**Spatial distribution and transformation of dissolved organic matter along the aquatic continuum: from lakes to oceans**

Philippe Massicotte‡\*, Eero Asmala\*, Colin Stedmon†, Stiig Markager\*

\*Aarhus university, Department of Biosciences, DK-4000 Roskilde, Denmark

†National Institute for Aquatic Resources, Technical University of Denmark, Charlottenlund, Denmark

**‡Corresponding author**

Philippe Massicotte

Aarhus university

Department of Biosciences

DK-4000 Roskilde, Denmark

Phone: +45 871 58 472 #59003

E-mail: pm@bios.au.dk

**Running head:** DOM along the aquatic continuum

**Keywords:** Dissolved organic matter (DOM), Absorption, Chromophoric dissolved organic matter (CDOM), Dissolved organic carbon (DOC), Specific UV absorbance (SUVA), Aquatic continuum, Rivers, Lakes, Oceans, ...

**Abstract**

We performed an extensive literature survey to highlight the global patterns in DOM along the aquatic continuum from lakes to oceans.

**Introduction**

Physico-chemical characteristic of dissolved organic matter (DOM) are important parameters that drive the functioning of aquatic ecosystems at different levels. For example, dissolved organic carbon (DOC) is the major fraction of the DOM pool and represents the main source of metabolic substrates for heterotrophic bacteria and influences the composition of aquatic microbial communities (Findlay2003). Additionally, the chromophoric fraction of the DOM pool (CDOM) is a major driver of underwater light characteristics (Kirk1994) which modulate many bio-optical processes such as primary production (Thrane2014, Seekell2015) and also constitute a natural screen protecting aquatic organisms against harmful ultraviolet (UV) radiations (**ref**).

In recent decades, climate changes, eutrophication and rapid transformations in land use have contributed to increase inputs of colored terrestrial DOM in aquatic ecosystems (Roulet2006, Massicotte2013RSE, Weyhenmeyer2014, Haaland2010). This has important consequences since the transformation of even a small fraction of the DOM pool can potentially have large impacts on ecosystem functioning (Prairie2008). Increases in CO2 emissions (Lapierre2013) and reduction in primary production due to light shading (Seekell2015, Thrane2014) have been already identified as consequences of this generalized increase in terrestrial DOM at local and regional scales. However, generalizing these effects from local to global scales is a difficult task because our current understanding on the fate and dynamics of DOM along the aquatic continuum gradient (from headwater lakes to oceans) is limited. Since most studies about the fate of DOM are either ecosystem or site specific, there is an indisputable need for integrative studies that will unify existing knowledge to better understand the fate and the dynamic of DOM from a broader perspective during its transport from headwaters lakes to oceans.

DOC and CDOM properties are now routinely measured in most ecological studies. This creates an opportunity to explore the factors regulating the spatial distribution, the fate and the dynamics of the DOM pool during its transition along the aquatic continuum. We performed an extensive literature survey (n = **xxx**) to extract datasets containing both DOC and CDOM absorption measurements to gain insights about the spatial distribution and the compositional characteristics of the DOM in different ecosystems along the aquatic continuum. We hypothesized that a strong relationship between DOC and absorption properties of CDOM would be a common characteristic in aquatic ecosystems receiving large amount of colored DOM from their surrounding terrestrial environments. Because photochemical processes are known to remove CDOM preferentially over DOC, we further expected that the robustness of the observed relationship would waken as DOM is gradually degraded during its transition toward ocean. A second key objective of this work was to gather and harmonize available spectral information to help the community to derive algorithms for predicting both CDOM and DOC across the aquatic continuum.

**Methods**

**Literature survey and spatial coverage**

Web of science, Google Scholar as well as public data repositories were searched using terms “cdom”, “doc”, “dissolved organic carbon”, “absorption” for datasets presenting raw (i.e. not summarized) values of DOC and optical properties of CDOM. The minimum variables required to be included in the dataset were DOC, absorbance or absorption of CDOM, geographical coordinates and time of the sampling. When not explicitly provided, geographical coordinates were estimated by hand using available sampling map in each study. For CDOM data, wavelengths and cuvette size used for the optical measurements were also required. Using these criteria, we compiled 43 datasets containing 11920 unique observations of simultaneous DOC and absorption properties of DOM measured between **1991** and **2014** (Fig. 1, Table 1). A total of 4308 observations with complete CDOM spectra (ie continuous measurements along a range of wavelengths) and 6518 observations with CDOM absorption measurement at discrete wavelengths were extracted.

Observations were localized all around the world distributed on five different continents and in seven oceans (Fig. 1, Supplementary Fig. 1A). A large proportion of the data were located in river and ocean ecosystems and to a lesser extend in estuaries, wetlands and lakes regions (Supplementary Fig. 1B). Oceanic observations were spread all around the North Atlantic, North Pacific, South Pacific, Arctic, South Atlantic, Southern and Indian oceans. In North America, dense clusters of observations were mostly located in large rivers and estuaries of the East Coast and along the Gulf of Mexico from the Rio Grande, Texas, to Anclote Island, Florida (Fig. 1). In northern America, observations were mainly located in Alaska along the Mackenzie and the Tanana rivers. Few observations were also extracted in the St-Lawrence river, around the Great Lakes and in the Hudson Bay. In Europe, a large fraction of the samples came from Baltic sea, North sea and Kattegat sea as well as in main rivers of Sweden. Few observations were also located around Greenland. In Russia, observations were mostly located in the Lana and Kolyma rivers as well as around the Laptev and Siberia seas. In Asia, observations were located in South Korea rivers, in lake Taihu in China and on west coast of Taiwan. In Africa, most of the observations were located in Congo, Niger, Zambezi and the Ogooué rivers. In Australia, data were located in the St. Vincent gulf and on the border of Timor Sea.

**Ecosystem classification**

Each observation were assigned to a defined ecosystem using either the sampling location or the salinity when available (supplementary Fig. 1B). Observations presenting salinity values were classified as follows: river (salinity ≤ 0.5), estuary (0.5 < salinity ≤ 5), coastal (5 < salinity ≤ 30), ocean (salinity > 30). Based available information, observations were classified as follows (Supplementary Fig. 1A): river (*n* = **4527**), ocean (*n* = **3143**), estuary (*n* = **1273**), wetlands (*n* = **954**), coastal (*n* = 866), lake (*n* = **627**), brines (*n* = **58**) and sewage (*n* = **32**). Brines and sewage have categorized as ecosystems because they represent systems which are outside the typical hydrological continuum freshwater-oceans.

**Data processing and metrics calculation**

Absorbance by CDOM were converted to absorption coefficients and expressed per meter using equation 1 (Kirk1994):

where aCDOM(λ) is the absorption coefficient (m−1) at wavelength λ, A(λ) the absorbance at wavelength λ and L the path length of the optical cell in meters. Given that UV–visible absorption spectra of CDOM decrease approximately exponentially with increasing wavelength, a simple exponential model (equation 2) has been used to extract quantitative information about optical properties of CDOM (Jerlov1968, Bricaud1981, Stedmon2001).

where aCDOM is the absorption coefficient (m−1), λ is the wavelength (nm), λ0 is a reference wavelength (nm), K is a background constant (m−1) accounting for scatter in the cuvette and drift of the instrument. S is the spectral slope (nm−1) that describes the approximate exponential rate of decrease in absorption with increasing wavelengths. Specific UV absorbance (SUVA254, L × mgC-1 × m-1) was calculated by dividing absorbance at 254 nm by DOC content in mgC × L-1 (Weisharr2003). For other analysis, DOC were expressed in µm of carbon assuming that one mole of carbon weight 12 grams. Spectral slope curves based on CDOM absorption spectra were calculating as described in Loiselle2009. Briefly, spectral slopes were calculated over a sliding window of 21 nm along the complete spectral range using equation 2. Each calculated value was associated to the middle wavelength of the current sliding window.

Given the wide range of wavelengths used in each study, absorption spectra were filtered to keep measurements between 250 and 600 nm at 1 nm increment. For the Nelson et al. (2001) dataset (Table 1), absorption was only available between 275 and 600 nm, and not included in the spectral analysis. Five criterion were used to control the quality of absorption spectra: (1) SUVA254 had to be smaller or equal to 6, (2) the spectral slope (S, equation 2) had to be smaller than 0.08 nm-1, (3) the determination coefficient of the fit (R2, equation 2) needed to be at least 0.95, (4) the value of aCDOM(440) needed to be positive and (5) the value of aCDOM(350) needed to be > 0.01 m-1. Based on these criteria, a total of 119 absorption spectra were discarded from further analyzes.

**Estimation of aCDOM(350)**

In the extracted data, we found out that a wide range of different wavelengths (between 253 and 443 nm) were used to report absorption coefficients of CDOM (Supplementary Table 2). To make absorption coefficient comparable among studies, an interpolation procedure was used to estimate aCDOM(350) independently of the wavelength used in each study. This choice was motivated because absorption at 350 nm was among the most reported wavelength in the available data. To achieve this, we used complete absorption spectra (n = **2387**) to predict the value of aCDOM(350) from observations measured at other wavelengths (Fig. 2). This was done by regressing all the absorption values at a specific wavelength against that measured at 350 nm (ex. aCDOM(254) vs. aCDOM(350)). Then, the slope (Fig. 2B) and the intercept (Fig. 2C) of the linear regression were used to predict aCDOM(350) from aCDOM(λ). Based on reported wavelengths, a total of **17** linear models were made (Supplementary Table 2). A minimum value of 0.98 for the determination coefficient (*R2*) was used as a threshold to discard observations that were too far from the targeted wavelength of 350 nm.

**Estimation of DOM biogeochemical processing history**

We used the distance to the closest shoreline and the SUVA254 to estimate DOM biogeochemical processing history (bleaching) and reactivity for river and marine samples along the aquatic continuum from rivers to oceans. The distance to the closest coastline was calculated using ocean shapefiles (resolution = 1:110000000) openly available on the Natural Earth website (<http://www.naturalearthdata.com/>) and the **rworldmap** R package (South2011). For inland samples (rivers) the measured distance have been assigned to positive values whereas for marine samples, the measured distance were assigned to negative values. Lakes were not included in this analysis, since their connectivity to larger-scale aquatic continuum is less obvious than that of rivers. Because precise geographical coordinates were not always available and thus often estimated by hand using available maps, calculated distances have been pooled using 150 km bins which was found to roughly distribute the observations equally in each bin.

**Statistical analysis**

Segmentation analysis were performed using the **segmented** R package (Muggeo2003, Muggeo2008). CDOM metrics were calculated using the **cdom** R package (Massicotte2016MC). Geographical analyses were done using the **rgeos** R package (Bivand2016). All statistical analysis were performed in R 3.3.1 (RCoreTeam2016).

**Results**

**Estimation of aCDOM(350)**

Results of the linear regressions used to predict aCDOM(350) from other wavelengths are shown in Fig. 2. Between 250 and 350 nm, determination coefficients (R2) gradually increased from 0.987 to 1 (Fig. 2A). After 350 nm, R2 decreased rapidly to reach 0.86 at 500 nm. Between 250 and 500 nm, the slope values increased almost exponentially (0.28-6.99, Fig. 2B) whereas the intercepts increased linearly between -1.4 and 1.5 m-1 (Fig. 2C). Perfect fit at 350 nm is identified using vertical dashed lines where R2 = 1, slope = 1 and intercept = 0. Note that the 95% confidence interval of the estimated slope values is hardly distinguishable which emphasis the robustness of the generated models (shaded areas in Fig. 2B). Regression coefficients used to estimate aCDOM(350) in this study are presented in Supplementary Table 2. Absorption coefficients measured higher than **412** nm were discarded because of the R2 bellow the selected threshold of 0.98. A heat map plot showing the R2 of the regressions between all possible pairs of wavelengths between 250 nm and 500 nm (1 nm increment) is presented in supplementary Fig. 2 (*n* = 63001). Corresponding coefficients are provided as a supplementary comma-separated values (CSV) file for enabling calculation of a given wavelength from another in the range of 250-500 nm.

**Estimation of DOC using absorption coefficients**

Estimation of DOC concentration using CDOM absorption measurements is a commonly used technique in many ecological and remote sensing based studies. However, the relationship between these two key parameters is rarely evaluated on global scale. The robustness of the DOC predictions from absorption measured at different wavelengths are presented in Fig. 3. For this exercise, ecosystems were reclassified in general end-members as follow: freshwater (lakes, rivers, sewage, wetlands), coastal (coasts, estuaries) and ocean. For all these end member ecosystems, prediction of DOC as a function of aCDOM(λ) was found to decrease monotonically with increasing wavelengths but with varying magnitude (Fig. 3). Due to the high humic content of the DOM pool in freshwater and coastal ecosystems, the goodness of the predictions remained relatively high along the complete spectral range. In freshwater, the robustness of the relationship between aCDOM(λ) and DOC remained relatively high and stable between 250 and 400 nm with and averaged 0.98 before decreasing to 0.68 at 500 nm. Prediction of DOC was also relatively high for coastal samples where R2 varied between 0.82 and 0.64. For ocean samples, the prediction of DOC from absorption measurements was much lower and R2 decreased rapidly from 0.63 at 250 nm to 0.023 at 500 nm, possibly due to the colorless nature of the organic molecule characterizing marine DOM (**refs**).

**Global relationship between DOC and aCDOM(350)**

A strong positive log-linear relationship was found between aCDOM(350) and DOC (Fig. 4A, n = 11562, R2 = 0.93, p < 0.0001). At low value of DOC (35 µmol), predicted value of aCDOM(350) was 0.03 m-1. As DOC increased to a maximum of **44600** µmol in wetlands, predicted aCDOM(350) reached **1097** m-1 (Fig. 4A). The derived equation from the general log-linear model indicated that aCDOM(350) increases by xxx m-1 for each unit of increase in DOC.

EQUATION FOR PREDICTING DOC

The robustness of the global relationship was found to vary greatly among the different ecosystems and R2 averaged 0.68 (Fig. 4B, supplementary Fig. 3). The individual relationships between DOC and aCDOM(350) for observations in ocean, coastal, lake and sewage was found to be weaker than the calculated average which caused larger scattering around the regression line at low DOC values (Fig. 4A, **Supplementary Fig requested by Eero**). The weakest relationship between DOC and aCDOM(350) was found in the ocean ecosystem (R2 = 0.44) whereas the strongest one was found in the wetland ecosystem (R2 = 0.94) which presented similarities with river and estuary ecosystems.

**Distribution of aCDOM(350), DOC and SUVA350 along the aquatic continuum**

The distributions of the principal variables used to characterize the DOM pool along the freshwater-marine continuum are presented in Fig. 5. DOC concentrations ranged from 19.2 to 44600 µm C *×* L-1, with a mean value of 769.3 ± 1854.8 µm C *×* L-1. Absorption coefficients at 350 nm varied by three orders of magnitude between wetland and ocean ecosystems (Fig. 5A). In wetlands, median aCDOM(350) was 87.2 m-1 and decreased linearly along the freshwater-marine continuum to reach 0.08 m-1 in ocean ecosystems. DOC concentration showed a similar negative trend along the aquatic continuum with median value decreasing from 3250 µm C *×* L-1 in the wetlands to reach 50.20 µm C *×* L-1 in oceans (Fig. 5B). Median value of SUVA350 was found to vary between 1.34 L × mgC-1 × m-1 m2 g-1 C in the wetlands and 0.09 L × m2 g-1 C mgC-1 × m-1 in oceans (Fig. 5C). Whereas aCDOM(350) and DOC values both negatively decreased along the freshwater-marine gradient, SUVA350 showed similar values among inland water ecosystems (wetland, pond, lake, river and sewage) with a median value of 0.86 L × m2 g-1 CmgC-1 × m-1. For marine-like ecosystems (coastal, brine, estuary and ocean) median SUVA350 was 0.15 L × m2 g-1 C (Fig. 3C).

**DOM reactivity along the aquatic continuum**

SUVA254 was used as a proxy for characterizing DOM chemical composition and reactivity (Weishaar2003) over 4000 km along the aquatic continuum (Fig. 6). A piecewise regression was found to adequately model the pattern observed in the data (R2 = 0.95, p < 0.0001). A significant breakpoint was found at 370 ± 103 km towards the ocean from the coastline at the interface between freshwater and marine ecosystems. Between 1500 and -360 km (mostly inland waters), mean SUVA254 decreased rapidly from 4.79 to 1.68 L × m2 g-1 C indicating an important loss in DOM aromaticity. Beyond the identified breakpoint onward to the open ocean, no visible trend was observed where SUVA254 remained stable at an average of **XXX**.

**Conservative mixing of DOM**

The linear trend of the specific UV absorbance at 254 nm (SUVA254) along the salinity gradient was modeled using a piecewise regression where two different breakpoints at salinity 8.7 ± 0.3 and 26.8 ± 0.8 were found (Fig. 7, R2 = 0.74, p < 0.0001). Between salinity 0 and 8.7, the slope of the linear regression was -0.3 indicating that SUVA254 decreased by this amount for each unit increase in salinity. Between salinity 8.7 and 26.8, SUVA254 remained stable and the slope of the regression was not significantly different from 0 (**p = xxx**). Another significant slope with a value of -0.09 was found after salinity of 26.8. **More details?**

**Spectral differences between freshwater and marine ecosystems**

Spectral slope (Sλ) curves calculated on normalized and averaged CDOM spectra in both end-member ecosystems showed contrasting patterns (Fig. 8). Sλ calculated from the freshwater samples looked like an inverted parabola. Spectral slopes showed a linear increase in the UV-C region (< 280 nm) before reaching a plateau in the UV-B and UV-A regions (280-400 nm). Above 400 nm, Sλ decreased rapidly to 0.0095 nm-1. For the marine end-member, a dominant peak in the spectral slope curve was found at approximately 280 nm. Contrary to the freshwater curve, the quality of the computed slope, estimated using R2, was found to decrease from 400 nm indicating that the exponential model used to calculate Sλ were not fully capturing the general pattern in the DOC-aCDOM(λ) relationship (Fig. 8).

**Discussion**

Based on an extensive literature survey, our results shows that the spatial distribution of DOM is highly heterogeneous along the aquatic continuum. **TODO**

**Relationships between DOC and aCDOM along the aquatic continuum**

DOC and aCDOM measurements are often used as proxies to characterize the quantity and the quality of the DOM pool in aquatic ecosystems. In a context of rapid climate change at global scale, there is an increasing interest to predict and estimate DOC and aCDOM with remote sensing techniques (Kutser2005, Olmanson2016, Slonecker2016). However, absorption coefficients of CDOM are often reported at different wavelength (Supplementary Table 1) which make cross-studies comparisons difficult given the non-linear pattern of absorption spectra (equation 2). Using an extensive dataset that includes **4802** complete profiles of absorption measurements (250-500 nm) from a wide range of ecosystems, we found out that the ability to predict DOC from aCDOM(λ) decreased with increasing wavelengths (Fig. 3). In this study, we found that aCDOM(350) correlated strongly to the DOC content of the DOM pool at global scale (Fig. 4, supplementary Fig. 3). However, the relationships between DOC and aCDOM(350) became gradually decoupled as the DOM pool was transported toward the oceans. One possible explanation is that photochemical processes are more efficient at removing CDOM over DOC in headwater streams compared in open oceans (Moran2000, Bittar2015, Hansen2016). Indeed, in wetlands, rivers and estuaries a large fraction of the DOM pool is controlled by terrestrial inputs originating from the tributaries, soil erosion and surface runoff (Massicotte2011, Lambert2015). This humic-like DOM is continually degraded by UV radiations due to its larger molecules and its higher aromaticity (Moran1997). Once entering the open oceans, a large proportion of the DOM pool has likely integrated a long history of photobleaching, causing aCDOM to be removed at a higher rate than that of DOC. As pointed out by Kowalczuk2010, this suggest that processes responsible for production, decomposition and distribution of the bulk DOC and CDOM components are decoupled in oceanic systems.

* Talk about Fig. 5
  + Large variability but clear pattern.
  + Link with sup fig asked by Eero with Residual from the “big” model.
  + An interesting observation is that aCDOM(350) decreased by more than 3 order magnitude along the aquatic continuum (Fig. 5A) where it was “only” by 2 order for DOC (Fig. 5B).
  + Gives another evidence that aCDOM is being removed faster than DOC.

**DOM reactivity during transition from headwaters to oceans**

During its transport from headwaters to oceans, microbial respiration and production (del Giorgio et al. 1997; Kritzberg et al. 2006b; Berggren et al. 2010); sedimentation and flocculation (Sholkovitz 1976; von Wachenfeldt and Tranvik 2008), production by photosynthetic organisms (Descy et al. 2002; Kritzberg et al. 2005; Lapierre and Frenette 2009); and UV photodegradation (Benner and Ziegler 1999; Amado et al. 2006; Zhang et al. 2009) operate simultaneously and drive the fate of DOM transiting in the various habitats of aquatic ecosystems (Massicotte2013LOFE). Based on the size-reactivity continuum model proposed by Amon1996, Benner2015, it was expected that these degradation processes would actively act to decrease the molecular weight of organic matter and subsequently its reactivity along the aquatic continuum. In agreement with this conceptual model, we found out that in headwater streams, up to 1500 km toward inland continents, the DOM pool was highly aromatic (Fig. 6), presumably due to elevated lateral connectivity with surrounding terrestrial landscape and organic matter inputs from the tributaries (Massicotte2011EA, Lambert2016). Lapierre2013 found a positive relationship between aCDOM(440) and the concentrations of biologically and photochemically degradable DOC in boreal aquatic ecosystems which suggest that the highest rates of photochemical and biological processing are occurring when the DOM pool is dominated by macro molecules originating from the landscape. Hence, as this terrestrial DOM is transiting in the various aquatic ecosystems, high molecular weight molecules are degraded into smaller and more refractory molecules. Interestingly, we found that the observed decrease in reactivity was also occurring in a buffer zone of 370 km around the continents once DOM entered marine water (Fig. 6, supplementary Fig. 4). After this breakpoint, it is likely that freshly produced material from primary producer become the dominant fraction of the DOM pool which is known to be characterized by low molecular weight and colorless molecules (**refs**).

**Dynamic of DOM in oceans: more than just conservative mixing**

A rich literature uses salinity as a tracer to estimate the conservative behavior of optical properties of CDOM (Kowalczuk2010, Asmala2016). Contrary to most studies (but see Goncalves2015), our results show that CDOM dynamics deviates from the expected conservative behavior (**Fig. XXX**). Rather than completely discarding the effect of water mixing, the fitted piecewise regression suggests that there are active removal processes that are taking place once DOM enters marine water. For instance, SUVA254 decreased more rapidly than the hypothetical conservative mixing line (dashed line in Fig. 7). In agreement with finding from Goncalves2015, our results evidence that there are two distinct phases of processing at low and high salinity.

Although still debated (Markager2004, DelCastillo2000), it was pointed out that flocculation could be an important DOM removal process (Sholkovitz1976), especially at low salinity (Asmala2014a). Between salinity 0 and 8.7 we observed a rapid decrease in SUVA254 that reinforce the idea that flocculation is an efficient process responsible for the removal of high molecular weight humic substance. However, it is unlikely that flocculation is the sole process involved in the observed decrease of CDOM. Because that residence time can be quite high along marine coastlines (**refs**), we can not discard that photodegradation (Helms2013) and bacterial degradation (**ref**) are also actively contributing to degrade CDOM. A more plausible scenario is that the observed kinetic of CDOM upon intrusion in marine ecosystems is a result of the combined effect of flocculation, photodegradation and biodegradation processes (see conceptual plots in Fig. 7). Interestingly, no change in DOM aromaticity was observable between salinity 8.7 and 26.8. In this segment, we hypothesize that previous exposure to photochemical and biological processes decreased the reactivity of CDOM to a point where that degradation processes are less effective are removing aromatic compounds from the DOM pool (Amon1995, Benner2015). This could explain why production (mainly primary production) and removal processes are balancing out (**Fig. 7**). At high salinity (> 26.8), the last slope of the piecewise regression lined up with the hypothetical conservative mixing line between SUVA254 and salinity. One important characteristic of deep ocean water is that the concentration of biomolecules decreases to almost unrecognizable concentrations due to a long history of degradation that took place over centuries and millennium time scale (see refs in Dittmar2014). This is suggesting that medium-aged water containing freshly and colorless DOM is being diluted once entering in contact with deep ocean water.

**Main differences between absorption spectra from freshwater and seawater**

Loiselle2009 developed an approach to characterize the absorption characteristics of CDOM. This method, based on the derivative signal of the absorption signal, allows to determine the wavelength intervals where there are change in the spectral slope (parameter S in equation 2).

1. Based on the study of the complete absorption spectra, we found out that the most “active” wavelengths were highly different between freshwater and marine spectra (Fig. 8).
2. These ranges are the most dynamic, i.e. show the largest changes between samples
3. Present updated slope ranges (260-280 for freshwater; 260-280 and 280-340 for marine)
4. Helms2008 does not work for marine samples (not sure finally...)

**Spatial and temporal data coverage, and methodological uncertainties**

Different conclusions can be drawn based on the analysis of the spatial and temporal distribution of the data extracted from this study (supplementary Fig. 5). Despite our effort to process all available information from the literature and open repositories, a first striking evidence is that southern aquatic ecosystems (**n = 1047, 8%**) are highly under represented compared to those in north hemisphere (**n = 11889, 92%**) and no data prior to 2006 was available. Furthermore, continental Africa (n = **603**), Asia (n = **423**) and South America (n = 0) are poorly represented in the dataset, all of which contain significant inland waters and are of major significance in global carbon cycle.

1. Increase of open data over time
2. Large proportion in boreal systems → high seasonality
3. The majority of samples taken during summer, which is the productive season → might lead to bias towards autochthonous signal
   1. uncertainties arising from the conversion to a350
   2. **uncertainties in compiling large datasets from various authors**

**Future research**

The collecting of the dataset used in this study has shown that there are serious shortcomings in our current practices in making the scientific data openly available. We fully acknowledge the considerable amount of work and funds used to acquire the data, but the fact that there have been a lot of resources spent to get the data just emphasizes the rationale to make the best possible use of it. One of the first steps to make the data available (after a reasonable period of exclusive use) would be to use the existing data portals (Pangaea etc.) for uploading and storing the data. After making the data available in any of the public repositories, new cross-platform solutions could be used by the community to use and analyses the stored data. From the analysis of spectral CDOM absorption and its relationship with DOC, it is evident that the level of detail acquired from single wavelength measurements is considerably inferior to the use of spectral information. As the computational capabilities are not likely to present obstacles for utilizing spectral analyses, we strongly recommend researchers to use and develop methods that use the full potential contained in the CDOM spectra such as the spectral slope curve (Loiselle2009) and the Gaussian decomposition (Massicotte2016).