Remote Sensing of Environment 270 (2022) 112879

Contents lists available at ScienceDirect

Remote Sensing of Environment

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

Modeling surface ocean phytoplankton pigments from hyperspectral remote sensing reflectance on global scales

Sasha J. Kramer [a,b,*], David A. Siegel [a], Stéphane Maritorena [a], Dylan Catlett [a,b,1]

[a] Earth Research Institute, University of California Santa Barbara, Santa Barbara, CA, USA
[b] Interdepartmental Graduate Program in Marine Science, University of California Santa Barbara, Santa Barbara, CA, USA

ARTICLE INFO

Editor: Menghua Wang

Keywords:
Phytoplankton
HPLC pigments
Remote sensing reflectance
Principal components regression modeling
Bio-optics

ABSTRACT

Phytoplankton community composition impacts food webs, climate, and fisheries on regional and global scales, and can be assessed at coarse taxonomic resolution from biomarker pigments measured using high-performance liquid chromatography (HPLC). Presently, satellite ocean color provides unprecedented coverage of the global surface ocean and offers reliable estimates of bulk biological properties; however, existing multispectral sensors have limited ability to provide information about phytoplankton community composition. Satellite ocean color at hyperspectral resolution (e.g., NASA's upcoming Plankton, Aerosol, Cloud, and ocean Ecosystem sensor, PACE) is expected to improve estimates of phytoplankton community composition from space. Phytoplankton impact ocean color via contributions to absorption and fluorescence (through phytoplankton pigments) and scattering, especially on narrow spectral scales (5–100 nm). Here, a global open ocean dataset of concurrent HPLC pigments and hyperspectral remote sensing reflectance ($R_{rs}(\lambda)$) observations is used to model phytoplankton pigment composition from optical data. Phytoplankton pigments are reconstructed from $R_{rs}(\lambda)$ using optimized principal components regression modeling. This work demonstrates that thirteen phytoplankton pigments, representing five phytoplankton pigment groups (e.g., diatoms, dinoflagellates, haptophytes, green algae, and cyanobacteria), can be modeled from hyperspectral $R_{rs}(\lambda)$. Spectral information needed to model each phytoplankton pigment concentration is found throughout the entire visible spectrum and the model results are best at high spectral resolution ($\leq$5 nm). The resulting model recreates observed relationships among pigment concentrations, providing support for the designation of five pigment-based phytoplankton groups for the global open ocean. This work represents a step toward developing robust, global spectral models for phytoplankton pigment composition. However, more high-quality data from a wide range of ecosystems and environments are still needed to achieve this goal.

1. Introduction

Phytoplankton community composition has a strong influence on the structure of planktic ecosystems, global biogeochemical cycles, and the ecosystem services that the oceans provide (Legendre, 1990; Vanni and Findlay, 1990; Le Quéré et al., 2005; Falkowski and Oliver, 2007). Characterizing the diversity of phytoplankton is crucial to develop marine food web and ocean carbon cycle models with improved accuracy (e.g., Legendre, 1990; Siegel et al., 2014). Satellite ocean color sensors observe surface ocean properties on unparalleled spatiotemporal scales, including parameters relevant to phytoplankton abundance and community composition, such as chlorophyll-*a* concentration (e.g., O'Reilly et al., 1998; Hu et al., 2012), colored dissolved and detrital materials (e.g., Siegel et al., 2002; Morel and Gentili, 2009), particulate backscattering (e.g., Stramski et al., 2001; Kostadinov et al., 2010), and particulate absorption (e.g., Ciotti and Bricaud, 2006; Chase et al., 2013). Many methods have also been developed to characterize phytoplankton community composition from ocean color measurements, including both phytoplankton abundance-based (e.g., Brewin et al., 2010; Hirata et al., 2011) and radiance-based (e.g., Alvain et al., 2008; Bracher et al., 2009; Uitz et al., 2015; Chase et al., 2017) approaches (see Mouw et al., 2017 and Bracher et al., 2017 for reviews of these approaches). With the upcoming launch of NASA's Plankton, Aerosol, Cloud, and ocean Ecosystem (PACE) mission, the spectral resolution and range of satellite

* Corresponding author at: Earth Research Institute, University of California Santa Barbara, Santa Barbara, CA, USA.
E-mail address: sasha.kramer@lifesci.ucsb.edu (S.J. Kramer).
[1] Present address: Biology Department, Woods Hole Oceanographic Institution, Woods Hole, MA, USA.

https://doi.org/10.1016/j.rse.2021.112879
Received 8 September 2021; Received in revised form 24 December 2021; Accepted 27 December 2021
Available online 6 January 2022

ocean color data will increase dramatically (Werdell et al., 2019). Improving the spectral resolution of ocean color measurements from multispectral to hyperspectral is expected to provide improved estimates of phytoplankton community composition from satellites (Wolanin et al., 2016; Xi et al., 2017; Werdell et al., 2018; Cael et al., 2020), highlighting the need for new phytoplankton community composition algorithms that take advantage of this higher spectral resolution.

Many ocean color models that separate groups of phytoplankton target spectral variations in remote sensing reflectance ($R_{rs}(\lambda)$) to retrieve information about phytoplankton community composition, relying on differences in the shape and magnitude of $R_{rs}(\lambda)$ introduced by phytoplankton pigment absorption (e.g., Alvain et al., 2005; Torrecilla et al., 2011; Bracher et al., 2015a; Uitz et al., 2015; Chase et al., 2017). The shape and magnitude of $R_{rs}(\lambda)$ are also dependent on other absorbing and scattering components in the ocean, including seawater, non-algal particles (NAP), and colored dissolved organic matter (CDOM). The optical properties of many of these oceanic constituents are either well characterized (i.e., absorption and scattering by seawater) or have simple spectral shapes that change over long ($\geq$100 nm) spectral scales (i.e., absorption by CDOM and NAP, scattering by NAP). Conversely, variability in phytoplankton absorption and some scattering features occurs on narrower spectral scales ($<$100 nm; Bidigare et al., 1989; Bricaud et al., 2004). Improvements in assessing phytoplankton abundance and composition from hyperspectral reflectance may be made by first accounting for the broader absorption and scattering signals associated with CDOM and NAP, and then isolating and enhancing the phytoplankton-specific features in absorption and scattering.

Ocean color modeling approaches to describe phytoplankton communities must be carefully constructed to account for both the input $R_{rs}(\lambda)$ data quality and the phytoplankton community metrics targeted (e.g., cell size, pigment composition, functional traits, etc.). In addition to the variability in $R_{rs}(\lambda)$ shape and magnitude caused by oceanic constituents other than phytoplankton, further uncertainty and variation is introduced to satellite-derived $R_{rs}(\lambda)$ by atmospheric correction (Werdell et al., 2018). Derivative methods that isolate spectral features of interest are therefore well suited to high spectral resolution data: these methods are less sensitive to the uncertainties in spectral magnitude introduced by other optically-relevant components of the surface ocean and atmosphere and magnify the variations in spectral shape (e.g., Tsai and Philpot, 1998; Taylor et al., 2011; Torrecilla et al., 2011; Xi et al., 2015; Uitz et al., 2015; Catlett and Siegel, 2018). However, spectral derivative methods can also accentuate instrument- and dataset-specific noise in bio-optical measurements (Tsai and Philpot, 1998), emphasizing the need to evaluate the utility of spectral derivative methods in approaches to reconstruct phytoplankton pigments and assess phytoplankton pigment composition from hyperspectral optics.

The validation method for any ocean color phytoplankton composition model is also important, as it determines the taxonomic scope and resolution of the model. While there are many available methods of characterizing phytoplankton community composition in situ, high performance liquid chromatography (HPLC) measurements of phytoplankton pigment concentrations are currently the most globally-available, consistent, quality-controlled data for validating phytoplankton community composition models (Mouw et al., 2017; Kramer and Siegel, 2019). HPLC pigment measurements are widespread in the global surface ocean relative to other characterizations of phytoplankton community composition and offer taxonomic information to broad group levels (see Kramer and Siegel, 2019). While pigments offer limited taxonomic resolution of phytoplankton composition compared to other, more taxonomically resolved methods (i.e., quantitative cell imaging [Chase et al., 2020], next generation sequencing [Lin et al., 2019], etc.) and inference of pigment-based taxonomy is not straightforward, retrieval of phytoplankton pigment concentrations from ocean color data is the first step required to assess phytoplankton composition from space.

Here, we quantify phytoplankton pigment concentrations using principal components regression modeling applied to a global surface ocean dataset of hyperspectral $R_{rs}(\lambda)$ spectra. The models are developed and validated using a paired dataset of globally-distributed HPLC pigment samples. Reflectance residuals were calculated between measured $R_{rs}(\lambda)$ data and $R_{rs}(\lambda)$ constructed from a generic reflectance model. The use of residual spectra removes many of the optical features that vary on long spectral scales (e.g., absorption and/or scattering by seawater, NAP, and CDOM) while enhancing the narrower spectral features, which may be associated with variations in absorption and scattering for the different pigment-based phytoplankton groups. Derivative analysis was then performed on the residual spectra to further enhance these narrow spectral features. $R_{rs}(\lambda)$ residual derivatives were used in an optimized principal components regression modeling framework to retrieve the concentrations of various phytoplankton pigments. This approach reconstructs representative pigment concentrations from five pigment-based phytoplankton groups and preserves the co-variability between and among phytoplankton pigment concentrations. Ultimately, the phytoplankton pigment composition model presented here demonstrates the utility of the spectral gap hypothesis for modeling phytoplankton pigments from hyperspectral data. Specifically, it shows that phytoplankton pigment concentrations can be successfully estimated from hyperspectral $R_{rs}(\lambda)$ when the fine-scale features most strongly correlated with phytoplankton absorption and scattering are isolated and compositional differences from base-state conditions are accentuated, while other features that vary on long spectral scales are removed.

2. Materials and methods

2.1. HPLC dataset construction and quality control

The global HPLC pigment dataset used in this analysis was constructed following the criteria defined in Kramer and Siegel (2019). Samples from the surface ocean (depths of 7 m or less) were analyzed at a small number of labs to reduce lab-dependent variability in the dataset. All samples had a consistent suite of HPLC pigments measured between samples. The initial dataset (from Kramer and Siegel, 2019) included 4480 samples. 70 additional surface samples collected as part of the EXport Processes in the Ocean from RemoTe Sensing (EXPORTS) North Pacific field campaign in August–September 2018 and analyzed at NASA Goddard Space Flight Center (GSFC) following Van Heukelem and Thomas (2001) were added to the Kramer and Siegel (2019) dataset for 4550 samples total. All pigment values below established HPLC method detection limits were set to zero (Van Heukelem and Thomas, 2001). If replicate samples of HPLC pigments were taken at a given site, an average of the replicates was used before the matchup procedure was applied.

The thirteen HPLC pigments used in all subsequent analyses (and their abbreviations) include: total chlorophyll-*a* (Tchla), 19′-hexanoyloxyfucoxanthin (HexFuco), 19′-butanoyloxyfucoxanthin (ButFuco), alloxanthin (Allo), fucoxanthin (Fuco), peridinin (Perid), zeaxanthin (Zea), divinyl chlorophyll *a* (DVchla), monovinyl chlorophyll *b* (MVchlb), chlorophyll $c_1 + c_2$ (Chlc12), chlorophyll c_3 (Chlc3), neoxanthin (Neo), and violaxanthin (Viola). Several pigments were measured in all datasets but not included for analysis, including: pigments that were redundant or not useful as taxonomic markers (total chlorophyll *b*, total chlorophyll *c*, alpha-beta carotene, diatoxanthin, diadinoxanthin; Kramer and Siegel, 2019); degradation pigments (chlorophyllide, phaeophytin, phaeophorbide); and pigments that were not detected or measured below established method detection limits (defined following Van Heukelem and Thomas, 2001) in >75% of samples in the final matchup dataset (divinyl chlorophyll *b*, lutein, and prasinoxanthin).

2.2. Hyperspectral $R_{rs}(\lambda)$ dataset construction and quality control

Model development and validation requires concurrent samples of HPLC phytoplankton pigments and hyperspectral $R_{rs}(\lambda)$ spectra. Hyperspectral $R_{rs}(\lambda)$ spectra were considered concurrent with HPLC samples if measurements were made within ±2 h at the same geographic location. Of the 4550 quality-controlled surface ocean HPLC samples, 178 samples had concurrent observations of hyperspectral $R_{rs}(\lambda)$ spectra, including spectra from eight oceanographic field campaigns (Table 1).

Details of initial $R_{rs}(\lambda)$ data processing can be found in: Chase et al., 2017 (Tara Oceans, Tara Mediterranean, SABOR, RemSensPOC, NAAMES, EXPORTS); Uitz et al., 2015 (BIOSOPE); and Bracher et al., 2015a (ANT). All spectra were interpolated to 1 nm resolution and smoothed using a 5 nm moving mean bandpass filter before subsequent analyses. Following this smoothing procedure, the first and last 4 nm of all spectra were removed. As some field campaigns measured a wider spectral range than others, the range of $R_{rs}(\lambda)$ in the final dataset was then restricted to 400–700 nm to match the range common to all campaigns.

Following this consistent smoothing approach, each individual $R_{rs}(\lambda)$ spectrum was visually inspected for quality control. Some $R_{rs}(\lambda)$ spectra in the original datasets exhibited extremely high noise-to-signal ratios in the ~610–660 nm range, where relatively low variance was expected. For these spectra, multiple large (e.g., a factor of 2- to 5-fold larger than the mean value) departures from the mean $R_{rs}(\lambda)$ value over this spectral range were observed, and thus these spectra were removed from this analysis (Table 1). The number of spectra used in each dataset are indicated in Table 1, and the number of spectra removed from each dataset is indicated in parentheses; ultimately, 33 of the 178 samples were removed following this quality control approach (~19% of the initial dataset), resulting in 145 valid matchup samples between HPLC and quality-controlled, hyperspectral $R_{rs}(\lambda)$.

The matched HPLC and $R_{rs}(\lambda)$ dataset is composed mostly of open ocean samples from the Atlantic, Pacific, and Indian Oceans as well as the Mediterranean Sea (Table 1). The dataset encompasses a broad range of chlorophyll-*a* concentrations, from 0.019–4.15 mg m^{-3} (Fig. 1; Table 1); however, the median chlorophyll-*a* concentration is relatively low (0.110 mg m^{-3}).

2.3. Hyperspectral reflectance model construction

A generic hyperspectral reflectance model was developed with the goal of enhancing the spectrally narrow phytoplankton signals associated with phytoplankton pigment variability. The generic formulation of the hyperspectral reflectance model is based on the quadratic relationship between reflectance measured just below the surface ($r_{rs}(0^-,\lambda)$), absorption (a), and backscattering (b_b), developed from radiative transfer theory by Gordon et al. (1988):

$$r_{rs}(0^-,\lambda) = \sum_{i=1}^{2} g_i \left(\frac{b_{bw}(\lambda) + b_{bp}(\lambda)}{a_w(\lambda) + a_{ph}(\lambda) + a_{dg}(\lambda) + b_{bw}(\lambda) + b_{bp}(\lambda)} \right)^i \quad (1)$$

where $r_{rs}(0^-,\lambda)$ is related to remote sensing reflectance measured just above the surface ($R_{rs}(0^+,\lambda)$) following Lee et al. (2002):

$$r_{rs}(0^-,\lambda) = R_{rs}(0^+,\lambda)/[0.52 + 1.7{*}R_{rs}(0^+,\lambda) \quad (2)$$

In Eq. (1), the g_i coefficients are the same as those used in the original Gordon et al. (1988) model. The components of backscattering and absorption are parameterized as follows. Backscattering by seawater, $b_{bw}(\lambda)$, is computed as in Zhang et al. (2009) using temperature and salinity values from the NOAA NODC World Ocean Atlas ¼° resolution statistical mean climatology (Locarnini et al., 2013; Zweng et al., 2013). Pure water absorption, $a_w(\lambda)$, is taken from Mason et al. (2016). Phytoplankton absorption, $a_{ph}(\lambda)$, is expressed as a power law function of Tchla:

$$a_{ph}(\lambda) = A(\lambda){*}Tchla^{B(\lambda)} \quad (3)$$

The $A(\lambda)$ and $B(\lambda)$ coefficients were derived from regressions performed at each wavelength using a large, global, multispectral (18 wavelengths) dataset extracted from the NASA SeaBASS bio-optical data repository (NOMAD; Werdell and Bailey, 2005) interpolated to 1 nm resolution between 350 and 700 nm using cubic spline interpolation. The $A(\lambda)$ and $B(\lambda)$ coefficients used here are shown between 400 and 700 nm in Table S6. The NOMAD data used to determine the $a_{ph}(\lambda)$ parameterization are independent from the paired $R_{rs}(\lambda)$-HPLC dataset constructed here.

The combined absorption of non-algal particles and dissolved matter, $a_{dg}(\lambda)$, is expressed as:

$$a_{dg}(\lambda) = a_{dg}(443){*}exp\left(S_{dg}(\lambda - 443)\right) \quad (4)$$

where the slope in the exponential term, S_{dg}, is a linear function of the $R_{rs}(490)/R_{rs}(555)$ ratio (as in Carder et al., 1999):

$$S_{dg} = -0.01447 + 0.00033{*}R_{rs}(490)/R_{rs}(555) \quad (5)$$

This relationship was also obtained from a large dataset of reflectance and $a_{dg}(\lambda)$ data from SeaBASS (Werdell and Bailey, 2005).

Finally, particulate backscattering is expressed as:

$$b_{bp}(\lambda) = b_{bp}(443){*}(\lambda/443)^{\eta} \quad (6)$$

where the exponent, η, is a function of the below-surface $r_{rs}(490)/r_{rs}(555)$ ratio, following Lee et al. (2002).

The hyperspectral $R_{rs}(\lambda)$ model first solves for three parameters in reconstructing the measured spectra: chlorophyll-*a* concentration (*Tchla*), non-algal absorption excluding water at 443 nm (combined CDOM and NAP absorption, $a_{dg}(443)$), and particulate backscattering at 443 nm ($b_{bp}(443)$) through a non-linear fit between measured and modeled reflectance, as in Maritorena et al. (2002). In that process, full spectra for $b_{bp}(\lambda)$, $a_{dg}(\lambda)$, and ultimately $R_{rs}(\lambda)$ are reconstructed using the expressions described above (Eqs. (1)-(6)).

The resulting modeled R_{rs} spectra ($R_{rs,\ mod}(\lambda)$; Fig. 2B) were subtracted from the measured R_{rs} spectra ($R_{rs,\ meas}(\lambda)$; Fig. 2A) to create the $R_{rs}(\lambda)$ residual: $\delta R_{rs}(\lambda)$ (Fig. 2C). The second derivative of the $R_{rs}(\lambda)$ residual, δR_{rs} " (λ), was used in subsequent analyses to maximize the

Table 1
Summary table for the eight field campaigns represented in the matched HPLC and $R_{rs}(\lambda)$ dataset. All data are cited in Kramer et al. (2021); campaign-specific citations: [1]Bracher et al. (2015b), [2]Behrenfeld et al. (2014a), [3]Cetinić (2013), [4]Behrenfeld et al. (2014b), [5]Boss and Claustre (2009), [6]Boss and Claustre (2014), [7]Claustre and Sciandra (2004) and Casey et al. (2019), [8]Behrenfeld et al. (2018).

Cruise name	# samples (# removed)	Geographic region	Chl range (mg m^{-3})	Median chl (mg m^{-3})	Mean chl (mg m^{-3})
ANT[1]	26 (28)	Atlantic	0.033–4.15	0.232	0.648
NAAMES[2]	11 (1)	Northwest Atlantic	0.094–0.987	0.496	0.540
RemSensPOC[3]	27	Northwest Atlantic & equatorial Pacific	0.049–1.09	0.090	0.173
SABOR[4]	9	Northwest Atlantic	0.070–1.31	0.252	0.471
Tara Oceans[5]	16 (3)	Global	0.021–0.950	0.168	0.194
Tara Med[6]	29	Mediterranean Sea	0.026–0.170	0.055	0.064
BIOSOPE[7]	23 (1)	Southeast Pacific	0.019–1.47	0.069	0.326
EXPORTS[8]	4	Northeast Pacific	0.172–0.292	0.224	0.228

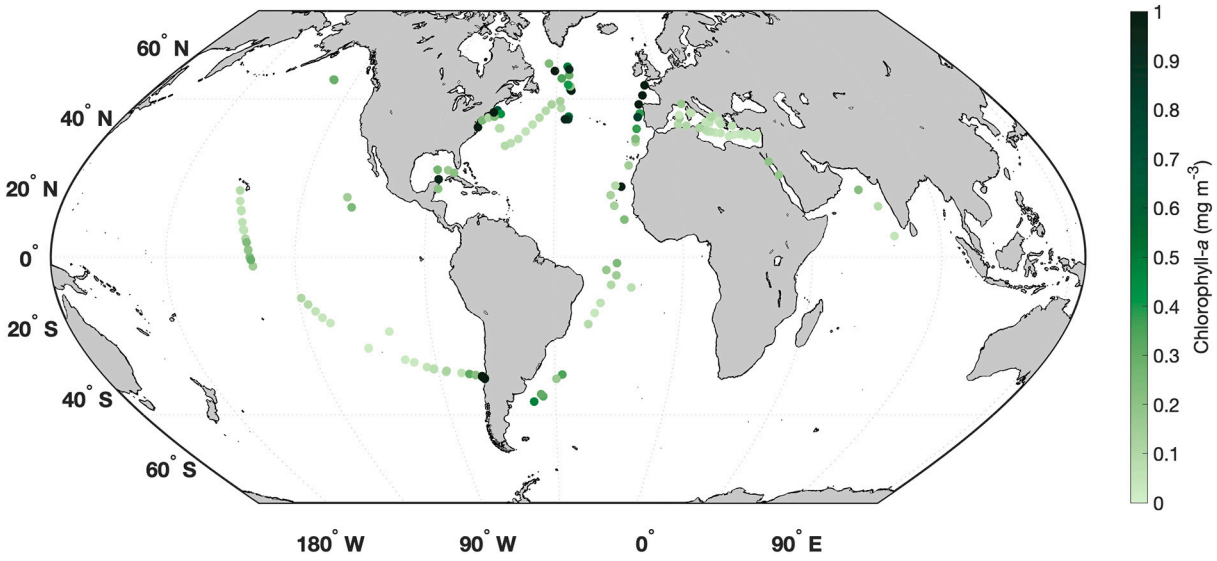

Fig. 1. Global distribution of 145 matched HPLC and hyperspectral $R_{rs}(\lambda)$ samples, colored by chlorophyll-*a* concentration (Tchla).

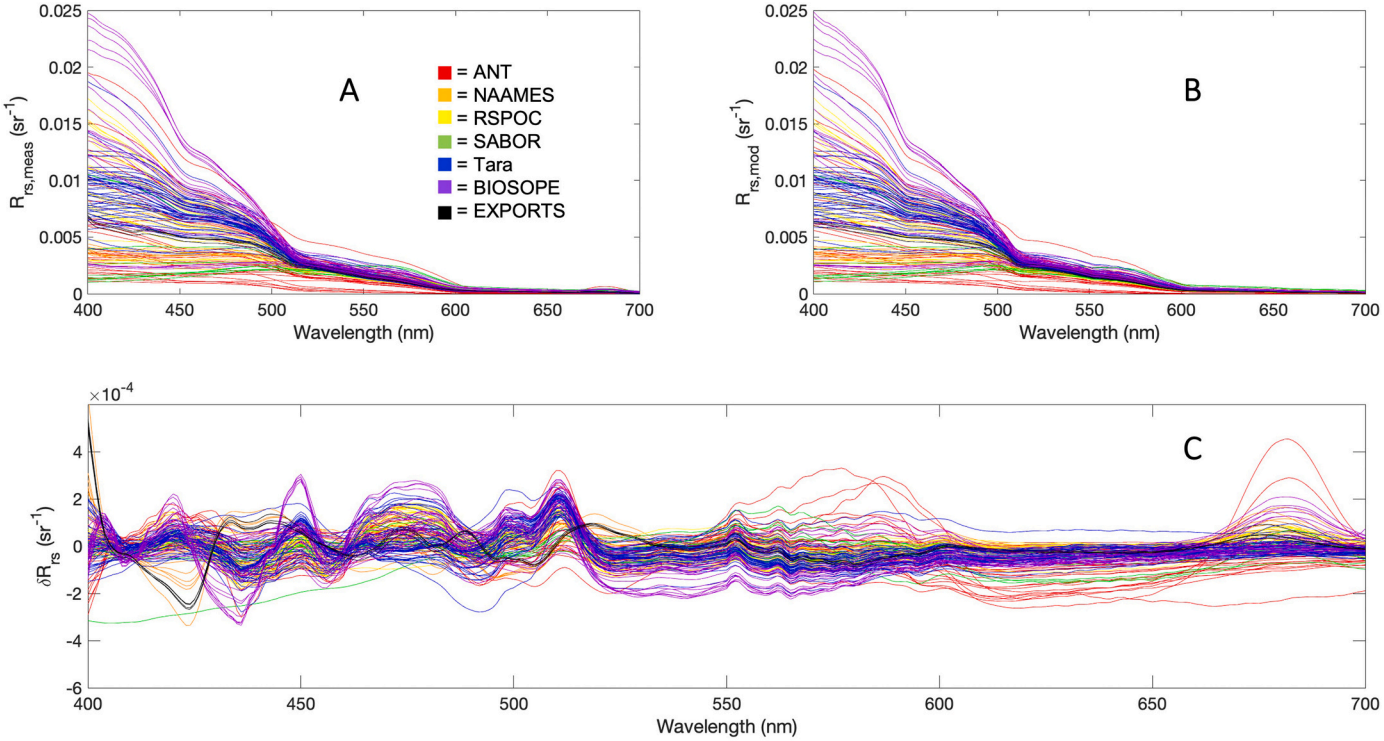

Fig. 2. (A) Measured ($R_{rs,\ meas}(\lambda)$) and (B) modeled ($R_{rs,\ mod}(\lambda)$) hyperspectral $R_{rs}(\lambda)$ spectra and (C) the residual spectrum ($\delta R_{rs}(\lambda)$) between measured and modeled $R_{rs}(\lambda)$. All spectra are colored by source (red = ANT, orange = NAAMES, yellow = RemSensPOC [RSPOC], green = SABOR, blue = Tara, purple = BIOSOPE, black = EXPORTS). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

narrow spectral features most related to phytoplankton absorption and scattering. As in Catlett and Siegel (2018), δR_{rs} " (λ) spectra were calculated using a second-order finite difference approximation.

2.4. Hierarchical clustering and empirical orthogonal function (EOF) analysis of HPLC data

Hierarchical cluster analysis of thirteen HPLC phytoplankton accessory pigment ratios to Tchla was performed following Catlett and Siegel (2018) and Kramer and Siegel (2019), using Ward's linkage method (the inner squared distance) and the correlation distance (1-R, where R is Pearson's correlation coefficient between phytoplankton pigment ratios). The dendrogram for all pigment ratios was then divided into distinct taxonomic clusters using a linkage cutoff distance of 0.65. The same linkage and distance methods were used to cluster the modeled pigments. The taxonomic utility of groups of phytoplankton pigments was assumed following Catlett and Siegel (2018) and Kramer and Siegel (2019).

Empirical orthogonal function (EOF) analysis was also performed following Kramer and Siegel (2019) and Kramer et al. (2020). Briefly, this analysis aims to decompose the data into the dominant orthogonal functions that describe the major modes of variability in the dataset.

Here, the EOF loadings, which describe the correlation between each mode of variability and ratios of phytoplankton pigments to Tchla, are considered. Phytoplankton pigment concentrations were normalized to Tchla concentration, then mean-centered and normalized by their standard deviation before the EOF analysis was performed. The same approach was repeated for the modeled pigment dataset.

2.5. Principal components regression model

A number of statistical methods were considered to model pigments from $R_{rs}(\lambda)$, including hierarchical cluster analysis of spectra (as in Torrecilla et al., 2011; Uitz et al., 2015) and network-based community detection approaches (as in Kramer et al., 2020). Ultimately, following the approach of Catlett and Siegel (2018), a principal components regression model was constructed. Here, the model used the second derivative of the $R_{rs}(\lambda)$ residual (δR_{rs} " (λ)). Principal components regression modeling was selected as this method accounts for the high degree of collinearity across phytoplankton bio-optical signatures that arises due to the co-variability among phytoplankton groups and accessory pigments (e.g., Massy, 1965; Catlett and Siegel, 2018). This

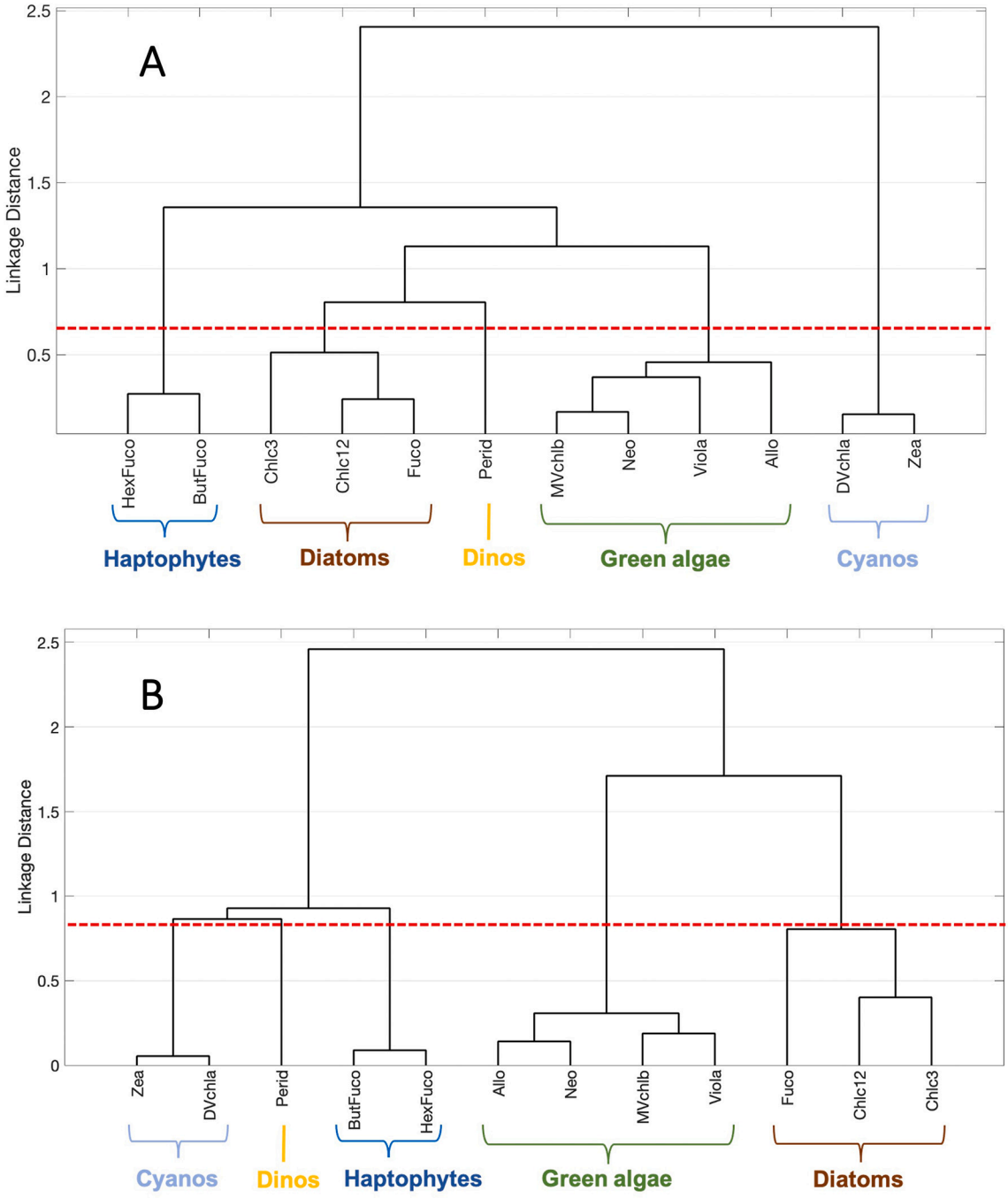

Fig. 3. Hierarchical cluster analysis of thirteen pigment ratios to Tchla. (A) Results for measured HPLC pigments: using a linkage distance of 0.65 (red dashed line), five distinct groups emerge and are annotated here with their assumed taxonomic representation: haptophytes (dark blue), diatoms (brown), dinoflagellates (gold), green algae (green), and cyanobacteria (light blue). (B) Results for principal components regression modeled pigments from δR_{rs} " (λ): using a linkage distance of 0.80 (red dashed line), the same five pigment groups identified in (A) emerge. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

approach reduced the inter-relatedness of the datasets (that is to say, the high correlations between pigment concentrations and δR_{rs} " (λ)) prior to modeling. Many other principal components regression models were tested, including models reliant on both the first and second derivatives of the measured hyperspectral reflectance ($R_{rs,\ meas}$ ′ (λ) and $R_{rs,\ meas}$ " (λ)) and models that varied the spectral resolution of the input data (see Supporting Information for details regarding these model constructions and results). The performance of the best of these models was similar, and thus we chose to highlight the results of the model constructed using δR_{rs} " (λ) at 1 nm, which had excellent performance and one spectral input.

Optimized principal component regression coefficients were determined following Catlett and Siegel (2018) and transformed into spectral coefficients for δR_{rs} " (λ). Pigment concentrations were modeled as:

$$\widehat{p}_m = \sum_{i=1}^{N} A_m(\lambda_i) * \delta R_{rs}''(\lambda_i) + C_m \quad (7)$$

where $A_m(\lambda_i)$ is the wavelength-specific coefficient applied to δR_{rs} " (λ_i) at the ith wavelength (λ) for a given pigment concentration ($\widehat{p}_m$), and C_m is an intercept. Resulting pigment values were constrained to be positive values (or zero) before computing goodness-of-fit statistics.

We employed the cross-validation-based model optimization and validation procedures described in Catlett and Siegel (2018), with some adjustments. The modeling approach was validated using a 100-fold cross-validation procedure for each pigment. 75% of the dataset was used for model training, while 25% of the dataset was used for model performance evaluation. Principal components are computed from standardized (z-scored; mean-centered and divided by the variance) δR_{rs} " (λ) spectra included in the training set. Principal components regression models are then optimized based on the training set by minimizing the mean absolute difference (*MAD*) following Seegers et al. (2018) and McKinna et al. (2021):

$$MAD = \frac{1}{N} \sum_{i=1}^{N} \left| \widehat{p}_{m,i} - p_{m,i} \right| \quad (8)$$

where N is the number of samples in the model training dataset (25% of 145, or 36 samples), $p_{m,\ i}$ is the measured HPLC pigment concentration, and $\widehat{p}_{m,i}$ is the corresponding modeled pigment concentration, each for the ith observation. This approach differs from Catlett and Siegel (2018) where models were optimized by maximizing Pearson's squared correlation coefficient (R^2).

Pigment concentrations were reconstructed for the entire dataset (see Figs. 3 and 6 below). For this exercise, the 100 quasi-independent sets of optimized coefficients ($A_m(\lambda_i)$ and C_m) determined from the 100 cross-validations were applied to all δR_{rs} " (λ_i) spectra used here, following Eq. (8). The median pigment value of those 100 modeled values was used in further analyses. Any modeled pigment values that were below the standard HPLC pigment detection limits (Van Heukelem and Thomas, 2001) were again set to zero before subsequent analyses. It should be noted that the goodness-of-fit statistics are expected to improve in this exercise relative to those determined from the 100-fold cross-validation procedure employed above since the training and validation datasets are not independent in this step.

3. Results

3.1. HPLC pigments

The relationships between and among phytoplankton pigment ratios to Tchla in the measured HPLC pigment dataset constrain the number of distinct groups that can be identified from any subsequent modeling using the $R_{rs}(\lambda)$ data (Kramer and Siegel, 2019; Kramer et al., 2020). In this HPLC dataset, hierarchical cluster analysis separates five distinct phytoplankton pigment groups (Fig. 3A), each of which can be distinguished by one biomarker pigment (with assumed taxonomic representation): Fuco (diatoms), Perid (dinoflagellates), HexFuco (haptophytes), MVchlb (green algae), and Zea (cyanobacteria).

The connections between and among the phytoplankton pigment groups that emerge here are very similar to those identified in the global analysis by Kramer and Siegel (2019); conclusions drawn there would be applicable to this subset of their data. The groups identified here also broadly separate along (widely-assumed) phytoplankton size class lines, with diatoms and dinoflagellates mostly comprising the micro- and nano-sized phytoplankton groups, while haptophytes, green algae, and cyanobacteria mostly comprise the nano- to pico-sized groups. The same phytoplankton pigment groups emerged from the EOF analysis (Fig. S1A—D), with the first mode separating cyanobacterial pigments from all other groups, the second mode separating haptophyte pigments from green algal pigments, the third mode separating diatom pigments from all other groups, and the fourth mode separating dinoflagellate and cyanobacteria pigments from all other groups.

3.2. Hyperspectral reflectance spectra

The hyperspectral reflectance modeling used here aims to reproduce the spectral shape and magnitude of the $R_{rs,\ meas}(\lambda)$ data (Fig. 2A) using a generic, data- and literature-based parameterization of the model components. The $R_{rs,\ mod}(\lambda)$ data (Fig. 2B) match the range of spectral shapes and magnitudes of the $R_{rs,\ meas}(\lambda)$ data quite well. The broadly similar patterns in spectral shape and relatively low magnitude of the residual spectra ($\delta R_{rs}(\lambda)$) show that most of the differences between the measured and modeled $R_{rs}(\lambda)$ are in the blue and red wavelengths (Fig. 2C), where phytoplankton accessory pigment absorption is highest and most variable in shape, and in the red, where chlorophyll fluorescence is active. The $\delta R_{rs}(\lambda)$ spectra are relatively flat in the ~520–550 and ~ 600–660 regions. The similarity in the shapes of the $\delta R_{rs}(\lambda)$ spectra qualitatively validates the approach taken here, to remove much of the signal from $R_{rs}(\lambda)$ that varies on broader spectral scales (e.g., $a_{NAP}(\lambda)$, $a_{CDOM}(\lambda)$, $b_{bp}(\lambda)$) and preserve the signal that varies on narrower spectral scales (e.g., due to PCC differences).

The performance of the hyperspectral reflectance model was further evaluated by comparing the model retrieval of Tchla with measured HPLC Tchla (Fig. 4).

Measured Tchla was compared to both Tchla derived from the OC4v6 chlorophyll *a*lgorithm (Fig. 4A; O'Reilly et al., 1998) and from the hyperspectral reflectance model used here (Fig. 4B). While both models produce Tchla concentrations that are well correlated with the measured HPLC Tchla ($R^2 = 0.75$ and 0.86, respectively), the performance of the hyperspectral reflectance model improves upon the OC4v6 algorithm performance both in terms of the model fit to the measured data and its adherence to the 1:1 line (slope = 0.96 vs. slope = 0.87). This result is consistent with previous findings showing that multispectral Tchla models perform better if the effects of $b_{bp}(\lambda)$ and $a_{dg}(\lambda)$ are accounted for (i.e., Siegel et al., 2005, 2013).

3.3. Correlations between $\delta R_{rs}(\lambda)$ and HPLC pigments

In order to assess the nature of phytoplankton pigment signals contained in $\delta R_{rs}(\lambda)$ spectra, correlations were examined between the $\delta R_{rs}(\lambda)$ spectra and pigment concentrations (Fig. 5A&D diatom and cyanobacteria pigments; Fig. S2 all other pigments), the first derivative of $\delta R_{rs}(\lambda)$ and pigments (δR_{rs} ′ (λ), Fig. 5B&E; Fig. S3), and the second derivative of $\delta R_{rs}(\lambda)$ and pigments (δR_{rs} " (λ), Fig. 5C&F; Fig. S4).

Correlations were considered between $\delta R_{rs}(\lambda)$ and Tchla and between $\delta R_{rs}(\lambda)$ and each of the five groups of biomarker pigments that broadly describe the five major pigment groups based on the results of the hierarchical cluster analysis presented in Fig. 3A. For $\delta R_{rs}(\lambda)$, δR_{rs} ′ (λ), and δR_{rs} " (λ), high correlations ($|R| \geq 0.5$) were found between reflectance spectra and pigments across the range of wavelengths considered in this analysis.

Strongly positive or negative relationships were not restricted to

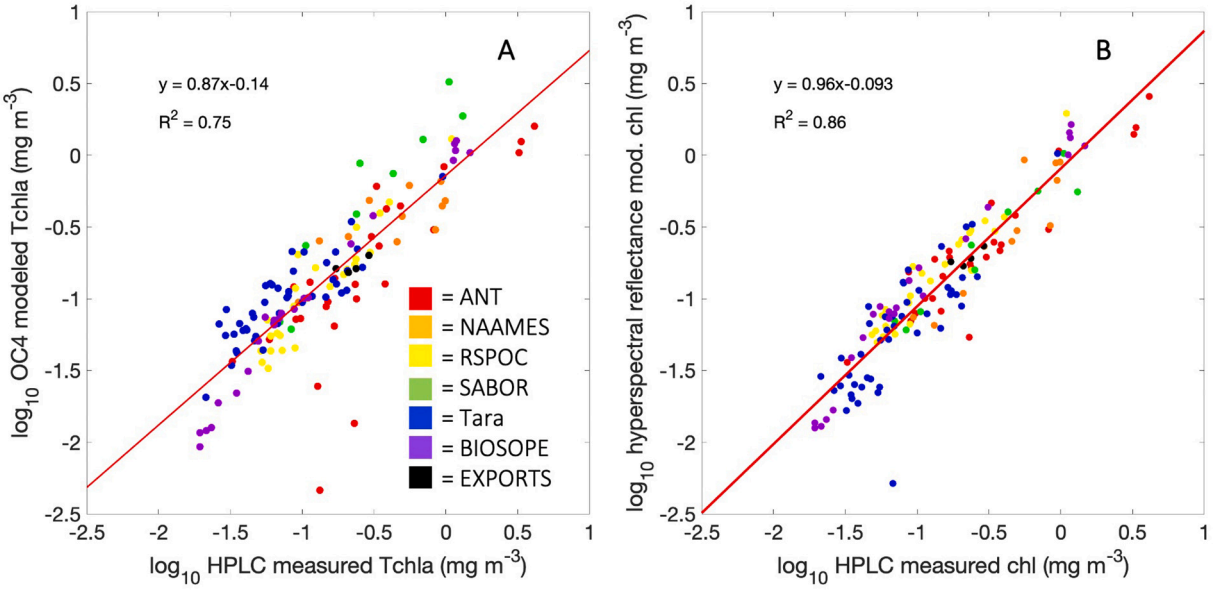

Fig. 4. Correlation between measured Tchla and Tchla modeled according to (A) the OC4 chlorophyll *a*lgorithm and (B) the hyperspectral GSM-like model used here. Samples are colored by source (red = ANT, orange = NAAMES, yellow = RemSensPOC [RSPOC], green = SABOR, blue = Tara, purple = BIOSOPE, black = EXPORTS). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

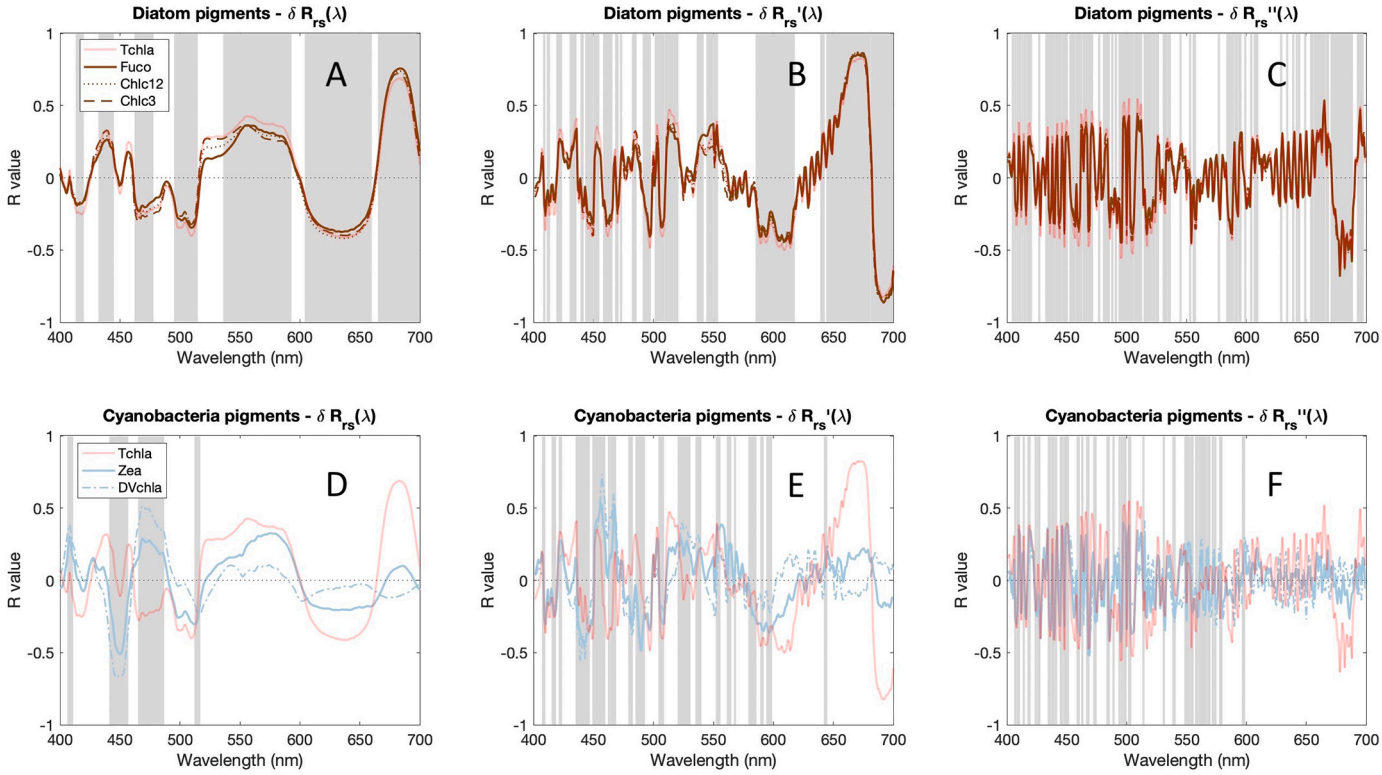

Fig. 5. Pearson's correlation coefficients (R) between (A & D) $\delta R_{rs}(\lambda)$ spectra and pigments, (B & E) $\delta R_{rs}{}'(\lambda)$ spectra and pigments, (C & F) $\delta R_{rs}{}''(\lambda)$ spectra and pigments, grouped based on the results of hierarchical cluster analysis (Fig. 3): (A, B, C) diatom pigments and (D, E, F) cyanobacterial pigments. Grey bars indicate wavelengths at which the correlation coefficients for all pigments are significantly different from zero. The correlation with Tchla (in red) is included on each panel for comparison. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

wavelengths where $\delta R_{rs}(\lambda)$ was necessarily more positive or negative (e.g., blue and red wavelengths; Fig. 2C); rather, nearly all pigments were significantly correlated with $\delta R_{rs}(\lambda)$ and its first and second derivatives across the visible spectrum (Figs. 5A&D, S2). Generally, correlations were high in the blue, through the green, and into the red part of the spectrum for most pigment groups (excluding cyanobacterial pigments). Some of the strongest correlations (both positive and negative) between $\delta R_{rs}(\lambda)$ (or its derivatives) and pigments were in the red, where chlorophyll both absorbs and fluoresces, which has an impact on the spectral shape and magnitude of both measured and modeled $R_{rs}(\lambda)$. The

correlation spectra for some pigment groups (for instance, diatom pigments; Fig. 5A-C) were almost identical to that of Tchla; however, there were differences in the ranges of wavelengths for which these pigments are most strongly correlated with $\delta R_{rs}(\lambda)$, indicated by the regions in which pigment correlations are significantly different from zero. Other pigment groups (such as cyanobacterial pigments; Fig. 5D-F) have correlation spectra that vary in spectral shape and magnitude from that of Tchla, often presenting an inverse correlation to that of Tchla. Ultimately, the strong correlations between most pigments and $\delta R_{rs}(\lambda)$ (and its derivative spectra) across nearly all wavelengths suggested that hyperspectral reflectance residuals are well suited to pigment modeling using all measured wavelengths.

3.4. Modeling phytoplankton pigments from hyperspectral $\delta R_{rs}''(\lambda)$

The concentrations of all thirteen phytoplankton pigments considered here were estimated from the $\delta R_{rs}''(\lambda)$ principal components regression modeling approach with relatively high accuracy and low error (Table 2; Fig. 6; $R^2 \geq 0.5$ for all pigments except Zea and the green algal pigments). Given the large differences in concentration of Tchla and each accessory pigment, the *MAD* presented in Table 2 was normalized to the average retrieved pigment concentration for each pigment to facilitate comparison of the model performance between pigments.

The mean model summary statistics from the 100-fold cross-validation exercise (Table 2) provide estimates of the central tendency of the model performance when extrapolated to novel observations (e.g., the randomly selected 25% of the dataset used for testing model performance for each cross-validation). The normalized mean absolute difference (MAD) was lowest for Tchla and red algal and cyanobacterial pigments and higher for green algal pigments. The relationships between measured and modeled pigments were quite strong when the entire pigment dataset was reconstructed from median modeled values across the 100 cross-validations (Fig. 6): the slopes of the relationship between measured and modeled pigments for Tchla and five of the major biomarker pigments (excluding Zea) are close to 1 (0.74–0.94), while the R^2 values for these linear fits are also high (0.51–0.73). There were no clear relationships between the data source (e.g., the individual field campaign) and the pigment reconstruction (Fig. 6). Specifically, the relationships between and among pigments were conserved through this modeling exercise and the same five pigment clusters found in the measured pigment dataset (Fig. 3A) are also identified from hierarchical cluster analysis of the modeled pigment dataset (Fig. 3B).

Five phytoplankton pigment groups can generally be distinguished by the co-variability between the ratios of five biomarker pigments to Tchla (Fig. 3). These same five pigment groups emerged from analyses of both the measured and modeled pigment analyses. The modeled pigments showed reasonably good correspondence with the measured pigments for most biomarker pigments (Table 2), particularly for Fuco ($R^2 = 0.65$). The order of some of the branches of the dendrogram shifted between the measured (Fig. 3A) and modeled (Fig. 3B) pigment datasets. Most notably, the modeled Perid clustered more closely with the modeled (assumed) cyanobacterial pigments, while measured Perid clustered more closely with measured (assumed) diatom pigments. However, the broad pigment groups remained the same between these analyses at high (>0.5) linkage distance thresholds, and the five groups of covarying pigments remain consistent. Similarly, the same major pigment-based taxonomic groups separated from the EOF analysis, but with different groups dominating different modes between the measured (Fig. S1A—D) and modeled (Fig. S1 *E*-H) datasets. The first mode separated green algal pigments from all other groups, the second mode separated haptophyte pigments from dinoflagellate pigments, the third mode separated diatom pigments from all other groups, and the fourth mode separated cyanobacterial pigments from all other groups.

Even the accessory and biomarker pigments with relatively poor model performance were reconstructed accurately enough that the patterns of covariation among those pigment ratios to Tchla, and between those pigment ratios and pigment ratios modeled with higher skill, were consistently recovered (Figs. 3, S1). For instance, Zea was retrieved with lower accuracy than many other pigments (Table 2; $R^2 = 0.37$); however, the strong covariation between Zea and DVchla meant that these reconstructed pigments still clustered closely together and away from all other pigments (Figs. 3B, S1H). Similarly, many of the green algal pigments were not as accurately modeled as many other pigments (Table 2; MVchlb $R^2 = 0.42$, Neo $R^2 = 0.42$, Viola $R^2 = 0.38$), but these pigments covary with each other and with Allo ($R^2 = 0.40$) and thus still clustered together as a distinct pigment group (Figs. 3B, S1E).

Table 2
Average summary statistics (R^2 and normalized MAD) and standard deviations of summary statistics across 100 model cross-validations for all modeled pigments. MAD and its standard deviation are normalized to the mean retrieved pigment concentration for each pigment. All statistics were assessed on a linear scale.

Pigment	Mean R^2	SD R^2	Mean normalized MAD	SD normalized MAD
Allo	0.40	0.19	1.221	0.400
But	0.62	0.16	0.588	0.185
Chlc3	0.68	0.13	0.639	0.212
Chlc12	0.70	0.13	0.703	0.235
DVchla	0.55	0.12	0.594	0.103
Fuco	0.65	0.15	0.844	0.274
Hex	0.54	0.16	0.692	0.201
MVchlb	0.42	0.19	0.975	0.295
Neo	0.42	0.21	1.127	0.354
Perid	0.49	0.13	0.783	0.166
Tchla	0.72	0.15	0.498	0.127
Viola	0.38	0.18	1.101	0.370
Zea	0.37	0.10	0.472	0.071

4. Discussion

The goal of this analysis was to model phytoplankton pigment concentrations from hyperspectral optics and use those modeled pigments to reconstruct relationships between and among groups of pigments that describe open ocean phytoplankton pigment composition. To achieve this goal, principal components regression was employed to model pigment concentrations from the second derivative of the residual spectra between measured and modeled hyperspectral remote sensing reflectance ($\delta R_{rs}''(\lambda)$). From a hierarchical cluster analysis of the measured HPLC pigment data, five distinct phytoplankton pigment groups were identified (diatoms, dinoflagellates, haptophytes, green algae, and cyanobacteria), constraining the number of groups that could be identified by the reflectance modeling approach to these same five (or fewer) groups. Ultimately, the principal components regression modeling approach reconstructed the measured pigment dataset, such that the same five pigment-based phytoplankton groups were identified again. The resulting modeled pigment dataset both reconstructs the patterns of covariability between and among phytoplankton pigments, and recreates the qualitative descriptions of five phytoplankton pigment groups determined from hierarchical cluster and EOF analyses. While the analyses presented here used the residual between the measured and modeled reflectance ($\delta R_{rs}''(\lambda)$), principal components regression modeling was repeated using the combined first and second derivatives of the measured hyperspectral reflectance ($R_{rs,\,meas}'(\lambda)$ and $R_{rs,\,meas}''(\lambda)$) with comparable results (Figs. S6–8, Table S1).

Here, we consider the strengths and limitations of the modeling approach and the results presented in this work. Since the derivative approach is sensitive to measurement noise in addition to variations in spectral shape, this analysis required the curation of a highly quality-controlled dataset. Data were limited for hyperspectral $R_{rs}(\lambda)$ matchups with HPLC pigments to 145 samples; more high quality data will improve this analysis and future analyses that use hyperspectral optics to model phytoplankton pigment concentrations. The results of

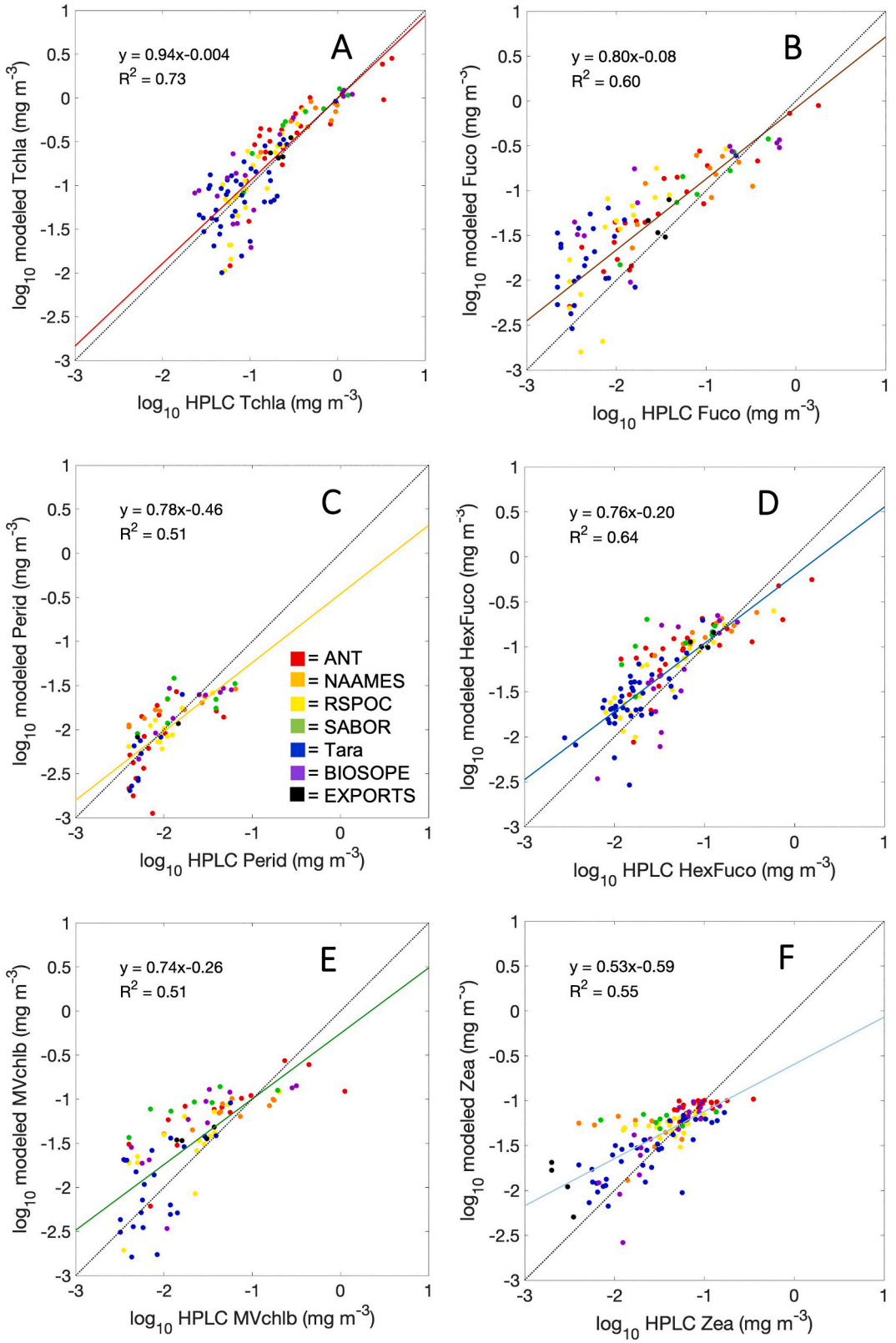

Fig. 6. Relationships between HPLC measured pigments and principal components regression modeled pigments using the median model result of all 100 cross-validations: (A) Tchla, (B) Fuco, (C) Perid, (D) HexFuco, (E) MVchlb, (F) Zea. The 1:1 line is shown in black; the linear fit is shown in red for Tchla, brown for Fuco, gold for Perid, dark blue for HexFuco, green for MVchlb, and light blue for Zea. Samples are colored by source (red = ANT, orange = NAAMES, yellow = RemSensPOC [RSPOC], green = SABOR, blue = Tara, purple = BIOSOPE, black = EXPORTS). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

the principal components regression models (or any bio-optical model) are constrained by the validation dataset used in the analysis. In this case, the taxonomic groups determined from the associations between and among HPLC phytoplankton pigments restricted the pigment groups that could be identified from optics to the five identified here. These five groups represent the extent to which phytoplankton pigment composition can be resolved within the global open ocean HPLC dataset assembled here. Finally, while this analysis aims to describe the central tendencies of the dataset used here, analyses that include different taxonomic or optical regimes than those included in this dataset (particularly inland or coastal waters) might need to combine approaches to fully describe the surface ocean phytoplankton pigment composition from optics. This approach describes a "base state" in the global surface ocean, while rare or more extreme departures from that base state will have divergent optical properties and will likely require more targeted approaches.

4.1. Quality controlling a global dataset from multiple sources

The robustness of any modeling approach is limited by the dataset used to construct and test that model. Here, data from eight field campaigns were combined, most of which had already been published in previous analyses (e.g., Uitz et al., 2015; Bracher et al., 2015a; Chase et al., 2017) or had been collected by those same groups using identical methods (e.g., the EXPORTS samples). The HPLC pigment dataset dictates the potential and limitations of the resulting optical model—here, the results were limited to five distinct pigment groups (Fig. 3A, Fig. S1A—D). The derivative analysis approach magnifies narrow spectral features, including measurement noise and error; thus, quality control of the $R_{rs}(\lambda)$ spectra was crucially important to ensure that the model results were influenced by real features rather than artifacts. Strict quality control will be particularly important for ocean color sensors such as PACE, particularly considering the potential effects of imperfect atmospheric corrections on reflectance data from these missions. It is likely that imperfections in atmospheric correction will occur on broader spectral scales (as is expected from the shapes of aerosol absorption and scattering; Werdell et al., 2019). Thus, the approach used here will negate many of these issues.

The quality control approach employed here aimed to remove any spectra with spurious features that would be amplified in the present approach; thus, some samples were removed from the datasets that were suitable for other analyses. Similarly, the wavelength range of the $R_{rs}(\lambda)$ spectra was selected to maximize overlap between different sampling approaches; all eight field campaigns measured reflectance between 400 and 700 nm, while some field campaigns had a larger range of measurements. There is undoubtedly useful phytoplankton community information in the UV and specific spectral features in the UV region have been shown to covary with specific biomarker pigments (e.g., Barrón et al., 2014; Kahru et al., 2021). Ideally, future $R_{rs}(\lambda)$ datasets will include high-quality measurements over a broader spectral range for full consideration of the impact of phytoplankton pigments on spectral data. Our results show that the model coefficients in this analysis vary across the visible spectrum (Fig. S5), not just in a narrow wavelength range. This result supports the importance of rigorous quality control for the spectral data used here; even small variations on short (5–10 nm) spectral scales are ultimately important in this pigment modeling approach. Similarly, the noise-to-signal ratio across the visible spectrum for in situ $R_{rs}(\lambda)$ data (as were used here) is much lower than for remotely sensed $R_{rs}(\lambda)$ data. Thus, spatiotemporal aggregation of remotely-sensed $R_{rs}(\lambda)$ will likely be required to improve and increase the signal-to-noise ratio to a level that can be tolerated by the approach presented here.

4.2. The need for more high-quality, paired global data

While the dataset used in this analysis was limited by the stringent quality control approach for both the HPLC pigment samples and $R_{rs}(\lambda)$ spectra, it was also limited by the available data that fit these requirements. There are abundant HPLC pigment samples with high data quality in the surface ocean (e.g., Kramer and Siegel, 2019). However, of the 4550 HPLC pigment samples in that analysis, only 145 had co-located, hyperspectral $R_{rs}(\lambda)$ spectra that passed the present quality control process. The distributions of both Tchla and the major accessory pigments varied in the 145 HPLC samples with corresponding $R_{rs}(\lambda)$ spectra, relative to the larger 4550 sample dataset analyzed previously (Fig. 7).

The mean pigment concentrations and ranges are significantly different for Tchla, Fuco, Perid, and HexFuco (two-sample *t*-test; $p < 0.01$). The mean values and range of the pigment concentrations in the global dataset were higher for Tchla and all accessory pigments except Zea compared to this dataset. The dataset used in this analysis was skewed more to samples with lower average Tchla concentrations that contained higher concentrations of Zea, but the difference in the mean Zea concentration between the two datasets was not significant (Fig. 7F). While the pigment-based statistical analyses from this dataset were comparable to the results of Kramer and Siegel (2019) in identifying nearly the same five groups of phytoplankton pigments (this analysis separated diatom pigments from dinoflagellate pigments; Fig. 3A), the bio-optical models that were constructed for this dataset fit a specific subset of the global dataset. Further model optimization may be required to apply this model accurately to all samples in that dataset, given the differences in dataset characteristics. However, despite the lower concentrations of most accessory pigments in this dataset, the model still reasonably reconstructed the concentrations of most accessory pigments.

There are many datasets that contain paired HPLC pigment samples and multispectral optics and/or radiometry (e.g., Werdell and Bailey, 2005). Similarly, some datasets include paired HPLC pigment samples (or other measurements of phytoplankton community composition) and hyperspectral optics (such as absorption by phytoplankton or other oceanic constituents), though few include hyperspectral reflectance as noted above (e.g., Valente et al., 2019; Casey et al., 2020). These datasets are also limited by their sampling locations—it is operationally more straightforward to collect both water samples and spectral measurements in inland and coastal waters than in the open ocean, so open ocean observations are more limited. The ratio of coastal to open ocean samples in most bio-optical datasets is not representative of the fraction of coastal to open ocean ecosystems on Earth (Mouw et al., 2017). The work presented here demonstrates conclusively the need for more and consistently collected, *paired* measurements of phytoplankton community composition (including, but not limited to, HPLC pigments) and hyperspectral $R_{rs}(\lambda)$ data (and, ideally, hyperspectral optical data) from diverse environments. Since all models, including the principal component regression model used here, are constrained by the quality and content of the datasets used to train and test those models, efforts to reconstruct phytoplankton community indices from hyperspectral reflectance can only be strengthened by the addition of more, high-quality open ocean hyperspectral optical and pigment data (e.g., Bracher et al., 2017).

4.3. The importance of spectral resolution

The quality and content of the model input data is also determined by the spectral resolution of that data. Hyperspectral data provide more degrees of freedom for modeling phytoplankton accessory pigments from $R_{rs}(\lambda)$ (Wolanin et al., 2016; Werdell et al., 2018; Cael et al., 2020). However, there are also high degrees of correlation between measurements made at similar wavelengths, which dilutes the statistical power of individual wavelengths (Cael et al., 2020). Thus, with these potential strengths and limitations in mind, this analysis was replicated for δR_{rs} " (λ) using 5 nm and 10 nm resolution rather than 1 nm resolution. The results demonstrate very little loss of qualitative or quantitative power

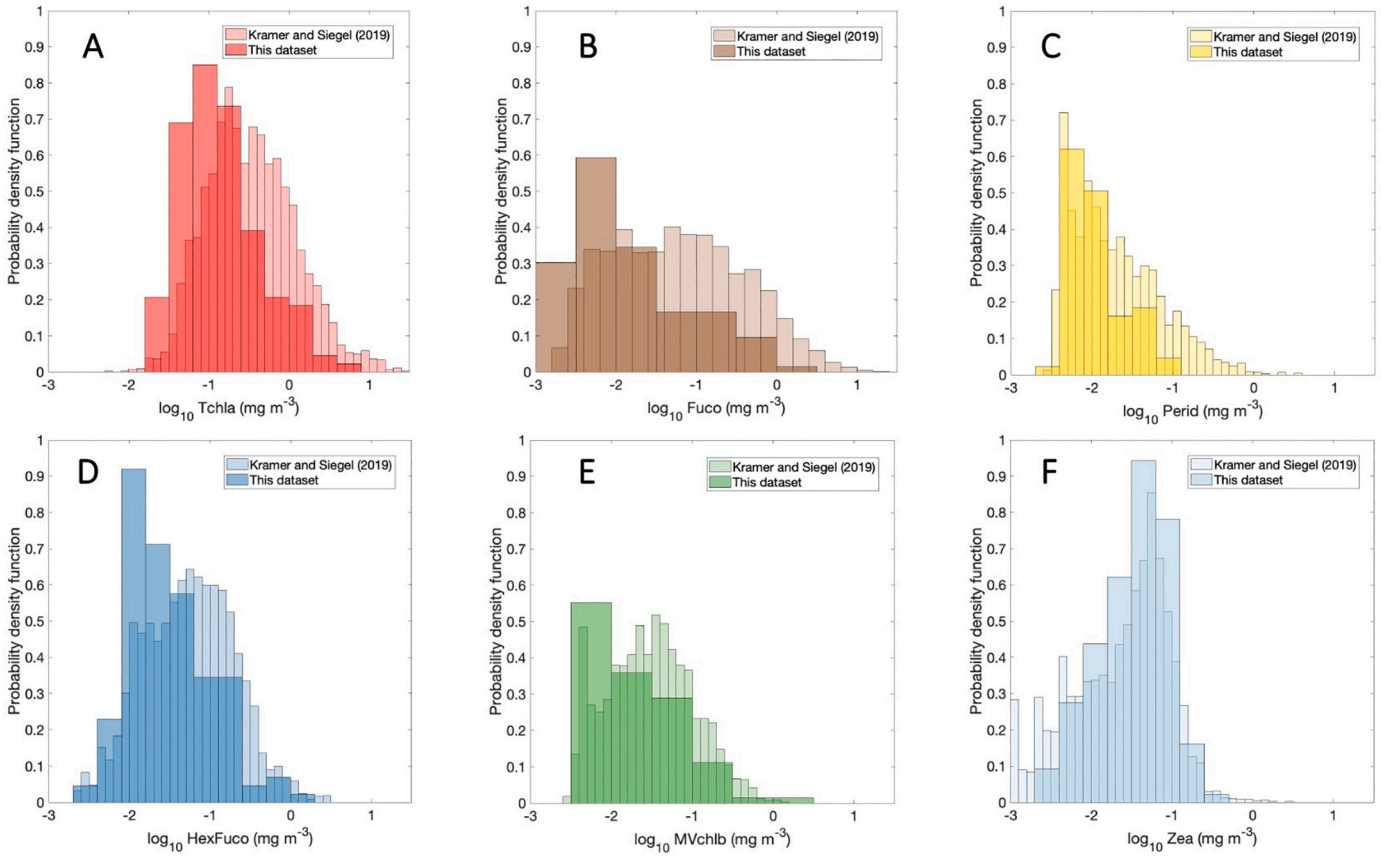

Fig. 7. Histograms of measured HPLC pigment concentrations from this analysis and from Kramer and Siegel (2019): (A) Tchla, (B) Fuco, (C) Perid, (D) HexFuco, (E) MVchlb, (F) Zea.

for pigment reconstruction between 1 nm and 5 nm resolution: the same 5 pigment groups separate (Figs. S9, S10), the relationships between measured and modeled pigments are comparably strong (Table S2, Fig. S11), and there is still predictive power across the visible spectrum that can be used for pigment modeling (Fig. S12). However, at 10 nm resolution, the results are notably worse for all modeled pigments (Table S3). This result is encouraging for existing and future ocean color remote sensing missions with high (~5 nm) spectral resolution (e.g., Werdell et al., 2019). These results can be replicated using both the first and second derivatives of the measured hyperspectral reflectance, $R_{rs,\ meas}{}'(\lambda)$ and $R_{rs,\ meas}{}''(\lambda)$ at varying spectral resolution in principal component regression models (Figs. S13–15; Tables S4, S5).

4.4. The potential of the spectral gap hypothesis

The present results highlight the benefit of removing $R_{rs}(\lambda)$ variability at broad spectral scales to accentuate those spectral variations that should be better associated with optical features caused by changes in phytoplankton pigment composition. Central to this approach is the hypothesis that phytoplankton optical signals can be useful for quantifying phytoplankton pigment composition by maximizing the variability in $\delta R_{rs}(\lambda)$ on narrow spectral scales (<100 nm) and reducing or removing the broad scale (>100 nm) signals that dominate the major optical properties in the ocean (e.g., CDOM, NAP). By removing broad-scale spectral signals, the $\delta R_{rs}(\lambda)$ spectra (and its derivatives) should accentuate the optical signals associated with the phytoplankton community. The major variations in the magnitude and shape of $\delta R_{rs}{}''(\lambda)$ were predominantly in the blue and red wavelengths (Fig. 2C), where phytoplankton accessory pigment absorption and fluorescence are the highest. However, the results of the principal components regression modeling approach demonstrate that relevant information for modeling pigments from the second derivative of $\delta R_{rs}(\lambda)$ is not just contained in the spectral regions where many phytoplankton pigments absorb, but across the whole visible spectrum (Figs. 5, S4). These results demonstrate the covariation among pigments and their absorption features, but also the co-variability of pigments with other phytoplankton pigment group-specific optical properties (e.g., fluorescence, scattering, packaging, etc.). The model coefficients also have power across the visible spectrum (Fig. S5), demonstrating the importance of using data from 400 to 700 nm in this modeling approach (see also discussion in Catlett and Siegel, 2018).

4.5. Further applications of $\delta R_{rs}(\lambda)$ for PACE

To a large extent, the residuals between measured and modeled $R_{rs}(\lambda)$, $\delta R_{rs}(\lambda)$, represent the differences in the relationship between Tchla and accessory pigments and their influence on phytoplankton absorption, $a_{ph}(\lambda)$, in the measured and modeled dataset (e.g., NOMAD; Werdell and Bailey, 2005) and might not accurately reflect the relationships between Tchla and accessory pigments (which influence the shape and magnitude of $a_{ph}(\lambda)$) in the present dataset. Thus, the residual reflectance spectrum, $\delta R_{rs}(\lambda)$, is a useful tool to quantify the shape differences of a given $R_{rs}(\lambda)$ spectrum—particularly when combined with a derivative analysis that accentuates the fine-scale features related to phytoplankton absorption and scattering. The usefulness of the reflectance residual approach in bio-optical oceanography has been established before (e.g., Roesler and Perry, 1995; Alvain et al., 2005), though it has not been applied for modeling phytoplankton pigment concentrations. This approach could be further applied to hyperspectral ocean color data to classify and cluster optical data and describe broad patterns in the global surface ocean (e.g., Siegel et al., 2005; Blondeau-Patissier et al., 2014). Through measurements and modeling of surface ocean

reflectance, the $\delta R_{rs}(\lambda)$ parameter could describe similarities and differences in the shapes of measured hyperspectral $R_{rs}(\lambda)$. Statistical analyses, such as EOFs or cluster analysis, could then partition these optical communities into broadly similar groups, and allow for a deeper investigation of the phytoplankton pigment composition underlying these similar optical regimes. Ultimately, this type of approach would aim to describe the central tendency in the dataset by classifying groups of spectra that were correlated with similar surface ocean patterns and ecosystems.

4.6. Combining principal components regression modeling with other remote sensing phytoplankton community composition algorithms

The model that was constructed here describes a statistical approach for predicting phytoplankton biomarker pigment concentrations from reflectance spectra. The modeling approach implemented with this dataset is empirical, and thus it was only able to reconstruct the phytoplankton pigment communities represented in *this* dataset. While many remote sensing algorithms have similarly been constructed to retrieve various optical parameters (e.g., Maritorena et al., 2002; Werdell and Bailey, 2005; Uitz et al., 2015; Chase et al., 2017; etc.), some remote sensing algorithms for detecting phytoplankton community composition aim to identify the cases that deviate from standard oceanic conditions. In those models, the aim is to identify *the* phytoplankton group that dominates the optical signal in a given ecosystem, often in the case of a monospecific phytoplankton bloom. This information is likely not retrievable using empirical techniques aimed at quantifying the central tendencies in a dataset. Approaches exist to quantify or identify blooms of coccolithophores (Brown and Yoder, 1994; Sadeghi et al., 2012) or *Trichodesmium* spp. (Westberry et al., 2005; Westberry and Siegel, 2006) on global scales, as well as *Phaeocystis* spp. (Lubac et al., 2008), harmful algal blooms (i.e., *Karenia brevis*, Stumpf et al., 2003; *Pseudo-nitzschia* spp., Smith and Bernard, 2020; etc.), and diatoms (Sathyendranath et al., 2004; Soppa et al., 2014; Kramer et al., 2018) on local scales. It is important to note that the approach developed here is not comparable to these methods, as it does not attempt to identify the dominant phytoplankton group within a community, but rather reconstructs individual phytoplankton pigment concentrations from $R_{rs}(\lambda)$ and $\delta R_{rs}(\lambda)$. Reconstructed pigment compositions and concentrations can then be used to estimate phytoplankton community composition. In other ecosystems or regions, the combinations of reconstructed pigments might cluster differently to form distinct phytoplankton pigment groups from the ones identified here (e.g., Kramer and Siegel, 2019). By aiming to describe variability in suites or communities of biomarker pigment concentrations, the principal components regression modeling approach used here describes a central tendency in the dataset, and is complimentary to ocean color algorithms that attempt to identify outliers dominated by a single phytoplankton type.

Combining the method presented here with one of the above more targeted methods may provide insight into how well the reconstructed pigment suites match the distinct optical signals associated with a given phytoplankton group. For example, in an ecosystem where a "coccolithophore bloom" (Brown and Yoder, 1994) can be identified from remote sensing, would the reconstructed pigment modeling also retrieve high concentrations of HexFuco and Chlc3? In this case, the principal components regression modeling approach could serve to describe a community in which the optics were more useful for describing phytoplankton community composition than the pigments. These combined approaches could also give insights into bloom succession, and the strengths or weaknesses of individual models as the optical properties of a bloom change. Alternately, the Westberry et al. (2005) approach can identify a *Trichodesmium* bloom from ocean color based on optical anomalies above a defined threshold value. Using pigment data, *Trichodesmium* could be distinguished by the cyanobacterial biomarker pigments considered here (Zea, DVchla), but also by phycobilins, which are not measured by traditional HPLC methods, but can be modeled by similar approaches to those employed here (Taylor et al., 2013). Again, the optics may provide more information than the pigment-based taxonomy, and thus the methods would be stronger when combined.

5. Conclusions

This analysis demonstrates the potential and limitations of hyperspectral remote sensing reflectance data for reconstructing phytoplankton pigment composition. Five pigment groups were separated from the validation dataset of HPLC pigments and are assumed to represent diatoms, dinoflagellates, haptophytes, green algae, and cyanobacteria. Thirteen pigments were then modeled from a matched-up dataset of reflectance data, resulting in the same five pigment groups. The approach used here tested the spectral gap hypothesis—i.e., that phytoplankton signals useful for characterizing phytoplankton pigment composition are contained on spectral scales narrower than the scale of other factors influencing optical properties (<100 nm). Overall, our results suggest that principal components regression modeling is a strong candidate for retrieving phytoplankton pigment composition from hyperspectral remote sensing data. The success of this model depended in part on rigorous quality control applied to both datasets before modeling, which ensured that only real features were magnified by the residual and derivative methods. Furthermore, the model works best at high (1–5 nm) spectral resolutions, and model performance decreases at coarser (10+ nm) resolution, which is relevant to future remote sensing instruments with improved spectral resolution (e.g., NASA's PACE sensor). Finally, this model is limited to the dataset for which it was developed; however, in combination with other remote sensing algorithms that target specific phytoplankton taxa, it would offer more information about both surface ocean optics and phytoplankton ecology, as it could help to illuminate some of the assumptions underlying both types of approaches. More high-quality, paired datasets from a range of different ecosystems and environments will also improve this approach and future global models for phytoplankton pigment composition.

Author credit

All authors conceived the study and developed the methods. SJK aggregated the dataset, performed the data analysis, and wrote the paper. SM implemented the reflectance model. All authors commented on and edited the manuscript.

Acknowledgments, funding, and data availability

- All data used in this work are publicly available, as referenced in Table 1. A compilation dataset is published at PANGAEA (Kramer et al., 2021: https://doi.pangaea.de/10.1594/PANGAEA.937536).
- Funding for this work was provided by NASA 80NSSC20M0226 to DAS and SM. SJK was supported by a National Defense Science and Engineering Graduate (NDSEG) fellowship through ONR.
- We are grateful for essential comments, suggestions, and support from Ivona Cetinić, Jeremy Werdell, and Emmanuel Boss. Thank you also to Ali Chase, Kelsey Bisson, B.B. Cael, and Mike Behrenfeld for many useful conversations about data, methods, and theory. Thank you to Ron Zaneveld for introducing DAS to the idea of the "mustache man," which is explored here.
- Thank you to three anonymous reviewers and to the editor, Frédéric Mélin, for thoughtful comments and suggestions that greatly improved this work.

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.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.rse.2021.112879.

References

Alvain, S., Moulin, C., Dandonneau, Y., Bréon, F.M., 2005. Remote sensing of phytoplankton groups in case 1 waters from global SeaWiFS imagery. Deep-Sea Res. I Oceanogr. Res. Pap. 52 (11), 1989–2004. https://doi.org/10.1016/j.dsr.2005.06.015.

Alvain, S., Moulin, C., Dandonneau, Y., Loisel, H., 2008. Seasonal distribution and succession of dominant phytoplankton groups in the global ocean: a satellite view. Glob. Biogeochem. Cycles 22 (3), 1–25. https://doi.org/10.1029/2007GB003154.

Barrón, R.K., Siegel, D.A., Guillocheau, N., 2014. Evaluating the importance of phytoplankton community structure to the optical properties of the Santa Barbara Channel, California. Limnol. Oceanogr. 59 (3), 927–946.

Behrenfeld, M.J., Bidle, K.D., Boss, E., Carlson, C.A., Gaube, P., Giovannoni, S., et al., 2014a. North Atlantic Aerosols and Marine Ecosystems Study (NAAMES). NASA Ocean Biology DAAC, SeaBASS. https://doi.org/10.5067/SeaBASS/NAAMES/DATA001.

Behrenfeld, M.J., Cetinić, I., Gilerson, A., Twardowski, M.S., 2014b. Ship-Aircraft Bio-Optical Research (SABOR). NASA Ocean Biology DAAC, SeaBASS. https://doi.org/10.5067/SeaBASS/SABOR/DATA001.

Behrenfeld, M.J., Benitez-Nelson, C.R., Boss, E., Buesseler, K.O., Carlson, C.A., Cassar, N., et al., 2018. EXPORTS. NASA Ocean Biology DAAC, SeaBASS. https://doi.org/10.5067/SeaBASS/EXPORTS/DATA001.

Bidigare, R.R., Morrow, J.H., Kiefer, D.A., 1989. Derivative analysis of spectral absorption by photosynthetic pigments in the western Sargasso Sea. J. Mar. Res. 47 (2), 323–341. https://doi.org/10.1357/002224089785076325.

Blondeau-Patissier, D., Gower, J.F., Dekker, A.G., Phinn, S.R., Brando, V.E., 2014. A review of ocean color remote sensing methods and statistical techniques for the detection, mapping and analysis of phytoplankton blooms in coastal and open oceans. Prog. Oceanogr. 123, 123–144. https://doi.org/10.1016/j.pocean.2013.12.008.

Boss, E., Claustre, H., 2009. Tara Oceans Expedition. NASA Ocean Biology DAAC, SeaBASS. https://doi.org/10.5067/SeaBASS/TARA_OCEANS_EXPEDITION/DATA001.

Boss, E., Claustre, H., 2014. Tara Mediterranean. NASA Ocean Biology DAAC, SeaBASS. https://doi.org/10.5067/SeaBASS/TARA_MEDITERRANEAN/DATA001.

Bracher, A., Vountas, M., Dinter, T., Burrows, J.P., Röttgers, R., Peeken, I., 2009. Quantitative observation of cyanobacteria and diatoms from space using PhytoDOAS on SCIAMACHY data. Biogeosciences 6, 751–764.

Bracher, A., Taylor, M.H., Taylor, B., Dinter, T., Röttgers, R., Steinmetz, F., 2015a. Using empirical orthogonal functions derived from remote-sensing reflectance for the prediction of phytoplankton pigment concentrations. Ocean Sci. 11, 139–158. https://doi.org/10.5194/os-11-139-2015.

Bracher, A., Taylor, M.H., Taylor, B., Dinter, T., Röttgers, R., Steinmetz, F., 2015b. Phytoplankton pigments, hyperspectral downwelling irradiance and remote sensing reflectance during POLARSTERN cruises ANT-XXIII/1, ANT-XXIV/1, ANT-XXIV/4, ANT-XXVI/4, and Maria S. Merian cruise MSM18/3. In: PANGAEA. https://doi.org/10.1594/PANGAEA.847820.

Bracher, A., et al., 2017. Obtaining phytoplankton diversity from ocean color: a scientific roadmap for future development. Front. Mar. Sci. 4, 1–15. https://doi.org/10.3389/fmars.2017.00055.

Brewin, R.J.W., Sathyendranath, S., Hirata, T., Lavender, S.J., Barciela, R.M., Hardman-Mountford, N., 2010. A three-component model of phytoplankton size class for the Atlantic Ocean. Ecol. Model. 221, 1472–1483. https://doi.org/10.1016/j.ecolmodel.2010.02.014.

Bricaud, A., Claustre, H., Ras, J., Oubelkheir, K., 2004. Natural variability of phytoplanktonic absorption in oceanic waters: influence of the size structure of algal populations. J. Geophys. Res. 109, 1–12. https://doi.org/10.1029/2004JC002419.

Brown, C.W., Yoder, J.A., 1994. Coccolithophorid blooms in the global ocean. J. Geophys. Res. 99 (C4), 7467–7482. https://doi.org/10.1029/93JC02156.

Cael, B.B., Chase, A.P., Boss, E.S., 2020. Information content of absorption spectra and implications for ocean color inversion. Appl. Opt. 39 (13), 3971–3984. https://doi.org/10.1364/AO.389189.

Carder, K.L., Chen, F.R., 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 (C3), 5403–5421. https://doi.org/10.1029/1998JC900082.

Casey, K.A., Rousseaux, C.S., Gregg, W.W., Boss, E., Chase, A.P., Craig, S.E., et al., 2019. In situ high spectral resolution inherent and apparent optical property data from diverse aquatic environments. PANGAEA. https://doi.org/10.1594/PANGAEA.902230.

Casey, K.A., Rousseaux, C.S., Gregg, W.W., Boss, E., Chase, A.P., Craig, S.E., et al., 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. https://doi.org/10.5194/essd-12-1123-2020.

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

Cetinić, I., 2013. RemSensPOC. NASA Ocean Biology DAAC, SeaBASS. https://doi.org/10.5067/SeaBASS/REMSENSPOC/DATA001.

Chase, A.P., Boss, E., Zaneveld, R., Bricaud, A., Claustre, H., Ras, J., et al., 2013. Decomposition of in situ particulate absorption spectra. Method Oceanogr. 7, 110–124. https://doi.org/10.1016/j.mio.2014.02.002.

Chase, A.P., Boss, E., Cetinić, I., Slade, W., 2017. Estimation of phytoplankton accessory pigments from hyperspectral reflectance spectra: toward a global algorithm. J. Geophys. Res. Oceans 122, 1–19. https://doi.org/10.1002/2017JC012859.

Chase, A.P., Kramer, S.J., Haëntjens, N., Boss, E.S., Karp-Boss, L., Edmondson, M., Graff, J.R., 2020. Evaluation of diagnostic pigments to estimate phytoplankton size classes. Limnol. Oceanogr. Methods 18, 570–584. https://doi.org/10.1002/lom3.10385.

Ciotti, A.M., Bricaud, A., 2006. Retrievals of a size parameter for phytoplankton and spectral light absorption by colored detrital matter from water-leaving radiances at SeaWiFS channels in a continental shelf region off Brazil. Limnol. Oceanogr. Methods 4 (7), 237–253. https://doi.org/10.4319/lom.2006.4.237.

Claustre, H., Sciandra, A., 2004. BIOSOPE cruise, RV L'Atalante. Sismer. https://doi.org/10.17600/4010100.

Falkowski, P.G., Oliver, M.J., 2007. Mix and match: how climate selects phytoplankton. Nat. Rev. Microbiol. 5, 813–819. https://doi.org/10.1038/nrmicro1751.

Gordon, H.R., Brown, O.B., Evans, R.H., Brown, J.W., Smith, R.C., Baker, K.S., Clark, D. K., 1988. A semianalytic radiance model of ocean color. J. Geophys. Res. 93 (D9), 10909–10924. https://doi.org/10.1029/JD093iD09p10909.

Hirata, T., Hardman-Mountford, N., Brewin, R.J.W., Aiken, J., Barlow, R., Suzuki, K., et al., 2011. Synoptic relationships between surface chlorophyll-a and diagnostic pigments specific to phytoplankton functional types. Biogeosciences 8, 311–327. https://doi.org/10.5194/bg-8-311-2011.

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 (C1), 1–25. https://doi.org/10.1029/2011JC007395.

Kahru, M., Anderson, C.R., Barton, A.D., Carter, M., Catlett, D., Send, U., Sosik, H.M., Weiss, E.L., Mitchell, B.G., 2021. Satellite detection of dinoflagellate blooms off California by UV reflectance ratios. Element Sci. Anthrop. 9 (1), 1–10. https://doi.org/10.1525/elementa.2020.00157.

Kostadinov, T.S., Siegel, D.A., Maritorena, S., 2010. Global variability of phytoplankton functional types from space: assessment via the particle size distribution. Biogeosciences 7, 3239–3257. https://doi.org/10.5194/bg-7-3239-2010.

Kramer, S.J., Siegel, D.A., 2019. How can phytoplankton pigments be best used to characterize surface ocean phytoplankton groups for ocean color remote sensing algorithms? J. Geophys. Res. Oceans 124, 7557–7574. https://doi.org/10.1029/2019JC015604.

Kramer, S.J., Roesler, C.S., Sosik, H.M., 2018. Bio-optical discrimination of diatoms from other phytoplankton in the surface ocean: evaluation and refinement of a model for the Northwest Atlantic. Remote Sens. Environ. 217, 126–143. https://doi.org/10.1016/j.rse.2018.08.010.

Kramer, S.J., Siegel, D.A., Graff, J.R., 2020. Phytoplankton community composition determined from co-variability among phytoplankton pigments from the NAAMES field campaign. Front. Mar. Sci. 7, 1–15. https://doi.org/10.3389/fmars.2020.00215.

Kramer, S.J., Siegel, D.A., Maritorena, S., Catlett, D., 2021. Global surface ocean HPLC phytoplankton pigments and hyperspectral remote sensing reflectance. PANGAEA. https://doi.org/10.1594/PANGAEA.937536.

Le Quéré, C., et al., 2005. Ecosystem dynamics based on plankton functional types for global ocean biogeochemistry models. Glob. Chang. Biol. 11, 2016–2040. https://doi.org/10.1111/j.1365-2486.2005.1004.x.

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 (27), 5755–5772. https://doi.org/10.1364/AO.41.005755.

Legendre, L., 1990. The significance of microalgal blooms for fisheries and for the export of particulate organic carbon in oceans. J. Plankton Res. 12 (4), 681–699. https://doi.org/10.1093/plankt/12.4.681.

Lin, Y., Gifford, S., Ducklow, H., Schofield, O., Cassar, N., 2019. Towards quantitative microbiome community profiling using internal standards. Appl. Environ. Microbiol. 85 (5), 1–14. https://doi.org/10.1128/AEM.02634-18.

Locarnini, R.A., Mishonov, A.V., Antonov, J.I., Boyer, T.P., Garcia, H.E., Baranova, O.K., et al., 2013. World Ocean Atlas 2013, Volume 1: Temperature (NOAA Atlas NESDIS 73), pp. 1–40.

Lubac, B., Loisel, H., Guiselin, N., Astoreca, R., Artigas, L.F., Mériaux, X., 2008. Hyperspectral and multispectral ocean color inversions to detect *Phaeocystis globosa* blooms in coastal waters. J. Geophys. Res. 113 (C06026), 1–17. https://doi.org/10.1029/2007JC004451.

Maritorena, S., Siegel, D.A., Peterson, A.R., 2002. Optimization of a semianalytical ocean color model for global-scale applications. Appl. Opt. 41 (15), 2705–2714. https://doi.org/10.1364/AO.41.002705.

Mason, J.D., Cone, M.T., Fry, E.S., 2016. Ultraviolet (250–550 nm) absorption spectrum of pure water. Appl. Opt. 55, 7163–7172. https://doi.org/10.1364/AO.55.007163.

Massy, W.F., 1965. Principal components regression in exploratory statistical research. J. Am. Stat. Assoc. 60 (309), 234–256. https://doi.org/10.2307/2283149.

McKinna, L.I., Cetinić, I., Werdell, P.J., 2021. Development and validation of an Empirical Ocean color algorithm with uncertainties: a case study with the particulate backscattering coefficient. J. Geophys. Res. Oceans 126 (5), 1–21. https://doi.org/10.1029/2021JC017231.

Morel, A., Gentili, B., 2009. A simple band ratio technique to quantify the colored dissolved and detrital organic material from ocean color remotely sensed data. Remote Sens. Environ. 113 (5), 998–1011. https://doi.org/10.1016/j.rse.2009.01.008.

Mouw, C.B., Hardman-Mountford, N., Alvain, S., Bracher, A., Brewin, R.J.W., Bricaud, A., Ciotti, A.M., Devred, E., Fujiwara, A., Hirata, T., Hirawake, T., Kostadinov, T.S., Roy, S., Uitz, J., 2017. A consumer's guide to satellite remote sensing of multiple phytoplankton groups in the global ocean. Front. Mar. Sci. 4, 1–19. https://doi.org/10.3389/fmars.2017.00041.

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

Roesler, C.S., Perry, M.J., 1995. In situ phytoplankton absorption, fluorescence emission, and particulate backscattering determined from reflectance. J. Geophys. Res. 100 (C7), 13279–13294. https://doi.org/10.1029/95JC00455.

Sadeghi, A., Dinter, T., Vountas, M., Taylor, B., Altenburg-Soppa, M., Bracher, A., 2012. Remote sensing of coccolithophore blooms in selected oceanic regions using the PhytoDOAS method applied to hyper-spectral satellite data. Biogeosciences 9 (6), 2127–2143. https://doi.org/10.5194/bg-9-2127-2012.

Sathyendranath, S., Watts, L., Devred, E., Platt, T., Caverhill, C., Maass, H., 2004. Discrimination of diatoms from other phytoplankton using ocean-colour data. Mar. Ecol. Prog. Ser. 272, 59–68. https://doi.org/10.3354/meps272059.

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. https://doi.org/10.1364/OE.26.007404.

Siegel, D.A., Maritorena, S., Nelson, N.B., Hansell, D.A., Lorenzi-Kayser, M., 2002. Global distribution and dynamics of colored dissolved and detrital organic materials. J. Geophys. Res. Oceans 107 (C12), 1–14. https://doi.org/10.1029/2001JC000965.

Siegel, D.A., Maritorena, S., Nelson, N.B., Behrenfeld, M.J., McClain, C.R., 2005. Colored dissolved organic matter and its influence on the satellite-based characterization of the ocean biosphere. Geophys. Res. Lett. 32 (L20605), 1–4. https://doi.org/10.1029/2005GL024310.

Siegel, D.A., Behrenfeld, M.J., Maritorena, S., McClain, C.R., Antoine, D., Bailey, S.W., et al., 2013. Regional to global assessments of phytoplankton dynamics from the SeaWiFS mission. Remote Sens. Environ. 135, 77–91. https://doi.org/10.1016/j.rse.2013.03.025.

Siegel, D.A., Buesseler, K.O., Doney, S.C., Sailley, S.F., Behrenfeld, M.J., Boyd, P.W., 2014. Global assessment of ocean carbon export by combining satellite observations and food-web models. Glob. Biogeochem. Cycles 28, 181–196. https://doi.org/10.1002/2013GB004743.

Smith, M.E., Bernard, S., 2020. Satellite Ocean color based harmful algal bloom indicators for aquaculture decision support in the southern Benguela. Front. Mar. Sci. 7 (61), 1–13. https://doi.org/10.3389/fmars.2020.00061.

Soppa, M.A., Hirata, T., Silva, B., Dinter, T., Peeken, I., Wiegmann, S., Bracher, A., 2014. Global retrieval of diatom abundance based on phytoplankton pigments and satellite data. Remote Sens. 6 (10) https://doi.org/10.3390/rs61010089.

Stramski, D., Bricaud, A., Morel, A., 2001. Modeling the inherent optical properties of the ocean based on the detailed composition of the planktonic community. Appl. Opt. 40 (18), 2929–2945. https://doi.org/10.1364/AO.40.002929.

Stumpf, R.P., Culver, M.E., Tester, P.A., Tomlinson, M., Kirkpatrick, G.J., Pederson, B.A., Truby, E., Ransibrahmanakul, V., Soracco, M., 2003. Monitoring *Karenia brevis* blooms in the Gulf of Mexico using satellite ocean color imagery and other data. Harmful Algae 2 (2), 147–160. https://doi.org/10.1016/S1568-9883(02)00083-5.

Taylor, B.B., Torrecilla, E., Bernhardt, A., Taylor, M.H., Peeken, I., Röttgers, R., et al., 2011. Bio-optical provinces in the eastern Atlantic Ocean and their biogeographical relevance. Biogeosciences 8 (12), 3609–3629. https://doi.org/10.5194/bg-8-3609-2011.

Taylor, B.B., Taylor, M., Dinter, T., Bracher, A., 2013. Estimation of relative phycoerythrin concentrations from hyperspectral underwater radiance measurements – a statistical approach. J. Geophys. Res. Oceans 118, 2948–2960. https://doi.org/10.1002/jgrc.20201.

Torrecilla, E., Stramski, D., Reynolds, R.A., Millán-Núñez, E., Piera, J., 2011. Cluster analysis of hyperspectral optical data for discriminating phytoplankton pigment assemblages in the open ocean. Remote Sens. Environ. 115, 2578–2593. https://doi.org/10.1016/j.rse.2011.05.014.

Tsai, F., Philpot, W., 1998. Derivative analysis of hyperspectral data. Remote Sens. Environ. 66 (1), 41–51. https://doi.org/10.1016/S0034-4257(98)00032-7.

Uitz, J., Stramski, D., Reynolds, R.A., Dubranna, J., 2015. Assessing phytoplankton community composition from hyperspectral measurements of phytoplankton absorption coefficient and remote-sensing reflectance in open-ocean environments. Remote Sens. Environ. 171, 58–74. https://doi.org/10.1016/j.rse.2015.09.027.

Valente, A., Sathyendranath, S., Brotas, V., Groom, S., Grant, M., Taberner, M., et al., 2019. A compilation of global bio-optical in situ data for ocean-colour satellite applications – version two. Earth Syst. Sci. Data 11 (3), 1037–1068. https://doi.org/10.5194/essd-11-1037-2019.

Van Heukelem, L., Thomas, C.S., 2001. Computer-assisted high-performance liquid chromatography method development with applications to the isolation and analysis of phytoplankton pigments. J. Chromatogr. A 910. https://doi.org/10.1016/S0378-4347(00)00603-4.

Vanni, M.J., Findlay, D.L., 1990. Trophic cascades and phytoplankton community structure. Ecology 71 (3), 921–937. https://doi.org/10.2307/1937363.

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. https://doi.org/10.1016/j.rse.2005.07.001.

Werdell, P.J., McKinna, L.I.W., Boss, E., Ackleson, S.G., Craig, S.E., Gregg, W.W., et al., 2018. An overview of approaches and challenges for retrieving marine inherent optical properties from ocean color remote sensing. Prog. Oceanogr. 160, 186–212. https://doi.org/10.1016/j.pocean.2018.01.001.

Werdell, P.J., Behrenfeld, M.J., Bontempi, P.S., Boss, E., Cairns, B., Davis, G.T., et al., 2019. The plankton, aerosol, cloud, ocean ecosystem (PACE) mission: status, science, advances. Bull. Am. Meteorol. Soc. 1–59 https://doi.org/10.1175/BAMS-D-18-0056.1.

Westberry, T.K., Siegel, D.A., 2006. Spatial and temporal distribution of Trichodesmium blooms in the world's oceans. Glob. Biogeochem. Cycles 20 (GB4016), 1–13. https://doi.org/10.1029/2005GB002673.

Westberry, T.K., Siegel, D.A., Subramaniam, A., 2005. An improved bio-optical algorithm for the remote sensing of Trichodesmium spp. blooms. J. Geophys. Res. Oceans 110 (C6), 1–11. https://doi.org/10.1029/2004JC002517.

Wolanin, A., Soppa, M.A., Bracher, A., 2016. Investigation of spectral band requirements for improving retrievals of phytoplankton functional types. Remote Sens. 8 (871), 1–21. https://doi.org/10.3390/rs8100871.

Xi, H., Hieronymi, M., Röttgers, R., Krasemann, H., Qiu, Z., 2015. Hyperspectral differentiation of phytoplankton taxonomic groups: a comparison between using remote sensing reflectance and absorption spectra. Remote Sens. 7, 14781–14805. https://doi.org/10.3390/rs71114781.

Xi, H., Hieronymi, M., Krasemann, H., Röttgers, R., 2017. Phytoplankton group identification using simulated and in situ hyperspectral remote sensing reflectance. Front. Mar. Sci. 4 (272), 1–13. https://doi.org/10.3389/fmars.2017.00272.

Zhang, X., Hu, L., He, M.-X., 2009. Scattering by pure seawater: effect of salinity. Opt. Express 17, 5698–5710. https://doi.org/10.1364/OE.17.005698.

Zweng, M.M., Reagan, J.R., Antonov, J.I., Locarnini, R.A., Mishonov, A.V., Boyer, T.P., et al., 2013. World Ocean Atlas 2013, Volume 2: Salinity (NOAA Atlas NESDIS 74), pp. 1–39.