



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](https://doi.org/10.1021/acs.est.7b03260)

Publication date:
2017

Document Version
Peer reviewed version

[Link back to DTU Orbit](#)

Citation (APA):
Wünsch, U., Murphy, K. R., & Stedmon, C. (2017). The one-sample PARAFAC approach reveals molecular size distributions of fluorescent components in dissolved organic matter. *Environmental Science & Technology (Washington)*, 51(20), 11900-11908. <https://doi.org/10.1021/acs.est.7b03260>

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

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

Downloaded from <http://pubs.acs.org> on September 27, 2017

Just Accepted

“Just Accepted” manuscripts have been peer-reviewed and accepted for publication. They are posted online prior to technical editing, formatting for publication and author proofing. The American Chemical Society provides “Just Accepted” as a free service to the research community to expedite the dissemination of scientific material as soon as possible after acceptance. “Just Accepted” manuscripts appear in full in PDF format accompanied by an HTML abstract. “Just Accepted” manuscripts have been fully peer reviewed, but should not be considered the official version of record. They are accessible to all readers and citable by the Digital Object Identifier (DOI®). “Just Accepted” is an optional service offered to authors. Therefore, the “Just Accepted” Web site may not include all articles that will be published in the journal. After a manuscript is technically edited and formatted, it will be removed from the “Just Accepted” Web site and published as an ASAP article. Note that technical editing may introduce minor changes to the manuscript text and/or graphics which could affect content, and all legal disclaimers and ethical guidelines that apply to the journal pertain. ACS cannot be held responsible for errors or consequences arising from the use of information contained in these “Just Accepted” manuscripts.

1 The one-sample PARAFAC approach reveals molecular size 2 distributions of fluorescent components in dissolved organic 3 matter

4 Urban J. Wünsch^{1*}, Kathleen R. Murphy², Colin A. Stedmon¹

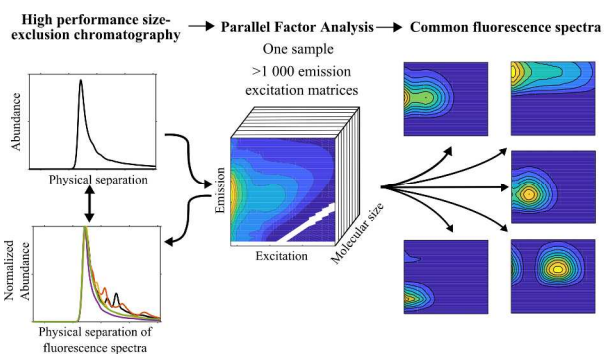
5 ¹National Institute of Aquatic Resources, Section for Oceans and Arctic, Technical University of
6 Denmark, Kemitorvet, Building 201, 2800 Kgs. Lyngby, Denmark

7 ²Chalmers University of Technology, Water Environment Technology, Sven Hultins Gata 6, 41296
8 Gothenburg, Sweden

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

10 Abstract

11 Molecular size plays an important role in dissolved
12 organic matter (DOM) biogeochemistry, but its
13 relationship with the fluorescent fraction of DOM
14 (FDOM) remains poorly resolved. Here high-
15 performance size exclusion chromatography
16 (HPSEC) was coupled to fluorescence emission-



17 excitation (EEM) spectroscopy in full spectral (60 emission and 34 excitation wavelengths) and
18 chromatographic resolution (< 1Hz), to enable the mathematical decomposition of fluorescence on an
19 individual sample basis by parallel factor analysis (PARAFAC). The approach allowed cross-system
20 comparisons of molecular size distributions for individual fluorescence components obtained from
21 independent datasets. Spectra extracted from allochthonous DOM were highly similar. Allochthonous and

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

31 Introduction

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

41 Optical properties of different DOM size fractions have provided evidence for the supramolecular
42 assembly hypothesis,¹⁷ whereby individual DOM moieties recur in non-covalently bound assemblies of
43 varying molecular size. In support of this hypothesis, highly similar optical properties are seen across the
44 entire molecular-size gradient of DOM.^{11,18-20} Apparent molecular size distributions of DOM are typically
45 analyzed by high-performance size exclusion chromatography (HPSEC) to study changes in the size
46 distribution of DOM as a function of biogeochemical and physical factors.²⁰⁻²⁴ Molecular size
47 distributions depend on the instrument used for its measurement: Mass spectrometry most often shows an
48 average molecular weight around 400 Da,²⁵ while HPSEC or ultrafiltration show higher averages.^{4,26,27}
49 However, many studies utilizing HPSEC are based on measurements of discrete sample fractionations,
50 and hence have relatively low sample (i.e. chromatographic) resolution. Recent studies used online
51 detectors to provide high resolution data (i.e. < 1 Hz);^{14,20,28} however in fluorescence studies, a limited
52 number of excitation wavelengths were monitored simultaneously by these detectors, and a systematic,
53 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.

95 **Materials and Methods**

96 **Sample Collection**

97 Four allochthonous samples (Lake Lillsjön, Sweden; Rio Negro, Brazil; Svartan River, Sweden; Rio
98 Tapajós, Brazil) and two autochthonous samples (Pacific Ocean & Pony Lake) were extracted with PPL
99 and XAD8 resins, respectively, using established methods (see Supporting Information (SI) S1 and Table
100 S1 for further information).^{52–55} 200 μL of the PPL extract was dried and SPE-DOM was reconstituted in
101 0.15 M ammonium acetate (pH 7). DOM of two XAD-8 extracts was dissolved in 0.15 M ammonium
102 acetate at concentrations of 1.4 mg L^{-1} and 0.25 mg L^{-1} for samples originating from Pony Lake and the
103 Pacific Ocean, respectively.

104 **High performance size-exclusion chromatography**

105 Full details pertaining to the HPSEC equipment, and methodology are provided in the Supporting
106 Information (S1, SI Figures S1-S6). Briefly, HPSEC was performed using a Shimadzu Nexera X2 UFLC
107 system equipped with a TSKgel SuperAWM-H column. DOM was eluted with 0.15 M ammonium acetate
108 (pH 7), and two sequential detectors were used. Absorbance was measured between 240 and 600 nm at
109 2 nm intervals using a Shimadzu SPD-M30. Fluorescence emission was then detected between 300 and
110 600 nm at 5 nm increments across excitation wavelengths from 240 – 450 nm at 5-10 nm increments
111 using a Shimadzu RF-30Axs by combining measurements from separate runs. For every sample, a
112 sequence of runs was performed whereby the same sample was injected while instrument parameters were
113 systematically changed between runs. In total, the chromatographic run was repeated 35 times to allow
114 the determination of absorbance properties (one run) and fluorescence properties at an EEM-like spectral
115 resolution (34 runs, one for every excitation wavelength at a constant emission wavelength range). This
116 resulted in around 1500 individual absorbance spectra and fluorescence emission scans (each of the 1500

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

129 Chemometric analysis

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

$$135 \quad x_{ijk} = \sum_{f=1}^F a_{if} b_{jf} c_{kf} + e_{ijk} \quad (\text{Eq. 1})$$

136 $i = 1 \dots I; j = 1 \dots J; k = 1 \dots K$

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

139 analyte in sample i . The term b_{if} represents the least-square estimate of the emission spectrum of the f th
140 analyte, while c_{kf} is the least-square estimate of the excitation spectrum of the f th analyte at wavelength k .
141 The term e_{ijk} represents the residual matrix that contains unexplained dataset variability. Importantly, the
142 successful decomposition of chemical datasets into underlying factors using the PARAFAC model hinges
143 on three assumptions: (1) Variability: No two compounds can have the same exact spectral properties and
144 identical fluorescence intensities; (2) Additivity: The total fluorescence intensities observed are the result
145 of the fluorescence of a finite number of analytes that do not interact electronically; and (3) Trilinearity:
146 The signal of a given analyte is linearly related to its invariant excitation and emission spectrum, i.e. one
147 component describes an analyte in all three modes. The combination of applying PARAFAC to
148 decompose HPSEC-derived EEMs is represented herein using the terminology HPSEC-EEM-PARAFAC.

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

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

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

170 Results and Discussion

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

172 The unique coupling of HPSEC and full-resolution EEM spectroscopy in this study presented the
173 opportunity to investigate the additive behavior of DOM fluorescence (i.e. compliance with Beer's law).
174 The sum of fluorescence emission from size-separated EEMs was compared to the bulk EEM obtained on
175 the same instrument without the chromatographic column installed. This was performed for two
176 representative samples (Fig. 1, SI Figure S8). Size separation did not produce substantial changes of
177 fluorescence in the visible fluorescence emission region (excitation > 300 nm), as might be expected from
178 intermolecular charge transfer or fluorescence quenching.^{46,62-64} However, at excitation wavelengths
179 below 300 nm, two regions deviated from the otherwise randomly-distributed residuals. In the UVA
180 region, a negative residual of less than 20 % was observed relative to the bulk EEM, indicating loss of
181 fluorescence during separation. This was likely caused by adsorption of small monomers onto the
182 analytical column due to secondary interactions, since pure tryptophan and salicylic acid also showed
183 secondary retention (data not shown). Secondly, a positive residual of <4 % indicating a gain of
184 fluorescence was seen in the emission range between 360 and 470 nm when excitation was below
185 300 nm. This small gain is likely attributable to a weak background fluorescence signal emitted by the
186 mobile phase, which constantly eluted from the analytical column despite an auto-zero blank subtraction
187 at the beginning of each run. Despite these minor differences, the spectral shape of bulk and size-
188 separated EEMs was highly similar. TCCs between fluorescence emission at all excitations was higher
189 than 0.9997, but did show lower values at low excitation wavelengths (<300 nm) due to the lack of
190 protein-like fluorescence in the size-separated EEMs (SI Figure S8).

191 Overall, the additivity of fluorescence within the framework of HPSEC separation was
192 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.

203 Spectral conformity among allochthonous samples

204 The four allochthonous DOM extracts originated from freshwater environments in Sweden and Brazil
205 receiving a large proportion of terrestrial organic matter. They were analyzed independently using
206 HPSEC-EEM-PARAFAC to decompose EEMs into independent fluorescence components. On average,
207 PARAFAC models with two to five components explained respectively 99.51, 99.78, 99.88, and 99.92 %
208 of variability in each dataset, and all models with the same number of components (two to five
209 components) had highly congruent underlying fluorescence spectra (SI Figure S9, SI Table S2). For all
210 four samples, the five-component PARAFAC model (Fig. 2) best represented their fluorescence
211 properties (SI Figure S10), since four component models did not adequately represent protein-like
212 fluorescence, while core consistencies, sum-of-squared-errors, and spectral loadings of six component
213 models frequently implied over-factorization. The spectral congruence between the independent datasets is
214 interpreted as compelling evidence for the validity of the four individual models (i.e. similar to a
215 conventional split-half analysis). The five components of the validated PARAFAC model were named
216 according to their fluorescence emission maximum as follows: C₃₅₀, C₄₀₅, C₄₃₀, C₄₅₀, C₅₁₀ (Fig. 2).

217 Despite high overall similarity between components in all four models (Fig. 2), their spectral
218 congruence did not always meet the criterion of $TCC \geq 0.95$ that is often applied to identify
219 interchangeable spectra (SI Table S2).^{41,45} However, in all comparisons, TCC exceeded the threshold of
220 “fair” similarity (0.85) and in all but one cases exceeded 0.90 (C₃₅₀ between Lake Lillsjön and Svartan
221 River).⁶¹ Common sources that might result in minor spectral changes, such as sample pH and metal-
222 quenching, were eliminated by the combination of DOM extraction and controlled buffer conditions
223 during chromatographic separations.^{65,66} Thus, the spectral differences observed between otherwise
224 similar models must have originated from other sources. Since fluorescence properties of aromatic
225 structures are influenced by conjugation and substitution, structural variations in similar fluorophores

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

244 **Comparison between autochthonous, allochthonous and community-derived fluorescence spectra**

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

260 Despite greater variability, the fluorescence spectra derived from the autochthonous samples
261 partially matched with spectra derived from the allochthonous samples. Components closely matching
262 allochthonous C₅₁₀ (as identified in Lake Lillsjön) were found in Pony Lake and Pacific Ocean FDOM.
263 Thus, C₅₁₀ was the only ubiquitous component across all investigated samples (Fig. 4). Additionally,
264 components closely resembling C₄₀₅ and C₄₃₀ were present in FDOM from Pony Lake and the Pacific
265 Ocean, respectively (Fig. 4).

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

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

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

290 In the context of HPSEC-EEM-PARAFAC, component scores represent molecular size as the primary
291 chromatographic separation mechanism. As stated above, spectral loadings of some components
292 originating from individual PARAFAC models were strongly congruent and therefore warranted further
293 comparison to examine apparent molecular size distributions between samples originating from different
294 aquatic environments. The supramolecular assembly hypothesis states that individual DOM moieties (e.g.
295 fluorescing compounds) form non-covalently bound assemblies (including non-fluorescing compounds)
296 of varying molecular size.^{17,74} Evidence supporting this hypothesis is based on the highly similar character
297 of DOM obtained from HPSEC-based fractions as observed by mass spectrometry,⁷⁵ infrared
298 spectroscopy,¹⁷ and fluorescence spectroscopy.^{11,18–20} In this light, the apparent molecular size
299 distributions are expected to be broad and unresolved. In agreement with this hypothesis, components
300 originating from allochthonous DOM showed highly similar molecular size distributions with poor
301 physical separation (Fig. 5, SI Fig. S13). Components generally exhibited a single peak with tailing
302 towards higher elution volumes (low apparent molecular size). The molecular size distributions of
303 PARAFAC components other than C₄₀₅ were very similar across samples (TCC>0.98, SI Fig. S13). The
304 observation of broad, overlapping distributions instead of distinct, resolved peaks thus aligns with earlier
305 findings, although the combined chromatographic and mathematic approach employed here provides
306 unprecedented detail due to the utilization of online detectors (< 1 Hz) instead of discrete fractionation.

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

311 components were not correlated to fluorescence emission. Across the allochthonous samples, the average
312 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
313 components C_{510} , C_{350} , C_{430} , C_{450} , and C_{405} , respectively ($R^2 = -0.22$, $p > 0.1$). Moreover, no relationship
314 between the FDOM composition (as observed by relative contributions of PARAFAC components to the
315 total HPSEC-EEM fluorescence) and molecular size of total fluorescence was apparent (Fig. 3, dashed
316 line and red dots). Although a direct correlation between fluorescence emission and molecular size might
317 be expected for simple mixtures, our results suggest that this was not the case for the complex mixtures
318 analyzed in this study. Our findings rather suggest that FDOM components were possibly associated with
319 non-fluorescing organic matter with a range of three-dimensional structures / sizes, thus convoluting the
320 relationship between fluorescence emission and molecular size. The contradictory results may arise at
321 least in part from the different analytical approaches between studies, but may also result from differences
322 in sample preparation, choice of analytical column, or the overall degree of compositional variability in
323 each dataset. While results in this study suggest that bulk FDOM is an unreliable indicator of the average
324 molecular size of DOM, further investigation with additional samples is warranted. This finding also
325 highlights the need for additional analytical detectors (such as refractive index or mass spectrometry) to
326 be included in HPSEC analyses, since a combination of detectors with overlapping analytical windows
327 will provide deeper insights into the molecular assemblies of DOM.

328 HPSEC-EEM-PARAFAC demonstrated that apparent molecular size distributions of spectrally congruent
329 fluorescence spectra may differ between samples. We identified differences in apparent size distributions,
330 most notably in the low molecular size range, in particular for the poorly-resolved peaks of components
331 C_{350} , C_{405} and C_{450} in several samples (Fig. 5 inserts). Notably, experiments with pure fluorophores
332 suggested the presence of secondary column interactions with compounds of low molecular size. Thus, a
333 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).

354 HPSEC-EEM-PARAFAC: Implications and future directions

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

371 Acknowledgements

372 This work is dedicated to George Aiken in recognition of his contribution to the field of DOM
373 biogeochemistry and characterization, and his role as a mentor for so many young scientists. This work
374 was in part funded by the Danish Council for Independent Research-Natural Sciences Grant DFF-1323-
375 00336 and Nordic5Tech collaborative funding (Technical University of Denmark). KRM acknowledges
376 support by the Swedish Research Council (Formas 2013–1214) and the DRICKS drinking water research
377 center. The authors would like to thank Michael Gonsior (University of Maryland) for providing organic
378 matter extracts, as well as Diane McKnight (University of Colorado Boulder) and George Aiken (United
379 States Geological Survey) for providing XAD-8 organic matter extracts. UJW would like to thank Adam
380 Hambly for proof-reading the manuscript.

381 Supporting Information

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

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

384 **References**

- 385 (1) Hansell, D. A.; Carlson, C. A.; Repeta, D. J.; Schlitzer, R. Dissolved Organic Matter in the
386 Ocean A Controversy Stimulates New Insights. *Oceanography* **2009**, *22* (4), 202–211.
- 387 (2) Cole, J. J.; Prairie, Y. T.; Caraco, N. F.; McDowell, W. H.; Tranvik, L. J.; Striegl, R. G.;
388 Duarte, C. M.; Kortelainen, P.; Downing, J. A.; Middelburg, J. J.; et al. Plumbing the
389 global carbon cycle: Integrating inland waters into the terrestrial carbon budget.
390 *Ecosystems* **2007**, *10* (1), 171–184.
- 391 (3) Tranvik, L. J.; Downing, J. A.; Cotner, J. B.; Loiselle, S. A.; Striegl, R. G.; Ballatore, T.
392 J.; Dillon, P.; Finlay, K.; Fortino, K.; Knoll, L. B. Lakes and reservoirs as regulators of
393 carbon cycling and climate. *Limnol. Oceanogr.* **2009**, *54* (6, part 2), 2298–2314.
- 394 (4) Helms, J. R.; Stubbins, A.; Ritchie, J. D.; Minor, E. C.; Kieber, D. J.; Mopper, K.
395 Absorption spectral slopes and slope ratios as indicators of molecular weight, source, and
396 photobleaching of chromophoric dissolved organic matter. *Limnology Oceanogr.* **2008**,
397 *53* (3), 955–969.
- 398 (5) Fichot, C. G.; Benner, R. The spectral slope coefficient of chromophoric dissolved organic
399 matter (S₂₇₅₋₂₉₅) as a tracer of terrigenous dissolved organic carbon in river-influenced
400 ocean margins. *Limnol. Oceanogr.* **2012**, *57* (5), 1453–1466.
- 401 (6) Weishaar, J. L.; Aiken, G. R.; Bergamaschi, B. A.; Fram, M. S.; Fujii, R.; Mopper, K.
402 Evaluation of specific ultraviolet absorbance as an indicator of the chemical composition
403 and reactivity of dissolved organic carbon. *Environ. Sci. Technol.* **2003**, *37* (20), 4702–
404 4708.
- 405 (7) Amon, R. M. W.; Benner, R. Rapid cycling of high-molecular-weight dissolved organic
406 matter in the ocean. *Nature* **1994**, *369* (6481), 549–552.
- 407 (8) Amon, R. M. W.; Benner, R. Bacterial utilization of different size classes of dissolved
408 organic matter. *Limnol Ocean.* **1996**, *41* (1), 41–51.
- 409 (9) Benner, R.; Amon, R. M. W. The Size-Reactivity Continuum of Major Bioelements in the
410 Ocean. *Ann. Rev. Mar. Sci.* **2015**, *7*, 185–205.
- 411 (10) Tranvik, L. J. Bacterioplankton growth on fractions of dissolved organic carbon of
412 different molecular weights from humic and clear waters. *Appl. Environ. Microbiol.* **1990**,
413 *56* (6), 1672–1677.
- 414 (11) Her, N.; Amy, G.; McKnight, D. M.; Sohn, J.; Yoon, Y. Characterization of DOM as a
415 function of MW by fluorescence EEM and HPLC-SEC using UVA, DOC, and
416 fluorescence detection. *Water Res.* **2003**, *37*, 4295–4303.

- 417 (12) Wu, F. C.; Evans, R. D.; Dillon, P. J. Separation and Characterization of NOM by High-
418 Performance Liquid Chromatography and On-Line Three-Dimensional Excitation
419 Emission Matrix Fluorescence Detection. *Environ. Sci. Technol.* **2003**, *37* (16), 3687–
420 3693.
- 421 (13) Stubbins, A.; Lapierre, J.-F.; Berggren, M.; Prairie, Y. T.; Dittmar, T.; del Giorgio, P. A.
422 What's in an EEM? Molecular Signatures Associated with Dissolved Organic
423 Fluorescence in Boreal Canada. *Environ. Sci. Technol.* **2014**, *48*, 10598–10606.
- 424 (14) Cuss, C. W.; Guéguen, C. Relationships between molecular weight and fluorescence
425 properties for size-fractionated dissolved organic matter from fresh and aged sources.
426 *Water Res.* **2015**, *68*, 487–497.
- 427 (15) Cuss, C. W.; Guéguen, C. Determination of relative molecular weights of fluorescent
428 components in dissolved organic matter using asymmetrical flow field-flow fractionation
429 and parallel factor analysis. *Anal. Chim. Acta* **2012**, *733*, 98–102.
- 430 (16) Stolpe, B.; Zhou, Z.; Guo, L.; Shiller, A. M. Colloidal size distribution of humic- and
431 protein-like fluorescent organic matter in the northern Gulf of Mexico. *Mar. Chem.* **2014**,
432 *164*, 25–37.
- 433 (17) Peuravuori, J.; Pihlaja, K. Preliminary study of lake dissolved organic matter in light of
434 nanoscale supramolecular assembly. *Environ. Sci. Technol.* **2004**, *38* (22), 5958–5967.
- 435 (18) Romera-Castillo, C.; Chen, M.; Yamashita, Y.; Jaffé, R. Fluorescence characteristics of
436 size-fractionated dissolved organic matter: Implications for a molecular assembly based
437 structure? *Water Res.* **2014**, *55*, 40–51.
- 438 (19) Boehme, J.; Wells, M. Fluorescence variability of marine and terrestrial colloids:
439 Examining size fractions of chromophoric dissolved organic matter in the Damariscotta
440 River estuary. *Mar. Chem.* **2006**, *101* (1–2), 95–103.
- 441 (20) Li, W. T.; Xu, Z. X.; Li, A. M.; Wu, W.; Zhou, Q.; Wang, J. N. HPLC/HPSEC-FLD with
442 multi-excitation/emission scan for EEM interpretation and dissolved organic matter
443 analysis. *Water Res.* **2013**, *47* (3), 1246–1256.
- 444 (21) Allpike, B. P.; Heitz, A.; Joll, C. A.; Kagi, R. Size Exclusion Chromatography To
445 Characterize DOC Removal in Drinking Water Treatment. *Environ. Sci. Technol.* **2005**,
446 *39* (7), 2334–2342.
- 447 (22) Kothawala, D. N.; Evans, R. D.; Dillon, P. J. Changes in the molecular weight distribution
448 of dissolved organic carbon within a Precambrian shield stream. *Water Resour. Res.* **2006**,
449 *42* (5).
- 450 (23) Wagner, S.; Jaffé, R. Effect of photodegradation on molecular size distribution and quality

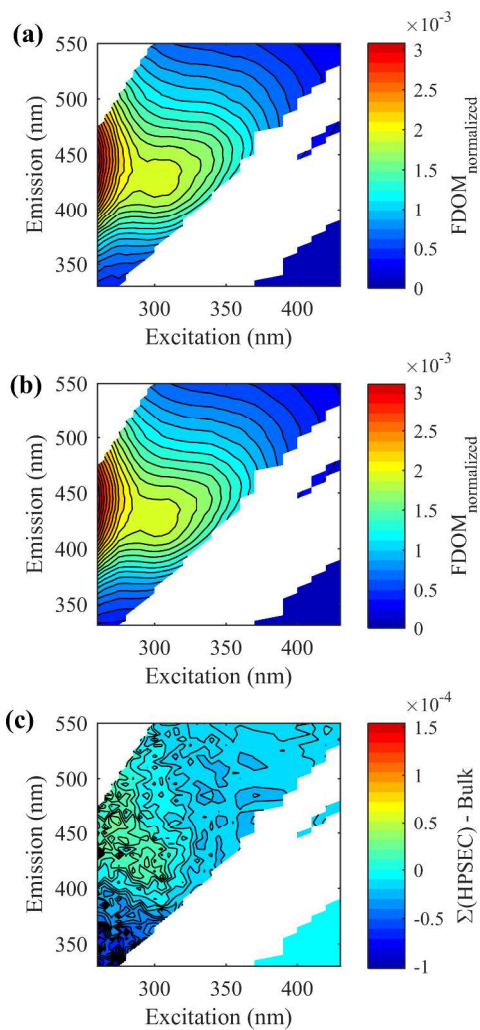
- 451 of dissolved black carbon. *Org. Geochem.* **2015**, *86*, 1–4.
- 452 (24) Li, W. T.; Chen, S. Y.; Xu, Z. X.; Li, Y.; Shuang, C. D.; Li, A. M. Characterization of
453 dissolved organic matter in municipal wastewater using fluorescence PARAFAC analysis
454 and chromatography multi-excitation/emission 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

- 518 dissolved organic matter (CDOM) in the Equatorial Atlantic Ocean: Optical properties
519 and their relation to CDOM structure and source. *Mar. Chem.* **2013**, *148*, 33–43.
- 520 (47) Andrew, A. A.; Del Vecchio, R.; Zhang, Y.; Subramaniam, A.; Blough, N. V. Are
521 Extracted Materials Truly Representative of Original Samples? Impact of C18 Extraction
522 on CDOM Optical and Chemical Properties. *Front. Chem.* **2016**, *4* (February), 1–12.
- 523 (48) McKay, G.; Couch, K. D.; Mezyk, S. P.; Rosario-Ortiz, F. L. Investigation of the Coupled
524 Effects of Molecular Weight and Charge-Transfer Interactions on the Optical and
525 Photochemical Properties of Dissolved Organic Matter. *Environ. Sci. Technol.* **2016**, *50*
526 (15), 8093–8102.
- 527 (49) Zhang, Y.; Del Vecchio, R.; Blough, N. V. Investigating the mechanism of hydrogen
528 peroxide photoproduction by humic substances. *Environ. Sci. Technol.* **2012**, *46* (21),
529 11836–11843.
- 530 (50) Schendorf, T. M.; Del Vecchio, R.; Koech, K.; Blough, N. V. A standard protocol for
531 NaBH₄ reduction of CDOM and HS. *Limnol. Oceanogr. Methods* **2016**, *14* (6), 414–423.
- 532 (51) Halim, M.; Spaccini, R.; Parlanti, E.; Amezghal, A.; Piccolo, A. Differences in
533 fluorescence properties between humic acid and its size fractions separated by preparative
534 HPSEC. *J. Geochemical Explor.* **2013**, *129*, 23–27.
- 535 (52) Dittmar, T.; Koch, B. P.; Hertkorn, N.; Kattner, G. A simple and efficient method for the
536 solid-phase extraction of dissolved organic matter (SPE-DOM) from seawater. *Limnol.*
537 *Ocean. Methods* **2008**, *6*, 230–235.
- 538 (53) Aiken, G. R. Isolation and concentration techniques for aquatic humic substances. In
539 *Humic Substances in Soil, Sediment, and Water*; Aiken, G. R., McKnight, D. M.,
540 Wershaw, R. L., MacCarthy, P., Eds.; New York, 1985; pp 363–385.
- 541 (54) Cawley, K. M.; McKnight, D. M.; Miller, P.; Cory, R.; Fimmen, R. L.; Guerard, J.;
542 Dieser, M.; Jaros, C.; Chin, Y.-P.; Foreman, C. Characterization of fulvic acid fractions of
543 dissolved organic matter during ice-out in a hyper-eutrophic, coastal pond in Antarctica.
544 *Environ. Res. Lett.* **2013**, *8* (4), 1–10.
- 545 (55) Malcolm, R. L. The uniqueness of humic substances in each of soil, stream and marine
546 environments. *Anal. Chim. Acta* **1990**, *232* (C), 19–30.
- 547 (56) Koch, B. P.; Ludwichowski, K. U.; Kattner, G.; Dittmar, T.; Witt, M. Advanced
548 characterization of marine dissolved organic matter by combining reversed-phase liquid
549 chromatography and FT-ICR-MS. *Mar. Chem.* **2008**, *111* (3–4), 233–241.
- 550 (57) Sandron, S.; Rojas, A.; Wilson, R.; Davies, N. W.; Haddad, P. R.; Shellie, R. A.;
551 Nesterenko, P. N.; Kelleher, B. P.; Paull, B. Chromatographic methods for the isolation,

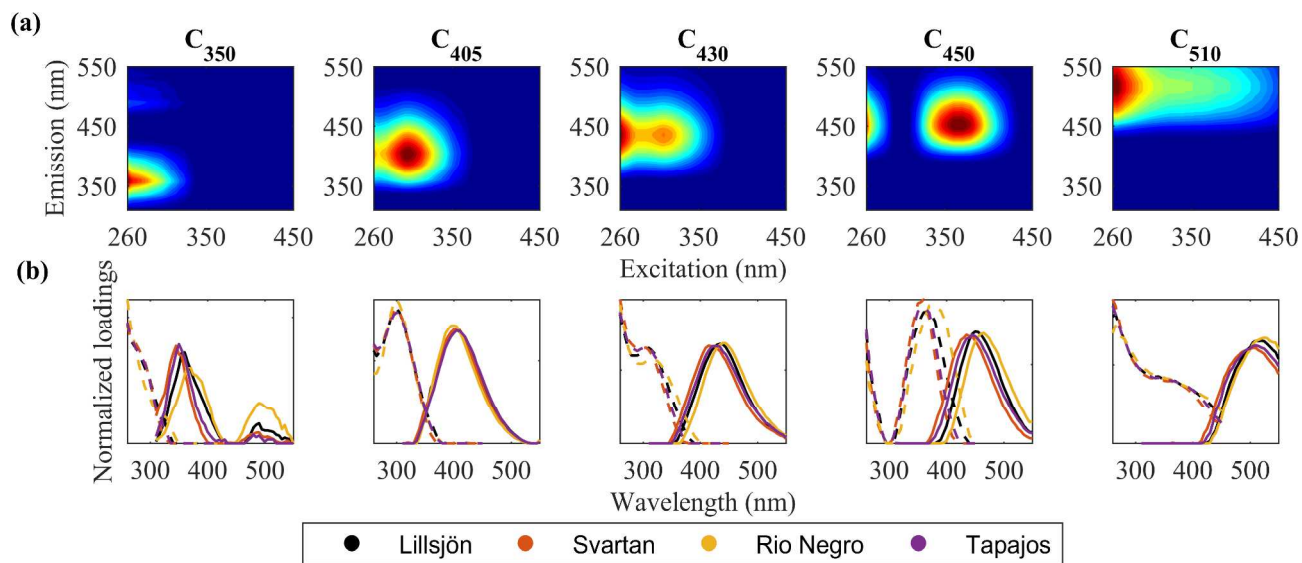
- 552 separation and characterisation of dissolved organic matter. *Environ. Sci. Process. Impacts*
553 **2015**, *17* (9), 1531–1567.
- 554 (58) Hertkorn, N.; Frommberger, M.; Witt, M.; Koch, B. P.; Schmitt-Kopplin, P.; Perdue, E.
555 M. Natural Organic Matter and the Event Horizon of Mass Spectrometry. *Anal. Chem.*
556 **2008**, *80* (23), 8908–8919.
- 557 (59) Schmidt, B.; Jaroszewski, J. W.; Bro, R.; Witt, M.; Stærk, D. Combining PARAFAC
558 analysis of HPLC-PDA profiles and structural characterization using HPLC-PDA-SPE-
559 NMR-MS experiments: Commercial preparations of St. John's wort. *Anal. Chem.* **2008**,
560 *80* (6), 1978–1987.
- 561 (60) Andersson, C. A.; Bro, R. The N-way Toolbox for MATLAB. *Chemom. Intell. Lab. Syst.*
562 **2000**, *52* (1).
- 563 (61) Lorenzo-Seva, U.; ten Berge, J. M. F. Tucker's congruence coefficient as a meaningful
564 index of factor similarity. *Methodology* **2006**, *2* (2), 57–64.
- 565 (62) Boyle, E. S.; Guerriero, N.; Thiallet, A.; Del Vecchio, R.; Blough, N. V. Optical
566 properties of humic substances and CDOM: Relation to structure. *Environ. Sci. Technol.*
567 **2009**, *43* (7), 2262–2268.
- 568 (63) Ma, J.; Del Vecchio, R.; Golanoski, K. S.; Boyle, E. S.; Blough, N. V. Optical properties
569 of humic substances and CDOM: Effects of borohydride reduction. *Environ. Sci. Technol.*
570 **2010**, *44* (14), 5395–5402.
- 571 (64) Sharpless, C. M.; Blough, N. V. The importance of charge-transfer interactions in
572 determining chromophoric dissolved organic matter (CDOM) optical and photochemical
573 properties. *Environ. Sci. Process. Impacts* **2014**, *16* (4), 654–671.
- 574 (65) Spencer, R. G. M.; Bolton, L.; Baker, A. Freeze/thaw and pH effects on freshwater
575 dissolved organic matter fluorescence and absorbance properties from a number of UK
576 locations. *Water Res.* **2007**, *41* (13), 2941–2950.
- 577 (66) Poulin, B. A.; Ryan, J. N.; Aiken, G. R. The effects of iron on optical properties of
578 dissolved organic matter. *Environ. Sci. Technol. Technol.* **2014**, *48* (17), 10098–10106.
- 579 (67) Wünsch, U. J.; Murphy, K. R.; Stedmon, C. A. Fluorescence Quantum Yields of Natural
580 Organic Matter and Organic Compounds: Implications for the Fluorescence-based
581 Interpretation of Organic Matter Composition. *Front. Mar. Sci.* **2015**, *2* (November), 1–
582 15.
- 583 (68) Zepp, R. G.; Sheldon, W. M.; Moran, M. A. Dissolved organic fluorophores in
584 southeastern US coastal waters: Correction method for eliminating Rayleigh and Raman
585 scattering peaks in excitation-emission matrices. *Mar. Chem.* **2004**, *89* (1–4), 15–36.

- 586 (69) Elcoroaristizabal, S.; Bro, R.; García, J. A.; Alonso, L. PARAFAC models of fluorescence
587 data with scattering: A comparative study. *Chemom. Intell. Lab. Syst.* **2015**, *142*, 124–130.
- 588 (70) Andersen, C. M.; Bro, R. Practical aspects of PARAFAC modeling of fluorescence
589 excitation-emission data. *J. Chemom.* **2003**, *17* (4), 200–215.
- 590 (71) Wünsch, U. J.; Stedmon, C. A.; Tranvik, L. J.; Guillemette, F.; Guillemette, F. Unraveling
591 the size-dependent optical properties of dissolved organic matter. *Limnol Ocean.* **2017**.
- 592 (72) Gao, Y.; Yan, M.; Korshin, G. V. Effects of Ionic Strength on the Chromophores of
593 Dissolved Organic Matter. *Environ. Sci. Technol.* **2015**, *49* (10), 5905–5912.
- 594 (73) Del Vecchio, R.; Blough, N. V. On the origin of the optical properties of humic
595 substances. *Environ. Sci. Technol.* **2004**, *38*, 3885–3891.
- 596 (74) Sutton, R.; Sposito, G. Critical Review Molecular Structure in Soil Humic Substances :
597 The New View. *Environ. Sci. Technol.* **2005**, *39* (510), 9009–9015.
- 598 (75) Reemtsma, T.; These, A.; Springer, A.; Linscheid, M. Differences in the molecular
599 composition of fulvic acid size fractions detected by size-exclusion chromatography-on
600 line Fourier transform ion cyclotron resonance (FTICR-) mass spectrometry. *Water Res.*
601 **2008**, *42* (1–2), 63–72.
- 602 (76) Zsolnay, Á. Dissolved organic matter: Artefacts, definitions, and functions. *Geoderma*
603 **2003**, *113* (3–4), 187–209.
- 604

605 **Figure legends**

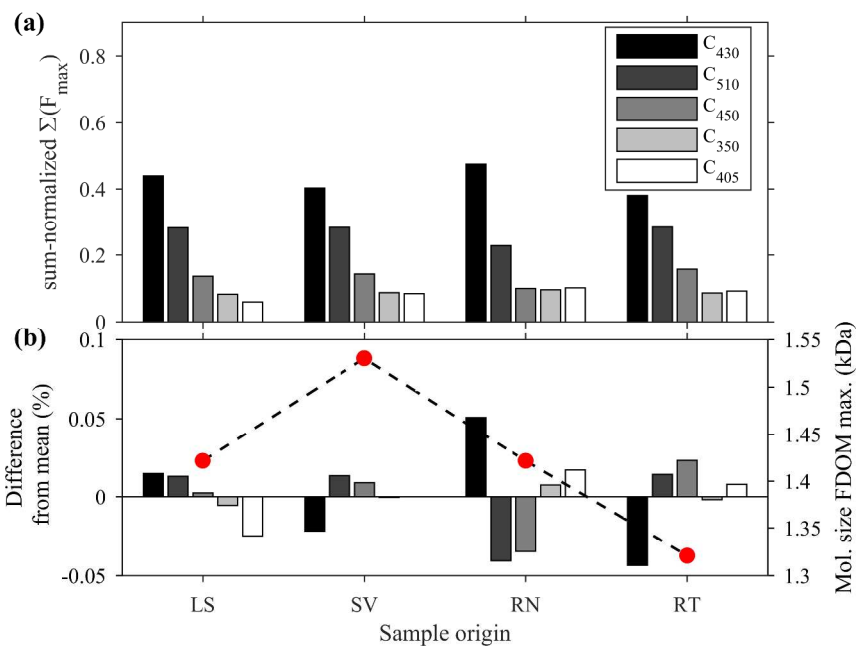
606

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

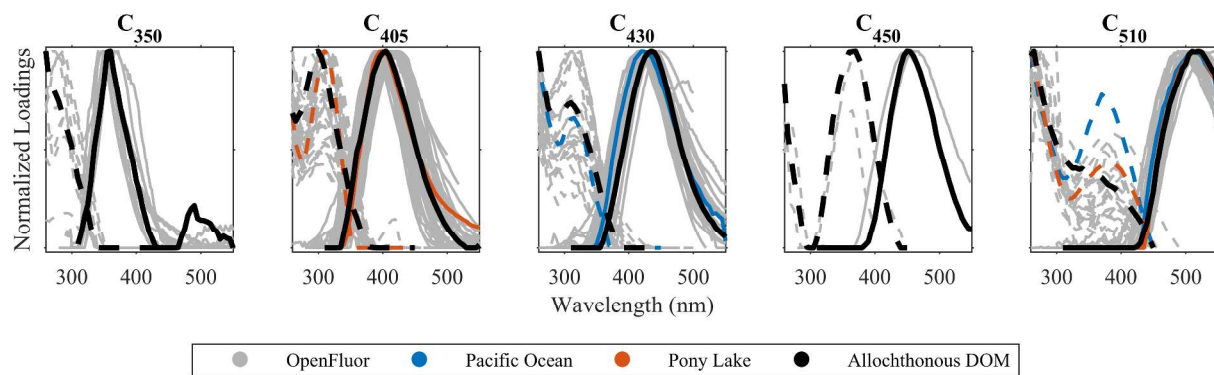
611 **Figure 2. Contour plots of five allochthonous freshwater PARAFAC-derived fluorescence spectra (sample from Lillsjön**
612 **(a)) and comparison between spectral properties of five spectra originating from four different samples and their**
613 **respective models (b). Components are ranked and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



615

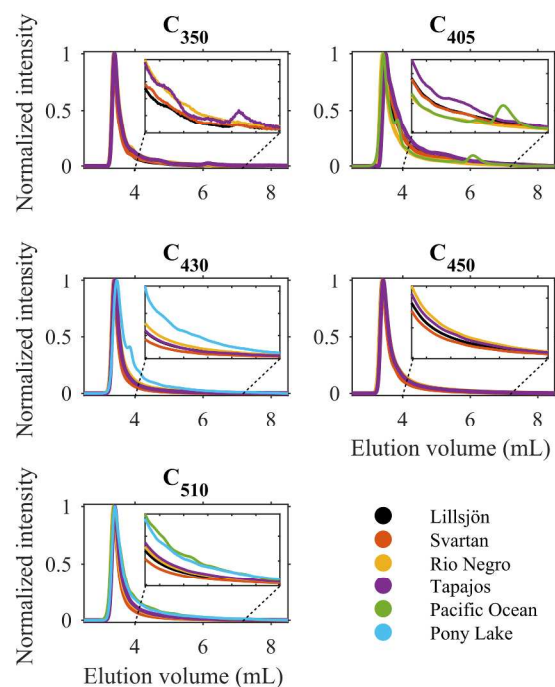
616 **Figure 3. Relative contributions of PARAFAC components to the total fluorescence in the four allochthonous samples (a),**
 617 **as well as deviation of the relative contribution of PARAFAC components from the average composition per component**
 618 **(b, left axis & bars LS = Lillsjön, SV = Svartan, 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 C₃₅₀, 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.