Nonlinear climate data analysis: A complex systems perspective

Cristina Masoller
Departamento de Fisica
Universitat Politecnica de Catalunya

UNIVERSITAT POLITÈCNICA DE CATALUNYA BARCELONATECH

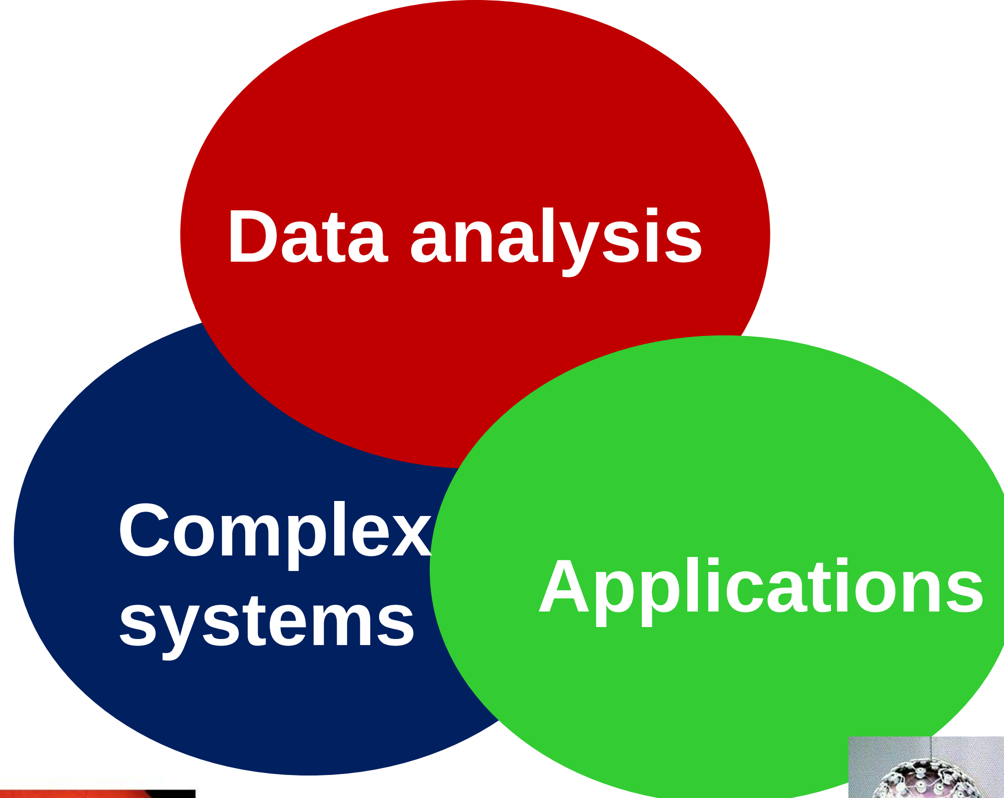
Challenges in Data Science: a Complex Systems Perspective, Torino, November 16, 2023

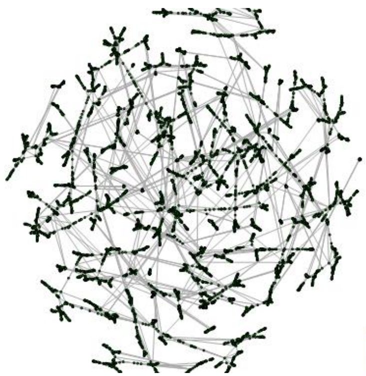
Campus d'Excel·lència Internacional

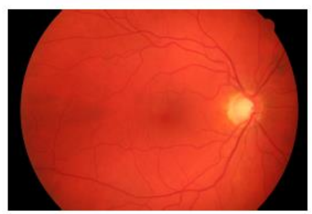
cristina.masoller@upc.edu @cristinamasoll1

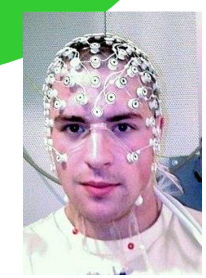
Research lines

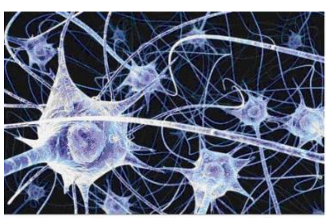
- Laser nonlinear dynamics
- Neuronal dynamics
- Complex networks
- Climate data analysis
- Biomedical data analysis
- Tipping points
- Extreme events

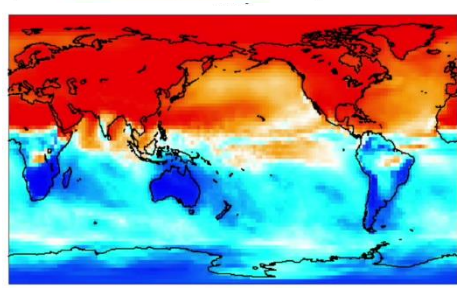


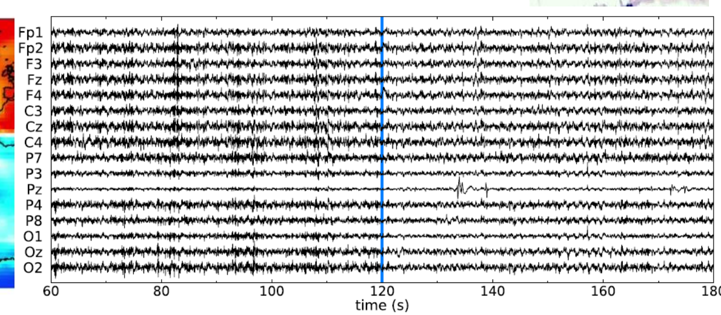








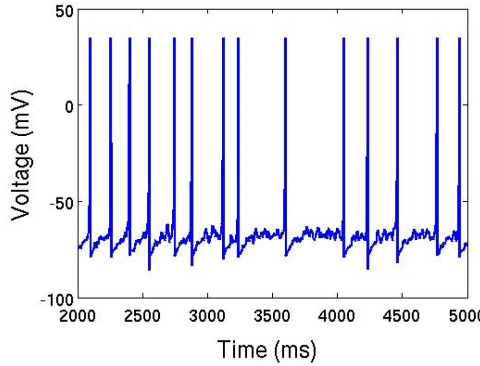




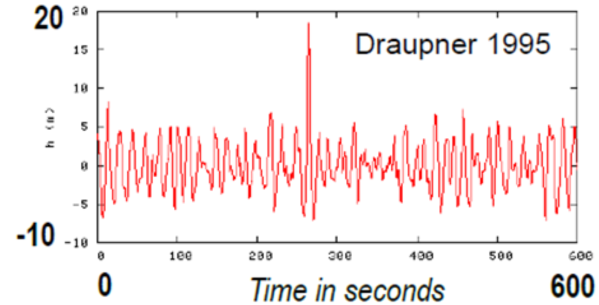
cristina.masoller@upc.edu @cristinamasoll1

Lasers, neurons, climate, complex systems?

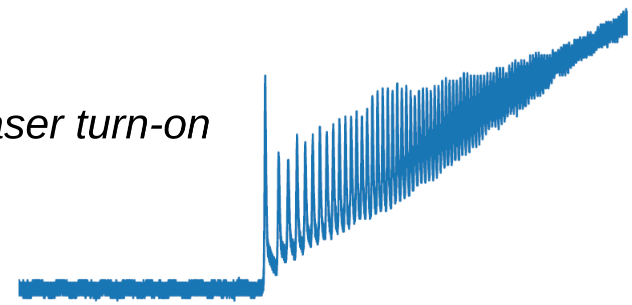
Laser & neuronal spikes



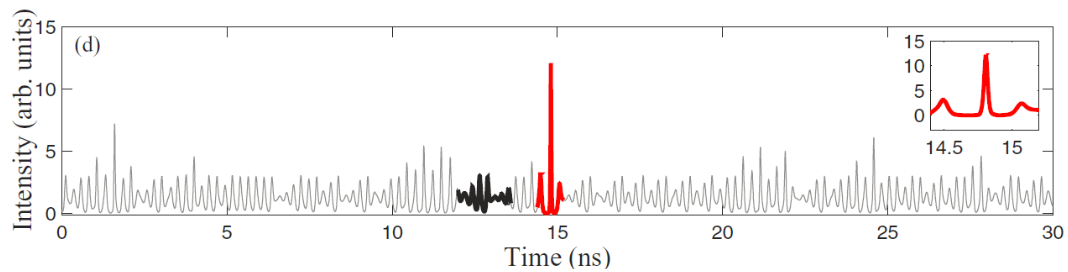
Ocean rogue wave (sea surface elevation in meters)



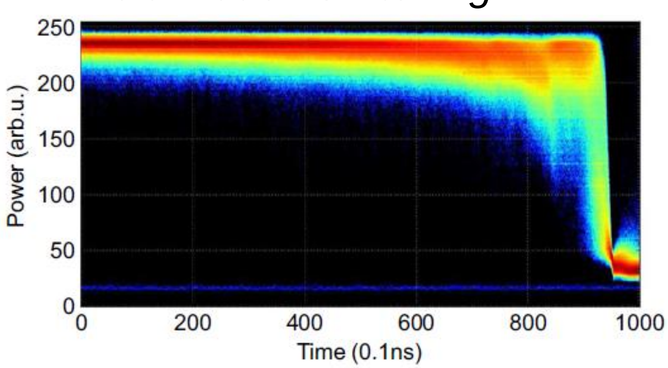
Laser turn-on



Extreme optical pulse (optical rogue wave)



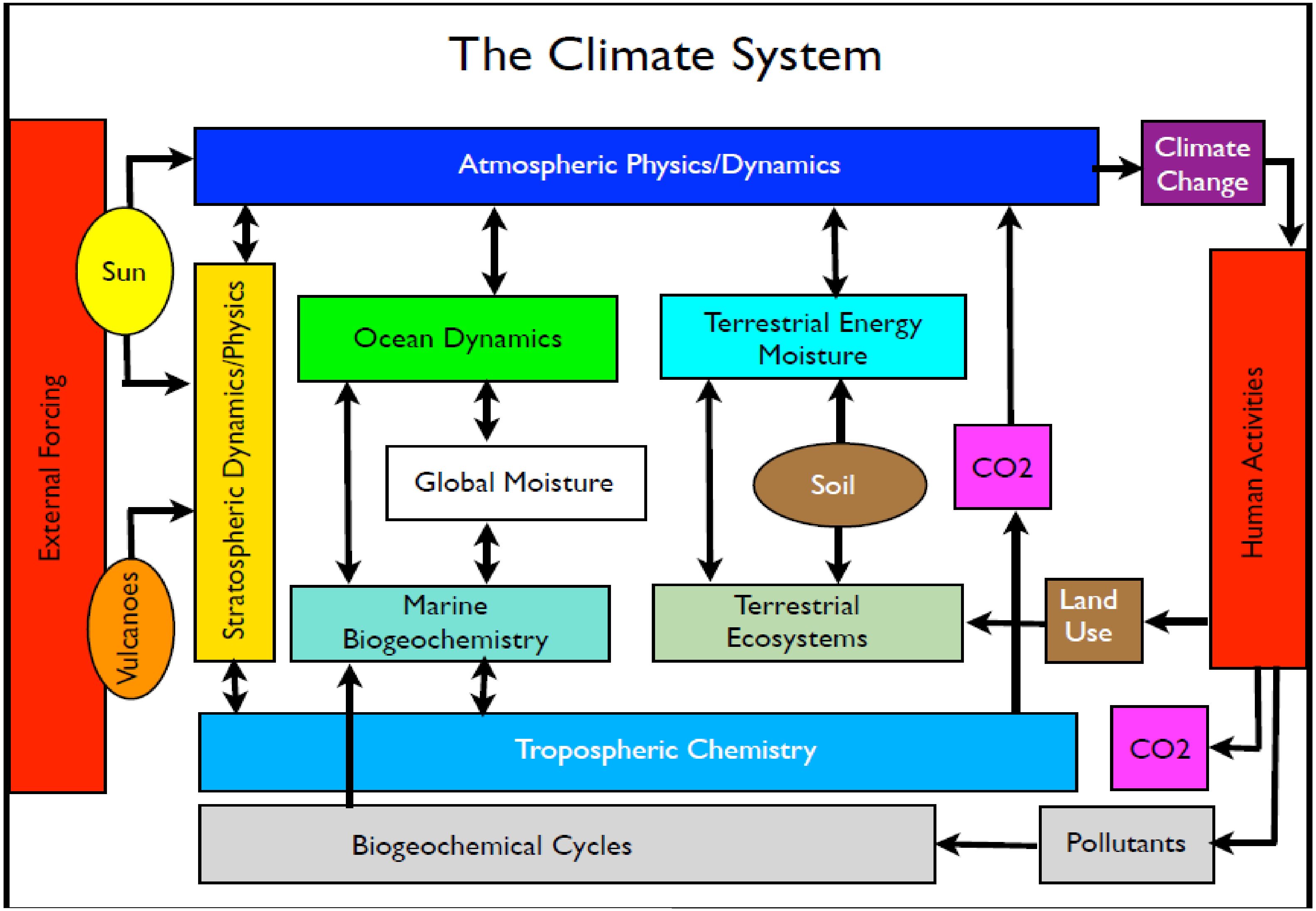
Polarization switching



cristina.masoller@upc.edu @cristinamasoll1

Nonlinear climate data analysis

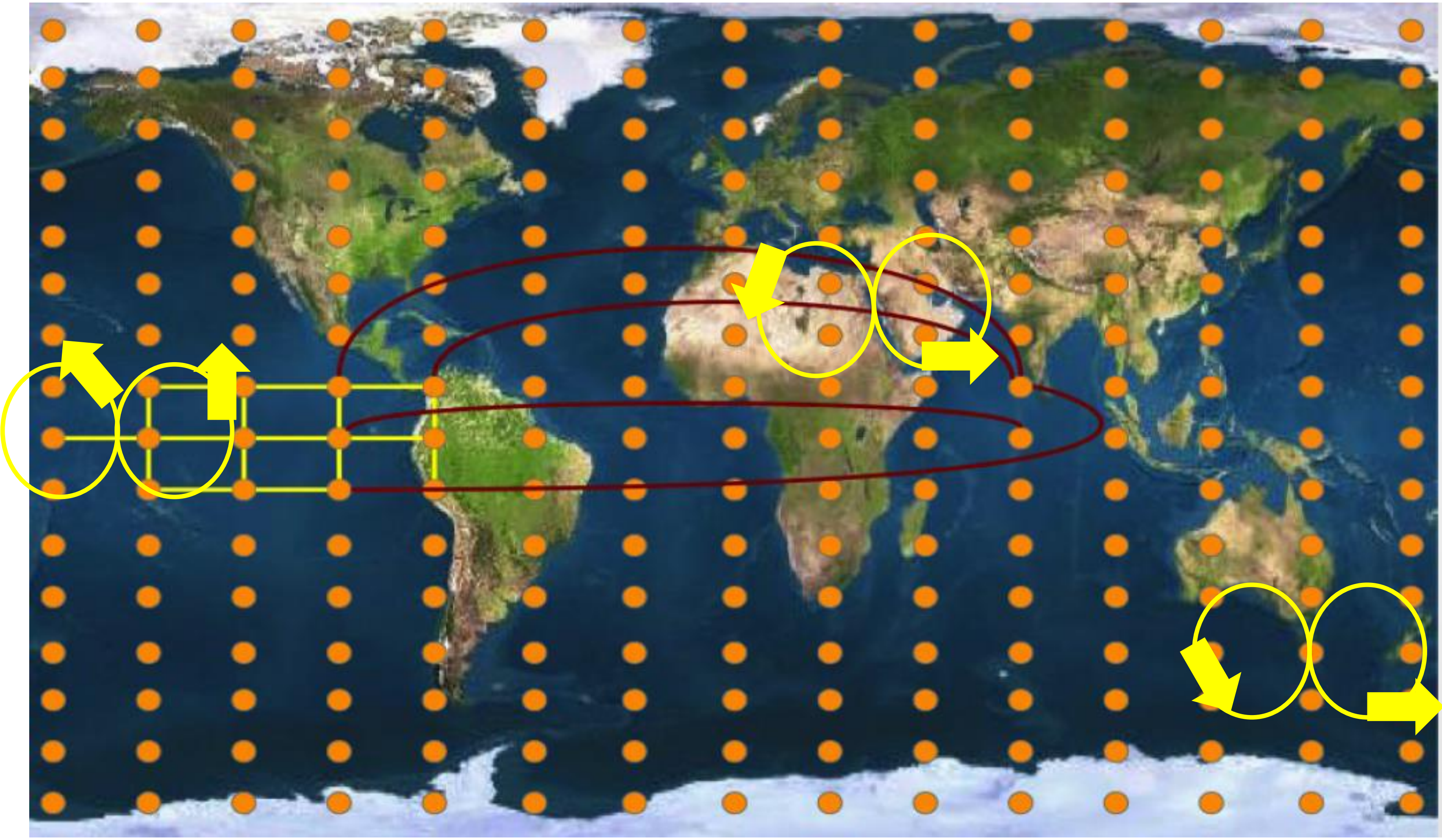
Why a complex systems perspective?



Courtesy of Henk Dijkstra (Ultrech University)

cristina.masoller@upc.edu @cristinamasoll1

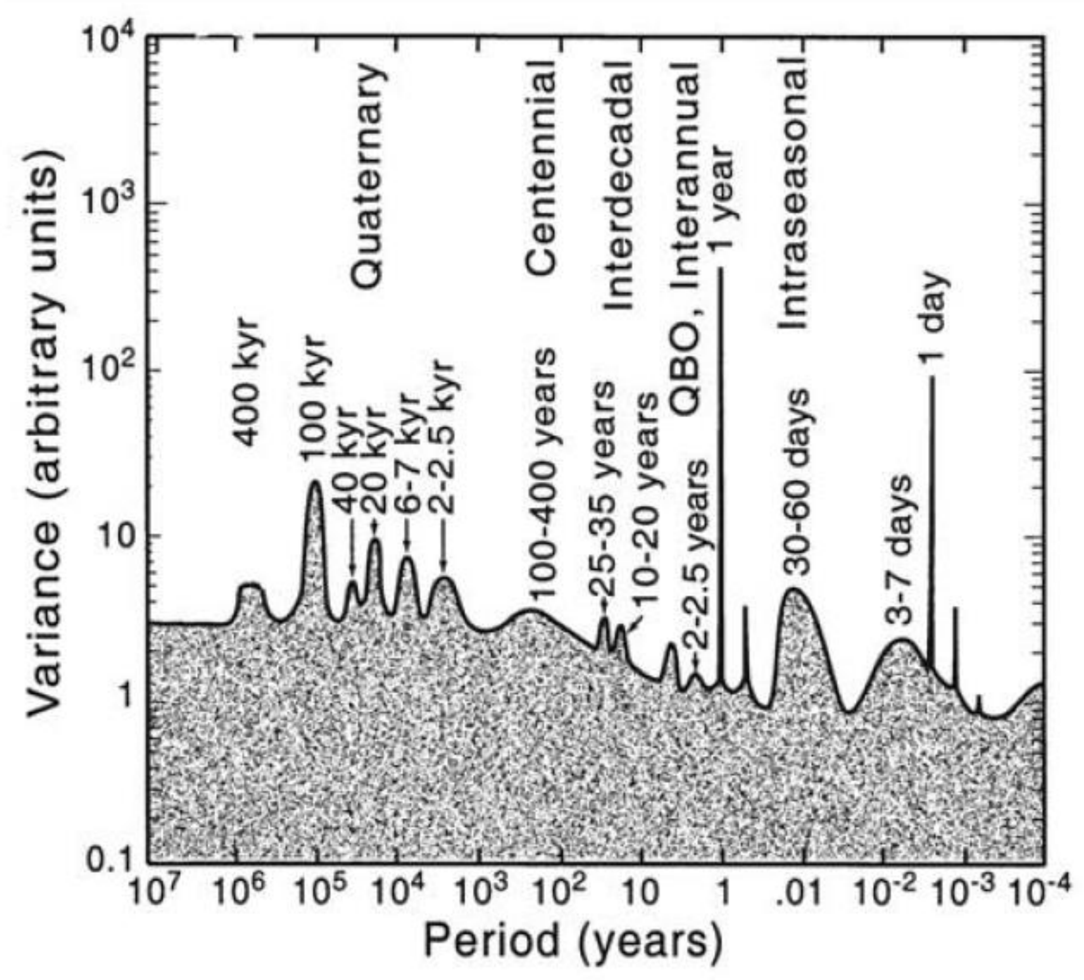
Time series of climatic variables are available with high spatial and temporal resolution



Credit: G. Tirabassi (UPC)

cristina.masoller@upc.edu @cristinamasoll1

Climate dynamics has a wide range of temporal scales



An "artistic representation" of the power spectrum of climate variability (M. Ghil 2002).

Outline of this talk:

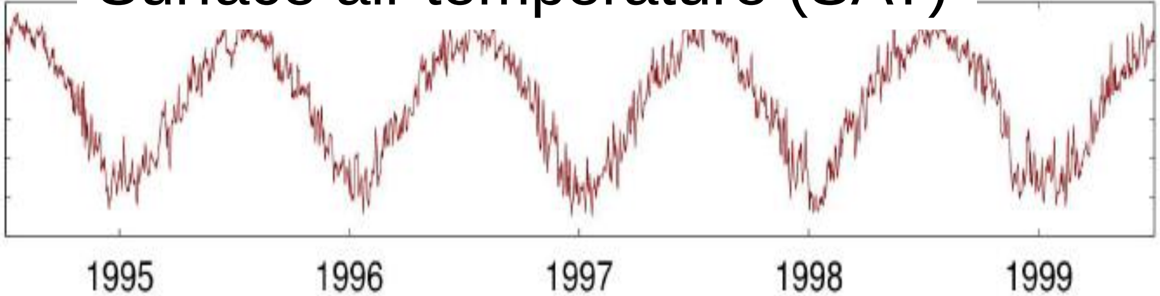
- Univariate analysis
  - Hilbert analysis
- Bivariate analysis
  - Mutual information, Transfer entropy
- Multivariate analysis
  - Climate networks

cristina.masoller@upc.edu @cristinamasoll1

1) Univariate analysis tool: The Hilbert Transform (HT)

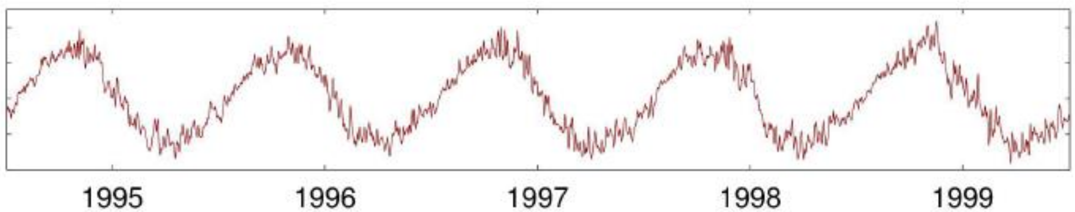
Surface air temperature (SAT)

x

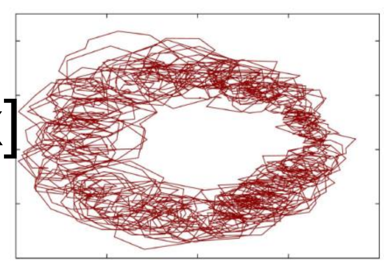


HT[sin(ωt)]=cos(ωt)

HT[x]



y=HT[x]



x

$$a(t) = \sqrt{[x(t)]^2 + [y(t)]^2}$$

$$\varphi(t) = \arctan[y(t)/x(t)]$$

cristina.masoller@upc.edu @cristinamasoll1

We analyzed a global dataset of surface air temperature

- Spatial resolution 2.5^0 x $2.5^0 \Rightarrow$ 10226 time series

- Daily resolution **1979 – 2016** $\Rightarrow$ 13700 data points

Where does the data come from?

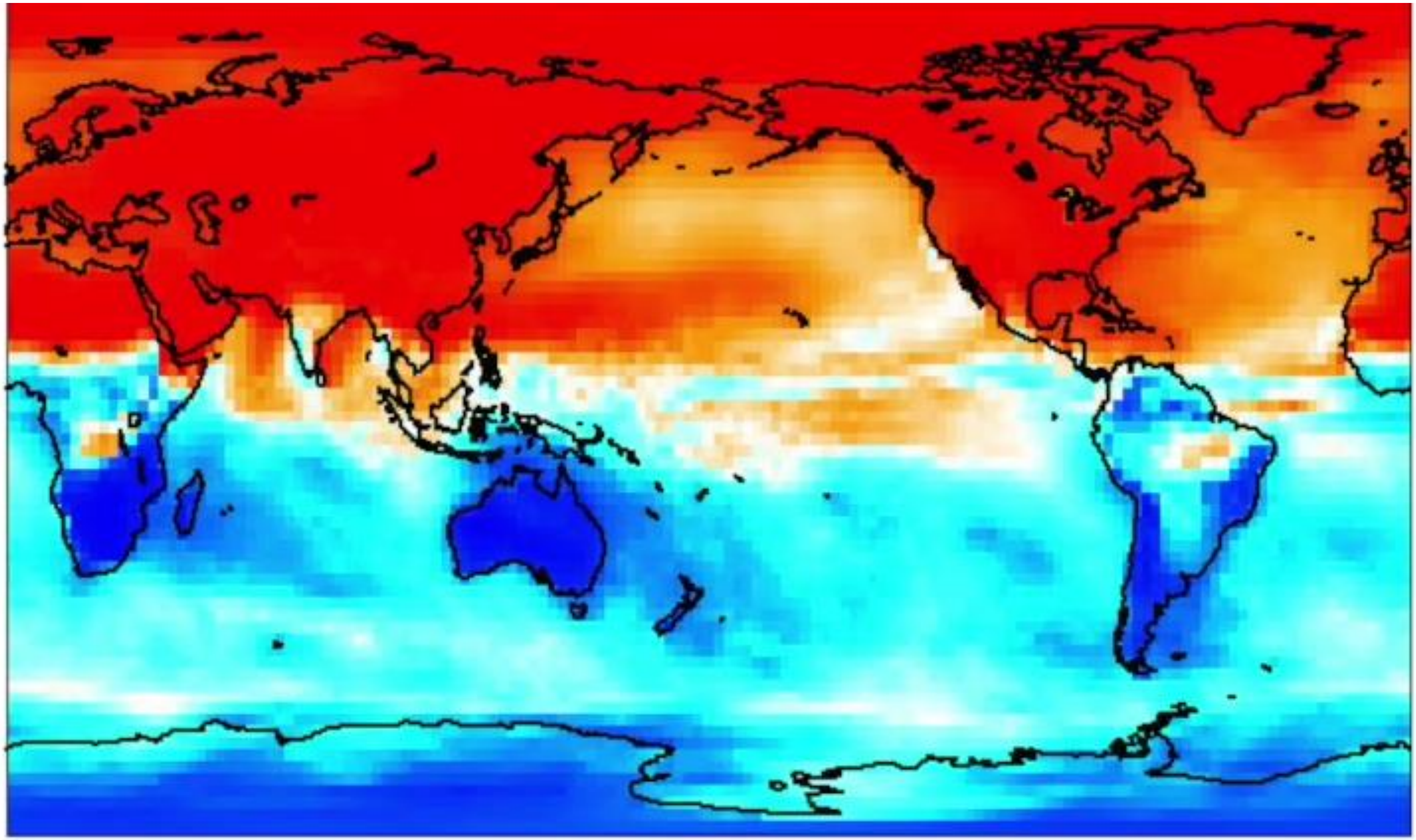
- Freely available from European Centre for Medium-Range Weather Forecasts (ECMWF).

- Reanalysis = general atmospheric circulation model feed with empirical data, where and when available (data assimilation).

cristina.masoller@upc.edu @cristinamasoll1

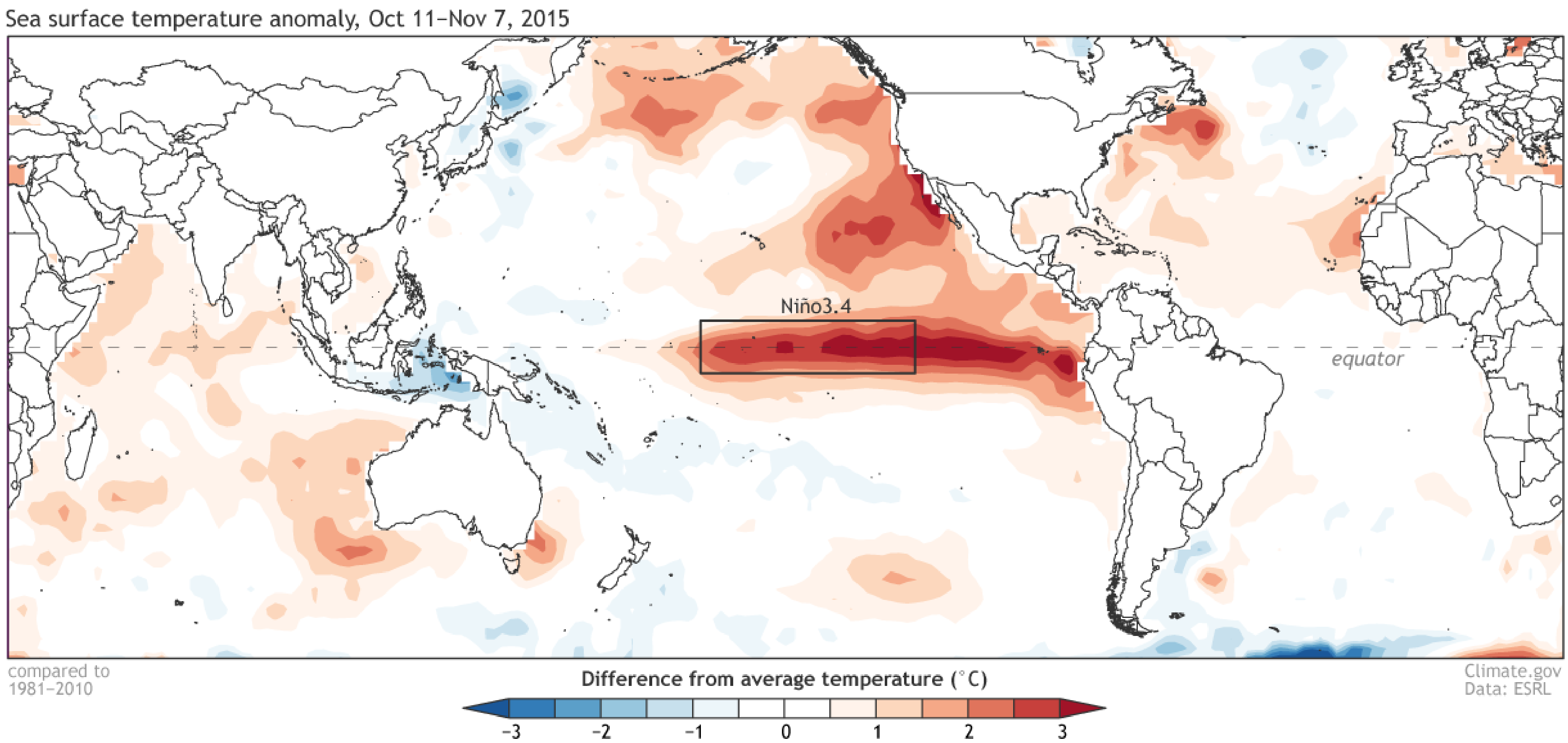
Which information carries the Hilbert phase?

In color code the cos(φ) averaged over all **July 1** in the period 1979 – 2016.



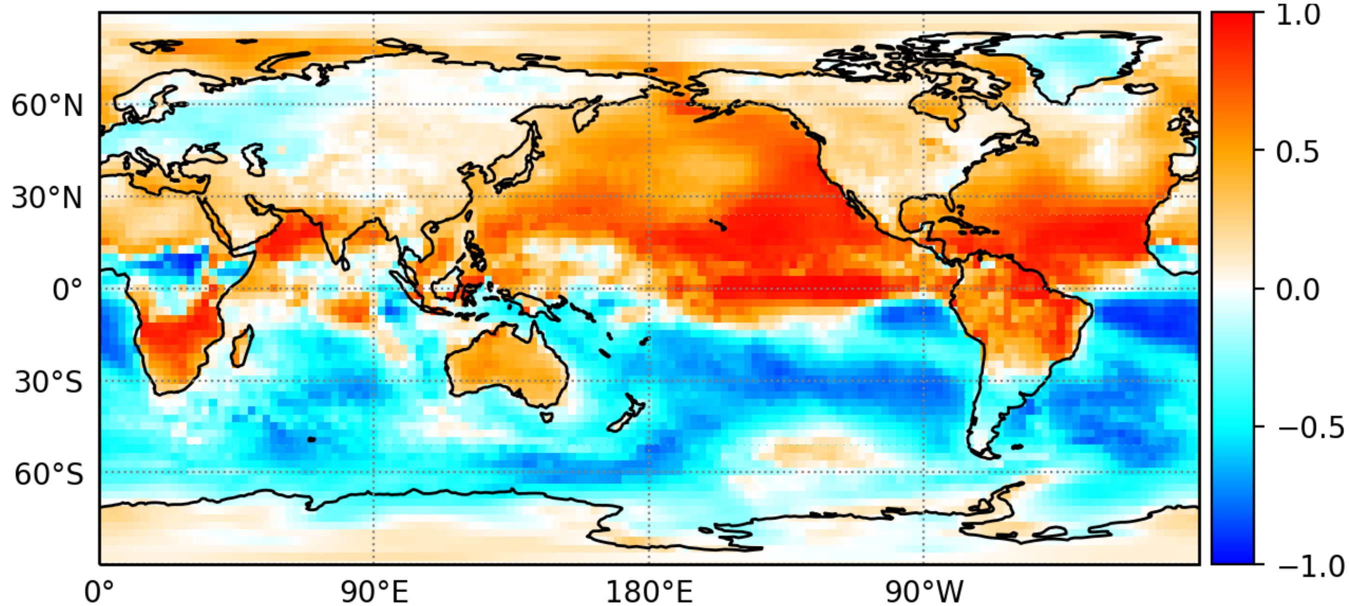
cristina.masoller@upc.edu @cristinamasoll1

ENSO (El niño / southern oscillation)

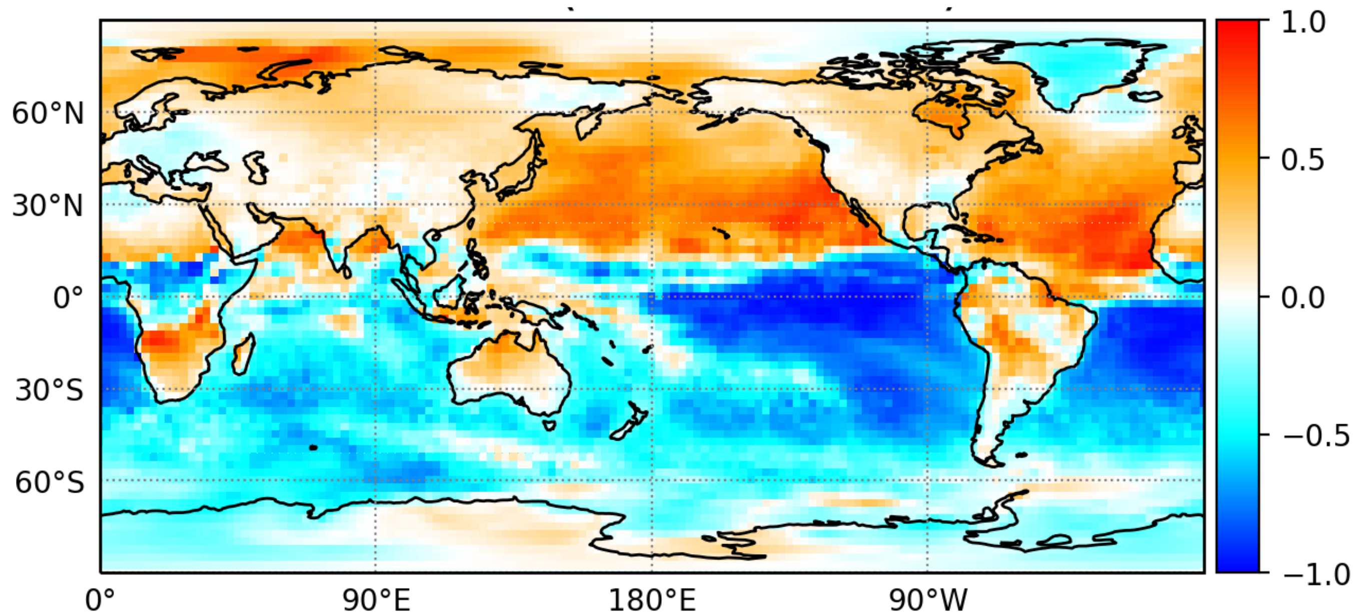


cristina.masoller@upc.edu @cristinamasoll1

El Niño period
(October 2015)

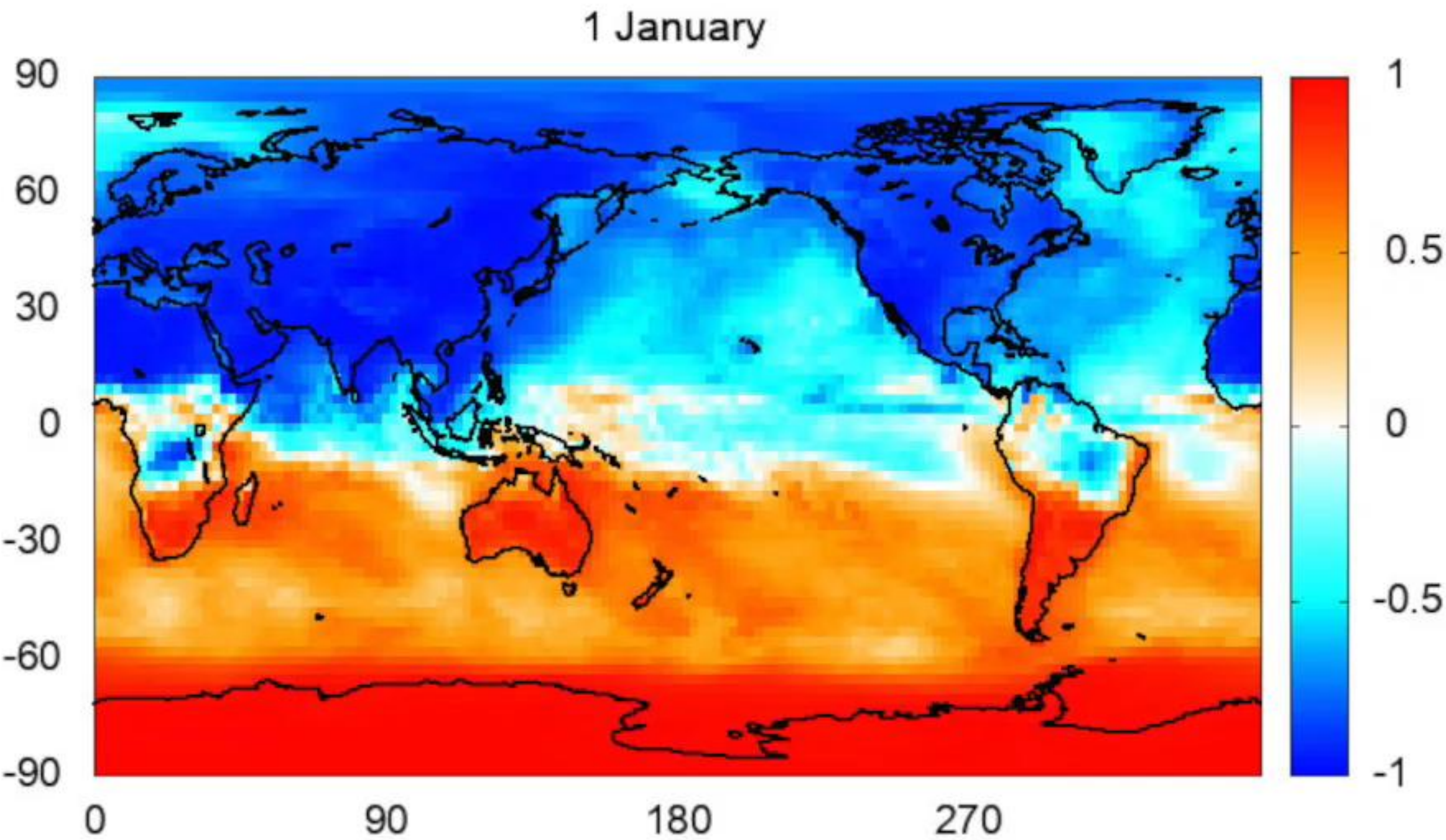


La Niña period
(Octubre 2011)



cristina.masoller@upc.edu @cristinamasoll1

How do the seasons evolve?
Temporal evolution of the cosine of the Hilbert phase



cristina.masoller@upc.edu @cristinamasoll1

Research questions

Can we use the Hilbert amplitude to identify regions where the effects of climate change are more pronounced?

Can we quantify these effects?

Relative decadal variations:

$$\Delta a = \langle a \rangle_{2016-2007} - \langle a \rangle_{1988-1979} \qquad \frac{\Delta a}{\langle a \rangle_{2016-1979}}$$

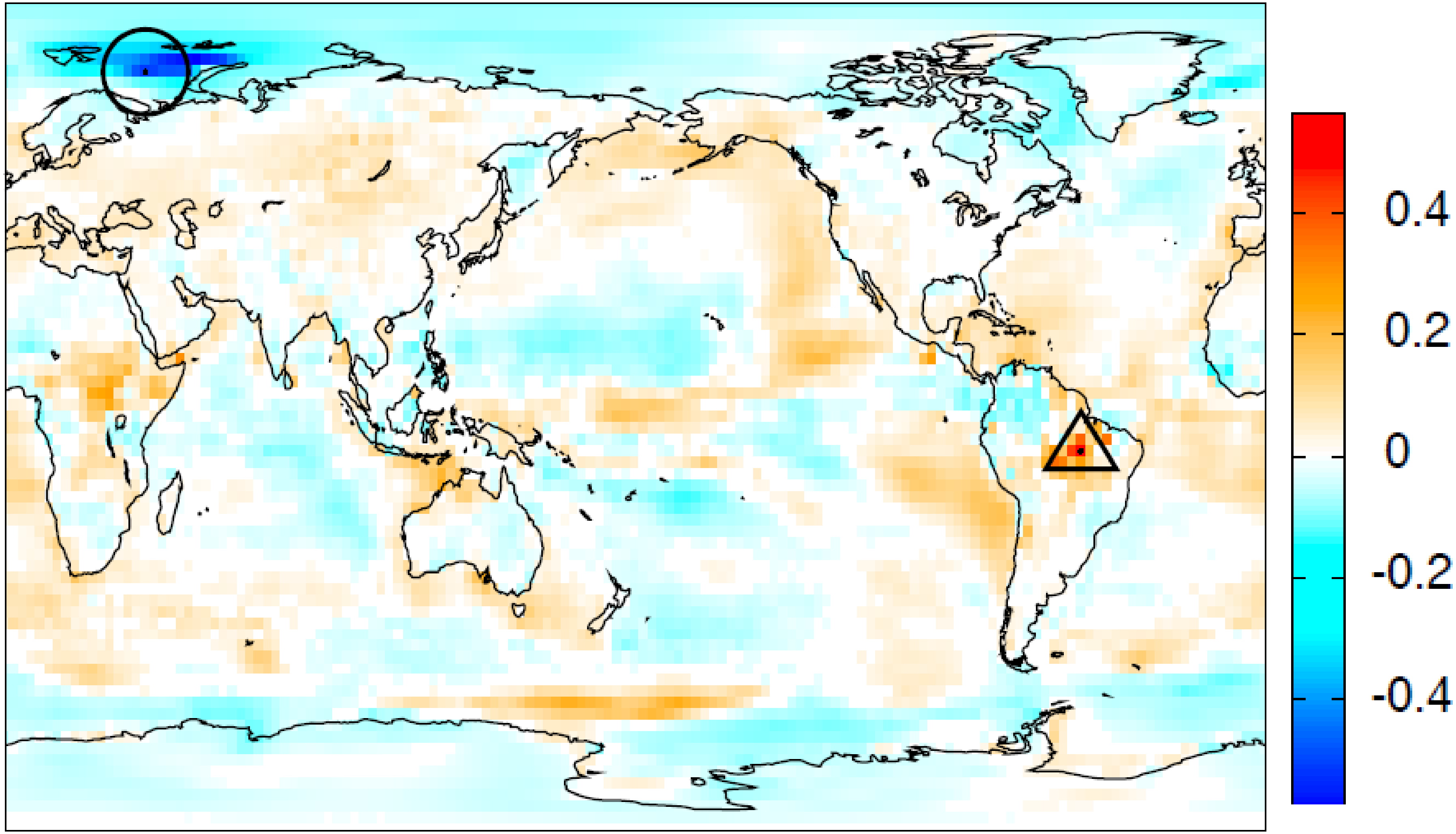
Relative variation is considered **significant** if:

$$\frac{\Delta a}{\langle a \rangle} \geq \langle . \rangle_s + 2\sigma_s \quad \text{or} \quad \frac{\Delta a}{\langle a \rangle} \leq \langle . \rangle_s - 2\sigma_s$$

100 "surrogates"

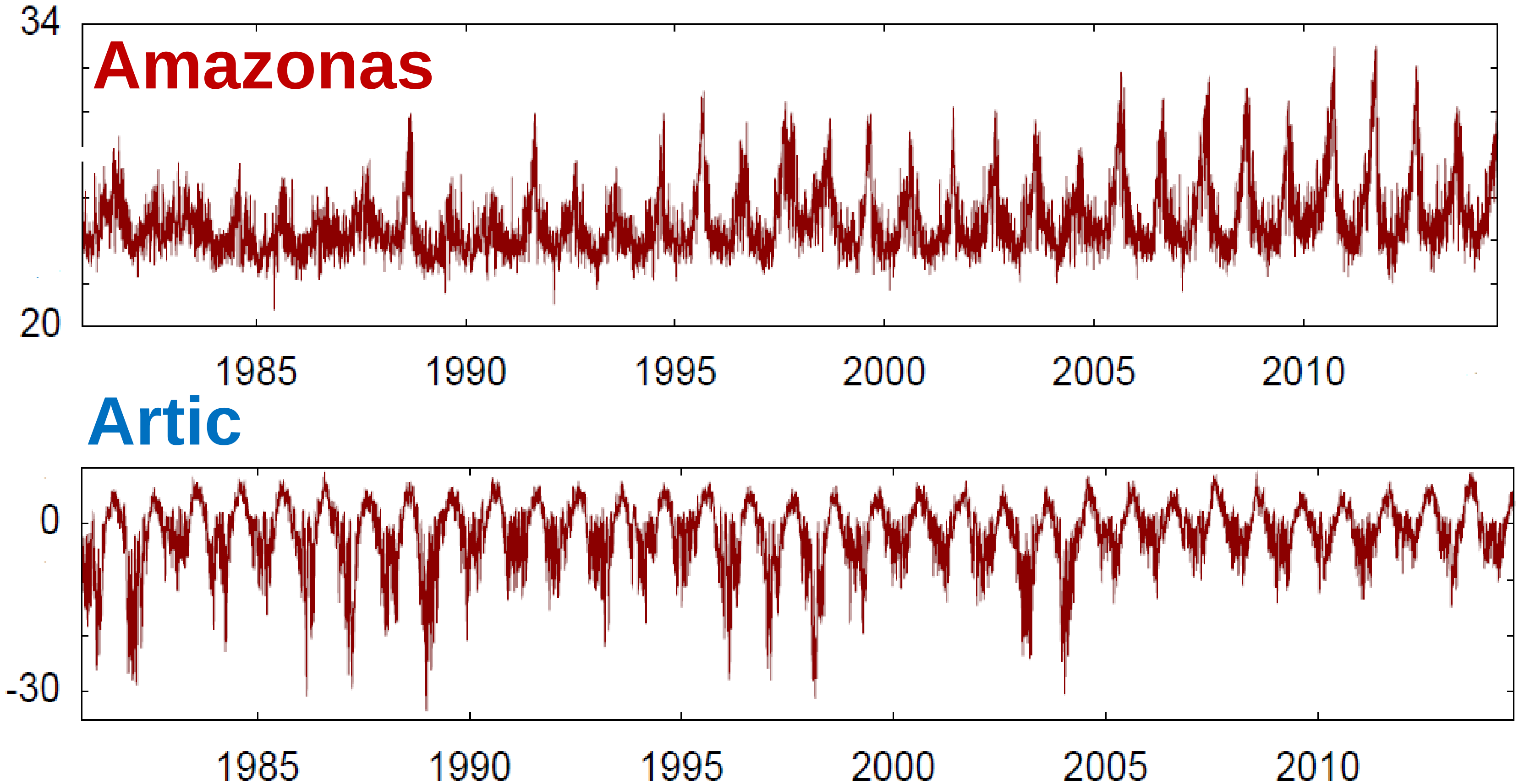
cristina.masoller@upc.edu @cristinamasoll1

Relative decadal variations



D. A. Zappala, M. Barreiro, C. Masoller, Earth Syst. Dynamics 9, 383 (2018)

cristina.masoller@upc.edu @cristinamasoll1

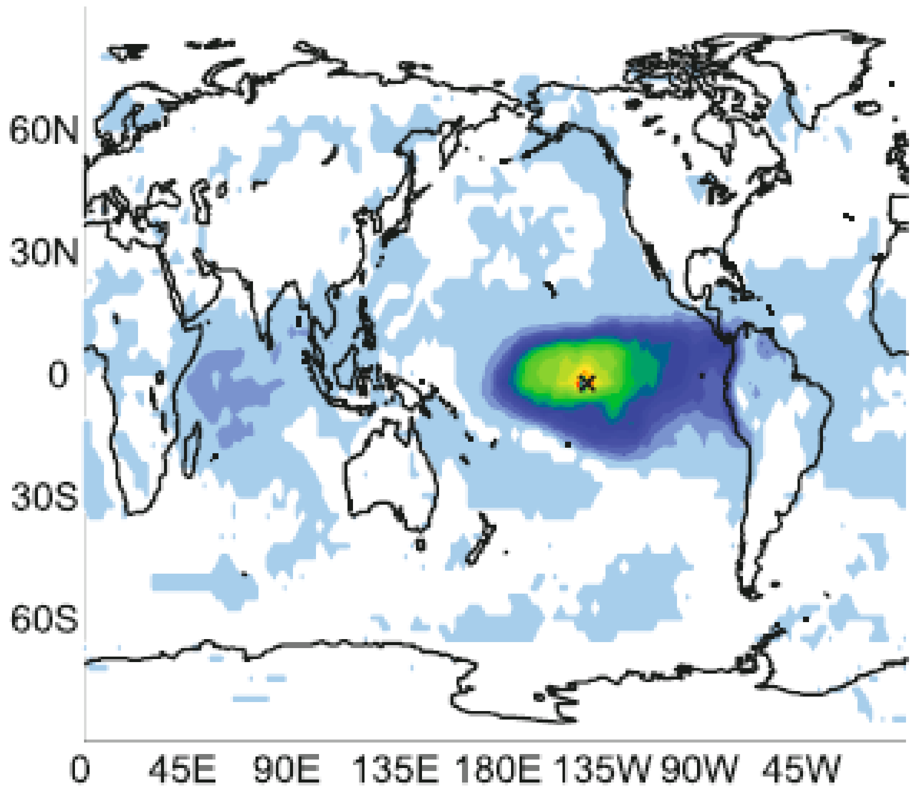


- **Decrease of precipitation** (due deforestation in Amazonas), solar radiation heats the ground.
- **Melting of sea ice**: during winter the air temperature is mitigated by the sea and tends to be more moderated.

cristina.masoller@upc.edu @cristinamasoll1

2) Bivariate analysis of surface air temperature anomalies

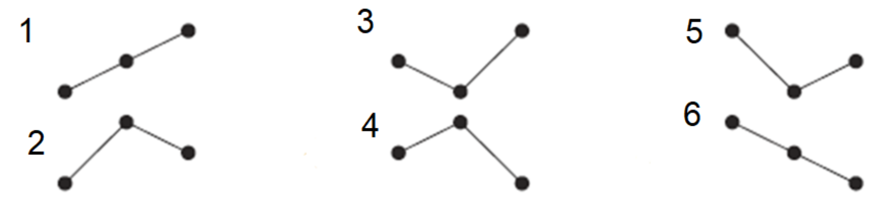
Mutual Information $$MI = \sum_{i \in x} \sum_{j \in y} p(x,y) \ \log \left(\frac{p(x,y)}{p(x)p(y)} \right)$$



J. I. Deza, M. Barreiro, and C. Masoller, “*Assessing the direction of climate interactions by means of complex networks and information theoretic tools*”, Chaos 25, 033105 (2015).

cristina.masoller@upc.edu @cristinamasoll1

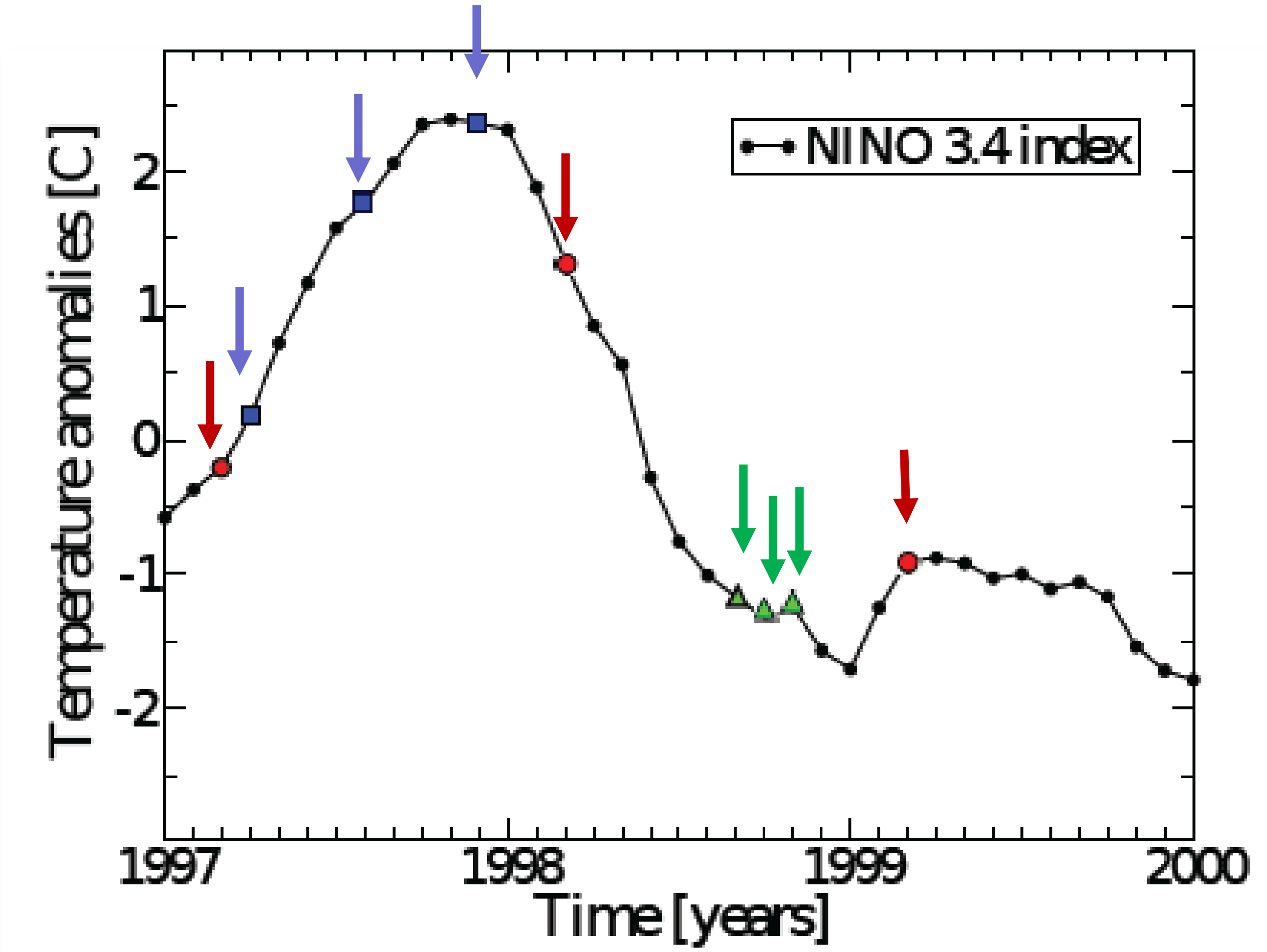
The Mutual Information can be computed from the probabilities of symbols (ordinal patterns)



Intra-season
102 (3)

Intra-annual
012 (1)

Inter-annual
120 (4)



cristina.masoller@upc.edu @cristinamasoll1

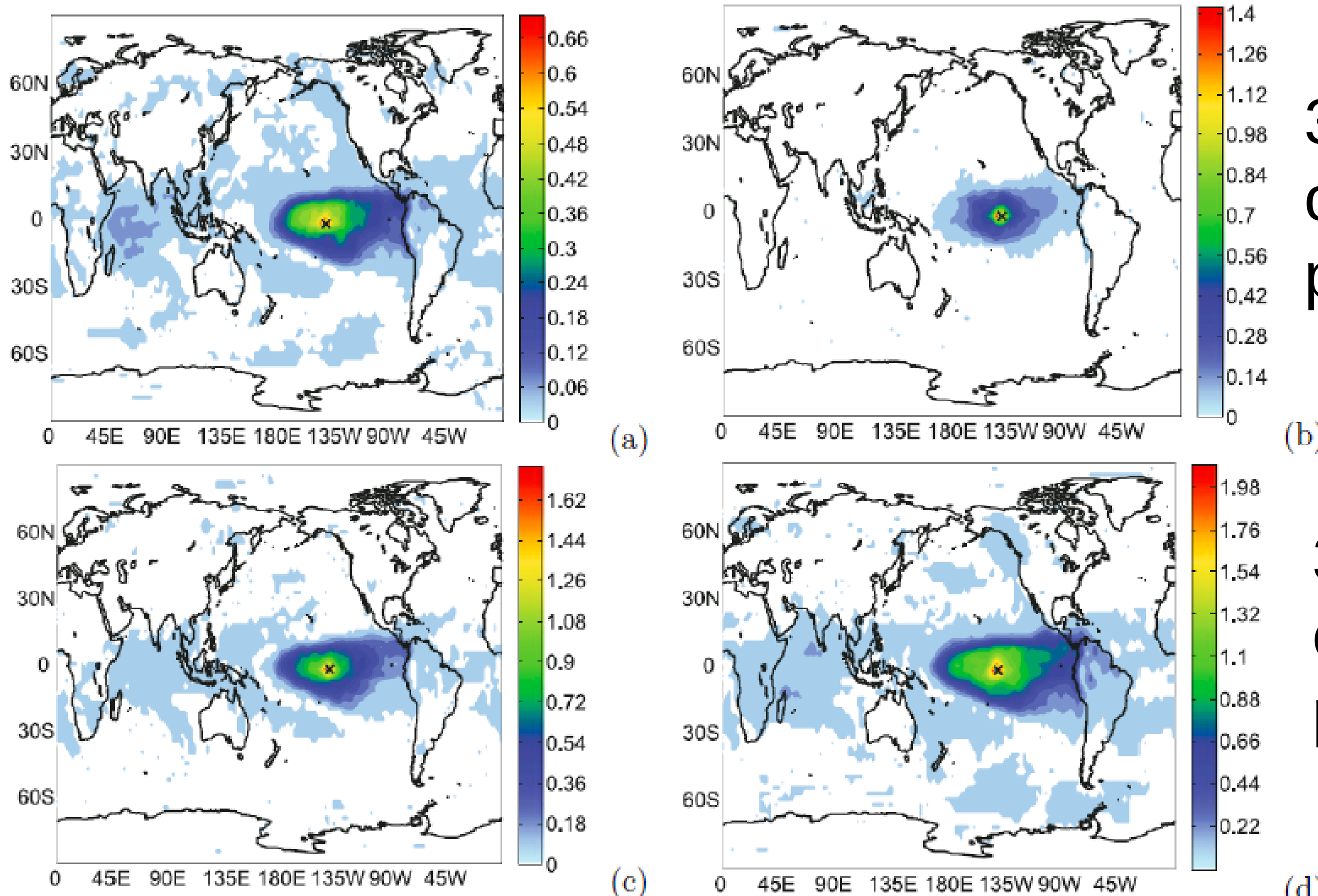
$$MI = \sum_{i \in x} \sum_{j \in y} p(x,y) \, \log \left(\frac{p(x,y)}{p(x)p(y)} \right)$$

Pdf of data values

3 months ordinal patterns

Inter-annual ordinal patterns

3 years ordinal patterns



Ordinal analysis separates the times-scales of the interactions

J. I. Deza, M. Barreiro, C. Masoller, "*Inferring interdependencies in climate networks constructed at inter-annual, intra-season and longer time scales*", Eur. Phys. J. Special Topics (2013).

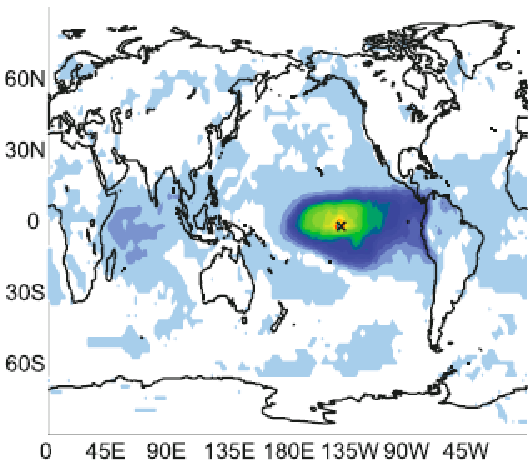
cristina.masoller@upc.edu @cristinamasoll1

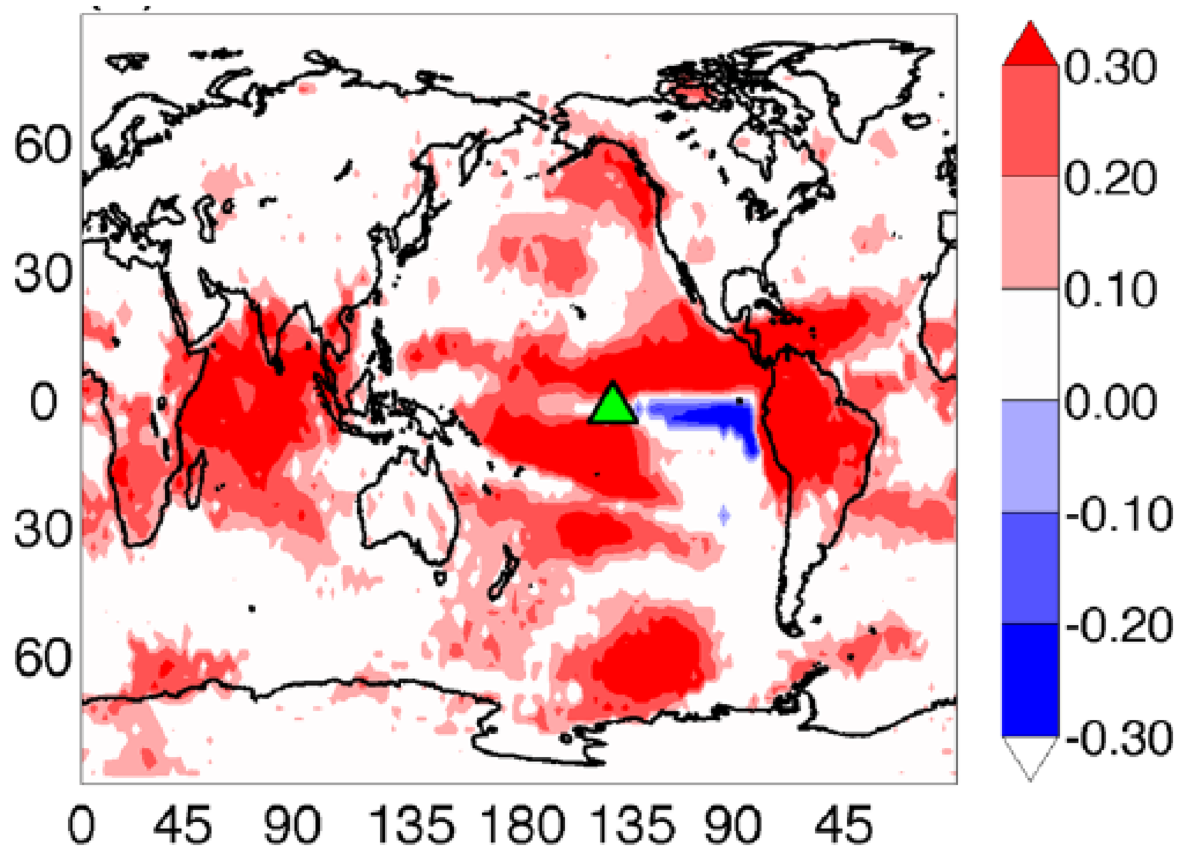
Directionality analysis using Transfer Entropy

$$TE(x,y) = MI(x, y|x_\tau)$$

$$TE(y,x) = MI(y, x|y_\tau)$$

$$DI: TE(x,y) - TE(y,x)$$

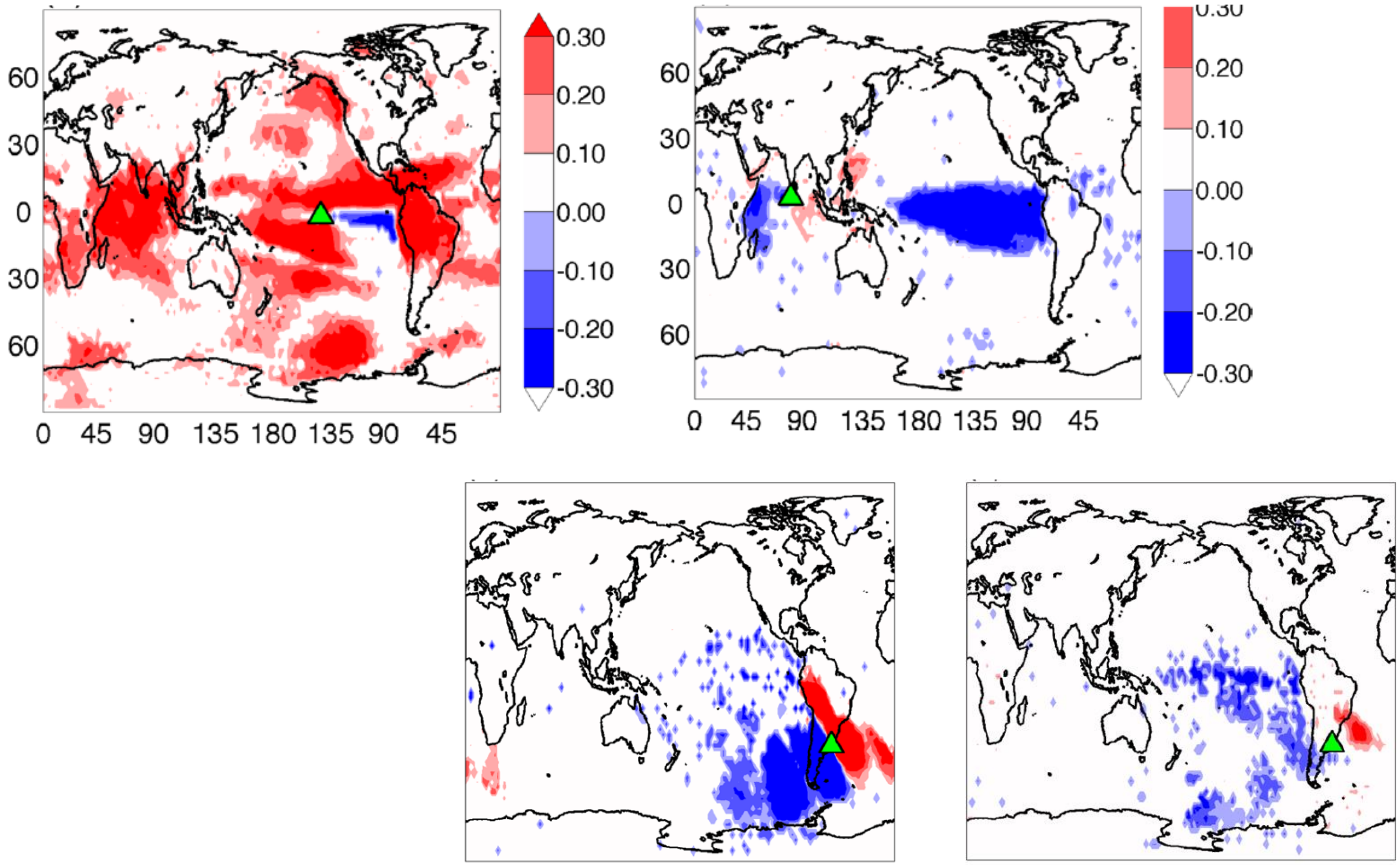




J. I. Deza, M. Barreiro, and C. Masoller, Chaos 25, 033105 (2015).

cristina.masoller@upc.edu @cristinamasoll1

Directionality Index: results consistent with known climatic interactions



J. I. Deza, M. Barreiro, and C. Masoller, Chaos 25, 033105 (2015).

cristina.masoller@upc.edu @cristinamasoll1

Problem: Transfer Entropy is computationally demanding

A "simple" solution: use the TE expression that is valid for Gaussian processes [MI = $-1/2 \log(1-\rho^2)$]

For Gaussian processes, TE is equivalent to Granger causality.

Does this work? Check it out:

scientific reports

Check for updates

OPEN **Fast and effective pseudo transfer entropy for bivariate data-driven causal inference**

Riccardo Silini & Cristina Masoller

https://doi.org/10.1038/s41598-021-87818-3

cristina.masoller@upc.edu @cristinamasoll1

Granger Causality

Hypothesis: X_1 and X_2 are described by linear autoregressive processes

$$X_1(t) = \underbrace{\sum_{j=1}^{p} A_{11,j}\, X_1(t-j)}_{\text{past of } X_1} + \underbrace{E_1(t)}_{\text{Residual error}}$$

$$X_1(t) = \underbrace{\sum_{j=1}^{p} A_{11,j}\, X_1(t-j)}_{\text{past of } X_1} + \underbrace{\sum_{j=1}^{p} A_{12,j} X_2(t-j)}_{\text{past of } X_2} + \underbrace{E'_1(t)}_{\text{Residual error}}$$

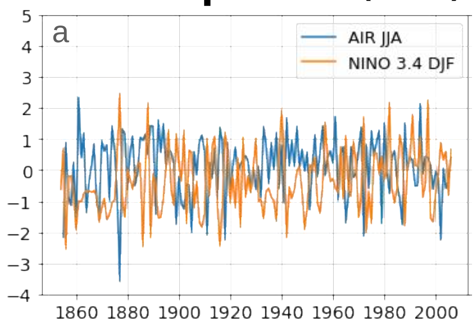
If $\langle E'_1(t)\rangle < \langle E_1(t)\rangle \Rightarrow X_2 \rightarrow X_1$

Granger, C. W. J. *Investigating causal relations by econometric models and cross-spectral methods*. *Econometrica* 37, 424–438 (1969).

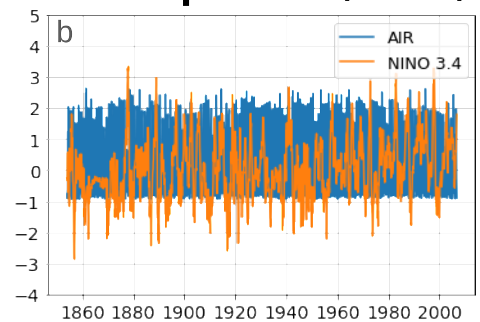
cristina.masoller@upc.edu @cristinamasoll1

Application to NINO3.4 ← → All India Rainfall

Yearly sampled (152)



Monthly sampled (1836)



IAAF

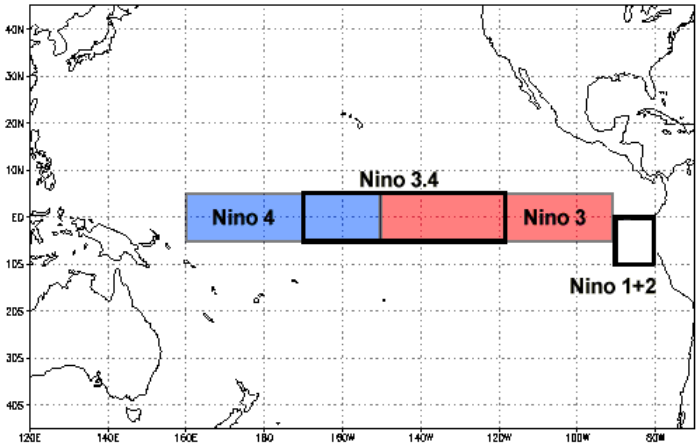
	Yearly		Monthly	
pTE	NINO3.4 ⟶ AIR	0.04 s	NINO3.4 ⟵ AIR	0.5 s
GC	NINO3.4 ⟶ AIR	0.4 s	NINO3.4 ⟵ AIR	0.9 s
TE	NINO3.4 ⇆ AIR	1 s	NINO3.4 ⇆ AIR	68 s

cristina.masoller@upc.edu @cristinamasoll1

How much time can we save?

For two time-series of 500 data points (1 data point per month, 40 years):

TE:**112 ms** but pTE: **4 ms**



8000 grid points (high resolution)
$\Rightarrow$ 64 x 10^6 pairs

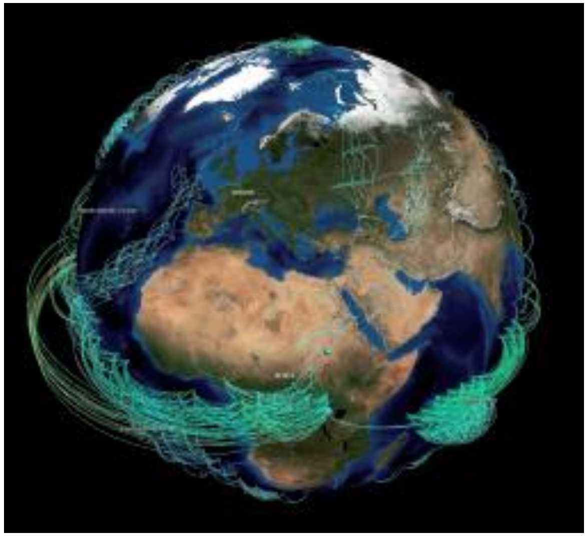
$\Rightarrow$ **829 days** (TE) vs. **29 days** (pTE).

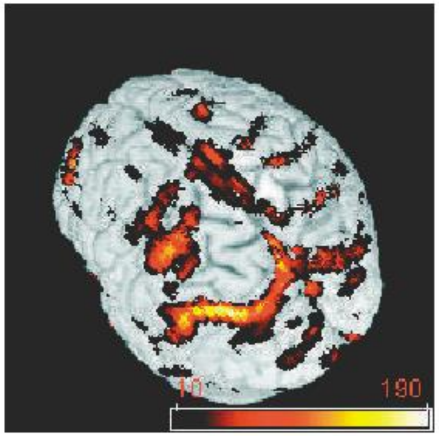
(without “surrogate” analysis)

https://github.com/riccardosilini/pTE

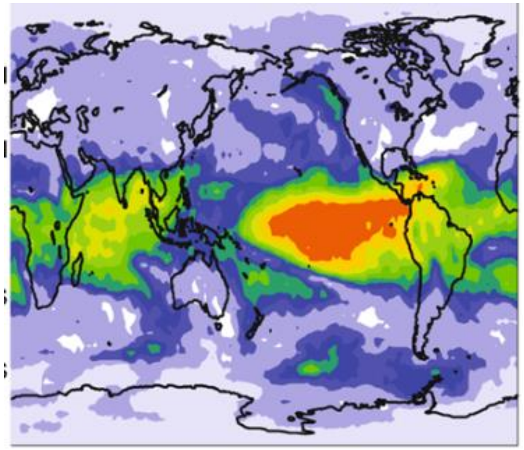
cristina.masoller@upc.edu @cristinamasoll1

3) Multivariate analysis: Brain / climate networks





"Degree"
(number of links)



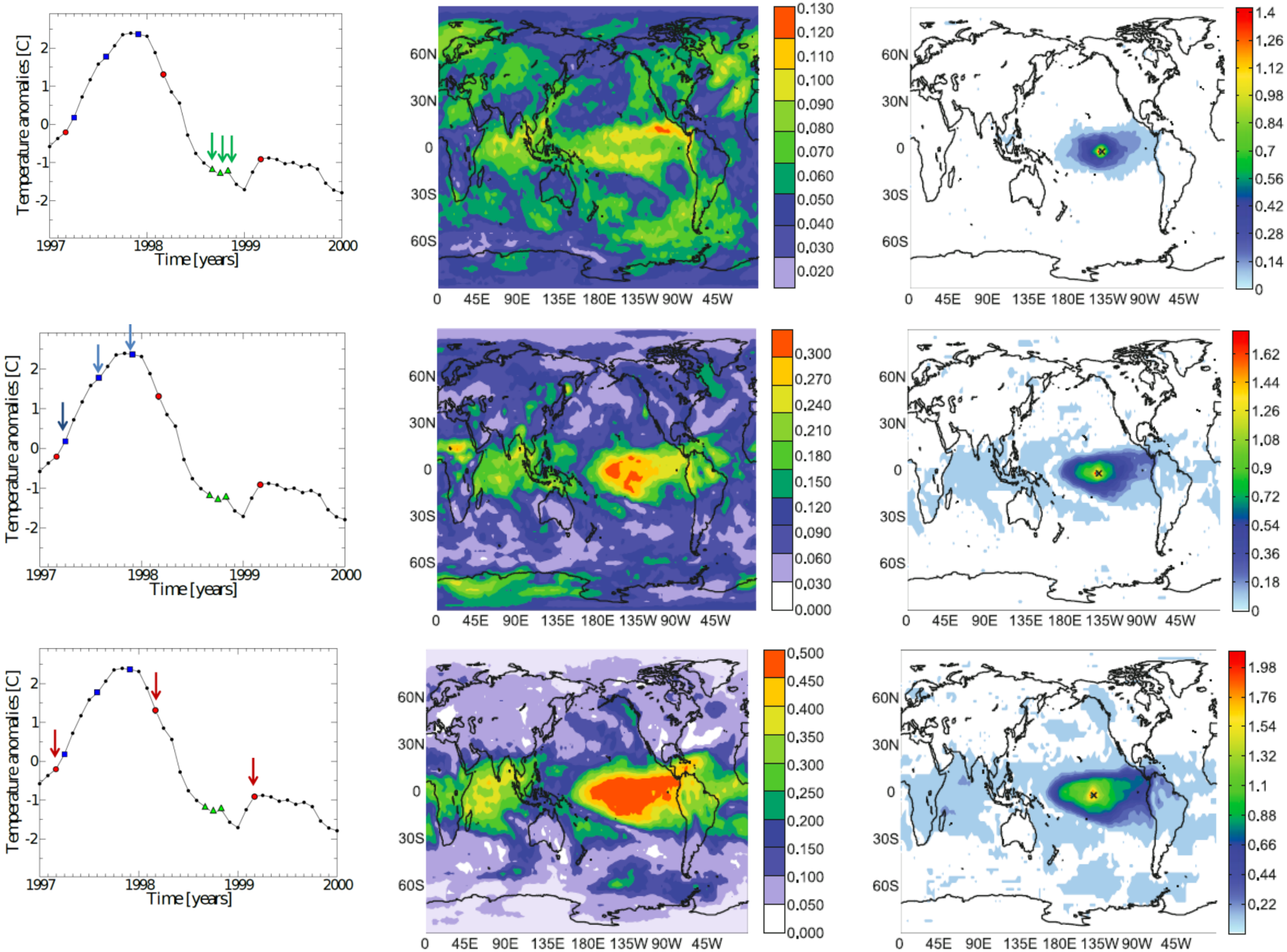
"Area
weighted
degree"

cristina.masoller@upc.edu @cristinamasoll1

Ordinal analysis allows to identify the time scale of the links

Area weighted degree

Links of a node in El Niño area

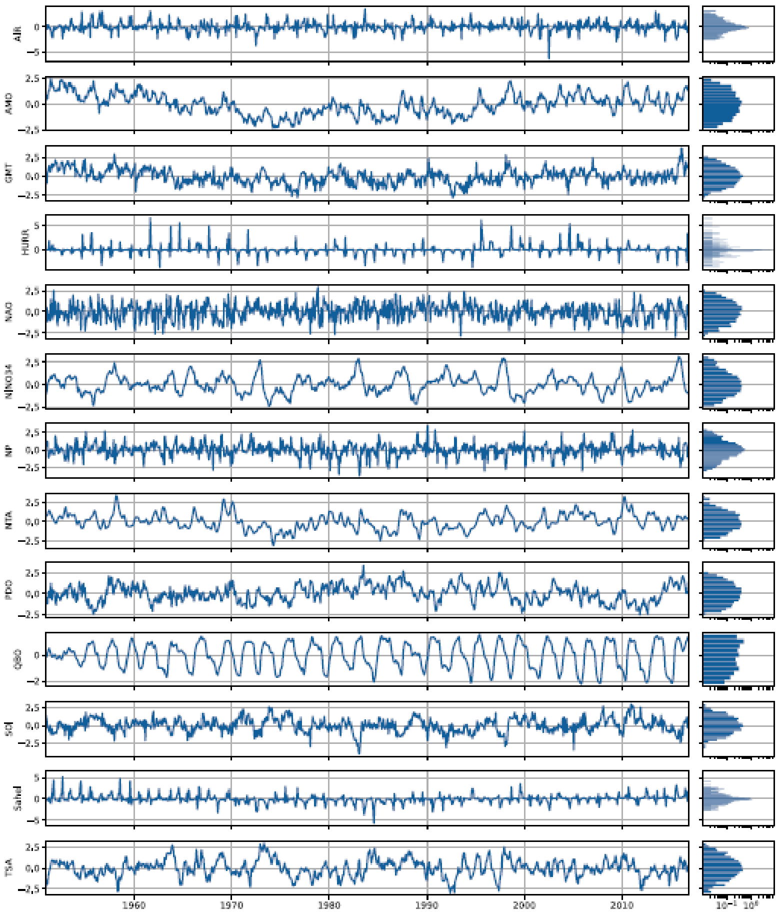


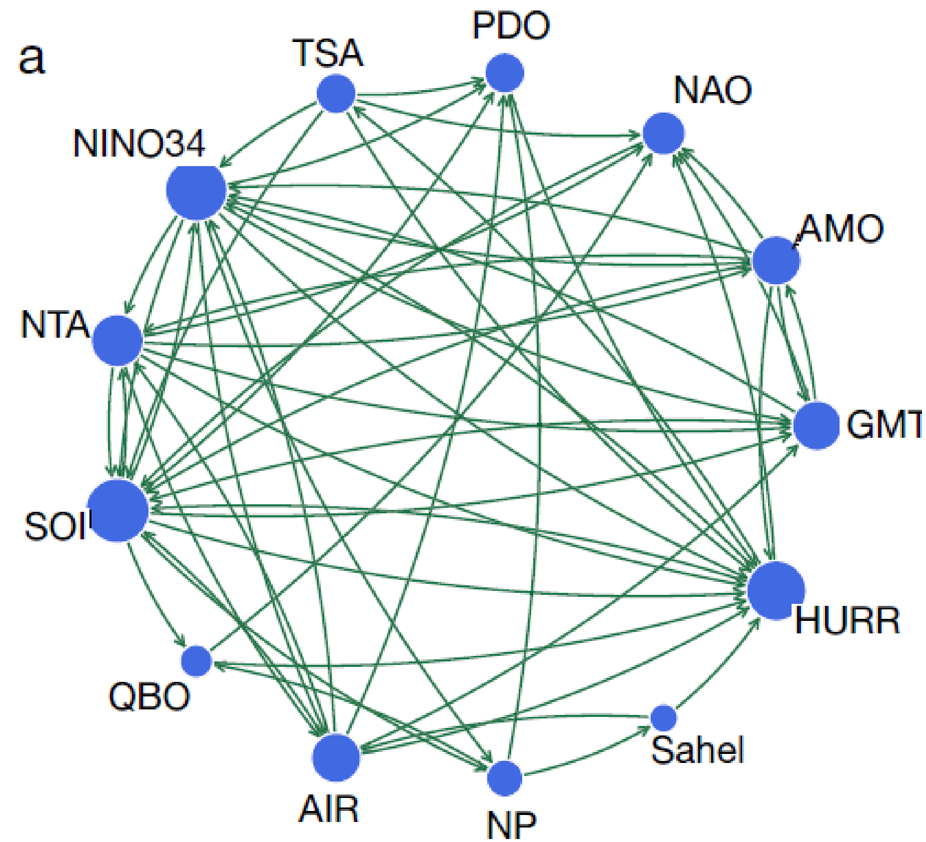
Longer time scale ⇒ increased connectivity

cristina.masoller@upc.edu @cristinamasoll1

Another approach: Directed network from climatic indices

Links defined with pTE, using different lags





R. Silini, G. Tirabassi, M Barreiro, L. Ferranti, C. Masoller, "*Assessing causal dependencies in climatic indices*", Climate Dynamics 61, 79 (2023).

cristina.masoller@upc.edu @cristinamasoll1

Take home messages

1. Hilbert analysis uncovers changes in large-scale patterns of atmospheric variability.
2. A reliable detection of bivariate (weak) climatic interactions is challenging; results depend on the method, data resolution, temporal scale, etc.
3. Next steps: compare different methods, identify high-order causal interactions, establish links to extreme events.

Nonlinear climate data analysis: very active and growing research field, with many applications and many challenges!

cristina.masoller@upc.edu @cristinamasoll1

Funding

"LINC"
Learning about Interacting Networks in Climate

CAFE
Climate Advanced Forecasting of sub-seasonal Extremes

Agència de Gestió d'Ajuts Universitaris i de Recerca

ICREA

GOBIERNO DE ESPAÑA
MINISTERIO DE CIENCIA E INNOVACIÓN

cristina.masoller@upc.edu @cristinamasoll1

Job announcement

Doctoral network BE-LIGHT offers
11 PhD positions, 3-year contracts

BE-LIGHT
Improving biomedical diagnosis
through light-based technologies
and machine learning

https://belightproject.eu/

cristina.masoller@upc.edu @cristinamasoll1