

# Hyperspectral Remote Sensing for Harmful Algal Bloom Detection: *Pseudo-nitzschia* in the Northeast Pacific

Alexander L. Bailess<sup>1</sup>, Nicholas Baetge<sup>1</sup>, Andrew H. Barnard<sup>1</sup>, Nicholas Tuffillaro<sup>2</sup>, Michael Behrenfeld<sup>1</sup>, Brian Bill<sup>3</sup>, Raphael Kudela<sup>4</sup>, Jason Graff<sup>1</sup>, and Maria Kavanaugh<sup>1</sup>

<sup>1</sup>Oregon State University

<sup>2</sup>Aqualytics

<sup>3</sup>Environmental Chemistry Division, Northwest Fisheries Science Center, National Marine Fisheries Service, National Oceanic and Atmospheric Administration

<sup>4</sup>University of California, Santa Cruz

## 1 Abstract

Diatoms are microscopic marine algae that are critical for global primary production, carbon sequestration, and fisheries productivity. However, select diatoms may form harmful algal blooms, which threaten marine ecosystems and the fisheries they sustain. Rapidly identifying harmful blooms is necessary to effectively manage marine resources, yet current identification methods are limited by expensive and labor-intensive *in situ* point sampling.

Hyperspectral remote sensing enables scalable monitoring, but its ability to resolve taxonomic shifts within phytoplankton groups (e.g. diatoms) is largely unknown. To investigate this uncertainty, we cultured four dominant diatom genera from the California Current upwelling system, including this systems' most abundant harmful algae, *Pseudo-nitzschia*. The hyperspectral absorption and backscatter of these taxa were measured and used to model

spectral reflectances that remote sensing platforms (satellites/drones) might detect. Differences between fingerprints of these taxa were quantified using vector-based and statistical analyses.

Mean spectral differences of 48% were observed between the most dominant diatom, *Thalassiosira*, and the most toxic diatom, *Pseudo-nitzschia*. Differences of approximately 30% were found between *Pseudo-nitzschia* and the second and third most abundant diatoms, *Chaetoceros* and *Asterionellopsis*. Successful identification of *Pseudo-nitzschia*'s reflectance fingerprint was driven by the presence of a unique feature around 560 nm. The distinct spectral fingerprint of *Pseudo-nitzschia* indicates that it can be distinguished from benign diatom blooms using hyperspectral remote sensing platforms.

## 2 Introduction

Diatoms are an important subgroup of phytoplankton, comprising 10 and 20% of global and marine primary production, respectively (Behrenfeld et al. 2021, Tréguer et al. 1995, Moore et al. 2001b, Aumont et al. 2003), and up to 40% of the marine biological pump (Jin et al. 2006, Tréguer et al. 2018). Large-bodied diatoms also support healthy fisheries by mediating efficient energy transfer to higher trophic levels (Chavez, Messié, and Pennington 2011). Of the 1000 plus known genera of diatoms (Fourtanier and Kociolek 2003), most facilitate beneficial ecosystem services. However, a few diatoms can be harmful, producing toxins or physical structures detrimental to fisheries and human health.

Harmful algal blooms (HABs) are aggregations of algae that threaten human health, ecosystems, or human activities (Kudela et al. 2015). In upwelling systems, some of the most toxic HABs come from the genus *Pseudo-nitzschia*. Half of *Pseudo-nitzschia* species produce a potent neurotoxin, domoic acid, that accumulates in primary consumers (e.g. bivalves, zooplankton, small fish) before being concentrated in marine mammals, seabirds, and fishes of commercial importance. For marine mammals and humans, domoic acid poisoning is devastating, causing the swelling and death of neurons, which induces disorientation, seizures, temporary or permanent memory loss, and in some cases, coma and death (Bates et al. 2018).

The world's largest fisheries are found within the highly productive Eastern Boundary Upwelling Systems (Ryther 1969, Chavez and Messié 2009). Here, large-scale nutrient injection supports plankton blooms, which ulti-

mately sustain these fisheries. However, these blooms can be episodically dominated by toxigenic *Pseudo-nitzschia*, with devastating ecological and economic effects—for example the coastwide *Pseudo-nitzschia* bloom in 2015 that spread from Baja California to British Columbia (McCabe et al. 2016). This bloom closed recreational and commercial fisheries, including mussels, razor clams, oysters, Dungeness and rock crabs, anchovies, and sardines. The closures resulted in an estimated revenue loss of \$170 million for the Dungeness Crab fishery alone (Pacific States Marine Fisheries Dungeness Crab Report 2014) (inflation adjusted to ~\$230 million in 2025), placing stress on fishing communities and coastal towns (Moore et al. 2020). The high levels of domoic acid caused severe damage to marine ecosystems with effects propagating throughout the food web. Mass mortalities and strandings of tens of thousands of seabirds, seals, and sea lions occurred, with toxins also detected in whales and dolphins (McCabe et al. 2016, Gobble et al. 2018).

Accurate and timely detection of large *Pseudo-nitzschia* blooms is critical for fisheries management, aquaculture operations, and public health. Current monitoring methods leverage discrete samples of domoic acid accumulation in shellfish tissues with microscopy-based identification of *Pseudo-nitzschia* cells. While these methods provide essential information, they suffer from limited spatial and temporal resolution and delayed detection relative to contamination (Lefebvre et al. 2017). Annual costs of this continuous *in-situ* monitoring are estimated in the millions of dollars (Anderson et al. 2000). These challenges underscore the need for contemporary methods—such as emerging remote sensing approaches—that can accurately detect *Pseudo-nitzschia* blooms across broader spatial and temporal scales.

Harmful algal bloom identification is the first step in ecological analysis, fisheries management, and early warning response. Identification using ocean color is likely the most cost-effective way to assess ecosystem threats over large spatial and temporal scales. To identify HABs using ocean color, we must first develop spectral ‘fingerprints’. Fingerprints are derived by measuring the optical properties (absorption and backscatter) of a specific taxon which, when combined (equation 1), mirror the signal an ocean color remote sensing platform will observe (Gordon et al. 1988, Zaneveld 1995). In practice, isolating fingerprints from the ocean is often confounded by the overlapping optical properties of dissolved substances, detritus, and other phytoplankton. However, the bloom-forming diatoms cultured in this study form dense surface aggregations *in-situ* (Trainer et al. 2009) that dominate the ocean color signal. This dominance makes them particularly well-suited

for identification using spectral fingerprints. Moreover, with the launch of NASA's latest satellite mission, PACE (Werdell et al. 2019), we now have unprecedented spectral coverage of the global oceans. This hyperspectral resolution provides significantly more information about the phytoplankton within the water, illuminating features and subtleties previously unseen.

Satellite-measured ocean color is strongly influenced by particulate backscatter. Backscattered light is all light that returns in the direction of its source after interaction with a particle or surface. The physical properties of this particle or surface (shape, size, and refractive index) influences the intensity, geometry, and spectral characteristics of the backscattered light (Boss et al. 2004). Unlike absorption spectra, whose features are easily associated to their physical drivers (e.g. pigments), the relationships between backscatter spectral features and their physical drivers are more nuanced. Consequently, there is limited understanding of how hyperspectral backscatter's features relate to their physical drivers. Historically, partly due to limited spectral resolution, spectral backscatter was assumed to be a monotonic function, without pronounced spectral features (e.g. Lee, Carder, and Arnone 2002, Lee et al. 2009, Kostadinov, Siegel, and Maritorena 2009, Morel and Prieur 1977). While this assumption works well for clear-oceanic waters (Smith and Baker 1981), coastal and upwelling systems with non log-linear particle size distributions (e.g. plankton blooms) and non-spherical particle shapes violate the basis for this assumption. Indeed, hyperspectral backscatter contains substantially more diagnostic information than spectral slope alone, with unique and more pronounced spectral features than the spectrum of total scatter (this study, Bricaud, Morel, and Prieur 1983, and Morel and Bricaud 1981). The advent of new hyperspectral backscatter technologies (e.g. Novak, Burmeister, and Röttgers 2024) will provide a wealth of new data to link *in situ* biogeochemistry and community composition (including HAB dominance) to hyperspectral backscatter features—which may be derived from NASA PACE. This will advance ocean color inversion models significantly, increasing their utility in detecting HABs by incorporating the information from empirically derived backscatter spectral shapes.

## 3 Methods

### 3.1 Cultures

Absorption and backscatter spectra were measured from acclimated cultures of four diatom genera, known to be the most dominant within the California Current. They are, in order of typical *in-situ* biovolume: *Thalassiosira rotula*, *Chaetoceros affinis*, *Asterionellopsis glacialis*, and *Pseudo-nitzschia spp.* (Du, Peterson, and O’Higgins 2015, Lassiter et al. 2006). Isolates were obtained from the Bigelow National Center for Marine Algae, except for *Pseudo-nitzschia*, which were clonal isolates from the Washington and Oregon coast (*P. fraudulenta*, *P. pungens*, and *P. seriata*). Each taxon was kept in exponential growth in triplicate 2 L Nalgene bottles as semicontinuous batch cultures. All cultures were grown in an environmental control room at 15°C with a salinity of 33 PSU, using L1 media (Guillard and Hargraves 1993). Cultures were exposed to 400  $\mu\text{mol}$  photons  $\text{m}^2/\text{s}$  with a relatively flat, white spectrum. The intensity of the grow lights (a Phyton-Systems light bank) were measured using a QSL-scalar PAR sensor. This light bank was scheduled on a 12:12 sinusoidal light-dark cycle to simulate a natural day-night cycle. Cultures were kept in suspension by bubbling with two aquarium air pumps (ProFILE 5500). Pumped air was filtered through a hydrocarbon trap (Restek) before entering the cultures.

Cells were acclimated to growth conditions (light, nutrients, temperature) for two months ( $\gg 30$  divisions) before experimental sampling to ensure steady-state, photoacclimated growth. Cell and detrital abundance were quantified daily via an Imaging Flow CytoBot (IFCB) to derive growth rates and ensure detrital production remained low. Using light microscopy, the presence of detritus too small for the IFCB was observed, and quantified using a Coulter Counter (Beckman Multisizer 3, 70  $\mu\text{m}$  aperture). The biomass of detritus was small compared to that of the live cells and detrital particles were almost entirely  $< 3 \mu\text{m}$ , much smaller than this study’s phytoplankton which were  $> 10 \mu\text{m}$  (Table 1). Phytoplankton photophysiology was monitored using a fast repetition rate fluorometer (FRRf). Fast repetition rate fluorometry samples were exposed to 15  $\mu\text{mol}$  photons/ $\text{m}^2/\text{s}$  at 480 nm for 10 minutes to relax non-photochemical quenching (NPQ) before measurements were made (Milligan, Aparicio, and Behrenfeld 2012). The amount of time and total irradiance to fully relax NPQ varied by taxon and time of day. We conducted several experiments and found 15  $\mu\text{mol}$  photons/ $\text{m}^2/\text{s}$  for 10

minutes to yield the highest Fv/Fm ( $\sim 0.6 \pm 0.05$ ) for these taxa on average; Fv/Fm values  $> 0.55$  indicated healthy diatoms ready for inline system measurements (Tan et al. 2019, Dijkman and Kromkamp 2006).

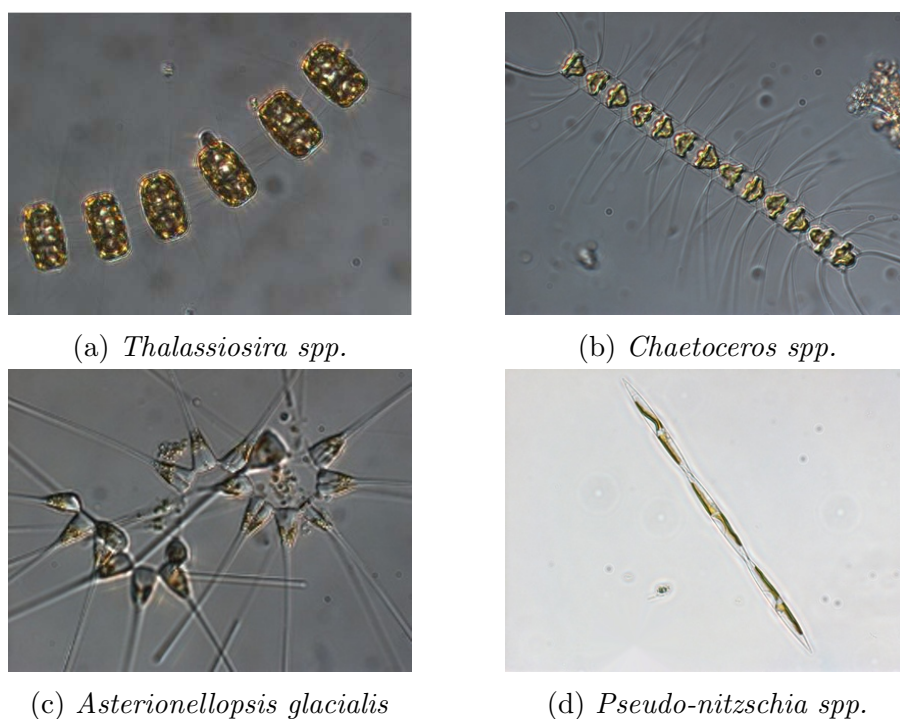


Figure 1: Representative diatom genera (reproduced with permission from the Kudela Lab).

Table 1: Approximate cell size ranges for representative diatom species.

Species	Individual Cell Sizes	
	width ( $\mu\text{m}$ )	length ( $\mu\text{m}$ )
<i>Thalassiosira rotula</i>	8 - 60	5 - 20
<i>Chaetoceros affinis</i>	8 - 15	7 - 30
<i>Asterionellopsis glacialis</i>	5 - 8	30 - 150
<i>Pseudo-nitzschia</i> spp.	3 - 8	70 - 145

## 3.2 Inline system

Measurements of phytoplankton absorption, scatter, and attenuation coefficients were made using a SeaBird (formerly WETLabs) AC-S (4 nm spectral sampling frequency). Backscatter was measured using a Sequoia Hyper-BB (5 nm spectral sampling frequency). All bio-optical measurements were made following the best practices for IOP measurements (Boss et al. 2019). Instruments were configured in a darkened recirculating flow-through inline system driven by a peristaltic pump (Slade et al. 2010). The inline system was filled with 33 PSU, 0.2  $\mu\text{m}$  filtered seawater, leaving sufficient headspace to add 200mL of phytoplankton culture. After inoculation, the solution was allowed to recirculate and homogenize until AC-S and HyperBB data approached a stable asymptote. Measurements were recorded for 10 to 15 minutes and the median derived for each wavelength to reduce the influence of outliers and noise. This process was repeated using a 5.0  $\mu\text{m}$  filtrate of the sample to estimate the optical contributions of detritus and colored dissolved organic matter (CDOM) to the culture. Subtraction of the 5  $\mu\text{m}$  filtrate’s optical signature from that of the total culture isolated the diatom-specific optical signature.

Between daily sampling, the inline system was flushed with deionized water. The optical windows and tubes of the AC-S were cleaned with 98% isopropyl alcohol and Milli-Q water, while the plastic HyperBB windows were cleaned with soap and rinsed with Milli-Q. After cleaning, degassed Milli-Q blanks were recorded for both the AC-S and HyperBB, which were then taken apart and cleaned again until two baseline measurements agreed at the 0.005  $\text{m}^{-1}$  level. All bio-optical data were recorded using the software Inlinino (<https://github.com/OceanOptics/Inlinino>, Haëntjens and Boss 2020).

Measurements of absorption, backscatter, and total scatter were recorded for triplicates of all genera, except *Asterionellopsis*. Due to a gain issue with the HyperBB, only a single backscatter spectrum was obtained for *Thalassiosira* and due to a pump failure at the end of sampling, the backscattering coefficients for the final *Pseudo-nitzschia* triplicate, PN\_3, were measured by placing this sample directly into the HyperBB light trap. The solution was stirred thoroughly to homogenize and then measured (this sample had no detrital blank). To align the spectral sampling frequency (4 nm for the AC-S and 5 nm for the HyperBB) and center wavelengths of both instruments, AC-S data were linearly interpolated to 1 nm resolution and then down-sampled to 5 nm. Intrataxon backscatter and absorption data were then bootstrapped

to provide more spectra for clustering algorithm analysis. All backscatter spectra from triplicate cultures were divided by all absorption spectra of the same genera (e.g. the backscatter spectra of PN\_1 was divided by absorption of PN\_1, 2, and 3. The same was done for the backscatter spectra of PN\_2 and so on). The primary clustering algorithm used was an unsupervised machine learning algorithm, k-means, from the open-source Python machine learning library Scikit-learn. The optimal number of clusters (k) was determined using an elbow method and silhouette scores. To isolate differences in spectral shape, independent of magnitude, a mean-normalization was performed where each value within a spectrum was divided by the mean of the entire spectrum.

### 3.3 Modeling

The spectral fingerprints of our *in vivo* cultures, measured using the inline system, can be related to *in situ* reflectances using the following equation (Gordon et al. 1988, Zaneveld 1995):

$$R_{rs\lambda} \approx \frac{f}{Q} \frac{b_{b\lambda}}{a_{\lambda} + b_{b\lambda}}, \quad (1)$$

where  $R_{rs}$ (reflectance) describes the intensity and color of light, at wavelength  $\lambda$ , leaving the ocean's surface. The backscatter coefficient at wavelength  $\lambda$  is represented by  $b_{b(\lambda)}$ , while  $a_{(\lambda)}$  denotes the absorption coefficient at wavelength  $\lambda$ . The symbols  $f$  and  $Q$  represent scaling factors which do not affect the spectral shape. Because  $a_{(\lambda)}$  is generally multiple orders of magnitude greater than  $b_{b(\lambda)}$ , we can further approximate the equation to:

$$\frac{b_{b\lambda}}{a_{\lambda}}, \quad (2)$$

as the contribution of backscatter in the denominator is often negligible.

### 3.4 Distance Metrics

Distance functions were used to distinguish between modeled spectral fingerprints in a reliable and quantitative way. These functions are able to define the degree of similarity between complex, non-linear spectra by assessing differences in translation (center wavelength or hue), and standard deviation

(full-width half max (FWHM) or peakedness), of spectral features. Deborah, Richard, and Hardeberg 2015 and Deborah 2016 evaluated thirty-one distance functions for hyperspectral analysis. Of these, we chose the Spectral Correlation Mapper (SCM) (Carvalho Júnior and Meneses 2000) as the optimal tool for calculating the differences between phytoplankton fingerprints. Spectral Correlation Mapper builds from vector-based angular methods which have a history of successful applications in remote sensing (e.g. Wei et al. 2022, Kruse et al. 1993). These tools are unaffected by changes in spectral magnitude and provide absolute differences between spectra. Spectral Correlation Mapper uses Pearson’s correlation coefficient ( $r$ ):

$$\text{Spectral Correlation Mapper } (r) = \frac{\sum_{i=1}^n (A_i - \bar{A})(B_i - \bar{B})}{\sqrt{\sum_{i=1}^n (A_i - \bar{A})^2 \sum_{i=1}^n (B_i - \bar{B})^2}} \quad (3)$$
$$\bar{A} = \frac{1}{n} \sum_{i=1}^n A_i \quad \bar{B} = \frac{1}{n} \sum_{i=1}^n B_i ,$$

to evaluate the degree of positive or negative linear correlation between mean-standardized data. Spectral Correlation Mapper processes spectra as a distribution, appropriate for the highly correlated nature of hyperspectral data.

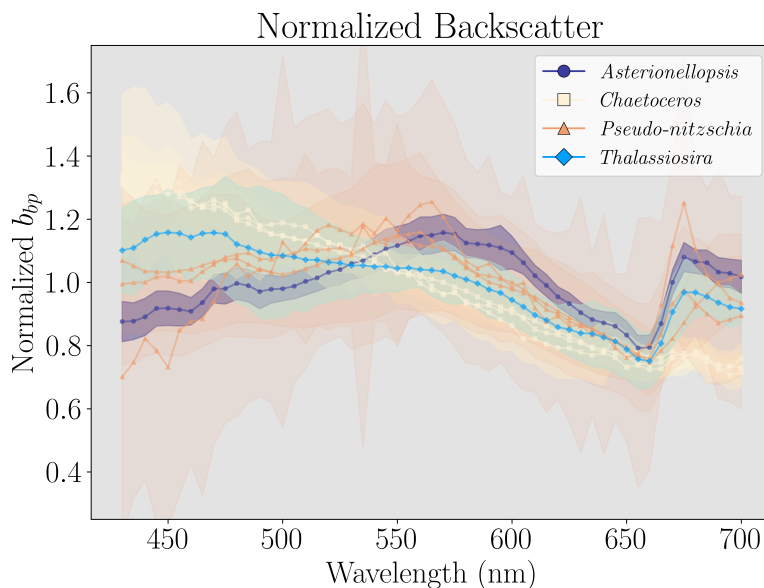
Spectral Correlation Mapper’s sensitivity to differences between spectral signatures can be further improved through the application of derivative transformations. However, while derivatives excel at highlighting small features, the increased sensitivity also amplifies noise, which must first be removed from the spectra. To remove noise while preserving as many of the spectral features as possible, we used a Savitzky-Golay filter. The Savitzky-Golay method preserves large spectral features better than other techniques (such as mean or median filtering) by fitting an ‘x’ order polynomial to the data within the filter window. This preserves spectrally-wide nonlinear features while spectrally-narrow variations (noise) are smoothed out. For these reasons, Savitzky-Golay smoothing is often used in hyperspectral applications (e.g. Vandermeulen et al. 2017, Xi et al. 2015, Hunt 2024). The Savitzky-Golay filter has two tunable parameters, window length and polynomial order. The goal is to find the ‘goldilocks’ parameters where the data are neither oversmoothed (thus losing important information) nor overfit (turning noise into features). We used the Scipy.signal function `savgol_filter` and chose a window length of 13 (5 nm spacing x 13 = 65 nm windows) and a polynomial order of 4 based on the iterations seen in the supplemental materials (figure 11). `Savgol_filter` ‘mode’ was set to ‘nearest’ to reduce boundary

artifacts from the filter window interacting with the end of the spectra (mode ‘nearest’ pads the end with values identical to the last encountered). The option ‘deriv’ was set to 2 (takes the second derivative of the spectra).

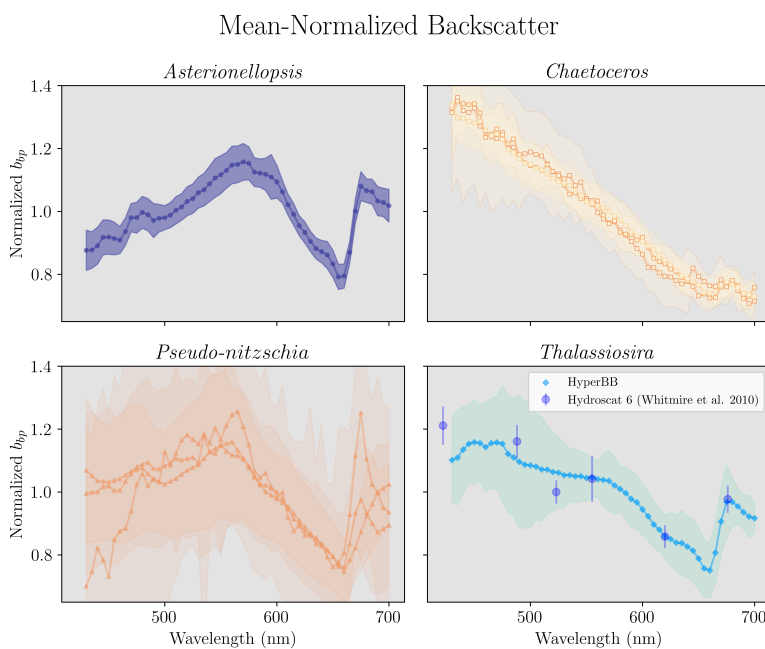
## 4 Results

All cultured phytoplankton share local maxima of backscatter around 560/570 nm (figure 2), the yellow-green region of these spectra. These backscatter maxima are likely amplified by the intersection of absorbance features, from fucoxanthin and the secondary peaks of chlorophylls b and c, which form a local absorbance minima (figure 3). However, the absolute lowest absorbance for all species throughout most of the visible range lies near 600 nm. While we observe local maxima in backscatter at 600 nm (figure 2), these wavelengths do not correspond to the greatest backscatter for any species. Instead, local and total maxima around 560/570 nm appear in nearly all species, being most pronounced for *Asterionellopsis* and *Pseudo-nitzschia*. For *Pseudo-nitzschia*, the 560/570 nm peak is followed by a steep decline that exhibits a subtle inflection near 585 nm before decreasing further. This subtle inflection coupled with the aforementioned peak in *Pseudo-nitzschia*’s backscatter spectra creates an important feature in *Pseudo-nitzschia*’s reflectance fingerprint, seen as a bifurcated peak in figure 4. After the maximum green-yellow (560/570 nm) backscatter peak, *Pseudo-nitzschia* and *Asterionellopsis* backscatter less for bluer wavelengths, contrary to *Thalassiosira* and *Chaetoceros* (figure 2). There is also a consistent trough for all spectra around 660 nm followed by a large peak that overlaps with the chlorophyll Qy absorption band around 670 nm (figure 3).

Figure 3 shows the absorption spectra for each of the cultured diatoms. Triplicate cultures within each taxon exhibit consistent spectral shapes across wavelengths. Differences in the absorption spectra between taxa are less visually pronounced than those observed in normalized backscatter (figure 2) or total scatter spectra (figure 9, supplemental material). There are small deviations in the absorption spectra of the third triplicate sample of *Pseudo-nitzschia* (PN\_3) from the first and second replicates (PN\_1, PN\_2), with slightly elevated mean-normalized values in the blue region (400 – 440 nm), likely related to the aforementioned pump failure for this sample. *Thalassiosira* has the strongest shoulder at 470 nm from the pigment fucoxanthin. *Asterionellopsis* has the largest ratio of Qy to Soret band absorption (chloro-



(a) Mean-normalized backscatter spectra comparison.



(b) Mean-normalized backscatter by genus.

Figure 2: Normalized backscatter spectra for each species. Dots represent the normalized median value of backscatter at each wavelength measured. Ribbons represent one standard deviation from the median, colors distinguish each species. Triplicate samples were measured for *Pseudo-nitzschia* and *Chaetoceros*. A multispectral comparison of *Thalassiosira rotula* from Whitmire et al. 2010 is included in dark blue, with bars representing 1 standard deviation from the median. Multispectral and hyperspectral data were aligned at 620 nm for comparison.

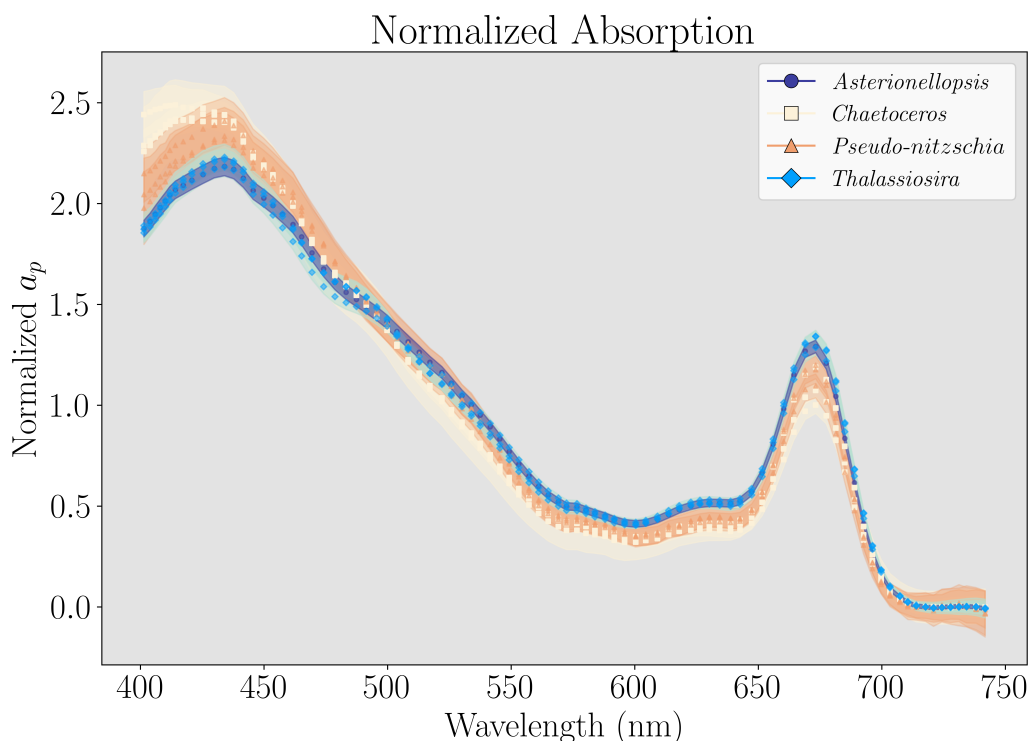
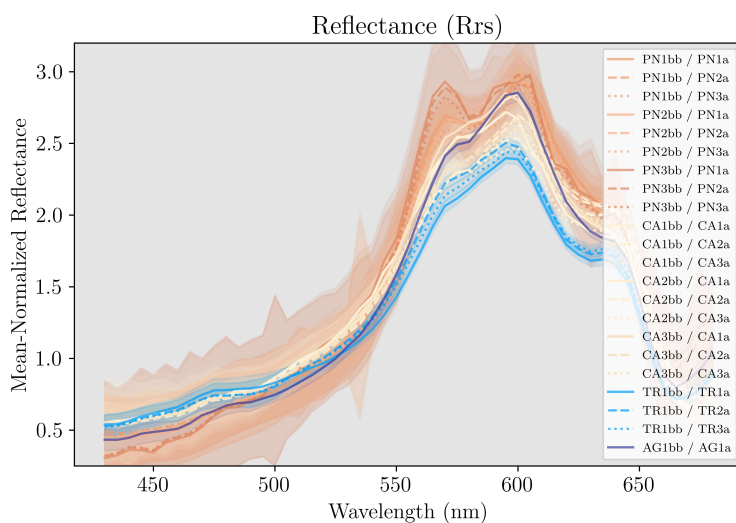


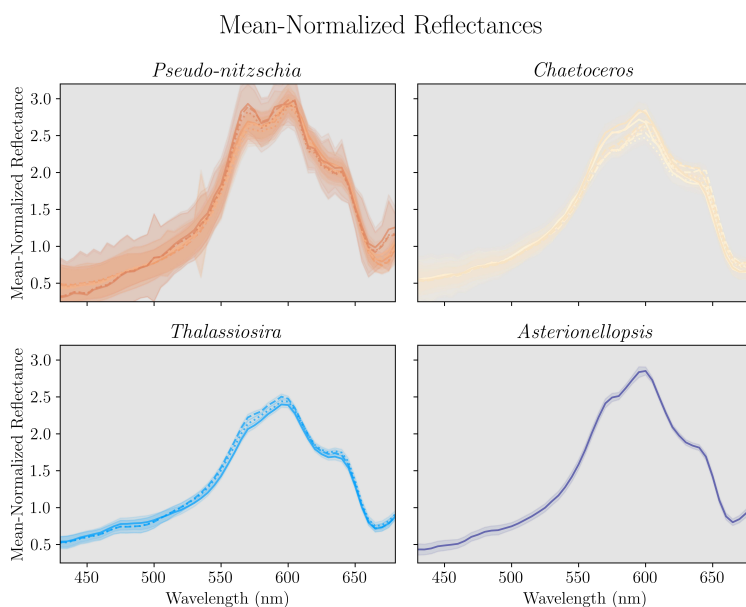
Figure 3: All cultures absorption spectra using the AC-s, spectra are mean-normalized. The peak centered on 440 nm is the Soret band, and the peak over 670 nm is the Qy band.

phyll a absorption in the red vs blue, respectively). *Chaetoceros* appears to have the broadest peak for the Soret band and the lowest relative absorption of carotenoids (pigments in roughly the 470 - 650 nm range).

The most pronounced distinction in the spectral shapes of reflectance (figure 4) is for *Pseudo-nitzschia*, which forms distinct bifurcated peaks around 570 and 600 nm, with a trough at 585 nm. This feature provides robust separation of the spectral fingerprint of *Pseudo-nitzschia* from all other diatoms sampled. In contrast, *Chaetoceros* forms a shoulder near 570 before peaking around 600 nm, while *Thalassiosira* remains linear throughout this range, monotonically increasing. The spectral fingerprint of *Asterionellopsis* is similar to that of *Chaetoceros*, with a shoulder at 570 nm and a peak at 600 nm, however the height of the peak relative to the shoulder is greater for *Asterionellopsis*. For all genera, there are only small differences in slope



(a) Mean-normalized reflectance fingerprints.



(b) All reflectance fingerprints in a 2x2 grid.

Figure 4: Nineteen bootstrapped spectra of the four diatom genera measured in this experiment: *Asterionellopsis* (AG) is purple, *Chaetoceros* (CA) is yellow, *Pseudo-nitzschia* (PN) is orange, and *Thalassiosira* (TR) is blue. Each spectrum is the mean value of hundreds to thousands of absorption and backscatter measurements (for each  $\lambda_i$ ), converted to reflectance spectra using equation 2.

from the deep-blue to cyan ( $\sim 430$  to  $510$  nm), with the largest differences in yellow and orange ( $\sim 560$  to  $620$  nm). There is relative uniformity at all other wavelengths. Because diatom absorption spectra are highly similar, genus-level discrimination is primarily driven by spectral backscatter. Unique backscatter features are most visible in reflectance spectra when centered within the absorbance minima of  $550$  to  $650$  nm.

Unsupervised machine learning algorithms were able to successfully distinguish between the spectral fingerprints of *Pseudo-nitzschia*, *Chaetoceros*, and *Thalassiosira* (figure 5). The highest silhouette coefficient found used  $k=5$  (0.5), but  $k=2-4$  were all within 8% of this score (0.46 - 0.48). Using  $k=2$  (clustering the spectra into two groups) isolates the *Pseudo-nitzschia* spectra, separating them from all other taxa. Using  $k=3$  provides nearly perfect clustering of spectra, with the exception that *Asterionellopsis* is misclassified as *Chaetoceros*. Using  $k=4$  yields similar results, however *Pseudo-nitzschia* is divided into two clusters. Lastly, using  $k=5$  divides *Chaetoceros* into two clusters (CA1bb triplicates and AG, versus CA2bb and CA3bb triplicates). Other non-Euclidean machine learning algorithms were applied and results from these are provided in the appendix.

To accentuate spectral features, the second derivatives of reflectance fingerprints from figure 4 were calculated and are shown in figure 6. These smoothed, second derivative fingerprints were then compared using the Spectral Correlation Mapper (SCM) distance metric. A heatmap of the evaluated differences is shown in figure 7. The SCM performed better when subsetting the spectra in figure 6 from  $550$  to  $600$  nm, wavelengths where the majority of spectral differences are found. Spectral Correlation Mapper calculates all *Pseudo-nitzschia* spectra as having a mean intertaxon correlation of 98% with a standard deviation of  $\pm 1\%$  (upper-left, red, right triangle), while *Thalassiosira* spectra are deemed nearly identical to each other (99% or greater correlation). *Chaetoceros* spectra are the most variable with a standard deviation of 7% and an average correlation of 91%. Spectral correlation mapper finds average differences of  $29\% \pm 13\%$  between *Pseudo-nitzschia* and *Chaetoceros* (top, black box), and  $34\% \pm 7\%$  between *Pseudo-nitzschia* and *Asterionellopsis* (bottom, black rectangle). Spectral correlation mapper performed exceedingly well for *Pseudo-nitzschia* and *Thalassiosira* (middle, black rectangle)—the most toxic and most common diatoms, respectively—with average differences of  $48\% \pm 10\%$ .

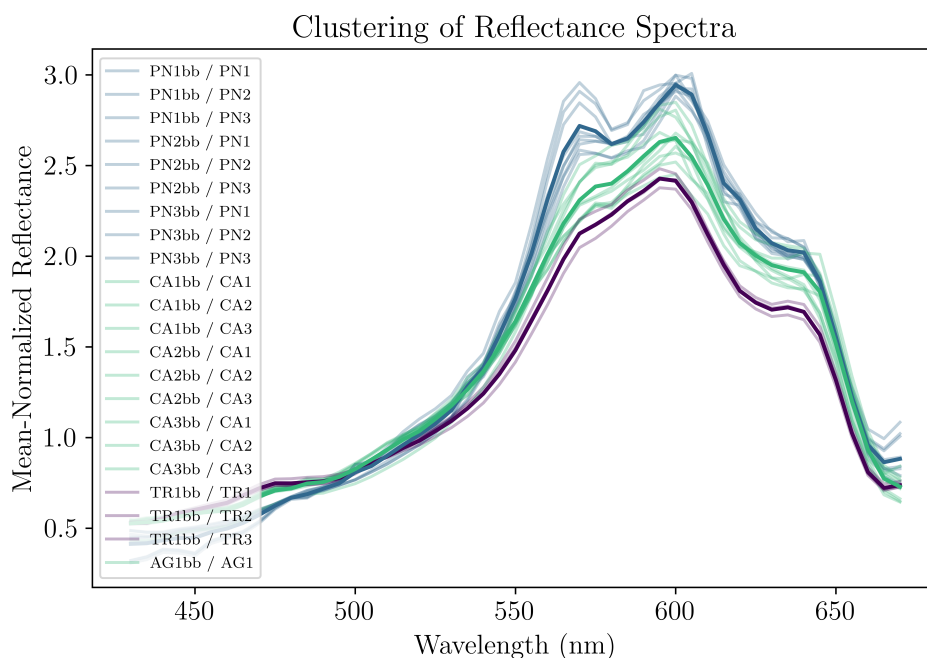


Figure 5: All bootstrapped spectra clustered using the k-means algorithm for  $k=3$ . *Pseudo-nitzschia* is abbreviated as 'PN', *Chaetoceros* as 'CA', *Thalassiosira* as 'TR', and *Asterionellopsis* as 'AG'. The number following the two letter phytoplankton ID code corresponds to the replicate culture number. 'bb' refers to backscatter. Therefore, **PN1bb/PN2** corresponds to the backscatter spectra of the first replicate culture of *Pseudo-nitzschia* divided by the absorption spectra of the second replicate culture of *Pseudo-nitzschia* (a bootstrapped sample). Measured spectra are shown semi-transparent while the derived mean of each cluster is opaque. Ignoring *Asterionellopsis* (only one sample), the clustering algorithm successfully classified all taxa.

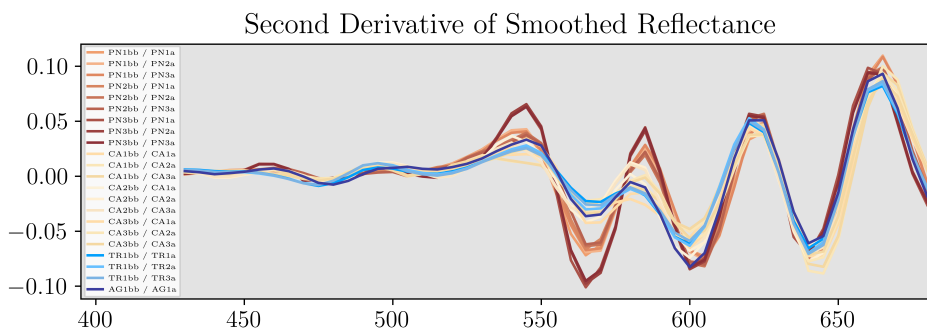


Figure 6: Second derivative of the smoothed spectra.

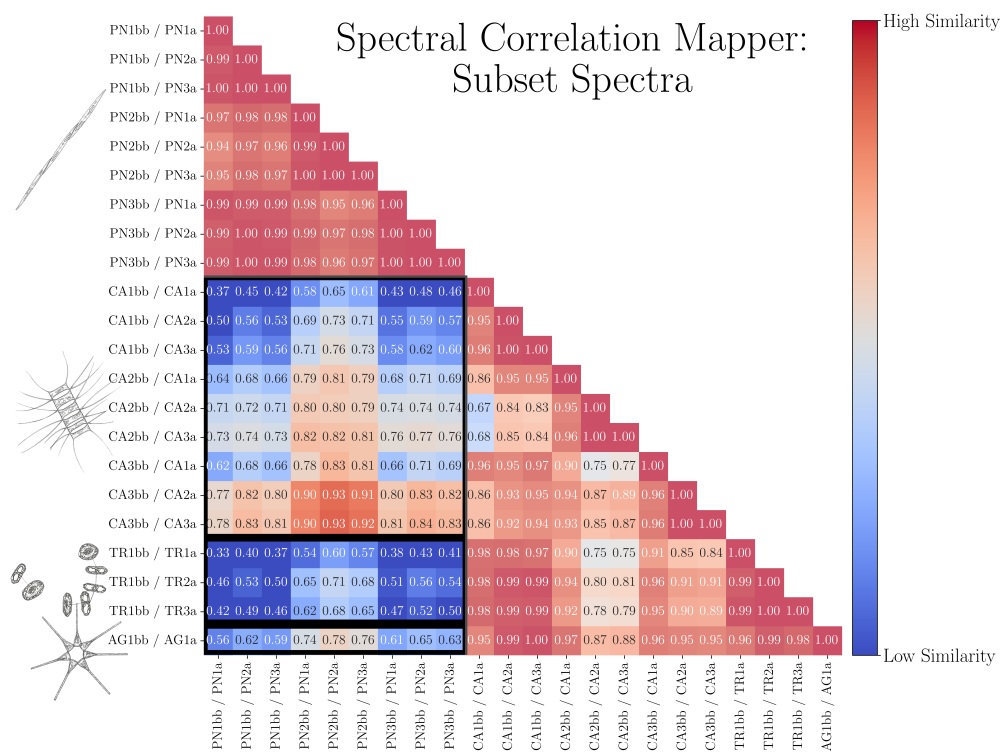


Figure 7: Spectral Correlation Mapper applied to a subset of the reflectance spectra (550 to 600 nm) from figure 6 to focus on regions of maximum differences. The black boxes highlight comparisons between the fingerprints of *Pseudo-nitzschia* (PN) and other genera—summarized in table 2.

Table 2: Mean spectral differences between *Pseudo-nitzschia* and other genera, with standard deviations.

Mean Difference and Standard Deviation of Spectral Fingerprints: Comparing <i>Pseudo-Nitzschia</i> and other Diatoms	
Taxa	<i>Pseudo-nitzschia</i> spp.
<i>Chaetoceros affinis</i>	29% ± 13%
<i>Thalassiosira rotula</i>	48% ± 10%
<i>Asterionellopsis glacialis</i>	34% ± 7%

## 5 Discussion

Our approach yielded distinct reflectance fingerprints, each related to a unique phytoplankton genera under bloom conditions. We identified visually discernible, group-specific spectral features and quantified these differences. *Pseudo-nitzschia*, the dominant HAB taxa, exhibits a unique feature in its modeled reflectance fingerprint, a bifurcated peak between 570 and 600 nm. This feature easily distinguishes the spectral fingerprint of *Pseudo-nitzschia* from those of the three most abundant diatoms in this system—both through visual assessment and using unsupervised classifying algorithms. *Pseudo-nitzschia*'s fingerprint was most easily distinguished from *Thalassiosira*'s, which was nearly flat between 565 and 600 nm. This distinction is important because *Thalassiosira* is the most abundant bloom-forming diatom in the Northern California Current. The ability to discern between *Thalassiosira* and *Pseudo-nitzschia* based on spectral fingerprints is valuable for a number of applications, such as productivity modeling, species distribution mapping, and HAB monitoring. While *Chaetoceros* fingerprints were more variable between these wavelengths, visual assessments and clustering algorithms were still able to distinguish between the fingerprint of *Chaetoceros*, the second most abundant diatom, and *Pseudo-nitzschia*.

Unsupervised clustering algorithms (figure 5 and appendix, figure 10) successfully isolated the unique spectral fingerprints of *Pseudo-nitzschia* from those of *Thalassiosira*, and *Chaetoceros*. These algorithms proved robust and versatile, successfully classifying *Pseudo-nitzschia*'s fingerprint regardless of spectral permutations (full spectral range or subset spectra, raw spectra or the second derivative, smoothed spectra or unsmoothed). Clustering algorithm's success with these fingerprints suggests an automated spectral classification framework will be able to distinguish between monospecific or near monospecific blooms of the most common diatoms and *Pseudo-nitzschia* using the PACE satellite.

### 5.1 Importance of backscatter

These experiments reveal the importance of integrating measured spectral backscatter into ocean color models. Rethinking the assumption that spectral backscatter is relatively featureless, being either flat or a monotonic function, may help us to disentangle the complexity of case 2 waters (optically-complex waters influenced by terrigenous inputs, namely CDOM and sediments). Un-

der the assumption of spectrally flat backscatter, reflectances are merely the inverse of absorption (figure 8). Spectral shapes in these ‘pseudoreflectances’ are nearly identical, with no translational differences between spectral features and only minor changes in the FWHM of peaks at 600 nm. Meanwhile, the variability and features seen in figure 4 highlight the complexity of parameters that define the spectra of backscatter: unique shapes, sizes, and intracellular structures. It is due to the unique morphological features, stemming from the great diversity of the diatoms, that the complexity in the spectral shapes of backscatter—and consequently, reflectance fingerprints—enables the identification of *Pseudo-nitzschia* from other diatom assemblages.

Our results indicate that absorption-based models will be unable to discern *Pseudo-nitzschia* in mixed diatom assemblages, while reflectance-based models that integrate spectral backscatter have potential. This is because spectral backscatter presents unique features and shapes (figure 2). Modeling the backscatter of these diatom’s complex frustule geometries and intracellular structures is prohibitively computationally intensive, making direct attribution of individual spectral features to morphologies challenging. However, we hypothesize that the continuous increase of backscatter coefficients in bluer wavelengths for *Chaetoceros* is due to the presence of many thin setae (spines) whose small size should preferentially scatter shorter (bluer) wavelengths of light. A similar increase in backscattering coefficients of bluer wavelengths is observed for *Thalassiosira*, due to the presence of many small chitinous thread structures which adorn the frustule (Tran et al. 2023). Many of these threads are joined together in the center of the frustule to bind *Thalassiosira* cells in chain formations. We propose that chitins’ increased absorption in blue-green wavelengths (Azofeifa, Arguedas, and Vargas 2012), relative to silica, makes these thread structures less efficient backscatters than setae. The varying efficiencies of backscatter for these distal structures may explain the observed differences in figure 2a between *Thalassiosira* and *Chaetoceros* in blue-green wavelengths. We note that natural populations of *Asterionellopsis glacialis* often have large siliceous spikes, similar to *Chaetoceros* setae, however in our cultures the length of these spikes were reduced compared to natural samples. This may explain why *Asterionellopsis* does not have the same high backscatter coefficients in blue wavelengths as *Chaetoceros* and *Thalassiosira*. However, both *Chaetoceros* and *Thalassiosira* possess numerous setae or threads per cell, whereas *Asterionellopsis* has but a single, large spike per individual. As a result, the backscatter spectrum of *Asterionellopsis* in blue-green wavelengths may be more strongly influenced

by absorption features—the Soret band and fucoxanthin—than by elevated backscatter from a single siliceous spike, leading to the spectral shape observed in figure 2.

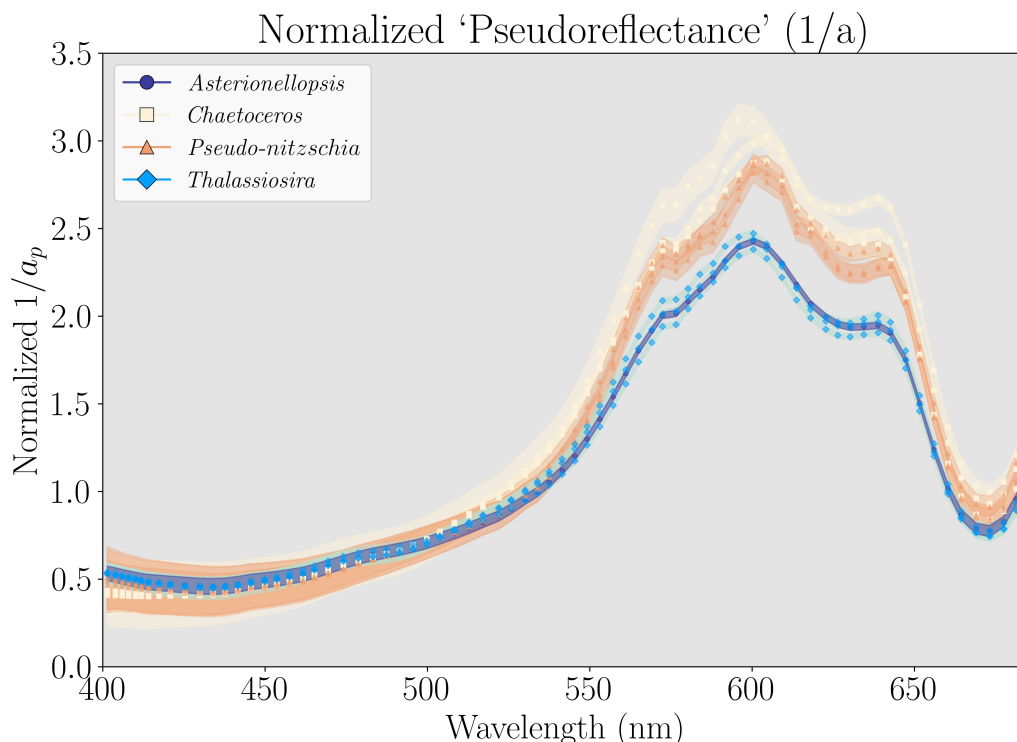


Figure 8: What if backscatter was flat? Pseudoreflectances shown assuming backscatter is spectrally flat for all samples.

The peak in backscatter for all species past 660 nm (figure 2a) is likely due to inelastic scatter in the form of fluorescence, rather than true backscatter. This fluorescence signal results from the large FWHM ( $\sim 25$  nm) of light emitted and accepted by the HyperBB in red wavelengths. Past 660 nm, the Hyperbb efficiently excites phytoplankton photosystems due to a high absorption coefficient from the chlorophyll  $Q_y$  band. Some of the absorbed photons are subsequently re-emitted by the phytoplankton at slightly redder wavelengths as fluorescence, which are then detected by the instrument and registered as backscatter. This effect appears to be an instrument-specific artifact of the HyperBB's detection window, rather than a property of phytoplankton scattering itself. Nevertheless, agreeing with the postulates of

Bricaud, Morel, and Prieur, 1983, the backscatter spectra presented more spectral complexity and shape, and therefore information, than the spectra of total scatter alone (figure 2 versus supplemental material, figure 9).

## 5.2 Orientation of cells

During these experiments, diatoms were measured in a turbulent environment that randomized their orientation. Since these taxa have complex, non-spherical shapes, stochastic orientation creates a plethora of apparent sizes and shapes (cross-sectional areas and geometries, relative to the incoming light source and detector). These apparent shapes and sizes will affect the optical properties of the cells (Marcos et al. 2011).

We observed an increase in the variability of backscatter coefficients with bluer wavelengths in all taxa, but especially *Pseudo-nitzschia*. Mie theory—which can be used to derive the optical properties of a single, homogeneous sphere—offers an explanation, citing backscatter coefficients of bluer wavelengths to be more sensitive to changes in apparent size (Mie 1908, Bohren and Huffman 2004, Van de Hulst 2012). Mie’s derivations are based on spherical particles, which do not adequately represent the complex morphologies of diatoms examined here. However, backscatter coefficients and spectral features for complex (non-spherical) particles are poorly constrained (Clavano, Boss, and Karp-Boss 2007), making Mie theory an imperfect but indispensable framework for inference in particle-light interactions.

Hence, we theorize that the large variability in backscatter coefficients for bluer wavelengths is due to changes in apparent size. This is supported by *Pseudo-nitzschia*’s backscatter spectra, which possesses the greatest variability of all backscatter spectra, especially for bluer wavelengths. Accordingly, *Pseudo-nitzschia* cells present the largest differences in apparent size between the apical axis and top-down, valve-apex view. The apical axis presents a large cross-sectional area when facing the light-source and detector, an oblate spheroid. While in a top-down view of the valve-apex, even *Pseudo-nitzschia* forming long chains present but a small, nearly spheroid, cross-sectional-area to the sensor. This small, spherical cross section would elevate backscatter in bluer wavelengths, while reducing overall scatter (Mie 1908, Marcos et al. 2011). The difference in apparent size between apical and valve-apex views of *Pseudo-nitzschia* is greater than that of any other diatom sampled in this experiment and will increase multiplicatively with chain length. Indeed, due to *Pseudo-nitzschia*’s unique chain morphology, this difference in apparent size,

and the resulting variability in backscatter coefficients, is likely higher than the vast majority of other marine phytoplankton, excluding a few penates with similar morphologies and chain-formation patterns (e.g. *Rhizosolenia*). In a turbulent medium (e.g. our inline system) where orientation is randomized, the direct valve-apex view will be significantly less probable than some angle of the planar view.

In their natural environment, cells are often assumed to be well mixed in the euphotic zone, but this is not always the case. During periods of laminar flow or stratification, some phytoplankton will orient themselves to the currents or sunlight, respectively (Karp-Boss and Jumars 1998, Nayak et al. 2018). Aligned orientation alters the apparent cross-sectional area of cells and can substantially modify both the magnitude (sometimes by more than 30% (Marcos et al. 2011)) and spectral characteristics of backscattered light. The optical effects of orientation will be most strongly expressed in the polarization of backscatter. Multi-angle polarimetry from platforms such as PACE offers a way to assess orientation-dependent scattering and polarization in natural assemblages, which will help to constrain uncertainties associated with phytoplankton taxonomic retrievals.

### 5.3 Distance metrics and proposed use

To quantify differences between spectra, we used the distance function SCM. Compared to other distance functions, SCM is easy to interpret (Pearson's correlation coefficient ( $r$ )) and insensitive to changes in magnitude, while still responsive to differences in spectral shape (Carvalho Júnior and Meneses 2000). Spectral Correlation Mapper performed best after applying the second derivative to the reflectance fingerprints as spectral features were accentuated. This quantification of spectral differences is important for establishing baselines to compare ocean color data to modeled reflectance fingerprints.

The low standard deviation between mean reflectance fingerprints of *Pseudo-nitzschia* replicates (1%, figure 7) suggests that stringent cutoffs can be applied when looking for *Pseudo-nitzschia*-dominated blooms. Differences between the most abundant diatoms and *Pseudo-nitzschia* ranged from 29 to 48% (table 2). Thus, when comparing ocean color data to the modeled fingerprints of *Pseudo-nitzschia*, the threshold for *similarity* between *in-situ* and modeled spectra should be greater than 71%. Ocean color datasets can be flagged for pixels with high spectral similarity to the modeled reflectance fingerprints, providing an automated and scalable framework.

These distance metrics perform best under bloom-conditions when the optical properties of the dominant phytoplankton are the primary contributor to ocean color. Removing the reflectance spectrum of pure water from the ocean color signal should produce a spectrum that closely matches the modeled fingerprints—if the bloom is dominated by one of these modeled taxa.

However, there will be greater spectral variation in natural samples due to mixed assemblages and other optically active components (colored dissolved organic matter (CDOM), detritus, sediments). This variability will make it difficult for distance metrics to select pixels where key phytoplankton groups are present, as the influence of other optically active components will shift the spectral shape further from that of the modeled fingerprints. Fortunately, the spectral shapes of CDOM and detritus are generally featureless, monotonic functions in visible wavelengths (Aurin, Mannino, and Lary 2018). Additionally, while the optical signatures of CDOM and detritus can have significant influence on the magnitude of ultraviolet and blue wavelengths, these spectral regions are ignored in this proposed framework for diatom identification. Corrections can also be made to account for CDOM contributions using widely-accepted remote sensing algorithms (e.g. Zhu et al. 2011, Aurin, Mannino, and Lary 2018, Bélanger, Babin, and Larouche 2008, Lee et al. 2009). The resuspension of sediments in the nearshore environment will be more difficult to account for, as the reflectance of these materials align with the yellow-orange wavelengths (560 - 630 nm) used for *Pseudo-nitzschia* identification. This may prevent identification in turbid estuaries and the very immediate nearshore environment. However, these spaces account for a small fraction of the total area within upwelling systems.

The primary obstacle in diatom identification will be the competing optical properties of mixed assemblages. While distance metrics should still be used to pull out pixels (spatial bins) with high similarity to modeled fingerprints, this should be followed up with further analysis into the spectra of pixels that do not pass the similarity threshold. This can be done by first using a clustering algorithm to subset the region of interest into ‘k’ bins with similar spectral shapes. From these k-bins, a visual analysis or band ratios near 570 and 585 nm, may be used to identify *Pseudo-nitzschia* within mixed assemblages. The competing optical properties of many unique phytoplankton taxa may denature the structure of the *in-situ* reflectance too much for distance metrics to discern. However, if *Pseudo-nitzschia* is present in sufficient concentration, the distinct ‘M’-shape seen in figure 4 should be

present.

Inverse models and hyperspectral libraries will benefit from future studies investigating the detection limits of unique taxa within mixed community assemblages, as well as additional measurements of taxa-specific hyperspectral backscatter. As higher spectral resolution and lower bandpass sensors become available, specifically open-source and accessible tools (e.g. Novak, Burmeister, and Röttgers 2024), we anticipate other genus-specific spectral features to become known. This will further constrain uncertainties in ocean color models including the model proposed in this study.

## 5.4 Integration with Existing Monitoring Frameworks

The California—Harmful Algae Risk Mapping (C-HARM) model was developed to combat the uncertainty in *Pseudo-nitzschia*'s presence and variable toxin production along the California coast (Anderson et al. 2016, 2011, 2009, 2006). California—Harmful Algae Risk Mapping integrates regional ocean circulation and statistical models with remote sensing data to provide forecasted predictions for the presence of toxic *Pseudo-nitzschia*. The remote sensing data integrated into the model predicts *Pseudo-nitzschia* cellular abundance using the ratio of reflectance from 488 to 555 nm, while the intensity of reflectance at 555 nm is coupled with reflectance-derived chlorophyll and other products to estimate toxin concentration. These remote sensing products and bands, while useful correlative variables within the Southern California Current, are not strictly tied to *Pseudo-nitzschia*. Hence, extrapolation to higher latitudes within the California Current may be less reliable if the biogeochemical relationships to *Pseudo-nitzschia* and its toxicity are dynamic.

The PACE satellite's hyperspectral resolution will allow us to better target the spectral regions of interest identified above—providing an optical proxy *directly tied* to the inherent optical properties of *Pseudo-nitzschia* itself. We anticipate that these findings can augment the current C-HARM model by integrating hyperspectral genus-specific fingerprinting using PACE data. This may increase C-HARM's skill when extrapolating to new environments, e.g. the Northern California Current (Oregon, Washington, British Columbia). This is especially important because these more temperate zones must be sampled for *Pseudo-nitzschia* and domoic acid with greater frequency than Central and Southern California, due to strong seasonality and rapid bloom progression (Frolov, Kudela, and Bellingham 2013). This di-

rect spectral pathway for detection may further allow the extrapolation of this approach to other systems where *Pseudo-nitzschia* is present in harmful concentrations.

## 5.5 Conclusion

In this study, we measured the hyperspectral backscatter coefficients of the dominant diatom genera within the California Current, including the region's most prominent HAB genera, *Pseudo-nitzschia*. Hyperspectral reflectances were then modeled and spectral differences derived through quantitative comparison. Our results support future algorithm development using remote sensing for the identification of harmful *Pseudo-nitzschia* blooms, particularly using hyperspectral observations from missions such as NASA PACE. This capability may be used to inform local agencies of HAB events in near-real time and facilitate targeted *in situ* sampling of toxin levels.

While additional field-based studies will be necessary to validate the fingerprints derived from these experiments, our data provide evidence that significant differences exist in the spectral shapes of these key plankton that are sufficient to discern between them using the methods provided. These results offer a new approach for monitoring biodiversity and detecting harmful algal blooms.

## 5.6 Acknowledgments

We would like to thank all students and staff of the 2023 Ocean Optics Class and the 2022 Inline Workshop at the University of Maine. These courses fostered a close-knit community, provided mentorship on best practices in ocean optics, and introduced students to valuable open-source resources such as *In-linino*. Emmanuel Boss, Collin Roesler, Meg Estapa, Ivona Cetinić, Wayne Slade, Nils Haëntjens, Patrick Gray, Charlotte Begouen Demeaux, and Guillaume Bourdin lead lectures and labs, giving students the knowledge that made this work possible. We are also indebted to many for conversations and advice that greatly benefited this work, including Mallie Hunt, Tara Conrad, Kim Halsey, Allen Milligan, Robert Frouin, Ashley Ohall, and Luke Carberry. We are deeply grateful to friends who helped us with sampling, setup, and late-night culture maintenance: Kelly George, Finley Dibert-Burgwin, Wave Moretto, and Ian Black. We are grateful for support from the OSU Seascape Ecology Lab, OSU Phytoplankton Ecophysiology Lab, and resources from the UCSC Kudela Lab, and the OSU Giovannoni lab.

## Funding

Support for this work was provided by the following grants. The Hatfield Marine Science Center Mamie Markham Award to ALB. NSF-OCE-2306993 and SFI-LS-BIOS-SCOPE-00017493 to NB. NASA grant 80NSSC25K0512 to MB. NASA grant 80NSSC21K1651 to RK. The Wayne & Gladys Valley Charitable Foundation PNW-HOPe Award to MTK, and NOAA grant NA24NOSX012C0022 to MTK.

## Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this article.

## References

Ahn, Y.-H., A. Bricaud, and A. Morel (Nov. 1992). “Light backscattering efficiency and related properties of some phytoplankters”. en. In: *Deep Sea Research Part A. Oceanographic Research Papers* 39.11, pp. 1835–1855. ISSN: 0198-0149. DOI: 10.1016/0198-0149(92)90002-B. URL: [https://doi.org/10.1016/0198-0149\(92\)90002-B](https://doi.org/10.1016/0198-0149(92)90002-B).

- [//www.sciencedirect.com/science/article/pii/S019801499290002B](https://www.sciencedirect.com/science/article/pii/S019801499290002B) (visited on 06/05/2023).
- Anderson, C. R., R. M. Kudela, C. Benitez-Nelson, E. Sekula-Wood, C. T. Burrell, Y. Chao, G. Langlois, J. Goodman, and D. A. Siegel (2011). “Detecting toxic diatom blooms from ocean color and a regional ocean model”. en. In: *Geophysical Research Letters* 38.4. eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1029/2010GL045858>. ISSN: 1944-8007. DOI: 10.1029/2010GL045858. URL: <https://onlinelibrary.wiley.com/doi/abs/10.1029/2010GL045858> (visited on 04/25/2025).
- Anderson, C. R., R. M. Kudela, M. Kahru, Y. Chao, L. K. Rosenfeld, F. L. Bahr, D. M. Anderson, and T. A. Norris (Nov. 2016). “Initial skill assessment of the California Harmful Algae Risk Mapping (C-HARM) system”. In: *Harmful Algae* 59, pp. 1–18. ISSN: 1568-9883. DOI: 10.1016/j.hal.2016.08.006. URL: <https://www.sciencedirect.com/science/article/pii/S1568988315301037> (visited on 02/11/2025).
- Anderson, C. R., D. A. Siegel, R. M. Kudela, and M. A. Brzezinski (Feb. 2009). “Empirical models of toxigenic *Pseudo-nitzschia* blooms: Potential use as a remote detection tool in the Santa Barbara Channel”. In: *Harmful Algae* 8.3, pp. 478–492. ISSN: 1568-9883. DOI: 10.1016/j.hal.2008.10.005. URL: <https://www.sciencedirect.com/science/article/pii/S1568988308001285> (visited on 04/25/2025).
- Anderson, C., M. Brzezinski, L. Washburn, and R. Kudela (Dec. 2006). “Circulation and environmental conditions during a toxigenic *Pseudo-nitzschia australis* bloom in the Santa Barbara Channel, California”. en. In: *Marine Ecology Progress Series* 327, pp. 119–133. ISSN: 0171-8630, 1616-1599. DOI: 10.3354/meps327119. URL: <http://www.int-res.com/abstracts/meps/v327/p119-133/> (visited on 04/25/2025).
- Anderson, D. M., E. Fensin, C. J. Gobler, A. E. Hoeglund, K. Hubbard, D. M. Kulis, J. H. Landsberg, K. A. Lefebvre, P. Provoost, M. L. Richlen, J. L. Smith, A. R. Solow, and V. L. Trainer (Feb. 2021). “Marine harmful algal blooms (HABs) in the United States: history, current status and future trends”. In: *Harmful algae* 102, p. 101975. ISSN: 1568-9883. DOI: 10.1016/j.hal.2021.101975. URL: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8058451/> (visited on 05/16/2025).
- Anderson, P. Hoagland, Y. Kaoru, and A. W. White (Sept. 2000). *Estimated Annual Economic Impacts From Harmful Algal Blooms (Habs) In The United States*. en. Tech. rep. Woods Hole Oceanographic Institution

- (WHOI). URL: <https://repository.library.noaa.gov/view/noaa/34913> (visited on 04/23/2025).
- Aumont, O., E. Maier-Reimer, S. Blain, and P. Monfray (2003). “An ecosystem model of the global ocean including Fe, Si, P colimitations”. en. In: *Global Biogeochemical Cycles* 17.2. eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1029/2001GB001745>. ISSN: 1944-9224. DOI: 10.1029/2001GB001745. URL: <https://onlinelibrary.wiley.com/doi/abs/10.1029/2001GB001745> (visited on 05/15/2025).
- Aurin, D., A. Mannino, and D. J. Lary (Dec. 2018). “Remote Sensing of CDOM, CDOM Spectral Slope, and Dissolved Organic Carbon in the Global Ocean”. en. In: *Applied Sciences* 8.12. Number: 12 Publisher: Multidisciplinary Digital Publishing Institute, p. 2687. ISSN: 2076-3417. DOI: 10.3390/app8122687. URL: <https://www.mdpi.com/2076-3417/8/12/2687> (visited on 04/24/2025).
- Azofeifa, D. E., H. J. Arguedas, and W. E. Vargas (Dec. 2012). “Optical properties of chitin and chitosan biopolymers with application to structural color analysis”. In: *Optical Materials* 35.2, pp. 175–183. ISSN: 0925-3467. DOI: 10.1016/j.optmat.2012.07.024. URL: <https://www.sciencedirect.com/science/article/pii/S0925346712003345> (visited on 04/15/2025).
- Bates, S. S., K. A. Hubbard, N. Lundholm, M. Montresor, and C. P. Leaw (Nov. 2018). “Pseudo-nitzschia, Nitzschia, and domoic acid: New research since 2011”. en. In: *Harmful Algae* 79, pp. 3–43. ISSN: 15689883. DOI: 10.1016/j.hal.2018.06.001. URL: <https://linkinghub.elsevier.com/retrieve/pii/S156898831830091X> (visited on 12/11/2024).
- Bates, S. S., N. Lundholm, K. A. Hubbard, M. Montresor, and C. P. Leaw (2019). “Toxic and Harmful Marine Diatoms”. en. In: *Diatoms: Fundamentals and Applications*. Section: 17 eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/9781119370741.ch17>. John Wiley & Sons, Ltd, pp. 389–434. ISBN: 978-1-119-37074-1. DOI: 10.1002/9781119370741.ch17. URL: <https://onlinelibrary.wiley.com/doi/abs/10.1002/9781119370741.ch17> (visited on 05/16/2025).
- Behrenfeld, M. and P. Falkowski (1997). “Photosynthetic rates derived from satellite-based chlorophyll concentration”. en. In: *Limnology and Oceanography* 42.1. eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.4319/lo.1997.42.1.0001>, pp. 1–20. ISSN: 1939-5590. DOI: 10.4319/lo.1997.42.1.0001. URL:

- <https://onlinelibrary.wiley.com/doi/abs/10.4319/lo.1997.42.1.0001> (visited on 06/05/2023).
- Behrenfeld, M., P. Bontempi, H. Dierssen, S. Dutkiewicz, L. Lorenzoni, M. Omand, J. Scott, and J. Werdell (2022). *Earth's Living Ocean: Vast, Dynamic, Essential to Humanity*. URL: [https://oceancolor.gsfc.nasa.gov/OBB\\_ScienceVision/](https://oceancolor.gsfc.nasa.gov/OBB_ScienceVision/).
- Behrenfeld, M., K. H. Halsey, E. Boss, L. Karp-Boss, A. J. Milligan, and G. Peers (2021). "Thoughts on the evolution and ecological niche of diatoms". en. In: *Ecological Monographs* 91.3. eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/ecm.1457>, e01457. ISSN: 1557-7015. DOI: 10.1002/ecm.1457. URL: <https://onlinelibrary.wiley.com/doi/abs/10.1002/ecm.1457> (visited on 03/18/2025).
- Bélanger, S., M. Babin, and P. Larouche (2008). "An empirical ocean color algorithm for estimating the contribution of chromophoric dissolved organic matter to total light absorption in optically complex waters". en. In: *Journal of Geophysical Research: Oceans* 113.C4. eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1029/2007JC004436>. ISSN: 2156-2202. DOI: 10.1029/2007JC004436. URL: <https://onlinelibrary.wiley.com/doi/abs/10.1029/2007JC004436> (visited on 04/24/2025).
- Bell, G. R. (Oct. 1961). "Penetration of Spines from a Marine Diatom into the Gill Tissue of Lingcod (*Ophiodon elongatus*)". en. In: *Nature* 192.4799. Publisher: Nature Publishing Group, pp. 279–280. ISSN: 1476-4687. DOI: 10.1038/192279b0. URL: <https://www.nature.com/articles/192279b0> (visited on 05/15/2025).
- Bhattacharjya, R., T. Kiran Marella, A. Tiwari, A. Saxena, P. Kumar Singh, and B. Mishra (Dec. 2020a). "Bioprospecting of marine diatoms *Thalassiosira*, *Skeletonema* and *Chaetoceros* for lipids and other value-added products". In: *Bioresource Technology* 318, p. 124073. ISSN: 0960-8524. DOI: 10.1016/j.biortech.2020.124073. URL: <https://www.sciencedirect.com/science/article/pii/S0960852420313456> (visited on 04/15/2025).
- (Dec. 2020b). "Bioprospecting of marine diatoms *Thalassiosira*, *Skeletonema* and *Chaetoceros* for lipids and other value-added products". en. In: *Bioresource Technology* 318, p. 124073. ISSN: 09608524. DOI: 10.1016/j.biortech.2020.124073. URL: <https://linkinghub.elsevier.com/retrieve/pii/S0960852420313456> (visited on 05/15/2025).

- Bohren, C. F. and D. R. Huffman (2004). *Absorption and Scattering of Light by Small Particles*. Weinheim: Wiley-VCH. 530 pp. ISBN: 978-0-471-05772-7 978-0-471-29340-8.
- Boss, E., N. Haëntjens, S. Ackleson, B. Balch, A. Chase, G. Dall’Olmo, S. Freeman, L. Yangyang, J. Loftin, W. Neary, N. Nelson, M. Novak, W. Slade, C. Proctor, P. Tortell, and T. Westberry (2019). *Ocean Optics & Biogeochemistry Protocols for Satellite Ocean Colour Sensor Validation, Volume 4.0. Inherent Optical Property Measurements and Protocols: Best Practices for the Collection and Processing of Ship-Based Underway Flow-Through Optical Data (V. 4.0)*. en. Report. Accepted: 2019-11-12T15:50:49Z. International Ocean Colour Coordinating Group (IOCCG). DOI: 10.25607/OBP-664. URL: <https://repository.oceanbestpractices.org/handle/11329/1144> (visited on 05/17/2025).
- Boss, E., W. S. Pegau, W. D. Gardner, J. R. V. Zaneveld, A. H. Barnard, M. S. Twardowski, G. C. Chang, and T. D. Dickey (2001). “Spectral particulate attenuation and particle size distribution in the bottom boundary layer of a continental shelf”. en. In: *Journal of Geophysical Research: Oceans* 106.C5. eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1029/2000JC900077>, pp. 9509–9516. ISSN: 2156-2202. DOI: 10.1029/2000JC900077. URL: <https://onlinelibrary.wiley.com/doi/abs/10.1029/2000JC900077> (visited on 10/13/2023).
- Boss, E., D. Stramski, T. Bergmann, W. Pegau, and M. Lewis (June 2004). “Why Should We Measure the Optical Backscattering Coefficient?” In: *Oceanography (Washington D.C.)* 17. DOI: 10.5670/oceanog.2004.46.
- Boyd, C. E., A. A. McNevin, and R. P. Davis (June 2022). “The contribution of fisheries and aquaculture to the global protein supply”. en. In: *Food Security* 14.3, pp. 805–827. ISSN: 1876-4525. DOI: 10.1007/s12571-021-01246-9. URL: <https://doi.org/10.1007/s12571-021-01246-9> (visited on 05/15/2025).
- Bricaud, A., A. Morel, and L. Prieur (1983). “Optical efficiency factors of some phytoplankters<sup>1</sup>”. en. In: *Limnology and Oceanography* 28.5. eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.4319/lo.1983.28.5.0816>, pp. 816–832. ISSN: 1939-5590. DOI: 10.4319/lo.1983.28.5.0816. URL: <https://onlinelibrary.wiley.com/doi/abs/10.4319/lo.1983.28.5.0816> (visited on 06/14/2024).

- Carvalho Júnior, O. de and P. Meneses (Jan. 2000). “Spectral Correlation Mapper (SCM): An Improvement on the Spectral Angle Mapper (SAM)”. In.
- Castagna, A., H. Dierssen, E. Organelli, M. Bogorad, J. Mortelmans, W. Vyverman, and K. Sabbe (Dec. 2021). “Optical Detection of Harmful Algal Blooms in the Belgian Coastal Zone: A Cautionary Tale of Chlorophyll c3”. English. In: *Frontiers in Marine Science* 8. Publisher: Frontiers. ISSN: 2296-7745. DOI: 10.3389/fmars.2021.770340. URL: <https://www.frontiersin.org/journals/marine-science/articles/10.3389/fmars.2021.770340/full> (visited on 10/30/2024).
- CDFW and ODFW (2014). “PACIFIC STATES MARINE FISHERIES COMMISSION DUNGENESS CRAB REPORT 2014”. In.
- Chavez, F. P. and M. Messié (Dec. 2009). “A Comparison of Eastern Boundary Upwelling Ecosystems”. In: *Progress in Oceanography*. Eastern Boundary Upwelling Ecosystems: Integrative and Comparative Approaches 83.1, pp. 80–96. ISSN: 0079-6611. DOI: 10.1016/j.pocean.2009.07.032. (Visited on 10/05/2025).
- Chavez, F. P., M. Messié, and J. T. Pennington (2011). “Marine Primary Production in Relation to Climate Variability and Change”. In: *Annual Review of Marine Science* 3.1. eprint: <https://doi.org/10.1146/annurev.marine.010908.163917>, pp. 227–260. DOI: 10.1146/annurev.marine.010908.163917. URL: <https://doi.org/10.1146/annurev.marine.010908.163917> (visited on 10/12/2023).
- Ciotti, Á. M., M. R. Lewis, and J. J. Cullen (2002). “Assessment of the relationships between dominant cell size in natural phytoplankton communities and the spectral shape of the absorption coefficient”. en. In: *Limnology and Oceanography* 47.2, pp. 404–417. ISSN: 1939-5590. DOI: 10.4319/lo.2002.47.2.0404. URL: <https://onlinelibrary.wiley.com/doi/abs/10.4319/lo.2002.47.2.0404> (visited on 03/14/2024).
- Clavano, W. R., E. Boss, and L. Karp-Boss (2007). “Inherent Optical Properties of Non-Spherical Marine-like Particles — From Theory to Observation”. en. In.
- Costalago, D., I. Forster, N. Nemcek, C. Neville, R. I. Perry, K. Young, and B. P. V. Hunt (May 2020). “Seasonal and spatial dynamics of the planktonic trophic biomarkers in the Strait of Georgia (northeast Pacific) and implications for fish”. en. In: *Scientific Reports* 10.1. Publisher: Nature Publishing Group, p. 8517. ISSN: 2045-2322. DOI: 10.1038/s41598-020-

- 65557-1. URL: <https://www.nature.com/articles/s41598-020-65557-1> (visited on 04/15/2025).
- Deborah, H. (2016). “Hilda Deborah - Towards spectral mathematical morphology - UPthèses - Les thèses en ligne de l’Université de Poitiers”. Doctoral Dissertation. Norwegian Institute of Science and Technology. URL: <http://theses.univ-poitiers.fr/notice/view/57878> (visited on 05/17/2025).
- Deborah, H., N. Richard, and J. Y. Hardeberg (June 2015). “A Comprehensive Evaluation of Spectral Distance Functions and Metrics for Hyperspectral Image Processing”. en. In: *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 8.6, pp. 3224–3234. ISSN: 1939-1404, 2151-1535. DOI: 10.1109/JSTARS.2015.2403257. URL: <https://ieeexplore.ieee.org/document/7061924/> (visited on 11/20/2024).
- Dijkman, N. and J. Kromkamp (Nov. 1, 2006). “Photosynthetic Characteristics of the Phytoplankton in the Scheldt Estuary: Community and Single-Cell Fluorescence Measurements”. In: *European Journal of Phycology* 41, pp. 425–434. DOI: 10.1080/09670260600937791.
- Douay, F., C. Verpoorter, G. Duong, N. Spilmont, and F. Gevaert (Jan. 2022). “New Hyperspectral Procedure to Discriminate Intertidal Macroalgae”. en. In: *Remote Sensing* 14.2. Number: 2 Publisher: Multidisciplinary Digital Publishing Institute, p. 346. ISSN: 2072-4292. DOI: 10.3390/rs14020346. URL: <https://www.mdpi.com/2072-4292/14/2/346> (visited on 04/23/2025).
- Du, X., W. Peterson, and L. O’Higgins (Jan. 2015). “Interannual variations in phytoplankton community structure in the northern California Current during the upwelling seasons of 2001-2010”. In: *Marine Ecology Progress Series* 519, pp. 75–87. ISSN: 0171-8630, 1616-1599. DOI: 10.3354/meps11097. URL: <https://www.int-res.com/abstracts/meps/v519/p75-87/> (visited on 05/03/2024).
- Duyens, L. N. M. (Jan. 1956). “The flattering of the absorption spectrum of suspensions, as compared to that of solutions”. In: *Biochimica et Biophysica Acta* 19, pp. 1–12. ISSN: 0006-3002. DOI: 10.1016/0006-3002(56)90380-8. URL: <https://www.sciencedirect.com/science/article/pii/0006300256903808> (visited on 04/25/2024).
- Fourtanier, E. and P. Kociolek (Nov. 1, 2003). “Catalogue of the Diatom Genera (Vol 14, Pg 190, 1999)”. In: *Diatom Research* 18, pp. 245–258.

- Frolov, S., R. M. Kudela, and J. G. Bellingham (Jan. 2013). “Monitoring of harmful algal blooms in the era of diminishing resources: A case study of the U.S. West Coast”. In: *Harmful Algae* 21-22, pp. 1–12. ISSN: 1568-9883. DOI: 10.1016/j.hal.2012.11.001. URL: <https://www.sciencedirect.com/science/article/pii/S1568988312001503> (visited on 02/23/2025).
- Gibble, C., R. Duerr, B. Bodenstein, K. Lindquist, J. Lindsey, J. Beck, L. Henkel, J. Roletto, J. Harvey, and R. Kudela (Mar. 2018). “Investigation of a Largescale Common Murre (*Uria aalge*) Mortality Event in California, USA, in 2015”. In: *Journal of Wildlife Diseases* 54.3. Publisher: Wildlife Disease Association, pp. 569–574. ISSN: 0090-3558, 1943-3700. DOI: 10.7589/2017-07-179. URL: <https://bioone.org/journals/journal-of-wildlife-diseases/volume-54/issue-3/2017-07-179/Investigation-of-a-Largescale-Common-Murre-Uria-aalge-Mortality-Event/10.7589/2017-07-179.full> (visited on 04/15/2025).
- Gittings, J. A., D. E. Raitsos, R. J. W. Brewin, and I. Hoteit (Jan. 2021). “Links between Phenology of Large Phytoplankton and Fisheries in the Northern and Central Red Sea”. en. In: *Remote Sensing* 13.2. Number: 2 Publisher: Multidisciplinary Digital Publishing Institute, p. 231. ISSN: 2072-4292. DOI: 10.3390/rs13020231. URL: <https://www.mdpi.com/2072-4292/13/2/231> (visited on 02/12/2025).
- Gordon, H. R., O. B. Brown, R. H. Evans, J. W. Brown, R. C. Smith, K. S. Baker, and D. K. Clark (1988). “A semianalytic radiance model of ocean color”. en. In: *Journal of Geophysical Research: Atmospheres* 93.D9. eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1029/JD093iD09p10909>, pp. 10909–10924. ISSN: 2156-2202. DOI: 10.1029/JD093iD09p10909. URL: <https://onlinelibrary.wiley.com/doi/abs/10.1029/JD093iD09p10909> (visited on 02/28/2024).
- Grattan, L. M., L. Kaddis, J. K. Tracy, and J. G. Morris (Jan. 2021). “Long Term Memory Outcome of Repetitive, Low-Level Dietary Exposure to Domoic Acid in Native Americans”. In: *International Journal of Environmental Research and Public Health* 18.8, p. 3955. ISSN: 1660-4601. DOI: 10.3390/ijerph18083955. URL: <https://www.mdpi.com/1660-4601/18/8/3955> (visited on 09/01/2025).
- Guillard, R. R. L. and P. E. Hargraves (May 1993). “Stichochrysis immobilis is a diatom, not a chrysophyte”. In: *Phycologia* 32.3. Publisher: Taylor &

- Francis \_eprint: <https://doi.org/10.2216/i0031-8884-32-3-234.1>, pp. 234–236. ISSN: 0031-8884. DOI: 10.2216/i0031-8884-32-3-234.1. URL: <https://doi.org/10.2216/i0031-8884-32-3-234.1> (visited on 05/17/2025).
- Haëntjens, N. and E. Boss (Mar. 2020). “Inlinino: A Modular Software Data Logger for Oceanography”. en. In: *Oceanography* 33.1, pp. 80–84. ISSN: 10428275. DOI: 10.5670/oceanog.2020.112. URL: <https://tos.org/oceanography/article/inlinino-a-modular-software-data-logger-for-oceanography> (visited on 11/21/2023).
- Horner, R. A., D. L. Garrison, and F. G. Plumley (July 1997). “Harmful algal blooms and red tide problems on the U.S. west coast”. en. In: *Limnology and Oceanography* 42.5part2, pp. 1076–1088. ISSN: 0024-3590, 1939-5590. DOI: 10.4319/lo.1997.42.5\_part\_2.1076. URL: [https://onlinelibrary.wiley.com/doi/10.4319/lo.1997.42.5\\_part\\_2.1076](https://onlinelibrary.wiley.com/doi/10.4319/lo.1997.42.5_part_2.1076) (visited on 05/15/2025).
- Hunt, M. (May 2024). “Using the Phytoplankton Light Scattering (PPLS) Prototype to Investigate Hyperspectral Backscattering of Dinoflagellate Species”. MA thesis. University of Gorgia.
- Jin, X., N. Gruber, J. P. Dunne, J. L. Sarmiento, and R. A. Armstrong (2006). “Diagnosing the contribution of phytoplankton functional groups to the production and export of particulate organic carbon, CaCO<sub>3</sub>, and opal from global nutrient and alkalinity distributions”. en. In: *Global Biogeochemical Cycles* 20.2. \_eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1029/2005GB002532>. ISSN: 1944-9224. DOI: 10.1029/2005GB002532. URL: <https://onlinelibrary.wiley.com/doi/abs/10.1029/2005GB002532> (visited on 05/17/2025).
- Karp-Boss, L. and P. A. Jumars (1998). “Motion of diatom chains in steady shear flow”. en. In: *Limnology and Oceanography* 43.8. \_eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.4319/lo.1998.43.8.1767>, pp. 1767–1773. ISSN: 1939-5590. DOI: 10.4319/lo.1998.43.8.1767. URL: <https://onlinelibrary.wiley.com/doi/abs/10.4319/lo.1998.43.8.1767> (visited on 10/20/2023).
- Kavanaugh, M. T., T. Bell, D. Catlett, M. A. Cimino, and S. C. Doney (Nov. 2021). “Satellite Remote Sensing and the Marine Biodiversity Observation Network: Current Science and Future Steps”. In: *Oceanography* 34.2, pp. 62–79. DOI: 10.5670/oceanog.2021.215. URL: <https://tos.org/oceanography/article/satellite-remote-sensing->

- and - the - marine - biodiversity - observation - network - current - science - and - future - steps (visited on 01/14/2025).
- Kirk, J. T. O. (2010). *Light and Photosynthesis in Aquatic Ecosystems*. 3rd ed. Cambridge: Cambridge University Press. ISBN: 978-0-521-15175-7. DOI: 10.1017/CB09781139168212. URL: <https://www.cambridge.org/core/books/light-and-photosynthesis-in-aquatic-ecosystems/0FA122C5EE317E4B27BCE4F63B304999> (visited on 05/17/2025).
- Kooistra, W. H., R. Gersonde, L. K. Medlin, and D. G. Mann (2007). “The Origin and Evolution of the Diatoms: Their Adaptation to a Planktonic Existence”. en. In: *Evolution of Primary Producers in the Sea*. Elsevier, pp. 207–249. ISBN: 978-0-12-370518-1. DOI: 10.1016/B978-012370518-1/50012-6. URL: <https://linkinghub.elsevier.com/retrieve/pii/B9780123705181500126> (visited on 04/14/2025).
- Kostadinov, T. S., D. A. Siegel, and S. Maritorena (2009). “Retrieval of the particle size distribution from satellite ocean color observations”. en. In: *Journal of Geophysical Research: Oceans* 114.C9. eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1029/2009JC005303>. ISSN: 2156-2202. DOI: 10.1029/2009JC005303. URL: <https://onlinelibrary.wiley.com/doi/abs/10.1029/2009JC005303> (visited on 10/09/2023).
- Kruse, F. A., A. B. Lefkoff, J. W. Boardman, K. B. Heidebrecht, A. T. Shapiro, P. J. Barloon, and A. F. H. Goetz (May 1993). “The spectral image processing system (SIPS)—interactive visualization and analysis of imaging spectrometer data”. In: *Remote Sensing of Environment*. Airbone Imaging Spectrometry 44.2, pp. 145–163. ISSN: 0034-4257. DOI: 10.1016/0034-4257(93)90013-N. URL: <https://www.sciencedirect.com/science/article/pii/003442579390013N> (visited on 12/10/2024).
- Kudela, R. M., S. L. Palacios, D. C. Austerberry, E. K. Accorsi, L. S. Guild, and J. Torres-Perez (Sept. 15, 2015). “Application of Hyperspectral Remote Sensing to Cyanobacterial Blooms in Inland Waters”. In: *Remote Sensing of Environment*. Special Issue on the Hyperspectral Infrared Imager (HyspIRI) 167, pp. 196–205. ISSN: 0034-4257. DOI: 10.1016/j.rse.2015.01.025. URL: <https://www.sciencedirect.com/science/article/pii/S0034425715000437> (visited on 08/22/2025).
- Lassiter, A. M., F. P. Wilkerson, R. C. Dugdale, and V. E. Hogue (Dec. 2006). “Phytoplankton assemblages in the CoOP-WEST coastal upwelling area”. In: *Deep Sea Research Part II: Topical Studies in Oceanography*. The Role of Wind-Driven Flow in Shelf Productivity 53.25,

- pp. 3063–3077. ISSN: 0967-0645. DOI: 10.1016/j.dsr2.2006.07.013. URL: <https://www.sciencedirect.com/science/article/pii/S0967064506002256> (visited on 05/17/2025).
- Lee, Z., K. L. Carder, and R. A. Arnone (Sept. 2002). “Deriving inherent optical properties from water color: a multiband quasi-analytical algorithm for optically deep waters”. en. In: *Applied Optics* 41.27, p. 5755. ISSN: 0003-6935, 1539-4522. DOI: 10.1364/AO.41.005755. URL: <https://opg.optica.org/abstract.cfm?URI=ao-41-27-5755> (visited on 05/17/2025).
- Lee, Z., B. Lubac, J. Werdell, and R. Arnone (2009). “An Update of the Quasi-Analytical Algorithm (QAA\_v5)”. en. In.
- Lefebvre, K. A., P. S. Kendrick, W. Ladiges, E. M. Hiolski, B. E. Ferriss, D. R. Smith, and D. J. Marcinek (Apr. 2017). “Chronic low-level exposure to the common seafood toxin domoic acid causes cognitive deficits in mice”. In: *Harmful algae* 64, pp. 20–29. ISSN: 1568-9883. DOI: 10.1016/j.hal.2017.03.003. URL: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5548283/> (visited on 04/15/2025).
- Lomas, M. W., A. R. Neeley, R. Vandermeulen, A. Mannino, C. Thomas, M. G. Novak, and S. A. Freeman (Feb. 2024). “Phytoplankton optical fingerprint libraries for development of phytoplankton ocean color satellite products”. en. In: *Scientific Data* 11.1, p. 168. ISSN: 2052-4463. DOI: 10.1038/s41597-024-03001-z. URL: <https://www.nature.com/articles/s41597-024-03001-z> (visited on 02/18/2024).
- Marcos, J. R. Seymour, M. Luvar, W. M. Durham, J. G. Mitchell, A. Macke, and R. Stocker (Mar. 2011). “Microbial alignment in flow changes ocean light climate”. en. In: *Proceedings of the National Academy of Sciences* 108.10, pp. 3860–3864. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.1014576108. URL: <https://pnas.org/doi/full/10.1073/pnas.1014576108> (visited on 01/27/2025).
- McCabe, R. M., B. M. Hickey, R. M. Kudela, K. A. Lefebvre, N. G. Adams, B. D. Bill, F. M. D. Gulland, R. E. Thomson, W. P. Cochlan, and V. L. Trainer (2016). “An unprecedented coastwide toxic algal bloom linked to anomalous ocean conditions”. en. In: *Geophysical Research Letters* 43.19. eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/2016GL070023>, pp. 10, 366–10, 376. ISSN: 1944-8007. DOI: 10.1002/2016GL070023. URL: <https://onlinelibrary.wiley.com/doi/abs/10.1002/2016GL070023> (visited on 06/19/2024).

- Mie, G. (1908). “Beiträge zur Optik trüber Medien, speziell kolloidaler Metallösungen”. en. In: *Annalen der Physik* 330.3. eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/andp.19083300302>, pp. 377–445. ISSN: 1521-3889. DOI: 10.1002/andp.19083300302. URL: <https://onlinelibrary.wiley.com/doi/abs/10.1002/andp.19083300302> (visited on 02/11/2026).
- Millie, D., O. Schofield, G. Kirkpatrick, G. Johnson, P. Tester, and B. Vintard (Jan. 1997). “Detection of Harmful Algal Blooms Using Photopigments and Absorption Signatures: A Case Study of the Florida Red Tide Dinoflagellate, *Gymnodinium breve*”. In: *Limnology and Oceanography* 42.5. DOI: 10.4319/lo.1997.42.5\_part\_2.1240. URL: [https://digitalcommons.odu.edu/oeas\\_fac\\_pubs/348](https://digitalcommons.odu.edu/oeas_fac_pubs/348).
- Milligan, A. J., U. A. Aparicio, and M. Behrenfeld (Feb. 2012). “Fluorescence and nonphotochemical quenching responses to simulated vertical mixing in the marine diatom *Thalassiosira weissflogii*”. en. In: *Marine Ecology Progress Series* 448, pp. 67–78. ISSN: 0171-8630, 1616-1599. DOI: 10.3354/meps09544. URL: <https://www.int-res.com/abstracts/meps/v448/p67-78/> (visited on 05/17/2025).
- Mobley, C., Z. Ahmad, D. Antoine, S. Bailey, P. Bissett, E. Boss, D. Bowers, I. Cetinić, B. Franz, H. Gordon, N. Haëntjens, V. Hill, C. D. Mobley, B. Monosmith, C. Roesler, K. Voss, J. Werdell, X. Zhang, and R. Zimmerman (2022). *The Ocean Optics Book*. Sequoia Scientific, Inc. URL: <http://www.oceanopticsbook.info/>.
- Moore, J. K., S. C. Doney, D. M. Glover, and I. Y. Fung (Jan. 2001a). “Iron cycling and nutrient-limitation patterns in surface waters of the World Ocean”. In: *Deep Sea Research Part II: Topical Studies in Oceanography*. The US JGOFS Synthesis and Modeling Project: Phase 1 49.1, pp. 463–507. ISSN: 0967-0645. DOI: 10.1016/S0967-0645(01)00109-6. URL: <https://www.sciencedirect.com/science/article/pii/S0967064501001096> (visited on 05/15/2025).
- Moore, J. K., S. C. Doney, J. A. Kleypas, D. M. Glover, and I. Y. Fung (Jan. 2001b). “An intermediate complexity marine ecosystem model for the global domain”. In: *Deep Sea Research Part II: Topical Studies in Oceanography*. The US JGOFS Synthesis and Modeling Project: Phase 1 49.1, pp. 403–462. ISSN: 0967-0645. DOI: 10.1016/S0967-0645(01)00108-4. URL: <https://www.sciencedirect.com/science/article/pii/S0967064501001084> (visited on 05/15/2025).

- Moore, S. K., S. J. Dreyer, J. A. Ekstrom, K. Moore, K. Norman, T. Klinger, E. H. Allison, and S. L. Jardine (June 1, 2020). “Harmful Algal Blooms and Coastal Communities: Socioeconomic Impacts and Actions Taken to Cope with the 2015 U.S. West Coast Domoic Acid Event”. In: *Harmful Algae* 96, p. 101799. ISSN: 1568-9883. DOI: 10.1016/j.hal.2020.101799. URL: <https://www.sciencedirect.com/science/article/pii/S1568988320300792> (visited on 07/25/2025).
- Morel, A. and A. Bricaud (Jan. 1981). “Theoretical Results Concerning the Optics of Phytoplankton, with Special Reference to Remote Sensing Applications”. In: pp. 313–327. ISBN: 978-1-4613-3317-3. DOI: 10.1007/978-1-4613-3315-9\_35.
- Morel, A. and L. Prieur (1977). “Analysis of variations in ocean color”. en. In: *Limnology and Oceanography* 22.4. eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.4319/lo.1977.22.4.0709>, pp. 709–722. ISSN: 1939-5590. DOI: 10.4319/lo.1977.22.4.0709. URL: <https://onlinelibrary.wiley.com/doi/abs/10.4319/lo.1977.22.4.0709> (visited on 05/17/2025).
- Munk, W. and G. Riley (Jan. 1952). “Absorption of nutrients by aquatic plants”. In: *Journal of Marine Research* 11.2. URL: [https://elischolar.library.yale.edu/journal\\_of\\_marine\\_research/772](https://elischolar.library.yale.edu/journal_of_marine_research/772).
- Nayak, A. R., M. N. McFarland, J. M. Sullivan, and M. S. Twardowski (Jan. 2018). “Evidence for ubiquitous preferential particle orientation in representative oceanic shear flows”. en. In: *Limnology and Oceanography* 63.1, pp. 122–143. ISSN: 0024-3590, 1939-5590. DOI: 10.1002/lno.10618. URL: <https://aslopubs.onlinelibrary.wiley.com/doi/10.1002/lno.10618> (visited on 12/10/2024).
- Novak, M. G., H. Burmeister, and R. Röttgers (June 2024). “Hyperspectral measurements of light backscattering by particles in water with a fixed angle setup: proof of concept and instrument calibration”. EN. In: *Optics Express* 32.13. Publisher: Optica Publishing Group, pp. 23722–23735. ISSN: 1094-4087. DOI: 10.1364/OE.529061. URL: <https://opg.optica.org/oe/abstract.cfm?uri=oe-32-13-23722> (visited on 05/17/2025).
- Peterson, W. T., J. L. Fisher, P. T. Strub, X. Du, C. Risien, J. Peterson, and C. T. Shaw (2017). “The pelagic ecosystem in the Northern California Current off Oregon during the 2014–2016 warm anomalies within the context of the past 20 years”. en. In: *Journal of Geophysical Research: Oceans* 122.9. eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/2017JC012952>,

- pp. 7267–7290. ISSN: 2169-9291. DOI: 10.1002/2017JC012952. URL: <https://onlinelibrary.wiley.com/doi/abs/10.1002/2017JC012952> (visited on 04/15/2025).
- Pope, R. M. and E. S. Fry (Nov. 1997). “Absorption spectrum (380–700 nm) of pure water. II. Integrating cavity measurements”. EN. In: *Applied Optics* 36.33, pp. 8710–8723. ISSN: 2155-3165. DOI: 10.1364/AO.36.008710. URL: <https://opg.optica.org/ao/abstract.cfm?uri=ao-36-33-8710> (visited on 02/11/2026).
- Prasad, A. K. S. K., J. A. Nienow, and E. Lochner (Mar. 2018). “Thalassiosira mala (Bacillariophyta), a potentially harmful, marine diatom from Chilka Lake and other coastal localities of Odisha, India: Nomenclature, frustule morphology and global biogeography”. en. In: *Journal of Biosciences* 43.1, pp. 59–74. ISSN: 0973-7138. DOI: 10.1007/s12038-018-9730-0. URL: <https://doi.org/10.1007/s12038-018-9730-0> (visited on 04/15/2025).
- Ryther, J. H. (1969). “Photosynthesis and Fish Production in the Sea”. In: *Science, New Series* 166.3901, pp. 72–76. JSTOR: 1727735.
- Silsbe, G. M., M. J. Behrenfeld, K. H. Halsey, A. J. Milligan, and T. K. Westberry (2016). “The CAFE model: A net production model for global ocean phytoplankton”. en. In: *Global Biogeochemical Cycles* 30.12. eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/2016GB005521>, pp. 1756–1777. ISSN: 1944-9224. DOI: 10.1002/2016GB005521. URL: <https://onlinelibrary.wiley.com/doi/abs/10.1002/2016GB005521> (visited on 03/18/2025).
- Slade, W. H., E. Boss, G. Dall’Olmo, M. R. Langner, J. Loftin, M. J. Behrenfeld, C. Roesler, and T. K. Westberry (Oct. 2010). “Underway and Moored Methods for Improving Accuracy in Measurement of Spectral Particulate Absorption and Attenuation”. EN. In: *Journal of Atmospheric and Oceanic Technology* 27.10. Publisher: American Meteorological Society Section: Journal of Atmospheric and Oceanic Technology, pp. 1733–1746. ISSN: 0739-0572, 1520-0426. DOI: 10.1175/2010JTECH0755.1. URL: [https://journals.ametsoc.org/view/journals/atot/27/10/2010jtecho755\\_1.xml](https://journals.ametsoc.org/view/journals/atot/27/10/2010jtecho755_1.xml) (visited on 04/24/2025).
- Smith, R. C. and K. S. Baker (Jan. 15, 1981). “Optical Properties of the Clearest Natural Waters (200–800 Nm)”. In: *Applied Optics* 20.2, pp. 177–184. ISSN: 2155-3165. DOI: 10.1364/AO.20.000177. URL: <https://opg.optica.org/ao/abstract.cfm?uri=ao-20-2-177> (visited on 05/13/2025).

- Smith, T. and J. Guild (Jan. 1931). “The C.I.E. colorimetric standards and their use”. en. In: *Transactions of the Optical Society* 33.3, p. 73. ISSN: 1475-4878. DOI: 10.1088/1475-4878/33/3/301. URL: <https://dx.doi.org/10.1088/1475-4878/33/3/301> (visited on 04/15/2025).
- Stock, C. A., J. G. John, R. R. Rykaczewski, R. G. Asch, W. W. L. Cheung, J. P. Dunne, K. D. Friedland, V. W. Y. Lam, J. L. Sarmiento, and R. A. Watson (Feb. 2017). “Reconciling fisheries catch and ocean productivity”. In: *Proceedings of the National Academy of Sciences* 114.8. Publisher: Proceedings of the National Academy of Sciences, E1441–E1449. DOI: 10.1073/pnas.1610238114. URL: <https://www.pnas.org/doi/full/10.1073/pnas.1610238114> (visited on 04/15/2025).
- Tan, L., W. Xu, X. He, and J. Wang (Nov. 30, 2019). “The Feasibility of Fv/Fm on Judging Nutrient Limitation of Marine Algae through Indoor Simulation and *in Situ* Experiment”. In: *Estuarine, Coastal and Shelf Science* 229, p. 106411. ISSN: 0272-7714. DOI: 10.1016/j.ecss.2019.106411. URL: <https://www.sciencedirect.com/science/article/pii/S0272771419307309> (visited on 08/22/2025).
- Trainer, V. L., M. L. Wells, W. P. Cochlan, C. G. Trick, B. D. Bill, K. A. Baugh, B. F. Beall, J. Herndon, and N. Lundholm (2009). “An Ecological Study of a Massive Bloom of Toxigenic *Pseudo-nitzschia Cuspidata* off the Washington State Coast”. In: *Limnology and Oceanography* 54.5, pp. 1461–1474. ISSN: 1939-5590. DOI: 10.4319/lo.2009.54.5.1461. URL: <https://onlinelibrary.wiley.com/doi/abs/10.4319/lo.2009.54.5.1461> (visited on 08/28/2025).
- Tran, Q. D., T. R. Neu, S. Sultana, H.-A. Giebel, M. Simon, and S. Billerbeck (2023). “Distinct glycoconjugate cell surface structures make the pelagic diatom *Thalassiosira rotula* an attractive habitat for bacteria”. en. In: *Journal of Phycology* 59.2. eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1111/jpy.13308>, pp. 309–322. ISSN: 1529-8817. DOI: 10.1111/jpy.13308. URL: <https://onlinelibrary.wiley.com/doi/abs/10.1111/jpy.13308> (visited on 03/03/2025).
- Tréguer, P., C. Bowler, B. Moriceau, S. Dutkiewicz, M. Gehlen, O. Aumont, L. Bittner, R. Dugdale, Z. Finkel, D. Iudicone, O. Jahn, L. Guidi, M. Lasbleiz, K. Leblanc, M. Levy, and P. Pondaven (Jan. 2018). “Influence of diatom diversity on the ocean biological carbon pump”. en. In: *Nature Geoscience* 11.1. Publisher: Nature Publishing Group, pp. 27–37. ISSN:

- 1752-0908. DOI: 10.1038/s41561-017-0028-x. URL: <https://www.nature.com/articles/s41561-017-0028-x> (visited on 05/17/2025).
- Tréguer, P., D. M. Nelson, A. J. Van Bennekom, D. J. DeMaster, A. Leynaert, and B. Quéguiner (Apr. 1995). “The Silica Balance in the World Ocean: A Reestimate”. In: *Science* 268.5209. Publisher: American Association for the Advancement of Science, pp. 375–379. DOI: 10.1126/science.268.5209.375. URL: <https://www.science.org/doi/10.1126/science.268.5209.375> (visited on 05/15/2025).
- Van de Hulst, H. C. (June 2012). *Light Scattering by Small Particles*. en. Google-Books-ID: 777DAgAAQBAJ. Courier Corporation. ISBN: 978-0-486-13975-3.
- Vandermeulen, R. A., A. Mannino, A. Neeley, J. Werdell, and R. Arnone (Aug. 2017). “Determining the optimal spectral sampling frequency and uncertainty thresholds for hyperspectral remote sensing of ocean color”. EN. In: *Optics Express* 25.16. Publisher: Optica Publishing Group, A785–A797. ISSN: 1094-4087. DOI: 10.1364/OE.25.00A785. URL: <https://opg.optica.org/oe/abstract.cfm?uri=oe-25-16-A785> (visited on 02/27/2024).
- Waite, A., A. Fisher, P. Thompson, and P. Harrison (Oct. 1997). “Sinking rate vs volume relationships illuminate sinking control mechanisms in marine diatoms”. In: *Marine Ecology-progress Series - MAR ECOL-PROGR SER* 157, pp. 97–108. DOI: 10.3354/meps157097.
- Wei, J., M. Wang, K. Mikelsons, L. Jiang, S. Kratzer, Z. Lee, T. Moore, H. M. Sosik, and D. Van Der Zande (Dec. 2022). “Global satellite water classification data products over oceanic, coastal, and inland waters”. en. In: *Remote Sensing of Environment* 282, p. 113233. ISSN: 00344257. DOI: 10.1016/j.rse.2022.113233. URL: <https://linkinghub.elsevier.com/retrieve/pii/S003442572200339X> (visited on 08/03/2023).
- Werdell, P. J., M. J. Behrenfeld, P. S. Bontempi, E. Boss, B. Cairns, G. T. Davis, B. A. Franz, U. B. Gliese, E. T. Gorman, O. Hasekamp, K. D. Knobelspiesse, A. Mannino, J. V. Martins, C. R. McClain, G. Meister, and L. A. Remer (Sept. 2019). “The Plankton, Aerosol, Cloud, Ocean Ecosystem Mission: Status, Science, Advances”. In: *Bulletin of the American Meteorological Society* 100.9, pp. 1775–1794. ISSN: 0003-0007, 1520-0477. DOI: 10.1175/BAMS-D-18-0056.1. URL: <https://journals.ametsoc.org/view/journals/bams/100/9/bams-d-18-0056.1.xml> (visited on 08/23/2025).

- Westberry, T., M. Behrenfeld, D. A. Siegel, and E. Boss (2008). “Carbon-based primary productivity modeling with vertically resolved photoacclimation”. en. In: *Global Biogeochemical Cycles* 22.2. eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1029/2007GB003078>. ISSN: 1944-9224. DOI: 10.1029/2007GB003078. URL: <https://onlinelibrary.wiley.com/doi/abs/10.1029/2007GB003078> (visited on 05/15/2025).
- Westberry, T. K., G. M. Silsbe, and M. Behrenfeld (Feb. 2023). “Gross and net primary production in the global ocean: An ocean color remote sensing perspective”. en. In: *Earth-Science Reviews* 237, p. 104322. ISSN: 0012-8252. DOI: 10.1016/j.earscirev.2023.104322. URL: <https://www.sciencedirect.com/science/article/pii/S0012825223000119> (visited on 07/28/2023).
- Whitmire, A. L., E. Boss, T. J. Cowles, and W. S. Pegau (2007). “Spectral variability of the particulate backscattering ratio”. en. In: *Optics Express* 15.11, p. 7019. ISSN: 1094-4087. DOI: 10.1364/OE.15.007019. URL: <https://opg.optica.org/oe/abstract.cfm?uri=oe-15-11-7019> (visited on 04/21/2025).
- Whitmire, A. (June 2008). “The Spectral Backscattering Properties of Marine Particles”. Doctoral Dissertation. Oregon State University.
- Whitmire, A. L., W. S. Pegau, L. Karp-Boss, E. Boss, and T. J. Cowles (July 2010). “Spectral backscattering properties of marine phytoplankton cultures”. EN. In: *Optics Express* 18.14. Publisher: Optica Publishing Group, pp. 15073–15093. ISSN: 1094-4087. DOI: 10.1364/OE.18.015073. URL: <https://opg.optica.org/oe/abstract.cfm?uri=oe-18-14-15073> (visited on 02/06/2025).
- Xi, H., M. Hieronymi, R. Röttgers, H. Krasemann, and Z. Qiu (Nov. 2015). “Hyperspectral Differentiation of Phytoplankton Taxonomic Groups: A Comparison between Using Remote Sensing Reflectance and Absorption Spectra”. en. In: *Remote Sensing* 7.11, pp. 14781–14805. ISSN: 2072-4292. DOI: 10.3390/rs71114781. URL: <https://www.mdpi.com/2072-4292/7/11/14781> (visited on 04/05/2024).
- Yang, C. and L. Albright (1992). “Effects of the harmful diatom *Chaetoceros concavicornis* on respiration of rainbow trout *Oncorhynchus mykiss*”. en. In: *Diseases of Aquatic Organisms* 14, pp. 105–114. ISSN: 0177-5103, 1616-1580. DOI: 10.3354/dao014105. URL: <http://www.int-res.com/articles/dao/14/d014p105.pdf> (visited on 05/15/2025).

- Yi, Z., M. Xu, X. Di, S. Brynjolfsson, and W. Fu (Jan. 2017). “Exploring Valuable Lipids in Diatoms”. en. In: *Frontiers in Marine Science* 4. ISSN: 2296-7745. DOI: 10.3389/fmars.2017.00017. URL: <http://journal.frontiersin.org/article/10.3389/fmars.2017.00017/full> (visited on 05/15/2025).
- Zaneveld, J. R. V. (1995). “A theoretical derivation of the dependence of the remotely sensed reflectance of the ocean on the inherent optical properties”. en. In: *Journal of Geophysical Research: Oceans* 100.C7. eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1029/95JC00453>, pp. 13135–13142. ISSN: 2156-2202. DOI: 10.1029/95JC00453. URL: <https://onlinelibrary.wiley.com/doi/abs/10.1029/95JC00453> (visited on 04/17/2025).
- Zhang, X. and L. Hu (Jan. 2021). “Light Scattering by Pure Water and Seawater: Recent Development”. en. In: *Journal of Remote Sensing* 2021, pp. 2021/9753625. ISSN: 2694-1589. DOI: 10.34133/2021/9753625. URL: <https://spj.science.org/doi/10.34133/2021/9753625> (visited on 04/21/2025).
- Zhu, W., Q. Yu, Y. Q. Tian, R. F. Chen, and G. B. Gardner (2011). “Estimation of chromophoric dissolved organic matter in the Mississippi and Atchafalaya river plume regions using above-surface hyperspectral remote sensing”. en. In: *Journal of Geophysical Research: Oceans* 116.C2. eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1029/2010JC006523>. ISSN: 2156-2202. DOI: 10.1029/2010JC006523. URL: <https://onlinelibrary.wiley.com/doi/abs/10.1029/2010JC006523> (visited on 04/24/2025).

## 6 Supplemental Materials

**Additional Distance Metrics** Additional distance metrics explored included: cosine similarity, Euclidean distance of cumulative spectra, and Kullback-Leibler pseudo-divergence.

Cosine similarity is defined as:

$$\text{Cosine Similarity} = \cos(\theta) = \frac{\mathbf{A} \cdot \mathbf{B}}{\|\mathbf{A}\| \|\mathbf{B}\|}$$

$$\mathbf{A} \cdot \mathbf{B} = \sum_{i=1}^n A_i B_i, \quad \|\mathbf{A}\| = \sqrt{\sum_{i=1}^n A_i^2}, \quad \|\mathbf{B}\| = \sqrt{\sum_{i=1}^n B_i^2},$$

where  $A_i$  and  $B_i$  are two distinct spectra. Cosine similarity takes one point of a spectrum (i) and places it in one-dimensional space (a floating dot), it then iterates upon this, placing another point (i+1) and adding a dimension (d+1). We now have two coordinates in 2-dimensional space to which we can draw a line from the origin (i.e. a vector in an x - y plane). The algorithm continues to add coordinates and subsequent dimensions until we have a space with equal dimensions and coordinates as our spectrum (in this case we are evaluating wavelengths 430 - 670 at 5 nm increments (i & d = 48)). For these spectra we create a single vector in 48-dimensional space derived from the coordinates and take the cosine of this vector, spectrum a, to that derived for spectrum b; where spectra a and b represent the ‘fingerprints’ of two unique phytoplankton taxa. When we apply a weighting function to the spectra (to isolate colors with the most inter-taxa variability), we reduce the dimensionality to the number of wavelengths being analyzed. It should be noted that cosine similarity is processed in Euclidean space and therefore assumes different wavelengths are statistically independent from one another. We know of course that this is untrue, as the distinct bands that compose hyperspectral data are highly correlated (440 nm is deep blue, and 441 nm is just slightly less deep blue).

In cosine space when two vectors are orthogonal to each other (at a 90° angle), we infer that they are entirely different by the metrics that cosine similarity measures. In contrast, vectors with a 0° angle indicate identical spectra. Cosine similarity is a non-linear metric, for example, a cosine value of 0.9 corresponds to an  $\arccos(0.9) = 24.8^\circ$  angular difference in n-dimensional space (a 0.9 - 1 = 10% difference in cosine similarity space), while a value of 0.99 would be an  $\arccos(0.99) = 8.1^\circ$  angular difference (0.99 - 1 = 1% difference in cosine similarity space).

Deborah 2016 found cosine similarity to work well with theoretical data but when applied to real spectra, noise contaminated accurate measurements. This issue can be alleviated by a Savitzky-Golay smoothing function. Both cosine similarity and SCM are common in the literature and were designed to ignore magnitude changes—we often only care about spectral shape and not intensity.

While cosine similarity and SCM respond well to overlapping translational changes (wavelength/hue changes), they saturates when these changes become too large (when spectral features no longer overlap). This is not an issue for our comparisons as the center wavelengths of spectral features are well-aligned, largely due to the common pigment classes shared between diatoms. Thus, the dominant parameter in our distance metric is the standard deviation of a spectral features, which cosine similarity and SCM excel at measuring.

The two following methods are sensitive to changes in magnitude and do not saturate with increased distance between features. These are the Euclidean distance of Cumulative Sum (ECS) and the Spectral Kullback-Leibler Pseudo-Divergence (KLPD). Both of these metrics fully satisfy the three fundamental criterion (magnitude, translation, and standard deviation) for a distance metric. One caveat to ECS is that this function integrates the spectra from the lowest to highest values, placing more emphasis on differences in bluer wavelengths. The integration can be flipped to place more emphasis on the red wavelengths, however, the derived differences between spectra will change depending on how the metric is integrated. Kullback-Leibler pseudo-divergence is a variation on the standard Kullback-Leibler divergence function, which is used for probability density functions, hence *pseudo*-divergence. Kullback-Leibler pseudo-divergence violates the theory of triangular inequality which means that we expect different, albeit similar, responses when comparing spectrum i to spectrum j as when comparing spectrum j to spectrum i. However, all spectral features are evenly weighted. For both ECS and KLPD, a value closer to 0 means the spectra are more similar, unlike cosine and SCM where a value closer to 0 means the spectra are more dissimilar.

Euclidean distance of Cumulative Spectra (ECS), processes the spectra as a distribution:

$$d_{\text{ECS}}(S_1, S_2) = \left( \int_{\lambda_{\min}}^{\lambda_{\max}} (s'_1(\lambda) - s'_2(\lambda))^2 d\lambda \right)^{\frac{1}{2}} \quad (4)$$

$$s'_i(\lambda_c) = \int_{\lambda_{\min}}^{\lambda_c} s_i(\lambda) d\lambda, \quad (5)$$

where  $s_1$  and  $s_2$  are two distinct spectra, or ‘fingerprints’ to be compared.

Kullback-Leibler Pseudo-Divergence (KLPD) is similar to ECS, but is not

spectrally weighted:

$$div_{KL}(S_1, S_2) = k_1 \cdot \text{KL}(\bar{S}_1, \bar{S}_2) + k_2 \cdot \text{KL}(\bar{S}_2, \bar{S}_1) + (k_1 - k_2) \log \left( \frac{k_1}{k_2} \right) \quad (6)$$

$$\text{shape} : k_1 \cdot \text{KL}(\bar{S}_1, \bar{S}_2) + k_2 \cdot \text{KL}(\bar{S}_2, \bar{S}_1) \quad (7)$$

$$\text{energy or intensity} : (k_1 - k_2) \log \left( \frac{k_1}{k_2} \right) . \quad (8)$$

Both ECS and KLPD performed well (intertaxon differences were large, while intrataxon differences were small) but are only able to provide relative differences, rather than absolute values like cosine similarity and SCM.

The value in both ECS and KLPD comes from their ability to show the relative differences between fingerprints. Euclidean distance of Cumulative Spectra presents the largest dynamic range with the center, identical values (whited out) at 0 while maximum differences are interpreted for *Thalassiosira* and *Pseudo-nitzschia* around 850. KLPD presents lower dynamic range, reaching a maximum of 59 for *Thalassiosira* and *Pseudo-nitzschia*, but presents the same general patterns. Looking at figure 13 we can see three red hotspots; one in the top left, the center-bottom-right, and nearly the bottom right. These are emphasizing the intrataxon similarities between *Pseudo-nitzschia*, *Chaetoceros*, and *Thalassiosira*.

On the full raw spectra, cosine similarity performed poorly in terms of percentage differences, all spectra were given identical values to the 0.00 level (spectral differences were all  $< 1\%$  in cosine space (heatmap not shown)). However, when we increased the decimal limit, we found that although differences were  $< 1\%$ , these small differences were still sufficient to successfully cluster most of the reflectance fingerprints into their respective taxonomic groups (figure 10). Both ECS and KLPD were performed on the full raw spectra because they are able to measure differences in magnitude and do not saturate with distance for translational changes. Cosine similarity and SCM perform best after a derivative is applied and the spectra is subset to a region of interest.

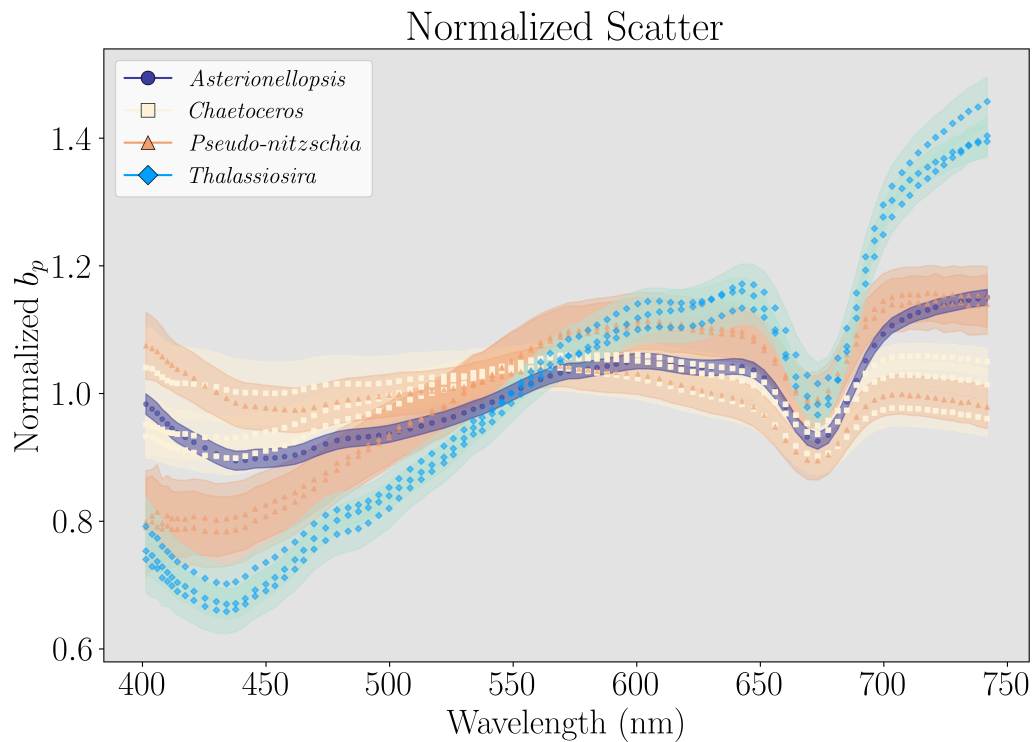


Figure 9: Normalized total scatter spectra for all species.

All species have total scatter spectra that visually appear to be more similar intraspecifically than interspecifically. The standard deviations are low across all species. PN\_3 is different from other triplicates due to a pump failure (the more flat, orange line). These spectra of total scatter closely mimic an inversion of the absorption spectra.

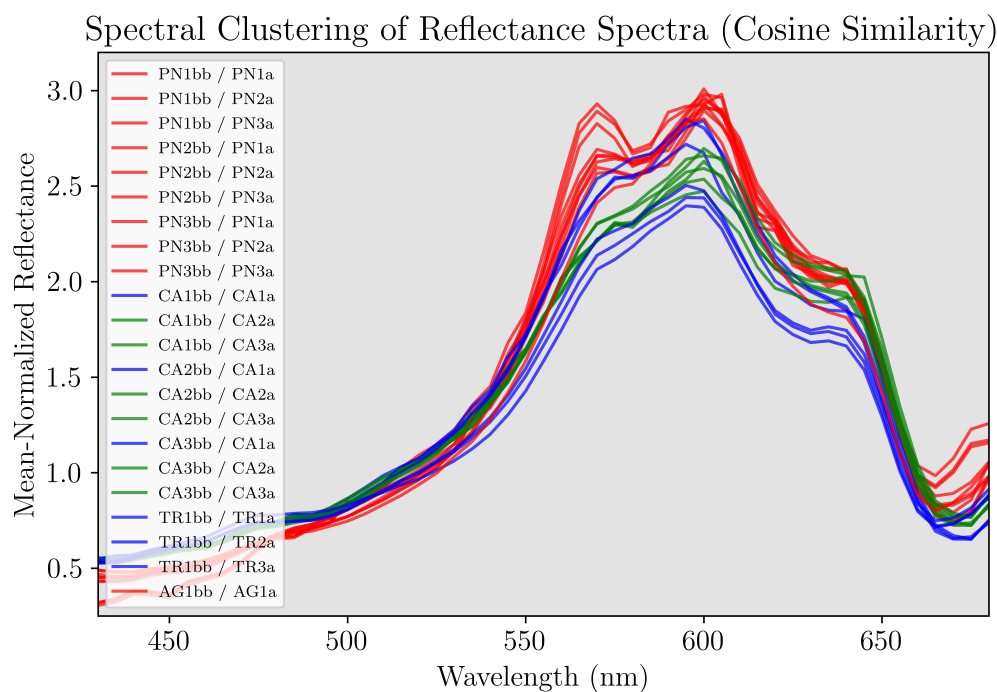


Figure 10: Spectral clustering: unsupervised machine learning using cosine similarity on all raw spectra. Cluster number set to 3. The absorption spectra of the first triplicate of *Chaetoceros* (CA1a) is misclassified as *Thalassiosira* (TR) for all bootstrapped spectra. *Asterionellopsis* is misclassified as *Pseudonitzschia*.

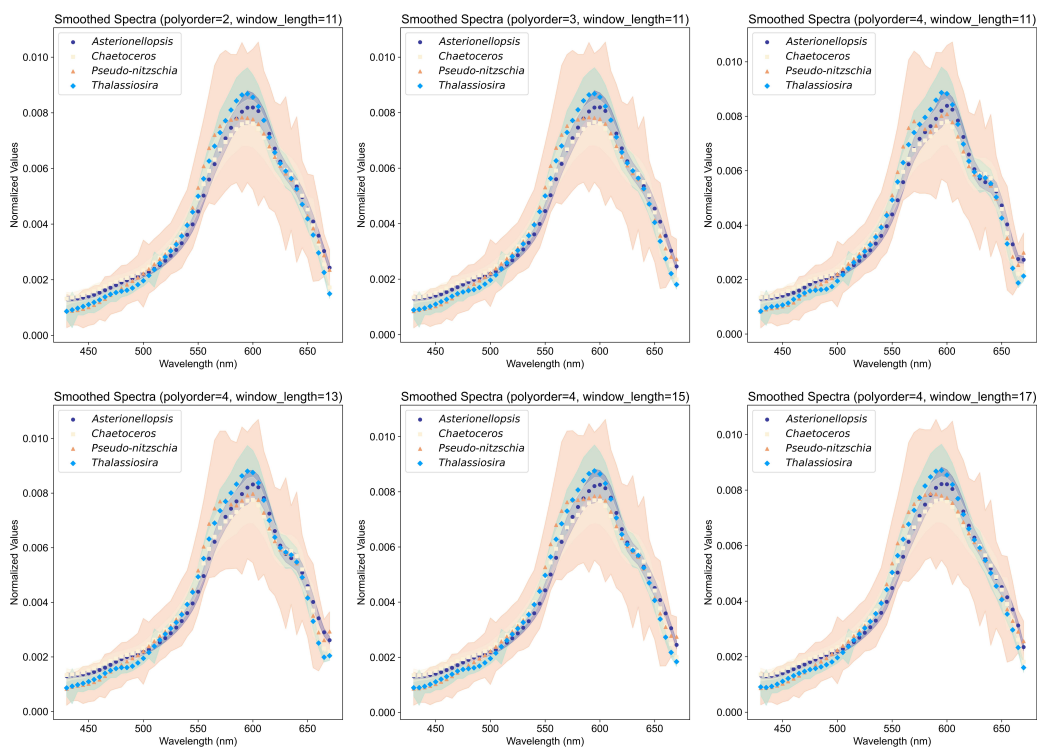


Figure 11: Iterations of the Savitzky-Golay smoothing algorithm applied to the reflectance spectra. The top row shows iterations of the smoothing function varying the polynomial order with a fixed window length of 11 (55 nm). The bottom row shows iterations changing the window-length from 13 (65 nm) to 17 (85 nm) with a fixed polynomial of 4. We selected the combination that preserved the most spectral features while removing noise. The two combinations that did this best were of polynomial size 4, with a window size of 11 or 13. We chose window size 13, matching Vandermeulen et al. 2017.

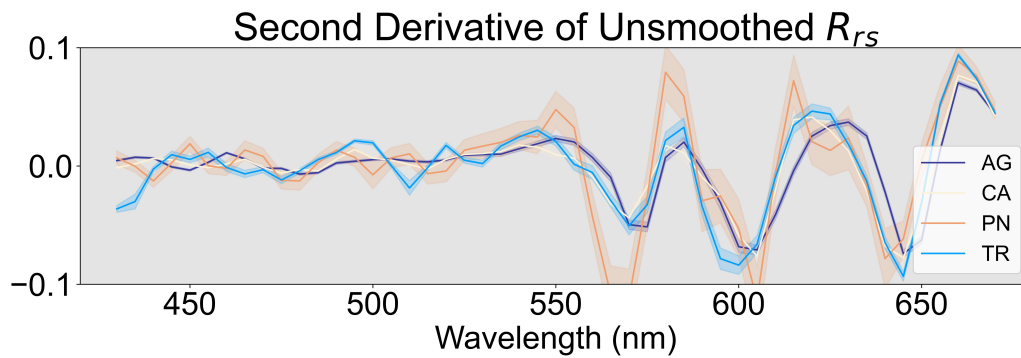


Figure 12: Second derivative of unsmoothed (raw) spectra.

One sample taken from each taxon. Note the accentuation of ‘features’ in wavelengths shorter than 550 nm. Whether these are real spectral features that we can glean information from or noise is unknown. Each of these lines are the median values of hundreds to thousands of absorption and backscatter measurements. Noise should be minimal. Yet, smoothing provides a cautionary step to ensure we are not mistaking noise for ‘true’ signal.

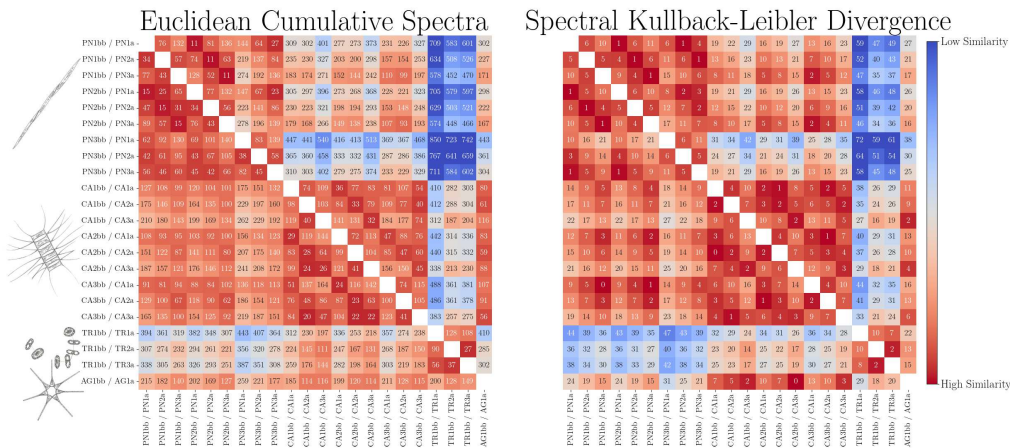


Figure 13: This figure displays two confusion matrices for the ECS and KLPD functions. There is no smoothing or alternate processing. Note the asymmetry in these heatmaps due to differences in integration for ECS (bottom is integrating from blue to red, top is integration from red to blue) and the triangular inequality of KLPD - the triangular inequality means we expect different results based on order of comparison (i.e. spectra a versus spectra b  $\neq$  spectra b versus spectra a). Kullback-Leibler pseudo-divergence performed better when comparing spectra of different diatoms against *Pseudo-nitzschia*, rather than comparing *Pseudo-nitzschia*'s spectra against other diatoms. Euclidean distance of cumulative spectra performs better when integrating the spectra from red to blue wavelengths, rather than blue to red. All values for KLPD and ECS are returned in absolute values since we are interested in the magnitude of the differences and not the sign.