NATIONAL ENVIRONMENTAL RESEARCH INSTITUTE
AARHUS UNIVERSITY

Characterising dissolved organic amtter fluorescence with parallel factor analysis

Tutorial comments

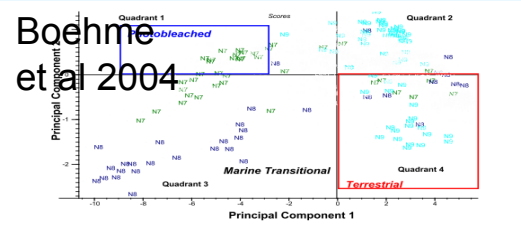
Spectroscopy workshop: Granada, Spain, 19-21 May 2010.

http://spectroscopyworkshop.weebly.com

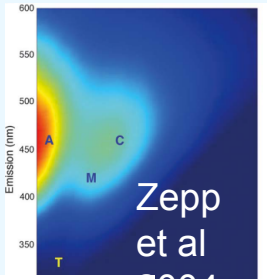
Colin A. Stedmon

cst@dmu.dk

Approaches to characterising DOM fluorescence

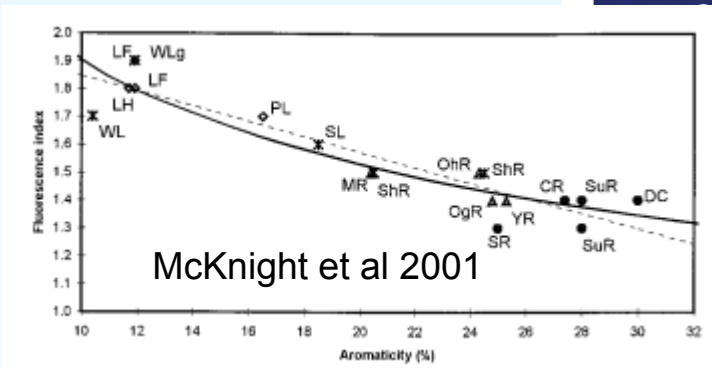


Boehme et al 2004

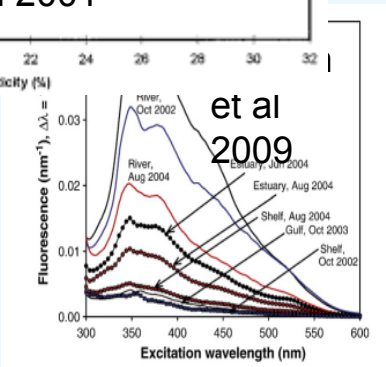


Zepp et al 2004

- **Uni-variate**
  - **Peak/shoulder intensities**
  - **Ratios**
  - **Specific spectra (ex, em or sync)**
- **Multi-variate**
  - **PCA**
  - **PARAFAC**



McKnight et al 2001

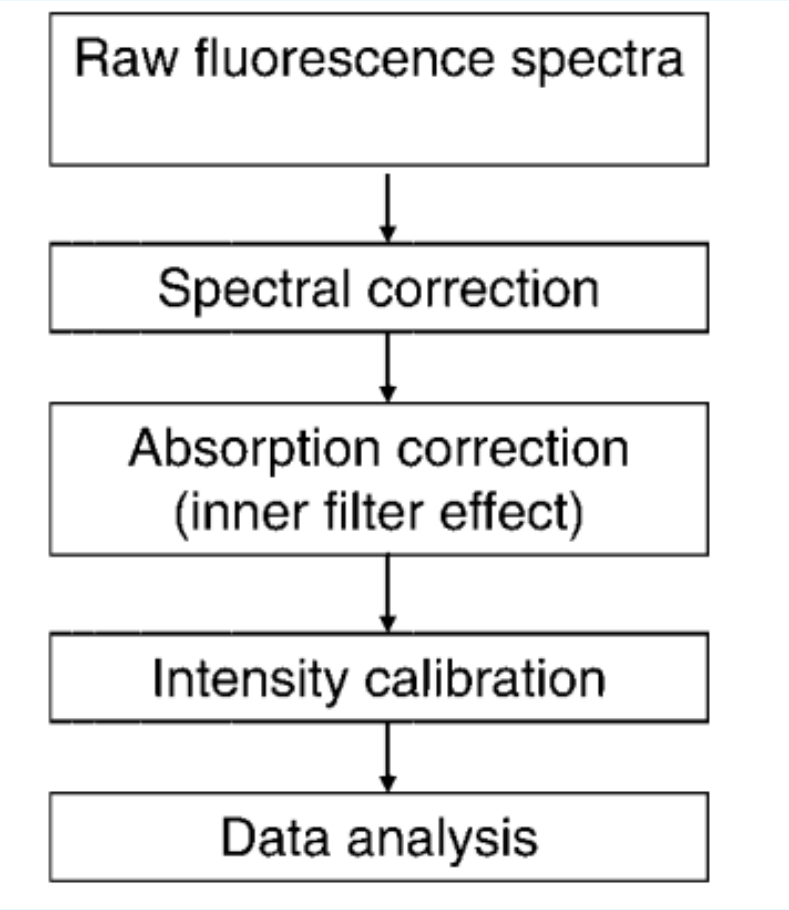


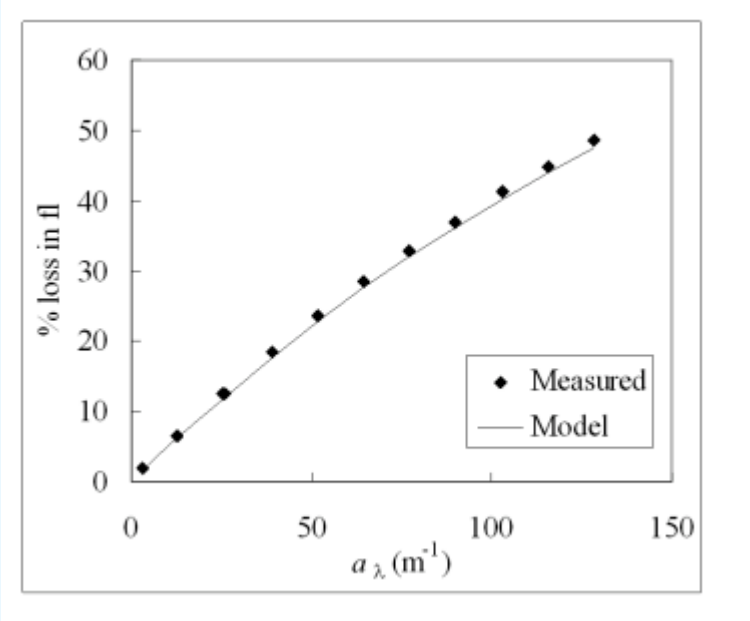
et al 2009

- **Choose the right approach for each specific study**
- **Design the study with the analysis approach in mind**

Data quality

› **The results are only as good as the quality of data used. No magic involved**

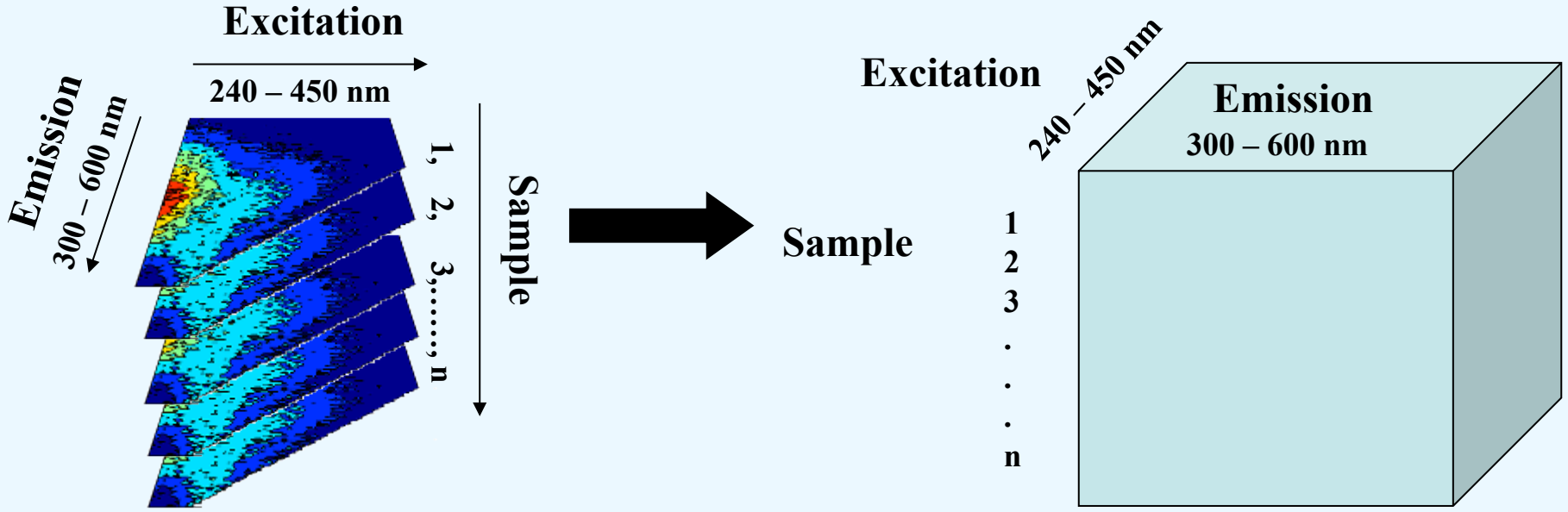




› **Absorbance$_\lambda$ <0.04 cm^{-1} (a_λ = 10 m^{-1}), inner filter is <5%**

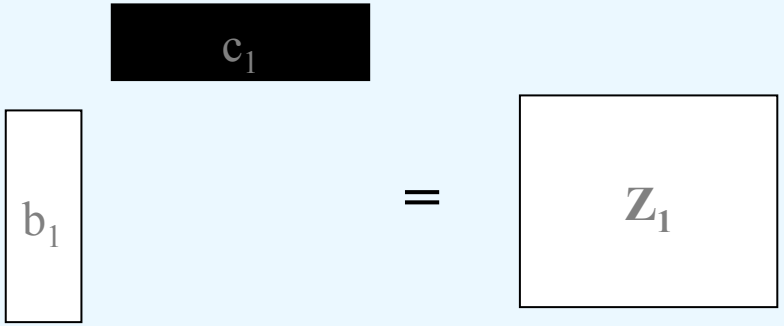
Background on Multi-way analysis?

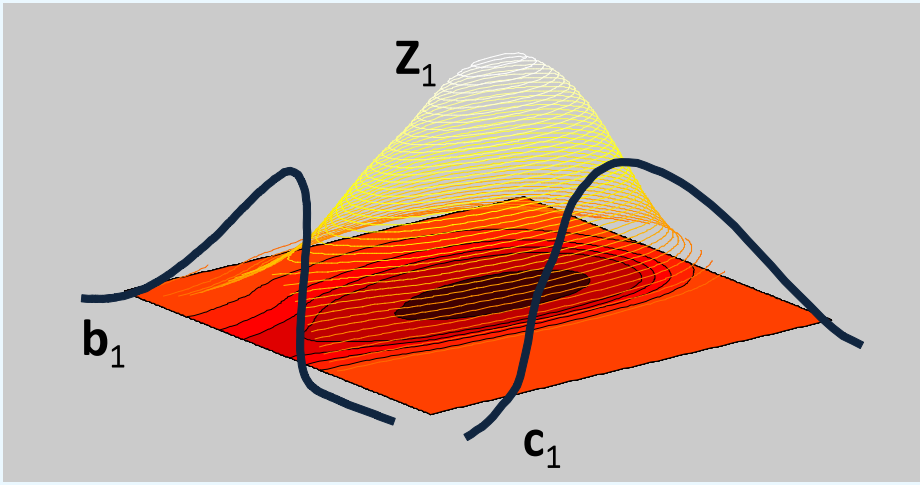
- **PARAFAC is one of a range of multi-way data analysis techniques.**
- **The term multi-way describes data with more than two dimensions (modes).**
- **Spectral fluorescence data is multi-way (3-way) as it varies as a function of excitation and emission wavelength**
- **Combining the data results in a box of data**



Understanding the PARAFAC algorithm

- **EEM of a fluorophore (1) is the product of its emission (b_1) and excitation (c_1) spectra. ($Z_1 = b_1 c_1^T$).**





- **The fluorescence intensity will also vary with fluorophore concentration (a_1)**

$$X = a_1 Z_1 = a_1 b_1 c_1^T$$

- **For each element of matrix X this can be re-written as**

$$x_{jk} = a_1 b_{1j} c_{1k}$$

where j and k refer to emission and excitation wavelengths.

Understanding the PARAFAC algorithm

- From the previous slide we now have the fluorescence of a fluorophore as $x_{jk} = a_1 b_{1j} c_{1k}$
- If there are two fluorophores it becomes

$$x_{jk} = a_1 b_{1j} c_{1k} + a_2 b_{2j} c_{2k}$$

- If there are three…

$$x_{jk} = a_1 b_{1j} c_{1k} + a_2 b_{2j} c_{2k} + a_3 b_{3j} c_{3k}$$

- Which simplifies to…

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

- For more than one sample (i=1, 2…I) it becomes…

$$x_{ijk} = \sum_{f=1}^{F} a_{if} b_{jf} c_{kf}$$

- Which is the equation behind PARAFAC

PARAFAC analysis

› **The model is fitted using an alternating least squares approach. i.e. consecutively varying the values of the matrices A, B and C in order to minimize the unexplained data (residuals)**

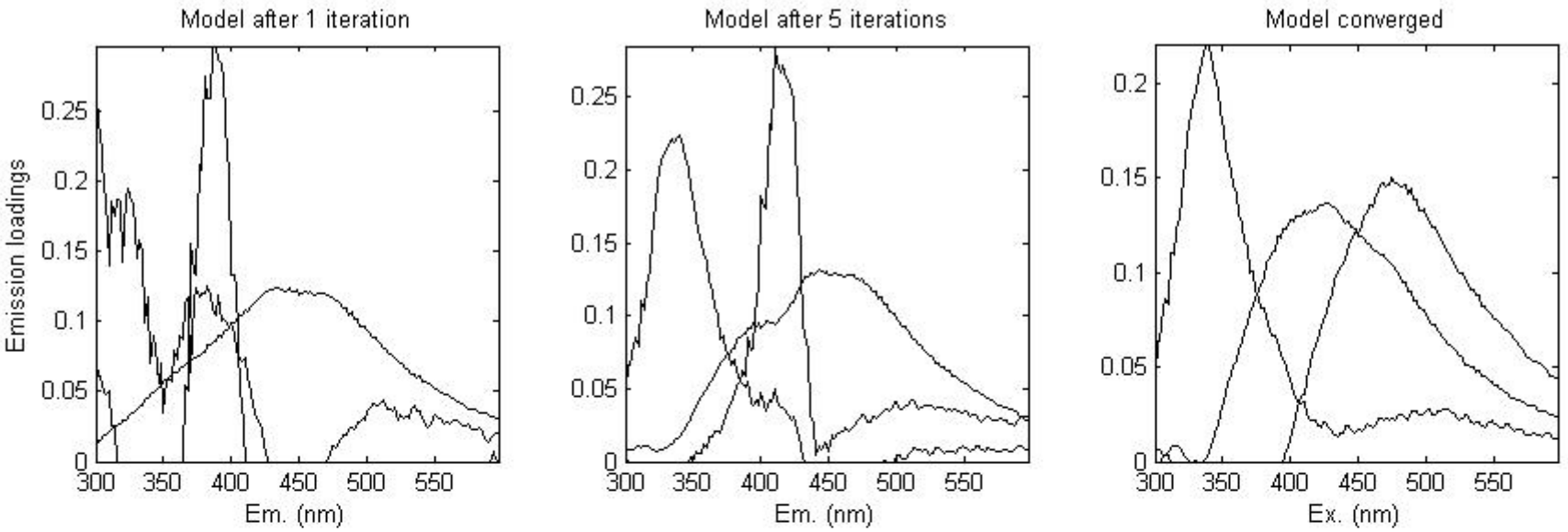
$$x_{ijk} = \sum_{f=1}^{F} a_{if} b_{jf} c_{kf} + \varepsilon_{ijk}$$

Measured data
i samples
j emission wavelengths
k excitation wavelengths

Model data
Concentrations of each component
emission properties of each
excitation properties of each

Residuals
Unexplained data
(e.g. remaining fluorescence, scatter, instrument noise)

Example of fitting procedure - changing loading estimates



- **As the model fits the values of A, B and C are varied sequentially repeatedly until no improvement in fit is achieved (convergence).**
- **Above is an example of the emission loadings for a three component model changing as the iterations progress.**
- **No assumption on the shape of the loadings**

Assumptions

- **Change in concentration of an analyte only influences its fluorescence intensity (not characteristics i.e. "shape"):**
  - **$b_{\lambda Em}$ $c_{\lambda Ex}$ are fixed for each component.**
- **Beer Lamberts law**
  - **Linear dependence between fluorescence and concentration**
- **Components are independent of each other**
- **Don't over interpret.**
  - **simple mixtures: components can be fluorophores**
  - **complex mixture (DOM): we have little knowledge of the structures or phenomena responsible.**

Data considerations

- **No magic number**
- **>30 generally a good start**
- **Large data set much easier to work with**
- **Try to have a dataset than spans a gradient or development (mixing, seasonal, t) rather than a collection of individuals**
  - **e.g. one sample from 60 different lakes will be difficult to characterise with PARAFAC. PCA probably best for grouping data.**
- **Global/regional fixed components.**
  - **apply with caution**
  - **maybe…but are we there yet? What about intercalibration?**

Approach

- **Explorative data analysis**
  - **determine outliers (due to error or just a unique sample or λ region)**
  - **ensure robustness**
  - **arrive at first estimate of suitable number of components (e.g. 3-5, or 5-7)**
  - **iterative approach,-get to know your data- add/remove samples and/or wavelengths**
- **Model validation**
  - **if the earlier step is carried out thoroughly, this is very simple**
  - **Residual analysis**
  - **Component spectra: do they look sensible? i.e. organic fluorescence**
  - **Split half analysis**