University of Glasgow | College of Science & Engineering

Water @ Glasgow

Establishing the evidence base: quality requirements for monitoring networks

E Marian Scott

School of Mathematics and Statistics

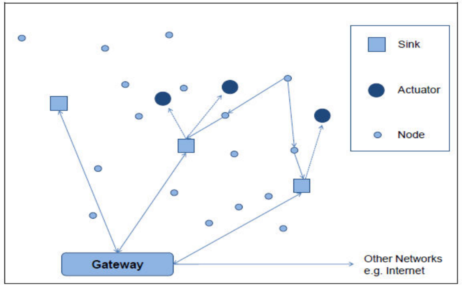
WWC, May 2015

With Adrian Bowman, Claire Miller, Kelly Gallacher, Ruth O'Donnell, Mengyi Gong, Craig Wilkie, Alistair Rushworth

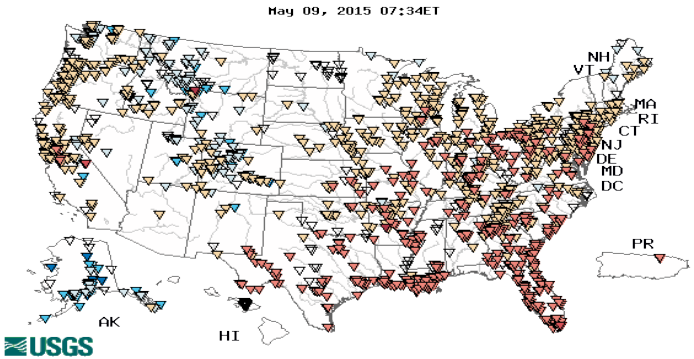
Networks in space and through time

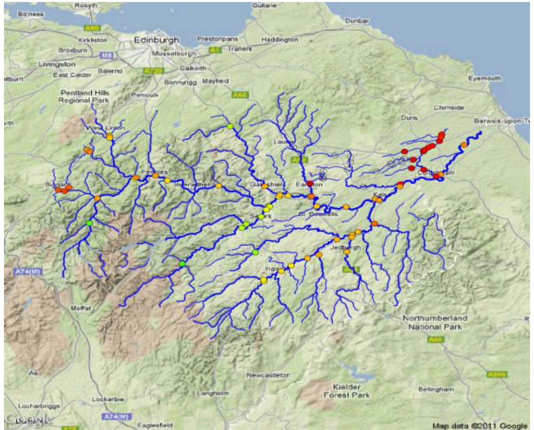
A set of monitoring sites- which may be physically connected

Sampling through time and over space, often sparse in space and not always representative



Source: OECD based on Verdone *et al.*, 2008.



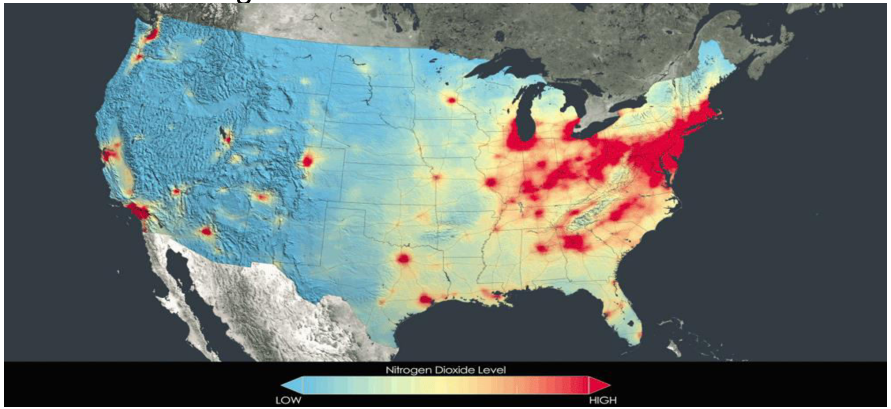


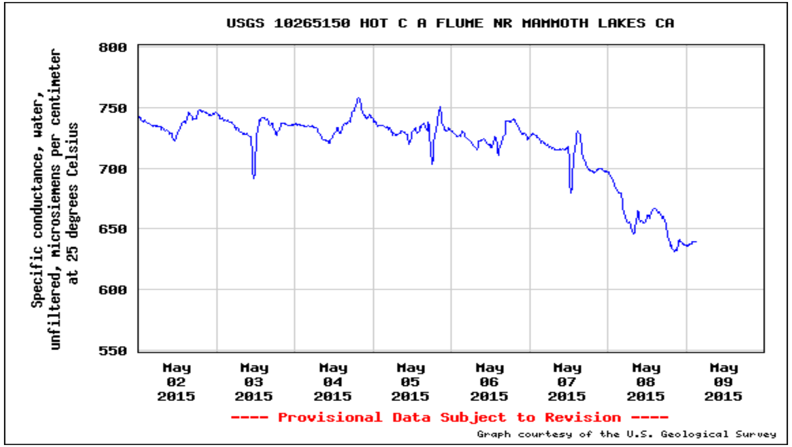
University of Glasgow | College of Science & Engineering

Networks in space and through time

[Water-Quality Watch]() displays real-time water quality data collected remotely by sensors installed in rivers, and lakes. Readings taken every 5 to 60 minutes are transmitted via satellite to the USGS National Water Information System (NWIS). Data include water temperature, pH, specific conductance, turbidity, dissolved oxygen, and (or) nitrate depending on the site.

www.nasa.gov





analysis

- Visualisation and modelling needs to include space and time, and also should reflect the connectedness of the sites.
- Spatial interpolation, temporal trends, spatio-temporal modelling-all deliver tools to address questions about change and trends, effects of interventions but
- (*things close together in space, or time are more alike than things which are far apart*), so our models must account for the existence of spatial and temporal correlations
- Quality and quantity of data can challenge- dealing with missingness and sparseness

River networks

With SEPA

Our goals

- Estimation at unmonitored locations
- Determination of overall changes in mean levels
- Quantification of uncertainty
- Insight into current sampling schemes

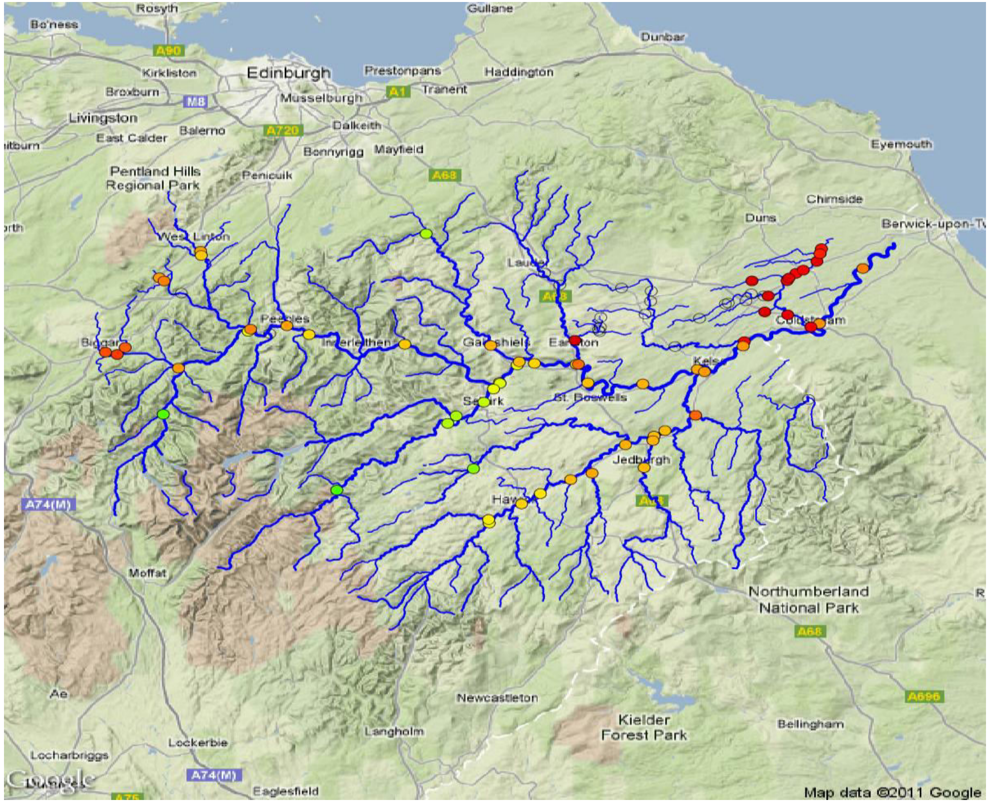
Example

River Tweed, 83 sites

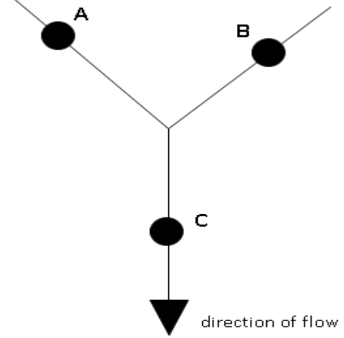
Nitrate data from 1986 - present

>16,000 data points

River networks



Spatial models for stream networks using stream distance rather than Euclidean distance .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)

networks-temporal patterns

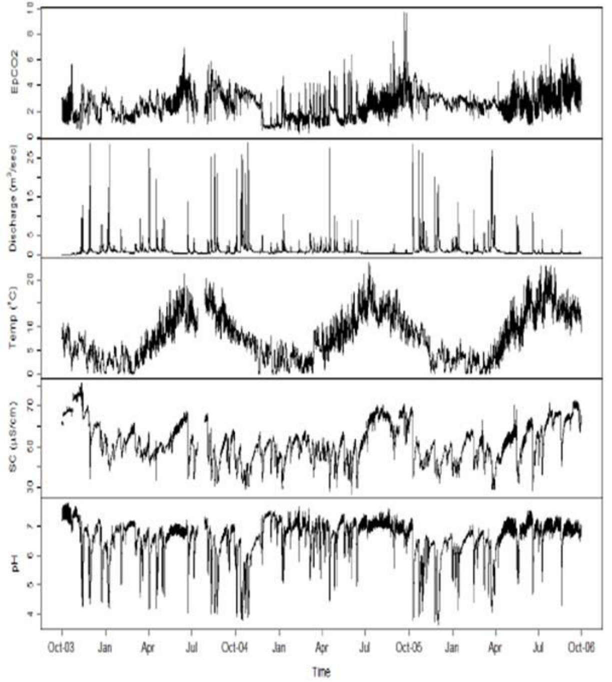
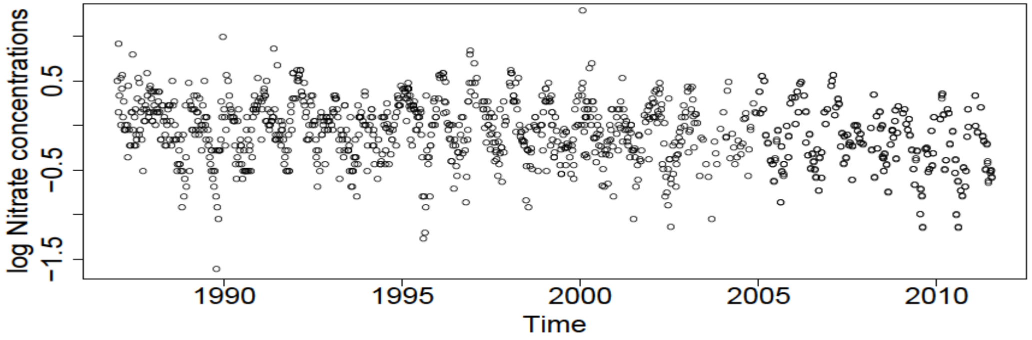


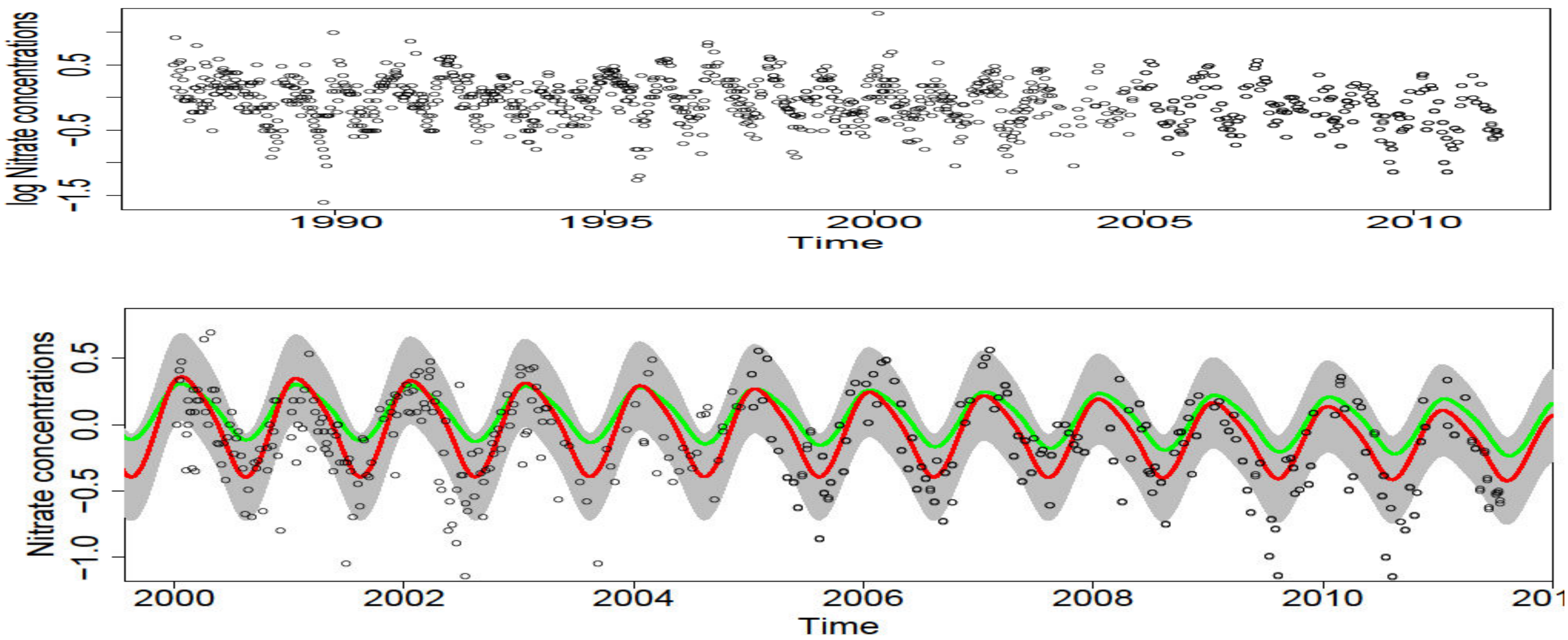
Fig. 1 Time plots of the 15 minute frequency series of the $EpCO_2$, Flow, Temperature, pH and SC series from October 2003 to September 2006.

- Irregular sampling in time, with more regular sampling frequency - large amounts of data
- Irregular in space; preferentially sampled, monitoring sites change in time

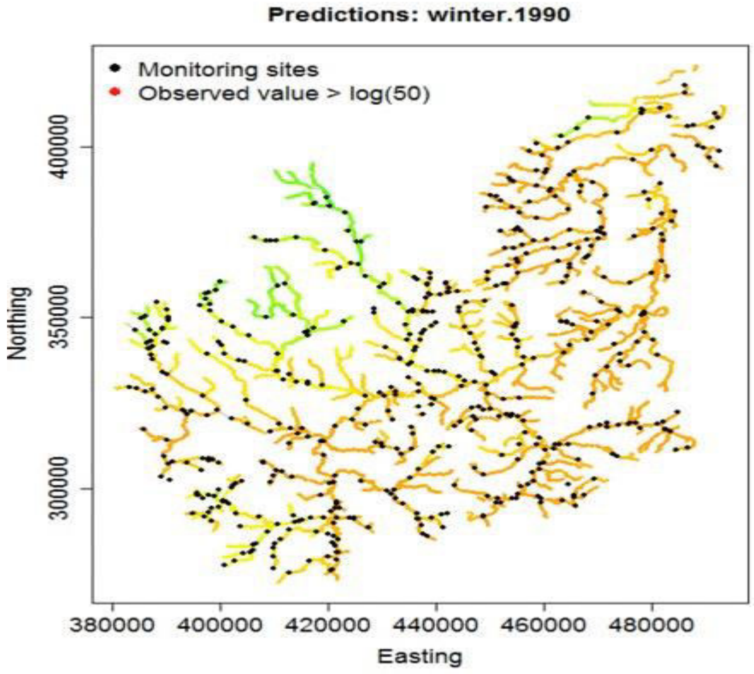


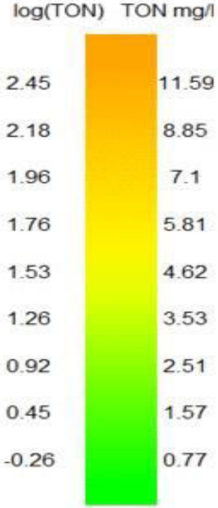
River networks- building models, single site model

Pollutant = Spatial + Seasonal + Trend + Local Effects+ Error



River networks, multiple sites





The following models were fitted to the dominant network in Trent:

$$log(TON) = Easting + Northing + \varepsilon$$

$$log(TON) = Easting + Northing + z_e + \varepsilon$$

$$log(TON) = Easting + Northing + z_u + \varepsilon$$

$$log(TON) = Easting + Northing + z_u + z_e + \varepsilon$$

Kelly Gallacher, PhD student, joint with Claire Miller, Robert Willows (EA)

University of Glasgow | College of Science & Engineering

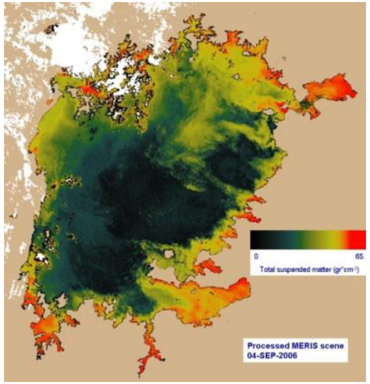
Networked data streams

GloboLakes

Earth Observation systems and the increased capability to retrieve in-water constituents provides exciting new data sets and statistical challenges. A hierarchy of linked data streams.

Globolakes is a 5 year consortium NERC project, to investigate the state of 1000 lakes using a 20 year archive of satellite based observations.

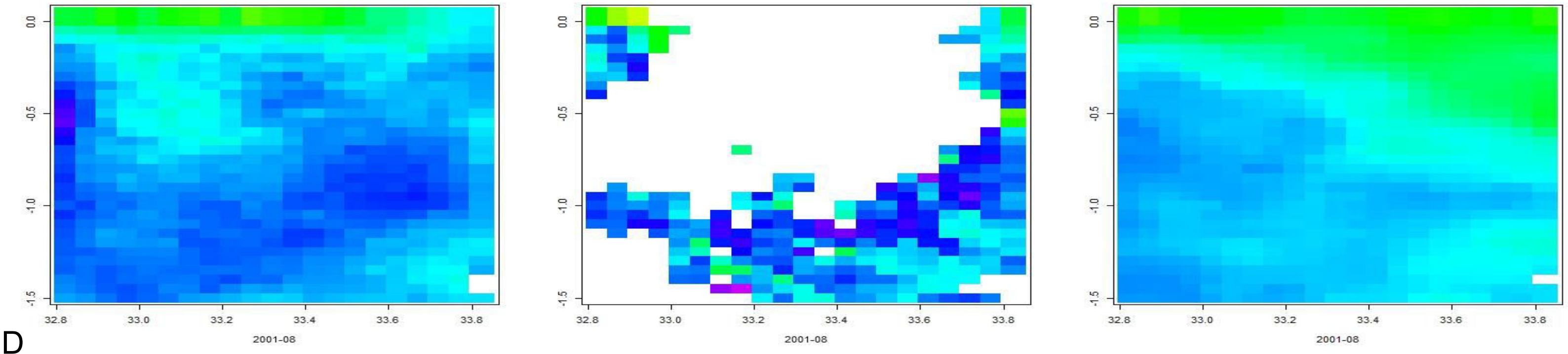
A key aim is to identify patterns of temporal coherence for individual remotely sensed lake characteristics and the spatial extent of coherence.



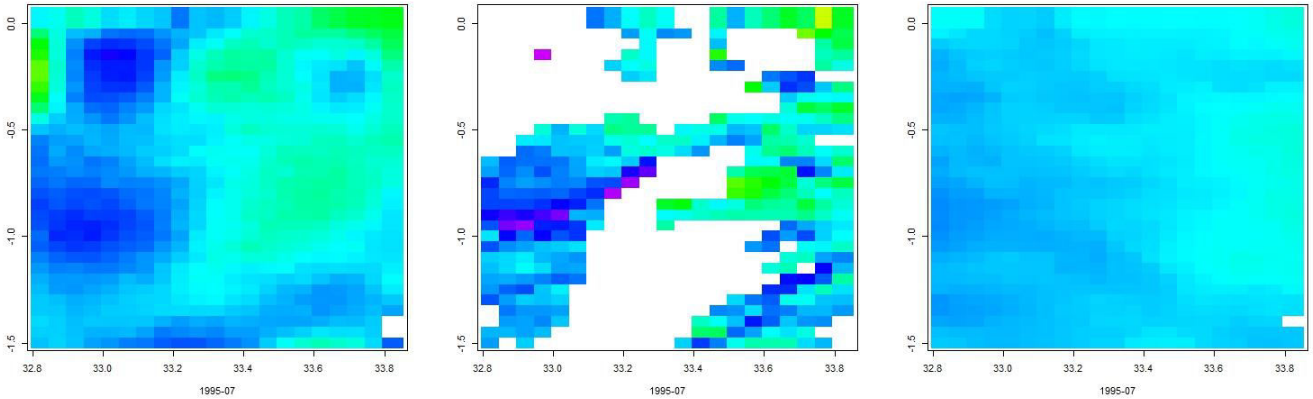
Centre for Ecology & Hydrology
NATURAL ENVIRONMENT RESEARCH COUNCIL
University of Dundee
University of Reading
University of Glasgow
PML Plymouth Marine Laboratory
University of Stirling

University of Glasgow | College of Science & Engineering

Lake Victoria



Mengyi Gong, PhD

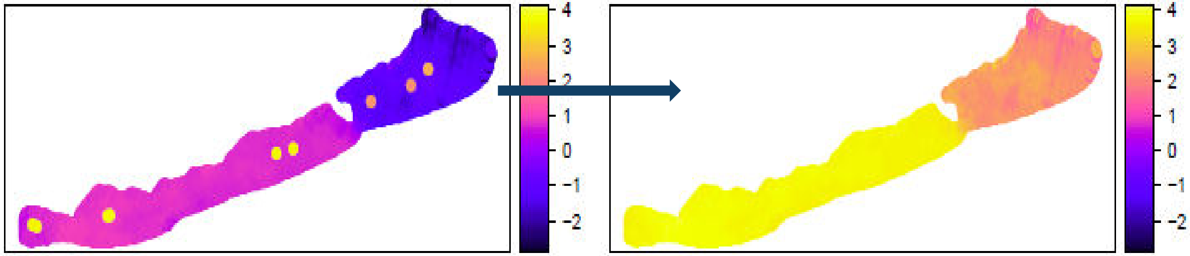


Missing data challenges

a) Statistical model,
b) data,
c) reconstruction

Lake Balaton

In-situ, satellite (left) through data assimilation and calibration to chlorophyll (right)



Network challenges: data linkage of satellite (grids of 300m or 1km), in-situ (spot), cloud cover hence missingness, spatially sparse, temporal heterogeneity. Statistical hierarchical models to combine the data streams and to quantify uncertainty

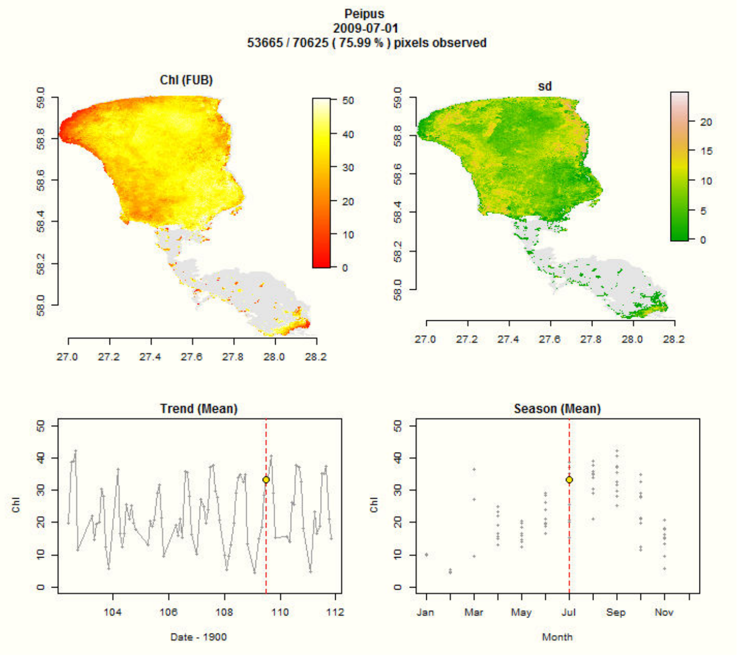
GloboLakes

Craig Wilkie, PhD

Data challenges

Dealing with the quantity of data in terms of the number of pixels, combined with the sparsity of the data, in terms of the time series observed will be a huge challenge.

Modern functional data approaches are suitable for large numbers of time series of potentially noisy data and enable clusters of curves to be identified which are coherent in terms of temporal dynamics.



Chlorophyll at Lake Peipsi, Estonia
Monthly obs, 2002-2012
(Data from DIVERSITY II project)

Opportunities and challenges

From a network we can learn:

- whether there are seasonal patterns and trend
- Whether the patterns are different at different locations
- How to make predictions at any point on the network, regardless of whether it is near a monitoring site

challenges

- Data characteristics- quantity and quality and relatedness
- Non stationary, complex nature of the relationships
- For networks, how to build fast and efficient spatio-temporal models,
- Designing the network and the resulting power to detect change

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*
- 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.*

Acknowledgements

SEPA for data and support (RO), Environment Agency (KG), Globolakes for CM, GU for CW and MG

Software:

smnet an R package for additive spatial modelling of river network data and SSN