Inversion of spectral absorption coefficients to infer phytoplankton size classes, chlorophyll concentration, and detrital matter

XIAODONG ZHANG,1,* YANNICK HUOT,2 ANNICK BRICAUD,3 AND HEIDI M. SOSIK4

1Department of Earth System Science and Policy, University of North Dakota, Grand Forks, North Dakota 58202, USA
2Centre d’applications et de recherches en télédétection, Université de Sherbrooke, Sherbrooke QC J1K 2R1, Canada
3Laboratoire d’Océanographie de Villefranche, UMR 7093, Université Pierre et Marie Curie-Paris 6, CNRS, Villefranche-sur-Mer 06230, France
4Biology Department, MS 32, Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543, USA
*Corresponding author: zhang@aero.und.edu

Received 21 April 2015; revised 27 May 2015; accepted 27 May 2015; posted 28 May 2015 (Doc. ID 237888); published 18 June 2015

Measured spectral absorption coefficients were inverted to infer phytoplankton concentration in three size classes (picoplankton, nanoplanckton, and microplankton), chlorophyll concentration [Chl], and both magnitude and spectral shape of absorption by colored detrital matter (CDM). Our algorithm allowed us to solve for the non-linear factor of CDM absorption slope separately from the other linear factors, thus fully utilizing the additive characteristic inherent in absorption coefficients. We validated the inversion with three datasets: two spatially distributed global datasets, the Laboratoire d’Océanographie de Villefranche dataset and the NASA bio-Optical Marine Algorithm Dataset, and a time series coastal dataset, the Martha’s Vineyard Coastal Observatory dataset. Comparison with high performance liquid chromatography analyses showed that the phytoplankton size classes can be retrieved with correlation coefficients \( r > 0.7 \), root mean square errors of 0.2, and median relative errors of 20% in oceanic waters and with similar performance in coastal waters. Much improved agreement was found for the entire phytoplankton population, with \( r > 0.90 \) for [Chl] and absorption coefficients \( (a_{ph}) \) for all three datasets. The inferred \( a_{CDM}(400) \) and CDM spectral slope agree within ±4% of measurements in both oceanic and coastal waters. The results indicate that the chlorophyll-a specific absorption spectra used as an inversion kernel represent well the global mean states for each of the three phytoplankton size classes. The method can be applied to either bulk or particulate absorption data and is spectrally flexible.

1. INTRODUCTION

Phytoplankton are a highly diverse group of photosynthetic microorganisms [1] that occupy vastly different ecological niches and impact global biogeochemical cycles in fundamentally different ways [2]. This is evident in the biological pump [3], one of the most important biologically mediated processes that modulates the consequences and response of the ocean to increasing atmospheric CO2. The efficiency of this “pump” varies greatly with phytoplankton of different sizes and types [4,5]. To represent this diversity and its impact on biogeochemistry and food web structure, phytoplankton groups characterized by their size or function and an increasing number of particulate components (bacteria, viruses, micrograzers, and nonliving particles) have been incorporated into biogeochemical or ecosystem models that better quantify interactions between ocean ecosystems and climate (e.g., [6–8]). Validation of these models is challenging, however, and requires observations of how phytoplankton groups vary in space and time.

Because the influence of size on the ecology and physiology of phytoplankton is well established [9], autotrophic phytoplankton are often classified into, and studied separately in, three class sizes: picoplankton (<2 μm), nanoplanckton (2–20 μm), and microplankton (>20 μm) [10]. Many bio-optical techniques, aiming for applications with satellite remote sensing, have been developed for detecting and quantifying these phytoplankton groups. In general, phytoplankton of different sizes are classified based on either their biomass (according to indices such as the magnitude of absorption at one particular wavelength or overall pigment concentration) or the spectral shape of their absorption [11–13]. The former approach exploits observed relationships between the trophic status of the environment and the type of phytoplankton expected under...
that trophic status [14]. In this case, trophic status is characterized by autotrophic biomass indicated by pigment concentration [15–17], the difference of absorption coefficients at 443 and 510 nm [18], or the combination of pigment concentration and absorption coefficient [19].

Different phytoplankton species have evolved different pigment assemblages optimized for photosynthesis and photoprotection, resulting in various and distinctive absorption spectra, which, in turn, can serve as signatures for their retrieval from bulk spectra. This is the basis for the second approach. In addition to the spectral absorption coefficient [20–22], the spectra of normalized water-leaving radiance anomaly [23] and remote sensing reflectance [24,25] have also been utilized to derive information on phytoplankton groups. Compared to radiance or reflectance, the absorption coefficient, as an inherent optical property, is independent of the ambient light field, unaffected by variable backscattering, and determined by additive components. Therefore, it is conceptually simpler and more straightforward to infer information on phytoplankton directly from the spectral absorption coefficient. The absorption coefficient can be readily measured in the field on discrete samples or continuously with flow-through instruments in profiling mode, on buoys or inline from a ship surface water intake (e.g., [22]), or can be derived from ocean color remote sensing [26,27].

In addition to phytoplankton, colored nonalgal particulate and dissolved material [colored detrital matter (CDM)], typically made of degraded organic matter or minerals, also absorbs light, with a spectrum typically decreasing exponentially with wavelength [28–30]. Many efforts have focused on partitioning nonwater absorption into contributions by the phytoplankton assemblage and CDM with varying degrees of success [29,31–34], but few have attempted to infer phytoplankton in different size classes and CDM simultaneously, which leads to the objective of this study. In earlier work exploring this challenge, Ciotti and Bricaud [21] tested two methods. In the first method, they used a two-step approach: initial separation of particle absorption into phytoplankton and detrital components, followed by estimation of a “size parameter” reflecting the proportional contributions of two extreme size classes. A limitation of this method is that, to drive the separation in the first step, assumptions are required regarding the absorption by the bulk phytoplankton particles. In the second method, Ciotti and Bricaud used nonlinear optimization to derive CDM [aCDM(443)] and the spectral slope and the size parameter simultaneously. However, this method requires an a priori estimate of the chlorophyll concentration.

In this study, we assess an optical inversion method to infer the concentrations of phytoplankton in three size fractions and CDM absorption simultaneously from measured nonwater absorption spectra. Our algorithm allowed us to solve for the nonlinear factor of CDM absorption slope separately from the other linear factors, thus fully utilizing the additive characteristic inherent in absorption coefficients. We validated the results with three datasets, including two global datasets: the NASA bio-Optical Marine Algorithm Dataset (NOMAD) [35], the Laboratoire d’Océanographie de Villefranche (LOV) dataset [36,37], and a coastal dataset from the Martha’s Vineyard Coastal Observatory [38].

2. DATA AND METHOD

As an inherent optical property, the absorption coefficient, $a(\lambda)$ (m$^{-1}$, where $\lambda$ is wavelength), reflects additive contributions by its constituents

$$a(\lambda) = \sum_{i=1}^{M} m_i a_i^*(\lambda),$$  

(1)

where $M$ is the number of constituents and, for constituent $i$, $a_i^*(\lambda)$ is its specific absorption at $\lambda$, and $m_i$ is its specific contribution. With multispectral or hyperspectral measurements of $a$ (at BN number of bands), Eq. (1) forms a linear system

$$a = a^* m,$$  

(2)

where $a$ is measured spectral absorption (a column vector with BN elements), i.e., $a = [a(\lambda_1), a(\lambda_2), \ldots a(\lambda_{BN})]^T$; $m$ is the specific contribution from each constituent (a row vector with M elements), i.e., $m = [m_1, m_2, \ldots m_M]$; and $a^*$ is the specific absorption spectrum for each constituent (a matrix with BN × M elements), i.e., $a^* = [a^*_1, a^*_2, \ldots a^*_M]$ and $a^*_i = [a^*_i(\lambda_1), a^*_i(\lambda_2), \ldots a^*_i(\lambda_{BN})]^T$. We can solve Eq. (2) for $m$ with measured absorption spectra $a$ and a prescribed kernel $a^*$.

A. Construction of Kernel $a^*$

Analyzing the LOV dataset of simultaneous measurements of phytoplankton spectral absorption coefficients and high performance liquid chromatography (HPLC) analysis of pigments, Utz et al. [39] derived the chlorophyll-specific absorption spectra from 400 to 700 nm at 2 nm increment for each of three phytoplankton classes: picoc <2 $\mu$m, denoted as $a^*_{picoc}$), nano (2–20 $\mu$m, $a^*_{nano}$), and micro (20–100 $\mu$m, $a^*_{micro}$). Phytoplankton are classified into each size class on the basis of diagnostic pigments measured by HPLC analysis of water samples. Utz et al. [39] also described changes in the spectra with optical depth, reflecting effects of photoacclimation and photoadaptation by phytoplankton. In this study, we restrict analysis to surface waters and hence consider one chlorophyll-specific spectrum for each size class (Fig. 1). For CDM, we represent the absorption spectrum normalized at 400 nm as wavelength-dependent exponential $a^*_{CDM} = \exp(-a(\lambda - 500))$. Combination of the normalized spectrum for CDM and the chlorophyll-specific absorption spectra for picoplankton, nanoplankton, and microplankton forms the inversion kernel, i.e., $a^* = [a^*_{CDM}, a^*_{picoc}, a^*_{nano}, a^*_{micro}]$ (Fig. 1).

To ensure a robust solution to Eq. (2), the kernel $a^*$ should be constructed to avoid either singularity or ill-condition, which requires that the shapes of any two spectra forming the kernel, $a^*_i$ and $a^*_j$ ($i \neq j$), must be sufficiently different from each other. The threshold of similarity is determined by two factors: the uncertainty of the instrument measuring the absorption and the number and locations of the spectral bands at which the absorption is measured or estimated. We used an index estimated as $S_{ij} = \frac{2}{BN} \sum_{\lambda=1}^{BN} \frac{a^*_i(\lambda) - a^*_j(\lambda)}{a^*_i(\lambda) + a^*_j(\lambda)}$ to represent the similarity between two specific spectra $a^*_i$ and $a^*_j$; smaller $S_{ij}$ values indicate more similar spectra. Generally, two initially distinct spectra would become increasingly similar as the uncertainty of measurements increases or the number of bands decreases. In this study, we set the similarity threshold at
S = 0.1 or 10%, below which two spectra are deemed the same and the respective particle groups that they cannot be separated optically. The value of 10% is roughly the uncertainty of the absorption coefficient that can be derived from ocean color measurements at blue–green wavelengths [41,42]. The kernel spectra we used (Fig. 1) differ from each other by more than 10% whether the differences are evaluated for the original bands (i.e., every 1–2 nm from 400 to 700 nm) or only for the SeaWiFS spectral bands. For example, S was 0.2 between picoplankton and nanoplanckton spectra (the two most similar ones among the four spectra in Fig. 1) for the hyperspectral data and 0.16 for the SeaWiFS bands. Even though in this case the similarity test does not affect our choice of kernel function, we emphasize that this test is an important step in any inversion scheme, regardless of the method used to execute the inversion.

### B. Measurements of Absorption and Test Datasets

We used three datasets (Fig. 2) to develop and evaluate the inversion. The LOV dataset [36,37] has hyperspectral measurements for particulate absorption, with phytoplankton (\(a_{ph}\)) absorption coefficients measured every 2 nm from 400 to 700 nm and at various depths from near surface to below the euphotic zone. We only used data collected within the first optical depth and nearest to the surface. The NASA NOMAD dataset [35] includes spectral absorption coefficients for colored dissolved organic matter (\(a_d\)), detrital particles (\(a_d\)), and phytoplankton (\(a_{ph}\)), from which we also computed the nonwater absorption coefficient, \(a_{nw} = a_d + a_d + a_{ph}\). The NOMAD optical data are available at 20 nominal wavebands. We used those at SeaWiFS bands, with central wavelengths at 411, 443, 489, 510, 555, and 670 nm. Since NOMAD was intended for ocean color algorithm development and validation, the absorption data (and other IOPs [inherent optical properties]) are provided as optically weighted values in the first optical depth estimated following Gordon and Clark [40]. The MVCO dataset includes measurements of \(a_{nw}\), \(a_{ph}\), and \(a_{ph}\) with 1 nm resolution from 300 to 850 nm at various depths [43]. We only used data for depths <3 m. We extracted values between 400 and 700 nm and computed the nonwater absorption coefficients \(a_{nw}\). The \(a_{ph}\) values from the LOV dataset and \(a_{nw}\) values from the NOMAD and MVCO datasets were used as inputs to drive the inversion.

The HPLC data collocated and concurrent with the absorption measurements were used for validation of chlorophyll concentration and size fractions of phytoplankton. We estimated fractional concentrations of picoplankton, nanoplanckton, and microplankton on the basis of relative proportions of seven diagnostic pigments [36,39,44]. According to Vidussi et al. [44], these pigments are fucoxanthin, peridinin, alloxanthin, 19′-butanoyloxyfucoxanthin (19′-BF), 19-hexanoyloxyfucoxanthin (19′-HF), zeaxanthin, and chlorophyll-b + divinyl chlorophyll b. The fractions of picoplankton (\(f_{pico}\)), nanoplanckton (\(f_{nano}\)), and microplankton (\(f_{micro}\)) are estimated following Uitz et al. [17]

\[
\begin{align*}
    f_{micro} &= (1.41[\text{fuco}]+1.41[\text{peri}]) / wDP \\
    f_{nano} &= (0.60[\text{allo}]+0.35[19′-BF]+1.27[19′-HF]) / wDP \\
    f_{pico} &= (0.86[\text{zea}]+1.01[\text{Chlb}+\text{div-Chlb}]) / wDP, \\
\end{align*}
\]

where \(wDP\) is the weighted sum of these seven diagnostic pigments

\[
wDP = 1.4[\text{fuco}]+1.4[\text{peri}]
+0.60[\text{allo}]+0.35[19′-BF]+1.27[19′-HF]
+0.86[\text{zea}]+1.01[\text{Chlb}+\text{div-Chlb}].
\]

Several factors need to be pointed out regarding Eqs. (3) and (4). First, classification from diagnostic pigments does not strictly reflect the true size of phytoplankton. For example, fucoxanthin, the main indicator of diatoms, is used to identify microplankton, but it may also be found in some nanoplanckton. These kinds of ambiguities will not affect the
interpretation of our results, however, because the $\alpha^*$ spectra included in the kernel function for the three phytoplankton size fractions were derived with the same classification method [39]. Second, modifications to Eq. (3) have been proposed to improve the estimate of nanoplanктон and picoplankton fractions at low [Chl] [15,18,45]. We chose to use Eq. (3), again, to be consistent with Üitz et al. [39]. Last, the uncertainty in Eq. (3) is typically <15% (ranging from 2% to 30%) [46]. The wide range of uncertainty occurs because the levels of certain diagnostic pigments are sometimes too low for accurate estimation, such as when detecting fucoxanthin and peridinin in an oligotrophic regime.

We estimated the absorption coefficient for CDM at 400 nm, $a_{CDM}(400)$, and the spectral slope $\alpha$ by applying a nonlinear regression to the measured $a_{CDM}(\lambda)$

$$a_{CDM}(\lambda) = a_{CDM}(400) \exp(-\alpha(\lambda - 400)),$$

where $a_{CDM} = a_\phi + a_d$. Values of $a_{CDM}(400)$ and $\alpha$ estimated from Eq. (5) were considered as “measured” values and used for validation of CDM inversion, which is only applicable to the NOMAD and MVCO tests.

A final quality control was applied to the LOV dataset by visually inspecting each $a_{ph}$ spectrum; those that were deemed too noisy or had a significant portion with negative values in the green part of the spectrum were removed. The number of records removed by this performance were 47 out of a total of 757. No quality control was applied to the NOMAD or MVCO dataset. The numbers of data in the final datasets are 710 for LOV, 270 for NOMAD, and 780 for MVCO.

The three datasets, two of which (LOV and NOMAD) have been used in numerous studies, offer several contrasts that help to better evaluate our inversion method. While the MVCO dataset provides temporal resolution over a decade in a coastal regime, the NOMAD and LOV datasets are distributed over the world ocean and cover many different water types (Fig. 2). The NOMAD dataset includes measurements from open ocean as well as coastal waters [35], whereas the LOV dataset was collected mainly in oceanic waters with chlorophyll concentration covering three orders of magnitude [36,37]. While most of the NOMAD and LOV data cover typical case 1 waters where phytoplankton species dominate spectral absorption, the MVCO is located in a coastal environment where CDM dominates absorption at wavelengths up to 600 nm. The absorption coefficients in both the LOV and MVCO datasets were measured with high spectral resolution (every 2 nm for LOV and 1 nm for MVCO), whereas the NOMAD data were extracted only at the six SeaWiFS bands. We used the measurements of $a_{ph}$ to drive the inversion for the LOV dataset, whereas the total absorption minus water ($a_{nw}$) was used for the NOMAD and MVCO data. This allowed us to evaluate not only the efficacy of the algorithm in retrieving CDM from $a_{nw}$, but also its performance when CDM is negligible since $a_{ph}$ is expected to contain little or no signal from CDM. Finally, since the kernel spectra $a_{pico}^*$, $a_{nano}^*$, and $a_{micro}^*$ were derived from the LOV dataset [39], applying our method to the dataset will only allow us to verify the applicability of the inversion method; it will not allow us to validate whether the $a^*$ spectra are representative and applicable beyond the LOV dataset.

Application to the NOMAD and MVCO datasets allows this kind of validation.

C. Performing the Inversion

For the inversion of the NOMAD dataset developed for remote sensing applications, the input was $a_{nw}$, and the kernel included $a_{CDM}^*$ and $a_{nano}^*$, $a_{micro}^*$ (the dashed curves in Fig. 1). These phytoplankton spectra were optically weighted within the first optical depth following Gordon and Clark [40] to match the weighting applied to the NOMAD data. All spectra were extracted at the SeaWiFS spectral bands. For the LOV dataset, the kernel included $a_{CDM}^*$ and the surface values of $a_{pico}^*$, $a_{nano}^*$, and $a_{micro}^*$ (the solid curves in Fig. 1), and the input was $a_{ph}$. Even though we expect a minimal or negligible amount of CDM present in measured $a_{ph}$, we included $a_{CDM}^*$ in the kernel as a test of robustness of the inversion algorithm. For the test with the MVCO dataset, the kernel was the same as for the LOV test, but the input was $a_{nw}$. With a prescribed kernel and measured spectral absorption coefficients, we used the singular value decomposition (SVD) least-squares method [47] to estimate $\alpha$ by inverting the linear system of Eq. (2) with the constraint $m \geq 0$. The same method has been used to invert measured volume scattering functions to infer particles of different types and sizes [48,49]. For a solution of $m$, $m_i = 0$ would mean that constituent $i$ is not present in the sample or makes negligible contribution to the observed absorption. In the final solution of $m$, $m_1$ represents $a_{CDM}(400)$ (m$^{-1}$), while $m_2$, $m_3$, and $m_4$ represent, respectively, chlorophyll-a concentration of pico-plankton ([Chl]$_{pico}$), Hg/L), nanoplanктон ([Chl]$_{nano}$), and microplanктон ([Chl]$_{micro}$). It is straightforward to derive $a_{ph}$ from the inversion results as $m_2a_{pico}^* + m_3a_{nano}^* + m_4a_{micro}^*$.

For CDM, the spectral slope $\alpha$ is an additional unknown and typically varies between 0.004 and 0.02 nm$^{-1}$ [50]. The absorption by CDM is generally dominated by colored dissolved organic matter (CDOM) [51], which typically exhibits higher $\alpha$ values than nonalgal particles [50,52]. While the values of $\alpha$ and $m$ can be solved simultaneously via nonlinear optimization, e.g., as in method 2 of Ciotti and Bracado [21] or in Zhang et al. [53], we adopted a different two-step approach to solve the nonlinear equation: (1) invert the linear system of Eq. (2), and (2) search for the optimal $\alpha$ value. For an arbitrary value of $\alpha$ within the range of $[0.004, 0.02]$ nm$^{-1}$, we computed $a_{CDM}^*$, constructed $a^*$, performed the linear inversion as described above to estimate $m$, and estimated the residual in $a$ as $\|a - a^*m\|$. We then conducted a bounded global minimum search (using the MATLAB function fminbnd) to find the $\alpha$ value that gave the minimum residual. The inversion result corresponding to this minimum is the final solution. The advantage of this approach is that the linearity resulting from the additive rule of IOPs is fully utilized in the inversion process; the algorithm is in essence identifying the best (linear) solution to Eq. (2) for all possible $\alpha$ values between 0.004 and 0.02 nm$^{-1}$. This is possible because a test with all three datasets showed that the residual after solving the linear system Eq. (2) is a well-behaved function of $\alpha$ (Fig. 3), with a single minimum value within the range $[0.004, 0.02]$ nm$^{-1}$. The similar approach was also used in Wang et al. [41] for solving the
nonlinear spectral shape parameters from measured spectral reflectance.

With the tolerance for $\alpha$ set at $1 \times 10^{-6}$ nm$^{-1}$, the residual as a function of $\alpha$ (Fig. 3) allows a rapid convergence of the minimum search, which typically takes less than 10 steps. In the few cases where convergence could not be achieved (only in the MVCO test), the search always ended at the upper end of the range, an indication of possibly steeper slopes for CDM absorption in these waters. Also, the SVD least-squares method ensures a unique solution even for underdetermined linear systems [54,55] because the solution always has the minimal $\|w\|$. While the datasets in this study are all overdetermined systems, the NOMAD test has only seven wavelengths to derive five parameters, and using the SVD method ensures unambiguous solutions.

3. RESULTS

Through the inversion, measured absorption spectra were partitioned into individual contributions by CDM and phytoplankton in three size fractions (see Fig. 4 for example spectra). The reconstructed absorption spectra (dotted black lines in Fig. 4) are in excellent agreement with the measurements (solid black lines). On average, the inversion can explain $\sim 99\%$, $\sim 95\%$, and $\sim 92\%$ of the variability observed in the measured absorption spectra in the NOMAD, LOV, and MVCO datasets, respectively. In terms of the variability explained in the measured absorption spectra, the seemingly better inversion efficiency for the NOMAD-SeaWiFS dataset appears because the absorption spectra sampled at six bands contain less information (or variability) than either the LOV or MVCO data where absorption spectra were measured at much finer spectral resolution (1–2 nm). Because both the NOMAD and LOV datasets have global coverage mostly over oceanic waters, we will first evaluate the inversion results comparatively between these two datasets, followed by a close examination of results for the MVCO site.

A. Phytoplankton Size Fractions

From the inversion results, the fractional contribution of each size class to the total chlorophyll is estimated as $m_i/(m_1 + m_2 + m_3)$, where $i = 2$, 3, or 4 for picoplankton, nanoplankton, and microplankton, respectively. For the HPLC results, the size fractions are estimated following Eqs. (3) and (4).

We compared the size fractions derived from the inversion with the HPLC estimates (Fig. 5) using a variety of statistical measures (Table 1). Between the two estimates of size fractions, correlation coefficients ($r$) are $>0.70$ for all cases except for nanoplankton in the NOMAD data (for which $r = 0.52$). Also, the median relative errors (MRE) are $< \sim 20\%$, and absolute errors as measured by root mean square error (RMSE) are $<0.2$ for all cases. Considering the inherent uncertainty of $\sim 15\%$ for HPLC-derived phytoplankton size fractions [46], the two estimates agree with each other reasonably well. More detailed regional comparison reveals some notable patterns (Fig. 5). In equatorial and South Pacific waters, picoplankton typically dominate ($f_{\text{pico}} > 0.6$), whereas near the coast in the north Pacific, microplankton often dominate. Meridional transects in the Atlantic (NOMAD dataset) show that each of the three size fractions can dominate. The LOV data collected in Atlantic waters are concentrated in two regions: the mid-North Atlantic and the Benguela upwelling waters. In the mid-North Atlantic, either picoplankton or nanoplankton dominate, with $f_{\text{micro}}$ never exceeding 50%.

![Fig. 3. Example showing typical $\alpha$ dependence of the residual ($\|a - a^* \|a\|$) after solving Eq. (2). The final solution is determined by selecting the $\alpha$ value corresponding to the minimum residual, which in this example occurs at $\alpha = 0.011$ nm$^{-1}$.](image)

![Fig. 4. Examples from the (a) NOMAD, (b) LOV, and (c) MVCO datasets showing how the inversion algorithm partitions measured absorption spectra (black lines) into contributions by CDM and phytoplankton in three size fractions. In (b), the inferred $f_{\text{CDM}}(400)$ is zero; in (c), the inferred $f_{\text{nano}}$ is zero. Dashed lines are the reconstructed absorption spectra—the sum of the absorption coefficients of the three size fractions and CDM. Locations of the three examples are indicated in Fig. 2 as an orange cross, circle, and star for (a), (b), and (c), respectively.](image)
whereas in the Benguela, either nanoplanckton or microplankton dominate, with \( f_{\text{pico}} \) seldom exceeding 40%. A range of conditions occur in the Mediterranean Sea where waters can be dominated by each of the three size fractions. These patterns of variability are consistent in both the HPLC and the inverted estimates, suggesting that the specific absorption spectra derived from Ulitz et al. [39] effectively represent the mean state of light absorption for each phytoplankton size class in surface waters. Nonetheless, there are deviations from these mean states. For example, the inversion overestimated microplankton in both the mid-North Atlantic [cyan in Fig. 5(b)] and North Atlantic transects [blue in Fig. 5(a)]. There is also noticeable unexplained variability in nanoplanckton and picoplankton in the equatorial and South Pacific oceans [purple in Figs. 5(d) and 5(f)].

A comparison of the performance of the inversion method in inferring phytoplankton size fractions as a function of trophic state for the NOMAD and LOV datasets shows a general agreement with the HPLC-based estimates (Fig. 6). The results from both datasets show that microplankton dominate \( (f_{\text{micro}} > 60\%) \) in eutrophic waters with [Chl] > 2 µg/L and account for less than 20% in waters with [Chl] < 0.2 µg/L. Dominance of picoplankton occurs mainly in oligotrophic waters, which also occasionally show a high fraction of nanoplankton. In mesotrophic waters \( (0.2 < \text{[Chl]} < 2 \text{µg/L}) \), picoplankton account for <30% and either microplankton or nanoplanckton can dominate. It is also noteworthy that the LOV dataset shows a clear dominance by nanoplanckton in mesotrophic waters [Figs. 6(b) and 6(d)]. As a separate point of comparison, we also considered size fractions predicted by the Brewin et al. [15] model and found favorable agreement with our inversion results [Figs. 6(e) and 6(f)]. The Brewin et al. model predicts the fractions of nanoplanckton, nanoplankton, and picoplankton according to an empirical relationship with

\[ r, \text{Slope (SE)}, \text{Int (SE)}, \text{RMSE}, \text{MRE} \text{ are, respectively, the correlation coefficient; the type-II regression slope (standard error); the type-II regression intercept (standard error); the root mean square error; and the median relative error. In type-II regression, we used observed (O) as y values and inversion-derived or predicted (P) as x values following the recommendations of Piñeiro et al. [56]. RMSE = \sqrt{\sum (P_i - O_i)^2}/N \text{ and MRE = median}(P_i/O_i - 1). The underlined values are estimated for log-transformed data since [Chl], \( a_{\text{ph}} \), and \( a_{\text{CDM}} \) typically follow log-normal distributions in the oceans [57].} \]
by picoplankton decreases with [Chl], while the nanoplanckton contribution tends to dominate in mesotrophic conditions. Notably, however, both the inversion and HPLC results show that this transition is far from monotonic and exhibits significant variability that cannot be captured by the empirical relationship.

The two datasets considered thus far differ in the inputs driving the inversion ($a_{\text{nw}}$ for NOMAD versus $a_{\text{ph}}$ for LOV), but the results are very similar in their accuracy in retrieval of the size fractions of phytoplankton. The distribution of errors ($= f_{\text{inversion}} - f_{\text{HPLC}}$) for the phytoplankton size fractions as a function [Chl] for the NOMAD and LOV datasets are almost the same, with mean $\approx 0$ and RMSE$\approx 0.2$ (Fig. 7). It is noteworthy that RMSE values estimated for each size fraction are also $\approx 0.2$ (Table 1). This pattern of error distribution indicates that errors in inferring the size fraction contributions largely cancel out when considering total phytoplankton biomass. The errors in both tests are greater at low [Chl] as compared to more eutrophic conditions. The similar error distribution in the two datasets suggests that the chlorophyll-specific absorption spectra of Uitz et al. [39] that we used in the kernel provide an effective representation of the size fractions beyond the LOV dataset.

**B. Chlorophyll-a Concentration**

The total chlorophyll-a concentration ([Chl]) was estimated by summing $m_2$, $m_3$, and $m_4$. Both datasets cover a wide dynamic range of chlorophyll values (0.06–10 µg/L in the NOMAD dataset and 0.01–30 µg/L in the LOV dataset). The inferred [Chl] agreed very well with the HPLC estimates (Fig. 8 and Table 1); the correlation coefficients ($r$) are $>0.95$ and the median percentage errors are about 11%, slightly higher than the 7% methodological uncertainty of [Chl] estimated from an intercomparison of HPLC pigment methods [46]. Type-II regression analysis also indicated agreement between the inferred and HPLC-determined (log-transformed) [Chl] values (Table 1) (slope is $1.10 \pm 0.02$ for the NOMAD dataset and $0.99 \pm 0.01$ for the LOV dataset; both intercepts are not significantly different from zero). While the excellent agreement for the LOV test is more or less expected since the same dataset was used to derive $a_{\text{micro}}^*$, $a_{\text{nano}}^*$, and $a_{\text{ph}}^*$ used in the inversion kernel, the results from the NOMAD test further confirm that these chlorophyll-specific absorption spectra are globally representative, and the individual errors associated with each size fraction are largely cancelled out [Figs. 7(a) and 7(b)]. Furthermore, the success in retrieving [Chl] from both tests indicates that the phytoplankton variability is well captured by the inversion.

**C. Phytoplankton Absorption Coefficient**

The absorption spectra for phytoplankton were estimated as $m_2 a_{\text{micro}}^* + m_3 a_{\text{nano}}^* + m_4 a_{\text{ph}}^*$. Since $a_{\text{ph}}$ was used as input to the LOV test, we expect that the inferred $a_{\text{ph}}$ would match the input [Fig. 4(b)]. Therefore, we only examine $a_{\text{ph}}$ inferred from the NOMAD test, which used $a_{\text{nw}}$ as input. The overall agreement between the two is excellent for all SeaWiFS bands (see Fig. 9 and Table 1 for details at 443 nm). Slightly degraded agreement was found for the 555 nm band, where the...
absorption due to phytoplankton is normally at a minimum and measurements are often more noisy. We noticed a small underestimation of $a_{ph}$, mostly at the lowest $a_{ph}$ measured, which slightly increases with decreasing wavelength. This could be due to slight overestimation of the CDM absorption slope (see Fig. 10, which is described later) or the assumption of a single slope for the entire CDM spectrum, whereas there is often a slight increase in the slope approaching the UV [58].

D. Nonalgal Particles

From the inversion, $m_1$ represents $a_{CDM}(400)$ (m$^{-1}$). For inversion of the NOMAD dataset, the input was $a_{nw}$, which includes contributions by both phytoplankton and CDM (i.e., $a_{nw} = a_{ph} + a_{CDM}$). Therefore, we expected the inversion results to show the influence of CDM. On the other hand, for the inversion of the LOV dataset, $a_{ph}$ was used; therefore, we expected a negligible presence of CDM.

Both retrieval of $a_{CDM}(400)$ ($r = 0.99$) and $\alpha$ ($r = 0.88$) are well correlated with measurements, with relative differences of ±4% (Fig. 10). While the search range for the spectral slope $\alpha$ was 0.004–0.02, 80% of the values of $\alpha$ are between 0.001 and 0.016 nm$^{-1}$, and 50% of the $\alpha$ values are between 0.001 and 0.016 nm$^{-1}$, agreeing well with the global average values of 0.011 nm$^{-1}$ and 0.016 nm$^{-1}$ for detritus and CDOM, respectively [50].

For inversion of the LOV dataset, out of a total of 710 inferred values for $a_{CDM}(400)$, nearly 60% are zeros and over...
90% have values less than $0.002 \text{ m}^{-1}$ [Fig. 10(c)]. For reference, the range of $a_{\text{CDM}}(400)$ for the NOMAD dataset is $0.01$–$0.6 \text{ m}^{-1}$ [Fig. 10(a)]. This distribution is consistent with the fact that the inversion with the LOV dataset started with $a_{\text{ph}}$ as the input, which is not expected to contain CDM. As a further test, we ran the inversion with $a_{\text{ph}}$ values from NOMAD and found that 60% of retrieved $a_{\text{CDM}}$ values were zero and 97% were $<0.01 \text{ m}^{-1}$. This and Fig. 10(c) indicate that the algorithm is robust, and its performance does not depend on the types of inputs.

**E. MVCO Test**

So far we have focused on validating the inversion results across different oceanic waters, spanning conditions for which the inversion kernel of $a_{\text{pico}}$, $a_{\text{nano}}$, and $a_{\text{micro}}$ was derived. Compared to the NOMAD and LOV datasets that include dramatic variations in phytoplankton size classes between different water types, the MVCO dataset shows more subtle changes in phytoplankton through time at one location. Also CDM contributes significantly to the absorption at the MVCO site (e.g., see Fig. 4). At this location, the phytoplankton community is dominated all year by microplankton ($f_{\text{micro}}$ is always $>0.5$), though picoplankton and nanoplankton do increase during the less eutrophic periods of the year (Fig. 11). HPLC analysis shows that the average fractions of microplankton, nanoplankton, and picoplankton at the MVCO site over the past 10 years are 0.74, 0.10, and 0.16, respectively. From inversion of $a_{\text{pico}}$, we were able not only to retrieve the average relative abundance of each size class ($r = 0.72$ for microplankton and $r = 0.54$ for both nanoplankton and picoplankton), but also to track small systematic seasonal shifts between microplankton and picoplankton (Fig. 11). The average fractions estimated from the inversion were 0.83, 0.06, and 0.11 for microplankton, nanoplankton, and picoplankton, respectively. The inversion was able to track the seasonal variability of microplankton, the dominant size class at the MVCO site, especially well, while variability in the nanoplankton was more difficult to recover (many of the retrievals were $\sim 0$), mainly because the fraction, and hence its signal, was too low to be detected reliably.

Inversion-based partitioning of the absorption spectra between phytoplankton and CDM in the MVCO dataset was surprisingly effective (Fig. 11), considering the lack of high CDM conditions used in kernel development. Over the time period of the observations, [Chl] varied nearly two orders of magnitude ($0.1$–$10 \mu \text{g L}^{-1}$), and the inferred [Chl] agreed with the observations ($r$ of 0.89 and MRE $= -0.03$; see Fig. 8(c)). Despite the fact that the absorption at wavelengths shorter than 600 nm is dominated by CDM at the MVCO site (e.g., see Fig. 4), inferred $a_{\text{ph}}$ compare well with the measurements ($r$ ranging from 0.88 to 0.93 and RMSE from 0.10 to 0.17; see Fig. 12), only slightly worse than the NOMAD test (Fig. 9). At the 670 nm band, where the impact of CDM is less significant, the inferred $a_{\text{ph}}$ compared better with the measurements than at the other bands. Particularly good agreement was found for CDM ($r = 0.99$ and 0.82 for $a_{\text{CDM}}(400)$ and $a$, respectively), with performance comparable to the NOMAD test [compare Figs. 10(a) and 10(b) with Figs. 10(d) and 10(e)].
4. DISCUSSION AND CONCLUSIONS

Even though the variables $\alpha$ and $m$ that we seek to solve are related in a nonlinear fashion, we found that solving the linear system described by Eq. (2) reveals residuals that are a well-behaved function of $\alpha$ (Fig. 3). This allows us to separate the inherently unstable nonlinear inversion into two separately stable steps: linear inversion for $m$, followed by a global search for the optimal $\alpha$. As an IOP, the absorption coefficient follows the additive rule, i.e., the total is a simple sum of constituent contributions. Consequently, when the bulk absorption coefficient is used to drive the inversion of the components, the system is inherently linear [Eq. (2)], and its least-squares solution represents a global minimum in the error function. If the spectral normalized water-leaving radiance or reflectance $[30, 59]$ are used to drive the inversion, the system is inherently nonlinear, and its solution is sensitive to the initial values $[60]$. Since the spectral variability exhibited in radiance or reflectance is largely attributable to the selective absorption by phytoplankton and detritus particles $[61, 62]$, using the absorption coefficient to derive different phytoplankton groups not only has mathematical convenience, but also is physically intuitive.

For phytoplankton, the kernels are chlorophyll-specific absorption spectra (m$^2$ (mg Chl a)$^{-1}$), representing the absorption cross-sectional area per unit mass of chlorophyll-a; for CDM, it is absorption per unit absorption at 400 nm. Because of this, the solutions for $m$ represent the chlorophyll concentrations for each of the three phytoplankton size classes and absorption at 400 nm for CDM, which is typically used as a surrogate for CDM concentration. In the case of $\alpha$, we constrained values to $0.004 < \alpha < 0.02$, thereby limiting the range only to previously measured values. Notably, few of the $\alpha$ values retrieved were at the extremes of the range even at the MVCO site where CDM dominated absorption at shorter wavelengths; this highlights the well-behaved nature of the method and that the chosen phytoplankton spectra in the kernel were appropriate and did not lead to compensation by an extreme $a_{\text{CDM}}$ slope.

To the best of our knowledge, the method we presented is unique in that the concentrations of three phytoplankton size classes and both concentration and spectral shape of CDM are retrieved simultaneously from the total nonwater absorption spectrum. We need to emphasize that the size classes for phytoplankton are based on pigment analysis by the HPLC method, which has been most widely used for validating the various inversion methods $[11, 15]$. While we recognize that the pigment-derived classes do not correspond strictly to the true size of phytoplankton $[17, 44]$, all other in situ methods have their own uncertainties (e.g., Table 2.3 in $[63]$), and it is beyond the scope of this study to further explore this issue. Also, ignoring this issue does not affect the interpretation of our results because the kernel we used and the reported results we compared to are all based on the size classes derived from HPLC pigment analysis.

Adopting a pigment-based size classification, our results compare well not only within the LOV dataset used to derive the phytoplankton kernel, but also with independent measurements (Figs. 5–12 and Table 1). Our algorithm searches for the optimal values for both the magnitude and spectral shape of CDM absorption in a linear manner (see discussion above). The dynamic retrieval of spectral slopes of CDM provides not only critical information on composition of CDM $[58]$, such as the ratio of humic acids to fulvic acids $[64]$, but also a better kernel for CDM absorption (i.e., $a_{\text{CDM}}$). This leads to an effective partition of the absorption spectrum into two broad components of CDM and phytoplankton. Furthermore, our algorithm seems to work for both oceanic and coastal waters. The RMSEs for log-transformed $a_{\text{ph}}$ (443) and $a_{\text{CDM}}$ (400) are 0.08 and 0.04, respectively, for the NOMAD test and are 0.17 and 0.03, respectively, for the MVCO test. For phytoplankton size classes, our method can infer each fraction with an uncertainty of 0.2 (measured as RMSE) or 20% (measured as MRE) in case 1 waters. In coastal waters, as tested for the MVCO site, the same level of performance can be achieved for the dominant size class.

In the present study, we found no significant difference in performance between multispectral and hyperspectral data because the results are comparable between the NOMAD-SeaWiFS and MVCO datasets. We also tested NOMAD datasets using all 20 bands, the MODIS bands only, and the MERIS bands only with very similar results. For example, the $r$ values for [Chl] are 0.94, 0.95, and 0.96 for the full NOMAD, MODIS, and MERIS bands, respectively. For comparison, $r = 0.96$ for the SeaWiFS bands. This is, of course, partly attributable to the optimal and similar spectral band placement of these ocean color sensors.

In summary, the inversion method we presented is able to simultaneously partition the measured nonwater absorption spectrum into contributions by phytoplankton in three major size classes and by CDM. The inversion further provides estimates of bulk chlorophyll-a concentration, spectral absorption by phytoplankton, and both the magnitude and the spectral slope of CDM absorption, all of which agree well with independent measurements in the tested datasets (Table 1). The performance of the inversion method indicates that the kernel function constructed with the chlorophyll-a specific absorption spectra derived by Uitz et al. $[39]$ adequately represents the mean absorption of the different sizes classes across all trophic states.

Compared with typical two-component absorption models that include phytoplankton and CDM, inclusion of three phytoplankton size classes (instead of one fixed bulk phytoplankton spectrum) provides additional freedom for inversion, which leads to better performance in retrieval for both phytoplankton and CDM components. While we tested the inversion for SeaWiFS wavelengths, it is straightforward to apply the method to other sensors by simply extracting the kernel at, for example, MODIS bands. The method can also be applied to in situ sensors such as absorption meters (e.g., WETLabs, Inc. ac-9 or ac-s), again by simply building the kernel at the appropriate bands. Application to such in situ sensors opens the door to continuous (inline) measurements of phytoplankton size classes during scientific cruises when biomass is sufficient to obtain accurate measurements of $a_{\text{nw}}$.

While this new inversion method can separate phytoplankton and CDM, we recognize that the errors in predicting the exact amount of a phytoplankton size class can be significant...
and vary with regions (Figs. 5 and 11). A likely explanation for much of this unexplained variability is that, even within one size class, significant variability is expected in the light absorption efficiency by phytoplankton of different species or under different environmental conditions. Thus, for improved estimation of multiple phytoplankton groups from optical observations, a regional approach, such as that based on ecological provinces [55] where a kernel function can be fine-tuned, may perform better. Another challenge is to account for the packaging effect [66,67] resulting from photoadaptation and photo-acclimation. This can lead to changes in phytoplankton specific absorption spectra [36] that have been assumed fixed in our method. In future work, it might be possible to incorporate some of this variability by allowing the kernel to vary depending on appropriate environmental properties (similar to the optical depth dependence documented by Uitz et al. [39]).


We thank two anonymous reviewers for their valuable comments. The LOV database has been assembled thanks to several projects (EUMELI, EPOPE, FRONTAL, PROSOPE, POMME, BIOSOSE) funded by the French PROOF National Program. The authors thank the chief scientists and crews of the cruises during which these data were collected, as well as the numerous colleagues in LOV who participated in sample collection, HPLC measurements, or absorption measurements. We thank the MVCO Operations Team for their dedication and skill in maintaining the facility. We are indebted to E. Taylor Crockford and Emily Peacock for excellent technical support in field collection and laboratory analysis of water samples from MVCO.

REFERENCES
