Nonlinear time series analysis

Bivariate analysis

Cristina Masoller

Universitat Politecnica de Catalunya, Terrassa, Barcelona, Spain

Cristina.masoller@upc.edu

www.fisica.edu.uy/~cris

UNIVERSITAT POLITÈCNICA DE CATALUNYA BARCELONATECH

Campus d'Excel·lència Internacional

Outline

- **Introduction**
  - Historical developments: from dynamical systems to complex systems
- **Univariate analysis**
  - Methods to extract information from a time series.
  - Applications to climate data.
- **Bivariate analysis**
  - Extracting information from two time series.
  - Correlation, directionality and causality.
  - Applications to climate data.
- **Multivariate analysis**
  - Many time series: complex networks.
  - Network characterization and analysis.
  - Climate networks.

Cross-correlation of two time series *X* and *Y* of length *N*

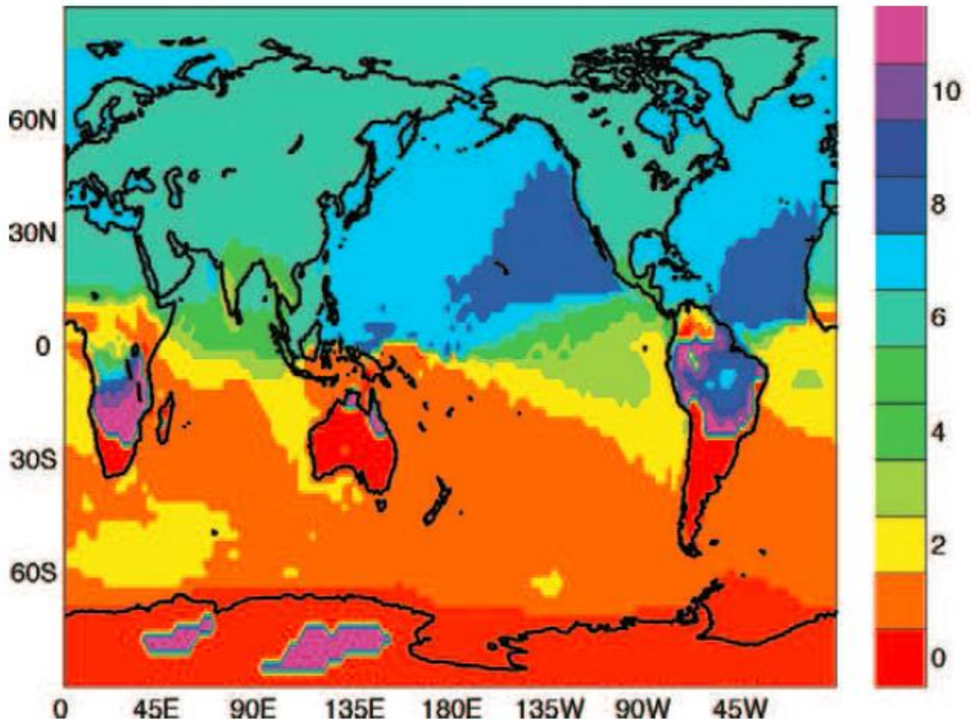
$$C_{xy}(\tau) = \frac{1}{N-\tau}\sum_{k=1}^{N-\tau} x(k+\tau)y(k)$$

the two time series are normalized to zero-mean and unit variance

- $-1 \leq C_{X,Y} \leq 1$
- $C_{X,Y} = C_{Y,X}$
- The maximum of $C_{X,Y}(\tau)$ indicates the **lag** that renders the time series *X* and *Y* best aligned.
- Pearson coefficient: $CC = |C_{X,Y}|$

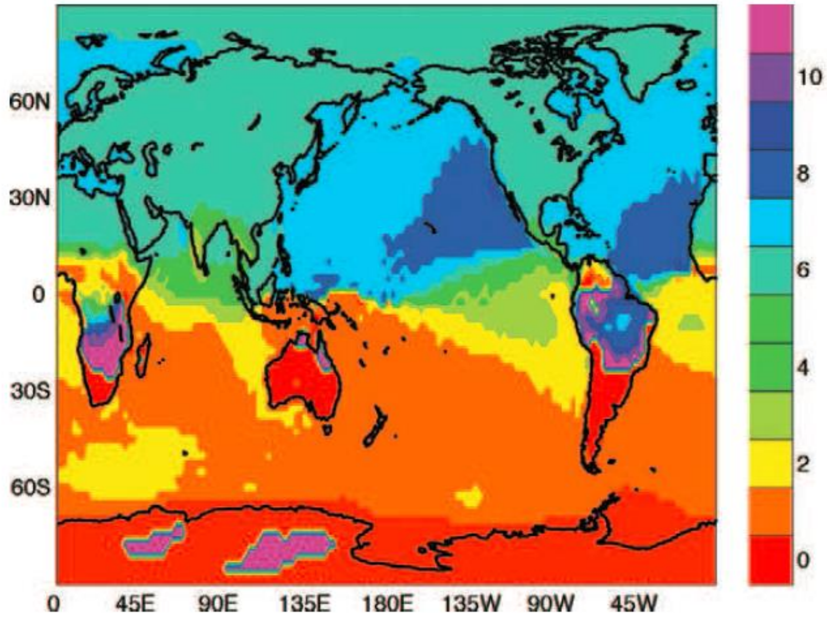
Map of lags between two surface air temperature series

Lag (in months) between the SAT at a reference point in Australia, and all the other time-series.

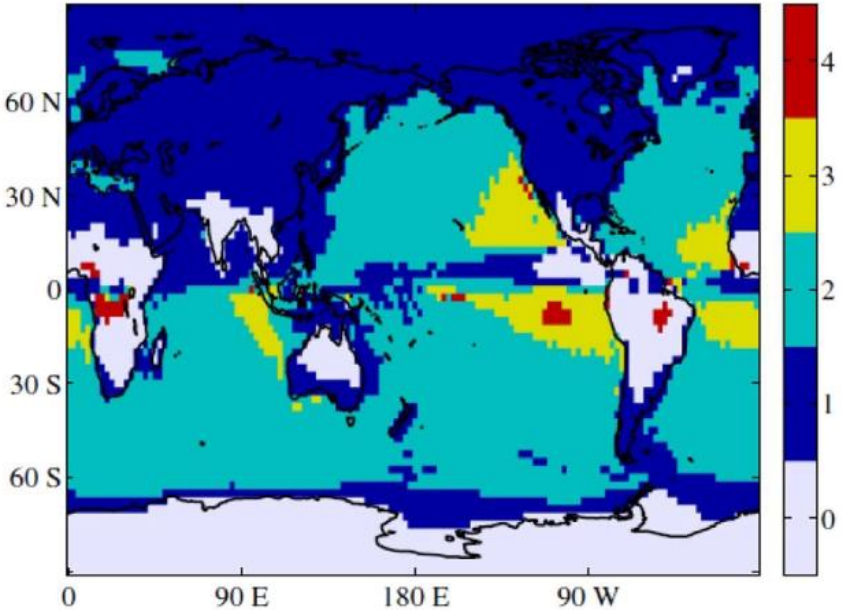


Tirabassi and Masoller EPL 102, 59003 (2013)

Lag times between SAT in different regions

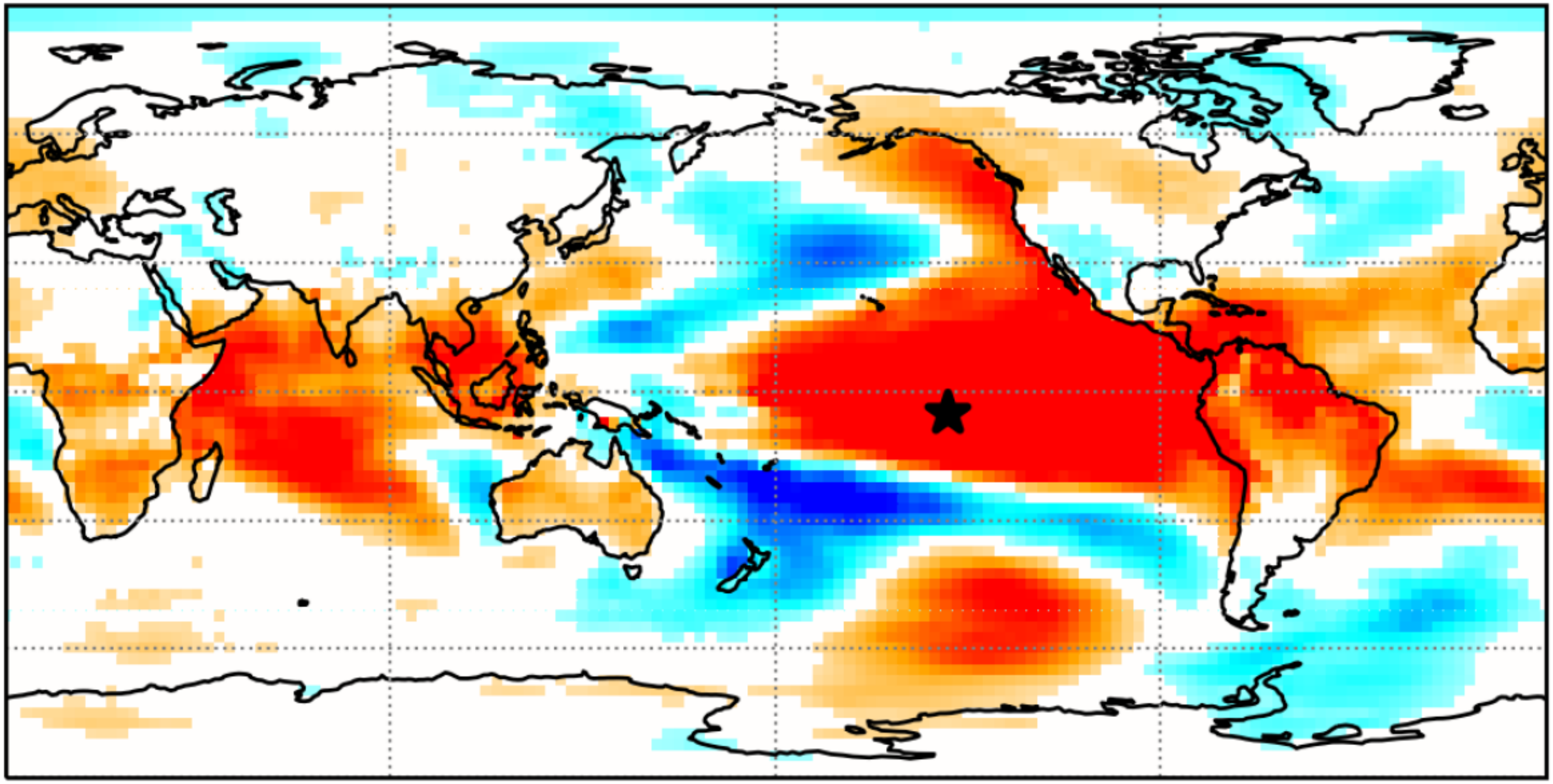


Lag times that minimize the distance between SAT and insolation in the same region



[F. Arizmendi, et al.
Sci. Rep. 7, 45676 (2017).]

An example of cross correlation map: monthly surface air temperature (SAT) anomalies

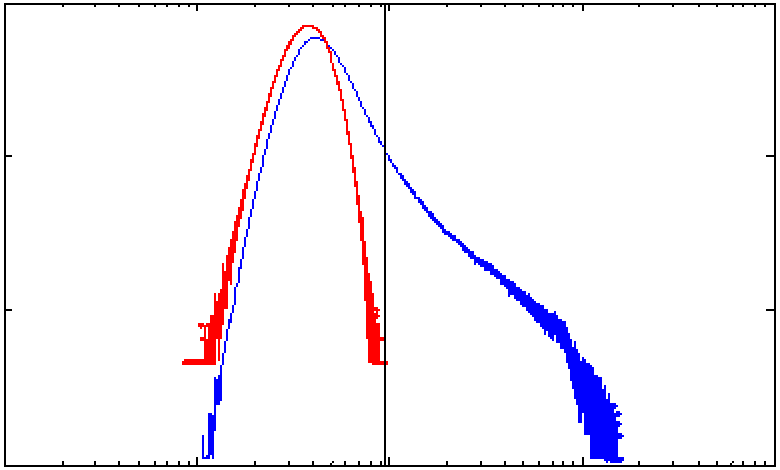


Nonlinear color scale represents the Pearson coefficient: CC = $|C_{X,Y}|$

Are these cross-correlation values significant?

Simplest option: consider statistically significant the values that are larger than those obtained with surrogates.

Prob.
Distrib.
Function
(PDF)

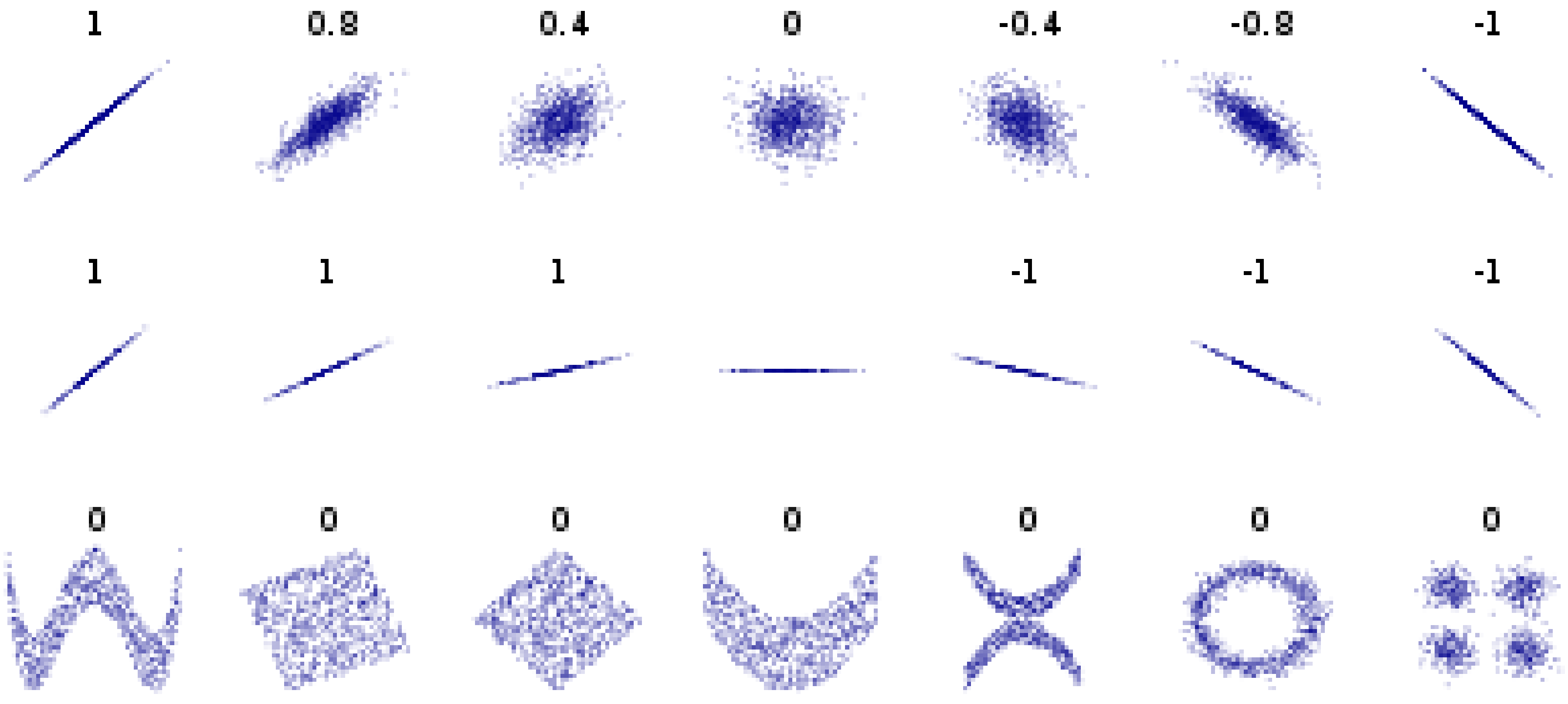


|Cross correlation|

Problems:

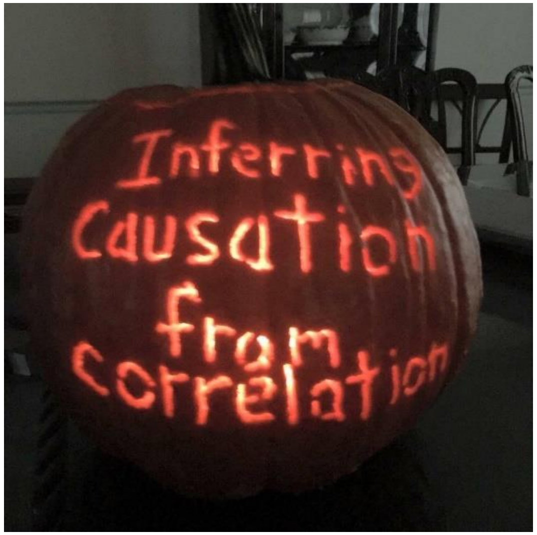
- significant weak links might be hidden by noise
- because of geographical proximity, the strongest CC values are those of neighboring points

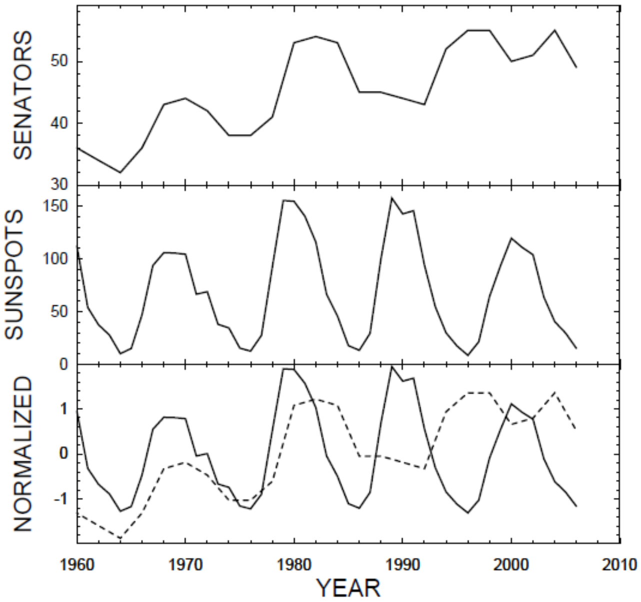
Cross-correlation analysis detects linear relationships only



Source: wikipedia

Correlation is NOT causality





An illustrative example: the number of sunspots and the number of the Republicans in the U.S. Senate in the years 1960-2006.

Considering the period 1960 to 1986 (biannual sampling, 14 points): **C=0.52** **Is this significant?**

Null hypotesis

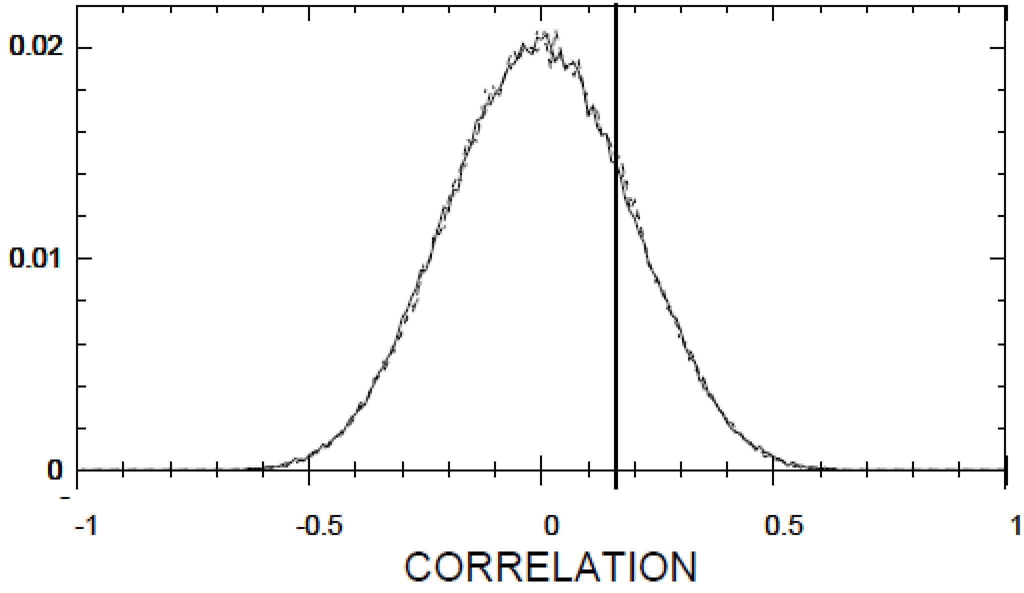
- Assuming the data sets were sampled from *independent, identically distributed* (IID) Gaussian populations and a significance level of 95%, then the significance threshold value of C is 0.458.

- Therefore, the null hypothesis (Gaussian IID) should be rejected.

- Something is wrong!

The analysis of surrogate data produces three identical distributions

- Between the number of the Republican senators in the period 1960-2006 (24 samples) with 24-sample sets randomly drawn from the Gaussian distribution (dashed);
- Between the number of the Republican senators in the period 1960-2006 (24 samples) with the 24-sample segment of the sunspot numbers randomly permutated in the temporal order (IID surrogate, dash-and-dotted)
- Two 24-sample sets randomly drawn from a Gaussian distribution (solid).



Vertical line: correlation between the number of the Republican senators and the sunspot numbers for the period 1960-2006.

What was wrong?

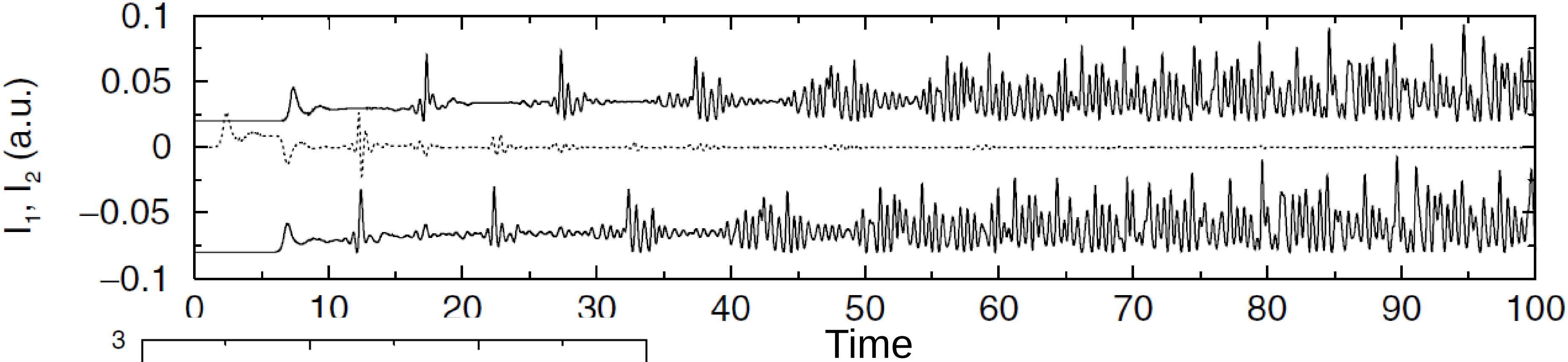
- The significance criterion C> 0.458 is not valid because the two datasets do not meet the *independent, identically distributed* (IID) criterion.
- IID samples: there is no relation between any x_i and x_{i+j}.
- But in both datasets there are **autocorrelations**.
- No universal table of critical values can be derived for testing the independence of serially correlated data sets.

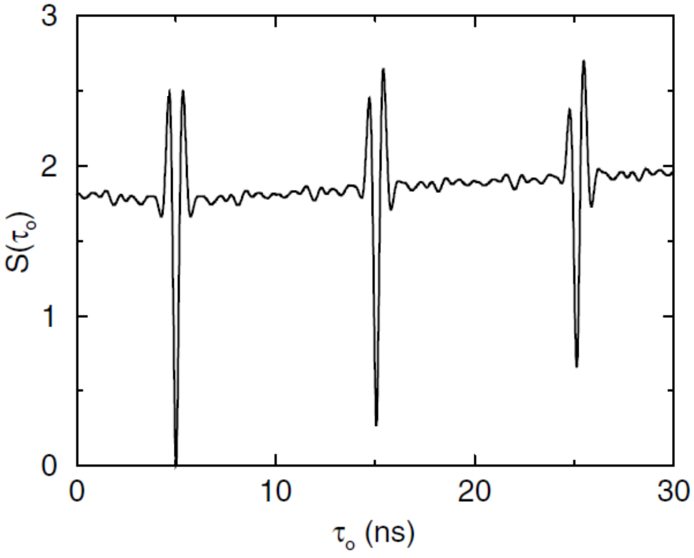
Read more: M. Palus, *From Nonlinearity to Causality: Statistical testing and inference of physical mechanisms underlying complex dynamics*. Contemporary Physics 48(6) (2007) 307-348.

Similarity function: similar to cross-correlation

$$S^2(\tau_0) = \frac{\langle [I_1(t + \tau_0) - I_2(t)]^2 \rangle}{[\langle I_1(t)^2 \rangle \langle I_2(t)^2 \rangle]^{1/2}}$$

Example: detection of anticipated synchronization in one-way coupled ($I_1 \rightarrow I_2$) chaotic systems.





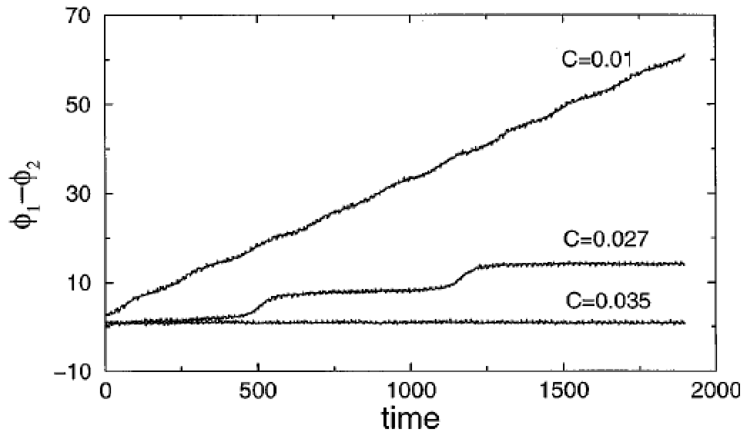
C. Masoller, Phys. Rev. Lett. 86, 2782 (2001)

Phase synchronization (PS)

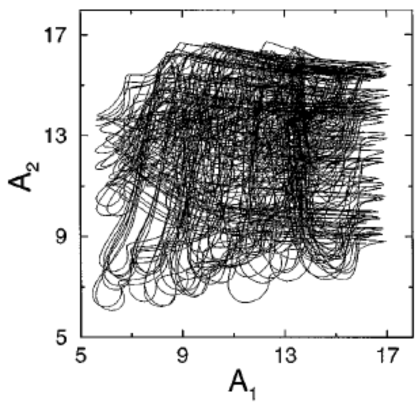
- The phase difference (relative phase) between two oscillators is bounded but their amplitudes are not synchronized.

Mutually coupled
Rossler systems

$$\dot{x}_{1,2} = -\omega_{1,2} y_{1,2} - z_{1,2} + C(x_{2,1} - x_{1,2}),$$
$$\dot{y}_{1,2} = \omega_{1,2} x_{1,2} + 0.15 y_{1,2},$$
$$\dot{z}_{1,2} = 0.2 + z_{1,2}(x_{1,2} - 10).$$



C=0.035

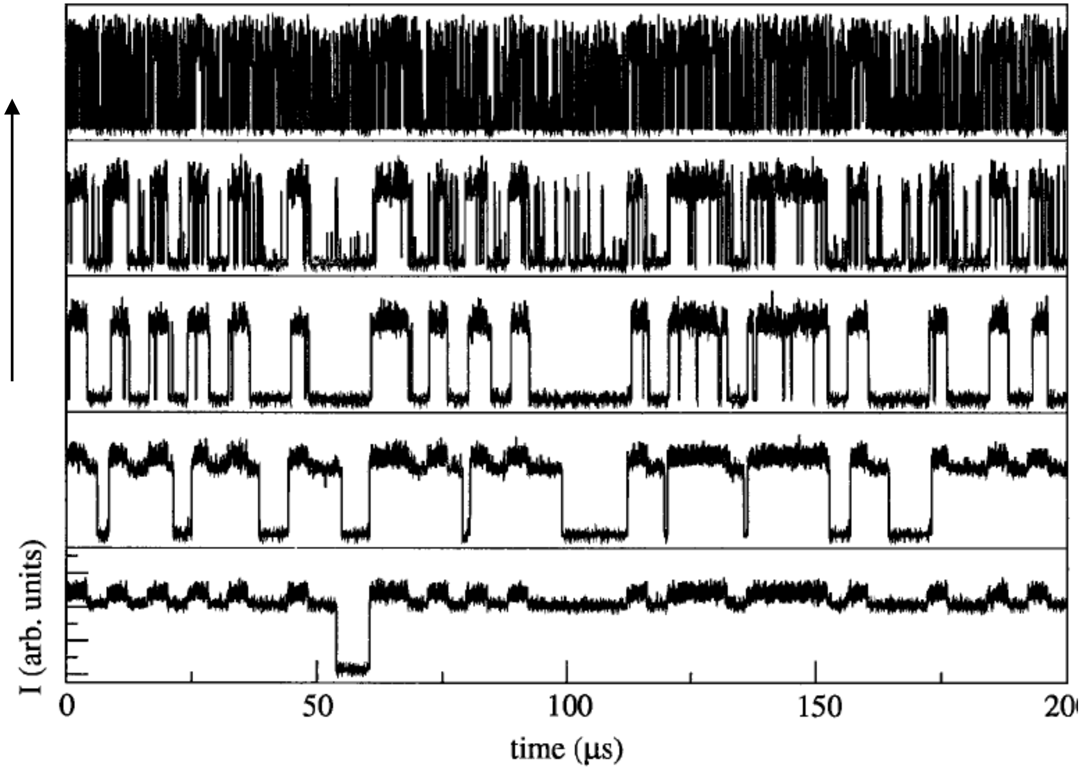


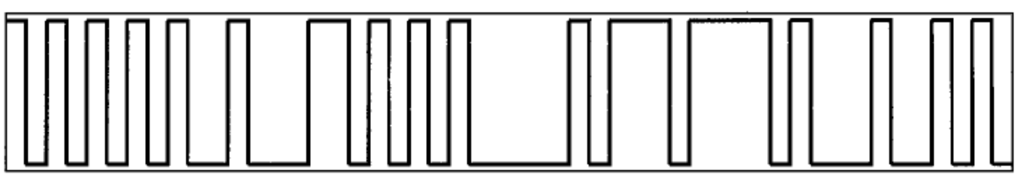
- Several measures have been proposed to detect PS in real signals.
- Main Idea: If two signals are phase synchronized, the relative phase will occupy a small portion of the unit circle, while the lack of PS gives a relative phase that spreads out over the entire unit circle.

Rosenblum et al., Phys. Rev. Lett. 76, 1804 (1996)
E. Pereda et al., Progress in Neurobiology 77, 1 (2005)

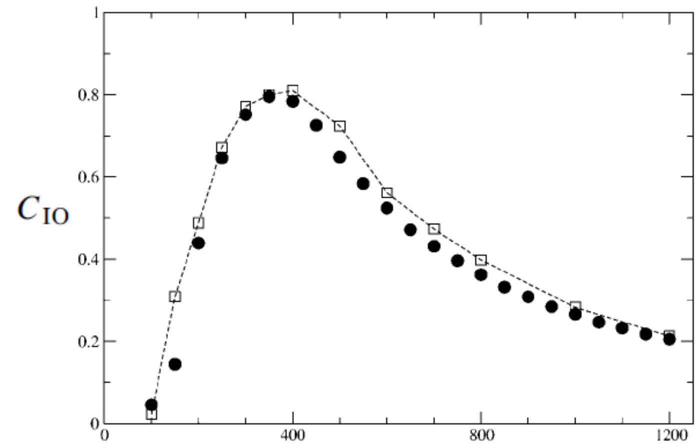
Example: stochastic resonance
Response of a bistable system to an aperiodic signal

Noise level





$$C_{IO} = \max_{\tau}\{\overline{[x_{in}(t) - \overline{x}_{in}][x_{out}(t + \tau) - \overline{x}_{out}]}\}$$



Noise level

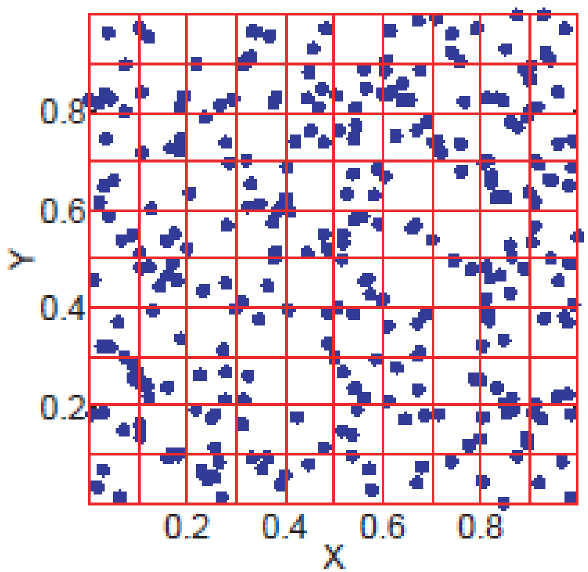
Barbay et al, PRL 85, 4652 (2000)

Mutual Information

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

- *MI (x,y) = MI (y,x)*
- *p(x,y) = p(x) p(y)* $\Rightarrow$ *MI = 0*, else ***MI >0***
- *MI* can also be computed with a lag-time.

MI values are systematically overestimated



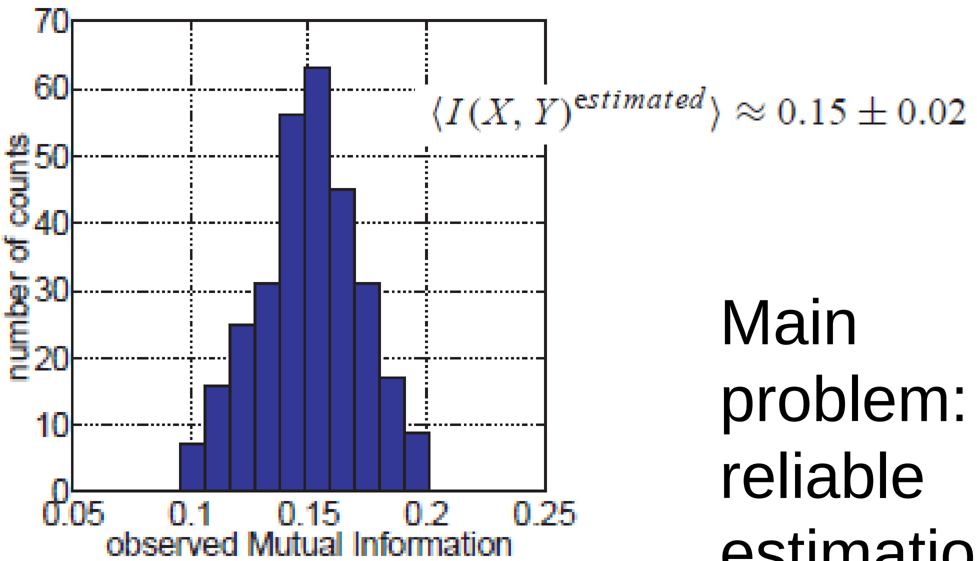


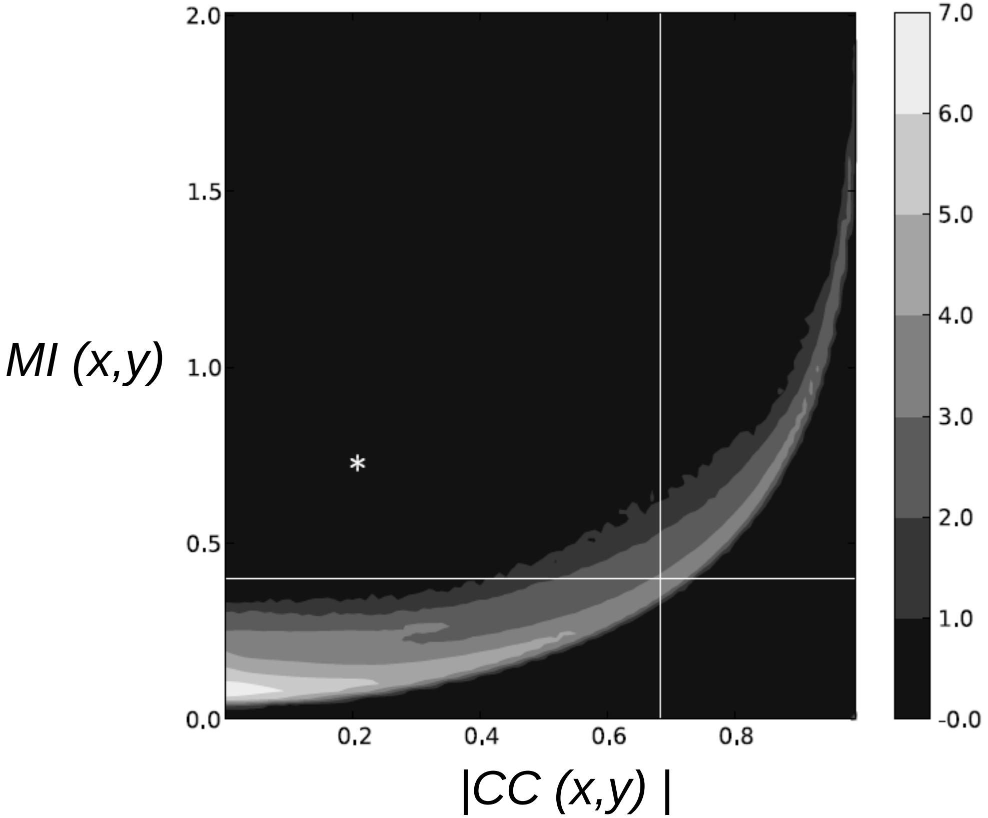
Fig. 1. Naive estimation of the mutual information for finite data. Left: The dataset consists of $N = 300$ artificially generated independent and equidistributed random numbers. The probabilities are estimated using a histogram which divides each axis into $M_X = M_Y = 10$ bins. Right: The histogram of the estimated mutual information $I(X, Y)$ obtained from 300 independent realizations.

Main problem: a reliable estimation of MI requires a large amount of data

R. Steuer et al, Bioinformatics 18, suppl 2, S231 (2002).

Relation between cross-correlation and mutual information

- Depends on the data.
- Here computed from 6816 x 6816 SAT anomaly series.



2D histogram; the color represents the number of elements in each bin in log scale

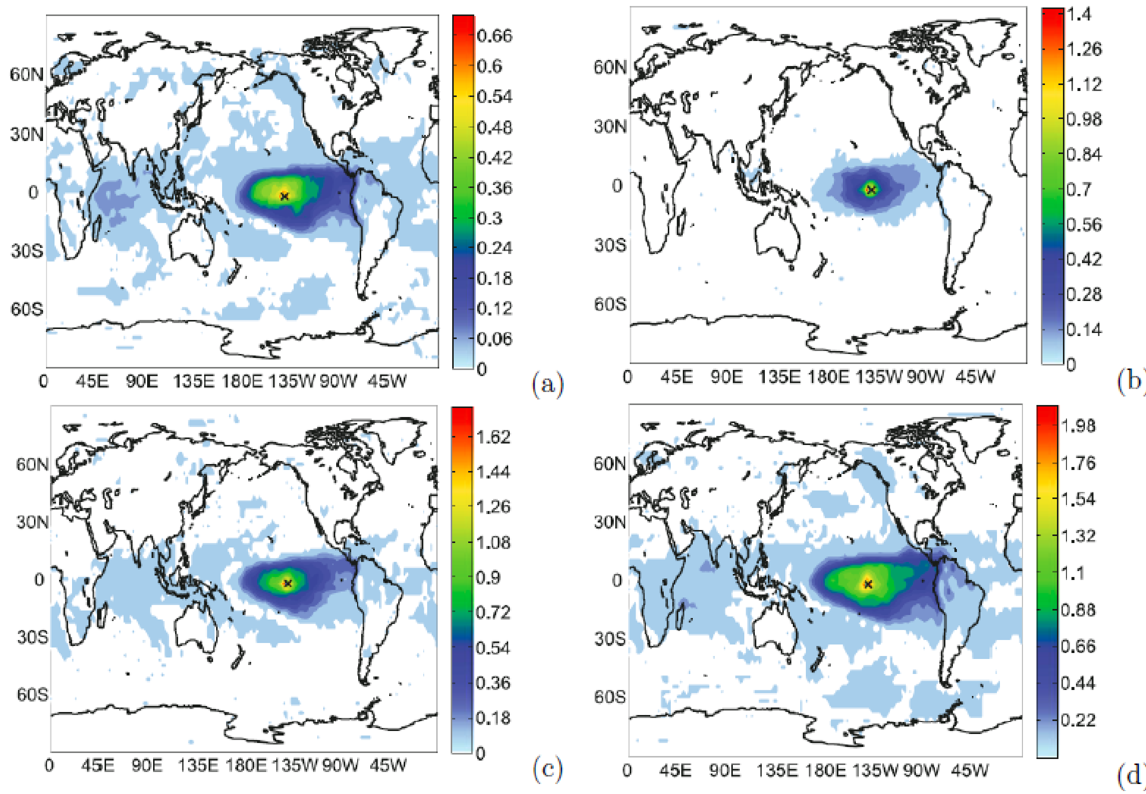
Donges et al, Eur. Phys. J. Special Topics 174, 157 (2009)

Mutual information maps

- MI between SAT anomalies time-series at a reference point located in El Niño, and all the other time-series.

Histograms

Inter-annual ordinal patterns

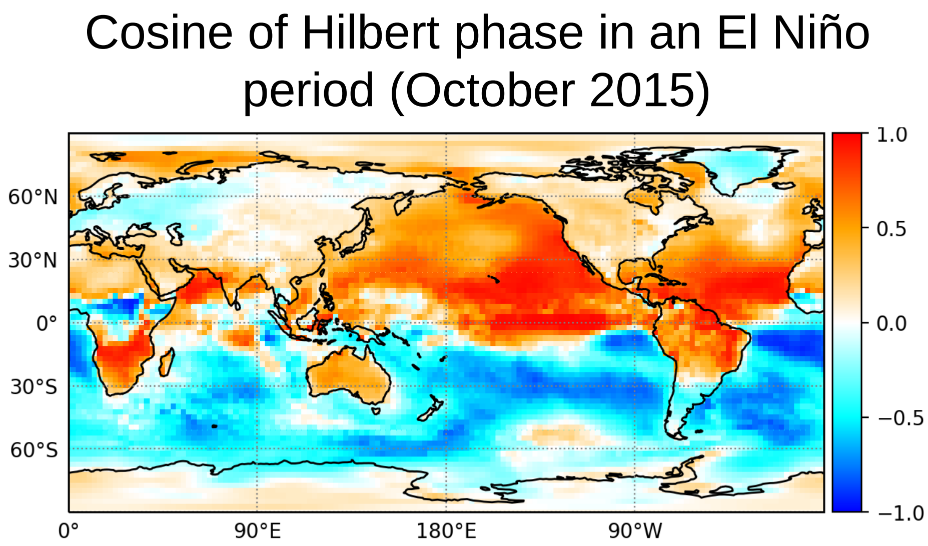


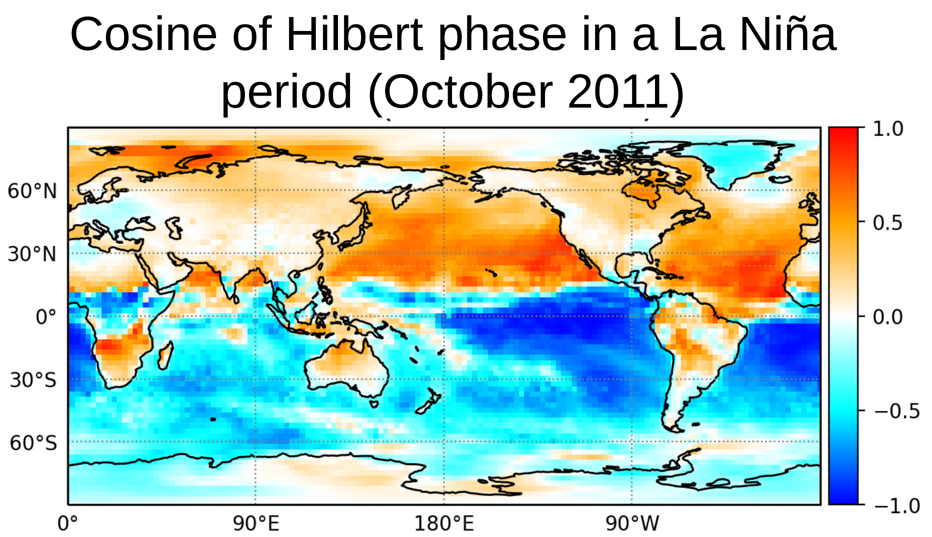
3 months ordinal patterns

3 years ordinal patterns

Ordinal analysis separates the times-scales of the interactions

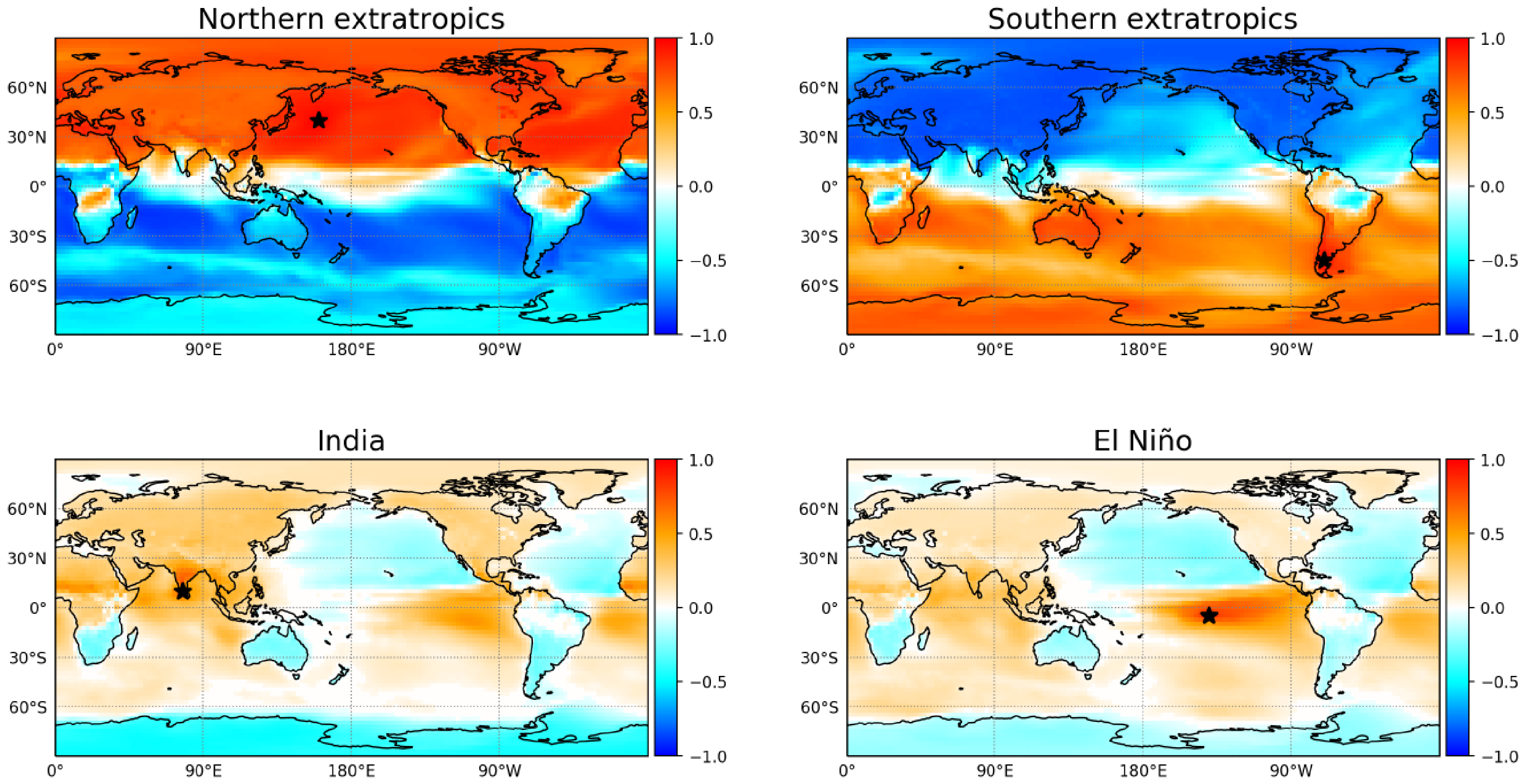
Deza, Barreiro and Masoller, Eur. Phys. J. ST 222, 511 (2013)





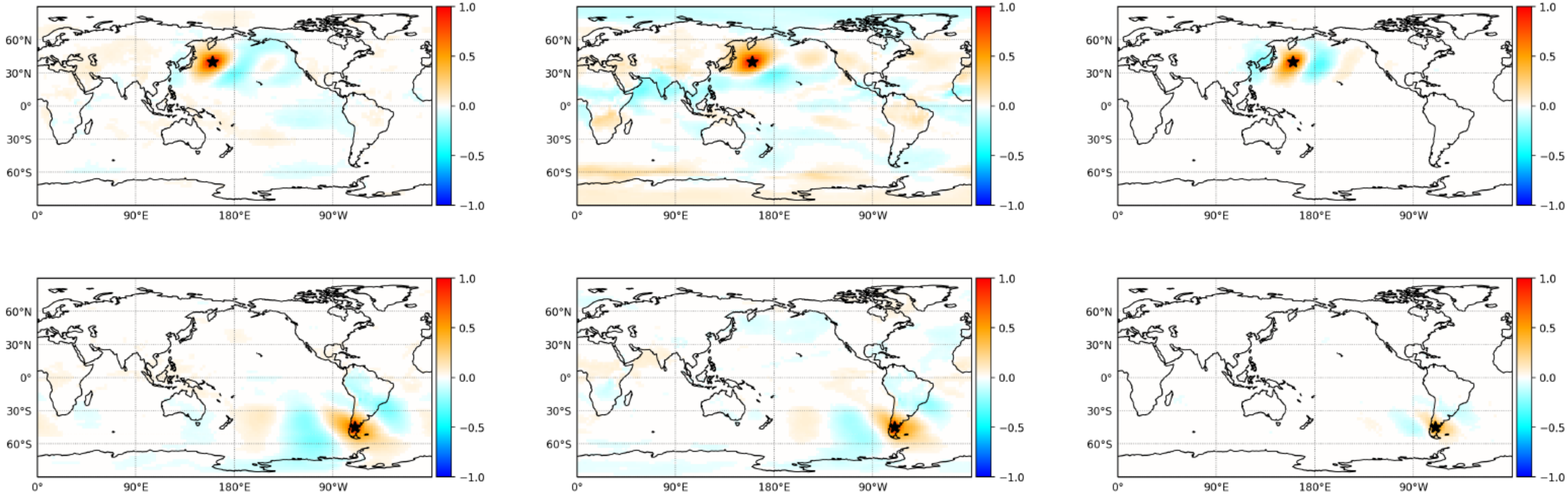
What connectivity patterns we infer using Hilbert analysis?

Cross-correlation of cosine of Hilbert phase



Cross-correlations in the extra-tropics

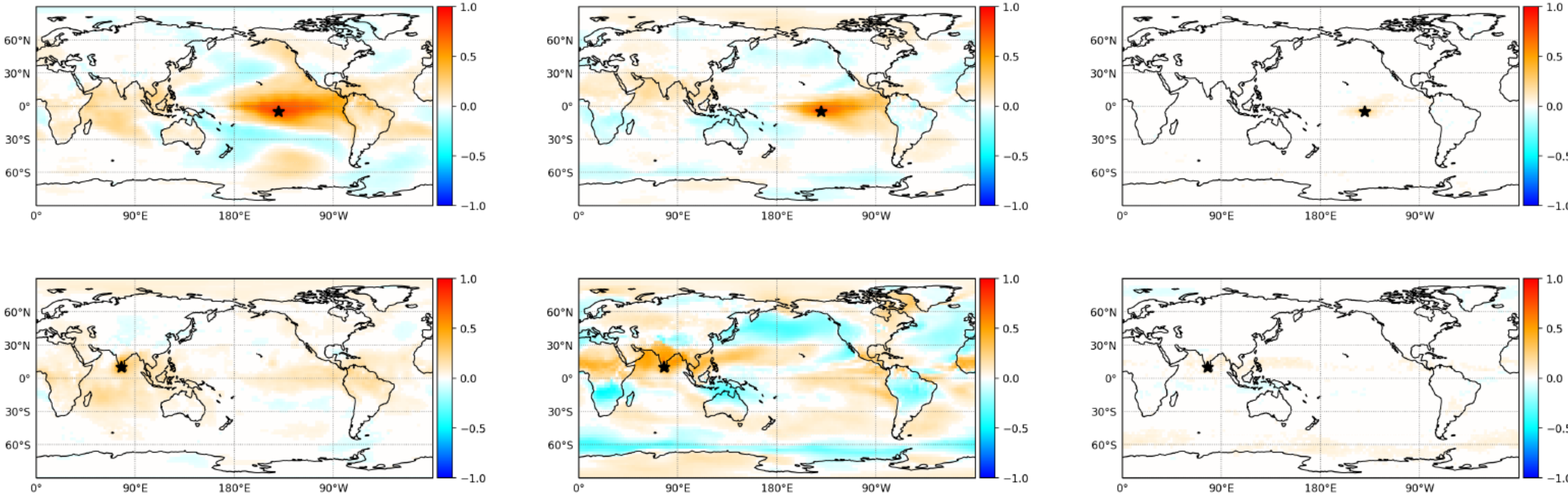
SAT Anomalies | Hilbert amplitude | Hilbert frequency



Zappala, Barreiro & Masoller (in preparation)

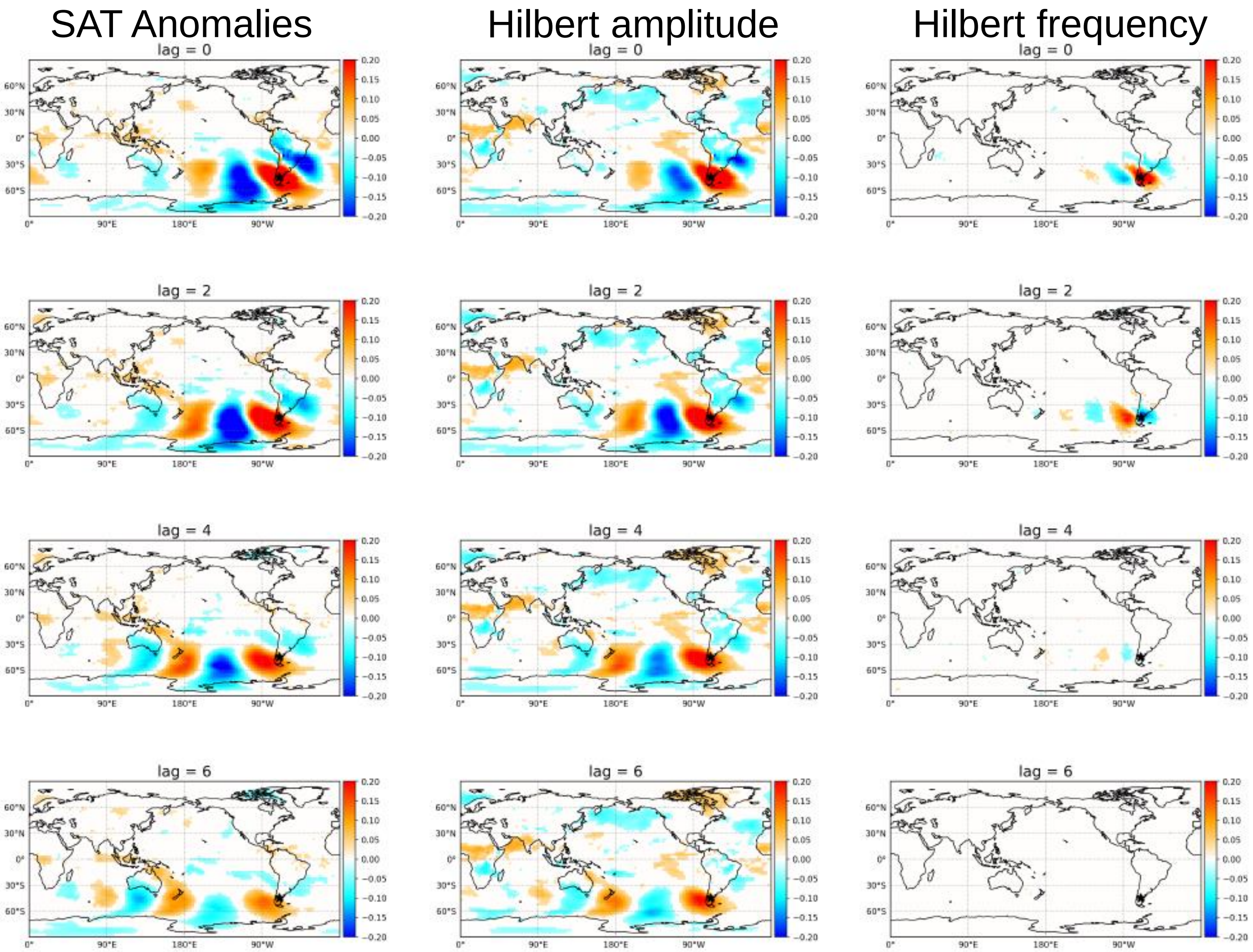
Cross-correlations in the tropics

SAT Anomalies | Hilbert amplitude | Hilbert frequency



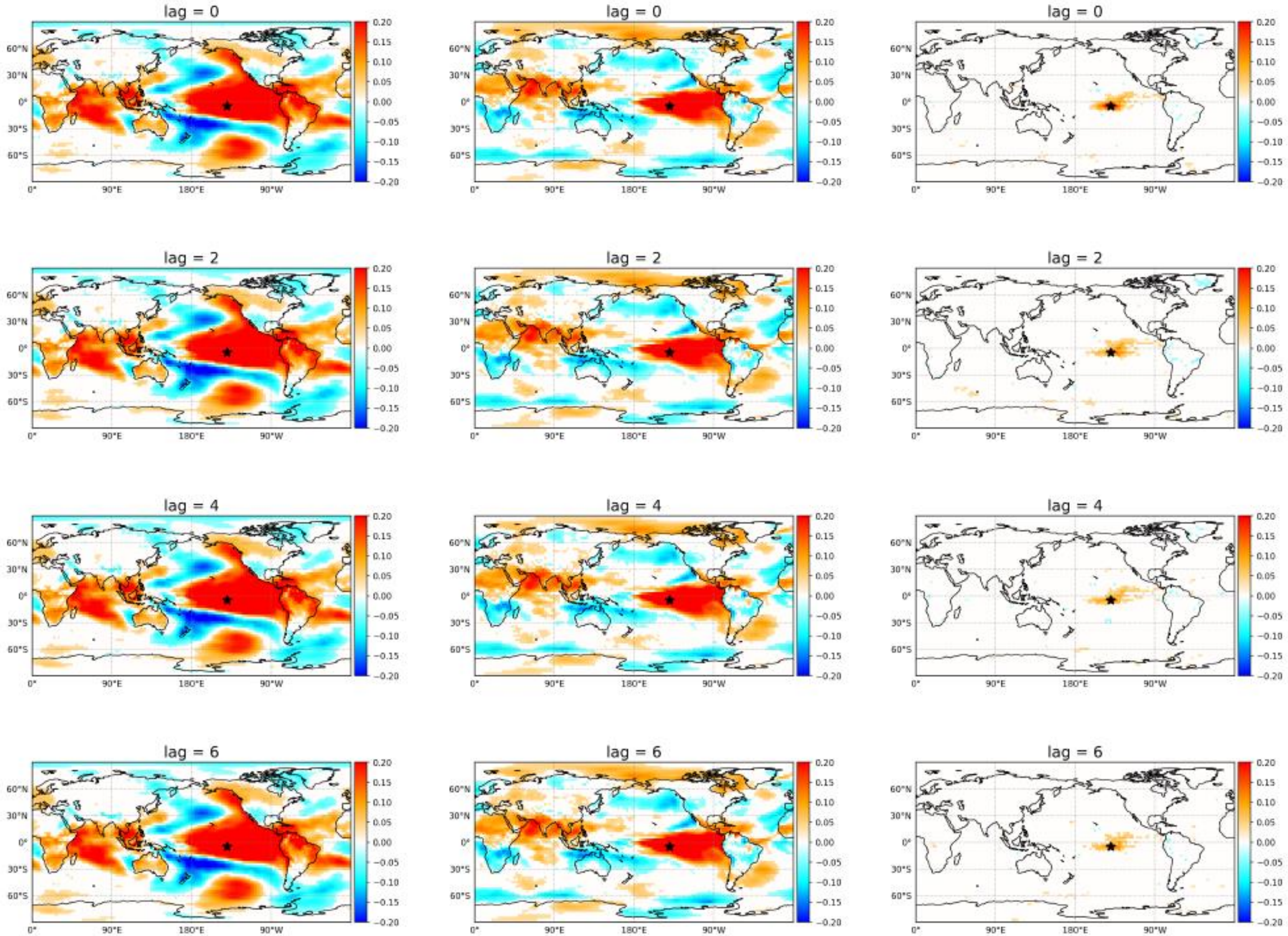
Zappala, Barreiro & Masoller (in preparation)

Influence of the lag time in extra-tropics



Influence of the lag time in El Niño region

SAT Anomalies | Hilbert amplitude | Hilbert frequency



Directionality of information transfer?

Conditional mutual information (CMI) and transfer entropy (TE)

- CMI measures the amount of information shared between two time series *i(t)* and *j(t)*, given the effect of a third time series, *k(t)*, over *j(t)*.

$$M_I(i;j|k) = \sum_{m,n,l} p_{ijk}(m,n,l) \log \frac{p_k(l)p_{ijk}(m,n,l)}{p_{ik}(m,l)p_{jk}(n,l)}$$

- Transfer entropy = CMI with the third time series, *k(t)*, replaced by the *past* of *i(t)* or *j(t)*.

$$\mathrm{TE}_{ij}(\tau) \equiv M_I(i;j|i_\tau) \qquad \mathrm{TE}_{ji}(\tau) \equiv M_I(j;i|j_\tau)$$

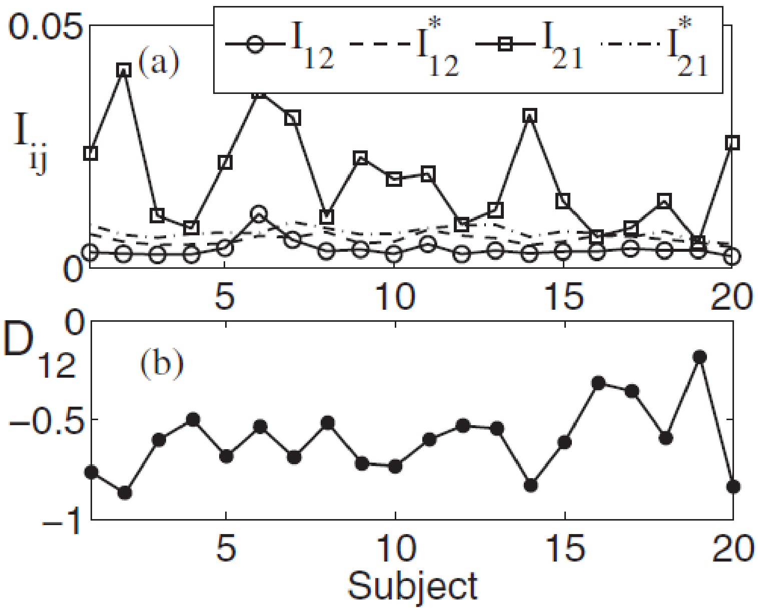
Directionality index

- τ: *time-scale* of information transfer
- *DI*: *net direction* of information transfer
- $DI_{ij} > 0 \rightarrow i$ drives j.

$$\mathrm{DI}_{ij}(\tau) = \frac{\mathrm{TE}_{ij}(\tau) - \mathrm{TE}_{ji}(\tau)}{\mathrm{TE}_{ij}(\tau) + \mathrm{TE}_{ji}(\tau)}$$

Problem: $x \rightarrow i$, $x \rightarrow j$ $\quad i \leftrightarrow j$??

Application to cardiorespiratory data measured from 20 healthy subjects:
(a) TEs (dashed lines: surrogate data)
(b) D_{12} (1 = heart; 2 = respiration).



$D_{12} < 0 \rightarrow$ respiration is drives cardiac activity.

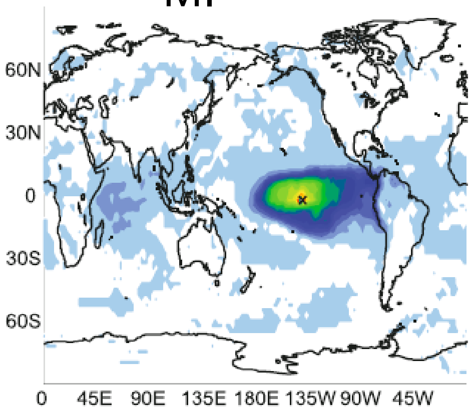
TEs were computed from ordinal probabilities and averaged over a short range of lags to decrease fluctuations.

A. Bahraminasab et al., PRL 100, 084101 (2008)

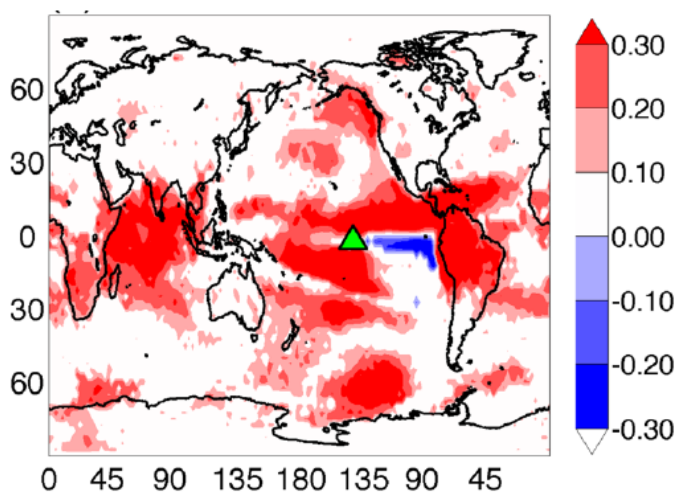
Application to climate data

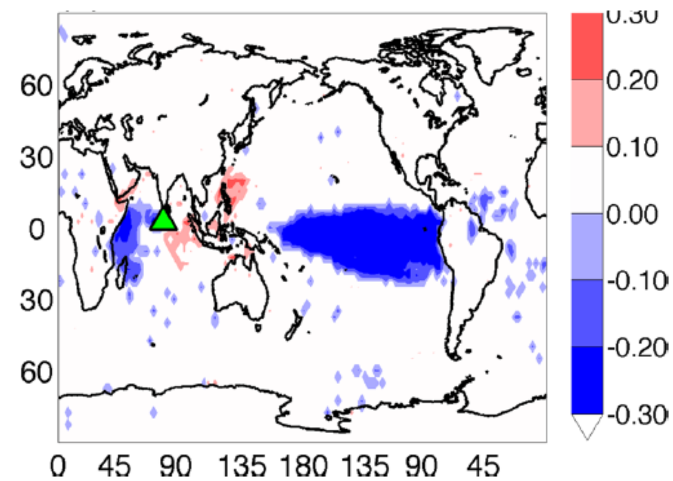
DI computed from daily SAT anomalies, PDFs estimated from histograms of values.
MI and DI are *both significant* (>3σ, bootstrap surrogates), τ=30 days.

MI



DI





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).

Influence of the time-scale of information transfer

τ=1 day

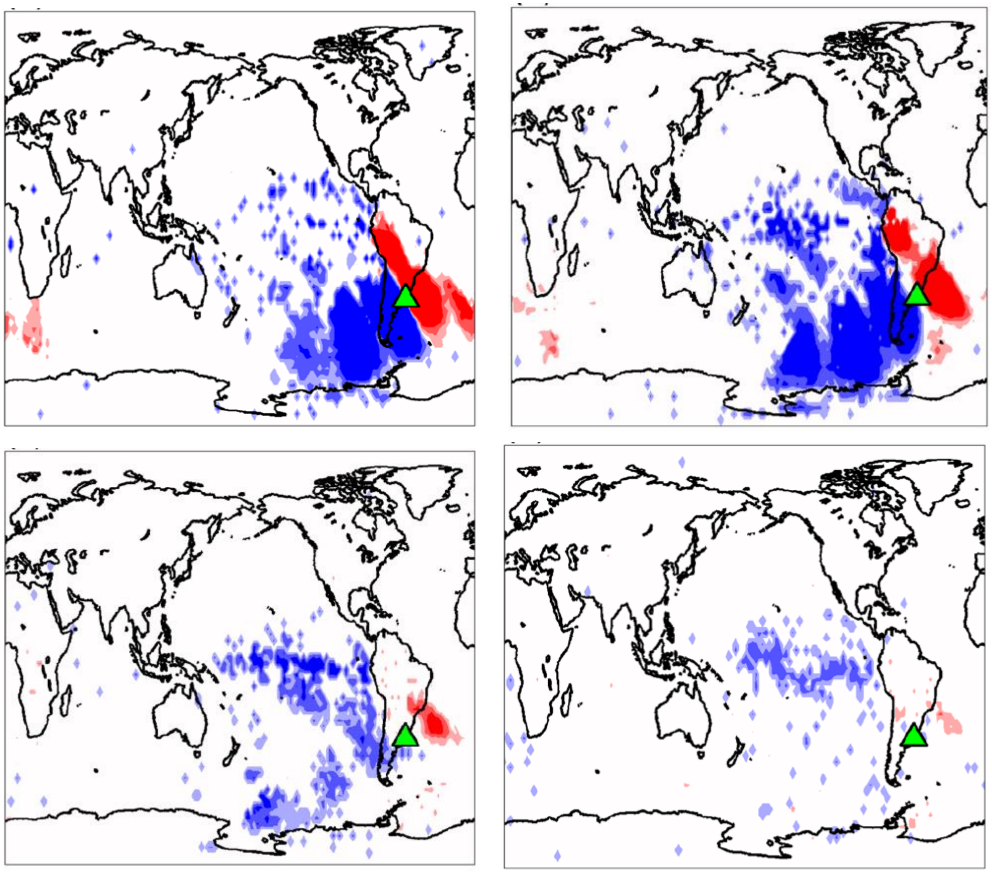
τ=3 days

Video SH

Video NH

τ=7 days

τ=30 days

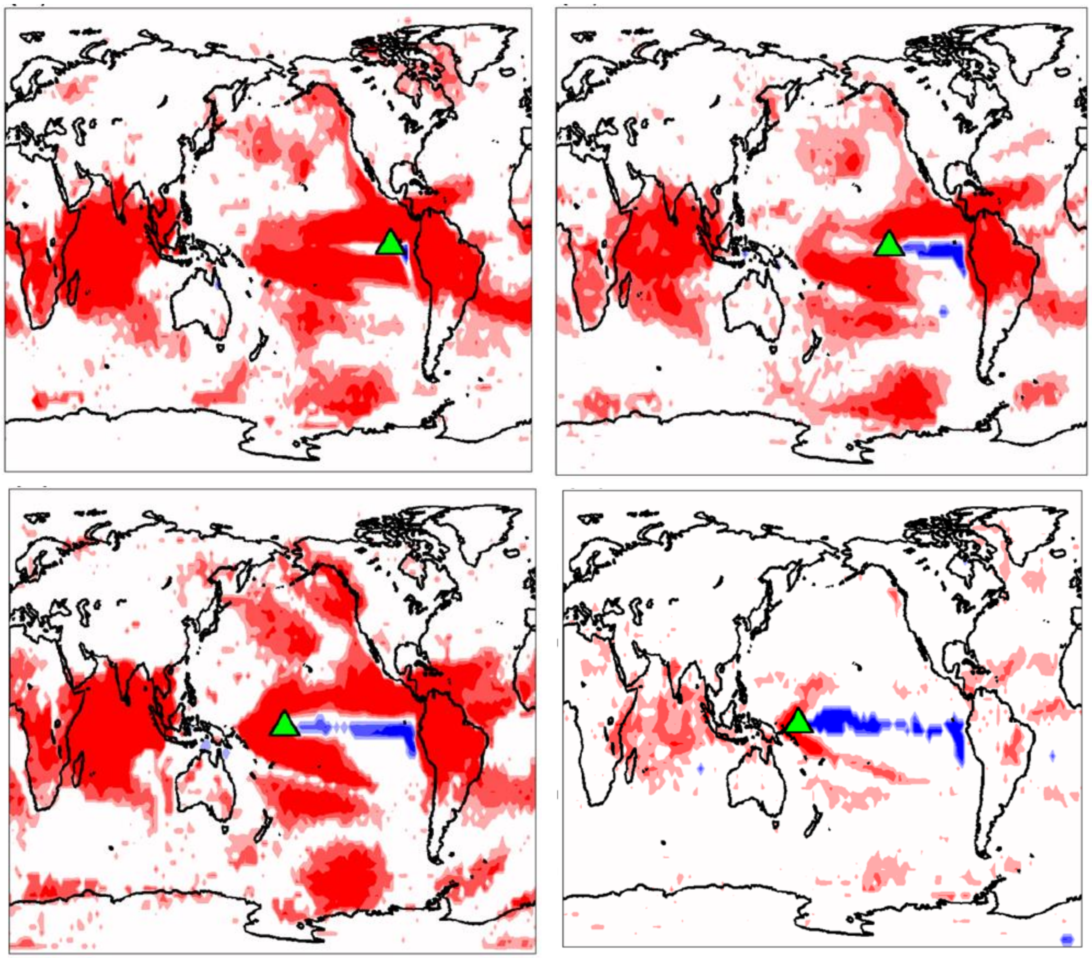


Link directionality reveals wave trains propagating from west to east.

Videos in El Niño, El Labrador and Rio de la Plata, when τ increases from 1 to 180 days

Deza, Barreiro and Masoller, Chaos 25, 033105 (2015)

Link directionality in El Niño area (τ=30 days)



[Deza, Barreiro and Masoller, Chaos 25, 033105 (2015)](#) [Video](#)

Causality?

Granger causality

- Main idea: A time series *X* is Granger causal to a time series *Y* (*X*→*Y*) if the information given by *X* allows for a more precise prediction of *Y*.
- Method: model *Y* as a AR(d) processes forced by *X* with residual noise ε

$$Y_t = \sum_{i=1}^{d} a_i Y_{t-i} + \sum_{i=1}^{d} b_i X_{t-i} + \epsilon_t$$

- Test the hypothesis *b* ≠ 0 against the null hypothesis *b*=0.

  To do this

  - Fit vectors *a* and *b* with a linear regression and compute the variance of the residual: $\sigma^2_{\text{coupled}}$
  - Repeat with *b*=0 and compute: $\sigma^2_{\text{uncoupled}}$

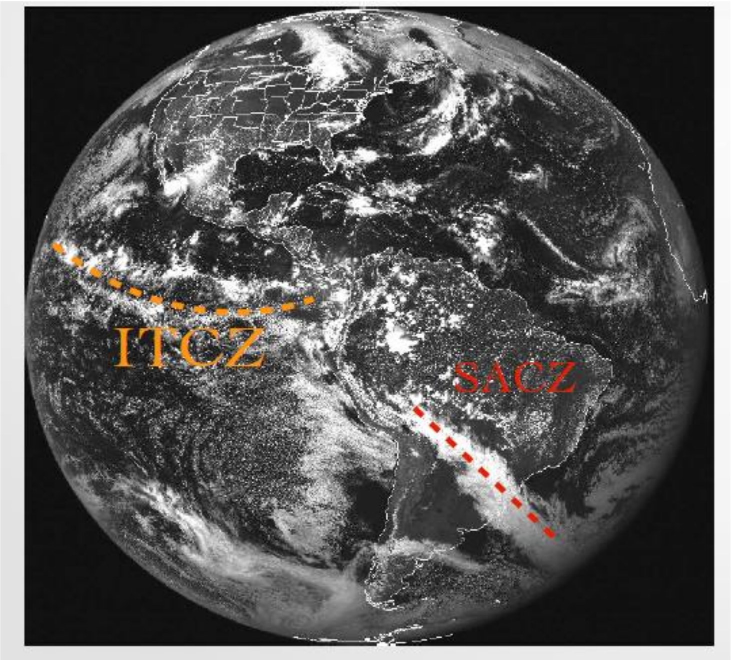
Granger causality estimator

$$GCE = \frac{\sigma^2_{\text{uncoupled}} - \sigma^2_{\text{coupled}}}{\sigma^2_{\text{uncoupled}}}$$

- If *GCE*>0 the information given by *X* allowed for a more precise prediction of *Y*.
- Problems:
  - how to select the dimension *d*?
  - how to test the statistical significance of the GCE value?

Read more: G. Tirabassi, C. Masoller and M. Barreiro, "*A study of the air-sea interaction in the South Atlantic Convergence Zone through Granger Causality*", Int. J. of Climatology, 35, 3440 (2015)

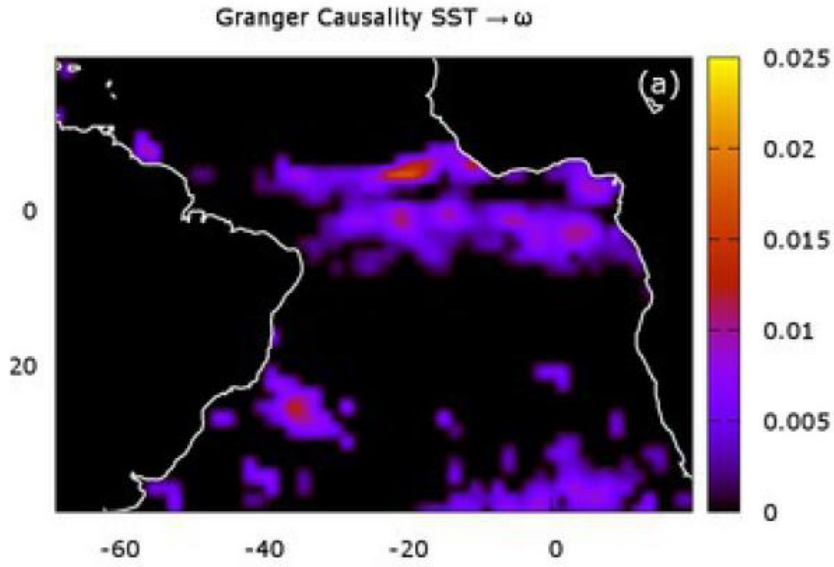
Other methods to detect causality: G. Tirabassi, L. Sommerlade and C. Masoller, "*Inferring directed climatic interactions with renormalized partial directed coherence and directed partial correlation*", Chaos 27, 035815 (2017)



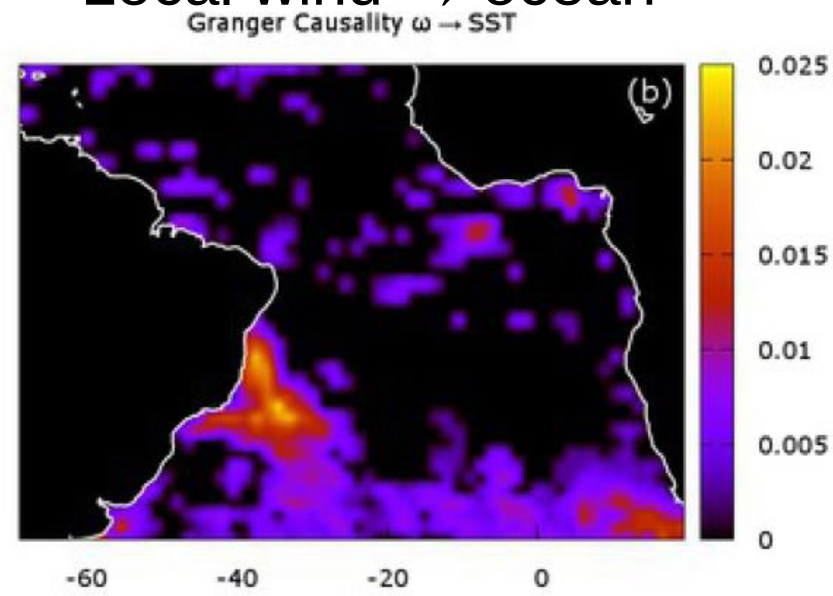
Application to climate data: rain-ocean interaction in the South Atlantic Convergence Zone

Data: two time series at the same geographical location.
SST = Surface sea temperature
ω = vertical wind velocity at 500 hPa (precipitation proxy)

Local ocean → wind



Local wind → ocean



The color code represents GCE values (only values significant at 99% confidence)

Ocean forces the atmosphere in the tropics and the subtropical waters of Brazil.

The atmosphere also forces a localized region of the ocean in front of Brazil.

G. Tirabassi et al, Int. J. of Climatology, 35, 3440 (2015)

How to find “synchronized events” in two time series?

Measures of event synchronization

- Define "events" in each time series. m_x, m_y are the number of events in each time series.
- Count $c^\tau(x|y)$ = number of times an event appears in x shortly after an event appears in y. Analogous for $c^\tau(y|x)$.
- Measures:

$$Q_\tau = \frac{c^\tau(y|x) + c^\tau(x|y)}{\sqrt{m_x m_y}} \qquad q_\tau = \frac{c^\tau(y|x) - c^\tau(x|y)}{\sqrt{m_x m_y}}$$

- Q_τ = 1 if and only if the events of the signals are fully synchronized.
- q_τ =1 if the events in x always precede those in y.
- Many applications. Further reading: Quian Quiroga et al, PRE 66, 041904 (2002).

Take home message

- Cross-correlation: detects linear interdependencies.
- Mutual information: can detect nonlinear interdependencies.
- The MI computed from the probabilities of ordinal patterns allows to select the time-scale of the analysis.
- The directionality index detects the net direction of the information flow.
- Granger causality can “disentangle” mutual interactions.

References

- M. Palus, Contemporary Physics 48, 307(2007)

- M. Barreiro, et. al, Chaos 21, 013101 (2011)
- Deza, Barreiro and Masoller, Eur. Phys. J. ST 222, 511 (2013)
- Tirabassi and Masoller, EPL 102, 59003 (2013)
- Deza, Barreiro and Masoller, Chaos 25, 033105 (2015)
- Tirabassi, Masoller and Barreiro, Int. J. of Climatology, 35, 3440 (2015)

<cristina.masoller@upc.edu>

http://www.fisica.edu.uy/~cris/