1-9-2018

Characterizing CDOM Spectral Variability Across Diverse Regions and Spectral Ranges

Brice K. Grunert

Colleen B. Mouw
University of Rhode Island, cmouw@uri.edu

Audrey B. Ciochetto
University of Rhode Island, audreyciochetto@uri.edu

Citation/Publisher Attribution

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. https://doi.org/10.1002/2017GB005756

Characterizing CDOM Spectral Variability Across Diverse Regions and Spectral Ranges

Characterizing CDOM Spectral Variability Across Diverse Regions and Spectral Ranges

Brice K. Grunert[1*], Colleen B. Mouw[2], Audrey B. Ciochetto[2]

[1]Michigan Technological University, Department of Geological and Mining Engineering and Sciences, 1400 Townsend Drive, Houghton, MI 49931, USA

[2] University of Rhode Island, Graduate School of Oceanography, 215 South Ferry Road, Narragansett, RI 02882, USA

[*]Corresponding author: bgrunert@mtu.edu, +1 414-322-7506

Keywords: colored dissolved organic matter, carbon cycling, ocean biogeochemistry

Key points

- CDOM spectral slope spatial variability is dependent on spectral range used
- CDOM spectral slope variability is not necessarily reflected in CDOM absorption values in ocean regions with little terrestrial influence
- Gaussian peaks fitting lignin absorption show a distinct blue shift from inland to ocean waters

Abstract

Satellite remote sensing of colored dissolved organic matter (CDOM) has focused on CDOM absorption (a_{CDOM}) at a reference wavelength, as its magnitude provides insight into the underwater light field and large-scale biogeochemical processes. CDOM spectral slope, S_{CDOM}, has been treated as a constant or semi-constant parameter in satellite retrievals of a_{CDOM} despite significant regional and temporal variability. S_{CDOM} and other optical metrics provide insights into CDOM composition, processing, food web dynamics, and carbon cycling. To date, much of this work relies on fluorescence techniques or a_{CDOM} in spectral ranges unavailable to current and planned satellite sensors (e.g. <300 nm). In preparation for anticipated future hyperspectral satellite missions, we take the first step here of exploring global variability in S_{CDOM} and fit deviations in the a_{CDOM} spectra using the recently proposed Gaussian decomposition method. From this, we investigate if global variability in retrieved S_{CDOM} and Gaussian components is significant and regionally distinct. We iteratively decreased the spectral range considered and analyzed the number, location and magnitude of fitted Gaussian components to understand if a reduced spectral range impacts information obtained within a common spectral window. We compared the fitted slope from the Gaussian decomposition method to absorption-based indices that indicate CDOM composition to determine the ability of satellite-derived slope to inform the analysis and modeling of large-scale biogeochemical processes. Finally, we present implications of the observed variability for remote sensing of CDOM characteristics via S_{CDOM}.

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}), where S_{CDOM} is derived from fitting the CDOM absorption spectra to an exponential curve (see Eq. 1) [*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 and lignin derivatives) 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].

Linking a_{CDOM} optical proxies with future hyperspectral satellite data presents a potentially powerful tool to assess spatially and temporally distinct biogeochemical processes associated with a_{CDOM} variability. To date, this connection is complicated by a lack of consistency in the spectral range used to calculate S_{CDOM} and a general focus on a spectral range that best represents the entire a_{CDOM} spectra over the utility of that spectral range [*Schwarz et al.*, 2002; *Twardowski et al.*, 2004]. Broad range S_{CDOM} (e.g. $S_{240:700}$) that best characterize the entire a_{CDOM} spectra have not been considered extensively for mechanistic relationships with CDOM composition. Most studies track S_{CDOM} across salinity gradients [*Kowalczuk et al.*, 2006; *Stedmon and Markager*, 2003; *Stedmon et al.*, 2011] or distinct in-water processes [*Nelson et al.*, 1998; *Organelli et al.*, 2014], producing correlations between S_{CDOM} and CDOM source or diagenetic state across geographically or temporally unique environments [*Organelli et al.*, 2014; *Stedmon and*

93 *Markager*, 2003]. Some work has suggested that $S_{300:600}$ is a useful indicator of CDOM
94 molecular weight [*Stedmon and Nelson 2015*], while the general understanding is that
95 broad range S_{CDOM} increases with photodegradation as chromophores are bleached,
96 particularly in the UV region [*Del Vecchio and Blough*, 2002; *Reche et al.*, 2000]. Optical
97 properties of source material have been considered [*Hansen et al.*, 2016; *Visser*, 1983]
98 and two major components of CDOM, humic acid and fulvic acid, display slopes of
99 approximately 0.011 nm^{-1} and 0.019 nm^{-1} [*Carder et al.*, 1989]. However, observed slope
100 values across broad spectral ranges can vary above and below these benchmarks and the
101 optical properties of source material are rapidly altered through poorly understood
102 processes [*Sharpless et al.*, 2014], suggesting that the complexity of material and its
103 spectral signal is still poorly constrained.

104 Based on this, optically estimating CDOM composition from hyperspectral ocean
105 color data requires validation of significant temporal and regional variability in S_{CDOM}
106 and a potential link between observable spectral features and spectral indices capable of
107 estimating CDOM composition. Maximizing CDOM compositional information that can
108 be retrieved by optical methods will enhance our ability to track changes in the CDOM
109 pool through autonomous and remote sensing platforms, providing a way to observe large
110 scale changes in ocean biogeochemical processes and circulation [*Nelson et al.*, 2010].
111 However, this first requires determination of what information regarding CDOM
112 composition is garnered from a specific spectral range and fully utilizing the information
113 contained within a_{CDOM} spectra to optically estimate CDOM composition. Here, we
114 utilize a_{CDOM} spectra available from public data repositories to address knowledge gaps
115 that have evolved from an inconsistency in spectral range used. We utilize the Gaussian

decomposition approach of Massicotte and Markager [2016] to better estimate S_{CDOM} and further characterize the CDOM pool and report differences between this approach and the standard exponential model traditionally used for fitting a_{CDOM} spectra. We describe global variability in S_{CDOM} within the context of biogeochemical provinces and relate several broad range S_{CDOM} to $S_{275:295}$, $S_{350:400}$, and a_{CDOM} at 350 nm ($a_{CDOM}(350)$) to define the efficacy of each of these parameters to optically estimate CDOM composition based on observations in the literature. 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}.

2. Methods

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:

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

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

139 spectra representing a broad quasi-exponential function: if the fit with Equation 1
140 displayed an $r^2 > 0.9$ (suggesting no contamination or instrumentation issues), spectra
141 were kept for further analysis [*Massicotte and Frenette*, 2011].

142 Deviations from the decaying expression in Eq. 1 can result from absorption by
143 specific chromophores (e.g. lignin), and can vary in degree and location [*Massicotte and*
144 *Markager*, 2016]. To detect regions where such deviations exist, spectra were fit with the
145 standard exponential in Eq. 1 and then trimmed by removing data with residuals greater
146 than the mean absolute residual multiplied by the weighting factor, *C*. *C* is dependent on
147 the quality of the data (signal to noise ratio), the environment and spectral range
148 considered and governs the number of residuals considered for Gaussian fitting by
149 defining the threshold for excluding these points from the baseline exponential fit; we
150 utilized a *C* of 1 to maintain consistency across spectra. With the remaining points,
151 spectra were again fit with Eq. 1 with the result defined as the baseline exponential.
152 Massicotte and Markager [2016] utilized a spectral range from 240-700 nm with
153 Gaussian fitting performed on residuals from 250-500 nm. S_{CDOM} is sensitive to the
154 spectral range used for calculation when using the exponential model [e.g. Twardowski et
155 al. 2004]. To determine the influence of spectral range on S_{CDOM} using the Gaussian
156 decomposition model, we considered spectral ranges of 240-700 nm, 300-700 nm and
157 350-550 nm, with Gaussian fitting performed on residuals below 500 nm to avoid
158 concerns with a low signal-to-noise (SNR) ratio at longer wavelengths [*Massicotte and*
159 *Markager*, 2016]. For spectral ranges below 300 nm, we found including wavelengths
160 down to 240 nm, rather than a 250 nm cutoff, better accounted for absorption deviations
161 likely due to lignin or lignin-derivatives. This and altering the spectral range considered

were the only deviations 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:

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

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.

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. Grouping samples for the INLAND province was justified by two considerations: 1) sampling locations were all in the eastern half of the United States or the Laurentian Great Lakes, and 2) retrieved spectral

208 slope was tightly grouped, ranging from 0.01-0.025 m^{-1} (see the discussion below
209 regarding Figure 5). To be considered for analysis, we required each province to contain
210 5% of the maximum number of samples found within the most sampled province. While
211 sample size is typically treated through consideration of group variance, we found
212 undersampled provinces displayed little variability typically due to all samples coming
213 from a single oceanographic cruise. In effect, our threshold considered provinces with at
214 least 100 spectra for spectral ranges <300 nm (less frequently sampled spectral range)
215 and provinces with at least 200 spectra for spectral ranges ≥300 nm (more frequently
216 sampled spectral range).

217 Absorption spectra were also considered within three depth classes defined as the
218 first and second optical depths, calculated as 2.3 and 4.6 divided by $K_d(490)$,
219 respectively, and below the photic zone (BPZ), comprising depths greater than the second
220 optical depth but less than 1500 m [*Kirk,* 1994]. The majority of spectra below 1500 m
221 were sampled at the Bermuda Atlantic Times Series (BATS) site; hence, we imposed the
222 1500 m threshold to avoid a potential bias from deep water spectra in this region. The
223 depth classes for each province were identified by the average light extinction coefficient
224 at 490 nm, $K_d(490)$, determined from MODIS-Aqua seasonal climatologies for each
225 province. Seasonal $K_d(490)$ was used to ensure samples from different seasons fell within
226 a similar underwater light field to maintain relatively constant photodegradation rates
227 across samples. Thus, a boreal winter sample near the lower limit of the first optical
228 depth in a province could be a few meters deeper than a boreal summer sample near the
229 upper limit of the second optical depth in the same province. Biogeochemical provinces
230 for each depth class were included in the analysis using the same 5% of maximum

231 sampling criterion described above for the complete dataset, resulting in similar sample
232 sizes. For most analyses, only the first optical depth is presented in the figures due to
233 space constraints. Additional depth data for these analyses can be found in the
234 Supplemental Material Tables 1-3. Seasonal distribution of data within the provinces did
235 not allow representation across all four seasons, thus, results focus only on spatial trends
236 in CDOM metrics.

237 **2.3 Analysis of CDOM Metrics**

238 Spectral slope calculated for the Gaussian decomposition and standard
239 exponential methodologies were compared using linear regression. $S_{G240:700}$, $S_{G300:700}$ and
240 $S_{G350:550}$ were compared to $S_{E275:295}$ and $S_{E350:400}$ to determine their potential for
241 estimating CDOM molecular weight, source, and degradation state using linear
242 regression. Non-linear least squares fits were used to assess the ability of log-normalized
243 $a_{CDOM}(\lambda)$ at wavelengths of 350, 412, and 443 nm to estimate $S_{G240:700}$, $S_{G300:700}$, $S_{G350:550}$,
244 $S_{E350:400}$ and $S_{E275:295}$.

245 We sought to determine if observed differences in S_{CDOM} between biogeochemical
246 provinces were significant. We first applied a one-way analysis of variance (ANOVA) to
247 S_{CDOM} categorized by biogeochemical province to determine if the mean slope values
248 between biogeochemical provinces were significantly different ($p < 0.01$). The least
249 significant difference multiple comparison of means (MCM) test was then applied using
250 the results of the ANOVA to determine if the distribution of spectral slope mean values
251 between provinces was significantly different ($p < 0.05$) [*Rafter et al.*, 2002]. It is
252 recognized the least significant difference (LSD) methodology offers looser statistical
253 thresholds for significance than the Tukey-Kramer methodology [*Hayter*, 1986].

254 However, as an exploratory exercise aimed at defining potential differences, the more
255 liberal LSD methodology was deemed more suitable here.

256 Throughout the Results section, multiple comparison of means tests are presented
257 on bioplots as a way to visualize the degree of similarity in S_{CDOM} between different
258 biogeochemical provinces. A line between boxes on a bioplot indicates mean S_{CDOM}
259 values between those provinces are statistically similar, while no connection indicates
260 that mean S_{CDOM} values between considered provinces are statistically unique. These
261 results are further clarified within maps of the regions displaying which provinces are
262 statistically similar and how similar (the number of related provinces) within groups
263 displaying connectivity (statistical similarity). The bioplots illustrate which provinces
264 displayed mean S_{CDOM} values that are explicitly similar while the maps give a global
265 representation of provinces that are statistically affiliated (all provinces are related to at
266 least one other province within that group).

267

268 **3. Results**

269 **3.1 Comparison of CDOM Models**

270 We compared spectral slope retrieved with the standard exponential to the
271 Gaussian decomposition approach (Fig. 3). For all spectral ranges, the majority of points
272 fell at or above the 1:1 line in Fig. 3, demonstrating the Gaussian decomposition fit
273 resulted in identical or larger slope values for a given spectral range. When considering
274 the spectral ranges investigated, differences between $S_{G240:700}$ and $S_{E240:700}$ were smallest
275 for inland and coastal provinces and largest for open ocean provinces (Fig. 4a), although
276 this comparison relies on only one open ocean province (NPTG). Absolute differences

277 between $S_{G300:700}$ and $S_{E300:700}$ did not show a distinct spatial trend, as both NWCS
278 (coastal) and NASW (open ocean) contained the largest absolute differences between the
279 methodologies (0.025 and 0.0279 nm^{-1}, respectively). However, overall mean differences
280 between $S_{G300:700}$ and $S_{E300:700}$ were smallest for inland and coastal regions and increased
281 in open ocean regions, while environments characterized by a greater magnitude of a_{CDOM}
282 (Arctic, inland and coastal waters) more frequently displayed no difference in calculated
283 slope between the two methodologies (Fig. 4b). This is not surprising, as the Gaussian
284 peaks fitted in the 300-700 and 350-550 nm spectral ranges for these environments were
285 relatively small in magnitude. Following this expectation, differences between $S_{G350:550}$
286 and $S_{E350:550}$ were quite small overall and did not display any regional trends (Fig. 4c).

287 **3.2.1 Slope Values: Spectral and Spatial Trends**

288 $S_{E275:295}$ displayed the largest variability in values across all biogeochemical
289 provinces with coastal waters displaying smaller values, typically 0.02-0.03 nm^{-1}.
290 Oceanic, Atlantic regions displayed larger values relative to all Pacific regions except
291 NPSW (Fig.5a,f; Supplementary Material Table 1). $S_{G240:700}$ displayed large differences
292 between provinces influenced by terrestrial CDOM (INLAND, coastal and Arctic waters)
293 and slope values observed in NPTG. $S_{G300:700}$ displayed similar trends, with INLAND,
294 coastal and Arctic waters displaying slope values typically < 0.025 nm^{-1}; however, it
295 should be noted that coastal, sub-Arctic and Arctic provinces displayed more variability
296 in slope values than the INLAND province, particularly with a longer tailed distribution
297 towards higher $S_{G300:700}$ values (Fig. 5d; Supplementary Material Table 1). Open ocean
298 regions tended to display higher slope values but variability and distribution of variables
299 was higher and irregular. In contrast to $S_{G300:700}$, $S_{G350:550}$ trended towards higher values

in coastal provinces and lower values in open ocean provinces, with some open ocean provinces displaying higher variability and irregular distributions (Fig. 5e; Supplementary Material Table 1). Overall, $S_{E275:295}$ displayed the broadest distribution of S_{CDOM} values while the other spectral ranges clustered more tightly around a group mean in generally a normal distribution when considered across all biogeochemical provinces. Mean slope values for each spectral range were 0.036, 0.018, 0.021, 0.020 and 0.016 nm^{-1} for $S_{E275:295}$, $S_{E350:400}$, $S_{G240:700}$, $S_{G300:700}$ and $S_{G350:550}$, respectively, suggesting that S_{CDOM} values decrease as the spectral range considered shifts to longer wavelengths, an important bias to consider when choosing a spectral range for a given application.

While previous studies have shown a generally decreasing slope value with depth, $S_{E350:400}$ was the only spectral range that showed a consistent relationship between increasing depth zone and decreasing spectral slope value. $S_{G350:550}$ decreased with depth zone, $S_{E275:295}$ decreased below the first optical depth but did not display a trend between the second optical depth and below the photic zone for most provinces, while $S_{G240:700}$ and $S_{G300:700}$ displayed relatively consistent slope values across all depth zones for most provinces (Supplementary Material Table 1). Overall, variability in the slope value decreased with depth for all spectral ranges considered.

3.2.2 Slope Values: Global Trends

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} mean values are presented as clusters where each province presented displays similar mean S_{CDOM} values to at least one other province for that spectral range. $S_{E275:295}$ showed the clearest spatial relationships, with provinces dominated by high a_{CDOM} and

riverine input displaying lower mean $S_{E275:295}$ values (0.024 nm^{-1}; Fig. 6a,b) and open ocean regions, presumably dominated by photobleaching, displaying high mean $S_{E275:295}$ values (0.039 nm^{-1}, Fig. 6c,d; 0.044 nm^{-1}, Fig. 6e,f; 0.048 nm^{-1}, Fig. 6g,h). Lower $S_{E275:295}$ values (0.026 nm^{-1}; Fig. 6i,j) are found in provinces with a significant portion of samples taken in upwelling regions or regions where deep ocean waters pass over continental shelves (within the first optical depth). The remaining regions were either independent in their mean $S_{E275:295}$ value or displayed similarity with one other province (SANT and ANTA). INLAND displayed the lowest mean $S_{E275:295}$ value of 0.014 nm^{-1}. Spatial trends for $S_{E350:400}$ were not as clear, with high a_{CDOM} coastal, sub-Arctic and Arctic provinces showing similar spectral slope values as all Atlantic provinces considered ($S_{E350:400}$=0.017 nm^{-1}, Fig. 7a,b). Open ocean provinces from the Pacific, Indian and Southern Oceans clustered with slope values either higher ($S_{E350:400}$=0.021 nm^{-1}, Fig. 7c,d) or lower ($S_{E350:400}$=0.014 nm^{-1}, Fig. 7e,f) than the $S_{E350:400}$ value from higher a_{CDOM} provinces, suggesting differences in the balance between photo- and microbial degradation relative to source signal. $S_{G300:700}$ showed little variation in mean slope value across all provinces while still producing three distinct clusters and INLAND as an independent province ($S_{G300:700}$=0.021 nm^{-1}, Fig. 8a,b; $S_{G300:700}$=0.023 nm^{-1}, Fig. 8c,d; $S_{G300:700}$=0.019 nm^{-1}, Fig. 8e,f; $S_{G300:700}$=0.016 nm^{-1}, Fig. 8g). Clustering appears to be a mix of results from $S_{E275:295}$ and $S_{E350:400}$, suggesting that a lack of contribution from wavelengths below 300 nm and a broad spectral range lead to greater similarities between disparate CDOM pools. While also displaying a relatively narrow range of mean slope values, $S_{G350:550}$ showed the most connectivity between proximal provinces and region while also separating into four distinct clusters ($S_{G350:550}$=0.016 nm^{-1}, Fig. 9a,b;

346 $S_{G350:550}$=0.017 nm^{-1}, Fig. 9c,d; $S_{G350:550}$=0.015 nm^{-1}, Fig. 9e,f; $S_{G350:550}$=0.014 nm^{-1}, Fig.
347 9g,h). APLR was the only independent province for this spectral range, displaying a
348 significantly higher mean slope value ($S_{G350:550}$=0.023 nm^{-1}, Fig. 9i).

349 **3.3 Gaussian Components**

350 The number of Gaussian components fit was highly dependent on the spectral
351 range considered (Fig. 10). $S_{G240:700}$ peak fitting was dominated by the lignin peak with
352 the mean first Gaussian component location occurring at 276 nm for all 6 biogeochemical
353 provinces considered. Fewer Gaussian components were fit in INLAND and coastal
354 provinces due to lignin peak fitting dramatically improving the goodness of fit (Fig.
355 10a,11a; Supplementary Material Table 2). This is also evidenced by NPTG having the
356 most Gaussian components fit relative to other provinces while peak height was generally
357 small for this province. Arctic and sub-Arctic provinces also saw an increase in fitted
358 components relative to other provinces with relatively high a_{CDOM} (e.g. INLAND). For
359 this spectral range, the location of the first component was strongly influenced by sample
360 source and proximity to terrestrial material, as the first Gaussian component location was
361 always below 300 nm for INLAND, NWCS, CARB and BERS while the first Gaussian
362 component location ranged up to 385 and 381 nm for BPLR and NPTG, respectively. For
363 $S_{G300:700}$, INLAND, sub-Arctic, Arctic and Atlantic provinces were fit with more
364 Gaussian components than Pacific and Southern Ocean provinces (Fig. 10b,11d;
365 Supplementary Material Table 2). This spectral range also observed the most fitted peaks
366 across all provinces. $S_{G350:550}$ observed similar trends relative to ocean province as
367 $S_{G300:700}$, but with fewer fitted peaks on average, likely due to peak height for components
368 located in this fitting range being less prominent features of the absorption spectra (Fig.

369 10c,11g; Supplementary Material Table 2). Interestingly, BPLR and BERS were fitted
370 with significantly more peaks in this spectral range, potentially due to relatively high
371 productivity and strong riverine influence in these regions leading to a complex CDOM
372 pool. BPLR and BERS were fit with large Gaussian peaks below 350 nm, also suggesting
373 that fitting strictly above 350 nm capture smaller deviations in the absorption spectrum
374 that are effectively masked with the improvement in fitting single, large deviations at
375 shorter wavelengths.

376 Gaussian peak height, φ, was found to correlate quite well with $a_{CDOM}(350)$
377 (r^2=0.88), although the distribution appears bifurcated with a break-point between
378 $a_{CDOM}(350) \leq 15$ m^{-1} (r^2=0.76) and $a_{CDOM}(350) > 15$ m^{-1} (r^2=0.90) (Supplementary
379 Material Figure 1). Both relationships suggest a large dependency between φ and the
380 magnitude of a_{CDOM} in the system. This effect was accounted for by normalizing φ with
381 the modeled a_{CDOM} at the wavelength associated with the Gaussian peak location as
382 described in the Methods section. Normalized φ showed that open ocean Gaussian
383 components are more prominent relative to the magnitude of CDOM in the system when
384 compared to terrestrial components, with INLAND mean normalized φ=0.03 while ocean
385 provinces had a mean normalized φ=0.1 (data not shown). Overall, provinces with high
386 a_{CDOM} tended to be fit with Gaussian peaks with larger peak height and width. Gaussian
387 component location did not change significantly with depth for all spectral ranges
388 considered (Fig. 11, Supplementary Material Table 2), while Gaussian component
389 location only displayed unique mean locations for INLAND ($S_{G300:700}$) and NASW
390 ($S_{G300:700}$,$S_{G350:550}$), suggesting that these features originate from a relatively small pool of
391 chromophores that are likely unique to terrestrial and ocean regions. Across all spectral

ranges, chromophores tended to be fit within spectral regions known to contain lignin species, photorefractory chromophores and degradation products produced *in situ* (see Discussion).

3.4 CDOM Metric Comparisons

3.4.1 Comparison of Broad Range Slopes to $S_{275:295}/S_{350:400}$

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; Supplementary Material Table 3). 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 ($r^2 < 0.2$; Supplementary Material Table 4), 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 (Supplementary Material Table 3).

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

415 reference wavelength to estimate S_{CDOM}. 412 and 443 nm, or nearby wavelengths, have
416 applications to heritage satellite sensors while 350 nm is the shortest wavelength
417 anticipated to be available for NASA's Plankton, Aerosol, Cloud, ocean Ecosystem
418 (PACE) sensor, providing maximum signal-to-noise for retrieving a_{CDOM} while also
419 avoiding overlap with phytoplankton pigments, including mycosporine-like amino acid
420 absorption peaks around 330 and 360 nm [*Pavlov et al.*, 2014]. Relationships were best
421 fit using a non-linear least squares exponential fit. The goodness of fit did not change
422 significantly between different reference wavelengths; subsequently, we only focus on
423 $a_{CDOM}(350)$ in anticipation of future sensor capabilities. $S_{E275:295}$ had the strongest
424 relationships with $a_{CDOM}(350)$, $S_{G300:700}$ relationships were highly variable, while $S_{E350:400}$
425 and $S_{G350:550}$ did not correlate with $a_{CDOM}(350)$ for almost all regions and depths
426 considered (Supplementary Material Table 3).

427 We also assessed the error introduced in satellite-derived estimates of $a_{CDOM}(412)$
428 by using an assumed S_{CDOM} value. CDOM has traditionally been considered alongside
429 non-algal particulate (NAP) material, as both have absorption spectra that follow an
430 approximately exponentially increasing with decreasing wavelength relationship. These
431 absorption terms, a_{CDOM} and $a_{NAP,}$ are combined into a single term, colored detrital matter
432 (CDM) absorption such that $a_{CDM}=a_{CDOM}+a_{NAP}$ and a_{CDM} has an average spectral slope
433 (S_{CDM}) representative of the slope and percent contribution of each component. The GSM
434 algorithm and the Quasi-Analytical Algorithm (QAA) assume an S_{CDM} of 0.015 and
435 0.013-0.017 nm^{-1}, respectively [*Lee et al.*, 2002; *Maritorena et al.*, 2002]. Global
436 observations of a_{NAP} suggest it accounts for 10-20% of the a_{CDOM} signal, and a typical
437 spectral slope for a_{NAP} is 0.011 nm^{-1} with a range of values much smaller than S_{CDOM}

438 [*Dong et al.*, 2013; *Hoepffner and Sathyendranath*, 1993; *Roesler et al.*, 1989]. Assuming
439 an S_{CDOM} of 0.015 nm^{-1}, a contribution of 10% and 20% of a_{NAP} with spectral slope of
440 0.011 nm^{-1} results in an assumed S_{CDOM} of 0.016 and 0.0154 nm^{-1}, respectively. We
441 assessed the percent error between mean $a_{CDOM}(412)$ observed for each province and
442 $a_{CDOM}(412)$ calculated for the province using the mean $a_{CDOM}(443)$ and these assumed
443 S_{CDOM} values. Error rates were similar when assuming 10% and 20% contributions from
444 a_{NAP} to a_{CDOM}. When assuming a 20% contribution and an S_{CDOM} of 0.0154 nm^{-1},
445 estimated $a_{CDOM}(412)$ error varied from 4-22% of the observed mean $a_{CDOM}(412)$ (Fig.
446 12).
447
448 **4. Discussion**
449 **4.1 CDOM Models**
450 Bias introduced in the S_{CDOM} parameter from varying the spectral range has been
451 widely recognized for some time [*Twardowski et al.*, 2004]. Despite this understanding,
452 only a few narrow range spectral slope parameters have been adopted consistently,
453 $S_{275:295}$ and $S_{350:400}$ [*Helms et al.*, 2008], with highly variable broad range slopes reported
454 throughout the literature. Regardless of the CDOM model used, exponential or Gaussian
455 decomposition, the bias due to spectral range considered remains an issue. S_{CDOM}
456 differences between the exponential CDOM and Gaussian decomposition models are
457 relatively minor for most spectra, with S_{CDOM} values generally higher for the Gaussian
458 decomposition model. However, differences between the models suggest that removing
459 deviations from the exponential model and modeling these features with Gaussian curves
460 allows for a better characterization of the underlying exponential signal and a better fit of

461 the measured CDOM spectra through improved r-squared and RMSE. This, in turn, will
462 allow more consistent comparisons of S_{CDOM} from the same spectral range and a stronger
463 basis for estimating CDOM composition optically [*Del Vecchio and Blough*, 2004a]. In
464 particular, we found the Gaussian decomposition method to be crucial for accurately
465 modeling CDOM spectra strongly influenced by terrestrial material.

466 **4.2 S_{CDOM}**

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

475 CDOM absorption at a reference wavelength is often used as an indicator of
476 CDOM composition as it relates well with $S_{E275:295}$ and lignin concentration [*Fichot et*
477 *al.*, 2016; *Mannino et al.*, 2014]. Past studies have shown strong relationships between
478 a_{CDOM} and S_{CDOM} along transects from lower salinity coastal waters to higher salinity
479 offshore waters where the range in $a_{CDOM}(\lambda)$ typically varies by orders of magnitude
480 [*Kowalczuk et al.*, 2006; *Pavlov et al.*, 2016; *Stedmon and Markager*, 2003], while a
481 study in the Arctic Ocean found that CDOM absorption and S_{CDOM} do not relate well
482 between marine and terrestrially-derived CDOM pools [*Granskog et al.*, 2012]. For our
483 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 within the context of low $a_{CDOM}(350)$ values. 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, reinforcing a disconnect between the magnitude of $a_{CDOM}(350)$ and S_{CDOM}.

Existing methods that estimate 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 NASA's PACE sensor. 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

530 contrasts between distinct CDOM pools in the global ocean. Neither $S_{E350:400}$ or $S_{G350:550}$
531 related well with $S_{E275:295}$; thus, both will likely be poor proxies for evaluating terrestrial
532 contribution and, potentially, molecular weight [*Helms et al.*, 2008].

533 Biogeochemical models assume an S_{CDOM} slope [*Xiu and Chai*, 2014] with some
534 models accounting for microbial and photodegradation of CDOM over appropriate time
535 scales [*Dutkiewicz et al.*, 2015]. While the inclusion of optical parameters in global
536 biogeochemical ocean models is a recent development, further partitioning the spectral
537 properties of CDOM by region would enhance the accuracy when modeling the
538 underwater light field. Past studies have found changes in CDOM spectral characteristics
539 with depth to be significant [*Hickman et al.*, 2010; *Pérez et al.*, 2016]. We found changes
540 in mean CDOM spectral characteristics to change up to 0.034, 0.02, 0.008 and 0.009 nm^{-1}
541 between biogeochemical provinces while changes across depth ranges within a given
542 biogeochemical province varied up to 0.016, 0.003, 0.005 and 0.004 nm^{-1} for $S_{E275:295}$,
543 $S_{G240:700}$, $S_{G300:700}$, and $S_{G350:550}$, respectively. While spatial variability in S_{CDOM} were
544 greater between provinces than variability by depth within provinces, $S_{E350:400}$ displayed
545 ranges up to 0.007 nm^{-1} between provinces and by depth within provinces. From this, we
546 suggest that *in situ* production pathways for CDOM are more variable by depth than
547 across global ocean provinces. To aid in future modeling and satellite remote sensing
548 efforts, we've presented slope values for each spectral range, province and depth range
549 (Supplementary Material Table 3).

550 We observed differences between slope values measured across the spectral
551 ranges for each province, suggesting that each spectral range does not convey the same
552 information about CDOM composition. To understand what could be determined about

553 CDOM composition from each spectral range, we considered how $S_{G240:700}$, $S_{G300:700}$ and
554 $S_{E350:550}$ related to $S_{E275:295}$, which is regarded as an indicator of source, molecular weight
555 and photobleaching of CDOM [*D'Sa et al.*, 2014; *Helms et al.*, 2008], and $S_{E350:400}$, a
556 slope range less entrenched in the literature but indicative of photobleaching, molecular
557 weight and microbial processing and production of CDOM [*Helms et al.*, 2008; *Helms et*
558 *al.*, 2013; *Helms et al.*, 2015; *Matsuoka et al.*, 2015]. We found that $S_{E350:400}$ likely
559 characterizes *in situ* production and degradation of CDOM with potential links to
560 microbial processes [*Matsuoka et al.*, 2015; *Nelson et al.*, 2004; *Seidel et al.*, 2015] as
561 this parameter varied throughout the global ocean, showed consistency across
562 presumably diverse CDOM pools (e.g. INLAND and ocean provinces; Fig. 7) and
563 consistently increased with depth. $S_{G240:700}$ is strongly correlated with $S_{E275:295}$ in most
564 regions (r^2=0.48-0.95, excluding BERS) suggesting that wavelengths below 300 nm
565 strongly influence S_{CDOM} in this spectral range. $S_{G300:700}$ displayed variable relationships
566 with $S_{E275:295}$ (r^2=0-0.90) and $S_{E350:400}$ (r^2=0.01-0.86). Typically, a region characterized by
567 a poor relationship between $S_{300:700}$ and $S_{E275:295}$ displayed a strong relationship between
568 $S_{300:700}$ and $S_{E350:400}$ (e.g. BPLR). A strong relationship with each parameter would be
569 expected if $S_{300:700}$ represents CDOM composition in a similar manner to $S_{E275:295}$ or
570 $S_{E350:400}$. However, considering that a consistent trend with one over the other was not
571 observed, it is likely that $S_{G300:700}$ blurs characteristics of each into a single parameter that
572 may not be particularly effective at characterizing the CDOM pool except under ideal
573 circumstances such as a single, dominant process contributing to the CDOM pool. Thus,
574 past work that has found potential relationships between molecular weight and $S_{G300:600}$, a

575 parameter we found to behave quite similarly to $S_{G300:700}$, may be contingent on the
576 environment [e.g. *Stedmon and Nelson 2015*].

577 Considering the variability and trends in slope for the different spectral ranges and
578 relationships with $S_{E275:295}$, it is likely that $S_{G350:550}$ will be sensitive to a smaller suite of
579 processes than spectral ranges that extend down to or below 300 nm. While $S_{G350:550}$
580 could be useful for accurately modeling a_{CDOM} above 350 nm, it appears to be limited for
581 extending a_{CDOM} lower than 350 nm. $S_{G350:550}$ showed greater differentiation between
582 biogeochemical provinces while maintaining a strong relationship with $S_{E350:400}$ in most
583 provinces. Few studies have reported spectral slope from 350-550 nm to our knowledge
584 [*Hancke et al.*, 2014; *Kowalczuk et al.*, 2006], making it difficult to directly assess what
585 $S_{G350:550}$ estimates about CDOM composition. We propose that $S_{E350:400}$ is a better metric
586 for tracking compositional changes in the CDOM pool affiliated with photodegradation
587 and *in situ* production of CDOM as suggested by Helms et al. [2013, 2015]. However,
588 $S_{G350:550}$ displayed greater uniqueness between provinces while $S_{E350:400}$ displayed more
589 consistent trends with depth across all spectral ranges. If these trends are consistent in
590 future datasets, we propose that $S_{G350:550}$ will be an ideal parameter for directly estimating
591 compositional changes in autochthonous CDOM from hyperspectral ocean color data
592 while changes in $S_{E350:400}$ will be indicative of vertical transport of unique CDOM or
593 distinct *in situ* production pathways. Based on previous studies, $a_{CDOM}(\lambda)$ will likely
594 remain a useful parameter for estimating terrestrial CDOM contributions from
595 hyperspectral satellite observations [*Fichot et al.*, 2014; *Mannino et al.*, 2014]. These
596 parameters appear to relate well with oceanic, *in situ* processes with the potential for

597 tracking vertical movement of the CDOM pool and informing estimates of CDOM
598 composition.

599 **4.3 Gaussian Components**

600 The spectral range used to fit a_{CDOM} strongly impacted the number of Gaussian
601 components fitted. When the first absorption peak was below 300 nm, presumed to be
602 lignin [*McKnight and Aiken*, 1998], the goodness of fit increased so significantly that
603 smaller peaks at longer wavelengths that were fitted in the 300-700 nm spectral range
604 were no longer fitted. Fitting from 300-700 nm resulted in the most peaks fitted for all
605 provinces, while 350-550 nm fitted fewer peaks than 300-700 nm suggesting that
606 chromophores between 300-350 nm can be fit while still fitting features at wavelengths >
607 350 nm. In waters where lignin is a strong or moderate contributor to a_{CDOM} below 300
608 nm, fitting from 240-500 nm, then fitting from 300-500 nm, may be a better approach,
609 although blending models could become an issue. The majority of waters sampled to 240
610 nm in the SeaBASS dataset occurred in waters that historically have a strong terrestrial
611 component [*Benner et al.*, 2005; *D'Sa et al.*, 2014; *D'Sa and DiMarco*, 2009; *Del*
612 *Vecchio and Blough*, 2004b], precluding a thorough analysis of the 240-700 nm spectral
613 range across the full range of oceanic conditions observed in the entire SeaBASS dataset.
614 However, $S_{E275:295}$ suggests that the trend in spectral slope values that account for a_{CDOM}
615 below 300 nm prevails in the global oceans. For spectra fitted with a peak below 300 nm,
616 mean peak location was 286 nm, suggesting that $S_{E275:295}$ is strongly influenced by the
617 shape of the Gaussian component, if present, rather than the underlying exponential
618 curve. It is likely that the large, complex molecular structure of lignin and the absorption
619 peak associated with lignin drive the relationship between $S_{E275:295}$ and CDOM molecular

620 weight [*McKnight and Aiken*, 1998]. Additionally, we observed a shift in the location of
621 the lignin absorption peak from terrestrial waters to oceanic waters (Fig. 11a) consistent
622 with photodegradation of this component [*Del Vecchio and Blough*, 2004a].

623 Absorption peaks at wavelengths less than 300 nm can extend beyond these
624 wavelengths through a complex process of charge-transfer interactions in the CDOM
625 pool [*Del Vecchio and Blough*, 2004a]. For waters strongly impacted by terrestrial
626 material and displaying a low $S_{275:295}$ spectral signature, the first Gaussian component
627 occurred at a much smaller wavelength in the spectra. Past studies have shown that
628 terrestrial material absorption is dominated by lignin which absorbs below 300 nm
629 [*McKnight and Aiken*, 1998; *Spencer et al.*, 2008]; however, the deviation from the
630 baseline associated with this peak extends beyond 300 nm [*Fichot et al.*, 2016], resulting
631 in a distortion of $S_{300:700}$ in these waters. Fitting Gaussian peaks provides a method to
632 pick out unique components within the CDOM pool, similar to fitting fluorescence peaks
633 in excitation-emission matrix spectroscopy (EEMs) and accounts for deviations that
634 impact S_{CDOM} in the spectral range considered. More work is required to determine the
635 significance of these absorption-based features to particular groups of molecules in the
636 CDOM pool, including whether features fitted between 300-325 nm in the 300-700 nm
637 spectral range are unique chromophores or a residual effect from the strong absorption of
638 lignin absorption extending above 300 nm.

639 Across all regions, spectral ranges that were commonly fitted were associated
640 with spectral locations (~350 nm, 375 nm) of chromophores that are likely photorefactory
641 [*Helms et al.*, 2013] or chromophores that have been found to be produced from
642 photobleaching of autochthonously produced CDOM [*Swan et al.*, 2012]. The latter, a

643 feature observed between 410-420 nm and noted in previous studies [*Bricaud et al.*,
644 2010; *Swan et al.*, 2012], was the most commonly fitted peak across all provinces in the
645 300-700 nm spectral range. This feature was noticeably present across all spectral ranges,
646 typically representing the second peak fitted in the 240-700 nm spectral range when more
647 than one peak was fitted to the spectrum in this spectral range.

648 Two provinces that stood out regarding the average number of Gaussian peaks
649 fitted per spectra were BPLR and BERS (Fig. 10). BPLR was dominated by
650 measurements in the Chukchi Sea, a region with a greater magnitude of CDOM than the
651 global ocean but less than most Arctic shelf regions due to predominantly
652 authochthonously-produced CDOM [*Dainard and Guéguen*, 2013]. However, both
653 BPLR and BERS have elevated levels of CDOM, likely a contributing factor to
654 significantly more fitted peaks on average than other provinces. Both provinces were
655 predominantly sampled during or shortly after the intense spring phytoplankton blooms
656 associated with ice melt in these regions [*Arrigo et al.*, 2014; *Goes et al.*, 2013].
657 Phytoplankton absorption spectra in this region frequently display absorption in UV
658 wavelengths due to the presence of mycosporine-like amino acids, a feature previously
659 observed in the CDOM pool in Arctic habitats [*Pavlov et al.*, 2014]. This, in conjunction
660 with an active microbial community [*Matsuoka et al.*, 2015] are potential explanations
661 for why these regions were dominated by spectra with an average of more than four
662 Gaussian components. The dramatic reduction in number of Gaussian peaks fitted to
663 a_{CDOM} spectra in BERS and BPLR from 300-700 nm to 350-550 nm despite a relatively
664 constant mean location of peaks (when fitted) suggests that further consideration can be
665 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 while not distorting the BIC analysis. Locating and observing changes in these peaks can potentially provide 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 primarily 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. 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.

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 $S_{E275:295}$ [e.g. *Mannino et al. 2014*]. For our analysis, $S_{E275:295}$ 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

712 relationships, suggesting high variability in CDOM pools and significant differences in
713 CDOM characteristics with similar $a_{CDOM}(\lambda)$ values. Thus, some regions would be
714 precluded from this methodology. Considering that most regions displayed unique trends
715 and distributions of slope, a global relationship is anticipated to poorly predict $S_{E275:295}$.

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

732 We also investigated the impact of inaccurate S_{CDOM} values for estimating
733 $a_{CDOM}(412)$ by comparing calculated $a_{CDOM}(412)$ to mean $a_{CDOM}(412)$ for each province.
734 The resultant errors, from 4-22%, suggest that poorly parameterizing S_{CDOM} lends

735 significant uncertainty to satellite estimates of IOP's strictly introduced through S_{CDOM} or
736 S_{CDM}. For multispectral algorithms, we suggest accounting for differences in S_{CDOM}
737 between geographical regions, such as the provinces presented here, and continued
738 consideration of the mean ratio of a_{NAP}:a_{CDOM} across distinct biogeochemical regions for
739 methods that utilize S_{CDM}.
740

741 **5. Conclusions**

742 In an attempt to close the knowledge gap and move towards a common
743 methodology, we have presented S_G values for three broad spectral ranges, 240-700 nm,
744 300-700 nm and 350-550 nm as well as S_E for 275-295 nm and 350-400 nm. We also
745 explored the ability of the Gaussian decomposition method to expand our optical
746 understanding of CDOM composition from a global database. Ideally, S_{CDOM} can both
747 adequately model the a_{CDOM} spectra and relate to CDOM characteristics. We presented
748 S_{CDOM} as a diagnostic tool that can provide insights into CDOM composition with the
749 potential to be applied to hyperspectral ocean color applications for optical estimates of
750 CDOM composition.

751 Of the slope ranges considered, we found $S_{E275:295}$ and $S_{E350:550}$ display clear,
752 unique spatial trends that can be exploited for optically estimating changes in CDOM
753 across diverse open ocean environments. $S_{E350:400}$ displayed potential for tracking changes
754 to *in situ* production of CDOM, particularly with depth at a given location. $S_{G300:700}$ is a
755 useful metric for particular regions if the underlying CDOM pool is known; however, this
756 metric displayed good relationships with both $S_{E275:295}$ and $S_{E350:400}$, implying that the

757 metric itself does not clearly delineate specific compositional traits that impact the optical
758 signature of CDOM.

759 Hyperspectral capabilities may allow for direct estimates of S_{CDOM}, providing
760 insight into CDOM degradative state and *in situ* production pathways. However, S_{CDOM}
761 calculated using wavelengths anticipated to be available from NASA's PACE sensor (>
762 ~350 nm) differs significantly from $S_{E275:295}$, requiring alternative methods for estimating
763 terrestrial contribution, lignin content and molecular weight. The divergence in optical
764 properties of $S_{E350:400}$ and $S_{E275:295}$ observed here and documented in the literature suggest
765 that $S_{E350:400}$ can provide insight into *in situ* production pathways; however, further
766 consideration should be given to this parameter in future studies along with whether a
767 broader spectral range such as $S_{G350:550}$ provides more insight into these processes or
768 clearer regional distinctions.

769 Beyond tracking changes in CDOM and presenting a way to estimate CDOM
770 composition, S_{CDOM} is also important for accurately modeling a_{CDOM}. We considered the
771 mean $a_{CDOM}(443)$ for each province and assessed how much error is introduced to the
772 a_{CDOM} parameter using an assumed S_{CDOM} of 0.0154 nm^{-1} and propagating a_{CDOM} to 412
773 nm. We found this assumed slope introduced errors in $a_{CDOM}(412)$ of 4-22% across all
774 provinces relative to the mean $a_{CDOM}(412)$ observed (Fig. 12). Thus, poorly
775 parameterizing S_{CDOM} with a constant global slope can introduce a similar, if not greater,
776 level of uncertainty in retrieved a_{CDOM} to mischaracterizing the percent contribution of
777 a_{NAP} to $a_{CDM}(412)$.

778 Ultimately, CDOM will be best considered using a suite of metrics applied to
779 ocean color imagery. Past work focusing on relationships between a_{CDOM} at a given

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

Acknowledgements

We gratefully acknowledge contributors to the SeaBASS data set (https://seabass.gsfc.nasa.gov/). Piotr Kowalczuk and an anonymous reviewer were crucial to shaping the final version of the manuscript. Philippe Massicotte was integral for clarifying the Gaussian decomposition method and sharing source code. Karl Meingast provided comments that shaped the final version of this manuscript. Research conducted here was funded by a NASA Earth and Space Science Fellowship awarded to Grunert.

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.

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

836 Del Vecchio, R., and N. V. Blough (2002), Photobleaching of chromophoric dissolved
837 organic matter in natural waters: kinetics and modeling, *Marine Chemistry*, *78*,
838 231-253.

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

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

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

847 Dutkiewicz, S., A. E. Hickman, O. Jahn, W. W. Gregg, C. B. Mouw, and M. J. Follows
848 (2015), Capturing optically important constituents and properties in a marine
849 biogeochemical and ecosystem model, *Biogeosciences*, *12*(14), 4447-4481.

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

853 Fichot, C. G., R. Benner, K. Kaiser, Y. Shen, R. M. W. Amon, H. Ogawa, and C.-J. Lu
854 (2016), Predicting Dissolved Lignin Phenol Concentrations in the Coastal Ocean
855 from Chromophoric Dissolved Organic Matter (CDOM) Absorption Coefficients,
856 *Frontiers in Marine Science*, *3*.

857 Goes, J. I., H. do Rosario Gomes, E. Haugen, K. McKee, E. D'Sa, A. M. Chekalyuk, D.
858 Stoecker, P. Stabeno, S. Saitoh, and R. Sambrotto (2013), Fluorescence, pigment
859 and microscopic characterization of Bering Sea phytoplankton community
860 structure and photosynthetic competency in the presence of a Cold Pool during
861 summer, *Deep Sea Research Part II: Topical Studies in Oceanography*.

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

866 Hancke, K., E. K. Hovland, Z. Volent, R. Pettersen, G. Johnsen, M. Moline, and E.
867 Sakshaug (2014), Optical properties of CDOM across the Polar Front in the

868 Barents Sea: Origin, distribution and significance, *Journal of Marine Systems*, *130*,
869 219-227.

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

872 Hansen, A. M., T. E. C. Kraus, B. A. Pellerin, J. A. Fleck, B. D. Downing, and B. A.
873 Bergamaschi (2016), Optical properties of dissolved organic matter (DOM):
874 Effects of biological and photolytic degradation, *Limnology and Oceanography*,
875 *61*(3), 1015-1032.

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

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

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

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

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

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

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

898 Kim, G. E., A. Gnanadesikan, and M.-A. Pradal (2016), Increased Surface Ocean
899 Heating by Colored Detrital Matter (CDM) Linked to Greater Northern
900 Hemisphere Ice Formation in the GFDL CM2Mc ESM, *Journal of Climate*, *29*(24),
901 9063-9076.

902 Kowalczuk, P., C. A. Stedmon, and S. Markager (2006), Modeling absorption by
903 CDOM in the Baltic Sea from season, salinity and chlorophyll, *Marine Chemistry*,
904 *101*(1-2), 1-11.

905 Kowalczuk, P., J. Stoń-Egiert, W. J. Cooper, R. F. Whitehead, and M. J. Durako (2005),
906 Characterization of chromophoric dissolved organic matter (CDOM) in the Baltic
907 Sea by excitation emission matrix fluorescence spectroscopy, *Marine Chemistry*,
908 *96*(3-4), 273-292.

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

912 Mannino, A., M. G. Novak, S. B. Hooker, K. Hyde, and D. Aurin (2014), Algorithm
913 development and validation of CDOM properties for estuarine and continental
914 shelf waters along the northeastern U.S. coast, *Remote Sensing of Environment*,
915 *152*, 576-602.

916 Maritorena, S., D. A. Siegel, and A. R. Peterson (2002), Optimization of a
917 semianalytical ocean color model for global-scale applications, *Applied optics*,
918 *41*(15), 2705-2714.

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

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

925 Matsuoka, A., S. B. Hooker, A. Bricaud, B. Gentili, and M. Babin (2013), Estimating
926 absorption coefficients of colored dissolved organic matter (CDOM) using a semi-
927 analytical algorithm for southern Beaufort Sea waters: application to deriving
928 concentrations of dissolved organic carbon from space, *Biogeosciences*, *10*(2),
929 917-927.

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

934 McKnight, D. M., and G. R. Aiken (1998), Sources and Age of Aquatic Humus, in
935 *Aquatic Humic Substances: Ecology and Biogeochemistry*, edited by D. O. Hessen
936 and L. J. Tranvik, Springer-Verlag Berlin Heidelberg, Germany.

937 Meler, J., P. Kowalczuk, M. Ostrowska, D. Ficek, M. Zabłocka, and A. Zdun (2016),
938 Parameterization of the light absorption properties of chromophoric dissolved
939 organic matter in the Baltic Sea and Pomeranian lakes, *Ocean Science*, *12*(4),
940 1013-1032.

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

943 Nelson, N. B., C. A. Carlson, and D. K. Steinberg (2004), Production of chromophoric
944 dissolved organic matter by Sargasso Sea microbes, *Marine Chemistry*, *89*(1-4),
945 273-287.

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

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

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

956 Ortega-Retuerta, E., T. Frazer, C. M. Duarte, S. Ruiz-Halpern, A. Tovar-Sanchez, J. M.
957 Arrieta, and I. Reche (2009), Biogeneration of chromophoric dissolved organic
958 matter by bacteria and kirll in the Southern Ocea, *Limnology & Oceanography*,
959 *54*(6), 1941-1950.

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

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

968 Pérez, G. L., M. Galí, S.-J. Royer, H. Sarmento, J. M. Gasol, C. Marrasé, and R. Simó
969 (2016), Bio-optical characterization of offshore NW Mediterranean waters:
970 CDOM contribution to the absorption budget and diffuse attenuation of
971 downwelling irradiance, *Deep Sea Research Part I: Oceanographic Research*
972 *Papers*, *114*, 111-127.

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

975 Reche, I., M. L. Pace, and J. J. Cole (2000), Modeled Effects of Dissolved Organic
976 Carbon and Solar Spectra on Photobleaching in Lake Ecosystems, *Ecosystems*,
977 *3*(5), 419-432.

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

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

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

987 Sharpless, C. M., M. Aeschbacher, S. E. Page, J. Wenk, M. Sander, and K. McNeill
988 (2014), Photooxidation-induced changes in optical, electrochemical, and
989 photochemical properties of humic substances, *Environ Sci Technol*, *48*(5), 2688-
990 2696.

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

994 Spencer, R. G. M., G. R. Aiken, K. P. Wickland, R. G. Striegl, and P. J. Hernes (2008),
995 Seasonal and spatial variability in dissolved organic matter quantity and
996 composition from the Yukon River basin, Alaska, *Global Biogeochemical Cycles*,
997 *22*(4).

998 Stedmon, C. A., and S. Markager (2003), Behaviour of the optical properties of
999 coloured dissolved organic matter under conservative mixing, *Estuarine, Coastal*
1000 *and Shelf Science*, *57*(5-6), 973-979.

1001 Stedmon, C. A., and S. Markager (2005), Resolving the variability in dissolved
1002 organic matter fluorescence in a temperate estuary and its catchment using
1003 PARAFAC analysis, *Limnology & Oceanography*, *50*(2), 686-697.

1004 Stedmon, C. A., R. M. W. Amon, A. J. Rinehart, and S. A. Walker (2011), The supply
1005 and characteristics of colored dissolved organic matter (CDOM) in the Arctic
1006 Ocean: Pan Arctic trends and differences, *Marine Chemistry*, *124*(1-4), 108-118.

1007 Swan, C. M., N. B. Nelson, D. A. Siegel, and T. S. Kostadinov (2012), The effect of
1008 surface irradiance on the absorption spectrum of chromophoric dissolved
1009 organic matter in the global ocean, *Deep Sea Research Part I: Oceanographic*
1010 *Research Papers*, *63*, 52-64.

1011 Twardowski, M. S., E. Boss, J. M. Sullivan, and P. L. Donaghay (2004), Modeling the
1012 spectral shape of absorption by chromophoric dissolved organic matter, *Marine*
1013 *Chemistry*, *89*(1-4), 69-88.

1014 Visser, S. A. (1983), Fluorescence phenomena of humic matter of aquatic origin and
1015 microbial cultures, *Aquatic and terrestrial humic materials*, 183-202.

1016 Walker, S. A., R. M. W. Amon, and C. A. Stedmon (2013), Variations in high-latitude
1017 riverine fluorescent dissolved organic matter: A comparison of large Arctic rivers,
1018 *Journal of Geophysical Research: Biogeosciences*, *118*(4), 1689-1702.

1019 Werdell, P. J., S. W. Bailey, G. S. Fargion, C. Pietras, K. D. Knobelspiesse, G. C. Feldman,
1020 and C. R. McClain (2003), Unique data repository facilitates ocean color satellite
1021 validation, *EOS Trans. AGU*, *84*(38), 377.

1022 Wünsch, U. J., K. R. Murphy, and C. A. Stedmon (2015), Fluorescence Quantum Yields
1023 of Natural Organic Matter and Organic Compounds: Implications for the
1024 Fluorescence-based Interpretation of Organic Matter Composition, *Frontiers in*
1025 *Marine Science*, *2*.

1026 Xiu, P., and F. Chai (2014), Connections between physical, optical and
1027 biogeochemical processes in the Pacific Ocean, *Progress in Oceanography*, *122*,
1028 30-53.

1029

1030

Figure Captions

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.

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.

3. 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.

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.

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

6. Global maps displaying provinces grouped based on results from the multiple comparison of means test. Connecting lines indicate provinces within the group are statistically similar. Province color indicates the number of provinces found to have a statistically similar mean slope based on the multiple comparison of means test. Provinces that appear more than once were found to have a mean and uncertainty in the mean (described in the Methods section) overlapping with provinces in two different groups. Subplots display corresponding maps and bioplots as (a,b) Cluster 1 ($S_{E275:295}$=0.0244 nm^{-1}), (c,d) Cluster 2 ($S_{E275:295}$=0.0440 nm^{-1}), (e,f) Cluster 3 ($S_{E275:295}$=0.0399 nm^{-1}), (g,h) Cluster 4 ($S_{E275:295}$=0.0485 nm^{-1}), (i,j) Cluster 5 ($S_{E275:295}$=0.0267 nm^{-1}), (k, bioplot only) SANT and ANTA ($S_{E275:295}$=0.0313 nm^{-1}) and (l) indicates provinces with unique mean $S_{275:295}$ 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}).

7. Global maps displaying provinces grouped based on results from the multiple comparison of means test. Connecting lines indicate provinces within the group are statistically similar. Province color indicates the number of provinces found to have a statistically similar mean slope based on the multiple comparison of means test. Provinces that appear more than once were found to have a mean and uncertainty in the mean (described in the Methods section) overlapping with

provinces in two different groups. Subplots display corresponding maps and bioplots as (a,b) Cluster 1 ($S_{E350:400}$=0.0176 nm^{-1}), (c,d) Cluster 2 ($S_{E350:400}$=0.0210 nm^{-1}) and (e,f) Cluster 3 ($S_{E350:400}$=0.0230 nm^{-1}) to indicate groups with 2 or more statistically similar province means while (g) APLR displayed a unique mean $S_{E350:400}$ value (0.0230 nm^{-1}).

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

9. Global maps displaying provinces grouped based on results from the multiple comparison of means test. Connecting lines indicate provinces within the group are statistically similar. Province color indicates the number of provinces found to have a statistically similar mean slope based on the multiple comparison of means test. Provinces that appear more than once were found to have a mean and uncertainty in the mean (described in the Methods section) overlapping with provinces in two different groups. Subplots display corresponding maps and bioplots as (a,b) Cluster 1 ($S_{G350:550}$=0.0167 nm^{-1}), (c,d) Cluster 2

($S_{G350:550}$=0.0177 nm^{-1}), (e,f) Cluster 3 ($S_{G350:550}$=0.0156 nm^{-1}) and (g,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}).

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

11. 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 (a,d,g) and second (b,e,h) optical depths and below the photic zone (c,f,i).

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}.

Figure 1.

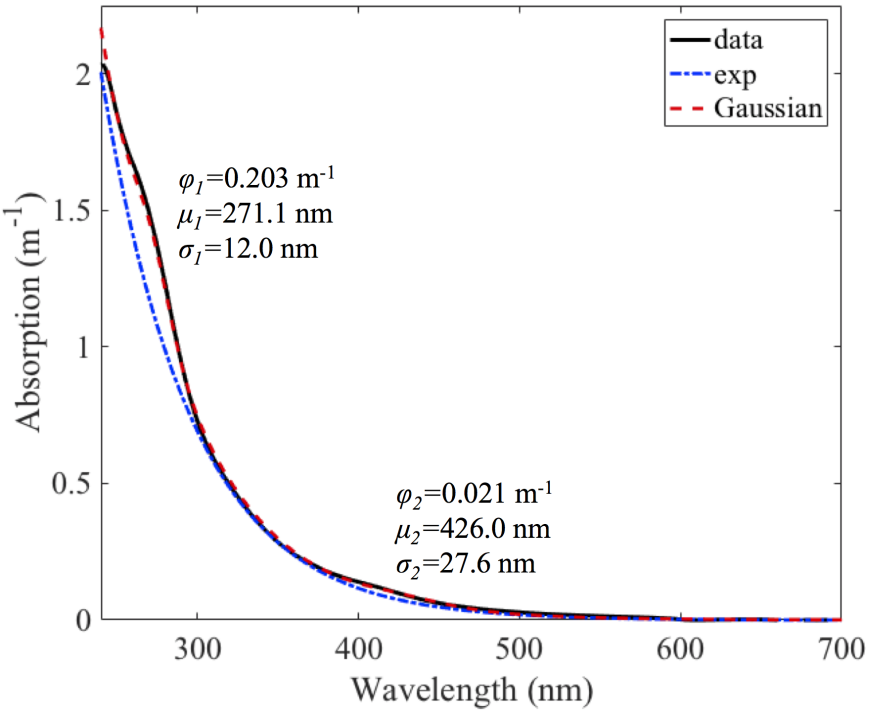

data
exp
Gaussian
φ_1=0.203 m^{-1}
μ_1=271.1 nm
σ_1=12.0 nm
φ_2=0.021 m^{-1}
μ_2=426.0 nm
σ_2=27.6 nm
Absorption (m^{-1})
Wavelength (nm)
0
0.5
1
1.5
2
300
400
500
600
700

Figure 2.

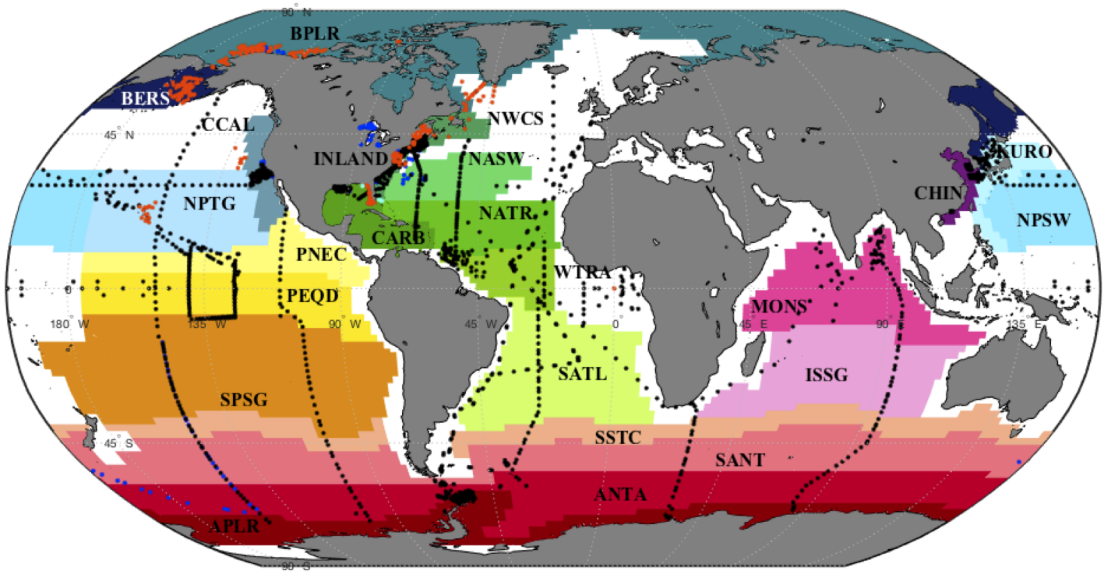

90° N
BPLR
BERS
CCAL
45° N
NWCS
INLAND
NASW
KURO
CHIN
NPTG
NATR
NPSW
CARB
PNEC
WTRA
PEQD
0°
180° W
135° W
90° W
45° W
0°
45° E
90°
135° E
MONS
SATL
ISSG
SPSG
SSTC
45° S
SANT
ANTA
APLR
90° S

Figure 3.

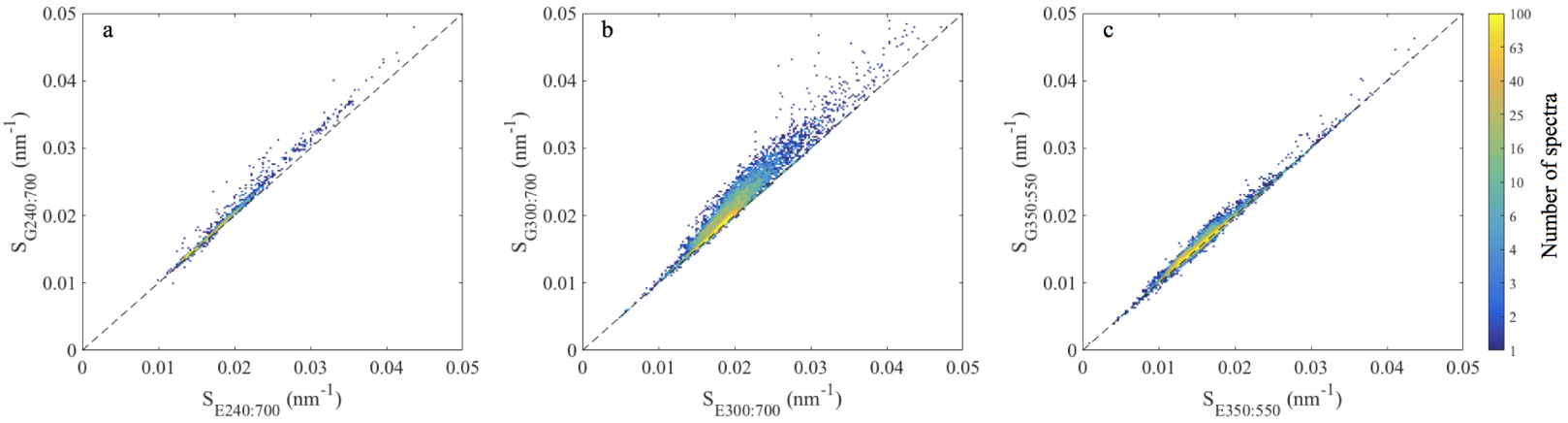

a
b
c
$S_{G240:700}$ (nm^{-1})
$S_{E240:700}$ (nm^{-1})
$S_{G300:700}$ (nm^{-1})
$S_{E300:700}$ (nm^{-1})
$S_{G350:550}$ (nm^{-1})
$S_{E350:550}$ (nm^{-1})
0
0.01
0.02
0.03
0.04
0.05
Number of spectra
100
63
40
25
16
10
6
4
3
2
1

Figure 4.

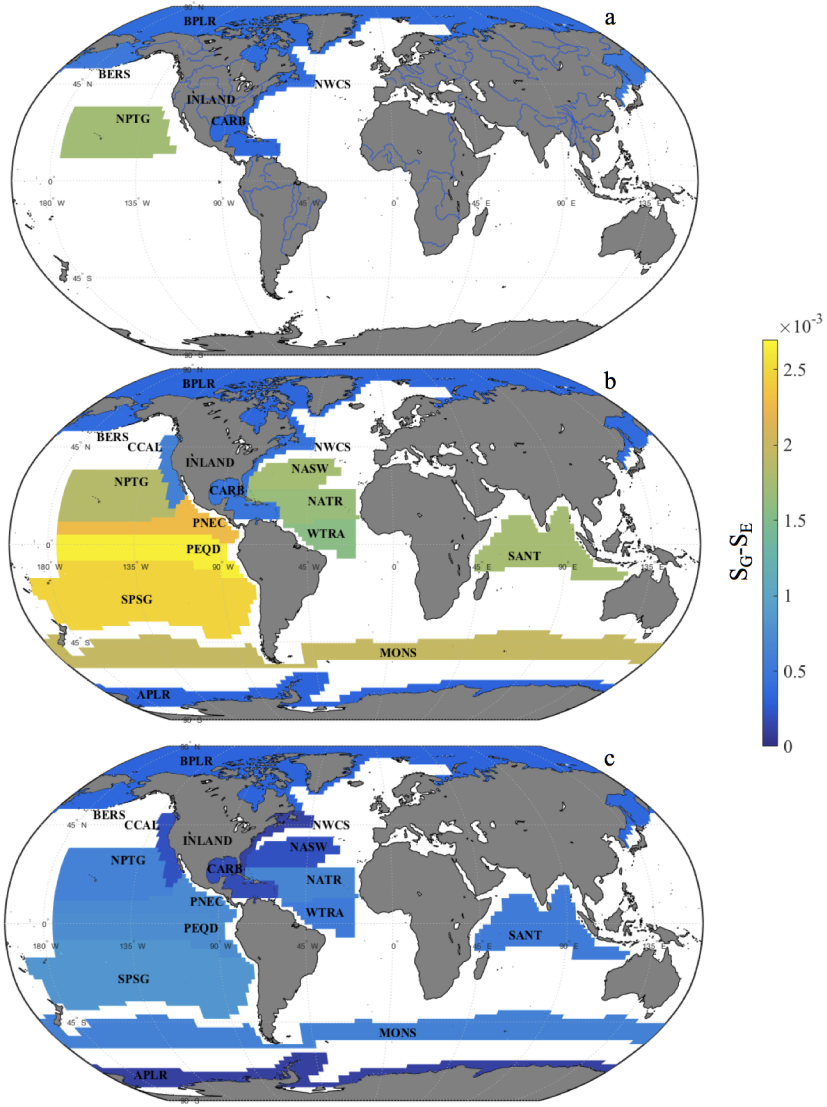

a
BPLR
BERS
NWCS
INLAND
NPTG
CARB
b
BPLR
BERS
CCAL
NWCS
INLAND
NASW
NPTG
CARB
NATR
PNEC
WTRA
PEQD
SANT
SPSG
MONS
APLR
c
BPLR
BERS
CCAL
NWCS
INLAND
NASW
NPTG
CARB
NATR
PNEC
WTRA
PEQD
SANT
SPSG
MONS
APLR
×10-3
2.5
2
1.5
1
0.5
0
SG-SE

Figure 5.

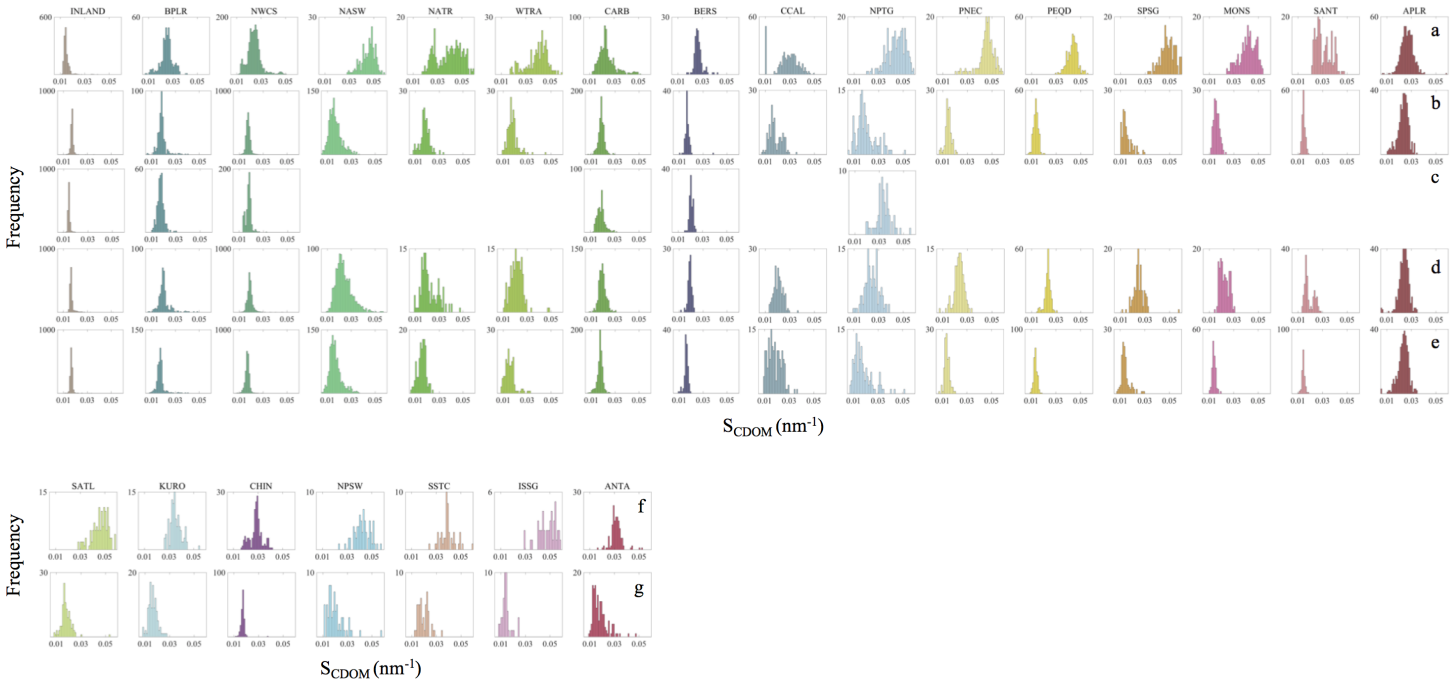

INLAND
BPLR
NWCS
NASW
NATR
WTRA
CARB
BERS
CCAL
NPTG
PNEC
PEQD
SPSG
MONS
SANT
APLR
a
b
c
d
e
Frequency
S_{CDOM} (nm^{-1})
SATL
KURO
CHIN
NPSW
SSTC
ISSG
ANTA
f
g
Frequency
S_{CDOM} (nm^{-1})

Figure 6.

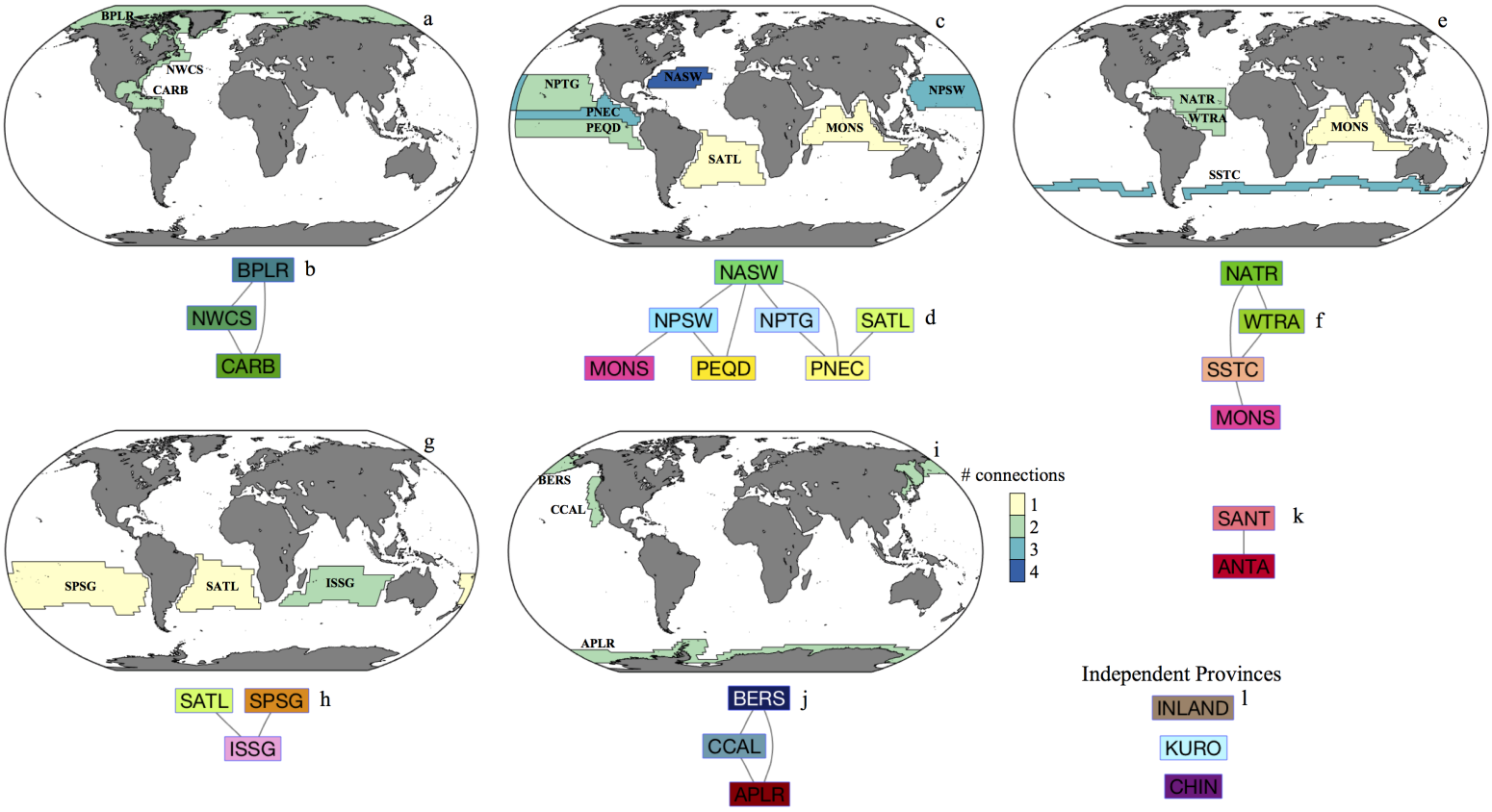

a
BPLR
NWCS
CARB
BPLR
NWCS
CARB
b
c
NPTG
NASW
NPSW
PNEC
PEQD
MONS
SATL
NASW
NPSW
NPTG
SATL
d
MONS
PEQD
PNEC
e
NATR
WTRA
MONS
SSTC
NATR
WTRA
f
SSTC
MONS
g
SPSG
SATL
ISSG
SATL
SPSG
h
ISSG
i
BERS
CCAL
APLR
connections
1
2
3
4
BERS
CCAL
APLR
j
SANT
k
ANTA
Independent Provinces
INLAND
l
KURO
CHIN

Figure 7.

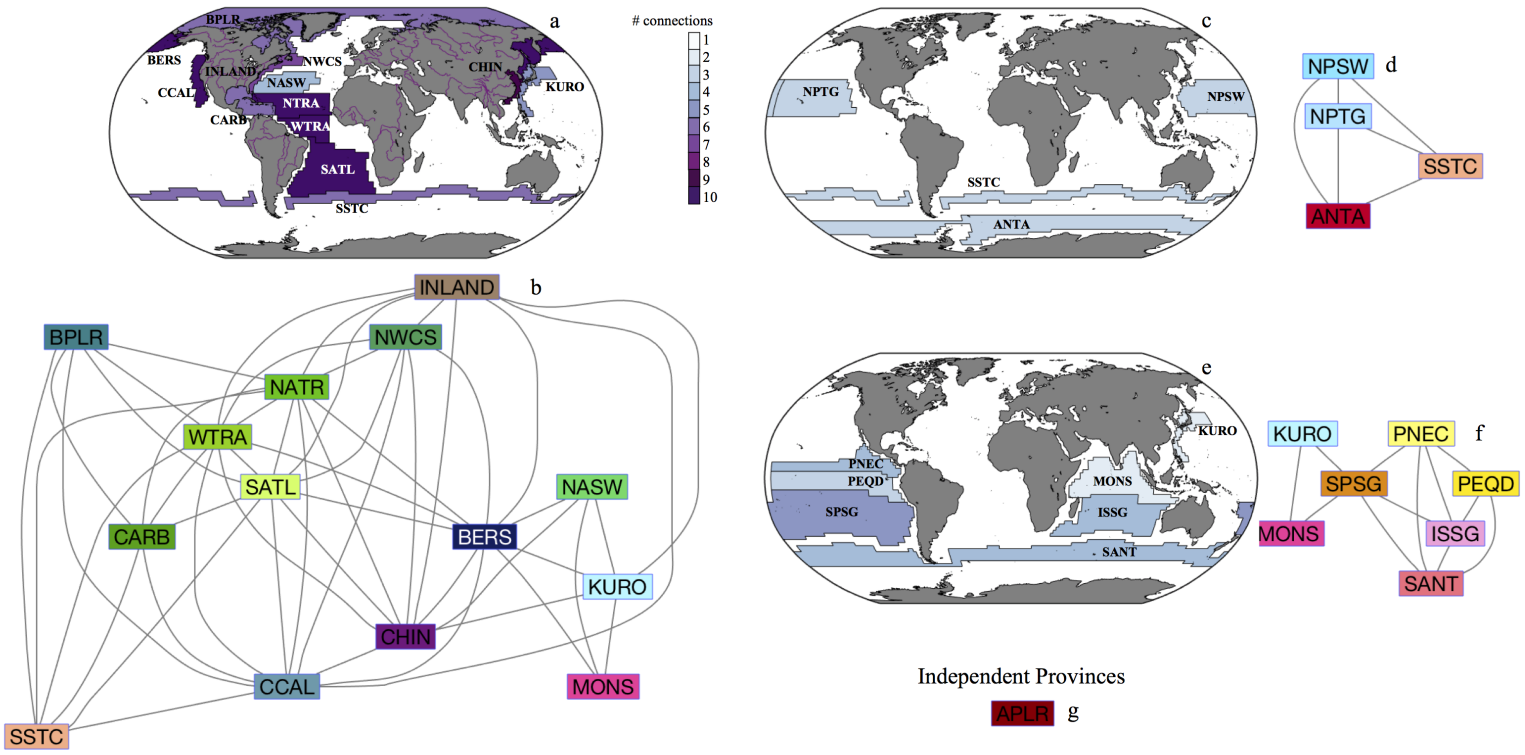
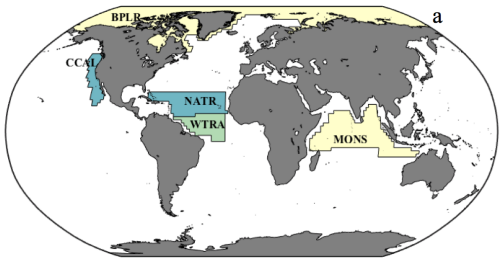
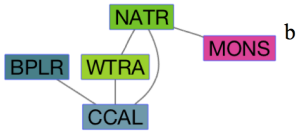

a
BPLR
BERS
INLAND
NWCS
CHIN
NASW
KURO
CCAL
NTRA
CARB
WTRA
SATL
SSTC
connections
1
2
3
4
5
6
7
8
9
10
b
INLAND
BPLR
NWCS
NATR
WTRA
SATL
NASW
CARB
BERS
KURO
CHIN
CCAL
MONS
SSTC
c
NPTG
NPSW
SSTC
ANTA
d
NPSW
NPTG
SSTC
ANTA
e
KURO
PNEC
PEQD
MONS
SPSG
ISSG
SANT
f
KURO
PNEC
SPSG
PEQD
MONS
ISSG
SANT
Independent Provinces
APLR
g

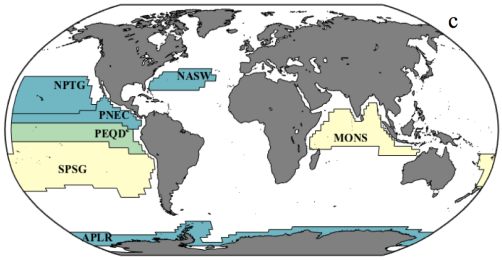
Figure 8.



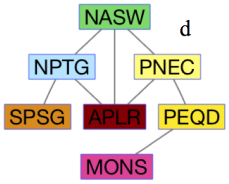
a



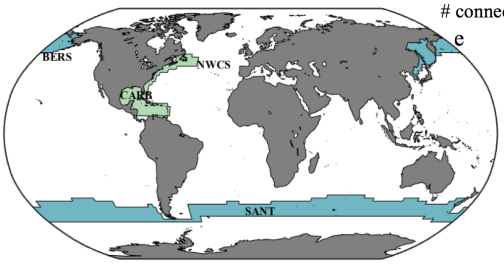
b



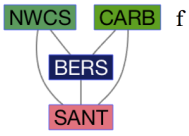
c



d



e

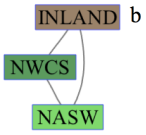


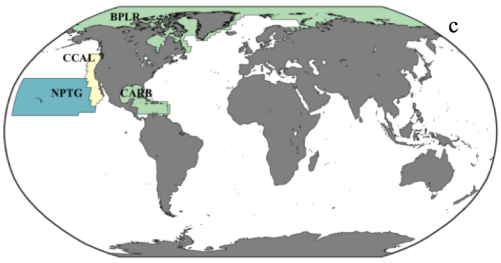
f

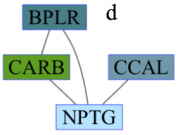
Independent Provinces

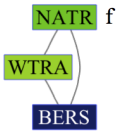
g

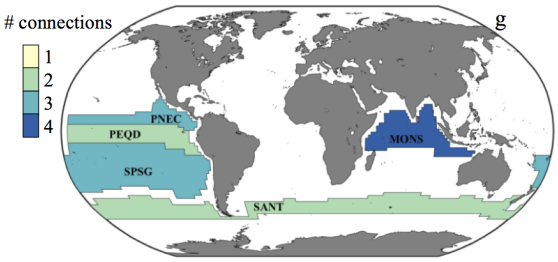
Figure 9.

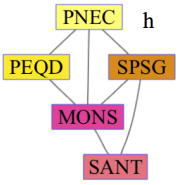












Independent Provinces

Figure 10.

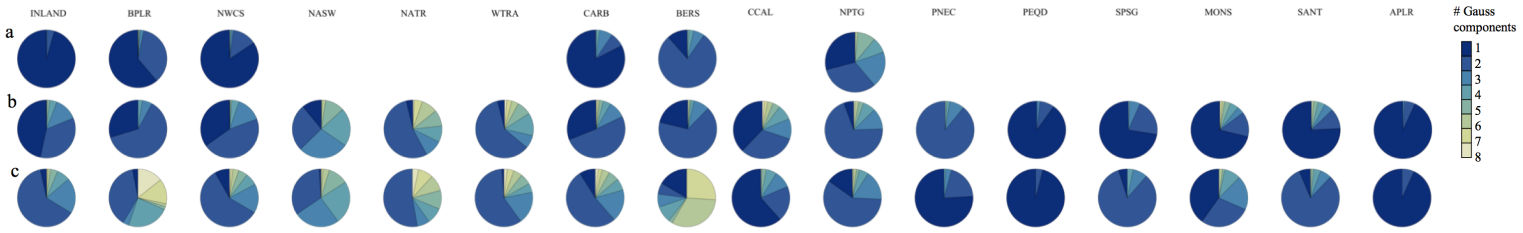

INLAND
BPLR
NWCS
NASW
NATR
WTRA
CARB
BERS
CCAL
NPTG
PNEC
PEQD
SPSG
MONS
SANT
APLR
a
b
c
Gauss components
1
2
3
4
5
6
7
8

Figure 11.

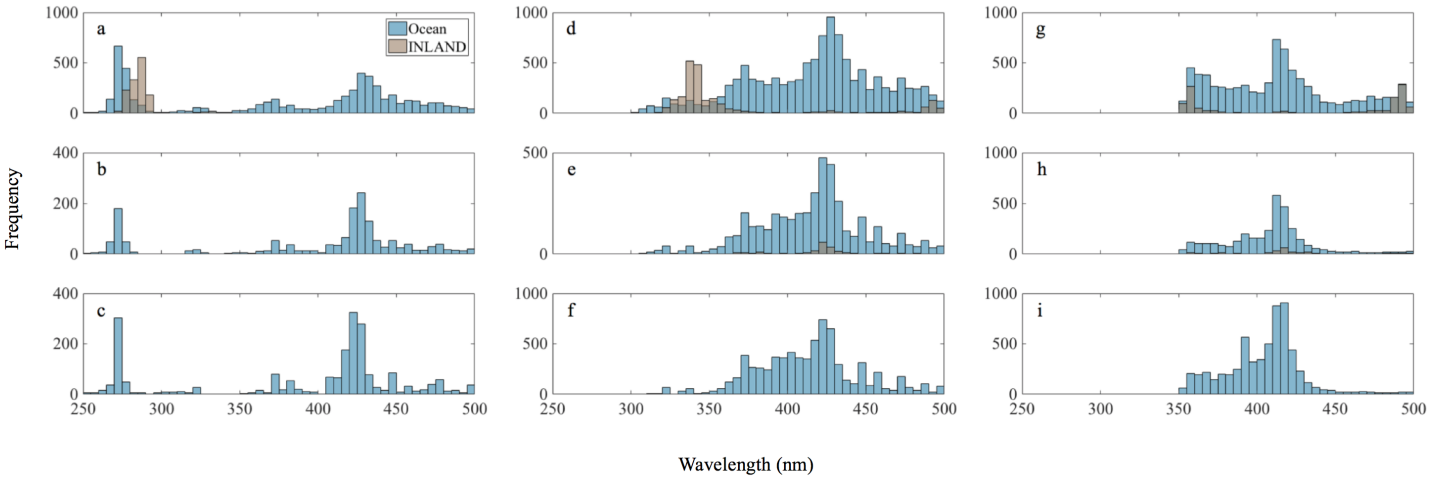
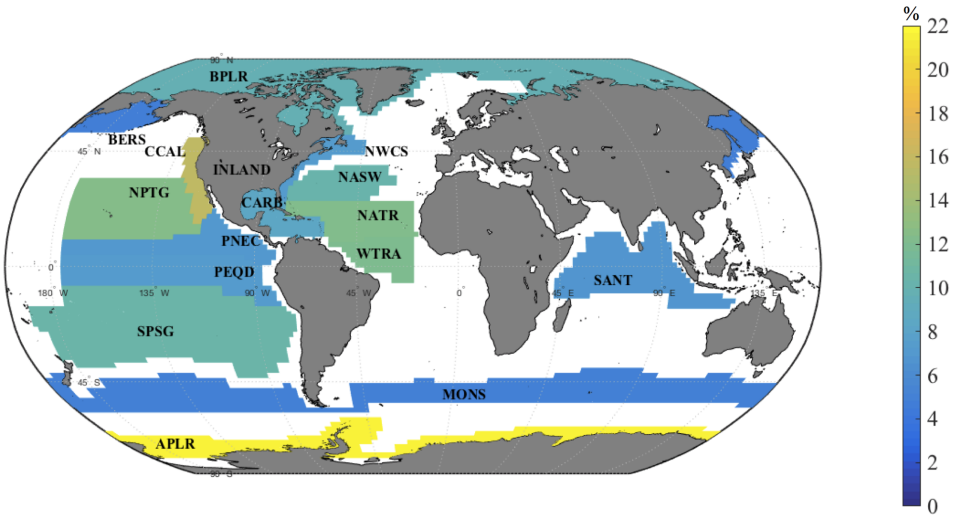

a
b
c
d
e
f
g
h
i
Ocean
INLAND
Frequency
Wavelength (nm)

Figure 12.


90 N
BPLR
BERS
45 N
CCAL
INLAND
NWCS
NASW
NPTG
CARB
NATR
PNEC
WTRA
0
PEQD
180 W
135 W
90 W
45 W
0
45 E
90 E
SANT
SPSG
45 S
MONS
APLR
90 S
%
22
20
18
16
14
12
10
8
6
4
2
0