REDES Y CLIMA (y métodos de análisis de datos)

Cristina Masoller

Departamento de Fisica

Universitat Politecnica de Catalunya

Slides:
http://www.fisica.edu.uy/~cris/enredando.pdf

UNIVERSITAT POLITÈCNICA DE CATALUNYA
BARCELONATECH

Campus d'Excel·lència Internacional

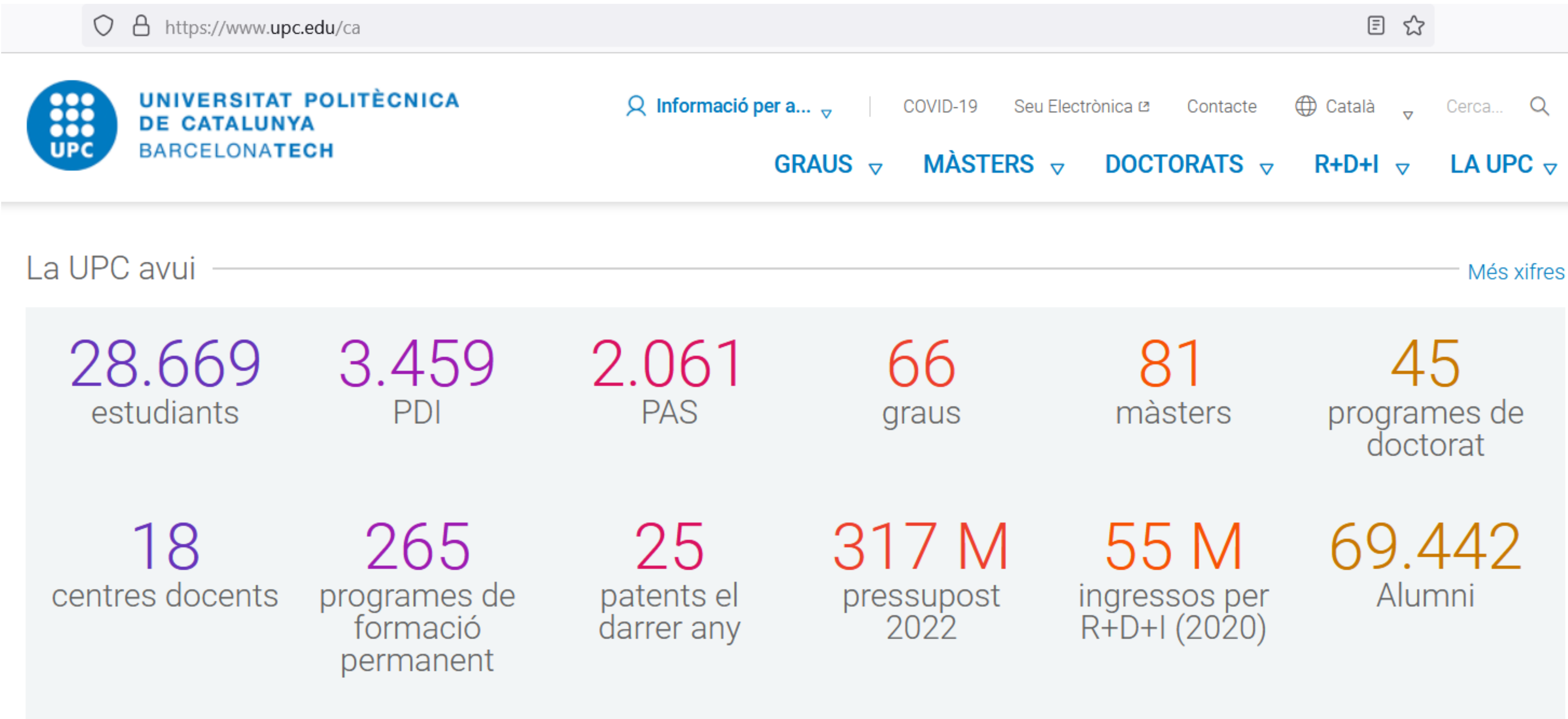
cristina.masoller@upc.edu @cristinamasoll1

Presentacion

- Professor of Physics, Universitat Politecnica de Catalunya.
- Originally from Montevideo, Uruguay.
- PhD in physics (nonlinear laser dynamics, Bryn Mawr College, USA).
- Research group: Dynamics, Nonlinear Optics and Lasers

cristina.masoller@upc.edu @cristinamasoll1

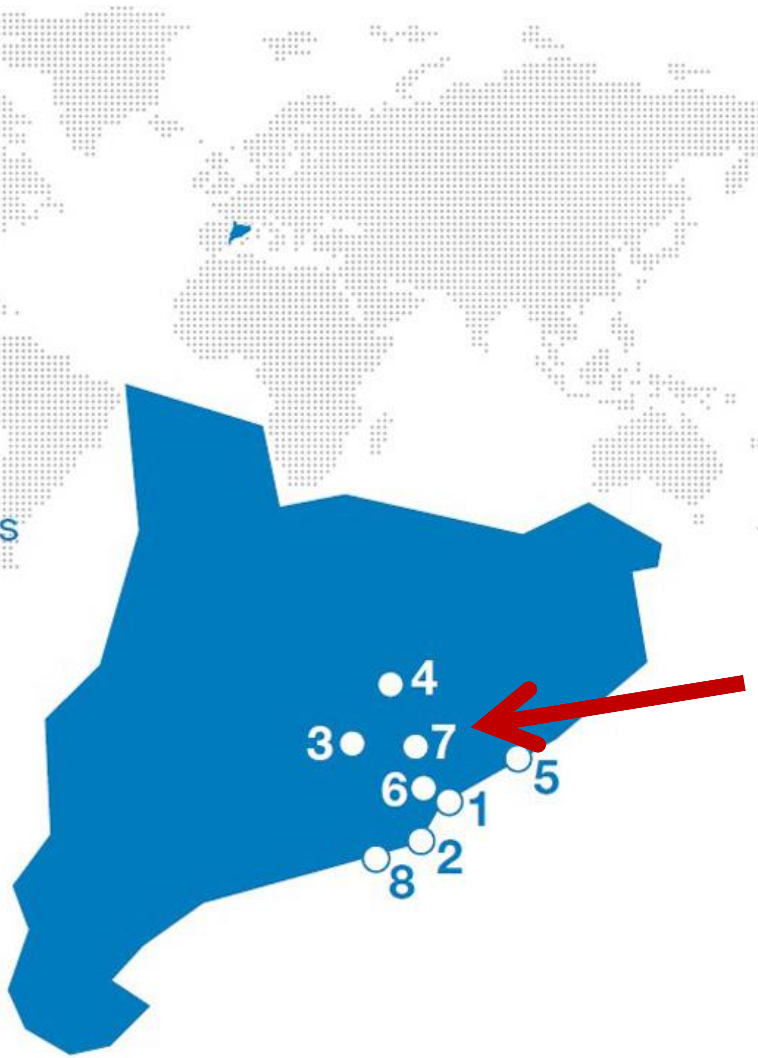
UPC: one of the largest technical universities in Spain



cristina.masoller@upc.edu @cristinamasoll1

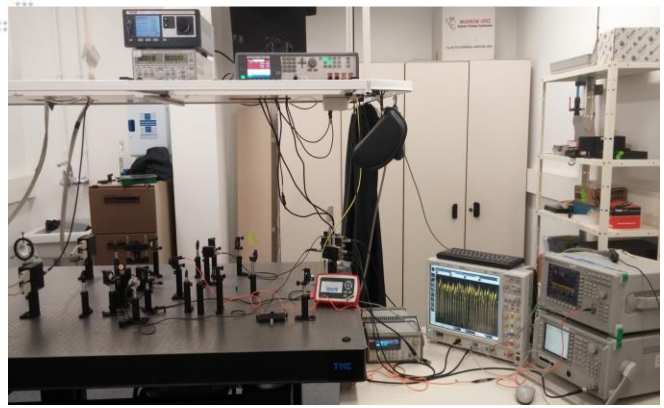
Where are we? UPC Campus Terrassa

1. Barcelona
2. Castelldefels
3. Igualada
4. Manresa
5. Mataró
6. Sant Cugat del Vallès
7. Terrassa
8. Vilanova i la Geltrú



Viernes, 25 de septiembre de 2009 Diari de Terrassa

El edificio Gaia centraliza grupos científicos consolidados y emergentes.



Laser lab in Gaia Building, UPC Campus Terrassa

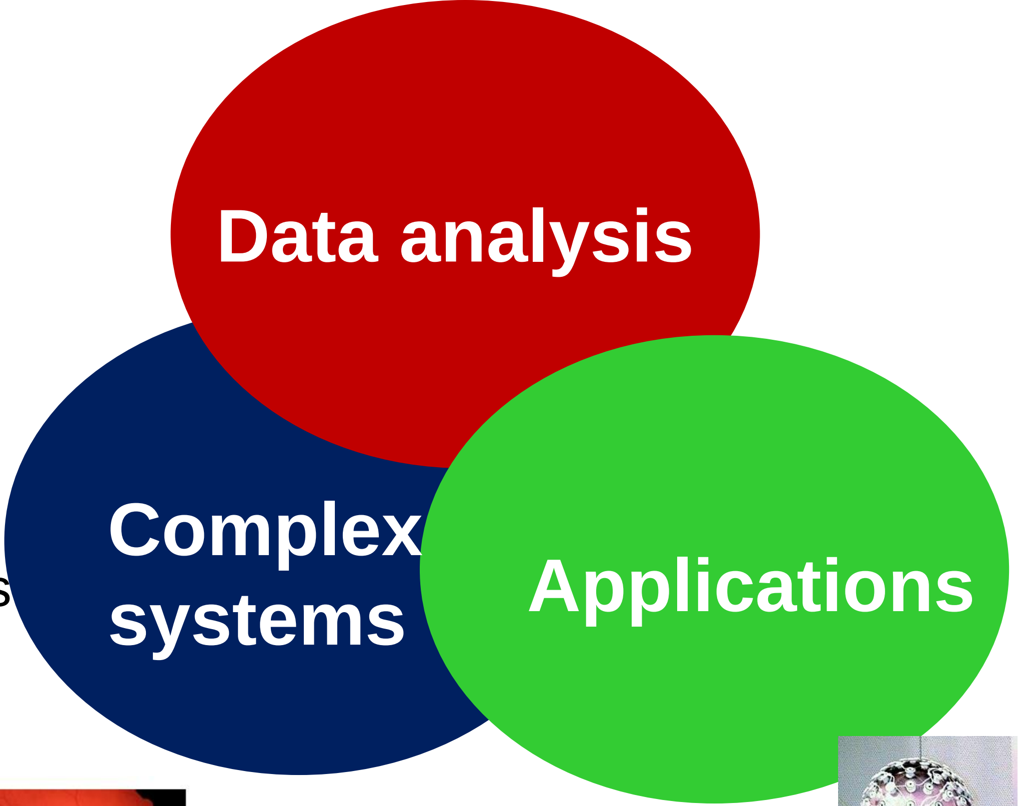
cristina.masoller@upc.edu @cristinamasoll1

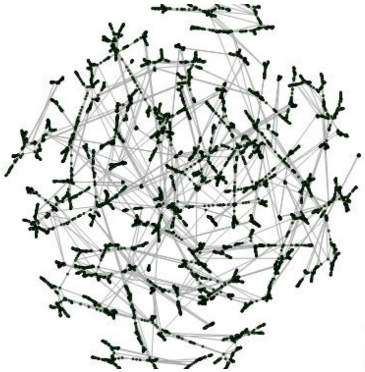
Research lines

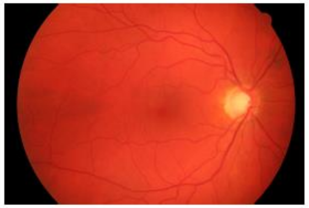
- Laser dynamics
- Neuronal dynamics
- Complex networks
- Data analysis

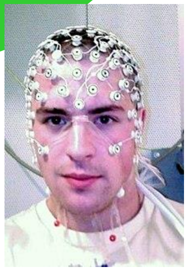
  climate and biomedical data

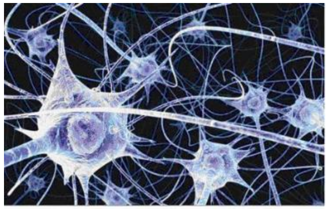
  tipping points, extreme events

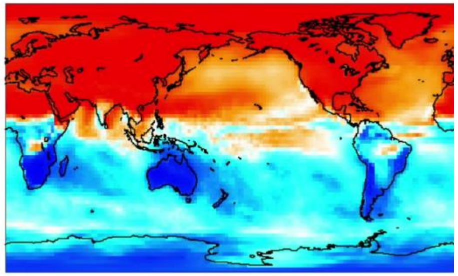


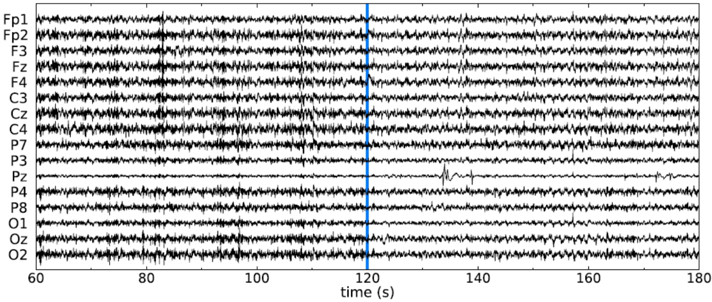








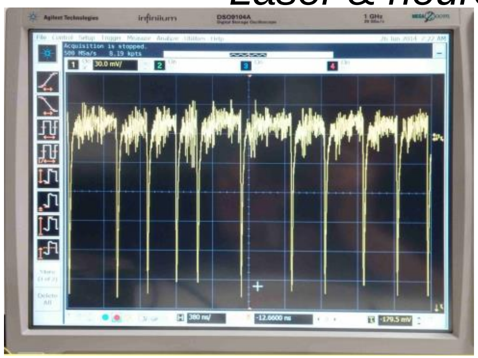


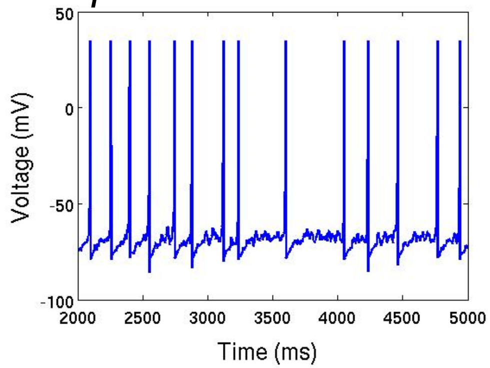


cristina.masoller@upc.edu @cristinamasoll1

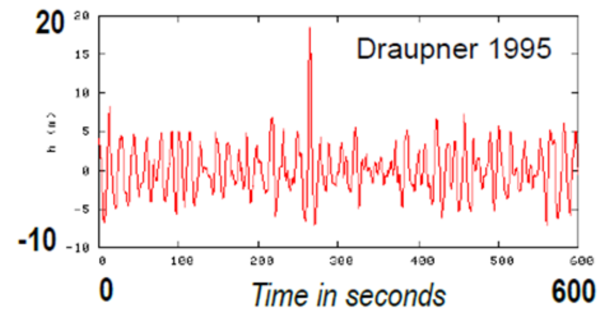
Lasers, neurons, climate, complex systems?

Laser & neuronal spikes

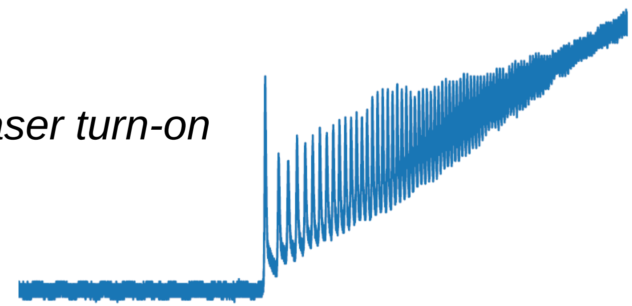




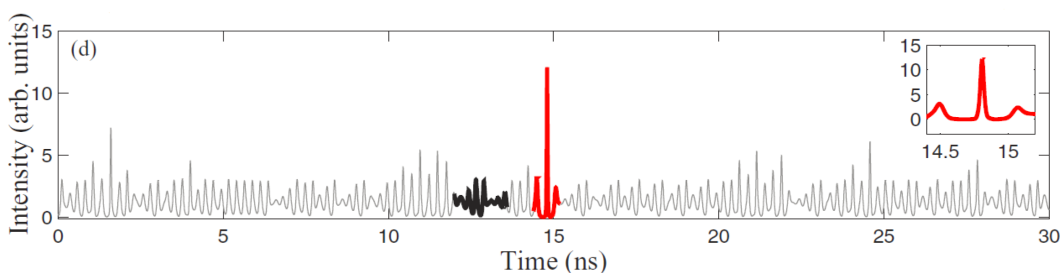
Ocean rogue wave (sea surface elevation in meters)



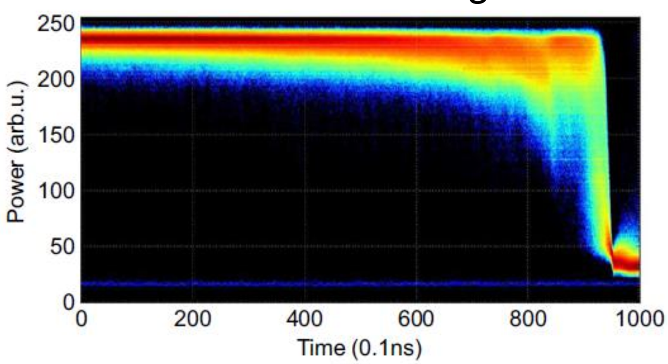
Laser turn-on



Extreme optical pulse (optical rogue wave)

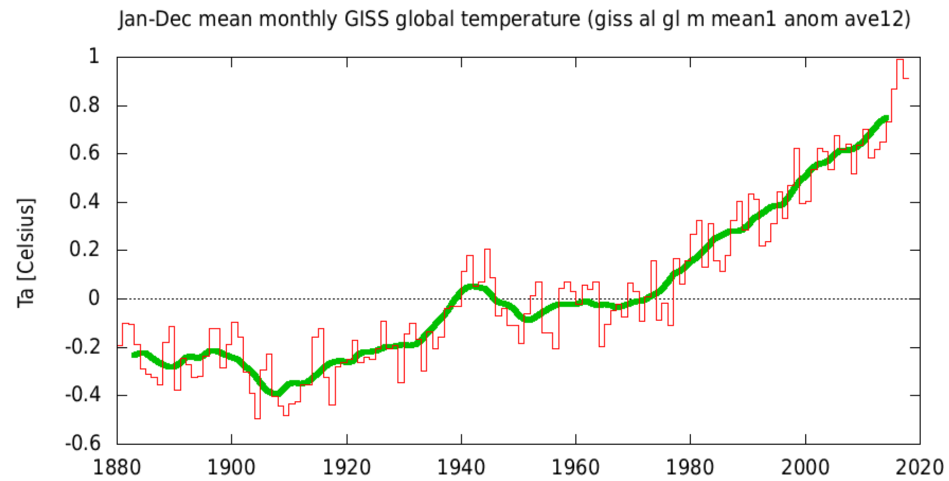


Polarization switching

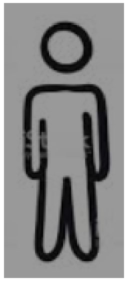


cristina.masoller@upc.edu @cristinamasoll1

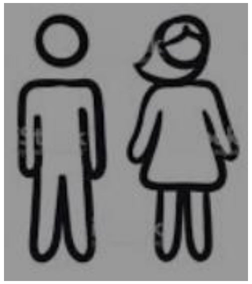
Methods of time series analysis $\{x_1, x_2, \ldots x_N\}$



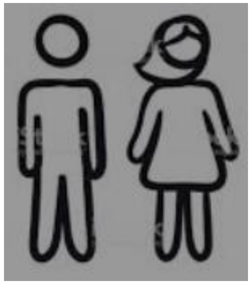
- Univariate analysis

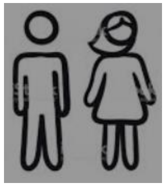


- Bivariate analysis



- Multivariate analysis
  - Complex networks

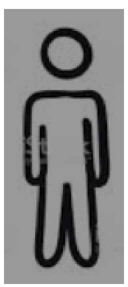




cristina.masoller@upc.edu @cristinamasoll1

1) “Univariate” time-series analysis tool: Hilbert analysis

(for oscillatory time series)

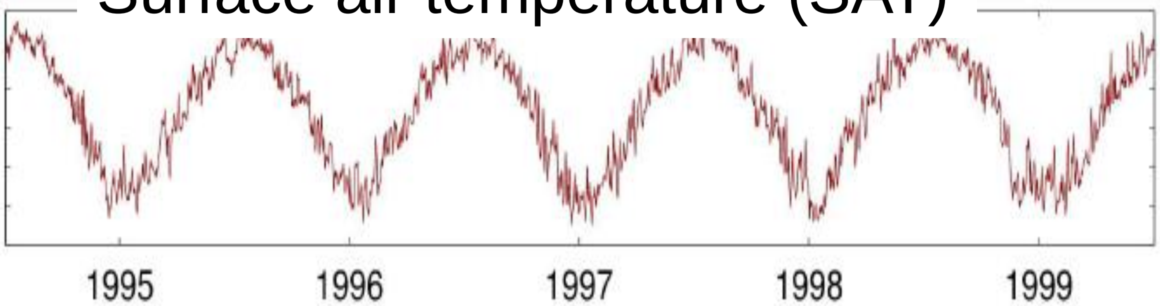


cristina.masoller@upc.edu @cristinamasoll1

The Hilbert Transform (HT)

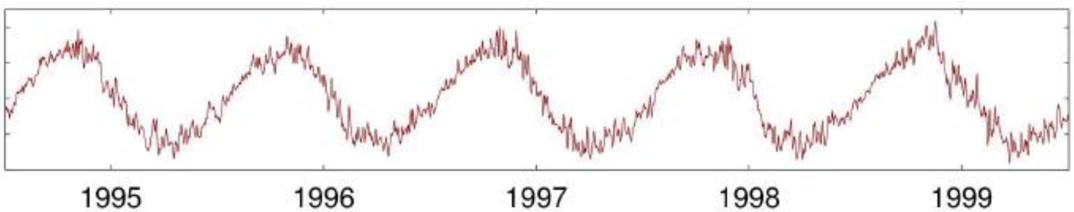
Surface air temperature (SAT)

x

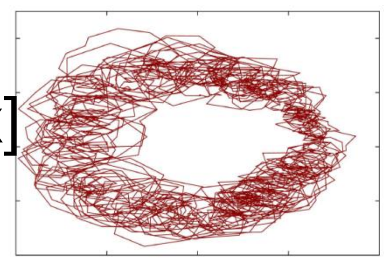


HT[sin(ωt)]=cos(ωt)

HT[x]



y=HT[x]



x

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

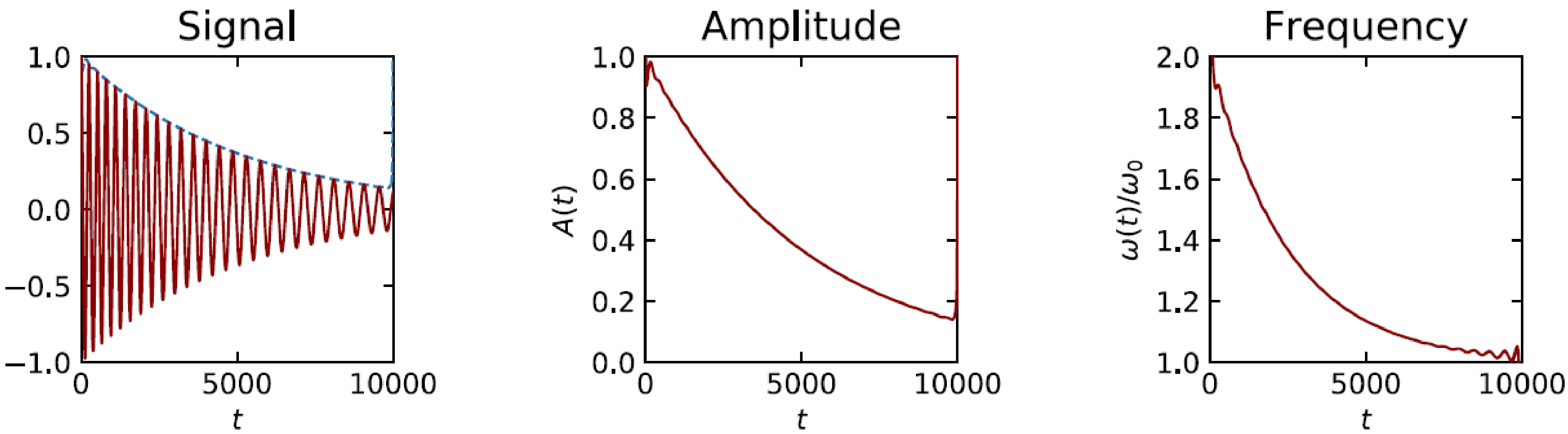
$$\varphi(t) = \arctan[y(t)/x(t)]$$

$\omega(t) = d\varphi/dt$

cristina.masoller@upc.edu @cristinamasoll1

Example

$$x(t) = e^{-\alpha t} \cos\left[\left(1 + e^{-2\alpha t}\right) \omega_0 t\right].$$

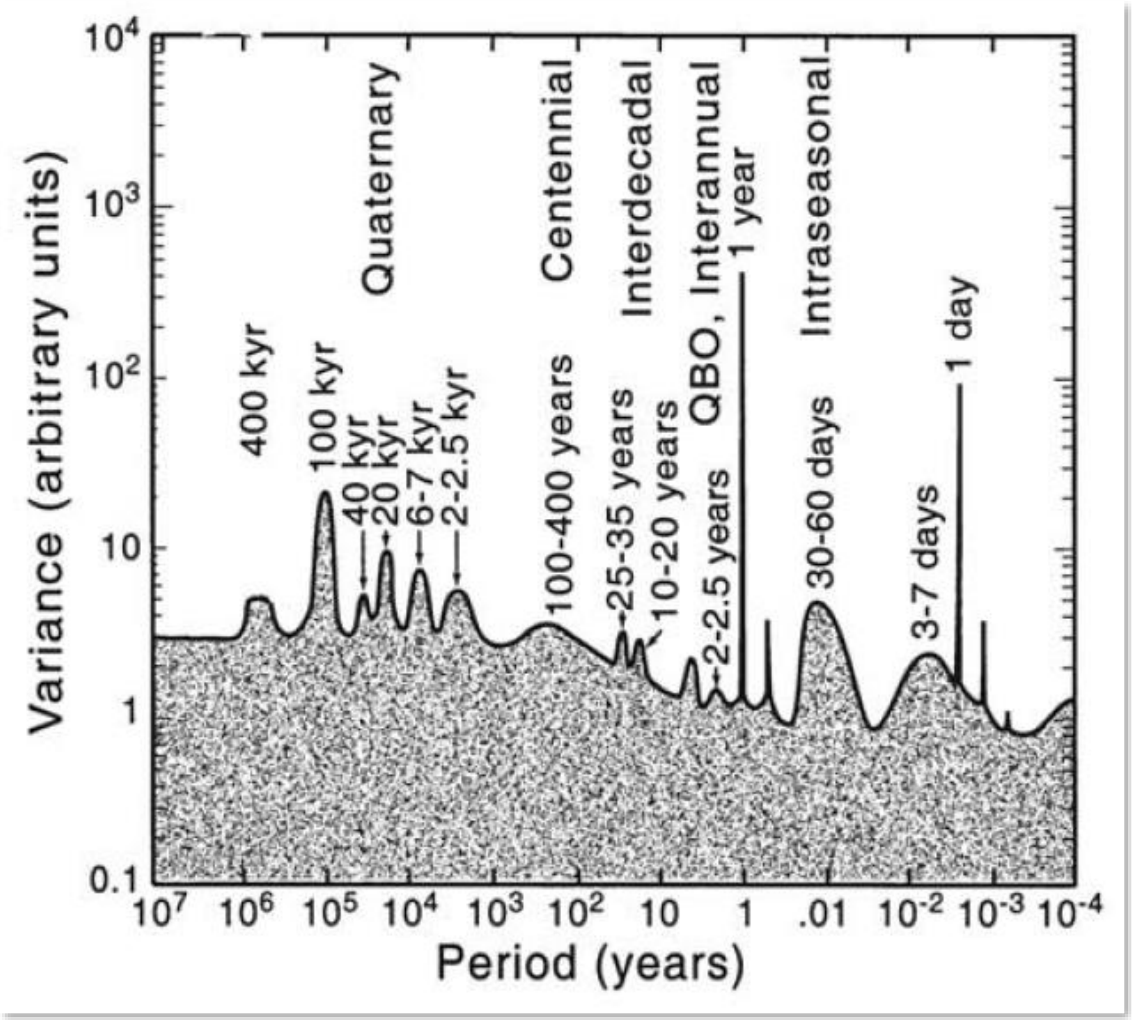


A word of warning: only if *x(t)* is a "narrow-band" signal then *a(t)* and $\omega(t) = d\varphi/dt$ have clear physical meaning

- ***a(t)*** is the envelope of *x(t)*
- ***ω(t)*** is the main frequency in the Fourier spectrum

cristina.masoller@upc.edu @cristinamasoll1

PROBLEM ! Climatic time series are NOT narrow-band.



An "artist's representation" of the power spectrum of climate variability (M. Ghil 2002).

cristina.masoller@upc.edu @cristinamasoll1

Solution ?

- Isolate a narrow frequency band (usual solution for EEG signals).

- However, I will show that HT directly applied to raw surface air temperature (SAT) returns meaningful results.

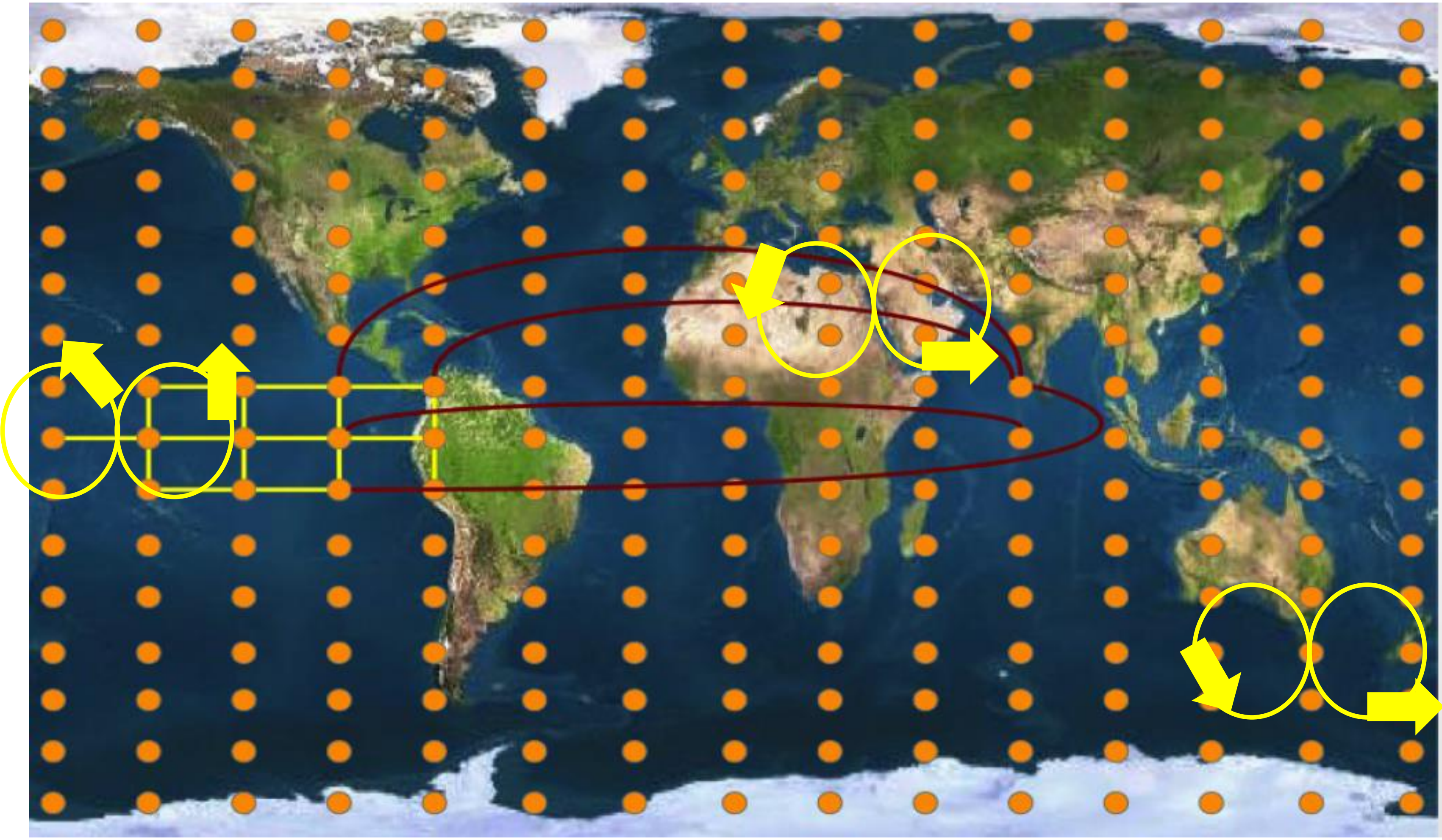
SAT data

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

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

cristina.masoller@upc.edu @cristinamasoll1

Oscillatory time series of 13700 data points in each "node" (more than 10000 nodes)



Credit: G. Tirabassi

cristina.masoller@upc.edu @cristinamasoll1

Where does the data come from?

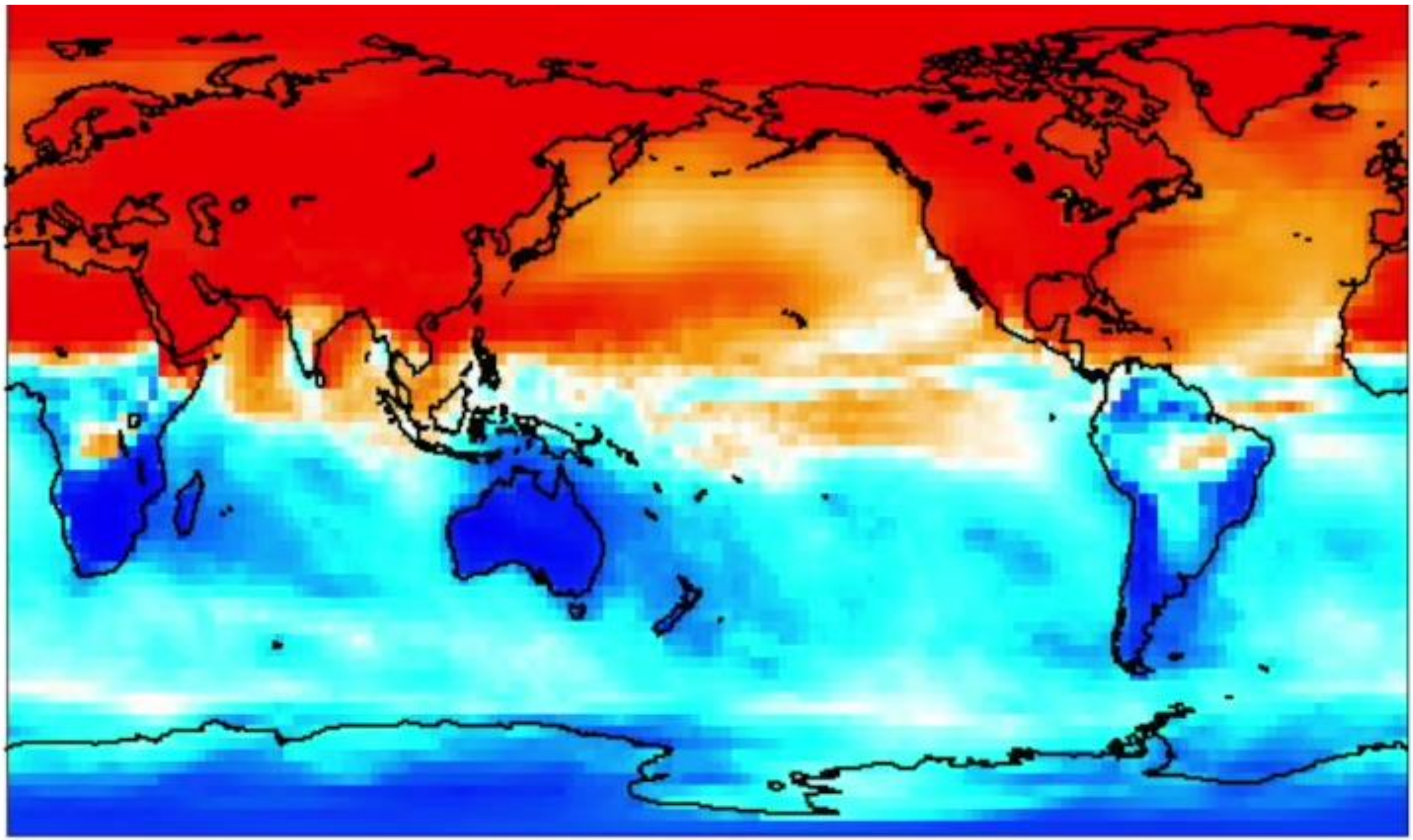
- European Centre for Medium-Range Weather Forecasts (ECMWF).
- Freely available.
- <u>Reanalysis</u> = general atmospheric circulation model feed with empirical data, where and when available (data assimilation).

ECMWF datacenter,
Reading, UK, March 2022
Meeting of EU project CAFE

cristina.masoller@upc.edu @cristinamasoll1

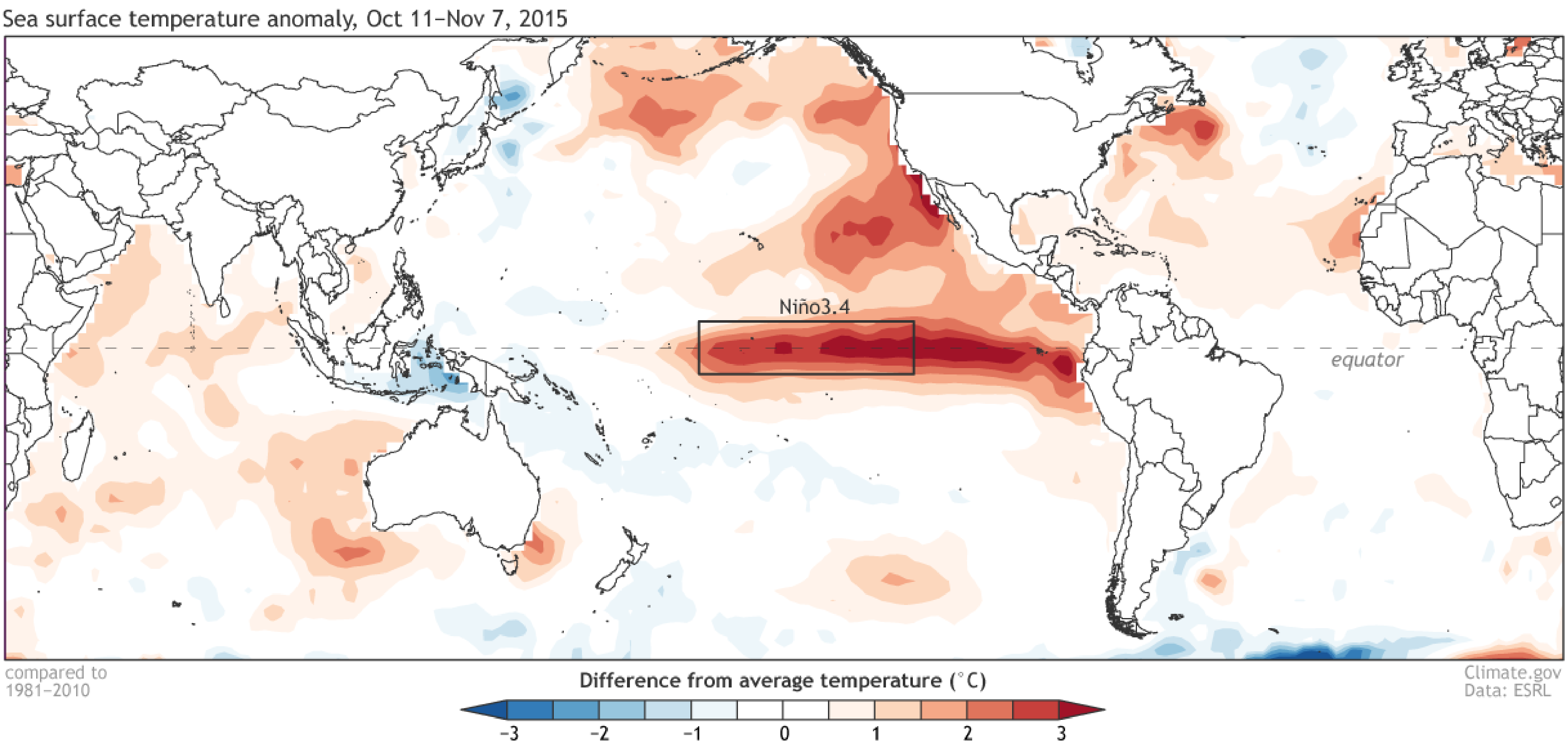
Which information carries the Hilbert phase?

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



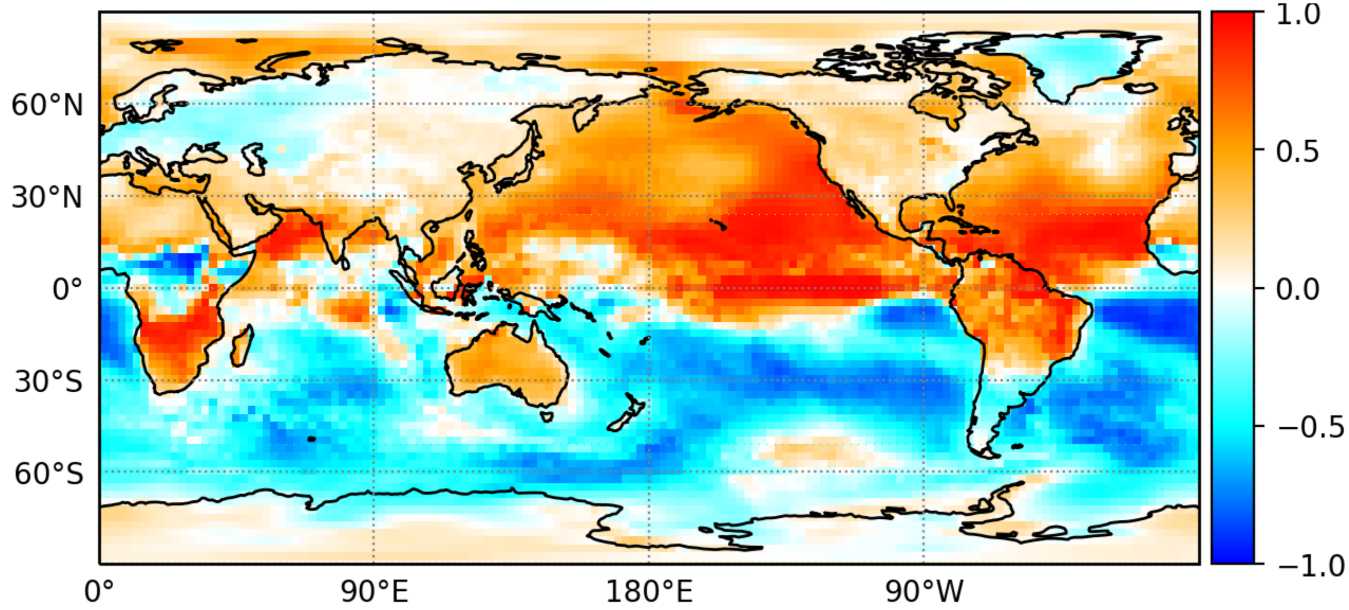
cristina.masoller@upc.edu @cristinamasoll1

ENSO (El niño / southern oscillation)

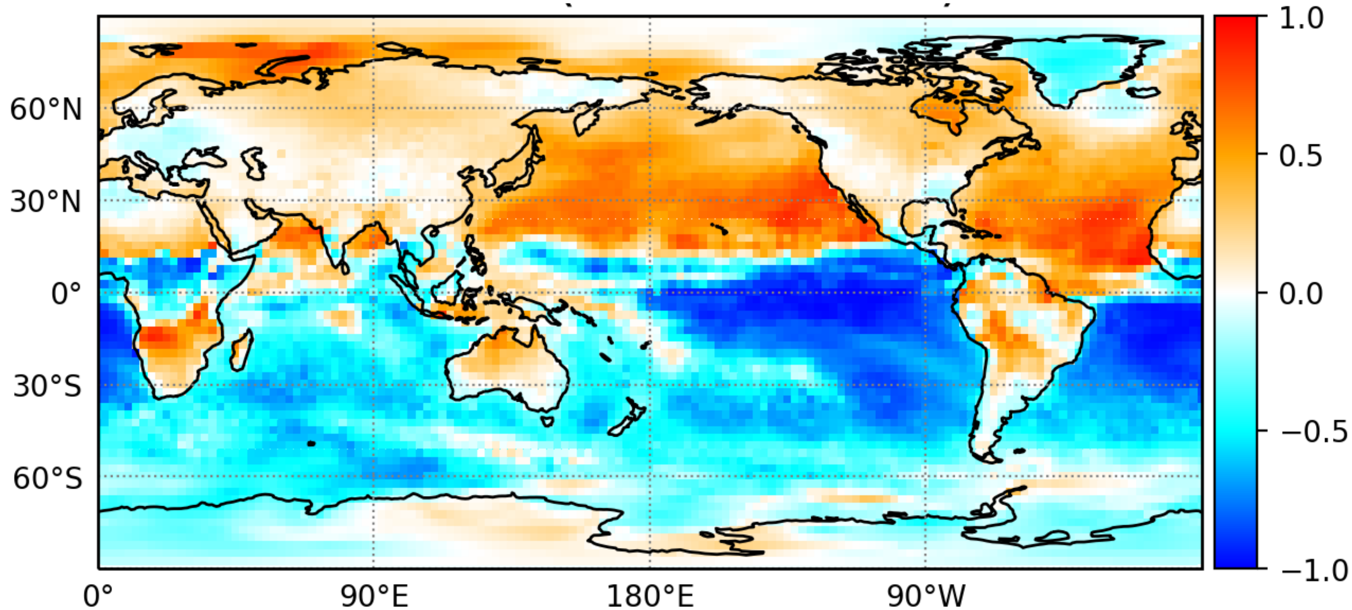


cristina.masoller@upc.edu @cristinamasoll1

El Niño period
(October 2015)

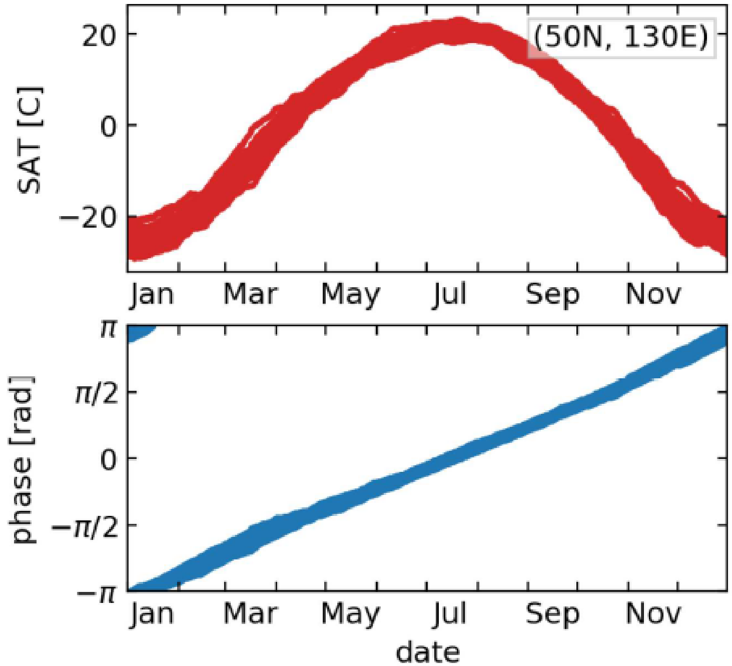


La Niña period
(Octubre 2011)

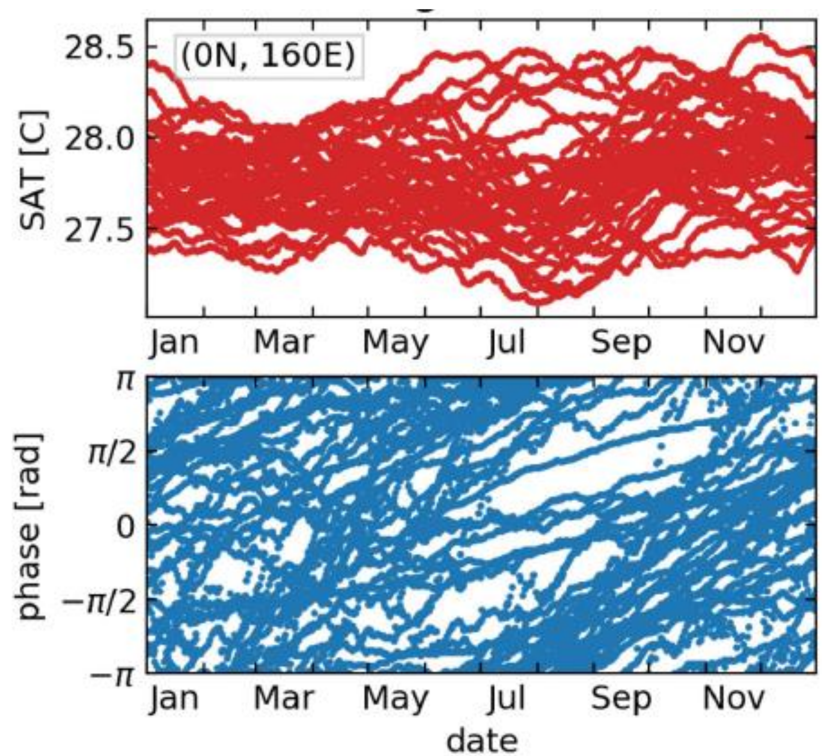


cristina.masoller@upc.edu @cristinamasoll1

Hilbert phase vs day of the year (1979 – 2016)



in a continental
“regular” region / node



And in an irregular region?

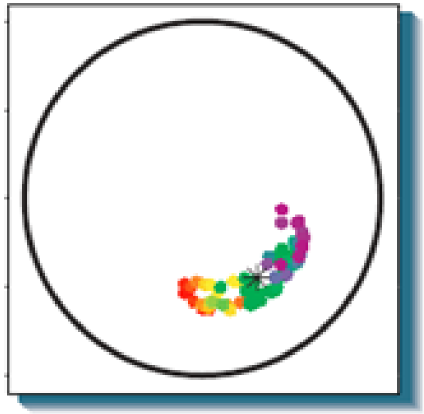
D. A. Zappala, M. Barreiro, C. Masoller, Chaos 29, 051101 (2019).

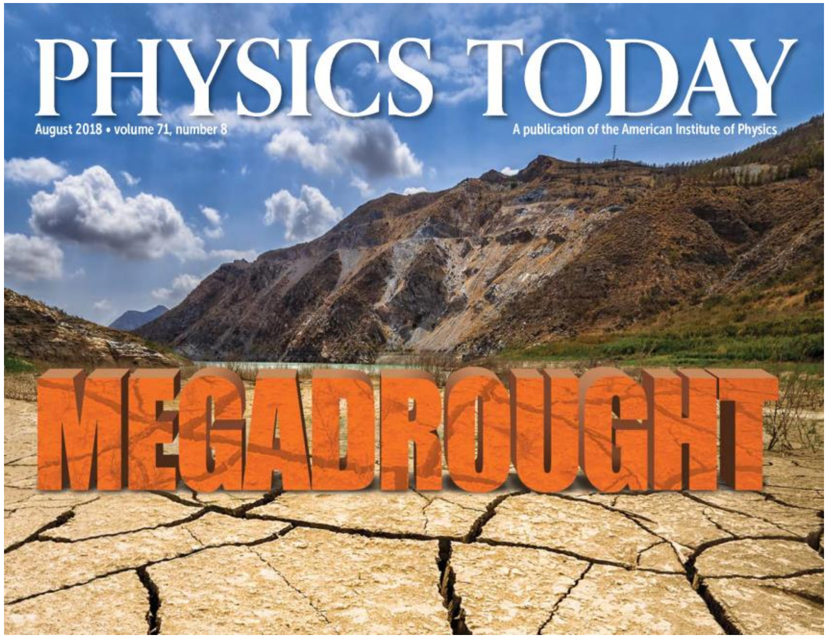
cristina.masoller@upc.edu @cristinamasoll1

Questions?

1) Can we use the Hilbert amplitude, phase, frequency, to identify and quantify regional "climate change"?

2) Can we identify and quantify synchronized oscillations?





cristina.masoller@upc.edu @cristinamasoll1

Relative decadal variations in each region ("node")

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

$$\frac{\Delta a}{\langle a \rangle_{2016-1979}}$$

Relative variation is considered **significant** if:

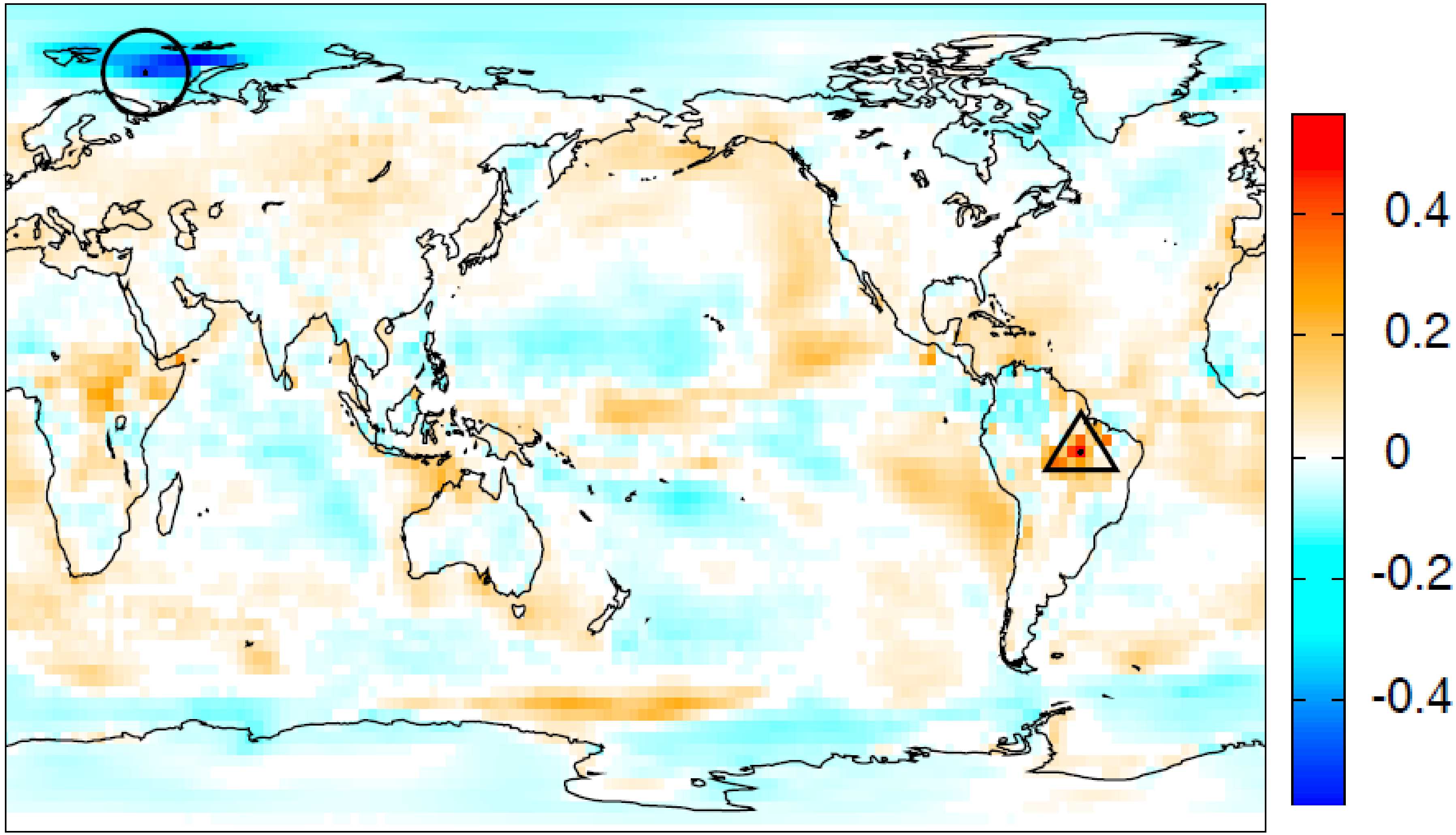
$$\frac{\Delta a}{\langle a \rangle} \geq \langle . \rangle_s + 2\sigma_s \quad \text{or} \quad \frac{\Delta a}{\langle a \rangle} \leq \langle . \rangle_s - 2\sigma_s$$

100 "surrogates"

G. Lancaster et al, "Surrogate data for hypothesis testing of physical systems", Physics Reports 748, 1 (2018).

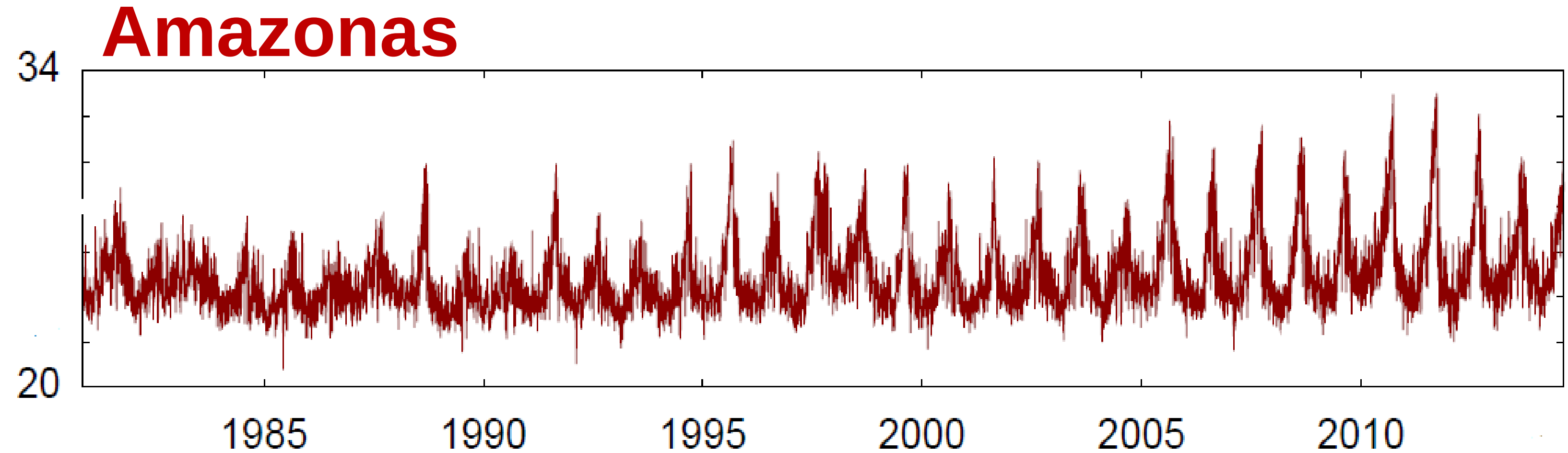
cristina.masoller@upc.edu **@cristinamasoll1**

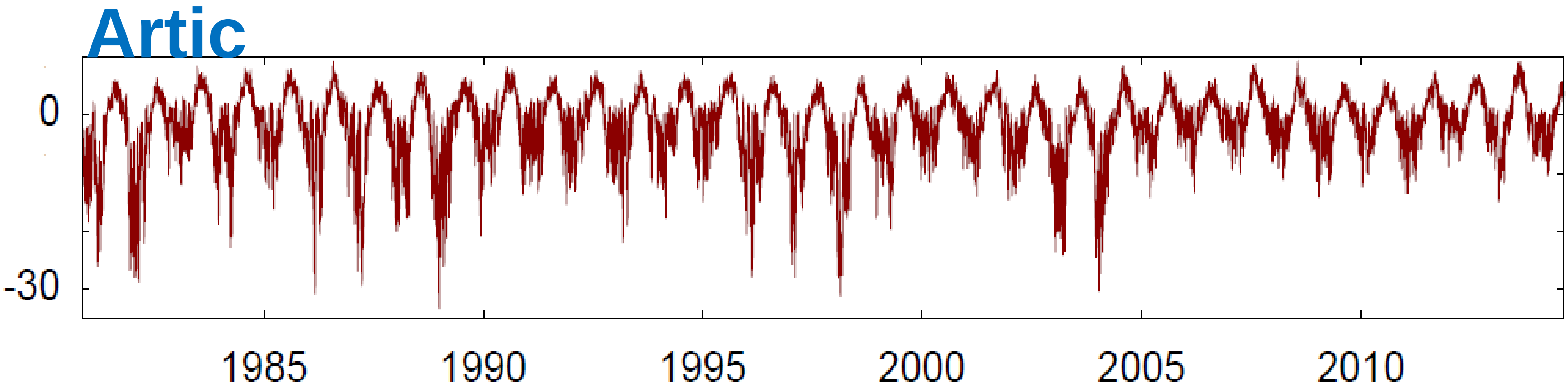
Relative decadal variations



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

cristina.masoller@upc.edu @cristinamasoll1





- **Decrease of precipitation**: the solar radiation that is not used for evaporation is used to heat the ground.

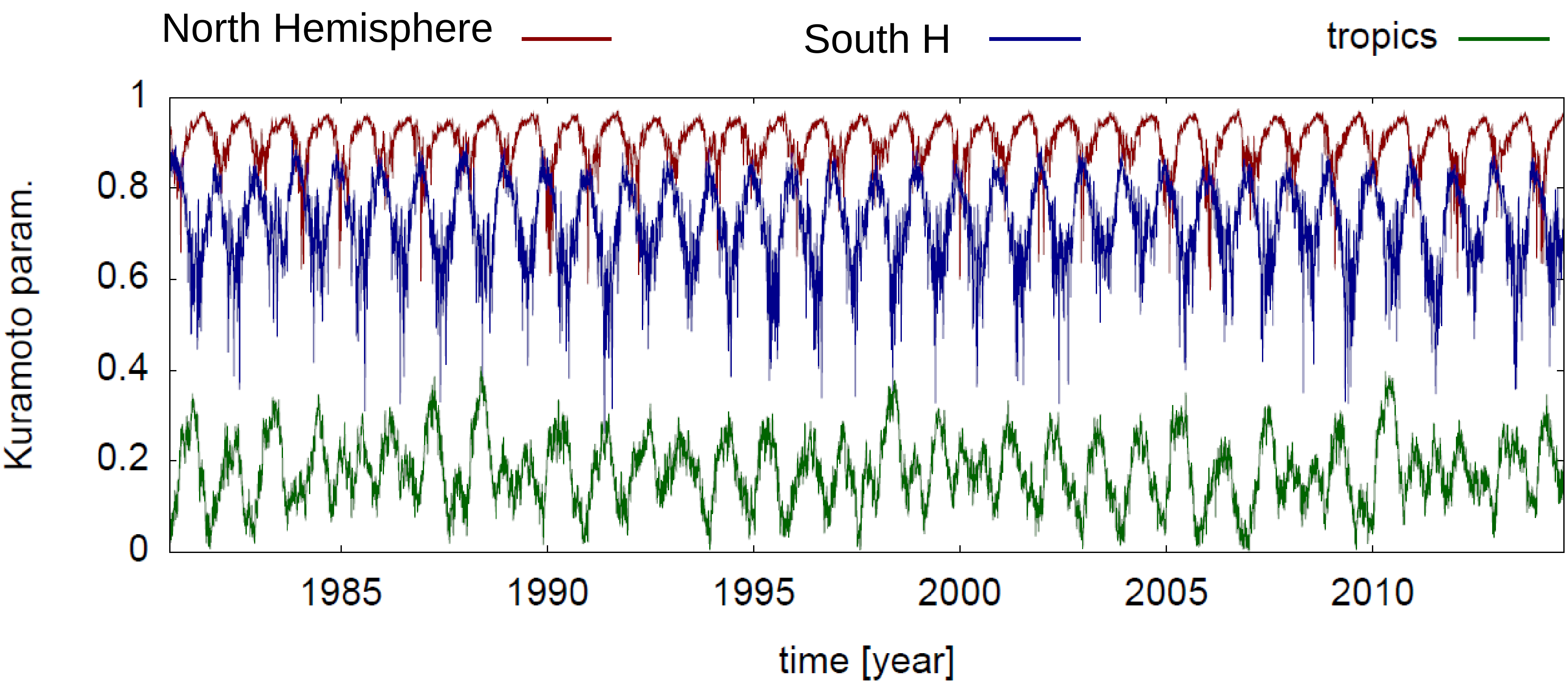
- **Melting of sea ice**: during winter the air temperature is mitigated by the sea and tends to be more moderated.

cristina.masoller@upc.edu @cristinamasoll1

Quantifying synchronization of air temperature oscillations

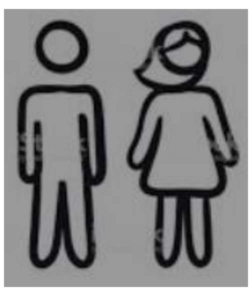
Kuramoto order parameter

$$r(t) = \left| \frac{1}{N} \sum_{j=1}^{N} e^{i\theta_j(t)} \right|$$



cristina.masoller@upc.edu @cristinamasoll1

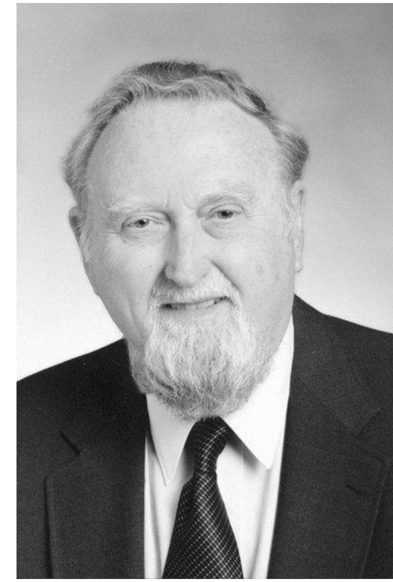
2) “Bivariate” time-series analysis tool: pseudo Transfer Entropy (pTE)



- For time series with a distribution of values that is approximately Gaussian
- It provides “causal” information (whether or not knowledge of Y improves the forecast of X)

cristina.masoller@upc.edu @cristinamasoll1

Granger Causality



Hypothesis: X_1 and X_2 can be described by autoregressive linear models

past of X_1 — Residual error

$$X_1(t) = \sum_{j=1}^{p} A_{11,j}\, X_1(t-j) \qquad + E_1(t)$$

past of X_1 — past of X_2 — Residual error

$$X_1(t) = \sum_{j=1}^{p} A_{11,j}\, X_1(t-j) + \sum_{j=1}^{p} A_{12,j} X_2(t-j) + E'_1(t)$$

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

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

Software: Fulton, C, https://github.com/statsmodels/statsmodel (2020)

cristina.masoller@upc.edu @cristinamasoll1

Entropy (disorder) and information

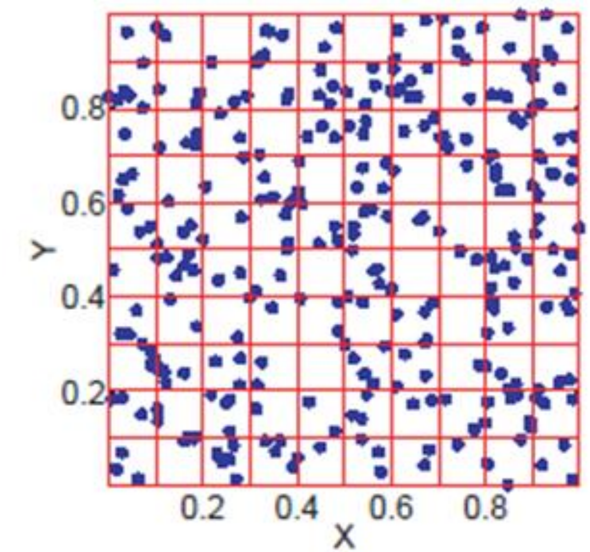
HIGH entropy LOW information

LOW entropy HIGH information

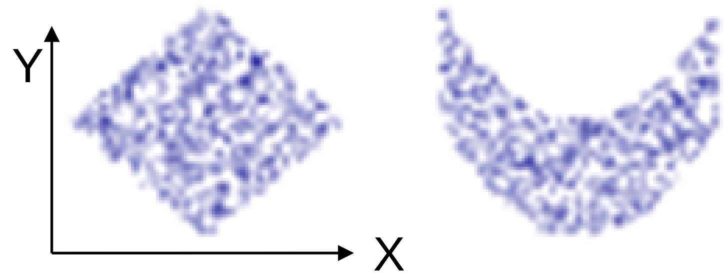
https://imgur.com/gallery/Otg97

cristina.masoller@upc.edu @cristinamasoll1

Mutual Information (MI)



Quantifies the reduction in uncertainty of one variable by knowing the other variable.



- *MI is calculated from probability distributions, p(X), p(Y) and p(X,Y)*
- *If X, Y are independent, MI = 0*, else ***MI >0***
- For Gaussian distributions: MI = $-1/2 \log(1-\rho^2)$

where ρ is the cross-correlation coefficient.

K. Hlaváčková-Schindler et al. / Physics Reports 441 (2007) 1–46

cristina.masoller@upc.edu @cristinamasoll1

Transfer Entropy (TE) and Directionality Index (DI)

- TE: is the Conditional Mutual information, given the "past" of one of the variables.

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

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

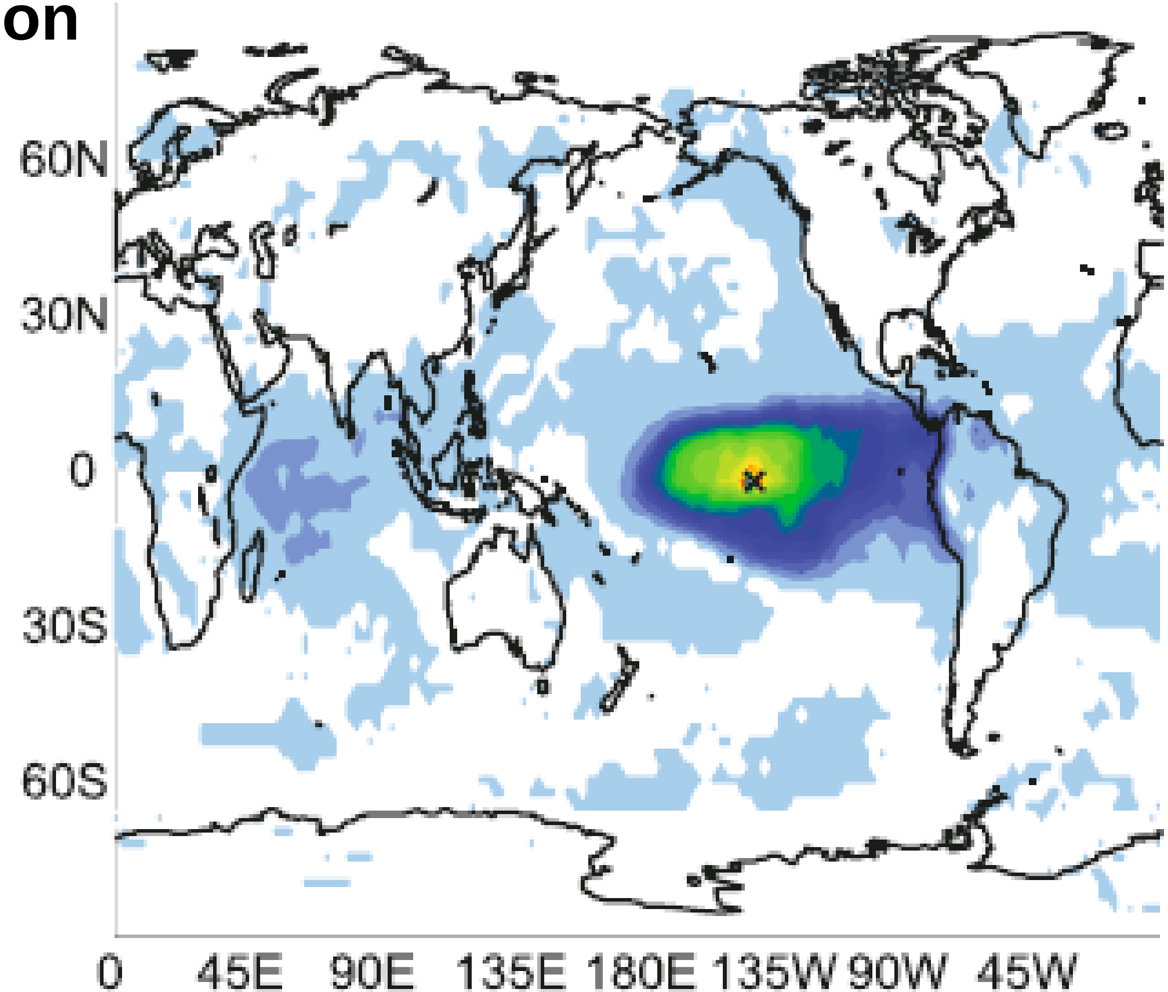
- MI (x,y) = MI (y,x) but TE (x,y) $\neq$ TE(y,x)

- Directionality Index: TE(x,y)-TE(y,x)

K. Hlaváčková-Schindler et al. / Physics Reports 441 (2007) 1–46

cristina.masoller@upc.edu @cristinamasoll1

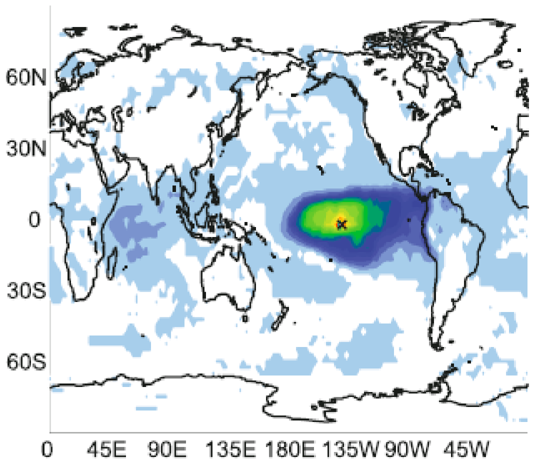
Analysis of surface air temperature anomalies: Mutual Information

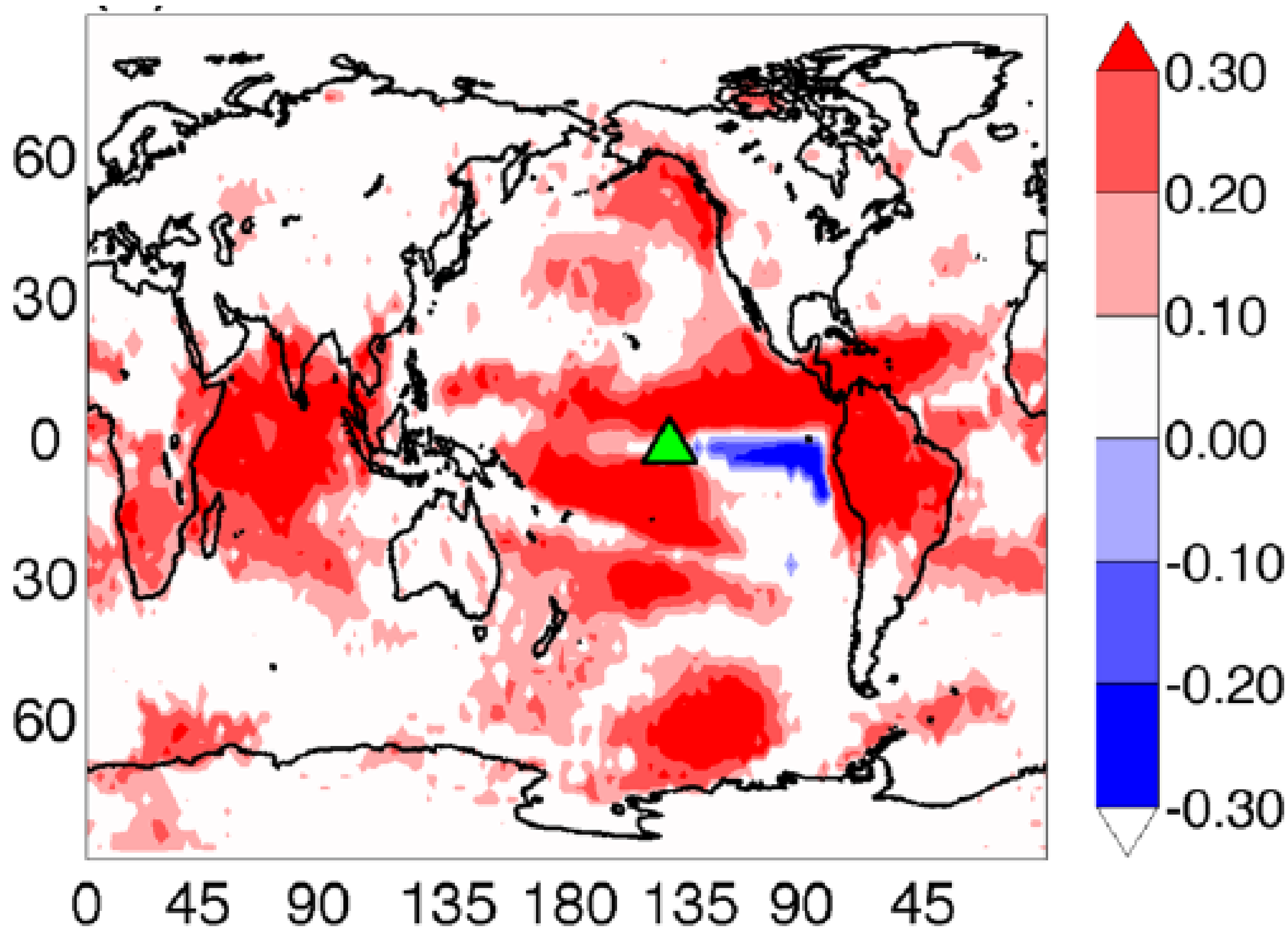


J. I. Deza, M. Barreiro, and C. Masoller, "*Assessing the direction of climate interactions by means of complex networks and information theoretic tools*", Chaos 25, 033105 (2015).

cristina.masoller@upc.edu @cristinamasoll1

Analysis of surface air temperature anomalies: Directionality Index

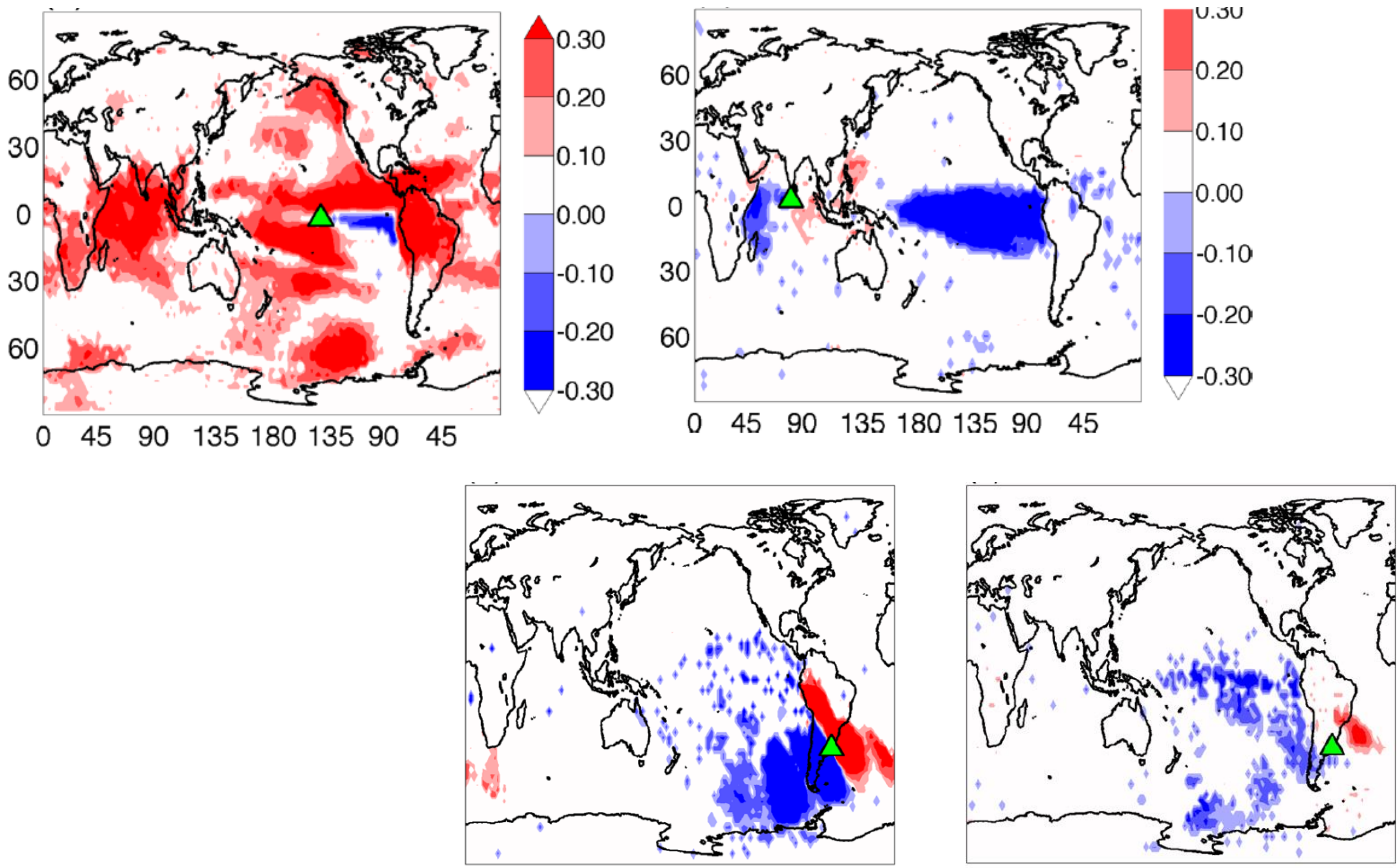




J. I. Deza, M. Barreiro, and C. Masoller, Chaos 25, 033105 (2015).

cristina.masoller@upc.edu @cristinamasoll1

Analysis of surface air temperature anomalies: Directionality Index



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

cristina.masoller@upc.edu @cristinamasoll1

Problem: Transfer Entropy is computationally demanding

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

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

Data Generating Processes and Performance Quantification

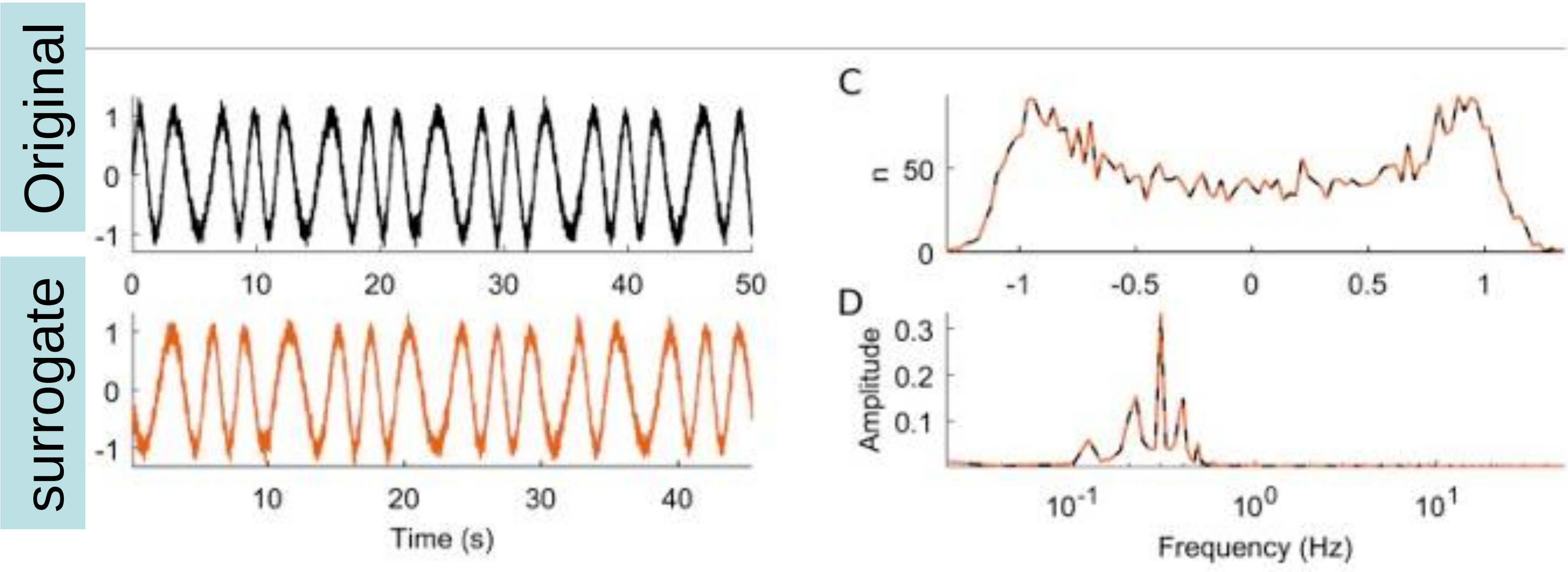
Power: True Positives

Size: False Positives

	Model	
Y X	M0 M1 M2	$x_t = (0.01 + 0.5\, x_{t-1}^2)^{0.5} + E_{1t}$ $\quad y_t = 0.5\, y_{t-1} + E_{2t}$
Y → X	M3 M4 M5 M6 M7 M8 M9 M10 M11 M12	$x_t = 0.6\, x_{t-1} + 0.5\, y_{t-1} + E_{1t}$ $\quad y_t = 0.5\, y_{t-1} + E_{2t}$
Y ⇆ X	M13 M14	$x_t = 0.15\, x_{t-1} + 0.7\, y_{t-1} + E_{1t}$ $y_t = 0.1\, y_{t-1} + 0.8\, x_{t-1} + E_{2t}$

cristina.masoller@upc.edu @cristinamasoll1

Time-shifted surrogates: “cheap” option for causality testing



Quiroga R.Q., Kraskov A., Kreuz T., Grassberger P. *Performance of different synchronization measures in real data: A case study on electroencephalographic signals,* Phys. Rev. E, 65 (4) (2002)

cristina.masoller@upc.edu @cristinamasoll1

Results

	Model	pTE	
		$Y \rightarrow X$	$X \rightarrow Y$
	M0	3.8	3.9
Y X	M1	2.3	2.6
	M2	4.2	4.7
	M3	100	4.5
	M4	80.7	3.8
	M5	100	2.2
	M6	100	1.8
	M7	100	2.8
$Y \rightarrow X$	M8	100	4.5
	M9	100	0.1
	M10	62.6	3.1
	M11	46.1	43.1
	M12	99.9	1.0
	M13	100	100
$Y \leftrightarrows X$	M14	100	100

cristina.masoller@upc.edu @cristinamasoll1

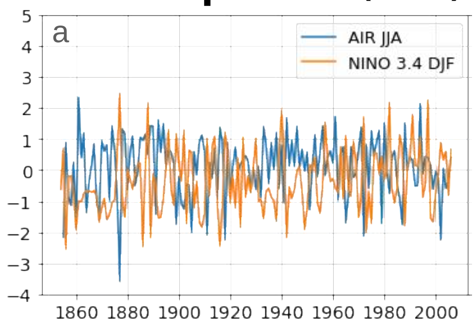
Comparison with Granger Causality and Transfer Entropy

	Model	pTE		GC		TE		DI		
		$Y \to X$	$X \to Y$	$Y \to X$	$X \to Y$	$Y \to X$	$X \to Y$	pTE	GC	TE
	M0	3.8	3.9	5.1	5.0	4.4	4.4	−0.01	0.01	0.00
Y X	M1	2.3	2.6	3.3	3.1	100	100	−0.06	0.03	0.00
	M2	4.2	4.7	5.5	5.9	4.7	4.9	−0.06	−0.04	−0.02
	M3	100	4.5	100	4.8	70.2	5.6	0.91	0.91	0.85
	M4	80.7	3.8	84.2	4.9	96.0	4.7	0.91	0.89	0.91
	M5	100	2.2	100	3.1	100	3.8	0.96	0.94	0.93
	M6	100	1.8	100	2.8	100	4.3	0.96	0.95	0.92
	M7	100	2.8	100	3.4	100	4.0	0.95	0.93	0.92
$Y \to X$	M8	100	4.5	100	5.6	100	100	0.91	0.89	0.00
	M9	100	0.1	100	0.1	100	100	1.00	1.00	0.00
	M10	62.6	3.1	67.3	4.3	12.2	4.5	0.91	0.88	0.46
	M11	46.1	43.1	53.1	49.8	37.8	45.0	0.03	0.03	−0.09
	M12	99.9	1.0	100	0.9	100	0	1.0	1.0	1.0
	M13	100	100	100	100	100	100	0.00	0.00	0.00
$Y \leftrightarrows X$	M14	100	100	100	100	100	100	0.00	0.00	0.00

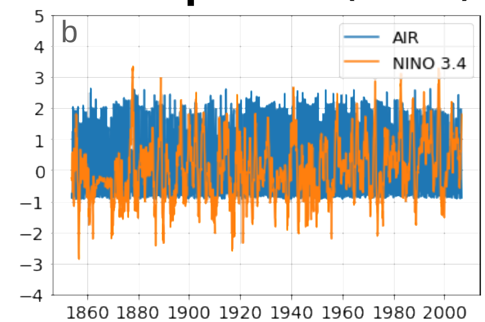
cristina.masoller@upc.edu @cristinamasoll1

Application to real data NINO3.4 ← → All India Rainfall

Yearly sampled (152)



Monthly sampled (1836)



IAAF

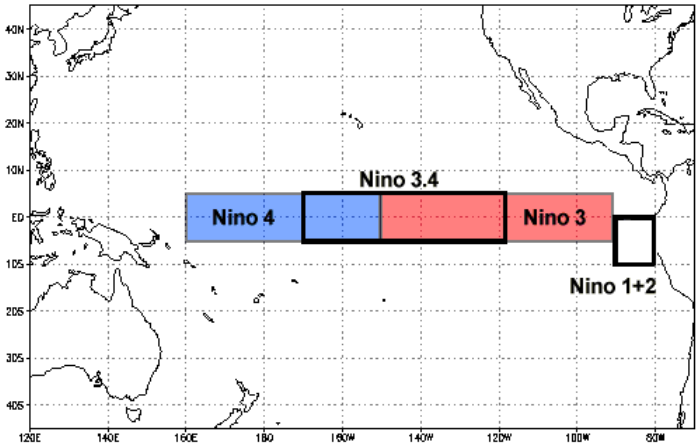
	Yearly sampled (152)	Monthly sampled (1836)
pTE	**NINO3.4 ⟶ AIR** 0.04 s	**NINO3.4 ⟵ AIR** 0.5 s
GC	**NINO3.4 ⟶ AIR** 0.4 s	**NINO3.4 ⟵ AIR** 0.9 s
TE	**NINO3.4 ⇆ AIR** 1 s	**NINO3.4 ⇆ AIR** 68 s

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).

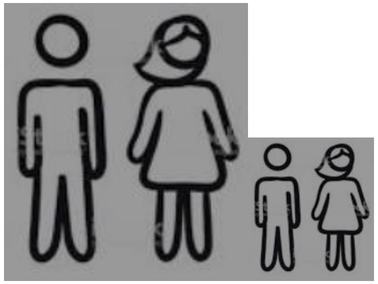
(without “surrogate” analysis)

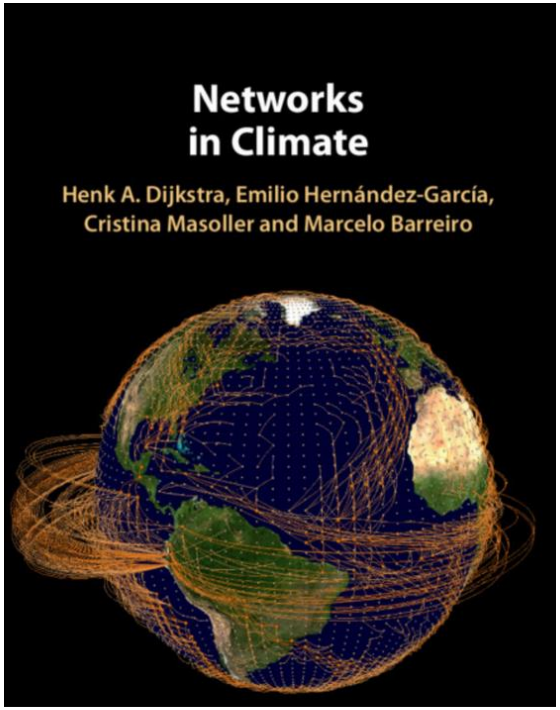
But, there is a price to pay, no “free lunch”.

https://github.com/riccardosilini/pTE

cristina.masoller@upc.edu @cristinamasoll1

3) "Multivariate" time-series analysis

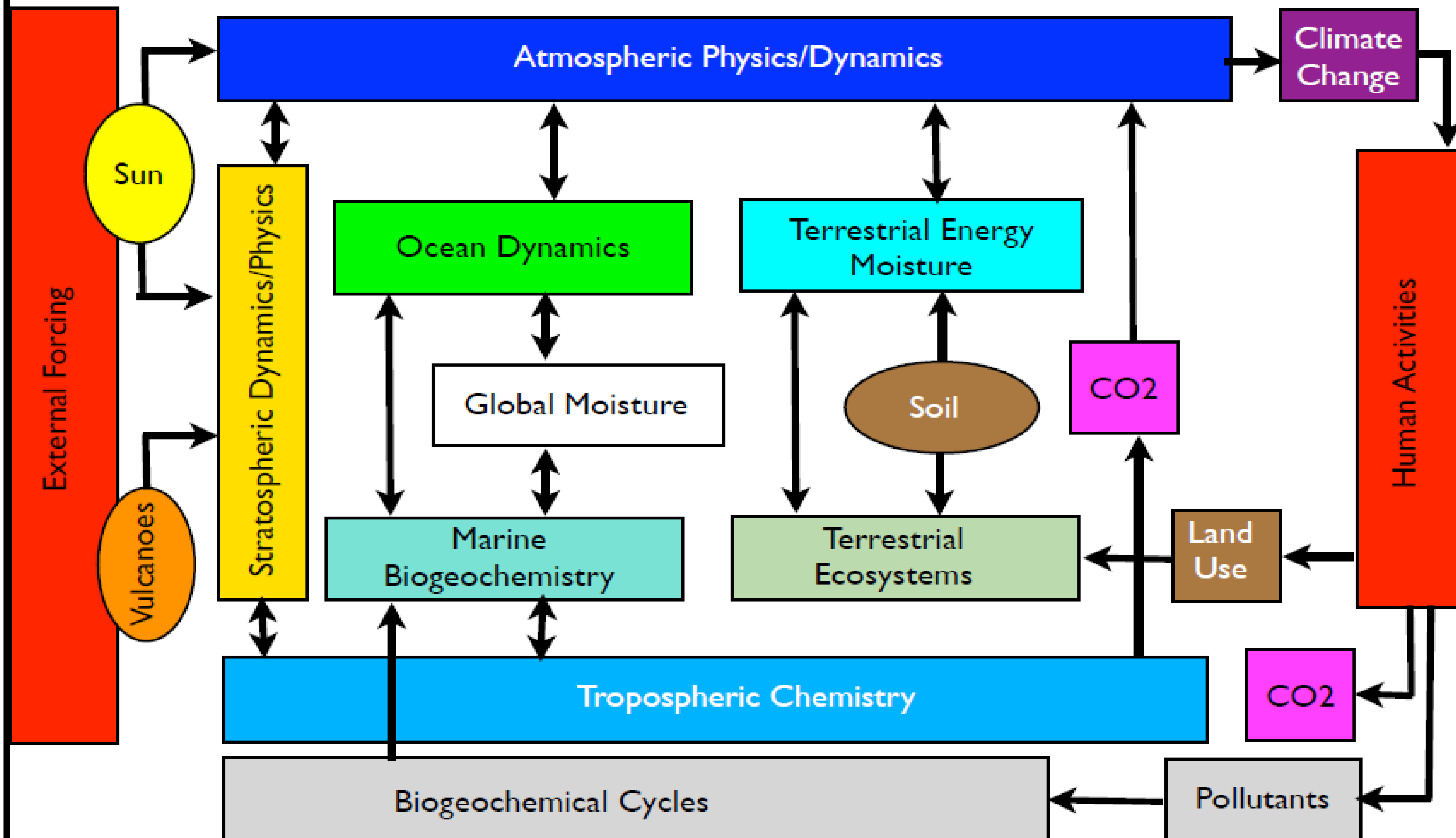




Cambridge University Press 2019

cristina.masoller@upc.edu @cristinamasoll1

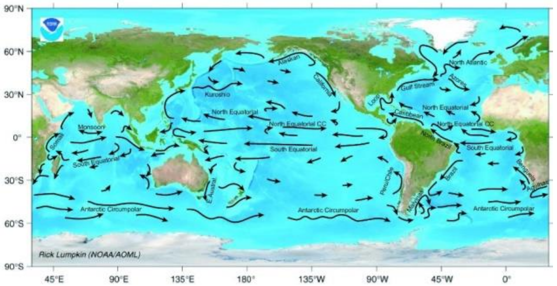
The Climate System

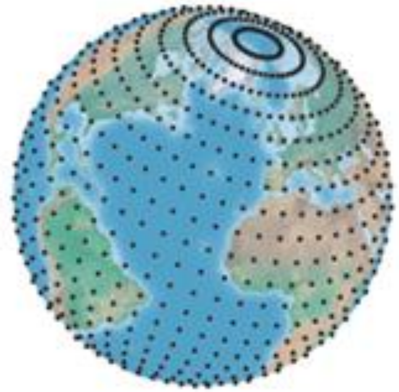


Courtesy of Henk Dijkstra (Ultrech University)

cristina.masoller@upc.edu @cristinamasoll1

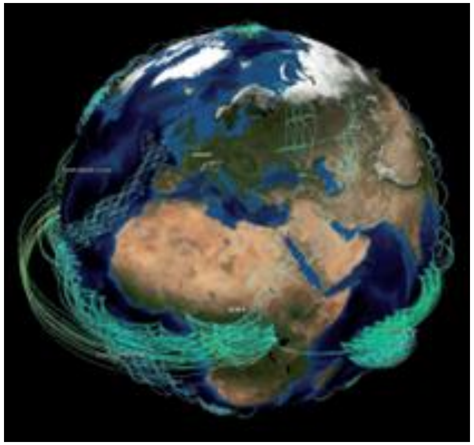
Complex network representation of the climate system

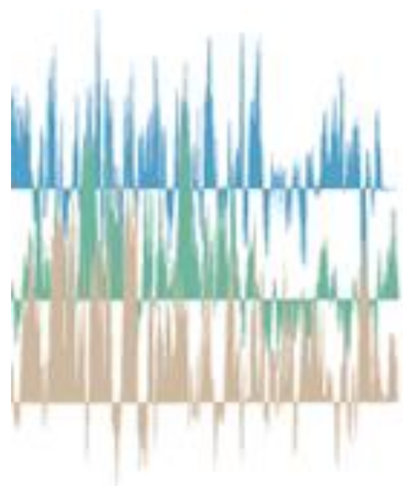




More than 10000 nodes.

Back to the climate system: interpretation (currents, winds, etc.)





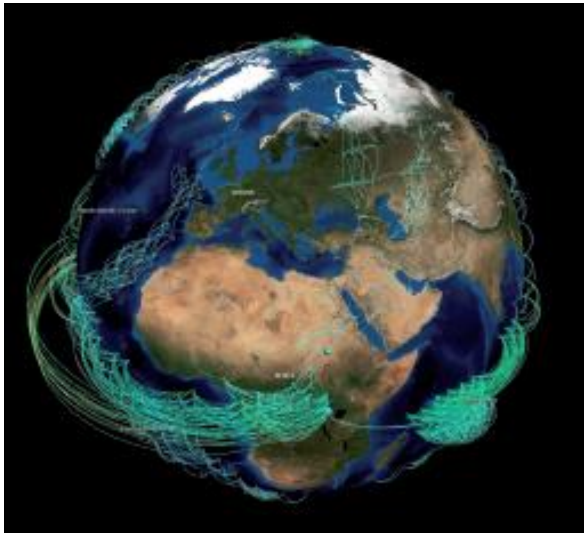
Daily resolution: more than 13000 data points in each time series

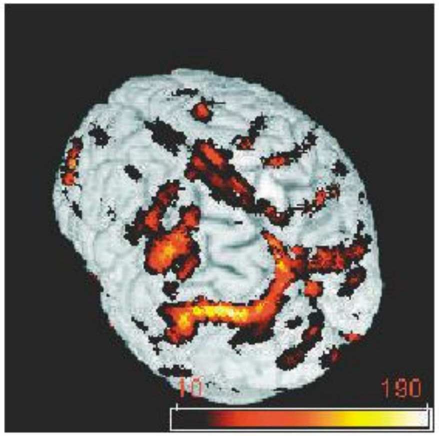
J. F. Donges et al, Chaos 25, 113101 (2015).

cristina.masoller@upc.edu @cristinamasoll1

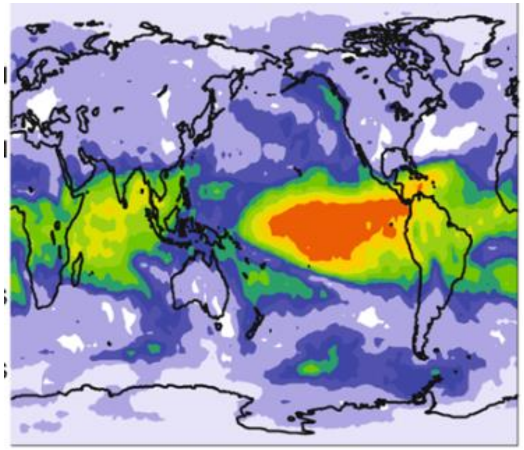
Brain network

Climate network





“Degree”
(number of links)



cristina.masoller@upc.edu @cristinamasoll1

In the analysis of climate data, depending on temporal resolution, lags need to be taken into account

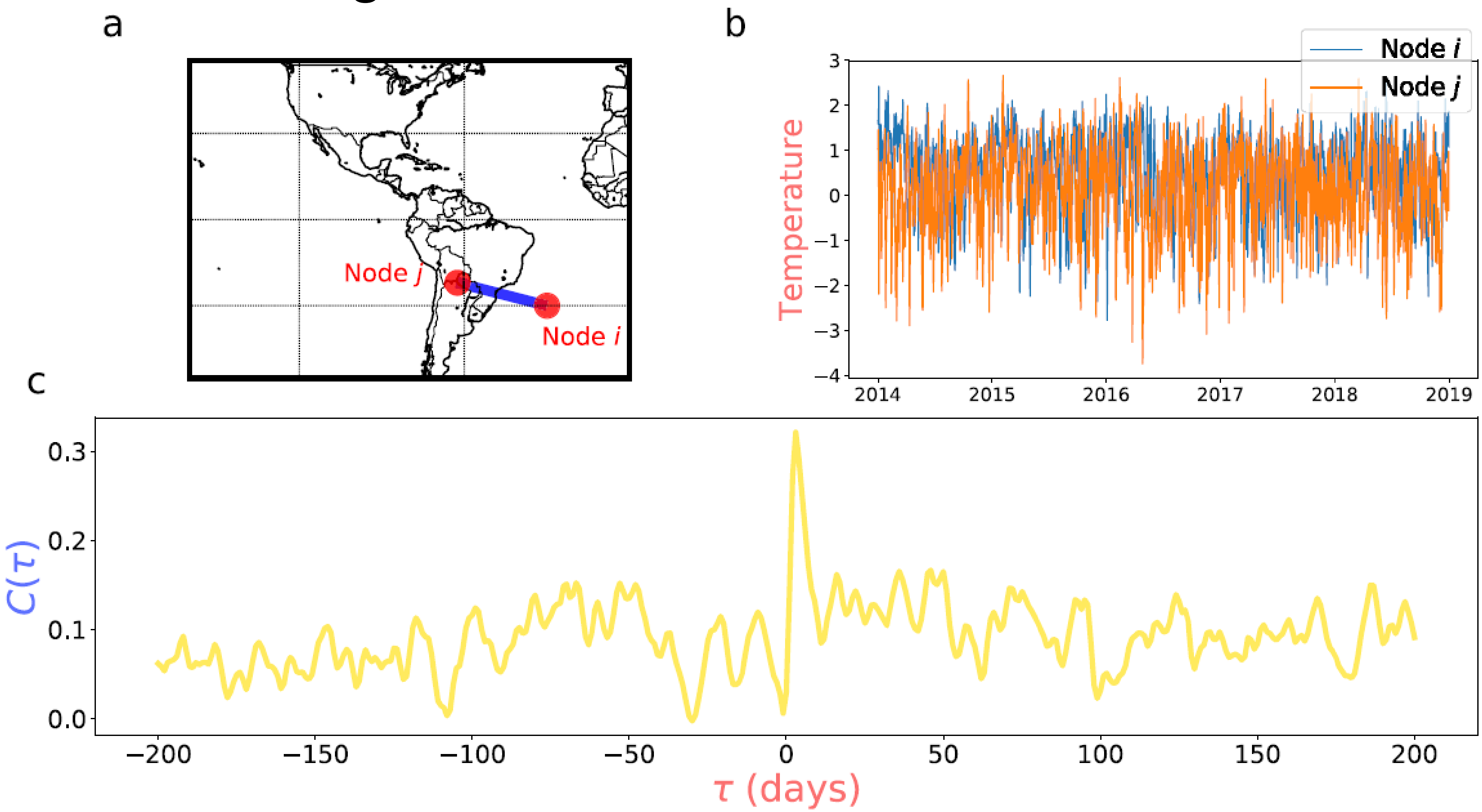


Fig. 6. Typical weighted and directed link in a Pearson Correlation Climate Network. (a) Node i is located on the Southwest Atlantic and node j is in the South American continent. (b) The near surface daily air temperature anomalies for the period [2014,2018]. (c) The cross-correlation function between the time series shown in (b). The direction of this link is from j to i with weight $W^{+}_{i,j} = 5.71$.

J. Fan, J. Meng, J. Ludescher et al. Physics Reports 896, 1 (2021).

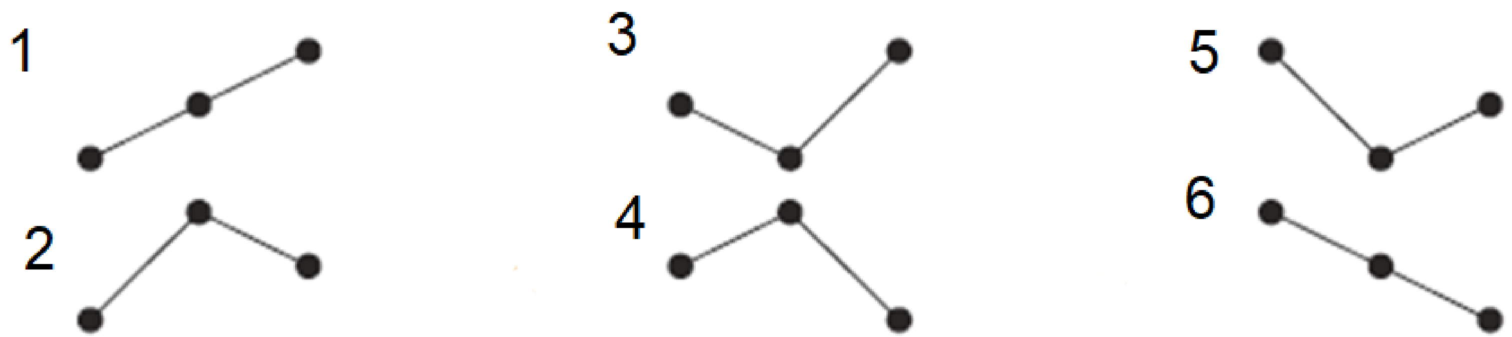
cristina.masoller@upc.edu @cristinamasoll1

Ordinal analysis provides a nonlinear way to consider lags

$\{\ldots x_i, x_{i+1}, x_{i+2}, \ldots\}$

How can three data points (let's say 2, 5, 7) be ordered?

{…2, 5, 7…} {…5, 2, 7…} {…7, 2, 5 …}

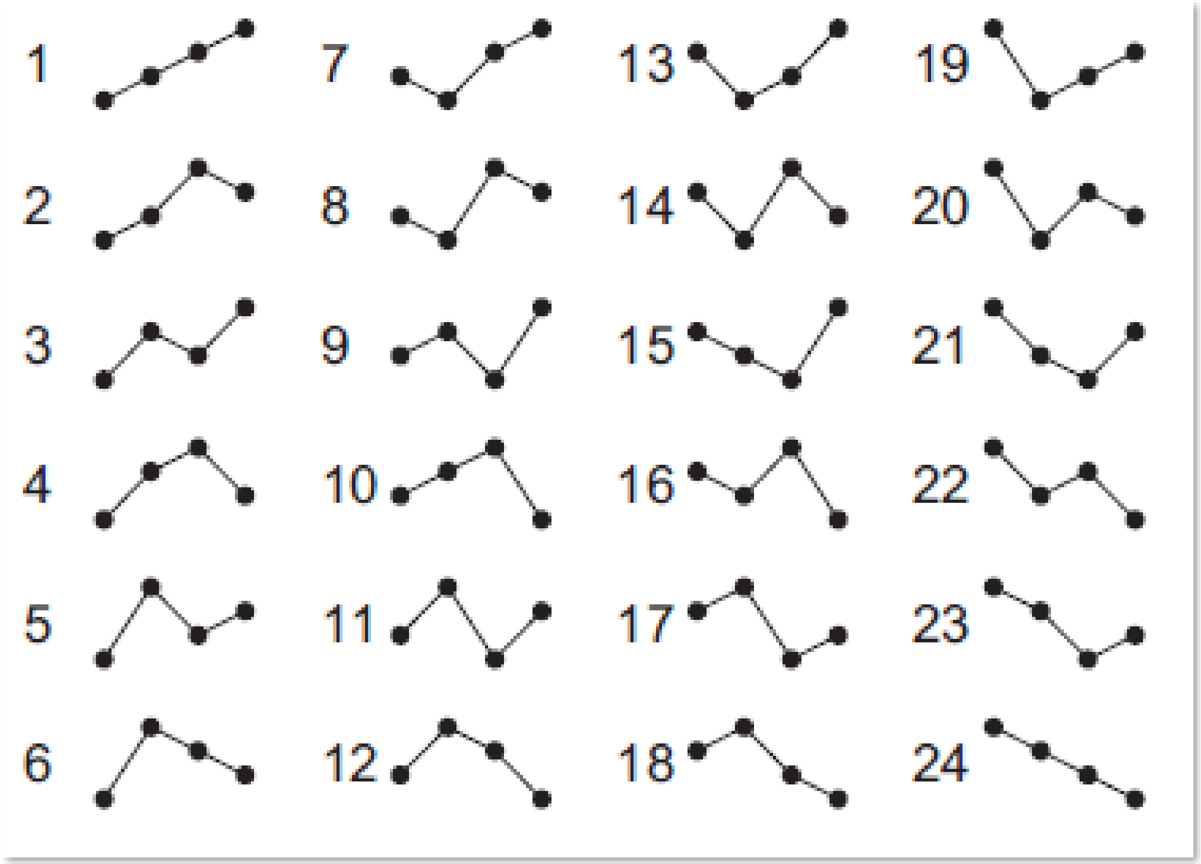


{…2, 7, 5…} {…5, 7, 2…} {…7, 5, 2…}

Bandt and Pompe: Phys. Rev. Lett. 2002

cristina.masoller@upc.edu @cristinamasoll1

How many possibilities for ordering four data points ?
$\{\ldots x_i, x_{i+1}, x_{i+2}, x_{i+3}, \ldots\}$

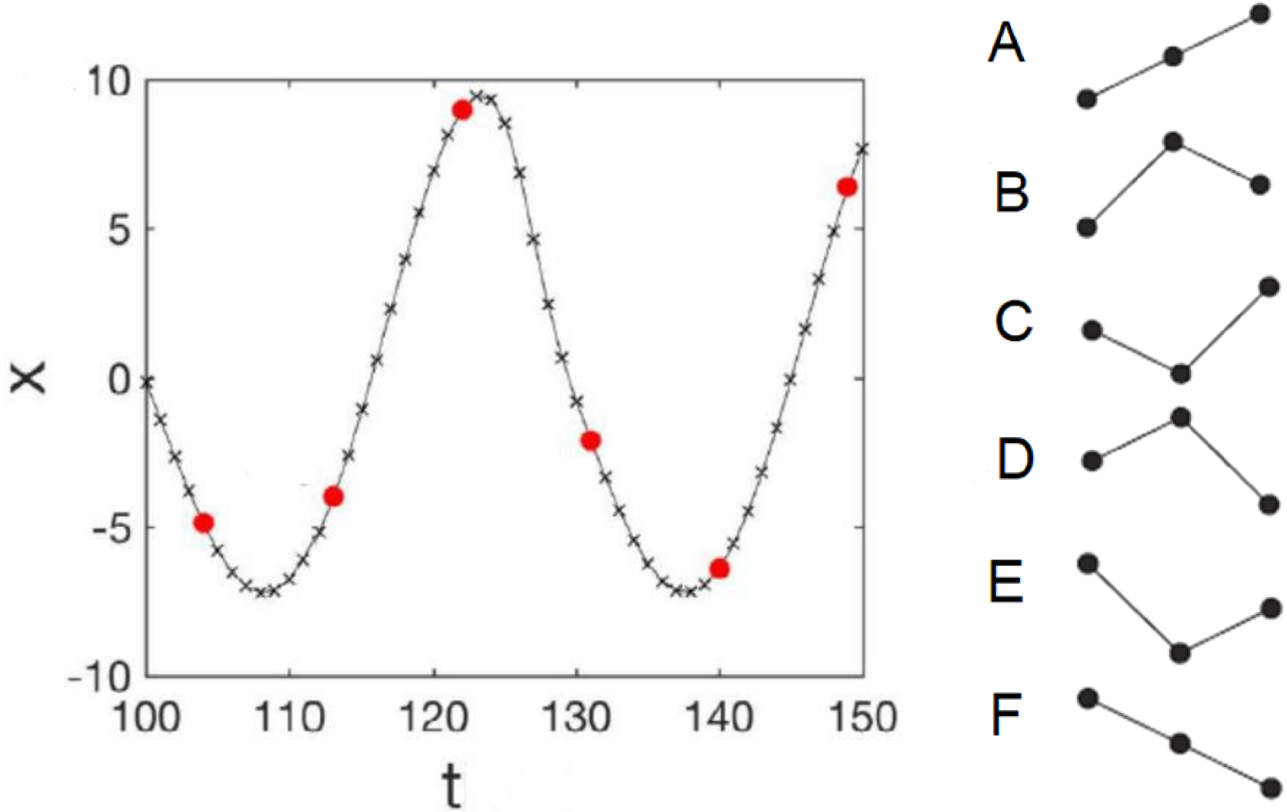


Python and Matlab codes for computing the ordinal pattern index available at:
U. Parlitz et al. Computers in Biology and Medicine 42, 319 (2012)

cristina.masoller@upc.edu @cristinamasoll1

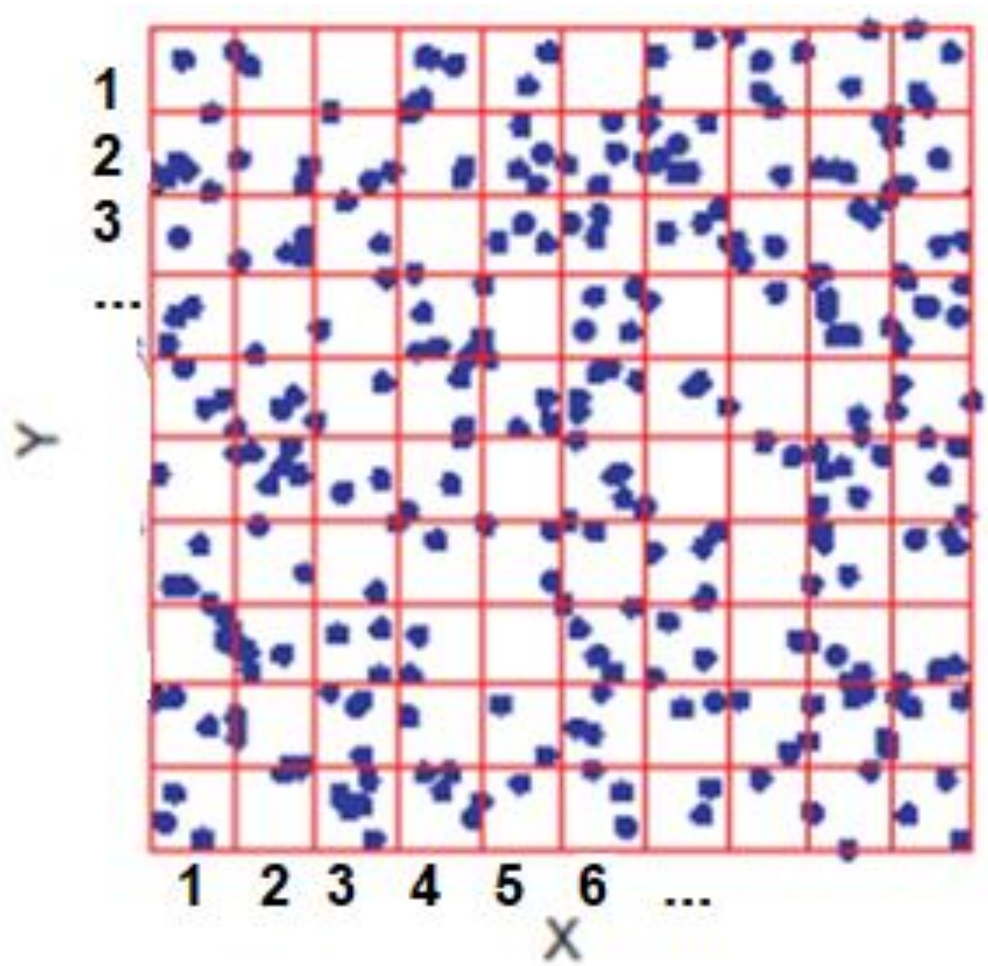
Which is the "message" "encoded" in the red dots?

s={A, B, F, C}



cristina.masoller@upc.edu @cristinamasoll1

Given two time series, X and Y, we can compute each sequence of "ordinal" patterns, s_x and s_y, and then, their mutual "ordinal" information.



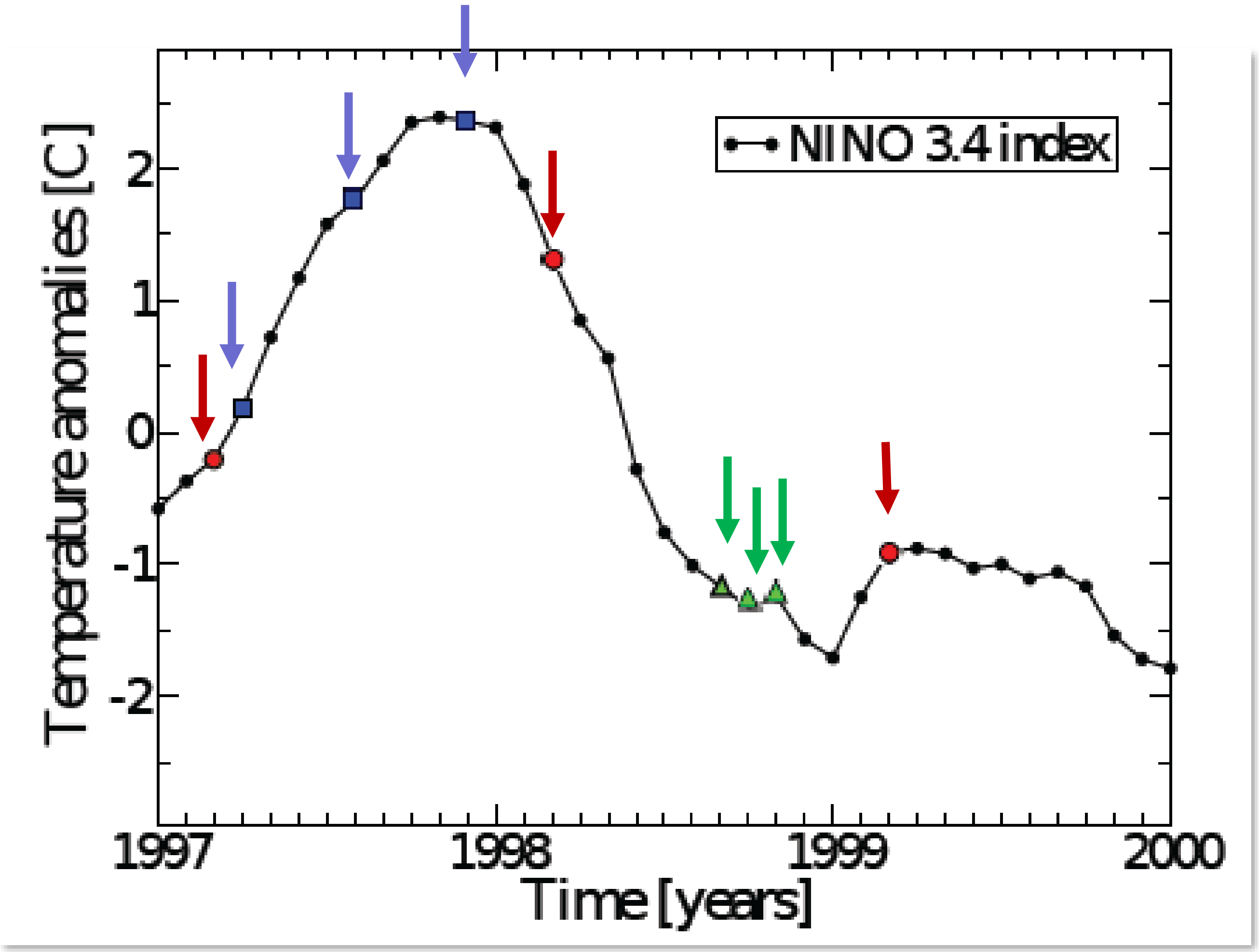
cristina.masoller@upc.edu @cristinamasoll1

Ordinal analysis allows to study different time scales

Intra-season
102

Intra-annual
012

Inter-annual
120

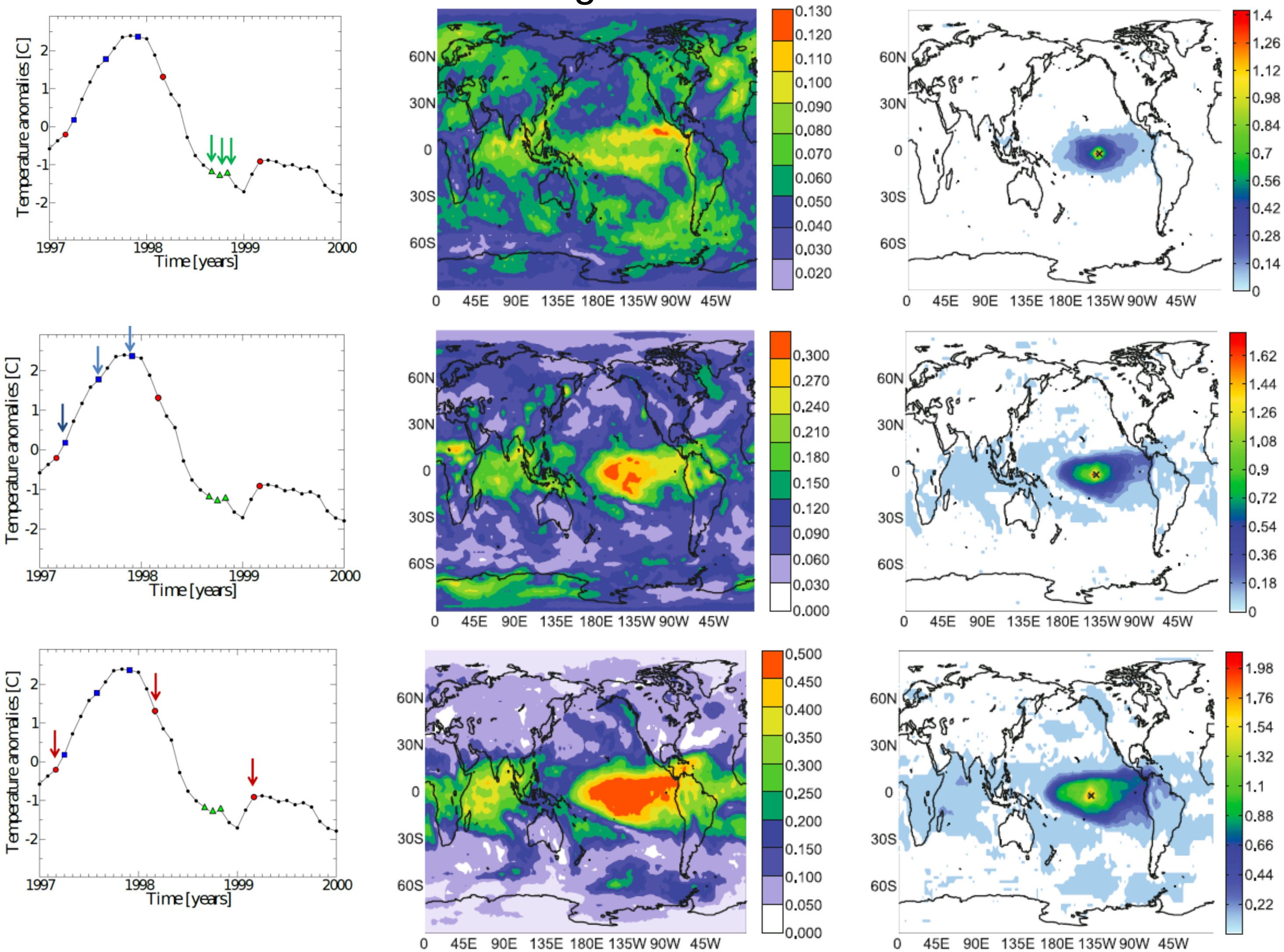


cristina.masoller@upc.edu @cristinamasoll1

Ordinal analysis detects teleconnections with different “time-scales”

Degree

Links of a node in El Niño area



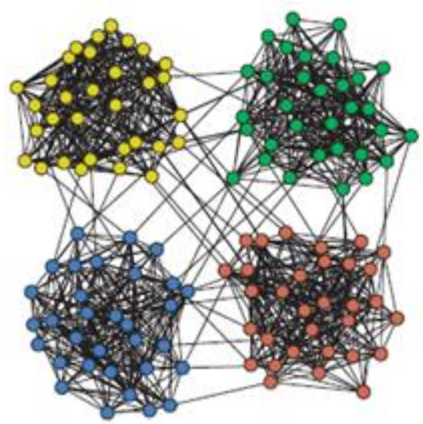
Longer time-scale ⇒ increased connectivity

cristina.masoller@upc.edu @cristinamasoll1

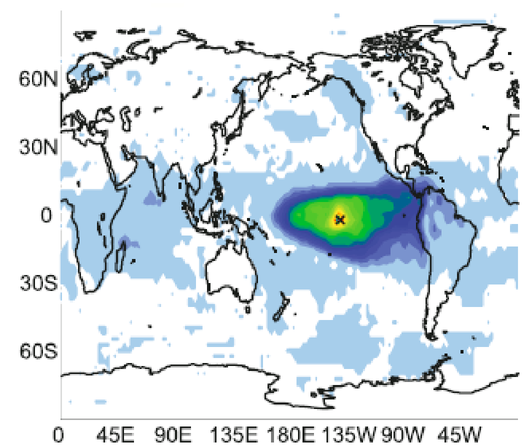
Climate "communities"

How to identify regions with similar climate?

- Goal: to construct a network in which regions with similar climate (e.g., continental) are in the same "community".



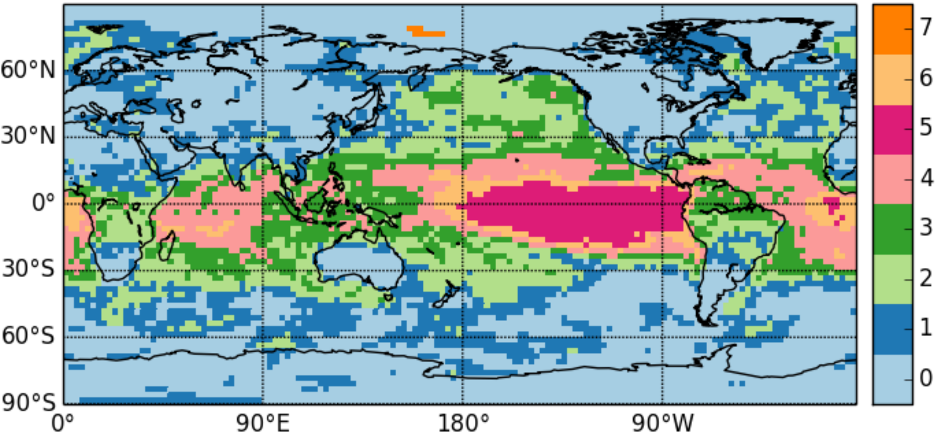
- Problem: not possible with the "usual" correlation-based method to construct the network because NH and SH are only indirectly connected.



cristina.masoller@upc.edu @cristinamasoll1

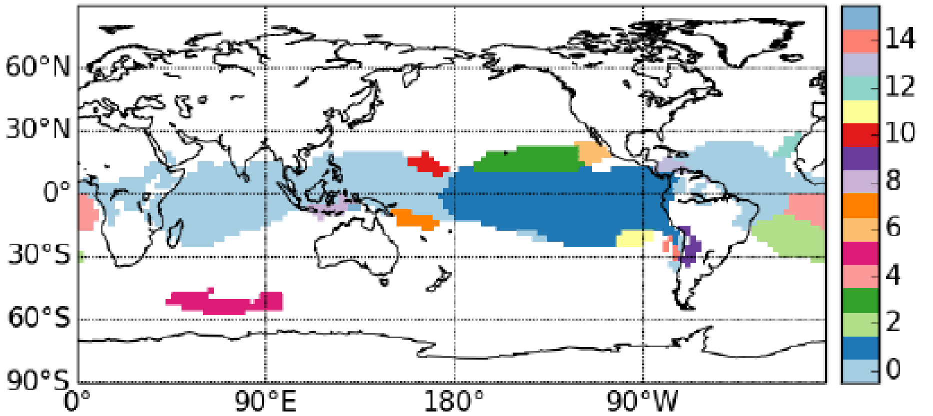
Results

Ordinal Network



Correlation network

(only the largest 16 communities)



G. Tirabassi and C. Masoller, "Unravelling the community structure of the climate system by using lags and symbolic time-series analysis", Sci. Rep. 6, 29804 (2016).

cristina.masoller@upc.edu @cristinamasoll1

Summarizing

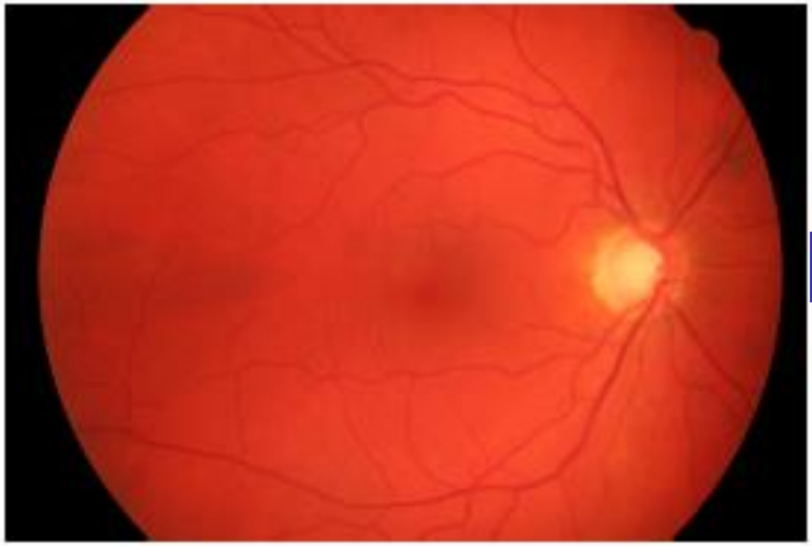
cristina.masoller@upc.edu @cristinamasoll1

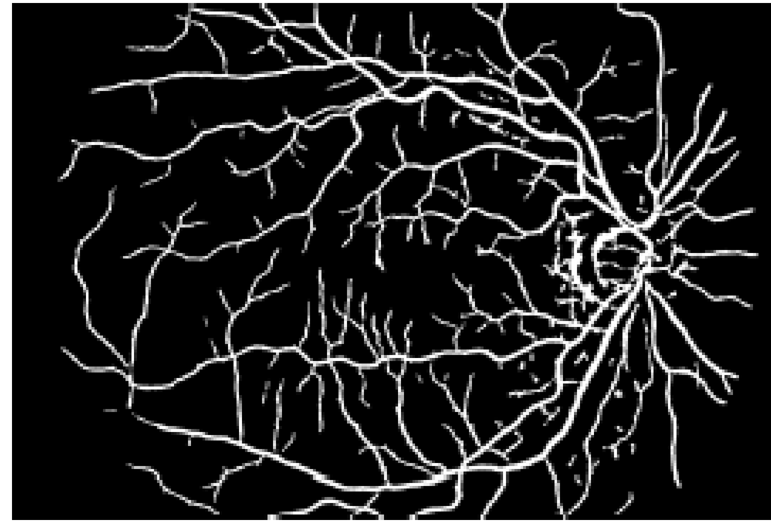
Take home messages

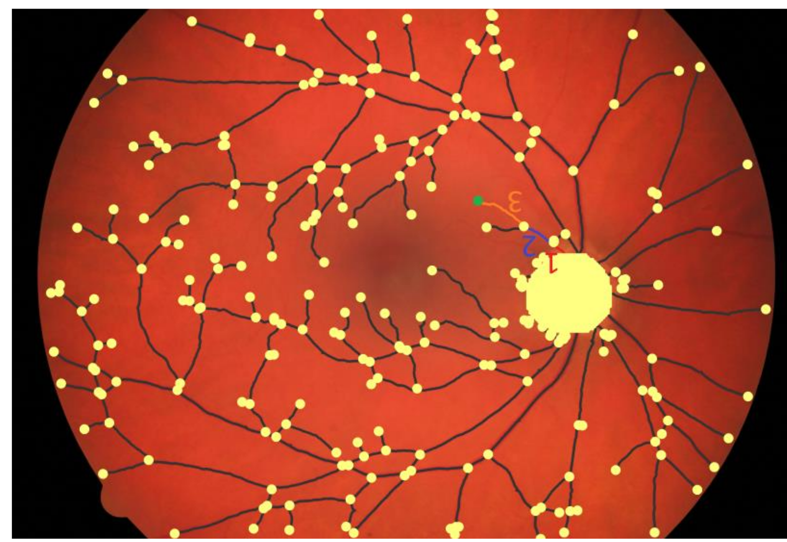
- Many measures are available to uncover inter-relationships in datasets.
- Each dataset has its own peculiarities.
- Different measures can uncover different properties.
- Hidden variables, hidden “nodes”, common “drivers” can make impossible to understand the network structure.
- Network science: many applications and challenges!

cristina.masoller@upc.edu @cristinamasoll1

“Extra bonus”: application of network science to fundus image analysis







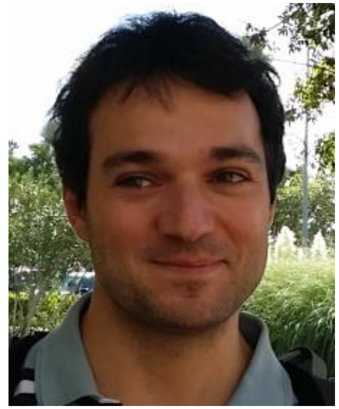
P. Amil et al., PLoS ONE 14, e0220132 (2019).

cristina.masoller@upc.edu @cristinamasoll1

THANKS TO

Ignacio Deza

Giulio Tirabassi



Dario Zappala

Riccardo Silini

Marcelo Barreiro

UNIVERSIDAD DE LA REPÚBLICA URUGUAY

Agència de Gestió d'Ajuts Universitaris i de Recerca

ICREA

GOBIERNO DE ESPAÑA — MINISTERIO DE CIENCIA E INNOVACIÓN

cristina.masoller@upc.edu @cristinamasoll1

Thank you for your attention

Slides: http://www.fisica.edu.uy/~cris/enredando.pdf

- J. I. Deza, M. Barreiro, C. Masoller, Chaos 25, 033105 (2015)
- G. Tirabassi, C. Masoller, Sci. Rep. 6, 29804 (2016)
- D. A. Zappala, M. Barreiro, C. Masoller, Earth Syst. Dynamics 9, 383 (2018)
- D. A. Zappala, M. Barreiro, C. Masoller, Chaos 29, 051101 (2019)
- R. Silini, C. Masoller, Scientific Reports 11 8423 (2021)
- Dijkstra, Barreiro, Hernandez-Garcia Masoller, *Networks in Climate*, Cambridge University Press (2019)

Agència de Gestió d'Ajuts Universitaris i de Recerca
ICREA
GOBIERNO DE ESPAÑA MINISTERIO DE CIENCIA E INNOVACIÓN

cristina.masoller@upc.edu @cristinamasoll1