

Practical guide on dissolved organic matter (DOM) optic

Philippe Massicotte

2016-09-01

Contents

1	Introduction and motivations	5
2	Measurements	7
2.1	Practical guide on DOM measurements	7
3	Absorbance	9
3.1	Introduction	9
3.2	How to write stuff	9
3.3	Spectral slopes	11
3.4	Modeling absorption spectra in R	11
3.5	Metrics	12
4	Fluorescence	15
4.1	Fluorescence of DOM: theoretical and mathematical background	16
4.2	R code and study case	19
4.3	Using R pipeline	23
4.4	Conclusion	24

Chapter 1

Introduction and motivations

Dissolved organic matter (DOM) plays a central role in the functioning of aquatic ecosystems. For example, characteristics of the DOM pool (quantity and quality) determine underwater light climate (Kirk, 1994), the composition of aquatic microbial communities (Foreman and Covert, 2003; Kritzberg et al., 2006) and the carbon cycling on local to global scales (Cole et al., 2007). Chemically, the DOM pool is complex (> 1500 compounds) and analytical methods used to characterize it are relatively complex, time-consuming and costly (Benner, 2002; Seitzinger et al., 2005; Fellman et al., 2010). This situation called for the development of rapid and cost effective characterization techniques. Because optical properties of DOM can be related to its chemical properties, optical techniques such as fluorescence spectroscopy have been developed and rapidly adopted by the community to characterize the DOM pool in aquatic ecosystems (Coble et al., 1990; Coble, 1996; McKnight et al., 2001).

DOM is a complex mixture containing thousands of different chemical compounds (ref). For this reasons, traditional chemical analysis used to characterize DOM are both expensive and time consuming.

- Optical methods have been developed.
- Why this project? Already few good books (Lakowicz, 2006)
- Need for unified way to present stuff with correct citation.
- R code examples are also provided.
- Wiki-like document that can serve as a starter for students.

Chapter 2

Measurements

2.1 Practical guide on DOM measurements

- How DOM is measured (different filter sizes, paper in L&O).
- Pump pression

Chapter 3

Absorbance

3.1 Introduction

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 (Boily2012). Because of its colored nature, CDOM is known to strongly absorb UV light. As wavelength increase, light absorbed by CDOM decrease exponentially 3.1.

```
library(cdom)
data(spectra)

spectra <- spectra %>% filter(wavelength <= 500)

p <- ggplot(spectra, aes(x = wavelength, y = spc1)) +
  geom_line() +
  xlab("Wavelength (nm.)") +
  ylab(bquote(Absorption~(nm^{-1})))

p
```

3.2 How to write stuff

People have been using different terminology:

- $a_g(\lambda)$
- $a(\lambda)$
- $A(\lambda)$
- $CDOM(\lambda)$
- Gelbstoff, yellow substance, gilvin, and humic substances) has more recently been provided the name chromophoric dissolved organic matter (CDOM).

$$a_{CDOM}(\lambda) = \frac{A(\lambda) \times 2.303}{L} \quad (3.1)$$

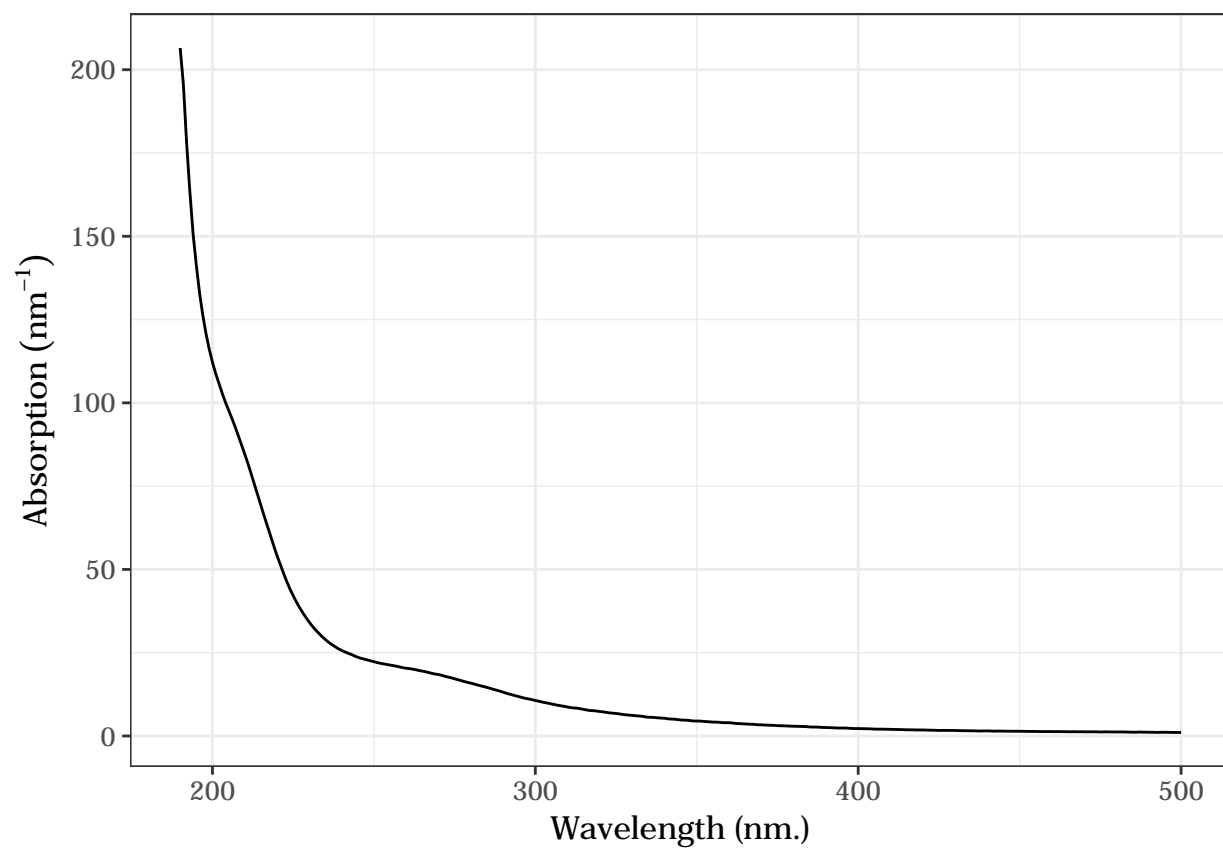


Figure 3.1: Example of an absorption spectrum of CDOM.

3.3 Spectral slopes

As observed in figure 3.1, absorption decrease exponentially with increasing wavelengths. Jerlov (1968) and Bricaud et al. (1981) first proposed to use a simple exponential formulation to model absorption (equation 3.2).

$$a_{\text{CDOM}}(\lambda) = a_{\text{CDOM}}(\lambda_0)e^{-S(\lambda-\lambda_0)} \quad (3.2)$$

Where $a_{\text{CDOM}}(\lambda)$ is the absorption coefficient (m^{-1}), λ is the wavelength (nm), λ_0 is a reference wavelength (nm) and S is the spectral slope (nm^{-1}) that describes the approximate exponential rate of decrease of absorption with increasing wavelength. Higher slopes indicate a more rapid decrease in absorption with increasing wavelength. The S parameter is frequently used as a proxy for tracing photochemical and microbial-induced changes of CDOM [Moran2000,Twardowski2004,Helms2013] or to determine its origin [Stedmon2001].

In 2001, equation 3.2 was modified by Stedmon and Markager (2001) which introduced k , a background constant (m^{-1}) accounting for scatter in the cuvette and drift of the instrument (equation 3.3).

$$a_{\text{CDOM}}(\lambda) = a_{\text{CDOM}}(\lambda_0)e^{-S(\lambda-\lambda_0)} + k \quad (3.3)$$

There are other mathematical formulations that can be used to model CDOM spectra. These are reviewed in Twardowski et al. (2004).

3.4 Modeling absorption spectra in R

Because of its inverse exponential shape, absorption spectra are best modeled using non-linear fittings. In R this is done easily using the `nls()` function. As we can see from equation 3.3, the most commonly equation used to model CDOM spectra contains three parameters: `a0`, `s`, `k`.

```
mod <- nls(spc1 ~ a0 * exp(-s * (wavelength - 250)) + k, # define the formula
          data = spectra, # where the data comes from
          start = list(a0 = 5, s = 0.02, k = 0)) # initial guesses.

# show the summary information
summary(mod)
```

```
##
## Formula: spc1 ~ a0 * exp(-s * (wavelength - 250)) + k
##
## Parameters:
##      Estimate Std. Error t value Pr(>|t|)
## a0 1.536e+01  4.778e-01  32.15  <2e-16 ***
## s  4.070e-02  5.989e-04  67.96  <2e-16 ***
## k  4.154e+00  2.871e-01  14.47  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.187 on 308 degrees of freedom
##
## Number of iterations to convergence: 7
## Achieved convergence tolerance: 3.221e-06
```

The value of a_0 is 15.36 represent the absorption coefficient at the reference wavelength of 250 nm (Fig. 3.2). The value of S is 0.04 means that absorption is decreasing at a rate of 0.04 m^{-1} for each increase of 1 nm.

- k parameter (scattering, introduced by Stedmon and Markager (2001))

```
p + geom_line(aes(y = predict(mod)), col = "red") +  
  geom_hline(yintercept = coef(mod)[1] + coef(mod)[3], lty = 2) +  
  geom_vline(xintercept = 250, lty = 2)
```

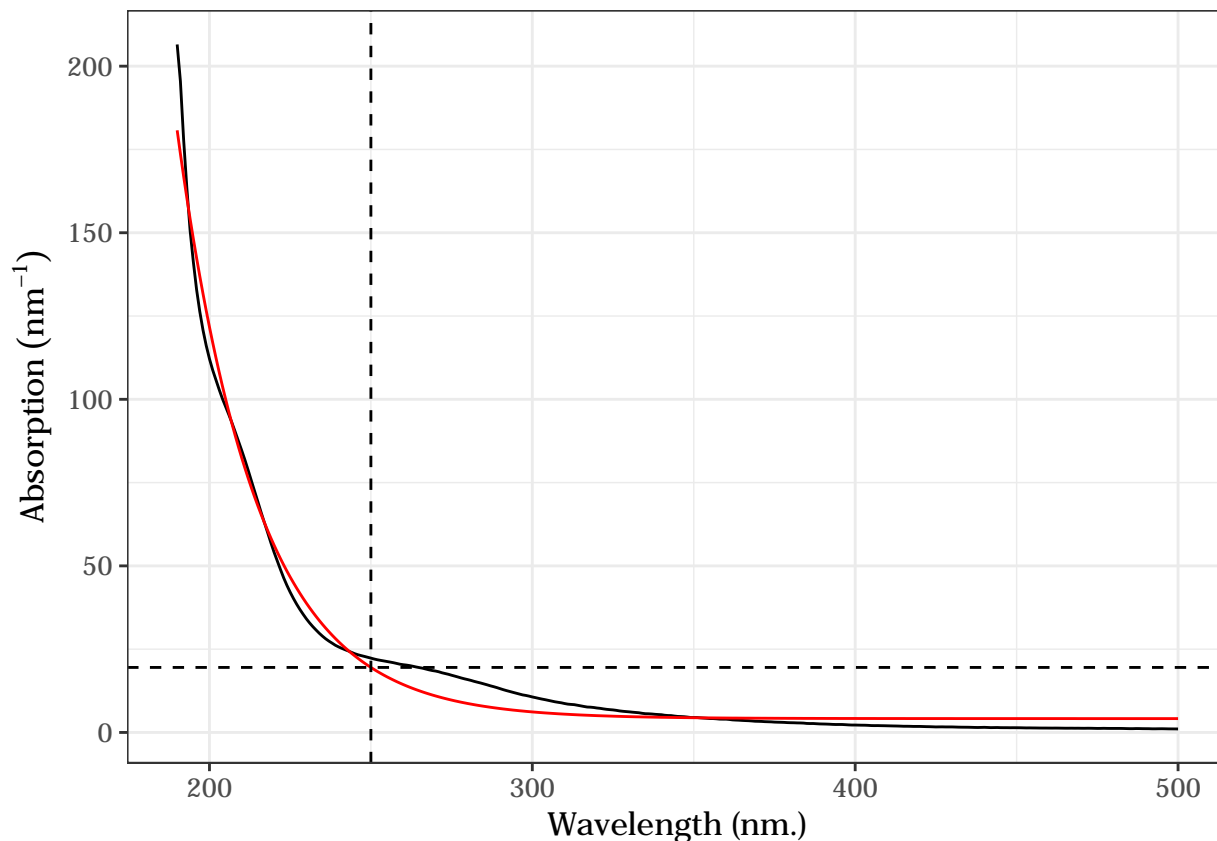


Figure 3.2: Predicted values of the exponential model.

3.5 Metrics

- $S_{300-600}$ linked to DOM molecular weight (Stedmon and Nelson, 2015).

3.5.1 Slope ratio

Equation 3.4 shows how the slope ratio (S_R) is calculated.

$$S_R = \frac{S_{275-295}}{S_{350-400}} \quad (3.4)$$

By calculating the ratio of the slope of the shorter wavelength region (275–295 nm) to that of the longer wavelength region (350–400 nm), a dimensionless parameter called “slope ratio” or S_R is defined. This approach avoids the use of spectral data near the detection limit of the

instruments used, and focuses on absorbance values that shift dramatically during estuarine transit and photochemical alteration of CDOM (Helms et al., 2008).

Figure 3.3 shows in red the 275-295 and 350-400 nm spectral range.

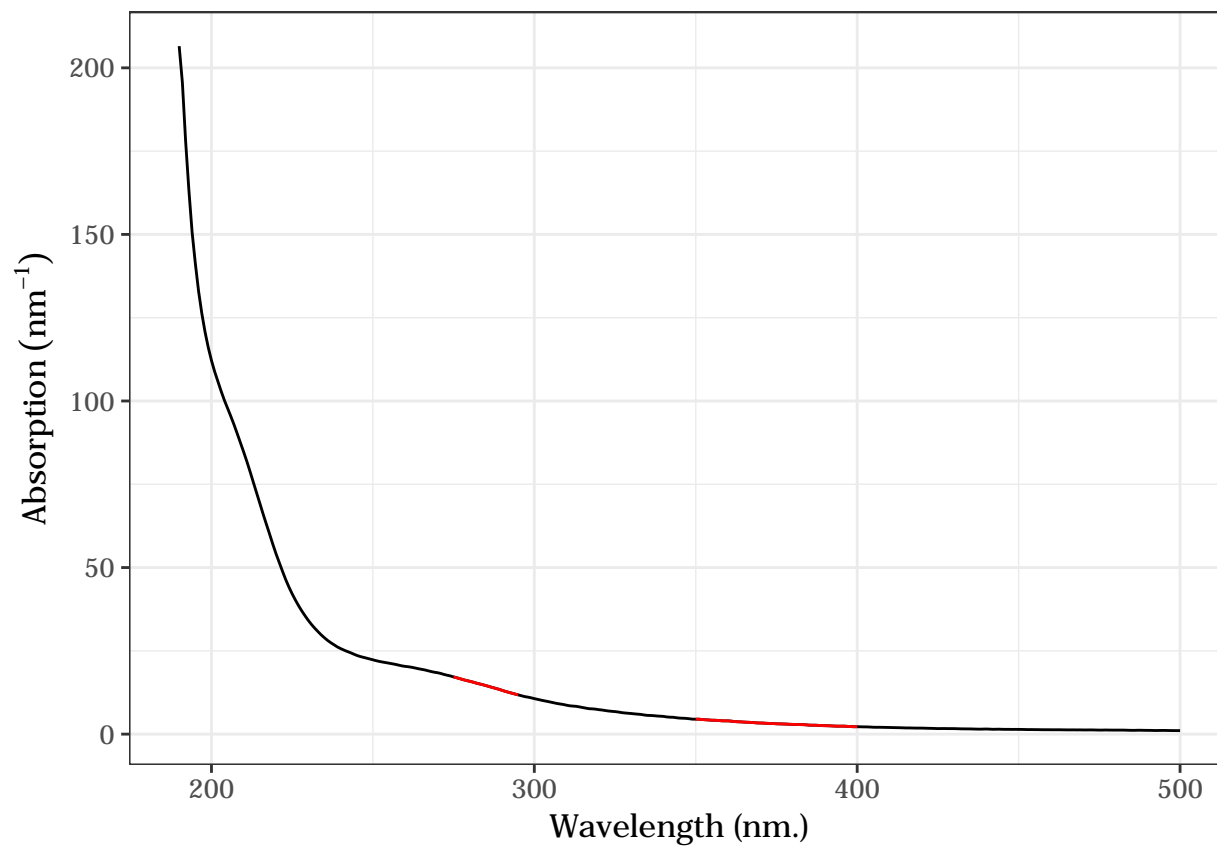


Figure 3.3: Spectral range used to calculate the slope ratio.

Chapter 4

Fluorescence

- What is a EEM
- Principal metrics (BIX, HIX, FI, etc.)
- PARAFAC

The seminal paper of (Stedmon et al., 2003) put at the forefront the use of parallel factor analysis (PARAFAC) to aid the characterization of fluorescent DOM. Briefly, this three-way technique allows the decomposition of complex DOM fluorescence signals contained in the excitation-emission matrix (EEM, Fig. 1) into a set of individual chemical components and provides estimations of their relative contribution to the total fluorescence (Bro, 1997; Fellman et al., 2010; Stedmon et al., 2003).

The PARAFAC model is described as (Bro, 1997; Harshman, 1970):

$$x_{ijk} = \sum_{f=1}^F a_{ij} b_{jf} c_{kf} + e_{ijk} \quad (4.1)$$

where $i = 1, \dots, I$; $j = 1, \dots, J$; $k = 1, \dots, K$, x_{ijk} is the intensity of fluorescence of the i^{th} sample at the j^{th} emission wavelength at the k^{th} excitation wavelength. a_{ij} is directly proportional to the concentration of the f^{th} component in the sample i . Although PARAFAC gained a lot of attention in environmental sciences, it is also widely used in other research fields such as medical, pharmaceutical, food, social and information sciences (Murphy et al., 2013). Until today, more than 1850 published scientific papers relying on PARAFAC have been identified on Web of Science.

Although PARAFAC was made easier using the **drEEM** MATLAB toolbox (Murphy et al., 2013), preprocessing of EEMs prior to the analysis is still not straightforward. EEM preprocessing is an important part of PARAFAC since it aims to correct any systematic bias in the measurements and to remove signal unrelated to DOM fluorescence (Murphy et al., 2013). Biased models can be produced if these steps are not conducted carefully (see Hiriart-Baer et al. (2008) where scattering fluorescence signals have been modeled and wrongly interpreted). Such data processing is cumbersome as it involves many steps (Stedmon and Bro, 2008; Murphy et al., 2013) which are usually executed by hand or within in-house scripting and therefore prone to introduce errors. Another important drawback limiting effective preprocessing of EEMs arise from the wide variety of file formats provided by the different manufacturers of spectrofluorometers that makes data importation difficult to generalize.

Possibly reflecting these difficulties, it was recently pointed out that characterization of DOM using fluorescence spectroscopy is still not routinely included in ecological studies (Fellman et al., 2010). Given the increasing interest for fluorescence spectroscopy in ecology, tools are needed to unify the main preprocessing steps needed for further analyzes such as PARAFAC or metric calculations. The purpose of the **eemR** R package is to provide a rapid and an elegant interface to perform preprocessing of EEMs as well as to extract common fluorescence-based metrics proposed in the literature to obtain quantitative information about the DOM pool.

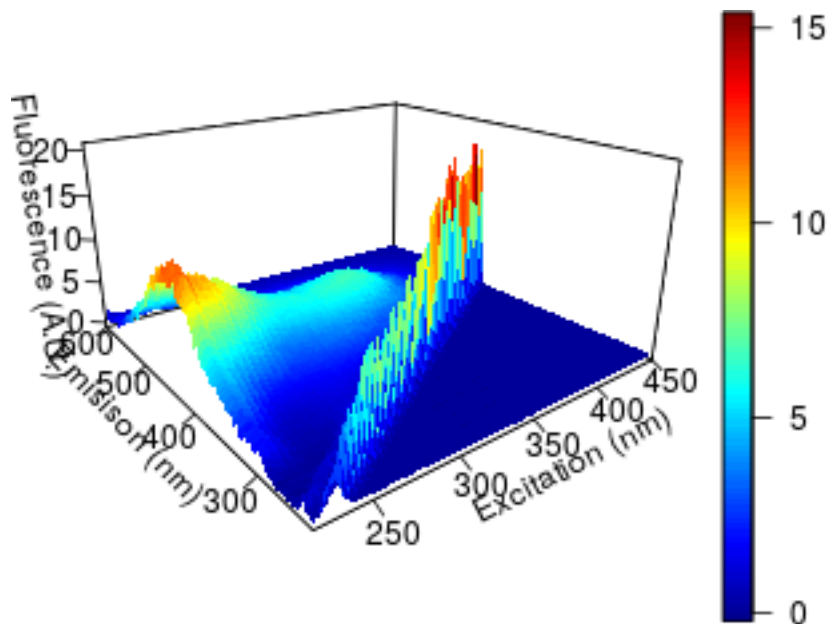


Figure 4.1: Example of an excitation-emission fluorescence matrix (EEM). The diagonal structure with high fluorescence corresponds to the first order of Rayleigh scattering.

This paper presents theoretical and mathematical background of the main PARAFAC preprocessing steps and metric calculations with concrete code examples.

4.1 Fluorescence of DOM: theoretical and mathematical background

Let us define X , an EEM of fluorescence intensities measured along a vector of excitation wavelengths (ex) at emission wavelengths (em). Usually, ex and em vary, respectively, between 200-500 nm and 220-600 nm (Fig. 1). $X_{ex,em}$ denotes the fluorescence intensity measured at excitation ex and emission em (ex.: $X_{250,400}$).

The following sections present the main correction steps for fluorescence data aiming to correct any systematic bias in the measurements and remove signal unrelated to fluorescence prior to any analysis.

Correction	Description
Blank subtraction	Subtract a pure water sample blank from the fluorescence data to help the removal of Raman and Rayleigh scattering peaks.
Scattering removal	Remove the the so-called scattering bands caused by first and second order of Raman and Rayleigh scattering.
Inner-filter effect correction	Correct for reabsorption of light occurring at both the excitation and emission wavelengths during measurement.
Raman normalization	Remove the dependency of fluorescence intensities from the measuring equipments thus allowing cross-study comparisons.

4.1.1 Scattering correction

Rayleigh and Raman scattering are optical processes by which some of the incident energy can be absorbed and converted into vibrational and rotational energy (Lakowicz, 2006). The resulting scattered energy produce the so-called scattering bands which are visually easily identifiable (Figs. 1 and 2). Given that both types of scattering are repeated across EEMs, it is important to remove such artifacts prior to analysis (Bahram

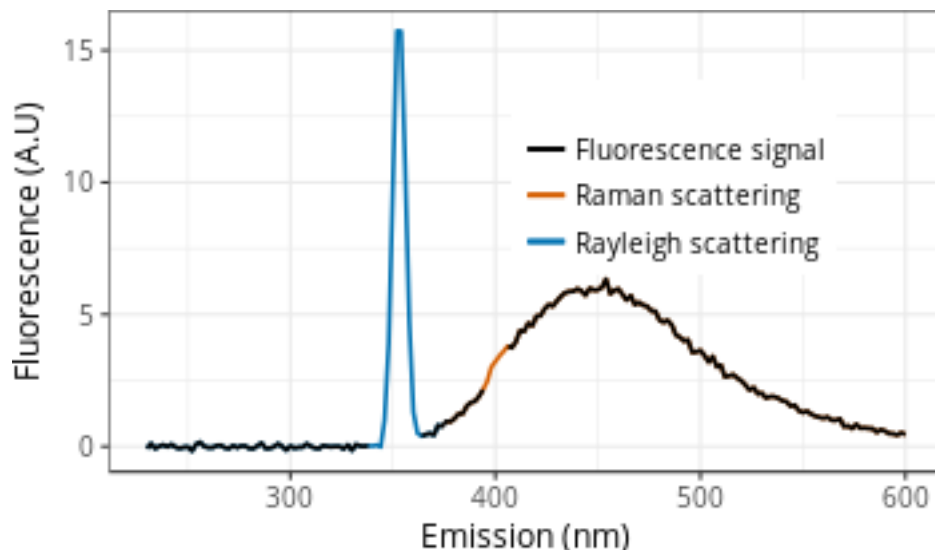


Figure 4.2: Emission fluorescence emitted at excitation $ex = 350$. First order of Rayleigh and Raman scattering regions are identified in blue and red.

et al., 2006). However, it is a common practice to simply remove the scattering-bands by inserting missing values (Fig. 3) at the corresponding positions (Murphy et al., 2013; Stedmon and Bro, 2008).

4.1.2 Inner-filter effect correction

The inner-filter effect (IFE) is an optical phenomenon of reabsorption of emitted light and occurs particularly in highly concentrated samples (Fig. 4). IFE is known to cause underestimation of fluorescence intensities especially at shorter wavelengths and even to alter the shape and the positioning of fluorescence spectra by shifting peak positions toward lower wavelengths (Fig. 4) with increasing concentration (Mobed et al., 1996; Kothawala et al., 2013). However, it was shown that the loss of fluorescence due to IFE could be estimated from absorbance spectra measured on the same sample using Equation (?) (Ohno, 2002; Parker and Barnes, 1957):

$$(@ife)X_0 = \frac{X}{10^{-b(A_{ex}+A_{em})}} \quad (4.4)$$

where X_0 is the fluorescence in the absence of IFE, X is the measured fluorescence intensity, b is half the cuvette pathlength (usually 0.5 cm) for excitation and emission absorbance, A_{ex} is the absorbance at the excitation wavelength ex and A_{em} the absorbance at the emission wavelength em (Fig. 4B).

It was recently shown that IFE corrected algebraically was not appropriate when total absorbance, defined as $A_{total} = A_{ex} + A_{em}$ (see Equation (?)), is greater than 1.5 (Kothawala et al., 2013). Under this circumstance, a two-fold dilution of the sample has been recommended. If this happens, a warning message will be displayed by the package during the correction process.

4.1.3 Raman calibration

The same DOM sample measured on different spectrofluorometers (or even the same but with different settings) can give important differences in fluorescence intensities (Lawaetz and Stedmon, 2009; Coble et al., 1993). The purpose of the Raman calibration is to remove the dependency of fluorescence intensities on the measuring equipment, thus allowing cross-study comparisons. Given that the Raman peak position of a water

sample is located at a fixed position, (Lawaetz and Stedmon, 2009) proposed to use the Raman integral of a blank-water sample measured the same day as the EEM to perform calibration. Moreover, the area of the Raman peak (A_{rp} , Fig. 5) is defined as the area of the emission profile between 371 and 428 nm at a fixed excitation of 350 nm (Lawaetz and Stedmon, 2009).

Mathematically, the value of A_{rp} is calculated using the following integral (Equation(?)):

$$(@arp)A_{rp} = \int_{\lambda_{em371}}^{\lambda_{em428}} W_{350,\lambda} d\lambda \quad (4.5)$$

where $W_{350,\lambda}$ is the fluorescence intensity of a pure water sample (preferably deionized and ultraviolet exposed, Lawaetz and Stedmon (2009)) at excitation $ex = 350$ nm and at emission $em = \lambda$ nm. Each values of the EEM X are then normalized using the scalar value of A_{rp} accordingly to Equation (?):

$$(@ramanormalisation)X_0 = \frac{X}{A_{rp}} \quad (4.6)$$

where X_0 is the normalized EEM with fluorescence intensities now expressed as Raman Units (R.U.), X are the unnormalized measured fluorescence intensities and A_{rp} is the Raman peak area.

4.1.4 Metrics

A wide range of different metrics obtained from EEMs have been proposed to characterize the DOM pool in aquatic ecosystems. These metrics extract quantitative information in specific regions (wavelengths) in EEMs. The following sections present an overview of the principal metrics supported by the package.

4.1.4.1 Coble's peaks

The following table presents the five major fluorescent components identified by (Coble, 1996) in marine EEMs. Peaks **B** and **T** represent protein-like compounds (tyrosine and tryptophane), peaks **A** and **C** are indicators of humic-like components whereas peak **M** was associated to marine humic-like fluorescence.

Peak	Ex (nm)	Em (nm)
B	275	310
T	275	340
A	260	380-460
M	312	380-420
C	350	420-480

4.1.4.2 Fluorescence, humification and biological indices

Three main indices have been proposed to trace the diagnostic state of the DOM pool in aquatic ecosystems. The fluorescence index (FI) was shown to be a good indicator of the general source and aromaticity of DOM in lakes, streams and rivers (McKnight et al., 2001). This index is calculated as the ratio of fluorescence at emission 450 nm and 500 nm, at fixed excitation of 370 nm (Equation (?)).

$$(@fi)FI = \frac{X_{370,450}}{X_{370,500}} \quad (4.7)$$

The humification index (HIX) is a measure of the complexity and the aromatic nature of DOM (Ohno, 2002). HIX calculated as the ratio of the sum of the fluorescence between 435 and 480 nm and between 300 and 345 nm at a fixed excitation of 254 nm (Equation (?)).

$$(@hix)HIX = \frac{\sum_{em=435}^{480} X_{254,em}}{\sum_{em=300}^{345} X_{254,em}} \quad (4.8)$$

The biological index (BIX) is a measure to characterize biological production of DOM (Huguet et al., 2009). BIX is calculated at excitation 310 nm, by dividing the fluorescence intensity emitted at emission 380 nm and at 430 nm (Equation (?)).

$$(@bix)BIX = \frac{X_{310,380}}{X_{310,430}} \quad (4.9)$$

4.2 R code and study case

Main preprocessing steps using the **eemR** package are illustrated using a subset of three EEMs from (Massicotte and Frenette, 2011). Briefly, these EEMs (see Fig. 1 for an example) have been sampled in the St. Lawrence River, one of the largest rivers in North America. Fluorescence matrices of DOM were measured on a Cary Eclipse spectrofluorometer (Varian, Mississauga, Ontario, Canada) over excitation wavelengths between 220 and 450 nm (5-nm increment) and emission wavelengths between 230 and 600 nm (2-nm increment). All functions from the package start with the prefix 'eem_'.

```
library(eemR)
ls("package:eemR")

## [1] "absorbance"           "eem_bind"
## [3] "eem_biological_index" "eem_coble_peaks"
## [5] "eem_cut"              "eem_export_matlab"
## [7] "eem_extract"          "eem_fluorescence_index"
## [9] "eem_humification_index" "eem_inner_filter_effect"
## [11] "eem_names"            "eem_names<-"
## [13] "eem_raman_normalisation" "eem_read"
## [15] "eem_read_parafac"      "eem_remove_blank"
## [17] "eem_remove_scattering" "eem_set_wavelengths"
```

4.2.1 Data importation and plotting

Importation of EEMs into R is done using the **eem_read()** function. Given that fluorescence files are dependent on the spectrofluorometer used, **eemR** will determine automatically from which manufacturer the files are from and load them accordingly.

```
file <- system.file("extdata/cary/scans_day_1", package = "eemR")
eems <- eem_read(file)
```

The generic **summary()** function displays useful information such as: (1) the wavelength ranges used in both emission and excitation modes, (2) the manufacturer from which the file was read and (3) the state of the EEM which indicate which corrections have been applied.

```
summary(eems)
```

```
##      sample ex_min ex_max em_min em_max is_blank_corrected
## 1      nano    220    450    230    600                FALSE
## 2 sample1    220    450    230    600                FALSE
## 3 sample2    220    450    230    600                FALSE
## 4 sample3    220    450    230    600                FALSE
##      is_scatter_corrected is_ife_corrected is_raman_normalized manufacturer
## 1                        FALSE            FALSE            FALSE Cary Eclipse
## 2                        FALSE            FALSE            FALSE Cary Eclipse
## 3                        FALSE            FALSE            FALSE Cary Eclipse
## 4                        FALSE            FALSE            FALSE Cary Eclipse
```

A surface plot of EEMs is made using the `plot(x, which = 1)` function where `which` is the index of the EEM to be plotted (see Fig. 3).

```
plot(eems, which = 3)
```

Interactive plots using a simple shiny app can be launched to interactively browse EEMs.

```
plot(eems, interactive = TRUE)
```

4.2.2 Blank subtraction

Subtraction of a water blank from the measured samples may help to reduce scattering (Murphy et al., 2013; Stedmon and Bro, 2008). In `eemR`, this is done using the `eem_remove_blank(eem, blank)` function where `eem` is a list of EEMs and `blank` is a water blank.

```
file <- system.file("extdata/cary/scans_day_1", "nano.csv", package = "eemR")
blank <- eem_read(file)

eems <- eem_remove_blank(eems, blank)
```

4.2.3 Raman and Rayleigh scattering removal

Scattering removal (Equation (?) and Equation (?)) is performed using the `eem_remove_scattering(eem, type, order, width)` function where `eem` is a list of EEMs, `type` is the scattering type (`raman` or `rayleigh`), `order` is the order of the scattering (1 or 2) and `width` the width in nanometers of the slit windows to be removed. In the following example, only first order and Raman and Rayleigh scattering are removed using a bandwidth of 10 nm (Fig. 3).

```
eems <- eem_remove_scattering(eems, "rayleigh", 1, 10) %>%
  eem_remove_scattering("raman", 1, 10)

plot(eems, which = 3)
```

4.2.4 Inner-filter effect correction

IFE correction requires the use of absorbance data (Equation (?)). For each EEM, an absorbance spectra must be supplied. The easiest way to provide absorbance is to use a data frame with column names matching EEMs names. In the following data frame, the first column represents the wavelengths at which absorbance have been measured whereas the remaining columns are absorbance spectra for `sample1`, `sample2` and `sample3`.

```
data("absorbance")
head(absorbance)
```

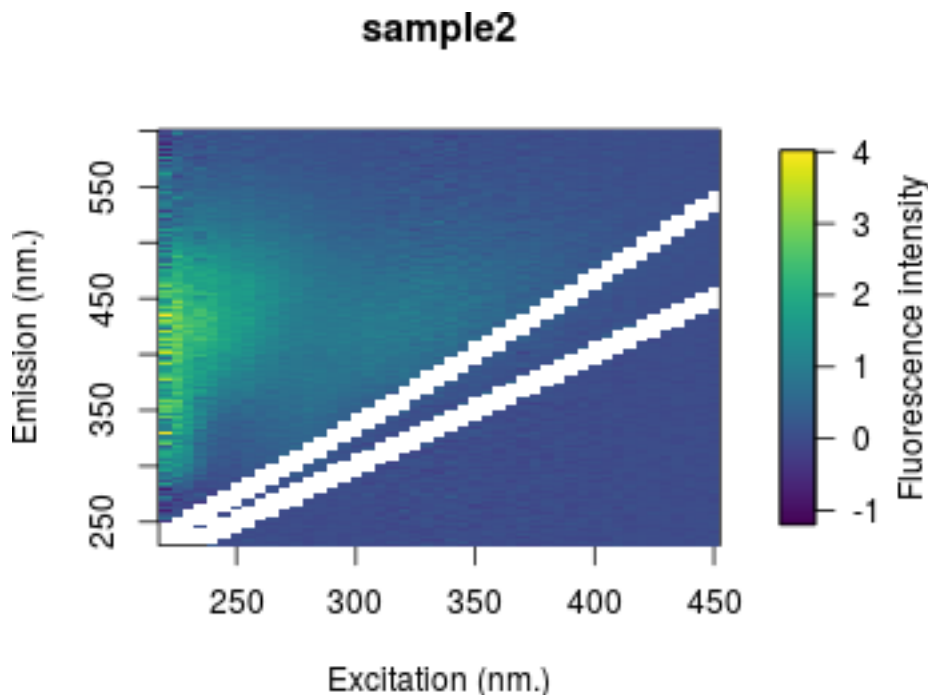


Figure 4.3: Fluorescence profile of a pure water sample at excitation 350 nm between 300 and 450 nm emission. The area of the Raman peak is identified by the shaded polygon and is calculated using Equation (@arp).

```
## wavelength sample1 sample2 sample3
## 1          190 0.89674 1.02927 1.19405
## 2          191 0.84894 0.96381 1.13721
## 3          192 0.77267 0.85339 1.04520
## 4          193 0.70967 0.75627 0.96782
## 5          194 0.65459 0.67145 0.90092
## 6          195 0.61371 0.60745 0.85054
```

Note that EEM names can be obtained using the `eem_sample_names()` function.

```
eem_names(eems)
```

```
## [1] "nano"      "sample1" "sample2" "sample3"
```

IFE correction is performed using the `eem_inner_filter_effect(eem, absorbance, pathlength)` function where `eem` is a list of EEMs, `absorbance` is a data frame containing absorbance spectra and `pathlength` is the absorbance cuvette pathlength expressed in *cm* (Fig. 4B). For each EEM contained in `eem`, the ranges spanned by the IFE correction factors and total absorbance A_{total} (Equation (?)) are displayed to the user. This can serve as diagnostic tool to determine if the mathematical correction was the appropriate method to use to handle IFE.

```
eems <- eem_inner_filter_effect(eem = eems,
                                absorbance = absorbance,
                                pathlength = 1)
```

```
## Warning: Absorbance spectrum for nano was not found. Returning uncorrected
## EEM.
```

```
## Range of IFE correction factors: 0.6433 0.989
## Range of total absorbance (Atotal) : 0.0096 0.3832
## Range of IFE correction factors: 0.7619 0.994
```

```
## Range of total absorbance (Atotal) : 0.0053 0.2362
## Range of IFE correction factors: 0.4217 0.9843
## Range of total absorbance (Atotal) : 0.0138 0.75
plot(eems, which = 3)
```

Fig. 4 presents intermediate results obtained for the correction of `sample3`. Note the nonlinearity of the correction with higher effect at lower wavelengths (bottom-left corner in panel C). The corrected EEM is presented in Fig. 4D which is the result of the operation of dividing matrix in 4A by 4C.

4.2.5 Raman normalization

The last step of the correction process consist to calibrate fluorescence intensities using the Raman scatter peak of water (Lawaetz and Stedmon, 2009). This is performed using the `eem_raman_normalisation(eem, blank)` function where `eem` is a list of EEMs and `blank` is a water blank measured the same day. Here, the same water-blank is used for the three EEMs. Note that the value of the Raman area (A_{rp} , Equation(?)) is printed.

```
eems <- eem_raman_normalisation(eems, blank)

## Raman area: 9.540904
## Raman area: 9.540904
## Raman area: 9.540904
plot(eems, which = 3)
```

At this stage, all corrections have been performed and EEMs are ready to be exported into MATLAB for PARAFAC analysis. The state of the EEMs can be verified using the `summary()` function.

```
summary(eems)
```

##	sample	ex_min	ex_max	em_min	em_max	is_blank_corrected
## 1	nano	220	450	230	600	FALSE
## 2	sample1	220	450	230	600	TRUE
## 3	sample2	220	450	230	600	TRUE
## 4	sample3	220	450	230	600	TRUE

##	is_scatter_corrected	is_ife_corrected	is_raman_normalized	manufacturer
## 1	TRUE	FALSE	FALSE	Cary Eclipse
## 2	TRUE	TRUE	TRUE	Cary Eclipse
## 3	TRUE	TRUE	TRUE	Cary Eclipse
## 4	TRUE	TRUE	TRUE	Cary Eclipse

4.2.6 Exporting to MATLAB

The `drEEM` MATLAB toolbox (Murphy et al., 2013) used to perform PARAFAC analysis requires data in a specific format (structure). The `eem_export_matlab(file, ...)` function can be used to export corrected EEMs into a PARAFAC ready format. The first `file` argument is the mat file where to export the structure and the second argument `...` is one or more `eem` object.

```
eem_export_matlab("myfile.mat", eems)
```

Once exported, one can simply import the generated mat file in MATLAB using `load('myfile.mat');`.

4.2.7 Metric extraction

Coble's peaks can be extracted using the `eem_coble_peaks(eem)` function. Note that for peaks **A**, **M**, **C**, the maximum fluorescence intensity in the range of emission region is returned.

```
file <- system.file("extdata/cary/scans_day_1", package = "eemR")
eems <- eem_read(file)
```

```
eem_coble_peaks(eems, verbose = FALSE)
```

```
##      sample      b      t      a      m      c
## 1      nano 0.8745673 0.1401188 0.140175 0.09653326 0.1255788
## 2 sample1 1.5452981 1.0603312 3.731836 2.42409567 1.8149415
## 3 sample2 1.2629968 0.6647042 1.583489 1.02359302 0.7709074
## 4 sample3 1.4740862 1.3162812 8.416034 6.06335506 6.3179129
```

Fluorescence (FI), humification (HIX) and biological (BIX) indices can be extracted as follow.

```
eem_fluorescence_index(eems, verbose = FALSE)
```

```
##      sample      fi
## 1      nano -0.5932057
## 2 sample1  1.2647823
## 3 sample2  1.4553330
## 4 sample3  1.3294132
```

```
eem_humification_index(eems, verbose = FALSE)
```

```
##      sample      hix
## 1      nano  0.5568136
## 2 sample1  6.3795618
## 3 sample2  4.2548483
## 4 sample3 13.0246234
```

```
eem_biological_index(eems, verbose = FALSE)
```

```
##      sample      bix
## 1      nano 2.6812045
## 2 sample1 0.7062640
## 3 sample2 0.8535423
## 4 sample3 0.4867927
```

It should be noted that different excitation and emission wavelengths are often used to measure EEMs. Hence, it is possible to have mismatch between measured wavelengths and wavelengths used to calculate specific metrics. In these circumstances, EEMs are interpolated using the `pracma` package (Borchers, 2015). A message warning the user will be displayed if data interpolation is performed. This behavior can be controlled using the `verbose = TRUE/FALSE` parameter.

4.3 Using R pipeline

Note that it is also possible to use the magrittr pipe line with most functions from `eemR`.

```
library(magrittr)
```

```
file <- system.file("extdata/cary/scans_day_1/", package = "eemR")
file %>%
  eem_read(recursive = TRUE) %>%
```

```
eem_raman_normalisation() %>%  
eem_remove_scattering(type = "raman", order = 1, width = 10) %>%  
eem_remove_scattering(type = "rayleigh", order = 1, width = 10) %>%  
plot(2)
```

4.4 Conclusion

eemR provides a flexible interface for manipulating and preprocessing fluorescence matrices based on theoretical and mathematical foundations of fluorescence spectroscopy (Lakowicz, 2006). Furthermore, this R package removes the drawbacks associated with EEM manipulation (dependent on spectrofluorometer manufacturer) and unifies the most important steps involved in EEM preparation in order to correct and remove systematic bias in fluorescence measurements. This will likely contribute to promote the use of fluorescence spectroscopy in various fields.

Bibliography

- Bahram, M., Bro, R., Stedmon, C., and Afkhami, A. (2006). Handling of Rayleigh and Raman scatter for PARAFAC modeling of fluorescence data using interpolation. *Journal of Chemometrics*, 20(3-4):99–105.
- Benner, R. (2002). Chemical composition and reactivity. In Hansell, D. A. and Carlson, C. A., editors, *Biogeochemistry of marine dissolved organic matter*, pages 59–90. Academic Press, San Diego, CA.
- Borchers, H. W. (2015). *pracma: Practical Numerical Math Functions*.
- Bricaud, A., Morel, A., and Prieur, L. (1981). Absorption by dissolved organic matter of the sea (yellow substance) in the UV and visible domains. *Limnology and Oceanography*, 26(1):43–53.
- Bro, R. (1997). PARAFAC. Tutorial and applications. *Chemometrics and Intelligent Laboratory Systems*, 38(2):149–171.
- Coble, P. G. (1996). Characterization of marine and terrestrial DOM in seawater using excitation-emission matrix spectroscopy. *Marine Chemistry*, 51(4):325–346.
- Coble, P. G., Green, S. A., Blough, N. V., and Gagosian, R. B. (1990). Characterization of dissolved organic matter in the Black Sea by fluorescence spectroscopy. *Nature*, 348(6300):432–435.
- Coble, P. G., Schultz, C. a., and Mopper, K. (1993). Fluorescence contouring analysis of DOC intercalibration experiment samples: a comparison of techniques. *Marine Chemistry*, 41(1-3):173–178.
- Cole, J. J., Prairie, Y. T., Caraco, N. F., McDowell, W. H., Tranvik, L. J., Striegl, R. G., Duarte, C. M., Kortelainen, P., Downing, J. A., Middelburg, J. J., and Melack, J. (2007). Plumbing the Global Carbon Cycle: Integrating Inland Waters into the Terrestrial Carbon Budget. *Ecosystems*, 10(1):172–185.
- Fellman, J. B., Hood, E., and Spencer, R. G. M. (2010). Fluorescence spectroscopy opens new windows into dissolved organic matter dynamics in freshwater ecosystems: A review. *Limnology and Oceanography*, 55(6):2452–2462.
- Foreman, C. and Covert, J. (2003). Linkages between Dissolved Organic Matter Composition and Bacterial Community Structure. In *Aquatic ecosystems : interactivity of dissolved organic matter*, chapter 14, pages 343–362. Elsevier.
- Harshman, R. a. (1970). Foundations of the PARAFAC procedure: Models and conditions for an “explanatory” multimodal factor analysis. *UCLA Working Papers in Phonetics*, 16(10):1– 84.
- Helms, J. R., Stubbins, A., Ritchie, J. D., Minor, E. C., Kieber, D. J., and Mopper, K. (2008). Absorption spectral slopes and slope ratios as indicators of molecular weight, source, and photobleaching of chromophoric dissolved organic matter. *Limnology and Oceanography*, 53(3):955–969.
- Hiriart-Baer, V. P., Diep, N., and Smith, R. E. H. (2008). Dissolved Organic Matter in the Great Lakes: Role and Nature of Allochthonous Material. *Journal of Great Lakes Research*, 34(3):383–394.
- Huguet, A., Vacher, L., Relexans, S., Saubusse, S., Froidefond, J. M., and Parlanti, E. (2009). Properties of fluorescent dissolved organic matter in the Gironde Estuary. *Organic Geochemistry*, 40(6):706–719.

- Jerlov, N. (1968). *Optical oceanography*. Elsevier Publishing Company, New York.
- Kirk, J. T. O. (1994). *Light and photosynthesis in aquatic ecosystems*. Cambridge University Press, Cambridge [England] ; New York, 2nd edition.
- Kothawala, D. N., Murphy, K. R., Stedmon, C. A., Weyhenmeyer, G. A., and Tranvik, L. J. (2013). Inner filter correction of dissolved organic matter fluorescence. *Limnology and Oceanography: Methods*, 11(12):616–630.
- Kritzberg, E. S., Cole, J. J., Pace, M. M., and Granéli, W. (2006). Bacterial Growth on Allochthonous Carbon in Humic and Nutrient-enriched Lakes: Results from Whole-Lake ^{13}C Addition Experiments. *Ecosystems*, 9(3):489–499.
- Lakowicz, J. R. (2006). *Principles of Fluorescence Spectroscopy*. Springer-Verlag, Boston, MA.
- Lawaetz, A. J. and Stedmon, C. A. (2009). Fluorescence Intensity Calibration Using the Raman Scatter Peak of Water. *Applied Spectroscopy*, 63(8):936–940.
- Massicotte, P. and Frenette, J.-J. (2011). Spatial connectivity in a large river system: resolving the sources and fate of dissolved organic matter. *Ecological Applications*, 21(7):2600–2617.
- McKnight, D. M., Boyer, E. W., Westerhoff, P. K., Doran, P. T., Kulbe, T., and Andersen, D. T. (2001). Spectrofluorometric characterization of dissolved organic matter for indication of precursor organic material and aromaticity. *Limnology and Oceanography*, 46(1):38–48.
- Mobed, J. J., Hemmingsen, S. L., Autry, J. L., and McGown, L. B. (1996). Fluorescence Characterization of IHSS Humic Substances: Total Luminescence Spectra with Absorbance Correction. *Environmental Science & Technology*, 30(10):3061–3065.
- Murphy, K. R., Stedmon, C. a., Graeber, D., and Bro, R. (2013). Fluorescence spectroscopy and multi-way techniques. PARAFAC. *Analytical Methods*, 5(23):6557.
- Ohno, T. (2002). Fluorescence Inner-Filtering Correction for Determining the Humification Index of Dissolved Organic Matter. *Environmental Science & Technology*, 36(4):742–746.
- Parker, C. a. and Barnes, W. J. (1957). Some experiments with spectrofluorimeters and filter fluorimeters. *The Analyst*, 82(978):606.
- Seitzinger, S. P., Hartnett, H., Lauck, R., Mazurek, M., Minegishi, T., Spyres, G., and Styles, R. (2005). Molecular-level chemical characterization and bioavailability of dissolved organic matter in stream water using electrospray-ionization mass spectrometry. *Limnology and Oceanography*, 50(1):1–12.
- Stedmon, C. A. and Bro, R. (2008). Characterizing dissolved organic matter fluorescence with parallel factor analysis: a tutorial. *Limnology and Oceanography: Methods*, 6(11):572–579.
- Stedmon, C. A. and Markager, S. (2001). The optics of chromophoric dissolved organic matter (CDOM) in the Greenland Sea: An algorithm for differentiation between marine and terrestrially derived organic matter. *Limnology and Oceanography*, 46(8):2087–2093.
- Stedmon, C. A., Markager, S., and Bro, R. (2003). Tracing dissolved organic matter in aquatic environments using a new approach to fluorescence spectroscopy. *Marine Chemistry*, 82(3-4):239–254.
- Stedmon, C. A. and Nelson, N. B. (2015). The Optical Properties of DOM in the Ocean. In Hansell, D. A. and Carlson, C. A., editors, *Biogeochemistry of Marine Dissolved Organic Matter*, chapter 10, pages 481–508. Elsevier, Burlington, academic p edition.
- Twardowski, M. S., Boss, E., Sullivan, J. M., and Donaghay, P. L. (2004). Modeling the spectral shape of absorption by chromophoric dissolved organic matter. *Marine Chemistry*, 89(1-4):69–88.
- Zepp, R. G., Sheldon, W. M., and Moran, M. A. (2004). Dissolved organic fluorophores in southeastern US coastal waters: correction method for eliminating Rayleigh and Raman scattering peaks in excitation–emission matrices. *Marine Chemistry*, 89(1-4):15–36.