Investigating large-scale atmospheric phenomena using complex networks and nonlinear time series analysis tools

Cristina Masoller
Departamento de Fisica
Universitat Politecnica de Catalunya

ICTS Climate Dynamics and Networks
November 11, 2024, Online

UNIVERSITAT POLITÈCNICA DE CATALUNYA
BARCELONATECH
Campus d'Excel·lència Internacional
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

Methods for climatological time series analysis

Outline

- Univariate analysis: Hilbert analysis

  $\{x_1, x_2, \ldots x_N\}$

- Bivariate analysis: Correlation and causality measures

  $\{x_1, x_2, \ldots x_N\}$ $\{y_1, y_2, \ldots y_N\}$

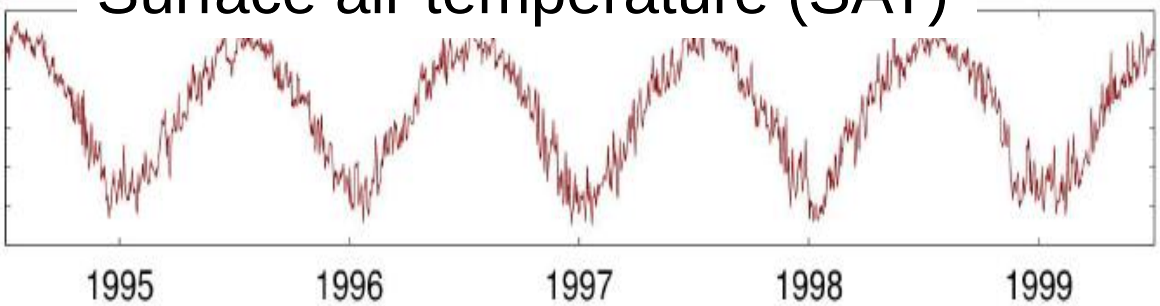
- Multivariate analysis: Ordinal analysis and complex networks

cristina.masoller@upc.edu @cristinamasoll1

The Hilbert Transform (HT)

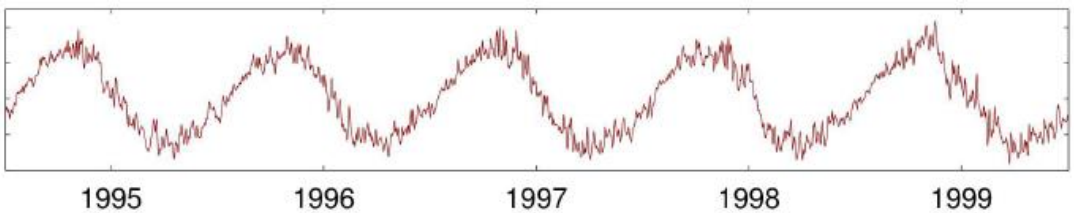
Surface air temperature (SAT)

x

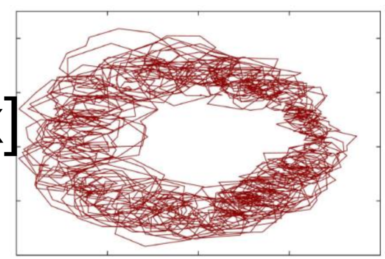


$HT[\sin(\omega t)]=\cos(\omega t)$

HT[x]



y=HT[x]



x

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

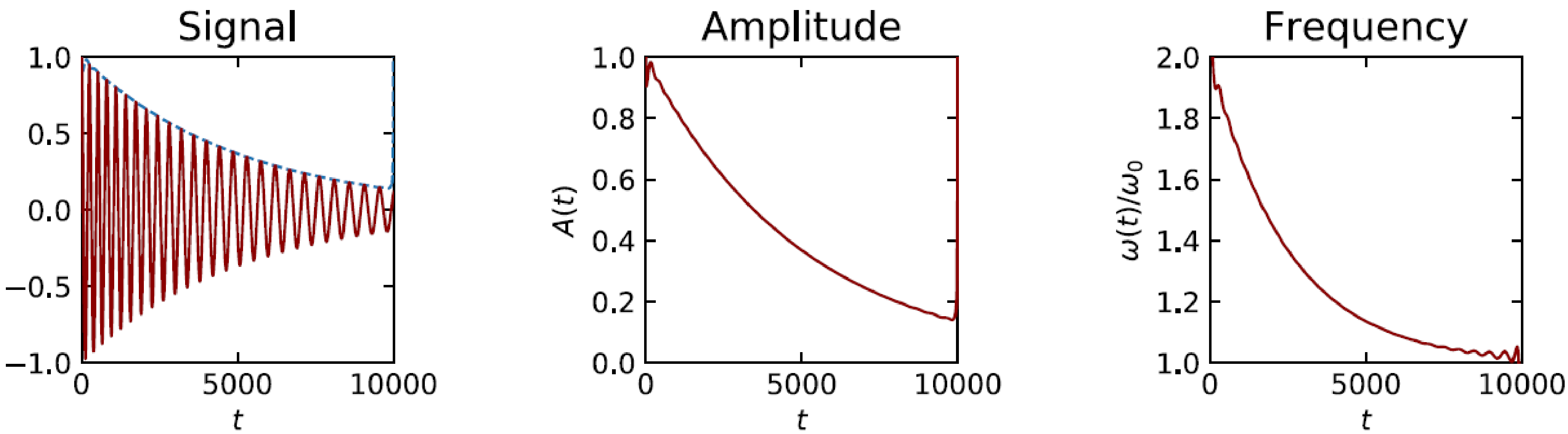
$$\varphi(t) = \arctan[y(t)/x(t)]$$

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

cristina.masoller@upc.edu @cristinamasoll1

Example

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



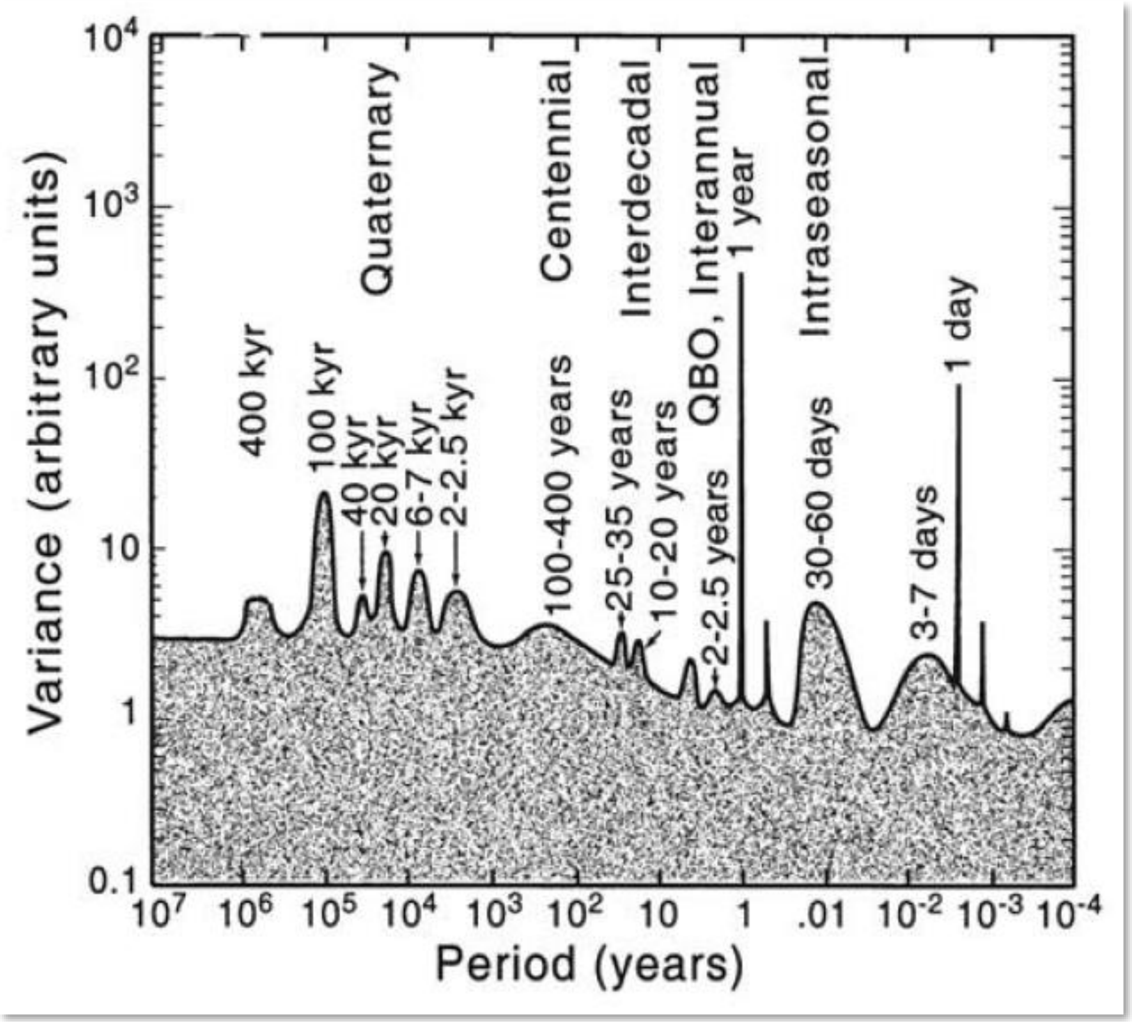
<u>A word of warning</u>: only if *x(t)* is a "narrow-band" signal then *a(t)* and $\omega(t) = d\varphi/dt$ have clear physical meaning

– ***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 for EEG analysis).

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

Surface Air Temperature (SAT) data

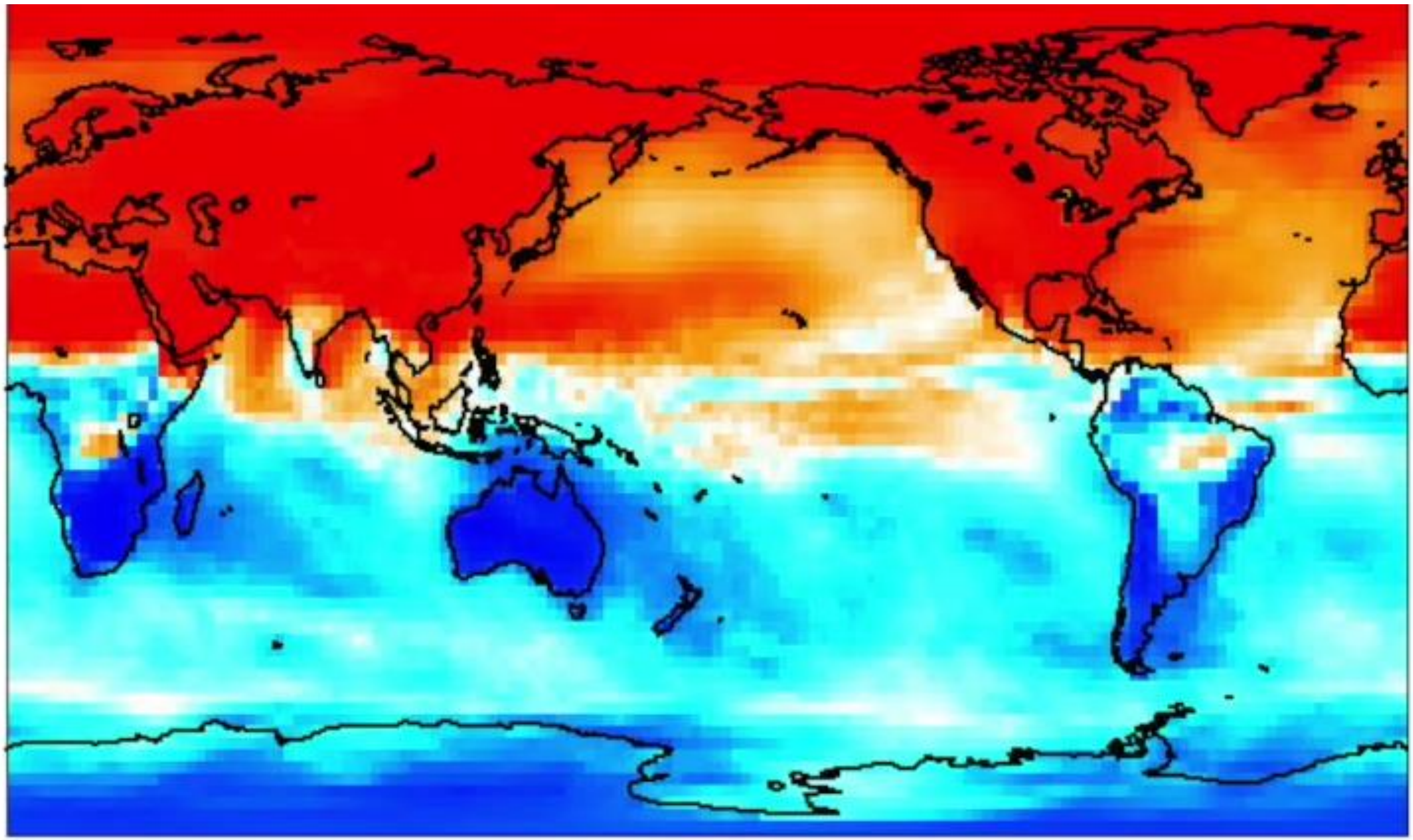
- Spatial grid $2.5^0 \times 2.5^0 \Rightarrow$ 10226 time series

- Daily resolution **1979 – 2016** $\Rightarrow$ each time series has 13700 data points

cristina.masoller@upc.edu @cristinamasoll1

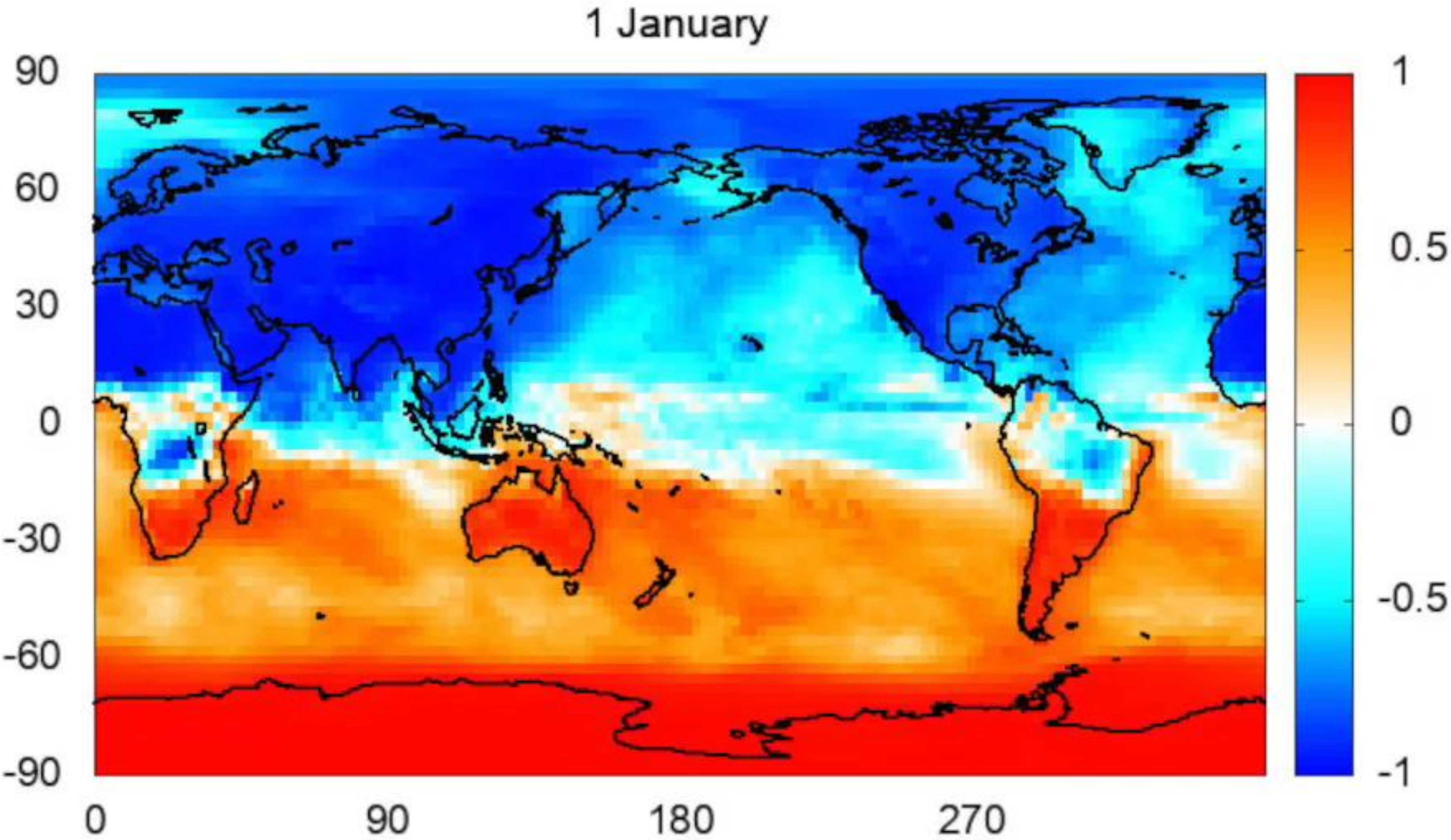
Which information carries the <u>Hilbert phase</u>?

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



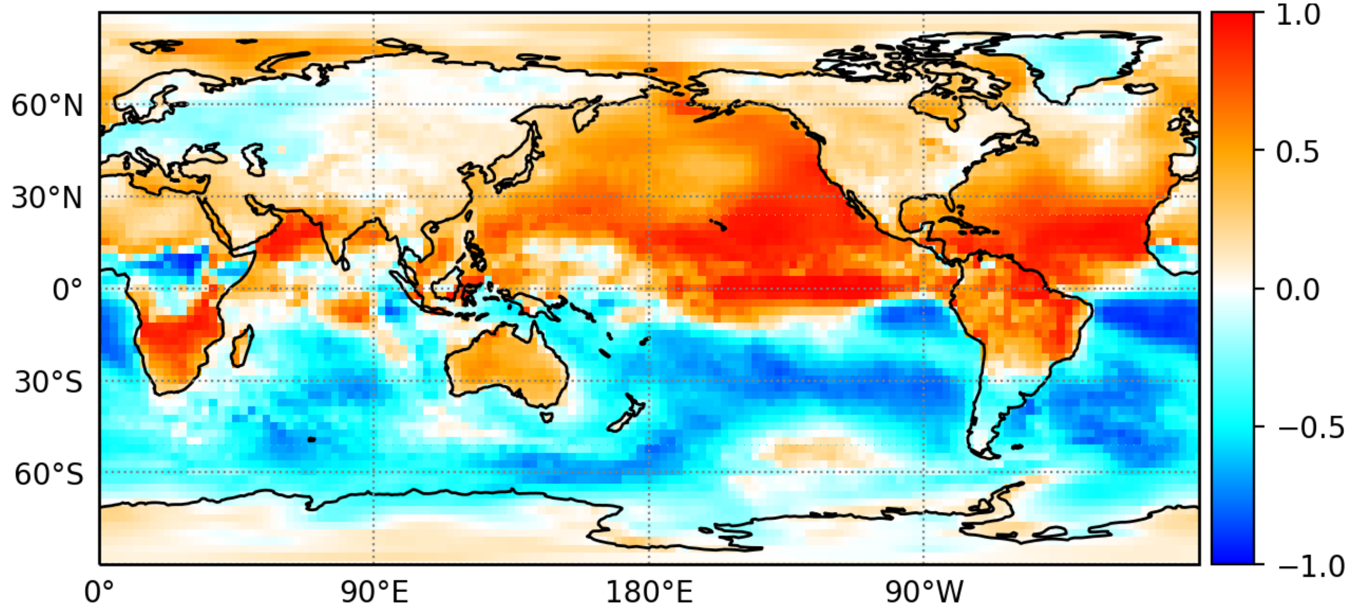
cristina.masoller@upc.edu @cristinamasoll1

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

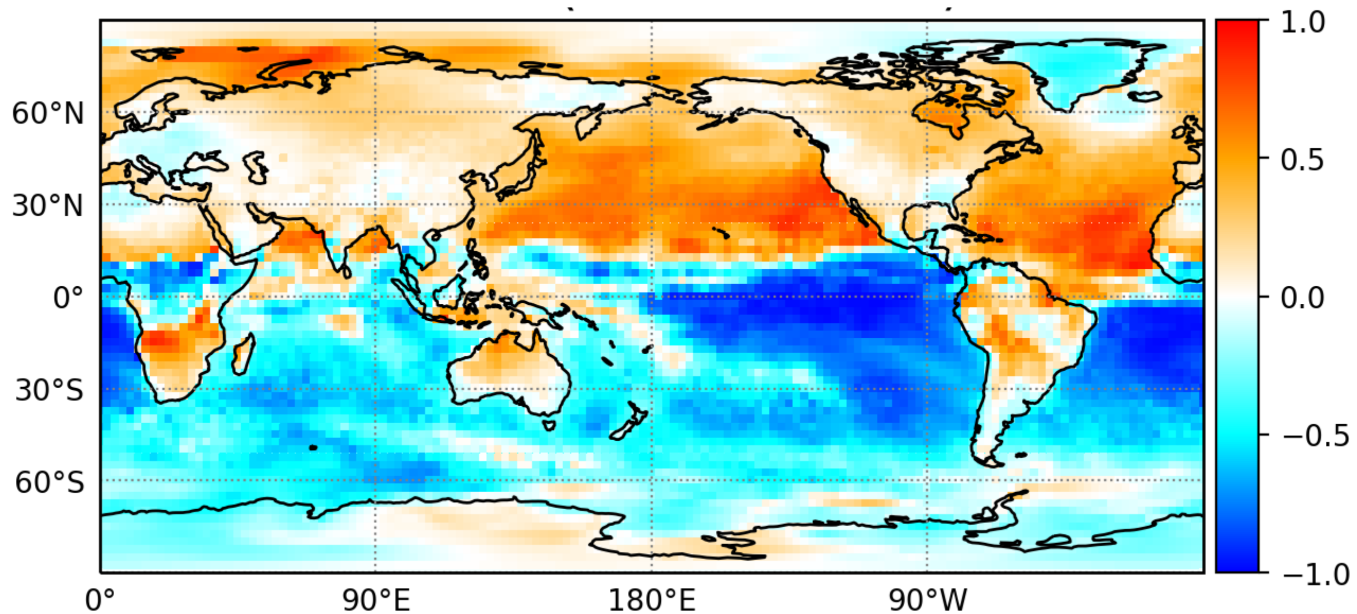


cristina.masoller@upc.edu @cristinamasoll1

El Niño period
(October 2015)



La Niña period
(Octubre 2011)



cristina.masoller@upc.edu @cristinamasoll1

Can we use Hilbert analysis to identify and quantify regional “climate change”?

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:

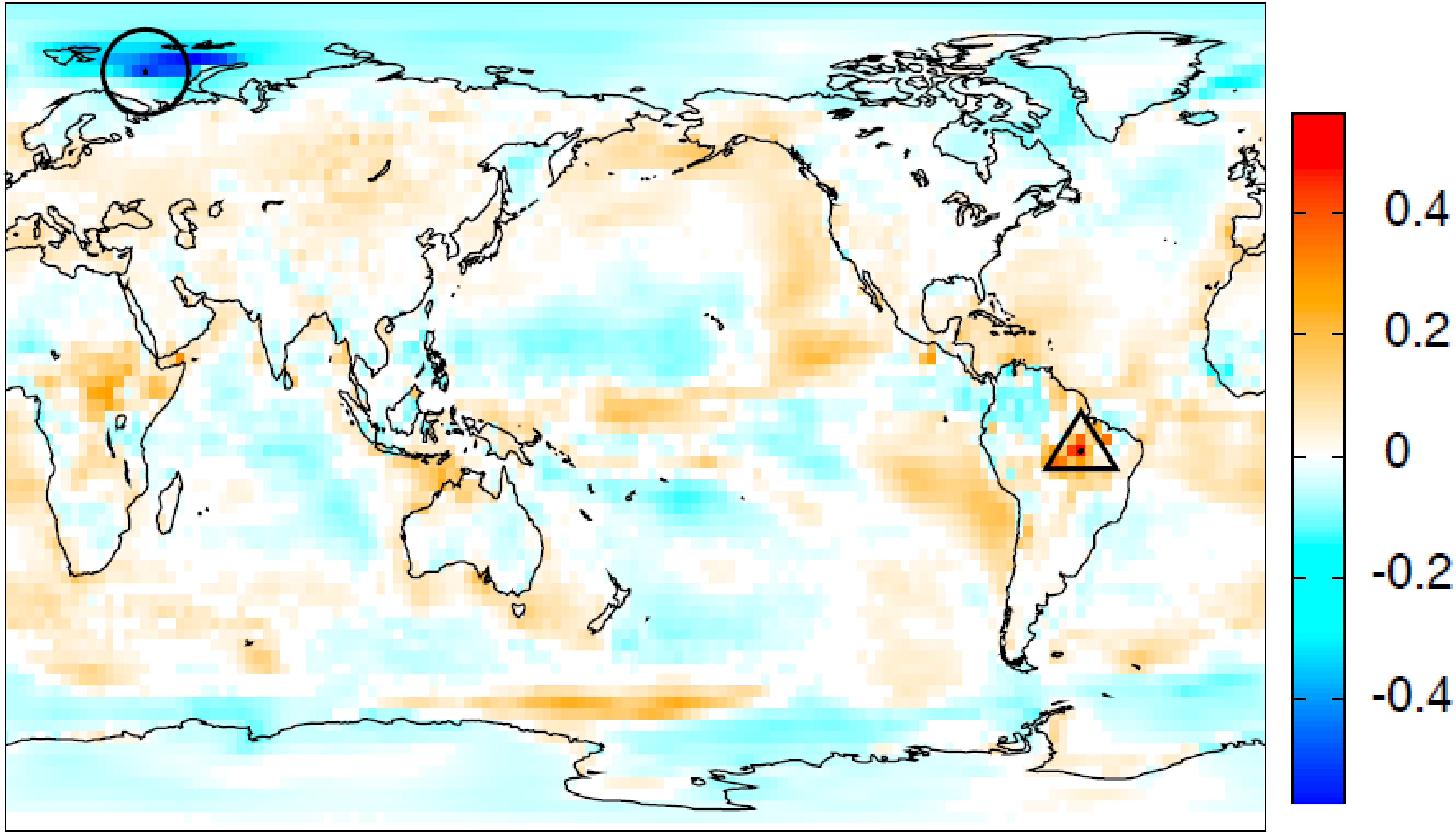
$$\frac{\Delta a}{\langle a \rangle} \geq \langle . \rangle_s + 2\sigma_s \quad \text{or} \quad \frac{\Delta a}{\langle a \rangle} \leq \langle . \rangle_s - 2\sigma_s$$

100 "surrogates"

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

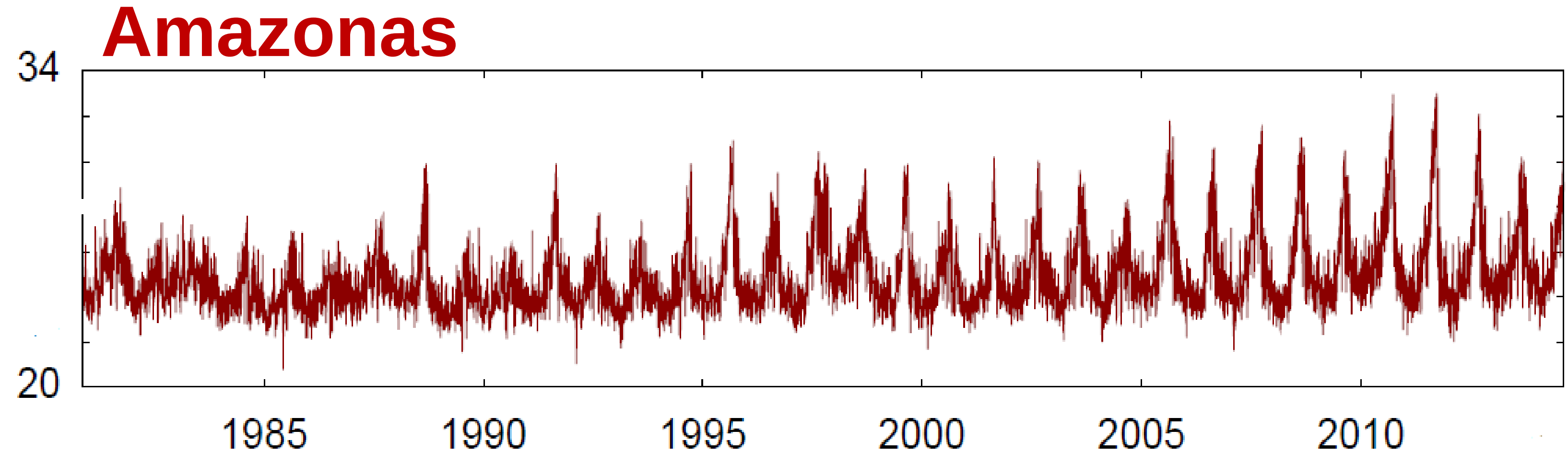
cristina.masoller@upc.edu @cristinamasoll1

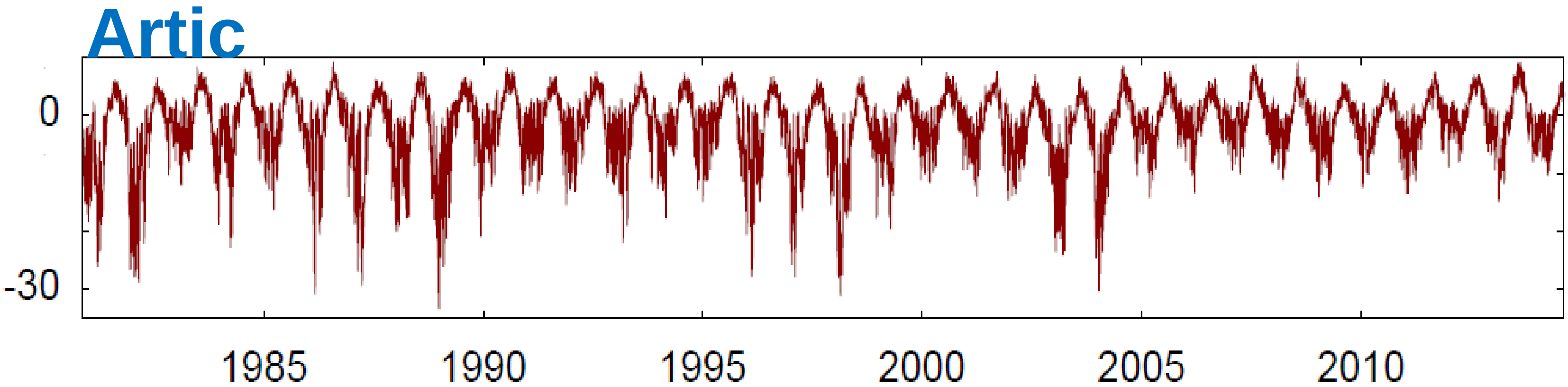
Relative decadal variations



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

Outline

- Univariate analysis: Hilbert analysis

  $\{x_1, x_2, \ldots x_N\}$

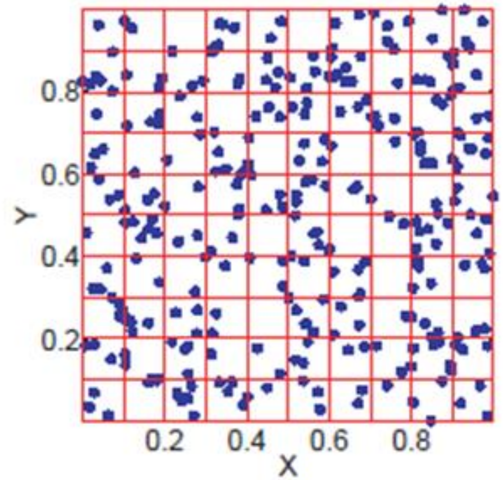
- Bivariate analysis: correlation and causality measures

  $\{x_1, x_2, \ldots x_N\}$ $\{y_1, y_2, \ldots y_N\}$

- Multivariate analysis: Ordinal analysis and complex networks

cristina.masoller@upc.edu @cristinamasoll1

Mutual Information (MI)



- *MI is calculated from probability distributions* $$M_{ij} = \sum_{m,n} p_{ij}(m,n) \log \frac{p_{ij}(m,n)}{p_i(m)p_j(n)}$$
- It quantifies the reduction in uncertainty of one variable by knowing the other variable.
- *If X, Y are independent, MI = 0*, else ***MI >0***
- For Gaussian processes: MI = -1/2 log(1-ρ^2) where ρ is the cross-correlation coefficient.

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)

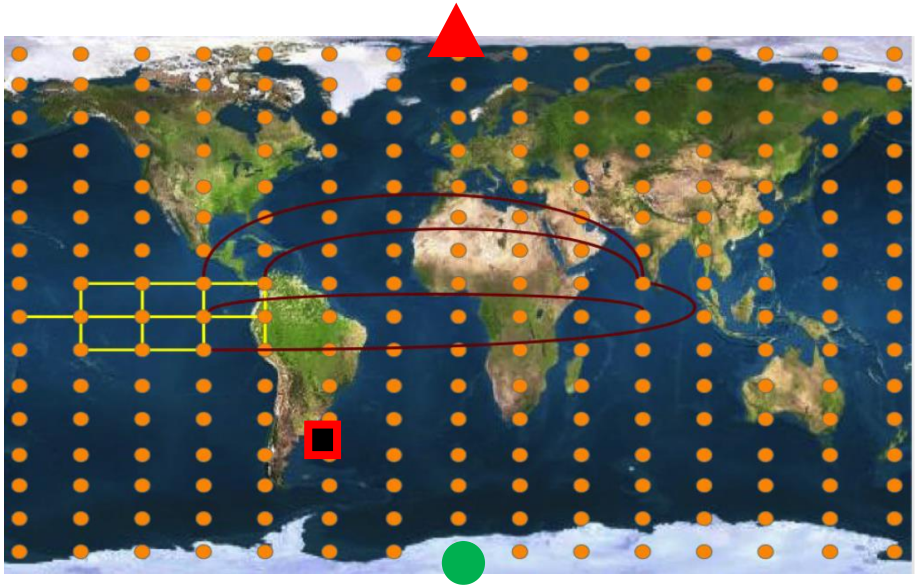
K. Hlaváčková-Schindler et al. / Physics Reports 441 (2007) 1–46

cristina.masoller@upc.edu @cristinamasoll1

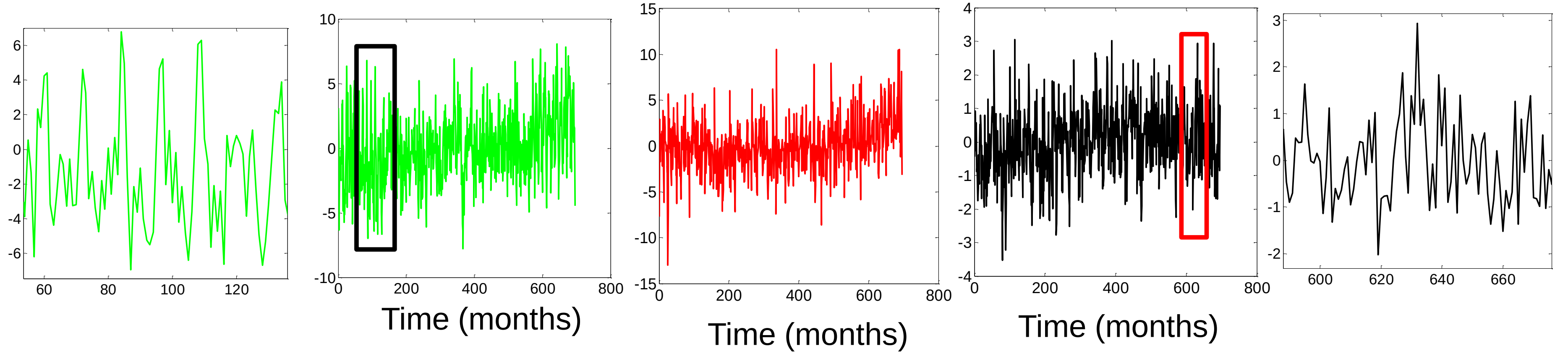
Application: analysis of surface air temperature (SAT) anomalies.

Anomaly = annual solar cycle removed

2.5º x 2.5º =10226 grid points
Monthly data ⇒ In each time series: 696 data points (1949-2006: 58 years x 12 months)

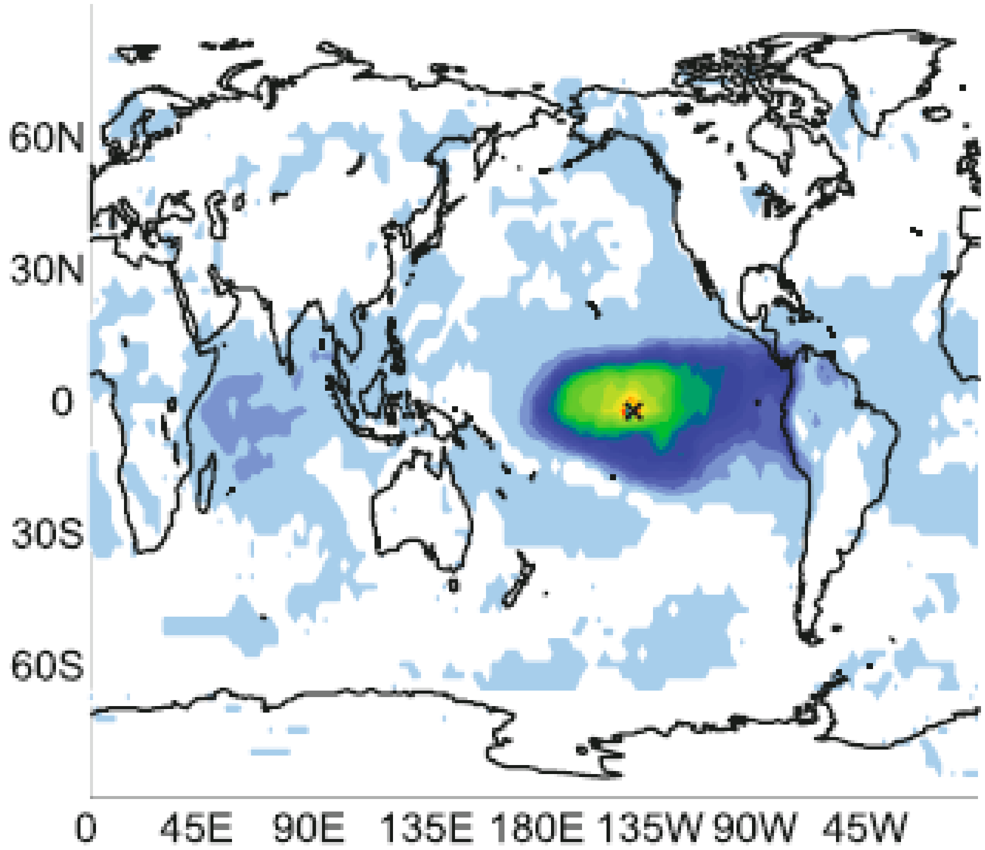


How does the data look like?



cristina.masoller@upc.edu @cristinamasoll1

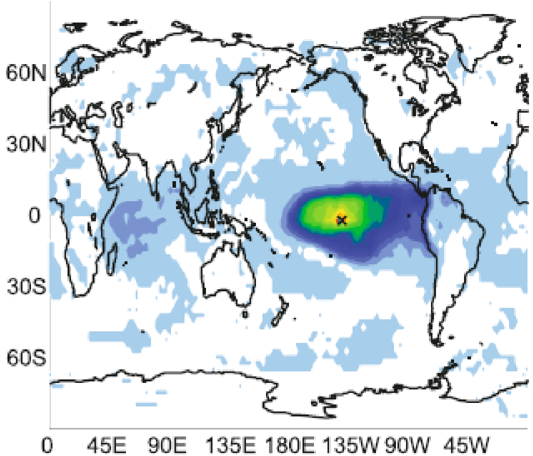
Mutual Information of SAT anomaly in El Niño region and other regions (white: MI not significant)

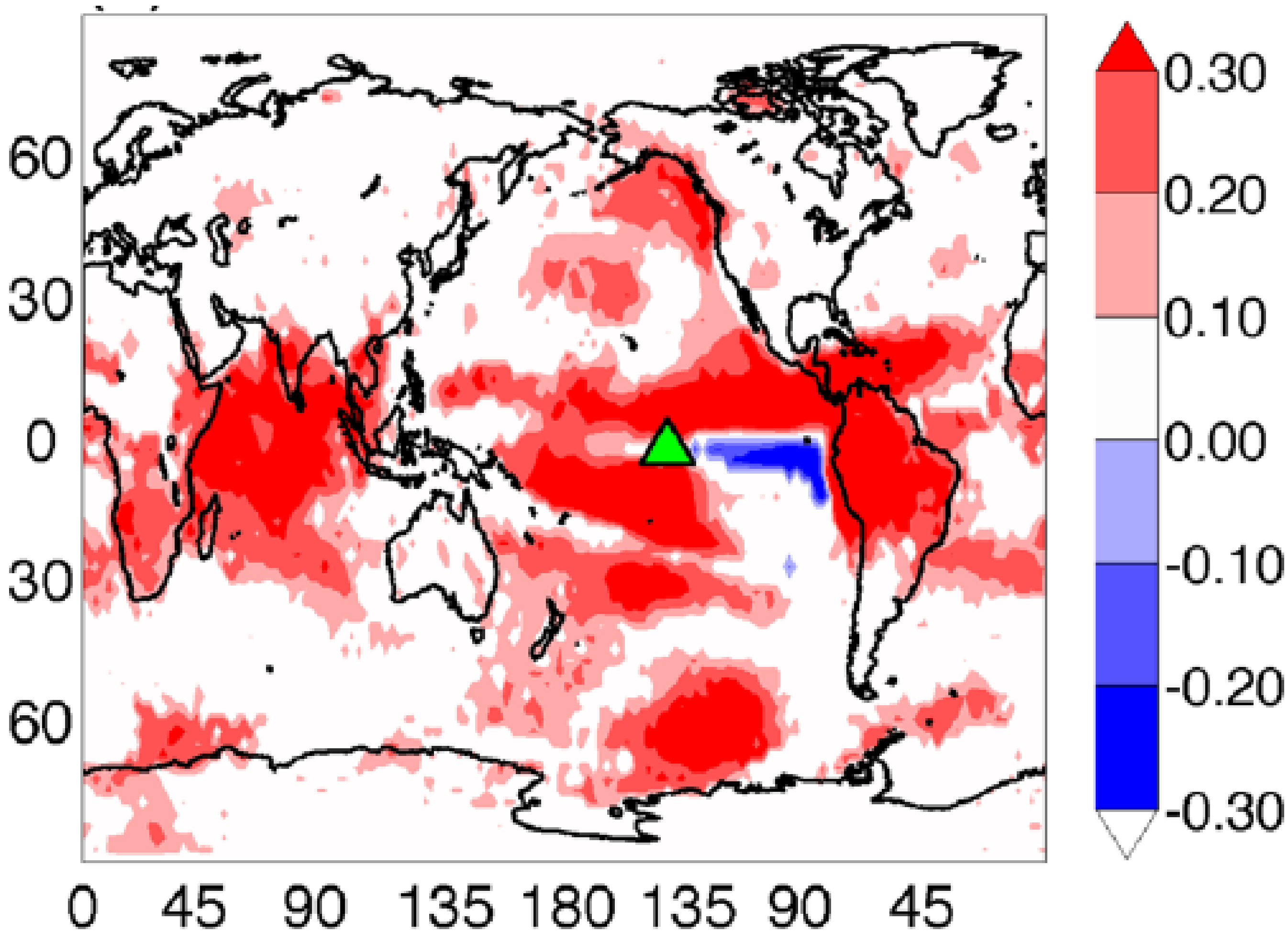


J. I. Deza, M. Barreiro, C. Masoller, "Inferring interdependencies in climate networks constructed at inter-annual, intra-season and longer time scales", Eur. Phys. J. ST 222, 511 (2013).

cristina.masoller@upc.edu @cristinamasoll1

Directionality Index

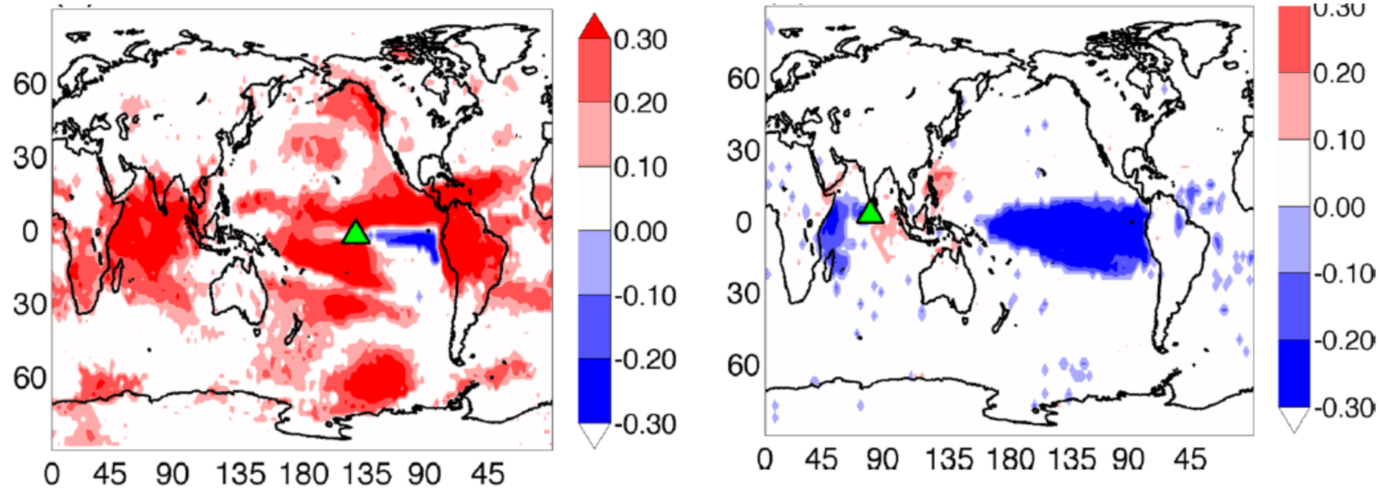


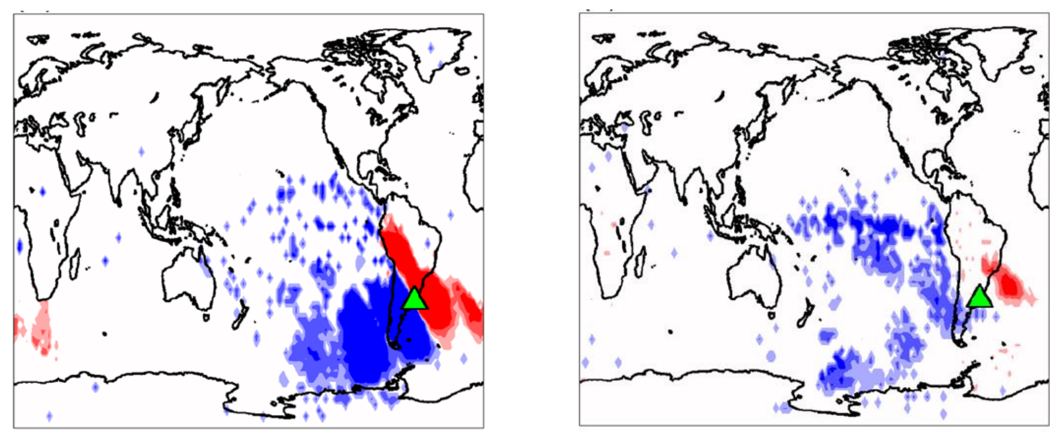


J. I. Deza, M. Barreiro, and C. Masoller, "*Assessing the direction of climate interactions by means of complex networks and information theoretic tools*", Chaos 25, 033105 (2015).

cristina.masoller@upc.edu @cristinamasoll1

Directionality Index





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

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

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

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

cristina.masoller@upc.edu @cristinamasoll1

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	Equations
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

Results

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

cristina.masoller@upc.edu @cristinamasoll1

Comparison with Granger Causality and Transfer Entropy

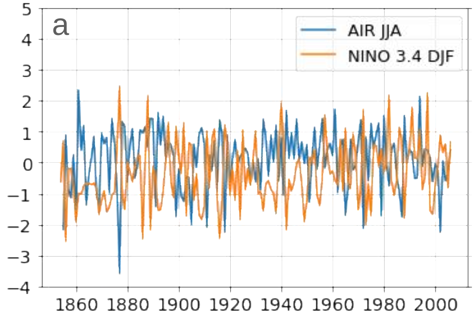
	Model	pTE		GC		TE		DI		
		Y → X	X → Y	Y → X	X → Y	Y → X	X → Y	pTE	GC	TE
Y X	M0	3.8	3.9	5.1	5.0	4.4	4.4	−0.01	0.01	0.00
	M1	2.3	2.6	3.3	3.1	100	100	−0.06	0.03	0.00
	M2	4.2	4.7	5.5	5.9	4.7	4.9	−0.06	−0.04	−0.02
Y → X	M3	100	4.5	100	4.8	70.2	5.6	0.91	0.91	0.85
	M4	80.7	3.8	84.2	4.9	96.0	4.7	0.91	0.89	0.91
	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
	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
Y ⇆ X	M13	100	100	100	100	100	100	0.00	0.00	0.00
	M14	100	100	100	100	100	100	0.00	0.00	0.00

cristina.masoller@upc.edu @cristinamasoll1

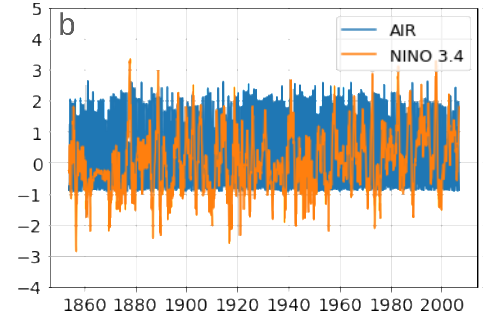
Application to real data NINO3.4 ← → All India Rainfall

IAAF

Yearly sampled (152)



Monthly sampled (1836)



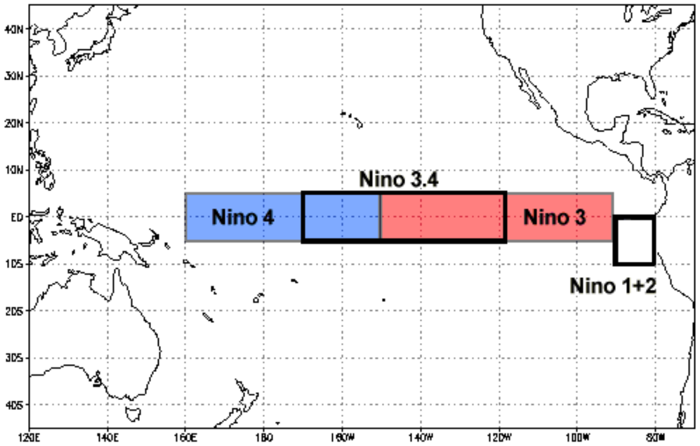
	Yearly sampled (152)	Monthly sampled (1836)
pTE	NINO3.4 ⟶ AIR (0.04 s)	NINO3.4 ⟵ AIR (0.5 s)
GC	NINO3.4 ⟶ AIR (0.4 s)	NINO3.4 ⟵ AIR (0.9 s)
TE	NINO3.4 ⇆ AIR (1 s)	NINO3.4 ⇆ AIR (68 s)

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)
⇒ 64 x 10^6 pairs

⇒ **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

Outline

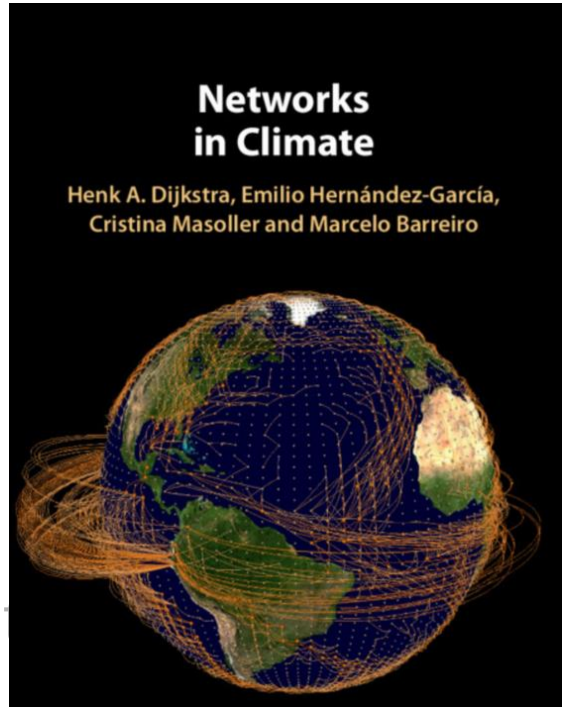
- Univariate analysis: Hilbert analysis

  $\{x_1, x_2, \ldots x_N\}$

- Bivariate analysis: correlation and causali

  $\{x_1, x_2, \ldots x_N\}$ $\{y_1, y_2, \ldots y_N\}$

- Multivariate analysis: Ordinal analysis and complex networks

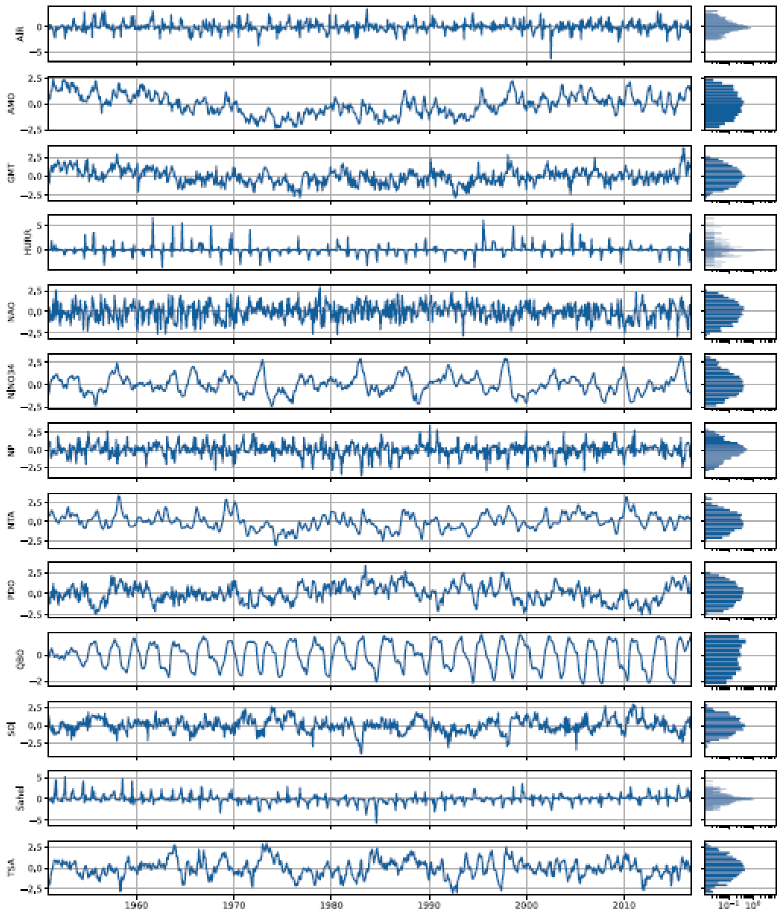


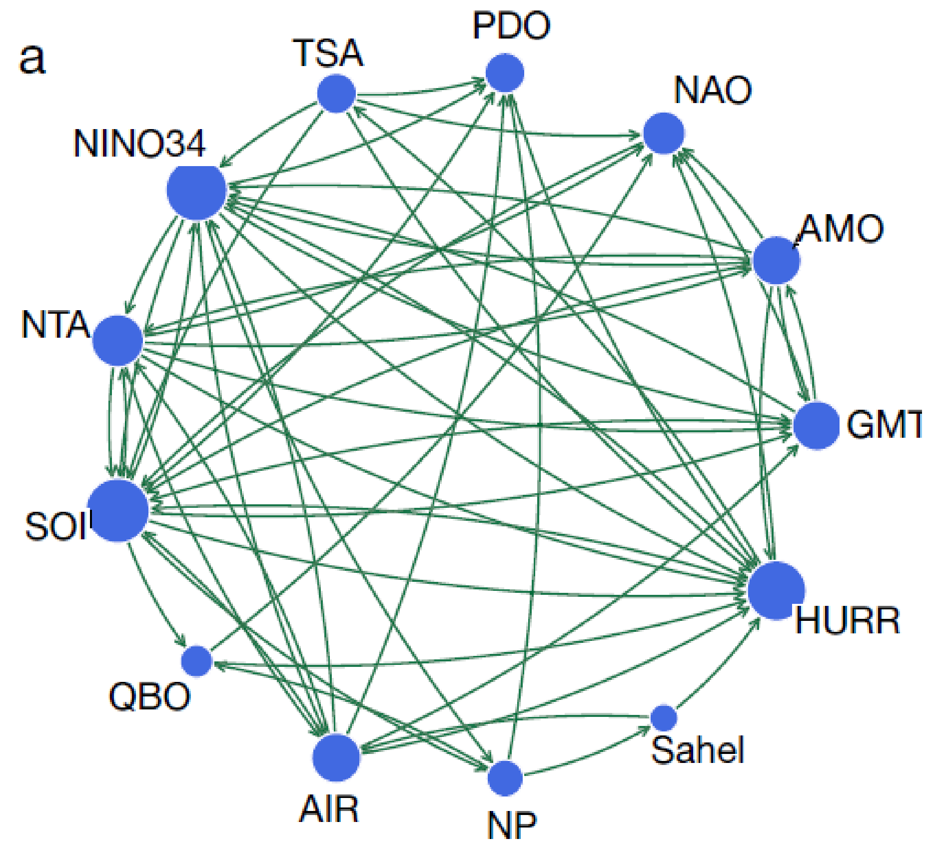
Cambridge University Press 2019

cristina.masoller@upc.edu @cristinamasoll1

Directed network of climatic indices

Constructed calculated pTE with different lags





R. Silini, G. Tirabassi, M Barreiro, L. Ferranti, C. Masoller, “*Assessing causal dependencies in climatic indices*”, Climate Dynamics 61, 79–89 (2023).

cristina.masoller@upc.edu @cristinamasoll1

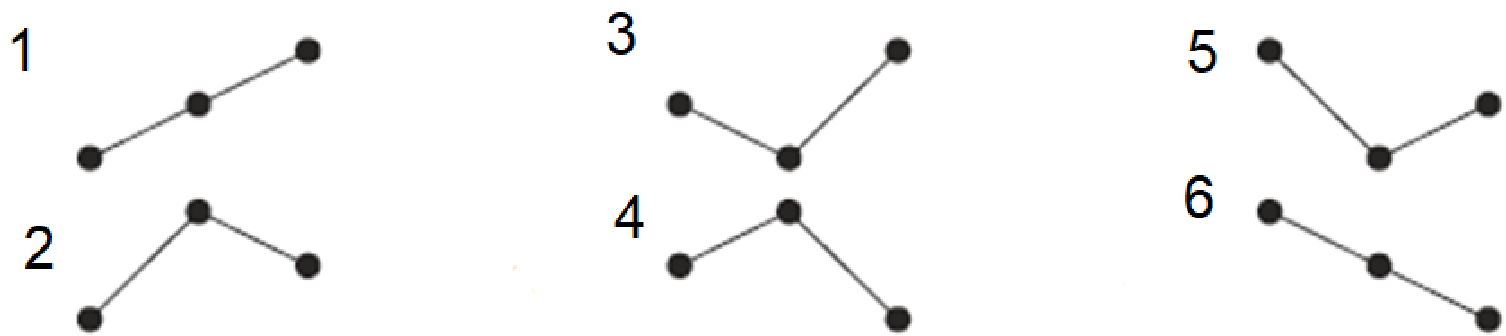
Ordinal analysis:
A nonlinear way to select the time-scale of the analysis

Ordinal analysis

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

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

{…2, 5, 7…} {…5, 2, 7…} {…7, 2, 5 …}



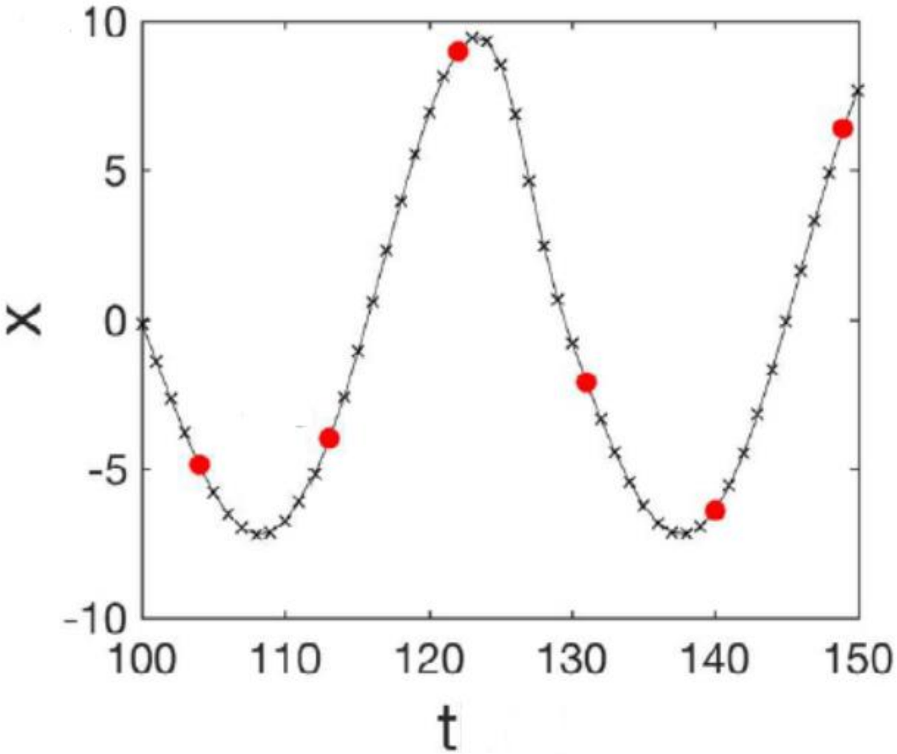
{…2, 7, 5…} {…5, 7, 2…} {…7, 5, 2…}

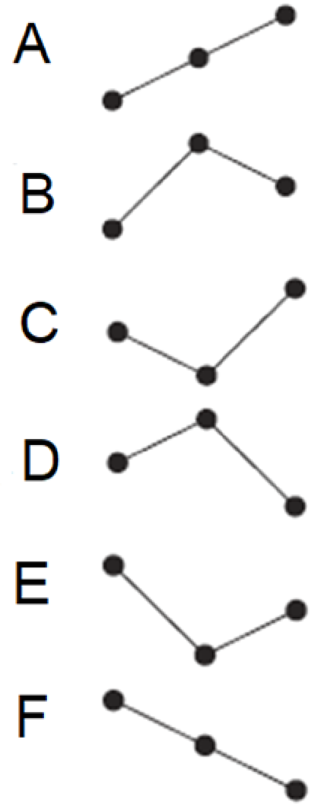
Bandt and Pompe: Phys. Rev. Lett. 2002

cristina.masoller@upc.edu @cristinamasoll1

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

{A, B, F, C}

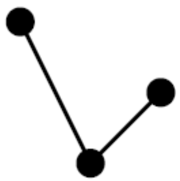




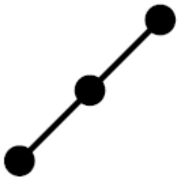
cristina.masoller@upc.edu @cristinamasoll1

Using lagged points to define the patterns allows to select the time scale of the analysis, useful for seasonal data

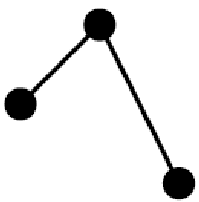
Intra-season

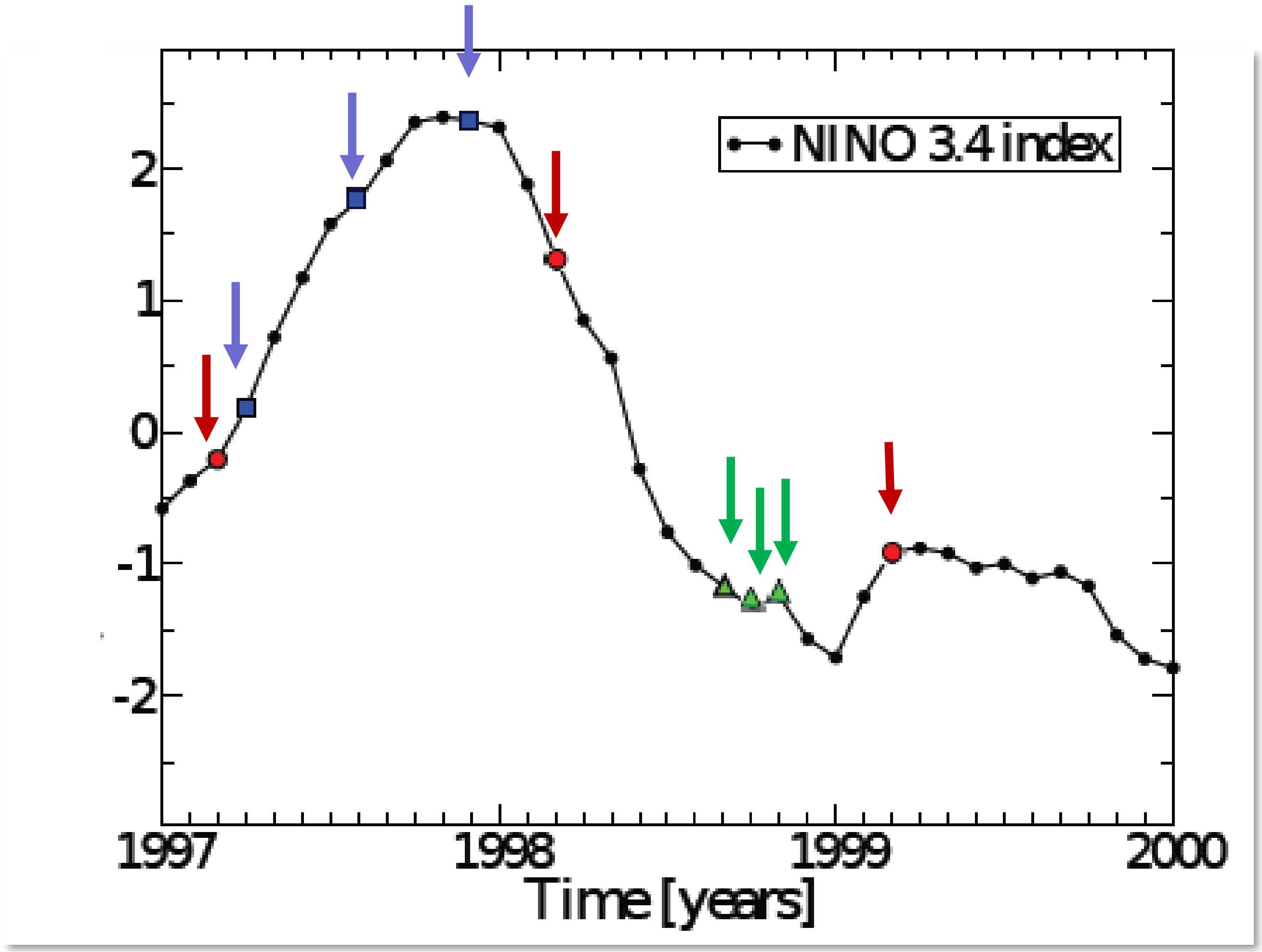


Intra-annual



Inter-annual



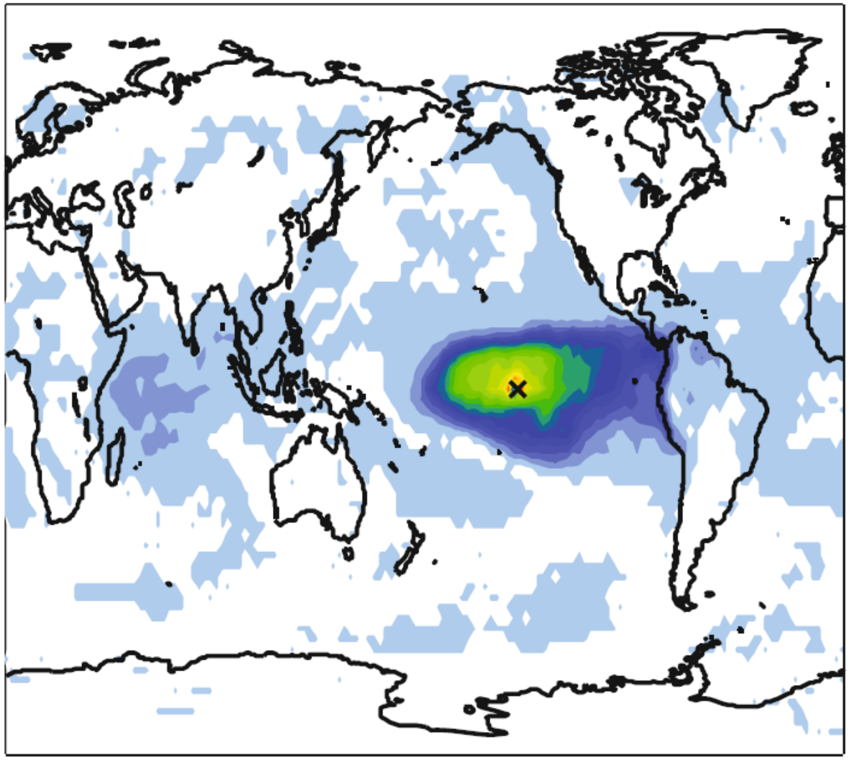


cristina.masoller@upc.edu @cristinamasoll1

Mutual Information of SAT anomaly in El Niño region and other regions (shown before, white: MI not significant)

$$M_{ij} = \sum_{m,n} p_{ij}(m,n) \log \frac{p_{ij}(m,n)}{p_i(m)p_j(n)}$$

MI from probabilities of SAT values



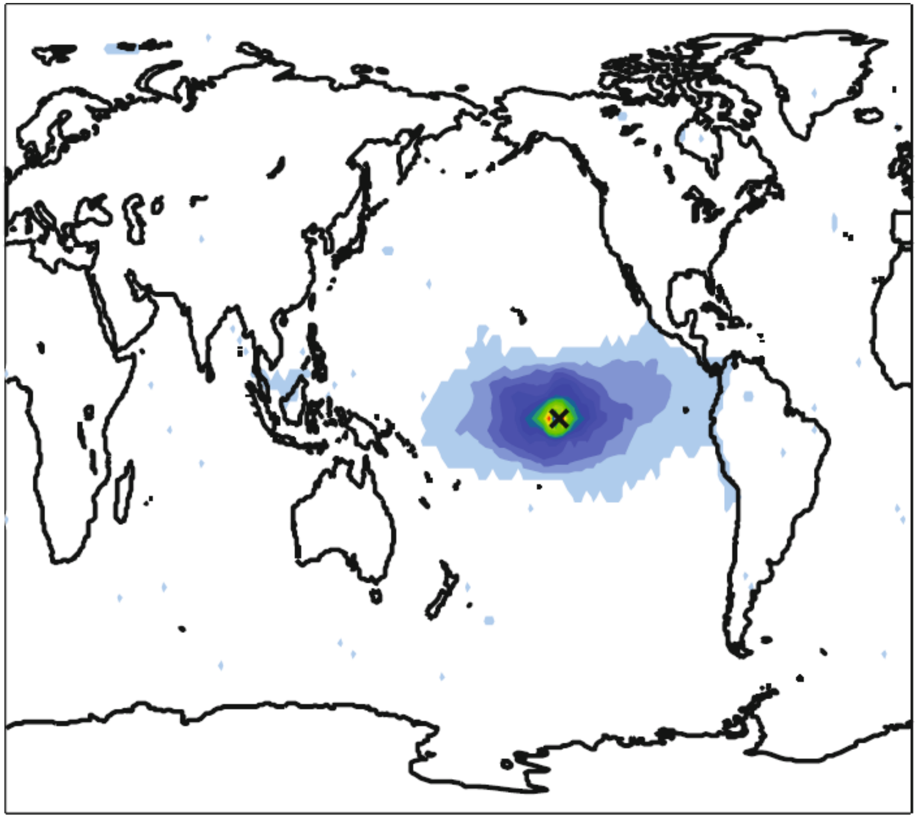
J. I. Deza, M. Barreiro, C. Masoller, Eur. Phys. J. ST 222, 511 (2013)

cristina.masoller@upc.edu @cristinamasoll1

Mutual Information (color code) from probabilities of ordinal patterns (white: MI not significant)

$$M_{ij} = \sum_{m,n} p_{ij}(m,n) \log \frac{p_{ij}(m,n)}{p_i(m)p_j(n)}$$

MI from probabilities of ordinal patterns defined by values in 3 consecutive months.



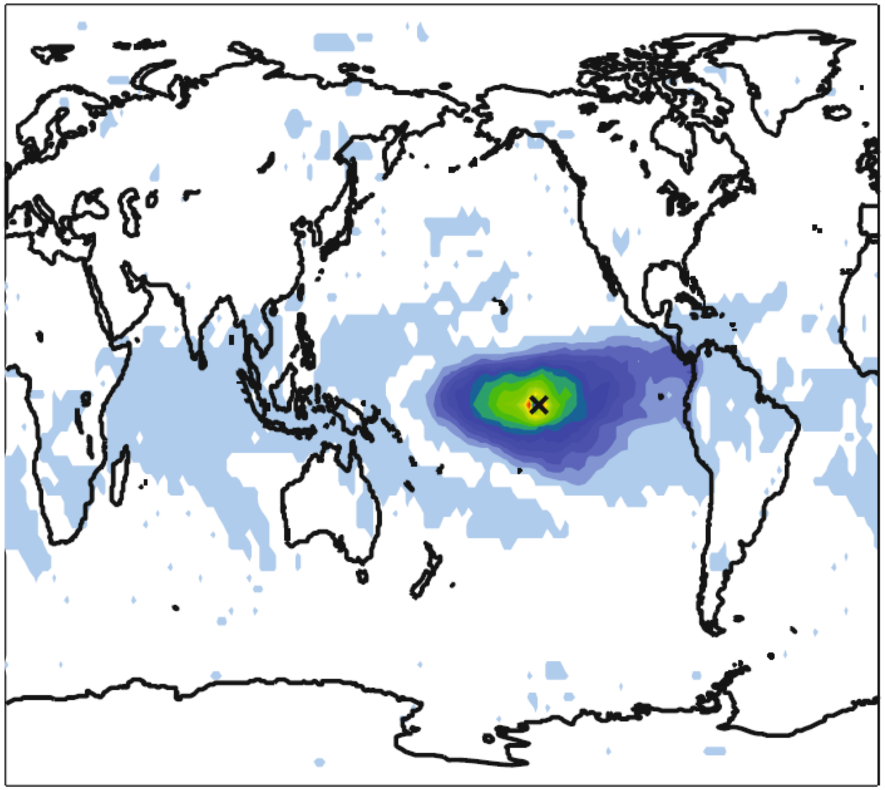
J. I. Deza, M. Barreiro, C. Masoller, Eur. Phys. J. ST 222, 511 (2013).

cristina.masoller@upc.edu @cristinamasoll1

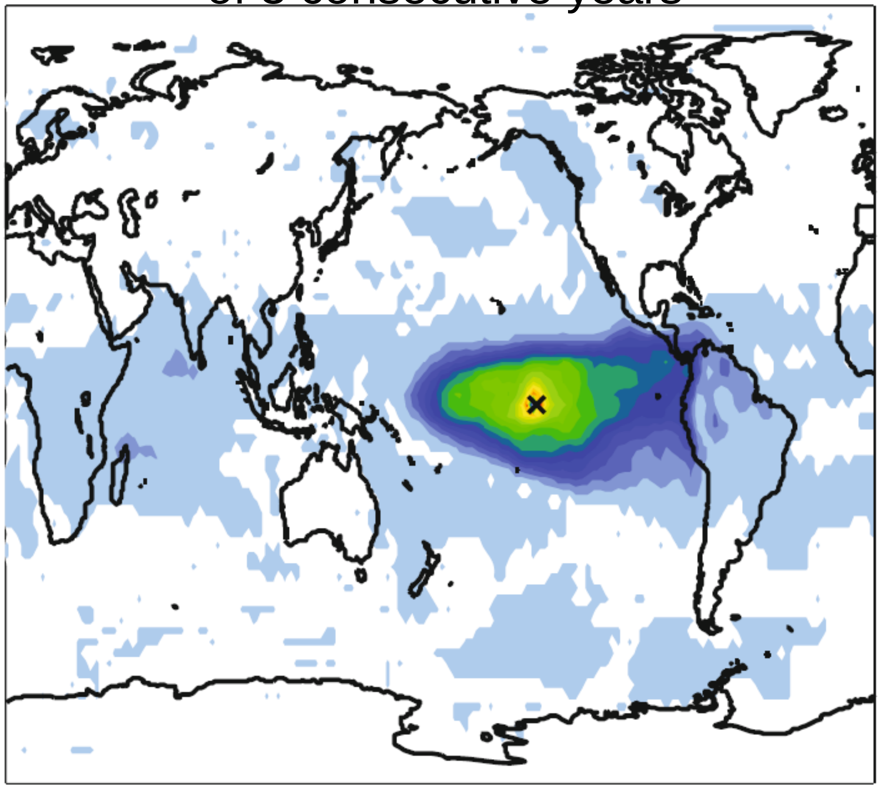
Comparison of two ways to calculate the ordinal probabilities, used to calculate the mutual information

$$M_{ij} = \sum_{m,n} p_{ij}(m,n) \log \frac{p_{ij}(m,n)}{p_i(m)p_j(n)}$$

Patterns defined by 3 values in a year



Patterns defined by data values of 3 consecutive years



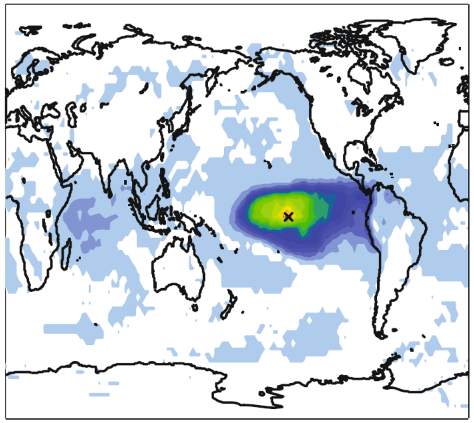
J. I. Deza, M. Barreiro, C. Masoller, Eur. Phys. J. ST 222, 511 (2013).

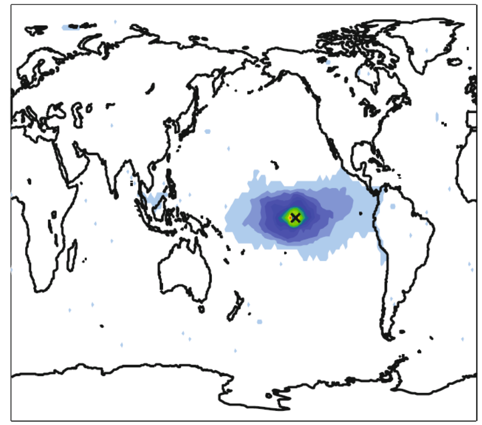
cristina.masoller@upc.edu @cristinamasoll1

Comparison

$$M_{ij} = \sum_{m,n} p_{ij}(m,n) \log \frac{p_{ij}(m,n)}{p_i(m)p_j(n)}$$

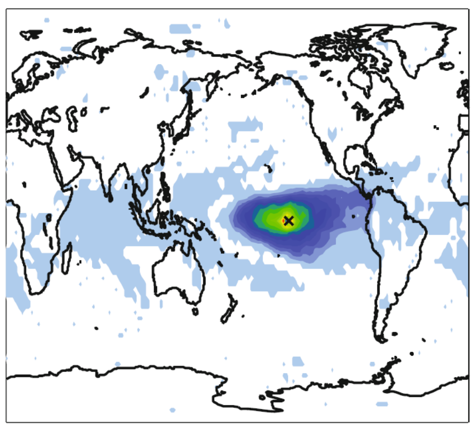
probabilities of SAT values

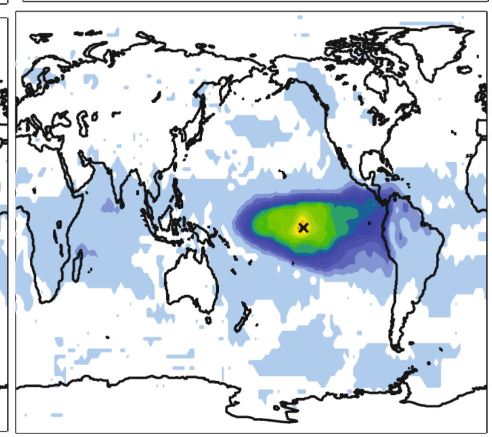




probabilities of ordinal patterns defined by values in 3 consecutive months.

probabilities of patterns defined by 3 values in a year.





probabilities of patterns defined by values in 3 consecutive years.

J. I. Deza, M. Barreiro, C. Masoller, Eur. Phys. J. ST 222, 511 (2013).

cristina.masoller@upc.edu @cristinamasoll1

Take home messages

cristina.masoller@upc.edu @cristinamasoll1

- Hilbert analysis and ordinal analysis are versatile tools that can provide new insights into climate phenomena.
- Mutual information and other information-based measures can be calculated in terms of the probabilities of ordinal patterns, allowing to select the time-scale of the analysis.
- Different large-scale spatial structures are uncovered when using different lags between the data points that define the ordinal patterns.

cristina.masoller@upc.edu @cristinamasoll1

Thank you for your attention

- J. I. Deza, M. Barreiro, C. Masoller, Eur. Phys. J. ST 222, 511 (2013)
- J. I. Deza, M. Barreiro, C. Masoller, Chaos 25, 033105 (2015)
- D. A. Zappala, M. Barreiro, C. Masoller, Earth Syst. Dynamics 9, 383 (2018)
- R. Silini and C. Masoller, Scientific Reports 11, 8423 (2021)
- R. Silini, G. Tirabassi, M Barreiro, L. Ferranti, C. Masoller, Climate Dynamics 61, 79 (2023)

H. Dijkstra, M. Barreiro, E. Hernandez-Garcia, C. Masoller, *Networks in Climate*, Cambridge University Press (2019)

cristina.masoller@upc.edu @cristinamasoll1