Challenges in KDD and ML for Sustainable Development

Laure Berti-Equille

IRD ESPACE-DEV
Montpellier, France
laure.berti@ird.fr

David Dao

ETH Zürich,
Zürich, Switzerland
david.dao@inf.ethz.ch

Stefano Ermon

Stanford University
Stanford, CA, USA
ermon@cs.stanford.edu

Bedartha Goswami

University of Tübingen
Tübingen, Germany
bedartha.goswami@uni-tuebingen.de

ML-based climate data analytics

PART III

Bedartha Goswami


SUSTAINABLE DEVELOPMENT GOALS
1 NO POVERTY
2 ZERO HUNGER
3 GOOD HEALTH AND WELL-BEING
4 QUALITY EDUCATION
5 GENDER EQUALITY
6 CLEAN WATER AND SANITATION
7 AFFORDABLE AND CLEAN ENERGY
8 DECENT WORK AND ECONOMIC GROWTH
9 INDUSTRY, INNOVATION AND INFRASTRUCTURE
10 REDUCED INEQUALITIES
11 SUSTAINABLE CITIES AND COMMUNITIES
12 RESPONSIBLE CONSUMPTION AND PRODUCTION
13 CLIMATE ACTION
14 LIFE BELOW WATER
15 LIFE ON LAND
16 PEACE, JUSTICE AND STRONG INSTITUTIONS
17 PARTNERSHIPS FOR THE GOALS
SUSTAINABLE DEVELOPMENT GOALS

machine learning in climate science

EBERHARD KARLS UNIVERSITÄT TÜBINGEN

ML

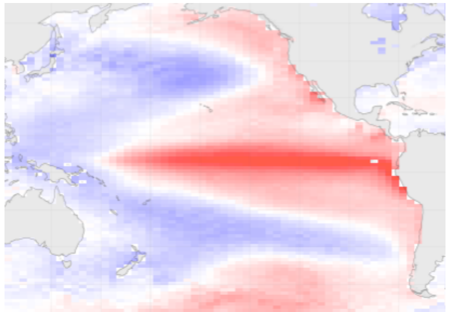
@bedartha

bedartha.goswami@uni-tuebingen.de

Outline

Outline

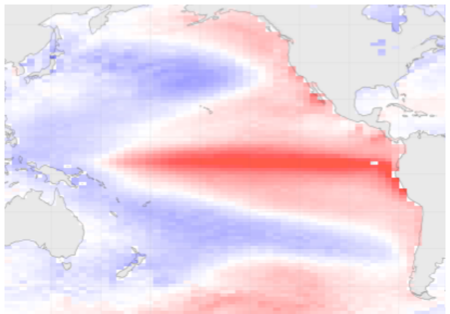
Part I Empirical orthogonal function analysis
→ principal components of the covariances in climate data

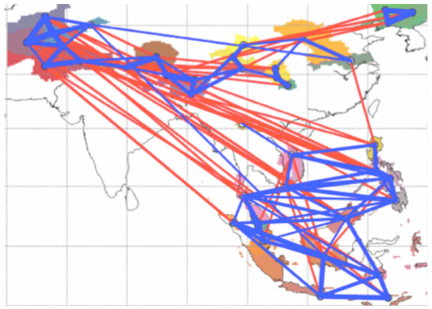


Outline

Part I Empirical orthogonal function analysis
→ principal components of the covariances in climate data

Part II Climate networks
→ interactions between climate data represented as graphs





Outline

Part I Empirical orthogonal function analysis

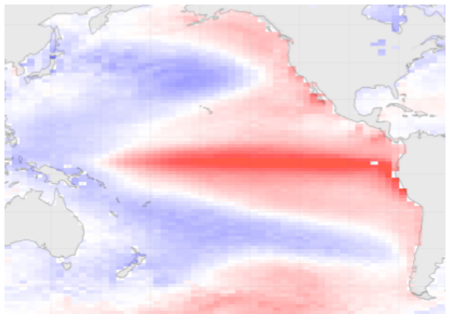
→ principal components of the covariances in climate data

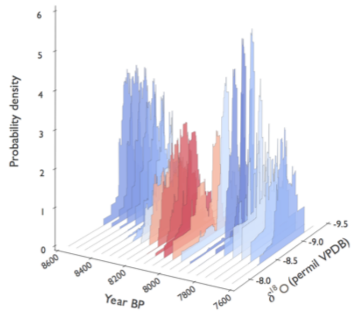
Part II Climate networks

→ interactions between climate data represented as graphs **Part III**

Climate data with uncertainties

→ detecting abrupt transitions in time series with uncertainties

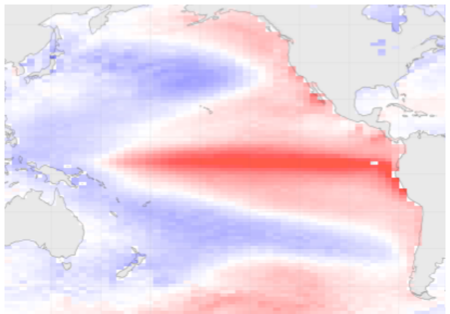


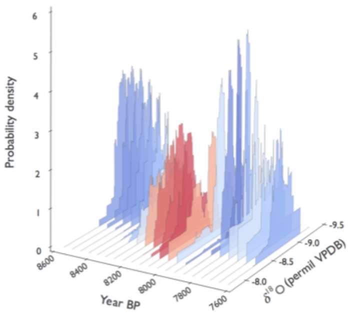


Outline

Part I Empirical orthogonal function analysis
→ principal components of the covariances in climate data **Part II**
Climate networks
→ interactions between climate data represented as graphs **Part III**
Climate data with uncertainties
→ detecting abrupt transitions in time series with uncertainties

Outlook Deep learning for uncovering climate patterns
→ graph neural networks to reveal climate interations





Outline

Part I Empirical orthogonal function analysis

→ principal components of the covariances in climate data **Part II**

Climate networks

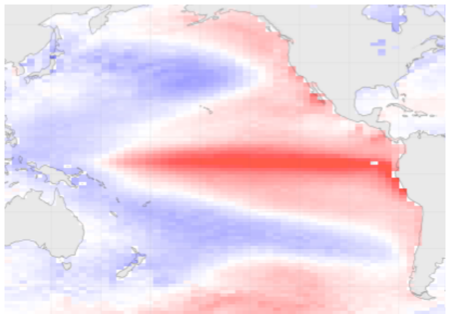
→ interactions between climate data represented as graphs **Part III**

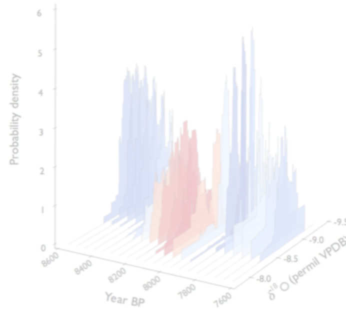
Climate data with uncertainties

→ detecting abrupt transitions in time series with uncertainties

Outlook Deep learning for uncovering climate patterns

→ graph neural networks to reveal climate interations





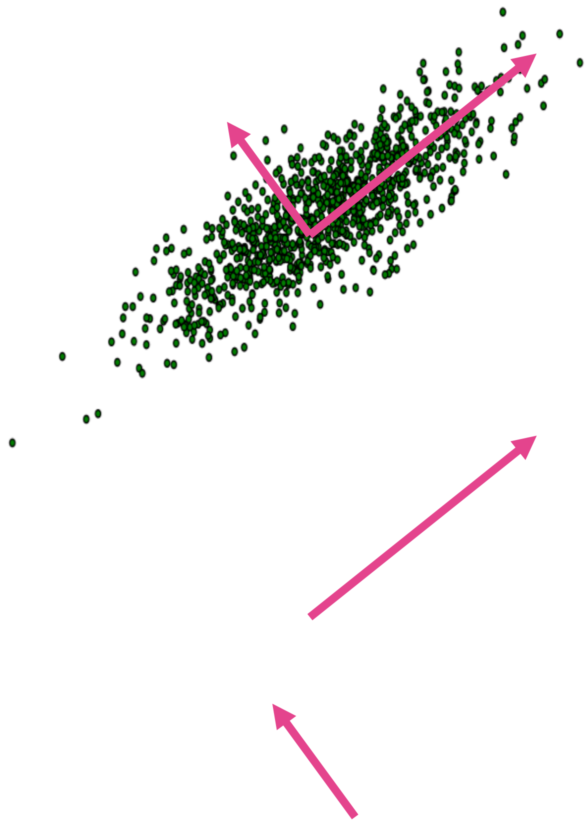
Empirical Orthogonal Functions (EOFs)

- Consider the data matrix
$$X = (\ x_1, x_2, \dots, x_p\ ),$$
- where p is the number of locations and
$$x_j = (x_{1j}, x_{2j}, \dots, x_{nj})^T$$
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



Empirical Orthogonal Functions (EOFs)

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

- For a single time instant,

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

- That is, for the entire time series

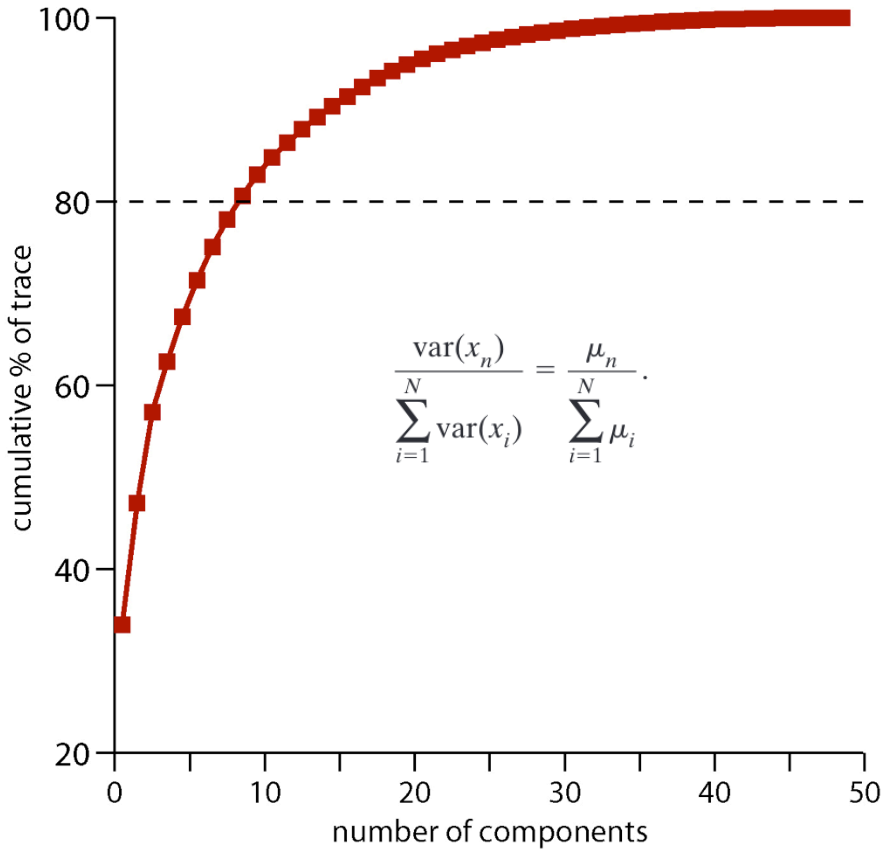
- $\boldsymbol{y}^1 = \mathbf{X}\ \boldsymbol{e}_1$

- gives the EOF time series for the first eigen direction

Considerations in estimating EOFs

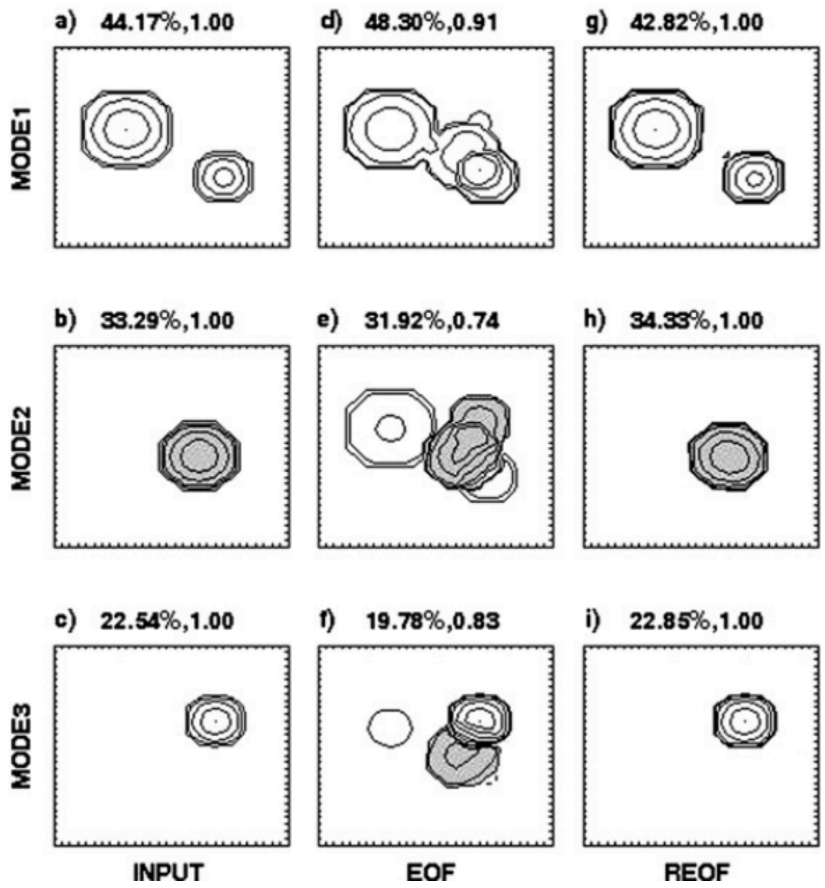
Considerations in estimating EOFs

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



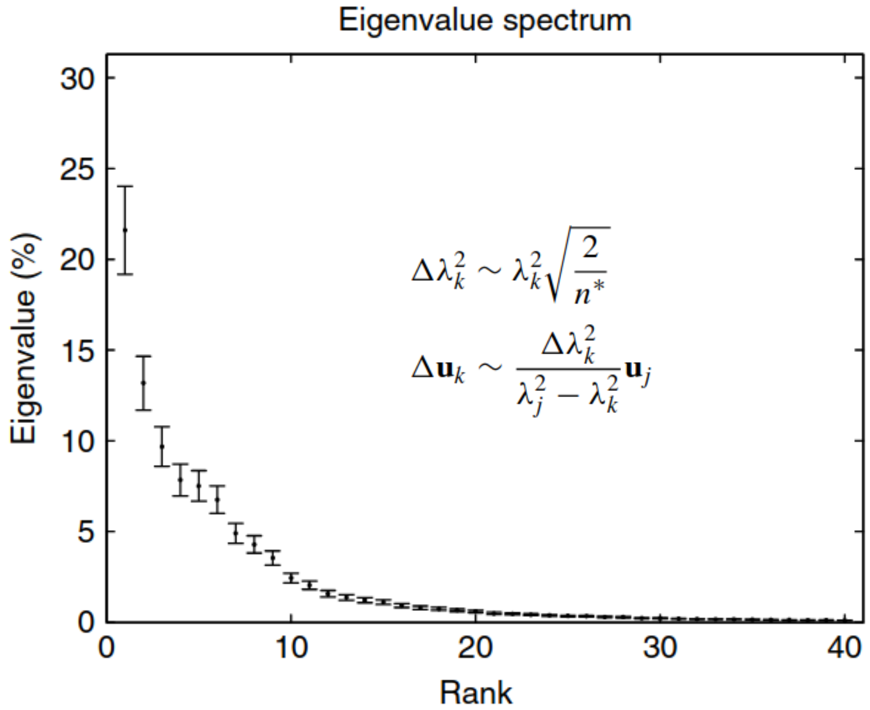
Considerations in estimating EOFs

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



Considerations in estimating EOFs

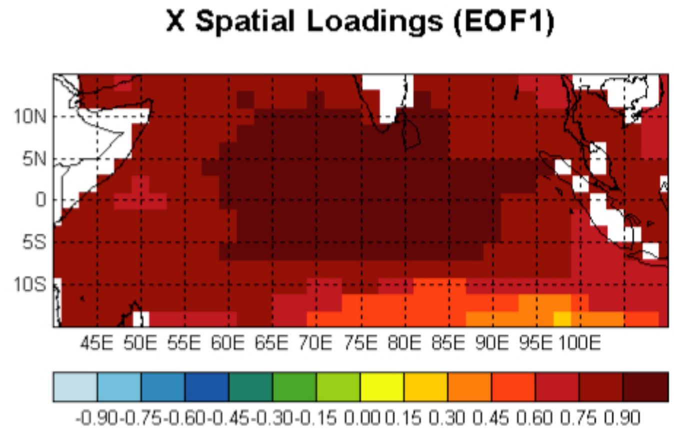
- **Truncation:**
  - Use only k << p leading eigenvalues
- **Rotation:**
  - Obtain 'simple' structures
  - Minimally overlapping EOFs
  - Ease of interpretation
- **Uncertainty:**
  - North's rule of thumb (North et al., 1982)
  - Monte Carlo sampling

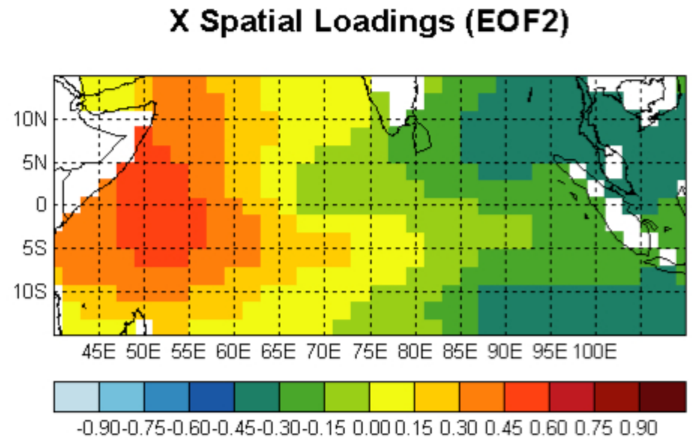


Spectrum, in percentage, of the covariance matrix of wintermonthly (DJF) SLP as per North's rule of thumb

Considerations in estimating EOFs

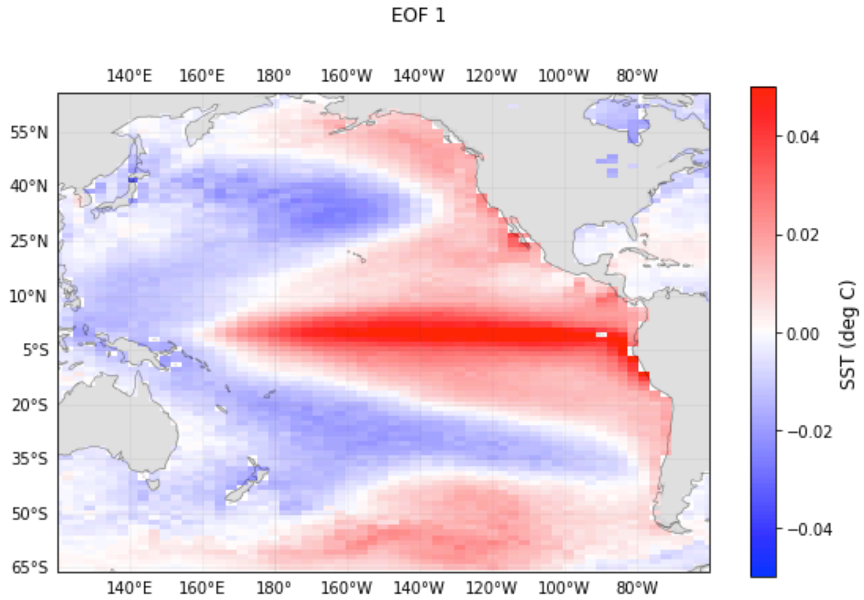
- **Truncation:**
  - Use only $k << p$ leading eigenvalues
- **Rotation:**
  - Obtain 'simple' structures
  - Minimally overlapping EOFs
  - Ease of interpretation
- **Uncertainty:**
  - North's rule of thumb (North et al., 1982)
  - Monte Carlo sampling
- **Spatial effects:**
  - For a rectangular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

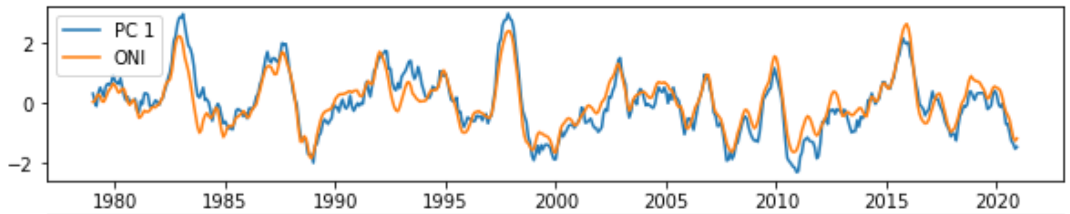




EOF example: Sea surface temperatures in the Pacific

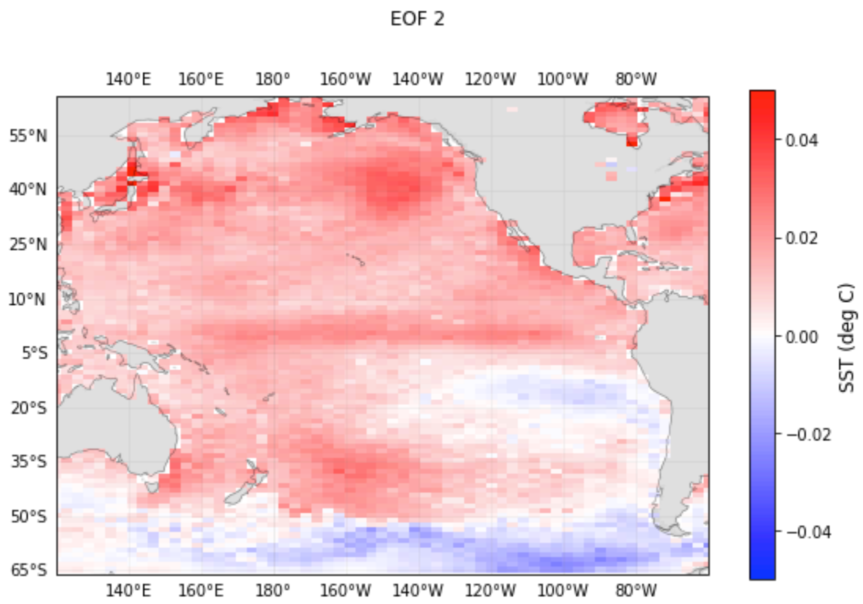
EOF example: Sea surface temperatures in the Pacific

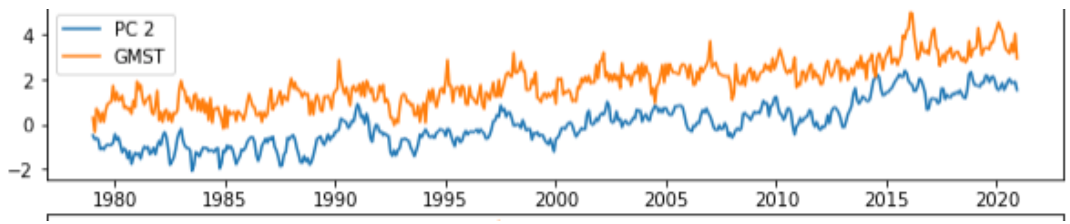




https://github.com/mlcs/mlcs.github.io/blob/master/files/sose2021/tutorial/pca_pacific.ipynb

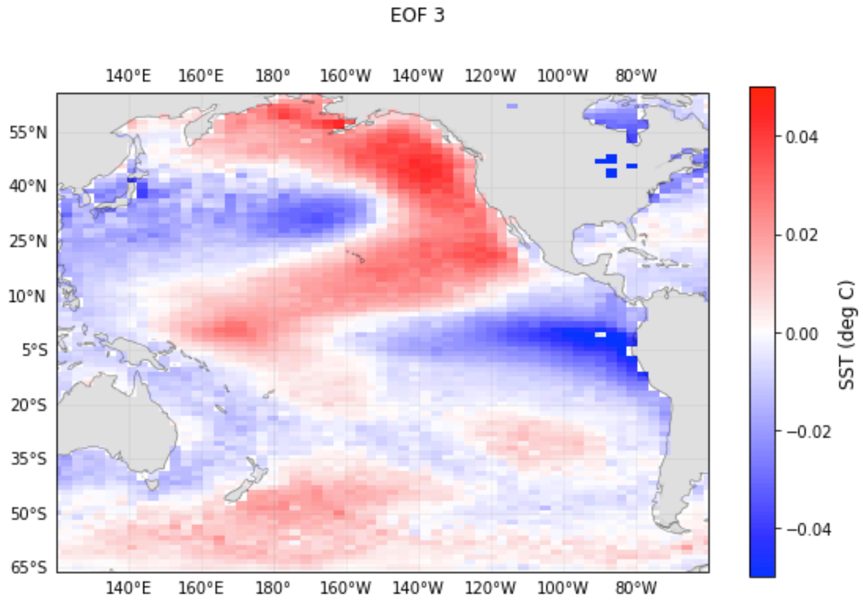
EOF example: Sea surface temperatures in the Pacific

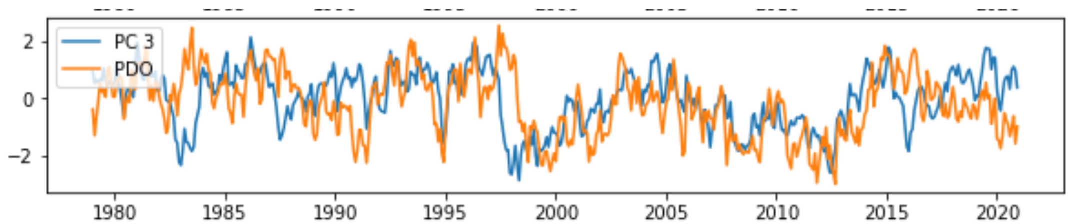




https://github.com/mlcs/mlcs.github.io/blob/master/files/sose2021/tutorial/pca_pacific.ipynb

EOF example: Sea surface temperatures in the Pacific





https://github.com/mlcs/mlcs.github.io/blob/master/files/sose2021/tutorial/pca_pacific.ipynb

Outline

Part I Empirical orthogonal function analysis

→ principal components of the covariances in climate data **Part II**

Climate networks

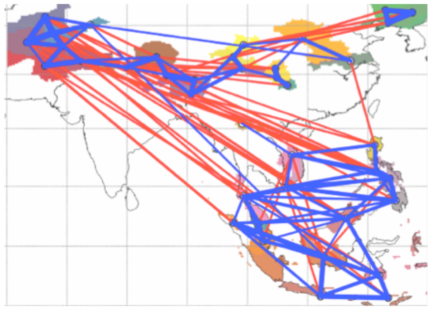
→ interactions between climate data represented as graphs **Part III**

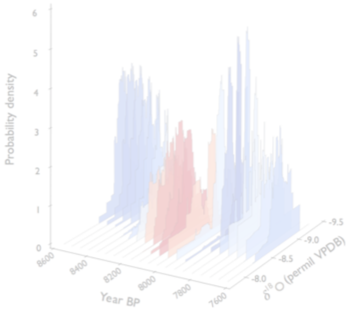
Climate data with uncertainties

→ detecting abrupt transitions in time series with uncertainties

Outlook Deep learning for uncovering climate patterns

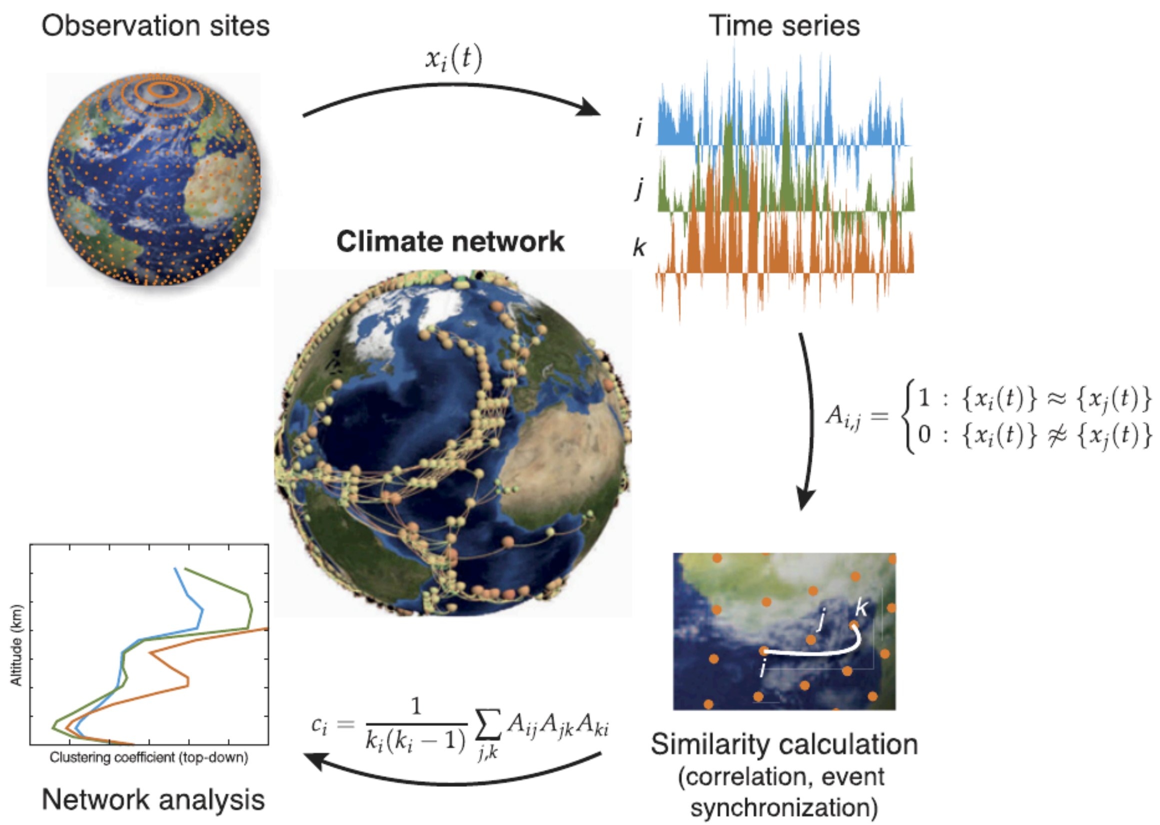
→ graph neural networks to reveal climate interations





Boers et al., J. Phys. Complex., 2021

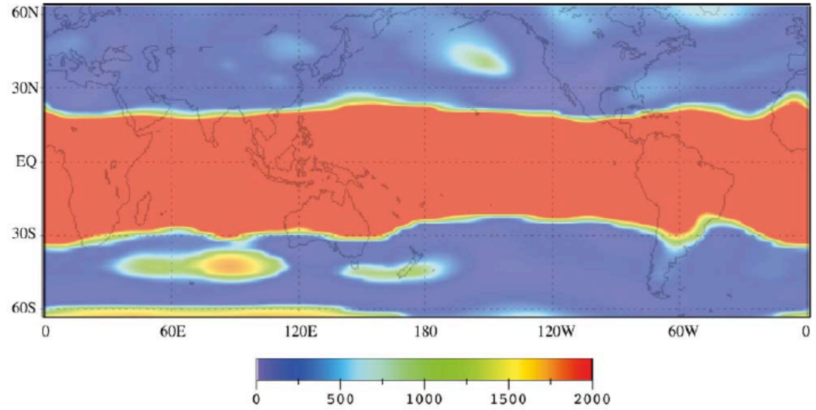
What are climate networks?



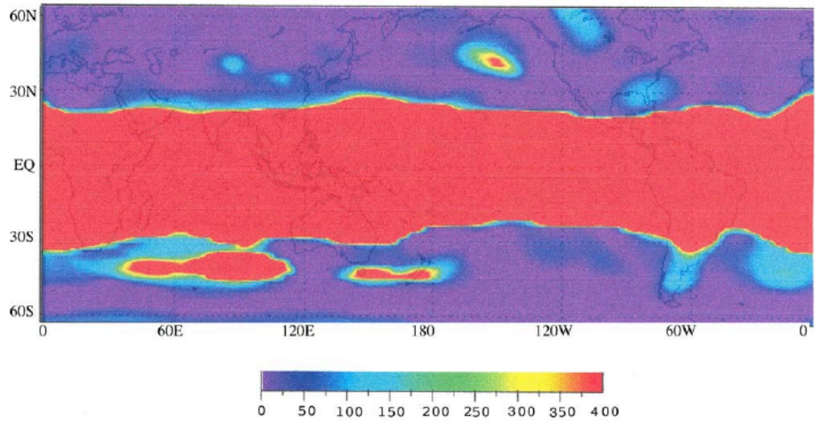
What do climate networks tell us?

What do climate networks tell us?

Monthly Z500, 5 deg lat-lon grid, correlation coeff., 1% significance (threshold rho = 0.5)



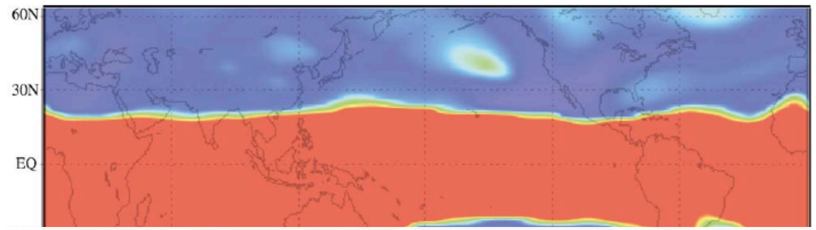
Degree



Degree > 5000 km

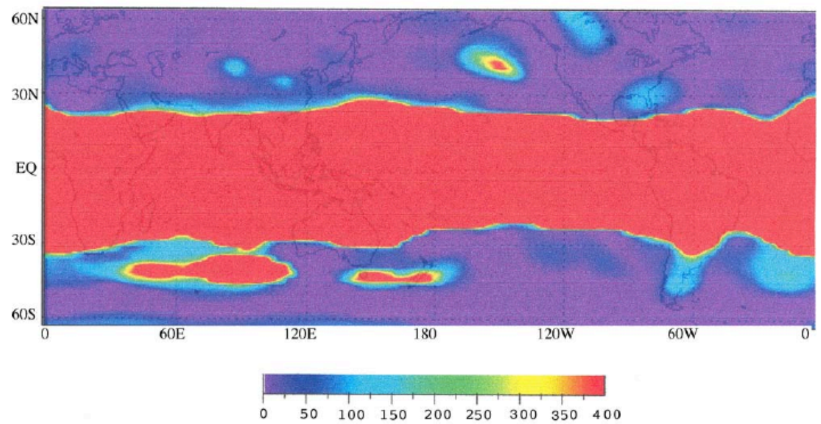
What do climate networks tell us?

Monthly Z500, 5 deg lat-lon grid, correlation coeff., 1% significance (threshold rho = 0.5)



Degree

"The physical interpretation [...] is that the climate system exhibits properties of stable networks [...] where information is transferred efficiently [...] 'information' should be regarded as 'fluctuations' from any source ([e.g.] the tropics, El Niño, etc.). "

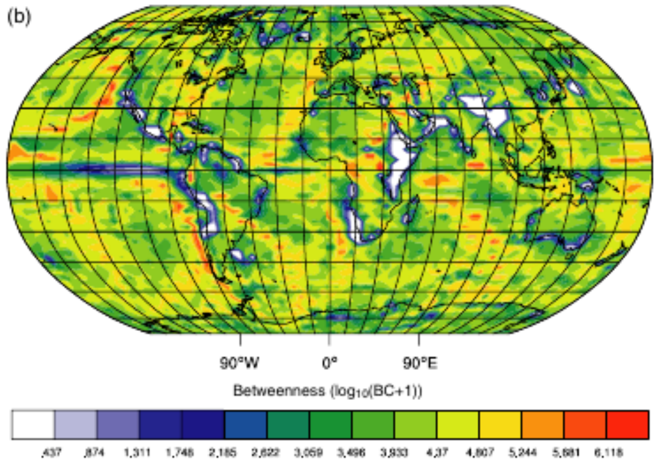


Degree > 5000 km

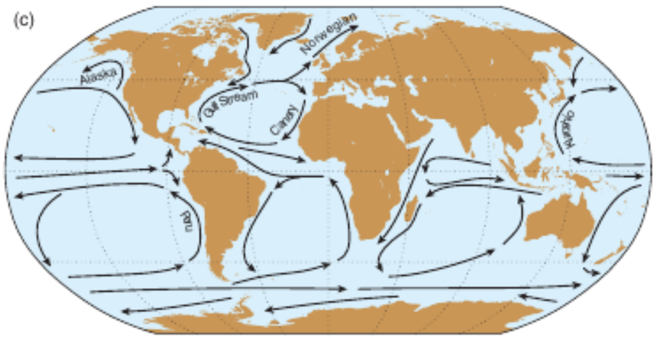
Donges et al., Europhys Lett, 2009

What do climate networks tell us?

Monthly SAT, 2.5 deg lat-lon grid, Mutual Info., link density 0.5 %



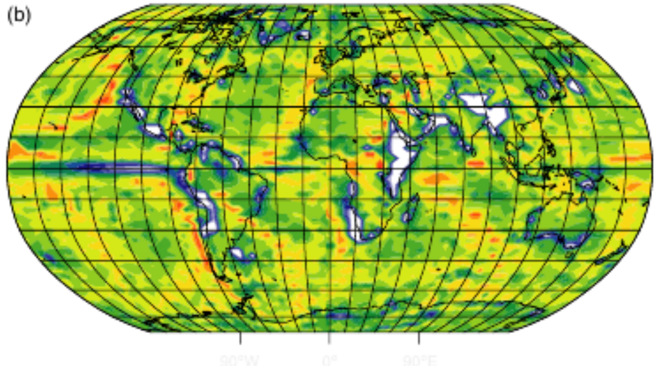
Shortest path betweenness



Ocean currents (schematic)

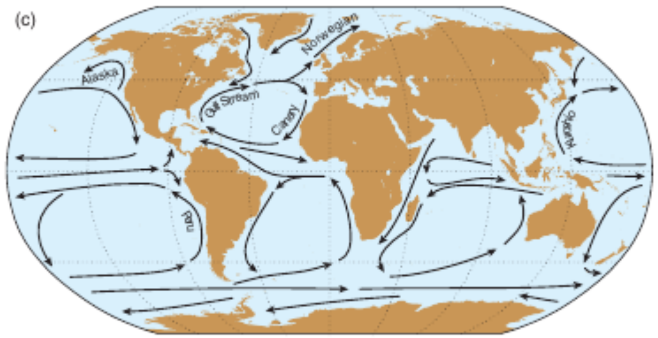
What do climate networks tell us?

Monthly SAT, 2.5 deg lat-lon grid, Mutual Info., link density 0.5 %



Shortest path betweenness

"In analogy with the internet, we call the network of these channels of high-energy flow the backbone of the climate network."

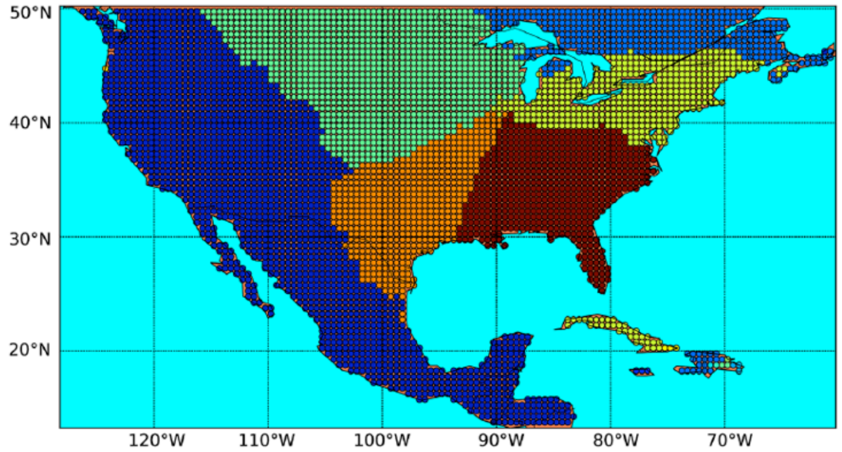


Ocean currents (schematic)

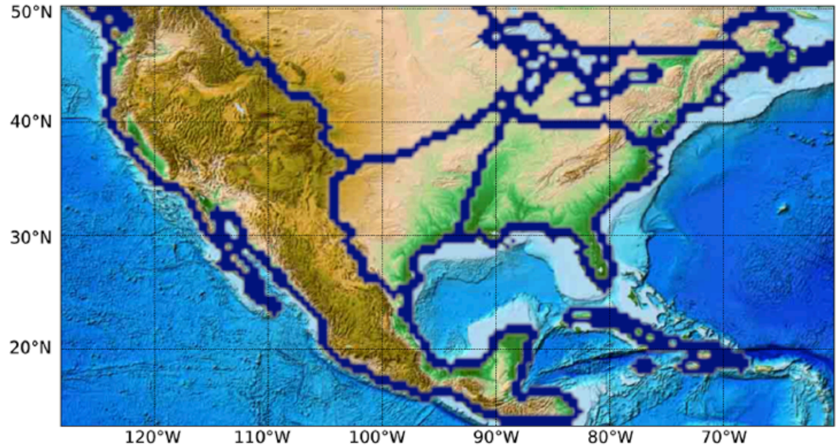
Peron et al., Nonlin. Proc. Geophys., 2014

What do climate networks tell us?

Monthly Surface T, 0.5 deg lat-lon grid, Cross correlation, link density 5 %



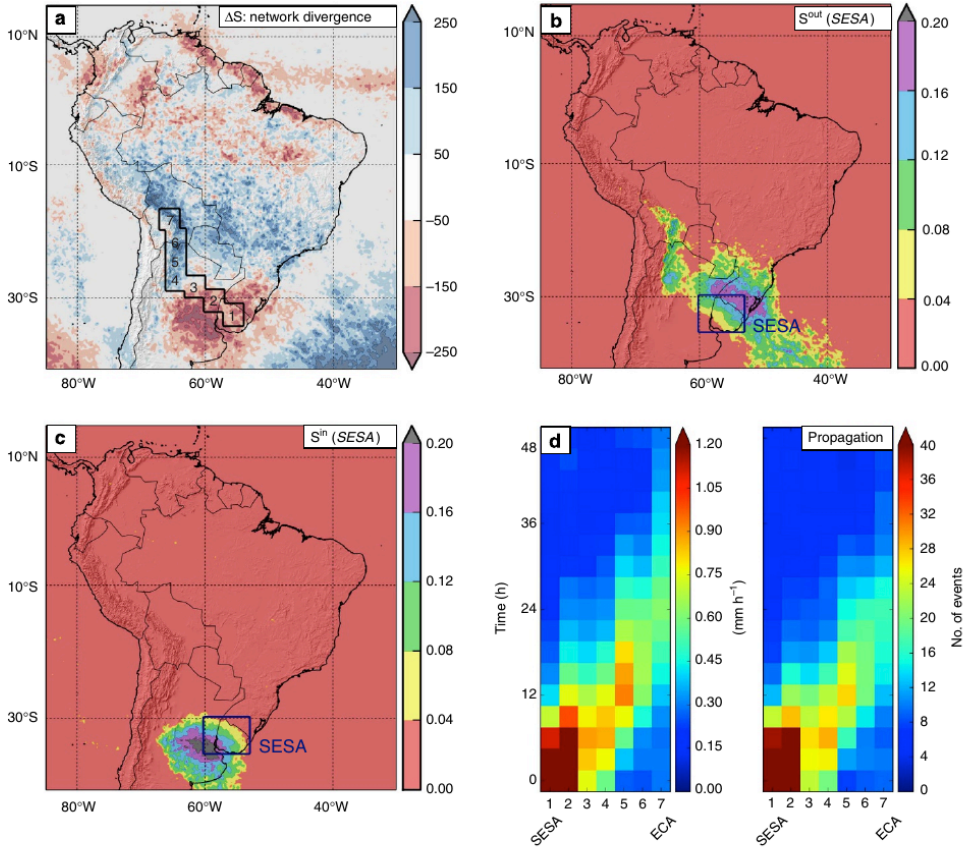
Network communities



Underlying topography

What do climate networks tell us?

3-hourly extreme rainfall, 0.25 deg lat-lon grid, Event Synchronisation, link density 2 %



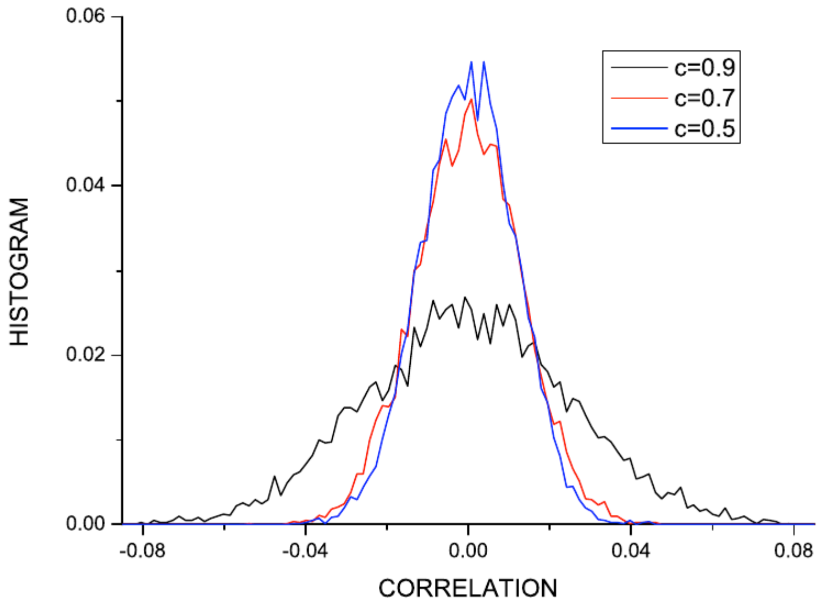
Outgoing links from SESA

Propagation of events

Potential pitfalls in climate networks

Palus et al., Nonlin. Proc. Geophys., 2011

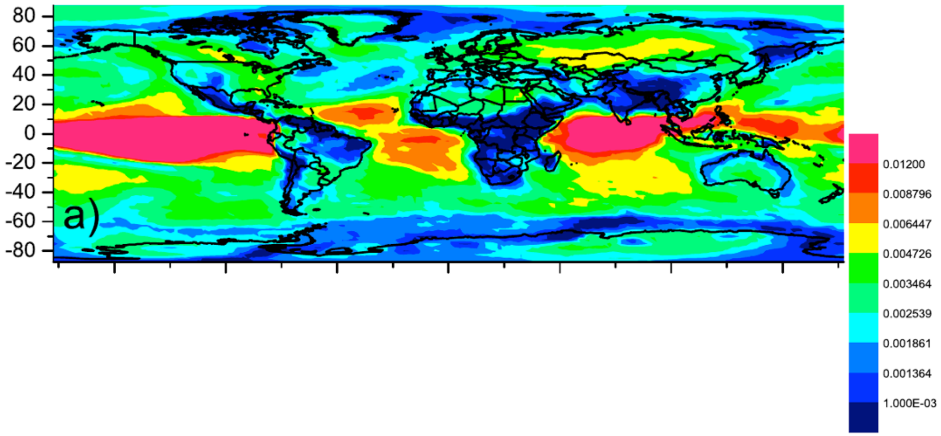
Potential pitfalls in climate networks: *Autocorrelation*



Potential pitfalls in climate networks: *Autocorrelation*

Monthly SAT, 2.5 deg lat-lon grid, absolute correlation coeff., link density 0.5 %

Area-weighted degree (original data)

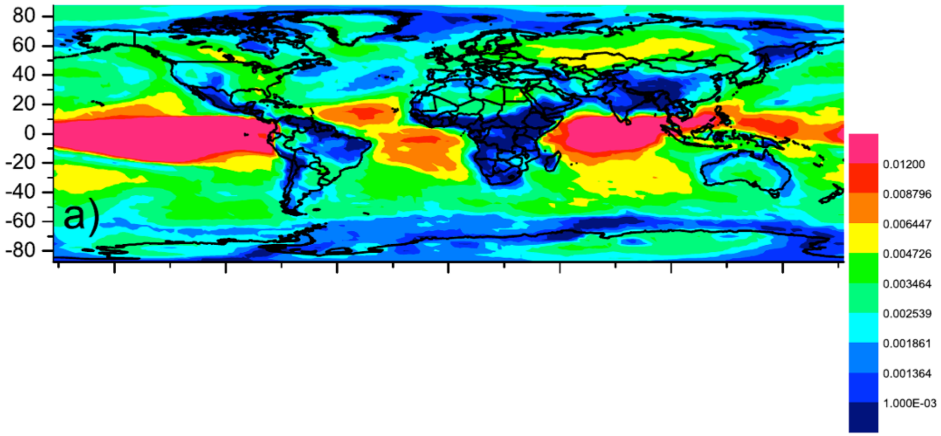


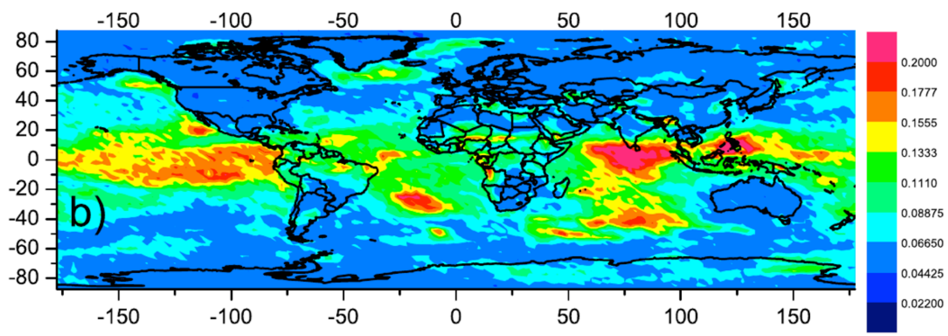
Palus et al., Nonlin. Proc. Geophys., 2011

Potential pitfalls in climate networks: *Autocorrelation*

Monthly SAT, 2.5 deg lat-lon grid, absolute correlation coeff., link density 0.5 %

Area-weighted degree (original data)





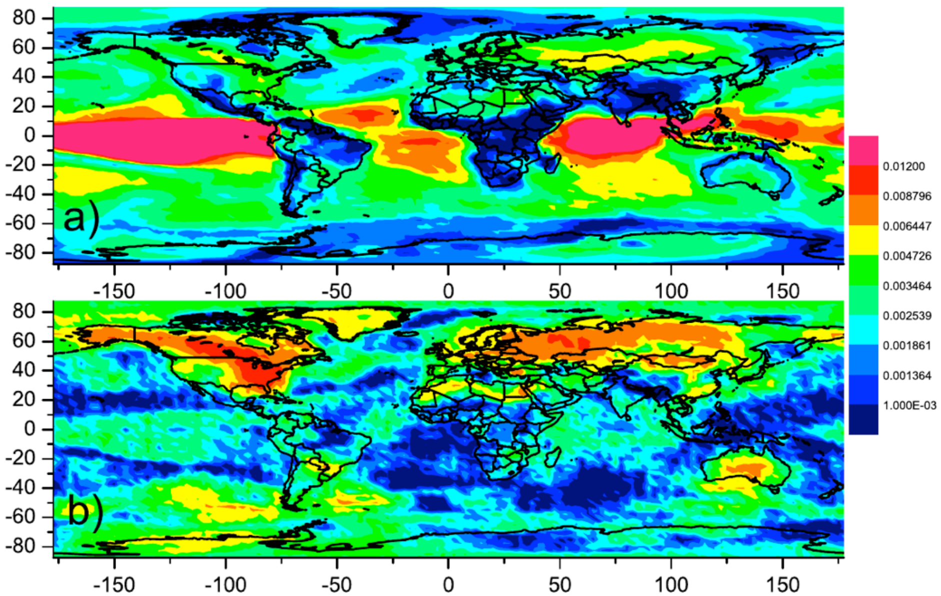
Area-weighted degree (randomised data)

Palus et al., Nonlin. Proc. Geophys., 2011

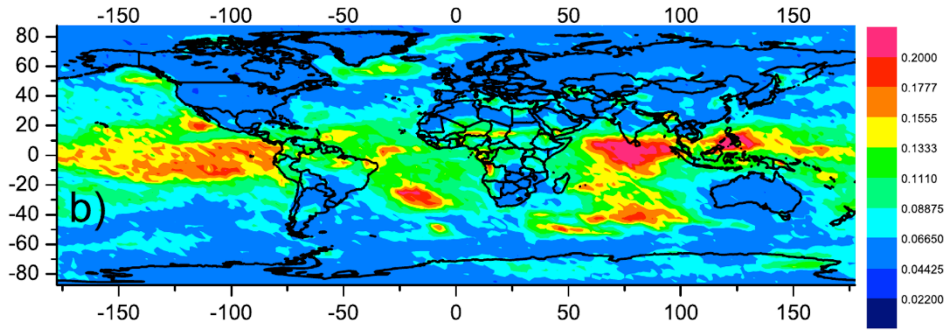
Potential pitfalls in climate networks: *Autocorrelation*

Monthly SAT, 2.5 deg lat-lon grid, absolute correlation coeff., link density 0.5 %

Area-weighted degree (original data)



Z-score of area-weighted degree (original data)
(w.r.t. the distribution from randomised data)



Area-weighted degree (randomised data)

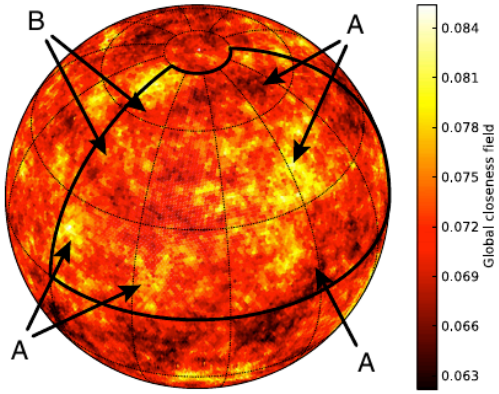
Potential pitfalls in climate network: *Boundary effects*

Rheinwalt et al., Europhys Lett., 2012

Potential pitfalls in climate network: *Boundary effects*

Closeness centrality on a random network on a sphere

Closeness centrality
(Full data)

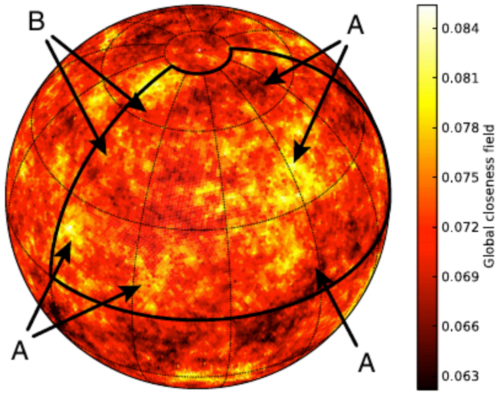


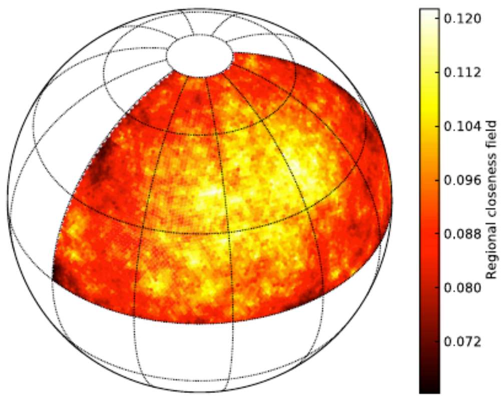
Rheinwalt et al., Europhys Lett., 2012

Potential pitfalls in climate network: *Boundary effects*

Closeness centrality on a random network on a sphere

Closeness centrality
(Full data)





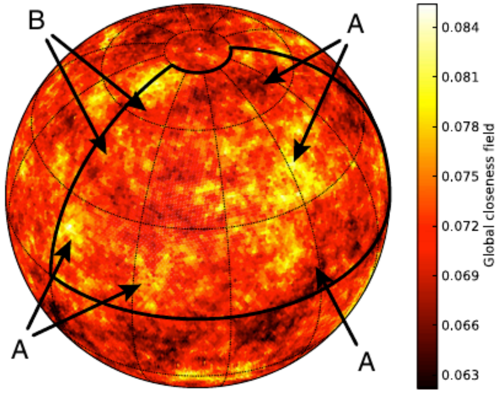
Closeness centrality
(boxed data)

Rheinwalt et al., Europhys Lett., 2012

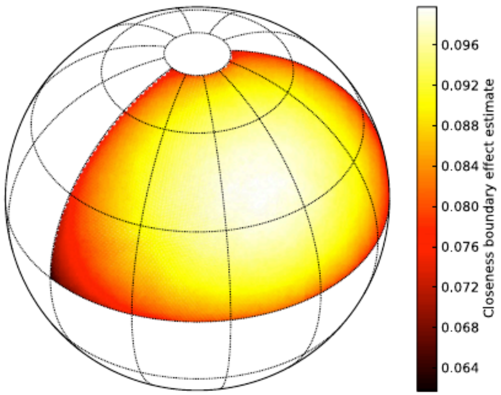
Potential pitfalls in climate network: *Boundary effects*

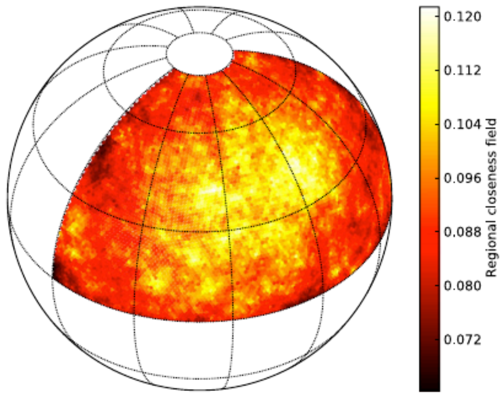
Closeness centrality on a random network on a sphere

Closeness centrality
(Full data)



Closeness centrality
(random model preserving link
length distribution)





Closeness centrality
(boxed data)

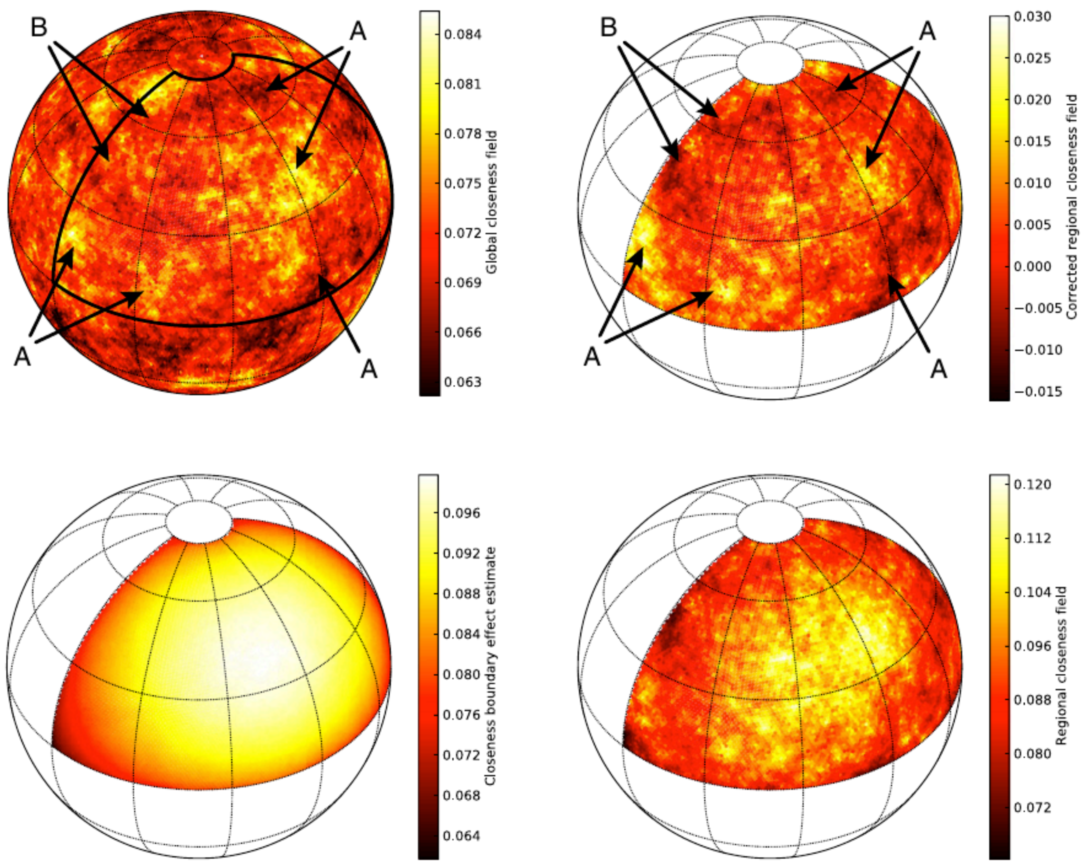
Rheinwalt et al., Europhys Lett., 2012

Potential pitfalls in climate network: *Boundary effects*

Closeness centrality on a random network on a sphere

Closeness centrality
(Full data)

Closeness centrality
(boxed data, corrected)



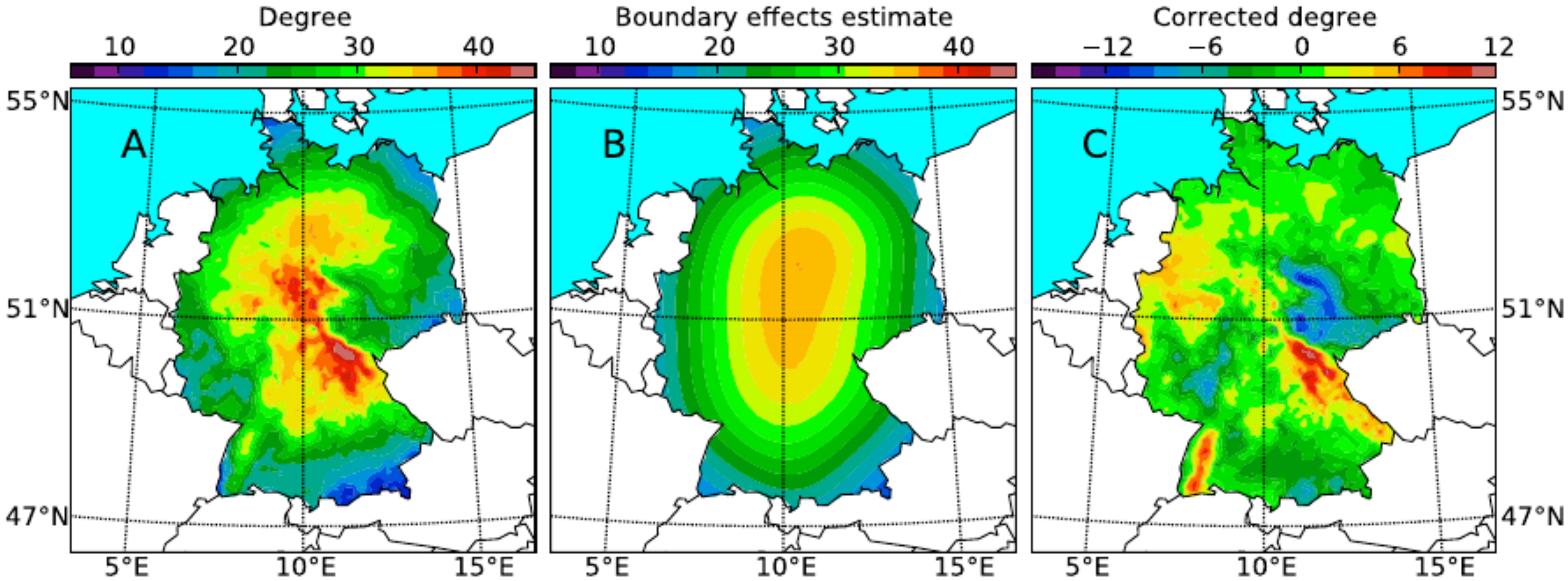
Closeness centrality
(random model preserving link
length distribution)

Closeness centrality
(boxed data)

Rheinwalt et al., Europhys Lett., 2012

Potential pitfalls in climate network: *Boundary effects*

Daily extreme rainfall, 2000 weather stations, Event Synchronisation, 50% link density

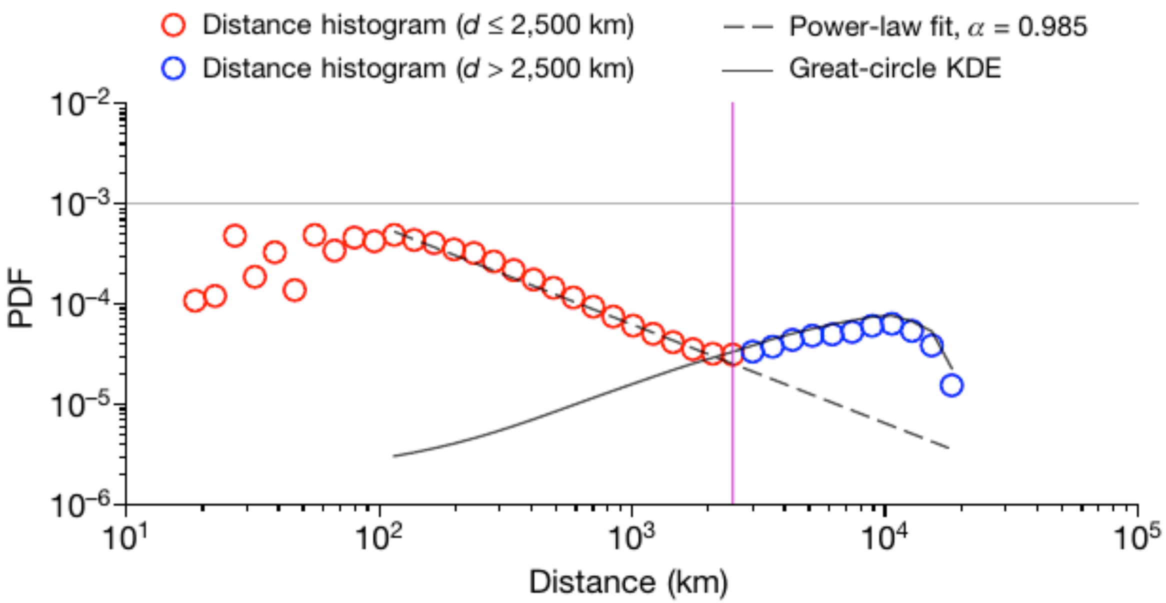


Potential pitfalls in climate network: *Spatial embedding*

Boers et al., Nature, 2019

Potential pitfalls in climate network: *Spatial embedding*

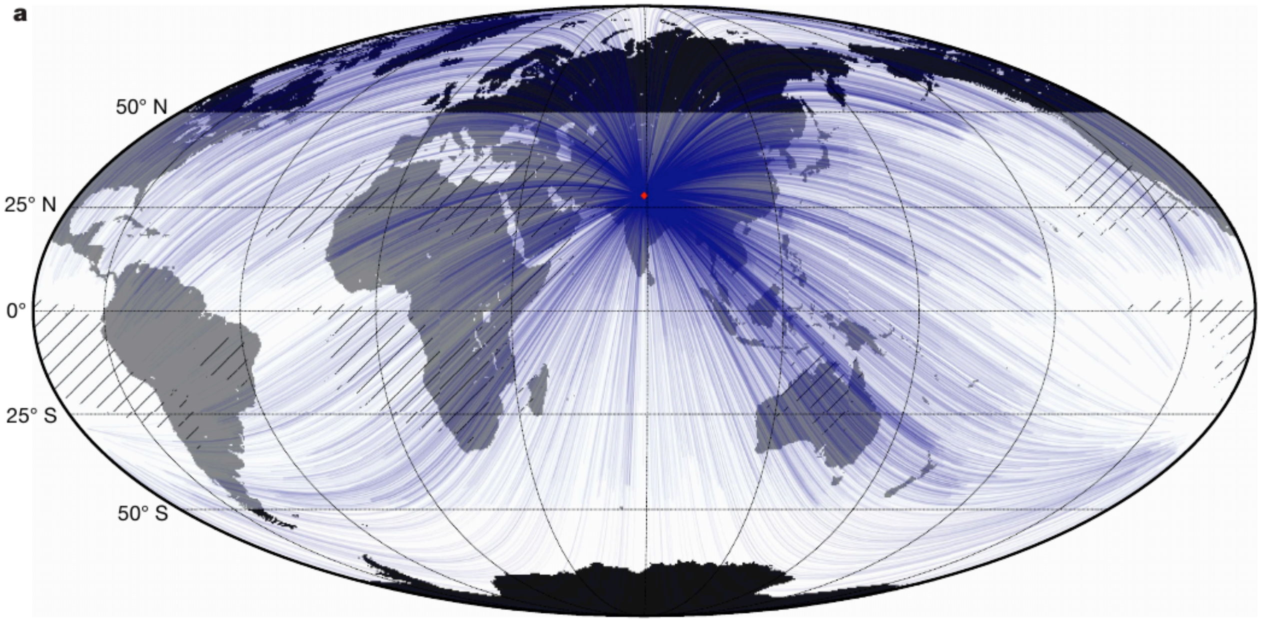
Daily extreme rainfall, 0.25 deg lat-lon grid, Event Synchronisation, 0.5% significance



Boers et al., Nature, 2019

Potential pitfalls in climate network: *Spatial embedding*

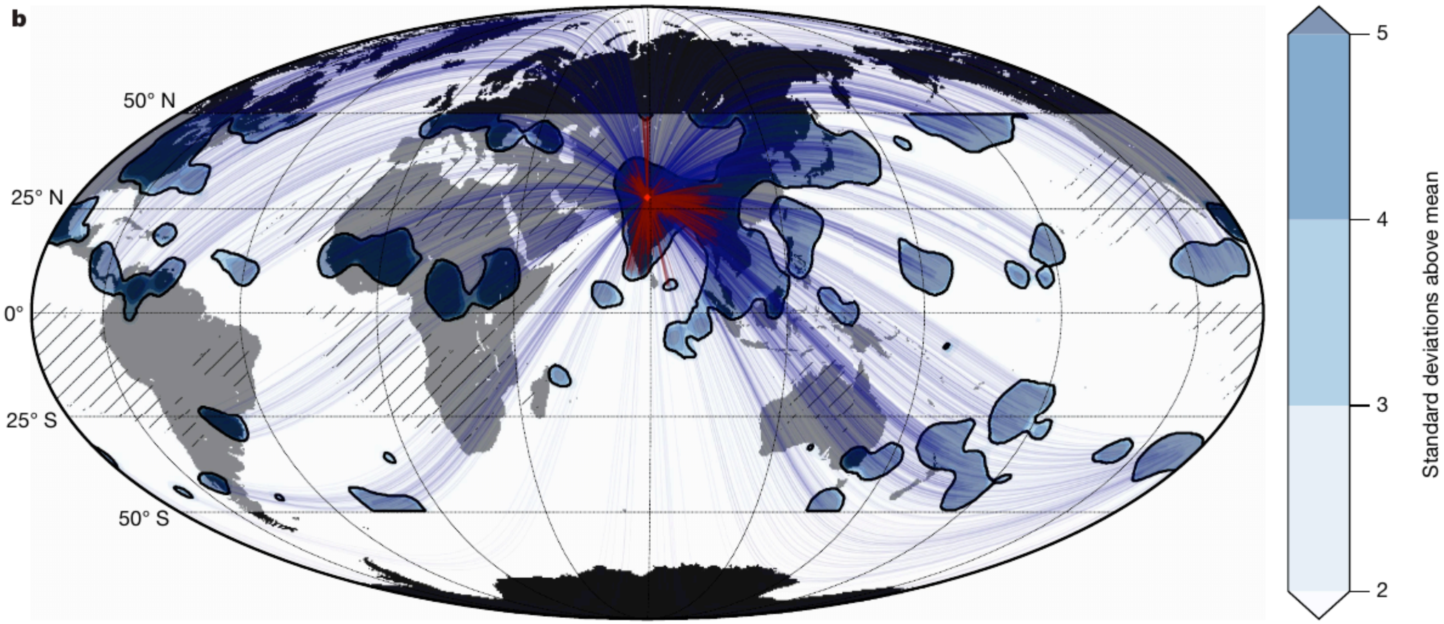
Daily extreme rainfall, 0.25 deg lat-lon grid, Event Synchronisation, 0.5% significance



Boers et al., Nature, 2019

Potential pitfalls in climate network: *Spatial embedding*

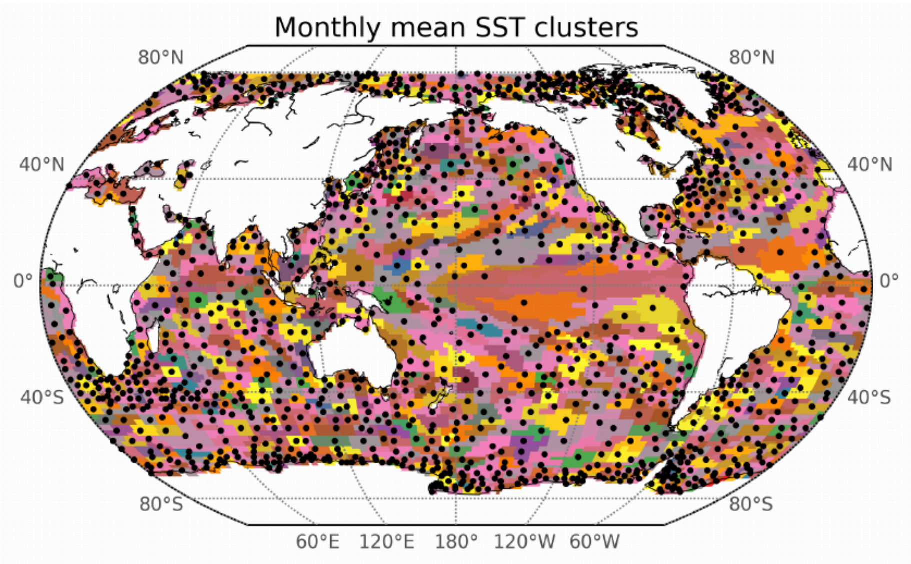
Daily extreme rainfall, 0.25 deg lat-lon grid, Event Synchronisation, 0.5% significance

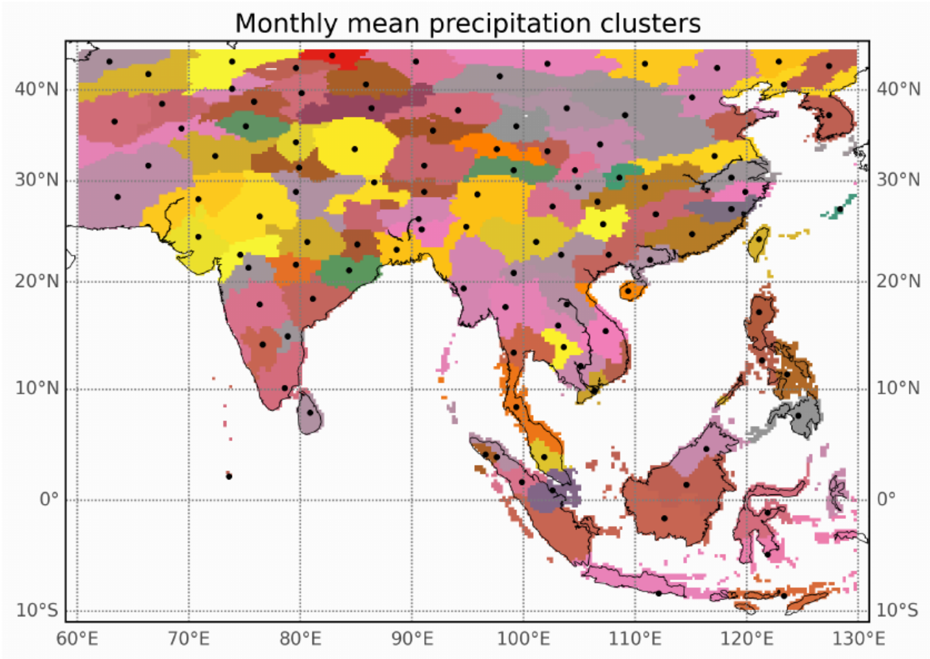


Teleconnection networks, or: a climate network of networks

Rheinwalt et a., Machine Learning and Data Mining Approaches to Climate Science (2015)

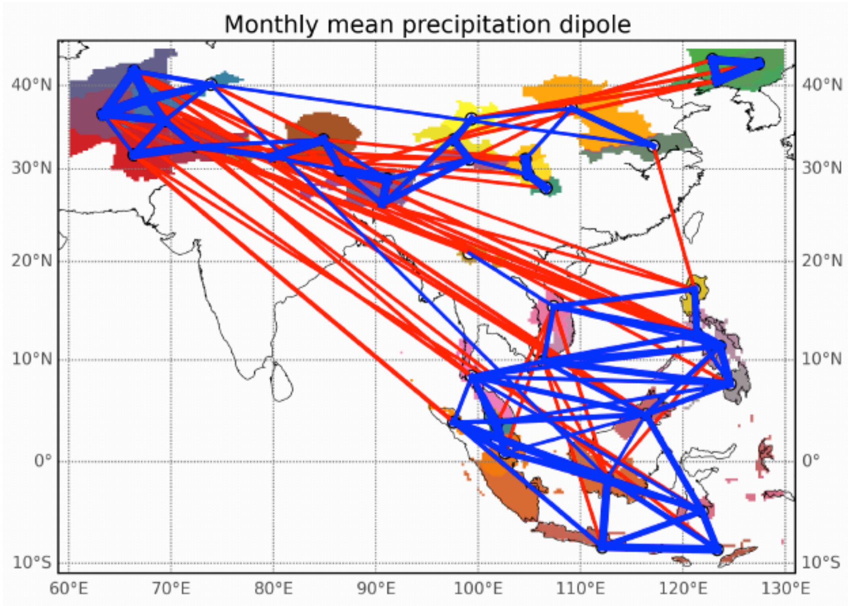
Teleconnection networks, or: a climate network of networks





Monthly mean data, arccos distance metric based on Spearman's correlation, complete linkage clustering

Teleconnection networks, or: a climate network of networks



Monthly mean data, arccos distance metric based on Spearman's correlation, complete linkage clustering

Outline

Part I Empirical orthogonal function analysis

→ principal components of the covariances in climate data **Part II**

Climate networks

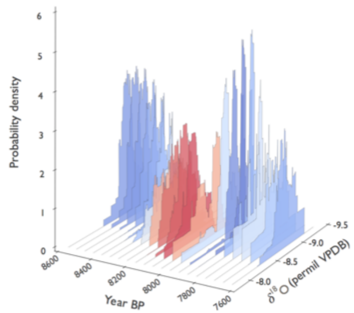
→ interactions between climate data represented as graphs **Part III**

Climate data with uncertainties

→ detecting abrupt transitions in time series with uncertainties

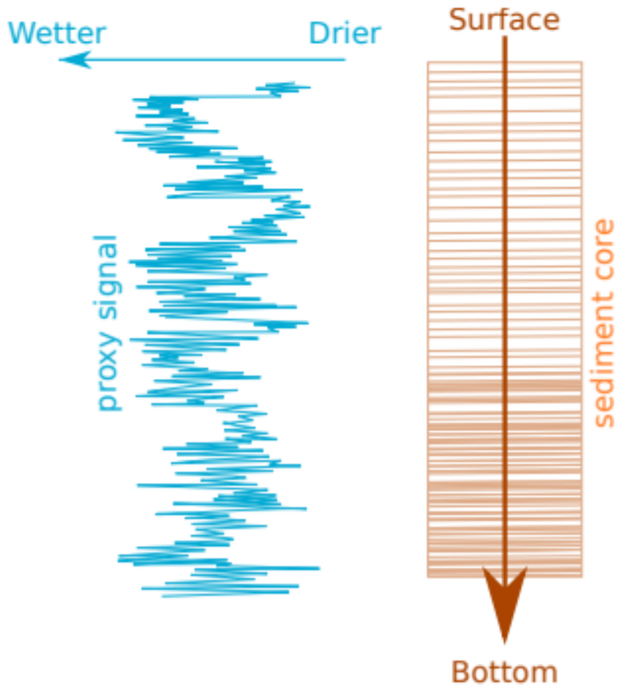
Outlook Deep learning for uncovering climate patterns

→ graph neural networks to reveal climate interations

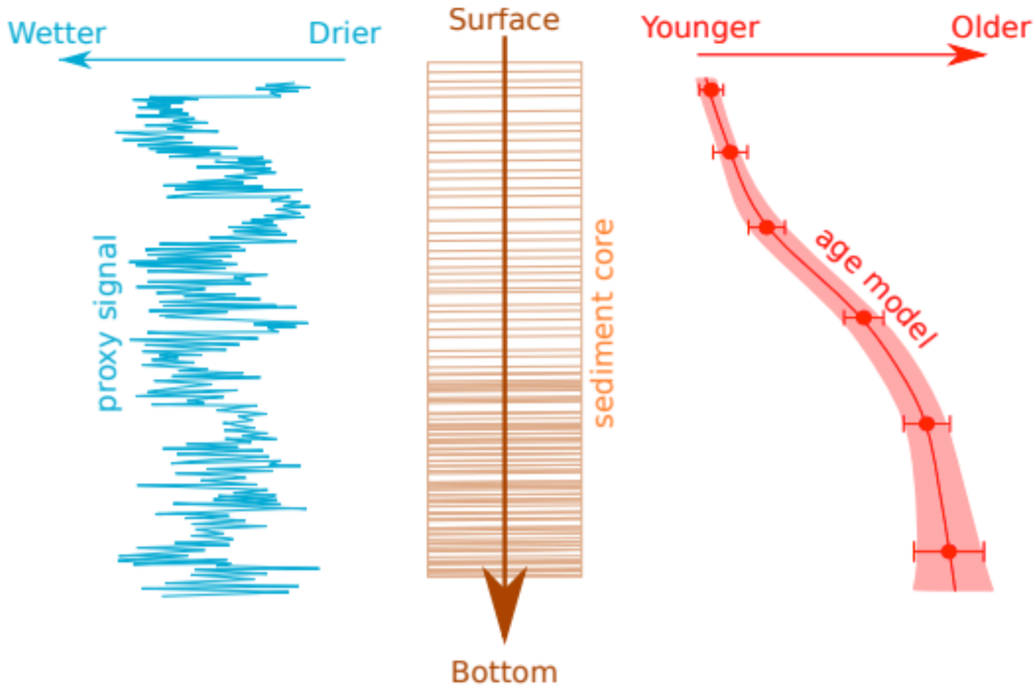


Uncertainties in paleoclimate proxy time series

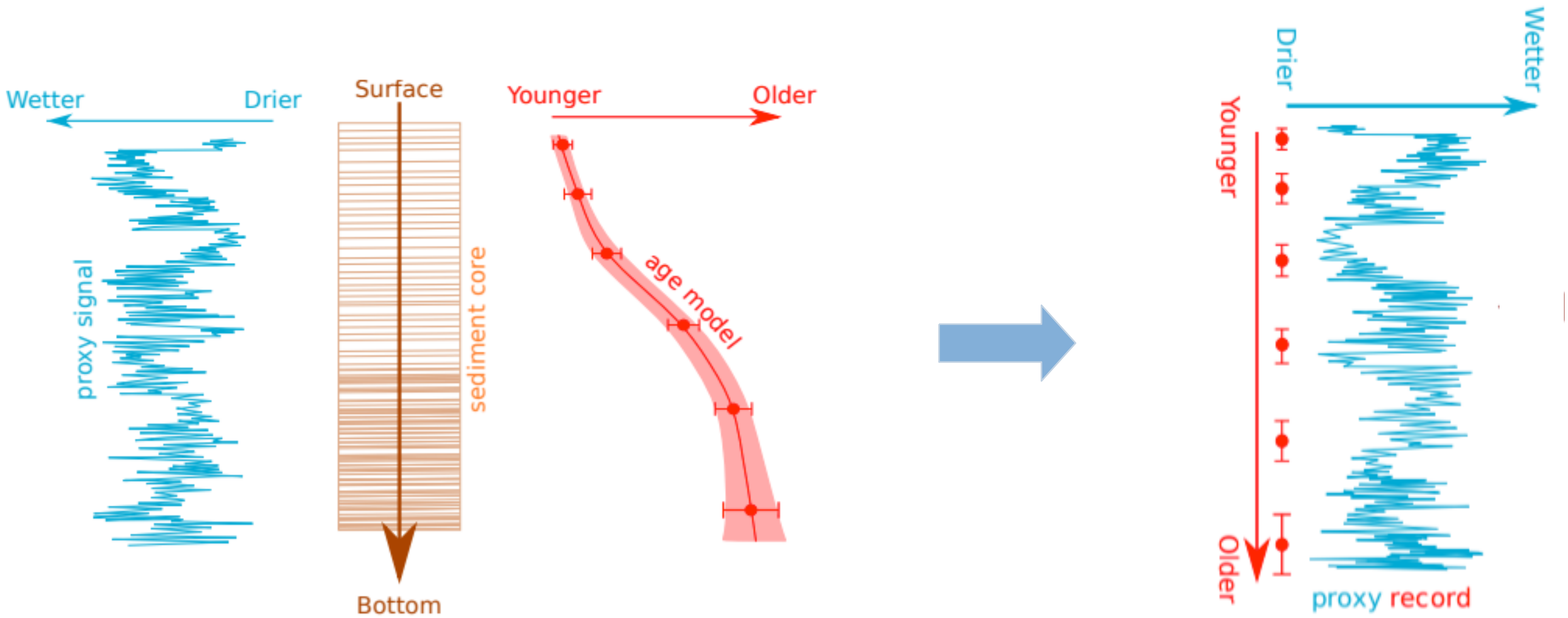
Uncertainties in paleoclimate proxy time series



Uncertainties in paleoclimate proxy time series

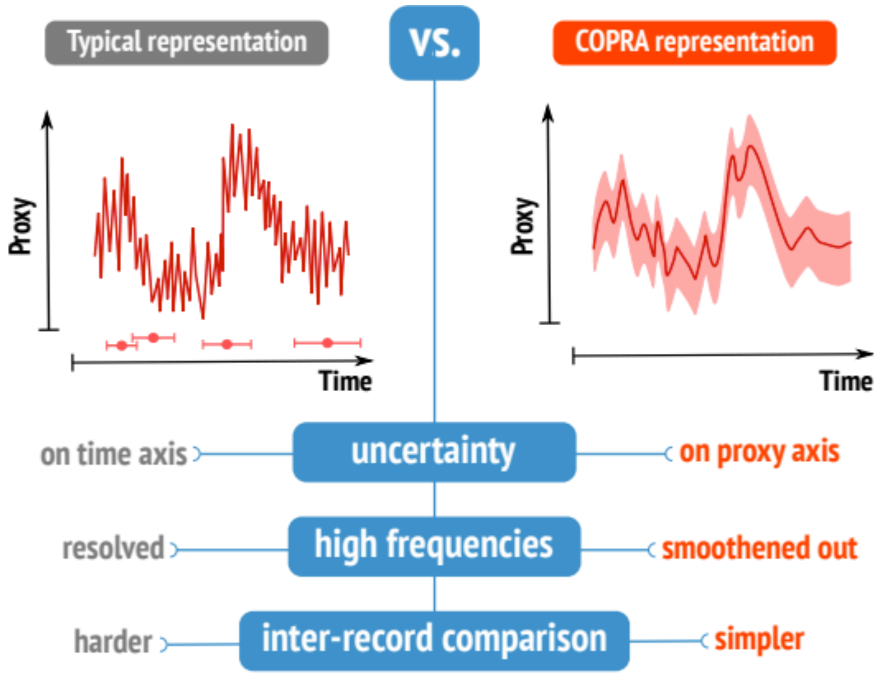


Uncertainties in paleoclimate proxy time series



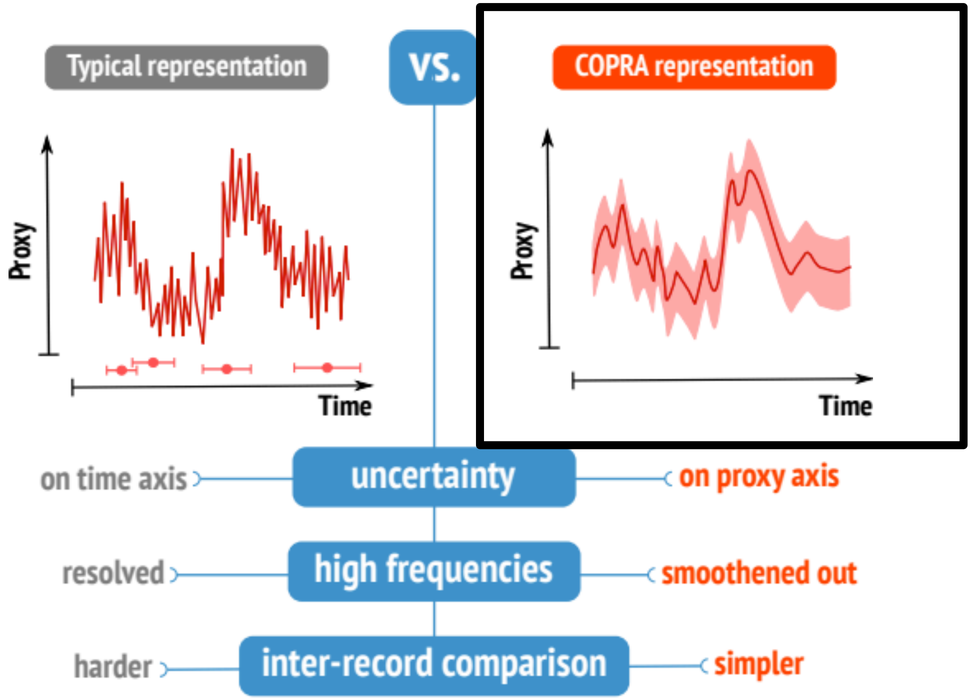
Breitenbach et al., *Clim. Past*, 2012

Uncertainties in paleoclimate proxy time series



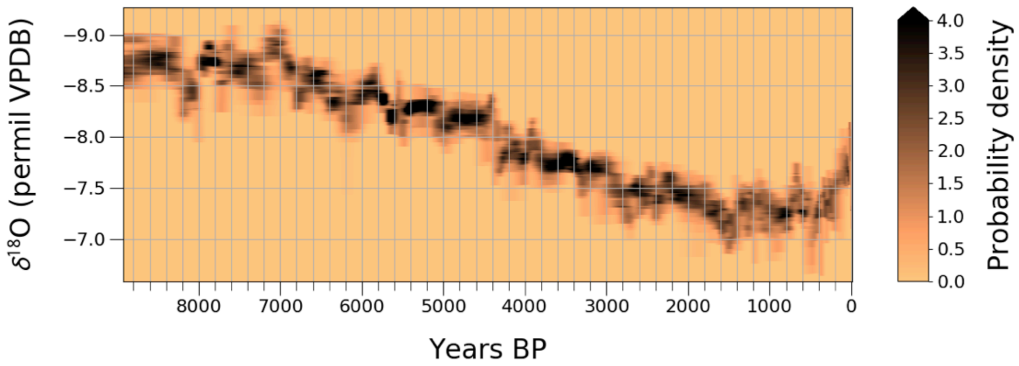
Breitenbach et al., Clim. Past, 2012

Uncertainties in paleoclimate proxy time series

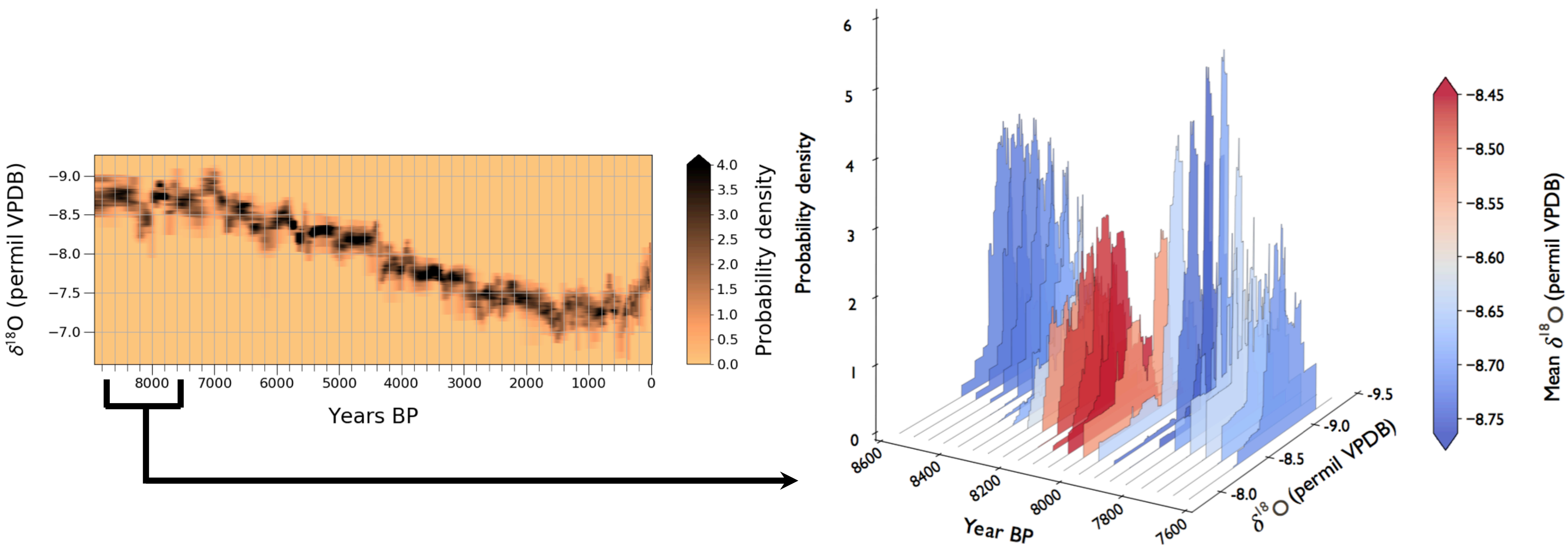


Time series as a sequence of probability distributions

Time series as a sequence of probability distributions

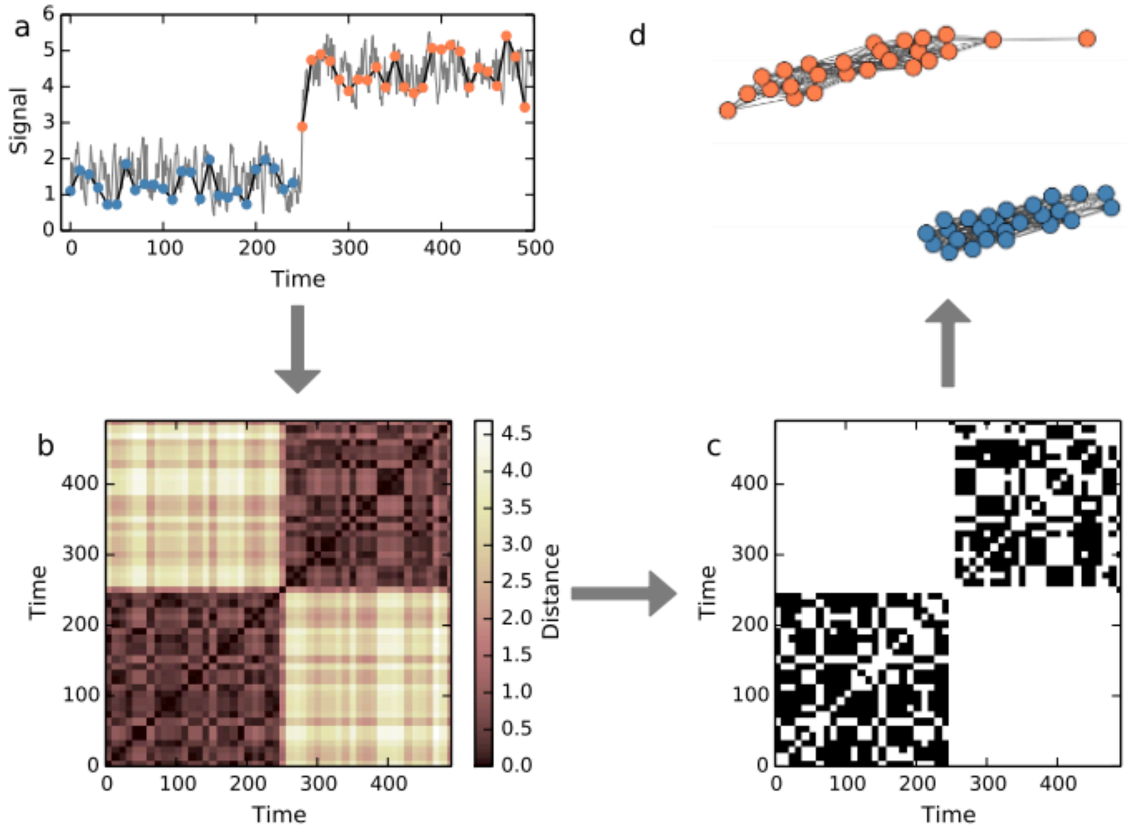


Time series as a sequence of probability distributions



Abrupt transitions as a community detection problem

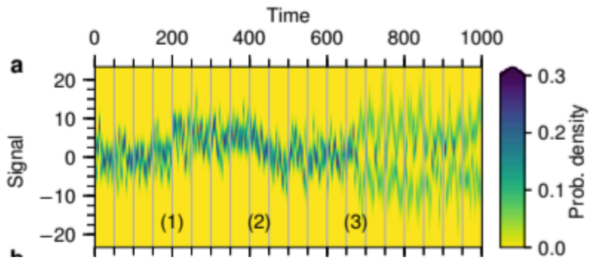
Abrupt transitions as a community detection problem

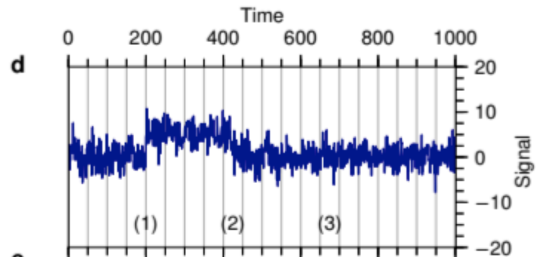


Put it together: Abrupt transitions in time series with uncertainties

Put it together: Abrupt transitions in time series with uncertainties

Time series as PDFs





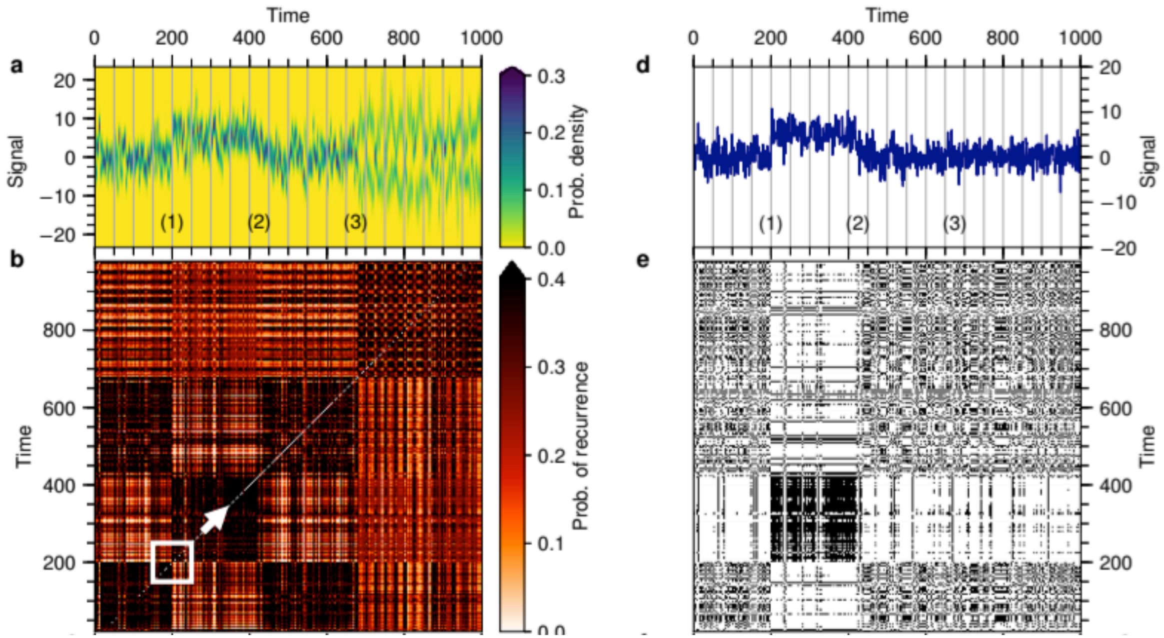
Classical tme series (mean)

Goswami et al., Nature Commnications, 2018

Put it together: Abrupt transitions in time series with uncertainties

Time series as PDFs

Probability of recurrence matrix



Classical tme series (mean)

Recurrence matrix

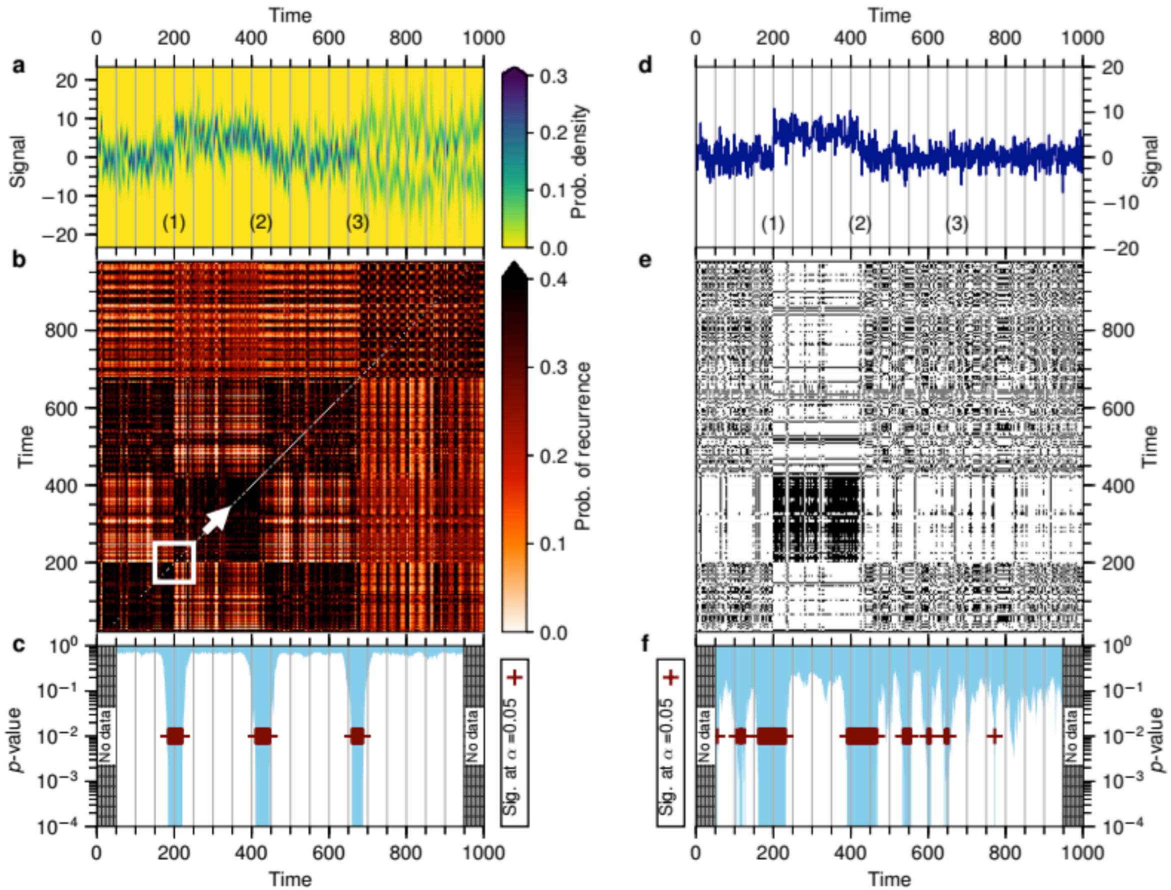
Goswami et al., Nature Commnications, 2018

Put it together: Abrupt transitions in time series with uncertainties

Time series as PDFs

Probability of recurrence matrix

p-values of test in each window



Classical tme series (mean)

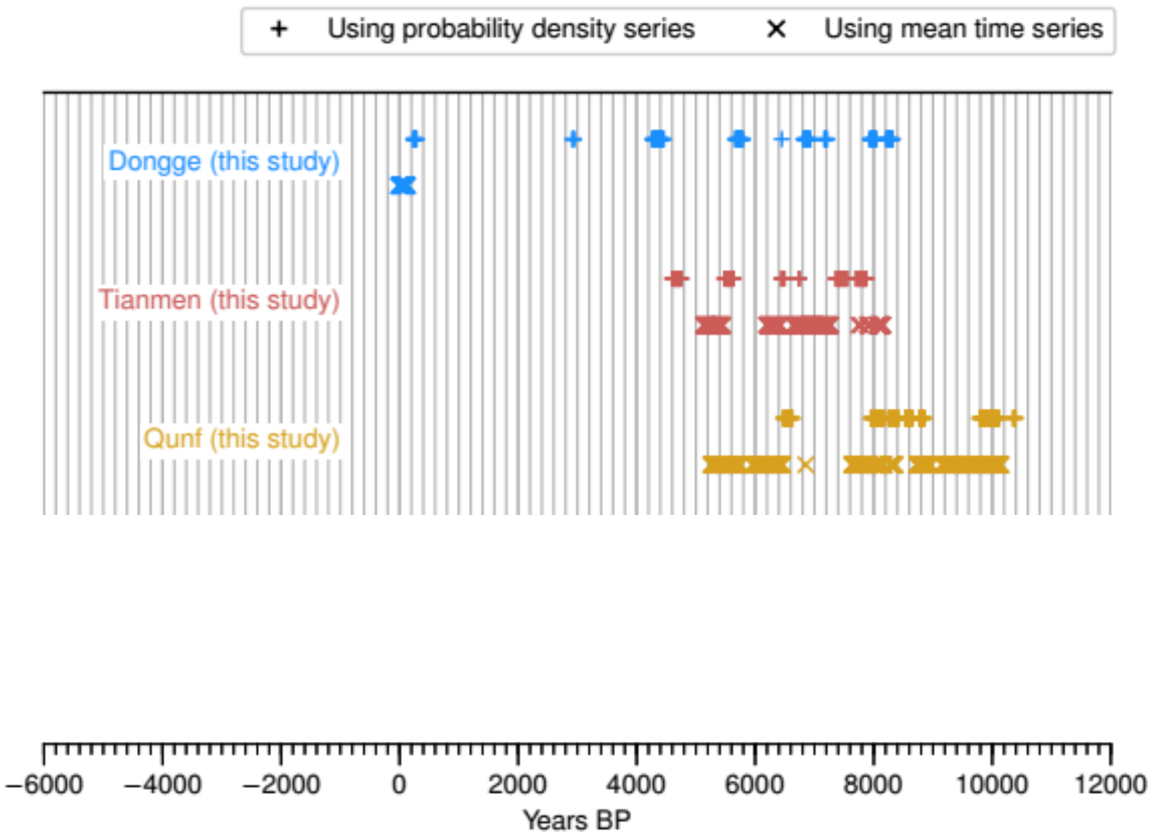
Recurrence matrix

p-values of test in each window

Transitions in the East Asian Summer Monsoon in the Holocene

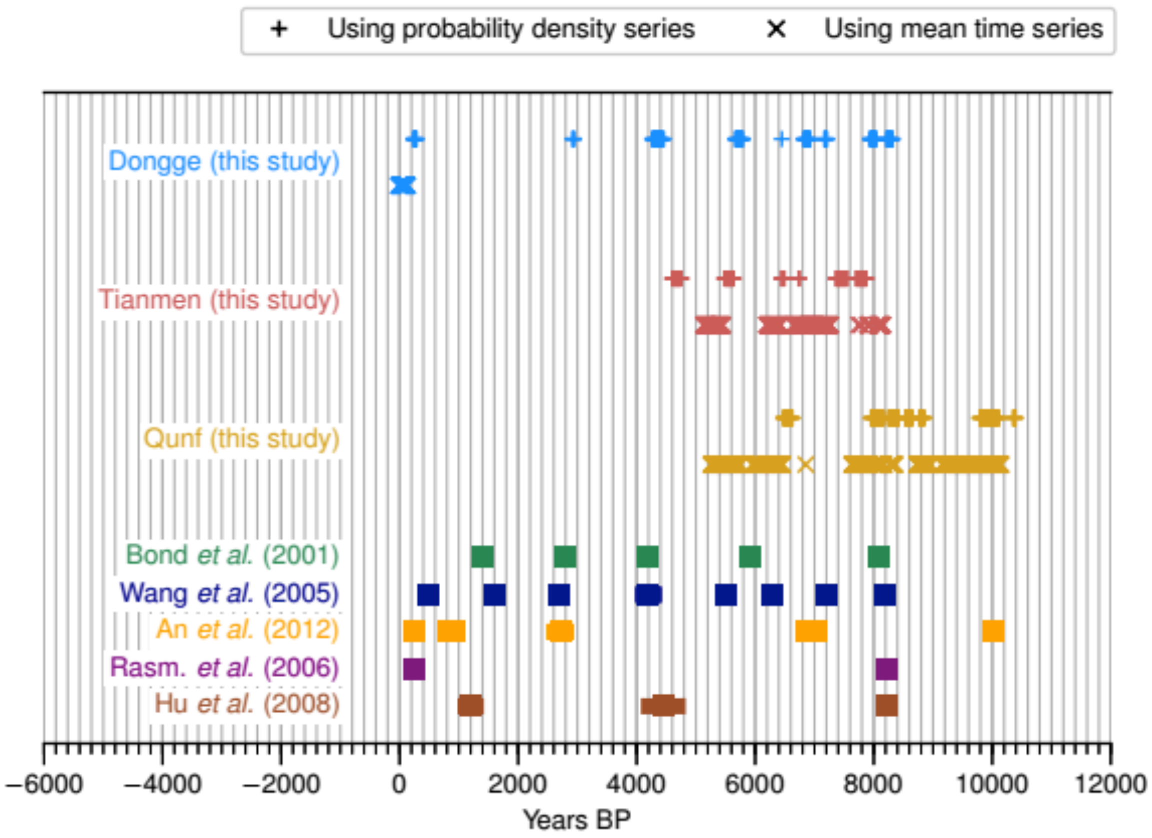
Goswami et al., Nature Commnications, 2018

Transitions in the East Asian Summer Monsoon in the Holocene



Goswami et al., *Nature Commnications*, 2018

Transitions in the East Asian Summer Monsoon in the Holocene



Outline

Part I Empirical orthogonal function analysis

→ principal components of the covariances in climate data **Part II**

Climate networks

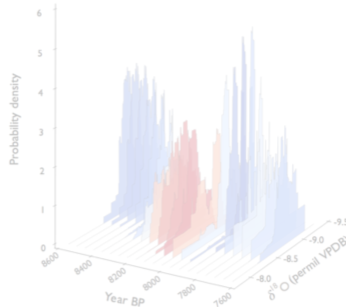
→ interactions between climate data represented as graphs **Part III**

Climate data with uncertainties

→ detecting abrupt transitions in time series with uncertainties

Outlook Deep learning for uncovering climate patterns

→ graph neural networks to reveal climate interations



The world as graph: A Graph Neural Network for climate data

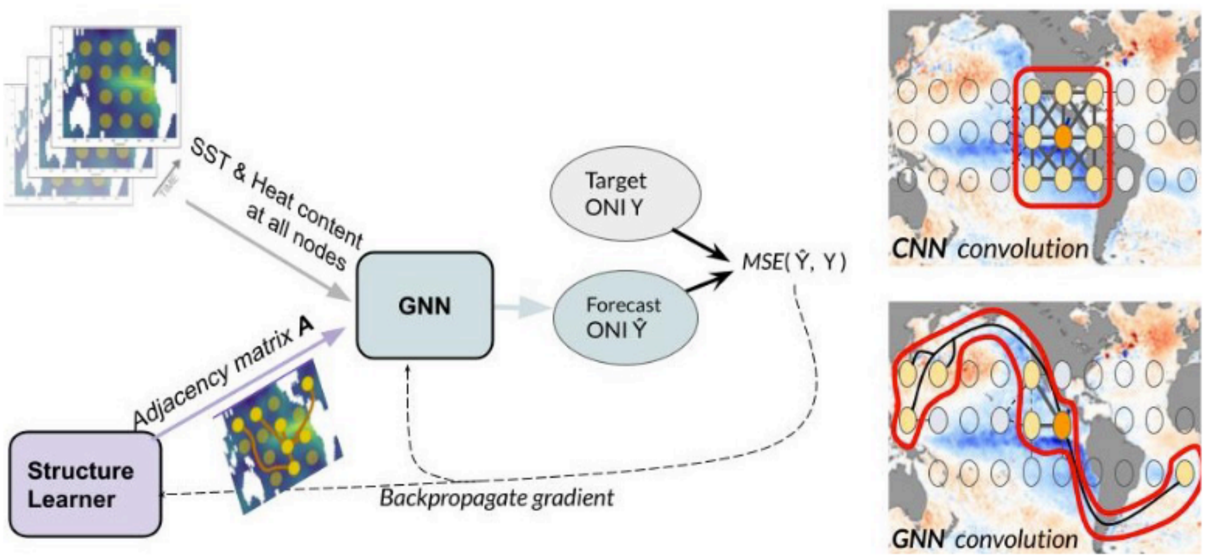
The World as a Graph: Improving El Niño Forecasts with Graph Neural Networks

Salva Rühling Cachay[1], Emma Erickson*[2],
Arthur Fender C. Bucker*[3, 4], Ernest Pokropek*[5], Willa Potosnak*[6],
Suyash Bire[8], Salomey Osei[7], and Björn Lütjens[8]

[1]Technical University of Darmstadt, [2]University of Illinois at Urbana-Champaign,
[3]University of São Paulo, [4] Technical University of Munich, [5]Warsaw University of Technology,
[6]Duquesne University, [7]African Institute for Mathematical Sciences, [8]Massachusetts Institute of Technology

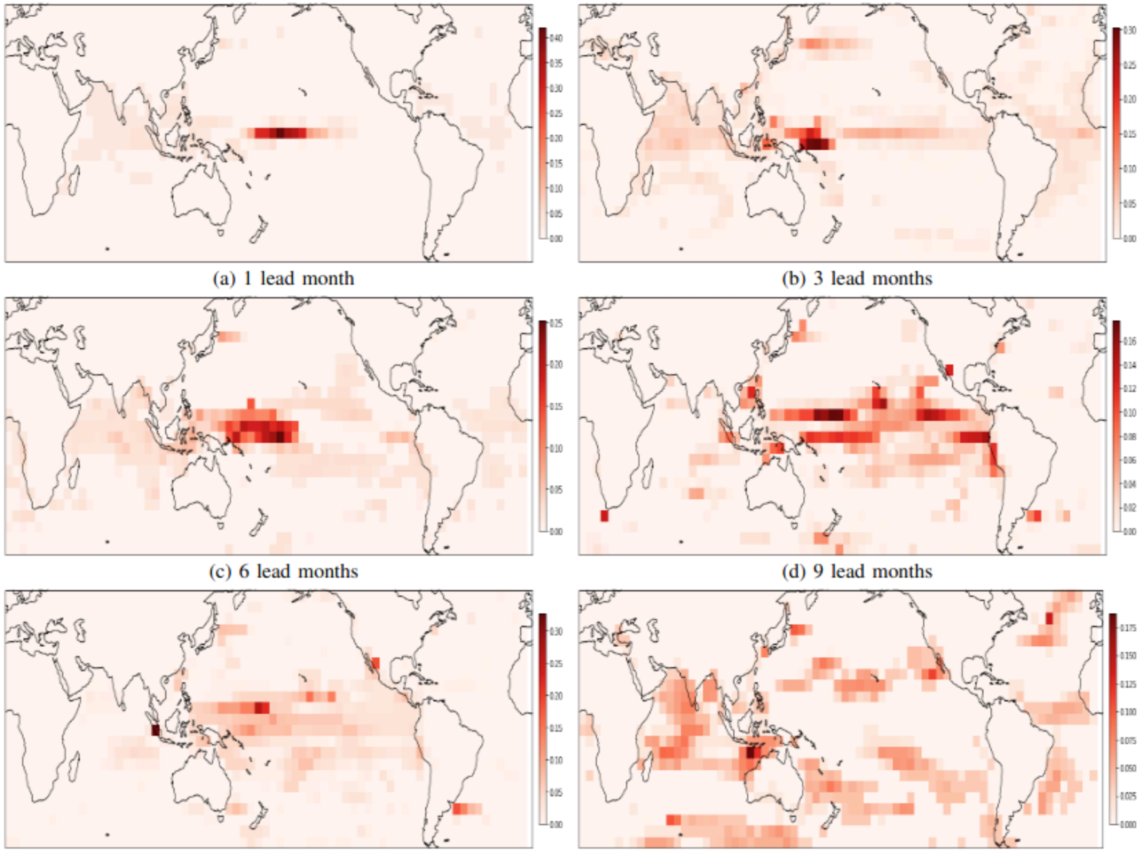
Cachay et al., arXiv, 2021

The world as graph: A Graph Neural Network for climate data



Cachay et al., arXiv, 2021

The world as graph: A Graph Neural Network for climate data



(a) 1 lead month (b) 3 lead months

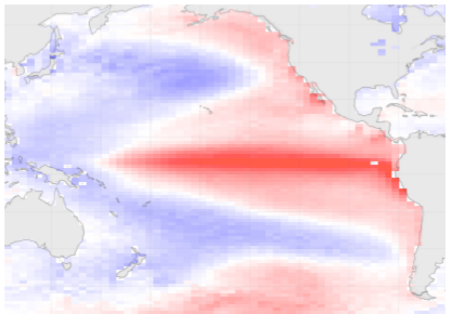
(c) 6 lead months (d) 9 lead months

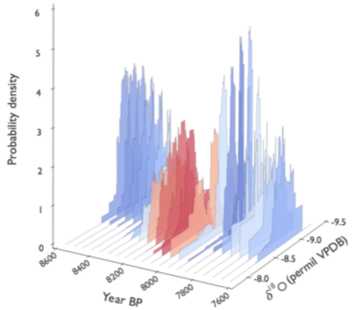
(e) 12 lead months (f) 23 lead months

Outline

Part I Empirical orthogonal function analysis
→ principal components of the covariances in climate data **Part II**
Climate networks
→ interactions between climate data represented as graphs **Part III**
Climate data with uncertainties
→ detecting abrupt transitions in time series with uncertainties

Outlook Deep learning for uncovering climate patterns
→ graph neural networks to reveal climate interations





References

1 Lian, T., & Chen, D. (2012). An evaluation of rotated EOF analysis and its application to tropical Pacific SST variability. *Journal of Climate*, 25(15), 5361-5373.
2 Hannachi, A., Jolliffe, I. T., & Stephenson, D. B. (2007). Empirical orthogonal functions and related techniques in atmospheric science: A review. *International Journal of Climatology*, 27(9), 1119-1152.
3 Boers, N., Kurths, J., & Marwan, N. (2021). Complex systems approaches for Earth system data analysis. *Journal of Physics: Complexity*, 2(1), 011001.
4 Tsonis, A. A., & Roebber, P. J. (2004). The architecture of the climate network. *Physica A*, 333, 497-504.
5 Donges, J. F., Zou, Y., Marwan, N., & Kurths, J. (2009). The backbone of the climate network. *Europhysics Letters*, 87(4), 48007.
6 Peron, T. K. D., Comin, C. H., Amancio, D. R., Da F Costa, L., Rodrigues, F. A., & Kurths, J. (2014). Correlations between climate network and relief data. *Nonlinear Processes in Geophysics*, 21(6), 1127-1132.
7 Boers, N., Bookhagen, B., Barbosa, H. M., Marwan, N., Kurths, J., & Marengo, J. A. (2014). Prediction of extreme floods in the eastern Central Andes based on a complex networks approach. *Nature Communications*, 5(1), 1-7.
8 Paluš, M., Hartman, D., Hlinka, J., & Vejmelka, M. (2011). Discerning connectivity from dynamics in climate networks. *Nonlinear Processes in Geophysics*, 18(5), 751-763.
9 Rheinwalt, A., Marwan, N., Kurths, J., Werner, P., & Gerstengarbe, F. W. (2012). Boundary effects in network measures of spatially embedded networks. *Europhysics Letters*, 100(2), 28002.
10 Boers, N., Goswami, B., Rheinwalt, A., Bookhagen, B., Hoskins, B., & Kurths, J. (2019). Complex networks reveal global pattern of extreme-rainfall teleconnections. *Nature*, 566(7744), 373-377.
11 Breitenbach, S. F., Rehfeld, K., Goswami, B., Baldini, J. U., Ridley, H. E., Kennett, D. J., ... & Marwan, N. (2012). Constructing proxy records from age models (COPRA). *Climate of the Past*, 8(5), 1765-1779.
12 Goswami, B., Boers, N., Rheinwalt, A., Marwan, N., Heitzig, J., Breitenbach, S. F., & Kurths, J. (2018). Abrupt transitions in time series with uncertainties. *Nature Communications*, 9(1), 1-10.
13 Cachay, S. R., Erickson, E., Bucker, A. F. C., Pokropek, E., Potosnak, W., Bire, S., ... & Lütjens, B. (2021). The World as a Graph: Improving El Niño Forecasts with Graph Neural Networks. *arXiv preprint*, arXiv:2104.05089.