



Analysis of raw EEM fluorescence spectra - ICA and PARAFAC capabilities

Jorge Costa Pereira*, Alberto A.C.C. Pais, Hugh D. Burrows

CQC, Department of Chemistry, University of Coimbra, Coimbra P-3004 535, Portugal

ARTICLE INFO

Article history:

Received 12 April 2018

Received in revised form 3 July 2018

Accepted 8 July 2018

Available online 11 July 2018

Keywords:

EEM

Fluorescence spectra

PARAFAC

ICA

Performance comparison

Semi-empirical simulations

ABSTRACT

Excitation-Emission fluorescence spectroscopy is a versatile technique and has been used to detect, characterize and quantify residual Dissolved Organic Matter (DOM) in aquatic domains. PARALLEL FACTOR ANALYSIS (PARAFAC) has been extensively used in the analysis of excitation-emission matrices (EEM), allowing for a better identification and quantification of contributions resulting from spectral decomposition.

In this work we have adapted Independent Component Analysis (ICA) in order to make it suitable to the analysis of three-way EEM datasets, and tested ICA and PARAFAC performances for the study of three available datasets (Claus, Dorrit and drEEM). Semi-empirical simulation allowed us to assess the impact of (a) sample size, (b) signal sources and (c) composition dependencies, and the presence of (d) unspecific signal contributions (e.g. light scattering) upon both algorithms.

PARAFAC and ICA have similar performances in processing ideal three-way EEM datasets but, in the presence of non-trilinear responses, ICA leads to a more realistic approach, yielding a better decomposition of contributing sources and their identification and quantification. This makes this algorithm more suitable for the analysis of real, raw EEM data, without the need of preprocessing to remove any unspecific contributions.

© 2018 Elsevier B.V. All rights reserved.

1. Introduction

Fluorescence is a very powerful technique, able to detect a large range of residual organic material dissolved in water [1]. This has the major advantages of low analytical costs, very rapid analysis, very high sensitivity and large linear analytical range, which makes it very convenient for the characterization of residual organic matter dissolved in aquatic environments [2].

This technique is versatile, since it allows the study and characterization of samples using different types of spectra such as excitation, where fluorescence is measured at a fixed wavelength while excitation range is scanned, emission, where a given excitation wavelength is used and the respective system emission measured, synchronous fluorescence, with simultaneous scan of excitation and emission wavelengths keeping the wavelength difference constant. In this work, sample information details are enhanced by recording Excitation-Emission matrices (EEM) to produce a fluorescence surface. The decomposition of these EEM datasets allows to characterize and quantify dissolved organic matter (DOM) present in environmental samples and other matrices [2,3]. Advances in electronics and

signal processing abilities are leading to standard use of, other techniques, such as time-resolved fluorescence [1], which can also use the two numerical techniques focused here.

1.1. EEM Analysis

1.1.1. PARAFAC

PARALLEL FACTOR ANALYSIS (PARAFAC) is a very powerful multivariate data analysis method suited for the decomposition of 3-way and 4-way multivariate systems into lower dimensional matrices [4,5]. Because of its ability to directly extract information with chemical meaning [5], PARAFAC is nowadays considered a standard algorithm to process EEM data [6]. In the case of tridimensional EEM datasets, PARAFAC performs a three-way tensor decomposition

$$\mathbf{EEM} = \mathbf{A} \otimes \mathbf{B} \otimes \mathbf{C} + \mathbf{U} \quad (1)$$

In case the first array dimension of **EEM** is related to the sample identification, matrix **A** contains sample scores and **B** and **C** the respective loadings, related with excitation and emission spectra. **U** represents the unjustified residual information of the initial dataset.

Solving Eq. (1) requires an alternating least-squares approach (ALS), which permits the imposition of various restrictions [5]. In the case of EEM, non-negativity provides PARAFAC with the ability

* Corresponding author.

E-mail address: jcpereira@qui.uc.pt (J. Costa Pereira).

to directly retrieve information with chemical meaning. The main drawback with PARAFAC is related with its difficulty in dealing with non-trilinear information such as light scattering phenomena (Rayleigh and Raman) [5,7], although several approaches have been described in the literature in order to compensate for this deficiency [7–9].

1.1.2. Independent Component Analysis

Independent Component Analysis (ICA) is based on the Blind Source Separation (BSS) algorithm [10], and has been developed to deconvolute mixed signals by maximizing their respective independence [11]. Measured signals (\mathbf{X}) may be decomposed into signal sources, representing specific individual signal contributions, and respective weights, using the following representation [12]

$$\mathbf{X} = \mathbf{S}\mathbf{A} + \mathbf{V} \quad (2)$$

where \mathbf{S} and \mathbf{A} stand for the loadings of the signal sources, and respective mixing information matrix, the scores, respectively. Similarly to Eq. (1), \mathbf{V} matrix corresponds to the unjustified residual information.

In previous work, ICA signal deconvolution [13–15] was shown to be a very powerful method for dealing with spectral information, allowing the identification and quantification of the contributing species present in different systems. However, the ICA algorithm requires \mathbf{X} to be a bidimensional (two-way) dataset. In order to process EEM information, ICA imposes a pre- and post-processing array unfolding and folding back i.e., an array reshape. In the absence of further restrictions, ICA solutions have to be carefully assessed before converting to chemical information. Data simulation and the use of adequate blind samples may be advised in order to familiarize with ICA [13,14].

In this study, we aim to establish a parallel between PARAFAC and ICA to compare the methods and their modeling performances,

and establish the correspondence between Eqs. (1) and 2 in terms of loadings and scores.

2. Procedures

2.1. Datasets

Selected datasets are all representative of EEM information. The first two, Claus and Dorrit, correspond to very simple systems, synthetic mixtures, simulated in the lab, in which the number of components and respective composition are known and are used as blind samples. In order to better evaluate the two algorithms, extra datasets are simulated using semi-empirical and other simulations.

2.1.1. Claus

The Claus dataset [5] consists of five simple laboratory-made samples ($I = 5$). Each sample contains different amounts of tryptophan, tyrosine and phenylalanine ($k^0 = 3$) dissolved in phosphate buffered water. The samples were measured by fluorescence (excitation 240–300 nm, emission 250–450 nm, 1 nm intervals) on a PE LS50B spectrofluorometer with excitation slit-width of 2.5 nm, an emission slit-width of 10 nm and a scan-speed of 1500 nm/s. This dataset is available online [16].

Fig. 1 represents all samples ($I = 5$) contained in the Claus dataset, with the respective mixture composition detailed in Table 1.

2.1.2. Dorrit

The Dorrit dataset, EEM(I,J,Q) [7], consists of $I = 27$ synthetic samples containing different concentrations of four analytes (hydroquinone, tryptophan, phenylalanine and DOPA) ($k^0 = 4$) measured on a Perkin-Elmer LS50 B fluorescence spectrometer. Each fluorescence landscape corresponding to an individual sample consists of $J = 121$ emission wavelengths (241–481 nm) and $Q = 24$ excitation wavelengths (200–315 nm taken every 5 nm). According to the

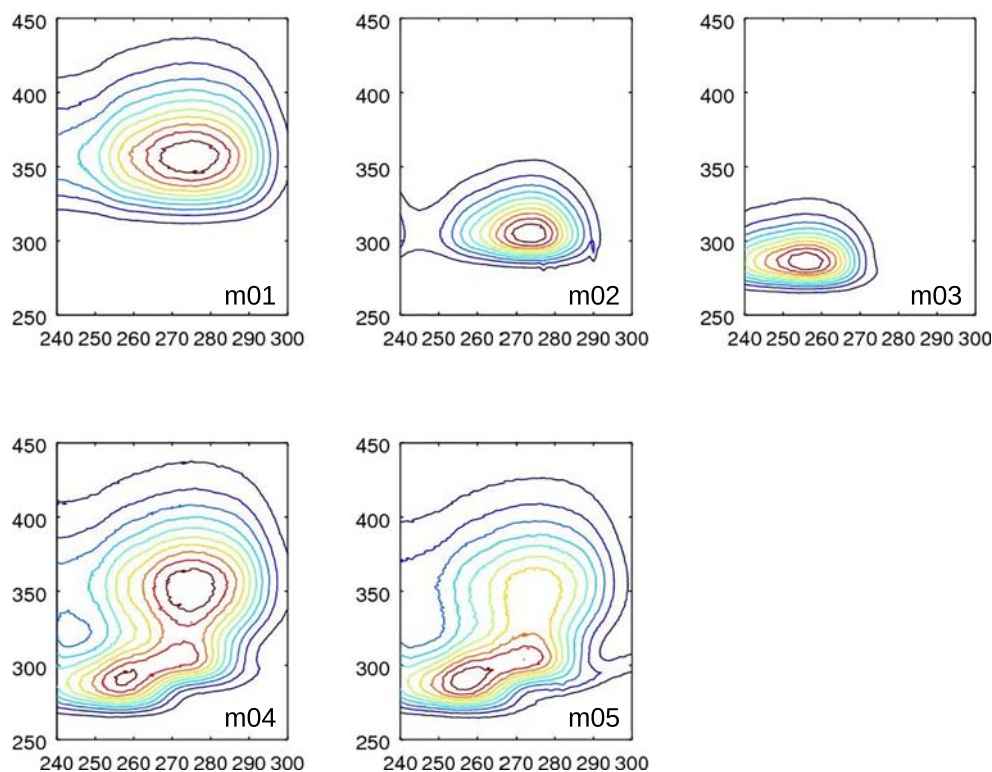


Fig. 1. Representation of excitation-emission fluorescence matrix (EEM) contour plots of samples ($I = 5$) contained in Claus dataset; horizontal scale refers to excitation (240–300 nm) and vertical scale to emission (250–450 nm). Higher fluorescence intensities in red.

Download English Version:

<https://daneshyari.com/en/article/7667559>

Download Persian Version:

<https://daneshyari.com/article/7667559>

[Daneshyari.com](https://daneshyari.com)