University of Glasgow | College of Science & Engineering

Water @ Glasgow

Water analytics- the evidence chain from measurement to model to decision

E Marian Scott

School of Mathematics and Statistics

WWC, May 2015

University of Glasgow | College of Science & Engineering

Co-authors

C. A. Miller, R. A. O'Donnell, K. Gallacher, A. Elayouty

School of Mathematics and Statistics, University of Glasgow, UK

Maria Franco Villoria

Department of Economics and Statistics, University of Turin, Italy

Francesco Finazzi

Department of Statistics, University of Bergamo, Italy

Acknowledgements

SEPA for data and support (MFV), Susan Waldron, University of Glasgow funding (AE and KG) , EA (KG), Scottish Water (support for MFV, RO)

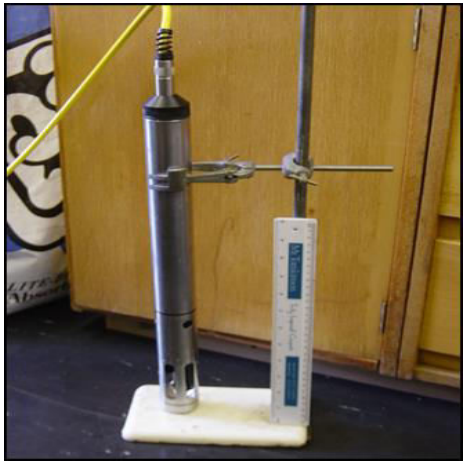
University of Glasgow | College of Science & Engineering

The state of the environment

Four questions:

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

Sensor technology delivering enhanced dynamic detail of environmental systems at unprecedented scale



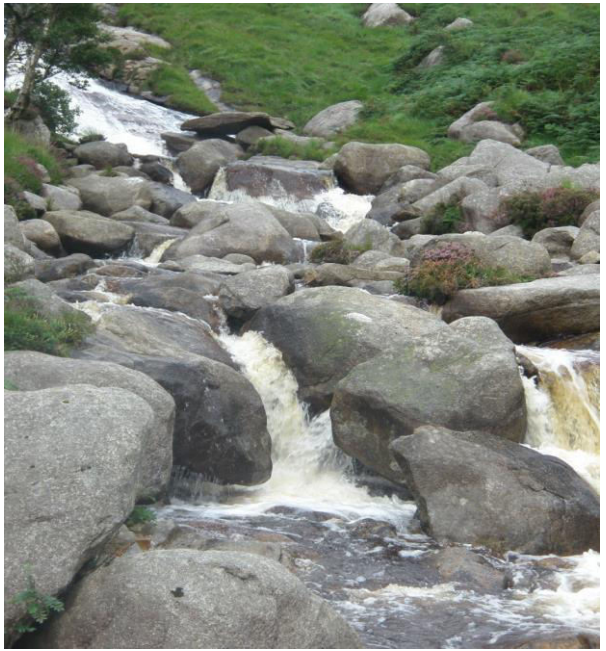
University of Glasgow | College of Science & Engineering

What analytic challenges?

The data needs are driven by policy, regulation and management.

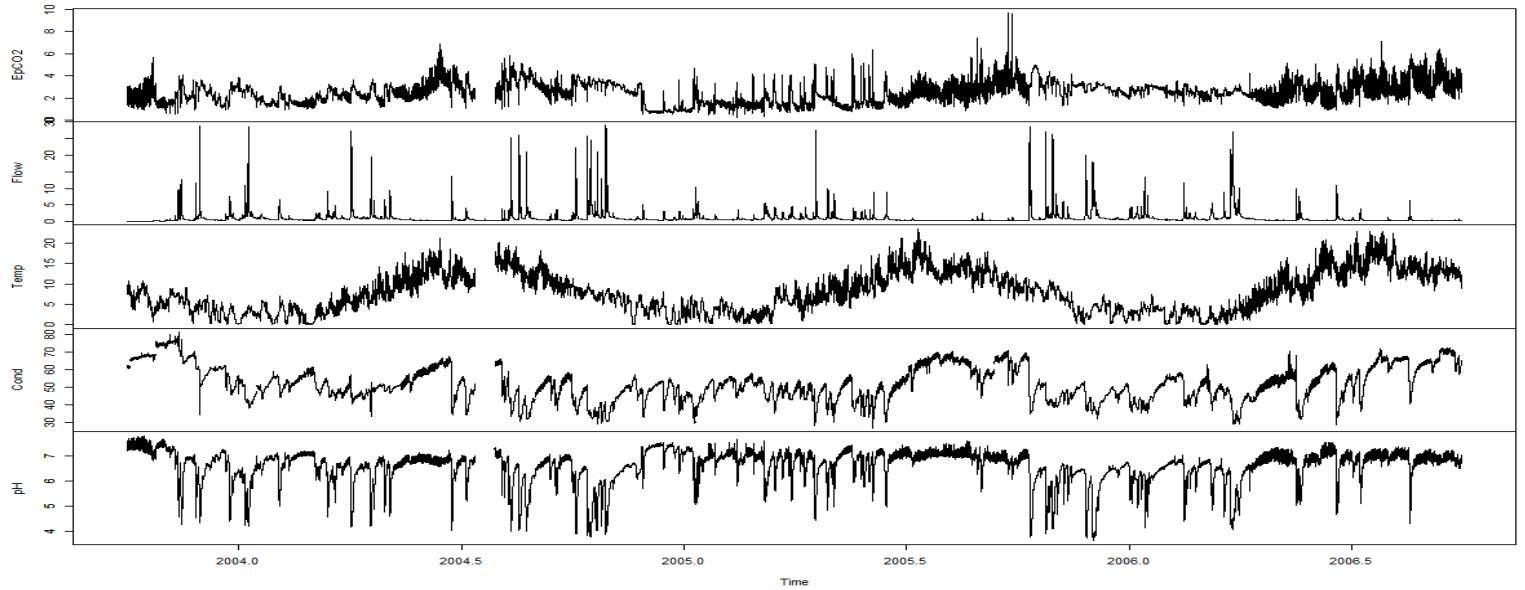
Analytic challenges:

- multi-pollutant data from monitoring networks
- Many covariates: meteorological, land morphology, from different data streams
- Dealing with uncertain climate change (mitigation and adaptation)
- BIG data-Real time analysis for control and management



University of Glasgow | College of Science & Engineering

patterns of interest



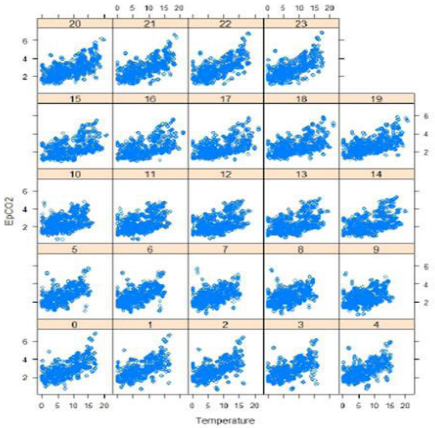
Partial pressure CO_2, flow, temperature, conductivity and pH over 3 years- Why? understanding carbon dynamics in small streams

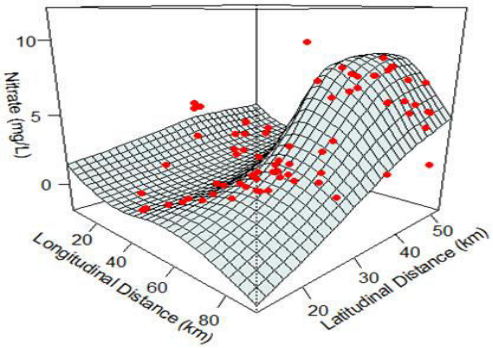
Amira El-Ayouty, current PhD student

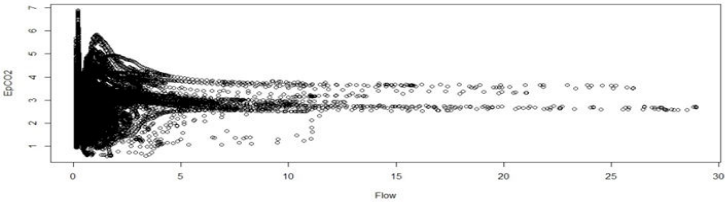
University of Glasgow | College of Science & Engineering

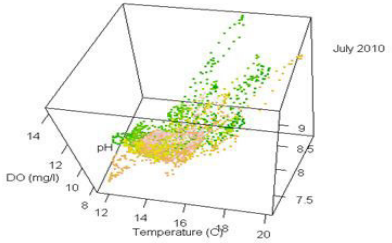
What are the questions?

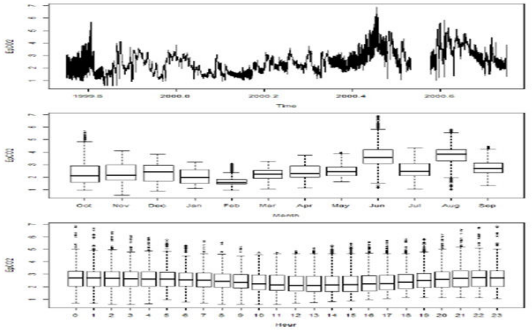
- How is $epCO_2$ (or the measurand) changing?
- What are the drivers of change?
- Events, anomalies, unusual conditions





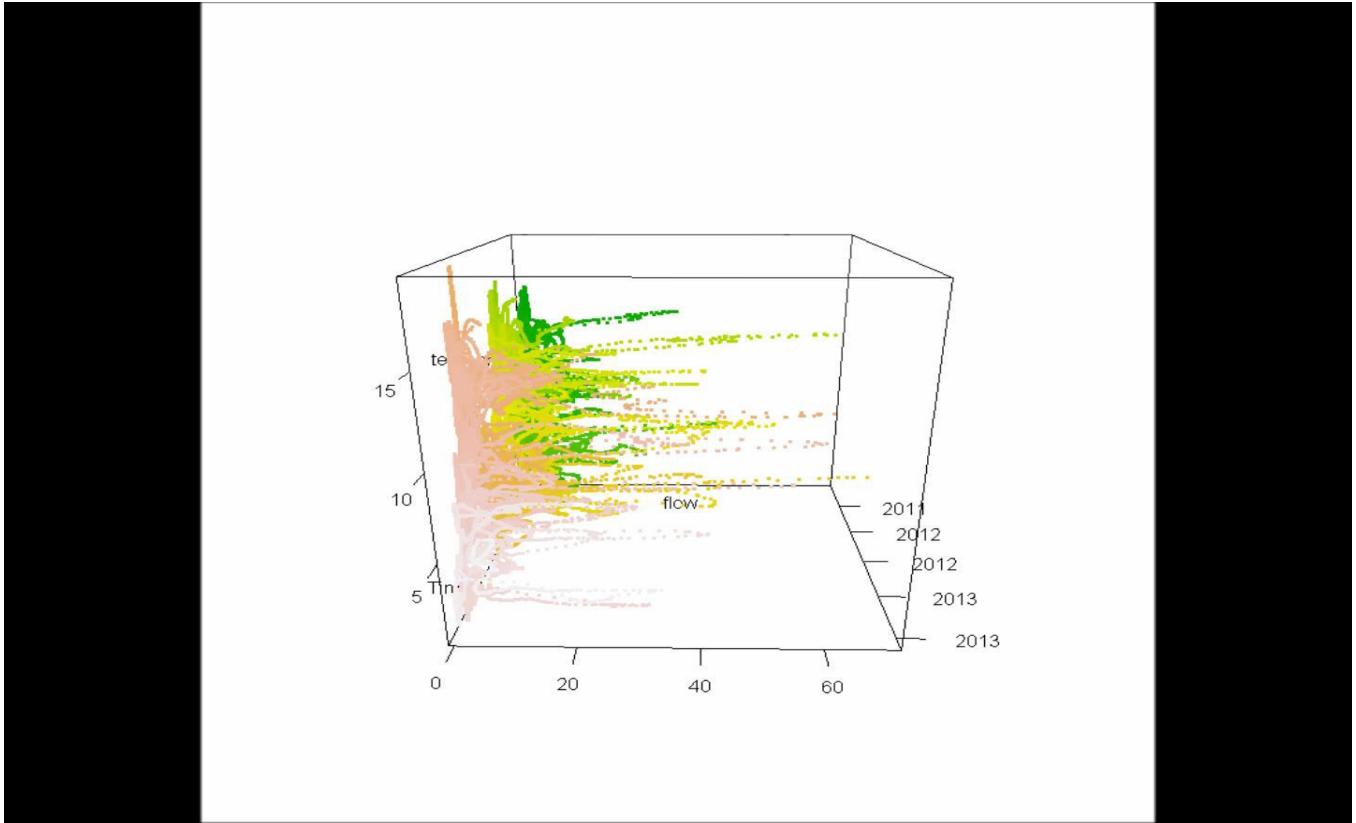






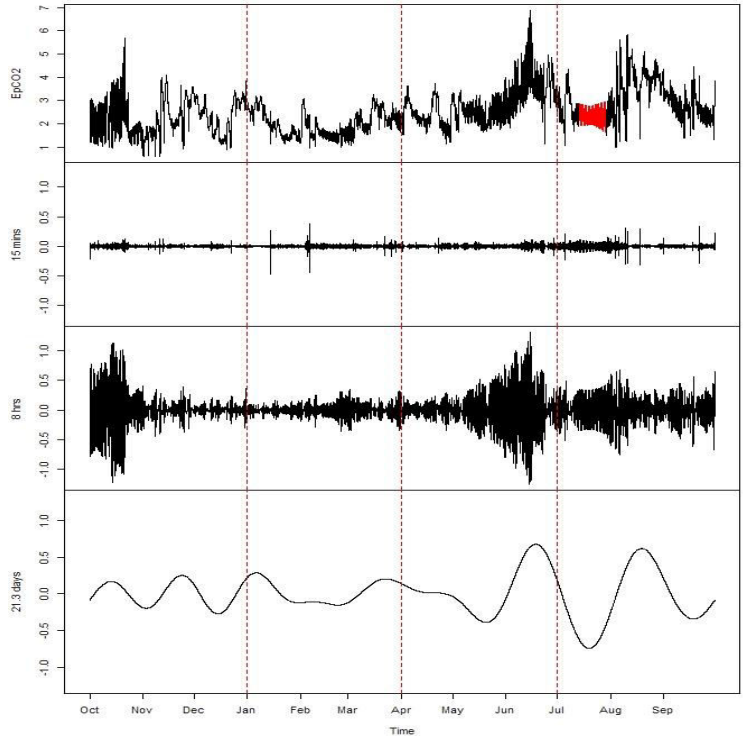
University of Glasgow | College of Science & Engineering

Visualising and animating



University of Glasgow | College of Science & Engineering

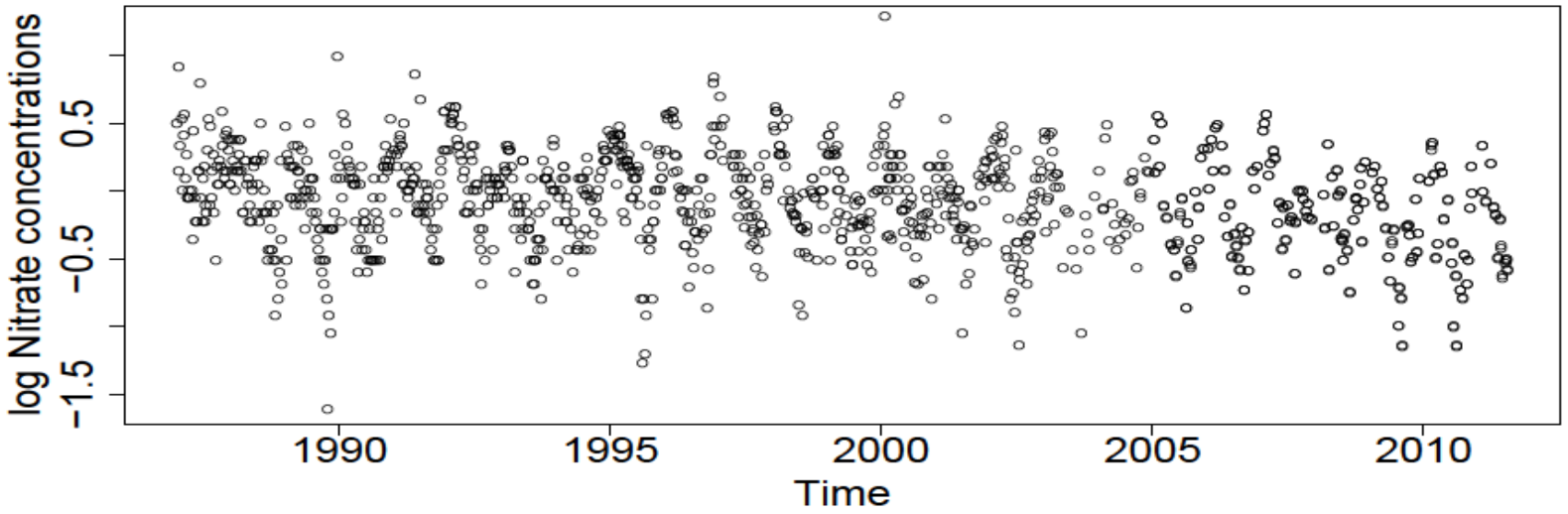
Making sense of the environment



- 15 minute data, 1 hydrological year- a wavelet analysis
- The highest variability is in the 8 hour signal. This reflects changes in photosynthetic / respiratory dominance, changing seasonally.
- more variation during summer than winter reflecting differences in CO_2 input versus consumption

Building models

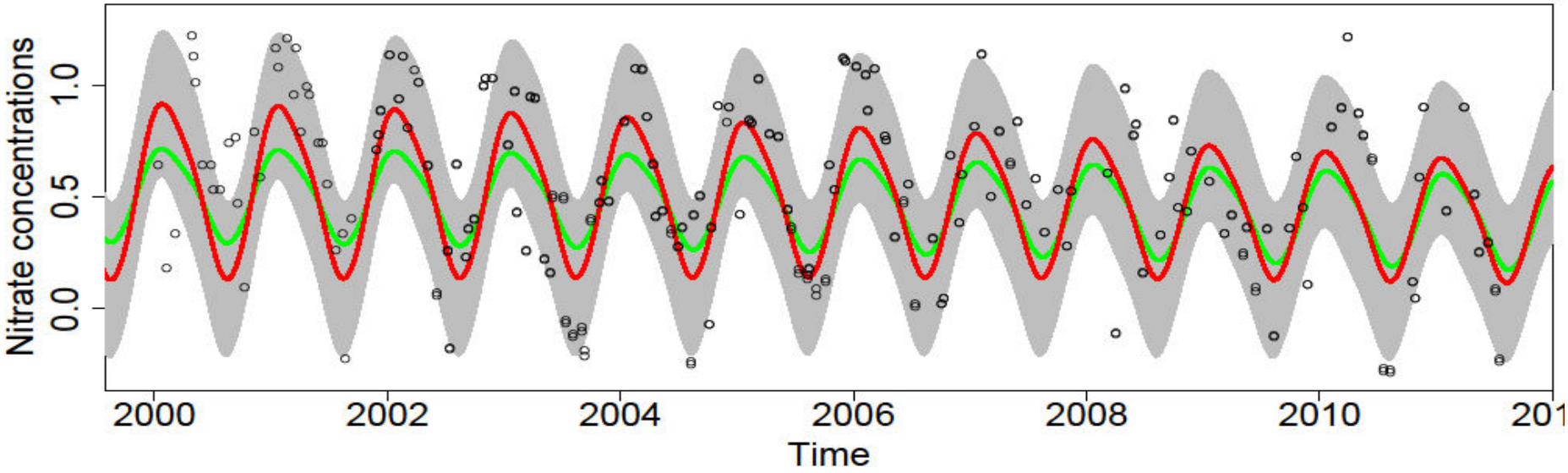
- How should we describe the trend? Linear or smooth? – how smooth?
- **are there seasonal components?** Constant in time? Changing phase or amplitude over time?



University of Glasgow | College of Science & Engineering

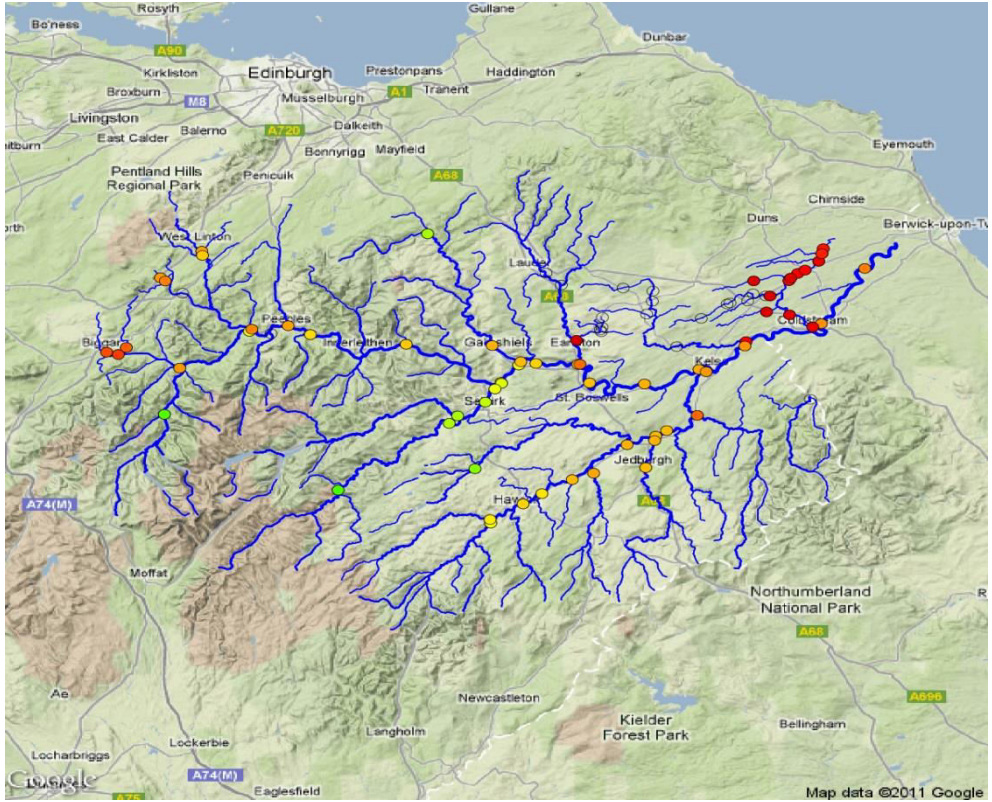
Building models

- How should we describe the trend and the seasonal components? Using smooth functions (based in splines) we can capture complex patterns, test different models and capture our uncertainty.

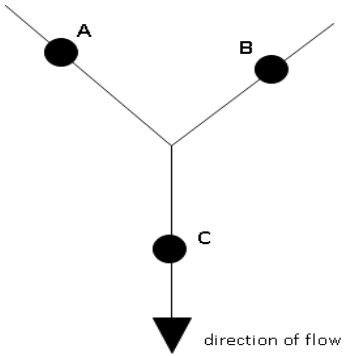


University of Glasgow | College of Science & Engineering

River networks



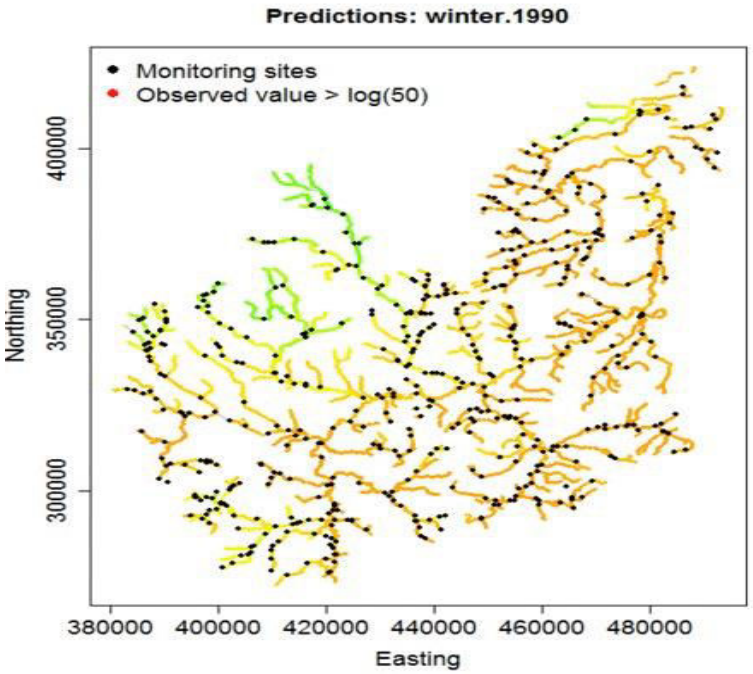
Spatial models for stream networks using stream distance rather than Euclidean distance .The user can specify if monitoring sites are ‘flow connected’ (A and C or B and C) or ‘flow unconnected’ (A and B).

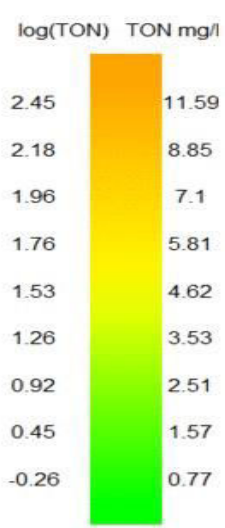


Flexible regression models over river networks, O’Donnell, Rushworth, Bowman, Scott and Hallard (2014)

University of Glasgow | College of Science & Engineering

Looking within one Large Hydrological Area





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

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

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

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

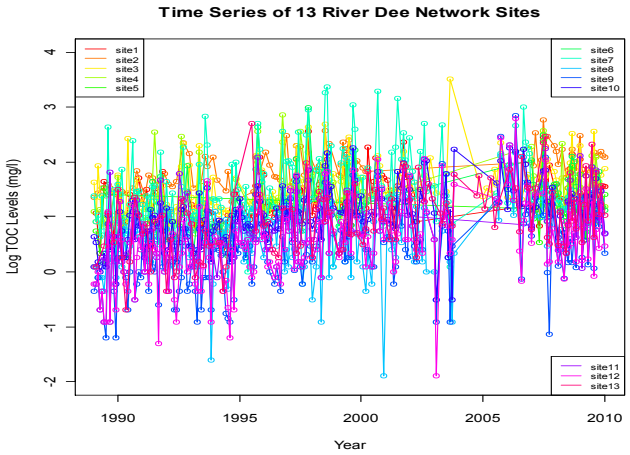
$$log(TON) = Easting + Northing + z_u + z_e + \varepsilon$$

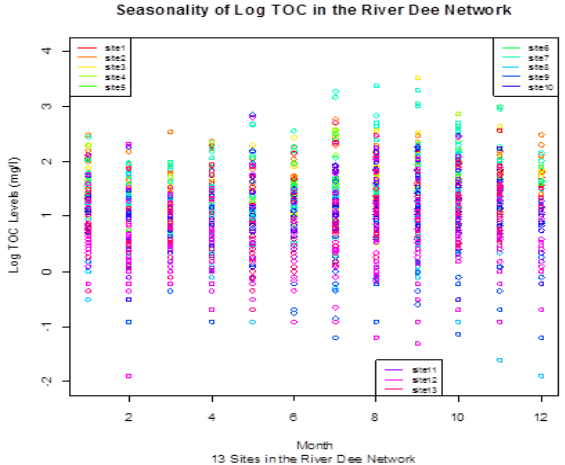
Kelly Gallacher, PhD student, joint with Claire Miller, Robert Willows (EA)

University of Glasgow | College of Science & Engineering

BIG data modelling

With a network of monitoring sites- *a data deluge*, the time series at each site can be regarded as a curve, the curve then becomes the "*data point*". The statistical model is based on the curves or functions which are assumed to be smooth. This is known as functional data analysis (FDA).

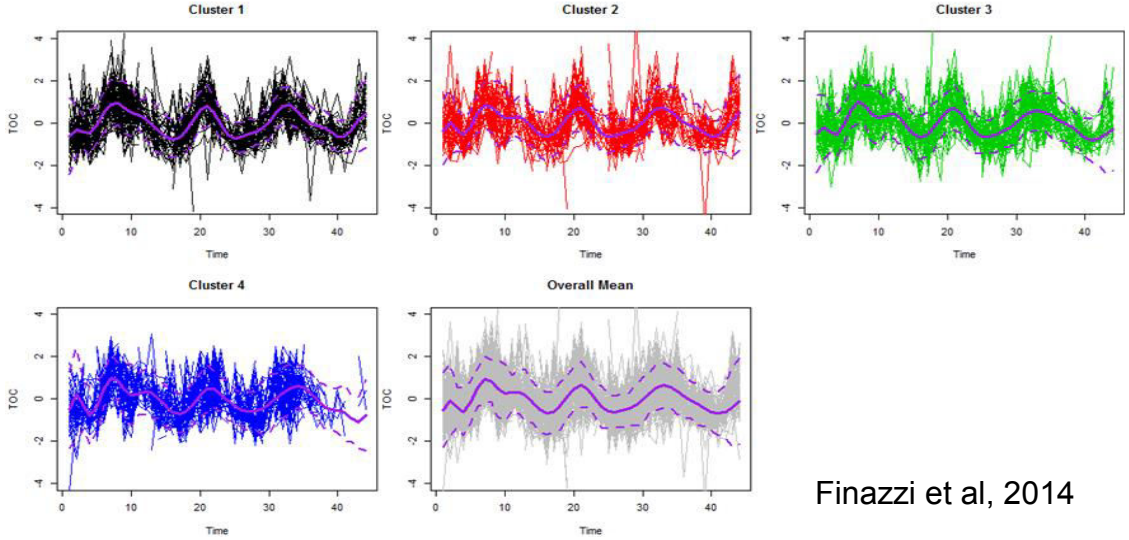


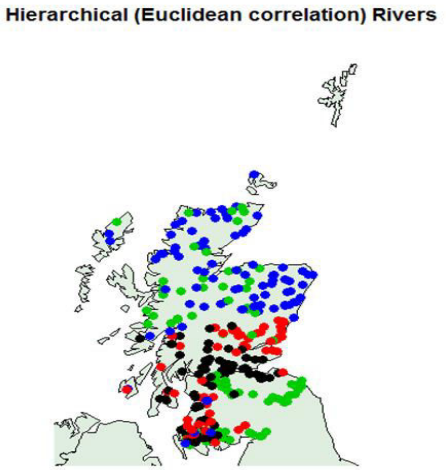


University of Glasgow | College of Science & Engineering

TOC in Scottish rivers

Functional clustering methodology has been applied to Total Organic Carbon (TOC) data from 333 monitoring locations across rivers in Scotland over 44 months, covering the period January 2007 - August 2010. (in collaboration with SEPA).



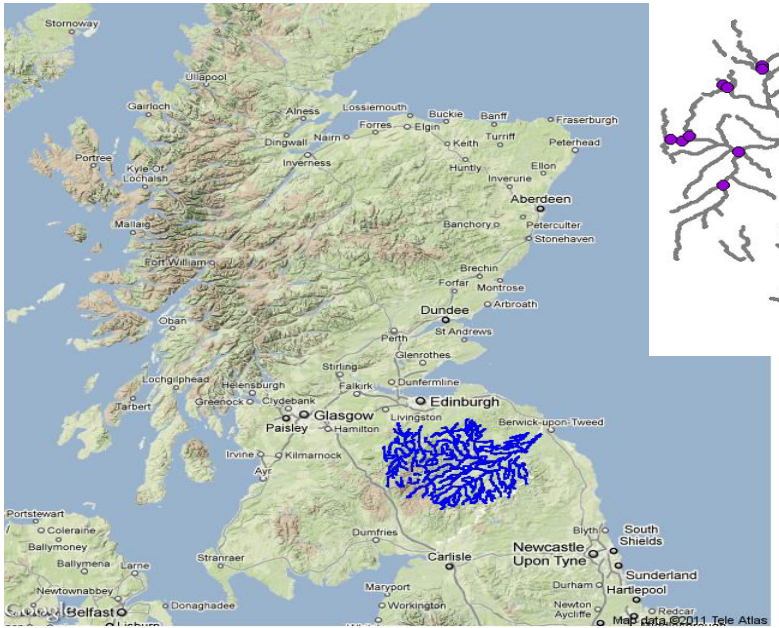


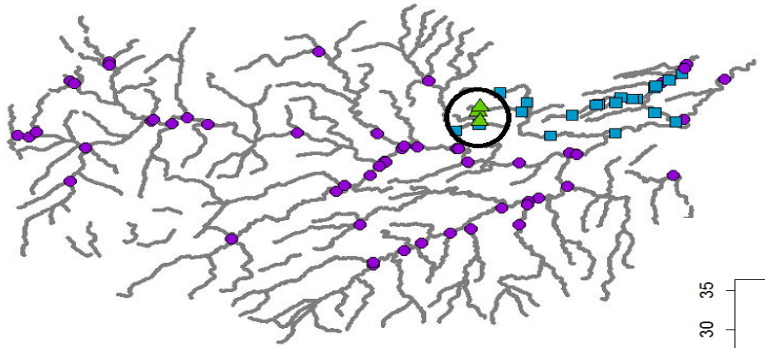
Finazzi et al, 2014

University of Glasgow | College of Science & Engineering

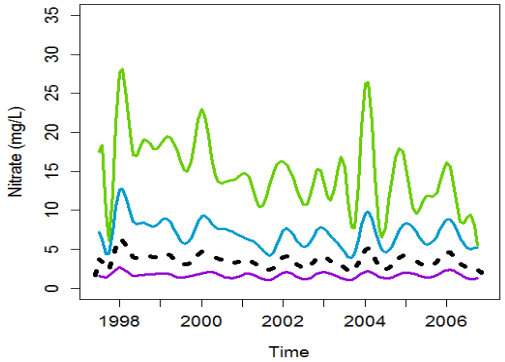
Nitrate in Tweed river basin

Functional clustering methodology has been applied to nitrate data in Tweed river basin- 77 monitoring sites, 10 years of data. Why? Can we simplify the network?





O'Donnell et al, 2014



challenges

- Data characteristics- quantity and quality, missingness, limits of detection.
- Non stationary, complex nature of the relationships
- For networks of sensors- building fast and efficient spatio-temporal models which scale, functional data analysis provides part of that solution
- uncertainty evaluation and visualisation

University of Glasgow | College of Science & Engineering

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* (2013) .
- 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.*
- El-Ayouty A, Scott E M, Miller C, Waldron S, Franco-Villoria M . Challenges in Modelling Detailed and Complex Environmental Data Sets: A Case Study Modelling the Excess Partial Pressure of Fluvial CO2, submitted J Ecological and Environmental Statistics