ITS proposal 2015

Scientific analysis	Long-term goals	Near-term science objectives (2015)
Data synthesis	Identify key parameters and condense the time series information	(1) Separately synthesize surface, midwater and benthic time series into a manageable number of representative data sets
Synchronous variability	Identify cause/effect relationships between the time series	(2) Compare the time series from (1) with each other and with climate indices to identify similar variability at the seasonal and interannual scale
Prediction	Predict key parameters in the future, a few years ahead and under climate change	(3) Identify lagged-relationships based on (2) and identify species for which those are robust enough to enable prediction
State of the ecosystem	Define and share proxies related to ecosystem health	(4) Investigate calculations of biodiversity indices
Outside datasets	Enlarge the above studies to the California Current	*- none for 2015 -*
Carbon budget	Quantify carbon fluxes from surface to bottom and harvest	(5) Start combining midwater animal counts with biomass and respiration numbers

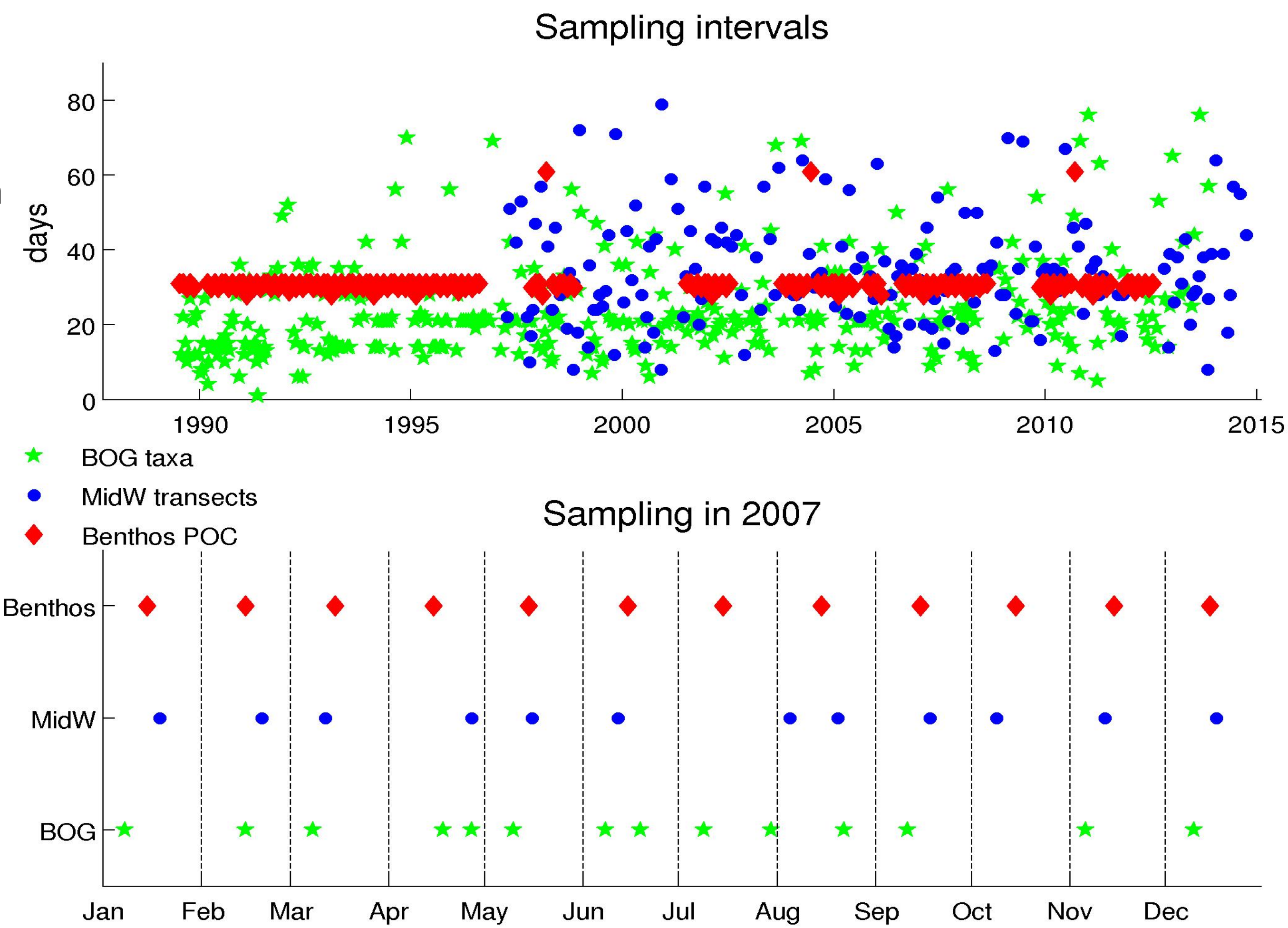
→ 1st objective is to generate comparable **“ecosystem indices”** for the 3 time series that capture **key ecological patterns** – can be:
counts/biomass of key species, biodiversity indices, indices of common variability (PCA), productivity (surface), carbon flux, vertical distribution (midwater), etc

Methods to synthesize ITS

- Data harmonization *(generating monthly time series of various parameters)*
  - temporal coverage
  - treatment of missing data
  - vertical averages (MidW)
- Data preprocessing *(generating standardized time series)*
  - log transform
  - remove mean / seasonal cycle / trend
  - smoothing
  - species relative weights
- Data synthesis *(generating a manageable number of ecosystem indices)*
  - principal component analysis
  - biodiversity indices
  - clustering

Temporal coverage & missing data

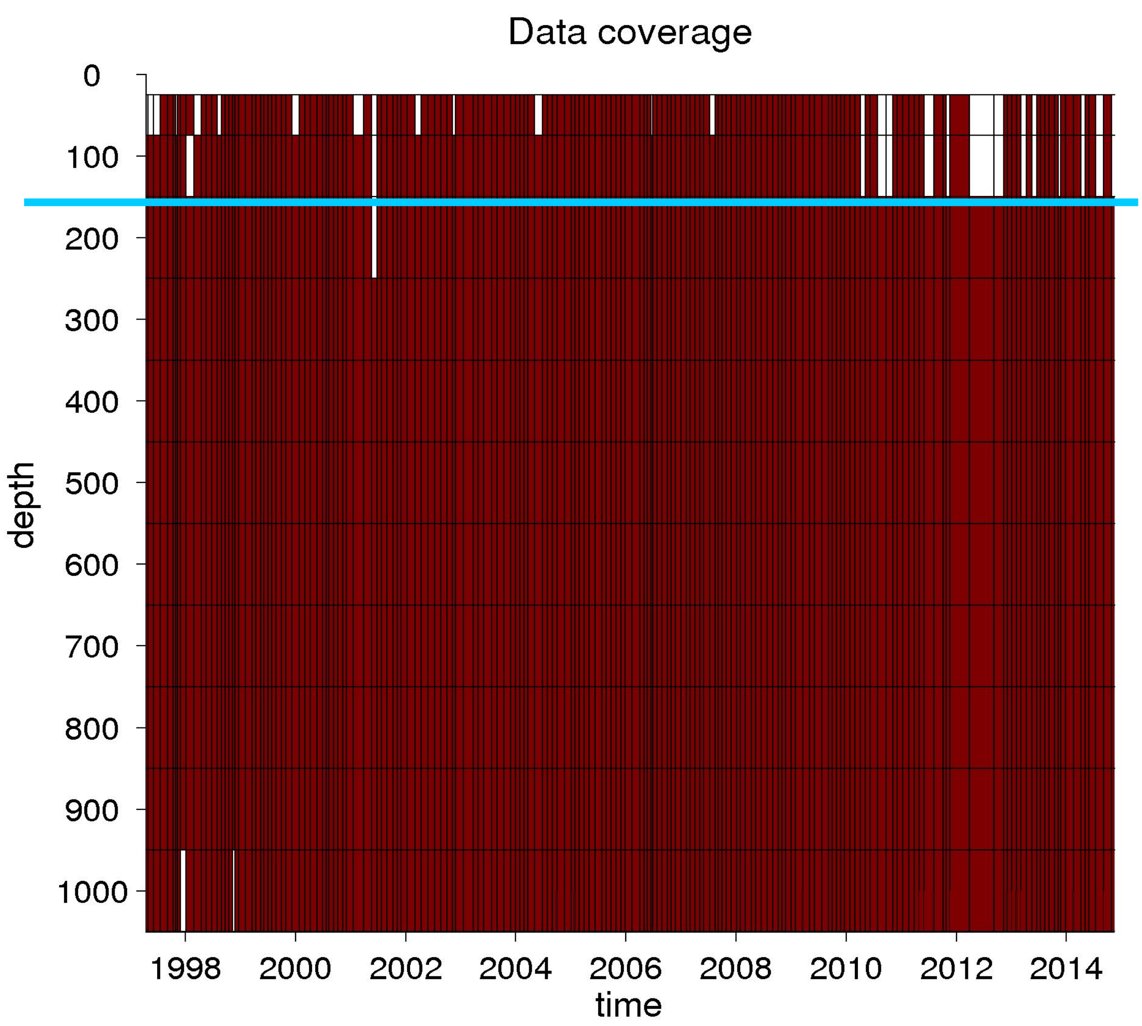
Time intervals between 2 data points



→ average data sets on monthly intervals (binning to conserve most of original data)

→ *linear interpolation on missing months if gaps less than xx months?*

Vertically integrated counts (MidW)



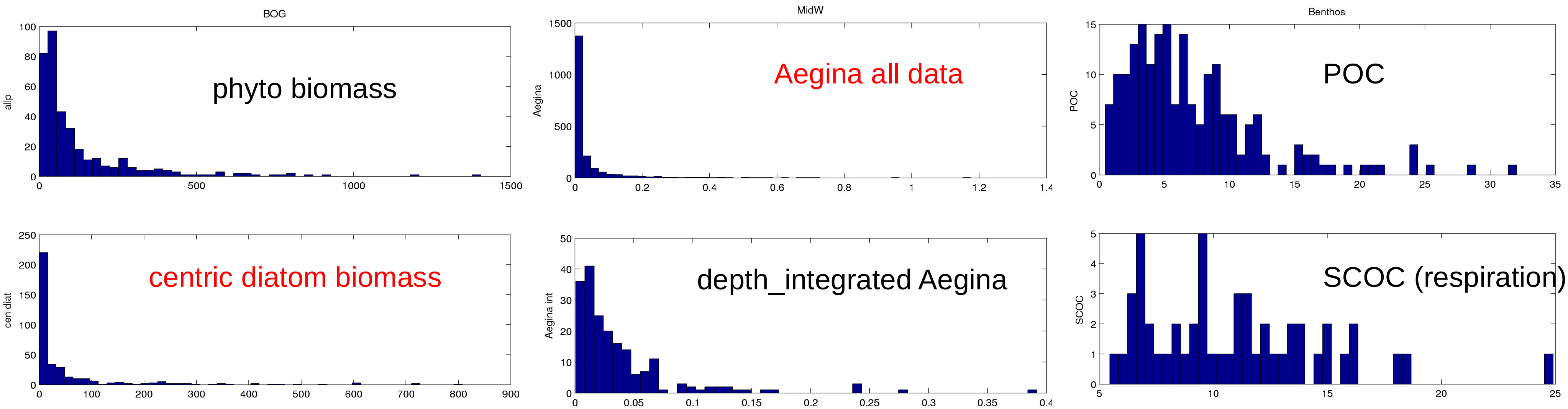
→ average data sets on monthly intervals
→ only consider depths 200-1000
→ temporal linear interpolation to fill shallow/deep missing depths
→ vertical average

Much better than previously because of combining / removing ROV dives

- *Data preprocessing*

Data distribution (histograms)

(original data) BOG MidW Benthos

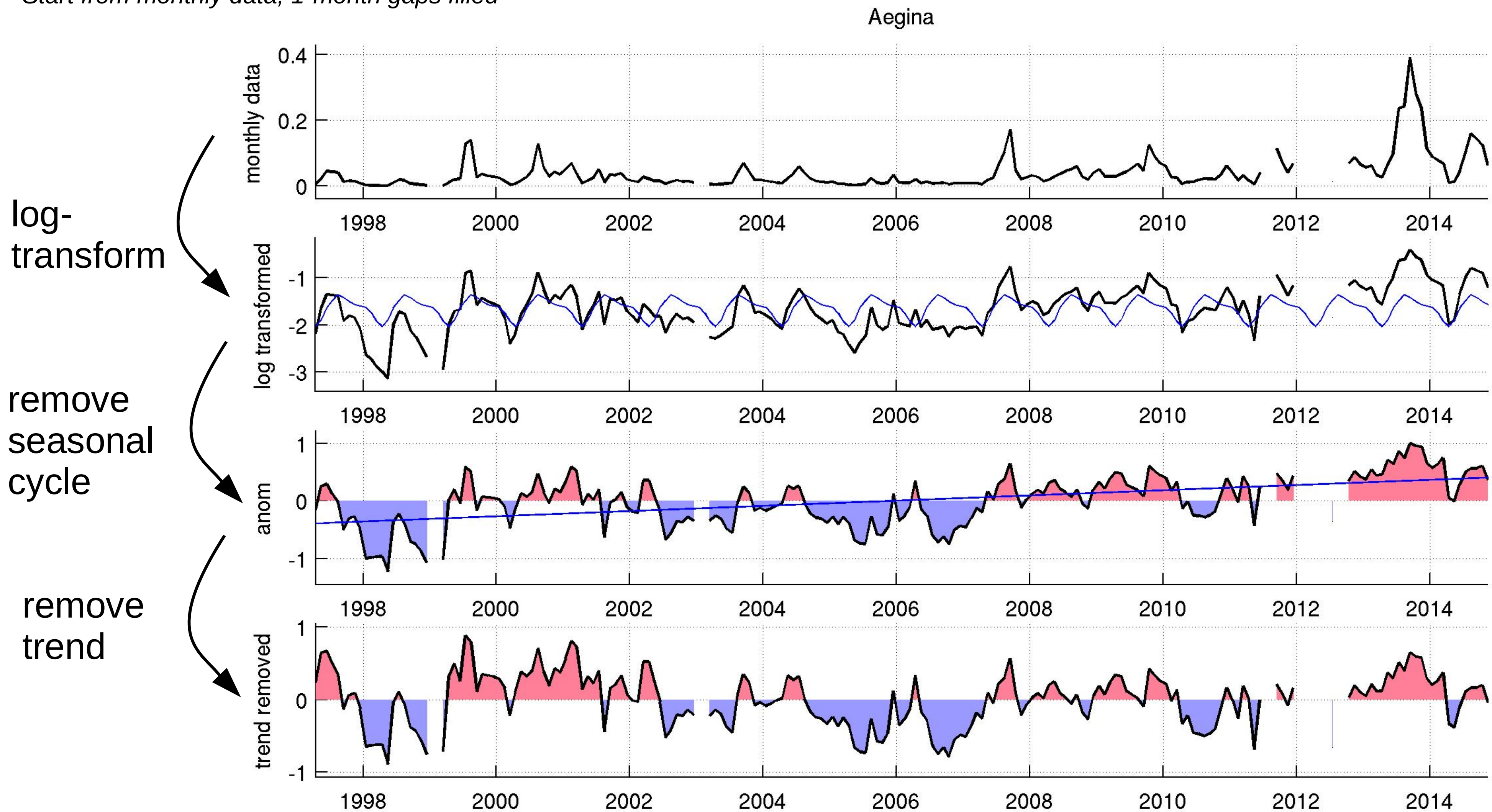


How to deal with blooms?
→ log-transforms is a possibility.

- *Data preprocessing*

Generating anomalies

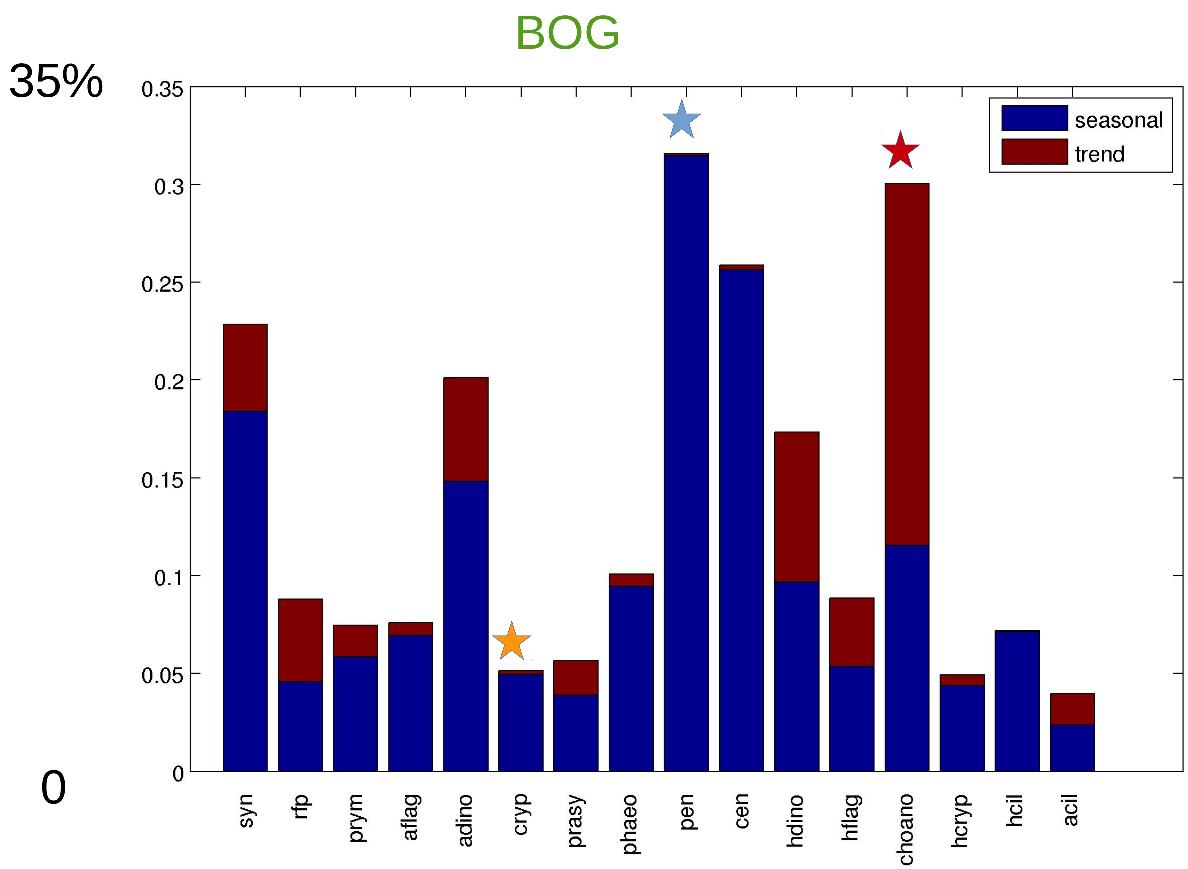
Start from monthly data, 1-month gaps filled



- *Data preprocessing*

Seasonal cycle & trend

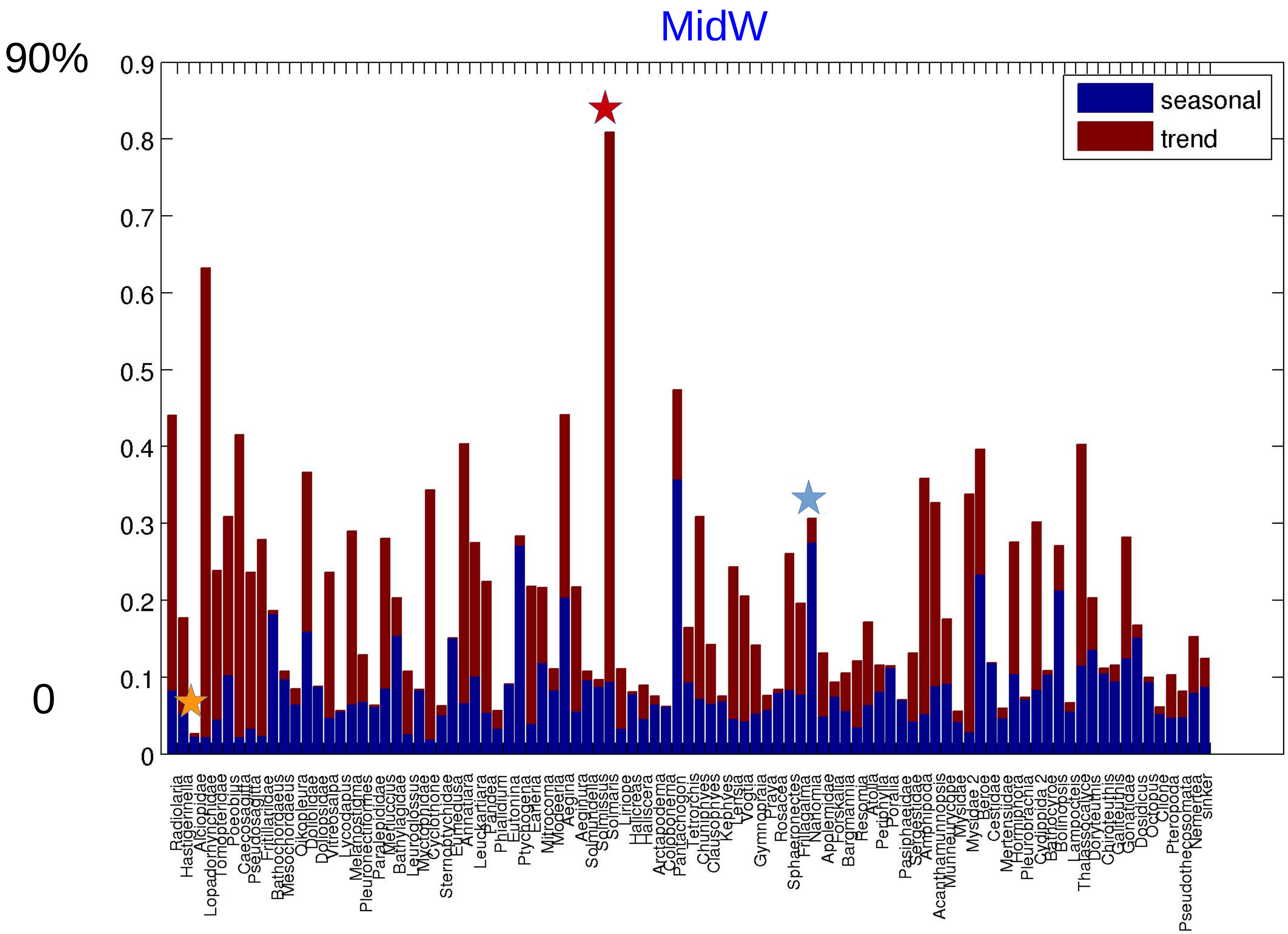
(regrid monthly & log-transform)



Percentage of variance explained by:
seasonal cycle
trend

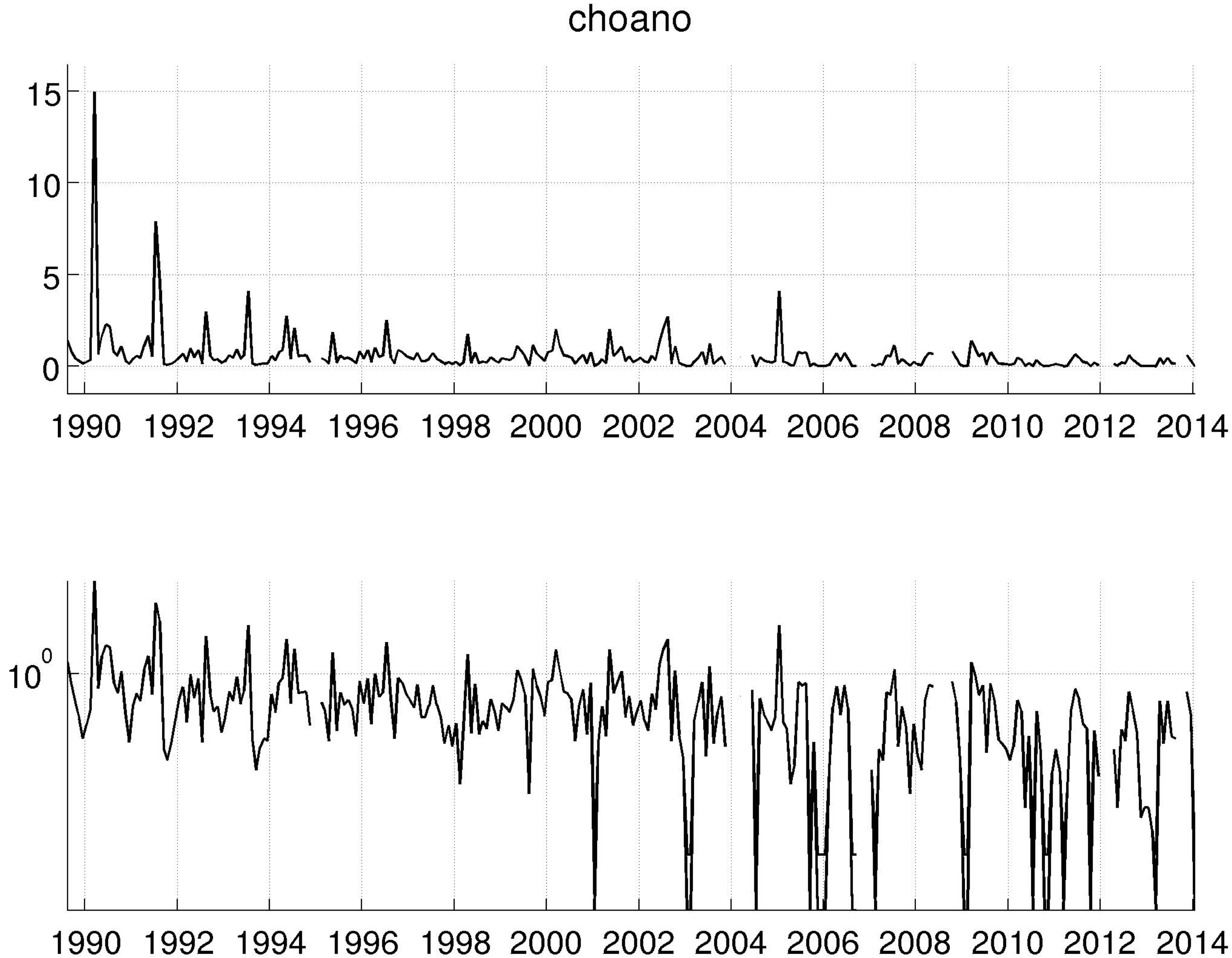
- *Data preprocessing*

Seasonal cycle & trend



!! trend from 1997 (MidW) instead of 1989 (BOG) !!

Issues with log-transform

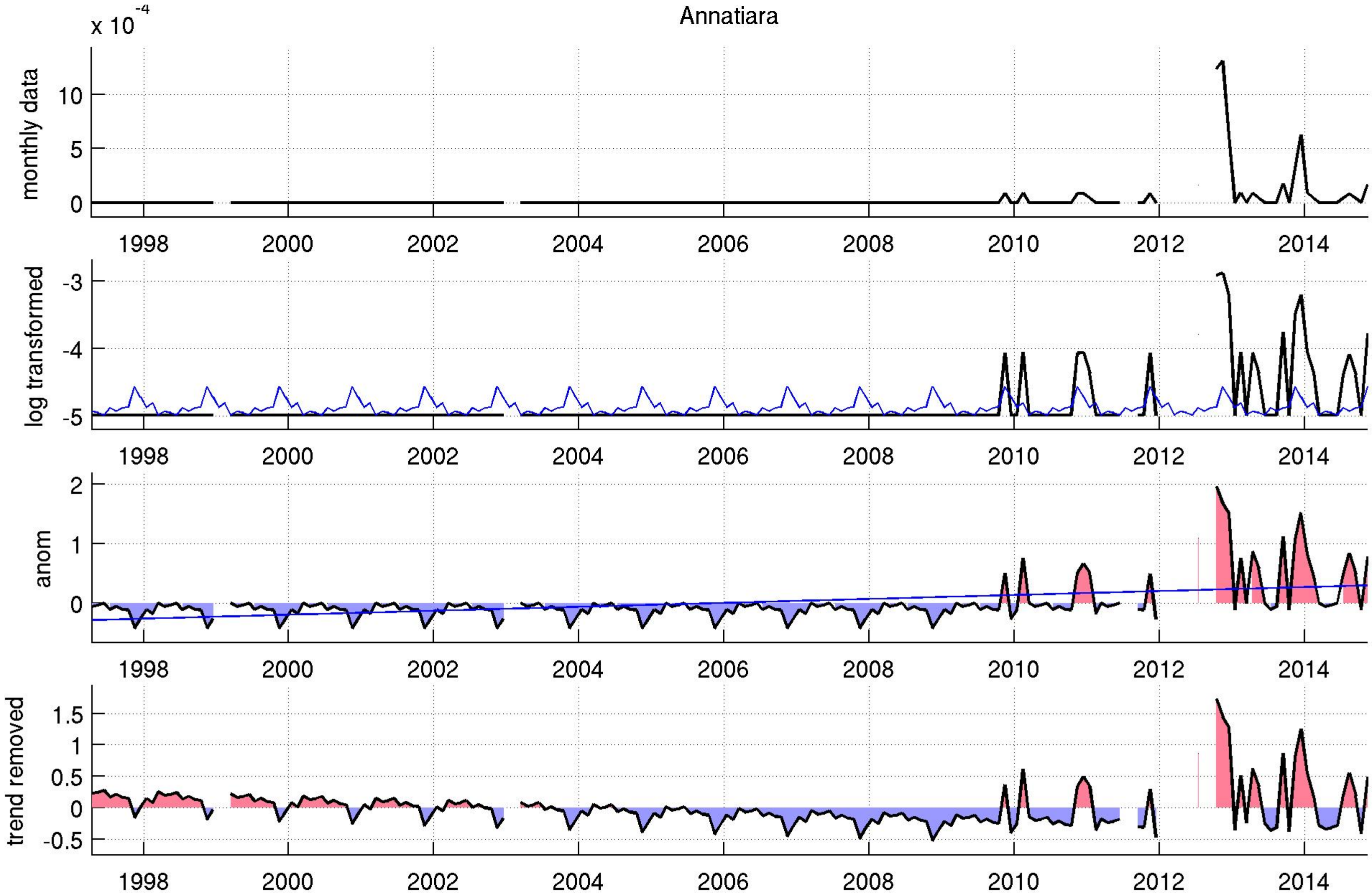


strongly weighs towards low-values variability, which may not mean much.

+ how to treat zeros? For now minimum value /2.

• *Data preprocessing*

Issues with removing the seasonal cycle & trend



→ not removing the trend (time series are too short) – what about the seasonal cycle?

Other preprocessing questions

- Smoothing?
  - removes noise... or information?
  - method? FFT, window smoothing, 3-points smoothing, else?
  - when?

- How to weigh the time series??
  - required for statistical analysis such as PCA
  - for now, normalization by standard deviation
  - else? e.g. species more ecologically important than others? use the trophic level? the classification level?

Ecosystem indices

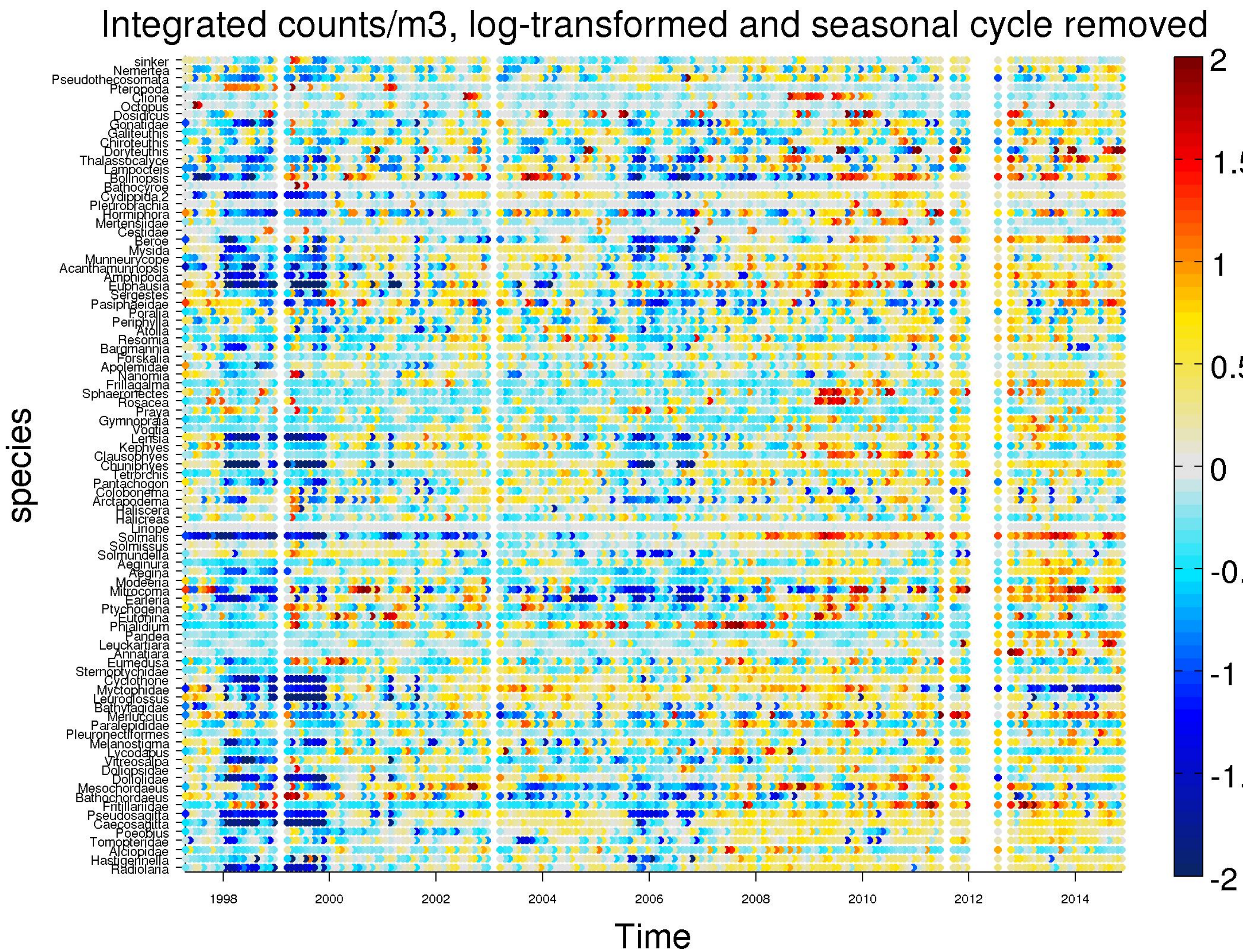
- Goal: *“Separately synthesize surface, midwater and benthic time series into a manageable number of representative data sets”* (ITS proposal)

- Methods:
  - Principal component analysis
  - Biodiversity indices
  - Key species?
  - Clustering?

• *Data analysis*

Principal component analysis

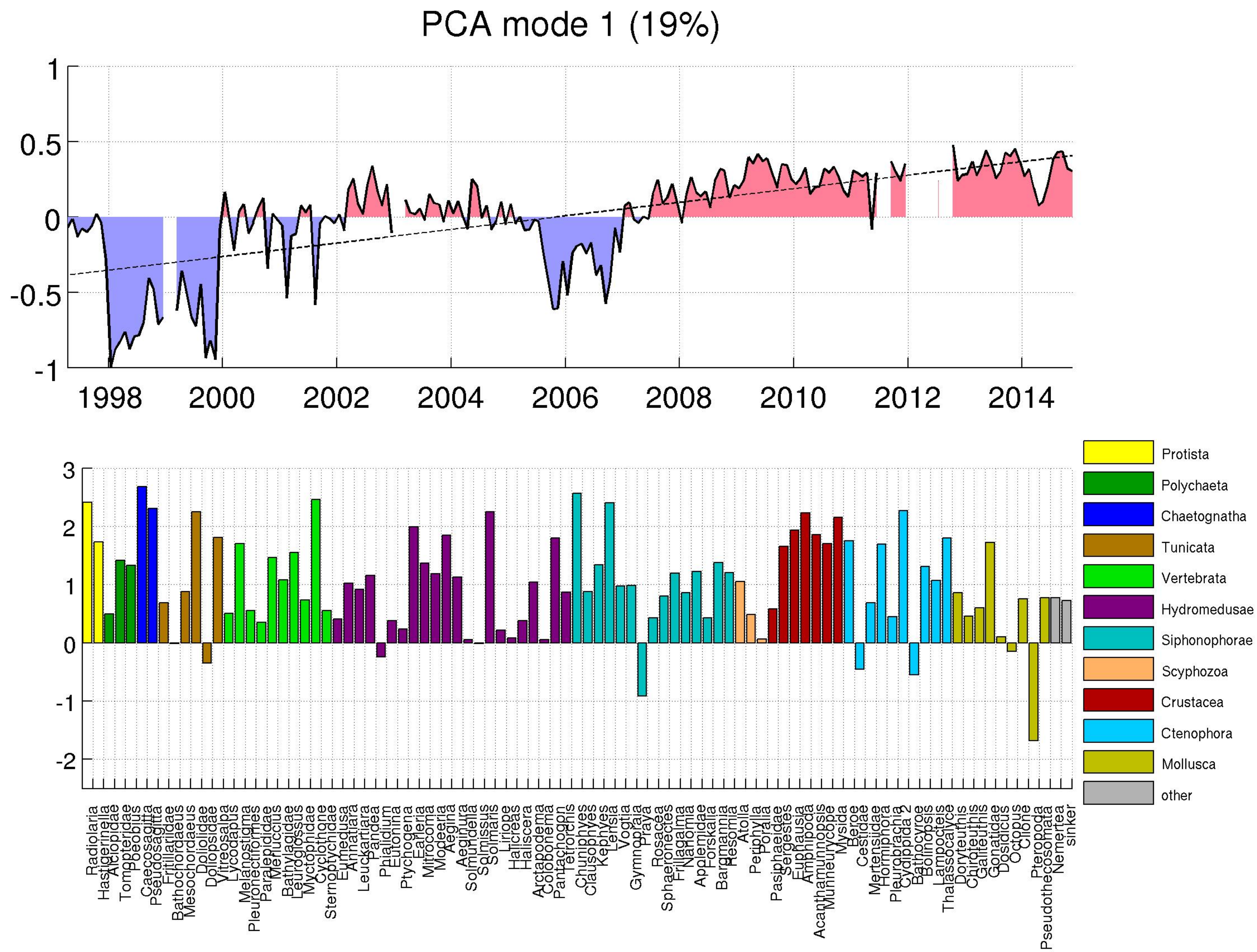
Start point = monthly anomalies (depth integrated)



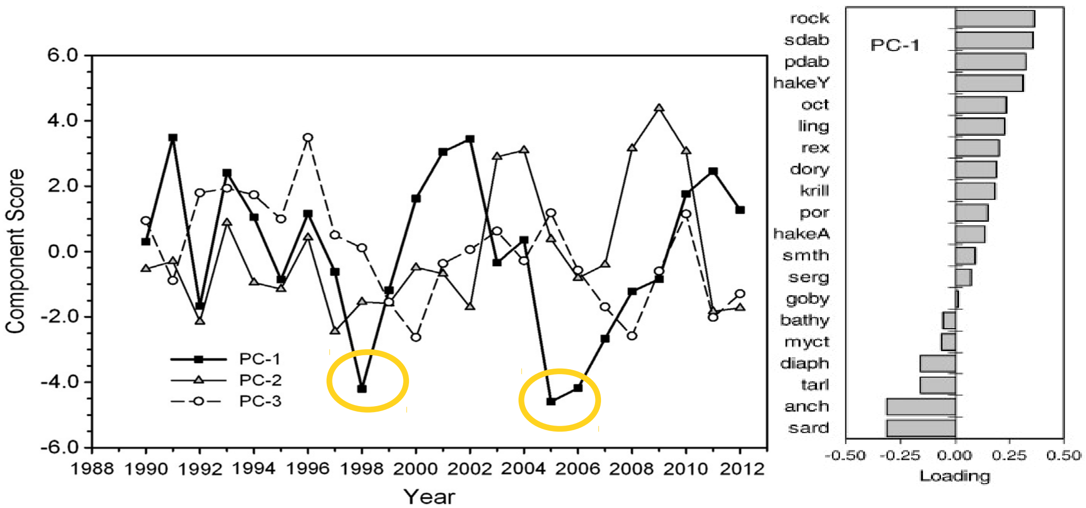
- *Data analysis*

Principal component analysis

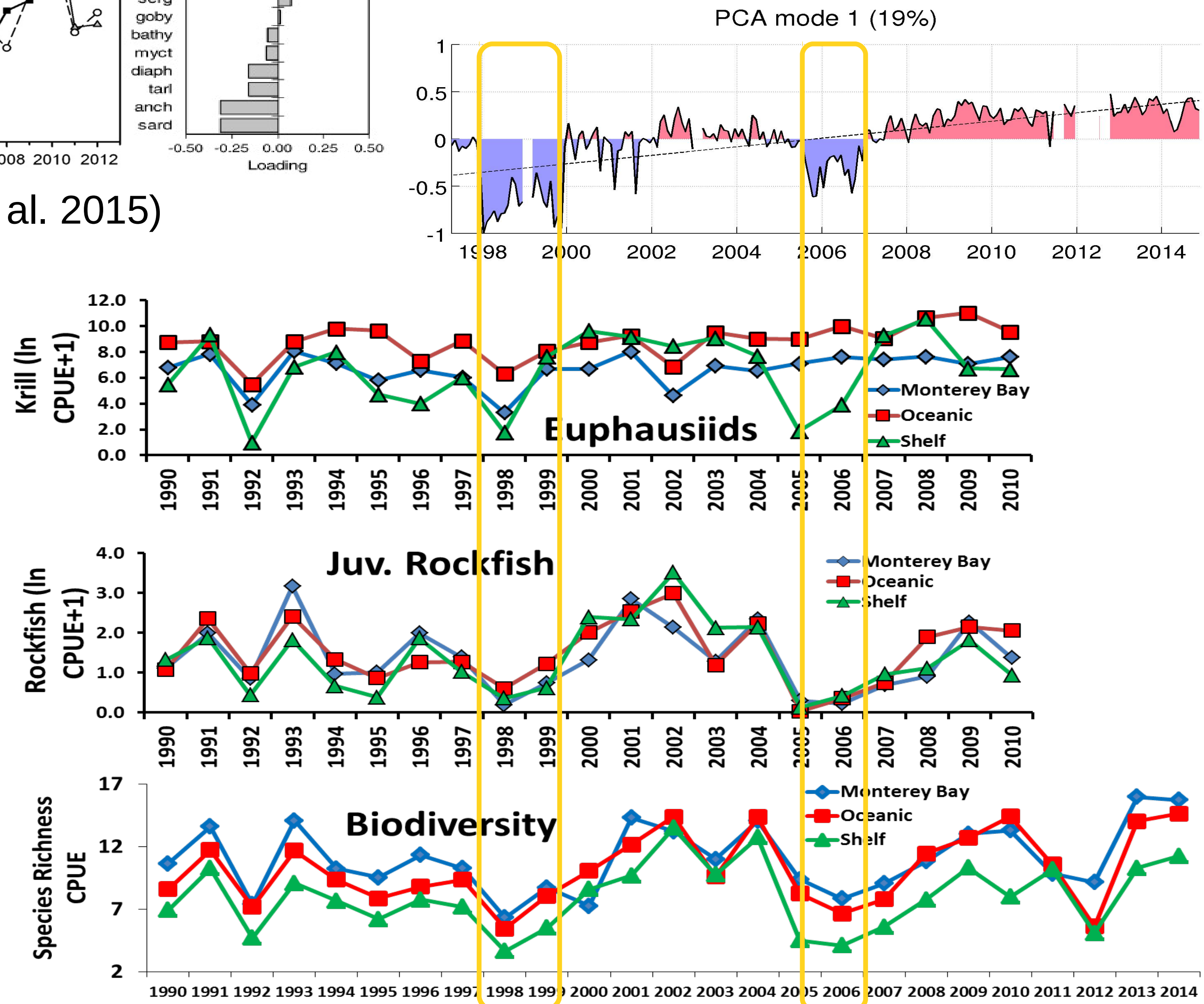
→ extracts common variations between time series
→ sensitive to normalization / weighting



- *Data analysis*



PCA on 20 taxa (Ralston et al. 2015)



Rockfish Recruitment and Ecosystem Assessment Survey (NOAA)

- *Data analysis*

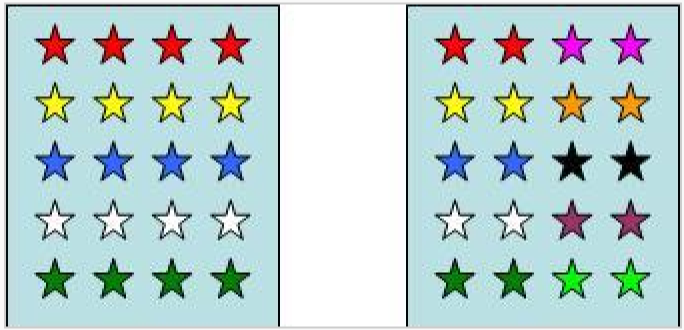
Biodiversity indices

1. Species richness indices: Species richness is a measure for the total number of the species in a community. However, complete inventories of all species present at a certain location, is an almost unattainable goal in practical applications.

e.g., Margalef's diversity index: $\frac{S-1}{\ln(N)}$

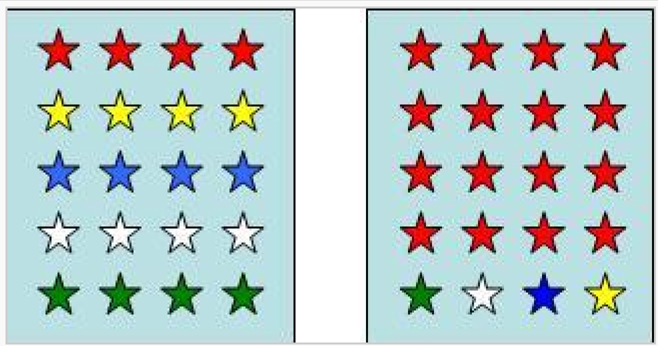
Menhinick's diversity index: $\frac{S}{\sqrt{N}}$

S=number of species,
N=total number of individuals
p_i= proportion of individuals found in the ith species



A visualization of the species richness: with respectively 5 and 10 species.

2. Evenness indices: Evenness expresses how evenly the individuals in a community are distributed among the different species.



A visualization of the evenness of 5 species.

Combined: heterogeneity measures

e.g., Shannon-Wiener diversity index $-\sum p_i * \ln(p_i)$

Simpson 's index $\sum p_i^2$

3. Taxonomic indices: These indices take into account the taxonomic relation between different organisms in a community. Taxonomic diversity, for example, reflects the average taxonomic distance between any two organisms, chosen at random from a sample. The distance can be seen as the length of the path connecting these two organisms along the branches of a phylogenetic tree.

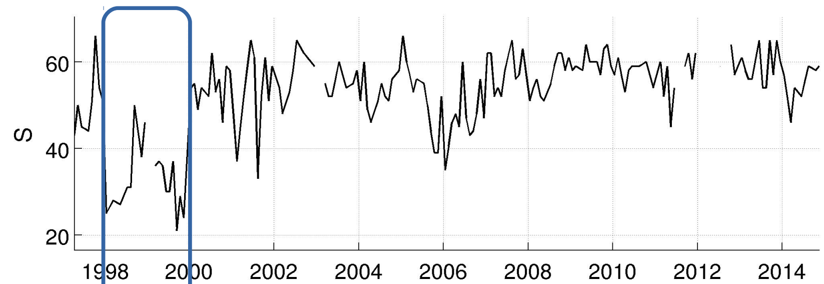
Source: http://www.coastalwiki.org/wiki/measurements_of_biodiversity

- *Data analysis*

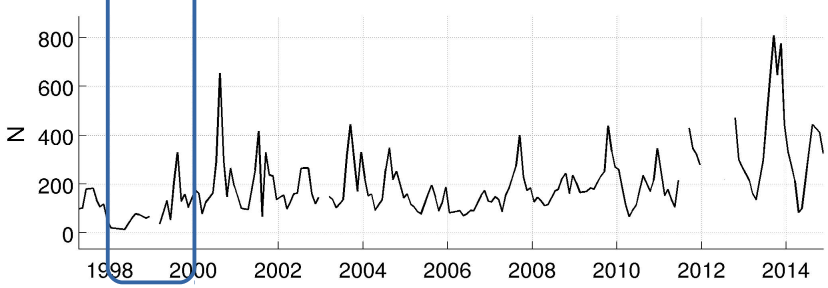
Species richness (MidW)

Biodiversity computed on integrated counts (not per volume, and no interpolation)

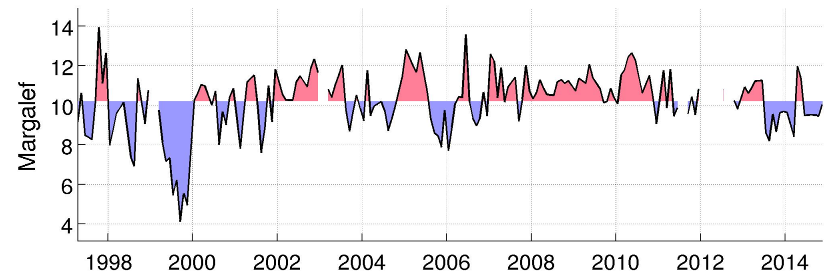
number of species



number of individuals

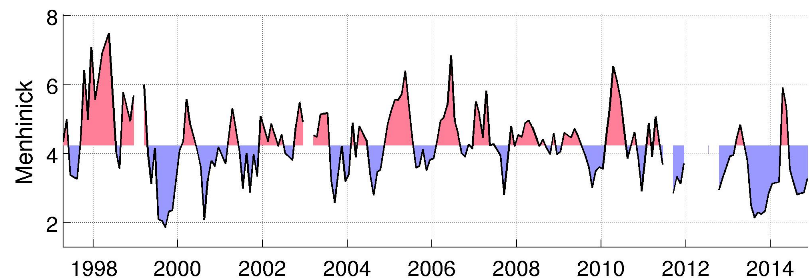


Margalef's diversity index



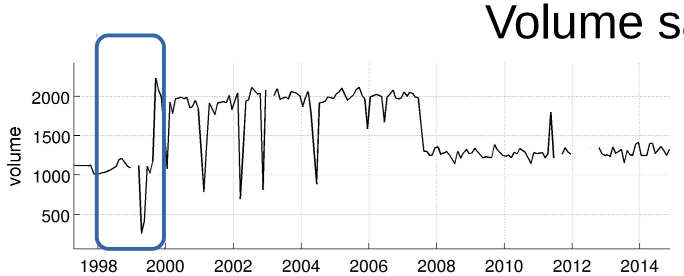
$$\frac{S-1}{\ln(N)}$$

Menhinick's diversity index



$$\frac{S}{\sqrt{N}}$$

Volume sampled



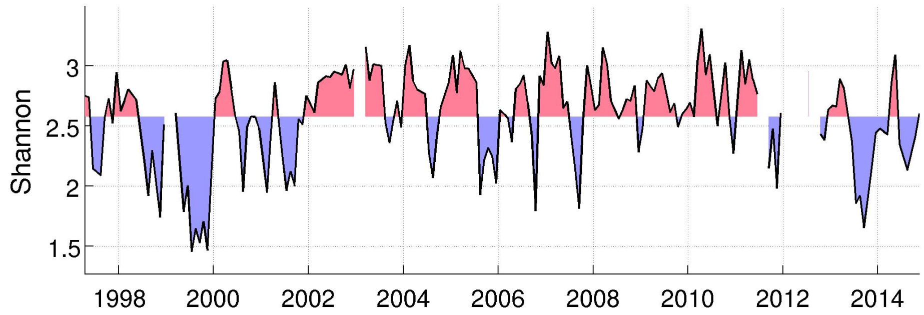
- *Data analysis*

Heterogeneity measure

p_i= proportion of individuals found in the ith species

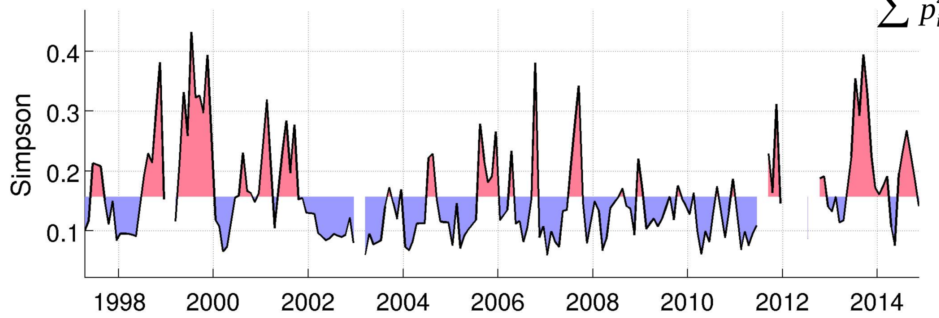
Shannon-Wiener’s diversity index

$$-\sum p_i * \ln(p_i)$$



Simpson's index

$$\sum p_i^2$$



Simpson = *probability that any two individuals drawn at random from an infinitely large community belong to the same species*
i.e. high = low biodiversity

- *Data analysis*

Depth & biodiversity