1. Global patterns in dissolved organic matter along the aquatic continuum: from lakes to oceans
2. Global relationships between dissolved organic carbon and CDOM spectral absorption along the aquatic continuum: from headwaters to oceans (we only have 1 graph about DOC/CDOM in the paper, I think the paper is more about the spatial distribution and fate of DOM)
3. Spatial distribution and transformation of dissolved organic matter along the aquatic continuum: from lakes to oceans

**Authors:** Philippe Massicotte, Eero Asmala, Colin Stedmon, Stiig Markager

**Keywords:** Dissolved organic matter (DOM), Chromophoric dissolved organic matter (CDOM), Dissolved organic carbon (DOC), Absorption, SUVA...

**Abstract**

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

**Introduction**

Dissolved organic matter (DOM) plays a fundamental role in aquatic ecosystems. Physico-chemical characteristics of the DOM in aquatic ecosystems are important parameters that drive its functioning in different processes. 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) radiation (Boily2012).

In recent decades, climate change, eutrophication and rapid transformations in land use have contributed to increased inputs of colored terrestrial DOM in aquatic ecosystems (Roulet2006, Massicotte2013RSE, Weyhenmeyer2014, Haaland2010). This has important consequences since transformations 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 the effects of this *brownification* 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 streams 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 currently routinely measured in most ecological studies. This creates an opportunity to explore the factors regulating the spatial distribution, the transformation and the dynamics of the DOM pool during its transition along the complete aquatic continuum from lakes and oceans. We performed an extensive literature survey to extract datasets containing both DOC and CDOM absorption measurements at global scale to gain insights about the compositional characteristics of the DOM in different ecosystems along the aquatic continuum.

We hypothesized that the CDOM/DOC relationship would be stronger in freshwater end-members (lakes and rivers), and gradually weaken as the DOM pool is processed and transformed during its transition to the oceans. We had X objectives:

1. Gather and harmonize available global DOC and CDOM data for further analysis.
2. Investigate the relationship between DOC and CDOM, in connection with potential factors affecting the DOM transformation, such as ecosystem type and salinity.
3. Use DOC-specific absorption to study the transformation of the DOM pool along the aquatic continuum.
4. Utilize spectral CDOM information to derive algorithms for predicting both CDOM and DOC.

**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 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). In North America, dense clusters of observations were mostly located along the coastline and in large river estuaries of the East Coast and along the Gulf of Mexico from the Rio Grande, Texas, to Anclote Island, Florida (Fig. 1B).

1. Mackenzie River Delta
2. Alaska (around Tanana river)
3. Few observations were also extracted in the St-Lawrence River and in Hudson Bay.

In Africa, most of the observations were located in Congo, Niger, Zambezi and the Ogooué rivers are among the largest rivers of Africa, both in terms of drainage area and discharge (Lambert2015).

1. North EU: Kattegat, Baltic Sea, Greenland
2. Lena Delta (Russia)
3. Kolyma river
4. Aisia: South Korea, Lake Taihu, West coast of Taiwan

**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), pond (*n* = 84), 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):

EQ1 HERE

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).

EQ2 HERE

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. For the reminder of the text, S300-600 denotes the spectral slope calculated between 300 and 600 nm, S275-295 the spectral slope calculated between 275 and 295 nm and S350-400 the spectral slope calculated between 350 and 400 nm. The slope ratio (SR) is the ratio between S275-295 and S350-400 which is a proxy for DOM molecular weight (Helms2008). 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 µmol C assuming that one mol 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 measurement 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. Four 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 1) needed to be at least 0.95 and (4) the value of aCDOM(440) needed to be positive. 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 and the intercept 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 distances whereas for marine samples, the measured distance were assigned to negative distances. Lakes were not included to 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 (n range here?) 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 (vertical dashed lines) shows values of R2, slope and intercept a 1, 1 and 0. Note that the 95% confidence interval of the estimated slope values is hardly distinguishable (shaded areas in Fig. 2B). Regression coefficients used to interpolate aCDOM(λ) in this study are presented in Supplementary Table 2. Absorption coefficients measured at 420, 440 nm and 443 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 is presented in supplementary Fig. 2 (*n* = 63001) and 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..

**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. 3. 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. 3A). 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. 3B). Median value of SUVA350 was found to vary between 1.34 L × m2 g-1 C in the wetlands and 0.09 L × m2 g-1 C in oceans (Fig. 3C). Whereas aCDOM(350) and DOC values both negatively decreased along the freshwater-marine gradient, SUVA350 showed similar values among terrestrial-like ecosystems (wetland, pond, lake, river and sewage) with a median value of 0.86 L × m2 g-1 C. For marine-like ecosystems (coastal, brine, estuary and ocean) median SUVA350 was 0.15 L × m2 g-1 C (Fig. 3C).

**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 log-linear model indicated that aCDOM(350) increases by xxx m-1 for each unit of increase in 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). The weakest relationship between DOC and aCDOM(350) was found in the ocean ecosystem (R2 = 0.31) whereas the strongest one was found in the wetland ecosystem (R2 = 0.94) which presented similarities with pond and estuary ecosystems.

**Estimation of DOC using absorption coefficients**

Estimation of DOC concentration using CDOM absorption measurements is a commonly used technique in many ecological studies but to the date has not beensystematically evaluated on global scale. The robustness of the DOC predictions from absorption measured at different wavelengths are presented in Fig. 5. For this exercise, ecosystems were reclassified in general end-members as follow: freshwater (lakes, rivers, sewage, ponds, 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 with increasing wavelengths but with varying magnitude (Fig. 5). 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 marine 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.

**DOM reactivity along the aquatic continuum**

SUVA254 was used as a proxy for characterizing DOM chemical composition and reactivity over 4000 km along the aquatic continuum (Fig. 6). A piecewise regression was found to adequately model the *hockey stick* pattern observed in the data (R2 = 0.95, p < 0.0001). A significant breakpoint was found at 360 ± 101 (SE) 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 a loss in DOM aromaticity. Beyond the identified breakpoint onward to the open ocean, no visible trend was observed where SUVA254 remained stable.

**SUVA vs salinity**

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 and 26.8 were found (Fig. 5B, 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. Another significant slope with a value of -0.09 was found after salinity 26.8.

**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. 8B). Sλ calculated from the freshwater samples 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. 8B).

**Discussion**

Tentative storyline for discussion:

1. **Spatial and temporal data coverage, and methodological uncertainties**
   1. figures 1 & 2
   2. potential bias from 1) place and 2) time of observations
      1. high number of boreal systems  high seasonality
      2. majority of samples taken during summer, which is the productive season  might lead to bias towards autochthonous signal
   3. uncertainties arising from the conversion to a350
   4. uncertainties in compiling large datasets from various authors
2. **DOC-CDOM relationship**
   1. figures 4 & 5 (to be changed to 3 & 4)
   2. strong global trend with decreasing R2 with increasing salinity
   3. wavelength-dependent relationship, with decreasing R2 with increasing wavelength
   4. (RSD/SE figure?)
   5. These are the key points of the study, so they need to be discussed thoroughly
3. **DOM changes along transition from headwaters to ocean**
   1. figures 3, 6 & 7 (to be changed to 5, 6 & 7)
   2. decreasing R2 we see earlier, is a result of decoupling of CDOM and DOC
   3. processing vs. mixing
   4. what processes lead to 1) bleaching (lower SUVA), and 2) decoupling (lower R2)
4. **differences in spectral slope between freshwater and seawater**
   1. figure 8
   2. the most “active” wavelengths are the ones indicated by large changes in Loiselle curve
   3. these ranges are the most dynamic, i.e. show largest changes between samples
   4. present updated slope ranges (260-280 for freshwater; 260-280 and 280-340 for marine)
   5. Helms2008 does not work for marine samples

* Dissolved organic carbon (DOC, the main constituent of DOM) and CDOM are often used as proxies to characterize the quantity and the quality of the DOM pool in aquatic ecosystems. It was found that CDOM absorption coefficients correlate linearly to the DOC content of the DOM pool, although the relationship vary greatly among ecosystems (Spencer2012a). One explanation is that the dynamic of DOC and CDOM can be decoupled during its transport from lakes to oceans (Coble2007, Spencer2012a, Helms2013). For example, photochemical processes are likely to have a greater influence on the CDOM fraction as it absorb more UV light compared to the colorless fraction of the DOM pool.
* Additionally, absorption coefficients of CDOM are often reported at different wavelength (254, 350, 400, 440) which make cross-studies comparisons difficult given the non-linear pattern in aCDOM absorption (equation 1).
* Bleaching of DOM from lakes → open oceans

**Link with salinity and breakpoint**

Lapierre2013 found a positive relationship between aCDOM(440) and the concentrations of biologically and photochemically degradable DOC in boreal aquatic ecosystems. Based on their results we can conclude the highest rates of bio and physical processing are occurring when DOC is highly colored. Here we found that estuaries-rivers-wetlands were dominated with highly colored DOC suggesting that these ecosystems are strongly contributing to highly process DOC before it reaches the open ocean.

**Outlook and future research**

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.

Collecting 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 analyse the stored data.

The collected data is biased towards northern hemisphere, as the vast majority of the observations (83%; 9841/11878) are coming from the northern hemisphere (Supp. Fig. 1). 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.

Highly aromatic groups of DOM, as indicated by high a254 and SUVA254, can be photoreactive and degrade quickly when exposed to sunlight (Stubbins et al., 2008, 2010; Spencer et al., 2009; Stubbins and Dittmar, 2015). (From Wagner 2015)

**Extra**

**Principal component analysis**

A principal component analysis on complete absorption spectra were carried out using variables that presented the highest number of observations (Fig. 6). The first two principal components of the PCA explained 67.5% of the total variance in the dataset. The first principal component (45.76%) was mostly capturing variation in SUVA254, S, salinity, S275-295, DOC and SR. The second principal component (21.74%) was mostly associated to S350-400, DOC, S275-295 and SR. Brines, oceans and coastal observations clustered together and correlated positively to high spectral slope values (S, S275-295, S350-400) and SR. On the other hand, lakes, rivers and estuaries correlated positively with high SUVA254 and DOC concentration. On the first PC, DOC and SUVA254 were found to co-vary and where negatively correlated to salinity, spectral slopes and SR. On the second PC, S350-400 and SR were found to be negatively correlated (Fig. 6). Based on these results, first PC was found to describe the freshwater-marine continuum gradient whereas as the second PC was associated to DOM quality gradient.

Helm2008, both slopes seem to covariate, but not in our PCA…