Micó Tormos, P.; García-Ballesteros, S.; Mora Carbonell, M.; Vicente Candela, R.; Amat Payá, AM.; Arqués Sanz, A. (2019). EEMlab: A graphical user-friendly interface for fluorimetry experiments based on the drEEM toolbox. Chemometrics and Intelligent Laboratory Systems. 188:6-13. https://doi.org/10.1016/j.chemolab.2019.03.001

EEMlab: a graphical user-firendly interface for fluorimetry experiments based on the drEEM toolbox

P. Micó,[a*] *S. García-Ballesteros,*[b] *M. Mora,*[c] *R. Vicente,*[b] *A. M.Amat,*[b] *A. Arques,*[b]

[a] Grupo de Procesos de Oxidación Avanzada, Departamento de Informática de Sistemas y Computadores, Universitat Politècnica de València - Campus d'Alcoi, Alcoi, Spain.
[b] Grupo de Procesos de Oxidación Avanzada, Departamento de Ingeniería Textil y Papelera, Universitat Politècnica de València - Campus d'Alcoi, Alcoi, Spain.
[c] Grupo de Procesos de Oxidación Avanzada, Departamento de Matemática Aplicada, Universitat Politècnica de València - Campus d'Alcoi, Alcoi, Spain.

*Corresponding author
E-mail addresses: aarques@txp.upv.es; pabmitor@disca.upv.es

Abstract

Fluorescence has been widely employed for the characterization of organic matter. In particular, excitation emission matrixes (EEM) provide important qualitative information on its composition. However, the application of this technique is limited by the mathematical complexity involved, which requires the use of PARAFAC for deconvolution of the EEM in their components. To overcome the numerical problem specific MATLAB toolboxes for the PARAFAC deconvolution have been implemented (e.g. drEEM). This toolbox is widely used by the scientific community but its intrinsic complexity in terms of programming knowledge makes it difficult to use. In this regard and in order to facilitate the first approximation to the PARAFAC programming problem, this paper describes and offers to the community the EEMlab software application: a graphical user-firendly interface for fluorimetry experiments based on the drEEM toolbox. The interface is developed in order to facilitate not only the intuitive use of the drEEM (no previous MATLAB knowledge is needed) but also to automate many repetitive tasks (as the data load or the modeling loop) or even to manage the different formats of files being produced by all the devices involved in the process. In order to validate the EEMlab, the same experiment documented by the drEEM is reproduced. In addition, the EEMlab is

tested again with conducting a new fluorimetry experiment and the results are presented at the end of the paper. Finally to appoint a reference to the public web site pabmitor.webs.upv.es/eemlab in where all the components of the EEMlab GUI (software, tutorial and datasets) are publicly available to the readers.

Keywords
Excitation-Emission Matrix (EEM); fluorescent experiment; composition of complex samples; mathematical deconvolution of data; multifluorophoric mixtures.

Abbreviations

EEM: excitation emission matrixes

GUI: graphical user interface

PARAFAC: Parallel Factor Analysis

EEMlab : GUI that manages the drEEM.

1. Introduction

Flourescence is a non-destructive spectroscopic method that is widely used in the analysis of complex samples. Some potential uses of this technique, because of its high sensitivity and specificity, are quality control of food and commercial products, as well as examination of natural water. Regarding foods (in particular meat, fish, cereals, fruits vegetables or sugars) fluorescence is a promising method of analysis, due to the presence of fluorescent substances such as aminoacids, vitamins and cofactors, nucleic acids, polyphenols of chlorophylls [1]. In addition, fluorescence is being employed to guarantee the place of origin and quality of products with highly added value, such as wine [2], oil [3], vinegar [4] or honey [5]. In the field of water monitoring, it has been applied to examine contamination of superficial waters

[6], analysis of specific industrial pollutants [7] or to detect highly toxic species (e.g. disinfection by-products, DBPs) that can be generated in the treatment of drinking water [8]. This methodology is interesting because it is cheap, provides a fast response and does not require a complex sample preparation.

Fluorescence spectroscopy and in particular bi-dimensional excitation emission matrix has been used to characterize the dissolved organic matter (DOM) [9-11], as it can be estimated that 40-60% of natural organic matter is fluorescent [12]. Fluorescent moieties are found in humic substances, (e.g. humic or fulvic acids) but also in the protein derived fraction, due to the fluorescence of tryptophan, tyrosine and phenylalanine. For this reason, this technique can provide valuable qualitative and even semi quantitative information about the nature of the DOM in freshwater or coastal and marine environments, in particular on their humic-like or protein-like fractions [13-15].

Among the fluorescence techniques, the use of excitation emission matrixes (EEM) is now gaining momentum. They consist in bi-dimensional sets of data, in which the fluorescence emission at a given wavelength is plotted vs. the excitation wavelength, giving information that is more valuable that a single excitation of emission spectrum. In fact, EEM might be considered as a fingerprint of some types of organic components. However, the fluorescence is affected by different parameters such as pH, complexation of the fluorophore, scattering or overlapping of signals, what is a major problem for the identification or quantitation of the signal corresponding to each component present in the sample [16]. Furthermore, the use of different spectrometers can result in artefacts in the EEM due to physical imperfections in the optical components, light sources or monochromators, what makes it difficult to compare results obtained with different apparatus [17]. Hence, correction of sample matrix effects and

standardization of results to allow comparison of results recorded with different spectrophotometers is needed [18].

Regarding the normalization of the signal given by different spectrophotometers, this problem is commonly solved by the internal processing of the apparatus; however, it can be manually done by the analyst using standards (BAM in Germany or NIST in USA [19]). In addition to this, normalization of fluorescence intensity is required. For this purpose, a standard (e.g. quinine sulphate) can be employed; however, this normalization is more usually based on the Raman scattering [20].

Absorption of light by the sample might result in an inner filter effect, decreasing the intensity of fluorescence. In this case, further correction is required, and for this purpose, several methods are available, namely Gauthier [21], Lakowicz [22] or Larsson [23]. Finally, a black subtraction is also needed.

EEM can provide further information on the composition of complex samples and its variation along time or during a treatment. However, in this case application of complex mathematical techniques are required [24]. For instance, EEMs consists in matrices that contains a response (fluorescence emission) for each combination of two wavelengths (excitation and emission). A mathematical deconvolution of these data can give different groups of substances with similar characteristics of fluorescence, also providing information of the concentration, and emission and excitation spectra of each group [25].
Different methods have been tested for deconvolution of EEM [26], being the most widely employed PCA (Principal Component Analysis) [27], PARAFAC (Parallel Factor Analysis) [28, 29], N-PLS (N-way Partial Least-Square Regression) [30] or MCR-ALS (Multivariate

Curve Resolution using Alternating Least Square) [31, 32]. Among them, maybe PARAFAC is the most widely used to discover the underlying components hidden by groups of fluorophores with similar features [33-35]; however, the mathematical complexity and the highly time-consuming preparation of data that is required is a major drawback for this technique to be employed for non-skilled researchers.

In his paper we present EEMlab, a graphical user interface (GUI) for experimental research. The paper is divided as follows: (i) a short review on the drEEM toolbox on which the EEMlab processing core is based; (ii) a description of the EEMlab application, the core scripts, the GUI, workflow, structural design, programming improvements and installation details; (iii) an experimental part that illustrates the EEMlab use with conducting experiments on the old (drEEM) and newly produced (EEMlab) datasets; (iv) the conclusions about the presented GUI and (v) the future work and EEMlab improvements.

2. Theory/calculation

2.1. A short review on the drEEM toolbox

The drEEM toolbox is a complete collection of Matlab® functions implemented not only for the PARAllel FACtor analysis (PARAFAC) but also for the fluorimetry EEM datasets load, correction and preprocessing. For all that and nowadays, drEEM becomes the reference tool in the scope of PARAFAC fluorimetry applications. The toolbox scripts, its description with the instructions for the correct use in fluorimetry experiments and an example dataset are publicly available elsewhere [36]. The drEEM example is oriented to highlight how important the correction and preprocessing of the raw dataset are for a correct PARAFAC modeling of the fluorescence sources. Although the undeniable success of the use of the

drEEM toolbox in most of the fluorimetry applications published in the recent times, the toolbox and the example dataset present two drawbacks that are outlined next.

- the toolbox is developed in flavor of script programming and processing. This becomes a handicap for all those users that are not confident with Matlab programming language
- the dataset collection of EEM files does not match with the experiment information detailed in the samplelog file. The most of the times, mismatches are unnoticed by the user

Then, a proposal to improve the use of the toolbox remains in the implementation of a GUI that manages the drEEM and handles all the (frequent) errors in the experiment definition and the formats of the raw files. From now on, we will refer to this GUI as EEMlab.

2.2. The EEMlab application

The EEMlab is a GUI developed with the Matlab® programming language that manages the drEEM toolbox. This application is developed with the aim of improving the drawbacks outlined in the previous section. For a quick understanding the EEMlab is structured in two parts: (i) the processing core and (ii) the GUI.

The EEMlab's processing core is based on the drEEM toolbox and extends its functionality. On the other hand, the GUI is designed from the scratch and presents a clear workflow that facilitates the load, correction, preprocessing and PARAFAC modeling in fluorimetry experiments.

Then, we consider EEMlab as a user-oriented layer that internally manages the drEEM toolbox and hidden its complexity to the user. Thanks to that, chemical engineers do not need any previous Matlab knowledge to successfully conduct a fluorimetry experiment, focusing their efforts on solving chemical challenges instead of spending time on programming tasks.

2.2.1 Workflow

The EEMlab implementation follows the software prototyping methodology. A basic workflow is defined and small-scale mock-ups are developed. This iterates until the prototype evolves to meet the authors' requirements. A first prototype was based on the specific features of a concrete acquisition device (Quantamaster fluorometer). The final prototype includes all the stages needed for the complete processing of a fluorimetry experiment. All the design is based on the functionality provided by the drEEM fluorimetry toolbox and hence, all the EEMlab workflow is organized in a logical way following the demo appendix example in [33]. The EEMlab workflow results as follows:

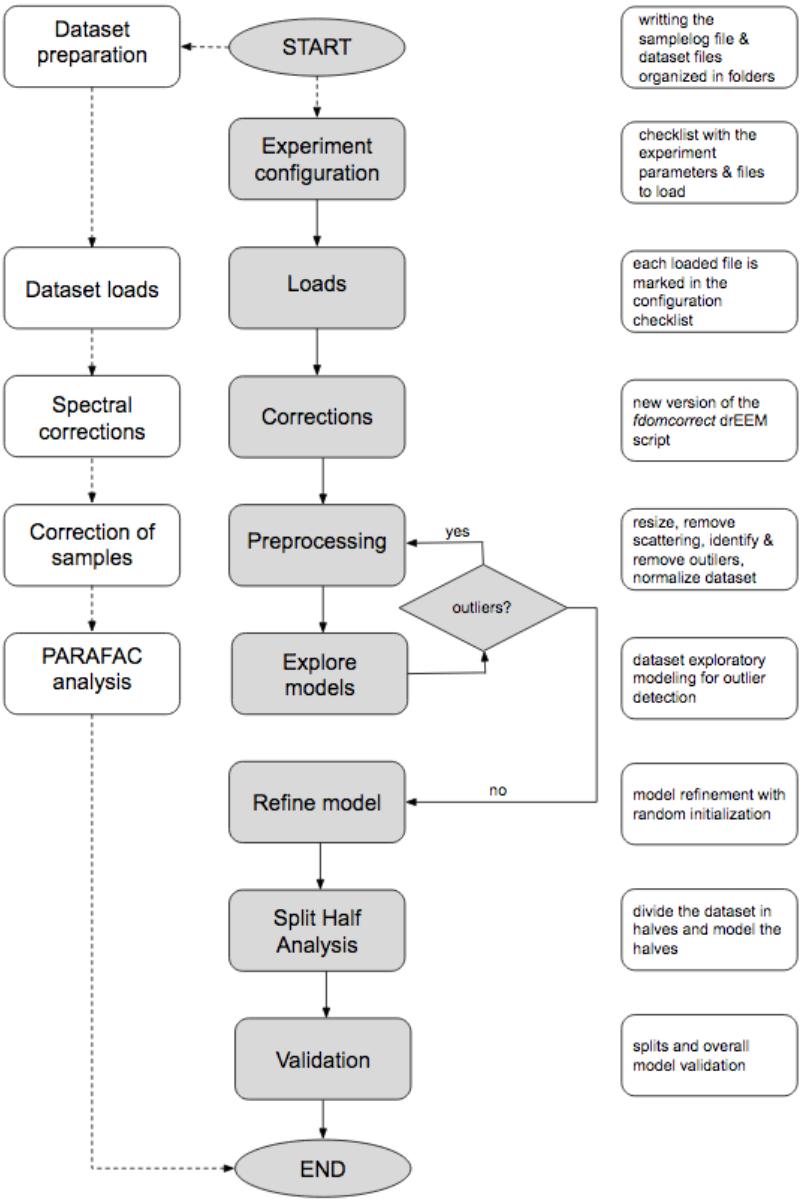

Fig. 1. EEMlab workflow

Dataset preparation

This is the step previous to working with EEMlab. In here, the experimental dataset is defined and the metadata information is recorded in a specific samplelog file. The samplelog includes the experiment parameters and the relations between the different files that compose it (raw EEMs, spectral corrections, blanks, Raman scans, normalization factors, etc.). It also

includes labels and tags associated to the data samples. All the example datasets cited and/or provided in this paper include a complete samplelog file.

Experiment configuration

To start processing a fluorimetry experiment, EEMlab needs to know in advance, which kind of files are going to be processed, how have they been produced and some other information. All these issues are defined in the EEMlab's configuration assistant, that presents a checklist in where the experiment is properly configured.

Load of files

Once the experiment is configured, we proceed with the load of all the requested files. In this regard, the user has to do the load manually, looking for the correct location of the files that are organized in folders. This is one of the hotspots in EEMlab as a series of new scripts have been implemented in order to extend the load functionality to new formats. Then, not only the load of CSV files (SPEX Fluorolog-3, Fluoromax, Hitachi, Varian, AquaLog) but also the load of TXT files (QuantaMaster fluorometer) is also supported by the system. This extends the range of acquisition devices that are valid for EEMlab. The progress in the loading process is recorded and displayed to the user by means of the config assistant.

Spectral corrections

The EEMlab corrects the spectrum of the dataset in the way defined by the samplelog. The script used is an improved version of the drEEM's fdom correct script. Depending on the acquisition device corrections can be extended to excitation and emission spectrums, to the inner filter effect and to the dilution blanks. In addition, EEMlab normalizes fluorescence intensities to Raman units and, if possible, a calibration to quinine-sulfate equivalent (QSE) units for interlaboratory comparison [17] is also done.

Preprocessing

The preprocessing consists on a manual inspection of the dataset in order to correct external artifacts (scattering), to identify and delete outlier samples and to detect corrupted samples that can still been recovered. The final preprocessing step before modeling is to normalize the dataset to unit norm, the normalization is done employing the function NORMEEM from drEEM. This function normalized each EEM to its total signal giving high and low-concentration samples similar weighting [33].

PARAFAC analysis

Once the dataset is suitable for modeling, we run PARAFAC algorithm [33], onto the dataset to find (model) all the individual fluorophores (sources) involved in the process. This analysis is divided in several parts depending on what the user is trying to achieve: the (i) exploratory modeling, the (ii) model refinement, a (iii) split half analysis and the final (iv) model validation.

Exploratory modeling

This stage enables to quickly get a model to work with. When a dataset is modeled, we can use the plotting tools to explore many aspects of the dataset: spectral loadings, correlations between components, wavelengths and samples loadings and leverages, spectral sum of squared errors, core consistency, etc. A simple manual inspection on the plots can lead the user to identify new outlier samples that remained hidden in the preprocessing stage. It also helps to decide the best number of underlying sources (fluorophores) involved in the process. Be aware of reversing the dataset normalization before a new iteration on the preprocessing.

Model refinement

When the dataset is suitable for modeling, (there are no outliers and we assume a probable number of model components) a more restrictive PARAFAC modeling loop is run. All the iterations are randomly initialized and the model with a minimum residual error is chosen.

Split half analysis

Once a dataset is modeled, we have to ensure that the model is appropriate and independent from subsamples of the dataset [37-40]. From a wide range of possibilities and taking advantage of the already implemented scripts in the drEEM toolbox, in this occasion the EEMlab enables the $S_4C_6T_3$ split half analysis while generating six split-half independent subdatasets. These subdatasets are suitable to infer the subsequent models for the next validation stage.

Model validation

The main goal of a PARAFAC analysis is to validate the overall model (the refined one). To do that, the first step is to validate some of the models inferred from the six split-half independent subdatasets. As the subdatasets composition is not controlled and following the recommendation in [38], we consider that the split half analysis is valid when validating just one of the split-half subdatasets. On the other hand, a test is valid when the product of the Tucker congruence between the loadings of the two models remains stable under a certain threshold (i.e. Tucker's correlation coefficient). In this sense, the EEMlab enables the user to define this threshold. Then, the user has to repeat the validation between the refined model and the previously validated split-half subdatasets. Finally and if the overall model is valid we can revert the normalization to manually inspect the model fingerprints and final spectral loadings.

2.2.2. Improvements

The EEMlab GUI represents an improvement by itself as reduces the programming background of the researcher. The extended functionality in the core programming scripts make this GUI a powerful tool for the correction of generic fluorimetry experiments. The main improvements are listed next.

Configuration of experiments

Sometimes, the task of loading all the files needed for a correct dataset configuration becomes a little bit messy. To help the user in checking the metadata information that defines the experiment, the EEMlab includes a configuration assistant. The assistant not only helps the researchers thinking about the requirements but it is also the way to feedback them about all the progress in the experiment loads.

Loading files

- automatic identification of data types in CSV files (getlogtypes function)
- overload of the readlogfile function to import samplelogs in XLSX format
- loads extended to both CSV and TXT formatted datasets
- in this moment, the loading functionality is limited to the format of certain fluorometers but it can be easely extended to new formats
- interaction with already EEMlab's processed MAT files with load/store options for MAT formatted datasets
- results can be saved and/or loaded in a MAT formatted file

Programming

- a new version of the drEEM's fdomcorrect to enable an individual correction of each EEM with individual spectrum correction files, in accordance to the metadata information
- neutral excitation and emission spectral correction scans are automatically provided in the case of working with spectrally corrected EEM datasets (i.e. EEMs acquisition with Quantamaster)
- automatic calculation of the optimal wavelength integration range for the Raman normalization

Validation

The drEEM's splitvalidation script is reviewed to let the user to modify the Tucker's correlation coefficient (the default value is set to a restrictive 0.95). This will affect to the validation in both, split-halves and the overall model.

Log file

In complex experiments in where many processing and iterations are considered, it is very important to report all the steps to make it reproducible. EEMlab automatically records all this information in a log file that can be listed by the user at any time. The log file also supports the error handling including all the error messages displayed during the processing. In addition, the log file is editable from both within and outside EEMlab, allowing the researcher to add personal information, annotations, comments, suggestions and/or metadata about the experiments for its best understanding.

Error handling

The EEMlab is not fault-tolerant with data. Consequently, the application includes an error checking functionality that aborts the experiment if a non-supported error is detected. The error, error location in code (script name) and cause of error are reported to the user in a log file. It is the user's responsibility to act according to the log file.

Displaying plots

The EEMlab supports and displays both, the dataset loads and the resulting models in several ways:

- numerical results are displayed throughout the Matlab's console and also recorded in the system's log file

- 2D, 3D and scatter plots are supported by the GUI itself or displayed in external figures. Plots for individual samples are embedded in the GUI. Collective plots (collections, scatter plots) are displayed out of the GUI

2.2.3. EEMlab components

The EEMlab software is composed by the processing core, the GUI bundle, two training datasets and a tutorial. The tutorial becomes a useful tool for the new users to start training with EEMlab. All the components are publicly available in the link pabmitor.webs.upv.es/eemlab. Once downloaded, the only requirement to work with the application is to include all the components into the Matlab's path. The EEMlab project website includes:

- the processing core is the drEEM's toolbox. This toolbox is public and downloads are available in [36]. The processing core is already included within the EEMlab's bundle
- the EEMlab bundle, that is composed by a series of scripts to generate the visual part of the application
- the drEEM example dataset. You can ask for the original drEEM example dataset files to the authors in http://www.models.life.ku.dk/drEEM or download them from the EEMlab's project
- the EEMlab example dataset, that becomes a controlled experiment used to validate the EEMlab application. The details about this dataset are detailed in a next section
- the EEMlab tutorial, that illustrates how to use the GUI to process the drEEM dataset

2.3. Training datasets

Two example datasets are available for training: a simplified version of the drEEMs dataset and a new EEMlab's dataset with EEM samples acquired with the QuantaMaster fluorometer. The main features of both are detailed next.

2.3.1. The drEEM dataset

This is the dataset that originally illustrates the use of the drEEM's toolbox in fluorimetry experiments. If working from the scratch with the original files, the dataset presents several mismatches between the samples and the metainformation: some referenced files are left (i.e. spectral correction scans) and some other are extra files. Another files present misleading values. A different problem is when the metainformation includes different spectral correction scans, which are not supported by the drEEM toolbox. Then, the drEEM original dataset seems to be inconsistent with the toolbox. In flavor of overcome all these handicaps the user can download and follow the EEMlab tutorial (see sect. 3.3) that illustrates and details how to use the GUI to process the drEEM dataset.

2.3.2. The EEMlab dataset

The EEMlab dataset is created with the aim of checking the performance of the PARAFAC deconvolution in a controlled environment. In this sense, a set of samples with known composition is manually designed and assembled in laboratory. Consequently, results coming from the PARAFAC analysis for this dataset are known in advance.

The EEMs of the samples are acquired with a QuantaMaster fluorometer in order to illustrate how EEMlab manages different formats of files. Then, there are two important features in the dataset to consider: (i) the TXT formatted samples and (ii) the automatic spectral correction made by the acquisition device. It becomes also important that absorbance scans for IFE correction be acquired in CSV format with a Hitachi spectrophotometer.

The dataset is composed by: 34 EEM files, 34 absorbance files, 11 blank files, 11 water Raman scans and the slope of a Quinine sulfate dilution series. The Raman normalization is

set in 350 nm excitation wavelength. All the files in the dataset are distributed into the folders' structure described above.

3. Experimental

The EEMlab GUI has been tested with the drEEM and EEMlab example datasets.

3.1. Experiments with the drEEM example dataset

The drEEM dataset is loaded in EEMlab when the experiment is properly configured in accordance to the acquisition devices (SPEX Fluorolog-3 Horiba fluorometer and Cary 4 Varian spectrophotometer) and the excitation Raman wavelength (275 nm). Next, if following the EEMlab tutorial, the user can easily overcome the problems reported in the loads of files about misleading information and samplelog mismatches. Once all the files are loaded, EEMlab can spectrally correct the dataset with no errors. The next steps are to preprocess the EEMs following the indications in the tutorial in order to get a clean dataset proper to be modeled. Then, the modeling and further model analysis is very easy to do making use of all the plotting tools integrated in the GUI. In sight of the modeling results, the number of underlying components is decided (six) and the final split half analysis is processed for the six-component refined model validation. In fact, the three split half tests are valid and, consequently the overall model it is too.

3.2 Experiments with the EEMlab example dataset

Reagent

All aqueous solutions were prepared with Milli-Q grade water. Ellagic, tannic, sinapic and syringic acids were purchased from sigma-Aldrich and used as received.

Analytical techniques

Solutions with a mixture of the four polyphenolics compounds (ellaginc, tannic, sinapic and syringic acids) at different concentrations (from 0.5 mg/L to 4 mg/L) were prepared. EEMs for the solutions and blanks were recorded using a modular QuantaMaster spectrofluorometer by subsequent emission scanning from 300 to 600 nm at 5 nm increments by varying the excitation wavelength from 250 to 550 nm at 5 nm increments. Raman scans at an excitation wavelength of 350 nm were also recorded as well as the Slope of a quinine sulfate dilution series. Finally, absorbance spectrum was obtained for each sample with a Hitachi spetrophotometer. The dataset was generated following the procedure described in [17]. EEMlab is used to load, correct, preprocess, explore, refine and validate models for the whole dataset.

4. Results and discussion

The discussion about the processing and results obtained from the PARAFAC modeling of the drEEM example dataset is absolutely justified in [33]. Unlike that it is reported in the original tutorial, in this case the EElab makes very easy the treatment of the many files involved in all the processing. To this respect, the only novelty in the experiment is the use of the GUI instead of programming the processing in script mode, directly in MATLAB language. Interestingly, when reproducing the experiment with EEMlab, we obtainthe same results than when processing the dataset directly with the drEEM toolbox. Then, we can validate the correctness of the functionality of the EEMlab application.

On the other hand, we also present a new (EEMlab) dataset to work with. In this case, the main challenge is to deal with files acquired with a non-standard fluorometer (QuantaMaster), that present TXT format and have been automatically corrected by the acquisition device. In addition, this dataset was designed to know a priori the number of components involved in the samples.

EEMlab example dataset

Once the dataset was loaded, corrected, preprocessed and normalized we start to explore the models, PARAFAC models with 3 to 6 components were computed. After the exploratory phase the EEMs were resized, excitation wavelength below 285 and above 400 and emission wavelength below 320 and above 470 were eliminated. The determination to resize the EEM in this range was assessed by evaluation of the distribution of residuals errors and the leverage (Fig. 2).

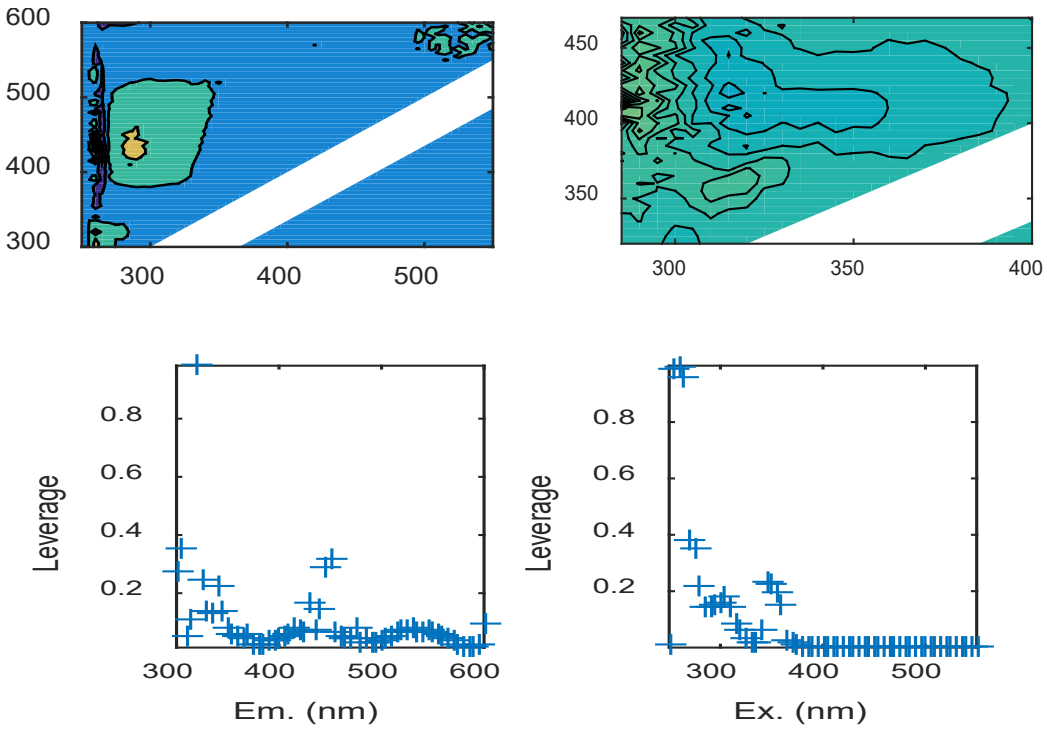

Fig. 2 (above right) Residual for a poorly modelled sample and (above left) a more adequately modelled sample (no. 1). Leverage plots indicate: (bellow right) emission wavelength with high influence especially near 315 nm; (bellow left) excitation wavelength with high influence especially near 250-270 nm.

The right number of PARAFAC component in the dataset was determined by the evaluation of the randomness of residuals (Figure 2 above) and the inspection of the physical sense of

the spectral loadings (Fig. 3 solid lines). The CONCORDIA index is not used due to the data set is small and core consistency is not always a reliable diagnostic of the number of PARAFAC components needed. It may provide too much protection against over-fitting and not enough against under-fitting [33].

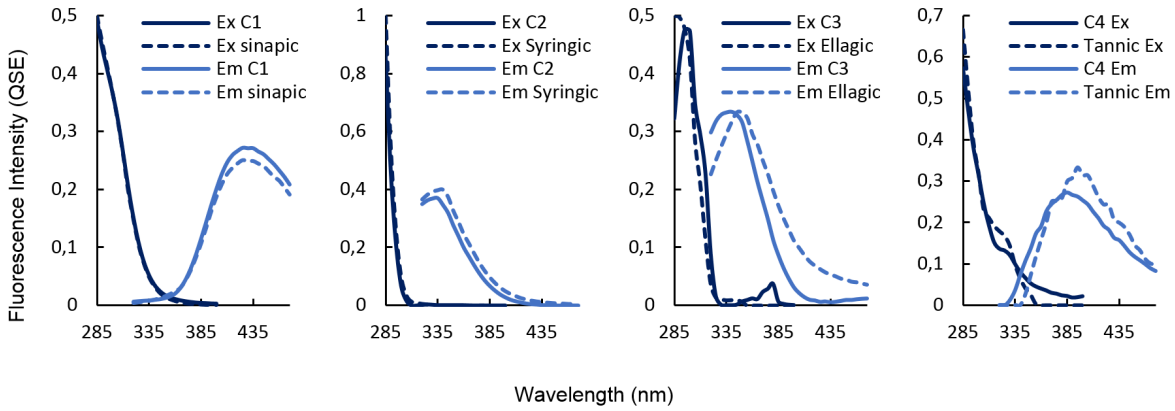

Fig. 3 Excitation and emission Spectra for the four-component PARAFAC model of the EEMlab dataset (solid line) and for the polyphenols compounds.

As in this case, the dataset comes from solutions with mixtures of known compounds after the model refinement, and we compare the fingerprint of the 4 PARAFAC components with the EEMs of the pure compounds (at 1 mg/L) (Fig. 4a y Fig. 4b respectively and the ex/em spectra for both PARAFAC components and pure compounds (Fig.3). As we can observe, similar fingerprints and ex/em spectra are obtained.

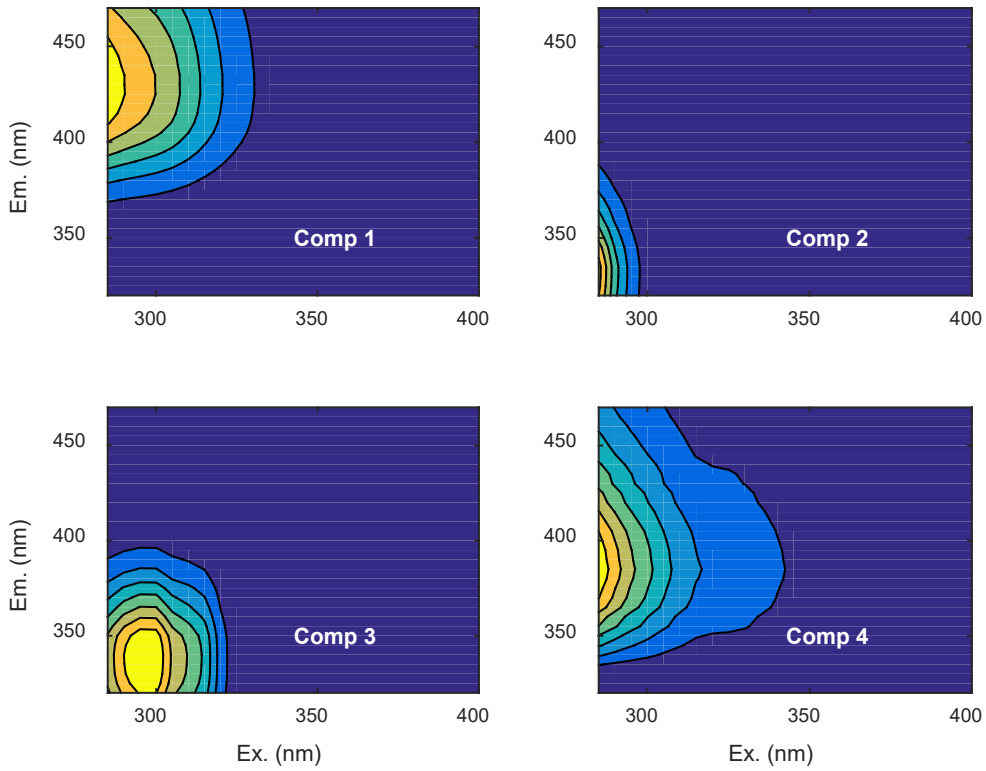

Fig. 4a Fingerprints obtained for the four-component PARAFAC model of the EEMlab dataset.

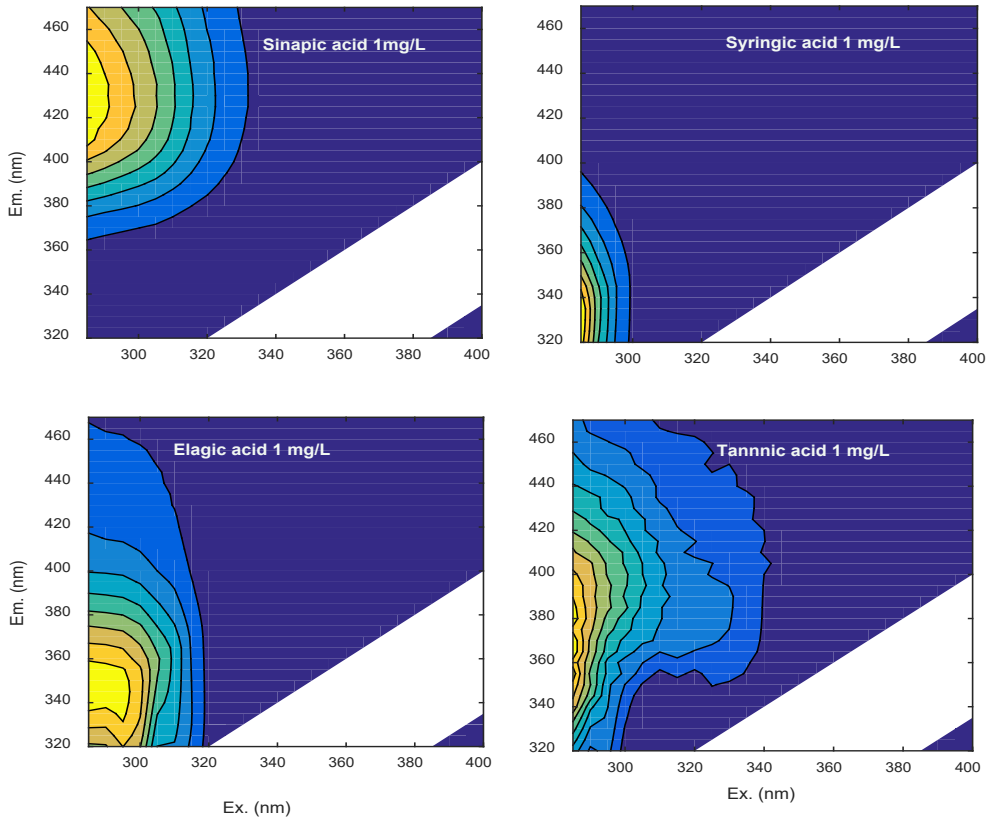

Fig. 4b Excitation Emission Matrices of the four compounds employed to create the EEMlab datasets at 1mg/L.

Finally, split-half analysis was used to confirm that 4-component PARAFAC model is appropriate for our dataset. The data set is split in 4 (A, B, C and D) and combined into 6 (AB, AC, AD, BC, BD, CD) different halves. One we have the splits they are modelled and validated (the Ex/Em spectra is compared between models, they are valid if the component are congruent). Finally, the model is validated for the all dataset comparing the Ex/Em spectra (Fig. 5), as the number of samples is a limiting condition we validate the model with the valid option for the splits (AC vs. BD) with a Tucker correlation coefficient of 0.9.

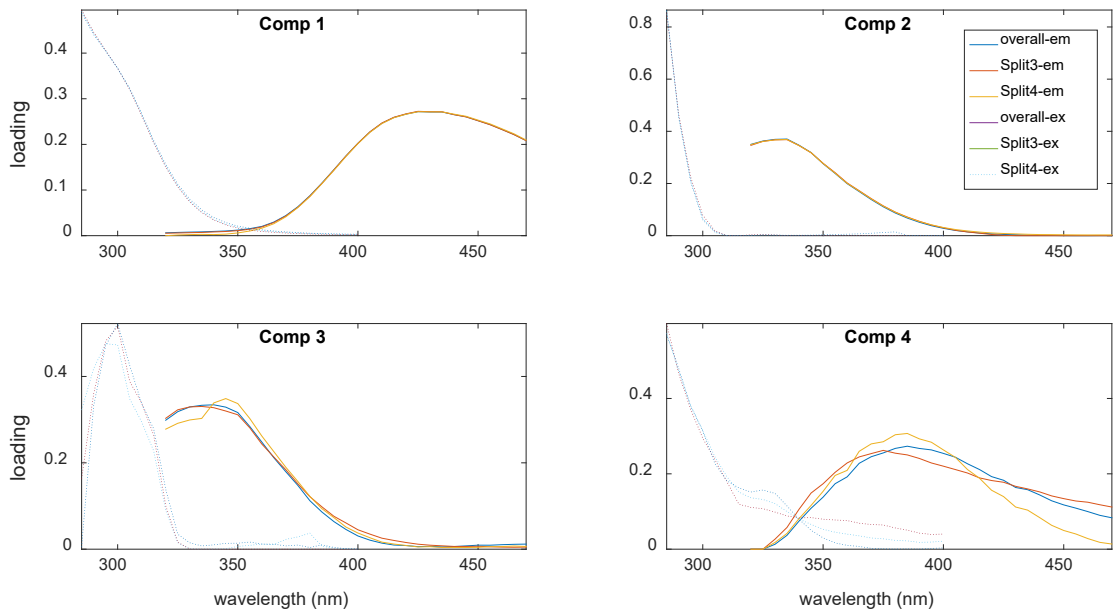

Fig. 5 Validation of the EEMlab dataset with two halves created randomly

5. Conclusions and future work

This work presents a graphical user interface based on the drEEM toolbox. The main objective of the application is to facilitate the use of the referred toolbox in fluorimetry trials. In this sense, the EEMlab includes not only all the functionality offered by drEEM but also the new functionality implemented for the load of files, the management of errors or the plot of the results. On the other hand, the EEMlab workflow has been designed in order to clarify all the experiment stages and with the aim to help the user in the complexity of all the processing. The results obtained from the processing of both example datasets also help to validate the use of EEMlab on conducting fluorescence experiments.

For all these reasons, the authors thing on EEMlab with the natural way to propagate a so powerful tool as the PARAFAC is. And maybe EEMlab can manage many other experiments different than fluorometric in where unrevealed sources are present.

In respect to the future of the EEMlab, the authors think on making the application compatible with the Openfluor web fluorescence spectra repository [39]. An interesting goal

could be the use EEMlab to analyze a dataset that could be automatically compared with all the individual components in the OpenFluor database, reporting a result about the real composition of the dataset. On the other hand, the EEMlab can also evolve to a web-app version in the same way of OpenFluor. In this manner, the use of the front-end application is detached from the Matlab framework and the application can be widely spread with the only requirement of having internet connectivity. Finally, the authors want to highlight the modularity in the design of the EEMlab that makes this application easy to evolve to enable, for example, the load and preprocessing of new raw formats (coming from any different acquisition device) or to meet with any other user requirements.

Independently tested by

The EEMlab software application has been independently tested by F.S. García Einschlag, adjoint professor in the Universidad Nacional de La Plata, Argentina. His review and comments are presented next:

“PARAFAC algorithm is increasingly used to decompose fluorescence excitation emission matrices (EEMs) into their underlying chemical components. However, the complexity of the mathematical concepts involved and the requirement a solid programming background still represents an important barrier to overcome. The recently developed drEMM toolbox has boosted the application of PARAFAC for the study of datasets associated with dissolved organic matter fluorescence. Due to the growing demand for easily available software capable of facilitating the application of the latter powerful tools the authors here present the EEMlab application, a graphical interface that combines several of the drEEM toolbox capabilities with very interesting new features that facilitate and automate many repetitive tasks usually required for both preprocessing and modeling of the fluorimetric datasets. EEMlab provides a user-friendly environment which assists the researcher in the creation of a preprocessed EEM dataset from the raw data files. The software runs according to the

features described in the documentation and leads the user through several tasks such as the importation of the required files, the application of correction factors (equipment spectral responses, inner-filter effects, blank subtraction, etc), the definition of wavelength ranges for analysis in both excitation and emission modes, the elimination of signal contributions associated to non-linear scattering effects and the removal of outliers. Moreover, with the EMlab GUI beginners are clearly directed through the essential steps required for the development, refinement and validation of PARAFAC models. Finally, it is important to note that the application also offers a wide range of options for the visualization and analysis of the results obtained"

Acknowledgements

- The authors want to thank the financial support of the Generalitat Valenciana, Conselleria d'Educació, Cultura i Esport [GV/2015/074]
- The authors want to thank the financial support of Spanish Ministerio de Educación y Ciencia (CTQ2015-69832-C4-4-R). Sara García-Ballesteros would like to thank Spanish Ministerio de Economía y Competitividad for her fellowship (BES-2013-066201)
- The authors want to thank Dr. F.S. García Einschlag who has independently tested the EEMlab and helped the authors to improve and validate the final version of the application

Figure Captions

Fig. 1. EEMlab workflow

Fig. 2 (above right) Residual for a poorly modelled simple and (above left) a more adequately modelled sample (no. 1). Leverage plots indicate: (bellow right) emission wavelength with

high influence especially near 315 nm; (bellow left) excitation wavelength with high influence especially near 250-270 nm

Fig. 3 Excitation and emission Spectra for the four-component PARAFAC model of the EEMlab dataset (solid line) and for the polyphenols compounds.

Fig. 4a Fingerprints obtained for the four-component PARAFAC model of the EEMlab dataset.

Fig. 4b Excitation Emission Matrices of the four compounds employed to create the EEMlab datasets at 1mg/L.

Fig. 5 Validation of the EEMlab dataset with two halves created randomly.

References

[1] J. Christensen, L. Nørgaard, R. Bro and S. Balling Engelsen, Multivariate autofluorescence of intact food systems, Chemical Reviews 106-6 (2006) 1979-1994.

[2] D. Airado-Rodriguez, I. Durán-Merás, T. Galeano-Díaz and J. Petter Wold, Front-fece fluorescence spectroscopy: a new tool for control in the wine industry, Journal of Food Composition and Analysis 24 (2011) 257-264.

[3] E. Guzmán, V. Baeten, J.A. Fernández and J.A. García-Mesa, Evaluation of the overall quality of olive oil using fluorescence spectroscopy, Food Chemistry 173 (2015) 927-934.

[4] R. Ríos-Reina, S. Elcoroaristizabal, J.A. Ocaña-González, D.L. García-González, J.M. Amigo, and R.M. Callejón, Characterization and authentication of Spanish PDO wine vinegars using multidimensional fluorescence and chemometrics, Food Chemistry 230 (2017) 108–116.

[5] L. Lenhardt, R. Bro, I. Zekovic, T. Dramicanin and M.D. Dramicanin, Fluorescence spectroscopy coupled with PARAFAC and PLS DA for characterization and classification of honey, Food Chemistry 175 (2015) 284–291.

[6] M. Sgroi, P. Roccaro, G.V. Korshin, F.G.A. Vagliasindi, Monitoring the Behavior of Emerging Contaminants in Wastewater-Impacted Rivers Based on the Use of Fluorescence Excitation Emission Matrixes (EEM), Environ. Sci. Technol., 51, 8 (2017) 4306–4316.

[7] E.M. Carstea, J. Bridgeman, A. Baker, D.M. Reynolds, Fluorescence spectroscopy for wastewater monitoring: A review, Water Research, 95 (2016) 205-219.

[8] B.F. Trueman, S.A. Macisaac, A.K. Stoddart, G.A. Gagnon, Prediction of disinfection by-product formation in drinking water via fluorescence spectroscopy, Environmental Science: Water Research & Technology, 2, 2 (2016) 383-389.

[9] A. Baker, R.G.M. Spencer. Characterization of dissolved organic matter from source to sea using fluorescence and absorbance spectroscopy, Sci. Total Environ. 333 (2004) 217–232.

[10] D. Lanciné Gonea, J.L. Seidelb, C. Batiot, K. Bamorya, R. Ligbana, J. Biemic, Using fluorescence spectroscopy EEM to evaluate the efficiency of organic matter removal during coagulation–flocculation of a tropical surface water (Agbo reservoir), J. of Hazardous Materials 172 (2009) 693–699.

[11] L. Yang, J. Hur and W. Zhuang, Occurrence and behaviors of fluorescence EEM-PARAFAC components in drinking water and wastewater treatment systems and their applications: a review, Environ Sci Pollut Res (2015) 22:6500–6510.

[12] S. Green, N. Blough, Optical absorption and fluorescence properties of chromophoric disolved organic matter in natural waters, Limnol. Oceanogr. (1994) *39*, 1903-1916.

[13] P. G. Coble. Characterization of marine and terrestrial DOM in seawater using excitation-emission matrix spectroscopy, Marine Chemistry 51 (1996) 325-346.

[14] W. Chen, P. Westerhoff, J.A. Leenheer, K. Booksh, Fluorescence Excitation - Emission Matrix Regional Integration to Quantify Spectra for Dissolved Organic Matter, Environ. Sci. Technol., 37 (2003) 5701-5710.

[15] D.M. McKnight, E.W. Boyer, P.K. Westerhoff, P.T. Doran, T. Kulbe, D.T. Andersen, Spectrofluorometric characterization of dissolved organic matter for indication of precursor organic material and aromaticity, Limnol. Oceanogr. (2001) 46, 38–48.

[16] M. Yan, O. Fu, D. Li, G. Gao, D. Wang, Study of the pH influence on the optical properties of dissolved organic matter using fluorescence excitation–emission matrix and parallel factor analysis, Journal of Luminescence, 142 (2013) 103-109.

[17] K. R. Murphy, K. D. Butler, R.G.M. Spencer, C.A. Stedmon, J. R. Boehme and G. R. Aiken, Measurement of dissolved organic matter fluorescence in aquatic environments: an interlaboratory comparison, Environ. Sci. Technol. 44 (2010) 9405–94.

[18] C. Goletz, M. Wagner, A. Grübel, W. Schmidt, N. Korf and P. Werner, Standardization of fluorescence excitation-emission-matrices in aquatic milieu, Talanta 85 (2011) 650-656.

[19] U. Resch-Renger, D. Pfeifer, C. Monte, W. Pilz, A. Hoffmann, M. Spieles, K. Rurack, J. Hollandt, D.Taubert, B.Schönenberger, P. Nording , The Calibration Kit Spectral Fluorescence Standards-A Simple and Certified Tool for the Standardization of the Spectral Characteristics of Fluorescence Instruments, J. Fluoresc. 15 (2005) 315-336.

[20] A. J. Lawaetz, C. A. Stedmon. Fluorescence Intensity Calibration Using the Raman Scatter Peak of Water, Applied Spectroscopy 63, 8 (2009) 936-940.

[21] T.D.Gauthier, E.C.Shane, W.F.Guerin, C.L.Seitz, C.L.Grant, Fluorescence quenching method for determining equilibrium constants for polycyclic aromatic hydrocarbons binding to dissolved humic materials. Environ. Sci. Technol. 20 (1986) 1162-1166.

[22] J.R. Lakowicz, Principles of Fluorescence Spectroscopy, Plenum Press, New York, 1983.

[23] T. Larsson, M. Wedborg, D. Turner, Correction of inner-filter effect in fluorescence excitation-emission matrix spectrometry using Raman scatter, Analytica Chimica Acta 583 (2007) 357–363.

[24] K.S. Booksh, B.R. Kowalski, Theory of analytical chemistry, Analytical Chemistry 66 (1994) 782A–791A.

[25] Kumar Mishra A., Analysis of dilute aqueous multifluorophoric mixtures using excitation–emission matrix fluorescence (EEMF) and total synchronous fluorescence (TSF) spectroscopy: A comparative evaluation, Talanta 117, 2013, 209-220.

[26] V. Gómez, M. P. Callao, Review: Analytical applications of second-order calibration methods, Analytica Chimica Acta 627 (2008) 169-183.

[27] T. Persson, M. Wedborg, Multivariate evaluation of the fluorescence of aquatic organic matter, Analytica Chimica Acta 434 (2001) 179–192.

[28] R. Bro. PARAFAC: Tutorial and applications, Chemometrics and Intelligent Laboratory Systems 38, 2, (1997) 149-171.

[29] C.M. Andersen, R. Bro, Practical aspects of PARAFAC modeling of fluorescence excitation- emission data, J. of Chemometrics, 17 (2003) 200-215.

[30] R. M. Maggio, A. Muñoz de la Peña, A. C. Olivieri, Unfolded partial least-squares with residual quadrilinearization: A new multivariate algorithm for processing five-way data achieving the second-order advantage. Application to fourth-order excitation-emission-kinetic-pH fluorescence analytical data, Chemometrics and Intelligent Laboratory Systems 109 (2011) 178–185.

[31] K. Kumar, A.K.Mishra, Application of multivariate curve resolution alternating least square (MCR–ALS) analysis to extract pure component synchronous fluorescence spectra at various wavelength offsets from total synchronous fluorescence spectroscopy (TSFS) data set

of dilute aqueous solutions of fluorophores, Chemometrics and Intelligent Laboratory Systems Volume 116 (2012) 78–86.

[32] T.Azzouz, R.Tauler, Application of multivariate curve resolution alternating least squares (MCR-ALS) to the quantitative analysis of pharmaceutical and agricultural samples, Talanta 74 (2008) 1201–1210.

[33] K.R. Murphy, C.A. Stedmon, D. Graeber, R. Bro, Fluorescence spectroscopy and multi-way techniques. PARAFAC, Analytical Methods, 5 (2013) 6557-6566.

[34] C.A. Stedmon, R. Bro, Characterizing dissolved organic matter fluorescence with paralel factor analysis: a tutorial, Limnol. Oceanogr. Methods, 6 (2008) 572-579.

[35] K.R. Murphy, J.R. Boheme, C. Brown, M. Noble, G. Smith, D. Sparks, G.M. Ruiz, Exploring the limits of dissolved organic matter fluorescence for determining seawater sources and ballast water exchange on the US Pacific coast, J. Mar. Syst., 111-112 (2013) 157-166.

[36] K.R. Murphy, C.A. Stedmon, D. Graeber, R. Bro, The drEEM toolbox for MATLAB, (2013) online http://www.models.life.ku.dk/dreem (accessed 01.03.2017).

[37] R.A. Harshman, Research methods for multimode data analysis, ed. H.G. Law, J.C.W. Snyder, J. Hattie and R.P. McDonald, Praeger, New York, 1984, pp. 602-642.

[38] R. Bro, M. Vidal, EEMizer: Automated modeling of fluorescence EEM data, Chemometrics and Intelligent Laboratory Systems, 106 (2011) 86-92.

[39] K.R. Murphy, C.A. Stedmon, P. Wenig, R. Bro, OpenFluor, an online specral library of auto-fluorescence by organic compounds in the environment, Anal. Methods, 6 (2014) 658 - 661.

[40] R.A. Harshman, in Research Methods for Multimode Data Analysis, ed. H.G. Law, J.C.W. Snyder, J. Hattie and R.P. McDonald, Praeger, New York, 1984, 566-591.