, F^{-1} is commonly not an exact function, that is, approximations and constant factors limit the uniqueness of the solution.

Different schemes have been proposed for the solution of the inverse problem of IOPs or the associated constituent concentrations from R_{rs}. The solution methods fall into three categories: a) semi-analytical, b) look-up-table (LUT), and c) empirical and/or statistical approaches. The most robust inversion methods are the ones that attempt to derive IOPs using mechanistic approaches, where approximations associated with some of the unknown parameters exist, i.e., spectral variability of specific IOPs (e.g., (Roesler and Boss, 2003)). These techniques either solve for IOP components (like a_{ph}, a_{cdom}, a_{nap}) simultaneously through an optimization scheme (e.g., (Brando et al., 2012; Gege, 2014)) or follow a step-wise process (Lee et al., 2002) to sequentially estimate IOPs. The LUT-based methods apply a pre-generated LUT, where each R_{rs} spectrum is associated with a combination of IOP and constituent concentrations (as well as bottom substrate properties in optically shallow environments) (Gerace et al., 2013; Hedley et al., 2009; Mobley et al., 2005; Moses et al., 2012a; Pahlevan and Schott, 2013). The empirical and/or statistical approaches range from methods that relate a number of principle components of R_{rs} spectrum to IOPs through multiple linear regressions (Barnes et al., 2014; Craig et al., 2012; Härmä et al., 2001), to neural networks (NN) that are optimized given R_{rs} and IOPs (Doerffer and Schiller, 2007; Ioannou et al., 2011; Jamet et al., 2005; Schiller and Doerffer, 1999), and to methods that utilize search mechanisms to identify optimal fitting coefficients to arrive at a solution for inverse problems (Chami and Robilliard, 2002; Pitarch et al., 2014; Song et al., 2013; Zhan et al., 2003).

Yet, most of the studies have been devised to function with current multispectral remote sensing technologies and are constrained to the number of unknown parameters for which they attempt to solve. Further, the widely used inversion methods apply various degrees of empiricism applicable to open ocean waters, limiting their utility in bio-geo-optically and spatio-temporally diverse coastal and inland waters (Mishra et al., 2013a). The global applicability of empirical approaches or NNs may be even more constrained for regions where IOPs and concentrations employed in the training are not adequately representative of conditions under evaluation. We develop an MDN model (Section 4) that largely surmounts this ill-posed problem. In essence, MDNs learn the likelihoods associated with different solutions from which an optimal estimation is computed.

2.2. Chlorophyll-a

For years, studies have proven that Chl*a* in optically complex waters can be estimated with various degrees of precision and/or accuracy from

R_{rs} (Gitelson, 1992; O'Reilly et al., 1998). These studies, to a large extent, have been based on available multispectral images (mostly acquired by the Medium Resolution Imaging Spectrometer; MERIS) and have often used a two- or three-band index to estimate Chl*a* (Gilerson et al., 2010; Moses et al., 2012b). With the anticipation of enhancements in the spectral capability of future global aquatic remote sensing missions (e.g., Plankton, Aerosol, Cloud, and ocean Ecosystem; PACE (Werdell et al., 2019)), it is essential to devise novel approaches to predict water quality (WQ) parameters, such as Chl*a*.

A few review studies have conjectured the expected improvements in retrievals of WQ parameters and other bio-optical properties (Devred et al., 2013; Dierssen et al., 2020; Giardino et al., 2019; Hestir et al., 2015). For instance, Giardino et al. (2019) speculated that Chl*a* could be better estimated if contiguous bands were present allowing a switching between different bands to capture the relative maximum and/or minimum of a reflectance spectrum. Hyperspectral data have enabled studies focusing on Chl*a* estimations as well as multiple optically significant constituents in both optically deep and optically shallow waters (Albert and Gege, 2006; Lee et al., 2007). A bulk of previous studies were dedicated to estimating Chl*a* or SPM using select spectral bands most sensitive to either parameter (e.g., Hakvoort et al. (2002); Kallio et al. (2001)). Other studies that demonstrate the full potential of hyperspectral imagery applying all available spectral bands are commonly based on spectral matching techniques, i.e., curve fitting, or matrix inversion methods (MIM) (Brando and Dekker, 2003; Hoogenboom et al., 1998; Mobley et al., 2005). Both approaches require assumptions on, or some knowledge of, the spectral slopes and/or magnitudes of specific IOPs. Using MIM, Brando and Dekker (2003) mapped Chl*a*, a_{cdom}, and SPM from a Hyperion scene of Deception Bay, Australia. Fichot et al. (2015) applied a partial least square regression to hyperspectral images data to retrieve Chl*a*, turbidity, and dissolved organic carbon in the San Francisco Bay-Delta Estuary. Other studies have implemented sensitivity analyses to identify optimal bands for Chl*a* retrievals (Sun et al., 2009). In this study, we will train an MDN model with high-fidelity *in situ* R_{rs} data, allowing the model to learn the spectral regions that contribute most to Chl*a* estimations.

3. Datasets

3.1. Development data

For a_{ph} retrievals, we used the SeaWiFS Bio-optical Archive and Storage System (SeaBASS) dataset in addition to data collected in several inland water bodies ($N = 1444$) shown in Fig. 1. The majority of the radiometric data were collected using the free-fall approach (Mueller et al., 2003). HyperOCR® radiometers were commonly used to collect data via the free-fall profiling and/or sky-block techniques (Lee et al., 2013). A small fraction of the data had been collected using measurements of the water surface, sky, and plaque radiances by handheld spectrometry (e.g., ASD FieldSpec®) following the ocean optics protocol (Mueller et al., 2003). The discarded data amounted to only a small subset (<1%) of the initial database, because the data had undergone preliminary quality screenings by the data providers. A data-quality control was further carried out to exclude outliers identified through visual inspections of R_{rs} data exhibiting abnormal spectral features, highly inconsistent with known spectral properties of constituents measured in the water.

A large portion of the data came from highly eutrophic or turbid waters of the Chesapeake Bay (Zheng et al., 2015), Mississippi aquaculture ponds (Mishra et al., 2013a; Mishra et al., 2013b; Wang et al., 2016), Lake Erie (Binding et al., 2019), Florida shelves (Cannizzaro et al., 2008; Soto et al., 2015), and small lakes in Wisconsin, Nebraska (Gurlin et al., 2011), and Indiana (Li et al., 2015; Tan et al., 2015) (Fig. 1). Moreover, data from three lakes in Asia, including Lake Taihu ($N = 45$) in China (Cao et al., 2020; Jiang et al., 2020; Xue et al., 2019), Lake Kasumigaura ($N = 26$), and Lake Suwa ($N = 8$) in Japan (Yang et al., 2012), were added to enhance the generalizability of the model.

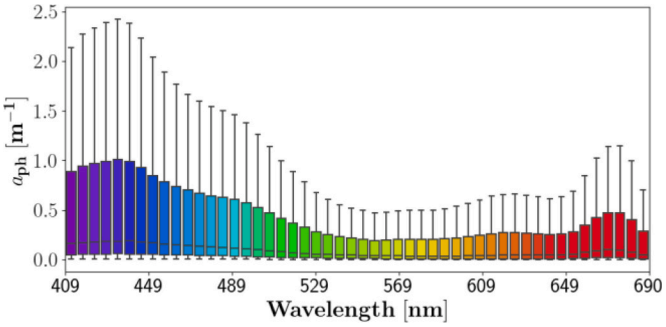

Fig. 2. Spectral distribution of a_{ph} data ($N = 1444$). The whiskers refer to interquartile ranges with the bars denoting the minimum and maximum for each HICO band center.

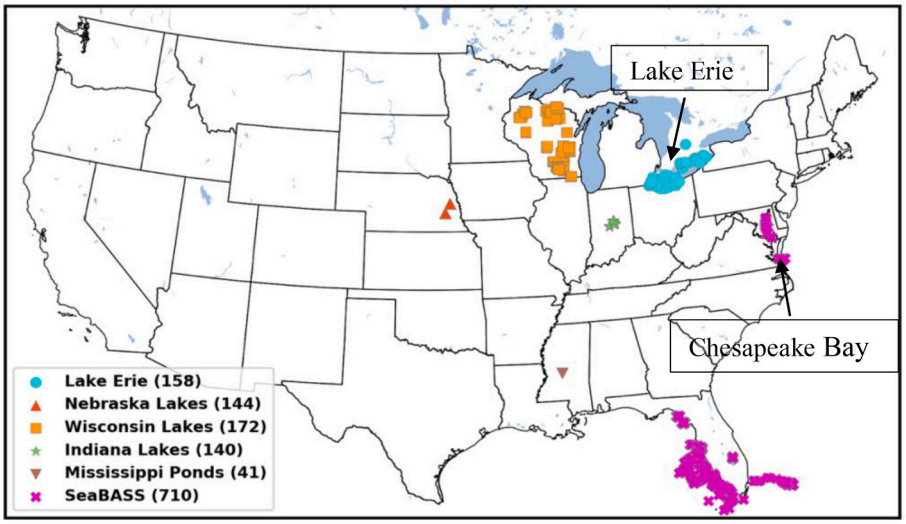

Fig. 1. Spatial distribution of the $R_{rs} - a_{ph}$ dataset ($N = 1444$) used for training and validating the MDN model. The total number of samples available in each region is provided in round brackets. Additional data from Lake Taihu ($N = 45$) (China) and Lakes Kasumigaura ($N = 26$) and Suwa ($N = 8$) (Japan) are not shown on the map.

The spectral distribution of a_{ph} data are illustrated in Fig. 2. While the quantitative filter pad technique was applied to measure a_{ph} in all research experiments (Roesler, 1998), we refer readers to the above-noted references for more details. The R_{rs} and a_{ph} were provided at various spectral resolutions (between 1 and 3.3 nm) and mostly available for the 400–800 and 400–690 nm ranges, respectively.

An augmented database of paired R_{rs} – Chl*a* data ($N = 3804$) covering a remarkably wide range of trophic conditions was assembled (Pahlevan et al., 2020). Compared to our R_{rs} – a_{ph} dataset, this database provides a better representation of global Chl*a* conditions (see Fig. 1 in (Pahlevan et al., 2020)). The recent update to the R_{rs} – Chl*a* dataset was made by adding 875 *in situ* samples from lakes in Italy (Bresciani et al., 2020; Warren et al., 2019) and Brazil (Cairo et al., 2020; da Silva et al., 2020; Maciel et al., 2019), as well as from western Lake Erie (Moore et al., 2017), Indiana lakes (Song et al., 2013), and SeaBASS.

The hyperspectral *in situ* data were resampled with the relative spectral responses of HICO, i.e., full-width-half-maximum of ~ 5.7 *nm* (Lucke et al., 2011), to simulate band-averaged R_{rs} spectra for model training and testing throughout this research (Section 4).

3.2. Image data

Images acquired by HICO with a nominal 100 m spatial resolution were utilized to demonstrate the reliability of our algorithm in the presence of uncertainties in atmospheric correction (AC) and instrument noise (Ibrahim et al., 2018). HICO data have been used in several coastal and inland water studies (Keith et al., 2014; Mishra et al., 2014; Moses et al., 2013; Ryan et al., 2014; Tufillaro and Davis, 2012). Most studies, however, applied existing algorithms that make use of select bands (Lee et al., 2002; O'Reilly et al., 1998) to obtain WQ and/or IOP maps. Here, we apply MDN models to atmospherically corrected HICO images and perform a matchup analysis using an independent dataset (Section 3.3). Furthermore, four HICO images over Lake Erie and the Chesapeake Bay are examined to demonstrate consistency in the spatial contexts of our generated products (Section 5.3).

The HICO images were atmospherically corrected using the SeaWiFS Data Analysis System (SeaDAS) (Ibrahim et al., 2018), which, by default, retrieves R_{rs} (hereafter $\widehat{R}_{rs}$; Table 1) at 65 spectral bands of HICO within the 358–719 nm range. Of these bands, due to increased uncertainties in the ultra-violet region, we mainly focused on spectral bands >400 nm leaving a total of 54 spectral bands for algorithm testing and/or evaluations. To process the images, all the SeaDAS default processing options (e.g., the 747–787 nm band pair for aerosol removal, water vapor correction using the 705, 725, 745 nm band combination, O_2 correction) were adopted.

Due to substantial misregistration errors associated with the image products (Garcia et al., 2014; Keith et al., 2014), an automatic registration pipeline was developed to refine the geolocation accuracy of ~150 HICO $\widehat{R}_{rs}$ products evaluated throughout this study. First, HICO swaths were projected onto a rectangular coordinate grid using their supplied geographic geolocation. The land-water boundaries were then estimated using the normalized difference vegetation index, and a Gaussian averaging filter was subsequently applied to ensure minimal artifacts are present. The resulting swaths, along with a permanent global water cover dataset (Buchhorn et al., 2020), were then used to estimate a homography matrix (Alcantarilla and Solutions, 2011; Fischler and Bolles, 1981); thereby determining a projection from the geolocated water cover dataset into HICO image coordinates. This projection was used to seed a second round of homography estimation, using the original (unfiltered) HICO swaths. Finally, this transformation was applied to the water cover geolocation grid, projecting it onto the HICO swaths and resulting in per-pixel geographic geolocation. Experimentation with accurately geolocated images (e.g. Landsat-8 scenes) suggested that the described registration process is able to obtain sub-pixel registration accuracy (i.e. < 100 m for HICO products).

3.3. Matchup data

To assess the performance of MDNs in practical applications, we identified near-coincident *in situ* measurements made within ±3 hr of HICO image acquisitions. Note that this dataset is independent of the development data described in Section 3.1. Overall, 65 R_{rs} matchups, 105 Chl*a* matchups, and 29 a_{ph} matchups over the western basin of Lake Erie, lower Chesapeake Bay, and Florida (FL) estuaries, including Pensacola, St. Andrew, and Choctawhatchee Bays (Casey et al., 2019; Keith et al., 2014; Schaeffer et al., 2015), were located. Of the 65 R_{rs} samples, 38 and 40 samples were accompanied with Chl*a* and a_{ph} measurements, respectively, providing an opportunity to evaluate the retrievals from both R_{rs} and $\widehat{R}_{rs}$ (Tables 4 and 5). This subset – all of which were acquired in FL estuaries (Keith et al., 2014) – enabled a better understanding of how uncertainties in AC affect algorithm performances. Our matchup datasets were collected by the Great Lakes Environmental Research, Environment and Climate Change Canada, Maryland Department of Natural Resources (MD DNR), Chesapeake Bay Program (CBP), and U.S. Environment Protection Agency through routine monitoring exercises or funded research projects. While all the a_{ph} matchups were available through multi-year field campaigns in FL estuaries (Keith et al., 2014), ~ 50% of Chl*a* matchups originated in the lower Chesapeake Bay area through CBP (https://www.chesapeakebay.net) and MD DNR (http://eyesonthebay.dnr.maryland.gov). Considering R_{rs} matchups, only six samples were from the western Lake Erie, and the rest were acquired in FL estuaries. A summary of descriptive statistics attributed to these matchups is provided in Table 2. Section 3.1

4. Approach

A full description of MDNs designed for Chl*a* retrievals from the MultiSpectral Instrument (MSI) and Ocean and Land Color Instrument (OLCI), as well as for approximating particulate backscattering (b_{bp}) is given in (Pahlevan et al., 2020) and (Balasubramanian et al., 2020). Nonetheless, we will provide a concise overview of the two MDN models for estimating a_{ph} and Chl*a*.

4.1. Mixture density networks

Conventional NNs directly predict target variables by approximating the conditional average of target data. However, in approximating continuous target variables (e.g., a_{ph} or Chl*a*), the conditional average is a limited representation of the statistical properties of the target space, and often fails to produce practical solutions (Bishop, 1994). This problem frequently arises when the problem at hand is a multi-valued mapping (e.g., IOP inversion) (Sydor et al., 2004; Yang et al., 2011). MDNs surmount this limitation by modeling conditional probabilities of the target variables given input data to obtain a more complete picture of the probability distribution of the data in the target space. MDNs produce three statistical measures of a target variable (e.g., a_{ph}), namely a mean vector (μ), a covariance matrix (Cov), and a mixing coefficient (α) for each probability density function modeled as Gaussian functions.

Table 2
Statistical attributes of HICO matchups used to assess the performance of the atmospheric correction and MDN models. Of the 65 Rrs matchups, 38 and 40 samples had an associated Chl*a* and a_{ph}, respectively. Of those 40 paired R_{rs} – a_{ph}, 29 were acquired near-coincident with HICO overpasses. The full spectral coverage for both R_{rs} and a_{ph} data was 400–735 nm. STD stands for one standard deviation.

	Median	Mean	Min	Max	STD	N
R_{rs} (443)	0.0019	0.0026	0.0000	0.0091	0.0019	65
R_{rs} (670)	0.0025	0.0030	0.00089	0.0119	0.0019	
a_{ph}(443)	0.2694	0.2780	0.0972	0.5780	0.0980	40
a_{ph}(670)	0.0681	0.0796	0.0314	0.1571	0.0310	
Chl*a*	6.0	13.93	0.249	173.0	23.237	105

The learned probability density via MDNs is as follows

$$p\left(\widehat{a}_{\mathrm{ph}} \mid R_{\mathrm{rs}}\right) = \sum_{i=1}^{c} \alpha_i(R_{\mathrm{rs}})\, \phi_i\left(a_{\mathrm{ph}} \mid R_{\mathrm{rs}}\right) \quad (1)$$

in which $\widehat{a}_{\mathrm{ph}}$ is the estimated a_{ph} and ϕ_i is a Gaussian distribution with μ and Cov as the mean vector and the covariance matrix, respectively. Fig. 3 illustrates a block diagram of an MDN model designed for $\widehat{a}_{\mathrm{ph}}$ retrievals. Once trained, the MDN predicts the conditional probability density of $\widehat{a}_{\mathrm{ph}}$ (or estimated Chl*a*; Chla^{e}) for each input R_{rs}. Having obtained a proper representation of the conditional density of $\widehat{a}_{\mathrm{ph}}$ via its inter-band correlations (Cov), it is straightforward to compute any desired statistical parameters. The Gaussians are combined to form the final output estimation via maximum likelihood, which represents the estimate in the area of highest probability mass:

$$\widehat{a}_{\mathrm{ph}} = \mu_i(R_{\mathrm{rs}}) : i = argmax\ \alpha(R_{\mathrm{rs}}) \quad (2)$$

In this study, two groups of MDN models, MDN-I (*in situ*) and MDN-S (satellite), were developed. The MDN-I models are the ones trained with 722 and 1902 randomly chosen pairs of R_{rs} – a_{ph} and R_{rs} – Chl*a* (Section 3.1), respectively, leaving out the rest half of the data (722 and 1902) for testing. Following several experiments on the size of the training subset, we found that increasing the proportion of training data versus the testing data introduces insignificant improvements in the model performance and, thus approximately only one-half of the development data were used to estimate $\widehat{a}_{\mathrm{ph}}$ and Chla^{e}. Input to MDN-I was composed of 50- (409–690 nm) and 67-element R_{rs} vectors (409–787 nm) for modeling $\widehat{a}_{\mathrm{ph}}$ and Chla^{e}, respectively. The former choice was made to include a subset of the a_{ph} dataset ($N = 140$) available only for this spectral range, discarding the 690–700 nm range.

Following the validation of MDN-I models (Section 5.1), MDN-S models were trained with the entire development data for applications to HICO $\widehat{R}_{\mathrm{rs}}$ maps (Sections 5.2 and 5.3). In this case, however, we removed 9 and 14 samples from our paired R_{rs} – a_{ph} and R_{rs}–Chl*a* datasets that originated from the western Lake Erie, leading to 1435 and 3785 records. This was carried out to ensure matchup data (Section 3.3) are excluded in our total training set. For HICO-derived $\widehat{a}_{\mathrm{ph}}$ products, i.e., $\widehat{a}^{\mathrm{r}}_{\mathrm{ph}}$, input features to MDN-S consisted of a 50-element $\widehat{R}_{\mathrm{rs}}$ vector. Note that if $\widehat{R}_{\mathrm{rs}}$ were invalid in any bands, no $\widehat{a}^{\mathrm{r}}_{\mathrm{ph}}$ retrievals would be attempted. For Chla^{e} retrieval, to maximize the number of valid retrievals (due to frequent invalid retrievals in the blue bands), we trained an MDN-S model that accepts input spectra ranging only from 501 to 713 nm, i.e., 38 bands.

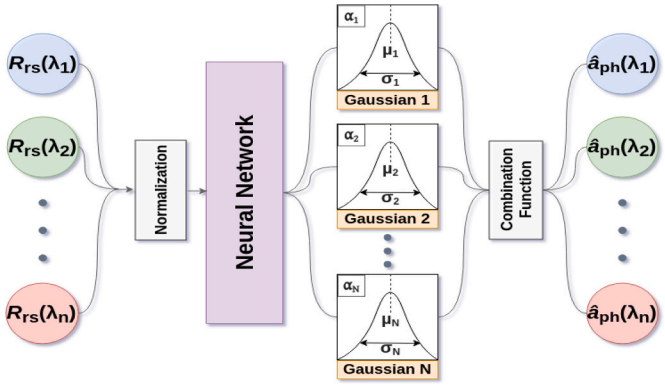

Fig. 3. Block diagram illustrating the primary elements of the Mixture Density Network (MDN) designed in this study for $\widehat{a}_{\mathrm{ph}}(\lambda)$ retrieval. Three probability density functions with their associated parameters (μ, σ) and mixing coefficients (α) are defined to arrive at a best solution.

4.2. Benchmarking

To analyze the performance of MDNs, select heritage IOP and Chl*a* algorithms were used for comparisons against our algorithms. Given their proven performance, the Quasi Analytical Algorithm (QAA) (Lee et al., 2002) and Generalized IOP inversion (GIOP) (Werdell et al., 2013) were adopted as benchmark algorithms. Although these two models are primarily intended for moderately turbid coastal and clear ocean waters, due to their extensive use in aquatic studies and straightforward implementation (IOCCG, 2006), we considered these algorithms as reference models for the wide range of trophic conditions of inland and coastal waters (Pahlevan et al., 2020).

For Chl*a*, we implemented a total of seven state-of-the-art algorithms (Hu et al., 2012; Mishra and Mishra, 2012; Moses et al., 2009; Moses et al., 2012b; O'Reilly et al., 1998; O'Reilly and Werdell, 2019; Smith et al., 2018). The algorithms examined were OC6 (O'Reilly and Werdell, 2019), Blend (Smith et al., 2018), 3-Band (Moses et al., 2012b), and the Normalized Difference Chl*a* Index (NDCI) (Mishra and Mishra, 2012) (see Appendix in (Pahlevan et al., 2020)). We further evaluated the Gons (Gons et al., 2002) and Gilerson 2-Band (GI2B) algorithms (Gilerson et al., 2010). For conciseness, the demonstration is provided for the five best-performing algorithms (Section 5.1). It should be noted that the NDCI, Gons, and GI2B models were originally calibrated and tested with Chl*a* within the range of 1–60, 1–185, 2–100 [mg m^{-3}], respectively, and no attempt was made to re-calibrate them in this study.

4.3. Performance metrics

Two metrics explaining error and bias in retrievals proposed in Morley et al. (2018) were adopted. These measures are calculated as below

$$\epsilon = 100 \times \left(10^{Y} - 1\right)\ [\%] \text{ where } Y = Median\ |log_{10}(\widehat{q}/q)| \quad (3)$$

$$\beta = 100 \times sign(\mathrm{z})\ \left(10^{|Z|} - 1\right)\ [\%] \text{ where } Z = Median\ \left(log_{10}(\widehat{q}/q)\right) \quad (4)$$

where $\widehat{q}$ and q are estimated and *in situ* quantities, respectively, Median is the median operator, ϵ represents the median symmetric accuracy (MdSA), and β is the symmetric signed percentage bias. These metrics are simple for interpretation, fairly robust against outliers, and zero-centered compared to those in Seegers et al. (2018). In addition to the above metrics, we will also report the slope of linear regression (*S*) to provide further insights into the goodness of retrievals.

5. Result

5.1. Performance assessment

The performance of MDN-I against that of QAA and GIOP assessed at 443 and 530 nm is illustrated in Fig. 4. Evidently, the generic models perform reasonably well in the blue bands in oligotrophic waters (i.e., $a_{\mathrm{ph}} < 0.1$ [m^{-1}] (Roesler and Perry, 1995)), with $\widehat{a}_{\mathrm{ph}}$ tending to diverge from the 1:1 line with an increase in absorption. In particular, QAA with ϵ of ~48% outperforms GIOP at the 443 nm band and performs satisfactorily within the 0.1–1.0 [m^{-1}] range. For $\widehat{a}_{\mathrm{ph}}(530)$, QAA tends to overestimate whereas GIOP exhibits 10–20% larger errors than that for the 443 nm band, even though it consistently provides fair predictions for < 0.01 [m^{-1}]. The performance of both models degrades significantly in the red portion of the spectrum (not shown here). The data distributions suggest that GIOP $\widehat{a}_{\mathrm{ph}}$ are spectrally correlated (Werdell et al., 2013), verifying potential improvements are possible if more representative initial phytoplankton specific absorption (a^{*}_{ph}) were available. On the contrary, worsening QAA retrievals in the green (and red) corroborates that the model assumptions (Lee et al., 1998) do not hold in eutrophic waters; although Lee and Carder (2004) surmised that

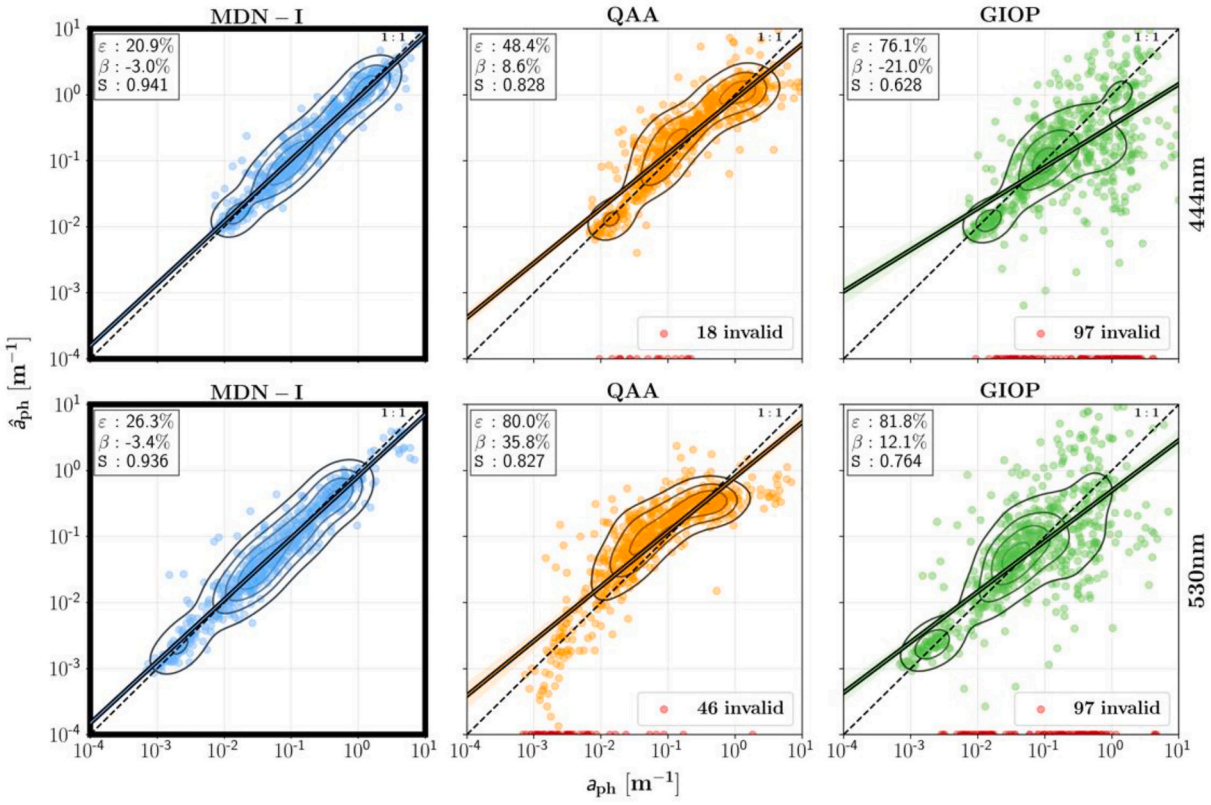

Fig. 4. Performance of MDN-I ($N = 722$) at two spectral bands compared to that of QAA and GIOP commonly used for $\hat{a}_{ph}$ retrievals in fairly clear ocean waters. The MDN model returns nearly zero biases and relatively small overall errors. Included metrics are error (ϵ), bias (β), and slopes of linear regression (S).

an increase in the pure water absorption hampers $\hat{a}_{ph}$ inversion in the red region.

From Fig. 4, it is inferred that MDN-I offers significantly more accurate and precise $\hat{a}_{ph}$ for the two visible bands with small biases and fairly low overall errors. Fig. 5 further illustrates a band-by-band performance assessment of MDN-I expressed through ϵ and β. This assessment suggests that errors do not fluctuate drastically, i.e., from 20 to 30%, across the 409–690 nm range, and there is a minimal bias in $\hat{a}_{ph}$ throughout the spectrum. Of particular interest for freshwater HAB studies (i.e., cyanobacteria) is the retrieval accuracy at the characteristic 620 nm peak, which is found to be within the 20 to 25% range. The apparent elevated uncertainties in the 520–600 nm region are found to

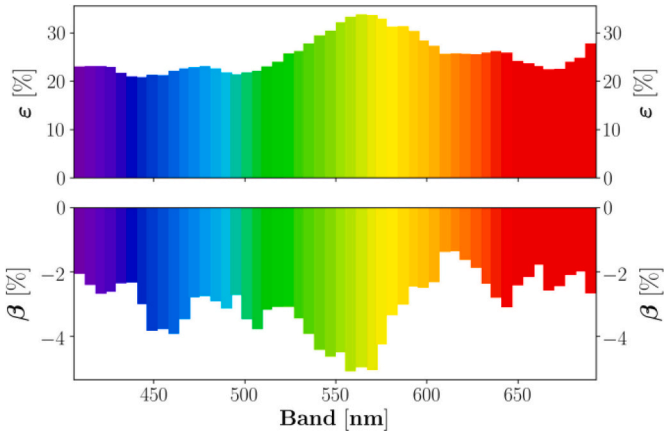

Fig. 5. Performance assessment (Eqs. (3) and (4)) of MDN-I ($N = 722$) for $\hat{a}_{ph}$ inversion across HICO visible bands. The larger errors are due to the lower magnitude of the signal within the 500–620 nm range.

be related to the generally low magnitude of a_{ph} (Fig. 2) or higher variability in the normalized R_{rs} in the green region, suggesting that the error metrics used are somewhat sensitive to the signal intensity.

The performance of MDN-I for Chla^e is evaluated against that of other widely used methods in Fig. 6. The MDN model remarkably outperforms other approaches across a wide range of Chl$a$ values, from moderately eutrophic coastal waters to hypereutrophic lake waters. This is, however, anticipated as MDN utilizes the entire spectral content of a given R_{rs} spectrum, whereas band-ratio models make use of select bands correlated with the variability in phytoplankton biomass (Clarke et al., 1970; Gitelson et al., 2007; Gordon et al., 1980; Mittenzwey et al., 1992). In particular, the Blend and GI2B models are found to perform fairly well in the range of 10 < Chl$a$ < 80 [mg m^{-3}] but otherwise return unreliable Chla^e. Similarly, the Gons model provides fairly robust retrievals within the 10 to 40 [mg m^{-3}] range.

5.2. HICO matchup analysis

A total of 12 paired $R_{rs} - \hat{R}_{rs}$ spectra are depicted in Fig. 7, where station identifications (IDs), *in situ* Chl$a$, and that estimated from both R_{rs} (Chla^e) and $\hat{R}_{rs}$ (Chla^r) are also denoted. Despite noticeable errors in the blue bands, major improvements in $\hat{R}_{rs}$ products through SeaDAS are achieved with respect to those reported in previous studies (Keith et al., 2014; Mishra et al., 2014). Table 3 includes performance statistics for five selected visible bands. By comparing the three metrics, one may infer that there is mostly a band-dependent underestimation in $\hat{R}_{rs}$. Note however that while instrument artifacts are noticeable in some of the $\hat{R}_{rs}$ bands (e.g., the peak around 416 and 490 nm at PB05), potential effects of land adjacency, in particular in fairly small FL estuaries (Keith et al., 2014), have not been accounted for. From the reported Chla^e in Fig. 7, it can be realized that the MDN model is sensitive to uncertainties in $\hat{R}_{rs}$ – more details are provided in Table 5. One should also note that the R_{rs}

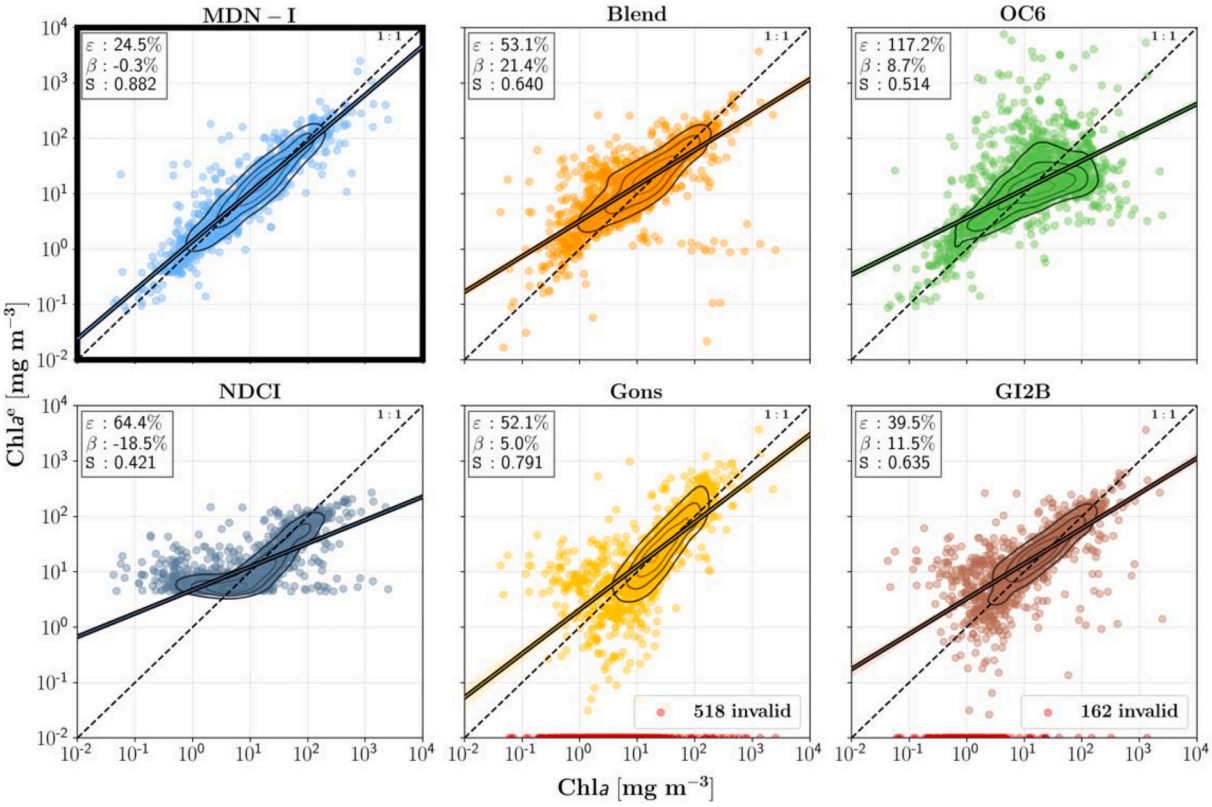

Fig. 6. Chla^e (estimated Chl$a$) from MDN-I (N = 1902) trained with HICO-simulated R_{rs} bands (409–787 nm) (Section 4.1) evaluated against the state-of-the-art algorithms (Gilerson et al., 2010; Gons et al., 2002; Mishra and Mishra, 2012; O'Reilly and Werdell, 2019). Included metrics are error (ϵ), bias (β), and slopes of linear regression (S).

measurements may be contaminated with skyglint effects (Kutser et al., 2013) resulting in increased estimated errors in $\widehat{R}_{rs}$ and/or in Chla^e.

The performance of MDNs for $\widehat{a}_{ph}$ at eight stations in FL estuaries is illustrated in Fig. 8. The error and bias metrics (Table 4) indicate degraded performances of MDNs for both $\widehat{a}_{ph}$ and $\widehat{a}^r_{ph}$ with respect to the statistics reported in Figs. 4 and 5, warranting the need for a further generalization of MDN through an extended training dataset. A further examination of $\widehat{a}_{ph}$ and $\widehat{a}^r_{ph}$ underscores the importance of highly accurate $\widehat{R}_{rs}$ products to allow for high-fidelity $\widehat{a}^r_{ph}$ retrievals (e.g., ϵ are two-to-four times larger for $\widehat{a}^r_{ph}$). The performance assessment of MDNs and the benchmark algorithms for Chl$a$ retrievals is reported in Table 5. Despite the strong evidence favoring MDN to return most accurate Chla^e (< 20% errors), its performance degrades more than three times for Chla^r. This loss in performance is pronounced for all the algorithms except for NDCI, which, surprisingly, yields better overall estimates, suggesting its lower sensitivity to uncertainties in AC. Regarding the NDCI performance, one should, however, note that a) the associated errors are >40%, which may render its utility for scientific studies limited, b) the linear regression slopes are fairly low (similar to the one in Fig. 6), and c) this assessment may not have global implications (N = 105), as opposed to the analysis presented in Fig. 6.

5.3. Demonstration: map products

Knowing the uncertainties in the atmospherically corrected products (Table 3) and the resultant uncertainties in Chla^r and $\widehat{a}^r_{ph}$, the MDN-S models were applied to $\widehat{R}_{rs}$ maps to demonstrate the algorithm performances for practical considerations. Fig. 9 illustrates two Chla^r maps over the western basin of Lake Erie. The spatial distribution of Chla^r on Sep 3rd 2011 in Lake Erie conforms to the historic knowledge of cyanobacteria distribution that tends to accumulate and intensify in the western edge of the lake (Binding et al., 2019). The bloom is found less severe across this section of the lake on Sep 8th 2014. The frontal feature separating the Detroit River plume from the rest of the lake (Station B3) is also consistent with prior observations (Chaffin et al., 2011; Moore et al., 2017). The $\widehat{R}_{rs}$ spectra associated with arbitrarily chosen stations denoted in Fig. 9A and B are illustrated in Fig. 10 where the corresponding $\widehat{a}^r_{ph}$ and $\widehat{a}^*_{ph}$(= $\widehat{a}^r_{ph}$/Chla^e), are also presented. The spectra extracted at Stations A1-2 and B2 clearly exhibit areas impacted by high concentrations of Chl$a$ and phytoplankton absorption spectra characteristic of cyanobacteria (Matthews and Bernard, 2013; Zhang et al., 2012).

The broad peaks in $\widehat{a}^r_{ph}$ around the 620 nm band, for example, correspond well with the depression bands in the associated $\widehat{R}_{rs}$. Further, the spectra extracted at Station B2 – with Chla^r = 160 [mg m^{-3}] – in the Sandusky Bay known to commonly host the cyanobacterium *Planktothrix sp.* (Davis et al., 2015) exhibit the expected spectral variability. The very low, and fairly "flat", $\widehat{a}^*_{ph}$ at this station implies phytoplankton packaging effect and presence of picoplankton as conjectured in Binding et al. (2019). The $\widehat{a}^r_{ph}$ spectra at Stations A1-2 influenced by the blooms from the Maumee River are similar to that of B2 with slight shifts in the depression areas around the 550 nm region. The $\widehat{a}^r_{ph}$ spectra at Stations A3-4 and B3 show distinct shapes similar to those typically observed in oligotrophic lakes (Effler et al., 1998) and closely match the spectra reported in Moore et al. (2017). The spectra extracted from the near-shore pixels (B4) (Chla^r= 18.3 [mg m^{-3}]) also appear to capture the spectral signature around ~620 nm, suggesting high concentrations of cyanobacteria that advected along the southwest shorelines of the lake (Berry et al., 2017).

A Chla^r map and locations of sample spectra extracted for four

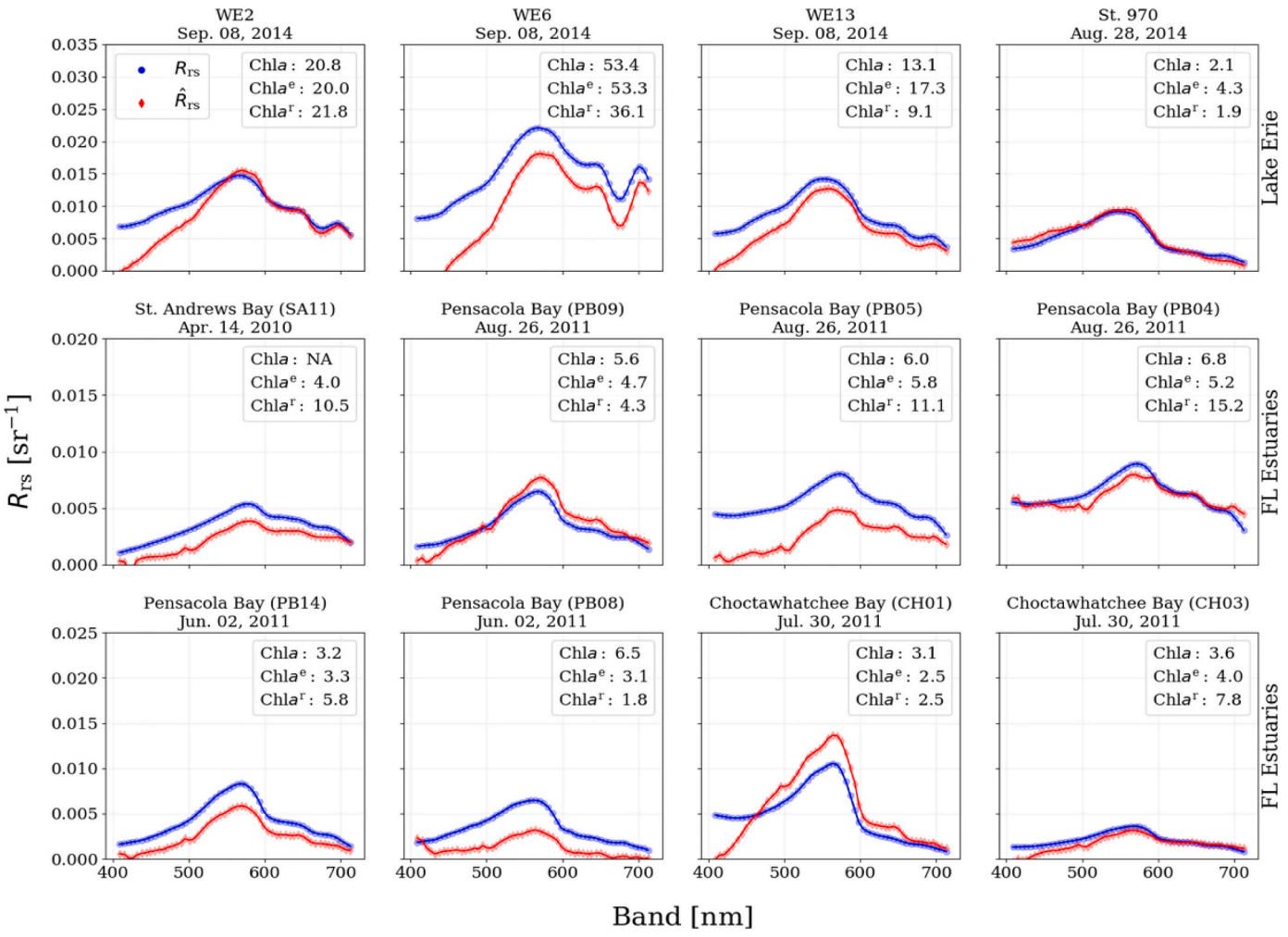

Fig. 7. Selected R_{rs}– $\hat{R}_{rs}$ (HICO) matchups in Lake Erie and FL estuaries. *In situ* measured Chl*a* and Chl*a* derived from both R_{rs} (Chla^e) and $\hat{R}_{rs}$ (Chla^r) derived via MDN-I and MDN-S, respectively, are denoted. Station IDs are also included for future references (Keith et al., 2014). Full performance statistics for $N = 65$ matchups are provided in Table 3. Note that R_{rs} measurements may include uncertainties due to skyglint contaminations. The assocaited a_{ph} spectra for the second and third rows are illustrated in Fig. 8.

Table 3
Performance statistics of the atmospheric correction obtained through near-coincident R_{rs} matchups acquired in FL estuaries. S stands for the slope of linear regression.

Band [nm]	ϵ [%]	β [%]	S	N
409	84.7	6.0	0.66	65
444	131.3	−118.6	1.08	
558	37.5	−36.0	1.14	
621	35.7	−29.4	1.01	
667	36.7	−27.4	0.98	

locations across a HICO image of April 1st 2010 over the lower Chesapeake Bay area is illustrated in Fig. 11A. The image captures spring blooms along the tributaries of the lower bay (Harding et al., 1994; Marshall and Egerton, 2010; Marshall and Nesius, 1996). The bloom at the mouth of the James River (A1) is manifested in the extracted spectra (Fig. 12) with an overall high magnitude of $\hat{a}^r_{ph}$ and a major trough around 675 nm. While $\hat{R}_{rs}$ at Stations A2 and A4 are, in general, very similar in shape and magnitude, the magnitudes of $\hat{a}^r_{ph}$ and the shapes of $\hat{a}^*_{ph}$ imply different concentrations and groups of phytoplankton. These differences in $\hat{a}^r_{ph}$ can be corroborated by examining Chla^r, i.e., 4.4 [mg m^{-3}] at Station A2 versus 8.9 [mg m^{-3}] at Station A4. Further, the differences in the magnitude and shapes of $\hat{R}_{rs}$ within the <450 nm interval for these stations reflect in the associated $\hat{a}^*_{ph}$. The spectra at Station 3 are typical of moderately eutrophic coastal waters with Chla^r of 1.6 [mg m^{-3}] and low $\hat{a}^r_{ph}$. The $\hat{a}^*_{ph}$ is also similar to the average a_{ph}* reported in Babin et al. (2003).

The $\hat{R}_{rs}$ and $\hat{a}^r_{ph}$ in Fig. 11B correspond to Chla^r ranging from 0.77 to 21.2 [mg m^{-3}]. Specifically, $\hat{a}^*_{ph}$ associated with Station B1 (Rappahannock River) evidently shows a peak around 620 nm, suggesting a considerable presence of cyanobacteria (Magnuson et al., 2004; Marshall, 1994). This peak is less pronounced at Station B2 situated in the main stem of the bay and disappears for Station 4, whose spectrum exhibits the largest ratio of $\hat{a}^r_{ph}(444)/\hat{a}^r_{ph}(670)$, typically used to determine phytoplankton size structure (Babin et al., 2003; Bricaud et al., 1995). The cyanobacteria-induced dips are also found to proportionally correspond with $\hat{R}_{rs}$ at Stations B1 through B3. The variability in accessory pigment concentrations and compositions may also be inferred through examining $\hat{a}^*_{ph}$ spectra (Marshall and Nesius, 1996). The spectrum at Station 2, for instance, displays a peak ~560 nm related to phycoerythrin (Ray et al., 1989).

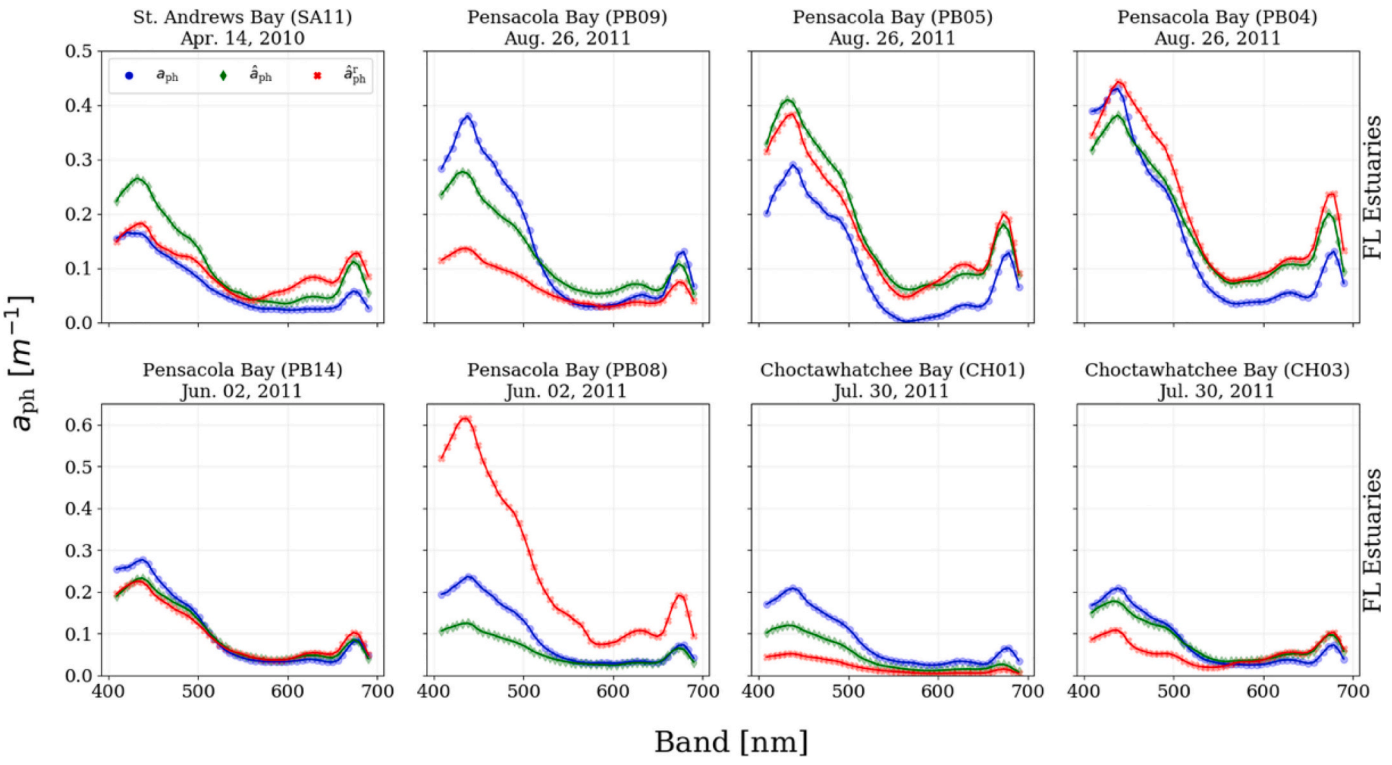

Fig. 8. *In situ* a_{ph} and MDN-derived $\hat{a}_{ph}$ and $\hat{a}^{r}_{ph}$, obtained by incorporating R_{rs} and $\widehat{R}_{rs}$, respectively, for selected stations in FL estuaries. More detailed performance analyses for $N = 40$ and $N = 29$ *in situ* sets and matchups are provided in Table 4. The corresponding R_{rs} and $\widehat{R}_{rs}$ spectra are illustrated in Fig. 7.

Table 4
Performance statistics of MDNs for a_{ph} retrievals from R_{rs} ($\hat{a}_{ph}$) and $\widehat{R}_{rs}$ ($\hat{a}^{r}_{ph}$) computed via *in situ* samples and matchups in FL estuaries. QAA and GIOP estimates were highly erroneous (e.g., band average errors >200% for QAA supplied with R_{rs}) and are excluded.

	$\hat{a}_{ph}$				$\hat{a}^{r}_{ph}$			
Band [nm]	ϵ [%]	β [%]	S	N	ϵ [%]	β [%]	S	N
409	37.5	−24.5	0.66	40	96.4	−76.4	−0.38	29
444	36.6	−27.6	0.78		108.6	−90.5	−0.28	
558	38.6	31.4	0.21		94.2	19.1	−0.49	
621	46.6	40.4	0.49		113.4	42.2	−0.40	
667	24.2	20.1	0.78		63.2	40.8	−0.15	

6. Discussion

6.1. Hyperspectral vs. multispectral

It is widely accepted that satellite-based hyperspectral observations have the potential to boost our knowledge of aquatic ecosystems at global scales. Using a fairly large database of *in situ* data, we showed that our developed machine-learning models outperform existing Chl*a* retrievals methods and IOP inversion models. We demonstrated that 20 to 25% error in Chl*a* retrievals is achievable. Nevertheless, compared to uncertainties attributed to retrievals from multispectral sensors like OLCI and MSI (Pahlevan et al., 2020), the performance of MDN is found to be only marginally improved when >35 bands are utilized. Although our dataset does not represent all global aquatic ecosystems, it is inferred that the spectral information content provided through OLCI and even MSI may be adequate for the retrieval of Chl*a* in these eutrophic waters (Gitelson et al., 2007). We speculate that expanding the validation data may further favor hyperspectral observations.

The performance of MDN models for $\hat{a}_{ph}$ retrievals using multispectral observations, such as those made by the Moderate Resolution Imaging Spectroradiometer (MODIS), is provided in Fig. 13. Using the development data (Section 3.1), we devised a model with simulated MODIS R_{rs} and a_{ph} spectra (both ocean and land bands) as input and output features, respectively. In comparison to Figs. 4 and 5, while it is evident that MDN considerably outperforms heritage QAA and GIOP models, it exhibits 3 to 25% larger errors than those obtained through the model trained with simulated HICO spectra. This brief analysis supports that it is primarily the model itself that enables high-quality retrievals of $\hat{a}_{ph}$, though it tends to somewhat leverage inter-band correlations present in the hyperspectral domain. A more extensive dataset

Table 5
Performance statistics associated with Chl*a* retrievals from R_{rs} (Chla^{e}) and $\widehat{R}_{rs}$ (Chla^{r}) calculated for *in situ* samples and HICO matchups in FL estuaries. Top algorithms are reported with boldfaced fonts.

	Chla^{e}					Chla^{r}			
	ϵ [%]	β [%]	S	N		ϵ [%]	β [%]	S	N
MDN-I	**17.8**	**−0.8**	**0.91**	38	MDN-S	65.4	49.5	0.41	105
Blend	136.7	136.7	0.34	38	Blend	239.8	239.8	0.29	104
OC6	121.3	87.9	0.22	38	OC6	452.8	383.3	0.35	103
NDCI	44.1	−2.3	0.51	38	NDCI	**51.2**	**40.0**	**0.44**	105
Gons	38.9	−38.9	1.27	5	Gons	79.8	56.6	0.62	97
GI2B	29.1	2.7	0.95	37	GI2B	117.9	98.5	0.52	103

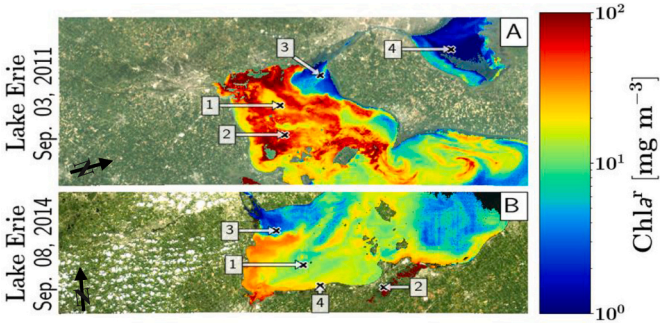

Fig. 9. HICO Chla^r maps derived by implementing MDN-S over the western basin of Lake Erie. The images have been processed using an identical atmospheric correction processing pipeline (Ibrahim et al., 2018). The station numbers correspond to the $\widehat{R}_{rs}$ and $\widehat{a}^r_{ph}$ spectra evaluated in Fig. 10.

is required to arrive at a general conclusion on the extent to which the model fully utilizes the hyperspectral information content.

The main advantage of hyperspectral data in optically deep waters is in their ability to address aquatic biodiversity (Muller-Karger et al.

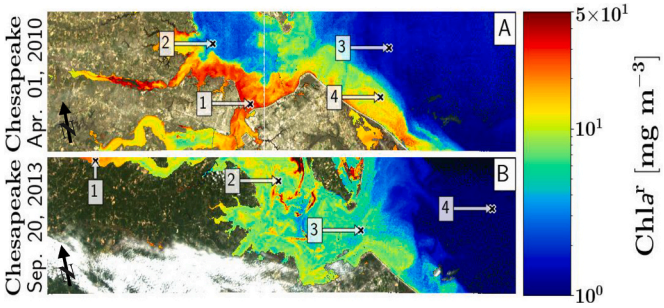

Fig. 11. Same as Fig. 9 but for the lower Chesapeake Bay region. The station numbers correspond to the $\widehat{R}_{rs}$ and $\widehat{a}^r_{ph}$ spectra in Fig. 12.

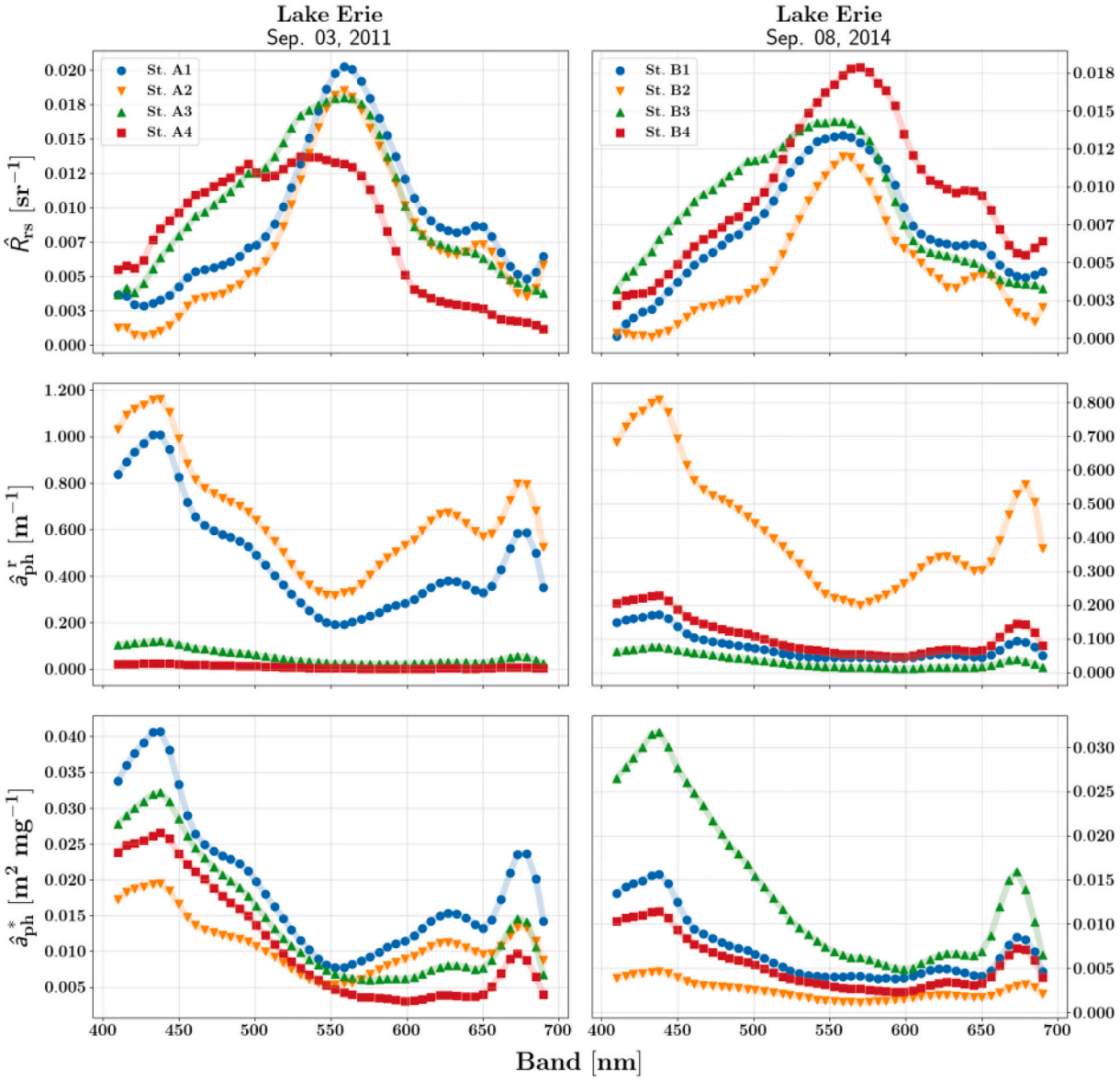

Fig. 10. HICO $\widehat{R}_{rs}$ and the corresponding $\widehat{a}^r_{ph}$ and $\widehat{a}^*_{ph}$ at four arbitrary locations (St.)in Fig. 9. Note the differences in the range of y-axes. The spectral bands beyond 700 nm are truncated to highlight $\widehat{a}^r_{ph}$ retrieval in the visible spectrum.

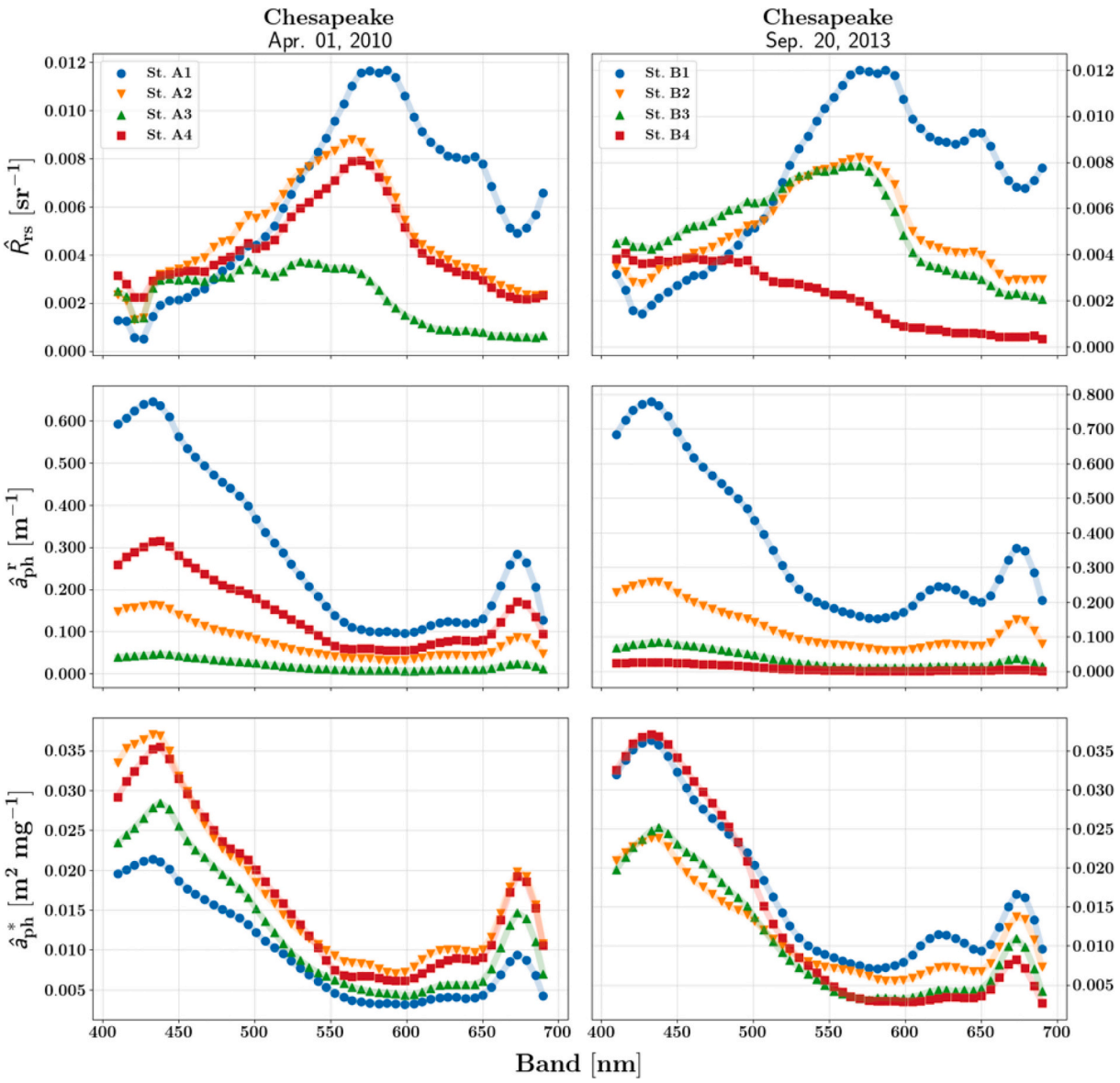

Fig. 12. HICO $\widehat{R}_{rs}$ and the corresponding $\widehat{a}_{ph}^{r}$ and $\widehat{a}_{ph}^{*}$ at four arbitrary locations (St.) in the maps shown in Fig. 11. Note the differences in the range of y-axes. The spectral bands beyond 700 nm are truncated to highlight $\widehat{a}_{ph}^{r}$ retrievals in the visible spectrum.

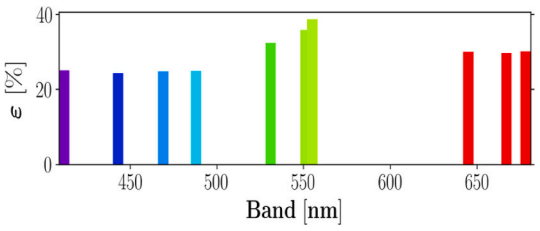

Fig. 13. MDN performance assessment associated with $\widehat{a}_{ph}$ for a multispectral mission (i.e., MODIS). An MDN model was trained with simulated MODIS R_{rs} - a_{ph} spectra. See Fig. 5 for comparison.

2018) (e.g., pigment and phytoplankton community composition) through revealing subtle spectral features in a_{ph} and b_{bp} (Roelke et al., 1999; Vandermeulen et al., 2017; Wolanin et al., 2016). Such spatially explicit products together with ecosystem models combined through data assimilation techniques are anticipated to improve HAB forecasting skills in the future (Schofield et al., 1999). Further, more accurate retrievals of a_{ph} and b_{bp} may lead to an improved differentiation of organic and inorganic particles and offer flexibility in the choice of algorithm solutions, including machine-learning models. Nonetheless, the MDN model is expected to enable equally robust a_{ph} retrievals from existing multispectral missions; although spectral information would be inadequate to fully infer phytoplankton community structure.

6.2. MDN models

Following their first successful demonstration (Pahlevan et al., 2020), MDNs are found to be capable of handling inverse problems to the extent that they are trained with a representative dataset. This model shows major promise in retrieving a_{ph} compared to the heritage inversion techniques constructed for retrievals in fairly clear oceanic environments. While it is essential for the model performance to be validated extensively, we anticipate that the performance is enhanced when simultaneous retrievals of multiple IOP components and WQ parameters are attempted. Under such scenarios, statistical approaches, such as multiple imputations (Sovilj et al., 2016), could be utilized to fill missing features (e.g., paired $R_{rs} - a_{ph}$ measurements are available, but a_{cdom} is

not measured). With this strategy, the trained model will learn correlations among IOPs, Chl*a*, and other parameters of interest to further constrain the solution space. Two other aspects of our research should be highlighted. First, despite the use of an independent *in situ* matchup dataset in FL estuaries to assess the model performances, future research should be dedicated to appraising the performance of MDN in other regions (Dierssen et al., 2020). This may only be possible through major airborne campaigns or by relying on hyperspectral imagery gathered by current sensors (e.g., PRecursore IperSpettrale della Missione Applicativa (PRISMA) (Candela et al., 2016), DLR Earth-Sensing Imaging Spectrometer (DESIS); (Alonso et al., 2019)) prior to the launch of future hyperspectral missions driven by aquatic requirements (e.g., PACE; (Cetinic et al., 2019)). Second, although our achieved uncertainties in $\widehat{a}_{\mathrm{ph}}$ (Fig. 5) indicate significant enhancements ($20 < \epsilon < 30\%$) with respect to those obtained from generic algorithms, given the retrieval statistics and illustrations in Table 4 and Fig. 8, for a robust assessment of aquatic biodiversity more accurate estimates are desired, i.e., $\epsilon < 10\%$ (CEOS, 2018; Muller-Karger et al., 2018; Roelke et al., 1999). This uncertainty requirement should be precisely characterized through end-to-end modeling exercises focusing on categorizing phytoplankton taxa, genera, and species via their optical properties.

6.3. Atmospheric correction and instrument performance

Despite the observation that heritage AC provided improved $\widehat{R}_{\mathrm{rs}}$ products with respect to previous studies, negative retrievals in $\widehat{R}_{\mathrm{rs}}(\lambda < 500\ \mathrm{nm})$ were frequently identified in highly eutrophic and/or turbid waters, which may have been induced by either uncertainties in the extrapolation of the aerosol contribution into the blue spectral region or inadequate accounting of backscattering in the near-infrared bands (Frouin et al., 2019). As expected, these uncertainties have explicit implications in downstream products. For instance, in this study, underestimated $\widehat{R}_{\mathrm{rs}}$ yielded ~50% overestimations of Chla^{e} r (Tables 3 and 5). The corresponding impacts on $\widehat{a}_{\mathrm{ph}}$ were on the order of two-to-four times larger errors at selected HICO bands (Table 4). Needless to state that inadequate vicarious calibrations (due to limited number of matchups), residual biases in HICO's TOA measurements (Ibrahim et al., 2018; Pahlevan et al., 2017), and adjacency effects (Sterckx et al., 2011) may also contribute to additional uncertainties in these products. It is also worthwhile noting that rigorous corrections for gaseous absorption, in particular for bands utilized in the aerosol estimation (e.g., 747 nm), were found to be crucial in producing realistic $\widehat{R}_{\mathrm{rs}}$.

Although HICO was a proof-of-concept mission and lacked radiometric stability required for aquatic studies (Ibrahim et al., 2018), we found that its signal-to-noise (SNR), i.e., >200:1 for the 400–600 nm bands and > 100:1 for 600–700 nm (Lucke et al., 2011), should provide adequate products for robust analyses of aquatic ecosystems (Moses et al., 2012a). Nevertheless, one may attempt to boost SNR by spatial and/or spectral band aggregations assuming the systematic noise (e.g., striping, electronic/optical crosstalk, poor dark-current correction) is negligible. That said, MDN appeared to be insensitive to spectral artifacts in HICO products (Section 5.2), where fairly smooth $\widehat{a}^{\mathrm{r}}_{\mathrm{ph}}$ spectra are derived despite the presence of unnatural features in $\widehat{R}_{\mathrm{rs}}$.

7. Conclusion

This study serves as the first demonstration for the hyperspectral retrieval of a_{ph} and Chl*a* through mixture density networks (MDNs). To do so, hyperspectral *in situ* data as well as spaceborne radiometric observations acquired by the proof-of-concept HICO mission were employed. The validation data ($N = 722$ for a_{ph} and $N = 1902$ for Chl*a*) confirmed that the MDN models significantly outperform the existing global models with 20 to 30% overall median errors and near-zero biases in retrievals. The models implemented on atmospherically corrected HICO images were analyzed for their performances using an independent matchup dataset. The performance loss due to uncertainties in $\widehat{R}_{\mathrm{rs}}$ was quantified to range from three-to-four times for Chla^{e} and $\widehat{a}^{\mathrm{r}}_{\mathrm{ph}}$, respectively. Despite the uncertainties in the atmospheric correction, the extracted $\widehat{a}^{\mathrm{r}}_{\mathrm{ph}}$ spectra were in agreement with previously published spectra for selected locations across four HICO swaths in Lake Erie and the Chesapeake Bay. It is concluded that to enable rigorous HAB characterizations and analyses of phytoplankton biodiversity and their community composition, further improvements in a_{ph} retrievals are required, i.e., < 10% error across the visible bands. Future research is anticipated to enhance the model performance by simultaneously retrieving IOPs, water quality indicators, and physical and/or environmental variables (e.g., skin temperature, salinity) to constrain the solution space. Nevertheless, the performance of the atmospheric correction in coastal and freshwater ecosystems still appears to be the primary challenge limiting practical uses of hyperspectral satellite-derived IOP products, such as a_{ph}.

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.

Acknowledgement

We offer our sincere appreciation to all the individuals and entities that supported this study through their directly or indirectly supplied *in situ* data. The principal investigators providing the data included Blake Schaeffer, John Schalles, Roghua Ma, Krista Alikas, Steven Greb, Bunkei Matsushita, Anatoly Gitelson, Wesley Moses, Deepak Mishra, and Moritz Lehman. We also acknowledge NASA's Ocean Ecology Lab for creating and maintaining SeaBASS and incorporating HICO processing capability into SeaDAS. We are also grateful to two anonymous reviewers for their constructive comments. Nima Pahlevan is funded under NASA ROSES contracts/grants # 80HQTR19C0015, Remote Sensing of Water Quality Element, # 80NSSC20M0235, PACE Science and Applications Team, as well as the USGS Landsat Science Team Award # 140G0118C0011. The codes for retrieving phytoplankton biomass and absorption properties from both simulated and recorded HICO data are available through https://github.com/STREAM-RS/STREAM-RS

References

Albert, A., Gege, P., 2006. Inversion of irradiance and remote sensing reflectance in shallow water between 400 and 800 nm for calculations of water and bottom properties. Appl. Opt. 45, 2331–2343.
Alcantarilla, P.F., Solutions, T., 2011. Fast explicit diffusion for accelerated features in nonlinear scale spaces. IEEE Trans. Patt. Anal. Mach. Intell 34, 1281–1298.
Alonso, K., Bachmann, M., Burch, K., Carmona, E., Cerra, D., de los Reyes, R., Dietrich, D., Heiden, U., Hölderlin, A., Ickes, J., 2019. Data products, quality and validation of the DLR earth sensing imaging spectrometer (DESIS). Sensors 19, 4471.
Alvain, S., Moulin, C., Dandonneau, Y., Bréon, F.-M., 2005. Remote sensing of phytoplankton groups in case 1 waters from global SeaWiFS imagery. Deep-Sea Res. I Oceanogr. Res. Pap. 52, 1989–2004.
Anderson, D.M., 2004. The growing problem of harmful algae. Oceanus magazine 43 (1), 1–5.
Babin, M., Stramski, D., Ferrari, G.M., Claustre, H., Bricaud, A., Obolensky, G., Hoepffner, N., 2003. Variations in the light absorption coefficients of phytoplankton, nonalgal particles, and dissolved organic matter in coastal waters around Europe. J. Geophys. Res. 108.
Balasubramanian, S.V., Pahlevan, N., Smith, B., Binding, C., Schalles, J., Loisel, H., Gurlin, D., Greb, S., Alikas, K., Randla, M., Bunkei, M., Moses, W., Nguyễn, H., Lehmann, M.K., O'Donnell, D., Ondrusek, M., Han, T.-H., Fichot, C.G., Moore, T., Boss, E., 2020. Robust algorithm for estimating total suspended solids (TSS) in inland and nearshore coastal waters. Remote Sens. Environ. 246, 111768.
Barnes, B.B., Hu, C., Cannizzaro, J.P., Craig, S.E., Hallock, P., Jones, D.L., Lehrter, J.C., Melo, N., Schaeffer, B.A., Zepp, R., 2014. Estimation of diffuse attenuation of ultraviolet light in optically shallow Florida Keys waters from MODIS measurements. Remote Sens. Environ. 140, 519–532.
Berry, M.A., Davis, T.W., Cory, R.M., Duhaime, M.B., Johengen, T.H., Kling, G.W., Marino, J.A., Den Uyl, P.A., Gossiaux, D., Dick, G.J., 2017. Cyanobacterial harmful

algal blooms are a biological disturbance to Western Lake Erie bacterial communities. Environ. Microbiol. 19, 1149–1162.
Binding, C., Zastepa, A., Zeng, C., 2019. The impact of phytoplankton community composition on optical properties and satellite observations of the 2017 western Lake Erie algal bloom. J. Great Lakes Res. 45, 573–586.
Bishop, C.M., 1994. Mixture Density Networks. NCRG/94/004. Aston University, Birmingham. http://www.ncrg.aston.ac.uk.
Brando, V.E., Dekker, A.G., 2003. Satellite hyperspectral remote sensing for estimating estuarine and coastal water quality. IEEE Trans. Geosci. Remote Sens. 41, 1378–1387.
Brando, V.E., Dekker, A.G., Park, Y.J., Schroeder, T., 2012. Adaptive semianalytical inversion of ocean color radiometry in optically complex waters. Appl. Opt. 51, 2808–2833.
Bresciani, M., Giardino, C., Bartoli, M., Tavernini, S., Bolpagni, R., Nizzoli, D., 2011. Recognizing harmful algal bloom based on remote sensing reflectance band ratio. J. Appl. Remote. Sens. 5, 053556.
Bresciani, M., Pinardi, M., Free, G., Luciani, G., Ghebrehiwot, S., Laanen, M., Peters, S., Della Bella, V., Padula, R., Giardino, C., 2020. The use of multisource optical sensors to study phytoplankton spatio-temporal variation in a Shallow Turbid Lake. Water 12, 284.
Brewin, R.J., Devred, E., Sathyendranath, S., Lavender, S.J., Hardman-Mountford, N.J., 2011. Model of phytoplankton absorption based on three size classes. Appl. Opt. 50, 4535–4549.
Bricaud, A., Babin, M., Morel, A., Claustre, H., 1995. Variability in the chlorophyll-specific absorption coefficients of natural phytoplankton: analysis and parameterization. J. Geophys. Res. Oceans 100, 13321–13332.
Bricaud, A., Mejia, C., Blondeau-Patissier, D., Claustre, H., Crepon, M., Thiria, S., 2007. Retrieval of pigment concentrations and size structure of algal populations from their absorption spectra using multilayered perceptrons. Appl. Opt. 46, 1251–1260.
Brooks, B.W., Lazorchak, J.M., Howard, M.D., Johnson, M.V.V., Morton, S.L., Perkins, D. A., Reavie, E.D., Scott, G.I., Smith, S.A., Steevens, J.A., 2016. Are harmful algal blooms becoming the greatest inland water quality threat to public health and aquatic ecosystems? Environ. Toxicol. Chem. 35, 6–13.
Buchhorn, M., Lesiv, M., Tsendbazar, N.-E., Herold, M., Bertels, L., Smets, B., 2020. Copernicus global land cover layers—collection 2. Remote Sens. 12, 1044.
Bukata, R.P., Jerome, J.H., Kondratyev, K.Y., Pozdnyakov, D.V., 1995. Optical Properties and Remote Sensing of Inland and Coastal Waters. CRC Press, New York.
Cairo, C., Barbosa, C., Lobo, F., Novo, E., Carlos, F., Maciel, D., Flores Júnior, R., Silva, E., Curtarelli, V., 2020. Hybrid chlorophyll-a algorithm for assessing trophic states of a tropical Brazilian reservoir based on MSI/Sentinel-2 data. Remote Sens. 12, 40.
Candela, L., Formaro, R., Guarini, R., Loizzo, R., Longo, F., Varacalli, G., 2016. The PRISMA mission. In: 2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS). IEEE, pp. 253–256.
Cannizzaro, J.P., Carder, K.L., Chen, F.R., Heil, C.A., Vargo, G.A., 2008. A novel technique for detection of the toxic dinoflagellate, Karenia brevis, in the Gulf of Mexico from remotely sensed ocean color data. Cont. Shelf Res. 28, 137–158.
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.
Casey, K.A., Rousseaux, C.S., Gregg, W.W., Boss, E., Chase, A.P., Craig, S.E., Mouw, C.B., Reynolds, R.A., Stramski, D., Ackleson, S.G., Bricaud, A., Schaeffer, B., Lewis, M.R., Maritorena, S., 2019. A global compilation of in situ aquatic high spectral resolution inherent and apparent optical property data for remote sensing applications. Earth Syst. Sci. Data Discuss. 2019, 1–29.
CEOS, 2018. In: Dekker, A.G.P. (Ed.), Feasibility Study for an Aquatic Ecosystem Earth Observing System. CSIRO.
Cetinic, I., McClain, C.R., Werdell, P.J., Ahmad, Z., Franz, B.A., Karakoylu, E.M., McKinna, L.I., Patt, F.S., 2019. PACE Technical Report Series, Volume 6: Data Product Requirements and Error Budgets Consensus Document.
Chaffin, J.D., Bridgeman, T.B., Heckathorn, S.A., Mishra, S., 2011. Assessment of Microcystis growth rate potential and nutrient status across a trophic gradient in western Lake Erie. J. Great Lakes Res. 37, 92–100.
Chami, M., Robilliard, D., 2002. Inversion of oceanic constituents in case I and II waters with genetic programming algorithms. Appl. Opt. 41, 6260–6275.
Clarke, G.L., Ewing, G.C., Lorenzen, C.J., 1970. Spectra of backscattered light from the sea obtained from aircraft as a measure of chlorophyll concentration. Science 167, 1119–1121.
Craig, S.E., Jones, C.T., Li, W.K., Lazin, G., Horne, E., Caverhill, C., Cullen, J.J., 2012. Deriving optical metrics of coastal phytoplankton biomass from ocean colour. Remote Sens. Environ. 119, 72–83.
da Silva, E.F.F., de Moraes Novo, E.M.L., de Lucia Lobo, F., Barbosa, C.C.F., Noernberg, M.A., da Silva Rotta, L.H., Cairo, C.T., Maciel, D.A., Júnior, R.F., 2020. Optical water types found in Brazilian waters. Limnology 1–12.
Davis, T.W., Bullerjahn, G.S., Tuttle, T., McKay, R.M., Watson, S.B., 2015. Effects of increasing nitrogen and phosphorus concentrations on phytoplankton community growth and toxicity during Planktothrix blooms in Sandusky Bay, Lake Erie. Environ. Sci. Technol. 49, 7197–7207.
Defoin-Platel, M., Chami, M., 2007. How ambiguous is the inverse problem of ocean color in coastal waters? J. Geophys. Res. 112.
Devred, E., Turpie, K., Moses, W., Klemas, V., Moisan, T., Babin, M., Toro-Farmer, G., Forget, M.-H., Jo, Y.-H., 2013. Future retrievals of water column bio-optical properties using the Hyperspectral infrared imager (HyspIRI). Remote Sens. 5, 6812–6837.
Dierssen, H., Bracher, A., Brando, V.E., Loisel, H., Ruddick, K.G., 2020. Data needs for hyperspectral detection of algal diversity across the globe. Oceanography 33.
Doerffer, R., Schiller, H., 2007. The MERIS case 2 water algorithm. Int. J. Remote Sens. 28, 517–535.
Effler, S., Perkins, M., Johnson, D., 1998. The optical water quality of Cannonsville Reservoir: spatial and temporal patterns, and the relative roles of phytoplankton and inorganic tripton. Lake Reserv. Manag. 14, 238–253.
Eleveld, M.A., Ruescas, A.B., Hommersom, A., Moore, T.S., Peters, S.W., Brockmann, C., 2017. An optical classification tool for global lake waters. Remote Sens. 9, 420.
Falkowski, P.G., Laws, E.A., Barber, R.T., Murray, J.W., 2003. Phytoplankton and their role in primary, new, and export production. In: Ocean Biogeochemistry. Springer, pp. 99–121.
Feng, H., Campbell, J.W., Dowell, M.D., Moore, T.S., 2005. Modeling spectral reflectance of optically complex waters using bio-optical measurements from Tokyo Bay. Remote Sens. Environ. 99, 232–243.
Fichot, C.D.G., Downing, B.D., Bergamaschi, B.A., Windham-Myers, L., Marvin-DiPasquale, M., Thompson, D.R., Gierach, M.M., 2015. High-resolution remote sensing of water quality in the San Francisco Bay–Delta Estuary. Environ. Sci. Technol. 50, 573–583.
Fischler, M.A., Bolles, R.C., 1981. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. Commun. ACM 24, 381–395.
Frouin, R.J., Franz, B.A., Ibrahim, A., Knobelspiesse, K., Ahmad, Z., Cairns, B., Chowdhary, J., Dierssen, H.M., Tan, J., Dubovik, O., 2019. Atmospheric correction of satellite ocean-color imagery during the PACE era. Front. Earth Sci. 7, 145.
Garcia, R.A., Fearns, P.R., McKinna, L.I., 2014. Detecting trend and seasonal changes in bathymetry derived from HICO imagery: a case study of Shark Bay, Western Australia. Remote Sens. Environ. 147, 186–205.
Gege, P., 2014. WASI-2D: A software tool for regionally optimized analysis of imaging spectrometer data from deep and shallow waters. Comput. Geosci. 62, 208–215.
Gerace, A.D., Schott, J.R., Nevins, R., 2013. Increased potential to monitor water quality in the near-shore environment with Landsat's next-generation satellite. J. Appl. Remote. Sens. 7, 073558.
Giardino, C., Brando, V., Gege, P., Pinnel, N., Hochberg, E., Knaeps, E., Reusen, I., Doerffer, R., Bresciani, M., Braga, F., 2019. Imaging spectrometry of inland and coastal waters: state of the art, achievements and perspectives. Surv. Geophys. 40, 401–429.
Gilerson, A.A., Gitelson, A.A., Zhou, J., Gurlin, D., Moses, W., Ioannou, I., Ahmed, S.A., 2010. Algorithms for remote estimation of chlorophyll-a in coastal and inland waters using red and near infrared bands. Opt. Express 18, 24109–24125.
Gitelson, A., 1992. The peak near 700 nm on radiance spectra of algae and water: relationships of its magnitude and position with chlorophyll concentration. Int. J. Remote Sens. 13, 3367–3373.
Gitelson, A.A., Schalles, J.F., Hladik, C.M., 2007. Remote chlorophyll-a retrieval in turbid, productive estuaries: Chesapeake Bay case study. Remote Sens. Environ. 109, 464–472.
Gons, H.J., Rijkeboer, M., Ruddick, K.G., 2002. A chlorophyll-retrieval algorithm for satellite imagery (medium resolution imaging spectrometer) of inland and coastal waters. J. Plankton Res. 24, 947–951.
Gordon, H.R., Clark, D.K., Mueller, J.L., Hovis, W.A., 1980. Phytoplankton pigments from the Nimbus-7 Coastal Zone Color Scanner: comparisons with surface measurements. Science 210, 63–66.
Gower, J., King, S., Goncalves, P., 2008. Global monitoring of plankton blooms using MERIS MCI. Int. J. Remote Sens. 29, 6209–6216.
Gurlin, D., Gitelson, A.A., Moses, W.J., 2011. Remote estimation of chl-a concentration in turbid productive waters—return to a simple two-band NIR-red model? Remote Sens. Environ. 115, 3479–3490.
Hakvoort, H., de Haan, J., Jordans, R., Vos, R., Peters, S., Rijkeboer, M., 2002. Towards airborne remote sensing of water quality in the Netherlands—validation and error analysis. ISPRS J. Photogramm. Remote Sens. 57, 171–183.
Harding, L.W., Itsweire, E.C., Esaias, W.E., 1994. Estimates of phytoplankton biomass in the Chesapeake Bay from aircraft remote sensing of chlorophyll concentrations, 1989–92. Remote Sens. Environ. 49, 41–56.
Härmä, P., Vepsäläinen, J., Hannonen, T., Pyhälahti, T., Kämäri, J., Kallio, K., Eloheimo, K., Koponen, S., 2001. Detection of water quality using simulated satellite data and semi-empirical algorithms in Finland. Sci. Total Environ. 268, 107–121.
Hedley, J., Roelfsema, C., Phinn, S.R., 2009. Efficient radiative transfer model inversion for remote sensing applications. Remote Sens. Environ. 113, 2527–2532.
Hestir, E.L., Brando, V.E., Bresciani, M., Giardino, C., Matta, E., Villa, P., Dekker, A.G., 2015. Measuring freshwater aquatic ecosystems: the need for a hyperspectral global mapping satellite mission. Remote Sens. Environ. 167, 181–195.
Hoepffner, N., Sathyendranath, S., 1993. Determination of the major groups of phytoplankton pigments from the absorption spectra of total particulate matter. J. Geophys. Res. Oceans 98, 22789–22803.
Hoogenboom, H., Dekker, A., De Haan, J., 1998. Retrieval of chlorophyll and suspended matter from imaging spectrometry data by matrix inversion. Can. J. Remote. Sens. 24, 144–152.
Hu, C., Lee, Z., Franz, B., 2012. Chlorophylla algorithms for oligotrophic oceans: A novel approach based on three-band reflectance difference. J. Geophys. Res. 117.
Ibrahim, A., Franz, B., Ahmad, Z., Healy, R., Knobelspiesse, K., Gao, B.-C., Proctor, C., Zhai, P.-W., 2018. Atmospheric correction for hyperspectral ocean color retrieval with application to the Hyperspectral Imager for the Coastal Ocean (HICO). Remote Sens. Environ. 204, 60–75.
Ioannou, I., Gilerson, A., Gross, B., Moshary, F., Ahmed, S., 2011. Neural network approach to retrieve the inherent optical properties of the ocean from observations of MODIS. Appl. Opt. 50, 3168–3186.

IOCCG, 2000. Remote Sensing of Ocean Colour in Coastal, and Other Optically-Complex, Waters. S. Sathyendranath Reports of the International Ocean-Colour Coordinating Group.
IOCCG, 2006. In: Lee, Z.-P. (Ed.), Remote Sensing of Inherent Optical Properties: Fundamentals, Tests of Algorithms, and Applications. IOCCG, p. 5.
IOCCG, 2014. In: Sathyendranath, S. (Ed.), Phytoplankton Functional Types from Space. IOCCG, p. 15.
Jamet, C., Thiria, S., Moulin, C., Crépon, M., 2005. Use of a neurovariational inversion for retrieving oceanic and atmospheric constituents from ocean color imagery: a feasibility study. J. Atmos. Ocean. Technol. 22, 460–475.
Jiang, G., Loiselle, S.A., Yang, D., Ma, R., Su, W., Gao, C., 2020. Remote estimation of chlorophyll a concentrations over a wide range of optical conditions based on water classification from VIIRS observations. Remote Sens. Environ. 241, 111735.
Kallio, K., Kutser, T., Hannonen, T., Koponen, S., Pulliainen, J., Vepsäläinen, J., Pyhälahti, T., 2001. Retrieval of water quality from airborne imaging spectrometry of various lake types in different seasons. Sci. Total Environ. 268, 59–77.
Keith, D.J., Schaeffer, B.A., Lunetta, R.S., Gould Jr., R.W., Rocha, K., Cobb, D.J., 2014. Remote sensing of selected water-quality indicators with the hyperspectral imager for the coastal ocean (HICO) sensor. Int. J. Remote Sens. 35, 2927–2962.
Kutser, T., 2004. Quantitative detection of chlorophyll in cyanobacterial blooms by satellite remote sensing. Limnol. Oceanogr. 49, 2179–2189.
Kutser, T., Vahtmäe, E., Paavel, B., Kauer, T., 2013. Removing glint effects from field radiometry data measured in optically complex coastal and inland waters. Remote Sens. Environ. 133, 85–89.
Lee, Z., Carder, K.L., 2004. Absorption spectrum of phytoplankton pigments derived from hyperspectral remote-sensing reflectance. Remote Sens. Environ. 89, 361–368.
Lee, Z., Carder, K.L., Steward, R., Peacock, T., Davis, C., Patch, J., 1998. An empirical algorithm for light absorption by ocean water based on color. J. Geophys. Res. Oceans 103, 27967–27978.
Lee, Z., Carder, K.L., Arnone, R.A., 2002. Deriving inherent optical properties from water color: a multiband quasi-analytical algorithm for optically deep waters. Appl. Opt. 41, 5755–5772.
Lee, Z., Weidemann, A., Kindle, J., Arnone, R., Carder, K.L., Davis, C., 2007. Euphotic zone depth: Its derivation and implication to ocean-color remote sensing. J. Geophys. Res. 112 (n/a-n/a).
Lee, Z., Pahlevan, N., Ahn, Y.-H., Greb, S., O'Donnell, D., 2013. Robust approach to directly measuring water-leaving radiance in the field. Appl. Opt. 52, 1693–1701.
Li, L., Li, L., Song, K., 2015. Remote sensing of freshwater cyanobacteria: an extended IOP Inversion Model of Inland Waters (IIMIW) for partitioning absorption coefficient and estimating phycocyanin. Remote Sens. Environ. 157, 9–23.
Lucke, R.L., Corson, M., McGlothlin, N.R., Butcher, S.D., Wood, D.L., Korwan, D.R., Li, R. R., Snyder, W.A., Davis, C.O., Chen, D.T., 2011. Hyperspectral imager for the Coastal Ocean: instrument description and first images. Appl. Opt. 50, 1501–1516.
Maciel, D., Novo, E., Sander de Carvalho, L., Barbosa, C., Flores Júnior, R., de Lucia Lobo, F., 2019. Retrieving total and inorganic suspended sediments in Amazon Floodplain Lakes: a multisensor approach. Remote Sens. 11, 1744.
Magnuson, A., Harding Jr., L.W., Mallonee, M.E., Adolf, J.E., 2004. Bio-optical model for Chesapeake Bay and the middle Atlantic bight. Estuar. Coast. Shelf Sci. 61, 403–424.
Marshall, H., 1994. Chesapeake bay phytoplankton. 1. Composition. In: Proceedings of the Biological Society of Washington, 107, pp. 573–585.
Marshall, H., Egerton, T., 2010. Increasing occurrence and development of potentially harmful algal blooms in Virginia tidal rivers. In: Water Resources in Changing Climates Richmond. Water Resources Research Center, Virginia.
Marshall, H., Nesius, K., 1996. Phytoplankton composition in relation to primary production in Chesapeake Bay. Mar. Biol. 125, 611–617.
Matthews, M., Bernard, S., 2013. Characterizing the absorption properties for remote sensing of three small optically-diverse south African reservoirs. Remote Sens. 5, 4370–4404.
Mishra, S., Mishra, D.R., 2012. Normalized difference chlorophyll index: a novel model for remote estimation of chlorophyll-a concentration in turbid productive waters. Remote Sens. Environ. 117, 394–406.
Mishra, S., Mishra, D.R., Lee, Z., 2013a. Bio-optical inversion in highly turbid and cyanobacteria-dominated waters. IEEE Trans. Geosci. Remote Sens. 52, 375–388.
Mishra, S., Mishra, D.R., Lee, Z., Tucker, C.S., 2013b. Quantifying cyanobacterial phycocyanin concentration in turbid productive waters: a quasi-analytical approach. Remote Sens. Environ. 133, 141–151.
Mishra, D.R., Schaeffer, B.A., Keith, D., 2014. Performance evaluation of normalized difference chlorophyll index in northern Gulf of Mexico estuaries using the Hyperspectral Imager for the Coastal Ocean. GISci. Remote Sens. 51, 175–198.
Mittenzwey, K.H., Ullrich, S., Gitelson, A., Kondratiev, K., 1992. Determination of chlorophyll a of inland waters on the basis of spectral reflectance. Limnol. Oceanogr. 37, 147–149.
Mobley, C.D., 1994. Light and Water: Radiative Transfer in Natural Waters. Academic Press, Inc.
Mobley, C.D., 1999. Estimation of the remote-sensing reflectance from above-surface measurements. Appl. Opt. 38, 7442–7455.
Mobley, C.D., Sundman, L.K., Davis, C.O., Bowles, J.H., Downes, T.V., Leathers, R.A., Montes, M.J., Bissett, W.P., Kohler, D.D.R., Reid, R.P., Louchard, E.M., Gleason, A., 2005. Interpretation of hyperspectral remote-sensing imagery by spectrum matching and look-up tables. Appl. Opt. 44, 3576–3592.
Moore, T.S., Mouw, C.B., Sullivan, J.M., Twardowski, M.S., Burtner, A.M., Ciochetto, A. B., McFarland, M.N., Nayak, A.R., Paladino, D., Stockley, N.D., Johengen, T.H., Yu, A.W., Ruberg, S., Weidemann, A., 2017. Bio-optical properties of cyanobacteria blooms in Western Lake Erie. Front. Mar. Sci. 4.
Morley, S.K., Brito, T.V., Welling, D.T., 2018. Measures of model performance based on the log accuracy ratio. Space Weath. 16, 69–88.
Moses, W.J., Gitelson, A.A., Berdnikov, S., Povazhnyy, V., 2009. Satellite estimation of chlorophyll-a concentration using the red and NIR bands of MERISThe Azov Sea case study. IEEE Geosci. Remote Sens. Lett. 6, 845–849.
Moses, W.J., Bowles, J.H., Lucke, R.L., Corson, M.R., 2012a. Impact of signal-to-noise ratio in a hyperspectral sensor on the accuracy of biophysical parameter estimation in case II waters. Opt. Express 20, 4309–4330.
Moses, W.J., Gitelson, A.A., Berdnikov, S., Saprygin, V., Povazhnyi, V., 2012b. Operational MERIS-based NIR-red algorithms for estimating chlorophyll-a concentrations in coastal waters—the Azov Sea case study. Remote Sens. Environ. 121, 118–124.
Moses, W.J., Gitelson, A.A., Berdnikov, S., Bowles, J.H., Povazhnyi, V., Saprygin, V., Wagner, E.J., Patterson, K.W., 2013. HICO-based NIR–red models for estimating chlorophyll-$ a $ concentration in productive coastal waters. IEEE Geosci. Remote Sens. Lett. 11, 1111–1115.
Mueller, J.L., Morel, A., Frouin, R., Davis, C., Arnone, R., Carder, K., Lee, Z., Steward, R., Hooker, S., Mobley, C., 2003. Ocean Optics Protocols for Satellite Ocean Color Sensor Validation, Revision 4. Volume III: Radiometric Measurements and Data Analysis Protocols.
Muller-Karger, F.E., Hestir, E., Ade, C., Turpie, K., Roberts, D.A., Siegel, D., Miller, R.J., Humm, D., Izenberg, N., Keller, M., 2018. Satellite sensor requirements for monitoring essential biodiversity variables of coastal ecosystems. Ecol. Appl. 28, 749–760.
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.
O'Reilly, J.E., Werdell, P.J., 2019. Chlorophyll algorithms for ocean color sensors-OC4, OC5 & OC6. Remote Sens. Environ. 229, 32–47.
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. Oceans 103, 24937–24953.
Pahlevan, N., Schott, J., 2013. Leveraging EO-1 to evaluate capability of new generation of landsat sensors for coastal/inland water studies. Select. Topics Appl. Earth Observ. Remote Sens. 6, 360–374.
Pahlevan, N., Roger, J.-C., Ahmad, Z., 2017. Revisiting short-wave-infrared (SWIR) bands for atmospheric correction in coastal waters. Opt. Express 25, 6015–6035.
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.
Pitarch, J., Odermatt, D., Kawka, M., Wüest, A., 2014. Retrieval of particle scattering coefficients and concentrations by genetic algorithms in stratified lake water. Remote Sens. 6, 9530–9551.
Ray, R.T., Haas, L.W., Sieracki, M.E., 1989. Autotrophic picoplankton dynamics in a Chesapeake Bay sub-estuary. Mar. Ecol. Prog. Ser. 52, 273.
Roelke, D., Kennedy, C., Weidemann, A., 1999. Use of discriminant and fourth-derivative analyses with high-resolution absorption spectra for phytoplankton research: limitations at varied signal-to-noise ratio and spectral resolution. Gulf Mexico Sci. 17, 2.
Roesler, C.S., 1998. Theoretical and experimental approaches to improve the accuracy of particulate absorption coefficients derived from the quantitative filter technique. Limnol. Oceanogr. 43, 1649–1660.
Roesler, C.S., Boss, E., 2003. Spectral beam attenuation coefficient retrieved from ocean color inversion. Geophys. Res. Lett. 30.
Roesler, C.S., Perry, M.J., 1995. In situ phytoplankton absorption, fluorescence emission, and particulate backscattering spectra determined from reflectance. J. Geophys. Res. Oceans 100, 13279–13294.
Ryan, J.P., Davis, C.O., Tufillaro, N.B., Kudela, R.M., Gao, B.-C., 2014. Application of the hyperspectral imager for the coastal ocean to phytoplankton ecology studies in Monterey Bay, CA, USA. Remote Sens. 6, 1007–1025.
Schaeffer, B.A., Conmy, R.N., Duffy, A.E., Aukamp, J., Yates, D.F., Craven, G., 2015. Northern Gulf of Mexico estuarine coloured dissolved organic matter derived from MODIS data. Int. J. Remote Sens. 36, 2219–2237.
Schiller, H., Doerffer, R., 1999. Neural network for emulation of an inverse model operational derivation of case II water properties from MERIS data. Int. J. Remote Sens. 20, 1735–1746.
Schofield, O., Grzymski, J., Bissett, W.P., Kirkpatrick, G.J., Millie, D.F., Moline, M., Roesler, C.S., 1999. Optical monitoring and forecasting systems for harmful algal blooms: possibility or pipe dream? J. Phycol. 35, 1477–1496.
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, 7404–7422.
Sinha, E., Michalak, A., Balaji, V., 2017. Eutrophication will increase during the 21st century as a result of precipitation changes. Science 357, 405–408.
Smith, M.E., Lain, L.R., Bernard, S., 2018. An optimized chlorophyll a switching algorithm for MERIS and OLCI in phytoplankton-dominated waters. Remote Sens. Environ. 215, 217–227.
Song, K., Li, L., Tedesco, L., Li, S., Duan, H., Liu, D., Hall, B., Du, J., Li, Z., Shi, K., 2013. Remote estimation of chlorophyll-a in turbid inland waters: three-band model versus GA-PLS model. Remote Sens. Environ. 136, 342–357.
Soto, I.M., Cannizzaro, J., Muller-Karger, F.E., Hu, C., Wolny, J., Goldgof, D., 2015. Evaluation and optimization of remote sensing techniques for detection of Karenia brevis blooms on the West Florida Shelf. Remote Sens. Environ. 170, 239–254.
Sovilj, D., Eirola, E., Miche, Y., Björk, K.-M., Nian, R., Akusok, A., Lendasse, A., 2016. Extreme learning machine for missing data using multiple imputations. Neurocomputing 174, 220–231.

Spyrakos, E., O'Donnell, R., Hunter, P.D., Miller, C., Scott, M., Simis, S.G., Neil, C., Barbosa, C.C., Binding, C.E., Bradt, S., 2018. Optical types of inland and coastal waters. Limnol. Oceanogr. 63, 846–870.

Sterckx, S., Knaeps, E., Ruddick, K., 2011. Detection and correction of adjacency effects in hyperspectral airborne data of coastal and inland waters: the use of the near infrared similarity spectrum. Int. J. Remote Sens. 32, 6479–6505.

Strömbeck, N., Pierson, D.C., 2001. The effects of variability in the inherent optical properties on estimations of chlorophyll a by remote sensing in Swedish freshwaters. Sci. Total Environ. 268, 123–137.

Sun, D., Li, Y., Wang, Q., 2009. A unified model for remotely estimating chlorophyll a in Lake Taihu, China, based on SVM and in situ hyperspectral data. IEEE Trans. Geosci. Remote Sens. 47, 2957–2965.

Sydor, M., Gould, R.W., Arnone, R.A., Haltrin, V.I., Goode, W., 2004. Uniqueness in remote sensing of the inherent optical properties of ocean water. Appl. Opt. 43, 2156–2162.

Tan, J., Cherkauer, K.A., Chaubey, I., 2015. Using hyperspectral data to quantify water-quality parameters in the Wabash River and its tributaries, Indiana. Int. J. Remote Sens. 36, 5466–5484.

Tilstone, G., Mallor-Hoya, S., Gohin, F., Couto, A.B., Sá, C., Goela, P., Cristina, S., Airs, R., Icely, J., Zühlke, M., 2017. Which ocean colour algorithm for MERIS in North West European waters? Remote Sens. Environ. 189, 132–151.

Tomlinson, M.C., Stumpf, R.P., Ransibrahmanakul, V., Truby, E.W., Kirkpatrick, G.J., Pederson, B.A., Vargo, G.A., Heil, C.A., 2004. Evaluation of the use of SeaWiFS imagery for detecting Karenia brevis harmful algal blooms in the eastern Gulf of Mexico. Remote Sens. Environ. 91, 293–303.

Tufillaro, N.B., Davis, C.O., 2012. Derivative spectroscopy with HICO®. In: Optical Remote Sensing of the Environment (p. RTu2E. 5). Optical Society of America.

Uitz, J., Claustre, H., Gentili, B., Stramski, D., 2010. Phytoplankton class-specific primary production in the world's oceans: seasonal and interannual variability from satellite observations. Glob. Biogeochem. Cycles 24.

Uitz, J., Stramski, D., Reynolds, R.A., Dubranna, J., 2015. Assessing phytoplankton community composition from hyperspectral measurements of phytoplankton absorption coefficient and remote-sensing reflectance in open-ocean environments. Remote Sens. Environ. 171, 58–74.

Vandermeulen, R.A., Mannino, A., Neeley, A., Werdell, J., Arnone, R., 2017. Determining the optimal spectral sampling frequency and uncertainty thresholds for hyperspectral remote sensing of ocean color. Opt. Express 25, A785–A797.

Wang, G., Lee, Z., Mishra, D.R., Ma, R., 2016. Retrieving absorption coefficients of multiple phytoplankton pigments from hyperspectral remote sensing reflectance measured over cyanobacteria bloom waters. Limnol. Oceanogr. Methods 14, 432–447.

Warren, M.A., Simis, S.G., Martinez-Vicente, V., Poser, K., Bresciani, M., Alikas, K., Spyrakos, E., Giardino, C., Ansper, A., 2019. Assessment of atmospheric correction algorithms for the Sentinel-2A MultiSpectral Imager over coastal and inland waters. Remote Sens. Environ. 225, 267–289.

Werdell, P.J., Behrenfeld, M., Bontempi, P.S., Boss, E., Cairns, B., Davis, G.T., Franz, B.A., Gliese, U.B., Gorman, EE., Hasekamp, O., Knobelspiesse, K.D., Mannino, A., Vanderlei Martins, J., McClain, C., Meister, G., Remer, L.A., 2019. The Plankton, Aerosol, Cloud, ocean Ecosystem mission: status, science, advances. Bulletin of the American Meteorological Society 100, 1775–1794. https://doi.org/10.1175/BAMS-D-18-0056.1.

Werdell, P.J., Franz, B.A., Bailey, S.W., Feldman, G.C., Boss, E., Brando, V.E., Dowell, M., Hirata, T., Lavender, S.J., Lee, Z., 2013. Generalized ocean color inversion model for retrieving marine inherent optical properties. Appl. Opt. 52, 2019–2037.

Werdell, P.J., McKinna, L.I., Boss, E., Ackleson, S.G., Craig, S.E., Gregg, W.W., Lee, Z., Maritorena, S., Roesler, C.S., Rousseaux, C.S., 2018. An overview of approaches and challenges for retrieving marine inherent optical properties from ocean color remote sensing. Prog. Oceanogr. 160, 186–212.

Wolanin, A., Soppa, M.A., Bracher, A., 2016. Investigation of spectral band requirements for improving retrievals of phytoplankton functional types. Remote Sens. 8, 871.

Xue, K., Ma, R., Wang, D., Shen, M., 2019. Optical classification of the remote sensing reflectance and its application in deriving the specific phytoplankton absorption in optically complex lakes. Remote Sens. 11, 184.

Yang, W., Matsushita, B., Chen, J., Fukushima, T., 2011. Estimating constituent concentrations in case II waters from MERIS satellite data by semi-analytical model optimizing and look-up tables. Remote Sens. Environ. 115, 1247–1259.

Yang, W., Matsushita, B., Chen, J., Yoshimura, K., Fukushima, T., 2012. Retrieval of inherent optical properties for turbid inland waters from remote-sensing reflectance. IEEE Trans. Geosci. Remote Sens. 51, 3761–3773.

Zhan, H., Lee, Z., Shi, P., Chen, C., Carder, K.L., 2003. Retrieval of water optical properties for optically deep waters using genetic algorithms. IEEE Trans. Geosci. Remote Sens. 41, 1123–1128.

Zhang, Y., Yin, Y., Wang, M., Liu, X., 2012. Effect of phytoplankton community composition and cell size on absorption properties in eutrophic shallow lakes: field and experimental evidence. Opt. Express 20, 11882–11898.

Zheng, G., Stramski, D., DiGiacomo, P.M., 2015. A model for partitioning the light absorption coefficient of natural waters into phytoplankton, nonalgal particulate, and colored dissolved organic components: a case study for the Chesapeake Bay. J. Geophys. Res. Oceans 120, 2601–2621.