2018

CHARACTERIZING CDOM SPECTRAL VARIABILITY FROM SEAS TO SPACE

Brice Grunert
Michigan Technological University, bgrunert@mtu.edu

CHARACTERIZING CDOM SPECTRAL VARIABILITY FROM SEAS TO SPACE

By

Brice K. Grunert

A DISSERTATION

Submitted in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

In Geology

MICHIGAN TECHNOLOGICAL UNIVERSITY

2018

This dissertation has been approved in partial fulfillment of the requirements for the Degree of DOCTOR OF PHILOSOPHY in Geology.

Department of Geological and Mining Engineering and Sciences

Dissertation Advisor:	*Dr. Colleen B. Mouw*
Committee Member:	*Dr. Joaquim Goes*
Committee Member:	*Dr. Evan S. Kane*
Committee Member:	*Dr. Alex Mayer*
Department Chair:	*Dr. John S. Gierke*

Table of Contents

Author contribution statement vi

Acknowledgements vii

Abstract viii

1 Optically estimating CDOM composition across diverse regions and spectral ranges 1

1.1 Introduction 1

1.2 Methods 3
1.2.1 CDOM Absorption Spectra 3
1.2.2 Data Distribution 5
1.2.3 Analysis of CDOM Metrics 7

1.3 Results 8
1.3.1 Comparison of CDOM Models 8
1.3.2 Spatial Trends in Slope Values 10
1.3.2.1 $S_{E275:295/E350:400}$ 10
1.3.2.2 $S_{G240:700}$ 10
1.3.2.3 $S_{G300:700}$ 10
1.3.2.4 $S_{G350:550}$ 11
1.3.2.5 Spatial Relationships 11
1.3.3 Gaussian Components 11
1.3.3.1 Spatial Trends in Gaussian Components: $S_{G240:700}$ 13
1.3.3.2 Spatial Trends in Gaussian Components: $S_{G300:700}$ 13
15
17
1.3.3.3 Spatial Trends in Gaussian Components: $S_{G350:550}$ 18
1.3.3.4 Spatial Trends in Gaussian Component Spectral Location: $S_{G240:700}$ 18
1.3.3.5 Spatial Trends in Gaussian Component Spectral Location: $S_{G300:700}$ 19
1.3.3.6 Spatial Trends in Gaussian Component Spectral Location: $S_{G350:550}$ 19
1.3.3.7 Trends in Gaussian Component Metrics 21
1.3.4 CDOM Metric Comparisons 21
1.3.4.1 Comparison of Broad Range Slopes to $S_{E275:295}/S_{E350:400}$ 21
1.3.4.2 Comparison between a_{CDOM} and S_{CDOM} 21

1.4 Discussion 22
1.4.1 CDOM Models 22
1.4.2 S_{CDOM} 23
1.4.3 Gaussian Components 26

1.4.4 Applications to Remote Sensing 29
1.5 Conclusions 30
1.6 References 31

2 Sensitivity of hyperspectral reflectance to colored dissolved organic matter spectral variability 38
2.1 Introduction 38
2.2 Methods 40
2.2.1 Study Site 40
2.2.2 Data 40
2.2.3 HydroLight 41
2.2.4 Data Analysis 42
2.3 Results 43
2.3.1 Field Data 43
2.3.2 HydroLight Simulations 45
2.3.2.1 Simulated $R_{rs}(\lambda)$ 45
2.3.2.2 S_g Sensitivity 47
2.3.3 Observed S_g Variability 48
2.4 Discussion 48
2.5 Conclusions 53
2.6 References 54

3 Deriving inherent optical properties from decomposition of hyperspectral non-water absorption 62
3.1 Introduction 62
3.2 Methods 64
3.2.1 Data 64
3.2.2 DAISEA Algorithm Development 65
3.2.2.1 Low $a_{ph}(\lambda)$ waters 74
3.2.3 Data Analysis 74
3.3 Results 75
3.3.1 DAISEA Performance 75
3.3.2 S_{dg} Estimation 78
3.3.3 Consistency in Gaussian Features 80
3.4 Discussion 81
3.4.1 DAISEA 81
3.4.1.1 Application 81
3.4.1.2 General Framework and Empirical Relationships 82
3.4.2 CDOM 85
3.4.3 Phytoplankton 87

3.4.4 NAP 89
3.5 Conclusions 90
3.6 References 91
A Copyright documentation 101

Author contribution statement

Chapter 1 was previously published in *Global Biogeochemical Cycles*. Chapter 2 will be submitted and Chapter 3 has been submitted to *Remote Sensing of Environment*. Pending acceptance, these chapters in their published form should be considered as the authoritative version. All contents of the dissertation were primarily authored by B. K. Grunert with assistance from Dr. Colleen Mouw, Audrey Ciochetto and committee members.

Acknowledgements

Thank you to NASA for generously funding this work through a Earth and Space Science Fellowship. Colleen, thank you for teaching me more than just how to answer research questions – from being a committed teacher to running a lab to effectively networking and creating connections in our field, you've been an incredible mentor. To both Colleen and Audrey, you have been absolutely crucial to the success of my career, both in assisting my research endeavors and leading by example. I feel fortunate to have a sound approach to science as the foundation of my career thanks to your mentoring and collaboration. Joaquim, thank you for always being willing to help advance my career and getting me to think about my research in new ways. Evan, thank you for broadening my view of dissolved organic matter, making me think hard about the chemistry and getting me involved with other Tech researchers. Alex, thank you for broadening my view, mentoring me on how to effectively manage research projects and being a phenomenal instructor – if I ever become a good teacher, I'll be sure to thank you again.

Thank you to Kelly, Brittany and Dr. Gierke for always being willing to help and taking care of all the things that weren't my research. My life has been far easier thanks to their efforts. To all my department colleagues for support, thank you.

Finally, a huge shout out to my family. Erika has been incredible and never complains through all the ups and downs. She's allowed me to have a wonderful family while prioritizing my career with time away for research cruises, conferences and workshops (just ask her about San Francisco!). Angie is always ready to help and never complains about long hours or time away. Ray and Seby, thank you for always greeting me with a smile, besos y brazos. Mom and Dad, I can't thank you enough for getting me where I am today.

Abstract

Colored dissolved organic matter (CDOM) absorption varies significantly across the global oceans, presumably due to differences in source and degradation pathways. Tracking this variability on a global, or even regional, scale requires broad temporal and spatial sampling at high frequency. Satellite remote sensing provides this platform; however, current and near future sensors are/will be limited to measurements within the UV and visible wavelengths (> 350 nm) while most optical proxies estimating CDOM composition, and relevant for understanding largescale biogeochemical processes, use wavelengths less than 350 nm. This dissertation examines global variability in CDOM spectral variability utilizing a variety of optical metrics. After assessing global variability in these optical metrics, we considered the ability to observe changes in remotely-sensed reflectance ($R_{rs}(\lambda)$) strictly due to S_g variability. Using the radiative transfer software, HydroLight, and data from Lake Superior, modeled $R_{rs}(\lambda)$ showed that S_g variability significantly alters $R_{rs}(\lambda)$ in waters where $a_g(\lambda)$ contributes >20% to total non-water absorption ($a_{t\text{-}w}(\lambda)$) at 440 nm. Based on the proposed signal-to-noise ratio of NASA's proposed Plankton, Aerosol, Cloud and ocean Ecosystem (PACE) hyperspectral sensor, S_g variability on the order of 0.001 nm^{-1} is an observable feature in these waters. We then developed an capable of estimating S_{dg} free of bias on hyperspectral absorption data. The algorithm shows that the increased spectral resolution of hyperspectral sensors should allow for remote estimation of S_{dg} and potentially S_g, providing a broad view of biogeochemical variability reflected by S_g.

1 Optically estimating CDOM composition across diverse regions and spectral ranges

1.1 Introduction

Dissolved organic matter (DOM) is the largest pool of reduced carbon in the ocean [*Hansell et al., 2009*]. Slight alterations in DOM remineralization can result in significant increases or decreases in the release of CO_2 [*Mannino et al.*, 2014] with remineralization rates highly dependent on DOM composition. DOM constituents vary from labile, readily metabolized compounds (e.g. phytoplankton exudates) important to short-term climate dynamics and food web processes to refractive compounds (e.g. microbial by-products) de-coupled from modern climate dynamics and largely precluded from food web dynamics [*Hansell*, 2013]. DOM contains optically active components in colored and fluorescent dissolved organic matter (CDOM and FDOM, respectively), with CDOM a small, variable portion of the total DOM pool and FDOM a small, variable portion of the CDOM pool [*Stedmon and Nelson, 2015*]. CDOM impacts the spectral quality and intensity of the underwater light field, surface ocean heating and plays direct and indirect roles in biogeochemical cycling [*Andrew et al.*, 2013; *Hickman et al.*, 2010; *Kim et al.*, 2016].

Advances in optical characterization through absorption and fluorescence techniques have illuminated mechanistic relationships between CDOM molecular composition and optical properties such as the shape of the CDOM absorption spectra indicated by the spectral slope of CDOM (S_{CDOM}) [*Helms et al.*, 2008; *Stedmon and Markager*, 2005; *Walker et al.*, 2013]. Changes to CDOM optical characteristics have been related to general classifications of the composition of CDOM including molecular weight and origin [*Spencer et al.*, 2008]. The strongest absorption-based indicators of these properties rely on S_{CDOM} from 275-295 nm and 350-400 nm ($S_{275:295}$ and $S_{350:400}$, respectively) since the largest divergence in spectral properties across compositionally unique CDOM samples are found in these spectral regions [*Helms et al.*, 2008]. Low values of $S_{275:295}$ (~0.01 nm^{-1}) generally represent high molecular weight material (e.g. lignin) that decreases in molecular size primarily from photodegradation and secondarily from microbial degradation [*Del Vecchio and Blough*, 2002]. $S_{275:295}$ increases with photodegradation, with terrestrial material often showing markedly different $S_{275:295}$ values due to the presence of lignin. Conversely, $S_{350:400}$ generally decreases with photodegradation, potentially due to the presence of photorefractory chromophores that absorb in this portion of the spectrum [*Helms et al.*, 2013].

Beyond $S_{275:295}$ and $S_{350:400}$, consistency is lacking in the spectral range used to calculate S_{CDOM}. Typically, a spectral range that best represents the entire a_{CDOM} spectra is used, as broadening the spectral range considered tends to decrease overall error in representing the a_{CDOM} spectrum [*Schwarz et al.*, 2002]. However, broad range S_{CDOM} (e.g. $S_{240:700}$) has not been considered extensively for mechanistic relationships with CDOM composition. Rather, most studies track S_{CDOM} values across geographically or temporally unique CDOM pools allowing for general statements about the diagenetic state of CDOM

across different environments [*Organelli et al.*, 2014; *Stedmon and Markager*, 2003]. Some work has suggested that $S_{300:600}$ is a useful indicator of CDOM molecular weight [*Stedmon and Nelson 2015*], while the general understanding is that broad range S_{CDOM} increases with photodegradation as chromophores are bleached, particularly in the UV region [*Del Vecchio and Blough*, 2002; *Reche et al.*, 2000]. Optical properties of source material have been considered [*Hansen et al.*, 2016; *Visser*, 1983]; however, much of the literature relies on optical observations of CDOM in a specific geographic region and relates the optical signal to known source waters without clearly disentangling the molecular basis of the optical signal or diagenetic state of the CDOM. Two major components of CDOM, humic acid and fulvic acid, display slopes of approximately 0.011 nm^{-1} and 0.019 nm^{-1} [*Carder et al.*, 1989]. However, observed slope values across broad spectral ranges can vary above and below these benchmarks, suggesting that the complexity of material and its signal requires further attention. If possible, it is important to distinguish low molecular weight material such as phytoplankton exudates readily reincorporated into the food web from low molecular weight, recalcitrant material based on optical data and calculated metrics (e.g. $S_{275:295}$) as these compounds constitute different roles in food web processes and carbon cycling.

CDOM has been quantified from satellite remote sensing by estimating CDOM absorption (a_{CDOM}) using empirical algorithms, semi-analytical algorithms or optimization routines such as neural networks [*Ioannou et al.*, 2013; *Mannino et al.*, 2008; *Matsuoka et al.*, 2013; *Stedmon et al.*, 2011]. Past work has shown relatively strong relationships between a_{CDOM} and common CDOM compositional metrics, including $S_{275:295}$ and lignin content [*Fichot et al.*, 2016; *Mannino et al.*, 2014]. Additionally, a_{CDOM} shows a tight correlation with dissolved organic carbon (DOC) at the terrestrial-aquatic interface [*Fichot and Benner*, 2011], resulting in reliable relationships for estimating DOC delivery to coastal oceans from major river systems via ocean color remote sensing [*Matsuoka et al.*, 2013]. However, this relationship rapidly deteriorates as the terrestrial component of CDOM, shown to behave semi-conservatively with mixing [*Stedmon and Markager*, 2003], diminishes and gives way to *in situ* processes that both degrade and produce CDOM. Open ocean environments do not display a relationship between CDOM and DOC [*D'Sa and Kim*, 2017; *Nelson et al.*, 2010], while the percent contribution of CDOM to the total DOM pool can be quite variable across environments but is generally relatively small in the global oceans [*Nelson and Siegel*, 2013].

Maximizing CDOM compositional information that can be retrieved by optical methods will enhance our ability to track changes in the CDOM pool through autonomous and remote sensing platforms, providing a way to observe large scale changes in ocean biogeochemical processes and circulation [*Nelson et al.*, 2010]. However, this first requires determination of what information regarding CDOM composition is garnered from a specific spectral range and fully utilizing the information contained within a_{CDOM} spectra to optically estimate CDOM composition. To this end, Massicotte and Markager [2016] introduced a new methodology for fitting deviations from the traditional exponential model often observed in a_{CDOM} spectra that are attributed to specific

chromophores. Additionally, this method allows for a more accurate basis of the S_{CDOM} parameter as it better fits the entire a_{CDOM} spectra.

Here, we utilize a_{CDOM} spectra available from public data repositories to address knowledge gaps that have evolved from an inconsistency in spectral range used. We relate several broad range S_{CDOM} to $S_{275:295}$, $S_{350:400}$, and a_{CDOM} at 350 nm ($a_{CDOM}(350)$) to best interpret and define the efficacy of each of these parameters to estimate CDOM composition from a_{CDOM} based on observations in the literature. We utilize the Gaussian decomposition approach of Massicotte and Markager [2016] to further characterize the CDOM pool and report differences between this approach and the standard exponential model traditionally used for fitting a_{CDOM} spectra. Finally, we offer insight into what can be determined about CDOM composition from $S_{350:400}$ and $S_{350:550}$ to step toward defining a methodology that can be applied to hyperspectral, satellite remotely-sensed retrievals of a_{CDOM}.

1.2 Methods

1.2.1 CDOM Absorption Spectra

CDOM absorption spectra were obtained from NASA SeaWiFS Bio-optical Archive and Storage System (SeaBASS, https://seabass.gsfc.nasa.gov/) on April 13, 2016 [*Werdell et al.*, 2003]. Only data measured on a bench top spectrophotometer were utilized. The CDOM absorption spectrum is typically modeled using an exponential decay function:

$$a_{CDOM}(\lambda)=a_{CDOM}(\lambda_0)e^{-S_E(\lambda-\lambda_0)}+K \qquad (1)$$

where λ_0 represents a reference wavelength for initializing the spectrum (nm), S_E is the spectral slope coefficient for an exponential CDOM model (nm^{-1}), and K is a constant addressing background noise and potential instrument bias (m^{-1}), calculated as the average a_{CDOM} from 690-700 nm (Fig. 1). Spectra were quality controlled by selecting spectra representing a broad quasi-exponential function: if the fit with Equation 1 displayed an $r^2 > 0.9$ (suggesting no contamination or instrumentation issues), spectra were kept for further analysis [*Massicotte and Frenette*, 2011].

Deviations from the decaying expression in Eq. 1 can result from absorption by specific chromophores (e.g. lignin), and can vary in degree and location [*Massicotte and Markager*, 2016]. To detect regions where such deviations exist, spectra were fit with the standard exponential in Eq. 1 and then trimmed by removing data with residuals greater than the mean absolute residual multiplied by the weighting factor, *C*. *C* is dependent on the quality of the data (signal to noise ratio), the environment and spectral range considered and directly influences the number of residuals considered for Gaussian fitting by defining the threshold for excluding these points from the baseline exponential fit; we utilized a *C* of 1 to maintain consistency across spectra. With the remaining points, spectra were again

fit with Eq. 1 with the result defined as the baseline exponential. Massicotte and Markager [2016] utilized a spectral range from 240-700 nm with Gaussian fitting performed on residuals from 250-500 nm. The influence of spectral range on the value of S_{CDOM} for an exponential model has been well established [e.g. Twardowski et al. 2004]. To determine the influence of spectral range on S_{CDOM} using the Gaussian decomposition model, we considered spectral ranges of 240-700, 300-700 and 350-550 nm, resulting in spectral ranges of 240-500 nm, 300-500 nm and 350-500 nm for fitting Gaussian components. We found fitting beginning at 240 nm for spectral ranges below 300 nm fit absorption deviations likely due to lignin or lignin-derivatives better than beginning at 250 nm, the only significant deviation from the methodology as described by Massicotte and Markager [2016]. The optimal number of Gaussian components modeled for each spectrum was chosen based on minimizing the Bayesian information criterion (BIC) score, with the final equation represented as:

$$a_{CDOM}(\lambda)=a_{CDOM}(\lambda_0)e^{-S_G(\lambda-\lambda_0)} + K + \sum_{i=0}^{n} \varphi_i e^{-\frac{(x-\mu_i)^2}{2\sigma_i^2}} + \varepsilon \qquad (2)$$

where S_G (nm^{-1}) is the spectral slope coefficient for a Gaussian decomposition CDOM model, φ (m^{-1}) is the height of the Gaussian peak, μ (nm) is the position of the center of the peak, σ (nm) is the width of the peak and ε are the residuals after fitting of the full model. An example spectra fitted with and without the Gaussian decomposition approach is shown (Fig. 1), illustrating the change in slope. For this approach, fitted K as described for Eq. 1 is also used as a threshold for the minimum height of fitted Gaussian components to add certainty that the Gaussian components are fitting chromophores and not instrument noise. We also acknowledged that many different instruments with varying accuracy were used to obtain this large data set. Considering this, we calculated the mean K value across all measured spectra (0.0004) and applied this as a threshold value for fitting Gaussian components: if the spectra-specific K value was smaller than the mean K, the mean K was used for fitting Gaussian components; the spectra-specific K value was used for all other spectra.

The inclusion of K in equations 1 and 2 is often debated, with many researchers considering it an unknown variable that has no physical basis. Others have pointed out its ability to account for instrument noise in regions of the spectrum that should have no, or very little, signal from CDOM absorption. We have included it in this analysis for several reasons: 1) many of the spectra that we considered had some noise at longer wavelengths (> 600 nm) within an otherwise good absorption spectra, suggesting some instrument noise in measuring a very low signal; 2) our analysis focuses primarily on the shape of the spectra, which is not altered by including a K term; 3) our a_{CDOM} analyses are focused on shorter wavelengths (e.g. 350 nm) where the percent contribution of the K term to the overall signal analyzed is generally quite small (mean of 0.6%); and 4) K was used as the threshold for fitting Gaussian components, leading to stricter fitting.

Spectral slope was retrieved for the following wavelength ranges: 275-295 nm, 350-400 nm, 240-700 nm, 300-700 nm and 350-550 nm. We also considered spectral slope

from 300-600 nm, 350-600 nm, and 350-700 nm; however, slope values and Gaussian statistics for these spectral range were not significantly different from 300-700 nm and thus were not considered further. Throughout the manuscript, slope coefficient subscripts indicate fitting procedure (E for exponential decay with Eq. 1 and G for Gaussian with Eq. 2) along with the spectral range of data utilized for the fit. For example, $S_{G300:700}$ would indicate results from a fit with Eq. 2 from 300 to 700 nm.

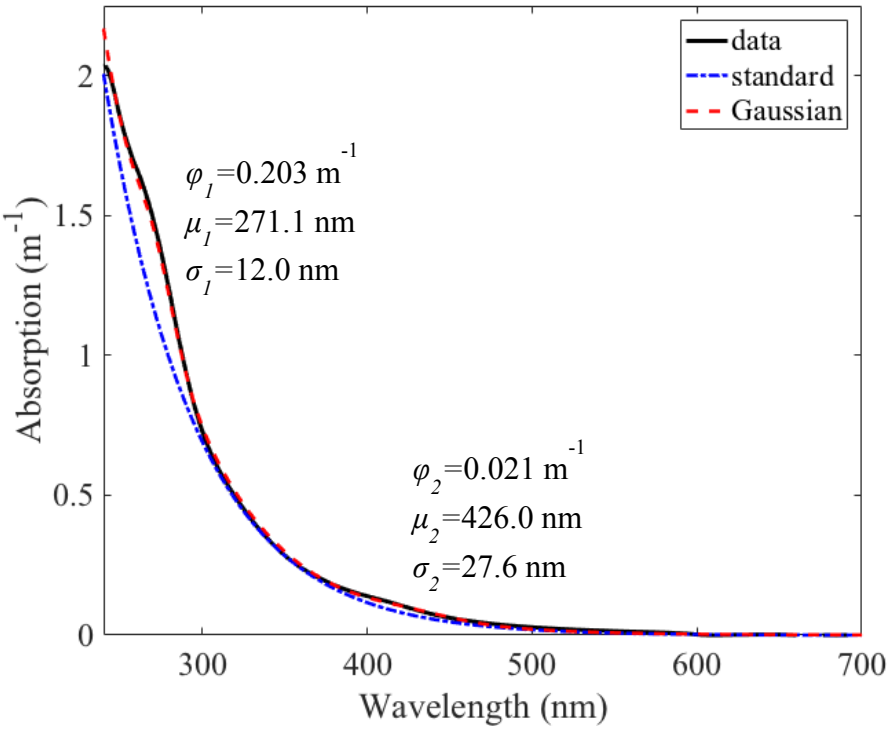


Figure 1. Example of exponential (Equation 1) and Gaussian decomposition fitting (Equation 2) of a_{CDOM} spectra from the Chukchi Sea. The black line is the observed data, the blue dash-dot line is the exponential fit for the spectra from 240-700 nm ($S_{E240:700}$=0.0178 nm^{-1}), and the red dashed line is the Gaussian decomposition fit from 240-700 nm ($S_{G240:700}$=0.0180 nm^{-1}). Two Gaussian components were fitted to this spectra, with φ, μ and σ for each component indicated within the figure.

1.2.2 Data Distribution

We focused on aggregating a_{CDOM} spectra into biogeochemical provinces following Longhurst [*2006*] biogeochemical province designations (provided by VLIZ [*2009*]). We considered all 54 Longhurst biogeochemical provinces and aggregated all inland water samples into an 'INLAND' biogeochemical province, resulting in 55 potential regions. Of these, 35 were represented within the SeaBASS dataset: 34 Longhurst provinces and the INLAND province. To be considered for analysis, we required each region to contain 5% of the maximum number of samples found within the most sampled region. For example, the NASW province (U.S. east coast) contained the maximum number of samples at 3,942

for the 300-700 nm spectral range. Thus, all regions considered in the $S_{G300:700}$ analyses contained at least 197 samples. Data availability allowed for the analysis of 6 provinces for the 240-700 nm spectral range, 16 provinces for 300-700 and 350-550 nm spectral ranges, and 23 provinces for 275-295 and 350-400 nm spectral ranges (Fig. 2). We considered $S_{350:400}$ in all provinces analyzed for $S_{275:295}$.

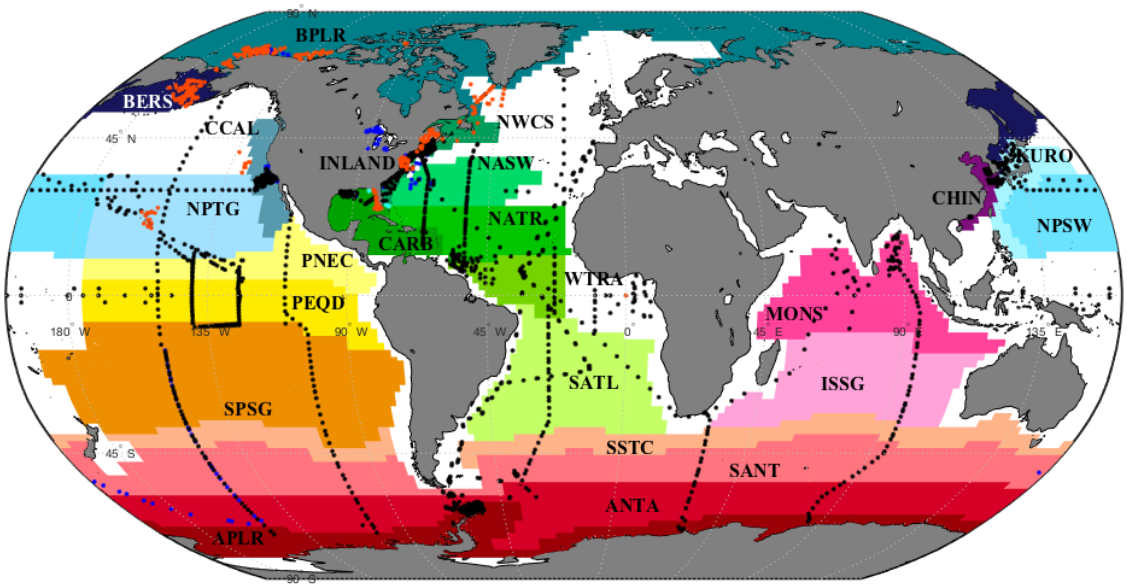


Figure 2. Map of a_{CDOM} observation locations for all spectral ranges considered: 350-550 nm (cyan, blue, black, red), 300-700 nm (blue, black, red), 275-295 nm (black, red), and 240-700 nm (red). The majority of INLAND samples are taken from rivers and lakes along the east and southeastern coast of the United States or from the Laurentian Great Lakes. Colored regions indicate Longhurst provinces used for analysis of spatial trends in the data set [*Longhurst, 2006*]. SATL, KURO, CHIN, NPSW, SSTC, ISSG and ANTA were only used for $S_{350:400}$ and $S_{275:295}$ analyses due to data restrictions described in Section 2.2. Following those restrictions, data points indicated here in uncolored regions are only used for aggregate slope comparisons depicted in Fig. 3a-c and are not used for any other analysis.

Absorption spectra were also considered within three depth classes defined as the first and second optical depths, calculated as 2.3 and 4.6 divided by $K_d(490)$, respectively, and below the photic zone (BPZ), comprising depths greater than the second optical depth but less than 1500 m [*Kirk,* 1994]. The majority of spectra below 1500 m were sampled at the Bermuda Atlantic Times Series (BATS) site; hence, we imposed the 1500 m threshold to avoid a potential bias from deep water spectra in this region. The depth classes for each province were identified by the average light extinction coefficient at 490 nm, $K_d(490)$, determined from MODIS-Aqua seasonal climatologies for each province. Seasonal $K_d(490)$ was used to ensure samples from different seasons fell within a similar underwater light field to ensure photodegradation rates are relatively constant within a given depth range used to group samples. Thus, a boreal winter sample near the lower limit of the first optical depth in a province could be a few meters deeper than a boreal summer sample near the upper limit of the second optical depth in the same province. Biogeochemical provinces for each depth class were included in the analysis using the same 5% of maximum sampling

criterion described above for the complete dataset. For most analyses, only the first optical depth is presented in the figures due to space constraints. Additional depth data for these analyses can be found in Grunert et al. (2018). Seasonal distribution of data within the provinces did not allow representation across all four seasons, thus, results focus on spatial trends in CDOM metrics.

1.2.3 Analysis of CDOM Metrics

Spectral slope calculated for the Gaussian decomposition and standard exponential methodologies were compared using linear regression. $S_{G240:700}$, $S_{G300:700}$ and $S_{G350:550}$ were compared to $S_{E275:295}$ and $S_{E350:400}$ to determine their potential for estimating CDOM molecular weight, source, and degradation state using linear regression. Non-linear least squares fits were used to assess the ability of log-normalized $a_{CDOM}(\lambda)$ at wavelengths of 350, 412, and 443 nm to estimate $S_{G240:700}$, $S_{G300:700}$, $S_{G350:550}$, $S_{E350:400}$ and $S_{E275:295}$.

We sought to determine if observed differences in S_{CDOM} between biogeochemical provinces were significant by applying a one-way analysis of variance (ANOVA) followed by a least significant difference multiple comparison of means test. The ANOVA was applied to S_{CDOM} categorized by biogeochemical province to determine if the mean slope values between biogeochemical provinces were significantly different ($p < 0.01$). The least significant difference multiple comparison of means test was then applied using the results of the ANOVA to determine if the distribution of spectral slope values is significantly different ($p < 0.05$). It is recognized the least significant difference (LSD) methodology offers looser statistical thresholds for significance than the Tukey-Kramer methodology [*Hayter*, 1986]. However, as an exploratory exercise aimed at defining potential differences, the more liberal LSD methodology was deemed more suitable here.

The results of the multiple comparison of means test are presented on bioplots as a way to visualize the degree of similarity in S_{CDOM} between different biogeochemical provinces. A line between boxes on a bioplot indicates that observed S_{CDOM} values between the considered provinces is statistically similar, while no connection indicates that observed S_{CDOM} values between considered provinces are statistically unique. These results are further clarified within maps of the regions displaying which provinces are statistically similar and how similar (the number of related provinces) within groups displaying connectivity (statistical similarity). The bioplots illustrate which provinces displayed S_{CDOM} values that are explicitly similar while the maps give a global representation of provinces that we consider as affiliated (all provinces are related to at least one other province within that group).

Average numbers of Gaussian components were calculated for provinces and by depth to determine the likelihood of fitting peaks within a province. While equation 2 requires a discrete number of components for fitting a single spectra, these non-discrete averages are derived from the number of fitted Gaussian components across all spectra for that province or depth range to indicate which regions are better fit using Gaussian fitting.

1.3 Results

1.3.1 Comparison of CDOM Models

We calculated the standard exponential slope for all spectra considered and compared them to slope values calculated using the Gaussian decomposition approach (Fig. 3). For all spectral ranges considered, the majority of Gaussian decomposition fitting resulted in same or larger slope values. For $S_{240:700}$, the mean difference, $S_{G240:700} - S_{E240:700}$, was 0.0005 nm^{-1} while the absolute differences for all spectra considered ranged from 0-0.0085 nm^{-1} (Fig. 3a). For $S_{300:700}$, the mean difference was 0.0012 nm^{-1} while the absolute differences for all spectra considered ranged from 0-0.0279 nm^{-1} (Fig. 3b). For $S_{350:550}$, the mean difference was 0.0002 nm^{-1} while the absolute differences for all spectra considered ranged from 0-0.0039 nm^{-1} (Fig. 3c). The sample size for $S_{240:700}$ fitting was significantly smaller; thus, differences may be more significant if more spectra across more provinces are considered.

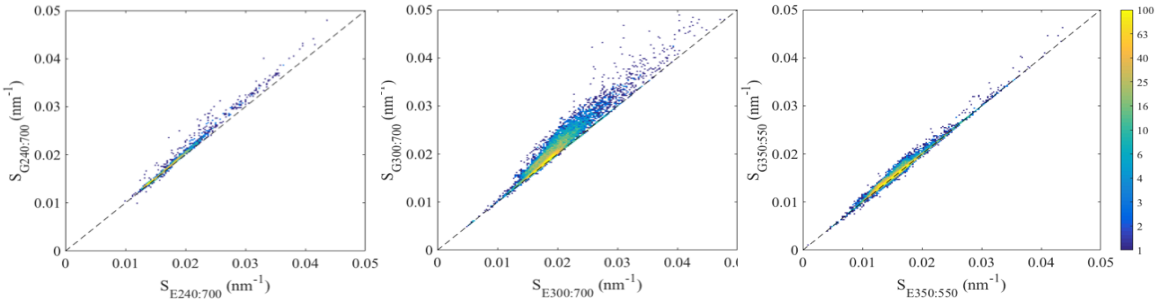


Figure 3a-c. Probability density functions comparing S_E (equation 1) vs. S_G (equation 2) for (a) 240-700 nm, (b) 300-700 nm and (c) 350-550 nm spectral ranges. Dashed lines indicate a 1:1 relationship.

When considering the spectral ranges investigated, differences between $S_{G240:700}$ and $S_{E240:700}$ were smallest for inland and coastal provinces and largest for open ocean provinces (Fig. 4a), although this comparison relies on only one open ocean province (NPTG). Absolute differences between $S_{G300:700}$ and $S_{E300:700}$ did not show a distinct spatial trend, as both NWCS (coastal) and NASW (open ocean) contained the largest absolute differences between the methodologies (0.025 and 0.0279 nm^{-1}, respectively). However, overall mean differences between $S_{G300:700}$ and $S_{E300:700}$ were smallest for inland and coastal regions and increased in open ocean regions, while environments characterized by a greater magnitude of a_{CDOM} (Arctic, inland and coastal waters) more frequently displayed no difference in calculated slope between the two methodologies (Fig. 4b). This is not surprising, as the Gaussian peaks fitted in the 300-700 and 350-550 nm spectral ranges for these environments were relatively small. Following this expectation, differences between $S_{G350:550}$ and $S_{E350:550}$ were quite small overall and did not display any regional trends (Fig. 4c).

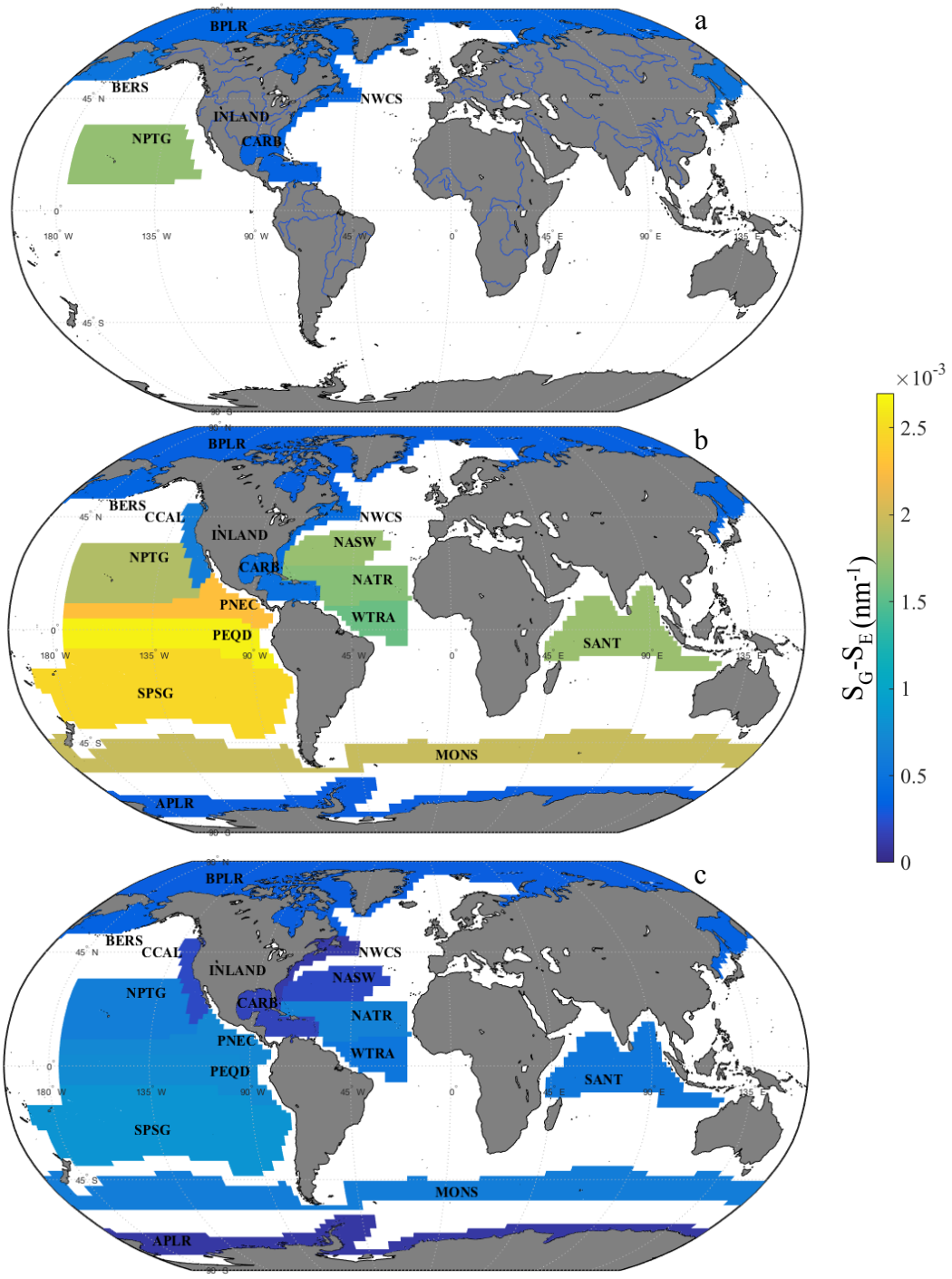


Figure 4. Mean difference between S_G and S_E for each province in the (a) 240-700 nm, (b) 300-700 nm and (c) 350-550 nm spectral ranges.

1.3.2 Spatial Trends in Slope Values

1.3.2.1 $S_{E275:295/E350:400}$

$S_{E275:295}$ in the first optical depth had a mean value of 0.036 nm^{-1} with a range in mean values of 0.035 nm^{-1} across all provinces in the first optical depth (Fig. 5a,f; Grunert et al. 2018). $S_{E275:295}$ was quite variable across all biogeochemical provinces, with coastal waters displaying smaller values, typically 0.02-0.03 nm^{-1}. Oceanic, Atlantic regions displayed larger values relative to all Pacific regions except NPSW. Southern Ocean regions displayed smaller values relative to the Atlantic and Pacific, with APLR displaying a mean $S_{E275:295}$ of 0.027 nm^{-1}. $S_{E275:295}$ decreased with depth for all regions except NWCS, CARB, KURO and SANT (Grunert et al. 2018). The range in values was relatively consistent between the second optical depth and below the photic zone, with both depth zones displaying a much smaller range in mean values than the first optical depth, 0.018 and 0.019, respectively. $S_{E350:400}$ values displayed less variability than $S_{E275:295}$, with a mean value of 0.018 nm^{-1} and a range of mean values of 0.01 nm^{-1} across all provinces in the first optical depth (Fig. 5b,g; Grunert et al. 2018). This spectral range observed the most consistent trend in decreasing slope with depth across regions (Grunert et al. 2018) with few provinces deviating slightly from this trend (ISSG, SPSG, PEQD, CCAL and SATL).

1.3.2.2 $S_{G240:700}$

$S_{G240:700}$ had a mean value of 0.020 nm^{-1} and a mean range of 0.019 nm^{-1} across all provinces in the first optical depth (Fig. 5c; Grunert et al. 2018). Differences between regions that are strongly influenced by terrestrial CDOM (inland, coastal and Arctic waters) displayed markedly different slope values from those observed in NPTG. Only 3 provinces (BPLR, BERS, and NWCS) had samples in all 3 depth ranges; for those provinces, slope values were relatively constant or increased slightly with depth.

1.3.2.3 $S_{G300:700}$

$S_{G300:700}$ displayed slightly less variability in mean slope, with a mean value of 0.021 nm^{-1} and a mean range of 0.008 nm^{-1} across all provinces in the first optical depth (Fig. 5d; Grunert et al. 2018). INLAND spectra displayed a relatively narrow slope range, with the bulk of observations between 0.01 and 0.02 nm^{-1} regardless of spectral range; however, the absolute range was quite broad (<0.01-~0.05 nm^{-1}). Coastal and sub-polar regions displayed mean $S_{G300:700}$ values ≤ 0.020 nm^{-1}, with BPLR (Arctic Ocean) also falling below 0.020 nm^{-1}. Minimum $S_{G300:700}$ was consistently below 0.015 nm^{-1} for all regions, with most regions falling below 0.010 nm^{-1}. Maximum $S_{G300:700}$ values showed considerably more variability, with $S_{G300:700}$ displaying the largest absolute range of all spectral ranges considered. While many studies observe a decrease in broad range S_{CDOM} with depth, some regions considered here maintained relatively consistent $S_{G300:700}$ across depth zones including NWCS, NATR and BERS. Regions that displayed an increase in

slope with depth include CARB and SANT. The range in $S_{G300:700}$ was generally consistent or showed a decline with depth.

1.3.2.4 $S_{G350:550}$

$S_{G350:550}$ values displayed similar variability to $S_{G300:700}$, with a mean $S_{G350:550}$ of 0.016 nm^{-1} and a mean range of 0.007 nm^{-1} across all provinces in the first optical depth (Fig. 5e; Grunert et al. 2018). While open ocean regions displayed larger $S_{G300:700}$ values, coastal regions generally displayed larger $S_{G350:550}$ values. $S_{G350:550}$ was more consistent with depth within regions than $S_{300:700}$. NASW, WTRA, CARB, NPTG, MONS and APLR $S_{G350:550}$ decreased with depth while $S_{G350:550}$ increased with depth in CCAL, PEQD, SPSG and SANT, albeit comparatively less than regions where $S_{G350:550}$ decreased with depth. $S_{G350:550}$ behaved similarly to $S_{350:400}$, however, with less range in mean values and less consistency in trends with respect to depth.

1.3.2.5 Spatial Relationships

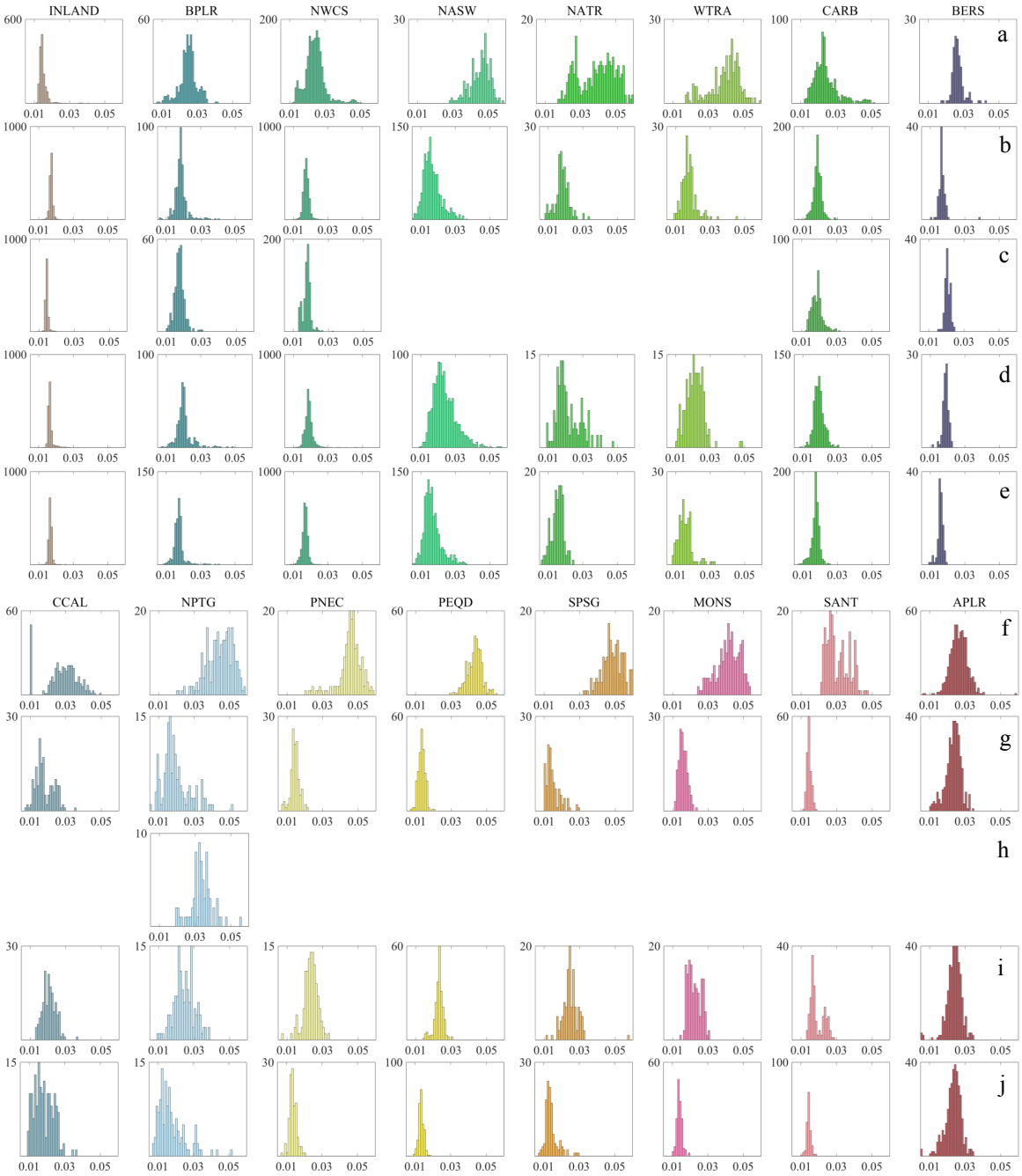
The relationship between S_{CDOM} values and biogeochemical province was considered for $S_{E275:295}$, $S_{E350:400}$, $S_{G300:700}$ and $S_{G350:550}$. Provinces with statistically similar S_{CDOM} values are presented as clusters where each province presented displays similar observed S_{CDOM} values to at least one other province.

$S_{E275:295}$ spatial relationships were unexpected, as provinces were not clustered within specific oceans or basins and displayed 4 distinct clusters (Fig. 6). BERS, KURO and APLR displayed unique $S_{E275:295}$ values relative to all other provinces. $S_{E350:400}$ showed the greatest similarity between spectral slope values across provinces, with two distinct clusters and APLR as an outlier (Fig. 7). $S_{G300:700}$ showed the greatest similarity across provinces, suggesting that a lack of contribution from wavelengths below 300 nm and a broad spectral range lead to greater similarities between disparate CDOM pools (Fig. 8). It should be noted that if Laurentian Great Lakes stations are excluded from the $S_{G300:700}$ analysis, INLAND is statistically unique from all other regions as well. $S_{G350:550}$ showed the most connectivity between proximal provinces, although none of the spectral ranges appear to be related primarily by province proximity in the global ocean (Fig. 9).

For all parameters considered except $S_{E275:295}$, S_{CDOM} similarity increased with depth. For $S_{G300:700}$ and $S_{G350:550}$, this resulted in only one bioplot cluster ($S_{G300:700}$) or one bioplot cluster with APLR (both) and CCAL (second optical depth) as outliers ($S_{G350:550}$). For all spectral ranges, the degree of uniqueness in mean slope and slope distribution decreased with depth.

1.3.3 Gaussian Components

The number of Gaussian components fit was highly dependent on the spectral range considered (Fig. 10). For 240-700 and 350-550 nm, fewer Gaussian components were fit


INLAND
BPLR
NWCS
NASW
NATR
WTRA
CARB
BERS
a
b
c
d
e
CCAL
NPTG
PNEC
PEQD
SPSG
MONS
SANT
APLR
f
g
h
i
j

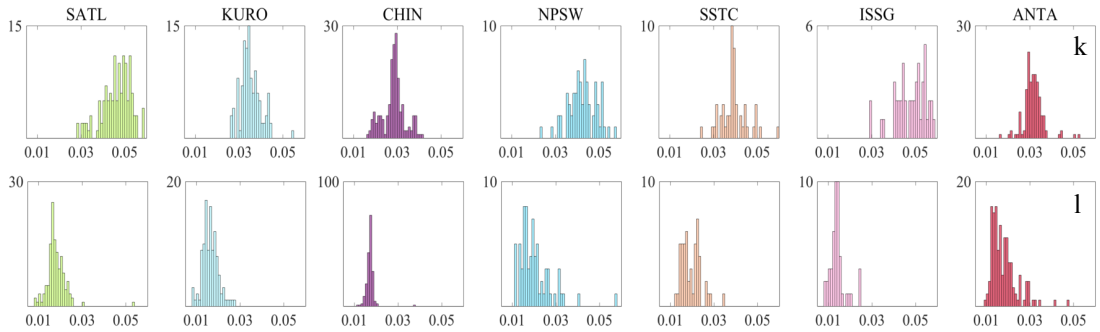


Figure 5. Spectral slope frequency distribution for the first optical depth of each province with spectral range for each row as (a,f,k) $S_{E275:295}$, (b,g,l) $S_{E350:400}$, (c,h) $S_{G240:700}$, (d,i) $S_{G300:700}$and (e,j) $S_{G350:550}$. Missing histograms indicate provinces that did not have enough spectra measured down to 240 nm to be considered here.

due to lignin peak fitting (240-700) dramatically improving the goodness of fit and relatively small peak height for components located in the fitting range used for 350-550 nm.

1.3.3.1 Spatial Trends in Gaussian Components: $S_{G240:700}$

The minimum number of Gaussian components fitted for all regions in the first optical depth for $S_{G240:700}$ was 1, with the maximum number ranging from 4 to 7 for INLAND and BPLR, respectively (Fig. 10a; Grunert et al. 2018). The range of fitted components stayed the same or decreased with depth (data not shown). For the INLAND province, it appears that the goodness of fit was improved drastically by fitting a lignin peak, to the point that fitting peaks at longer wavelengths did not produce a statistically better fit.

1.3.3.2 Spatial Trends in Gaussian Components: $S_{G300:700}$

The minimum number of Gaussian components fitted for each province in the first optical depth for $S_{G300:700}$ was 0 or 1, with a maximum number of fitted components ranging from 2 to 10 for APLR and NWCS, respectively (Fig. 10b; Grunert et al. 2018). Overall, most provinces were fit with less Gaussian components with depth (data not shown). APLR was an outlier across all depths with the majority of spectra not fitted with any Gaussian components. BERS and BPLR displayed the most fitted components with the number of fitted components increasing with each depth zone considered. Atlantic regions consistently had a mean of 2 or more Gaussian components, while NWCS and CARB had an average of 1.7 and 1.8 Gaussian components, respectively. Pacific regions had from 1.1-1.5 Gaussian components on average. The number of fitted Gaussian components typically decreased with depth in Atlantic regions while the number of Gaussian components in Pacific regions were relatively consistent with depth.

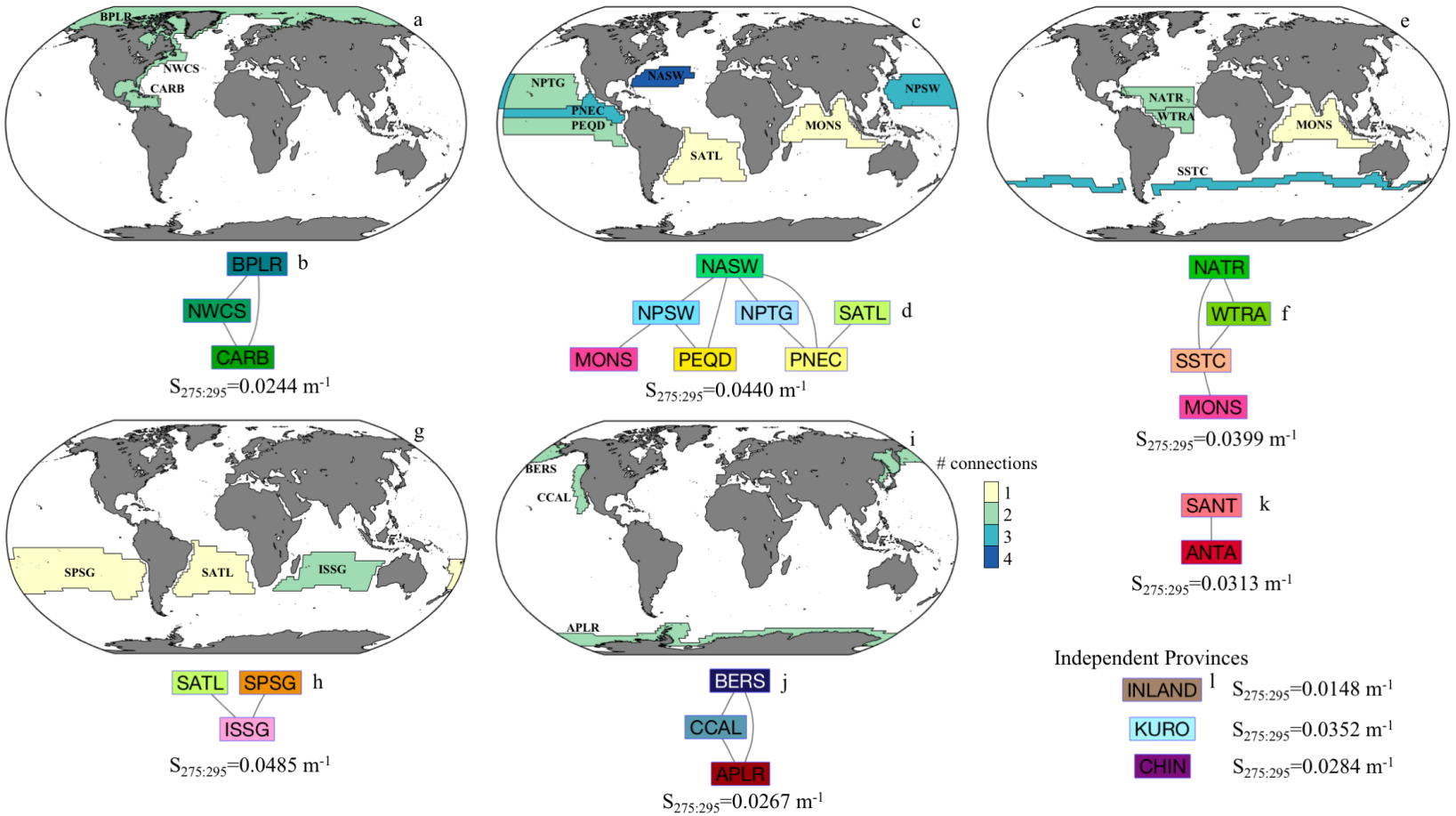


Figure 6. Global maps displaying provinces grouped based on results from the multiple comparison of mean test. The connecting lines indicate that provinces within the group are statistically similar. The province color indicates the number of provinces found to have a statistically similar mean slope based on the multiple comparison of mean test. The provinces that appear more than once were found to have a mean and uncertainty in the mean (described in section 2) overlapping with provinces in two different groups. Subplots display corresponding maps and bioplots as (a and b) Cluster 1 ($S_{E275:295}$=0.0244 nm^{-1}), (c and d) Cluster 2 ($S_{E275:295}$=0.0440 nm^{-1}), (e and f) Cluster 3 ($S_{E275:295}$=0.0399 nm^{-1}), (g and h) Cluster 4 ($S_{E275:295}$=0.0485 nm^{-1}), (i and j) Cluster 5 ($S_{E275:295}$=0.0267 nm^{-1}), and (k and bioplot only) SANT and ANTA ($S_{E275:295}$=0.0313 nm^{-1}) and (l) indicate provinces with unique mean $S_{E275:295}$ values: INLAND ($S_{E275:295}$=0.0148 nm^{-1}), KURO ($S_{E275:295}$=0.0352 nm^{-1}), and CHIN ($S_{E275:295}$=0.0284 nm^{-1}).

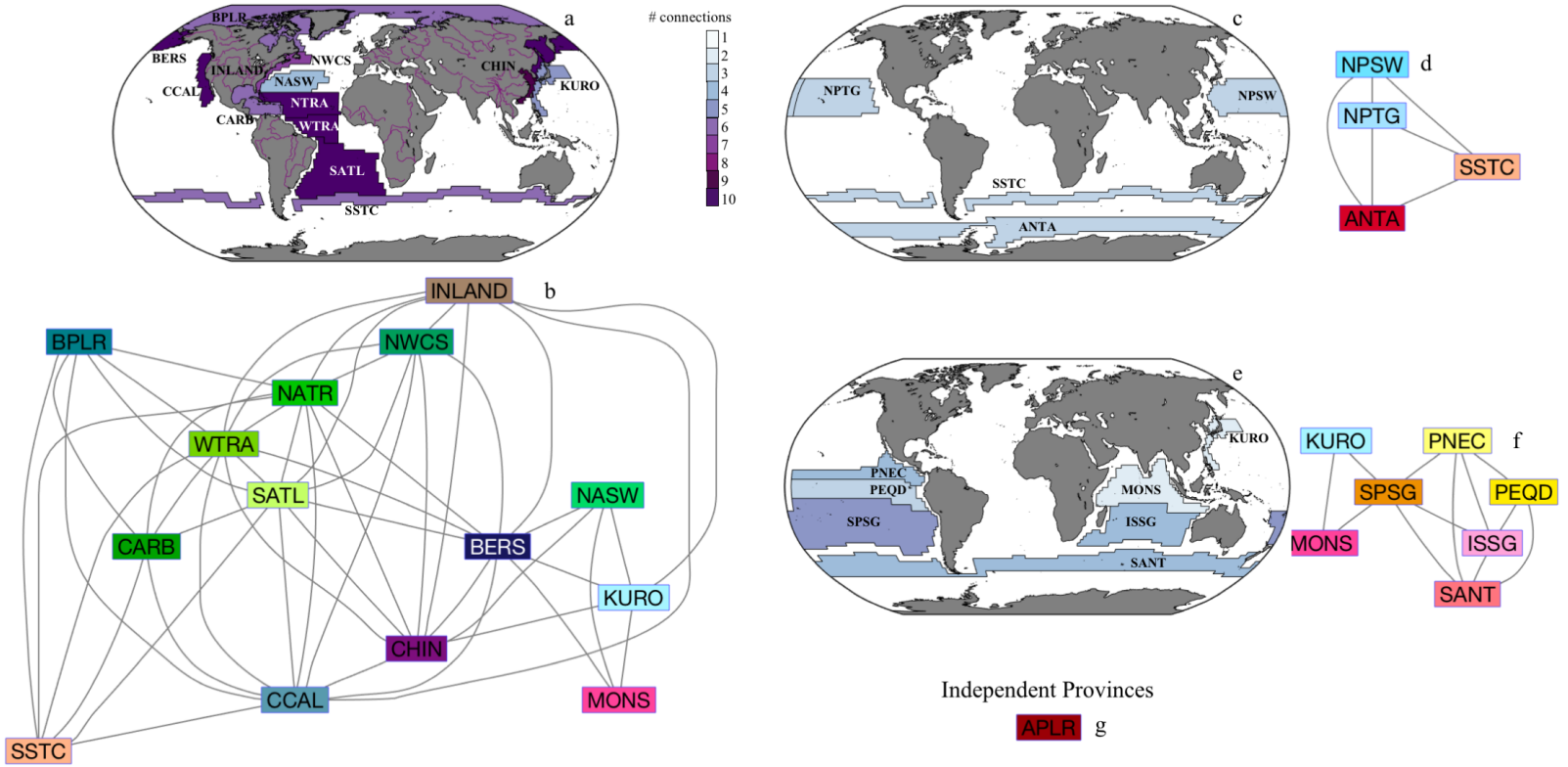


Figure 7. Global maps displaying provinces grouped based on results from the multiple comparison of means test. The connecting lines indicate that provinces within the group are statistically similar. The province color indicates the number of provinces found to have a statistically similar mean slope based on the multiple comparison of mean test. Provinces that appear more than once were found to have a mean and uncertainty in the mean (described in section 2) overlapping with provinces in two different groups. The subplots display corresponding maps and bioplots as (a and b) Cluster 1 ($S_{E350:400}$ =0.0176 nm^{-1}), (c and d) Cluster 2 ($S_{E350:400}$=0.0210 nm^{-1}), and (e and f) Cluster 3 ($S_{E350:400}$=0.0147 nm^{-1}) to indicate groups with two or more statistically similar province means, while (g) APLR displayed a unique mean $S_{E350:400}$ value (0.0230 nm^{-1}).

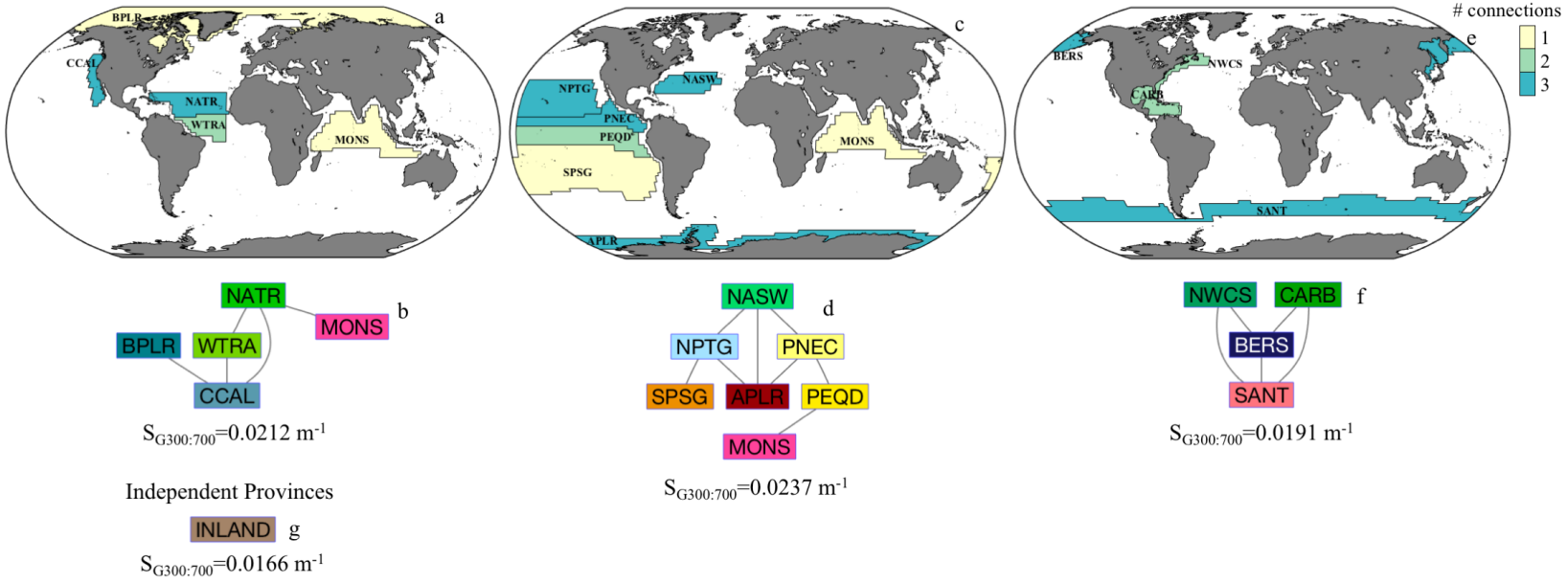


Figure 8. Global maps displaying provinces grouped based on results from the multiple comparison of mean test. The connecting lines indicate that provinces within the group are statistically similar. The province color indicates the number of provinces found to have a statistically similar mean slope based on the multiple comparison of mean test. Provinces that appear more than once were found to have a mean and uncertainty in the mean (described in section 2) overlapping with provinces in two different groups. The subplots display corresponding maps and bioplots as (a and b) Cluster 1 ($S_{G300:700}$=0.0212 nm^{-1}), (c and d) Cluster 2 ($S_{G300:700}$=0.0237 nm^{-1}), and (e and f) Cluster 3 ($S_{G300:700}$ =0.0191 nm^{-1}), while (g) INLAND displayed a unique mean $S_{G300:700}$ (0.0166 nm^{-1}).

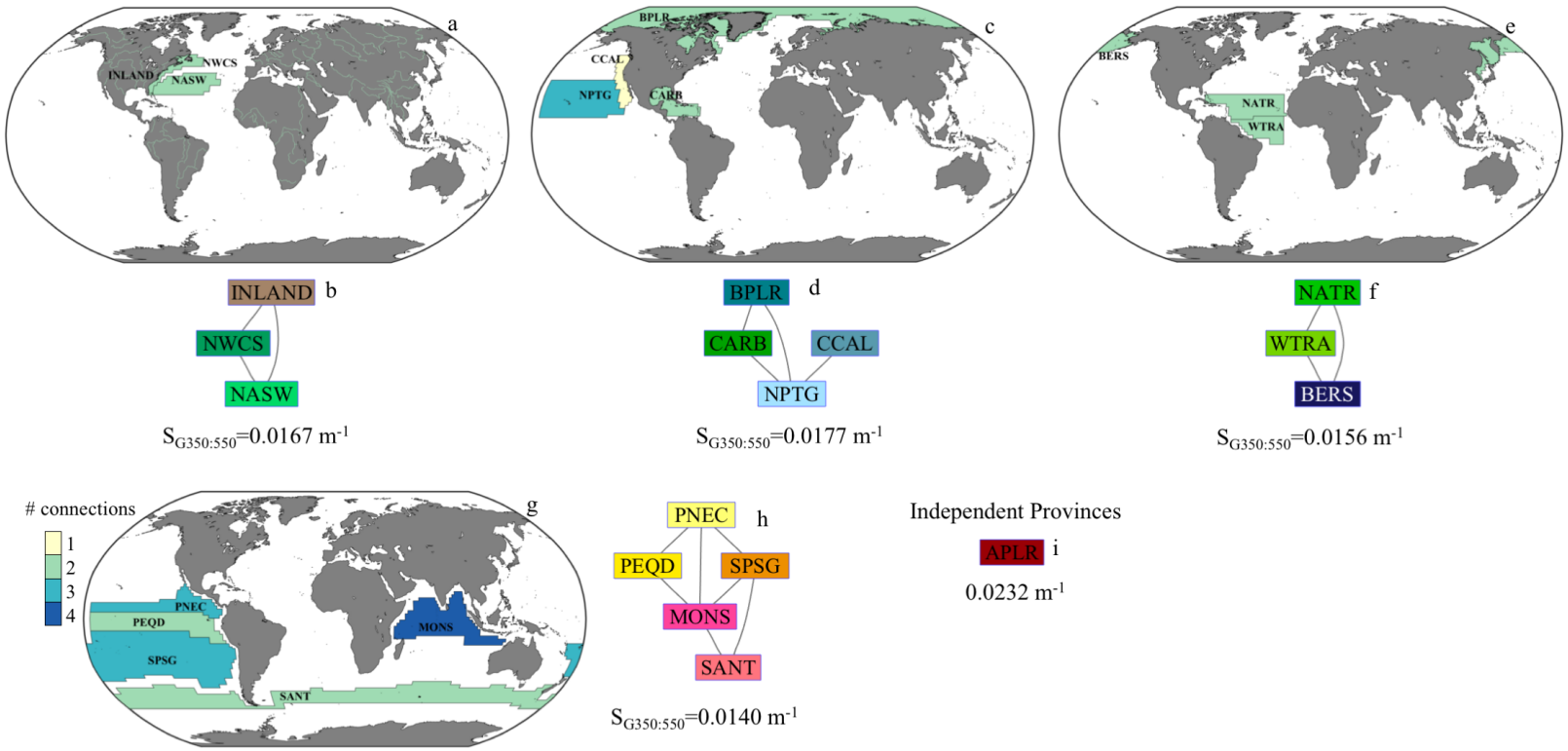


Figure 9. Global maps displaying provinces grouped based on results from the multiple comparison of mean test. The connecting lines indicate that provinces within the group are statistically similar. The province color indicates the number of provinces found to have a statistically similar mean slope based on the multiple comparison of mean test. Provinces that appear more than once were found to have a mean and uncertainty in the mean (described in section 2) overlapping with provinces in two different groups. The subplots display corresponding maps and bioplots as (a and b) Cluster 1 ($S_{G350:550}$=0.0167 nm^{-1}), (c and d) Cluster 2 ($S_{G350:550}$=0.0177 nm^{-1}), (e and f) Cluster 3 (SG350:550 = 0.0156 nm^{-1}), and (g and h) Cluster 4 ($S_{G350:550}$=0.0140 nm^{-1}), while (i) APLR displayed a unique mean $S_{G350:550}$ value (0.0232 nm^{-1}).

1.3.3.3 Spatial Trends in Gaussian Components: $S_{G350:550}$

The minimum number of Gaussian components fitted for each province in the first optical depth for $S_{G350:550}$ was 0 or 1, with a maximum number of fitted components ranging from 1 to 8 for APLR and WTRA, respectively (Fig 10c; Grunert et al. 2018). Overall, most provinces were fit with the same or less Gaussian components with depth (data not shown). APLR remained an outlier across all depths in this spectral range as well. BERS and BPLR were fitted with 1 and 0.8 mean components in the first optical depth with little change with depth. This is in stark contrast to fitting from 300-700 nm, suggesting that 300-350 nm is an important spectral range to consider in this region and contains more prominent deviations from the baseline exponential model in this spectral range.

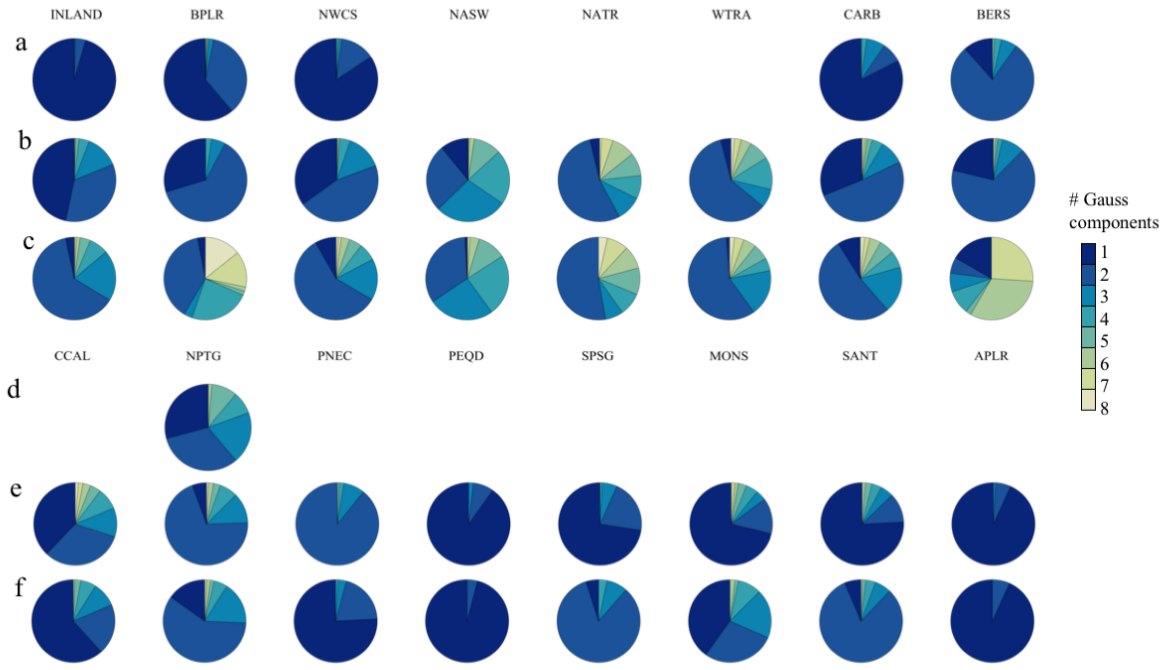


Figure 10. Number of Gaussian components indicated by color in the first optical depth for (a, d) 240–700 nm, (b, e) 300–700 nm, and (c, f) 350–550 nm.

1.3.3.4 Spatial Trends in Gaussian Component Spectral Location: $S_{G240:700}$

Mean Gaussian component location in the first optical depth across all regions was 317.8 nm, with a minimum and maximum mean location of 286.9 nm for INLAND and 350.6 nm for NPTG (Fig. 11a-c; Grunert et al. 2018). Fitting of spectra was dominated by fitting of the lignin peak with the mean first Gaussian component location occurring at 276 nm for all 6 biogeochemical provinces considered. The location of the first component was strongly influenced by sample source and proximity to terrestrial material, as the first Gaussian component location was always below 300 nm for INLAND, NWCS, CARB and BERS while the first Gaussian component location ranged up to 385 and 381 nm for BPLR and NPTG, respectively. The location of the first peak dominated the location of peaks

across all provinces, while the spectral region from 420-425 nm was also frequently fitted in this spectral range. The location of fitted peaks was relatively consistent between depth ranges (Fig. 11a-c).

1.3.3.5 Spatial Trends in Gaussian Component Spectral Location: $S_{G300:700}$

Mean Gaussian component location in the first optical depth for $S_{G300:700}$ across all regions was 415 nm (Fig. 11d-f), with a minimum and maximum mean location of 366 nm for INLAND and 429 nm for BERS (Grunert et al. 2018). INLAND and NASW were the only regions that were statistically unique in their mean location based on one-way ANOVA and multiple comparison of means tests, with a mean location of 397 nm in NASW. All other regions had a mean location between 413 and 429 nm. For INLAND spectra, the majority of Gaussian components were fitted from 330-350 nm (57%). NASW Gaussian components did not display a bias towards a particular spectral region, with nearly 75% of peaks fitted between 365-425 nm. For all regions other than INLAND, spectral regions most frequently fitted included 370-375 nm (4%) and 410-440 nm (35%), with 425-430 nm displaying the most fitted peaks at 9% of all peaks fitted. The number of fitted peaks significantly decreased below 360 nm (Fig. 11d).

In the second optical depth, the most frequently fitted region was 415-435 nm with 34% of fitted peaks occurring in this region (Fig. 11e). 370-375 nm was an important spectral range with 5% of peaks fitted in this region. Below the photic zone, 415-430 nm contained 17% of all fitted peaks and 370-375 nm was an important spectral range with 6% of fitted peaks (Fig. 11f). In this depth zone, 390-415 nm was also important, with 13% of all peaks fitted in this region.

1.3.3.6 Spatial Trends in Gaussian Component Spectral Location: $S_{G350:550}$

Mean Gaussian component location in the first optical depth for $S_{G350:550}$ across all regions was 424 nm, with a minimum and maximum mean location of 395 nm for NASW and 444 nm for BERS. NASW was the only region that had a statistically unique mean peak location. All other regions had a mean location between 413 and 444 nm. However, INLAND Gaussian components were most frequently fitted from 355-360 nm (22%) and 490-495 nm (24%), resulting in a mean location of 431 nm. NASW Gaussian components did not display a bias towards a particular spectral region, with over 80% of peaks fitted between 360 and 425 nm. For all regions other than INLAND, spectral regions most frequently fitted included 355-370 nm (16%) and 410-430 nm (28%), with 410-415 nm displaying the most fitted peaks at 10% of all peaks fitted (Fig. 11f).

In the second optical depth, the most frequently fitted region was 410-420 nm with 31% of fitted peaks occurring in this region. A local peak in the distribution of fitted components occurred from 390-395 nm with 6% of all peaks fitted in this spectral region (Fig. 11g). Below the photic zone, 410-420 nm contained 29% of all fitted peaks while 405-425 nm contained 44% of all fitted peaks (Fig. 11h). 390-395 nm was an important

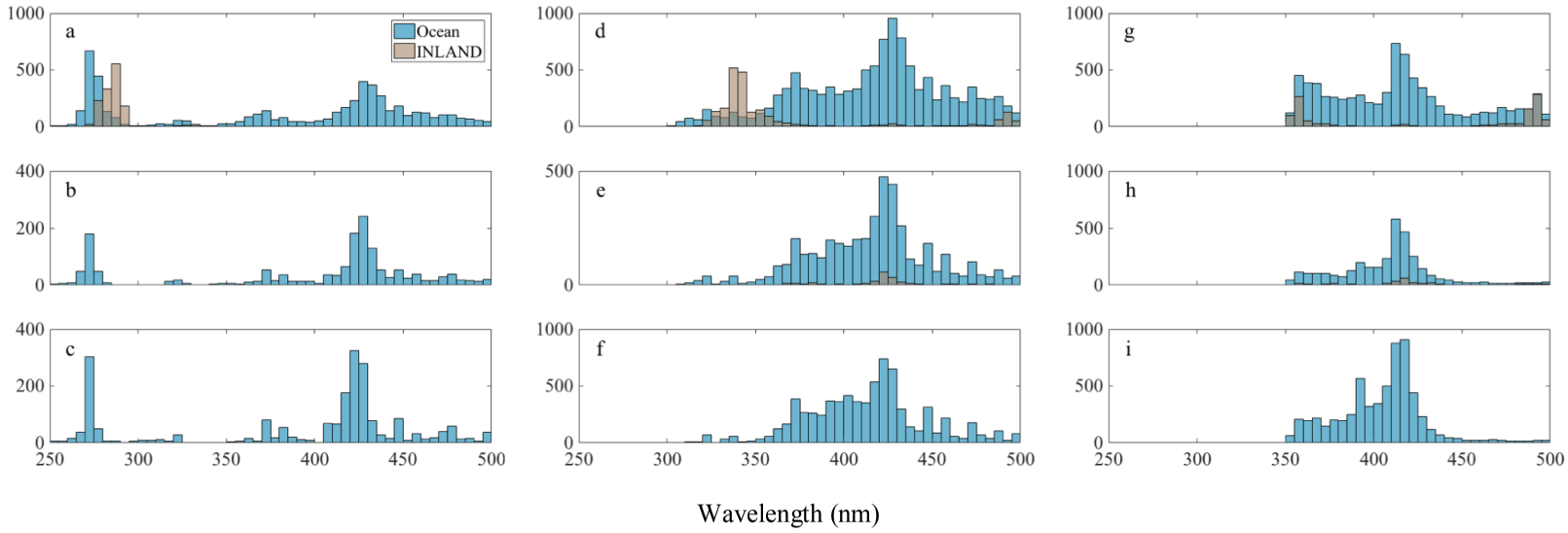


Figure 11. Frequency distribution of Gaussian component spectral location between ocean and inland observations for (a–c) 240–700 nm, (d–f) 300–700 nm, and (g–i) 350–550 nm for the first (Figures 11a, 11d, and 11g) and second (Figures 11b, 11e, and 11h) optical depths and below the photic zone (Figures 11c, 11f, and 11i).

spectral range with 9% of all fitted peaks. In this depth zone, relatively few peaks were fitted after 430 nm.

1.3.3.7 Trends in Gaussian Component Metrics

Gaussian peak height, φ, was found to correlate quite well with $a_{CDOM}(350)$ (r^2=0.88), although the distribution appears bifurcated with a break-point between $a_{CDOM}(350) \leq 15$ m^{-1} (r^2=0.76) and $a_{CDOM}(350) > 15$ m^{-1} (r^2=0.90) (Grunert et al. 2018). Both relationships suggest a large dependency between φ and the magnitude of a_{CDOM} in the system. We accounted for this effect by normalizing φ with the modeled a_{CDOM} at the wavelength associated with the Gaussian peak location as described in the Methods section. Normalized φ showed that open ocean Gaussian components are more prominent relative to the magnitude of CDOM in the system when compared to terrestrial components, with INLAND mean normalized φ=0.03 while ocean provinces had a mean normalized φ=0.1 (data not shown). For $S_{G300:700}$ (excluding INLAND), the first, second and BPZ depths were fitted with an average of 1.8, 1.7 and 2.2 Gaussian components, respectively. For $S_{G350:550}$ (excluding INLAND), the first, second and BPZ depths were fitted with an average of 1.0, 1.2 and 1.8 Gaussian components, respectively.

1.3.4 CDOM Metric Comparisons

1.3.4.1 Comparison of Broad Range Slopes to $S_{E275:295}$/$S_{E350:400}$

The influence of capturing absorption below 300 nm and the likely influence of lignin for relating broad range spectral slopes to absorption from $S_{275:295}$ is clear, as regions dominated by terrestrially-sourced CDOM show tight relationships between $S_{E275:295}$ and $S_{G240:700}$ in the first optical depth (r^2=0.83-0.95; Grunert et al. 2018). Relationships between $S_{E275:295}$ and $S_{G300:700}$ were generally poor while relationships between $S_{E275:295}$ and $S_{G350:550}$ entirely deteriorate for most regions and depths ($r^2 < 0.2$; Grunert et al. 2018), suggesting a distinct difference in the processes these spectral ranges describe. Notably, APLR a_{CDOM} spectra displayed consistent slope values across diverse spectral ranges.

As expected, $S_{E350:400}$ related inversely to broad slope ranges relative to $S_{E275:295}$. Most $S_{G240:700}$ correlations were relatively poor (r^2=0-0.34) with INLAND and CARB displaying relationships of 0.49 and 0.82, respectively, suggesting a broad influence of absorption by lignin or lignin-derivatives in these regions. Overall, $S_{E350:400}$ showed distinct trends from spectral ranges capturing terrestrial features, suggesting a utility for monitoring *in situ* processes within this spectral range (Grunert et al. 2018).

1.3.4.2 Comparison between a_{CDOM} and S_{CDOM}

We considered relationships between a_{CDOM} at 350, 412, and 443 nm and the calculated slope values to determine the feasibility of using the magnitude of a_{CDOM} at a reference wavelength to estimate S_{CDOM}. 412 and 443 nm, or nearby wavelengths, have

applications to heritage satellite sensors while 350 nm is the shortest wavelength anticipated to be available for the PACE sensor, providing maximum signal-to-noise for retrieving a_{CDOM} while also avoiding overlap with phytoplankton pigments, including mycosporine-like amino acid absorption peaks around 330 and 360 nm [*Pavlov et al.*, 2014]. Relationships were best fit using a non-linear least squares exponential fit. The goodness of fit did not change significantly between different reference wavelengths; subsequently, we only focus on $a_{CDOM}(350)$ in anticipation of future sensor capabilities.

$S_{275:295}$ had the strongest relationships with $a_{CDOM}(350)$, although APLR showed a poor relationship ($r^2 < 0.15$). Relationships between $a_{CDOM}(350)$ and $S_{G300:700}$ were highly variable, with SANT the only station that showed strong to relatively strong relationships across all depths ($r^2 > 0.7$). $S_{350:400}$ and $S_{G350:550}$ did not correlate with $a_{CDOM}(350)$ for almost all regions and depths considered (Grunert et al. 2018).

We also assessed the error introduced in satellite-derived estimates of $a_{CDOM}(412)$ by using an assumed S_{CDOM} value. CDOM has traditionally been considered alongside non-algal particulate (NAP) material, as both have absorption spectra that follow an approximately exponentially increasing with decreasing wavelength relationship. These absorption terms, a_{CDOM} and $a_{NAP,}$ are combined into a single term, colored detrital matter (CDM) absorption such that $a_{CDM}=a_{CDOM}+a_{NAP}$ and a_{CDM} has an average spectral slope (S_{CDM}) representative of the slope and percent contribution of each component. The GSM algorithm and the Quasi-Analytical Algorithm (QAA) assume an S_{CDM} of 0.015 and 0.013-0.017, respectively [*Lee et al.*, 2002; *Maritorena et al.*, 2002]. Global observations of a_{NAP} suggest it accounts for 10-20% of the a_{CDOM} signal, and a typical spectral slope for a_{NAP} is 0.011 nm^{-1} with a range of values much smaller than S_{CDOM} [*Dong et al.*, 2013; *Hoepffner and Sathyendranath*, 1993; *Roesler et al.*, 1989]. Assuming an S_{CDOM} of 0.015 nm^{-1}, a contribution of 10% and 20% of a_{NAP} with spectral slope of 0.011 nm^{-1} results in an assumed S_{CDOM} of 0.016 and 0.0154 nm^{-1}, respectively. We assessed the percent error between mean $a_{CDOM}(412)$ observed for each province and $a_{CDOM}(412)$ calculated for the province using the mean $a_{CDOM}(443)$ and these assumed S_{CDOM} values. Error rates were similar when assuming 10% and 20% contributions from a_{NAP} to a_{CDOM}. When assuming a 20% contribution and an S_{CDOM} of 0.0154 nm^{-1}, estimated $a_{CDOM}(412)$ error varied from 4-22% of the observed mean $a_{CDOM}(412)$ (Fig. 12).

1.4 Discussion

1.4.1 CDOM Models

Bias introduced in the S_{CDOM} parameter from varying the spectral range has been widely recognized for some time [*Twardowski et al.*, 2004]. Despite this understanding, only a few narrow range spectral slope parameters have been adopted consistently, $S_{275:295}$ and $S_{350:400}$ [*Helms et al.*, 2008], with broad range slopes highly variable throughout the

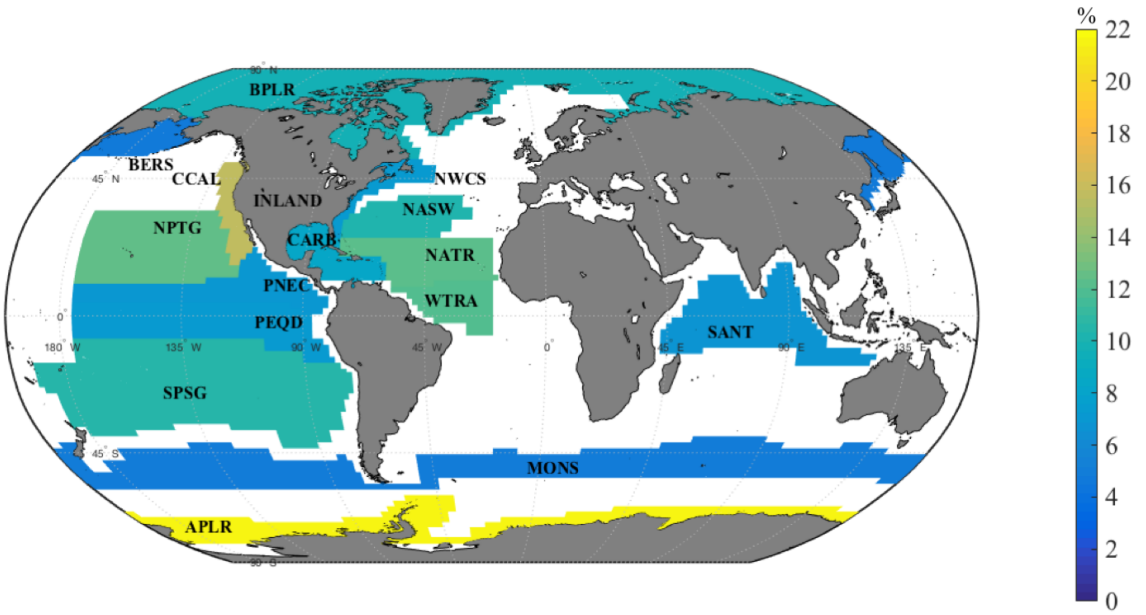


Figure 12. Percent error between mean observed $a_{CDOM}(412)$ for each province and calculated $a_{CDOM}(412)$ using mean observed $a_{CDOM}(443)$ and an assumed S_{CDOM} of 0.0154 nm^{-1}.

literature. Regardless of the CDOM model used, exponential or Gaussian decomposition, the bias due to spectral range considered remains an issue. S_{CDOM} differences between the exponential CDOM and Gaussian decomposition models are relatively minor for most spectra, with S_{CDOM} values generally higher for the Gaussian decomposition model. However, differences between the models suggest that removing deviations from the exponential model and modeling these features with Gaussian curves allows for a better characterization of the underlying exponential signal and a better fit of the measured CDOM spectra through improved r-squared and RMSE. This, in turn, will allow for more consistent comparisons of S_{CDOM} from the same spectral range and a stronger basis for estimating CDOM composition optically [*Del Vecchio and Blough*, 2004a]. In particular, we found the Gaussian decomposition method to be crucial for accurately modeling CDOM spectra strongly influenced by terrestrial material.

1.4.2 S_{CDOM}

Various spectral ranges have been used to characterize S_{CDOM}, with narrow range slopes typically focusing on specific attributes related to CDOM source, diagenetic state and molecular weight [*Helms et al.*, 2008; *Spencer et al.*, 2008]. The basis for linking optical properties to estimates of CDOM composition has strong backing in the literature [*Boyle et al.*, 2009; *Coble*, 2007; *Hernes and Benner*, 2003; *Kowalczuk et al.*, 2005; *Wünsch et al.*, 2015]. Here, we assessed how broad spectral range S_{CDOM} relates to common optical indicators of CDOM composition (e.g. $a_{CDOM}(\lambda)$, $S_{275:295}$) and varies across spectral ranges, ocean biogeochemical provinces and depth ranges.

CDOM absorption at a reference wavelength is often used as an indicator of CDOM composition as it relates well with $S_{E275:295}$ and lignin concentration [*Fichot et al.*, 2016; *Mannino et al.*, 2014]. Past studies have shown strong relationships between a_{CDOM} and S_{CDOM} along transects from lower salinity coastal waters to higher salinity offshore waters where the range in $a_{CDOM}(\lambda)$ typically varies by orders of magnitude [*Kowalczuk et al.*, 2006; *Pavlov et al.*, 2016; *Stedmon and Markager*, 2003], while a study in the Arctic Ocean found that CDOM absorption and S_{CDOM} do not relate across marine and terrestrially-derived CDOM pools [*Granskog et al.*, 2012]. For our study, $S_{E275:295}$ displayed relatively strong relationships with $a_{CDOM}(350)$ for most provinces ($r^2 > 0.5$). Notably, APLR displayed a poor relationship between $a_{CDOM}(350)$ and S_{CDOM} across all spectral ranges, while most provinces displayed highly variable or poor relationships between $a_{CDOM}(350)$ and $S_{350:400}$. Relationships between broad range S_{CDOM} and a_{CDOM} were highly variable and were generally poor, even within the INLAND province where variability in $a_{CDOM}(350)$ is high across similar S_{CDOM} values, an observation also seen by *Meler et al.* [2016] . We consider these inconsistencies within the literature and our findings as representative of a decoupling between $a_{CDOM}(\lambda)$ and S_{CDOM} in systems that display higher variability in one parameter over the other or when data sets are not taken within a single sampling period. Open ocean CDOM has been characterized as a mix of degraded terrestrial material and contributions from *in situ* production [*Andrew et al.*, 2013], suggesting that spectral shape due to CDOM composition is reflected by varying contributions from these processes with a low magnitude of $a_{CDOM}(350)$ across spectra considered for these provinces. We hypothesize that within oceanic regions with little terrestrial influence, process-specific S_{CDOM} variability (e.g. photodegradation and alteration by the microbial community) is not necessarily reflected in $a_{CDOM}(350)$ values at different times. These degradation processes occur regardless of the amount of CDOM; thus, an $a_{CDOM}(350)$ value of 0.1 m^{-1} could be affiliated with a spectral shape consistent with microbial alteration of the CDOM pool or with a spectral shape consistent with photodegraded terrestrial material depending on the time and/or location sampled. This is a potential bias of ship-based sampling that should be accounted for when attempting to accurately estimate a_{CDOM} from satellite-based remote sensing.

Spatial trends across all spectral ranges indicate that S_G is quite variable across the global ocean (Fig. 5). Typically, global distributions of CDOM are presented as satellite-derived climatologies of colored dissolved and detrital matter (CDM) absorption (a_{CDM}), where non-algal particulate (NAP) absorption (a_{NAP}) and CDOM are retrieved together and a_{NAP} typically accounts for <20% of CDM in the global ocean [*Nelson et al.*, 1998]. These climatologies typically display low $a_{CDM}(\lambda)$ in subtropical ocean gyres and higher $a_{CDM}(\lambda)$ along continental margins, the equator, and high latitudes [*Siegel et al.*, 2005]. S_G did not follow spatial patterns observed in satellite observations of a_{CDM} for any spectral range considered.

Estimating a_{CDOM} from satellite radiometry requires assuming $S_{CDOM/CDM}$ or estimating a_{CDOM} by parameterizing non-algal particulate absorption (a_{NAP}) through empirical relationships and treating a_{CDOM} as the residual [*Matsuoka et al.*, 2013]. Attempts to estimate S_{CDOM} directly have relied on removal of the a_{NAP} signal and adjusting an

initially-assumed S_{CDOM} using ratios of remotely-sensed reflectance [*Dong et al.*, 2013]. We found that assuming an S_{CDOM} of 0.0154 nm^{-1}, equivalent to a 90% contribution of a_{CDOM} to the a_{CDM} signal [*Nelson et al.*, 2010], and using $a_{CDOM}(443)$ results in errors in $a_{CDOM}(412)$ of 4-22% relative to the average $a_{CDOM}(412)$ observed in the provinces, suggesting that the assumed values or starting points used for S_{CDOM} should be regionalized to reduce uncertainty (Fig. 12).

Retrieving S_{CDOM} independently without assumptions would be a first step towards estimating CDOM composition through a direct, rather than parameterized, observation. $S_{E350:400}$ can potentially be retrieved using PACE. However, we found clearer separation between regions using $S_{G350:550}$, suggesting that accounting for a broader spectral range and Gaussian components within that range draws sharper contrasts between distinct CDOM pools in the global ocean. Neither $S_{E350:400}$ or $S_{G350:550}$ related well with $S_{E275:295}$; thus, both will likely be poor proxies for evaluating terrestrial contribution and molecular weight [*Helms et al.*, 2008].

Biogeochemical models assume an S_{CDOM} slope [*Xiu and Chai*, 2014] with some models accounting for microbial and photodegradation of CDOM over appropriate time scales [*Dutkiewicz et al.*, 2015]. While the inclusion of optical parameters in global biogeochemical ocean models is a recent development, further partitioning the spectral properties of CDOM by region would enhance the accuracy when modeling the underwater light field. Past studies have found changes in CDOM spectral characteristics with depth to be significant [*Hickman et al.*, 2010; *Pérez et al.*, 2016]; we found changes in mean CDOM spectral characteristics to change up to 0.034, 0.02, 0.008 and 0.009 nm^{-1} between biogeochemical provinces while changes across depth ranges within a given biogeochemical province varied up to 0.016, 0.003, 0.005 and 0.004 nm^{-1} for $S_{E275:295}$, $S_{G240:700}$, $S_{G300:700}$, and $S_{G350:550}$, respectively. While spatial variability in S_{CDOM} were greater between provinces than variability by depth within provinces, $S_{E350:400}$ displayed ranges up to 0.007 nm^{-1} between provinces and by depth within provinces. From this, we suggest that *in situ* production pathways for CDOM are more variable by depth than across global ocean provinces.

We observed differences between slope values measured across the spectral ranges for each province, suggesting that each spectral range does not convey the same information about CDOM composition. To understand what could be determined about CDOM composition from each spectral range, we considered how $S_{G240:700}$, $S_{G300:700}$ and $S_{E350:550}$ related to $S_{E275:295}$, which is regarded as an indicator of source, molecular weight and photobleaching of CDOM [*D'Sa et al.*, 2014; *Helms et al.*, 2008], and $S_{E350:400}$, a slope range less entrenched in the literature but indicative of photobleaching and likely microbial processing and production of CDOM [*Helms et al.*, 2013; *Helms et al.*, 2015; *Matsuoka et al.*, 2015]. We found that $S_{E350:400}$ can characterize *in situ* production and degradation of CDOM with potential links to microbial processes [*Matsuoka et al.*, 2015; *Nelson et al.*, 2004; *Seidel et al.*, 2015] as this parameter varied throughout the global ocean and consistently increased with depth. $S_{G240:700}$ is strongly correlated with $S_{E275:295}$ in most regions (r^2=0.48-0.95, excluding BERS) suggesting that wavelengths below 300 nm

strongly influence S_{CDOM} when calculated using these wavelengths. $S_{G300:700}$ displayed variable relationships with $S_{E275:295}$ (r^2=0-0.90) and $S_{E350:400}$ (r^2=0.01-0.86), suggesting that this parameter is largely dependent on local processes. Typically, a region characterized by a poor relationship between $S_{300:700}$ and $S_{E275:295}$ displayed a strong relationship between $S_{300:700}$ and $S_{E350:400}$ (e.g. BPLR). A strong relationship with each parameter would be expected if $S_{300:700}$ represents CDOM composition in a similar manner to $S_{E275:295}$ or $S_{E350:400}$. However, considering that a consistent trend with one over the other was not observed, it is likely that $S_{G300:700}$ blurs characteristics of each into a single parameter that may not be particularly effective at characterizing the CDOM pool except under ideal circumstances such as a single, dominant process contributing to the CDOM pool. Thus, past work that has found potential relationships between molecular weight and $S_{G300:600}$, a parameter we found to behave quite similarly to $S_{G300:700}$, may be contingent on the environment [e.g. *Stedmon and Nelson 2015*].

Considering the variability and trends in slope for the different spectral ranges and relationships with $S_{E275:295}$, it is likely that $S_{G350:550}$ will be sensitive to a smaller suite of processes than spectral ranges that extend down to or below 300 nm. While $S_{G350:550}$ could be useful for accurately modeling a_{CDOM} above 350 nm, it appears to be limited for extending a_{CDOM} lower than 350 nm. $S_{G350:550}$ showed greater differentiation between biogeochemical provinces while maintaining a strong relationship with $S_{E350:400}$ in most provinces. Few studies have reported spectral slope from 350-550 nm to our knowledge [*Hancke et al.*, 2014; *Kowalczuk et al.*, 2006], making it difficult to directly assess what $S_{G350:550}$ estimates about CDOM composition. We propose that $S_{E350:400}$ is a better metric for tracking compositional changes in the CDOM pool affiliated with photodegradation and *in situ* production of CDOM as suggested by Helms et al. [2013, 2015]. However, $S_{G350:550}$ displayed greater uniqueness between provinces while $S_{E350:400}$ displayed more consistent trends with depth across all spectral ranges. If these trends are consistent in future datasets, we propose that $S_{G350:550}$ will be an ideal parameter for directly estimating compositional changes in autochthonous CDOM from hyperspectral ocean color data while changes in $S_{E350:400}$ will be indicative of vertical transport of unique CDOM or distinct *in situ* production pathways. Based on previous studies, $a_{CDOM}(\lambda)$ will likely remain a useful parameter for estimating terrestrial CDOM contributions from hyperspectral satellite observations [*Fichot et al.*, 2014; *Mannino et al.*, 2014]. These parameters appear to relate well with oceanic, *in situ* processes with the potential for tracking vertical movement of the CDOM pool with the added benefit of estimating CDOM composition.

1.4.3 Gaussian Components

The spectral range used to fit a_{CDOM} strongly impacted the number of Gaussian components fitted. When the first absorption peak was below 300 nm, presumed to be lignin [*McKnight and Aiken*, 2010], the goodness of fit increased so significantly that smaller peaks at longer wavelengths that were fitted in the 300-700 nm spectral range were no longer fitted. Fitting from 300-700 nm resulted in the most peaks fitted for all provinces, while 350-550 nm fitted fewer peaks than 300-700 nm. In waters where lignin is a strong

or moderate contributor to a_{CDOM} below 300 nm, fitting from 240-500 nm, then fitting from 300-500 nm, may be a better approach, although blending models could become an issue. The majority of waters sampled to 240 nm in the SeaBASS dataset occurred in waters that historically have a strong terrestrial component [*Benner et al.*, 2005; *D'Sa et al.*, 2014; *D'Sa and DiMarco*, 2009; *Del Vecchio and Blough*, 2004b], precluding a thorough analysis of the 240-700 nm spectral range across the full range of oceanic conditions observed in the entire SeaBASS dataset. However, $S_{E275:295}$ suggests that the trend in spectral slope values that account for a_{CDOM} below 300 nm prevails in the global oceans. For spectra fitted with a peak below 300 nm, mean peak location was 286 nm, suggesting that $S_{E275:295}$ is strongly influenced by the shape of the Gaussian component, if present, rather than the underlying exponential curve. It is likely that the large, complex molecular structure of lignin and the absorption peak associated with lignin drive the relationship between $S_{E275:295}$ and CDOM molecular weight [*McKnight and Aiken*, 2010]. Additionally, we observed a shift in the location of the lignin absorption peak from terrestrial waters to oceanic waters (Fig. 11a) consistent with photodegradation of this component [*Del Vecchio and Blough*, 2004a].

Absorption peaks at wavelengths less than 300 nm can extend beyond these wavelengths through a complex process of charge-transfer interactions in the CDOM pool [*Del Vecchio and Blough*, 2004a]. For waters strongly impacted by terrestrial material and displaying a low $S_{275:295}$ spectral signature, the first Gaussian component occurred at a much smaller wavelength in the spectra. Past studies have shown that terrestrial material absorption is dominated by lignin which absorbs below 300 nm [*McKnight and Aiken*, 2010; *Spencer et al.*, 2008]; however, the deviation from the baseline associated with this peak extends beyond 300 nm [*Fichot et al.*, 2016], resulting in a distortion of $S_{300:700}$ in these waters. Fitting Gaussian peaks provides a method to pick out unique components within the CDOM pool, this is similar to fitting fluorescence peaks in excitation-emission matrix spectroscopy (EEMs) and accounts for deviations that impact S_{CDOM} in the spectral range considered. More work is required to determine the significance of these absorption-based features to particular groups of molecules in the CDOM pool, including whether features fitted between 300-325 nm in the 300-700 nm spectral range are unique chromophores or a residual effect from the strong absorption of lignin absorption extending above 300 nm.

Across all regions, spectral ranges that were commonly fitted were associated with spectral locations (~350 nm, 375 nm) of chromophores that are likely photorefactory [*Helms et al.*, 2013] or chromophores that have been found to be produced from photobleaching of autochthonously produced CDOM [*Swan et al.*, 2012]. The latter, a feature observed between 410-420 nm and noted in previous studies [*Bricaud et al.*, 2010; *Swan et al.*, 2012], was the most commonly fitted peak across all provinces in the 300-700 nm spectral range. This feature was noticeably present across all spectral ranges, typically representing the second peak fitted in the 240-700 nm spectral range when more than one peak was fitted to the spectrum in this spectral range.

Two provinces that stood out regarding the average number of Gaussian peaks fitted per spectra were BPLR and BERS (Fig. 10). BPLR was dominated by measurements in the Chukchi Sea, a region with a greater magnitude of CDOM than the global ocean but less than most Arctic shelf regions due to predominantly authochthonously-produced CDOM [*Dainard and Guéguen*, 2013]. However, both BPLR and BERS have elevated levels of CDOM, likely a contributing factor to significantly more fitted peaks on average than other provinces. Both provinces were predominantly sampled during or shortly after the intense spring phytoplankton blooms associated with ice melt in these regions [*Arrigo et al.*, 2014; *Goes et al.*, 2013]. Phytoplankton absorption spectra in this region frequently display absorption in UV wavelengths due to the presence of mycosporine-like amino acids, a feature previously observed in the CDOM pool in Arctic habitats [*Pavlov et al.*, 2014]. This, in conjunction with an active microbial community [*Matsuoka et al.*, 2015] are potential explanations for why these regions were dominated by spectra with an average of more than four Gaussian components. The dramatic reduction in number of Gaussian peaks fitted to a_{CDOM} spectra in BERS and BPLR from 300-700 nm to 350-550 nm despite a relatively constant mean location of peaks (when fitted) suggests that further consideration can be given to the weighting factor in environments with a diverse CDOM pool if spectral range is a limiting factor. In these environments, reducing the weighting factor removes more residuals, allowing for more, smaller peaks to potentially be fit if the data presents an appropriate signal-to-noise (SNR) ratio. Locating and observing changes in these peaks will lend insight into links between the phytoplankton community and CDOM as well as the degradative state of CDOM.

APLR spectra were typically fitted with few features and stood out as a unique province in most analyses (Fig. 6-9). This is potentially due to low a_{CDOM} for some samples, but many samples contained $a_{CDOM}(350)$ greater than 0.1 m^{-1}. This province is characterized by autochthonously produced CDOM, with a distinct $S_{275:295}$ signal and a high correlation of $a_{CDOM}(325)$ with chlorophyll concentrations and upwelled waters transporting subsurface water with elevated levels of CDOM into the photic zone [*D'Sa and Kim*, 2017; *Ortega-Retuerta et al.*, 2010; *Ortega-Retuerta et al.*, 2009]. For many of these spectra, the lack of components is likely due to old, upwelled CDOM that behaves remarkably consistent across all spectral ranges, evidenced as an approximate one-to-one line in slope comparisons across different spectral ranges (not shown). This feature was also seen in CCAL, another province seasonally driven by upwelling and displaying different CDOM signatures between upwelled waters and waters dominated by phytoplankton blooms [*Day and Faloona*, 2009]. Considering that the number of Gaussian components decreased with depth for most provinces, zones of upwelling will likely display unique CDOM characteristics that include relatively uniform spectra with deviations resulting from recent biological contributions.

The utility of Gaussian peak height, φ, and width, σ, are less certain from our analysis. When we normalized for peak height by modeled a_{CDOM} at the location of μ, oceanic waters presented a broader range of normalized φ including larger values, suggesting that these peaks are more prominent features relative to other chromophores contributing to a_{CDOM} in these regions. This suggests these regions contain chromophores

that are consistently produced amid a background of relatively degraded CDOM or are photorefractory in nature, consistent with the spectral locations of the peaks relative to previous studies [*Helms et al.*, 2013; *Swan et al.*, 2012].

We did not find any significant trends in σ for any spectral ranges considered. While μ and φ are relatively intuitive features, the parameterization of σ carries more uncertainty related to the methodology. While peak location and height can shift due to changes in C (weighting factor for residual removal) and spectral range used for fitting, they are largely grounded in features of the a_{CDOM} spectra as evidenced by similarity in location and height across spectral ranges used. Peak width can change dramatically based on C and, to a lesser extent, spectral range used, suggesting less interpretability. For our analysis, C was held constant at one for the entire dataset, weighing spectra evenly across all wavelengths. Thus, while our results are not impacted by changes in C, it is an important consideration for the community as the method becomes utilized more broadly.

1.4.4 Applications to Remote Sensing

Previous studies have found reliable relationships between a_{CDOM} at specific wavelengths, including $a_{CDOM}(412)$ and $a_{CDOM}(443)$, and spectral slope values, including $S_{275:295}$ [e.g. *Mannino et al. 2014*]. For our analysis, $S_{275:295}$ could potentially be retrieved with reasonable accuracy in most biogeochemical provinces assuming an accurate retrieval of $a_{CDOM}(\lambda)$ and a predetermined relationship between these two parameters. However, the INLAND and BPLR regions displayed particularly poor relationships, suggesting high variability in CDOM pools and significant differences in CDOM characteristics with similar $a_{CDOM}(\lambda)$ values. Thus, some regions would be precluded from this methodology. Considering that most regions displayed unique trends and distributions of slope, a global relationship would poorly predict $S_{275:295}$.

NASA's Plankton, Aerosol, Cloud, ocean Ecosystem (PACE) sensor is anticipated for launch in the near future and is expected to have hyperspectral (every 5 nm) capabilities down to 350 nm. Considering this, it is prudent to advance the knowledge of what can be determined regarding CDOM composition using a_{CDOM} in this spectral range. To this end, we applied the methodology of Massicotte and Markager [2016] to identify spectral regions frequented by deviations that can impact satellite-derived estimates of chlorophyll-a and phytoplankton pigments, particularly methods utilizing band ratios as the most common region of deviations occurred from 410-440 nm. Hyperspectral sensors will allow for a baseline exponential absorption spectra, such as $S_{E350:550}$, to be fitted to a_{CDOM} spectra, allowing for these features to be ignored if the SNR of the sensor doesn't allow for confident fitting of these features. An accurate $S_{E350:550}$ directly estimated from hyperspectral satellite data will also aid in accounting for for deviations in a_{CDOM} that are currently centered on or near multispectral bands. Our approach allows for these spectral regions and their frequency to be assessed to determine if regional accounting for these features can decrease error propagated through the spectrum by an ill-defined $a_{CDOM}(\lambda_0)$.

We also investigated the impact of inaccurate S_{CDOM} values for estimating $a_{CDOM}(412)$ by comparing calculated $a_{CDOM}(412)$ to mean $a_{CDOM}(412)$ for each province. The resultant errors, from 4-22%, suggest that poorly parameterizing S_{CDOM} lends significant uncertainty to satellite estimates of IOP's strictly introduced through S_{CDOM}. For multispectral algorithms, we suggest accounting for differences in S_{CDOM} between provinces, as presented here, and continued consideration of the mean ratio of a_{NAP}:a_{CDOM} across distinct biogeochemical regions for methods that utilize S_{CDM}.

1.5 Conclusions

In an attempt to close the knowledge gap and move towards a common methodology, we have presented S_G values for three broad spectral ranges, 240-700 nm, 300-700 nm and 350-550 nm as well as S_E for 275-295 nm and 350-400 nm. We also explored the ability of the Gaussian decomposition method to expand our optical understanding of CDOM composition from a global database. Ideally, S_{CDOM} can both adequately model the a_{CDOM} spectra and relate to CDOM characteristics. We presented S_{CDOM} as a diagnostic tool that can provide insights into CDOM composition with the potential to be applied to hyperspectral ocean color applications for optical estimates of CDOM composition.

Of the slope ranges considered, we found $S_{275:295}$ and $S_{350:550}$ display clear, unique spatial trends that can be exploited for optically estimating changes in CDOM across diverse open ocean environments. $S_{350:400}$ displayed potential for tracking changes to *in situ* production of CDOM, particularly with depth at a given location. $S_{300:700}$ is a useful metric for particular regions if the underlying CDOM pool is known; however, this metric displayed good relationships with both $S_{275:295}$ and $S_{350:400}$, implying that the metric itself does not clearly delineate specific compositional traits that impact the optical signature of CDOM.

Hyperspectral capabilities allow for direct estimates of S_{CDOM}, providing insight into CDOM degradative state and *in situ* production pathways. However, S_{CDOM} calculated using wavelengths anticipated to be available from PACE (> ~350 nm) differs significantly from $S_{E275:295}$, requiring alternative methods for estimating terrestrial contribution, lignin content and molecular weight. The divergence in optical properties of $S_{E350:400}$ and $S_{E275:295}$ observed here and documented in the literature suggest that $S_{E350:400}$ can provide insight into *in situ* production pathways; however, further consideration should be given to this parameter in future studies along with whether a broader spectral range such as $S_{G350:550}$ provides more insight into these processes.

Beyond tracking changes in CDOM and presenting a way to estimate CDOM composition, S_{CDOM} is also important for accurately modeling a_{CDOM}. We considered the mean $a_{CDOM}(443)$ for each province and assessed how much error is introduced to the a_{CDOM} parameter using an assumed S_{CDOM} of 0.0154 nm^{-1} and propagating a_{CDOM} to 412 nm. We found this assumed slope introduced errors in $a_{CDOM}(412)$ of 4-22% across all provinces relative to the mean $a_{CDOM}(412)$ observed (Fig. 12). Thus, poorly parameterizing

S_{CDOM} with a constant global slope can introduce a similar, if not greater, level of uncertainty in retrieved a_{CDOM} to mischaracterizing the percent contribution of a_{NAP} to $a_{CDM}(412)$.

Ultimately, CDOM will be best considered using a suite of metrics applied to ocean color imagery. Past work focusing on relationships between a_{CDOM} at a given reference wavelength to estimate $S_{E275:295}$ and lignin content can continue to be improved using *in situ* data and are anticipated to provide additional information not directly available from the anticipated PACE mission. The emphasis should be on continuing to relate quality, *in situ* measurements with $S_{E350:400}$ and $S_{G350:550}$, two metrics anticipated to be directly available via NASA's PACE sensor in the near future, to maximize data potential from remotely-sensed imagery. It is our view that, prior to mission launch, the community will be well served with a better understanding of what information is directly retrievable with S_{CDOM} and which spectral range is best suited for discrimination between distinct CDOM pools within the spectral capability of the mission.

1.6 References

Andrew, A. A., R. Del Vecchio, A. Subramaniam, and N. V. Blough (2013), Chromophoric dissolved organic matter (CDOM) in the Equatorial Atlantic Ocean: Optical properties and their relation to CDOM structure and source, *Marine Chemistry*, *148*, 33-43.

Arrigo, K. R., et al. (2014), Phytoplankton blooms beneath the sea ice in the Chukchi sea, *Deep Sea Research Part II: Topical Studies in Oceanography*, *105*, 1-16.

Benner, R., P. Louchouarn, and R. M. W. Amon (2005), Terrigenous dissolved organic matter in the Arctic Ocean and its transport to surface and deep waters of the North Atlantic, *Global Biogeochemical Cycles*, *19*(2).

Boyle, E. S., N. Guerriero, A. Thiallet, R. D. Vecchio, and N. V. Blough (2009), Optical Properties of Humic Substances and CDOM: Relation to Structure, *Environmental Science & Technology*, *43*(7), 2262-2268.

Bricaud, A., M. Babin, H. Claustre, J. Ras, and F. Tièche (2010), Light absorption properties and absorption budget of Southeast Pacific waters, *Journal of Geophysical Research*, *115*(C8).

Carder, K. L., R. G. Steward, G. R. Harvey, and P. B. Ortner (1989), Marine humic and fulvic acids: Their effects on remote sensing of ocean chlorophyll, *Limnology and Oceanography*, *34*(1), 68-81.

Coble, P. G. (2007), Marine Optical Biogeochemistry: The Chemistry of Ocean Color, *Chemical Reviews*, *107*(2), 402-418.

D'Sa, E. J., and H.-c. Kim (2017), Surface Gradients in Dissolved Organic Matter Absorption and Fluorescence Properties along the New Zealand Sector of the Southern Ocean, *Frontiers in Marine Science*, *4*.

D'Sa, E. J., J. I. Goes, H. Gomes, and C. Mouw (2014), Absorption and fluorescence properties of chromophoric dissolved organic matter of the eastern Bering Sea in the summer with special reference to the influence of a cold pool, *Biogeosciences*, *11*(12), 3225-3244.

D'Sa, E. J., and S. F. DiMarco (2009), Seasonal variability and controls on chromophoric dissolved organic matter in a large river-dominated coastal margin, *Limnology & Oceanography*, *54*(6), 2233-2242.

Dainard, P. G., and C. Guéguen (2013), Distribution of PARAFAC modeled CDOM components in the North Pacific Ocean, Bering, Chukchi and Beaufort Seas, *Marine Chemistry*, *157*, 216-223.

Day, D. A., and I. Faloona (2009), Carbon monoxide and chromophoric dissolved organic matter cycles in the shelf waters of the northern California upwelling system, *Journal of Geophysical Research*, *114*(C1).

Del Vecchio, R., and N. V. Blough (2002), Photobleaching of chromophoric dissolved organic matter in natural waters: kinetics and modeling, *Marine Chemistry*, *78*, 231-253.

Del Vecchio, R., and N. V. Blough (2004a), Spatial and seasonal distribution of chromophoric dissolved organic matter and dissolved organic carbon in the Middle Atlantic Bight, *Marine Chemistry*, *89*(1-4), 169-187.

Del Vecchio, R., and N. V. Blough (2004b), On the Origin of the Optical Properties of Humic Substances, *Environmental Science & Technology*, *38*(14), 3885-3891.

Dong, Q., S. Shang, and Z. Lee (2013), An algorithm to retrieve absorption coefficient of chromophoric dissolved organic matter from ocean color, *Remote Sensing of Environment*, *128*, 259-267.

Dutkiewicz, S., A. E. Hickman, O. Jahn, W. W. Gregg, C. B. Mouw, and M. J. Follows (2015), Capturing optically important constituents and properties in a marine biogeochemical and ecosystem model, *Biogeosciences*, *12*(14), 4447-4481.

Fichot, C. G., and R. Benner (2011), A novel method to estimate DOC concentrations from CDOM absorption coefficients in coastal waters, *Geophysical Research Letters*, *38*(3).

Fichot, C. G., S. E. Lohrenz, and R. Benner (2014), Pulsed, cross-shelf export of terrigenous dissolved organic carbon to the Gulf of Mexico, *Journal of Geophysical Research: Oceans*, *119*(2), 1176-1194.

Fichot, C. G., R. Benner, K. Kaiser, Y. Shen, R. M. W. Amon, H. Ogawa, and C.-J. Lu (2016), Predicting Dissolved Lignin Phenol Concentrations in the Coastal Ocean from Chromophoric Dissolved Organic Matter (CDOM) Absorption Coefficients, *Frontiers in Marine Science*, *3*.

Goes, J. I., H. do Rosario Gomes, E. Haugen, K. McKee, E. D'Sa, A. M. Chekalyuk, D. Stoecker, P. Stabeno, S. Saitoh, and R. Sambrotto (2013), Fluorescence, pigment

and microscopic characterization of Bering Sea phytoplankton community structure and photosynthetic competency in the presence of a Cold Pool during summer, *Deep Sea Research Part II: Topical Studies in Oceanography*.

Granskog, M. A., C. A. Stedmon, P. A. Dodd, R. M. W. Amon, A. K. Pavlov, L. de Steur, and E. Hansen (2012), Characteristics of colored dissolved organic matter (CDOM) in the Arctic outflow in the Fram Strait: Assessing the changes and fate of terrigenous CDOM in the Arctic Ocean, *Journal of Geophysical Research*, *117*(C12).

Grunert, B. K., C. B. Mouw, A. B. Ciochetto (2018), Characterizing CDOM variability across diverse regions and spectral ranges, *Global Biogeochemical Cycles, 32,* 57-77.

Hancke, K., E. K. Hovland, Z. Volent, R. Pettersen, G. Johnsen, M. Moline, and E. Sakshaug (2014), Optical properties of CDOM across the Polar Front in the Barents Sea: Origin, distribution and significance, *Journal of Marine Systems*, *130*, 219-227.

Hansell, D. A. (2013), Recalcitrant dissolved organic carbon fractions, *Annual review of marine science*, *5*, 421-445.

Hansen, A. M., T. E. C. Kraus, B. A. Pellerin, J. A. Fleck, B. D. Downing, and B. A. Bergamaschi (2016), Optical properties of dissolved organic matter (DOM): Effects of biological and photolytic degradation, *Limnology and Oceanography*, *61*(3), 1015-1032.

Hayter, A. (1986), The Maximum Familywise Error Rate of Fisher's Least Significant Difference Test, *Journal of American Statistical Association*, *81*(396), 1000-1004.

Helms, J. R., A. Stubbins, J. D. Ritchie, E. C. Minor, D. J. Kieber, and K. Mopper (2008), Absorption spectral slopes and slope ratios as indicators of molecular weight, source, and photobleaching of chromophoric dissolved organic matter, *Limnology and Oceanography*, *53*(3), 955-969.

Helms, J. R., A. Stubbins, E. M. Perdue, N. W. Green, H. Chen, and K. Mopper (2013), Photochemical bleaching of oceanic dissolved organic matter and its effect on absorption spectral slope and fluorescence, *Marine Chemistry*, *155*, 81-91.

Helms, J. R., J. Mao, H. Chen, E. M. Perdue, N. W. Green, P. G. Hatcher, K. Mopper, and A. Stubbins (2015), Spectroscopic characterization of oceanic dissolved organic matter isolated by reverse osmosis coupled with electrodialysis, *Marine Chemistry*, *177*, 278-287.

Hernes, P. J., and R. Benner (2003), Photochemical and microbial degradation of dissolved lignin phenols: Implications for the fate of terrigenous dissolved organic matter in marine environments, *Journal of Geophysical Research*, *108*(C9).

Hickman, A. E., S. Dutkiewicz, R. G. Williams, and M. J. Follows (2010), Modelling the effects of chromatic adaptation on phytoplankton community structure in the oligotrophic ocean, *Marine Ecology Progress Series*, *406*, 1-17.

Hoepffner, N., and S. Sathyendranath (1993), Determination of the major groups of phytoplankton pigments from the absorption spectra of total particulate matter, *Journal of Geophysical Research*, *98*(C12), 22789.

Ioannou, I., A. Gilerson, B. Gross, F. Moshary, and S. Ahmed (2013), Deriving ocean color products using neural networks, *Remote Sensing of Environment*, *134*, 78-91.

Kim, G. E., A. Gnanadesikan, and M.-A. Pradal (2016), Increased Surface Ocean Heating by Colored Detrital Matter (CDM) Linked to Greater Northern Hemisphere Ice Formation in the GFDL CM2Mc ESM, *Journal of Climate*, *29*(24), 9063-9076.

Kowalczuk, P., C. A. Stedmon, and S. Markager (2006), Modeling absorption by CDOM in the Baltic Sea from season, salinity and chlorophyll, *Marine Chemistry*, *101*(1-2), 1-11.

Kowalczuk, P., J. Stoń-Egiert, W. J. Cooper, R. F. Whitehead, and M. J. Durako (2005), Characterization of chromophoric dissolved organic matter (CDOM) in the Baltic Sea by excitation emission matrix fluorescence spectroscopy, *Marine Chemistry*, *96*(3-4), 273-292.

Lee, Z., K. L. Carder, and R. A. Arnone (2002), Deriving inherent optical properties from water color: a multiband quasi-analytical algorithm for optically deep waters, *Applied optics*, *41*(27), 5755-5772.

Mannino, A., M. E. Russ, and S. B. Hooker (2008), Algorithm development and validation for satellite-derived distributions of DOC and CDOM in the U.S. Middle Atlantic Bight, *Journal of Geophysical Research*, *113*(C7).

Mannino, A., M. G. Novak, S. B. Hooker, K. Hyde, and D. Aurin (2014), Algorithm development and validation of CDOM properties for estuarine and continental shelf waters along the northeastern U.S. coast, *Remote Sensing of Environment*, *152*, 576-602.

Maritorena, S., D. A. Siegel, and A. R. Peterson (2002), Optimization of a semianalytical ocean color model for global-scale applications, *Applied optics*, *41*(15), 2705-2714.

Massicotte, P., and J.-J. Frenette (2011), Spatial connectivity in a large river system: resolving the sources and fate of dissolved organic matter, *Ecological Applications*, *21*(7), 2600-2617.

Massicotte, P., and S. Markager (2016), Using a Gaussian decomposition approach to model absorption spectra of chromophoric dissolved organic matter, *Marine Chemistry*, *180*, 24-32.

Matsuoka, A., S. B. Hooker, A. Bricaud, B. Gentili, and M. Babin (2013), Estimating absorption coefficients of colored dissolved organic matter (CDOM) using a semi-analytical algorithm for southern Beaufort Sea waters: application to deriving concentrations of dissolved organic carbon from space, *Biogeosciences*, *10*(2), 917-927.

Matsuoka, A., E. Ortega-Retuerta, A. Bricaud, K. R. Arrigo, and M. Babin (2015), Characteristics of colored dissolved organic matter (CDOM) in the Western Arctic Ocean: Relationships with microbial activities, *Deep Sea Research Part II: Topical Studies in Oceanography*, *118*, 44-52.

McKnight, D. M., and G. R. Aiken (2010), Sources and Age of Aquatic Humus, in *Aquatic Humic Substances: Ecology and Biogeochemistry*, edited by D. O. Hessen and L. J. Tranvik, Springer-Verlag Berlin Heidelberg, Germany.

Meler, J., P. Kowalczuk, M. Ostrowska, D. Ficek, M. Zabłocka, and A. Zdun (2016), Parameterization of the light absorption properties of chromophoric dissolved organic matter in the Baltic Sea and Pomeranian lakes, *Ocean Science*, *12*(4), 1013-1032.

Nelson, N. B., D. A. Siegel, and A. F. Michaels (1998), Seasonal dynamics of colored dissolved material in the Sargasso Sea, *Deep-Sea Research I*, *45*, 931-957.

Nelson, N. B., C. A. Carlson, and D. K. Steinberg (2004), Production of chromophoric dissolved organic matter by Sargasso Sea microbes, *Marine Chemistry*, *89*(1-4), 273-287.

Nelson, N. B., D. A. Siegel, C. A. Carlson, and C. M. Swan (2010), Tracing global biogeochemical cycles and meridional overturning circulation using chromophoric dissolved organic matter, *Geophysical Research Letters*, *37*(3).

Organelli, E., A. Bricaud, D. Antoine, and A. Matsuoka (2014), Seasonal dynamics of light absorption by chromophoric dissolved organic matter (CDOM) in the NW Mediterranean Sea (BOUSSOLE site), *Deep Sea Research Part I: Oceanographic Research Papers*, *91*, 72-85.

Ortega-Retuerta, E., I. Reche, E. Pulido-Villena, S. Agustí, and C. M. Duarte (2010), Distribution and photoreactivity of chromophoric dissolved organic matter in the Antarctic Peninsula (Southern Ocean), *Marine Chemistry*, *118*(3-4), 129-139.

Ortega-Retuerta, E., T. Frazer, C. M. Duarte, S. Ruiz-Halpern, A. Tovar-Sanchez, J. M. Arrieta, and I. Reche (2009), Biogeneration of chromophoric dissolved organic matter by bacteria and kirll in the Southern Ocea, *Limnology & Oceanography*, *54*(6), 1941-1950.

Pavlov, A. K., A. Silyakova, M. A. Granskog, R. G. J. Bellerby, A. Engel, K. G. Schulz, and C. P. D. Brussaard (2014), Marine CDOM accumulation during a coastal Arctic mesocosm experiment: No response to elevated pCO2levels, *Journal of Geophysical Research: Biogeosciences*, *119*(6), 1216-1230.

Pavlov, A. K., C. A. Stedmon, A. V. Semushin, T. Martma, B. V. Ivanov, P. Kowalczuk, and M. A. Granskog (2016), Linkages between the circulation and distribution of dissolved organic matter in the White Sea, Arctic Ocean, *Continental Shelf Research*, *119*, 1-13.

Pérez, G. L., M. Galí, S.-J. Royer, H. Sarmento, J. M. Gasol, C. Marrasé, and R. Simó (2016), Bio-optical characterization of offshore NW Mediterranean waters:

CDOM contribution to the absorption budget and diffuse attenuation of downwelling irradiance, *Deep Sea Research Part I: Oceanographic Research Papers*, *114*, 111-127.

Rafter, J. A., M. L. Abell, and J. P. Braselton (2002), Multiple comparison methods for means, *SIAM Review*, *44*(2), 259-278.

Reche, I., M. L. Pace, and J. J. Cole (2000), Modeled Effects of Dissolved Organic Carbon and Solar Spectra on Photobleaching in Lake Ecosystems, *Ecosystems*, *3*(5), 419-432.

Roesler, C. S., M. J. Perry, and K. L. Carder (1989), Modeling in situ phytoplankton absorption from total absorption spectra in productive inland marine waters, *Limnology & Oceanography*, *34*(8), 1510-1523.

Schwarz, J. N., et al. (2002), Two models for absorption by coloured dissolved organic matter (CDOM), *Oceanologia*, *44*(2), 209-241.

Seidel, M., P. L. Yager, N. D. Ward, E. J. Carpenter, H. R. Gomes, A. V. Krusche, J. E. Richey, T. Dittmar, and P. M. Medeiros (2015), Molecular-level changes of dissolved organic matter along the Amazon River-to-ocean continuum, *Marine Chemistry*, *177*, 218-231.

Siegel, D. A., S. Maritorena, N. B. Nelson, M. Behrenfeld, and C. R. McClain (2005), Colored dissolved organic matter and its influence on the satellite-based characterization of the ocean biosphere, *Geophysical Research Letters*, *32*(20).

Spencer, R. G. M., G. R. Aiken, K. P. Wickland, R. G. Striegl, and P. J. Hernes (2008), Seasonal and spatial variability in dissolved organic matter quantity and composition from the Yukon River basin, Alaska, *Global Biogeochemical Cycles*, *22*(4).

Stedmon, C. A., and S. Markager (2003), Behaviour of the optical properties of coloured dissolved organic matter under conservative mixing, *Estuarine, Coastal and Shelf Science*, *57*(5-6), 973-979.

Stedmon, C. A., and S. Markager (2005), Resolving the variability in dissolved organic matter fluorescence in a temperate estuary and its catchment using PARAFAC analysis, *Limnology & Oceanography*, *50*(2), 686-697.

Stedmon, C. A., R. M. W. Amon, A. J. Rinehart, and S. A. Walker (2011), The supply and characteristics of colored dissolved organic matter (CDOM) in the Arctic Ocean: Pan Arctic trends and differences, *Marine Chemistry*, *124*(1-4), 108-118.

Swan, C. M., N. B. Nelson, D. A. Siegel, and T. S. Kostadinov (2012), The effect of surface irradiance on the absorption spectrum of chromophoric dissolved organic matter in the global ocean, *Deep Sea Research Part I: Oceanographic Research Papers*, *63*, 52-64.

Twardowski, M. S., E. Boss, J. M. Sullivan, and P. L. Donaghay (2004), Modeling the spectral shape of absorption by chromophoric dissolved organic matter, *Marine Chemistry*, *89*(1-4), 69-88.

Visser, S. A. (1983), Fluorescence phenomena of humic matter of aquatic origin and microbial cultures, *Aquatic and terrestrial humic materials*, 183-202.

Walker, S. A., R. M. W. Amon, and C. A. Stedmon (2013), Variations in high-latitude riverine fluorescent dissolved organic matter: A comparison of large Arctic rivers, *Journal of Geophysical Research: Biogeosciences*, *118*(4), 1689-1702.

Werdell, P. J., S. W. Bailey, G. S. Fargion, C. Pietras, K. D. Knobelspiesse, G. C. Feldman, and C. R. McClain (2003), Unique data repository facilitates ocean color satellite validation, *EOS Trans. AGU*, *84*(38), 377.

Wünsch, U. J., K. R. Murphy, and C. A. Stedmon (2015), Fluorescence Quantum Yields of Natural Organic Matter and Organic Compounds: Implications for the Fluorescence-based Interpretation of Organic Matter Composition, *Frontiers in Marine Science*, *2*.

Xiu, P., and F. Chai (2014), Connections between physical, optical and biogeochemical processes in the Pacific Ocean, *Progress in Oceanography*, *122*, 30-53.

2 Sensitivity of hyperspectral reflectance to colored dissolved organic matter spectral variability

2.1 Introduction

Ocean color remote sensing is a powerful tool for observing marine biogeochemistry, including monitoring the aquatic carbon cycle, managing fisheries and monitoring socio-economic hazards including harmful algal blooms (Friedland et al. 2012; Mannino et al. 2014; Stumpf et al. 2003; Vantrepotte et al. 2015). The utility of these instruments is largely derived from the ability to observe variability in remotely-sensed reflectance due to the primary absorbing constituents in seawater, or inherent optical properties (IOPs), of the water column – namely, phytoplankton, non-algal particles (NAP) and colored dissolved organic matter (CDOM). IOPs directly relate to many biological and biogeochemical ocean processes due to the magnitude and spectral shape of their spectra. CDOM absorption coefficients have been used to estimate dissolved organic carbon (DOC) export in the coastal zone (Cao et al. 2018; Mannino et al. 2008; Matsuoka et al. 2014; Vantrepotte et al. 2015), degradation of terrestrially-derived material (Fichot et al. 2014; Fichot et al. 2016) and *in situ* production of dissolved organic matter (Asmala et al. 2018; Danhiez et al. 2017; Nelson et al. 2010). Phytoplankton absorption coefficients and chlorophyll-a concentration (Chl) have been used to parameterize primary production (Behrenfeld and Falkowski 1997; Silsbe et al. 2016), constrain deep ocean carbon flux estimates (e.g. Mouw et al. 2016), optically estimate phytoplankton community composition or functional type via pigment identification (Bracher et al. 2017; Mouw et al. 2017b; Mouw and Yoder 2010; Uitz et al. 2015) and aid in accurately modeling sustainable fisheries harvest (e.g. Friedland et al. 2012).

Addressing biogeochemical variability represented by IOP's requires accurate retrieval of CDOM, phytoplankton and NAP absorption ($a_g(\lambda)$, $a_{ph}(\lambda)$, and $a_d(\lambda)$, respectively; λ indicates wavelength of measurement). Historically, observation of these IOP's at visible wavelengths has resulted in a lack of single solutions to inversion of remotely-sensed reflectance ($R_{rs}(\lambda)$; Defoin-Platel and Chami 2007; Sydor et al. 2004). From this, researchers have considered how the addition of ultraviolet (UV) wavelengths aid in solving inversion of $R_{rs}(\lambda)$ problems due to increased sensitivity of these wavelengths to $a_g(\lambda)$ (Wei et al. 2016). The assumption for this is that $a_{ph}(\lambda)$ decreases in the UV, while $a_g(\lambda)$ and $a_d(\lambda)$ exponentially increases with decreasing wavelength. However, phytoplankton absorption in the UV can be a significant contributor, particularly in the presence of mycosporine-like amino acids (MAAs) that limit UV damage to phytoplankton cellular structures (e.g. Carreto et al. 1990). Additionally, modeling sensitivity to CDOM spectral shape, described by the spectral slope parameter (S_g), is complicated by sensitivity to the initial wavelength used to propagate theoretical $a_g(\lambda)$ spectra (e.g. Wei et al. 2016). Beyond consideration of wavebands that help in constraining inverted $R_{rs}(\lambda)$ towards a more limited set of solutions, the signal-to-noise (SNR) of a sensor also determines the ability to view spectral features. Satellite-derived $R_{rs}(\lambda)$ is

notoriously noisy when compared to *in situ* measurements; thus, while optical techniques established for *in situ* data may prove quite promising, application to a sensor requires an understanding of sensor limitations alongside constraining the observed signal to a limited set of solutions.

A significant parameter in limiting uncertainty in absorption retrieved from satellite-derived $R_{rs}(\lambda)$ is the shape of the exponential absorption signal due to $a_g(\lambda)$ and $a_d(\lambda)$, $a_{dg}(\lambda)$, parameterized as spectral slope (S_{dg}) following the exponential model:

$$a_{dg}(\lambda) = a_{dg}(\lambda_0)e^{-S_{dg}(\lambda-\lambda_0)} \quad (1)$$

where λ_0 is a reference wavelength (e.g. 440 nm). Current algorithms utilize observed $R_{rs}(\lambda)$ at specific wavelengths to characterize $a_{dg}(\lambda_0)$, relying on either an assumed S_{dg} or an S_{dg} estimated from $R_{rs}(\lambda)$ ratios to model $a_{dg}(\lambda)$ (Lee et al. 2002; Mannino et al. 2014; Maritorena et al. 2002; Matsuoka et al. 2013; Werdell et al. 2013). Depending on the magnitude of $a_{dg}(\lambda)$ in the system and relative accuracy of S_{dg}, considerable uncertainty can be added to retrievals of $a_{ph}(\lambda)$ and products derived from $a_{ph}(\lambda)$ (e.g. Organelli et al. 2016). Due to the problem of more unknowns than knowns in inversion procedures, S_{dg} is typically assumed at a value from 0.015-0.018 nm^{-1} (Lee et al. 2002; Werdell et al. 2013; Maritorena et al. 2002). The spectral shape of NAP (S_d) is considerably less variable in the global ocean than S_g (Babin et al. 2003; Siegel et al. 2002), while S_g has been linked to specific production and degradation processes of CDOM, particularly in riverine, inland and coastal waters (Danhiez et al. 2017; Helms et al. 2008). Thus, S_g contains biogeochemical information relevant to understanding turnover production and degradation dynamics of the marine DOM pool, a pool of carbon that accounts for ~98% of oceanic organic carbon (citation). From this, researchers have attempted to estimate, rather than assume S_g, although to date, no method provides an estimation free of explicit or semi-explicit assumptions (e.g. Dong et al. 2013; Matsuoka et al. 2013).

NASA is planning the Plankton, Aerosol, Cloud and ocean Ecosystem (PACE) sensor with hyperspectral capability ($R_{rs}(\lambda)$ resolution $\leq$ 5 nm). This sensor has been posited to provide an enormous leap forward in our ability to accurately constrain biogeochemical parameters currently retrieved as well as providing the ability to observe new parameters, including S_g (e.g. Wei et al. 2016; Vandermeulen et al. 2017). Our understanding of the sensitivity of $R_{rs}(\lambda)$ to changes in S_g, however, has not been considered. Considering the future capabilities of PACE and useful biogeochemical information provided by S_g, we detail the impact of changes in S_g on $R_{rs}(\lambda)$ across a large lake (Lake Superior). We relate changes in $R_{rs}(\lambda)$ to the proposed SNR of PACE to determine spectral ranges that could be utilized to estimate S_g as well as the sensitivity of an estimated S_g across diverse environmental conditions. Finally, we relate the proposed accuracy of determining S_g to observed variability in Lake Superior and relate this to our ability to detect biogeochemical variability affiliated with S_g.

2.2 Methods

2.2.1 Study Site

Lake Superior is one of the largest lakes in the world and dominates the local climate. Despite its size, carbon cycling in Lake Superior is strongly influenced by terrestrial contributions (Minor et al. 2012; Urban et al. 2005). These contributions cause CDOM to dominate the absorption budget (Mouw et al. 2013) with degradation of terrestrial CDOM largely driven by photodegradation (Ma and Green 2004; MacDonald and Minor 2013) and long-term DOM transformations driven by microbial use of DOM, particularly photodegraded material (Biddanda and Cotner 2003; Hiriart-Baer et al. 2011; McManus et al. 2003). Lake Superior is an oligotrophic system (Barbiero and Tuchman 2004), which would suggest limited contribution of marine-produced CDOM to the total DOM budget; however, annual autochthonous DOC production is estimated to be quite similar to annual terrestrial DOC loading (Cotner et al. 2004; Sterner 2010; Urban et al. 2005). Lake Superior is also experiencing the most rapid warming of all the Laurentian Great Lakes (Austin and Colman 2007). This has been largely attributed to decreases in ice extent, although both ice cover and surface heat flux play a large role in temperature fluctuations for this system (Austin and Colman 2007). Due to large perturbations from climate change, an increase in CDOM quantity and quality from terrestrial systems is anticipated, as well as changes in photochemistry due to a decrease in winter ice cover (e.g. Creed et al. 2017; Austin and Colman 2007). These biogeochemical changes should be readily observable as changes in $a_g(\lambda)$ magnitude and spectral shape; thus, Lake Superior provides an ideal testing environment for PACE capabilities related to use of $a_g(\lambda)$ to observe sensitivity of carbon cycling to climate warming (Danhiez et al. 2017; Romera-Castillo et al. 2011; Williamson et al. 2015).

2.2.2 Data

A detailed description of the full dataset collected in Lake Superior, of which this data is a subset, can be found in Mouw et al. (2017a). Here we provide a brief overview of the observed parameters and their collection methods used in this analysis. Optical data were collected during ice-free months (May-October) of 2014-2016 (Fig. 1). Radiometric measurements were made with Hyper-OCR spectral radiometers (Satlantic Inc.) that measure between 350-800 nm at an ~3 nm resolution (137 total wavelengths). All $R_{rs}(\lambda)$ were derived from surface casts where the radiometer was equipped with a flotation collar by which the surface light field is characterized by collecting surface irradiance, $E_s(\lambda)$ (μW cm^{-2}) and upwelling radiance, $L_u(\lambda)$ (μW cm^{-2} sr^{-1}) below the surface is used to calculate $R_{rs}(\lambda)$ (sr^{-1}) as described in Mouw et al. (2017a). CDOM, non-algal particulate and phytoplankton absorption were measured spectrophotometrically (PerkinElmer; Lambda 35 UV/Vis dual beam) for wavelengths between 300 and 800 nm. CDOM absorption was measured in a 10 cm cuvette following NASA ocean optics protocols (Mueller et al. 2003). Particulate absorption (a_p) and a_d were measured following the transmission-reflectance (T-R) method (Lohrenz 2000; Lohrenz et al. 2003; Tassan and Ferrari 2002) using an

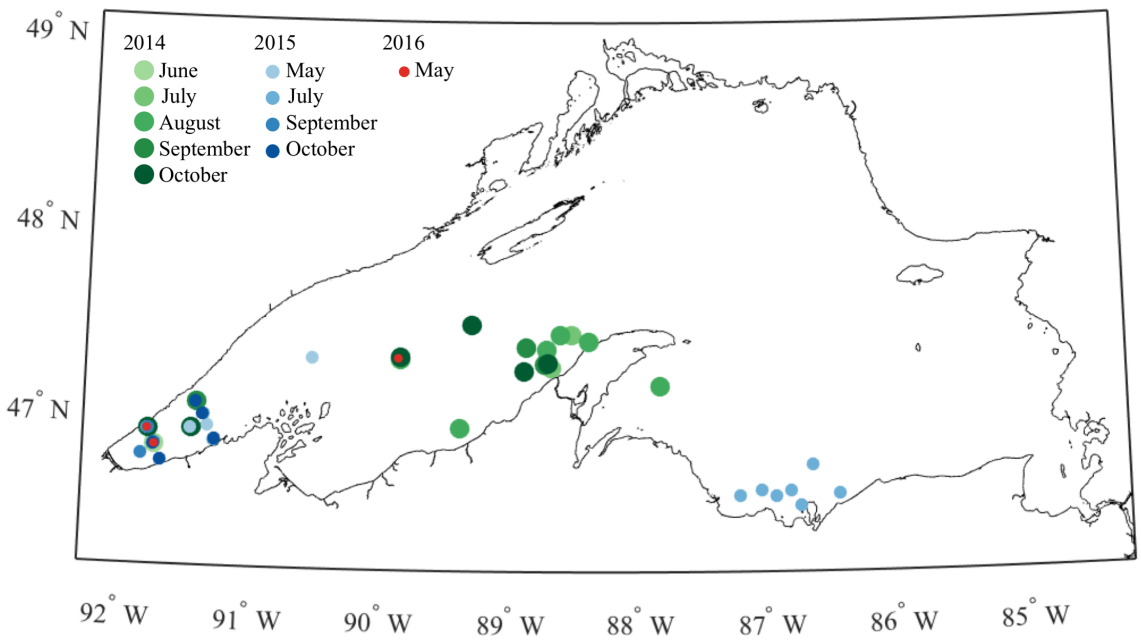


Figure 1. Stations map with marker color and size indicating year and marker color and shade indicating month of a given year.

integrating sphere. Blanks were prepared and measured daily to normalize T-R data. Phytoplankton absorption was derived as the difference between before (a_p) and after (a_d) exposure to sodium hypochlorite. Particulate backscattering ($b_{bp}(\lambda)$) was characterized using an ECO-BB9 sensor (Wetlabs) on board a profiling bio-optical package by removing scattering due to pure water (b_{bw}) from total measured backscattering. All available measurements (binned to 1 m values; Mouw et al. 2017a) in the top 10 m were averaged to produce a $b_{bp}(\lambda)$ spectra representative of the upper euphotic zone; $b_{bp}(\lambda)$ typically displayed consistent values in this depth range.

2.2.3 HydroLight

We used the radiative transfer software HydroLight (Mobley 1988) to simulate remotely-sensed reflectance for each station sampled in Lake Superior. Simulations were initialized with discrete absorption spectra collected within 10 m of the surface, discrete Chl collected coincident with absorption spectra and mean $b_p(535)$ and $b_{bp}(535)$ collected in the top 10 m of the water column (input for HydroLight as the ratio of $b_{bp}(\lambda)/b_p(\lambda)$ and observed $b_{bp}(\lambda)$). Particulate scattering and backscattering displayed consistent values in this depth range.

Radiative transfer simulations were run assuming that the mixed layer was homogeneous and within an infinitely deep water column and model outputs were obtained at a depth of 0 m (lake-atmosphere interface). Sky conditions were based on observed cloud cover and visibility. Solar angle was calculated using a NOAA solar angle calculator

(https://www.esrl.noaa.gov/gmd/grad/solcalc/) from the latitude, longitude and date of sampling. All samples were taken between 900 and 1500 local time and solar angle at 1200 was assumed for model input. Wind speed was set to 3 $m \cdot s^{-1}$, a value generally representative of conditions during sampling. Raman scattering and chlorophyll-a fluorescence were included in all runs. We initially included CDOM fluorescence but this resulted in a significantly higher simulated $R_{rs}(\lambda)$ for all stations relative to observed $R_{rs}(\lambda)$, suggesting that the quantum yield used to parameterize CDOM fluorescence in HydroLight dramatically overestimates actual CDOM fluorescence in Lake Superior. Model runs were performed from 350-700 nm at 5 nm increments to assess sensitivity and retrieval improvement from PACE. For *in situ* data that were not observed at these wavelengths, data were linearly interpolated (e.g. $b_{bp}(\lambda)$ measured at 9 wavelengths). Data that were sampled at a higher frequency (1 nm) were downsampled to 5 nm resolution from 350-700 nm for a total of 71 data points (e.g. $a_g(\lambda)$).

We assessed the impact of varying S_g on $R_{rs}(\lambda)$ by simulating identical conditions for each station but altered input $a_g(\lambda)$ by taking the ensemble mean of $a_g(\lambda)$ spectra calculated using λ_0 of 350, 400, 450 and 500 nm and the respective S_g. We varied S_g ±0.005 nm^{-1} from observed S_g calculated from 350-700 nm ($S_{350:700}$), varying in 0.001 nm^{-1} increments. This range of S_g values was chosen based on the approximate range of observed values in the dataset, with this range consistent across different ultraviolet-visible (UV-vis) wavelength ranges considered and used in the literature (e.g. 350-400 nm, 300-600 nm). In effect, this created 10 additional simulations per sampled station to assess the sensitivity of $R_{rs}(\lambda)$ to changes in S_g. We utilized $S_{350:700}$ for all HydroLight simulations, but present results for *in situ* S_g in the spectral range of 300-600 nm, as this spectral range captures additional spectral variability and has been linked to CDOM bulk molecular weight (Stedmon and Nelson 2014; Wünsch et al. 2018).

2.2.4 Data Analysis

We considered the accuracy of HydroLight simulations through comparison of simulated and measured $R_{rs}(\lambda)$. Bias and mean absolute percentage error (MAPE) were calculated as:

$$Bias = \frac{1}{n}\sum_{i=1}^{n}\left(x_i^{estimated} - x_i^{observed}\right) \quad (2)$$

$$MAPE = \frac{\sum_{i=1}^{n}\left(\left|x_i^{estimated} - x_i^{observed}\right|\right)}{n} \cdot 100 \quad (3)$$

There is an evident bias in simulation results when altering $a_g(\lambda)$ based on the initial λ_0 chosen to represent a spectrally-altered a_g spectra (e.g. Wei et al. 2016). We attempted to address this through an ensemble mean of alterations to $a_g(\lambda)$ modeled with Eq. 1. However, ideally we could consider how changes in observed $R_{rs}(\lambda)$ are linked to S_g variability. We identified two sets of stations with very similar $a_{phy}(\lambda)$, $a_d(\lambda)$ and $b_{bp}(\lambda)$ to compare how $a_g(\lambda)$ and S_g impact $R_{rs}(\lambda)$. For both sets of data, $a_g(\lambda)$ played the primary role in variability of $R_{rs}(\lambda)$ and this impact needs to be accounted for to consider the impact of S_g. We did this through the following expression:

$$R_{rs\,norm}(\lambda) = R_{rs}(\lambda) \cdot a_g(\lambda) \tag{4}$$

where $R_{rs\_norm}(\lambda)$ is considered as a normalized $R_{rs}(\lambda)$. We then calculated the difference between $R_{rs\_norm}(\lambda)$ ($\Delta R_{rs\_norm}(\lambda)$) for each site. At this point, units of $\Delta R_{rs\_norm}(\lambda)$ are $sr^{-1} \cdot m^{-1}$, precluding a direct comparison to PACE SNR (sr^{-1}). To account for this we multiplied $\Delta R_{rs\_norm}(\lambda)$ by 1 m^{-1}, resulting in $\Delta R_{rs\_norm}(\lambda)$ with units of sr^{-1} and theoretically representative of the impact of a change in S_g relative to a given magnitude of $a_g(\lambda)$. In effect, this produces a comparison of observed change in $R_{rs}(\lambda)$ due to S_g variability relative to PACE SNR.

2.3 Results

2.3.1 Field Data

Absorption parameters displayed several remarkably uniform patterns across sampling regions and years: $a_g(\lambda)$ and S_g displayed a clear inverse relationship across all wavelengths sampled (300-700 nm), suggesting a strong role for photodegradation within the system; $a_{ph}(\lambda)$ and $a_d(\lambda)$ displayed a positive covariation with $a_g(\lambda)$ suggesting that elevated CDOM, despite decreasing light availability, is present alongside positive growth conditions (increased nutrients, photoprotection) for phytoplankton or light-adapted phytoplankton communities (higher $cell^{-1}$ absorption rate). Some relatively low S_g values were observed in offshore waters during August and October. These S_g values were very close to S_g typically observed at depth, suggesting mixing of deep waters into the surface. Lake Superior is seasonally stratified, but stratification in the middle of the lake can be quite weak, particularly in summers following high ice, extended winter periods (e.g. summer of 2014; Austin and Colman 2007). NAP absorption co-varied with $a_{ph}(\lambda)$, suggesting that the bulk of $a_d(\lambda)$ in Lake Superior is sourced from phytoplankton-derived organics (Fig. 2); alternatively, the majority of $a_d(\lambda)$ could also derive from mineral particles that have displayed enhanced backscattering in this system (Peng et al. 2009), as $b_{bp}(\lambda)$ displayed a consistent decrease in magnitude with decreasing $a_d(\lambda)$.

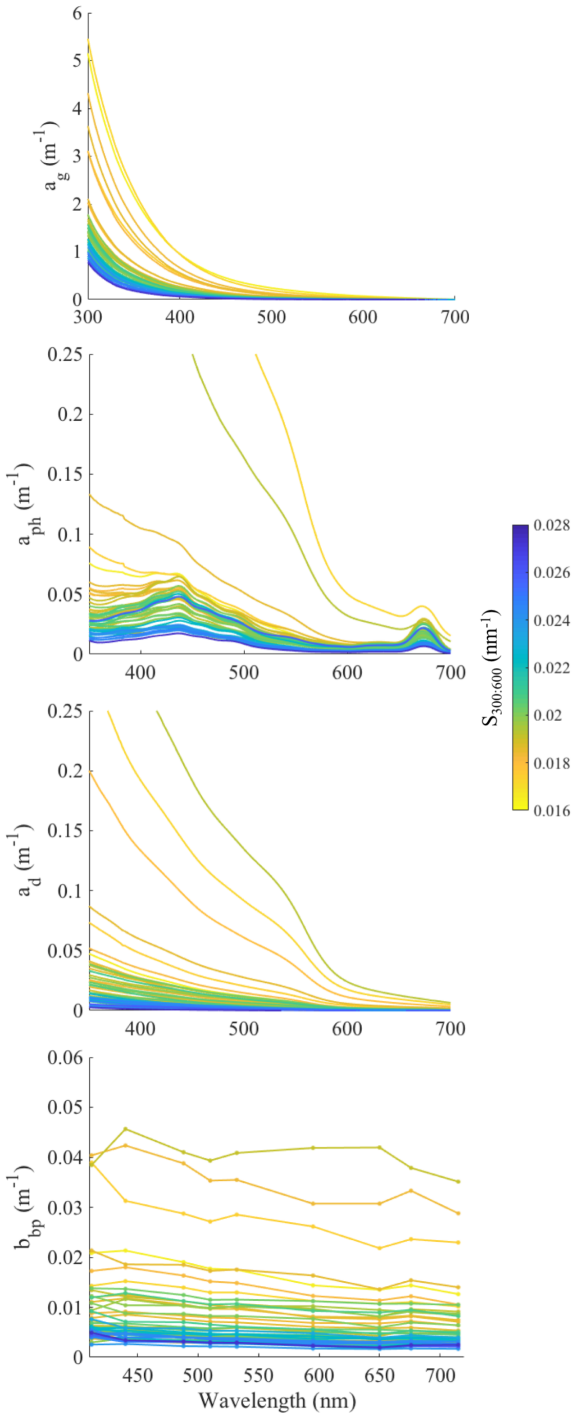


Figure 2. Observed (a) $a_g(\lambda)$, (b) $a_{ph}(\lambda)$, (c) $a_d(\lambda)$ and (d) $b_{bp}(\lambda)$ for all stations and years. Line color indicates $S_{300:600}$ observed at a given station.

$R_{rs}(\lambda)$ did not display a consistent relationship with $a_g(\lambda)$ or $S_{300:600}$, although this is somewhat expected (Fig. 3). While waters with higher magnitudes of $a_g(\lambda)$ are expected to absorb more light, resulting in a lower overall $R_{rs}(\lambda)$ signal, they also typically had higher $b_{bp}(\lambda)$ values resulting in more light reflected from the euphotic zone. Some stations were visibly turbid, but the majority of waters sampled were low turbidity with relatively deep Secchi depths (mean=13 m, n=20). The result was a dome-shaped $R_{rs}(\lambda)$ for all stations, with the rate of decrease at shorter wavelengths (380-490 nm) dictated by the magnitude of $a_g(\lambda)$. This dome-shape is consistent for oligotrophic, northern latitude lakes dominated by terrestrial $a_g(\lambda)$, as pure water absorption dominates absorption at red and infrared wavelengths and $a_g(\lambda)$ dominates absorption at UV wavelengths, while $a_{ph}(\lambda)$ is low. For waters where observed $a_g(\lambda)$ was high (e.g. $a_g(350) > 1\ m^{-1}$), $R_{rs}(\lambda)$ at wavelengths < 400 nm was often only an order of magnitude larger than PACE SNR for these wavelengths, displaying the difficulty in relying on UV wavelengths to estimate $a_g(\lambda)$ and S_g in waters dominated by this IOP.

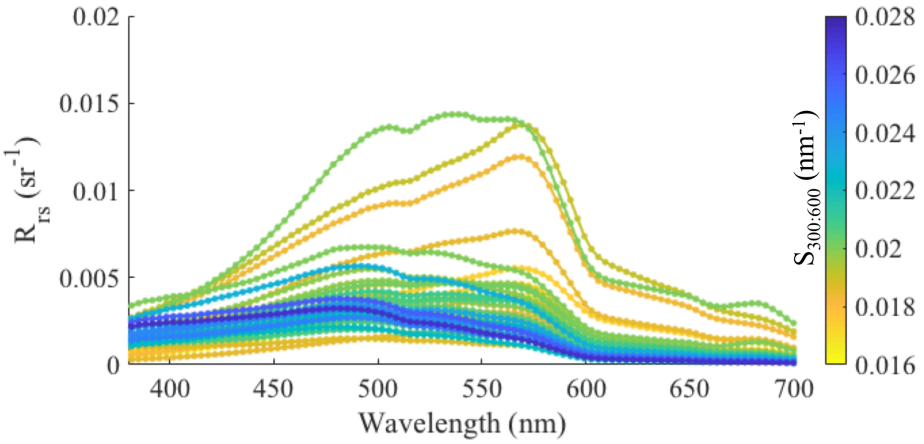


Figure 3. Observed $R_{rs}(\lambda)$ for all stations and years with line color indicating observed $S_{300:600}$.

2.3.2 HydroLight Simulations

2.3.2.1 Simulated $R_{rs}(\lambda)$

The main focus of this work is to consider how varying S_g impacts $R_{rs}(\lambda)$ and estimate the ability to detect changes in $R_{rs}(\lambda)$ explicitly due to changes in S_g. Thus, we did not pursue optical closure but did assess HydroLight performance relative to observed $R_{rs}(\lambda)$. Overall, simulated $R_{rs}(\lambda)$ displayed a similar distribution regarding magnitude and relationship with CDOM variability (a_g and S_g; Fig. 4). The absolute minimum and maximum simulated $R_{rs}(\lambda)$ fell very close to within observed values, suggesting that simulation results were reasonable (Fig. 4b). MAPE was predominantly <40% for wavelengths up to 650 nm (Fig. 4c). Calculated bias shows that simulated $R_{rs}(\lambda)$ was

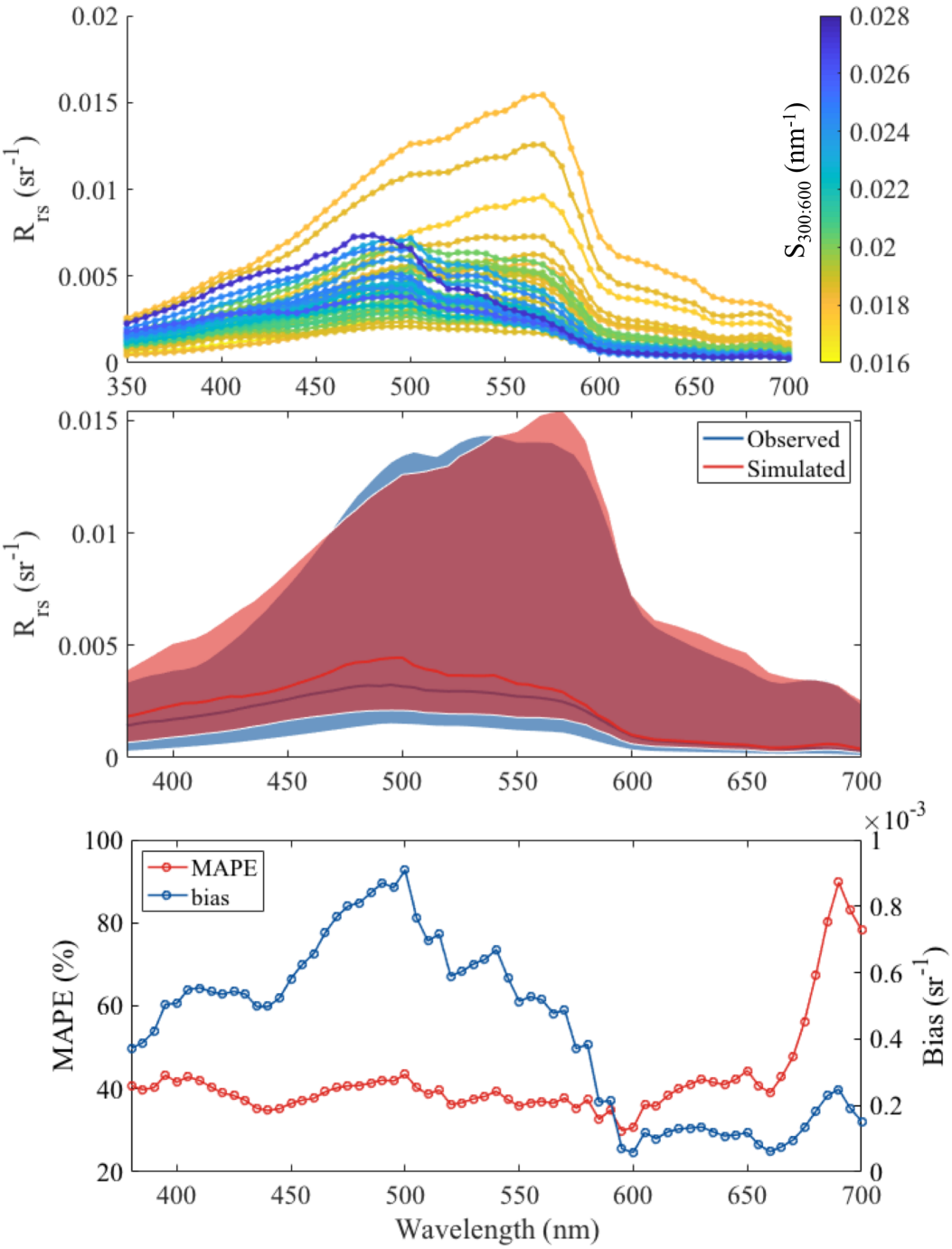


Figure 4. (a) HydroLight simulated $R_{rs}(\lambda)$ for all stations and years with line color indicating observed $S_{300:600}$. (b) Comparison of mean and range of observed and simulated $R_{rs}(\lambda)$. (c) Mean absolute percent error and bias for all simulated $R_{rs}(\lambda)$.

typically larger than observed values for a given spectra. Error tended to be an order of magnitude smaller than observed $R_{rs}(\lambda)$.

2.3.2.2 S_g Sensitivity

In order to more accurately assess the influence of variable S_g on changes in $R_{rs}(\lambda)$, we modeled changes in $a_g(\lambda)$ as an ensemble mean of $a_g(\lambda)$ modeled using λ_0 values of 350, 400, 450 and 500 nm (Fig. 5a). This results in changes in $a_g(\lambda)$ more representative of observed changes in $a_g(\lambda)$ affiliated with variability in molecular size and S_g on a single sample (Wunsch et al. 2018) while also avoiding the issue of a "pinch point" where all simulation results converge at the point of common λ_0 (Wei et al. 2016). The impact of S_g variability on $R_{rs}(\lambda)$ for the example spectra in Fig. 5a is presented (Fig. 5b) and displays the largest magnitude of change from ~450-600 nm, coincident with the largest magnitude of $R_{rs}(\lambda)$.

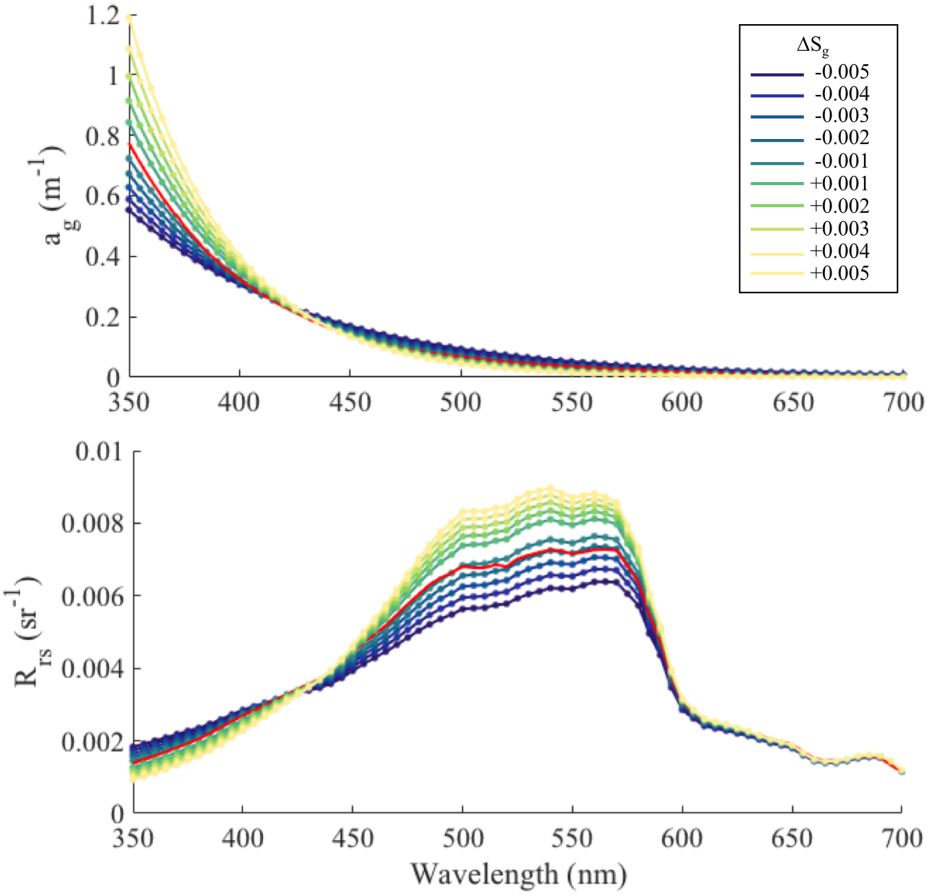


Figure 5. Examples of altered (a) $a_g(\lambda)$ relative to the observed $a_g(\lambda)$ (red line) and (b) resulting $R_{rs}(\lambda)$ for the respective change in S_g and altered $a_g(\lambda)$.

We considered the influence of changes in S_g affiliated with the relative variability in S_g observed in the system (Fig. 6). Observed S_g values were altered by ±0.005 nm^{-1} in 0.001 nm^{-1} increments, with the minimum and maximum change in $R_{rs}(\lambda)$ (ΔR_{rs}) illustrated alongside median ΔR_{rs} relative to PACE SNR. We considered variability in S_g that corresponded to ΔR_{rs} of an order of magnitude larger than PACE SNR to be a readily viewable spectral change and consider "consistent viewability" as a spectral range of at least 50 nm, consistent with spectral ranges used in UV and visible wavelengths in the literature. Minimum ΔR_{rs} did not consistently pass this threshold until absolute changes in S_g were greater than 0.003 nm^{-1}, while median ΔR_{rs} was consistently an order of magnitude greater than PACE SNR for all changes in S_g considered. The spectral range of ~400-550 nm most consistently saw observable changes in S_g for this system.

2.3.3 Observed S_g Variability

HydroLight provides a unique ability to isolate changes in $R_{rs}(\lambda)$ due to specific IOP's or spectral variability, as evidenced here. Changes in S_g of 0.001 nm^{-1} are observable for many optical conditions sampled in Lake Superior; we sought to consider how relatable simulated results were to *in situ* variability and observed changes in $R_{rs}(\lambda)$ across stations where $a_g(\lambda)$ represented the primary optical variability between locations. S_g presents approximately 0.01 nm^{-1} variability regardless of spectral range used (350-700 nm, as used for HydroLight simulations, and 300-600 nm, as used to characterize *in situ* spectral variability). The western arm of Lake Superior near a large Lake Superior tributary, the St. Louis River, displayed relatively limited $a_g(\lambda)$ spectral variability across all sampled sites while also presenting cases of clear to very turbid water (Fig. 7). We considered stations where $a_{ph}(\lambda)$ and $a_d(\lambda)$ magnitude were similar (variability was at least one order of magnitude smaller than $a_g(\lambda)$ variability) and $b_{bp}(535)$ was the same or very similar (Fig. 8). In effect, this isolates the first order variability in $R_{rs}(\lambda)$ to changes in $a_g(\lambda)$ and S_g. We then compared ΔR_{rs_norm} between stations to isolate the impact of observed S_g variability on $R_{rs}(\lambda)$ (Fig. 8a-v,b-v) and compare this change relative to PACE SNR (Fig. 8a-vi,b-vi).

2.4 Discussion

The utility of optics-based metrics of CDOM composition is the low cost and speed of measurement and the potential for application across autonomous and satellite platforms. Historically, linking optical variability in a_g spectra to DOM composition has been limited by available tools. Recent advances in characterization, including various size fractionation techniques and mass spectrometry analysis alongside fluorescence and absorption characterization of CDOM has provided significant advances in our understanding of specific optics-based metrics relevant to understanding DOM composition, source, degradation state and molecular weight (Helms et al. 2008; Kellerman et al. 2018; Kellerman et al. 2015; Wünsch et al. 2018). Alongside these advances has been a more thorough understanding of the role of specific ecosystems and trophic levels (e.g. phytoplankton, bacteria) to optical signatures, including under various nutrient conditions

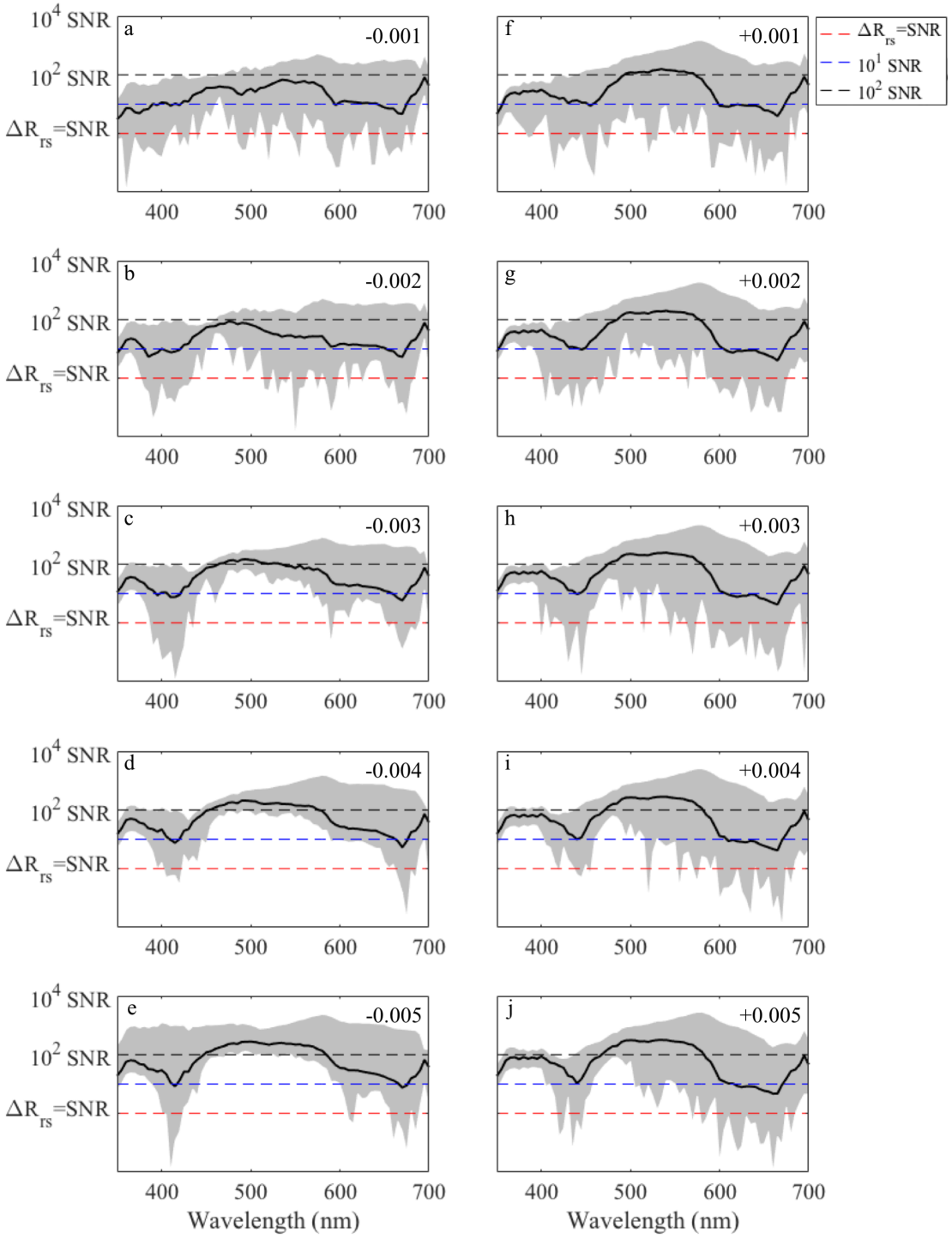


Figure 6. Median and range of ΔR_{rs} for all simulated $R_{rs}(\lambda)$ for each change in S_g.

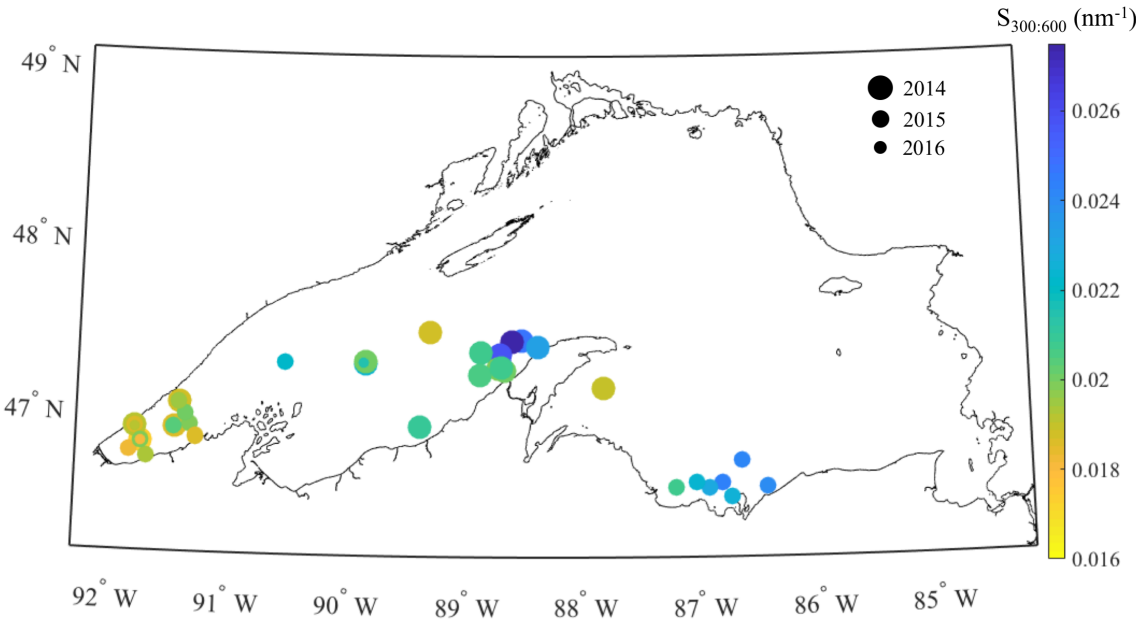


Figure 7. Distribution of $S_{300:600}$ for all sampled stations with marker size indicating year and marker color indicating $S_{300:600}$ values.

and degradation state (Asmala et al. 2014; Danhiez et al. 2017; Asmala et al. 2018; Kinsey et al. 2018). These advances provide tremendous insight on short timescales but lack the synoptic information available from satellite sensors. However, beyond the initial work of Wei et al. (2016), the ability of current and future hyperspectral sensors to detect changes in $R_{rs}(\lambda)$ explicitly due to S_g variability has not been assessed.

Here, we show that for an environment dominated by CDOM as the primary IOP, changes in S_g are a detectable feature in $R_{rs}(\lambda)$. We considered the sensitivity of $R_{rs}(\lambda)$ to both $a_g(\lambda)$ magnitude and S_g variability. From this, it appears likely that biogeochemically significant changes in S_g (~0.001 nm^{-1}) are an observable feature for many water conditions, a finding supported by analysis of *in situ* data (Fig. 9). From this, a few important points emerge: 1) Available wavelengths for detecting changes due to S_g typically occur at visible wavelengths (400-550 nm), a spectral range that has not been well considered for biogeochemical significance. Spectral variability from 350-400 nm has been previously linked to changes in molecular weight and degradation pathways (Helms et al. 2008) but has also been shown to correlate poorly with significant changes in DOM composition at the molecular level (Kellerman et al. 2018). Rates of change at various CDOM wavelengths directly impact observed S_g and are the origin of inconsistency between S_g calculated over different wavelength ranges (e.g. Twardowski et al. 2004). These changes are reflective of production and degradation of different chromophores, and appear to be linked to specific environmental conditions, phytoplankton bloom stages and degradation history (Asmala et al. 2018; Asmala et al. 2014; Danhiez et al. 2017; Yamashita et al. 2013). Much of this variability has been attributed to the contribution of humic or humic-like CDOM by bacteria (Kinsey et al. 2018; Romera-Castillo et al. 2011).

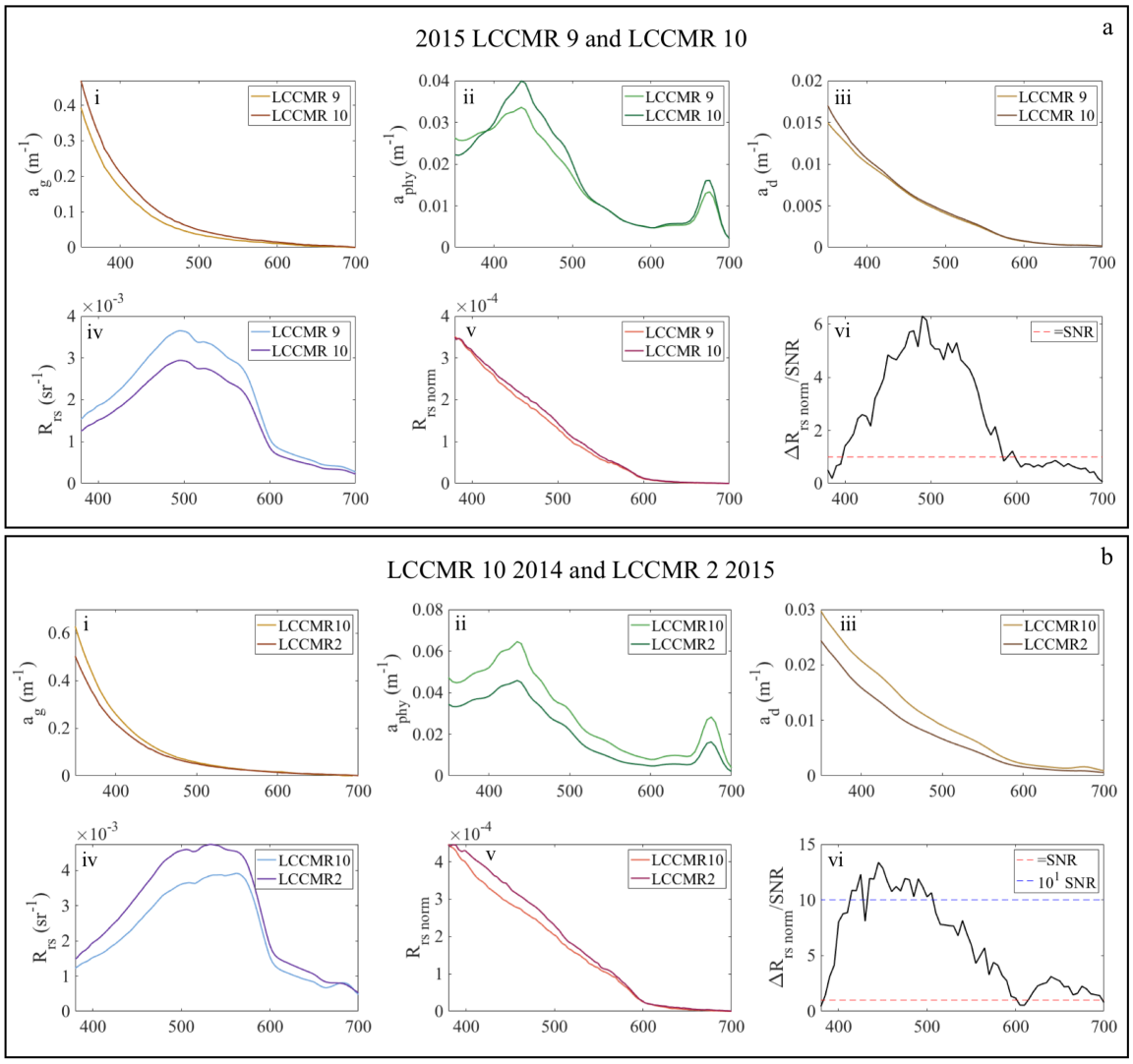


Figure 8. Example (a) $a_g(\lambda)$, (b) $a_{ph}(\lambda)$, (c) $a_d(\lambda)$, (d) $R_{rs}(\lambda)$, (e) ΔR_{rs} and (f) ΔR_{rs}/SNR between optically similar stations.

The molecular composition of phytoplankton exudates has also been observed to vary based on phytoplankton composition, physiology and nutrient stress (Asmala et al. 2018; Myklestad 1995), with observable changes in a_g and S_g affiliated with these changes (Danhiez et al. 2017; Organelli et al. 2014; Romera-Castillo et al. 2011). These optical changes are directly affiliated with changes in oxidation and reduction potential of DOM that are a function of background seawater chemistry, degradation state and interaction of newly produced DOM with the background DOM pool (i.e. priming effect; Aeschbacher et al. 2011; Boyd and Osburn 2004; Helms et al. 2014; Helms et al. 2013; Kujawinski et al. 2016; Mentges et al. 2017).

It seems difficult to constrain an immensely complicated matrix of organic compounds subject to varying forcing – terrestrial and marine signal; microbial and phytoplankton signal; environmental controls – with a single, relatively simple parameter such as S_g. However, the consistent relationship between optical metrics including S_g and DOM compositional variability across environments suggests that specific groups of molecules that behave in an optically similar fashion carry information about DOM composition, and these optical signatures can in turn be characterized (Catalá et al. 2016; Danhiez et al. 2017; Kellerman et al. 2018; Zhao et al. 2017). Indeed, spatial variability in DOM composition from the coastal to open and surface to deep ocean is consistent with optical considerations that view an increasingly humic, microbial-derived DOM pool (Catalá et al. 2015; Kaiser and Benner 2009; Mentges et al. 2017; Yamashita and Tanoue 2008). This opens a door for tracking the aquatic carbon cycle on global scales via ocean color data, provided molecular characterization of DOM alongside optical characterization continues *in situ*. The primary limitations, beyond launch of PACE and similar hyperspectral sensors (e.g. similar spatial and temporal resolution, SNR), are: 1) algorithms capable of estimating S_g in an unbiased manner and 2) relationships of UV-vis S_g metrics with DOM composition. To date, the majority of work has related $S_{275:295}$ to trends in DOM composition with promising results (Helms et al. 2008; Kellerman et al. 2018). Algorithms exist to relate $a_g(\lambda)$ to $S_{275:295}$ but are dependent on strong gradients in $a_g(\lambda)$ that typically occur across transects of predominantly terrestrial to marine CDOM or where photodegradation is the primary pathway of change (Fichot et al. 2013; Mannino et al. 2014). Alternately, algorithms assume a consistent relationship between variability between $a_g(\lambda)$ at visible wavelengths and S_g, including $S_{275:295}$ (Cao and Miller 2015; Matsuoka et al. 2013). While this approach works for some systems, analysis of rates of change in $a_g(\lambda)$ show that production and degradation vary at different wavelengths and under different environmental conditions, with visible wavelengths particularly insensitive to spectral changes in the UV at times (Asmala et al. 2014; Danhiez et al. 2017; Asmala et al. 2018). This suggests further consideration of directly observable S_g (i.e. 350-600 nm spectral range) and the relationship with changes in DOM composition. Also, consideration of alternate metrics capable of determining environmental conditions to assess the applicability of a specific algorithm could be useful; for example, one algorithm can accurately retrieve $S_{275:295}$ under the exponential growth phase of a phytoplankton bloom, but another algorithm is required for the decay phase.

Lake Superior is a unique environment to study CDOM dynamics as its size provides for significant production and processing of DOM but the large, predominantly forested watershed draining into the lake delivers a readily observed terrestrial signal to the lake. Lake Superior displays significant optical variability that is regionally and seasonally consistent (Trochta et al. 2015) with this optical variability largely driven by changes in $a_g(\lambda)$ (Mouw et al. 2013). Sampling of this system is largely limited to ice-free months during fair weather, leaving a large data gap that is well-addressed from satellite platforms (Mouw et al. 2013). Past success in retrieving optical variability and $a_g(\lambda)$ from multispectral platforms suggests that satellite-estimated $a_g(\lambda)$ can estimate changes to terrestrial systems draining into the lake as well as autochthonous processes manifested by

$a_g(\lambda)$ optical variability. Considering rapid warming of Lake Superior, understanding spectral variability of CDOM provides a unique tool to track climate-driven changes to the Lake Superior watershed and Lake Superior proper and how these changes drive variability in aquatic carbon cycling in this system (Austin and Colman 2007; Creed et al. 2018; Mouw et al. 2013; Williamson et al. 2015).

Carbon cycling into and within Lake Superior and CDOM optical properties for these systems have been considered and display interesting trends. DOM delivered to Lake Superior across a range of tributaries display similar optical properties including spectral slope and magnitude of $a_g(\lambda)$, regardless of tributary size and reach, as well as consistent biodegradation rates (Coble et al. 2016b; unpublished data). Many large Lake Superior tributaries are quite turbid (e.g. Sturgeon River), while smaller tributaries often flow through primarily forested watersheds, limiting within-stream photodegradation (Coble et al. 2016a). Additionally, nitrate concentrations in these watersheds are very low and microbial degradation has been shown to be enhanced by introducing nitrogen sources (Coble et al. 2015). This suggests that microbial degradation of DOM within streams is nutrient limited. Considering this and the extremely consistent trend between spectral slope and magnitude of $a_g(\lambda)$ across sampling sites and years, it appears the vast majority of CDOM in Lake Superior is delivered from terrestrial systems and progressively photodegraded (Fig. 2; Ma and Green 2004). DOM is a major component of carbon cycling in Lake Superior with photodegradation playing a large role in its bioavailability (Biddanda and Cotner 2003). Samples considered here focused on the sunlit surface waters where there is likely a tight feedback between photodegradation and microbial utilization of DOM in this large, oligotrophic lake (Biddanda et al. 2001). Coastal and open ocean DOM cycling can be quite complex; while we do not seek to reduce the complexity of Lake Superior DOM cycling, it appears that there are relatively clear trends in delivered material and its processing within the lake. If true, Lake Superior presents an ideal environment to observe climate-driven changes in CDOM cycling from a future hyperspectral sensor.

2.5 Conclusions

Characterizing CDOM spectral variability allows for rapid and effective estimation of aquatic DOM cycling. Hyperspectral satellite ocean color data provides a means of estimating this variability over broad spatial and temporal scales, provided S_g variability can be observed and estimated. We show that for Lake Superior and other systems where $a_g(\lambda)$ dominates the absorption budget, $R_{rs}(\lambda)$ variability solely due to changes in S_g is an observable feature, particularly at wavelengths from 400-550 nm. Sensitivity to S_g can be estimated to 0.001 nm^{-1} for many observed waters, suggesting that biogeochemical variability characterized by S_g can be estimated from PACE.

Much of the discussion considered here is very forward thinking, particularly in the context of current multispectral sensors. With hyperspectral sensors in orbit (e.g. Hyperspectral Imager for the Coastal Ocean (HICO), Scanning Imaging Absorption Spectrometer for Atmospheric Chartography (SCIAMACHY)) and slated for orbit (e.g.

PACE) in the next ~5 years, we suggest that what heritage sensors did for detecting chlorophyll, hyperspectral sensors are capable of doing for the aquatic carbon cycle and specifically DOM cycling. We show here that the SNR of PACE provides the ability to detect changes in $R_{rs}(\lambda)$ explicitly related to variability in S_g. In Lake Superior, changes in S_g were typically associated with changes in $a_g(\lambda)$, where $a_g(\lambda)$ variability accounts for the majority of change in $R_{rs}(\lambda)$ magnitude and, to a lesser extent, $R_{rs}(\lambda)$ spectral shape. However, variability in the global ocean of $a_g(\lambda)$ and S_g are often de-coupled. For these environments, accurate detection of S_g is crucial for monitoring changes in CDOM composition. In systems where $a_g(\lambda)$ and S_g variability are tightly coupled, such as Lake Superior, these variables can be seen as complimentary in nature with added value in accurate retrievals of S_g.

2.6 References

Aeschbacher, M., Vergari, D., Schwarzenbach, R.P., & Sander, M. (2011). Electrochemical analysis of proton and electron transfer equilibria of the reducible moieties in humic acids. *Environ Sci Technol, 45*, 8385-8394

Asmala, E., Autio, R., Kaartokallio, H., Stedmon, C.A., & Thomas, D.N. (2014). Processing of humic-rich riverine dissolved organic matter by estuarine bacteria: effects of predegradation and inorganic nutrients. *Aquatic Sciences, 76*, 451-463

Asmala, E., Haraguchi, L., Jakobsen, H.H., Massicotte, P., & Carstensen, J. (2018). Nutrient availability as major driver of phytoplankton-derived dissolved organic matter transformation in coastal environment. *Biogeochemistry, 137*, 93-104

Austin, J.A., & Colman, S.M. (2007). Lake Superior summer water temperatures are increasing more rapidly than regional air temperatures: A positive ice-albedo feedback. *Geophysical Research Letters, 34*

Babin, M., Stramski, D., Ferrari, G.M., Claustre, H., Bricaud, A., Obolensky, G., & Hoepffner, N. (2003). Variations in the light absorption coefficients of phytoplankton, nonalgal particles, and dissolved organic matter in coastal waters around Europe. *Journal of Geophysical Research, 108*

Barbiero, R.P., & Tuchman, M.L. (2004). The deep chlorophyll maximum in Lake Superior. *Journal of Great Lakes Research, 30 (Supplement 1)*, 256-268

Behrenfeld, M., & Falkowski, P.G. (1997). Photosynthetic rates derived from satellite-based chlorophyll concentration. *Limnology & Oceanography, 42*, 1-20

Biddanda, B.A., & Cotner, J.B. (2003). Enhancement of dissolved organic matter bioavailability by sunlight and its role in the carbon cycle of Lakes Superior and Michigan. *Journal of Great Lakes Research, 29*, 228-241

Biddanda, B.A., Ogdahl, M., & Cotner, J.B. (2001). Dominance of bacterial metabolism in oligotrophic relative to eutrophic waters. *Limnology and Oceanography, 46*, 730-739

Boyd, T.J., & Osburn, C.L. (2004). Changes in CDOM fluorescence from allochthonous and autochthonous sources during tidal mixing and bacterial degradation in two coastal estuaries. *Marine Chemistry, 89*, 189-210

Bracher, A., Bouman, H.A., Brewin, R.J.W., Bricaud, A., Brotas, V., Ciotti, A.M., Clementson, L., Devred, E., Di Cicco, A., Dutkiewicz, S., Hardman-Mountford, N.J., Hickman, A.E., Hieronymi, M., Hirata, T., Losa, S.N., Mouw, C.B., Organelli, E., Raitsos, D.E., Uitz, J., Vogt, M., & Wolanin, A. (2017). Obtaining Phytoplankton Diversity from Ocean Color: A Scientific Roadmap for Future Development. *Frontiers in Marine Science, 4*

Cao, F., & Miller, W.L. (2015). A new algorithm to retrieve chromophoric dissolved organic matter (CDOM) absorption spectra in the UV from ocean color. *Journal of Geophysical Research: Oceans, 120*, 496-516

Cao, F., Tzortziou, M., Hu, C., Mannino, A., Fichot, C.G., Del Vecchio, R., Najjar, R.G., & Novak, M. (2018). Remote sensing retrievals of colored dissolved organic matter and dissolved organic carbon dynamics in North American estuaries and their margins. *Remote Sensing of Environment, 205*, 151-165

Carreto, J.I., Carignan, M.O., Daleo, G., & Marco, S.G.D. (1990). Occurrence of mycosporine-like amino acids in the red-tide dinoflagellate Alexandrium excavatum: UV-photoprotective compounds? *Journal of Plankton Research, 12*, 909-921

Catalá, T.S., Reche, I., Fuentes-Lema, A., Romera-Castillo, C., Nieto-Cid, M., Ortega-Retuerta, E., Calvo, E., Alvarez, M., Marrasé, C., & Stedmon, C.A. (2015). Turnover time of fluorescent dissolved organic matter in the dark global ocean. *Nature communications, 6*, 5986

Catalá, T.S., Reche, I., Ramón, C.L., López-Sanz, À., Álvarez, M., Calvo, E., & Álvarez-Salgado, X.A. (2016). Chromophoric signatures of microbial by-products in the dark ocean. *Geophysical Research Letters, 43*, 7639-7648

Coble, A.A., Marcarelli, A.M., & Kane, E.S. (2015). Ammonium and glucose amendments stimulate dissolved organic matter mineralization in a Lake Superior tributary. *Journal of Great Lakes Research, 41*, 801-807

Coble, A.A., Marcarelli, A.M., Kane, E.S., & Huckins, C.J. (2016a). Uptake of ammonium and soluble reactive phosphorus in forested streams: influence of dissolved organic matter composition. *Biogeochemistry, 131*, 355-372

Coble, A.A., Marcarelli, A.M., Kane, E.S., Toczydlowski, D., & Stottlemyer, R. (2016b). Temporal patterns of dissolved organic matter biodegradability are similar across three rivers of varying size. *Journal of Geophysical Research: Biogeosciences, 121*, 1617-1631

Cotner, J.B., Biddanda, B.A., Makino, W., & Stets, E. (2004). Organic carbon biogeochemistry of Lake Superior. *Aquatic Ecosystem Health & Management, 7*, 451-464

Creed, I.F., Bergstrom, A.K., Trick, C.G., Grimm, N.B., Hessen, D.O., Karlsson, J., Kidd, K.A., Kritzberg, E., McKnight, D.M., Freeman, E.C., Senar, O.E., Andersson, A., Ask, J., Berggren, M., Cherif, M., Giesler, R., Hotchkiss, E.R., Kortelainen, P., Palta, M.M., Vrede, T., & Weyhenmeyer, G.A. (2018). Global change-driven effects on dissolved organic matter composition: Implications for food webs of northern lakes. *Glob Chang Biol*

Danhiez, F., Vantrepotte, V., Cauvin, A., Lebourg, E., & Loisel, H. (2017). Optical properties of chromophoric dissolved organic matter during a phytoplankton bloom. Implication for DOC estimates from CDOM absorption. *Limnology and Oceanography, 62*, 1409-1425

Defoin-Platel, M., & Chami, M. (2007). How ambiguous is the inverse problem of ocean color in coastal waters? *Journal of Geophysical Research, 112*

Dong, Q., Shang, S., & Lee, Z. (2013). An algorithm to retrieve absorption coefficient of chromophoric dissolved organic matter from ocean color. *Remote Sensing of Environment, 128*, 259-267

Fichot, C.G., Kaiser, K., Hooker, S.B., Amon, R.M., Babin, M., Belanger, S., Walker, S.A., & Benner, R. (2013). Pan-Arctic distributions of continental runoff in the Arctic Ocean. *Sci Rep, 3*, 1053

Friedland, K.D., Stock, C., Drinkwater, K.F., Link, J.S., Leaf, R.T., Shank, B.V., Rose, J.M., Pilskaln, C.H., & Fogarty, M.J. (2012). Pathways between primary production and fisheries yields of large marine ecosystems. *PLoS One, 7*, e28945

Helms, J.R., Mao, J., Stubbins, A., Schmidt-Rohr, K., Spencer, R.G.M., Hernes, P.J., & Mopper, K. (2014). Loss of optical and molecular indicators of terrigenous dissolved organic matter during long-term photobleaching. *Aquatic Sciences, 76*, 353-373

Helms, J.R., Stubbins, A., Perdue, E.M., Green, N.W., Chen, H., & Mopper, K. (2013). Photochemical bleaching of oceanic dissolved organic matter and its effect on absorption spectral slope and fluorescence. *Marine Chemistry, 155*, 81-91

Helms, J.R., Stubbins, A., Ritchie, J.D., Minor, E., Kieber, D.J., & Mopper, K. (2008). Absorption spectral slopes and slope ratios as indicators of molecular weight, source, and photobleaching of chromophoric dissolved organic matter. *Limnology and Oceanography, 53*, 955-969

Hiriart-Baer, V.P., Milne, J.E., & Marvin, C.H. (2011). Temporal trends in phosphorus and lacustrine productivity in Lake Simcoe inferred from lake sediment. *Journal of Great Lakes Research, 37*, 764-771

Kaiser, K., & Benner, R. (2009). Biochemical composition and size distribution of organic matter at the Pacific and Atlantic time-series stations. *Marine Chemistry, 113*, 63-77

Kellerman, A.M., Guillemette, F., Podgorski, D.C., Aiken, G.R., Butler, K.D., & Spencer, R.G.M. (2018). Unifying Concepts Linking Dissolved Organic Matter Composition to Persistence in Aquatic Ecosystems. *Environ Sci Technol, 52*, 2538-2548

Kellerman, A.M., Kothawala, D.N., Dittmar, T., & Tranvik, L.J. (2015). Persistence of dissolved organic matter in lakes related to its molecular characteristics. *Nature Geoscience, 8*, 454-457

Kinsey, J.D., Corradino, G., Ziervogel, K., Schnetzer, A., & Osburn, C.L. (2018). Formation of Chromophoric Dissolved Organic Matter by Bacterial Degradation of Phytoplankton-Derived Aggregates. *Frontiers in Marine Science, 4*

Kujawinski, E.B., Longnecker, K., Barott, K.L., Weber, R.J.M., & Kido Soule, M.C. (2016). Microbial Community Structure Affects Marine Dissolved Organic Matter Composition. *Frontiers in Marine Science, 3*

Lee, Z., Carder, K.L., & Arnone, R.A. (2002). Deriving inherent optical properties from water color: a multiband quasi-analytical algorithm for optically deep waters. *Appl Opt, 41*, 5755-5772

Lohrenz, S. (2000). A novel theoretical approach to correct for pathlength amplification and variable sampling loading in measurements of particulate spectral absorption by the quantitative filter technique. *Journal of Plankton Research, 22*, 639-657

Lohrenz, S., Weidemann, A., & Tuel, M. (2003). Phytoplankton spectral absorption influenced by community size structure and pigment composition. *Journal of Plankton Research, 25*, 35-61

Ma, X., & Green, S.A. (2004). Photochemical transformation of dissolved organic carbon in Lake Superior - an in-situ experiment. *Journal of Great Lakes Research, 30 (Supplement 1)*, 97-112

Macdonald, M.J., & Minor, E.C. (2013). Photochemical degradation of dissolved organic matter from streams in the western Lake Superior watershed. *Aquatic Sciences, 75*, 509-522

Mannino, A., Novak, M.G., Hooker, S.B., Hyde, K., & Aurin, D. (2014). Algorithm development and validation of CDOM properties for estuarine and continental shelf waters along the northeastern U.S. coast. *Remote Sensing of Environment, 152*, 576-602

Mannino, A., Russ, M.E., & Hooker, S.B. (2008). Algorithm development and validation for satellite-derived distributions of DOC and CDOM in the U.S. Middle Atlantic Bight. *Journal of Geophysical Research, 113*

Maritorena, S., Siegel, D.A., & Peterson, A.R. (2002). Optimization of a semianalytical ocean color model for global-scale applications. *Appl Opt, 41*, 2705-2714

Matsuoka, A., Babin, M., Doxaran, D., Hooker, S.B., Mitchell, B.G., Bélanger, S., & Bricaud, A. (2014). A synthesis of light absorption properties of the Arctic Ocean: application to semianalytical estimates of dissolved organic carbon concentrations from space. *Biogeosciences, 11*, 3131-3147

Matsuoka, A., Hooker, S.B., Bricaud, A., Gentili, B., & Babin, M. (2013). Estimating absorption coefficients of colored dissolved organic matter (CDOM) using a semi-analytical algorithm for southern Beaufort Sea waters: application to deriving concentrations of dissolved organic carbon from space. *Biogeosciences, 10*, 917-927

McManus, J., Heinen, E.A., & Baehr, M.M. (2003). Hypolimnetic oxidation rates in Lake Superior: Role of dissolved organic material on the lake's carbon budget. *Limnology and Oceanography, 48*, 1624-1632

Mentges, A., Feenders, C., Seibt, M., Blasius, B., & Dittmar, T. (2017). Functional Molecular Diversity of Marine Dissolved Organic Matter Is Reduced during Degradation. *Frontiers in Marine Science, 4*, 194

Minor, E.C., Steinbring, C.J., Longnecker, K., & Kujawinski, E.B. (2012). Characterization of dissolved organic matter in Lake Superior and its watershed using ultrahigh resolution mass spectrometry. *Organic Geochemistry, 43*, 1-11

Mobley, C. (1988). HydroLight. In

Mouw, C.B., Barnett, A., McKinley, G.A., Gloege, L., & Pilcher, D. (2016). Phytoplankton size impact on export flux in the global ocean. *Global Biogeochemical Cycles, 30*, 1542-1562

Mouw, C.B., Chen, H., McKinley, G.A., Effler, S., O'Donnell, D., Perkins, M.G., & Strait, C. (2013). Evaluation and optimization of bio-optical inversion algorithms for remote sensing of Lake Superior's optical properties. *Journal of Geophysical Research: Oceans, 118*, 1696-1714

Mouw, C.B., Ciochetto, A.B., Grunert, B.K., & Yu, A.W. (2017a). Expanding understanding of optical variability in Lake Superior with a 4-year dataset. *Earth System Science Data, 9*, 497-509

Mouw, C.B., Hardman-Mountford, N.J., Alvain, S., Bracher, A., Brewin, R.J.W., Bricaud, A., Ciotti, A.M., Devred, E., Fujiwara, A., Hirata, T., Hirawake, T., Kostadinov, T.S., Roy, S., & Uitz, J. (2017b). A Consumer's Guide to Satellite Remote Sensing of Multiple Phytoplankton Groups in the Global Ocean. *Frontiers in Marine Science, 4*

Mouw, C.B., & Yoder, J.A. (2010). Optical determination of phytoplankton size composition from global SeaWiFS imagery. *Journal of Geophysical Research, 115*

Mueller, J.L., Fargion, G.S., & McClain, C.R. (2003). Ocean optics protocols for satellite ocean color sensor validation. In, *revision 4, vol. IV: Inherent Optical Properties: Instruments, characterizations, field measurements and data analysis protocols, NASA/TM-2003-211621/Rev4-Vol. IV* (p. 76pp). National Aeronautical and Space Administration, Goddard Space Flight Center, Greenbelt, Maryland

Myklestad, S.M. (1995). Release of extracellular products by phytoplankton with special emphasis on polysaccharides. *Science of the total Environment, 165*, 155-164

Nelson, N.B., Siegel, D.A., Carlson, C.A., & Swan, C.M. (2010). Tracing global biogeochemical cycles and meridional overturning circulation using chromophoric dissolved organic matter. *Geophysical Research Letters, 37*

Organelli, E., Bricaud, A., Antoine, D., & Matsuoka, A. (2014). Seasonal dynamics of light absorption by chromophoric dissolved organic matter (CDOM) in the NW Mediterranean Sea (BOUSSOLE site). *Deep Sea Research Part I: Oceanographic Research Papers, 91*, 72-85

Organelli, E., Bricaud, A., Gentili, B., Antoine, D., & Vellucci, V. (2016). Retrieval of Colored Detrital Matter (CDM) light absorption coefficients in the Mediterranean Sea using field and satellite ocean color radiometry: Evaluation of bio-optical inversion models. *Remote Sensing of Environment, 186*, 297-310

Romera-Castillo, C., Sarmento, H., Alvarez-Salgado, X.A., Gasol, J.M., & Marrase, C. (2011). Net production and consumption of fluorescent colored dissolved organic matter by natural bacterial assemblages growing on marine phytoplankton exudates. *Appl Environ Microbiol, 77*, 7490-7498

Siegel, D.A., Maritorena, S., Nelson, N.B., Hansell, D.A., & Lorenzi-Kayser, M. (2002). Global distribution and dynamics of colored dissolved and detrital organic materials. *Journal of Geophysical Research: Oceans, 107*, 21-21-21-14

Silsbe, G.M., Behrenfeld, M.J., Halsey, K.H., Milligan, A.J., & Westberry, T.K. (2016). The CAFE model: A net production model for global ocean phytoplankton. *Global Biogeochemical Cycles, 30*, 1756-1777

Stedmon, C.A., & Nelson, N.B. (2014). The optical properties of DOM in the ocean. *Biogeochemistry of Marine Dissolved Organic Matter (Second Edition)* (pp. 481-508): Elsevier

Sterner, R.W. (2010). In situ-measured primary production in Lake Superior. *Journal of Great Lakes Research, 36*, 139-149

Stumpf, R.P., Culver, M.E., Tester, P.A., Tomlinson, M., Kirkpatrick, G.J., Pederson, B.A., Truby, E., Ransibrahmanakul, V., & Soracco, M. (2003). Monitoring Karenia brevis blooms in the Gulf of Mexico using satellite ocean color imagery and other data. *Harmful Algae, 2*, 147-160

Sydor, M., Gould, R.W., Arnone, R., Haltrin, V.I., & Goode, W. (2004). Uniqueness in remote sensing of the inherent optical properties of ocean water. *Appl Opt, 43*, 2156-2162

Tassan, S., & Ferrari, G.M. (2002). A sensitivity analysis of the "Transmittance-Reflectance" method for measuring light absorption by aquatic particles. *Journal of Plankton Research, 24*, 757-774

Trochta, J.T., Mouw, C.B., & Moore, T.S. (2015). Remote sensing of physical cycles in Lake Superior using a spatio-temporal analysis of optical water typologies. *Remote Sensing of Environment, 171*, 149-161

Twardowski, M.S., Boss, E., Sullivan, J.M., & Donaghay, P.L. (2004). Modeling the spectral shape of absorption by chromophoric dissolved organic matter. *Marine Chemistry, 89*, 69-88

Uitz, J., Stramski, D., Reynolds, R.A., & Dubranna, J. (2015). Assessing phytoplankton community composition from hyperspectral measurements of phytoplankton absorption coefficient and remote-sensing reflectance in open-ocean environments. *Remote Sensing of Environment, 171*, 58-74

Urban, N.R., Auer, M.T., Green, S.A., Lu, X., Apul, D.S., Powell, K.D., & Bub, L. (2005). Carbon cycling in Lake Superior. *Journal of Geophysical Research, 110*

Vandermeulen, R.A., Mannino, A., Neeley, A., Werdell, J., & Arnone, R. (2017). Determining the optimal spectral sampling frequency and uncertainty thresholds for hyperspectral remote sensing of ocean color. *Opt Express, 25*, A785-A797

Vantrepotte, V., Danhiez, F.P., Loisel, H., Ouillon, S., Meriaux, X., Cauvin, A., & Dessailly, D. (2015). CDOM-DOC relationship in contrasted coastal waters: implication for DOC retrieval from ocean color remote sensing observation. *Opt Express, 23*, 33-54

Wei, J., Lee, Z., Ondrusek, M., Mannino, A., Tzortziou, M., & Armstrong, R. (2016). Spectral slopes of the absorption coefficient of colored dissolved and detrital material inverted from UV-visible remote sensing reflectance. *Journal of Geophysical Research: Oceans*, n/a-n/a

Werdell, J., Franz, B.A., Bailey, S.W., Feldman, G.C., Boss, E., Brando, V., Dowell, M., Hirata, T., Lavender, S.J., Lee, Z., Loisel, H., Maritorena, S., Mélin, F., Moore, T.S., Smyth, T.J., Antoine, D., Devred, E., D'Andon, O.H.F., & Mangin, A. (2013). Generalized ocean color inversion model for retrieving marine inherent optical properties. *Appl Opt, 52*, 2019-2037

Williamson, C.E., Overholt, E.P., Pilla, R.M., Leach, T.H., Brentrup, J.A., Knoll, L.B., Mette, E.M., & Moeller, R.E. (2015). Ecological consequences of long-term browning in lakes. *Sci Rep, 5*, 18666

Wünsch, U.J., Stedmon, C.A., Tranvik, L.J., & Guillemette, F. (2018). Unraveling the size-dependent optical properties of dissolved organic matter. *Limnology and Oceanography, 63*, 588-601

Yamashita, Y., Nosaka, Y., Suzuki, K., Ogawa, H., Takahashi, K., & Saito, H. (2013). Photobleaching as a factor controlling spectral characteristics of chromophoric dissolved organic matter in open ocean. *Biogeosciences, 10*, 7207-7217

Yamashita, Y., & Tanoue, E. (2008). Production of bio-refractory fluorescent dissolved organic matter in the ocean interior. *Nature Geoscience, 1*, 579

Zhao, Z., Gonsior, M., Luek, J., Timko, S., Ianiri, H., Hertkorn, N., Schmitt-Kopplin, P., Fang, X., Zeng, Q., & Jiao, N. (2017). Picocyanobacteria and deep-ocean fluorescent dissolved organic matter share similar optical properties. *Nature communications, 8*, 15284

3 Deriving inherent optical properties from decomposition of hyperspectral non-water absorption

3.1 Introduction

Oceanic dissolved organic matter (DOM) comprises the largest pool of fixed carbon in the ocean, roughly equivalent to the reservoir of atmospheric CO_2 (~670 Pg; Hansell et al. 2009; Ogawa et al. 2001). Yet, sources and cycling of DOM in the global ocean remain poorly constrained due to difficulty in assigning origin and tracking changes to a complex mixture of organic compounds composed of up to ~20,000 molecular formulas in a sample (Andrew et al. 2013; Mentges et al. 2017; Riedel and Dittmar 2014). A portion of DOM is optically active, colored dissolved organic matter (CDOM), and displays distinct spectral variability between uniquely sourced material, namely terrestrial and marine-derived, and different degradation pathways, such as microbial or photodegradation (Catalá et al. 2016; Danhiez et al. 2017; Helms et al. 2013; Helms et al. 2008; Zhao et al. 2017). Due to its interaction with light, CDOM can be rapidly characterized using optical sensors and is observable from autonomous and satellite platforms (e.g. Siegel et al. 2005; Xing et al. 2012). These observations are crucial to adequately model ocean physical and biogeochemical processes due to the influence of CDOM on distribution and spectral quality of light in the water column and heating of the surface ocean (Chang and Dickey 2004; Dutkiewicz et al. 2015; Kim et al. 2016). Additionally, CDOM absorption ($a_g(\lambda)$, m^{-1}; λ denotes wavelength) at visible wavelengths tracks dissolved organic carbon concentration ([DOC], $mg \cdot L^{-1}$) in coastal waters where a strong gradient of relatively degraded, terrestrial-derived material and conservative mixing produce a clear, observable signal across unique pools of CDOM (Cory and Kling 2018; Fichot and Benner 2011; Mannino et al. 2014; Stedmon and Markager 2003). This continuous dilution of $a_g(\lambda)$ in coastal waters presents predictive capability of terrestrial biomarkers (e.g. lignin) using $a_g(\lambda)$ due to unique spectral features present in terrestrial material relative to CDOM of marine origin (Fichot et al. 2016; Fichot et al. 2013; Helms et al. 2008; Vantrepotte et al. 2015). While these relationships are strong in coastal waters, open ocean waters do not display a consistent relationship between $a_g(\lambda)$ and [DOC], likely due to relatively low production rates and strong photodegradation in surface ocean waters (Helms et al. 2013; Nelson et al. 2010).

Satellite remote sensing provides a means of estimating $a_g(\lambda)$ through its influence on the amount of light leaving the water, observable from a satellite platform as spectral remotely-sensed reflectance ($R_{rs}(\lambda)$. However, due to the number of unknowns versus knowns, retrieval of distinct absorbing constituents, or inherent optical properties (IOP's) that impact the spectral quality of $R_{rs}(\lambda)$, phytoplankton ($a_{ph}(\lambda)$, m^{-1}), non-algal particulate (NAP) absorption ($a_d(\lambda)$, m^{-1}) and $a_g(\lambda)$, requires assumptions or constraints imposed on solutions. Historically, this is typically addressed by lumping $a_g(\lambda)$ and $a_d(\lambda)$ into a single term, $a_{dg}(\lambda)$, as both approximately follow an exponential increase in absorption with

decreasing wavelength and the spectral slope parameter, S_{dg}, describes $a_{dg}(\lambda)$ spectral shape. Another near-universal assumption has been assuming a fixed value for S_{dg}. For past algorithms, $a_{dg}(\lambda)$ is retrieved using either an assumed S_{dg} ranging from 0.015-0.018 nm^{-1} (Lee et al. 2002; Maritorena et al. 2002; Werdell et al. 2013), or a quasi-assumed value, where 0.015 nm^{-1} is initially assumed and then allowed to vary based on $R_{rs}(\lambda)$ at various wavelengths (e.g. 443 and 555 nm; Quasi-Analytical Algorithm (QAA), version 6; www.ioccg.org/groups/Software_OCA/QAA_v6_2014209.pdf). It is important to note that for this situation, 0.015 nm^{-1} is the lowest S_{dg} capable within the refined QAA. Alternatively, S_{dg} can be decomposed into S_d and S_g through the extended QAA approach of Dong et al. (2013), also providing a quasi-assumed S_g value and an assumed S_d value.

Increasing observations of $a_g(\lambda)$ have shown that S_g displays significant variability and is also capable of characterizing CDOM of unique source and degradation state (Danhiez et al. 2017; Grunert et al. 2018; Helms et al. 2008) . Beyond being linked to uniquely sourced material, the spectral shape of $a_g(\lambda)$ has been linked to specific environmental conditions (e.g. phytoplankton blooms, nutrient conditions) that are not accurately quantified by changes in the magnitude of $a_g(\lambda)$ at visible wavelengths (e.g. 412 nm; Asmala et al. 2018; Danhiez et al. 2017). Considering that spectral variability in $a_g(\lambda)$ can be attributed to specific production and degradation processes (Helms et al. 2008, 2013) as well as different environmental conditions (Asmala et al. 2018), it is likely that this parameter contains very useful information regarding food web processes and marine carbon cycling relevant to understanding the balance of the marine DOM carbon reservoir.

To date, satellite oceanographers have recognized the importance of retrieving S_g for a fuller understanding of large-scale biogeochemical processes but have been limited by multispectral satellite observing capabilities. Hyperspectral capabilities are currently primarily limited to *in situ* and airborne sensors, with past targeted observations available from the National Aeronautics and Space Administration's (NASA) Hyperspectral Imager of the Coastal Ocean (HICO) and broad spatial resolution hyperspectral data available from the Scanning Imaging Absorption Spectrometer for Atmospheric Chartography (SCIAMACHY) that is capable of characterizing unique phytoplankton functional groups at this spatial resolution (Bracher et al. 2009; Sadeghi et al. 2012). Future hyperspectral sensors with increased spatial and/or temporal resolution are planned for launch in the near future (~1-5 years) including the German Aerospace Center's Environmental Mapping and Analysis Program (EnMAP) sensor and NASA's Plankton, Aerosol, Cloud and ocean Ecosystem (PACE) sensor. Specifically, PACE is anticipated to provide 5 nm spectral resolution from 350-890 nm, offering an unprecedented view of the global ocean on climate-relevant spatial (1 km resolution) and temporal (2-3 day revisit time) scales. In theory, this sensor could provide remote estimation of S_g. However, we are unaware of any algorithms capable of retrieving $a_{ph}(\lambda)$ and $a_{dg}(\lambda)$ from total non-water absorption ($a_{t\text{-}w}(\lambda)$) free of explicit assumptions regarding spectral shape of $a_{dg}(\lambda)$ or other parameters that influence spectral shape of IOP's (e.g. slope of backscattering). Beyond estimation of S_{dg} and more accurate spectral retrievals of a_{ph}, such a method would provide clearer spectral features for the derivation of specific phytoplankton functional types, including Gaussian

fitting and second or fourth derivative analysis of phytoplankton pigments (Chase et al. 2017; Vandermeulen et al. 2017; Wang et al. 2017).

Here, we present a new algorithm that estimates S_{dg}, $a_{dg}(\lambda)$ and $a_{ph}(\lambda)$ free of explicit assumptions from $a_{t\text{-}w}(\lambda)$ using derivative analysis, iterative spectral evaluation and Gaussian decomposition of total non-water absorption spectra. We focus on accurate retrieval of S_{dg} and $a_{dg}(\lambda)$ to represent biogeochemical variability in NAP and CDOM absorption represented by the spectral shape and magnitude of $a_{dg}(\lambda)$. Results show good success in retrieving IOP magnitude and spectral shape. We discuss potential biogeochemical variability inferred by S_{dg} along with the significance of accurate retrievals of both $a_{dg}(\lambda)$ and $a_{ph}(\lambda)$ from hyperspectral absorption spectra. Results from the algorithm described here, Derivative Analysis and Iterative Spectral Evaluation of Absorption (DAISEA), suggest that hyperspectral satellite ocean color data will improve our ability to track biogeochemical processes affiliated with variability in $a_{dg}(\lambda)$ and S_{dg}. Finally, we discuss the likelihood of utilizing independent satellite datasets to separate $a_{dg}(\lambda)$ into $a_d(\lambda)$ and $a_g(\lambda)$.

3.2 Methods

3.2.1 Data

In situ data were accessed from NASA's SeaWiFS Bio-optical Archive and Storage System (SeaBASS, https://seabass.gsfc.nasa.gov/) on January 12, 2018 (Werdell et al. 2003). We focused our collection on data where $a_{ph}(\lambda)$, $a_d(\lambda)$ and $a_g(\lambda)$ were all measured coincidentally on a benchtop spectrophotometer within 10 m of the surface (Fig. 1). We initially quality controlled each set of absorption spectra by considering if any values were below zero for an individual spectra, calculating an offset for the most negative value and shifting the entire spectrum by this amount. In doing so, spectral shape is retained while removing poorly defined absorption values that result in negative algorithm solutions. We removed any spectra where S_{dg} was less than 0.004 nm^{-1}, values unrealistic with historic observations and estimates (e.g. Siegel et al. 2002; Wang et al. 2005) . Additionally, spectra that had been sampled at a resolution less than 2 nm were not considered to ensure spectral shape was maintained when downsampling. After removing poor quality spectra, a total of 4,787 spectra remained. These spectra were randomly split into training (n=3,434; Fig. 1a) and test datasets (n=1,353; Fig. 1b) so that training spectra accounted for ~75% of total spectra. All absorption spectra were subsampled to 5 nm data either through direct sub-sampling or linear interpolation to avoid introducing artificial curvature, with the spectral range from 350-700 nm used (71 data points). Some spectra were not sampled down to exactly 350 nm but were measured at or below 355 nm (e.g. 350.7, 355 nm; n=79); for these spectra, we extrapolated to 350 nm using a discretized partial differential equation with an enhanced plate metaphor (D'Errico 2005). We focus on 5 nm spectral resolution here for an assessment of performance relative to the anticipated resolution of PACE.

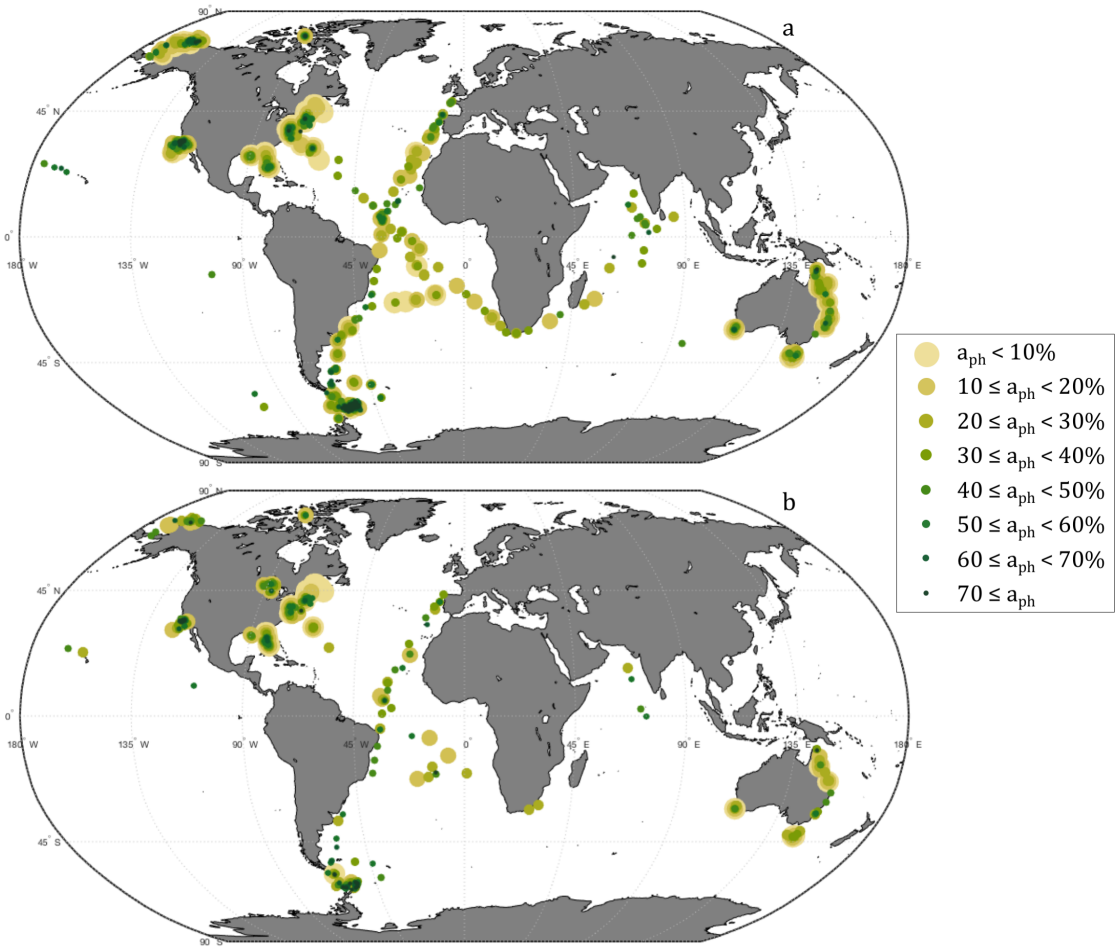


Figure 1. Locations of spectra utilized in the (a) training dataset and (b) test dataset where color and size represent spectra grouped by varying $a_{ph}(440)$ percent contribution.

3.2.2 DAISEA Algorithm Development

Our approach for decomposing $a_{t\text{-}w}$ focuses on estimating $a_{dg}(\lambda)$ first through derivative analysis, optimizing the fit of $a_{dg}(\lambda)$ through iterative spectral evaluation, then estimating $a_{ph}(\lambda)$ using Gaussian decomposition. Steps described in this section are summarized in a schematic and accompanied by figures illustrating the primary components of each step (Fig. 2). Steps 1-7 evalute $a_{t\text{-}w}(\lambda)$ to optimize estimates of $a_{dg}(\lambda)$ and $a_{ph}(\lambda)$ and Step 8 is a Gaussian decomposition of $a_{t\text{-}w}(\lambda)$ using estimated $a_{dg}(\lambda)$ and $a_{ph}(\lambda)$ with constraints defined below.

Step 1

To first parameterize $a_{dg}(\lambda)$, the second derivative of $a_{t\text{-}w}$ is calculated as

$$\frac{d^2 a_{t-w}}{d\lambda^2} \approx \frac{a_{t-w}(\lambda_i - 2a_{t-w}(\lambda_j) + a_{t-w}(\lambda_k)}{\Delta\lambda^2} \quad (5)$$

where $\Delta\lambda$ indicates the wavelength resolution used to measure $a_{t\text{-}w}(\lambda)$ (here, 5 nm), with Eq. 1 following the approach of Tsai and Philpot (1998). Points where the second derivative equals 0 indicate inflection points of the spectrum (Fig. 2a; Lee et al. 2007). For $a_{t\text{-}w}$, these are points where individual phytoplankton pigments least impact the underlying exponential signal and thus are considered as the observed signal most likely representative of $a_{dg}(\lambda)$. These points are defined as λ_{d0} and are found by rounding $d^2a_{t\text{-}w}(\lambda)$ to 0 at the decimal corresponding to the relative median magnitude of the second derivative (e.g. median $d^2a_{t\text{-}w}(\lambda)$=0.005, then round to the third decimal place), which itself is a function of the magnitude of observed absorption.

Step 2

Using these wavelengths, an initial exponential expression is fitted following

$$a_{t-w}(\lambda_{d0}) = a_{t-w}(\lambda_0)e^{-S(\lambda_{d0}-\lambda_0)} \quad (6)$$

where λ_0 is the minimum wavelength in λ_{d0} (Fig. 2b). S is used as the initial estimate of S_{dg} and $a_{dg}(\lambda_0)$ is estimated as $a_{t\text{-}w}(440)$ multiplied by the estimated contribution of $a_{dg}(\lambda)$ (decimal value) using a piece-wise exponential relationship derived from the training dataset as follows:

$$\%\, a_{phy}(440) = 1.038e^{-0.9257\left(\frac{a_{t-w}(555)}{a_{t-w}(680)}\right)} \text{where } \frac{a_{t-w}(555)}{a_{t-w}(680)} > 0.685 \quad (7)$$

or

$$\%\, a_{phy}(440) = 2.088e^{-1.946\left(\frac{a_{t-w}(555)}{a_{t-w}(680)}\right)} \text{ where } \frac{a_{t-w}(555)}{a_{t-w}(680)} \leq 0.685 \quad (8)$$

and

$$\%\, a_{dg}(440) = 100 - \%\, a_{phy}(440) \quad (9)$$

The spectra for $a_{dg}(\lambda)$ is then estimated (Fig. 2b) as follows

$$a_{dg}(\lambda) = \left(a_{t-w}(440) \cdot \%a_{dg}(440)\right) e^{-S_{dg}(\lambda-440)} \quad (10)$$

Step 3

To determine if the $a_{dg}(\lambda)$ estimate is feasible, we compare it to $a_{t\text{-}w}(\lambda)$:

$$a_{residual}(\lambda) = a_{t-w}(\lambda) - a_{dg}(\lambda) \quad (11)$$

If $a_{residual}(\lambda)$ is negative, an offset is calculated by finding the wavelength where $a_{dg}(\lambda)$ is most overestimated, from

$$a_{dg}(\lambda_0) = a_{t-w}(\lambda_0) - max\left|a_{t-w}(\lambda_{ind}) - a_{dg}(\lambda_{ind})\right| \quad (12)$$

where $\lambda_0=\lambda_{ind}$ corresponding to the wavelength where $a_{dg}(\lambda)$ is most overestimated. We then re-calculate $a_{dg}(\lambda)$ using the new λ_0 and the following expression:

$$a_{dg}(\lambda) = a_{dg}(\lambda_0)e^{-S_{dg}(\lambda-\lambda_0)} \quad (13)$$

If $a_{residual}(\lambda)$ is always positive, the previous variables - λ_0, $a_{dg}(\lambda_0)$, S_{dg} – are maintained at the current estimated values (e.g. λ_0=440 nm; Fig. 2c). If $a_{residual}(\lambda)$ is negative at any point, $a_{residual}(\lambda)$ is re-calculated following Eq. 7 for the new estimated $a_{dg}(\lambda)$. If $a_{residual}(\lambda)$ is still negative at any point, S_{dg} is incrementally adjusted by +0.0001 nm^{-1} to a maximum adjustment of +0.011 nm^{-1}. If a potential solution has not been found, S_{dg} is then incrementally adjusted by -0.0001 nm^{-1} to a minimum adjustment of -0.004 nm^{-1}. The difference in adjustment and focus on positive adjustment values first is discussed further in Section 4.1.2. If no valid solution is found through this routine, the initial estimate of $a_{dg}(\lambda)$ is used; if a valid solution is found, that is the new $a_{dg}(\lambda)$ estimate (e.g. Fig. 2c).

Step 4

Using the new or initial $a_{dg}(\lambda)$ estimate, $a_{ph}(\lambda)$ is estimated (Fig. 2d) following

$$a_{ph}(\lambda) = a_{t-w}(\lambda) - a_{dg}(\lambda) \quad (14)$$

Step 5

To determine if $a_{dg}(\lambda)$ was estimated reasonably well, we consider the ratio of $a_{ph}(350)$:$a_{ph}(440)$, where a value greater than 1.5 is used to indicate whether a significant portion of the $a_{dg}(\lambda)$ signal is still present in the residuals. While some waters with a significant pigment contribution below 400 nm (e.g. mycosporine-like amino acids) may violate this rule, it is generally applicable following discussion in Section 4.1.2.

If $a_{ph}(350)$:$a_{ph}(440)$ is greater than 1.5, a blended estimate of $a_{dg}(\lambda)$ is produced by fitting residuals from 350-400 nm with an exponential model (Fig. 2e) following

$$a_{residual}(\lambda) = a_{residual}(\lambda_0)e^{-S_{residual}(\lambda-\lambda_0)} \quad (15)$$

A new estimate of $a_{dg}(\lambda)$, denoted as $a_{dg2}(\lambda)$, is created from

$$a_{dg2}(\lambda) = a_{dg}(\lambda) + a_{residual}(\lambda) \quad (16)$$

A new S_{dg} is re-calculated for $a_{dg2}(\lambda)$ and the next iteration of $a_{dg}(\lambda)$ is estimated from

$$a_{dg}(\lambda) = \left(a_{t-w}(440) \cdot \%a_{dg}(440)\right) e^{-S_{dg\_new}(\lambda-440)} \quad (17)$$

The $a_{dg}(\lambda)$ estimated from Eq. 13 is then iteratively evaluated within by adjusting S_{dg} and assessing whether $a_{dg}(\lambda) > a_{t\text{-}w}(\lambda)$ at any wavelength. If it is, an offset is calculated following Eq. 8 where $a_{dg}(\lambda)$ is most overestimated (Step 3, Fig. 2). If this step is performed, a new λ_0 and $a_{dg}(\lambda_0)$ is estimated at the wavelength of the largest overestimation and $a_{dg}(\lambda_0)$ is no longer set to the empirically-derived estimate of $a_{dg}(440)$. These steps are performed in a step-wise manner until $a_{phy}(350)$:$a_{phy}(440)$ is less than 1.5 (Fig. 2f,g).

Step 6

Gaussian decomposition of the optimized residual from Eq. 7 is performed. For this, we utilize a generic version of Eq. 1 to calculate the second derivative of estimated $a_{ph}(\lambda)$ (Fig. 2h). The second derivative is smoothed with a linear Savitzky-Golay filter using a smoothing window as close to 9 nm as possible with the provided spectral resolution (e.g. 10 nm here). The smoothed second derivative is inverted and local maxima are identified using a peak finding function, where local maxima are identified as any value greater than the value before or after and peak width is identified as the best Gaussian fit of respective minima. These identified peaks are used as an initial estimate of the number of peaks and each peak's location and width (Fig. 2h,i) with each Gaussian curve modeled following

$$f(x, \varphi, \mu, \sigma) = \varphi e^{-\frac{(x-\mu)^2}{2\sigma}} \quad (18)$$

where σ (nm) is the width of the curve, φ (m^{-1}) is the height of the Gaussian curve defined as $\varphi = \frac{1}{\sigma\sqrt{2\pi}}$ and μ (nm) is the peak center position.

Step 7

Peak height is then estimated by prioritizing peaks based on their relative prominence, identified as the height determined from peak identification in the previous step. When identified in this manner, pigments that do not overlap, or overlap little, are identified first, followed by peaks that are observed as a shoulder (e.g. chlorophyll-a peak at 676 nm is typically prioritized for fitting first). Based on the order of prominence, $a_{ph}(\lambda)$ is iteratively fit following

$$a_{ph_i}(\lambda) = a_{ph_{i-1}}(\lambda) - \sum_{i=1}^{n} \varphi_i e^{-\frac{(x-\mu_i)^2}{2\sigma_i^2}} \quad (19)$$

Due to the additive nature of fitting Gaussian curves, there is potential for some peaks to have a negative height. After initially estimating the shape of each Gaussian curve, we filter

out peaks with negative heights and we cap the total possible number of peaks at 16 ; most Gaussian decomposition schemes assume the presence of ~12 peaks (e.g. Hoeppfner and Sathyendranath 1993; Wang et al. 2016; Chase et al. 2017) . These studies have considered similar peak locations with minor differences accounting for a total of 16 unique peak locations in the literature. From this, we assumed if more than 16 peaks are present and all have a positive peak height, some identified peaks are noise or signals not affiliated with phytoplankton pigments. We sort for likely pigment signals by prominence, using the same method described for peak height, and select the 16 most prominent identified peaks. Next, we use our estimated Gaussian curves as input into a least squares Gaussian decomposition model that best fits our initial $a_{ph}(\lambda)$ estimate (Eq. 10) with our estimated Gaussian curves and fitting constraints as described in Step 8 to define an updated set of Gaussian curves (Fig. 2j) following the expression:

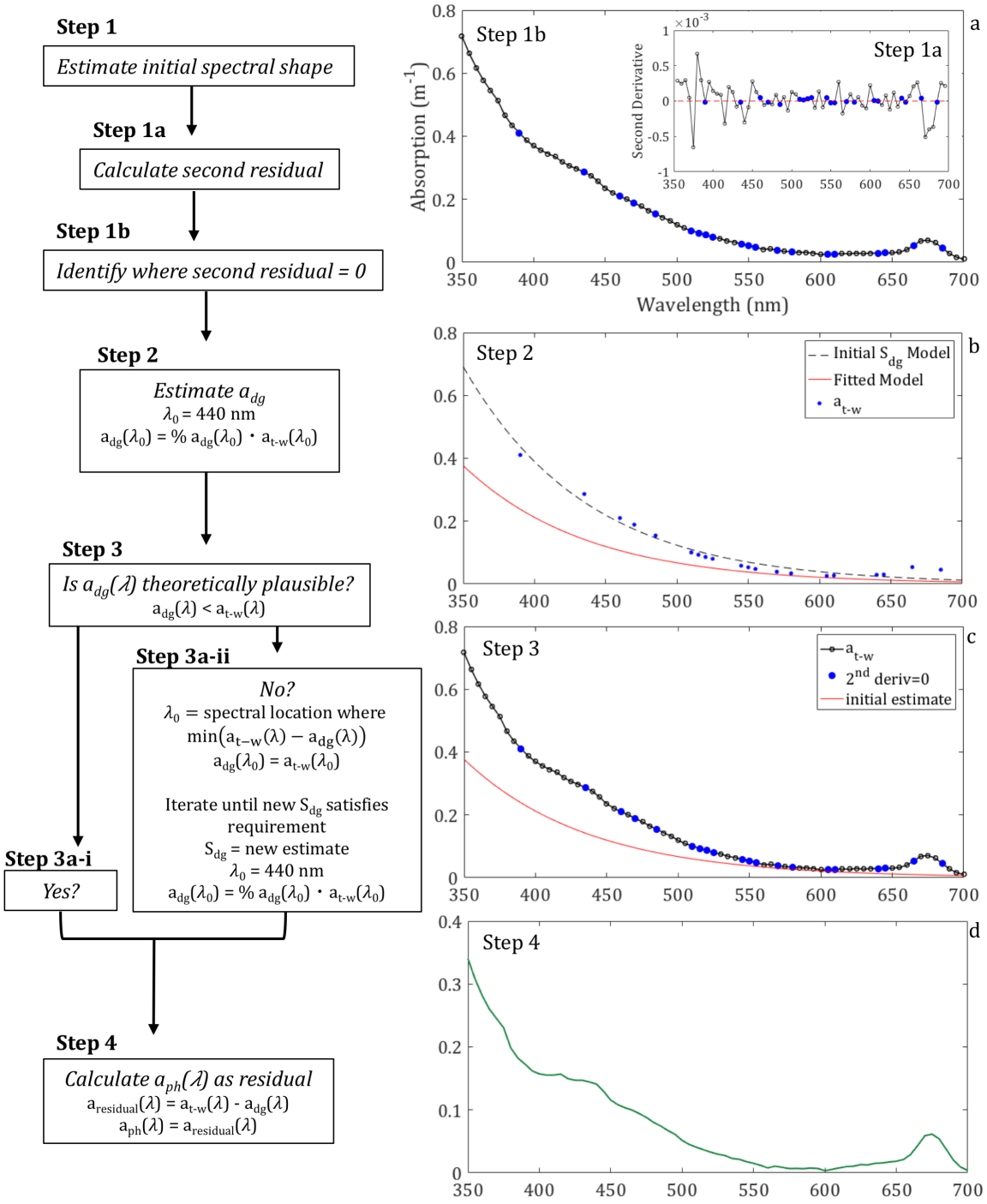
$$a_{ph}(\lambda) = \sum_{i=1}^{n} \varphi_i e^{-\frac{(x-\mu_i)^2}{2\sigma_i^2}} \tag{20}$$

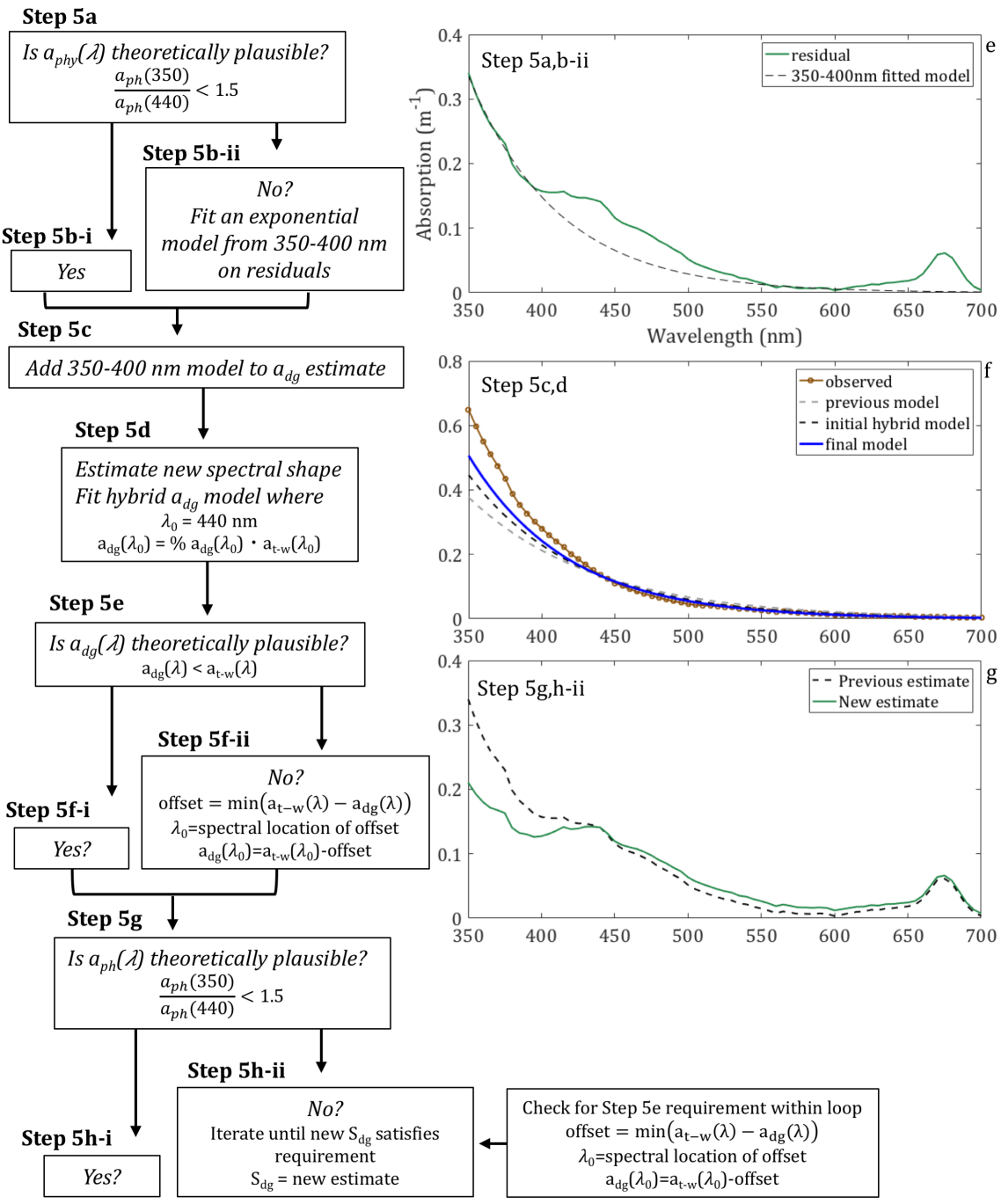
Step 8

Results from Steps 1-7 provide the start point for a combined retrieval of $a_{dg}(\lambda)$ and $a_{ph}(\lambda)$ from $a_{t\text{-}w}(\lambda)$. Using the current estimate of $a_{dg}(\lambda)$ from Steps 1-7 and an estimate for each identified Gaussian curve fitted to $a_{ph}(\lambda)$, a least squares fitting approach is performed using the following expression:

$$a_{t-w}(\lambda) = a_{dg}(\lambda_0)e^{-S_{dg}(\lambda-\lambda_0)} + \sum_{i=1}^{n} \varphi_i e^{-\frac{(x-\mu_i)^2}{2\sigma_i^2}} \tag{21}$$

Analogous to methods used for identifying poorly constrained features that deviate from an underlying exponential signal presented elsewhere (e.g. Massicotte and Markager 2016), the model decomposes $a_{t\text{-}w}(\lambda)$ by utilizing a baseline exponential (Eq. 9) accompanied by a pre-defined number of Gaussian components based on previous steps (Eq. 16). This method differs from other Gaussian decomposition methods applied to particulate absorption (a_p), in that those methods typically have a pre-defined number of Gaussian components based on analysis of separate $a_{ph}(\lambda)$ for the respective system (e.g. Chase et al. 2013; Wang et al. 2016). This methodology fits primary pigments with width estimated from spectral features identified in the second derivative of estimated $a_{ph}(\lambda)$, allowing for a constrained solution to decomposing $a_{t\text{-}w}(\lambda)$ while not assuming the presence any specific types of phytoplankton. Parameters in Eq. 17 are constrained utilizing results from Steps 1-7: $a_{dg}(\lambda_0)$ can vary from 0 m^{-1} to $a_{t\text{-}w}(\lambda_0)$, S_{dg} can vary by -0.002 nm^{-1} to +0.003 nm^{-1} from the input estimate, Gaussian peak width can vary from input width to 3 times the input width, Gaussian peak height can vary by 0.25 times input height to 3 times input height and μ is fixed at the identified location due to high confidence in the second derivative analysis. DAISEA output is as follows: $a_{dg}(\lambda)$ is that estimated in Eq. 17, while


Step 1
Estimate initial spectral shape
Step 1a
Calculate second residual
Step 1b
Identify where second residual = 0
Step 2
Estimate a_{dg}
λ_0 = 440 nm
$a_{dg}(\lambda_0)$ = % $a_{dg}(\lambda_0)$ • $a_{t\text{-}w}(\lambda_0)$
Step 3
Is $a_{dg}(\lambda)$ theoretically plausible?
$a_{dg}(\lambda) < a_{t\text{-}w}(\lambda)$
Step 3a-ii
No?
λ_0 = spectral location where $\min(a_{t-w}(\lambda) - a_{dg}(\lambda))$
$a_{dg}(\lambda_0) = a_{t\text{-}w}(\lambda_0)$
Iterate until new S_{dg} satisfies requirement
S_{dg} = new estimate
λ_0 = 440 nm
$a_{dg}(\lambda_0)$ = % $a_{dg}(\lambda_0)$ • $a_{t\text{-}w}(\lambda_0)$
Step 3a-i
Yes?
Step 4
Calculate $a_{ph}(\lambda)$ as residual
$a_{residual}(\lambda) = a_{t\text{-}w}(\lambda) - a_{dg}(\lambda)$
$a_{ph}(\lambda) = a_{residual}(\lambda)$
Step 1b
Step 1a
Second Derivative
Absorption (m^{-1})
Wavelength (nm)
a
Step 2
Initial S_{dg} Model
Fitted Model
$a_{t\text{-}w}$
b
Step 3
$a_{t\text{-}w}$
2nd deriv=0
initial estimate
c
Step 4
d


Step 5a
Is $a_{phy}(\lambda)$ theoretically plausible?
$\frac{a_{ph}(350)}{a_{ph}(440)} < 1.5$
Step 5b-ii
No?
Fit an exponential model from 350-400 nm on residuals
Step 5b-i
Yes
Step 5c
Add 350-400 nm model to a_{dg} estimate
Step 5d
Estimate new spectral shape
Fit hybrid a_{dg} model where
λ_0 = 440 nm
$a_{dg}(\lambda_0) = \% \, a_{dg}(\lambda_0) \cdot a_{t\text{-}w}(\lambda_0)$
Step 5e
Is $a_{dg}(\lambda)$ theoretically plausible?
$a_{dg}(\lambda) < a_{t\text{-}w}(\lambda)$
Step 5f-ii
No?
offset = $\min(a_{t-w}(\lambda) - a_{dg}(\lambda))$
λ_0=spectral location of offset
$a_{dg}(\lambda_0)=a_{t\text{-}w}(\lambda_0)$-offset
Step 5f-i
Yes?
Step 5g
Is $a_{ph}(\lambda)$ theoretically plausible?
$\frac{a_{ph}(350)}{a_{ph}(440)} < 1.5$
Step 5h-ii
No?
Iterate until new S_{dg} satisfies requirement
S_{dg} = new estimate
Step 5h-i
Yes?
Check for Step 5e requirement within loop
offset = $\min(a_{t-w}(\lambda) - a_{dg}(\lambda))$
λ_0=spectral location of offset
$a_{dg}(\lambda_0)=a_{t\text{-}w}(\lambda_0)$-offset
Step 5a,b-ii
residual
350-400nm fitted model
Absorption (m^{-1})
Wavelength (nm)
e
Step 5c,d
observed
previous model
initial hybrid model
final model
f
Step 5g,h-ii
Previous estimate
New estimate
g

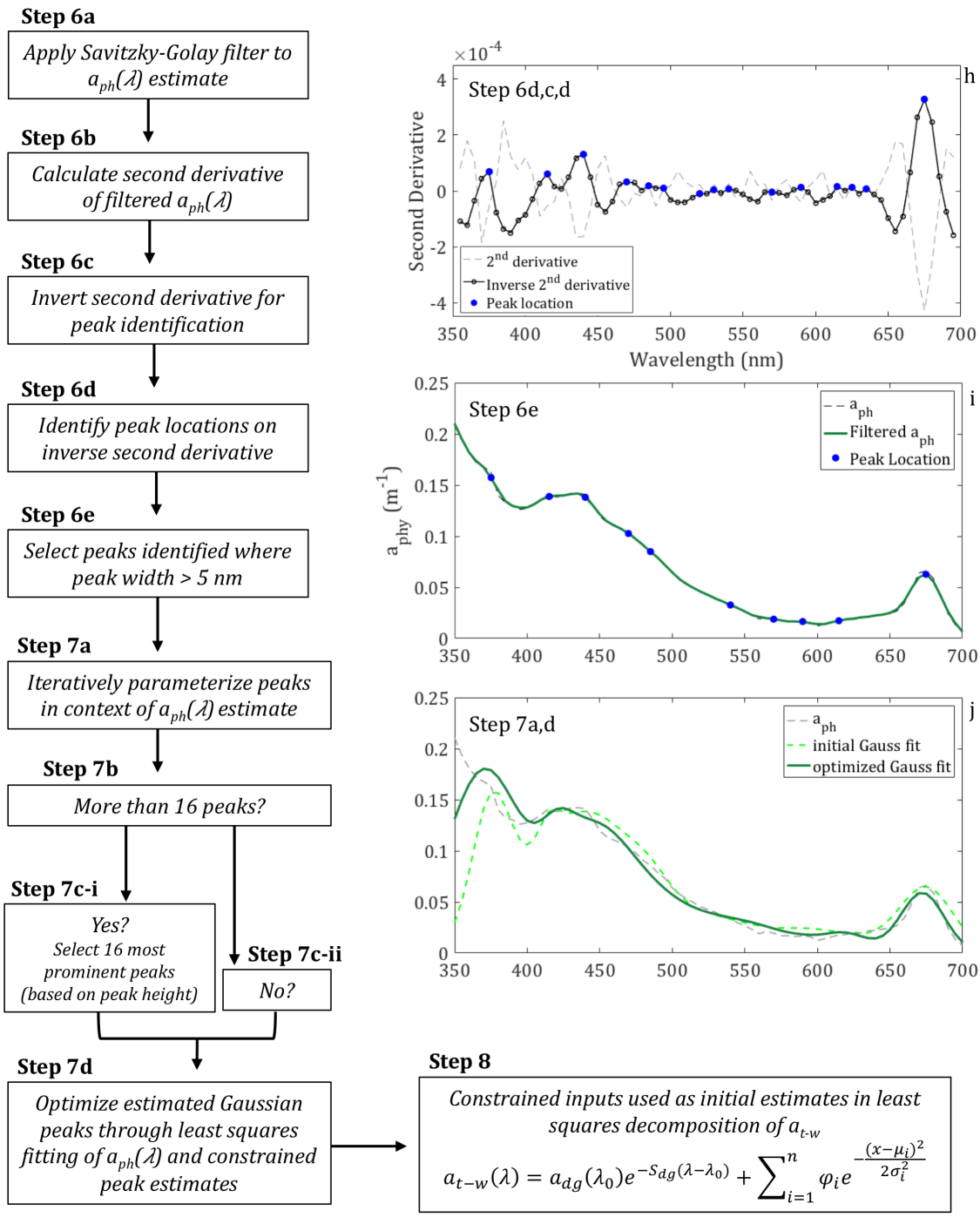


Figure 2. Schematic and figures illustrating primary steps for the Gaussian decomposition algorithm. This schematic is provided to aid in visualizing and organizing the steps detailed in the accompanying text (Section 2.2). Each figure illustrates the step as indicated for an example spectra. Not all spectra require all the steps depicted, while some spectra walk through all the steps (e.g. Fig. 2c shows a successful first guess, while some spectra required an iteration at this step).

$a_{ph}(\lambda)$ is the difference between observed $a_{t\text{-}w}(\lambda)$ and $a_{dg}(\lambda)$ from Eq. 17 (Fig. 3). Step 8 ensures coherence between the exponential signal and overlying deviations due to $a_{ph}(\lambda)$ as constrained through Steps 1-7 in a flexible manner, while not assuming that $a_{ph}(\lambda)$ can be best parameterized by 6-8 Gaussian curves. Fitting of secondary features is possible but also increases the probability of over-constraining a solution (i.e. less flexibility is adjustments to $a_{dg}(\lambda)$).

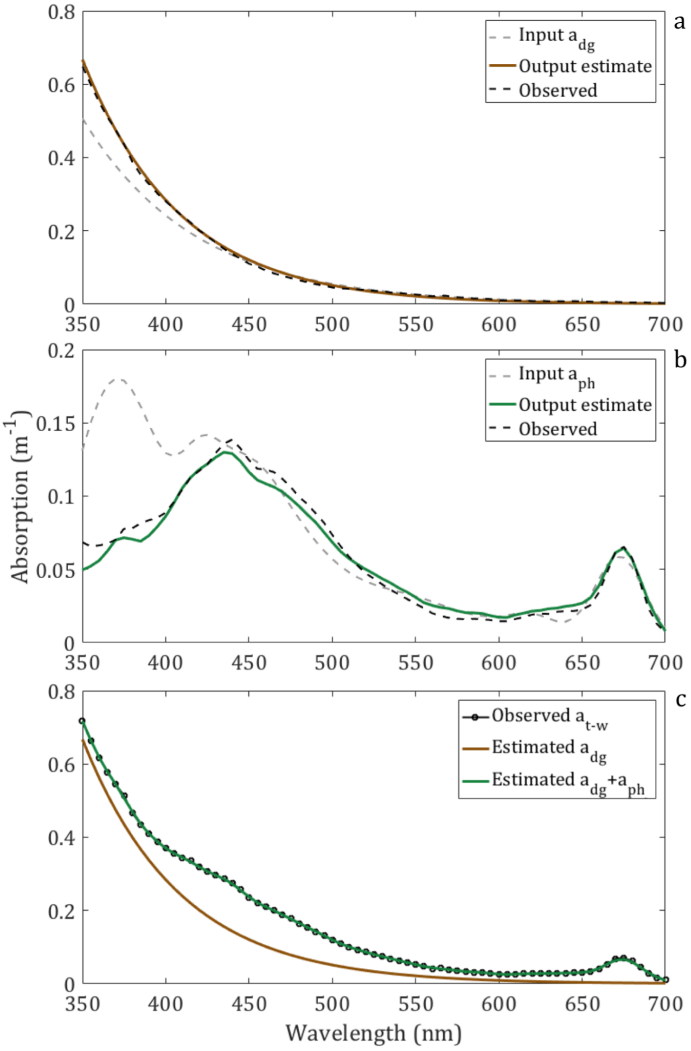


Figure 3. Algorithm output for the example spectra depicted in Fig. 2. Gray dashed lines indicate the estimated (a) $a_{dg}(\lambda)$ and (b) $a_{ph}(\lambda)$ used as input into the least squares Gaussian decomposition of observed $a_{t\text{-}w}(\lambda)$ and black dashed lines indicate the respective observed IOP. For (a) and (b), respective colored lines display algorithm output. For (c), the brown line represents $a_{dg}(\lambda)$ algorithm output, the green line represents $a_{dg}(\lambda) + a_{ph}(\lambda)$ algorithm output and the black line with circles indicates observed $a_{t\text{-}w}(\lambda)$. This example shows how a Gaussian component can be fitted to the residuals derived from Step 5 (Fig. 2), but is minimized due to a better fit of observed $a_{t\text{-}w}(\lambda)$ with an exponential curve.

3.2.2.1 Low $a_{ph}(\lambda)$ waters

We found that waters dominated by $a_{dg}(\lambda)$ were best decomposed by fitting an initial exponential function and adjusting to a realistic solution following Eq. 7 and 8. These cases are identified after Eq. 3 and 4; waters are considered dominated by $a_{dg}(\lambda)$ where the ratio of $a_{t\text{-}w}(555)$:$a_{t\text{-}w}(680) > 2.528$ (the empirical value indicating $a_{ph}(440) < 10\%$). For such situations, the algorithm opts out of the Gaussian decomposition routine and follows a simplified routine analogous to Steps 2-4, where S_{dg} is considered equivalent to S calculated for $a_{t\text{-}w}$ (Eq. 2), and magnitude is adjusted so that $a_{dg}(\lambda) \leq a_{t\text{-}w}(\lambda)$. While Eq. 3 and 4 are empirical, the basis for these relationships and the algorithm fitting routine in general follows expected spectral relationships between wavelengths assuming varying contributions of $a_{dg}(\lambda)$ and $a_{ph}(\lambda)$ and is currently the basis for fitting S_{dg} in current non-empirical schemes (e.g. Matsuoka et al. 2013; see section 4.1.2 for further discussion). Noise in this relationship is explained by variability in the exact shape of $a_{ph}(\lambda)$ due to varying phytoplankton composition, physiology and pigment packaging effects (Bricaud and Morel 1986; Bricaud et al. 1983; Ciotti et al. 2002; Johnsen et al. 1994) as well as variability in the spectral shape and features of $a_g(\lambda)$ and $a_d(\lambda)$ (Grunert et al. 2018). As the algorithm is currently optimized for a global approach, users may find that adjusting the empirical relationship used to initially estimate $a_{dg}(440)$ and filter out $a_{dg}(\lambda)$ dominant spectra, as well as adjusting the value of 1.5 for the ratio of $a_{ph}(350)$:$a_{ph}(440)$ (Step 5), for a value more representative of their study region results in better algorithm performance.

3.2.3 Data Analysis

To assess the performance of DAISEA across a variety of water conditions, we separated spectra into eight different categories based on the percent contribution of $a_{ph}(440)$ relative to $a_{t\text{-}w}(440)$, with the distribution of spectra within these classes shown in Fig. 1. Classes ranged from <10% to >70% contribution. This classification scheme emphasizes the relative, not the absolute, contribution of phytoplankton to the overall absorption signal. Thus, waters where $a_{ph}(440)$ is the dominant contributor to total absorption are not limited to highly productive waters. In this sense, algorithm performance is not assessed across classic definitions of Case 1 or Case 2 waters (Morel and Prieur 1977). Rather, the only group dominated by coastal and inland waters was the group where $a_{ph}(440) < 10\%$.

To determine whether $a_{ph}(\lambda)$ or $a_{dg}(\lambda)$ was retrievable we calculated the absolute error in the opposing metric and compare it to the observed value. For example, if $a_{ph\_obs}(\lambda) > |a_{dg\_obs}(\lambda)\text{-}a_{dg\_est}(\lambda)|$, we consider it retrievable at that wavelength. Within each $\%a_{ph}(440)$ group, we summed the total number of instances at each wavelength where $a_{ph}(\lambda)$ or $a_{dg}(\lambda)$ was greater than the absolute error in the opposing metric and divided by the total number of spectra to get a percent retrievable metric. We also calculated root mean square error (RMSE), normalized RMSE (NRMSE), bias, mean absolute error (MAE) and unbiased absolute percent difference (UAPD) using the following expressions:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}\left[\left(x_i^{estimated}\right) - \left(x_i^{observed}\right)\right]^2}{n}} \quad (22)$$

$$NRMSE\ (\%) = \frac{RMSE}{x_{max}^{observed} - x_{min}^{observed}} \times 100 \quad (23)$$

$$Bias = \frac{1}{n}\sum_{i=1}^{n}\left(x_i^{estimated} - x_i^{observed}\right) \quad (24)$$

$$MAE = \frac{\sum_{i=1}^{n}\left(\left|x_i^{estimated} - x_i^{observed}\right|\right)}{n} \quad (25)$$

$$UAPD\ (\%) = \frac{|S_{estimated} - S_{observed}|}{0.5(S_{estimated} - S_{observed})} \times 100 \quad (26)$$

3.3 Results

3.3.1 DAISEA Performance

Across the eight categories, DAISEA performed best in waters where $a_{phy}(440)$ contributed 50-60% of total absorption, relatively consistent with results for the training data set ($a_{phy}(440)$ of 40-50%), although performance was quite good for $a_{phy}(440)$ ranging from 20-60%, outside of the 350-400 nm spectral range (NRMSE < 20% for most wavelengths). The threshold of use for estimating $a_{dg}(\lambda)$ with DAISEA appears to be when $a_{ph}(440)$ contributes < 70%; for these conditions, $a_{dg}(\lambda)$ is estimated with NRMSE < 20% from 350-650 nm. NRMSE for $a_{ph}(\lambda)$ was < 20% for the majority of wavelengths between 400-650 nm when considering conditions where $a_{ph}(440)$ contributed more than 10%. This was also consistent when considering the retrievability of $a_{ph}(440)$ under different conditions and can be considered as the threshold for estimating $a_{ph}(440)$. S_{dg} uncertainty

increased with increasing contribution of $a_{ph}(440)$; however, error was reasonable across all water conditions and estimates (Table 1).

Table 1. Median and distribution of observed S_{dg} (1st and 3rd quartile) delineated by percent $a_{ph}(440)$ contribution. Relative accuracy of estimated S_{dg} is presented as the median and distribution of absolute error (estimated S_{dg} – observed S_{dg}).

Observed S_{dg} (nm^{-1})				***Relative estimated S_{dg} accuracy (nm^{-1})***		
1st quartile	**Median**	**3rd quartile**	**$a_{ph}(440)$**	**1st quartile**	**Median**	**3rd quartile**
0.0146	0.0153	0.0161	**<10%**	-0.0003	-0.0001	-0.0001
0.0143	0.0165	0.0176	**10-20%**	-0.0015	-0.0010	+0.0004
0.0141	0.0156	0.0175	**20-30%**	-0.0010	-0.0001	+0.0013
0.0127	0.0142	0.0159	**30-40%**	-0.0015	+0.0001	+0.0017
0.0126	0.0140	0.0150	**40-50%**	-0.0024	-0.0005	+0.0016
0.0128	0.0146	0.0160	**50-60%**	-0.0018	-0.0003	+0.0021
0.0120	0.0138	0.0167	**60-70%**	-0.0024	-0.0007	+0.0005
0.0139	0.0191	0.0211	**>70%**	-0.0037	-0.0022	-0.0002

We present general algorithm performance within each category by comparing the mean observed value at each wavelength relative to the mean estimated value at each wavelength (Fig. 4). It is evident from these graphs that the algorithm is biased to overestimate $a_{phy}(\lambda)$ from 350-400 nm for spectra where $a_{phy}(440)$ is less than 50%. For spectra where $a_{phy}(440)$ contributes less than 10%, $a_{phy}(440)$ is poorly resolved in inland and coastal waters, which is not surprising given the small contribution of $a_{ph}(\lambda)$ to the overall absorption budget. In these cases, the magnitude of $a_{dg}(\lambda)$ is typically quite high, with $a_{dg}(350)$ often greater than 10 m^{-1}. Waters where $a_{phy}(440)$ contributes from 10-20% also often failed, as the empirical relationship often miscategorized these spectra as having $a_{ph}(440) < 10\%$ resulting in the assumption that the majority of $a_{t\text{-}w}(440)$ could be accurately modeled with an exponential model. This highlights the primary drawback of utilizing empirical relationships, resulting in more than half of the $a_{phy}(\lambda)$ estimates in the 10-20% $a_{ph}(440)$ group to have negative values at wavelengths greater than 650 nm, outside of the chlorophyll-a (Chl) absorption peak at 676 nm (typically assigned to 680 nm within the algorithm framework). While $a_{phy}(\lambda)$ was typically overestimated when it was a non-dominant contributor, $a_{phy}(\lambda)$ was generally underestimated when it was a dominant contributor.

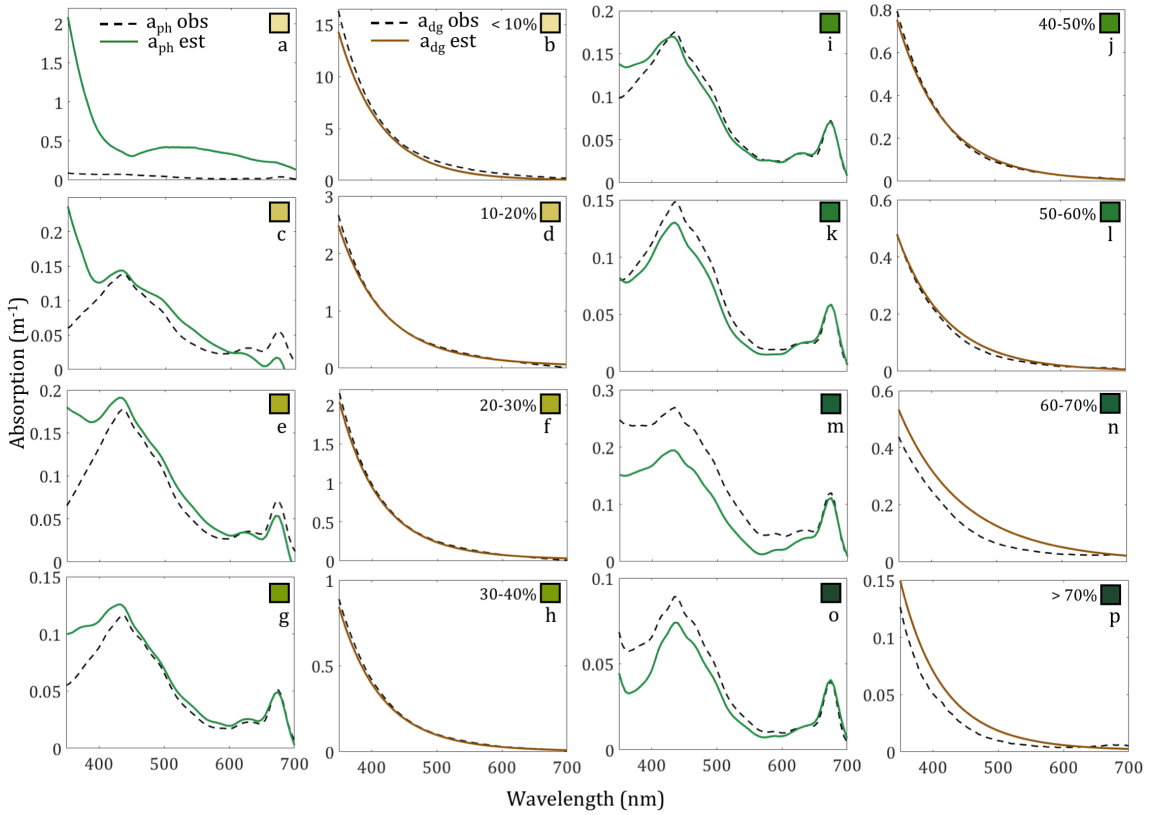


Figure 4. Mean performance of the algorithm for all spectra within each group of spectra delineated by $a_{ph}(440)$ percent contribution relative to mean observed (a,c,e,g,i,k,m,o) $a_{ph}(\lambda)$ and (b,d,f,h,j,l,n,p) $a_{dg}(\lambda)$.

Performance metrics were also delineated by $a_{ph}(440)$ contribution (Fig. 5). Across all groups, $a_{dg}(\lambda)$ was retrievable >80% of the time for wavelengths < 450 nm (Fig. 5a). For waters where $a_{dg}(\lambda)$ contributed greater than 60%, it was retrievable at a rate of >80% for all wavelengths up to 650 nm. For $a_{ph}(\lambda)$, local maxima in retrieval at the chlorophyll-a absorption peaks (~440 and 680 nm within DAISEA) were observed for all waters, with these wavelengths displaying a retrievability >80% for waters where $a_{ph}(440)$ contributed >10% (Fig. 5b). Relative error for $a_{ph}(\lambda)$ and $a_{dg}(\lambda)$ was parameterized as NRMSE and displayed excellent performance for both parameters across most wavelengths and environments. For all conditions except $a_{ph}(440)$ > 70%, $a_{dg}(\lambda)$ had a mean error of less than 20% for wavelengths from 350-650 nm (Fig. 5c), while $a_{ph}(\lambda)$ error was generally less than 20% for these wavelengths when $a_{ph}(440)$ contributed > 10% (Fig. 5d). As seen in Fig. 4, $a_{ph}(\lambda)$ was biased to greater than observed values when it was a non-dominant contributor at 440 nm and was biased towards values less than observed when it was a dominant contributor at 440 nm, and vice versa for $a_{dg}(\lambda)$ (Fig. 5e). Mean absolute error generally decreased as the contribution of $a_{ph}(440)$ increased (Fig. 5f).

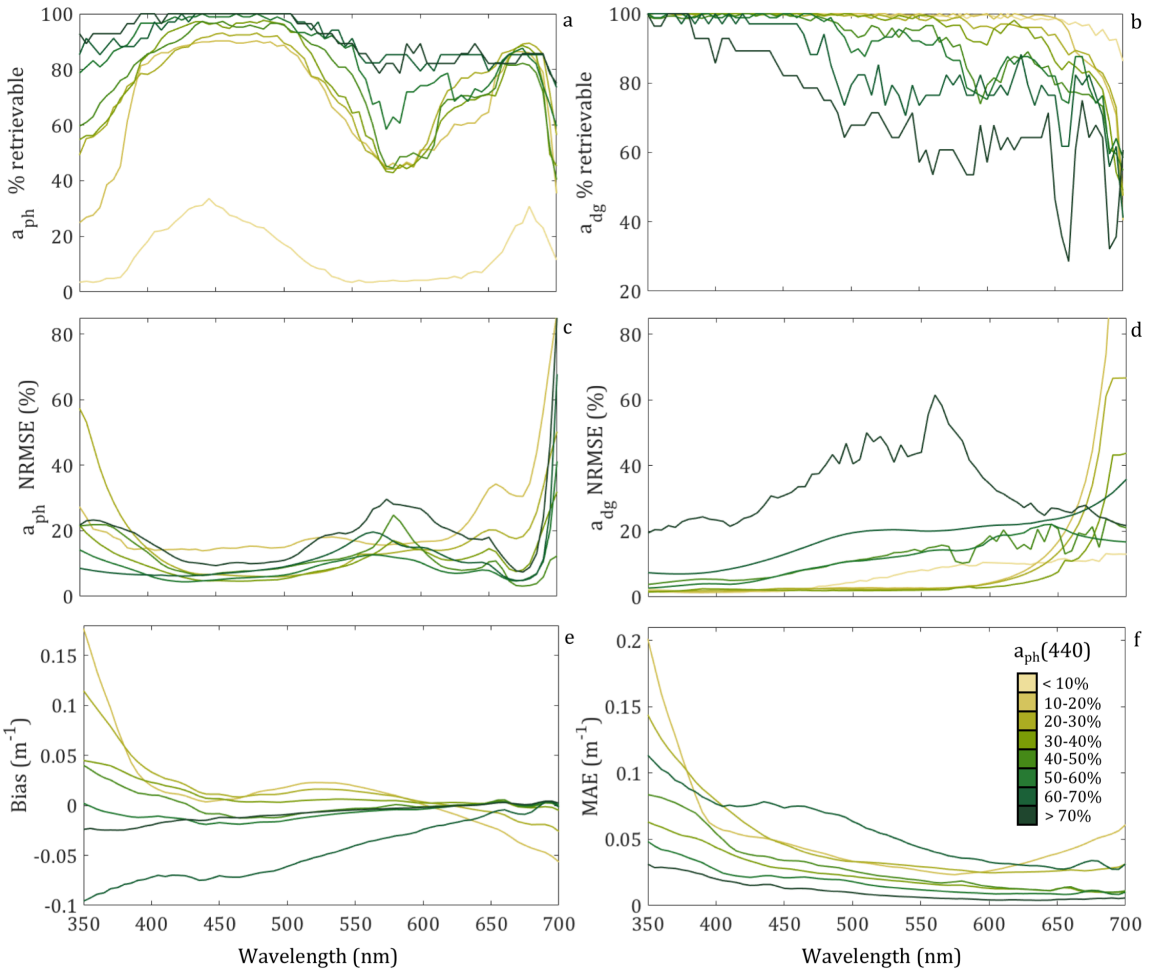


Figure 5. Performance metrics for each group delineated by $a_{ph}(440)$ percent contribution (indicated by color, from tan to dark green). Each plot corresponds to (a) percent retrievable $a_{ph}(\lambda)$, (b) percent retrievable $a_{dg}(\lambda)$, (c) $a_{ph}(\lambda)$ %NRMSE, (d) $a_{dg}(\lambda)$ %NRMSE, (e) $a_{ph}(\lambda)$ bias ($a_{dg}(\lambda)$ bias represented as inverse of each line) and (f) mean absolute error.

3.3.2 S_{dg} Estimation

One of the primary motivators for developing DAISEA was to accurately retrieve S_{dg} without any assumptions regarding spectral shape. Our results suggest that this is possible across a variety of optical conditions with a reasonable to excellent degree of accuracy, depending on the relative contribution of $a_{dg}(\lambda)$. Across the different groups of varying $a_{ph}(440)$ contribution, median error in S_{dg} varied from 0.9-17.7%, with third quartile errors ranging from 2.4-39.2% (Fig. 6a). Mean S_{dg} observed across all spectra in the test dataset was 0.0147 nm^{-1} compared to a mean estimated value of 0.0150 nm^{-1}, while median observed and estimated S_{dg} was 0.0152 and 0.0153 nm^{-1}, respectively. The relative error affiliated with each of these metrics suggests a very reasonable degree of accuracy relevant for estimating biogeochemical variability affiliated with CDOM production and degradation processes (see section 4.2). Across individual groups, we evaluated the

different errors and present anticipated accuracy for S_{dg} (Table 1). For most groups, median error was << 0.001 nm^{-1} and absolute errors affiliated with the 1st and 3rd quantiles ranged up to 0.0037 and 0.0021 nm^{-1}, respectively but were typically much smaller. We also considered distribution of error in S_{dg} across all groups and it followed a predominantly normal distribution (data not shown), without an obvious bias between observed and estimated S_{dg} regardless of percent $a_{ph}(440)$ contribution (Fig. 6b).

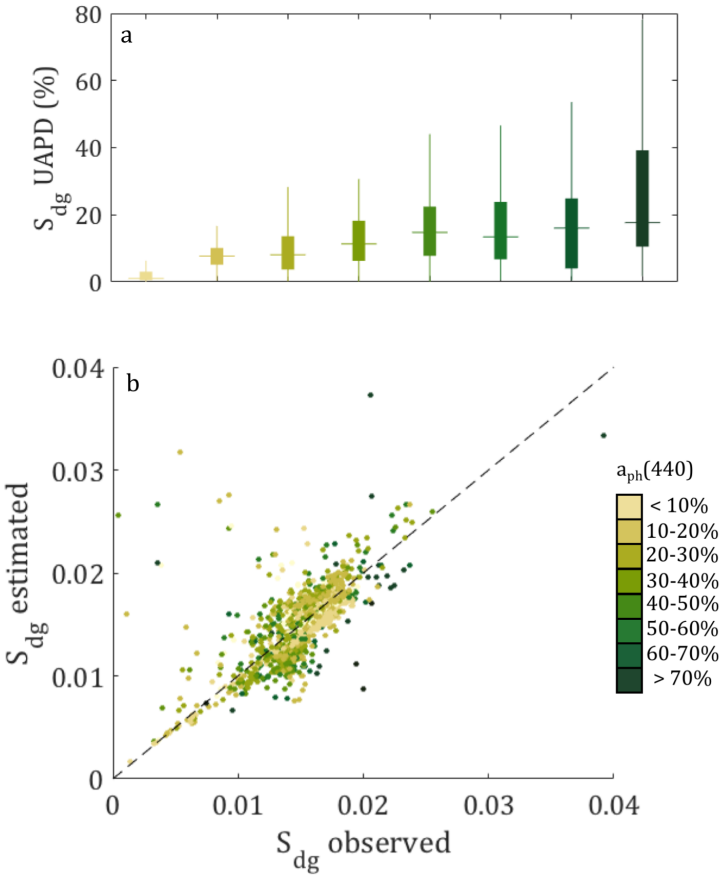


Figure 6. (a) Unbiased absolute percent difference of S_{dg} for each grouping delineated by $a_{ph}(440)$ percent contribution, indicated by the color (see legend) and (b) distribution and relationship between observed and estimated S_{dg}, with marker color indicating $\%a_{ph}(440)$ and the dashed black line (--) representing the 1:1 line.

Previous research has considered the use of UV wavelengths for estimating $a_{dg}(\lambda)$ as well as S_{dg} due to the reduced influence of a_{ph} at these wavelengths (Wei et al. 2016). While the wavelength range used to estimate S_{dg} influences its value on $a_{dg}(\lambda)$(e.g. Twardowski et al. 2004), satellite approaches focus on utilizing an S_{dg} value most applicable across visible wavelengths to maximize the accuracy of $a_{dg}(\lambda)$ estimation, as variability in this spectral range (400-700 nm) is typically quite small regardless of which wavelengths are used (e.g. $S_{400:500}$ or $S_{400:700}$). Significant spectral variability does occur from 350-400 nm (Helms et al. 2008); thus, we considered whether our approach could be simplified by focusing only on UV wavelengths (350-400 nm) to accurately retrieve S_{dg}

and maintain an accurate estimation of spectral variation of both $a_{dg}(\lambda)$ and $a_{ph}(\lambda)$ at visible wavelengths. We found that overall performance was decreased by only focusing on UV wavelengths due to the increasing rate of influence of $a_{ph}(\lambda)$ approaching 400 nm (see Section 4.1.2 for discussion) as well as a general bias towards larger $S_{350:400}$ values relative to $S_{350:700}$ on the same spectra, a relationship seen across all spectra (training and test datasets; Fig. 7). From this, we only focused on retrieving S_{dg} optimized for all wavelengths. It is very likely that separately retrieving $S_{350:400}$ will carry value, in addition to information retrieved by $S_{350:700}$ (Grunert et al. 2018).

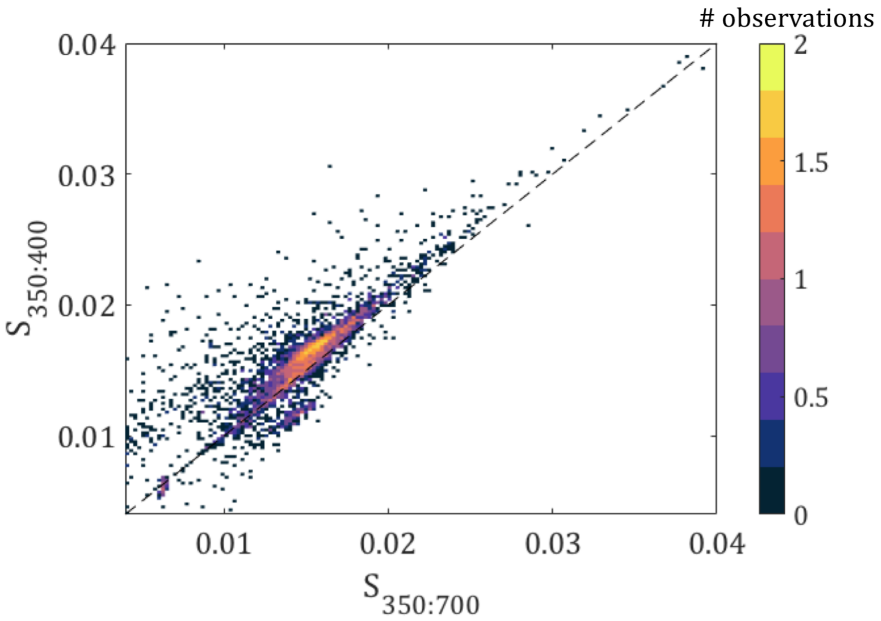


Figure 7. Probability density distribution plot displaying the typically positive bias of $S_{350:400}$ relative to $S_{350:700}$ individual $a_{dg}(\lambda)$ spectra. Colorbar indicates the number of spectra per bin and the dashed black line (--) represents the 1:1 line.

3.3.3 Consistency in Gaussian Features

We considered the accuracy of our Gaussian component locations within DAISEA by comparing to Gaussian component locations identified on observed $a_{ph}(\lambda)$ following the same approach for the test dataset (Fig. 8). Overall, fewer Gaussian components were identified in the $a_{ph}(\lambda)$ estimated within the algorithm (total peaks=7,539; 5.6 peaks/spectra) than were fitted on observed $a_{ph}(\lambda)$ spectra (total peaks=8,794; 6.5 peaks/spectra). However, when considering DAISEA output, the number of fitted peaks was higher than for the observed $a_{ph}(\lambda)$ spectra (total peaks=10,394; 7.7 peaks/spectra). Since $a_{ph}(\lambda)$ estimated from the algorithm is derived from the smoothed residuals of $a_{t\text{-}w}(\lambda) - a_{dg\_est}(\lambda)$, this means that the additional noise in the spectra is derived from deviations in $a_d(\lambda)$ and $a_g(\lambda)$ not accounted for by a strictly exponential fit. We discuss potential reasons for an increase in fitted peaks in DAISEA output over the observed in Section 4.3, as well

as fitting significantly fewer peaks under our approach than other Gaussian decomposition approaches (e.g. Chase et al. 2013).

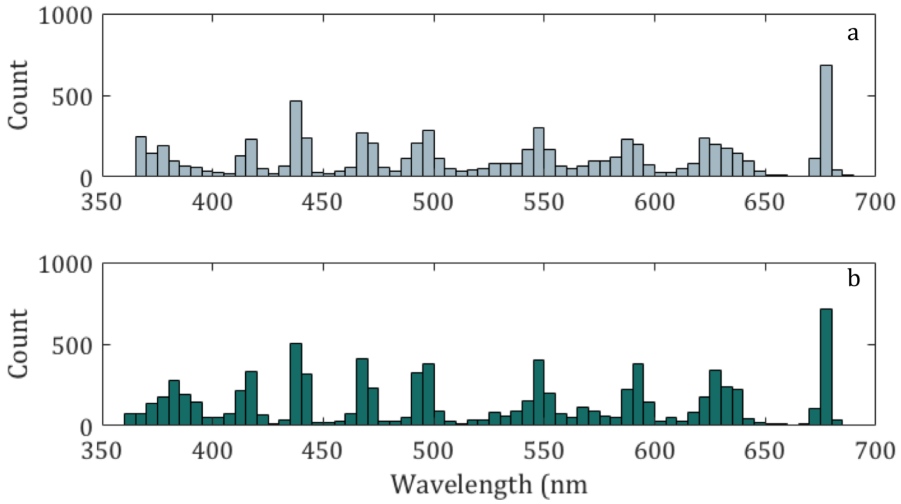


Figure 8. Distribution of identified peak locations for (a) observed $a_{ph}(\lambda)$ and (b) $a_{ph}(\lambda)$ estimated from $a_{t\text{-}w}(\lambda)$. Overall, identified peaks were quite consistent between the two signals displaying the strength of the scheme for initial estimates and constraints used for the Gaussian decomposition model.

3.4 Discussion

3.4.1 DAISEA

3.4.1.1 Application

As evidenced here and elsewhere, hyperspectral ocean color data theoretically provides a means for estimating more variables in a less constrained manner (Dierssen et al. 2015; Uitz et al. 2015; Vandermeulen et al. 2017; Wang et al. 2017). Global variability in water optical properties is significant yet the non-uniqueness of $R_{rs}(\lambda)$ hampers consistent interpretation across both empirical and semi-analytic methods (Babin et al. 2003; Bricaud et al. 1995; Grunert et al. 2018; Lee et al. 2002; Mannino et al. 2014; Mélin and Vantrepotte 2015). Previous concepts for working around this issue, particularly in light of multispectral limitations, have included screening $R_{rs}(\lambda)$ to most likely cases based on optical water types, linear matrix inversions and neural networks ((Brando et al. 2012; Hieronymi et al. 2017; Mélin and Vantrepotte 2015; Trochta et al. 2015). Increasingly, independent satellite parameters are used to compare IOP's or derived products, particularly satellite-based Light Detection And Ranging (lidar) systems and ocean color remote sensing (Behrenfeld et al. 2013; Behrenfeld et al. 2016). Lidar is a proposed sensing technology for PACE, offering the potential of hyperspectral absorption and lidar on-board the same satellite.

Ideally, step-wise, iterative approaches can be used to analyze a signal and retrieve IOP's with independent satellite observations used for cross-validation. This approach may seem quite optimistic but appears increasingly likely. NASA's Cloud-Aerosol Lidar with Orthogonal Polarization (CALIOP) sensor, despite focusing on retrieval of atmospheric products, provides a means of estimating particle absorption and backscattering in the upper surface ocean. By combining hyperspectral ocean color retrieved IOPs with IOP estimates derived from independent sensors such as lidar, it seems likely that an iterative, step-wise approach could be used to assess each retrieved parameter and the likelihood of co-occurrence (e.g. lidar-determined particulate backscattering magnitude and/or slope alongside an ocean color estimated $a_{ph}(\lambda)$), allowing for further constraining solutions using direct observations, rather than pre-defined relationships. Taking this a step further, we suggest that accurate estimates of individual parameters can help inform each other, e.g. observations of $a_{ph}(\lambda)$ indicative of nutrient stress and affiliated S_g values consistent with this observation, or ratios of $b_{bp}(\lambda)/a_p(\lambda)$ to indicate particulate composition, size distribution and bloom state (Asmala et al. 2018; Behrenfeld and Milligan 2013; Neukermans et al. 2016). Some work has considered the ability to more effectively constrain $a_{dg}(\lambda)$ and S_{dg} with UV wavelengths (Wei et al. 2016); research in this regard is relatively limited when compared to that addressing $a_{ph}(\lambda)$, despite dissolved organic matter accounting for ~98% of total organic carbon in the global oceans (Bishop 1999; Gardner et al. 2006) and the ability of optics to effectively characterize a large degree of production and degradation processes occurring in the global ocean (see Section 4.2 for a detailed discussion). We show here that accurate retrieval of S_{dg} free of explicit assumptions is possible within a step-wise framework, providing a novel, unbiased metric for further characterizing ocean biogeochemical processes with hyperspectral satellite data.

3.4.1.2 General Framework and Empirical Relationships

The general premise of DAISEA is that $a_{dg}(\lambda)$ can be accurately modeled using an exponential model and that deviations from this exponential model are solely due to $a_{ph}(\lambda)$. There are alternate explanations for both of these assumptions (e.g. Cael and Boss 2017; Catalá et al. 2016); however, there is biogeochemical significance in S_{dg}, while phytoplankton would presumably produce the largest deviation from an exponential signal as observable from satellite ocean color data. Beyond these basic assumptions, we also considered the relationship between $a_{dg}(\lambda)$ and $a_{ph}(\lambda)$ within a theoretical framework (Fig. 9). Based on this framework, it is important to recognize how varying contributions of each component will inherently lead to specific biases. For example, we noted that estimating $S_{350:400}$ did not improve algorithm performance despite the recognition that $a_{ph}(\lambda)$ typically has a limited role in total absorption at these wavelengths even when it is a dominant contributor at 440 nm. This is due to a very strong influence of $a_{ph}(\lambda)$ on absorption at 400 nm relative to 350 nm for most $a_{ph}(\lambda)$; since an exponential model is a relatively basic, least squares fit this sudden bias at one end of the spectral range decreases S_{dg} estimates. This is also why we increased estimates of S_{dg} first, then alternated to decreasing S_{dg}, as an exponential fit of $a_{t\text{-}w}$ will produce lower S values than that observed in S_{dg} when $a_{ph}(\lambda)$ is a contributor, particularly when $a_{ph}(440) > 10\%$. Finding where the second derivative of $a_{t\text{-}}$

$_w(\lambda)$ equals 0 and fitting an exponential at these points minimizes this impact (essentially "cutting through" primary pigment features for a least squares fit); however, there was a consistent bias towards lower S_{dg} values as $a_{ph}(440)$ contribution increased, as expected. The general framework illustrated in Fig. 9 is also the justification for setting a ratio of 1.5 to $a_{ph}(350)$:$a_{ph}(440)$; when the residual used to estimate $a_{ph}(\lambda)$ had a ratio higher than this, it was almost always indicative of a significant portion of the $a_{dg}(\lambda)$ signal remaining in the residual.

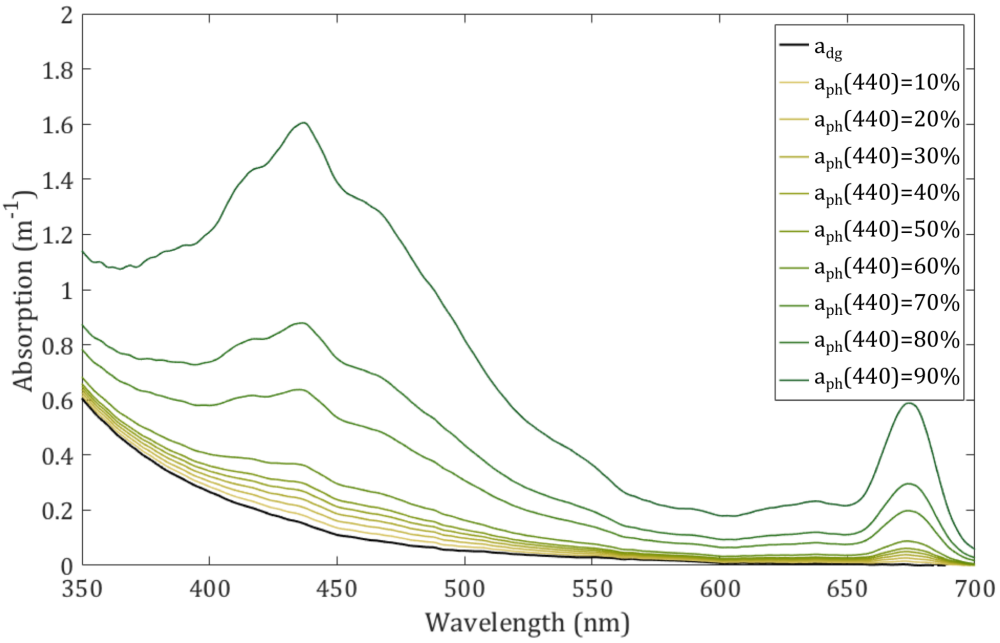


Figure 9. A theoretical representation of varying spectral shape of $a_{t\text{-}w}(\lambda)$ under varying contributions of $a_{dg}(\lambda)$ and $a_{ph}(\lambda)$. The base $a_{dg}(\lambda)$ and $a_{ph}(\lambda)$ spectra used for each curve are taken from measured spectra. We utilized this theoretical framework to develop the algorithm, namely understanding how changes in $a_{ph}(\lambda)$ percent contribution will inherently impact estimates of S_{dg}, how this inherent bias is impacted by wavelengths used and how to assess whether $a_{dg}(\lambda)$ has been accurately retrieved from $a_{t\text{-}w}(\lambda)$ free of an empirical relationship.

In short of independent variables to validate each component of interest, some explicit assumptions are required within any algorithm framework. Here, we chose to limit our solutions by constraining initial $a_{dg}(440)$ estimates by the empirical relationship between $a_{t\text{-}w}(555)/a_{t\text{-}w}(680)$ from the training dataset (Fig. 10a) and a theoretical ratio of 1.5 for $a_{residual}(350)/a_{residual}(440)$ (Eq. 7) to determine whether the contribution of $a_{dg}(\lambda)$ to $a_{t\text{-}w}(\lambda)$ from 350-400 nm had been reasonably estimated. These relationships do not explicitly dictate the final product, but guide the algorithm to reasonable estimates, at which point fitting is not constrained by these specific values. They do, however, leave an impact on how results are constrained. As we discussed previously, empirical relationships can often fall short of their intended accuracy. Despite a similar optical and geographical distribution between the training and test datasets (Fig. 1), the piece-wise exponential relationship derived from the training dataset to predict $a_{ph}(440)/a_{dg}(440)$ contribution to

$a_{t\text{-}w}(440)$ (r^2=0.91, RMSE=0.068) did not predict the same relationship nearly as well for the test dataset (r^2=0.58, RMSE=0.110; Fig. 10b). Additionally, this relationship for the training dataset is not perfect, likely due to differences in phytoplankton community composition, affiliated pigment packaging and spectral variability in $a_{dg}(\lambda)$ affiliated with varying CDOM composition and NAP particle composition (Babin et al. 2003; Bricaud et al. 1995; Wünsch et al. 2018). We did adjust the theoretical value of 1.5 but algorithm results were not very sensitive to changes in this parameter to values less than 1.5. This is primarily due to the fact that a significant exponential signal can remain in the $a_{ph}(\lambda)$ estimate used in the Gaussian decomposition of $a_{t\text{-}w}(\lambda)$ and the signal will be removed due

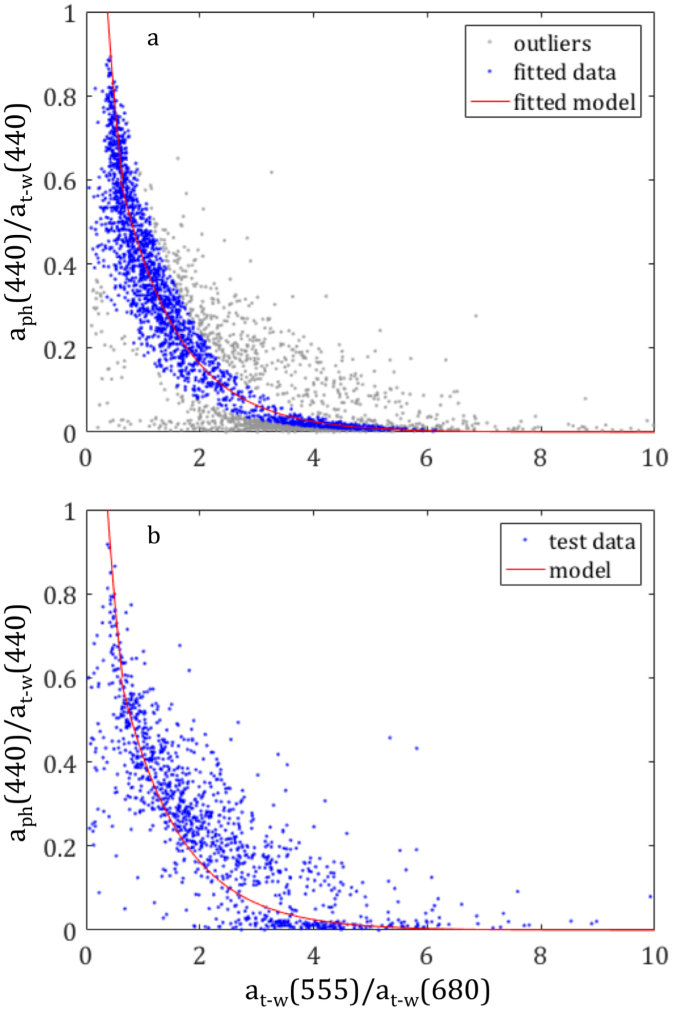


Figure 10. Relationship between $a_{t\text{-}w}(555)/a_{t\text{-}w}(680)$ and $a_{ph}(440)$ contribution for the (a) training dataset, where the piecewise exponential relationship used in the algorithm is represented by the red line, blue points indicate fitted data and gray points indicate values excluded from model fitting (r^2=0.91, RMSE=0.068). Outliers were defined as $1.5 \cdot 1^{st}/3^{rd}$ quartile and were used to remove the influence of the large spread in data points where $a_{ph}(440)$ contributed less than 10%, as these points represented nearly 25% of the dataset. (b) Test dataset points relative to the piecewise exponential relationship derived from the training dataset, displaying the primary weakness in empirical relationships (r^2=0.58, RMSE=0.110).

to a much better fit by incorporating this signal into an exponential model rather than attempting to fit with a Gaussian curve. However, this step is crucial as it: 1) further constrains S_{dg} to a more accurate solution and 2) inclusion of a Gaussian component in the $a_{ph}(\lambda)$ estimate requires fitting of this component within our scheme; no Gaussian components are dropped, as this would not provide for consistent and stable solutions. Thus, estimating $a_{dg}(\lambda)$ to the best ability results in fewer erroneous assignments of Gaussian components. Adjusting this value for a more regionally reflective value is quite easy within the provided code (Section 2.2.2). We maintained a value of 1.5 to account for some spectra that contain pigments below 400 nm (e.g. mycosporine-like amino acids) while maintaining strong performance in spectra without these pigments.

3.4.2 CDOM

Previous studies have shown that proper characterization of S_{dg} variability within a regional system accounts for an equivalent improvement in algorithm performance relative to optimizing multiple variables simultaneously, namely chlorophyll-normalized $a_{ph}(\lambda)$, NAP particle size distribution and S_{dg} (Organelli et al. 2016). For many systems, the bulk of S_{dg} variability is driven by S_g due to a relatively limited range of S_d values (typically 0.06-0.015 nm^{-1}) and a low contribution of $a_d(\lambda)$ to the total $a_{dg}(\lambda)$ signal (Babin et al. 2003; Nelson et al. 1998). S_g, in contrast, can vary from 0.01-0.03 nm^{-1} for most systems, with variability up to 0.05 nm^{-1} depending on the spectral range used to calculate S_g (Grunert et al. 2018). As seen in our results, properly characterizing S_g when $a_d(\lambda) << a_g(\lambda)$, and S_{dg} when $a_d(\lambda) \approx a_g(\lambda)$ is crucial to accurately estimating $a_{ph}(\lambda)$ spectral shape.

Beyond accurate retrievals of $a_{ph}(\lambda)$, variability in S_g can be used to characterize CDOM source and relative degradation state while also displaying a relatively consistent relationship with bulk molecular weight of the CDOM pool (Danhiez et al. 2017; Helms et al. 2013; Helms et al. 2008; Wünsch et al. 2018). Typically, lower S_g values indicate terrestrial material or material with a higher humic content and molecular weight (< ~0.015 nm^{-1}) while marine-derived material typically displays a S_g that varies from 0.015-0.03 nm^{-1}, depending on source and degradation state (Granskog 2012; Grunert et al. 2018; Wünsch et al. 2018). Narrow range S_g displays strong predictive power when considering molecular weight (e.g. $S_{275:295}$; Helms et al. 2008); recent consideration of broad wavelength S_g (e.g. $S_{300:600}$) shows these metrics have similar trends with CDOM molecular weight and degradation state, at least in observed samples (Stedmon and Nelson 2014; Wünsch et al. 2018). Across a variety of spectral ranges, mean S_g has been observed to range from 0.014-0.048 nm^{-1} ($S_{275:295}$) to 0.014-0.023 nm^{-1} ($S_{350:400}$) across inland waters and Longhurst biogeochemical provinces (Grunert et al. 2018; Longhurst 2006). Additionally, consideration of samples along river transects have shown changes in S_g varying from ~0.015-0.022 nm^{-1} and CDOM of different bulk molecular weight sampled from four ecologically distinct lakes displayed differences in S_g varying from ~0.015-0.021 nm^{-1} for $S_{350:400}$ and $S_{300:600}$ metrics (Helms et al. 2008; Wunsch et al. 2018). DAISEA estimated S_{dg} from 350-700 nm at a resolution relevant to these S_g ranges (Fig. 6; Table 1). While not identical metrics, this suggests that further separation of this metric into S_d and S_g can still

provide estimates of S_g at a resolution that carries confidence for estimating biogeochemical variability estimated by S_g at a global scale.

Across strong gradients in $a_g(\lambda)$ magnitude (e.g. coastal to open ocean transects), $a_g(\lambda)$, S_g and [DOC] strongly covary (Fichot et al. 2013; Kowalczuk et al. 2010). This has resulted in many empirical and semi-analytical approaches that link $a_g(\lambda)$ at a reference wavelength to rates of change in S_g and [DOC] (Fichot and Benner 2011; Fichot et al. 2014; Mannino et al. 2014; Mannino et al. 2008; Matsuoka et al. 2013; Vantrepotte et al. 2015). The utility of these relationships is the fact that S_g is relatively stable for high $a_g(\lambda)$ values, resulting in coherent variability across S_g calculated at various spectral ranges and $a_g(\lambda)$ at visible wavelengths (Babin et al. 2003; Grunert et al. 2018). However, the relationship between $a_g(\lambda)$ at visible wavelengths (e.g. 412 or 443 nm) and S_g and [DOC] does not generally hold up in most ocean environments, particularly the open ocean, where marine-produced CDOM has a different spectral signature, is produced at much lower rates and where degradation processes can dominate cycling for extended periods (Green and Blough 1994; Grunert et al. 2018; Meler et al. 2016; Nelson et al. 2010).

Retrieval of $a_g(\lambda)$ from ocean color remote sensing has been relevant to understanding distribution and spectral quality of light in the water column, magnitude of CDOM related to net production or degradation and ocean heating (Kim et al. 2016; Kim et al. 2015; Nelson and Siegel 2013; Swan et al. 2013). The de-coupling of $a_g(\lambda)$ and S_g in the open ocean has limited the role of $a_g(\lambda)$ in assessing more specific biogeochemical cycling of CDOM, including source, relative composition (e.g. low or high molecular weight material) and degradation state, all of which provide insight into ecosystem functioning and carbon cycling (Helms et al. 2013; Helms et al. 2008; Wünsch et al. 2018). Where $a_g(\lambda)$ may not vary significantly, S_g, fluorescence metrics and molecular characterization of CDOM and DOM often display significant spatial variability, suggesting a strong role for more nuanced optical characterizations of CDOM using absorption and fluorescence features (Kaiser and Benner 2009; Medeiros et al. 2015; Mentges et al. 2017; Nelson and Gauglitz 2016; Zhao et al. 2017). Recent work specifically considering marine-produced CDOM is also beginning to address causes for spectral variability across different wavelength ranges with trends linked to environmental variability including phytoplankton bloom state and nutrient stoichiometry (Asmala et al. 2018; Danhiez et al. 2017). These findings agree well with past studies that have found CDOM to peak approximately 1 month after the peak phytoplankton bloom, suggesting a role for phytoplankton degradation and microbial and zooplankton CDOM production at particular stages of a phytoplankton bloom (Kinsey et al. 2018; Organelli et al. 2014; Ortega-Retuerta et al. 2009; Rochelle-Newall and Fisher 2002; Romera-Castillo et al. 2011). This work is necessary to elucidate the cause of non-covarying differences in S_g estimated for different spectral ranges that can significantly impact our ability to accurately relate $a_g(\lambda)$ and S_g to biogeochemical variability under different environmental conditions.

The bulk of CDOM composition variability directly related to optical variability in S_g has been linked to spectral ranges either entirely or partially unavailable from the

proposed spectral capability of PACE (e.g. $S_{275:295}$). Some empirical or multivariate statistical techniques have been proposed to retrieve these parameters, as well as semi-analytical frameworks contingent on an empirical relationship to define S_g of biogeochemical interest (Cao and Miller 2015; Fichot et al. 2013; Mannino et al. 2014; Matsuoka et al. 2013); however, spectral variability at visible wavelengths is often disconnected from variability at wavelengths <400 nm due to very different sets of chromophores impacting each spectral range (Asmala et al. 2018; Danhiez et al. 2017). This suggests that relationships indirectly estimating $S_{275:295}$ will have limited efficacy unless environmental conditions are well-defined. From this, we suggest a renewed focus on relating spectral variability from directly viewable wavelengths (e.g. 350-400 nm) to CDOM composition. A common issue when decomposing $a_{t\text{-}w}$ spectra dominated by $a_{dg}(\lambda)$ was overestimation of $a_{ph}(\lambda)$ from 350-400 nm. While some of this error comes from the algorithm scheme, it should also be noted that $S_{350:400}$ tends to be higher across much of the spectra considered here. Spectral variability in this wavelength range is relatively high (Helms et al. 2008) and displays unique variability in the global ocean (Grunert et al. 2018). It is likely that better characterizing variability in this portion of the spectrum will improve estimation of both $a_{dg}(\lambda)$ and $a_{ph}(\lambda)$, as well as providing a metric carrying information about unique biogeochemical processes. Variability at wavelengths < 300 nm is largely attributed to terrestrial material and its degradation, while 300-400 nm wavelengths are subject to variability from both terrestrial and marine-sourced material (Danhiez et al. 2017; Helms et al. 2008). This suggests that S_g at wavelengths available from PACE can provide information on microbial carbon cycling. We show here that relatively accurate retrievals of S_{dg} are possible from hyperspectral absorption data, free of explicit assumptions, and are theoretically possible from satellite-retrieved $R_{rs}(\lambda)$. As we move forward, hyperspectral ocean color remote sensing will benefit from a common set of S_g metrics that relate to specific biogeochemical processes.

3.4.3 Phytoplankton

Phytoplankton community composition, grouped into functional types including taxonomic class, cell size, and particle size distribution have been successfully retrieved using multispectral satellite imagery with a variety of approaches (Mouw et al. 2017 and references therein). While proof of concept has been demonstrated for phytoplankton functional types, validation metrics vary (Mouw et al. 2017) and there are inconsistencies in magnitude, temporal and spatial variability between existing retrievals (Kostadinov et al. 2017). Greater spectral information with forthcoming missions is anticipated to improve and expand phytoplankton functional type retrievals (Isada et al. 2015; Mouw et al. 2017; Wolanin et al. 2016). Hyperspectral satellite imagery provides new avenues for identifying phytoplankton community composition and physiology through use of identified pigments (Bracher et al. 2017; Uitz et al. 2015), allowing for more accurate estimates of primary productivity, light and nutrient limitation and phytoplankton carbon content (Behrenfeld et al. 2015; Graff et al. 2016; Robinson et al. 2017; Roy et al. 2017; Westberry et al. 2008). Historically, Chl has been the primary variable of interest from satellite sensors, dating back to the Coastal Zone and Color Scanner (CZCS) proof-of-

concept mission. This is largely due to the ability to constrain ecosystem productivity from Chl (e.g. Behrenfeld and Falkowski 1997). Interest in the global carbon cycle, however, has increased interest in variables beyond Chl as phytoplankton carbon does not directly covary with Chl but is rather dependent on environmental and physiological factors (Behrenfeld et al. 2015; Burt et al. 2018; Roy et al. 2017). Additionally, the introduction of more accurate net production estimates are possible using IOP retrievals (Robinson et al. 2017; Silsbe et al. 2016).

As with Chl algorithms, accurately retrieving IOP's in productive and complex waters is quite difficult and a significant motivator for launching hyperspectral capabilities on a satellite platform (Bricaud et al. 2012; Mouw et al. 2015; Odermatt et al. 2012). Significant work in Gaussian decomposition of hyperspectral $a_p(\lambda)$ and $a_{ph}(\lambda)$ has been shown to yield good closure between modeled and measured $a_p(\lambda)/a_{ph}(\lambda)$ while also providing strong relationships between some phytoplankton pigments and fitted Gaussian parameters (Chase et al. 2013; Chase et al. 2017; Wang et al. 2016; Wang et al. 2017). Identifying phytoplankton pigments in a complex signal typically relies on derivative analysis provided noisy data can be adequately smoothed (Tsai and Philpot 1998; Lee et al. 2007). While previous methods focusing on hyperspectral capabilities have assumed an $S_{d/dg}$ value, we show the potential to retrieve S_{dg}, $a_{ph}(\lambda)$ and $a_{dg}(\lambda)$ with a high level of accuracy across a range of optical conditions. Accurately retrieving spectral shape of both $a_{ph}(\lambda)$ and $a_{dg}(\lambda)$ is critical for accurately estimating both parameters, while unbiased retrieval allows for utilizing both parameters to estimate biogeochemical variables free of an assumption of covariance that is often inaccurate for optically complex waters (e.g. Babin et al. 2003).

We considered the locations of fitted Gaussian components for observed $a_{ph}(\lambda)$ and parameters used to estimate $a_{ph}(\lambda)$ components within the algorithm; after removing $a_{dg}(\lambda)$, Gaussian component location was remarkably consistent with components fitted on observed $a_{ph}(\lambda)$ through the same methodology, albeit with fewer fitted peaks overall (Fig. 8). This is likely due to pigments centered at wavelengths < 400 nm, where initial estimates of $a_{ph}(\lambda)$ were typically smoother than observed, but final estimates more often captured spectral features. We did not utilize Gaussian decomposition to estimate the final $a_{ph}(\lambda)$ output, as we found that a smoothed residual more accurately represented observed $a_{ph}(\lambda)$ than $a_{ph}(\lambda)$ modeled from the Gaussian components. This could be due to fitting fewer Gaussian components than needed to accurately model $a_{ph}(\lambda)$ or constraints based on poor initial Gaussian curve parameters; our methodology did fit fewer peaks than alternate Gaussian decomposition schemes due to a difference in methodologies (Hoeppfner and Sathyendranath 1993; Chase et al. 2013). These algorithms typically identify pigments from first and second derivative analysis of an existing database of phytoplankton spectra then assign windows around these points (typically 12 peaks). Our approach focuses on identifying primary pigment features to best fit observed $a_{t\text{-}w}(\lambda)$ without assuming the locations of pigments, resulting in fewer identified peaks (~6 peaks). There is potential to increase the sensitivity of the peak finding step and to not filter out identified peaks that are small. Our focus was on retrieving $a_{dg}(\lambda)$ and $a_{ph}(\lambda)$ accurately, including spectral

shape, rather than individually parameterizing phytoplankton pigments. It is possible to utilize the $a_{ph}(\lambda)$ output in a separate Gaussian decomposition scheme, or other approach that identifies phytoplankton pigments. However, it should be noted that derivative analysis of the final $a_{ph}(\lambda)$ output, even after smoothing, resulted in more identified peaks than the observed $a_{ph}(\lambda)$ using our scheme. This is very likely due to the inclusion of chromophores in $a_g(\lambda)$ and $a_d(\lambda)$ that result in deviations from the typical exponential expression used to model these parameters, features visibly apparent in many of the $a_{dg}(\lambda)$ spectra. While often overlooked, these features have been recognized for some time (Babin et al. 2003; Schwarz et al. 2002) and a recent methodology for fitting these features provides a means of both quantifying them and more accurately modeling the underlying exponential signal (Catalá et al. 2016; Massicotte and Markager 2016; Grunert et al. 2018). This approach is useful for *in situ* data, but not practical for our proposed methodology and likely a non-factor when considering $a_{t\text{-}w}(\lambda)$ derived from satellite $R_{rs}(\lambda)$.

3.4.4 NAP

Of the IOP's we focus on here, understanding spectral variability in $a_d(\lambda)$ is perhaps the most ambiguous. NAP absorption typically follows an exponential increase with decreasing wavelength, similar to $a_g(\lambda)$, but within a much narrower range of spectral slope values across diverse environments (Babin et al. 2003; Neukermans et al. 2016). S_d has been hypothesized to vary with minerogenic and organic content, with some studies showing differences in mean S_d supporting this theory (Babin and Stramski 2004; Babin et al. 2003; Estapa et al. 2012). Additionally, recent work has shown differences in S_d across $a_d(\lambda)$ dominated by organic matter of different origin or degradation state, e.g. NAP affiliated with a phytoplankton bloom (Neukermans et al. 2016). Beyond broad spectral shape, $a_d(\lambda)$ spectra often display deviations from an exponential model. These features are sometimes associated with phytoplankton pigments remaining after the bleaching process but are often spectrally consistent with iron compounds (e.g. organic-complexed iron, iron oxides; Babin et al. 2003; Babin and Stramski 2004; Estapa et al. 2012). Recent approaches have also displayed optically consistent features between CDOM and base-extracted particulate organic matter (BEPOM), effectively CDOM extracted from phytoplankton and NAP (Brym et al. 2014; Osburn et al. 2015). CDOM and BEPOM collected from the same environment shows remarkable consistency in both absorption spectra and fluorescence features, suggesting that deviations from the exponential in $a_d(\lambda)$ are affiliated with unique chromophores or groups of chromophores also present in $a_g(\lambda)$. As with $a_g(\lambda)$, these deviations can be modeled through a Gaussian decomposition scheme following the approach of Massicotte and Markager (2016). However, an understanding of what these features are is even more limited than for $a_g(\lambda)$.

Separating $a_d(\lambda)$ from $a_{dg}(\lambda)$ relies on empirical relationships (Dong et al. 2013). With current multispectral limitations and algorithm capabilities, this will likely remain a limitation. However, work linking scattering and absorption properties of particulate material with its composition (minerogenic vs. organic), size and refractive index has shown promising capability to further constrain $a_d(\lambda)$ and $a_{ph}(\lambda)$ with hyperspectral

measurements by considering relationships between $b_{bp}(\lambda)/a_p(\lambda)$ (Neukermans et al. 2016). Consideration of lidar and independent $b_{bp}(\lambda)$ measurements provides a powerful means of characterizing particulate material including its refractive index (Behrenfeld et al. 2013; Boss et al. 2004; Reynolds et al. 2016; Zhang et al. 2014), with S_d appearing to show some relationship between minerogenic and organic content, even organic content presumably of variable degradation state (Babin et al. 2003; Estapa et al. 2012; Neukermans et al. 2016). Relationships between $b_{bp}(\lambda)$ and $a_p(\lambda)$ have also indicated various bloom stages, suggesting a further ability to constrain IOPs by considering independent measurements and relationships between IOPs (Neukermans et al. 2014). Currently, a robust relationship between particle size, composition and refractive index, as characterized with $b_{bp}(\lambda)$ and the slope of the $b_{bp}(\lambda)$ spectra, and S_d is not evident from the literature. This, and other possible means of differentiating S_d and S_g free of the restrictions of an empirical relationship, are crucial for advancing our understanding of ocean biogeochemistry as afforded through accurate retrievals of S_g. This work would also likely lend insight into the biogeochemical significance of S_d.

3.5 Conclusions

We show that across most water types considered, DAISEA can accurately estimate $a_{dg}(\lambda)$, S_{dg} and $a_{ph}(\lambda)$ magnitude and spectral features for all water types where $a_{ph}(\lambda)$ contributes at moderate (>10%) levels at 440 nm. We parameterized the percent of $a_{dg}(\lambda)$ and $a_{ph}(\lambda)$ estimates that were retrievable by comparing the error in one metric to the observed signal of the other. Consistent with the general accuracy of DAISEA, primary features (i.e. chlorophyll-a absorption peaks) of $a_{ph}(\lambda)$ were retrievable for greater than 80% of spectra across environments where $a_{ph}(440) > 10\%$; $a_{dg}(\lambda)$ was retrievable for at least 80% of spectra from 350-650 nm when $a_{ph}(440) < 70\%$. NRMSE metrics suggest strong algorithm performance across most optical variability from 350-650 nm. Algorithm bias shows a tendency to overestimate $a_{ph}(\lambda)$ when $a_{ph}(440) < 40\%$ and to underestimate $a_{ph}(\lambda)$ when $a_{ph}(440) > 60\%$.

Despite schemes for separation of $a_{dg}(\lambda)$ and S_{dg} into the component parts (NAP and CDOM; e.g. Dong et al. 2013), we did not pursue separation here. Currently, coincident hyperspectral measurements of $R_{rs}(\lambda)$, $b_{bp}(\lambda)$, $a_{ph}(\lambda)$, $a_d(\lambda)$ and $a_g(\lambda)$ down to a minimum wavelength of 350 nm, the proposed lower wavelength limit of PACE, are quite uncommon relative to coincident measurements at wavelengths ≥ 400 nm. We anticipate that hyperspectral inversion methods and coincident lidar measurements will enable a detailed analysis of particulate absorption that will lend insights into $a_d(\lambda)$ magnitude and NAP particle composition, with further insight into S_d. It is very likely that this will be a more robust means of individually characterizing S_g and S_d and should be the focus of future research efforts. Considering current algorithm performance, we anticipate that a well-performing scheme to separate S_{dg} into its component parts will allow for appropriate resolution in S_g to estimate source and degradation state of CDOM in the surface ocean.

3.6 References

Andrew, A.A., Del Vecchio, R., Subramaniam, A., & Blough, N.V. (2013). Chromophoric dissolved organic matter (CDOM) in the Equatorial Atlantic Ocean: Optical properties and their relation to CDOM structure and source. *Marine Chemistry, 148*, 33-43

Asmala, E., Haraguchi, L., Jakobsen, H.H., Massicotte, P., & Carstensen, J. (2018). Nutrient availability as major driver of phytoplankton-derived dissolved organic matter transformation in coastal environment. *Biogeochemistry, 137*, 93-104

Babin, M., & Stramski, D. (2004). Variations in the mass-specific absorption coefficient of mineral particles suspended in water. *Limnology & Oceanography, 49*, 756-767

Babin, M., Stramski, D., Ferrari, G.M., Claustre, H., Bricaud, A., Obolensky, G., & Hoepffner, N. (2003). Variations in the light absorption coefficients of phytoplankton, nonalgal particles, and dissolved organic matter in coastal waters around Europe. *Journal of Geophysical Research, 108*

Behrenfeld, M., & Falkowski, P.G. (1997). Photosynthetic rates derived from satellite-based chlorophyll concentration. *Limnology & Oceanography, 42*, 1-20

Behrenfeld, M.J., Hu, Y., Hostetler, C.A., Dall'Olmo, G., Rodier, S.D., Hair, J.W., & Trepte, C.R. (2013). Space-based lidar measurements of global ocean carbon stocks. *Geophysical Research Letters, 40*, 4355-4360

Behrenfeld, M.J., Hu, Y., O'Malley, R.T., Boss, E.S., Hostetler, C.A., Siegel, D.A., Sarmiento, J.L., Schulien, J., Hair, J.W., Lu, X., Rodier, S., & Scarino, A.J. (2016). Annual boom–bust cycles of polar phytoplankton biomass revealed by space-based lidar. *Nature Geoscience, 10*, 118-122

Behrenfeld, M.J., & Milligan, A.J. (2013). Photophysiological expressions of iron stress in phytoplankton. *Ann Rev Mar Sci, 5*, 217-246

Behrenfeld, M.J., O'Malley, R.T., Boss, E.S., Westberry, T.K., Graff, J.R., Halsey, K.H., Milligan, A.J., Siegel, D.A., & Brown, M.B. (2015). Revaluating ocean warming impacts on global phytoplankton. *Nature Climate Change, 6*, 323-330

Bishop, J.K. (1999). Transmissometer measurement of POC. *Deep Sea Research Part I: Oceanographic Research Papers, 46*, 353-369

Boss, E., Pegau, W.S., Lee, M., Twardowski, M.S., Shybanov, E., Korotaev, G., & Baratange, F. (2004). Particulate backscattering ratio at LEO 15 and its use to study particle composition and distribution. *Journal of Geophysical Research, 109*

Bracher, A., Bouman, H.A., Brewin, R.J.W., Bricaud, A., Brotas, V., Ciotti, A.M., Clementson, L., Devred, E., Di Cicco, A., Dutkiewicz, S., Hardman-Mountford, N.J., Hickman, A.E., Hieronymi, M., Hirata, T., Losa, S.N., Mouw, C.B., Organelli, E., Raitsos, D.E., Uitz, J., Vogt, M., & Wolanin, A. (2017). Obtaining Phytoplankton Diversity from Ocean Color: A Scientific Roadmap for Future Development. *Frontiers in Marine Science, 4*

Bracher, A., Vountas, M., Dinter, T., Burrows, J.P., Rüttgers, R., & Peeken, I. (2009). Quantitative observation of cyanobacteria and diatoms from space using PhytoDOAS on SCIAMACHY data. *Biogeosciences, 6*, 751-764

Brando, V., Dekker, A., Park, Y.J., & Schroeder, T. (2012). Adaptive semianalytical inversion of ocean color radiometry in optically complex waters. *Appl Opt, 51*

Bricaud, A., Babin, M., Morel, A., & Claustre, H. (1995). Variability in the chlorophyll-specific absorption coefficients of natural phytoplankton: Analysis and parameterization. *Journal of Geophysical Research, 100*, 13,321-313,332

Bricaud, A., Ciotti, A.M., & Gentili, B. (2012). Spatial-temporal variations in phytoplankton size and colored detrital matter absorption at global and regional scales, as derived from twelve years of SeaWiFS data (1998-2009). *Global Biogeochemical Cycles, 26*, n/a-n/a

Bricaud, A., & Morel, A. (1986). Light attenuation and scattering by phytoplanktonic cells: a theoretical modeling. *Appl Opt, 25*, 571-580

Bricaud, A., Morel, A., & Prieur, L. (1983). Optical efficiency factors of some phytoplankters. *Limnology & Oceanography, 28*, 816-832

Brym, A., Paerl, H.W., Montgomery, M.T., Handsel, L.T., Ziervogel, K., & Osburn, C.L. (2014). Optical and chemical characterization of base-extracted particulate organic matter in coastal marine environments. *Marine Chemistry, 162*, 96-113

Burt, W.J., Westberry, T.K., Behrenfeld, M.J., Zeng, C., Izett, R.W., & Tortell, P.D. (2018). Carbon: Chlorophyll Ratios and Net Primary Productivity of Subarctic Pacific Surface Waters Derived From Autonomous Shipboard Sensors. *Global Biogeochemical Cycles, 32*, 267-288

Cael, B.B., & Boss, E. (2017). Simplified model of spectral absorption by non-algal particles and dissolved organic materials in aquatic environments. *Opt Express, 25*, 25486-25491

Cao, F., & Miller, W.L. (2015). A new algorithm to retrieve chromophoric dissolved organic matter (CDOM) absorption spectra in the UV from ocean color. *Journal of Geophysical Research: Oceans, 120*, 496-516

Catalá, T.S., Reche, I., Ramón, C.L., López-Sanz, À., Álvarez, M., Calvo, E., & Álvarez-Salgado, X.A. (2016). Chromophoric signatures of microbial by-products in the dark ocean. *Geophysical Research Letters, 43*, 7639-7648

Chang, G.C., & Dickey, T.D. (2004). Coastal ocean optical influences on solar transmission and radiant heating rate. *Journal of Geophysical Research, 109*

Chase, A., Boss, E., Zaneveld, R., Bricaud, A., Claustre, H., Ras, J., Dall'Olmo, G., & Westberry, T.K. (2013). Decomposition of in situ particulate absorption spectra. *Methods in Oceanography, 7*, 110-124

Chase, A.P., Boss, E., Cetinić, I., & Slade, W. (2017). Estimation of Phytoplankton Accessory Pigments From Hyperspectral Reflectance Spectra: Toward a Global Algorithm. *Journal of Geophysical Research: Oceans, 122*, 9725-9743

Ciotti, A.M., Lewis, M., & Cullen, J.J. (2002). Assessment of the relationship between dominant cell size in natural phytoplankton communities and the spectral shape of the absorption coefficient. *Limnology & Oceanography, 47*, 404-417

Cory, R.M., & Kling, G.W. (2018). Interactions between sunlight and microorganisms influence dissolved organic matter degradation along the aquatic continuum. *Limnology and Oceanography Letters, 3*, 102-116

D'Errico, J. (2005). Surface Fitting using gridfit. In. (https://www.mathworks.com/matlabcentral/fileexchange/8998-surface-fitting-using-gridfit): MATLAB Central File Exchange

Danhiez, F., Vantrepotte, V., Cauvin, A., Lebourg, E., & Loisel, H. (2017). Optical properties of chromophoric dissolved organic matter during a phytoplankton bloom. Implication for DOC estimates from CDOM absorption. *Limnology and Oceanography, 62*, 1409-1425

Dong, Q., Shang, S., & Lee, Z. (2013). An algorithm to retrieve absorption coefficient of chromophoric dissolved organic matter from ocean color. *Remote Sensing of Environment, 128*, 259-267

Dutkiewicz, S., Hickman, A.E., Jahn, O., Gregg, W.W., Mouw, C.B., & Follows, M.J. (2015). Capturing optically important constituents and properties in a marine biogeochemical and ecosystem model. *Biogeosciences, 12*, 4447-4481

Estapa, M.L., Boss, E., Mayer, L.M., & Roesler, C.S. (2012). Role of iron and organic carbon in mass-specific light absorption by particulate matter from Louisiana coastal waters. *Limnology and Oceanography, 57*, 97-112

Fichot, C.G., & Benner, R. (2011). A novel method to estimate DOC concentrations from CDOM absorption coefficients in coastal waters. *Geophysical Research Letters, 38*

Fichot, C.G., Benner, R., Kaiser, K., Shen, Y., Amon, R.M.W., Ogawa, H., & Lu, C.-J. (2016). Predicting Dissolved Lignin Phenol Concentrations in the Coastal Ocean from Chromophoric Dissolved Organic Matter (CDOM) Absorption Coefficients. *Frontiers in Marine Science, 3*

Fichot, C.G., Kaiser, K., Hooker, S.B., Amon, R.M., Babin, M., Belanger, S., Walker, S.A., & Benner, R. (2013). Pan-Arctic distributions of continental runoff in the Arctic Ocean. *Sci Rep, 3*, 1053

Fichot, C.G., Lohrenz, S.E., & Benner, R. (2014). Pulsed, cross-shelf export of terrigenous dissolved organic carbon to the Gulf of Mexico. *Journal of Geophysical Research: Oceans, 119*, 1176-1194

Gardner, W., Mishonov, A., & Richardson, M. (2006). Global POC concentrations from in-situ and satellite data. *Deep Sea Research Part II: Topical Studies in Oceanography, 53*, 718-740

Graff, J.R., Westberry, T.K., Milligan, A.J., Brown, M.B., Dall'Olmo, G., Reifel, K.M., & Behrenfeld, M.J. (2016). Photoacclimation of natural phytoplankton communities. *Marine Ecology Progress Series, 542*, 51-62

Granskog, M.A. (2012). Changes in spectral slopes of colored dissolved organic matter absorption with mixing and removal in a terrestrially dominated marine system (Hudson Bay, Canada). *Marine Chemistry, 134-135*, 10-17

Green, S.A., & Blough, N.V. (1994). Optical absorption and fluorescence properties of chromophoric dissolved organic matter in natural waters. *Limnology & Oceanography, 39*, 1903-1916

Grunert, B.K., Mouw, C.B., & Ciochetto, A.B. (2018). Characterizing CDOM Spectral Variability Across Diverse Regions and Spectral Ranges. *Global Biogeochemical Cycles, 32*, 57-77

Hansell, D.A., Carlson, C.A., Repeta, D.J., & Schlitzer, R. (2009). Dissolved organic matter in the ocean: A controversy stimulates new insights. *Oceanography, 22*, 202-211

Helms, J.R., Stubbins, A., Perdue, E.M., Green, N.W., Chen, H., & Mopper, K. (2013). Photochemical bleaching of oceanic dissolved organic matter and its effect on absorption spectral slope and fluorescence. *Marine Chemistry, 155*, 81-91

Helms, J.R., Stubbins, A., Ritchie, J.D., Minor, E.C., Kieber, D.J., & Mopper, K. (2008). Absorption spectral slopes and slope ratios as indicators of molecular weight, source, and photobleaching of chromophoric dissolved organic matter. *Limnology and Oceanography, 53*, 955-969

Hieronymi, M., Müller, D., & Doerffer, R. (2017). The OLCI Neural Network Swarm (ONNS): A Bio-Geo-Optical Algorithm for Open Ocean and Coastal Waters. *Frontiers in Marine Science, 4*

Hoepffner, N., & Sathyendranath, S. (1993). Determination of the major groups of phytoplankton pigments from the absorption spectra of total particulate matter. *Journal of Geophysical Research, 98*, 22789

Isada, T., Hirawake, T., Kobayashi, T., Nosaka, Y., Natsuike, M., Imai, I., Suzuki, K., & Saitoh, S.-I. (2015). Hyperspectral optical discrimination of phytoplankton community structure in Funka Bay and its implications for ocean color remote sensing of diatoms. *Remote Sensing of Environment, 159*, 134-151

Johnsen, G., Nelson, N., Jovine, R.V.M., & Prezelin, B. (1994). Chromoprotein- and pigment dependent modeling of spectral light absorption in two dinoflagellates, Prorocentrum minimum and Heterocapsa pygmaea. *Marine Ecology Progress Series, 114*, 245-258

Kaiser, K., & Benner, R. (2009). Biochemical composition and size distribution of organic matter at the Pacific and Atlantic time-series stations. *Marine Chemistry, 113*, 63-77

Kim, G.E., Gnanadesikan, A., & Pradal, M.-A. (2016). Increased Surface Ocean Heating by Colored Detrital Matter (CDM) Linked to Greater Northern Hemisphere Ice Formation in the GFDL CM2Mc ESM. *Journal of Climate, 29*, 9063-9076

Kim, G.E., Pradal, M.A., & Gnanadesikan, A. (2015). Quantifying the biological impact of surface ocean light attenuation by colored detrital matter in an ESM using a new optical parameterization. *Biogeosciences, 12*, 5119-5132

Kinsey, J.D., Corradino, G., Ziervogel, K., Schnetzer, A., & Osburn, C.L. (2018). Formation of Chromophoric Dissolved Organic Matter by Bacterial Degradation of Phytoplankton-Derived Aggregates. *Frontiers in Marine Science, 4*

Kostadinov, T.S., Cabré, A., Vedantham, H., Marinov, I., Bracher, A., Brewin, R.J.W., Bricaud, A., Hirata, T., Hirawake, T., Hardman-Mountford, N.J., Mouw, C., Roy, S., & Uitz, J. (2017). Inter-comparison of phytoplankton functional type phenology metrics derived from ocean color algorithms and Earth System Models. *Remote Sensing of Environment, 190*, 162-177

Kowalczuk, P., Cooper, W.J., Durako, M.J., Kahn, A.E., Gonsior, M., & Young, H. (2010). Characterization of dissolved organic matter fluorescence in the South Atlantic Bight with use of PARAFAC model: Relationships between fluorescence and its components, absorption coefficients and organic carbon concentrations. *Marine Chemistry, 118*, 22-36

Lee, Z., Carder, K.L., Arnone, R., & He, M. (2007). Determination of primary spectral bands for remote sensing of aquatic environments. *Sensors, 7*, 3428-3441

Lee, Z., Carder, K.L., & Arnone, R.A. (2002). Deriving inherent optical properties from water color: a multiband quasi-analytical algorithm for optically deep waters. *Appl Opt, 41*, 5755-5772

Longhurst, A.R. (2006). *Ecological Geography of the Sea*. (2nd ed.). San Diego, CA: Academic Press

Mannino, A., Novak, M.G., Hooker, S.B., Hyde, K., & Aurin, D. (2014). Algorithm development and validation of CDOM properties for estuarine and continental shelf waters along the northeastern U.S. coast. *Remote Sensing of Environment, 152*, 576-602

Mannino, A., Russ, M.E., & Hooker, S.B. (2008). Algorithm development and validation for satellite-derived distributions of DOC and CDOM in the U.S. Middle Atlantic Bight. *Journal of Geophysical Research, 113*

Maritorena, S., Siegel, D.A., & Peterson, A.R. (2002). Optimization of a semianalytical ocean color model for global-scale applications. *Appl Opt, 41*, 2705-2714

Massicotte, P., & Markager, S. (2016). Using a Gaussian decomposition approach to model absorption spectra of chromophoric dissolved organic matter. *Marine Chemistry, 180*, 24-32

Matsuoka, A., Hooker, S.B., Bricaud, A., Gentili, B., & Babin, M. (2013). Estimating absorption coefficients of colored dissolved organic matter (CDOM) using a semi-analytical algorithm for southern Beaufort Sea waters: application to deriving concentrations of dissolved organic carbon from space. *Biogeosciences, 10*, 917-927

Medeiros, P.M., Seidel, M., Powers, L.C., Dittmar, T., Hansell, D.A., & Miller, W.L. (2015). Dissolved organic matter composition and photochemical transformations in the northern North Pacific Ocean. *Geophysical Research Letters, 42*, 863-870

Meler, J., Kowalczuk, P., Ostrowska, M., Ficek, D., Zabłocka, M., & Zdun, A. (2016). Parameterization of the light absorption properties of chromophoric dissolved organic matter in the Baltic Sea and Pomeranian lakes. *Ocean Science, 12*, 1013-1032

Mélin, F., & Vantrepotte, V. (2015). How optically diverse is the coastal ocean? *Remote Sensing of Environment, 160*, 235-251

Mentges, A., Feenders, C., Seibt, M., Blasius, B., & Dittmar, T. (2017). Functional Molecular Diversity of Marine Dissolved Organic Matter Is Reduced during Degradation. *Frontiers in Marine Science, 4*, 194

Morel, A., & Prieur, L. (1977). Analysis of variations in ocean color. *Limnology & Oceanography, 22*, 709-722

Mouw, C.B., Greb, S., Aurin, D., DiGiacomo, P.M., Lee, Z., Twardowski, M., Binding, C., Hu, C., Ma, R., Moore, T., Moses, W., & Craig, S.E. (2015). Aquatic color radiometry remote sensing of coastal and inland waters: Challenges and recommendations for future satellite missions. *Remote Sensing of Environment, 160*, 15-30

Mouw, C.B., Hardman-Mountford, N.J., Alvain, S., Bracher, A., Brewin, R.J.W., Bricaud, A., Ciotti, A.M., Devred, E., Fujiwara, A., Hirata, T., Hirawake, T., Kostadinov, T.S., Roy, S., & Uitz, J. (2017). A Consumer's Guide to Satellite Remote Sensing of Multiple Phytoplankton Groups in the Global Ocean. *Frontiers in Marine Science, 4*

Nelson, N.B., & Gauglitz, J.M. (2016). Optical signatures of dissolved organic matter transformation in the global ocean. *Frontiers in Marine Science, 2*, 118

Nelson, N.B., & Siegel, D.A. (2013). The global distribution and dynamics of chromophoric dissolved organic matter. *Ann Rev Mar Sci, 5*, 447-476

Nelson, N.B., Siegel, D.A., Carlson, C.A., & Swan, C.M. (2010). Tracing global biogeochemical cycles and meridional overturning circulation using chromophoric dissolved organic matter. *Geophysical Research Letters, 37*

Nelson, N.B., Siegel, D.A., & Michaels, A.F. (1998). Seasonal dynamics of colored dissolved material in the Sargasso Sea. *Deep-Sea Research I, 45*, 931-957

Neukermans, G., Reynolds, R.A., & Stramski, D. (2014). Contrasting inherent optical properties and particle characteristics between an under-ice phytoplankton bloom and open water in the Chukchi Sea. *Deep Sea Research Part II: Topical Studies in Oceanography, 105*, 59-73

Neukermans, G., Reynolds, R.A., & Stramski, D. (2016). Optical classification and characterization of marine particle assemblages within the western Arctic Ocean. *Limnology and Oceanography, 61*, 1472-1494

Odermatt, D., Gitelson, A., Brando, V.E., & Schaepman, M. (2012). Review of constituent retrieval in optically deep and complex waters from satellite imagery. *Remote Sensing of Environment, 118*, 116-126

Ogawa, H., Amagai, Y., Koike, I., Kaiser, K., & Benner, R. (2001). Production of refractory dissolved organic matter by bacteria. *Science, 292*, 917-920

Organelli, E., Bricaud, A., Antoine, D., & Matsuoka, A. (2014). Seasonal dynamics of light absorption by chromophoric dissolved organic matter (CDOM) in the NW Mediterranean Sea (BOUSSOLE site). *Deep Sea Research Part I: Oceanographic Research Papers, 91*, 72-85

Organelli, E., Bricaud, A., Gentili, B., Antoine, D., & Vellucci, V. (2016). Retrieval of Colored Detrital Matter (CDM) light absorption coefficients in the Mediterranean Sea using field and satellite ocean color radiometry: Evaluation of bio-optical inversion models. *Remote Sensing of Environment, 186*, 297-310

Ortega-Retuerta, E., Frazer, T., Duarte, C.M., Ruiz-Halpern, S., Tovar-Sanchez, A., Arrieta, J.M., & Reche, I. (2009). Biogeneration of chromophoric dissolved organic matter by bacteria and kirll in the Southern Ocea. *Limnology & Oceanography, 54*, 1941-1950

Osburn, C.L., Mikan, M.P., Etheridge, J.R., Burchell, M.R., & Birgand, F. (2015). Seasonal variation in the quality of dissolved and particulate organic matter exchanged between a salt marsh and its adjacent estuary. *Journal of Geophysical Research: Biogeosciences, 120*, 1430-1449

Reynolds, R.A., Stramski, D., & Neukermans, G. (2016). Optical backscattering by particles in Arctic seawater and relationships to particle mass concentration, size distribution, and bulk composition. *Limnology and Oceanography, 61*, 1869-1890

Riedel, T., & Dittmar, T. (2014). A method detection limit for the analysis of natural organic matter via Fourier transform ion cyclotron resonance mass spectrometry. *Analytical chemistry, 86*, 8376-8382

Robinson, C.M., Cherukuru, N., Hardman-Mountford, N.J., Everett, J.D., McLaughlin, M.J., Davies, K.P., Van Dongen-Vogels, V., Ralph, P.J., & Doblin, M.A. (2017). Phytoplankton absorption predicts patterns in primary productivity in Australian coastal shelf waters. *Estuarine, Coastal and Shelf Science, 192*, 1-16

Rochelle-Newall, E.J., & Fisher, T.R. (2002). Production of chromophoric dissolved organic matter fluorescence in marine and estuarine environments: an investigation into the role of phytoplankton. *Mar Chem, 77*, 7-21

Romera-Castillo, C., Sarmento, H., Alvarez-Salgado, X.A., Gasol, J.M., & Marrase, C. (2011). Net production and consumption of fluorescent colored dissolved organic matter by natural bacterial assemblages growing on marine phytoplankton exudates. *Appl Environ Microbiol, 77*, 7490-7498

Roy, S., Sathyendranath, S., & Platt, T. (2017). Size-partitioned phytoplankton carbon and carbon-to-chlorophyll ratio from ocean colour by an absorption-based bio-optical algorithm. *Remote Sensing of Environment, 194*, 177-189

Sadeghi, A., Dinter, T., Vountas, M., Taylor, B., Altenburg-Soppa, M., & Bracher, A. (2012). Remote sensing of coccolithophore blooms in selected oceanic regions using the PhytoDOAS method applied to hyper-spectral satellite data. *Biogeosciences, 9*, 2127-2143

Schwarz, J.N., Kowalczuk, P., Kaczmarek, S., Cota, G., Mitchell, B.G., Kahru, M., Chavez, F.P., Cunningham, A., McKee, D., Gege, P., Kishino, M., Phinney, D.A., & Raine, R. (2002). Two models for absorption by coloured dissolved organic matter (CDOM). *Oceanologia, 44*, 209-241

Silsbe, G.M., Behrenfeld, M.J., Halsey, K.H., Milligan, A.J., & Westberry, T.K. (2016). The CAFE model: A net production model for global ocean phytoplankton. *Global Biogeochemical Cycles, 30*, 1756-1777

Stedmon, C.A., & Markager, S. (2003). Behaviour of the optical properties of coloured dissolved organic matter under conservative mixing. *Estuarine, Coastal and Shelf Science, 57*, 973-979

Stedmon, C.A., & Nelson, N.B. (2014). The optical properties of DOM in the ocean. *Biogeochemistry of Marine Dissolved Organic Matter (Second Edition)* (pp. 481-508): Elsevier

Swan, C.M., Nelson, N.B., Siegel, D.A., & Fields, E.A. (2013). A model for remote estimation of ultraviolet absorption by chromophoric dissolved organic matter based on the global distribution of spectral slope. *Remote Sensing of Environment, 136*, 277-285

Trochta, J.T., Mouw, C.B., & Moore, T.S. (2015). Remote sensing of physical cycles in Lake Superior using a spatio-temporal analysis of optical water typologies. *Remote Sensing of Environment, 171*, 149-161

Tsai, F., & Philpot, W. (1998). Derivative analysis of hyperspectral data. *Remote Sensing of Environment, 66*, 41-51

Twardowski, M.S., Boss, E., Sullivan, J.M., & Donaghay, P.L. (2004). Modeling the spectral shape of absorption by chromophoric dissolved organic matter. *Marine Chemistry, 89*, 69-88

Uitz, J., Stramski, D., Reynolds, R.A., & Dubranna, J. (2015). Assessing phytoplankton community composition from hyperspectral measurements of phytoplankton absorption coefficient and remote-sensing reflectance in open-ocean environments. *Remote Sensing of Environment, 171*, 58-74

Vandermeulen, R.A., Mannino, A., Neeley, A., Werdell, J., & Arnone, R. (2017). Determining the optimal spectral sampling frequency and uncertainty thresholds for hyperspectral remote sensing of ocean color. *Opt Express, 25*, A785-A797

Vantrepotte, V., Danhiez, F.P., Loisel, H., Ouillon, S., Meriaux, X., Cauvin, A., & Dessailly, D. (2015). CDOM-DOC relationship in contrasted coastal waters: implication for DOC retrieval from ocean color remote sensing observation. *Opt Express, 23*, 33-54

Wang, G., Lee, Z., Mishra, D.R., & Ma, R. (2016). Retrieving absorption coefficients of multiple phytoplankton pigments from hyperspectral remote sensing reflectance measured over cyanobacteria bloom waters. *Limnology and Oceanography: Methods, 14*, 432-447

Wang, G., Lee, Z., & Mouw, C. (2017). Multi-Spectral Remote Sensing of Phytoplankton Pigment Absorption Properties in Cyanobacteria Bloom Waters: A Regional Example in the Western Basin of Lake Erie. *Remote Sensing, 9*

Wang, P., Boss, E., & Roesler, C.S. (2005). Uncertainties of inherent optical properties obtained from semianalytical inversions of ocean color. *Appl Opt, 44*, 4074-4085

Wei, J., Lee, Z., Ondrusek, M., Mannino, A., Tzortziou, M., & Armstrong, R. (2016). Spectral slopes of the absorption coefficient of colored dissolved and detrital material inverted from UV-visible remote sensing reflectance. *Journal of Geophysical Research: Oceans*, n/a-n/a

Werdell, J., Franz, B.A., Bailey, S.W., Feldman, G.C., Boss, E., Brando, V., Dowell, M., Hirata, T., Lavender, S.J., Lee, Z., Loisel, H., Maritorena, S., Mélin, F., Moore, T.S., Smyth, T.J., Antoine, D., Devred, E., D'Andon, O.H.F., & Mangin, A. (2013). Generalized ocean color inversion model for retrieving marine inherent optical properties. *Appl Opt, 52*, 2019-2037

Werdell, P.J., Bailey, S.W., Fargion, G.S., Pietras, C., Knobelspiesse, K.D., Feldman, G.C., & McClain, C.R. (2003). Unique data repository facilitates ocean color satellite validation. *EOS Trans. AGU, 84*, 377

Westberry, T.K., Behrenfeld, M., Siegel, D.A., & Boss, E. (2008). Carbon-based primary productivity modeling with vertically resolved photoacclimation. *Global Biogeochemical Cycles, 22*

Wolanin, A., Soppa, M., & Bracher, A. (2016). Investigation of Spectral Band Requirements for Improving Retrievals of Phytoplankton Functional Types. *Remote Sensing, 8*

Wünsch, U.J., Stedmon, C.A., Tranvik, L.J., & Guillemette, F. (2018). Unraveling the size-dependent optical properties of dissolved organic matter. *Limnology and Oceanography, 63*, 588-601

Xing, X., Morel, A., Claustre, H., D'Ortenzio, F., & Poteau, A. (2012). Combined processing and mutual interpretation of radiometry and fluorometry from autonomous profiling Bio-Argo floats: 2. Colored dissolved organic matter absorption retrieval. *Journal of Geophysical Research: Oceans, 117*, n/a-n/a

Zhang, X., Stavn, R.H., Falster, A.U., Gray, D., & Gould, R.W. (2014). New insight into particulate mineral and organic matter in coastal ocean waters through optical inversion. *Estuarine, Coastal and Shelf Science, 149*, 1-12

Zhao, Z., Gonsior, M., Luek, J., Timko, S., Ianiri, H., Hertkorn, N., Schmitt-Kopplin, P., Fang, X., Zeng, Q., & Jiao, N. (2017). Picocyanobacteria and deep-ocean fluorescent dissolved organic matter share similar optical properties. *Nature communications, 8*, 15284

A Copyright documentation