Global distribution of the spectral power coefficient of particulate backscattering coefficient obtained by a neural network scheme

Xiaolong Yu [*], Zhongping Lee, Wendian Lai

State Key Laboratory of Marine Environmental Science, College of Ocean and Earth Sciences, Xiamen University, Xiamen 361101, China

ARTICLE INFO

ABSTRACT

The spectral power coefficient (η) of the particulate backscattering coefficient ($b_{bp}(\lambda)$) is directly estimated from the remote sensing reflectance ($R_{rs}(\lambda)$) with a neural network-based scheme (NN_η) in this study. Evaluations with both synthetic dataset and *in-situ* measurements show that NN_η could significantly improve the accuracy of estimated η compared to several conventional schemes reported in the literature that are based on chlorophyll-a concentration (Chl), band ratios of $R_{rs}(\lambda)$, or remotely sensed $b_{bp}(\lambda)$. Demonstrations with measurements from MODerate resolution Imaging Spectroradiometer (MODIS) Aqua further confirm the robustness of NN_η, where reasonable spatial distribution and seasonality of η in the global oceans can be acquired by NN_η. High and low η values are observed in the oligotrophic gyres and the coastal zones, respectively, which are consistent with the current understanding of η distribution concluded from theoretic analysis and repeated field measurements in the global ocean. Implementation of NN_η to 19-year MODIS monthly composite measurements from 2003 to 2021 reveals strong seasonal variations of η in most of the global ocean, but the decadal changes of η are insignificant in the majority (~82.2%) of the global ocean. Similar to any empirical algorithms, the performance of NN_η is dependent on the training dataset, particularly its range, here a proper upper limit of η for natural waters is provided.

1. Introduction

1.1. Importance of η

The spectral backscattering coefficient of marine particles ($b_{bp}(\lambda)$, in m^{-1}) is of paramount importance to applications of optical remote sensing in oceanography as the remote sensing reflectance ($R_{rs}(\lambda)$, in sr^{-1}) of the ocean (ocean color) is proportional to $b_{bp}(\lambda)$ (Morel and Prieur, 1977; Gordon et al., 1988). Also, as a proxy of the suspended matter in the ocean, $b_{bp}(\lambda)$ allows for the synoptic estimation of suspended particle mass (Boss et al., 2009), particulate organic matter (POC) concentration (Stramski et al., 1999; Stramski et al., 2008), and primary production (PP) in global oceans (Behrenfeld et al., 2005; Westberry et al., 2008), which greatly promotes the understanding of the role of ocean biota in the carbon cycle and climate processes (Stramski et al., 2004). Theoretical simulations and field measurements indicate that $b_{bp}(\lambda)$, in general, decreases toward longer wavelengths (λ, in nm) and can be described as a power-law function of λ (Gordon and Morel, 1983; Gordon et al., 1988)

$$b_{bp}(\lambda) = b_{bp}(\lambda_0)(\lambda/\lambda_0)^{-\eta} \qquad (1)$$

where λ_0 is a reference wavelength and η (dimensionless) is the power coefficient of spectral $b_{bp}(\lambda)$, with $b_{bp}(\lambda_0)$ determining the magnitude of b_{bp}, while η determines the spectral shape. Thus, an accurate estimation of η is critical for the determination of $b_{bp}(\lambda)$. Also, because $R_{rs}(\lambda)$ is a function of the total absorption ($a(\lambda)$, in m^{-1}) and backscattering coefficients ($b_b(\lambda)$, in m^{-1}), which is calculated as the sum of $b_{bp}(\lambda)$ and the backscattering coefficients of pure seawater ($b_{bw}(\lambda)$, in m^{-1}), known $b_{bp}(\lambda)$ enables an algebraic calculation of $a(\lambda)$ from $R_{rs}(\lambda)$ (Lee et al., 2002; Lee et al., 2004). Therefore, η is also one of the uncertainty sources in remotely sensed $a(\lambda)$ (Lee et al., 2010). In addition, the selection of η value could be of particular importance for optical-biogeochemical modeling (Fujii et al., 2007; Terzić et al., 2021), as well as for the closure studies between apparent optical properties (AOPs) and inherent optical properties (IOPs) (Tzortziou et al., 2006; Chang et al., 2007; Gallegos et al., 2008; Chang and Whitmire, 2009).

Apart from modeling the spectrum of $b_{bp}(\lambda)$ from remote sensing, the value of η itself is very sensitive to the composition and size of marine particles, particularly the relative contribution of small-sized particles to

---

* Corresponding author.
  E-mail address: xlyu@xmu.edu.cn (X. Yu).

total particle concentrations, with small η values indicating the dominance of large-sized particles (Reynolds et al., 2001). Thus, η has been widely used to estimate the particle composition (Reynolds et al., 2016), the Junge slope of the particle size distribution (Kostadinov et al., 2009; Reynolds et al., 2010; Brewin et al., 2011; Kostadinov et al., 2012), and the phytoplankton function types (PFTs) (Kostadinov et al., 2010; Fujiwara et al., 2011). Thus, accurate estimation of η is also strongly desired for the characterization of particle-related properties in global oceans.

Acquiring η *via in-situ* measurements with oceanographic cruises could be quite limited in describing the temporal and large-scale spatial variability of η in the world's oceans, while remote sensing of η provides a more effective measure. Conventionally, η was estimated empirically from ocean color measurements, such as the chlorophyll-a concentration (Ciotti et al., 1999; Morel and Maritorena, 2001), $b_{bp}(\lambda)$ at a reference wavelength (Reynolds et al., 2001), and the band ratio of $R_{rs}(\lambda)$ (Carder et al., 1999; Lee et al., 2002). Alternatively, η can be calculated by fitting remotely sensed $b_{bp}(\lambda)$ spectrum to the power-law function of Eq. (1) over specific ranges of wavelengths (Loisel et al., 2006; Jorge et al., 2021), or obtained from the inversion of IOPs using spectral optimization when η is set as a free variable (Yu et al., 2016). However, η derived from the above-mentioned schemes could vary largely from each other in either the spatial distribution or the ranges of estimated η in global oceans (Jorge et al., 2021), it begs a further and more comprehensive investigation of the retrieval schemes of η, so that more consistent and reasonable η products can be used for subsequent applications. Our aim of this study is to propose a new scheme to directly estimate η from $R_{rs}(\lambda)$ based on a neural network and to evaluate its performance with both *in-situ* measurements and satellite observations from MODerate resolution Imaging Spectroradiometer (MODIS) Aqua. Detailed descriptions of these schemes are given below.

1.2. Candidate schemes to estimate η

1.2.1. Empirical scheme based on the band ratio of $R_{rs}(\lambda)$

The value of η could not be solved analytically from an $R_{rs}(\lambda)$ spectrum, rather it is commonly estimated empirically, *via* one step or multi-steps, from $R_{rs}(\lambda)$. In the quasi-analytical algorithm (QAA) for the inversion of IOPs (Lee et al., 2002; Lee et al., 2007), η is estimated directly from the blue-green band ratio of $R_{rs}(\lambda)$. For the band configuration of MODIS, η is estimated as

$$\eta = 2.0\left(1 - 1.2exp\left(-0.9\frac{r_{rs}(443)}{r_{rs}(547)}\right)\right) \tag{2}$$

where $r_{rs}(\lambda)$ is the remote sensing reflectance just beneath the water surface and can be converted from $R_{rs}(\lambda)$ (Lee et al., 2002). The scheme based on Eq. (2) is hereafter referred to as QAA, and the range of estimated η by Eq. (2) is from -0.5 to 2.0.

1.2.2. Empirical scheme based on chlorophyll-a concentration

For approaches taking multiple steps, one is to estimate η empirically from the remotely sensed chlorophyll-a concentration (Chl, in mg m^{-3}), with Chl obtained empirically from $R_{rs}(\lambda)$ in the first place using retrieval algorithms. Here, Chl is estimated by the Ocean Color Index (OCI) algorithm (O'Reilly et al., 1998; Hu et al., 2012), while the empirical formula of Morel and Maritorena (2001) is employed to estimate η from Chl,

$$\eta = 0.5(0.3 - log10(Chl)), 0.02 < Chl < 2.0 \text{ mg m}^{-3} \tag{3}$$

$$\eta = 0, Chl > 2.0 \text{ mg m}^{-3}$$

The scheme based on Eq. (3) is hereafter referred to as MM01, where the estimated η could range between 0 and 1.0. Note that for waters with estimated Chl <0.02 mg m^{-3}, predicted η *via* MM01 is set to 1.0 in this study.

1.2.3. The two-step scheme (LS2)

In a more complex scheme, η can be computed from $b_{bp}(\lambda)$ obtained from inverse algorithms that do not require any assumptions about the spectral shape of $b_{bp}(\lambda)$, while relying on empirically estimated diffuse attenuation coefficient ($K_d(\lambda)$, in m^{-1}) (Loisel and Stramski, 2000; Loisel et al., 2018). Specifically, this scheme takes the following steps to estimate η:

Step 1: $K_d(\lambda)$ at the bands of $R_{rs}(\lambda)$ are first estimated empirically from $R_{rs}(\lambda)$ using either an empirical algorithm (Loisel and Stramski, 2000; Loisel et al., 2001), or based on a neural network (Jamet et al., 2012; Loisel et al., 2018), with these two algorithms termed LS1 and LS2, respectively;

Step 2: $b_b(\lambda)$ are calculated semi-anlytically from $R_{rs}(\lambda)$ and $K_d(\lambda)$ with a look-up-table for empirical coefficients, where $b_{bp}(\lambda)$ can be subsequently obtained as the difference between $b_b(\lambda)$ and $b_{bw}(\lambda)$;

Step 3: η can then be estimated by fitting a linear least squares regression through log-transformed $b_{bp}(\lambda)$ as a function of log-transformed λ.

Since the empirically estimated $K_d(\lambda)$ (Step 1) plays a key role in the calculation of η, this two steps scheme, in essence, could be also considered empirical despite that Step 2 and Step 3 are rooted in ocean optics. Given that LS2 is an improved version of LS1, we only employ LS2 for the schemes inter-comparison in this study. As suggested in Loisel et al. (2006) and Jorge et al. (2021), $b_{bp}(\lambda)$ at blue and red bands are not recommended for the estimation of η, as satellite-measured $R_{rs}(\lambda)$ have relatively large uncertainties in the blue bands due to inadequate atmospheric correction (Wei et al., 2020), while having too low signal in the red bands in open ocean water. Thus, LS2-derived η was computed from $b_{bp}(\lambda)$ at 443, 488, 531, and 547 nm in this effort by linear least squares regression between $log(b_{bp}(\lambda))$ and $log(\lambda)$. Note that only estimated η from a good fit (*i.e.*, the determination coefficient of the linear fit $R^2 > 0.6$) was considered valid. It is worth pointing out that a poor fit does not necessarily mean questionable $b_{bp}(\lambda)$ retrievals by LS2, as many field observations have shown that the spectral $b_{bp}(\lambda)$ are not always in a hyperbolic shape (Kutser et al., 2009; Vadakke-Chanat and Shanmugam, 2019; Xu et al., 2021). However, we include such a quality control measure for a fair comparison with QAA or MM01, as their primary objectives of η estimations are for the inversion of IOPs, which assumed a hyperbolic shape of $b_{bp}(\lambda)$.

1.2.4. One-step scheme based on neural network

The above-mentioned schemes all estimate η from $R_{rs}(\lambda)$, directly or indirectly. As algorithms based on neural network (NN) have been widely accepted by the ocean color community to derive IOPs and AOPs (Ioannou et al., 2013; Chen et al., 2014; Wang et al., 2021), water constituents (Buckton et al., 1999; Tanaka et al., 2004), as well as the depths of optically shallow waters (Lai et al., 2022), it might be a plausible option to estimate η using the neural network by incorporating all available $R_{rs}(\lambda)$ (Lee et al., 2003). Here, we propose an NN-based scheme, termed NN$_\eta$, to predict η using $R_{rs}(\lambda)$ at the six MODIS visible bands centered at 412, 443, 488, 531, 547, and 667 nm. A detailed description of the structure of the neural network is given in Section 2.5.

2. Data and method

2.1. Satellite measurements

$R_{rs}(\lambda)$ products from MODIS Aqua are used in this study to obtain various bio-optical properties and the spatial-temporal variation of η in the global ocean. Specifically, the level-3 monthly $R_{rs}(\lambda)$ products with a spatial resolution of 4 km from January 2003 to December 2021 were acquired from NASA's OceanColor Web (oceancolor.gsfc.nasa.gov). Note that all MODIS data examined in this study were from the latest ocean color reprocessing (R2018.1) by NASA Ocean Biology Processing Group (OBPG). The MODIS 4 km resolution monthly Chl products, derived from $R_{rs}(\lambda)$ using the OCI algorithm (O'Reilly et al., 1998; Hu

et al., 2012), over the same period were also acquired from NASA's OceanColor Web for the analysis in this effort.

As $b_{bw}(\lambda)$ is sensitive to temperature and salinity (Zhang and Hu, 2009; Zhang et al., 2009), Sea Surface Temperature (SST, in $°$) and Sea Surface Salinity (SSS, in PSU or ‰) in the global oceans were also desired for η retrieval, particularly for the Step 2 of LS2. Here, MODIS level-3 monthly standard mapped SST products, as well as the climatology product, with a spatial resolution of 4 km, were acquired from NASA's OceanColor Web from 2003 to 2021. For SSS, ESA Climate Change Initiative (CCI) standard monthly SSS products (v03.21, available at https://climate.esa.int/fr/projects/sea-surface-salinity/), at a spatial resolution of 25 km, were acquired for the period between 2010 and 2020. The SSS data were then interpolated to the 4 km resolution to match the grid of MODIS measurements. The monthly climatology of SSS product over the global ocean was also calculated as the mean value of the monthly SSS data from 2010 to 2020. From the climatology SSS and SST data, the median SSS and SST in the global ocean were calculated, which are $15.2°$ and 34.5 PSU, respectively. The median SSS and SST values will be used to calculate $b_{bw}(\lambda)$ when *in-situ* or satellite-measured SSS and SST are absent unless otherwise stated.

In addition, the MODIS climatology product of phytoplankton absorption coefficient at 443 nm ($a_{ph}(443)$, in m^{-1}) and $b_{bp}(443)$, at 4 km spatial resolution, were also downloaded from NASA's OceanColor Web and used in this study. Note that the inversions of $a_{ph}(443)$ and $b_{bp}(443)$ for the MODIS climatology product employ the generalized IOP (GIOP) model (Werdell et al., 2013).

2.2. In-situ datasets

To acquire concurrent *in-situ* measurements of $R_{rs}(\lambda)$ and η, we assembled a few publicly available datasets for the evaluation of different η schemes, which include the NASA bio-Optical Marine Algorithm Dataset (NOMAD) (Werdell and Bailey, 2005), measurements from the Biogeochemistry and Optics South Pacific Experiment (BIOSOPE) (Claustre et al., 2008; Stramski et al., 2008), and two compiled datasets by Stramski and Reynolds (2018) and Casey et al. (2020). The latter two datasets are termed Stramski18 and Casey20, respectively. The NOMAD and BIOSOPE datasets can be obtained from the Seaviewing Wide Field of View Sensor (SeaWiFS) Bio-optical Archive and Storage System (SeaBASS, seabass.gsfc.nasa.gov), while the Stramski18 and Casey20 datasets can be acquired from their respective references. The following general rules were applied for data screening and quality control (QC) to construct the *in-situ* datasets desired in this effort.

First, all data records with valid observations of $R_{rs}(\lambda)$ at 412, 443, 488, 531, 547, and 667 nm (or nearby equivalent wavelengths) were screened. Second, the quality of spectral $R_{rs}(\lambda)$ was examined by the quality assurance score (QA), which ranges between 0 and 1 with a higher QA score standing for better $R_{rs}(\lambda)$ quality (Wei et al., 2016). Records with QA scores lower than 0.7 were then discarded. Third, $b_{bp}(\lambda)$ measurements should be available at three or more spectral bands, excluding the short-blue (412 nm and shorter wavelengths) and red bands (650 nm and longer wavelengths). η was computed by the linear fitting between $\log(b_{bp}(\lambda))$ and $\log(\lambda)$, where only records with the coefficient of determination (R^2) of the linear regression >0.6 remained.

For the NOMAD dataset, $R_{rs}(\lambda)$ were first screened at 411, 443, 489, 510, 555, and 670 nm from the raw archive (NOMAD v2.b_2010097), which were later interpolated to the six MODIS visible bands. $b_{bp}(\lambda)$ were acquired by subtracting $b_{bw}(\lambda)$ from the fitted $b_b(\lambda)$, at 20 spectral bands from 405 to 683 nm, with $b_{bw}(\lambda)$ calculated from Zhang et al. (2009) with *in-situ* S and T records. For measurements without matched salinity records, the global average salinity of 34.5 PSU was used, which was calculated from measurements with salinity records. η was later computed from $b_{bp}(\lambda)$ between 443 and 590 nm. After screening the QA score of spectral $R_{rs}(\lambda)$ and the R^2 of the linear regression between log ($b_{bp}(\lambda)$) and $\log(\lambda)$, 82 records were discarded and 283 matched $R_{rs}(\lambda)$ and η were retained for NOMAD datasets.

For the BIOSOPE dataset, vertical profiles of $b_b(\lambda)$ were downloaded from the SeaBASS archive (https://seabass.gsfc.nasa.gov/search /experiment/BIOSOPE), which include $b_b(\lambda)$ measurements by Hydroscat-6 at six wavebands (centered at 420, 442, 470, 550, 589, 620, and 671 nm) and two α-βeta sensors for 420 and 510 nm, respectively. In this effort, surface $b_b(\lambda)$ from both the downcast and upcast were first calculated by averaging $b_b(\lambda)$ measurements between 3 m and 5 m beneath the surface. Surface $b_{bp}(\lambda)$ were then obtained by subtracting the values of $b_{bw}(\lambda)$, which were calculated from *in-situ* temperature and salinity records at each station using Zhang et al. (2009). η was later computed from spectral $b_{bp}(\lambda)$ at 442, 470, 510, 550, and 589 nm. The final surface η that matches $R_{rs}(\lambda)$ measurements were averaged from the upcast and downcast measurements, where records with a deviation of η >50% were discarded, with the deviation defined as 200% $\times$ $|\eta_{upcast} -$ $\eta_{downcast}|$ / ($\eta_{upcast} + \eta_{downcast}$). Finally, 19 matched $R_{rs}(\lambda)$ and η records, out of a total number of 24, were retained after the QC of $R_{rs}(\lambda)$ and η calculation.

For the Stramski18 dataset, $R_{rs}(\lambda)$ were first screened at 412, 443, 490, 510, 555, and 670 nm), which were later interpolated to the six MODIS visible bands. Both $b_b(\lambda)$ and $b_{bp}(\lambda)$ are directly available in the Stramski18 dataset at 443, 510, 550, and 671 nm, but with missing $b_{bp}(\lambda)$ measurements at some of these wavelengths. Thus, we first filled in the missing $b_{bp}(\lambda)$ values by subtracting $b_{bw}(\lambda)$ from $b_b(\lambda)$ when $b_b(\lambda)$ are available, where $b_{bw}(\lambda)$ were calculated using the global median SSS and SST following Zhang et al. (2009). Note that inclusion of $b_{bp}(671)$ would result in overall higher R^2 from the linear regression between log ($b_{bp}(\lambda)$) and $\log(\lambda)$ for the Stramski18 dataset, and resulting in less unreasonable η calculations that are either >4 or less than -0.5. Thus, η was computed from $b_{bp}(\lambda)$ at all available wavelengths including 671 nm for the Stramski18 dataset. Note that the BIOSOPE measurements were also compiled in the Stramski18 dataset, which were then removed here. After the QC measures, removing the BIOSOPE measurements, and three outliers with η smaller than -0.5, 91 matched $R_{rs}(\lambda)$ and η, out of 117, were retained for the Stramski18 dataset.

For the Casey20 dataset, $R_{rs}(\lambda)$ were first extracted or interpolated to the six MODIS visible bands and then screened to match $b_{bp}(\lambda)$ measurements. $b_{bp}(\lambda)$, with valid observations at three or more wavelengths between 420 and 640 nm, were used to calculate η. The $R_{rs}(\lambda)$ and η matchups were further screened by the QC measures, with one outlier at station ir2 in Lake Superior removed as the computed η is >6. Finally, 75 matched $R_{rs}(\lambda)$ and η records, out of 93, remained. Note that 10 of the 70 quality-controlled matchups are repeated records of the Stramski18 dataset, and thus removed from the Casey20 dataset. The remaining 60 matchups in the Casey20 dataset are all from measurements in Lake Superior.

In total, 468 matched $R_{rs}(\lambda)$ and η from field measurements were acquired for the subsequent analysis in this effort, with this compiled dataset referred to as the *in-situ* dataset hereafter.

2.3. The lower and upper limits of η in natural waters

The range of η in the training dataset, especially its upper and lower boundaries, is important to the predicted range of η by an NN algorithm and its applicability. Historically, it is assumed that η is around 1.0 for oceanic waters (*i.e.*, λ^{-1} dependency) (Gordon et al., 1988; Morel and Maritorena, 2001), and ~ 0 for turbid coastal waters (Morel and Maritorena, 2001). In recent decades, with the development of multiband instruments to measure $b_{bp}(\lambda)$ *in situ*, such as Hydroscat-6 (HobiLabs, Inc) and ECO BB9 (Sea-Bird Scientific, Inc), values of η can be estimated from measured $b_{bp}(\lambda)$ spectra. The majority of $b_{bp}(\lambda)$ measurements in various oceanic environments showed that η values are generally <3 (Reynolds et al., 2001; Stramska et al., 2003; Aas et al., 2005; Snyder et al., 2008; Kostadinov et al., 2012), while some studies reported maximum η as high as ~ 4 (Loisel et al., 2006; Twardowski et al., 2007; Antoine et al., 2011).

However, our current understanding of the spectral behavior of

$b_{bp}(\lambda)$ in natural waters is quite limited, especially when the exact impacts of nonsphericity and anomalous dispersion from particulate absorption structure on the complex index of refraction spectra are still poorly understood to date (Twardowski et al., 2007). Theoretical modeling, such as the enhanced layered (coated) sphere model, still can not fully resolve the backscattering by real biological cells as the cell thickness, cell shape, and internal cell structure of different cells can not be simply represented in the model (Stramski et al., 2004). Thus, it is almost impossible to obtain a theoretical boundary of η for particles in natural waters from modeling. Moreover, field measurements of $b_{bp}(\lambda)$ in extremely clear waters, such as the ocean gyres, are quite scarce and challenges in obtaining consistent η from measured $b_{bp}(\lambda)$ in these waters further restrain our understanding of the possible upper boundary of η in the natural waters. For instance, η is quite sensitive to the wavelength range and $b_{bp}(\lambda)$ values from which it is computed (McKee et al., 2009). Thus, measurement uncertainties of $b_{bp}(\lambda)$ at individual bands could have impacts on the estimated η value, especially for measurements in extremely clear waters, where very high η values were reported (Morel et al., 2007; Twardowski et al., 2007). On one hand, $b_b(\lambda)$ measurements in the extremely clear waters (<0.001 m^{-1}) would be of relatively large uncertainties introduced by the random noise of the instrument (McKee et al., 2009). The random noise might be partly canceled out with a larger integrating time (Antoine et al., 2011), but still, consistent estimations of η from measured $b_{bp}(\lambda)$ in extremely clear waters are challenging. Take the GYR stations (GYR2 - GYR5) of the Biosope dataset for example, surface η calculated from the upcast and downcast bbp profiles, or deployments of different times at the same GYR station, could be differed by up to 50% even after the strict quality control described in Section 2.2, highlighting the difficulty in obtaining consistent η from field $b_{bp}(\lambda)$ measurements in extremely clear waters. On the other hand, as $b_{bw}(\lambda)$ is the dominant term of $b_b(\lambda)$ in such waters, there could be strong difficulties to get reliable $b_{bp}(\lambda)$ estimations from the difference between $b_b(\lambda)$ and $b_{bw}(\lambda)$, whereas the selection of $b_{bw}(\lambda)$ values could also affect the calculation of η (Morel et al., 2007; Twardowski et al., 2007; Reynolds et al., 2016). Here, we try to reach a reasonable upper boundary of η in natural waters from both field measurements and theoretical simulations.

2.3.1. The range of η from field measurements

We first analyzed the range of η from the field measurements using the *in-situ* dataset compiled in this study. Note that NOMAD, Stramski18, and Casey20 cover a wide range of water types from oceanic to coastal and inland lake waters, while BIOSOPE covers the ultra-oligotrophic ocean. Fig. 1 shows the scatterplots between η and Chl and $b_{bp}(443)$ for the *in-situ* dataset, from which we can conclude three major findings despite the potential uncertainties associated with the field measurements:

1) No clear, robust, relationships between Chl and η, or between $b_{bp}(443)$ and η. These data show that seemingly random variations of η for any given Chl or $b_{bp}(443)$ are remarkably significant, where the standard deviation of η for any given Chl or $b_{bp}(443)$ are 0.42 and 0.5, respectively. Here the standard deviation of η was calculated for different ranges of Chl (or $b_{bp}(443)$) with a spacing of log10(0.1) mg m^{-3} (or m^{-1});

2) η is tilted to higher values for low Chl or low $b_{bp}(443)$, but the deviation from a mean is large. Thus, the overall trend of higher η for oceanic waters and lower value for coastal waters is reasonable, as the sizes of particles in coastal waters are generally greater than that in oceanic waters;

3) It appears that the highest η from these measurements is ~3. More importantly, for ultra-oligotrophic waters in the South Pacific Gyre (SPG) covered in the BIOSOPE dataset, the highest η is only ~2.6, which is within a range of 2.1 and 3.1 after considering the standard deviation associated with these measurements (up to ~0.5 as discussed earlier).

Note that Twardowski et al. (2007) reported that η estimated from the green-red band ratio of measured $b_{bp}(\lambda)$ in the SPG could be up to 4 in the deep chlorophyll maximum (DCM). However, such estimates could be questionable since an λ^{-4} dependency of $b_{bp}(\lambda)$ is almost equivalent to molecule scattering (pure water) (Morel, 1974), which is less likely as DCM is rich in phytoplankton. Besides, as η is quite sensitive to wavelengths of $b_{bp}(\lambda)$ used for its calculation, it might not be a proper comparison between The larger η estimations in Twardowski et al. (2007) could be partly because they calculate η from the band ratio of $b_{bp}(\lambda)$ whereas η is quite sensitive to the wavelengths used for its calculation. For instance, the calculated η from the blue-green band ratio of $b_{bp}(\lambda)$ would be up to 4.6 at the GYR stations in the BIOSOPE dataset.

2.3.2. Implications based on a K_d-closure exercise in the south pacific gyre

Optical closure has been widely used to derive bio-optical properties or evaluate the credibility of *in-situ* bio-optical measurements (Twardowski et al., 2007; Lee et al., 2015). In comparison to the possible uncertainties associated with *in-situ* measured $b_{bp}(\lambda)$ in extremely clear waters, radiance measurements could be of much higher precision and consistency (Wei et al., 2021c). We here tried to gain a further understanding of the upper limit of η through a closure analysis between profiling-measured $K_d(\lambda)$ and that derived from $R_{rs}(\lambda)$ for measurements within SPG from the BIOSOPE dataset, which is considered the 'clearest' natural waters (Morel et al., 2007; Twardowski et al., 2007). Specifically, $a(\lambda)$ and $b_b(\lambda)$ were derived from the measured $R_{rs}(\lambda)$ spectrum
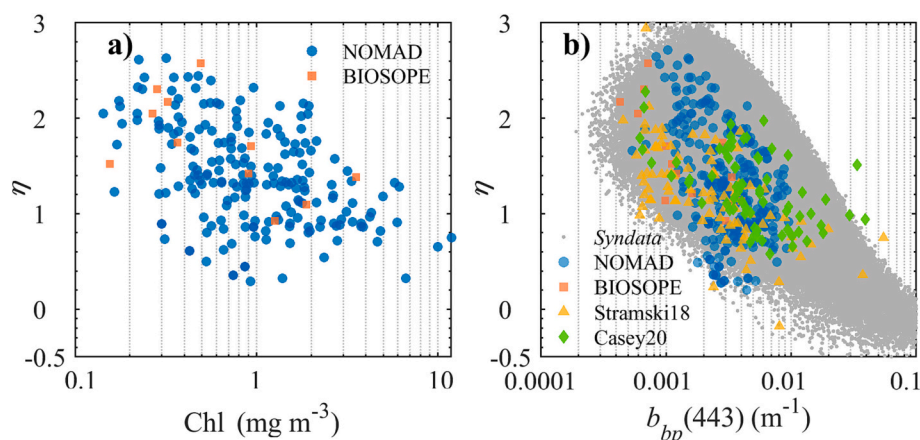

Fig. 1. η as a function of Chl (a) and $b_{bp}(\lambda)$ at 443 nm ($b_{bp}(443)$) (b) for *in-situ* measurements from the NOMAD, BIOSOPE, Stramski18, and Casey20 datasets. The range of η in the synthetic dataset of this study (*Syndata*) is also provided as a function of $b_{bp}(443)$ in panel b for comparison.

using the latest version of QAA (QAA_v6, available at http://www.ioccg.org/groups/Software_OCA/QAA_v6_2014209.pdf) and then used to model $K_d(\lambda)$ with a semi-analytical model (Lee et al., 2005a; Lee et al., 2005b), which is expressed as

$$K_d(\lambda) = (1+0.005\theta_s)a(\lambda) + 4.259\left(1 - 0.265\frac{b_{bw}(\lambda)}{b_b(\lambda)}\right)\left(1 - 0.52e^{-10.8a(\lambda)}\right)b_b(\lambda)$$

(4)

where θ_s (in deg) is the solar zenith angle in air and set as 20° here. Different from the original QAA_v6 where η is parameterized as a function of the band ratio of $R_{rs}(\lambda)$ (see Eq. (2)) for the estimation of spectral b_b, here we arbitrarily set η as 1, 2, and 3, respectively, and compared the resulted $K_d(\lambda)$ spectra with that from field measurements. Concurrent measurements of profile-measured $K_d(\lambda)$ and $R_{rs}(\lambda)$ at 7 stations within the SPG from the BIOSOPE dataset were used here to demonstrate the range of measured and derived $K_d(\lambda)$ spectra, with the station aliases denoted as STB6, STB7, STB8, GYR2, GYR3, GYR4, and GYR5, respectively (locations of these stations can be found in Fig. 5a in Loisel et al. (2006) and spectral $R_{rs}(\lambda)$ can be found in Fig. 2a). As shown in Fig. 2, measured $R_{rs}(\lambda)$ at these selected stations are quite consistent, while measured $K_d(\lambda)$ show significant variations in the blue-green domain, suggesting the potentially large uncertainties in the $K_d(\lambda)$ measurements, which are likely due to wave focusing. Thus, the median $K_d(\lambda)$ measurements are not used as the reference to evaluate the performance of derived $K_d(\lambda)$ in this effort. Alternatively, we assume that the 'true' $K_d(\lambda)$ of these waters should be within the range of measured $K_d(\lambda)$. As shown in Fig. 2b, increasing η values will elevate $K_d(\lambda)$ in the ultraviolet-blue domain. When η is set as 3, the derived $K_d(\lambda)$ values are barely within the range of field-measured $K_d(\lambda)$, suggesting that an η value higher than 3 could result in a more deviated $K_d(\lambda)$ spectrum from the reasonable $K_d(\lambda)$ range measured from the "clearest" natural waters. Thus, we conclude that the upper limit of η in natural waters could be ~3, or it should not be significantly >3. Nevertheless, more rigorous evaluations of the K_d-closure exercise for the "clearest" natural waters are still desired and recommended when high-quality measurements of both $R_{rs}(\lambda)$ and $K_d(\lambda)$ are acquired.

2.4. η for common size distribution based on Mie simulation

The particle size distributions (PSDs) of ocean surface waters have been found to be well approximated by a hyperbolic (Junge-type) distribution (Boss et al., 2001; Twardowski et al., 2001), and the Junge slope of the PSD (ξ) is closely related to η (Kostadinov et al., 2009). Thus, the range of measured ξ in global oceans could provide some insights into the natural range of η. As demonstrated by the simulations based on the forward Mie model (see Fig. 1 of Kostadinov et al. (2009)), an η value of 3 is equivalent to a ξ value of 6, which is less likely encountered in

natural waters. Repeated measurements of ξ in the global ocean showed that ξ is generally <5 for marine particle populations (Twardowski et al., 2001; Babin et al., 2003; Boss et al., 2004; Buonassissi and Dierssen, 2010; Reynolds and Stramski, 2021). More importantly, η for pure seawater (molecular scattering) is ~4 (Morel, 1974), with the size of water molecules in an order of 10^{-10} m. In contrast, common marine particles that have significant contributions to $b_{bp}(\lambda)$ have much larger sizes than water molecules (Stramski et al., 2004). Thus, η for common-size marine particles should be far <4. The large η values (up to 4) calculated from $b_{bp}(\lambda)$ measurements reported in the literature were most likely attributed to measurement uncertainties in $b_{bp}(\lambda)$ obtained in extremely clear waters.

Following the above observations, we capped the upper η value as ~3.0 in synthesizing the training data (see Section 2.4) but let η vary randomly within a common range for any given water type or Chl value. In other words, there was no pre-determination of η value for any given water type.

2.5. Synthesis of the training dataset

For any empirical algorithms aimed at global applications, no matter it is explicit empirical or implicit algorithms based on machine learning, a training dataset having "global" coverage is the key. Therefore, compared to the earlier practices where algorithms were calibrated or trained with limited field measurements, here we use a synthesized dataset with 400,000 simulations to account for possible R_{rs} spectra and combinations of IOPs that may be encountered in natural waters.

The synthetic dataset was generated following the same approach as the IOCCG synthetic dataset (IOCCG, 2006), but we employed an R_{rs} model (Lee et al., 2004), rather than Hydrolight (Mobley and Sundman, 2008), to simulate the spectral of $R_{rs}(\lambda)$ from synthesized IOPs for efficient computation. Specifically, the relationship between $R_{rs}(\lambda)$ and IOPs can be expressed as

$$r_{rs}(\lambda) = g_w\frac{b_{bw}(\lambda)}{a(\lambda)+b_b(\lambda)} + g_p(\lambda)\frac{b_{bp}(\lambda)}{a(\lambda)+b_b(\lambda)},$$

(5)

$$g_p(\lambda) = G_0\left[1 - G_1 exp\left(-G_2\frac{b_{bp}(\lambda)}{a(\lambda)+b_b(\lambda)}\right)\right],$$

(6)

$$R_{rs}(\lambda) = \frac{0.52r_{rs}(\lambda)}{1 - 1.7r_{rs}(\lambda)},$$

(7)

where $r_{rs}(\lambda)$ is the remote sensing reflectance just below the sea surface. g_w is the model parameter related to molecular scattering, while $g_p(\lambda)$ describes the contribution of particle scattering, with its model constants (G_{0-2}) depending on the light geometry and particle scattering phase function. The values of g_w and G_{0-2} are set to 0.113, 0.197, 0.636, and
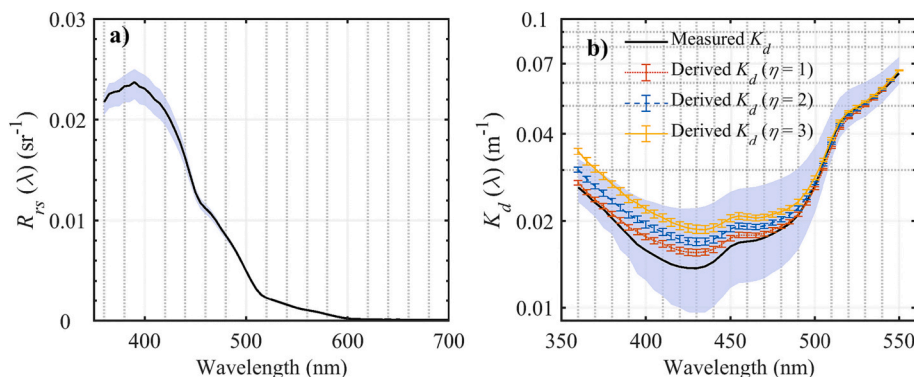

Fig. 2. Panel (a): spectral $R_{rs}(\lambda)$ measured at seven stations in the SPG from BIOSOPE with the black solid line representing the median $R_{rs}(\lambda)$ spectrum and the shaded area for the range of $R_{rs}(\lambda)$. Panel (b): comparison between field-measured $K_d(\lambda)$ and that derived from $R_{rs}(\lambda)$ with η set as constant values at 1, 2, and 3, respectively. The shaded area in panel (b) indicates the range of field-measured $K_d(\lambda)$, while the error bar stands for the standard deviation of derived $K_d(\lambda)$.

2.552, respectively, which were optimized from extensive Hydrolight simulations for various water types (Lee et al., 2004). Note that Hydrolight simulations in Lee et al. (2004) were run for clear skies, constant solar zenith angle (30°), viewing zenith angle (0°), wind speed (5 m/s), and the Petzold average particle scattering phase function, while the bottom reflectance and inelastic scatterings (such as Raman scattering and fluorescence) were excluded (Lee et al., 2004).

A detailed description of the parameterization of component IOPs from a known Chl value can be found in (IOCCG-OCAG, 2003). Consistent with Wang et al. (2021), the value of absorption coefficient by phytoplankton at 440 nm ($a_{ph}(440)$) was used as the input in this study, but it can be readily converted to Chl to model component IOPs following (IOCCG-OCAG, 2003)

$$\text{Chl} = \left(20 a_{ph}(440)\right)^{1.597}. \tag{8}$$

A stepwise description to generate the synthetic $R_{rs}(\lambda)$ using the R_{rs} model can be found in Appendix A of Wang et al. (2021), while steps to model $b_{bp}(\lambda)$ are briefly described here. Specifically, $b_{bp}(\lambda)$ can be modeled as

$$b_{bp}(\lambda) = b_{bph}(\lambda) + b_{bdm}(\lambda), \tag{9}$$

where the subscripts *ph* and *dm* stand for contributions from phytoplankton pigments and the sum of detritus and minerals, respectively. Spectral $b_{bph}(\lambda)$ can be modeled as

$$b_{bph}(\lambda) = B_{ph}\left(c_{ph}(\lambda) - a_{ph}(\lambda)\right), \tag{10}$$

$$c_{ph}(\lambda) = c_{ph}(550)\left(\frac{550}{\lambda}\right)^{n_1}, \tag{11}$$

where $c_{ph}(\lambda)$ and $a_{ph}(\lambda)$ are the attenuation and absorption coefficients of phytoplankton, respectively. B_{ph} is the backscattering-to-scattering ratio of phytoplankton, with a value of 1% employed in this effort. $c_{ph}(550)$ can be empirically estimated from Chl with a random term (IOCCG-OCAG, 2003). n_1 is the spectral power coefficient of $c_{ph}(\lambda)$, which largely determines the spectral shape of $b_{bph}(\lambda)$. Similarly, spectral $b_{bdm}(\lambda)$ can be modeled as

$$b_{bdm}(\lambda) = B_p \times b_{dm}(550)\left(\frac{550}{\lambda}\right)^{n_2}, \tag{12}$$

where B_p is the backscattering-to-scattering ratio of particles and was set as 1.83%, which is equivalent to the Petzold average particle scattering phase function (Petzold, 1972). $b_{bdm}(550)$ can also be estimated from Chl with a random term (IOCCG-OCAG, 2003), while n_2 is the spectral power of $b_{bdm}(\lambda)$. Here n_1 and n_2 were deliberately parameterized so that the range of filed-measured η in the four *in-situ* datasets could be well represented by the synthesized η (see Fig. 1b), with n_1 and n_2 expressed as

$$n_1 = -0.7 + \frac{3.0 + 1.5\Re}{1 + \sqrt{\text{Chl}}}, \tag{13}$$

$$n_2 = -0.6 + \frac{2.2 + 1.5\Re}{1 + \sqrt{\text{Chl}}}, \tag{14}$$

where $\Re$ is a random value between 0 and 1.

In this study, 400,000 sets of $a_{ph}(443)$ values were randomly selected from the MODIS $a_{ph}(443)$ monthly climatology product, which were approximated as $a_{ph}(440)$ and used to generate the hyperspectral $R_{rs}(\lambda)$ and $b_{bp}(\lambda)$ for the spectral range of 400–750 nm at an interval of 5 nm. The η values matching $R_{rs}(\lambda)$ simulations were computed from simulated $b_{bp}(\lambda)$ by the linear fit between $\log(b_{bp}(\lambda))$ and $\log(\lambda)$, where the computed η is generally within −0.7 and 3.1. The synthesized dataset is hereafter referred to as *Syndata*. Note that pure seawater absorption coefficient ($a_w(\lambda)$) and $b_{bw}(\lambda)$ are required to model $R_{rs}(\lambda)$ from IOPs, with $a_w(\lambda)$ values taken from Lee et al. (2015) for 350–550 nm and Pope

and Fry (1997) for 550 nm and longer wavelengths, while $b_{bw}(\lambda)$ values were calculated from Zhang et al. (2009) using the global median SSS (34.5 PSU) and SST (15.2°). Here, $a_w(\lambda)$ values of Lee et al. (2015) were employed due to that they are closer to the pure seawater absorption coefficient (Yu et al., 2019b), while the dependence of $a_w(\lambda)$ on temperature was omitted as it has negligible impacts on $a_w(\lambda)$ in the visible domain (Röttgers et al., 2014; Wei et al., 2021a).

The validity and performance of an NN-based scheme are largely dependent on the representativeness of the training dataset to the actual range and distribution of the NN-predicted properties. In this effort, the input $a_{ph}(440)$ values were obtained from satellite measurements to ensure IOPs and AOPs in the synthetic dataset are within reasonable ranges. The simulated η in *Syndata*, approximating a normal distribution as shown in Fig. 3a, agrees well with the η distribution in the global oceans from field measurements (Fig. 1b). In addition, the frequency distribution of simulated $R_{rs}(\lambda)$ and $b_{bp}(\lambda)$ are also in good agreement with that in the global ocean measured by MODIS (see comparisons in Fig. 3 with $R_{rs}(488)$ and $b_{bp}(443)$ as examples). More importantly, the ranges of simulated $R_{rs}(\lambda)$ and $b_{bp}(\lambda)$ are greater than that from the MODIS climatology products of $R_{rs}(\lambda)$ and $b_{bp}(\lambda)$, suggesting that the *Syndata* could represent most water types in natural environments. The frequency distribution of $a_{ph}(443)$ in *Syndata* is not shown here, as it has a similar pattern as that in Fig. 3d.

It is necessary to emphasize that for any given input $a_{ph}(440)$ (or equivalent Chl), there is no predefined value of η in *Syndata*, rather η could vary randomly within a commonly accepted range. For example, as shown in Fig. 1b, for $b_{bp}(443)$ within 0.0001–0.1 m^{-1}, the standard deviation of simulated η would be generally between 0.3 and 0.6 for a given $b_{bp}(443)$ value. However, given that the square root of Chl is used in the denominator to regulate the values of n_1 and n_2 (see Eq. (13) and Eq. (14)), a statistically decreasing trend of η with increasing Chl is assigned in the synthetic dataset, which is consistent with field observations (see Fig. 1a).

2.6. Structure of NN$_\eta$

Following the same neural network system used in Wang et al. (2021), the architecture of NN$_\eta$ consists of three units: input layer, hidden layer, and output layer. The input layer contains 6 neurons, corresponding to $R_{rs}(\lambda)$ at six MODIS bands centered at 412, 443, 488, 531, 547, and 667 nm, respectively. The number of hidden layers and the number of neurons within each layer were determined following the concept of minimum loss, where two hidden layers with 32 and 16 respective neurons were found optimum for NN$_\eta$. The output layer contains only one neuron, which is the η. Keras, a Python-implemented deep learning Application Programming Interface (https://keras.io/), was employed in this study for the training of the neural network, which is running on top of the machine-learning platform TensorFlow (Abadi et al., 2016). Other important configurations of the network are briefly described here. The learning rate was set to 0.001, while the optimizer employed the Adam algorithm (Kingma and Ba, 2014), with the loss function defined as the mean absolute error between the predicted and known η. In addition, the Rectified Linear Unit (ReLu) function was used as the activation function of each layer (Krizhevsky et al., 2012). The training phase is completed when the loss function converges, or the iterations reach the maximum of 200 epochs.

For *Syndata*, 80% of the simulations are randomly selected and used as the training dataset (termed *CalSet*, N = 320,000), while the rest simulations are used for validation (termed *ValSet*, N = 80,000). The performance of the trained network, termed NN$_\eta$ hereafter, is preliminarily evaluated with the *ValSet*, with scatterplots between known and predicted η shown in Fig. 4. As shown in the figure, the line representing the best linear fit is very close to 1:1 line, suggesting an excellent agreement between the predicted and known η.

The good performance of NN$_\eta$ is also confirmed by the statistical measures calculated from the simulated and predicted η. As every single
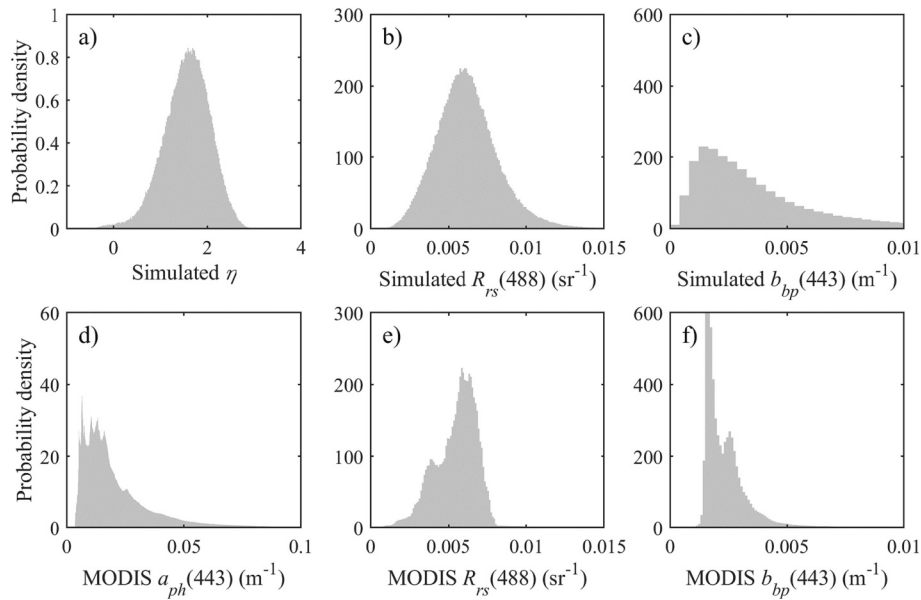
Fig. 3. Histograms showing the frequency distribution of (a) simulated η, (b) simulated $R_{rs}(488)$, and (c) simulated $b_{bp}(443)$ in the *Syndata* (N = 400,000). The lower panel shows the histogram of (d) $a_{ph}(443)$, (e) $R_{rs}(488)$, and (f) $b_{bp}(443)$ from MODIS climatology product (N ~ 23,000,000), respectively. Note that the histograms are normalized by the probability density function.
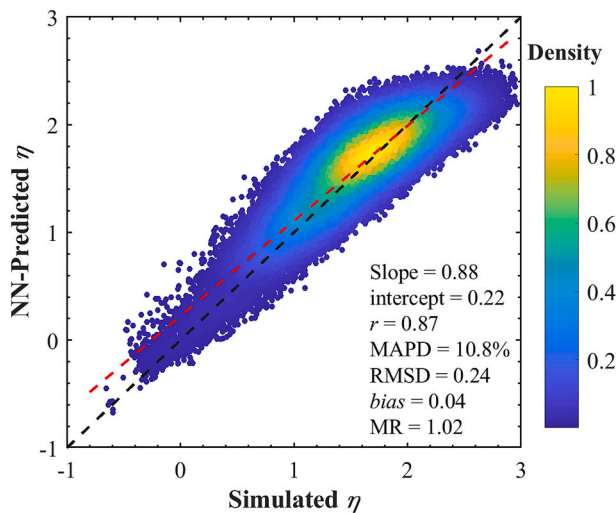

Fig. 4. Evaluation of NN$_\eta$ with the *ValSet* of *Syndata* (N = 80,000), with the color of scatter points indicating the frequency that is normalized to 1. The black dashed line indicates a 1:1 relationship, while the red dashed line indicates the best linear fit from Type-II regression. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

metric has its limitations (Brewin et al., 2015; Seegers et al., 2018), we, therefore, employed seven metrics to provide a relatively comprehensive evaluation of the model's performance, which include the Pearson correlation coefficient (r), the slope and intercept computed from the linear squares fitting using the Type-II regression, the Median Absolute Percentage Difference (MAPD), the Root Mean Square Deviation (RMSD), *bias*, and the Median Ratio (MR), with the latter four metrics defined as

$$\text{MAPD} = median\left\{ \left| \frac{y_i - x_i}{x_i} \right| \right\} \times 100\%, i = 1, 2, \ldots N \quad (15)$$

$$\text{RMSD} = \sqrt{\frac{\sum (y_i - x_i)^2}{N - 2}}, i = 1, 2, \ldots N \quad (16)$$

$$bias = median(y_i - x_i), i = 1, 2, \ldots N \quad (17)$$

$$MR = median(y_i/x_i), i = 1, 2, \ldots N \quad (18)$$

where y_i and x_i are the predicted and known η, respectively, and N is the number of data points used in the metrics calculation.

3. Results

3.1. Evaluation of NN$_\eta$ with in-situ measurements

Dataset, independent of those used in the algorithm's calibration or training, is usually demanded for convincing evaluations of any proposed algorithm. Here we apply the three conventional η schemes (LS2, MM01, and QAA, as described in Section 1.2) and NN$_\eta$ to the *in-situ* dataset (described in Section 2.2) to demonstrate their respective performance and applicability, with scatterplots and statistics shown in Fig. 5 for visual inspection.

As shown in Fig. 5, NN$_\eta$ generally has the best performance among these four schemes, with scatters well distributed along the 1:1 line and overall least uncertainties in the estimated η. For instance, the statistical measures for the estimated η by NN$_\eta$ suggest that NN$_\eta$ outperforms LS2 and MM01, with much smaller MAPD (= 24.6%), RMSD (= 0.45), and *bias* (= −0.03), larger r (= 0.62), almost nil intercept (= 0.1), and the close-to-unity slope (= 0.92) and MR (= 0.98). Though QAA has relatively comparable performance with NN$_\eta$ in terms of MAPD, RMSD, and *bias*, estimations from QAA are less confident in terms of the slope, intercept, and r. Also, it is evident that QAA-estimated η saturates at ~2 from mesotrophic to ultra-oligotrophic waters. Take the BIOSOPE dataset for example, the estimated η at 13 out of the 19 stations approximated to 2 (see the orange squares in Fig. 5d). Thus, QAA might have rather limited applicability in predicting η in mesotrophic to ultra-oligotrophic waters, as significant variabilities of η are observed in these waters (see Fig. 1b for example).

Predicted η by LS2 have relatively larger uncertainties than NN$_\eta$, with a more pronounced scatter around the 1:1 line (Fig. 5b). The
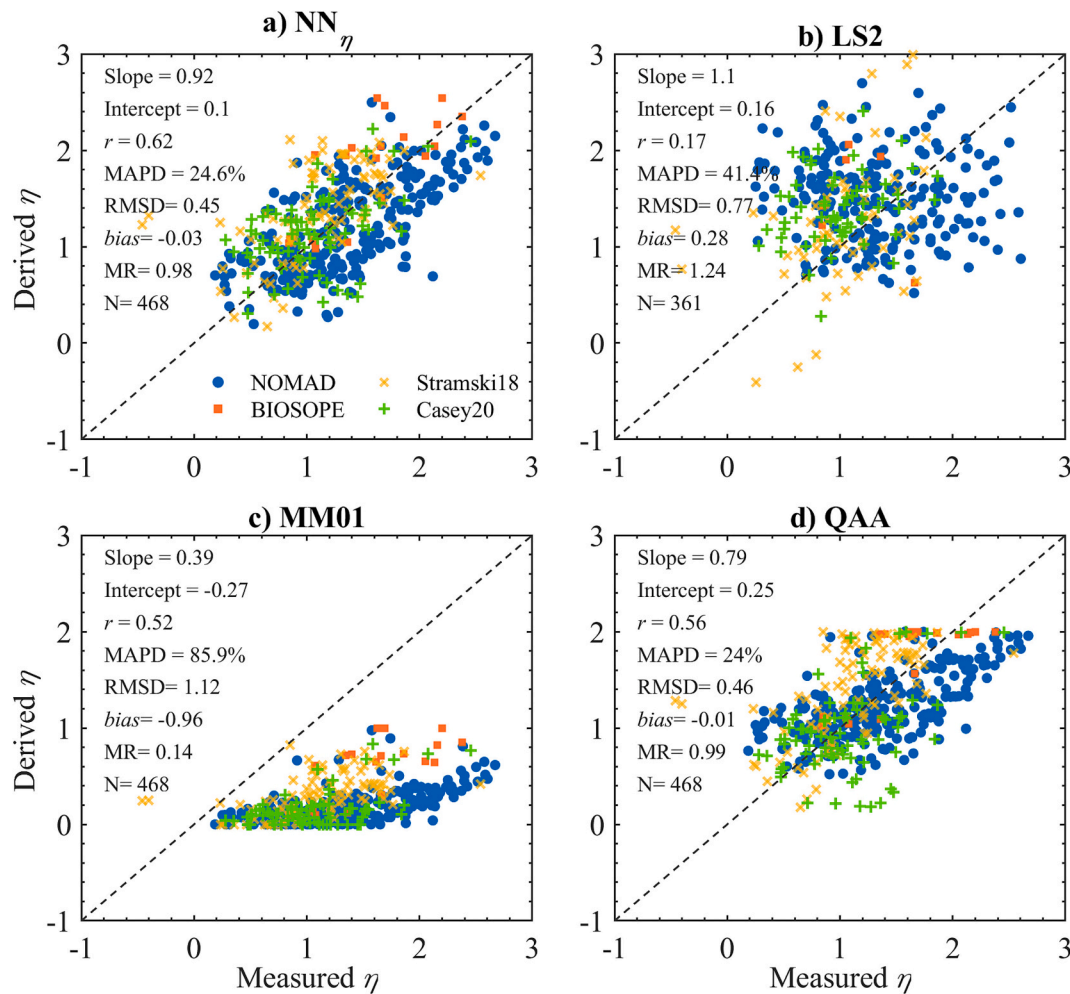
Fig. 5. Validation of (a) NN$_\eta$, (b) LS2, (c) MM01, and (d) QAA with the *in-situ* dataset, with different symbols indicating measurements from NOMAD (circles), BIOSOPE (squares), Stramski18 (crosses), and Casey20 (plus sign).

relatively poor performance of LS2 in the *in-situ* dataset is also supported by the statistical measures, with an MAPD of 41.4%, an RMSD of 0.77, a *bias* of 0.28, and an MR of 1.24 for derived η. More importantly, for the 468 quality-controlled $R_{rs}(\lambda)$ and η matchups in the *in-situ* dataset, LS2 failed to provide reasonable estimations of η for 107 of these matchups, which is quite significant given the failure ratio ($\sim$ 22.9%). The questionable η estimations are mainly due to poor linear fit between log $(b_{bp}(\lambda))$ and log(λ) with $R^2 < 0.6$ ($N = 84$), or unreasonable estimates of η ($\eta < -0.5$ or $\eta > 4$, $N = 23$).

At last, estimated η by MM01 are systematically underestimated for the *in-situ* dataset (Fig. 5c), with an MR of only 0.13 between the estimated η to known η. Also, MAPD, RMSD, and the absolute *bias* of the estimated η by MM01 are significantly greater than the rest three schemes. As shown in Fig. 5c, it can be also concluded that setting η as 0 for waters with Chl > 2.0 mg m^{-3} is not practicable. For instance, there exists a large number of data in Fig. 5c with MM01-estimated η as 0 (*i.e.*, Chl > 2.0 mg m^{-3}), but the measured η spans a range roughly between 0 and 2 (see also Fig. 1a).

Detailed discussions on the limitations of these schemes are provided in Section 4, here we preliminarily conclude from Fig. 5 that NN$_\eta$ might be the most appropriate scheme for the estimation of η from ocean color remote sensing compared to the three conventional schemes.

3.2. Global distribution of η

The four schemes are further implemented to MODIS imagery to

examine their respective performance in the global oceans, with η products of the global ocean by these schemes presented in Fig. 6. Here MODIS monthly composite data of May 2018 is employed to facilitate direct comparisons with the results presented in Jorge et al. (2021), which were generated from monthly composite data of Ocean and Land Color Instrument (OLCI) in May 2018 (see their Fig. 12). Results in Fig. 6 demonstrate that derived η by the four schemes differ a lot from each other in both the magnitude and the spatial distribution pattern, which deserve further investigations on the applicability of these η schemes in global oceans.

As shown in Fig. 6, the predicted η by NN$_\eta$, LS2, and MM01 features a manifest spatial pattern with large η values observed in the oligotrophic gyres, while such a pattern is not observed by QAA. It is commonly acknowledged that small-sized phytoplankton dominates in oligotrophic water with low Chl, except for bloom scenarios, while coastal and nutrient-rich waters, such as the upwelling regions, are characterized by the prevalent presence of large-sized phytoplankton groups and much higher Chl (Loisel et al., 2006; Buonassissi and Dierssen, 2010). Thus, a general decreasing trend of η from the oligotrophic ocean gyres (low Chl) to eutrophic waters (high Chl) should be observed, where larger η values should be expected in the ocean gyre waters (Ciotti et al., 1999). Therefore, the estimated η by NN$_\eta$, LS2, and MM01 is at least reasonable in the spatial pattern. However, the derived η by MM01 is much smaller than the common η values in global oceans concluded from either field measurements (*e.g.*, Fig. 1b) or theoretical analysis (see Section 2.3 for example), while LS2 tends to predict high η values in the oligotrophic
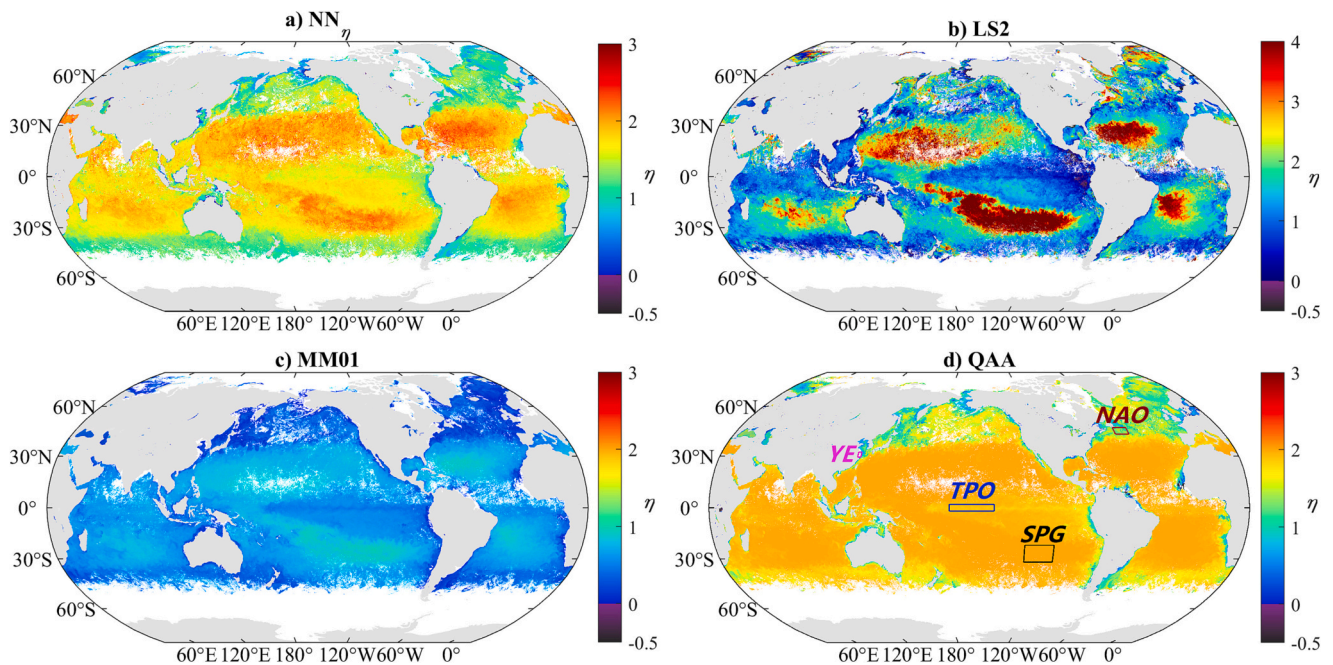
Fig. 6. Global distribution of η in May 2018 estimated by (a) NN_η, (b) LS2, (c) MM01, and (d) QAA using the MODIS monthly composite data. The four boxes in panel (d) highlight the locations of four regions of interest in the Tropic Pacific Ocean (TPO), the South Pacific Gyre (SPG), the North Atlantic Ocean (NAO), and the Yangtze Estuary (YE), respectively. Note that panel (b) has a different colorbar scale.

ocean gyres and the North Sea that >3 (Fig. 6b). In fact, a majority of η retrievals by LS2 in the ocean gyres are even >4, which accounts for ~4% of all retrievals in the global ocean. As discussed in Section 2.3, an upper boundary of η in natural waters could be ~3, or at least not significantly >3. Thus, those high η retrievals (> 4) from LS2 should be questionable. In addition, ~0.2% of LS2-derived η are found with η less than −1, which are most frequently encountered in the coastal regions. In contrast, derived η by NN_η is more reasonable in magnitude when compared with measurements from the global ocean as shown in the *in-situ* dataset (see Fig. 1b).

The estimated η by QAA is almost constant for a majority of oceanic waters, with 60.3% of all η retrievals by QAA between 1.9 and 2 for the MODIS measurement in May 2018. Such a limited spatial variability of η contradicts observations from field measurements. For instance, Chl and $b_{bp}(\lambda)$ from the BIOSOPE transect were found with an evident gradient for oceanic stations, with measured η from $b_{bp}(\lambda)$ generally varying

between 1 and 2.6 (see Fig. 1b). However, the predicted η by QAA at these stations all approximate 2 as shown in Fig. 6d. Thus, QAA fails to capture the detailed spatial feature of η in open oceans.

The ranges and frequency distribution of the estimated η by the four schemes in global oceans are further illustrated by the histogram of η retrievals with results presented in Fig. 7a. In general, the frequency distribution of estimated η by NN_η, LS2, MM01 could be well approximated by a normal distribution despite the existence of a small peak, but with different mean values and standard deviations to describe the probability density function. For QAA, it is not surprising that the majority of estimated η fall between 1.8 and 2, and the values of estimated η follow a truncated normal distribution with zero probability for η >2.

Statistically, the median η values in the global ocean are 1.67, 1.46, 0.59, and 1.94 for η retrievals from NN_η, LS2, MM01, and QAA, respectively. Here, the largest 0.1% and the smallest 0.1% η retrievals from these four schemes were excluded from these calculations to
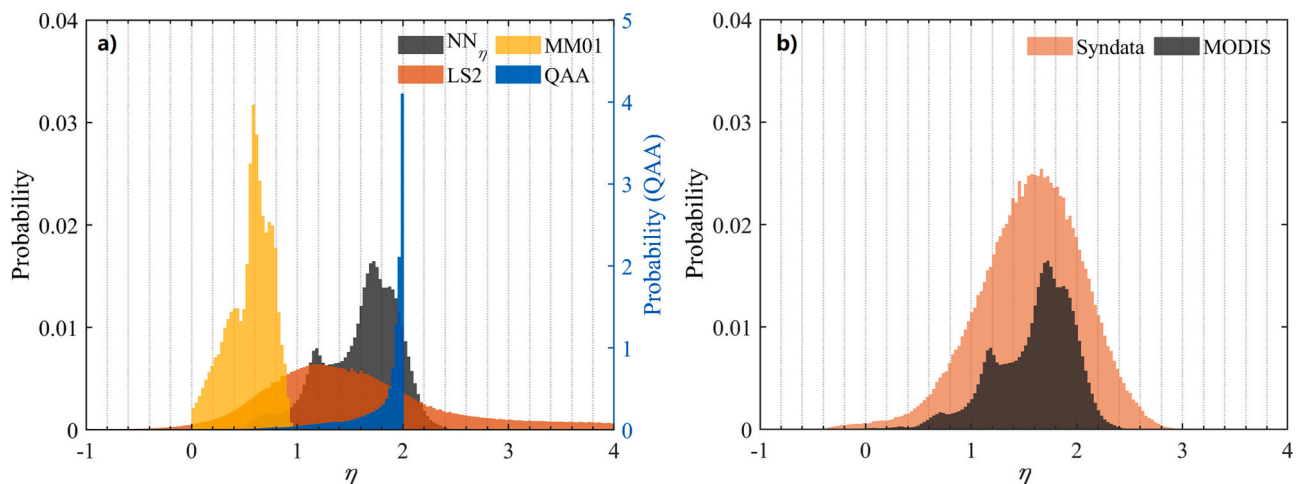

Fig. 7. Panel (a) shows the histogram of η values in global oceans in May 2018 estimated by NN_η, LS2, MM01, and QAA, with results from QAA aligned to the right y-axis. Panel (b) presents a direct comparison between the η distribution in *Syndata* and that in the global ocean obtained by NN_η using MODIS data of May 2018.

remove possible outliers. The range of estimated η by NN$_\eta$ for MODIS data in May 2018 is from -0.1 to 2.4, which is expected given the range of η in the training dataset (Fig. 7b). The ranges of estimated η by MM01 and QAA also fall into their theoretic boundaries described in Section 1.2, which are 0–1 and -0.2–2.0, respectively. In contrast, estimated η by LS2 in global oceans ranges between -4.2 and 6.1, where both the maximum and minimum η values far exceed the realistic boundary of η. Note that estimated η by LS2 with poor linear fit between $\log(b_{bp}(\lambda))$ and $\log(\lambda)$ were already removed from the calculations.

3.3. Monthly variation of η

The monthly variation of η in the four regions of interest (ROIs) characterizing distinguished bio-optical features is further analyzed using MODIS monthly composite data from January 2003 to December 2021. These four ROIs are the Tropical Pacific Ocean (TPO, 175°W–145°W; 2°S–2°N), the South Pacific Gyre (SPG, 124°W – 104°W, 32°S – 22°S), the North Atlantic Ocean (NAO, 55°W – 45°W, 43°N – 47°N), and the Yangtze Estuary (YE, 120.7°E – 123.0°E, 29.5°N – 32.5°N), respectively. Specifically, TPO represents oceanic waters with negligible seasonal cycles of bio-optical properties (Yu et al., 2022); SPG represents the 'clearest' natural waters (Morel et al., 2007; Twardowski et al., 2007); NAO represents optically complex waters with high colored dissolved organic matter (CDOM) to Chl ratio (Morel et al., 2010); while YE represents one of the most turbid waters in the globe (Yu et al., 2019a; Wei et al., 2021b). Here, results from all four schemes are included in Fig. 8 to demonstrate their applicability in capturing the long-term variation of η in the global ocean.

In TPO, η is relatively stable in the last two decades with no clear

seasonal cycle observed (see Fig. 8a). For instance, η is almost stabilized around 1.6, 0.5, and 1.8 for estimations from NN$_\eta$, MM01, and QAA, respectively. Estimated η by LS2, however, shows relatively larger monthly variabilities, with η varying between 0.7 and 2.1. The standard deviation for LS2-derived η in TPO at each month is ~0.3, which is also significantly greater than that from the rest three schemes. The prominent plummet and surge of η in 2010 and 2015 observed from the four schemes are highly coincident with the strongest La Niña and El Niño events in the last two decades, respectively, which is concluded from the sea surface temperature (SST) anomalies in the Niño3.4 region (5°N – 5°S, 170°W – 120°W) (data available at psl.noaa.gov/data/timeseries/monthly/NINO34/). During an El Niño event, warming of surface water would enhance ocean stratification, which suppresses nutrient mixing from subsurface layers and thus limits phytoplankton growth in the surface layer (Roxy et al., 2016; Gittings et al., 2018), resulting in decreased Chl (smaller phytoplankton) and elevated η. Contrarily, η is expected to decrease when SST increases during a La Niña event (*e.g.*, from 2010 to 2011). Take η derived from NN$_\eta$ for example, η in TPO is inversely correlated with Chl (see Fig. 9a), with a Pearson correlation coefficient (*r*) of -0.79, suggesting that Chl could be the main factor regulating the monthly variability of η. Here, Chl was also calculated as the median value of each ROI from the MODIS Chl monthly products. In contrast, the Pearson correlation coefficient between η and b_{bp} is only -0.54. The median b_{bp} at 547 nm ($b_{bp}(547)$) of each ROI is used for the scatterplots and correlation analysis in Fig. 9b, with $b_{bp}(547)$ derived from the monthly $R_{rs}(\lambda)$ using QAA_v6.

In contrast to the seasonal variation in TPO, η in SPG, NAO, and YE present strong seasonal patterns. In SPG, η peaks in austral summer (December to February) and attains a minimum in austral winter (June
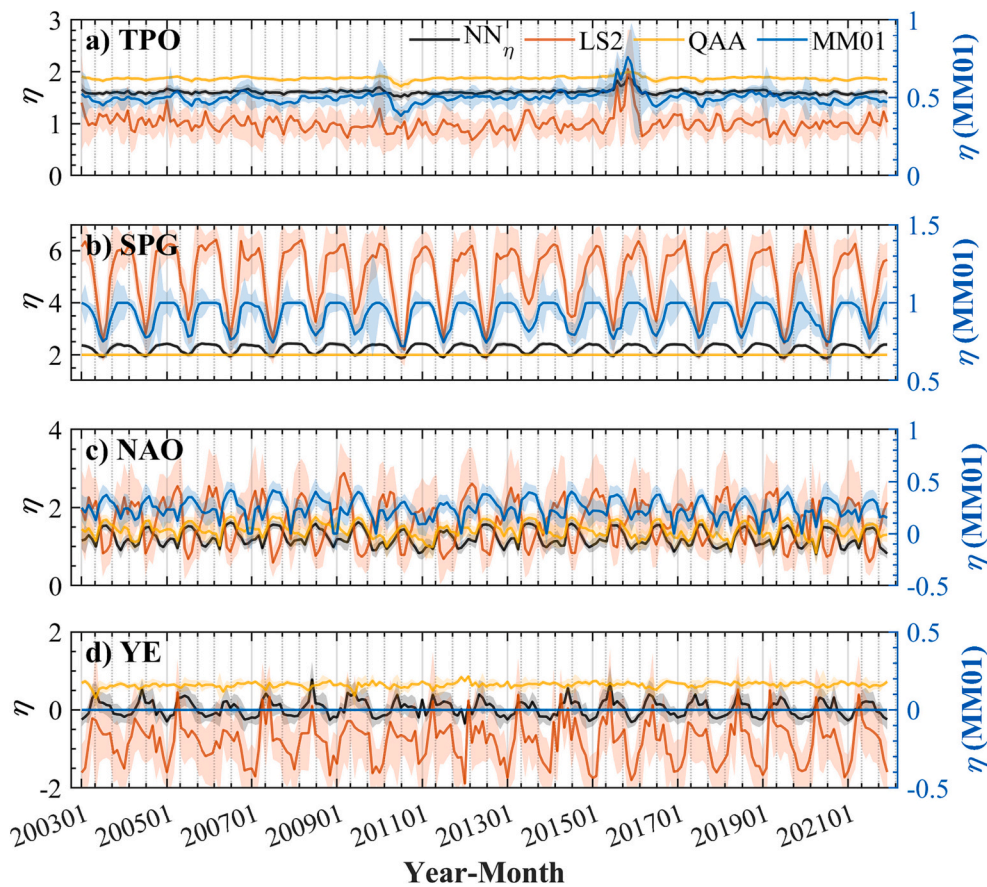

Fig. 8. Monthly variation of derived η by NN$_\eta$, LS2, MM01, and QAA from January 2003 to December 2021 in (a) TPO, (b)SPG, (c) NAO, and (d) YE, respectively. The median values of η in the respective ROI are plotted with the solid lines, while the shaded areas indicate the standard deviation. Results from MM01 are aligned to the right y-axis, while results from the rest schemes are aligned to the left y-axis.
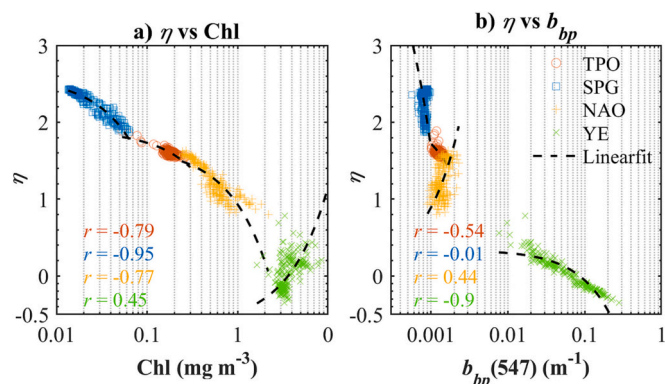
Fig. 9. Panel (a) shows the correlation between Chl and NN_η-derived η in the four ROIs using the MODIS monthly composite data from January 2003 to December 2021. The dashed lines represent the best linear fit between Chl and η using Type-II regression for each of the ROI. Panel (b) is the same as panel (a) but for the correlation between $b_{bp}(547)$ and η.

to August), with a much larger seasonal amplitude than that in TPO (see Fig. 8b). Take NN_η for example, estimated η varied between 1.9 and 2.5, with a seasonal amplitude of ~0.6. Estimated η by LS2 and MM01 present the same seasonal pattern as NN_η, but with overall much higher η estimations by LS2 and much smaller η estimations by MM01, respectively. For LS2, estimated η varied between 2.2 and 6.2 in the last two decades with a seasonal amplitude of ~4.0. As discussed in Section 2.3, natural waters with η over 4 are hardly possible, given the ~λ^{-4} dependency of the scattering by water molecules. Thus, predicted η by LS2 could be questionable in SPG. Consistent with the conclusion in Section 3.2, estimated η by MM01 are significantly underestimated in SPG, with η varying between 0.8 and 1. For QAA, it can be found that predicted η is constant at ~2 over the entire period. Thus, QAA fails to capture the seasonality of η in SPG. It is worth pointing out that η in SPG is highly correlated with Chl ($r = 0.95$), but is almost independent of $b_{bp}(547)$ ($r = -0.01$) (see Fig. 9). Thus, it can be concluded that the seasonality of η in SPG can be well explained by the changes in Chl.

The estimated η by NN_η and QAA are quite consistent in NAO in both the temporal variation and the magnitude, with η generally varying between 0.8 and 1.6. In NAO, η features a different seasonal pattern with two peaks around March and August, respectively, and the latter peak prevails. As shown in Fig. 9a, the seasonality of η in NAO is also largely determined by Chl ($r = -0.77$), but the correlation with Chl degrades compared to that in SPG. This is more likely because NAO is rich in CDOM and suspended matters originated from terrestrial inputs (Morel et al., 2010). Estimated η by MM01 exhibits similar temporal variations but with overall smaller values (~0.3). The seasonal variability of LS2-derived η is, however, almost opposite to the rest three schemes, and the seasonal amplitude (~2) is also much greater than by NN_η and QAA. As η is a second-order parameter in LS2, its value is quite sensitive to the empirically estimated $K_d(\lambda)$ and the subsequent $b_{bp}(\lambda)$ at the four selected wavelengths, which might have relatively large uncertainties in NAO. Note that the seasonal variability of LS2-derived η would be consistent with results from NN_η if employing the NN-K_d scheme by Jamet et al. (2012) (results not shown here). Thus, further evaluations of LS2-derived $K_d(\lambda)$ and $b_{bp}(\lambda)$ in NAO would be desired, particularly the applicability of the NN-K_d scheme by Jamet et al. (2012) and its updated version proposed in Loisel et al. (2018) that employed by LS2 in this effort.

For the turbid waters in YE, estimated η by both NN_η and LS2 presents a strong seasonal cycle (see Fig. 8d), with η peaking in boreal spring and summer from April to June and reaching a minimum in boreal winter from December to February. However, significant discrepancies are observed for the median value and the amplitude of derived η by NN_η and LS2 in the last two decades. In contrast, MM01

predicts a constant η of 0 in YE with no seasonal variability. Predicted η by QAA shows no clear seasonal pattern but with overall high η values (~0.6). On one hand, given that YE is characteristic of high turbidity and turbid waters usually have relatively flatter backscattering spectra (Lobo et al., 2014), estimated η by NN_η, with a median η value of ~ -0.05 in the last two decades, appears more reasonable than predictions by QAA and LS2. In particular, LS2-derived η in YE ranges between ~ -2 and 0, with a median value of -0.81, which could be questionable and probably a result of the limited applicability of LS2 in extremely turbid waters. On the other hand, the seasonal cycle of η obtained from NN_η is also more reasonable from the oceanographic point of view. Since higher turbidity results in flatter backscattering spectra (*i.e.*, smaller, close to nil, η) (Lobo et al., 2014), an inverse correlation between η and turbidity should be expected in turbid waters in YE, which is confirmed by the strong and inverse correlation between $b_{bp}(547)$ and η by NN_η ($r = -0.9$) (see Fig. 9b). The seasonal cycle of $b_{bp}(547)$ in YE can also be reasonably explained by the seasonal changes in turbidity observed from either satellite retrievals (Shen et al., 2010; Wei et al., 2021b) or field observations (Chen et al., 2003). Specifically, in boreal spring and summer, large amounts of freshwater discharge would dilute the concentration of suspended sediment (flood season), resulting in overall low turbidity in YE, while high turbidity in boreal winter is due to wind-driven resuspension of suspended sediment and less freshwater discharge (dry season) (Chen et al., 2003). Thus, we can conclude that the seasonal cycle of η in YE can be well represented by NN_η and the seasonality of η is mostly regulated by turbidity (or b_{bp}).

Comparisons among the four schemes suggest that NN_η might be the best scheme to approach reasonable η estimations in terms of the spatial pattern, magnitude, and seasonality of η in the global ocean, while the rest three schemes all have limitations. Detailed discussions on the limitations of all four schemes can be found in Sections 4.1 and 4.2.

3.4. Decadal changes of η

The long-term variations of η might be useful to interpret the changes in the suspended particulate matter pool in the global ocean (Vantrepotte et al., 2011). Thus, we employ a simple trend analysis to investigate the decadal changes of η in the global ocean using NN_η-derived η from the 19-year MODIS monthly series, with results demonstrated in Fig. 10. Here, the Matlab function *trend* within the Climate Data Toolbox was used for the trend analysis in this effort (Greene et al., 2019), and the trend of η was calculated only for grid cells with at least 50% valid η retrievals over the timespan.

As shown in Fig. 10, the annual trends of η in the majority of the global waters (~82.2%) were not statistically significant from 2003 to 2021 (*i.e.*, grey areas in Fig. 10), which were also reported in the previous calculations using the SeaWiFS monthly series from 1997 to 2007 (Vantrepotte et al., 2011). Take the four ROIs of this study as examples,
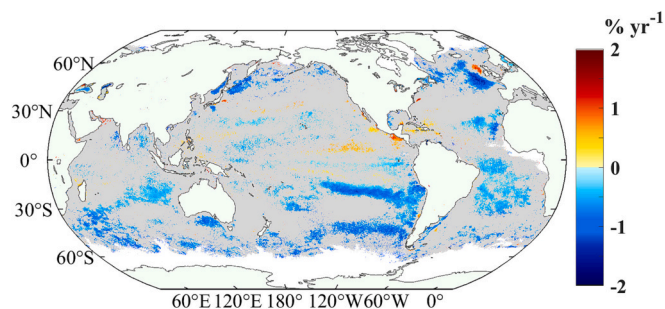

Fig. 10. Significant monotonic trends of NN_η-derived η ($p < 0.05$, in % yr^{-1}) in the global ocean from MODIS monthly composites (from January 2003 to December 2021). The white areas indicate grid cells without valid or enough ($>$ 50%) η retrievals over the timespan for the trend analysis, while the grey areas indicate no significant trends.

the slopes of the annual changes of η in TPO, SPG, NAO, and YE are 0.012% yr^{-1}, -0.15% yr^{-1}, -0.35% yr^{-1}, and -0.23% yr^{-1}, respectively, but these changes are not statistically significant ($p > 0.1$). On the other hand, the results of this effort and that of Vantrepotte et al. (2011) indicate overall decreasing trends of η for those waters with significant η changes ($p < 0.05$), especially in the southern hemisphere, such as the Southeast Pacific Ocean, the South Atlantic Ocean, and the South Indian Ocean.

However, discrepancies are also found between these two studies, most notably in the Western Pacific Warm Pool (WPWP), where a significant increasing trend was reported in Vantrepotte et al. (2011) ($>$ 4% yr^{-1}), while a slightly decreasing trend was observed in this effort ($\sim$ -0.2% yr^{-1}, see Fig. 10). These discrepancies might be largely due to different periods investigated in these two studies, as trend analysis could be drastically impacted by events occurring over specific periods. In addition, as discussed in Vantrepotte et al. (2011), the annual changes of η in WPWP could reflect large-scale oscillations in the physical environment, such as the El Nino-Southern Oscillation (ENSO). Thus, the discrepancies observed in WPWP between this study and that in Vantrepotte et al. (2011) could also be related to ENSO. Take the Nino3.4 index for example, a significant increasing trend of the Nino3.4 index was observed for the timespan (from November 1997 to October 2007) used for the trend analysis in Vantrepotte et al. (2011), with a slope of 5.6% yr^{-1} ($p = 0.04$). In contrast, the changes in the Nino3.4 index were not statistically significant for the timespan from January 2003 to December 2021 (slope = 0.0004% yr^{-1}, $p = 1$). However, to further explore the relationships between the seasonality of η and large-scale oscillations and climate indices, a comprehensive and dedicated analysis would be required, especially for different oceanic domains, which is out of the scope of this study.

4. Discussion

4.1. Limitations and uncertainties associated with conventional η schemes

The four schemes for η estimation examined in this study are fundamentally empirical, and thus their performances in global oceans are largely determined by the data used for model training and the algorithm architecture. For both MM01 and QAA, where the primary objective of η is to model spectral $b_{bp}(\lambda)$, η is simply estimated from the blue-green band ratio of $R_{rs}(\lambda)$, which, to the first order, represents an overall loading of the constituents suspended or dissolved in oceanic waters (Shang et al., 2019). Thus, a simple blue-green ratio of R_{rs} does not have the flexibility to estimate the widely varying η (Lee et al., 2003), which can be observed in Fig. 5d and Fig. 6d. For example, QAA failed to interpret the seasonality of η in ocean gyres while MM01 failed in turbid coastal waters (Fig. 8). Furthermore, when an intermediate variable, such as Chl in MM01 and $K_d(\lambda)$ in LS2, is introduced in the process, it would probably increase the uncertainties for the estimated η. For instance, uncertainties of derived Chl by the OCI algorithm could be considerably large, especially in optically complex waters (Cota et al., 2004; Le et al., 2013; Shang et al., 2019). Likewise, empirically estimated $K_d(\lambda)$ in LS2, and the subsequent semi-analytically estimated $b_{bp}(\lambda)$, could be also of large uncertainties, with these uncertainties being propagated to the estimated η.

Though the two steps scheme (LS2) could allow better interpretation of the spectral behavior of $b_{bp}(\lambda)$ as they are estimated independently at each band, the estimated η is more sensitive to noise by model construction, particularly the uncertainties in the empirically estimated $K_d(\lambda)$ and subsequent $b_{bp}(\lambda)$. As discussed earlier, the implementation of a different version of the NN-K_d scheme (*i.e.*, Jamet et al. (2012)) would even alter the seasonal variability LS2-derived η (see Section 3.3). In addition, as η is calculated from LS2-derived $\log(b_{bp}(\lambda))$ and $\log(\lambda)$ *via* linear regression, the wavelengths used in the linear fitting would significantly alter the interpretation of η in the global ocean (Loisel et al., 2006; Vantrepotte et al., 2011; Jorge et al., 2021). For instance, the

spatial distribution of LS2-derived η acquired from MODIS in this study is quite different from that observed using OLCI (Jorge et al., 2021), which could probably be attributed to the band configurations in these two sensors. In particular, OLCI has radiometric measurements at 510 nm, but not 532 nm compared to MODIS. Note that Dr. Daniel Jorge also implemented LS2 to the same MODIS $R_{rs}(\lambda)$ measurements in the SPG region, and consistent η estimations were obtained with the results of this study (results not shown here).

Moreover, the spatial pattern of LS2-derived η presented in this study is quite consistent with the result reported by Loisel et al. (2006) using SeaWiFS measurements at 490, 510, and 555 nm, with the latter presenting overall smaller η retrievals in the extremely clear waters (η = 3.5–4). The discrepancies between LS2-estimated η in this effort and that in Loisel et al. (2006) could be mainly attributed to the fact that Loisel et al. (2006) excluded $b_{bp}(\lambda)$ at $\sim$440 nm from the calculation of η due to strong absorption by phytoplankton at this band. For example, if $b_{bp}(443)$ is excluded in this effort, LS2-derived η from MODIS would be significantly increased, with a global mean of $\sim$3.5 for the same MODIS measurement in May 2018. In addition, the values of $b_{bw}(\lambda)$ employed in LS2 also have significant impacts on the estimated η. For example, employing the constant $b_{bw}(\lambda)$ values of Morel (1974) in this effort would result in $\sim$0.5 smaller η on average for the global ocean compared to that using salinity- and temperature-dependent $b_{bw}(\lambda)$. Thus, the selection of different $b_{bw}(\lambda)$ values could also partly explain the overall higher η estimation in this study compared to that in Loisel et al. (2006), with the latter employing the constant $b_{bw}(\lambda)$ values of Morel (1974). Thus, the different η estimations by LS2 in these reports could be mainly attributed to the inversion system, particularly the wavelengths and $b_{bw}(\lambda)$ values used for η calculation, rather than the differences in the satellite measurements.

4.2. Dependency of NN$_\eta$ on the training dataset

The applicability and uncertainty of the estimated η by NN$_\eta$ are largely determined by the representativeness of the dataset used in the training process. To evaluate the dependency of the proposed NN$_\eta$ on the training dataset, a new batch of 400,000 simulations was generated, where n_1 and n_2 were modeled following the default parametrizations in IOCCG-OCAG (2003),

$$n_1 = -0.4 + \frac{1.6 + 1.2\Re}{1 + \sqrt{Chl}}, \tag{19}$$

$$n_2 = -0.5 + \frac{2.0 + 1.2\Re}{1 + \sqrt{Chl}}. \tag{20}$$

With Eq. (19) and Eq. (20), the simulated η is found within in range of -0.6–2.3 and the resulting simulations are hereafter referred to as *Syndata*-new. Following the same network configuration as NN$_\eta$, a separate neural network, termed NN$_\eta$-new, was trained using 80% of the simulations in *Syndata*-new (randomly selected).

We implemented NN$_\eta$-new to the MODIS monthly composite data in May 2018 and compared the derived η in global oceans with that using NN$_\eta$. As shown in Fig. 11, the estimated η by NN$_\eta$-new in the global ocean is almost linearly correlated with that by NN$_\eta$ ($r = 0.99$), but NN$_\eta$-new predicts systematically smaller η than NN$_\eta$, especially in the oligotrophic oceans (high η values). The median ratio between estimated η by NN$_\eta$ and NN$_\eta$-new is 1.34 (see Fig. 11), which is highly coincident to the ratio ($\sim$1.38) of median synthetic η values in *Syndata* and *Syndata*-new, which are 1.59 and 1.15, respectively. Also, as shown in the insert figure of Fig. 11, the median η value in the global ocean estimated by NN$_\eta$ ($\sim$1.6) is $\sim$0.4 larger than that estimated by NN$_\eta$-new ($\sim$1.2) (see also the histogram for the difference between estimated η by NN$_\eta$ and NN$_\eta$-new), suggesting the dependency of NN$_\eta$ on the training dataset. However, it is worth pointing out that the overall changes in the estimated η by NN$_\eta$ and NN$_\eta$-new in global oceans are still within the uncertainties obtained from field measurements.
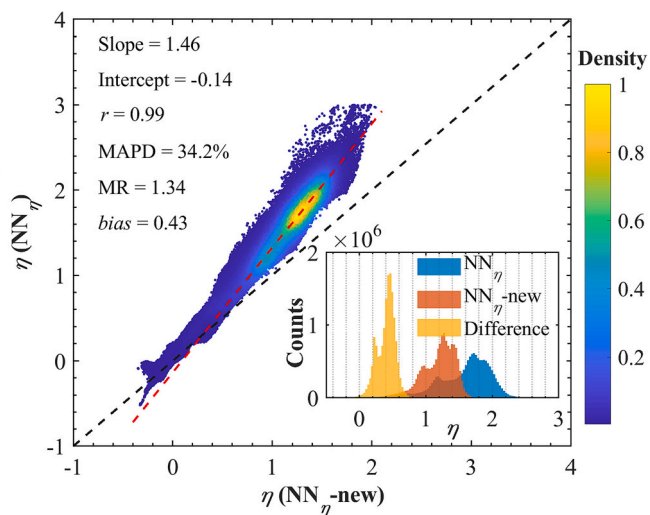
Fig. 11. Scatterplot between estimated η by NN_η and NN_η-new in global oceans using the MODIS monthly composite data in May 2018. The color of the scatter point indicates the frequency that is normalized to 1. The black dashed line indicates a 1:1 relationship, while the red dashed line indicates the best linear fit from Type-II regression. The insert figure shows the histograms of estimated η by NN_η and NN_η-new of the same MODIS data, as well as the difference between estimated η by the two schemes (*i.e.*, NN_η – NN_η-new). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

It is always preferred to have a training dataset that is as representative of the natural waters as possible for the robust performance of a neural network-based scheme. In this study, the uncertainties in the estimated η by NN_η associated with the dependency of the training dataset could be within an acceptable range. On one hand, the range of the synthetic η in *Syndata* was delicately simulated to cover the range of η obtained from the *in-situ* measurements in the global ocean (see Fig. 1b). On the other hand, the upper limit of η was determined by theoretic analysis (see Section 2.3). Thus, simulations in *Syndata* could be representative of most waters in the natural environment and the predicted η by NN_η could be more reliable than that by NN_η-new. However, the performance of NN_η could be further improved by updating the training dataset with more reliable field measurements of $R_{rs}(\lambda)$ and η from the global ocean.

4.3. Implication for the inversion of IOPs

In most IOPs retrieval algorithms, spectral $b_{bp}(\lambda)$ is obtained from $b_{bp}(\lambda_0)$ and η (Lee et al., 2002; Werdell et al., 2013). Thus, uncertainties in η would have direct impacts on the inverted $b_{bp}(\lambda)$ and $a(\lambda)$ (Lee et al., 2010), especially for the wavelengths that are far from the reference wavelength (Yu et al., 2021). Here we try to demonstrate the potential impacts of η on the inversed IOPs by evaluating the performance of QAA_v6 in the global ocean but with η adopting the values estimated from NN_η, with the new algorithm termed QAA_NN as in Lee et al. (2003). The performance of QAA_NN is evaluated by the median percentage difference (MPD) of derived IOPs by QAA_NN for $a(\lambda)$ and $b_{bp}(\lambda)$ at three MODIS bands centered at 412, 443, and 488 nm, with MPD calculated as 100% × median [(QAA_NN – QAA_v6) / QAA_v6]. Statistics of inversed IOPs for the four ROIs and the global oceans are tabulated in Table 1. Note that statistics for derived IOPs at 547 and 667 nm are not shown here as they are used as the reference wavelengths in QAA_v6 for clear and turbid waters, respectively. Thus, derived IOPs at 547 nm will not be affected by the selection of η values in clear waters and the same for derived IOPs at 667 nm for turbid waters.

For the four ROIs, η estimated by NN_η and QAA are overall comparable in SPG (MPD = 3.3%), while significant divergences are found in

Table 1

The median percentage difference (MPD) of derived IOPs by QAA_NN in comparison to that by QAA_v6. Statistics are presented for the four regions of interest and the global ocean for $a(\lambda)$ and $b_{bp}(\lambda)$ at 412, 443, and 488 nm, respectively. The last row presents the MPD of η estimated by NN_η compared to that by QAA_v6.

MPD (%)	TPO	SPG	NAO	YE	Global
$b_{bp}(412)$	−7.5	1.9	−3.3	−19.0	−5.5
$b_{bp}(443)$	−5.7	1.5	−2.5	−16.4	−4.1
$b_{bp}(488)$	−3.1	0.9	−1.4	−12.8	−2.3
$a(412)$	−3.1	0.9	−1.6	−18.0	−2.3
$a(443)$	−2.6	0.8	−1.3	−15.8	−1.9
$a(488)$	−1.5	0.6	−0.8	−12.4	−1.2
η	−14.5	3.3	−7.1	−64.2	−11.2

coastal and tropic oceans, especially in YE (MPD = −64.2%). As a result, derived $b_{bp}(\lambda)$ by QAA_NN differ largely from that by QAA_v6 in YE with MPD of derived $b_{bp}(412)$ approximating −19.0%. In comparison, the MPD of derived $b_{bp}(412)$ by QAA_NN in TPO, NAO, and SPG are −7.5%, −3.3, and 1.9, respectively, with the minimal difference found in SPG. On average, the estimated $b_{bp}(\lambda)$ by QAA_NN at 412 nm is 5.5% smaller than that by QAA_v6 in global oceans, and their differences narrow down toward longer wavelengths. Since $a(\lambda)$ is derived algebraically from $R_{rs}(\lambda)$ and $b_{bp}(\lambda)$, the changes in η would also affect the retrieved $a(\lambda)$, but the impacts are smaller than that on derived $b_{bp}(\lambda)$ as shown in Table 1. The negligible impact of η on retrieved $a(\lambda)$ in SPG (MPD < 1%) is mainly because $b_{bw}(\lambda)$ is the dominant factor for the inversion of $a(\lambda)$ as the values of $b_{bw}(\lambda)$ at these bands are overall greater than $b_{bp}(\lambda)$ in oligotrophic ocean gyres. Compared to the MPD values of η in the four ROIs, it can be found that η is a secondary parameter in regulating the retrievals of IOPs in QAA (Lee et al., 2002), especially for wavelengths close to the reference wavelength.

Nevertheless, as shown in Table 1, employing η values from NN_η would alter the retrieved values of IOPs, particularly for retrievals in turbid waters in YE, which will subsequently affect the quantitative evaluations of $b_{bp}(\lambda)$ derivative products, such as POC (Stramski et al., 1999) and particle mass concentrations (Boss et al., 2009). Thus, accurate estimations of η in the global ocean are certainly desired. Based on evaluations with the *in-situ* dataset and satellite imagery discussed in this effort, estimated η by NN_η could be more accurate and reasonable than that in QAA for the global ocean and is thus recommended for the retrieval of IOPs by QAA_v6 or GIOP in the future applications.

5. Conclusions

In this study, the spectral $R_{rs}(\lambda)$ at six MODIS visible bands were used to estimate η with a scheme based on a neural network (NN_η). Evaluation results with field measurements, synthetic datasets, and MODIS imagery show that NN_η could have overall better performance than the conventional schemes (LS2, MM01, and QAA) with improved retrieval accuracy when validated with quality-controlled field measurements. More importantly, η estimated by NN_η presents more reasonable patterns in both the spatial distribution and the seasonal variability, which are consistent with field observations and can be well explained. In contrast, the conventional schemes, with their limitation and uncertainties discussed in detail in this effort, predict questionable η in the global ocean in either the magnitude or the spatial distribution pattern. In particular, the comparison between NN_η and LS2 shows that it is better to directly estimate η from $R_{rs}(\lambda)$ when a neural network-based scheme is employed, as intermediate products, such as R_{rs}-derived $K_d(\lambda)$ and $b_{bp}(\lambda)$ in LS2, could introduce extra uncertainties to the estimated η. With the implementation of NN_η to the MODIS monthly composites from 2003 to 2021, we show that η features strong seasonal cycles in most of the global ocean, but the decadal changes in η are rather insignificant in the majority of the global oceans, at least for the period investigated in this effort.

It is worth pointing out that NN_η, the same as other algorithms based on NN, is dependent on the training dataset, particularly the range of η in the training dataset. As the training dataset (*Syndata*) for NN_η covers a rather broader range of water types that could be encountered in natural waters, the applicability of NN_η in the global ocean is ensured. Further efforts are still required to obtain reliable $b_{bp}(\lambda)$ measurements in various types of water in the global ocean to 1) evaluate the performance of NN_η, and 2) expand the training dataset used in this effort.

CRediT authorship contribution statement

Xiaolong Yu: Conceptualization, Writing – original draft, Methodology, Writing – review & editing. **Zhongping Lee:** Conceptualization, Writing – review & editing. **Wendian Lai:** Software.

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 Python scripts for the training of the neural network for NN_η, as well as the trained network, are made available at https://github.com/oceanopticsxmu/eta_backscattering to facilitate further evaluation and comparison. Associated data used in this study, including the compiled *in-situ* dataset and global mapping products of η obtained by multiple schemes in May 2018, can also be retrieved from the above link. Mapping products of monthly η derived by NN_η from January 2003 to December 2021 can be requested *via* email to the authors.

MATLAB scripts for LS2, provided by Dr. Hubert Loisel, can be accessed from https://github.com/SIO-Ocean-Optics-Research-Laboratory/Kd_NN_Distribution and https://github.com/SIO-Ocean-Optics-Research-Laboratory/LS2_Distribution for the inversion of K_d and IOPs, respectively.

References

Aas, E., Høkedal, J., Sørensen, K., 2005. Spectral backscattering coefficient in coastal waters. Int. J. Remote Sens. 26, 331–343.

Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M., Ghemawat, S., Irving, G., Isard, M., 2016. TensorFlow: A system for large-scale machine learning. In: 12th USENIX Symposium on Operating Systems Design and Implementation (OSDI 16), pp. 265–283.

Antoine, D., Siegel, D.A., Kostadinov, T., Maritorena, S., Nelson, N.B., Gentili, B., Vellucci, V., Guillocheau, N., 2011. Variability in optical particle backscattering in contrasting bio-optical oceanic regimes. Limnol. Oceanogr. 56, 955–973.

Babin, M., Morel, A., Fournier-Sicre, V., Fell, F., Stramski, D., 2003. Light scattering properties of marine particles in coastal and open ocean waters as related to the particle mass concentration. Limnol. Oceanogr. 48, 843–859.

Behrenfeld, M.J., Boss, E., Siegel, D.A., Shea, D.M., 2005. Carbon-Based Ocean productivity and phytoplankton physiology from space. Glob. Biogeochem. Cycles 19, GB1006.

Boss, E., Pegau, W., Lee, M., Twardowski, M., Shybanov, E., Korotaev, G., Baratange, F., 2004. Particulate backscattering ratio at LEO 15 and its use to study particle composition and distribution. J. Geophys. Re. Oceans 109, C01014.

Boss, E., Pegau, W.S., Gardner, W.D., Zaneveld, J.R.V., Barnard, A.H., Twardowski, M.S., Chang, G., Dickey, T., 2001. Spectral particulate attenuation and particle size distribution in the bottom boundary layer of a continental shelf. J. Geophys. Res. Oceans 106, 9509–9516.

Boss, E., Taylor, L., Gilbert, S., Gundersen, K., Hawley, N., Janzen, C., Johengen, T., Purcell, H., Robertson, C., Schar, D.W., 2009. Comparison of inherent optical properties as a surrogate for particulate matter concentration in coastal waters. Limnol. Oceanogr. Methods 7, 803–810.

Brewin, R.J., Hardman-Mountford, N.J., Lavender, S.J., Raitsos, D.E., Hirata, T., Uitz, J., Devred, E., Bricaud, A., Ciotti, A., Gentili, B., 2011. An intercomparison of bio-optical techniques for detecting dominant phytoplankton size class from satellite remote sensing. Remote Sens. Environ. 115, 325–339.

Brewin, R.J., Sathyendranath, S., Müller, D., Brockmann, C., Deschamps, P.-Y., Devred, E., Doerffer, R., Fomferra, N., Franz, B., Grant, M., 2015. The ocean colour climate change initiative: III. A round-robin comparison on in-water bio-optical algorithms. Remote Sens. Environ. 162, 271–294.

Buckton, D., Omongain, E., Danaher, S., 1999. The use of neural networks for the estimation of oceanic constituents based on the MERIS instrument. Int. J. Remote Sens. 20, 1841–1851.

Buonassissi, C., Dierssen, H., 2010. A regional comparison of particle size distributions and the power law approximation in oceanic and estuarine surface waters. J. Geophys. Res. Oceans 115, C10028.

Carder, K.L., Chen, F., Lee, Z., Hawes, S.K., Kamykowski, D., 1999. Semianalytic moderate-resolution imaging spectrometer algorithms for chlorophyll a and absorption with bio-optical domains based on nitrate-depletion temperatures. J. Geophys. Res. Oceans 104, 5403–5421.

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., 2020. A global compilation of in situ aquatic high spectral resolution inherent and apparent optical property data for remote sensing applications. Earth Syst.Sci. Data 12, 1123–1139.

Chang, G., Barnard, A., Zaneveld, J.R.V., 2007. Optical closure in a complex coastal environment: particle effects. Appl. Opt. 46, 7679–7692.

Chang, G., Whitmire, A.L., 2009. Effects of bulk particle characteristics on backscattering and optical closure. Opt. Express 17, 2132–2142.

Chen, J., Quan, W., Cui, T., Song, Q., Lin, C., 2014. Remote sensing of absorption and scattering coefficient using neural network model: development, validation, and application. Remote Sens. Environ. 149, 213–226.

Chen, S.L., Zhang, G.A., Yang, S.L., 2003. Temporal and spatial changes of suspended sediment concentration and resuspension in the Yangtze River estuary. J. Geogr. Sci. 13, 498–506.

Ciotti, Á.M., Cullen, J.J., Lewis, M.R., 1999. A semi-analytical model of the influence of phytoplankton community structure on the relationship between light attenuation and ocean color. J. Geophys. Res. Oceans 104, 1559–1578.

Claustre, H., Sciandra, A., Vaulot, D., 2008. Introduction to the special section bio-optical and biogeochemical conditions in the south East Pacific in late 2004: the BIOSOPE program. Biogeosciences 5, 679–691.

Cota, G.F., Wang, H., Comiso, J.C., 2004. Transformation of global satellite chlorophyll retrievals with a regionally tuned algorithm. Remote Sens. Environ. 90, 373–377.

Fujii, M., Boss, E., Chai, F., 2007. The value of adding optics to ecosystem models: a case study. Biogeosciences 4, 817–835.

Fujiwara, A., Hirawake, T., Suzuki, K., Saitoh, S.-I., 2011. Remote sensing of size structure of phytoplankton communities using optical properties of the chukchi and Bering Sea shelf region. Biogeosciences 8, 3567–3580.

Gallegos, C., Davies-Colley, R., Gall, M., 2008. Optical closure in lakes with contrasting extremes of reflectance. Limnol. Oceanogr. 53, 2021–2034.

Gittings, J.A., Raitsos, D.E., Krokos, G., Hoteit, I., 2018. Impacts of warming on phytoplankton abundance and phenology in a typical tropical marine ecosystem. Sci. Rep. 8, 1–12.

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.-Atmos. 93, 10909–10924.

Gordon, H.R., Morel, A.Y., 1983. Remote assessment of ocean color for interpretation of satellite visible imagery: A review. Springer-Verlag, New York.

Greene, C.A., Thirumalai, K., Kearney, K.A., Delgado, J.M., Schwanghart, W., Wolfenbarger, N.S., Thyng, K.M., Gwyther, D.E., Gardner, A.S., Blankenship, D.D., 2019. The climate data toolbox for MATLAB. Geochem. Geophys. Geosyst. 20, 3774–3781.

Hu, C., Lee, Z., Franz, B., 2012. Chlorophyll a algorithms for oligotrophic oceans: a novel approach based on three-band reflectance difference. J. Geophys. Res. Oceans 117, C01011.

Ioannou, I., Gilerson, A., Gross, B., Moshary, F., Ahmed, S., 2013. Deriving Ocean color products using neural networks. Remote Sens. Environ. 134, 78–91.

IOCCG-OCAG, 2003. Models, parameters, and approaches that used to generate wide range of absorption and backscattering spectra. available at:. IOCCG http://ioccg.org/wp-content/uploads/2016/2003/lee-data.pdf.

IOCCG, 2006. Remote sensing of inherent optical properties: Fundamentals, tests of algorithms, and applications. In: Lee, Z.P. (Ed.), Reports of the International Ocean-Colour Coordinating Group, No. 5. IOCCG, Dartmouth, Canada.

Jamet, C., Loisel, H., Dessailly, D., 2012. Retrieval of the spectral diffuse attenuation coefficient Kd(λ) in open and coastal ocean waters using a neural network inversion. J. Geophys. Res. 117, C10023.

Jorge, D.S., Loisel, H., Jamet, C., Dessailly, D., Demaria, J., Bricaud, A., Maritorena, S., Zhang, X., Antoine, D., Kutser, T., 2021. A three-step semi analytical algorithm (3SAA) for estimating inherent optical properties over oceanic, coastal, and inland waters from remote sensing reflectance. Remote Sens. Environ. 263, 112537.

Kingma, D.P., Ba, J., 2014. Adam: A method for stochastic optimization arXiv preprint arXiv:1412.6980.

Kostadinov, T., Siegel, D., Maritorena, S., 2009. Retrieval of the particle size distribution from satellite ocean color observations. J. Geophys. Res. Oceans 114, C09015.

Kostadinov, T., Siegel, D., Maritorena, S., 2010. Global variability of phytoplankton functional types from space: assessment via the particle size distribution. Biogeosciences 7, 3239–3257.

Kostadinov, T.S., Siegel, D.A., Maritorena, S., Guillocheau, N., 2012. Optical assessment of particle size and composition in the Santa Barbara Channel, California. Appl. Opt. 51, 3171–3189.

Krizhevsky, A., Sutskever, I., Hinton, G.E., 2012. Imagenet classification with deep convolutional neural networks. In: Advances in Neural Information Processing Systems, 25.

Kutser, T., Hiire, M., Metsamaa, L., Vahtmäe, E., Paavel, B., Aps, R., 2009. Field measurements of spectral backscattering coefficient of the Baltic Sea and boreal lakes. Boreal Environ. Res. 14, 305–312.

Lai, W., Lee, Z., Wang, J., Wang, Y., Garcia, R., Zhang, H., 2022. A portable algorithm to retrieve bottom depth of optically shallow waters from top-of-atmosphere measurements. J. Remote Sens. 2022 https://doi.org/10.34133/2022/9831947.

Le, C., Hu, C., Cannizzaro, J., English, D., Muller-Karger, F., Lee, Z., 2013. Evaluation of chlorophyll-a remote sensing algorithms for an optically complex estuary. Remote Sens. Environ. 129, 75–89.

Lee, Z., Arnone, R., Hu, C., Werdell, P.J., Lubac, B., 2010. Uncertainties of optical parameters and their propagations in an analytical ocean color inversion algorithm. Appl. Opt. 49, 369–381.

Lee, Z., Carder, K.L., Arnone, R.A., 2002. Deriving inherent optical properties from water color: a multiband quasi-analytical algorithm for optically deep waters. Appl. Opt. 41, 5755–5772.

Lee, Z., Carder, K.L., Du, K., 2004. Effects of molecular and particle scatterings on the model parameter for remote-sensing reflectance. Appl. Opt. 43, 4957–4964.

Lee, Z., Du, K., Arnone, R., 2005a. A model for the diffuse attenuation coefficient of downwelling irradiance. J. Geophys. Res. Oceans 110, C02016.

Lee, Z., Sandidge, J., Zhang, M., 2003. Ocean-color inversion: a combined approach by analytical solution and neural networks. In: Ocean Remote Sensing and Imaging II. SPIE, pp. 153–161.

Lee, Z., Wei, J., Voss, K., Lewis, M., Bricaud, A., Huot, Y., 2015. Hyperspectral absorption coefficient of "pure" seawater in the range of 350–550 nm inverted from remote sensing reflectance. Appl. Opt. 54, 546–558.

Lee, Z., Weidemann, A., Kindle, J., Arnone, R., Carder, K.L., Davis, C., 2007. Euphotic zone depth: its derivation and implication to ocean-color remote sensing. J. Geophys. Res. Oceans 112, C03009.

Lee, Z.P., Darecki, M., Carder, K.L., Davis, C.O., Stramski, D., Rhea, W.J., 2005b. Diffuse attenuation coefficient of downwelling irradiance: an evaluation of remote sensing methods. J. Geophys. Res. Oceans 110, C02017.

Lobo, F.L., Costa, M., Phillips, S., Young, E., McGregor, C., 2014. Light backscattering in turbid freshwater: a laboratory investigation. J. Appl. Remote. Sens. 8, 083611.

Loisel, H., Nicolas, J.M., Sciandra, A., Stramski, D., Poteau, A., 2006. Spectral dependency of optical backscattering by marine particles from satellite remote sensing of the global ocean. J. Geophys. Res. Oceans 111, C09024.

Loisel, H., Stramski, D., 2000. Estimation of the inherent optical properties of natural waters from the irradiance attenuation coefficient and reflectance in the presence of raman scattering. Appl. Opt. 39, 3001–3011.

Loisel, H., Stramski, D., Dessailly, D., Jamet, C., Li, L., Reynolds, R.A., 2018. An inverse model for estimating the optical absorption and backscattering coefficients of seawater from remote-sensing reflectance over a broad range of oceanic and coastal marine environments. J. Geophys. Res. Oceans 123, 2141–2171.

Loisel, H., Stramski, D., Mitchell, B.G., Fell, F., Fournier-Sicre, V., Lemasle, B., Babin, M., 2001. Comparison of the ocean inherent optical properties obtained from measurements and inverse modeling. Appl. Opt. 40, 2384–2397.

McKee, D., Chami, M., Brown, I., Calzado, V.S., Doxaran, D., Cunningham, A., 2009. Role of measurement uncertainties in observed variability in the spectral backscattering ratio: a case study in mineral-rich coastal waters. Appl. Opt. 48, 4663–4675.

Mobley, C., Sundman, L., 2008. Hydrolight 5 ecolight 5 users' guide. Sequoia Scientific Inc., US.

Morel, A., 1974. Optical properties of pure water and pure sea water. Opt. Aspects Oceanogr. 1, 1–24.

Morel, A., Claustre, H., Gentili, B., 2010. The most oligotrophic subtropical zones of the global ocean: similarities and differences in terms of chlorophyll and yellow substance. Biogeosciences 7, 3139–3151.

Morel, A., Gentili, B., Claustre, H., Babin, M., Bricaud, A., Ras, J., Tieche, F., 2007. Optical properties of the "clearest" natural waters. Limnol. Oceanogr. 52, 217–229.

Morel, A., Maritorena, S., 2001. Bio-optical properties of oceanic waters: a reappraisal. J. Geophys. Res. Oceans 106, 7163–7180.

Morel, A., Prieur, L., 1977. Analysis of variations in ocean color. Limnol. Oceanogr. 22, 709–722.

O'Reilly, J.E., Maritorena, S., Mitchell, B.G., Siegel, D.A., Carder, K.L., Garver, S.A., Kahru, M., McClain, C., 1998. Ocean color chlorophyll algorithms for SeaWiFS. J. Geophys. Res. Oceans 103, 24937.

Petzold, T.J., 1972. Volume scattering functions for selected ocean waters. In: DTIC Document.

Pope, R.M., Fry, E.S., 1997. Absorption spectrum (380–700 nm) of pure water. II. Integrating cavity measurements. Appl. Opt. 36, 8710–8723.

Reynolds, R., Stramski, D., Wright, V., Woźniak, S., 2010. Measurements and characterization of particle size distributions in coastal waters. J. Geophys. Res. Oceans 115, C08024.

Reynolds, R.A., Stramski, D., 2021. Variability in oceanic particle size distributions and estimation of size class contributions using a non-parametric approach. J. Geophys. Res. Oceans 126, e2021JC017946.

Reynolds, R.A., Stramski, D., Mitchell, B.G., 2001. A chlorophyll-dependent semianalytical reflectance model derived from field measurements of absorption and backscattering coefficients within the Southern Ocean. J. Geophys. Res. Oceans 106, 7125–7138.

Reynolds, R.A., Stramski, D., Neukermans, G., 2016. Optical backscattering by particles in Arctic seawater and relationships to particle mass concentration, size distribution, and bulk composition. Limnol. Oceanogr. 61, 1869–1890.

Röttgers, R., McKee, D., Utschig, C., 2014. Temperature and salinity correction coefficients for light absorption by water in the visible to infrared spectral region. Opt. Express 22, 25093–25108.

Roxy, M.K., Modi, A., Murtugudde, R., Valsala, V., Panickal, S., Prasanna Kumar, S., Ravichandran, M., Vichi, M., Lévy, M., 2016. A reduction in marine primary productivity driven by rapid warming over the tropical Indian Ocean. Geophys. Res. Lett. 43, 826–833.

Seegers, B.N., Stumpf, R.P., Schaeffer, B.A., Loftin, K.A., Werdell, P.J., 2018. Performance metrics for the assessment of satellite data products: an ocean color case study. Opt. Express 26, 7404–7422.

Shang, S., Lee, Z., Lin, G., Li, Y., Li, X., 2019. Progressive scheme for blending empirical ocean color retrievals of absorption coefficient and chlorophyll concentration from open oceans to highly turbid waters. Appl. Opt. 58, 3359–3369.

Shen, F., Verhoef, W., Zhou, Y.X., Salama, M.S., Liu, X.L., 2010. Satellite estimates of wide-range suspended sediment concentrations in changjiang (Yangtze) estuary using MERIS data. Estuar. Coasts 33, 1420–1429.

Snyder, W.A., Arnone, R.A., Davis, C.O., Goode, W., Gould, R.W., Ladner, S., Lamela, G., Rhea, W.J., Stavn, R., Sydor, M., 2008. Optical scattering and backscattering by organic and inorganic particulates in US coastal waters. Appl. Opt. 47, 666–677.

Stramska, M., Stramski, D., Hapter, R., Kaczmarek, S., Stoń, J., 2003. Bio-optical relationships and ocean color algorithms for the north polar region of the Atlantic. J. Geophys. Res. Oceans 108, 3143.

Stramski, D., Boss, E., Bogucki, D., Voss, K.J., 2004. The role of seawater constituents in light backscattering in the ocean. Prog. Oceanogr. 61, 27–56.

Stramski, D., Reynolds, R.A., 2018. DS3 ocean optics dataset. In: PANGAEA.

Stramski, D., Reynolds, R.A., Babin, M., Kaczmarek, S., Lewis, M.R., Röttgers, R., Sciandra, A., Stramska, M., Twardowski, M.S., Franz, B.A., Claustre, H., 2008. Relationships between the surface concentration of particulate organic carbon and optical properties in the eastern South Pacific and eastern Atlantic oceans. Biogeosciences 5, 171–201.

Stramski, D., Reynolds, R.A., Kahru, M., Mitchell, B.G., 1999. Estimation of particulate organic carbon in the ocean from satellite remote sensing. Science 285, 239–242.

Tanaka, A., Kishino, M., Doerffer, R., Schiller, H., Oishi, T., Kubota, T., 2004. Development of a neural network algorithm for retrieving concentrations of chlorophyll, suspended matter and yellow substance from radiance data of the ocean color and temperature scanner. J. Oceanogr. 60, 519–530.

Terzić, E., Miró, A., Organelli, E., Kowalczuk, P., D'Ortenzio, F., Lazzari, P., 2021. Radiative transfer modeling with biogeochemical-Argo float data in the Mediterranean Sea. J. Geophys. Res. Oceans 126, e2021JC017690.

Twardowski, M., Claustre, H., Freeman, S., Stramski, D., Huot, Y., 2007. Optical backscattering properties of the" clearest" natural waters. Biogeosciences 4, 1041–1058.

Twardowski, M.S., Boss, E., Macdonald, J.B., Pegau, W.S., Barnard, A.H., Zaneveld, J.R. V., 2001. A model for estimating bulk refractive index from the optical backscattering ratio and the implications for understanding particle composition in case I and case II waters. J. Geophys. Res. Oceans 106, 14129–14142.

Tzortziou, M., Herman, J.R., Gallegos, C.L., Neale, P.J., Subramaniam, A., Harding Jr., L. W., Ahmad, Z., 2006. Bio-optics of the Chesapeake Bay from measurements and radiative transfer closure. Estuar. Coast. Shelf Sci. 68, 348–362.

Vadakke-Chanat, S., Shanmugam, P., 2019. Retrieval of spectral backscattering from spectral scattering based on spectral partitioning technique. Estuar. Coast. Shelf Sci. 217, 196–205.

Vantrepotte, V., Loisel, H., Mélin, F., Desailly, D., Duforêt-Gaurier, L., 2011. Global particulate matter pool temporal variability over the SeaWiFS period (1997–2007). Geophys. Res. Lett. 38, L02605.

Wang, W., Lee, Z., Wei, J., Shang, S., Wang, M., Lai, W., 2021. Extending satellite ocean color remote sensing to the near-blue ultraviolet bands. Remote Sens. Environ. 253, 112228.

Wei, G., Lee, Z., Wu, X., Yu, X., Shang, S., Letelier, R., 2021a. Impact of temperature on absorption coefficient of pure seawater in the blue wavelengths inferred from satellite and in situ measurements. J. Remote Sens. 2021 https://doi.org/10.34133/32021/9842702.

Wei, J., Wang, M., Jiang, L., Yu, X., Mikelsons, K., Shen, F., 2021b. Global estimation of suspended particulate matter from Satellite Ocean color imagery. J. Geophys. Res. Oceans 126, e2021JC017303.

Wei, J., Wang, M., Lee, Z., Ondrusek, M., Zhang, S., Ladner, S., 2021c. Experimental analysis of the measurement precision of spectral water-leaving radiance in different water types. Opt. Express 29, 2780–2797.

Wei, J., Yu, X., Lee, Z., Wang, M., Jiang, L., 2020. Improving low-quality satellite remote sensing reflectance at blue bands over coastal and inland waters. Remote Sens. Environ. 250, 112029.

Wei, J.W., Lee, Z.P., Shang, S.L., 2016. A system to measure the data quality of spectral remote-sensing reflectance of aquatic environments. J. Geophys. Res. Oceans 121, 8189–8207.

Werdell, P.J., Bailey, S.W., 2005. An improved in-situ bio-optical data set for ocean color algorithm development and satellite data product validation. Remote Sens. Environ. 98, 122–140.

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

Westberry, T., Behrenfeld, M., Siegel, D., Boss, E., 2008. Carbon-based primary productivity modeling with vertically resolved photoacclimation. Glob. Biogeochem. Cycles 22, GB2024.

Xu, J., Bian, Y., Lyu, H., Miao, S., Li, Y., Liu, H., Xu, J., 2021. Estimation of particulate backscattering coefficient in turbid inland water using sentinel 3A-OLCI image. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 14, 8577–8593.

Yu, X., Lee, Z., Shang, S., Wang, M., Jiang, L., 2022. Estimating the water-leaving albedo from ocean color. Remote Sens. Environ. 269, 112807.

Yu, X., Lee, Z., Shang, Z., Lin, H., Lin, G., 2021. A simple and robust shade correction scheme for remote sensing reflectance obtained by the skylight-blocked approach. Opt. Express 29, 470–486.

Yu, X., Lee, Z., Shen, F., Wang, M., Wei, J., Jiang, L., Shang, Z., 2019a. An empirical algorithm to seamlessly retrieve the concentration of suspended particulate matter from water color across ocean to turbid river mouths. Remote Sens. Environ. 235, 111491.

Yu, X., Lee, Z., Wei, J., Shang, S., 2019b. Impacts of pure seawater absorption coefficient on remotely sensed inherent optical properties in oligotrophic waters. Opt. Express 27, 34974–34984.

Yu, X., Salama, M.S., Shen, F., Verhoef, W., 2016. Retrieval of the diffuse attenuation coefficient from GOCI images using the 2SeaColor model: a case study in the Yangtze estuary. Remote Sens. Environ. 175, 109–119.

Zhang, X., Hu, L., 2009. Scattering by pure seawater at high salinity. Opt. Express 17, 12685–12691.

Zhang, X., Hu, L., He, M., 2009. Scattering by pure seawater: effect of salinity. Opt. Express 17, 5698–5710.