Detecting spatial and temporal patterns- coherence in lake water quality

UK-India water centre; Stirling, June 2017

Marian Scott, Claire Miller and Ruth O’Donnell

Marian.scott@glasgow.ac.uk

GloboLakes

Global Observatory of Lake Responses to Environmental Change

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

EO has the potential to change how we monitor and manage our natural resources.

Using new data streams, we can assess how lakes in all parts of the world are responding to environmental and climate change

Natural Objectives

In GloboLakes, we want to

Identify patterns of temporal coherence for individual remotely sensed lake characteristics & the spatial extent of coherence.

Identify phenological patterns of change in remotely sensed lake characteristics, evaluate evidence about state and any long term trends

Big data, big challenges

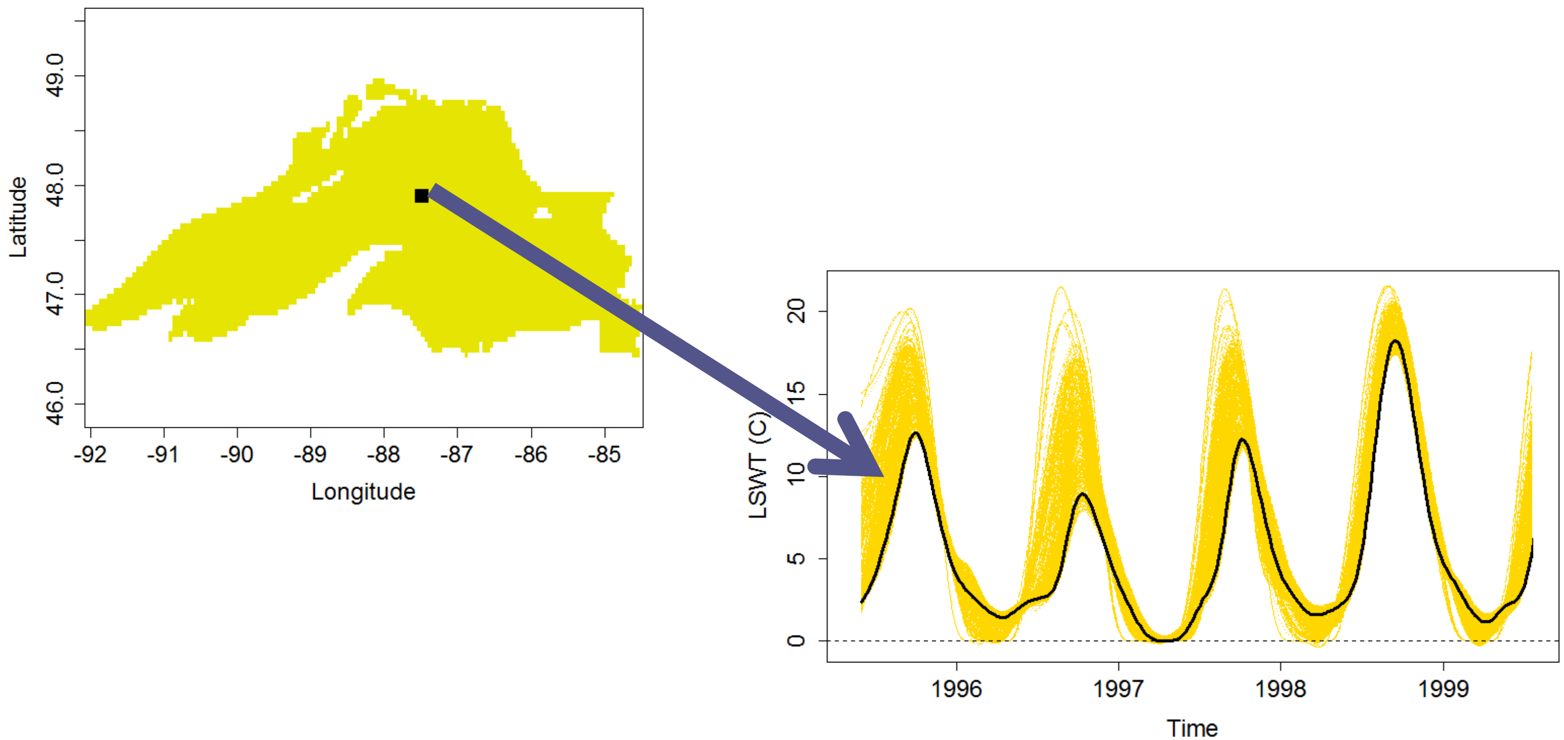
Huge volumes of data are being generated every day, we need to have statistical tools to allow us to explore, model and visualise the results

Our Glonolakes Data

For 1000 lakes, we will have temperature, chlorophyll, TSM, CDOM....

The data are available at pixel level, ranging from several thousands to millions, depending on the size of the lake and the satellite resolution
For each lake pixel we have a **time series** of measurements.

Big data, big challenges



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

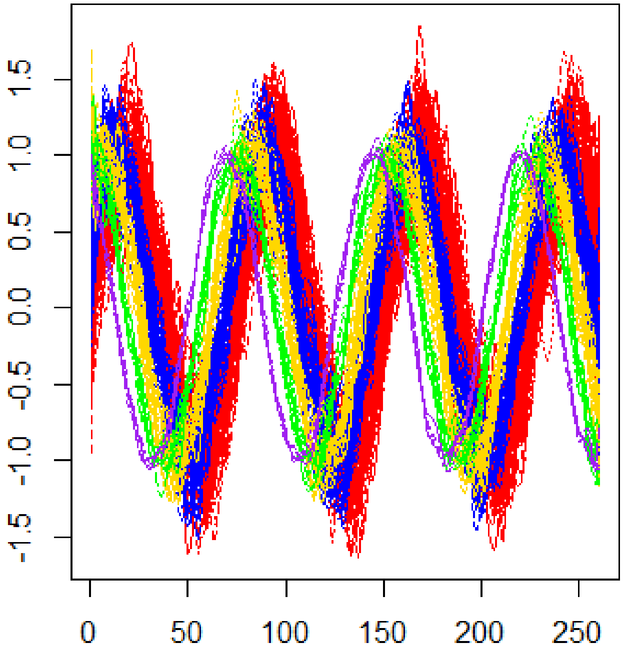
Why is temporal coherence of interest?

“Limnologists have regarded temporal coherence (synchrony) as a powerful tool for identifying the relative importance of local-scale regulators and regional climatic drivers on lake ecosystems” Yu et al, 2011

“It is well known that lakes can show a regionally coherent response to large-scale weather patterns. Lakes may also be synchronously affected by the effects of atmospheric deposition, meteorological effects on the catchment and hydrological connectivity. The degree of coherence is governed by the strength of local, lake-specific factors compared to the strength of regional drivers.”
UKLeon project

What is temporal Coherence?

- The synchrony between major fluctuations in a set of time series is often described as **temporal coherence. So we are trying to identify common patterns.**



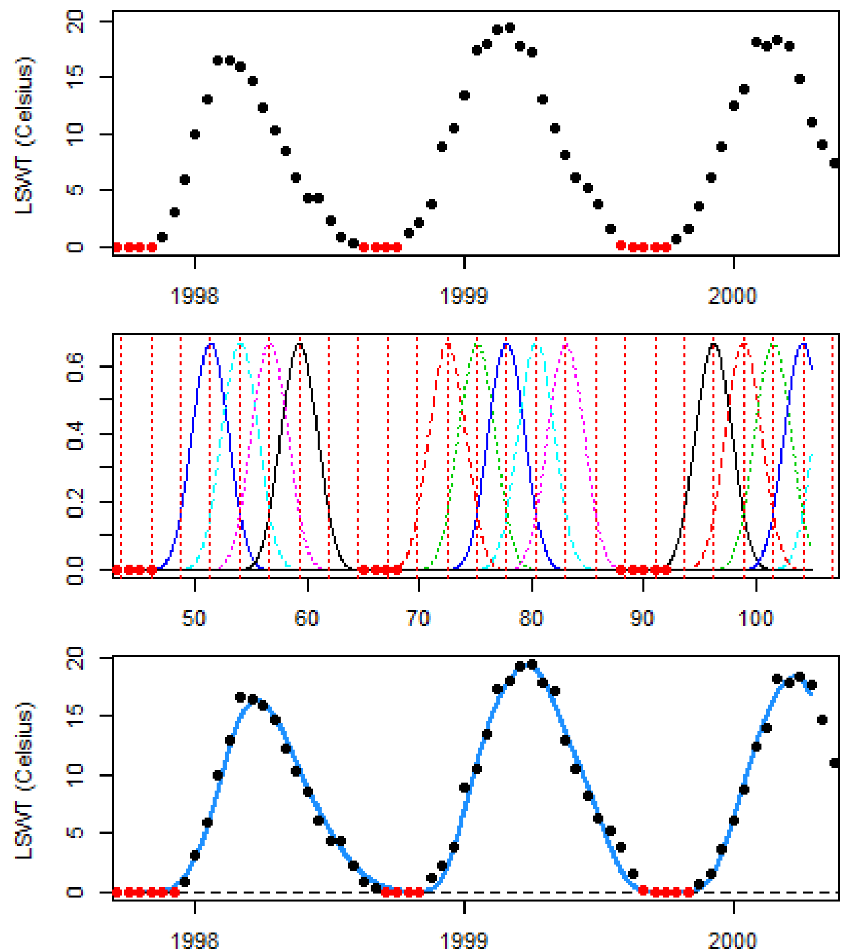
Centre for Ecology & Hydrology NATURAL ENVIRONMENT RESEARCH COUNCIL | University of Dundee | University of Reading | University of Glasgow | PML Plymouth Marine Laboratory | University of Stirling

Between and within Lake Coherence

- As well as looking at temporal coherence across lakes, it may also be of interest to investigate the similarities of different areas within lakes
- Large lakes may have several basins/areas within them that have different characteristics in terms of the trends, seasonal patterns and levels of determinands present
- How do we identify which lakes or which areas are coherent?
- Several statistical approaches are available;
  - ~~Correlation analysis~~- not recommended
  - Functional data analysis
  - Cluster Analysis
  - Principal Component Analysis

Between and within Lake Coherence

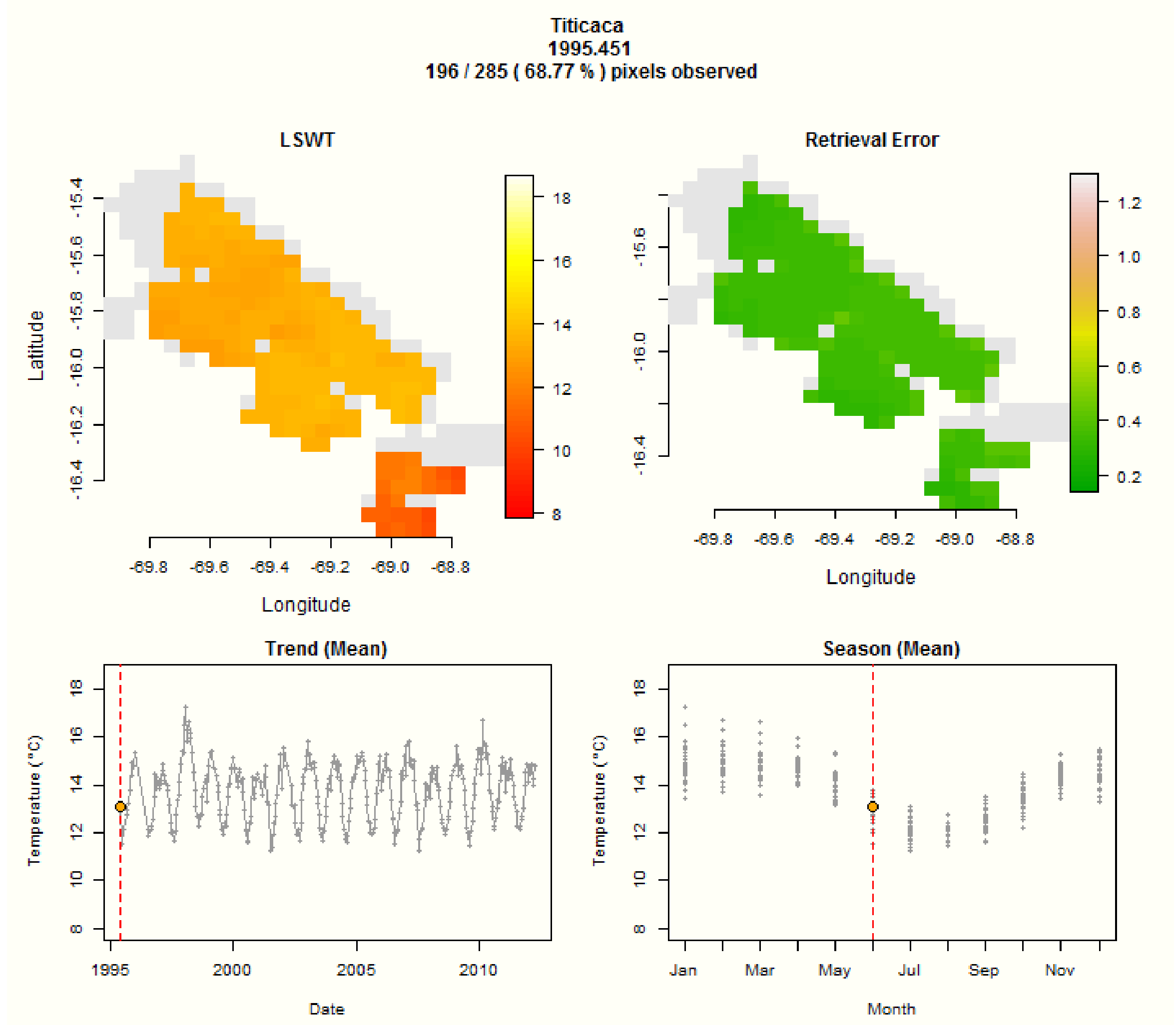
- We have developed some new statistical tools using functional data analysis
- For each lake pixel we have a **time series** of measurements. We view the time series as realisations of an underlying, **unknown smooth function**
- We estimate the smooth functions using a technique called **B-splines** and carry out statistical analysis on the fitted functions rather than a single summary value; known as Functional Data Analysis (FDA)



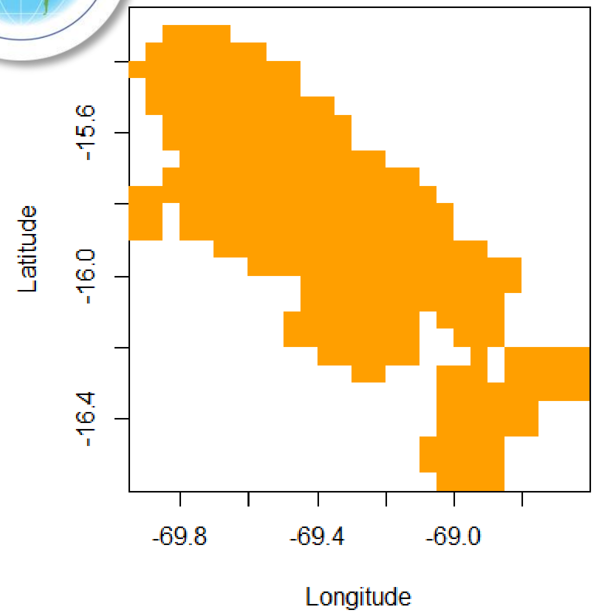
Centre for Ecology & Hydrology NATURAL ENVIRONMENT RESEARCH COUNCIL
UNIVERSITY OF DUNDEE
University of Reading
University of Glasgow
PML Plymouth Marine Laboratory
UNIVERSITY OF STIRLING

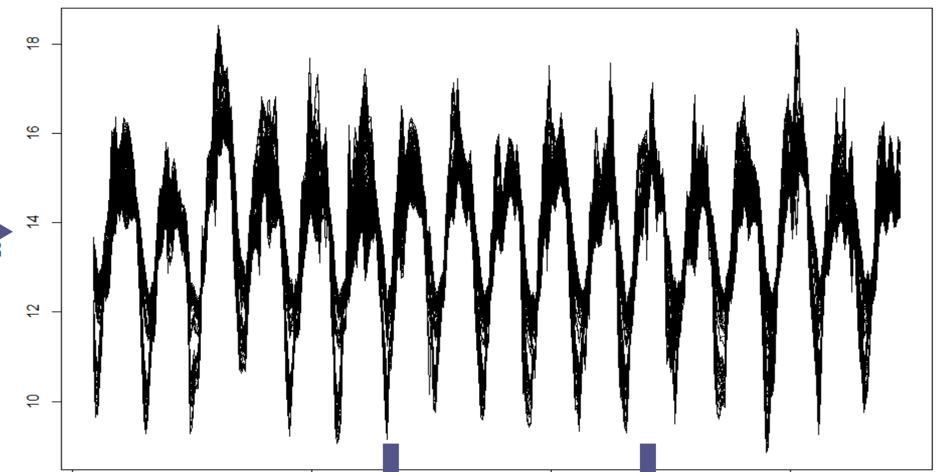
ARClake Data Example

- Lake Titicaca
- Twice monthly data
- 281 pixels (max)
- Lower panels show average pattern over time and seasonal pattern

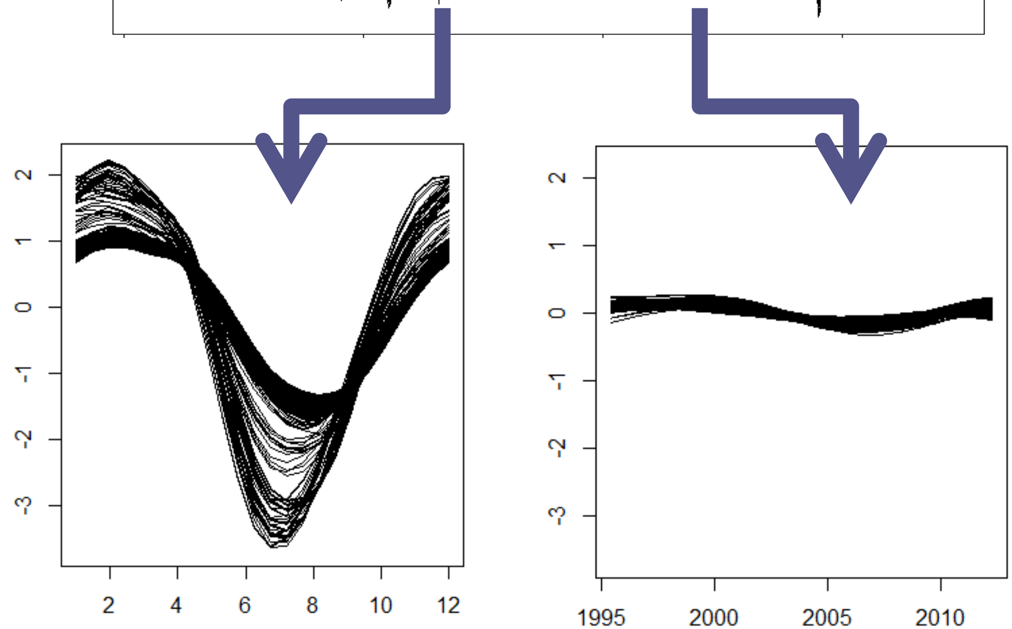


Pixels to Time Series





- 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

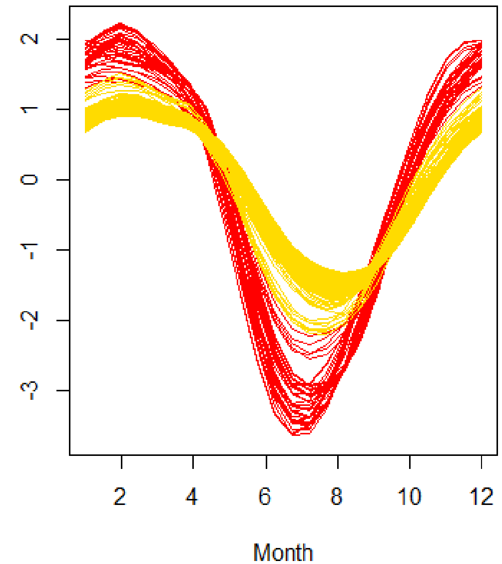


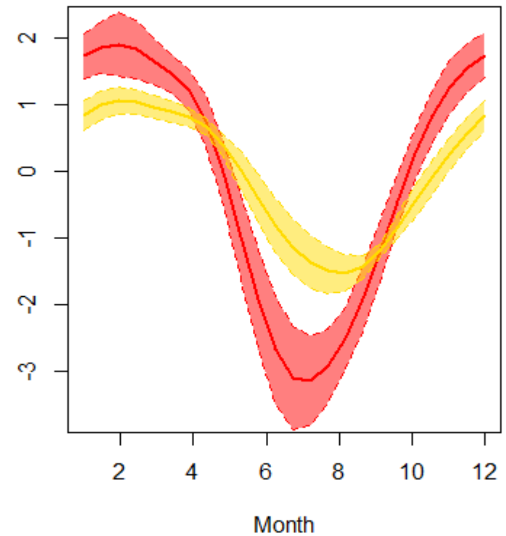
Functional Clustering

- We identify coherent (ie similar patterns) groups of lakes or clusters
- Identifying common patterns and changes is ecologically important, subsequently we can assess what is driving these changes
- Functional clustering means we obtain clusters of curves identified as coherent in terms of temporal dynamics e.g. seasonal patterns and long term trends.
- We have developed adaptive curve fitting to take into account unique features of the data such as discontinuities due to ice cover

Lake Titicaca:

- We have used reconstructed data so we have complete time series over the time period
- An automatic (data driven) procedure was used to determine the statistically optimal number of groups
- 2 groups of pixels were identified
- Colouring the seasonal curves by the cluster memberships highlights the differences between the two groups
- The bottom plot shows the cluster mean seasonal pattern for each group (solid line) with the shaded area representing the variability of the curves within each group



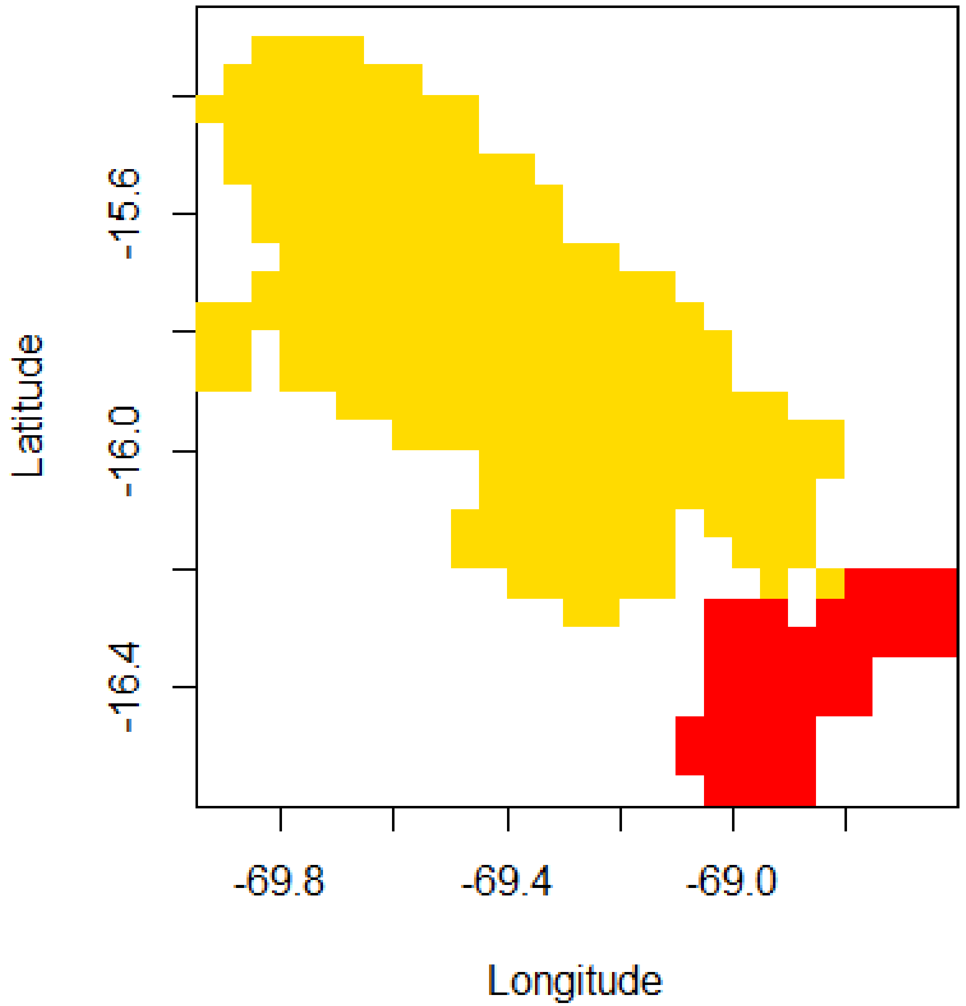


Lake Titicaca:

Map shows spatial distribution of the clusters

No spatial information was included in the clustering (but spatial correlation can be incorporated)

Clear separation of the lake into 2 distinct regions based on seasonal patterns

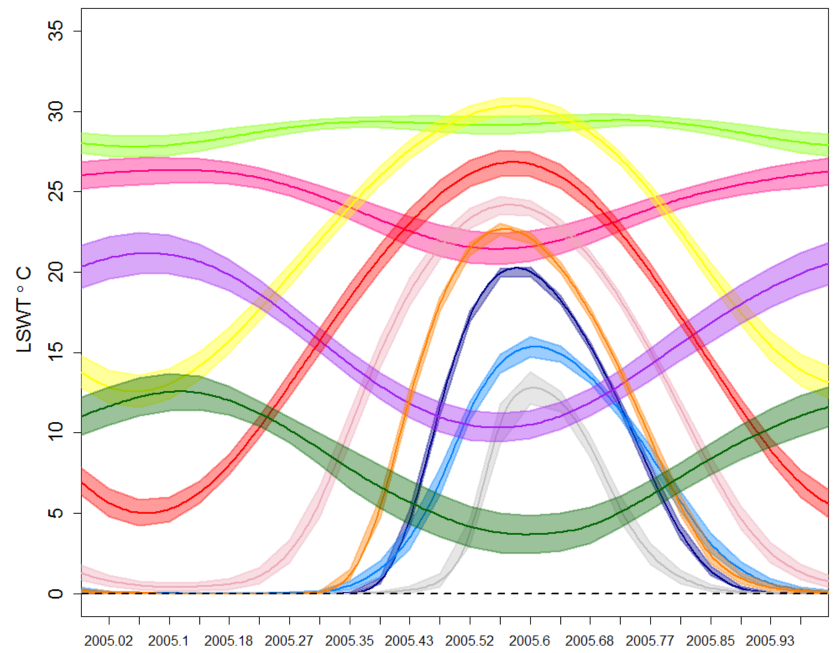


Functional Clustering on a global scale

- We show a simple illustration based on surface water temperature
- We use the smooth lake mean temperature curve as the data point
- We take account of ice, missing data and retrieval uncertainty
- We identify coherent groups of lakes or clusters
- We generate a map
- We can then think about what might be the global and regional effects driving the patterns- attribution

Arclake- Global temperature coherence example

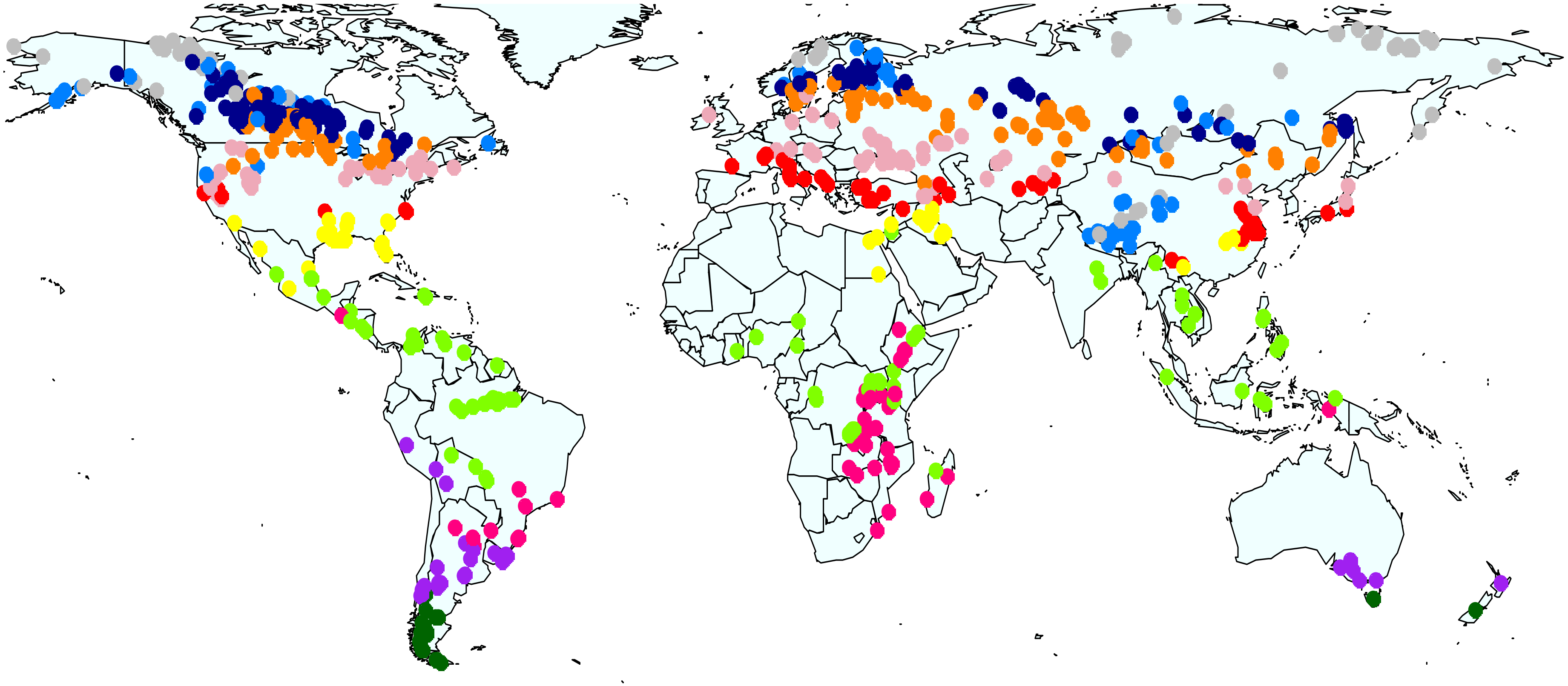
- Using a reduced set of 732 lakes and 17 years of bi-monthly lake surface water temperature.
- We identified 11 distinct groups.
- The figure on the right shows the temporal patterns in each of the lake clusters



- The map below shows functional clustering results

Centre for Ecology & Hydrology
NATURAL ENVIRONMENT RESEARCH COUNCIL
UNIVERSITY OF DUNDEE
University of Reading
University of Glasgow
PML Plymouth Marine Laboratory
UNIVERSITY OF STIRLING

Arclake- example of global lake temperature 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

Summary

- Lake studies are usually based on small numbers of lakes. EO lets us look at things on a global scale.
- We need appropriate statistical models and methods to take full advantage of the new data streams.
- The data streams present interesting statistical challenges including dealing with missing data, ice cover and the retrieval uncertainties. We have been developing new statistical tools based on functional data analysis to address these new data streams.

acknowledgements

- O'Donnell, Scott and Miller were funded for this work through the NERC GloboLakes project (NE/J022810/1).
- The authors gratefully acknowledge the ARC Lake project for access to the LSWT data (www.laketemp.net) and the rest of the GloboLakes team.

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