Discrimination of phytoplankton functional groups using an ocean reflectance inversion model

P. Jeremy Werdell,[1,2,*] Collin S. Roesler,[3] and Joaquim I. Goes[4]

[1]NASA Goddard Space Flight Center, Code 616, Greenbelt, Maryland 20771, USA

[2]School of Marine Sciences, University of Maine, Orono, Maine 04469, USA

[3]Department of Earth and Ocean Sciences, Bowdoin College, Brunswick, Maine 04011, USA

[4]Lamont Doherty Earth Observatory, Columbia University, Palisades, New York 10964, USA

*Corresponding author: jeremy.werdell@nasa.gov

Received 6 January 2014; revised 4 April 2014; accepted 8 May 2014;
posted 10 June 2014 (Doc. ID 204088); published 21 July 2014

Ocean reflectance inversion models (ORMs) provide a mechanism for inverting the color of the water observed by a satellite into marine inherent optical properties (IOPs), which can then be used to study phytoplankton community structure. Most ORMs effectively separate the total signal of the collective phytoplankton community from other water column constituents; however, few have been shown to effectively identify individual contributions by multiple phytoplankton groups over a large range of environmental conditions. We evaluated the ability of an ORM to discriminate between *Noctiluca miliaris* and diatoms under conditions typical of the northern Arabian Sea. We: (1) synthesized profiles of IOPs that represent bio-optical conditions for the Arabian Sea; (2) generated remote-sensing reflectances from these profiles using Hydrolight; and (3) applied the ORM to the synthesized reflectances to estimate the relative concentrations of diatoms and *N. miliaris*. By comparing the estimates from the inversion model with those from synthesized vertical profiles, we identified those conditions under which the ORM performs both well and poorly. Even under perfectly controlled conditions, the absolute accuracy of ORM retrievals degraded when further deconstructing the derived total phytoplankton signal into subcomponents. Although the absolute magnitudes maintained biases, the ORM successfully detected whether or not *Noctiluca miliaris* appeared in the simulated water column. This quantitatively calls for caution when interpreting the absolute magnitudes of the retrievals, but qualitatively suggests that the ORM provides a robust mechanism for identifying the presence or absence of species.

OCIS codes: (010.4450) Oceanic optics; (280.4991) Passive remote sensing.
http://dx.doi.org/10.1364/AO.53.004833

1. Introduction

Changes in marine phytoplankton species compositions can be triggered by changes in the Earth's climate [1–3]. Long-term changes in phytoplankton community composition contribute to changes in food webs and air–land–sea carbon cycles [4]. The northern Arabian Sea, for example, appears to be undergoing a shift in phytoplankton species composition from diatoms to *Noctiluca miliaris* during the annual winter (November to March) Northeast Monsoon (NEM) [5–7]. Blooms of *N. miliaris* are disrupting the traditional diatom-dominated food chain during the NEM and altering the magnitude of carbon export [8]. Ship- and aircraft-based sampling alone cannot produce sufficient data records to study such phytoplankton diversity shifts

1559-128X/14/224833-17$15.00/0

on large temporal and spatial scales. Following, the oceanographic community has invested in the development of methods to identify and discriminate between members of the phytoplankton community using remotely sensed data records [9]. Satellite ocean color instruments provide consistent and high-volume data records on scales that far exceed current ship and aircraft sampling strategies, with time-series of sufficient length to allow retrospective analysis of oceanographic trends. The imagery captured by the NASA Sea-Viewing Wide Field-of-View Sensor (SeaWiFS; 1997–2010) and the Moderate Resolution Imaging Spectroradiometer onboard Aqua (MODISA; 2002–present), for example, provide viable data records for observing almost two decades of changes in the biogeochemistry of both global and regional marine ecosystems [10–13]. The community considers ocean color satellite estimates of phytoplankton community composition to be sufficiently critical for advancing our understanding of biogeochemistry and carbon cycle science that all forthcoming ocean color satellite programs are now required to have the capability to discriminate between phytoplankton groups [14,15].

A variety of approaches exist for identifying phytoplankton communities from satellite ocean color data records. Briefly, satellite ocean color instruments measure the spectral radiance exiting the top of the atmosphere ($L_t(\lambda)$; $\mu W\ cm^{-2}\ nm^{-1}\ sr^{-1}$) at discrete visible and infrared wavelengths. Atmospheric correction algorithms are applied to $L_t(\lambda)$ to remove the contribution of the atmosphere from the total signal and produce estimates of remote-sensing reflectances $R_{rs}(\lambda)$; sr^{-1}), the light exiting the water mass was normalized to the hypothetical condition of an overhead Sun and no atmosphere [16,17]. Bio-optical algorithms are applied to the $R_{rs}(\lambda)$ to produce estimates of additional marine geophysical properties, such as the near-surface concentration of the phytoplankton pigment chlorophyll-a (C_ϕ; $mg\ m^{-3}$) [18,19] and inherent optical properties (IOPs: the spectral absorption and scattering characteristics of ocean water and its dissolved and particulate constituents) [20,21]. Several methods for identifying phytoplankton communities do so using empirically derived thresholds on the magnitudes of the derived C_ϕ or IOPs [22–27], whereas others interpret the retrieved radiometric spectral shapes [28–37]. In general, the former, abundance-based methods exploit observed relationships between the trophic status of the environment and the type of phytoplankton expected to be present, whereas the latter, spectral methods exploit differences in the optical signatures of specific size classes or functional groups to distinguish between phytoplankton types [9]. Corresponding studies related to changing ecosystems or the appearance of new species most often employ spectral methods, as abundance methods cannot capture the desired signals when the emerging environmental relationships do not appear in their *in situ* training data sets.

Ocean reflectance inversion models (ORMs) provide a common spectral method for inverting the "color" of the water observed by a satellite [e.g., $R_{rs}(\lambda)$] into marine IOPs through a combination of empiricism and radiative transfer theory [20,21]. Generally speaking, ORMs attribute variations realized in the spectral shape of $R_{rs}(\lambda)$ to varying marine populations of phytoplankton, nonalgal particles (NAP), and colored dissolved organic material (CDOM), all of which maintain unique optical signatures. They operate by assuming spectral shape functions of the constituent absorption and scattering components and retrieving the magnitudes of each constituent required to match the spectral distribution of $R_{rs}(\lambda)$. Therefore, unlike abundance methods, ORMs can, in principle, discriminate between phytoplankton groups with common abundances, provided the groups present have contrasting optical signatures within the spectral bands detected. This ability to further deconvolve $a_\phi(\lambda)$ into contributions by individual phytoplankton groups, however, remains inadequately demonstrated for a large range of environmental conditions [35,37]. When ORMs include only a single spectrum for a phytoplankton group, they remain confounded by natural variations in the spectral characteristics of that group due to growth stage, nutrient availability, and ambient light history. Furthermore, their solution can be statistically ambiguous [i.e., several combinations of IOPs can produce similar $R_{rs}(\lambda)$] and dependent upon the wavelength suite (number and position) used in the inversion [38,39].

Here, we evaluate the ability of a common ORM to discriminate between two phytoplankton groups residing under varied biophysical conditions. Our goal was twofold: evaluate the ability of an ORM to quantify blooms with variable depths, thicknesses, and ages; and verify the spectral requirements for success using an ORM. For this case study, we selected a well-vetted ORM algorithmic form that we parameterized specifically to identify *N. miliaris* in a mixed phytoplankton community, using satellite ocean color data records collected in the Arabian Sea [40]. Briefly, we: (1) synthesized a series of vertical profiles of IOPs that represent a wide variety of bio-optical conditions for the northern Arabian Sea; (2) generated $R_{rs}(\lambda)$ from these profiles using Hydrolight-Ecolight 5 (HE5) [41]; (3) applied the ORM to the synthesized $R_{rs}(\lambda)$ to estimate the relative presence of diatoms and *N. miliaris* for each example; and (4) repeated the third step using $R_{rs}(\lambda)$ with reduced spectral resolution. Comparing the estimates from the inversion model with those from the synthesized vertical profiles provided a mechanism for identifying those bio-optical conditions under which the ORM performed both well and poorly. While we focus on two specific phytoplankton populations, our ORM otherwise maintains the common form of many established approaches [20,21] and, thus, we expect our results to provide a case study that transfers well to other algorithms and phytoplankton populations.

2. Methods

A. Study Site and *In Situ* Data

Field experiments crafted to study the transition from *N. miliaris* to diatoms in the northern Arabian Sea were conducted during the 2011 NEM onboard the Indian Fishery Oceanographic Research Vessel *Sagar Sampada*, which traveled offshore following a northwest transect from Goa, India. Sampling dates ranged from March 7th–9th, 2011 and encompassed diatom-dominant, *N. miliaris*-dominant, and mixed population stations. We acquired measurements of $R_{rs}(\lambda)$ and spectral absorption coefficients during this field campaign, using data collection and processing methods that appear in Roesler *et al.* [40] and Thibodeau *et al.* [7], respectively. With regard to the latter, we measured the spectrophotometric absorption of particles from discrete water samples using the quantitative filter technique, as modified by Mitchell [42] and Roesler [43]. We determined phytoplankton and non-algal absorption fractions by extraction [44] and identified and quantified phytoplankton species microscopically [5]. We assigned taxonomic dominance to each *in situ* sample using the microscopic species enumerations. We also collected discrete water samples via Niskin bottles for analysis of C_ϕ and other pigments by the NASA high performance liquid chromatography facility at Goddard Space Flight Center. As in Thibodeau *et al.* [7], we used these measurements to derive average C_ϕ-specific absorption spectra for *N. miliaris* ($a^*_{\phi N}(\lambda)$; $m^2\ mg^{-1}$) and diatoms ($a^*_{\phi D}(\lambda)$; $m^2\ mg^{-1}$) (Fig. 1).

The species *N. miliaris* appearing in the Arabian Sea is a mixotrophic dinoflagellate with green prasinoxanthin-containing symbionts (*Pedinomonas noctilucae*), which is physiologically different than the pink, bioluminescent variety more commonly found in temperate coastal waters [45] and optically different than most other resident diatom species, including mixtures of *Rhizosolenia spp.*, *Thallassiothrix spp.*, *Skeletonema spp.*, and *Chaetoceros spp.*

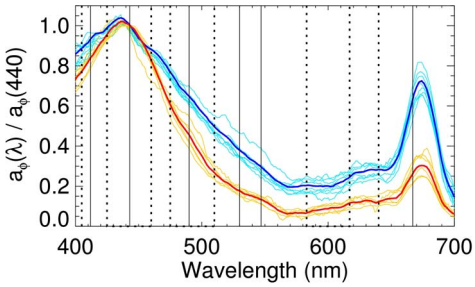

Fig. 1. Absorption spectra for *N. miliaris* and diatoms, normalized to 440 nm [7]. Thin blue lines show spectra from *in situ* stations dominated by diatoms, with the thick blue line indicating the mean spectrum. Thin orange lines show spectra from *in situ* stations dominated by *N. miliaris*, with the thick red line indicating the mean spectrum. The average $a^*_{\phi D}(440)$ and $a^*_{\phi N}(440)$ were 0.052 (±0.022) and 0.066 (±0.015) $m^2\ mg^{-1}$, respectively. Vertical solid lines indicate center MODISA wavelengths. Vertical dotted lines show the additional wavelengths considered in this study.

We acknowledge that *N. miliaris* and diatoms alone cannot universally represent the phytoplankton community of the northern Arabian Sea at all times and that single $a^*_\phi(\lambda)$ cannot universally represent a phytoplankton group at all times. However, the contrasting optical signatures of *N. miliaris* and diatoms and their known colocation under many circumstances provided an ideal "real world" scenario for our sensitivity analyses.

B. Ocean Reflectance Inversion Model

Our ORM adopts the generalized IOP (GIOP) form described in Werdell *et al.* [21]. Briefly, ocean color satellite instruments provide estimates of $R_{rs}(\lambda)$, which we convert to their subsurface values using the method presented in Lee *et al.* [46]:

$$r_{rs}(\lambda, 0^-) = \frac{R_{rs}(\lambda)}{0.52 + 1.7R_{rs}(\lambda)}. \tag{1}$$

Subsurface remote-sensing reflectances relate to marine IOPs, following Gordon *et al.* [47]:

$$r_{rs}(\lambda, 0^-) = 0.0949u(\lambda) + 0.0794u(\lambda)^2$$
$$u(\lambda) = \frac{b_b(\lambda)}{a(\lambda) + b_b(\lambda)}, \tag{2}$$

where $b_b(\lambda)$ is the total backscattering coefficient (m^{-1}) and $a(\lambda)$ is the total absorption coefficient (m^{-1}). Total absorption can be expanded as the sum of all absorbing components. Further, each component can be expressed as the product of its mass-specific absorption spectrum (eigenvector: a^*) and its magnitude or concentration (eigenvalue: M):

$$a(\lambda) = a_w(\lambda) + M_\phi a^*_\phi(\lambda) + M_{dg}a^*_{dg}(\lambda), \tag{3}$$

where the subscripts w, ϕ, and dg indicate contributions by water, phytoplankton, and NAP (d etritus) + CDOM (g elbstoff). In the remote sensing paradigm, absorption by NAP ($a_d(\lambda)$; m^{-1}) and absorption by CDOM ($a_g(\lambda)$; m^{-1}) cannot currently be effectively separated, as they maintain similar spectral shapes. We expressed $a^*_{dg}(\lambda)$ as $\exp(-S_{dg}\lambda)$, where S_{dg} describes the rate of exponential decay and typically varies between 0.01 and 0.02 nm^{-1} [48]. We further deconstructed $a^*_\phi(\lambda)$ into contributions by diatoms and *N. miliaris* (Fig. 1):

$$M_\phi a^*_\phi(\lambda) = M_{\phi D}a^*_{\phi D}(\lambda) + M_{\phi N}a^*_{\phi N}(\lambda), \tag{4}$$

where the subscripts D and N indicate diatoms and *N. miliaris*. We adopted the $a^*_{\phi D}(\lambda)$ and $a^*_{\phi D}(\lambda)$ described in Section 2.A. As these are C_ϕ-specific absorption coefficients, the eigenvalues $M_{\phi D}$ and $M_{\phi N}$ represent chlorophyll for diatoms ($C_{\phi D}$) and *N. miliaris* ($C_{\phi N}$).

Similar to absorption, total backscattering can be expanded to:

$$b_b(\lambda) = b_{bw}(\lambda) + M_{bp} b^*_{bp}(\lambda), \quad (5)$$

where the subscripts bw and bp indicate contributions by water and particles (= NAP+ phytoplankton), respectively. We expressed $b^*_{bp}(\lambda)$ as λ^η, respectively, where η defines the steepness of the power law and typically varies between −2 and 0 in natural waters [49]. Both $a_w(\lambda)$ and $b_{bw}(\lambda)$ are known, as are their temperature and salinity dependencies [50–52].

Four unknowns remain in Eqs. (1)–(5) after we assign the eigenvectors: $M_{dg}, M_{bp}, M_{\phi D}$ (hereafter $C_{\phi D}$), and $M_{\phi N}$ (hereafter $C_{\phi N}$). Using $R_{rs}(\lambda)$ as input, we estimate these eigenvalues via nonlinear least squares (Levenberg–Marquardt) inversion of Eq. (2) [21]. We retained only those solutions with viable estimates of M_{dg}, M_{bp}, $C_{\phi D}$, and $C_{\phi N}$ (e.g., $-0.05 a_w(\lambda) \le a_{dg}(\lambda) \le 5\ \mathrm{m}^{-1}$) that resulted in reconstructed $R_{rs}(\lambda)$, which differed from the input $R_{rs}(\lambda)$ by less than 33% for all wavelengths between 400 and 600 nm. We reconstructed $R_{rs}(\lambda)$ using the retrieved eigenvalues as input into Eqs. (1)–(5) and defined failure as nonconvergence in the inversion. Werdell *et al.* [21] outlined the similarities in form and accuracy shared by this ORM configuration and other common approaches (e.g., [19,20,31,37,46]).

C. Synthesis of IOP Profiles

We synthesized a series of vertical profiles of spectral IOPs, representing a wide variety of bio-optical conditions for the northern Arabian Sea, for use as input into the "IOP Data" model of HE5. Specifically, we constructed vertical profiles of absorption and attenuation by particles plus CDOM ($a_{pg}(\lambda)$ and $c_{pg}(\lambda)$, respectively; m^{-1}):

$$\begin{aligned} a_{pg}(\lambda,z) &= a_{dg}(\lambda,z) + a_{\phi D}(\lambda) + a_{\phi N}(\lambda) a_{pg}(\lambda,z) \\ &= a_{dg}(\lambda,z) + C_{\phi D}(z) a^*_{\phi D}(\lambda) + C_{\phi N}(z) a^*_{\phi N}(\lambda), \end{aligned} \quad (6)$$

$$c_{pg}(\lambda,z) = a_{pg}(\lambda,z) + b_p(\lambda,z), \quad (7)$$

where $b_p(\lambda,z)$ is total scattering by particles (m^{-1}):

$$\begin{aligned} b_p(\lambda,z) &= b_d(\lambda) + b_{\phi D}(\lambda) + b_{\phi N}(\lambda) \\ b_p(\lambda,z) &= M_d(z) b^*_d(\lambda) + C_{\phi D}(z) b^*_{\phi D}(\lambda) + C_{\phi N}(z) b^*_{\phi N}(\lambda). \end{aligned} \quad (8)$$

We expressed $M_d(z)$ as $b_d(555,z)$ and the shape of the eigenvectors as power laws. Note that $b^*_{\phi D}(555)$ and $b^*_{\phi N}(555)$ represent C_ϕ-specific scattering coefficients:

$$\begin{aligned} b_p(\lambda,z) = b_d(555,z)\left(\frac{\lambda}{555}\right)^{\eta_d} + C_{\phi D}(z) b^*_{\phi D}(555)\left(\frac{\lambda}{555}\right)^{\eta_{\phi D}} \\ + C_{\phi N}(z) b^*_{\phi N}(555)\left(\frac{\lambda}{555}\right)^{\eta_{\phi N}}. \end{aligned} \quad (9)$$

Unlike our ORM, we deconvolved $a_{dg}(\lambda,z)$ into its two components:

$$\begin{aligned} a_{dg}(\lambda,z) &= M_d(z) a^*_d(\lambda) + M_g(z) a^*_g(\lambda) \\ a_{dg}(\lambda,z) &= a_d(443,z)\exp(-S_d(\lambda - 443)) \\ &\quad + a_g(443,z)\exp(-S_g(\lambda - 443)). \end{aligned} \quad (10)$$

Construction of $a_{pg}(\lambda,z)$ and $c_{pg}(\lambda,z)$ using Eqs. (6)–(10) required defining 14 eigenvectors and eigenvalues: $C_{\phi D}(z), C_{\phi N}(z), a^*_{\phi D}(\lambda), a^*_{\phi N}(\lambda), b_d(555,z)$, $b^*_{\phi D}(555), b^*_{\phi N}(555), \eta_d, \eta_{\phi D}, \eta_{\phi N}, a_d(443,z), a_g(443,z)$, S_g, and S_d (Table 1). We only assigned depth dependence to a single term $C_{\phi N}(z)$, which we constructed as a Gaussian expression defined by a null background signal, a full width at half-maximum (N^W), and a maximum value ($N^{\max}$) at an assigned depth (N^Z). $C_{\phi N}(z)$ alone controlled the vertical structure of our simulated profiles (Fig. 2).

We carefully selected appropriate ranges of simulation parameters, based on extensive literature reviews and local knowledge from field sampling (Table 1). In addition to $a_{pg}(\lambda,z)$ and $c_{pg}(\lambda,z)$, running the HE5 "IOP Data" model also required generation of vertical profiles of $a_g(\lambda,z)$ [see Eq. (10)] and $C_\phi(z)$ $(= C_{\phi N}(z) + C_{\phi D}(z))$, plus spectra of $a^*_\phi(\lambda)$. For the latter, we generated $a^*_\phi(\lambda)$ for each simulation following:

$$a^*_\phi(\lambda) = \frac{C_{\phi N}}{C_{\phi N} + C_{\phi D}} a^*_{\phi N}(\lambda) + \frac{C_{\phi D}}{C_{\phi N} + C_{\phi D}} a^*_{\phi D}(\lambda). \quad (11)$$

Using all combinations of the values presented in Table 1 resulted in 1,920 simulated stations. We generated vertical profiles and spectra at 5 nm intervals from 400 to 750 nm to loosely mimic the output of a WET Labs, Inc. AC-S (a hyperspectral absorption and attenuation meter), which HE5 readily accepts as input into its "IOP Data" model. Figure 2 provides an example suite of profiles for one simulated station.

D. Modeling and Data Analysis

Using our simulated data as input into HE5, we generated simulated remote-sensing reflectances (Fig. 3) and depth profiles of spectral upwelling radiance and downwelling irradiance. We configured HE5 as follows: spectral output from 400 to 700 nm at 5 nm intervals; vertical output from 0 to 30 m at 3 m intervals; an infinitely deep bottom; cloud cover of 25%; wind speed of 5 $\mathrm{m\,s}^{-1}$; solar geometry specific to year day 45, latitude 20°N, and longitude 60°E at local noon; the RADTRAN sky model; all inelastic scattering options enabled; and a spectrally constant and depth-independent b_b/b of 0.01 [53]. Although some uncertainty accompanied our assignment of the latter, we repeated the analysis with other values and found our overall results to be unchanged.

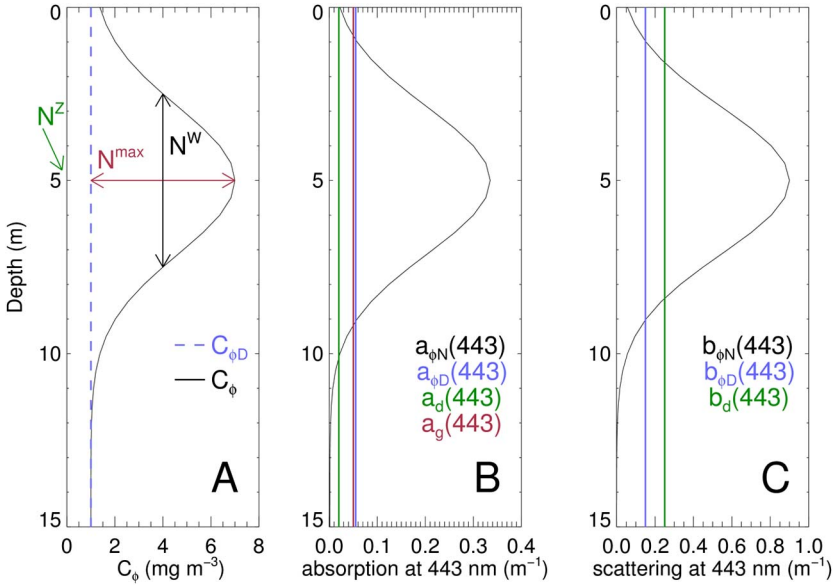

Fig. 2. Example simulated profiles for C_ϕ, absorption coefficients at 443 nm, and scattering coefficients at 443 nm. The simulation parameters for these profiles are: $C_{\phi D} = 1$ mg m^{-3}, $N^Z = 5$ m, $N^W = 5$ m, $N^{\max} = 6$ mg m^{-3}, $a_d(443) = 0.02$ m^{-1}, $a_g(443) = 0.05$ m^{-1}, and $b_d(555) = 0.2$ m^{-1} (Table 1). Panel (A) shows C_ϕ and $C_{\phi D}$ and indicates the three parameters that describe the Gaussian-shaped $C_{\phi N}$ (N^Z, N^W, and $N^{\max}$). Panel (B) shows the corresponding $a_{\phi N}(443)$, $a_{\phi D}(443)$, $a_d(443)$, and $a_g(443)$. Panel (C) shows the corresponding $b_{\phi N}(443)$, $b_{\phi D}(443)$, and $b_d(443)$.

We applied the ORM to each simulated $R_{rs}(\lambda)$ using two different suites of input wavelengths with relevance to current and planned satellite ocean color science. First, we used only six visible MODISA wavelengths to represent the present state of satellite ocean color (412, 443, 488, 531, 547, and 667 nm). Second, we considered a suite of 16 visible wavelengths in consideration for future ocean color satellite instruments (the six MODISA wavelengths plus 400, 425, 460, 475, 510, 583, 617, 640, 655, and 665 nm) [15,54]. Hereafter, we refer to this expanded suite of wavelengths as "all wavelengths" in figures and tables. While we acknowledge the value in evaluating alternate combinations of wavelengths to study phytoplankton species composition [55], our analysis of these two suites also provides practical results for existing and forthcoming satellite instruments. These two $R_{rs}(\lambda)$ suites serve well to explore what can be accomplished today (e.g., using MODISA) and what we might be able to accomplish in the near future {e.g., via the upcoming NASA Pre-Aerosols, Clouds, and Ocean Ecosystems (PACE) mission [15]}.

At this stage, our synthesized $C_\phi(z)$ and IOP(z) profiles provide the "ground truth" for comparison with the modeled values from the ORM. Approximately 90% of what a satellite ocean color "sees" includes weighted contributions from all water column constituents shallower than the first optical depth (z_{90}; m^{-1}); that is, the e-folding depth for diffuse attenuation coefficients for downwelling irradiance ($K_d(\lambda)$; m^{-1}) [56]. To properly account for the subsurface $C_{\phi N}(z)$ maxima, we optically weighted the $C_\phi(z)$

Table 1. Values Used to Simulate Vertical IOP Profiles for Input in HE5

Variable	Values	Units	References
$C_{\phi D}$	0.02, 0.1, 0.5, 1	mg m^{-3}	Gomes *et al.* [6], Thibodeau *et al.* [7]
$N^{\max}$	0, 0.5, 1, 3, 6	mg m^{-3}	Gomes *et al.* [6,8], Thibodeau *et al.* [7]
N^Z	0, 5, 10, 15	m	Gomes *et al.* [6,8], Thibodeau *et al.* [7]
N^W	1, 2, 5	m	Gomes *et al.* [6,8], Thibodeau *et al.* [7]
$a^*_{\phi D}(\lambda)$	Figure 1	m^2 mg^{-1}	This study (Fig. 1)
$a^*_{\phi N}(\lambda)$	Figure 1	m^2 mg^{-1}	This study (Fig. 1)
$b_d(555)$	0.1, 0.2	m^{-1}	Roesler *et al.* [40]
$b^*_{\phi D}(555)$	0.15	m^2 mg^{-1}	Morel [70], Parab *et al.* [5], Roesler *et al.* [40]
$b^*_{\phi N}(555)$	0.15	m^2 mg^{-1}	Roesler *et al.* [40]
η_d	−1	Unitless	Stramski *et al.* [49]
$\eta_{\phi D}$	0	Unitless	Stramski and Mobley [71]
$\eta_{\phi N}$	0	Unitless	Roesler *et al.* [40]
$a_d(443)$	0.002, 0.02	m^{-1}	Roesler *et al.* [40]
$a_g(443)$	0.005, 0.05	m^{-1}	Roesler *et al.* [40]
S_d	0.011	nm^{-1}	Roesler *et al.* [40,48]
S_g	0.018	nm^{-1}	Roesler *et al.* [40,48]

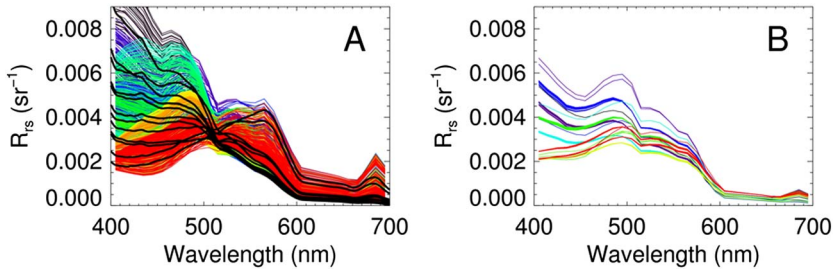

Fig. 3. $R_{rs}(\lambda)$ collected *in situ* (thick black lines) and synthesized using HE5 (colored thin lines). Panel (A) shows all synthesized and *in situ* $R_{rs}(\lambda)$. Panel (B) shows only synthesized $R_{rs}(\lambda)$ with optically weighted C_ϕ ranging from 0.95 to 1.05 mg m^{-3}. Colors are only used to visually distinguish between spectra.

and IOP(z)s using the HE5-generated radiance and irradiance profiles, following the method of Zaneveld *et al.* [57], which produced the corresponding pseudo-depth-integrated values for these variables detected by the satellite instrument. We hereafter imply that values are optically weighted, unless depth dependence is specifically indicated [e.g., by "(z)"]. As will be elaborated upon later, this weighting process highlights an inherent ambiguity within the ocean color paradigm; that is, identical C_ϕ can be produced by larger, deeper blooms and smaller, shallower blooms, simply because of their relative vertical location in the water column.

3. Results

A. Synthesized $R_{rs}(\lambda)$

The synthesized $R_{rs}(\lambda)$ reasonably reproduced the dynamic range of *in situ* values acquired in the Arabian Sea in March, 2011 (Fig. 3). Unusual $R_{rs}(\lambda)$ can exist in any suite of synthesized values when input parameters combine in ways that do not occur naturally. We identified unnatural simulations for our study area as those with single scattering albedos at 443 nm less than 0.7 [occurring for stations with combinations of the lowest C_ϕ and highest $a_{dg}(\lambda)$] or R_{rs} (490) greater than 0.008 sr^{-1} [occurring for stations with combinations of the lowest C_ϕ and lowest $a_{dg}(\lambda)$]. We considered simulations with low single scattering albedos to be overly atypical, as previous radiative transfer and *in situ* studies indicated it exceeds 0.8 in most natural waters [58,59]. Similarly, we considered simulations with $R_{rs}(490) > 0.008$ sr^{-1} to be overly unusual, as this threshold exceeds our *in situ* measurements by >40% and an extreme value for open ocean models by >10% [60]. Elimination of these unnatural simulations left 1437 stations for our analyses. As these remaining synthesized values fall within the envelope of spectra measured in the field, we propose that they adequately represent a natural range of environmental conditions for the northern Arabian Sea, at least for the purposes of this theoretical analysis.

B. Retrieval of Total IOPs

Retrievals of $b_{bp}(\lambda)$, $a_{pg}(\lambda)$, and $a_\phi(\lambda)$ compared favorably with the ground-truth-synthesized values (Fig. 4, Table 2). It is clear by difference [$a_{pg}(443) - a_\phi(443)$] that retrievals of $a_{dg}(443)$ also compared favorably. Using all available wavelengths yielded better results than only considering MODISA wavelengths. When using all wavelengths in the inversion, the comparisons maintained coefficients of determination (r^2), least squares regression slopes, and median ORM-to-synthesized ratios that ranged from 0.77 to 0.97, 0.65 to 0.95, and 0.85 to 1.21 respectively, indicating good performance over the full dynamic ranges of retrievals. The mean r^2, regression slopes, and ratios were 0.85, 0.83, and 1.07, respectively. The results changed somewhat

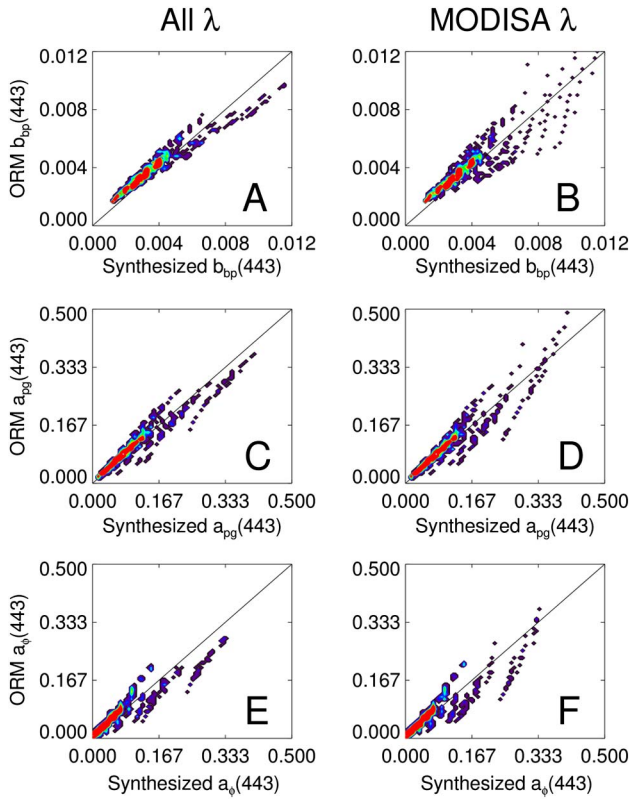

Fig. 4. Comparison of ground-truth (synthesized) and ORM-derived $b_{bp}(443)$ (A) and (B), $a_{pg}(443)$ (C) and (D), and $a_\phi(443)$ (E) and (F) using all available wavelengths in the inversion (left column) and only MODISA visible wavelengths in the inversion (right column). Colors show the numerical density of the retrievals, with red to purple indicating high to low volumes of sample sizes. We present comparative statistics in Table 2.

Table 2. Ordinary Least Squares Regression Statistics for Ground-Truth-Synthesized Versus ORM-Derived IOPs[a]

		All λ				MODISA λ			
		r^2	Slope (SE)	Ratio	RMSE	r^2	Slope (SE)	Ratio	RMSE
b_{bp}	412	0.92	0.79 (0.006)	1.19	0.00034	0.85	0.84 (0.009)	1.18	0.00054
	443	0.94	0.80 (0.005)	1.14	0.00031	0.86	0.85 (0.009)	1.13	0.00051
	488	0.95	0.81 (0.005)	1.06	0.00028	0.87	0.86 (0.009)	1.07	0.00048
	531	0.95	0.81 (0.005)	1.00	0.00027	0.88	0.86 (0.008)	1.01	0.00047
	547	0.95	0.81 (0.005)	0.98	0.00026	0.88	0.86 (0.008)	0.99	0.00046
	667	0.95	0.79 (0.005)	0.85	0.00026	0.88	0.85 (0.008)	0.88	0.00044
a_{pg}	412	0.97	0.95 (0.005)	1.08	0.01031	0.94	1.02 (0.007)	1.09	0.01509
	443	0.94	0.89 (0.006)	1.09	0.01246	0.91	0.94 (0.008)	1.09	0.01608
	488	0.91	0.83 (0.007)	1.03	0.00785	0.86	0.90 (0.010)	1.03	0.01118
	531	0.86	0.75 (0.008)	0.95	0.00468	0.77	0.85 (0.012)	0.96	0.00713
	547	0.85	0.72 (0.008)	0.91	0.00358	0.76	0.84 (0.013)	0.94	0.00567
	667	0.78	0.68 (0.010)	1.07	0.00871	0.68	0.74 (0.014)	1.11	0.01249
a_{ϕ}	412	0.89	0.80 (0.007)	1.19	0.01295	0.85	0.82 (0.009)	1.21	0.01601
	443	0.90	0.82 (0.007)	1.21	0.01438	0.86	0.82 (0.009)	1.22	0.01720
	488	0.86	0.77 (0.008)	1.17	0.00891	0.80	0.80 (0.011)	1.19	0.01166
	531	0.79	0.69 (0.009)	1.13	0.00513	0.69	0.75 (0.013)	1.16	0.00726
	547	0.79	0.65 (0.009)	1.08	0.00390	0.69	0.75 (0.013)	1.16	0.00578
	667	0.77	0.67 (0.010)	1.13	0.00888	0.67	0.74 (0.014)	1.16	0.01261

[a]We used all available wavelengths in the inversion (left columns) and only MODISA visible wavelengths in the inversion (right columns). The sample size (N) was 1437. r^2 is the regression coefficient of determination. *Slope(SE)* is the regression slope and standard error. We calculated *Ratio* as *median* (ORM/truth). *RMSE* is the RMSE (the root of the residual mean square, in units equal to that of the observation).

when using only MODISA wavelengths, with r^2, regression slopes, and median ratios ranging from 0.67 to 0.94, 0.74 to 1.02, and 0.88 to 1.22, respectively. The corresponding mean r^2, regression slopes, and ratios degraded slightly to 0.82, 0.84, and 1.09. The most notable statistical difference between the two wavelength suites was an increase in root mean square error (RMSE) of 69%, 45%, and 35% for $b_{bp}(\lambda)$, $a_{pg}(\lambda)$, and $a_{\phi}(\lambda)$, respectively, when using only MODISA wavelengths (Table 2).

Our results suggest that ORM performance increases with finer $R_{rs}(\lambda)$ spectral resolution, most notably with regard to retrieval variability (via RMSE). However, the remaining statistics (r^2, regression slopes, and ratios) suggest that data records from multispectral satellite instruments, such as MODISA, provide reliable ORM-derived IOPs on average for a significant dynamic range of water types. In general, these validation statistics fall well within the range of those presented in previous ORM studies [20,21]. To verify this, we re-ran our analyses using all available wavelengths as input into the quasi-analytical algorithm (QAA) of Lee [46]. The r^2, regression slopes, ratios, and RMSEs for QAA were: 0.93, 1.22, 1.18, and 0.0051 for $b_{bp}(443)$; 0.95, 1.23, 1.09, and 0.01617 for $a_{pg}(443)$; and, 0.93, 0.83, 0.97, and 0.01220 for $a_{\phi}(443)$. More often than not, the performance of our ORM matched or exceeded that of QAA, which confirms to a first-order that our ORM yields results comparable with those of alternate configurations [21] (Table 2).

C. Optically Weighted C_{ϕ}

The ORM demonstrated dependence on the depth of the subsurface maxima in its ability to estimate absolute magnitudes of *N. miliaris*. Before proceeding with our interpretation of this dependency, we feel it worthwhile to remind the reader how vertical structure in a water column constituent shapes its optically weighted value and the corresponding $R_{rs}(\lambda)$ [38,39]. Multiple variants of $C_{\phi N}(z)$ yielded common C_{ϕ}, the optically integrated values measured by an ocean color satellite instrument. Figure 5 presents C_{ϕ} as a function of bloom depth (N^Z) and magnitude ($N^{\max}$) for a subset of our synthesized stations with constant bloom thickness (N^W), $C_{\phi D}$, $a_d(443)$, $a_g(443)$, and $b_d(555)$. The deepest variants of $N_{\max}$ (3 and 6 mg m^{-3}), for example, produced C_{ϕ} that matched the shallower variants of $N_{\max}$ (0.5 and 1 mg m^{-3}) (Table 3). Generally speaking, C_{ϕ} for *N. miliaris* blooms of all magnitudes converged to the background diatom signal as N^Z increased. The weighted C_{ϕ} increasingly underestimated $N^{\max}$ as the subsurface maximum increased in magnitude and deepened. For the smallest $N^{\max}$, C_{ϕ} depended most significantly on the background $C_{\phi D}$ and showed very little dependence on the depth and thickness of the subsurface *N. miliaris* maxima (Table 3). Ultimately, the magnitude of the peak subsurface population exceeded its remote-sensing value (e.g., $N^{\max} > C_{\phi}$) for most examples.

Changing the vertical position of a subsurface bloom of fixed size and thickness yielded different C_{ϕ}. Consider, for example, the profiles with $N_{\max} =$ 6 mg m^{-3} in Figure 5(A) (red circles). The corresponding optically weighted C_{ϕ} dropped from 5.58 to 0.75 mg m^{-3} as this synthesized bloom progressed from the surface to a depth of 15 m (Table 3). Given that standard ocean color algorithms are tuned to optically weighted constituent stocks [18,61], it stands

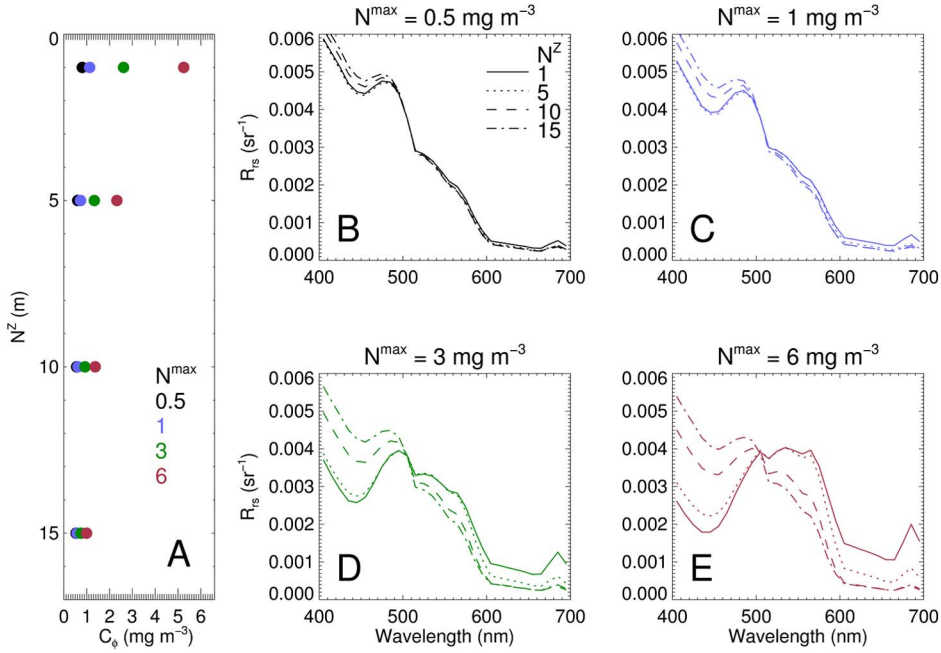

Fig. 5. Optically weighted C_ϕ and corresponding $R_{rs}(\lambda)$ as a function of *N. miliaris* bloom depth (N^Z) and magnitude (N^{max}). This subset of simulations has constant $C_{\phi D}$ (= 0.5 mg m^{-3}), N^W (= 5 m), $a_g(443)$ (= 0.005 m^{-1}), $a_d(443)$ (= 0.002 m^{-1}), and $b_p(555)$ (= 0.1 m^{-1}). Panel (A) shows optically weighted C_ϕ versus bloom depth, with colors representing different bloom magnitudes. Panel (B) shows $R_{rs}(\lambda)$ for a bloom magnitude of 0.5 mg m^{-3}, with different line styles representing different bloom depths. Panels (C)–(E) follow Panel (B), but for bloom magnitudes of 1, 3, and 6 mg m^{-3}. Colors in Panels (B)–(E) follow Panel (A) for clarity.

to reason that satellite retrievals of these stocks do not independently provide information about the vertical distribution of the constituents or the thicknesses of their layers. To circumvent this, previous studies used external hydrographic and environmental information to make assumptions about vertical distributions [27,62]. Uitz *et al.* [27], for example, suggested that the vertical structure of $C_\phi(z)$ could be inferred on average in the open ocean from its absolute magnitude and some knowledge of the local average mixed layer depth. Not surprisingly, the rate of change of C_ϕ with bloom depth (N^Z) varied strongly with the magnitude of the subsurface maximum (N^{max}).

Vertically displacing the subsurface *N. miliaris* maxima resulted in different $R_{rs}(\lambda)$ [Figs. 5(B)–5(E)]. The depths of subsurface features and the bulk attenuating (absorbing and scattering) properties of the water mass modulate the average contributions of each constituent to $R_{rs}(\lambda)$ [56,57]. Consider again the $N^{max} = 6$ mg m^{-3} in Fig. 5 (red circles). The color of the water appeared green under a near-surface bloom and progressively bluer as this bloom sank deeper into the water column [Fig. 5(E)]. All N^{max} demonstrated this, but the effect lessened as it decreased. For the case study presented in Fig. 5, $R_{rs}(\lambda)$ appeared similar for the multiple scenarios with C_ϕ between 0.6 and 0.75 mg m^{-3} (compare, e.g., the dashed lines in Figs. 5(B) and 5(C) with the dashed–dotted lines in Figs. 5(D) and 5(E). Within the full population of simulations, however, $R_{rs}(\lambda)$ differed for many instances of common C_ϕ. Fig. 3(B) shows $R_{rs}(\lambda)$ for simulations with C_ϕ near ~1 mg m^{-3}, many of which vary significantly in the blue–green part of the spectrum. This variability in $R_{rs}(\lambda)$ for common C_ϕ contributes to the overall variability seen in algorithm development data sets used to develop common bio-optical models (see, e.g., the vertical scatter of the radiometric data for each given optically weighted C_ϕ presented in Figs. 3 and 6 of O'Reilly [18] and Fig. 5(A) of Werdell and Bailey [61]).

D. Retrieval of Multiple Phytoplankton Species

The ORM demonstrated decent skill in identifying *N. miliaris* and diatoms, but with sufficient

Table 3. Optically weighted C_ϕ, $K_d(490)$, and z_{90} for the Example Simulations Presented in Fig. 5

N^{max} (mg m^{-3})	N^Z (m)	C_ϕ (mg m^{-3})	$K_d(490)$ (m^{-1})	z_{90} (m)
0.5	1	0.69	0.052	19.0
0.5	5	0.69	0.054	18.0
0.5	10	0.61	0.054	18.0
0.5	15	0.56	0.054	18.0
1	1	0.89	0.056	17.5
1	5	0.90	0.061	16.0
1	10	0.72	0.061	16.0
1	15	0.61	0.058	17.0
3	1	2.04	0.081	12.0
3	5	1.89	0.110	9.0
3	10	1.01	0.086	11.5
3	15	0.69	0.065	14.5
6	1	5.58	0.211	4.5
6	5	2.74	0.168	5.5
6	10	1.17	0.096	9.5
6	15	0.75	0.070	13.5

variability to indicate that the inversion cannot provide absolute magnitudes for both species at all times. All ORMs most effectively retrieve total absorption (e.g., [20,21]). Deconstructing total absorption into its CDOM + NAP and phytoplankton components remains more uncertain (Table 2) and, not surprisingly, decomposing the total phytoplankton signature into multiple components carries additional uncertainties [23,24,29]. In general, the ORM tended to underestimate *N. miliaris* and to overestimate diatoms. When using all wavelengths in the inversion, the ORM-to-synthesized bias, RMSE, and regression slope changed from 0.093 mg m^{-3}, 0.261 mg m^{-3}, and 0.82 for total C_ϕ to 0.097 mg m^{-3}, 0.215 mg m^{-3}, and 0.93 for $C_{\phi D}$ and −0.005 mg m^{-3}, 0.310 mg m^{-3}, and 0.64 for $C_{\phi N}$ (Fig. 6). We calculated bias as the average difference between modeled and measured values. The differences amplified when using only MODISA wavelengths, with biases, RMSEs, and slopes changing from 0.094 mg m^{-3}, 0.313 mg m^{-3}, and 0.82 for C_ϕ to 0.113 mg m^{-3}, 0.402 mg m^{-3}, and 1.02 for $C_{\phi D}$ and −0.019 mg m^{-3}, 0.345 mg m^{-3}, and 0.55 for $C_{\phi N}$. The ORM exclusively underestimated *N. miliaris* at concentrations greater than 2 mg m^{-3}.

The ORM frequently identified monospecific phytoplankton populations, where only *N. miliaris* or only diatoms were present. For the former, the ORM routinely detected *N. miliaris*, but often also returned false-positives for diatoms [Figs. 7(A) and 7(B)]. In general, the ORM underestimated $C_{\phi N}$ (most significantly when the *N. miliaris* bloom resided at the surface), which resulted in an overestimation of $C_{\phi D}$. The lower cluster of $C_{\phi N} \geq 2$ mg m^{-3} in Figs. 7(A) and 7(B) corresponds to the $C_{\phi D} \geq 0.5$ mg m^{-3} in these panels. Reducing the spectral resolution of $R_{rs}(\lambda)$ to MODISA wavelengths amplified the underestimation of $C_{\phi N}$ for concentrations greater than 4 mg m^{-3} and increased the magnitudes of false positives for diatoms. When only diatoms were present, however, the ORM never significantly detected *N. miliaris* (always <0.2 mg m^{-3}) and successfully estimated $C_{\phi D}$ over its full dynamic range [Figs. 7(C) and 7(D)]. Using only MODISA wavelengths did not significantly degrade these results. In general, the ORM did not identify *N. miliaris* when it was absent and often identified *N. miliaris* when it was present. Although the quantification of $C_{\phi N}$ and $C_{\phi D}$ may be imperfect, this suggests that a positive ORM-derived $C_{\phi N}$ reliably indicates the presence of *N. miliaris* somewhere in the upper water column.

Equally mixed populations of *N. miliaris* and diatoms produced $R_{rs}(\lambda)$ that occasionally challenged the ORM. Consider the following scenarios presented in Fig. 8: monospecific $C_{\phi D} = 0.5$ mg m^{-3}, monospecific $C_{\phi N} = 0.5$ mg m^{-3}, monospecific $C_{\phi D} = 1$ mg m^{-3}, monospecific $C_{\phi N} = 1$ mg m^{-3}, and mixed $C_{\phi D} = C_{\phi N} = 0.5$ mg m^{-3} (total $C_\phi = 1$ mg m^{-3}). The four monospecific populations produced sufficiently unique $R_{rs}(\lambda)$ such that the ORM effectively

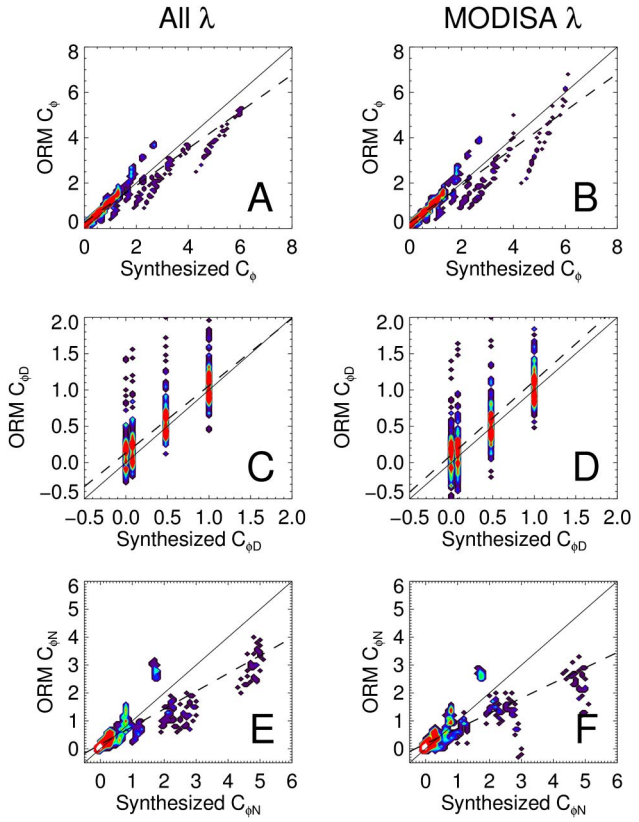

Fig. 6. Comparison of ground-truth (synthesized) and ORM-derived C_ϕ (A) and (B), $C_{\phi D}$ (C) and (D), and $C_{\phi N}$ (E) and (F) using all available wavelengths in the inversion (left column) and only MODISA visible wavelengths in the inversion (right column). Colors as in Fig. 4.

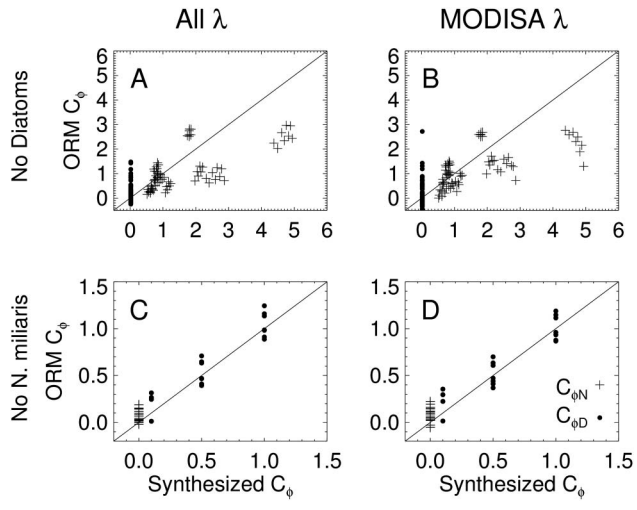

Fig. 7. Comparison of ground-truth (synthesized) and ORM-derived $C_{\phi D}$ and $C_{\phi N}$ for a mono-species subset of simulations using all available wavelengths in the inversion (left column) and only MODISA visible wavelengths in the inversion (right column). We considered only synthesized $C_{\phi D} = 0.02$ and $C_{\phi N} \geq 0.5$ mg m^{-3} in panels (A) and (B). We considered only synthesized $C_{\phi D} \geq 0.1$ and $C_{\phi N} = 0$ mg m^{-3} in panels (C) and (D). Crosses and filled circles show $C_{\phi N}$ and $C_{\phi D}$, respectively.

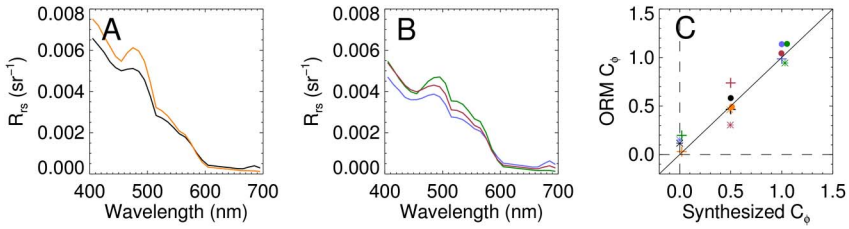

Fig. 8. Five example simulated $R_{rs}(\lambda)$ and their corresponding ORM-derived C_ϕ. Each simulation was created using $a_d(443) = 0.002$, $a_g(443) = 0.005$, and $b_d(555) = 0.1\ m^{-1}$. Panel (A) presents spectra for $C_{\phi D} \sim 0.5\ mg\,m^{-3}$ without *N. miliaris* (black) and $C_{\phi N} \sim 0.5\ mg\,m^{-3}$ without diatoms (orange). Panel (B) presents spectra for $C_{\phi D} \sim 1\ mg\,m^{-3}$ without *N. miliaris* (blue), $C_{\phi N} \sim 1\ mg\,m^{-3}$ without diatoms (green), and a mixed population with $C_{\phi D}$ and $C_{\phi N} \sim 0.5\ mg\,m^{-3}$ each (red). Panel (C) presents ground-truth (synthesized) versus ORM comparisons, with colors referring to the corresponding $R_{rs}(\lambda)$, and circles, crosses, and asterisks indicating C_ϕ, $C_{\phi D}$, and $C_{\phi N}$, respectively.

estimated C_ϕ, $C_{\phi N}$, and $C_{\phi D}$ without false positives for the absent species. In the mixed population, where *N. miliaris* and diatoms both contributed equally to C_ϕ, the ORM achieved accurate C_ϕ, but less effectively separated the two phytoplankton populations. The ORM attributed much of the C_ϕ to $C_{\phi D}$ ($0.75\ mg\,m^{-3}$), which is consistent with our previous general observation that the ORM tends to slightly over- and underestimate diatoms and *N. miliaris*, respectively. Equal contributions by *N. miliaris* and diatoms did not produce $R_{rs}(\lambda)$ that fell equally between the two monospecific spectra, suggesting to a first-order that the contributions of the two species to $R_{rs}(\lambda)$ were not strictly additive. The case studies with equivalent C_ϕ presented in Figs. 5(B) and 5(C) show a clean progression in $R_{rs}(\lambda)$ as a bloom deepened in the water column and the color of the water transitioned from green to blue. In contrast, the spectra in Fig. 8(B) do not demonstrate a clean transition as species mix.

The accuracy of the ORM showed dependency on the depth of the *N. miliaris* bloom when considering all available wavelengths and only MODISA wavelengths. On average, this dependency did not vary with the total magnitude of C_ϕ (Table 4). For the full population of synthesized data, the ORM consistently under- and over-estimated $C_{\phi N}$ and $C_{\phi D}$, respectively, when $N^Z = 1$ m (Fig. 9). This under- and overestimation of $C_{\phi N}$ and $C_{\phi D}$ generally reversed when N^Z deepened. In all cases, the differences in $C_{\phi N}$ exceeded those for $C_{\phi D}$. With one exception (the $C_\phi > 1\ mg\,m^{-3}$ subset at 15 m), the largest positive differences in $C_{\phi N}$ appeared at 5 m and decreased as the subsurface maxima deepened. Overall, the ORM overestimated *N. miliaris* by 23.8% to 77.0%, on average, for $N^Z \geq 5$ m. Absolute biases in $C_{\phi N}$ decreased with increasing bloom depth, which is not surprising given that the (optically weighted) total C_ϕ also decreased with increasing bloom depth (see, e.g., Fig. 5). In contrast to *N. miliaris*, the largest differences in $C_{\phi D}$ occurred near the surface. At depth, ORM retrievals of $C_{\phi D}$ differed from ground truth by <8%, with only one exception (the $C_\phi < 1\ mg\,m^{-3}$ subset at 15 m), and maintained null absolute biases ($<0.07\ mg\,m^{-3}$).

4. Discussion

A. Interpreting the ORM Retrievals

Ocean color satellite instruments provide rich data streams for studying phytoplankton dynamics. The community considers these data streams to be sufficiently critical to require that upcoming ocean color satellite programs include capabilities to discriminate between phytoplankton groups in support of advanced carbon studies [14,15]. Our study focused on a class of algorithm known as ORMs (also known as semi-analytic algorithms). Few studies have demonstrated the ability of an ORM to successfully deconvolve $a_\phi(\lambda)$ into contributions by individual phytoplankton groups over a large range of environmental conditions [37]. In practice, ORMs are currently applied to satellite measurements of $R_{rs}(\lambda)$ without exact *a priori* knowledge of phytoplankton physiological state or environmental conditions for a given pixel. They operate assuming that their

Table 4. Median Relative Percent Differences (MPD) and Absolute Biases between the Ground-Truth-Synthesized and ORM-Derived $C_{\phi N}$ and $C_{\phi D}$[a]

		$C_{\phi N}$		$C_{\phi D}$	
	N^Z	MPD	Bias	MPD	Bias
All C_ϕ	1	−30.6	−0.173	19.5	0.140
	5	73.5	0.231	−4.4	−0.031
	10	56.8	0.073	1.9	0.005
	15	35.5	0.023	7.9	0.030
$C_\phi < 1\ mg\,m^{-3}$	1	−23.3	−0.061	13.5	0.030
	5	73.8	0.150	1.9	−0.001
	10	52.3	0.058	−0.7	−0.001
	15	29.8	0.017	13.5	0.034
$C_\phi > 1\ mg\,m^{-3}$	1	−33.3	−0.677	27.1	0.181
	5	70.8	0.402	−7.0	−0.068
	10	60.7	0.119	2.8	0.028
	15	77.0	0.055	−0.1	−0.001

[a]Shown as a function of the depth of the subsurface *N. miliaris* maxima (N^Z). Differences were calculated as in Fig. 8. Biases were calculated as the average difference between modeled and ground-truth values. Units for MPD and bias are % and $mg\,m^{-3}$, respectively. Statistics are presented for all simulations, simulations with $C_\phi < 1\ mg\,m^{-3}$, and simulations with $C_\phi > 1\ mg\,m^{-3}$.

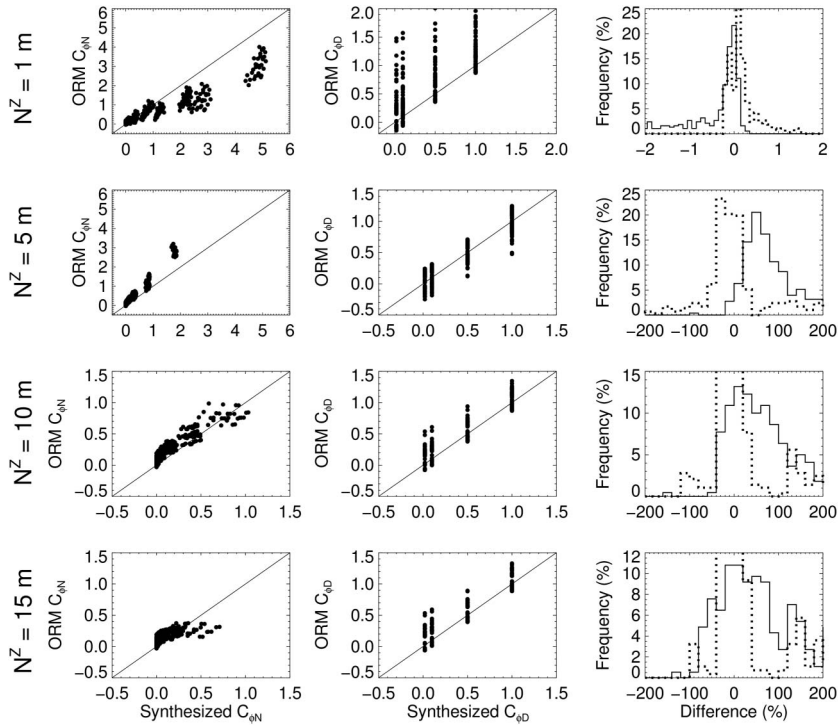

Fig. 9. Comparisons of ground-truth (synthesized) and ORM-derived $C_{\phi N}$ (left column) and $C_{\phi D}$ (center column) using all available wavelengths in the inversion. Results are stratified by depth of the *N. miliaris* subsurface maxima, with $N^Z = 1$, 5, 10, and 15 m shown from top to bottom. The right column presents frequency distributions of relative percent differences, calculated as 100%* (*ORM*/*synthesized*-1). Solid and dotted lines indicate differences for $C_{\phi N}$ and $C_{\phi D}$, respectively.

adopted eigenvectors adequately represent local conditions at the moment the measurement is made. We constructed a controlled environment where the eigenvectors used to parameterize the ORM were the same as those used to build the input $R_{rs}(\lambda)$. Doing so generated scenarios for which we had exact (simulated) knowledge of phytoplankton physiological state and environmental conditions. Comparisons of ORM retrievals and simulation inputs within this ideal environment should reveal weaknesses in ORM operation and interpretation that merit subsequent review. To ensure broad applicability of our results, we conducted our analyses using a very common ORM algorithmic form (built using the GIOP framework [21] and analogous in form to that of, e.g., Maritorena *et al.* [19], Roesler *et al.* [35], and Westberry *et al.* [37]).

Our results reaffirm previous conclusions that ORMs effectively separate the total signal of the collective phytoplankton community from other water column constituents, such as CDOM and NAP [20,21] (Fig. 4, Table 2). Our results suggest, however, that the absolute accuracy of ORM retrievals degrades when further deconstructing the derived C_ϕ and IOPs into additional subcomponents, even under perfectly controlled conditions (Fig. 6). Despite this, we view our results as positive. Although the absolute magnitudes of $C_{\phi N}$ and $C_{\phi D}$ maintained biases, the ORM successfully detected whether or not *N. miliaris* appeared in the simulated water column for both the limited (MODISA) and expanded (PACE-like) wavelength suites (Figs. 6–8). This quantitatively calls for caution when interpreting the absolute magnitudes of the retrievals, but qualitatively suggests that the ORM provides a robust mechanism for identifying the presence or absence of *N. miliaris*.

Let us briefly explore this suggestion further. A positive retrieval of $C_{\phi N}$ reliably indicated the presence of *N. miliaris* at some near surface (≤15 m) position in our simulations. The ORM effectively identified diatoms and often identified *N. miliaris* in mono- and mixed-populations, but often with imperfect absolute magnitudes (e.g., $C_{\phi N}$ showed negative biases when the bloom maxima resided near the surface and positive biases when it deepened). Despite this, *N. miliaris* was always present when the ORM retrieved positive $C_{\phi N}$. The ORM did fail to identify *N. miliaris* on occasion and produced corresponding false positives for diatoms, but it never substantially identified *N. miliaris* for simulations with only diatoms. In other words, when *N. miliaris* was present, the ORM often (but not always) estimated $C_{\phi N} > 0$ mg m^{-3}, but when *N. miliaris* was not present, the ORM never estimated $C_{\phi N} \gg 0$ mg m^{-3}. At a minimum then, this suggests that the ORM can be used to indicate whether or not *N. miliaris* is present in a satellite pixel. Indeed, we explored the qualitative use of satellite-derived $C_{\phi N}$ in a companion study via application of the ORM to MODISA to simply identify and catalog pixels with substantially positive $C_{\phi N}$.

To a first-order, applying this (binary absence or presence) identification scheme to the full MODISA data record permitted us to study the spatial and temporal distribution of *N. miliaris* appearances over a decade in the northern Arabian Sea [63].

B. Phytoplankton Absorption Eigenvectors

Both spectral resolutions yielded comparable regression statistics for the retrieved $b_{bp}(\lambda)$, $a_{pg}(\lambda)$, and $a_{\phi}(\lambda)$, as evidenced by r^2, regression slopes, and median ORM-to-synthesized ratios (Fig. 4, Table 2). Using all available wavelengths enabled somewhat improved separation of $C_{\phi D}$ and $C_{\phi N}$ (Fig. 6). The variability in all IOP and C_{ϕ} retrievals (not just $C_{\phi D}$ and $C_{\phi N}$), however, increased significantly when using $R_{\mathrm{rs}}(\lambda)$ with reduced spectral resolution (see, e.g., the RMSE in Table 2). The slopes from 400–440 nm and from 440–510 nm differ substantially for $a^*_{\phi D}(\lambda)$ and $a^*_{\phi N}(\lambda)$ (Fig. 1). The MODISA wavelength suite maintains less spectral information in these spectral ranges (and, thus, fewer degrees of freedom in the inversion) with which to discriminate between *N. miliaris* and diatoms. In other words, the optical signatures of $a^*_{\phi D}(\lambda)$ and $a^*_{\phi N}(\lambda)$ contrast less when viewed at only MODISA wavelengths versus at PACE-like wavelengths. This has implications for planning new satellite missions with mandates to deliver data records with improved capabilities for discriminating between phytoplankton species [14,15]. Certainly, the SeaWiFS and MODISA data records (and others) stimulated innovation and progress toward the remote identification of phytoplankton groups [22–31,34,36,37,40]. Given our results, however, we expect that new satellite instruments with finer spectral resolution will enable more quantitative resolution of multiple species with fewer uncertainties [32,35] (Figs. 6 and 7).

Single $a^*_{\phi}(\lambda)$ cannot perfectly represent any phytoplankton species under all conditions of growth, nutrient availability, and ambient light [29,64]. This presents problems for ocean color remote sensing, as the phytoplankton population under observation and its physiological state cannot be known *a priori*. Previous studies adopted varied methods of expressing phytoplankton absorption: (1) single $a^*_{\phi}(\lambda)$ derived from global assemblies of *in situ* measurements [19,46]; (2) multiple $a^*_{\phi}(\lambda)$ to simultaneously represent several phytoplankton size classes [23,24,29,35]; (3) dynamic assignment of $a^*_{\phi}(\lambda)$ based on estimates of C_{ϕ} [21]; and (4) iterations on ranges of $a^*_{\phi}(\lambda)$ [65,66]. Our ORM generically falls into method (2), but might realize future benefits from being refined to the forms of (3) or (4). Additional *in situ* measurements from various stages of diatom and *N. miliaris* blooms during the NEM could be used to call the ORM in an iterative fashion using ranges of observed eigenvectors. More sophisticated parameterizations of $a_{\phi}(\lambda)$ might also be developed to replace the linear expression we adopted ($= C_{\phi} a^*_{\phi}(\lambda)$) [64,67]. Nevertheless, these approaches all require some *a priori* assumption of phytoplankton spectral shape(s). The need to make such an assumption argues in favor of retrospective analysis when interpreting satellite time-series of C_{ϕ} and $a_{\phi}(\lambda)$ to verify that local hydrodynamics, environmental conditions, and biology support the *a priori* selection of eigenvectors.

With respect to the mechanics of our analyses, our choice in $a^*_{\phi}(\lambda)$ remains irrelevant. However, they merit additional discussion in the context of interpretation of satellite data records. We generated our simulated IOP profiles and configured the ORM using consistent $a^*_{\phi}(\lambda)$. Despite this, the ORM retrievals did not always exactly reproduce the inputs $C_{\phi N}$ and $C_{\phi D}$. We chose diatoms and *N. miliaris* because they best represent significantly different end-members in the Arabian Sea, and because the ability to detect *N. miliaris* using long-term satellite data records presents an opportunity to quantify its emerging presence over the past decade. We know, however, that additional phytoplankton species coexist and flourish in the Arabian Sea, such as dinoflagellates and *Trichodesmium spp.* Unfortunately, the ORM became unstable when we added a third phytoplankton component (raising the unknowns to five), particularly when limiting the runs to MODISA wavelengths. In the two-component system, however, certain combinations of our $a^*_{\phi N}(\lambda)$ and $a^*_{\phi D}(\lambda)$ yielded spectra that resembled that of dinoflagellates (see Thibodeau *et al.* [7]). Following, in scenarios where $a^*_{\phi}(\lambda)$ within an ORM do not represent the population under observation, the ORM defaults to combinations of its native phytoplankton components and misidentifies the species present. This supports our recommendation of caution when interpreting the absolute magnitudes of multiple phytoplankton eigenvalues. It also argues that, for targeted application of ORMs, a configuration for alternate phytoplankton species would produce more reliable and appropriate IOPs outside of the NEM, when we would not expect an abundance of *N. miliaris*.

Our choice in $a^*_{\phi D}(\lambda)$ and $a^*_{\phi N}(\lambda)$ contributed in part to the vertical biases seen in $C_{\phi N}$. Concentrations of absorbing constituents control the shape of $R_{\mathrm{rs}}(\lambda)$, whereas the magnitude of scattering relative to absorption modulates the brightness. Near the surface, pure seawater absorption dominates the total absorption signal at wavelengths 600 nm [51]. As CDOM and NAP contribute very little to total absorption in this spectral range [48], phytoplankton begin to shape $R_{\mathrm{rs}}(\lambda)$ in the absence of these other water column constituents as blooms near the surface. Blooms of sufficient magnitude also enhance backscattering and can notably fluoresce [68], which elevates red $R_{\mathrm{rs}}(\lambda)$, as demonstrated in Figs. 5(D) and 5(E). In these figures, the ratio of $R_{\mathrm{rs}}(443)$-to-$R_{\mathrm{rs}}(670)$ decreases toward unity as the simulated bloom maxima shallows. The $a^*_{\phi D}(\lambda)$ and $a^*_{\phi N}(\lambda)$ we adopted also maintain substantially different blue-to-red ratios, with diatoms more closely approaching a ratio of one (Fig. 1). We expect that these near-unity blue-to-red relationships favored the retrieval of $C_{\phi D}$ when $N^Z = 1$ m. This

favoritism switched to $C_{\phi N}$ as the bloom maxima deepened and the red $R_{rs}(\lambda)$ became water-dominated instead of phytoplankton-dominated.

Several of our processing choices complicate this interpretation of vertical biases. First, as clearly demonstrated near 683 nm in Figs. 5(D) and 5(E), we enabled the fluorescence capabilities of HE5. A discussion of the fluorescence model incorporated into HE5 exceeds the scope of this manuscript. In contrast to HE5, however, our ORM (as well as most others) does not account for phytoplankton fluorescence [20,21]. To our knowledge, most (quasi-single scattering) relationships between apparent and IOPs, e.g., the coefficients in Eq. (2), or simplify contributions by fluorescence to $R_{rs}(\lambda)$. Second, the Levenberg–Marquardt inversion scheme employs a chi-squared cost function that minimizes relative differences between the fit and measured $R_{rs}(\lambda)$ with equal consideration of all wavelengths [21]. Given the weakness of the red $R_{rs}(\lambda)$ signal in most open ocean waters, small absolute differences between fit and measured spectra amplify into substantial relative differences, which can overweight this region of the spectrum. Previous studies partially addressed this by constraining their inversion to 412–555 nm [19,65]. Doing so using our simulations increased the variability of our bulk-derived IOPs. For example, the RMSEs for $b_{bp}(443)$ and $a_{pg}(443)$ rose from 0.00031 to 0.00056 and 0.0125 to 0.0167, respectively. We recommend that subsequent studies explore revised cost functions that consider both absolute and relative goodness-of-fits [15] and/or alternate spectral weighting schemes (see e.g., the discussion of this in Werdell *et al.* [21]).

C. Future Directions

We conclude with a discussion of several general paths forward: (1) refining the ORM parameterization and evaluating alternate configurations (e.g., following Werdell *et al.* [21]); (2) using other ORM products, namely $b_{bp}(\lambda)$, to help detect *N. miliaris*; and (3) re-exploring the role of vertical structure in interpretations of ocean color data records. With regard to the ORM itself, we reaffirmed that a GIOP-like form can produce accurate estimates of bulk IOPs [$b_{bp}(\lambda)$, $a_{dg}(\lambda)$, and $a_{\phi}(\lambda)$] as has been repeatedly demonstrated (e.g., [19–21,35,37]). The estimates showed biases, whereas our ORM effectively discriminated between C_{ϕ} from diatoms and *N. miliaris*. With regard to improving the latter, reconsidering our treatment of phytoplankton absorption remains an obvious next step as discussed in Section 4.B. Other ORM parameterizations could also be refined, several of which we briefly explored. Using linear matrix inversion [65,66] *in lieu* of the nonlinear Levenberg–Marquardt inversion scheme yielded almost identical results, but with fewer overall viable retrievals. Alternate forms of $b^*_{bp}(\lambda)$ and $a^*_{dg}(\lambda)$ could also be adopted [31,46]. That said, ORM-retrieved $C_{\phi N}$ demonstrated no significant dependence on the magnitudes of $a_g(\lambda)$, $a_d(\lambda)$, or $b_d(\lambda)$ (Fig. 10). Higher $a_g(443)$ and $a_d(443)$ corresponded to slightly elevated ORM-derived $C_{\phi N}$, but these differences fell well below variances seen elsewhere in our analyses, such as by N^Z. We expect to pursue incorporating pixel classification [69] and ensemble (iterative or bootstrapping) [65,66] schemes into future studies, both of which provide vehicles for considering ranges of eigenvectors within the ORM.

Our analyses attempted to discriminate between diatoms and *N. miliaris* using phytoplankton absorption coefficients and C_{ϕ}. However, differences in diatom and *N. miliaris* backscattering properties could also be exploited to identify these species. Changes in the $b_{bp}(\lambda)$-to-C_{ϕ} ratio over time, for example, would indicate changes in the optics (e.g., average particle size) or biology (e.g., phytoplankton or nonpigmented grazers). Figure 11 presents the relationship between $b_{bp}(443)$ and C_{ϕ} as a function of the abundance of *N. miliaris* for both the simulated data set (panel A) and the ORM (panel B). The linear pattern of $b_{bp}(443)$ versus C_{ϕ} results from our parameterization of the simulated profiles (Table 1). Two features stand out. First, the simulated and ORM relationships exhibit consistent patterns in $C_{\phi N}$, dynamic ranges, and slopes. The simulated relationships represent a forward solution for the exact radiative transfer equations (RTE) while the ORM relationships represent an inversion solution for a simplified (single scattering) approximation of the RTE. The similarities in the patterns suggest that the ORM provides an adequate approximation of the RTE, with the smearing of points [relative to the clean slopes in Fig. 11(A)] resulting from the approximations made. Second, $b_{bp}(443)$ changes with $C_{\phi N}/C_{\phi}$;

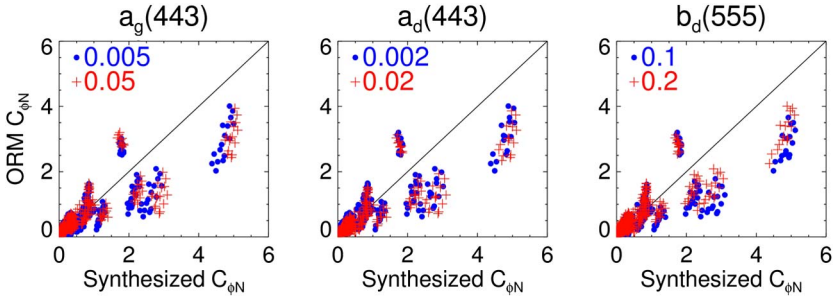

Fig. 10. Comparisons of ground-truth (synthesized) and ORM-derived $C_{\phi N}$ using all available wavelengths in the inversion. Results are stratified by the varied $a_g(443)$ (left), $a_d(443)$ (center), and $b_d(555)$ (right) used to generate the synthesized IOP profiles and $R_{rs}(\lambda)$.

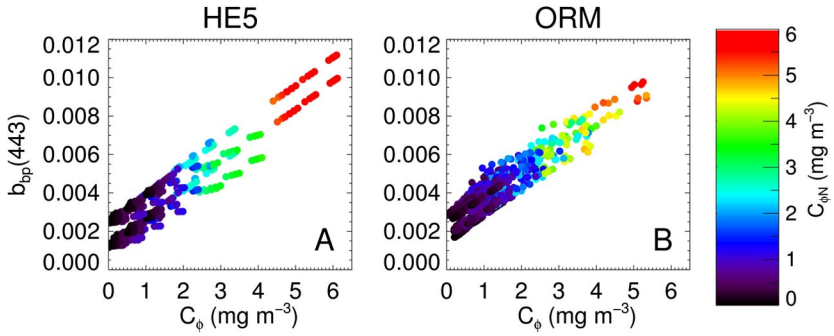

Fig. 11. C_ϕ versus $b_{bp}(443)$. Panel (A) shows C_ϕ from our simulated profiles versus backscattering from HE5. Panel (B) shows C_ϕ from the ORM versus backscattering from the ORM. Colors indicate the magnitude of $C_{\phi N}$ in mg m^{-3}.

although the direction of the change appears to reverse several times as C_ϕ rises from 0 to 6 mg m^{-3}. In practice, such a pattern could provide a useful indicator of *N. miliaris* if robust patterns in $b_{bp}(\lambda)$ versus $C_{\phi N}/C_\phi$ emerge. However, as no consensus covariance emerged from our synthesized data set, similar analyses of *in situ* data (as they become available) are warranted.

Synthesized data cannot represent all conditions at all times. We could endlessly refine the parameterizations used to generate the simulated IOP profiles (Table 1); however, the values we selected correspond closely to field measurements and produce complementary $R_{rs}(\lambda)$ to those measured *in situ* (Fig. 2) [7,40]. Regarding the role of scattering, we assumed that the *N. miliaris* and diatom species present during the NEM are large relative to visible wavelengths (thus, our choice of $\eta = 0$) [70,71]. Optically, *N. miliaris* remains poorly studied relative to other phytoplankton [49]; however, preliminary studies suggest inefficient light scattering, consistent with its cell-specific properties [40] (thus, our choice of $b^*(555) = 0.15$ m^2 mg^{-1}). Although our constant $b_b(\lambda)/b(\lambda)$ of 0.01 remains a vetted global average, we also acknowledge that this ratio ranges from 0.005 to 0.02 over a range of oceanographic conditions [53,72]. Currently, we have no knowledge of the scattering phase function of *N. miliaris*, but expect it might differ from that embedded in the Monte Carlo simulations used to generate the apparent-to-IOP relationship expressed in Eq. (2) [47]. An optical closure study following Tzortziou *et al.* [73] would be prudent as subsequent work to better constrain $b_b(\lambda)/b(\lambda)$ (and, thus, the scattering phase function) of species present during the NEM.

Finally, it remains well-known within the oceanographic community that ocean color estimates of water column properties represent optically weighted contributions that do not independently provide information on vertical structure [27,56,62]. While we did not originally intend to revisit this, our simulations provided an opportunity to pose a reminder and offer two related suggestions with potential for advancing our collective ability to remotely sense phytoplankton community composition. First, if the community wants true estimates of vertical structure from space, alternate technologies need to be developed [74] and/or methods that make use of complementary environmental information need to be nurtured [27,62]. The case studies presented in Fig. 5 demonstrate that optically weighted values cannot independently describe the vertical position, subsurface magnitude and thickness, or age of a bloom. Furthermore, our results provided little evidence that the ORM better detects *N. miliaris* as its bloom expands in abundance or thickness. While the known biases with N^Z were quantifiable in this controlled study, the vertical position of *N. miliaris* detected from space will not be known *a priori* and cannot currently be independently inferred in all water masses at all times using passive ocean color radiometry [27,62]. In practice, this simply means that absolute abundances for multiple ORM-derive phytoplankton communities can only be compared with the understanding that they represent pseudo-depth-integrated values.

Second, reconstructing the optically weighted C_ϕ seen by a satellite instrument requires collecting biogeochemical and radiometric data with sufficient vertical resolution to capture heterogeneity over the layer defined by z_{90} [57]. The vertical composition and structure of the water column shapes what the satellite measures and, therefore, drives how *in situ* measurements should be collected and prepared for use in both ORM development (parameterization) and ocean color satellite data product validation. For a heterogeneous water column with deep C_ϕ maxima appearing within the first e-folding depth, near surface measurements alone cannot be accurately compared with satellite-derived data products [61]. In practice, this argues for specialized *in situ* data collection and processing to support ocean color satellite validation activities.

5. Conclusions

We pursued this work for two primary reasons: (1) to contribute to the growing foundation upon which advanced methods for remote phytoplankton detection are being built and (2) to begin developing a method to identify *N. miliaris* in the northern Arabian Sea. We explored the capabilities of an ORM to discriminate between two distinct phytoplankton communities under a diverse array of biophysical conditions. As in previous studies, the

ORM successfully separated the total phytoplankton signal from those of other water column constituents. The ORM effectively separated the individual contributions of *N. miliaris* and diatoms; however, the absolute estimates showed biases even under our perfectly controlled conditions. The vertical structure of our synthesized blooms highly influenced these biases, analysis of which reaffirmed that ORM retrievals alone cannot provide information on the vertical structure of a phytoplankton bloom. In the end, our results quantitatively call for caution when interpreting the absolute magnitudes of the retrievals, but qualitatively suggest that the ORM provides a robust mechanism for identifying the presence or absence of *N. miliaris*. Not surprisingly, incorporating $R_{rs}(\lambda)$ at additional wavelengths improved the quality of the ORM retrievals, underscoring the benefit of additional spectral information on forthcoming satellite instruments with mandates to produce phytoplankton community data products. In the short-term, we propose that our ORM (and other ORMs of this form) can adequately support studies that require only identifications of phytoplankton groups. In the long-term, we anticipate the quantitative skills of ORMs to improve as additional technologies, relevant *in situ* data, and environmental data get routinely included in modeling and data analysis activities.

Many thanks to Sean Bailey, Gene Feldman, Curt Mobley, Mary Jane Perry, Andrew Thomas, Huijie Xue, and Emmanuel Boss for their advice throughout the development of this manuscript. Thanks also to Susan Drapeau, Kelly Keebler, Helga do Rosario Gomes, Prahbu Matondkar, and colleagues at the National Institute of Oceanography, India for their support with field experiments and data analysis. Finally, many thanks to Stéphane Maritorena and an anonymous reviewer for their exceptional comments and attention to detail. Support for this work was provided through the NASA Ocean Biology and Biogeochemistry Program under research grant NNX11AE22G.

References

1. E. Martinez, D. Antoine, F. D'Ortenzio, and B. Gentili, "Climate-driven basin-scale decadal oscillations of oceanic phytoplankton," Science **326**, 1253–1256 (2009).
2. D. G. Boyce, M. R. Lewis, and B. Worm, "Global phytoplankton decline over the past century," Nature **466**, 591–596 (2010).
3. S. Henson, R. Lampitt, and D. Johns, "Variability in phytoplankton community structure in response to the North Atlantic Oscillation and implications for organic carbon flux," Limnol. Oceanogr. **57**, 1591–1601 (2012).
4. V. Smetacek and J. E. Cloern, "On phytoplankton trends," Science **319**, 1346–1348 (2008).
5. S. G. Parab, S. G. Prabhu Matondkar, H. D. R. Gomes, and J. I. Goes, "Monsoon-driven changes in phytoplankton populations in the eastern Arabian Sea as revealed by microscopy and HPLC pigment analysis," Continent. Shelf Res. **26**, 2538–2558 (2006).
6. H. D. R. Gomes, J. I. Goes, S. G. P. Matondkar, S. G. Parab, A. R. N. Al-Azri, and P. G. Thoppil, "Blooms of *Noctiluca miliaris* in the Arabian Sea–an *in situ* and satellite study," Deep Sea Res. I **55**, 751–765 (2008).
7. P. S. Thibodeau, C. S. Roesler, S. Drapeau, P. Matondkar, J. I. Goes, and P. J. Werdell, "Locating Noctiluca miliaris in the Arabian Sea: an evaluation of in situ multispectral fluorescent signatures," Limnol. Oceanogr. (to be published).
8. H. D. R. Gomes, J. I. Goes, S. G. P. Matondkar, E. Buskey, S. Basu, S. Parab, and P. Thoppil, "Massive outbreaks of Noctiluca scintillans blooms in the Arabian Sea due to spread of hypoxia," Nat. Commun. (submitted).
9. R. J. W. Brewin, N. J. Hardman-Mountford, S. J. Lavender, D. E. Raitsos, T. Hirata, J. Uitz, E. Devred, A. Bricaud, A. Ciotti, and B. Gentili, "An intercomparison of bio-optical techniques for detecting dominant phytoplankton size class from satellite remote sensing," Remote Sens. Environ. **115**, 325–339 (2011).
10. M. J. Behrenfeld, R. T. O'Malley, D. A. Siegel, C. R. McClain, J. L. Sarmiento, G. C. Feldman, A. J. Milligan, P. G. Falkowski, R. M. Letelier, and E. S. Boss, "Climate-driven trends in contemporary ocean productivity," Nature **444**, 752–755 (2006).
11. P. J. Werdell, S. W. Bailey, B. A. Franz, L. W. Harding, Jr., G. C. Feldman, and C. R. McClain, "Regional and seasonal variability of chlorophyll-a in Chesapeake Bay as observed by SeaWiFS and MODIS-Aqua," Remote Sens. Environ. **113**, 1319–1330 (2009).
12. S. A. Henson, J. L. Sarmiento, J. P. Dunne, L. Bopp, I. Lima, S. C. Doney, J. John, and C. Beaulieu, "Detection of anthropogenic climate change in satellite records of ocean chlorophyll and productivity," Biogeoscience **7**, 621–640 (2010).
13. V. Vantrepotte and F. Mélin, "Inter-annual variations in the SeaWiFS global chlorophyll-*a* concentration (1997–2007)," Deep Sea Res. I **58**, 429–441 (2011).
14. National Research Council, *Assessing Requirements for Sustained Ocean Color Research and Operations* (National Academies, 2011).
15. NASA, "Pre-Aerosol, Clouds, and Ocean Ecosystem (PACE) Mission," Science Definition Team Report, 2012, http://decadal.gsfc.nasa.gov/pace.html .
16. H. R. Gordon and M. Wang, "Retrieval of water-leaving radiance and aerosol optical thickness over the oceans with SeaWiFS: a preliminary algorithm," Appl. Opt. **33**, 443–452 (1994).
17. B. A. Franz, S. W. Bailey, P. J. Werdell, and C. R. McClain, "Sensor-independent approach to the vicarious calibration of satellite ocean color radiometry," Appl. Opt. **46**, 5068–5081 (2007).
18. J. E. O'Reilly, S. Maritorena, B. G. Mitchell, D. A. Siegel, K. L. Carder, S. A. Garver, M. Kahru, and C. R. McClain, "Ocean color chlorophyll algorithms for SeaWiFS," J. Geophys. Res. **103**, 24937–24953 (1998).
19. S. Maritorena, D. A. Siegel, and A. Peterson, "Optimization of a semi-analytic ocean color model for global scale applications," Appl. Opt. **41**, 2705–2714 (2002).
20. IOCCG, Remote Sensing of Inherent Optical Properties: fundamentals, Tests of Algorithms, and Applications, Z.-P. Lee, ed., Reports of the International Ocean Colour Coordinating Group No. 5 (Dartmouth, 2006).
21. P. J. Werdell, B. A. Franz, S. W. Bailey, G. C. Feldman, E. Boss, V. E. Brando, M. Dowell, T. Hirata, S. J. Lavender, Z.-P. Lee, H. Loisel, S. Maritorena, F. Mélin, T. S. Moore, T. J. Smyth, D. Antoine, E. Devred, O. H. Fanton d'Andon, and A. Mangin, "A generalized ocean color inversion model for retrieving marine inherent optical properties," Appl. Opt. **52**, 2019–2037 (2013).
22. J. Aiken, J. R. Fishwick, S. Lavender, R. Barlow, G. F. Moore, H. Sessions, S. Bernard, J. Ras, and N. J. Hardman-Mountford, "Validation of MERIS reflectances and chlorophyll during the BENCAL cruise October 2002: preliminary validation of new demonstration products for phytoplankton functional types and photosynthetic parameters," Int. J. Remote Sens. **28**, 497–516 (2007).
23. R. J. W. Brewin, E. Devred, S. Sathyendranath, S. J. Lavender, and N. J. Hardman-Mountford, "Model of phytoplankton absorption based on three size classes," Appl. Opt. **50**, 4535–4549 (2011).

24. E. Devred, S. Sathyendranath, V. Stuart, H. Maass, O. Ulloa, and T. Platt, "A two-component model of phytoplankton absorption in the open ocean: theory and applications," J. Geophys. Res. **111**, C03011 (2006).
25. T. Hirata, J. Aiken, N. Hardman-Mountford, T. J. Smyth, and R. G. Barlow, "An absorption model to determine phytoplankton size classes from satellite ocean colour," Remote Sens. Environ. **112**, 3153–3159 (2008).
26. T. Hirata, N. J. Hardman-Mountford, R. J. W. Brewin, J. Aiken, R. Barlow, K. Suzuki, T. Isada, E. Howell, T. Hashioka, M. Noguchi-Aita, and Y. Yamanaka, "Synoptic relationships between surface chlorophyll-a and diagnostic pigments specific to phytoplankton functional types," Biogeoscience **8**, 311–327 (2011).
27. J. Uitz, H. Claustre, A. Morel, and S. B. Hooker, "Vertical distribution of phytoplankton communities in open ocean: an assessment based on surface chlorophyll," J. Geophys. Res. **111**, C08005 (2006).
28. S. Alvain, H. Loisel, and D. Dessailly, "Theoretical analysis of ocean color radiances anomalies and implications for phytoplankton groups detection in case 1 waters," Opt. Express **20**, 1070–1083 (2012).
29. A. M. Ciotti and A. Bricaud, "Retrievals of a size parameter for phytoplankton and spectral light absorption by colored detrital matter from water-leaving radiances at SeaWiFS channels in a continental shelf region off Brazil," Limnol. Oceanogr. **4**, 237–253 (2006).
30. T. S. Kostadinov, D. A. Siegel, and S. Maritorena, "Global variability of phytoplankton functional types from space: assessment via the particle size distribution," Biogeoscience **7**, 3239–3257 (2010).
31. H. Loisel, J.-M. Nicolas, A. Sciandra, D. Stramski, and A. Poteau, "Spectral dependency of optical backscattering by marine particles from satellite remote sensing of the global ocean," J. Geophys. Res. **111**, C09024 (2006).
32. B. Lubac, H. Loisel, N. Guiselin, R. Astoreca, L. F. Artigas, and X. Mériaux, "Hyperspectral and multispectral ocean color inversions to detect Phaeocystis globosa blooms in coastal waters," J. Geophys. Res. **113**, C06026 (2008).
33. A. Bracher, M. Vountas, T. Dinter, J. P. Burrows, R. Rottgers, and I. Peeken, "Quantitative observation of cyanobacteria and diatoms from space using PhytoDOAS on SCIAMACHY data," Biogeoscience **6**, 751–764 (2009).
34. C. Mouw and J. A. Yoder, "Optical determination of phytoplankton size composition from global SeaWiFS imagery," J. Geophys. Res. **115**, C12018 (2010).
35. C. S. Roesler, S. M. Etheridge, and G. C. Pitcher, "Application of an ocean color algal taxa detection model to red tides in Southern Benguela," in *Harmful Algae 2002*, K. A. Steidinger, J. H. Landsberg, C. R. Tomas, and G. A. Fargo, eds. (UNESCO, 2004), pp. 303–305.
36. S. Sathyendranath, L. Watts, E. Devred, T. Platt, C. Caverhill, and H. Maass, "Discrimination of diatoms from other phytoplankton using ocean-color data," Mar. Ecol. Prog. Ser. **272**, 59–68 (2004).
37. T. K. Westberry, D. A. Siegel, and A. Subramaniam, "An improved bio-optical model for the remote sensing of Trichodesmium spp. blooms," J. Geophys. Res. **110**, C06012 (2005).
38. M. Sydor, R. W. Gould, R. A. Arnone, V. I. Haltrin, and W. Goode, "Uniqueness in remote sensing of inherent optical properties of ocean water," Appl. Opt. **43**, 2156–2162 (2004).
39. M. Defoin-Platel and M. Chami, "How ambiguous is the inverse problem of ocean color in coastal waters?" J. Geophys. Res. **112**, C03004 (2007).
40. C. S. Roesler, P. J. Werdell, H. D. R. Gomes, J. I. Goes, and S. G. P. Matondkar, "Remote detection of Noctiluca miliaris in the Arabian Sea," Rem. Sens. Environ. (to be published).
41. C. D. Mobley and L. K. Sundman, *Hydrolight 5 Ecolight 5 Users' Guide* (Sequoia Scientific, 2008).
42. B. G. Mitchell, "Algorithms for determining the absorption coefficient for aquatic particulates using the quantitative filter technique," Proc. SPIE **1302**, 137 (1990).
43. C. S. Roesler, "Theoretical and experimental approaches to improve the accuracy of particulate absorption coefficients from the quantitative filter technique," Limnol. Oceanogr. **43**, 1649–1660 (1998).
44. M. Kishino, N. Okami, and S. Ichimura, "Estimation of the spectral absorption coefficients of phytoplankton in the sea," Bulletin of Marine Science **37**, 634–642 (1985).
45. M. Elbrachter and Y. Z. Qui, "Aspects of *Noctiluca* (*Dinophyceae*) population dynamics," in *Physiological Ecology of Harmful Algal Blooms*, M. D. Anderson, A. D. Cembella, and G. M. Hallegraeff, eds. (Springer, 1998), pp. 315–345.
46. Z.-P. Lee, K. L. Carder, and R. Arnone, "Deriving inherent optical properties form water color: a multi-band quasi-analytical algorithm for optically deep water," Appl. Opt. **41**, 5755–5772 (2002).
47. H. R. Gordon, O. B. Brown, R. H. Evans, J. W. Brown, R. C. Smith, K. S. Baker, and D. K. Clark, "A semianalytic radiance model of ocean color," J. Geophys. Res. **93**, 10909–10924 (1988).
48. C. S. Roesler, M. J. Perry, and K. L. Carder, "Modeling *in situ* phytoplankton absorption from total absorption spectra in productive inland marine waters," Limnol. Oceanogr. **34**, 1510–1523 (1989).
49. D. Stramski, E. Boss, D. Bogucki, and K. J. Voss, "The role of seawater constituents in light backscattering in the ocean," Prog. Oceanogr. **61**, 27–56 (2004).
50. X. Zhang, L. Hu, and M.-X. He, "Scattering by pure seawater: effect of salinity," Opt. Express **17**, 5698–5710 (2009).
51. R. M. Pope and E. S. Fry, "Absorption spectrum (380–700 nm) of pure water. II. integrating cavity measurements," Appl. Opt. **36**, 8710–8723 (1997).
52. P. J. Werdell, B. A. Franz, J. T. Lefler, W. D. Robinson, and E. Boss, "Retrieving marine inherent optical properties from satellite using temperature and salinity-dependent backscattering by seawater," Opt. Express **21**, 32611–32622 (2013).
53. A. Whitmire, E. Boss, T. J. Cowles, and W. S. Pegau, "Spectral variability of the particulate backscattering ratio," Opt. Express **15**, 7019–7031 (2007).
54. IOCCG, Mission Requirements for Future Ocean-Colour Sensors, C. R. McClain and G. Meister, eds. Reports of the International Ocean-Colour Coordinating Group No. 13 (Dartmouth, 2013).
55. Z.-P. Lee, K. Carder, R. Arnone, and M.-X. He, "Determination of primary spectral bands for remote sensing of aquatic environments," Sensors **7**, 3428–3441 (2007).
56. H. R. Gordon and W. R. McCluney, "Estimation of the depth of sunlight penetration in the sea for remote sensing," Appl. Opt. **14**, 413–416 (1975).
57. J. R. V. Zaneveld, A. H. Barnard, and E. Boss, "Theoretical derivation of the depth average of remotely sensed optical parameters," Opt. Express **13**, 9052–9061 (2005).
58. C. D. Mobley, *Light and Water: Radiative Transfer in Natural Waters* (Academic, 1994).
59. M. Babin, A. Morel, V. Fournier-Sicre, F. Fell, and D. Stramski, "Light scattering properties of marine particles in coastal and open ocean waters as related to the particle mass concentration," Limnol. Oceanogr. **48**, 843–859 (2003).
60. A. Morel and S. Maritorena, "Bio-optical properties of oceanic waters: a reappraisal," J. Geophys. Res. **106**, 7163–7180 (2001).
61. P. J. Werdell and S. W. Bailey, "An improved *in situ* bio-optical data set for ocean color algorithm development and satellite data product validation," Remote Sens. Environ. **98**, 122–140 (2005).
62. A. Morel and J.-F. Berthon, "Surface pigments, algal biomass profiles, and potential production of the euphotic layer: relationships reinvestigated in view of remote-sensing applications," Limnol. Oceanogr. **34**, 1545–1562 (1989).
63. P. J. Werdell, C. S. Roesler, and J. I. Goes, "Remotely searching for Noctiluca miliaris in the Arabian Sea," J. Geophys. Res. (submitted).
64. A. Bricaud, A. Morel, M. Babin, K. Allali, and H. Clauster, "Variations of light absorption by suspended particles with chlorophyll-a concentration in oceanic (case 1)

waters: analysis and implications for bio-optical models," J. Geophys. Res. **103**, 31033–31044 (1998).
65. P. Wang, E. Boss, and C. S. Roesler, "Uncertainties of inherent optical properties obtained from semi-analytical inversions of ocean color," Appl. Opt. **44**, 4074–4085 (2005).
66. V. E. Brando, A. G. Dekker, Y. J. Park, and T. Schroeder, "Adaptive semi-analytic inversion of ocean color radiometry in optically complex waters," Appl. Opt. **51**, 2808–2833 (2012).
67. P. J. Werdell, C. W. Proctor, E. Boss, T. Leeuw, and M. Ouhssain, "Underway sampling of marine inherent optical properties on the Tara Oceans expedition as a novel resource for ocean color satellite data product validation," Meth. Oceanogr. **7**, 40–51 (2013).
68. S. L. McLeroy-Etheridge and C. S. Roesler, "Are the inherent optical properties of phytoplankton responsible for the distinct ocean colors observed during harmful algal blooms?" Ocean Opt. **14**, 109–116 (1998).
69. T. S. Moore, J. W. Campbell, and M. D. Dowell, "A class-based approach to characterizing and mapping the uncertainty of the MODIS ocean chlorophyll product," Remote Sens. Environ. **113**, 2424–2430 (2009).
70. A. Morel, "Chlorophyll-specific scattering coefficient of phytoplankton: a simplified theoretical approach," Deep Sea Res. **34**, 1093–1105 (1987).
71. D. Stramski and C. D. Mobley, "Effects of microbial particles on ocean optics: a database of single-particle optical properties," Limnol. Oceanogr. **42**, 538–549 (1997).
72. M. S. Twardowski, E. Boss, J. B. Macdonald, W. S. Pegain, A. H. Barnard, and J. R. V. Zaneveld, "A model for estimating bulk refractive index from the optical backscattering ratio and the implications for understanding particle composition in case I and case II waters," J. Geophys. Res. **106**, 14129–14142 (2001).
73. M. Tzortziou, J. R. Hermn, C. L. Gallegos, P. J. Neale, A. Subramaniam, L. W. Harding, Jr., and Z. Ahmad, "Bio-optics of the Chesapeake Bay from measurements and radiative transfer closure," Estuarine, Coastal Shelf Sci. **68**, 348–362 (2006).
74. D. J. Bogucki and G. Spiers, "What percentage of the oceanic mixed layer is accessible to marine lidar? Global and the Gulf of Mexico prospective," Opt. Express **21**, 23997–24014 (2013).