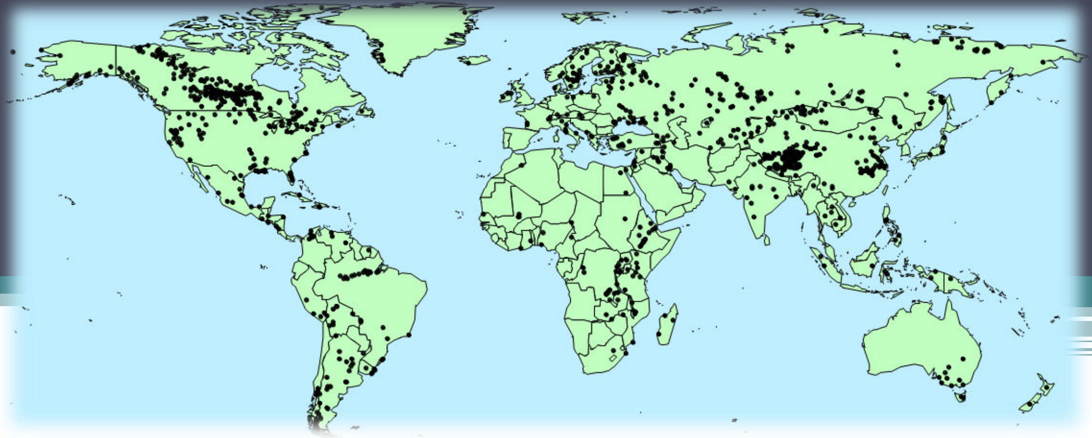
Globolakes : Global coherence of lake water quality

Ruth O'Donnell[1], Claire Miller[1], Marian Scott[1], Chris Merchant[2]
[1]School of Mathematics and Statistics, University of Glasgow
[2]Department of Meteorology, University of Reading



GloboLakes
Global Observatory of Lake Responses to Environmental Change

NERC SCIENCE OF THE ENVIRONMENT

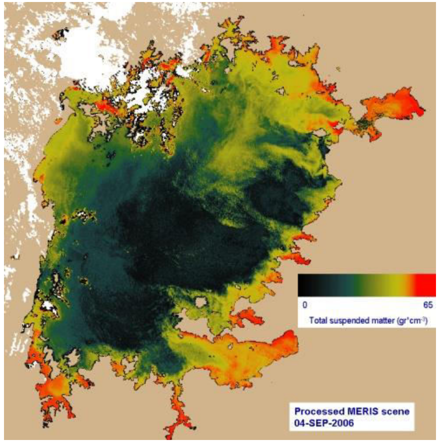
Centre for Ecology & Hydrology NATURAL ENVIRONMENT RESEARCH COUNCIL
UNIVERSITY OF DUNDEE
University of Reading
University of Glasgow
PML Plymouth Marine Laboratory
UNIVERSITY OF STIRLING

Globolakes (www.globolakes.ac.uk)

- **Globolakes** is a 5 year consortium project involving 6 UK research groups

- Our goal is to investigate the state of 1000 lakes and their response to climatic and other environmental drivers of change at a global scale using a 20 year archive of satellite based observations.

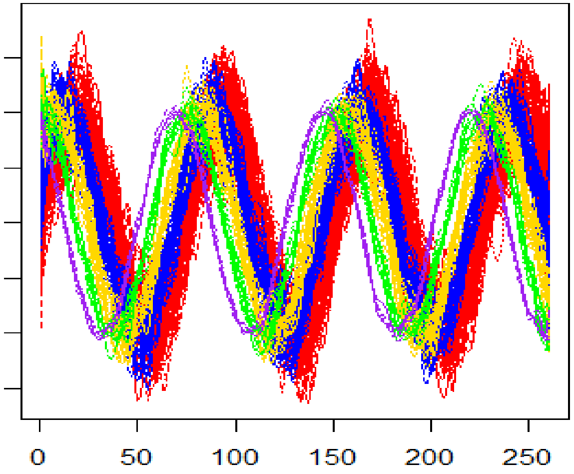
- One of our key aims is to identify patterns of temporal coherence for individual remotely sensed lake characteristics and the spatial extent of coherence for 1000 lakes.



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

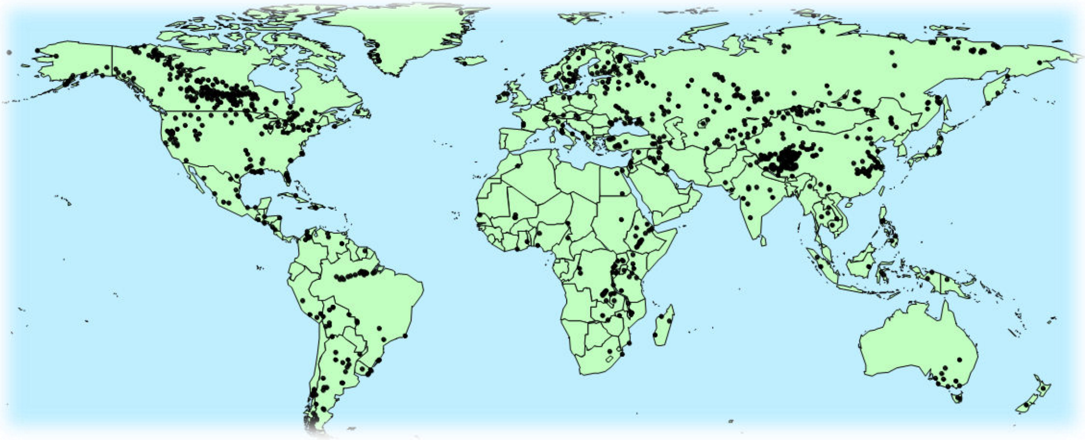
Coherence

- Identifying both long-term change and phenological/seasonal changes is of great ecological importance.
- The synchrony between major fluctuations in a set of time series is often described as **temporal coherence.**
- Aim: group time series into a suitable number of clusters where two time series belong to the same cluster if they are coherent with each other.
- Focus on comparing a large number of time series and obtain groupings based on common features across time.
- A **functional data analysis** approach has been taken



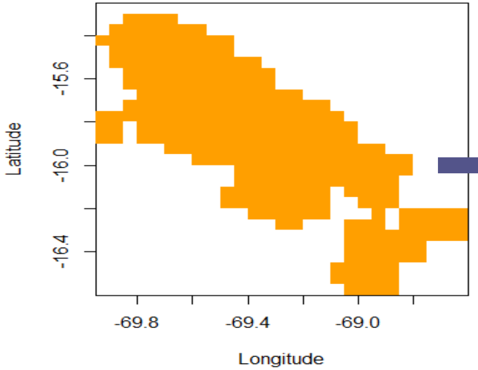
ARCLake data

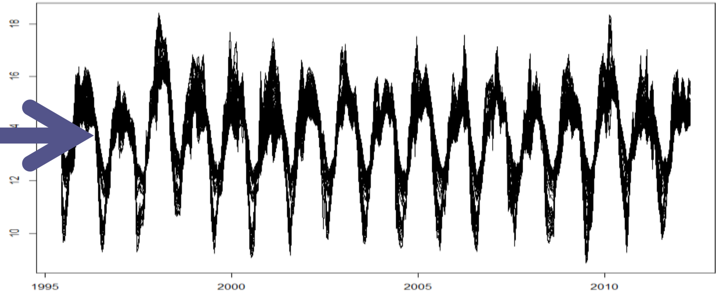
- Average lake surface water temperature data (processed using images from AATSR satellite)
- For global coherence, lake mean time series are considered
- Twice monthly observations covering an 18 year time period between 1995 and 2012
- Data on 700 lakes across entire time period
- Investigate clusters based on trend and seasonality together



ARClakes/Globolakes Lakes Locations

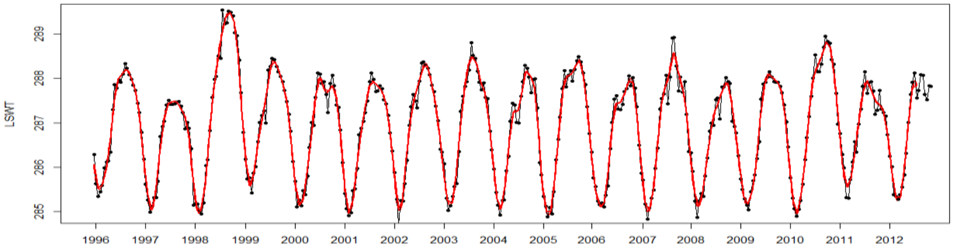
Functional Data Analysis





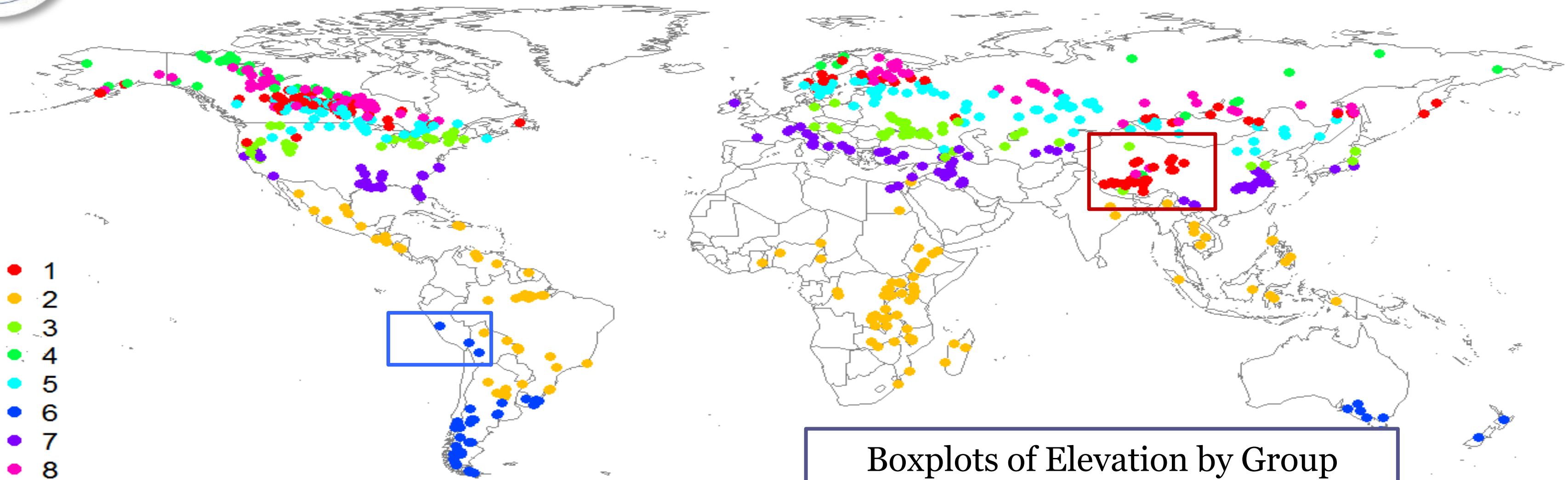
- Each time series corresponds to a pixel
- Compute mean curves
- Each time series can be smoothed – keep key features but remove local variability

- Smooth curves fitted to each lake means using a b-spline basis
- Model based clustering applied to the b-spline coefficients which define the smooth curves
- Results are data driven

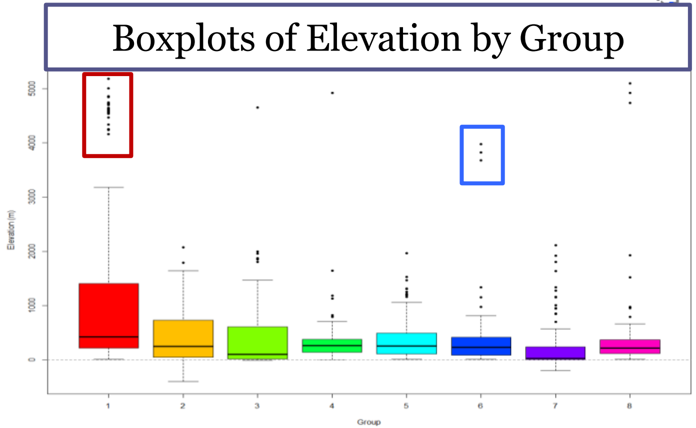


Centre for Ecology & Hydrology
NATURAL ENVIRONMENT RESEARCH COUNCIL
UNIVERSITY OF DUNDEE
University of Reading
University of Glasgow
PML Plymouth Marine Laboratory
UNIVERSITY OF STIRLING

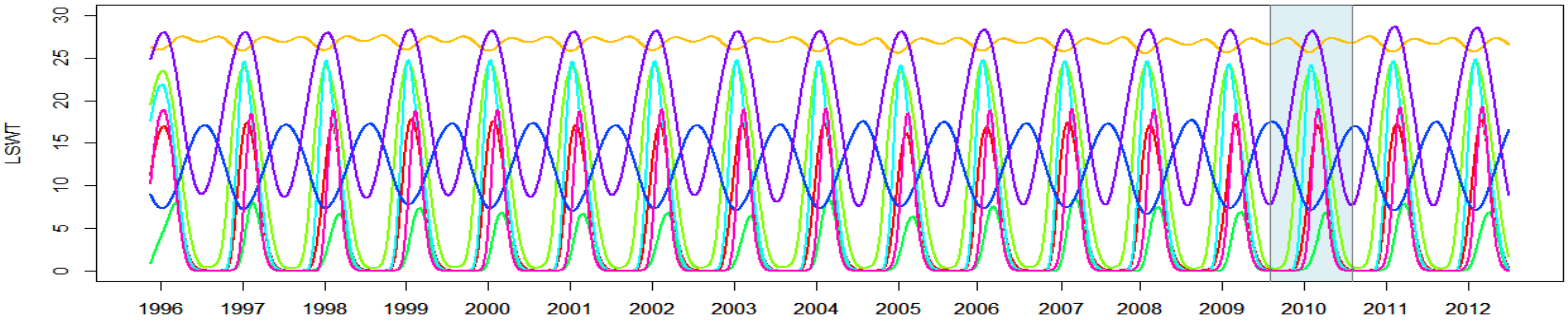
Results: Global Coherence

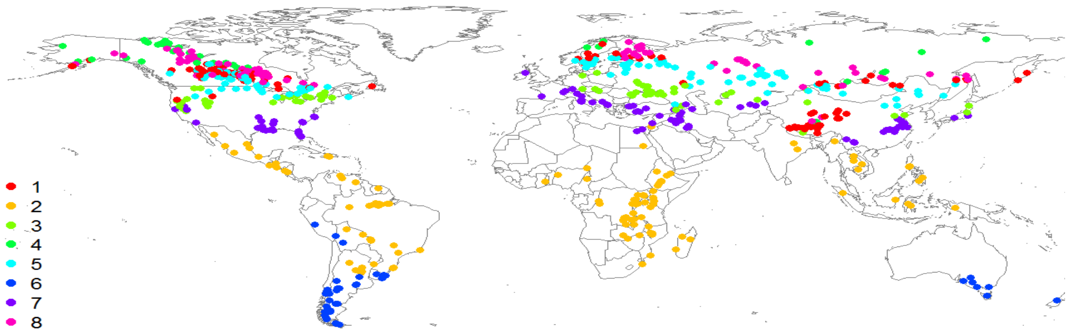


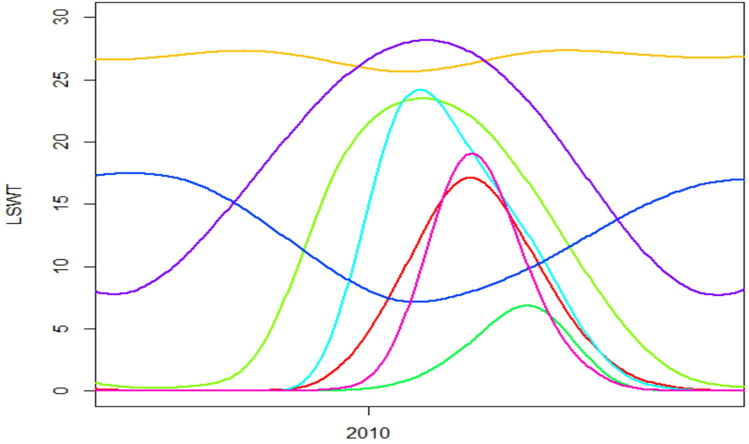
- 8 clusters identified as statistically optimal
- Map shows distribution of clusters
- Only temperature data underpins the formation of the clusters



Results: Global Coherence

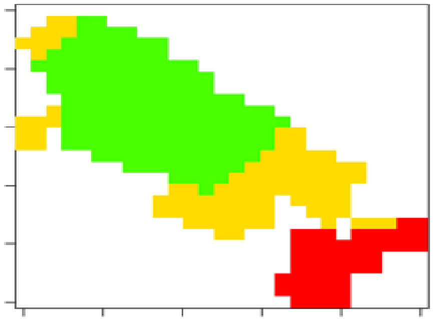


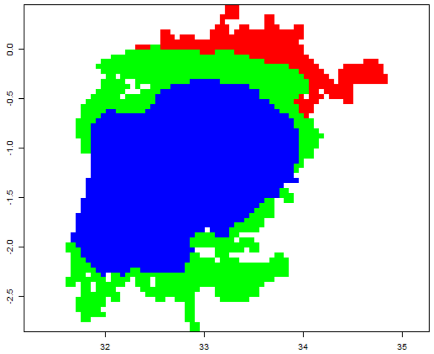


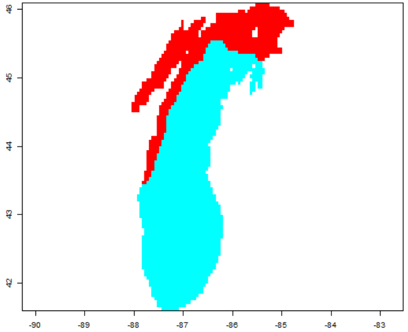


Within Lake Coherence

- 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 these different areas?
- We have taken a two stage approach;
  - Functional Principal Component Analysis (FPCA)
- Use FPCA to find dominant modes of variation in the data (known as principal components or PCs)
- Keep a small number of these PCs in order to approximate the original data



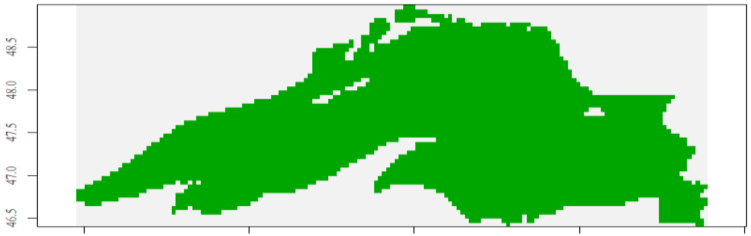


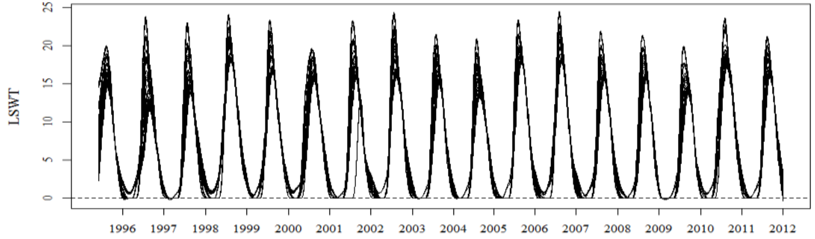


Centre for Ecology & Hydrology
NATURAL ENVIRONMENT RESEARCH COUNCIL
UNIVERSITY OF DUNDEE
University of Reading
University of Glasgow
PML Plymouth Marine Laboratory
UNIVERSITY OF STIRLING

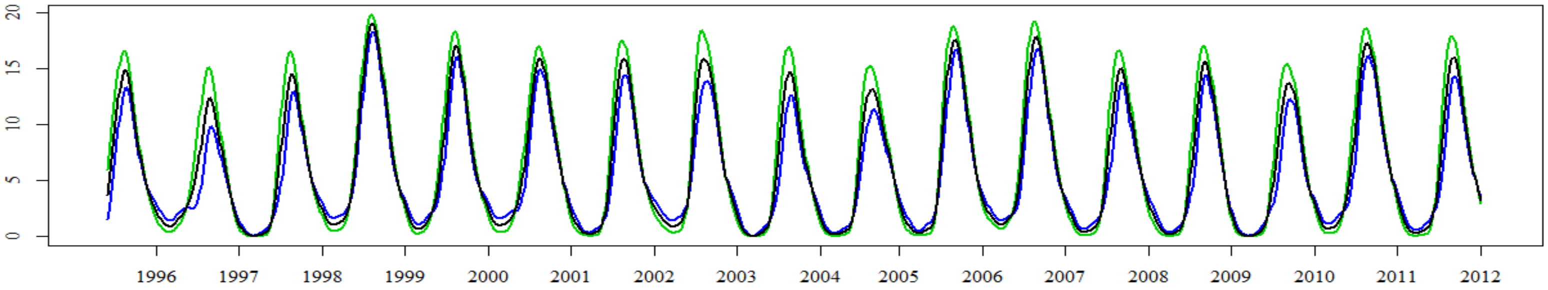
Within Lake Coherence: Example

- Example: Lake Superior - 4094 pixels
- Smooth the time series for each pixel – 4094 smooth curves
- First two FPC scores explain 86% of the variability in the curves



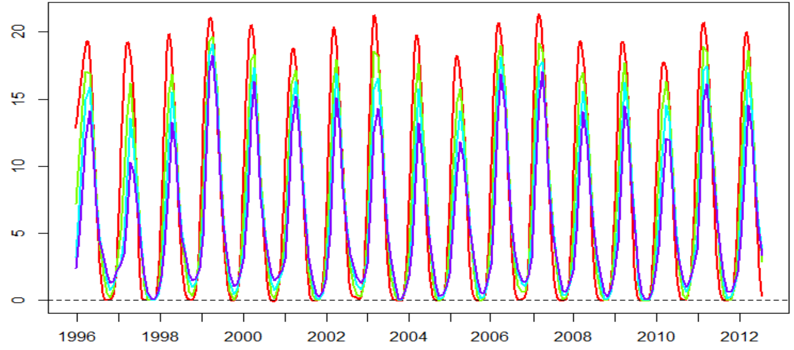


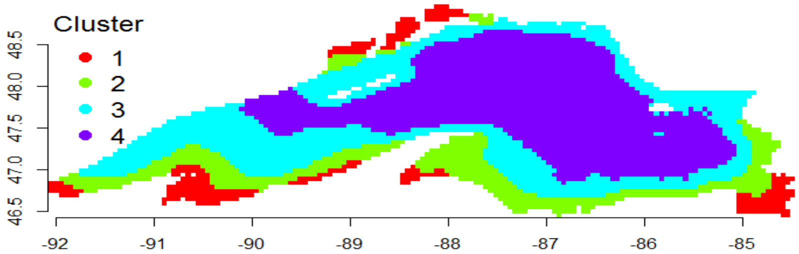
PC function
(Percentage of variability 77)

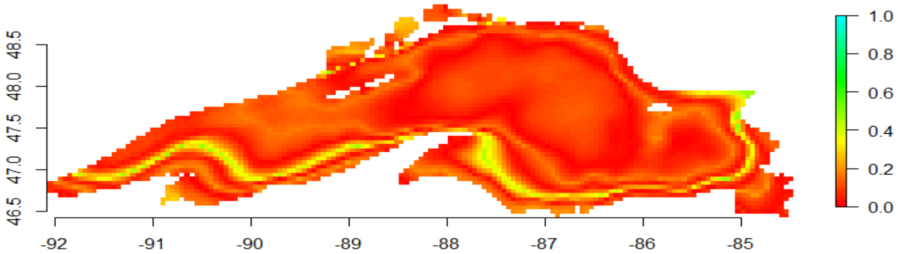


Within Lake Coherence: Example

- 4 clusters identified as statistically optimal
- Clear distinctions between the mean functions corresponding to these four groups
- Key discrepancy is amplitude of the seasonal patterns each year.





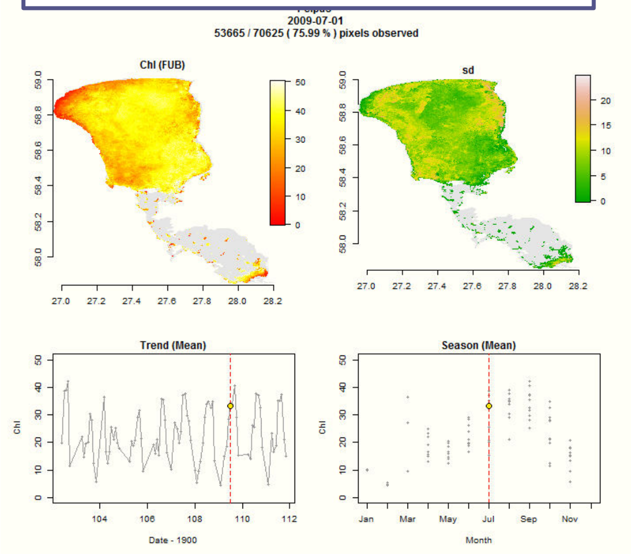


Summary

- Our approach enables clusters of curves to be identified which are coherent in terms of temporal dynamics.
- Functional data analysis reduces dimensionality while retaining important information
- The methods are very computationally efficient (run on standard desktop)
- Model based clustering can be used to identify statistically optimal (data driven) number of clusters
- Uncertainty in cluster classification is also obtained

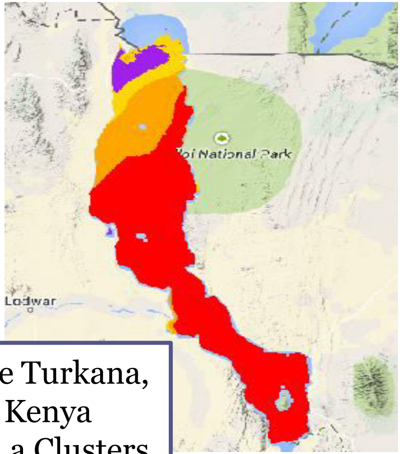
Next Steps

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



Analysis of lake mean curves gives us an overview of global coherence

Within lake clustering captures smaller scale detail - our ultimate goal is to incorporate within lake clustering results into global coherence



Lake Turkana, Kenya
Chl a Clusters

We will also explore more variables such as chlorophyll and coloured dissolved organic matter at a much finer resolution.

Missing data through time and in space present new statistical challenges

References

- Abraham, C., Cornillon, P.A., Matzner-Lber, E., Molinari, N. Unsupervised curve clustering using b-splines. *Scandinavian Journal of Statistics* 30(3), 581–595 (2003).
- 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,* 2014 .
- Henderson, B. Exploring between site differences in water quality trends: a functional data analysis approach. *Environmetrics* 17(1), 65–80 (2006).
- MacCallum, S., Merchant, C.: Arc-lake v2.0, 1995-2011 [alidxxxx plrec9d ts366lm]. University of Edinburgh, School of GeoSciences / European

- **Acknowledgements:** Natural Environment Research Council (NE/J022810/1)
  ARCLake Project
  Globolakes consortium
  ESA DUE DIVERSITY II project for providing ENVISAT data and derived indicator products.