An R Script for Visualising and Analysing Fluorescence Excitation–Emission Matrices (EEMs)

Dan J. Lapworth* and David G. Kinniburgh

British Geological Survey, Maclean building, Crowmarsh Gifford, Wallingford, Oxon, OX10 8BB
e-mail: djla@bgs.ac.uk
* Corresponding author

Keywords: Fluorescence spectroscopy, excitation-emission matrices, 3D visualisation

Introduction

Fluorescence spectroscopy has been used for characterising the composition and sources of dissolved organic matter (DOM) in many different environments and is an active area of research (Baker and Lamont-Black 2001, Hudson et al. 2008, Hall et al. 2005, Jones et al. 2002). Natural organic matter (NOM) is derived from the decay of organic material from both plant, animal and microbial sources and a significant proportion of NOM fluoresces (Senesi, 1993). Due to rapid technological advances in recent years modern fluorescence instruments can now characterise the excitation–emission properties of a sample at multiple wavelengths very rapidly (typically< 5 minutes) and are able to produce matrices which represent the three dimensional fluorescence landscape of a particular sample. Peaks in fluorescence intensity are related to the concentrations of a particular fluorophores (e.g. fulvic/humic-like compounds, tryptophan and tyrosine) and the method is able to detect low concentrations of these compounds in environmental samples.

As well as simply displaying results from excitation–emission matrices (EEMs) visualising the EEM data is also the first step in carrying out any sort of detailed data analysis or interpretations using EEMs. For example prior to carrying out statistical analysis on samples e.g. parallel factor analysis (Bro 1997, Bro et al. 2002, Stedmon

et al. 2003) the first step in the process is to visualise the matrices to assess if it is worth carrying out further analysis, i.e. to isolate and possibly remove obvious outliers or erroneous EEMs from the modelling process.

Three dimensional plots are used in many applications, common examples include mapping landscape topography, thermal imaging and NMR imaging (Fairley and Nicholson 2006, van Beek 2007). Many different commercial statistical and graphics packages are available to plot three dimensional data (e.g. Matlab, S-plus, Graphis) as well as freeware applications (e.g. openGL and R). However, in many cases considerable time and effort is required by the user to understand the programming language of the application in order to implement a particular application, and the initial cost of the application may also be prohibitive.

R is an open source statistical and graphics system that is a dialect of the S language, the basis of the commercial S-PLUS system (Maindonald and Braun 2003). It is a function based language that has an extensive and rapidly growing library of packages (Ihaka and Gentleman, 1996, R Development Core Team, 2008).

This computer note outlines a tool for producing a range of plots of fluorescence EEM data in the open–source statistical package R. A script has been written that allows the user to input data from Excel worksheets (CSV files) and process multiple input data files. The R script will then produce the plots and summary statistics, in the form of a CSV table, and write the output files to a user specified folder. Plot types include filled contour plots and 3D wireframe perspective plots for whole EEMs, filled contour plots for selected regions within an EEM and plots for specified slices across the EEM. This script provides a free tool to enable the user to rapidly visualise EEMs and produce tailored EEM plots and slices through EEMs of publication quality.

An R script for Plotting Fluorescence Data

The R script presented in this paper is designed to read in and process CSV files of

EEMs configured as a matrix. For the generic script presented in this paper the default

specification for the data input file is as follows: excitation wavelengths n= 40 (200–

400nm with 5nm interval, columns in CSV file), emission wavelenghts n= 111 (200–

500 nm with 2 nm interval, rows in CSV file), intensity range is set as -10–1000 (a.u).

Figure 1 shows the format of the sample file that is read into R as a CSV file. R needs

to be installed on the computer, and this can be done by downloading the latest

version from the Comprehensive R Archive Network (CRAN – see section on

software availability). Figure 2 shows the file structure that is required to run the

script. Prior to running the script the EEM CSV files are saved into the data folder.

The script enables multiple files to be processes at the same time. As a default the plot

file names and titles on plots are the same as the original data file names in the data

folder. While this is the default setup it is relatively simple to modify the R script to

be able to read in files from other instruments and customise where the data files are

read from and written to. The R script enables the user to have full control of the

plotting region e.g. the title, axis and colours can all be modified by the user to suit

different tastes. A range of 3D plotting functions are produced, filled contour plots

and wireframe plots of the whole EEM and filled contour plots of selected regions

within the EEM. Slices through the EEM are also produced at selected excitation and

emission wavelengths, these and the other plot options can also be easily modified by

the user. The script also writes results (mean, min, max, Σ intensity and standard

deviation) extracted from selected regions within the EEM to a CSV file called

"Summary_Results.csv" in the results folder. The user is provided with an R script, a

detailed tutorial on how to set up and run the R script, and some demonstration

sample files to familiarise the user with the tool.

A script is also available that will read in and process CSV files that are produced directly from the Varian Cary Eclipse Fluorescence Spectrometer™, the only difference being the script first manipulates the data produced by the instrument into the matrix format shown in Figure 2. It is relatively simple to modify the script presented in this paper to process non-standard fluorescence data which is not configured in a matrix.

User interface

Once R is installed and started the script is loaded into the R editor window by browsing to the R script in the R graphical user interface GUI: file, open script and browse to where the script was saved. Set 'history' in the graphics window to 'Recording' in the GUI this will record the plots generated in the graphics window and enable the user to page up and down to view all the plots. Once the script is loaded it is run by using the following commands, Ctrl A then Ctrl R from the editor window. Plots are automatically generated in the graphics window. Plots can be written to file from the GUI: file, save as, and choosing the output format (metafile, postscript, pdf, png, bmp, TIFF or jpeg). For details on how to write plots directly to file, change the selected plot zone and select your own 2D slices of the EEM (see the tutorial supplied with the script).

Example Application

The data used for this demonstration is an EEM scan of a shallow groundwater sample from the Permo-Triassic Sandstone in the UK (Lapworth et al. 2008). Figure 3a shows the whole EEM as a filled contour plot and 3b as a wireframe plot. Figure 3c shows cross-sections through the fulvic-acid like region of the EEM (Excitation wavelength = 330 nm). The peak on the left hand side of plot Figure 3c is the Raman line. Figure 3d shows a zoomed in plot of the tryptophan-like region (Ex250–300 nm,

Em300–370 nm). These were all generated from the default script using the demonstration sample file supplied with the R script. Ultrapure water (ASTM type I reagent grade water, including a UV cracker) was used for the blank sample and to clean the quartz cell between samples. All EEM plots have had the linear features of the Rayleigh-Tyndall and Raman lines removed. Blank subtraction is able to remove most of the linear features of the water Raman line and the Rayleigh-Tyndall lines were masked out by multiplying the sample matrix with a mask matrix containing values of 0 and 1. The mask matrix was developed by delineating the regions affected by the Rayleigh-Tyndall lines in the blank and samples and assigning them a value of 0, other regions of interest were assigned a value of 1. The blank and mask files are supplied with the demonstration files and the R script.

One possible future application of this script would be to allow rapid data analysis of fluorescence results from in-situ monitoring systems for waste water treatment processes. With only relatively small developments the script could rapidly extract and display results as a time series allowing the operators to monitor the performance of the treatment process by using fluorescence as a water quality indicator e.g. a proxy for biological oxygen demand (BOD) (Hudson., 2008).

Conclusions

This paper has demonstrated the use of open-source software to visualise and analyse raw 3D EEM data. The R package is quick and easy to install, and the R script is able to rapidly help the user assess fluorescence results and produce outputs of publication quality. The outputs can be tailored to the users requirements, and either visualised in the GUI or written directly to file for future use.

Software Availability

The latest version of R is available free of charge from the CRAN (http://www.r-project.org/) and the R script for visualising EEMs (PLOTEEM.R or the equivalent script that will process Varian CSV files), tutorial and example data files are available free upon request by contacting djla@bgs.ac.uk.

Acknowledgments

References

Bro, R. 1997. PARAFAC, tutorial and applications. Chemometrics and Intelligent Laboratory Systems 38 149–171.

Bro, R., Sidiropoulos, N., Smilde, A. 2002. Maximum likelihood fitting using ordinary least squares algorithms. Journal of Chemometrics 16 387-400.

Baker, A., Lamont-Black, J. 2001. Fluorescence of Dissolved Organic Matter as a Natural Tracer of Ground Water. Ground Water 39 (5), 745–750.

van Beek, J.D, 2007. matNMR: A flexible toolbox for processing, analyzing and visualizing magnetic resonance data in Matlab®. Journal of Magnetic Resonance 187 (1), 19–26.

CRAN, Comprehensive R Archive Network, http://cran.r-project.org/.

Fairley, J.P., Nicholson, K.N. 2006. Imaging lateral groundwater flow in the shallow subsurface using stochastic temperature fields. Journal of Hydrology 321 (1/4), 276–285.

Hall, G.J., Clow, K.E., Kenny, J.E. 2005. Estuarial fingerprinting through multidimensional fluorescence and multivariate analysis. Environment Science and Technology 39 (19), 7560–7567.

Hudson, N., Baker, A., Ward, D., Reynolds, D.M., Brunsdon C., Carliell-Marquet, C, Browning, S. 2008. Can fluorescence spectrometry be used as a surrogate for the Biochemical Oxygen Demand (BOD) test in water quality assessment? An example from South West England. Science of the Total Environment, 391 (1), 149–158.

Ihaka, R., Gentleman, R., 1996. R: A language for data analysis and graphics. Journal of Computational and Graphical Statistics, 5 (3), 229-235.

Jones, D.L., Owen, A G., Farrar, J F. 2002. Simple method to enable high resolution determination of total free amino acids in soil solutions and soil extracts. Soil Biology & Biogeochemistry, 34 (12), 1893–1902.

Lapworth, D.J., Gooddy, D C., Butcher, A., Morris, B. 2008. Tracing groundwater flow and sources of organic carbon in sandstone aquifers using fluorescence properties of dissolved organic matter (DOM). Applied Geochemistry, doi:10.1016/j.apgeochem.2008.07.011.

Maindonald, J., and Braun, J. 2003. Data analysis and graphics using R – an example based approach. Cambridge University Press, pp 360.

R Development Core Team. 2008. The R foundation for statistical computing, Vienna University of Technology, Vienna, Austria.

Senesi, N. 1993. In: Organic substances in soil and water: natural constituents and their influences on contaminant behaviour; Bech, A.J., Jones, K.C., Hayes,

M.B.H., Mingelgrin, U., Eds.; The Royal Society of Chemistry: Cambridge, p

74.

Stedmon, C.A., Markager, S., Bro, R. 2003. Tracing dissolved organic matter in

aquatic environments using a new approach to fluorescence spectroscopy.

Marine Chemistry 82 (3/4), 239–254.

Figures

Figure 1. Matrix of intensity for R script

Figure 2. Default file structure that is needed to run R script

Figure 3. EEM plots produced: a) filled contour plot of whole EEM, b) perspective
plot of whole EEM, c) cross section through EEM at Ex330 nm, d) zoomed in plot of
tryptophan-like region of EEM.