The one-sample PARAFAC approach reveals molecular size distributions of fluorescent components in dissolved organic matter

Wünsch, Urban; Murphy, Kathleen R.; Stedmon, Colin

Published in:
Environmental Science & Technology (Washington)

Link to article, DOI:
10.1021/acs.est.7b03260

Publication date:
2017

Document Version
Peer reviewed version

Article

The one-sample PARAFAC approach reveals molecular size distributions of fluorescent components in dissolved organic matter

Urban Wünsch, Kathleen Murphy, and Colin Andrew Stedmon

Environ. Sci. Technol., **Just Accepted Manuscript** • DOI: 10.1021/acs.est.7b03260 • Publication Date (Web): 26 Sep 2017

The one-sample PARAFAC approach reveals molecular size distributions of fluorescent components in dissolved organic matter

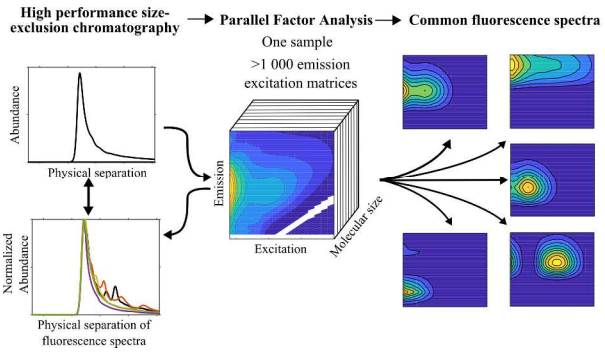
Urban J. Wünsch[1*], Kathleen R. Murphy[2], Colin A. Stedmon[1]

[1]National Institute of Aquatic Resources, Section for Oceans and Arctic, Technical University of Denmark, Kemitorvet, Building 201, 2800 Kgs. Lyngby, Denmark

[2]Chalmers University of Technology, Water Environment Technology, Sven Hultins Gata 6, 41296 Gothenburg, Sweden

**Corresponding Author*: Urban J. Wünsch (urbw@aqua.dtu.dk)

Abstract



Molecular size plays an important role in dissolved organic matter (DOM) biogeochemistry, but its relationship with the fluorescent fraction of DOM (FDOM) remains poorly resolved. Here high-performance size exclusion chromatography (HPSEC) was coupled to fluorescence emission-excitation (EEM) spectroscopy in full spectral (60 emission and 34 excitation wavelengths) and chromatographic resolution (< 1Hz), to enable the mathematical decomposition of fluorescence on an individual sample basis by parallel factor analysis (PARAFAC). The approach allowed cross-system comparisons of molecular size distributions for individual fluorescence components obtained from independent datasets. Spectra extracted from allochthonous DOM were highly similar. Allochthonous and

autochthonous DOM shared some spectra, but included unique components. In agreement with the supramolecular assembly hypothesis, molecular size distributions of the fluorescence fractions were broad and chromatographically unresolved, possibly representing reoccurring fluorophores forming non-covalently bound assemblies of varying molecular size. Samples shared underlying fluorescence components that differed in their size distributions but not their spectral properties. Thus, in contrast to absorption measurements, bulk fluorescence is unlikely to reliably indicate the average molecular size of DOM. The one-sample approach enables robust and independent cross-site comparisons without large-scale sampling efforts and introduces new analytical opportunities for elucidating the origins and biogeochemical properties of FDOM.

Introduction

Dissolved organic matter (DOM) represents a large pool of organic carbon in aquatic ecosystems of a magnitude comparable to atmospheric carbon dioxide.[1] DOM has a significant role in the continental-scale carbon balance, as well as influence at local scales.[2,3] Previous studies have shown direct links between the optical, physical and chemical properties of DOM, such as the molecular size,[4] lignin content,[5] and aromaticity.[6] The molecular size distribution of DOM as a whole, and the size of individual compounds within it, are a key trait that can be linked to its degradation susceptibility.[7–10] In particular, numerous studies suggest a positive relationship between the average molecular size of DOM and fluorescence emission maxima,[11–16] suggesting that "humic-like" fluorescence is the result of extended, conjugated aromatic structures.

Optical properties of different DOM size fractions have provided evidence for the supramolecular assembly hypothesis,[17] whereby individual DOM moieties recur in non-covalently bound assemblies of varying molecular size. In support of this hypothesis, highly similar optical properties are seen across the entire molecular-size gradient of DOM.[11,18–20] Apparent molecular size distributions of DOM are typically analyzed by high-performance size exclusion chromatography (HPSEC) to study changes in the size distribution of DOM as a function of biogeochemical and physical factors.[20–24] Molecular size distributions depend on the instrument used for its measurement: Mass spectrometry most often shows an average molecular weight around 400 Da,[25] while HPSEC or ultrafiltration show higher averages.[4,26,27] However, many studies utilizing HPSEC are based on measurements of discrete sample fractionations, and hence have relatively low sample (i.e. chromatographic) resolution. Recent studies used online detectors to provide high resolution data (i.e. < 1 Hz);[14,20,28] however in fluorescence studies, a limited number of excitation wavelengths were monitored simultaneously by these detectors, and a systematic, comprehensive data analysis approach has yet to be established. Moreover, the determination of

54 molecular size distributions of emission excitation matrix (EEM) fluorescence in a continuous fashion
55 (<1 Hz) at full spectral resolution remains unachieved.

56 Absorbance and fluorescence spectroscopy allow the rapid and sensitive determination of chromophoric
57 and fluorescent DOM (CDOM and FDOM, respectively).[4,29,30] Due to the higher sensitivity and
58 selectivity of fluorescence over absorbance spectroscopy, FDOM properties measured as EEMs are
59 widely used as a proxy for DOM quality in aquatic environments.[31–34] EEM fluorescence spectroscopy
60 produces three-dimensional datasets that can be decomposed mathematically with methods such as
61 Parallel Factor Analysis (PARAFAC)[35,36], to obtain chemically and mathematically independent
62 fluorescence spectra. Targeted analysis of specific DOM compounds (such as amino acids) and
63 untargeted analysis of DOM using e.g. mass spectrometry, have suggested the presence of a common, or
64 even ubiquitous, fraction of chemical compounds within DOM.[37,38] These findings may also extend to the
65 optical properties of DOM and indicate the possible presence of common fluorophores within the global
66 FDOM pool.[39,40] Since 2014, the OpenFluor database has enabled systematic comparisons between
67 PARAFAC-derived DOM fluorescence spectra;[41] however, comparisons between studies and systems are
68 often hampered by instrumental, methodological and inter-laboratory variation.[42] In order to achieve
69 systematic and robust comparisons and locate common fluorescence spectra in the global FDOM pool, it
70 is crucial to establish analytical frameworks that mitigate such disturbances, whilst also avoiding
71 excessive sampling and measurement efforts.

72 PARAFAC is frequently used to interpret bulk DOM fluorescence datasets, though a number of
73 practical and theoretical hurdles still limit its application. For example, the fluorescence dataset must
74 contain sufficient spectral variation to produce meaningful, stable, and verifiable models, hence large
75 sample sizes are preferred.[43] In studies involving a relatively low number of samples, this requirement
76 often inhibits the use of PARAFAC or prevents model validation. Additionally, the mathematical

77 decomposition of EEMs assumes the superposition of a finite number of independently fluorescing
78 compounds (following Beer's Law), i.e. that the fluorescence spectrum of a mixture arises from the
79 spectra of its individual constituents.[44,45] However, verifying this assumption is difficult for datasets
80 containing environmental samples. While multiple studies have questioned the superposition assumption
81 due to electronic interactions between chromophores,[46–49] evidence of electronic interactions have mainly
82 been observed under conditions of strong chemical oxidation,[50] although one study reported self-
83 quenching of humic acid in HPSEC separations.[51] The extent to which electronic interactions undermine
84 the use of PARAFAC under environmental conditions is still unknown.

85 The goal of this study was to establish a new analytical framework based on HPSEC that can
86 reveal the molecular size distributions of the underlying DOM fluorescence components in individual
87 environmental samples using the full spectral resolution of EEMs. We further aimed to identify whether
88 or not FDOM separated by HPSEC follow Beer's law (i.e. behaves additively). Once this was confirmed,
89 the goal was to mathematically decompose EEMs from individual samples and compare the underlying
90 fluorescence spectra. Moreover, we aimed to assess the supramolecular assembly hypothesis using our
91 analytical framework. Finally, we aimed to identify consistent trends between fluorescence emission
92 maxima and molecular size of statistically derived components in individual samples. If found, these
93 would offer the opportunity to use the bulk FDOM composition to gain direct insights into the average
94 molecular size of FDOM.

Materials and Methods

Sample Collection

Four allochthonous samples (Lake Lillsjön, Sweden; Rio Negro, Brazil; Svartan River, Sweden; Rio Tapajós, Brazil) and two autochthonous samples (Pacific Ocean & Pony Lake) were extracted with PPL and XAD8 resins, respectively, using established methods (see Supporting Information (SI) S1 and Table S1 for further information).[52–55] 200μL of the PPL extract was dried and SPE-DOM was reconstituted in 0.15 M ammonium acetate (pH 7). DOM of two XAD-8 extracts was dissolved in 0.15 M ammonium acetate at concentrations of 1.4 mg L^{-1} and 0.25mg L^{-1} for samples originating from Pony Lake and the Pacific Ocean, respectively.

High performance size-exclusion chromatography

Full details pertaining to the HPSEC equipment, and methodology are provided in the Supporting Information (S1, SI Figures S1-S6). Briefly, HPSEC was performed using a Shimadzu Nexera X2 UFLC system equipped with a TSKgel SuperAWM-H column. DOM was eluted with 0.15 M ammonium acetate (pH 7), and two sequential detectors were used. Absorbance was measured between 240 and 600 nm at 2 nm intervals using a Shimadzu SPD-M30. Fluorescence emission was then detected between 300 and 600 nm at 5 nm increments across excitation wavelengths from 240 – 450 nm at 5-10 nm increments using a Shimadzu RF-30Axs by combining measurements from separate runs. For every sample, a sequence of runs was performed whereby the same sample was injected while instrument parameters were systematically changed between runs. In total, the chromatographic run was repeated 35 times to allow the determination of absorbance properties (one run) and fluorescence properties at an EEM-like spectral resolution (34 runs, one for every excitation wavelength at a constant emission wavelength range). This resulted in around 1500 individual absorbance spectra and fluorescence emission scans (each of the 1500

emission scan subsequently combined with those of the other injections to form EEMs as shown in SI Figure S6) per sample (total of ~ 26 hrs measurement time per sample). To reduce the dataset size, a subset of 250 evenly-spaced emission scans were extracted for the chemometric analysis (see below). After compilation of fluorescence emission scans into EEMs, each EEM presents the fluorescence composition of a measured sample at a given elution volume (or apparent molecular size). The analytical column was calibrated using polystyrene sulfonate, which serves only to report the approximate apparent molecular size of peak maxima assuming identical chromatographic separation of chemically diverse DOM and uniform polystyrene standards. In this regard, whole chromatograms are plotted by elution volumes rather than size. Data processing steps included dataset alignment (elimination of inter-detector tubing volume), as well as artefact removal (self-shading and physical scatter) as detailed in the Supporting Information. Fluorescence data were normalized to the Raman peak area at excitation wavelength 350 nm.

Chemometric analysis

HPSEC of natural DOM seldom results in the clear chromatographic separation of different DOM fractions, since the mixture represents an overlapping continuum of compounds.[56–58] The separation of co-eluting analytes can be achieved with mathematical deconvolution approaches such as PARAFAC.[59] In this study, PARAFAC was applied using the drEEM toolbox (v0.3.0) to mathematically decompose the three-way data array as described previously:[35,45,60]

$$x_{ijk} = \sum_{f=1}^{F} a_{if} b_{jf} c_{kf} + e_{ijk} \quad \text{(Eq. 1)}$$

i = 1…I; j = 1…J; k = 1…K

PARAFAC models the fluorescence emission of the *i*th sample (representing discrete elution volumes) at excitation *k*, and emission *j*. The term a_{if} is proportional to the abundance of the *f*th chromatographic

analyte in sample *i*. The term b_{jf} represents the least-square estimate of the emission spectrum of the *f*th analyte, while c_{kf} is the least-square estimate of the excitation spectrum of the *f*th analyte at wavelength *k*. The term e_{ijk} represents the residual matrix that contains unexplained dataset variability. Importantly, the successful decomposition of chemical datasets into underlying factors using the PARAFAC model hinges on three assumptions: (1) Variability: No two compounds can have the same exact spectral properties and identical fluorescence intensities; (2) Additivity: The total fluorescence intensities observed are the result of the fluorescence of a finite number of analytes that do no interact electronically; and (3) Trilinearity: The signal of a given analyte is linearly related to its invariant excitation and emission spectrum, i.e. one component describes an analyte in all three modes. The combination of applying PARAFAC to decompose HPSEC-derived EEMs is represented herein using the terminology HPSEC-EEM-PARAFAC.

In agreement with the analytical nature of the HPLC dataset, PARAFAC models in this work were constrained to non-negativity, i.e. component scores and loadings were forced to be positive. Model fits were stopped when the relative reduction in fitting error from one iteration to the next did not exceed 10^{-7}. Since HPSEC chromatograms typically feature analyte abundances that vary over several orders of magnitude, pretreatment of data is critical to avoid extremely different leverages across the dataset gradient.[45] However, normalizing HPSEC EEMs to unit variance is problematic since early- and late-eluting EEMs with fluorescence close to zero are amplified, preventing efficient PARAFAC modelling. Instead, fluorescence intensities were $\log_{10}$ normalized, which reduced the effect of peak-to-baseline concentration gradients, limited covariance between simultaneously eluting analytes, and limited the effect of noise (SI Figure S7).

The Tucker congruence coefficient (TCC) was used to assess spectral congruence between components derived from different samples and models.[61] A classic (dataset-, i.e. sample-specific) split-half validation was performed for the two autochthonous samples ($TCC_{combined} > 0.95$), while a more stringent external

(i.e. cross-dataset and -sample) comparison was performed for the four allochthonous FDOM models since these samples appeared to be highly similar. Since these allochthonous samples were collected by different scientists and at different locations and times, this approach represents a more stringent approach to assessing model validity. For the external comparison, a slightly lower TCC threshold of > 0.95 for emission and excitation spectra (i.e. $TCC_{combined} > 0.9$) set. This threshold represents a compromise between the more rigid threshold employed by OpenFluor ($TCC_{combined} > 0.95$) and a lower threshold that takes into account variability that can arise from modeling and comparing two completely independent datasets.

Results and Discussion

HPSEC optical properties: Beer's law or charge transfer?

The unique coupling of HPSEC and full-resolution EEM spectroscopy in this study presented the opportunity to investigate the additive behavior of DOM fluorescence (i.e. compliance with Beer's law). The sum of fluorescence emission from size-separated EEMs was compared to the bulk EEM obtained on the same instrument without the chromatographic column installed. This was performed for two representative samples (Fig. 1, SI Figure S8). Size separation did not produce substantial changes of fluorescence in the visible fluorescence emission region (excitation > 300 nm), as might be expected from intermolecular charge transfer or fluorescence quenching.[46,62–64] However, at excitation wavelengths below 300 nm, two regions deviated from the otherwise randomly-distributed residuals. In the UVA region, a negative residual of less than 20 % was observed relative to the bulk EEM, indicating loss of fluorescence during separation. This was likely caused by adsorption of small monomers onto the analytical column due to secondary interactions, since pure tryptophan and salicylic acid also showed secondary retention (data not shown). Secondly, a positive residual of <4 % indicating a gain of fluorescence was seen in the emission range between 360 and 470 nm when excitation was below 300 nm. This small gain is likely attributable to a weak background fluorescence signal emitted by the mobile phase, which constantly eluted from the analytical column despite an auto-zero blank subtraction at the beginning of each run. Despite these minor differences, the spectral shape of bulk and size-separated EEMs was highly similar. TCCs between fluorescence emission at all excitations was higher than 0.9997, but did show lower values at low excitation wavelengths (<300 nm) due to the lack of protein-like fluorescence in the size-separated EEMs (SI Figure S8).

Overall, the additivity of fluorescence within the framework of HPSEC separation was confirmed. The application of superposition-based decomposition models such as PARAFAC was

193 therefore deemed to be appropriate. Since HPSEC possibly disrupts intermolecular charge transfer and
194 fluorescence quenching based on partial physical separation, the absence of substantial differences
195 between separated and bulk EEMs indicates that such interactions between fluorescence components were
196 not occurring to any significant extent. Conversely, this result does not provide information on
197 intramolecular charge-transfer or quenching interactions, since chromatographic separation would not be
198 expected to disrupt their occurrence. Any effects of these intramolecular interactions may thus remain
199 embedded in the spectral signatures of components identified in Fig. 2. However, since charge-transfer
200 interactions are embedded in the extracted spectral signatures, HPSEC-EEM-PARAFAC might help to
201 systematically investigate and identify such interactions by comparing the optical properties of chemically
202 contrasting samples in future studies.

Spectral conformity among allochthonous samples

The four allochthonous DOM extracts originated from freshwater environments in Sweden and Brazil receiving a large proportion of terrestrial organic matter. They were analyzed independently using HPSEC-EEM-PARAFAC to decompose EEMs into independent fluorescence components. On average, PARAFAC models with two to five components explained respectively 99.51, 99.78, 99.88, and 99.92 % of variability in each dataset, and all models with the same number of components (two to five components) had highly congruent underlying fluorescence spectra (SI Figure S9, SI Table S2). For all four samples, the five-component PARAFAC model (Fig. 2) best represented their fluorescence properties (SI Figure S10), since four component models did not adequately represent protein-like fluorescence, while core consistencies, sum-of-squared-errors, and spectral loadings of six component models frequently implied over-factorization The spectral congruence between the independent datasets is interpreted as compelling evidence for the validity of the four individual models (i.e. similar to a conventional split-half analysis). The five components of the validated PARAFAC model were named according to their fluorescence emission maximum as follows: C_{350}, C_{405}, C_{430}, C_{450}, C_{510} (Fig. 2).

Despite high overall similarity between components in all four models (Fig. 2), their spectral congruence did not always meet the criterion of $TCC \geq 0.95$ that is often applied to identify interchangeable spectra (SI Table S2).[41,45] However, in all comparisons, TCC exceeded the threshold of "fair" similarity (0.85) and in all but one cases exceeded 0.90 (C_{350} between Lake Lillsjön and Svartan River).[61] Common sources that might result in minor spectral changes, such as sample pH and metal-quenching, were eliminated by the combination of DOM extraction and controlled buffer conditions during chromatographic separations.[65,66] Thus, the spectral differences observed between otherwise similar models must have originated from other sources. Since fluorescence properties of aromatic structures are influenced by conjugation and substitution, structural variations in similar fluorophores

between samples may explain the slight spectral shifts and shape variations.[67] While this explanation cannot be denied without further experiments, we hypothesize that the spectral differences may have also been caused by the unavoidable spectral limitations of the highly sensitive HPSEC fluorescence detector. Compared to traditional EEMs,[68] HPSEC EEMs are affected by larger areas of scatter and a slightly reduced spectral range, both of which may influence the mathematical decomposition.[69] While the general influence of these factors has been investigated previously,[70] the detailed influence of variable size of scatter excision and changing spectral ranges on fluorescence modeling remains poorly understood and should be investigated further. While the identification of the primary factor responsible for the observed spectral differences between congruent fluorescence components is currently not identifiable, the one-sample framework is best suited to investigate such issues since it would otherwise not be possible to make such observations using such a limited number of environmental samples. Despite this, a high degree of overall similarity between the fluorescence compositions of independent samples from geographically contrasting sites was observed (Fig. 3a, SI Figure S11), since relative component contributions were within 5 % of the respective mean contributions (Fig. 3b). These deviations are especially low compared to a recent HPSEC-based study of boreal lake DOM that indicated variations of more than 50 % for some humic-like components between samples from different lakes.[71] This compositional and spectral similarity is striking and suggests that globally, the bulk optical properties of terrestrial DOM may arise from very similar chemical structures.

Comparison between autochthonous, allochthonous and community-derived fluorescence spectra

Since the spectral properties of four allochthonous samples were strikingly similar, it was hypothesized that spectral properties of autochthonous FDOM would also be similar across samples. To test this, the size-dependent optical properties of the autochthonous extracts from the Pacific Ocean and Pony Lake samples were analyzed using the same approach employed for the allochthonous samples. This offered the opportunity to compare fluorescence components originating from lateral terrestrial inputs in rivers and lakes with fluorescence components produced *in situ*. For both autochthonous samples, a six-component PARAFAC model best described the size-dependent optical properties (SI Figure S12). The spectral properties of the autochthonous extracts visibly differed from the allochthonous extracts and unlike the allochthonous extracts, contained mostly unique fluorescence spectra. Only two components (emission maxima at 510 and 430 nm) derived from the Pacific Ocean and Pony Lake sample were spectrally congruent. Unique spectra in both autochthonous samples consisted of three protein-like fluorophores with emission maxima below 400 nm and five humic-like components with emission maxima between 400 and 500 nm. Thus, the hypothesis that autochthonous FDOM components are spectrally similar across samples was rejected. However, it should be noted that the two autochthonous samples were extracted using different resins, potentially affecting this result.

Despite greater variability, the fluorescence spectra derived from the autochthonous samples partially matched with spectra derived from the allochthonous samples. Components closely matching allochthonous C_{510} (as identified in Lake Lillsjön) were found in Pony Lake and Pacific Ocean FDOM. Thus, C_{510} was the only ubiquitous component across all investigated samples (Fig. 4). Additionally, components closely resembling C_{405} and C_{430} were present in FDOM from Pony Lake and the Pacific Ocean, respectively (Fig. 4).

All five freshwater-derived fluorescence components identified in this study correlated with fluorescence spectra in the OpenFluor database.[41] Components C_{350}, C_{405}, C_{430}, C_{450}, and C_{510} yielded matches with components from a total of 10, 38, 31, 2, and 24 studies, respectively (Fig. 4, grey lines). Considering the current total of 62 models with 4 or more components in the database (as of June 2017), the five allochthonous fluorescence spectra thus showed spectral correlation with a significant proportion of previous studies (except in the case of C_{450}). C_{510} and C_{405} also showed striking similarity with two components previously listed by Ishii & Boyer (2012) as reoccurring humic-like FDOM components.[39] The fact that C_{510} and C_{405} also represent the components with the highest number of matches in the OpenFluor database confirms these earlier observations and the presence of reoccurring PARAFAC components across aquatic environments.

Compared to the bulk-sample PARAFAC approach,[45] the one-sample modeling approach described here offers critical advantages. First, our approach does not require large-scale sampling efforts. Secondly, HPSEC offers the unique opportunity to confirm the additive behavior of DOM fluorescence and thus ensures the applicability of mathematical decomposition routines. Moreover, EEMs originating from HPSEC separations are not influenced by disturbances common to environmental gradients, such as pH,[65] metal-quenching,[66] ionic strength,[72] and charge-transfer.[73] Thus, we propose that the described one-sample modeling framework offers a systematic approach to investigate the commonality of fluorescence spectra across different aquatic environments. However, it should be noted that the shown sample characteristics strictly apply to the time of sampling. DOM composition may change with season and sampling location. Nevertheless, similarities between samples were found despite factors such as time of sampling and seasonality of the individual systems, spatial differences in DOM biogeochemistry, and methodological differences in sampling.

Physical separation and mathematical decomposition: Molecular size distributions of fluorescence spectra

In the context of HPSEC-EEM-PARAFAC, component scores represent molecular size as the primary chromatographic separation mechanism. As stated above, spectral loadings of some components originating from individual PARAFAC models were strongly congruent and therefore warranted further comparison to examine apparent molecular size distributions between samples originating from different aquatic environments. The supramolecular assembly hypothesis states that individual DOM moieties (e.g. fluorescing compounds) form non-covalently bound assemblies (including non-fluorescing compounds) of varying molecular size.[17,74] Evidence supporting this hypothesis is based on the highly similar character of DOM obtained from HPSEC-based fractions as observed by mass spectrometry,[75] infrared spectroscopy,[17] and fluorescence spectroscopy.[11,18–20] In this light, the apparent molecular size distributions are expected to be broad and unresolved. In agreement with this hypothesis, components originating from allochthonous DOM showed highly similar molecular size distributions with poor physical separation (Fig. 5, SI Fig. S13). Components generally exhibited a single peak with tailing towards higher elution volumes (low apparent molecular size). The molecular size distributions of PARAFAC components other than C_{405} were very similar across samples (TCC>0.98, SI Fig. S13). The observation of broad, overlapping distributions instead of distinct, resolved peaks thus aligns with earlier findings, although the combined chromatographic and mathematic approach employed here provides unprecedented detail due to the utilization of online detectors (< 1 Hz) instead of discrete fractionation.

A direct link between fluorescence emission maximum and molecular size would provide evidence that the chemical structure of larger fluorophores results in “humic-like” fluorescence through extended conjugation of aromatic structures.[76] Contrary to findings in earlier studies that reported direct correlations between molecular size and fluorescence emission maximum,[11–16] peak molecular sizes of

components were not correlated to fluorescence emission. Across the allochthonous samples, the average peak molecular size was 1.54 ± 0.15, 1.45 ± 0.05, 1.42 ± 0.09, 1.30 ± 0.12, 0.89 ± 0.24 kDa for components C_{510}, C_{350}, C_{430}, C_{450}, and C_{405}, respectively ($R^2 = -0.22$, $p > 0.1$). Moreover, no relationship between the FDOM composition (as observed by relative contributions of PARAFAC components to the total HPSEC-EEM fluorescence) and molecular size of total fluorescence was apparent (Fig. 3, dashed line and red dots). Although a direct correlation between fluorescence emission and molecular size might be expected for simple mixtures, our results suggest that this was not the case for the complex mixtures analyzed in this study. Our findings rather suggest that FDOM components were possibly associated with non-fluorescing organic matter with a range of three-dimensional structures / sizes, thus convoluting the relationship between fluorescence emission and molecular size. The contradictory results may arise at least in part from the different analytical approaches between studies, but may also result from differences in sample preparation, choice of analytical column, or the overall degree of compositional variability in each dataset. While results in this study suggest that bulk FDOM is an unreliable indicator of the average molecular size of DOM, further investigation with additional samples is warranted. This finding also highlights the need for additional analytical detectors (such as refractive index or mass spectrometry) to be included in HPSEC analyses, since a combination of detectors with overlapping analytical windows will provide deeper insights into the molecular assemblies of DOM.

HPSEC-EEM-PARAFAC demonstrated that apparent molecular size distributions of spectrally congruent fluorescence spectra may differ between samples. We identified differences in apparent size distributions, most notably in the low molecular size range, in particular for the poorly-resolved peaks of components C_{350}, C_{405} and C_{450} in several samples (Fig. 5 inserts). Notably, experiments with pure fluorophores suggested the presence of secondary column interactions with compounds of low molecular size. Thus, a combination of secondary interaction (possibly of hydrophobic nature) and molecular size might be

334 responsible for peaks at low apparent molecular size. Nonetheless, these observations all point towards
335 distinct compositional differences between samples.

336 Molecular size distributions of corresponding fluorescence spectra extracted from autochthonous samples
337 (Pony Lake and the Pacific Ocean) visibly differed compared to allochthonous samples (Fig. 5). Although
338 molecular size peaks were similar, the size distribution of C_{510} was shifted toward low apparent molecular
339 size. For the Pacific Ocean sample, C_{405} showed two distinct peaks at high elution volume (~3.8, and
340 ~6.1 mL) that did not occur in allochthonous samples. Similarly, C_{430} of the Pony Lake sample showed a
341 peak at elution volume 3.9 mL that was not visible in allochthonous samples.

342 These small, but significant differences present novel insights into the chemical properties of spectrally
343 interchangeable fluorescence components. For example, according to the size-reactivity continuum,[7]
344 chemical compounds at contrasting ends of the marine DOM molecular size distribution are utilized by
345 bacteria at drastically different rates.[8,9] In this light, our findings suggest that interchangeable fluorescence
346 spectra may inadvertently be proxies for chemical assemblies of different molecular size and thus
347 different biogeochemical reactivity. The inherent inability of bulk measurements to provide such
348 information highlights the need to incorporate further analytical dimensions in the characterization of
349 DOM in order to unravel the biogeochemical role of the various DOM fractions. Similar to the systematic
350 investigation of spectral properties of FDOM, we propose the one-sample modeling approach as
351 framework to provide novel insights into the relationship between DOM (as analyzed by various
352 instruments, such as spectrofluorometers or mass spectrometers) and physical and chemical properties of
353 DOM (as determined by e.g. HPSEC or reverse-phase liquid chromatography).

HPSEC-EEM-PARAFAC: Implications and future directions

The combination of physical and mathematical chromatography (HPSEC-EEM-PARAFAC) presents an advantageous framework for the systematic investigation of fluorescence properties of single environmental samples. To date, the application of PARAFAC has been hindered by necessity to attain a large dataset spanning a relevant gradient in composition. The opportunity to now assess cross-system variability of DOM in a standardized, robust fashion represents a significant advance in the characterization of DOM. At the same time, this approach provides numerous additional opportunities: Firstly, the HPSEC-base single-sample approach offers detailed insights into molecular size distributions of fluorophores. This analytical advance will improve the understanding of fluorophores as proxies for DOM biogeochemistry. Secondly, the fact that spectral decomposition / characterization can now be performed on individual samples increases the potential utility of the PARAFAC-EEM approach in experimental manipulations with limited samples or for studies focused on characterizing trends across independent systems (e.g. suite of isolated lakes or biomes). Finally, beyond fluorescence spectroscopy, the single-sample approach opens up opportunities for a systematic comparison of data originating from different analytical techniques (such as fluorescence spectroscopy, nuclear magnetic resonance spectroscopy, and mass spectrometry). The fusion of data from multiple analytical approaches may produce new insights into the composition of DOM that are inaccessible from any technique on its own.

Acknowledgements

This work is dedicated to George Aiken in recognition of his contribution to the field of DOM biogeochemistry and characterization, and his role as a mentor for so many young scientists. This work was in part funded by the Danish Council for Independent Research-Natural Sciences Grant DFF-1323-00336 and Nordic5Tech collaborative funding (Technical University of Denmark). KRM acknowledges support by the Swedish Research Council (Formas 2013–1214) and the DRICKS drinking water research center. The authors would like to thank Michael Gonsior (University of Maryland) for providing organic matter extracts, as well as Diane McKnight (University of Colorado Boulder) and George Aiken (United States Geological Survey) for providing XAD-8 organic matter extracts. UJW would like to thank Adam Hambly for proof-reading the manuscript.

Supporting Information

The Supporting Information contains extended methods (S1), eleven figures (S2), and two tables (S3).

This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Hansell, D. A.; Carlson, C. A.; Repeta, D. J.; Schlitzer, R. Dissolved Organic Matter in the Ocean A Controversy Stimulates New Insights. *Oceanography* **2009**, *22* (4), 202–211.

(2) Cole, J. J.; Prairie, Y. T.; Caraco, N. F.; McDowell, W. H.; Tranvik, L. J.; Striegl, R. G.; Duarte, C. M.; Kortelainen, P.; Downing, J. A.; Middelburg, J. J.; et al. Plumbing the global carbon cycle: Integrating inland waters into the terrestrial carbon budget. *Ecosystems* **2007**, *10* (1), 171–184.

(3) Tranvik, L. J.; Downing, J. A.; Cotner, J. B.; Loiselle, S. A.; Striegl, R. G.; Ballatore, T. J.; Dillon, P.; Finlay, K.; Fortino, K.; Knoll, L. B. Lakes and reservoirs as regulators of carbon cycling and climate. *Limnol. Oceanogr.* **2009**, *54* (6, part 2), 2298–2314.

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

(5) Fichot, C. G.; Benner, R. The spectral slope coefficient of chromophoric dissolved organic matter (S275-295) as a tracer of terrigenous dissolved organic carbon in river-influenced ocean margins. *Limnol. Oceanogr.* **2012**, *57* (5), 1453–1466.

(6) Weishaar, J. L.; Aiken, G. R.; Bergamaschi, B. A.; Fram, M. S.; Fujii, R.; Mopper, K. Evaluation of specific ultraviolet absorbance as an indicator of the chemical composition and reactivity of dissolved organic carbon. *Environ. Sci. Technol.* **2003**, *37* (20), 4702–4708.

(7) Amon, R. M. W.; Benner, R. Rapid cycling of high-molecular-weight dissolved organic matter in the ocean. *Nature* **1994**, *369* (6481), 549–552.

(8) Amon, R. M. W.; Benner, R. Bacterial utilization of different size classes of dissolved organic matter. *Limnol Ocean.* **1996**, *41* (1), 41–51.

(9) Benner, R.; Amon, R. M. W. The Size-Reactivity Continuum of Major Bioelements in the Ocean. *Ann. Rev. Mar. Sci.* **2015**, *7*, 185–205.

(10) Tranvik, L. J. Bacterioplankton growth on fractions of dissolved organic carbon of different molecular weights from humi and clear waters. *Appl. Environ. Microbiol.* **1990**, *56* (6), 1672–1677.

(11) Her, N.; Amy, G.; McKnight, D. M.; Sohn, J.; Yoon, Y. Characterization of DOM as a function of MW by fluorescence EEM and HPLC-SEC using UVA, DOC, and fluorescence detection. *Water Res.* **2003**, *37*, 4295–4303.

(12) Wu, F. C.; Evans, R. D.; Dillon, P. J. Separation and Characterization of NOM by High-Performance Liquid Chromatography and On-Line Three-Dimensional Excitation Emission Matrix Fluorescence Detection. *Environ. Sci. Technol.* **2003**, *37* (16), 3687–3693.

(13) Stubbins, A.; Lapierre, J.-F.; Berggren, M.; Prairie, Y. T.; Dittmar, T.; del Giorgio, P. A. What's in an EEM? Molecular Signatures Associated with Dissolved Organic Fluorescence in Boreal Canada. *Environ. Sci. Technol.* **2014**, *48*, 10598–10606.

(14) Cuss, C. W.; Guéguen, C. Relationships between molecular weight and fluorescence properties for size-fractionated dissolved organic matter from fresh and aged sources. *Water Res.* **2015**, *68*, 487–497.

(15) Cuss, C. W.; Guéguen, C. Determination of relative molecular weights of fluorescent components in dissolved organic matter using asymmetrical flow field-flow fractionation and parallel factor analysis. *Anal. Chim. Acta* **2012**, *733*, 98–102.

(16) Stolpe, B.; Zhou, Z.; Guo, L.; Shiller, A. M. Colloidal size distribution of humic- and protein-like fluorescent organic matter in the northern Gulf of Mexico. *Mar. Chem.* **2014**, *164*, 25–37.

(17) Peuravuori, J.; Pihlaja, K. Preliminary study of lake dissolved organic matter in light of nanoscale supramolecular assembly. *Environ. Sci. Technol.* **2004**, *38* (22), 5958–5967.

(18) Romera-Castillo, C.; Chen, M.; Yamashita, Y.; Jaffé, R. Fluorescence characteristics of size-fractionated dissolved organic matter: Implications for a molecular assembly based structure? *Water Res.* **2014**, *55*, 40–51.

(19) Boehme, J.; Wells, M. Fluorescence variability of marine and terrestrial colloids: Examining size fractions of chromophoric dissolved organic matter in the Damariscotta River estuary. *Mar. Chem.* **2006**, *101* (1–2), 95–103.

(20) Li, W. T.; Xu, Z. X.; Li, A. M.; Wu, W.; Zhou, Q.; Wang, J. N. HPLC/HPSEC-FLD with multi-excitation/emission scan for EEM interpretation and dissolved organic matter analysis. *Water Res.* **2013**, *47* (3), 1246–1256.

(21) Allpike, B. P.; Heitz, A.; Joll, C. A.; Kagi, R. Size Exclusion Chromatography To Characterize DOC Removal in Drinking Water Treatment. *Environ. Sci. Technol.* **2005**, *39* (7), 2334–2342.

(22) Kothawala, D. N.; Evans, R. D.; Dillon, P. J. Changes in the molecular weight distribution of dissolved organic carbon within a Precambrian shield stream. *Water Resour. Res.* **2006**, *42* (5).

(23) Wagner, S.; Jaffé, R. Effect of photodegradation on molecular size distribution and quality

451 of dissolved black carbon. *Org. Geochem.* **2015**, *86*, 1–4.

452 (24) Li, W. T.; Chen, S. Y.; Xu, Z. X.; Li, Y.; Shuang, C. D.; Li, A. M. Characterization of
453 dissolved organic matter in municipal wastewater using fluorescence PARAFAC analysis
454 and chromatography multi-excitation/emission scan: A comparative study. *Environ. Sci.*
455 *Technol.* **2014**, *48* (5), 2603–2609.

456 (25) Koch, B. P.; Witt, M.; Engbrodt, R.; Dittmar, T.; Kattner, G. Molecular formulae of
457 marine and terrigenous dissolved organic matter detected by electrospray ionization
458 Fourier transform ion cyclotron resonance mass spectrometry. *Geochim. Cosmochim. Acta*
459 **2005**, *69* (13), 3299–3308.

460 (26) Everett, C. R.; Chin, Y. P.; Aiken, G. R. High-pressure size exclusion chromatography
461 analysis of dissolved organic matter isolated by tangential-flow ultrafiltration. *Limnol.*
462 *Oceanogr.* **1999**, *44* (5), 1316–1322.

463 (27) Lepane, V.; Persson, T.; Wedborg, M. Effects of UV-B radiation on molecular weight
464 distribution and fluorescence from humic substances in riverine and low salinity water.
465 *Estuar. Coast. Shelf Sci.* **2003**, *56* (1), 161–173.

466 (28) Huang, H.; Sawade, E.; Cook, D.; Chow, C. W. K.; Drikas, M.; Jin, B. High-performance
467 size exclusion chromatography with a multi-wavelength absorbance detector study on
468 dissolved organic matter characterisation along a water distribution system. *J. Environ.*
469 *Sci. (China)* **2015**, *44*, 235–243.

470 (29) Stedmon, C. A.; Nelson, N. B. The Optical Properties of DOM in the Ocean. In
471 *Biogeochemistry of Marine Dissolved Organic Matter*; Hansell, D. A., Carlson, C. A.,
472 Eds.; Elsevier Inc.: San Diego, 2015; pp 481–508.

473 (30) Coble, P. G. Marine optical biogeochemistry: The chemistry of ocean color. *Chem. Rev.*
474 **2007**, *107* (2), 402–418.

475 (31) Coble, P. G. Characterization of marine and terrestrial DOM in seawater using excitation
476 emission matrix spectroscopy. *Mar Chem* **1996**, *51* (4), 325–346.

477 (32) McKnight, D. M.; Boyer, E. W.; Westerhoff, P. K.; Doran, P. T.; Kulbe, T.; Andersen, D.
478 T. Spectrofluorometric characterization of dissolved organic matter for indication of
479 precursor organic material and aromaticity. *Limnol. Oceanogr.* **2001**, *46* (1), 38–48.

480 (33) Parlanti, E.; Wörz, K.; Geoffroy, L.; Lamotte, M. Dissolved organic matter fluorescence
481 spectroscopy as a tool to estimate biological activity in a coastal zone submitted to
482 anthropogenic inputs. *Org. Geochem.* **2000**, *31* (12), 1765–1781.

483 (34) Boyd, T. J.; Osburn, C. L. Changes in CDOM fluorescence from allochthonous and
484 autochthonous sources during tidal mixing and bacterial degradation in two coastal

485 estuaries. *Mar. Chem.* **2004**, *89* (1–4), 189–210.

486 (35) Bro, R. PARAFAC. Tutorial and applications. *Chemom. Intell. Lab. Syst.* **1997**, *38* (2),
487 149–171.

488 (36) Stedmon, C. A.; Markager, S.; Bro, R. Tracing dissolved organic matter in aquatic
489 environments using a new approach to fluorescence spectroscopy. *Mar. Chem.* **2003**, *82*
490 (3–4), 239–254.

491 (37) Dittmar, T.; Fitznar, H. P.; Kattner, G. Origin and biogeochemical cycling of organic
492 nitrogen in the eastern Arctic Ocean as evident from D-and L-amino acids. *Geochim.*
493 *Cosmochim. Acta* **2001**, *65* (22), 4103–4114.

494 (38) Gonsior, M.; Valle, J.; Schmitt-Kopplin, P.; Hertkorn, N.; Bastviken, D.; Luek, J.; Harir,
495 M.; Bastos, W.; Enrich-Prast, A. Chemodiversity of dissolved organic matter in the
496 Amazon Basin. *Biogeosciences* **2016**, *13* (14), 4279–4290.

497 (39) Ishii, S. K. L.; Boyer, T. H. Behavior of Reoccurring PARAFAC Components in
498 Fluorescent Dissolved Organic Matter in Natural and Engineered Systems: A Critical
499 Review. *Environ. Sci. Technol.* **2012**, *46* (4), 2006–2017.

500 (40) Cory, R. M.; McKnight, D. M. Fluorescence spectroscopy reveals ubiquitous presence of
501 oxidized and reduced quinones in dissolved organic matter. *Environ. Sci. Technol.* **2005**,
502 *39* (21), 8142–8149.

503 (41) Murphy, K. R.; Stedmon, C. A.; Wenig, P.; Bro, R. OpenFluor- an online spectral library
504 of auto-fluorescence by organic compounds in the environment. *Anal. Methods* **2014**, *6*
505 (3), 658–661.

506 (42) Murphy, K. R.; Butler, K. D.; Spencer, R. G. M.; Stedmon, C. A.; Boehme, J. R.; Aiken,
507 G. R. Measurement of dissolved organic matter fluorescence in aquatic environments: An
508 interlaboratory comparison. *Environ. Sci. Technol.* **2010**, *44* (24), 9405–9412.

509 (43) Murphy, K. R.; Bro, R.; Stedmon, C. A. Chemometric analysis of organic matter
510 fluorescence. In *Aquatic organic matter fluorescence*; Coble, P., Baker, A., Lead, J.,
511 Reynolds, D., Spencer, R., Eds.; Cambridge University Press: New York, 2014; pp 339–
512 375.

513 (44) Beer. Bestimmung der Absorption des rothen Lichts in farbigen Flüssigkeiten. *Ann. Phys.*
514 **1852**, *162* (5), 78–88.

515 (45) Murphy, K. R.; Stedmon, C. A.; Graeber, D.; Bro, R. Fluorescence spectroscopy and
516 multi-way techniques. PARAFAC. *Anal. Methods* **2013**, *5* (23), 6557.

517 (46) Andrew, A. A.; Del Vecchio, R.; Subramaniam, A.; Blough, N. V. Chromophoric

dissolved organic matter (CDOM) in the Equatorial Atlantic Ocean: Optical properties and their relation to CDOM structure and source. *Mar. Chem.* **2013**, *148*, 33–43.

(47) Andrew, A. A.; Del Vecchio, R.; Zhang, Y.; Subramaniam, A.; Blough, N. V. Are Extracted Materials Truly Representative of Original Samples? Impact of C18 Extraction on CDOM Optical and Chemical Properties. *Front. Chem.* **2016**, *4* (February), 1–12.

(48) McKay, G.; Couch, K. D.; Mezyk, S. P.; Rosario-Ortiz, F. L. Investigation of the Coupled Effects of Molecular Weight and Charge-Transfer Interactions on the Optical and Photochemical Properties of Dissolved Organic Matter. *Environ. Sci. Technol.* **2016**, *50* (15), 8093–8102.

(49) Zhang, Y.; Del Vecchio, R.; Blough, N. V. Investigating the mechanism of hydrogen peroxide photoproduction by humic substances. *Environ. Sci. Technol.* **2012**, *46* (21), 11836–11843.

(50) Schendorf, T. M.; Del Vecchio, R.; Koech, K.; Blough, N. V. A standard protocol for NaBH4 reduction of CDOM and HS. *Limnol. Oceanogr. Methods* **2016**, *14* (6), 414–423.

(51) Halim, M.; Spaccini, R.; Parlanti, E.; Amezghal, A.; Piccolo, A. Differences in fluorescence properties between humic acid and its size fractions separated by preparative HPSEC. *J. Geochemical Explor.* **2013**, *129*, 23–27.

(52) Dittmar, T.; Koch, B. P.; Hertkorn, N.; Kattner, G. A simple and efficient method for the solid-phase extraction of dissolved organic matter (SPE-DOM) from seawater. *Limnol. Ocean. Methods* **2008**, *6*, 230–235.

(53) Aiken, G. R. Isolation and concentration techniques for aquatic humic substances. In *Humic Substances in Soil, Sediment, and Water*; Aiken, G. R., McKnight, D. M., Wershaw, R. L., MacCarthy, P., Eds.; New York, 1985; pp 363–385.

(54) Cawley, K. M.; McKnight, D. M.; Miller, P.; Cory, R.; Fimmen, R. L.; Guerard, J.; Dieser, M.; Jaros, C.; Chin, Y.-P.; Foreman, C. Characterization of fulvic acid fractions of dissolved organic matter during ice-out in a hyper-eutrophic, coastal pond in Antarctica. *Environ. Res. Lett.* **2013**, *8* (4), 1–10.

(55) Malcolm, R. L. The uniqueness of humic substances in each of soil, stream and marine environments. *Anal. Chim. Acta* **1990**, *232* (C), 19–30.

(56) Koch, B. P.; Ludwichowski, K. U.; Kattner, G.; Dittmar, T.; Witt, M. Advanced characterization of marine dissolved organic matter by combining reversed-phase liquid chromatography and FT-ICR-MS. *Mar. Chem.* **2008**, *111* (3–4), 233–241.

(57) Sandron, S.; Rojas, A.; Wilson, R.; Davies, N. W.; Haddad, P. R.; Shellie, R. A.; Nesterenko, P. N.; Kelleher, B. P.; Paull, B. Chromatographic methods for the isolation,

separation and characterisation of dissolved organic matter. *Environ. Sci. Process. Impacts* **2015**, *17* (9), 1531–1567.

(58) Hertkorn, N.; Frommberger, M.; Witt, M.; Koch, B. P.; Schmitt-Kopplin, P.; Perdue, E. M. Natural Organic Matter and the Event Horizon of Mass Spectrometry. *Anal. Chem.* **2008**, *80* (23), 8908–8919.

(59) Schmidt, B.; Jaroszewski, J. W.; Bro, R.; Witt, M.; Stærk, D. Combining PARAFAC analysis of HPLC-PDA profiles and structural characterization using HPLC-PDA-SPE-NMR-MS experiments: Commercial preparations of St. John's wort. *Anal. Chem.* **2008**, *80* (6), 1978–1987.

(60) Andersson, C. A.; Bro, R. The N-way Toolbox for MATLAB. *Chemom. Intell. Lab. Syst.* **2000**, *52* (1).

(61) Lorenzo-Seva, U.; ten Berge, J. M. F. Tucker's congruence coefficient as a meaningful index of factor similarity. *Methodology* **2006**, *2* (2), 57–64.

(62) Boyle, E. S.; Guerriero, N.; Thiallet, A.; Del Vecchio, R.; Blough, N. V. Optical properties of humic substances and CDOM: Relation to structure. *Environ. Sci. Technol.* **2009**, *43* (7), 2262–2268.

(63) Ma, J.; Del Vecchio, R.; Golanoski, K. S.; Boyle, E. S.; Blough, N. V. Optical properties of humic substances and CDOM: Effects of borohydride reduction. *Environ. Sci. Technol.* **2010**, *44* (14), 5395–5402.

(64) Sharpless, C. M.; Blough, N. V. The importance of charge-transfer interactions in determining chromophoric dissolved organic matter (CDOM) optical and photochemical properties. *Environ. Sci. Process. Impacts* **2014**, *16* (4), 654–671.

(65) Spencer, R. G. M.; Bolton, L.; Baker, A. Freeze/thaw and pH effects on freshwater dissolved organic matter fluorescence and absorbance properties from a number of UK locations. *Water Res.* **2007**, *41* (13), 2941–2950.

(66) Poulin, B. A.; Ryan, J. N.; Aiken, G. R. The effects of iron on optical properties of dissolved organic matter. *Environ. Sci. Technol. Technol.* **2014**, *48* (17), 10098–10106.

(67) Wünsch, U. J.; Murphy, K. R.; Stedmon, C. A. Fluorescence Quantum Yields of Natural Organic Matter and Organic Compounds: Implications for the Fluorescence-based Interpretation of Organic Matter Composition. *Front. Mar. Sci.* **2015**, *2* (November), 1–15.

(68) Zepp, R. G.; Sheldon, W. M.; Moran, M. A. Dissolved organic fluorophores in southeastern US coastal waters: Correction method for eliminating Rayleigh and Raman scattering peaks in excitation-emission matrices. *Mar. Chem.* **2004**, *89* (1–4), 15–36.

(69) Elcoroaristizabal, S.; Bro, R.; García, J. A.; Alonso, L. PARAFAC models of fluorescence data with scattering: A comparative study. *Chemom. Intell. Lab. Syst.* **2015**, *142*, 124–130.

(70) Andersen, C. M.; Bro, R. Practical aspects of PARAFAC modeling of fluorescence excitation-emission data. *J. Chemom.* **2003**, *17* (4), 200–215.

(71) Wünsch, U. J.; Stedmon, C. A.; Tranvik, L. J.; Guillmette, F.; Guillemette, F. Unraveling the size-dependent optical properties of dissolved organic matter. *Limnol Ocean.* **2017**.

(72) Gao, Y.; Yan, M.; Korshin, G. V. Effects of Ionic Strength on the Chromophores of Dissolved Organic Matter. *Environ. Sci. Technol.* **2015**, *49* (10), 5905–5912.

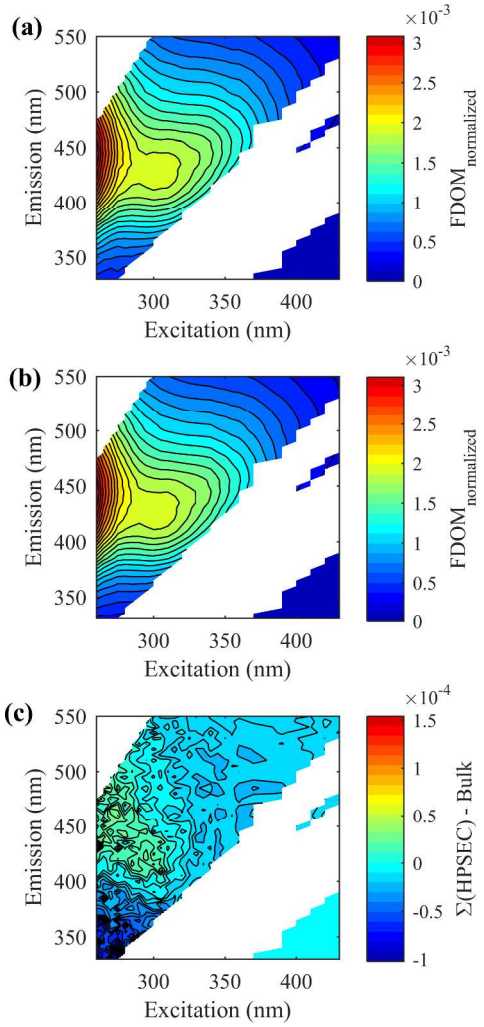
(73) Del Vecchio, R.; Blough, N. V. On the origin of the optical properties of humic substances. *Environ. Sci. Technol.* **2004**, *38*, 3885–3891.

(74) Sutton, R.; Sposito, G. Critical Review Molecular Structure in Soil Humic Substances : The New View. *Environ. Sci. Technol.* **2005**, *39* (510), 9009–9015.

(75) Reemtsma, T.; These, A.; Springer, A.; Linscheid, M. Differences in the molecular composition of fulvic acid size fractions detected by size-exclusion chromatography-on line Fourier transform ion cyclotron resonance (FTICR-) mass spectrometry. *Water Res.* **2008**, *42* (1–2), 63–72.

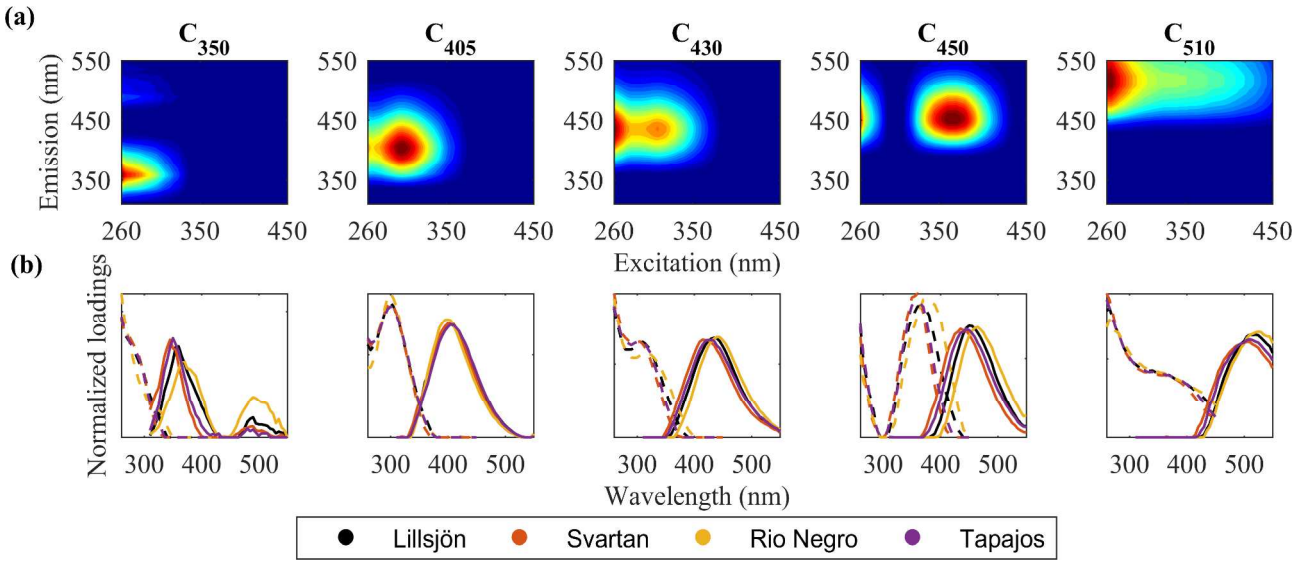
(76) Zsolnay, Á. Dissolved organic matter: Artefacts, definitions, and functions. *Geoderma* **2003**, *113* (3–4), 187–209.

605 **Figure legends**



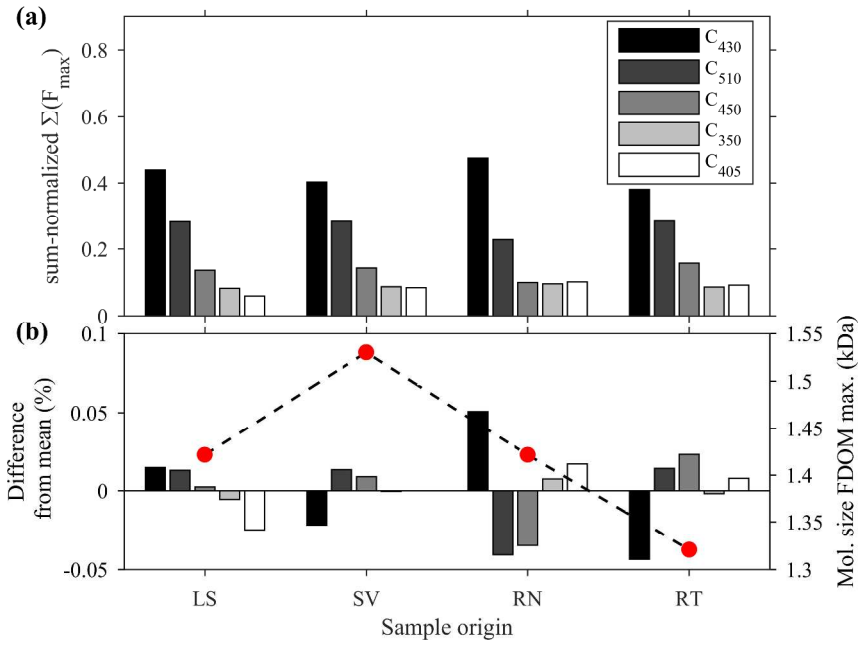
606

607 **Figure 1. Comparison between sum-normalized bulk sample fluorescence (a) and the sum of size separated EEMs (b) of**
608 **lake Lillsjön DOM. The difference between both EEMs is shown in (c). Fluorescence in (a) and (b) was normalized to the**
609 **sum of fluorescence in each EEM.**



610

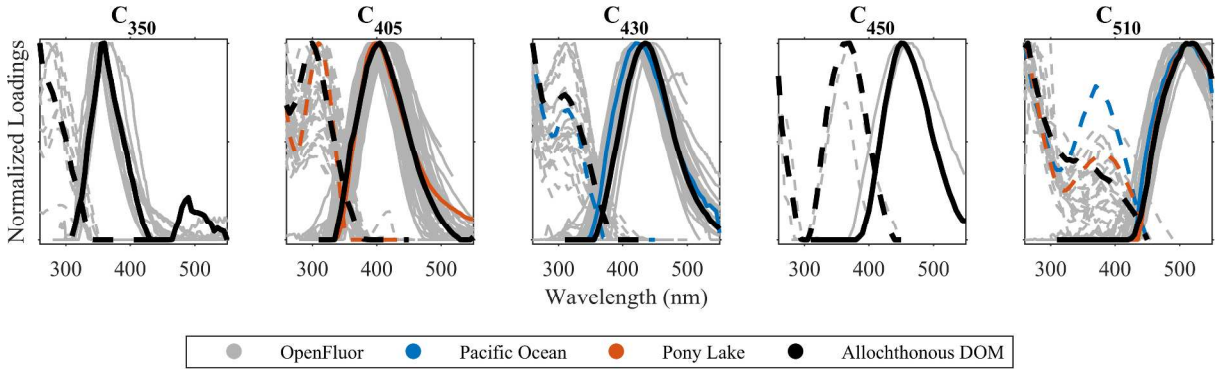
611 **Figure 2. Contour plots of five allochthonous freshwater PARAFAC-derived fluorescence spectra (sample from Lillsjön**
612 **(a)) and comparison between spectral properties of five spectra originating from four different samples and their**
613 **respective models (b). Components are ranked and named according to their respective emission maxima. Tucker**
614 **congruence coefficients are shown in the SI Table S2.**



615

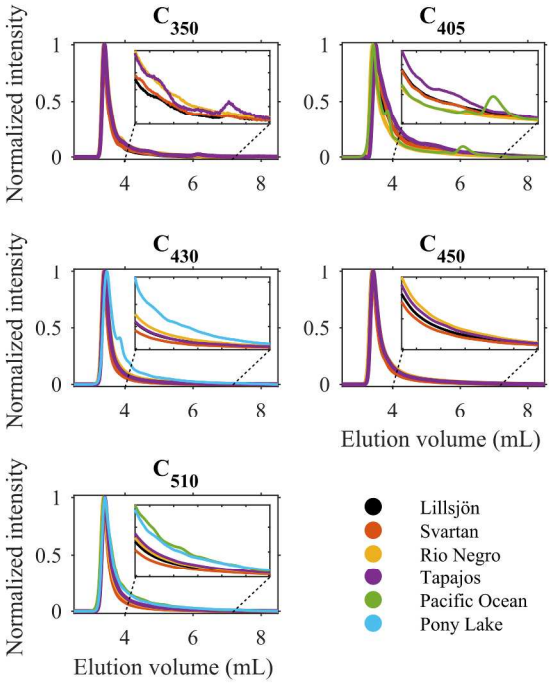
616 **Figure 3. Relative contributions of PARAFAC components to the total fluorescence in the four allochthonous samples (a),**
617 **as well as deviation of the relative contribution of PARAFAC components from the average composition per component**
618 **(b, left axis & bars LS = Lillsjön, SV = Svartan, RN = Rio Negro, RT = Rio Tapajos) against the molecular size peak**
619 **maximum obtained from the total fluorescence chromatogram (right axis, red dots & dotted line).**

620



621

622 **Figure 4. Spectral congruence between five PARAFAC-derived fluorescence spectra of allochthonous DOM from Lake**
623 **Lillsjön (boreal lake, black line), spectra extracted from the OpenFluor database (gray), and two autochthonous DOM**
624 **samples (Pacific Ocean and Pony Lake, blue and orange lines, respectively). For C_{350}, the emission spectrum above 450nm**
625 **was set to missing numbers since data above that emission wavelength likely represented an artefact related to leftover**
626 **physical scatter.**



627

628 **Figure 5. Comparison of chromatograms of five PARAFAC components from four allochthonous samples and two**
629 **autochthonous samples (only for components with sufficient spectral similarity). To mitigate the high degree of**
630 **correlation seen in most components, all datasets were log-normalized prior to modeling and the normalization was**
631 **reversed post-fitting to obtain the original chromatograms. Inserts show elution profiles between 4 and 7 mL.**