Remote Sensing of Environment 217 (2018) 126–143

Contents lists available at ScienceDirect

Remote Sensing of Environment

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

Bio-optical discrimination of diatoms from other phytoplankton in the surface ocean: Evaluation and refinement of a model for the Northwest Atlantic

Sasha J. Kramer[a,b,c,*], Collin S. Roesler[b], Heidi M. Sosik[c]

[a] Interdepartmental Graduate Program in Marine Science, University of California Santa Barbara, Santa Barbara, CA, United States of America
[b] Department of Earth and Oceanographic Science, Bowdoin College, Brunswick, ME, United States of America
[c] Biology Department, Woods Hole Oceanographic Institution, Woods Hole, MA, United States of America

ARTICLE INFO

Keywords:
Phytoplankton
Community structure
Ocean color
Diatoms

ABSTRACT

Diatoms dominate global silica production and export production in the ocean; they form the base of productive food webs and fisheries. Thus, a remote sensing algorithm to identify diatoms has great potential to describe ecological and biogeochemical trends and fluctuations in the surface ocean. Despite the importance of detecting diatoms from remote sensing and the demand for reliable methods of diatom identification, there has not been a systematic evaluation of algorithms that are being applied to this end. The efficacy of these models remains difficult to constrain in part due to limited datasets for validation. In this study, we test a bio-optical algorithm developed by Sathyendranath et al. (2004) to identify diatom dominance from the relationship between ratios of remote sensing reflectance and chlorophyll concentration. We evaluate and refine the original model with data collected at the Martha's Vineyard Coastal Observatory (MVCO), a near-shore location on the New England shelf. We then validated the refined model with data collected in Harpswell Sound, Maine, a site with greater optical complexity than MVCO. At both sites, despite relatively large changes in diatom fraction (0.8–82% of chlorophyll concentration), the magnitude of variability in optical properties due to the dominance or non-dominance of diatoms is less than the variability induced by other absorbing and scattering constituents of the water. While the original model performance was improved through successive re-parameterizations and re-formulations of the absorption and backscattering coefficients, we show that even a model originally parameterized for the Northwest Atlantic and re-parameterized for sites such as MVCO and Harpswell Sound performs poorly in discriminating diatom-dominance from optical properties.

1. Introduction

Phytoplankton comprise only 0.2% of photosynthetically active biomass on Earth, yet they are responsible for half of global primary production (Behrenfeld and Falkowski, 1997; Falkowski et al., 1998; Field et al., 1998). In addition to forming the base of the marine food web, these organisms represent an essential source of elemental compounds and nutrients to the ocean (Redfield, 1934; Arrigo, 2005). There are thousands of known phytoplankton species, but the expansive taxonomic diversity of phytoplankton can be simplified by combining groups of species according to their functional or biogeochemical roles in an ecosystem (Le Quéré et al., 2005). The diatoms comprise a major group of phytoplankton: despite physiological and morphological differences between species, all diatoms contribute to oceanic silica production. As the phytoplankton group that contributes the most to phytoplankton carbon, diatoms efficiently support higher trophic levels and dominate export production in the global ocean (Cushing, 1989; Smetacek, 1999). Thus, understanding the distribution and abundance of diatoms within broader phytoplankton communities is essential to quantifying the impacts of this functional group on macronutrient cycles, trophic transfer, carbon export, and fisheries (Legendre, 1990; Arrigo, 2005; Falkowski and Oliver, 2007; Guidi et al., 2009).

Detecting the distribution of diatoms is difficult, however, as many factors confound the direct sampling of global phytoplankton communities. Developing in situ methods for sampling phytoplankton on large scales represents a logistical and financial challenge that can be prohibitive for answering questions about the distributions and environmental impacts of different phytoplankton functional groups. Recently,

* Corresponding author at: Interdepartmental Graduate Program in Marine Science, University of California Santa Barbara, Santa Barbara, CA, United States of America.
E-mail address: sasha.kramer@lifesci.ucsb.edu (S.J. Kramer).

https://doi.org/10.1016/j.rse.2018.08.010
Received 6 December 2017; Received in revised form 3 August 2018; Accepted 10 August 2018
Available online 15 August 2018

there has been a great deal of interest in developing methods to use satellite-based ocean color remote sensors to study phytoplankton diversity on broad spatial and temporal scales in the surface ocean. These methods exploit spectral differences in remote sensing data to retrieve properties of the phytoplankton community (IOCCG, 2014 and references therein; Bracher et al., 2017; Mouw et al., 2017). Some bio-optical algorithms target phytoplankton size structure (e.g., Ciotti et al., 2002; Devred et al., 2006; Uitz et al., 2006; Kostadinov et al., 2009) or seek to identify multiple phytoplankton types at once (e.g., Alvain et al., 2008; Hirata et al., 2008; Nair et al., 2008; Bracher et al., 2009). Other models work to distinguish one dominant phytoplankton type from all other phytoplankton (e.g., Gordon et al., 2001; Westberry and Siegel, 2006).

One such algorithm is presented in Sathyendranath et al. (2004). This algorithm (hereafter denoted as S04) uses two curves of remote sensing reflectance ratios, $R(\lambda_1)/R(\lambda_2)$, computed as a function of chlorophyll-*a* concentration to distinguish between diatom-dominated surface ocean waters and waters containing mixed phytoplankton species (Fig. S1). The upshot of this model is that, for a given chlorophyll concentration, diatom-dominated waters will appear bluer compared to those dominated by a mixed phytoplankton composition.

We reconstructed the S04 model from multiple sources as described in detail in the Supplementary material (Section S1). In brief, the authors built a forward model of remote sensing reflectance as a function of absorption and scattering spectra. Inherent in this forward model are spectral differences in the phytoplankton absorption coefficients that reflect differences in the optical signatures of diatoms compared to all other phytoplankton groups. Variations in phytoplankton size, pigment composition and density, and the degree of pigment packaging within cells give phytoplankton groups distinct optical signatures that may affect the water-leaving radiance signal detected by satellites (Morel and Bricaud, 1981; Roesler et al., 1989; Bidigare et al., 1989; Hoepffner and Sathyendranath, 1993; Sosik and Mitchell, 1994; Ulloa et al., 1994). The model contains terms for absorption and scattering by pure seawater, absorption by yellow matter, and scattering by phytoplankton and other particles. The S04 model relies on an inherent optical property (IOP)-based approximation to the radiative transfer equation to compute reflectance ratio curves (Fig. S1). In the forward component of the model, absorption and scattering by phytoplankton and other particles and absorption by yellow matter all vary as a function of chlorophyll-*a* and thus are biomass-dependent parameters—however, as the forward model produces two reflectance ratios for diatom-dominated and mixed phytoplankton groups at each chlorophyll concentration, the model is considered radiance-based rather than abundance-based (in the terminology reviewed by Mouw et al., 2017; for full consideration of this issue, please see the Discussion).

The S04 model was originally designed for application in the Northwest Atlantic Zone, which is an oceanographic region that encompasses several biogeographic provinces (Longhurst et al., 1995; Sathyendranath et al., 1995; Longhurst, 1998). Many of the samples used to develop the model were collected from the Northwest Coastal Shelf province (Sathyendranath et al., 1995; Sathyendranath et al., 2004; Platt et al., 2005). This model has been implemented several times in the region for which it was developed, but no further evaluation of the model performance has been published to date (Platt et al., 2005; Son et al., 2007; Zhai et al., 2008; Platt et al., 2010; Trzcinski et al., 2013; Budge et al., 2014). Sathyendranath et al. (2004) suggest that the algorithm should be modified to reflect local conditions before application in other regions due to variations in the optical properties of diatoms. Accordingly, when the model was applied in the waters off the coast of Chile, Jackson et al. (2011) empirically tuned both the threshold of pigment-based diatom identification and the coefficients for the modeled phytoplankton absorption curves to match their measured pigment concentration and phytoplankton absorption data. The regionally-tuned model correctly identified the phytoplankton community as mixed or diatom-dominated at seven stations while the original model intended for the Northwest Atlantic misidentified two of the diatom-dominated stations as mixed (Jackson et al., 2011). Notably, however, this reported performance is not an independent validation since the same observations were used for model tuning.

Arguably, the ability to identify diatoms from ocean color remote sensing data would enhance current knowledge of the ecology and biogeochemistry of the surface ocean, and applications of the S04 algorithm go beyond simply identifying the presence or absence of diatoms at a given place and time. This model has been invoked in studies examining the power of hurricanes to shift phytoplankton community structure and nutrient concentration following a physical overturning (Son et al., 2007). The model has also been applied to investigate trophic exchange in the North Atlantic: the presence of diatoms may explain trends in cod and haddock recruitment to the coastal shelf region (Trzcinski et al., 2013). Similarly, in identifying the relative fraction of diatoms in the surface ocean, the model was used to estimate the total concentration of omega-3 fatty acids in the Northwest Atlantic Ocean (Budge et al., 2014). Thus, this remote sensing algorithm to identify diatoms (and others like it) is already being used to interpret ecological and biogeochemical trends and fluctuations in the surface ocean. However, without proper validation, it is difficult to identify the significance or uncertainties associated with a model output before it is used for other applications.

Most bio-optical models lack sufficient validation metrics or validation data products (Anderson, 2005; Bracher et al., 2017; Mouw et al., 2017). Complete validation of a bio-optical model requires a multifaceted independent dataset: IOP data (absorption and backscattering by seawater constituents), radiometry, and means of assessing the phytoplankton community composition directly rather than by proxy. Complete validation datasets are rare because these measurements are difficult to obtain, particularly on concurrent space and time scales. In the S04 model development, the authors validated their algorithm outputs with one year of data that was excluded from the dataset used to construct the model. The algorithm correctly identified seven out of ten validation stations as either diatom-dominated or containing mixed phytoplankton taxa. Considering that diatoms often dominate microplankton, Mouw and Yoder (2010) compared the output of the S04 model to the outputs of their bio-optical algorithm to identify the fraction of microplankton in the surface ocean. The S04 model outputs predicted diatom dominance in up to 75% of cases where the Mouw and Yoder (2010) model predicted microplankton dominance in the Northwest Atlantic, with highest correspondence in high chlorophyll locations. Thus, while the S04 algorithm has been widely applied and the outputs analyzed, the model itself has not been independently validated with a bio-optical dataset of in situ measurements including IOPs, apparent optical properties (AOPs), pigments, and microscopy.

In this study, we use an extensive dataset of bio-optical properties measured at the Martha's Vineyard Coastal Observatory (MVCO), which falls in the Northwest Coastal Shelf province within the Northwest Atlantic Zone, to evaluate the performance of the S04 model. The MVCO area features a systematic seasonal cycle with diatom-dominance in winter and smaller, mixed phytoplankton communities in summer (Sosik and Olson, 2008; Sosik et al., 2010; Peacock et al., 2014; Hunter-Cevera et al., 2016). The measurements at this site are well suited to test the model: the MVCO dataset includes optical measurements (reflectance, component absorption and backscattering) to evaluate functional relationships and algorithm parameters, as well as metrics describing the phytoplankton community (High Performance Liquid Chromatography or HPLC phytoplankton pigments, flow cytometry and cell imaging) to test the model outcome.

To evaluate the appropriateness of this study site for validation of the S04 model, we compared the reflectance-ratio-to-chlorophyll relationships at MVCO (Fig. 1A) to other Case 1 sites found in the NASA bio-Optical Marine Algorithm Dataset (NOMAD). Case 1 was originally defined to refer to green waters for which the reflectance spectrum

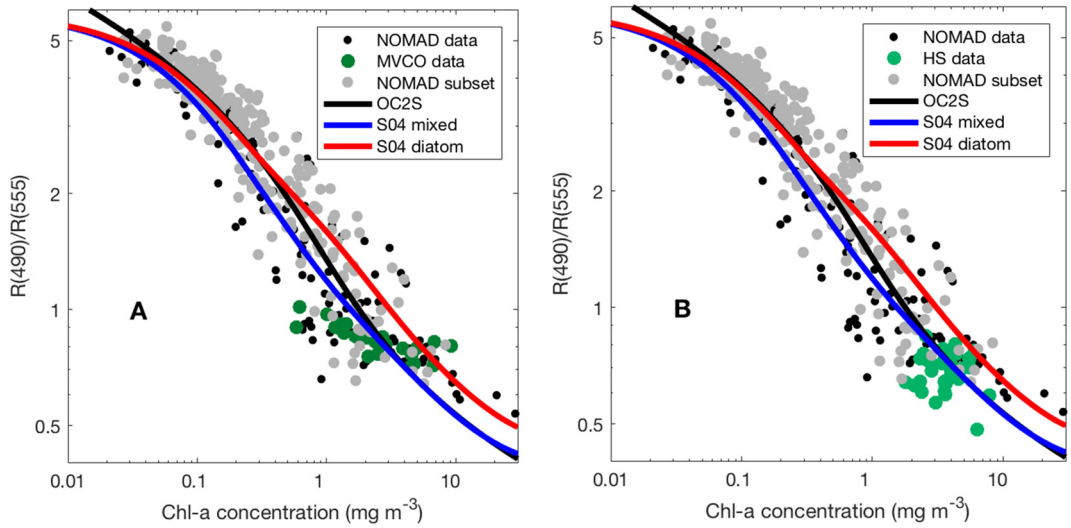

Fig. 1. Relationship between R490:R555 and chlorophyll-*a*. NASA OC2vS model for the SeaWiFS sensor (black; https://oceancolor.gsfc.nasa.gov/atbd/chlor_a/) and original S04 models for mixed phytoplankton populations (blue) and diatom-dominated populations (red). The models overlay all data from the NASA bio-Optical Marine Algorithm Dataset (NOMAD, in black; Werdell and Bailey, 2005); a Case 1 subset of NOMAD (defined following Lee and Hu, 2006, in grey); and data measured at (A) MVCO and (B) Harpswell Sound (in green). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

cannot be explained by simply the optical properties of pure water and particle backscattering, but for which the absorption is dominated by phytoplankton pigments rather than inorganic particles. In addition, Morel and Prieur (1977) acknowledged that "dissolved yellow substance is present in variable amounts and also contributes to absorption." However, over time, this metric has developed new nuance and implications that have resulted in misuse of the term or classification in some cases (Mobley et al., 2004 and references therein). To avoid complication, we here adhere to the definition that Case 1 waters are not just open ocean waters, waters with low chlorophyll concentrations (for instance, the range in chlorophyll concentration associated with Morel and Prieur's, 1977 Case 1 waters is 0.2 to 18.1 mg m^{-3}), waters without variation in CDOM (colored dissolved organic matter) absorption and non-algal particle absorption and scattering, or any combination of the above. Case 1 waters can be green, coastal waters with high chlorophyll concentration and varying concentrations of CDOM and NAP (Sauer et al., 2012; Antoine et al., 2014).

The S04 model was designed for Case 1 waters, but the sampling locations for model construction and validation vary from very nearshore to open ocean (Fig. 2A). While MVCO is a very nearshore site, the blue to green reflectance-ratio-to-chlorophyll relationships observed at MVCO fit well within the range of relationships observed across a global compilation of Case 1 sites. While the NOMAD dataset has been applied extensively for bio-optical algorithm development and validation (i.e., O'Reilly et al., 1998, Morel and Gentili, 2009, Lee et al., 2010, Maritorena et al., 2010, Sauer et al., 2012, Siegel et al., 2013, Werdell et al., 2013, etc.), the dataset overrepresents eutrophic and very coastal waters (Werdell and Bailey, 2005). Thus, we chose to use a Case 1 subset of the dataset as defined using the metric of Lee and Hu (2006). This method aims to classify sites as Case 1 or Case 2 based on the range of the reflectance ratio values, allowing for some spread around a mean value due to the variability of CDOM and non-algal particle absorption and scattering in natural waters. This spread can also be attributed to phytoplankton community composition, as it is in the S04 approach (for more consideration of this concept, please see the Discussion). The Lee and Hu (2006) method was also applied to our observational datasets: at MVCO, 50% of the data points used in this analysis (13/26) were considered to be Case 1 as defined by Lee and Hu (2006); in Harpswell Sound, 26% of the data points (6/23) were considered Case 1.

Despite every indication that the original S04 model should fit the data collected at MVCO given the range of data found at the site compared to other Case 1 sites (Fig. 1A), initial evaluation demonstrated that the model did not fit the MVCO observations. Thus, we used data collected at MVCO for a series of successive re-parameterizations of each component part of the original reflectance model using data collected at MVCO, with the goals of identifying which component was responsible for the poor fit and improving the fit between the model and the measured optical data at MVCO. The refined model (hereafter designated as K18) was then validated with data collected at the Bowdoin Buoy in Harpswell Sound (bowdoin.loboviz.com), a tidally-dominated productive inlet in eastern Casco Bay, Maine with similar optical complexity and phytoplankton community composition to MVCO (Chase et al., 2009) but a notably different range of blue to green reflectance-ratio-to-chlorophyll than what is found at MVCO (Fig. 1B).

In validating, re-parameterizing and re-formulating, and re-validating the S04 model with data collected in the region for which the model was originally developed, we demonstrate the importance of rigorously testing any model with in situ data before widespread model application. The resulting bio-optical algorithm combines empirical observations and measurements of biological and optical properties with knowledge of the underlying oceanographic system to create a new model, K18, that better fits the measured data. The process of refining the S04 model highlights the difficulty of attributing variability in the reflectance ratios measured by satellites to only one component of the ocean, here the dominant phytoplankton group, when in fact this variability can be attributed to multiple sources in most ocean ecosystems.

2. Methods

2.1. MVCO

To evaluate and then re-parameterize and refine the S04 model, we relied on an extensive dataset of biological and optical parameters, including IOPs, AOPs, pigments, flow cytometry, and automated microscopy at MVCO. Here, the spatial context of data used to construct and validate the S04 model is replaced by a temporal context at MVCO, as the data were collected in all seasons and show a large dynamic range over the ten year time series used here (Zhang et al., 2015; Supplementary material). The instruments used in this study were

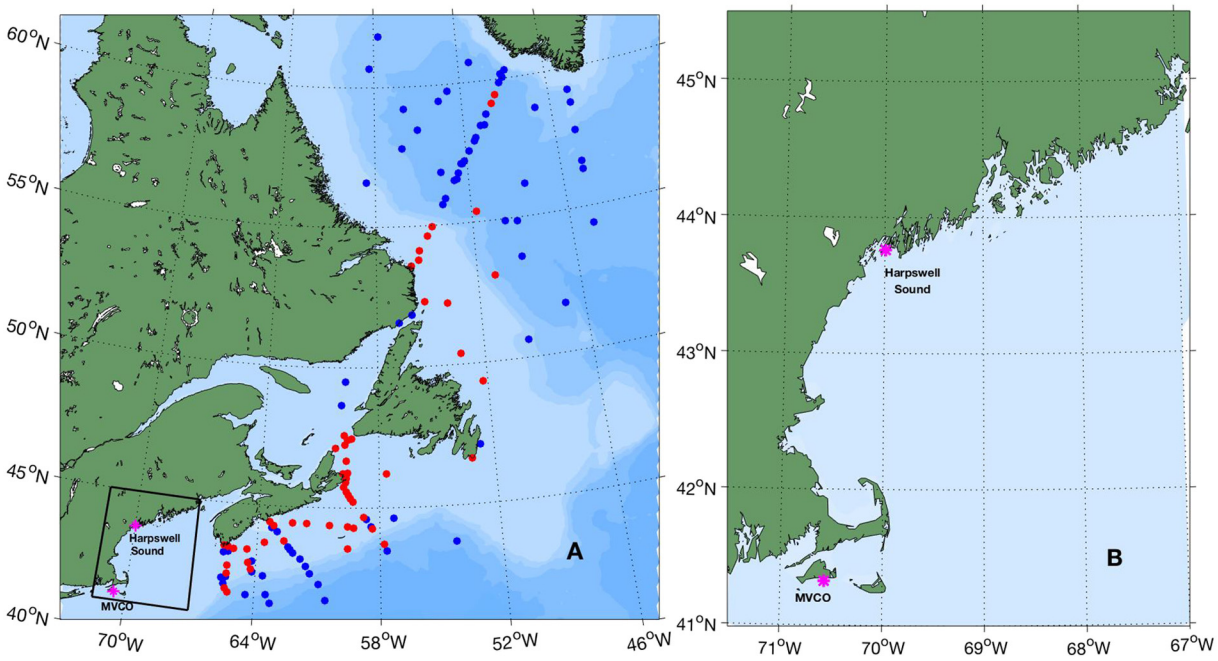

Fig. 2. (A) Sampling locations of Sathyendranath et al. (2004) where red points are diatom-dominated sites and blue points are mixed populations. The locations of MVCO and Harpswell Sound are in pink; detailed area shown in (B) outlined in black. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

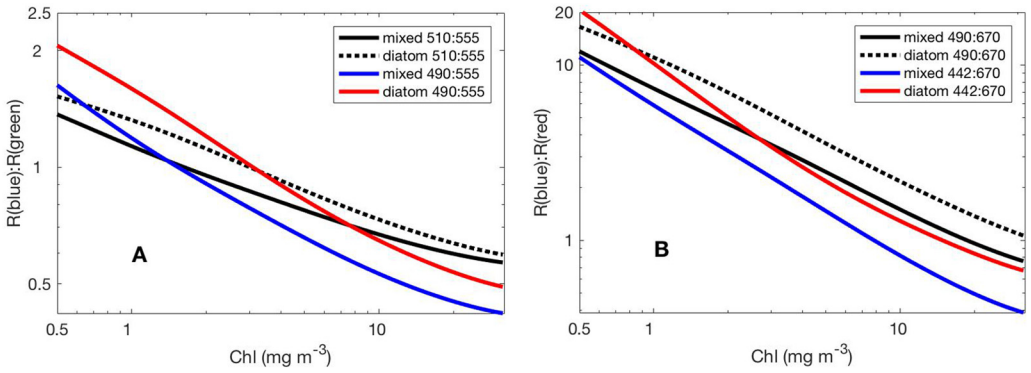

Fig. 3. Modeled relationships between ratios of reflectance and chlorophyll-*a* concentration. (A) Blue-to-green ratio and (B) blue-to-red ratio at wavelengths used in the original S04 model (solid black = mixed; dashed black = diatoms) and at MVCO measured wavelengths (blue = mixed; red = diatoms). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

deployed on the Air-Sea Interaction Tower (ASIT) at MVCO, which is located 3 km south of the island of Martha's Vineyard (41°19.5′N, 70°34.0′W; Fig. 2) in 15 m. deep water, which is between 4 and 15 optical depths, depending on wavelength and range of optical properties. To compare data collected from 2004 to 2015 on different instruments over a similar scale, a matchup routine was used to find datapoints that had been sampled within the same 4-hour window. For details on sample numbers used in each analysis resulting from these matchups, see Supplementary material Section S2; for details on uncertainty in measurements, see Supplementary material, Section S6, Table S3. Since the original S04 model is designed for application to satellite remote sensing ocean color data, we only used samples collected in the near surface (< 5 m).

The data products at MVCO used in this study include the following: remote sensing reflectance, discrete sample absorption, in situ measured backscattering coefficients, discrete phytoplankton pigments from HPLC analysis, and in situ flow-cytometry/cell imaging.

The AERONET-OC above-water radiometry system (i.e., SeaPRISM) measures downwelling irradiance and upwelling radiance just above the sea surface, which are used to compute remote sensing reflectance (Zibordi et al., 2010). The SeaPRISM has a full angle field of view of 1.2°. As mounted on the ASIT, it takes measurements from ~10 m above the water, allowing for minimal atmospheric interference. The instrument is configured to measure reflectance at eight distinct wavelengths: 412.7 nm, 442 nm, 490.9 nm, 531.4 nm, 555.1 nm, 668.1 nm, 870.2 nm, and 1019.8 nm. Raw data are uploaded to the AERONET data system maintained at NASA's Goddard Space Flight Center and calibrated, quality controlled standard products were subsequently downloaded from https://aeronet.gsfc.nasa.gov. As the SeaPRISM measurements can be impacted by surface reflectance, we only used measurements between 10:00 and 14:00 local time for the matchup. Reflectance at MVCO was not measured at the same wavelengths as were used to construct the S04 model. While the use of bluer wavelengths (490 nm compared to 510 nm and 442 nm compared to 490 nm) could make the ratios more vulnerable to the impacts of CDOM absorption, which decreases exponentially with wavelength, the use of different wavelengths does not change the critical feature of significant separation between the mixed and diatom-dominated model curves (Fig. 3).

Absorption coefficients were determined spectrophotometrically on discrete samples collected approximately monthly at the ASIT. Particles (typically from 500 mL water sample) were collected on GF/F glass fiber filters and stored at liquid nitrogen temperatures until analysis on a dual beam spectrophotometer (Perkin Elmer Lambda-18) equipped

with 60-mm integrating sphere attachment. Filters were mounted on quartz plates at the entrance to the sphere and scanned with a hydrated blank filter in the reference beam to determine total particle absorption (a_p). Following the method of Kishino et al. (1985), absorption by phytoplankton (a_{ph}) and non-algal particles (a_{NAP}) were separated by rescanning the sample filter after extraction of pigments with methanol. In all cases, the spectrophotometer was autozeroed with blank filters in both beams prior to sample scans. The average optical density in the range 780–800 nm was subtracted from each sample and the pathlength amplification factor of Mitchell (1990) was applied to correct for filter effects. Separate samples were collected for determination of absorption by colored dissolved organic matter (a_{CDOM}). Filtrate (0.2 μm) was stored in refrigerated detergent-soaked, acid-washed glass bottles until spectrophotometric analysis in 10-cm cuvettes, with pure water (Milli-Q) in the reference beam (after autozero with pure water in both sample and reference cuvettes). Samples were warmed to room temperature before analysis and care was taken to maintain sample water and reference water at the same temperature. A blank scan (pure water in both cuvettes after autozero) was subtracted from each sample spectrum and then the average optical density in the range 660–670 nm was subtracted to minimize any residual effects of temperature and salinity differences.

Additional discrete water samples were taken at the ASIT for determination of phytoplankton pigment concentrations (chlorophylls and accessory pigments). Particles were collected on GF/F filters and stored at liquid nitrogen temperatures until HPLC analysis by the NASA GSFC field support group (https://oceancolor.gsfc.nasa.gov/fsg/hplc). We used the CHEMTAX matrix inversion method (Mackey et al., 1996) to estimate the contribution of diatoms to the total chlorophyll-*a* concentration. Measured pigment concentrations (chlorophyll-*a*, chlorophyll *b*, chlorophyll *c*1 + *c*2, chlorophyll *c*3, 19′-hexanoyloxyfucoxanthin, 19′-butanoyloxyfucoxanthin, fucoxanthin, peridinin, alloxanthin, diatoxanthin + diadinoxanthin, lutein, neoxanthin, prasinoxanthin, violaxanthin, and zeaxanthin) were provided as input, along with initial taxon-specific pigment ratios as in Pan et al. (2011). The algorithm searches for optimal pigment ratios and associated taxon-specific pigment concentrations for diatoms, cryptophytes, cyanophytes, two types of dinoflagellates, two types of haptophytes, and two types of chlorophytes. To minimize problems associated with local minima (Latasa, 2007), we ran the CHEMTAX optimization 108 separate times with random noise (± 35%) added to the initial pigment ratios, then computed the average of the 10 solutions with the lowest overall residuals. Inspection of the residuals confirmed an absence of seasonal- or annual-scale biases that would warrant splitting the dataset to allow for non-stationary pigment ratios.

A HydroScat-6 sensor (HOBI Labs) was deployed at 4 m depth at the ASIT for in situ determination of backscattering coefficients (b_b) at multiple wavelengths (420 nm, 442 nm, 470 nm, 510 nm, 550 nm, 590 nm, and 700 nm). The sensor was regularly calibrated at HOBI Labs Inc. prior to 2012 and at HOBI Instrument Services LLC thereafter. Calibrated backscattering coefficients were computed with HydroSoft (version 2.95). Hourly averaged values were computed from observations made at 15-second intervals.

In situ characterization of the phytoplankton community was carried out by automated flow cytometry with FlowCytobot (FCB; Olson et al., 2003) and imaging-in-flow cytometry with the Imaging FlowCytobot (IFCB; Olson and Sosik, 2007). FCB observations were used to count and characterize pico- and small nanoplankton (2–10 μm) and IFCB was used to count and characterize large nano- and microplankton (> 10 μm), including separation of diatoms from other taxa (Sosik and Olson, 2007). Biomass of each cell (or chain/colony in the case of many diatoms) was estimated by converting bio-volume to carbon according to the meta-analysis of Menden-Deuer and Lessard (2000), which showed that a single relationship holds for all cell types (pico-cyanobacteria and protists) except large diatoms. Biovolume for each cell (or chain/colony) was determined from light scattering measurements for FCB data (calibrated with a wide range of cultured cell types; Olson et al., 2003) and from image analysis for IFCB data (Moberg and Sosik, 2012). In a given sample, diatom carbon was determined by summing the contributions of individual diatom cells and colonies/chains. The diatom fraction of total phytoplankton carbon was then computed by dividing by the corresponding sum for all phytoplankton.

2.2. Harpswell Sound

The K18 model was further validated with data collected from 2008 to 2016 in productive Harpswell Sound in eastern Casco Bay, Maine at the Bowdoin Buoy (GoMICOOS Buoy D), which features optically complex Case 2 waters (Fig. 1B). Separate data types collected in Harpswell Sound were matched to find data-points sampled on the same day at the same time, from depths above 5 m. The data products from Harpswell Sound used in this study include reflectance and HPLC phytoplankton pigments. As the goal with the Harpswell Sound dataset was to validate and not retune the model, absorption and backscattering observations, while available, were not used.

Upwelling radiance (L_u) and downwelling irradiance were measured with a hyperspectral radiometer buoy (HTSRB; Satlantic Instruments, Sea-Bird Scientific), which was deployed on weekly cruises in Harpswell Sound each summer. The buoy was deployed off the sunny side of the ship at a distance of at least 3 optical depths away to avoid both ship shadow and ship reflection. L_u is measured 0.63 m below the base of the buoy and was propagated to the surface with an estimate of spectral diffuse attenuation, $K_u(\lambda)$. As we did not have profiles of $L_u(\lambda, z)$, we assumed that $K_u(\lambda)$ could be approximated by the diffuse attenuation of downward irradiance, which we estimated from Morel et al. (2007).

HPLC-based pigment concentrations and CHEMTAX-based diatom fraction were determined in the same manner as described above for MVCO.

3. Results

3.1. Defining diatom dominance

As our validation of the model output depends on the identification of diatom-dominated waters, the method used to determine phytoplankton taxonomy could significantly impact the model validation. The gold standard for quantifying phytoplankton taxonomic composition is microscopy and cell imaging. However, optically-based models such as S04 rely on light absorption features that arise due to both the taxonomically distinct pigment composition between phytoplankton groups and the degree of pigment packaging in cells. It may therefore be preferable to validate a bio-optical model with an optical proxy for phytoplankton taxonomy: for the S04 model, an appropriate proxy is the shape of the phytoplankton absorption spectra in response to diatom-dominated communities versus mixed phytoplankton communities (Fig. S2). These modeled spectra should capture the predominance of fucoxanthin as a biomarker pigment to indicate the presence or absence of diatoms, as well as the large cell size and high chlorophyll-to-carbon ratio of typical diatom cells, which leads to a higher degree of pigment packaging and thus a flattened absorption spectrum. As phytoplankton pigments are the agents that drive absorption, they are the first proxy for absorption. Pigments are also taxonomically distinct to first order and serve as proxies for taxonomy. Thus, pigments provide the analytic link between the modeled optical property (absorption) and the measured phytoplankton taxonomy (imaging).

We compared three pigment proxies for diatom fraction to the carbon-based estimate from FCB and IFCB at MVCO. The first pigment proxy is a simple fucoxanthin-to-chlorophyll-*a* pigment ratio (Fig. 4A). The second pigment proxy is the one from the S04 method, which considers both the ratio of fucoxanthin-to-chlorophyll-c_3 and

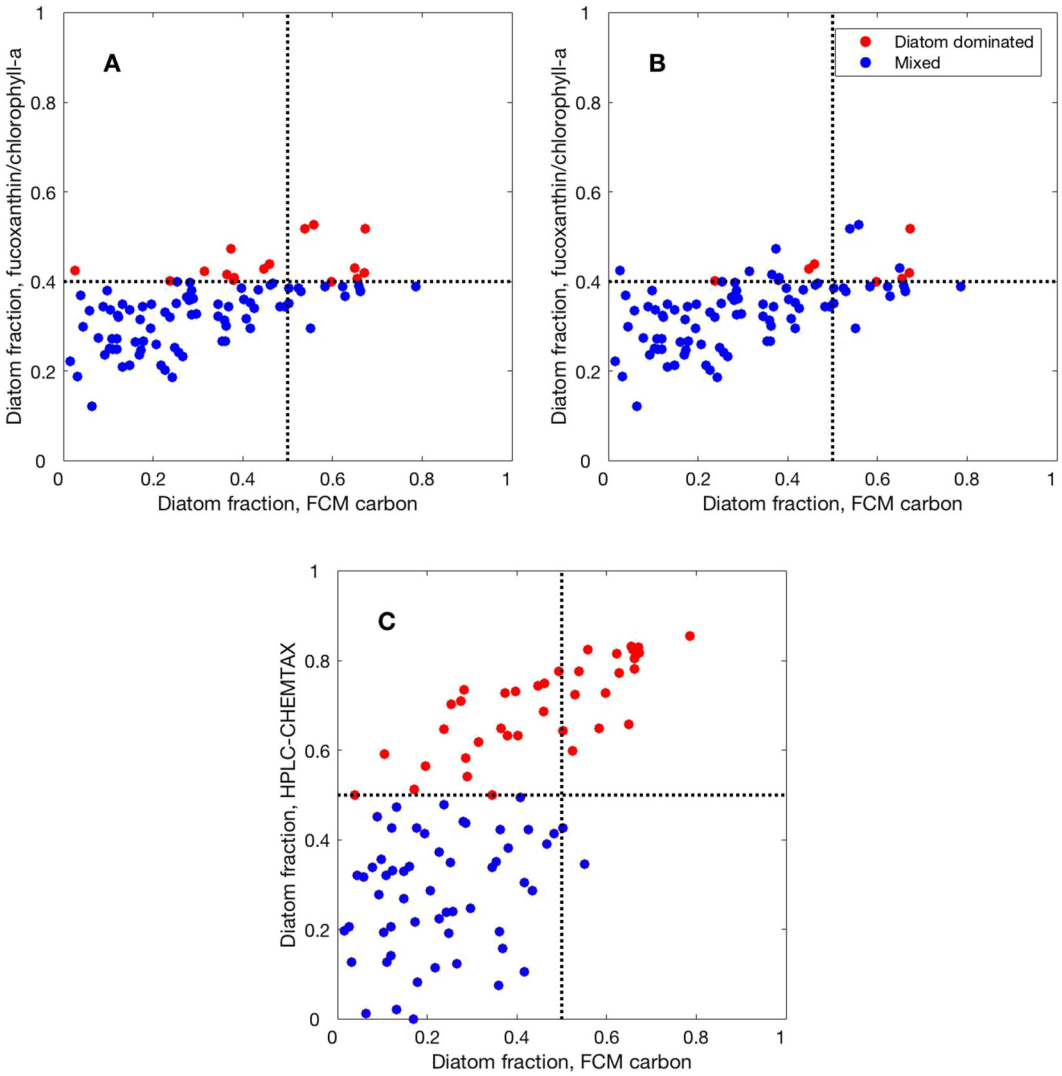

Fig. 4. Comparisons between pigment-based metrics for diatom fraction and the carbon-based estimate from measurements of phytoplankton at MVCO. (A) Fucoxanthin-to-chlorophyll-*a* ratio where fuco:chl > 0.4 is defined as diatom dominated, in red (16 points); fuco:chl < 0.4 is mixed, in blue (78 points). (B) Fucoxanthin-to-chlorophyll-*a* ratio where the S04 metric is applied: only points with fuco:chl > 0.4 and chlc3:chl < 0.02 are defined as diatom dominated (red; 7 points) and all other points are mixed (blue; 87 points). (C) HPLC CHEMTAX where > 50% diatoms determined from CHEMTAX are in red (38 points), ≤50% diatoms are in blue (56 points). Dashed lines show the threshold of diatom dominance in each case. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

chlorophyll-c_3-to-chlorophyll-*a* to help separate diatoms from prymnesiophytes, which also contain fucoxanthin (Fig. 4B). The third proxy uses a full suite of pigments through the CHEMTAX inversion (Fig. 4C).

This comparison between pigment proxies and the FCB/IFCB data shows that while the fucoxanthin to chlorophyll ratio proxies exhibit a linear relationship to the FCB/IFCB diatom carbon estimates with smaller spread in the data, CHEMTAX correctly identifies the most diatom-dominated points that were also identified as > 50% diatoms by flow cytometry and imaging. Thus, we chose to use CHEMTAX as our method of diatom identification for model validation. There are still a number of points that are identified as diatom-dominated by CHEMTAX but are < 50% diatoms by the fraction of phytoplankton carbon (Fig. 4C). For the purposes of this reflectance-based model, it was preferable to use a pigment-based proxy (i.e., pigments) to identify diatoms for validation. However, the discrepancy between pigment-based methods and FCB/IFCB data suggests caution when choosing a method for phytoplankton community identification, as well as a source of potential uncertainty in the process of model validation. For more consideration of these issues, please see the Discussion.

3.2. Original model performance

To test the original S04 model performance at MVCO, we first reconstructed the model as described explicitly in Sathyendranath et al. (2004) with the reflectance ratio curves for R510:R555 and R490:R670 computed as a function of chlorophyll (Supplementary material, Section S1) for diatom-dominated or mixed phytoplankton populations (i.e., phytoplankton populations dominated by a group other than diatoms or with no dominant group). We then modified the S04 model at the wavelengths measured at MVCO and constructed reflectance ratio curves for R490:R555 and R442:R670 (Fig. 3). Next, we compared these curves to observations at MVCO. For each of the MVCO matchup data points, we plotted the observed reflectance ratios against HPLC-determined chlorophyll-*a* concentration with the points color-coded according to the CHEMTAX-determined diatom fraction (Fig. 5); this approach acknowledges a continuum of diatom fraction rather than simply designating a point “diatom-dominated” or “mixed” (as was done in S04). This method was chosen to elucidate trends in the model that might be concealed by the binary approach. However, it does not preclude binary analysis as values greater than or less than 50%

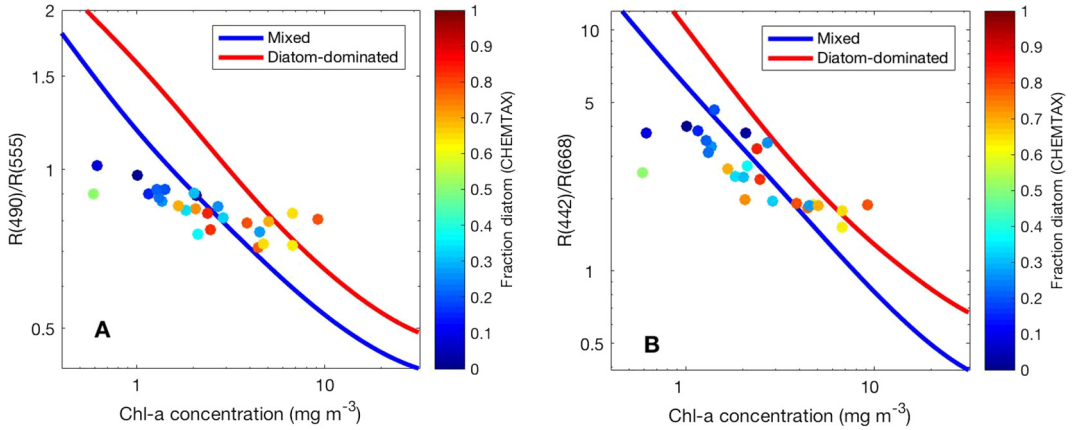

Fig. 5. Relationships between remote sensing reflectance ratios (A) R(490)/R(555) and (B) R(443)/R(668) and chlorophyll-*a* concentration. The original S04 model results for diatom-dominated and mixed communities are shown as red and blue lines, respectively. Observed values obtained from matchups at MVCO (26 total points) are shown by symbols colored coded by CHEMTAX-determined diatom fraction. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

diatoms (green symbols) are easily identified.

We quantitatively compared the difference between the measured data and the model curves by considering a continuum of diatom dominance between the two model curves, with the blue mixed curve representing the reflectance ratio associated with 0% diatoms and the red diatom-dominated curve representing the reflectance ratio associated with 100% diatoms. Each validation data point has a measured chlorophyll concentration, a measured reflectance ratio, and an associated diatom fraction determined from CHEMTAX. These data were used to linearly interpolate between the two extreme curves to find the corresponding reflectance ratio value at a given diatom fraction and chlorophyll concentration. Finally, the predicted reflectance ratio value was compared to the measured reflectance ratio value in the form of the root mean square deviation (RMSD):

$$\mathrm{RMSD} = \sqrt{\frac{\sum_{i=1}^{n} (R_{obs} - R_{pred})^2}{n}} \quad (1)$$

where n is the number of samples in the validation dataset, R_{obs} is the observed reflectance ratio value, and R_{pred} is the predicted reflectance ratio value. The RMSD was calculated for both reflectance ratios (Table 1). We chose to express the validation statistic in terms of the difference between measured and modeled reflectance ratio values rather than measured and modeled diatom fractions as the reflectance value was measured directly and the associated diatom fraction was modeled using CHEMTAX. However, as this work is an exercise in validation, we feel it is important to note that there are associated uncertainties with both the measured data (see Supplementary material, Section S6, Table S3) and the models used to describe the data.

The modified S04 model (i.e., different wavelengths) overestimates the magnitude of the reflectance ratios at low chlorophyll concentrations and underestimates the magnitude at high chlorophyll concentrations. This result indicates that the water at MVCO is less blue (or greener or redder with the same blue reflectance) than the S04 model predicted at low chlorophyll concentrations (Table 1), implying stronger absorption and/or weaker scattering of blue light at MVCO than predicted by the model.

To investigate the source or sources of optical variations between the MVCO and S04 relationships between reflectance ratios and chlorophyll-*a*, we analyzed each term in the model equations (Eqs. (S4)–(S13)) in comparison with the extensive MVCO validation dataset. The contributions of Raman scattering (Eqs. (S14) and (S15)) to reflectance are well understood and can be described with a robust analytical function (Marshall and Smith, 1990). In addition, the contribution of Raman scattering to total reflectance is on the order of 2–10% (Westberry et al., 2013). Thus, we focused on rebuilding the two essential components of the elastic reflectance term: absorption and backscattering coefficients.

Table 1
Root mean squared deviations (RMSD) between K18 and S04 model results and MVCO measurements shown in Fig. 10. The "Model adjustment" column refers to the term that was altered for that model run: for instance, in the "a_{ph}" column, only the phytoplankton absorption term was changed from the original S04 model; all other terms remained the same.

Model adjustment	RMSD 490:555	Percent change from S04	RMSD 442:668	Percent change from S04
S04 model unchanged	0.2606	N/A	2.7581	N/A
a_{ph}	0.1765	32.29	1.9325	29.93
a_y	0.0828	68.22	0.7943	71.20
High chl b_{bp}	0.1749	32.91	1.3707	50.30
Low chl b_{bp}	0.1619	37.89	1.2765	53.72
a_{ph} + a_y	0.1031	60.42	0.8419	69.47
a_{ph} + high chl b_{bp}	0.1148	55.97	1.0195	63.04
a_{ph} + low chl b_{bp}	0.1078	58.63	1.0027	63.60
a_y + high chl b_{bp}	0.1103	57.68	1.1741	57.43
a_y + low chl b_{bp}	0.1161	55.45	1.2853	53.40
Sum − high chl	0.1349	48.26	1.2287	55.45
Sum − low chl	0.1408	45.97	1.3364	51.55

3.3. Phytoplankton absorption

The difference in the two reflectance ratio curves generated by the S04 algorithm relies on the difference in magnitude between a_{ph} for waters dominated by diatoms compared to waters with mixed phytoplankton communities (Figs. S2; 6B and D). At MVCO, the mean measured a_{ph} for diatom-dominated (> 50% diatoms from CHEMTAX) and mixed community conditions separate by both magnitude and spectral shape (Fig. 6A and C). The mean measured spectra at MVCO are significantly different from each other (two-sample *t*-test, $p < 0.01$).

At MVCO, absorption per chlorophyll is lower for diatoms than for mixed phytoplankton taxa by about a factor of two, depending on wavelength (Fig. 6A). The spectral shape also varies (Fig. 6C): while the absorption spectra are quite similar in the green wavelengths, where absorption is minimum, the spectra are significantly different in the blue wavelengths (400–475 nm; two sample *t*-test, $p < 0.001$) and in the red wavelengths (600–700 nm; two sample *t*-test, $p < 0.01$). Mixed phytoplankton communities absorb much more in the blue relative to the red. The differences between the shapes of the curves can be exploited to determine community structure with optical methods, as

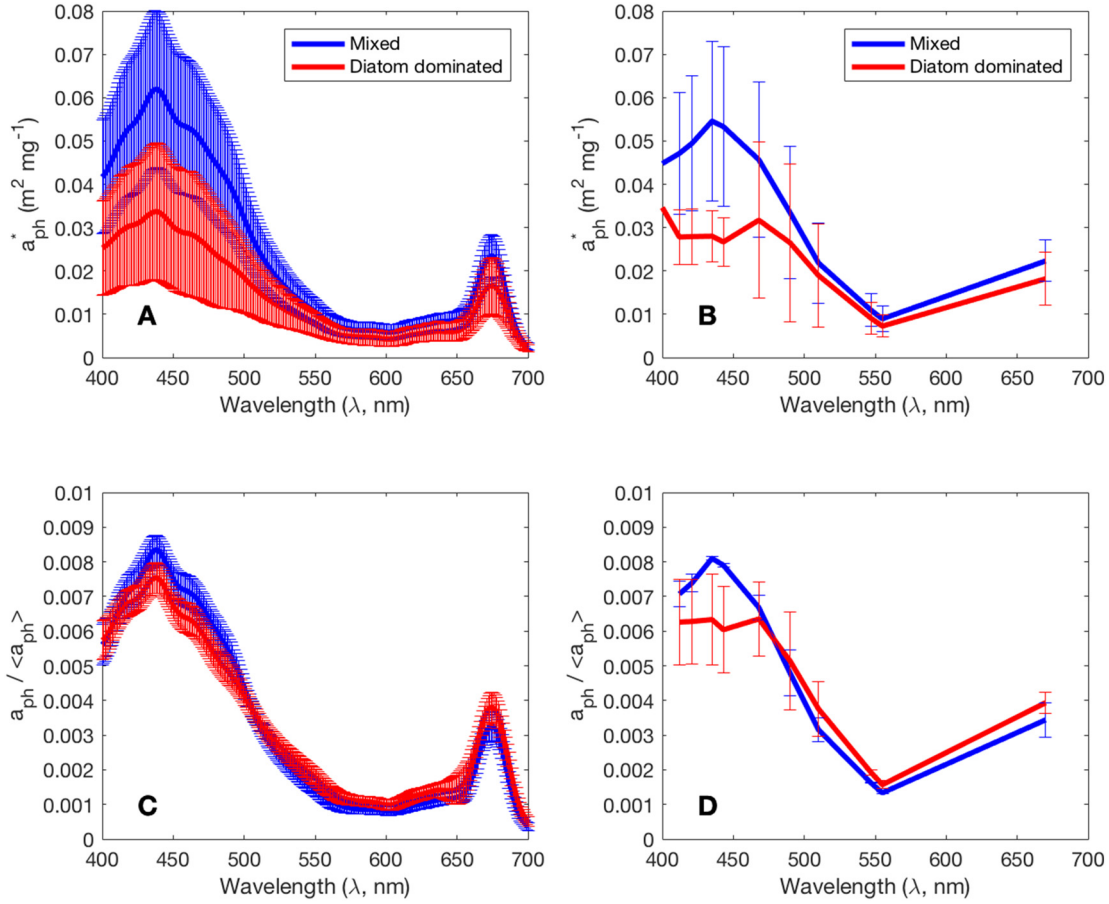

Fig. 6. Mean chlorophyll-specific phytoplankton absorption (a_{ph}^*; phytoplankton absorption normalized to chlorophyll-*a* concentration) for mixed (blue) and diatom-dominated (red) phytoplankton communities; (A) measured at MVCO; (B) modeled as in S04. Mean a_{ph} normalized to the area under the spectrum ($< a_{ph} >$) for mixed and diatom-dominated phytoplankton; (C) measured at MVCO; (D) modeled as in S04. Error bars show ± one standard deviation from the mean for the range of chlorophyll concentrations measured at MVCO. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

proposed by Sathyendranath et al. (2004), as reflectance will vary both in magnitude and spectral shape depending upon both the concentration and composition of the phytoplankton community (Fig. S2; Prieur and Sathyendranath, 1981; Roesler and Perry, 1995; Bricaud et al., 1995).

Notably, the mean measured chlorophyll-specific phytoplankton absorption spectra (a_{ph}^*) at MVCO for both diatom-dominated and mixed phytoplankton at MVCO were significantly different from the S04 modeled spectra (Fig. 6; two-sample *t*-test: mixed $p < 0.05$, diatom $p < 0.01$). The S04 model parameterizes the phytoplankton absorption coefficient as a function of chlorophyll but assumes empirical relationships that do not exist at MVCO (Eq. (S5); Fig. S2). We thus compared the results of optimization for multiple models for phytoplankton absorption, including a re-parameterization of the S04 model and the Ciotti et al. (2002) model. Ultimately, the best fit was the Bricaud et al. (1995, 2004) model for phytoplankton absorption, which uses wavelength-specific power-law relationships between absorption-to-chlorophyll to explain variations in pigment packaging and capture biomass-dependent variations in the spectral shape of the phytoplankton absorption coefficient over a global range of chlorophyll concentrations.

The Bricaud et al. (1995, 2004) model has been applied with success in coastal regions, including the Northwest Coastal Shelf province (Devred et al., 2006). As with the S04 model, phytoplankton absorption is chlorophyll-dependent but features two wavelength-dependent parameters defined over the spectrum of 400 nm to 700 nm. One parameter, *A*, defines the magnitude of the specific absorption and the other parameter, *B*, provides the non-linear dependence on chlorophyll concentration (reflecting impact factors such as variation in pigment packaging and accessory pigments). We use the Bricaud et al. (1995) model for the chlorophyll-specific absorption coefficient (a_{ph}^*):

$$a_{ph}^* = A * Chl^B \quad (2)$$

where a_{ph}^* has units of $m^2\,mg^{-1}$ chl-*a* and the parameters *A* and *B* are unitless. *A* and *B* were quantified by a non-linear least-squares optimization approach to match the shape of the measured spectra at MVCO (mixed and diatom-dominated *A* and *B* values are given in Table S2). The resulting re-formulated a_{ph}^* spectra following the approach of Bricaud et al. (1995, 2004) were statistically identical to the mean measured spectra (two-sample *t*-test: mixed $p \gg 0.999$, diatom $p = 1$). We compared the measured relationship between the phytoplankton absorption coefficient at 442 nm and chlorophyll-*a* to both the S04 model and the Bricaud et al. (1995) model: the Bricaud et al. (1995) model resulted in a better fit with the measured data (mixed $r^2 = 0.85$, diatom $r^2 = 0.75$, both $p \gg 0.999$; Fig. 7). Thus, these re-formulated a_{ph} coefficients were incorporated into the K18 model.

3.4. Absorption by CDOM and non-algal particles

The second component we investigated in the S04 model is the absorption by yellow substances (CDOM plus non-algal particles or NAP). The S04 model produces two a_{ph} coefficients (diatom-dominated and mixed) and therefore two associated a_y coefficients (see Supplementary material, Section S1 for details). Thus, for the same

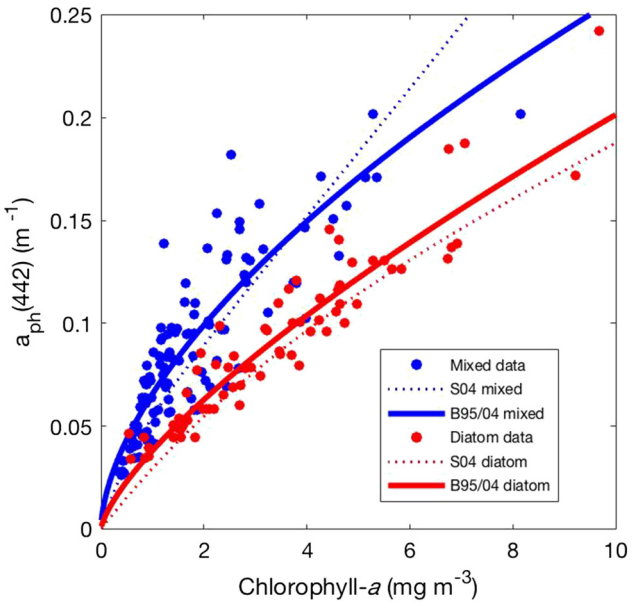

Fig. 7. Relationship between a_{ph} at 442 nm and chlorophyll-*a* concentration for diatom-dominated populations (red) and mixed phytoplankton (blue) at MVCO. Dashed lines show the S04 model; solid lines show the Bricaud et al. (1995, 2004) model parameterized to fit the MVCO observations. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

chlorophyll concentration, the model predicts different concentrations of dissolved material in the water depending on the phytoplankton community. This assumption implies that mixed populations have associations with greater CDOM concentrations compared to diatoms. Ecologically, diatoms dominate earlier in the seasonal succession and mixed populations dominate later, when more grazing and trophic transfer may enhance CDOM.

Absorption by CDOM and NAP were measured separately at MVCO, which allows the variability of the sum or the components (dissolved and particulate) to be examined. In keeping with S04, we first considered the two components together ($a_y = a_{CDOM} + a_{NAP}$) as a function of chlorophyll: the measured values were significantly different from the S04 modeled absorption curves (two-sample *t*-test: mixed < 0.01, diatom $p < 0.01$). The lack of correlation between a_y and chlorophyll-*a* is best understood by examining the two components independently. Across all points at MVCO, a_{CDOM} does not exhibit significant correlation with chlorophyll concentration (Fig. 8A). Most of the a_{CDOM} at 440 nm observations fall within a narrow range of values; thus, for the K18 model, we selected the mean value of 0.13 m^{-1} ($\pm$ 0.05) to represent the a_{CDOM} coefficient at 440 nm (following Eq. (S6)).

In contrast, we found that a_{NAP} could be described by a linear function with chlorophyll concentration (Fig. 8B), albeit with substantial unexplained variance. As a first order description, we fit two linear regressions (mixed $r^2 = 0.28$, $p < 0.001$; diatom $r^2 = 0.13$, $p < 0.01$):

$$a_{NAP}(440)_{Mixed} = 0.009{*}Chl + 0.033 \quad (3)$$

$$a_{NAP}(440)_{Diatom} = 0.0074{*}Chl + 0.062 \quad (4)$$

We combined the mean a_{CDOM} value at 440 nm with a_{NAP} at 440 nm from Eqs. (3) and (4) to provide the reference value (at wavelength of 440 nm; Eq. (S7)) needed for the exponential a_y model (Eq. (S6)):

$$a_y(\lambda)_{Mixed} = [a_{CDOM}(440) + a_{NAP}(440)_{Mixed}]{*}e^{[-0.014(\lambda-440)]} \quad (5)$$

$$a_y(\lambda)_{Diatom} = [a_{CDOM}(440) + a_{NAP}(440)_{Diatom}]{*}e^{[-0.014(\lambda-440)]} \quad (6)$$

The slope of the exponential decay (-0.014) is consistent with the observed values. These final re-parameterized a_y coefficients were incorporated into the K18 model.

3.5. Particle backscattering

Particle backscattering at 660 nm, $b_{bp}(660)$, and chlorophyll-*a* concentration are uncorrelated ($r^2 = 0.10$) in the MVCO dataset. Separating the points based on phytoplankton community structure did not change this result (mixed $r^2 = 0.12$, diatom $r^2 = 0.01$), nor did the wavelength of backscattering (Fig. 9A). Furthermore, b_{bp} at MVCO exhibited significantly larger values than predicted by the Sathyendranath et al. (2004) b_{bp} model used in S04 (Eq. (S10)). That model appears to represent a lower limit to the MVCO data (Fig. 9A).

To better understand the variability in the magnitude of b_{bp} at MVCO, we examined the relationships between $b_{bp}(550)$ and other measured physical and biological properties, including day of the year and dominant taxa. We found that community structure and season explained little of the variability in measured b_{bp}, which we thought was likely due to resuspension. However, we were not able to explain the variability in b_{bp} with resuspension, as b_{bp} was uncorrelated with tidal phase or season (i.e., winter storms). On scales of minutes to years, the magnitude of b_{bp} at MVCO at any given wavelength varies by a factor of two to four. Notably, the variation in backscattering observed over short timescales of minutes to hours is on the same order as seasonal differences, implying a high degree of patchiness. Over seasonal timescales, high phytoplankton biomass is generally associated with higher backscattering but with a large uncertainty.

In finding no clear driver for backscattering based upon phytoplankton composition or physical dynamics, we explored a driver based on other biological factors. We noted that b_{bp} and chlorophyll appear temporally lagged. If we consider that particulate matter is generated mainly by phytoplankton blooms at MVCO, backscattering by particulate matter may be out of phase with chlorophyll-*a*, which is used as a proxy for phytoplankton biomass. While chlorophyll-*a* is related to phytoplankton biomass, b_{bp} depends on all particle types, many of which are likely generated during bloom decline or associated with phenomena such as resuspension of senescent blooms.

We used the ratio of chlorophyll to phaeophytin (measured with HPLC) as an ecological proxy for bloom stage to explore the temporal mismatch between particle backscattering and chlorophyll concentration. Phaeophytin is a degradation product of phytoplankton pigments, generally found in higher concentrations in nature as a phytoplankton bloom comes to its end (Jensen and Sakshuag, 1973). At high chlorophyll-to-phaeophytin ratios, there are little to no phytoplankton degradation products in the water (i.e., a healthy bloom). Most of the particles in the water are living phytoplankton with high chlorophyll that absorb more light and scatter little. At lower chlorophyll-to-phaeophytin ratios, there are more phytoplankton degradation products in the water, likely associated with grazing and cell death (i.e., a bloom that is ending). The particles in the water include more dead and decaying phytoplankton cells, fecal pellets, and heterotrophic bacteria with low absorption, but high scattering. Ultimately, we found that a chlorophyll-to-phaeophytin ratio criterion helped to explain some of the variance in the measured $b_{bp}(550)$ vs. chlorophyll relationship (Fig. 9B).

We determined two phytoplankton bloom states (high chlorophyll-to-phaeophytin, Hchl, and low chlorophyll-to-phaeophytin, Lchl) and represented the $b_{bp}(550)$ vs. chlorophyll relationship for each condition with a linear regression (Hchl $r^2 = 0.60$; Lchl $r^2 = 0.30$):

$$b_{bp}^{HC}(550) = 0.0020{*}Chl + 0.0066 \quad (7)$$

$$b_{bp}^{LC}(550) = 0.0073{*}Chl + 0.0058 \quad (8)$$

This method offers an approach appropriate for forward modeling of remote sensing reflectance ratios as a function of chlorophyll, but inverse applications of the S04 and K18 models are also desirable in cases

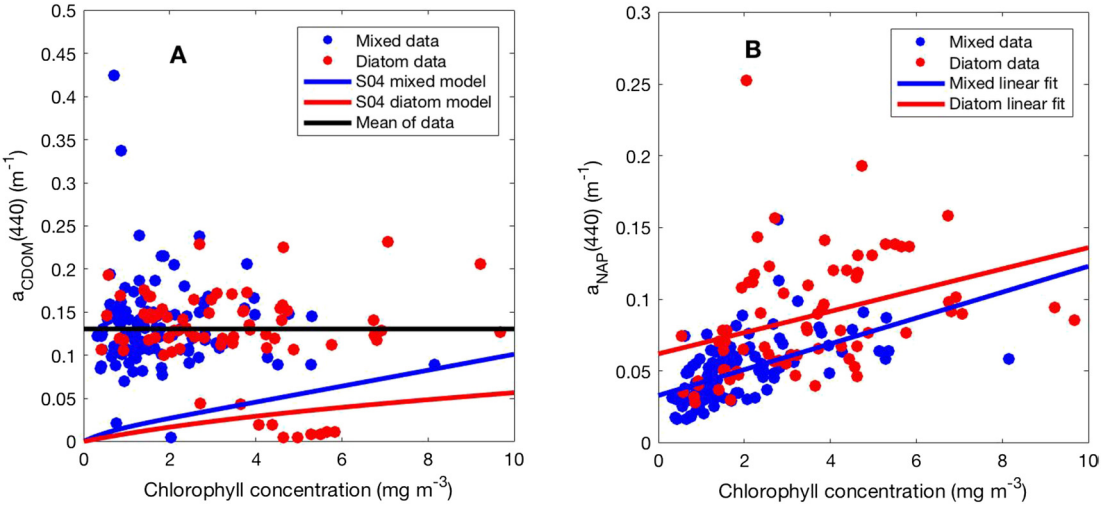

Fig. 8. (A) Relationship between a_{CDOM} at 440 nm and chlorophyll-*a* concentration at MVCO for mixed phytoplankton (blue) and diatom-dominated phytoplankton populations (red). Thick colored lines show the functional form proposed by S04. Thick black line indicates the mean value of the observations. (B) Relationship between a_{NAP} at 440 nm and chlorophyll-*a* concentration at MVCO. Solid lines show least squares linear regression results. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

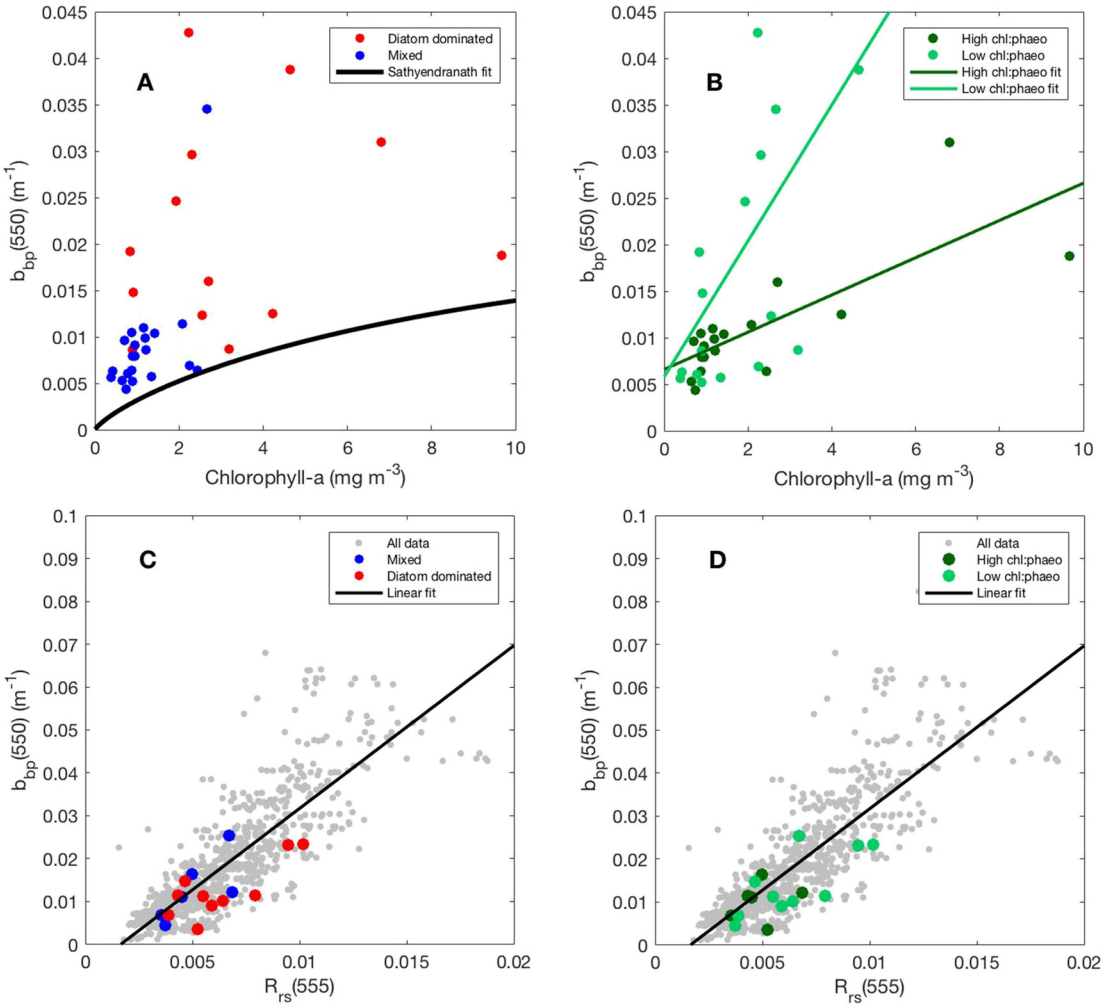

Fig. 9. (A) Relationship between chlorophyll-*a* concentration and b_{bp} at 550 nm at MVCO. Colors indicate mixed phytoplankton populations (blue) and diatom-dominated phytoplankton populations (red) determined from HPLC CHEMTAX. Thick black line represents the model fit between $b_{bp}(550)$ and chlorophyll given by Sathyendranath et al. (2001). (B) Point colors indicate chlorophyll-to-phaeophytin ratio (high chl:phaeo in dark green, low chl:phaeo in light green). Solid lines represent linear fit. (C) Relationship between remote sensing reflectance (R_{rs}) at 555 nm and $b_{bp}(550)$ at MVCO. Colors indicate mixed (blue) and diatom-dominated (red) phytoplankton populations. (D) Colors indicate chlorophyll-to-phaeophytin ratio (high chl:phaeo in dark green, low chl:phaeo in light green). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

where no chlorophyll-to-phaeophytin ratio is available to select the appropriate ecosystem state. When an inverse modeling approach is applied in the case of an input remote sensing reflectance spectrum, our observations suggest that b_{bp} magnitude can be derived directly from the relationship with remote sensing reflectance. Specifically, at MVCO we find (Fig. 9C and D):

$$b_{bp}(550) = 3.8 * R_{rs}(555) - 0.0063 \quad (9)$$

For the mean spectral slope parameter for particle backscattering, η, we found a significant difference between diatom-dominated and mixed conditions (two-sample t-test, $p < 0.01$). Optimized non-linear least-squares regression yielded values of 0 for diatom-dominated (i.e., spectrally flat backscattering), suggesting mostly large particles and larger phytoplankton (i.e., diatoms) (Whitmire et al., 2010). The spectral slope for mixed conditions was highly variable but always negative, consistent with smaller particles and smaller phytoplankton (Whitmire et al., 2010). The mean slope was -0.57 ± 0.28, slightly flatter than the global median reported by Boss et al. (2013).

Our reformulated b_{bp} model encompasses two ecosystem states with two phytoplankton community structures. The magnitude of $b_{bp}(550)$ depends on the ecosystem state: high chlorophyll-to-phaeophytin (onset of bloom) and low chlorophyll-to-phaeophytin (decline of bloom). The slope of b_{bp}, η, depends on the phytoplankton community:

$$b_{bp,Mixed} = b_{bp}(550) * \left[\frac{550}{\lambda}\right]^{-0.57} \quad (10)$$

$$b_{bp,Diatom} = b_{bp}(550) * \left[\frac{550}{\lambda}\right]^{0} \quad (11)$$

These final re-parameterized and re-formulated b_{bp} terms were input into the K18 model.

3.6. Applying the K18 model at MVCO

We investigated the degree to which re-parameterization and re-formulation of each absorption and backscattering term affected the fit of the K18 model to the relationships between reflectance ratios and chlorophyll observed at MVCO. We did this by changing one term at a time and maintaining the other terms unchanged from S04. However, because the a_y term depends on the a_{ph} term in S04, re-formulating only the a_{ph} term also introduces variations in the a_y term. We performed the analysis for both ecosystem states (high chlorophyll-to-phaeophytin conditions and low chlorophyll-to-phaeophytin) and phytoplankton community states (Fig. 10).

For chlorophyll concentrations $< 5\,mg\,m^{-3}$ (where mixed phytoplankton taxa were typically observed at MVCO), the re-parameterized and re-formulated models for phytoplankton absorption and backscattering are all significantly different from each other with the exception of the two backscattering models which are not significantly different form each other (two-sample t-test; $p < 0.01$ for all cases, $p = 0.77$ for backscattering models). Generally, these refined models exhibited a similar slope to S04 but with slightly lower reflectance ratios. However, despite the significant differences between the refined model curves, these new models did not describe the MVCO observations significantly better than S04 (Fig. 10A and B). The re-parameterized models for phytoplankton absorption and backscattering in the diatom-dominated cases also did not result in significantly different model shapes compared to S04 and thus did not perform significantly better than S04 in describing the MVCO observations (Fig. 10C and D). However, for both mixed populations and diatom-dominated populations, the re-parameterized a_y model resulted in a change in slope that reflects the distribution of observations better than any of the other cases (Fig. 10E and F; Table 1). The re-parameterized a_y model is significantly different from S04 and better describes the relationship in the MVCO observations for both mixed and diatom-dominated communities, specifically at low to mid chlorophyll values ($\leq 8\,mg\,m^{-3}$).

Next, we investigated whether additive re-parameterization yielded significant improvements to model fit over the single a_y improvement by considering each combination of paired re-parameterized models and the sum of all. All combinations yielded better fits to the MVCO observations compared to the original S04 parameterizations (Table 1), with RMSD reductions exceeding 30% and approaching 60% in some cases. However, no combination yielded better agreement with observations re-parameterizing the a_y model alone.

3.7. Applying the K18 model in Harpswell Sound

A model cannot be validated with the data used to construct that model. Thus, although the K18 model was a better fit with the data measured at MVCO, the generalization of the K18 model can only be assessed by validation with an independent dataset. We compared both the S04 model and the K18 model to data collected at the Bowdoin Buoy, a site in Harpswell Sound in Casco Bay in the eastern Gulf of Maine that, like MVCO, lies within the Northwest Coastal Shelf province, but is much more coastal than MVCO with higher and less variable CDOM coefficients. While we knew that the original S04 model would likely not fit the Harpswell Sound data, which lies at the edge of the Case 1 range (Fig. 1B), we included the S04 model in our comparison as a baseline for the performance of the K18 model at a different site.

While MVCO and Harpswell Sound have some general ecological and oceanographic conditions in common, the optical and taxonomic data show some differences. The data measured in Harpswell Sound have a narrower range of chlorophyll-a concentrations, as the data in Harpswell Sound were only collected during the late spring to summer season while the data at MVCO were collected year-round and thus have a larger dynamic range (Supplementary material, Section S4; Fig. S3). The Harpswell Sound data also have lower overall reflectance ratios than the MVCO data (associated principally with more absorption by CDOM) and are clearly outside of the range of MVCO and other Case 1 sites (Fig. 1B). While the data at MVCO show a trend in reflectance ratio as a function of phytoplankton composition, the data at Harpswell Sound do not show the same trend in diatom dominance with increasing chlorophyll concentration (Fig. 11).

For both diatom-dominated and mixed phytoplankton populations, each re-parameterization in the K18 model describes the Harpswell Sound observations significantly better than S04, as the K18 model re-parameterizations account for higher absorption (by a_y) and variable scattering, which changes the slope of the modeled relationships to better fit the measured Harpswell Sound data (Fig. 11). The best fit version of the K18 model at Harpswell Sound was again the model with only a re-parameterized a_y term. While the K18 model was tuned with the absorption and backscattering coefficients measured at MVCO, the resulting model represents substantial improvement in fit at Harpswell Sound as well (Table 2).

Although the observations in Harpswell Sound are different from the observations at MVCO (Fig. 1), every term in the K18 model that was re-parameterized to fit the conditions at MVCO fits the Harpswell Sound data significantly better than the original S04 model (Table 2).

4. Discussion

The S04 model as originally formulated does not accurately predict the observed variations in phytoplankton community structure at MVCO (Fig. 5). However, successive re-parameterization and re-formulation of the model components resulted in demonstrable improvements in fit with the K18 model (Table 1). The largest improvement in the model fit is with the revised a_y term. The IOP parameterizations of the S04 model have been indirectly evaluated in terms of AOP relationships (e.g., Sathyendranath and Platt (1988) for K_d and in Sathyendranath et al. (2001) for reflectance). Importantly, however, there have not previously been direct comparisons to observed CDOM

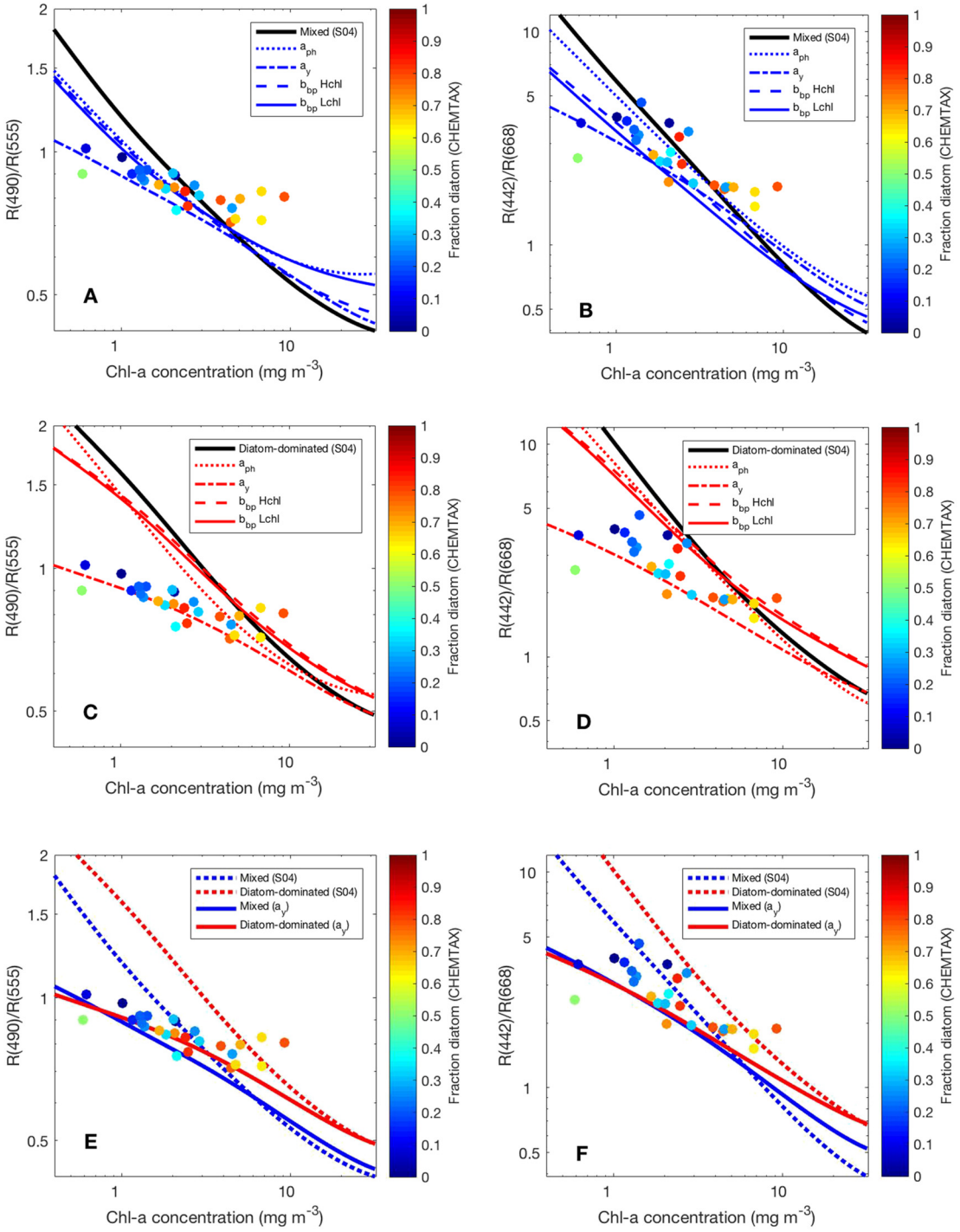

Fig. 10. Relationship between observed R(490)/R(555) (left panels) and R(443)/R(668) (right panels) and chlorophyll-*a* concentration at MVCO. Symbol colors indicate CHEMTAX-determined diatom fraction. The original S04 model for (A, B) diatom-dominated and (C, D) mixed populations are shown as black lines, along with new model lines where only the parameter indicated in the legend was changed from the published S04 model. (E, F) The best fit model to the MVCO observations was the re-parameterized a_y model; original S04 models shown in dotted lines for comparison.

and NAP absorption (Eqs. (S6) & (S7)). The restructuring of the functions describing CDOM and NAP absorption based on measurements at MVCO had a significant impact on the overall agreement between the modeled and observed relationships between reflectance ratios and chlorophyll-*a*. CDOM and NAP absorb much more strongly at blue wavelengths compared to green and red; thus, it is unsurprising that these re-parameterizations would alter the blue-to-green and blue-to-red ratios.

Across all points, there is no clear relationship between CDOM absorption and chlorophyll at MVCO (Fig. 8A), nor between particle backscattering and chlorophyll (Fig. 9A and B). This site does not experience regular resuspension or inputs from shore, except in the event of large storms; there is no river outflow or tidal connection to the patterns of CDOM absorption (Supplementary material, Section S5, Fig. S4). CDOM absorption (Fig. S4) and particle backscattering (Fig. 9A and B) both vary at times either as a function of chlorophyll and independent of variations in chlorophyll at MVCO, but the site is overall consistent with Case 1 conditions. However, the variability in these

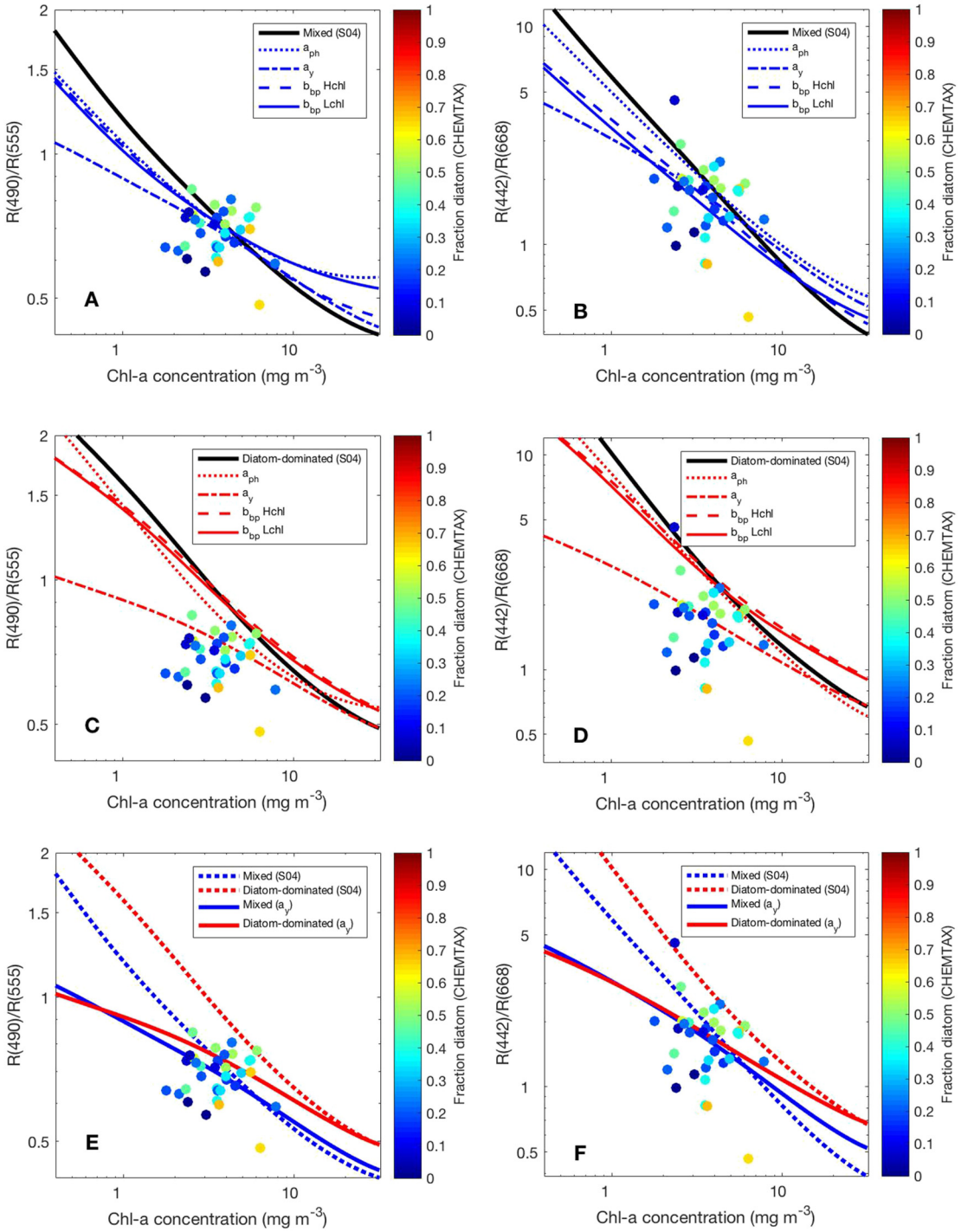

Fig. 11. Relationship between observed values of R(490)/R(555) (left panels) and R(443)/R(668) (right panels) and chlorophyll-*a* concentration at Harpswell Sound. Symbol colors indicate CHEMTAX-determined diatom fraction. The original S04 models for (A, B) diatom-dominated and (C, D) mixed populations are shown as black lines along with new model lines. (E, F) The best fit model to the Harpswell Sound observations was the re-parameterized a_y model; original S04 models shown in dotted lines for comparison.

relationships suggests that there are times when MVCO represents a system that would be considered Case 2 by Morel and Prieur (1977), when CDOM is supplied by a source other than phytoplankton and is high at low chlorophyll concentrations. At other times, MVCO would be considered Case 1 by Morel and Prieur (1977), when CDOM varies directly with chlorophyll concentration and phytoplankton outnumber other particles (Fig. 1A). While it would not be appropriate to apply the model results incorporating the revised a_y presented here to the entire Northwest Atlantic, which was the range of the original S04 model, we suggest that nearshore regions or regions with temporal variation in CDOM concentration will especially benefit from careful parameterization of this variable term.

Our results confirm the importance of local re-parameterization (and, where necessary, re-formulation) of empirical models and of quantifying which term has the greatest impact on model parameters. Compared to S04, CDOM and NAP at MVCO are significant contributors to the factor-of-two-reduced blue reflectance, and are responsible for more variations in the blue-to-green or blue-to-red reflectance ratios

Table 2
Root mean squared deviations (RMSD) between K18 and S04 model results and Harpswell Sound measurements shown in Fig. 11. The "Model adjustment" column refers to the term that was altered for that model run: for instance, in the "a_{ph}" column, only the phytoplankton absorption term was changed from the original S04 model; all other terms remained the same.

Model adjustment	RMSD 490:555	Percent change from S04	RMSD 442:668	Percent change from S04
S04 model unchanged	0.1637	N/A	1.0458	N/A
a_{ph}	0.1226	25.11	0.9351	10.59
a_y	0.0871	46.79	0.6859	34.41
High chl b_{bp}	0.1406	14.08	0.8381	19.86
Low chl b_{bp}	0.1365	16.60	0.7674	26.62
a_{ph} + a_y	0.0809	50.58	0.6835	34.64
a_{ph} + high chl b_{bp}	0.1049	35.92	0.7756	25.83
a_{ph} + low chl b_{bp}	0.1023	37.48	0.7288	30.32
a_y + high chl b_{bp}	0.0781	52.32	0.7158	31.55
a_y + low chl b_{bp}	0.0763	53.37	0.7586	27.46
Sum − high chl	0.0790	52.86	0.7260	30.58
Sum − low chl	0.0758	53.69	0.7719	26.19

than are variations in phytoplankton community structure. Application of the K18 model to a second independent validation dataset collected in Harpswell Sound showed that the new model improvements extended beyond MVCO. In fact, the RMSD are smaller for the Harpswell Sound data (Table 2) than for the MVCO data (Table 1) for every re-parameterization step. This result validates the effectiveness of the re-parameterized model terms and shows that they better capture the bio-optical complexity and variability not just at MVCO but also at another nearshore site in the same biogeographic province.

Sathyendranath et al. (2004) recommend model tuning to match local conditions and suggest that future model refinements will be necessary as the quality of observations and capabilities of technology improve. Our study shows that the S04 model was indeed greatly improved after re-parameterization and re-formulation to match the measured conditions at MVCO, a Case 1 site in the same biogeographic province for which the model was developed. Results with the revised a_y model demonstrate that the K18 model exhibits a better fit to the data, albeit with reduced distinction between the mixed and diatom-dominated model curves (Fig. 10E and F). As the goal of the S04 and K18 models is to identify diatom dominance from other phytoplankton community states, any reduction in the distinction between the two model curves complicates this process. While parameterizations of the component parts of the original model improved the K18 model fit with the measured data at both sites, the measured phytoplankton community structure data at MVCO and in Harpswell Sound emphasize a fundamental challenge with identifying phytoplankton composition from the relationship between reflectance ratios and chlorophyll-*a*.

This challenge can be easily summarized as the biomass effect on community composition. The differences between the diatom-dominated and mixed populations at MVCO have less to do with differences in reflectance ratios and more to do with biomass: as is common in nature, diatom-dominated conditions occur at higher chlorophyll-*a* concentrations while mixed communities occur at lower chlorophyll concentrations. While the PFT community defines the S04 algorithm as a spectral-response model, the model shows clear trends with changing biomass: when the original S04 algorithm was applied to satellite imagery, regions with high derived chlorophyll concentration (Fig. 12A) were generally predicted to be dominated by diatoms (Fig. 12B). The model rarely if ever identifies the opposite condition, when chlorophyll is high but mixed phytoplankton dominate (using the definition of "mixed" from Sathyendranath et al. (2004) to mean either no dominant group or dominated by a group other than diatoms, such as prymnesiophytes). Since this model uses the designation of diatom-dominated or mixed to determine the chlorophyll concentration, these two figures are not independent; rather, they show the difficulty of the model to separate community structure and biomass. The distinctions between curves based only on a spectral absorption signature between diatoms and mixed phytoplankton disappears when there are any other absorbing or scattering components of the water, even if the concentration of these components is not higher than phytoplankton (the definition of Case 1 waters).

Sathyendranath et al. (2004) computed chlorophyll-*a* concentrations as a function of community structure, using their algorithm to distinguish diatoms from other phytoplankton (Fig. 12A). The NASA chlorophyll-*a* algorithms are designed to derive chlorophyll concentration from remote sensing reflectance band ratios (O'Reilly et al., 1998). The resulting ocean color model curve fits the measured global data: however, there is substantial variance in the reflectance ratio that is not described by chlorophyll-*a*. The S04 model ascribes the variability in this relationship to differences in phytoplankton community structure (Fig. 1) when in fact it is controlled by a number of different inherent optical properties (Dierssen et al., 2006; Dierssen, 2010; Sauer et al., 2012). The S04 model predicts that for the same measured reflectance ratio, points with higher in situ chlorophyll-*a* concentration are associated with diatom dominance while lower chlorophyll-*a* conditions reflect mixed communities.

The observations at MVCO show a gradient of diatom concentration: the proportion of diatoms increases as chlorophyll concentration increases. However, the data are not separated along the S04 model curves but rather follow a much flatter slope for which relative invariance in the reflectance ratio is associated with an order of magnitude increase in chlorophyll concentration and increasing diatom dominance. Despite re-parameterization and re-formulation of the S04 model to create the K18 model, the measured data do not separate based on the magnitude of reflectance in a way that would allow a tuned version of this model to identify phytoplankton composition. The difference in the in situ chlorophyll-*a* concentrations associated with one measured reflectance ratio value at MVCO cannot be attributed only to phytoplankton community structure. Still, variations in particle backscattering, CDOM absorption, and NAP absorption at MVCO can significantly alter the reflectance ratio, in addition to changes in the chlorophyll concentration and phytoplankton community composition.

Despite the unprecedented spatiotemporal potential of satellite ocean color algorithms to characterize the phytoplankton community structure from remote sensing, algorithms should be implemented with caution and must be validated to quantify the conditions appropriate for application. A model is only as good as the information used to develop it: both the measured data and the validation data products ought to be robust and individually examined in relation to the model assumptions. The original S04 model provided the first framework for identifying diatom-dominated regions from satellite ocean color data; the K18 model builds on this approach to include re-parameterized and re-formulated input functions based upon measured data, application of relevant ecological and oceanographic theory, and further validation with a third independent dataset. Outside of the re-parameterization and re-formulation of the S04 model, this work suggests that validating the empirical relationships in all bio-optical models of dominant phytoplankton groups is a crucial step that cannot be overlooked and ought to be undertaken regularly as improved data products and methods become available.

The exercise of validation can be complicated by uncertainties. In this case, both general uncertainties associated with the methods used to evaluate and refine the S04 model (see Supplementary material, Section S6, Table S3) must be considered, along with the uncertainties associated with the method of taxonomic identification. The uncertainties associated with each measurement used for the algorithm development are relatively small compared to the magnitude of the measurements. As many of the components considered here are spectral, the associated uncertainty is wavelength dependent, but does not vary greatly across wavelengths over the set of measurements used. Furthermore, our selection of linear interpolation to distinguish the

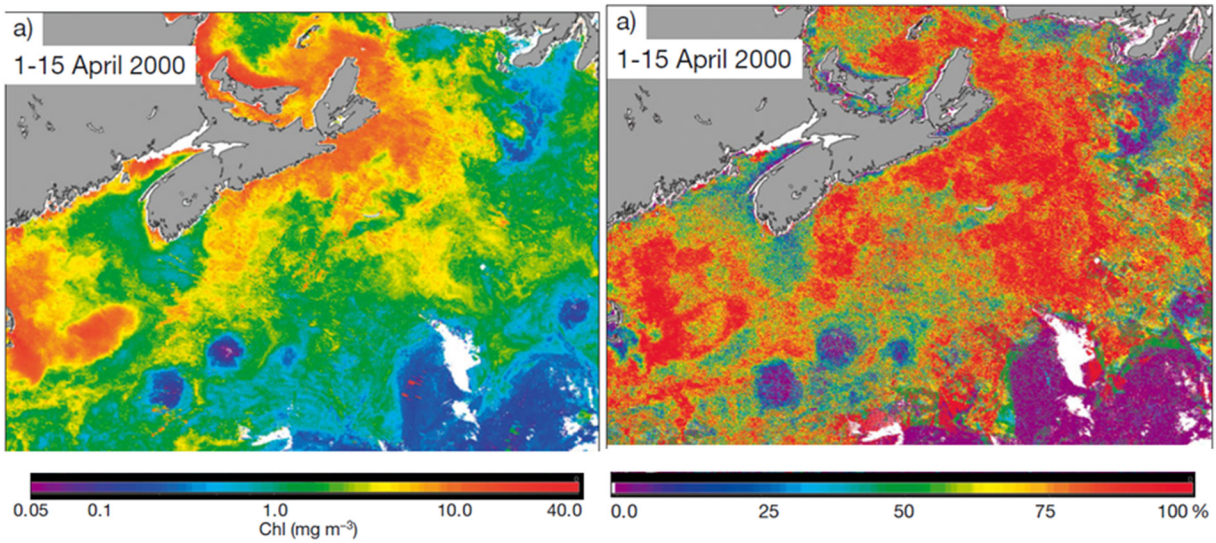

Fig. 12. (A) Composite image of satellite-derived chlorophyll concentration (Sathyendranath et al., 2004 Fig. 6a) and (B) probability that a phytoplankton population is dominated by diatoms (their Fig. 4a).

gradient of diatom dominance between the diatom and mixed phytoplankton community composition curves (Figs. 5, 10, 11) represents an assumption that the reflectance ratio changes linearly with community composition, or another source of potential uncertainty.

Given the relatively small contribution of measurement error to overall uncertainty in the algorithm testing, the main source of uncertainty in this work is the taxonomic identification of diatoms, in part because the choice of a taxonomic proxy for validation is not straightforward. It is critically important to consider the method used to determine taxonomy—in this specific model application, to confirm the presence or absence of diatoms. Taxonomic characterization of individual cells (as with cytometry and imaging) is the gold standard for quantifying the fraction of total phytoplankton biomass that is assigned to a target group such as diatoms. However, bio-optical models are based upon the absorption features of phytoplankton pigments, which are likely more closely related to the concentrations of those pigments and thus to the pigment-based taxonomic proxies. Thus, there is a second validation step to quantify the uncertainty between pigment- and image-based taxonomic composition.

CHEMTAX, the method used here to determine the fraction of diatoms in a sample, gave quite different results for diatom identification when compared to simpler pigment ratio methods. Of the three pigment-based identification methods, CHEMTAX identified the most diatom dominated points; this method also incorporates the most pigments in its determination of the presence or absence of diatoms in a sample. The S04 criteria, which involve both the fucoxanthin-to-chlorophyll-*a* ratio and the chlorophyll-c_3-to-chlorophyll-*a* ratio, identified the fewest diatom dominated points in the MVCO data. It removes many of the points that CHEMTAX identified to be diatom dominated by associating a fraction of the fucoxanthin in the sample with prymnesiophytes (Stuart et al., 2000; Sathyendranath et al., 2004). The relationship between the pigment-based diatom identification and flow-cytometry carbon-based diatom identification is not the same between the methods.

Based on these results, we recommend caution in selecting the tool used for identifying phytoplankton composition for model validation. The way that phytoplankton composition is defined and the methods that are used to determine phytoplankton community composition from in situ data can have a significant impact on the results of a validation study. For our validation method, we chose a pigment-based proxy due to the relationships between pigments and phytoplankton absorption and between absorption and reflectance. This reflectance ratio-based model aimed to identify diatom dominance, and here we have quantified the uncertainty in identifying diatoms from pigment-based approaches compared to flow-cytometric and imaging-based approaches. These results also suggest that there is a need to develop new metrics for phytoplankton communities and/or to strengthen methods of comparisons between existing metrics for community structure.

The reflectance ratio approach used in S04 to determine phytoplankton community composition relies heavily on pigment absorption features that lead to variations in the blue-to-green and blue-to-red reflectance signals. Variations in blue-to-green absorption are driven to first order by variations in phytoplankton biomass (Evers-King et al., 2014), but the presence or absence of other blue absorbing material (such as CDOM and NAP) additionally influences the blue-to-green absorption signal (Sauer et al., 2012). These other absorbing materials may not always co-vary with phytoplankton, even in Case 1 waters, due to ecological processes accompanying bloom onset and decline. Thus, taxonomic variations in accessory pigments are often a third order source of variability that may not be independent of the other two. At nearshore sites, phytoplankton community structure may be a third order source of variability more so than at offshore sites.

This variability in the reflectance-ratio-to-chlorophyll concentration relationship is captured in the spread of points in the NOMAD and MVCO datasets around the OC algorithm curve (Fig. 1A). While the reflectance ratio-to-chlorophyll relationships at MVCO fall within the range of other Case 1 sites (Fig. 1A), this location is quite coastal; unsurprisingly, the model fit was much improved when CDOM absorption was better constrained in the model (Table 1). The change in phytoplankton biomass concentration drives the ratio, but the variability around the curve is driven (to second and third order) by variations in phytoplankton community structure and in other particles and dissolved matter. Thus, although the reflectance-ratios-to-chlorophyll relationships at Harpswell Sound fall at the boundary of the range of Case 1 sites (marking it more typically Case 2), the K18 model with an improved CDOM and scattering term better fit the Harpswell Sound data (Table 2). Finally, the ecological patterns whereby phytoplankton taxonomy is not independent of biomass (i.e., blooms are dominated by diatoms while non-bloom conditions are mixed) further muddle the interpretation of blue-to-green reflectance variations, as was the case for MVCO. Any empirical or biomass-dependent model will be challenged to interpret taxonomy in the presence of these common (yet complex) optical regimes.

Bio-optical models are notoriously under-validated (IOCCG, 2014)

due to the scarcity of in situ datasets that include IOPs, AOPs, pigments, and microscopy. This analysis demonstrates the importance of using in situ data for both sufficient empirical model parameterization and re-formulation as well as complete model testing and validation. Even if a model is applied in the geographic region and optical regime for which it was designed, as in this study, our results show the need for careful stepwise validation before widespread implementation. We further demonstrate the difficulty of testing a model meant to discriminate phytoplankton taxonomy when the methods of validating phytoplankton taxonomic identification are also uncertain and poorly constrained. Thus, whenever possible, we suggest that independent validation should be undertaken with in situ bio-optical and taxonomic data products to uphold the assumptions inherent in the model development and compare the model output to measured data.

The applications of satellite ocean color models to identify the dominant phytoplankton group in the surface ocean naturally go beyond ecological surveys: for instance, several studies have implemented the S04 model in the Northwest Atlantic Zone to diagnose the probability of diatoms in the surface ocean and then use that information to examine broader concepts in ecology or trophic modeling (Platt et al., 2005; Son et al., 2007; Zhai et al., 2008; Platt et al., 2010; Trzcinski et al., 2013; Budge et al., 2014). Independent validation exercises such as the one undertaken here can only strengthen the robustness of these model applications to problems beyond simple classification of phytoplankton community structure. Going forward, the ability of satellite ocean color models to diagnose more complex problems such as the strength of the biological carbon pump will depend on the capacity of models to describe the phytoplankton community in the surface ocean. As we have shown here, that success is highly dependent on re-parameterization and re-formulation of existing satellite ocean color algorithms, and development of new models for determining phytoplankton community structure from satellites, as improved data products and methods become available.

Declarations of interest

None.

Acknowledgments

Many thanks to Taylor Crockford and Emily Peacock at WHOI and Sue Drapeau, Emily Kallin, and Luke Carberry at Bowdoin for their support with field experiments and data analysis. We gratefully acknowledge the dedication and hard work of the MVCO operations team, notably Janet Fredericks, Jay Sisson, Hugh Poponoe, and Steve Faluotico. We are indebted to Hui Feng for support of early years of the AERONET-OC observations at MVCO. Thank you to Kelsey Bisson and Dave Siegel for edits and comments. This work was started as part of a Summer Student Fellowship for S. Kramer in H. Sosik's lab at WHOI and completed in fulfillment of the Honors thesis research requirements at Bowdoin College with C. Roesler. We thank Shubha Sathyendranath, Bob Brewin, Áurea Ciotti, and an anonymous reviewer for helping us to significantly improve an earlier version of this manuscript.

Funding

This work was supported by: a Woods Hole Oceanographic Institution Summer Student Fellowship (NSF REU award #1156952) and a Bowdoin College Grua/O'Connell Research Award to SJK; grants to HMS from NASA (Ocean Biology and Biogeochemistry program and Biodiversity and Ecological Forecasting program), NSF (Ocean Sciences), the Gordon and Betty Moore Foundation, the Simons Foundation, and NOAA through the Cooperative Institute for the North Atlantic Region (CINAR) under Cooperative Agreement NA14OAR4320158; and grants to CSR from NASA (Ocean Biology and Biogeochemistry program). A Student Travel award from The Oceanography Society provided funds for SJK to present this work at the 2016 Ocean Sciences Meeting.

Notation

λ	wavelength, nm
λ_e	excitation wavelengths for Raman scattering, nm
$a(\lambda)$	total absorption coefficient, m^{-1}
$a_{CDOM}(\lambda)$	absorption by chromophoric dissolved organic matter (CDOM), m^{-1}
$a_{NAP}(\lambda)$	absorption by non-algal particles (NAP), m^{-1}
$a_{ph}(\lambda)$	absorption by phytoplankton, m^{-1}
$a_w(\lambda)$	absorption by pure seawater, m^{-1}
$a_y(\lambda)$	absorption by particulate and dissolved yellow substances (summed contributions of absorption by CDOM, a_{CDOM}, and absorption by non-algal particles, a_{NAP}), m^{-1}
a_{ph}^*	chlorophyll-specific phytoplankton absorption, m^2 mg chlorophyll-a^{-1}
$b_b(\lambda)$	total backscattering coefficient, m^{-1}
b_{bw}	backscattering by pure seawater, m^{-1}
$b_p(\lambda)$	particle scattering coefficient, m^{-1}
b_{bp}	particle backscattering coefficient, m^{-1}
Chl	concentration of chlorophyll-a per volume, mg chlorophyll-a m^{-3}
K_d	diffuse attenuation of downward irradiance, m^{-1}
$R_{rs}(\lambda)$	remote sensing reflectance, sr^{-1}

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.rse.2018.08.010.

References

Alvain, S.v., 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 (GB3001), 1–15. https://doi.org/10.1029/2007GB003154.

Anderson, T.R., 2005. Phytoplankton functional type modelling: Running before we can walk? J. Plankton Res. 27 (11), 1073–1081. https://doi.org/10.1093/plankt/fbi076.

Antoine, D., Babin, M., Berthon, J.-F., Bricaud, A., Gentili, B., Loisel, H., Maritorena, S., Stramski, D., 2014. Shedding light on the sea: André Morel's legacy to optical oceanography. Annu. Rev. Mar. Sci. 6, 1–21. https://doi.org/10.1146/annurev-marine-010213-135135.

Arrigo, K.R., 2005. Marine microorganisms and global nutrient cycles. Nature 437, 349–355. https://doi.org/10.1038/nature04158.

Behrenfeld, M.J., Falkowski, P.G., 1997. Photosynthetic rates derived from satellite-based chlorophyll concentration. Limnol. Oceanogr. 42 (1), 1–20. https://doi.org/10.4319/lo.1997.42.1.0001.

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, 323–341. https://doi.org/10.1357/002224089785076325.

Boss, E., Picheral, M., Leeuw, T., Chase, A.P., Karsenti, E., Gorsky, G., Taylor, L., Slade, W., Ras, J., Claustre, H., 2013. The characteristics of particulate absorption, scattering and attenuation coefficients in the surface ocean; contribution of the Tara Oceans expedition. Methods Oceanogr. 7, 52–62. https://doi.org/10.1016/j.mio.2013.11.002.

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. www.biogeosciences.net/6/751/2009/.

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.

Bricaud, A., Babin, M., Morel, A., Claustre, H., 1995. Variability in the chlorophyll-specific absorption coefficients of natural phytoplankton: analysis and parameterization. J. Geophys. Res. 100 (C7), 13321–13332. https://doi.org/10.1029/95JC00463.

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.

Budge, S.M., Devred, E., Forget, M.-H., Stuart, V., Trzcinski, M.K., Sathyendranath, S., Platt, T., 2014. Estimating concentrations of essential omega-3 fatty acids in the ocean: supply and demand. ICES J. Mar. Sci. 71 (4), 1885–1893. https://doi.org/10.1093/icesjms/fsu003.

Chase, A.P., Roesler, C.S., Laine, E.P., Teegarden, G.J., Hankinson, S.J., 2009. Spectral

reflectance-based observations of diatoms and dinoflagellates in Harpswell Sound, Maine. In: Geological Society of America Northeast Section 44th Annual Meeting, (edited, p. 41, GSA Abstracts with Programs).
Ciotti, Á.M., Lewis, M.R., Cullen, J.J., 2002. Assessment of the relationships between dominant cell size in natural phytoplankton communities and the spectral shape of the absorption coefficient. Limnol. Oceanogr. 47 (2), 404–417. https://doi.org/10.4319/lo.2002.47.2.0404.
Cushing, D.H., 1989. A difference in structure between ecosystems in strongly stratified waters and in those that are only weakly stratified. J. Plankton Res. 11 (1), 1–13. https://doi.org/10.1093/plankt/11.1.1.
Devred, E., Sathyendranath, S., Stuart, V., Maass, H., Ulloa, O., Platt, T., 2006. A two-component model of phytoplankton absorption in the open ocean: theory and applications. J. Geophys. Res. 111 (C030111), 1–11. https://doi.org/10.1029/2005JC002880.
Dierssen, H.M., 2010. Perspectives on empirical approaches for ocean color remote sensing of chlorophyll in a changing climate. Proc. Natl. Acad. Sci. 107 (40), 17072–17078. https://doi.org/10.1073/pnas.0913800107.
Dierssen, H.M., Kudela, R.M., Ryan, J.P., Zimmerman, R.C., 2006. Red and black tides: quantitative analysis of water-leaving radiance and perceived color for phytoplankton, colored dissolved organic matter, and suspended sediments. Limnol. Oceanogr. 51 (6), 2646–2659.
Evers-King, H., Bernard, S., Lain, L.R., Probyn, T.A., 2014. Sensitivity in reflectance attributed to phytoplankton cell size: forward and inverse modelling approaches. Opt. Express 22 (10), 11536–11551. https://doi.org/10.1364/OE.22.011536.
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.
Falkowski, P.G., Barber, R.T., Smetacek, V., 1998. Biogeochemical controls and feedbacks on ocean primary productivity. Science 281, 200–206. https://doi.org/10.1126/science.281.5374.200.
Field, C.B., Behrenfeld, M.J., Randerson, J.T., Falkowski, P., 1998. Primary production of the biosphere: integrating terrestrial and oceanic components. Science 281, 237–240. https://doi.org/10.1126/science.281.5374.237.
Gordon, H.R., Boynton, G.C., Balch, W.M., Groom, S.B., Harbour, D.S., Smyth, T.J., 2001. Retrieval of coccolithophore calcite concentration from SeaWiFS imagery. Geophys. Res. Lett. 28 (8), 1587–1590. https://doi.org/10.1029/2000GL012025.
Guidi, L., Stemmann, L., Jackson, G.A., Ibanez, F., Claustre, H., Legendre, L., Picheral, M., Gorsky, G., 2009. Effects of phytoplankton community on production, size and export of large aggregates: a world-ocean analysis. Limnol. Oceanogr. 54 (6), 1951–1963. https://doi.org/10.4319/lo.2009.54.6.1951.
Hirata, T., Aiken, J., Hardman-Mountford, N., Smyth, T.J., Barlow, R., 2008. An absorption model to determine phytoplankton size classes from satellite ocean colour. Remote Sens. Environ. 112, 3153–3159. https://doi.org/10.1016/j.rse.2008.03.011.
Hoepffner, N., Sathyendranath, S., 1993. Determination of the major groups of phytoplankton pigments from the absorption spectra of total particulate matter. J. Geophys. Res. 98 (C12), 22,789–722,803. https://doi.org/10.1029/93JC01273.
Hunter-Cevera, K.R., Neubert, M.G., Olson, R.J., Solow, A.R., Shalapyonok, A., Sosik, H.M., 2016. Physiological and ecological drivers of early spring blooms of a coastal phytoplankter. Science 354 (6310), 326–329. https://doi.org/10.1126/science.aaf8536.
IOCCG, et al., 2014. Phytoplankton functional types from space. Rep. 15 (163 pp., Dartmouth, Canada).
Jackson, T., Bouman, H.A., Sathyendranath, S., Devred, E., 2011. Regional-scale changes in diatom distribution in the Humboldt upwelling system as revealed by remote sensing: implications for fisheries. ICES J. Mar. Sci. 68 (4), 729–736. https://doi.org/10.1093/icesjms/fsq181.
Jensen, A., Sakshuag, E., 1973. Studies on the phytoplankton ecology of the Trondheimsfjord. II. Chloroplast pigments in relation to abundance and physiological state of the phytoplankton. J. Exp. Mar. Biol. Ecol. 11, 137–155. https://doi.org/10.1016/0022-0981(73)90052-X.
Kishino, M., Takahashi, M., Okami, N., Ichimura, S., 1985. Estimation of the spectral absorption coefficients of phytoplankton in the sea. Bull. Mar. Sci. 37 (2), 634–642.
Kostadinov, T.S., Siegel, D.A., Maritorena, S., 2009. Retrieval of the particle size distribution from satellite ocean color observations. J. Geophys. Res. 114 (C09015), 1–22. https://doi.org/10.1029/2009JC005303.
Latasa, M., 2007. Improving estimations of phytoplankton class abundances using CHEMTAX. Mar. Ecol. Prog. Ser. 329, 13–21. https://doi.org/10.3354/meps329013.
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., Hu, C., 2006. Global distribution of Case-1 waters: an analysis from SeaWiFS measurements. Remote Sens. Environ. 101, 270–276. https://doi.org/10.1016/j.rse.2005.11.008.
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 (3), 369–381. https://doi.org/10.1364/AO.49.000369.
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.
Longhurst, A., 1998. Ecological Geography of the Sea. Academic Press, Burlington, MA.
Longhurst, A., Sathyendranath, S., Platt, T., Caverhill, C., 1995. An estimate of global primary production in the ocean from satellite radiometer data. J. Plankton Res. 17 (6), 1245–1271. https://doi.org/10.1093/plankt/17.6.1245.
Mackey, M.D., Mackey, D.J., Higgins, H.W., Wright, S.W., 1996. CHEMTAX - a program for estimating class abundances from chemical markers: application to HPLC measurements of phytoplankton. Mar. Ecol. Prog. Ser. 114, 265–283.
Maritorena, S., D'Andon, O.H.F., Mangin, A., Siegel, D.A., 2010. Merged satellite ocean color data products using a bio-optical model: characteristics, benefits and issues. Remote Sens. Environ. 114, 1791–1804. https://doi.org/10.1016/j.rse.2010.04.002.
Marshall, B.R., Smith, R.C., 1990. Raman scattering and in-water ocean optical properties. Appl. Opt. 29 (1), 71–84. https://doi.org/10.1364/AO.29.000071.
Menden-Deuer, S., Lessard, E.J., 2000. Carbon to volume relationships for dinoflagellates, diatoms, and other protist plankton. Limnol. Oceanogr. 45 (3), 569–579. https://doi.org/10.4319/lo.2000.45.3.0569.
Mitchell, B.G., 1990. Algorithms for Determining the Absorption Coefficient of Aquatic Particulates Using the Quantitative Filter Technique (QFT), Paper Presented at Ocean Optics X, International Society for Optics and Photonics, Orlando, FL, 16–20 April.
Moberg, E.A., Sosik, H.M., 2012. Distance maps to estimate cell volume from two-dimensional plankton images. Limnol. Oceanogr. Methods 10, 278–288. https://doi.org/10.4319/lom.2012.10.278.
Mobley, C.D., Stramski, D., Bissett, W.P., Boss, E., 2004. Optical modeling of ocean waters: is the Case 1-Case 2 classification still useful? Oceanography 17 (2), 60–67. https://doi.org/10.5670/oceanog.2004.48.
Morel, A., Bricaud, A., 1981. Theoretical results concerning light absorption in a discrete medium, and application of specific absorption to phytoplankton. Deep-Sea Res. 28A (11), 1375–1393. https://doi.org/10.1016/0198-0149(81)90039-X.
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, 998–1011. https://doi.org/10.1016/j.rse.2009.01.008.
Morel, A., Prieur, L., 1977. Analysis of variations in ocean color. Limnol. Oceanogr. 22 (4), 709–722. https://doi.org/10.4319/lo.1977.22.4.0709.
Morel, A., Huot, Y., Gentili, B., Werdell, P.J., Hooker, S.B., Franz, B.A., 2007. Examining the consistency of products derived from various ocean color sensors in open ocean (Case 1) waters in the perspective of a multi-sensor approach. Remote Sens. Environ. 111, 69–88. https://doi.org/10.1016/j.rse.2007.03.012.
Mouw, C.B., Yoder, J.A., 2010. Optical determination of phytoplankton size composition from global SeaWiFS imagery. J. Geophys. Res. 115 (C12018), 1–20. https://doi.org/10.1029/2010JC006337.
Mouw, C.B., et al., 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.
Nair, A., Sathyendranath, S., Platt, T., Morales, J., Stuart, V., Forget, M.-H., Devred, E., Bouman, H., 2008. Remote sensing of phytoplankton functional types. Remote Sens. Environ. 112, 3366–3375. https://doi.org/10.1016/j.rse.2008.01.021.
Olson, R.J., Sosik, H.M., 2007. A submersible imaging-in-flow instrument to analyze nano- and microplankton: imaging FlowCytobot. Limnol. Oceanogr. Methods 5 (6), 195–203. https://doi.org/10.4319/lom.2007.5.195.
Olson, R.J., Shalapyonok, A., Sosik, H.M., 2003. An automated submersible flow cytometer for analyzing pico-and nanophytoplankton: FlowCytobot. Deep-Sea Res. I Oceanogr. Res. Pap. 50 (2), 301–315. https://doi.org/10.1016/S0967-0637(03)00003-7.
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. 103 (C11), 24937–24953. https://doi.org/10.1029/98JC02160.
Pan, X., Mannino, A., Marshall, H.G., Filippino, K.C., Mulholland, M.R., 2011. Remote sensing of phytoplankton community composition along the northeast coast of the United States. Remote Sens. Environ. 115 (12), 3731–3747. https://doi.org/10.1016/j.rse.2011.09.011.
Peacock, E.E., Olson, R.J., Sosik, H.M., 2014. Parasitic infection of the diatom *Guinardia delicatula*, a recurrent and ecologically important phenomenon on the New England Shelf. Mar. Ecol. Prog. Ser. 503, 1–10. https://doi.org/10.3354/meps10784.
Platt, T., Bouman, H., Devred, E., Fuentes-Yaco, C., Sathyendranath, S., 2005. Physical forcing and phytoplankton distributions. Sci. Mar. 69 (Suppl. 1), 55–73. https://doi.org/10.3989/scimar.2005.69s155.
Platt, T., Sathyendranath, S., White III, G.N., Fuentes-Yaco, C., Zhai, L., Devred, E., Tang, C., 2010. Diagnostic properties of phytoplankton time series from remote sensing. Estuar. Coasts 33, 428–439. https://doi.org/10.1007/s12237-009-9161-0.
Prieur, L., Sathyendranath, S., 1981. An optical classification of coastal and oceanic waters based on the specific spectral absorption curves of phytoplankton pigments, dissolved organic matter, and other particulate materials. Limnol. Oceanogr. 26 (4), 671–689.
Redfield, A.C., 1934. On the proportions of organic derivatives in sea water and their relation to the composition of plankton. In: James Johnstone Memorial Volume, Edited. University Press of Liverpool, pp. 176–192.
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.
Roesler, C.S., Perry, M.J., Carder, K.L., 1989. Modeling in situ phytoplankton absorption from total absorption spectra in productive inland marine waters. Limnol. Oceanogr. 34 (8), 1510–1523. https://doi.org/10.4319/lo.1989.34.8.1510.
Sathyendranath, S., Platt, T., 1988. The spectral irradiance field at the surface and in the interior of the ocean: a model for applications in oceanography and remote sensing, J. Geophys. Res. Oceans 93 (C8), 9270–9280.
Sathyendranath, S., Longhurst, A., Caverhill, C., Platt, T., 1995. Regionally and seasonally differentiated primary production in the North Atlantic. Deep-Sea Res. 42 (10), 1773–1802. https://doi.org/10.1016/0967-0637(95)00059-F.
Sathyendranath, S., Cota, G., Stuart, V., Maass, H., Platt, T., 2001. Remote sensing of phytoplankton pigments: a comparison of empirical and theoretical approaches. Int. J. Remote Sens. 22 (2–3), 249–273. https://doi.org/10.1080/014311601449925.
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.
Sauer, M.J., Roesler, C.S., Werdell, P.J., Barnard, A., 2012. Under the hood of satellite

empirical chlorophyll *a* algorithms: revealing the dependencies of maximum band ratio algorithms on inherent optical properties. Opt. Express 20 (19), 20920–20933. https://doi.org/10.1364/OE.20.020920.

Siegel, D.A., 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.

Smetacek, V., 1999. Diatoms and the ocean carbon cycle. Protist News 150, 25–32. https://doi.org/10.1016/S1434-4610(99)70006-4.

Son, S., Platt, T., Fuentes-Yaco, C., Bouman, H., Devred, E., Wu, Y., Sathyendranath, S., 2007. Possible biogeochemical response to the passage of Hurricane Fabian observed by satellites. J. Plankton Res. 29 (8), 687–697. https://doi.org/10.1093/plankt/fbm050.

Sosik, H.M., Mitchell, B.G., 1994. Effects of temperature on growth, light absorption, and quantum yield in *Dunaliella tertiolecta* (chlorophyceae). J. Phycol. 30, 833–840. https://doi.org/10.1111/j.0022-3646.1994.00833.x.

Sosik, H.M., Olson, R.J., 2007. Automated taxonomic classification of phytoplankton sampled with imaging-in-flow cytometry. Limnol. Oceanogr. Methods 5, 204–216. https://doi.org/10.4319/lom.2007.5.204.

Sosik, H.M., Olson, R.J., 2008. Phytoplankton community regulation on the New England shelf: insights from automated submersible flow cytometry. In: Proceedings of Ocean Optics XIX, Edited, pp. 1–12 (Barga, Italy).

Sosik, H.M., Olson, R.J., Armbrust, E.V., 2010. Flow cytometry in phytoplankton research. In: Suggett, D.J., Prasil, O., Borowitzka, M.A. (Eds.), Chlorophyll-a Fluorescence in Aquatic Science: Methods and Applications. Developments in Applied Phycology 4 Springer, pp. 171–185. https://doi.org/10.1007/978-90-481-9268-7_8.

Stuart, V., Sathyendranath, S., Head, E.J.H., Platt, T., Irwin, B.D., Maass, H., 2000. Bio-optical characteristics of diatom and prymnesiophyte populations in the Labrador Sea. Mar. Ecol. Prog. Ser. 201, 91–106. https://doi.org/10.3354/meps201091.

Trzcinski, M.K., Devred, E., Platt, T., Sathyendranath, S., 2013. Variation in ocean colour may help predict cod and haddock recruitment. Mar. Ecol. Prog. Ser. 491, 187–197. https://doi.org/10.3354/meps10451.

Uitz, J., Claustre, H., Morel, A., Hooker, S.B., 2006. Vertical distribution of phytoplankton communities in open ocean: An assessment based on surface chlorophyll. J. Geophys. Res. 111 (C08005), 1–23. https://doi.org/10.1029/2005JC003207.

Ulloa, O., Sathyendranath, S., Platt, T., 1994. Effect of the particle size distribution on the backscattering ratio in seawater. Appl. Opt. 33 (30), 7070–7077.

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., et al., 2013. Generalized ocean color inversion model for retrieving marine inherent optical properties. Appl. Opt. 52 (10), 2019–2037. https://doi.org/10.1364/AO.52.002019.

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., Boss, E., Lee, Z., 2013. Influence of Raman scattering on ocean color inversion models. Appl. Opt. 52 (22), 5552–5561.

Whitmire, A.L., Pegau, W.S., Karp-Boss, L., Boss, E., Cowles, T.J., 2010. Spectral backscattering properties of marine phytoplankton cultures. Opt. Express 18 (14), 15073–15093. https://doi.org/10.1364/OE.18.015073.

Zhai, L., Platt, T., Tang, C., Dowd, M., Sathyendranath, S., Forget, M.-H., 2008. Estimation of phytoplankton loss rate by remote sensing. Geophys. Res. Lett. 35 (L23606), 1–5. https://doi.org/10.1029/2008GL035666.

Zhang, X., Huot, Y., Bricaud, A., Sosik, H.M., 2015. Inversion of spectral absorption coefficients to infer phytoplankton size classes, chlorophyll concentration, and detrital matter. Appl. Opt. 54 (18), 5805–5816. https://doi.org/10.1364/AO.54.005805.

Zibordi, G., Holben, B., Mélin, F., D'Alimonte, D., Berthon, J.-F., Slutsker, I., Giles, D., 2010. AERONET-OC: an overview. Can. J. Remote. Sens. 36 (5), 488–497. https://doi.org/10.5589/m10-073.