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

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