Ariel

Rapid #: -4150181 **IP: 128.138.154.145**

CALL #: http://www.libproxy.wvu.edu/login?url=http://aslo.org/lometh ...

LOCATION: **WVU :: Electronic :: EJournal**

TYPE: Article CC:CCG

JOURNAL TITLE: Limnology and oceanography, methods

USER JOURNAL TITLE: LIMNOLOGY AND OCEANOGRAPHY-METHODS

WVU CATALOG TITLE: Limnology and oceanography, methods

ARTICLE TITLE: Characterizing dissolved organic matter fluorescence with parallel factor analysis: a tutorial

ARTICLE AUTHOR: Stedmon,

VOLUME: 6

ISSUE:

MONTH:

YEAR: 2008

PAGES: 572-579

ISSN: 1541-5856

OCLC #:

CROSS REFERENCE ID: [TN:380935][ODYSSEY:128.138.154.194/ILL]

VERIFIED:

BORROWER: **COD :: Main Library**

PATRON: **Rachel Gabor**

PATRON ID: gabor

PATRON ADDRESS:

PATRON PHONE:

PATRON FAX:

PATRON E-MAIL:

PATRON DEPT:

PATRON STATUS:

PATRON NOTES:

RAPID ILL

This material may be protected by copyright law (Title 17 U.S. Code)
System Date/Time: 2/17/2011 11:24:40 AM MST

LIMNOLOGY
and
OCEANOGRAPHY: METHODS

Limnol. Oceanogr.: Methods 6, 2008, 572–579
© 2008, by the American Society of Limnology and Oceanography, Inc.

Characterizing dissolved organic matter fluorescence with parallel factor analysis: a tutorial

Colin A. Stedmon[1]* *and Rasmus Bro*[2]

[1]Department of Marine Ecology, National Environmental Research Institute, Aarhus University, Frederiksborgvej 399, Roskilde, Denmark

[2]Dept. Food Science, Faculty of Life Sciences, University Copenhagen, Rolighedsvej 30, DK-1958, Frederiksberg, Denmark

Abstract

A sub-fraction of dissolved organic matter fluoresces when excited with ultraviolet light. This property is used to quantify and characterize changes in dissolved organic matter (DOM) in aquatic environments. Detailed mapping of the fluorescence properties of DOM produces excitation emission matrices (EEM), which are well suited to multi-way data analysis techniques (chemometrics). Techniques such as parallel factor analysis (PARAFAC) are increasingly being applied to characterize DOM fluorescence properties. Here, an introduction to the technique and description of the advantages and pitfalls of its application to DOM fluorescence is presented. Additionally a MATLAB based tutorial and toolbox specific to PARAFAC analysis of DOM fluorescence is presented.

Introduction

Fluorescence spectroscopy is a sensitive technique often used for tracing the dynamics of dissolved organic matter (DOM) in marine and freshwaters. When irradiated by UV and blue light, a sub-fraction of the DOM pool fluoresces. The concentration and chemical composition of DOM influences the intensity and shape of the fluorescence spectra. In early studies, DOM fluorescence was used for following the supply of terrestrial organic matter via rivers into coastal waters (Kalle 1966). More recently, studies have shown how different fractions of the DOM pool can be distinguished using detailed spectral fluorescence measurements (Coble et al. 1990; Coble 1996). Over the last decade, fluorescence excitation emission matrix (EEM) spectroscopy has been successfully applied for the identification of terrestrial, marine, and anthropogenic components of DOM (Coble 1996; Baker 2001; Stedmon et al. 2003; Cory and McKnight 2005). EEMs are obtained by combining fluorescence (emission) spectra measured from a series of different excitation wavelengths. In general, DOM fluorescence has a broad excitation between 250 and 400 nm and a broad emission from 350 to 500 nm. The location of the excitation and emission peaks varies with the composition of DOM.

EEMs provide a wealth of information about DOM, which in itself, can be very difficult to interpret. Until recently, the techniques for characterizing EEM's have generally relied on visual identification of peaks and ratios of fluorescence in different regions of the spectrum (e.g., Coble 1996; McKnight et al. 2001). Recently multivariate data analysis techniques have been applied to the study of DOM's complex mixture of fluorescence signals (Persson and Wedborg 2001; Stedmon et al. 2003; Boehme et al. 2004; Hall et al. 2007). An excellent example is parallel factor analysis (PARAFAC) which can decompose the fluorescence signal into underlying individual fluorescent phenomena (Bro 1997). This is a valuable tool for characterizing and quantifying changes in DOM fluorescence enabling the tracing of different fractions in the natural environment (Cory and McKnight 2005; Hall et al. 2005; Stedmon and Markager 2005a,b; Murphy et al. 2006).

Principal Component Analysis (PCA) is another commonly used technique for modeling and visualizing complex, multivariate data (Martens and Næs 1989). However, with regards to analyzing EEMs, it has significant limitations compared with PARAFAC. PCA provides only a qualitative characterization of the data and the model is expressed as principal components, which are largely abstract mathematical entities and

*Corresponding author: E-mail: cst@dmu.dk; tel: + 45 46301805

Acknowledgments

This work was supported by the Danish Research Council (grant nr 274-05-0064 (CS)) and Carlsberg Foundation (CS). Samples were collected via the Danish National Marine Monitoring program funded by the Danish Ministry of the Environment. Charlotte Ørsted Andersen and the crew and laboratory technicians onboard R/V Gunnar Thorson are thanked for their assistance in sample collection and analysis. Sally Walker is thanked for her comments on the manuscript and tutorial. Members of the NERC U.K FluoroNet network are thanked for their comments on the tutorial.

contain little chemical information. In contrast, PARAFAC provides both a quantitative and qualitative model of the data and separates the complex signal measured into its individual underlying fluorescent phenomena with specific excitation and emission spectra.

The aim of the paper is to present a tutorial for the analysis of DOM EEMs using parallel factor analysis and discuss some aspects of the approach that need to be considered. Although the discussion is focused on the application of the technique to DOM fluorescence, elements are also relevant for other fluorescence applications. A series of papers on different aspects of applying PARAFAC to fluorescence data have been published (Bro 1997, 1999; Andersen and Bro 2003), however the majority of these deal with simple mixtures or specific complex analytes in very different scientific fields, such as food science. Furthermore, these prior papers have mainly been devoted to overall methodological descriptions, lacking the detail needed for the reader to actually perform practical data analysis. The complexity of DOM fluorescence warrants special attention and indications of obvious pitfalls and gains apparent from the last 5 y of modeling DOM fluorescence. A critical look at the current state and development of a common approach is now required.

Multi-way data analysis—The term multi-way is used to describe data with more than two dimensions, for example, where data can be expressed as a box with three dimensions (e.g., sample by "property 1" by "property 2"). Spectral fluorescence data are multi-way (three-way) as the fluorescence of a sample varies depending on the wavelength of light absorbed (excitation) and the wavelength at which fluorescence is observed (emission). Combining the data from a series of samples results in a three-way box of data (Fig. 1).

PARAFAC models three-way data using Eq. 1, fitting the equation by minimizing the sum of squares of the residuals (ε_{ijk}).

$$x_{ijk} = \sum_{f=1}^{F} a_{if} b_{jf} c_{kf} + \varepsilon_{ijk}, \; i = 1,..,I; \; j = 1,..,J; \; k = 1,..,K; \qquad (1)$$

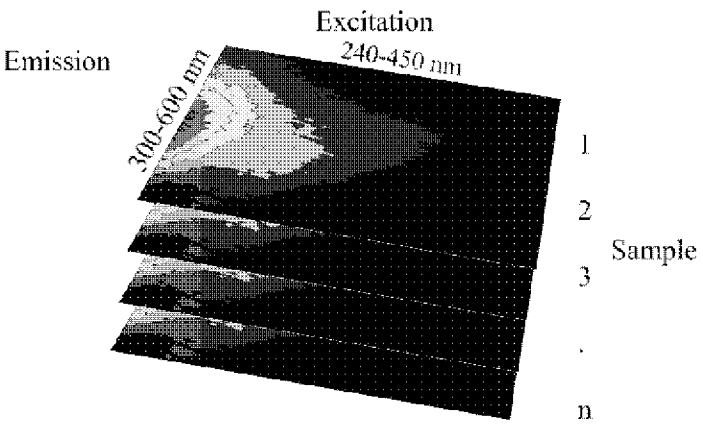

Fig. 1. Example of combining the EEMs into a three-dimensional box of data (x_{ijk} in Eq. 1)

x_{ijk} is one element of the three-way data array with dimensions I, J, and K. In the analysis of EEMs, the number x_{ijk} is the fluorescence intensity of sample i measured at emission wavelength j and excitation wavelength k. The final term ε_{ijk} represents the unexplained signal (residuals containing noise and other un-modeled variation). The outcome of the model are the parameters a, b, and c. Ideally, these respectively represent the concentration, emission spectra, and excitation spectra of the underlying fluorophores. Equation 1 is identical to that of fluorescence of a mixture of fluorophores, assuming that they behave according to Beer-Lamberts law and that there are no interactions between them. If there are F fluorophores in the mixture, the measured signal is the sum of the contribution from each. The fluorescence characteristics (shape) of each fluorophore do not change except for its fluorescence intensity, which is dependent on its concentration in the respective sample. This means that the wavelength position of the fluorescence peaks of each fluorophore do not "shift," however, the fluorescence maxima of the mixture will shift depending on the relative contribution (concentration) of each of the fluorophores.

This application of Beer-Lamberts law in Eq. 1 can be further illustrated by imagining that one EEM has been measured, with J emission and K excitation wavelengths. This data are a table of fluorescence intensities of size $J \times K$, which we can refer to as matrix X. If there is only one fluorophore present in the sample, then the EEM is the product of the emission (the vector b_1) and excitation (the vector c_1) spectra. The matrix $Z_1 = b_1 c_1^T$ is thus the EEM of the fluorophore at unit concentration (Fig. 2). If the concentration of this fluorophore is a_1 then the EEM can be described as $x = a_1 Z_1$. It then follows that if there are more than one fluorophore and Beer-Lamberts law is valid, then the measured EEM is the sum of contributions of all F fluorophores each weighted by the respective concentration (a_f)

$$X = \sum_{f=1}^{F} a_f Z_f \qquad (2)$$

For each element of X (x_{jk}), this can be written

$$x_{jk} = \sum_{f=1}^{F} a_f b_{jf} c_{kf} \qquad (3)$$

which is Eq.1 but where there is only one sample ($i = I$).

As with many analytical and statistical approaches, there are some fundamental assumptions that are necessary for the application of PARAFAC to DOM fluorescence. One is that a change in analyte concentration results only in a change in its fluorescence intensity and not in the shape of the excitation (c) and emission (b) spectra. Alterations to the spectral fluorescence of a fluorophore as a result of inner filter effects, changes in the local environment of the analyte (temperature, metal concentrations, proximity to, etc.), or intra-molecular processes are assumed to be minimal. Although some factors, such as inner filter effects and effects due to variable temperature, can be eliminated or at least minimized, it is very difficult if not impossible to eliminate the effects of all factors. However, this still

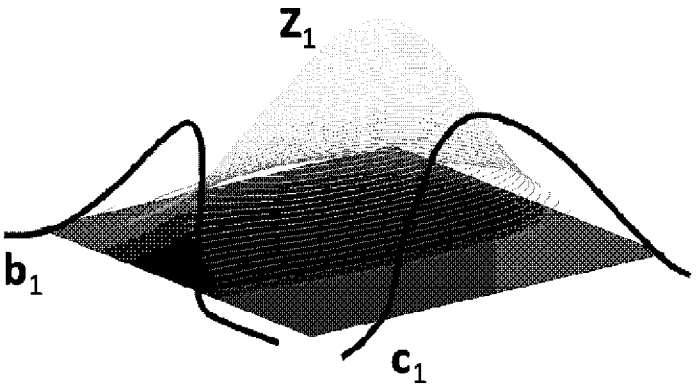

Fig. 2. Example of how the EEM of a fluorophore is the product of its excitation and emission spectra ($Z_1 = b_1 c_1^T$)

does not invalidate the use of PARAFAC to characterize DOM fluorescence, provided the user is aware of these issues, and careful model validation and interpretation of results is applied.

When dealing with simple mixtures, the PARAFAC components identified are often the individual fluorophores present in the mixture (e.g., Bro 1997). However, this is most likely not fully possible with DOM fluorescence due to its complex nature (Del Vecchio and Blough 2004). The components identified (if a valid and robust model is derived) represent fluorescent phenomena, some of which may be individual fluorophores, but may just as well be approximations of the effects of other local processes (quenching or intra molecular charge transfer) occurring.

Approach

This section contains a discussion of various aspects of the procedure and recommendations based on our experience with applying PARAFAC to DOM fluorescence. Fig. 3 shows a flow chart outlining the overall stages in the approach. A detailed step by step hands-on tutorial is presented in Appendix 1.

Data considerations—The first question that is often asked while preparing to use PARAFAC is as follows: how many EEMs are required in order to obtain a robust and adequate PARAFAC model? There is no simple answer to this question as it is dependent both on the nature of the data and the focus of the study. When dealing with complex mixtures such as DOM, where both the number and characteristics of the underlying fluorescent signals are unknown, it is generally preferable to model datasets with 20-100 samples and being close to or even far above 100 samples generally makes modeling simpler. This is largely due to the model validation process, where split half analysis is often the most suitable to apply (see section *Model Validation*). In our experience with DOM fluorescence, best results are often obtained with datasets spanning a gradient (e.g., mixing) or following a process (e.g., production or removal), depending on the focus of the study. This makes both the PARAFAC modeling and the interpretation of results easier. For example, if a study is examining the changes in

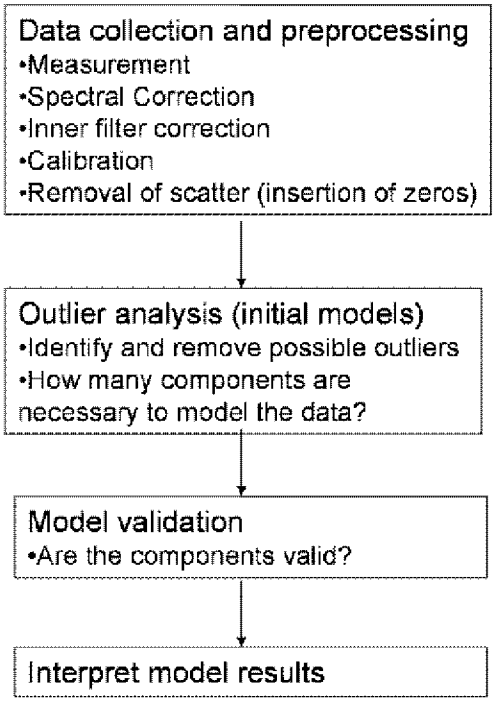

Fig. 3. A diagram summarizing the different necessary steps in the fluorescence-PARAFAC analysis

DOM fluorescence during estuarine mixing, an ideal set of samples would span the relevant salinity range. Arriving at a robust model based on a dataset with, for example, 20 river samples (freshwater) and 20 marine samples (salinity > 30) may be difficult and is also actually basically inappropriate for the study focus (estuarine mixing). Similarly, if studying the production or removal of fluorescent DOM in an experiment, best results are often arrived at when several time points are sampled rather than just at the start and finish of the experiment.

A second question that arises is as follows: if we create a model based on a wide range of samples, is it then justified to fit this model to all future samples rather than derive and validate a new model for each dataset? Although this is a tempting idea, it is also, at present, inappropriate on a global scale. The fluorescence of DOM is dependent on a range of environmental and analytical factors. How can one be sure that a new batch of samples has a similar fluorescence to the training data the model was derived from? In this regard, this remains to be systematically tested. It is already clear from comparing the results of different studies that there are certain common components across broad geographical regions (see Stedmon et al. 2007 sup. info.). More work is needed before a global perspective of the individual DOM fluorescence signals is or even can be obtained. On a more local system scale, this approach may be more appropriate but should only be applied after vigorous testing.

Preliminary data treatment—In any kind of quantitative analysis of fluorescence data, careful calibration of the instrument

data are necessary. It is important to correct the fluorescence measurements for instrument specific biases. These corrections have both a quantitative (signal intensity) and qualitative (spectral shape) aspect. Quantitative calibration is a simple issue for applications where one is working with known fluorophores. The routine measurement of Chlorophyll *a* fluorescence as an indicator of phytoplankton biomass is an example of this. With DOM fluorescence however, the responsible fluorophores or phenomena are unknown, and there is a requirement to calibrate the signal intensity to enable comparison of measurements. To date, there are two methods that are routinely applied. The first involves calibrating the fluorescence to that of quinine sulfate measured with an excitation at 350 nm and emission at 450 nm (Coble 1993). Another approach is to use to the Raman signal from pure water (e.g., Determann et al. 1994; Nieke et al. 1997; Stedmon et al. 2003).

Spectral corrections take into account deviations in the spectral output of the light source and small imperfections in the instrument's components ability to transmit light. However, most modern fluorescence spectrophotometers have a built-in reference detector which is used to normalize the measured signal to the lamp output and hereby correct for changes in the spectral properties of the light source. This is sometimes referred to in manuals as operating in "ratio" mode, and the correction removes a large part of the spectral error. Minor deviations usually remain as the light paths from the lamp to the sample and from the lamp to the reference detector are slightly different. Therefore it is still recommended to correct for and periodically check for additional deviations due to other factors. This can be done using a concentrated solution of Rhodamine B or 101 (Melhuish 1975; Karstens and Kob 1980; Stedmon et al. in prep.).

Similar to the issues with excitation light, every fluorometer is not capable of transmitting the sample emission (fluorescence) from the cuvette to the detector equally efficiently at all wavelengths. Therefore an emission correction spectrum must be applied. This is often supplied with the instrument and is considered to be less variable. However, this should be monitored.

Possible inner filter effects also have to be corrected for. They result from the absorption of the incoming excitation light and to a lesser extent, absorption of the emission light. For DOM, the greatest effects are seen on the excitation signal as this is where DOM itself absorbs the most light. The emission signal is influenced to a lesser extent as the absorption by DOM is considerably lower at these wavelengths (300-600 nm). For samples and wavelengths where DOM absorption coefficients are above approximately 10 m^{-1} (corresponding to an absorbance of 0.04 in a 1 cm cuvette), considerable reductions in the fluorescence signal are expected and observed (Fig. 4). A simple correction can be applied and account for the most of this effect (Mobed et al. 1996; Lakowicz 1999; Ohno 2002).

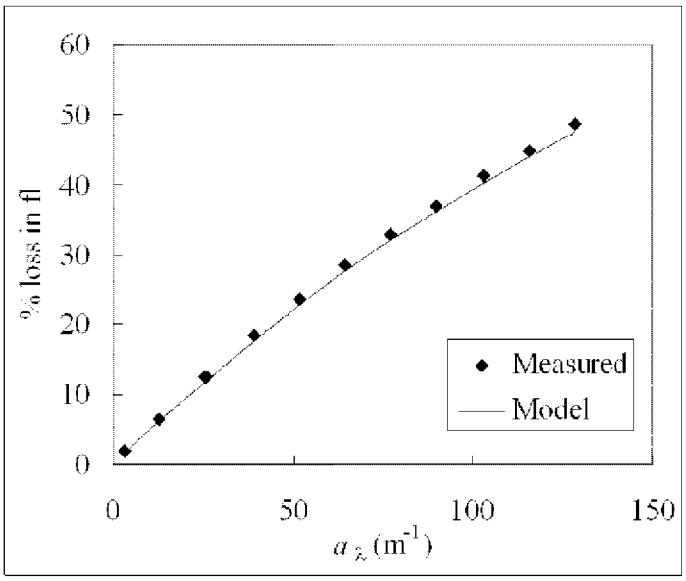

Fig. 4. Example of inner filter effects on fluorescence. It is seen that if the absorption coefficient is below 10 m^{-1} (absorbance 0.04 cm^{-1}) inner filter effects become negligible (<5%). Data are derived from a dilution series of the Suwannee River Fulvic Acid (International Humic Substances Society), starting with a stock solution of 24 mg/L. Points represent measured data. Curve represents predicted effect using the correction equations (e.g., Ohno 2002).

The final step of the data pretreatment is to handle the effects of Raman and Rayleigh scatter (Fig. 5). These scatter signals do not conform to the PARAFAC model (Eq. 1). The position of the scatter peaks change depending on the excitation wavelength. If they were to be modeled with PARAFAC, a new factor would be needed for each excitation wavelength. In addition, the scatter signal from Rayleigh is often very large, and it is therefore most appropriate to delete it before modeling (Fig. 5). The region of the EEMs below the Rayleigh peak should be set to zero. The majority of Raman effects can be removed by subtracting the pure water spectrum from the sample spectrum (Fig. 5). The Rayleigh removal results in a region of the EEMs not containing any data, which can slow the PARAFAC modeling procedure considerably as it tries to extrapolate across the region. Recent work has shown that this can be overcome by inserting a series of zero values in the region of no fluorescence (excitation wavelength << emission wavelength) (see Stedmon et al. 2003; Andersen and Bro 2003; Rinnan and Andersen 2005).

Exploratory analysis—The initial stage of the PARAFAC analysis consists of fitting a series of models to the data using one, two, three components, etc. Modeling is typically done by increasing the number of components from too few up to the supposedly correct number of components. Part of the reason for this is that if too many components are used, the resulting models are meaningless for mathematical reasons and hence complicated to interpret. It is important to stress that the right number of components has to be determined in

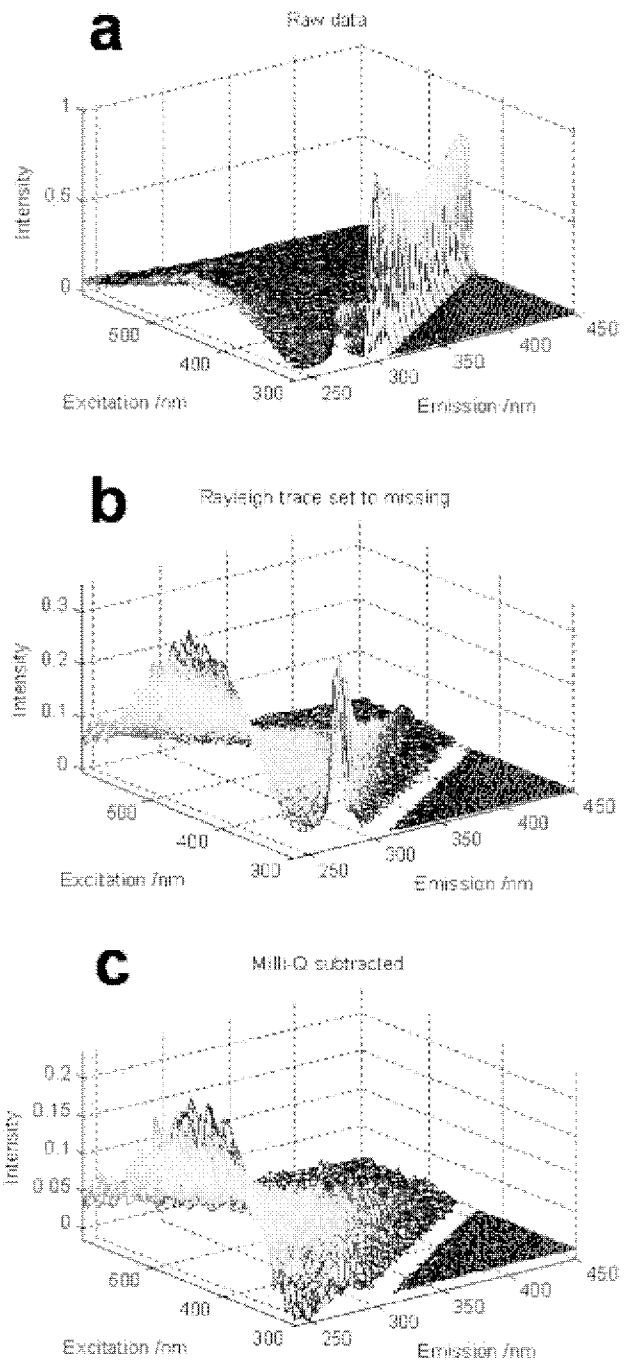

Fig. 5. Example of EEMs before and after removing the effects of Rayleigh and Raman scatter. a) Original sample, b) Rayleigh scatter removed, and c) Raman peak removed by subtracting a MilliQ blank.

the end. If too few or too many components are used, all estimated parameters are incorrect from a chemical point of view because they all depend on each other during the estimation.

Typically, the complexity (number of components) is increased until a reasonable fit has been obtained. Depending on the data, a reasonable fit for EEM data is normally above 99% variance explained, unless the signal is very close to the measurement noise. During these initial models, the shape of the estimated spectra are also considered and extreme outliers (e.g., samples that are completely different) are observed and possibly tentatively removed. This initial analysis can be considered as a combination of outlier identification and explorative data analysis. This step is crucial for later arriving at a valid model. The term outlier is used in its broadest sense. With regards to sample, an outlier may either be a sample that contains some measurement error or artifact, or a correctly measured sample, which is just very different from all the other samples. Outlier wavelengths could, for example, be excitation wavelengths below 240 nm, where most instrument light sources perform poorly, resulting in much lower signal to noise ratios relative to the other wavelengths.

Ideally the removal of one sample from a data set should not influence the shape of the components derived (the model result). If a very different result is obtained, this most likely indicates that the sample is an outlier or that the data set is extremely poorly sampled. The same is applicable in simple linear regression analysis. An effective tool for outlier identification is to calculate the leverage that each sample and wavelength has. The leverage relates to what degree a sample (or wavelength) deviates from the average distribution for all the data. Its value can vary between zero and one, with zero representing a sample that is identical to the average and one representing a sample that is unique and hence completely different from the other samples. Ideally, there should not be a sample that has a notably higher leverage than the others. A sample with a high leverage should first be examined to ensure that there is not analytical (measurement/calibration/correction) error. Second, one can experiment with removing the sample and remodeling. If this does not have any effect on the model result then this indicates that the sample can remain in the data set. If a notably different model is obtained and no analytical error is observed, this indicates that the sample is too unique and that either more samples similar to it need to be included in the data set or that it should be removed. Note that outlying behavior is often based on model parameters (such as leverage) and hence depends on having determined the correct number of components. Likewise, it is not possible to determine the right number of components before outliers have been handled. In practice, one therefore has to work iteratively by tentatively identifying (gross) outliers and tentatively guessing the approximately right number of components.

Model validation—This is the most important part of the analysis. Essentially the goal of this step in the procedure is to identify how many components are present and detectable in the data and to prove the validity of these components. This is no easy feat and at present is the focus of much research in chemometric data analysis. There are a range of tools available to assess the appropriate number of components, however not all function well when applying PARAFAC to model DOM fluorescence. Assuming that a thorough outlier analysis has been carried out, then a model can be validated using four approaches.

Residual analysis. Are there systematic patterns in the residuals? The residuals should be characterized by instrument noise and contain little structure. The presence of a region with a peak next to a trough (negative values) in the residuals indicates that the model is not fitting well (Fig. 6a). If they are present in many samples, this suggests that an inappropriate number of components has been fitted. If they are only present in one or two samples, then this would suggest that these samples might be considered as, possibly moderate, outliers.

Examine the spectral properties of each component. Do the components have spectral properties expected of fluorophores? Excitation spectra can have one or more maxima, but emission spectra should only exhibit one emission maximum. Additionally, the excitation and emission spectra of organic fluorphores often overlap slightly. The spectra should also be smooth, although this depends to a certain extent on the spectral resolution of the fluorescence data (Fig. 5b).

Split half analysis (Harshman 1984). The data are split into two halves (with respect to samples), and each half is modeled independently. Then the spectral properties of the components derived from each half are compared (b_{jf} and c_{kf} in Eq. 1). If found to be identical, then the model is considered robust. It is important to consider how the data are split to ensure that both halves contain similar variability. If the data set is too small and spans a large range in samples, it will be difficult to validate using this approach.

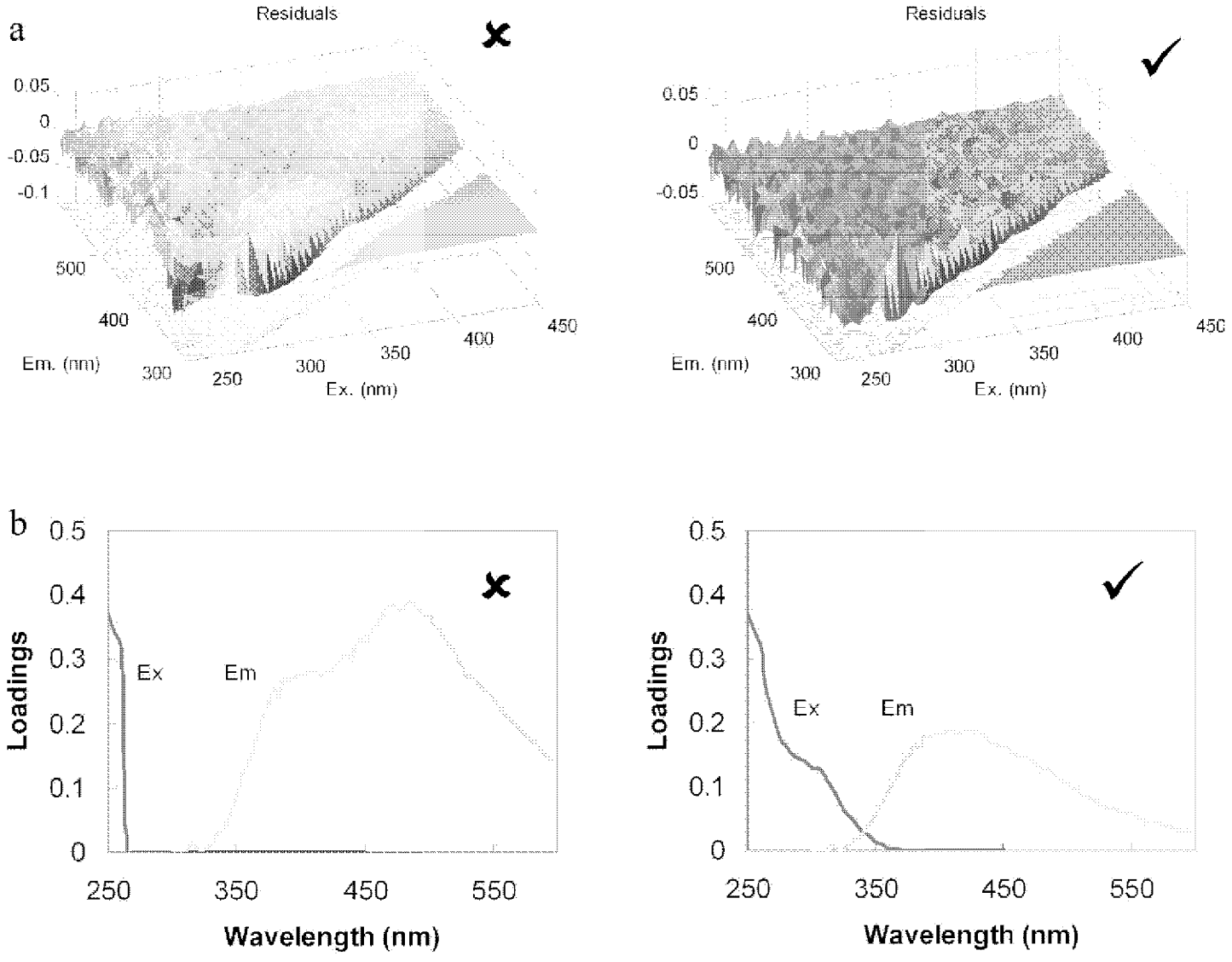

Fig. 6. a) Example of the residuals from two different models fitted to a DOM EEM. The graph on the left has systematic deviation suggesting that the model is inadequate. The graph on the right resembles more instrument noise with no clear peaks or troughs. b) Example of the excitation and emission spectra (loadings) of a component derived from PARAFAC modeling. The lefthand graph exhibits properties unlike a fluorophore, whereas the righthand side graph resembles that of an organic fluorophore.

Random initialization. For this a series of models are fitted using random numbers as the initial estimates. By comparing the fit (sum of squared error) from each model, one can ensure that the result is in fact the least squares results and not a local minimum.

Model refinement—The algorithms used for fitting PARAFAC have their imperfections. This is partly due to the algorithms themselves but also partly due to the actual mathematical problem of fitting PARAFAC models. This has one important repercussion on practical data analysis which is that local minima solutions can exist. This is particularly the case when a solution is difficult to estimate. This is often observed indirectly by estimated concentrations or spectra being very similar. This is a sign of a potentially difficult problem, and in such case, it pays to re-estimate the solution several times (as mentioned in point four above) and observe if the fit is the same every time. If not, there are local minima and only the best-fitting estimates can be assumed to represent a solution. It may also be indicative of a problem that is mathematically difficult, e.g., due to very similar spectral profiles. Increasing the number of samples may then help.

Data interpretation—Once the model has been validated, the fluorescence data can now be examined with respect to the variability in each component found. As the actual identity of the components is unknown, the fluorescence of each component cannot be converted into a concentration. Instead the fluorescence intensity at the maximum can be derived for each component and has the same units as the measurements were calibrated with, e.g., Quinine Sulfate Equivalents or Raman units. Care must be taken when interpreting this data. The fact that the fluorescence signal from one component (A) is greater than that from another (B) does not equate to A having a higher concentration than B, only higher fluorescence. Fluorescence is not only dependent on concentrations but also on the molar absorptivity and quantum efficiency, which is unknown. However, relative changes and ratios between component fluorescence can be used to illustrate the quantitative and qualitative differences between samples. If certain PARAFAC components can be identified as specific chemical analytes, absolute calibration can be performed, for example, through standard addition.

Perspectives

The combination of EEM fluorescence and parallel factor analysis has proved to be a promising tool for studying DOM. These relatively inexpensive fluorescence measurements can be used in combination with other measurements to rapidly quantify and characterize DOM across a range of environments. This approach also has great potential in a wide range of other applications, such as drinking water monitoring, waste water treatment control, and the evaluation of ballast water exchange in ships. However, as with many new techniques, it is not as straightforward as could be desired, and there are a range of potential pitfalls. For the approach to persist and become a trustworthy method, a basis for a standardized approach was required. We hope that the discussion presented in this paper and the tutorial in the appendix provides a suitable starting point for developing this approach further.

References

Andersen, C. M., and R. Bro. 2003. Practical aspects of PARAFAC modeling of fluorescence excitation-emission data. J. Chemom.17:200-217.

Baker, A. 2001. Fluorescence Excitation-Emission matrix characterization of some sewage impacted rivers. Environ. Sci. Technol. 35:948-953.

Boehme, J., P. Coble, R. Conmy, and A. Stovall-Leonard. 2004. Examining CDOM fluorescence variability using principal component analysis: seasonal and regional modeling of three-dimensional fluorescence in the Gulf of Mexico. Mar. Chem. 89: 3-14.

Bro R. 1999. Exploratory study of sugar production using fluorescence spectroscopy and multi-way analysis. Chemom. Intell. Lab. Syst. 46: 133-147.

Bro R. 1997. PARAFAC tutorial and applications. Chemom. Intell. Lab. Syst. 38: 149-171.

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

———. 1996. Characterization of marine and terrestrial DOM in seawater using excitation-emission matrix spectroscopy. Mar. Chem. 51:325-346.

Cory, R. M., and D. M. McKnight. 2005. Fluorescence spectroscopy reveals ubiquitous presence of oxidized and reduced quinones in dissolved organic matter. Environ. Sci. Technol. 39:8142-8149.

Del Vecchio, R., and N. V. Blough. 2004. On the origin of the optical properties of humic substances. Environ. Sci. Technol. 38:3885-3891.

Determann, S., R. Reuter, P. Wagner, R. Willkomm. 1994. Fluorescent matter in the eastern Atlantic Ocean. Part 1: Method of measurement and near-surface distribution. Deep Sea Res. 41:659-675.

Hall, G. J. and J. E. Kenny. 2007. Estuarine water classification using EEM spectroscopy and PARAFAC–SIMCA. Anal. Chim. Acta 581:118-124.

Hall, G. J., K. E. Clow, and J. E. Kenny. 2005. Estuaral fingerprinting through multidimentional fluorescence and multivariate analysis. Environ. Sci. Technol. 39:7560-7567.

Harshman, R. 1984. "How can I know if it's real?" A catalog of diagnostics for use with three-mode factor analysis and multidimensional scaling, pp. 566-591. *In* H. G. Law, C. W. Snyder, J. A. Hattie, and R. P. McDonal [eds.], Research methods for multimode data analysis. Praeger.

Kalle, K. 1966. The problem of gelbstoff in the sea. Oceanogr. Mar. BiolA. Rev. 4:91-104.

Karstens, T., and K. Kobs. 1980. Rhodamine B and Rhodamine 101 as reference substances for fluorescence quantum yield measurements. J. Phys. Chem. 84:1871- 1872.

Lakowicz, J. R. 1999. Principles of fluorescence spectroscopy, 2nd ed. Kluwer Academic/Plenum Publishers.

Martens, H. and T. Naes. 1989. Multivariate Calibration, Chichester, John Wiley & Sons.

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

Melhuish, W. H. 1975. Modified technique for determining the wavelength-sensitivity curve of a spectrofluorimeter. App. Optics 14:26-27.

Mobed, J. J., S. L. Hemmingsen, J. L. Autry, and L. B. McGown. 1996. Fluorescence characterization of IHSS humic substances: Total luminescence spectra with absorbance correction. Environ. Sci. Technol. 30:3061-3065.

Murphy, K. R., G. M. Ruiz, W. T. M. Dunsmuir, and T. D. Waite. 2006. Optimized parameters for rapid fluorescence-based verification of ballast water exchange by ships. Environ. Sci. Technol. 40:2357-2362.

Nieke, B., R. Reuter, R. Heuermann, H. Wang, M. Babin and J.C. Therriault. 1997. Light absorption and fluorescence properties of chromophoric dissolved organic matter (CDOM), in the St. Lawrence Estuary (Case 2 waters). Cont. Shelf Sci. 17:235-252.

Ohno, T. 2002. Fluorescence inner-filtering correction for determining the humification index of dissolved organic matter. Environ. Sci. Technol. 36:742-746.

Persson, T., and M. Wedborg. 2001. Multivariate evaluation of the fluorescence of aquatic organic matter. Anal. Chim. Acta 434:179-192.

Rinnan, A., K. S. Booksh, and R. Bro. 2005. First order Rayleigh scatter as a separate component in the decomposition of fluorescence landscapes. Anal. Chim. Acta 537:349-358.

Stedmon, C. A., D. N. Thomas, M. Granskog, H. Kaartokallio, S. Papadimitriou, and H. Kuosa. 2007. Characteristics of Dissolved Organic Matter in Baltic Coastal Sea Ice: Allochthonous or Autochthonous Origins? - Environ Sci Technol. 41(21): 7273-7279.

Stedmon, C.A. and S.S. Markager. 2005a. Resolving the variability in dissolved organic matter fluorescence in a temperate estuary and its catchment using PARAFAC analysis. Limnol. Oceanogr. 50(2):686-697.

Stedmon, C.A. and S.S. Markager. 2005b. Tracing the production and degradation of autochthonous fractions of dissolved organic matter by fluorescence analysis. Limnol. Oceanogr. 50(5):1415-1426.

Stedmon, C. A., S. Markager, and R. Bro. 2003. Tracing dissolved organic matter in aquatic environments using a new approach to fluorescence spectroscopy. Mar. Chem. 82:239-254.

Submitted 2 July 2008
Revised 17 October 2008
Accepted 22 October 2008