Clustering Lake Responses to Environmental Change

Mengyi Gong, Ruth O'Donnell, Claire Miller & Marian Scott | University of Glasgow

Stephen Maberley, Laurence Carvalho, Steve Thackerey & Ian Jones | Centre for Ecology & Hydrology

Stefan Simis | Plymouth Marine Laboratory

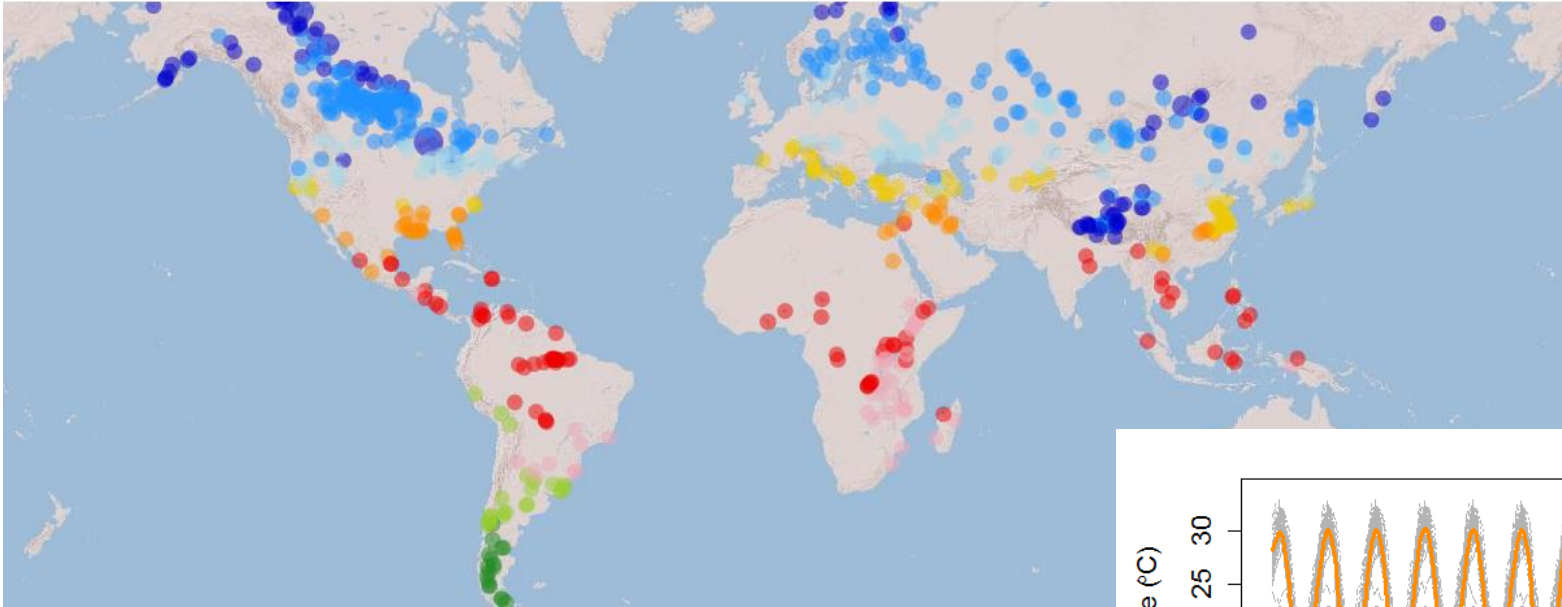Laura Carrera & Chris Merchant | University of Reading

Peter Hunter, Vagelis Spyrakos & Andrew Taylor | University of Stirling

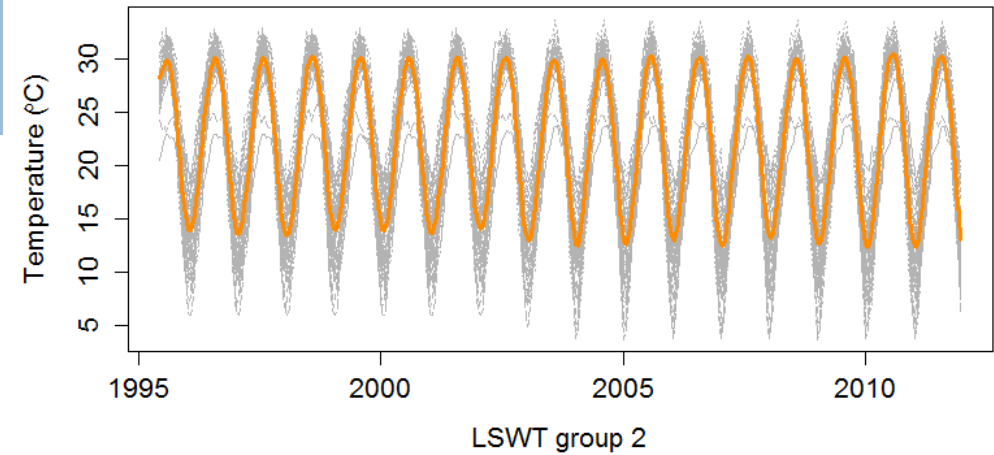Eirini Politi, Mark Cutler & John Rowan | University of Dundee

Coherence – Lake surface water temperature

GloboLakes ARCLAKE Time Series

This map is centered at (0, 0).

LSWT group 2

The synchrony between major fluctuations in a set of time series is often described as **temporal coherence.**

Coherence

- Identifying both long-term change and phenological/seasonal changes is of great ecological importance.

- We aim to group lake time series for lake surface water temperature and chlorophyll into a suitable number of clusters.

- Two time series belong to the same cluster if they are coherent with each other.

- A **functional data analysis** approach has been taken here to analyse the data.
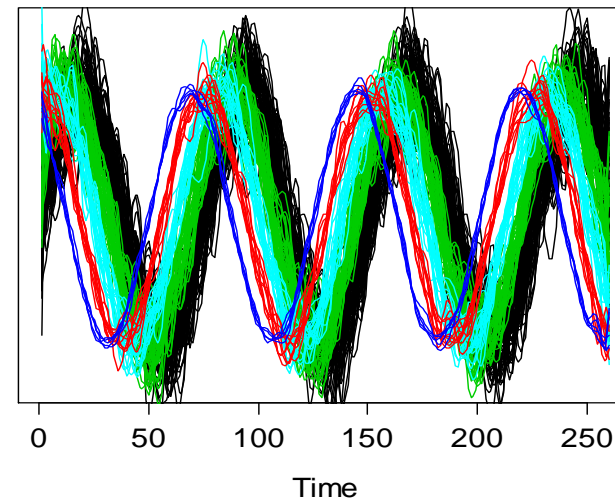
# Data – locations of 1000 lakes
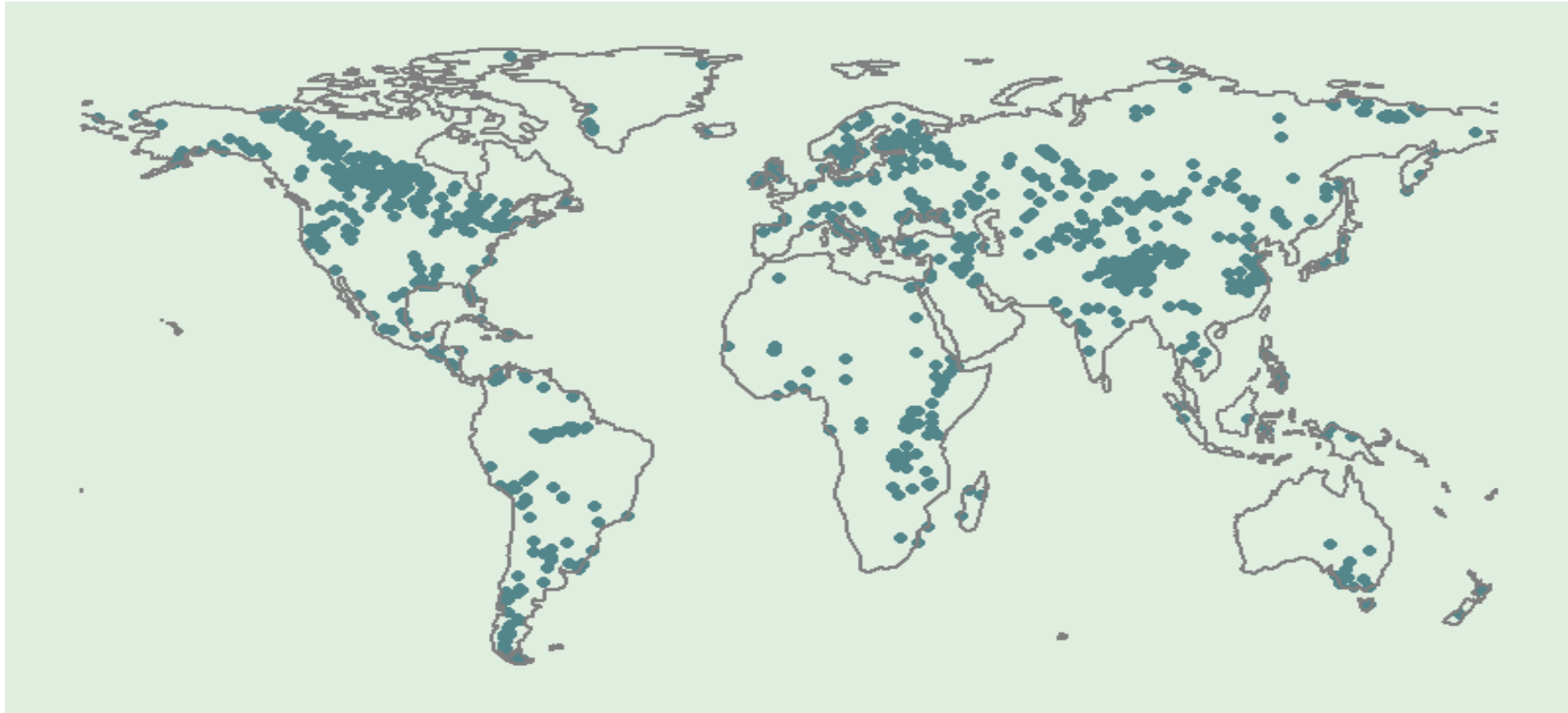
ArcLake v3.0 data: lake surface water temperature (LSWT)

- 0.05 degree resolution (approx 5km sq)
- 961 Globolakes matches, reduced to 732 based on data availability
- Bi-monthly reconstructed data, from June 1995 to Dec 2011 (398 time points)

Calimnos v1.04 data: chlorophyll-a (Chl-a)

- 0.0027 degree resolution (approx 300m sq)
- 1000 lakes in total, reduced to 535 lakes based on data availability
- Monthly data, from June 2003 to Oct 2011 (101 time points)

Data to smooth curves

Smoothing **reduces dimensionality and captures key underlying patterns** in the data - removing any local variability which may not be of interest
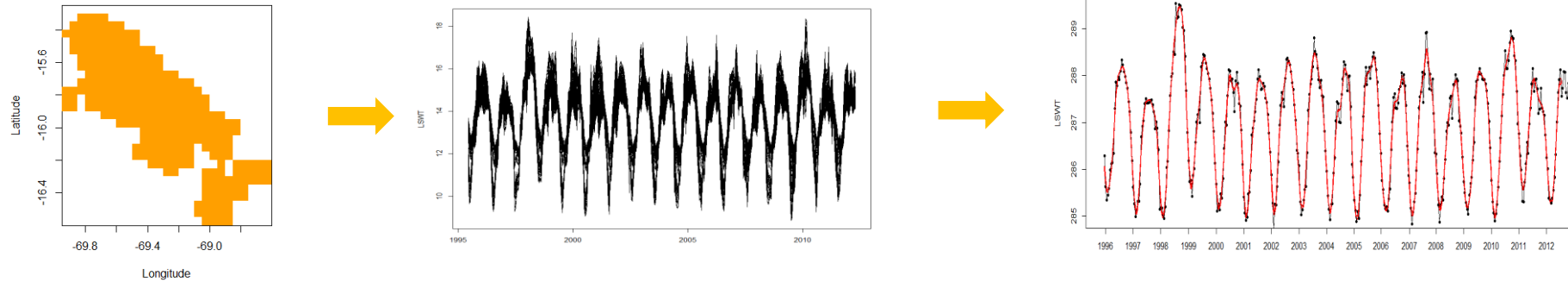
- There is a time series for each pixel
- We can average over all pixels to get the mean time series for the lake
- Each lake mean time series, or pixel time series, can be thought of as a noisy realisation of an unobservable function

Adaptive smoothing

- We can **adjust the smoothness** based on the **proportion of missing data** in each monthly image, to reflect the level of 'confidence' we have in the mean values, which sometimes comes from limited data

- A heavier penalty is applied to time points with less data, so the resulting curve would not track these points too closely.
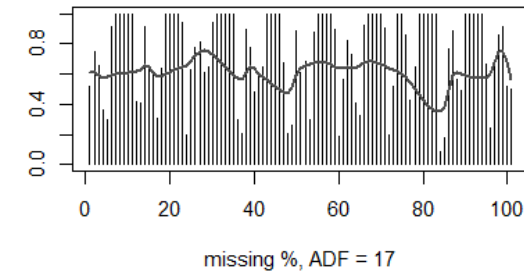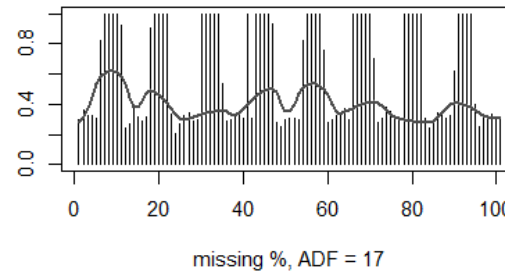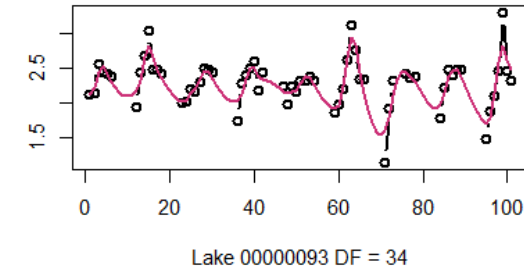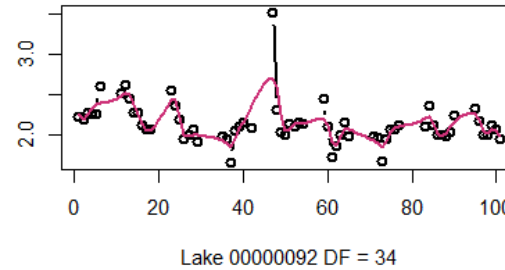
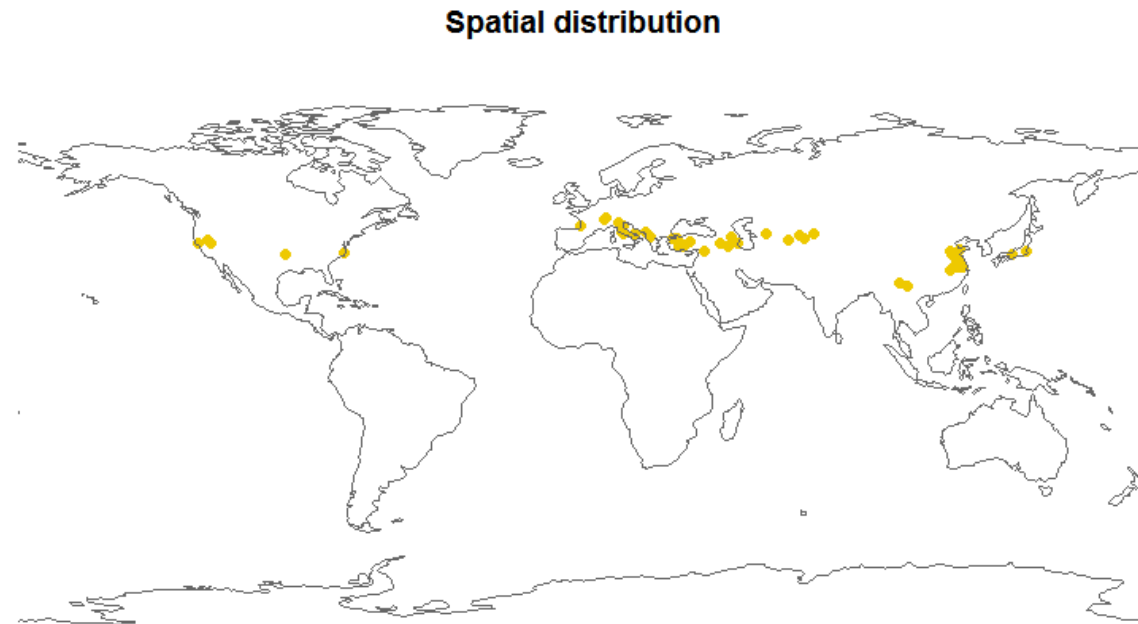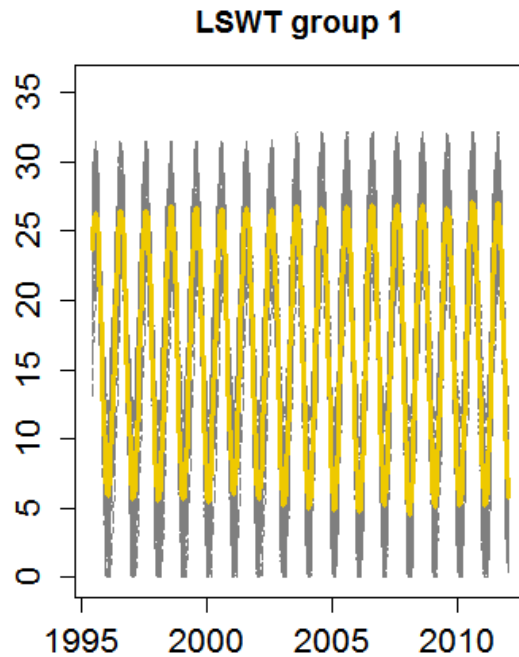Clustering can be applied to the attributes of the curves
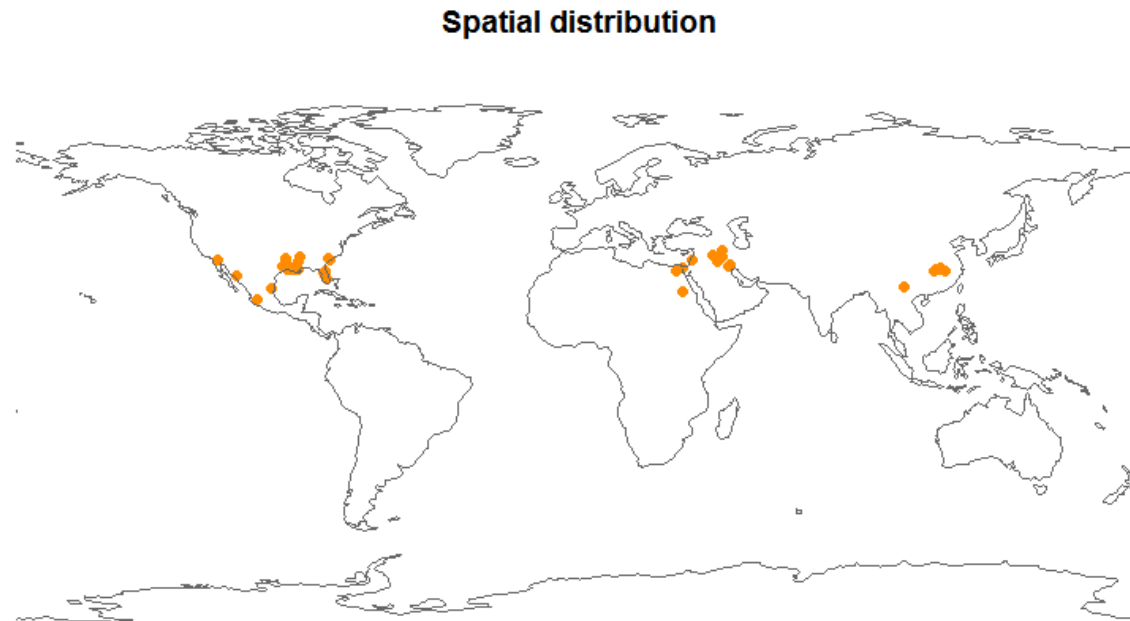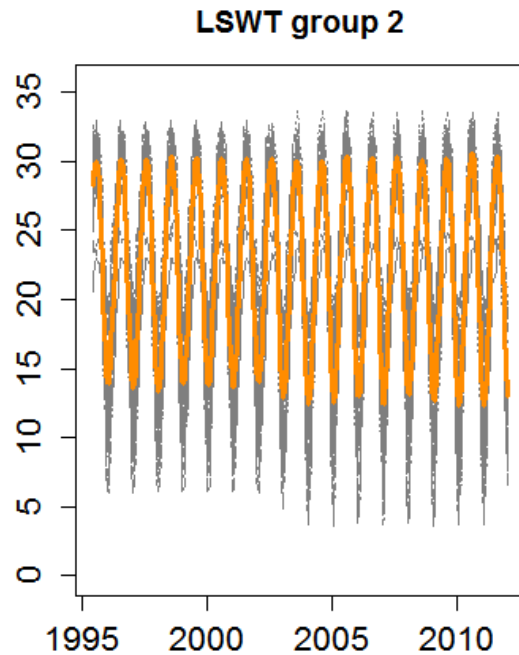
K-means Clustering
Other Clustering Methods
- Hierarchical clustering
- Model based clustering (e.g. Gaussian mixture model)

Data driven methods can be used to select the number of clusters
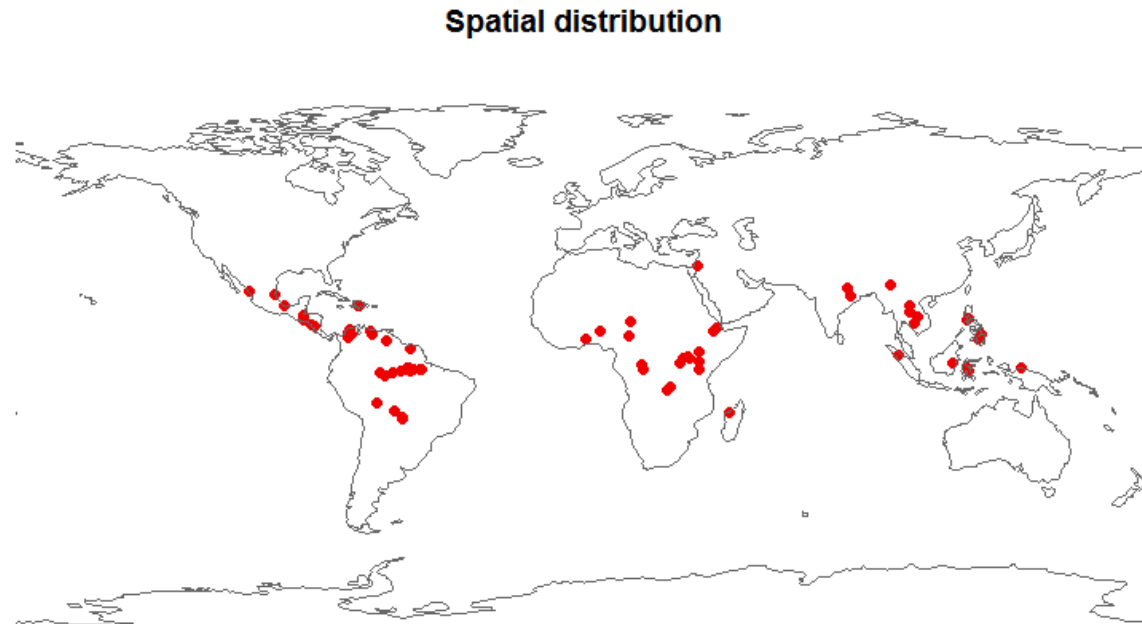
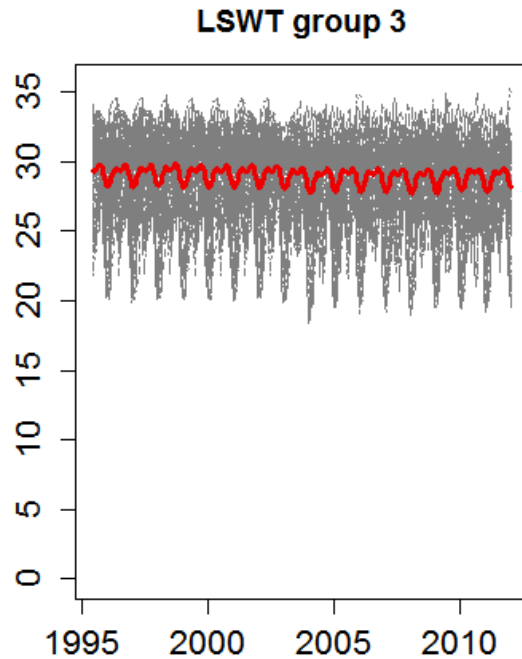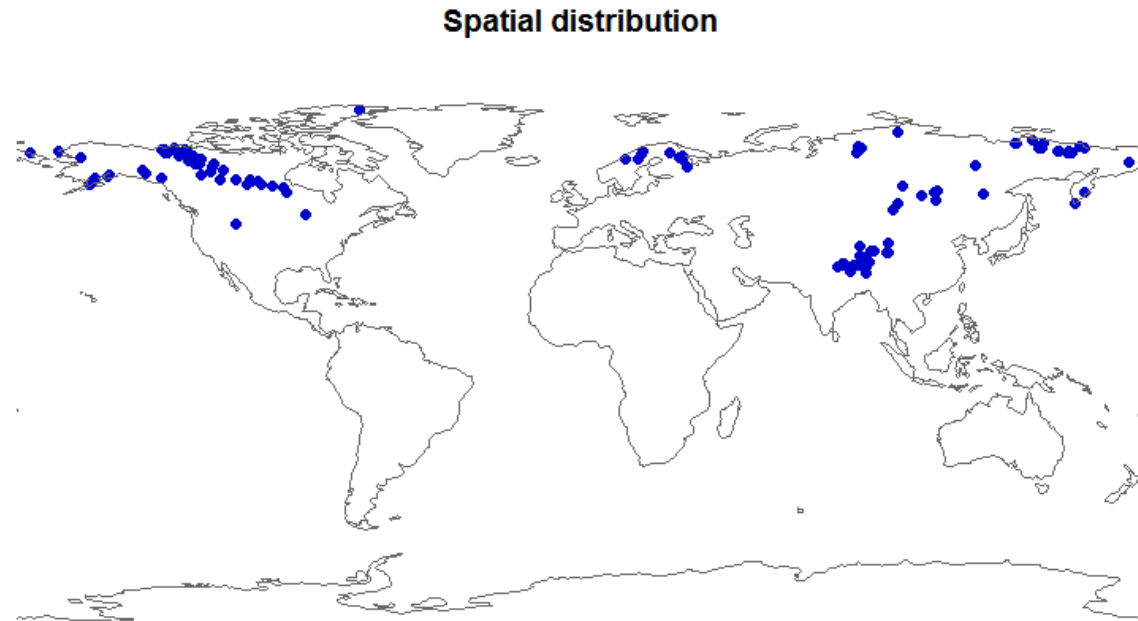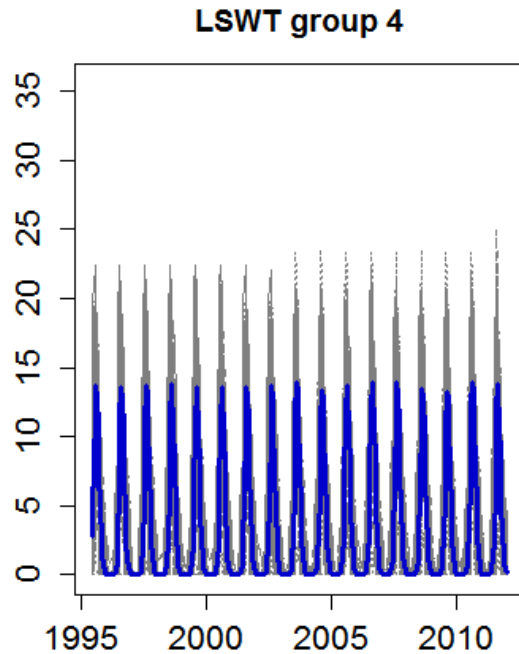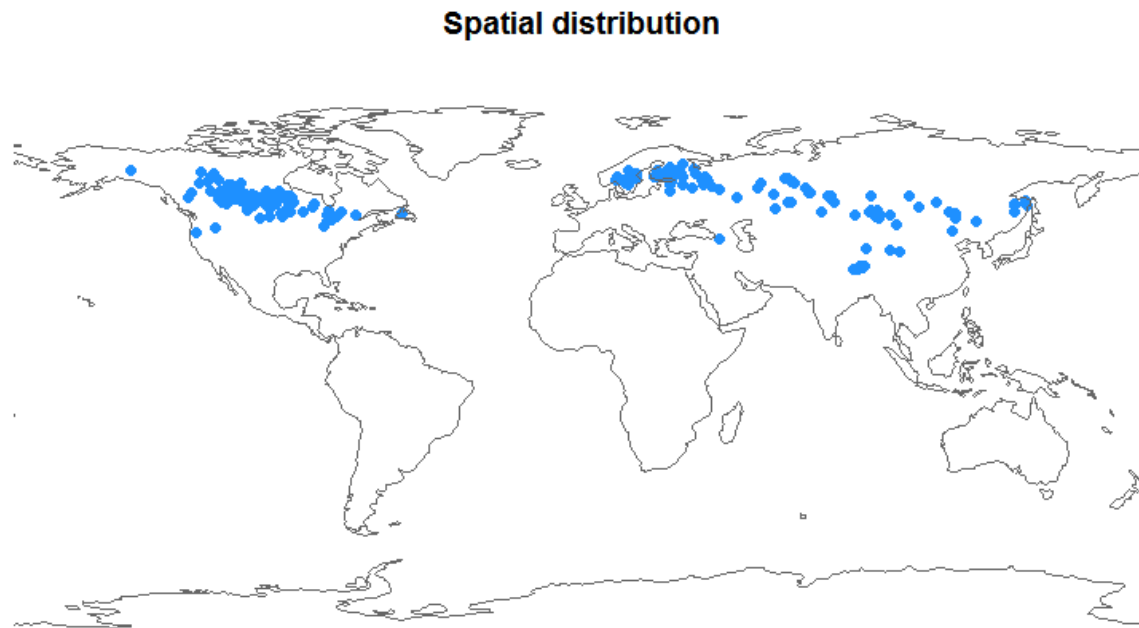# Clustering LSWT
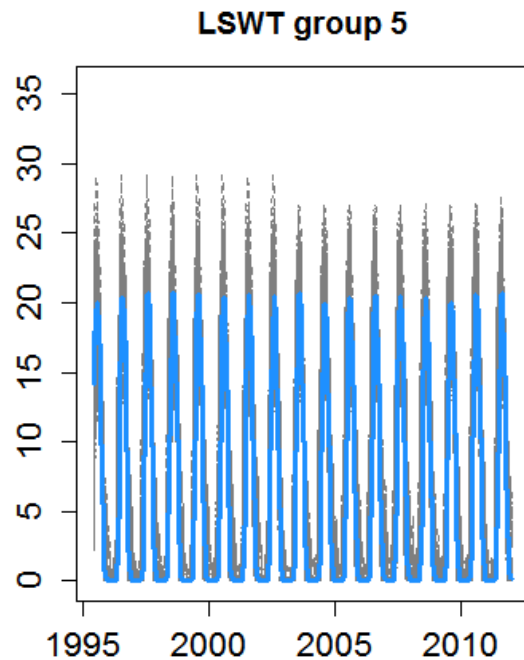
- Smoothed **LSWT time series** with cluster mean curves for 9 clusters
- The spatial distribution of the clusters are shown as maps

LSWT group 1

Spatial distribution
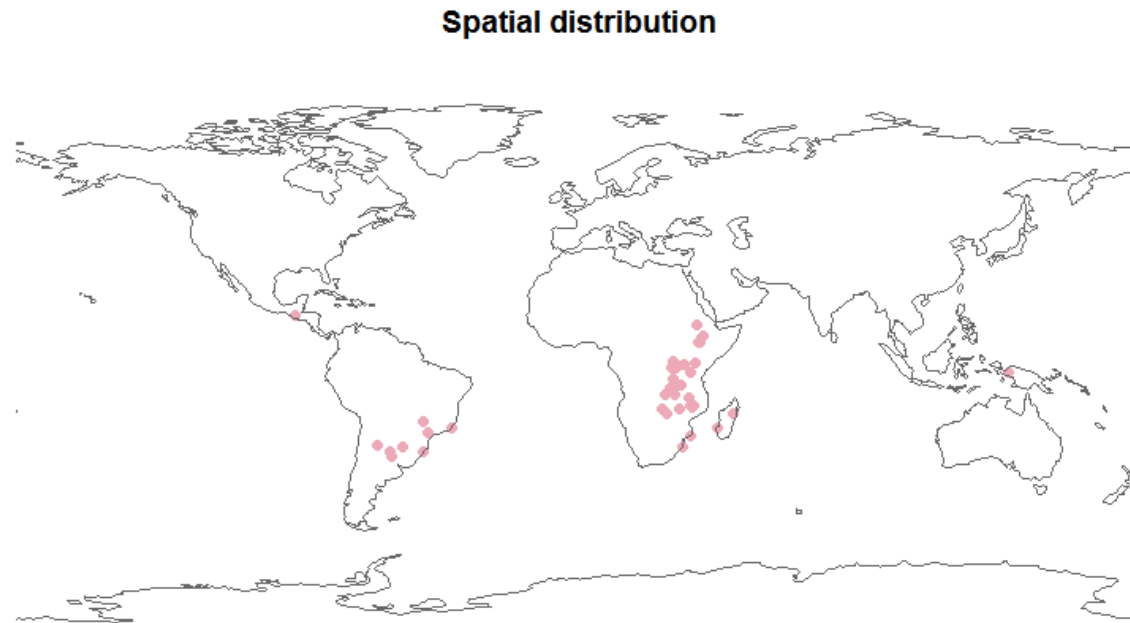
Clustering LSWT
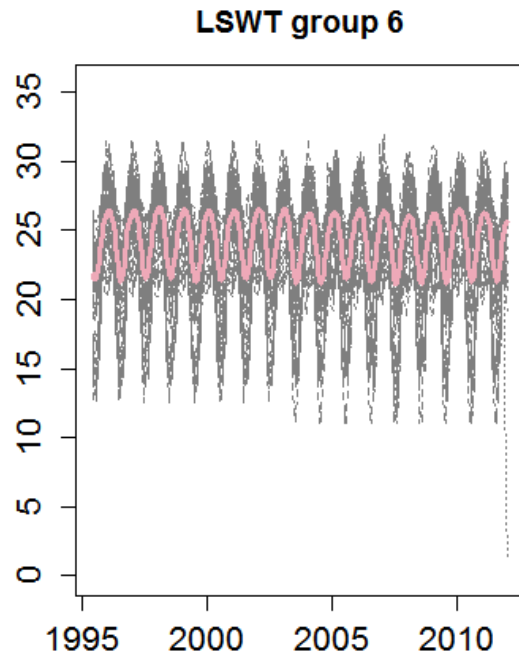
- Smoothed **LSWT time series** with cluster mean curves for 9 clusters
- The spatial distribution of the clusters are shown as maps

LSWT group 2

Spatial distribution

Clustering LSWT

- Smoothed **LSWT time series** with cluster mean curves for 9 clusters
- The spatial distribution of the clusters are shown as maps

LSWT group 3

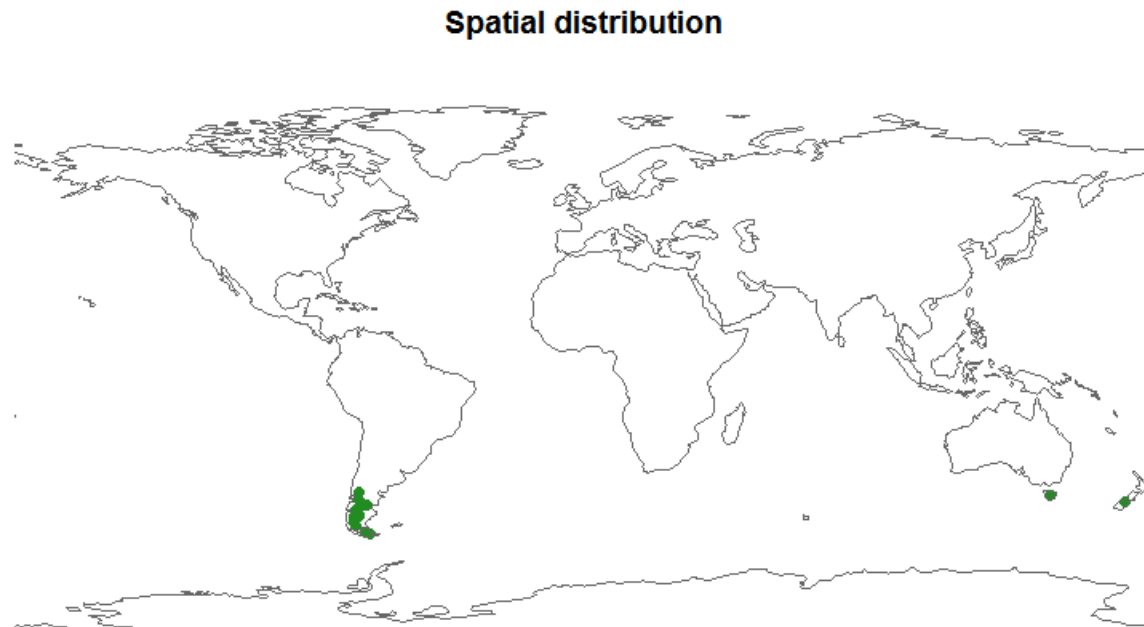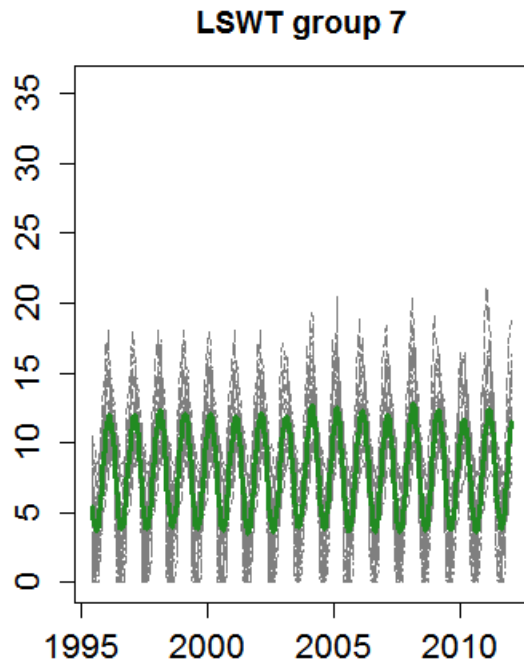Spatial distribution

Clustering LSWT

- Smoothed **LSWT time series** with cluster mean curves for 9 clusters
- The spatial distribution of the clusters are shown as maps

LSWT group 4

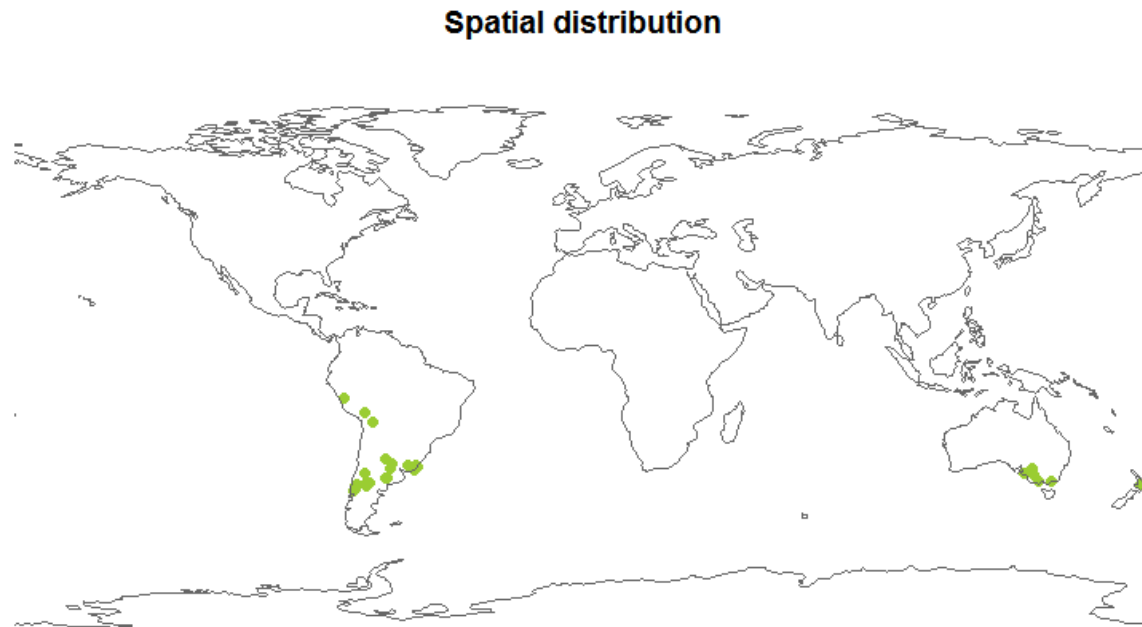Spatial distribution

Clustering LSWT

- Smoothed **LSWT time series** with cluster mean curves for 9 clusters
- The spatial distribution of the clusters are shown as maps
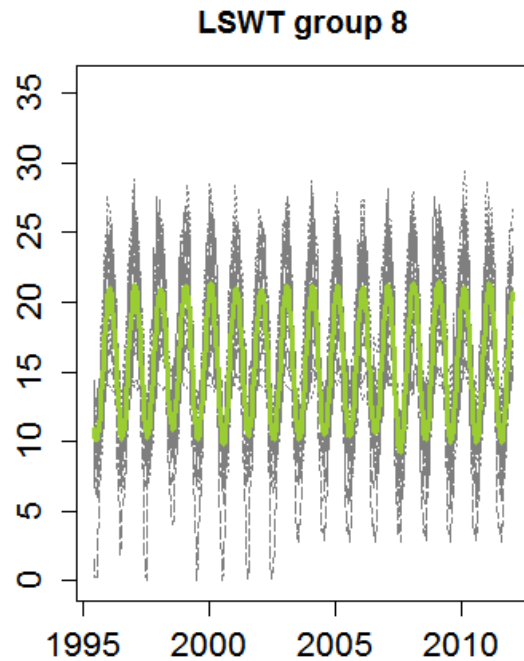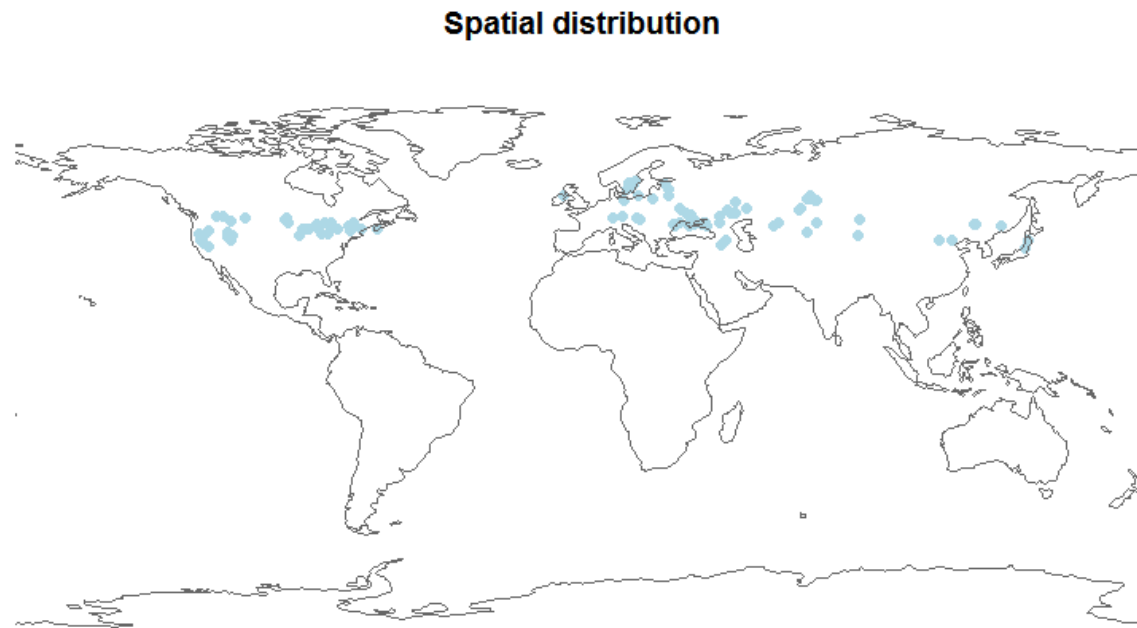

LSWT group 5

Spatial distribution

Clustering LSWT

- Smoothed **LSWT time series** with cluster mean curves for 9 clusters
- The spatial distribution of the clusters are shown as maps
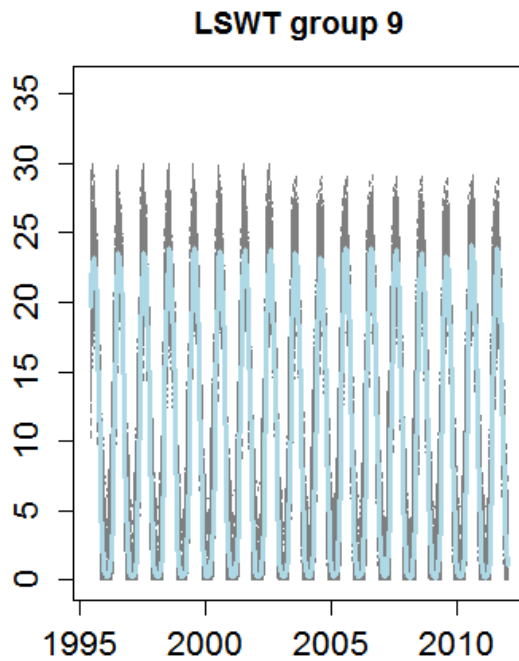

LSWT group 6

Spatial distribution

Clustering LSWT

- Smoothed **LSWT time series** with cluster mean curves for 9 clusters
- The spatial distribution of the clusters are shown as maps

LSWT group 7

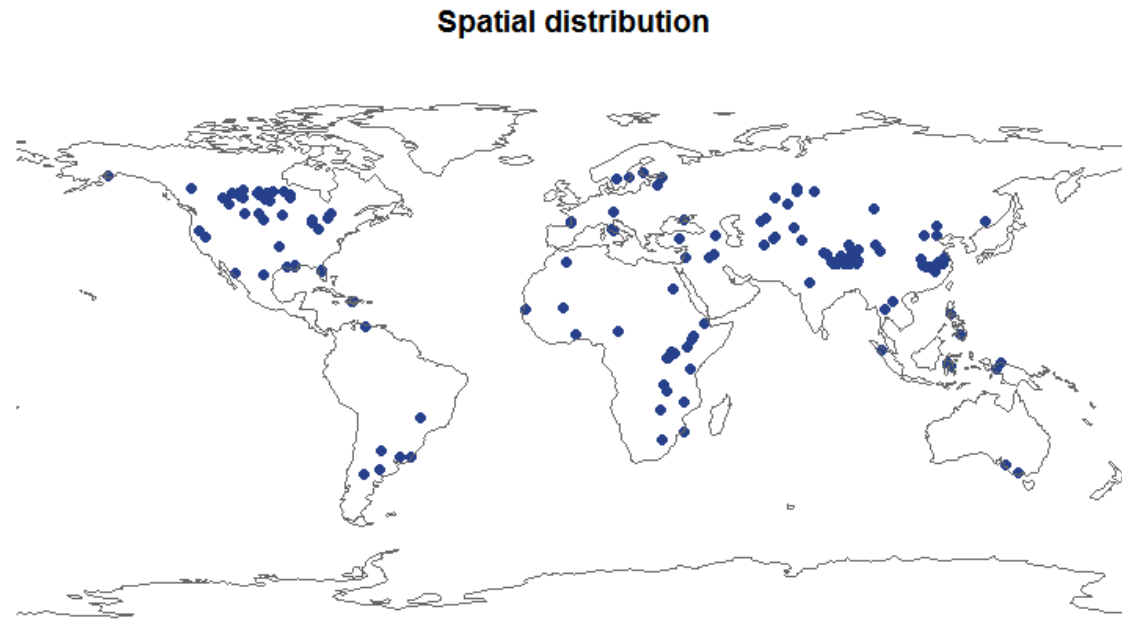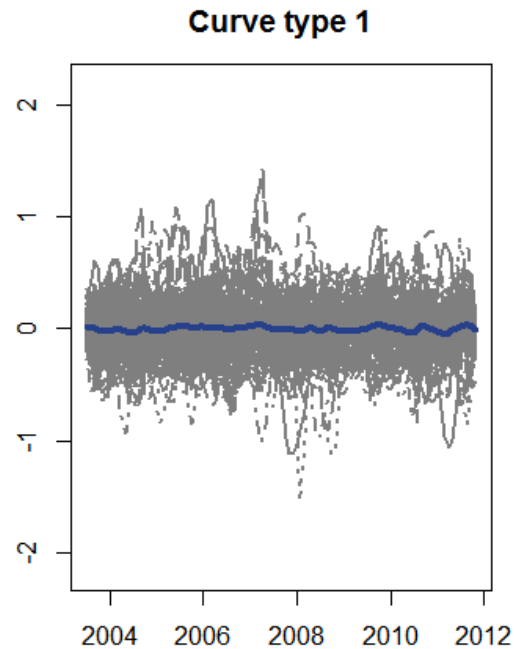Spatial distribution

Clustering LSWT

- Smoothed **LSWT time series** with cluster mean curves for 9 clusters
- The spatial distribution of the clusters are shown as maps

LSWT group 8

Spatial distribution

Clustering LSWT

- Smoothed **LSWT time series** with cluster mean curves for 9 clusters
- The spatial distribution of the clusters are shown as maps

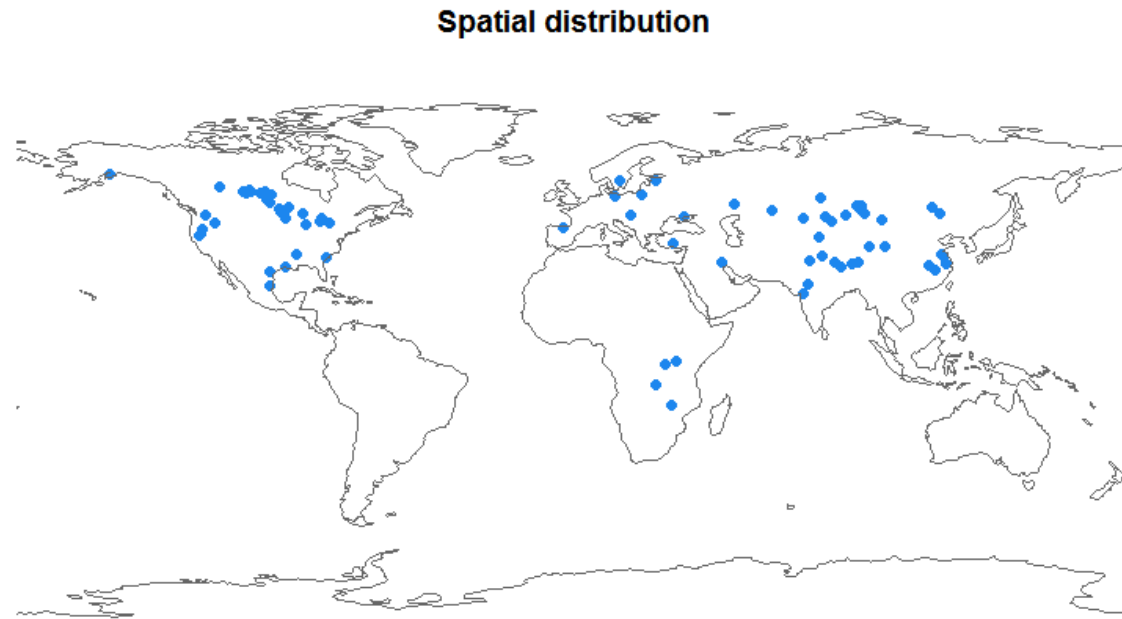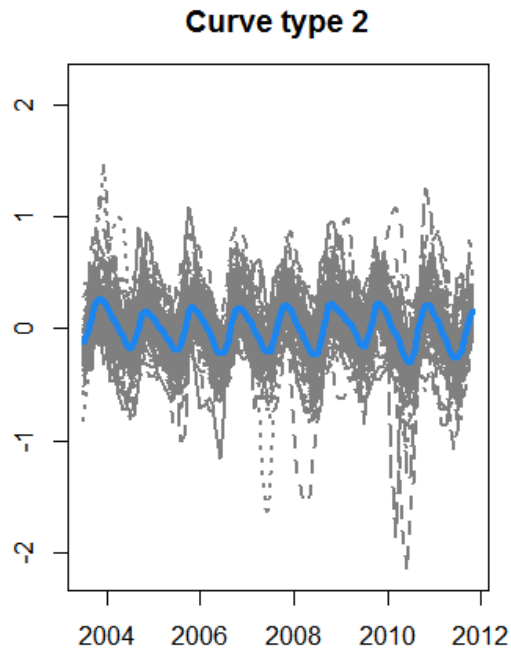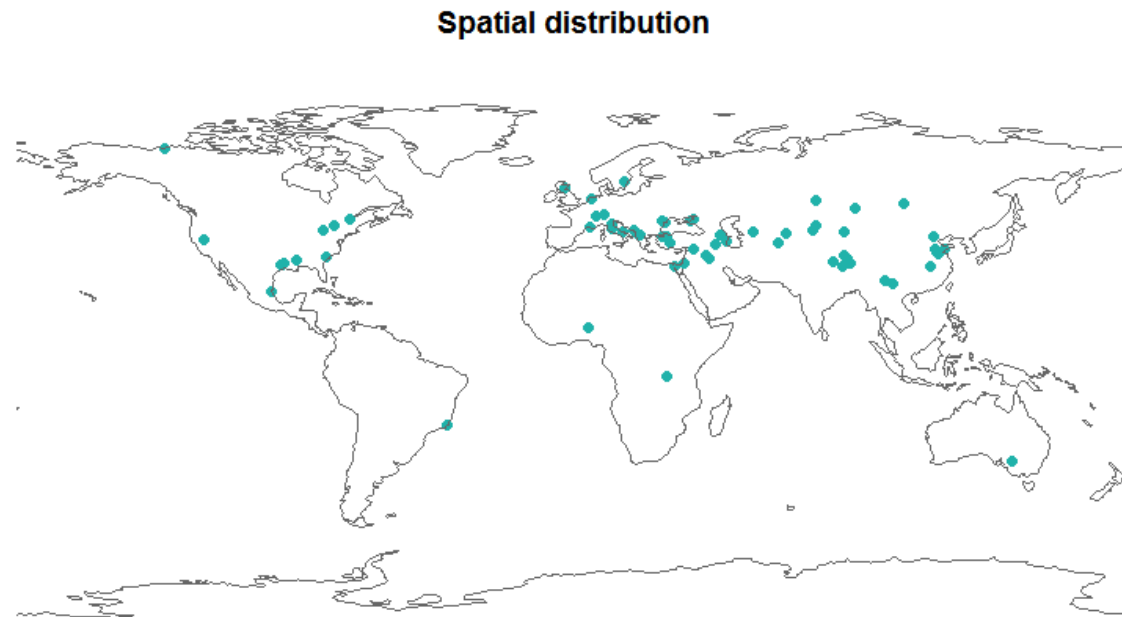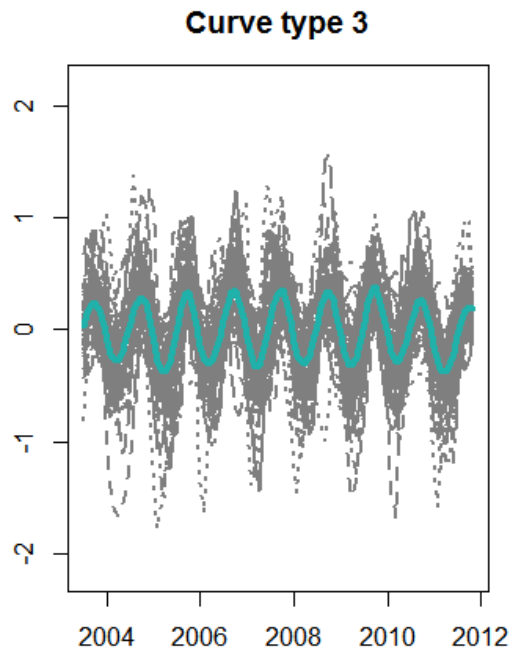LSWT group 9

Spatial distribution

Due to the large variability in the data, we investigate the **seasonality and trend** separately.

- Smoothed Chl-a **seasonal signals** with cluster mean curves and spatial distribution map (10 clusters)
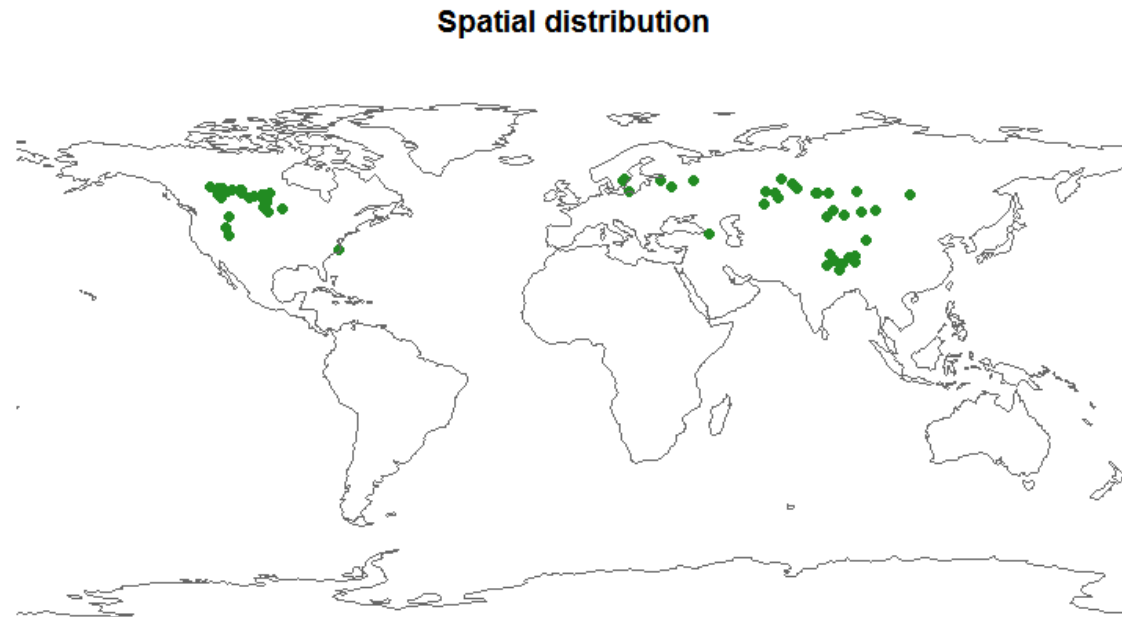
Clustering Chl-a

Due to the large variability in the data, we investigate the **seasonality and trend** separately.

- Smoothed Chl-a **seasonal signals** with cluster mean curves and spatial distribution map (10 clusters)
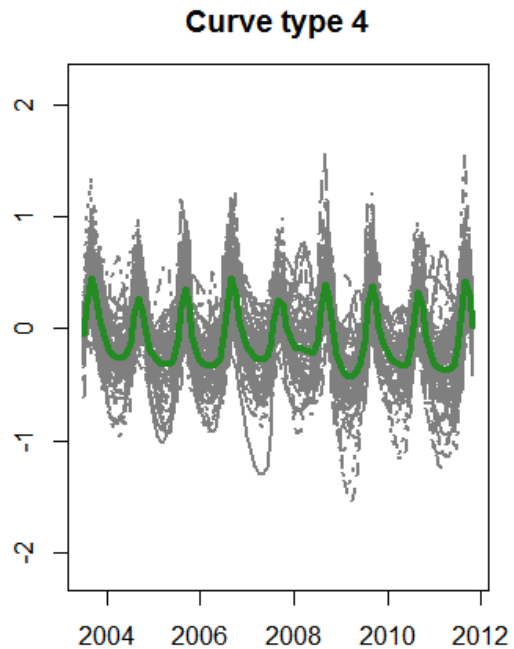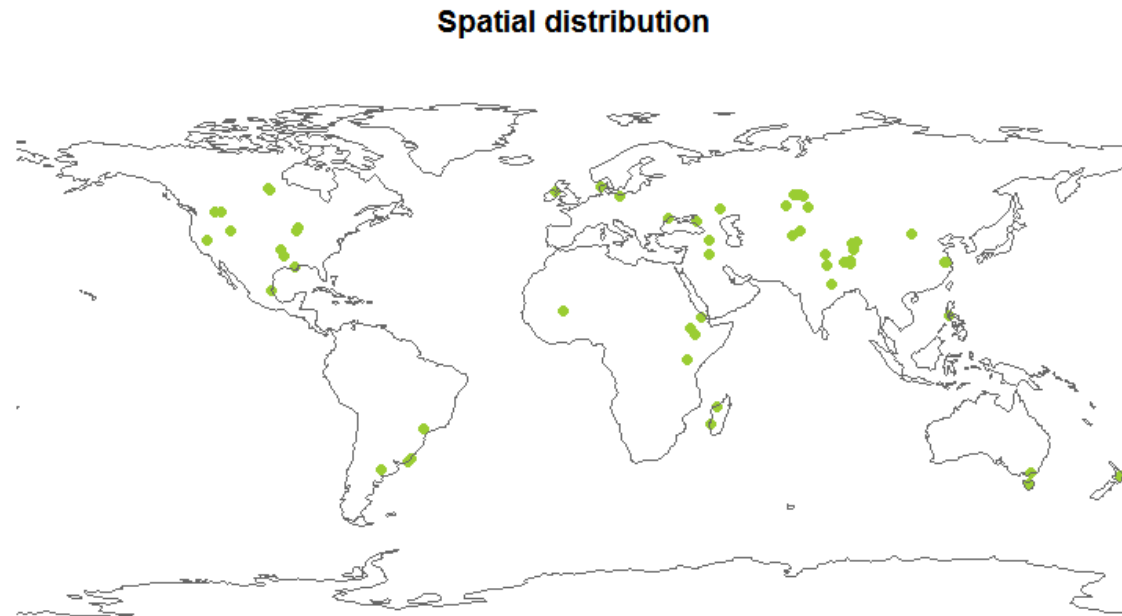
Clustering Chl-a

Due to the large variability in the data, we investigate the **seasonality and trend** separately.

- Smoothed Chl-a **seasonal signals** with cluster mean curves and spatial distribution map (10 clusters)
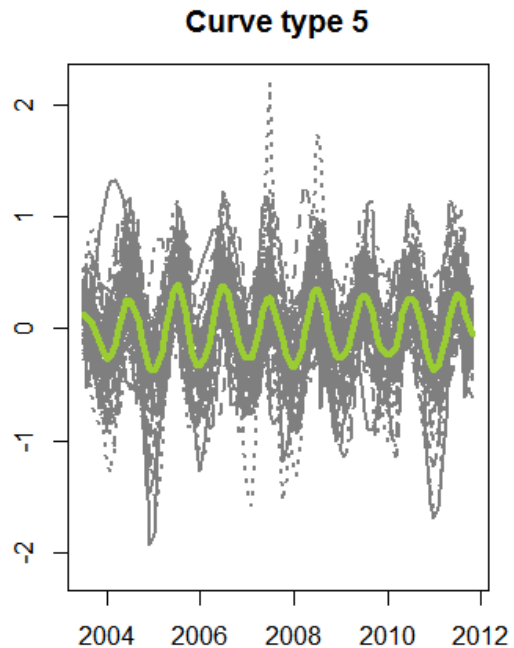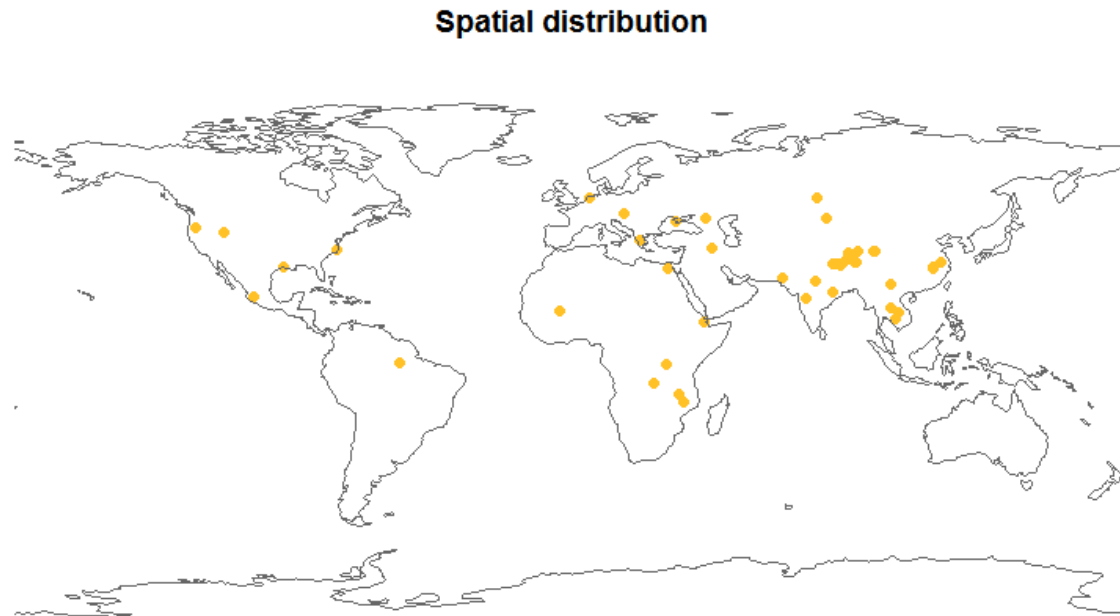
Clustering Chl-a

Due to the large variability in the data, we investigate the **seasonality and trend** separately.

- Smoothed Chl-a **seasonal signals** with cluster mean curves and spatial distribution map (10 clusters)

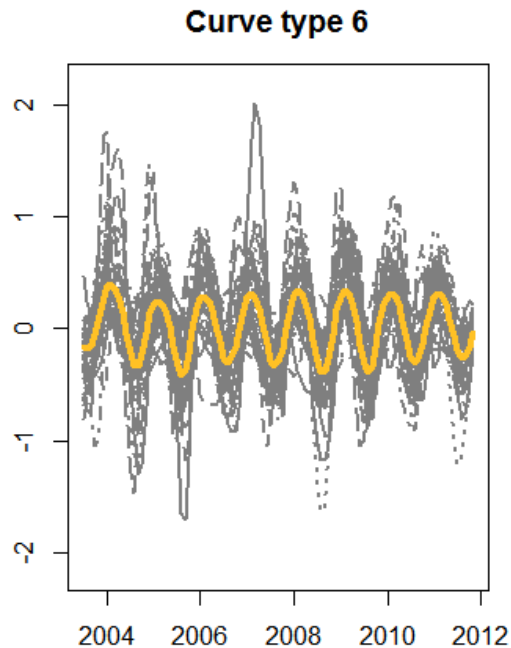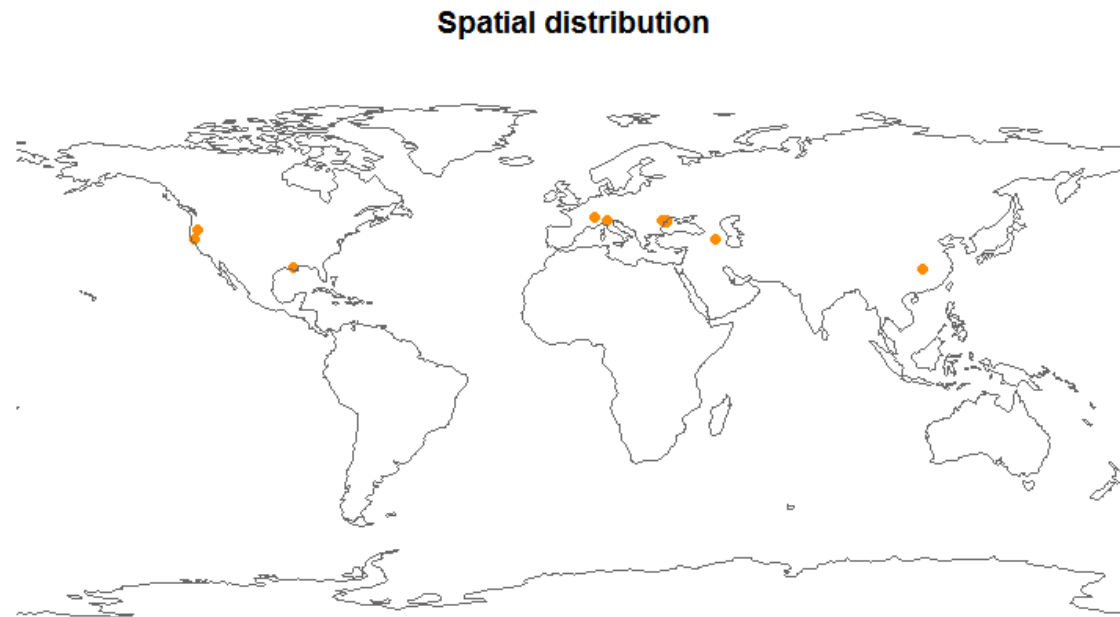Curve type 4

Spatial distribution

Clustering Chl-a

Due to the large variability in the data, we investigate the **seasonality and trend** separately.

- Smoothed Chl-a **seasonal signals** with cluster mean curves and spatial distribution map (10 clusters)
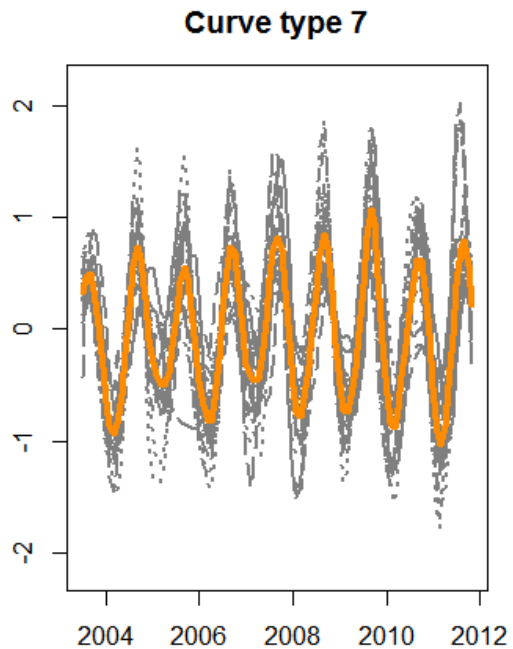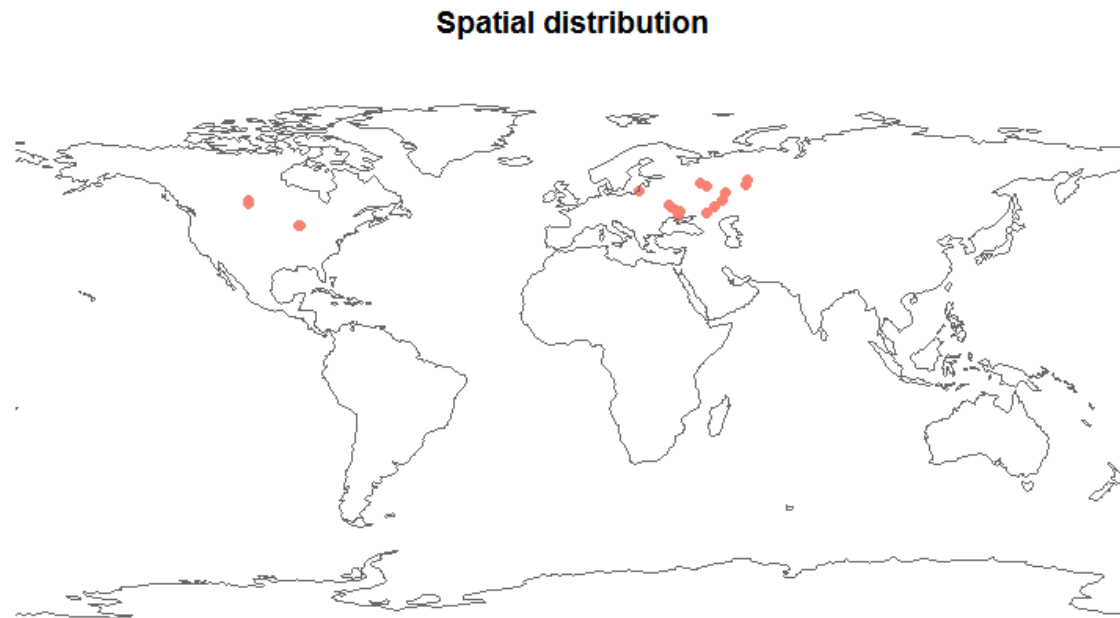
Clustering Chl-a

Due to the large variability in the data, we investigate the **seasonality and trend** separately.

- Smoothed Chl-a **seasonal signals** with cluster mean curves and spatial distribution map (10 clusters)

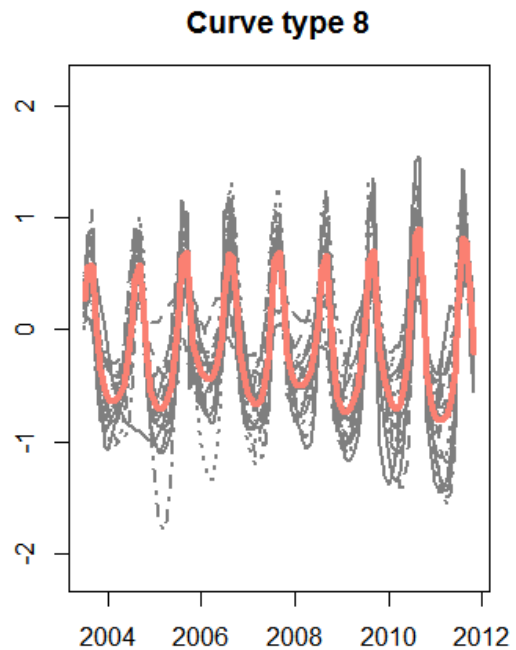Curve type 6

Spatial distribution

Clustering Chl-a

Due to the large variability in the data, we investigate the **seasonality and trend** separately.

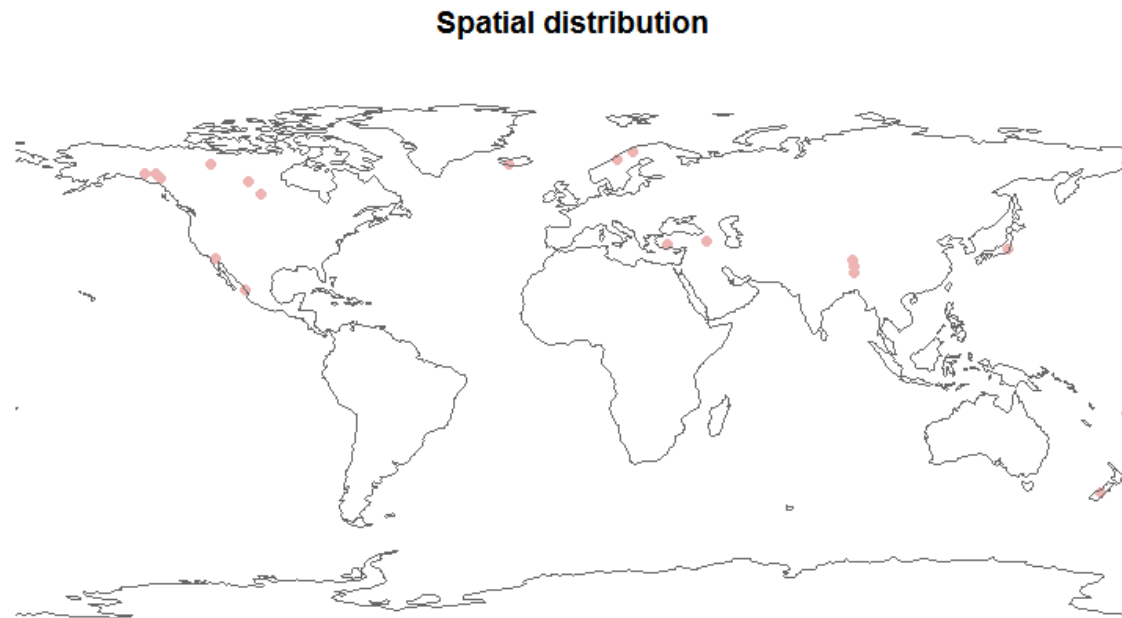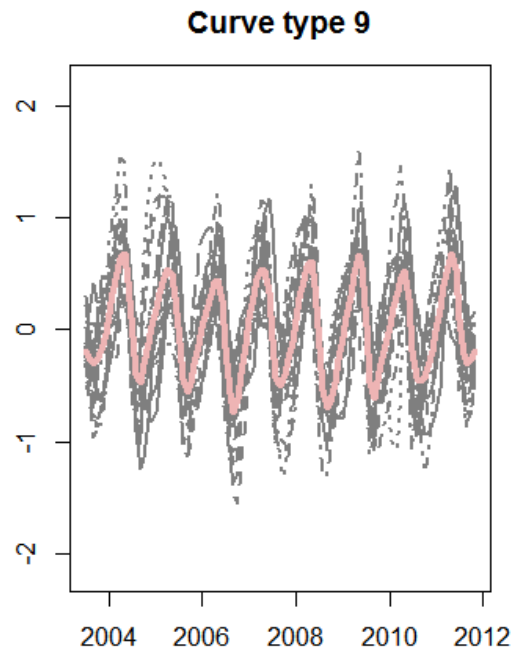- Smoothed Chl-a **seasonal signals** with cluster mean curves and spatial distribution map (10 clusters)

Due to the large variability in the data, we investigate the **seasonality and trend** separately.

- Smoothed Chl-a **seasonal signals** with cluster mean curves and spatial distribution map (10 clusters)

Curve type 8

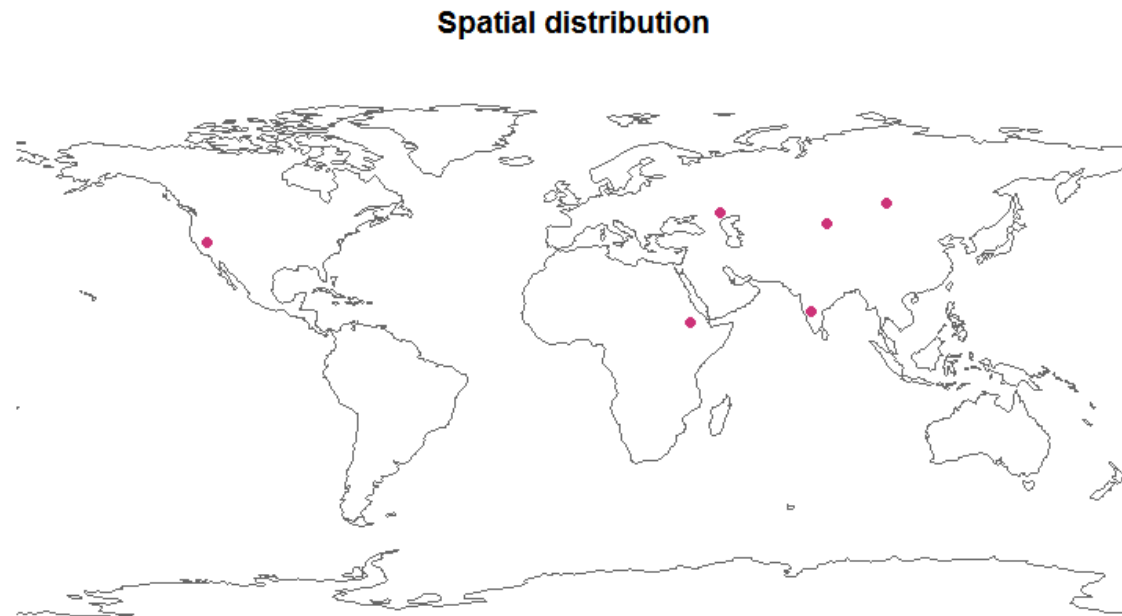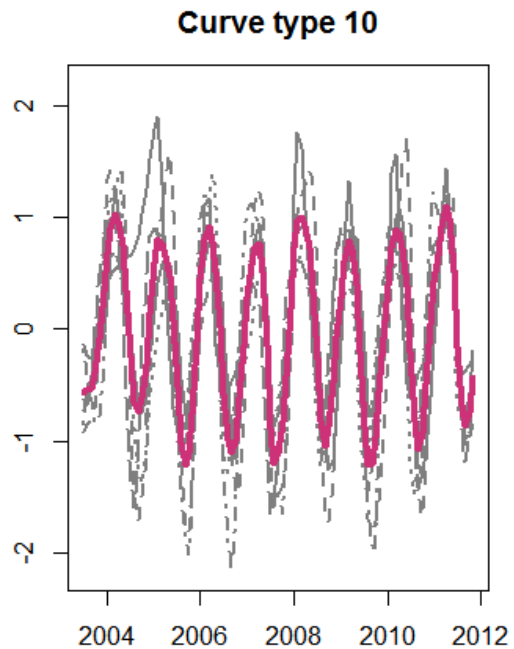Spatial distribution

Clustering Chl-a

Due to the large variability in the data, we investigate the **seasonality and trend** separately.

- Smoothed Chl-a **seasonal signals** with cluster mean curves and spatial distribution map (10 clusters)
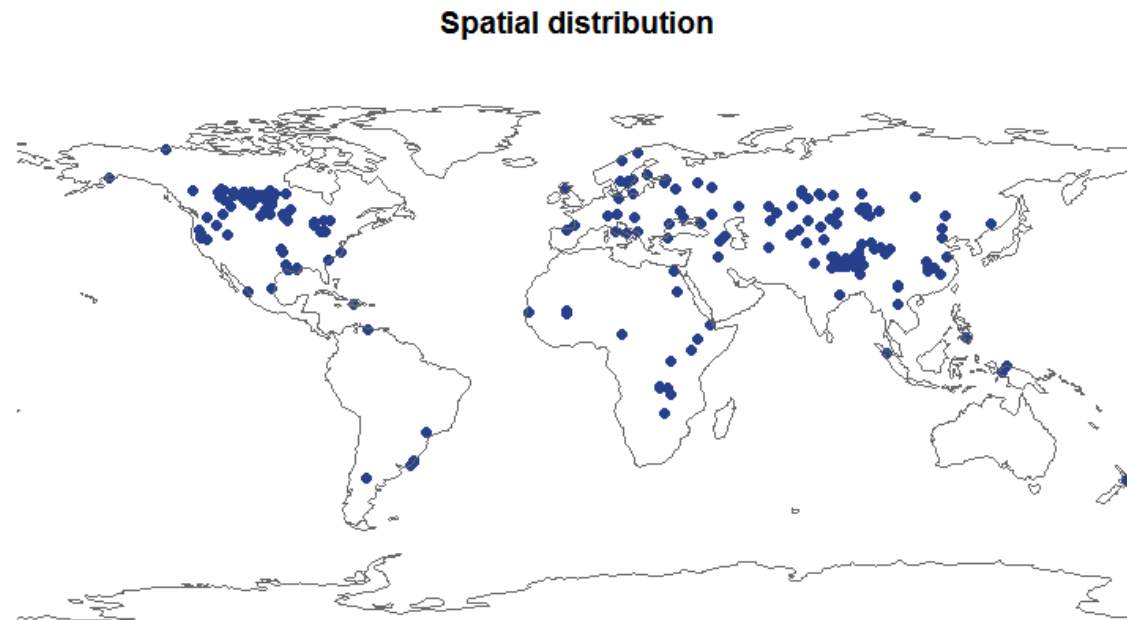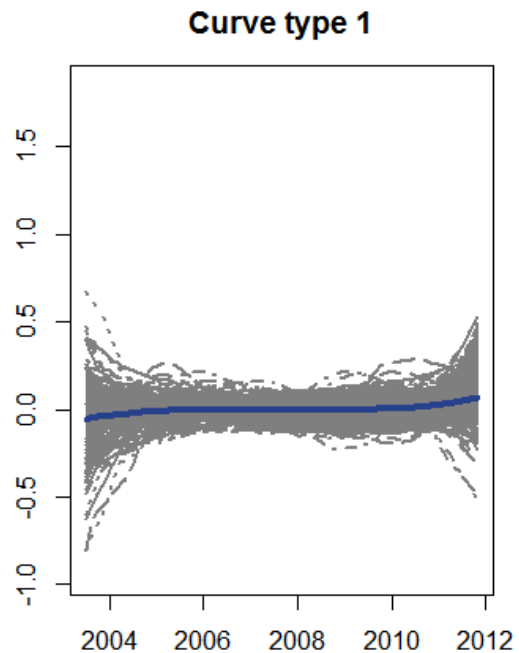
Clustering Chl-a

Due to the large variability in the data, we investigate the **seasonality and trend** separately.
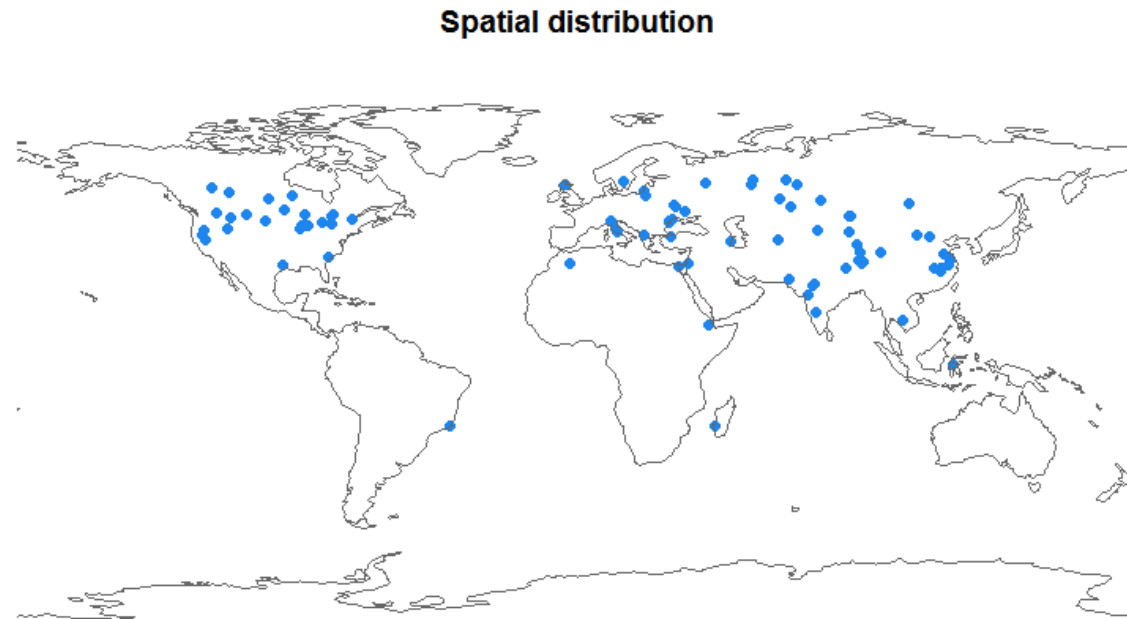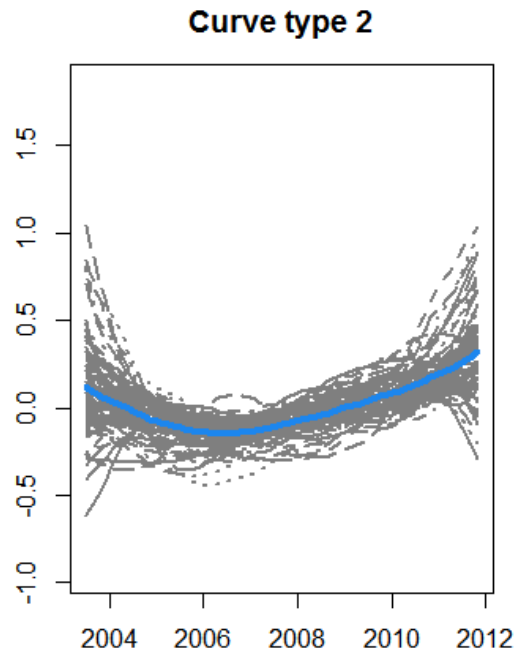
- Smoothed Chl-a **seasonal signals** with cluster mean curves and spatial distribution map (10 clusters)
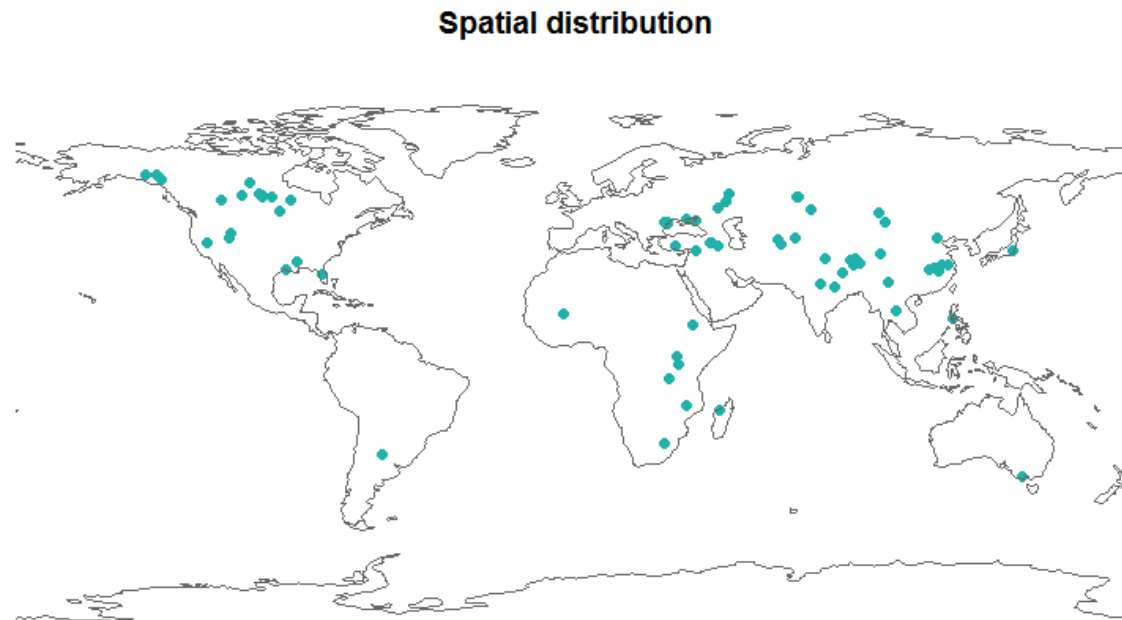
Clustering Chl-a

- Smoothed Chl-a **trend signals** with cluster mean curves and spatial distribution map (8 clusters)
- Tested the significance of the differences in the cluster mean curves

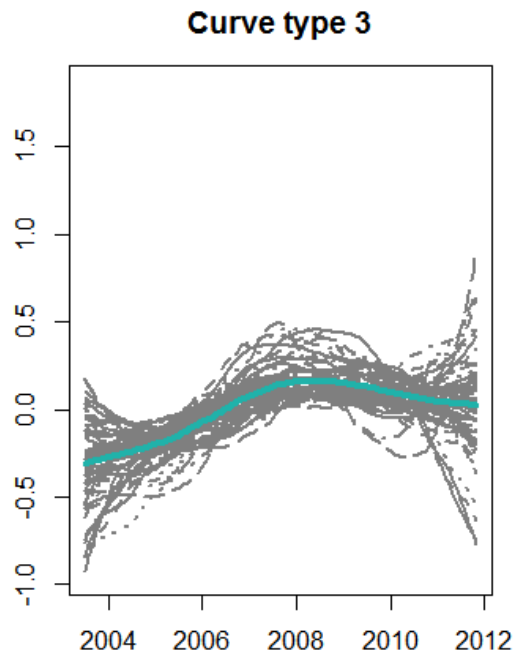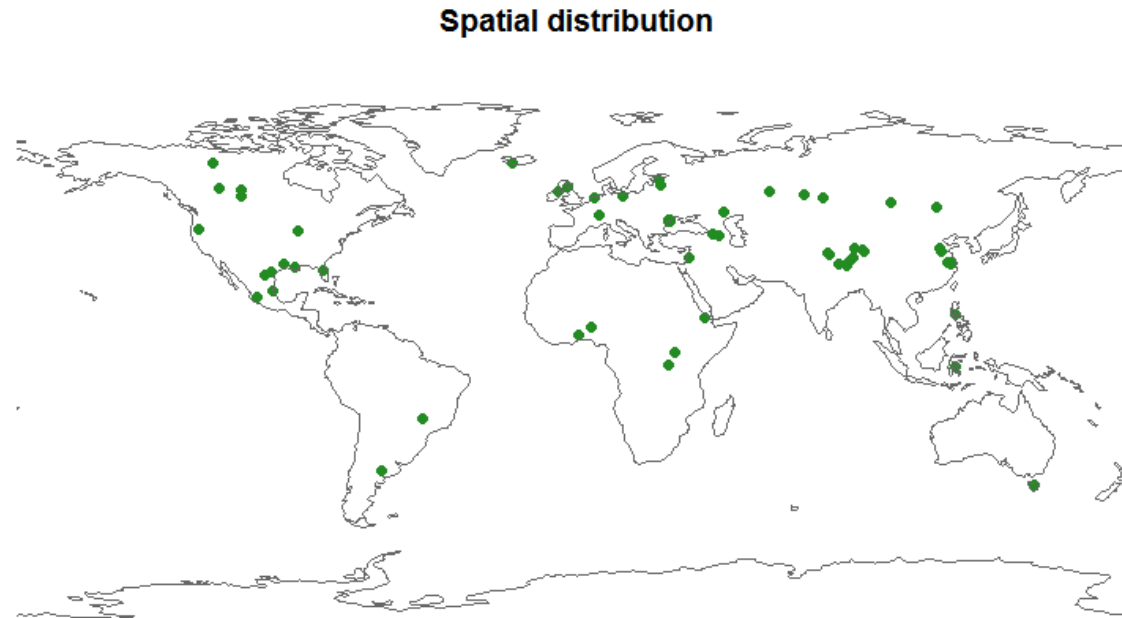Curve type 1

Spatial distribution

Clustering Chl-a

- Smoothed Chl-a **trend signals** with cluster mean curves and spatial distribution map (8 clusters)
- Tested the significance of the differences in the cluster mean curves

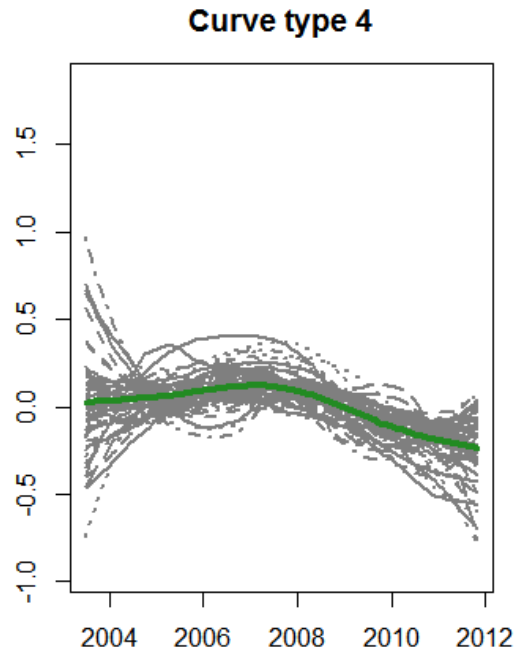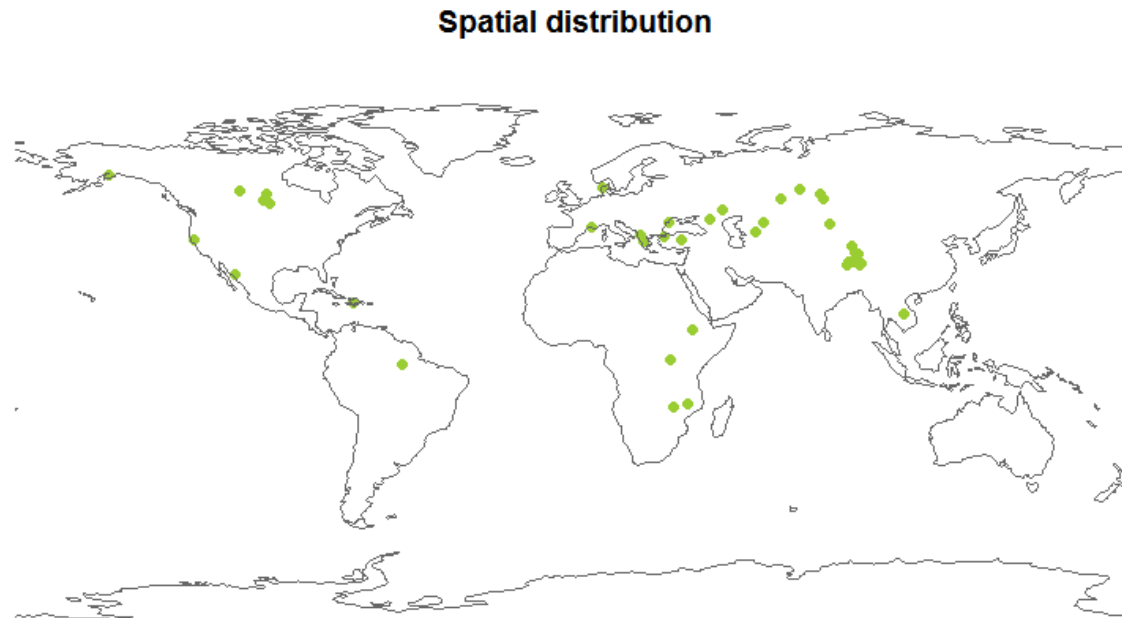Curve type 2

Spatial distribution

Clustering Chl-a

- Smoothed Chl-a **trend signals** with cluster mean curves and spatial distribution map (8 clusters)
- Tested the significance of the differences in the cluster mean curves

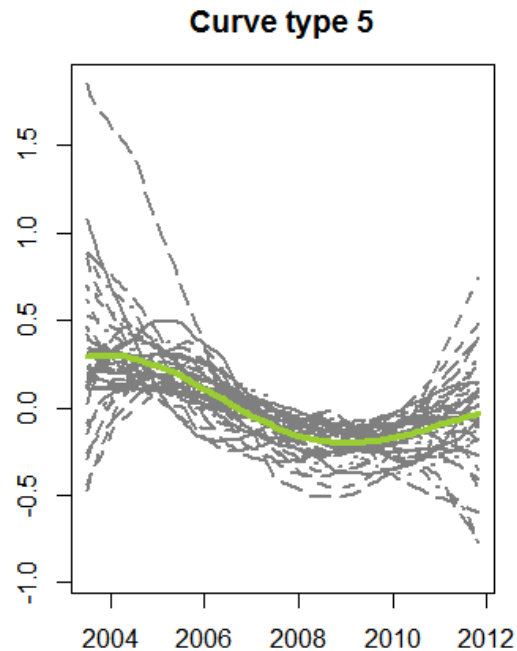Curve type 3

Spatial distribution

Clustering Chl-a

- Smoothed Chl-a **trend signals** with cluster mean curves and spatial distribution map (8 clusters)
- Tested the significance of the differences in the cluster mean curves

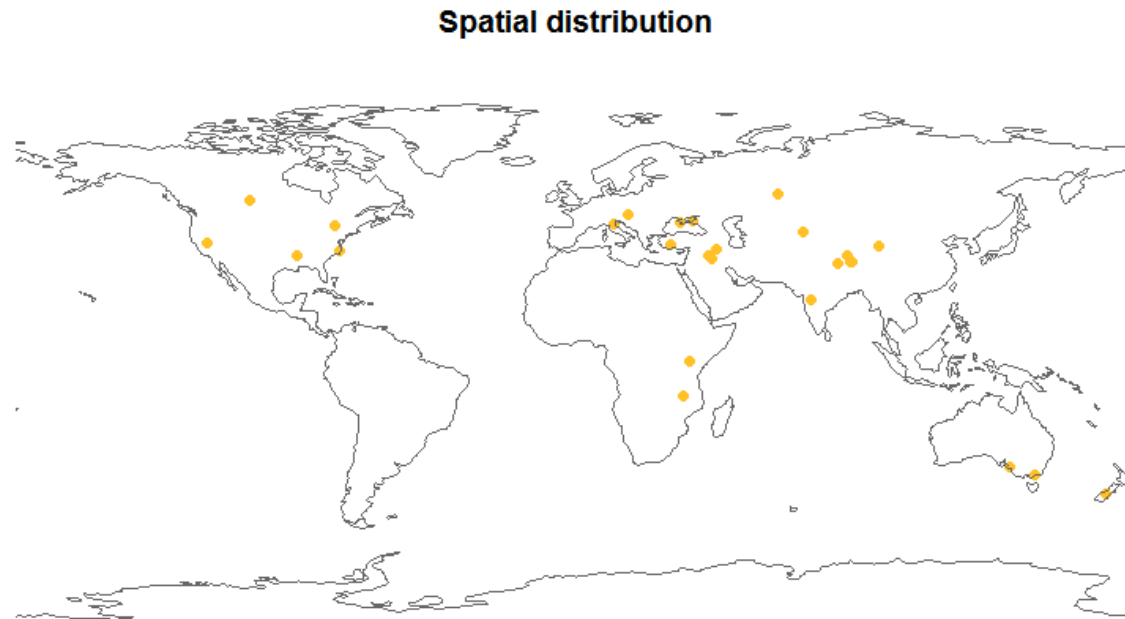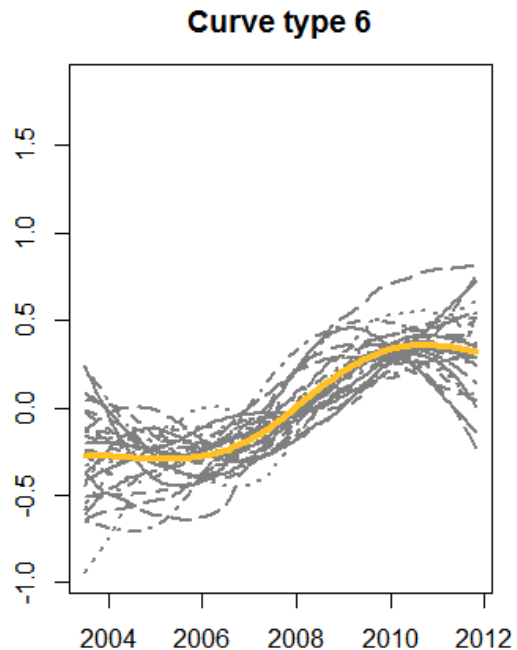Curve type 4

Spatial distribution

Clustering Chl-a

- Smoothed Chl-a **trend signals** with cluster mean curves and spatial distribution map (8 clusters)
- Tested the significance of the differences in the cluster mean curves

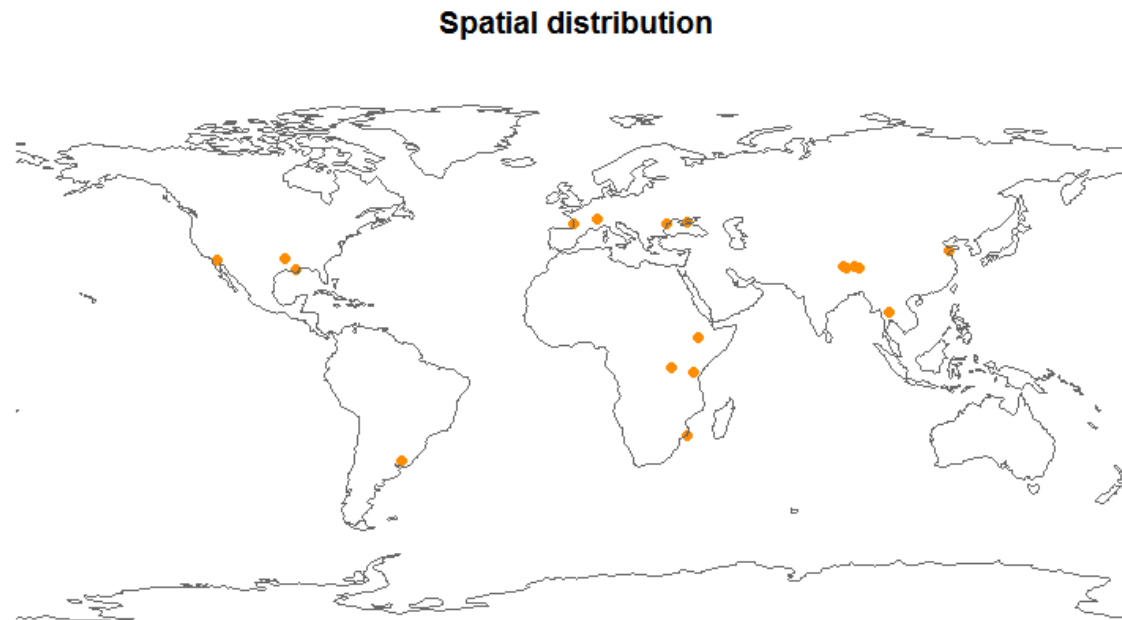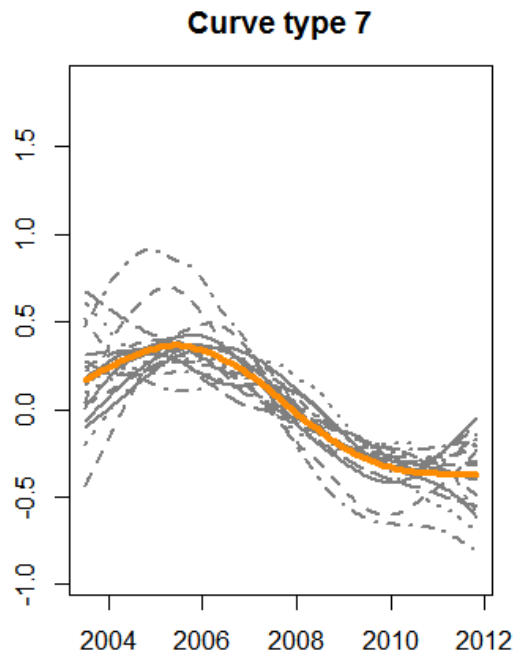Curve type 5

Spatial distribution

Clustering Chl-a

- Smoothed Chl-a **trend signals** with cluster mean curves and spatial distribution map (8 clusters)
- Tested the significance of the differences in the cluster mean curves
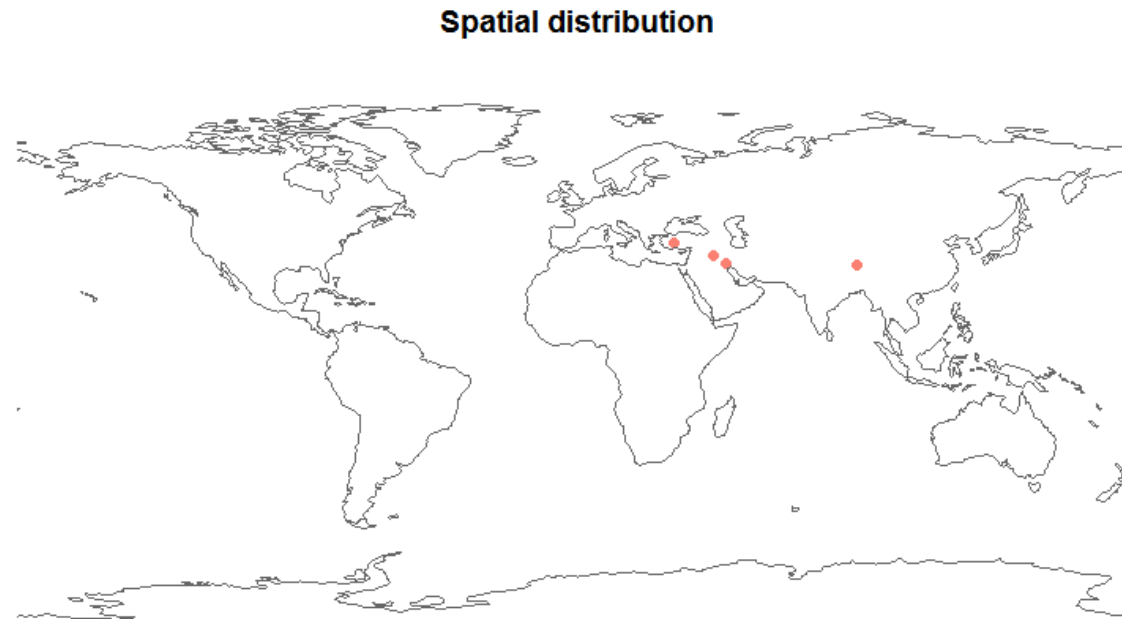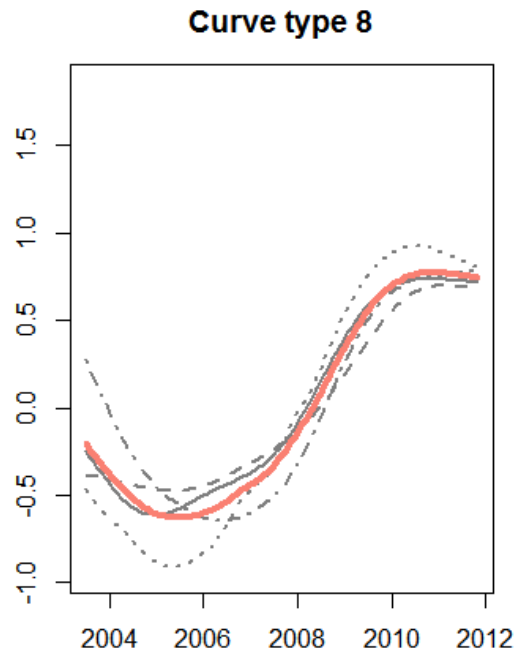
Clustering Chl-a

- Smoothed Chl-a **trend signals** with cluster mean curves and spatial distribution map (8 clusters)
- Tested the significance of the differences in the cluster mean curves

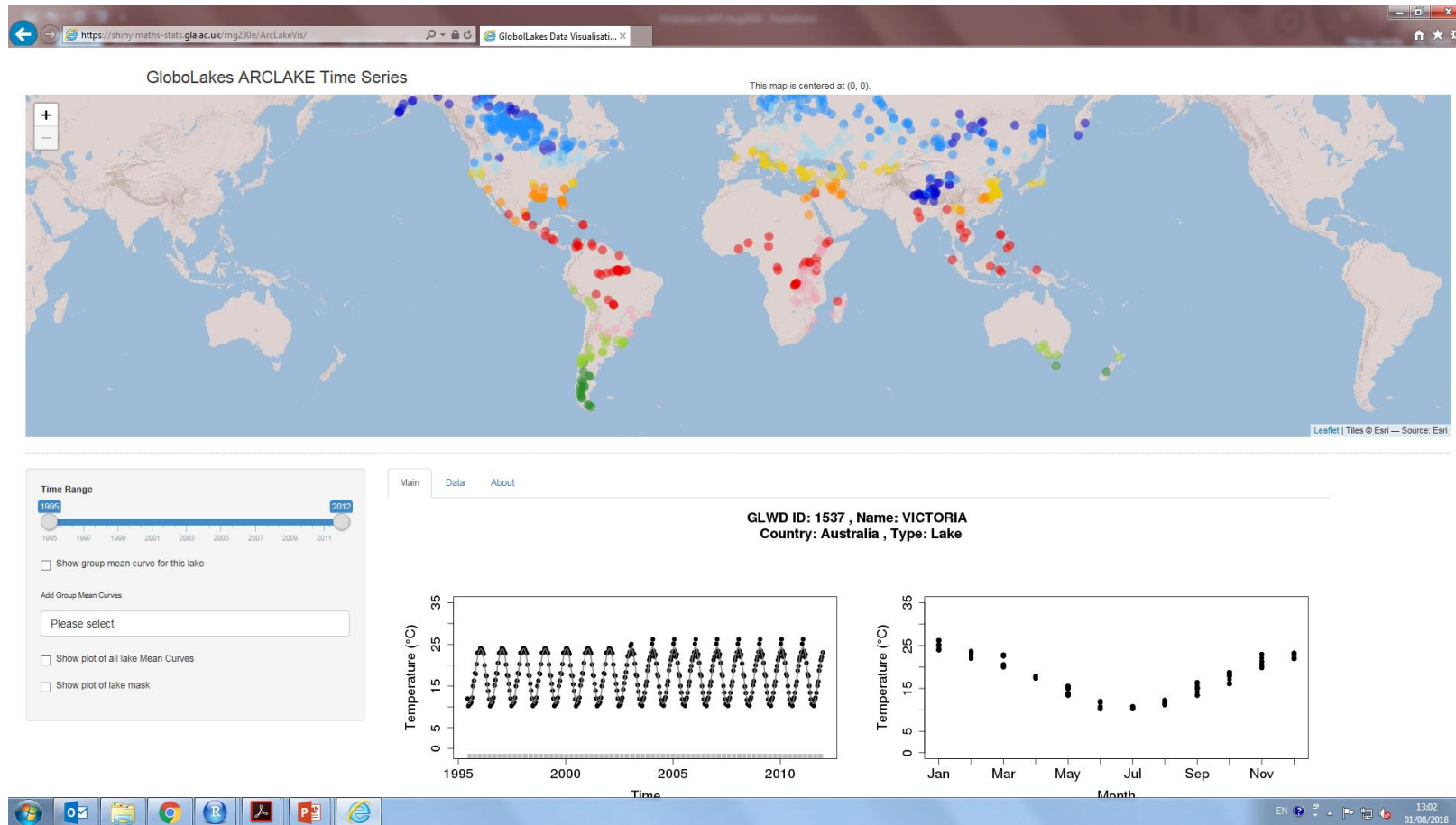Curve type 7

Spatial distribution

Clustering Chl-a

- Smoothed Chl-a **trend signals** with cluster mean curves and spatial distribution map (8 clusters)
- Tested the significance of the differences in the cluster mean curves

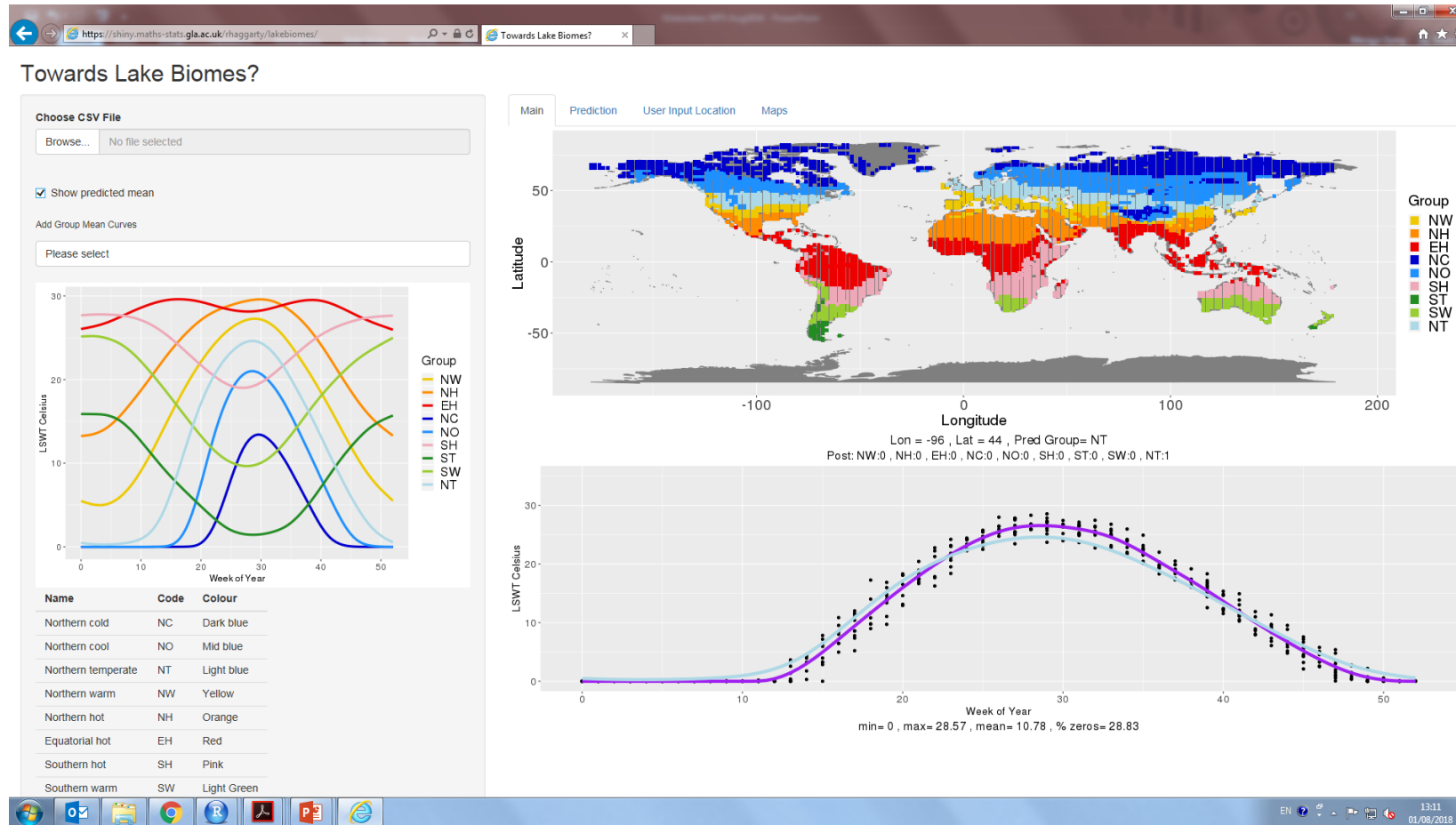Curve type 8

Spatial distribution

Analyses using the smoothed and/or clustered LSWT and Chl-a data and other attribution variables are investigated in the following presentations

- Status and Attribution of lake Water Quality

- Lake surface water temperature dynamics

Visualization – R shiny apps

https://shiny.maths-stats.gla.ac.uk/mg230e/ArcLakeVis/

Visualization – R shiny apps

https://shiny.maths-stats.gla.ac.uk/rhaggarty/lakebiomes/

Visualization – R shiny apps

https://shiny.maths-stats.gla.ac.uk/mg230e/ChlCluster/

Visualization – R shiny apps

Alternatively, these 3 apps can be accessed from GitHub using simple R command. The related links are given below.

https://github.com/GMY2018/ArcLakeVis

https://github.com/GMY2018/Lakebiomes

https://github.com/GMY2018/ChlCluster

Within Lake Clustering

- 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

- Where appropriate, within lake clustering can be applied in order to **identify groups of pixels which behave similarly** in terms of their temporal dynamics.

- This can be **computationally demanding** - there are potentially hundreds of thousands of time series to be clustered. One way of dealing with this is to look at **dimension reduction techniques** such as functional principal components analysis.

Within Lake Clustering – a case study

Lake Balaton Chl-a - 118 months between July 2002 and April 2012. 7943 pixels after 2 pixel boundary removed.

3 clusters identified

- (Top right) spatial distribution of the clusters
- (Bottom right) uncertainty in cluster membership
- (Bottom left) cluster mean curves
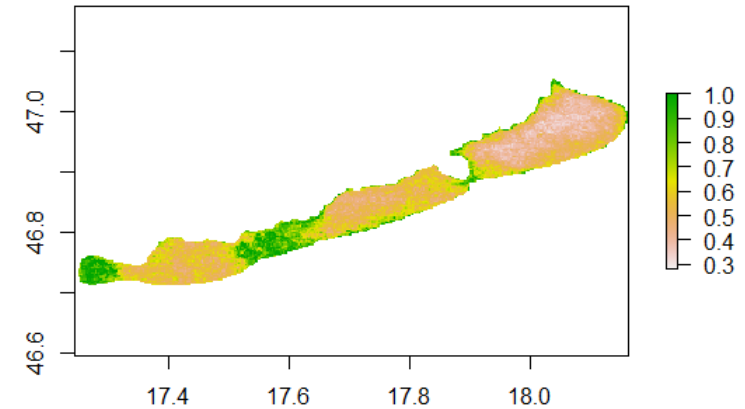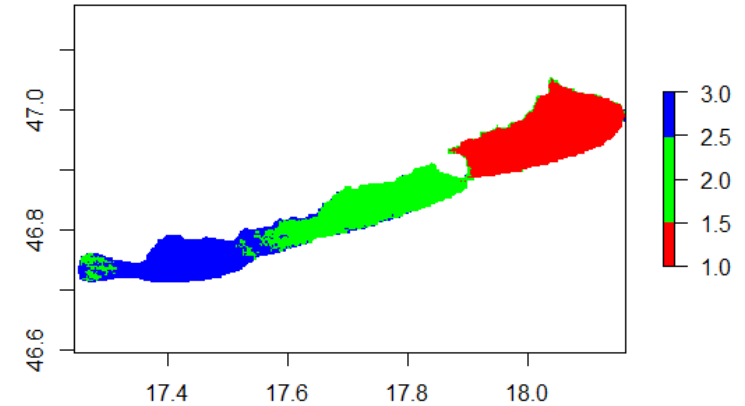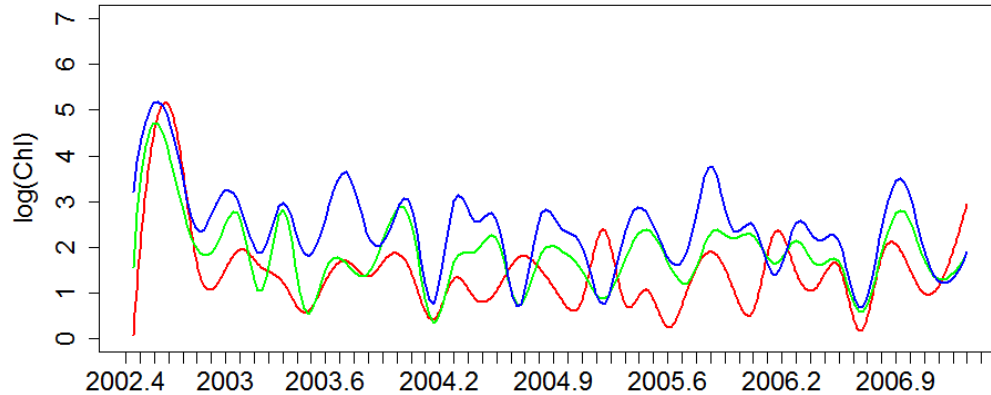
Thank you

Mengyi Gong
Room 228
School of Mathematics and Statistics
University of Glasgow

e mengyi.gong@glasgow.ac.uk
w www.gla.ac.uk

www.globolakes.ac.uk

@globolakes

NERC
SCIENCE OF THE
ENVIRONMENT