AN ABSTRACT OF THE THESIS OF

Alexander Bailess for the degree of Master of Science in Ocean, Earth, and Atmospheric Sciences presented on May 7, 2025.

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

Abstract approved: ______________________________

Maria Kavanaugh

Diatoms are microscopic marine algae that are critical for global primary production, carbon sequestration, and fisheries productivity. However, marine ecosystems, and the fisheries they sustain, are threatened by select diatoms that form harmful algal blooms. Identifying harmful blooms is necessary to effectively manage marine resources, but our current framework for identification is limited by expensive and labor-intensive in situ point sampling. These point samples offer only a fragmented view of harmful algal bloom distribution. To effectively monitor these dynamic events in real time, we need a framework that combines wide-area coverage with high spatial and temporal resolution. Recent advances in hyperspectral remote sensing offer a promising path toward scalable, cost-effective monitoring - but remote sensing's ability to detect changes in dominance within phytoplankton groups, e.g. diatoms, is currently unknown.

To address this, we cultured the three most dominant diatom genera within the Northeast Pacific's upwelling system, and this systems' most abundant harmful algae, *Pseudo-nitzschia*. We measured the hyperspectral absorption and backscatter of these taxa, which were then used to model a spectral reflectance 'fingerprint' that a remote observing platform (satellite/drone) might detect. Differences between fingerprints of these taxa were quantified using vector-based and statistical analyses.

We found mean-spectral differences of 48% +/- 5% between the most dominant diatom, *Thalassiosira*, and the most toxic diatom, *Pseudo-nitzschia*. Mean-spectral differences of 29% +/- 11% and 34% were found between *Pseudo-nitzschia* and the second, and third most abundant diatoms, *Chaetoceros*, and *Asterionellopsis*, respectively. Successful identification of *Pseudo-nitzschia's* reflectance fingerprint was largely driven by the presence of a unique feature around 560 nm. The large differences observed between spectral fingerprints suggest identification by remote sensing is possible. This research leveraged new technologies to improve early detection of harmful algal blooms, with the potential to provide advanced knowledge for fisheries and aquaculture managers in a cost-effective and scalable way.

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

by

Alexander Bailess

A THESIS

submitted to

Oregon State University

in partial fulfillment of
the requirements for the
degree of

Master of Science

Presented May 7, 2025
Commencement June 2026

Master of Science thesis of Alexander Bailess presented on May 7, 2025.

APPROVED:

Major Professor, representing Ocean, Earth, and Atmospheric Sciences

Dean of the College of Earth, Ocean, and Atmospheric Sciences

Dean of Graduate Education

I understand that my thesis will become part of the permanent collection of Oregon State University libraries. My signature below authorizes release of my thesis to any reader upon request.

Alexander Bailess, Author

ACKNOWLEDGEMENTS

There are a great deal of people who deserve acknowledgment for their contributions to this work, indeed, if I were to name them all and their contributions in full, this section would likely be longer than the thesis itself. I will attempt to be brief, for your sake.

First, I thank my ever patient co-authors. Nicholas Baetge, who exudes inspirational positivity and insightfulness. Nick B. made significant contributions to the experimental design of this project and (along with Tara Conrad) taught me to culture phytoplankton like a microbiologist, 'are they healthy, what is their physiological state, and have we used the best sterile techniques possible?'. Nick B. has provided exceptional reviews for this work and taught me to be a better reviewer myself, he is not just a mentor, but a friend. I thank Nick Tufillaro for funding my first years at Oregon State and spending up to four hours a week with me my first quarter (alongside Andrew Barnard) teaching me optics. For the price of one beer, or perhaps two, Nick T. will make the time to meet me at any pub in Corvallis and field all the mathematical or physics-related questions I have. I have pictures of every chalkboard and piece of scratch paper he has scribbled on, explaining these concepts to me. I often spend a few weeks reviewing the concepts Nick T. jots down in a matter of minutes before I can fully grasp them - I shudder to think how long it would take me to learn these things without him. Thank you for all the times you had me over for dinner explaining distance metrics and hyperspectral analyses, for the field work with Gybe and VIIRS where I learned so much, and for showing me how to build a bike when mine was stolen. I thank Mike Behrenfeld for taking the time to make me a better writer, as I learn to speak my thoughts clearly. I also thank Mike for his generosity (along with all those in the phytoplankton eco-physiology lab) for allowing me to be a part of their lab, sharing their resources and time, and in doing so, teaching me many things about phytoplankton and light. I thank Jason Graff for reviews and feedback and most of all for his unwavering positivity; nothing gets this guy down, and as a result, everyone is brought up. I thank Brian Bill for taking a chance on me and replying to my emails about *Pseudo-nitzschia*. From these emails and a meeting in Seattle, he has provided a wealth of knowledge on this infamous toxicogenic diatom. We have Brian to thank for isolating the various *Pseudo-nitzschia* species we investigated in this experiment and for providing guidance on how to culture them well.

To my advisors, Maria Kavanaugh and Andrew Barnard. Maria Kavanaugh has given me freedom to explore during my time at Oregon State. At first I didn't know what to do with it - I spun out in every direction, sinking my teeth into each new and interesting thing without taking the time to chew (I hear my parents echo, 'Chew your food!'). But with time, this freedom taught me how to prioritize and how to focus. I learned to isolate the important things. This freedom provided me time to deeply ruminate on ideas, which, combined with many failed experiments, resulted in the thesis below. I greatly appreciate the trust Maria placed in me, allowing me to explore many topics and work at my own pace. Maria taught me to build a collaborative network, filled with those who excel where I struggle, which allowed me to tackle such a multidisciplinary topic. I am fortunate to have an advisor who makes me a better scientist, collaborator, and helps me to find balance. Maria always reminds us to take time for ourselves, be kind to ourselves, and keep hold of the passion we have for our work. Andrew Barnard has made time for me, selflessly blocking out at least one (and often many more) hours of his busy weekly-schedule to teach me about optics - no small feat considering I was an ecologist! Along with Nick T., Andrew has taught me almost everything that I know about optics. Throughout this time Andrew has become a close friend; I cherish the ability to freely discuss any matters with him, professional or personal. Andrew has taught me how to navigate the complex nuances of academia, and has consistently supported me through graduate school's ups and downs. I thank Andrew for funding much of my time, facilitating my journey into ocean optics, and always encouraging me to pursue fruitful opportunities.

To my optics friends and family. First and foremost I want to thank the students and staff of the 2023 Ocean Optics class. Especially, Matthew Kehrli, Corentin Subirade, and Ashley Ohall. I thank Matt for his friendship and intellectual camaraderie, and for the time spent teaching me how to code; I cannot thank you (and Meredith Kime!) enough, you really saved me that first week. I thank Coco for all our late nights in the optics lab and our time spent traveling - you are such a fun loving free spirit, the smile on your face always makes me beam inside. I thank Ash for being able to make everything fun, even when we are mired in optics work at 12 am, one can never have a bad time when Ash is around. I thank Luke Carberry for swimming with me nearly every day in Maine, our time exploring different islands in Harpswell Sound was among my favorite of the optics class. I learned a lot from Luke and Ash with their initial HyperBB experiments, indeed, their discoveries served as the foundation for the idea behind this work. I owe great debts to Nils Haëntjens and Guillaume Bourdin for mentorship and for developing software that kept me from my wits end. Thank you to Collin Roesler, Meg Estapa, and Emmanuel Boss. Collin for being one of the best teachers I have ever had - I have yet to find another who can explain such complex topics with perfect clarity, breaking them down piece by piece, "It really can never be too simple!". Meg for all her work making everything happen and run smoothly, she is the calm in the chaos - serving as both a mentor and guide for new students in the field. Thank you for your patience, kindness, and support! Emmanuel for the energy he brings to the field, he knows so much and generously shares it with all his students. The community that he and the Ocean Optics Class staff have created and fostered among my peers is the number one thing I value about my life as a scientist. Emmanuel has also exemplified the importance of work-life balance (but with all gas and no breaks for both work and play). I thank the CalVal VIIRS crew, Harrison Smith for giving me my first lessons on light scattering and absorption as we watched the sunsets in the middle of the Pacific. Mike Ondrusek for mentorship, guidance, patience, Sherwin Lander for his wit and humor, Charles Kovach for his great music and enchanting life stories, Joaquim Goes for his kindness, and Joaquin Chavez for his passion, work ethic, and kindness to me. Scott Freeman for his mentoring role and great attitude on life. All these people of the CalVal VIIRS team are 'optics father-figures' if you will and have taken great care of me.

To my Corvallis buds; Finley Dibert-Burgwin, Larry Ninh, Wave Moretto, Ian Black, Haley Carlton, Adrian Heath, Matthew Gschwend, Tara Conrad, Rick Colwell, Kelly George, Alex Soderquist, Riley Short, Marine Andriot, Lauren Horstmeyer, Kassondra Watson, and my late dog Jasmine to name but a few. You have all gotten me through the past few years here. I thank you all for keeping me sane during the 100+ hour weeks of 24 hour sampling and reminding me to get outside. Thank you for time spent foraging, snowboarding, SCUBA diving, spearfishing, disc golfing, playing pool, or just chatting. You all mean a very great deal to me.

Last, but certainly not least, I thank my folks. Mom and Pop, thank you for always supporting me in any endeavors I wish to pursue, no matter how much they stress you guys out. I love you both dearly (-:

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF APPENDIX FIGURES

Jasmine

You are missed by many

I love you lots

Chapter 1: The world's most ubiquitous primary producers

Algae live, in one from or another, nearly everywhere liquid water is present. Like plants, algae trap carbon dioxide from their environment and in doing so produce oxygen, while storing the carbon in their tissues. Large, multi-cellular algae make up the seaweeds, while small, single-celled algae are called phytoplankton.

Phytoplankton are responsible for nearly half of the global net primary production, assimilating roughly 52 petagrams (petagram = 10^{15} grams) of carbon annually (Silsbe et al. 2016, T. Westberry et al. 2008). To put this in perspective, the weight of 8 billion humans on earth would only be about half a petagram. Oxygen production and carbon capture sustains nearly all heterotrophic marine life, as phytoplankton are the primary source of fixed carbon for aquatic consumers. Through a series of predator-prey interactions, this carbon makes it to the commercially fished species that sustain the human populations and economies (Costalago et al. 2020). Food from the sea currently accounts for around 15-17% of total human protein consumption (Boyd, McNevin, and Davis 2022) with the expectation that this will rise in the future to meet increased demand.

1.1 Diatoms

Diatoms form a cosmopolitan subgroup of phytoplankton and play important roles in productive ecosystems. Large diatoms (those greater than 10 um) quickly export carbon to the benthic environment and deep ocean (Munk and Riley 1952, Waite et al. 1997), while mediating efficient food webs (Chavez, Messié, and Pennington 2011). Diatoms are estimated to account for up to 25 percent of *all oceanic primary production* (Tréguer et al. 1995, Moore et al. 2001b, Aumont et al. 2003, M. J. Behrenfeld et al. 2021) and exhibit incredible diversity—indeed, the estimated number of diatom species exceeds that of all other phytoplankton groups combined (Kooistra et al. 2007). This species richness is partly due to diatom's success in upwelling systems, where intermittent, but intense injection of nutrients promote bloom-conditions, facilitating genetic mixing. While the diatoms are characterized by siliceous frustules and large vacuoles, extensive speciation has given rise to a wide array of shapes, chain formations, and frustule characteristics that distinguish individual taxa. These morphological differences often reflect unique physiologies and ecological roles, with different genera contributing distinct ecological functions.

While blooming, diatoms greatly alter the marine environment (e.g. changes in underwater light field, prey availability, biogeochemistry) and ecosystems therein. *Thalassiosira* holds significant ecological importance due to its high lipid content - up to 52% of its dry-weight (Yi et al. 2017, Bhattacharjya et al. 2020a). These lipid-rich algae are key food sources for large zooplankton which sustain commercially important fish species (Stock et al. 2017, W. T. Peterson et al. 2017). *Pseudo-*

nitzschia is also ecologically significant, not for its nutritional value, but for its ability to produce domoic acid - a compound capable of reshaping entire ecosystems in a very different way.

1.2 Harmful Algae

Harmful algal blooms (HABs) are aggregations of algae that pose a threat to human/ecosystem activities and health. Harmful algae alter organism/ecosystem function through physical disruptions or chemical production. For example, *Chaetoceros* setae can inflict damage to sensitive tissues like gills and have caused mass mortalities (e.g. Bell 1961, Yang and Albright 1992). Mucus from *Thalassiosira* blooms have had similar effects (Prasad, Nienow, and Lochner 2018). While most diatoms do not produce toxins, *Pseudo-nitzschia* is a notable exception. Certain species of *Pseudo-nitzschia* produce domoic acid, a potent neurotoxin that has profound effects on both ecosystems and human health. Domoic acid begins to accumulate in filter feeders like shellfish and forage fish (anchovies, sardines). These filter feeders are then eaten by marine mammals, seabirds, and fishes of commercial importance, in which toxicity becomes acute. For marine mammals and humans, domoic acid poisoning is devastating, leading to neurological damage, disorientation, seizures, and amnesia. In extreme cases it causes coma and death (Bates et al. 2018). Of the 29 known toxin-producing *Pseudo-nitzschia* species, at least 21 have been found off the west coast of North America (Bates et al. 2019).

In 2015, the west coast of North America experienced record-breaking toxin levels from a *Pseudo-nitzschia* bloom that spread from British Columbia to Baja California. The 2015 bloom closed recreational and commercial fisheries including: mussels, razor clams, oysters, Dungeness and rock crabs, anchovies, and sardines. This corresponded to 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 reliant on tourism. The high levels of domoic acid caused severe damage to marine ecosystems with effects propagating throughout the food web - many of which we are yet to understand. Mass moralities 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, Gibble et al. 2018). It is of utmost importance for fisheries managers, aquaculturists, and coastal communities to know when and where *Pseudo-nitzschia* is blooming in dense concentrations.

Current methods of detection require point samples of shellfish which has three clear drawbacks. First, point samples result in poor spatial and temporal resolution of toxin distribution. Second, positive toxin detection means seafood is already contaminated, and managerial decisions lag behind toxin accumulation - even low concentrations of domoic acid have harmful effects (Lefebvre et al. 2017). Finally, year-round point sampling is expensive (millions of dollars per year for US coastal states (D. M. Anderson et al. 2000)). Alternatively, one can look for *Pseudo-nitzschia* in water samples. Observed individually, the trained eye can easily identify *Pseudo-nitzschia*, however, trained eyes (phytoplankton taxonomists) are expensive, as manual identification using a microscope can take days to get through but a handful of samples. Clearly, there is a strong need for a framework that can rapidly identify *Pseudo-nitzschia* in real time over broad spatial scales to mitigate human and fishery impacts.

1.3 Bio-Optics and the Spectral Fingerprint

Bio-optical tools allow us to observe the color and intensity of light in ways we cannot imagine using the human eye. Our eyes have 3 color receptors, red, green, and blue (Smith and Guild 1931), from which different combinations of intensity form the rainbow of colors we may observe. Our eyes are multi-spectral, multi meaning multiple (in this case 3), and spectral meaning color, for our 3 color receptors. In contrast, bio-optical tools may be hyperspectral, with new instruments having hundreds, to thousands, of spectral channels. A picture taken from a hyperspectral platform (satellite, drone, handheld radiometer/camera) will provide a spectrum, a combination of the color and intensity of reflected light. This spectrum is often a complex, non-linear function. This function provides information about a subject's interaction with light, which can be a proxy for other physical processes of interest. A spectrum therefore provides a sort of fingerprint, that can be used to infer the presence or absence of a physical substance and/or process. Fingerprints of unique phytoplankton populations can be derived semi-empirically, via inversion modeling (Gordon et al. 1988, Zaneveld 1995). Such models relate in-situ measurements of light absorption and backscatter to the color and intensity of light an observer would see (reflectance).

It should be noted, however, that the many unique particles and dissolved substances within the first optical depth (depth to which a satellite can see) each contribute to ocean color based on their concentration, depth, and orientation. Hence, one must consider the competing optical signatures of other constituents when looking for the fingerprint of a specific group within any spectrum.

NASA's latest satellite mission, PACE (Plankton, Aerosols, Clouds, and Ecosystems), possesses a hyperspectral sensor with over 100 spectral channels. The high-resolution spectra this provides, coupled with existing knowledge of common marine constituent's optical properties, may help to identify the fingerprints of functionally diverse phytoplankton and even HABs from space.

1.4 Absorption

One contributor to the optical properties of phytoplankton (how they interact with light) are their pigments. Eukaryotic phytoplankton pigments are molecular complexes held in the thylakoid membrane by adjoining proteins that serve to capture as much visible light from the environment as possible and funnel this energy to the photosystems to perform photosynthesis or disperse excess energy as heat.

The way pigments interact with the surrounding light field is influenced by how they are arranged within cells. Overlap of pigments results in intra-cellular self-shading, known as the packaging effect (Duyens 1956). The size and internal structuring of chloroplasts, individual cells, and colonies, also influences the packaging effect, as large and more densely packed morphological features increase self-shading. (Kirk 2010, Ciotti, Lewis, and Cullen 2002). This effect results in a broadening of spectral peaks and shallowing of troughs, as wavelengths of light that are statistically less likely to be absorbed (such as green light for chlorophyll a) have more material to pass through before exiting the cell, increasing the probability of absorption. Alternatively, wavelengths of light where absorption is highly probable (blue or red light for chlorophyll a) are already absorbed entirely after passing through a much shorter path-length in this medium.

Absorption is a large contributor to reflectance spectra because all light that is absorbed is eliminated from the light field, and therefore cannot be scattered to reach our eyes or the satellite. Hence, for wavelengths of light where absorption is high (dependent on the phytoplankton-specific pigments) we have reduced reflectance.

1.5 Backscatter

Backscatter is all scatter (redirection) of light that is in the backwards direction (towards its source) which occurs after interactions with a particle or surface. In the ocean, the shape, size, and orientation of the cell dictates the way light is scattered. Smaller cells (e.g. bacteria) tend to scatter light more isotropically (in all directions) and preferentially scatter shorter, higher energy (bluer) wavelengths of visible light. Conversely, light interacting with relatively larger particles (e.g. diatoms) tends to scatter a larger proportion in the forward (relative to backwards) direction, and has weak dependence on wavelength (color) (Mobley et al. 2022, Van de Hulst 2012, Boss et al. 2001). The great diversity of shape and intracellular structure of these taxa means the magnitude and direction of light scattering is often unpredictable, however, there are some inferences we can make in relation to the size of morphological features. For example, long, thin setae (spines) extend from *Chaetoceros spp.* which we expect to preferentially scatter bluer wavelengths due to their small size. The thin chitinous threads that bind *Thalassiosira spp.* (Tran et al. 2023) might have similar effects - but these are likely weaker due to chitin's blue-green absorbance (Azofeifa, Arguedas, and Vargas 2012).

Backscatter measurements are challenging due to low scattering efficiencies. For phytoplankton, we expect less than 0.1% of the total scattered light to be in the backwards direction (Bricaud, Morel, and Prieur 1983). As a result of this small signal, backscatter sensors often only measure 1 to 3 wavelengths with large spectral bandwidths to measure as much of the available light as possible. This low spectral resolution has limited our knowledge of backscatter spectral shapes for complex (nonspherical) particles - despite models showing that backscatter should have greater spectral variation (and therefore information) than the spectrum of total scatter alone (Morel and Bricaud 1981). This lack of intuition for the spectral shape of backscatter means it is often assumed to be a monotonic function (e.g. Lee, Carder, and Arnone 2002, Lee et al. 2009, Kostadinov, Siegel, and Maritorena 2009, Gordon and Morel, 1983, Morel and Prieur 1977); while this assumption works well for clear-oceanic waters, we anticipate it is hampering ocean color models in highly productive systems dominated by non-normal particle size distributions (i.e. phytoplankton blooms). Backscatter forms the backbone of ocean color (Boss et al. 2004), comprising all observable light leaving the ocean. As such, it is critical that we deepen our understanding of backscatter's spectral behavior to monitor these highly-productive systems using remote sensing.

1.6 Summary

Phytoplankton are globally important primary producers. Diatoms are especially important phytoplankton for their role in marine food webs, carbon cycling, and oxygen production. Some of these diatoms can be harmful, but the majority are beneficial and essential to marine ecosystem function. Early warning systems on how marine ecosystems - and the resources they provide - may change is of great importance. One way to assess ecosystem health and function is by monitoring biodiversity. Biodiversity acts as a canary, with community assemblage shifts being early indicators of ecosystem change (Kavanaugh et al. 2021). But observing the biodiversity of phytoplankton in the ocean is both difficult and costly, especially considering the large spatial scales (71% of the planet). The need to understand global biodiversity, but the great difficulty in monitoring it, is one of NASA's 'Grand Challenges'. NASA states that in order to better understand ocean ecosystem change and plankton diversity we will need to integrate targeted process studies with modeling and remote sensing (M. Behrenfeld et al. 2022). The body of work to come details a targeted study on the bio-optical signatures of four key phytoplankton taxa, which through inversion modeling, form the basis of a framework that aims to monitor their biodiversity using remote sensing.

This chapter has served to provide a brief overview of phytoplankton, harmful algal blooms, and the theory behind their detection using remote sensing. Chapter 2 will use this theory to construct spectral fingerprints for four ecologically significant marine phytoplankton and explore several methods to distinguish between fingerprints. We will then propose a framework for the identification of *Pseudo-nitzschia* using ocean color. Chapter 3 will summarize this work and outline next steps.

Chapter 2: Hyperspectral Remote Sensing for Harmful Algal Bloom Detection: *Pseudo-nitzschia* in the Northeast Pacific

2.1 Abstract

Diatoms are microscopic marine algae that are critical for global primary production, carbon sequestration, and fisheries productivity. However, marine ecosystems, and the fisheries they sustain, are threatened by select diatoms that form harmful algal blooms. Identifying harmful blooms is necessary to effectively manage marine resources, but our current framework for identification is limited by expensive and labor-intensive in situ point sampling. These point samples offer only a fragmented view of harmful algal bloom distribution. To effectively monitor these dynamic events in real time, we need a framework that combines wide-area coverage with high spatial and temporal resolution. Recent advances in hyperspectral remote sensing offer a promising path toward scalable, cost-effective monitoring - but remote sensing's ability to detect changes in dominance within phytoplankton groups, e.g. diatoms, is currently unknown.

To address this, we cultured the three most dominant diatom genera within the Northeast Pacific's upwelling system, and this systems' most abundant harmful algae, *Pseudo-nitzschia*. We measured the hyperspectral absorption and backscatter of these taxa, which were then used to model a spectral reflectance 'fingerprint'

that a remote observing platform (satellite/drone) might detect. Differences between fingerprints of these taxa were quantified using vector-based and statistical analyses.

We found mean-spectral differences of 48% ± 10% between the most dominant diatom, *Thalassiosira*, and the most toxic diatom, *Pseudo-nitzschia*. Mean-spectral differences of 29% ± 13% and 34% ± 7% were found between *Pseudo-nitzschia* and the second, and third most abundant diatoms, *Chaetoceros*, and *Asterionellopsis*, respectively. Successful identification of *Pseudo-nitzschia's* reflectance fingerprint was largely driven by the presence of a unique feature around 560 nm. The large differences observed between spectral fingerprints suggest identification by remote sensing is possible. This research leveraged new technologies to improve early detection of harmful algal blooms, with the potential to provide advanced knowledge for fisheries and aquaculture managers in a cost-effective and scalable way.

2.2 Introduction

Phytoplankton are responsible for approximately half of global net primary production, assimilating around 52 Pg of carbon annually (Silsbe et al. 2016, T. Westberry et al. 2008). Diatoms form an important subgroup of phytoplankton, accounting for roughly 20 percent of all marine primary production (M. J. Behrenfeld et al. 2021, Tréguer et al. 1995, Moore et al. 2001b, Aumont et al. 2003). Diatoms' success in the marine environment is reflected in their incredible diversity, yet all diatoms can be characterized by the presence of vacuoles and siliceous frustules.

For large diatoms, the combination of size and mineral ballasting (from the frustule) leads to significant carbon export. Carbon export from diatoms may comprise up to 40 percent of the biological pump (Jin et al. 2006, Tréguer et al. 2018). Large-bodied diatoms also mediate efficient energy transfer to higher trophic levels by reducing the number of trophic interactions needed from primary producers to final consumers, bypassing microbe-microbe interactions in favor of direct consumption by zooplankton (Chavez, Messié, and Pennington 2011).

There are a few diatoms, however, that are noxious. Harmful algal blooms (HABs) are defined as any aggregation of algae that poses a threat to human/ecosystem health or is a nuisance to human activity. The diatom *Pseudo-nitzschia* is one of the deadliest, with certain species able to produce a potent neurotoxin, domoic acid, that propagates through marine food webs. Domoic acid bioaccumulates in primary consumers such as shellfish and forage fish (anchovies, sardines) before being eaten by 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).

HAB identification is the first step in ecological analysis, fisheries management, and early warning/management 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 develop spectral (color) ‘fingerprints’. These fingerprints are derived by measuring the optical properties (absorption and backscatter) of a specific taxon which are converted to what an ocean color remote sensing platform might see using a reflectance proxy (Gordon et al. 1988, Zaneveld 1995). In practice, however, isolating fingerprints from the ocean is often confounded by the overlapping optical properties of dissolved substances, detritus, and other phytoplankton. Fortunately, the bloom-forming diatoms cultured in this study form dense surface aggregations that dominate the ocean color signal. This dominance makes them particularly well-suited for identification using spectral fingerprints.

To derive the spectral fingerprints, we need to measure backscatter. Backscatter is a key component of ocean color and consists of all scattered light in the backwards direction. Redirection occurs after interactions with a particle or surface (Boss et al. 2004). Phytoplankton are highly-absorbing particles, making them poor backscatters. As a result of this low signal, backscatter of marine phytoplankton has rarely been measured in more than a few wavelengths. Consequently, our knowledge of backscatter spectral shapes is limited. However, theoretical models, and a handful of empirical studies, have shown that hyperspectral backscatter should have more pronounced features than total scatter (Morel and Bricaud 1981). This conflicts with the common assumption that spectral backscatter lacks complex features and shapes. Spectral backscatter is often assumed to be a monotonic function, or even flat. We believe ocean color inversion models will advance significantly, increasing their utility in detecting HABs, by incorporating the information found in empirically derived backscatter spectral shapes.

In this study, we derived hyperspectral backscatter coefficients of the most dominant phytoplankton genera from a large upwelling system in the Northeast Pacific, including the region's most prominent HAB genera. This work was inspired by that of Bricaud, Morel, and Prieur in 1981, the first to measure the hyperspectral backscatter of marine phytoplankton, which noted that spectral backscattering exhibits complex shapes, and contains more pronounced spectral features than the spectrum of total scattering alone. Spectral backscatter values were divided by concurrent measurements of hyperspectral absorption to derive the reflectance proxy. This modeled reflectance mimics the *in situ* reflectance an ocean-color satellite would observe.

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 applied statistical analyses to quantify these differences. *Pseudo-nitzschia*, the dominant HAB taxa, exhibits a unique feature in its modeled reflectance fingerprint, a bi-furcated peak near 560 nm. Due to this feature, the spectral fingerprint of *Pseudo-nitzschia* is readily distinguishable from those of the three most abundant diatoms in this system—both through visual assessment and using unsupervised classifying algorithms. These results support future studies using remote sensing for the identification of harmful *Pseudo-nitzschia* blooms in the Northeast Pacific.

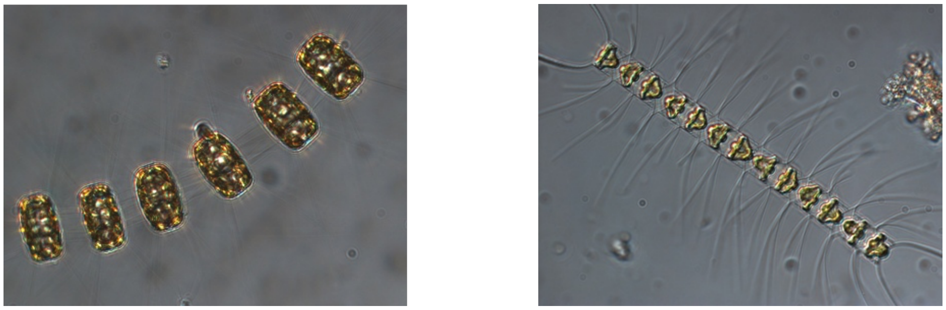
2.3 Methods

2.3.1 Cultures

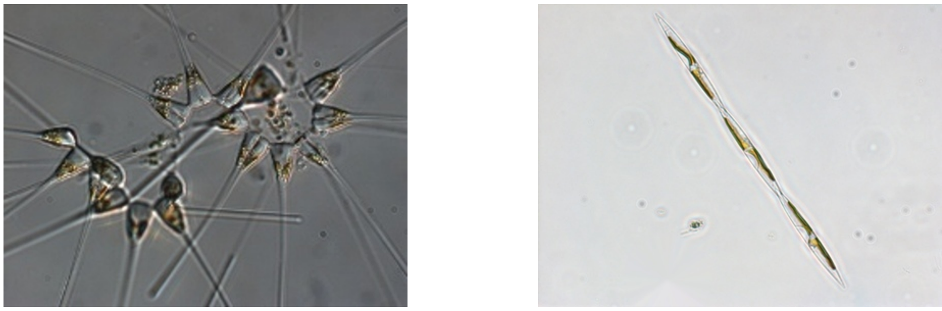
Four of the most abundant, bloom-forming genera of diatoms within the Northeast Pacific's large upwelling system were selected for this experiment. They are, in order of cell abundance: *Thalassiosisra rotula*, *Chaetoceros affinis*, *Asterionellopsis glacialis*, and *Pseudo-nitzschia spp.* (Du, W. Peterson, and O'Higgins 2015, Lassiter et al. 2006). These plankton 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 μmol photons m^2/s with a relatively flat white spectrum. The intensity and spectral distribution of the grow lights (a Phyton-Systems light bank) were measured using a Satlantic Hyper-OCR and 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 before experimental sampling to ensure steady-state, photoacclimated growth. Cell and detrital abundance were quantified daily via an Imaging Flow

CytoBot (IFCB). We also observed, using light microscopy, the presence of detritus too small for the IFCB, and quantified it using a Coulter Counter (Beckman Multisizer 3, 100 μm aperture). The biomass of detritus was small compared to that of the live cells and detrital particles were almost entirely < 3 μm, much smaller than this study's phytoplankton which were > 10 μm (Table 1). Phytoplankton photophysiology was monitored using a fast repetition rate fluorometer (FRRf). FRRf samples were exposed to 15 μmol photons/m^2/s at 480 nm for 10 minutes to relax non-photochemical quenching (NPQ) before measurements were made (Milligan, Aparicio, and M. Behrenfeld 2012). The amount of time and total irradiance to fully relax NPQ varied by taxon, time of day, and cell physiology. We conducted several experiments and found 15 μmol photons/m^2/s for 10 minutes to yield the highest Fv/Fm (˜0.6 ± 0.05) for these taxa on average.



(a) *Thalassiosira spp.* (b) *Chaetoceros spp.*



(c) *Asterionellopsis glacialis* (d) *Pseudo-nitzschia spp.*

Figure 2.1: Representative diatom genera (reproduced with permission from the Kudela Lab).

Table 2.1: Approximate cell size ranges for representative diatom species.

Individual Cell Sizes		
Species	width(μm)	length(μm)
Thalassiosira rotula	8 - 60	5 - 20
Chaetoceros affinis	8 - 15	7 - 30
Asterionellopsis glacialis	5 - 8	30 - 150
Pseudo-nitzschia spp.	3 - 8	70 - 145

2.3.2 Inline system

Measurements of phytoplankton absorption, scatter, and attenuation coefficients were made using an 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 μm 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 to reduce the influence of outliers and noise. This process was repeated using a 5.0 μm filtrate of the sample to estimate the optical contributions of detritus and colored dissolved organic matter (CDOM) to the culture. Subtraction of the 5 μm 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 m^{-1} and 0.0005 m^{-1}

level, respectively. 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 was linearly interpolated to 1 nm resolution and then down-sampled to 5 nm. Intrataxon data were then bootstrapped to provide more spectra for clustering algorithm analysis. The primary clustering algorithm used was an unsupervised machine learning algorithm, k-means, from the open-source Python machine learning library Scikit-learn. Bootstrapping consisted of backscatter spectra from triplicate cultures being 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). To better compare spectral fingerprints between taxa, each value within a spectrum was normalized by dividing it by the mean of the entire spectrum. This method highlights variations in spectral shape rather than overall magnitude.

2.3.3 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):

$$R_{rs\lambda} \approx \frac{f}{Q}\frac{b_{b\lambda}}{a_{\lambda} + b_{b\lambda}} \ , \tag{2.1}$$

where R_{rs}(reflectance) is the parameter that describes the intensity and color of light, at wavelength λ, leaving the ocean's surface. The backscatter coefficient at wavelength λ is represented by $b_{b(\lambda)}$, while $a_{(\lambda)}$ denotes the absorption coefficient at wavelength λ. 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:

$$\frac{b_{b\lambda}}{a_{\lambda}} \ , \tag{2.2}$$

as the contribution of backscatter in the denominator is negligible.

2.3.4 Distance Metrics

Distance functions were used to distinguish between modeled spectral fingerprints in a reliable, quantitative way. These functions are able to define the degree of similarity between complex, non-linear spectra by assessing differences in translation (center wavelength/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. Kruse et al. 1993, Wei et al. 2022). These tools are unaffected by changes in spectral magnitude and provide absolute differences between spectra. Spectral Correlation Mapper uses Pearson's correlation coefficient (r):

$$\text{Spectral Correlation Mapper (r)} = \frac{\sum_{i=1}^{n}(A_i - \bar{A})(B_i - \bar{B})}{\sqrt{\sum_{i=1}^{n}(A_i - \bar{A})^2 \sum_{i=1}^{n}(B_i - \bar{B})^2}} \tag{2.3}$$

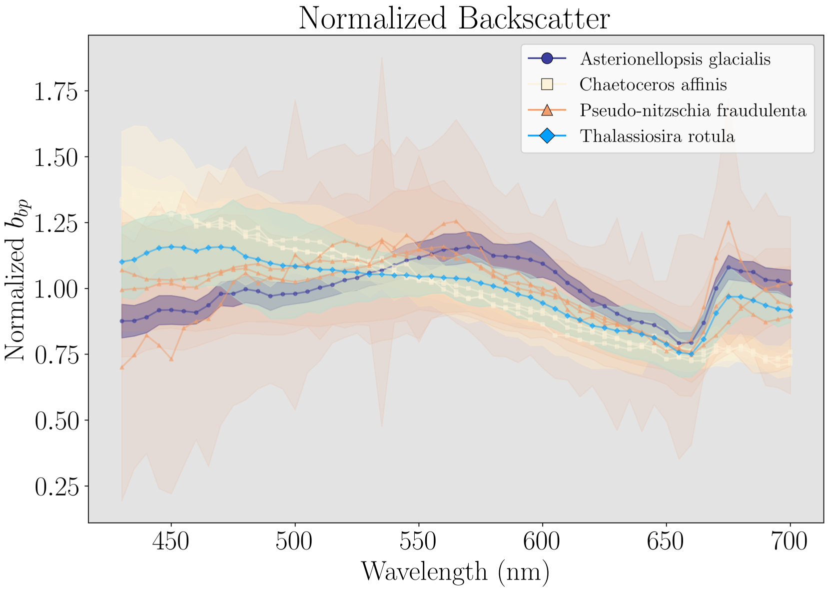
$$\bar{A} = \frac{1}{n}\sum_{i=1}^{n} A_i \qquad \bar{B} = \frac{1}{n}\sum_{i=1}^{n} B_i \ ,$$

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.

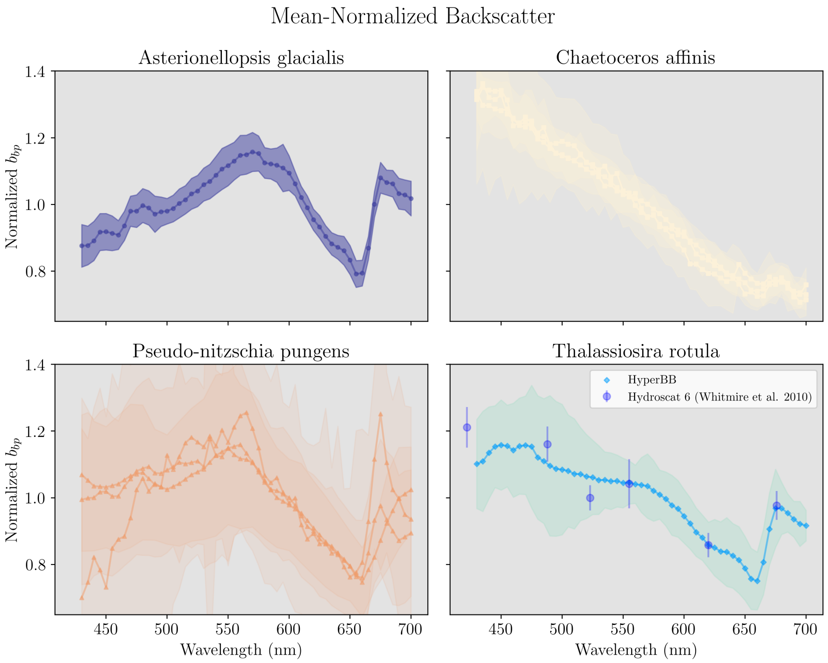
Spectral Correlation Mapper's sensitivity to differences between spectral signatures can be further improved through the application of derivative transformations. However, while derivatives excel at highlighting small features, the increased sensitivity also amplifies noise, which must first be removed 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 allows for spectrally wide nonlinear features to be conserved while low-level 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 savigol 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. Savigol 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).

2.4 Results

All cultured phytoplankton share local maxima of backscatter around 560/570 nm (figure 2.2), the yellow-green region of our spectra. This feature likely corresponds to the intersection between the absorbance of fucoxanthin and the secondary peaks of chlorophyll b and c, forming an absorbance minimum. 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 2.2), 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 sharp decline to a local minima near 585 nm. The 585 nm feature in *Pseudo-nitzschia's* backscatter spectra creates an important feature in *Pseudo-nitzschia's* reflectance fingerprint, seen as a bifurcated peak in figure 4. The 585 nm feature appears more pronounced in *Pseudo-nitzschia* than other taxa with similar absorption spectra. 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 2.3).



(a) Mean-normalized backscatter spectra comparison.



(b) Mean-normalized backscatter by genera.

Figure 2.2: 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 *Chaetoceos*. A multispectral comparison of *Thalassiosira rotula* from A. L. 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 2.3 shows the absorption spectra for each of the cultured diatoms. While the absorption spectra of triplicate cultures are very similar, we also observe a greater similarity between taxa than in normalized backscatter (figure 2.2) or total scatter spectra (figure 1, appendix). There are small deviations in the absorption spectra of the third triplicate sample of *Pseduo-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 (chlorophyll 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 greatest and most noteworthy difference in spectral shape of figure 2.4 is for *Pseudo-nitzschia*, which forms distinct bifurcated peaks at 570 and 600 nm, with a trough at 585 nm. This noticeable feature easily distinguishes 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 peak at 600 nm, however the height of this peak, relative to the shoulder, is greater for *Asterionellopsis.* For all genera, there are only small differences in slope from the cyan to deep-blue (˜430 to 510 nm), with the largest differences in yellow and orange (˜560 to 620 nm). There is relative uniformity at

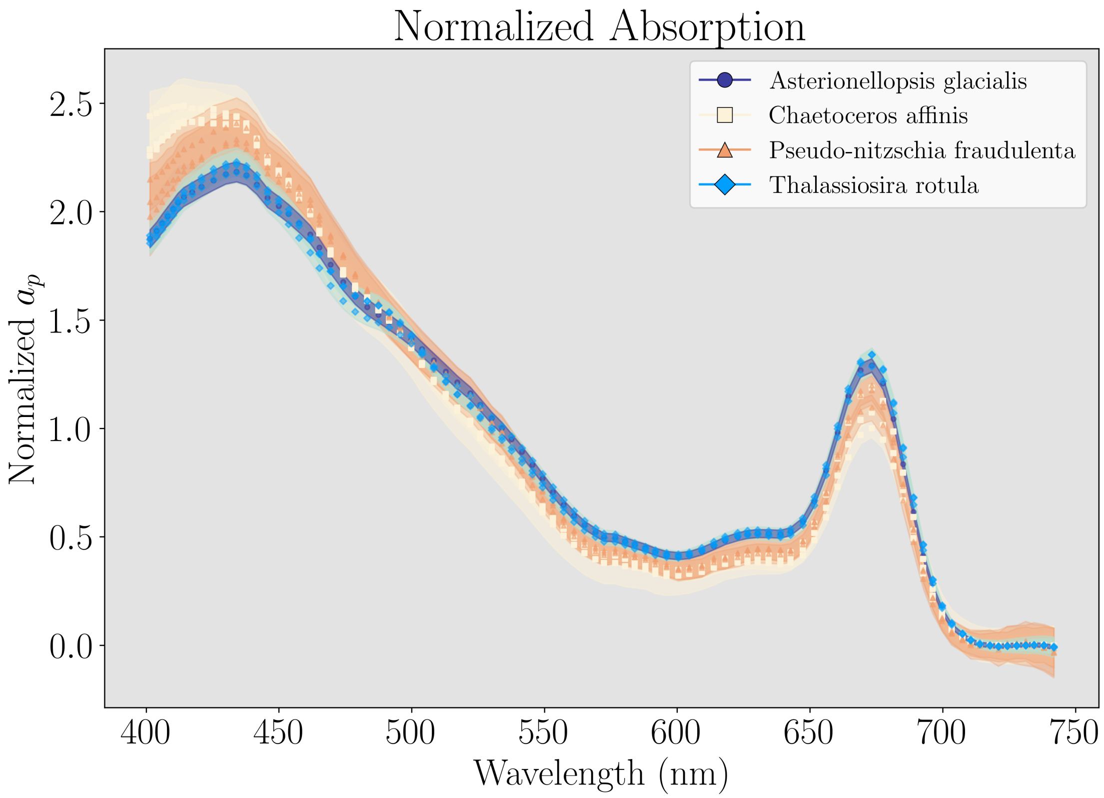
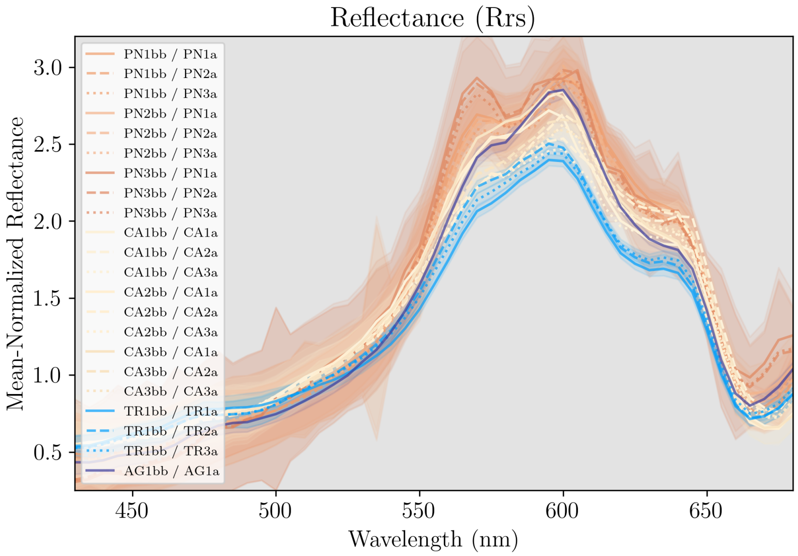
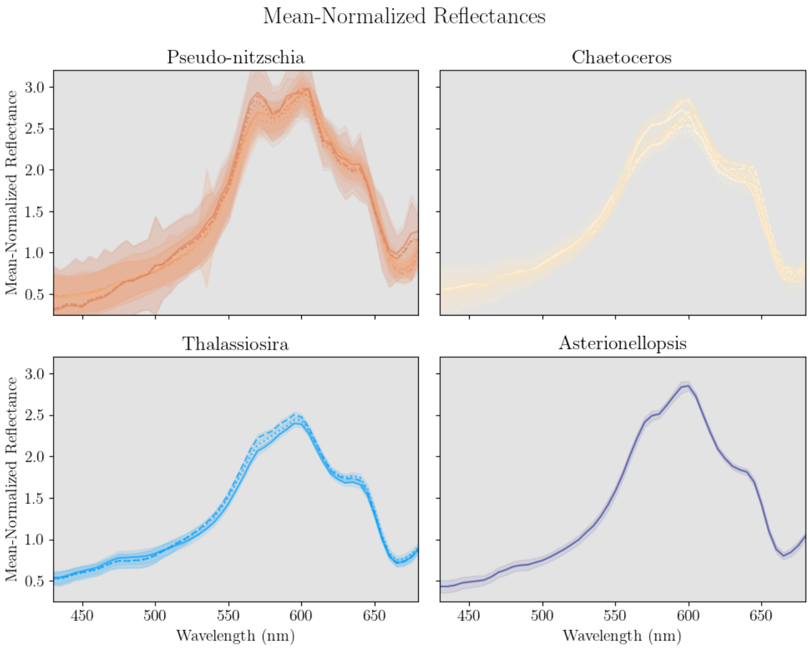


Figure 2.3: All cultures absorption spectra using the AC-s, spectra are mean-normalized. The peak centered on 440 nm is the Soret band, and the peak over 670 nm is the Qy band.

all other wavelengths. Thus, due to the similarities in the absorption spectra of diatoms, genus-specific identification through ocean color will be most easily derived from unique features in spectral backscatter centered within the absorbance minima of 550 to 650 nm.



(a) Mean-normalized reflectance fingerprints.



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

Figure 2.4: 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 λ_i), converted to reflectance spectra using equation 2.2.

Unsupervised machine learning algorithms were able to successfully distinguish between the spectral fingerprints of *Pseudo-nitzschia*, *Chaetoceros*, and *Thalassiosira* (figure 2.5). *Asterionellopsis* was routinely misclassified as *Chaetoceros*. We found the k-means algorithm performed optimally using k=3 (clustering into 3 groups). Using k=4 yielded similar results with the exception that *Pseudo-nitzschia* was divided into two clusters, while *Asterionellopsis* was still misclassified as *Chaetoceros*. Other non-Euclidean machine learning algorithms were applied and results from these are provided in the appendix.

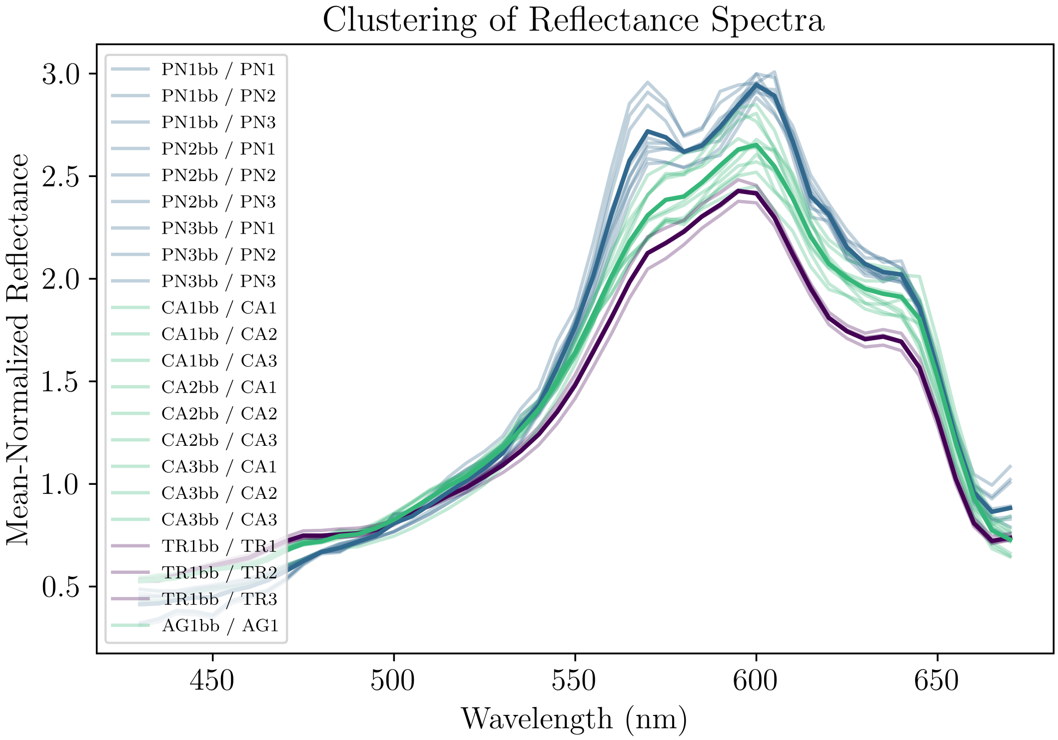


Figure 2.5: All bootstrapped spectra clustered using the k-means algorithm for k=3. *Pseudo-nitzschia* is abbreviated as 'PN', *Chaetoceros* as 'CA', Thalassiosira as 'TR', and Asterionellopsis as 'AG'. The number following the two letter phytoplankton ID code corresponds to the replicate culture number. '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.

To accentuate spectral features, the second derivatives of reflectance fingerprints from figure 2.4 were calculated and are shown in figure 2.6. These smoothed, second derivative fingerprints were then compared using the Spectral Correlation

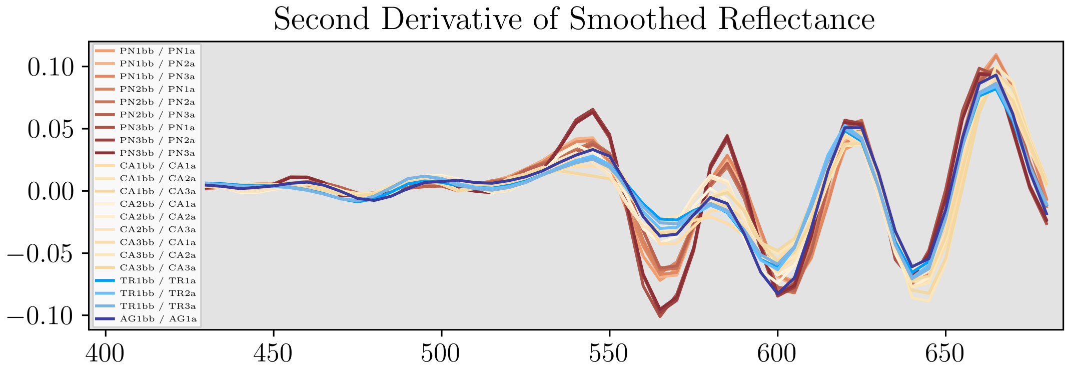
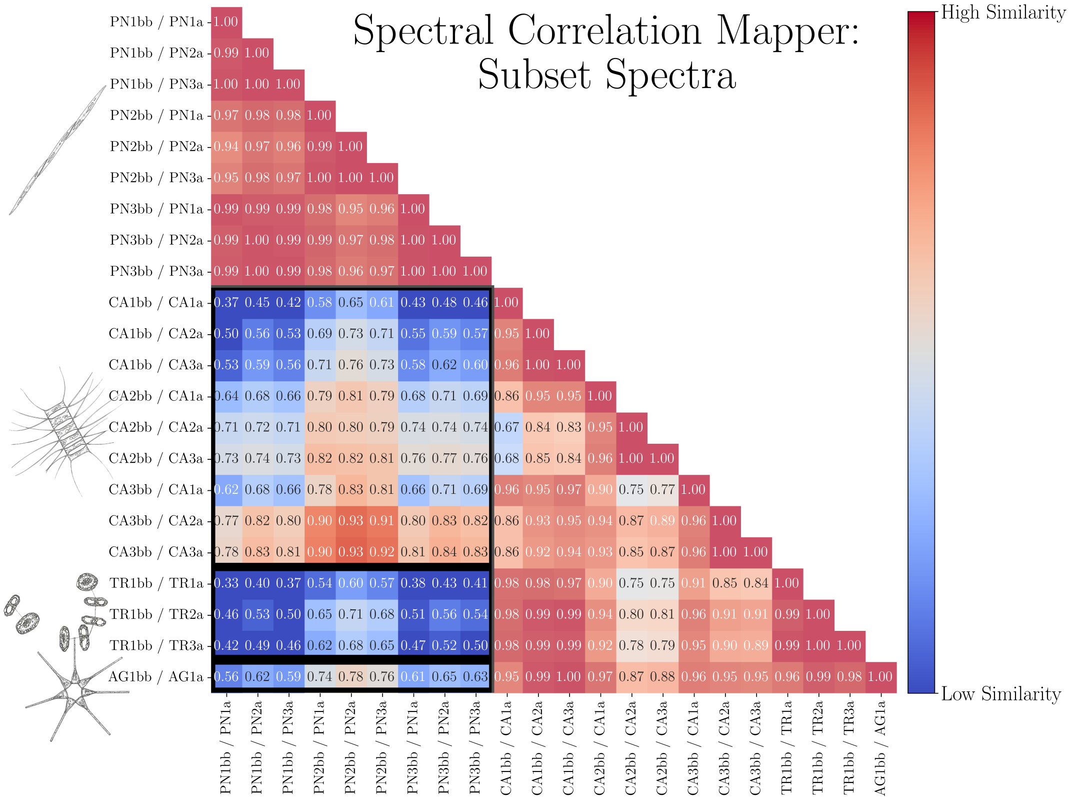


Figure 2.6: Second derivative of the smoothed spectra.

Mapper (SCM) distance metric. A heatmap of the evaluated differences is shown in figure 2.7. SCM performed better when subsetting the spectra in figure 2.6 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% +/- 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% +/- 13% between *Pseudo-nitzschia* and *Chaetoceros* (top, black box), and 34% +/- 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% +/- 10%.



	PN1bb / PN1a	PN1bb / PN2a	PN1bb / PN3a	PN2bb / PN1a	PN2bb / PN2a	PN2bb / PN3a	PN3bb / PN1a	PN3bb / PN2a	PN3bb / PN3a	CA1bb / CA1a	CA1bb / CA2a	CA1bb / CA3a	CA2bb / CA1a	CA2bb / CA2a	CA2bb / CA3a	CA3bb / CA1a	CA3bb / CA2a	CA3bb / CA3a	TR1bb / TR1a	TR1bb / TR2a	TR1bb / TR3a	AG1bb / AG1a
PN1bb / PN1a	1.00																					
PN1bb / PN2a	0.99	1.00																				
PN1bb / PN3a	1.00	1.00	1.00																			
PN2bb / PN1a	0.97	0.98	0.98	1.00																		
PN2bb / PN2a	0.94	0.97	0.96	0.99	1.00																	
PN2bb / PN3a	0.95	0.98	0.97	1.00	1.00	1.00																
PN3bb / PN1a	0.99	0.99	0.99	0.98	0.95	0.96	1.00															
PN3bb / PN2a	0.99	1.00	0.99	0.99	0.97	0.98	1.00	1.00														
PN3bb / PN3a	0.99	1.00	0.99	0.98	0.96	0.97	1.00	1.00	1.00													
CA1bb / CA1a	0.37	0.45	0.42	0.58	0.65	0.61	0.43	0.48	0.46	1.00												
CA1bb / CA2a	0.50	0.56	0.53	0.69	0.73	0.71	0.55	0.59	0.57	0.95	1.00											
CA1bb / CA3a	0.53	0.59	0.56	0.71	0.76	0.73	0.58	0.62	0.60	0.96	1.00	1.00										
CA2bb / CA1a	0.64	0.68	0.66	0.79	0.81	0.79	0.68	0.71	0.69	0.86	0.95	0.95	1.00									
CA2bb / CA2a	0.71	0.72	0.71	0.80	0.80	0.79	0.74	0.74	0.74	0.67	0.84	0.83	0.95	1.00								
CA2bb / CA3a	0.73	0.74	0.73	0.82	0.82	0.81	0.76	0.77	0.76	0.68	0.85	0.84	0.96	1.00	1.00							
CA3bb / CA1a	0.62	0.68	0.66	0.78	0.83	0.81	0.66	0.71	0.69	0.96	0.95	0.97	0.90	0.75	0.77	1.00						
CA3bb / CA2a	0.77	0.82	0.80	0.90	0.93	0.91	0.80	0.83	0.82	0.86	0.93	0.95	0.94	0.87	0.89	0.96	1.00					
CA3bb / CA3a	0.78	0.83	0.81	0.90	0.93	0.92	0.81	0.84	0.83	0.86	0.92	0.94	0.93	0.85	0.87	0.96	1.00	1.00				
TR1bb / TR1a	0.33	0.40	0.37	0.54	0.60	0.57	0.38	0.43	0.41	0.98	0.98	0.97	0.90	0.75	0.75	0.91	0.85	0.84	1.00			
TR1bb / TR2a	0.46	0.53	0.50	0.65	0.71	0.68	0.51	0.56	0.54	0.98	0.99	0.99	0.94	0.80	0.81	0.96	0.91	0.91	0.99	1.00		
TR1bb / TR3a	0.42	0.49	0.46	0.62	0.68	0.65	0.47	0.52	0.50	0.98	0.99	0.99	0.92	0.78	0.79	0.95	0.90	0.89	0.99	1.00	1.00	
AG1bb / AG1a	0.56	0.62	0.59	0.74	0.78	0.76	0.61	0.65	0.63	0.95	0.99	1.00	0.97	0.87	0.88	0.96	0.95	0.95	0.96	0.99	0.98	1.00

Figure 2.7: Spectral Correlation Mapper applied to a subset of the reflectance spectra (550 to 600 nm) from figure 2.6 to focus on regions of maximum differences (visually assessed). The black boxes highlight comparisons between the fingerprints of *Pseudo-nitzschia* (PN) and *Chaetoceros* (CA), *Thalassiosira* (TR), and *Asterionellopsis* (AG), in descending order—summarized in table 2.2.

Table 2.2: Mean spectral differences between *Pseudo-nitzschia* and other genera, with standard deviations.

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

2.5 Discussion

We found the spectral fingerprint of *Pseudo-nitzschia* forms a distinctive bifurcated peak between 570 and 600 nm, which easily distinguishes this genus from the others cultured under these ideal conditions. We also found that *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 *Thalassiosisra* is the most abundant bloom-forming diatom in the California Current. The ability to discern between *Thalassiosira* and *Pseudo-nitzschia* based on spectral fingerprints is valuable for a number of applications, e.g. productivity models, mapping species distributions, 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 2.5 and appendix, figure 2) successfully isolated the unique spectral fingerprints of *Pseudo-nitzschia* from those of *Thalassiosira*, *Asterionellopsis*, 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). The potential of automated classifiers to identify unique water masses based on their optical properties, and the scalability of these tools with increased computational power, will be of great use for ocean color applications classifying unique biogeochemical provinces based on spectral signatures. Unsupervised clustering algorithm's success with these diatoms spectral fingerprints suggests an automated bloom-tracking system would be able to identify the most common diatoms, and *Pseudo-nitzschia*, using the PACE satellite.

2.5.1 Importance of backscatter

The experiments described here reveal the importance of integrating spectral backscatter in 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). For an example of what the reflectance fingerprints in figure 2.4 would look like if backscatter was spectrally flat, see figure 2.8. 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 2.4 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* within diatom assemblages.

From this knowledge we assert that absorption-based models, which have dominated the literature thus far, will be unable to identify *Pseudo-nitzschia* remotely in diatom-dominated systems. The absorption coefficients of diatoms are too similar for the current spectral resolution of satellites to discriminate between. Alternatively, the spectral shapes of backscatter between just these four genera presented unique features and shapes (figure 2.2b). Due to the high computational power required to model the spectra of backscatter for phytoplankton with such highly detailed frustules and intracellular structures, we cannot directly attribute each feature within spectral backscatter to a physical property of the phytoplankton with absolute certainty, but we can make several educated guesses. We theorize 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. We expect, and observe, a similar increase in backscattering coefficients of bluer wavelengths for *Thalassiosira*, due to the presence of many small chitinous thread structures which adorn the frustule. Many of these threads are joined together in the center of the frustule to bind *Thalassiosira* cells in chain formations. We propose that chitons' increased absorption in blue-green wavelengths, 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 2.2a 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 diminished compared to natural samples. This may explain why *Asterionellopsis* does not have the same high backscatter coefficients in blue wavelengths as *Chaetoceros* and *Thalassiosira.* That said, both *Chaetoceros* and *Thalassiosira* possess numerous setae or threads per cell, whereas *Asterionellopsis* has only 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 2.2.

The peak in backscatter for all species past 660 nm (figure 2.2a) is likely due to inelastic scatter in the form of fluorescence, rather than true backscatter. This fluorescence signal is likely a factor of the large FWHM (˜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 Qy 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 2.2b versus appendix, figure 1).

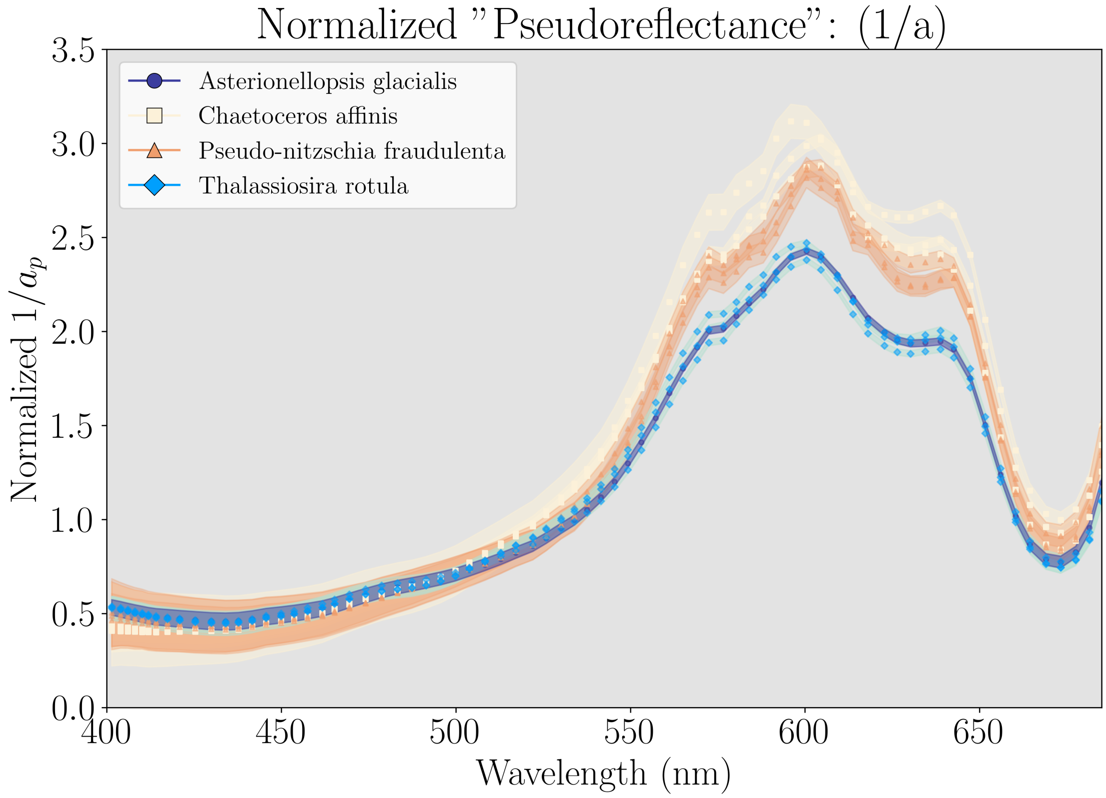


Figure 2.8: What if backscatter was flat? Pseudoreflectances shown assuming bb is spectrally flat for all samples.

2.5.2 Orientation of cells

The backscatter spectra, and derived reflectance fingerprints, of *Pseudo-nitzschia* have large standard deviations, which exceed that of all other phytoplankton sampled. The variability in backscattering coefficients increased with bluer wavelengths. This phenomenon was observed in all taxa. We theorize that these uncertainties may lie in the orientation of the cells. These diatoms were measured in a turbulent environment, and while the surface of the ocean is often assumed to be well-mixed, this is not always the case. During periods of laminar flow, or stratification, we expect phytoplankton to orient themselves to the currents or sunlight, respectively (Nayak et al. 2018, Karp-Boss and Jumars 1998). Aligned orientation may affect the spectrum and intensity of backscattered light (orientation can increase the intensity of backscatter by more than 30% (Marcos et al. 2011)). The backscattering coefficients, and resulting spectral features, for non-spherical particles, are still largely unknown (Clavano, Boss, and Karp-Boss 2007). We do know however, that shape affects scattering coefficients, and that for non-spherical particles, the apparent shape (cross-sectional area, relative to incoming light source and detector) changes with orientation. We hypothesize that the large differences in cross-sectional area between the apical axis and top-down valve-apex view of *Pseudo-nitzschia* contributed to its large variability in backscatter values. The apical axis presents the largest 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 and reduce overall scatter. The difference in apparent size and shape between apical and valve-apex views of *Pseudo-nitzschia* will be greater than that of any other diatom sampled in this experiment. Indeed, due to *Pseudo-nitzschia's* unique method of chain formation, this difference in apparent size and shape, and the proposed resulting variability in backscatter coefficients, would likely be higher than for the vast majority of other marine phytoplankton, excluding a few penates with similar morphologies and behaviors (e.g. Rhizosolenia). In a turbulent medium, e.g. our inline system, orientation is randomized, the direct valve-apex view will be significantly less probable than some angle of the planar view. Future studies of the scattering properties of these plankton under laminar flows or stratified conditions will be of benefit as we expect large differences in the magnitude of backscatter and potentially even spectral shape in different orientations. We expect the largest differences to be in the signature of polarization.

2.5.3 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 (Pearsons' correlation coefficient (r)) and insensitive to changes in magnitude, while still responsive to differences in spectral shape. Spectral Correlation Mapper performed best after applying the second derivative to the reflectance fingerprints as spectral features were accentuated. The quantification of spectral differences is important for establishing baselines to compare ocean color and modeled reflectance fingerprints.

The low standard deviation in mean reflectance fingerprints for *Pseudo-nitzschia* (1%) 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 49%. When comparing ocean color to the molded fingerprints of *Pseudo-nitzschia*, the threshold for similarity should therefore be greater than 70%. Ocean color datasets can be flagged for pixels with high spectral similarity to the modeled reflectance fingerprints, providing a scalable framework.

By using distance metrics, we reduced a spectrum, or a comparison of spectra, to one number. In doing so, we limit the information that can be derived. This reductionist approach is a double edge sword, facilitating easy cross-comparisons and data processing, but potentially removing key sources of information. Relying solely on distance metrics may be naive.

We theorize that distance metrics will work well under bloom-conditions when the optical properties of the dominant phytoplankton are the primary contributor to ocean color. In this regard, after removing the reflectance spectrum of pure water from the ocean color signal, we should have a spectrum that matches closely to the modeled fingerprints - if the bloom is dominated by the most abundant taxa found in the system, or *Pseudo-nitzschia*.

However, there will be greater spectral variation in natural samples due to mixed assemblages and other optically active components (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 (creating larger differences). Fortunately, the spectral shapes of CDOM and detritus are often featureless, mononic functions (or so we assume!) which should not greatly influence the retrievals of our target phytoplankton. Furthermore, these components have the greatest influence at ultraviolet and blue wavelengths - spectral regions which we disregard 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. Fortunately, these spaces encompass but a small fraction of the total area of our system in which this framework should be used (the California Current).

Therefore, we anticipate the primary obstacle in diatom identification will be due to competing optical properties from 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 by further analysis into the spectra of pixels that do not pass the threshold similarity. This should be done by first using a clustering algorithm to subset the region of interest into 'k' bins with similar spectral shapes - so that one does not have to look at every single pixel, but a representative sample. From these k-bins a visual analysis, or simple band ratios near 570, 585, and 600 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 2.4 should be present. If k is very large and visual assessment of all bins is impractical, band ratios of 570:585 nm and 585:600 nm can be used to search for the bifurcated peaks. In situ preliminary reflectance data of mixed assemblages containing *Pseudo-nitzschia* have displayed the distinct bifurcated peaks to be present even in the face of many competing optical signals. Future studies may investigate the concentrations required for, and the detection limits of, these plankton with respect to community assemblages. These future studies and others measuring hyperspectral backscatter will be useful for the development of hyperspectral libraries and inverse models. 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 serve to further constrain uncertainties in all ocean color models - including this study.

2.5.4 Conclusion

Diatoms provide a wide range of ecosystem services and sustain commercial fisheries success in upwelling systems. But there is one diatom, *Pseudo-nitzschia*, that has the potential to shut down these fisheries over large spatial scales. Marine resource managers, commercial fishermen, aquaculturists, and coastal communities need a framework to identify when and where *Pseudo-nitzschia* is blooming to mitigate economic losses and ecosystem damage. Current practices to monitor *Pseudo-nitzschia* are costly and spatially fragmented. Hyperspectral remote sensing platforms, which can continuously sample large areas, have been proposed as tools to identify harmful algal blooms, but their ability to distinguish genera within groups was previously uncertain. Our models show that genus-specific identification and HAB monitoring using hyperspectral remote sensing platforms is possible. This capability will be used to inform local agencies of HAB events in near-real time and facilitate targeted in situ sampling of toxin levels.

This work quantified the differences in spectral shapes of key diatom genera. The pronounced differences enable the distinction of the most common genera in the California Current—*Thalassiosira*, *Chaetoceros*, and *Asterionellopsis*—from the toxigenic *Pseudo-nitzschia* using ocean color data. This work is the first to show the unique reflectance fingerprints for the dominant primary producers in one of the world's most productive regions, the California Current. It is also the first to quantitatively discern the degree of difference between them. This was only possible due to the high spectral resolution of new hyperspectral sensors, which measure significantly more wavelengths of light than their historic counterparts. The continued development of comprehensive hyperspectral libraries will enhance our ability to identify phytoplankton species and detect HABs remotely and over broad spatial scales. While field-based studies will be necessary to validate the fingerprints derived from these experiments, our data provide evidence that the significant differences in the spectral shapes of these key plankton are sufficient to discern between them using the methods provided. These results offer a new approach for monitoring biodiversity and detecting harmful algal blooms.

Chapter 3: Reflections and Future Directions

3.0.1 In Situ Validation

The next step forward will undoubtedly be thorough testing of the proposed model's skill through in-situ validation. These validations will require taking water samples, then enumerating and identifying the phytoplankton taxa therein. This will provide an estimate of the phytoplankton community composition and allow us to determine *Pseudo-nitzschia*'s potential abundance in these samples.

As the water samples are taken, concurrent measurement of the water body's optical properties should be recorded. This can be done in a number of ways. First, the user can record the *inherent optical properties* of the water. Inherent optical properties (IOPs) include absorption, and backscatter. The absorption and backscatter coefficients do not change with the surrounding light field – rather these are related to the *inherent* efficiency, or ability, of a substance (a molecule of water, or a *Thalassiosira* cell) to absorb or scatter light. Theoretically, the IOPs of a substance remain constant as long as that substance remains in the same state (e.g. temperature and salinity for a water body, physiological state for a cell).

Second, the user can use an above, or in-water radiometer to measure the reflectance of the water body directly. This is essentially using a hyperspectral camera to take a picture of the water – and record the spectra of light that is being emitted from the surface. If this is done, the spectrum of light from the sun and sky must also be recorded as this version of in-situ reflectance is an *apparent optical property*. This means that the spectrum of light leaving the water's surface is influenced by the apparent light field – which is the light field of the sun and sky. This apparent light field changes with cloud cover, time of day, and atmospheric aerosols.

The last option is the most rewarding – but also challenging. Measuring the reflectance of the in-situ sample using a remote sensing platform. The reward scales with difficulty. **Easiest**: drones are affordable, easy to deploy, and modular – as they can be fitted with a variety of radiometers. However, drones only allow us to sample relatively small sections of the nearshore environment. **Intermediate**: low-flying aircraft have a larger spatial footprint and can cover offshore waters. However, these operations are costly and take significant effort to organize. **Most difficult**: ocean color satellites cost billions of dollars and can take decades of planning to materialize. Once in orbit, these satellites still require millions of dollars in calibration and validation efforts to produce reliable data – which must be corrected for atmospheric influence.

The reward? Global coverage of ocean reflectance every few days. In particularly cloudy systems (such as the Pacific Northwest), drones and low flying aircraft may still be advantageous during episodic upwelling-driven blooms, but satellites, such as NASA PACE, provides data at no cost to the user. The no-cost entry to use PACE ocean color products makes this system particularly well suited for an accessible and scalable HAB early warning system.

The collection of optical data using any of these three methods in combination with taxonomic classification of the phytoplankton community will allow us to validate the relationship between *Pseudo-nitzschia* and reflectance found in chapter 2. The spectra of water samples containing *Pseudo-nitzschia* should be analyzed for the bifurcated peak displayed in figure 2.4. Since all taxa measured had a peak in reflectance at 600 nm, a ratio of the reflectance values at ∼560 and 585 nm should suffice to identify the *Pseudo-nitzschia*-specific optical fingerprint. Ratios greater than 1 will be indicative of *Pseudo-nitzschia* presence. In our preliminary validation results (not shown), we have subtracted the modeled reflectance signature of pure water from our in-situ samples. This allows us to detect *Pseudo-nitzschia* in low concentrations where the dominant optical signals are from the water itself.

3.0.2 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, along with its 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 at 488 to 555 nm, while the intensity of reflectance at 555 nm is coupled with a reflectance-derived chlorophyll product to estimate toxin concentrations. These remote sensing products and bands, while useful correlative variables within the Southern California Current, are not strictly tied *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 in chapter 2 – providing an optical proxy directly tied to the inherent optical properties of *Pseudo-nitzschia* themselves. We hope that the findings of chapter 2 can therefore serve to augment the current C-HARM model by integrating these newly discovered hyperspectral bio-optical fingerprints through PACE data. This may increase C-HARM's skill in the Northern California Current (Oregon, Washington, British Columbia). This is especially important because these temperate zones must be sampled for *Pseudo-nitzschia* and domoic acid much more frequently than Central and Southern California due to strong seasonality and high variability (Frolov, Kudela, and Bellingham 2013). This direct spectral pathway for detection may further allow the extrapolation of this approach to other systems where *Pseudo-nitzschia* is present in harmful concentrations.

Bibliography

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://www.sciencedirect.com/science/article/pii/019801499290002B` (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, D. M., P. Hoagland, Y. Kaoru, and A. W. White (Sept. 2000). *Estimated AnnualcEconomic 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., 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/`.

Behrenfeld, M. J. and P. G. 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. J., 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).

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 phytoplankters1”. 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., 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).

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).

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).

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://aslopubs.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, CaCO3, 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/CBO9781139168212`. 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).

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. Luhar, 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).

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`.

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).

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`.

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).

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).

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: `10.1175/2010JTECHO755.1`. URL: `https://journals.ametsoc.org/view/journals/atot/27/10/2010jtecho755_1.xml` (visited on 04/24/2025).

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).

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).

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).

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).

APPENDICES

Additional Distance Metrics Additional distance metrics we 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, or 'fingerprints'. 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 which we can draw a line to 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° 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° angular difference (0.99 - 1 = 1% difference in cosine similarity space).

Deborah et al. 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 Savityzky-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. This means the dominant parameter in our distance metric is standard deviation, 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, 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 importance on the bluer wavelengths. The integration can be flipped to place more importance on the red wavelengths, but this means that the derived differences between spectra will change depending on how the metric is integrated. Kullback-Leibler pseduo-divergence is a variation on the standard Kullback-Leibler divergence function, which is used for probability density functions, hence *pseudo*-divergence. Kullback-Leibler pseduo-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:

$$d_{\mathrm{ECS}}(S_1, S_2) = \left(\int_{\lambda_{\mathrm{min}}}^{\lambda_{\mathrm{max}}} \left(s_1'(\lambda) - s_2'(\lambda) \right)^2 d\lambda \right)^{\frac{1}{2}} \tag{1}$$

$$s_i'(\lambda_c) = \int_{\lambda_{\min}}^{\lambda_c} s_i(\lambda) d\lambda \ , \tag{2}$$

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:

$$div_{KL'}(S_1, S_2) = k_1 \cdot \mathrm{KL}(\bar{S}_1, \bar{S}_2) + k_2 \cdot \mathrm{KL}(\bar{S}_2, \bar{S}_1) + (k_1 - k_2) \log\left(\frac{k_1}{k_2}\right) \tag{3}$$

$$\text{shape} : k_1 \cdot \mathrm{KL}(\bar{S}_1, \bar{S}_2) + k_2 \cdot \mathrm{KL}(\bar{S}_2, \bar{S}_1) \tag{4}$$

$$\text{energy or intensity} : (k_1 - k_2) \log\left(\frac{k_1}{k_2}\right) \ . \tag{5}$$

Both ECS and KLPD performed well (intertaxon differences were large, while intrataxon differences were small) but are only 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. ECS 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 5 we can see three red hotspots; one in the top left, the center-bottom-right, and nearly the bottom

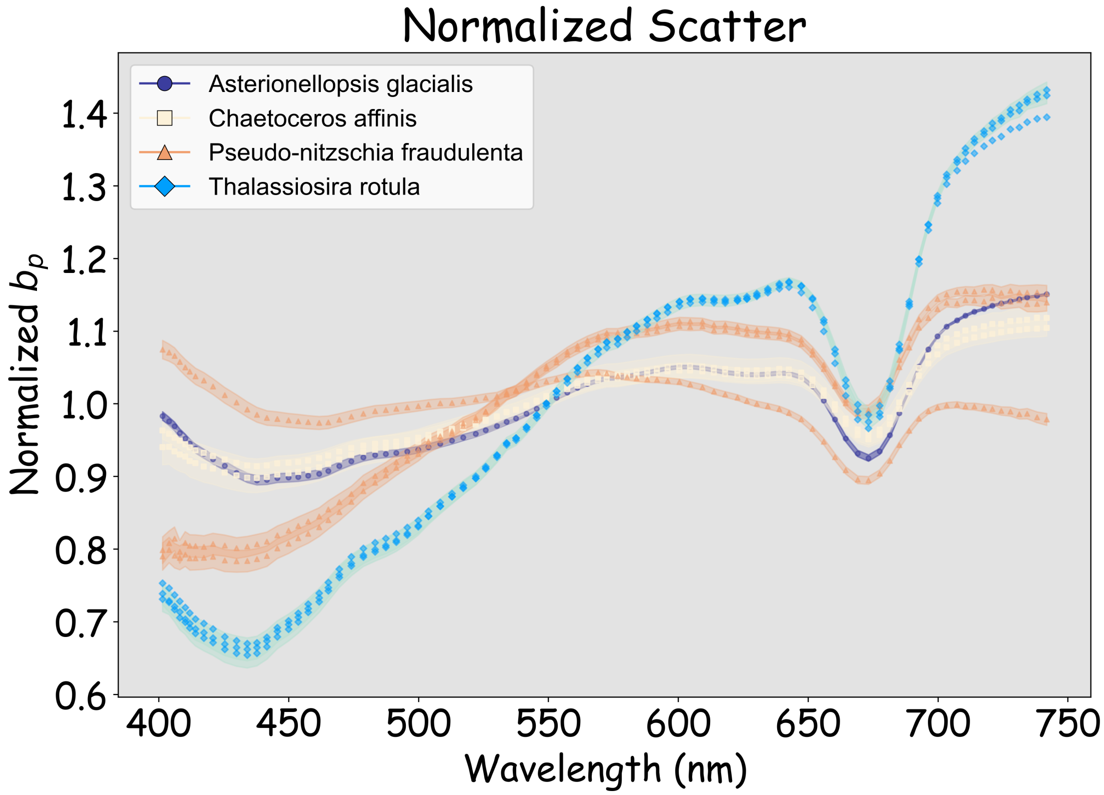


Figure 1: Normalized scatter spectra for all species
. All species have scatter spectra that visually appear to be more similar intraspecifically than interspecifically. The standard deviations are low and similar across species. PN_3 is different from other triplicates due to a pump failure (the more flat, orange line). These spectra of total scatter mimic an inversion of the absorption spectra.

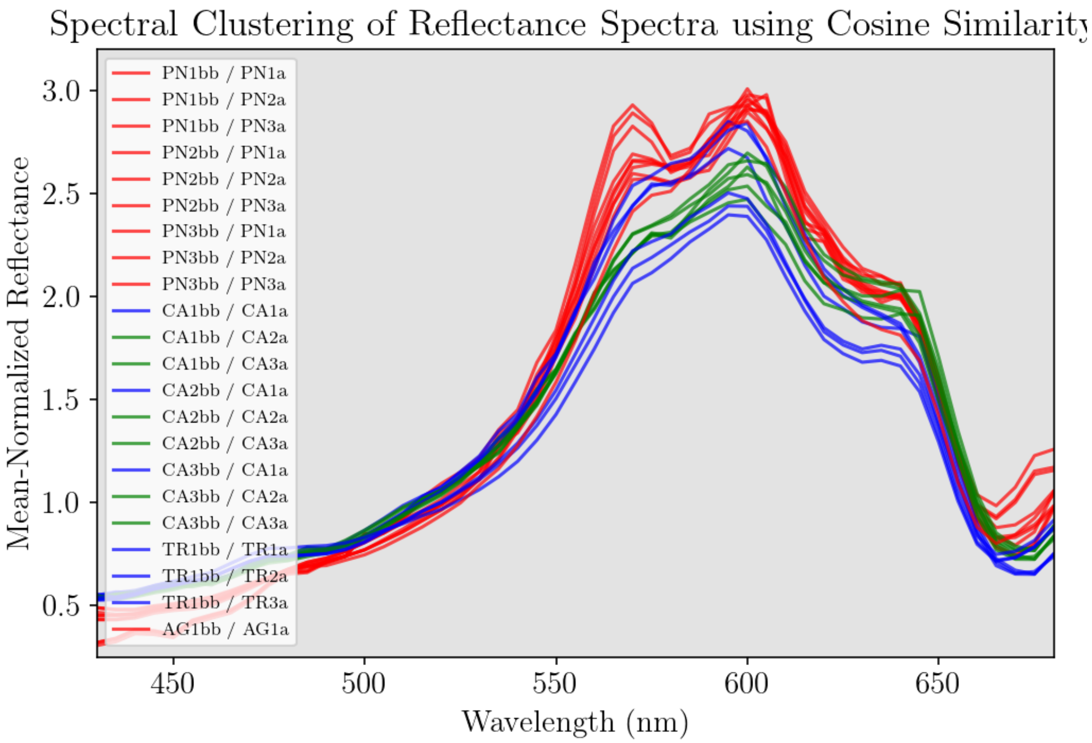


Figure 2: Spectral clustering: unsupervised machine learning using cosine similarity on all raw spectra. Cluster number set to 3. The absorption spectra of the first triplicate of Chaetoceros (CA1a) is misclassified as Thalassiosira (TR) for all bootstrapped spectra. Asterionellopsis is misclassified as Pseudo-nitzschia.

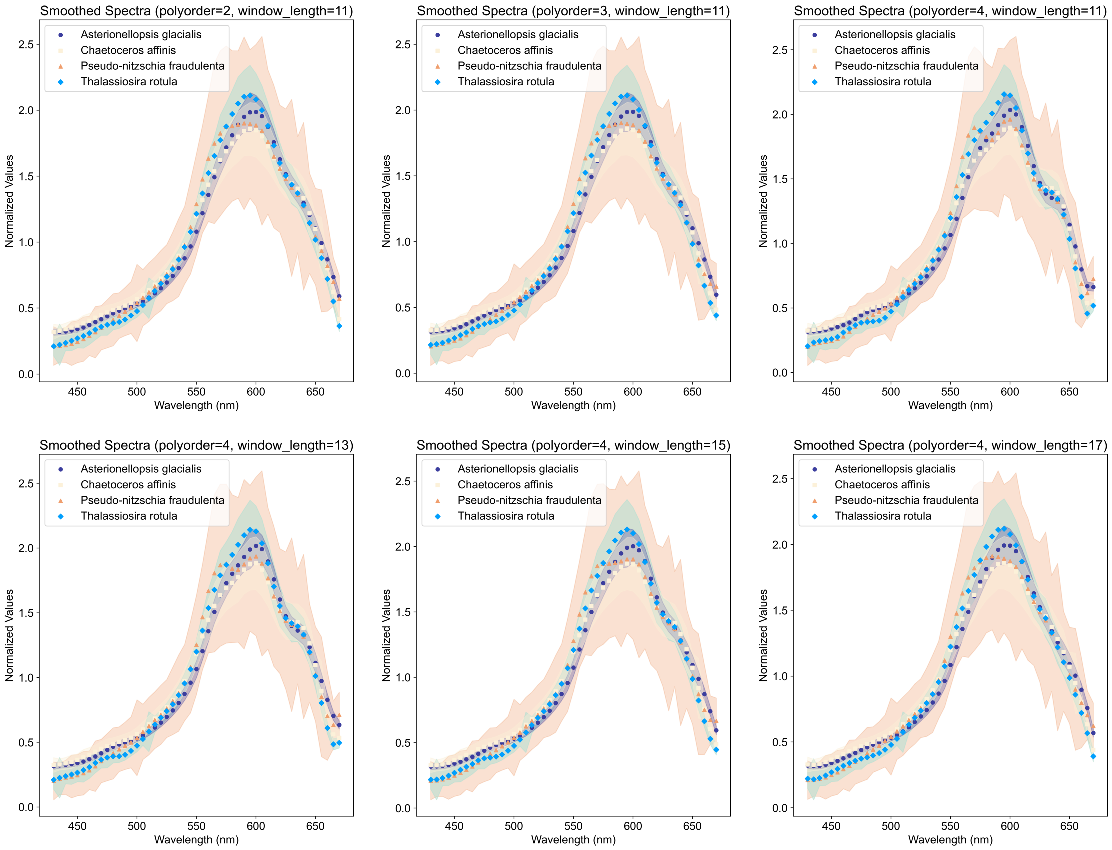


Figure 3: Iterations of the Savitzky-Golay smoothing algorithm applied to the reflectance spectra. The top row shows iterations of the smoothing function varying the polynomial order with a fixed window length of 11 (55 nm), while the bottom row shows iterations changing the window-length from 13 (65 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.

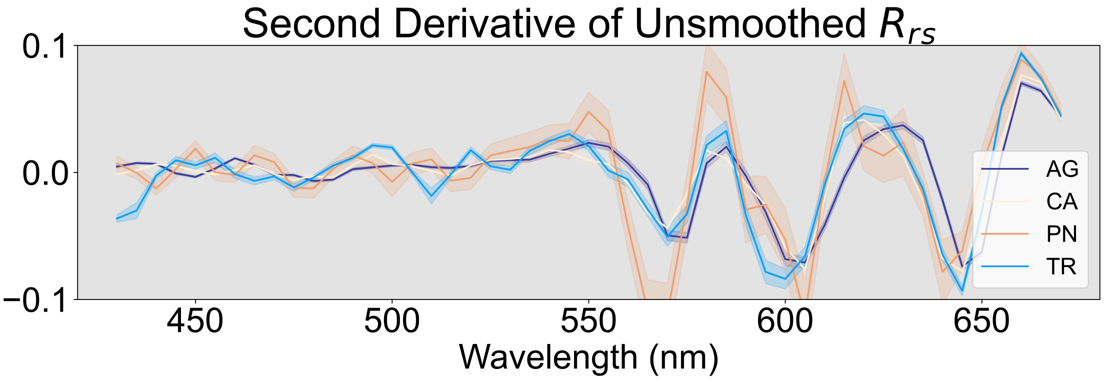


Figure 4: Second derivative of unsmoothed (raw) spectra
, one sample taken from each taxon. Note the accentuation of 'features' in wavelengths shorter than 550 nm. Whether these are real spectral features that we can glean information from or noise is unknown. By taking the mean of many samples, noise should be removed, but smoothing is a cautionary step to ensure we are not mistaking noise for 'true' signal.

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 less ¡ 1%, these small differences were still sufficient to successfully cluster most of the reflectance fingerprints into their respective taxonomic groups (figure 2). Both ECS and KLPD perform best on the full raw spectra, likely 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.

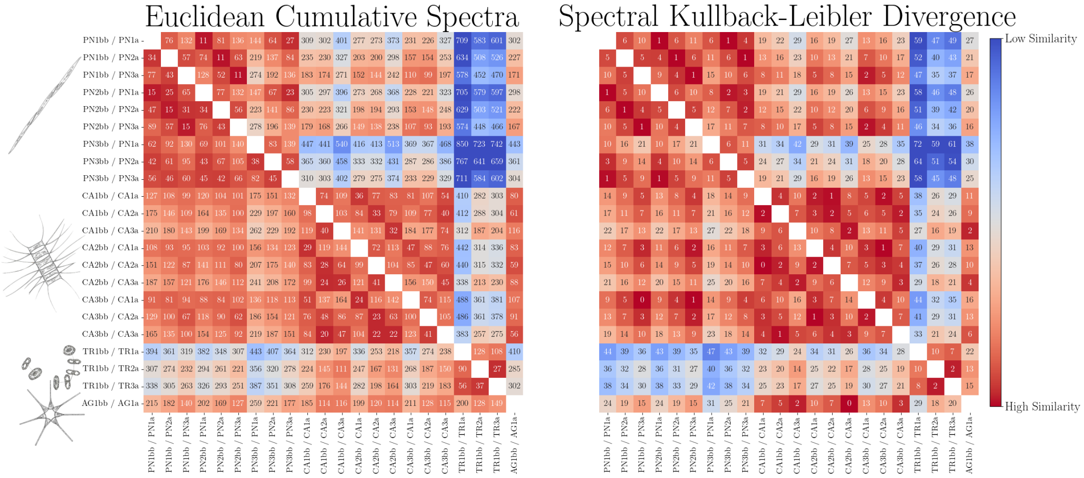


Figure 5: This figure displays two confusion matricies for the ECS and KLPD fucntions. There is no smoothing or alternate processing. Note the asymmetry in these heatmaps due to differences in integration for ECS (bottom is integrating from blue to red, top is integration from red to blue) and the triangular inequality of KLPD - the triangular inequality means we expect different results based on order of comparison (i.e. spectra a vs spectra b $\neq$ spectra b vs spectra a). It is important to note that KLPD performs better when comparing spectra of different diatoms against *Pseudo-nitzschia*, rather than comparing *Pseudo-nitzschia* spectra against the 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.