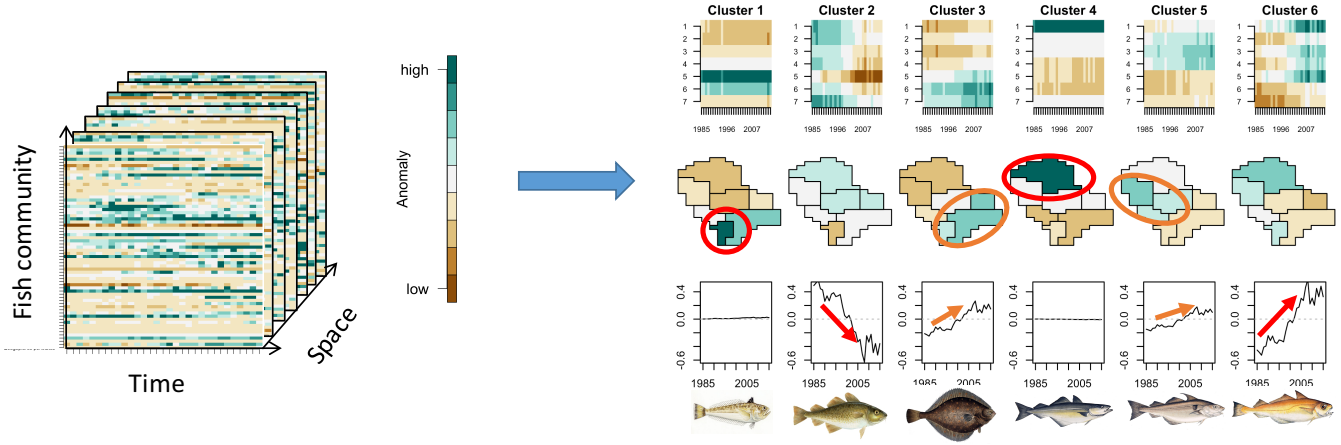
Universität Hamburg

MARmaED

Tensor Decomposition reveals spatiotemporal dynamics of fish communities

From a spatiotemporal dataset to well interpreted clusters



Romain Frelat, Martin Lindegren, Jens Floeter, Camilla Sguotti, Saskia Otto and Christian Möllmann

Bremen, 20th October 2016

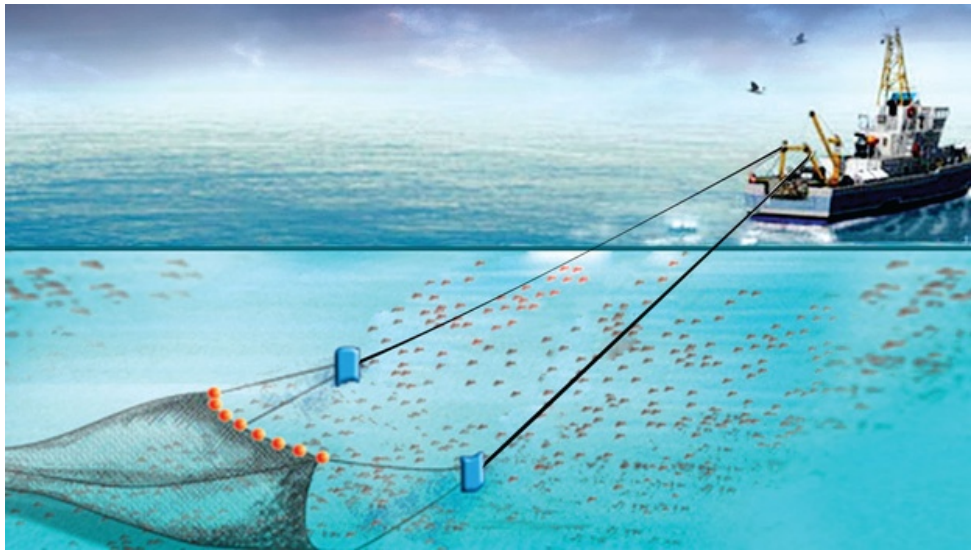
Zentrum für Marine Tropenökologie

Objectives

- What is the distribution of demersal fish communities in time and space ?
- What are the drivers of the spatiotemporal dynamics of fish communities?

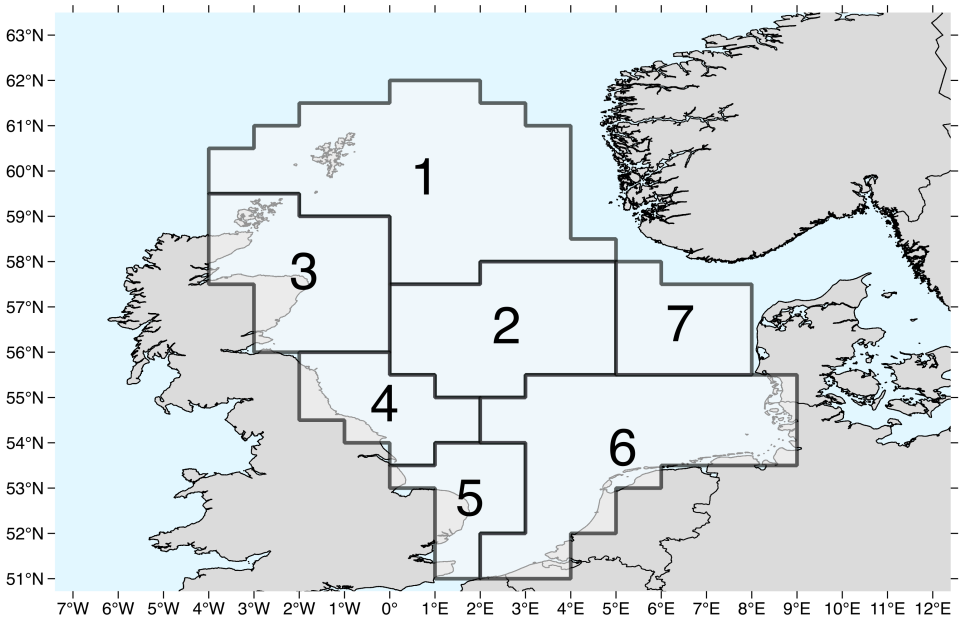
Empirical approach based on routinely collected dataset in the North Sea

Data



Source & structure

- ICES DATRAS data, NS-IBTS
- 7 roundfish areas
- 1985-2015



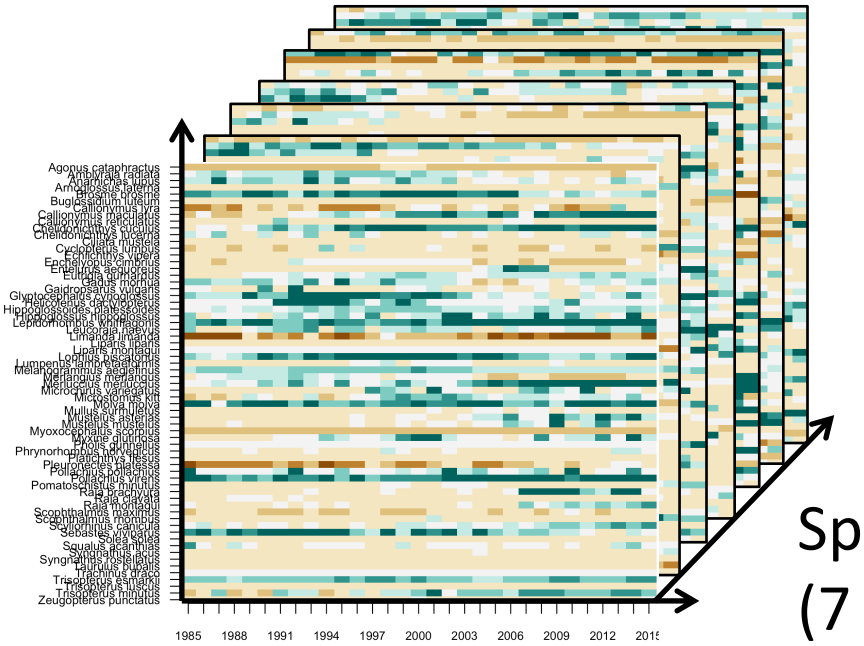
Pre processing

- Cleaning
- Abundance of species
  CPUE in number per hour,
  log transformed
- Remove rare species (30%)

Overview of dataset & method

Abundance of fishes in the North Sea

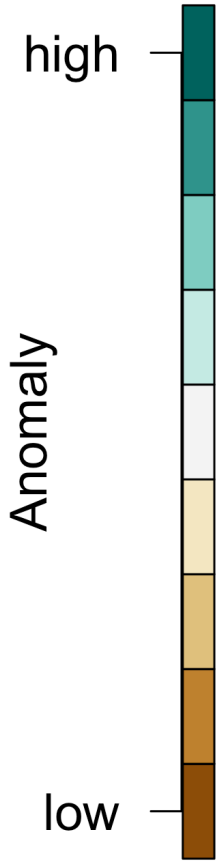
Fish community (65 species)



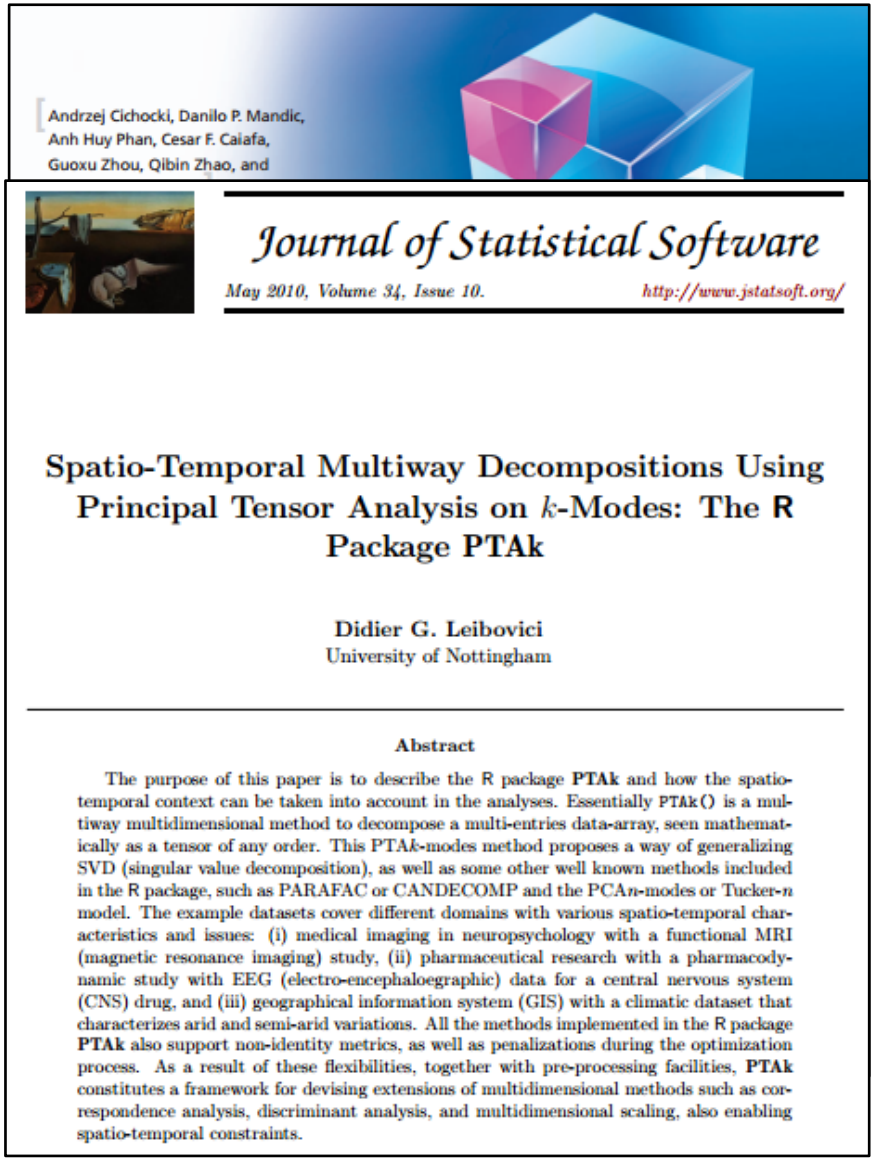
Time
(1985-2015 : 31 years)

Tensor

Space
(7 roundfish area)



Quick review and terminology



Andrzej Cichocki, Danilo P. Mandic,
Anh Huy Phan, Cesar F. Caiafa,
Guoxu Zhou, Qibin Zhao, and

Journal of Statistical Software

May 2010, Volume 34, Issue 10. *http://www.jstatsoft.org/*

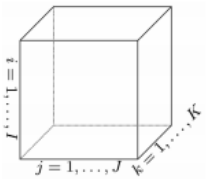
Spatio-Temporal Multiway Decompositions Using Principal Tensor Analysis on k-Modes: The R Package PTAk

Didier G. Leibovici
University of Nottingham

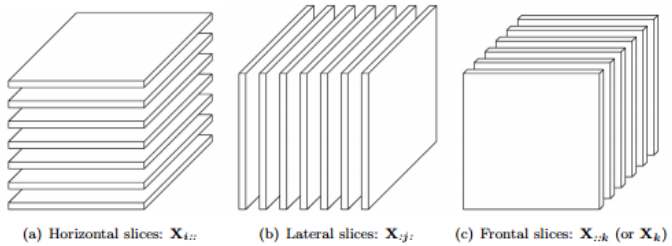
Abstract

The purpose of this paper is to describe the R package **PTAk** and how the spatio-temporal context can be taken into account in the analyses. Essentially `PTAk()` is a multiway multidimensional method to decompose a multi-entries data-array, seen mathematically as a tensor of any order. This PTAk-modes method proposes a way of generalizing SVD (singular value decomposition), as well as some other well known methods included in the R package, such as PARAFAC or CANDECOMP and the PCAn-modes or Tucker-n model. The example datasets cover different domains with various spatio-temporal characteristics and issues: (i) medical imaging in neuropsychology with a functional MRI (magnetic resonance imaging) study, (ii) pharmaceutical research with a pharmacodynamic study with EEG (electro-encephaloegraphic) data for a central nervous system (CNS) drug, and (iii) geographical information system (GIS) with a climatic dataset that characterizes arid and semi-arid variations. All the methods implemented in the R package **PTAk** also support non-identity metrics, as well as penalizations during the optimization process. As a result of these flexibilities, together with pre-processing facilities, **PTAk** constitutes a framework for devising extensions of multidimensional methods such as correspondence analysis, discriminant analysis, and multidimensional scaling, also enabling spatio-temporal constraints.

60's Tucker Decomposition
70's PARAFAC and CANDECOMP
90's Principal Tensor Analysis

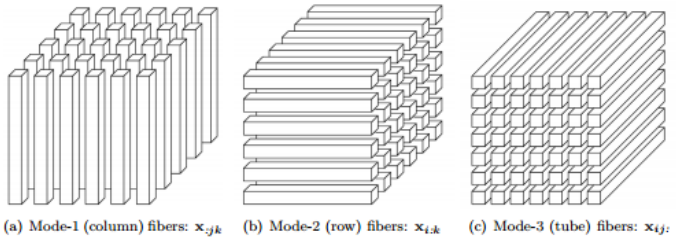


Tensor (3D)



(a) Horizontal slices: $\mathbf{X}_{i::}$ (b) Lateral slices: $\mathbf{X}_{:j:}$ (c) Frontal slices: $\mathbf{X}_{::k}$ (or $\mathbf{X}_k$)

Slice (2D),

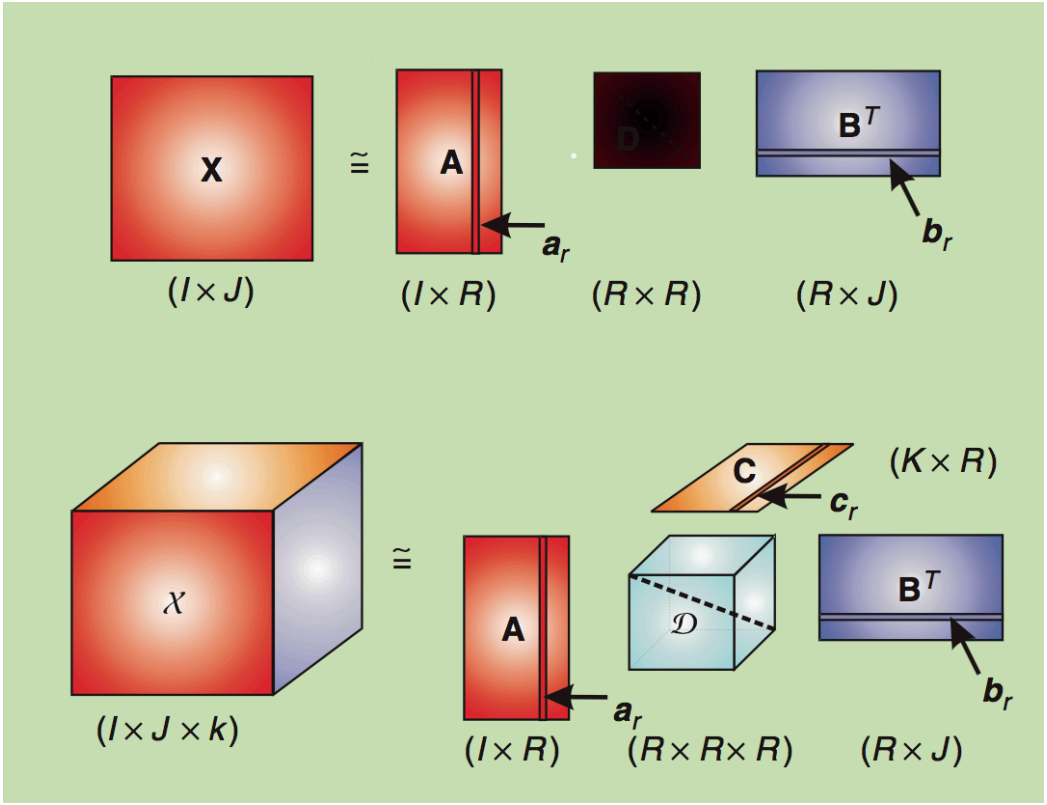


(a) Mode-1 (column) fibers: $\mathbf{x}_{:jk}$ (b) Mode-2 (row) fibers: $\mathbf{x}_{i:k}$ (c) Mode-3 (tube) fibers: $\mathbf{x}_{ij:}$

Fibers (1D)

Cichocki A. et al. 2015, IEEE Signal Processing
Leibovici D. 2010, Journal of Statistical Software

Illustration of multivariate analysis



Objectives :

- Simplification of the data by revealing the main patterns 'hidden' in the data.
- Explain the maximum of the variability in the minimum number of components

Cichocki A. et al. 2015, IEEE Signal Processing

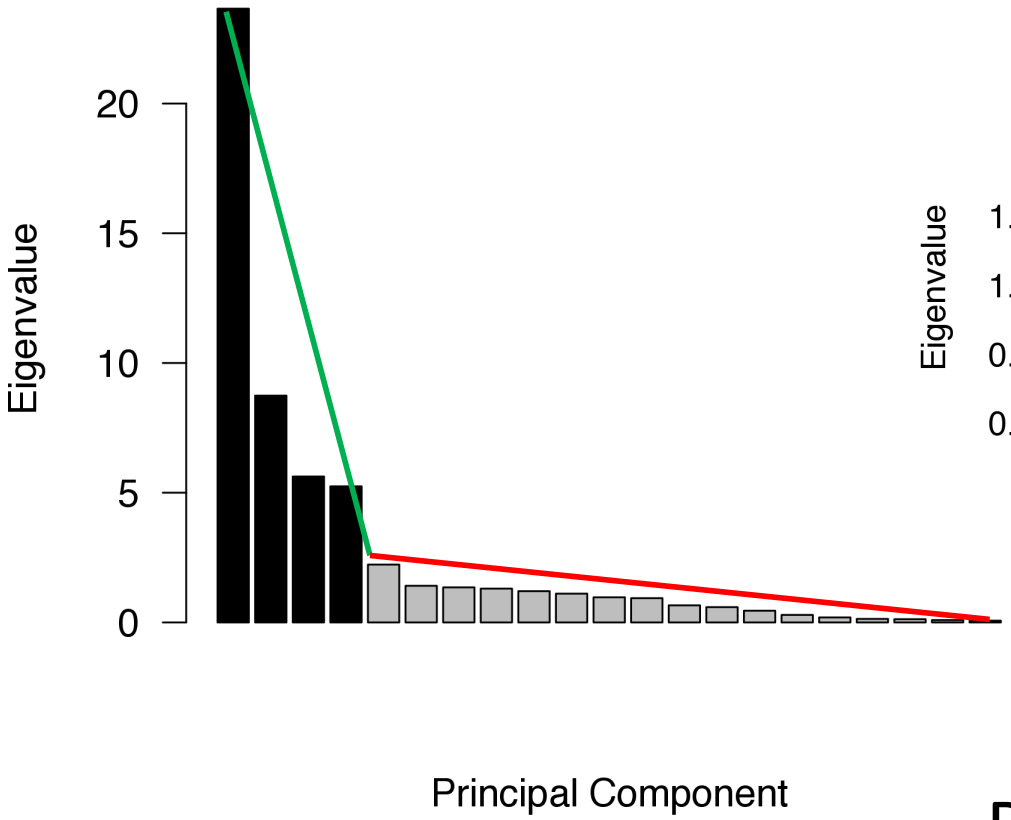
5 steps for multivariate analysis

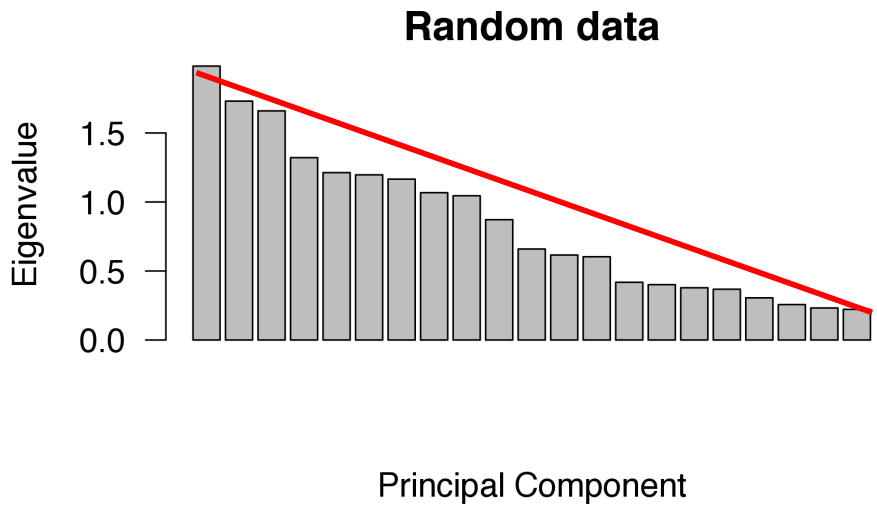
2D - Matrix
Factor analysis, Principal Component Analysis (PCA), Correspondence analysis (CA), Non negative Matrix Factorization (NMF), Singular Value Decomposition (SVD), Empirical Orthogonal Function (EOF), …

3D – Tensor, Array
Tensor Decomposition, Multiway multidimensional analysis, Tucker decomposition, Parafac/CANDECOMP, 3DPCA, High- order SVD (HOSVD), Principal Tensor Analysis,

1. Select the variables and check the distribution (log transform if too skewed)
2. Scale the variables
3. Run the analysis and select the number of “components” with a scree test
4. Interpret the selected components
5. (optional) cluster the individuals based on their projection on the components

Principal Tensor Analysis



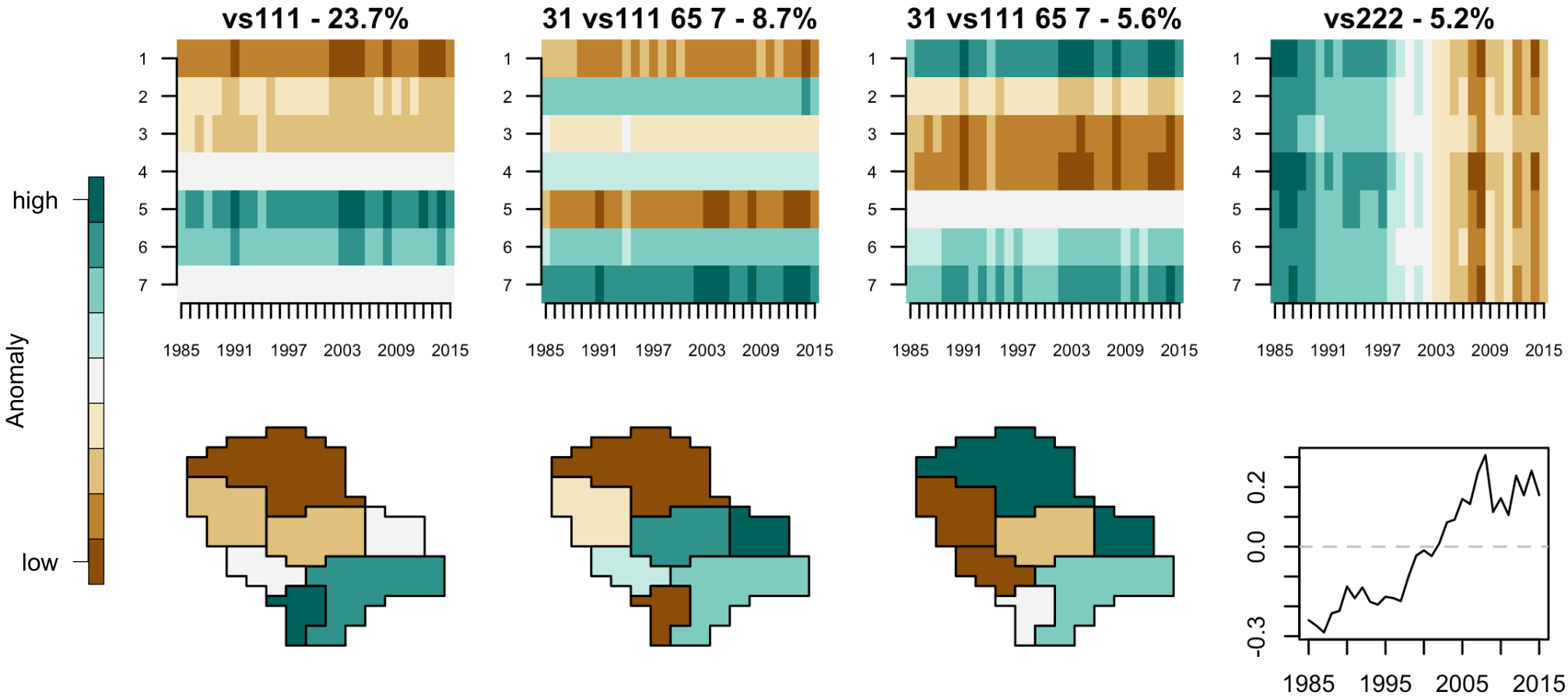


PTA-k found 4 significant “PCs”, explaining together 43% of the variability in the data set.

Principal Tensor Analysis

Corr. Chl a, SBS, depth

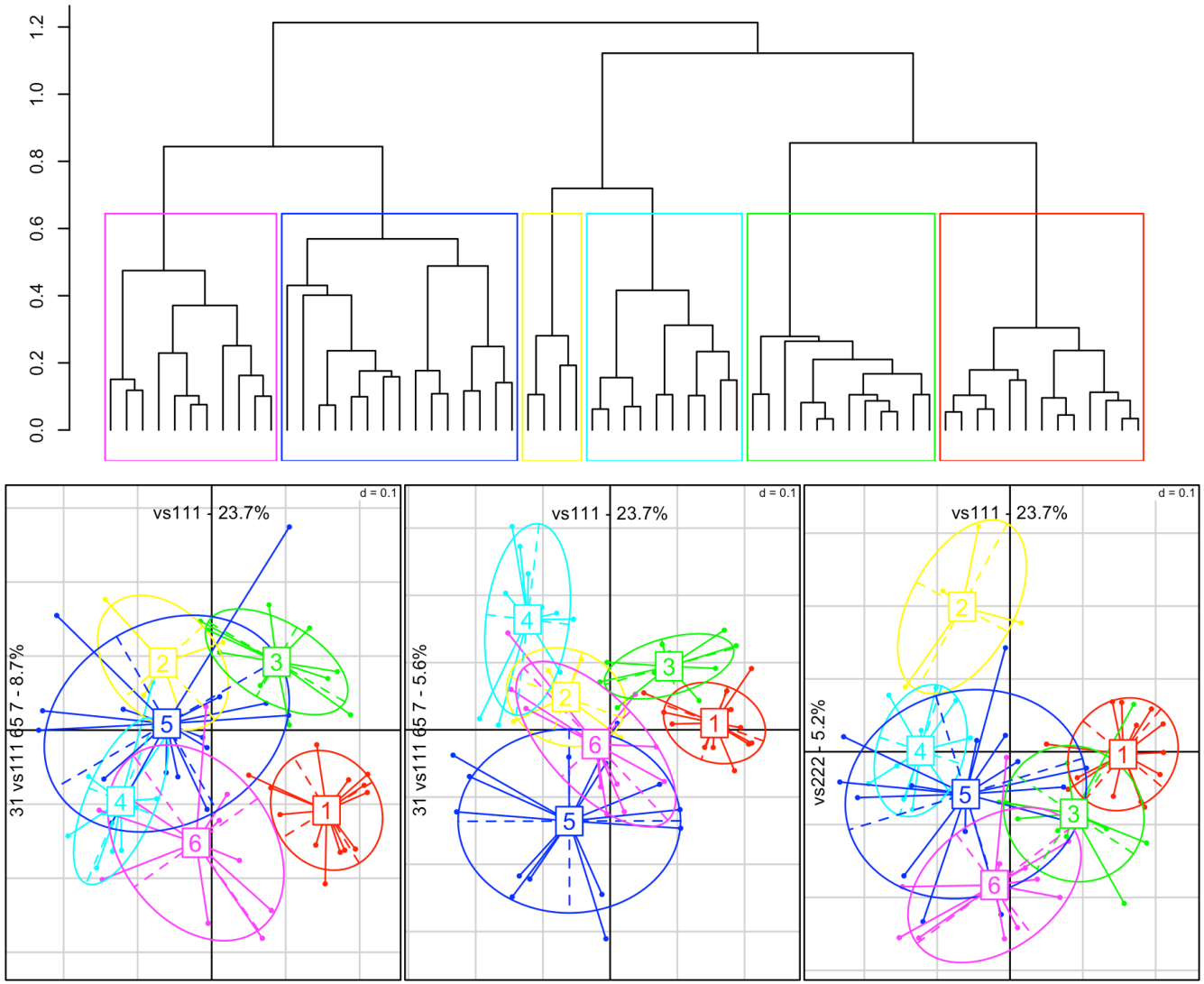
Corr. with AMO



Spatial pattern (38%, 3PCs) > Temporal pattern (5%, 1PC)

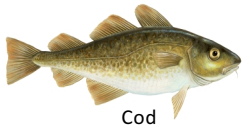
Clustering

Identifying sub-communities with similar dynamics

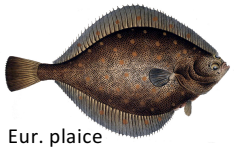


Clustering

Lesser weever

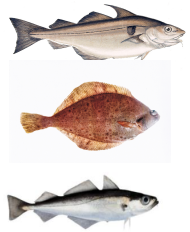


Cod

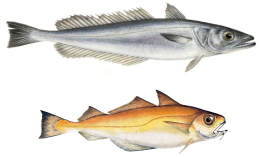


Eur. plaice

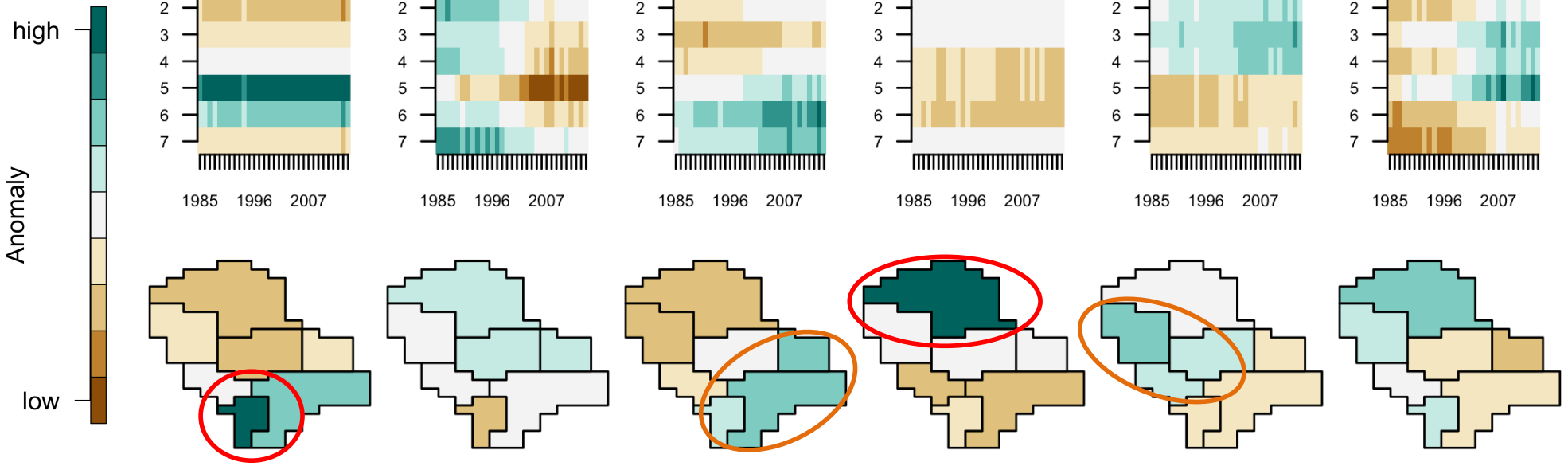
Saithe (Pollock)

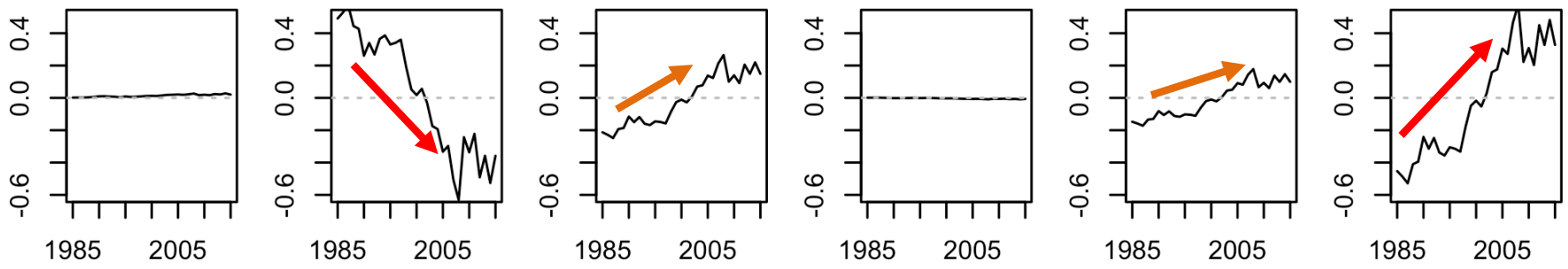


Norway pout
Whiting, Dab
Haddock



Hake, Poor cod





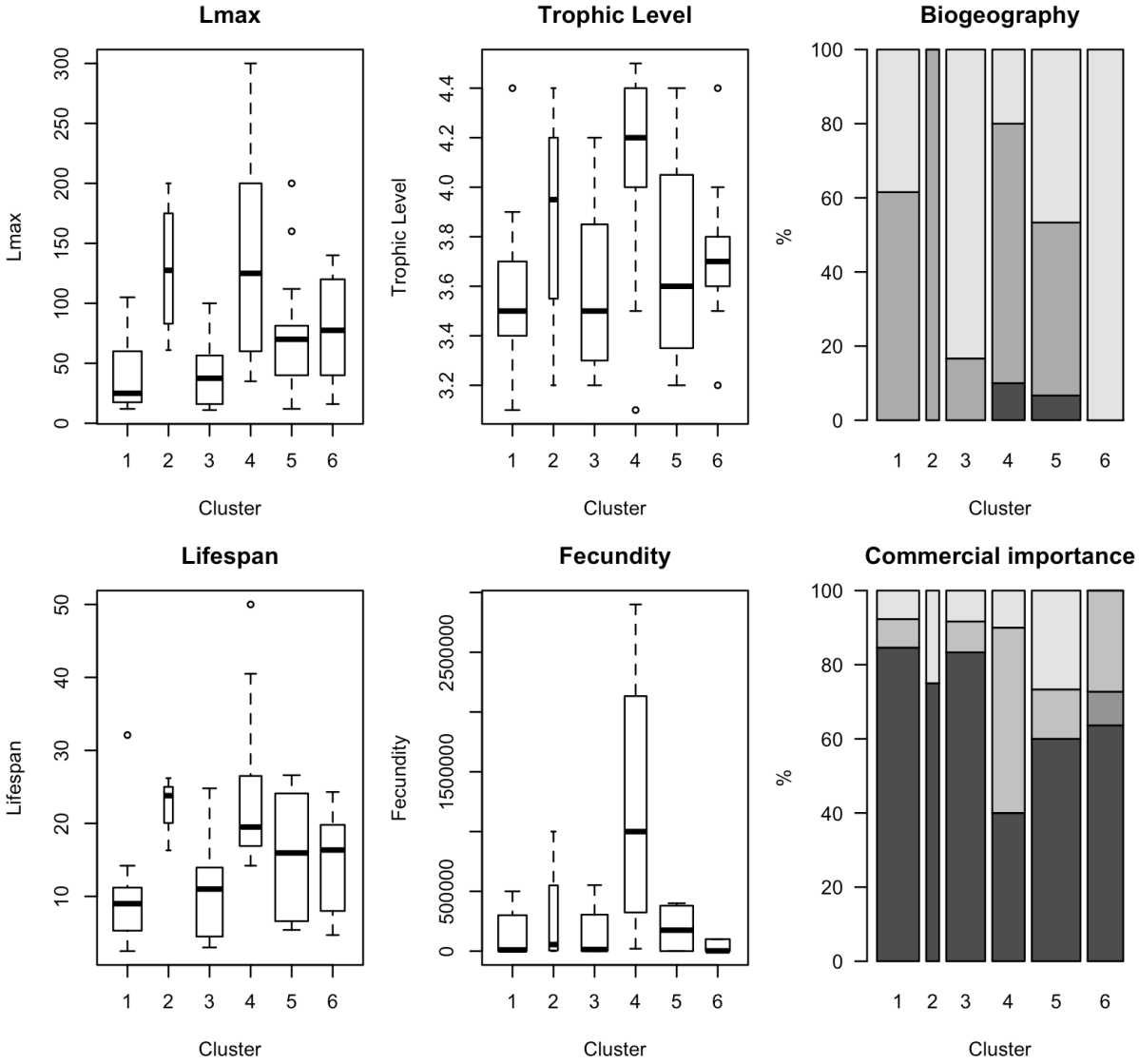
CHL **
SBT *

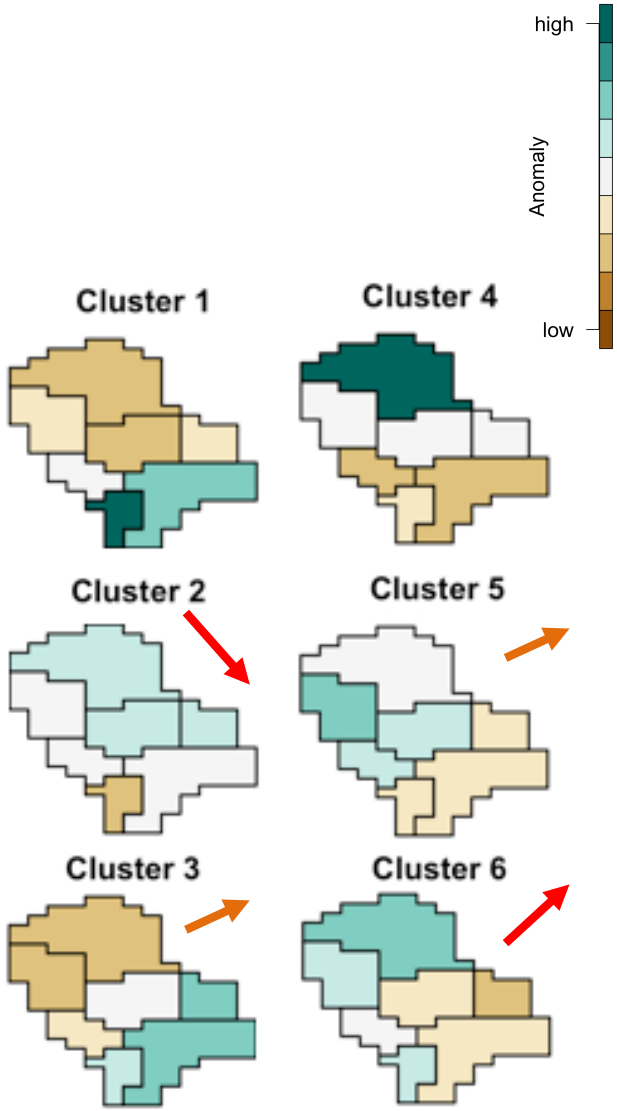
SSS *

SBS *

CHL *
SBT *

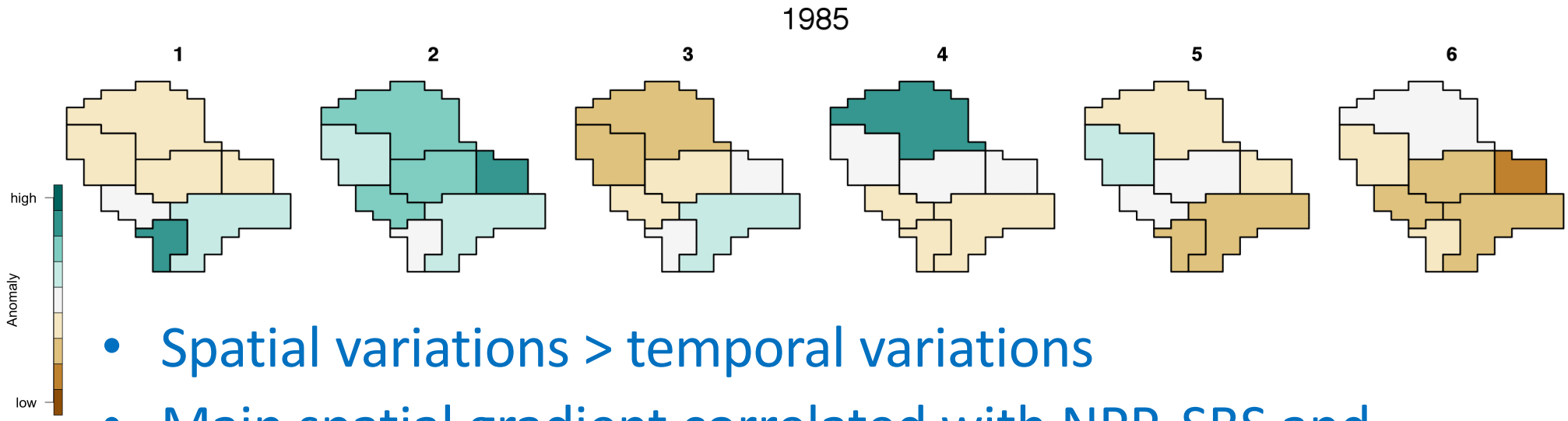
Biological characteristics





Clusters based on spatiotemporal dynamics show differences in traits

Conclusion



- Spatial variations > temporal variations
- Main spatial gradient correlated with NPP, SBS and depth and temporal gradient correlated with AMO
- Six sub communities of demersal fishes, influenced by different drivers, and showing differences in traits.

Thank you for your attention

romain.frelat@uni-hamburg.de

Universität Hamburg