



# Quantification of pH-dependent speciation of organic compounds with spectroscopy and chemometrics



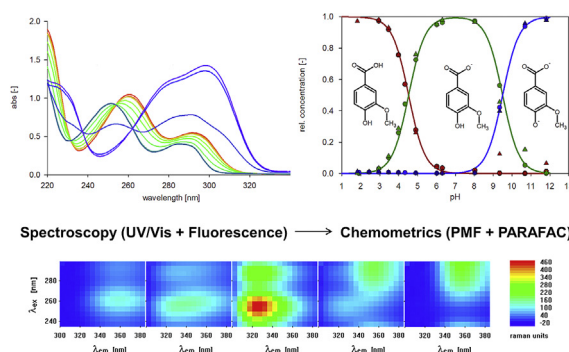
Thomas Ritschel\*, Kai Uwe Totsche

Department of Hydrogeology, Institute for Geosciences, Friedrich-Schiller-University Jena, Burgweg 11, 07749 Jena, Germany

## HIGHLIGHTS

- Speciation of organic acids was reconstructed from UV/Vis and fluorescence spectra.
- PMF and PARAFAC also permit quantification in multi-component solutions.
- Fixed components were used to reconstruct the pH from speciation of organic acids.

## GRAPHICAL ABSTRACT



## ARTICLE INFO

### Article history:

Received 10 November 2016  
 Received in revised form  
 23 December 2016  
 Accepted 28 December 2016  
 Available online 30 December 2016

Handling Editor: I. Cousins

### Keywords:

Parallel factor analysis (PARAFAC)  
 Positive matrix factorization (PMF)  
 Organic acid  
 Environmental tracer  
 Number of components

## ABSTRACT

Fluorescence and UV/Vis spectra of aqueous solutions with numerous organic compounds are a superposition of single spectra of the chemical species present. Thus, an isolation of individual spectra with chemometrics is required for their quantification. We investigated UV/Vis spectra and fluorescence excitation-emission matrices of vanillic acid, salicylic acid, phenoxyacetic acid and phthalic acid with positive matrix factorization (PMF) and non-negativity constrained parallel factor analysis (PARAFAC) in combination with the law of mass action. In consideration of the pH-dependent speciation of organic acids, we first reconstructed the pH-specific spectra of each compound. Using these spectra as known components in a constrained algorithm, we could successfully quantify species of multiple compounds and reconstruct the solution pH. In addition, we estimated the uncertainty of reconstructed spectra and concentrations in order to assess the most probable number of components for PMF/PARAFAC. Therefore, we could derive a framework to reconstruct the number of relevant species and their individual concentration present in spectroscopic data of aqueous solutions containing multiple organic compounds.

© 2016 Elsevier Ltd. All rights reserved.

## 1. Introduction

The understanding of the fate of dissolved organic compounds is a crucial prerequisite for the assessment of soil functions and

processes (e.g. Kalbitz et al., 2000). For this reason, transport experiments with a broad spectrum of organic tracers are commonly conducted at the field and laboratory scale in environmental sciences. However, an unambiguous reconstruction of processes in natural porous media requires a long-term, yet temporally highly resolved, screening of seepage or effluent water with respect to the tracer concentration (e.g. Weigand et al., 2002). Spectroscopic methods such as photometry and fluorimetry are well suited for

\* Corresponding author.

E-mail address: [Thomas.Ritschel@uni-jena.de](mailto:Thomas.Ritschel@uni-jena.de) (T. Ritschel).

this purpose since they allow for the rapid collection of data on organic compounds in aqueous solutions without sample preparation or sample consumption. As a consequence, fluorescent dyes are frequently used as tracers in hydrology, hydrogeology and soil science (e.g. Huang et al., 2002; Kasnavia et al., 1999) or as pH probes in biology (e.g. Mao et al., 2015). However, the corresponding measurement of absorption or fluorescence emission is frequently conducted at single wavelengths, which do not reveal the full information potential that can be obtained from absorption and emission spectra. In addition, each compound of a solution contributes linearly to the total spectrum according to Lambert Beer's Law and thus, the spectra of solutions containing multiple fluorophores or chromophores represent a superposition of all individual compound spectra. Therefore, a single-wavelength or spectral region integral quantification for aqueous solutions containing multiple compounds is susceptible to misinterpretation. Yet, even for single compound solutions, the spectrum may alter due to the changes in the chemical milieu of the solution. For this reason, absorption (e.g. Flexser et al., 1935) and fluorescence spectroscopy (e.g. Rosenberg et al., 1979; Wolfbeis et al., 1983; Gningue and Aaron, 1985) has been used for a long time to determine  $pK_a$  values of acid-base reactions. The methodology applied in this context is diverse and under constant development (e.g. Kara and Alkan, 2000; Meloun et al., 2007; Zevatskiy et al., 2015).

Consequently, when the extraction of full quantitative information from multi-component mixtures is desired in environmental studies, the entire spectra have to be considered for the clean separation of all individual compounds including their species. To carry out this decomposition without any a priori knowledge on the compound spectra, methods from chemometrics have to be applied (e.g. Geladi, 2003; Geladi et al., 2004). In the form of parallel factor analysis (PARAFAC), this is already frequently applied to fluorescence excitation-emission matrices (EEM), e.g., to characterize dissolved organic matter (DOM) (Bro, 1997; Fellman et al., 2009). Studies involving UV/Vis spectra of DOM (e.g. Gao et al., 2015; Dryer et al., 2008) sometimes rely on application of spectral subtraction, which might not reveal as much information as factor analysis. Therefore, it is highly desirable to derive a physically meaningful model for absorption spectra that is as predictive as PARAFAC is for fluorescence EEM. However, the two-dimensional analogue of PARAFAC suffers from non-uniqueness of factor loadings due to rotational freedom, which also renders the reconstructed factors arbitrary if no additional constraints are given (Paatero et al., 2002). Yet, with the theoretical knowledge on the pH-dependent speciation of chemicals, a meaningful reconstruction of absorption spectra is possible by rotating PCA reconstructions until an agreement of the factorization and the law of mass action is achieved, as shown by Kubista et al. (1993) for fluorescein and benzoic acid. The same authors have demonstrated that the reconstruction also allows for a quantification (Kubista et al., 1995) and that an application to fluorescence spectra of fluorescein is possible (Sjöback et al., 1995).

With this paper, we investigate UV/Vis spectra and fluorescence EEM of selected organic substances with potential use as reactive tracers in environmental sciences, i.e. vanillic acid, salicylic acid, phenoxyacetic acid and phthalic acid. Measurements were conducted at different pH values in order to derive explanatory models that are capable of also quantifying the pH-dependent speciation of these substances. We approached the problem of rotational freedom by imposing additional constraints based on chemical considerations. In addition to a comparison to the law of mass action as done by Kubista et al. (1993), we used fixed components in order to create a model that is also predictive for known components in more complex samples and that can be used to reconstruct

the pH value. The reliability and robustness of the models is analyzed and possible applications are discussed.

## 2. Material and methods

### 2.1. Matrix factorization

For data  $\mathbf{X}$  ( $I \times J$ ) consisting of  $I$  samples and  $J$  variables, a two-way factorization with  $F$  factors is given as

$$x_{ij} = \sum_{f=1}^F a_{if}b_{jf} + e_{ij} \quad \text{with } i = 1 \dots I, j = 1 \dots J \quad (1)$$

$$\mathbf{X} = \mathbf{A}\mathbf{B}' + \mathbf{E} \quad (2)$$

Here, small letters denote scalars, bold letters denote vectors and bold capitals denote matrices. The same letters refer to the same entity. So,  $x_{ij}$  is the scalar at row  $i$  and column  $j$  in the matrix  $\mathbf{X}$ .

The matrices  $\mathbf{A}$  ( $I \times F$ ) and  $\mathbf{B}$  ( $J \times F$ ) are referred to as scores matrix and loadings matrix, respectively (Geladi, 2003). Here,  $\mathbf{A}$  is the matrix of concentrations of each component in the samples and  $\mathbf{B}$  is the matrix of the pure spectra of the basic components. The matrix  $\mathbf{E}$  ( $I \times J$ ) carries the residuals of  $\mathbf{X}$  which are not explained by  $\mathbf{A}\mathbf{B}'$ . The subscripted vectors  $\mathbf{x}_i$  and  $\mathbf{a}_i$  denote the  $i$ -th row and  $\mathbf{x}_j$  and  $\mathbf{b}_j$  the  $j$ -th column of  $\mathbf{X}$  and  $\mathbf{A}$  or  $\mathbf{B}$ . A row-wise or column-wise formulation of Equation (2) is therefore

$$\mathbf{x}_i = \mathbf{a}_i\mathbf{B}' + \mathbf{e}_i \quad (3)$$

$$\mathbf{x}_j = \mathbf{A}\mathbf{b}_j' + \mathbf{e}_j \quad (4)$$

Factor analysis aims at finding the set of scores  $\mathbf{A}$  and loadings  $\mathbf{B}$  which minimizes  $\mathbf{E}$  and therefore provides the best fit of the observed data  $\mathbf{X}$ . The algorithm to solve for  $\mathbf{A}$  and  $\mathbf{B}$  is based on alternating least squares (ALS), which divides the solution to Equation (2) into sub-problems according to

$$\mathbf{A} = \mathbf{X}\mathbf{B}^{\dagger} = \mathbf{X}\mathbf{B}(\mathbf{B}'\mathbf{B})^{-1} \quad (5)$$

$$\mathbf{B} = \mathbf{X}'\mathbf{A}^{\dagger} = \mathbf{X}'\mathbf{A}(\mathbf{A}'\mathbf{A})^{-1}, \quad (6)$$

in which  $\dagger$  denotes the Moore-Penrose pseudoinverse. ALS algorithms are commonly used in factor analysis and can be easily extended to higher order factorizations (Bro, 1997). Once again, the solution can be calculated row-wise or column-wise with respect to  $\mathbf{X}$ .

$$\mathbf{a}_i = \mathbf{x}_i\mathbf{B}^{\dagger} = \mathbf{x}_i\mathbf{B}(\mathbf{B}'\mathbf{B})^{-1} \quad (7)$$

$$\mathbf{b}_j = \mathbf{x}_j'\mathbf{A}^{\dagger} = \mathbf{x}_j'\mathbf{A}(\mathbf{A}'\mathbf{A})^{-1} \quad (8)$$

Bro and De Jong (1997) or Kim and Park (2008) describe algorithms to subject Equation (7) and Equation (8) to a non-negativity constraint with an active set method. In our study we based all calculations on these algorithms to comply with the knowledge on the positive values of spectra as well as concentrations. Thus, our general approach is classified as positive matrix factorization (PMF) as described in detail by Paatero and Tapper (1994). The processing of fluorescence EEM was carried out with PARAFAC, the three-dimensional analogue of previous equations (Harshman, 1970), also constrained by non-negativity (Bro and De Jong, 1997). As for

Download English Version:

<https://daneshyari.com/en/article/5747066>

Download Persian Version:

<https://daneshyari.com/article/5747066>

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