International Symposium on
Climate Change Effects on Fish and Fisheries

On the utility of self-organizing maps (SOM) and k-means clustering to characterize and compare low frequency spatial and temporal climate impacts on marine ecosystem productivity

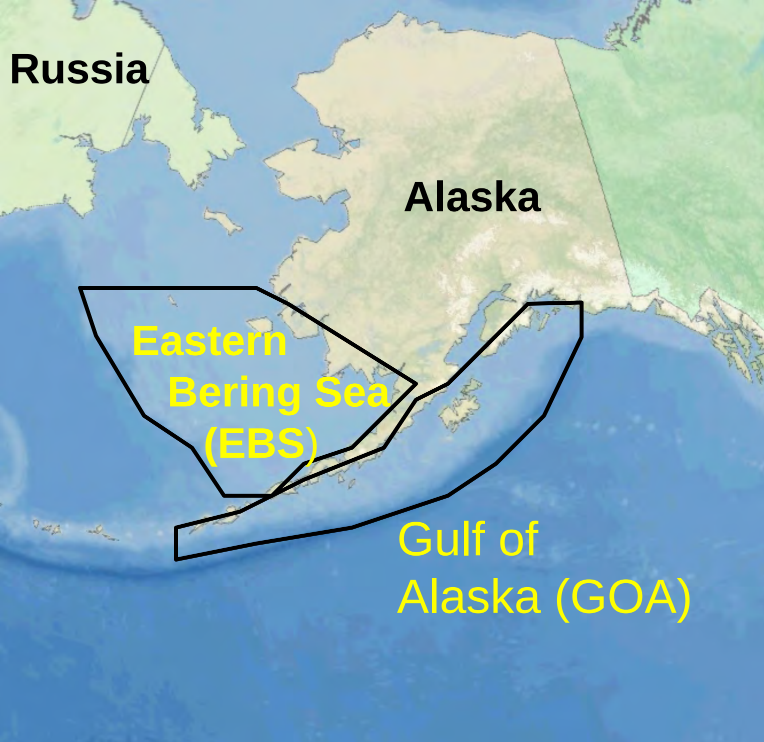
Jae-Bong Lee[1] and Bernard A. Megrey[2]

[1] National Fisheries Research and Development Institute, Busan, Korea

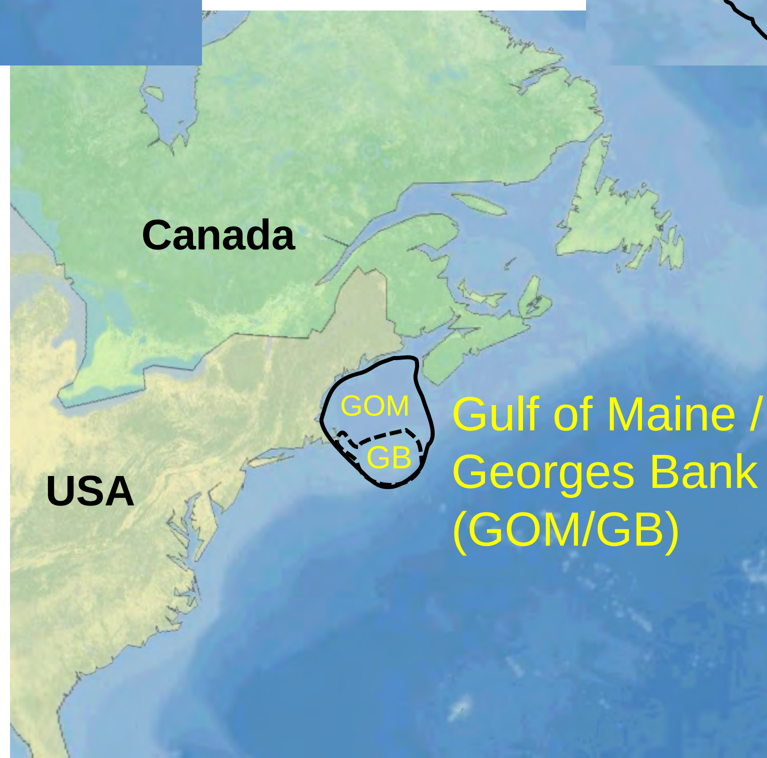
[2] Alaska Fisheries Science Center, Seattle, USA

Outline

- Motivation and rational: temporal and spatial of R and S variability compared
- Describe Self organizing maps (SOM)
- Apply to time series trends in the ln(R/S) response variable
- Discuss the utility of SOM



4 Major Ecosystems

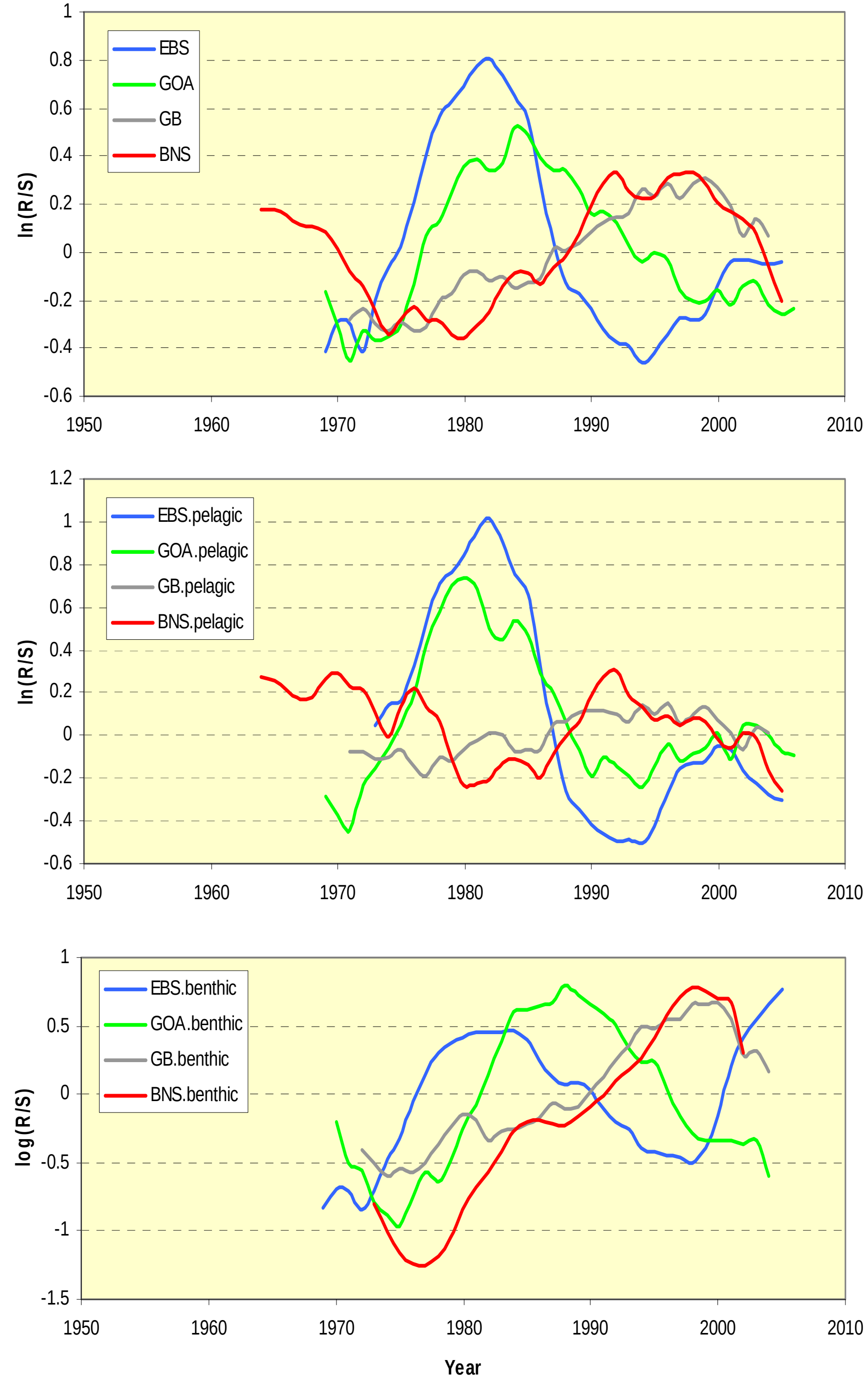


Trends in Recruitment Time Series Anomalies

- Temporal and spatial patterns of R and S variability from 17 stocks were compared among functionally analogous species and similar feeding guilds from four marine ecosystems.
- Calculate the anomaly in ln(R/S),
- Pool by feeding guild (benthic vs. pelagic) or ecosystem (EBS, GOA, GB/GOM, BNS) by calculating the average anomaly per year
- Look for within and cross ecosystem trends

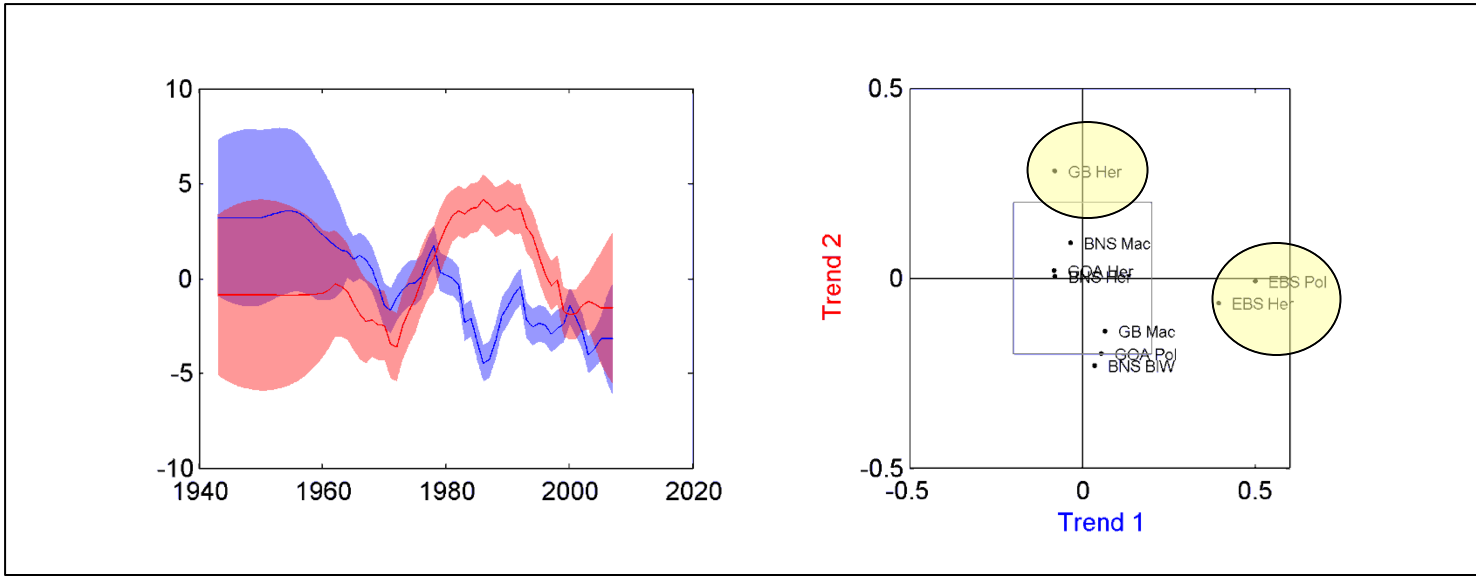
ln(R/S) Response Variable

- EBS and GOA follow same trend (r=0.64, p<0.01)
- GB and BNS follow same trend (r=0.89, p<0.01)
- EBS/GOA out of phase with GB/BNS (r=-0.67, p<0.01)
- Declining survival since mid-90's for all but EBS. Declining pelagic survival for GB and BNS.
- Improving benthic survival for GB and BNS since late 1970's. Declining benthic survival for EBS and GOA over same period.
- Recent upturn in benthic survival in EBS since late 1990's and in pelagic survival since mid 1990's.

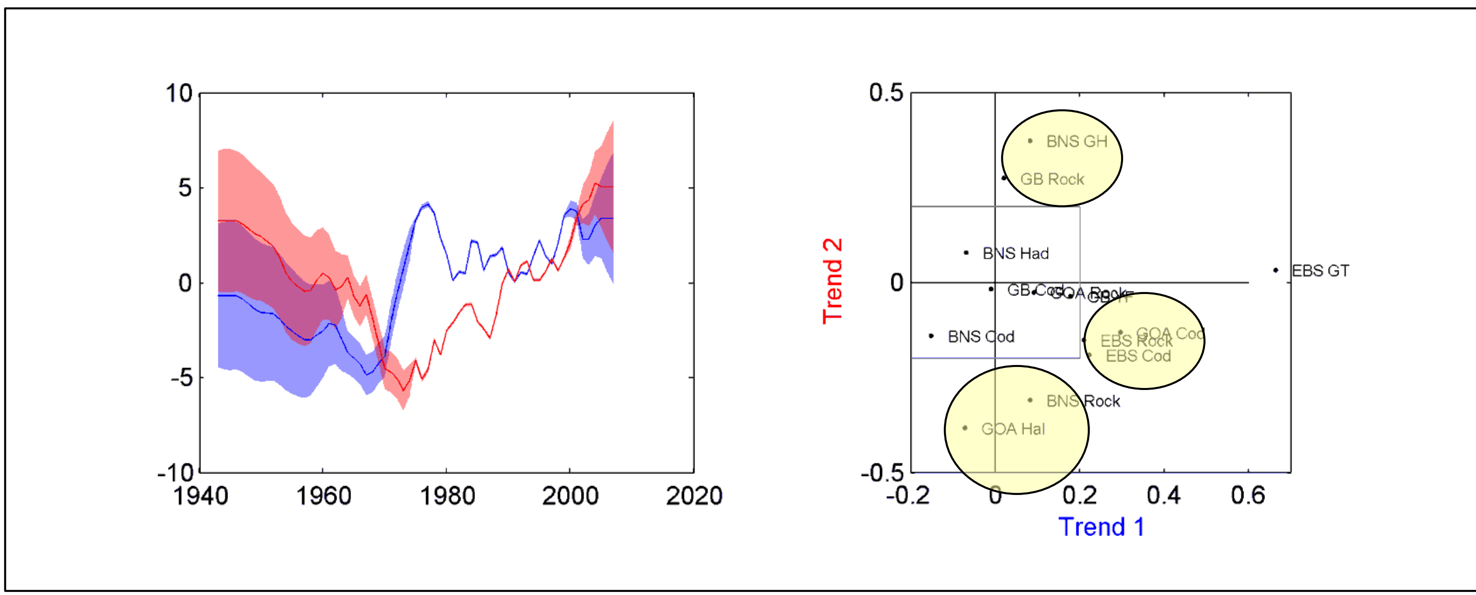


Dynamic Factor Analysis ln(R/S)

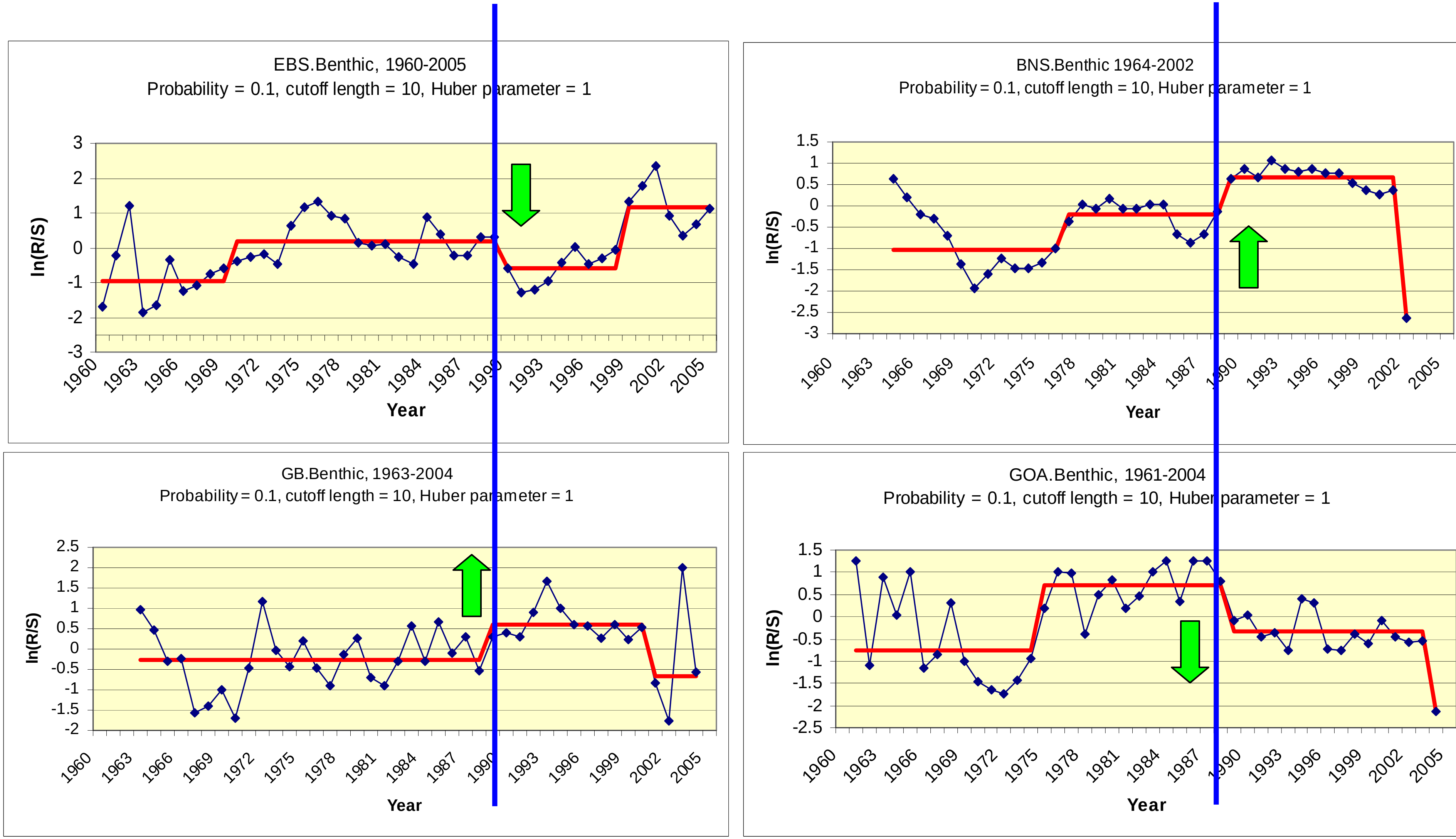
Pelagic Species



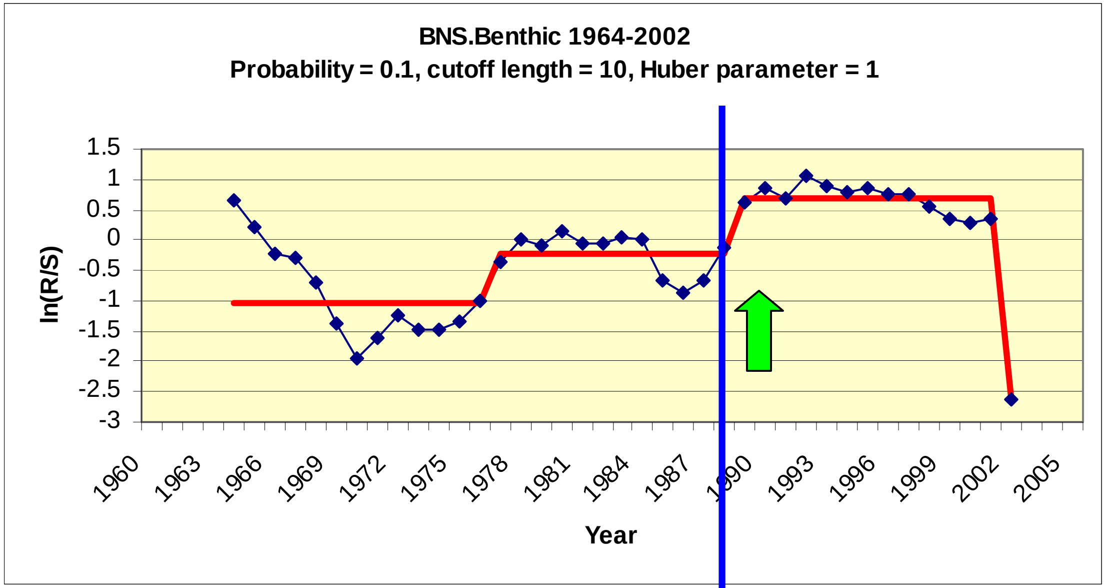
Benthic Species



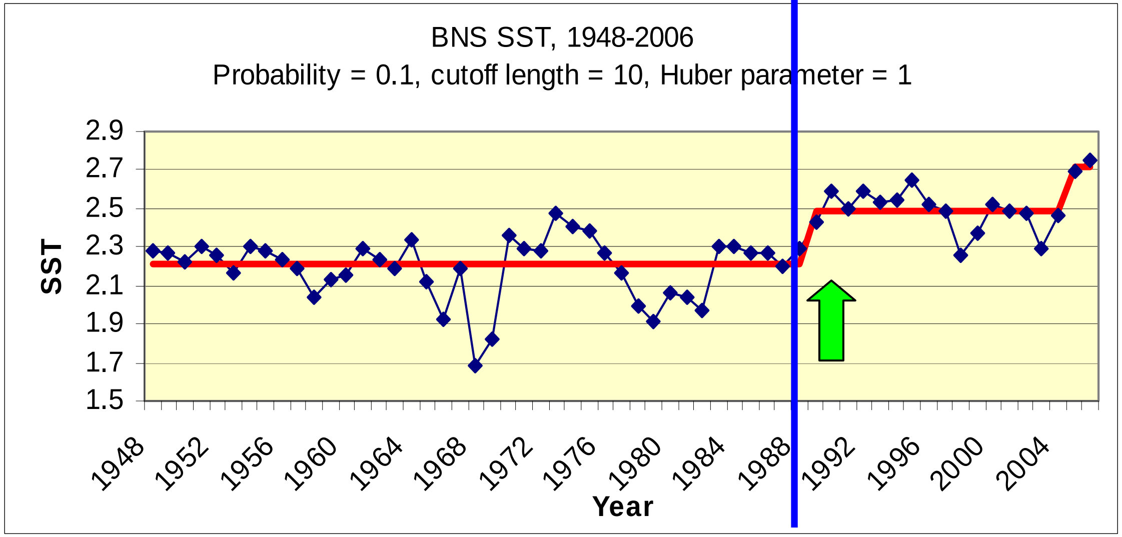
Survival Changes all Four Ecosystems: Benthic Regime Shift ~ 1989?



Climate forcing of the Barents/Norwegian Sea?



BNS Pelagic ln(R/S)



BNS SST

Purpose

- Motivation and rational: temporal and spatial of R and S variability compared
- Describe Self organizing maps (SOM)
- Apply to time series trends in the ln(R/S) response variable
- Discuss the utility of SOM

Methods

- Artificial neural network (ANN)
  -- Self organizing mapping (SOM)

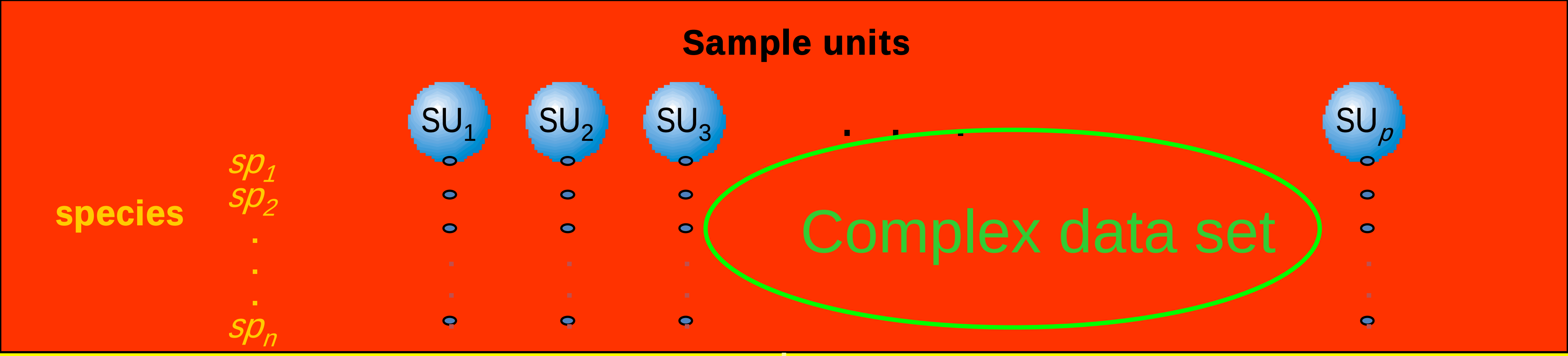
Wikipedia http://en.wikipedia.org/wiki/Self-organizing_map

A **self-organizing map (SOM)** or **self-organizing feature map (SOFM)** is a type of artificial neural network that is trained using unsupervised learning to produce a low-dimensional (typically two-dimensional) map. Self-organizing maps are different from other artificial neural networks in the sense that they use a neighborhood function to preserve the topological properties of the input space.

… This makes SOM useful for visualizing low-dimensional views of high-dimensional data, akin to multidimensional scaling. The model was first described as an artificial neural network by the Finnish professor Teuvo Kohonen, and is sometimes called a **Kohonen map**.

… Therefore, SOM forms a semantic map where similar samples are mapped close together and dissimilar apart. This may be visualized by a U-Matrix (Euclidean distance between weight vectors of neighboring cells) of the SOM.

SOM may be considered a nonlinear generalization of Principal components analysis (PCA) …. SOM has many advantages over the conventional feature extraction methods such as Empirical Orthogonal Functions (EOF) or PCA.



- **Ordination**

  Polar Ordination (PO)

  Principal Component Analysis (PCA)

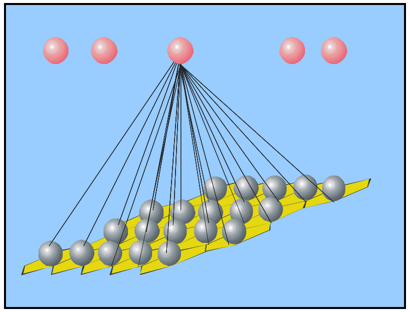
  Correspondence Analysis (CoA)

  Nonlinear Multidimensional Scaling (NMDS)

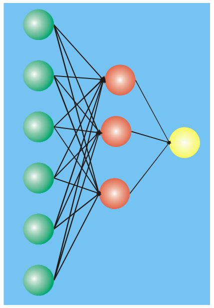
  ...

- **Classification**

- **Artificial Neural Network (ANN)**



SOM

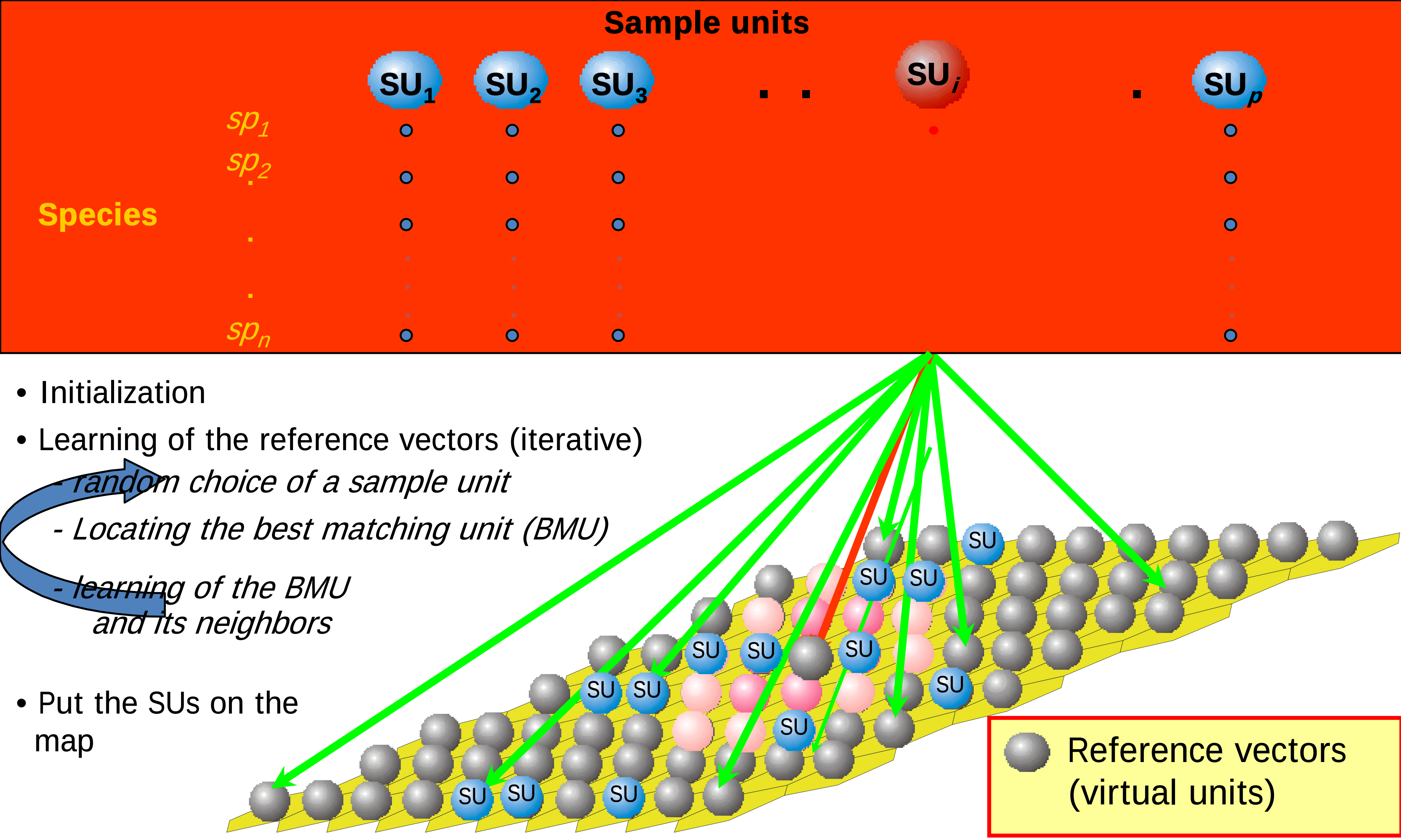


BP

Interpretation

Self Organizing Maps, a good tool to simplify high dimensional dataset.

Bacpropagation neural network, for prediction and discrimination.



In each hexagon, a virtual unit (VU) will be considered. The virtual units are virtual sites with species abundance to be computed. The modifications of the VUs are made through an ANN.

Clustering with Self-Organizing Maps

- Teaching the SOM: the species abundance is computed for each virtual units.

- Computing the U-matrix.

- Mapping the Sample Units onto the U-matrix.

- Making the clustering structure apparent for the human expert of the dataset by selecting the brightness of the display.

Clustering with Self-Organizing Maps

- With a large dataset, when dendrograms become very difficult to read, the SOM and the U-matrix are able to provide a very convenient visualisation.
- U-matrix is not a "ready made" clustering algorithm but rather a tool for the inspection of high dimensional data.
- The clusters have to be « seen » on the map by the human dataset expert.
- The expert can define all types of clusters including the non-convex ones.

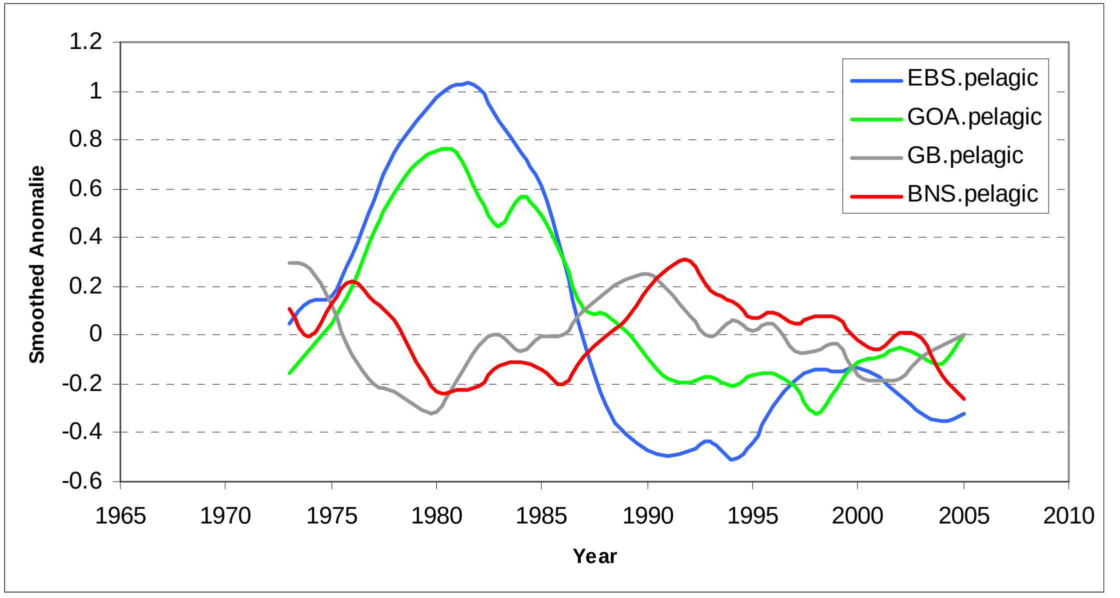
K-means Clustering

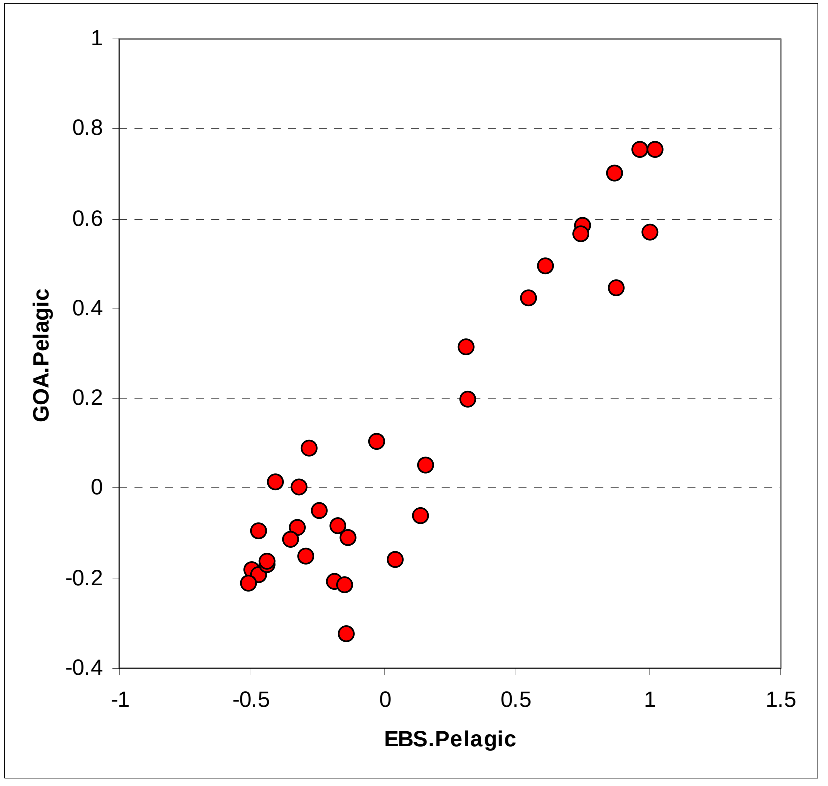
***k*-means clustering** is a method of cluster analysis which aims to partition *n* observations into *k* clusters in which each observation belongs to the cluster with the nearest mean. It is similar to the expectation-maximization algorithm for mixtures of Gaussians in that they both attempt to find the centers of natural clusters in the data.

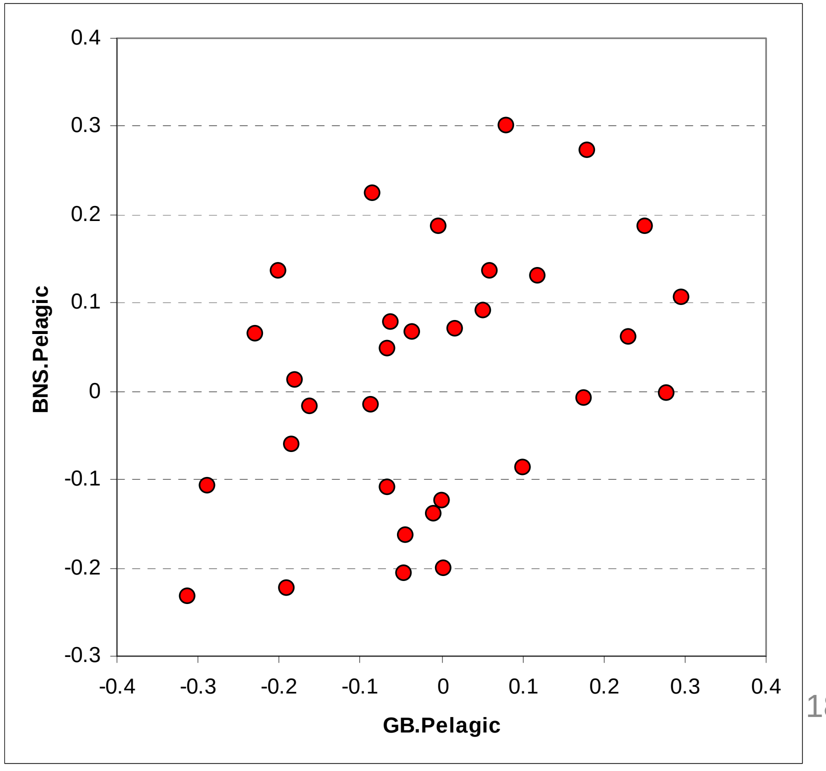
It has been shown that the relaxed solution of *k*-means clustering, specified by the cluster indicators, is given by the PCA (principal component analysis) principal components, and the PCA subspace spanned by the principal directions is identical to the cluster centroid subspace specified by the between-class scatter matrix.

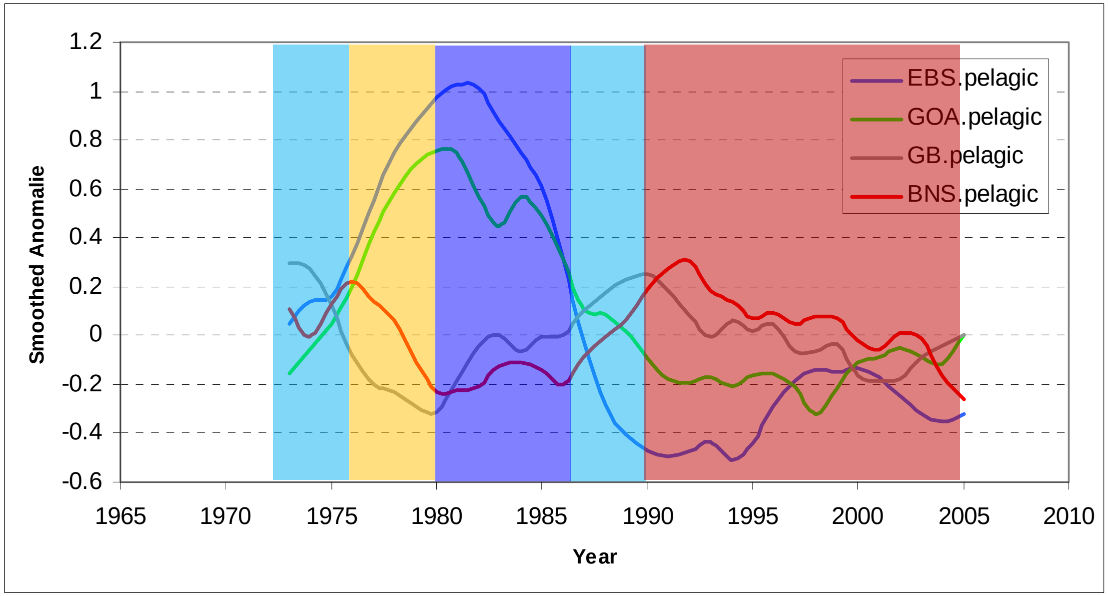
Apply model to four ecosystem data

- Ln(R/S) of pelagic
- Ln(R/S) of benthic
- Ln(R/S) of the pooled for ecosystems

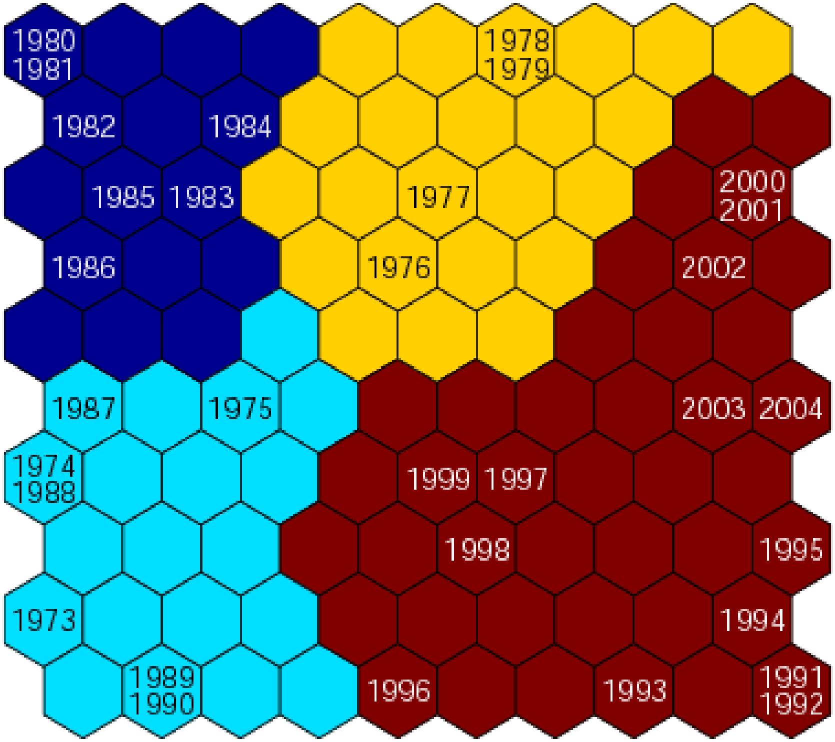

Smoothed Anomalie
Year
EBS.pelagic
GOA.pelagic
GB.pelagic
BNS.pelagic
1.2
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
1965
1970
1975
1980
1985
1990
1995
2000
2005
2010


GOA.Pelagic
EBS.Pelagic
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-1
-0.5
0
0.5
1
1.5

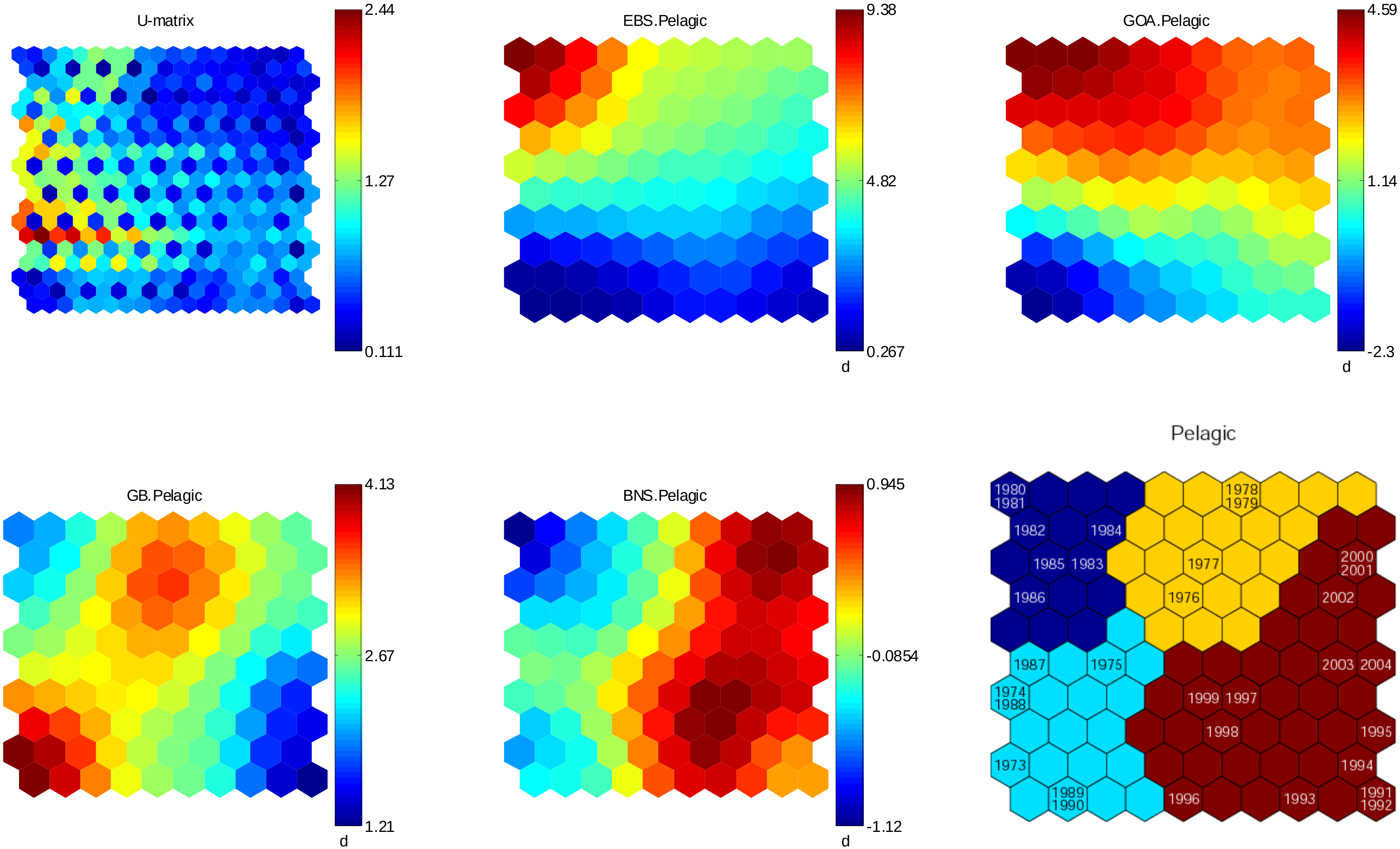

BNS.Pelagic
GB.Pelagic
0.4
0.3
0.2
0.1
0
-0.1
-0.2
-0.3
-0.4
-0.3
-0.2
-0.1
0
0.1
0.2
0.3
0.4



Smoothed Anomalie
Year
EBS.pelagic
GOA.pelagic
GB.pelagic
BNS.pelagic
1.2
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
1965
1970
1975
1980
1985
1990
1995
2000
2005
2010

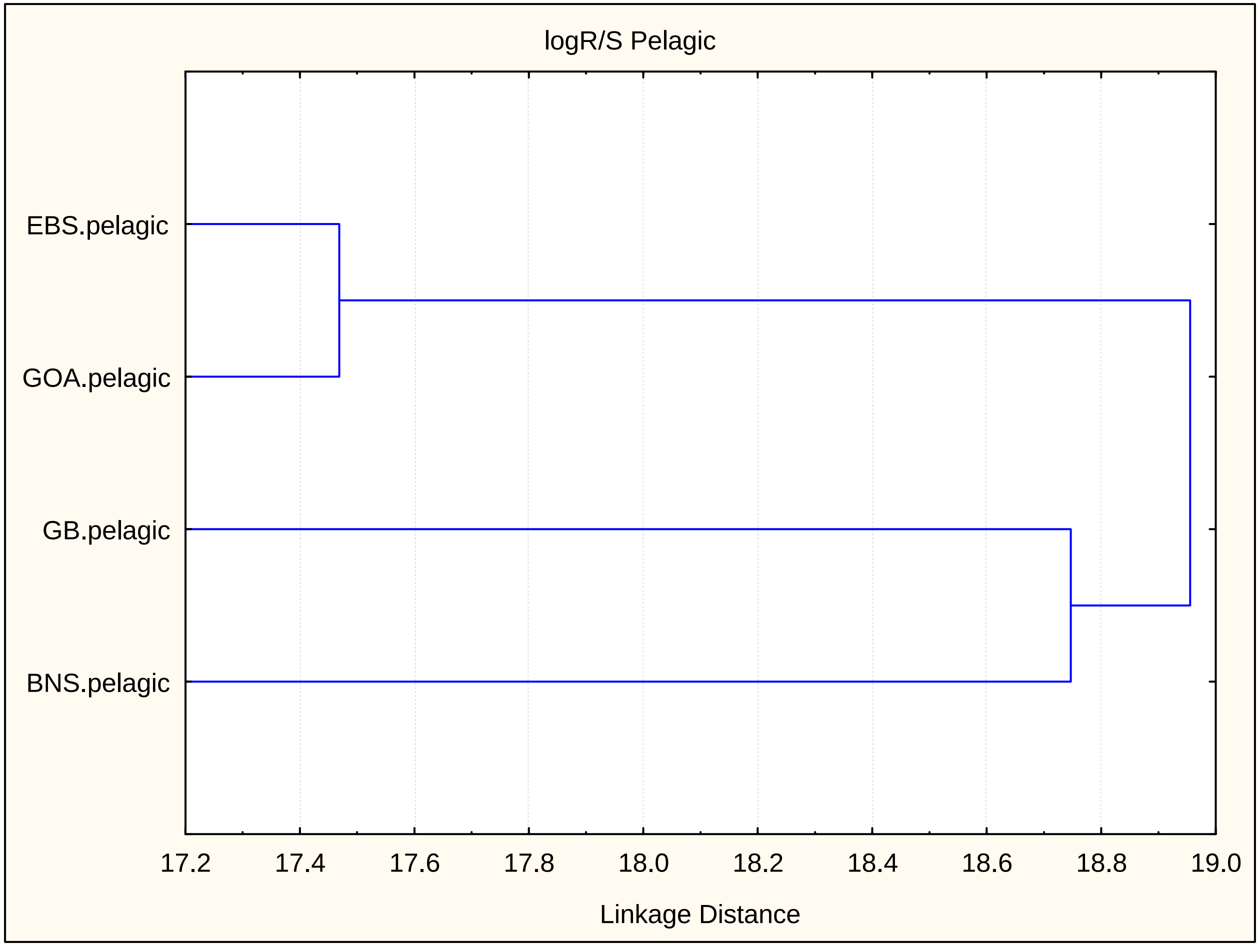


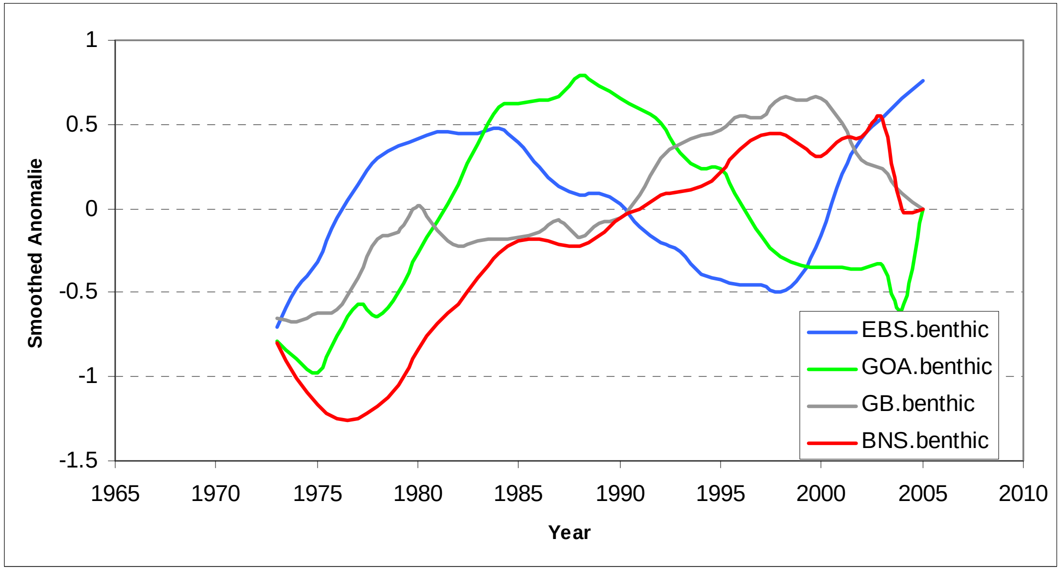
1980
1981
1978
1979
1982
1984
1985
1983
1977
2000
2001
1986
1976
2002
1987
1975
2003
2004
1974
1988
1999
1997
1998
1995
1973
1994
1989
1990
1996
1993
1991
1992

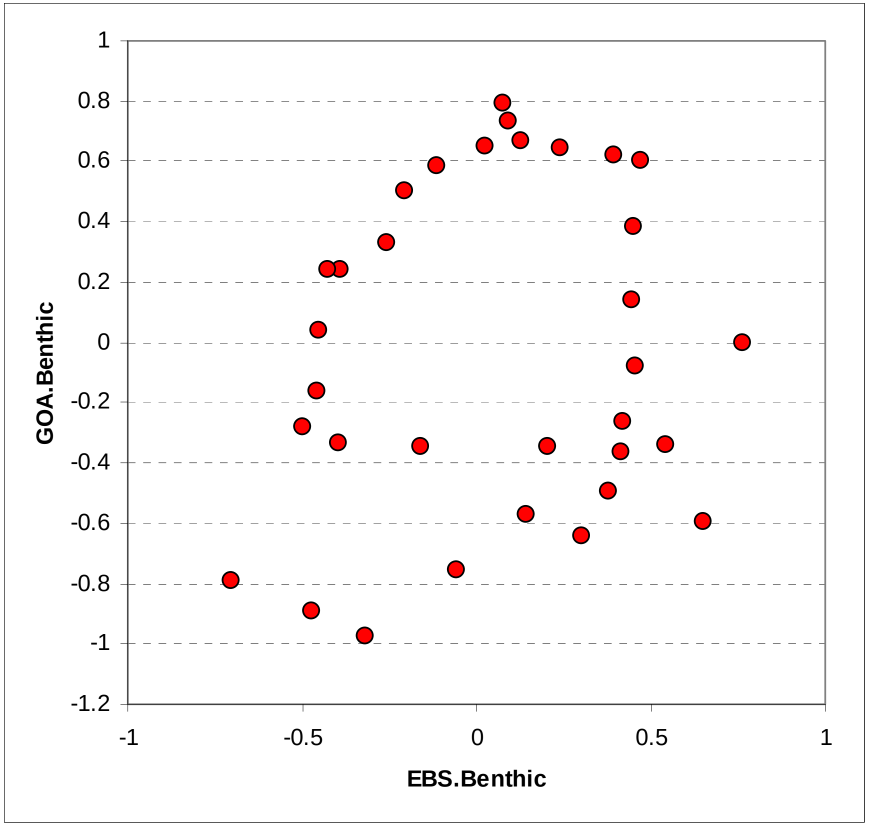


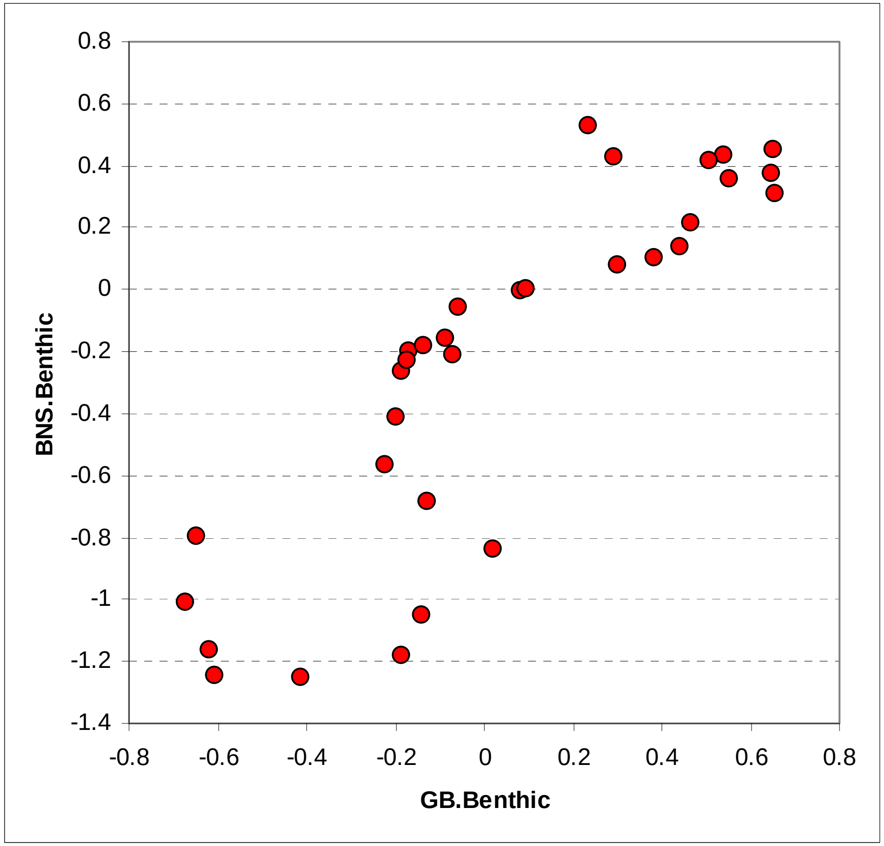
U-matrix
2.44
1.27
0.111
EBS.Pelagic
9.38
4.82
0.267
d
GOA.Pelagic
4.59
1.14
-2.3
d
GB.Pelagic
4.13
2.67
1.21
d
BNS.Pelagic
0.945
-0.0854
-1.12
d
Pelagic
1980
1981
1978
1979
1982
1984
2000
2001
1985
1983
1977
1986
1976
2002
1987
1975
2003
2004
1974
1988
1999
1997
1998
1995
1973
1994
1989
1990
1996
1993
1991
1992

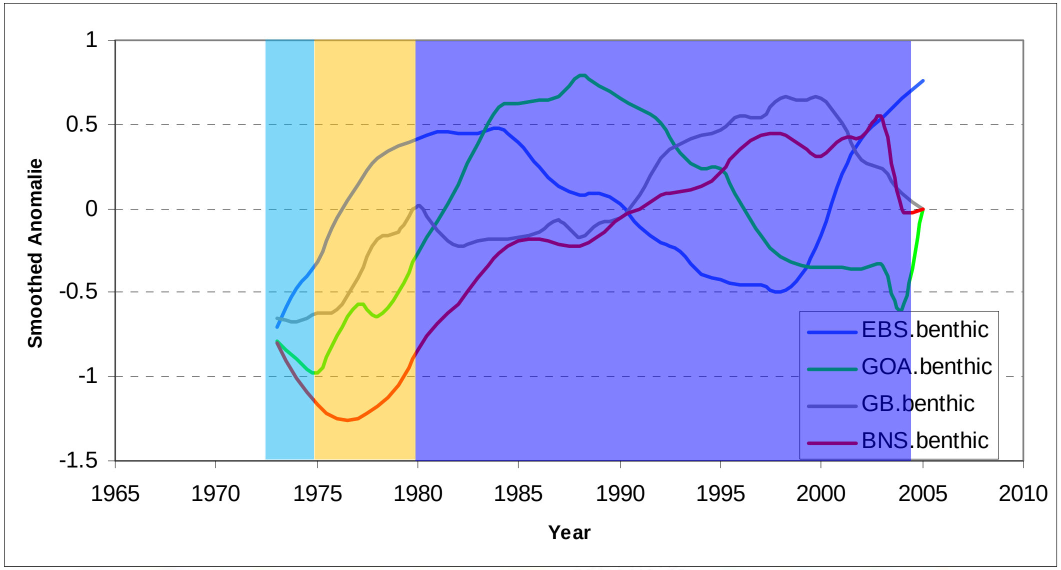
gic-CA


logR/S Pelagic
EBS.pelagic
GOA.pelagic
GB.pelagic
BNS.pelagic
17.2
17.4
17.6
17.8
18.0
18.2
18.4
18.6
18.8
19.0
Linkage Distance

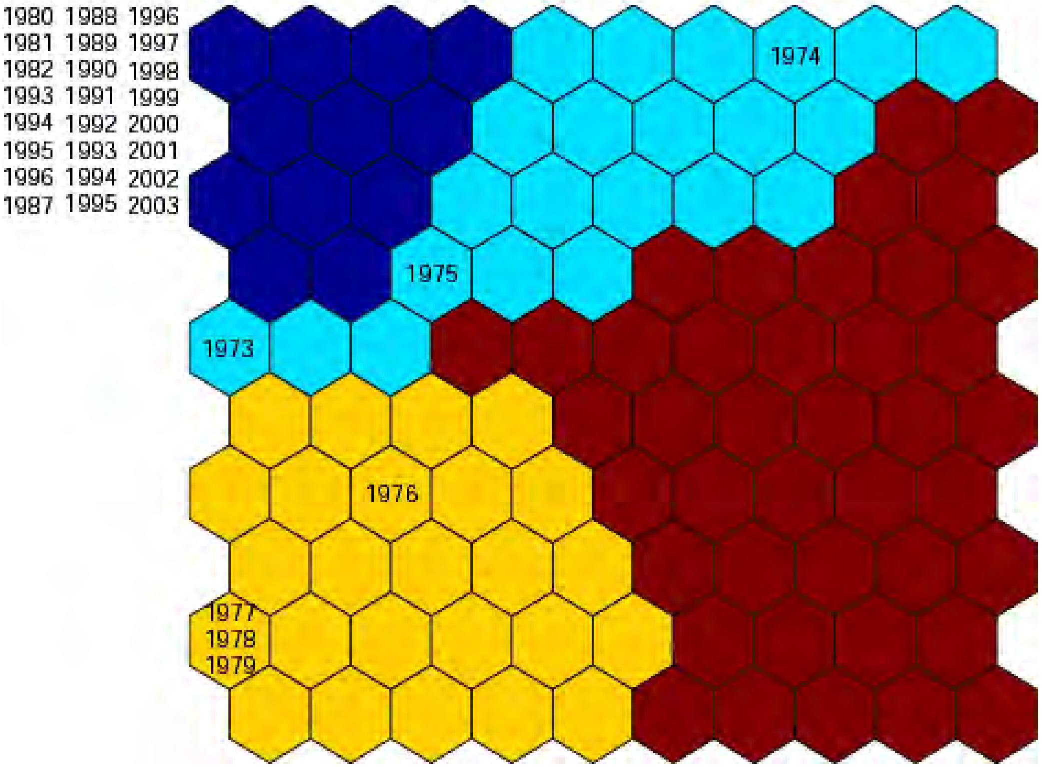

Smoothed Anomalie
Year
EBS.benthic
GOA.benthic
GB.benthic
BNS.benthic
1
0.5
0
-0.5
-1
-1.5
1965
1970
1975
1980
1985
1990
1995
2000
2005
2010


GOA.Benthic
EBS.Benthic
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1
-1.2
-1
-0.5
0
0.5
1

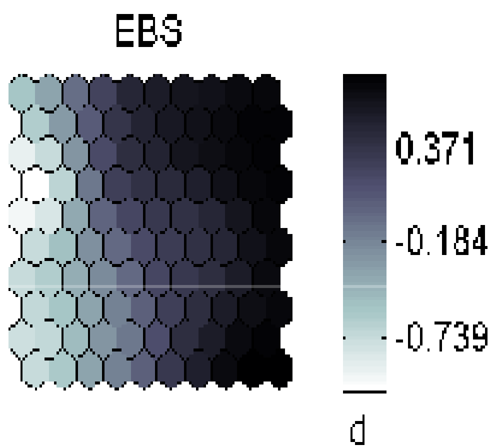

BNS.Benthic
GB.Benthic
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1
-1.2
-1.4
-0.8
-0.6
-0.4
-0.2
0
0.2
0.4
0.6
0.8

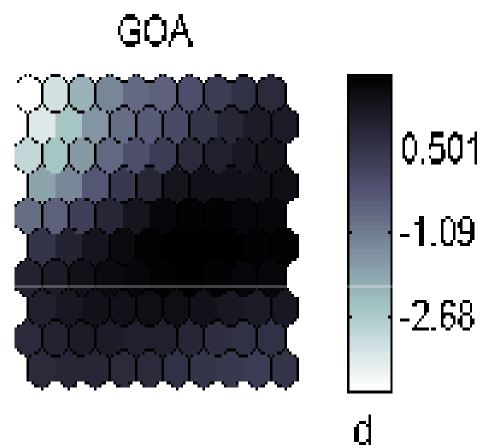


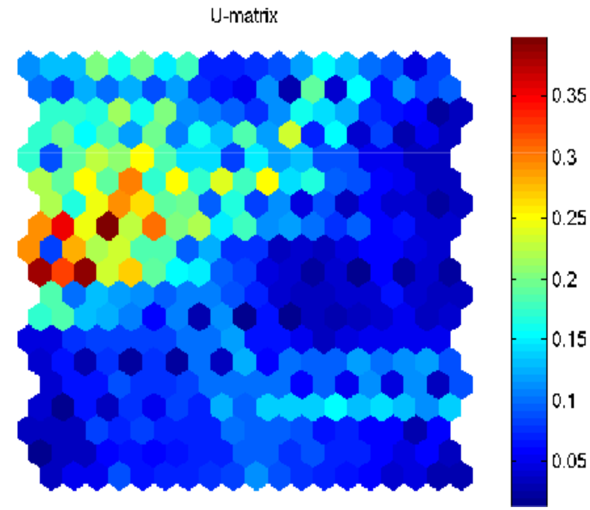
Smoothed Anomalie
1
0.5
0
-0.5
-1
-1.5
1965
1970
1975
1980
1985
1990
1995
2000
2005
2010
Year
EBS.benthic
GOA.benthic
GB.benthic
BNS.benthic

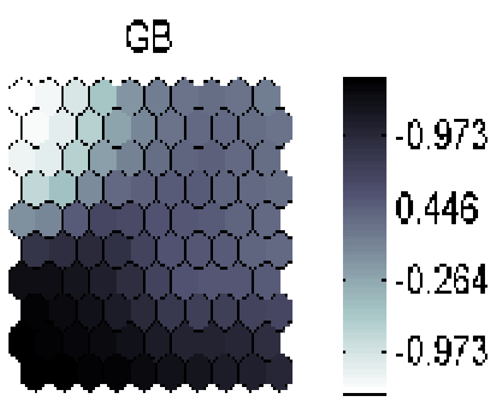


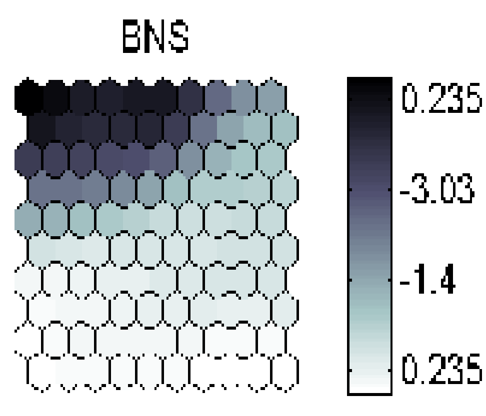
1980 1988 1996
1981 1989 1997
1982 1990 1998
1993 1991 1999
1994 1992 2000
1995 1993 2001
1996 1994 2002
1987 1995 2003
1974
1975
1973
1976
1977
1978
1979

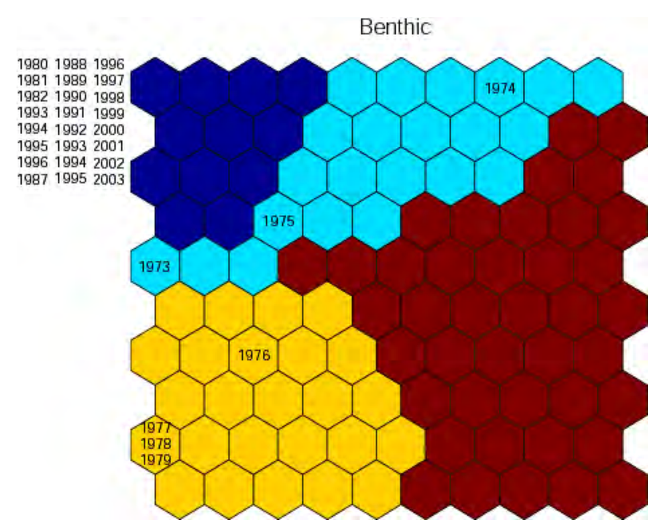

EBS
0.371
-0.184
-0.739
d


GOA
0.501
-1.09
-2.68
d

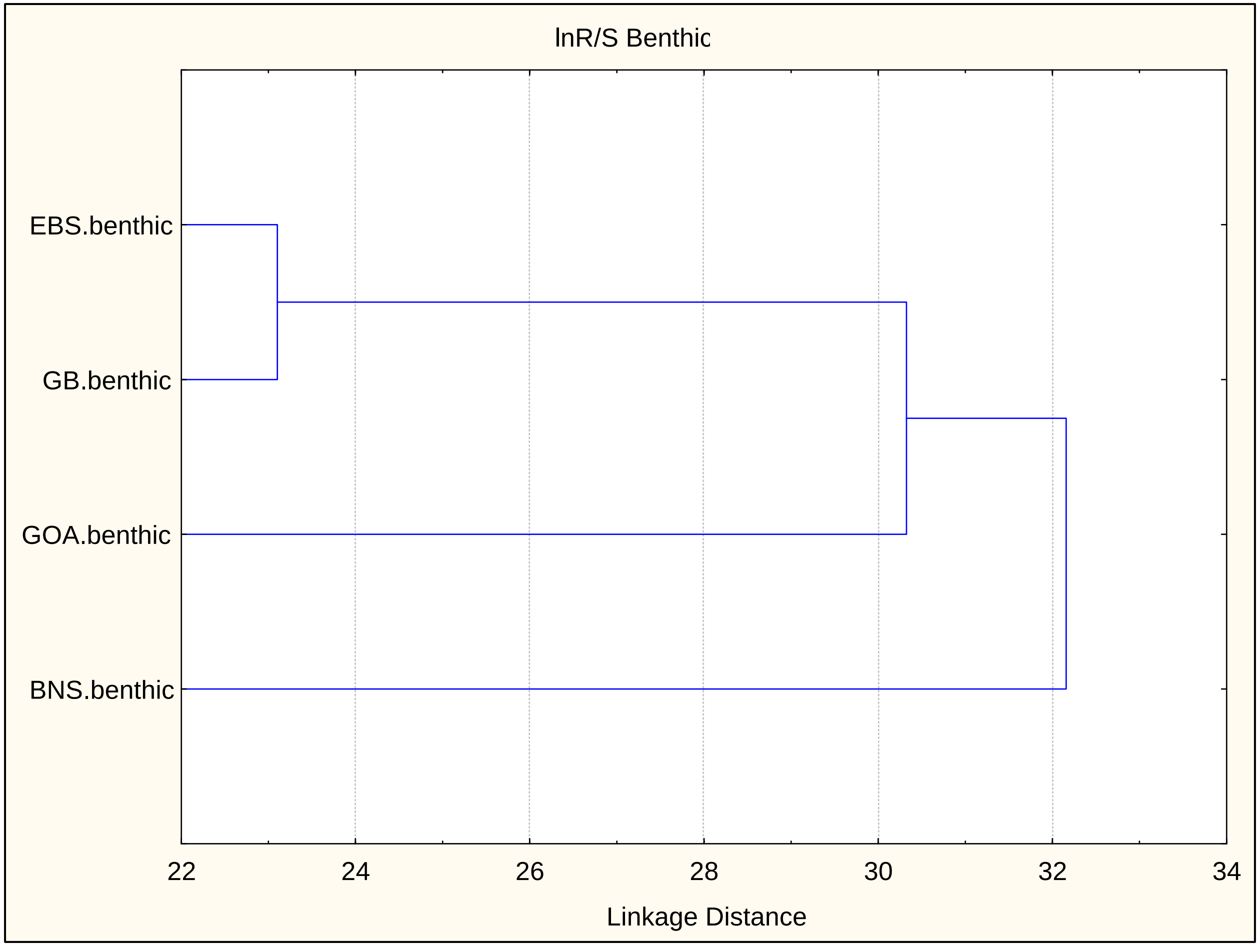

U-matrix
0.35
0.3
0.25
0.2
0.15
0.1
0.05

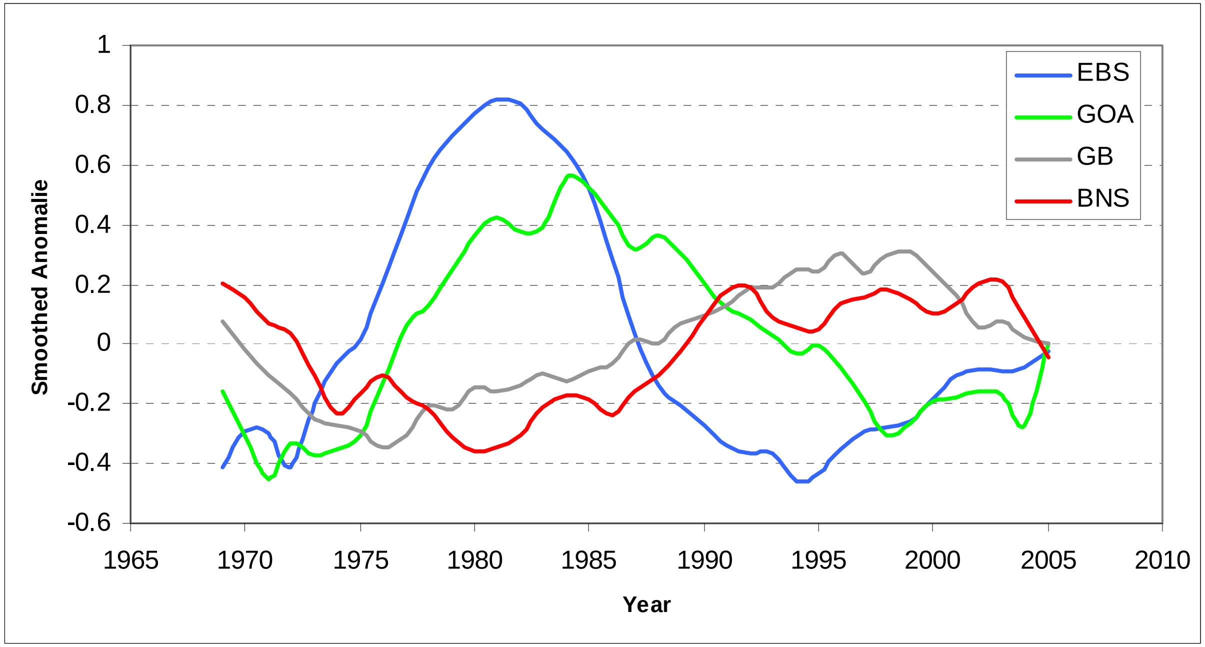

GB
-0.973
0.446
-0.264
-0.973


BNS
0.235
-3.03
-1.4
0.235

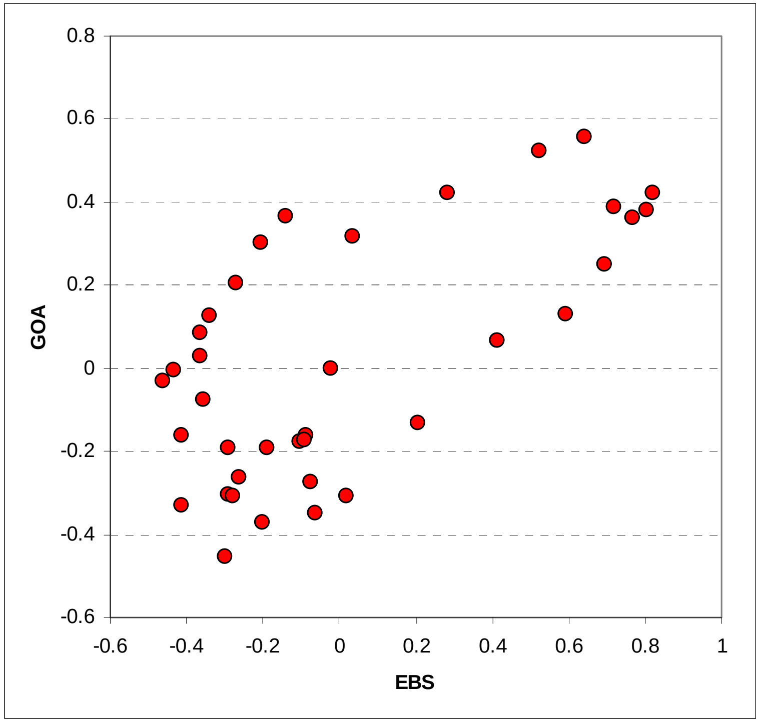

Benthic
1980 1988 1996
1981 1989 1997
1982 1990 1998
1993 1991 1999
1994 1992 2000
1995 1993 2001
1996 1994 2002
1987 1995 2003
1974
1975
1973
1976
1977
1978
1979

ic-CA

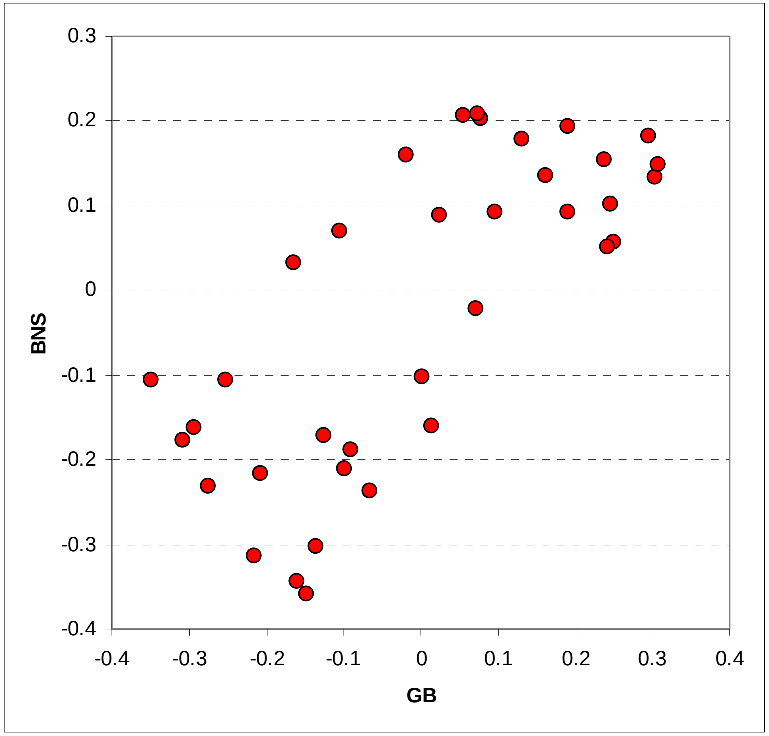

lnR/S Benthic
EBS.benthic
GB.benthic
GOA.benthic
BNS.benthic
22
24
26
28
30
32
34
Linkage Distance



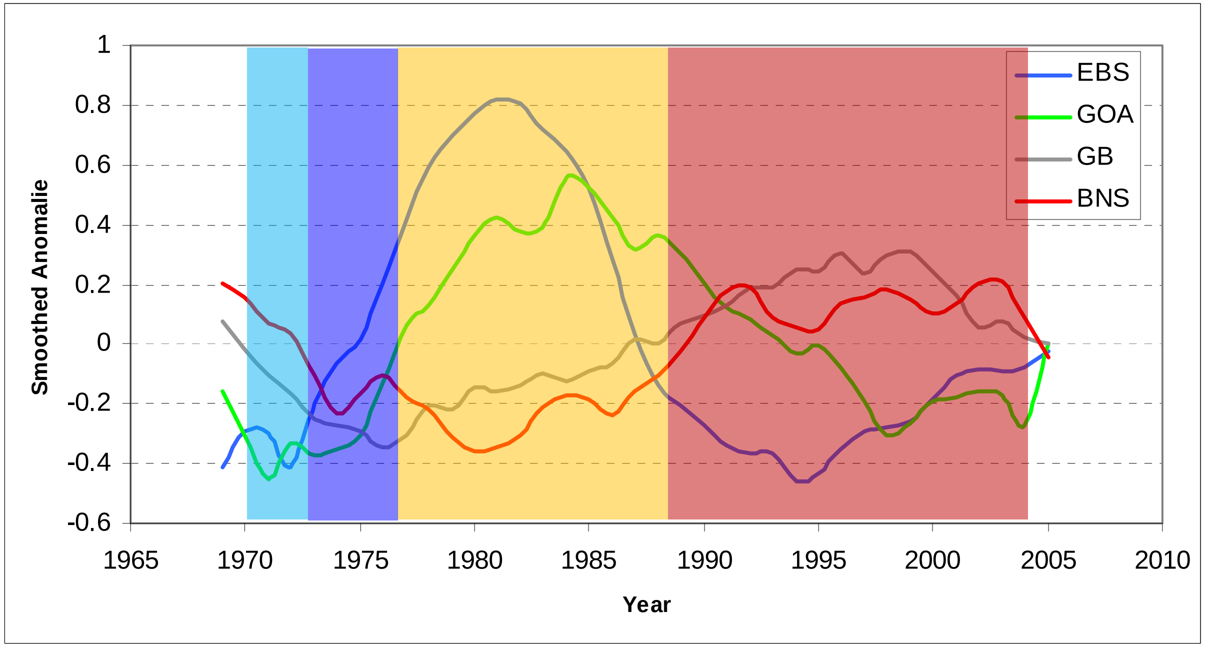
Smoothed Anomalie
Year
EBS
GOA
GB
BNS
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
1965
1970
1975
1980
1985
1990
1995
2000
2005
2010

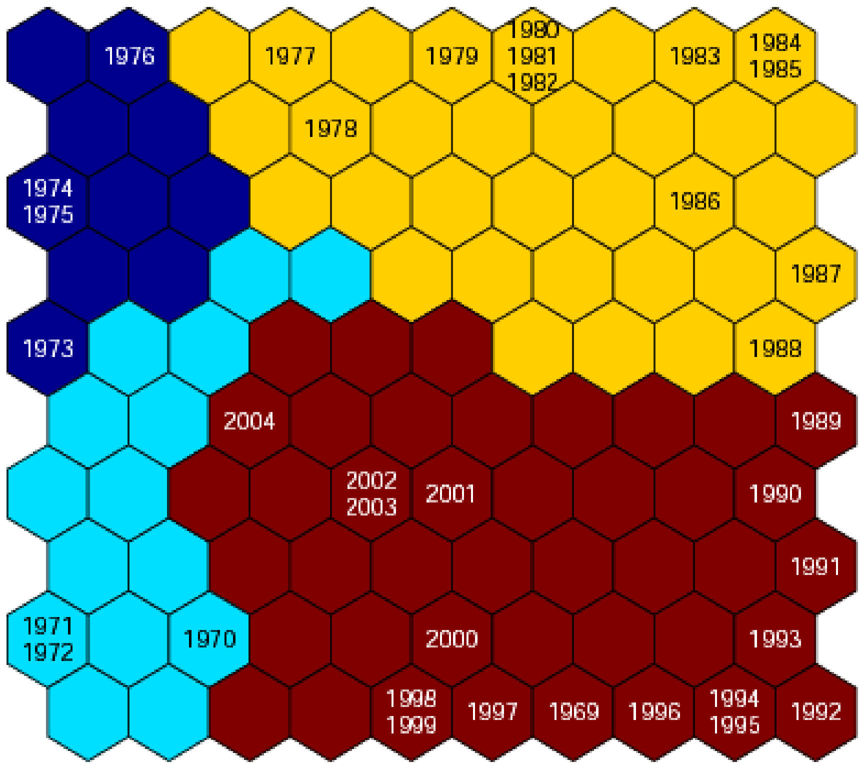


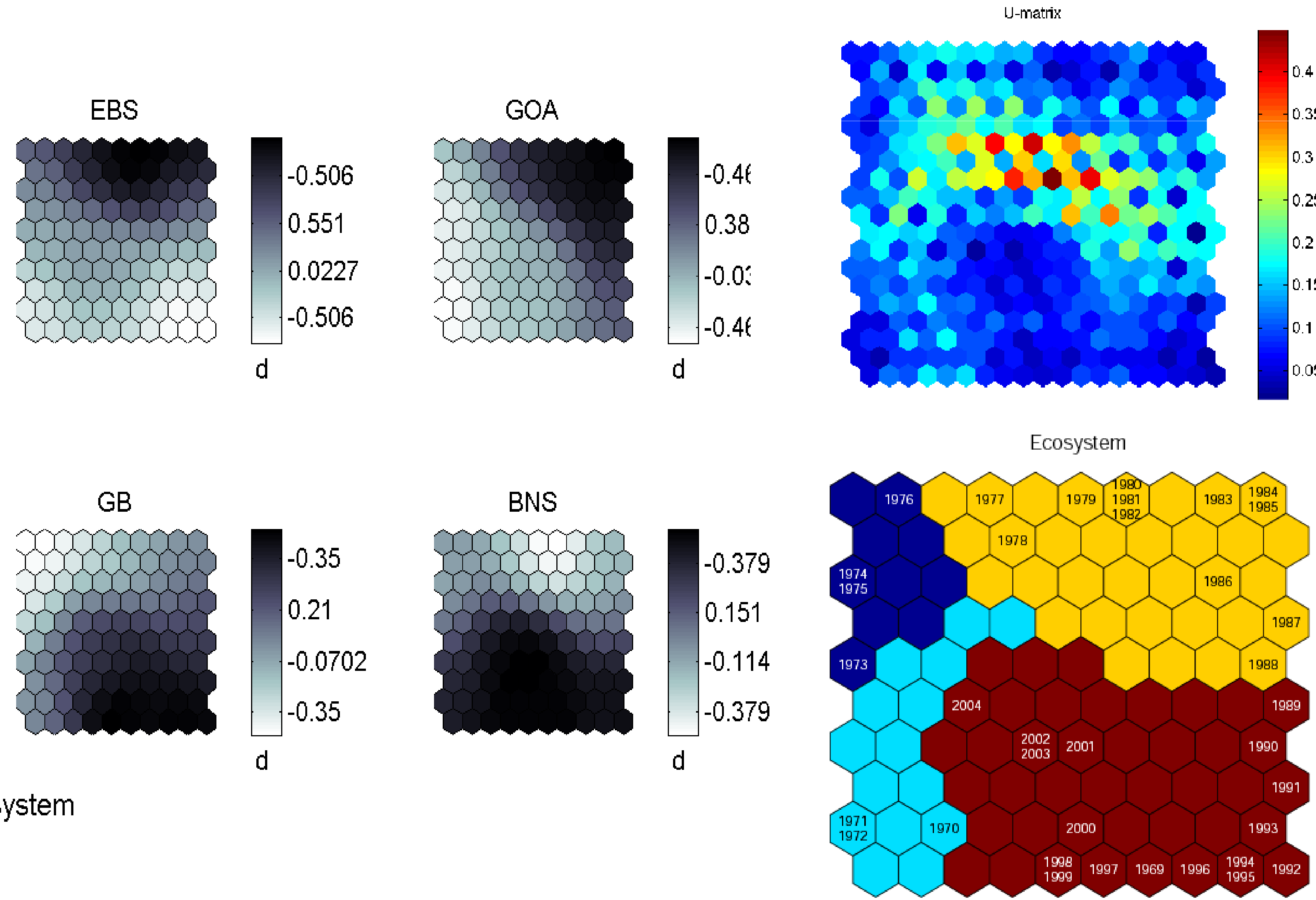
GOA
EBS
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.6
-0.4
-0.2
0
0.2
0.4
0.6
0.8
1



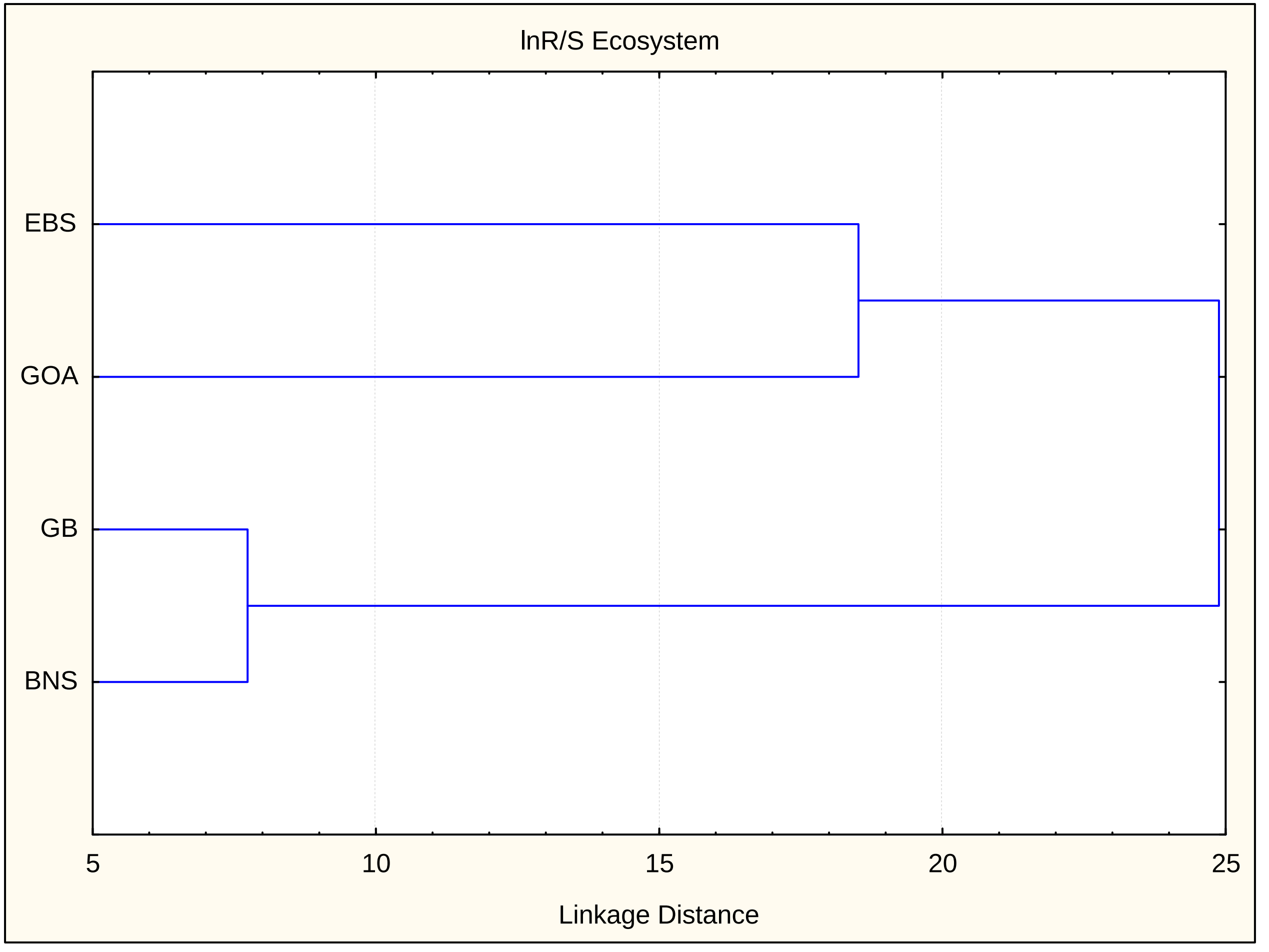
BNS
GB
0.3
0.2
0.1
0
-0.1
-0.2
-0.3
-0.4
-0.4
-0.3
-0.2
-0.1
0
0.1
0.2
0.3
0.4


Smoothed Anomalie
Year
EBS
GOA
GB
BNS
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
1965
1970
1975
1980
1985
1990
1995
2000
2005
2010


1976
1977
1979
1980
1981
1982
1983
1984
1985
1978
1974
1975
1986
1987
1973
1988
2004
1989
2002
2003
2001
1990
1991
1971
1972
1970
2000
1993
1998
1999
1997
1969
1996
1994
1995
1992

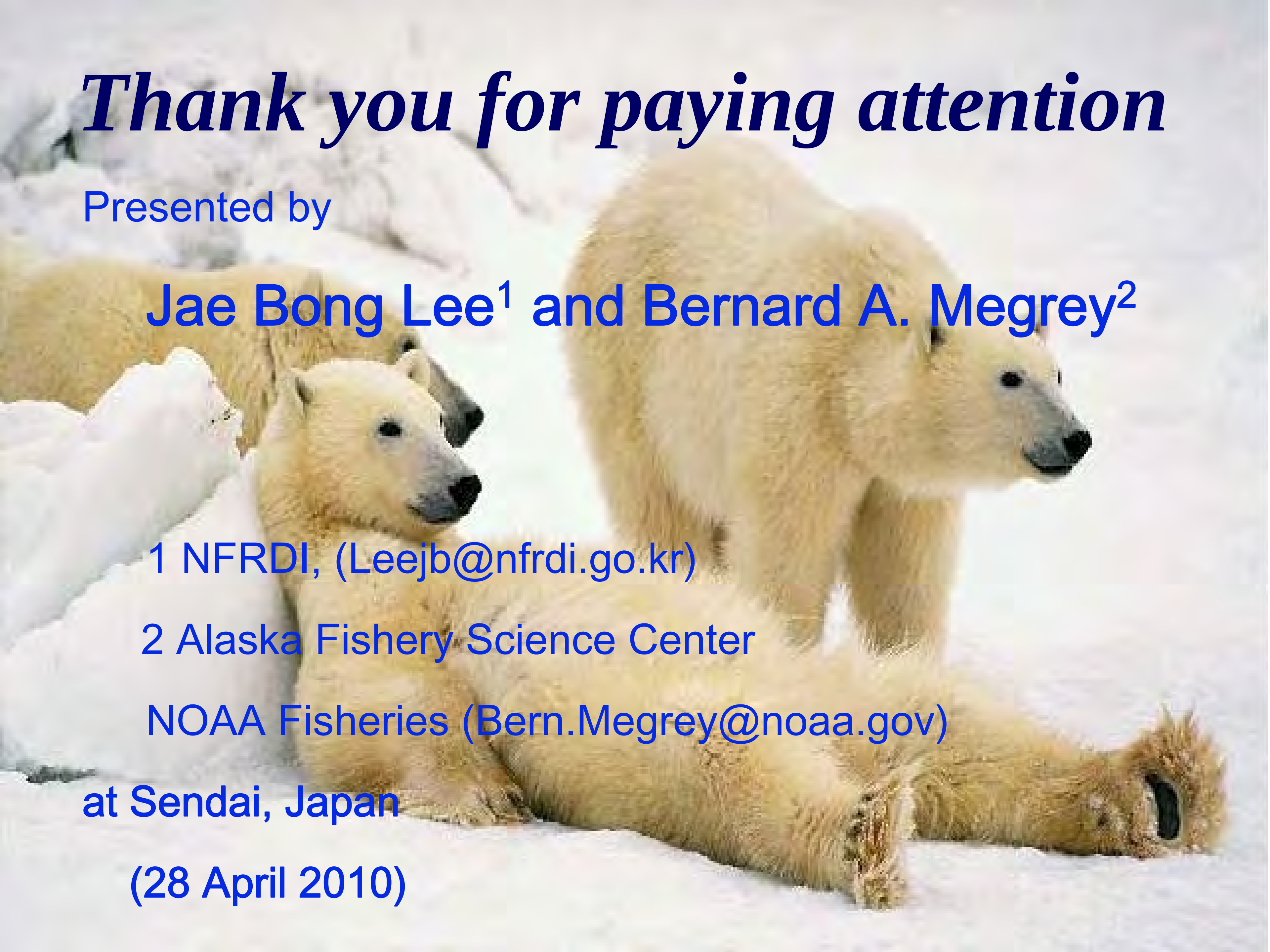

EBS
-0.506
0.551
0.0227
-0.506
d
GOA
0.38
d
GB
-0.35
0.21
-0.0702
-0.35
d
BNS
-0.379
0.151
-0.114
-0.379
d
U-matrix
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
Ecosystem
1976
1977
1979
1980
1981
1982
1983
1984
1985
1978
1974
1975
1986
1987
1973
1988
2004
1989
2002
2003
2001
1990
1991
1971
1972
1970
2000
1993
1998
1999
1997
1969
1996
1994
1995
1992

Ecosystem


lnR/S Ecosystem
EBS
GOA
GB
BNS
5
10
15
20
25
Linkage Distance

Conclusions

- SOMS and k-means clustering provide a highly visual tool to easily identify patterns in the timing of high or low productivity years across both species and ecosystems.
- Many of the peaks in the time series were synchronous within an ocean basin and opposing alternations in patterns of productivity were observed in ecosystems in between the Atlantic and Pacific Ocean basins.
- Basin-scale results (similar within but different between) suggest that productivity in the two geographically broad areas are connected by unknown climatic mechanisms that, depending on the year, generate low frequency opposing alternations after 1976 and 1988 in the two basins.



Thank you for paying attention

Presented by

Jae Bong Lee[1] and Bernard A. Megrey[2]

1 NFRDI, (Leejb@nfrdi.go.kr)

2 Alaska Fishery Science Center

NOAA Fisheries (Bern.Megrey@noaa.gov)

at Sendai, Japan

(28 April 2010)