Efstratia Tramountani	Ricarda Hogl	Model mortality rates in WE using cold / warm extremes	Model economic indicators of WE countries using climatic observables
Michael Gröger	Carmen Gil	Predict spatial pattern of WE surface temperatures using CNNs	Predict spatial pattern of average wind speeds using CNNs
Stefan Zürn	Alexander Braun	Predict spatial pattern of WE surface temperatures using CNNs	Predict spatial pattern of WE rainfall using CNNs
Leonard Siegert	Rosanna Krebs	Predict WE surface temperatures using LSTMs	Predict WE rainfall using LSTMs
Ludwig Bald	Christian Fröhlich	Model rainfall over WE as a correlate of North Atlantic Oscillation (NAO)	Model temperatures in WE as a correlate of the Arctic Oscillation (AO)
Jolanda Dünner	Robert Neulen	Predict WE rainfall using LSTMs	Predict WE rainfall using Gaussian processes
Gereon Recht	Josua Stadelmaier	Estimate regions of similar behaviour for average wind speeds over WE using climate net	Estimate regions of similar behaviour for WE rainfall using climate network communities
Johannes Schulz	Adrian Stock	Estimate regions of similar behaviour for WE rainfall using climate network communities	Estimate regions of similar behaviour for WE surface temperatures with climate network communities
Felix Strnad	Moritz Haas	Other (i.e. your own idea, which you can inform me via email / Discord)	Estimate regions of similar behaviour for WE rainfall using climate network communities
Merle Kammer	Mara Seyfert	Predict spatial pattern of WE rainfall using CNNs	Predict spatial pattern of WE surface temperatures using CNNs
Kari Gustedt	Naman	Model mortality rates in WE using cold / warm extremes	Predict spatial pattern of WE surface temperatures using CNNs
Dorothee Sigg	Frieder Göppert	Estimate latent factors underlying WE surface temperatures using EOFs	Estimate latent factors underlying WE rainfall using VAEs
shiaw-shiuan Chuang	only me	Predict spatial pattern of WE rainfall using CNNs	Estimate latent factors underlying WE surface temperatures using VAEs
Julian Petruck	Dexter Früh	Model surface temperature over WE as a correlate of atmospheric CO2 levels	Estimate latent factors underlying WE surface temperatures using VAEs

LECTURE 6:
Non-neural network approaches

ML-4430: Machine learning approaches in climate science

2 June 2021

Empirical Orthogonal Functions

1

- What are EOFs?
- Considerations in estimating EOFs
- Examples

Cluster Analysis

2

- What is cluster analysis?
- Hierarchical clustering
- Non-hierarchical clustering

Bedartha Goswami

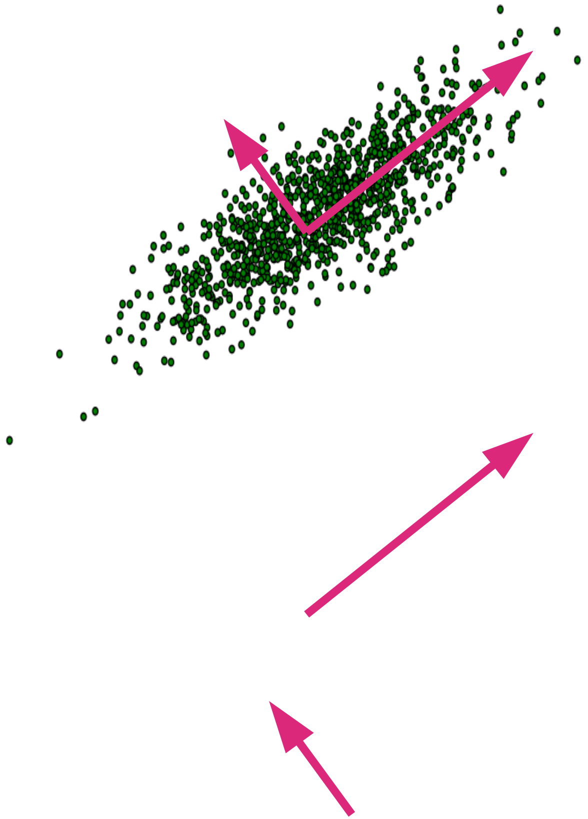
EOF <=> PCA

- Consider the data matrix
  $$X = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_p),$$
  where p is the number of locations and
  $$\mathbf{x}_j = (x_{1j}, x_{2j}, \dots, x_{nj})^T$$
  is the time series of length *n* at location *j*

- Esimate the covariance matrix of size *p* x *p*
  $$C = X^T X$$

- Empirical orthogonal functions are the eigenvectors of **C** → allowing a change of basis

- It identifies the dominant directions of variability in the data



EOF <=> PCA

- Projecting the data onto each eigen direction reveals the different dominant “modes” of variability

- For a single time instant,

$$y^1_t = \boldsymbol{X}_t \rightarrow \boldsymbol{e}_1 = [x_{t1}, x_{t2}, \dots , x_{tp}]\, [e_{11}, e_{12}, \dots , e_{p1}]^T$$

- That is, for the entire time series

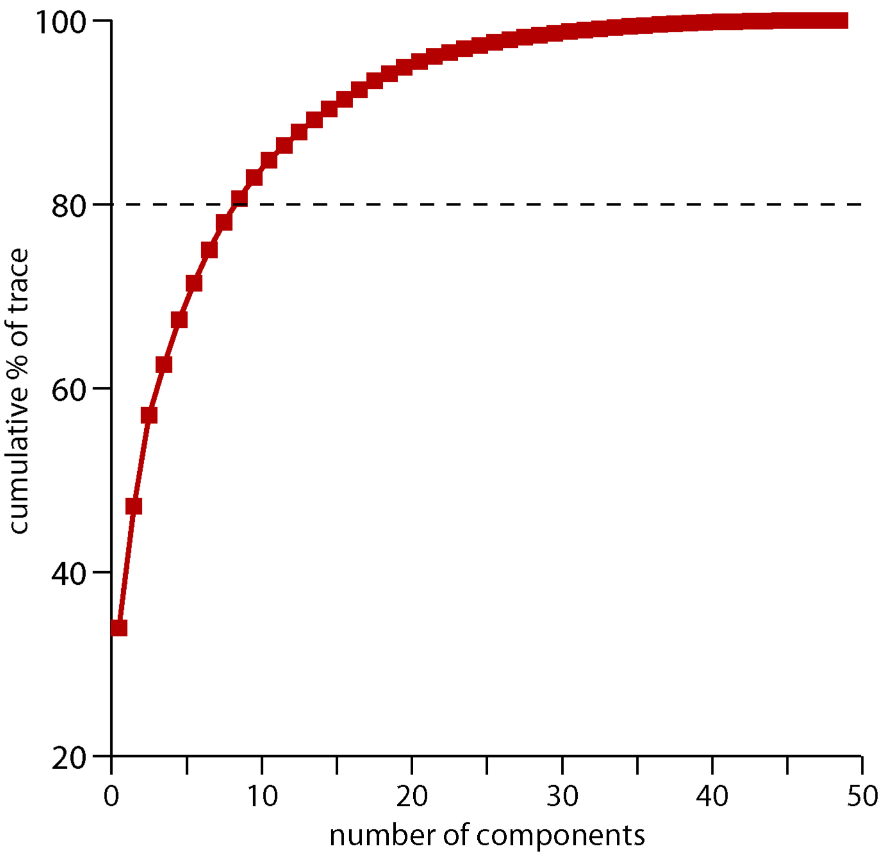
$$\boldsymbol{y}^1 = \mathrm{X}\, \boldsymbol{e}_1$$

gives the EOF time series for the first eigen direction

Considerations in estimating EOFs

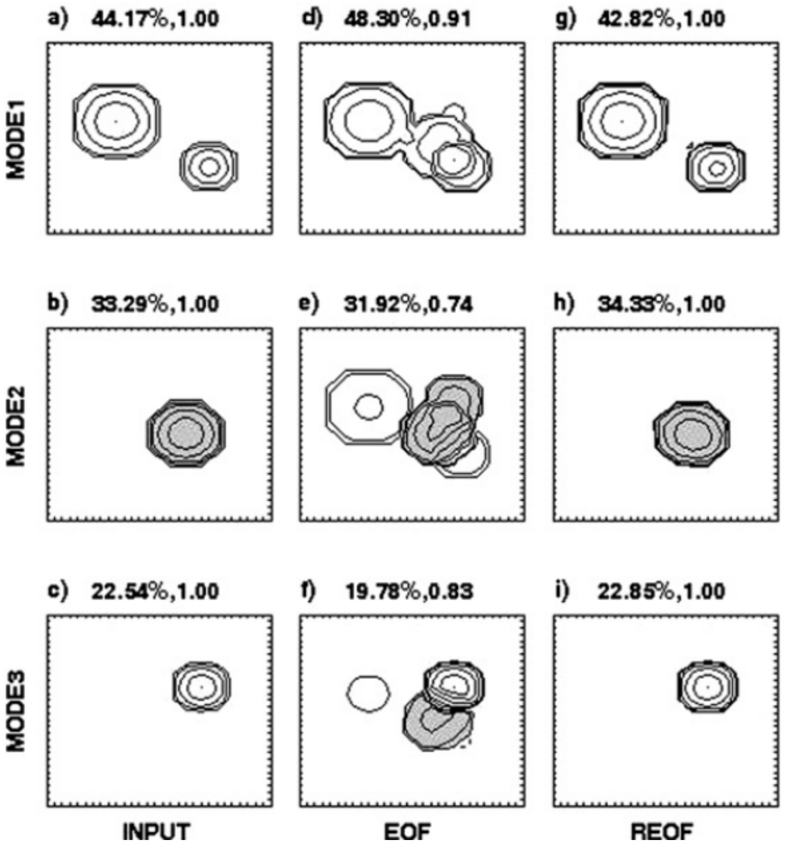
- Truncation:
  - Use only $k << p$ leading eigenvalues

$$\frac{\mathrm{var}(x_n)}{\sum_{i=1}^{N}\mathrm{var}(x_i)} = \frac{\mu_n}{\sum_{i=1}^{N}\mu_i}.$$



Considerations while estimating EOFs

- Truncation:
  - Use only $k << p$ leading eigenvalues
- Rotation:
  - Obtain ‘simple’ structures
  - Minimally overlapping EOFs
  - Ease of interpretation



1. Empirical Orthogonal Functions → Considerations in estimating EOFs

Considerations while estimating EOFs

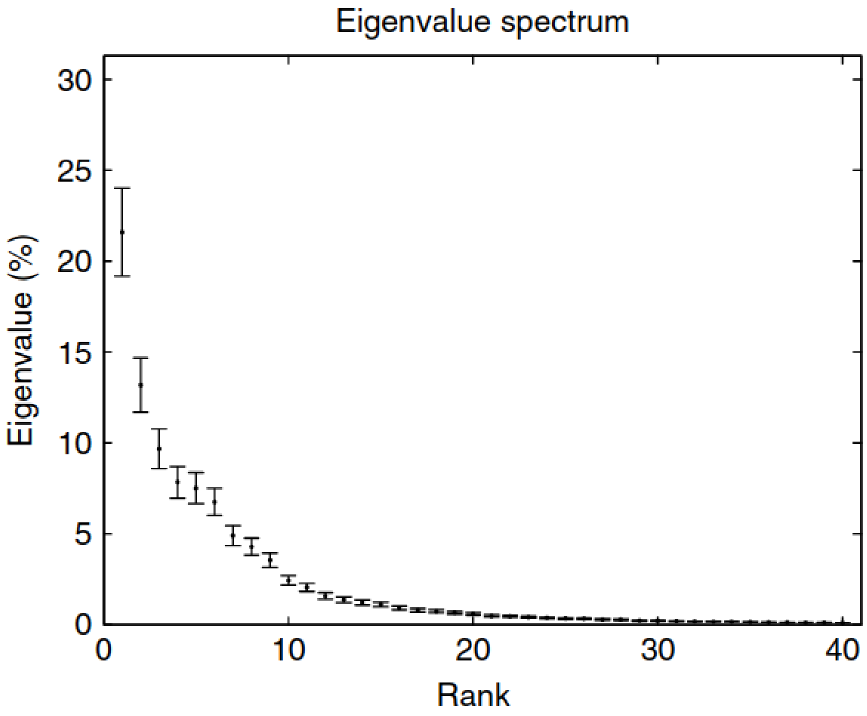
- Truncation:
  - Use only $k << p$ leading eigenvalues

- Rotation:
  - Obtain ‘simple’ structures
  - Minimally overlapping EOFs
  - Ease of interpretation

- Uncertainty
  - North’s rule of thumb (North et al., 1982)
  - Monte Carlo sampling

$$\Delta\lambda_k^2 \sim \lambda_k^2\sqrt{\frac{2}{n^*}}$$

$$\Delta\mathbf{u}_k \sim \frac{\Delta\lambda_k^2}{\lambda_j^2 - \lambda_k^2}\mathbf{u}_j$$



Spectrum, in percentage, of the covariance matrix of wintermonthly (DJF) SLP as per North’s rule of thumb

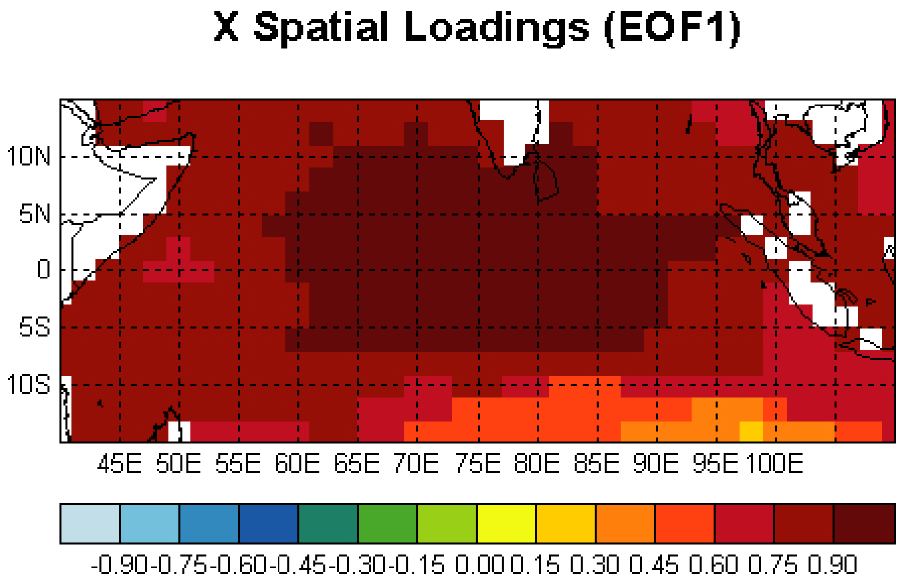
Considerations while estimating EOFs

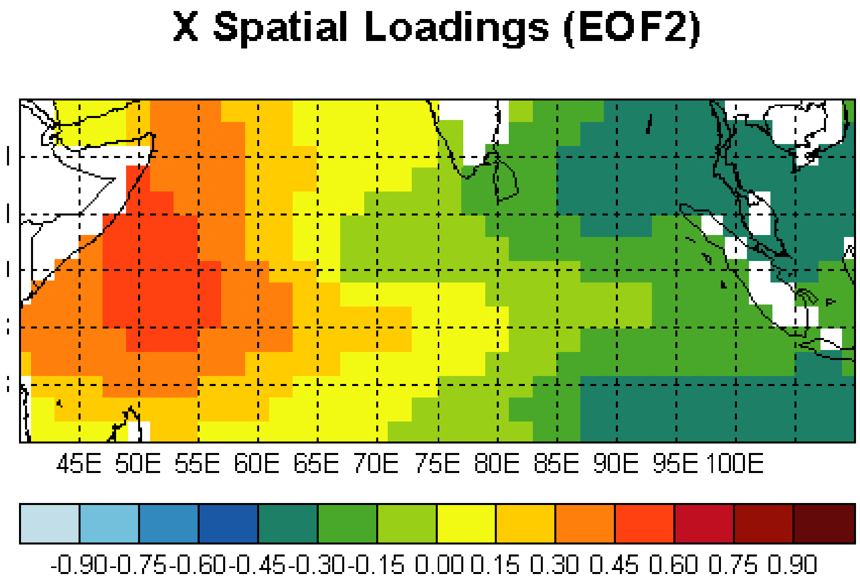
- Truncation:
  - Use only $k << p$ leading eigenvalues

- Rotation:
  - Obtain 'simple' structures
  - Minimally overlapping EOFs
  - Ease of interpretation

- Uncertainty
  - North's rule of thumb (North et al., 1982)
  - Monte Carlo sampling

- Spatial effects
  - For a rectangular grid, scale (co)variances by latitudes
  - Buell Patterns

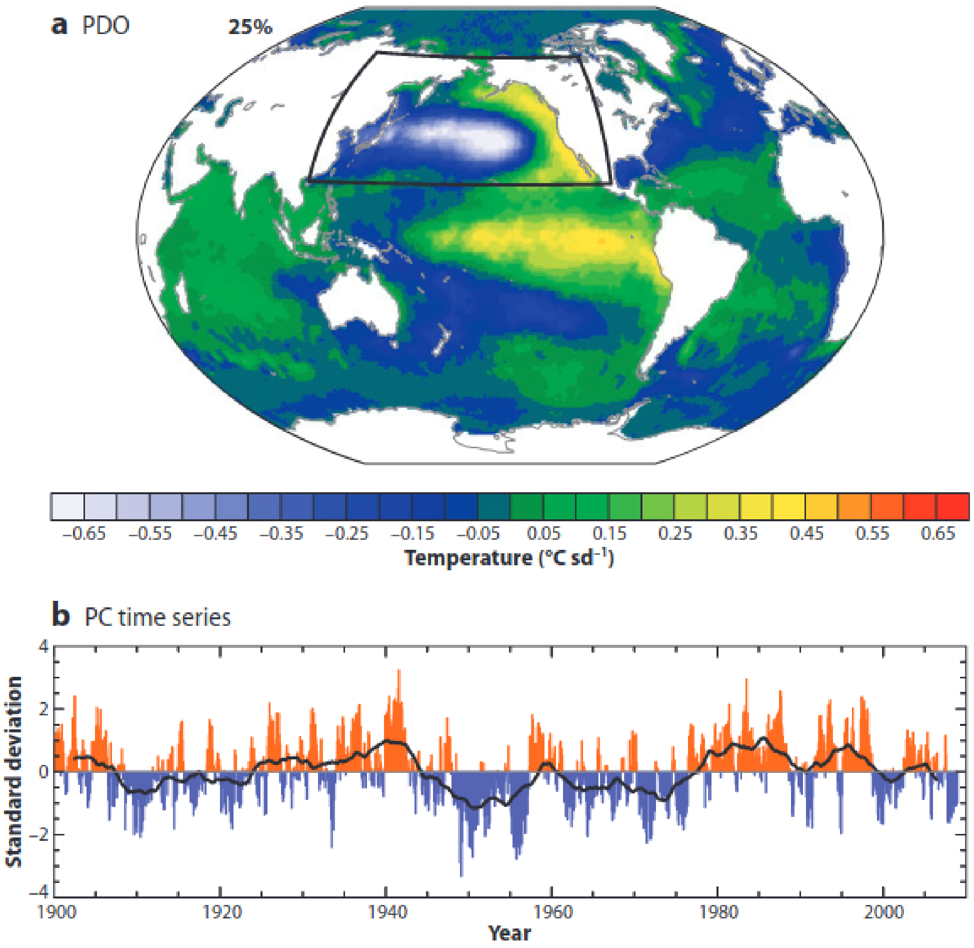




Bedartha Goswami

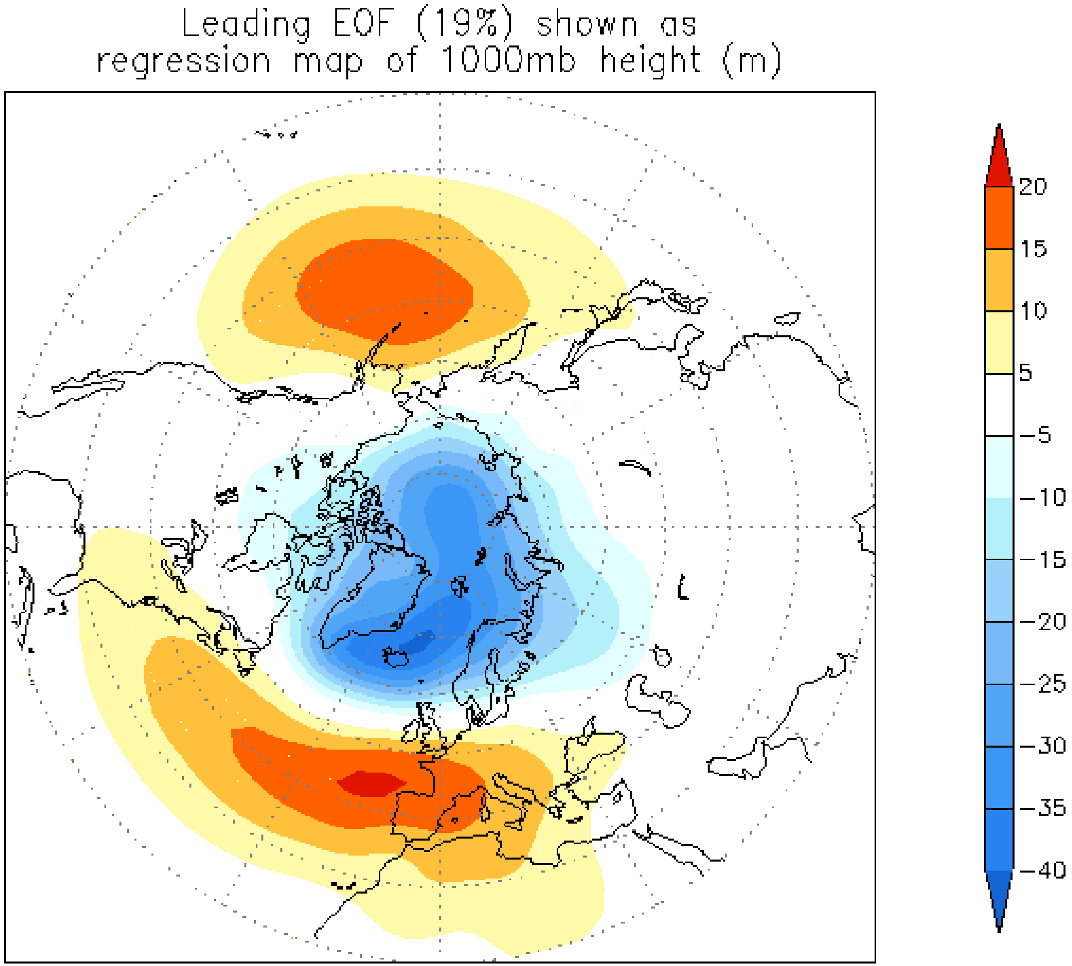
Considerations while estimating EOFs

- Truncation:
  - Use only *k* << *p* leading eigenvalues
- Rotation:
  - Obtain ‘simple’ structures
  - Minimally overlapping EOFs
  - Ease of interpretation
- Uncertainty
  - North’s rule of thumb (North et al., 1982)
  - Monte Carlo sampling
- Spatial effects
  - For a rectangular grid, scale (co)variances by latitudes
  - Buell Patterns



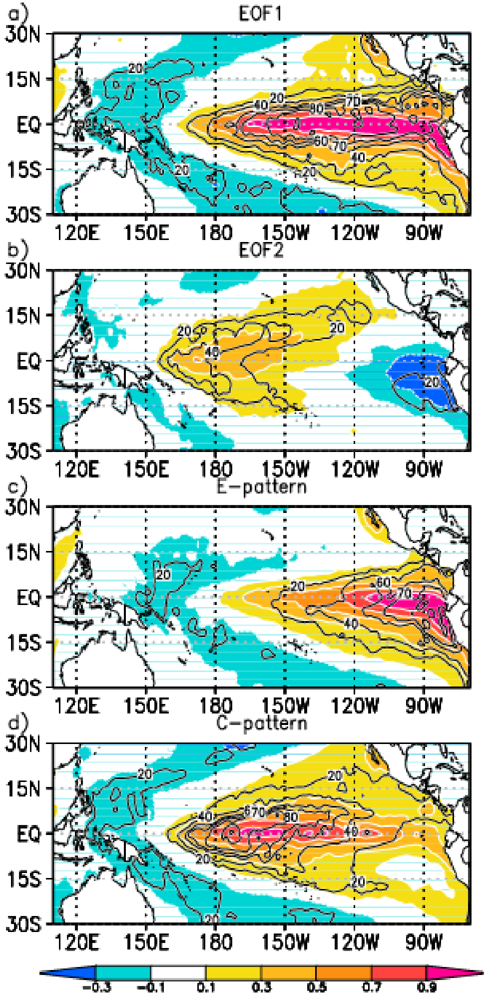
1. Empirical Orthogonal Functions → Examples: Pacific decadal oscillation

https://www.cpc.ncep.noaa.gov/products/precip/CWlink/daily_ao_index/ao.loading.shtml



1. Empirical Orthogonal Functions → Examples: Pacific decadal oscillation

Takahashi et al., Geophys. Res. Lett., 2011



Linear regression coefficients between SST data and the first two principal components of the SST data

Linear regression coefficients between SST data and Eastern Pacific Nino Index

Linear regression coefficients between SST data and Central PacificNino Index

1. Empirical Orthogonal Functions → Examples: ENSO

Empirical Orthogonal Functions

1

- What are EOFs?
- Considerations in estimating EOFs
- Examples

Cluster Analysis

2

- What is cluster analysis?
- Hierarchical clustering
- Non-hierarchical clustering

Outline

Clustering ...

- Once again, consider the data matrix
  $$\boldsymbol{X} = (\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_p),$$
  where p is the number of locations and
  $$\boldsymbol{x}_j = (x_{1j}, x_{2j}, \ldots, x_{nj})^T$$
  is the time series of length *n* at location *j*

- The goal of cluster analysis is to group the locations *p* into *k* groups on the basis of statistical notions of similarity between the locations

- For this we typically need a distance metric to quantify how similar (or ‘close’) two time series $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ are

Clustering ...

- We can choose different forms of distance metrics between two time series

- Eulidean metric

$$d_{ij} = [\ \Sigma_k\ (\ x_{ik} - x_{jk}\ )^{\ 2}\ ]^{1/2}$$

- Cosine metric

$$d_{ij} = \text{arccos}\ (\ \text{corr}\ (x_i, x_j)\ )$$

- Once you have a distance metric, you can choose between hierarchical and non-hierarchical methods

Hierarchical clustering ...

- Start with p clusters (as many clusters as locations)
  - each containing only one member, i.e., the location itself
- Merge the two clusters closest to each other
  - Now you have p – 1 clusters
- Next, merge the next two closest clusters
  - Now you have p – 2 clusters
- Keep merging until you have only one (trivial) cluster, which contains all nodes

Agglomerative clustering (based on distance between clusters)

- Single (or minimum) linkage

  $d(C_1, C_2) = \min d_{ij},$ for all i in C_1, j in C_2

- Complete (or maximum) linkage

  $d(C_1, C_2) = \max d_{ij},$ for all i in C_1, j in C_2

- Average linkage

  $d(C_1, C_2) = 1 / (n_1 n_2) \ \Sigma_k \Sigma_j \ d_{ij},$ for all i in C_1, j in C_2

- Centroid linkage

  $d(C_1, C_2) = || x^g_1 - x^g_2 ||$

  where x^g_1 and x^g_2 are the centroids of groups 1 and 2

Ward's method (based on minimum variance)

- *Does not need a distance metric*
- Start with p clusters, and keep merging until you have one cluster
- Merge so that the sum of squared distances of each point with respect to the centroid of its cluster is minimised
- Thus, at each step, find the merge that minimises

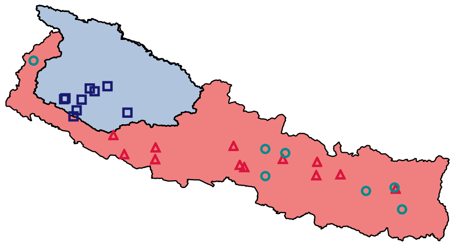
$$W = \Sigma_c \Sigma_i \Sigma_t \, || \, x_{it} - x^g_t \, ||^2$$

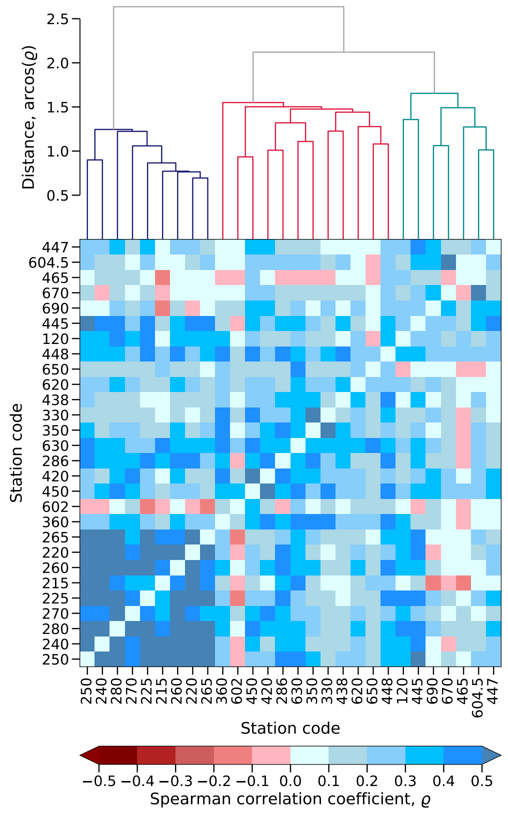
where t is the time index, i is the index for time series in each group, and c is the index that goes over the number of clusters

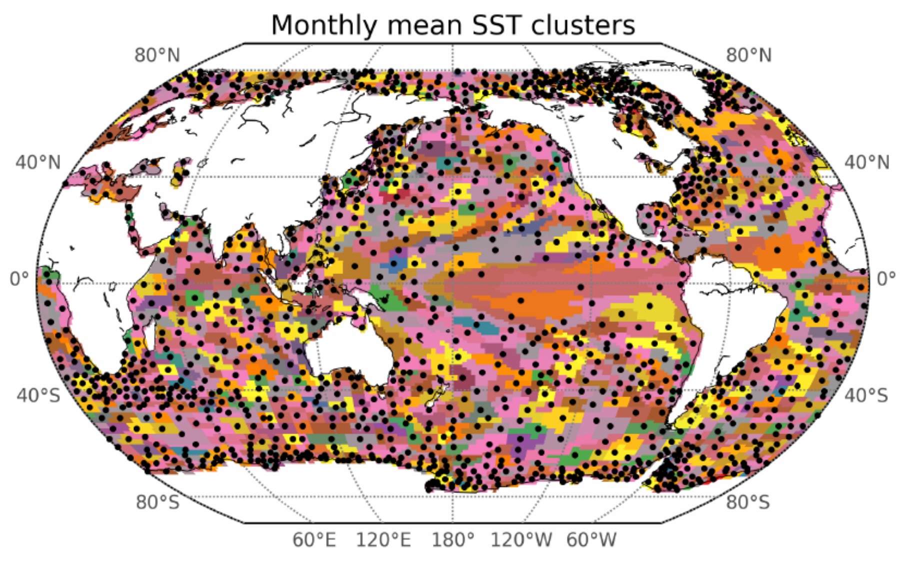
Bedartha Goswami

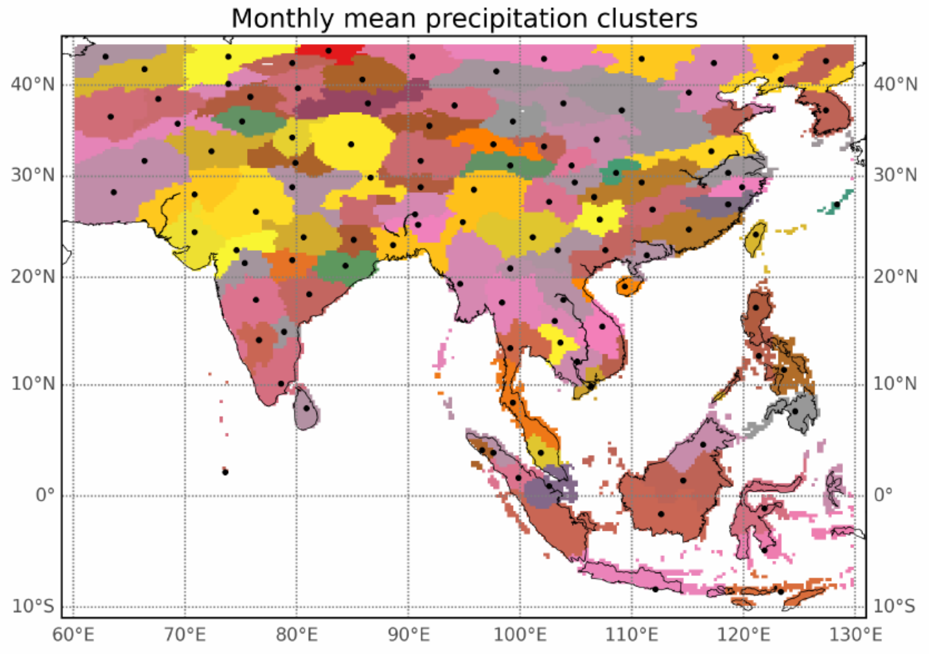
Dendrogram

- Visual representation of the merges
- Also shows the distance beteen clusters
- Can be helpful to decide an appropriate number of clusters







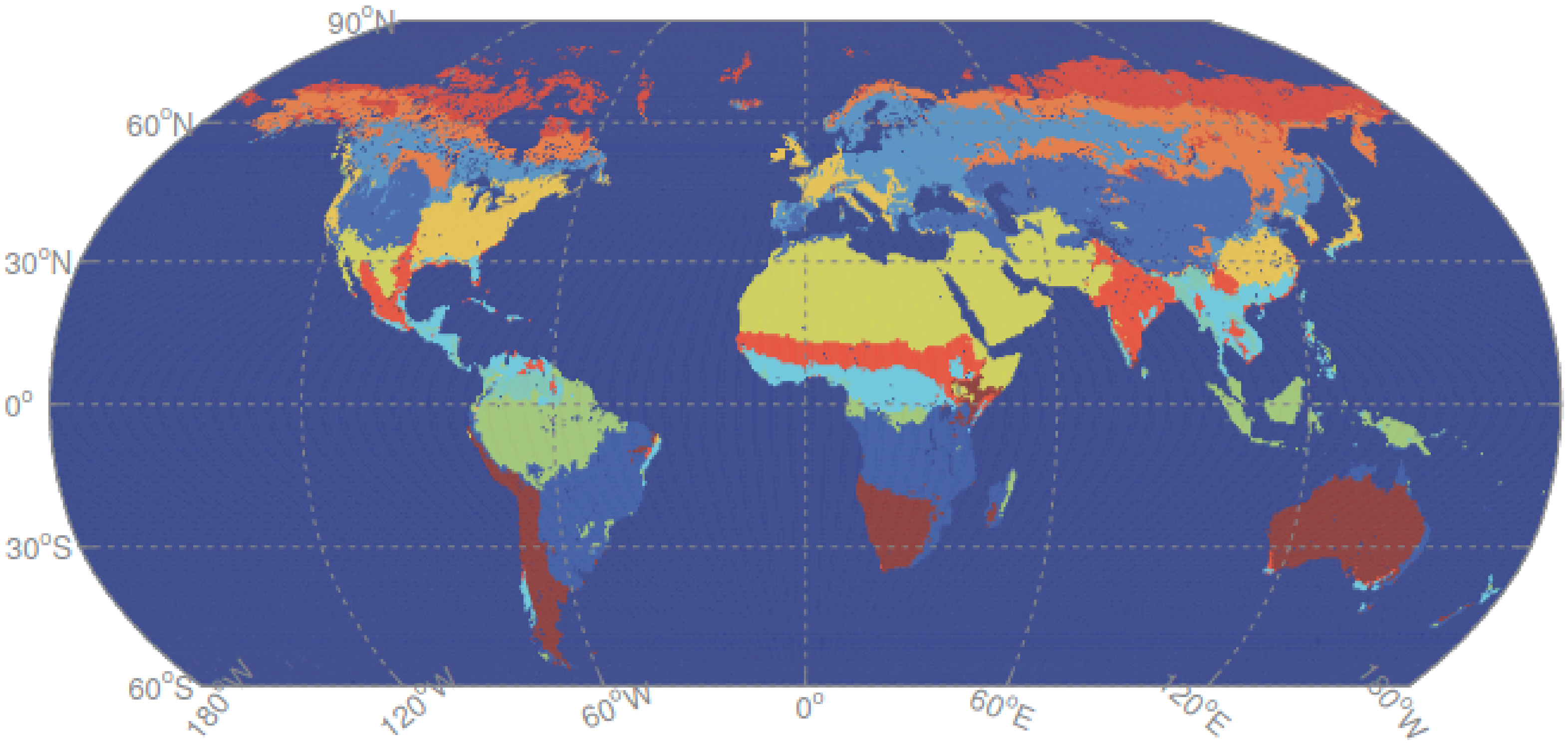


Monthly mean data, ar-coss distance metric based on Spearman's correlation, complete linkage clustering

2. Cluster Analysis → Hierarchical clustering

K-means clustering

- Pre-specify that we need *K* clusters
- Start with a rangom grouping of nodes into *K* clusters
- Iterate over nodes
  - At node $\boldsymbol{x}_i$, compute the distance from that node to the centroids and find the cluster which is closest
  - If $\boldsymbol{x}_i$ belongs to that cluster already, move to the next node
  - Else, assign $\boldsymbol{x}_i$ to the closest cluster and move to next node
- Keep iterating until a full rotation over all nodes results in zero reassignments



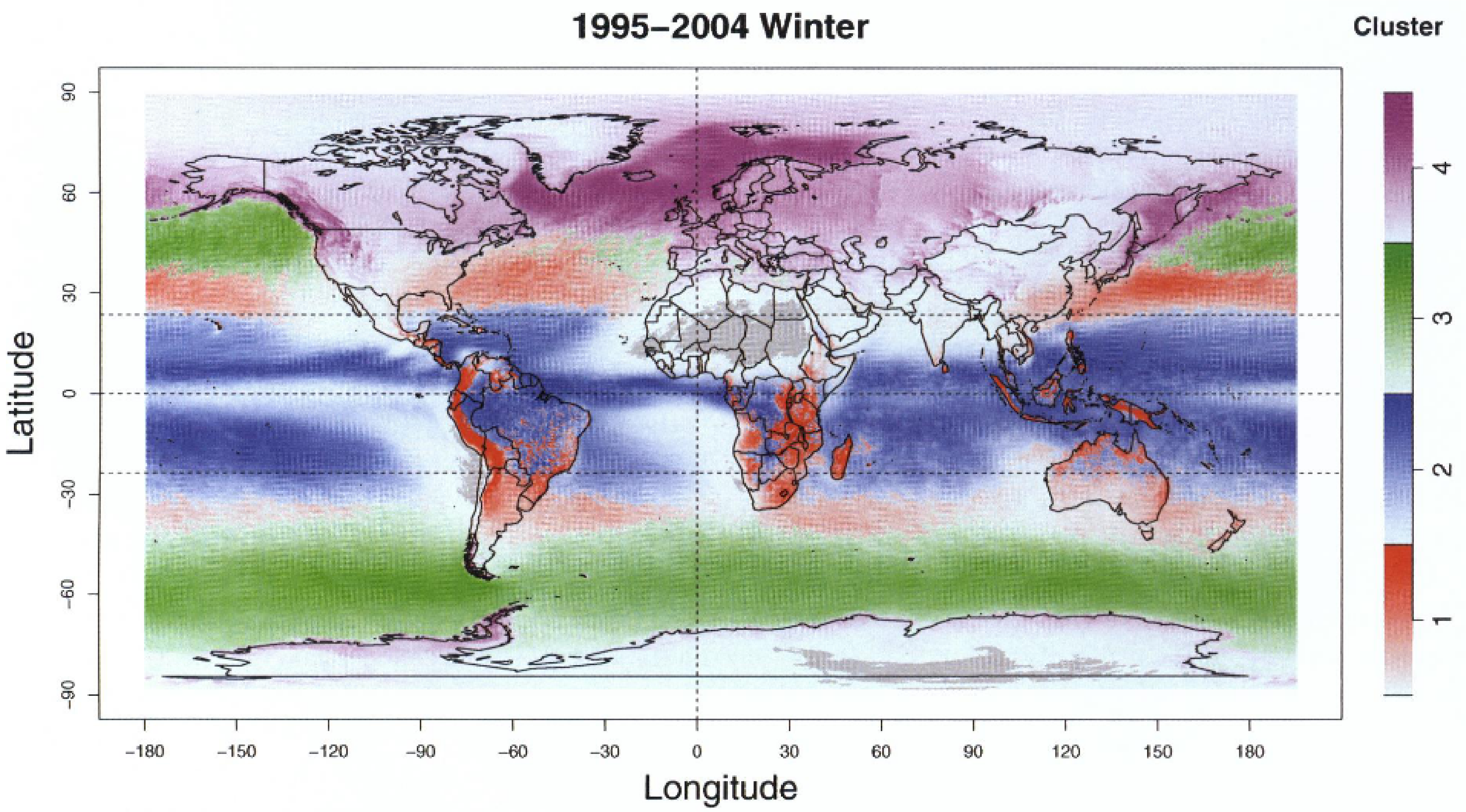
Map of *k*-means clustering withk=12 for the variables P, T, SW, EVI, FAPAR

2. Cluster Analysis → Non-Hierarchical clustering

Gaussian mixture model

- Model the data as belonging to a Gaussian mixture composed of *K* individual Gaussian distributions
- Use a Bayesian approach to determine the liklihood that a data point belongs to a particular component of the mixture
- Use the expectation minimization (EM) algorithm to find the best fit
- Result: Posterior probabilities denoting membership of each data point to each group (fuzzy clustering)
- *K* has to be pre-specified

`



4 clusters of extreme precipitation type, colored according to frequency of extreme events in winter, between 1995 - 2004

2. Cluster Analysis → Non-Hierarchical clustering

Empirical Orthogonal Functions

1

- What are EOFs?
- Considerations in estimating EOFs
- Examples

Cluster Analysis

2

- What is cluster analysis?
- Hierarchical clustering
- Non-hierarchical clustering

Q
A