The primacy of dissolved organic matter to aquatic light variability

Henry F. Houskeeper[1] and Stanford B. Hooker[2]

[1]Department of Applied Ocean Physics & Engineering, Woods Hole Oceanographic Institution, Woods Hole, Massachusetts, U.S.A.

[2]Independent Researcher, U.S.A.

Correspondence: Henry F. Houskeeper (henry.houskeeper@whoi.edu)

Abstract. Absorption and scattering by optically active constituents (OACs) modify the sunlit aquatic light environment, facilitating the derivation of biogeochemical data products at scales spanning *in situ* to satellite observations. Excluding solar illumination, plus geometric and atmospheric effects, variability in an optical parameter arises from changing OAC concentrations, wherein observed patterns in the spectral evolution of data products are associated with the connectivity and spatiotemporal dynamics of OACs. In open-ocean waters far from terrestrial and riverine inputs, the content and mixture of OACs principally relates to the dynamics of phytoplankton and the microbial loop—a trophic pathway describing the cycling of microbial primary producers, remineralizers (e.g., bacteria and archaea), plus dissolved organic and inorganic materials (as applicable). Historical bio-optical models for the open ocean primarily invoke chlorophyll a concentration (C_a)—a commonly used proxy for phytoplankton biomass—as the ubiquitous independent variable governing optical data products such as the normalized water-leaving radiance, $[L_W(\lambda)]_{\mathrm{N}}$. Formulation of $[L_W(\lambda)]_{\mathrm{N}}$ as a function of C_a invokes an idealized food chain, wherein phytoplankton are the dominant control of OACs, including the colored (or chromophoric depending on the literature) portion of the dissolved organic matter (DOM) pool, hereafter CDOM. This prescription, in which C_a maximally explains oceanic light variability (hereafter primacy), is tested herein using eigenanalysis—e.g., an Empirical Orthogonal Function (EOF) analysis, Principal Component Analysis (PCA), or other eigendecomposition depending on the literature. Analyses using three independent bio-optical datasets assess the shapes and associations of the principal and secondary eigenfunctions of aquatic $[L_W(\lambda)]_{\mathrm{N}}$ observations. The analyses reveal $[L_W(\lambda)]_{\mathrm{N}}$ variations to be more strongly associated with changes in CDOM rather than C_a—even for purely oceanic datasets—indicating that CDOM dynamics are more variable and exhibit greater independence from C_a than formerly ascribed. Blue and green band-ratio algorithms routinely used for remote sensing of C_a are found to be maximally sensitive to variability in CDOM—rather than C_a—based on validation tests of OC algorithm performance (e.g., R^2 of 0.85 versus 0.78), plus partial correlation coefficients relating eigenfunction scalar amplitude functions to field or derived observations. Eigenanalyses applied to spectral subsets of the data indicate expansive spectral range observing improves the independence in retrieving CDOM absorption and C_a. The spectral subset comparisons indicate expanded spectral observations supported by recent domestic and international satellite missions constitute a new and unique opportunity to optically characterize surface ocean phytoplankton stocks without relying on explicit or implied empiricisms requiring CDOM and other OACs to covary with C_a. The shapes and associations of the eigenfunctions suggest a greater diversity of trophic pathways determine OAC dynamics—e.g., in addition to phytoplankton contributing CDOM via cellular lysis, excretion, and grazing—

and are consistent with advancing knowledge of the microbial loop in the decades after bio-optical formulations based on C_a were proposed.

1 Introduction

More than seven decades ago, transitions in the visual perception of electromagnetic emission (hereafter *color*) from the deepest blue waters to the much shallower green coastal zone were attributed primarily to changes in the colored dissolved organic matter (CDOM) content (Jerlov 1951; Yentsch 1960; Kalle 1966). Known absorption characteristics of photosynthetic pigments, principally absorption associated with the content of chlorophyll a, C_a, supported analogous dependencies of color on phytoplankton content, which motivated investigations into optical approaches for retrieving surface ocean phytoplankton abundance (Ryther and Yentsch 1957; Tyler 1961). Significant overlap in C_a and CDOM spectral absorption characteristics, combined with limited spectral resolution of early optical technologies, however, prevented the partition of C_a and CDOM signals (Yentsch 1960). Contents of C_a and CDOM covary in oceanic environments away from the confounding effects of terrestrial and riverine inputs, and a generalized relationship between color and C_a content was demonstrated using airborne surveys (Clarke et al. 1970)—despite acknowledged confounding factors, including (but not exclusively) solar glint, sky state, and natural variability in the content of other optically active constituents (OACs) such as CDOM or suspended inorganic particles. Natural variability in OACs was temporarily managed by stipulating that OACs were adherent to stable empirical relationships in open-ocean environments, which were operationally defined as *case-1* waters (Morel and Prieur 1977). The case-1 prescription circumvents uncertainty associated with natural variability in the relative content of different OACs so that the color of seawater could be related to individual biogeochemical parameters, namely C_a (Morel 1980; Grew and Mayo 1983; Gordon et al. 1983a). The approach was consistent with contemporary knowledge of the grazing food chain—wherein phytoplankton growth and grazing regulate oceanic ecosystems (Kalle 1966; Steele 1974)—prior to growing recognition of the importance of microbial loop dynamics (Azam et al. 1983; Azam 1998).

The launch of the Coastal Zone Color Scanner (CZCS) satellite mission in 1978 enabled remote assessment of electromagnetic flux emitted from aquatic surface waters in the visible (VIS) spectral domain—hereafter *ocean color*—at large spatial scales (Hovis 1981). CZCS imagery corroborated linkages between ocean color and C_a, e.g., by demonstrating elevated oceanic C_a within eddy-like turbidity patterns (Gordon et al. 1980). Satellite remote sensing of oceanic apparent optical products (AOPs) relies principally on the normalized water-leaving radiance, $[L_W(\lambda)]_{\mathrm{N}}$, or its conjugate product the remote-sensing reflectance, $R_{\mathrm{rs}}(\lambda)$, where λ denotes wavelength. Both were—and continue to be—derived for ocean color satellite observations by ascribing null oceanic flux at the longest signal-limited wavelength to facilitate the partitioning of oceanic and atmospheric observed signals (Gordon et al. 1983b). As technology inexorably advanced, the wavelength associated with null flux was incrementally adjusted from the VIS (i.e., red) to the near-infrared (NIR), and then to the shortwave infrared (SWIR) domain (Gordon 2021)—necessary, in pertinent part, because null flux approximations are inconsistent with the optical properties of the sunlit aquatic environment (Houskeeper and Hooker 2023). Inversions relating $[L_W(\lambda)]_{\mathrm{N}}$ or $R_{\mathrm{rs}}(\lambda)$ waveband ratios to C_a—hereafter ocean chlorophyll (OC) algorithms—were defined for the CZCS mission (Clark 1981) and subse-

quently extended to, for example, the Sea-viewing Wide Field-of-view Sensor (SeaWiFS), the MODerate resolution Imaging Spectroradiometer (MODIS), the Second-Generation Global Imager (SGLI), and other dedicated ocean color missions (Aiken et al. 1996; O'Reilly et al. 1998; Hirata et al. 2014; O'Reilly and Werdell 2019; Isada et al. 2022).

As satellite missions progressed, pigment quantitation from water samples—requisite for developing and validating empirical OC algorithms—evolved from spectrophotometric and fluorometric techniques to high performance liquid chromatography (HPLC) methods (Tyler 1961; Smith et al. 1981; Hooker et al. 2000) and ultimately advanced to ultra HPLC (Suzuki et al. 2015). The trajectory in advancing laboratory methods allowed a concomitant improvement in OC algorithms by including chlorophyll allomers, epimers, and degradation products (Claustre et al. 2004), while lowering pigment uncertainties (Van Heukelem and Hooker 2011; Hooker et al. 2012a). This era of development also included contemporaneous—i.e., occurring during the same time—advances in commercial-off-the-shelf (COTS) *in situ* optical instrumentation. The fixed-wavelength microradiometer (Morrow et al. 2010) reduced the size and weight of instruments while providing purely digital operations with improved dynamic range and accuracy. When coupled with a compact backplane equipped with hydrobaric buoyancy and small digital thrusters (Hooker et al. 2018a), oceanic, coastal, and inland (hereafter *global*) waters could be sampled with the same instrument suite. The new capabilities included autonomous at-sea operations (Hooker et al. 2018b)—plus onshore support measurements and semiautonomous airborne surveys (Guild et al. 2020)—to provide data products for the calibration and validation of the next generation of ocean color satellite missions (Hooker et al. 2018c).

Over the time span of advancing satellite, laboratory, and field capabilities, the architecture of OC algorithms remained largely unchanged, i.e., OC algorithms continued to rely on VIS band ratios (O'Reilly and Werdell 2019), although improvements were obtained, e.g., in oligotrophic waters (Hu et al. 2012) and by simultaneous (VIS) retrieval of multiple OACs (e.g., Garver and Siegel 1997). Continuity of satellite OC algorithm retrievals supports global monitoring of oceanic C_a needed to assess planetary change (McClain et al. 2006), but correspondence between the OC algorithm data product and oceanic C_a content varies significantly in space and time (Dierssen 2010; Sauer et al. 2012). For example, natural variability in OAC relationships (Dierssen and Smith 2000; Siegel et al. 2005) injects regional biases into OC algorithmic products, even for large oligotrophic (purportedly case-1) water bodies, such as the Mediterranean Sea (Claustre et al. 2002; Morel and Gentili 2009). Evolving climate and biological conditions further alter OAC relationships and derivation of optical data products, thereby modifying OC algorithm performance (Sauer et al. 2012). Variability in OAC relationships confounds detection of long-term C_a trends (Dierssen 2010), although trends in the fundamental remote sensing optical data products, e.g., $[L_W(\lambda)]_{\mathrm{N}}$ or $R_{\mathrm{rs}}(\lambda)$, are detectable (Dutkiewicz et al. 2019). Relating changes in oceanic optical properties to specific OACs—e.g., robust estimation of C_a—would help to assess changes in oceanic ecosystem functionality and provide context for cause-and-effect relationships in aquatic environments coinciding with a changing global climate.

Interpreting changes in oceanic AOPs remains challenging—as noted in earlier historical studies, e.g., Yentsch (1960)—due to confounding similarities in the spectral absorption of OACs within the VIS domain relative to the spectral range, resolution, and signal-to-noise ratio (SNR) of legacy ocean color datasets combined with non-adherence to an absolute radiometric scale (Houskeeper and Hooker 2023). An eigenanalysis for modeled aquatic spectra demonstrated optical data products require the shortest available wavelength (modeled to 400 nm at the time), i.e., the spectral end member, to support discrimination of in-

dependently varying CDOM and C_a (Sathyendranath et al. 1987). As field technologies advanced, spectrally expansive above- and in-water observations were shown to provide optimal estimation of CDOM across global conservative waters with maximal independence from C_a (Hooker et al. 2020; Houskeeper et al. 2020a)—with spectral expansivity referring to the spectral range of compliant data products, with the most spectrally expansive observations characterizing the greatest spectral range. Conservative waters—wherein the inflow and outflow of properties constrain the range in the gradient of a constituent (Hooker et al. 2020)—include all case-1 or otherwise oligotrophic waters, while also including significant representation from optically complex, hereafter case-2 (e.g., including coastal and inland), waters. The inclusion of case-2 waters means that regional (or global average) empirical OAC relationships are not invoked, and global algorithms are, thus, more capable of partitioning OACs and maintaining consistent performance under changing environmental conditions (Houskeeper et al. 2020a). New algorithmic approaches included using ratios of the most spectrally separated optical data products (Hooker et al. 2013), especially those from the UV and NIR spectral domains, an approach hereafter termed end-member analysis (EMA) following Hooker et al. (2020). EMA analyses were subsequently extended to single-channel UV algorithms (Hooker et al. 2021b), indicating the importance of accurately deriving UV data products for robust CDOM estimation. VIS algorithms, conversely, were shown to exhibit decreased robustness to variability in OAC relationships relative to non-visible—or invisible (INV)—algorithms (Hooker et al. 2020, 2021b; Houskeeper et al. 2020, 2021, 2022).

In February 2024, the launch of the Plankton, Aerosol, Cloud, ocean Ecosystem (PACE) satellite mission provided state-of-the-art hyperspectral and spectrally extended oceanographic observing using the Ocean Color Instrument (OCI). Hyperspectral observations obtained using OCI are anticipated to reveal new opportunities for satellite observing of the microbial community (Cetinić et al. 2022), but the fidelity of inversions relating hyperspectral features to biological parameters is mostly unknown. The challenges stem, in part, from the use of legacy *in situ* instrumentation, e.g., wherein COTS spectrometers lacking paired microradiometers—i.e., hybridspectral configuration—to mitigate low sampling rate and SNR (Hooker et al. 2022; Houskeeper et al. 2024; Kudela et al. 2024) preclude most INV data products and do not support adherence to an absolute radiometric scale. For example, non-hybridspectral data products frequently rely on a null long-wavelength bias. Reporting of observations at resolutions finer than the nominal resolution of the spectrograph—often in addition to spectral smoothing—further challenges assessments of spectral content and independence of adjacent waveband observations.

Nonetheless, assessments of optical information content of the sunlit aquatic environment are ongoing and indicate adequate derivation of independent OACs is plausible—but not guaranteed. Hooker et al. (2021a) assessed spectral variability of global conservative plus nonconservative water bodies, with the latter modified by internal processes rather than linear mixing of parent or source waters, and demonstrated that a finite number of spectral modes adequately describe the majority of spectral variability across globally representative—and including case-2—waters. Cael et al. (2023) applied an eigenanalysis to legacy smoothed and interpolated hyperspectral observations and demonstrated two components capture over 80% of the variance—with similar information constraint also previously shown for hyperspectral observations of particulate absorption (Cael et al. 2020). Prochaska and Frouin (2025) applied a Bayesian framework testing whether hyperspectral observations could overcome an acknowledged degeneracy in radiative transfer (wherein multiple combinations of spectral absorption and backscattering are possible and reasonable) and reasoned that hyperspectral ocean color inversions are ill-posed.

Quantifying information content using eigenanalyses can be challenging: dimensionality bias can produce an incomplete perspective of information content by decreasing the variance captured by spectrally broadband versus spectrally narrow features. The latter—when not spectrally diluted—often provide informative and exploitable information (Houskeeper et al. 2020b). Furthermore, despite known challenges in finite information content and spectral degeneracy, new information sources have steadily been discovered in aquatic optics throughout the field's long history. Legacy satellite observations of phytoplankton fluorescence added new biological observables complementary to—but distinct from—OC observations, while also minimizing vulnerabilities to atmospheric correction (Letelier and Abbott 1996). Hyperspectral airborne investigations leveraged spatially cohesive optical signatures associated with harmful algal bloom (HAB) events (e.g., Lee and Carder 2005; Kudela et al. 2015) and retrieved macroalgal physiological condition (Bell and Siegel 2022). Mixture density networks (MDNs) and inverse models that incorporate prior knowledge (e.g., leveraging a Bayesian framework) improve the management of degeneracy in radiative transfer and are forthcoming for PACE science objectives (O'Shea et al. 2021; Erickson et al. 2020, 2023).

A fundamental and timely question remains how to best leverage PACE observations within the trajectory of ocean color satellite missions to improve robust quantification of phytoplankton biomass and support characterization of biogeochemical changes in the sunlit ocean. Advancing this trajectory—wherein confounding effects of CDOM have been reported for at least over seven decades (e.g., Jerlov 1951; Kalle 1966; Yentsch 1960; Sathyendranath et al. 1987) and recent studies suggest high spectral autocorrelation and degeneracy (Hooker et al. 2021a; Cael et al. 2023; Prochaska and Frouin 2025)—requires improvements in understanding the primary sources of marine spectral variability. Aquatic spectral variability is assessed herein for natural aquatic environments using three independent bio-optical datasets, and analyses are replicated using the datasets' oligotrophic subsets, as applicable. Associations relating spectral variability to contemporaneous OAC field observations indicate the relative importance of individual constituents in modifying AOP spectral shapes. The key findings document, as follows: continuing challenges in retrieving C_a using observations constrained to the VIS domain; opportunities to advance independent OAC retrieval using spectrally expansive observations; greater variability in OAC relationships for oceanic environments than formerly ascribed; and primacy of CDOM absorption in spectrally modifying AOPs. The outcomes reflect a greater complexity of OAC dynamics consistent with advancing knowledge of the microbial loop, wherein new trophic pathways have been discovered contemporaneously with the aforementioned advances in optical oceanography (Azam et al. 1983; Azam 1998; Fenchell 2008).

2 Methods

Aquatic observations presented herein correspond to published datasets of above- and in-water observations of the emergent aquatic light field. The datasets rely on instrumentation, including legacy and state-of-the-art (SOTA) radiometers and spectrometers, corresponding to a technological trajectory spanning decades of improvements applicable across the hardware, software, and processing domains. For example, legacy oceanic radiometric observations were obtained using in-water optical instrument suites consisting of primarily analog radiometers mounted on rapidly descending, so-called rocket-shaped, profilers (Morel et al. 2007). Rocket-shaped profilers confer coarse vertical sampling resolution (VSR), deep extrapolation intervals,

and a limited spectral range of observations. These deficiencies prevent observations in shallow, highly turbid, or high-flow-rate waters, meanwhile introducing uncertainties including, but not limited to, those arising from depth aliasing, expanded extrapolation distances, and low signal (Hooker et al. 2002). The Compact-Optical Profiling System (C-OPS) introduced a novel kite-shaped backplane (Morrow et al. 2010) to slow the rate of descent of the profiler while ensuring planar stabilities, along with corresponding improvements in microradiometer technology to advance sampling rates from usually less than 1 Hz for legacy spectrometers to routinely 15 Hz, with improved SNR and dynamic range. The addition of digital thrusters, i.e., the Compact-Propulsion Option for Profiling Systems (C-PrOPS), ensures planar stability at the initiation of a cast and mitigates adjacency effects by supporting navigation away from a ship or dock (Hooker et al. 2018a). The C-OPS with C-PrOPS technologies greatly expand the spectral range of in-water observations and reduce uncertainties in optical data products relative to legacy observations, in part, by supporting improved VSR (as fine as 0.9 mm) to enable accurate derivation of the central tendency of the wave-focusing field (e.g., Zaneveld et al. 2001) and reducing the initial depths of upwelling radiance observations, $L_u(z,\lambda)$, to as shallow as 0.3 m, i.e., approximately the length of the downward-pointing radiance radiometer (Hooker et al. 2020). Improvements in sampling rates support management of high-frequency, non-Gaussian variability in flux observed by an above- or in-water instrument (due to glint and wave focusing, respectively), thereby expanding the spectral range of optical data products to preserve INV information and supporting necessary corrections to retain information associated with spectral signal amplitudes or brightness (Houskeeper et al. 2023, 2024).

SOTA advances in hyperspectral instrumentation include hybridspectral sensing configurations (Hooker et al. 2018c, 2022), wherein microradiometer and spectrograph observations are obtained in concert, with the multispectral radiometer providing necessary quality control for the hyperspectral spectrography. The quality control is desirable, in part, because COTS spectrographs suffer from slower integration times, narrower dynamic range, and a degraded signal-to-noise ratio (SNR) relative to COTS microradiometers (Houskeeper et al. 2024; Kudela et al. 2019, 2024). In addition, a radiance control arm positions the $L_u(z,\lambda)$ aperture near the water surface and approximately aligned with the upward-pointing irradiance radiometer (Hooker et al. 2018a). Improvements in the number of spectrograph pixels to as high as 2,048, denoted $\lambda_{2,048}$, with 3 nm resolution is in keeping with PACE mission requirements for vicarious calibration exercises (Hooker et al. 2012b). For comparison, spectral resolution of legacy spectrometers include λ_{256} spectrographs sampling the visible domain as often as every 3 nm, although bandwidths can be on the order of 10 nm resolution (Seabird 2024), i.e., comparable to COTS microradiometers but conferring lower SNR, dynamic range, and sampling rates. Legacy spectrometer observations are often interpolated to synthetic 1 nm intervals—which eases archiving requirements for dissimilar spectral configurations but over-reports the spectral resolution—and smoothed, e.g., using a 5 nm moving mean, which removes spectral information. Spectral smoothing and interpolation creates challenges for comparisons of spectral resolution between instruments.

2.1 Datasets

Three AOP datasets are considered herein, with biogeochemical parameters included when available. No dataset includes SOTA hybridspectral observations, because hybridspectral observations are presently too sparse in terms of the number of

observations. Naming conventions for the datasets correspond to the applicable published literature, and their descriptions plus quality control are briefly summarized, as follows:

RSE2007 Oceanic $[L_W(\lambda)]_{\mathrm{N}}$ observations obtained using primarily analog free-falling instrument suites during open-ocean field campaigns, including a minority of stations in the coastal zone, are described by Morel et al. (2007). Three stations (0.5% of the dataset) corresponding to surface mucilage and coccolithophore blooms were removed, and observations were thinned so that only one spectrum was retained per station, although at least three spectra were generally obtained per station. The dataset is defined as RSE2007 and consists of contemporaneous observations of C_a and $[L_W(\lambda)]_{\mathrm{N}}$ spanning the VIS domain.

RSE2021 Global $[L_W(\lambda)]_{\mathrm{N}}$ observations obtained using the digital C-OPS with C-PrOPS instrument suite deployed at oceanic (31.2%), coastal (36.6%), and inland (32.2%) water sites are presented by Houskeeper et al. (2021). Optical observations featured average VSR of 6.0 mm (maximum VSR of 0.9 mm) and initial $L_u(z,\lambda)$ observations routinely obtained at 0.3 m, i.e., the depth corresponding to the length of the $L_u(z,\lambda)$ radiance radiometer. Contemporaneous observations of C_a and the absorption coefficient of CDOM at 440 nm, $a_{\mathrm{CDOM}}(440)$, supported global algorithm development (Hooker et al. 2020; 2021b; Houskeeper et al. 2021; 2022) and confirmed expansive ranges for OACs. For example, observed $a_{\mathrm{CDOM}}(440)$ spanned 0.001–2.305 m^{-1}, and the spectral slope of CDOM spanned 0.0095–0.0410 nm^{-1}. Conservative versus nonconservative designations were objectively determined prior to sampling (Hooker et al. 2020) and only the conservative observations are used herein. The dataset is defined as RSE2021 and consists of contemporaneous observations of C_a, $a_{\mathrm{CDOM}}(440)$, and $[L_W(\lambda)]_{\mathrm{N}}$ spanning the INV plus VIS domains.

RSE2022 Oceanic hyperspectral observations with concurrent samples of phytoplankton pigments are described in Kramer et al. (2022) and accessed via Kramer et al. (2021). Observations span the open and coastal oceans, with C_a ranging from 0.019–4.150 $\mathrm{mg\,m}^3$. Optical observations were obtained using rocket-shaped profilers and buoys equipped with legacy COTS spectrometers, including systems with 10 nm bandwidths (Seabird 2024). Vertical profiles of measurements obtained deep in the water column were averaged to 2 m or coarser depth bins and extrapolated to the surface (Taylor et al. 2011). The compiled dataset overcomes differences in spectral configuration between various COTS systems by spectral interpolation to 1 nm resolution and mitigates optical data product artifacts associated with legacy profiling systems by subsetting spectra to VIS wavelengths, smoothing using a 5 nm moving mean, and additionally removing 19% of the observations by visual inspection (Kramer et al. 2022). The dataset is defined as RSE2022 and consists of contemporaneous observations of C_a and $[L_W(\lambda)]_{\mathrm{N}}$ spanning the VIS domain.

The field campaign supporting the RSE2021 dataset (Hooker et al. 2020; Houskeeper et al. 2021a) was designed to capture maximal range in global OAC variability, including the clearest waters to the most sedimented, eutrophic, or humic-rich waters. The clearest waters were observed, e.g., near Kawaihae, HI, and in Crater Lake, OR, both of which corresponded to an $a_{\mathrm{CDOM}}(440)$ value of 0.001 m^{-1}, approaching so-called pure seawater. The turbidity limit is inherently undefined, e.g., there

is no known natural maximum to turbidity, but the RSE2021 dataset greatly exceeds the turbidity range of the RSE2007 and RSE2022 datasets because it contains significant representation from inland waters. The absolute radiometric data products from RSE2021, therefore, constitute a reasonably expansive range in optical properties such that open and coastal ocean optical data products—collected with legacy systems and with degraded SNR and VSR relative to the RSE2021 dataset—that lie substantially outside of the range of global conservative data products indicate likely nonphysical or biased observations. Quality control based on the threshold removal of nonphysical radiometric brightness was thus applied, albeit only using VIS data products (given that RSE2007 and RSE2022 are restricted to the VIS domain), following sensitivity testing. Briefly, more stringent quality control metrics (e.g., requiring observations to comply with the brightness ranges of the globally representative RSE2021 dataset) produced similar eigenfunction shapes but reduced the size of RSE2022 by approximately 50%. Alternately, relaxed quality control metrics permitting radiometric values up to 50 times darker than the global dataset did not alter the primary findings of RSE2022, and no threshold relaxation affected the primary findings of RSE2007. Considering acknowledged uncertainties relating to hardware, processing, data acquisition, and water mass differences, the quality control thresholds were performed within this insensitivity range by filtering: RSE2022 observations radiometrically darker than the darkest RSE2021 observations (by at least approximately 50%); and RSE2007 observations radiometrically darker or brighter than RSE2021 observations (approximately) at corresponding green (555 nm) or red (683 nm) wavebands, respectively. The brightness filtering resulted in the removal of 18 of 590 observations (3%) from RSE2007 and 22 of 144 observations (15%) from RSE2022. Because both RSE2007 and RSE2022 include observations obtained using rocket-shaped profilers, the difference in compliance is likely due to differences in the optical sensing technologies—i.e., COTS spectrometers (RSE2022) are more degraded by slow sampling rates, reduced dynamic range, and low SNR (Houskeeper et al. 2024; Kudela et al. 2024). The final number of compliant observations for the RSE2007, RSE2021, and RSE2022 datasets are 572, 612, and 122, respectively. The quartile summary of the biochemical quantities corresponding to each dataset is shown in the supplemental materials.

2.2 Biogeochemical algorithms

OC algorithms are derived for each dataset using the OC4 formulation (O'Reilly and Werdell, 2019), a fourth-order polynomial with coefficients configured to match the spectral configurations of the SeaWiFS mission and others, as follows:

$$C_a \;=\; a_0 \;+\; \sum_{n=1}^{4} a_n \left(\log_{10} \frac{R_{\mathrm{rs}}(\lambda_b)}{R_{\mathrm{rs}}(555)} \right)^n, \qquad [\mathrm{mg\ m^{-3}}] \qquad (1)$$

where the a terms are fitting constants, n is the order index, and λ_b indicates the wavelength corresponding to the radiometrically brightest $R_{\mathrm{rs}}(\lambda)$ observation at applicable blue SeaWiFS wavelengths (i.e., 443, 489, and 510 nm). The OC4 formulation was selected to ensure maximal applicability to RSE2007, which supported the fewest spectral channels, while maintaining a consistent formulation across datasets. Nonetheless, other OC formulations are anticipated to produce similar results.

The RSE2021 dataset contains contemporaneous observations of $a_{\mathrm{CDOM}}(440)$, but contemporaneous $a_{\mathrm{CDOM}}(440)$ observations were not contained in the other datasets. Consequently, an EMA algorithm derived using both case-1 plus case-2 waters to minimize biases in algorithm fitting corresponding to regional OAC covariances (Houskeeper et al. 2021) is applied to the RSE2007 and RSE2022 datasets to estimate $a_{\mathrm{CDOM}}(440)$. The approach leverages ratios of $[L_W(\lambda)]_{\mathrm{N}}$, hereafter $\Gamma_{\lambda_2}^{\lambda_1}$, using

the most spectrally separated wavelengths, which mitigates confounding signals from internal VIS wavelengths to improve the independence of $a_{\mathrm{CDOM}}(440)$ retrieval from C_a variability (Houskeeper 2020a; 2021). The EMA formulation for above-water, spectrally expansive observations is modeled as a power-law relationship, as follows:

$$a_{\mathrm{CDOM}}(440) \;=\; 0.242 \left[\Gamma_{\lambda_2}^{\lambda_1}\right]^{-0.961} . \qquad [\mathrm{m}^{-1}] \qquad (2)$$

The λ_1 and λ_2 values correspond to the shortest and longest wavelengths, respectively, supported by the given dataset. EMA algorithm performance improves with expanding spectral range (Hooker et al. 2020; Houskeeper et al. 2021), but robust performance of VIS algorithms has been documented for ocean color observations (Houskeeper et al. 2021; Hooker et al. 2021b), and simulated geostationary weather satellite observations (Houskeeper et al. 2022). The uncertainty of EMA, however, is anticipated to increase when applied to optical data products obtained using spectrometers and rocket-shaped profilers (e.g., RSE2022) due to increasing extrapolation depths combined with strong attenuation of signal, particularly within the end-member spectral domains (Kudela et al. 2019, 2024; Hooker et al. 2020; Houskeeper et al. 2021, 2024).

2.3 Spectral comparisons

The dimensionality for each of the RSE2007, RSE2021, and RSE2022 datasets is reduced using an eigenanalysis of observations of the square root of $[L_W(\lambda)]_{\mathrm{N}}$, hereafter $[L_W(\lambda)]_{\mathrm{N}}^{0.5}$, because the square-root transformation was found to improve normality in the blue spectral domain based on the Shapiro-Wilk test statistic, W. Nonetheless, sensitivity testing was performed by reproducing the eigenanalyses using natural-log transformation, as well as no transformation. The eigenanalysis captures linear patterns of variability in the datasets and is performed, as follows:

$$\boldsymbol{C}\boldsymbol{\psi} \;=\; \boldsymbol{\Lambda}\boldsymbol{\psi} \,, \qquad [\mu\mathrm{W\ cm^{-2}\ nm^{-1}\ sr^{-1}}] \qquad (3)$$

in which $\boldsymbol{C}$ is the covariance matrix of the $[L_W(\lambda)]_{\mathrm{N}}^{0.5}$ values for each dataset, with square dimensions corresponding to the number of wavelengths in each dataset. The $\boldsymbol{\psi}$ value denotes the eigenfunction matrix (matrices are denoted by bold symbols). The diagonal matrix $\boldsymbol{\Lambda}$ contains scalar information corresponding to the variance captured by each eigenfunction. The mean spectra from each dataset, plus reconstructed spectra indicating the effects of the leading and secondary eigenfunctions, are shown in the supplemental materials.

Scalar amplitude functions quantifying the stretching and compressing necessary to represent the $[L_W(\lambda)]_{\mathrm{N}}^{0.5}$ values in the coordinates defined by $\boldsymbol{\psi}$ are derived by projecting the optical observations onto $\boldsymbol{\psi}$, as follows:

$$\boldsymbol{S} \;= [L_W(\lambda)]_{\mathrm{N}}^{0.5}\ \boldsymbol{\psi} \,, \qquad [\mu\mathrm{W\ cm^{-2}\ nm^{-1}\ sr^{-1}}]^{0.5} \qquad (4)$$

in which $\boldsymbol{S}$ is the scalar amplitude function matrix with columns, $\boldsymbol{S}_i$, corresponding to scalar amplitude functions for individual eigenfunctions, $\boldsymbol{\psi}_i$ (where i is the eigenfunction index). The length of each $\boldsymbol{S}_i$ column is the number of observations in the underlying dataset.

Comparisons between available biogeochemical parameters—e.g., field plus algorithm values of C_a and $a_{\mathrm{CDOM}}(440)$—and the $\boldsymbol{S}$ matrices are derived using Pearson's correlation coefficient, $\rho_{i,x}$, where i is the eigenfunction index and x represents

the biogeochemical quantity under comparison with the corresponding column i of the $\boldsymbol{S}$ matrix. Because biogeochemical parameters covary, e.g., C_a and $a_{\mathrm{CDOM}}(440)$ are strongly correlated (Morel et al. 1977), partial correlation coefficients (Fisher 1924) are derived using notation of $\rho_{i,x|y}$ to indicate correlation of $\boldsymbol{S}_i$ with biogeochemical quantity x after adjusting for covariance with the biogeochemical quantity y. The subscript representations used herein are g (for the legacy linkage to *gelbstoff* absorption), C_a, EMA, and OC4, which correspond to field observations of $a_{\mathrm{CDOM}}(440)$, field observations of C_a, EMA algorithm observations of $a_{\mathrm{CDOM}}(440)$, and OC algorithm observations of C_a, respectively. The R^2 value is also utilized when quantifying variance in biogeochemical field observations captured by a remote sensing algorithm. Pearson's correlation coefficients comparing C_a and $a_{\mathrm{CDOM}}(440)$—measured or derived algorithmically—for the RSE2021, RSE2022, and RSE2007 datasets are 0.812, 0.912, and 0.909, respectively.

3 Results

Eigenanalyses indicate that the first, second, and third eigenfunctions capture 60%, 32%, and 5% ($\pm$ 1%), respectively, of the variance within each data set. The decomposition of the majority of the variance in the aquatic spectral dataset into three functions is consistent with the findings of Hooker et al. (2021a) and Cael et al. (2023), as follows: Hooker et al. (2021a) demonstrated that spectral changes corresponding to increasing optical complexity were represented using only five spectral modifications—two of which are similar but in opposing directions and would therefore correspond to a single eigenfunction, plus one which is characterized by the emergence of multiple peaks and corresponds to severely complex waters; and Cael et al. (2023) showed the aquatic optical spectra could be compressed into three (or four) principal components with minimal information loss. The Cael et al. (2023) dataset includes significant representation by RSE2022, and similar results obtained herein using RSE2022 provide support for the normalizations and methodologies are comparable.

3.1 Eigenfunction characteristics

The three eigenfunctions capturing maximal variance for the RSE2021, RSE2022, and RSE2007 datasets are presented as a function of λ in Fig. 1. Similarities in the eigenfunction spectral shapes are apparent across datasets for each of $\boldsymbol{\psi}_1$, $\boldsymbol{\psi}_2$, and $\boldsymbol{\psi}_3$, and similar shapes are likewise produced by the eigenanalyses for the oligotrophic (C_a less than 0.5 mg m^{-3}) datasets (gray). Although the tertiary eigenfunction explains only approximately 5% of the variance for each dataset, the spectral shapes represented by $\boldsymbol{\psi}_3$ are consistent across all datasets. The similarities in $\boldsymbol{\psi}$ are summarized, as follows:

- $\boldsymbol{\psi}_1$ Reversed sign for the anomaly between shorter (UV/blue) and longer (red/NIR) wavelengths;
- $\boldsymbol{\psi}_2$ Maximum amplitude for the anomaly in the blue/green domain (i.e., a peak), plus the emergence of spectral features in the spectral vicinity of the C_a fluorescence peak (except for RSE2007 which lacks the requisite spectral resolution in the applicable wavelength domain); and
- $\boldsymbol{\psi}_3$ Reversed sign for the anomaly between interior VIS and exterior VIS plus INV wavelengths.

Individual eigenfunctions are not anticipated to be singularly associated with variability in a specific OAC (e.g., because OACs covary), but the spectral shapes of the eigenfunctions may nonetheless be assessed qualitatively in light of known spectral properties of individual OACs to identify similarities to spectral properties of OACs. The spectral shapes of the eigenfunctions are primarily considered as a function of absorption processes, which generally—notwithstanding absorption band effects (Zaneveld and Kitchen 1995)—confer stronger spectral dependencies than scattering processes (Kirk 2011). Differences in brightness or peak height, however, could produce reversals in the VIS and INV anomalies, possibly relevant to $\boldsymbol{\psi}_3$, but are not assessed herein.

Spectral features associated with C_a absorption—e.g., a relative maximum or minimum in the blue and red domains (Brigidare et al. 1990), or a fluorescence peak (Letelier and Abbott 1996)—are most clearly visible in $\boldsymbol{\psi}_2$. For example, although both $\boldsymbol{\psi}_1$ and $\boldsymbol{\psi}_2$ indicate spectral dependencies in the vicinity of 670 nm for RSE2022, the amplitude of the anomaly (as a percentage of the range expressed in $\boldsymbol{\psi}$) is on the order of 10% versus 50% for $\boldsymbol{\psi}_1$ and $\boldsymbol{\psi}_2$, respectively. Similarly, all datasets indicate a blue- or green-domain peak (or trough) in $\boldsymbol{\psi}_2$ of greater magnitude than in $\boldsymbol{\psi}_1$. Conversely, spectral darkening of shorter wavelengths with minimal spectral features is most consistent with CDOM absorption (Kalle 1966; Jerlov 1968; Siegel et al. 2013), which approximately describes $\boldsymbol{\psi}_1$. The similarity of spectral shapes for $\boldsymbol{\psi}$ between the complete and oligotrophic (gray) datasets—as well as the differences in the range and mean properties of the water bodies sampled across the three datasets—indicates that spectral variability is consistent across datasets and not primarily a function of differences in sampling within each dataset.

The spectral shape comparisons, while qualitative, can be verified using partial correlation coefficients comparing the $\boldsymbol{S}$ amplitude functions and corresponding values of C_a and $a_{\mathrm{CDOM}}(440)$, while controlling for $a_{\mathrm{CDOM}}(440)$ and C_a, respectively. RSE2021 includes field observations of $a_{\mathrm{CDOM}}(440)$ and C_a, whereas RSE2022 and RSE2007 include field observations of C_a but not $a_{\mathrm{CDOM}}(440)$. Consequently, the RSE2021 results are considered first. Briefly, partial correlation coefficients for $\boldsymbol{S}_1$ indicate maximal correlation to $a_{\mathrm{CDOM}}(440)$ values ($\rho_{1,g\,|\,C_a} = 0.80$) compared to C_a values ($\rho_{1,C_a\,|\,g} = 0.11$), whereas partial correlation coefficients for $\boldsymbol{S}_2$ indicate maximal correlation to C_a values ($\rho_{2,C_a\,|\,g} = -0.22$) compared to $a_{\mathrm{CDOM}}(440)$ values ($\rho_{2,g\,|\,C_a} = 0.02$). Both of the maximal correlation results are highly significant ($\mathrm{P} < 0.01$), whereas the non-maximal comparisons are not ($\mathrm{P} > 0.18$). These results are in keeping with the qualitative assessments of the spectral shapes of $\boldsymbol{\psi}$, in which $\boldsymbol{\psi}_1$ was shown to produce spectral dependencies more consistent with that of CDOM as documented in the literature, whereas $\boldsymbol{\psi}_2$ was shown to produce spectral dependencies with features most closely matching the spectral domains of C_a absorption or fluorescence properties. The sign of the significant partial correlation coefficients is also consistent with the $\boldsymbol{\psi}$ phase and the spectral shape of absorption for $a_{\mathrm{CDOM}}(440)$ and C_a. For example, negative anomalies in the blue and UV indicated by $\boldsymbol{S}_1$ produce a positive correlation with $a_{\mathrm{CDOM}}(440)$ values, whereas positive (peak-like) anomalies in blue wavelengths indicated by $\boldsymbol{S}_2$ produce a negative correlation with C_a values.

Estimates of $a_{\mathrm{CDOM}}(440)$ are also obtained for RSE2022 and RSE2007 using EMA. The partial correlation values derived using the EMA product for the RSE2022 dataset are very similar to those derived using the RSE2021 dataset. Briefly, partial correlation coefficients for $\boldsymbol{S}_1$ indicate maximal correlation to $a_{\mathrm{CDOM}}(440)$ values, whereas partial correlation coefficients for $\boldsymbol{S}_2$ indicate maximal correlation to C_a values. Both of the maximal correlation results are highly significant ($\mathrm{P} \leq 0.01$),

whereas the other comparisons are not ($P > 0.38$). The results from the RSE2007 dataset are different from those of RSE2022 and RSE20221 as follows: Partial correlation coefficients for $\boldsymbol{S}_1$ still indicate maximal correlation to $a_{\mathrm{CDOM}}(440)$ values, but correlation to C_a is also significant, albeit lower. Partial correlation coefficients for $\boldsymbol{S}_2$ indicate slightly higher correlation to $a_{\mathrm{CDOM}}(440)$ values compared to C_a values, and both are highly significant ($P < 0.01$). The differences are consistent with decreasing independence of EMA to C_a variability with decreasing spectral range (Houskeeper et al. 2021), plus maximal representation of open-ocean waters in RSE2007.

Correlation tests comparing the $\boldsymbol{S}$ functions with OC algorithm values find the strongest relationships for all datasets correspond to $\boldsymbol{S}_1$, ranging from 0.89 to 0.96. Less strong, albeit significant, relationships are found for $\boldsymbol{S}_2$, ranging from -0.38 to 0.16. The significant results for comparing OC algorithm values with $\boldsymbol{S}_2$ are likely due to the test not accounting for covariance in C_a and $a_{\mathrm{CDOM}}(440)$. For example, the correlation coefficients using RSE2021—the highest quality dataset and the only dataset to contain contemporaneous C_a and $a_{\mathrm{CDOM}}(440)$ observations—still indicate strong significant relationships when comparing OC algorithm values and $\boldsymbol{S}_1$ while controlling for C_a, but indicate weak and insignificant relationships when comparing OC algorithm values and $\boldsymbol{S}_2$ while controlling for $a_{\mathrm{CDOM}}(440)$.

In summary, analyzing the shapes of the eigenfunctions and comparing correlation matrices using the eigenfunction amplitude functions and contemporaneous OAC field observations yields five observations, as follows: a) spectral shapes of the eigenfunctions for an individual dataset are distinct from one another and characterized by maxima and minima consistent with OAC absorption properties described in the applicable literature; b) the transition in spectral shapes for the first, second, and third eigenfunctions are similar across all datasets and the oligotrophic subsets of those same datasets; c) the interpretation of the eigenfunction shapes based on the spectral positioning of observed minima and maxima is consistent with the results of the partial correlation tests (e.g., $\boldsymbol{\psi}_1$ confers spectral similarities with CDOM absorption and the corresponding amplitude function produces highly significant correlation to CDOM absorption); d) the OC algorithm produces stronger correlation to $\boldsymbol{S}_1$ than $\boldsymbol{S}_2$; and e) $\boldsymbol{S}_1$ produces stronger correlation to field observations of $a_{\mathrm{CDOM}}(440)$ than C_a values. Although RSE2007 indicates relatively more OAC covariance than RSE2021 and RSE2022, as anticipated, the key findings are nonetheless consistent.

3.2 OC algorithm retrievals

Section 3.1 documents the primary mode of variability in $[L_W(\lambda)]_{\mathrm{N}}$ for all datasets is maximally correlated with $a_{\mathrm{CDOM}}(440)$ and OC algorithm retrievals (Fig. 1). The observed linkage between OC algorithm sensitivity and $a_{\mathrm{CDOM}}(440)$ variability is further investigated by comparing OC algorithm retrievals with field observations of C_a, plus field or algorithmic $a_{\mathrm{CDOM}}(440)$ observations (Fig. 2), as applicable. The coherence of a relationship between the OC algorithm and C_a observations is assessed using R^2, and differences across datasets in the coherences observed are consistent with differences in the oceanic contributions of each dataset, as follows: a) the least coherence is associated with RSE2021, which contains similar representation of oceanic, coastal, and inland waters and, therefore, consists of a high percentage of optically complex water types; b) intermediate coherence is associated with RSE2022, which contains coastal and inland water observations and also consists of data products obtained with coarse VSR, low SNR, and deep sampling intervals far from the surface; and c) the most coherence is associated

with RSE2007, which corresponds to primarily open-ocean waters and was obtained with technology most similar to that used in the earliest OC algorithm derivations.

The coherence of the relationship between the OC algorithm and $a_{\mathrm{CDOM}}(440)$ observations is also assessed using R^2, and improvements in R^2 are observed for the $a_{\mathrm{CDOM}}(440)$ comparison relative to the C_a comparison. The magnitude of the improvements are likewise consistent with the differences in the oceanic contributions of each dataset, as follows: a) maximal improvement in coherence is indicated by RSE2021; b) intermediate improvement in coherence is indicated by RSE2022; and c) minimal improvement in coherence is indicated by RSE2007. Although the differences observed are slight for RSE2007 (indicating the dataset adheres well to case-1 waters), all datasets nonetheless produce stronger R^2 values when comparing the OC algorithm values to $a_{\mathrm{CDOM}}(440)$ versus C_a. The findings are consistent with those of Sect. 3.1, in which the OC algorithm products produced maximal correlation to $\boldsymbol{\psi}_1$, which in turn was maximally associated with $a_{\mathrm{CDOM}}(440)$—and not with C_a.

3.3 Information content of invisible and visible data products

Information content as a function of spectral range is further investigated using the RSE2021 dataset, which confers the most expansive spectral range (313–865 nm). Scree plots indicating variance captured for eigenfunctions $\boldsymbol{\psi}_1$–$\boldsymbol{\psi}_6$ are shown for three spectral partitions of the RSE2021 dataset in Fig. 3, as follows: UVN21 corresponds to the full spectral range of RSE2021 data products (i.e., UVN21 is identical to RSE2021); INV21 corresponds to the invisible data products of RSE2021 (i.e., wavelengths shorter than 400 nm or longer than 700 nm); and VIS21 corresponds to the visible data products of RSE2021 (i.e., wavelengths within 400–700 nm). Pearson's correlation coefficients in Fig. 3 compare the $\boldsymbol{S}_1$ and $\boldsymbol{S}_2$ amplitude functions with C_a and $a_{\mathrm{CDOM}}(440)$ values. The comparisons explore the results of Fig. 1, wherein $\boldsymbol{\psi}_1$ and $\boldsymbol{\psi}_2$ are found to confer information likely most related to $a_{\mathrm{CDOM}}(440)$ and C_a, respectively.

Variances captured by the UVN21, INV21, and VIS21 datasets decrease rapidly with increasing eigenfunction number. Considering the UVN scree plot as a baseline, $\boldsymbol{\psi}_1$ captures similar variance and $\boldsymbol{\psi}_2$ captures more variance for the VIS21 dataset, whereas $\boldsymbol{\psi}_1$ captures more variance and $\boldsymbol{\psi}_2$ captures less variance for the INV21 dataset. Information content of INV21 is thus compressed more fully into a singular eigenfunction, which captures nearly 78% of the variance in the dataset. The INV21 leading eigenfunction is very strongly associated with variability in $a_{\mathrm{CDOM}}(440)$, and the correlation coefficient comparing $\boldsymbol{S}_1$ and $a_{\mathrm{CDOM}}(440)$ for the INV21 dataset, -0.96, approaches that of a perfect inverse correlation. The INV21 secondary eigenfunction indicates minimal adherence to either $a_{\mathrm{CDOM}}(440)$ or C_a, with correlation values of -0.07 and -0.09, respectively. Very strong correlation of $\boldsymbol{\psi}_1$ with $a_{\mathrm{CDOM}}(440)$, no correlation of $\boldsymbol{\psi}_2$ with C_a, and very low explanatory power of the subsequent eigenfunctions for INV21 indicate photosynthetic pigmentation confers minimal spectral modifications in the INV domain besides those covarying with $a_{\mathrm{CDOM}}(440)$, consistent with previous EMA assessments (Hooker et al. 2020; Housekeeper et al. 2021). Inorganic particles, not measured in any of the datasets assessed herein, modify INV data products (e.g., Doron et al. 2011) and may contribute to the variance captured in INV21 $\boldsymbol{\psi}_2$. Inorganic particles are not assessed herein due to the sparsity of independent particle field observations contemporaneous and co-registered with the AOP observations.

Information content of VIS21 is less compressed into a singular eigenfunction (although the leading eigenfunction still dominates), and the correlation coefficients comparing $\boldsymbol{\psi}_1$ and $\boldsymbol{\psi}_2$ to $a_{\mathrm{CDOM}}(440)$ and C_a are more similar than for the

INV21 dataset. Similar correlation coefficient values indicate spectral modifications associated with $a_{\mathrm{CDOM}}(440)$ and C_a are less readily partitioned using only VIS data products, despite the high SNR of the RSE2021 dataset. The result suggests independent derivation of C_a is more challenging using VIS observations in the absence of UV and NIR information. Spectrally expansive (UV, VIS, and NIR) observations thus support more accurate and independent retrievals of $a_{\mathrm{CDOM}}(440)$ compared to VIS observations wherein $a_{\mathrm{CDOM}}(440)$ and C_a signals are maximally confounded.

4 Discussion

4.1 Historical formulations invoke primacy of chlorophyll *a*

The generalized transition from bluer oceanic to greener coastal waters is routinely mapped with OC algorithms to increases in the near-surface abundance of oceanic phytoplankton (McClain et al. 2006). The theoretical basis for OC algorithms is the absorption of blue light by biogenic compounds (Morel 1983; Gordon 1983a), including C_a but also other covarying OACs, e.g., $a_{\mathrm{CDOM}}(440)$, various secondary photopigments, and detritus. OC algorithms and the ocean color activities supporting their development and validation historically pose C_a as the predominant determinant of oceanic optical properties, demonstrated in pertinent part, as follows: a) OC algorithms regress ratios of oceanic reflectances to C_a only and do not include secondary photopigments or dissolved compounds (O'Reilly et al. 1998); b) C_a field observations are more ubiquitous than $a_{\mathrm{CDOM}}(440)$ field observations (Werdell and Bailey 2002); c) spatiotemporal differences in OC algorithm performance are routinely mitigated using regional tunings, in which OC algorithms are refit for a subset of global waters using regional field observations of R_{rs} and C_a—but not other parameters such as $a_{\mathrm{CDOM}}(440)$; and d) primacy of C_a is consistent with the published literature regarding the modeling of optical properties of case-1 water masses. For example, Morel 1988 leverages C_a for quantifying oceanic living and detrital algal material, as well as the depth of the euphotic layer.

Bio-optical formulations to derive optical properties as a function of C_a do not explicitly require primacy of C_a. Rather, algorithms parameterize the evolution in optical properties—notwithstanding observational challenges—based on empirically observed relationships (Morel and Prieur 1977; Morel 2009). Nonetheless, variability in OAC relationships exists for *in situ* water bodies (Hansell and Orellana 2021). Empirical approaches such as the OC family of algorithms must mitigate observational artifacts (Uitz et al. 2006; Kudela et al. 2019; 2024), plus regional variability and nonlinearity in OAC relationships (Morel 2009). Regional variability in OAC relationships manifests with the need to tune VIS algorithms for specific water bodies. For example, tunings for arctic waters account for higher $a_{\mathrm{CDOM}}(440)$ relative to C_a content (Matsuoka et al. 2013; Lewis and Arrigo 2020), and, conversely, tunings for antarctic waters account for lower $a_{\mathrm{CDOM}}(440)$ relative to C_a (Dierssen and Smith 2000). Nonlinear OAC relationships are in keeping with highly nonlinear formulations of OC algorithms, e.g., the number of power terms used in the polynomial model to fit C_a to ratios of R_{rs} is routinely four (Morel 2009; O'Reilly and Werdell 2019). Linearity and loglinearity of CDOM algorithms were shown previously to be a function of the spectral separation of algorithm end members (Hooker et al. 2020, 2021a; Houskeeper et al. 2021). The results herein suggest that differences in linearity correspond, in part, to differences in the separability of signals associated with the spectral range.

4.2 Observations indicate primacy of CDOM over C_a

The eigenanalyses herein assess the primacy of C_a using three independent datasets, plus the oligotrophic subsets of each dataset. Phytoplankton absorption—rather than C_a—was not tested due to the limitations of the datasets. Although phytoplankton absorption is generally considered to correspond to C_a, processes such as pigment packaging can add variability relevant to the correlations presented herein (Bricaud et al. 2004). Nonetheless, despite significant differences in the spectral range and uncertainties of the datasets due to differences in hardware, software, and processing implemented, as well as the water bodies assessed plus variability associated with pigment packaging, the eigenanalyses show qualitative and quantitative similarities for all datasets and oligotrophic subsets (Fig. 1). The similarities favor primacy of $a_{\mathrm{CDOM}}(440)$—rather than C_a—as the predominant driver of optical variability for waters, including those of the open ocean, summarized briefly, as follows: a) the leading eigenvector from each dataset—capturing approximately 60% of the variance—indicates opposing anomalies for longer versus shorter wavelengths with minimal amplitude in the blue-green transition domain. The secondary eigenvector from each dataset—capturing approximately 32% of the variance—indicates internal VIS spectral dependencies characteristic of C_a absorption, including a maximum near the blue-green transition; b) the $\boldsymbol{S}_1$ term representing the stretching and compressing necessary to best represent the data using the leading eigenfunction is always more strongly associated with variability in $a_{\mathrm{CDOM}}(440)$ than C_a values. The $\boldsymbol{S}_2$ term more strongly associates with field observations of C_a than $a_{\mathrm{CDOM}}(440)$ for two of the three datasets, likely due to reduced applicability of the EMA algorithm for the less spectrally expansive dataset; and c) the signs of the partial correlation coefficients derived comparing the leading and secondary eigenfunctions to C_a and $a_{\mathrm{CDOM}}(440)$ variability are in all cases consistent considering the short wavelength anomaly in $\boldsymbol{\psi}_1$ and the blue-green anomaly in $\boldsymbol{\psi}_2$ compared to the spectral absorption profiles of CDOM and C_a. Observations of detrital absorption were not available, and detritus and CDOM share general similarities in the spectral shape of absorption, i.e., absorption increases for both at shorter wavelengths. Some key factors relevant to the focus herein on primacy of $a_{\mathrm{CDOM}}(440)$ rather than detritus include the leading eigenfunction capturing approximately (within 1% of) 60% of the variance for each dataset, whereas each dataset corresponds to dissimilar contributions from inland, coastal, and open-ocean waters. Absorption by CDOM and detritus also exhibit variable correlation patterns (Babin et al. 2003; Twardowski et al. 2004), which would degrade the correlation between the leading eigenfunction and $a_{\mathrm{CDOM}}(440)$ if the leading eigenfunctions were most strongly associated with detritus. The comparisons between $\boldsymbol{S}_1$ and $a_{\mathrm{CDOM}}(440)$ indicate loglinear relationships, which are shown—along with Spearman's rank correlation coefficient values—in the supplemental materials.

Although primacy of $a_{\mathrm{CDOM}}(440)$ is conceptually distinct from formulations in the published literature posing C_a as an independent variable primarily determining optical properties (e.g., Gordon 1988; Morel 1988), the results do not necessarily conflict with published observations in the bio-optical literature. For example, the inadequacy of multispectral VIS observations to partition C_a and $a_{\mathrm{CDOM}}(440)$ signals was noted at least six decades ago (Yentsch 1960), and the case-1 prescription does not explicitly require primacy of C_a (Morel and Prieur 1977). Rather, primacy of C_a is invoked by models representing the oceanic light field and its OACs as a function of C_a (Morel 1988), plus the implementation of OC algorithms wherein regional tunings refit R_{rs} ratios and C_a to overcome regional differences in relationships between OACs. The principal finding wherein

CDOM—rather than C_a—variability is most directly associated with optical data products emerging from the aquatic surface, e.g., $[L_W(\lambda)]_{\mathrm{N}}$ or R_{rs}, may reflect differences in temporal and vertical dynamics of particulate versus dissolved constituents, including via the complex pathways dissolved materials cycle in the microbial loop (Azam et al. 1983) that would contribute to decoupling CDOM dynamics from C_a.

For example, the connectivity, directionality, and rates of fluxes connecting phytoplankton, dissolved organics, nutrients, and other parameters determine the degree to which one parameter could feasibly predict another. The case-1 model, wherein C_a predominantly modifies the aquatic light environment and predicts CDOM and other OACs, is perhaps most logically consistent with historical understandings of the so-called grazing food chain, a simplified oceanic trophic cascade wherein bacteria were considered negligible in the cycling of organic materials (Steele 1974) and phytoplankton degradation produces CDOM (Kalle 1966). Increasingly complex microbial loop dynamics—wherein phytoplankton contribute dissolved organic matter (DOM) directly through cellular lysis, excretion, and grazing, and wherein bacteria regulate the remineralization of DOM that fuels phytoplankton growth—support an increasing diversity of pathways representing the flow of energy and materials through marine systems (Azam et al. 1983; Azam 1998). Additional discoveries, including new functional groups (e.g., viruses) and pathways (e.g., direct phytoplankton uptake of DOM), further complicate oceanic food webs (Granéli et al. 1999; Fenchell 2008) relative to those of the grazing food chain.

Stratification and mixing modify vertical time scales of sinking and have been related to global-scale patterns in surface concentrations of DOM, e.g., wherein stratified mid-latitude waters are characterized by higher concentrations of DOM and well-mixed waters of higher latitudes—with the exception of the arctic—contain relatively lower concentrations of DOM (Hansell and Orellana 2021). The optically active DOM subset, i.e., CDOM, is further influenced by the kinematic processes governing photobleaching and degradation rates, affected by mixed layer depth, seasonality, latitude, sea surface temperature, and sky condition (Gonsior et al. 2014; Del Vecchio and Blough 2002). Changing ocean temperatures and acidity also relate to temporal dynamics in DOM via the restructuring of microbial food webs through altering stratification (and therefore nutrient availability), microbial metabolic rates plus nutrient demands, remineralization of DOM, and perhaps, species distributions (Sala et al. 2016). Horizontal spatial comparisons are not assessed, but the results are, perhaps, in keeping with recent findings that patchiness in oceanic CDOM is not well explained by patchiness in C_a and exhibits a distribution more similar to sea surface temperature (Gray et al. 2025). Vertical differences may also be relevant although neither the depth dimension nor vertical flux are investigated here. Negative buoyancy of particles—e.g., sinking velocity increases proportionally with density differences and the squared particle radius (Bach et al. 2012)—or nutricline structure may drive particulate organic distributions away from the surface, i.e., away from where remote sensing signals are weighted (e.g., Morel and Berthon 1989). DOM—including a refractory pool that also persists with lower sensitivity to the dynamics of microbial food webs (Jiao et al. 2010)—is transported primarily through advection with water masses.

The eigenanalyses presented herein additionally suggest case-1 conditions are not representative of global water bodies nor selective oceanic subsets, because the partial correlation coefficients indicate the eigenfunctions associate—to a greater extent than anticipated—with individual OACs, and not with bulk combinations. For example, considering an idealized case-1 dataset wherein all OACs are accurately derived from C_a, the leading eigenfunction would be anticipated to produce similar

rank correlation coefficients for each OAC. In contrast to expectations that water mass OACs should be modeled from C_a, the eigenfunction analyses indicate that the majority of the variance (approximately 60%) in oceanic optical data products most strongly corresponds to $a_{\mathrm{CDOM}}(440)$, and an orthogonal mode capturing less variance (approximately 32%) may correspond to residual variability in C_a. Finally, although the results suggest a case-1 model with $a_{\mathrm{CDOM}}(440)$ as the fundamental variable rather than C_a might better represent the three datasets and their oligotrophic subsets, the underlying assumption that a single OAC captures oceanic light variability would limit the model to representing only 60% of the variance observed. The result suggests oceanic trophic cascades are characterized by greater complexity than can be represented by a univariate model.

Overall, the findings demonstrate high variance captured by the leading eigenfunction and the strong association of the leading eigenfunction with $a_{\mathrm{CDOM}}(440)$ variability, but the findings would be strengthened by adding field sampling for additional biochemical or bio-optical parameters. For example, RSE2021 demonstrates the importance of adding observations of $a_{\mathrm{CDOM}}(440)$ plus spectrally expansive radiometry to observations of C_a. Also adding observations of non-algal particles (NAPs) and detrital absorption would improve clarity on the eigenfunction interpretations and provide another option for investigating subsequent modes. The findings presented herein correspond to datasets spanning globally representative or oceanic waters and do not rank the importance of constituents in any specific region. For example, regional scenarios wherein high variability of NAPs is associated with riverine flux or resuspension would not be anticipated to reflect the findings herein supporting $a_{\mathrm{CDOM}}(440)$ primacy.

4.3 OC algorithm retrievals

High sensitivity of OC algorithms to regional differences in bio-optics is likewise consistent with the findings herein of $\boldsymbol{S}_1$ producing maximal correlation to $a_{\mathrm{CDOM}}(440)$ and the OC algorithm products. For example, regional tunings of OC algorithms have proven necessary even at spatial scales as large as currents (Kahru and Mitchell 1999), seas (Morel and Gentili 2009), and oceans (Johnson et al. 2013), revealing an ever diminishing fraction of oceanic surface environments wherein OC algorithms are robust to natural variability in the mixtures and dynamics of OACs.

Sensitivity of the OC algorithm to $a_{\mathrm{CDOM}}(440)$ is supported by the R^2 metrics, which indicate the OC algorithm captures greater variance in $a_{\mathrm{CDOM}}(440)$ compared to C_a, consistent with the results of the eigenanalyses. The assessment warrants caution, as follows: $a_{\mathrm{CDOM}}(440)$ is less commonly obtained in bio-optical field sampling than C_a, which is why two of the datasets herein do not include contemporaneous $a_{\mathrm{CDOM}}(440)$ observations; and the RSE2021 dataset, which does include contemporaneous observations of both $a_{\mathrm{CDOM}}(440)$ and C_a, also includes representation from coastal and inland waters. Regarding the former, the high incidence of optical measurements with contemporaneous field observations of C_a but not $a_{\mathrm{CDOM}}(440)$ poses a persistent difficulty in aquatic optics, and the results herein indicate the importance of routine $a_{\mathrm{CDOM}}(440)$ field observations in bio-optical field activities. Regarding the latter, the water mass representation of RSE2021 is addressed to the extent practicable by considering the scatter of the data points in the separate quadrants of the Fig. 2 validation comparisons. For example, a hypothesis wherein the OC algorithm would produce maximal association to C_a in open-ocean waters would correspond to reduced scatter of the data points in the lower left quadrant in the C_a comparison (panel a) relative to the $a_{\mathrm{CDOM}}(440)$

comparison (panel d). Conversely, the Fig. 2 results show *tighter* association of the OC algorithm to $a_{\mathrm{CDOM}}(440)$ relative to C_a in the lower left quadrant.

Comparing between datasets, the RSE2007 observations produce the most similar R^2 values when comparing OC algorithm products to C_a and $a_{\mathrm{CDOM}}(440)$, consistent with the RSE2007 dataset being maximally represented by open-ocean waters far from coastal inputs and RSE2007 conferring the least expansive spectral range, which challenges the applicability of the EMA algorithm for retrieving $a_{\mathrm{CDOM}}(440)$ independently of C_a. Regarding the former, nearly identical agreement in R^2 comparing the OC algorithm products to C_a and $a_{\mathrm{CDOM}}(440)$ suggests RSE2007 waters are most consistent with the case-1 scenario wherein OACs covary. Nonetheless, $a_{\mathrm{CDOM}}(440)$ always produces higher R^2 values than C_a, with differences being insignificant for RSE2007 but increasing for RSE2022 and RSE2021.

The results of the algorithmic comparisons (Fig. 2) are consistent with those of the eigenanalyses (Fig. 1), and both suggest the OC algorithm more strongly associates with variability in $a_{\mathrm{CDOM}}(440)$ rather than C_a. The OC algorithm's applicability, therefore, depends strongly upon the relationships between OACs. The finding regarding the applicability of the OC algorithm is not the primary focus herein, and is consistent with previous studies demonstrating vulnerabilities in band-ratio retrievals of C_a to changing bio-optical relationships (Dierssen 2010). The comparisons further suggest estimation of $a_{\mathrm{CDOM}}(440)$ is more readily obtained than estimation of C_a using remote sensing. For example, the OC algorithm produces $a_{\mathrm{CDOM}}(440)$ estimation with an R^2 of 0.85 for the RSE2021 dataset wherein optically complex waters are included, although applying the OC algorithm for $a_{\mathrm{CDOM}}(440)$ estimation is not recommended. Improved R^2 values greater than 0.90 have been shown for remote sensing of $a_{\mathrm{CDOM}}(440)$ using end-member analysis with VIS or INV wavebands (Houskeeper et al. 2021).

4.4 Spectral information content of aquatic spectra

Maximal association of the leading eigenfunction for each dataset with $a_{\mathrm{CDOM}}(440)$—rather than C_a—indicates that accurate retrieval of $a_{\mathrm{CDOM}}(440)$ is fundamental to the objectives of the ocean color community to quantify oceanic productivity. Information extraction for subsequent eigenfunctions—including the second eigenfunction observed herein for most datasets to correlate most strongly with C_a—demands both increasing radiometric capabilities (e.g., higher accuracy, more expansive spectral range, and perhaps finer spectral resolution) but also accurate retrieval of the information associated with the preceding eigenfunctions. Regarding the latter, small uncertainties or residual errors in retrievals of information corresponding to a leading eigenfunction produce so-called noise amplification in retrieval of information corresponding to subsequent eigenfunctions, just as the mean spread of $\boldsymbol{S}_i$ functions decreases with increasing i. Accurate retrieval of $a_{\mathrm{CDOM}}(440)$ is, therefore, prerequisite to—not separate from—accurate retrieval of phytoplankton properties, in keeping with the trajectory of research into simultaneous inherent optical property (IOP) inversions (e.g., Maritorena et al. 2002; Lee et al. 2002; Werdell et al. 2013), machine learning approaches (O'Shea et al. 2021), and degeneracy (Prochaska and Frouin 2025).

Fine-scale spectral dependencies have previously been shown to confer useful and exploitable information—e.g., fluorescence (Letelier and Abbott 1996), absorption band effects (Houskeeper et al. 2020b), and macroalgal physiological state (Bell and Siegel 2022)—which in some cases may not emerge with strong explanatory power in an eigenanalysis using the full spectral range, because $\boldsymbol{\Lambda}$ is a metric of variance captured across the full spectral range of the analysis. Eigenfunction analyses

such as those performed here—as well as in Sathyendranath et al. (1987), Cael et al. (2023), and others—are useful tools for exploring the major patterns of a dataset. However, caution is also warranted because information only impacting a subset of the wavebands in an eigenanalysis may average out or suffer spectral dilution (Houskeeper et al. 2020b).

The eigenanalyses recreated for the RSE2021 dataset but using VIS or INV spectral subsets (denoted VIS21 and INV21, respectively) also support previous findings demonstrating improvements in $a_{\mathrm{CDOM}}(440)$ algorithmic relationships with increasing spectral separation of the algorithm wavebands from the VIS domain (Hooker et al. 2020, 2021b; Houskeeper et al. 2020a, 2021). For example, the results indicate INV information is more readily represented by a single mode of variability than VIS information, as follows: a) the leading eigenfunction of VIS21 captures similar variance to that of the spectrally expansive UVN21; b) the leading eigenfunction of INV21 captures approximately 17% more variance than that of either UVN21 or VIS21; and c) the second eigenfunction of VIS21 captures approximately 15% more variance than that of INV21. Pearson's correlation coefficients relating $\boldsymbol{S}_1$ to $a_{\mathrm{CDOM}}(440)$ increase strongly for the INV21 subset, approaching that of a perfectly coherent relationship (-0.96). The results indicate INV21 captures information more associated with $a_{\mathrm{CDOM}}(440)$ and less associated with C_a. Conversely, the leading eigenfunction of the VIS21 subset captures information more similarly associated with both $a_{\mathrm{CDOM}}(440)$ and C_a. The findings confirm previous literature based on theoretical formulations and observations, wherein spectrally expansive observations mitigate the confounding effects of spectral overlap in C_a and CDOM absorption (Houskeeper et al. 2021).

The recent launch of the PACE satellite mission supports novel spectrally expansive and hyperspectral observations of surface ocean environs. The spectral subset eigenanalyses indicate opportunities to leverage spectrally expansive PACE observations to advance OAC retrievals. For example, accurate retrieval of INV reflectances support retrieval of $a_{\mathrm{CDOM}}(440)$ with improved independence from C_a, and knowledge of $a_{\mathrm{CDOM}}(440)$ could then inform C_a retrievals using VIS information. Alternately, the results indicate that simultaneous retrieval of OACs, e.g., using inversions (Garver and Siegel 1997) or MDNs (O'Shea et al. 2021), may improve if (quality assured) input reflectances are spectrally expansive and include INV domains. However, spectrally expansive *in situ* datasets are presently sparse and corresponding improvements in atmospheric correction are likewise required to support spectrally expansive satellite ocean observing.

5 Conclusions

Spectral AOP data products of aquatic environments produce leading eigenfunctions more strongly correlated to variability in CDOM absorption than C_a, consistent with advancing knowledge of microbial loop dynamics and an increasing diversity of trophic pathways represented therein (Azam 1998). The similarity of results using three independent datasets strengthens support for CDOM primacy and suggests that accurate and independent estimation of CDOM is requisite to the retrieval of C_a and other phytoplankton parameters. Acknowledged limitations include the eigenanalyses capturing linear—but not nonlinear—patterns of variability, although the relationships between the leading eigenfunction and log-transformed $a_{\mathrm{CDOM}}(440)$ indicated linearity. The eigenanalyses results are insensitive to moderate changes in the quality control filtering, but not to complete removal of quality control filtering for the legacy data, which contain significant proportions of likely nonphysical data

products. Confounding signals identified for CDOM and C_a are consistent with early investigations into the drivers of color variability (Yentsch 1960), as well as subsequent work assessing vulnerabilities in band-ratio algorithms for characterizing C_a (Dierssen 2010; Sauer et al. 2012). Spectrally expansive data products have been shown to improve retrieval of $a_{\mathrm{CDOM}}(440)$ independent of C_a (Sathyendranath et al. 1987; Hooker et al. 2020; Houskeeper et al. 2021), attributed, in part, to improved robustness to variability in the relationships between OACs. The comparisons herein of eigenanalyses using the INV21 and VIS21 spectral subsets further support the hypothesis that an expansive spectral range for data products improves the separability of signals. Spectrally expansive data products have been demonstrated *in situ* spanning the UV to shortwave infrared (SWIR) wavelength domain, with the latter formerly ascribed as null (Houskeeper and Hooker 2023). The findings herein support opportunities associated with leveraging the spectrally expansive capabilities of PACE OCI, although further advances in image processing, atmospheric correction, algorithm design, and *in situ* datasets would be necessary.

Acknowledgements. The authors are grateful for contributions from the following researchers (in alphabetical order): Tom Bell (Woods Hole Oceanographic Institution), Carlos del Castillo (NASA's Goddard Space Flight Center), Liane Guild (NASA's Ames Research Center), Raphael Kudela (University of California, Santa Cruz), Randall Lind (Biospherical Instruments Inc.), and Kendra Negrey (University of California, Santa Cruz). Funding for this work was provided as follows: Woods Hole Oceanographic Institution Innovation Month (H.F.H.); the NASA Biodiversity Survey of the Cape (BioSCape) activity (H.F.H.); and the NASA ACE satellite mission office (SBH).

Author contributions. H.F.H. conceptualization; H.F.H. and S.B.H. methodology; H.F.H. formal analysis; H.F.H. and S.B.H. resources; H.F.H. writing–original draft; H.F.H. and S.B.H. writing–review and editing; H.F.H. and S.B.H. finding acquisition.

Competing interests. The authors declare no competing interests.

Code and data availability. Data to replicate figures is available in Dryad (Houskeeper and Hooker 2025 [Dataset]). All other data may be made available upon request.

References

Aiken, J., G.F. Moore, C.C. Trees, S.B. Hooker, and D.K. Clark. “The SeaWiFS CZCS-type pigment algorithm." *Oceanographic Literature Review* 3, no. 43 (1996): 315-316.

Aksnesa, D.L., and M.D. Ohman. “Multi-decadal shoaling of the euphotic zone in the southern sector of the California Current System." *Limnology and Oceanography* 54, no. 4 (2009): 1272-1281.

Azam, F., T. Fenchel, J.G. Field, J.S. Gray, L.A. Meyer-Reil, and F. Thingstad. “Microbial utilization of organic matter in the world’s oceans.” *Marine Ecology Progress Series* 10 (1983): 101-114.

Azam, F. “Microbial control of oceanic carbon flux: the plot thickens." *Science* 280, no. 5364 (1998): 694-696.

Babin, M., D. Stramski, G.M. Ferrari, H. Claustre, A. Bricaud, G. Obolensky, and N. Hoepffner. “Variations in the light absorption coefficients of phytoplankton, nonalgal particles, and dissolved organic matter in coastal waters around Europe." *Journal of Geophysical Research: Oceans* 108, no. C7 (2003).

Bach, L.T., U. Riebesell, S. Sett, S. Febiri, P. Rzepka, and K.G. Schulz. “An approach for particle sinking velocity measurements in the 3–400 μm size range and considerations on the effect of temperature on sinking rates." *Marine Biology* 159, no. 8 (2012): 1853-1864.

Bell, T.W., and D.A. Siegel. “Nutrient availability and senescence spatially structure the dynamics of a foundation species.” *Proceedings of the National Academy of Sciences* 119, no. 1 (2022): e2105135118.

Bricaud, Annick, Hervé Claustre, Josephine Ras, and Kadija Oubelkheir. “Natural variability of phytoplanktonic absorption in oceanic waters: Influence of the size structure of algal populations." *Journal of Geophysical Research: Oceans* 109, no. C11 (2004).

Bidigare, R.R., M.E. Ondrusek, J.H. Morrow, and D. Kiefer. “In vivo absorption properties of algal pigments.” *Ocean Optics X Proc. SPIE*, 1302, (1990): 290-302.

Cael, B.B., A. Chase, and E. Boss. “Information content of absorption spectra and implications for ocean color inversion." *Applied Optics* 59, no. 13 (2020): 3971-3984.

Cael, B.B., K. Bisson, E. Boss, and Z.K. Erickson. “How many independent quantities can be extracted from ocean color?" *Limnology and Oceanography Letters* 8, no. 4 (2023): 603-610.

Cetinić, I., et al. “Phytoplankton composition from sPACE: Requirements, opportunities, and challenges" *Remote Sensing of Environment* 302, no. 1 (2023): 113964.

Chang, G.C., T.D. Dickey, C.D. Mobley, E. Boss, and W.S. Pegau, “Toward closure of upwelling radiance in coastal waters.” *Applied Optics* 42, 1574-1582 (2003).

Chase, A. P., E. Boss, I. Cetinić, and W. Slade. “Estimation of phytoplankton accessory pigments from hyperspectral reflectance spectra: toward a global algorithm." *Journal of Geophysical Research: Oceans* 122, no. 12 (2017): 9725-9743.

Clarke, G.L., Gifford C. Ewing, and C.J. Lorenzen. “Spectra of backscattered light from the sea obtained from aircraft as a measure of chlorophyll concentration." *Science* 167, no. 3921 (1970): 1119-1121.

Clark, D.K. “Phytoplankton pigment algorithms for the Nimbus-7 CZCS." In *Oceanography from Space*, pp. 227-237. Boston, MA: Springer US, 1981.

Claustre, H., A. Morel, S.B. Hooker, M. Babin, D. Antoine, K. Oubelkheir, A. Bricaud, K. Leblanc, B. Quéguiner, and S. Maritorena. “Is desert dust making oligotrophic waters greener?” *Geophys. Rev. Lett.,* **29,** 107 (2002): 1–4.

Claustre, H., S.B. Hooker, L. Van Heukelem, J-F. Berthon, R. Barlow, J. Ras, H. Sessions, C. Targa, C. Thomas, D. van der Linde, and J-C. Marty. “An intercomparison of HPLC phytoplankton pigment methods using *in situ* samples: Application to remote sensing and database activities.” *Mar. Chem.,* **85,** (2004): 41–61.

Del Vecchio, R., and N.V. Blough. “Photobleaching of chromophoric dissolved organic matter in natural waters: kinetics and modeling." *Marine Chemistry* 78, no. 4 (2002): 231-253.

Dierssen, H.M. and R.C. Smith. “Bio-optical properties and remote sensing ocean color algorithms for Antarctic Peninsula waters.” *Journal of Geophysical Research: Oceans* 105, no. C11 (2000): 26301-26312.

Dierssen, H.M. “Perspectives on empirical approaches for ocean color remote sensing of chlorophyll in a changing climate." *Proceedings of the National Academy of Sciences* 107, no. 40 (2010): 17073-17078.

Doron, M., S. Bélanger, D. Doxoran, and M. Babin. “Spectral variations in the near-infrared reflectance." *Remote Sensing of Environment* 115, no. 7 (2011): 1617-1631.

Dutkiewicz, S., A.E. Hickman, O. Jahn, S. Henson, C. Beaulieu, and E. Monier. “Ocean colour signature of climate change." *Nature communications* 10, no. 1 (2019): 578.

Erickson, Z.K., P.J. Werdell, and I. Cetinić. “Bayesian retrieval of optically relevant properties from hyperspectral water-leaving reflectances." *Applied Optics* 59, no. 23 (2020): 6902-6917.

Erickson, Z.K., L. McKinna, P.J. Werdell, and I. Cetinić. “Bayesian approach to a generalized inherent optical property model." *Optics Express* 31, no. 14 (2023): 22790-22801.

Fenchel, T. “The microbial loop–25 years later." *Journal of Experimental Marine Biology and Ecology* 366, no. 1-2 (2008): 99-103.

Fisher, Ronald Aylmer. “The distribution of the partial correlation coefficient." *Metron* 3 (1924): 329-332.

Franz, B.A., S.W. Bailey, G. Meister, and P.J. Werdell. “Quality and consistency of the NASA ocean color data record." *Proc. Ocean Optics XXI* (2012).

Garver, S.A., and D.A. Siegel. “Inherent optical property inversion of ocean color spectra and its biogeochemical interpretation: 1. Time series from the Sargasso Sea." *Journal of Geophysical Research: Oceans* 102, no. C8 (1997): 18607-18625.

Gordon, H.R., D.K. Clark, J.L. Mueller, and W.A. Hovis. “Phytoplankton pigments from the Nimbus-7 Coastal Zone Color Scanner: comparisons with surface measurements." *Science* 210, no. 4465 (1980): 63-66.

Gordon, H.R., and A. Morel. “Remote Assessment of Ocean Color for Interpretation of Satellite Visible Imagery.” In *Lecture Notes on Coastal and Estuarine Studies*, 113 pp., Springer-Verlag, New York, 1983.

Gordon, H.R., D.K. Clark, J.W. Brown, O.B. Brown, R.H. Evans, and W.W. Broenkow. “Phytoplankton pigment concentrations in the Middle Atlantic Bight: comparison of ship determinations and CZCS estimates." *Applied Optics* 22, no. 1 (1983): 20-36.

Gordon, H.R. “Evolution of ocean color atmospheric correction: 1970–2005." *Remote Sensing* 13, no. 24 (2021): 5051.

Gonsior, M., N. Hertkorn, M.H. Conte, W.J. Cooper, D. Bastviken, E. Druffel, and P. Schmitt-Kopplin. “Photochemical production of polyols arising from significant photo-transformation of dissolved organic matter in the oligotrophic surface ocean.” *Marine Chemistry* 163 (2014): 10-18.

Granéli, E., P. Carlsson, and C. Legrand. “The role of C, N and P in dissolved and particulate organic matter as a nutrient source for phytoplankton growth, including toxic species.” *Aquatic Ecology* 33 (1999): 17-27.

Gray, P.C., E. Boss, G. Bourdin, and Y. Lehahn. “Emergent patterns of patchiness differ between physical and planktonic properties in the ocean." *Nature Communications* 16, no. 1 (2025): 1808.

Grew, G.W., and L.S. Mayo. “Ocean color algorithm for remote sensing of chlorophyll." No. NAS 1.60 (1983): 2164.

Haentjens, N. “Optical Signatures of Plankton in the Open Ocean: From Individual Cells to Global Patterns.” The University of Maine, 2020.

Hansell, D.A., and M.V. Orellana. “Dissolved Organic Matter in the Global Ocean: A Primer.” *Gels* 7, no. 3 (2021): 128.

Hirata, T., T. Hirawake, F. Sakaida, H. Yamaguchi, K. Suzuki, H. Murakami, J. Ishizaka, H. Kobayashi, A. Fujukara, M. Toratani, and S. Saitoh. “Development and Verification of SGLI/GCOM-C1 Ocean Algorithms.” *Journal of The Remote Sensing Society of Japan* 34, no. 4 (2014): 278-285.

Hooker, S.B., H. Claustre, J. Ras, L. Van Heukelem, J-F. Berthon, C. Targa, D. van der Linde, R. Barlow, and H. Sessions, 2000: The First SeaWiFS HPLC Analysis Round-Robin Experiment (SeaHARRE-1). *NASA Tech. Memo. 2000–206892, Vol. 14,* S.B. Hooker and E.R. Firestone, Eds., NASA Goddard Space Flight Center, Greenbelt, Maryland, 42 pp.

Hooker, S.B., and S. Maritorena. “An evaluation of oceanographic radiometers and deployment methodologies.” *J. Atmos. Oceanic Technol.,* **17,** (2000):811–830.

Hooker, S.B., G. Lazin, G. Zibordi, and S. McLean. “An evaluation of above- and in-water methods for determining water-leaving radiances. *J. Atmos. Oceanic Technol.,* **19,** (2002): 486–515.

Hooker, S.B., L. Clementson, C.S. Thomas, L. Schlüter, M. Allerup, J. Ras, H. Claustre, C. Normandeau, J. Cullen, M. Kienast, W. Kozlowski, M. Vernet, S. Chakraborty, S. Lohrenz, M. Tuel, D. Redalje, P. Cartaxana, C.R. Mendes, V. Brotas, S.G.P. Matondkar, S.G. Parab, A. Neeley, and E.S. Egeland, 2012: The Fifth SeaWiFS HPLC Analysis Round-Robin Experiment (SeaHARRE-5). *NASA Tech. Memo. 2012–217503,* NASA Goddard Space Flight Center, Greenbelt, Maryland, 98 pp.

Hooker, S.B., G. Bernhard, J.H. Morrow, C.R. Booth, T. Comer, R.N. Lind, and V. Quang. 2012: “Optical sensors for planetary radiant energy (OSPREy): calibration and validation of current and next-generation NASA missions.” TM-2012-215872, NASA Goddard Space Flight Center, Greenbelt, Maryland, 117 pp.

Hooker, S.B., J.H. Morrow, and A. Matsuoka. “Apparent optical properties of the Canadian Beaufort Sea–Part 2: The 1% and 1 cm perspective in deriving and validating AOP data products." *Biogeosciences* 10, no. 7 (2013): 4511-4527.

Hooker, S.B., R.N. Lind, J.H. Morrow, J.W. Brown, K. Suzuki, H.F. Houskeeper, T. Hirawake, and E.R. Maure, 2018a: Advances in Above- and In-Water Radiometry, Vol. 1: Enhanced Legacy and State-of-the-Art Instrument Suites. TP–2018–219033/Vol. 1, NASA Goddard Space Flight Center, Greenbelt, Maryland, 60 pp.

Hooker, S.B., R.N. Lind, J.H. Morrow, J.W. Brown, R.M. Kudela, H.F. Houskeeper, K. Suzuki, 2018b: Advances in Above- and In-Water Radiometry, Vol. 2: Autonomous Atmospheric and Oceanic Observing Systems. TP–2018–219033/Vol. 2, NASA Goddard Space Flight Center, Greenbelt, Maryland, 69 pp.

Hooker, S.B., R.N. Lind, J.H. Morrow, J.W. Brown, R.M. Kudela, and H.F. Houskeeper, K. Suzuki, 2018c: Advances in Above- and In-Water Radiometry, Vol. 3: Hybridspectral Next-Generation Optical Instruments. TP–2018–219033/Vol. 3, NASA Goddard Space Flight Center, Greenbelt, Maryland, 39 pp.

Hooker, S.B., A. Matsuoka, R.M. Kudela, Y. Yamashita, K. Suzuki, and H.F. Houskeeper. “A global end-member approach to derive $a_{\mathrm{CDOM}}(440)$ from near-surface optical measurements." *Biogeosciences* 17, no. 2 (2020): 475-497.

Hooker, S.B., H.F. Houskeeper, R.M. Kudela, A. Matsuoka, K. Suzuki, and T. Isada. “Spectral modes of radiometric measurements in optically complex waters." *Continental Shelf Research* 219 (2021a): 104357.

Hooker, S.B., H.F. Houskeeper, R.N. Lind, and K. Suzuki. “One-and two-band sensors and algorithms to derive $a_{\mathrm{CDOM}}(440)$ from global above- and in-water optical observations." *Sensors* 21, no. 16 (2021b): 5384.

Hooker, S.B., H.F. Houskeeper, R.N. Lind, R.M. Kudela, and K. Suzuki. “Verification and validation of hybridspectral radiometry obtained from an unmanned surface vessel (USV) in the open and coastal oceans.” *Remote Sensing* 14, no. 5 (2022): 1084.

Houskeeper, Henry F. “Advances in bio-optics for observing aquatic ecosystems." University of California, Santa Cruz, 2020a.

Houskeeper, H.F., D. Draper, R.M. Kudela, and E. Boss. “Chlorophyll absorption and phytoplankton size information inferred from hyperspectral particulate beam attenuation." *Applied Optics* 59, no. 22 (2020b): 6765-6773.

Houskeeper, H.F., S.B. Hooker, and R.M. Kudela. “Spectral range within global $a_{\mathrm{CDOM}}(440)$ algorithms for oceanic, coastal, and inland waters with application to airborne measurements." *Remote Sensing of Environment* 253 (2021): 112155.

Houskeeper, Henry F., Stanford B. Hooker, and Kyle C. Cavanaugh. “Spectrally simplified approach for leveraging legacy geostationary oceanic observations.” *Applied Optics* 61, no. 27 (2022): 7966-7977.

Houskeeper, H.F., and S.B. Hooker. “Extending aquatic spectral information with the first radiometric IR-B field observations." *PNAS Nexus* 2, no. 11 (2023): pgad340.

Houskeeper, H.F., S.B. Hooker, and R.N. Lind. “Expanded linear responsivity for Earth and planetary radiometry." *J. Atm. Ocea. Tech.* 41, no. 11 (2024): 1,093–1,105.

Houskeeper, H.F. and S.B. Hooker [Dataset]. “The primacy of dissolved organic matter to aquatic light variability [Dataset].” *Dryad* (2025): doi:10.5061/dryad.8pk0p2p14.

Hovis, W.A. “The Nimbus-7 coastal zone color scanner (CZCS) program." In *Oceanography from Space*, pp. 213-225. Boston, MA: Springer US, 1981.

Hu, C., Z. Lee, and B. Franz. “Chlorophyll a algorithms for oligotrophic oceans: A novel approach based on three-band reflectance difference." *J. Geophys. Res.*, 117 no. C1 (2012): 10.1029/2011JC007395.

Isada, T., S.B. Hooker, Y. Taniuchi, and K. Suzuki. “Evaluation of retrieving chlorophyll a concentration and colored dissolved organic matter absorption from satellite ocean color remote sensing in the coastal waters of Hokkaido, Japan.” *Journal of Oceanography* 78, no. 4 (2022): 263-276.

Jerlov, N.G. “Optical studies of ocean water." *Rept. Swedish Deep-Sea Exped.*, no. 3 (1951): 1-59.

Jerlov, N.G. “Optical Oceanography,” *Elsevier Oceanography*, Series 5 (1968).

Jiao, N., et al. “Microbial production of recalcitrant dissolved organic matter: Long-term carbon storage in the global ocean.” *Nature Reviews Microbiology*, 8 no. 8 (2010): 593-599.

Johnson, R., P.G. Strutton, S.W. Wright, A. McMinn, and K.M. Meiners. “Three improved satellite chlorophyll algorithms for the Southern Ocean.” *Journal of Geophysical Research: Oceans* 118, no. 7 (2013): 3,694–3,703.

Kahru, M., and B.G. Mitchell. “Empirical chlorophyll algorithm and preliminary SeaWiFS validation for the California Current.” *International Journal of Remote Sensing* 20, no. 17 (1999): 3,423–3,429.

Kahru, M., R.M. Kudela, M. Manzano-Sarabia, and B.G. Mitchell. “Trends in the surface chlorophyll of the California Current: Merging data from multiple ocean color satellites." *Deep Sea Research Part II: Topical Studies in Oceanography* 77 (2012): 89-98.

Kalle, K. “The problem of the Gelbstoff in the sea.” *Marine Biology* 4 (1966): 91-104.

Kirk, John TO. Light and photosynthesis in aquatic ecosystems. Cambridge university press, 2011.

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

Kramer, S.J., D.A. Siegel, S. Maritorena, and D. Catlett. “Modeling surface ocean phytoplankton pigments from hyperspectral remote sensing reflectance on global scales." *Remote Sensing of Environment* 270 (2022): 112879.

Kudela, R.M., S.L. Palacios, D.C. Austerberry, E.K. Accorsi, L.S. Guild, and J. Torres-Perez. “Application of hyperspectral remote sensing to cyanobacterial blooms in inland waters." *Remote Sensing of Environment* 167 (2015): 196-205.

Kudela, R.M., S.B. Hooker, H.F. Houskeeper, and M. McPherson. “The influence of signal to noise ratio of legacy airborne and satellite sensors for simulating next-generation coastal and inland water products." *Remote Sensing* 11, no. 18 (2019): 2071.

Kudela, R.M., S.B. Hooker, L.S. Guild, H.F. Houskeeper, and N. Taylor. “Expanded signal to noise ratio estimate for validating next-generation satellite sensors in oceanic, coastal, and inland waters." *Remote Sensing* 16 (2024): 1238.

Lee, Z.P., K.L. Carder, and R.A. Arnone. “Deriving inherent optical properties from water color: a multiband quasi-analytical algorithm for optically deep waters." *Applied optics* 41, no. 27 (2002): 5755-5772.

Lee, Z.P., and K.L. Carder. “Absorption spectrum of phytoplankton pigments derived from hyperspectral remote-sensing reflectance." *Remote Sensing of Environment* 89, no. 3 (2004): 361-368.

Letelier, R.M., and M.R. Abbott. “An analysis of chlorophyll fluorescence algorithms for the Moderate Resolution Imaging Spectrometer (MODIS)." *Remote Sensing of Environment* 58, no. 2 (1996): 215-223.

Lewis, K.M., and K.R. Arrigo. “Ocean color algorithms for estimating chlorophyll a, CDOM absorption, and particle backscattering in the Arctic Ocean." *Journal of Geophysical Research: Oceans* 125, no. 6 (2020): e2019JC015706.

Maritorena, S., D.A. Siegel, and A.R. Peterson. “Optimization of a semianalytical ocean color model for global-scale applications." *Applied Optics* 41, no. 15 (2002): 2705-2714.

Matsuoka, A., M. Babin, D. Doxaran, S.B. Hooker, B.G. Mitchell, S. Bélanger, and A. Bricaud. “A synthesis of light absorption properties of the Pan-Arctic Ocean: application to semi-analytical estimates of dissolved organic carbon concentrations from space." *Biogeosciences Discussions* 10, no. 11 (2013).

McClain, C., S. Hooker, G. Feldman, and P. Bontempi, 2006: Satellite data for ocean biology, biogeochemistry, and climate research. *Eos, Trans. Amer. Geophys. Union*, **87,** 337–343.

Morel, A. “In-water and remote measurements of ocean color." *Boundary-Layer Meteorology* 18, no. 2 (1980): 177-201.

Morel, A. “Optical modeling of the upper ocean in relation to its biogenous matter content (case I waters).” *Journal of Geophysical Research*, 93, no. C9. (1988): 10,749–10,768.

Morel, A. “Are the empirical relationships describing the bio-optical properties of case 1 waters consistent and internally compatible?” *Journal of Geophysical Research* 114, no. C01016 (2009).

Morel, A., and J.-F. Berthon. “Surface pigments, algal biomass profiles, and potential production of the euphotic layer: Relationships reinvestigated in view of remote-sensing applications." *Limnology and Oceanography* 34, no. 8 (1989): 1545-1562.

Morel, A., and B. Gentili. “The dissolved yellow substance and the shades of blue in the Mediterranean Sea." *Biogeosciences* 6, no. 11 (2009): 2625-2636.

Morel, A., and L. Prieur. “Analysis of variations in ocean color 1." *Limnology and Oceanography* 22, no. 4 (1977): 709-722.

Morrow, J.H., S.B. Hooker, C.R. Booth, G. Bernhard, R.N. Lind, and J.W. Brown, 2010: Advances in Measuring the Apparent Optical Properties (AOPs) of Optically Complex Waters. *NASA Tech. Memo. 2010–215856,* NASA Goddard Space Flight Center, Greenbelt, Maryland, 80 pp.

O’Reilly, J.E., S. Maritorena, B.G. Mitchell, D.A. Siegel, K.L. Carder, S.A. Garver, M.Kahru, and C. McClain. “Ocean color chlorophyll algorithms for SeaWiFS." *Journal of Geophysical Research: Oceans* 103, no. C11 (1998): 24937-24953.

O’Reilly, J.E., and P.J. Werdell. “Chlorophyll algorithms for ocean color sensors-OC4, OC5 & OC6." *Remote Sensing of Environment* 229 (2019): 32-47.

O’Shea, R.E., N. Pahlevan, B. Smith, M. Bresciani, T. Egerton, C. Giardino, L. Li, et al. “Advancing cyanobacteria biomass estimation from hyperspectral observations: Demonstrations with HICO and PRISMA imagery." *Remote Sensing of Environment* 266 (2021): 112693.

Prochaska, J.X., and R.J. Frouin. “On the Challenges of Retrieving Phytoplankton Properties from Remote-Sensing Observations.” *EGUsphere* 2025 (2025): 1-39.

Ryther, John H., and Charles S. Yentsch. “The estimation of phytoplankton production in the ocean from chlorophyll and light data 1.” *Limnology and Oceanography* 2, no. 3 (1957): 281-286.

Sala, M.M., F.L. Aparicio, V. Balagué, J.A. Boras, E. Borrull, C. Cardelús, L. Cros et al. “Contrasting effects of ocean acidification on the microbial food web under different trophic conditions.” *ICES Journal of Marine Science* 73, no. 3 (2016): 670-679.

Sarmiento, J. L., R. Slater, R. Barber, L. Bopp, S.C. Doney, A.C. Hirst, J. Kleypas, R. Matear, U. Mikolajewicz, P. Monfray, and V. Soldatov. "Response of ocean ecosystems to climate warming." *Global Biogeochemical Cycles* 18, no. 3 (2004).

Sathyendranath, S., L. Prieur, and A. Morel. "An evaluation of the problems of chlorophyll retrieval from ocean colour, for case 2 waters." *Advances in Space Research* 7, no. 2 (1987): 27-30.

Sauer, M.J., C.S. Roesler, P.J. Werdell, and A. Barnard. "Under the hood of satellite empirical chlorophyll a algorithms: revealing the dependencies of maximum band ratio algorithms on inherent optical properties." *Optics Express* 20, no. 19 (2012): 20920-20933.

Sea-Bird Scientific. "Hyperspectral Ocean Color Radiometer User Manual." HyperOCR 2024-07-10 Version C. Accessed: 2 August 2024.

Siegel, D.A., S. Maritorena, N.B. Nelson, and M.J. Behrenfeld. "Independence and interdependencies among global ocean color properties: Reassessing the bio-optical assumption." *Journal of Geophysical Research: Oceans* 110, no. C7 (2005).

Siegel, D.A., M.J. Behrenfeld, S. Maritorena, C.R. McClain, D. Antoine, S.W. Bailey, P.S. Bontempi et al. "Regional to global assessments of phytoplankton dynamics from the SeaWiFS mission." *Remote Sensing of Environment* 135 (2013): 77-91.

Smith, R.C., K.S. Baker, and P. Dustan. "Fluorometric techniques for the measurement of oceanic chlorophyll in the support of remote sensing." *Scripps Institution of Oceanography* Ref. 81-17 (1981).

Steele, J.H. The structure of marine ecosystems. *Harvard University Press*, 1974.

Suzuki, K., A. Kamimura, and S.B. Hooker, 2015: Rapid and highly sensitive analysis of chlorophylls and carotenoids from marine phytoplankton using ultra-high performance liquid chromatography (UHPLC) with the first derivative spectrum chromatogram (FDSC) technique. *Mar. Chem.*, **176,** 96–109, 10.1016/j.marchem.2015.07.010.

Taylor, B.B., E. Torrecilla, A. Bernhardt, M.H. Taylor, I. Peeken, R. Röttgers, J. Piera, and A. Bracher. "Bio-optical provinces in the eastern Atlantic Ocean and their biogeographical relevance." *Biogeosciences* 8, no. 12 (2011): 3609-3629.

Twardowski, M.S., E. Boss, J.M. Sullivan, and P.L. Donaghay. "Modeling the spectral shape of absorption by chromophoric dissolved organic matter." *Marine Chemistry* 89, no. 1-4 (2004): 69-88.

Tyler, J.E. "*In situ* detection and estimation of chlorophyll and other pigments in the ocean." *Proceedings of the National Academy of Sciences* 47, no. 11 (1961): 1726-1733.

Uitz, J., H. Claustre, A. Morel, and S.B. Hooker. "Vertical distribution of phytoplankton communities in open ocean: An assessment based on surface chlorophyll." *Journal of Geophysical Research: Oceans* 111, no. C8 (2006).

Van Heukelem, L., and S.B. Hooker, 2011: "The Importance of a Quality Assurance Plan for Method Validation and Minimizing Uncertainties in the HPLC Analysis of Phytoplankton Pigments." In: *Phytoplankton Pigments Characterization, Chemotaxonomy and Applications in Oceanography,* S. Roy et al. Eds., Cambridge University Press, Cambridge, 195–242.

Werdell, P.J., and S.W. Bailey. “The SeaWiFS Bio-optical Archive and Storage System (SeaBASS): Current architecture and implementation.” NASA Technical Memo: 2002-211617, G.S. Fargion and C.R. McClain, Eds., NASA Goddard Space Flight Center, Greenbelt, Maryland, 45 pp. (2002).

Werdell, P.J., B.A. Franz, S.W. Bailey, G.C. Feldman, E. Boss, V.E. Brando, M. Dowell et al. “Generalized ocean color inversion model for retrieving marine inherent optical properties." *Applied optics* 52, no. 10 (2013): 2019-2037.

Yentsch, C.S. “The influence of phytoplankton pigments on the colour of sea water." *Deep Sea Research* 7, no. 1 (1960): 1-9.

Zaneveld, J.R.V. and J.C. Kitchen, “The variation in the inherent optical properties of phytoplankton near an absorption peak as determined by various models of cell structure,” *J. Geophys. Res. Oceans* 100, 13309–13320 (1995).

Zaneveld, J.R.V., E. Boss, and A. Barnard. “Influence of surface waves on measured and modeled irradiance profiles.” *Applied Optics* 40, no. 9 (2001): 1442-1449.

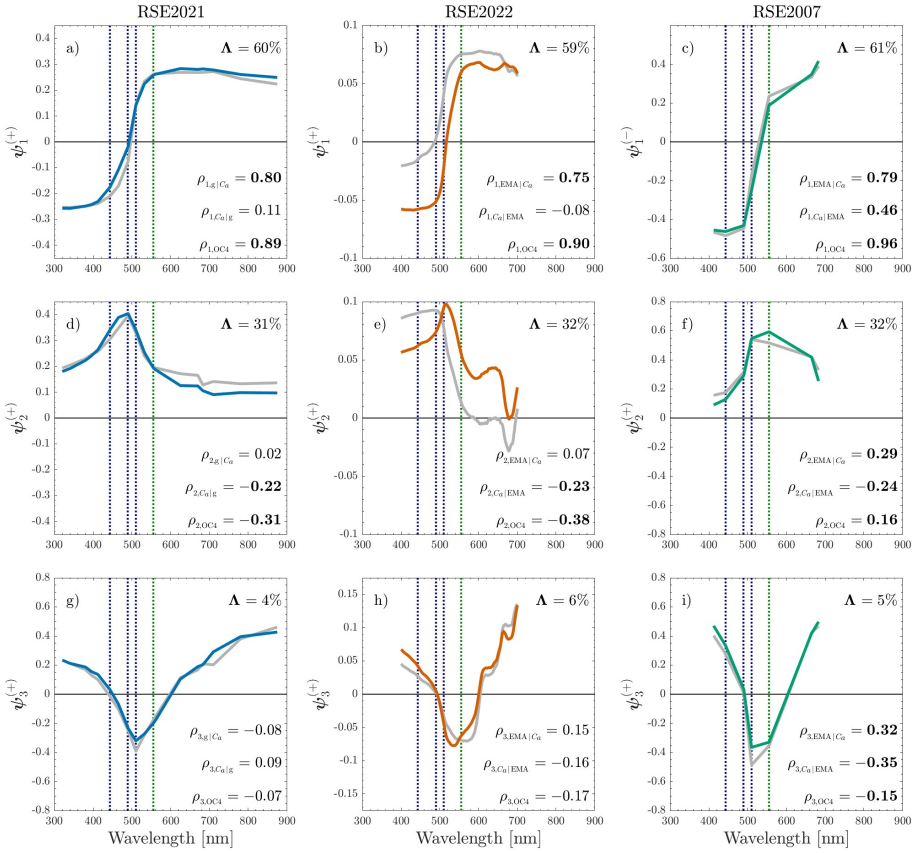

Figure 1. Spectral representations of the eigenfunction matrices corresponding to the RSE2021, RSE2022, and RSE2007 datasets are represented as blue, orange, and green lines, respectively. The first three eigenfunctions $\boldsymbol{\psi}_1$–$\boldsymbol{\psi}_3$ are organized vertically based on descending variance explained, $\boldsymbol{\Lambda}$. The y-axis label superscripts indicate the phase, i.e., either positive $(+)$ or negative $(-)$, in which the eigenvectors are presented, and y-axis label subscripts indicate the ordering of the eigenvector. Correlation coefficients relating $\boldsymbol{S}$ for each eigenfunction to recorded environmental (g and C_a) or algorithm (EMA and OC4) values are indicated using ρ notation, wherein the left subscript indicates the parameter under comparison with $\boldsymbol{S}$, given the parameter indicated by the right subscript, when applicable. Eigenfunctions for oligotrophic subsets (C_a less than $0.5\,\mathrm{mg\,m^{-3}}$) of the datasets are plotted on the same axes using gray lines. For clarity, all statistics presented correspond to the complete datasets (colored lines), and the sign for ρ corresponds to the phase in which $\boldsymbol{\psi}$ is plotted. The nominal locations of the blue and green wavebands used by the OC4 algorithm are indicated by blue and green vertical dashed lines, respectively.

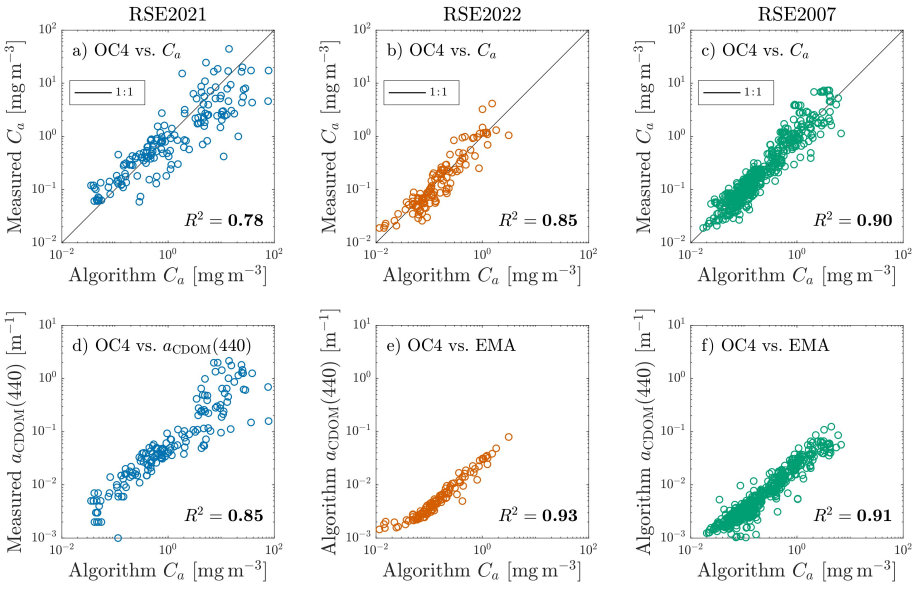

Figure 2. Algorithmic relationships for the RSE2021, RSE2022, and RSE2007 datasets are represented as blue, orange, and green open circles, respectively. The top panels indicate relationships between OC algorithms and field C_a, and a 1:1 correspondence line is shown in black. The bottom panels indicate relationships between OC algorithms and field (panel d) or algorithmic (panels e and f) observations of $a_{\mathrm{CDOM}}(440)$, with R^2 values shown. Algorithmic products (using an EMA relationship following Houskeeper et al. 2021) were required for the RSE2022 and RSE2007 datasets because only RSE2021 included routine field sampling of $a_{\mathrm{CDOM}}(440)$.

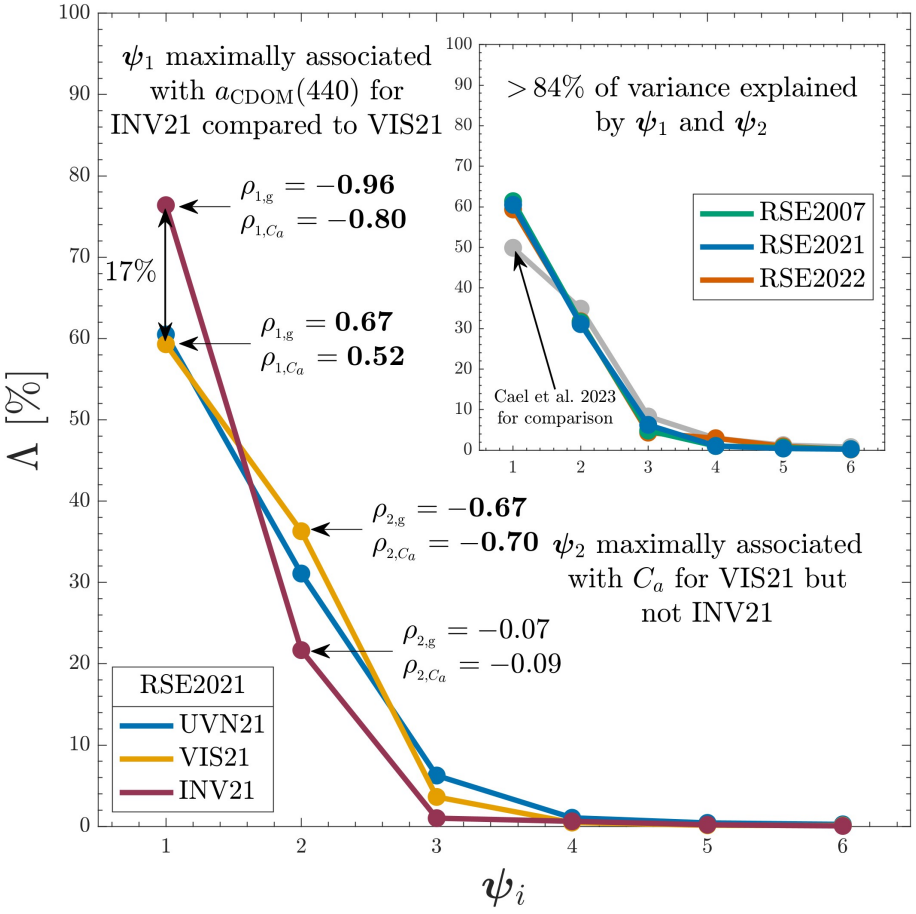

Figure 3. Variance explained by the first six eigenfunctions (ψ_1–ψ_6) for eigenanalyses using spectral subsets of the RSE2021 dataset, as follows: the full spectral range (320–875 nm) is shown in blue matching the presentations of Fig. 1 and 2 and denoted UVN21; the visible (412–683 nm) spectral subset results are shown in yellow and denoted VIS21; and the invisible (313–395 and 710–875 nm) spectral subset results are shown in red and denoted INV21. Pearson's correlation coefficients for $a_{\mathrm{CDOM}}(440)$ and C_a ($\rho_{1,g}$ and ρ_{1,C_a}, respectively) are denoted using arrows for the VIS21 and INV21 analyses, with bold font indicating significance ($P < 0.05$). The inlay panel shows that RSE2007, RSE2021, and RSE2022 datasets have similar maximal variance explained in ψ_1 and ψ_2.