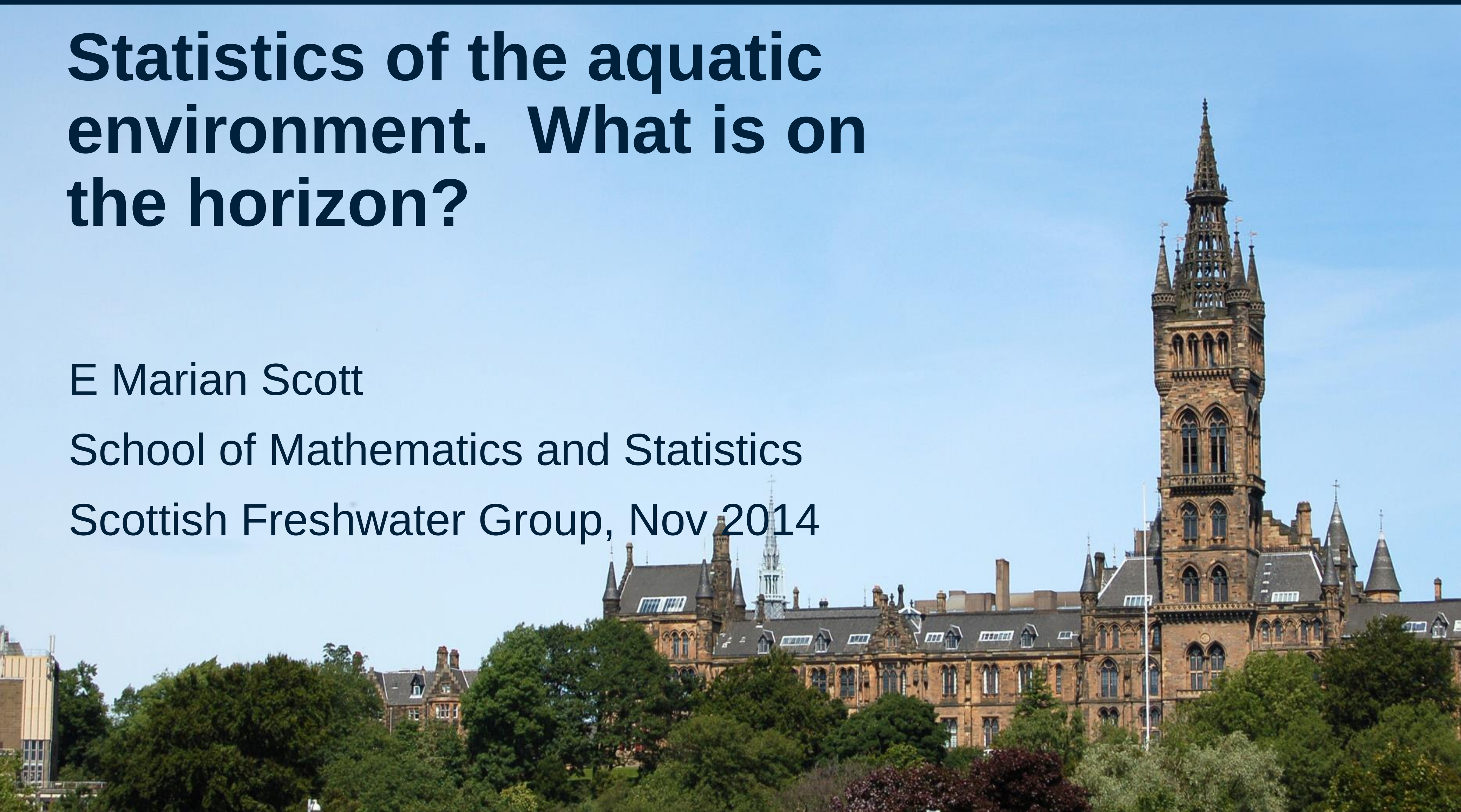
University of Glasgow | College of Science & Engineering

Statistics of the aquatic environment. What is on the horizon?

E Marian Scott

School of Mathematics and Statistics

Scottish Freshwater Group, Nov 2014



University of Glasgow | College of Science & Engineering

The changing data world?

- Emerging sensor technology is able to deliver enhanced dynamic detail of environmental systems at unprecedented scale and “will revolutionise our understanding of the environment by providing observations at .....expanding observational scales that will enable a deeper and broader understanding of environmental variability and change ... improving public awareness, enabling better informed public policies and addressing the intrinsic interdependence of human society and the natural environment.” (NSF, 2004).

The state of the environment

Four questions:

- What is changing?
- What are the changes?
- What is driving the changes?
- How certain are we?

The changing state of how and what we monitor

The changing nature of how we report the “state”

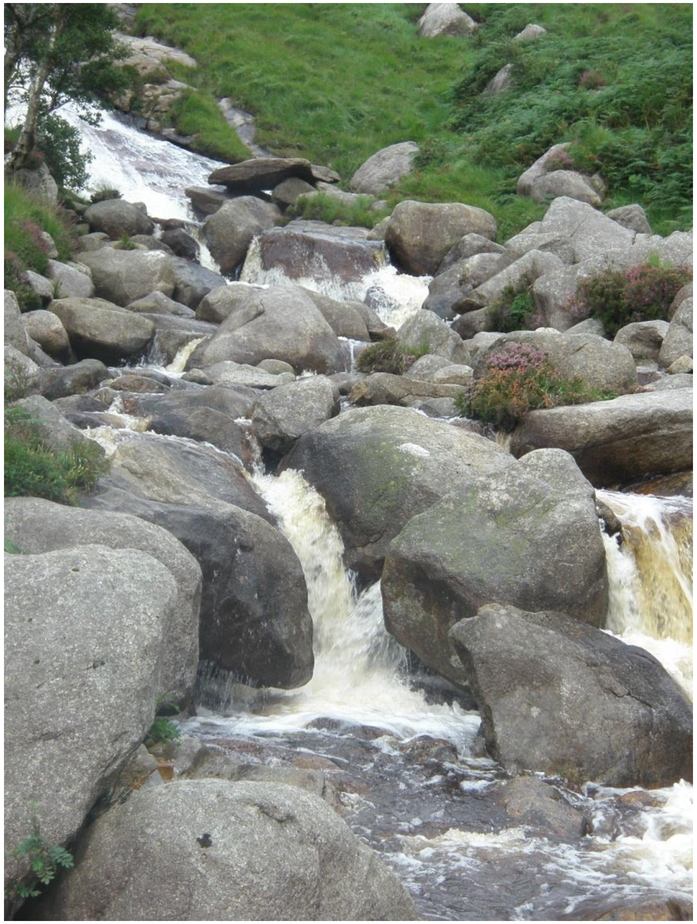
What new future challenges?

Data

- multi-pollutant concentration data from monitoring networks
- Many covariates: meteorological, land morphology, etc.
- Different data streams

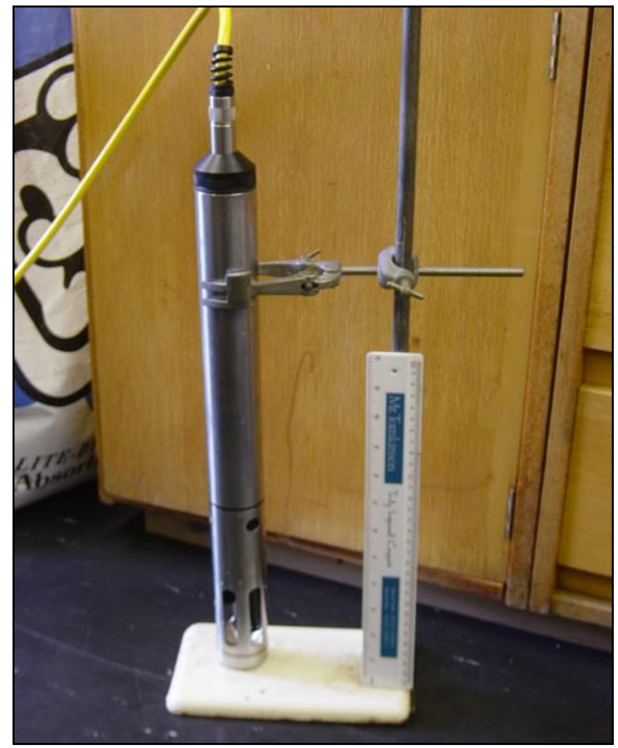
People and Society

- Ecosystems (services and values)-networks
- Social and economic factors
- Policy
- Sustainability



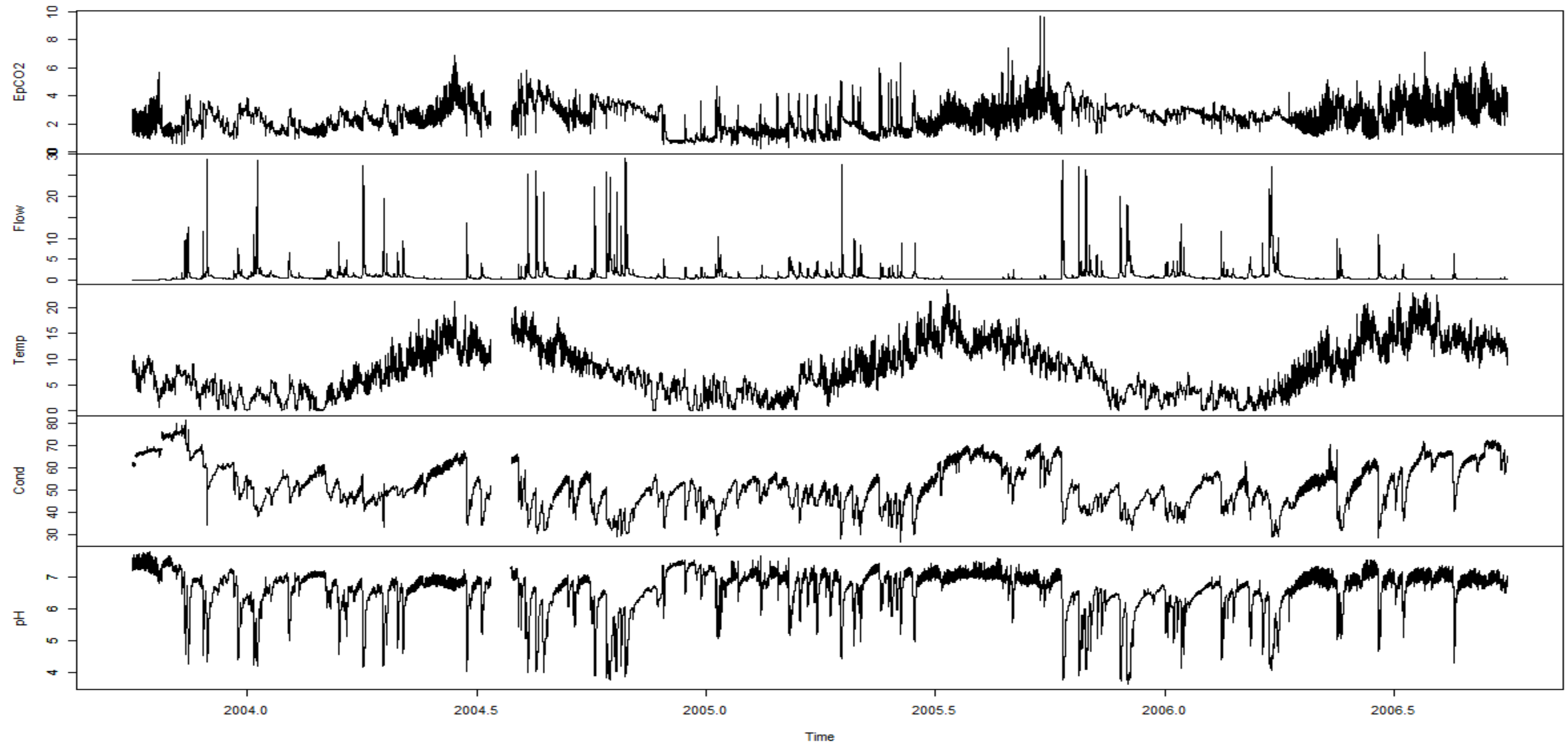
University of Glasgow | College of Science & Engineering

Making sense of the environment- partial pressure of carbon dioxide



Susan Waldron, carbon dioxide monitoring in freshwater catchments

patterns of interest



epCO$_2$, flow, temperature, conductivity and pH over 3 years

Amira El-Ayouty, current PhD student

University of Glasgow | College of Science & Engineering

What are the questions?

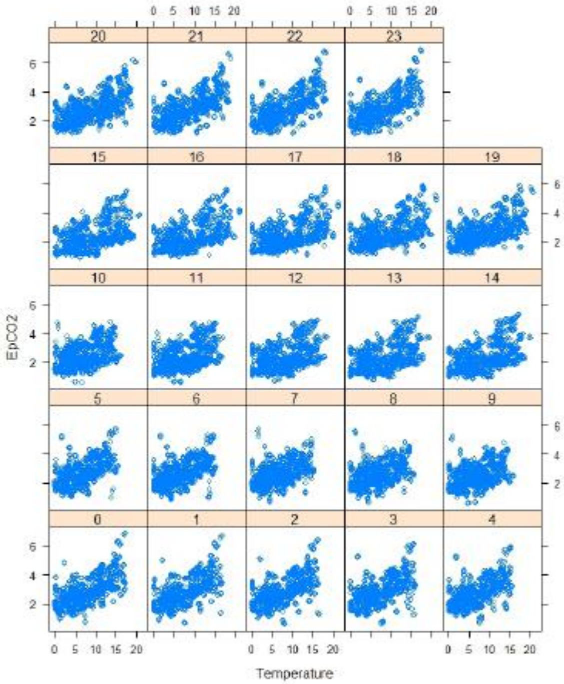
- **How is $epCO_2$ (or the measurand) changing?**
- **What are the drivers of change?**
- **What are the temporal and spatial scales of change?**

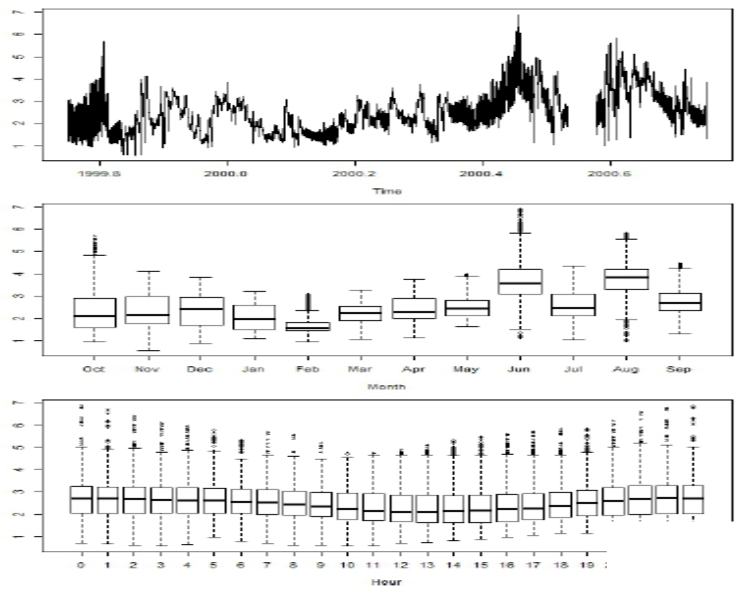
- **Events, anomalies, unusual conditions**

- **We begin by visualisation, and then follow with models to allow us to make inferences, predictions etc**

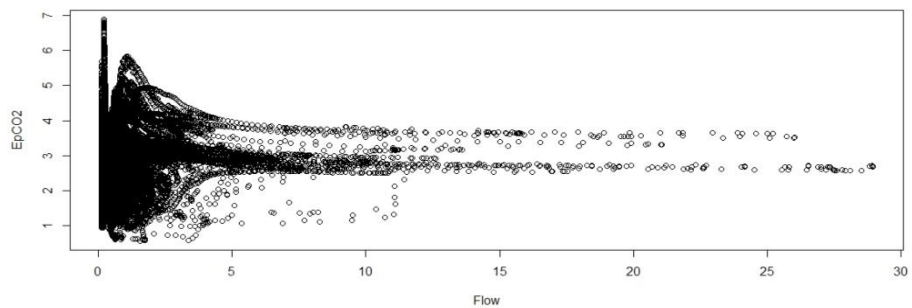
University of Glasgow | College of Science & Engineering

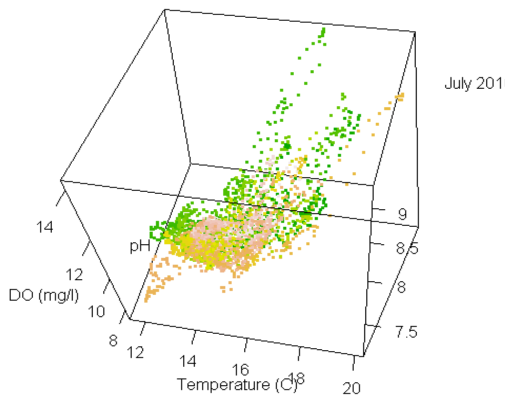
visualising the components (in time)



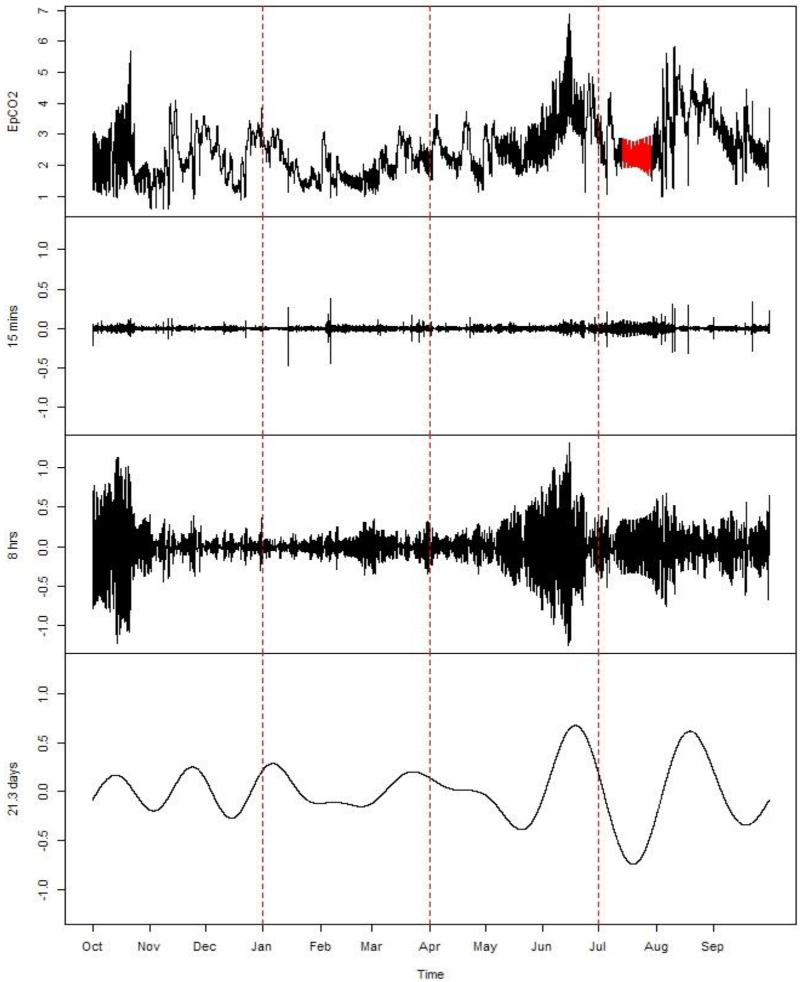


Complex and transient relationships





Making sense of the environment- the future



- 15 minute data, 1 hydrological year
- The highest variability is in the 8 hour signal. This reflects changes in photosynthetic / respiratory dominance, changing seasonally.
- The smoothed component relates to variations of 21+ days and higher- approx monthly?
- more variation during summer than winter reflecting differences in CO_2 input versus consumption.

Optimising sampling-functional data analysis

- 104 sites
- 30 distinct groupings
- 2 – 8 lochs in each group

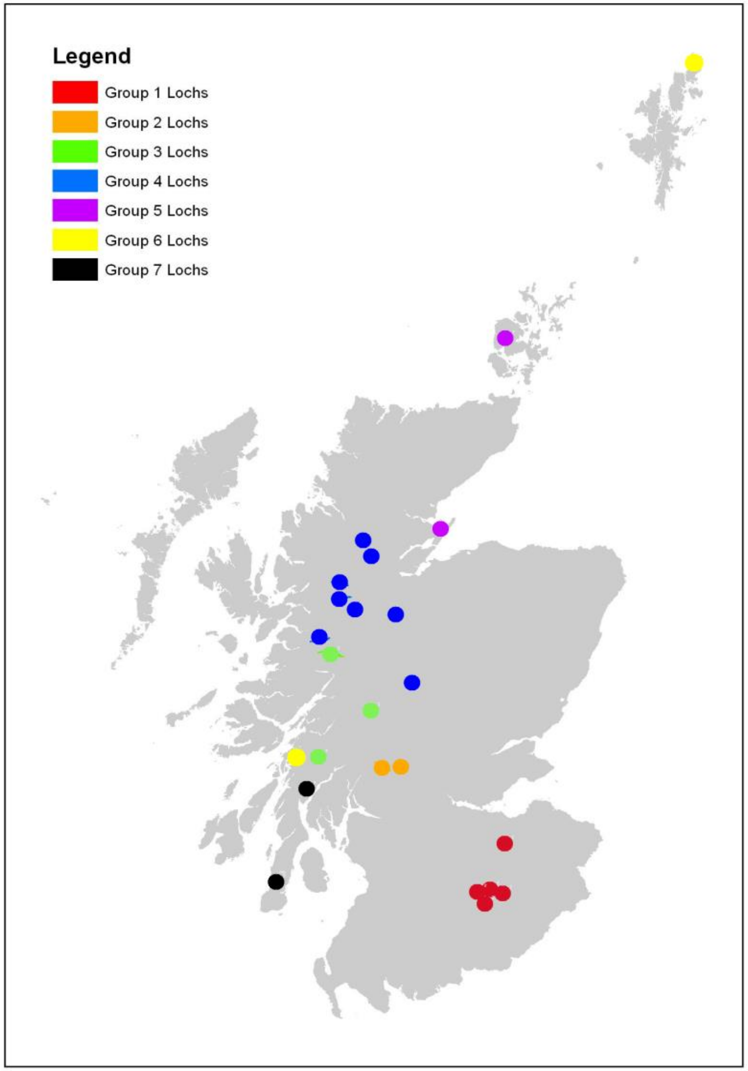
Focus on a subset of data;

- 7 groups
- 24 lochs

Alkalinity, TON, Nitrate, Phosphorus

Chlorophyll_a

Our goal: to use observed chemistry data to investigate different groups



Ruth Haggarty, PhD and SEPA, Haggarty et al, 2012

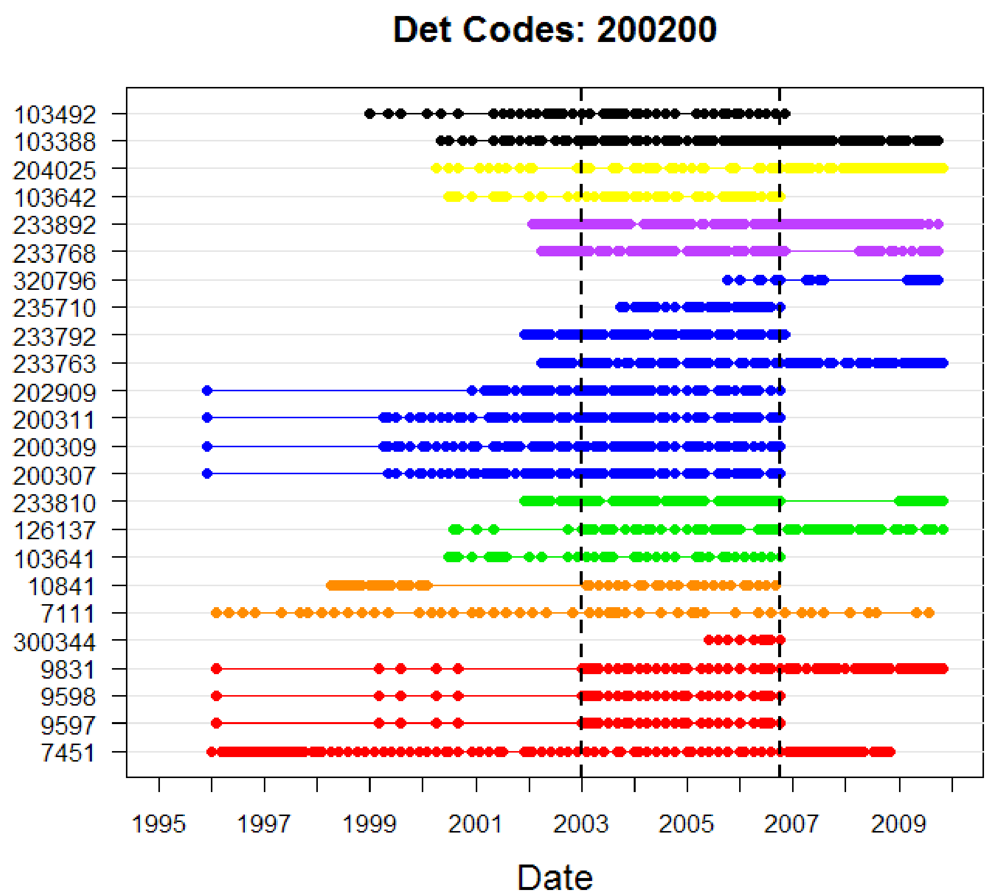
Comparing Sites

- Inconsistency in the quantity of data and time period covered

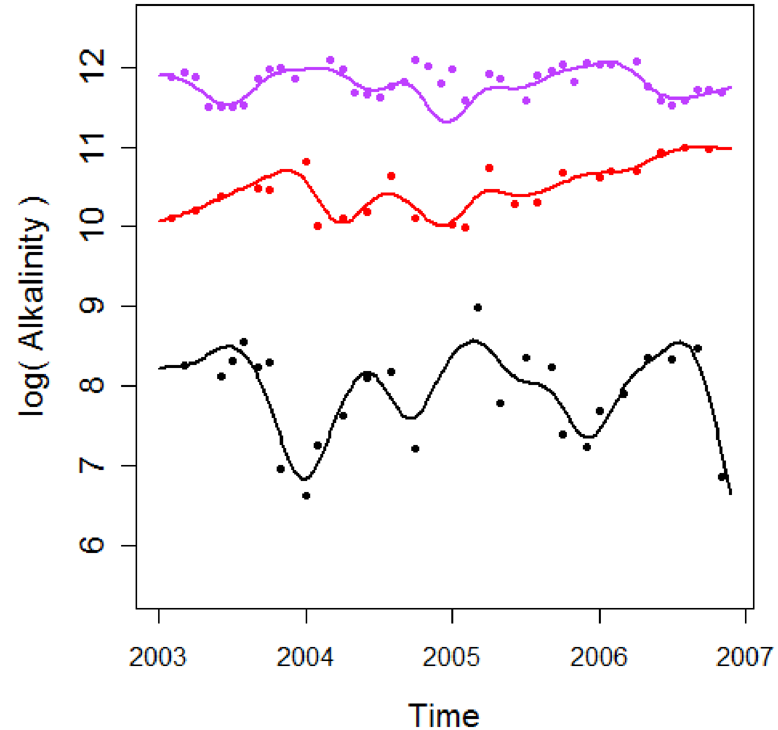
-matching by date or season inefficient

-More informative to compare trends and seasonal patterns in the lochs using smooth curves

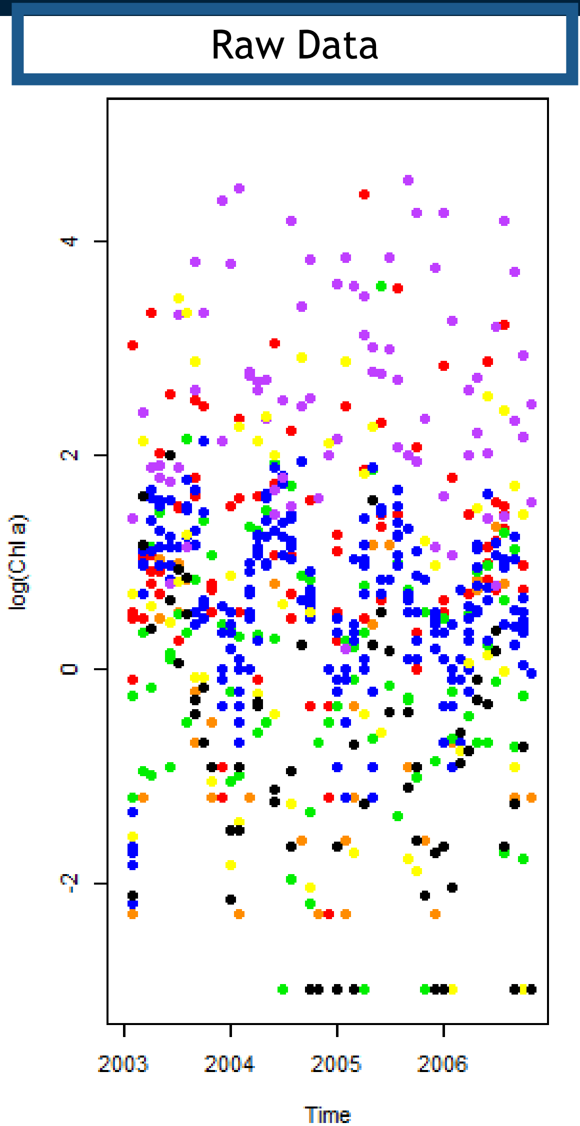
•Functional data analysis

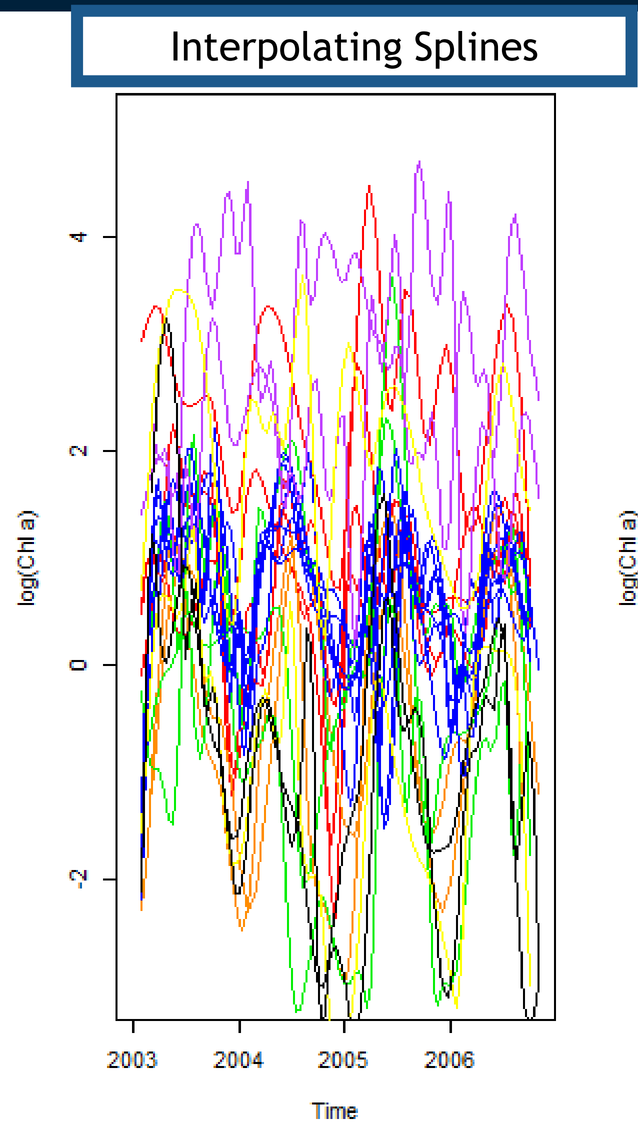


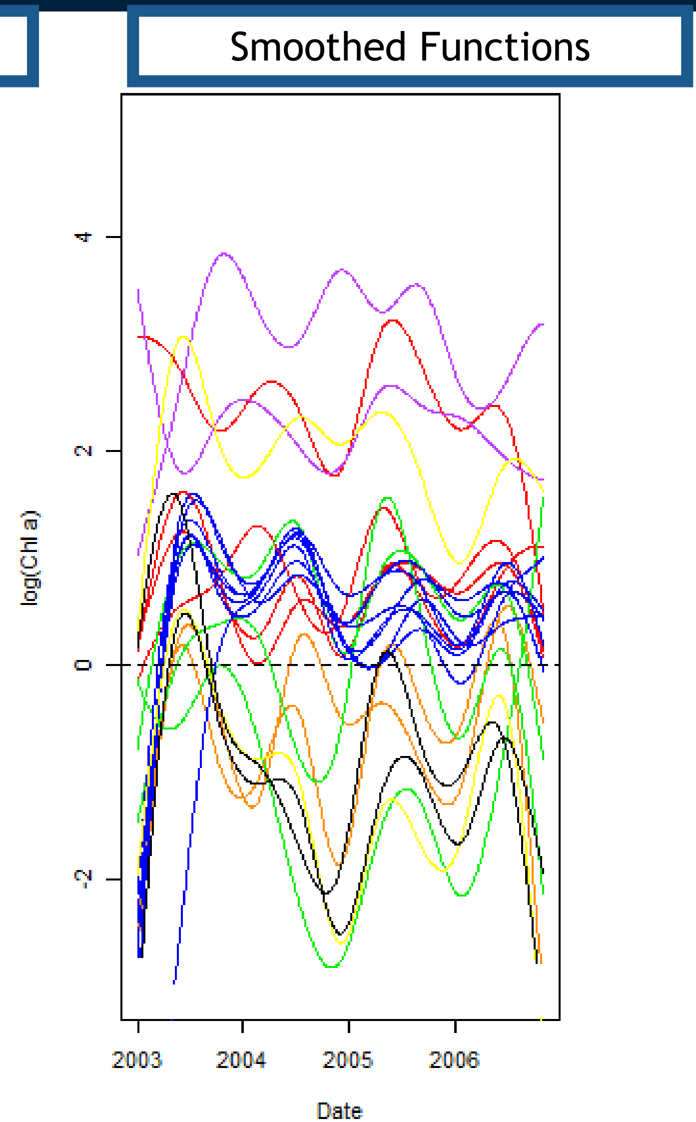
- **Compare seasonal patterns/trends in the data rather than observed values**
- **Overcomes some of the problems involved with matching dates**
  - Time series of data collected on each individual – each loch
  - These are measurements of a continuous function taken at a finite number of time points
  - Any observed trajectory can be viewed as a noisy measurement of an unobservable curve
  - **Fit a smooth curve to the points from each loch and cluster these curves**



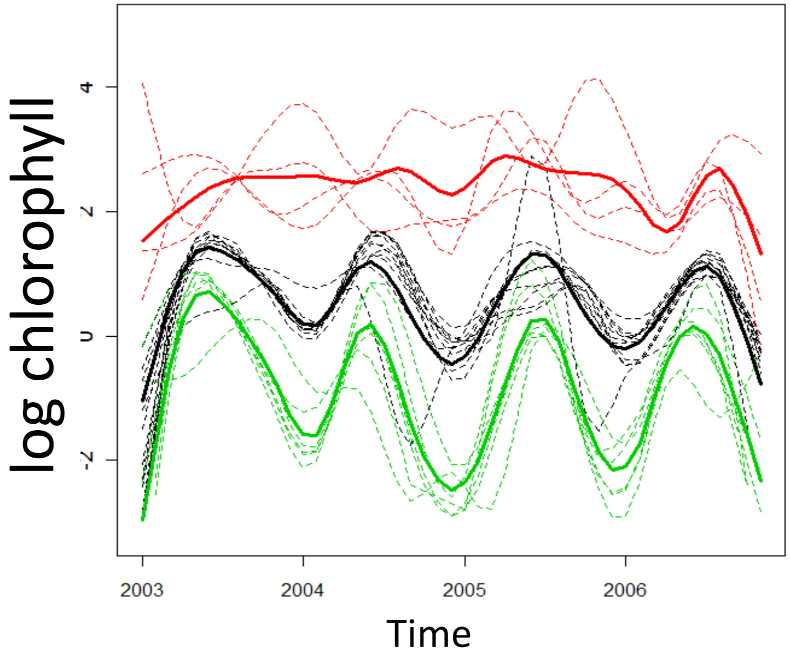
Functional Data





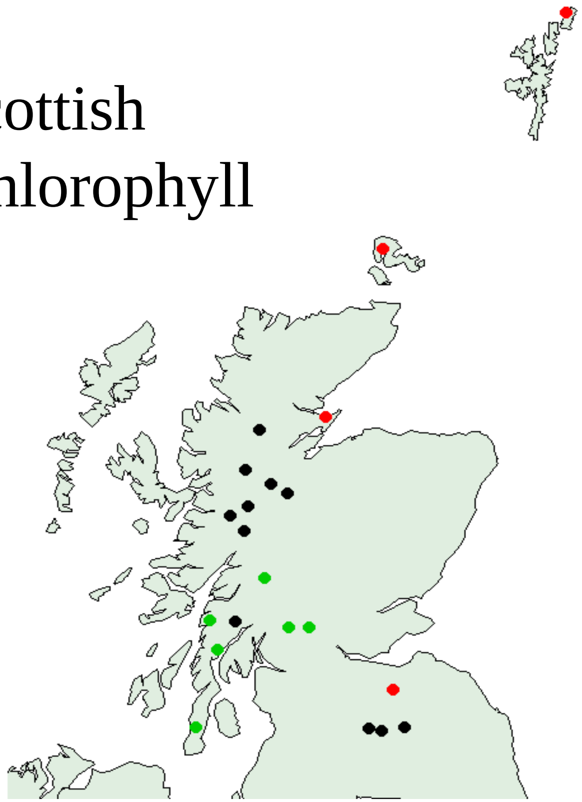


Identification of clusters of common signals by FDA



Cluster mean curves

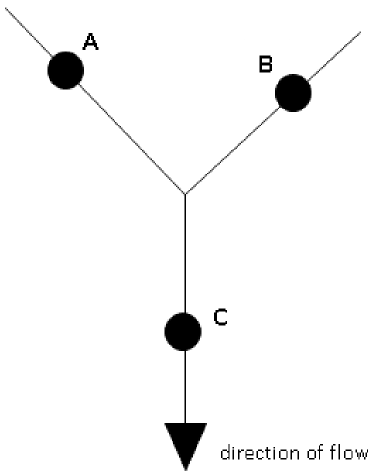
Clustering Scottish lakes using chlorophyll



Haggarty et. al (2012), Environmetrics

River networks

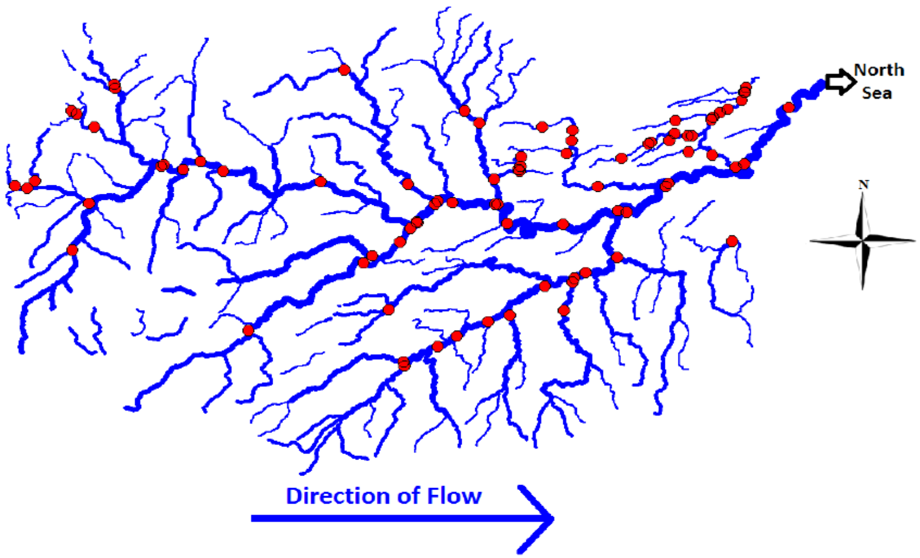
Spatial models for stream networks have recently been developed that include a spatial covariance structure based on stream distance rather than Euclidean distance (ver Hoef et al, (2006, 2010)).
The user can specify if monitoring sites are 'flow connected' (A and C or B and C) or 'flow unconnected' (A and B).

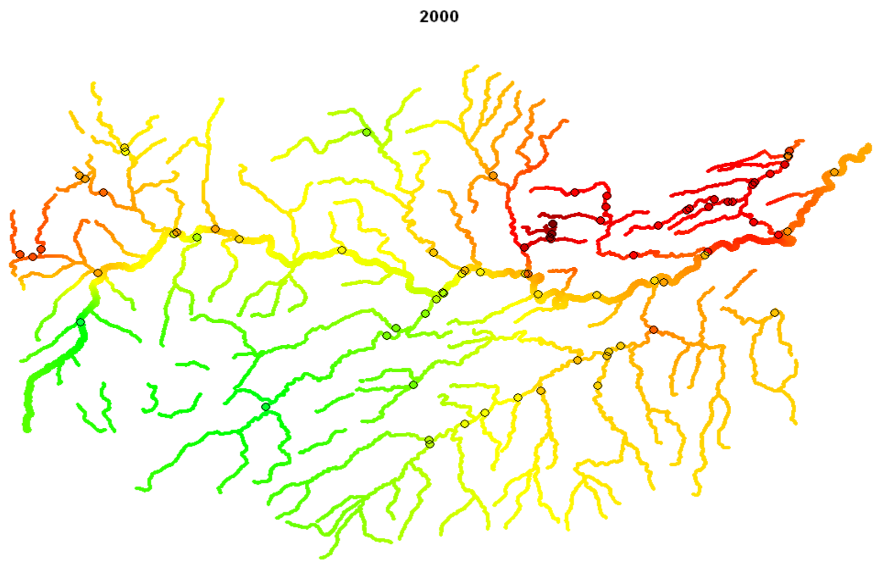


Flexible regression models over river networks, O'Donnell, Rushworth, Bowman, Scott and Hallard (2014)

University of Glasgow | College of Science & Engineering

Spatial patterns of change- at river basin scale





the circles represent the stations on the network, clearly not spatially representative

Joint work with David O'Donnell, Mark Hallard (SEPA), Adrian Bowman, Alastair Rushworth

Functional data analysis

The time series of data from each sensor can be regarded as a curve, the curve then becomes the "*data point*".

The statistical model is then based on the curves or functions which are assumed to be smooth. We can decompose the function to include trend, seasonal pattern, and relationships with other covariates.

FDA is very powerful and there are functional equivalents of many standard statistical techniques. We have been using such techniques to look at coherent spatial clusters.

University of Glasgow | College of Science & Engineering

TOC in Scottish rivers

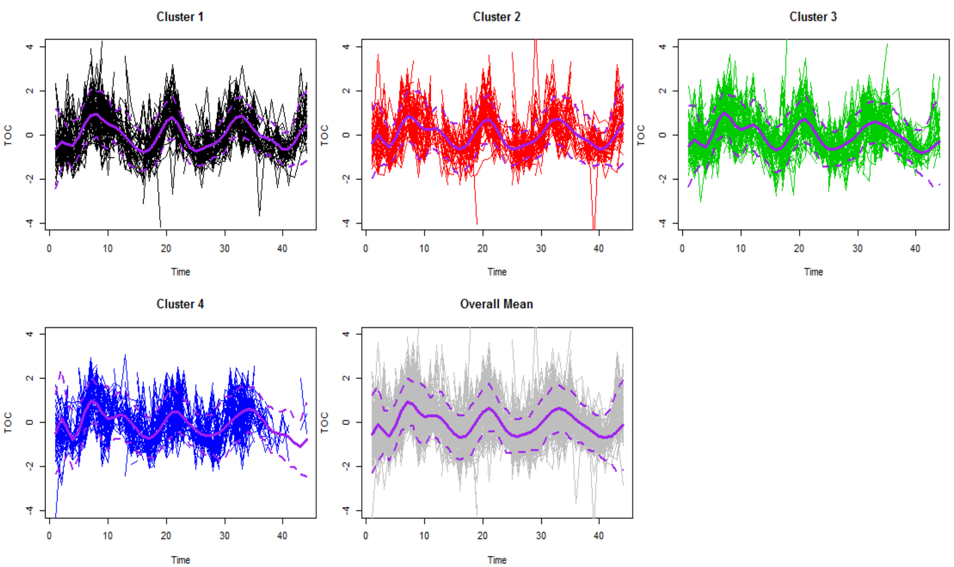
- **Functional clustering methodology has been applied to Total Organic Carbon (TOC) data from several hundreds of monitoring locations across rivers in Scotland over 44 months, covering the period January 2007 - August 2010.**
- **Each of the 333 river time series are first standardised, individually, to have zero mean and unit variance.**
- **Number of clusters and cluster membership are then estimated, and plotted spatially.**

Water @ Glasgow

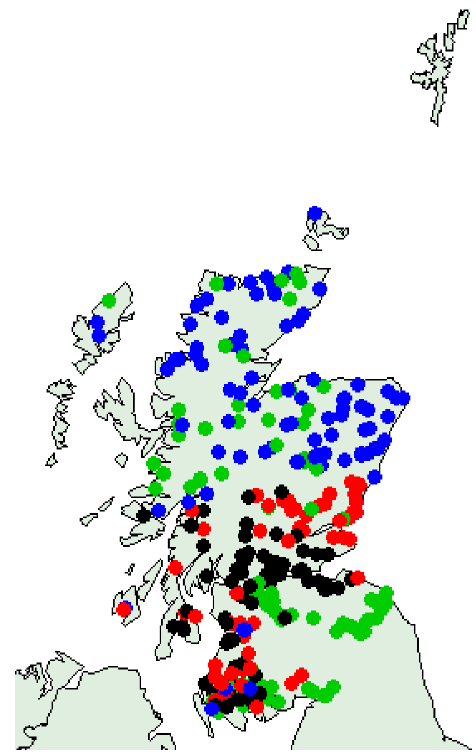
Stephen Reid, MSc, Francesco Finazzi, U of Bergamo, SEPA

University of Glasgow | College of Science & Engineering

TOC in Scottish rivers



Hierarchical (Euclidean correlation) Rivers

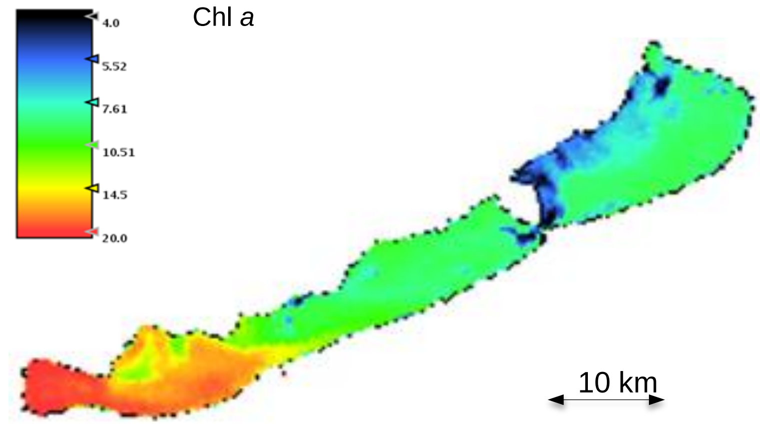


Water @ Glasgow

University of Glasgow | College of Science & Engineering

Globolakes

set up to investigate the state of lakes using satellite data for 20 years of lake temperature, suspended matter, chlorophyll to:



- Detect spatial and temporal trends and attribute causes of change
- Forecast lake sensitivity to environmental change

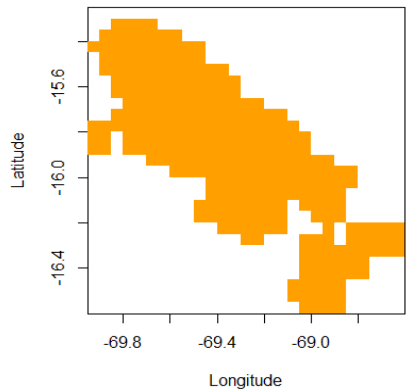
With Claire Miller and Ruth Haggarty, Globolakes consortium

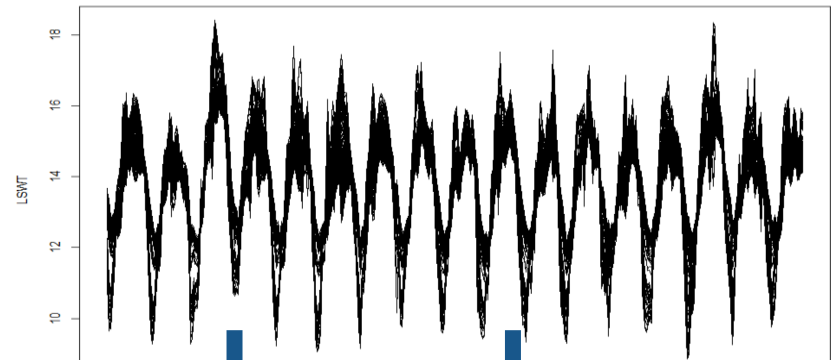
Centre for Ecology & Hydrology NATURAL ENVIRONMENT RESEARCH COUNCIL | University of Dundee | The University of Edinburgh | University of Glasgow | PML Plymouth Marine Laboratory | University of Stirling

University of Glasgow | College of Science & Engineering

GloboLakes

Pixels to Time Series



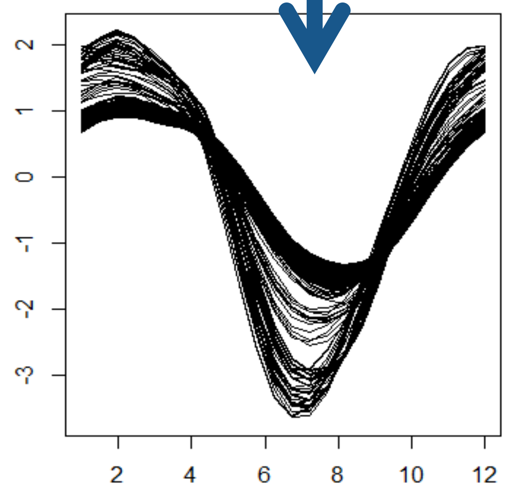


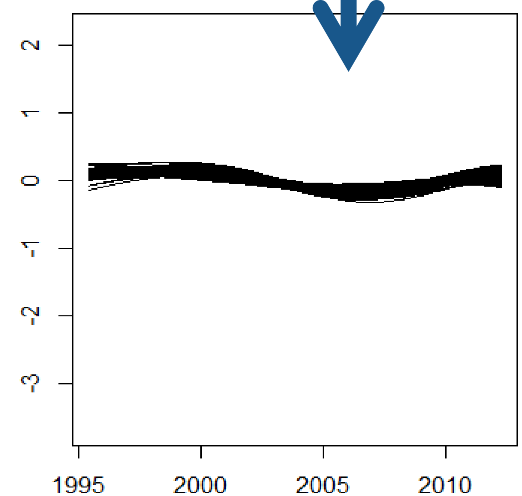
Each time series corresponds to a pixel

Each time series can be smoothed and an estimate of the trend and seasonal components can be obtained

Trend and seasonal components are centred

Seasonal pattern is dominant





Centre for Ecology & Hydrology
NATURAL ENVIRONMENT RESEARCH COUNCIL
University of Dundee
University of Reading
University of Glasgow
PML Plymouth Marine Laboratory
University of Stirling

On a global scale

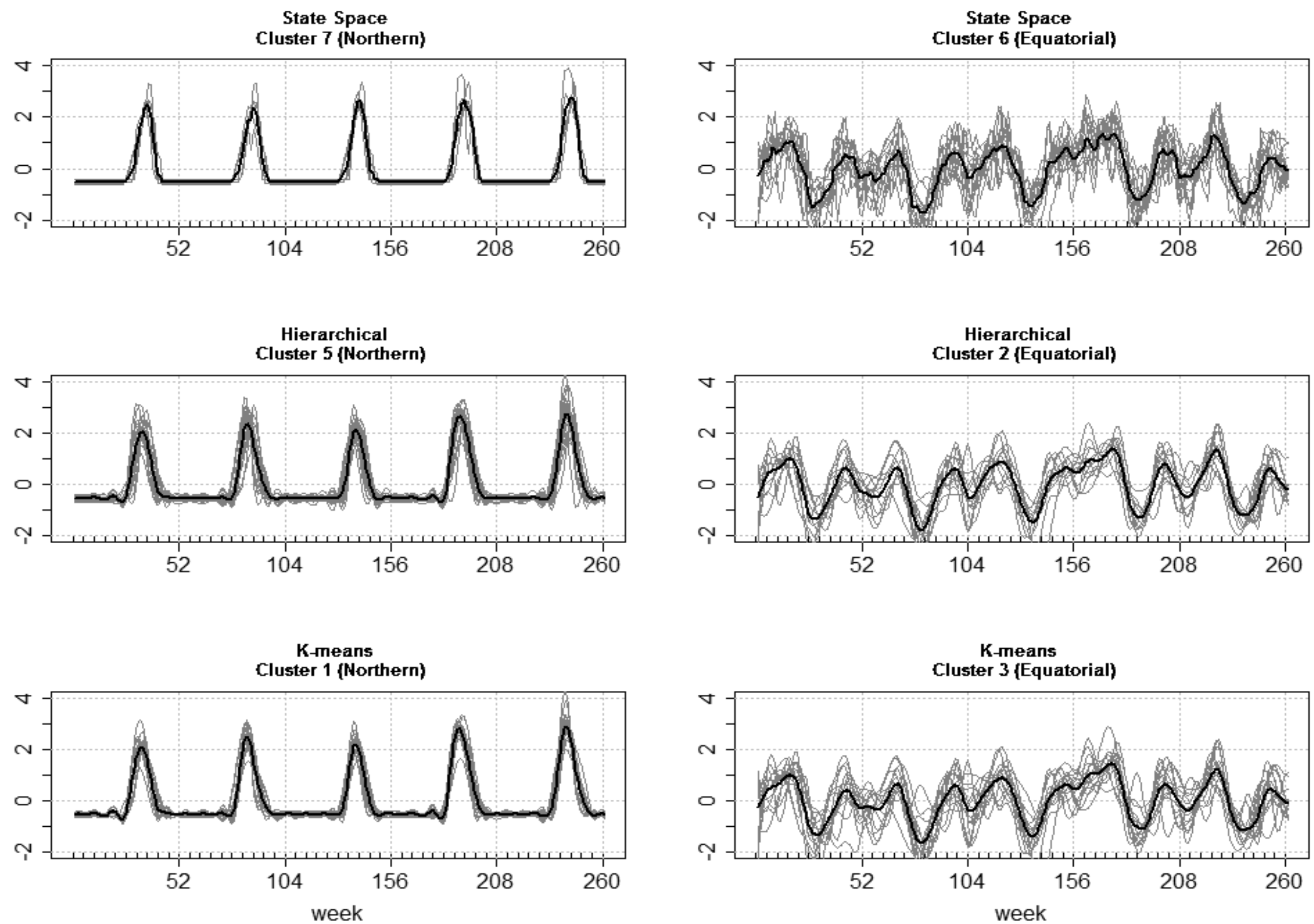
- Clustering approaches are applied to the LSWT time series of the ARC-Lake data set (www.geos.ed.ac.uk/arclake) in order to cluster the lakes into homogeneous groups with respect to their temporal coherence

- 5 years of weekly mean values were used in the analysis (2006-2010) for 261 lakes.

- functional clustering identified 11 clusters as optimal.

Cluster curves

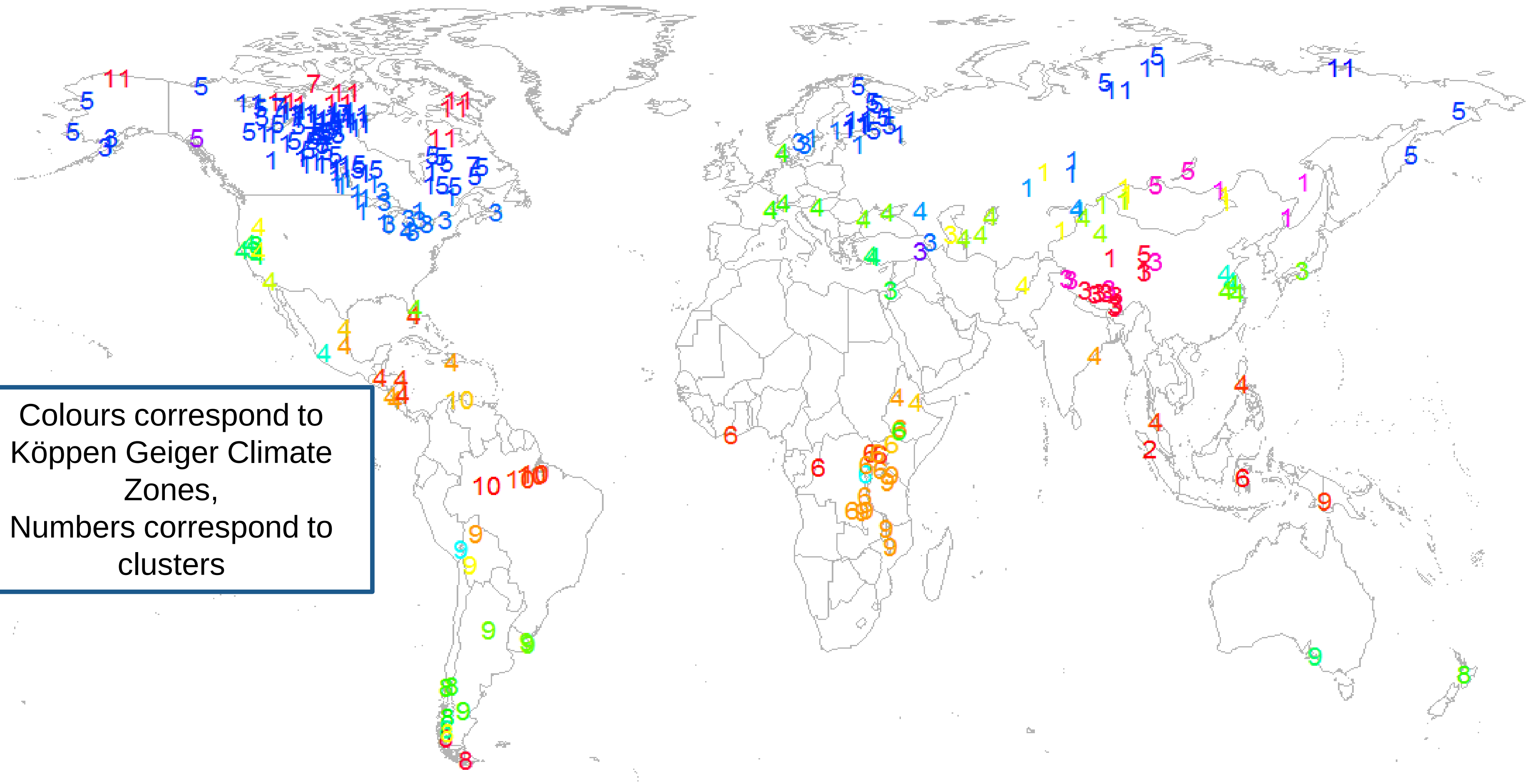
Each approach provides a different clustering result, however, the temporal patterns they identify are similar. Results for two clusters are shown.



Finazzi et al, 2014

University of Glasgow | College of Science & Engineering

Spatial cluster structure



Colours correspond to Köppen Geiger Climate Zones,
Numbers correspond to clusters

challenges

- **Data characteristics- quantity and quality, missingness**
- **Non stationary, complex nature of the relationships**
- **For networks of sensors- building fast and efficient spatio-temporal models, functional data analysis provides part of that solution**
- **Aspects of scale and aggregation**
- **uncertainty evaluation and visualisation multi-disciplinarity**

Water @ Glasgow

references

- O'Donnell D, Rushworth A, Bowman A W, Scott E M, Hallard M (2014) Flexible regression models over river networks. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*
- Haggarty R, Miller C A, Scott E M, Wylie F, Smith M (2012) Functional clustering of water quality data in Scotland. *Environmetrics*
- Finazzi, F., Haggarty, R., Miller, C., Scott, M., Fasso, A. A comparison of clustering approaches for the study of the temporal coherence of multiple time series, *Stochastic Environment Research and Risk Assessment* .
- Miller C, Magdalina A, Willows R, Bowman A, Scott E M, Lee D, Burgess C, Pope L, Pannullo F, Haggarty R (2014). Spatiotemporal statistical modelling of long term change in river nutrient concentrations in England and Wales. *Sci Tot Env, 466-467.*