

## Identification of dissolved organic matter size components in freshwater and marine environments

Eero Asmala , <sup>1\*</sup> Philippe Massicotte, <sup>2</sup> Jacob Carstensen <sup>3</sup>

<sup>1</sup>Tvärminne Zoological Station, University of Helsinki, Hanko, Finland

<sup>2</sup>Takuvik Joint International Laboratory / UMI 3376, Université Laval & Centre National de la Recherche Scientifique, Québec, Canada

<sup>3</sup>Department of Bioscience, Aarhus University, Roskilde, Denmark

### Abstract

Dissolved organic matter (DOM) in the transition zone from freshwater to marine systems was analyzed with a new approach for parameterizing the size distribution of organic compounds. We used size-exclusion chromatography for molecular size analysis and quantified colored DOM (CDOM) on samples from two coastal environments in the Baltic Sea (Roskilde Fjord, Denmark and Gulf of Gdansk, Poland). We applied a Gaussian decomposition method to identify peaks from the chromatograms, providing information beyond bulk size properties. This approach complements methods where DOM is separated into size classes with pre-defined filtering cutoffs, or methods where chromatograms are used only to infer average molecular weight. With this decomposition method, we extracted between three and five peaks from each chromatogram and clustered these into three size groups. To test the applicability of our method, we linked our decomposed peaks with salinity, a major environmental driver in the freshwater-marine continuum. Our results show that when moving from freshwater to low-salinity coastal waters, the observed steep decrease of apparent molecular weight is mostly due to loss of the high-molecular-weight fraction (HMW; >2 kDa) of CDOM. Furthermore, most of the CDOM absorbance in freshwater originates from HMW DOM, whereas the absorbing moieties are more equally distributed along the smaller size range (< 2 kDa) in marine samples.

The aquatic pool of organic carbon is one of the largest dynamic carbon reservoirs on Earth, comparable to the atmospheric CO<sub>2</sub> reservoir (Jiao et al. 2010). Most of this aquatic organic carbon is in dissolved form (dissolved organic carbon, DOC), which is the main fraction of dissolved organic matter (DOM) and often used as a proxy to understand overall DOM biogeochemistry. Riverine inputs are significant sources of terrestrial DOM to the oceanic carbon pool (Cole et al. 2007). During the passage along the estuarine gradient, the DOM dynamics are driven by two different general mechanisms: mixing of freshwater with seawater and biogeochemical processing (Stedmon and Markager 2003; Boyd and Osburn 2004; Asmala et al. 2016). Deviations from conservative mixing between fresh and saline end-members can be utilized as indicators for biogeochemical processing along the

salinity gradient (Massicotte et al. 2017). Previously, it has been shown that apparent molecular weight (i.e., the size of the DOM molecules) does not conform with conservative mixing, but decreases rapidly with increasing salinity, and faster than the colored fraction of DOM (CDOM) (Sholkovitz et al. 1978; Zhou et al. 2016; Asmala et al. 2018). This implies that the molecular composition of DOM can change in a profound way during transport from land to sea and importantly, these changes can be observed with optical measurements. There are large differences in typical DOM molecular size distributions between freshwater and marine systems. The proportion of large DOM molecules (molecular weight > 1 kDa) of the total DOM pool decreases from 50% to 20% along the estuarine gradient from freshwater source to sea (Benner and Opsahl 2001). Further, during the estuarine transit, the average DOM molecular weight decreases from 1000 to 1500 Da in the freshwater end-member to 300–500 Da in the coastal sea end-member (Asmala et al. 2016). The size spectrum of organic compounds in freshwater and marine systems has been proposed to follow the so-called size-reactivity continuum (Amon and Benner 1996), where larger dissolved molecules are more reactive and degraded or removed first by various biogeochemical processes such as photodegradation

\*Correspondence: eero.asmala@helsinki.fi

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

Additional Supporting Information may be found in the online version of this article.

(Dalzell et al. 2009), bacterial mineralization (Moran et al. 1999) and flocculation (Sholkovitz et al. 1978). The DOM pool can be classified according to the size of the molecules, which is typically carried out by choosing an operational cutoff based on, e.g., filter membrane pore sizes. For instance, low and high-molecular-weight fractions (LMW and HMW, respectively) can be discriminated using a cutoff at 1000 Da (Amon and Benner 1996). Additional intermediate classifications have been used, such as medium molecular weight (MMW) in between LMW and HMW (1000–4000 Da; Ogawa 2000; Malik et al. 2016), and even very low molecular weight (VLMW) has been used, which is in the lower end of the DOM size spectrum (< 200 Da; Mayorga and Aufdenkampe 2002). As these cutoffs are based on technical and operational selection processes, they might systematically ignore relevant information contained in the continuous size spectrum of DOM.

To overcome the limitations of applying such thresholds where only a few discrete DOM size classes are examined, methods have been employed to examine the continuous size spectrum of the natural organic matter; size-exclusion chromatography being one of the most common ones. In size-exclusion chromatography, molecules are separated according to their size when a sample solution flows through a column of porous gel material (Kirkland and Antle 1977; Kostanski et al. 2004). Principally, as the sample passes through the column, the small compounds permeate the pores of the matrix to a larger extent than the larger compounds and are retained longer within the column, resulting in the largest molecules eluting first, and the smallest last. This method is suitable for natural samples with high variability in DOM size (Hongve et al. 1996). There are multiple variations in size-exclusion chromatography methodology regarding the technical details of the analysis. Among them, the choice of size separation columns, eluents and detectors have important implications on the results derived from the analysis (Peuravuori and Pihlaja 1997; O'Loughlin and Chin 2001). The eluent is the aqueous solvent carrying analytes through the gel matrix of the size-exclusion chromatography column and past the detector. The choice of eluent plays a critical role in studying aquatic organic matter with size-exclusion chromatography methods (Peuravuori and Pihlaja 1997). The most important effect of the choice of the eluent is the resolution in the chromatograms of natural organic matter, as some eluents are able to produce more detailed chromatograms than others. This is due to the non-size-exclusion interactions of humic substances in DOM with the gel matrix in the column, which can be effectively suppressed by adding electrolytes and hydrophobic solvents to the eluent, leading to higher accuracy of the analysis (Swift 1999).

Size-separated DOM molecules are commonly detected and analyzed with absorbance or fluorescence detectors in size-exclusion chromatography setups (Shimotori et al. 2016). The chosen detector wavelength has also important consequences on the results. Most studies using size-exclusion

chromatography method have used a single wavelength (typically in the UV range, due to high signal-to-noise ratios) for organic matter detection. The few studies that have expanded the analytical window beyond one wavelength, have shown that the molecular size characteristics (such as apparent molecular weight) are wavelengths-dependent (O'Loughlin and Chin 2001; Yan et al. 2012; Wunsch et al. 2018). Other detection methods such as DOC detection (Huber and Frimmel 1994; Dittmar and Kattner 2003; Shimotori et al. 2016) and mass spectrometry (Minor et al. 2002; Wu et al. 2004) have been used to analyze the size-separated molecules. Each size-exclusion chromatography analysis yields a detailed chromatogram with a wealth of information, which is usually not utilized to its full potential with traditional metrics of, e.g., average molecular weight.

Optical measurements of bulk (non-size-separated) DOM are widely used in aquatic sciences due to their relatively low cost and labor intensity. A wide range of wavelengths (most commonly ranging between 250 and 450 nm) has been used to measure the absorbance of CDOM ( $a_{\text{CDOM}}$ ) for studying the CDOM dynamics in various aquatic environments. Typically, the highest level of detail in the  $a_{\text{CDOM}}$  spectra and best predictive power for other water quality variables (e.g., DOC concentration) have been found when using the ultraviolet (UV) spectral region (Massicotte et al. 2017). The choice of wavelengths (i.e., the spectral range) to measure  $a_{\text{CDOM}}$  can influence the results in aquatic biogeochemistry, but the underlying mechanisms behind this wavelength-dependency are uncertain. The tradeoff of the relative simplicity of this approach is its low chemical and structural specificity. However, in recent years, focused efforts have been made to link the optical properties of DOM to its chemical characteristics (Stubbins et al. 2015; Asmala et al. 2016; Osburn et al. 2016). Many recent advances in CDOM characterization have been made by moving on from single-wavelength proxies to a more in-depth analysis of the shape and features of the CDOM spectra (Loiselle et al. 2009; Reader et al. 2015; Massicotte and Markager 2016). Alongside with CDOM absorbance, fluorescence of CDOM has been used to study DOM biogeochemistry (Coble 1996; Stedmon et al. 2003). These approaches have been useful in assessing the biogeochemical processing of DOM in the aquatic environment.

The overall objective of this study was to examine the molecular weight dynamics of the DOM pool along two coastal salinity gradients using (a) multiple detector wavelengths in size-exclusion chromatography and (b) mathematical decomposition of the resulting size spectra. We aimed to increase the amount of information gained from individual size-exclusion chromatograms with the decomposition, while retaining the continuous spectral properties of natural DOM. With this novel approach, we further aimed to bridge the knowledge gap between the molecular weight and optical properties across the absorbance spectrum of DOM in coastal environments. Specifically, we wanted to resolve how

different size fractions of DOM contribute to the total absorbance of natural organic matter, and investigate the origin and fate of these different size fractions. We hypothesized that the relationship between apparent molecular weight and the chosen absorbance wavelength of natural DOM is different across the aquatic continuum from land to sea. To test this hypothesis, CDOM absorbance ( $a_{\text{CDOM}}$ ) was measured with both spectrofluorometric (bulk optical properties) and size-exclusion (size-resolved optical properties) methods. For additional insights about linkages between size and optical properties of DOM, we mathematically decomposed the chromatograms to individual components representing different size classes of chromophoric material. Finally, we evaluated the suitability of the discrete cutoff methods for studying DOM in the land-to-sea gradient.

## Materials and methods

### Study area

The two study sites were located in coastal Baltic Sea; Roskilde Fjord in Denmark and Gulf of Gdansk in Poland. The details and locations of the sampling campaigns can be found in Asmala et al. (2018) for the Roskilde Fjord and in Reader et al. (2019) for the Gulf of Gdansk. Briefly, samples from five streams and three marine sites in Roskilde Fjord were collected on eight sampling campaigns between November 2014 and November 2015, resulting in 43 samples analyzed with size-exclusion chromatography (27 marine samples and 16 freshwater samples). Alongside the sampling, basic physico-chemical parameters (temperature, pH, salinity) were measured with a hand-held multiparameter logger (ProDSS, YSI, Inc.). The sampling campaign in the Gulf of Gdansk was carried out in the plume of the Vistula River onboard the R/V Alkor in February 2015, resulting in 20 surface samples analyzed with size-exclusion chromatography. Water column properties were sampled using a Seabird 911plus CTD-rossette water sampling system. The dataset consisted of 63 samples in total, which were grouped into three salinity categories: freshwater (salinity 0), oligohaline (6–11) and mesohaline (14–25). Dissolved organic carbon (DOC), a proxy for DOM concentration, ranged from  $723 \pm 143$  (mean  $\pm$  1SD) in the freshwater end-member to  $401 \pm 112 \mu\text{mol L}^{-1}$  in the mesohaline end-member (Asmala et al. 2018).

### Laboratory analyses

For analysis of DOM properties, water samples were filtered through pre-combusted (450°C for 4 h) 0.7  $\mu\text{m}$  glass fiber filters (Whatman). An aliquot of the filtered sample for CDOM was stored in acid-washed HDPE vials at 4°C until absorbance measurements within 2 weeks from sampling. Another filtered fraction for the size-exclusion chromatography (SEC) measurement was stored in acid-washed glass vials at -20°C until analysis. The size-exclusion chromatography analyzer setup consisted of a Shimadzu HPLC system (Shimadzu Corporation, Kyoto, Japan) equipped with a

linear-type column (TSK G2000SW<sub>XL</sub> column, 7.8  $\times$  300 mm, 5  $\mu\text{m}$  particle size, Tosoh Bioscience GmbH), a guard column (Tosoh Bioscience GmbH), and a UV-Vis diode array (Shimadzu SPD-M10AVP) set to measure a range between 250 and 400 nm with 5 nm intervals. The eluent was 0.01 M acetate buffer at a pH of 7.00 (Vartiainen et al., 1987). This eluent type has been shown to retain a high-resolution level in the resulting chromatograms, compared to, e.g., concentrated NaCl-based eluents with higher ionic strength (Peuravuori and Pihlaja 1997). The system was calibrated using acetone (58 Da) and differently sized polystyrene sulphonate standards of 1.1, 3.61, 4.23, 6.52, and 10.6 kDa. Sample runs were calibrated daily. A log-linear calibration relationship was used over the apparent molecular weight (AMW) range for each wavelength individually (Fig. S1). From baseline-corrected and calibrated chromatograms we calculated the number-averaged and weight-averaged apparent molecular weights (AMW<sub>n</sub> and AMW<sub>w</sub>, respectively; Chin et al. 1994). Polydispersity ( $p$ ) was calculated as the ratio between these two (AMW<sub>w</sub>/AMW<sub>n</sub>) to assess the molecular weight distribution of the mixture of organic compounds of varying sizes (O'Loughlin and Chin 2001). Dissolved organic carbon (DOC), dissolved organic nitrogen, CDOM absorbance and fluorescence (peaks C and T; Coble 1996) data were from Asmala et al. (2018). Comparison between integrated area under the size-exclusion chromatogram curve and CDOM absorbance measurements yielded a linear correlation coefficient that ranged between 0.952 and 0.975 over the wavelength range 250–400 nm, indicating a high recovery of CDOM with the size-exclusion chromatography method (Fig. S2).

### Sample baseline and instrumental drift corrections

To correct for possible baseline drift occurring during the measurement after calibration, baseline correction was performed by using a linear regression calculated on the average absorbance values calculated at the beginning (60–120 s) and the end (1680–1740 s) of each measurement, when no signal from the absorbance detector was expected. Day-to-day instrumental drift was corrected for using the time difference between the daily acetone reference sample and the acetone sample in the calibration set (Fig. S3). The timing of the acetone peak was used in scaling chromatography column retention times to actual molecular weight values (in Da). Examples of calibrated chromatograms are shown in Fig. S4.

### Gaussian decomposition

A Gaussian decomposition of the chromatograms was performed based on a modified version of the method used to decompose CDOM absorbance spectra proposed in Massicotte and Markager (2016). Briefly, the procedure aims to decompose the measured absorbance of the chromatogram ( $A(x)$ ,  $x \in [0.8; 4]$  kDa) into a distinct number of fundamental Gaussian

components, each described by the probability density function with three parameters:

$$f(x, \mu, \sigma; \varphi) = \varphi e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (\text{Eq1})$$

where  $\mu$  is the position parameter for the center of the peak,  $\sigma$  is the SD parameter controlling the width of the component, and  $\varphi$  is the height parameter of the component peak. The chromatogram was estimated as a linear combination of a varying number ( $n_C$ ; up to 5) of Gaussian components:

$$A(x; \mu, \sigma, \varphi) = \sum_{i=0}^{n_C} \varphi_i e^{-\frac{(x-\mu_i)^2}{2\sigma_i^2}} + \varepsilon \quad (\text{Eq2})$$

where  $i = 1, \dots, n_C$  denotes a particular Gaussian component and  $\varepsilon$  are the residuals representing the variability not accounted for by the Gaussian components. The Bayesian information criterion was used to identify the optimal number of Gaussian components (Schwarz 1978). The Bayesian information criterion is based on the principle of parsimony, helping to identify the model that accounts for the most variation with the fewest parameters (or the fewest number of Gaussian components; Fig. 1). The optimal number of Gaussian components ranged between two and five. The parameters of the Gaussian components ( $\mu$ ,  $\sigma$ ,  $\varphi$ ) were estimated in Matlab using the “peakfit.m” toolbox (O’Haver 2020).

### Statistical analyses

Clustering analysis was used to classify identified Gaussian peaks. First, a hierarchical cluster analysis using the *hclust* function in R software (R Core Team 2019) was used to determine the number of peak classes, based on peak characteristics (position, width, and height) from the Gaussian decomposition. Using silhouette analysis (which measures how well an observation is clustered by estimating the average distance between clusters; Kaufman and Rousseeuw 2009), the optimal number of three clusters was defined. This is visualized by the silhouette plot, displaying a measure of how close each point in one cluster is to points in the neighboring clusters (Fig. S5). Second, a *k*-means clustering with the predefined number of clusters from the previous analysis was carried out using the *kmeans* function in R software. The *k*-means analysis partitions the points into groups by minimizing the sum of squares from points to their assigned cluster centers using the algorithm of Hartigan and Wong (1979). Each peak was assigned to one of the three distinct size classes based on the Gaussian decomposition metrics.

## Results

### Dissolved organic matter size distribution along the land-to-sea gradient

Apparent molecular weight measured at 250 nm ( $AMW_{250}$ ) along the salinity ranged between 120 and 1252 Da (Fig. 2a). The mean  $AMW_{250}$  value for freshwater samples was 1010 Da

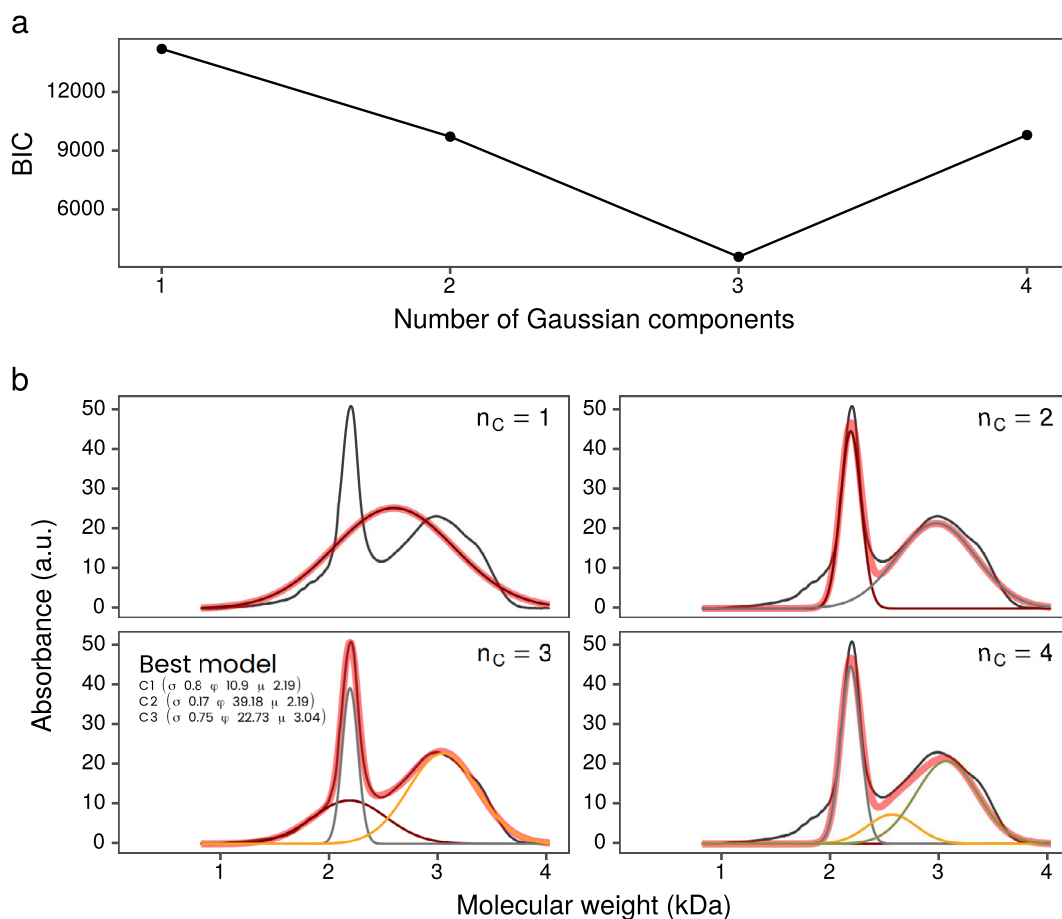
(range 670–1252 Da), while for estuarine and marine samples the mean value was 240 Da (range 120–489 Da). Strong non-conservative behavior of  $AMW_{250}$  was observed, as even the lowest freshwater values were considerably higher than predicted by linear regression of the marine  $AMW_{250}$  values vs. salinity (Fig. 2a; intercept at  $378 \pm 37$  Da). Polydispersity, a proxy for heterogeneity of molecular weight distribution within samples, ranged from 2.08 to 3.21 (higher values indicating higher heterogeneity; Fig. 2b). The lowest values were observed in the freshwater endmember, and the highest values in mid-salinities (salinity 5–15), after which polydispersity decreased again. Despite that samples were collected over an annual cycle in Roskilde Fjord, no apparent seasonality was observed (Fig. S6).

### Effect of detector wavelength on bulk size parameters

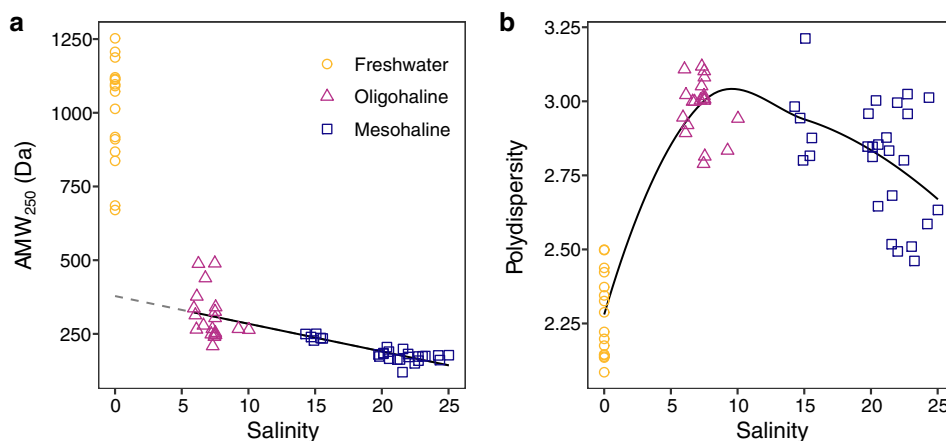
Estimated AMW varied along the spectral range used to measure CDOM (Fig. 3a). In all sample types, the lowest average AMW values were observed at 250 nm. In freshwater samples, AMW increased linearly with wavelength from  $\sim 1000$  Da at 250 nm to  $\sim 1300$  Da at 400 nm. With increasing salinity, the position of maximum AMW shifted towards lower wavelengths: 320 and 280 nm for oligohaline and mesohaline samples, respectively. Overall, there was a linear decrease of the detector wavelength at the maximum value of  $AMW_\lambda$  across the salinity range 0–25 (Fig. 3b).

### Decomposition of size-exclusion chromatograms

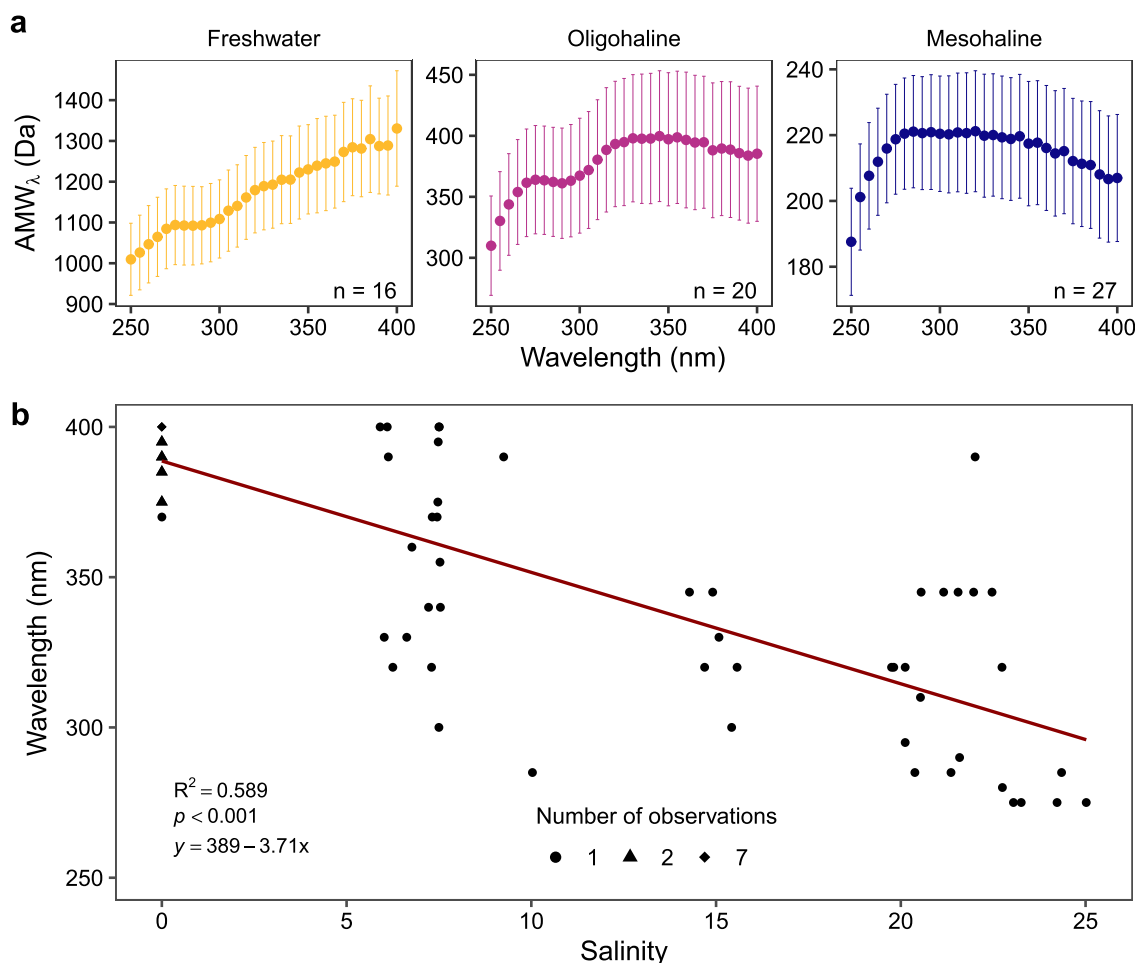
We decomposed the 63 chromatograms into 3–5 distinct peaks per sample, the total number of peaks ranging from 158 at detector wavelength of 250 nm to 111 at 350 nm. The identified Gaussian peaks measured at 250 nm were distributed along a size range from 60 to 4000 Da (Fig. S7). Higher local densities of peaks were found at around 130, 1100, and 3100 Da. First, Gaussian peaks were assigned to three predefined size classes (Fig. S7a): LMW, low molecular weight (< 1000 Da), MMW, medium molecular weight (1000–2000 Da); and HMW, high molecular weight (> 2000 Da), based on classifications used in previous studies (Amon and Benner 1996; Ogawa 2000; Mayorga and Aufdenkampe 2002; Malik et al. 2016). Interestingly, the Gaussian peaks were also optimally partitioned into three separate clusters (*k*-means clustering; Fig. S7) with boundaries shifting slightly compared to those established in the literature. The boundary between LMW and MMW shifted from 1000 to 800 Da, whereas the boundary between MMW and HMW did not change (Fig. S7). Our revised grouping criteria were used in subsequent analysis. Detector wavelength had an influence on the peak position, as all peak clusters shifted towards higher molecular weights with increasing wavelength (Table 1). Peaks belonging to LMW DOM were identified from all samples. On the other hand, peaks of the largest DOM size class (HMW) were found in all freshwater samples, but only in 3 out of 27 of marine samples.



**Fig. 1.** Illustration of the Gaussian decomposition method. Equation ((2)) was estimated for  $n_C = 1-4$  Gaussian components and the optimal model had the lowest Bayesian information criterion ( $n_C = 3$ ; panel **a**). In panel **b**, the black line is the measured absorbance and the thick red line is the sum of all Gaussian components (thin lines of different colors). Parameters of the three Gaussian components are inserted for  $n_C = 3$ .



**Fig. 2.** DOM molecular weight characteristics across salinity: **(a)** apparent molecular weight and **(b)** polydispersity at 250 nm detector wavelength. Freshwater and mesohaline samples are from Roskilde Fjord and oligohaline from Vistula Estuary. The linear regression line (excluding freshwater samples) in panel **(a)** emphasizes the non-conservative behavior of apparent molecular weight (AMW) along the salinity gradient, with its extrapolation shown with a dashed line. In panel **(b)**, a non-parametric local smoothing function (LOESS) is applied for visualizing a possible relationship.



**Fig. 3.** Relationships between (a) apparent molecular weight (AMW<sub>λ</sub>) and detector wavelength (UV range; 250–400 nm) in freshwater, oligohaline, and mesohaline samples, and (b) detector wavelength of the maximum AMW<sub>λ</sub> and salinity. Symbols in panel (a) are mean values with error bars indicating one SD of the samples. Some observations in panel (b) were overlapping, as indicated with different symbols.

**Table 1.** Classification and summary statistics of the three clusters identified with the k-means clustering analysis of the positions of the Gaussian components. LMW, low molecular weight; MMW, moderate molecular weight; and HMW, high molecular weight. Mean value ± SD and range of the peak position for each Gaussian component is given for three detector wavelengths (250, 300, and 350 nm). For each salinity group, the number and proportion of samples with identified Gaussian components in the three weight clusters are shown (number of components in each sample did not change with detector wavelengths). Note that as 3–5 peaks were identified from each sample, multiple peaks from the same sample may be included in the same weight cluster.

Classification	Peak position (Da)			Number of samples with an identified component in the cluster		
	At 250 nm	At 300 nm	At 350 nm	Freshwater (n=16)	Oligohaline (n=20)	Mesohaline (n=27)
LMW	187 ± 130 (60–682)	221 ± 148 (67–740)	365 ± 256 (104–931)	18 (100%)	27 (100%)	32 (100%)
MMW	1233 ± 235 (801–1753)	1426 ± 287 (878–2114)	1664 ± 306 (1087–2340)	18 (100%)	27 (100%)	30 (94%)
HMW	2695 ± 419 (2073–3488)	3096 ± 450 (2365–3903)	3326 ± 366 (2677–4005)	18 (100%)	12 (44%)	3 (9%)

### Effect of salinity and detector wavelength on Gaussian peaks

The peak positions of the Gaussian components decreased with increasing salinity for each size cluster across measured wavelengths (Fig. 4). The peak positions of the largest DOM size class (HMW) decreased more steeply with increasing salinity, about three times faster than for MMW and four times faster than LMW. The mean molecular sizes decreased almost 1000 Da for HMW, and about 350 and 250 Da for MMW and LMW, respectively, across the salinity range. The peak position increased in all three size groups with increasing wavelength from 250 to 350 nm. Inversely, peak height decreased with increasing wavelength in all size groups, which was also the case with peak width. The other peak metrics (height and width) also varied along the salinity gradient. Overall, peak height of HMW decreased steeply with increasing salinity across wavelengths, MMW also decreased, but LMW either increased or did not change with salinity. The Gaussian components were also in general getting broader when moving towards the sea, but with some variability among size classes and detector wavelengths.

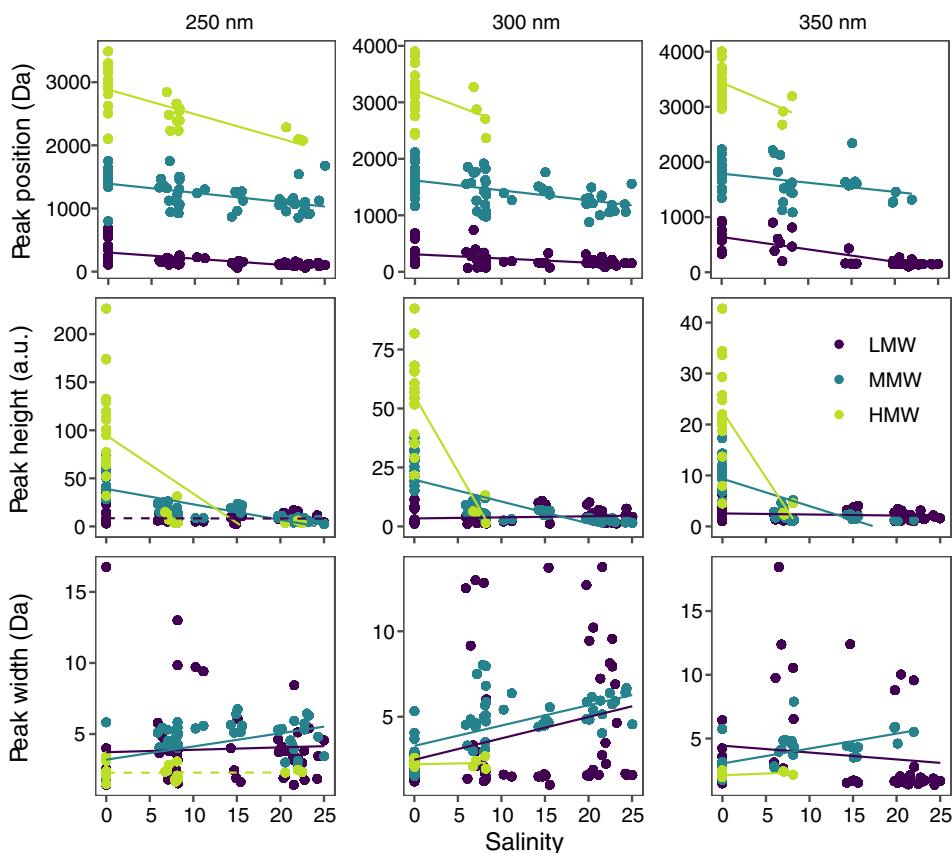
The contribution of each decomposed peak to the total absorbance of the sample was calculated from the integrated

area under the peak. The absorbance for each size class and their relative contribution to the total absorbance varied between sample types (Fig. 5). In freshwater, the absorbance of different size classes increased from 3.4 a.u. in LMW to 16.3 a.u. in MMW, and up to 40.6 a.u. in HMW (70% of the total freshwater absorbance). In saline samples, the highest absorbances were observed in MMW size class with 11.3 and 7.7 a.u. for oligo- and mesohaline samples, respectively (corresponding to 67 and 59% of the total absorbance). The contribution of the HMW size class to total absorbance was minor in saline samples (8 and 2% for oligo- and mesohaline samples, respectively).

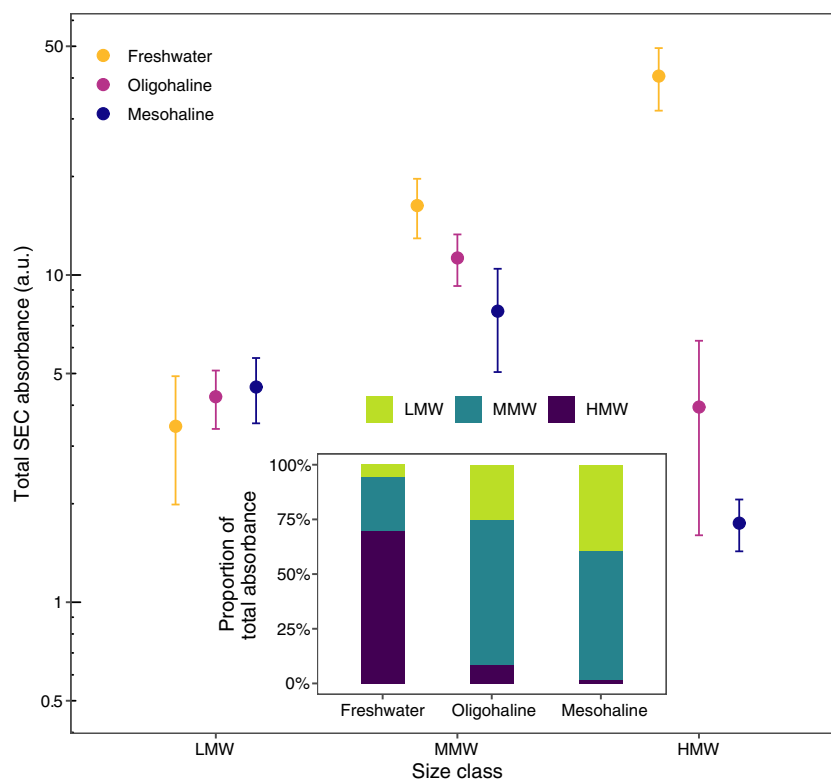
### Discussion

#### Chromatogram decomposition and peak clustering

Traditional methods focusing on DOM molecular size are typically not capturing the continuous nature of the size distribution, but relying on either operational cutoffs (e.g., tangential ultrafiltration techniques) or simplifying indices of the size spectrum (e.g., averaged molecular weight from size-exclusion chromatography). However, Shimotori et al. (2016) showed that photochemical characteristics of natural DOM depend on molecular sizes. In order to examine the



**Fig. 4.** Relationship between peak position, height and width of the Gaussian components and salinity for each size cluster at detector wavelengths 250, 300 and 350 nm. Regression values are presented in Table S1.



**Fig. 5.** Total absorbance (area under the Gaussian peak) of each size class at detector wavelength 250 nm. Insert shows the proportion of each size class of total absorbance at 250 nm for each three salinity classes. Error bars indicate 1 SD.

continuous DOM size distribution in more detail, we applied a decomposition method that aims to utilize all the potential information contained in the size-exclusion chromatograms. Decomposing analyte chromatograms is an established technique (Maeder 1987), but only recently it has been applied to natural CDOM absorbance spectra (Massicotte and Markager 2016; Omanović et al. 2019). By using these spectral decomposition techniques with size-exclusion spectra, we were able to separate distinct chromophore groups based on their size. We fitted between three and five Gaussian peaks to each chromatogram, and these peaks potentially reflect the number of different chromophore groups in the sample. The number of peaks was similar among the different detector wavelengths tested, indicating that each identified size-resolved chromophore group absorbs light to some extent across the measured CDOM spectrum (Sharpless and Blough 2014). The decomposed peaks were typically highly overlapping (Fig. 1), making it very challenging to obtain similar results from the complete chromatograms, thus emphasizing the value of information gained from the mathematical decomposition of the chromatograms. The peaks also show the heterogeneity of the chromophore sizes in the CDOM pool, resulting from its chemical diversity (Gonsior et al. 2017).

Despite the inherent chemical and compositional diversity of natural organic matter (Hertkorn et al. 2013; Kellerman et al. 2014), it can be expected that some compounds are more abundant than others in the environment. We used clustering analysis to group the decomposed peaks, and three groups with different peak positions emerged from the analysis (Fig. S7). This number of different size groups is consistent with previous size classifications based mostly on ultrafiltration methods (Amon and Benner 1996; Mayorga and Aufdenkampe 2002; Malik et al. 2016). Also the size ranges aligned well with the previous studies, and the changes in cut-offs separating groups from the cluster analysis were <200 Da. The distribution of the peaks was not consistent throughout the size range measured, as some sizes appear more frequently than others (Table 1). For instance, a peak at around 0.2 kDa (LMW) was found in all samples. This suggests a common presence of a chromophore or group of chromophores in this size range, such as fragments of lipids or carbohydrates from microbial DOC degradation (Ali and Tremblay 2019). On the other hand, the largest peak around 3 kDa was found in all freshwater samples, but only from a minor proportion of marine samples. This peak likely represents terrestrial “humic” substances, such as fulvic acids (Aiken et al. 1989; Huber et al. 2011).



### DOM molecular size distribution in different salinity regimes

Our results support earlier findings suggesting that the apparent molecular weight of DOM at 250 nm ( $AMW_{250}$ ) decreases drastically when moving from freshwater to marine systems (Fig. 2). A likely cause for this phenomenon is salt-induced flocculation, which is known to preferentially remove large compounds from the DOM pool (Sholkovitz et al. 1978; Asmala et al. 2014). Further, the hydrodynamic diameter of terrestrial humic substances may decrease due to intramolecular contraction or coiling induced by increasing ionic strength (Dittmar and Kattner 2003). Simultaneously, biological degradation and photolytic processes are changing the DOM pool towards smaller size and decreased reactivity (Amon and Benner 1996; Moran et al. 2000; Aarnos et al. 2012). As the apparent molecular weight for marine samples was lower than in freshwater samples, the contribution of small molecules to the bulk optical properties increased with increasing salinity (Batchelli et al. 2009). In addition to the average size of the DOM molecules, the size distribution of molecules is also an important characteristic of the DOM pool, and polydispersity is so far the most widely used indicator of the heterogeneity in molecular weight distribution (Chin et al. 1994). In both freshwater and mesohaline samples, the size spectra were dominated by one peak (Fig. S4), whereas in oligohaline samples there were two discernible peaks, resulting in a relatively large difference between number- and weight-averaged molecular weights, and thus high  $p$ . This supports the hypothesis about enhanced transformation of the DOM pool at low salinities, resulting in smaller average DOM molecules and higher heterogeneity in the size distribution. A potential mechanism for this is that with increasing salinity the proportion of smaller molecules increases in the DOM pool as large molecules are removed in early stages of mixing with seawater (Asmala et al. 2014). But part of the larger, terrestrial molecules escape this estuarine filter and results in increasing difference between number- and weight-averaged MW (i.e., polydispersity).

### Differences in spectral properties between freshwater and marine environments

By extending the analytical window beyond a single wavelength, we could confirm the previous findings of apparent molecular weight increasing almost linearly with increasing wavelengths in freshwater DOM (Fig. 3; see also Zhou et al. 2000; O'Loughlin and Chin 2001). This is the result of chromophores (i.e., compounds in DOM pool responsible for its color) absorbing light at higher wavelengths being larger in size compared to those absorbing in lower wavelengths. However, our results show that in saline coastal waters this relationship changes so that the  $AMW_{\lambda}$  does not increase linearly with wavelength, but the highest  $AMW_{\lambda}$  in coastal samples is observed at detection wavelengths between 280 and 320 nm (Fig. 3). In other words, CDOM molecules absorbing in UV-B

region (280–315 nm) are on average larger than molecules absorbing at UV-A wavelengths (315–400 nm) in marine samples.

The difference between freshwater and marine samples could be attributed to the effective removal of allochthonous material in the early stages of the estuarine salinity gradient. Terrestrial humic substances are characterized by larger molecules that absorb across a wide spectral range at UV-A and visible wavelengths (O'Loughlin and Chin 2001). Marine humic-like material, on the other hand, is derived from autochthonous processes and has different optical properties compared to terrestrial humic substances (Murphy et al. 2008). Our findings underline the importance of the selection of detection wavelength, as the inferred molecular weight of DOM will be strongly influenced by the wavelength. As this relationship is linear only in freshwater samples, extrapolation of molecular weight information from measurements done on one wavelength to another in saline samples is rather challenging. We speculate that the relatively large DOM molecules absorbing at higher wavelengths are present in freshwater, but removed in early stages in estuarine mixing (Kowalczyk et al. 2010; Massicotte et al. 2017).

The average molecular size of DOM decreases rapidly along the coastal salinity gradient from freshwater towards the open sea (Sholkovitz et al. 1978; Asmala et al. 2014). This change is typically non-conservative, indicating that the observed decrease cannot be attributed solely on the mixing with seawater, but other biogeochemical processes are affecting the DOM size distribution (Asmala et al. 2018). Salt-induced flocculation leads to partial removal of the largest DOM molecules via sedimentation to the seafloor (Jilbert et al. 2018), and larger DOM molecules have higher propensity for both photolytic and bacterial degradation (Lepane et al. 2003; Dalzell et al. 2009; Asmala et al. 2013). Our decomposed data confirm that the largest DOM size class (HMW) decreased most rapidly when moving from freshwater to marine environment (Fig. 4). It should be noted, that there are gaps in the salinity range our data covers, resulting in some uncertainties about the continuity of the processes along the coastal salinity gradient. Decreases in MMW and LMW are considerably smaller, indicating selective removal of molecules in the largest size class. Peaks were highest (i.e., the highest maximum absorbance) in freshwater samples in MMW and HMW size groups, but relatively constant in marine samples. The results also show that LMW components are present in all samples in all three salinity groups, and MMW size class in almost all samples (Table 1). The HMW size class is very rare in mesohaline samples, but still abundant in freshwater samples. In other words, the abundance of LMW remains high throughout the salinity gradient, MMW decreases to a minor extent towards higher salinities and the abundance of HMW decreases substantially. This explains the observed decrease in average molecular weight, which is not the result of the bulk of DOM getting smaller, but a disproportional removal of the largest

molecules. Indeed, the removal of these large compounds can be quite effective, as the HMW peaks in freshwater samples were not always present in oligo- and mesohaline samples. Such effective removal, observed in the transition from fresh to salt water, could be caused by salt-induced flocculation and subsequent aggregation into sinking particles (Forsgren et al. 1996; Jokinen et al. 2020).

The total absorbance of the samples was distributed unevenly among size classes. The peak area (i.e., total absorbance of the chromophore) varied 12-fold among size classes in freshwater samples, but only three- and fourfold in oligohaline and mesohaline samples, respectively (Fig. 5). Marine DOM is typically processed more extensively (i.e., is further down the diagenetic continuum) compared to freshwater DOM (Amon and Benner 1996). This is reflected in the proportional contribution of each size class to total absorbance, as the largest size class (HMW) is the most important in freshwater and MMW in marine waters. This is consistent with previous findings showing that molecules larger than 1 kDa are responsible for the majority of the total absorbance (Helms et al. 2008). As a result, the relative importance of small molecules in CDOM absorbance increases with distance from the river mouth towards the sea. Even the wavelength at which absorbance is measured had an effect on the observed DOM size characteristics (Fig. 4). Increasing detector wavelength typically leads to higher AMW values (Fig. 3; O'Loughlin and Chin 2001). These findings also provide potential explanations about the mechanism behind the observed tight relationship between molecular weight and spectral CDOM absorbance characteristics,  $E2 : E3$  ratio ( $A_{250} : A_{365}$ ) and slope  $S_{275-295}$  (De Haan and De Boer 1987; Peuravuori and Pihlaja 1997; Helms et al. 2008). Essentially, the relative contribution of chromophores absorbing at higher wavelengths, such as at 350 nm, have lesser impact on spectral CDOM as  $S_{275-295}$  and  $E2 : E3$  expectedly increase with salinities (Fig. 4 and Table S1). Moreover, as indicated by the regression intercept of the peak position, the apparent molecular weight of these chromophores is in general larger, which in turn results in smaller apparent molecular weight at high salinities, linked with higher  $S_{275-295}$  and  $E2 : E3$  values. Our data shows that the linear increase in AMW is the result of peak position shifting towards larger sizes across all detector wavelengths used, but holds true only in freshwater (Fig. 3).

## Conclusions

Despite that mathematical approaches for deconvoluting chromatograms have been used for decades in other disciplines, here we used the approach for the first time to assess the size-dependent changes in the optical properties of natural DOM in the aquatic continuum from land to sea. Our data show that different chromophores have different size ranges, and thus using different wavelengths for measurements will result in different CDOM molecules being detected. CDOM

absorbance results from different size groups in freshwater and marine waters, and there is no “universal” CDOM molecule, but instead a wide range of different absorbing molecules (chromophores). We also observed notable shifts in molecular weight from freshwater to the marine environment, which indicate large changes in the composition of chromophores responsible for CDOM absorbance. It is apparent that CDOM in freshwater results from relatively large molecules, whereas in marine systems small molecules are responsible for most CDOM absorbance. Our findings show that the relatively large molecules in freshwater samples are transformed and/or removed effectively in the early stages of estuarine mixing. The largest molecules are uncommon in the marine samples and hence, they are unlikely to be produced autochthonously in the coastal environment and most likely of terrestrial origin. We argue that studies using different CDOM wavelengths are essentially detecting different pools of chromophores, which are however partially overlapping. Since the molecular size is of key importance when analyzing biogeochemical cycling of DOM in aquatic environments, further information is needed to link size, chemical properties and optical characteristics of DOM, which could be achieved by the mathematical decomposition methods that provide details about the DOM size spectrum beyond the bulk properties.

## References

- Aarnos, H., P. Ylöstalo, and A. V. Vähätalo. 2012. Seasonal phototransformation of dissolved organic matter to ammonium, dissolved inorganic carbon, and labile substrates supporting bacterial biomass across the Baltic Sea. *J. Geophys. Res. Biogeo.* **117**: 1–14. <https://doi.org/10.1029/2010JG001633>
- Aiken, G., P. Brown, T. Noyes, and D. Pinckney. 1989. Molecular size and weight of fulvic and humic acids from the suwannee river, p. 163–178. *In* Humic substances in the Suwannee River, Georgia: Interactions, properties, and proposed structures. Citeseer.
- Ali, D. A. B., and L. Tremblay. 2019. HPLC-SEC-FTIR characterization of the dissolved organic matter produced by the microbial carbon pump. *Mar. Chem.* **215**: 103668. <https://doi.org/10.1016/j.marchem.2019.103668>
- Amon, R. M., and R. Benner. 1996. Bacterial utilization of different size classes of dissolved organic matter. *Limnol. Oceanogr.* **41**: 41–51. <https://doi.org/10.4319/lo.1996.41.1.0041>
- Asmala, E., D. G. Bowers, R. Autio, H. Kaartokallio, and D. N. Thomas. 2014. Qualitative changes of riverine dissolved organic matter at low salinities due to flocculation. *J. Geophys. Res. Biogeo.* **119**: 1919–1933. <https://doi.org/10.1002/2014JG002722>
- Asmala, E., H. Kaartokallio, J. Carstensen, and D. N. Thomas. 2016. Variation in riverine inputs affect dissolved organic matter characteristics throughout the estuarine gradient.

- Front. Mar. Sci. **2**: 125. <https://doi.org/10.3389/fmars.2015.00125>
- Asmala, E., L. Haraguchi, S. Markager, P. Massicotte, B. Riemann, P. A. Staehr, and J. Carstensen. 2018. Eutrophication leads to accumulation of recalcitrant autochthonous organic matter in coastal environment. *Global Biogeochem. Cycles* **32**: 1673–1687. <https://doi.org/10.1029/2017GB005848>
- Asmala, E., Autio, R., Kaartokallio, H., Pitkänen, L., Stedmon, C. A., and Thomas, D. N. 2013. Bioavailability of riverine dissolved organic matter in three Baltic Sea estuaries and the effect of catchment land use. *Biogeosciences* **10**, 6969–6986, <https://doi.org/10.5194/bg-10-6969-2013>.
- Batchelli, S., F. L. Muller, M. Baalousha, and J. R. Lead. 2009. Size fractionation and optical properties of colloids in an organic-rich estuary (Thurso, UK). *Mar. Chem.* **113**: 227–237. <https://doi.org/10.1016/j.marchem.2009.02.006>
- Benner, R., and S. Opsahl. 2001. Molecular indicators of the sources and transformations of dissolved organic matter in the Mississippi river plume. *Org. Geochem.* **32**: 597–611. [https://doi.org/10.1016/S0146-6380\(00\)00197-2](https://doi.org/10.1016/S0146-6380(00)00197-2)
- Boyd, T. J., and C. L. Osburn. 2004. Changes in CDOM fluorescence from allochthonous and autochthonous sources during tidal mixing and bacterial degradation in two coastal estuaries. *Mar. Chem.* **89**: 189–210. <https://doi.org/10.1016/j.marchem.2004.02.012>
- Chin, Y., G. Aiken, and E. O'Loughlin. 1994. Molecular weight, polydispersity, and spectroscopic properties of aquatic humic substances. *Environ. Sci. Technol.* **28**: 1853–1858. <https://doi.org/10.1021/es00060a015>
- Coble, P. G. 1996. Characterization of marine and terrestrial DOM in seawater using excitation-emission matrix spectroscopy. *Mar. Chem.* **51**: 325–346. [https://doi.org/10.1016/0304-4203\(95\)00062-3](https://doi.org/10.1016/0304-4203(95)00062-3)
- Cole, J.J., Prairie, Y.T., Caraco, N.F., et al. 2007. Plumbing the Global Carbon Cycle: Integrating Inland Waters into the Terrestrial Carbon Budget. *Ecosystems* **10**: 172–185. <https://doi.org/10.1007/s10021-006-9013-8>
- Dalzell, B. J., E. Minor, and K. Mopper. 2009. Photodegradation of estuarine dissolved organic matter: A multi-method assessment of DOM transformation. *Org. Geochem.* **40**: 243–257. <https://doi.org/10.1016/j.orggeochem.2008.10.003>
- De Haan, H., and T. De Boer. 1987. Applicability of light absorbance and fluorescence as measures of concentration and molecular size of dissolved organic carbon in humic Lake Tjeukemeer. *Water Res.* **21**: 731–734. [https://doi.org/10.1016/0043-1354\(87\)90086-8](https://doi.org/10.1016/0043-1354(87)90086-8)
- Dittmar, T., and G. Kattner. 2003. Recalcitrant dissolved organic matter in the ocean: Major contribution of small amphiphilics. *Mar. Chem.* **82**: 115–123. [https://doi.org/10.1016/S0304-4203\(03\)00068-9](https://doi.org/10.1016/S0304-4203(03)00068-9)
- Forsgren, G., M. Jansson, and P. Nilsson. 1996. Aggregation and sedimentation of iron, phosphorus and organic carbon in experimental mixtures of freshwater and estuarine water. *Estuar. Coast. Shelf Sci.* **43**: 259–268. <https://doi.org/10.1006/ecss.1996.0068>
- Gonsior, M., J. Luek, P. Schmitt-Kopplin, J. M. Grebmeier, and L. W. Cooper. 2017. Optical properties and molecular diversity of dissolved organic matter in the Bering Strait and Chukchi Sea. *Deep-Sea Res. II Top. Stud. Oceanogr.* **144**: 104–111. <https://doi.org/10.1016/j.dsr2.2017.01.003>
- Hartigan, J. A., and M. A. Wong. 1979. Algorithm AS 136: A k-means clustering algorithm. *J. R. Stat. Soc. Ser. C (Appl. Stat.)* **28**: 100–108. <https://doi.org/10.2307/2346830>
- Helms, J. R., A. Stubbins, J. D. Ritchie, E. C. Minor, D. J. Kieber, and K. Mopper. 2008. Absorption spectral slopes and slope ratios as indicators of molecular weight, source, and photobleaching of chromophoric dissolved organic matter. *Limnol. Oceanogr.* **53**: 955–969. <https://doi.org/10.4319/lo.2008.53.3.0955>
- Hertkorn, N., M. Harir, B. Koch, B. Michalke, and P. Schmitt-Kopplin. 2013. High-field NMR spectroscopy and FTICR mass spectrometry: Powerful discovery tools for the molecular level characterization of marine dissolved organic matter. *Biogeosciences* **10**: 1583–1624. <https://doi.org/10.5194/bg-10-1583-2013>
- Hongve, D., J. Baann, G. Becher, and S. Lømo. 1996. Characterization of humic substances by means of high-performance size exclusion chromatography. *Environ. Int.* **22**: 489–494. [https://doi.org/10.1016/0160-4120\(96\)00046-3](https://doi.org/10.1016/0160-4120(96)00046-3)
- Huber, S. A., and F. H. Frimmel. 1994. Direct gel chromatographic characterization and quantification of marine dissolved organic carbon using high-sensitivity DOC detection. *Environ. Sci. Technol.* **28**: 1194–1197. <https://doi.org/10.1021/es00055a035>
- Huber, S. A., A. Balz, M. Abert, and W. Pronk. 2011. Characterisation of aquatic humic and non-humic matter with size-exclusion chromatography–organic carbon detection–organic nitrogen detection (LC-OCD-OND). *Water Res.* **45**: 879–885. <https://doi.org/10.1016/j.watres.2010.09.023>
- Jiao, N., Herndl, G., Hansell, D. et al. 2010. Microbial production of recalcitrant dissolved organic matter: long-term carbon storage in the global ocean. *Nat. Rev. Microbiol.* **8**: 593–599. <https://doi.org/10.1038/nrmicro2386>
- Jilbert, T., Asmala, E., Schröder, C., Tiihonen, R., Myllykangas, J.-P., Virtasalo, J. J., Kotilainen, A., Peltola, P., Ekholm, P., Hietanen, S. 2018. Impacts of flocculation on the distribution and diagenesis of iron in boreal estuarine sediments. *Biogeosciences* **15**: 1243–1271. <http://dx.doi.org/10.5194/bg-15-1243-2018>
- Jokinen, S. A., T. Jilbert, R. Tiihonen-Filppula, and K. Koho. 2020. Terrestrial organic matter input drives sedimentary trace metal sequestration in a human-impacted boreal estuary. *Sci. Total Environ.* **717**: 137047. <https://doi.org/10.1016/j.scitotenv.2020.137047>

- Kaufman, L., and P. J. Rousseeuw. 2009. Finding groups in data: An introduction to cluster analysis, v. **344**. John Wiley & Sons.
- Kellerman, A. M., T. Dittmar, D. N. Kothawala, and L. J. Tranvik. 2014. Chemodiversity of dissolved organic matter in lakes driven by climate and hydrology. *Nat. Commun.* **5**: 3804. <https://doi.org/10.1038/ncomms4804>
- Kirkland, J., and P. Antle. 1977. High-performance size-exclusion chromatography of small molecules with columns of porous silica microspheres. *J. Chromatogr. Sci.* **15**: 137–147. <https://doi.org/10.1093/chromsci/15.5.137>
- Kostanski, L. K., D. M. Keller, and A. E. Hamielec. 2004. Size-exclusion chromatography—A review of calibration methodologies. *J. Biochem. Biophys. Methods* **58**: 159–186. <https://doi.org/10.1016/j.jbbm.2003.10.001>
- Kowalczyk, P., M. Zablocka, S. Sagan, and K. Kulinski. 2010. Fluorescence measured in situ as a proxy of CDOM absorption and DOC concentration in the Baltic Sea. *Oceanologia* **52**: 431–471.
- Lepane, V., T. Persson, and M. Wedborg. 2003. Effects of UV-B radiation on molecular weight distribution and fluorescence from humic substances in riverine and low salinity water. *Estuar. Coast. Shelf Sci.* **56**: 161–173. [https://doi.org/10.1016/S0272-7714\(02\)00154-3](https://doi.org/10.1016/S0272-7714(02)00154-3)
- Loiselle, S. A., L. Bracchini, A. M. Dattilo, M. Ricci, A. Tognazzi, A. Cózar, and C. Rossi. 2009. The optical characterization of chromophoric dissolved organic matter using wavelength distribution of absorption spectral slopes. *Limnol. Oceanogr.* **54**: 590–597. <https://doi.org/10.4319/lo.2009.54.2.0590>
- Maeder, M. 1987. Evolving factor analysis for the resolution of overlapping chromatographic peaks. *Analytical Chemistry* **59**: 527–530. <http://dx.doi.org/10.1021/ac00130a035>
- Malik, A. A., V. Roth, M. Hébert, L. Tremblay, T. Dittmar, and G. Gleixner. 2016. Linking molecular size, composition and carbon turnover of extractable soil microbial compounds. *Soil Biol. Biochem.* **100**: 66–73. <https://doi.org/10.1016/j.soilbio.2016.05.019>
- Massicotte, P., and S. Markager. 2016. Using a Gaussian decomposition approach to model absorption spectra of chromophoric dissolved organic matter. *Mar. Chem.* **180**: 24–32. <https://doi.org/10.1016/j.marchem.2016.01.008>
- Massicotte, P., E. Asmala, C. Stedmon, and S. Markager. 2017. Global distribution of dissolved organic matter along the aquatic continuum: Across rivers, lakes and oceans. *Sci. Total Environ.* **609**: 180–191. <https://doi.org/10.1016/j.scitotenv.2017.07.076>
- Mayorga, E. and A. Aufdenkampe. 2002. Processing of bioactive elements in the Amazon River system, *The Ecohydrology of South American Rivers and Wetlands*, 6, 1–24.
- Minor, E., J. Simjouw, J. Boon, A. Kerckhoff, and J. Van der Horst. 2002. Estuarine/marine UDOM as characterized by size-exclusion chromatography and organic mass spectrometry. *Mar. Chem.* **78**: 75–102. [https://doi.org/10.1016/S0304-4203\(02\)00011-7](https://doi.org/10.1016/S0304-4203(02)00011-7)
- Moran, M. A., W. M. Sheldon, and J. E. Sheldon. 1999. Biodegradation of riverine dissolved organic carbon in five estuaries of the southeastern United States. *Estuaries* **22**: 55–64. <https://doi.org/10.2307/1352927>
- Moran, M. A., W. M. Sheldon, and R. G. Zepp. 2000. Carbon loss and optical property changes during long-term photochemical and biological degradation of estuarine dissolved organic matter. *Limnol. Oceanogr.* **45**: 1254–1264. <https://doi.org/10.4319/lo.2000.45.6.1254>
- Murphy, K. R., C. A. Stedmon, T. D. Waite, and G. M. Ruiz. 2008. Distinguishing between terrestrial and autochthonous organic matter sources in marine environments using fluorescence spectroscopy. *Mar. Chem.* **108**: 40–58. <https://doi.org/10.1016/j.marchem.2007.10.003>
- Ogawa, H. 2000. Bulk chemical aspects of dissolved organic matter in seawater review: The recent findings and unsolved problems, p. 311–337. *In Dynamics and Characterization of Marine Organic Matter*. Springer.
- O'Haver, T. 2020. peakfit.m (<https://www.mathworks.com/matlabcentral/fileexchange/23611-peakfit-m>), MATLAB Central File Exchange. Retrieved November 13, 2020.
- O'Loughlin, E., and Y. Chin. 2001. Effect of detector wavelength on the determination of the molecular weight of humic substances by high-pressure size exclusion chromatography. *Water Res.* **35**: 333–338. [https://doi.org/10.1016/S0043-1354\(00\)00228-1](https://doi.org/10.1016/S0043-1354(00)00228-1)
- Omanović, D., C. Santinelli, S. Marcinek, and M. Gonnelli. 2019. ASFit-An all-inclusive tool for analysis of UV-Vis spectra of colored dissolved organic matter (CDOM). *Comput. Geosci.* **133**: 104334. <https://doi.org/10.1016/j.cageo.2019.104334>
- Osburn, C. L., T. J. Boyd, M. T. Montgomery, T. S. Bianchi, R. B. Coffin, and H. W. Paerl. 2016. Optical proxies for terrestrial dissolved organic matter in estuaries and coastal waters. *Front. Mar. Sci.* **2**: 127. <https://doi.org/10.3389/fmars.2015.00127>
- Peuravuori, J., and K. Pihlaja. 1997. Molecular size distribution and spectroscopic properties of aquatic humic substances. *Anal. Chim. Acta* **337**: 133–149. [https://doi.org/10.1016/S0003-2670\(96\)00412-6](https://doi.org/10.1016/S0003-2670(96)00412-6)
- R Core Team. 2019. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>.
- Reader, H. E., C. A. Stedmon, N. J. Nielsen, and E. S. Kritzberg. 2015. Mass and UV-visible spectral fingerprints of dissolved organic matter: Sources and reactivity. *Front. Mar. Sci.* **2**: 88. <https://doi.org/10.3389/fmars.2015.00088>
- Reader, H. E., F. Thoms, M. Voss, and C. A. Stedmon. 2019. The influence of sediment-derived dissolved organic matter in the Vistula River estuary/gulf of Gdansk. *J. Geophys. Res. Biogeophys.* **124**: 115–126. <https://doi.org/10.1029/2018JG004658>

- Schwarz, G. 1978. Estimating the Dimension of a Model. *The Annals of Statistics* **6**: 461–464. <http://dx.doi.org/10.1214/aos/1176344136>
- Sharpless, C. M., and Blough, N. V. 2014. The importance of charge-transfer interactions in determining chromophoric dissolved organic matter (CDOM) optical and photochemical properties. *Environ. Sci.: Processes Impacts* **16**: 654–671. <http://dx.doi.org/10.1039/c3em00573a>
- Shimotori, K., T. Satou, A. Imai, N. Kawasaki, K. Komatsu, A. Kohzu, N. Tomioka, R. Shinohara, and S. Miura. 2016. Quantification and characterization of coastal dissolved organic matter by high-performance size exclusion chromatography with ultraviolet absorption, fluorescence, and total organic carbon analyses. *Limnol Oceanogr: Methods* **14**: 637–648. <https://doi.org/10.1002/lom3.10118>
- Sholkovitz, E., E. Boyle, and N. Price. 1978. The removal of dissolved humic acids and iron during estuarine mixing. *Earth Planet. Sci. Lett.* **40**: 130–136. [https://doi.org/10.1016/0012-821X\(78\)90082-1](https://doi.org/10.1016/0012-821X(78)90082-1)
- Stedmon, C. A., and S. Markager. 2003. Behaviour of the optical properties of coloured dissolved organic matter under conservative mixing. *Estuar. Coast. Shelf Sci.* **57**: 973–979. [https://doi.org/10.1016/S0272-7714\(03\)00003-9](https://doi.org/10.1016/S0272-7714(03)00003-9)
- Stedmon, C. A., S. Markager, and R. Bro. 2003. Tracing dissolved organic matter in aquatic environments using a new approach to fluorescence spectroscopy. *Mar. Chem.* **82**: 239–254. [https://doi.org/10.1016/S0304-4203\(03\)00072-0](https://doi.org/10.1016/S0304-4203(03)00072-0)
- Stubbins, A., R. G. Spencer, P. J. Mann, R. M. Holmes, J. W. McClelland, J. Niggemann, and T. Dittmar. 2015. Utilizing colored dissolved organic matter to derive dissolved black carbon export by arctic rivers. *Front. Earth Sci.* **3**: 63. <https://doi.org/10.3389/feart.2015.00063>
- Swift, R. S. 1999. Macromolecular properties of soil humic substances: Fact, fiction, and opinion. *Soil Sci.* **164**: 790–802.
- Vartiainen, T., Liimatainen, A., & Kauranen, P. 1987. The use of TSK size exclusion columns in determination of the quality and quantity of humus in raw waters and drinking waters. *Sci. Total Environ.* **62**: 75–84. [https://doi.org/10.1016/0048-9697\(87\)90484-0](https://doi.org/10.1016/0048-9697(87)90484-0)
- Wu, F., D. Evans, P. Dillon, and S. Schiff. 2004. Molecular size distribution characteristics of the metal–DOM complexes in stream waters by high-performance size-exclusion chromatography (HPSEC) and high-resolution inductively coupled plasma mass spectrometry (ICP-MS). *J. Anal. At. Spectrom* **19**: 979–983. <https://doi.org/10.1039/B402819H>
- Wünsch, U. J., C. A. Stedmon, L. J. Tranvik, and F. Guillemette. 2018. Unraveling the size-dependent optical properties of dissolved organic matter. *Limnol. Oceanogr.* **63**: 588–601. <https://doi.org/10.1002/lno.10651>
- Yan, M., G. Korshin, D. Wang, and Z. Cai. 2012. Characterization of dissolved organic matter using high-performance liquid chromatography (HPLC)–size exclusion chromatography (SEC) with a multiple wavelength absorbance detector. *Chemosphere* **87**: 879–885. <https://doi.org/10.1016/j.chemosphere.2012.01.029>
- Zhou, Q., S. E. Cabaniss, and P. A. Maurice. 2000. Considerations in the use of high-pressure size exclusion chromatography (HPSEC) for determining molecular weights of aquatic humic substances. *Water Res.* **34**: 3505–3514. [https://doi.org/10.1016/S0043-1354\(00\)00115-9](https://doi.org/10.1016/S0043-1354(00)00115-9)
- Zhou, Z., B. Stolpe, L. Guo, and A. M. Shiller. 2016. Colloidal size spectra, composition and estuarine mixing behavior of DOM in river and estuarine waters of the northern Gulf of Mexico. *Geochim. Cosmochim. Acta* **181**: 1–17. <https://doi.org/10.1016/j.gca.2016.02.032>

#### Acknowledgments

We are grateful for Lumi Haraguchi from Finnish Environment Institute, and Berit Langkilde Møller and Janus Larsen from Aarhus University for their assistance in the lab and in the field. We would like to thank Colin Stedmon from DTU Aqua for the DOC analysis, and the captain and crew of the R/V Alkor. EA and JC were supported by the BONUS COCOA project (grant agreement 2112932-1), funded jointly by the EU and Danish Research Council. EA was also supported by the Academy of Finland (Grant No. 309748).

#### Conflict of interest

None declared.

Submitted 16 March 2020

Revised 05 June 2020

Accepted 26 November 2020

Associate editor: Yunlin Zhang