

**Dissolved organic matter fluorescence**

A. Butturini and  
E. Ejarque

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



This discussion paper is/has been under review for the journal Biogeosciences (BG).  
Please refer to the corresponding final paper in BG if available.

# Technical Note: Dissolved organic matter fluorescence – a finite mixture approach to deconvolve excitation-emission matrices

**A. Butturini and E. Ejarque**

Departament d'Ecologia, Facultat de Biologia, Universitat de Barcelona (UB), Diagonal 643,  
08028 Barcelona, Spain

Received: 28 February 2013 – Accepted: 1 March 2013 – Published: 8 March 2013

Correspondence to: A. Butturini (abutturini@ub.edu)

Published by Copernicus Publications on behalf of the European Geosciences Union.

## Abstract

The analysis of the shape of excitation-emission matrices (EEMs) is a relevant tool for exploring the origin, transport and fate of dissolved organic matter (DOM) in aquatic ecosystems. Within this context, the decomposition of EEMs is acquiring a notable relevance. A simple mathematical algorithm that automatically deconvolves single EEM is described, creating new possibilities for the comparison of DOM fluorescence properties and EEMs that are very different from each other. A mixture model approach is adopted to decompose complex surfaces into sub-peaks. The laplacian operator and the Nelder–Mead optimization algorithm are implemented to individuate and automatically locate potential peaks in the EEM landscape. A small heterogeneous data set of 21 EEMs from a human-impacted Mediterranean river is used to describe the model application and to illustrate a strategy that optimises the search for the optimal output.

## 1 Introduction

Since the pioneering works of Traganza (1969) and Coble et al. (1990) the analysis of fluorescence properties of dissolved organic matter (DOM) in aquatic ecosystems has become an essential technique in the exploration of qualitative changes and fate of dissolved organic carbon in aquatic ecosystems (Hudson et al., 2007; Fellman et al., 2010; Ishii and Boyer, 2012).

This analytical technique has benefited considerably from instrumentation advances that facilitate the rapid analysis of large amount of samples in a short period of time and allow the storage of large data sets. This has motivated the generation of EEMs, which consist in three-dimensional matrices that arise from joining a set of fluorescence intensities emitted along a range of emission wavelength (typically from 300 to 600 nm) measured at different excitation wavelengths (typically from 240 to 450 nm).

EEMs are essentially displayed as topographic maps (contour plots) and the form of their landscape responds to a complex mixture of fluorescent compounds

**BGD**

10, 4711–4732, 2013

## Dissolved organic matter fluorescence

A. Butturini and  
E. Ejarque

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



(fluorophores) about which relatively little is known. The main challenge consists in individuating the location and relevance of fluorescence events that compose the fluorescence spectra. To date, decomposition of fluorescence spectra is performed with advanced supervised (PCA, N-PLS, PARAFAC) or unsupervised (self-organizing map) statistical multivariate techniques (Bierozza et al., 2009). These algorithms strongly enhanced the study of DOM. However, their use is a matter of debate (Felman et al., 2010) and deep analysis (Bierozza et al., 2009). Multivariate tools are usually executed with data sets that include a large number of EEMs (typically more than 100) and robust results are more easily obtained when a data set integrates samples that follow gradual gradients (Stedmon and Bro, 2008). Conversely, to our knowledge, an algorithm that decomposes the signal of individual EEM opening the perspective to work with a reduced number of EEMs and, at the same time, to compare deconvolution results from EEMs that do not follow necessarily any gradient is unavailable.

In this note, we introduce an alternative approach that integrates a simple surface analysis with the finite distribution mixture (FDM) approach. Similarly to the tools mentioned previously, FDM is widely used for data mining and pattern recognition. It assumes that a single complex surface (an EEM for example) can be deconvolved into  $n$  subjacent peaks (Frjgühwirth-Schnatter, 2006). In consonance with multiway techniques FDM assumes that peaks behave independently, without interference between them. A basic difference with respect to these tools lies in the assumption that peaks fit a predefined probabilistic density function. A peak is simply a mathematical unit that isolates a fluorescence event. This unit is not necessarily synonym of “fluorophore”. EEMs from pure fluorescent substances frequently showed single or multiple peaks that can be roughly approximated to a Gaussian bell (for some example, see Boehme and Coble, 2000; Yamashita and Tanoue, 2003; Hudson et al., 2007). On the other hand, due to their chemical intrinsic complexity, fluorescence events from natural DOM samples, can not be attributed solely to specific fluorophores (Del Vecchio and Blough, 2004; Chen and Kenny, 2007). However the Gaussian shapes frequently emerge when we observe fluorescence signal in natural samples. Boehme and Coble

**BGD**

10, 4711–4732, 2013

## Dissolved organic matter fluorescence

A. Butturini and  
E. Ejarque

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



(2000), reported the detection of a small pool of fluorophores with dual peaks with “circular contours”. EEM from algae extract show two-three clear peaks that call in mind a Gaussian bell (Her et al., 2003). Therefore, it is not surprising that researchers attempt to fit fluorescence signals with one-dimensional Gaussian distribution (Korshin et al., 1999; Westerhoff et al., 2001).

These preliminary considerations are at the heart of the idea to adopt the FDM to decompose the fluorescence events in individual EEM. In this note, besides the model description, a strategy to optimise the selection of an optimal model for an EEM is reported. Finally, the model is applied to a heterogeneous data set that includes 21 EEMs collected along the longitudinal continuum of an impacted Mediterranean river. Implementation of FDM is executed with Wolfram Mathematica® program (version 7 was used in this study). A didactical example of the model and its implementation with this software is available at the following link <http://hdl.handle.net/2445/33820>. However, FDM can be computed with any other mathematical software.

## 2 Model description

Within the FDM context, an EEM is a bivariate matrix ( $f_{(xy)}$ ) that can be described as the sum of  $n$  distributions ( $c_{(x,y)}$ , Eq. 1). In FDM research, Gaussian distributions are the most used probability models (Frjgühwirth-Schnatter, 2006). However, in this model we implemented the asymmetric Gaussian distribution (Kato et al., 2002) to capture peaks with eventual asymmetries and/or long tails (Eq. 2). The parameters that describe each distribution are their mean  $\mu_i$  ( $\mu_{ix}$ ,  $\mu_{iy}$ ), deviation  $\sigma_i$  ( $\sigma_{ix}$ ,  $\sigma_{iy}$ ), height  $a_i$  ( $a_{ix}$ ,  $a_{iy}$ ) and

BGD

10, 4711–4732, 2013

### Dissolved organic matter fluorescence

A. Butturini and  
E. Ejarque

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



skewness  $r_i$  ( $r_{ix}$ ,  $r_{iy}$ ):

$$Z_{(x,y)} = \sum_{i=1}^n c_{(x,y)_i} \quad (1)$$

$$c_{(x,y)_i} = a_i e^{-\left( \begin{cases} \frac{(\mu - \mu_i)^2}{2\sigma_i^2} & \text{if } \mu > \mu_i \\ \frac{(\mu - \mu_i)^2}{2r_i^2\sigma_i^2} & \text{otherwise} \end{cases} \right)} \quad (2)$$

5 where  $Z_{(x,y)}$  is the sum of  $n$  peaks  $c_{(x,y)}$  that fit a bivariate asymmetric Gaussian distribution model. If  $r = 1$ , Eq. (2) is equivalent to a gaussian distribution.

In Eqs. (1) and (2), estimates of the unknown parameters ( $\mu_i$ ,  $\sigma_i$ ,  $a_i$  and  $r_i$ ) are performed following two main steps:

Step A: Surface analysis to detect and locate the potential peaks in  $f_{(xy)}$ ,  $S_n =$

10  $\{\mu_1, \mu_2, \mu_3, \dots, \mu_n\}$ ;

Step B: Optimal model selection criteria and estimate of the parameters  $a_i$ ,  $\sigma_i$  and  $r_i$ .

To avoid chemically meaningless results, the only requirement is that all selected peaks must have a positive height ( $a_i > 0$ ).

15 The two steps are detailed below.

## 2.1 Step A: detection and location of candidate peaks

It consists of an analysis of the surface of  $f_{(xy)}$  to detect the position of potential peaks in a EEM. This step combines two search strategies:

(a) Detection of global and local maxima in the  $f_{(xy)}$ :

$$20 \text{ Max } f_{(xy)} = \{\mu_a, \mu_b, \mu_c, \dots, \mu_n\} \quad (3)$$

BGD

10, 4711–4732, 2013

## Dissolved organic matter fluorescence

A. Butturini and  
E. Ejarque

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



(b) Detection of local minima of the differential Laplacian operator of  $f_{(xy)}$  ( $\nabla^2 f$ ):

$$\text{Min } \nabla^2 f = \{\mu'_a, \mu'_b, \mu'_c, \dots, \mu'_n\} \quad (4)$$

$\nabla^2 f$  describes the sum of the second derivative of  $f_{(xy)}$  with respect to  $x$  and  $y$  (Ganza and Vorozhtsov, 1996). It is used to detect shoulders and edges in complex surfaces. In chemometrics, the local minimum of second derivative is used to identify the position of non-evident peaks in complex chromatograms (Stevenson et al., 2012). Here, we extend this idea to two dimensions.

The search for maxima in  $f_{(xy)}$  and minima in  $\nabla^2 f$  is performed with the Nelder–Mead optimisation algorithm under constrained conditions (Horst and Pardalos, 1995). The sensitivity of this algorithm can be increased or reduced by modifying selected parameters (namely: the contraction ratio, the expansion ratio, the reflection ratio and the shrink ratio). In our application we used the standard values for these parameters (0.5, 2, 1, and 0.5, respectively, Nelder and Mead, 1965) as they guaranteed an exhaustive search of main local minima in  $\nabla^2 f$  into a relatively short computational time. The  $\nabla^2 f$  operator is sensible to edges. Therefore minimum in  $\nabla^2 f$  surface found in the proximity of the Raman and Rayleigh–Tyndall scattering are omitted.

Once Max  $f_{(xy)}$  and Min  $\nabla^2 f$  are obtained, results are joined to sort all distinct coordinates that appear in the two lists:

$$L_n = \text{Max } f_{(xy)} \cup \text{Min } \nabla^2 f \quad (5)$$

where  $L_n$  is the list of the potential  $n$  peaks in  $f_{(x,y)}$ . In complex surfaces the Nelder–Mead algorithm can be easily trapped in local minima (or maxima) that are very close to each other and, presumably, are identifying the same peak. From a statistical perspective it is assumed that these neighbour peaks fall into the same cluster. In this case it is necessary to merge them into a single coordinate. The search for clusters is performed according to the *fixed radius near neighbour* approach (Bentley et al., 1977): at each detected coordinate ( $\mu_i$ ), a circular influence area (IA<sub>*i*</sub>) of radius  $R$  is

associated ( $IA_j = \pi R^2$ ), centred at the point  $\mu_j$ . The value of the radius  $R$  is the same for all detected  $\mu_j$  and is fixed to set the IA value to 10% of the planar area of the surface matrix. Those coordinates (different from  $\mu_j$ ) that fall within the area  $IA_j$  of  $\mu_j$  are automatically grouped into a same cluster. Two criteria are established to assign a coordinate to each cluster:

Criteria # 1 (applicable for Eqs. 3 and 5): the coordinate with the highest maxima is selected, the rest are discarded.

Criteria # 2 (applicable for Eq. 4): the coordinate with the lowest  $\nabla^2 f$  is selected, the rest are discarded.

## 2.2 Step B: Optimal model selection criteria

Once a set of  $n$  potential peaks and their coordinates ( $L_n$ ) have been individuated, it is necessary to choose the optimal dimension of the model (i.e. the optimal number of peaks) that fit the surface matrix  $f_{(x,y)}$ . This search is performed according the following procedure.

First all, we extract all  $i$  distinct proper subsets of  $L_n$ :

$$P(L_n) = \{\{\mu_1\}_1, \{\mu_2\}_2, \{\mu_3\}_3, \{\mu_1, \mu_2\}_4 \dots \dots \{\mu_1, \mu_2, \mu_3, \dots, \mu_n\}_i\} \quad (6)$$

where  $i = 2^n - 1$ .

Successively, FDM (Eqs. 1 and 2) is run for each  $i$  subset. The Bayesian Information Criterion (BIC) is used to reduce the risk of overestimating the model parameters (i.e. the number of peaks):

$$BIC_i = -2 \ln(L_i) + k_i \ln(O) \quad (7)$$

where  $L_i$  is the maximized likelihood of the model associated to the subset  $i$ ;  $k_i$  number of input parameters (i.e. number of element in the subset  $i$ );  $O$  is the sample size. The model with the smallest BIC value is selected as the optimal model (Schwarz, 1978). The optimal model is considered valid if it fits the observed data with  $r^2 > 0.99$ .

## Dissolved organic matter fluorescence

A. Butturini and  
E. Ejarque

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



Figure 1 shows this process for a real EEM from the data set. In this example the step A identifies nine potential peaks ( $n = 9$ ) that generate  $2^9 - 1 = 511$  subsets. Each subset has one model output. Out of them 53 ( $\sim 10\%$ ) generated “good” model outputs ( $r^2 > 0.99$ , the black dots in Fig. 1a). Number of peaks of these candidate “good” models ranged between five and nine. Within this reduced pool of models, the selected optimal model (lowest BIC value) is that one with eight peaks and are discarded all models that might be overestimating the number of peaks. The eight peaks are shown in Fig. 2.

This search can be accelerated by removing all subsets that do not have any chance to generate a reasonably satisfactory output. A criteria might be to remove all subsets shorter than the length of the list obtained with Max  $f_{(xy)}$  (Eq. 3). In fact it is highly improbable to fit reasonably well  $f_{(xy)}$  with less peaks that those detected with Eq. (3). In the preceding example, the numbers of subsets decrease to 64.

### 3 Model application to a data set

#### 3.1 The data set and fluorescence measurements

The data set integrates 21 EEMs (labelled from S1 to S21) obtained along the entire fluvial continuum of la Tordera river, a 60 km-long human-impacted Mediterranean river, which drains a catchment of 870 km<sup>2</sup> located 70 km to the north-east of Barcelona (Catalonia, Spain). Samples were collected in April 2012, under basal discharge conditions ( $2 \text{ m}^3 \text{ s}^{-1}$  at the outlet). DOC concentration ranged from 0.5 ppm in headwaters, to 2.9 ppm. Fluorescence analyses were performed with a Shimadzu RF-5301 PC spectrofluorometer. Raw EEM data were corrected and normalised to allow inter study comparison following the steps described in Goletz et al. (2011). Data were normalised by the area under the Raman peak of a deionised water sample at  $\lambda_{\text{ex}} = 350 \text{ nm}$  and  $\lambda_{\text{em}} = 371\text{--}428 \text{ nm}$  (Lawaetz and Stedmon, 2009). Inner filter effects were corrected by comparing absorbance measurements according to Lackowicz (2006). Absorption

BGD

10, 4711–4732, 2013

## Dissolved organic matter fluorescence

A. Butturini and  
E. Ejarque

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



spectra were measured with a UV-Visible spectrophotometer UV1700 Pharma Spec (Shimadzu). Each EEM consists of a  $\{x, y, z\}$  list of 1050 elements.

EEM shapes are extremely variable. Differences among EEMs do not follow a clear gradient along the river continuum due to a high anthropogenic influence and heterogeneity of land uses along the catchment (Fig. 3).

### 3.2 Deconvolution output

Table 1 summarizes deconvolution results for each EEM. Optimal models presented  $r^2$  values ranking between 0.993 (S3) and 0.999 (S6). Figure 3 allows comparing visually the original EEMs and their respective modelled versions for five samples.

At step A, the number  $n$  of potential peaks, in each EEM, ranges between seven and eleven and it decreases to five (S13) and ten (S19) after step B. The number of selected peaks tends to increase down river, however the trend is not significant ( $r = 0.53$ ,  $df = 19$ ,  $p > 0.01$ ).

When the coordinates of all peaks, from all samples, are represented in a unique Excitation vs. Emission scatter plot, data are grouped around eleven centers that allow to subdivide the plot into eleven sub-regions according the Voronoi diagram tessellation approach (Aurenhammer and Klein, 2000). All points that lie within a region are assigned to the centers of that region (Fig. 4).

Most of the identified peak coordinates matched well (or are close) to that found in the literature with visual peak-picking (Baker, 2002; Hudson et al., 2007) or more sophisticated multivariate methods (Cory and McKnight, 2005; Fellman et al., 2009; Table 2). Peaks from P1 to P3 encompass the protein-like region (frequently named peaks B, T1, and T2, Coble, 1998). The remaining peaks (from P4 to P11) are located in the humic/fulvic-like region (peaks A, C and M, according the nomenclature proposed by Coble et al., 1998).

Significant correlations between peak intensities ( $a_i$ ) are observed. Some of them related protein-like peaks (P3 and P2) with that of humic/fulvic like (P4, P11 and P10).

**BGD**

10, 4711–4732, 2013

## Dissolved organic matter fluorescence

A. Butturini and  
E. Ejarque

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



Others suggest a relationship between peaks within the humic/fluvic region (P4 with P11) or within the protein-like region (P1 with P3) (Table 3).

Peaks position and relevance (in terms of peak intensity,  $a_i$ ) change within a constrained region. For instance, in the broad region traditionally named peak “A” ( $230 \leq \lambda_{ex} \leq 260$ ,  $400 \leq \lambda_{em} \leq 500$ , Ishii and Boyer, 2012), the fluorescence maxima appears at  $\lambda_{ex}/\lambda_{em} \sim 246/441$  nm (P5) between the headwaters and 35 km. Downriver, it shifts to larger emission wavelength  $\lambda_{ex}/\lambda_{em} \sim 250/460$  nm (P4). These two peaks never coincided in the same sample, with the exception of S15 (Fig. 5a). This shift in peak positions and intensities might indicate changes in DOM molecular weight along the river ecosystem: from relatively small compounds with low concentration to larger and more concentrated ones (Ishii and Boyer, 2012).

In the protein-like region, the P2 ( $\lambda_{ex}/\lambda_{em} \sim 290/356$  nm) appears at head waters with a maximum between 24 and 40 km. It coexists with P1 ( $\lambda_{ex}/\lambda_{em} \sim 272/319$  nm; Fig. 5b). They are considered labile substrates related to different biological processes such as bacterial (Hudson et al., 2007), dead organisms or primary producers leachates (Fellman et al., 2010). Additionally, these signals are frequently associated to wastewater treatment plant effluents (Saadi et al., 2006; Baker, 2002). The elevated intensities of these protein-like peaks in the data set, presumably indicates the inputs of anthropogenic origins (Hudson et al., 2007; Saadi et al., 2006). The abrupt detection of P1 might indicate the increase of contribution of degraded proteins/peptides downriver (Fellman et al., 2010).

Finally, peaks 7 and 11 (within the humic/fulvic region) are detected abruptly in two different points of the river continuum: P7 appears at 24 km, meanwhile P11 at 37 km. Both peaks decreased gradually in intensity down waters (Fig. 5c). P7 has been detected in outlets of wastewater treatment plants (Saadi et al., 2006).

**BGD**

10, 4711–4732, 2013

## Dissolved organic matter fluorescence

A. Butturini and  
E. Ejarque

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



## 4 Conclusions

Advances in our knowledge of DOM fluorescence properties in aquatic ecosystems strongly benefit from technological advances in data acquisition of fluorescence data. In consequence deconvolution statistical tools are becoming indispensable to manage and extract information from the large amount of data generated by these instruments. However sophisticated spectrofluorometers are not widespread and the opportunity to generate large data set of EEMs is limited in several laboratories. Furthermore the community recognizes limitations of the multivariate tools (Fellman et al., 2009) especially when the data set is integrated by a heterogeneous pool of EEMs (Stedmon and Bro, 2008). Therefore a large portion of studies on DOM fluorescence does not implement the deconvolution tools to explore EEMs properties.

Finite mixture models are usual in several scientific disciplines (McLachlan and Peel, 2004). For instance, a recent implementation of FDM, that shows an evident similitude with that described in this note, consists in the identification of peaks/subpopulations in bidimensional cytograms (Boedigheimer and Ferbas, 2008).

FDM analyzes individual EEMs, but it is not designed to analyze differences among EEMs or dependencies among peaks within one EEM. Therefore successive data treatment is necessary to analyse these two essential aspects. Here, the information extracted by deconvolution of the data set is exposed in a simple and intuitive way for a descriptive purpose. Evidently, each researcher can adopt the appropriate statistics to attain all these aspects according to the objectives and the experimental design.

We conclude that the FDM expands the family of deconvolution tools opening the perspective to implement it with data sets composed by extremely different EEMs. The idea underlying the FDM is intuitive and the mathematical language is not excessively complex. Experts on data mining tools remark that sophisticated and complex deconvolution tools can produce similar results to those obtained with a simple peak-picking (Bieroza et al., 2011). In this framework the approach described here could be viewed as an intermediate step between the two extremes because it integrates an improved

BGD

10, 4711–4732, 2013

### Dissolved organic matter fluorescence

A. Butturini and  
E. Ejarque

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



version of the peak-picking into a relatively simple deconvolution algorithm. This aspect might further help to bring researchers closer to these techniques.

*Acknowledgements.* This research is funded by the Spanish Ministry of Education and Science (MEC) (CGL2011-30151-C02-02). Elisabet Ejarque's research was supported by an FPU doctoral scholarship from the MEC (AP2008-03431). Both authors are member of the Graccie consortium. We would like to thank Madgalena Bieroza, Jordi Flos and Eusebi Vazquez for useful comments and suggestions on an earlier version of this manuscript.

## References

Aurenhammer, F. and Klein, R.: Voronoi diagrams, in: Handbook of Computational Geometry, edited by: Sack, J. R. and Urrutia, J., North-Holland, Amsterdam, Netherlands, 201–290, 2000.

Baker A.: Fluorescence excitation–emission matrix characterization of river waters impacted by a tissue mill effluent, Environ. Sci. Technol., 36, 1377–1382, 2002.

Bentley J., Stanat, D., and Williams Jr., E.: The complexity of finding fixed-radius near neighbors, Inform. Process. Lett., 6, 209–213, 1977.

Bieroza, M., Baker, A., and Bridgeman, J.: Classification and calibration of organic matter fluorescence data with multiway analysis methods and artificial neural networks: an operational tool for improved drinking water treatment, Environmetrics, 22, 256–270, 2011.

Bieroza, M., Baker, A., and Bridgeman, J.: Exploratory analysis of excitation-emission matrix fluorescence spectra with self-organizing maps as a basis for determination of organic matter removal efficiency at water treatment works, J. Geophys. Res.-Biogeo., 114, G00F07, doi:10.1029/2009JG000940, 2009.

Boedigheimer, M. J. and Ferbas, J.: Mixture modelling approach to flow cytometry data, Cytom. Part A, 73A, 421–429, 2008.

Boheme, J. R. and Coble, P. G.: Characterization of coloured dissolved organic matter using high energy laser fragmentation, Environ. Sci. Technol., 34, 3283–3290, 2000.

Coble, P. G., Green, S. A., Blough, N. V., and Gagosian, R. B.: Characterization of dissolved organic-matter in the Black Sea by fluorescence spectroscopy, Nature, 348, 432–435, 1990.

Coble, P. G., Del Castillo, C. E., and Avril, B.: Distribution and optical properties of CDOM in the Arabian Sea during the 1995 SW monsoon, Deep-Sea Res. Pt. II, 45, 2195–2223, 1998.

## Dissolved organic matter fluorescence

A. Butturini and  
E. Ejarque

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



## Dissolved organic matter fluorescence

A. Butturini and  
E. Ejarque

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



Fellman, J. B., Hood, E., and Spencer, R. G. M.: Fluorescence opens new windows into dissolved organic matter dynamics in freshwater ecosystems: a review, *Limnol. Oceanogr.*, 55, 2452–2462, 2010.

Frgjuhwirth-Schnatter, S.: Finite Mixture and Markov Switching Models, Springer Series in Statistics, doi:10.1007/978-0-387-35768-3, 492 pp., 2006.

Ganza, V. G. and Vorozhtsov, E. V.: Numerical Solution for Partial Differential Equations. Problem Solving Using Mathematica, CRC Press, 347 pp., 1996.

Goletz, C., Wagner, M., Gruebel, A., Schmidt, W., Korf, N., and Werner, P.: Standardization of fluorescence excitation-emission-matrices in aquatic milieu, *Talanta*, 85, 650–656, doi:10.1016/j.talanta.2011.04.045, 2011.

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

Horst, R. and Pardalos, P. M.: Handbook of Global Optimization, Kluwer, Dordrecht, Netherlands, 900 pp., 1995,

Hudson, N., Baker, A., and Reynolds, D.: Fluorescence analysis of dissolved organic matter in natural, waste and polluted waters – a review, *River Res. Appl.*, 23, 631–649, 2007.

Ishii, S. K. and Boyer, T. H.: Behaviour or reoccurring PARAFAC components in fluorescent dissolved organic matter in natural and engineering systems: a critical review, *Environ. Sci. Technol.*, 46, 2006–2017, 2012.

Korshin, G. V., Kumke, M. U., Li, C. W., and Frimmel, F. H.: Influence of chlorination on chromophores and fluorophores in humic substances, *Environ. Sci. Technol.*, 33, 1207–1212, 1999.

Lakowicz, J. R.: Principles of fluorescence spectroscopy, 3rd Edn., Springer, 1255 pp., 2006.

Lawaetz, A. J. and Stedmon, C. A.: Fluorescence intensity calibration using the Raman scatter peak of water, *Appl. Spectrosc.*, 63, 936–940, 2009.

McLachlan G. and Peel, D.: Finite mixture models, in: Wiley Series in Probability and Statistics: Applied Probability and Statistics, Wiley-Interscience, 456 pp., 2004,

Nelder, J. A. and Mead, R.: A simplex method for function minimization, *Comput. J.*, 7, 308–313, 1965.

Saadi, I., Borisover, M., Armon, R., and Laor, Y.: Monitoring of effluent DOM biodegradation using fluorescence, UV and DOC measurements, *Chemosphere*, 63, 530–539, 2006.

**Dissolved organic matter fluorescence**A. Butturini and  
E. Ejarque

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



Stedmon, C. A. and Bro, R.: Characterizing dissolved organic matter fluorescence with parallel factor analysis: a tutorial, *Limnol. Oceanogr.-Meth.*, 6, 572–579, 2008.

Stevenson, P. G., Mnatsakanyan, M., Guichon, P. G., and Shalliker, R. A.: Peak picking and the assessment of separation performance in two-dimensional high performance liquid chromatography, *Analyst*, 135, 1541–1550, 2010.

Schwarz, G.: Estimating the dimension of a model, *Ann. Stat.*, 6, 461–464, 1978.

Traganza, E. D.: Fluorescence excitation and emission spectra of dissolved organic matter in sea water, *B. Mar. Sci.*, 9, 897–904, 1969.

Westerhoff, P., Chen, W., and Esparza, M.: Fluorescence analysis of a standard fulvic acid and tertiary treated wastewater, *J. Environ. Qual.*, 30, 2037–2046, 2001.

Yamashita, Y. and Tanoue, E.: Chemical characterization of protein-like fluorophores in DOM in relation to aromatic amino acids, *Mar. Chem.*, 82, 255–271, 2003.

## Dissolved organic matter fluorescence

A. Butturini and  
E. Ejarque

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



**Table 1.** DOC concentration and deconvolution results for each DOM sample. Number of initial potential peaks (step A) and selected peaks after step B, fit goodness parameters ( $r^2$  and BIC) of each one optima FDM model are provided. Numbers in parenthesis in the first column show the distance (km) of each sampling point from headwater.

| Sample   | DOC (ppm) | Surface Analysis: Step A         |                        | Deconvolution: Step B |       |       |
|----------|-----------|----------------------------------|------------------------|-----------------------|-------|-------|
|          |           | Max $f_{(x,y)}$ (#) <sup>a</sup> | $L_n$ (#) <sup>b</sup> | Selected Peaks (#)    | $r^2$ | BIC   |
| S1 (3)   | 0.5       | 3                                | 11                     | 8                     | 0.994 | -7606 |
| S2 (6)   | 0.6       | 2                                | 10                     | 6                     | 0.997 | -8753 |
| S3 (8)   | 0.7       | 4                                | 10                     | 6                     | 0.993 | -7077 |
| S4 (16)  | 0.7       | 2                                | 9                      | 7                     | 0.998 | -9595 |
| S5 (19)  | 0.9       | 2                                | 9                      | 6                     | 0.997 | -8319 |
| S6 (22)  | 1.3       | 2                                | 7                      | 7                     | 0.999 | -7965 |
| S7 (24)  | 1.8       | 2                                | 8                      | 6                     | 0.994 | -5687 |
| S8 (26)  | 2         | 5                                | 9                      | 7                     | 0.998 | -5362 |
| S9 (28)  | 2.7       | 3                                | 7                      | 7                     | 0.998 | -5844 |
| S10 (30) | 2.9       | 3                                | 8                      | 8                     | 0.997 | -5830 |
| S11 (32) | 2.9       | 3                                | 9                      | 8                     | 0.998 | -5743 |
| S12 (34) | 2.5       | 4                                | 7                      | 6                     | 0.997 | -5994 |
| S13 (35) | 2.5       | 3                                | 8                      | 5                     | 0.995 | -4816 |
| S14 (37) | 2.4       | 4                                | 7                      | 7                     | 0.995 | -4321 |
| S15 (39) | 2.3       | 5                                | 9                      | 8                     | 0.996 | -4621 |
| S16 (41) | 2.1       | 4                                | 9                      | 8                     | 0.995 | -4649 |
| S17 (42) | 1.9       | 4                                | 9                      | 8                     | 0.996 | -5086 |
| S18 (45) | 2.1       | 4                                | 9                      | 9                     | 0.998 | -5578 |
| S19 (51) | 1.9       | 4                                | 11                     | 10                    | 0.998 | -6023 |
| S20 (53) | 2.2       | 4                                | 11                     | 8                     | 0.998 | -6173 |
| S21 (58) | 1.8       | 4                                | 9                      | 8                     | 0.998 | -5908 |

<sup>a</sup> Length of list Max  $f_{(x,y)}$  (Eq. 3); <sup>b</sup> Number of potential peaks (length of list  $L_n$ , Eq. 5).

## Dissolved organic matter fluorescence

A. Butturini and  
E. Ejarque

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



**Table 2.** Coordinates and brief description of each one of the twelve main peaks detected in the data set after the deconvolution.

| Peak ID | $\lambda_{ex}$ (nm) | $\lambda_{em}$ (nm) | # cases <sup>c</sup> | Conventional class <sup>a</sup> |    |    |                   |   |   |   |
|---------|---------------------|---------------------|----------------------|---------------------------------|----|----|-------------------|---|---|---|
|         |                     |                     |                      | Protein like                    |    |    | Humic/Fulvic like |   |   |   |
|         |                     |                     |                      | B                               | T1 | T2 | A                 | M | C |   |
| P1      | 272                 | 319                 | 7                    | X                               |    |    |                   |   |   |   |
| P2      | 290                 | 356                 | 21                   |                                 | X  |    |                   |   |   |   |
| P3      | 231                 | 339                 | 17                   |                                 |    | X  |                   |   |   |   |
| P4      | 250                 | 460                 | 9                    |                                 |    |    | X                 |   |   |   |
| P5      | 246                 | 441                 | 14                   |                                 |    |    | X                 |   |   | X |
| P6      | 239                 | 385                 | 17                   |                                 |    |    | X                 |   |   | X |
| P7      | 269                 | 433                 | 5                    |                                 |    |    | X <sup>b</sup>    |   |   |   |
| P8      | 326                 | 402                 | 11                   |                                 |    |    |                   | X |   |   |
| P9      | 332                 | 431                 | 21                   |                                 |    |    |                   | X |   | X |
| P10     | 380                 | 471                 | 21                   |                                 |    |    |                   |   |   | X |
| P11     | 307                 | 471                 | 7                    |                                 |    |    | unknown           |   |   |   |

<sup>a</sup> Coble et al. (1998).

<sup>b</sup> Detected in a wastewater treatment plant (Saadi et al., 2006).

<sup>c</sup> Number of EEMs, in which the peak has been detected.

## Dissolved organic matter fluorescence

A. Butturini and  
E. Ejarque**Table 3.** Pairwise Pearson correlations values between the peak intensities identified in the 21 EEMs after the deconvolution. Bold values represent the significant correlations ( $P < 0.001$ ,  $df = 19$ ).

|     | P1          | P2          | P3          | P4          | P5    | P6   | P7    | P8    | P9   | P10  | P11 |
|-----|-------------|-------------|-------------|-------------|-------|------|-------|-------|------|------|-----|
| P1  | 1           | 0           | 0           | 0           | 0     | 0    | 0     | 0     | 0    | 0    | 0   |
| P2  | 0.19        | 1           | 0           | 0           | 0     | 0    | 0     | 0     | 0    | 0    | 0   |
| P3  | <b>0.89</b> | 0.45        | 1           | 0           | 0     | 0    | 0     | 0     | 0    | 0    | 0   |
| P4  | <b>0.73</b> | 0.29        | <b>0.83</b> | 1           | 0     | 0    | 0     | 0     | 0    | 0    | 0   |
| P5  | -0.10       | 0.29        | -0.06       | -0.34       | 1     | 0    | 0     | 0     | 0    | 0    | 0   |
| P6  | 0.52        | 0.26        | 0.39        | 0.43        | -0.12 | 1    | 0     | 0     | 0    | 0    | 0   |
| P7  | -0.27       | 0.19        | -0.33       | -0.33       | 0.28  | 0.20 | 1     | 0     | 0    | 0    | 0   |
| P8  | -0.16       | 0.01        | -0.15       | -0.32       | 0.29  | 0.18 | 0.29  | 1     | 0    | 0    | 0   |
| P9  | 0.25        | 0.50        | 0.43        | 0.35        | 0.27  | 0.20 | 0.07  | -0.02 | 1    | 0    | 0   |
| P10 | 0.17        | <b>0.75</b> | 0.38        | 0.35        | 0.30  | 0.34 | 0.35  | 0.30  | 0.61 | 1    | 0   |
| P11 | <b>0.70</b> | 0.22        | <b>0.74</b> | <b>0.74</b> | -0.44 | 0.56 | -0.23 | 0.03  | 0.18 | 0.17 | 1   |

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



## Dissolved organic matter fluorescence

A. Butturini and  
E. Ejarque

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

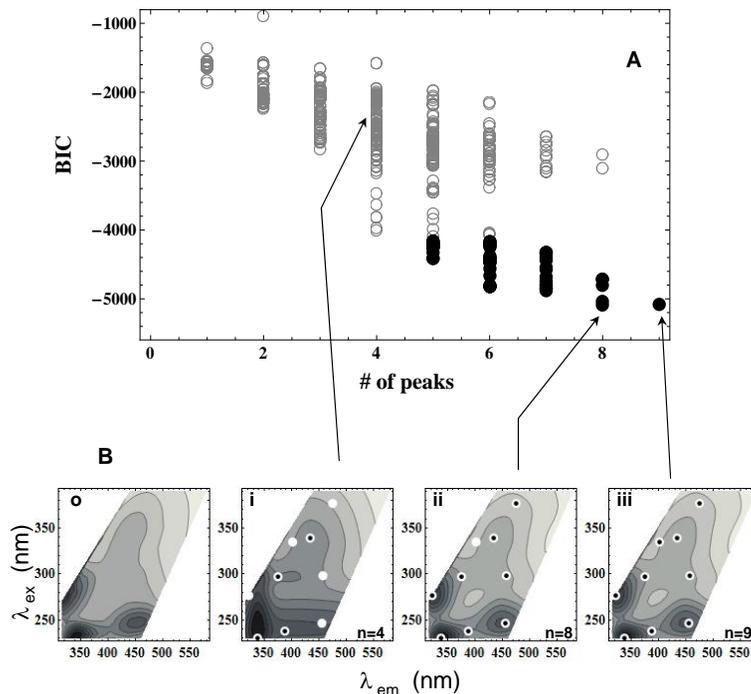
Back

Close

Full Screen / Esc

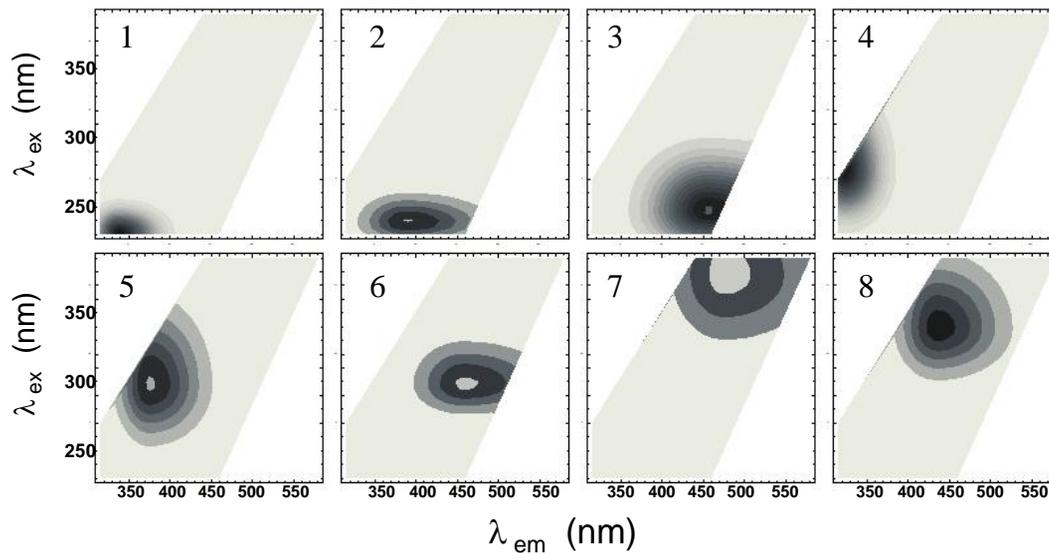
Printer-friendly Version

Interactive Discussion



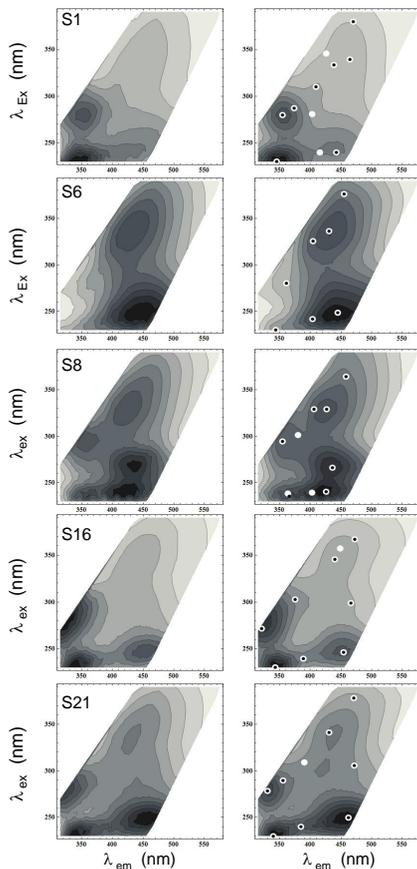
**Fig. 1.** Visual example of the optimal model selection process. This example refers to the sample S17 with nine potential peaks ( $n = 9$ ). Panel (A) shows the relationship between BIC values and number of peaks obtained executing the FDM  $z(x, y)$  (Eqs. 1 and 2) for all possible subsets  $i$  of the nine potential peaks, where  $i = 2^9 - 1 = 511$ . Gray disks and black dots discern modeled EEM adjust with  $r^2$  lower and higher than 0.99, respectively. Panel (B) shows the contour plots of original EEM sample (o), and three model outputs with a “poor” adjust (i), “optimal” adjust (i.e. lower BIC values, ii) and overfitted adjust (i.e. larger number of peaks, iii). Large white and small black dots in contours plots show location of potential and selected peaks, respectively. Contours indicates the 10% of the intensity peak.

## Dissolved organic matter fluorescence

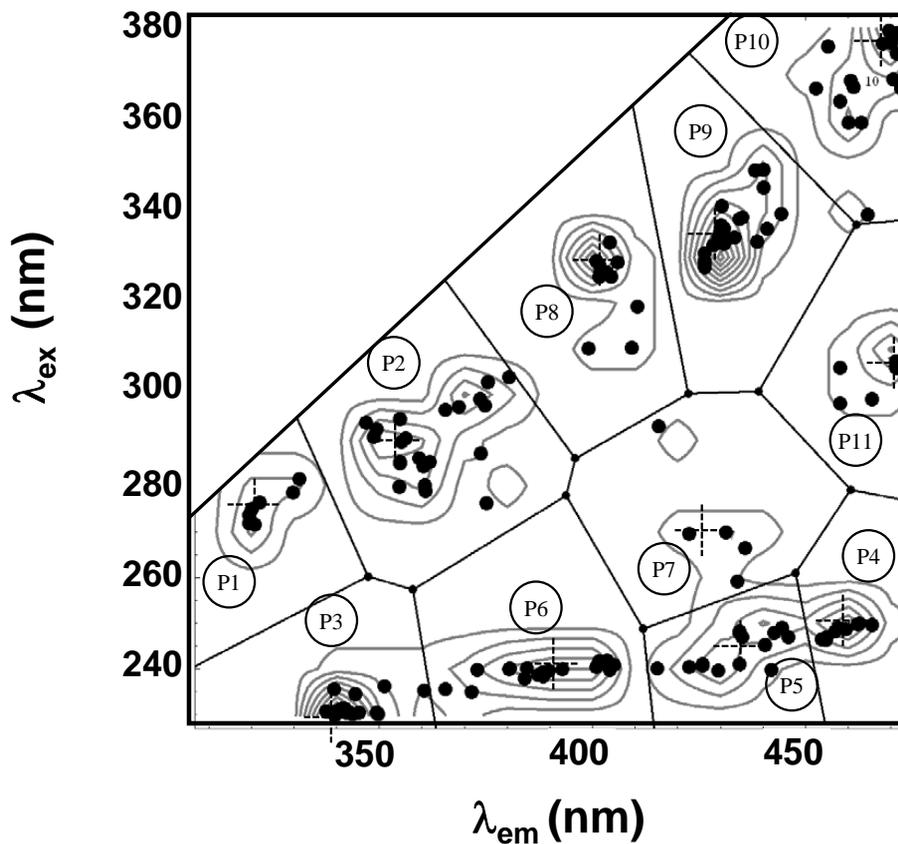
A. Butturini and  
E. Ejarque

**Fig. 2.** Shape of the height peaks individuated in sample S17 after the optimal selection model process illustrated in Fig. 2. Contours indicate the 10% of the intensity peak.

[Title Page](#)[Abstract](#)[Introduction](#)[Conclusions](#)[References](#)[Tables](#)[Figures](#)[⏪](#)[⏩](#)[◀](#)[▶](#)[Back](#)[Close](#)[Full Screen / Esc](#)[Printer-friendly Version](#)[Interactive Discussion](#)



**Fig. 3.** Examples of observed (left) and modelled (right) EEMs from la Tordera river from five sampling sites. Contours indicate the 10% of the intensity peak. Large white and small black dots in contours plots show location of potential and selected peaks, respectively.



**Fig. 4.** Distribution of peaks (black dots) within the Excitation vs. Emission plane. Contours helps to visualize the position of the centers (dotted cross) of each Voronoi polygons (black thin lines). Contour plot has been generated converting the two-dimensional data into an array of count of width of 10. Numbers into white circles label the eleven peaks in the data set.

**Dissolved organic matter fluorescence**

A. Butturini and  
E. Ejarque

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



## Dissolved organic matter fluorescence

A. Butturini and  
E. Ejarque

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

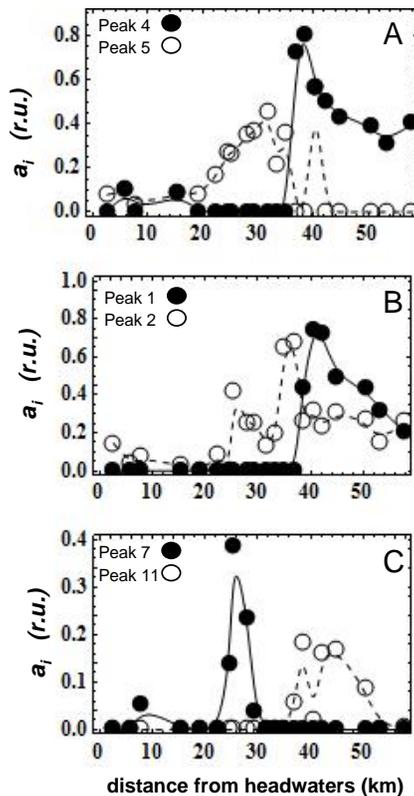
Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



**Fig. 5.** Intensity of several peaks (in Raman units) detected along the la Tordera continuum after the deconvolution. Solid and dotted lines show the smoothed trend of each peak. Lines are obtained with a cubic B-spline function and have only a descriptive purpose.