Data Handling Methods
Data Handling Methods A

- Peak picking
- Fluorescence indices
- Regional Integration
- Data reduction

Key considerations:

speed (on-line data collection vs manual post processing).
Simplicity (simple linear regression against water quality vs detailed post processing to obtain structural information)
Initial dataset characteristics (EEM vs line scans).
Peak Picking

Simplest data extraction method: finding the maximum intensity ($z$) in a given region $(x,y)$.

Intensity vs concentration relationships can be developed by linear regression.

Wavelength variations may also provide information about organic matter character.
Bieroza et al., 2009, STOTEN

Hudson et al., 2008, STOTEN

Baker et al., 2008, Chemosphere
Fluorescence Indices

Typically ratios of intensity at two x,y wavelength pairs.

They attempt to characterise the organic matter present. For example:

McKnight et al. (2001) FI: attempts to characterise microbial vs terrestrial organic matter.

Lower plot shows relationship with land cover (Wilson and Xenopolous, 2009).
Fluorescence Indices

Baker et al (2008): peak T/peak C intensity shown to correlate with organic matter function (e.g. hydrophobicity, aluminium adsorption and benzopyrene binding).
Fluorescence Indices

Parlanti et al (2000) developed a $\beta/\alpha$ index.

Ratio of ‘peak C’ fluorescence intensity at 380 nm to that between 420 and 435 nm.

For example, applied by Wilson and Xenopoulos (2009) to relate fluorescence to water quality parameters.
Fluorescence indices

Advantages:
Simple to use and understand
Quantifiable
Can be used to design in-situ fixed wavelength instruments

Disadvantages:
Choice of wavelength pairs often arbitrary
Wastes significant information
Extent to which indices are transferable is unclear

Advantages: uses whole EEM.
Disadvantages: choice and definitions of regions may not be transferrable, wavelength variability not considered.
Data Reduction

- Previously detailed techniques are all trying to reduce the data volume to something manageable, but they are doing so in a relatively arbitrary way.

- Statistical / operational techniques are available that would reduce the data to manageable components.

Principal Components Analysis
Principal Filter Analysis
Artificial Neural networks / Expert Systems
PCA

- A method to simplify a data set by means of linear transformations that chose a new coordinate system in which the first axis (first principal component, PC1) describes the greatest variance within the dataset.

- Example: Spencer et al (2007). PC1 correlates strongly positively with peaks C and A intensity and emission wavelengths, absorbance and DOC and strongly negatively with salinity and spectral slope. PC2 is correlated strongly positively with DON, TDN, NH$_4^+$ and peak T and B intensity.
PCA - summary

Advantages:
A powerful exploratory data analysis tool, can reveal relationships in the data that wouldn’t otherwise be observed
Can be combined with any of the previous techniques (peak picking, regional integration, indices), as well as other water quality parameters
Can be performed in standard statistics packages.

Disadvantages:
Some statistics experience needed
Off-line technique
Only as good as the input data.
This is just a brief overview of just a few of the techniques available. Further advanced techniques are available (expert systems, discriminant analysis, parallel factor analysis).

All approaches are a compromise between maximizing information extracted from an EEM and practical considerations (time, data available).

No one solution fits all datasets.