Remote Sensing of Environment 295 (2023) 113706

Contents lists available at ScienceDirect

Remote Sensing of Environment

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

A hyperspectral inversion framework for estimating absorbing inherent optical properties and biogeochemical parameters in inland and coastal waters

Ryan E. O'Shea [a,b,*], Nima Pahlevan [a,b], Brandon Smith [a,b], Emmanuel Boss [c], Daniela Gurlin [d], Krista Alikas [e], Kersti Kangro [e], Raphael M. Kudela [f], Diana Vaičiūtė [g]

[a] NASA Goddard Space Flight Center, Greenbelt, MD, USA
[b] Science Systems and Applications, Inc. (SSAI), Lanham, MD, USA
[c] School of Marine Sciences, University of Maine, Orono, ME, USA
[d] Wisconsin Department of Natural Resources, Madison, WI, USA
[e] Tartu Observatory of the University of Tartu, Toravere, Estonia
[f] Ocean Sciences Department, Institute of Marine Sciences, University of California-Santa Cruz, Santa Cruz, CA, USA
[g] Marine Research Institute, Klaipėda University, Klaipėda, Lithuania

ARTICLE INFO

Edited by Dr. Menghua Wang

Keywords:
Hyperspectral
Machine learning
Mixture density network
Aquatic remote sensing
Inherent optical properties
HICO
PRISMA
PACE

ABSTRACT

The simultaneous remote estimation of biogeochemical parameters (BPs) and inherent optical properties (IOPs) from hyperspectral satellite imagery of globally distributed optically distinct inland and coastal waters is a complex, unsolved, non-unique inverse problem. To tackle this problem, we leverage a machine-learning model termed Mixture Density Networks (MDNs). MDNs outperform operational algorithms by calculating the covariance between the simultaneously estimated products. We train the MDNs on a large ($N = 8237$) dataset of co-aligned, *in situ* measured, hyperspectral remote sensing reflectance (R_{rs}), BPs, and absorbing IOPs from globally representative optically distinct inland and coastal waters. The estimated IOPs include absorption due to phytoplankton (a_{ph}), chromophoric dissolved organic matter (a_{cdom}), and non-algal particles (a_{nap}). The estimated BPs include chlorophyll-*a*, total suspended solids, and phycocyanin (PC). MDNs dramatically reduce uncertainty in the retrievals, relative to operational algorithms, when using a 50/50 dataset split, where the MDNs are trained on a randomly selected half of the *in situ* dataset and validated on the other half. Our model is shown to have higher, or equivalent, generalization performance than the calculated operational algorithms available for all BPs and IOPs (except PC) *via* a leave-one-out cross-validation assessment. The MDNs are sensitive to uncertainties in the hyperspectral satellite R_{rs}, resulting from instrument noise and atmospheric correction; there is a difference of ~37.4–62.8% (using median symmetric accuracy) between the MDNs' estimates derived from co-located satellite-derived R_{rs} and *in situ* R_{rs}. Of the IOPs, a_{cdom} and a_{nap} are less sensitive to uncertainties in hyperspectral satellite imagery relative to a_{ph}, with remote estimates of a_{ph} exhibiting incorrect spectral shape and magnitude relative to *in situ* measured IOPs. Despite the uncertainties in satellite derived R_{rs}, the spatial distributions of BPs and IOPs in MDN-derived product maps of Lake Erie and the Curonian Lagoon, based on imagery taken with the Hyperspectral Imager for the Coastal Ocean (HICO) and PRecursore IperSpettrale della Missione Applicativa (PRISMA), are confirmed *via* co-aligned *in situ* measurements and agree with the literature's understanding of these well-studied regions. The consistency and accuracy of the model on HICO and PRISMA imagery, despite radiometric uncertainties, demonstrate its applicability to future hyperspectral missions, such as the Plankton, Aerosol, Cloud, ocean Ecosystem (PACE) mission, where the simultaneous estimation model will serve as a key part of phytoplankton community composition analysis.

* Corresponding author at: NASA Goddard Space Flight Center, Greenbelt, MD, USA.
E-mail addresses: ryan.oshea@ssaihq.com, ryan.e.o'shea@nasa.gov (R.E. O'Shea).

https://doi.org/10.1016/j.rse.2023.113706
Received 3 February 2023; Received in revised form 28 June 2023; Accepted 29 June 2023
Available online 19 July 2023
0034-4257/© 2023 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

1. Introduction

Multi- and hyperspectral satellite imagery can supplement *in situ* measurements, to achieve the spatiotemporal resolutions required to monitor inland and coastal aquatic ecosystem health, through the remote estimation of optically relevant biogeochemical parameters (BPs) and inherent optical properties (IOPs) (Binding et al., 2021; Devlin et al., 2013; El Serafy et al., 2021; Gohin et al., 2008; Schaeffer et al., 2013). The remote estimation of these BPs and IOPs in inland and coastal waters, on global scales, is complicated by a variety of factors. The first major factor is that the inverse problem (estimating BPs from hyperspectral imagery) is non-unique (particularly in optically complex inland and coastal waters), meaning multiple different combinations of BPs and IOPs in the water column can be associated with the same hyperspectral signal (Defoin-Platel and Chami, 2007). The second major factor is that on global scales these optically complex inland and coastal waters consist of a range of optical water types (Mélin and Vantrepotte, 2015; Spyrakos et al., 2018), so inversion algorithms designed to retrieve BPs and IOPs for one region may not generalize well to other optically distinct regions. Therefore, these inversion algorithms must be validated on a globally representative dataset (Palmer et al., 2015). The third major factor is that uncertainties in the hyperspectral imagery (such as instrument noise and imperfect atmospheric correction) can further impact the applicability of inversion algorithms designed from *in situ* measurements (Ibrahim et al., 2018; Moses et al., 2012; O'Shea et al., 2021). Although many alternative remote sensing algorithms exist for estimating individual BPs and IOPs (Section 2), and even the simultaneous estimation of many of these BPs and IOPs, their capability to generalize on optically distinct inland and coastal water bodies and to perform sufficiently well despite the uncertainties in hyperspectral satellite imagery has not been assessed.

The spectral scattering and absorbing IOPs, which are shaped by the absorption and scattering properties of the optically relevant BPs, determine the upwelling spectral radiance (color) of water. Satellite imagers can remotely estimate IOPs and BPs by applying spectral algorithms to the remote sensing reflectance ($R_{rs}(\lambda)$, where λ is the wavelength; hereafter (λ) will be dropped for brevity), which is defined as the normalized water-leaving radiance (corrected for atmospheric and solar geometry effects and set to the mean earth sun distance) divided by the mean extraterrestrial solar irradiance. Absorbing IOPs can be partitioned into water, phytoplankton, chromophoric dissolved organic matter (CDOM), and non-algal particles (NAP). Pure water absorption and its temperature dependence is well known in the visible part of the spectrum and can be easily removed from the total absorption (Pope and Fry, 1997). Phytoplankton absorption is a combination of the spectral absorption of all pigments, the proportions of which vary by species and, therefore, can offer a method for assessing phytoplankton community composition (Catlett and Siegel, 2018; IOCCG, 2014). One such pigment, chlorophyll-*a* (chl*a*), is ubiquitous to phytoplankton and commonly serves as a proxy for total phytoplankton biomass (Huot et al., 2007; Paltsev and Creed, 2022), though this relationship may be altered by factors such as community composition (Kasprzak et al., 2008). Absorption by CDOM is composed of the dissolved component (operationally defined as <0.2 μm) of chromophoric organic matter. CDOM is sourced, in part, from decaying leaves in lakes and rivers and decomposing phytoplankton in the ocean, which shape the absorption spectrum differently, and thus can serve as a metric for dissolved organic matter (DOM) source and composition (Carder et al., 1989; Grunert et al., 2018). Absorption by non-algal (non-pigmented) particles (often operationally defined as >0.2 μm) includes absorption due to (non-pigmented) phytoplankton cells as well as terrestrial and biogenic inorganic minerals (*e.g.*, silt and calcite, respectively). NAP absorption at 443 nm is highly correlated to the total suspended solids (TSS) concentration, and the slope of their regression may vary with inorganic *versus* organic composition (Babin, 2003; Binding et al., 2008). In summary, this suite of optically relevant BPs and IOPs can serve as a set of diverse proxies, providing a more complete picture of aquatic ecosystem health *via* the composition and abundance of its constituents.

The community composition and abundance of phytoplankton, a metric of such aquatic ecosystem health, can be assessed *via* measurements such as the hyperspectral absorption spectrum and individual pigment concentrations. The hyperspectral phytoplankton absorption spectrum (a_{ph}), which is a combination of the absorption of all phytoplankton pigments, can be estimated from hyperspectral remote sensing measurements (IOCCG, 2014). Individual pigments and phytoplankton size classes can be estimated directly from their absorption spectra using multilayer perceptrons (Bricaud et al., 2007). Additionally, remotely estimated a_{ph} can serve as input to algorithms to estimate the phytoplankton community composition (Catlett and Siegel, 2018; IOCCG, 2014). While pigment compositions of a_{ph} vary greatly between species, two important pigments for water quality management include chlorophyll-*a* (chl*a*) and phycocyanin (PC). PC is a pigment specific to (primarily) freshwater cyanobacteria, a type of aquatic organism that can produce dangerous toxins, and can, therefore, serve as a proxy for potential toxin-producing harmful algal blooms (HABs). Optically, chl*a* absorbs in the blue and red sections of the spectrum and fluoresces in the red (peak ~685 nm), whereas PC absorbs in the orange (620 nm) and fluoresces in the red (650 nm) (Becker et al., 2002; Dekker, 1993; Mobley, 2022; Schalles and Yacobi, 2000). In addition to serving as proxies for aquatic organisms, these pigments can heat the sea surface through enhanced absorption (Kahru et al., 1993), creating ideal conditions for further blooms. Overall, the abundance and community composition of phytoplankton can be estimated *via* their unique spectral signatures, which are defined by their pigment composition.

CDOM is the optically relevant component of dissolved organic matter. CDOM exhibits highest absorption in the blue bands, and this absorption exponentially decreases towards the red (Mobley, 2022; Twardowski et al., 2004). The light absorption by CDOM thereby limits light availability for primary production in benthic and pelagic environments and heats the sea surface (Campbell et al., 2002; Chang and Dickey, 2004; Hill, 2008). By leveraging the spectral absorption signal of CDOM (a_{cdom}), both DOM source and composition can be estimated (Carder et al., 1989; Grunert et al., 2018). CDOM also serves as a proxy for the main source of carbon available for primary production, dissolved organic carbon (DOC) (Hestir et al., 2015; Mannino et al., 2008; Spencer et al., 2012) and as a proxy for salinity in coastal ecosystems (Bowers and Brett, 2008). The lack of operationally available remote-sensing-retrieved spectral CDOM and NAP absorption may lead to inaccurate phytoplankton carbon fixation estimates for freshwater ecosystems with significant CDOM and/or NAP absorption (Sayers et al., 2021). To summarize, CDOM has a complex relationship with aquatic ecosystem health; as increases in DOM can provide heat and nutrients required for blooms, but also limit light availability for primary production.

Much like CDOM absorption, NAP absorption is exponentially decaying as a function of wavelength, with highest absorption in the UV and blue sections of the spectrum, and, therefore, has similar impacts on light availability in aquatic ecosystems (Mobley, 2022). Although the slope of NAP absorption has a narrow range of variations in coastal regions relative to its magnitude, it may be relatable (regionally) to the proportion of organic *versus* mineral components or to the composition of organic particles (Babin, 2003; Bélanger et al., 2013; Bricaud et al., 2010). However, in practice, relating the slope of NAP absorption to the proportion of organic to inorganic matter may vary regionally. In Lake Erie higher slopes had a larger proportion of inorganic matter, whereas in European coastal waters the opposite was true (Babin, 2003; Binding et al., 2008).

TSS is composed of both organic (phytoplankton cells and non-algal particles) and inorganic (primarily mineral soil particles) fractions. Scattering by TSS varies depending on the specific constituents, however both non-algal particles and minerals have been shown to have high scattering in the red and near-infrared (NIR) (Mobley, 2022; Stramski

et al., 2007; Sun et al., 2010). The red and NIR sections of the spectrum are minimally impacted by CDOM or phytoplankton absorption, making these sections of the spectrum ideal for remote sensing in optically complex waters (particularly at higher TSS concentrations (Novoa et al., 2017)). TSS is an important parameter to measure remotely, as TSS transports both trace elements (including heavy metals) and nutrients (Horowitz, 2008). Additionally, particulate organic carbon (the phytoplankton cell and non-algal particle fractions of TSS) is a major contributor to the carbon budget. Since TSS is transported *via* fluid motion, it can be used with geosynchronous satellites to track seawater circulation in turbid coastal waters (Yang et al., 2014), including sediment discharge and circulation in estuaries surrounding river mouths (Baban, 1995). Finally, TSS can serve as a proxy for water clarity, and has been used to study the impact of macroalgae on water clarity in turbid lakes using airborne remote sensing (Giardino et al., 2015). Concisely, TSS can be measured remotely *via* its scattering in the red and NIR, and serve as a proxy for, among other parameters, nutrient availability and water clarity.

The most developed semi- and quasi-analytical models that simultaneously estimate absorption due to phytoplankton, CDOM, and non-algal particles were designed for open ocean waters (Lee et al., 2002; Werdell et al., 2013). The models have been reparameterized for inland and coastal waters, either by shifting reference wavelengths in the red to the NIR or utilizing field data for regional tuning (Najah and Al-Shehhi, 2021; Yang et al., 2013). Unfortunately, these reparameterized models still fail to estimate specific parameters in many inland and coastal waters (Najah and Al-Shehhi, 2021). Since reparameterization of existing semi- and quasi-analytical models is insufficient to solve this complex non-unique inverse problem (Defoin-Platel and Chami, 2007; Sydor et al., 2004), additional constraints may be required to simultaneously estimate these IOPs and BPs in globally distributed optically unique inland and coastal waters.

MDNs solve non-unique inverse problems, such as estimating BPs from R_{rs}, by modeling the probability distribution of the outputs as a mixture of Gaussians (Bishop, 1994, 1995). MDNs then select the prediction with the highest likelihood of occurrence, instead of the average typically reported by multilayer perceptrons, to get a better overall estimate from the multimodal output space. Dedicated MDNs can estimate individual BPs such as chl*a* (Pahlevan et al., 2020, 2021; Smith et al., 2021), TSS (Balasubramanian et al., 2020), PC (O'Shea et al., 2021), and a_{ph} (Pahlevan et al., 2021) from both multispectral and hyperspectral satellite data, in inland and coastal waters. In fact, the same architecture has been successfully applied to simultaneously estimate two BPs and an IOP at a single wavelength (chl*a*, TSS, and $a_{cdom}(440)$) (Pahlevan et al., 2022). In addition to better fitting the multimodal output space for multiple BPs and IOPs, MDNs also have the key feature of estimating uncertainties of their predictions from the covariance matrices of the Gaussian mixture model (Brando Guillaumes, 2017; Choi et al., 2018; Saranathan et al., 2023). The pixel-by-pixel uncertainty product is a key requirement for a reliable interpretation of biogeochemical models that leverage aquatic science products (*e.g.*, chl*a*) (Gould et al., 2014; Sheng et al., 2014; Werdell et al., 2018) and for water quality management using satellite products (IOCCG, 2019). Overall, through their proven efficacy for the non-unique inverse estimation of available BPs and IOPs and their inherent ability to estimate uncertainties, MDNs are well suited for simultaneous retrievals of multiple BPs and IOPs for scientific and water quality management applications.

In preparation for hyperspectral observations by the Ocean Color Instrument (OCI) from the Plankton, Aerosol, Cloud, ocean Ecosystem (PACE) mission, this manuscript aims at demonstrating and validating an inversion scheme (MDNs) to enable the generation of viable products for aquatic science and applications across global inland and coastal waters. This manuscript builds upon previous research (O'Shea et al., 2021; Pahlevan et al., 2021, 2022), by creating MDNs that *simultaneously* estimate both relevant BPs (chl*a*, TSS, and PC) and absorbing IOPs (phytoplankton, NAP, and CDOM) from hyperspectral imagery. The MDNs are validated for product generation on the globally representative *in situ* dataset *via* a leave-one-out regional cross-validation analysis and comparison to operational algorithms. Finally, the MDN-derived products are demonstrated for use on hyperspectral missions (*e.g.*, PACE), *via* comparing *in situ* radiometric and biogeochemical/IOP measurements to product maps produced from imagery from heritage and current hyperspectral imagers (the Hyperspectral Imager for the Coastal Ocean (HICO) and the Italian Space Agency's (ISA's) PRecursore IperSpettrale della Missione Applicativa (PRISMA), respectively). The simultaneous estimation of BPs and absorbing IOPs from hyperspectral imagery *via* MDNs will serve a key role in phytoplankton community composition identification.

The manuscript is structured as follows. The Background section (Section 2) builds upon the Introduction (Section 1) to provide a more in-depth review of absorbing IOPs and retrieval algorithms. The Datasets section (Section 3) covers the spatial distribution and concentration range of the available *in situ* BPs, IOPs, and associated imagery for training and validating the MDN. The Methods section (Section 4) describes the MDN architecture and performance assessment. The Results section (Section 5) reports the performance of the MDN relative to operational algorithms over a variety of assessments. The Discussion section (Section 6) compares the presented MDN to previously developed dedicated models, identifies the major drivers of uncertainty in the model predictions, and discusses implications and future directions. Finally, the Conclusions section (Section 7) summarizes the improvements the presented model makes over the state-of-the-art for estimation of BPs and IOPs, and how this will impact key goals of future hyperspectral missions (*e.g.*, PACE).

2. Background

IOPs include the absorption (a), scattering (b), and total attenuation (c) of water and its constituents at a wavelength (λ) (Eq. 1). IOPs do not vary under changing illumination conditions.

$$a(\lambda) + b(\lambda) = c(\lambda) \tag{1}$$

Total absorption can be further broken down into absorption of water (a_w, known (Pope and Fry, 1997)) and the optically active constituents within the water column, including: absorption due to phytoplankton pigments (a_{ph}), absorption due to chromophoric dissolved organic matter (a_{cdom}), and absorption due to non-algal particles (a_{nap}) (Eq. 2, (Mobley, 2022)).

$$a(\lambda) = a_w(\lambda) + a_{ph}(\lambda) + a_{cdom}(\lambda) + a_{nap}(\lambda) \tag{2}$$

While a_{cdom} and a_{nap} are functions of wavelength (λ), they are often represented (in the visible) by a decreasing exponential (Eq. 3) using only their magnitude at a specific wavelength (λ_0, often ~440 nm) and slope (S, labeled as S-CDOM and S-NAP for CDOM and NAP, respectively) (Mobley, 2022; Twardowski et al., 2004).

$$a_{cdom}(\lambda) = a_{cdom}(\lambda_0)\, e^{-S(\lambda-\lambda_0)} \tag{3}$$

Since a_{cdom} and a_{nap} can both be fit by a decreasing exponential, and are often hard to deconvolve from one another, they are sometimes combined and represented as a single decreasing exponential (a_{dg}) (Cael and Boss, 2017).

Total scattering can be broken down into forward scattering and backward scattering. The backscattering coefficient ($b_b(\lambda)$) is most relevant from a remote sensing perspective, as it controls the magnitude of light backscattered into satellite imagery. Total spectral backscattering is assumed to be defined by water ($b_{bw}(\lambda)$, known, also a function of salinity (Zhang et al., 2009)), and particles ($b_{bp}(\lambda)$), while dissolved constituents other than dissolved ions are assumed not to contribute (Eq. 4 (Werdell et al., 2013)).

$$b_b(\lambda) = b_{bw}(\lambda) + b_{bp}(\lambda_0)\left(\frac{\lambda_0}{\lambda}\right)^{\eta} \tag{4}$$

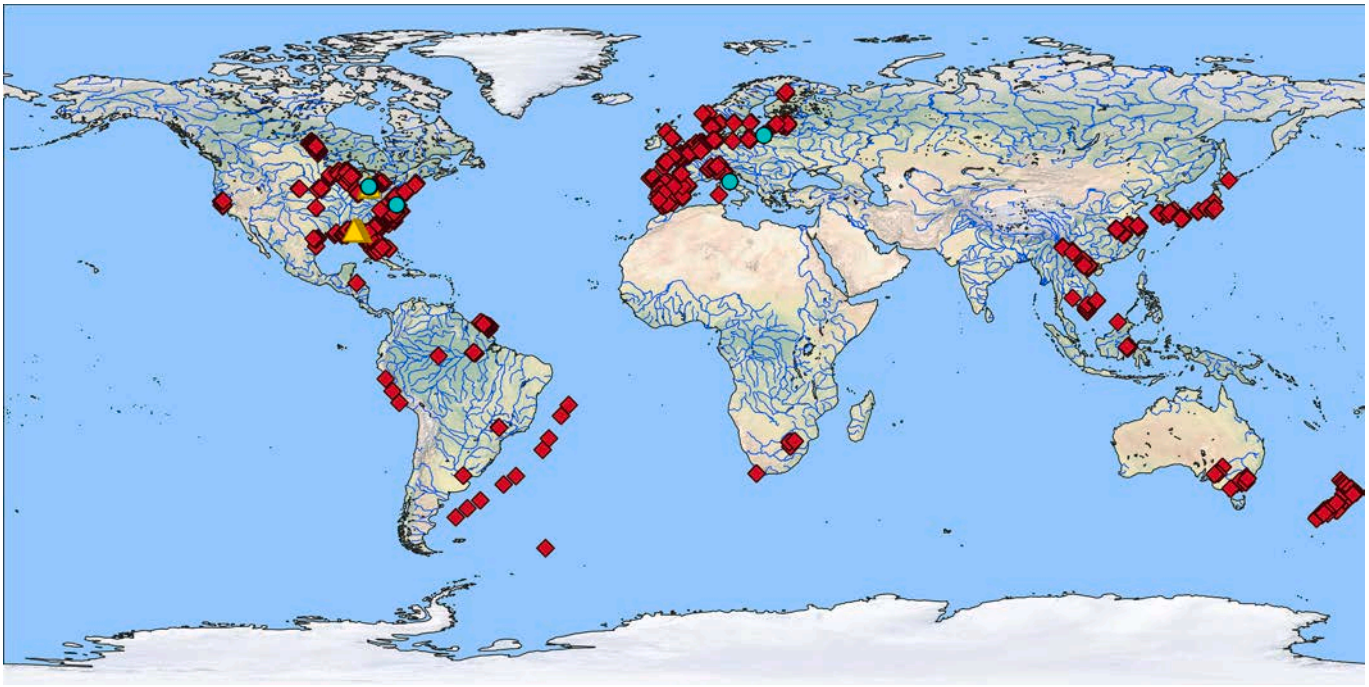

Fig. 1. Global distribution of *in situ* measured R_{rs} and co-located BPs and IOPs. Red diamonds show *in situ* BPs and IOPs co-aligned with *in situ* R_{rs}, cyan circles show *in situ* BPs and IOPs co-aligned with satellite-derived R_{rs}, yellow triangles show *in situ* R_{rs} co-aligned with satellite-derived R_{rs}. Not all samples of the *in situ* dataset have associated latitude and longitude measurements. The basemap was retrieved from: https://www.naturalearthdata.com/. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

where λ_0 is a constant wavelength, and η is the exponential slope (often estimated *via* an empirical relationship (Lee et al., 1996, 1999)).

The quasi-analytical algorithm (QAA) is one of the most commonly used semi-analytical algorithms in the ocean-color community for global IOP retrievals from $R_{rs}(\lambda)$ (Le et al., 2009; Lee et al., 2002, 2023; Pan et al., 2015). QAA is a two step inversion process which first derives the total absorption and backscattering (without leveraging assumed spectral signatures of the individual components) for optically deep waters, before deriving the specific absorption magnitudes of each of the water constituents. First, the total backscattering at a reference wavelength ($b_b(\lambda_0)$, Eq. 4) can be solved for by leveraging the R_{rs} at a known wavelength *via* one of two relationships, either Morel and Maritorena, 2001 ($\lambda_0 = 555$ nm) for open ocean and most coastal waters or Lee et al., 2002 ($\lambda_0 = 640$ nm) for high absorption (>0.5 m^{-1}) waters, or a combination of the two for total absorptions between 0.3 and 0.5 m^{-1}. Second, the total backscattering at the reference wavelength ($b_b(\lambda_0)$) can then be derived from solving Eq. 5 with the absorption for the specific reference wavelength, as $b_{bw}(\lambda)$ and η are known. Total spectral backscatter ($b_b(\lambda)$, Eq. 4) can then be derived as a function of wavelength (Gordon and Morel, 1983; Smith and Baker, 1981), as the exponential slope can be estimated from the remote-sensing reflectance just below the surface (Lee et al., 2002). Now $a(\lambda)$ can be derived *via* Eq. 5, as the $R_{rs}(\lambda)$ and $b_b(\lambda)$ are known.

$$u[R_{rs}(\lambda)] = \frac{b_b(\lambda)}{a(\lambda) + b_b(\lambda)} \tag{5}$$

where u[] is an empirical model that is a function of $R_{rs}(\lambda)$ (Gordon et al., 1988; Lee et al., 1999). Finally, spectral decomposition of the total absorption into a_{ph} and a_{dg} is performed by relating these terms to each other at two specific wavelengths (411 and 443 nm). This algorithm leverages models and empirical relationships that may not hold in all optically distinct aquatic regions.

An alternative, and perhaps even more commonly used IOP inversion technique, developed by the Ocean Biology Processing Group and implemented in NASA's data processing pipeline (Najah and Al-Shehhi, 2021; Werdell et al., 2013), is the Generalized IOP (GIOP) algorithm. The GIOP algorithm provides a framework for evaluating parameterizations of the many similarly formulated semi-analytical algorithms (SAAs) (Werdell et al., 2013). Therefore, GIOP can be reparameterized and tuned to specific regions, by adjusting parameters such as the phytoplankton-specific absorption coefficient (phytoplankton absorption normalized by the chl*a* concentration) and the combined CDOM and NAP slope (Favareto et al., 2018). Although not originally designed for coastal and inland waters, reparameterization could improve GIOPs performance in these regions.

Although QAA and GIOP were not originally designed or validated for turbid coastal and inland waters, many studies have performed validation approaches in globally distributed coastal and inland waters, and found regional adaptations to their algorithms can be made to improve performance (Najah and Al-Shehhi, 2021). Specifically, QAA has been applied to many turbid inland lakes in China (Le et al., 2009; Chu et al., 2020; Wei et al., 2016), where shifting the reference wavelength from the red to the NIR, or using multiple reference wavelengths (Pan et al., 2015), improved overall performance. This algorithm has also been applied, and the empirical relationships modified, for more accurate retrievals from coastal regions in the Yellow and East China Seas (Qing et al., 2011). GIOP has also been both applied and validated in coastal and inland waters (Aguilar-Maldonado et al., 2019). Though minimal adaptations have been performed for GIOP, it is possible that regional tuning of the phytoplankton-specific absorption, and NAP, CDOM, and particulate backscattering spectral slopes could improve regional performance (Favareto et al., 2018). Although the QAA and GIOP performed well in certain coastal and inland waters, with some adaptations, there were also regions experiencing poor performance, including the coastal waters of Australia, Bay of Bengal, Gulf of California, and Chesapeake Bay, to name a few (Betancur-Turizo et al., 2018; Lotliker et al., 2015; Najah and Al-Shehhi, 2021; Qin et al., 2007; Zheng and DiGiacomo, 2018). While QAA and GIOP can be regionally tuned, if suitable *in situ* datasets are available, these algorithms are not ideal for application to all turbid coastal and inland waters, on a global scale. We

Table 1
Location, number of samples, and provider of datasets added to GLORIA. If a dataset was already within GLORIA, but it was augmented with at least one spectral IOP, the number of samples is followed by: (IOPs).

Region	Location	Samples	Provider
Global	Global waters	289	Various providers (*via* SeaBASS)
Asia	Xingyun Lake (China)	19	Ronghua Ma & Zhigang Cao
	Taihu Lake (China)	303	Ronghua Ma & Zhigang Cao
	Vietnamese lakes	185	Hà Nguyễn
	Chinese & Japanese lakes	26 (IOPs)	Bunkei Matsushita & Dalin Jiang
	Indonesian waters	118	CoastColour (Nechad et al., 2015)
Europe	Curonian Lagoon (Lithuania)	55	Diana Vaičiūtė
	French and Belgian estuaries	79	David Doxaran & Kevin Ruddick
	Lake Trasimeno (Italy)	10	Claudia Giardino
	North Sea	48	CoastColour (Nechad et al., 2015)
	Dutch lakes	72 (IOPs)	Stefan G. H. Simis
	Lake Kummerow (Germany)	26 (IOPs)	Natascha Oppelt
	Italian lakes	73 (IOPs)	Claudia Giardino & Mariano Bresciani
	Estonian and Swedish lakes	162 (IOPs)	Krista Alikas, Kersti Kangro, & Martin Ligi
North America (N.A.)	Coastal Atlantic Ocean bordering Chesapeake Bay	60	Richard C. Zimmerman & Glenn Cota (*via* SeaBASS)
	Chesapeake Bay (U.S.)	43	Alexander Gilerson
	Chesapeake Bay (U.S.) & Gulf of Mexico	208	Richard W. Gould (*via* SeaBASS)
	North American coastal waters (predominantly Gulf of Mexico)	438	Blake Schaeffer
	Gulf of Mexico	34	Richard C. Zimmerman (*via* SeaBASS)
	Gulf of Mexico	528	Chuanmin Hu (*via* SeaBASS)
	Grizzly Bay (U.S.)	20	Cédric G. Fichot
	San Francisco Bay (U.S.)	21	Raphael Kudela
	Lake Erie (U.S.)	125	Caren Binding
	Great Lakes (U.S. & Canada)	9 (IOPs)	David M. O'Donnell
	Lake Erie (U.S.)	34 (IOPs)	Timothy S. Moore
	Massachusetts Bay (U.S.)	3	Nima Pahlevan
	Mississippi Aquaculture Ponds (U.S.)	41 (IOPs)	Deepak R. Mishra & Sachidananda Mishra
	Indianan reservoirs (U.S.)	192 (IOPs)	Lin Li
	Wisconsin Lakes (U.S.)	188 (IOPs)	Steven R. Greb & Daniela Gurlin
	Nebraskan Lakes (U.S.)	183 (IOPs)	Anatoly A. Gitelson, Daniela Gurlin, & Wesley J. Moses
Oceania (OC)	Australian Lakes	111 (IOPs)	Janet Anstee & Nathan Drayson

compare these algorithms against our own model in this research as they are widely used and their performance on inland and coastal waters is well documented in the literature.

Direct retrieval of specific BPs, such as pigments, from optically complex inland and coastal waters can be a much simpler process that generally relies on empirical relationships leveraging spectral band ratios. These band ratios focus on bands with unique spectral features associated with a specific pigment while avoiding spectral overlap with those of other optically relevant constituents, such as absorption or fluorescence characteristics of other pigments, which can change independently in optically complex inland and coastal waters. For chl*a* estimation in inland and coastal waters, instead of using the blue and green bands typical for chl*a* retrievals from open ocean waters, retrieval algorithms focus on band ratios between the red and NIR (Gilerson et al., 2010). Band ratios in the red (*e.g.*, 665 nm) and NIR (*e.g.*, 709 nm) are minimally impacted by CDOM and NAP absorption in these optically complex waters. For coastal regions that undergo algal blooms, variations can span three orders of magnitude, so blended switching algorithms have been developed that leverage the blue-green relationship at lower concentrations, and red-NIR at higher concentrations (associated with a higher NIR-red ratio) (Smith et al., 2018). Early phycocyanin algorithms leveraged the pigment's fluorescence (650 nm) and absorption (620–625 nm) peaks *via* a simple band ratio (Dekker, 1993; Schalles and Yacobi, 2000), but further advances implemented correction for absorption by chl*a* in these bands by leveraging the red (665 nm) and NIR (709 nm) bands (Simis et al., 2005, 2007). In even more optically complex inland waters, with wide variations in accessory pigments, their spectral overlap with the respective signatures of PC and chl*a* can further limit an algorithm's applicability, leading to an overestimation of PC (Ruiz-Verdú et al., 2008; Simis et al., 2007). Retrieval of TSS is also more complex in coastal and inland waters, where concentrations can span up to three orders of magnitude. While single-band algorithms leveraging red wavelengths (665 nm) can work well (Nechad et al., 2015), algorithms can also switch between using single band algorithms in the green, red, and NIR bands to attempt to achieve the highest possible accuracy in their retrievals (Novoa et al., 2017). Although these empirical relationships work well in multi-band operational algorithms, leveraging the full hyperspectral signal *via* machine learning models has increased the retrieval accuracy from globally distributed optically complex inland and coastal waters (O'Shea et al., 2021; Pahlevan et al., 2022).

3. Datasets

3.1. *In situ* R_{rs}, BPs, and absorbing IOPs

Our model development dataset (N = 8237 for HICO) consists of globally distributed (Fig. 1) samples of *in situ* R_{rs} (approximated as the water leaving radiance divided by the downwelling solar irradiance (Gordon and Voss, 2004; Mobley, 2022; Mobley, 1999)) co-located with at least one biogeochemical parameter (chl*a*, PC, TSS, or CDOM) or IOP (a_{ph}, a_{nap}, or a_{cdom}). Retrievals of the majority of the BPs and IOPs have been previously described in a variety of publications, each with maps of individual parameters by region (Balasubramanian et al., 2020; Jiang et al., 2021; O'Shea et al., 2021; Pahlevan et al., 2021, 2022; Smith et al., 2021). Our dataset (Table 1) augments the chl*a*, TSS, and CDOM presented in the GLObal Reflectance community dataset for Imaging and optical sensing of Aquatic environments (GLORIA) (Lehmann et al., 2023) with PC and spectral IOPs (a_{ph}, a_{nap}, or a_{cdom}). Additional datasets including IOPs come from a variety of globally distributed waters, either directly from contributors, or *via* compiled datasets (Casey et al., 2019; Nechad et al., 2015; Werdell and Bailey, 2002).

Some previously presented datasets were also pruned, including data from Dutch Lakes and data from coastal estuaries of the south and east U.S. coasts. Repeat PC measurements associated with each R_{rs} were removed from the Dutch Lakes dataset. Suspicious outliers were removed from the predominantly south and east U.S. coastal waters dataset (contributed by John F. Schalles). 84 suspicious outliers were removed, these samples had an Rrs(410)/Rrs(450) > 1.5 and at least one wavelength that did not agree with previously measured ocean color values (*i.e.*, they had a quality assurance score of <1, a metric assessing the proportion of spectral R_{rs} that agree with previously measured ocean color values (Wei et al., 2016)).

Total samples for each R_{rs} and BP/IOP pair are: 5740 for chl*a*, 5312 for TSS, 5165 for CDOM (a_{cdom} at 440 nm), 940 for PC, 2425 for a_{ph}, 2289 for a_{nap}, and 2764 for a_{cdom} (at any wavelength). **We output two**

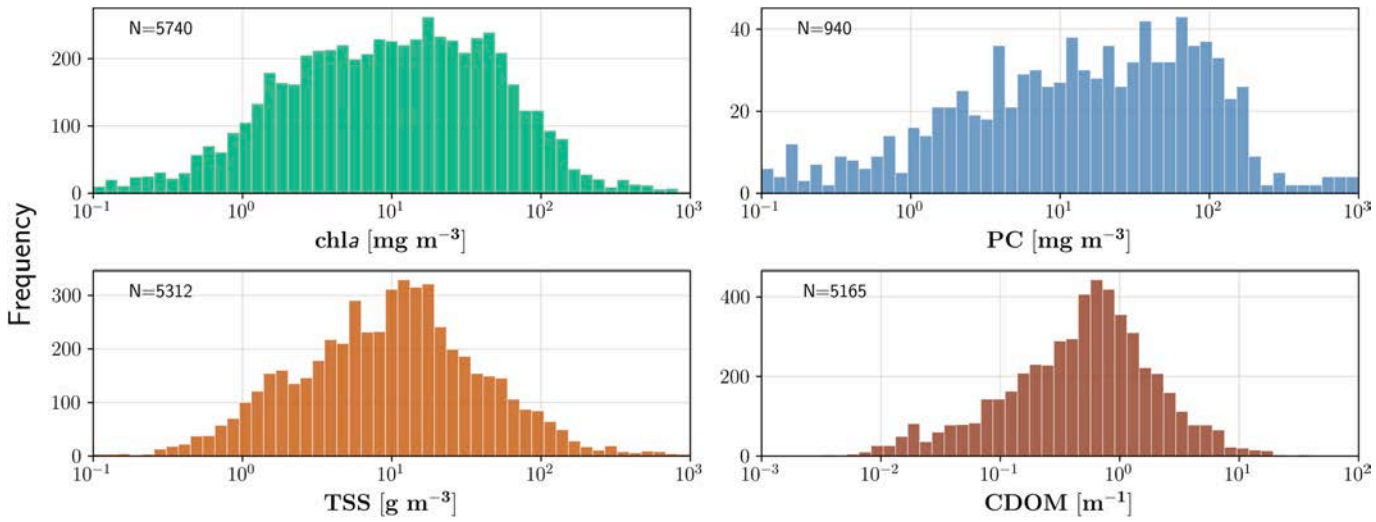

Fig. 2. Histograms of chl*a*, PC, TSS, and CDOM. Total samples (N) shown in top left of each subplot. The mean and median values for chl*a*, PC, TSS, and CDOM are: (41.6 mg/m^3, 10.5 mg/m^3), (52.5 mg/m^3, 15.1 mg/m^3), (30.1 g/m^3, 10.4 g/m^3), and (1.1 m^{-1}, 0.55 m^{-1}), respectively.

estimates for a_{cdom} at 440 nm due to the differences in the provided/measured data: 1) when a_{cdom} was measured/provided with **full hyperspectral information** (**a_{cdom}**, $N = 2764$, an IOP) or 2) when a_{cdom} was measured either at a **single wavelength or pulled from the hyperspectral absorption spectrum** (**CDOM**, $N = 5165$, grouped with BPs). As of publication, there are a small percentage of known duplicates and misaligned measurements within this dataset. While this serves as the final dataset for this manuscript, the model can be rapidly retrained on the evolving dataset as additional sources and quality control metrics are added, to improve model performance.

Our *in situ* dataset (measurements with co-located R_{rs} spanning 409–724 nm, predominantly at 10 nm spectral resolution) covers multiple orders of magnitude of BPs and IOPs (Figs. 2 & 3), representative of a wide range of optically distinct waters. Chl*a*, TSS, and CDOM each have >5000 samples, have a single peak, and are nearly symmetric on a logarithmic scale (Fig. 2). PC has a much more limited dataset (940 samples) and is less symmetric than the other products on the logarithmic scale, due to a high number of samples at higher concentrations (~10^2 mg/m^3). The mean and median values for each biogeochemical parameter are: chl*a* (41.6 mg/m^3, 10.5 mg/m^3), PC (52.5 mg/m^3, 15.1 mg/m^3), TSS (30.1 g/m^3, 10.4 g/m^3), and CDOM (1.1 m^{-1}, 0.55 m^{-1}). The >2200 a_{ph}, a_{nap}, and a_{cdom} measurements (Fig. 3) come from many optically distinct locations and different times. The target wavelengths for each IOP output from the model were constrained (to 409/411, 438/441, 490/485, 530/530, 558/563, 627/632, 650/651, & 673/679 nm for a_{nap} and a_{cdom} for HICO/PRISMA, and to 409–690 nm for a_{ph}) to reduce model complexity.

3.2. Co-located satellite R_{rs}, *in situ* R_{rs}, BPs, and IOPs

The impact of uncertainties in radiometric products (*e.g.*, image-derived R_{rs}) on MDN product estimates can be explored by comparing MDN-derived estimates from co-located *in situ* and satellite R_{rs}. We have assembled a dataset of 65 near-coincident (±3 h) co-located *in situ*, and SeaWiFS Data Analysis System (SeaDAS) corrected HICO R_{rs} from the Chesapeake Bay, Lake Erie, and Florida Estuaries (Casey et al., 2019; Keith et al., 2014; Schaeffer et al., 2015, summarized in Table 2). This dataset also includes co-located *in situ* matchups for some sampling stations, which allow for the direct impact assessment and comparison of uncertainties across product estimates, including spectral a_{ph}, a_{nap}, and a_{cdom}. To test the MDN on these measurements, it was trained without this data subset. However, the final operational MDN was trained on the entire dataset including these *in situ* measurements.

3.3. Hyperspectral satellite imagery: HICO and PRISMA

3.3.1. HICO

HICO imagery (~90 m spatial resolution, ~5.7 nm spectral resolution (Lucke et al., 2011)) of Lake Erie acquired on September 8th 2014 was chosen to demonstrate the spatial consistency of the product maps despite uncertainties in the hyperspectral satellite imagery. This particular image has minimal cloud cover, and five co-located *in situ* chl*a* and PC measurements to confirm the spatial distributions of these pigments. The MDN used to produce these maps was trained using the *in situ* hyperspectral dataset, except for the Lake Erie dataset containing the five co-located *in situ* chl*a*/PC measurements, resampled using the spectral response function of HICO. The HICO imagery was corrected using the Atmospheric Correction for OLI 'lite' (ACOLITE V20221114.0), which produced fewer invalid (~0 or negative) R_{rs} (particularly in the blue bands) within the scene relative to the SeaDAS v7.5.3 level 2 product (not shown). The atmospheric correction flags were adapted to the region, and the processing options (thresholds) used for atmospheric correction are available in Table A1.

3.3.2. PRISMA

PRISMA imagery (~30 m spatial resolution, ~10 nm spectral resolution) of the Curonian Lagoon acquired on September 20th, 2020, provides a second sensor, atmospheric correction algorithm, and region for which to test the MDN. To produce estimates for this region, the MDN was retrained using the full *in situ* dataset, except for the Curonian Lagoon dataset which holds the 10 *in situ* chl*a*, PC, TSS, CDOM, and a_{cdom} sampling stations co-located with the PRISMA imagery. The hyperspectral *in situ* dataset was resampled using the spectral response function of PRISMA, prior to training the MDN. The L1 top-of-atmosphere PRISMA imagery was re-projected using a Bowtie Correction implemented in prismaread (Busetto and Ranghetti, 2020) and atmospherically corrected using the Atmospheric and Topographic Correction (ATCOR v.9.3.0) (Richter and Schläpfer, 2002). The Curonian Lagoon is a shallow (3.8 m average depth) fresh-to-brackish water lagoon (Zemlys et al., 2013), with minimal additional *in situ* data in the training dataset, making it ideal for assessing the performance of the model in optically distinct regions.

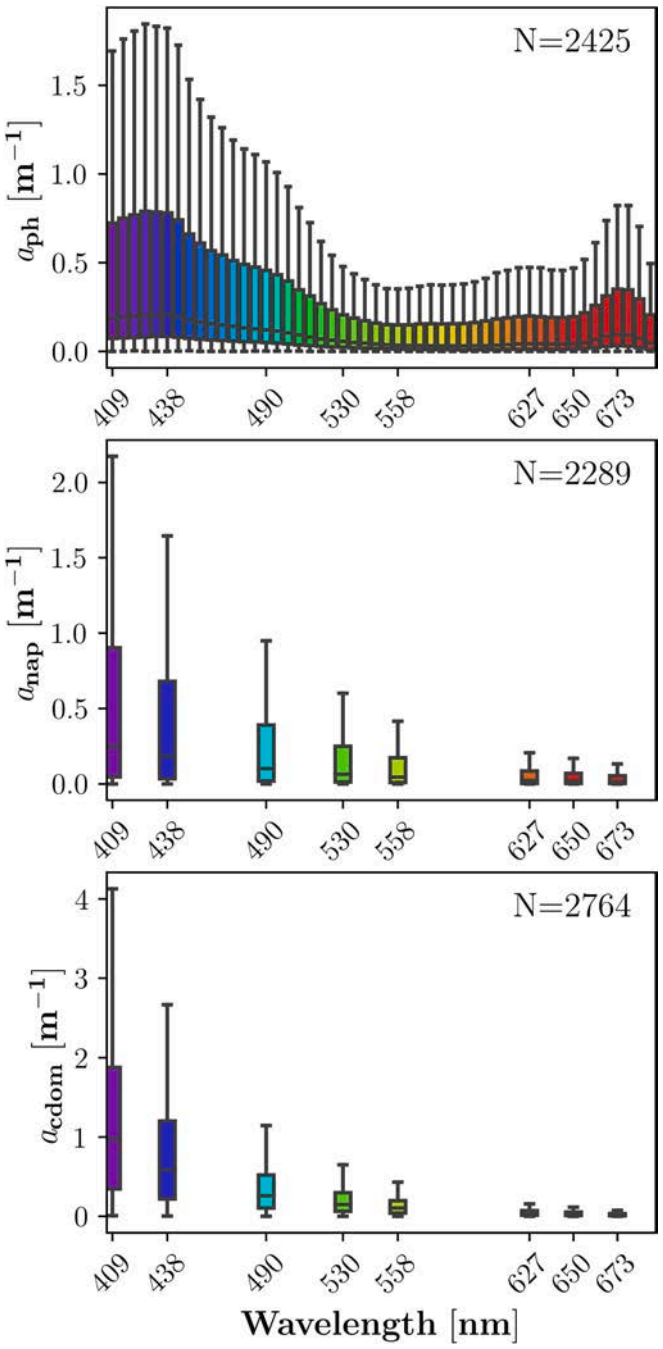

Fig. 3. Box and whisker plots of spectral a_{ph}, a_{nap}, and a_{cdom}. Only select wavelengths, which are used as the output of the MDN, are plotted. Figure differs from data in previous publication (Pahlevan et al., 2021) due to 1) increased dataset size and 2) increased required wavelength range for input R_{rs} (409–724 nm serve as input for our simultaneous estimation MDN).

4. Methods

4.1. MDN architecture, training, and testing

4.1.1. Overarching MDN architecture

The MDN (Fig. 4) follows the standard architecture for the retrieval of individual BPs (Pahlevan et al., 2020; Smith et al., 2021) and simultaneous retrieval of multiple BPs (Pahlevan et al., 2022). The spectral R_{rs} (409–724 nm for HICO, Table 3) serve as inputs to the model. The maximum of the input wavelength range (409–724 nm) was chosen to allow for features in the NIR, useful for PC and chl*a* retrieval (O'Shea et al., 2021; Pahlevan et al., 2022), while avoiding the higher wavelength oxygen-A bands. Nine band ratios (BRs) and line heights (LHs), listed in Table 3, are calculated from the R_{rs}, and also used as inputs to the model. The R_{rs}, BRs and LHs, are then normalized (scaled from −1 to 1) before being run through the weights of the neural network (architecture described in section 4.1.2 and summarized in Table 4). The final layer of the MDN is a mixture of five Gaussians, each with their own mean (μ_n), standard deviation (σ_n), and weighting (α_n). The Gaussians are input to a combination function, which adds the weighted probabilities derived from each of the five Gaussians together and then selects the highest probability estimate for each output parameter. The simultaneous outputs include: chl*a*, TSS, PC, CDOM (a_{cdom} at 440 nm, see section 3.1), a_{ph} from 409 to 690 nm, and a_{cdom} as well as a_{nap} at 409/411, 438/441, 490/485, 530/530, 558/563, 627/632, 650/651, & 673/679 nm for HICO/PRISMA. The output wavelengths were constrained for a_{ph} as measurements of a_{ph} at wavelengths >690 nm have increased noise, which reduces the ability of the model to converge to an optimal solution. The output wavelengths for a_{cdom} and a_{nap} were constrained (to 409/411, 438/441, 490/485, 530/530, 558/563, 627/632, 650/651, & 673/679 nm for HICO/PRISMA) as MDN complexity increases as a function of outputs due to calculation of covariance matrices between each product. Additionally, these eight wavelengths are sufficient, as they can be fit by a decreasing exponential (Eq. 3), and then interpolated, to estimate absorption at alternative wavelengths (Mobley, 2022; Mobley, 1994; Twardowski et al., 2004). Overall, the constrained wavelengths should not limit efficacy as pigment features of a_{ph} largely occur in this spectral range, and by fitting multispectral a_{cdom} and a_{nap} with exponential models, the hyperspectral a_{cdom} or a_{nap} can be interpolated or extrapolated. Training of this model is repeated 10 times, with different initialization weights, and the final estimate is the median of the estimated outputs.

4.1.2. Hyperparameter sweep and neural network architecture

Previous MDN training focused on single BPs (Smith et al., 2021) or simultaneous retrieval of only three BPs (Pahlevan et al., 2022). In each case, the model successfully converged using the same set of previously proven hyperparameters. However, training an MDN to simultaneously model ~55–70 outputs from multiple BPs and IOPs with the same hyperparameters led to poor convergence for a_{ph} (results not shown). To identify more viable hyperparameters, the MDN was retrained on the same training dataset with 375 different hyperparameter combinations as shown in Table 4, but only for a single training round. The hyperparameters of the model with the lowest median of the ensemble median symmetric accuracy (MdSA, 30.45%, hyperparameters in Table 4, 'Best' column) over all output products (with each spectral product being represented by a median value taken over all bands) was chosen for final training of the 10 round model in the neural network block (Fig. 4) presented in this study.

4.1.3. Imputation

Not all BPs and IOPs are simultaneously measured *in situ* at each site. In our framework, a process known as multiple imputation fills in for the missing values (Rubin, 2004), which has been successfully proven to work for the simultaneous estimation of three BPs from multispectral satellite imagery (Pahlevan et al., 2022). In multiple imputation, *m* values are randomly selected from the distribution of the input dataset for the missing parameter. Instead of naively drawing from the distribution of the input dataset, MDNs allow us to learn the joint probability distribution of all target parameters and draw from the learned posterior probability (Fig. 4). By drawing estimates from the learned posterior instead of the distribution of the input dataset, multiple imputation can not only represent missing completely at random and missing at random data (King et al., 2001), but in some instances, also missing not at random data, where the missing data depends upon its value (van Buuren and Groothuis-Oudshoorn, 2011; Galimard et al., 2018). This may be useful in monitoring cyanobacterial blooms, where practical

Table 2
Matchup datasets for HICO and PRISMA imagery.

Date (M/D/Y)	Location	Sensor	Atmospheric Correction	*In-situ* Matchups
09/08/2014	Lake Erie	HICO	ACOLITE	5 chl*a* & PC
09/20/2020	Curonian Lagoon	PRISMA	ATCOR	10 chl*a*, PC, TSS, CDOM, & a_{cdom}
04/14/2010, 06/02/2011, 07/30/2011, 08/24/2011, 08/26/2011, 09/09/2011, 08/28/2014, 09/08/2014	Chesapeake Bay, Lake Erie, & Florida Estuaries	HICO	SeaDAS	29 a_{ph}, 23 a_{nap}, & 23 a_{cdom} + 65 *in-situ* R_{rs}

limitations typically lead to sampling only in regions prone to these blooms, and only during peak bloom events for water quality monitoring decision making (Fig. 2, left-skewed distribution). The accuracy of the multiple imputation method improves with additional training, as the MDN better learns to represent the joint probability distribution of the target parameters. In summary, multiple imputation with MDNs more accurately fills in missing data, leading to a larger available dataset for training, and, in extension, potentially improved generalization.

4.1.4. Calculation of S-CDOM and S-NAP from spectral a_{cdom} and a_{nap}

S-CDOM and S-NAP can both be calculated by fitting a decreasing exponential (Eq. 3) to a_{cdom} and a_{nap} at the eight predicted wavelengths

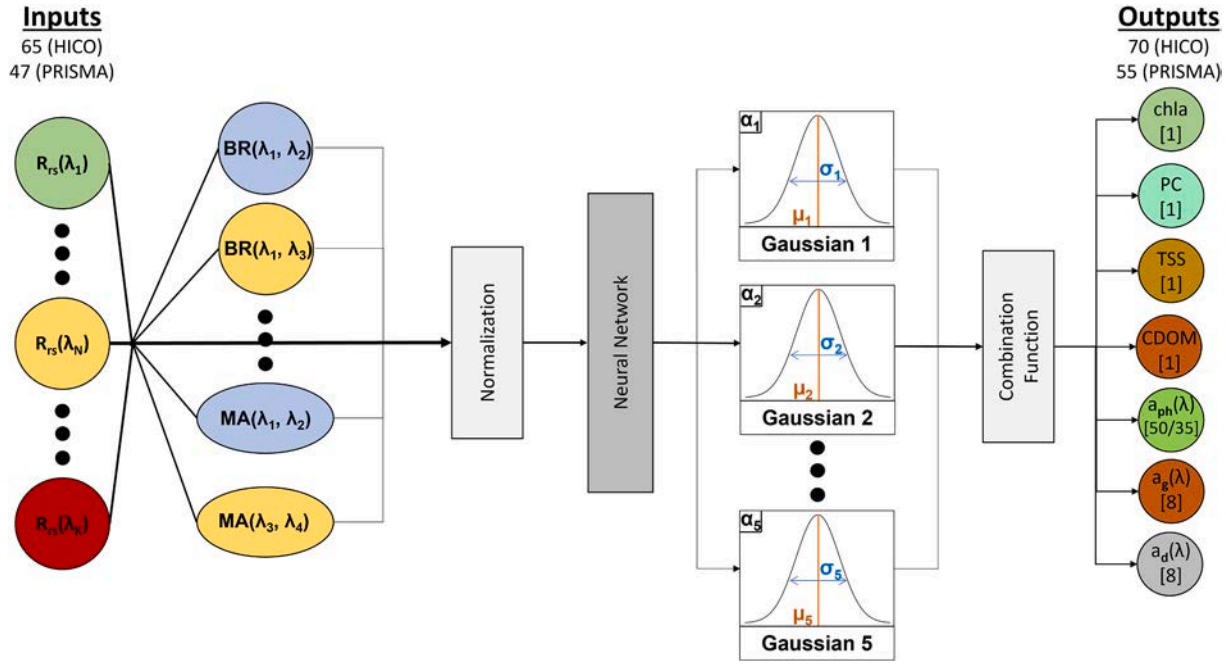

Fig. 4. A block diagram representing the MDN architecture. HICO/PRISMA spectral R_{rs} (with exact bands dependent on sensor) are inputs (features described in Table 3) to the MDN, in addition to standard band ratios (BRs) and line heights (LHs). The inputs are normalized and then run through the neural network's weights. The unique feature of MDNs is their output layer, where a mixture of Gaussians, with mean (μ_n), standard deviation (σ_n), and weighting (α_n) is formed to represent the output variable's probability distribution. A combination function selects the most probable value as an estimate of the output BP/IOP. The number of outputs is shown in brackets for each output parameter for HICO/PRISMA.

Table 3
Itemized input and output features for HICO MDN and PRISMA MDN. A limited set of a_{cdom} and a_{nap} wavelengths are output (choosing the sensor band nearest to the ac-9 (Sea-Bird Scientific's transmissometer) equivalent wavelengths) to reduce the required model complexity (Section 3.1).

Input or output?	Feature labels	HICO Features (Per feature (total))	PRISMA Features (Per feature (total))
Input	R_{rs}	409–724 (56)	411–719 (38)
Input	BRs	'667/707', '558/707', '558/667', '490/558', '444/490', '667/621', '621/558' (7)	'660/709', '563/709', '563/660', '493/563', '441/493', '660/623', '623/563', (7)
Input	LHs	'490–558–667', '558–621–667' (2)	'493–563–660', '563–623–660' (2)
Total Inputs		**65**	**47**
Output	BPs (chl*a*, PC, TSS, CDOM)	1 (4)	1 (4)
Output	a_{cdom} & a_{nap}	409, 438, 490, 530, 558, 627, 650, & 673 nm (16)	411, 441, 485, 530, 563, 632, 651, & 679 nm (16)
Output	a_{ph}	409–690 (50)	411–689 (35)
Total Outputs		**70**	**55**

Table 4
Hyperparameters, range of covered values in hyperparameter search, and hyperparameters of the model with the lowest single round median MdSA. Training uncertainty as a function of parameters is covered in Fig. B1.

Hyperparameter	Pahlevan et al., 2022 Standard Values	Range (steps)	Best
Iterations	10,000	10^3–10^5 (5)	31,622
Nodes	100	10^2–$2*10^3$ (5)	446
Layers	5	5–12 (5)	5
Learning Rate	0.001	0.001	0.001
Regularization (l2)	0.001	0.001	0.001
Epsilon	0.001	10^{-3}–10^{-2} (3)	0.001
Rounds	10	1	1

(409/411, 438/441, 490/485, 530/530, 558/563, 627/632, 650/651, & 673/679 nm for HICO/PRISMA) from the MDN (Twardowski et al., 2004). The output wavelengths for a_{cdom} and a_{nap} were chosen to stay within the ~400–650 nm range to 1) facilitate comparisons to computed slopes from an ac-9 instrument, 2) avoid artifacts in NAP at longer wavelengths, and 3) avoid longer wavelengths where scattering uncertainty may dominate the signal, as the model may not converge with such a low SNR target variable.

4.2. MDN performance assessment

An ideal performance assessment would include a comparison between satellite-derived products and co-located *in situ* measurements over a globally representative dataset of optically distinct water bodies. Unfortunately, the practical limitations of *in situ* sampling and demonstration space-borne mission overpass times make this currently infeasible. Instead, we rely on multiple different performance metrics to gain an understanding of overall performance: these include a 50/50 training/testing split, an iterative leave-one-out cross-validation assessment, a retrieved product comparison from co-located *in situ* R_{rs} and satellite R_{rs}, and, finally, a matchup analysis for a small dataset of *in situ* measurements and products derived from co-located satellite imagery, described below.

4.2.1. 50% training and 50% testing (50/50) split

The first performance metric to characterize idealized model accuracy is a 50% training and 50% testing split of our *in situ* dataset. In this assessment, the model is trained using a random half of the *in situ* dataset, and model performance is assessed on the test half of this dataset and further compared against operational algorithms for each of the BPs and IOPs. This performance assessment provides an idealized performance accuracy as 1) the test data may come from the same field campaigns (not different regions/seasons), 2) the MDN is trained on half of the data while the operational algorithms are not optimized to the training set (though minimal gains were found by optimizing these multispectral algorithms (O'Shea et al., 2021)), and 3) the radiometric test data (*in situ* R_{rs}) do not suffer from atmospheric correction uncertainties typical of satellite imagery. Overall, this performance assessment is useful to determine if the training dataset is sufficiently large to represent the solution space by estimating the generalization performance on the wide range of optical water quality conditions covered within the dataset. However, this performance assessment does not apply to model performance on out-of-training set hyperspectral satellite imagery.

4.2.2. Iterative leave-one-out cross-validation assessment

To better characterize the performance of the final model on out-of-training set regions, we retrain the model by leaving out data from one region (one or more datasets). We can then better compare the performance accuracy of the MDN against un-optimized operational algorithms, and determine the estimated uncertainty for a range of optically distinct water bodies. By leaving out entire datasets, this performance assessment accounts for the impact of uncertainties due to differences in sampling methods and instrumentation used by any individual lab on the product estimation accuracy. Some complicating factors to this method are that not all datasets are from single regions and some datasets have a disproportionately high number of overall samples for specific BPs or IOPs. We evaluated most datasets by region (not all samples include latitude and longitude), but some datasets span multiple regions (*e.g.*, the "Global Waters" dataset extracted *via* SeaBASS in Table 1). This offered the additional benefit of testing our algorithm with data collected by different scientists, as datasets are entirely contained. Overall, the leave-one-out cross-validation assessment is a more accurate assessment for the model's generalization performance on previously unseen *in situ* data from global inland and coastal waters analyzed at different labs, but it does not account for atmospheric correction uncertainties common to hyperspectral satellite imagery which will impact the satellite derived products (Ibrahim et al., 2018).

4.2.3. Product comparison from co-located in situ R_{rs} and hyperspectral satellite R_{rs}

One way to estimate the impact of atmospheric correction uncertainties (Ibrahim et al., 2018) on the estimated BPs and IOPs is to compare products derived from *in situ* R_{rs} (X^e) to products derived from co-located satellite-derived R_{rs} (X^r). We have a dataset of 65 *in situ* R_{rs} from Pensacola Bay, Florida, U.S.A., with co-located and near-coincident (+/− 3 h) hyperspectral R_{rs} from HICO (Keith et al., 2014), used to assess product performance in previous publications (O'Shea et al., 2021). The nearest pixel was selected and the R_{rs} has been atmospherically corrected using SeaDAS. Such performance assessment also helps to determine the sensitivity of MDNs to uncertainties in R_{rs}, which are unavoidable due to atmospheric correction and instrument artifacts. In summary, while this method assesses the relative bias and uncertainty from products derived *via in situ* vs. satellite R_{rs} in a single region, it does not offer absolute metrics of uncertainty, as it is not a direct comparison to *in situ* measured BPs and IOPs, and does not generalize to additional regions or atmospheric correction approaches where uncertainties may differ.

4.2.4. Matchup analysis: In situ biogeochemical parameter and IOP comparison to satellite products

The best estimate of MDN product uncertainty is a direct comparison of satellite-derived products to *in situ* measurements. HICO imagery co-located with *in situ* measurements of chl*a*, PC, a_{ph}, a_{nap}, and a_{cdom} are available from the previously described Pensacola Bay dataset (Keith et al., 2014). Additionally, PRISMA imagery co-located with *in situ* chl*a*, PC, TSS, CDOM, and a_{cdom} are available from the Curonian Lagoon. Finally, chl*a* and PC measurements are available *via* co-located matchups with HICO imagery of Lake Erie (Casey et al., 2019; Schaeffer et al., 2015). Although direct matchup analysis is ideal, practical limitations of available samples limit assessments over a wide range of optically distinct water bodies.

4.3. Uncertainty metrics

Common uncertainty metrics such as mean absolute percentage difference (MAPD) have issues such as: being undefined at a true value of 0, asymmetry with respect to overestimation and underestimation of

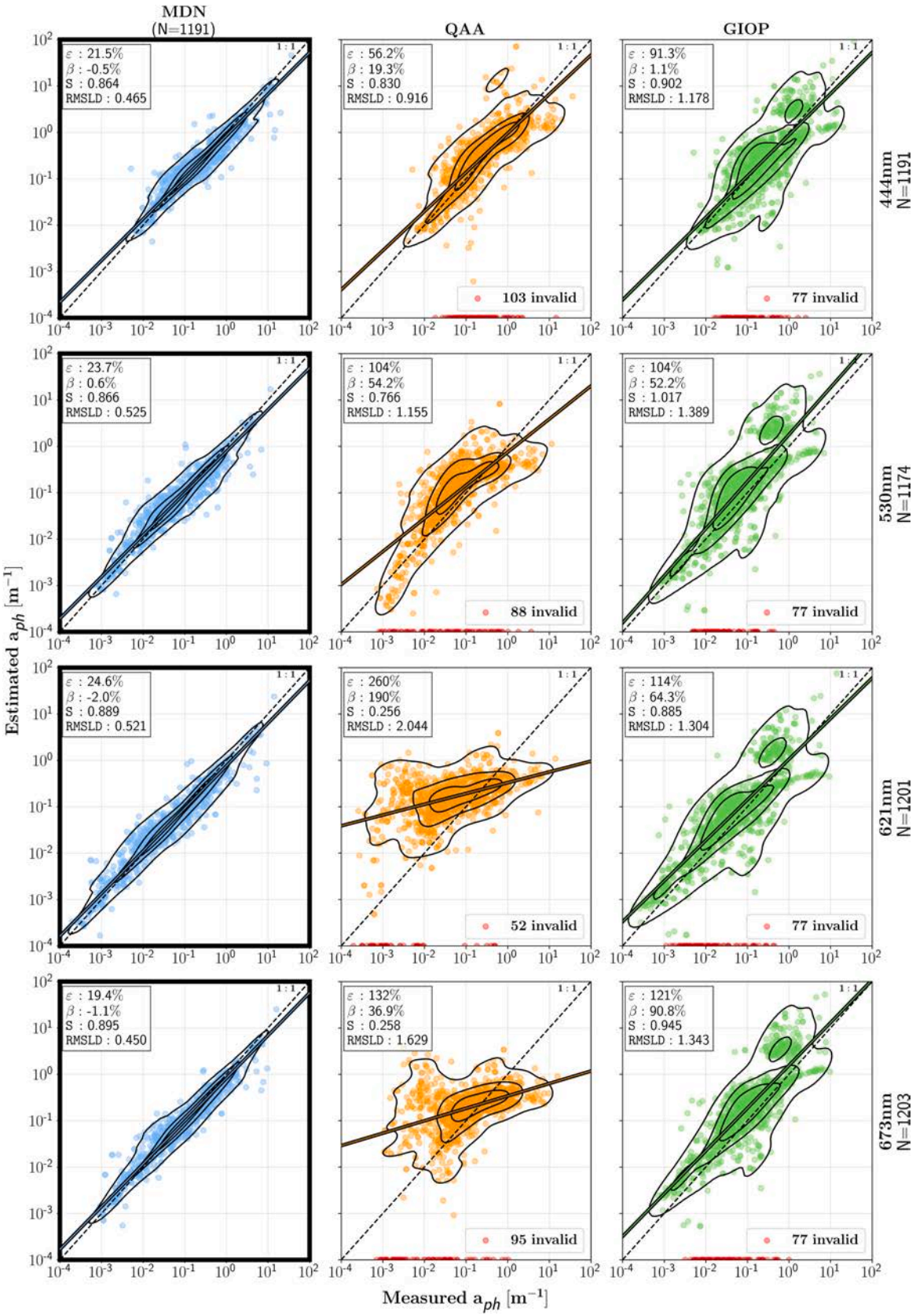

Fig. 5. Measured and estimated $a_{ph}(\lambda)$ from the test half of the 50/50 split at key spectral wavelengths for pigment analysis. A similar figure for a_{ph} retrievals from an MDN dedicated to a_{ph} retrieval can be found in (Pahlevan et al., 2021) for comparison. Additional BPs & IOPs can be found in the appendix, in Figs. C1–6. Contour lines demarcate equal proportions of the probability mass, to enhance the visualization of the data distributions.

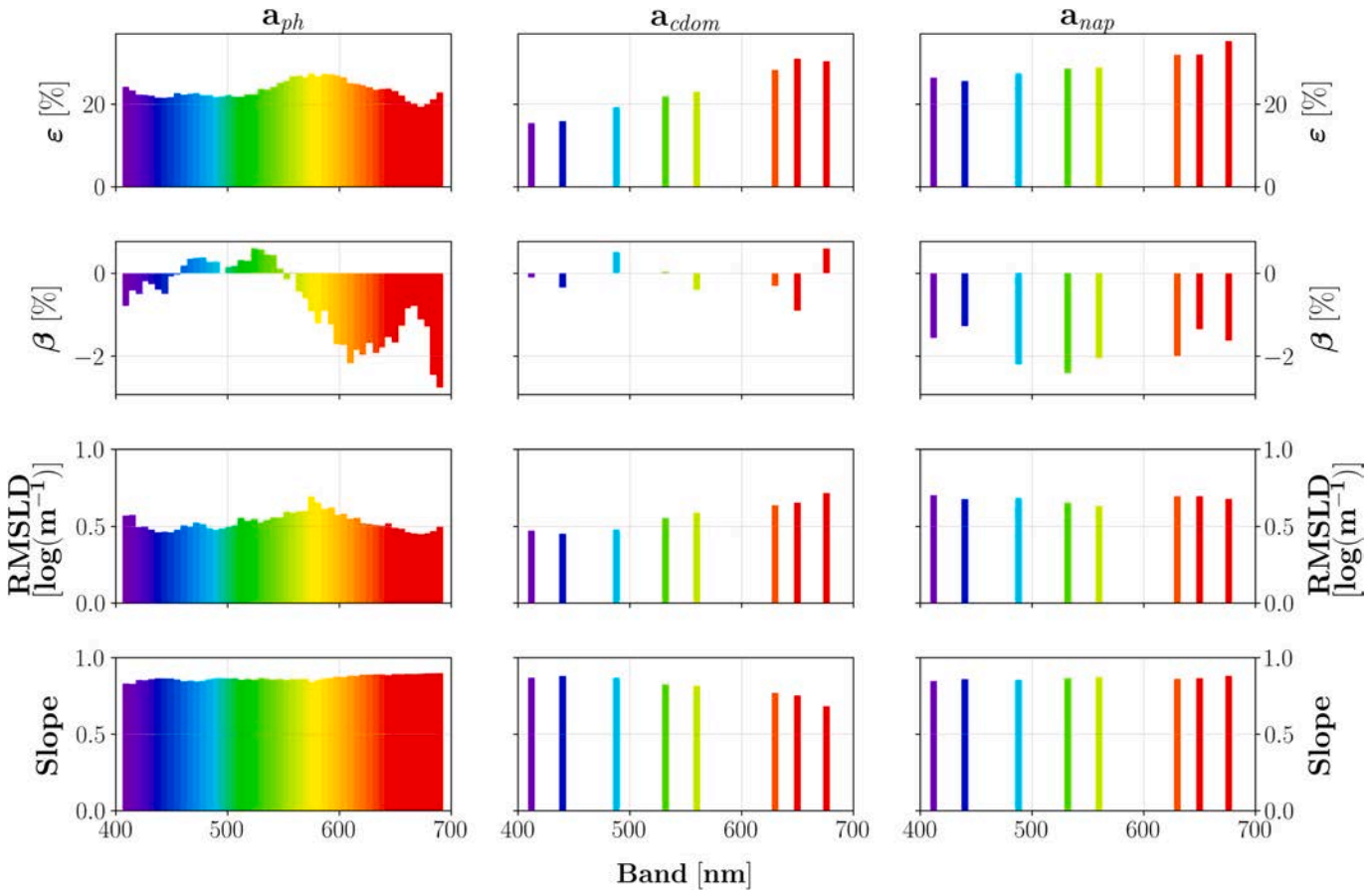

Fig. 6. ε(λ), β(λ), RMSLD, and slope for $a_{ph}(\lambda)$, $a_{nap}(\lambda)$, and $a_{cdom}(\lambda)$ from the test half of the *in situ* dataset. A graphic demonstrating similar uncertainties can be found for a_{ph} retrievals from a dedicated model in (Pahlevan et al., 2021).

target values, positive skew, and a lack of robustness to outliers (Morley et al., 2018). Instead of these common metrics, we leverage median symmetric accuracy (ε) and signed symmetric bias (β) for uncertainty quantification (Eqns. 6–8 Morley et al., 2018). ε and β are ideal uncertainty metrics because they are meaningful over multiple orders of magnitude (*i.e.*, they use relative uncertainty, Eq. 6), symmetric (underprediction and overprediction are penalized equally), robust to outliers (median, Eqns. 7 & 8), and easily interpretable. For completeness, and to better understand noise performance, the slope of the linear least-squares regression (in log space), the Root Mean Squared Log Difference (RMSLD, Eq. 9), and coefficient of determination (R^2) are also reported.

$$Q_i = y_i/x_i \tag{6}$$

$$\varepsilon = 100^{*}\left(e^{median(|log_e(Q_i)|)} - 1\right) \tag{7}$$

$$\beta = 100^{*}sgn(\text{median}(log_e(Q_i)))^{*}\left(e^{|median(log_e(Q_i))|} - 1\right) \tag{8}$$

where Q_i is the accuracy ratio, with y_i and x_i representing the predictions and *in situ* observations, respectively.

$$\text{RMSLD} = \frac{1}{n}\sum_{i=1}^{n}(log(y_i+1) - log(x_i+1))^2 \tag{9}$$

5. Results

5.1. Performance evaluation: 50/50 training/testing split and leave-one-out cross-validation

The MDN architecture achieves the lowest ε and β relative to all (non-optimized) operational algorithms, for all parameters on the test half of the 50/50 training/testing *in situ* dataset split (Fig. 5 for a_{ph}, see Figs. C1–6 for BPs, a_{cdom}, and a_{nap}). In particular, MDN estimates of a_{ph} at four key wavelengths for a_{ph} retrieval are more accurate across the full 4–5 orders of magnitude covered by the *in situ* measurements (Fig. 5). Also, the MDN has higher linearity and a lower number of outliers in the highest phytoplankton absorbing waters, characteristic of blooms, relative to the two (unoptimized) operational algorithms, QAA and GIOP. In addition to having fewer outliers than these algorithms, the MDN also produced no invalid estimates.

The spectral ε for the IOPs calculated from the test half of the *in situ* dataset (Fig. 6) are generally proportional to the *in situ* absorption magnitude (Fig. 3). For a_{ph}, ε generally stays within the ~20–30% bounds. The highest ε in a_{ph} retrievals occurs in the 550–600 nm range, where a_{ph} is typically lowest (Fig. 3), and the lowest ε in a_{ph} retrievals occurs in the 400–480 nm and 650–700 nm ranges, where a_{ph} is generally highest. Spectral ε of a_{cdom} and a_{nap} range from ~15–40%, are lowest in the blue and highest in the red (Fig. 6), and follow an inverse relationship with the spectral magnitude of a_{cdom} and a_{nap} (Fig. 3) due to the relative nature of the metric. Spectral β in the calculated products does not seem to have a consistent pattern but only minimally varies from ~1 to −3% for a_{ph}, a_{cdom}, and a_{nap} (Fig. 6). Overall, *in situ* retrievals from areas represented in the test half of the *in situ* dataset should have minimal spectral bias and differences in their uncertainty, but these results do not consider uncertainties from atmospheric correction or instrument noise within hyperspectral satellite imagery.

The simultaneous estimation MDN performs better than, or equivalent to, unoptimized operational retrieval algorithms for estimating all products (except for PC) from independent datasets (blue dashed lines, Fig. 7) from the leave-one-out cross-validation analysis. MDN retrieval

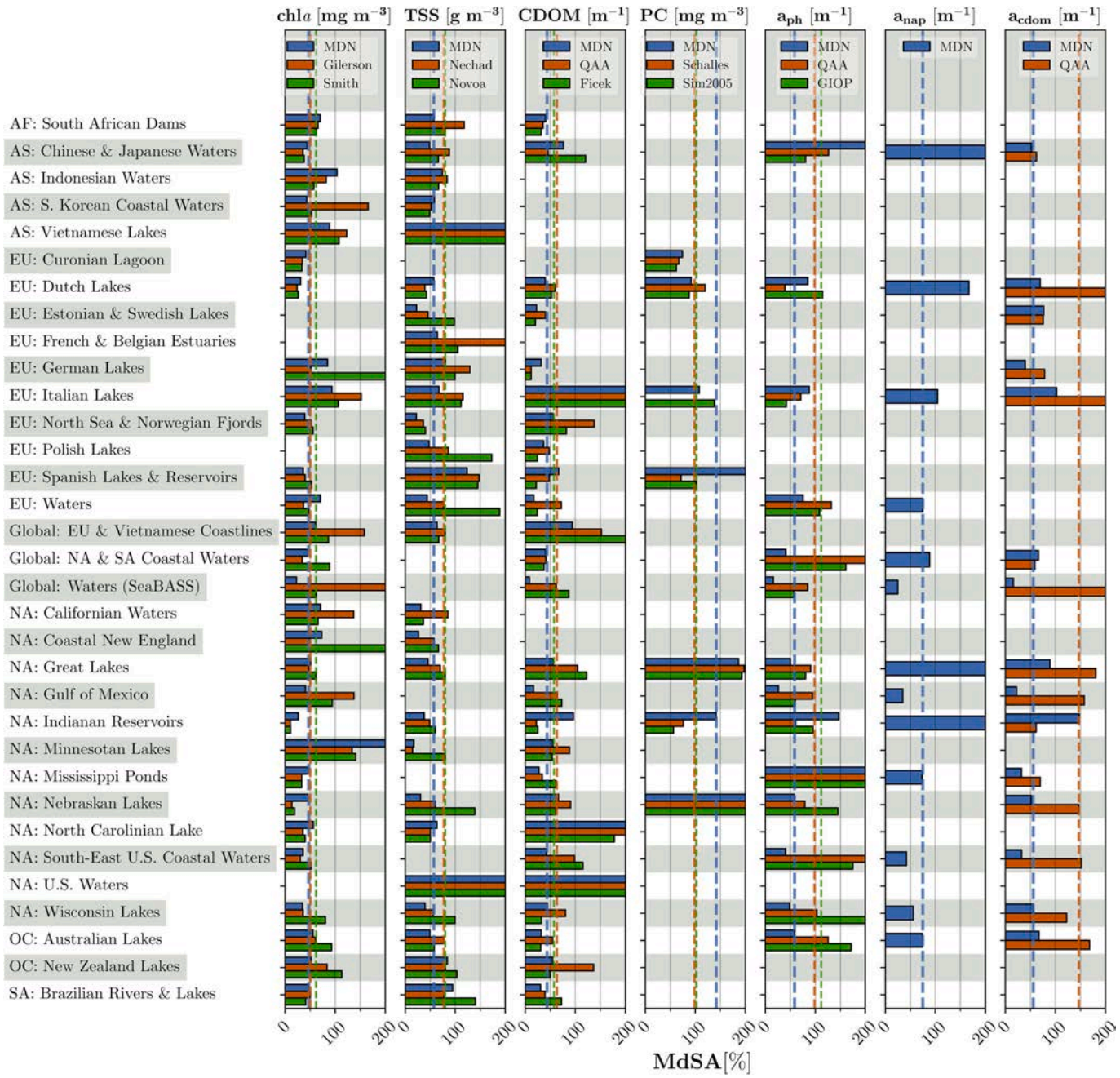

Fig. 7. Leave-one-out cross-validation results for the MDN using the hyperparameters resulting in the lowest median MdSA in the hyperparameter sweep (Table 4), but for 10 full rounds of training. Similar datasets and regions were combined to represent individual locations (*e.g.*, Erie, Taihu), though some conglomerate datasets (*e.g.*, SeaBASS) may contain samples from multiple regions. Dashed lines represent the median regional uncertainty (calculated across regions). Spectral IOPs are reported on the median over all retrieved wavelengths. Acronyms: AF - Africa, AS - Asia, EU - Europe, NA - North America, OC - Oceania, and SA - South America.

of PC seemingly offers the least benefit, however, the Schalles algorithm (Schalles and Yacobi, 2000) estimates are invalid for a substantial proportion (202/471 estimates from the test half of the 50/50 benchmark split, Fig. C4) of the dataset (notably in the lower concentration range). The datasets for which the MDN performs poorest either have a large composition of the overall dataset (NA: Great Lakes, EU: Spanish Lakes & Reservoirs) or the highest PC concentrations (NA: Nebraska). Another product having one of the lowest benefits from using the simultaneous MDN is chl*a*, where the Gilerson algorithm (Gilerson et al., 2010) performs only slightly worse than the simultaneous MDN (blue dashed line, leftmost column, Fig. 7). For the other products, TSS, CDOM, a_{ph}, a_{nap}, and a_{cdom}, the MDN provides lower uncertainty in its results than the presented alternative algorithms, though, as expected, specific datasets with unique optical properties still prove difficult for the MDN, for some products (*e.g.*, 'NA: Minnesotan Lakes' chl*a* and 'AS: Vietnamese Lakes' TSS).

5.2. *Impacts of uncertainties in satellite R_{rs} on MDN retrievals*

Comparison of the products calculated from *in situ* R_{rs} and those from co-located SeaDAS processed HICO imagery demonstrates the substantial impact of uncertainties in satellite R_{rs} on the retrieval of BPs and IOPs (Fig. 8). The remotely estimated BPs and IOPs exhibit ε (relative to values estimated from *in situ* R_{rs}) ranging from 37.4 to 62.8%, due to uncertainties in the satellite-derived R_{rs}. All products generally have a high slope (0.660–0.949) and the IOPs have high R^2 (0.603–0.905),

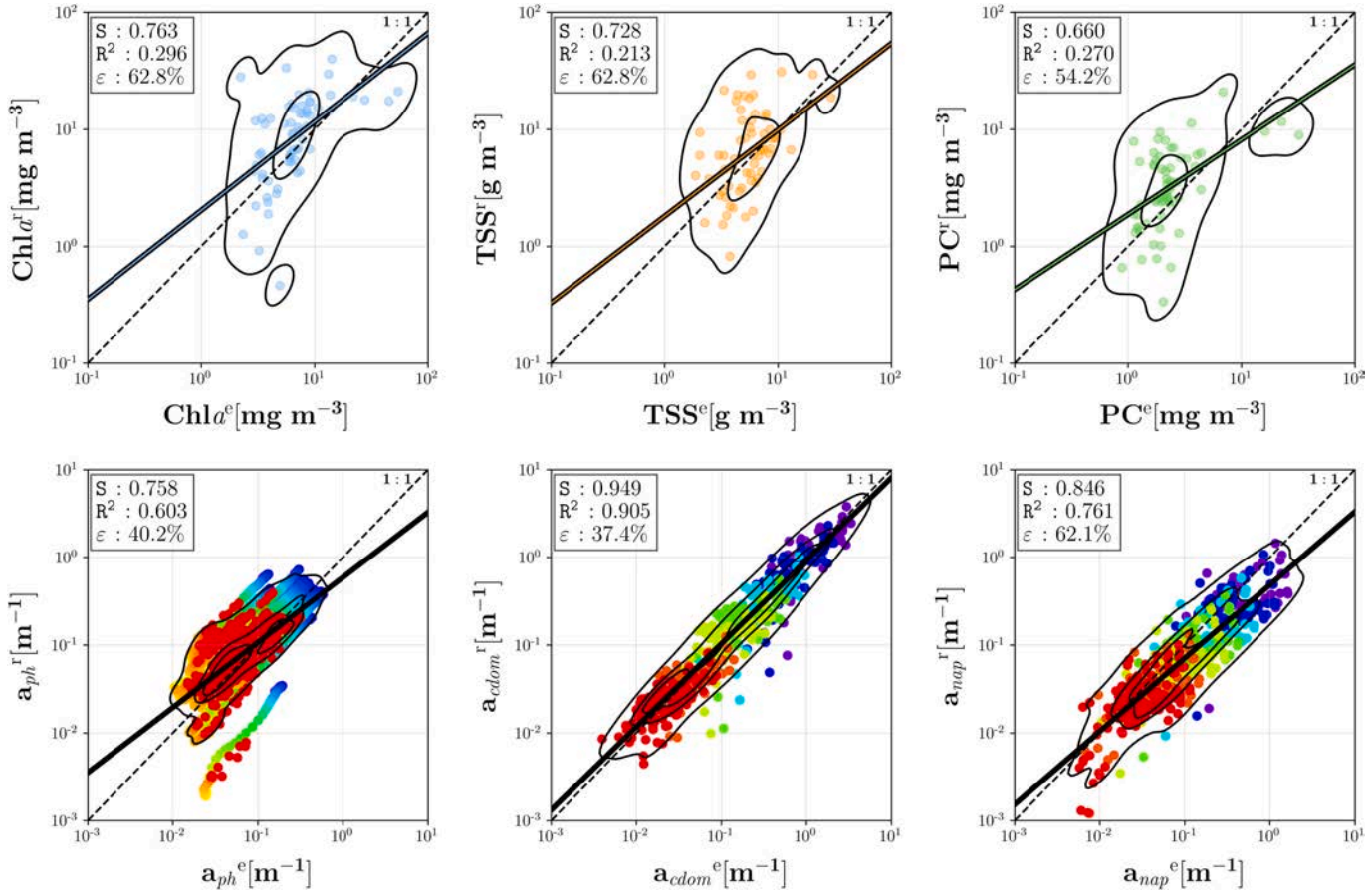

Fig. 8. BPs and absorbing IOPs calculated from 65 co-located remotely sensed (SeaDAS corrected HICO imagery, X^r [*e.g.*, aph^r]) and *in situ* (X^e [*e.g.*, aph^e]) R_{rs} (Section 3.2). IOP scatterplots include all retrieved wavelengths (data color coded by respective wavelength). Wavelengths for absorbing IOPs spanned 409–690 nm for a_{ph} and 409, 438, 490, 530, 558, 627, 650, & 673 nm for a_{cdom} and a_{nap} (as described in Section 4.1.1). Contour lines demarcate equal proportions of the probability mass, to enhance the visualization of the data distributions.

however the BPs have much lower R^2 (0.213–0.296). a_{ph} exhibits minimal sensitivity, except for some larger outliers (Fig. 8). Uncertainties in R_{rs} only minimally impact a_{cdom} and a_{nap}, which have a low uncertainty (37.4%, 62.1%), high slope (0.949, 0.846), and high R^2 (0.905, 0.761). The trends for all three IOPs hold across magnitude and wavelength (Fig. 8). Unfortunately, these results are limited to matchups from a few days in a single region/optical scenario, represent R_{rs} over very different spatial scales, and only report a relative uncertainty, but do not demonstrate absolute model accuracy through comparison to the actual *in situ* measurements and do not necessarily represent the impact of uncertainties from other atmospheric correction techniques (*e.g.*, ACOLITE).

Direct comparison between the *in situ* measurements and *in situ*- and remotely-derived products demonstrates an increased uncertainty in the remotely-derived products due to propagation of uncertainties from the atmospheric correction process to the product retrievals (Figs. 9 & 10). The R_{rs} are dominantly impacted by offsets and physically implausible results in the blue section of the spectrum (Ibrahim et al., 2018; Keith et al., 2014). In general, a_{ph}, is very well estimated by the *in situ* R_{rs}, capturing both the larger spectral shape and features at 425, 620, and 673 nm, though there is a notable offset in the magnitude within the 400–500 nm range in most retrievals. A very large offset in the magnitude of the R_{rs} combined with a physically unrealistic shape (a much higher $R_{rs}(410)$ than $R_{rs}(440)$) leads to a massive underestimation of a_{ph}, a_{nap}, and a_{cdom} (Fig. 10, PB08). Otherwise, uncertainties in R_{rs} lead to either overestimation of a_{ph} across the spectrum (Fig. 9, PB05 and PB04), underestimation of a_{ph} in the blue (Fig. 9, PB09), or a generally poor retrieval of the overall spectral shape of a_{ph} (Fig. 9, SA11). The spectral shapes of a_{nap} and a_{cdom} are much less sensitive to uncertainties in R_{rs}, generally staying within ~30% of the *in situ* measured absorption (or *in situ*-derived absorption, from R_{rs} input to the MDN, in the absence of an *in situ* measured absorption). The magnitude for both a_{cdom} and a_{nap} is much more heavily affected by the uncertainties, with three stations yielding derived magnitudes with uncertainties above 50% relative to the *in situ* measured absorption (Fig. 9, PB05 & PB09 and Fig. 10 PB08). Using the MDN-derived CDOM absorption in place of $a_{cdom}(440)$ performs as well or slightly worse in all cases (Figs. 9 & 10), despite the higher number of available samples (Figs. 2 & 3). In summary, the spectral shape of a_{ph} is the most sensitive to uncertainties in R_{rs}, while the spectral slopes of a_{cdom} and a_{nap} are relatively insensitive.

5.3. Product map assessment: HICO imagery of Lake Erie

MDN-derived BP product maps of Lake Erie from HICO imagery are spatially consistent and agree with our historical understanding of the region despite uncertainties in R_{rs} associated (mainly) with atmospheric correction (Figs. 11-13, D1–6). The Detroit River plume (DRP, northwestern section of Lake Erie), fed by the Detroit River from Lake St. Claire, historically has low chl*a* and PC concentrations (Binding et al., 2019; Moore et al., 2017). The MDN-derived product maps (Fig. 11) align very well with the literature measured *in situ* values (Table 5) for the DRP, though PC may be slightly overestimated (which is expected as much of the Lake Erie data, which contains a large percentage of our PC samples, particularly those with low concentrations, was left out of training). The product maps (Fig. 11) also agree quite well with literature measured values from prior years (Table 5) for Maumee Bay (MB)

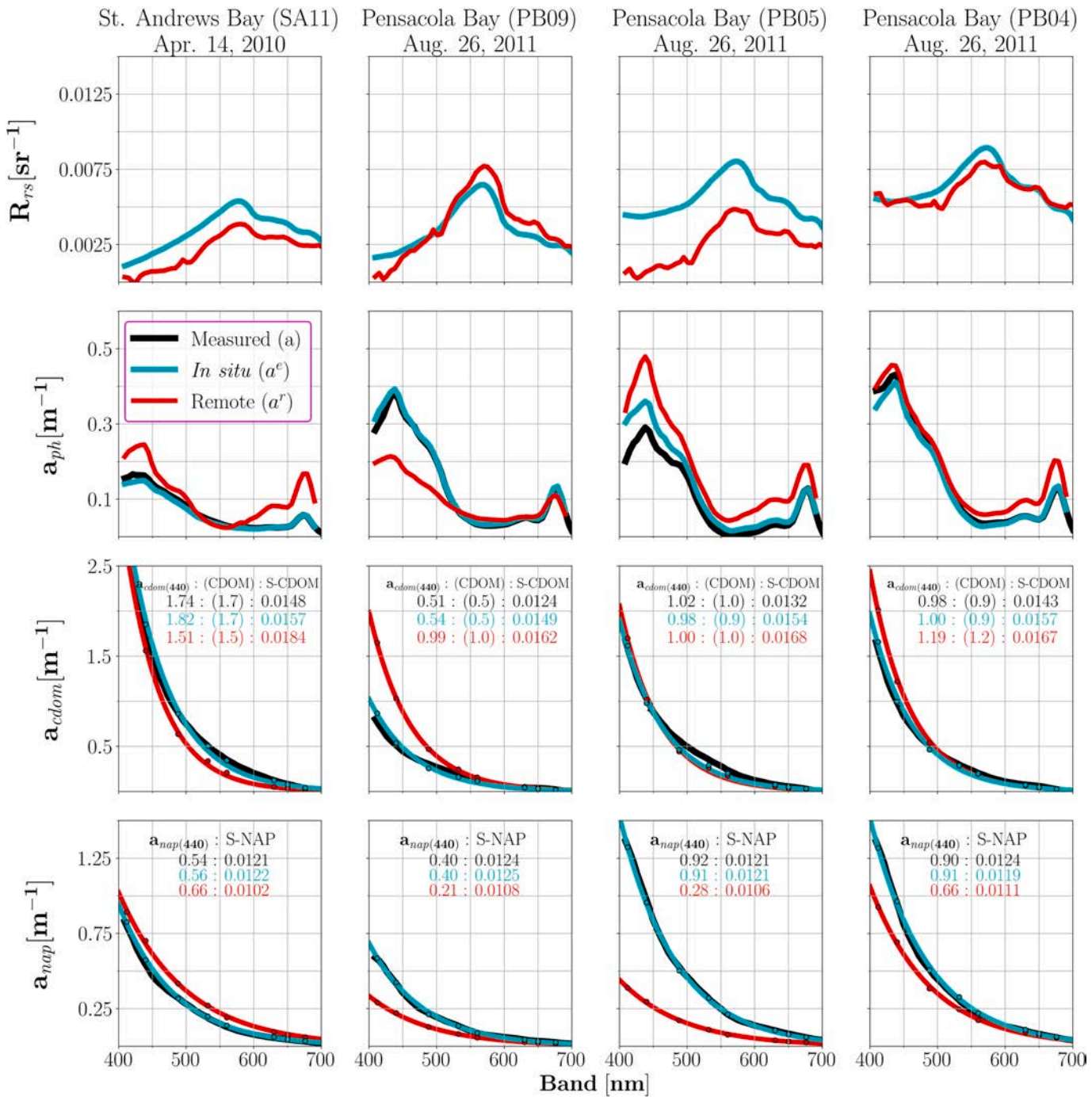

Fig. 9. *In situ* (cyan) and co-located HICO R_{rs} (red) spectra from select stations within Pensacola Bay (top row, matching row 2 from Fig. 7 of Pahlevan et al., 2021). The MDN-derived estimates of a_{ph}, a_{nap}, and a_{cdom} from the *in situ* (cyan) and co-located HICO R_{rs} (red) are compared against the *in situ* measured (black) absorptions. Text on the a_{nap} and a_{cdom} plots compares between the point and slope (derived from Eq. 3), as well as the MDN-estimated CDOM value ($a_{cdom}(440)$, but with more samples available for training than the multispectral curves). The calculated point and slope are used to generate the *in situ* and remote curves for a_{nap} and a_{cdom} with eq. 3. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

and Central Basin (CB).

While the products within Maumee Bay do not align perfectly with the historical measurements of the region, they do match the spatial distribution of the *in situ* observations quite well (black circles, Fig. 11). In addition to generally agreeing with historical measurements of the region and the *in situ* values in Maumee Bay, the product maps for chl*a* and PC also agree well with maps from dedicated MDNs (O'Shea et al., 2021, Fig. 8; Pahlevan et al., 2021, Fig. 9). Overall, the BP product maps align quite well with our co-located *in situ* measurements and historical understanding of the spatial distribution of these constituents, though, as may be expected, temporal and spatial fluctuations in bloom magnitude combined with uncertainties related to remote estimation lead to imperfect agreement.

IOP product maps produced from the hyperspectral inversion framework also agree with the historical understanding of the region (Figs. 11. & D1–3). Specifically, all three IOPs, a_{ph}, a_{nap}, and, a_{cdom}, are lowest in the DRP and CB of Lake Erie, and highest in SB and MB. The relative values align quite well with both the literature's reported spatial distributions in this region and with the historical absolute values (Moore et al., 2017). Of note, the lower concentrations of chl*a* and PC in

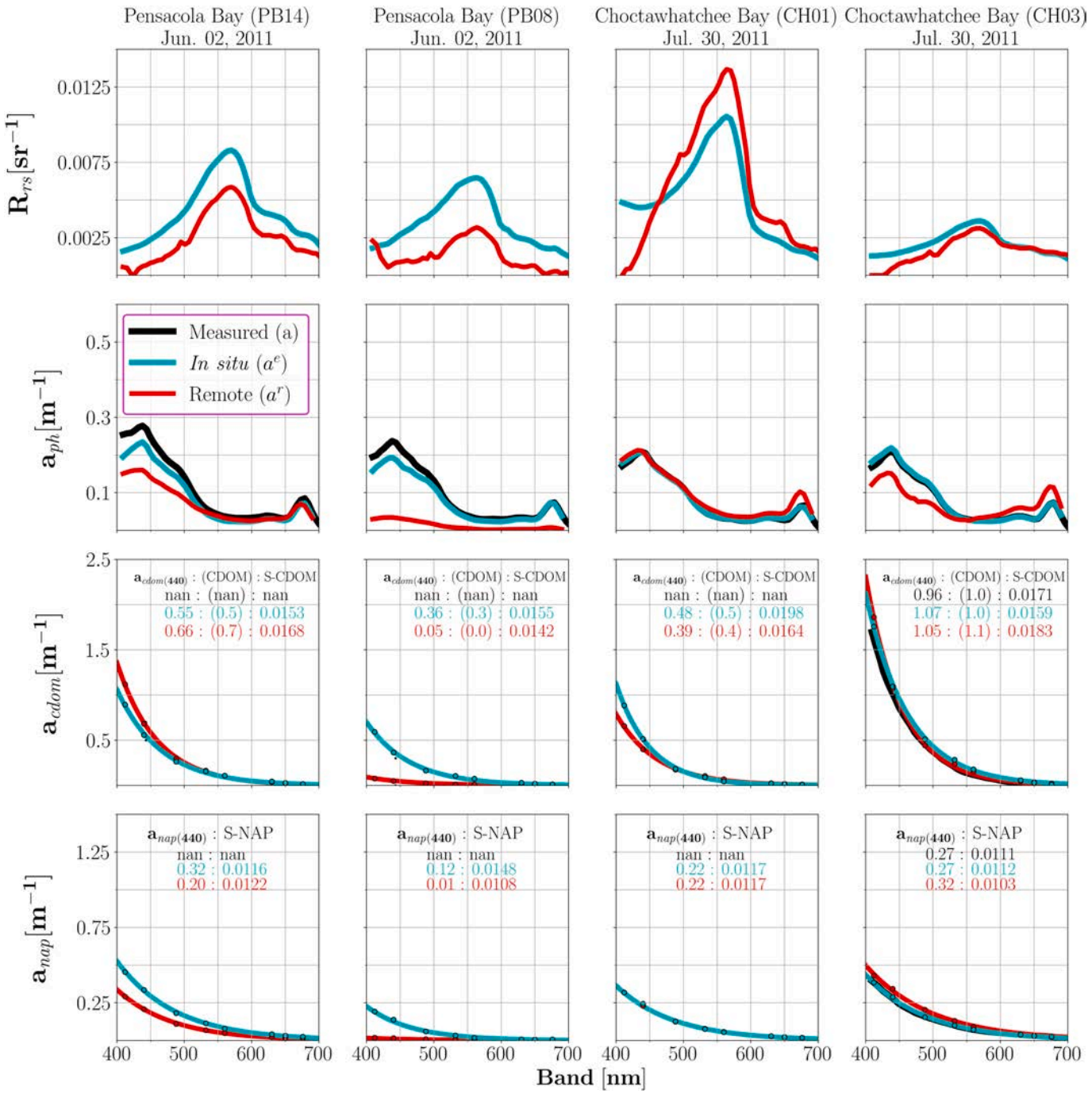

Fig. 10. Same as Fig. 9, but matching stations in row 3 from Fig. 7 of Pahlevan et al., 2021.

the DRP and CB are not only visible *via* the chl*a* and PC product maps, but also by a_{ph} at all bands, and most notably by negligible $a_{ph}(620)$, the absorption peak of phycocyanin, a pigment specific to cyanobacterial biomass. Meanwhile, absorption peaks of chl*a* and accessory pigments are visible in the spectral estimates for a_{ph} in the SB and MB (Fig. 12, row 2). The MDN-derived specific absorption coefficient of phytoplankton (Fig. 12, row 3) is similar to the shape (except with a lower peak at ~680 nm) and magnitude of the specific absorption coefficient of phytoplankton from Maumee Bay reported in the literature (Fig. 5c from Binding et al., 2019), though the MDN-derived specific absorption coefficient of phytoplankton (MDN-derived a_{ph}/chl*a*) is significantly lower for Sandusky Bay. Overall, the spatial distribution of the MDN-derived IOPs is underpinned by the literature.

5.4. Map product assessment: PRISMA imagery of the Curonian Lagoon

PRISMA-derived BPs and IOPs (Figs. 13 & D4–6) agree with *in situ* values (where available) and the literature's understanding of the hydrodynamic regime of the region. Product maps of chl*a* and CDOM match the magnitude of *in situ* values, but PC is significantly underestimated and TSS is slightly overestimated. The MDN-derived chl*a*, TSS, and PC maps, with lower values in the Baltic Sea (BS), Klaipeda Strait (KS), and Atmata-Nemunas River plume (ANRP) and higher values in the main body of the Curonian Lagoon (CL), largely agree with the typical climatology derived from remote sensing for summer months (Kowalczuk et al., 2010; Kyryliuk and Kratzer, 2019; Vaičiūtė et al., 2021; Woźniak et al., 2016). Maps of a_{cdom} at five wavelengths

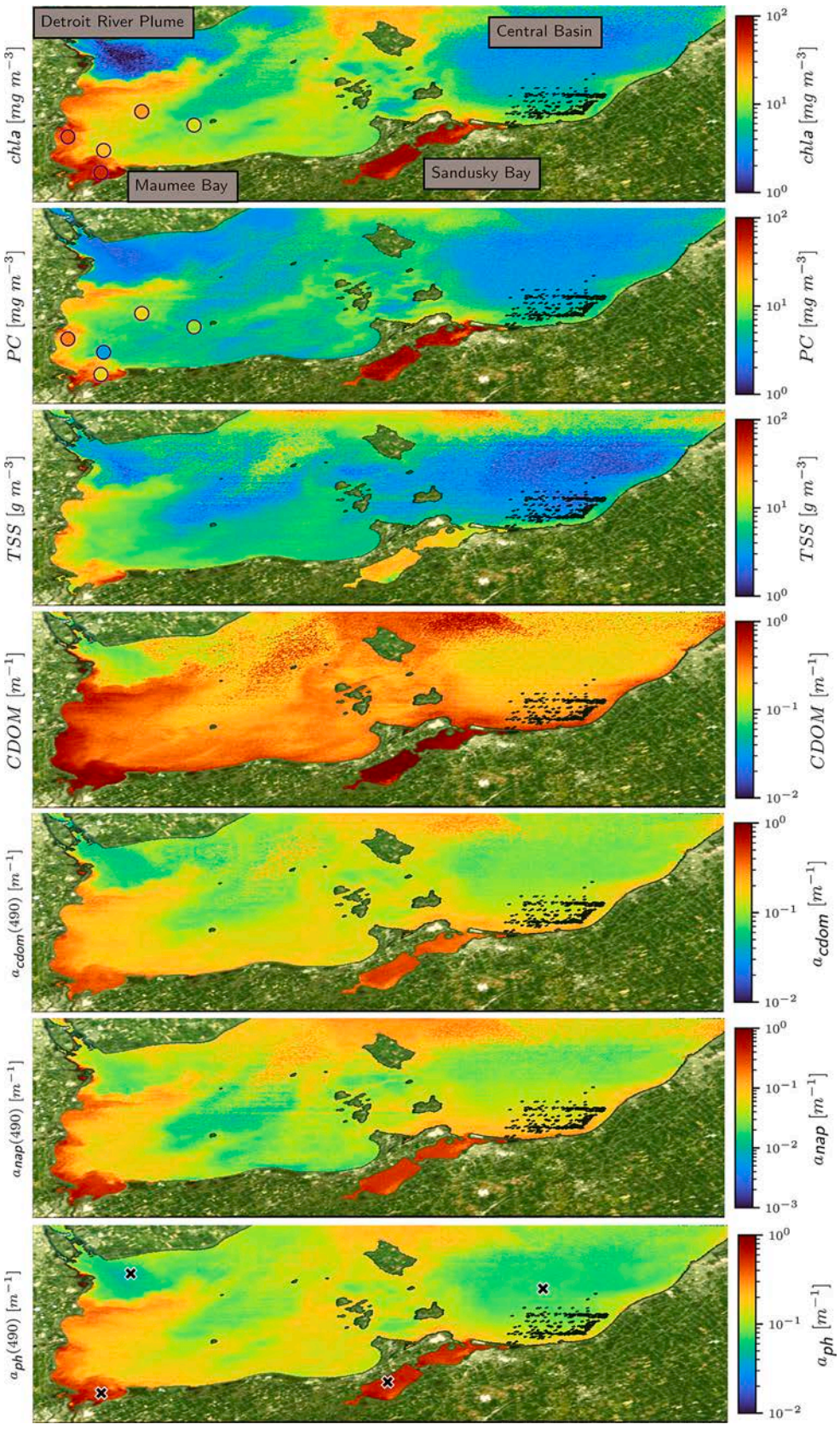

Fig. 11. MDN-derived product maps of chl*a*, PC, TSS, CDOM as well as a_{cdom}, a_{nap}, and a_{ph} at 490 nm from ACOLITE corrected HICO imagery of Lake Erie acquired on September 8th, 2014. Co-located *in situ* measurements are shown as black circles, with the inside fill color matching the *in situ* measurement. The black stippling in the Central Basin near Sandusky Bay is due to invalid R_{rs} retrievals in the blue bands. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

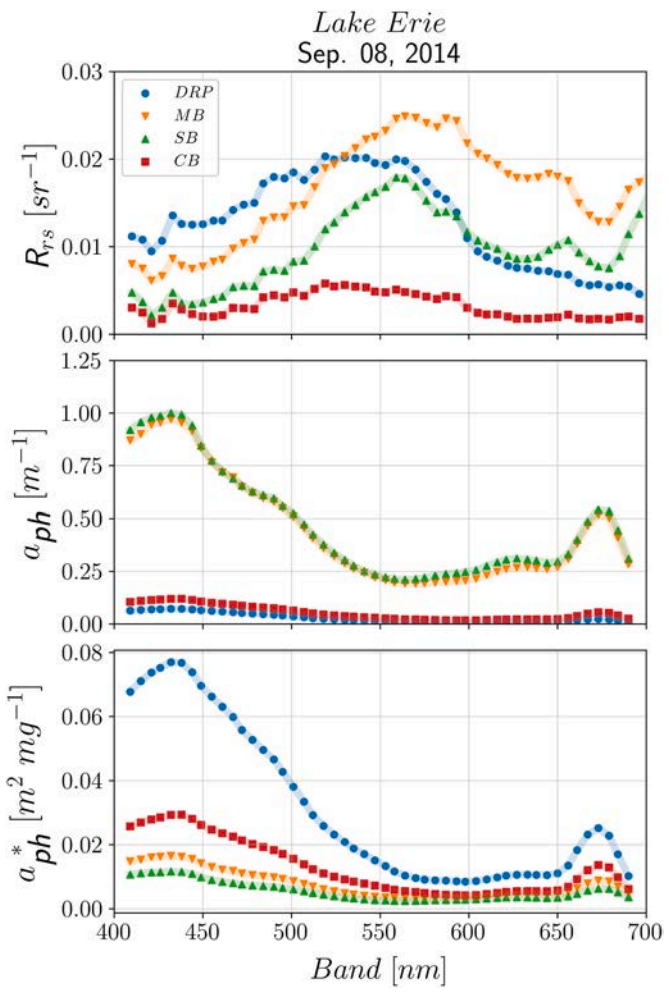

Fig. 12. Selected HICO R_{rs} from the Detroit River plume (DRP, blue circle), Maumee Bay (MB, orange upside-down triangle), Sandusky Bay (SB, green triangle), and Central Basin (CB, red square). MDN-derived a_{ph} and specific absorption ($a^*_{ph} = a_{ph}/chla$) are shown in the bottom two rows. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

demonstrate strong agreement with the magnitude of the associated *in situ* values, though there is over- and underestimation at the shorter and longer wavelengths, respectively (Fig. D4). Additionally, while the *in situ* a_{cdom} slightly increases across the spectrum in the ANRP, the MDN-derived a_{cdom} slightly decreases in this region. Although there are no *in situ* values for a_{ph}, the MDN-derived maps (Figs. 13 & D5) capture lower estimates in the BS and the KS connection to the CL, which are expected due to generally lower values in the BS. The substantially lower *in situ* PC concentration in the ANRP would suggest lower relative $a_{ph}(620)$ (due to phycocyanin's peak absorption at this wavelength), which is not the case (Figs. 13 & D5). While the chl*a* concentrations are slightly higher in the ANRP, the associated absorption at 673 nm does not increase (Figs. 13 & D5). Finally, absorption by NAP follows a similar trend, being lowest in the BS, KS, and ANRP and highest in the main body of the CL (Figs. 13 & D4), which generally agrees with the trends of the measured *in situ* TSS magnitude (Fig. 13).

6. Discussion

6.1. Simultaneous vs. dedicated models

Theoretically, simultaneously estimating additional parameters should enhance the overall performance of the MDN for each individual parameter, as the MDN learns the covariance between the estimated products. Indeed, we saw that the uncertainties of the BPs and IOPs matched well with the ones reported in previous multi- and hyperspectral applications while using imputation to fill in for missing values (Rubin, 2004; Sovilj et al., 2016), even when the missing values substantially outnumbered the *in situ* measurements. Not only did the simultaneous models generally beat the performance for the dedicated (and 3 parameter) multispectral models (Balasubramanian et al., 2020; Pahlevan et al., 2022; Smith et al., 2021) for the *in situ* estimates, but they also tied the hyperspectral models (Pahlevan et al., 2021), notably, on a significantly larger *in situ* dataset that spanned a wider distribution of optically unique scenarios. While the uncertainty may still be insufficient for aquatic biodiversity assessments (Dekker et al., 2018; Muller-Karger et al., 2018), our proposed inversion scheme significantly advances generalization accuracy in inland and coastal waters relative to operationally available algorithms. Our generalization performance improved substantially, both relative to widely used heritage algorithms (Fig. 7) and dedicated MDNs (Pahlevan et al., 2021, 2022). Interestingly, the simultaneous MDN (utilizing imputation) generalized better than the dedicated MDN. On improperly labeled training data, the dedicated MDN overfit training data that provided attenuation instead of absorption measurements, while the simultaneous MDN correctly underestimated them (not shown).

The major cost for the increased complexity of simultaneously derived spectral products comes during training, as the covariance must be learned between the different output variables. Since our model only estimates spectral a_{cdom} and a_{nap} at eight selected bands in the visible section of the spectrum, we lose information on potential real variability at non-represented wavelengths. The ideal model architecture to represent the additional hyperspectral bands, as we move forward with PACE's OCI sensor with higher spectral resolution and range, must be identified *via* a similar hyperparameter sweep as performed above (Table 4). Training with these additional parameters may require increased computational performance *via* distributed training of the 10 rounds on a high-performance computing system. While the training complexity increases substantially, runtimes on a standard computer are unlikely to increase dramatically (~30 min per HICO image).

6.2. Major drivers of uncertainty

Our current dataset has many limitations for application to globally distributed inland and coastal waters. One major limitation is the uneven global representation of inland and coastal waters. While there are substantial measurements from optically unique water bodies in North America and Europe, there are very limited contributions (~20%) from inland and coastal waters of South America, Asia, and Africa (Pahlevan et al., 2022). Although one may expect high epistemic uncertainty in these regions, due to limited training samples requiring the MDN to extrapolate to some degree, the MDN performed quite well on out-of-training samples across the globe (Fig. 7), with a performance comparable to that of operational algorithms on the undersampled regions (South America, Asia, Oceania, and Africa). These results, where the MDN matches operational algorithm accuracy in out-of-training set regions, aligns with previous leave-one-out assessments applied to simultaneous MDNs, even when compared against additional operational algorithms (Pahlevan et al., 2022). Not only are these models comparable to operational algorithms on regions with limited data, but their epistemic and aleatoric uncertainty can also be estimated in these regions (Saranathan et al., 2023), to better determine the validity of the estimates.

Other, more minor, limitations in our dataset exist. First, measurements of phycocyanin are left-skewed (Fig. 2), likely due to practical limitations of water quality managers only having sufficient funding to sample PC in water bodies susceptible to cyanobacterial blooms. Another limitation that may prevent comparison between PC samples representing different regions is the lack of standardized analytical

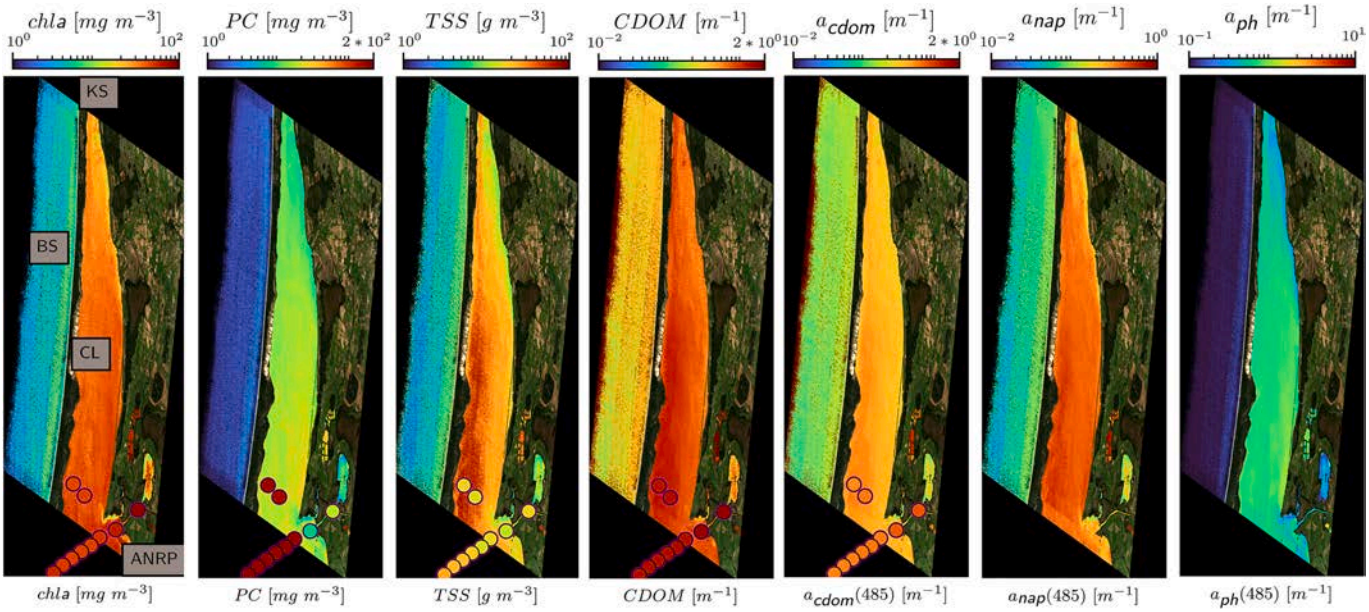

Fig. 13. MDN-derived product maps of chl*a*, PC, TSS, and CDOM as well as a_{cdom}, a_{nap}, and a_{ph} at 490 nm from PRISMA imagery of the Curonian Lagoon acquired on September 20th, 2020. Co-located *in situ* measurements are shown as black circles, with the inside fill color matching the *in situ* measurement.

Table 5
Literature measured (Moore et al., 2017) values in Lake Erie from the summers of 2013 and 2014, retabulated (from Table 1 in Moore et al., 2017) for visual comparison to MDN-derived product maps (Fig. 11). Range of literature measured values is shown in parenthesis, after the mean.

Product	Detroit River Plume	Central Basin	Maumee Bay
Chl*a* (mg/m^3)	5.4 (1.3–16.5)	17.2 (6.0–22.5)	54.7 (4.6–100)
TSS (g/m^3)	3.2 (2.4–3.8)	4.2 (2.7–5.5)	18.6 (1.8–57.8)
CDOM (m^{-1})	0.09 (0.06–0.11)	0.24 (0.13–0.39)	0.38 (0.08–1.18)
PC (mg/m^3)	0.2 (0–0.4)	3.5 (1.0–5.3)	40.5 (0.7–210)

methods (Zimba, 2012). Finally, due to the reduced revisit time of current and past proof-of-concept hyperspectral sensors, there are a limited number of *in situ* matchups available, and only from a specific few well-studied regions. With the upcoming PACE mission, additional efforts will have to be made to retrieve *in situ* measurements from inland and coastal waters that are co-aligned with satellite overpasses.

Besides the uncertainty associated with an imperfect *in situ* dataset, the majority of the remaining uncertainty is associated with the uncertainties in the satellite imagery, including uncertainties in the atmospheric correction process, imperfect radiometric calibration, and low signal-to-noise ratios in the hyperspectral imagery (Ibrahim et al., 2018; O'Shea et al., 2021; Pahlevan et al., 2020). We found differences in the range of 37.4–62.8% between products estimated from *in situ* R_{rs} and co-located satellite R_{rs} (Fig. 8), which generally exceed or match the uncertainties associated with the 50/50 training/testing split and leave-one-out cross-validation for most products (Figs. 6 & 7). Uncertainty related to the atmospheric correction process not only resulted in random errors, but incorrect spectral shapes in features for some products, such as a_{ph} (Figs. 9 & 10, stations SA11 & CH03), which can be even more deleterious for phytoplankton community composition retrieval. Fortunately, certain products, such as a_{cdom} and a_{nap}, were minimally affected by uncertainties in R_{rs} (Figs. 8, 9, & 10). Overall, the largest increase in phytoplankton composition analysis capabilities from hyperspectral satellite sensors will likely come from reduced uncertainty in the atmospheric correction process and increasing representation within *in situ* training datasets, and not direct improvements to inversion frameworks.

6.3. Implications for science and applications

Using our product suite, water resources managers studying and/or monitoring HABs (*e.g.*, cyanobacterial blooms) will have the ability to augment their existing sampling regimes to incorporate undersampled phytoplankton pigments, such as phycocyanin, and use their spatio-temporal distribution to target sites of interest for additional *in situ* monitoring, while managing their monitoring resources more conservatively (Stroming et al., 2020). Since the model is developed with global data, and shown to perform well in leave-one-out regional datasets (Fig. 7), the model does not need to be retuned for reasonable retrievals from new regions. While our product suite is currently available online *via* a codebase (https://github.com/STREAM-RS/STREAM-RS), we are also developing a web-based toolkit for displaying interpretation ready products. Our high-fidelity hyperspectral products will allow for the quantification and optical characterization of phytoplankton blooms and studying of phytoplankton diversity. Further, retrieving a_{nap} may permit the determination of relative proportions of inorganic and organic matter (Babin, 2003). In bloom conditions, a_{ph} products will also trigger developments of innovative algorithms for the retrievals of phytoplankton properties, such as size classes, pigment composition and concentration, and species abundance in highly eutrophic and/or sediment-rich waters (Chase et al., 2013; Ciotti et al., 2002). Moreover, using our IOP products, the scientific community may explore physics-based closure analyses (Mobley et al., 2002) to arrive at lesser-known variables, such as the phytoplankton Volume Scattering Function (VSF) and Bidirectional Reflectance Distribution Function (BRDF) (Twardowski et al., 2012; Twardowski and Tonizzo, 2018).

6.4. Future directions

The MDN's accuracy and applicability can be improved through the addition of input and output features. Additional metadata, such as water temperature, salinity, and latitude/longitude, can all be added as

input features to the MDN, to better constrain the possible spectral a_{ph} (*via* constraining the phytoplankton community composition) (Bouman et al., 2003; Brewin et al., 2019; López-Urrutia and Morán, 2015; Montes-Hugo and Xie, 2015; Olli et al., 2019; Robinson et al., 2021). Additionally, while the MDN may produce inaccurate estimates for any individual spectra, the outputs may serve as the initial parameterization of an optimization technique (*e.g.*, GIOP), to improve final estimates. In addition to potentially increasing accuracy, this would also allow for calculating additional IOPs, such as the particulate backscattering coefficient (b_{bp}) by leveraging pre-existing optimization techniques (*e.g.*, GIOP) (Werdell et al., 2013). Additionally, regional datasets could be used in transfer learning-based approaches to increase accuracy for parameter retrieval in optically unique regions. Finally, per-pixel uncertainty maps could be leveraged to identify areas with high confidence estimates for stakeholder use and biogeochemical modeling (Saranathan et al., 2023). Overall, machine learning models may be able to fully leverage available information and to better estimate phytoplankton pigments, which, when combined with optimization techniques, may achieve the highest accuracy estimates.

7. Conclusions

This study presents the first MDN-based hyperspectral inversion framework for simultaneous estimation of chl*a*, PC, TSS, CDOM, a_{ph}, a_{cdom}, and a_{nap} from inland and coastal water bodies. The model achieves similar or better accuracy (15–40% for all products, but PC, and $<$ 60% for PC) on a larger, more diverse, *in situ* dataset than previous, *dedicated* MDNs. When retrained using a leave-one-out cross-validation approach, the MDN had better generalization performance than operational algorithms for all BPs and IOPs (except for PC). Absolute model accuracy is sensitive to uncertainties typical of hyperspectral satellite imagery for all BPs and IOPs, with a_{cdom} and a_{nap} exhibiting reduced sensitivity relative to a_{ph}. Despite this sensitivity, product maps from two optically distinct regions, Lake Erie and the Curonian Lagoon, generally agree with both co-located *in situ* samples and the literature's understanding of these regions. The model was rapidly retrained and deployed for PRISMA, an additional, currently operational, hyperspectral imager, demonstrating the model's expeditious redeployment capabilities in preparation for the upcoming Plankton, Aerosol, Cloud, ocean Ecosystem (PACE) mission. The ability to rapidly retrain and deploy the model addresses stakeholders concerns about data continuity over multiple mission lifetimes (Schaeffer et al., 2013). Substantial uncertainty from the atmospheric correction process suggests that the largest accuracy improvement for recovering IOPs would come from increased accuracy in retrieval of the remote sensing reflectance, stemming from improvements in atmospheric correction techniques. Further improvements to the simultaneous MDN would include environmental and physical variables as input features, to enable phytoplankton type/species discrimination in a constrained region.

CRediT author statement

Ryan E. O'Shea: Conceptualization, Software, Data Curation, Formal Analysis, Writing
Nima Pahlevan: Supervision, Conceptualization, Writing
Brandon Smith: Software, Data Curation
Emmanuel Boss: Conceptualization, Writing – Review and Editing
Daniela Gurlin: Data Curation, Writing – Review and Editing
Krista Alikas: Data Curation, Writing – Review and Editing
Kersti Kangro: Data Curation, Writing – Review and Editing
Raphael M. Kudela: Conceptualization, Writing – Review and Editing
Diana Vaičiūtė: Data Curation, Writing – Review and Editing

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.

Data availability

The code for running the HICO and PRISMA models is located at https://github.com/STREAM-RS/STREAM-RS. Weights must be requested separately from the corresponding author.

Acknowledgements

This manuscript was supported through the NASA ROSES grants #80NSSC20M0235, PACE Science and Applications Team, and #80NSSC21K0499, Ocean Biology and Biogeochemistry (OBB) program. This work made use of PRISMA Products© of the Italian Space Agency (ASI) which were delivered under the agreement of an ASI license to use. We would like to thank all data contributors, including: Claudia Giardino and Mariano Bresciani for data collection and curation from the Curonian Lagoon, Bunkei Matsushita and Dalin Jiang for data collection and curation from Chinese and Japanese Lakes, Linwei Yue, Cédric G. Fichot, Arnold G. Dekker, Tim J. Malthus, Micheal Ondrusek, Hà Nguyễn, Marcel R. Wernand, Stéfani Novoa, Hans J. van der Woerd, Cláudio C.F. Barbosa, Daniel A. Maciel, Waterloo Pereira Filho, Evlyn M. L. M. Novo, Courtney Di Vittorio, Dariusz Ficek, Peter Gege, Ian Somlai-Schweiger, Stefan Plattner, Moritz K. Lehmann, Leif G. Olmanson, Patrick Brezonik, Antonio Ruiz Verdú, Glenn Cota, David Doxaran, Kevin Ruddick, all of the SeaBASS contributors, and all of the GLORIA dataset contributors (Lehmann et al., 2023). Specifically for IOPs, we would like to thank: Kimberly A. Casey (Casey et al., 2019), Caren Binding, Janet Anstee, Wesley J. Moses, Chuanmin Hu, Anatoly A. Gitelson, Steven R. Greb, Lin Li, Deepak R. Mishra, Sachidananda Mishra, Timothy S. Moore, David M. O'Donnell, Natascha Oppelt, Blake Schaeffer, Richard C. Zimmerman, John F. Schalles, Stefan G. H. Simis, Martin Ligi, Nathan Drayson, Ronghua Ma, and Zhigang Cao.

Appendix A. Atmospheric correction settings

The HICO imagery of Lake Erie in Figs. 11 and D1-D3 were processed with ACOLITE, using user defined settings as listed in Table A1.

Table A1
User defined ACOLITE settings for processing HICO imagery.

ACOLITE Setting	User Defined Value
l2w_mask	False
l2w_mask_water_parameters	False
l2w_mask_negative_rhow	False
l2w_mask_cirrus	False

(*continued on next page*)

Table A1 (*continued*)

ACOLITE Setting	User Defined Value
l2w_mask_high_toa	False
polylakes	False
l2w_mask_wave	794
l2w_mask_threshold	0.08
l2w_mask_cirrus_wave	1011
l2w_mask_cirrus_threshold	0.12
l2w_mask_high_toa_threshold	10.0
l2w_mask_smooth	False
l2w_mask_negative_wave_range	410,725
limit	41.3330468,-83.71757,42.31545,-80.677101

Appendix B. Hyperparameter optimization

The MDN hyperparameter search (covering the hyperparameters shown in Table 4) yielded hyperparameters suitable for ocean color remote sensing. The models with similar hyperparameters as the standard parameters used for estimating individual products (5 layers, 100 neurons, 10,000 training iterations, and an epsilon of 0.001) yielded higher uncertainty (Fig. B1, leftmost plot) than alternative hyperparameter combinations, demonstrating the insufficient complexity of this combination for the simultaneous estimation of such a large number of products. Changes in the epsilon value result in dramatically different uncertainty ranges for the respective hyperparameter combinations.

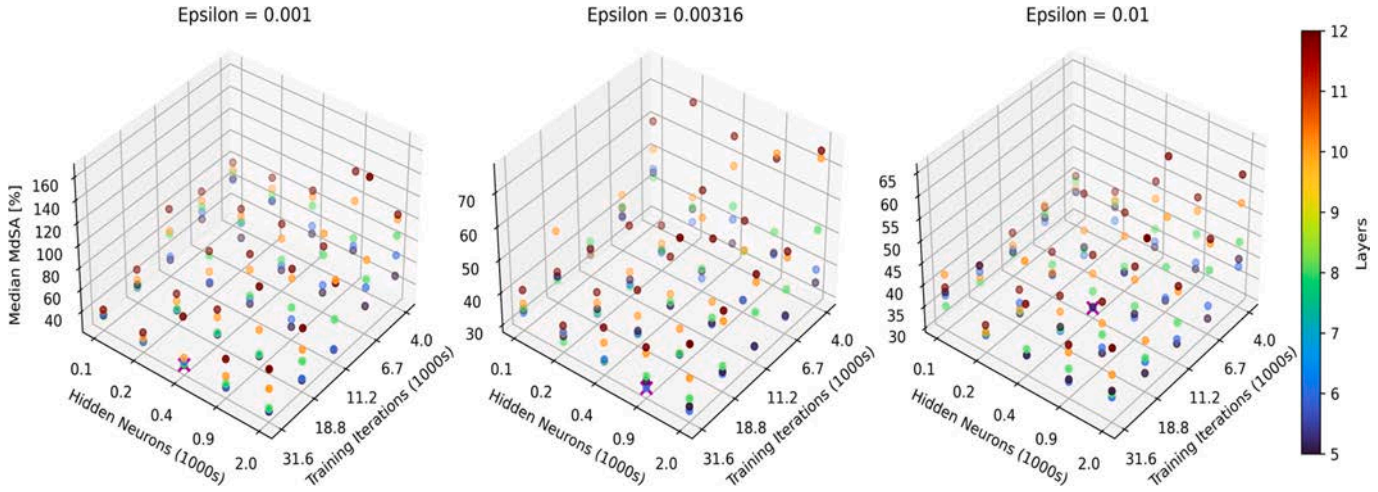

Fig. B1. MDN hyperparameter optimization for the hyperspectral simultaneous retrieval of BPs and IOPs. Hidden neurons, training iterations, layers, and epsilon values were varied. The lowest median MdSA for each epsilon value is shown with a magenta X. The overall lowest median MdSA (taken over all products, with spectral products represented by the median over all bands) is (30.45%) and occurs at the hyperparameters covered in Table 4. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Appendix C. Performance assessment - 50/50 training/testing split

Additional BP products and IOPs from the test half of the 50/50 split benchmark assessment demonstrate the MDNs improvement against (unoptimized) operational algorithms (C1–6).

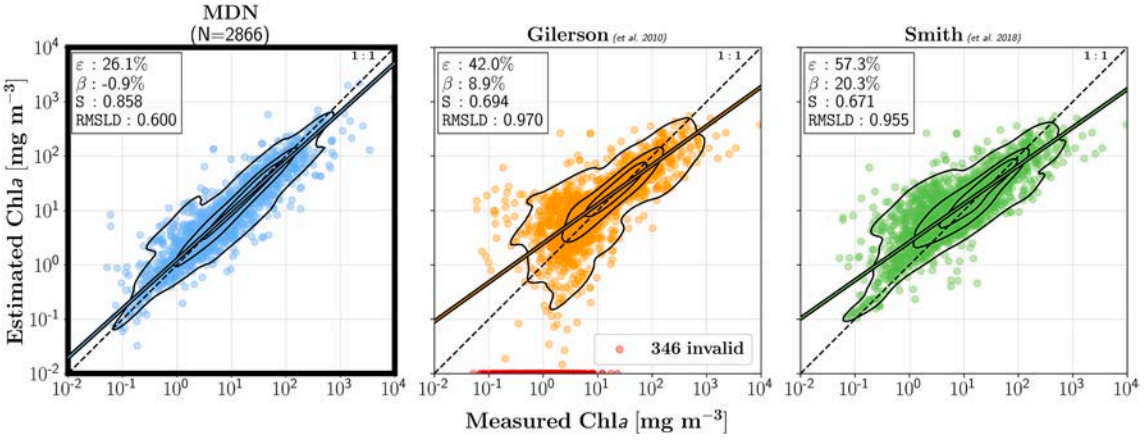

Fig. C1. Estimated *vs.* measured chl*a* from the test half of the 50/50 split, compared against operational algorithms (Gilerson et al., 2010; Smith et al., 2018).

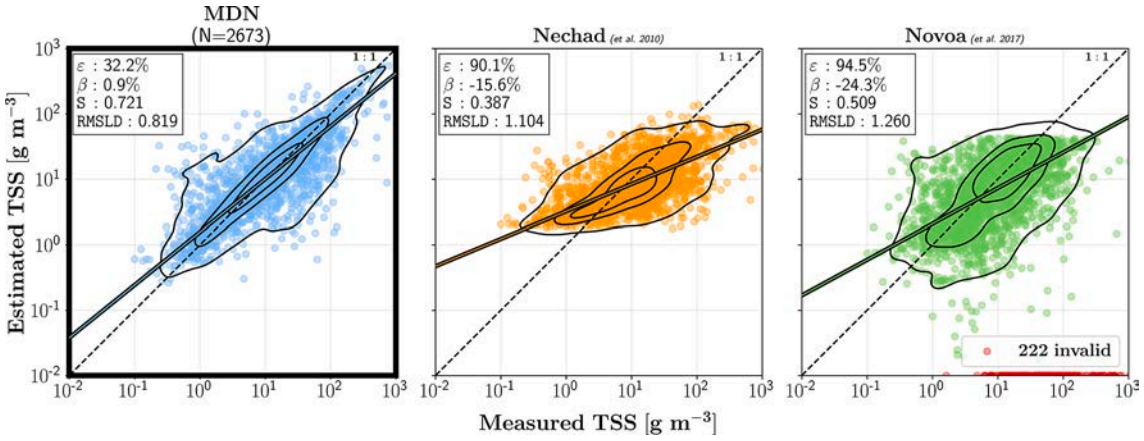

Fig. C2. Estimated *vs.* measured TSS from the test half of the 50/50 split, compared against operational algorithms (Nechad et al., 2010; Novoa et al., 2017). These operational algorithms were chosen as they perform well on globally distributed waters and have been previously used for comparison to TSS retrieval from MDNs (Balasubramanian et al., 2020; Pahlevan et al., 2022).

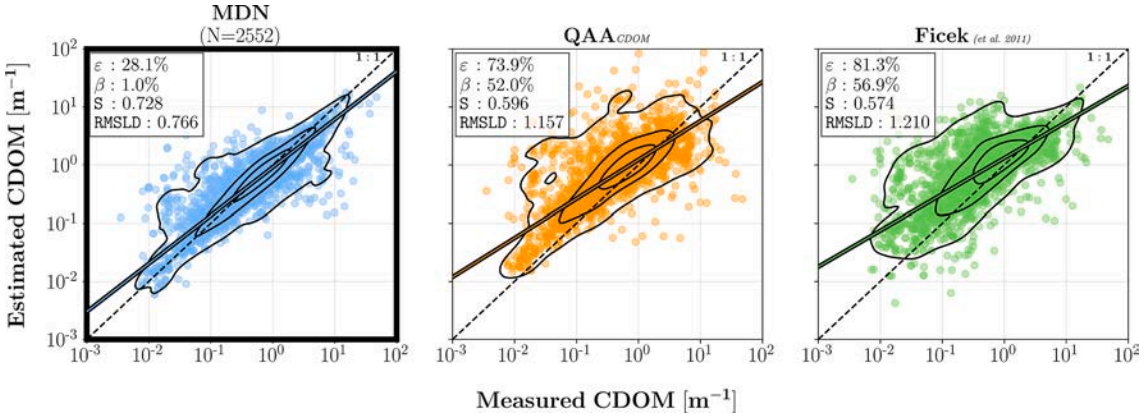

Fig. C3. Estimated *vs.* measured CDOM from the test half of the 50/50 split, compared against operational algorithms (Ficek et al., 2011; Zhu and Yu, 2013).

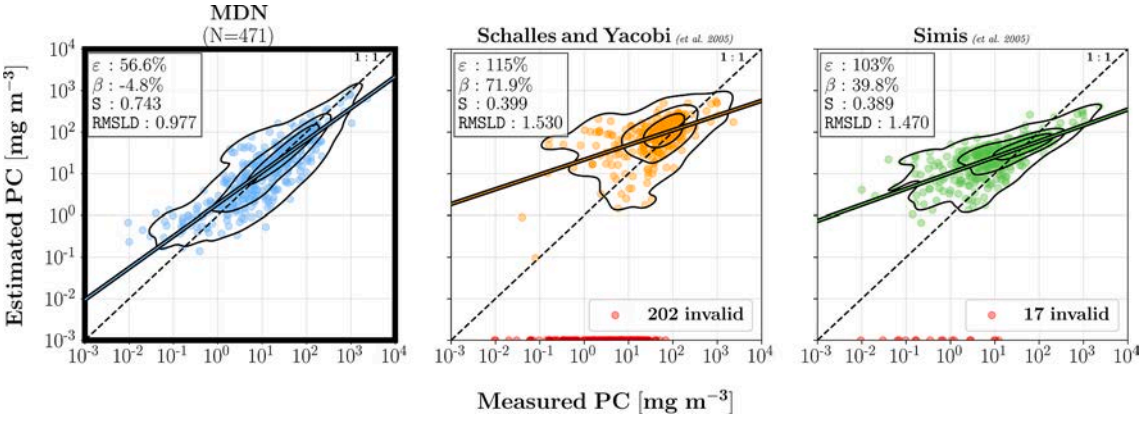

Fig. C4. Estimated *vs.* measured PC from the test half of the 50/50 split, compared against operational algorithms (Schalles and Yacobi, 2000; Simis et al., 2005).

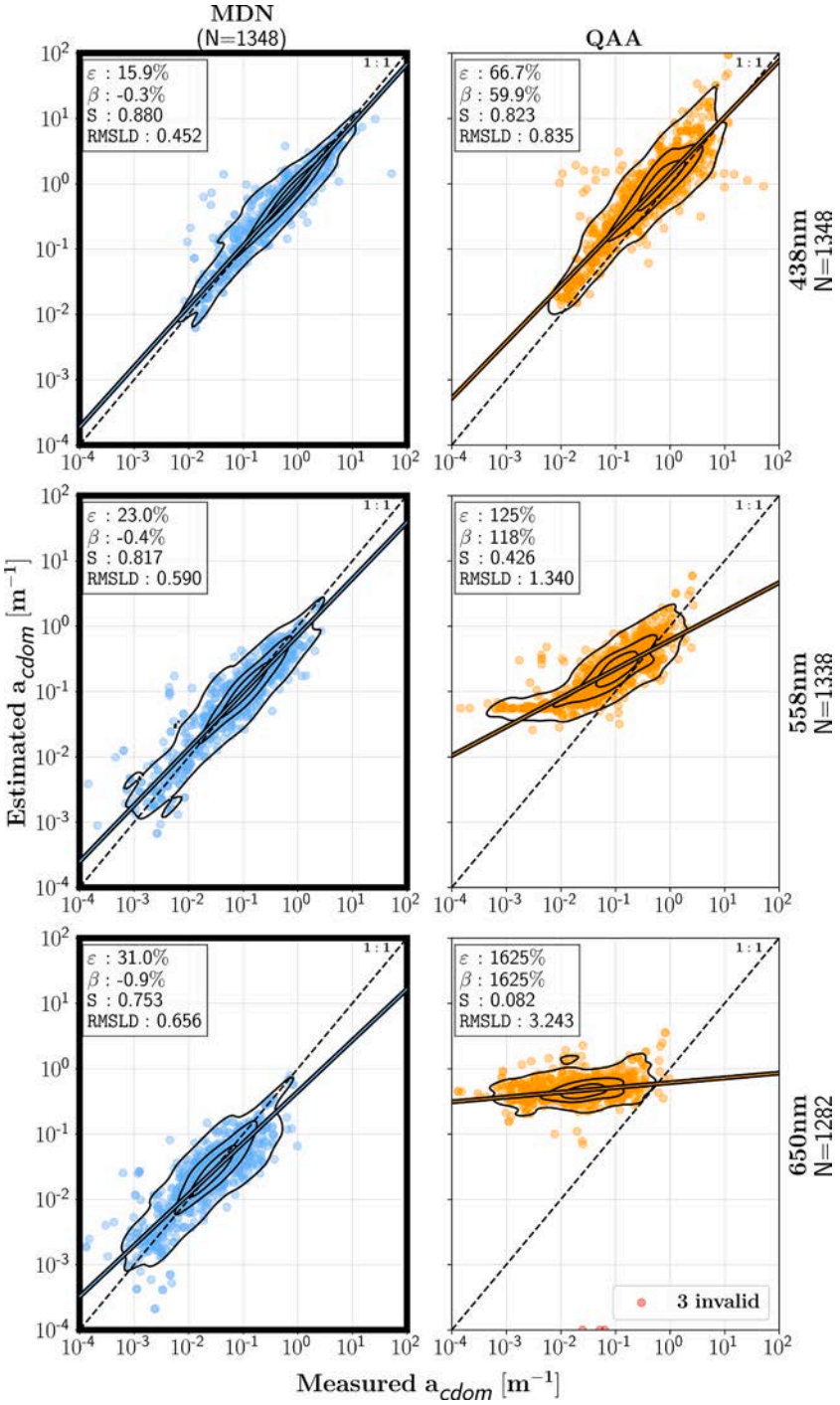

Fig. C5. Estimated *vs.* measured a_{cdom} from the test half of the 50/50 split, compared against operational algorithms.

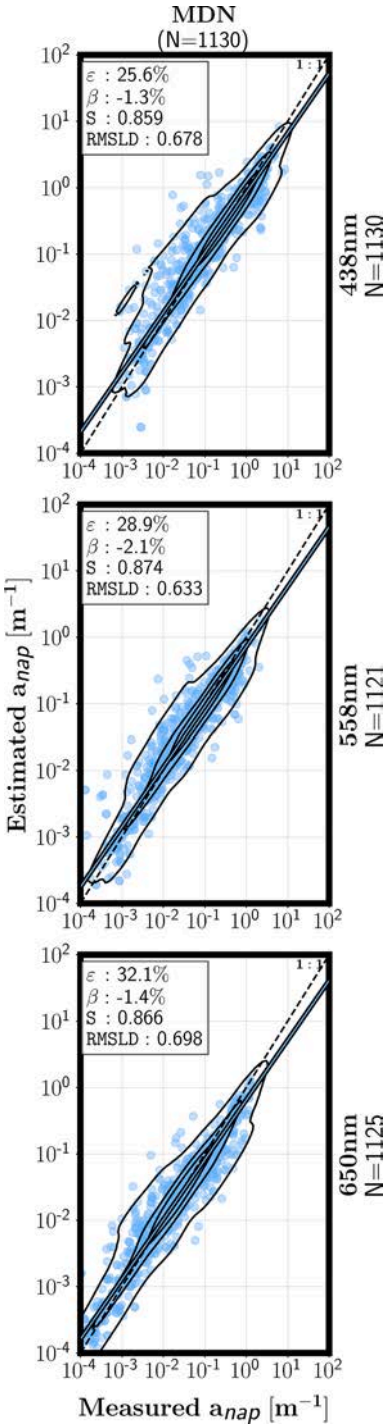

Fig. C6. Estimated *vs.* measured a_{nap} from the test half of the 50/50 split.

Appendix D. Additional HICO and PRISMA product maps

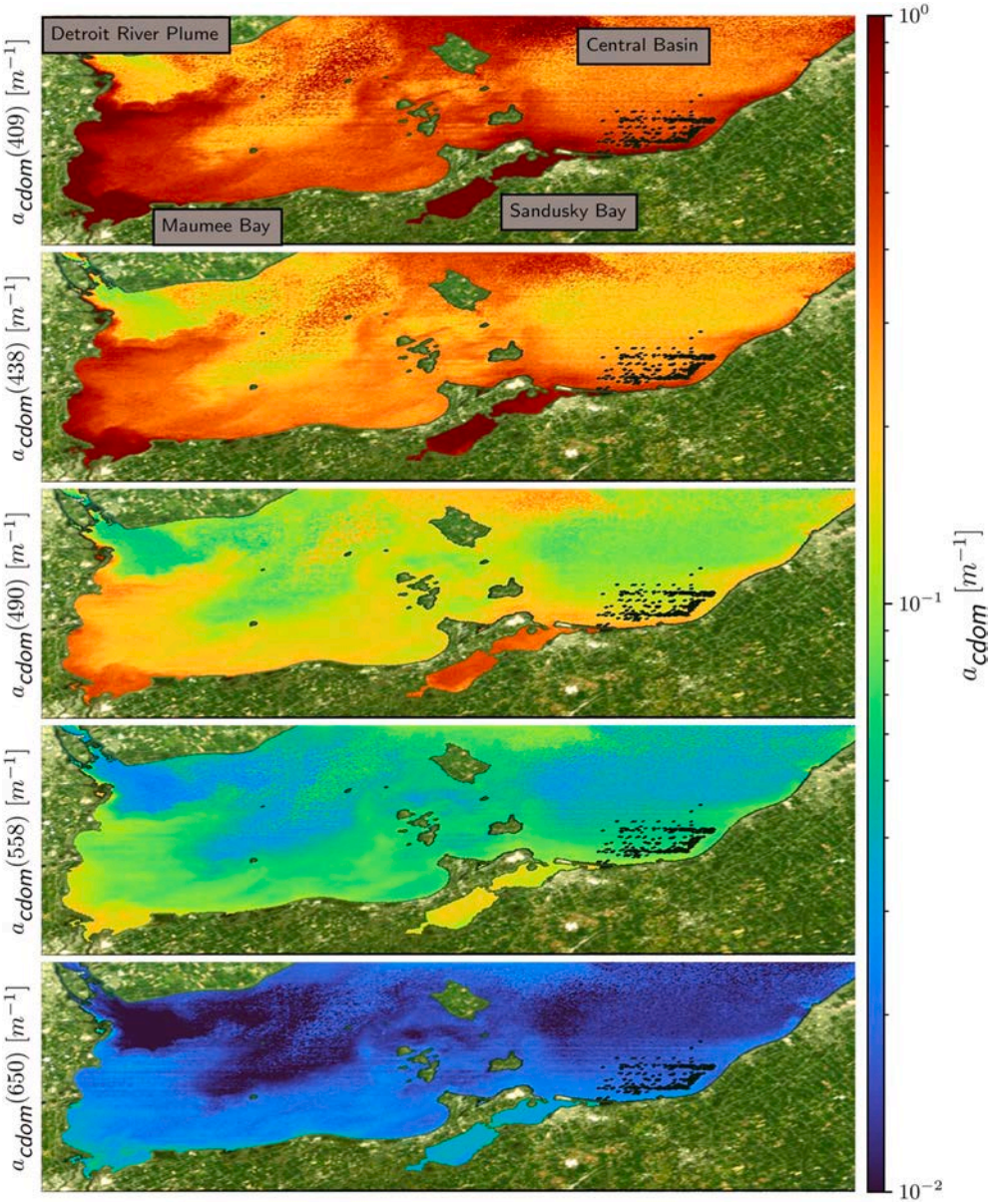

Fig. D1. MDN-derived product maps of a_{cdom} at 409, 438, 490, 558, & 650 nm (530, 626, & 673 nm not shown) from ACOLITE corrected HICO imagery of Lake Erie acquired on September 8th, 2014.

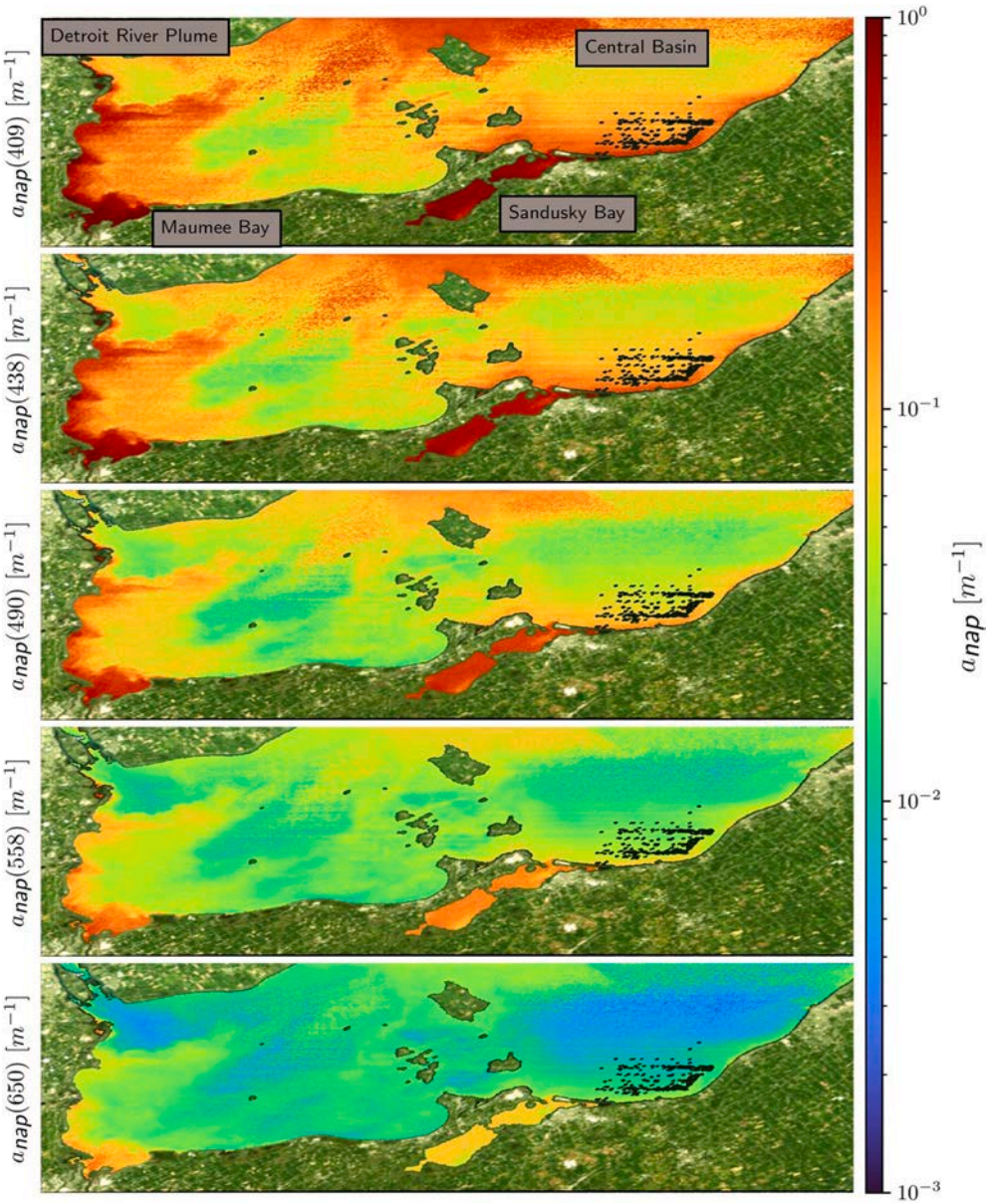

Fig. D2. MDN-derived product maps of a_{nap} at 409, 438, 490, 558, & 650 nm (530, 626, & 673 nm not shown) from ACOLITE corrected HICO imagery of Lake Erie acquired on September 8th, 2014.

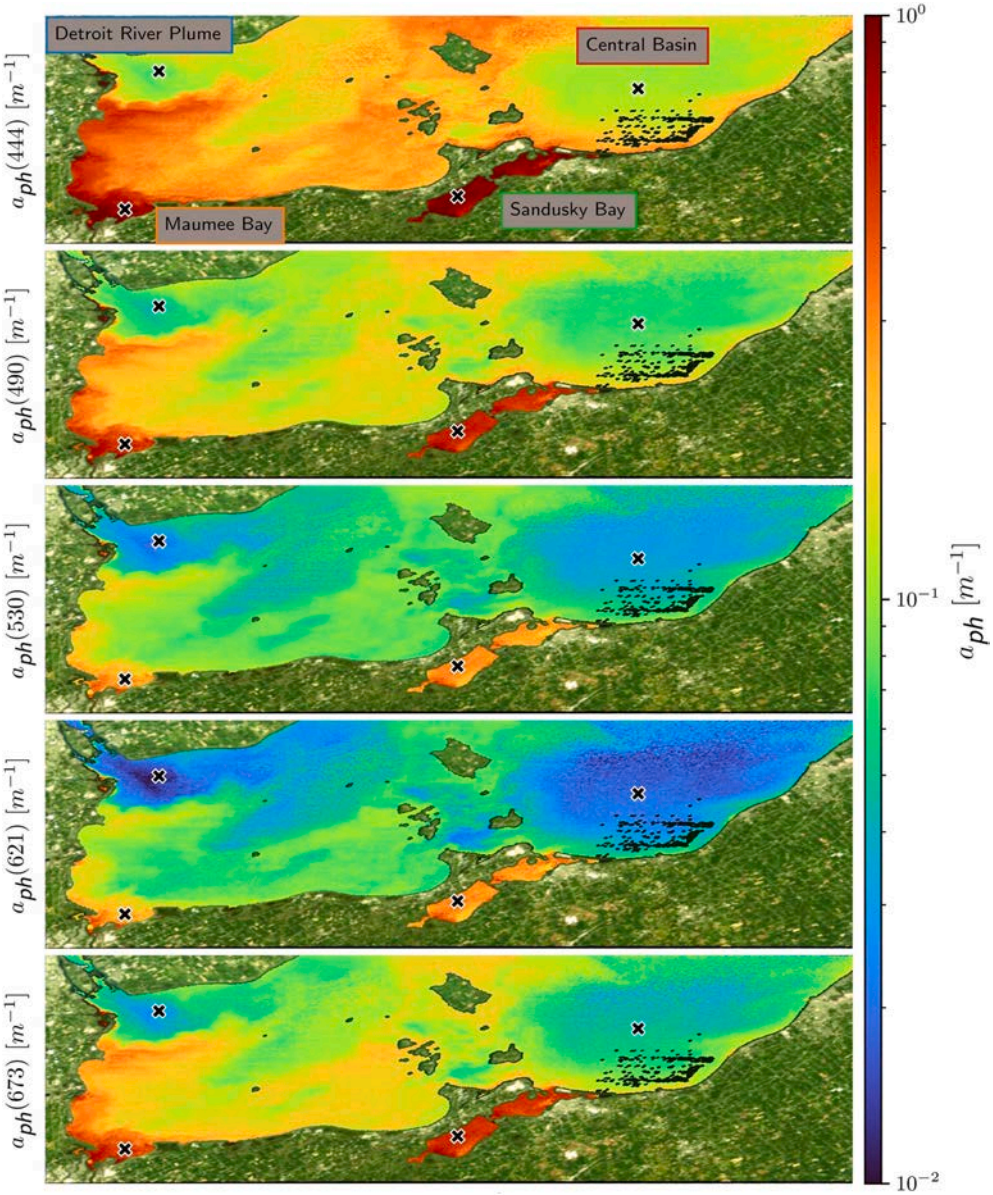

Fig. D3. MDN-derived product maps of a_{ph} at 444, 490, 530, 621, & 673 nm from ACOLITE-corrected HICO imagery of Lake Erie acquired on September 8th, 2014. Black saltires (x) with white borders mark the sample locations used in Fig. 12, the labels' border color corresponds to R_{rs} from that location in Fig. 12.

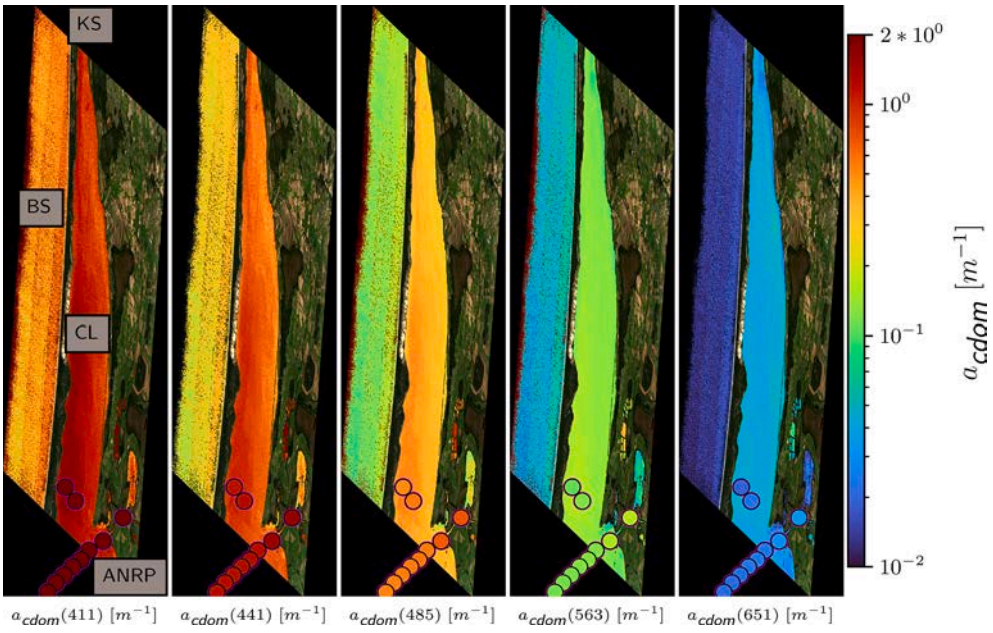

Fig. D4. MDN-derived product maps of a_{cdom} at at 411, 441, 485, 563, & 651 nm from PRISMA imagery of the Curonian Lagoon acquired on September 20th, 2020. Co-located *in situ* measurements are shown as black circles, with the inside fill color matching the *in situ* measurement.

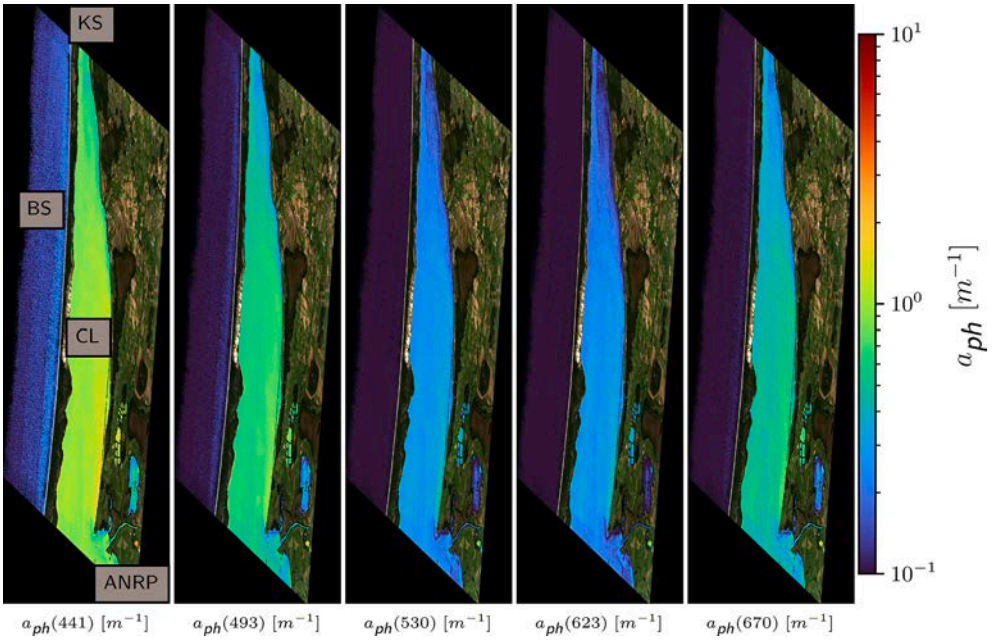

Fig. D5. MDN-derived product maps of a_{ph} at 441, 493, 530, 623, & 670 nm from PRISMA imagery of the Curonian Lagoon acquired on September 20th, 2020.

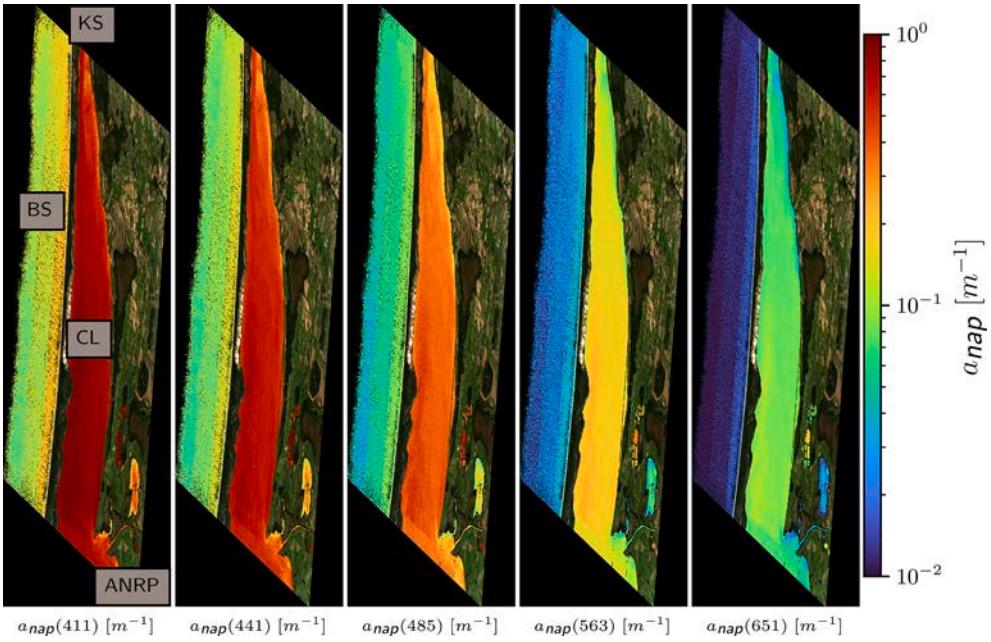

Fig. D6. MDN-derived product maps of a_{nap} at 411, 441, 485, 563, & 651 nm from PRISMA imagery of the Curonian Lagoon acquired on September 20th, 2020.

References

Aguilar-Maldonado, J.A., Santamaría-del-Ángel, E., Gonzalez-Silvera, A., Sebastiá-Frasquet, M.T., 2019. Detection of phytoplankton temporal anomalies based on satellite inherent optical properties: a tool for monitoring phytoplankton blooms. Sensors 19, 3339. https://doi.org/10.3390/s19153339.

Baban, S.M.J., 1995. The use of Landsat imagery to map fluvial sediment discharge into coastal waters. Mar. Geol. 123, 263–270. https://doi.org/10.1016/0025-3227(95)00003-H.

Babin, M., 2003. Variations in the light absorption coefficients of phytoplankton, nonalgal particles, and dissolved organic matter in coastal waters around Europe. J. Geophys. Res. 108, 3211. https://doi.org/10.1029/2001JC000882.

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 https://doi.org/10.1016/j.rse.2020.111768.

Becker, A., Meister, A., Wilhelm, C., 2002. Flow cytometric discrimination of various phycobilin-containing phytoplankton groups in a hypertrophic reservoir. Cytometry 48, 45–57. https://doi.org/10.1002/cyto.10104.

Bélanger, S., Cizmeli, S.A., Ehn, J., Matsuoka, A., Doxaran, D., Hooker, S., Babin, M., 2013. Light absorption and partitioning in Arctic Ocean surface waters: impact of multiyear ice melting. Biogeosciences 10, 6433–6452. https://doi.org/10.5194/bg-10-6433-2013.

Betancur-Turizo, S., González-Silvera, A., Santamaría-del-Ángel, E., Tan, J., Frouin, R., 2018. Evaluation of semi-analytical algorithms to retrieve particulate and dissolved absorption coefficients in gulf of California optically complex waters. Remote Sens. 10, 1443. https://doi.org/10.3390/rs10091443.

Binding, C., Jerome, J., Bukata, R., Booty, W., 2008. Spectral absorption properties of dissolved and particulate matter in Lake Erie. Remote Sens. Environ. 112, 1702–1711. https://doi.org/10.1016/j.rse.2007.08.017.

Binding, C.E., Pizzolato, L., Zeng, C., 2021. EOLakeWatch; delivering a comprehensive suite of remote sensing algal bloom indices for enhanced monitoring of Canadian eutrophic lakes. Ecol. Indic. 121, 106999 https://doi.org/10.1016/j.ecolind.2020.106999.

Binding, C.E., 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. Gt. Lakes Res. 45, 573–586. https://doi.org/10.1016/j.jglr.2018.11.015.

Bishop, C.M., 1995. Neural Networks for Pattern Recognition. Oxford University Press.

Bishop, C.M., 1994. Mixture Density Networks. Aston University, Birmingham, U.K.

Bouman, H., Platt, T., Sathyendranath, S., Li, W., Stuart, V., Fuentes-Yaco, C., Maass, H., Horne, E., Ulloa, O., Lutz, V., Kyewalyanga, M., 2003. Temperature as indicator of optical properties and community structure of marine phytoplankton: implications for remote sensing. Mar. Ecol. Prog. Ser. 258, 19–30. https://doi.org/10.3354/meps258019.

Bowers, D.G., Brett, H.L., 2008. The relationship between CDOM and salinity in estuaries: an analytical and graphical solution. J. Mar. Syst. 73, 1–7. https://doi.org/10.1016/j.jmarsys.2007.07.001.

Brando Guillaumes, A., 2017. Mixture Density Networks for Distribution and Uncertainty Estimation. Universitat Politècnica de Catalunya.

Brewin, R.J.W., Ciavatta, S., Sathyendranath, S., Skákala, J., Bruggeman, J., Ford, D., Platt, T., 2019. The influence of temperature and community structure on light absorption by phytoplankton in the North Atlantic. Sensors 19, 4182. https://doi.org/10.3390/s19194182.

Bricaud, A., Babin, M., Claustre, H., Ras, J., Tièche, F., 2010. Light absorption properties and absorption budget of Southeast Pacific waters. J. Geophys. Res. 115, C08009. https://doi.org/10.1029/2009JC005517.

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. https://doi.org/10.1364/AO.46.001251.

Busetto, L., Ranghetti, L., 2020. Prismaread: A Tool for Facilitating Access and Analysis of PRISMA L1/L2 Hyperspectral Imagery v1.0.0. Zenodo. https://doi.org/10.5281/ZENODO.4019081.

van Buuren, S., Groothuis-Oudshoorn, K., 2011. Mice : multivariate imputation by chained equations in R. J. Stat. Softw. 45 https://doi.org/10.18637/jss.v045.i03.

Cael, B.B., Boss, E., 2017. Simplified model of spectral absorption by non-algal particles and dissolved organic materials in aquatic environments. Opt. Express 25, 25486. https://doi.org/10.1364/OE.25.025486.

Campbell, J., Antoine, D., Armstrong, R., Arrigo, K., Balch, W., Barber, R., Behrenfeld, M., Bidigare, R., Bishop, J., Carr, M.-E., Esaias, W., Falkowski, P., Hoepffner, N., Iverson, R., Kiefer, D., Lohrenz, S., Marra, J., Morel, A., Ryan, J., Vedernikov, V., Waters, K., Yentsch, C., Yoder, J., 2002. Comparison of algorithms for estimating ocean primary production from surface chlorophyll, temperature, and irradiance: Comparison of primary productivity algorithms. Glob. Biogeochem. Cycles 16. https://doi.org/10.1029/2001GB001444, 9–1-9–15.

Carder, K.L., Steward, R.G., Harvey, G.R., Ortner, P.B., 1989. Marine humic and fulvic acids: their effects on remote sensing of ocean chlorophyll: marine humic and fulvic acids. Limnol. Oceanogr. 34, 68–81. https://doi.org/10.4319/lo.1989.34.1.0068.

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 (preprint). Data, Algorithms, and Models. https://doi.org/10.5194/essd-2019-105.

Catlett, D., Siegel, D.A., 2018. Phytoplankton pigment communities can be modeled using unique relationships with spectral absorption signatures in a dynamic coastal environment. J. Geophys. Res. Oceans 123, 246–264. https://doi.org/10.1002/2017JC013195.

Chang, G.C., Dickey, T.D., 2004. Coastal Ocean optical influences on solar transmission and radiant heating rate. J. Geophys. Res. 109, C01020. https://doi.org/10.1029/2003JC001821.

Chase, A., Boss, E., Zaneveld, R., Bricaud, A., Claustre, H., Ras, J., Dall'Olmo, G., Westberry, T.K., 2013. Decomposition of in situ particulate absorption spectra. Methods Oceanogr. 7, 110–124. https://doi.org/10.1016/j.mio.2014.02.002.

Le, Cheng Feng, Li, Yun Mei, Zha, Yong, Sun, Deyong, Yin, Bin, 2009. Validation of a quasi-analytical algorithm for highly turbid eutrophic water of Meiliang Bay in Taihu Lake, China. IEEE Trans. Geosci. Remote Sens. 47, 2492–2500. https://doi.org/10.1109/TGRS.2009.2015658.

Choi, S., Lee, K., Lim, S., Oh, S., 2018. Uncertainty-aware learning from demonstration using mixture density networks with sampling-free variance modeling. In: 2018 IEEE International Conference on Robotics and Automation (ICRA). Presented at the 2018 IEEE International Conference on Robotics and Automation (ICRA). IEEE, Brisbane, QLD, pp. 6915–6922. https://doi.org/10.1109/ICRA.2018.8462978.

Chu, Q., Zhang, Y., Ma, R., Hu, M., Jing, Y., 2020. MODIS-based remote estimation of absorption coefficients of an inland turbid Lake in China. Remote Sens. 12, 1940. https://doi.org/10.3390/rs12121940.

Ciotti, Á.M., Lewis, M.R., Cullen, J.J., 2002. Assessment of the relationships between dominant cell size in natural phytoplankton communities and the spectral shape of the absorption coefficient. Limnol. Oceanogr. 47, 404–417. https://doi.org/10.4319/lo.2002.47.2.0404.

Defoin-Platel, M., Chami, M., 2007. How ambiguous is the inverse problem of ocean color in coastal waters? J. Geophys. Res. 112, C03004. https://doi.org/10.1029/2006JC003847.

Dekker, A., Pinnel, N., Gege, P., Briottet, X., Court, A., Peters, S., Turpie, K., Sterckx, S., Costa, M., Giardino, C., Brando, V., Braga, F., Bergeron, M., Heege, T., Pflug, B., 2018. Feasibility Study of an Aquatic Ecosystem Earth Observing System, CEOS Report. CSIRO, Canberra, Australia.

Dekker, A.G., 1993. Detection of optical water quality parameters for eutrophic waters by high resolution remote sensing (PhD-Thesis - Research and graduation internal). Vrije Universiteit Amsterdam.

Devlin, M.J., da Silva, E.T., Petus, C., Wenger, A., Zeh, D., Tracey, D., Álvarez-Romero, J. G., Brodie, J., 2013. Combining in-situ water quality and remotely sensed data across spatial and temporal scales to measure variability in wet season chlorophyll-a: great barrier reef lagoon (Queensland, Australia). Ecol. Process. 2, 31. https://doi.org/10.1186/2192-1709-2-31.

El Serafy, G.Y.H., Schaeffer, B.A., Neely, M.-B., Spinosa, A., Odermatt, D., Weathers, K.C., Baracchini, T., Bouffard, D., Carvalho, L., Conmy, R.N., Keukelaere, L.D., Hunter, P. D., Jamet, C., Joehnk, K.D., Johnston, J.M., Knudby, A., Minaudo, C., Pahlevan, N., Reusen, I., Rose, K.C., Schalles, J., Tzortziou, M., 2021. Integrating inland and coastal water quality data for actionable knowledge. Remote Sens. 13, 2899. https://doi.org/10.3390/rs13152899.

Favareto, L., Rudorff, N., Kampel, M., Frouin, R., Röttgers, R., Doxaran, D., Murakami, H., Dupouy, C., 2018. Bio-optical characterization and ocean colour inversion in the Eastern Lagoon of New Caledonia, South Tropical Pacific. Remote Sens. 10, 1043. https://doi.org/10.3390/rs10071043.

Ficek, D., Zapadka, T., Dera, J., 2011. Remote sensing reflectance of pomeranian lakes and the Baltic. Oceanologia 53, 959–970. https://doi.org/10.5697/oc.53-4.959.

Galimard, J.-E., Chevret, S., Curis, E., Resche-Rigon, M., 2018. Heckman imputation models for binary or continuous MNAR outcomes and MAR predictors. BMC Med. Res. Methodol. 18, 90. https://doi.org/10.1186/s12874-018-0547-1.

Giardino, C., Bresciani, M., Valentini, E., Gasperini, L., Bolpagni, R., Brando, V.E., 2015. Airborne hyperspectral data to assess suspended particulate matter and aquatic vegetation in a shallow and turbid lake. Remote Sens. Environ. 157, 48–57. https://doi.org/10.1016/j.rse.2014.04.034.

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. https://doi.org/10.1364/OE.18.024109.

Gohin, F., Saulquin, B., Oger-Jeanneret, H., Lozac'h, L., Lampert, L., Lefebvre, A., Riou, P., Bruchon, F., 2008. Towards a better assessment of the ecological status of coastal waters using satellite-derived chlorophyll-a concentrations. Remote Sens. Environ. 112, 3329–3340. https://doi.org/10.1016/j.rse.2008.02.014.

Gordon, H., Voss, K., 2004. MODIS Normalized Water-leaving Radiance Algorithm Theoretical Basis Document.

Gordon, H.R., Brown, O.B., Evans, R.H., Brown, J.W., Smith, R.C., Baker, K.S., Clark, D. K., 1988. A semianalytic radiance model of ocean color. J. Geophys. Res. 93, 10909. https://doi.org/10.1029/JD093iD09p10909.

Gordon, H.R., Morel, A.Y., 1983. Remote Assessment of Ocean Color for Interpretation of Satellite Visible Imagery: A review: By H. R. Gordon and A. Y. Morel. Q.J.R. Meteorol. Soc. https://doi.org/10.1002/qj.49711146913.

Gould, R.W., McCarthy, S.C., Coelho, E., Shulman, I., Richman, J.G., 2014. Combining satellite ocean color and hydrodynamic model uncertainties in bio-optical forecasts. J. Appl. Remote. Sens. 8, 083652 https://doi.org/10.1117/1.JRS.8.083652.

Grunert, B.K., Mouw, C.B., Ciochetto, A.B., 2018. Characterizing CDOM spectral variability across diverse regions and spectral ranges: characterizing CDOM spectral variability. Glob. Biogeochem. Cycles 32, 57–77. https://doi.org/10.1002/2017GB005756.

Hestir, E.L., Brando, V., Campbell, G., Dekker, A., Malthus, T., 2015. The relationship between dissolved organic matter absorption and dissolved organic carbon in reservoirs along a temperate to tropical gradient. Remote Sens. Environ. 156, 395–402. https://doi.org/10.1016/j.rse.2014.09.022.

Hill, V.J., 2008. Impacts of chromophoric dissolved organic material on surface ocean heating in the Chukchi Sea. J. Geophys. Res. 113, C07024. https://doi.org/10.1029/2007JC004119.

Horowitz, A.J., 2008. Determining annual suspended sediment and sediment-associated trace element and nutrient fluxes. Sci. Total Environ. 400, 315–343. https://doi.org/10.1016/j.scitotenv.2008.04.022.

Huot, Y., Babin, M., Bruyant, F., Grob, C., Twardowski, M.S., Claustre, H., 2007. Relationship between photosynthetic parameters and different proxies of phytoplankton biomass in the subtropical ocean. Biogeosciences 4, 853–868. https://doi.org/10.5194/bg-4-853-2007.

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. https://doi.org/10.1016/j.rse.2017.10.041.

IOCCG, 2014. Phytoplankton functional types from space. In: Sathyendranath, S. (Ed.), Reports of the International Ocean-Colour Coordinating Group, No. 15. IOCCG, Dartmouth, Canada, p. 2014.

IOCCG, 2019. Uncertainties in ocean colour remote sensing. In: Mélin, F. (Ed.), IOCCG Report Series, No. 18, International Ocean Colour Coordinating Group, Dartmouth, Canada. https://doi.org/10.25607/OBP-696 n.d.

Jiang, D., Matsushita, B., Pahlevan, N., Gurlin, D., Lehmann, M.K., Fichot, C.G., Schalles, J., Loisel, H., Binding, C., Zhang, Y., Alikas, K., Kangro, K., Uusõue, M., Ondrusek, M., Greb, S., Moses, W.J., Lohrenz, S., O'Donnell, D., 2021. Remotely estimating total suspended solids concentration in clear to extremely turbid waters using a novel semi-analytical method. Remote Sens. Environ. 258, 112386 https://doi.org/10.1016/j.rse.2021.112386.

Kahru, M., Leppanen, J.-M., Rud, O., 1993. Cyanobacterial blooms cause heating of the sea surface. Mar. Ecol. Prog. Ser. 101, 1–7. https://doi.org/10.3354/meps101001.

Kasprzak, P., Padisák, J., Koschel, R., Krienitz, L., Gervais, F., 2008. Chlorophyll a concentration across a trophic gradient of lakes: an estimator of phytoplankton biomass? Limnologica 38, 327–338. https://doi.org/10.1016/j.limno.2008.07.002.

Keith, D.J., Schaeffer, B.A., Lunetta, R.S., Gould, 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. https://doi.org/10.1080/01431161.2014.894663.

King, G., Honaker, J., Joseph, A., Scheve, K., 2001. Analyzing incomplete political science data: an alternative algorithm for multiple imputation. Am. Polit. Sci. Rev. 95, 49–69. https://doi.org/10.1017/S0003055401000235.

Kowalczuk, P., Darecki, M., Zabłocka, M., Górecka, I., 2010. Validation of empirical and semi-analytical remote sensing algorithms for estimating absorption by coloured dissolved organic matter in the Baltic Sea from SeaWiFS and MODIS imagery. Oceanologia 52, 171–196. https://doi.org/10.5697/oc.52-2.171.

Kyryliuk, D., Kratzer, S., 2019. Summer distribution of Total suspended matter across the Baltic Sea. Front. Mar. Sci. 5, 504. https://doi.org/10.3389/fmars.2018.00504.

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. https://doi.org/10.1364/AO.41.005755.

Lee, Z., Carder, K.L., Mobley, C.D., Steward, R.G., Patch, J.S., 1999. Hyperspectral remote sensing for shallow waters: 2 deriving bottom depths and water properties by optimization. Appl. Opt. 38, 3831. https://doi.org/10.1364/AO.38.003831.

Lee, Z., Lubac, B., Werdell, J., Arnone, R., 2023. An Update of the Quasi-Analytical Algorithm (QAA_v5.

Lee, Z.P., Carder, K.L., Marra, J., Steward, R.G., Perry, M.J., 1996. Estimating primary production at depth from remote sensing. Appl. Opt. 35, 463. https://doi.org/10.1364/AO.35.000463.

Lehmann, M.K., Gurlin, D., Pahlevan, N., Alikas, K., Anstee, J., Balasubramanian, S.V., Barbosa, C.C.F., Binding, C., Bracher, A., Bresciani, M., Burtner, A., Cao, Z., Dekker, A.G., Drayson, N., Errera, R.M., Fernandez, V., Fichot, C.G., Gege, P., Giardino, C., Gitelson, A.A., Greb, S.R., Henderson, H., Higa, H., Irani Rahaghi, A., Jamet, C., Jiang, D., Kangro, K., Kudela, R., Li, L., Ligi, M., Loisel, H., Lohrenz, S., Ma, R., Maciel, D.A., Malthus, T.J., Matsushita, B., Minaudo, C., Mishra, D.R., Mishra, S., Moore, T., Moses, W.J., Nguyễn, H., Novo, E.M.L.M., Novoa, S., Odermatt, D., O'Donnell, D.M., Olmanson, L.G., Ondrusek, M., Oppelt, N., Pereira Filho, W., Plattner, S., Ruiz Verdú, A., Salem, S.I., Schalles, J.F., Simis, S.G.H., Siswanto, E., Smith, B., Somlai-Schweiger, I., Soppa, M.A., Spyrakos, E., van der Woerd, H.J., Vander Woude, A., Vantrepotte, V., Wernand, M.R., Werther, M., Yue, L., Jordan, T., Kravitz, J.A., Kristoffersen, A.S., Matthews, M., Tessin, E., Vandermeulen, R.A., Ficek, D., Di Vittorio, C., Young, K., 2023. GLORIA - A globally representative hyperspectral in situ dataset for optical sensing of water quality. Scientific Data (in press). https://doi.org/10.1038/s41597-023-01973-y.

López-Urrutia, Á., Morán, X.A.G., 2015. Temperature affects the size-structure of phytoplankton communities in the ocean: temperature and phytoplankton size. Limnol. Oceanogr. 60, 733–738. https://doi.org/10.1002/lno.10049.

Lotliker, A.A., Baliarsingh, S.K., Sahu, K.C., Kumar, T.S., 2015. Performance of semianalytical algorithm and associated inherent optical properties in coastal waters of North Western Bay of Bengal. J. Indian Soc. Remote Sens. 43, 143–149. https://doi.org/10.1007/s12524-014-0399-z.

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. https://doi.org/10.1364/AO.50.001501.

Mannino, A., Russ, M.E., Hooker, S.B., 2008. Algorithm development and validation for satellite-derived distributions of DOC and CDOM in the U.S. Middle Atlantic bight. J. Geophys. Res. 113, C07051. https://doi.org/10.1029/2007JC004493.

Mélin, F., Vantrepotte, V., 2015. How optically diverse is the coastal ocean? Remote SensEnviron. 160, 235–251. https://doi.org/10.1016/j.rse.2015.01.023.

924pp. In: Mobley, C.D. (Ed.), 2022. The Oceanic Optics Book. International Ocean Colour Co-ordinating Group (IOCCG), Dartmouth, NS, Canada. https://doi.org/10.25607/OBP-1710. n.d.

Mobley, C.D., 1999. Estimation of the remote-sensing reflectance from above-surface measurements. Appl. Opt. 38, 7442. https://doi.org/10.1364/AO.38.007442.

Mobley, C.D., 1994. Light and Water: Radiative Transfer in Natural Waters. Academic Press.

Mobley, C.D., Sundman, L.K., Boss, E., 2002. Phase function effects on oceanic light fields. Appl. Opt. 41, 1035. https://doi.org/10.1364/AO.41.001035.
Montes-Hugo, M., Xie, H., 2015. An inversion model based on salinity and remote sensing reflectance for estimating the phytoplankton absorption coefficient in the Saint Lawrence E stuary. J. Geophys. Res. Oceans 120, 6958–6970. https://doi.org/10.1002/2015JC011079.
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, 300. https://doi.org/10.3389/fmars.2017.00300.
Morel, A., Maritorena, S., 2001. Bio-optical properties of oceanic waters: a reappraisal. J. Geophys. Res. Oceans 106, 7163–7180. https://doi.org/10.1029/2000JC000319.
Morley, S.K., Brito, T.V., Welling, D.T., 2018. Measures of model performance based on the log accuracy ratio. Space Weather 16, 69–88. https://doi.org/10.1002/2017SW001669.
Moses, W.J., Bowles, J.H., Lucke, R.L., Corson, M.R., 2012. 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. https://doi.org/10.1364/OE.20.004309.
Muller-Karger, F.E., Hestir, E., Ade, C., Turpie, K., Roberts, D.A., Siegel, D., Miller, R.J., Humm, D., Izenberg, N., Keller, M., Morgan, F., Frouin, R., Dekker, A.G., Gardner, R., Goodman, J., Schaeffer, B., Franz, B.A., Pahlevan, N., Mannino, A.G., Concha, J.A., Ackleson, S.G., Cavanaugh, K.C., Romanou, A., Tzortziou, M., Boss, E.S., Pavlick, R., Freeman, A., Rousseaux, C.S., Dunne, J., Long, M.C., Klein, E., McKinley, G.A., Goes, J., Letelier, R., Kavanaugh, M., Roffer, M., Bracher, A., Arrigo, K.R., Dierssen, H., Zhang, X., Davis, F.W., Best, B., Guralnick, R., Moisan, J., Sosik, H.M., Kudela, R., Mouw, C.B., Barnard, A.H., Palacios, S., Roesler, C., Drakou, E.G., Appeltans, W., Jetz, W., 2018. Satellite sensor requirements for monitoring essential biodiversity variables of coastal ecosystems. Ecol. Appl. 28, 749–760. https://doi.org/10.1002/eap.1682.
Najah, A., Al-Shehhi, M.R., 2021. Performance of the ocean color algorithms: QAA, GSM, and GIOP in inland and coastal waters. Remote Sens. Earth Syst. Sci. 4, 235–248. https://doi.org/10.1007/s41976-022-00068-3.
Nechad, B., Ruddick, K., Schroeder, T., Oubelkheir, K., Blondeau-Patissier, D., Cherukuru, N., Brando, V., Dekker, A., Clementson, L., Banks, A.C., Maritorena, S., Werdell, P.J., Sá, C., Brotas, V., Caballero de Frutos, I., Ahn, Y.-H., Salama, S., Tilstone, G., Martinez-Vicente, V., Foley, D., McKibben, M., Nahorniak, J., Peterson, T., Siliò-Calzada, A., Röttgers, R., Lee, Z., Peters, M., Brockmann, C., 2015. CoastColour round Robin data sets: a database to evaluate the performance of algorithms for the retrieval of water quality parameters in coastal waters. Earth Syst. Sci. Data 7, 319–348. https://doi.org/10.5194/essd-7-319-2015.
Nechad, B., Ruddick, K.G., Park, Y., 2010. Calibration and validation of a generic multisensor algorithm for mapping of total suspended matter in turbid waters. Remote Sens. Environ. 114, 854–866. https://doi.org/10.1016/j.rse.2009.11.022.
Novoa, S., Doxaran, D., Ody, A., Vanhellemont, Q., Lafon, V., Lubac, B., Gernez, P., 2017. Atmospheric corrections and multi-conditional algorithm for multi-sensor remote sensing of suspended particulate matter in low-to-high turbidity levels coastal waters. Remote Sens. 9, 61. https://doi.org/10.3390/rs9010061.
Olli, K., Ptacnik, R., Klais, R., Tamminen, T., 2019. Phytoplankton species richness along coastal and estuarine salinity continua. Am. Nat. 194, E41–E51. https://doi.org/10.1086/703657.
O'Shea, R.E., Pahlevan, N., Smith, B., Bresciani, M., Egerton, T., Giardino, C., Li, L., Moore, T., Ruiz-Verdu, A., Ruberg, S., Simis, S.G.H., Stumpf, R., Vaičiūtė, D., 2021. Advancing cyanobacteria biomass estimation from hyperspectral observations: demonstrations with HICO and PRISMA imagery. Remote Sens. Environ. 266, 112693 https://doi.org/10.1016/j.rse.2021.112693.
Pahlevan, N., Smith, B., Alikas, K., Anstee, J., Barbosa, C., Binding, C., Bresciani, M., Cremella, B., Giardino, C., Gurlin, D., Fernandez, V., Jamet, C., Kangro, K., Lehmann, M.K., Loisel, H., Matsushita, B., Hà, N., Olmanson, L., Potvin, G., Simis, S. G.H., VanderWoude, A., Vantrepotte, V., Ruiz-Verdù, A., 2022. Simultaneous retrieval of selected optical water quality indicators from Landsat-8, Sentinel-2, and Sentinel-3. Remote Sens. Environ. 270, 112860 https://doi.org/10.1016/j.rse.2021.112860.
Pahlevan, N., Smith, B., Binding, C., Gurlin, D., Li, L., Bresciani, M., Giardino, C., 2021. Hyperspectral retrievals of phytoplankton absorption and chlorophyll-A in inland and nearshore coastal waters. Remote Sens. Environ. 253, 112200 https://doi.org/10.1016/j.rse.2020.112200.
Pahlevan, N., Smith, B., Schalles, J., Binding, C., Cao, Z., Ma, R., Alikas, K., Kangro, K., Gurlin, D., Hà, N., Matsushita, B., Moses, W., Greb, S., Lehmann, M.K., Ondrusek, M., Oppelt, N., Stumpf, R., 2020. Seamless retrievals of chlorophyll-a from Sentinel-2 (MSI) and Sentinel-3 (OLCI) in inland and coastal waters: a machine-learning approach. Remote Sens. Environ. 240, 111604 https://doi.org/10.1016/j.rse.2019.111604.
Palmer, S.C.J., Kutser, T., Hunter, P.D., 2015. Remote sensing of inland waters: challenges, progress and future directions. Remote Sens. Environ. 157, 1–8. https://doi.org/10.1016/j.rse.2014.09.021.
Paltsev, A., Creed, I.F., 2022. Multi-decadal changes in phytoplankton biomass in northern temperate lakes as seen through the prism of landscape properties. Glob. Change Biol. 28, 2272–2285. https://doi.org/10.1111/gcb.16079.
Pan, H., Lyu, H., Wang, Y., Jin, Q., Wang, Q., Li, Y., Fu, Q., 2015. An improved approach to retrieve IOPs based on a quasi-analytical algorithm (QAA) for turbid eutrophic inland water. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 8, 5177–5189. https://doi.org/10.1109/JSTARS.2015.2480102.
Werdell, P.J., Bailey, S.W., 2002. The SeaWiFS Bio-optical Archive and Storage System (SeaBASS): Current architecture and implementation, NASA Tech. Memo. 2002-211617, 45 pp. In: Fargion, G.S., McClain, C.R. (Eds.), NASA Goddard Space Flight Center, Greenbelt, Maryland. n.d.
Pope, R.M., Fry, E.S., 1997. Absorption spectrum (380–700 nm) of pure water II integrating cavity measurements. Appl. Opt. 36, 8710. https://doi.org/10.1364/AO.36.008710.
Qin, Y., Brando, V.E., Dekker, A.G., Blondeau-Patissier, D., 2007. Validity of SeaDAS water constituents retrieval algorithms in australian tropical coastal waters. Geophys. Res. Lett. 34, L21603. https://doi.org/10.1029/2007GL030599.
Qing, S., Tang, J., Cui, T., Zhang, J., 2011. Retrieval of inherent optical properties of the Yellow Sea and East China Sea using a quasi-analytical algorithm. Chin. J. Oceanol. Limnol. 29, 33–45. https://doi.org/10.1007/s00343-011-9967-z.
Richter, R., Schläpfer, D., 2002. Geo-atmospheric processing of airborne imaging spectrometry data. Part 2: Atmospheric/topographic correction. Int. J. Remote Sens. 23, 2631–2649. https://doi.org/10.1080/01431160110115834.
Robinson, C.M., Huot, Y., Schuback, N., Ryan-Keogh, T.J., Thomalla, S.J., Antoine, D., 2021. High latitude Southern Ocean phytoplankton have distinctive bio-optical properties. Opt. Express 29, 21084. https://doi.org/10.1364/OE.426737.
Rubin, D.B., 2004. Multiple Imputation for Nonresponse in Surveys. Wiley Classics Library. Wiley-Interscience, Hoboken, N.J.
Ruiz-Verdú, A., Simis, S.G.H., de Hoyos, C., Gons, H.J., Peña-Martínez, R., 2008. An evaluation of algorithms for the remote sensing of cyanobacterial biomass. Remote Sens. Environ. 112, 3996–4008. https://doi.org/10.1016/j.rse.2007.11.019.
Saranathan, A.M., Smith, B., Pahlevan, N., 2023. Per-pixel uncertainty quantification and reporting for satellite-derived chlorophyll-a estimates via misture density networks. Mixture + IEEE Transactions and Geoscience and Remote Sensing (Under Review).
Sayers, M.J., Fahnenstiel, G.L., Shuchman, R.A., Bosse, K.R., 2021. A new method to estimate global freshwater phytoplankton carbon fixation using satellite remote sensing: initial results. Int. J. Remote Sens. 42, 3708–3730. https://doi.org/10.1080/01431161.2021.1880661.
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. https://doi.org/10.1080/01431161.2015.1035408.
Schaeffer, B.A., Schaeffer, K.G., Keith, D., Lunetta, R.S., Conmy, R., Gould, R.W., 2013. Barriers to adopting satellite remote sensing for water quality management. Int. J. Remote Sens. 34, 7534–7544. https://doi.org/10.1080/01431161.2013.823524.
Schalles, J., Yacobi, Y.Z., 2000. Remote detection and seasonal patterns of phycocyanin, carotenoid and chlorophyll pigments in eutrophic waters. Arch.of Hydrobio. Spec. Issues: Advanc. Limnol. 153–168.
Sheng, M.A., Xiaofeng, Y., Zui, T., Ziwei, L., Xuan, Z., 2014. Assessment of uncertainties of ocean color parameters for the ocean carbon-based productivity model. IOP Conf. Ser. Earth Environ. Sci. 17, 012102 https://doi.org/10.1088/1755-1315/17/1/012102.
Simis, S.G.H., Peters, S.W.M., Gons, H.J., 2005. Remote sensing of the cyanobacterial pigment phycocyanin in turbid inland water. Limnol. Oceanogr. 50, 237–245. https://doi.org/10.4319/lo.2005.50.1.0237.
Simis, S.G.H., Ruiz-Verdú, A., Domínguez-Gómez, J.A., Peña-Martinez, R., Peters, S.W. M., Gons, H.J., 2007. Influence of phytoplankton pigment composition on remote sensing of cyanobacterial biomass. Remote Sens. Environ. 106, 414–427. https://doi.org/10.1016/j.rse.2006.09.008.
Smith, B., Pahlevan, N., Schalles, J., Ruberg, S., Errera, R., Ma, R., Giardino, C., Bresciani, M., Barbosa, C., Moore, T., Fernandez, V., Alikas, K., Kangro, K., 2021. A chlorophyll-a algorithm for Landsat-8 based on mixture density networks. Front. Remote Sens. 1, 623678 https://doi.org/10.3389/frsen.2020.623678.
Smith, M.E., Robertson Lain, L., Bernard, S., 2018. An optimized chlorophyll a switching algorithm for MERIS and OLCI in phytoplankton-dominated waters. Remote Sens. Environ. 215, 217–227. https://doi.org/10.1016/j.rse.2018.06.002.
Smith, R.C., Baker, K.S., 1981. Optical properties of the clearest natural waters (200–800 nm). Appl. Opt. 20, 177. https://doi.org/10.1364/AO.20.000177.
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. https://doi.org/10.1016/j.neucom.2015.03.108.
Spencer, R.G.M., Butler, K.D., Aiken, G.R., 2012. Dissolved organic carbon and chromophoric dissolved organic matter properties of rivers in the USA: DISSOLVED ORGANIC MATTER IN U.S. RIVERS. J. Geophys. Res. Biogeosciences 117. https://doi.org/10.1029/2011JG001928 n/a-n/a.
Spyrakos, E., O'Donnell, R., Hunter, P.D., Miller, C., Scott, M., Simis, S.G.H., Neil, C., Barbosa, C.C.F., Binding, C.E., Bradt, S., Bresciani, M., Dall'Olmo, G., Giardino, C., Gitelson, A.A., Kutser, T., Li, L., Matsushita, B., Martinez-Vicente, V., Matthews, M. W., Ogashawara, I., Ruiz-Verdú, A., Schalles, J.F., Tebbs, E., Zhang, Y., Tyler, A.N., 2018. Optical types of inland and coastal waters: optical types of inland and coastal waters. Limnol. Oceanogr. 63, 846–870. https://doi.org/10.1002/lno.10674.
Stramski, D., Babin, M., Woźniak, S.B., 2007. Variations in the optical properties of terrigenous mineral-rich particulate matter suspended in seawater. Limnol. Oceanogr. 52, 2418–2433. https://doi.org/10.4319/lo.2007.52.6.2418.
Stroming, S., Robertson, M., Mabee, B., Kuwayama, Y., Schaeffer, B., 2020. Quantifying the human health benefits of using satellite information to detect cyanobacterial harmful algal blooms and manage recreational advisories in U.S. lakes. GeoHealth 4. https://doi.org/10.1029/2020GH000254.
Sun, D., Li, Y., Wang, Q., Lv, H., Le, C., Huang, C., Gong, S., 2010. Partitioning particulate scattering and absorption into contributions of phytoplankton and non-algal particles in winter in Lake taihu (China). Hydrobiologia 644, 337–349. https://doi.org/10.1007/s10750-010-0198-7.
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. https://doi.org/10.1364/AO.43.002156.

Twardowski, M., Tonizzo, A., 2018. Ocean color analytical model explicitly dependent on the volume scattering function. Appl. Sci. 8, 2684. https://doi.org/10.3390/app8122684.

Twardowski, M., Zhang, X., Vagle, S., Sullivan, J., Freeman, S., Czerski, H., You, Y., Bi, L., Kattawar, G., 2012. The optical volume scattering function in a surf zone inverted to derive sediment and bubble particle subpopulations: volume scattering function and inversion. J. Geophys. Res. Oceans 117. https://doi.org/10.1029/2011JC007347 n/a-n/a.

Twardowski, M.S., Boss, E., Sullivan, J.M., Donaghay, P.L., 2004. Modeling the spectral shape of absorption by chromophoric dissolved organic matter. Mar. Chem. 89, 69–88. https://doi.org/10.1016/j.marchem.2004.02.008.

Vaičiūtė, D., Bučas, M., Bresciani, M., Dabulevičienė, T., Gintauskas, J., Mėžinė, J., Tiškus, E., Umgiesser, G., Morkūnas, J., De Santi, F., Bartoli, M., 2021. Hot moments and hotspots of cyanobacteria hyperblooms in the curonian lagoon (SE Baltic Sea) revealed via remote sensing-based retrospective analysis. Sci. Total Environ. 769, 145053 https://doi.org/10.1016/j.scitotenv.2021.145053.

Wei, J., Lee, Z., Shang, S., 2016. A system to measure the data quality of spectral remote sensing reflectance of aquatic environments. J. Geophys. Res. Oceans. https://doi.org/10.1002/2016JC012126.

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., Loisel, H., Maritorena, S., Mélin, F., Moore, T.S., Smyth, T.J., Antoine, D., Devred, E., d'Andon, O.H.F., Mangin, A., 2013. Generalized Ocean color inversion model for retrieving marine inherent optical properties. Appl. Opt. 52, 2019. https://doi.org/10.1364/AO.52.002019.

Werdell, P.J., McKinna, L.I.W., Boss, E., Ackleson, S.G., Craig, S.E., Gregg, W.W., Lee, Z., Maritorena, S., Roesler, C.S., Rousseaux, C.S., Stramski, D., Sullivan, J.M., Twardowski, M.S., Tzortziou, M., Zhang, X., 2018. An overview of approaches and challenges for retrieving marine inherent optical properties from ocean color remote sensing. Prog. Oceanogr. 160, 186–212. https://doi.org/10.1016/j.pocean.2018.01.001.

Woźniak, M., Bradtke, K., Darecki, M., Krężel, A., 2016. Empirical model for phycocyanin concentration estimation as an indicator of cyanobacterial bloom in the optically complex coastal waters of the Baltic Sea. Remote Sens. 8, 212. https://doi.org/10.3390/rs8030212.

Yang, H., Choi, J.-K., Park, Y.-J., Han, H.-J., Ryu, J.-H., 2014. Application of the Geostationary Ocean color imager (GOCI) to estimates of ocean surface currents. J. Geophys. Res. Oceans 119, 3988–4000. https://doi.org/10.1002/2014JC009981.

Yang, W., Matsushita, B., Chen, J., Yoshimura, K., Fukushima, T., 2013. Retrieval of inherent optical properties for turbid inland waters from remote-sensing reflectance. IEEE Trans. Geosci. Remote Sens. 51, 3761–3773. https://doi.org/10.1109/TGRS.2012.2220147.

Zemlys, P., Ferrarin, C., Umgiesser, G., Gulbinskas, S., Bellafiore, D., 2013. Investigation of saline water intrusions into the curonian lagoon (Lithuania) and two-layer flow in the Klaipėda Strait using finite element hydrodynamic model. Ocean Sci. 9, 573–584. https://doi.org/10.5194/os-9-573-2013.

Zhang, X., Hu, L., He, M.-X., 2009. Scattering by pure seawater: effect of salinity. Opt. Express 17, 5698. https://doi.org/10.1364/OE.17.005698.

Zheng, G., DiGiacomo, P.M., 2018. Detecting phytoplankton diatom fraction based on the spectral shape of satellite-derived algal light absorption coefficient: algae type from the absorption spectral shape. Limnol. Oceanogr. 63, S85–S98. https://doi.org/10.1002/lno.10725.

Zhu, W., Yu, Q., 2013. Inversion of chromophoric dissolved organic matter from EO-1 hyperion imagery for turbid estuarine and coastal waters. Geosci. Remote Sens. IEEE Trans. On 51, 3286–3298. https://doi.org/10.1109/TGRS.2012.2224117.

Zimba, P.V., 2012. An improved phycobilin extraction method. Harmful Algae 17, 35–39. https://doi.org/10.1016/j.hal.2012.02.009.