Symbolic and Network-based Analysis Tools

Cristina Masoller

Cristina.masoller@upc.edu

www.fisica.edu.uy/~cris

UNIVERSITAT POLITÈCNICA DE CATALUNYA
BARCELONATECH

Campus d'Excel·lència Internacional

First CAFE School
Sitges 18/11/2019

CAFE
Climate Advanced Forecasting of sub-seasonal Extremes

Introducing myself

- Originally from Montevideo, Uruguay
- PhD in physics (lasers, Bryn Mawr College, USA 1999)
- Since 2004 @ Universitat Politecnica de Catalunya
- Profesora Catedratica, Physics Department, research group on Dynamics, Nonlinear Optics and Lasers
- Web page: http://www.fisica.edu.uy/~cris/

Introducing our research group Dynamics, Nonlinear Optics and Lasers

Senior researchers / PhD students: 11/8

Where are we?

UNIVERSITAT POLITÈCNICA DE CATALUNYA
BARCELONATECH

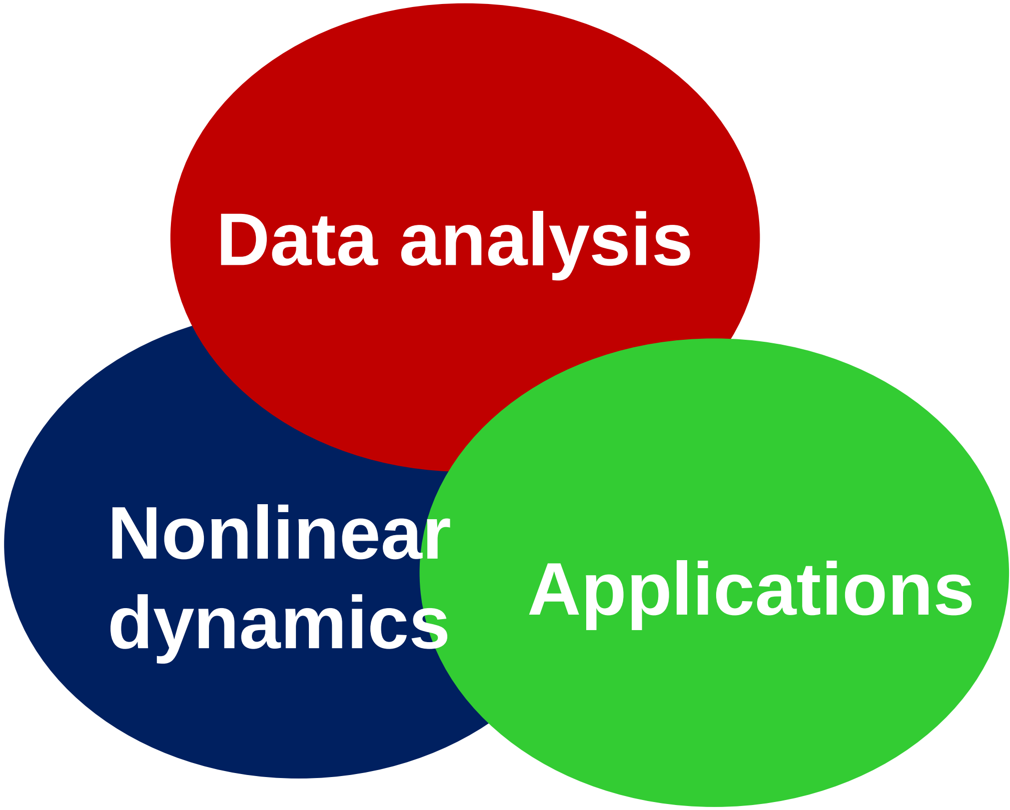
1. Barcelona
2. Castelldefels
3. Igualada
4. Manresa
5. Mataró
6. Sant Cugat del Vallès
7. Terrassa
8. Vilanova i la Geltrú

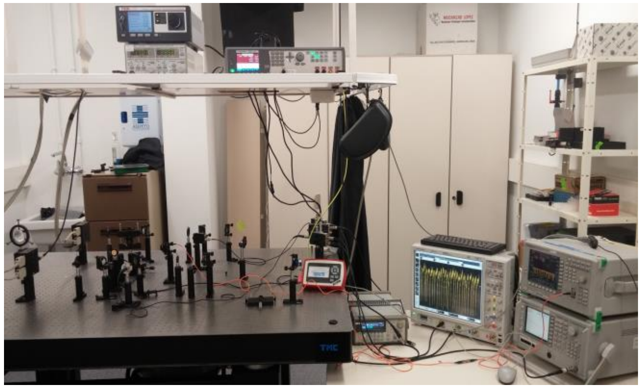
Viernes, 25 de septiembre de 2009 Diari de Terrassa

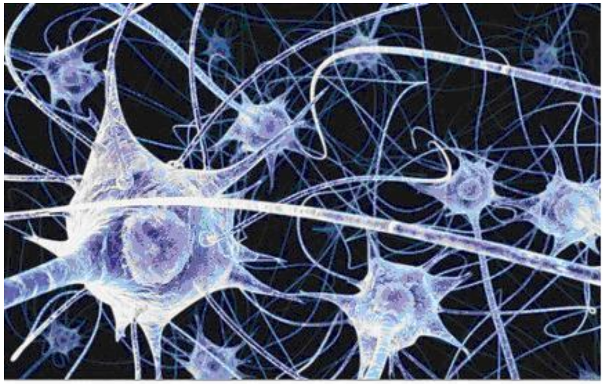
El edificio Gaia centraliza grupos científicos consolidados y emergentes.

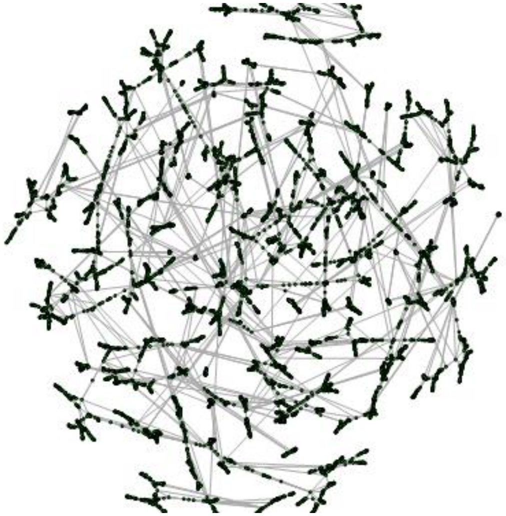
What do we study?

- Nonlinear and stochastic phenomena
  - laser dynamics
  - neuronal dynamics
  - complex networks
  - data analysis (climate, biomedical signals)





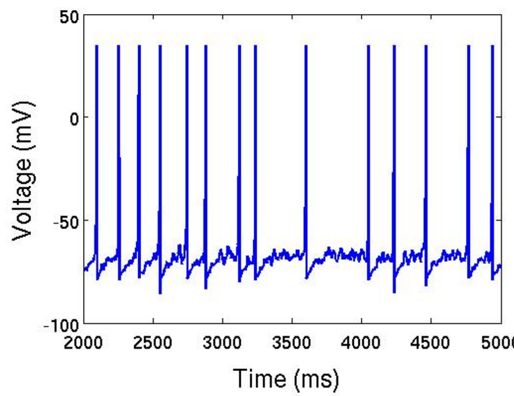




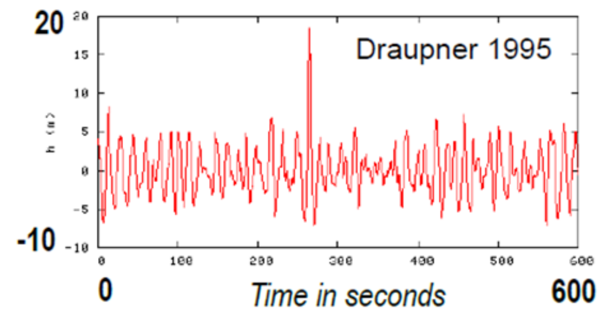
Lasers, neurons, climate, complex systems?

- Lasers allow us to study in a controlled way phenomena that occur in diverse complex systems.
- Laser experiments allow to generate sufficient data to test new methods of data analysis for prediction, classification, etc.

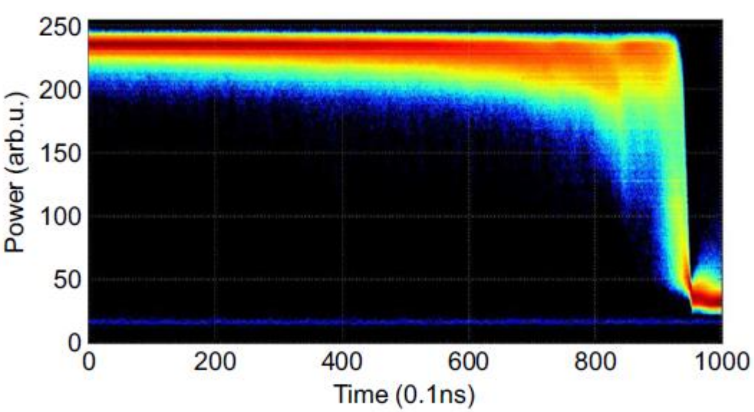
Laser & neuronal spikes



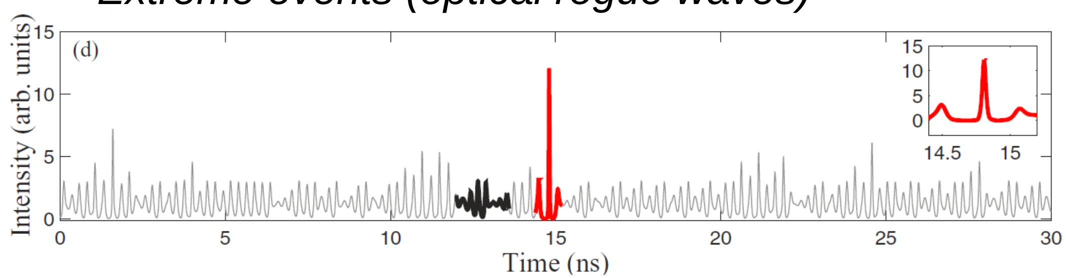
Ocean rogue wave (sea surface elevation in meters)



Abrupt switching

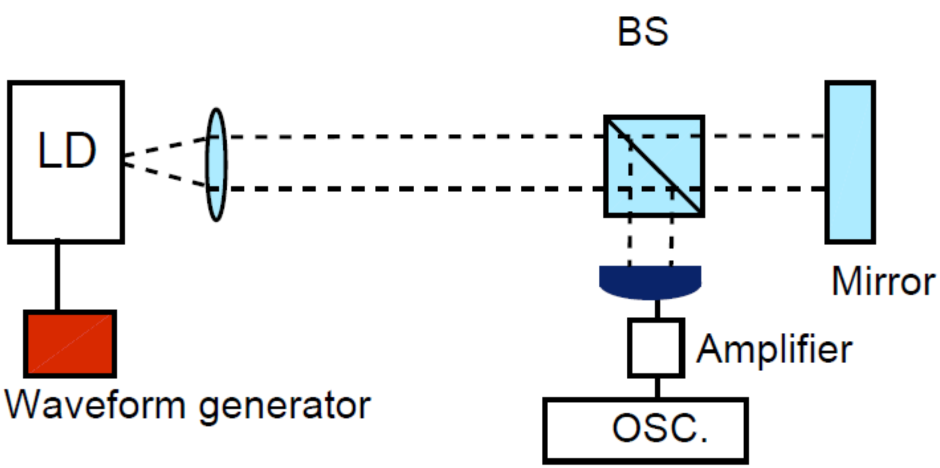


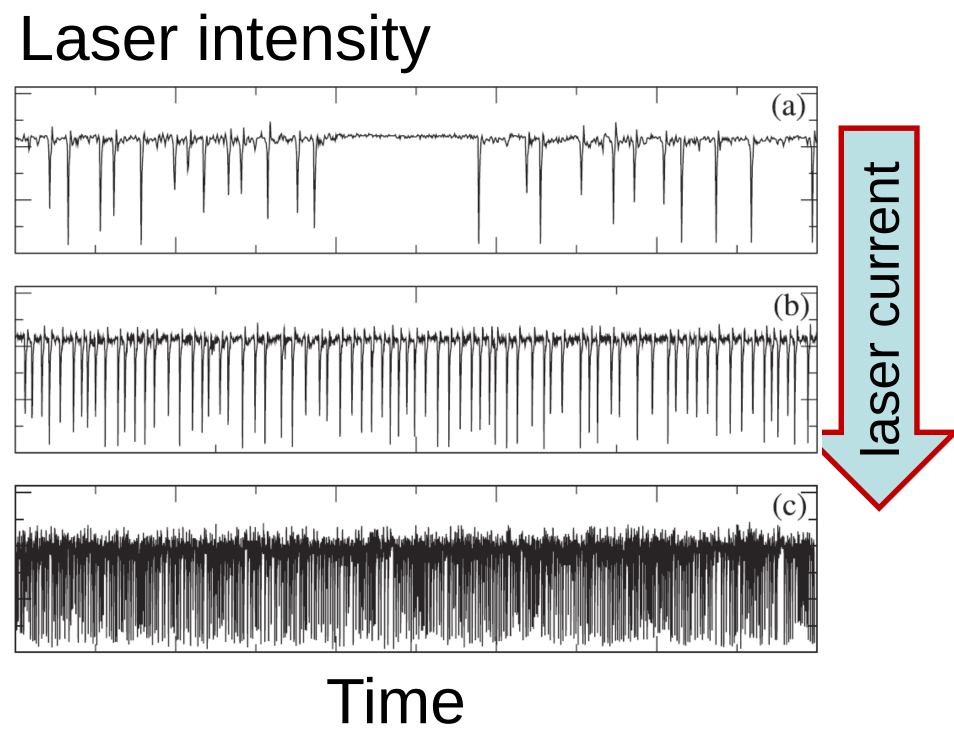
Extreme events (optical rogue waves)

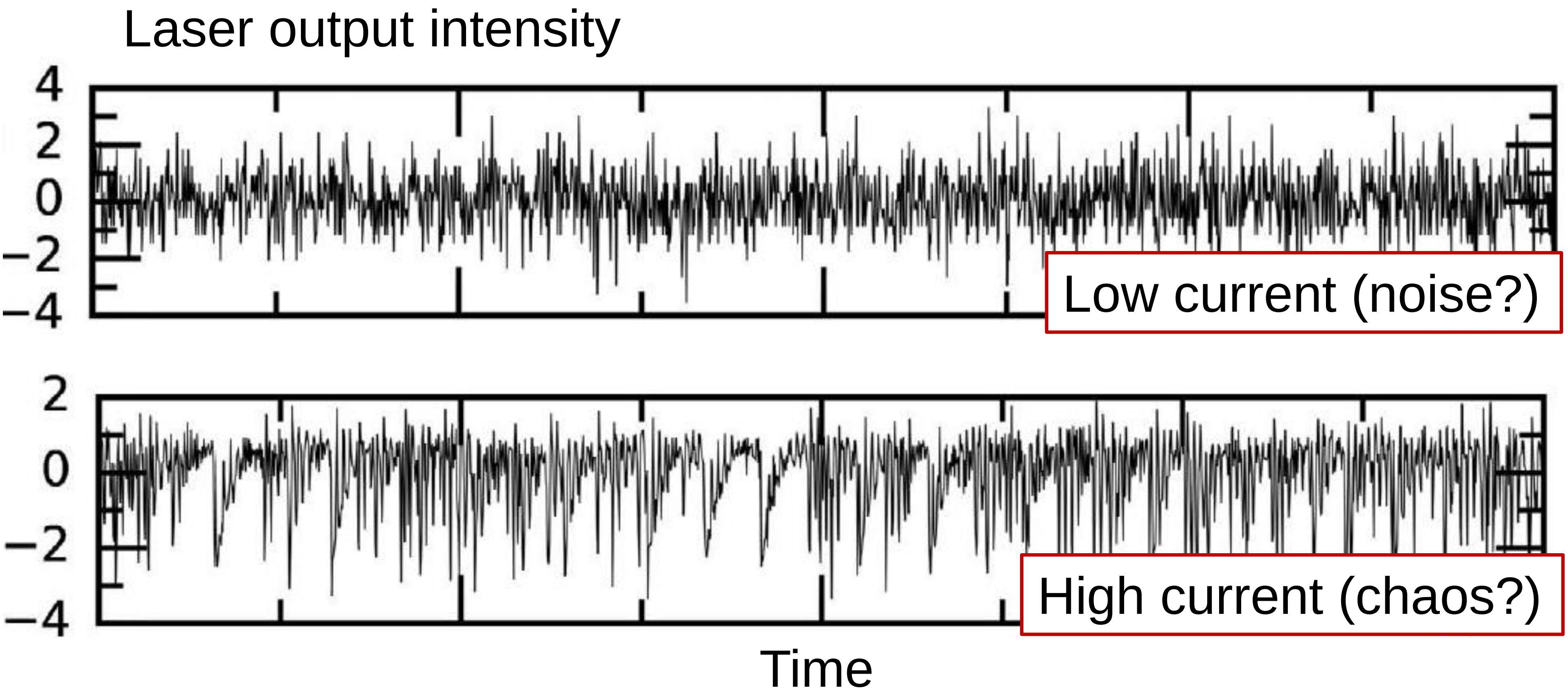


In complex systems dynamical transitions are difficult to identify and to characterize.

Example: the noise-chaos transition in a diode laser







Can differences be quantified? With what reliability?

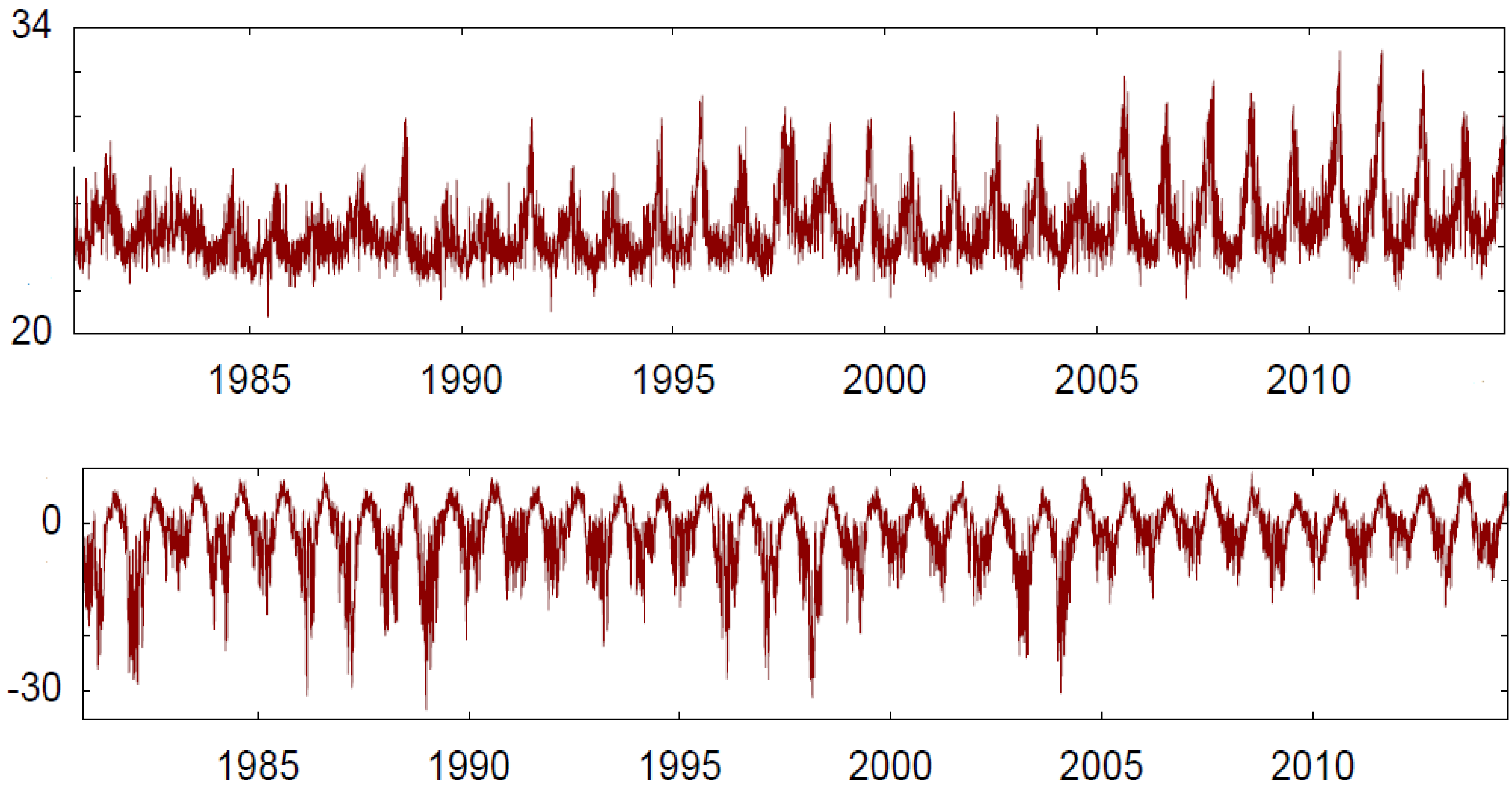
Are weather extremes becoming more frequent?
more extreme?

Physics Today, Sep. 2017

ECMWF

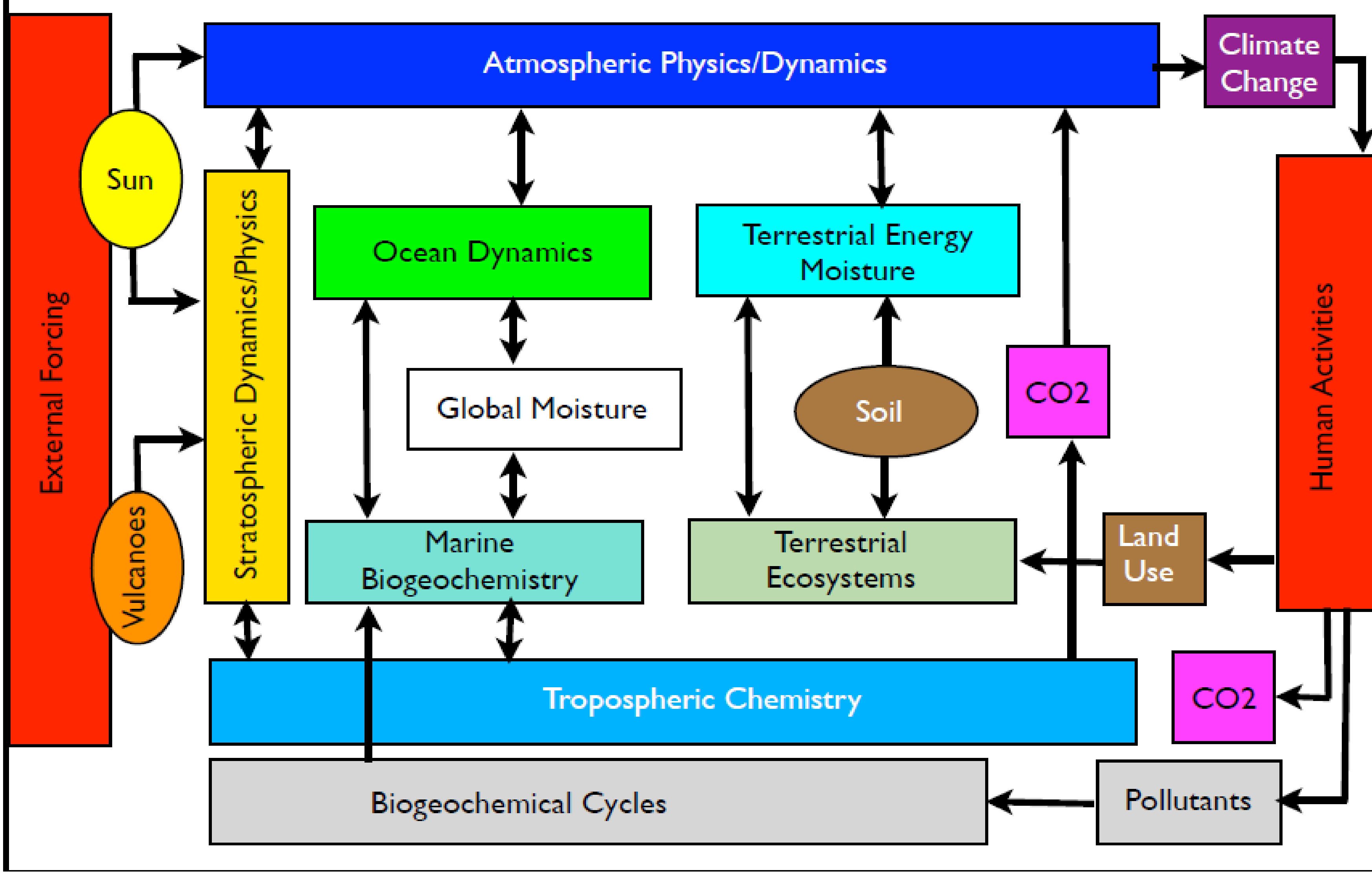
Credit: Richard Williams, North Wales, UK

Surface air temperature in two different regions



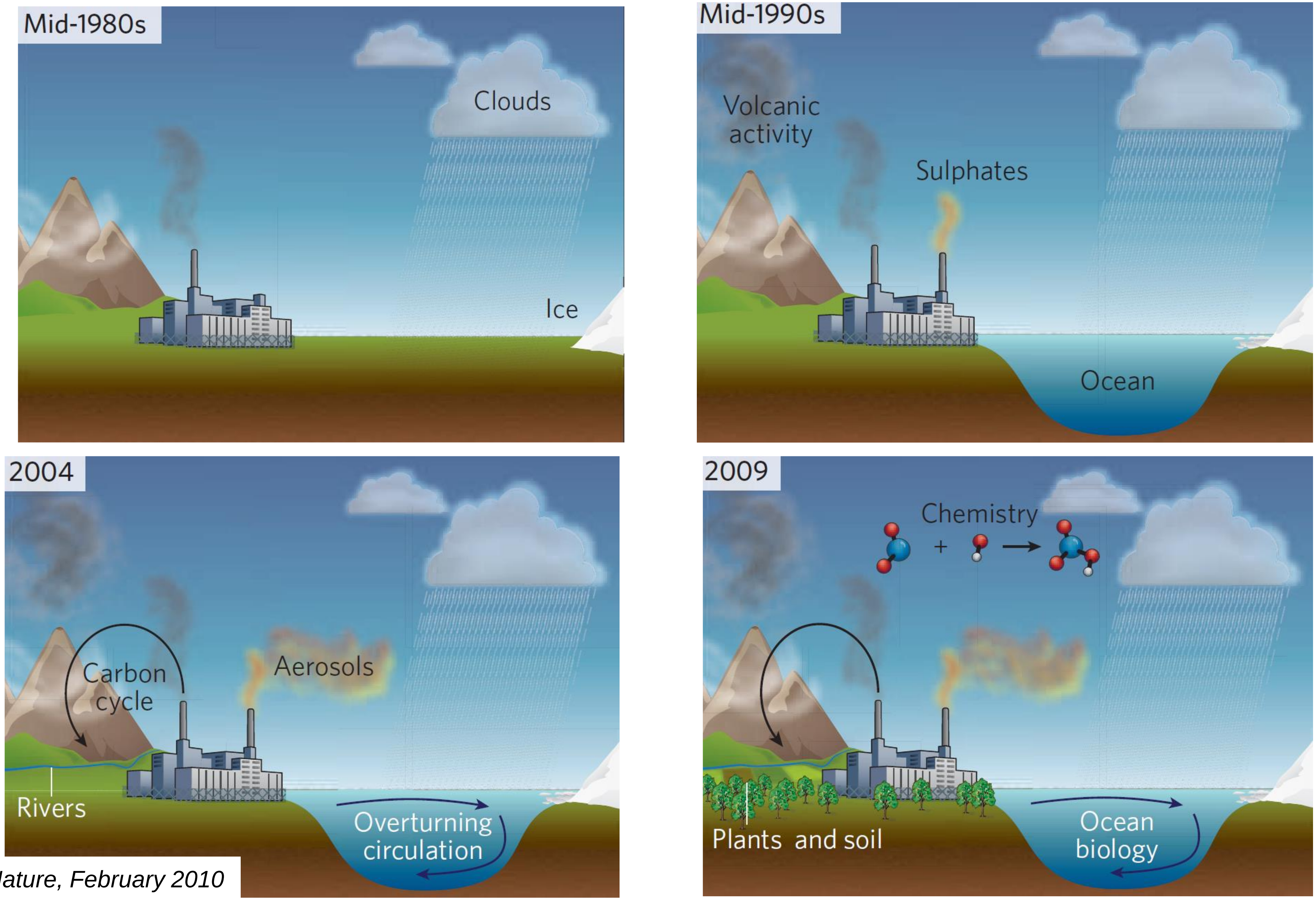
Can gradual changes be quantified? With what reliability?

The Climate System is a “complex system”



Courtesy of Henk Dijkstra (Ultrech University)

Thanks to advances in computer science, climate models now allow for good (?) weather forecasts



Nature, February 2010

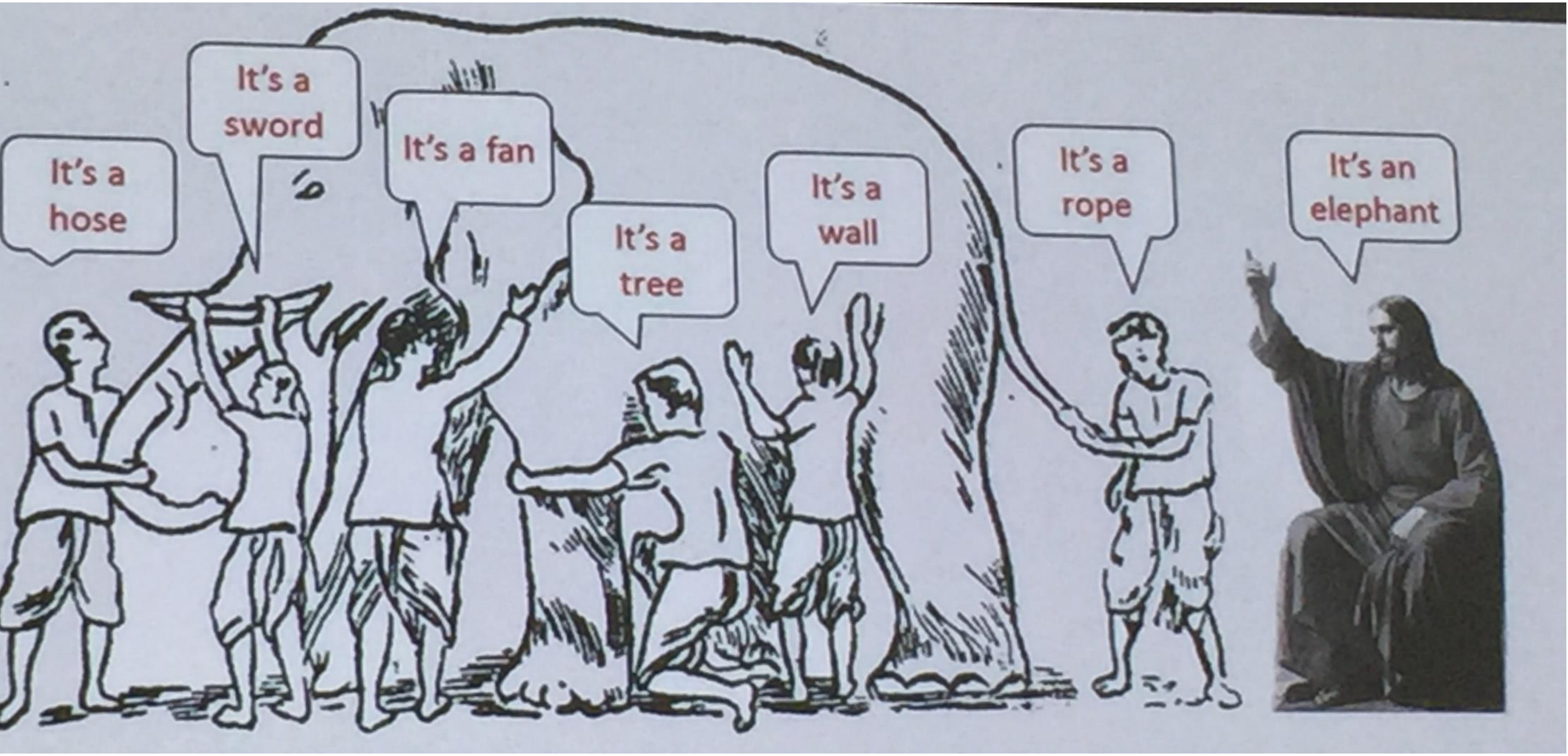
But detailed models are not very useful for improving our understanding

On the other hand, “over-simplified models” do not provide very useful information (e.g. the zero dimensional energy balance model that considers the whole Earth as a point mass).

In summer, 1996, milk production at a Wisconsin dairy farm was very low.
The farmer wrote to the state university, asking help from academia.
A multidisciplinary team of professors was assembled, headed by a theoretical physicist, and two weeks of intensive on-site investigation took place.
A few weeks later, a physicist phoned the farmer, "I've got the answer," he said, "But it only works when you consider spherical cows in a vacuum…”
Source: https://mirror.uncyc.org/wiki/Spherical_Cows

Strong need of nonlinear data-driven methods

Why nonlinear ?



Because the whole is not always equal to the sum of the parts

Methods

- Many methods have been developed to extract information from a time series (x_1, x_2, … x_N).
- The method to be used depends on the characteristics of the data
  - Length of the time series;
  - Stationarity;
  - Level of noise;
  - Temporal resolution;
  - etc.

- **Different methods provide complementary information.**

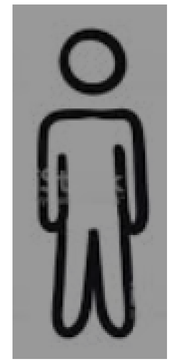
Where the data comes from?

- Modeling assumptions about the **type of dynamical system** that generates the data:
  - Stochastic or deterministic?
  - Regular or chaotic or “complex”?
  - Stationary or non-stationary? Time-varying parameters?
  - Low or high dimensional?
  - Spatial variable? Hidden variables?
  - Time delays? Etc.

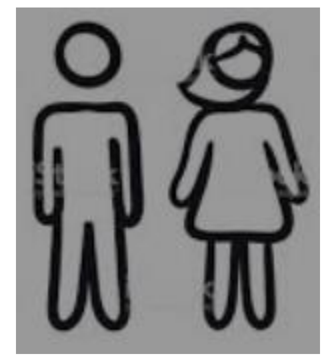
Outline

- **Introduction**
  - Historical developments: from dynamical systems to complex systems

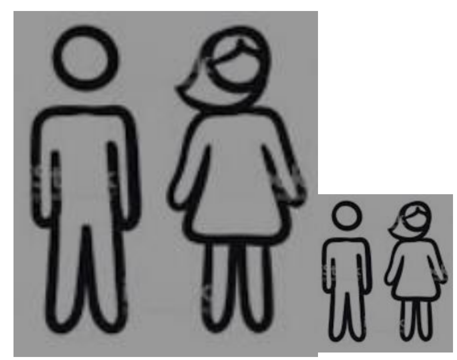
- **Univariate analysis**
  - Symbolic & network-based tools.
  - Applications.



- **Bivariate analysis**
  - Correlation, mutual information and directionality.
  - Applications.



- **Multivariate analysis**
  - Complex networks.
  - Network characterization and analysis.
  - Climate networks.



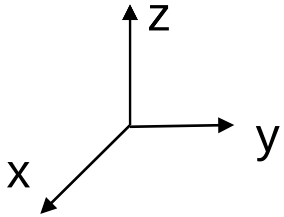
Late 1800s

- **Henri Poincare** (French mathematician).

  Instead of asking "*which are the exact positions of planets (trajectories)*?"

  he asked: "*is the solar system **stable** for ever, or will planets eventually run away*?"

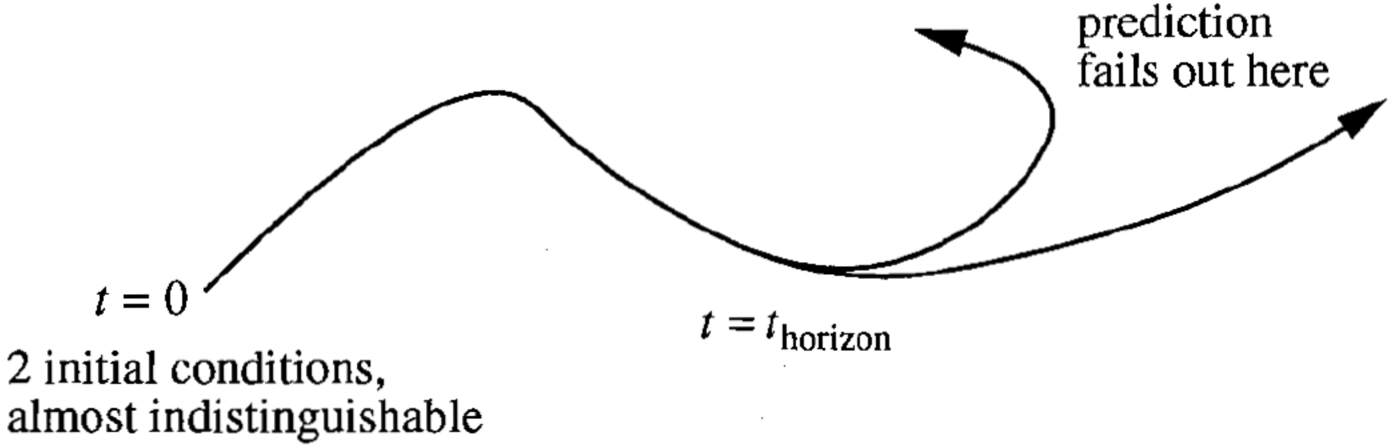
- He developed a **geometrical** approach to solve the problem.



- Introduced the concept of "phase space".

- He also had an intuition of the possibility of **chaos.**

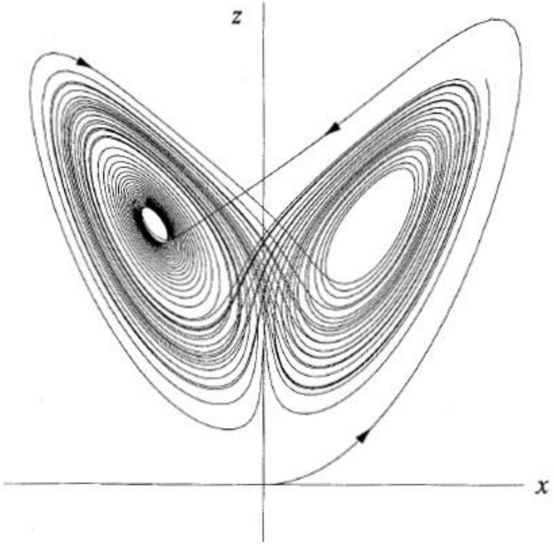
*Poincare: "The evolution of a **<u>deterministic</u>** system can be aperiodic, unpredictable, and strongly depends on the initial conditions"*



Deterministic system: the initial conditions fully determine the future state. **There is no randomness but the system can be unpredictable**.

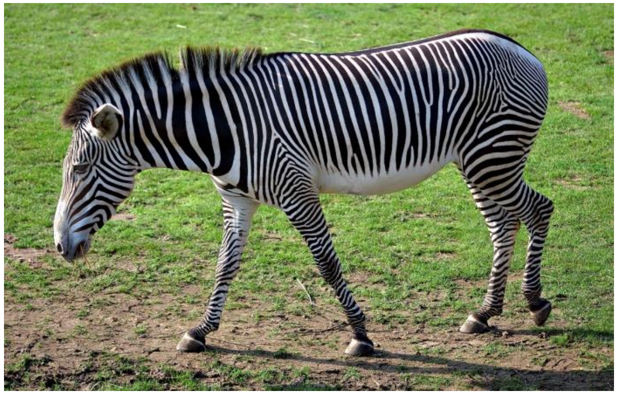
1950s: First computer simulations

- Computes allowed to experiment with equations.
- Huge advance in the field of "*Dynamical Systems*".
- 1960s: **Eduard Lorentz** (American mathematician and meteorologist at MIT): simple model of convection rolls in the atmosphere.
- **Chaotic** motion.



Order within chaos and self-organization

- **Ilya Prigogine** (Belgium, born in Moscow, Nobel Prize in Chemistry 1977)
- Thermodynamic systems far from equilibrium.
- Discovered that, in chemical systems, the interplay of (external) **input of energy** and **dissipation** can lead to “self-organized” patterns.



The study of spatio-temporal structures has uncovered striking similarities in nature

Honey bees do a spire wave to scare away predators

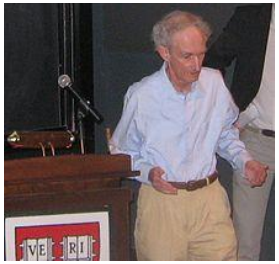
https://www.youtube.com/watch?v=Sp8tLPDMUyg

Rotating waves occur in the heart during ventricular fibrillation

Hurricane Maria (Wikipedia)

https://media.nature.com/original/nature-assets/nature/journal/v555/n7698/extref/nature26001-sv6.mov

The 1970s



- **Robert May** (Australian, 1936): population biology
- "*Simple mathematical models with very complicated dynamics*", *Nature* (1976).

$$x_{t+1} = f(x_t)$$

Example: $f(x) = r\,x(1-x)$

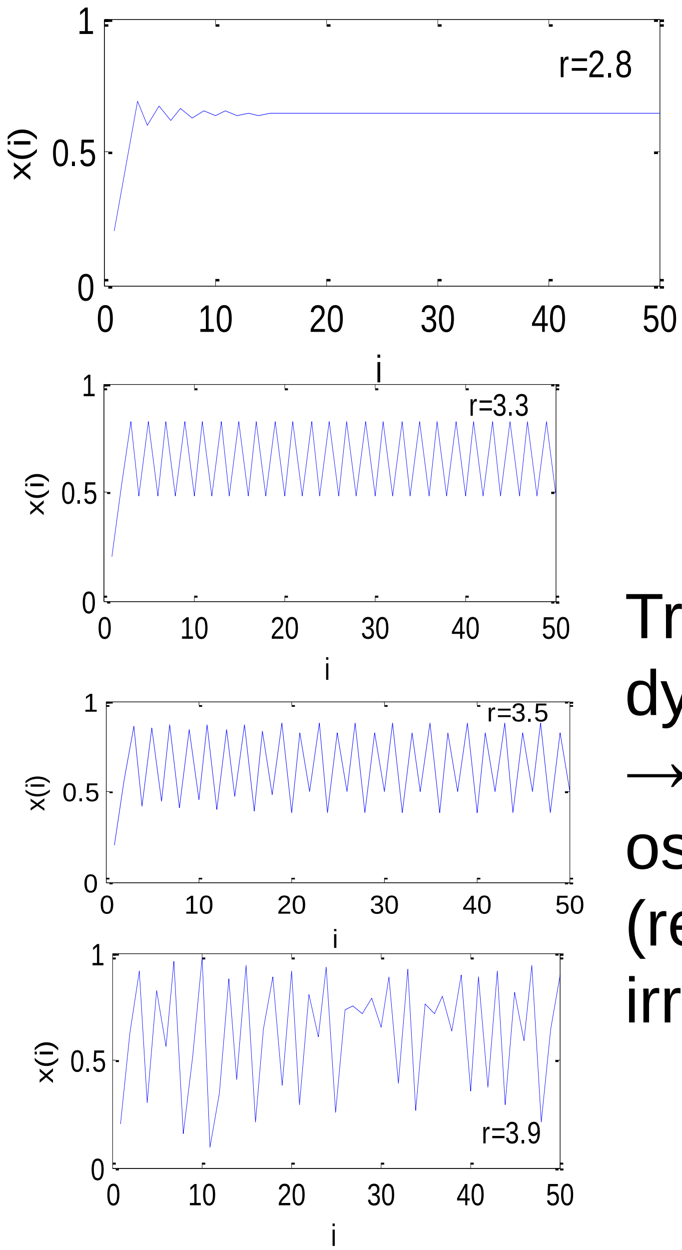
- Difference equations ("iterated maps"), even though simple and deterministic, can exhibit different types of dynamical behaviors, from **stable points**, to a bifurcating hierarchy of **stable cycles**, to **apparently random fluctuations**.

The logistic map

$$x(i+1) = r\,x(i)[1 - x(i)]$$

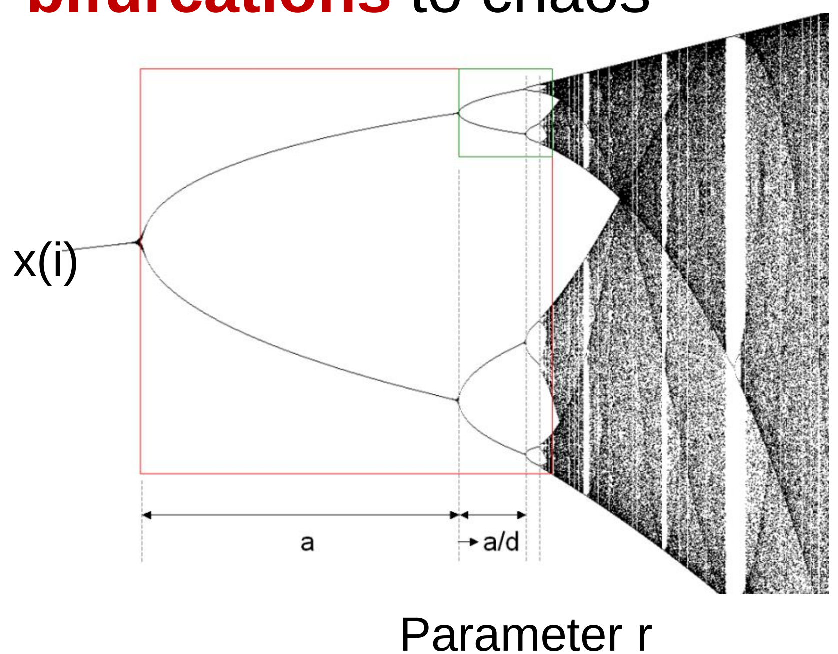
r=2.8, Initial condition: x(1) = 0.2

Transient relaxation → long-term stability



Transient dynamics → stationary oscillations (regular or irregular)

"period-doubling" **bifurcations** to chaos



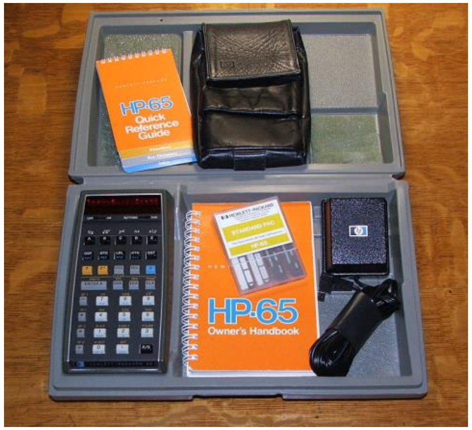
Universal route to chaos

- In 1975, **Mitchell Feigenbaum** (American mathematician and physicist 1944-2019), using a small HP-65 calculator, discovered the scaling law of the bifurcation points

$$\lim_{n\to\infty} \frac{r_{n-1} - r_{n-2}}{r_n - r_{n-1}} = 4.6692...$$

- Then, he showed that the same behavior, with the same mathematical constant, occurs within a wide class of functions, prior to the onset of chaos (**universality**).

Very different systems (in chemistry, biology, physics, etc.) go to chaos in the same way, quantitatively.



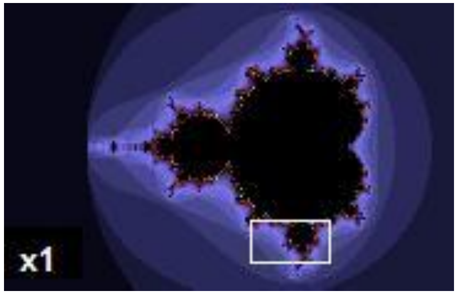
HP-65 calculator: the first magnetic card-programmable handheld calculator

The late 1970s

- **Benoit Mandelbrot** (Polish-born, French and American mathematician 1924-2010): “self-similarity” and **fractal objects:**

  each part of the object is like the whole object but smaller.

- Because of his access to IBM's computers, Mandelbrot was one of the first to use **computer graphics** to create and display fractal geometric images.



Fractal objects

- Are characterized by a “fractal” dimension that measures roughness.

Broccoli
D=2.66

Human lung
D=2.97

Coastline of Ireland
D=1.22

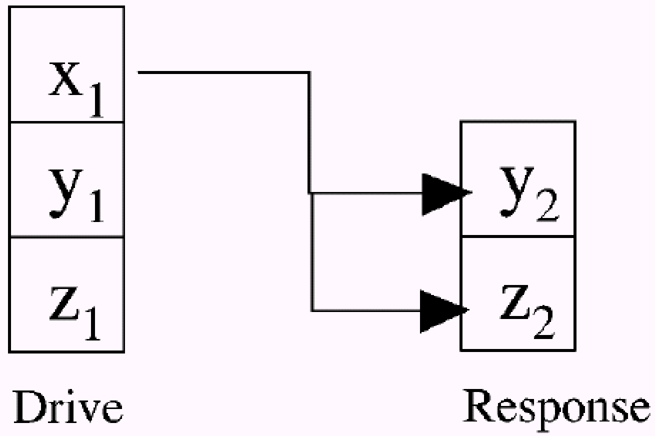
A lot of research is focused on detecting fractal objects underlying real-world signals.

Video: http://www.ted.com/talks/benoit_mandelbrot_fractals_the_art_of_roughness#t-149180

The 1990s: synchronization of chaotic systems

Pecora and Carroll, PRL 1990

Unidirectional coupling of two chaotic systems: one variable, '*x*', of the response system is **replaced** by the same variable of the drive system.



$$t \to \infty \quad |y_2 - y_1| \to 0, \; |z_2 - z_1| \to 0$$

First observation of synchronization: mutual *entrainment* of pendulum clocks

In mid-1600s **Christiaan Huygens** (Dutch mathematician) noticed that two pendulum clocks mounted on a common board synchronized with their pendulums swinging in opposite directions (in-phase also possible).

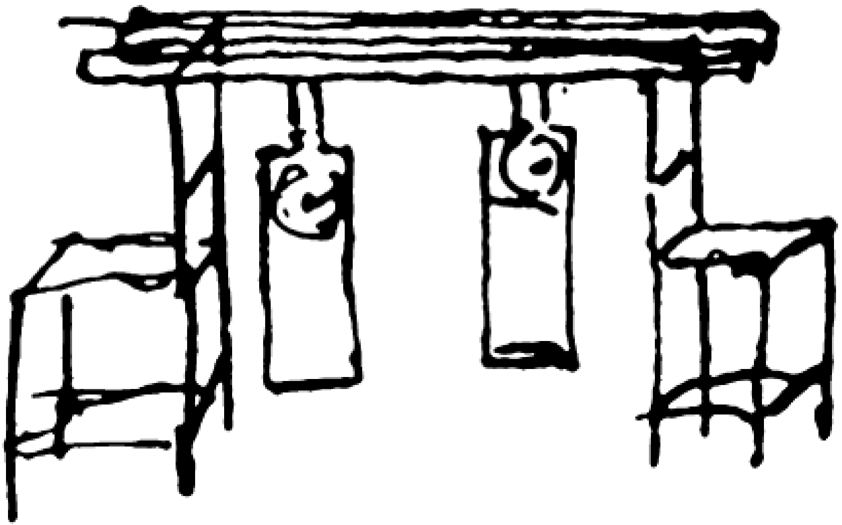
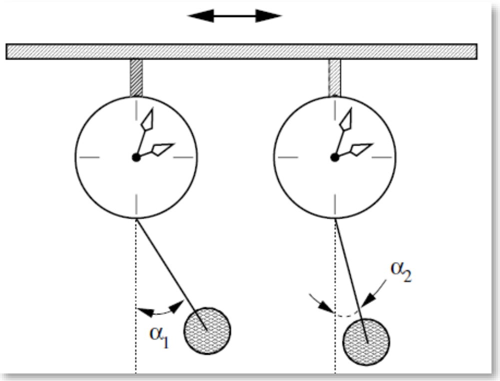


Figure 1.2. Original drawing of Christiaan Huygens illustrating his experiments with two pendulum clocks placed on a common support.



http://www.youtube.com/watch?v=izy4a5erom8

Different types of synchronization

$$dx_1 / dt = F(x_1)$$

$$dx_2 / dt = F(x_2) + \alpha E(x_1 - x_2)$$

- Complete: $x_1(t) = x_2(t)$ (identical systems)
- Phase: the phases of the oscillations synchronize, but the amplitudes are not.
- Lag: $x_1(t+\tau) = x_2(t)$
- Generalized: $x_2(t) = f(x_1(t))$ (*f* can depend on the strength of the coupling)

A lot of research is focused on detecting synchronization in real-world signals.

Synchronization of a large number of coupled oscillators

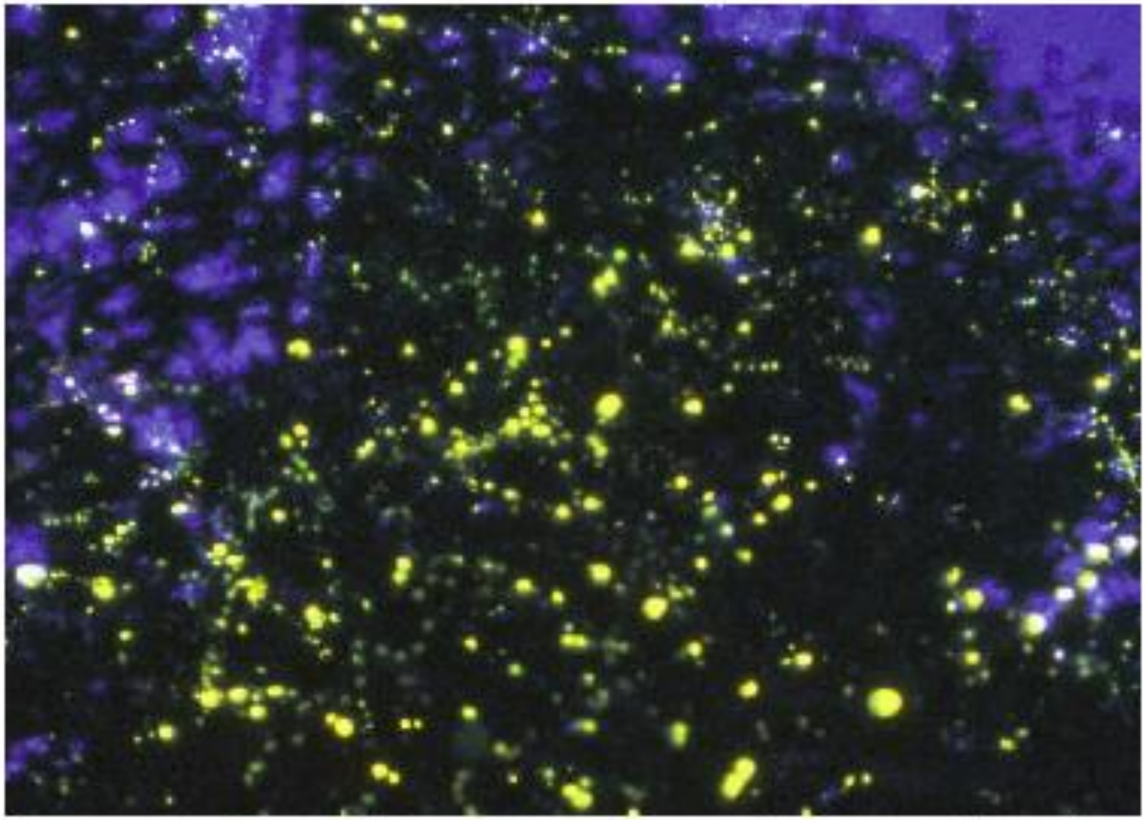
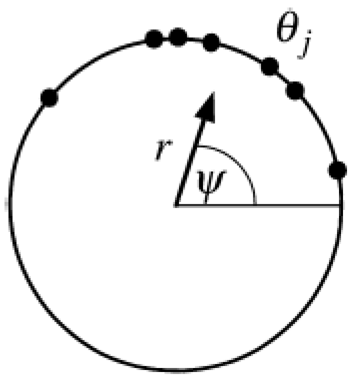


Figure 1 | Fireflies, fireflies burning bright. In the forests of the night, certain species of firefly flash in perfect synchrony — here *Pteroptyx malaccae* in a mangrove apple tree in Malaysia. Kaka *et al.*[2] and Mancoff *et al.*[3] show that the same principle can be applied to oscillators at the nanoscale.

Kuramoto model

(Japanese physicist, 1975)

Model of **all-to-all** coupled **phase oscillators**.



$$\frac{d\theta_i}{dt} = \omega_i + \frac{K}{N}\sum_{j=1}^{N}\sin(\theta_j - \theta_i) + \xi_i, \qquad i = 1 \ldots N$$

K = coupling strength, ξ_i = stochastic term (noise)

Describes the emergence of collective behavior

How to quantify?

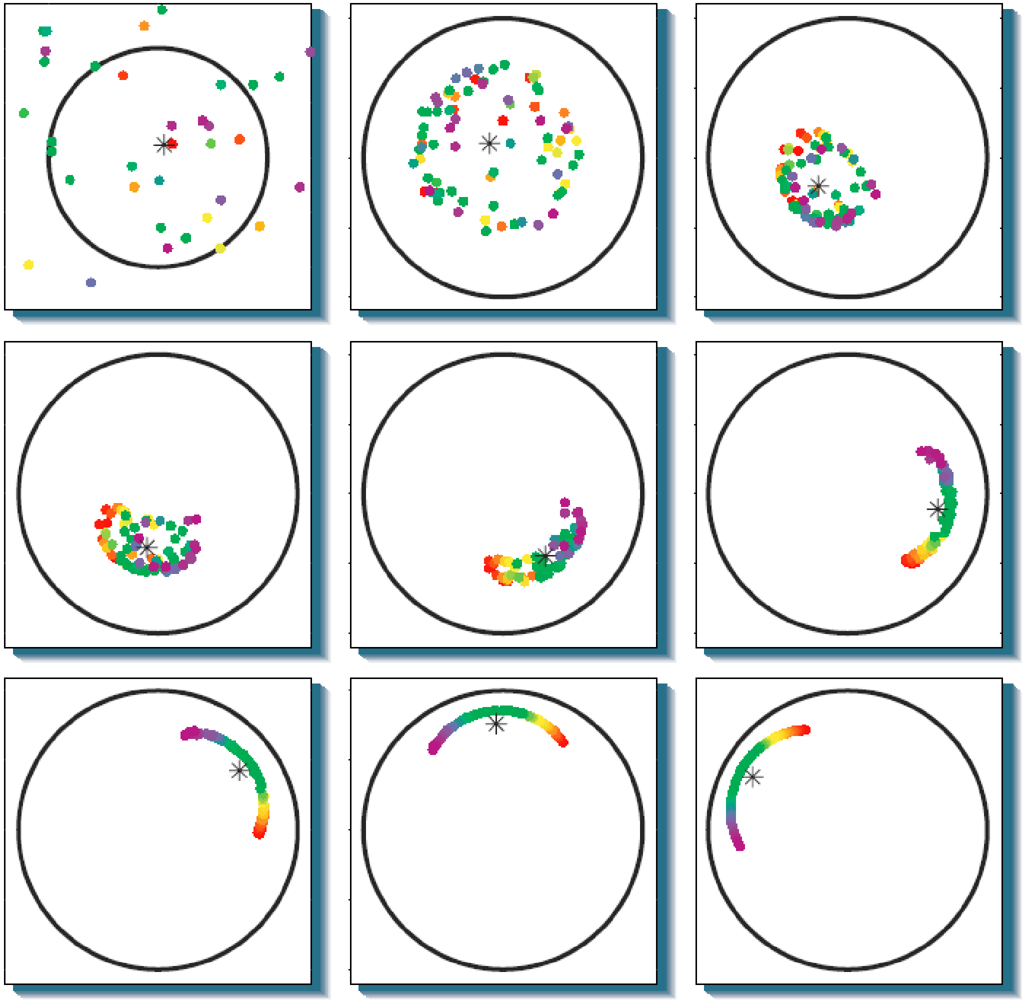
With the **order parameter**:

$$re^{i\psi} = \frac{1}{N}\sum_{j=1}^{N} e^{i\theta_j}$$

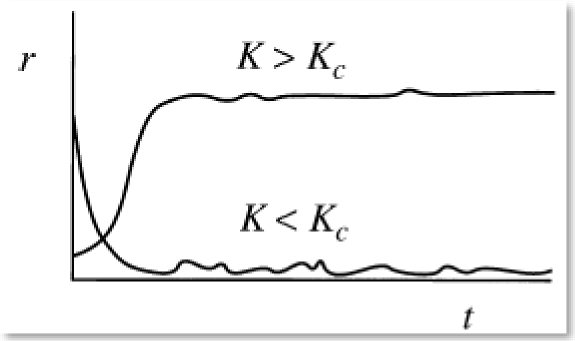
r =0 incoherent state (oscillators scattered in the unit circle)

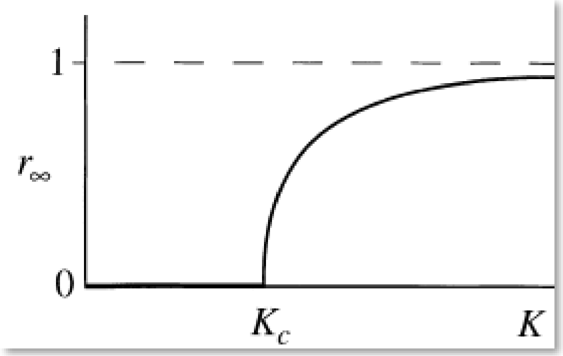
r =1 all oscillators are in phase ($\theta_i=\theta_j\ \forall$ i,j)

Synchronization transition as the coupling strength increases



Strogatz and others, late 90’





Strogatz, Nature 2001

Video: https://www.ted.com/talks/steven_strogatz_on_sync

End of 90's - present

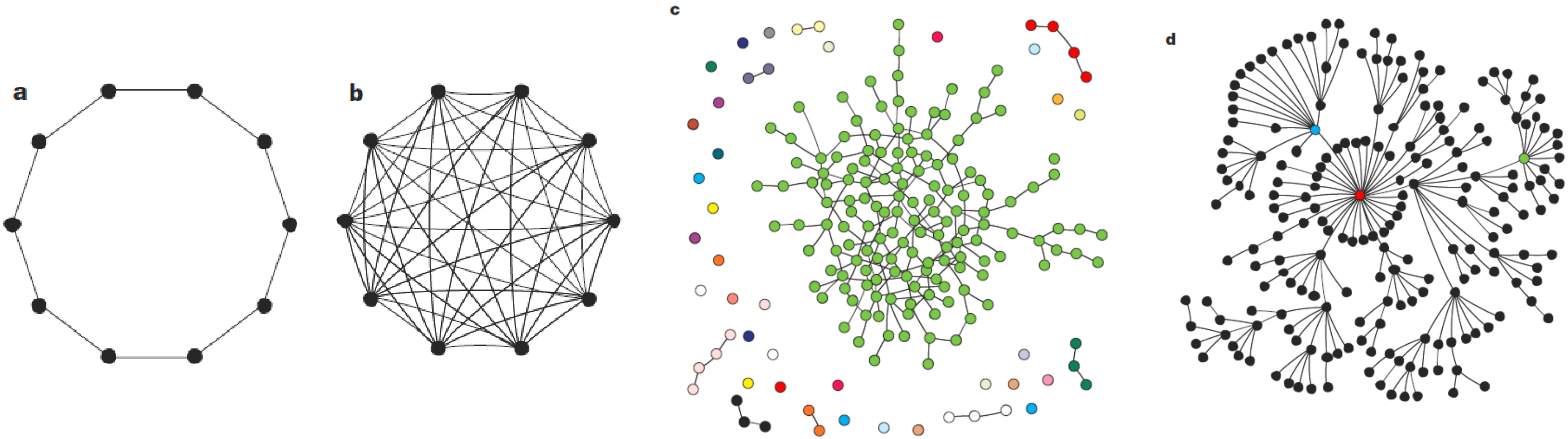
- Interest moves from <u>chaotic systems</u> to <u>complex systems</u> (small vs. very large number of variables).

- Networks (or graphs) of interconnected systems

- **Complexity science**: dynamics of emergent properties
  - Epidemics
  - Rumor spreading
  - Transport networks
  - Financial crises
  - Brain diseases
  - Etc.

Network science

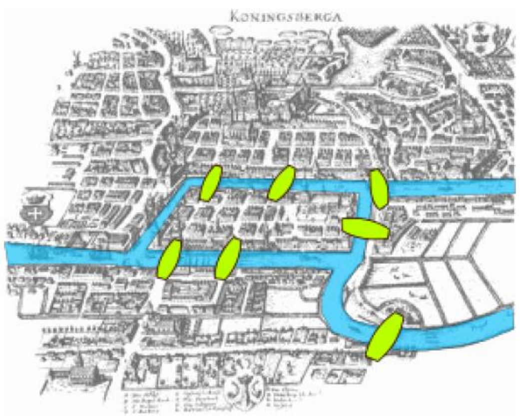
The challenge: to understand how the network **structure** and the **dynamics** (of individual units) determine the collective behavior.



Source: Strogatz
Nature 2001

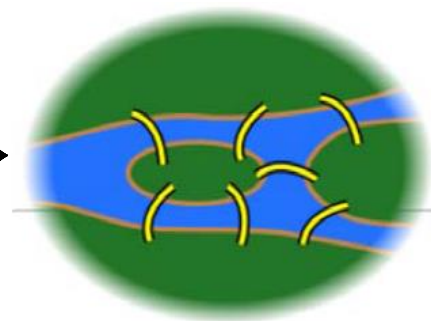
The start of Graph Theory:
The Seven Bridges of Königsberg (Prussia, now Russia)

- The problem was to devise a walk through the city that would cross each of those bridges once and only once.

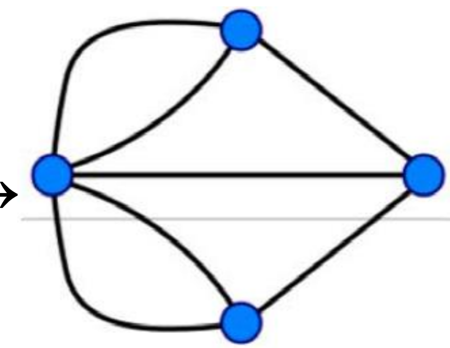


Source: Wikipedia

→



→



- By considering the number of odd/even links of each “node”, **Leonhard Euler** (Swiss mathematician) demonstrated in 1736 that is impossible.

Summary

- Dynamical systems allow to
  - understand low-dimensional systems,
  - uncover patterns and “order within chaos”,
  - characterize attractors, uncover universal features
- Synchronization: emergent behavior of interacting dynamical systems.
- Complexity and network science: emerging phenomena in large sets of interacting units.
- Time series analysis develops tools to characterize complex signals.
- Is an interdisciplinary research field with many applications.

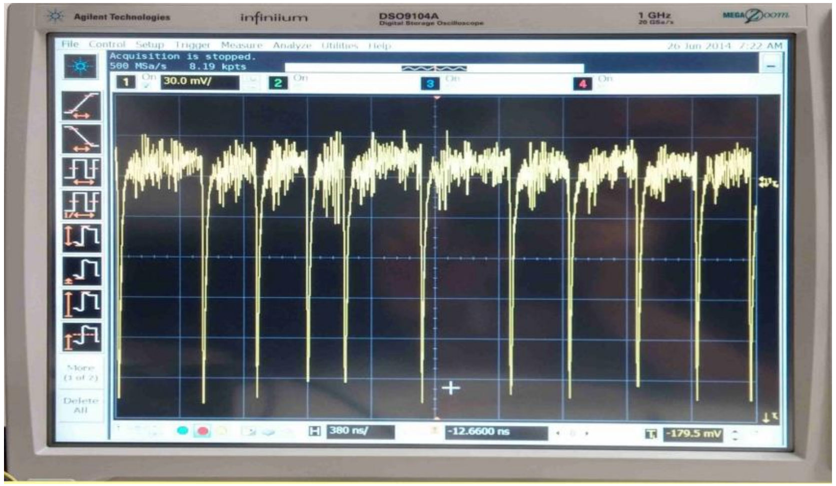
Outline

- **Introduction**
  - Historical developments: from dynamical systems to complex systems
- **Univariate analysis**
  - Symbolic & network-based tools.
  - Applications.
- **Bivariate analysis**
  - Correlation, mutual information and directionality.
  - Applications.
- **Multivariate analysis**
  - Complex networks.
  - Network characterization and analysis.
  - Climate networks.

Symbolic methods to identify patterns and structure in time series

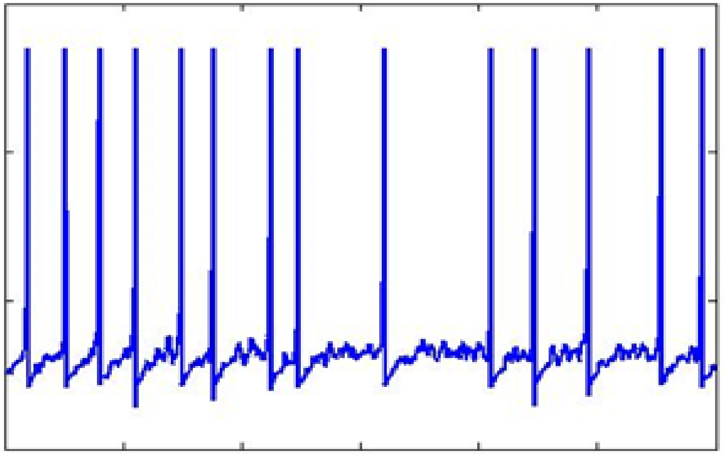
Can lasers mimic real neurons?

Laser spikes



Time (μs)

Neuronal spikes



Time (ms)

How to identify statistical similarities in the temporal sequence of “events”?

Symbolic analysis

- The time series $\{x_1, x_2, x_3, \ldots\}$ is transformed (using an appropriated **rule**) into a sequence of symbols $\{s_1, s_2, \ldots\}$
- taken from an "**alphabet**" of possible symbols.
- Then consider "blocks" of D symbols ("**patterns**" or "**words**").
- All the possible words form the "**dictionary**".

- Then analyze the "**language**" of the sequence of words
  - the probabilities of the words,
  - missing/forbidden words,
  - transition probabilities,
  - information measures (entropy, etc).

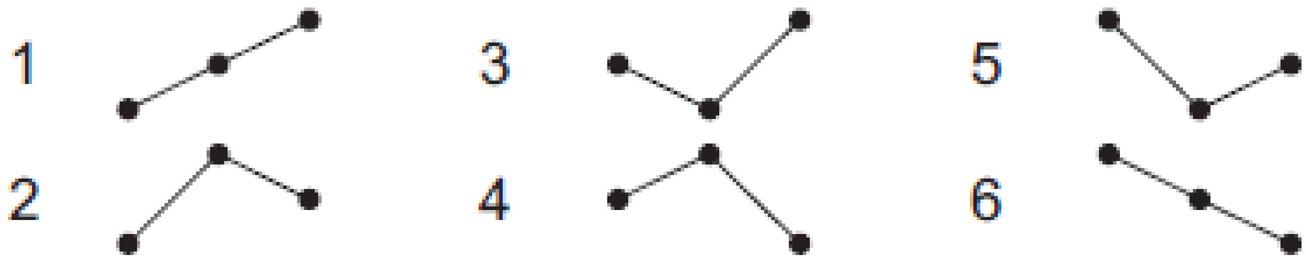
Threshold transformation: "partition" of the phase space

- if $x_i > x_{th} \Rightarrow s_i = 0$; else $s_i = 1$
  transforms a time series into a sequence of 0s and 1s, e.g., {011100001011111…}
- Considering "blocks" of D letters gives the sequence of words. Example, with D=3:
  {011   100   001   011   111 …}
- The number of words (patterns) grows as **2^D**
- More thresholds allow for more letters in the "alphabet" (and more words in the dictionary). Example:

$$\text{if } x_i > x_{th1} \Rightarrow s_i = 0;$$
$$\text{else if } x_i < x_{th2} \Rightarrow s_i = 2;$$
$$\text{else } (x_{th2} < x_i < x_{th1}) \Rightarrow s_i = 1.$$

Ordinal analysis: a method to find patterns in data

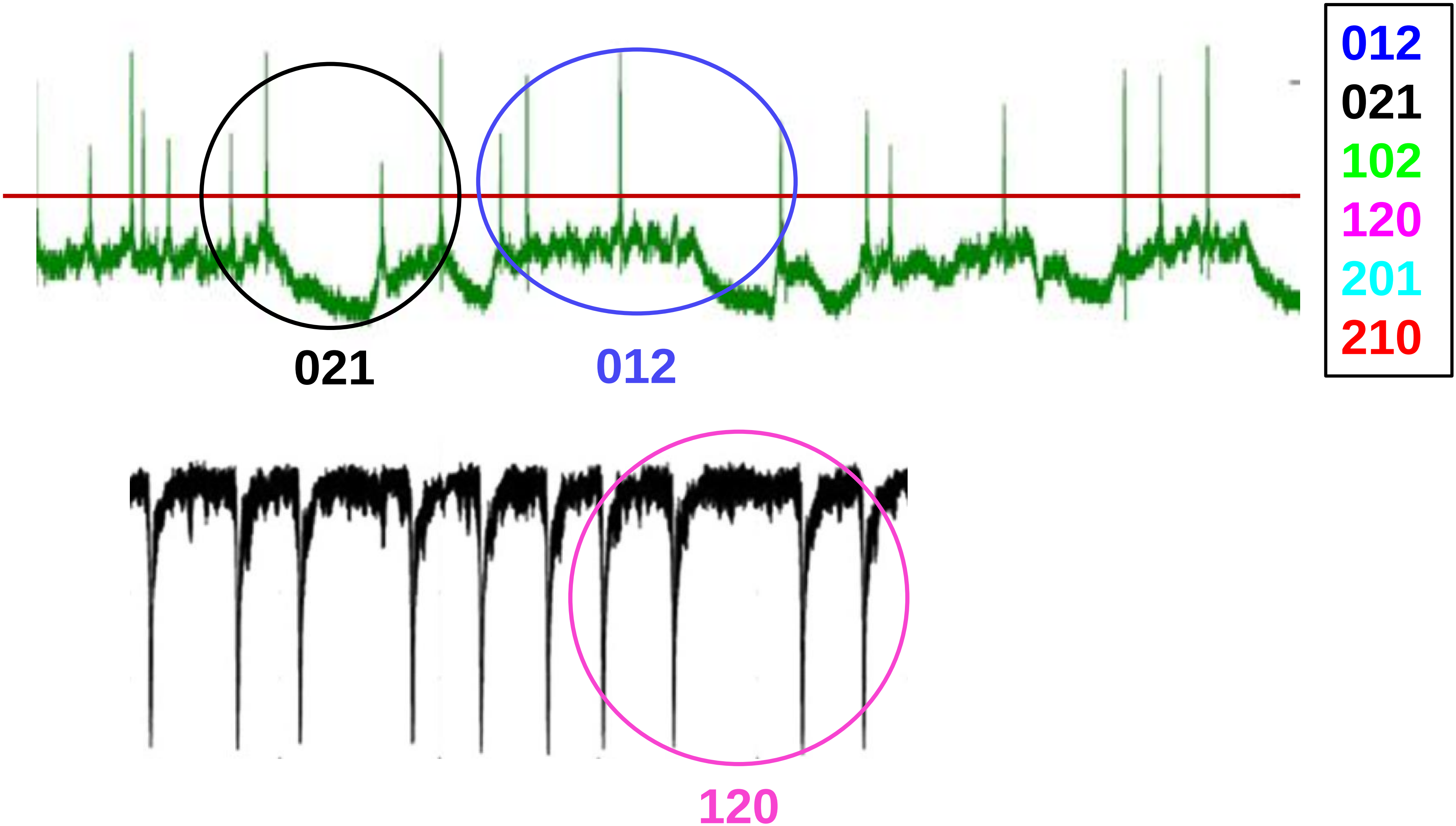
- Consider a time series $x(t)=\{\ldots x_i, x_{i+1}, x_{i+2}, \ldots\}$
- Which are the possible order relations among three data points?



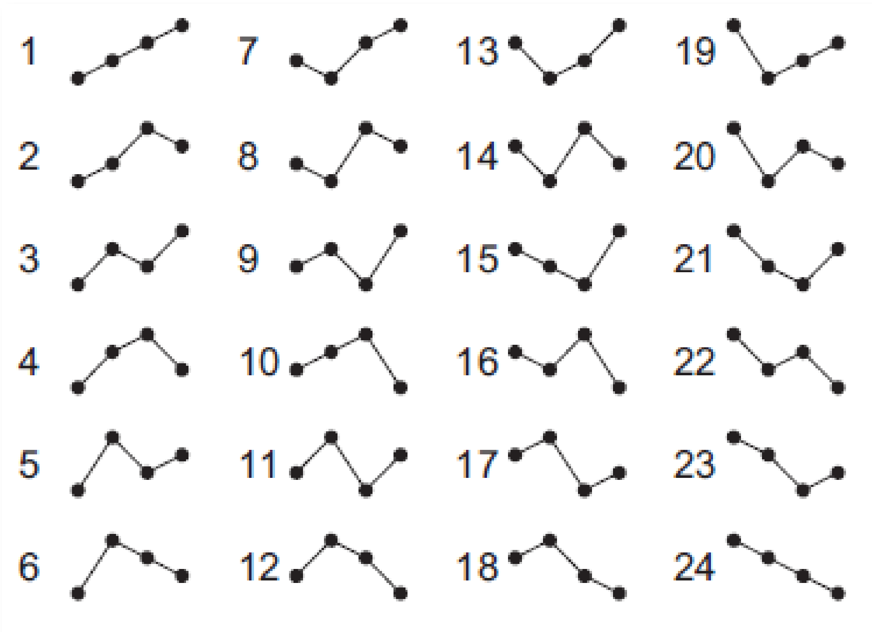
- Count how many times each "ordinal pattern" appears.
- Advantages: allows to identify temporal structures & is robust to noise.
- Drawback: information about actual data values is lost.

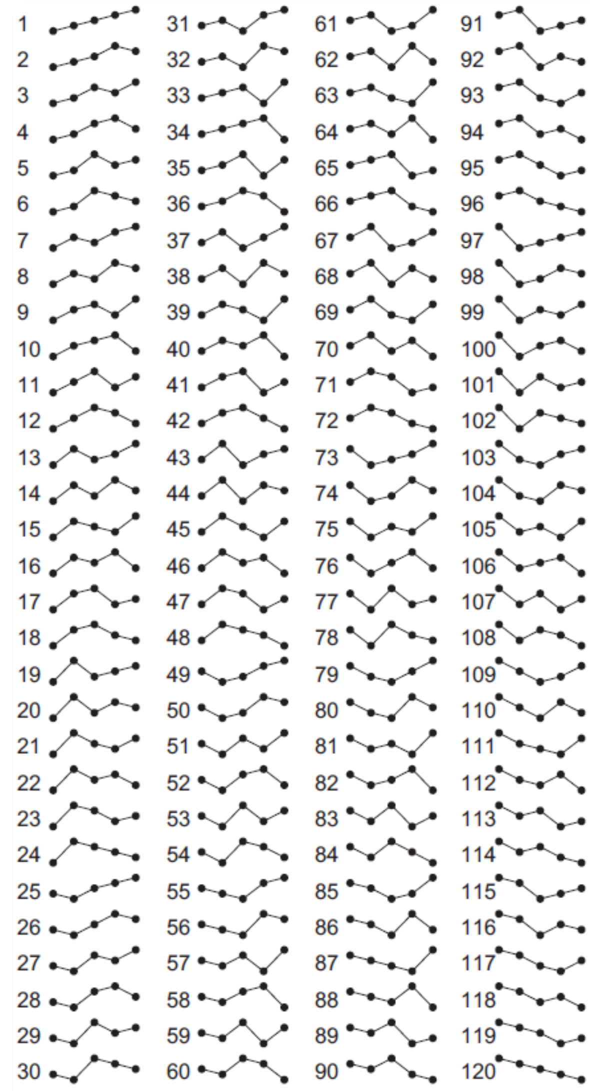
Bandt and Pompe PRL 88, 174102 (2002)

Analysis of D=3 patterns in spike sequences



The number of ordinal patterns increases as D!





- A problem for short datasets
- How to select optimal D?

  it depends on:
  - The length of the data
  - The length of the correlations

U. Parlitz et al. / Computers in Biology and Medicine 42 (2012) 319–327

Comparison

Threshold transformation:

if $x_i > x_{th} \Rightarrow s_i = 0$; else $s_i = 1$

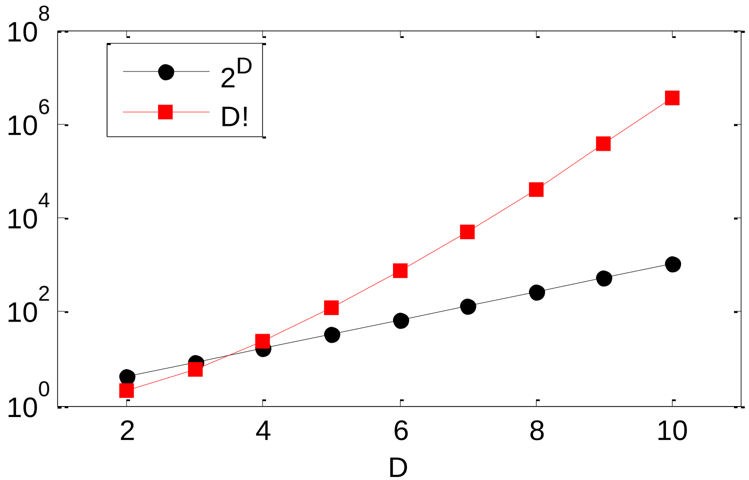
- Advantage: keeps information about the magnitude of the values.
- Drawback: how to select an adequate threshold ("partition" of the phase space).
- 2^D

Ordinal transformation:

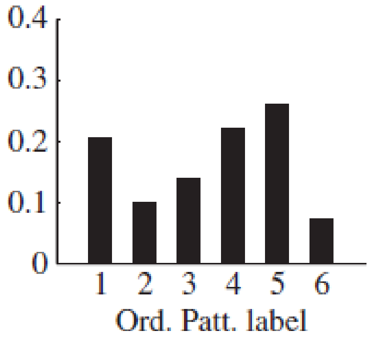
if $x_i > x_{i-1} \Rightarrow s_i = 0$; else $s_i = 1$

- Advantage: no need of threshold; keeps information about the temporal order in the sequence of values
- Drawback: no information about the actual data values
- D!

Number of symbols



Are the D! ordinal patterns equally probable?



- **Null hypothesis**:

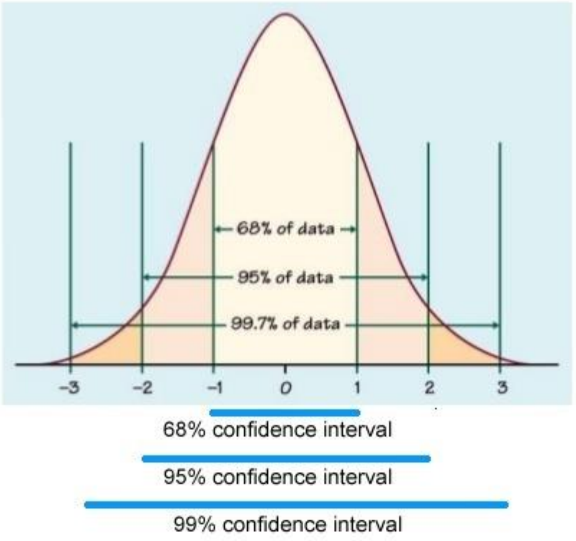
  $p_i = p = 1/D!$ for all $i = 1 \dots D!$

- If at least one probability **is not** in the interval $p \pm 3\sigma$ with $\sigma = \sqrt{p(1-p)/N}$ and N the number of ordinal patterns:

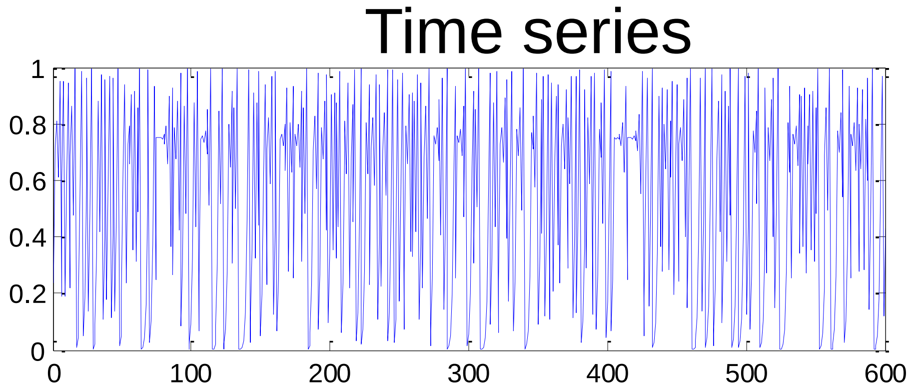
  We **reject** the NH with 99.74% confidence level.

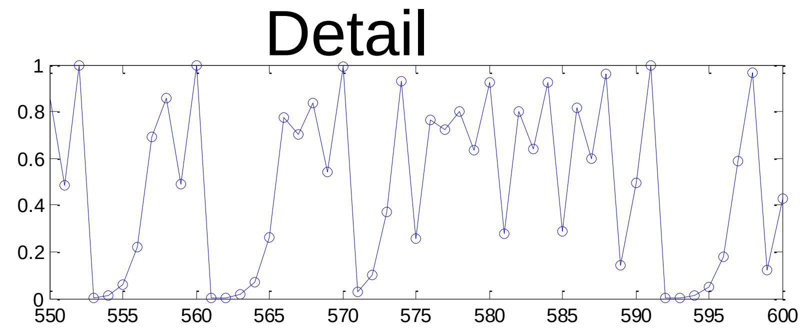
- Else

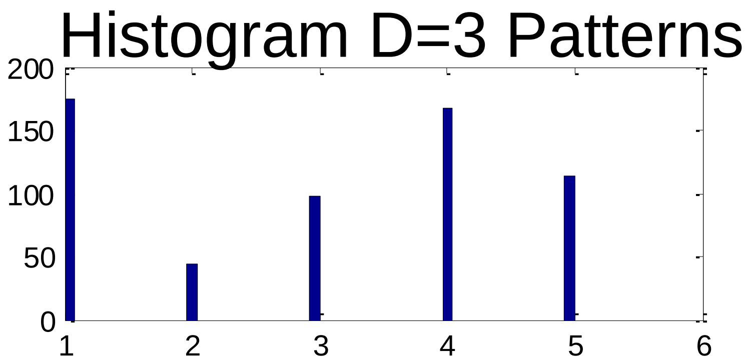
  We **fail to reject** the NH with 99.74% confidence level.

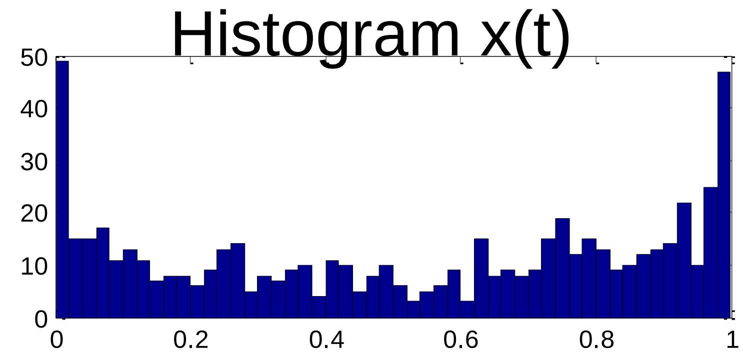


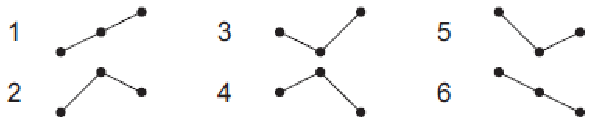
Logistic map $x(i+1) = r\, x(i)[1 - x(i)]$







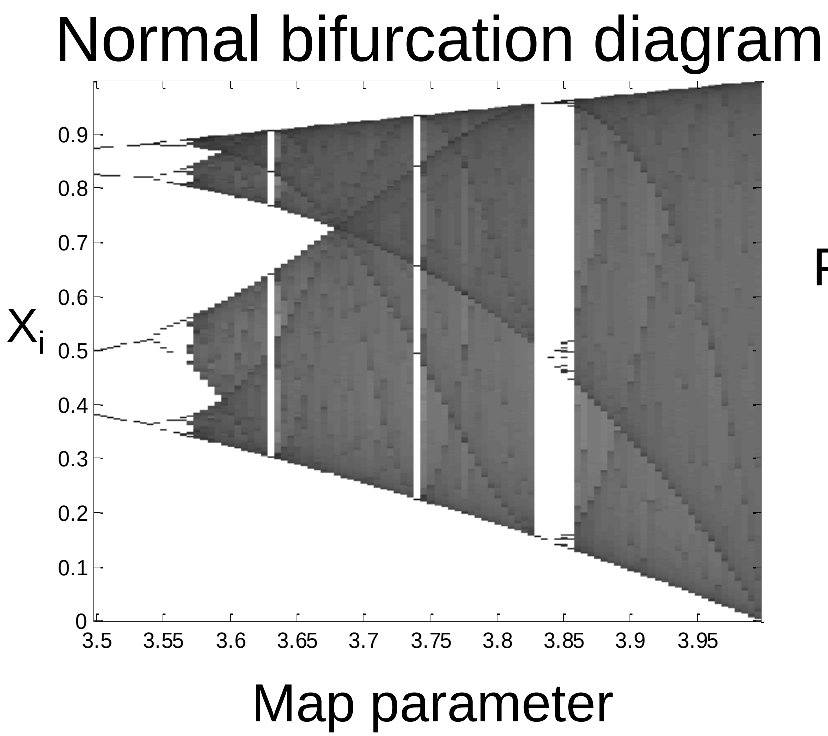


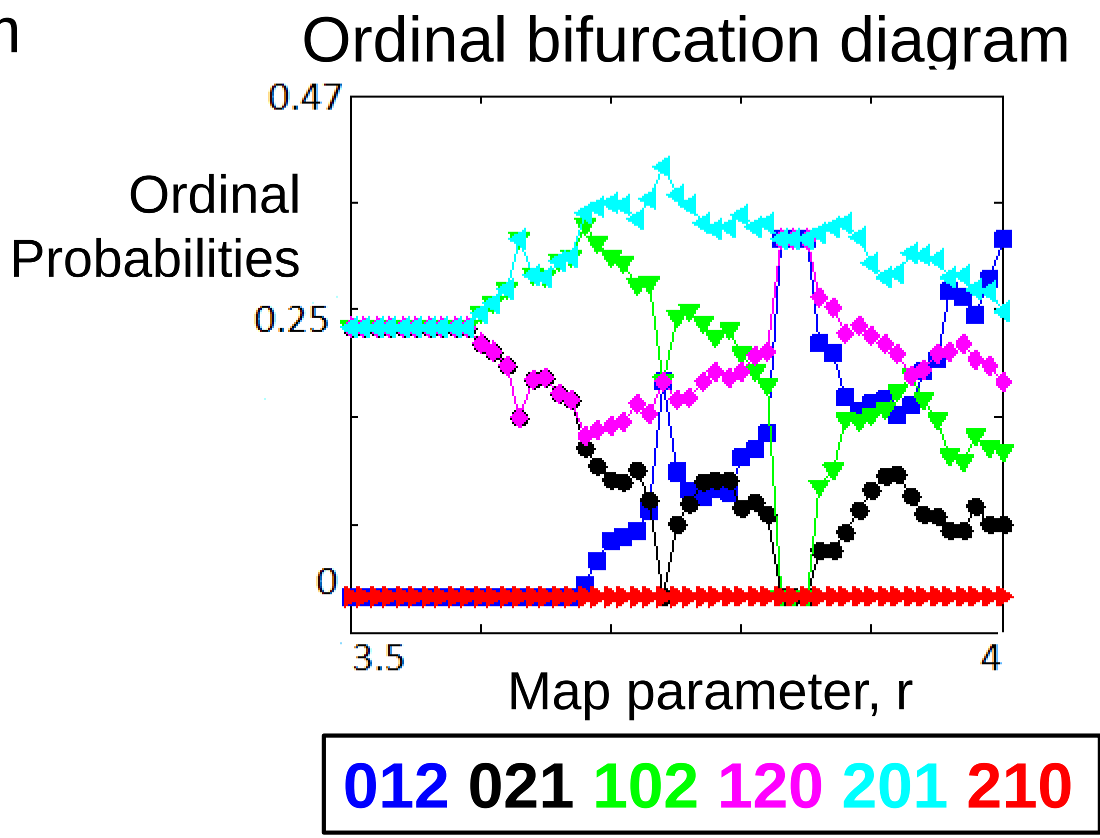


forbidden

Ordinal analysis yields information about more and less expressed patterns in the data

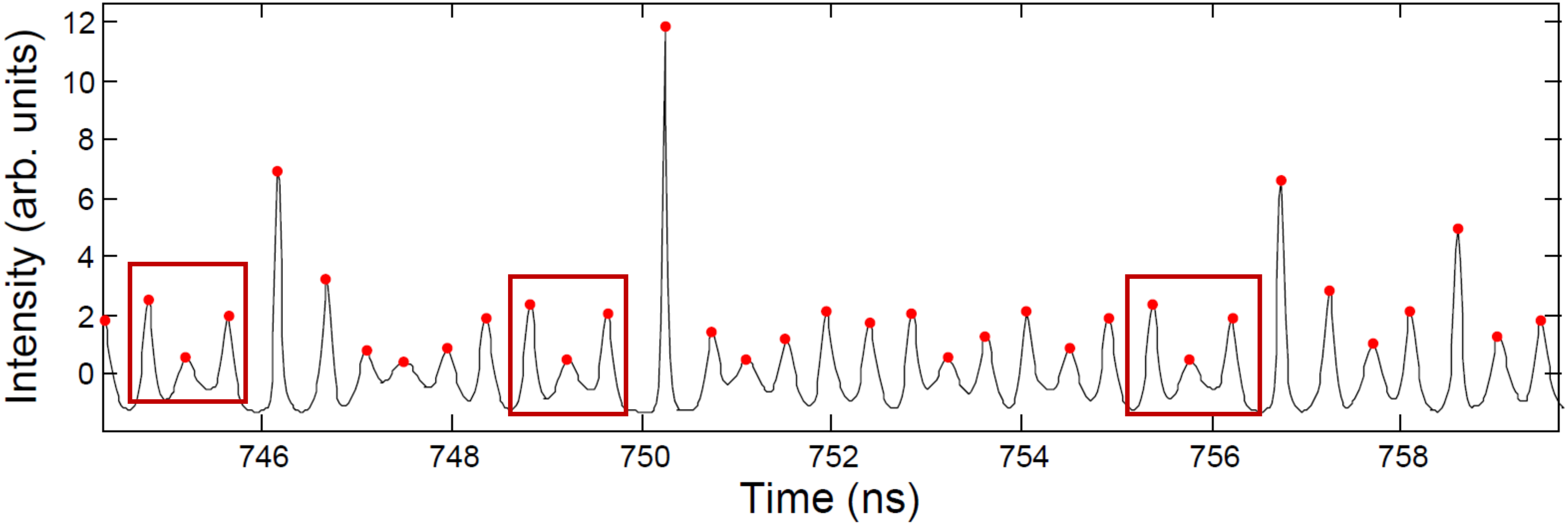
“ordinal” bifurcation diagram of the Logistic map with D=3





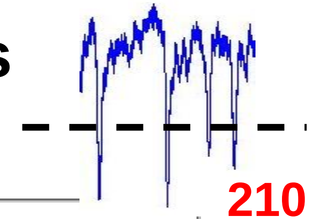
Pattern **210** is always forbidden; pattern **012** is more frequently expressed as r increases

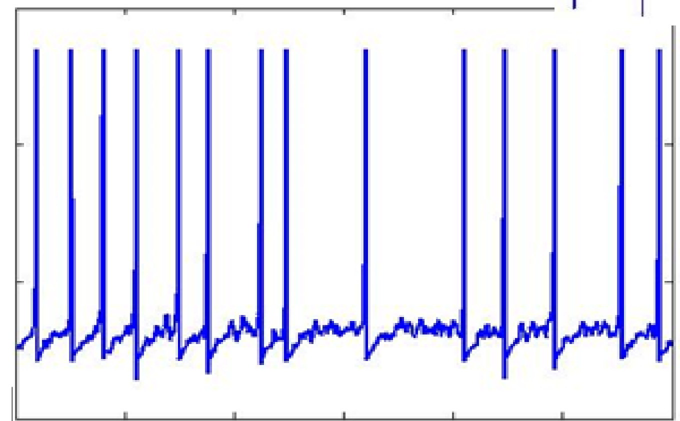
Example: intensity pulses emitted by a chaotic laser

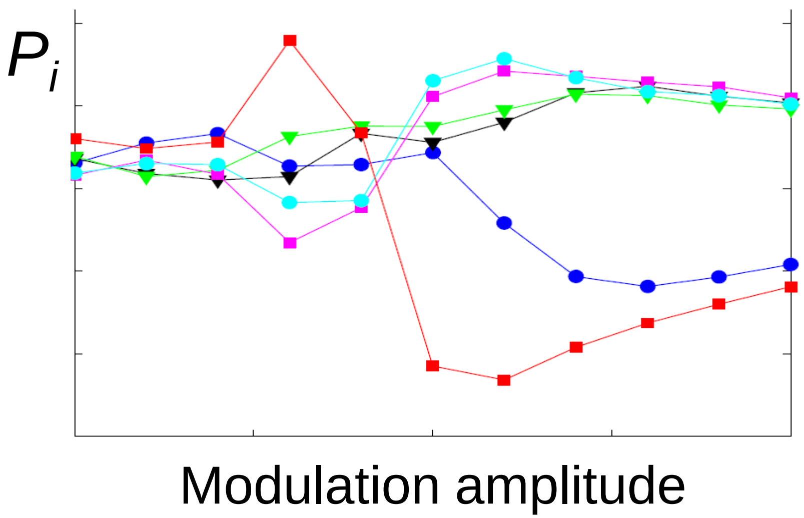


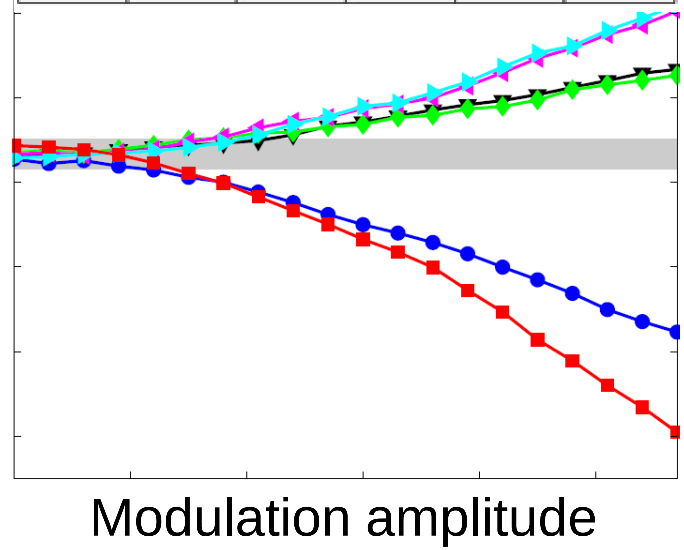
N. Martinez Alvarez, S. Borkar and C. Masoller, “Predictability of extreme intensity pulses in optically injected semiconductor lasers”
Eur. Phys. J. Spec. Top. 226, 1971 (2017).

Example: optical (laser) vs neuronal (model) spikes







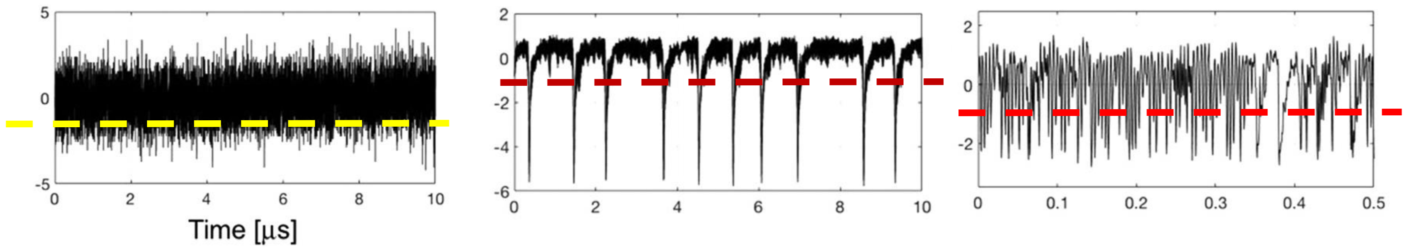


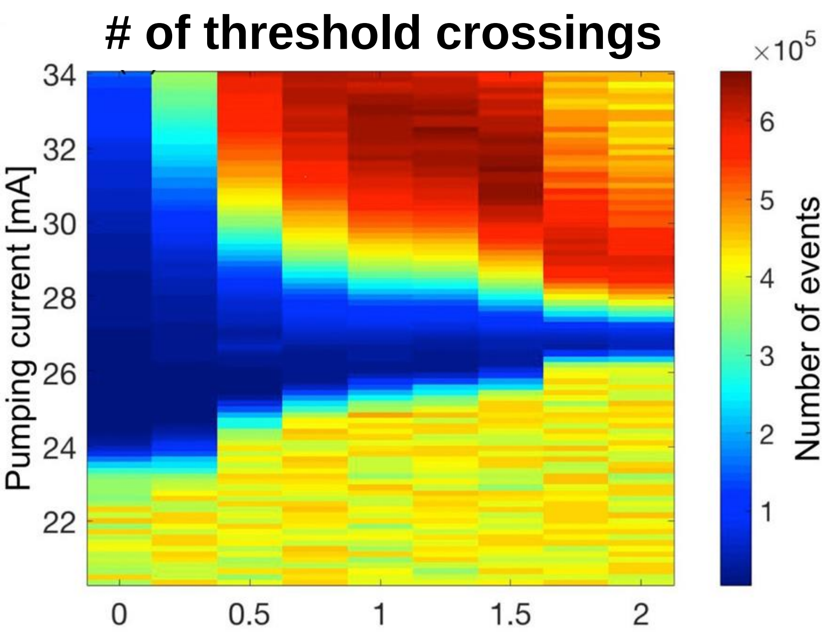
012 021 102 120 201 210

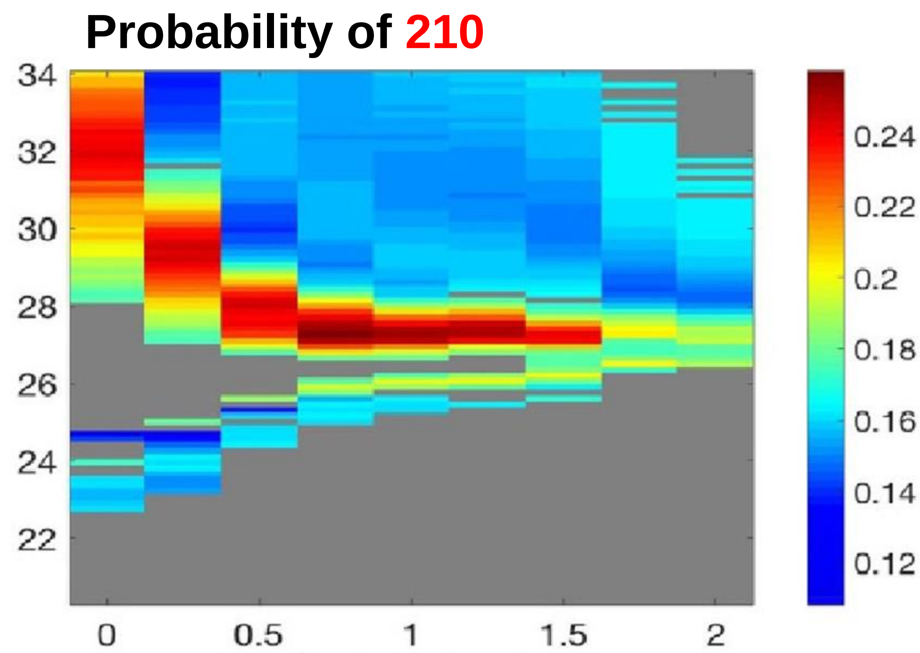
J. M. Aparicio-Reinoso, M. C. Torrent and C. Masoller, PRE 94, 032218 (2016)

Transition noise-chaos in experimental data (diode laser): identifying different dynamical regimes

Laser intensity normalized to μ=0, σ=1







Feedback strength

https://youtu.be/nltBQG_IIWQ

Panozzo et al, Chaos 27, 114315 (2017)

How to detect longer temporal correlations?

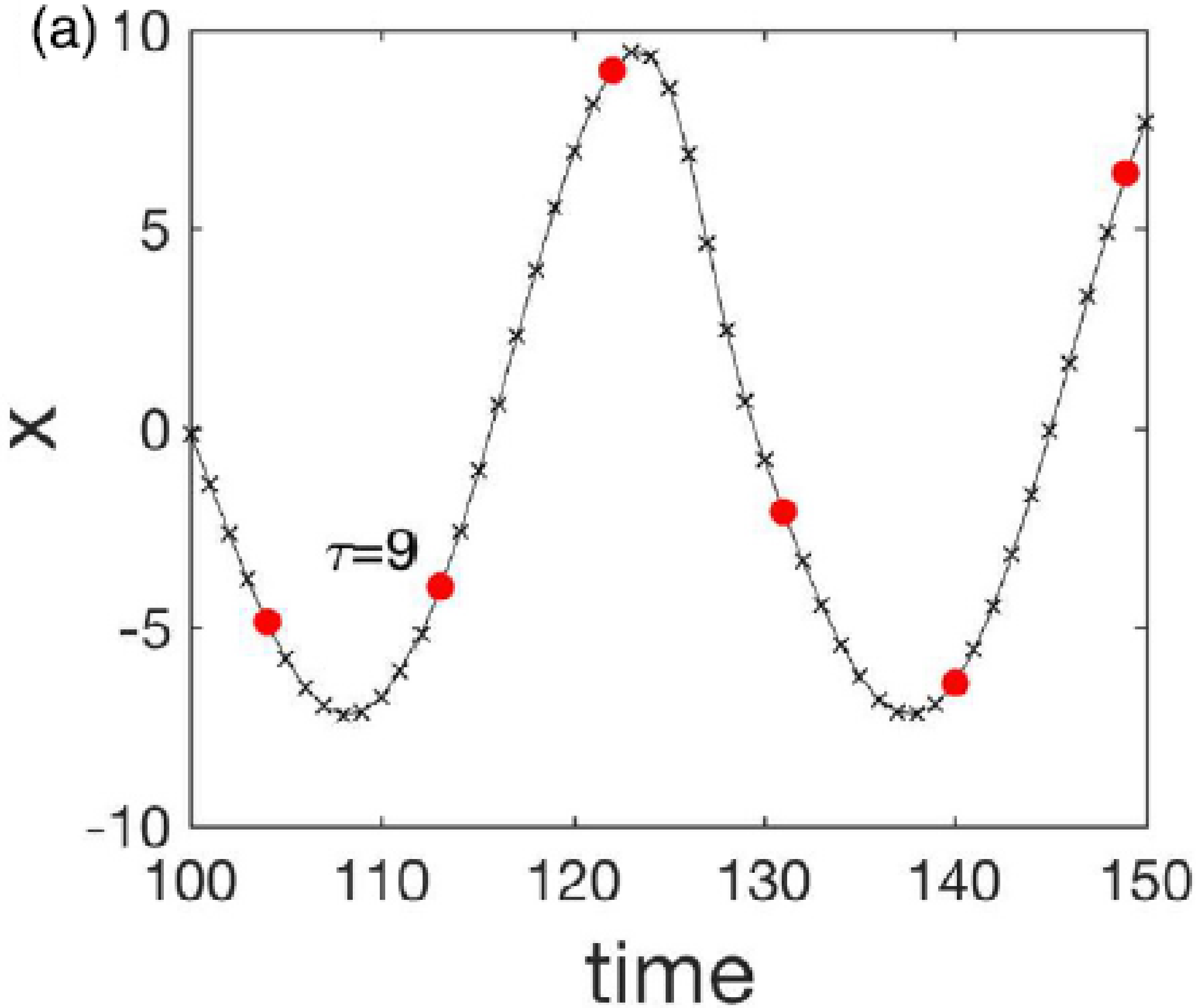
$$[\ldots\, x(t), x(t+1), x(t+2), x(t+3), x(t+4), x(t+5)\ldots]$$

- Problem: number of patterns increases as D!.

- Solution: a lag **τ** allows considering long time-scales without having to use words of many letters

$$[\ldots\, x(t), x(t+2), x(t+4),\ldots]$$

- Example: climatological data (monthly sampled)
  - Consecutive months: $[\ldots x_i(t),\, x_i(t+1),\, x_i(t+2)\ldots]$
  - Consecutive years: $[\ldots x_i(t),\ldots\, x_i(t+12),\ldots\, x_i(t+24)\ldots]$
- **Varying τ = varying temporal resolution (sampling time**)

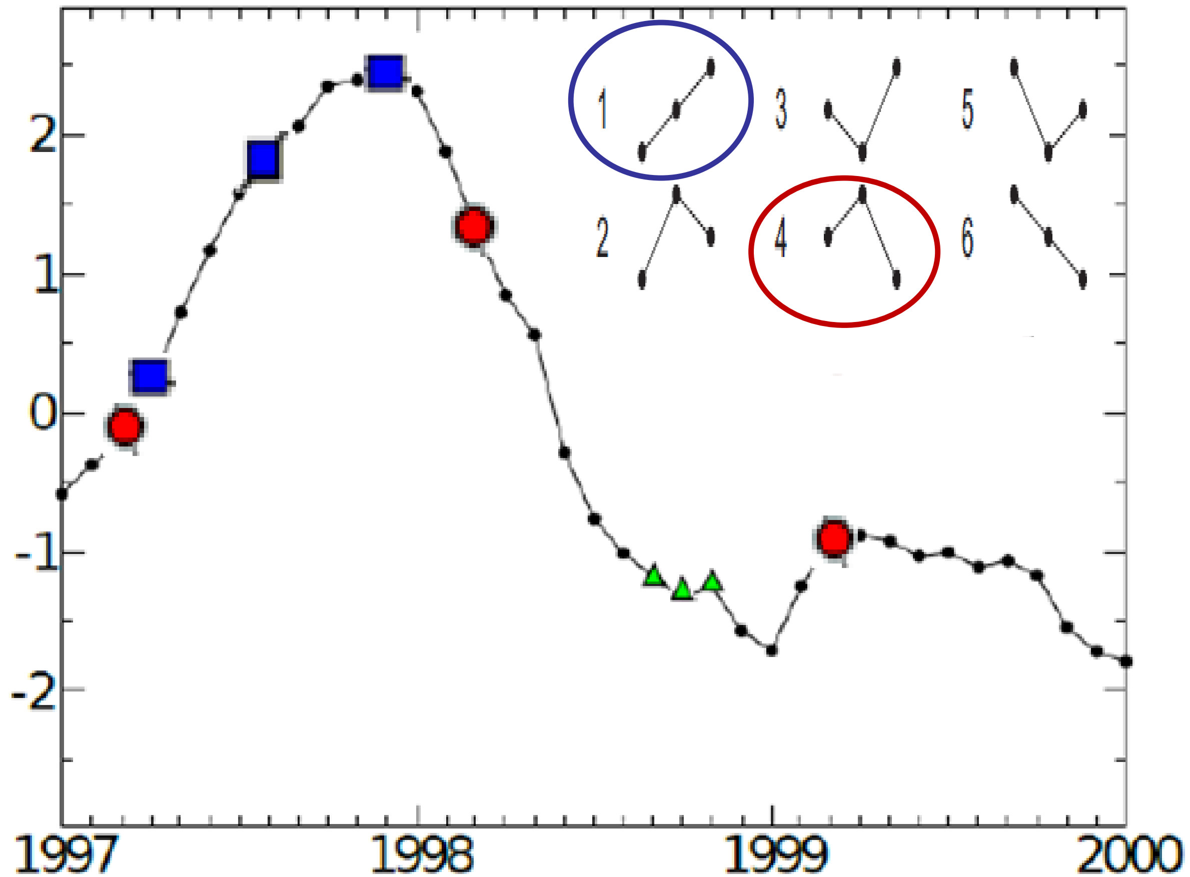


Y. Zou, R.V. Donner, N. Marwan et al. / Physics Reports 787 (2019) 1–97

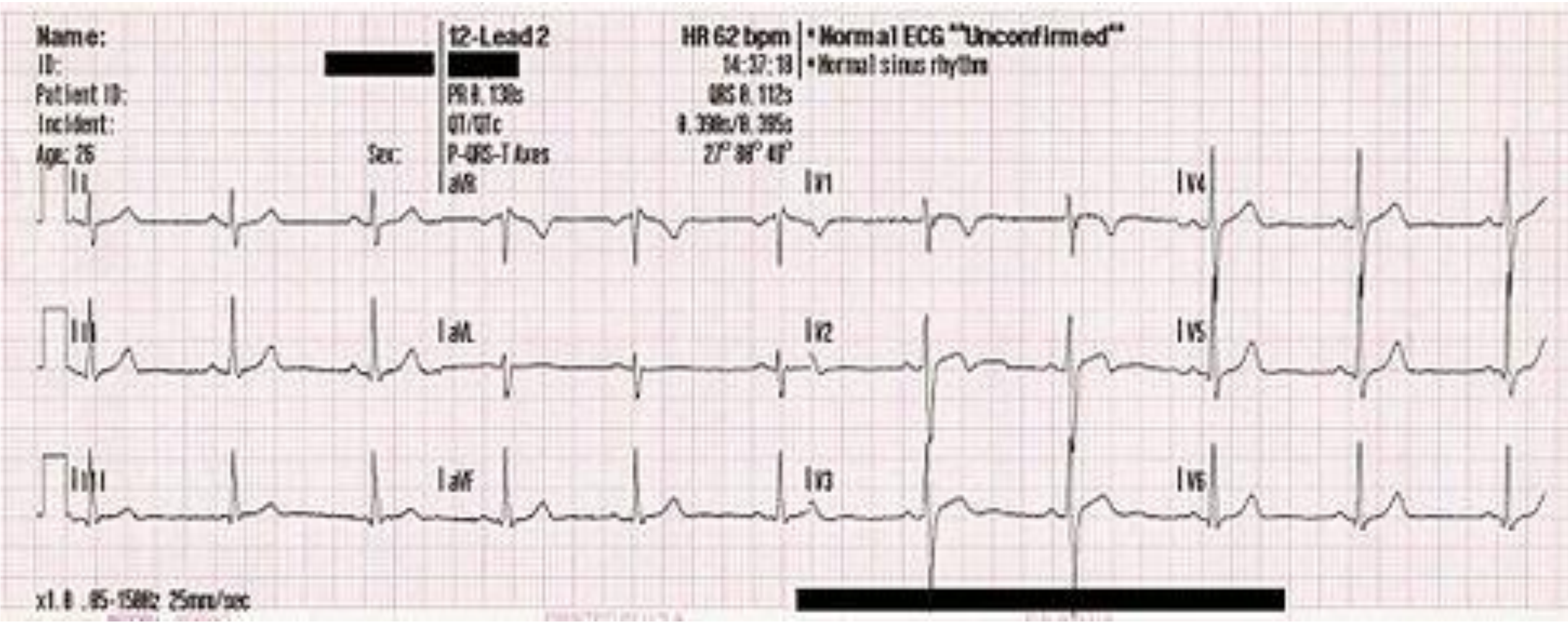
Very useful for “seasonal” data: allows to select the time scale of the analysis

- **Green** triangles: intra-seasonal pattern,
- **blue** squares: intra-annual pattern
- **red** circles: inter-annual pattern

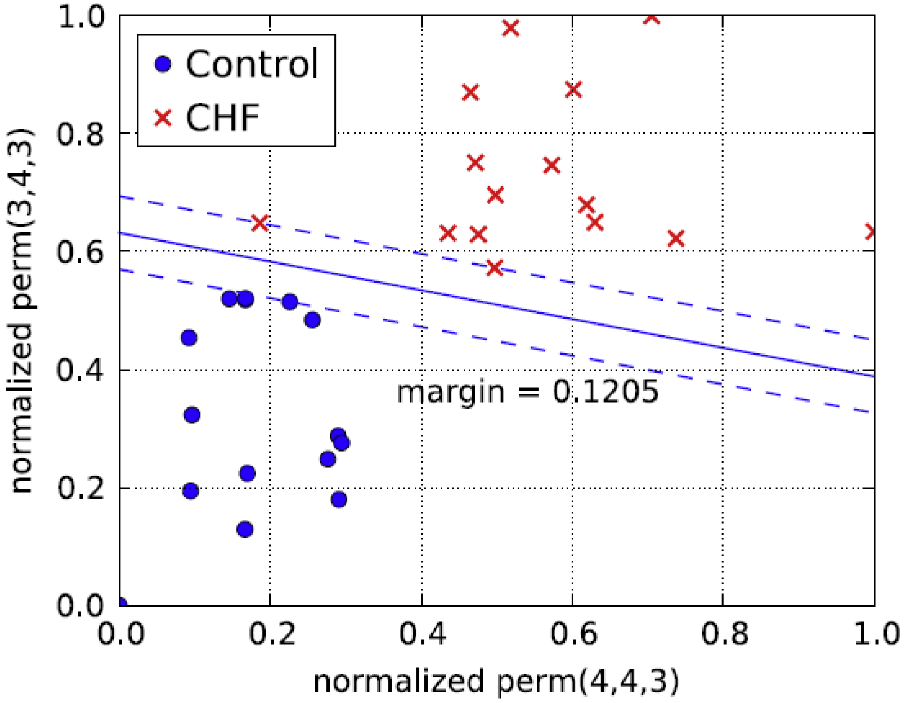
Example: el Niño index, monthly sampled

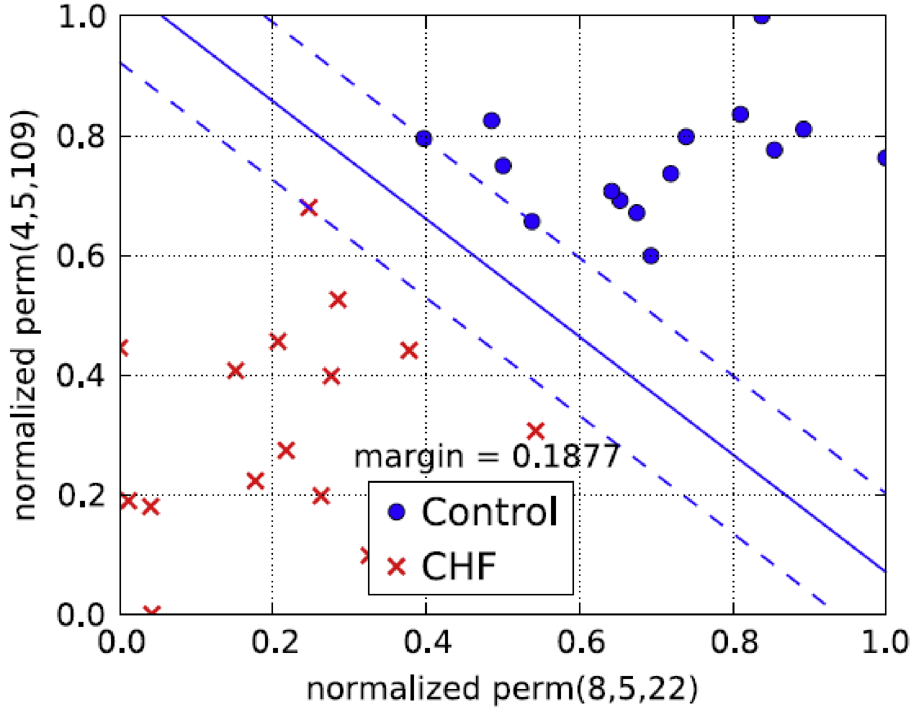


Example: analysis of time series of inter-beat intervals



Classifying ECG signals according to ordinal probabilities



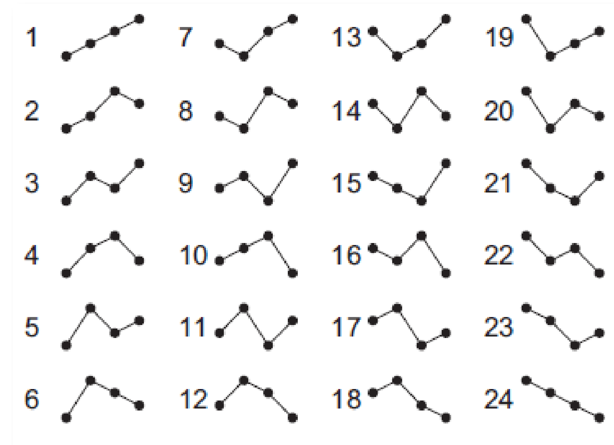


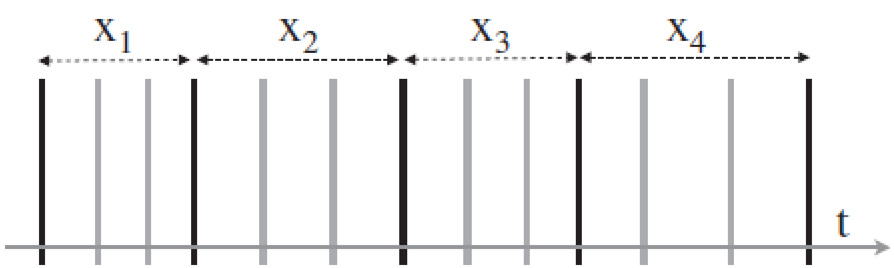
- Analysis of raw data (statistics of ordinal patterns is almost unaffected by a few extreme values)
- The probabilities are normalized with respect to the smallest and the largest value occurring in the data set.

U. Parlitz et al. Computers in Biology and Medicine 42, 319 (2012)

Software

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



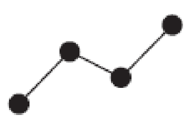


World length (wl): 4

Lag = 3 (skip 2 points)

Result:

indcs= 3

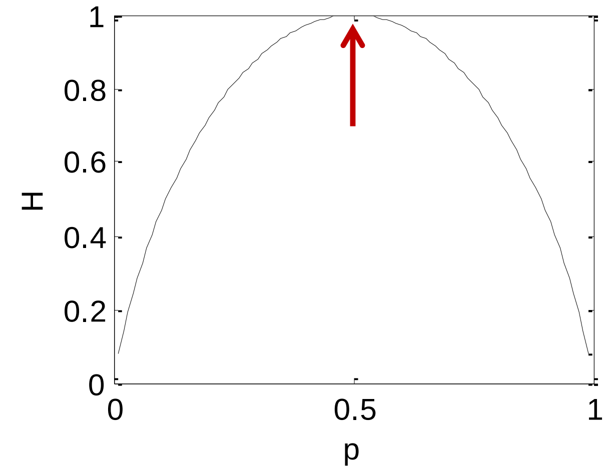


```
function indcs = perm_indices(ts, wl, lag) ;
m = length(ts)-(wl-1)*lag;
 indcs = zeros(m,1) ;
for i = 1:wl-1 ;
   st = ts(1+(i-1)*lag : m+(i-1)*lag) ;
   for j = i:wl-1 ;
      indcs = indcs+(st > ts(1+j*lag : m+j*lag));
   end
   indcs = indcs*(wl-i) ;
end
indcs=indcs + 1 ;
```

Permutation entropy: Shannon entropy computed from the probabilities of the ordinal patterns

- **Shannon** entropy: $H = -\sum_i p_i \log_2 p_i$ $\qquad \sum_{i=1}^{N} p_i = 1$
- What does the entropy represent?
- *Quantity of **surprise** one should feel upon reading the result of a measurement*

  K. Hlavackova-Schindler et al, Physics Reports 441 (2007)
- Simple example: a random variable takes values 0 or 1 with probabilities: p(0) = p, p(1) = 1 − p.



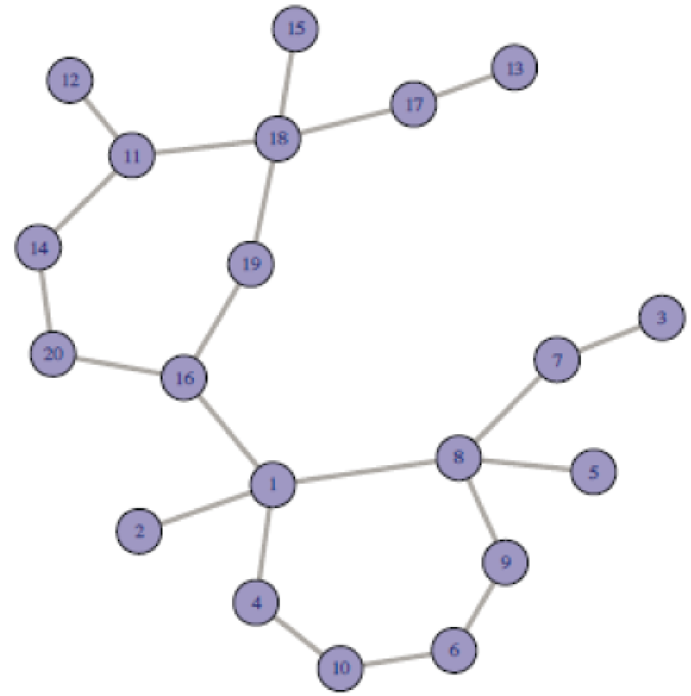
- $H = -p \log_2(p) - (1 - p) \log_2(1 - p)$.

  $\Rightarrow$ p=0.5: Maximum **unpredictability.**

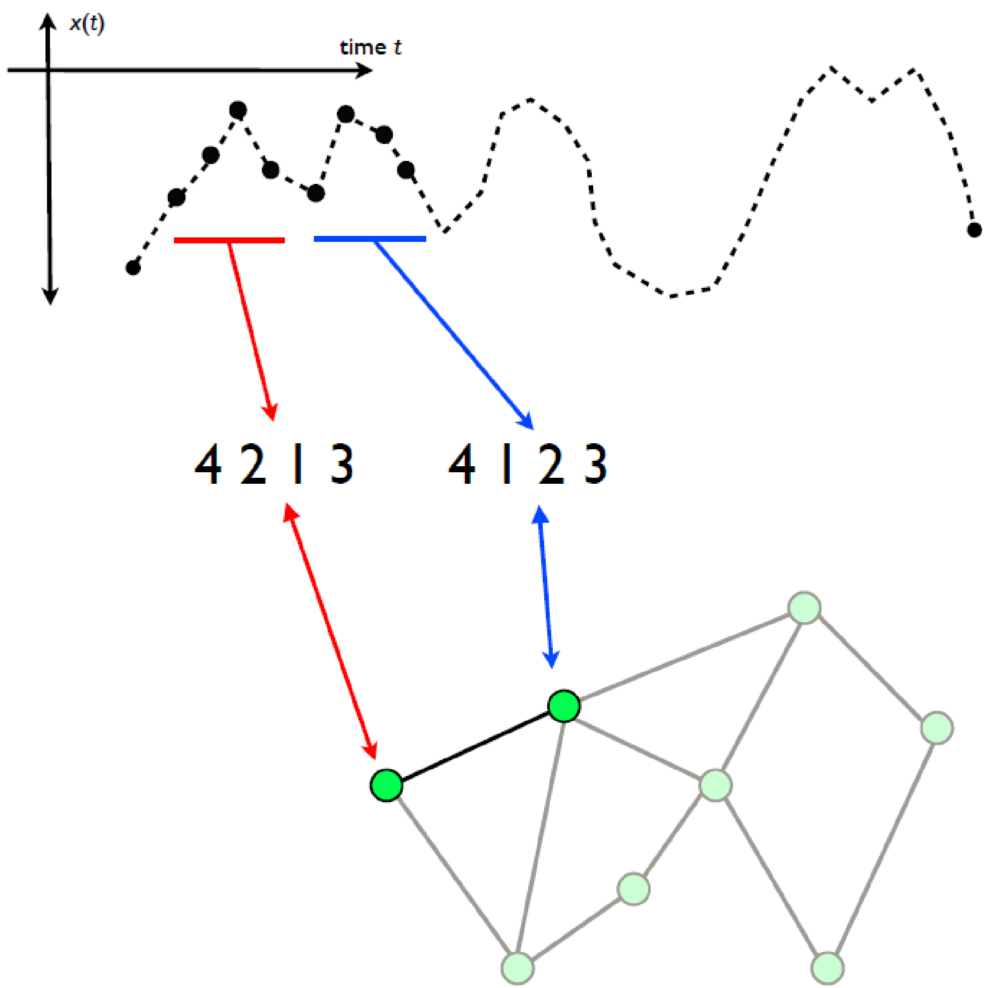
Network representation of a time-series

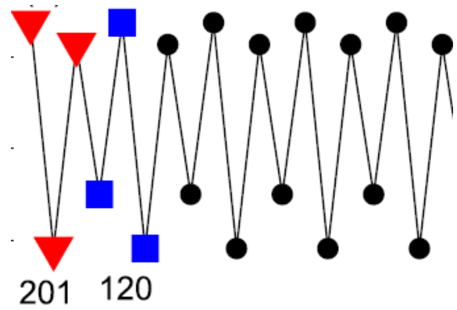
What is a network?

- A graph: a set of “nodes” connected by a set of “links”
- Nodes and links can be weighted or unweighted
- Links can be directed or undirected
- More in part 3 (multivariate analysis)



We use symbolic patterns as the *nodes* of the network. And the *links*? Defined as the transition probability $\alpha \rightarrow \beta$





$201 \rightarrow 120$

- In each node *i*:

  $\sum_j w_{ij}=1$

- <u>Weigh of node i</u>: the probability of pattern *i*

  $(\sum_i p_i=1)$

$\Rightarrow$ **Weighted and directed network**

Source: M. Small

Network-based diagnostic tools

- Entropy computed from node weights (**permutation entropy**)

$$s_p = -\sum p_i \log p_i$$

- Average node entropy (entropy of the link weights)

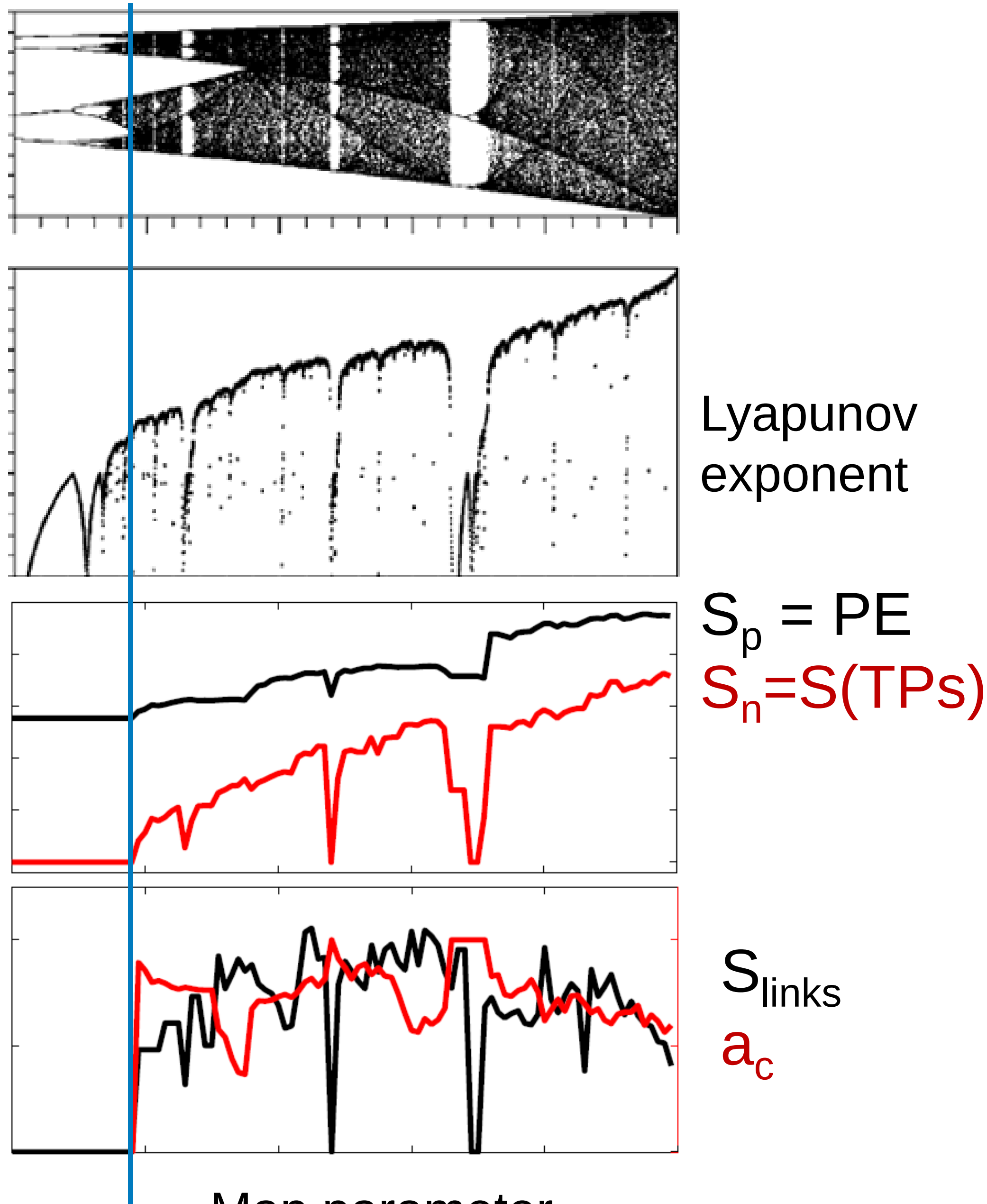
$$s_n = \frac{1}{M}\sum_{i=1}^{M} s_i \qquad s_i = -\sum w_{ij} \log w_{ij}$$

- Asymmetry coefficient: normalized difference of transition probabilities, P('01'→ '10') - P('10'→ '01'), etc.

$$a_c = \frac{\sum_i \sum_{j\neq i} \left| w_{ij} - w_{ji} \right|}{\sum_i \sum_{j\neq i} \left(w_{ij} + w_{ji} \right)}$$

(0 in a fully symmetric network;
1 in a fully directed network)

A first test with the Logistic map

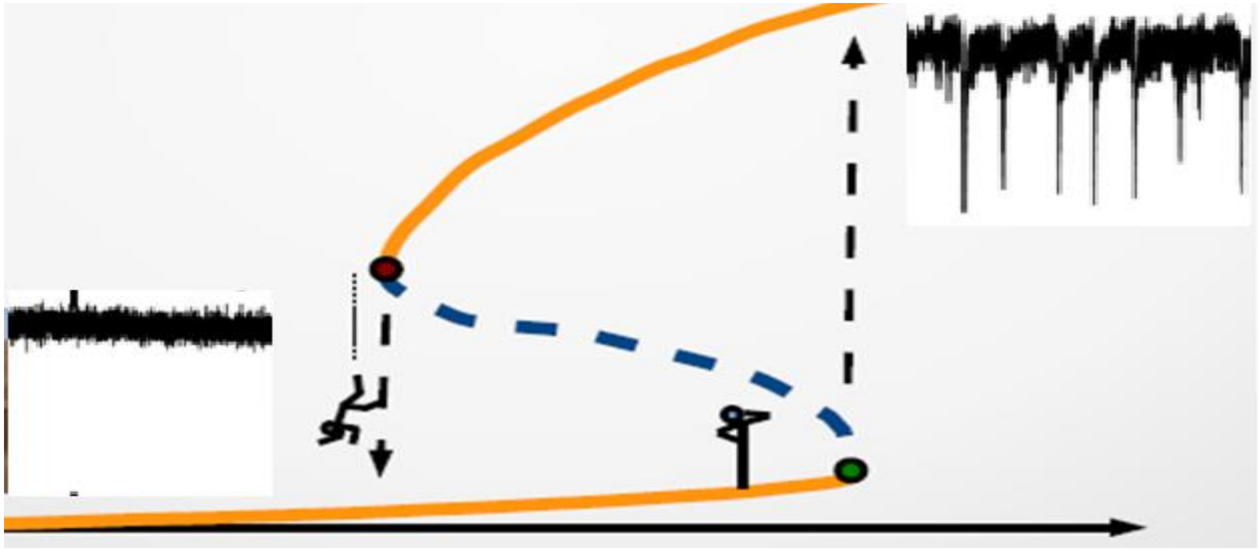


D=4

⇒ Detects the merging of four branches, not detected by the Lyapunov exponent.

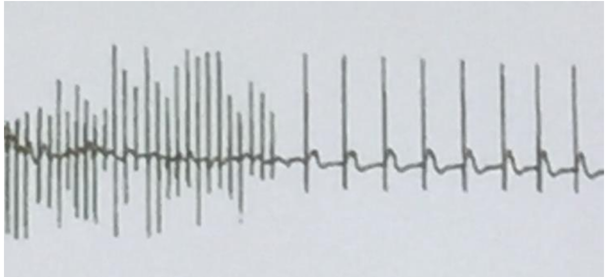
[C. Masoller et al, NJP (2015)]()

Approaching a "tipping point"



Control parameter

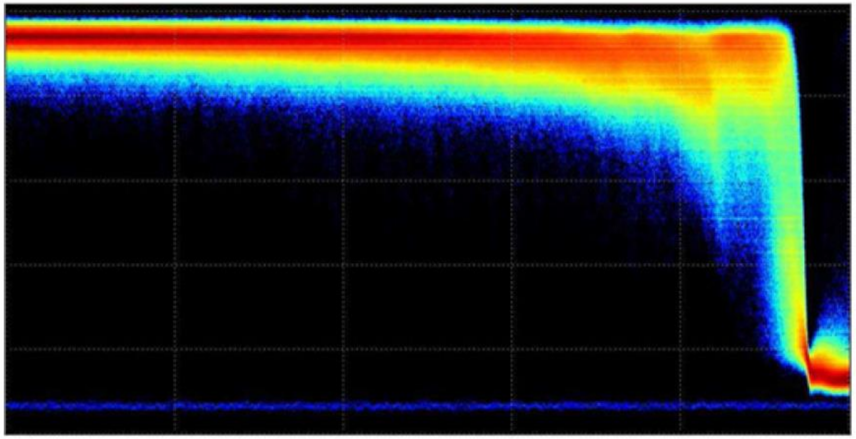
Can we use ordinal networks to detect transitions between different dynamical regimes?



Yes! Two examples: optical signals and brain signals

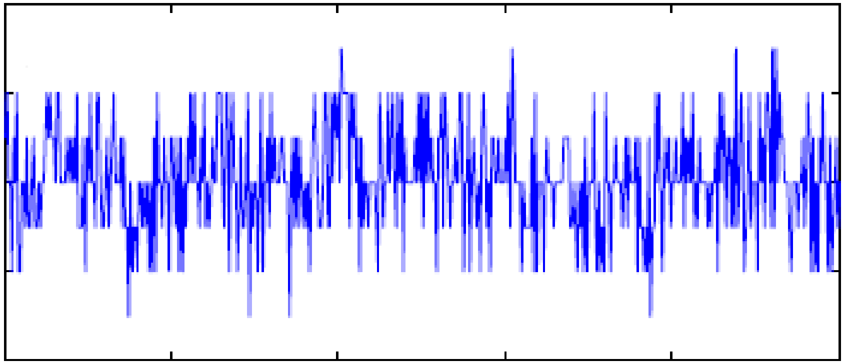
Apply the ordinal network method to laser data

As the laser current increases



Time

Intensity @ constant current



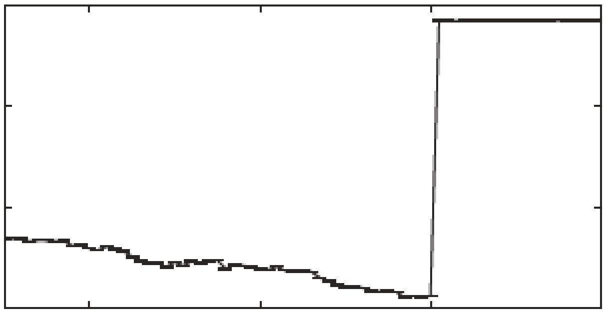
Time

- Two sets of experiments: intensity time series were recorded
  - keeping constant the laser current.
  - while increasing the laser current.
- We analyzed the polarization that turns on / turns off.

Is it possible to anticipate the switching?
No if the switching is fully stochastic.

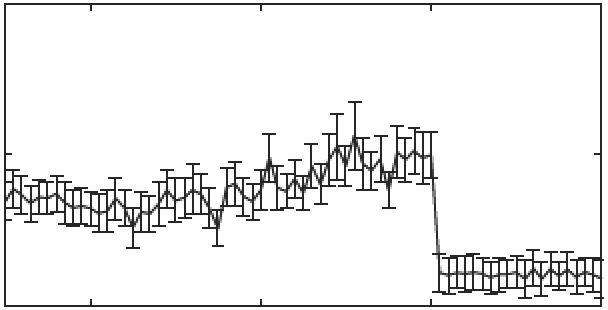
First set of experiments (the current is kept constant): despite of the stochasticity of the time-series, the node entropy “anticipates” the switching

$\langle I \rangle$



Laser current

σ_I

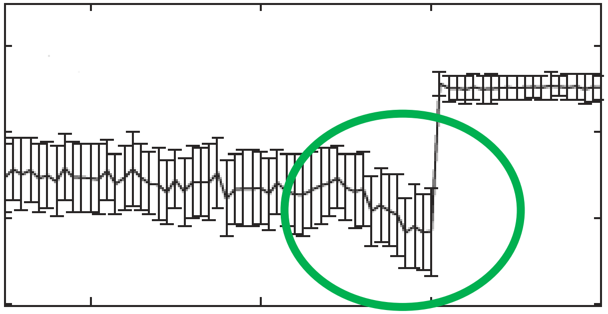


Laser current

L=1000
100 windows

$\Rightarrow$ **No warning**

Node entropy s_n (D=3)



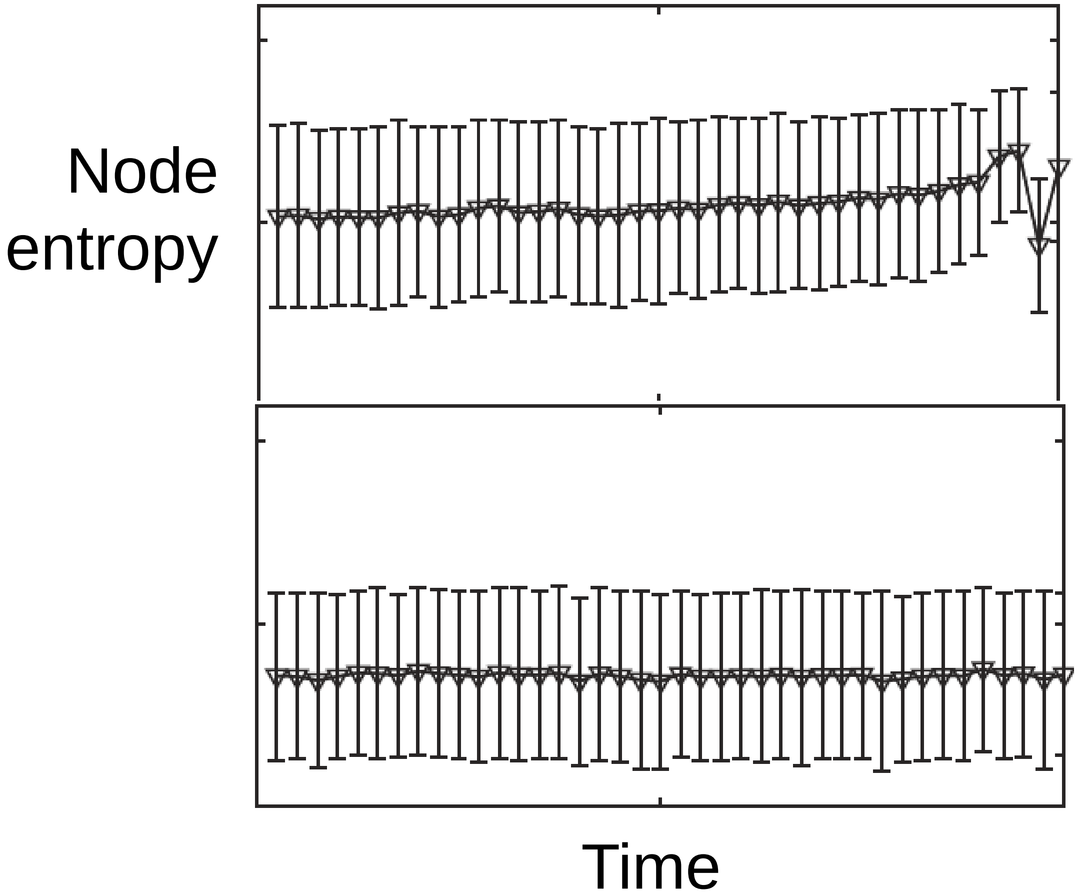
Laser current

Early warning

$\Rightarrow$ Deterministic mechanisms must be involved.

C. Masoller et al, NJP (2015)

In the second set of experiments (current increases linearly in time): an early warning is also detected



L=500, D=3
1000 time series

With slightly different experimental conditions: no switching.

C. Masoller et al, NJP (2015)

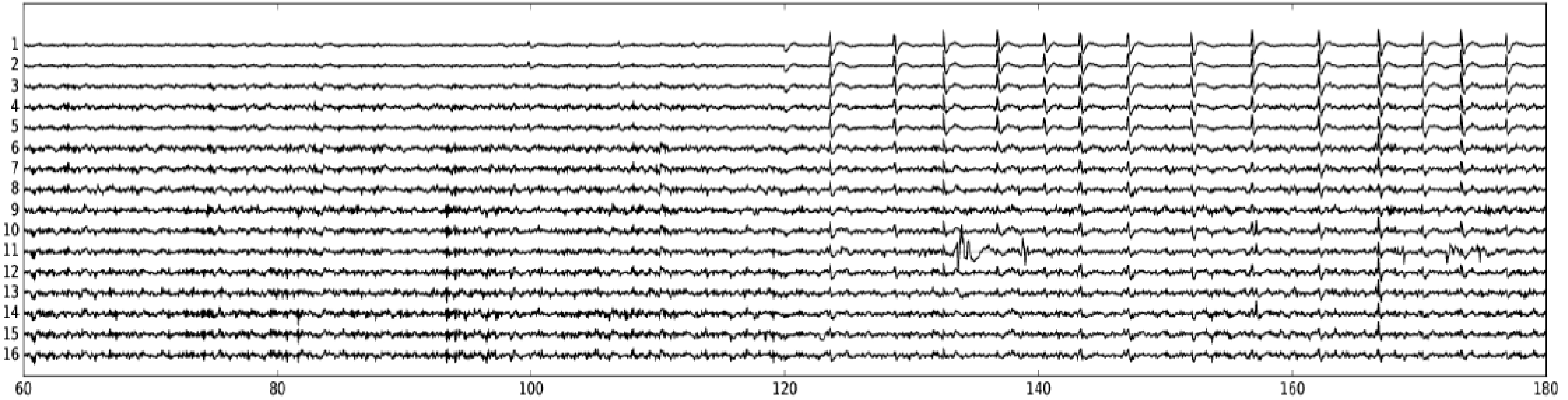
Second application of the ordinal network method: distinguishing *eyes closed* and *eyes open* brain states

Analysis of two EEG datasets

	BitBrain DTS1	PhysioNet DTS2
Sampling rate(Hz)	256	160
Time task(seg)	120	60
Total points	30720	9600
Number of electrodes	16	64
Number of subjects	70	109

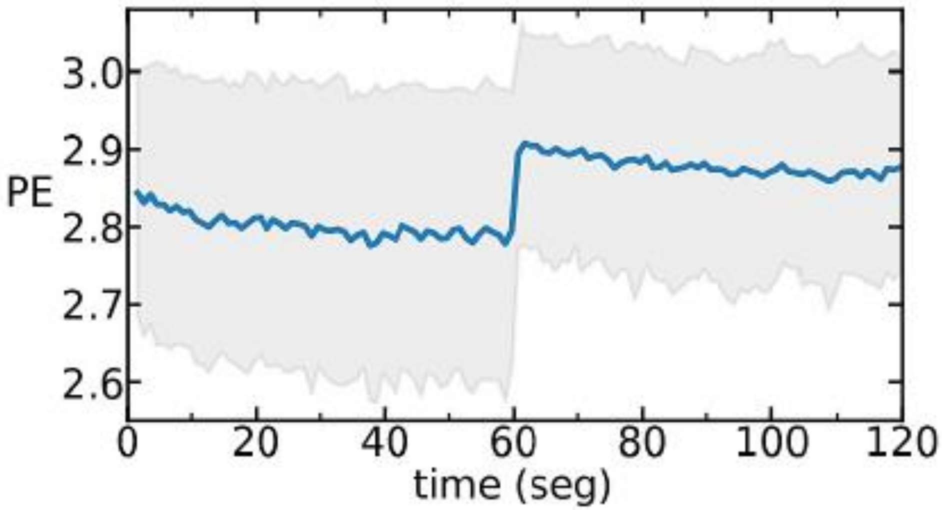
Eye closed

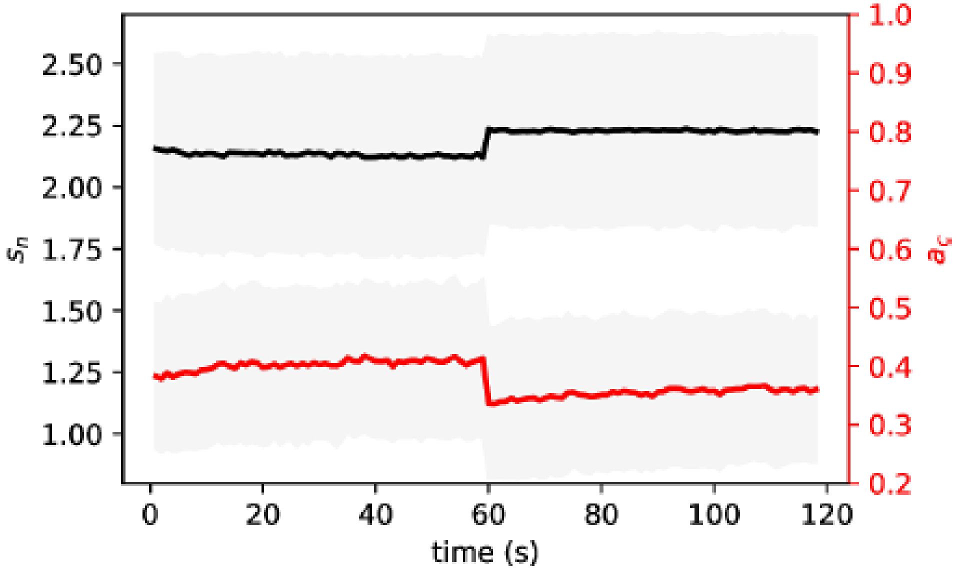
Eye open



- Symbolic analysis is applied to the raw data; similar results were found with filtered data using independent component analysis.

“Randomization”: the entropies increase and the asymmetry coefficient decreases

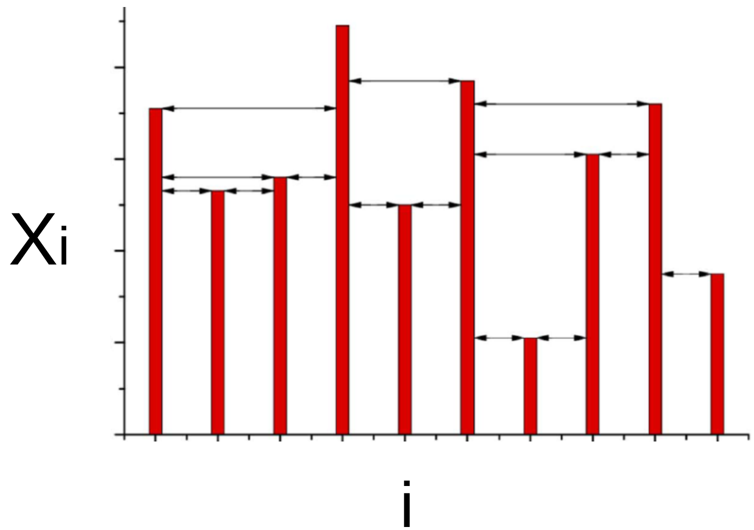




Time window = 1 s
(160 data points)

C. Quintero-Quiroz et al, “*Differentiating resting brain states using ordinal symbolic analysis*”, Chaos 28, 106307 (2018).

Another way to represent a time series as a network: the horizontal visibility graph (HVG)



Rule: data points *i* and *j* are connected if there is "visibility" between them

$\Rightarrow$ **Unweighted and undirected graph**

Parameter free!

Luque et al PRE (2009); Gomez Ravetti et al, PLoS ONE (2014)

Exercise

Consider the following time series:

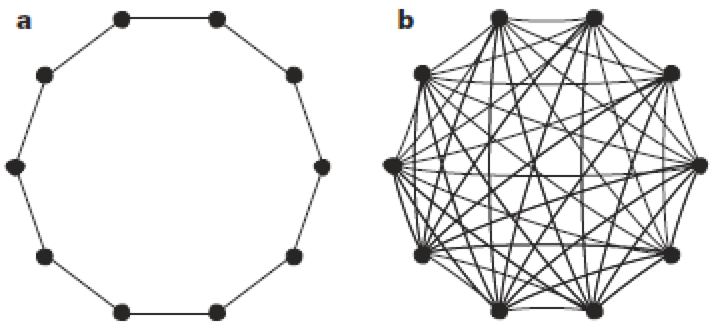
$$x = \{0.71, 0.53, 0.56, 0.89, 0.50,$$
$$0.77, 0.21, 0.6, 0.72, 0.35\}.$$

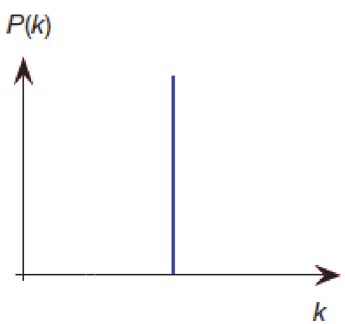
How many links ("degree") does each data point have?

$$k = \{3, 2, 3, 4, 2, 5, 2, 3, 3, 1\}$$

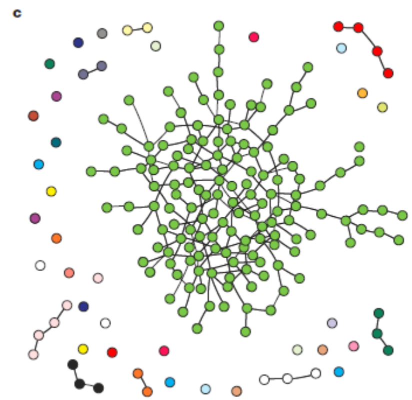
How to characterize the HV graph? The degree distribution

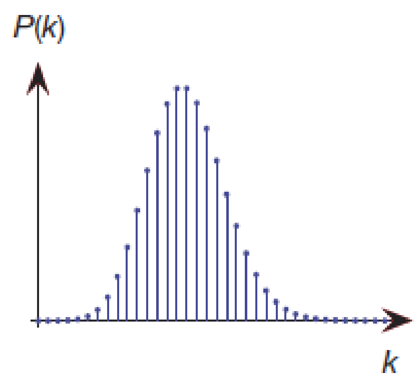
Regular



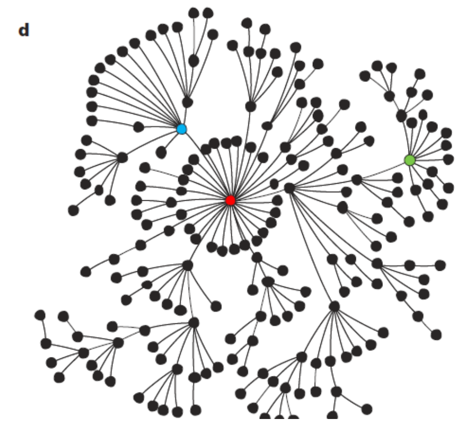


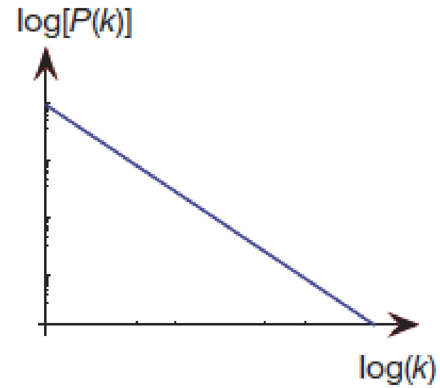
Random





Scale-free





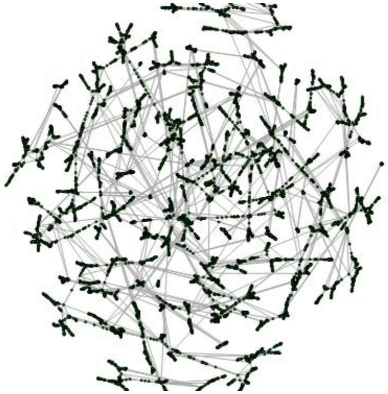
From the degree distribution of the horizontal visibility graph we calculate the entropy: **HVG entropy**

Strogatz, Nature 2001

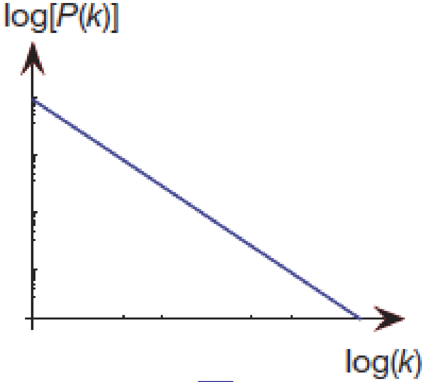
HVG entropy: computed from the degree distribution

Time series

HV graph

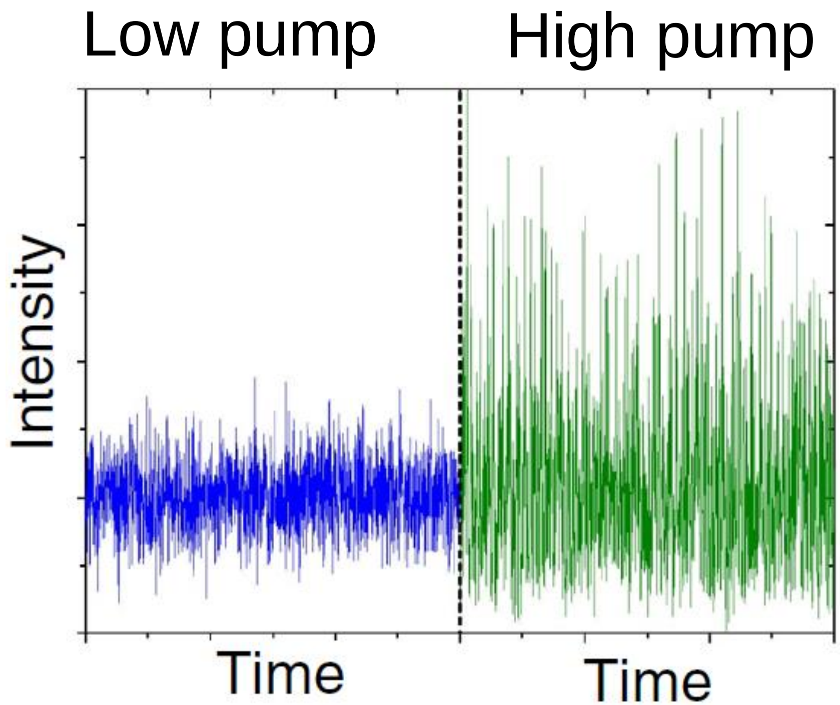


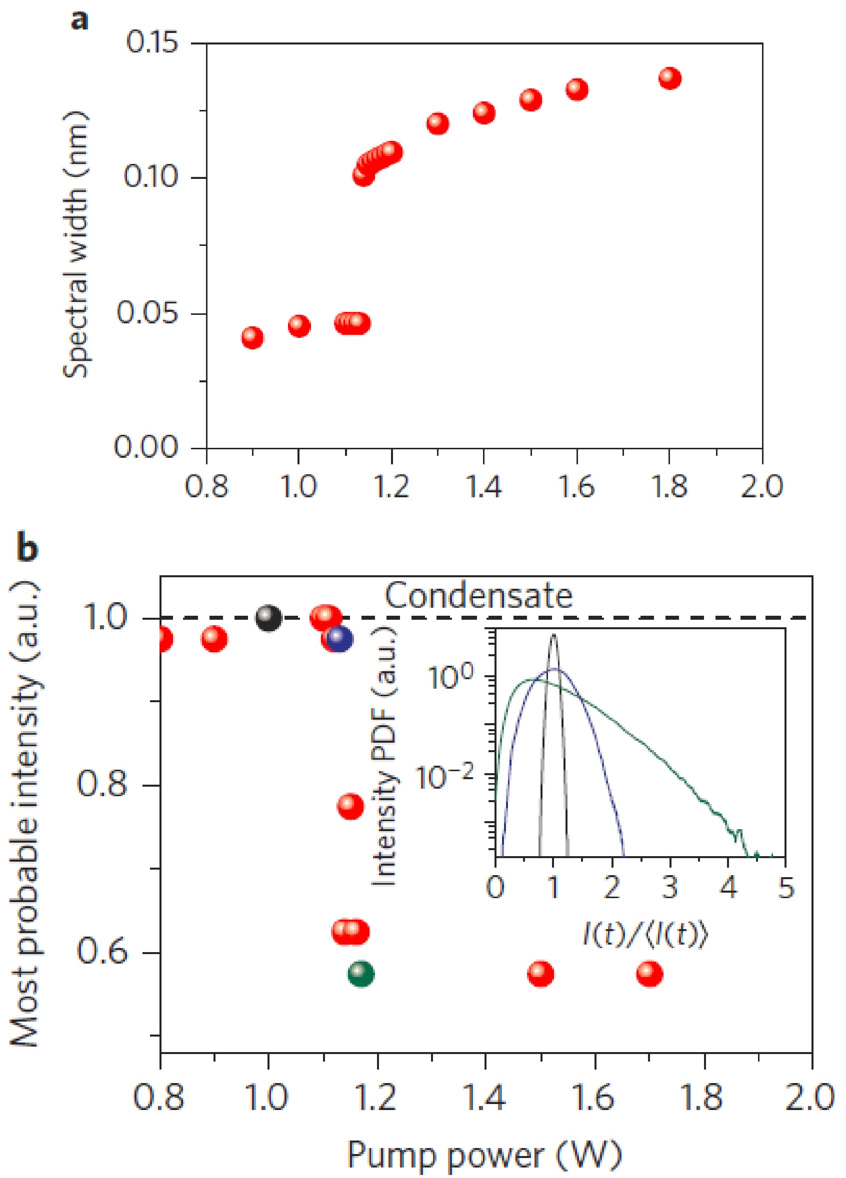
Degree distribution



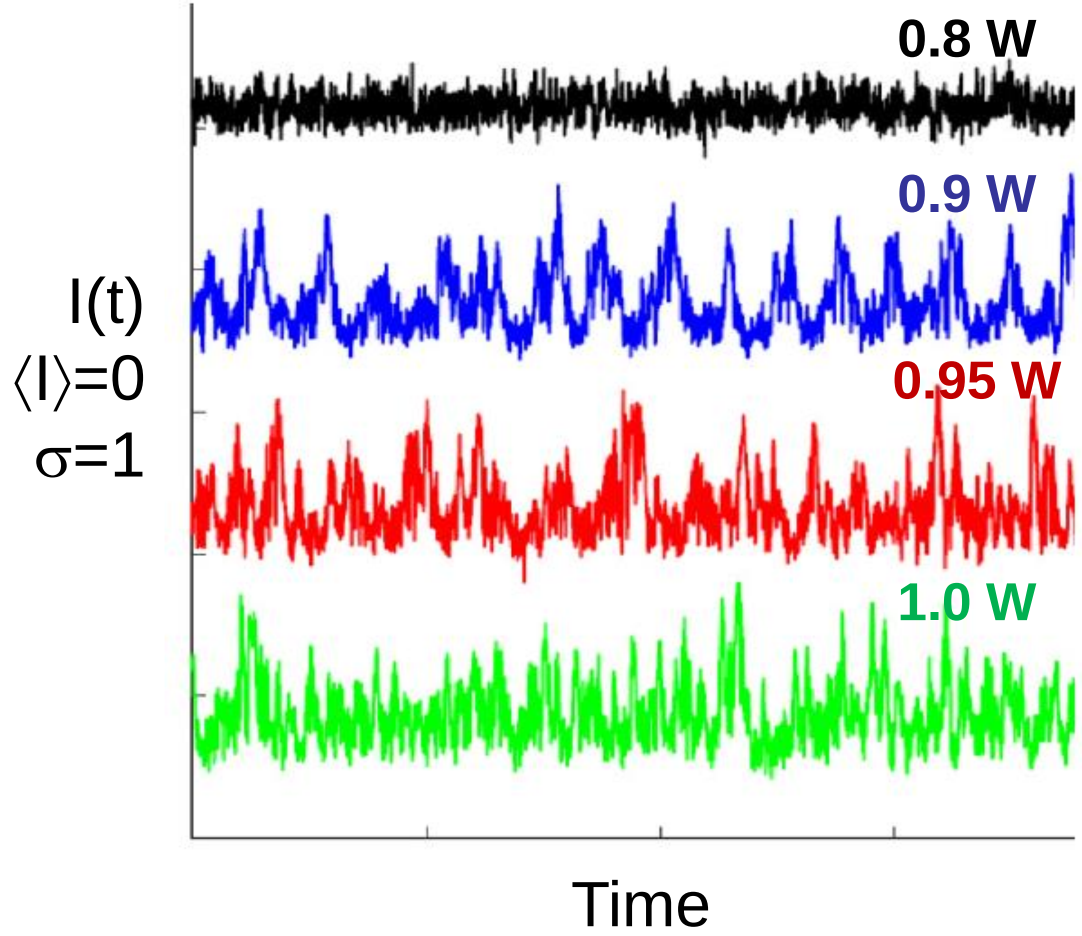
HVG entropy $H = -\sum_{k} p(k) \log p(k)$

Example of application: Laminar → Turbulence transition in a fiber laser as the pump (control parameter) increases



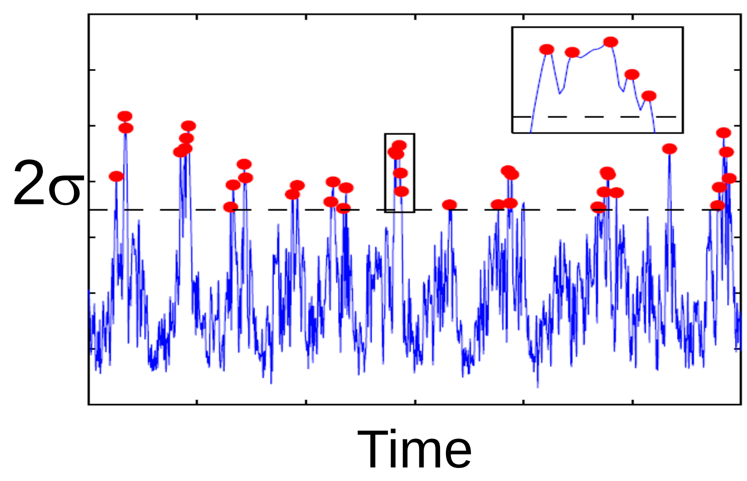


E. G. Turitsyna et al Nat. Phot. 7, 783 (2013)



Nonlinear temporal correlations?

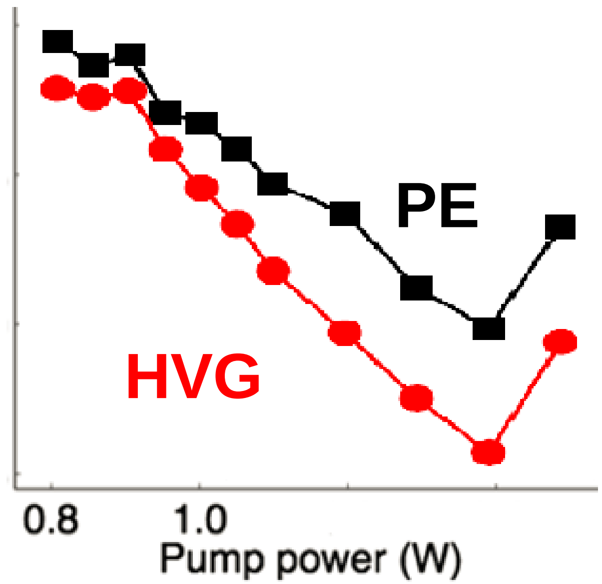
Raw and **thresholded** data



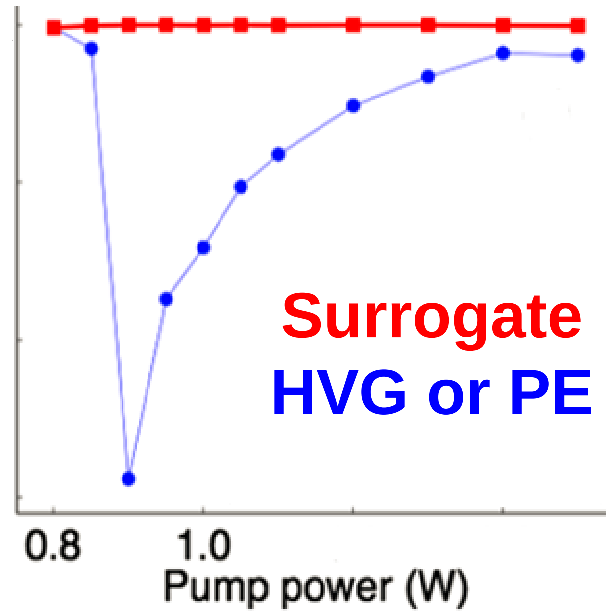
L. Carpi and C. Masoller, "*Persistence and stochastic periodicity in the intensity dynamics of a fiber laser during the transition to optical turbulence*", Phys. Rev. A **97**, 023842 (2018).

Four ways to compute the Entropy

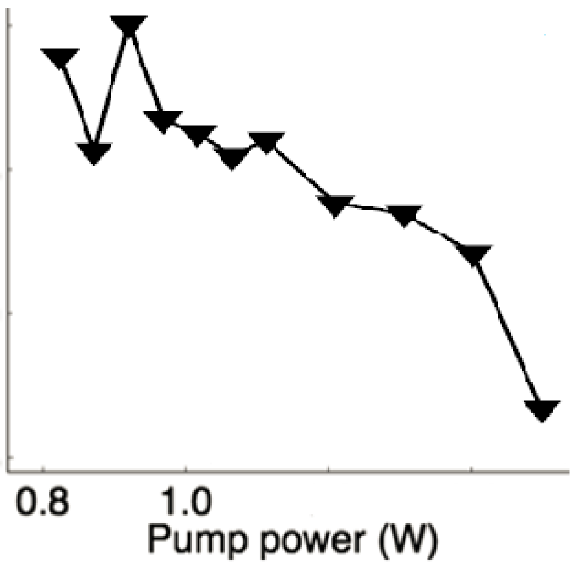
PE/HVG from
“raw” data



Using only the
“thresholded” data



From the histogram
of “raw” values



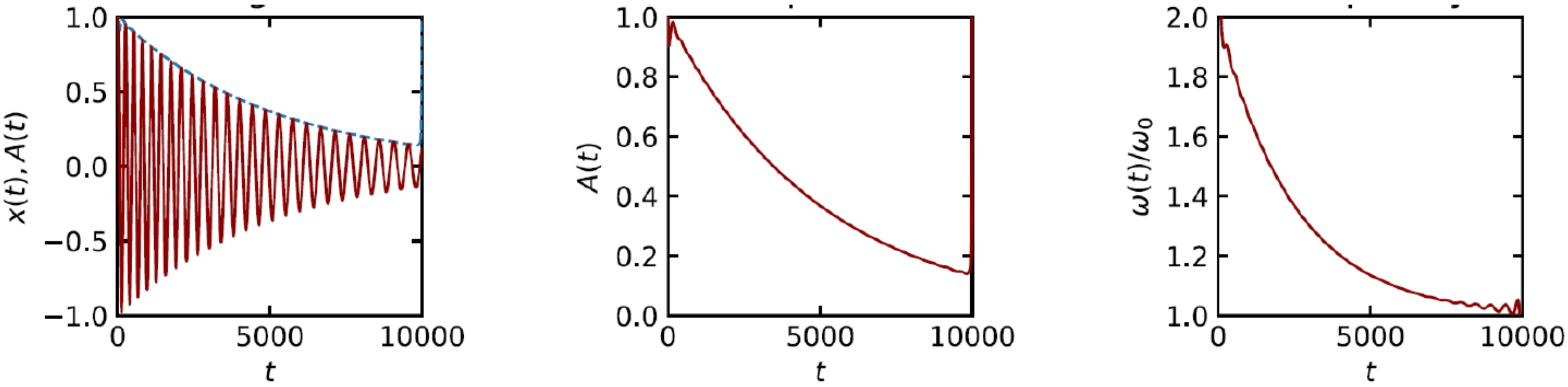
(the abrupt transition is
robust with respect to the
selection of the threshold)

[Aragoneses et al, PRL (2016)]()

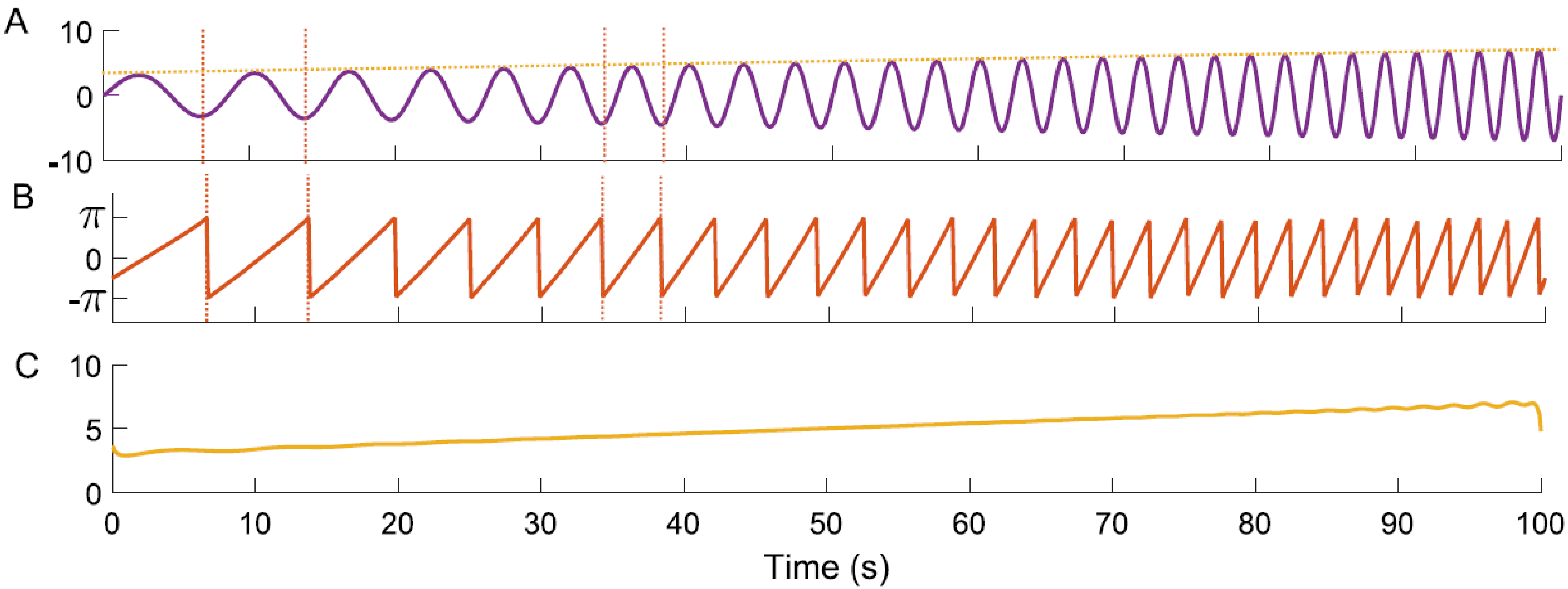
For the analysis of oscillatory signals: phase and amplitude information

How to obtain instantaneous amplitude and frequency information from a time series?

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



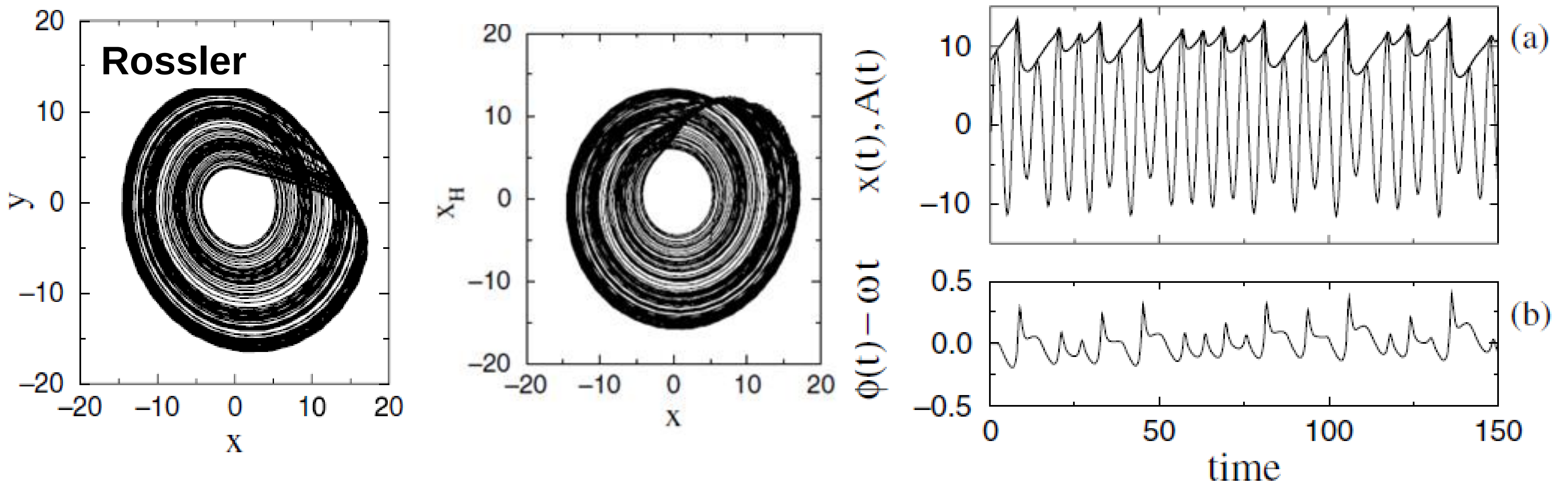
Example: sine wave with increasing amplitude and frequency



- (A) The original signal. (B) The instantaneous phase extracted using the Hilbert transform. (C) The instantaneous amplitude.
- A = C cos(B).

G. Lancaster et al, Physics Reports 748 (2018) 1–60

Second example



$$\begin{pmatrix} \dot{x} \\ \dot{y} \\ \dot{z} \end{pmatrix} = \begin{pmatrix} -y - z \\ x + ay \\ b + z(x - c) \end{pmatrix}$$

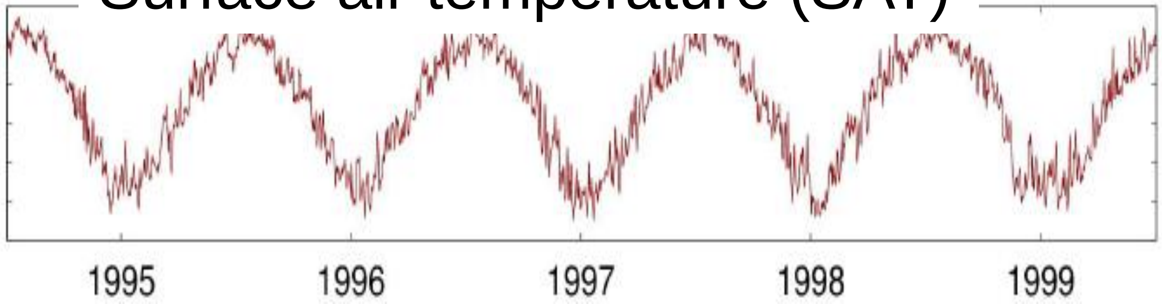
A. Pikovsky, M. Rosenblum and J. Kurths 2001

Third example

Normalization: μ=0, σ=1

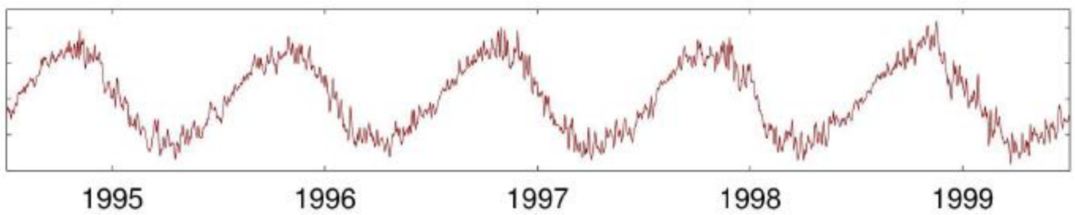
Surface air temperature (SAT)

x

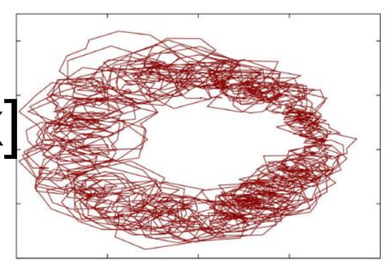


- HT[sin(ωt)]=cos(ωt)

HT[x]



y=HT[x]



x

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

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

Zappala, Barreiro and Masoller, Entropy (2016)

- For a real time series x(t) defines an *analytic signal*

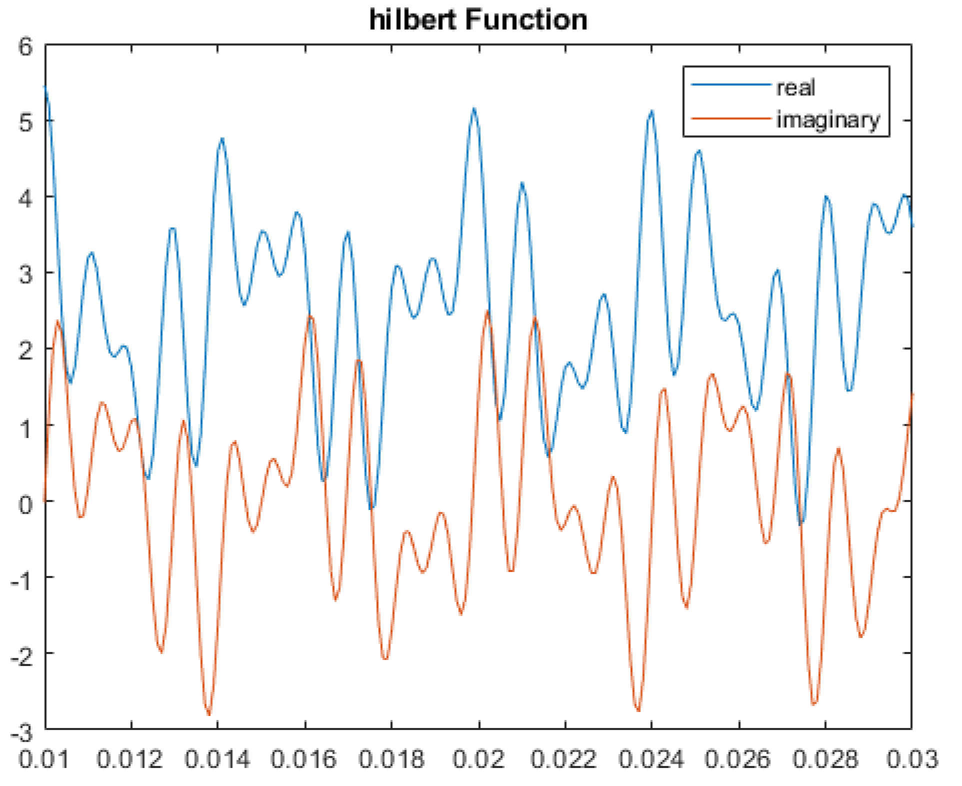
$$\zeta(t) = x(t) + iy(t) = a(t)e^{i\varphi(t)}$$

$$y(t) = H[x(t)] = \pi^{-1}\text{P.V.}\int_{-\infty}^{+\infty}\frac{x(\tau)}{t-\tau}d\tau$$

<u>A word of warning</u>:
Although formally *a(t)* and *φ(t)* can be defined for any *x(t)*, **they have a clear physical meaning only if *x(t)* is a narrow-band oscillatory signal**: in that case, the ***a(t)*** coincides with the envelope of *x(t)* and the **instantaneous frequency, *ω(t)=dφ/dt***, coincides with the dominant frequency in the power spectrum.

Hilbert with matlab

```
x = 2.5 + cos(2*pi*203*t) + sin(2*pi*721*t) + cos(2*pi*1001*t);
y = hilbert(x);
plot(t,real(y),t,imag(y))
xlim([0.01 0.03])
legend('real','imaginary')
title('hilbert Function')
```

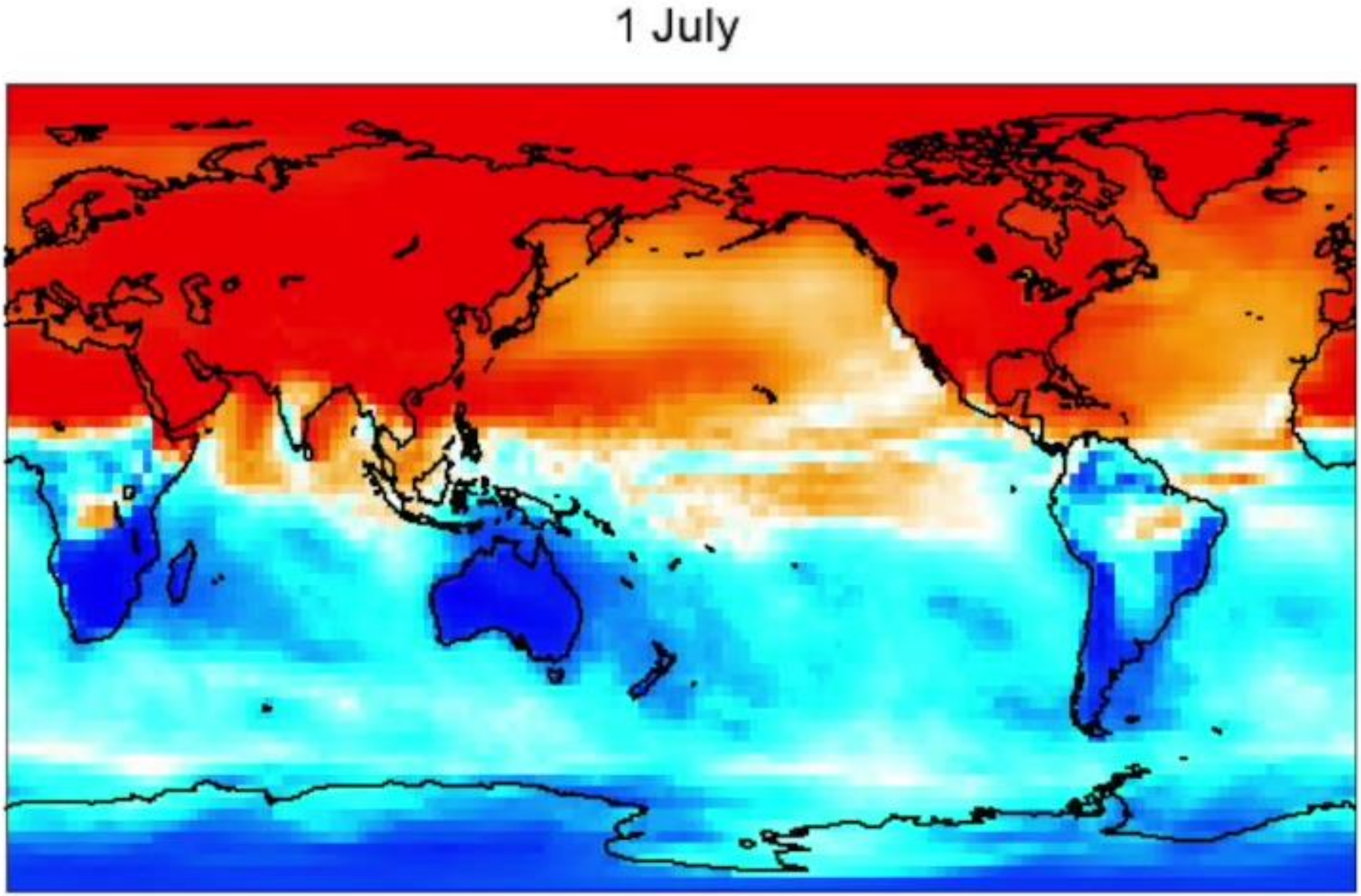


The sampling rate must be chosen in order to have at least 20 points per characteristic period of oscillation.

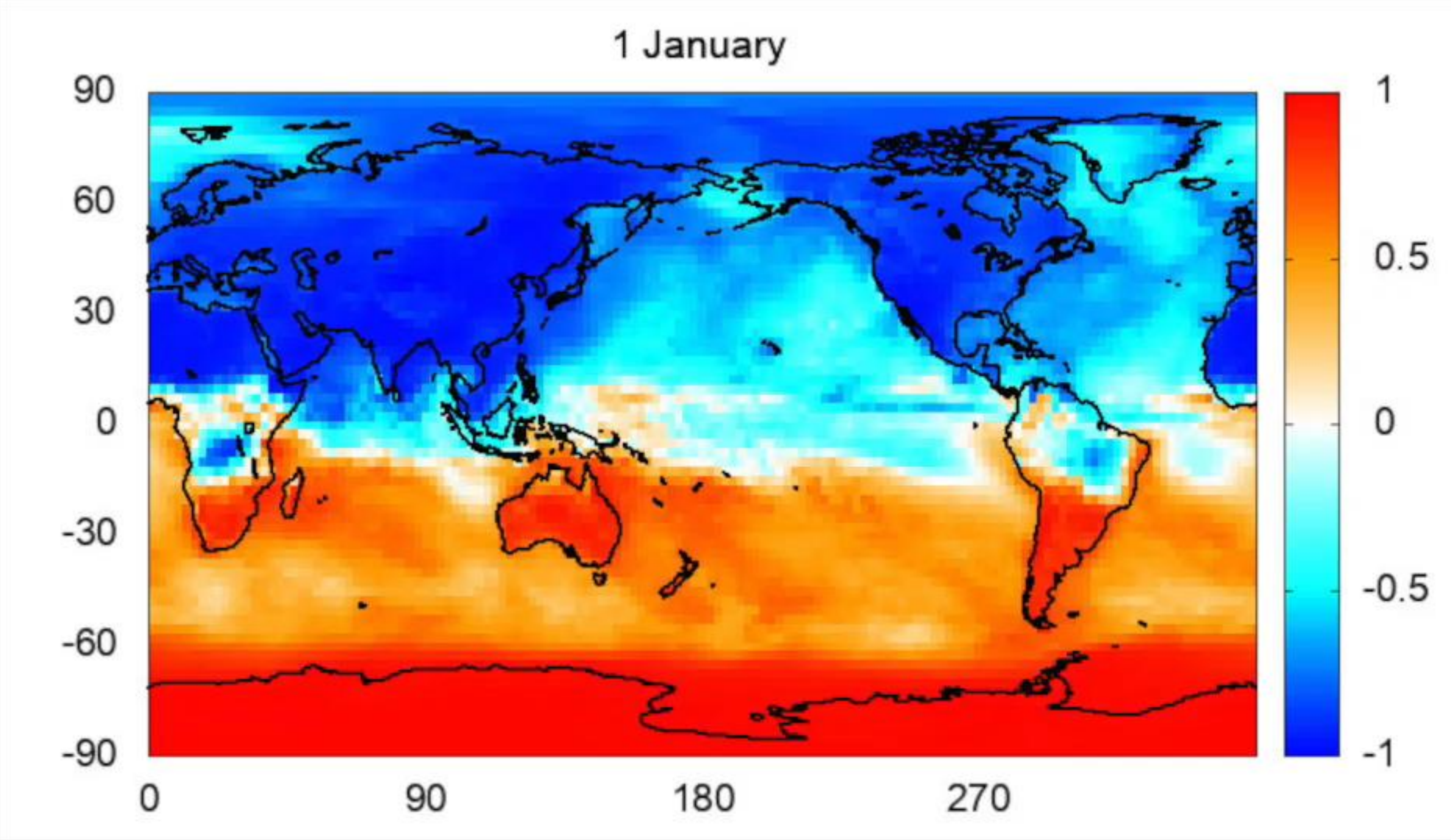
Application to climate data

- Can we use the Hilbert amplitude, phase, frequency, to :
  - Identify and quantify regional climate change?
  - Investigate synchronization in climate data?
- Problem: climate time series are not narrow-band.
- Usual solution (e.g. brain signals): isolate a narrow frequency band.
- However, the Hilbert transform applied to Surface Air Temperature time series yields meaningful insights.

Cosine of Hilbert phase



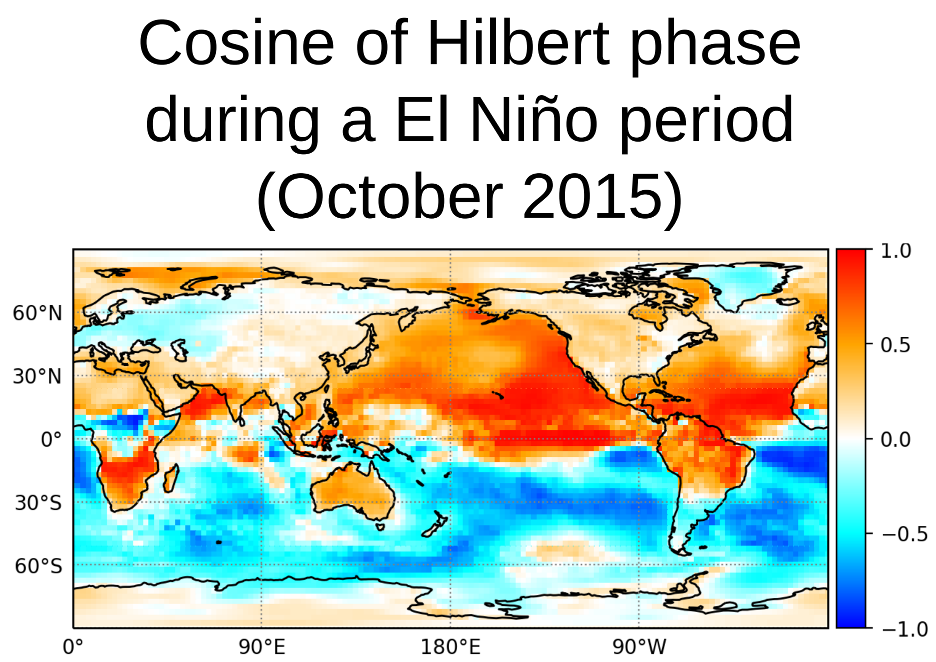
How do seasons evolve?
Temporal evolution of the cosine of the Hilbert phase

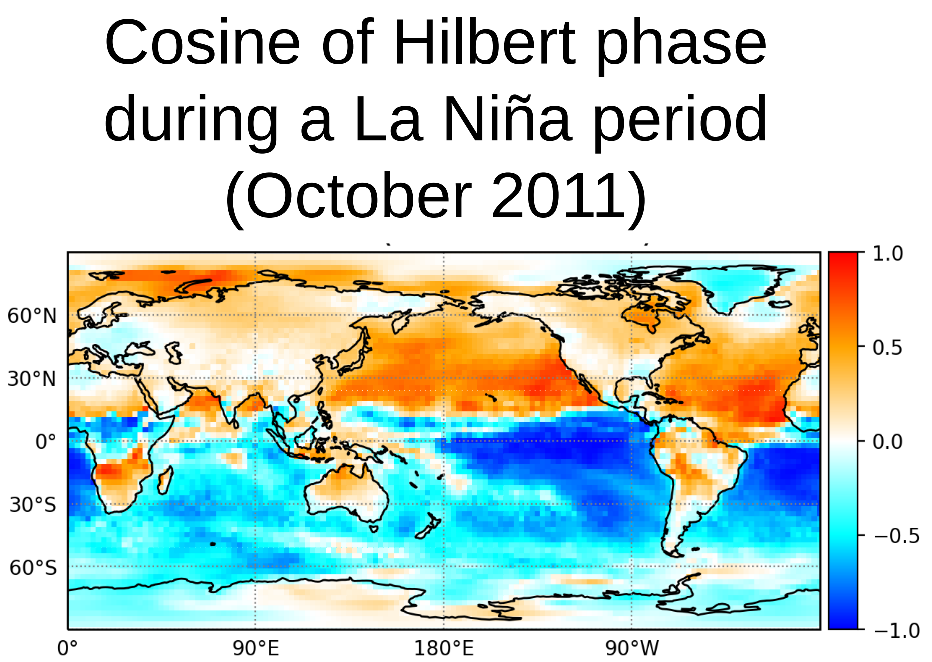


El Niño/La Niña-Southern Oscillation (ENSO)

Is the most important climate phenomena on the planet

- Occurs across the tropical Pacific Ocean with ≈ 3-6 years periodicity.
- Variations in the surface **temperature** of the tropical eastern Pacific Ocean (warming: El Niño, cooling: La Niña)
- Variations in the air surface **pressure** in the tropical western Pacific (the Southern Oscillation).
- These two variations are coupled:
  - El Niño (ocean warming) -- high air surface pressure,
  - La Niña (ocean cooling) -- low air surface pressure.


Cosine of Hilbert phase
during a El Niño period
(October 2015)
60°N
30°N
0°
30°S
60°S
0°
90°E
180°E
90°W
1.0
0.5
0.0
−0.5
−1.0


Cosine of Hilbert phase
during a La Niña period
(October 2011)
60°N
30°N
0°
30°S
60°S
0°
90°E
180°E
90°W
1.0
0.5
0.0
−0.5
−1.0

Changes in Hilbert amplitude and frequency detect inter-decadal variations in surface air temperature (SAT)

The data:

- Spatial resolution $2.5^0 \times 2.5^0 \Rightarrow$ 10226 time series
- Daily resolution 1979 – 2016 $\Rightarrow$ 13700 data points

Where does the data come from?

- European Centre for Medium-Range Weather Forecasts (ECMWF, ERA-Interim).
- Freely available.

"Features" extracted from each SAT time series

- Time averaged amplitude, $\langle a \rangle$
- Time averaged frequency, $\langle \omega \rangle$
- Standard deviations, σ_a, σ_ω

Relative decadal variations

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

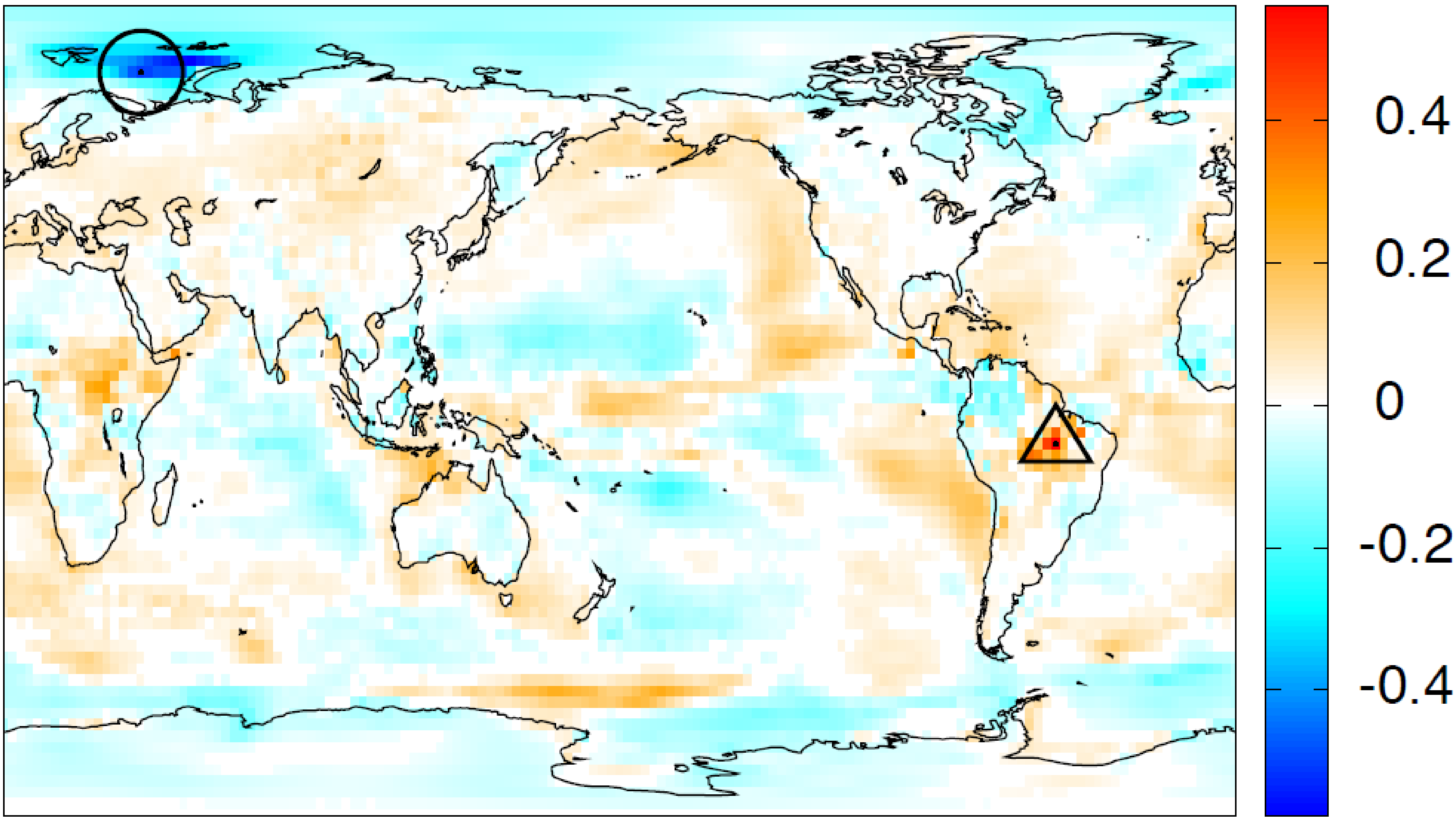
$$\frac{\Delta a}{\langle a \rangle_{2016-1979}}$$

Relative variation is considered significant if:

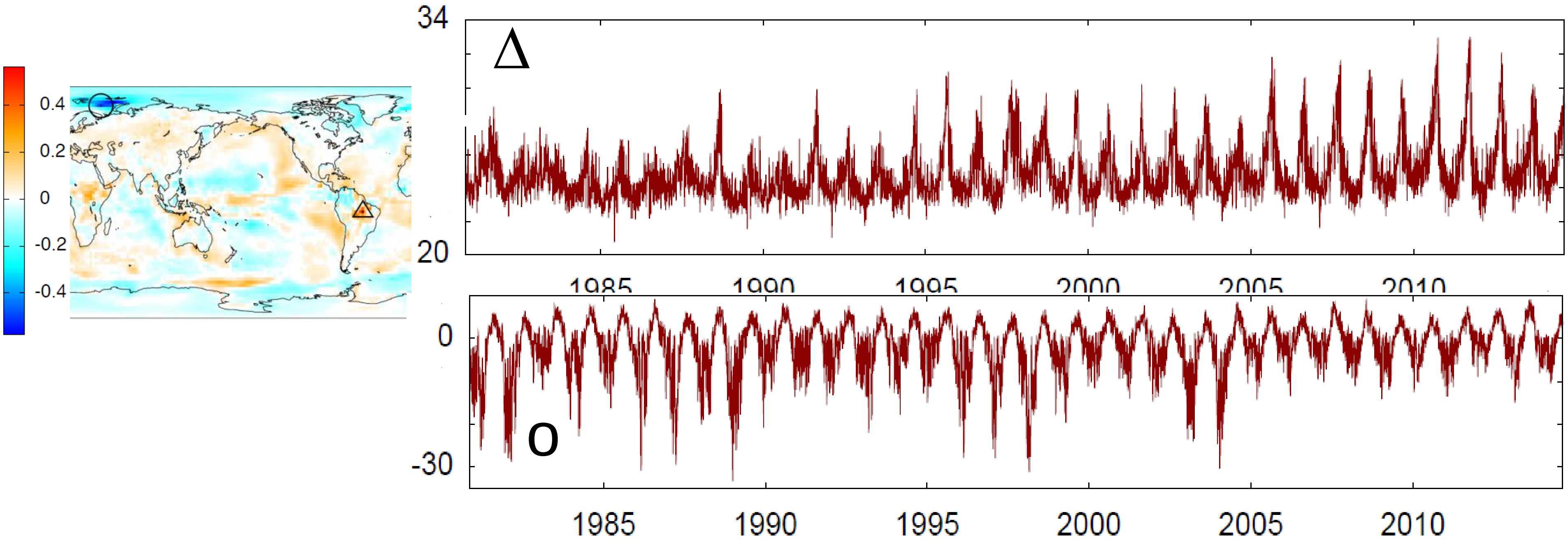
$$\frac{\Delta a}{\langle a \rangle} \geq \langle . \rangle_s + 2\sigma_s \quad \text{or} \quad \frac{\Delta a}{\langle a \rangle} \leq \langle . \rangle_s - 2\sigma_s$$

100 “block” surrogates

D. A. Zappala, M. Barreiro and C. Masoller, “*Quantifying changes in spatial patterns of surface air temperature dynamics over several decades*”, Earth Syst. Dynam. **9**, 383 (2018)

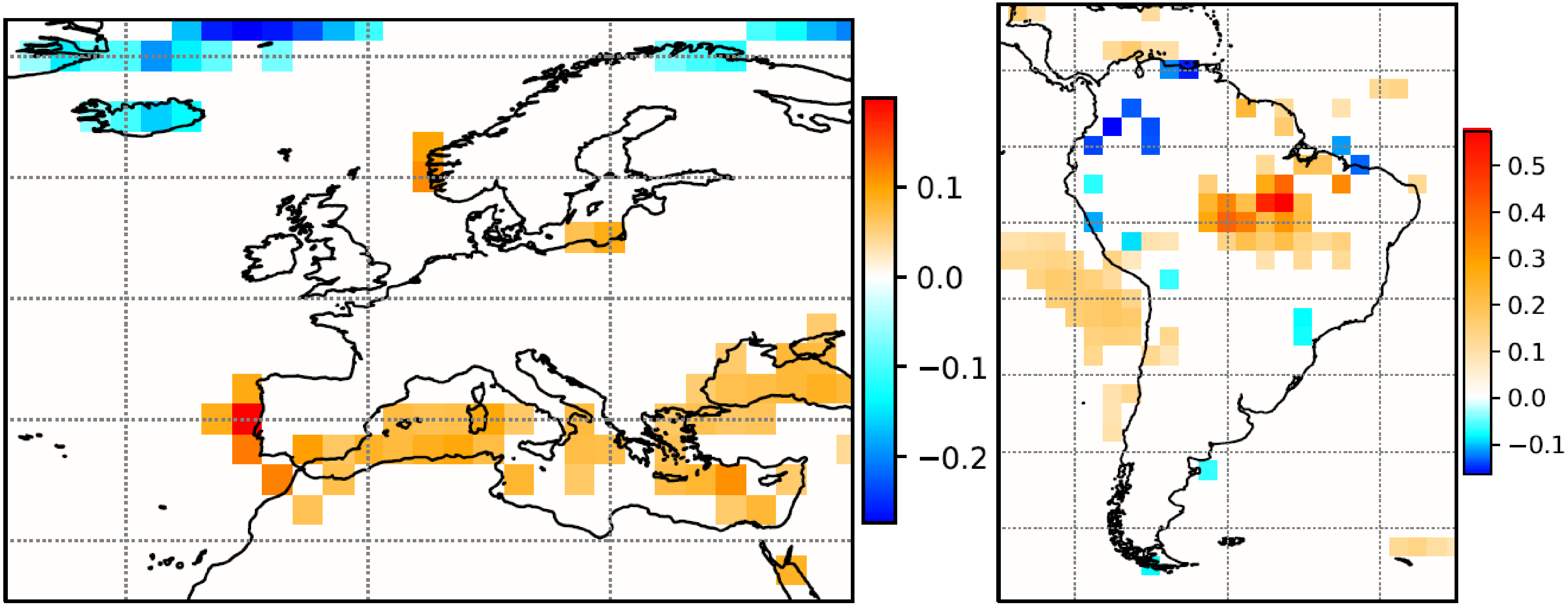

0.4
0.2
0
-0.2
-0.4

Relative variation of average Hilbert amplitude uncovers regions where the amplitude of the seasonal cycle increased or decreased



- **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.

D. A. Zappala et al., Earth Syst. Dynam. **9**, 383 (2018)



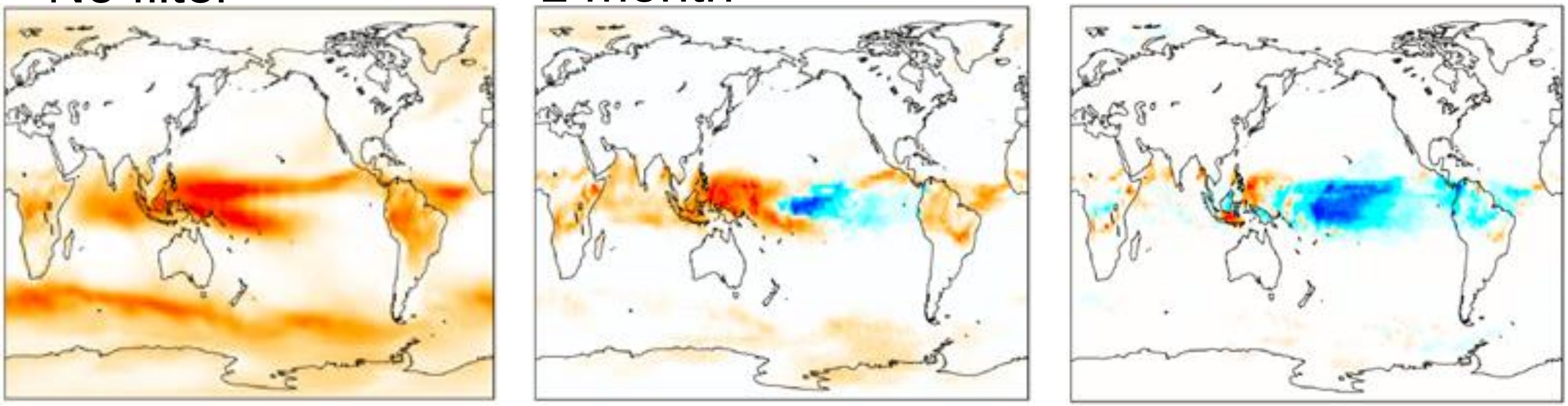
D. A. Zappala et al., Earth Syst. Dynam. **9**, 383 (2018)

Influence of pre-processing SAT time series by temporal averaging in a moving window: fast variability removed

SAT → average in a time window → Hilbert

The color-code shows the mean frequency (red fast, blue slow)

No filter 1 month 3 months



Take home messages

- Symbolic analysis, network representation, Hilbert analysis and many others, are useful tools for investigating complex signals.

- Different techniques provide *complementary* information.

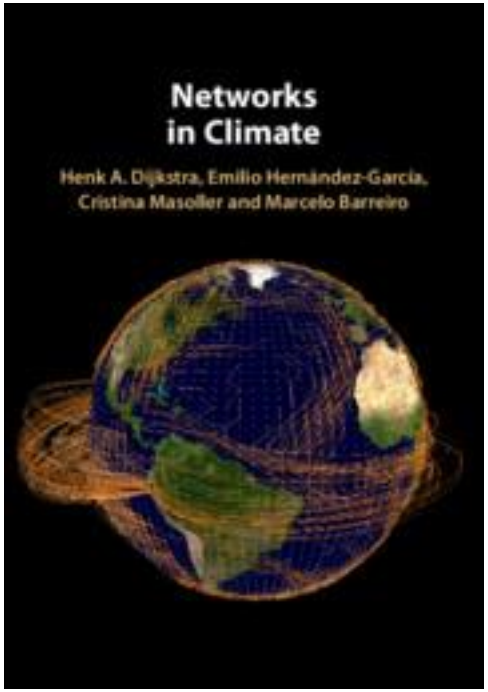
"…nonlinear time-series analysis has been used to great advantage on thousands of real and synthetic data sets from a wide variety of systems ranging from roulette wheels to lasers to the human heart. Even in cases where the data do not meet the mathematical or algorithmic requirements, the results of nonlinear time-series analysis can be helpful in understanding, characterizing, and predicting dynamical systems…"

Bradley and Kantz, CHAOS 25, 097610 (2015)

References

- Bandt and Pompe, PRL 88, 174102 (2002)
- U. Parlitz et al., Computers in Biology and Medicine 42, 319 (2012)

- C. Masoller et al, NJP 17, 023068 (2015)
- A. Aragoneses et al, PRL 116, 033902 (2016)
- Panozzo et al, Chaos 27, 114315 (2017)
- Zappala, Barreiro and Masoller, Entropy 18, 408 (2016)
- Zappala, Barreiro and Masoller, Earth Syst. Dynam. 9, 383 (2018)
- Zappala, Barreiro and C. Masoller, Chaos 29, 051101 (2019)



Cambridge University Press 2019

Outline

- **Introduction**
  - Historical developments: from dynamical systems to complex systems

- **Univariate analysis**
  - Symbolic & network-based tools.
  - Applications.

- **Bivariate analysis**
  - Correlation, mutual information and directionality.
  - Applications.

- **Multivariate analysis**
  - Complex networks.
  - Network characterization and analysis.
  - Climate networks.

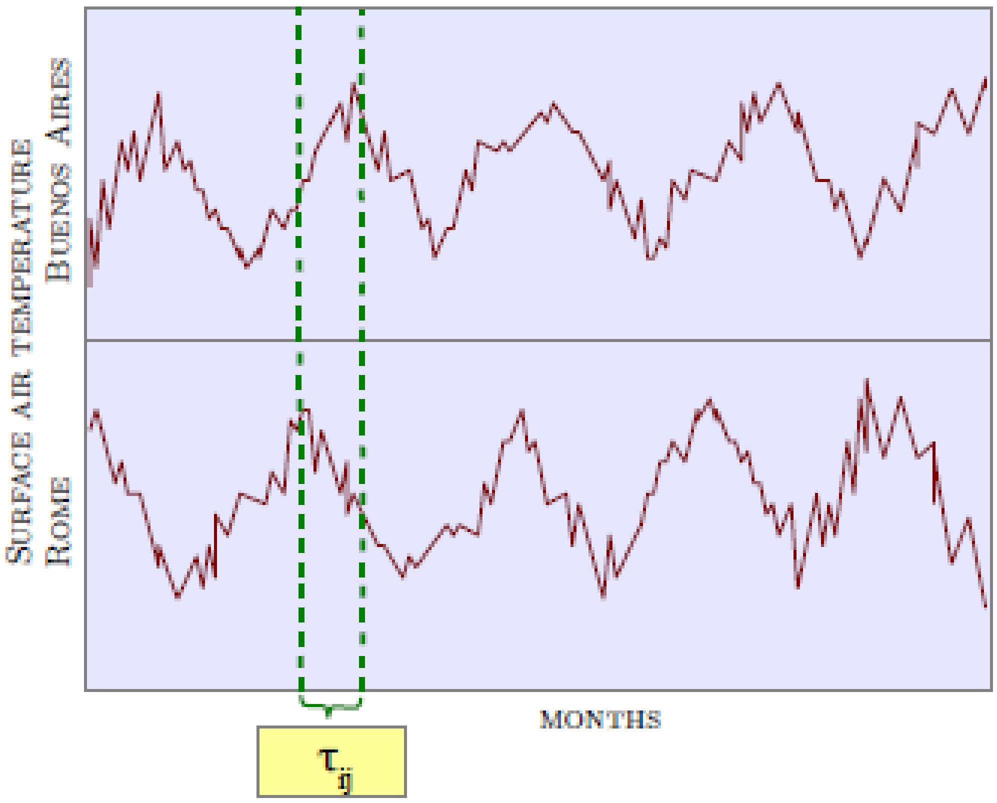
Cross-correlation of two time series *X* and *Y* of length *N*

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

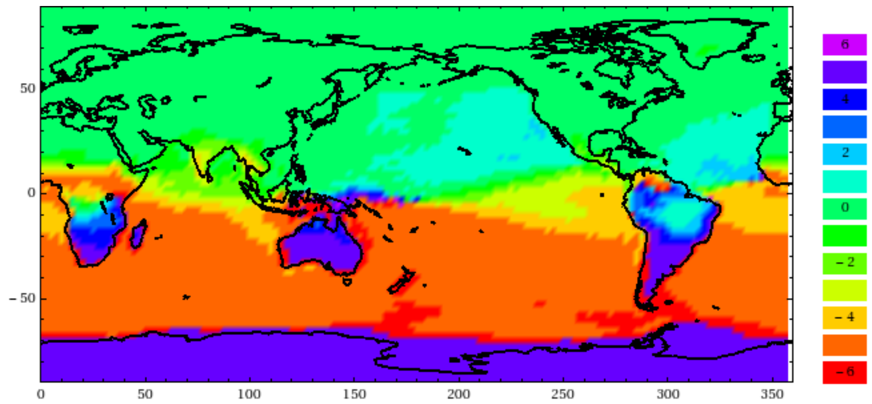
the two time series are normalized to zero-mean **μ=0** and unit variance, **σ=1**

- $-1 \leq C_{X,Y} \leq 1$
- $C_{X,Y} = C_{Y,X}$
- The maximum of $C_{X,Y}(\tau)$ indicates the **lag** that renders the time series *X* and *Y* best aligned.
- Pearson coefficient: $\rho = C_{X,Y}(0)$

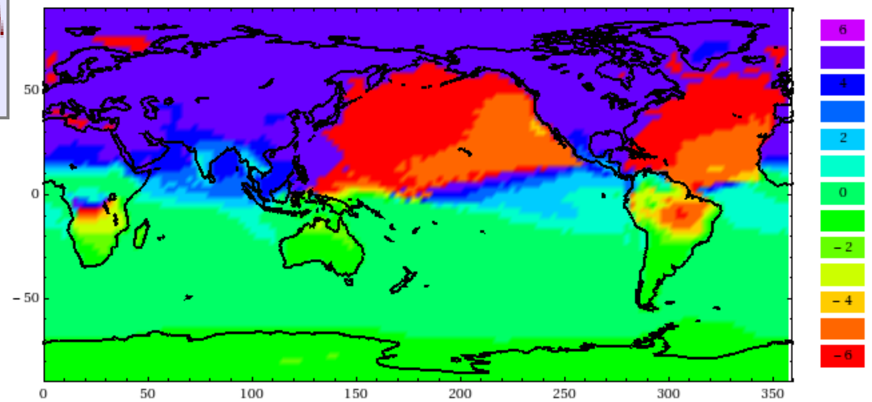
Correlation analysis of lag-times between seasonal cycles (Surface Air Temperature)



Rome

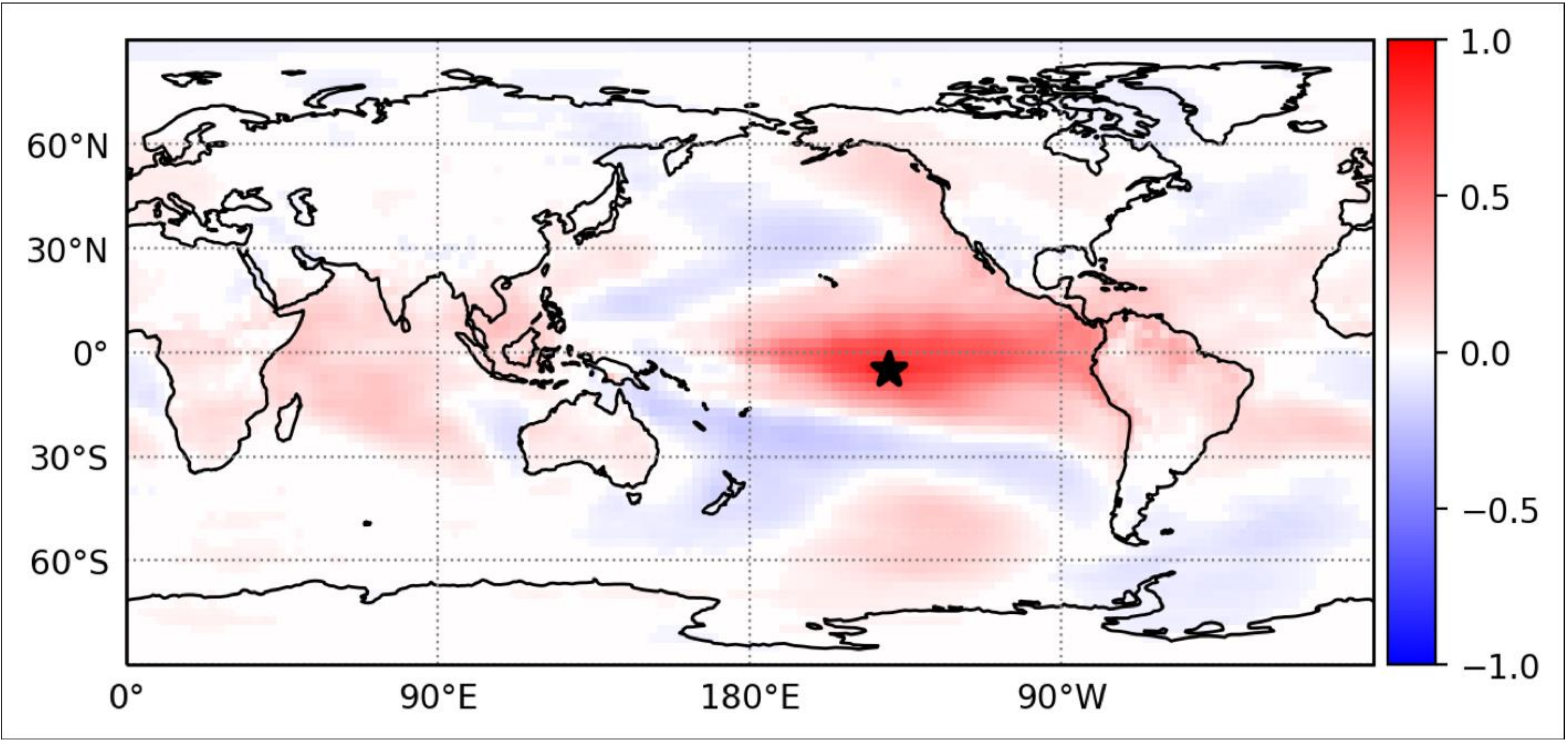


Buenos Aires



G. Tirabassi and C. Masoller, Sci. Rep. 6:29804 (2016)

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

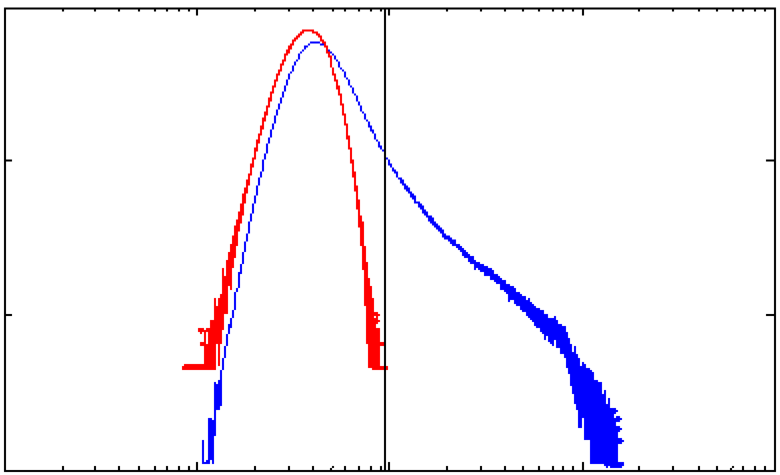


Color code represents the Pearson coefficient of ∗ and all the world

Are cross-correlation values significant?

Surrogate analysis needed.

Prob.
Distrib.
Function
(PDF)



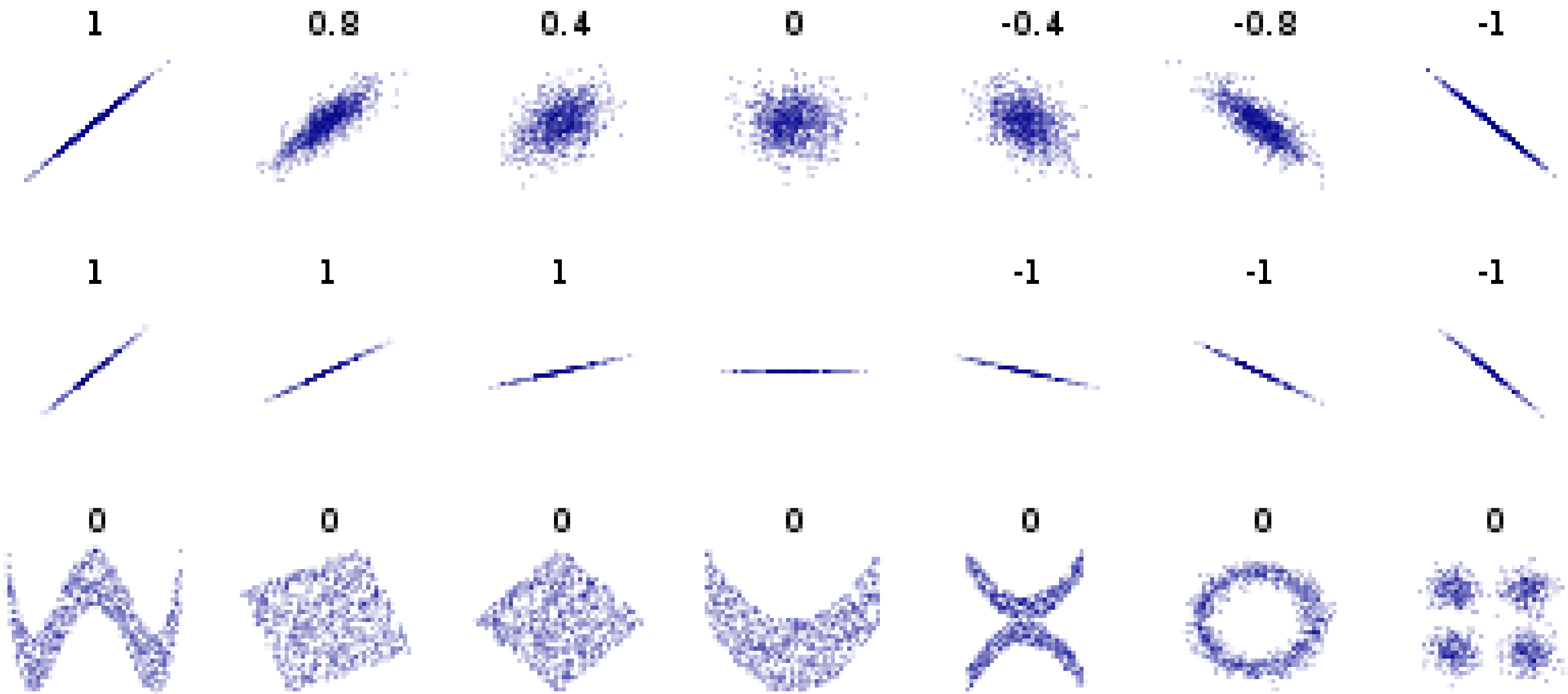
|Cross correlation|

Problems:

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

G. Lancaster et al, Physics Reports 748 (2018) 1–60

Cross-correlation analysis detects linear relationships only



Source: wikipedia

Nonlinear correlation measure based on information theory: the 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)*

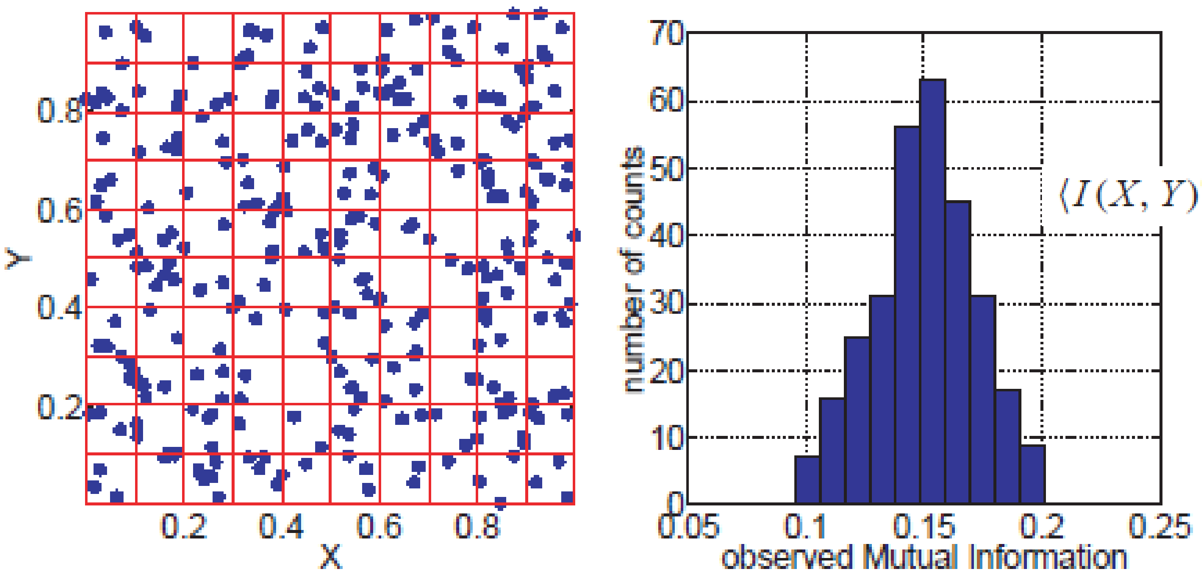
- $p(x,y) = p(x)\,p(y) \Rightarrow MI = 0$, else ***MI >0***

- *MI* can also be computed with a lag-time.
- If *p(x,y)* is a bivariate Gaussian distribution, then

  $$MI = -1/2 \log(1-\rho^2)$$

  where ρ is the Pearson coefficient.

MI values are systematically overestimated



$$\langle I(X, Y)^{estimated} \rangle \approx 0.15 \pm 0.02$$

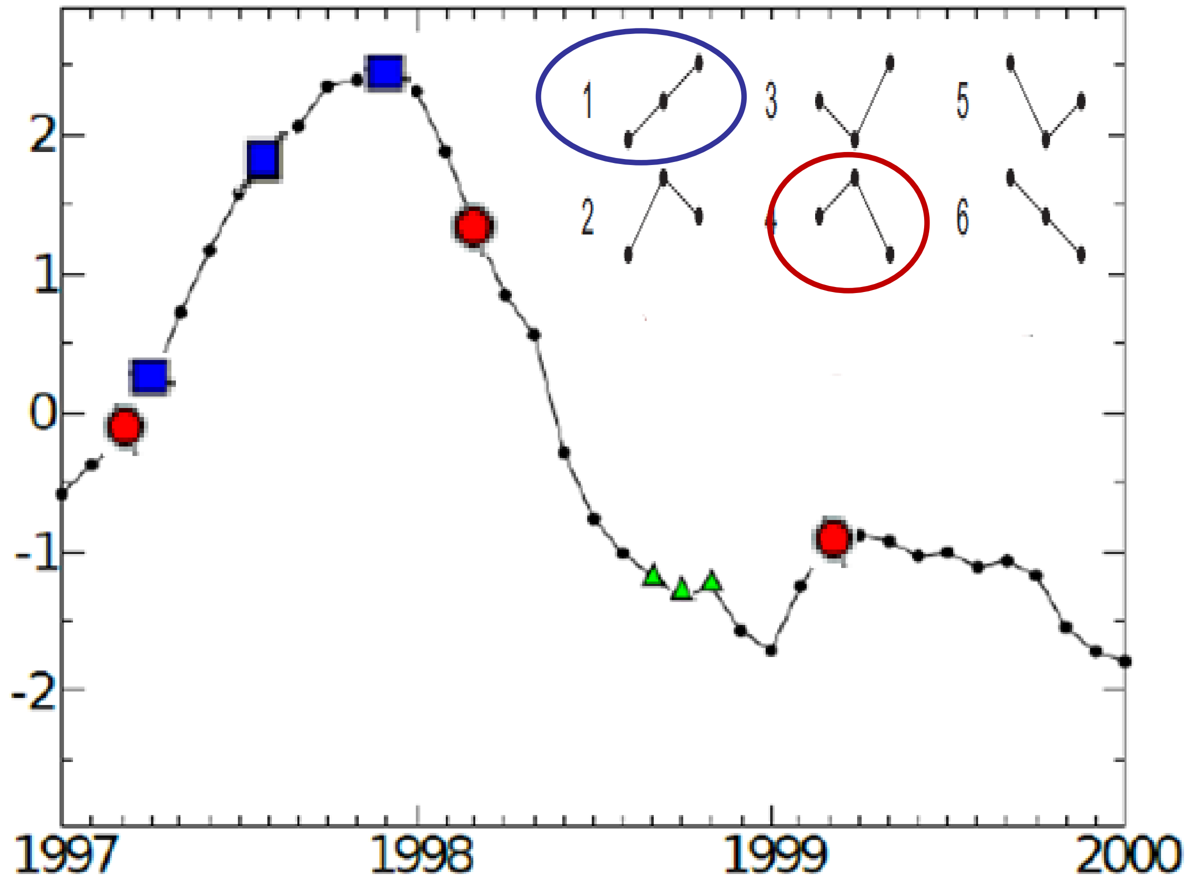
Main problem: a reliable estimation of MI requires a large amount of data

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.

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

The Mutual Information can be computed from the probabilities of ordinal patterns: the lag allows to select the time scale of the analysis

- **Green** triangles: intra-seasonal pattern,
- **blue** squares: intra-annual pattern
- **red** circles: inter-annual pattern



Example: el Niño index, monthly sampled

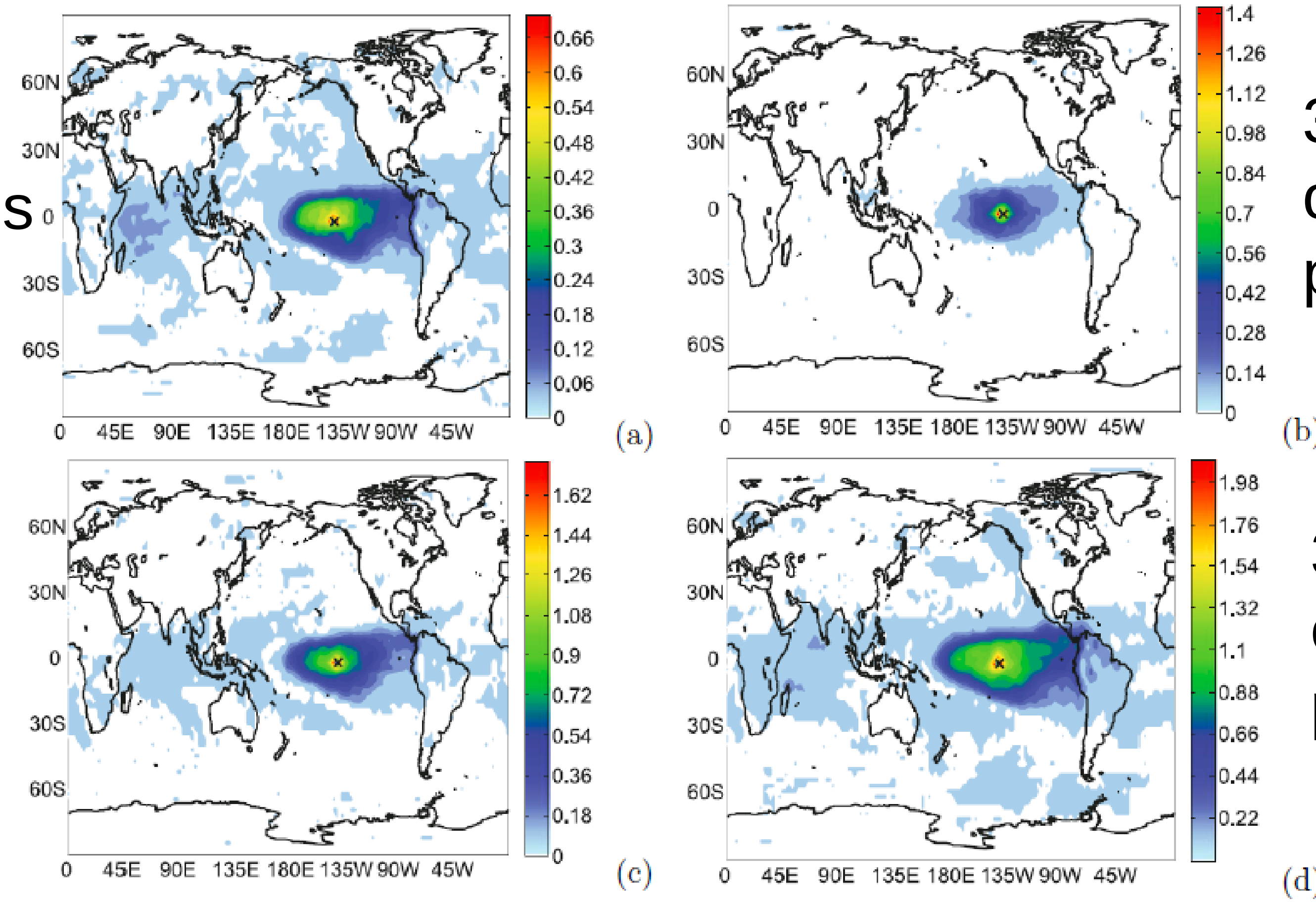
Mutual information maps: SAT anomalies at El Niño

Histograms

3 months ordinal patterns

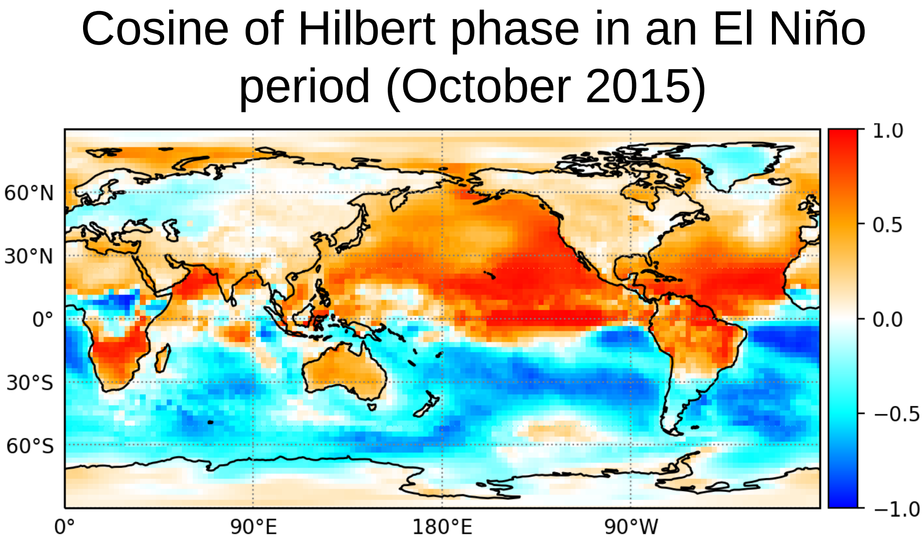
Inter-annual ordinal patterns

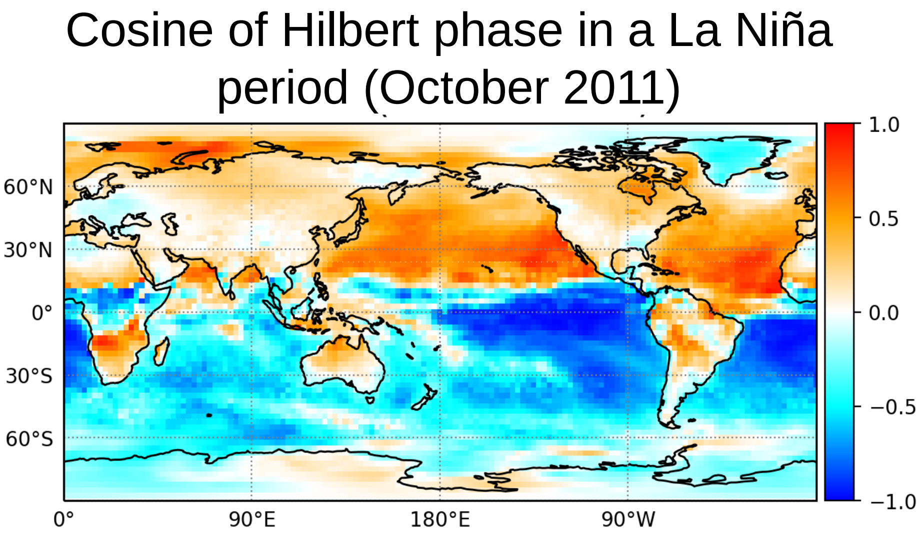
3 years ordinal patterns



Ordinal analysis separates the times-scales of the interactions

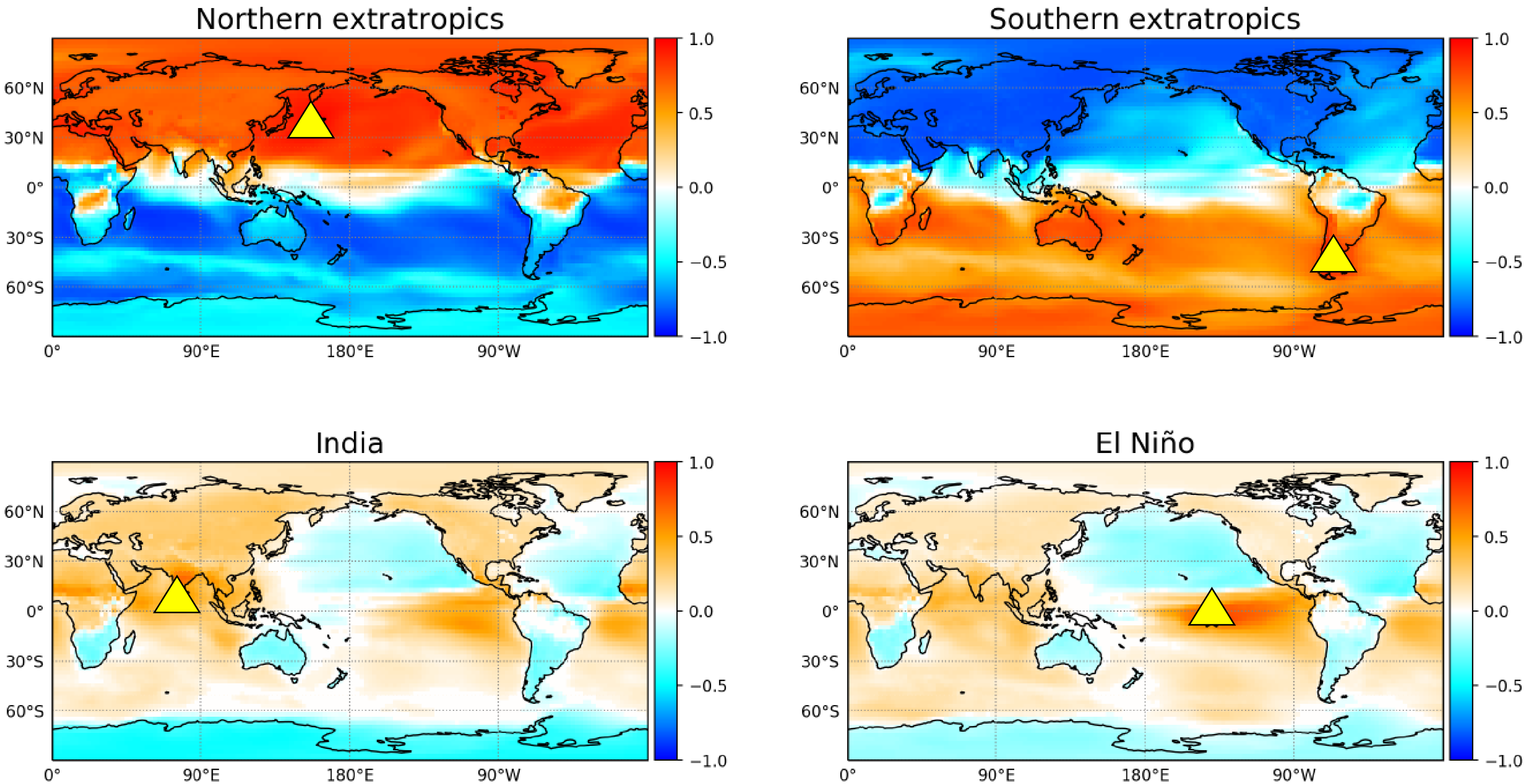
Deza, Barreiro and Masoller, Eur. Phys. J. ST 222, 511 (2013)





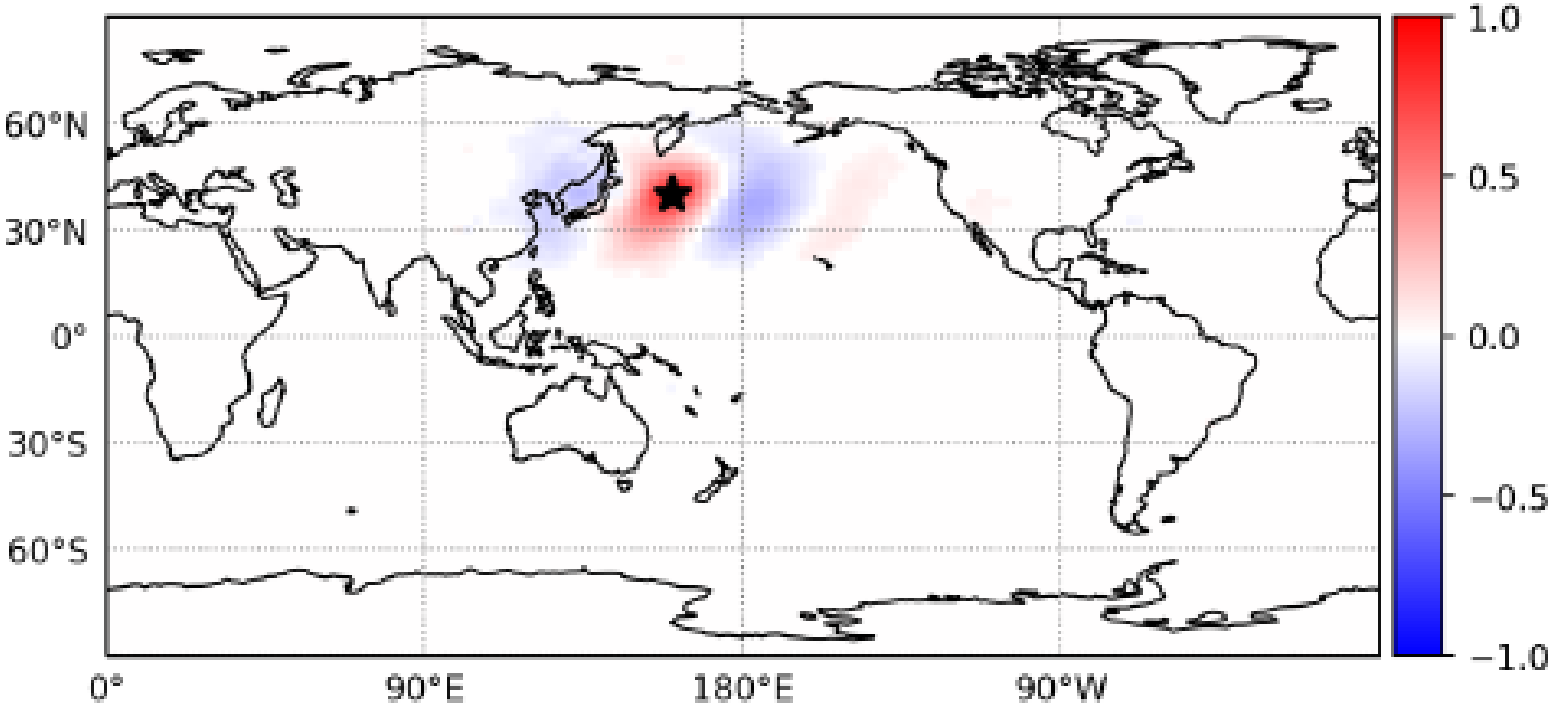
Cross correlation analysis of instantaneous phases, amplitudes and frequencies

Cross-correlation of cosine of Hilbert phase



Cross-correlation analysis of Hilbert frequencies identifies Rossby waves

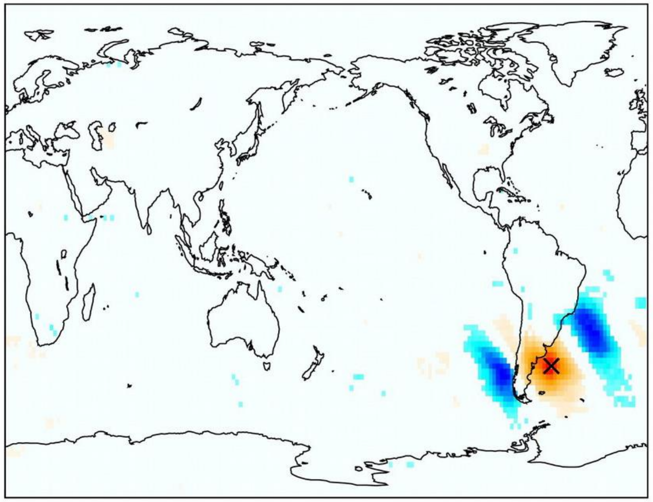
Cross-correlation in color code



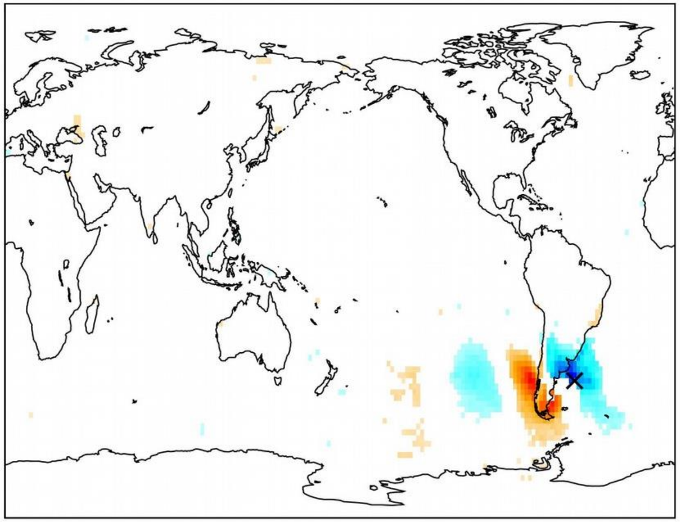
D. A. Zappala, M. Barreiro and C. Masoller, "*Quantifying phase synchronization and unveiling Rossby wave patterns in surface air temperature dynamics*", submitted (2019)

Lagged-cross correlation

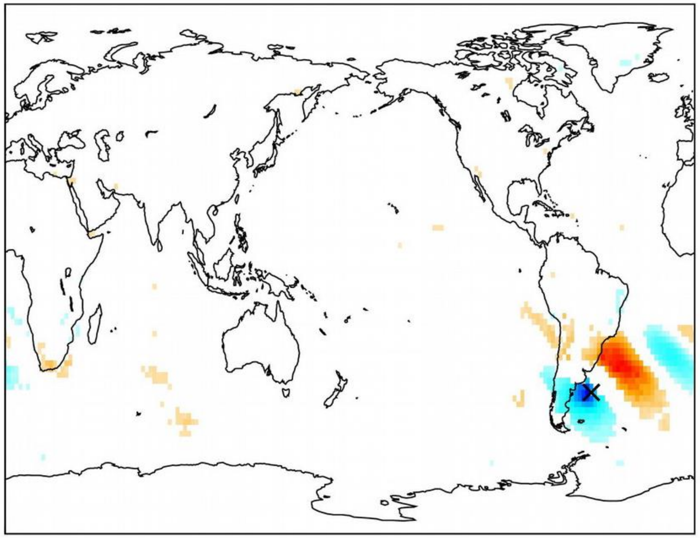
τ=0



τ= - 2 days



τ= + 2 days

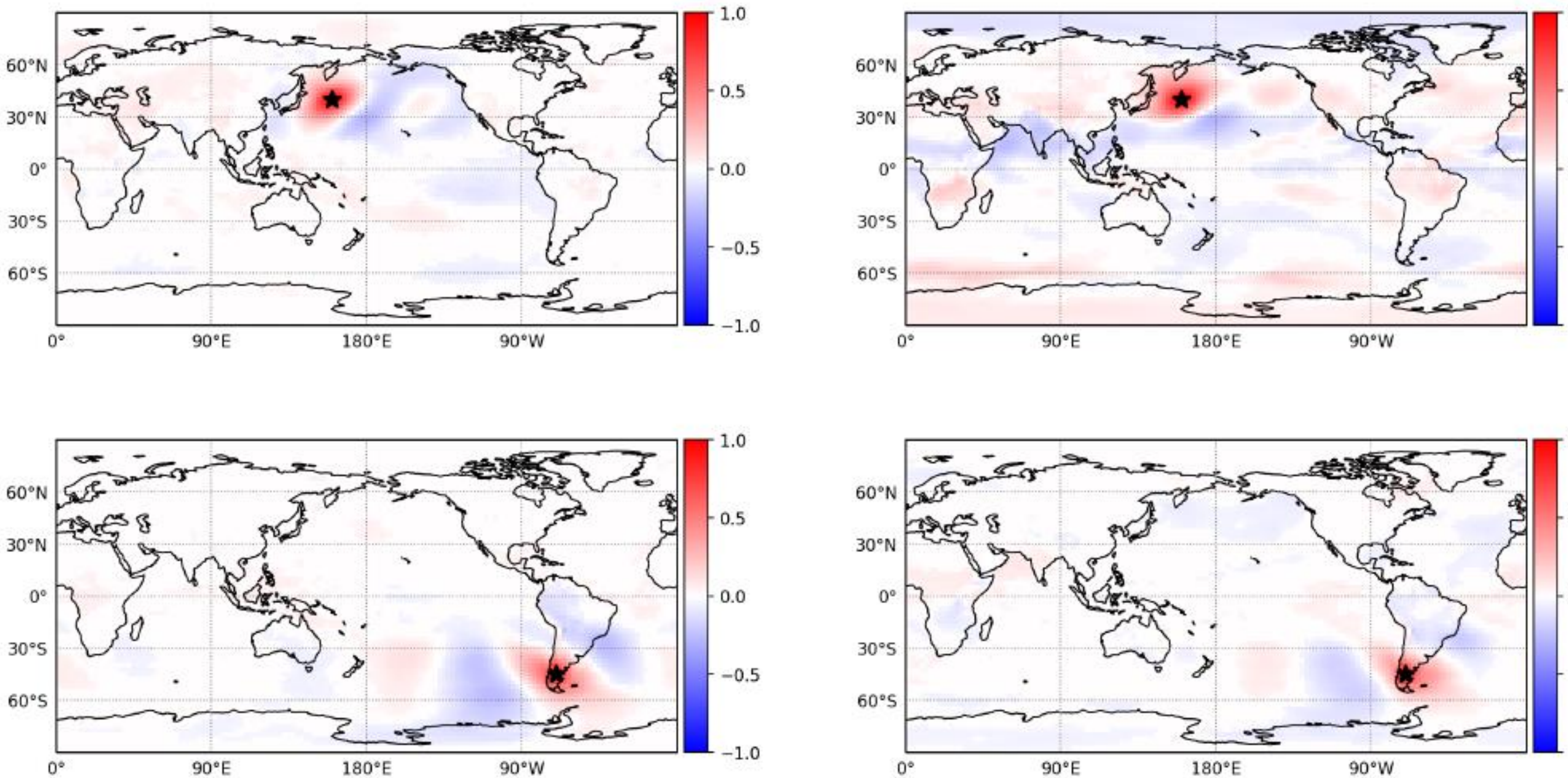


As expected, the wave pattern moves towards east

Clean wave pattern not seen in cross-correlation analysis of Hilbert amplitudes or anomalies

Cross-correlation in color code.

Anomalies **Hilbert amplitudes**



D. A. Zappala, M. Barreiro and C. Masoller, "*Quantifying phase synchronization and unveiling Rossby wave patterns in surface air temperature dynamics*", submitted (2019)

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)*.

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

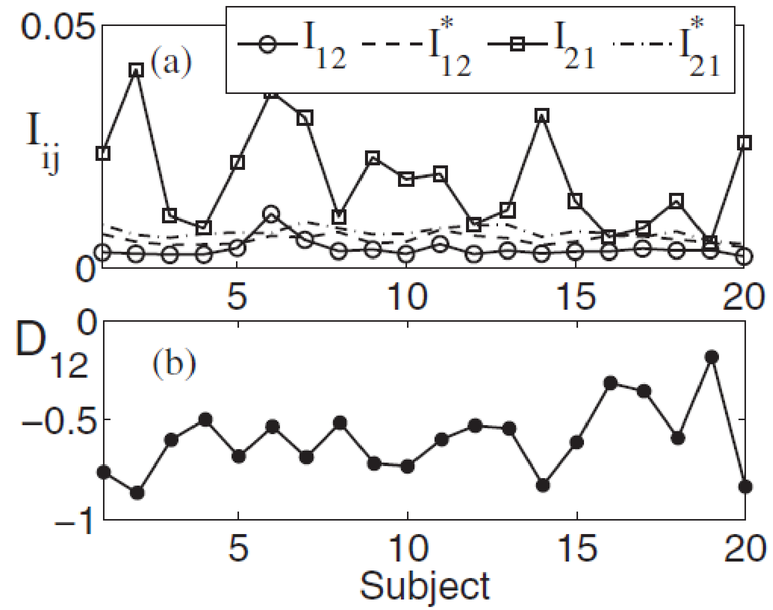
Directionality index

- 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.

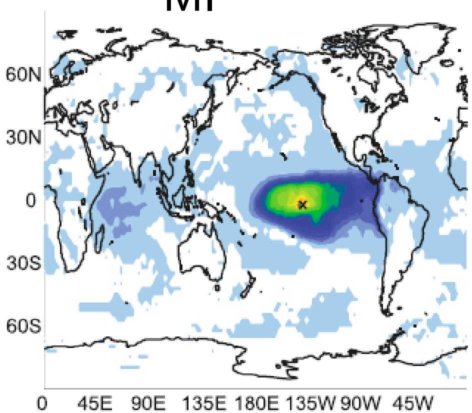
A. Bahraminasab et al., PRL 100, 084101 (2008)

Application to climate data

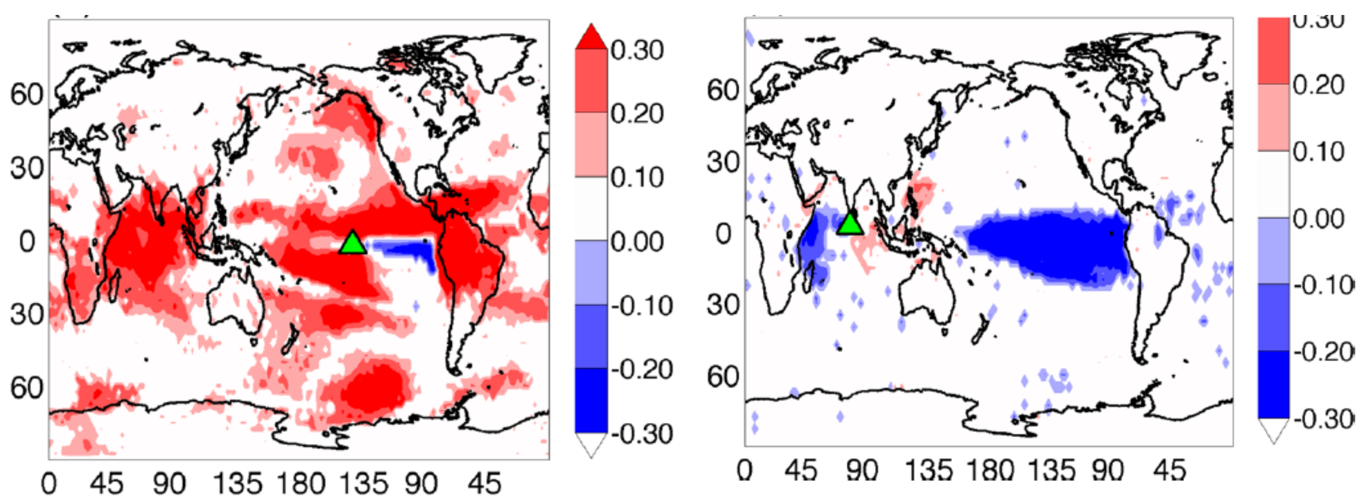
DI computed from daily SAT anomalies, PDFs estimated from histograms of values.

MI and DI are *both significant* ($>3\sigma$, bootstrap surrogates), τ=30 days.

MI

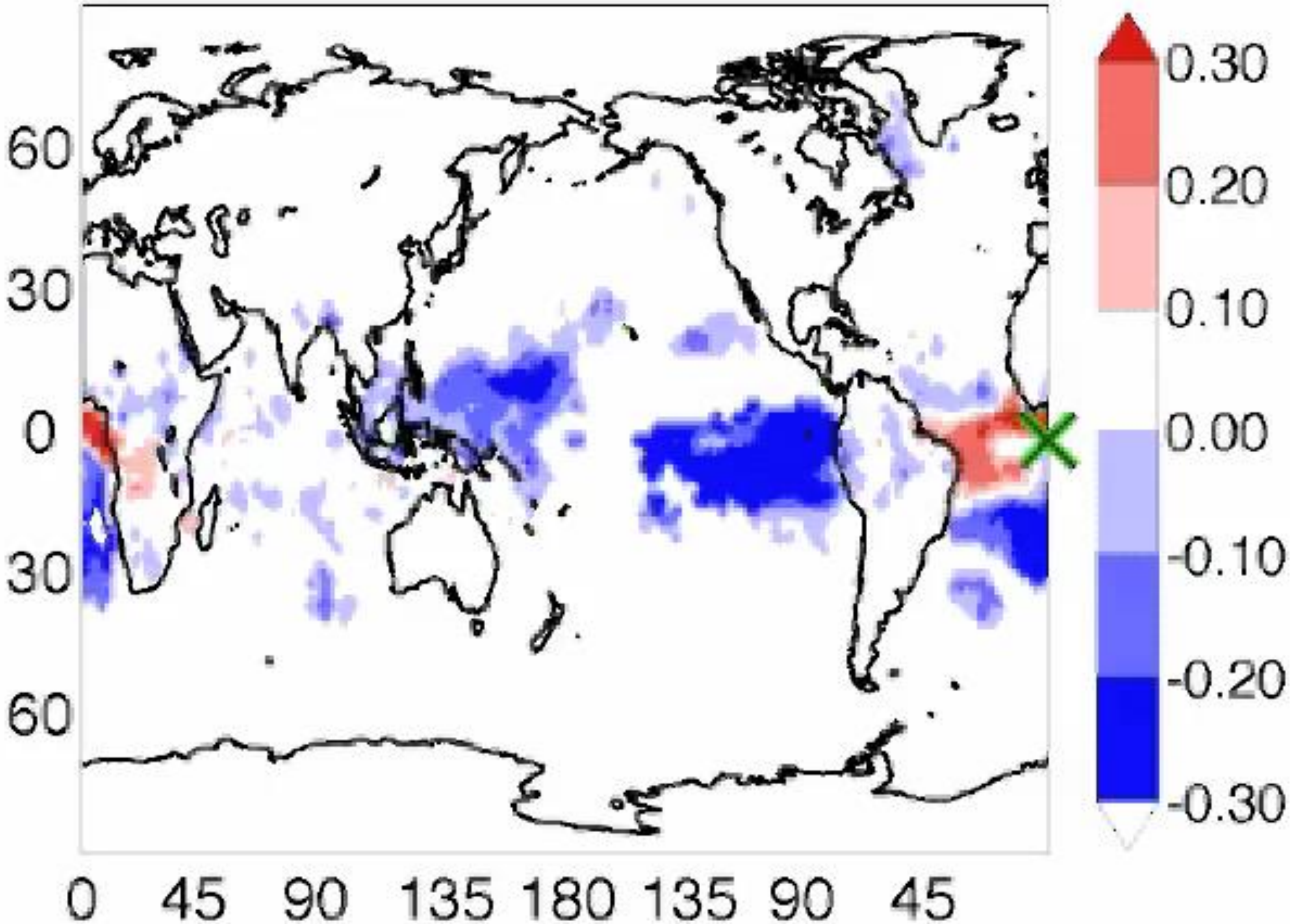


DI



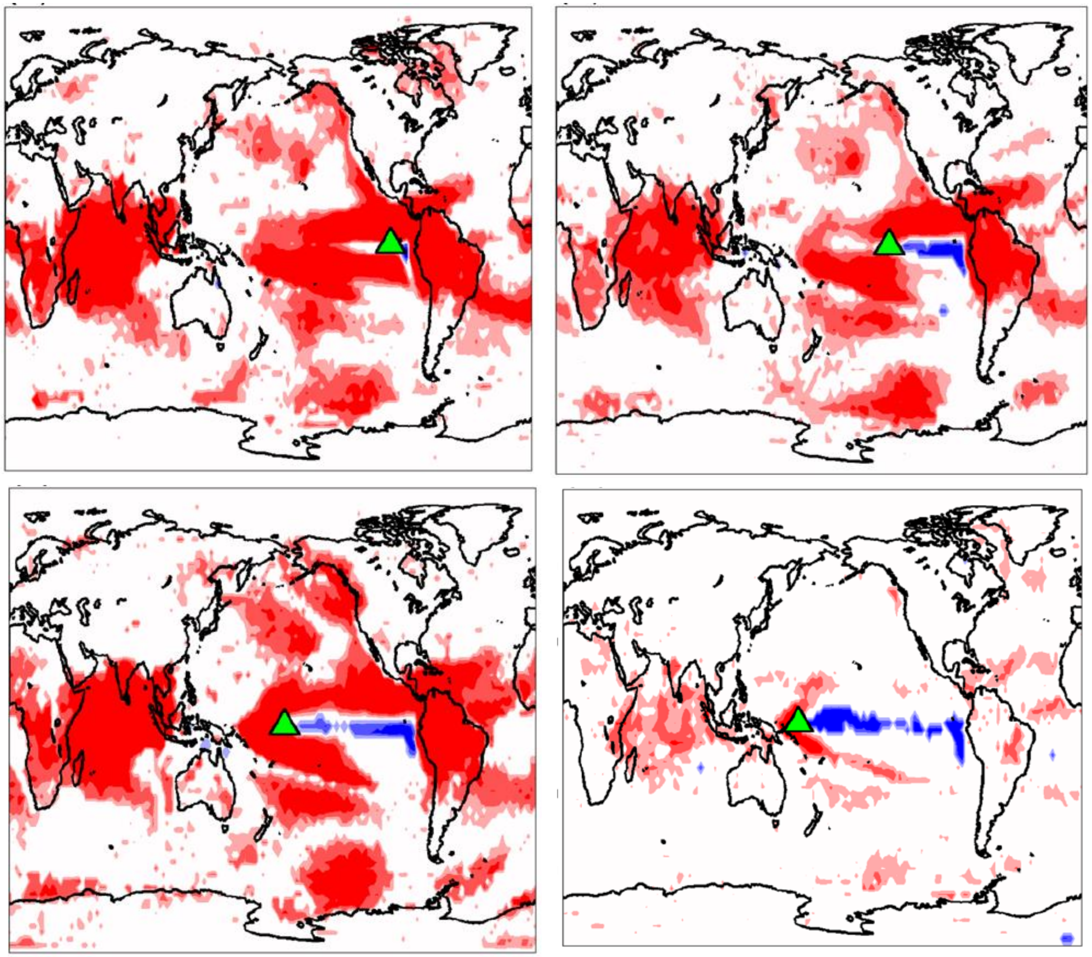
J. I. Deza, M. Barreiro, and C. Masoller, "*Assessing the direction of climate interactions by means of complex networks and information theoretic tools*", Chaos 25, 033105 (2015).

Directionality index along the equator



[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)

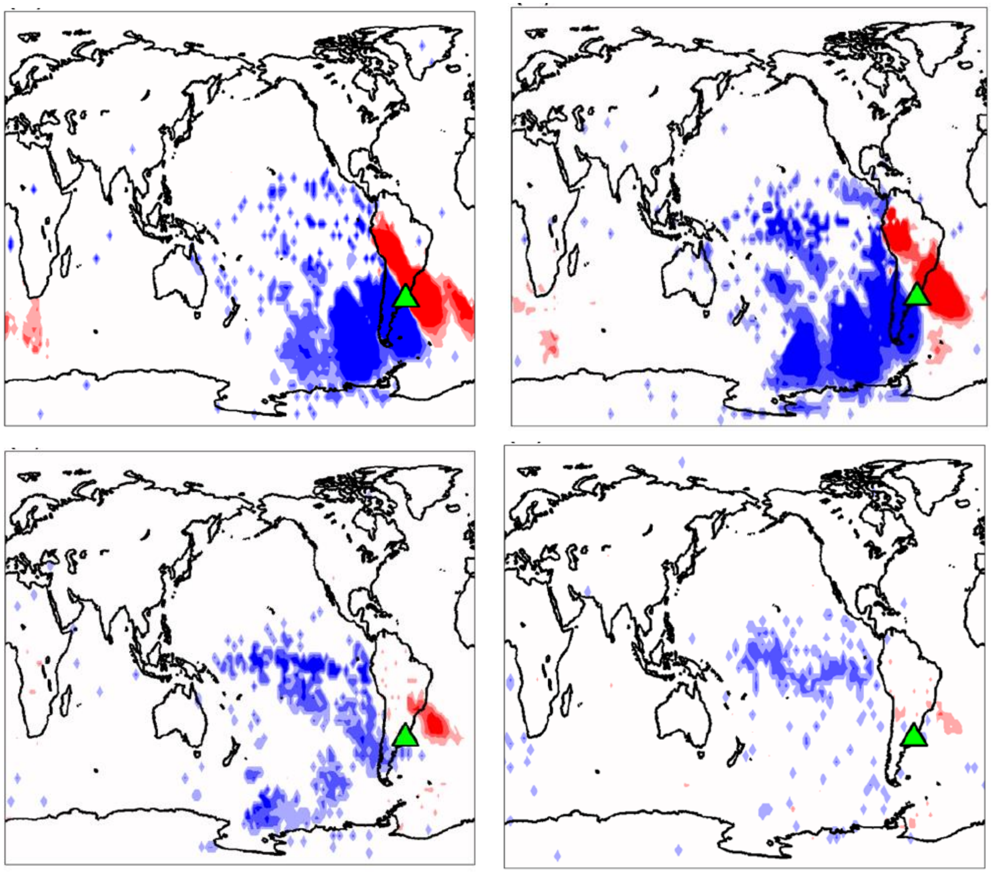
Influence of the time-scale of information transfer

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)

Summary

- Cross-correlation: detects linear interdependencies.
- Mutual information (MI): can detect some nonlinear interdependencies.
- When the MI computed from the probabilities of ordinal patterns, using a lag allows to select the time-scale of the analysis.
- The directionality index detects the net direction of the information flow.

References

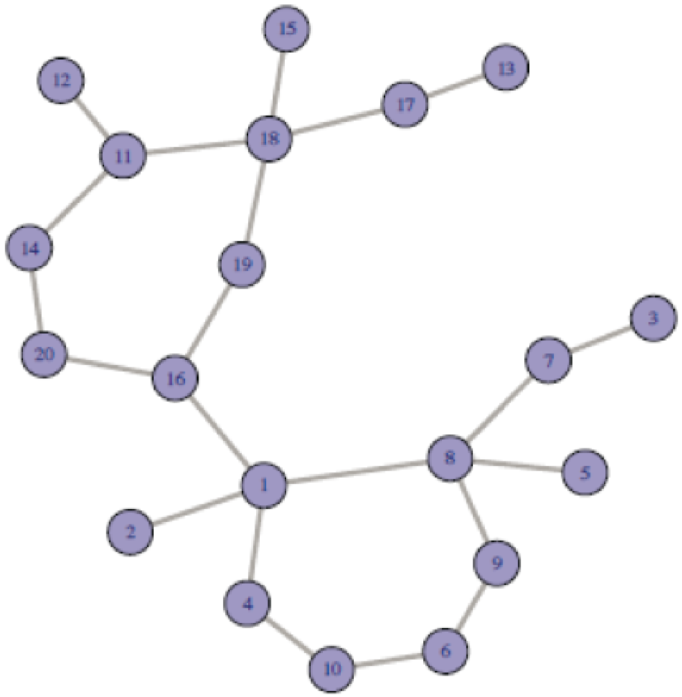
- M. Barreiro, et. al, Chaos 21, 013101 (2011)
- Deza, Barreiro and Masoller, Eur. Phys. J. ST 222, 511 (2013)
- Deza, Barreiro and Masoller, Chaos 25, 033105 (2015)
- Zappala, Barreiro and C. Masoller, Chaos 29, 051101 (2019)

Outline

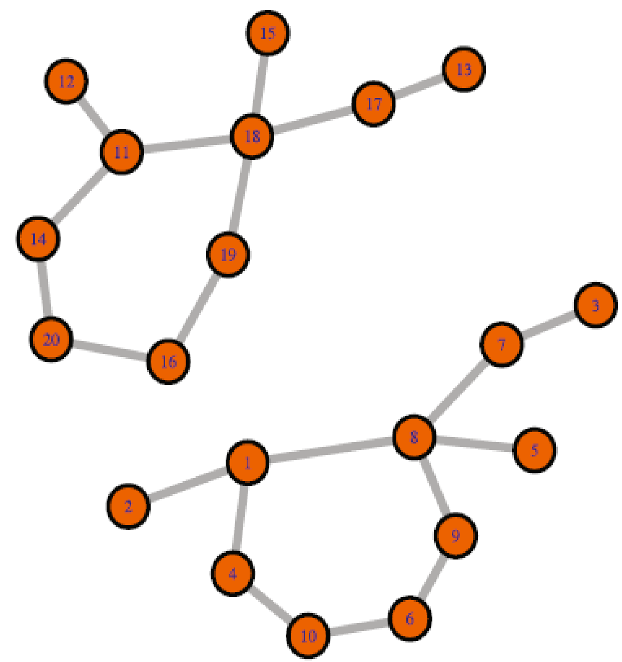
- **Introduction**
  - Historical developments: from dynamical systems to complex systems
- **Univariate analysis**
  - Symbolic & network-based tools.
  - Applications.
- **Bivariate analysis**
  - Correlation, mutual information and directionality.
  - Applications.
- **Multivariate analysis**
  - Complex networks.
  - Network characterization and analysis.
  - Climate networks.

Problem: how to infer the underlying interactions ("links") from observed signals at a set of units ("nodes") using bivariate similarity measures

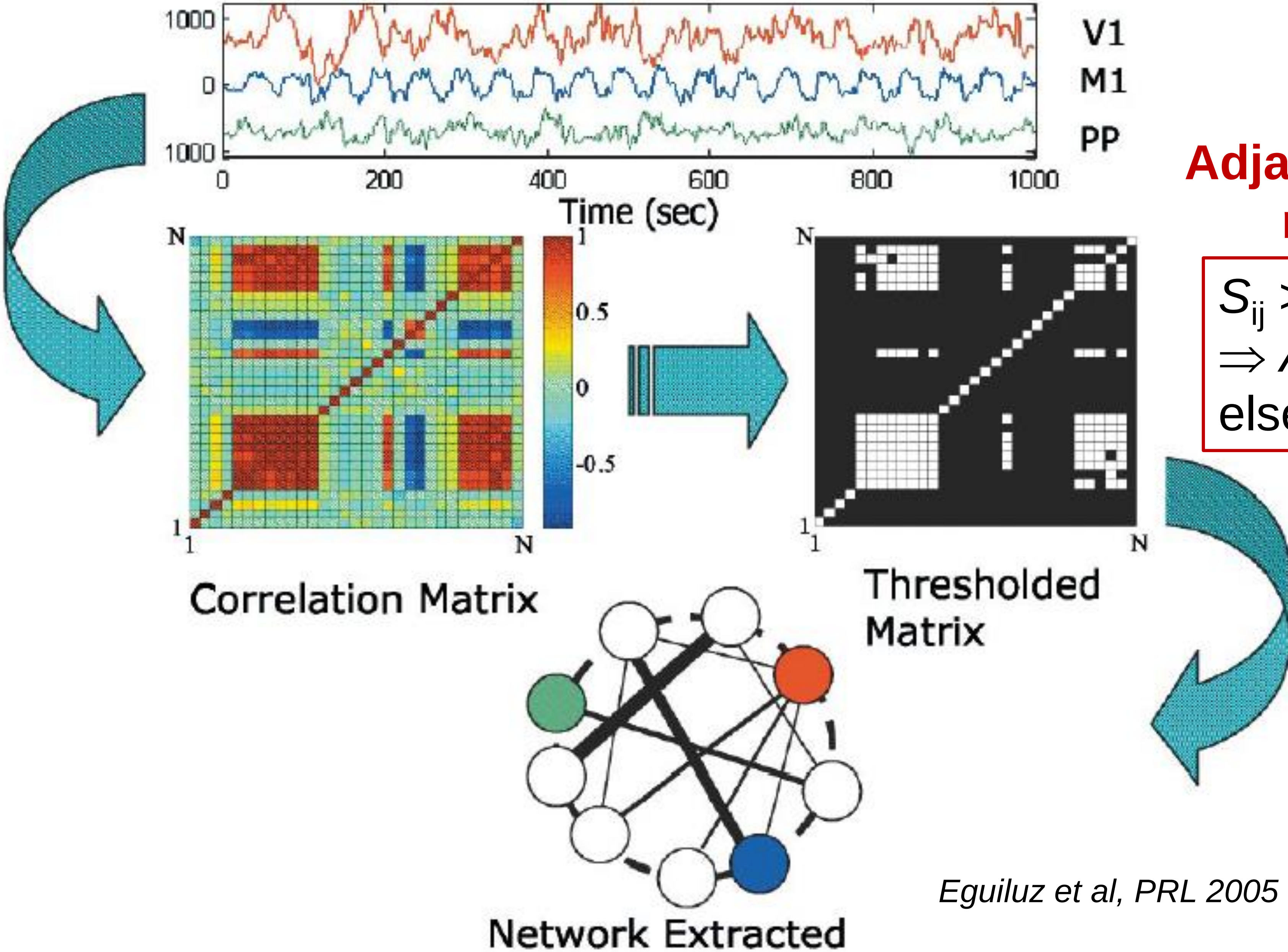
Structural (or real) network



Inferred correlation (or functional) network



Brain functional network



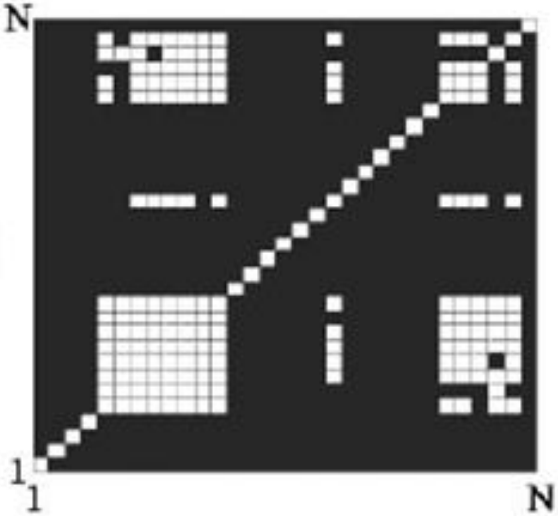
Adjacency matrix

$S_{ij} > Th$
$\Rightarrow A_{ij} = 1,$
else $A_{ij} = 0$

Eguiluz et al, PRL 2005

Graphical representation

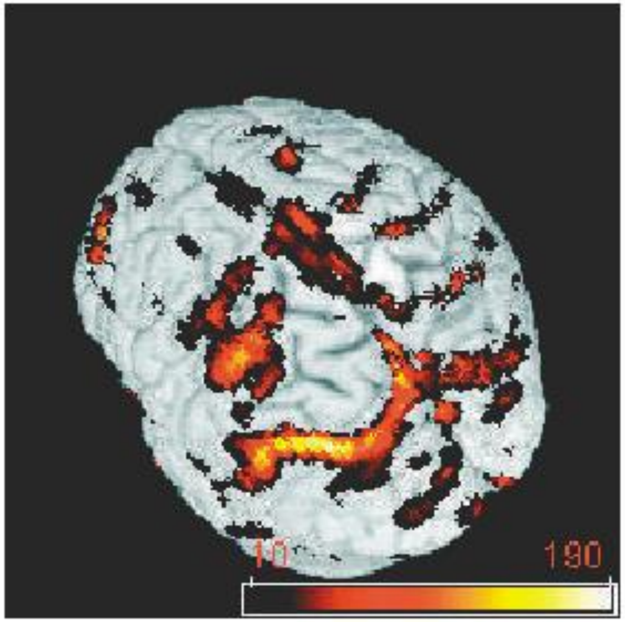
Adjacency matrix



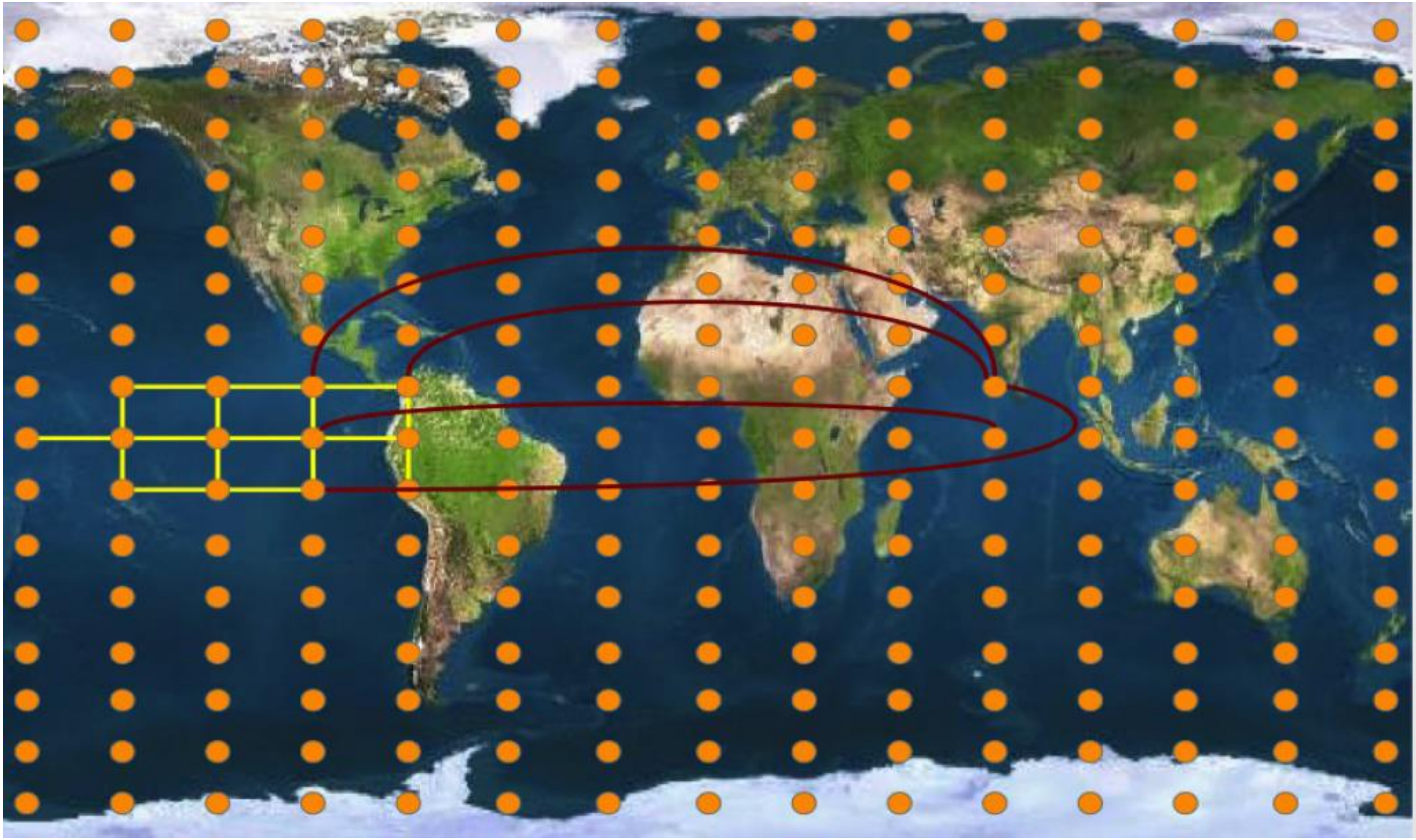
Thresholded matrix = inferred ("functional") network

Degree of a node: number of links

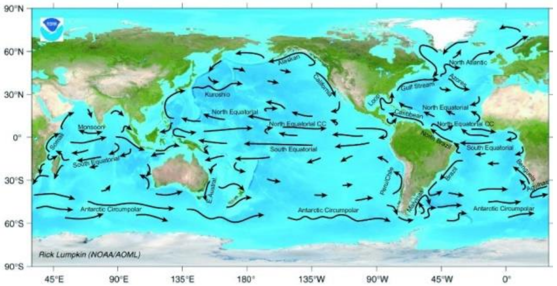
$$k_i = \Sigma_j A_{ij}$$



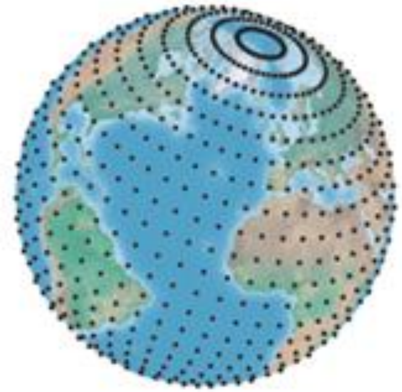
Same approach for climate time series ⇒ climate network



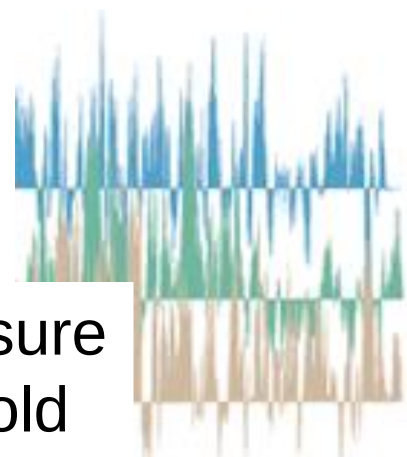
Complex network representation of the climate system



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



More than 10000 nodes (with different sizes).



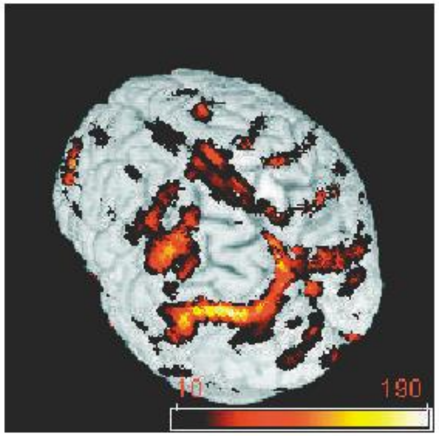
Daily resolution: more than 13000 data points in each TS

Surface Air Temperature <u>Anomalies</u> (solar cycle removed)

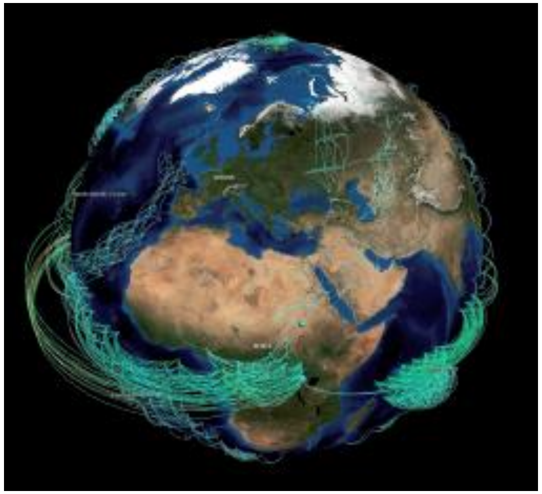
Sim. measure + threshold

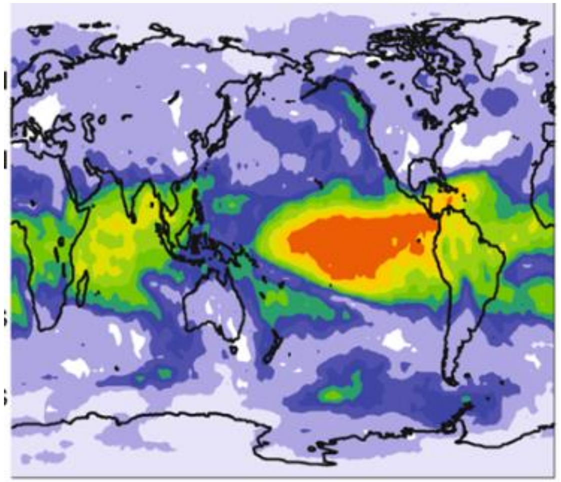
Donges et al, Chaos 2015

Brain network



Climate network





Area weighted connectivity (AWC):
weighted degree (nodes represent areas with different sizes)

$$\mathrm{AWC}_i = \frac{\sum_j^N A_{ij}\cos(\lambda_j)}{\sum_j^N \cos(\lambda_j)}$$

How to select the threshold ?

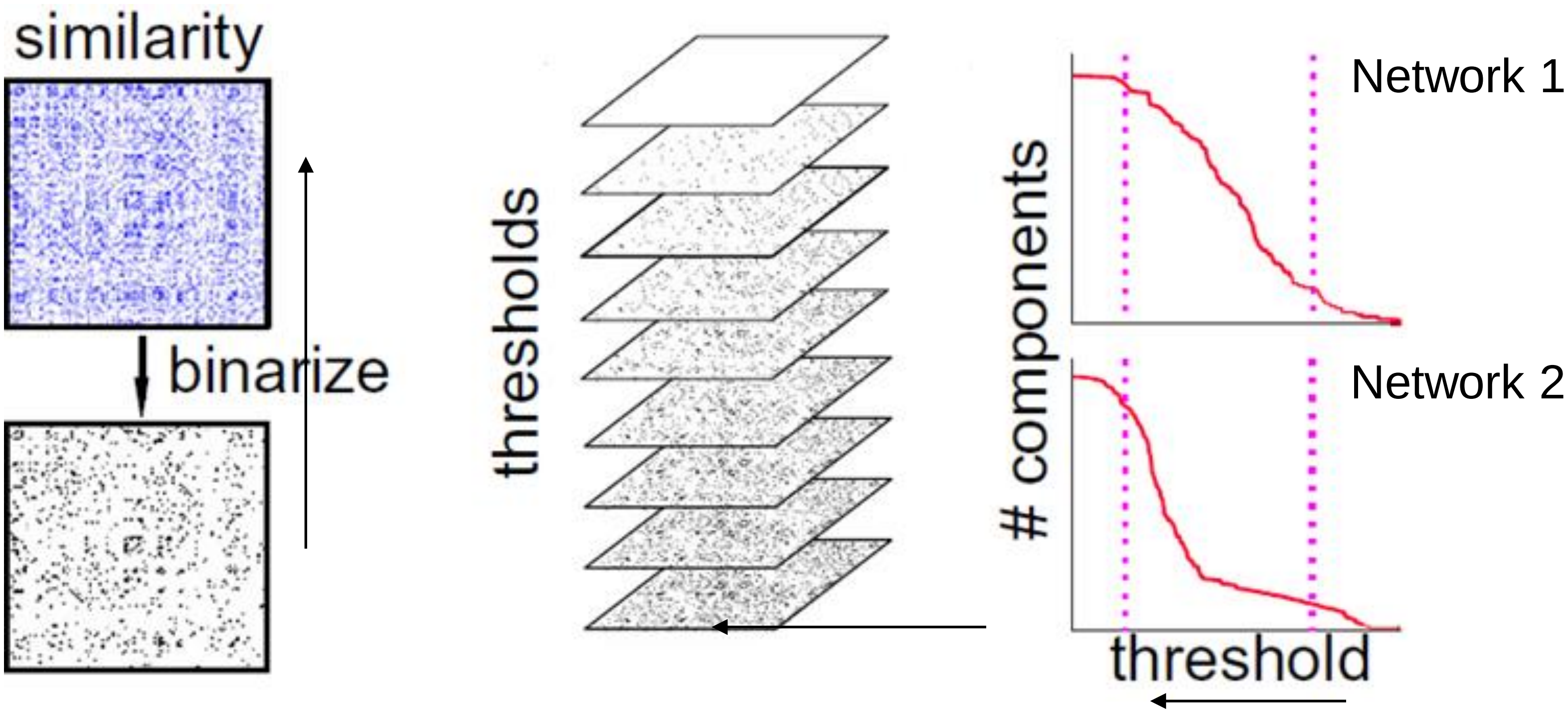
$S_{ij} > \mathrm{Th} \Rightarrow A_{ij} = 1$,
else $A_{ij}=0$

Three criteria are typically used:

- A significance level is used (typically 5%) in order to omit connectivity values that can be expected by chance;
- We select an arbitrary value as threshold, such that it gives a certain pre-fixed number of links (or link density);
- We define the threshold as large as possible while guaranteeing that all nodes are connected (or a so-called “giant component” exists).

C. M. van Wijk et al., “*Comparing Brain Networks of Different Size and Connectivity Density Using Graph Theory*”, PLoS ONE 5, e13701 (2010)

Problems with thresholding

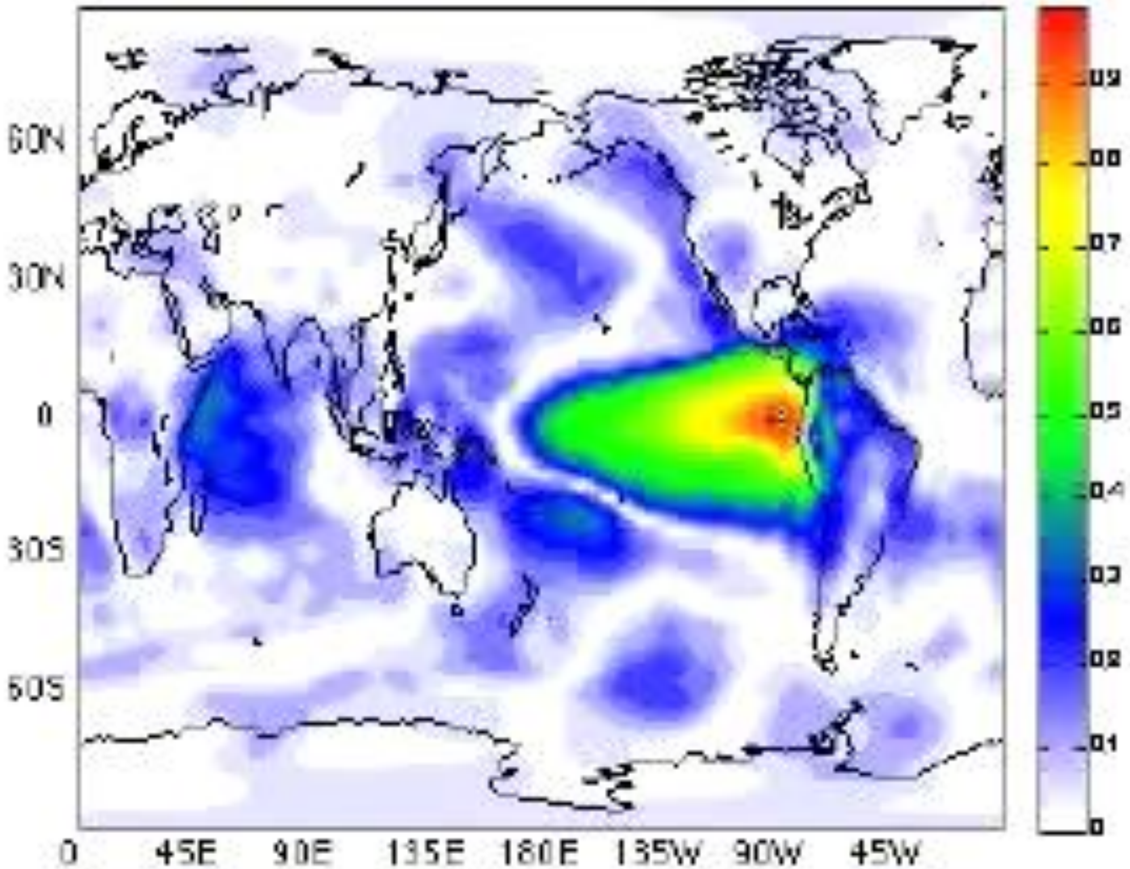


- The number of connected components as a function of threshold reveals different structures.
- But thresholding near the dotted lines indicates (inaccurately) that networks 1 and 2 have similar structures.
- Climate networks: too high threshold only keeps links between neighboring nodes.

Giusti et al., J Comput Neurosci (2016) 41:1–14

Influence of the threshold in the network connectivity

Similarity measure: Pearson coefficient

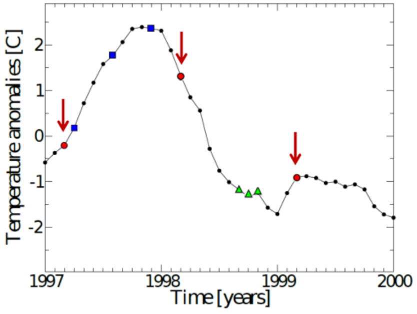


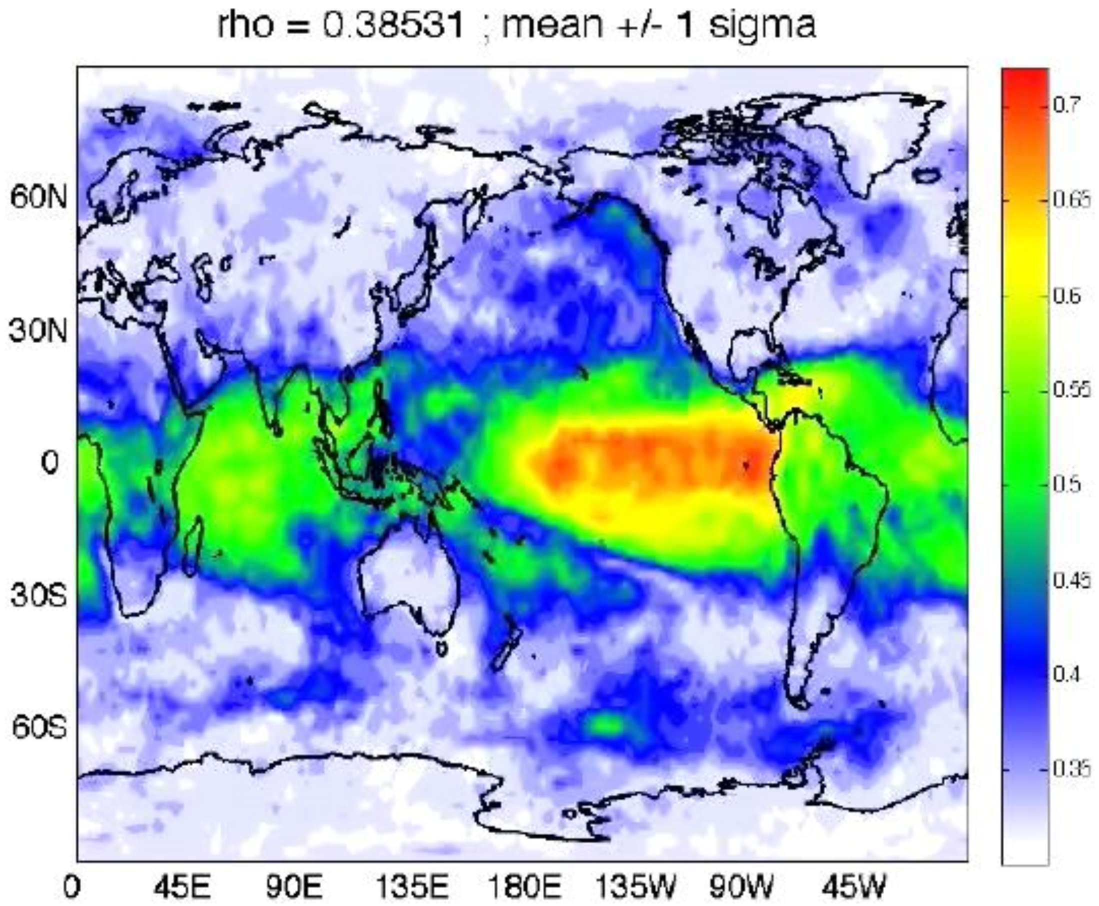
weighted degree: $$\mathrm{AWC}_i = \frac{\sum_j^N A_{ij}\cos(\lambda_j)}{\sum_j^N \cos(\lambda_j)}$$

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

Influence of the threshold in the network connectivity

Similarity measure: Mutual Information computed from the probabilities of "annual" ordinal patterns.





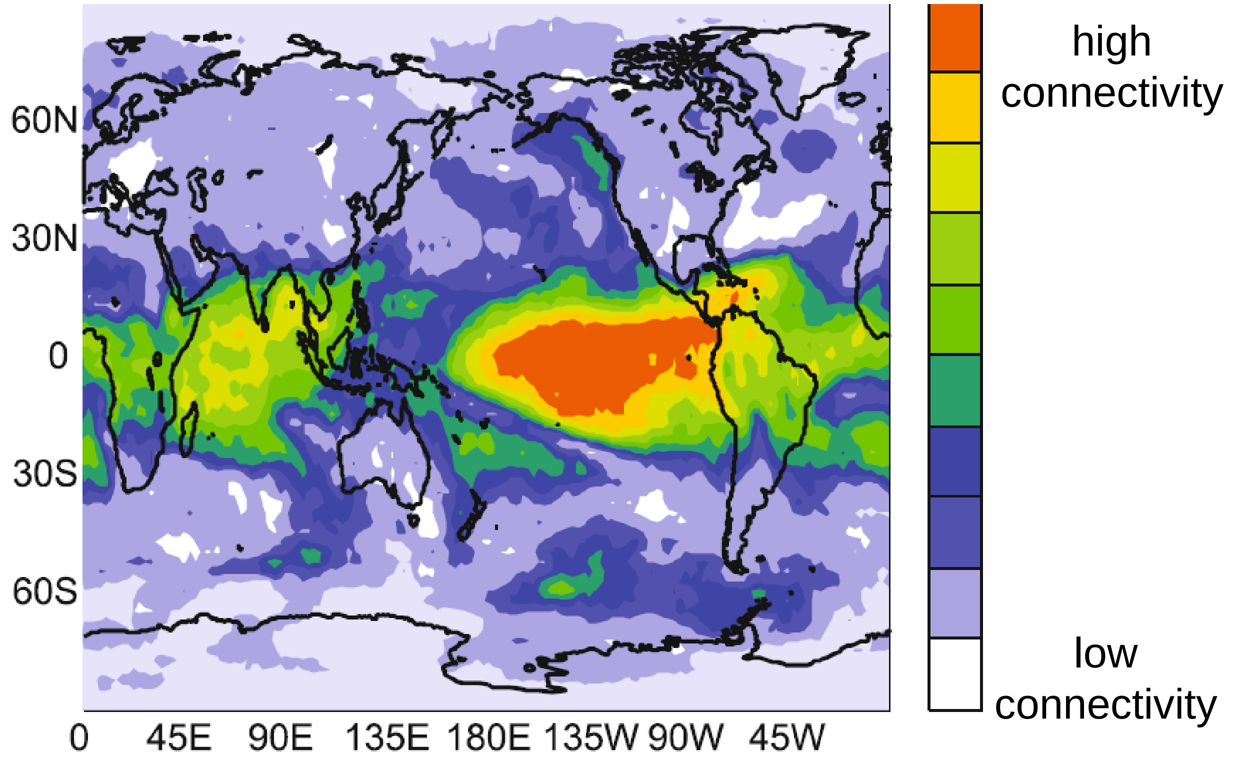
weighted degree: $\mathrm{AWC}_i = \frac{\sum_j^N A_{ij} \cos(\lambda_j)}{\sum_j^N \cos(\lambda_j)}$

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

Climate network with mutual information computed with probabilities of ordinal patterns

$$\mathrm{AWC}_i = \frac{\sum_j^N A_{ij} \cos(\lambda_j)}{\sum_j^N \cos(\lambda_j)}$$

inter-annual time-scale (3 consecutive years). The color-code indicates the Area Weighted Connectivity (weighted degree)

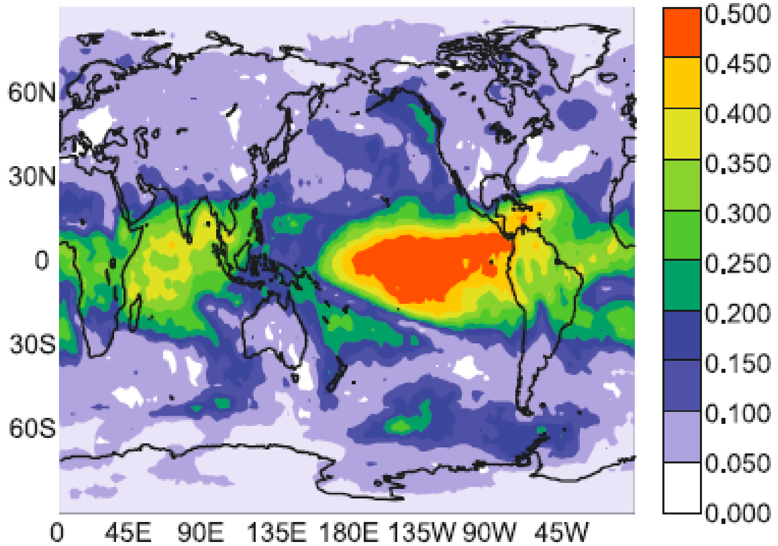


J. I. Deza, M. Barreiro, and C. Masoller, Eur. Phys. J. Special Topics 222, 511 (2013)

Comparison: ordinal probabilities vs. histogram of data values

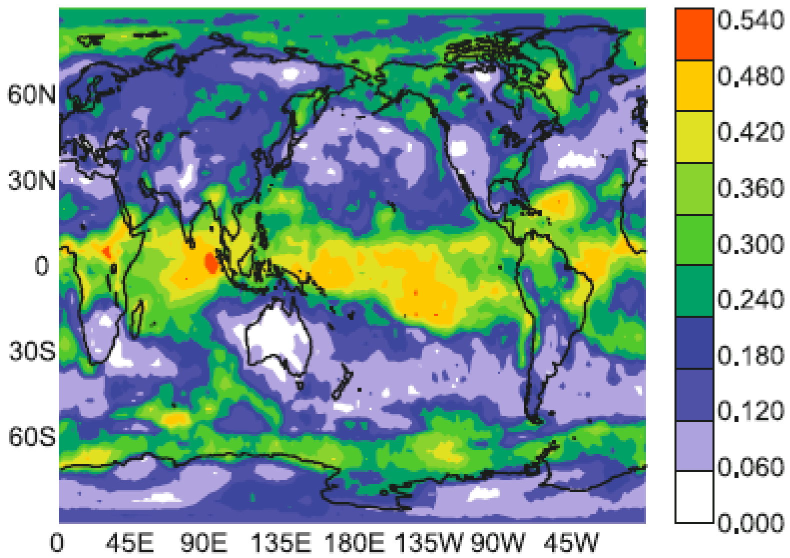
$$M_{ij} = \sum_{m,n} p_{ij}(m,n) \log \frac{p_{ij}(m,n)}{p_i(m)p_j(n)}$$

Network when the probabilities are computed with ordinal analysis

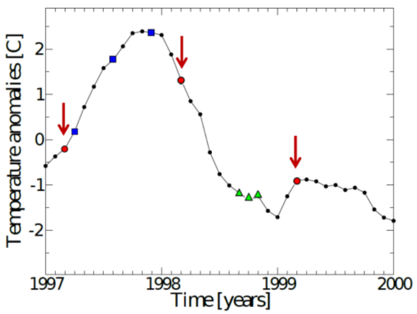


Color code indicates the area-weighted connectivity

Network when the probabilities are computed with histogram of values

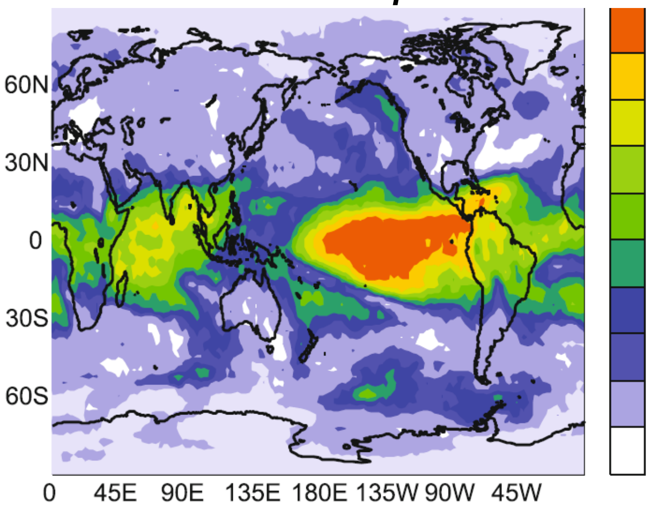


inter-annual time scale

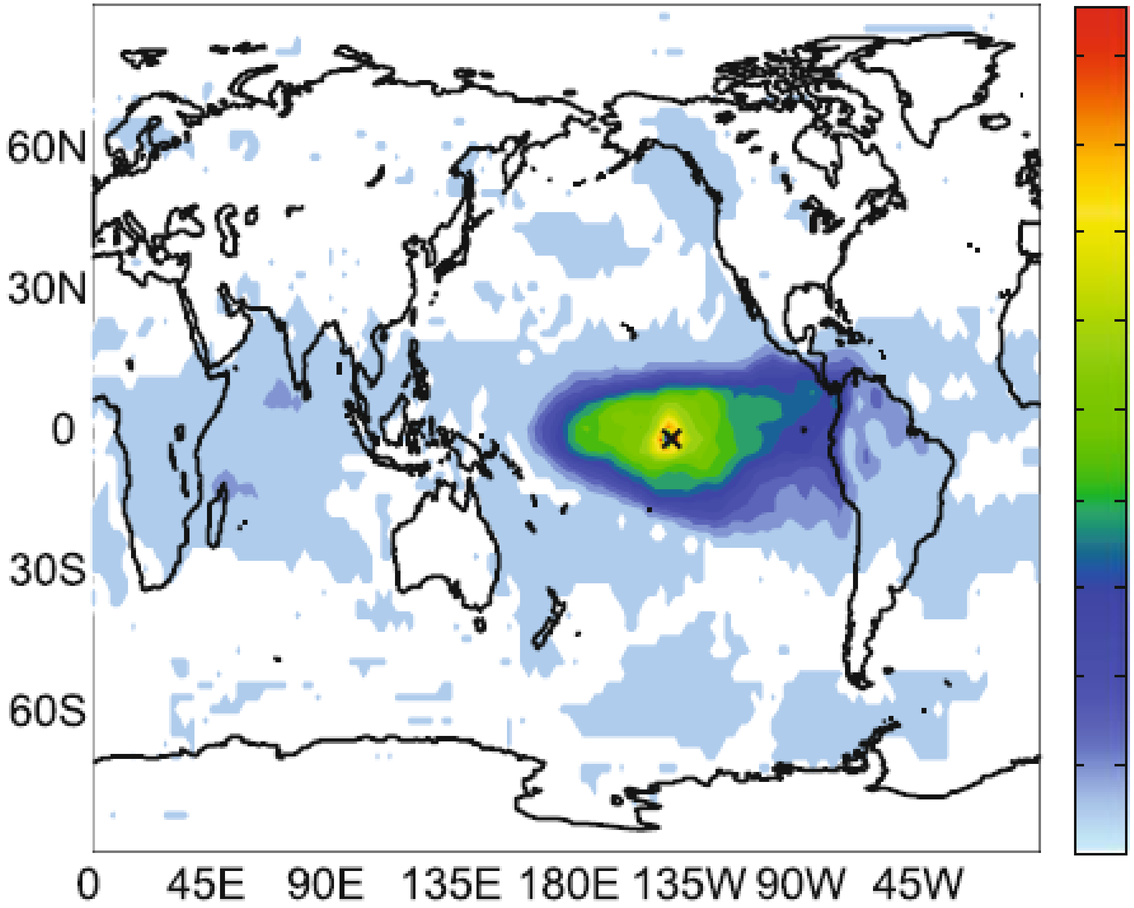


Who is connected to who?

AWC map

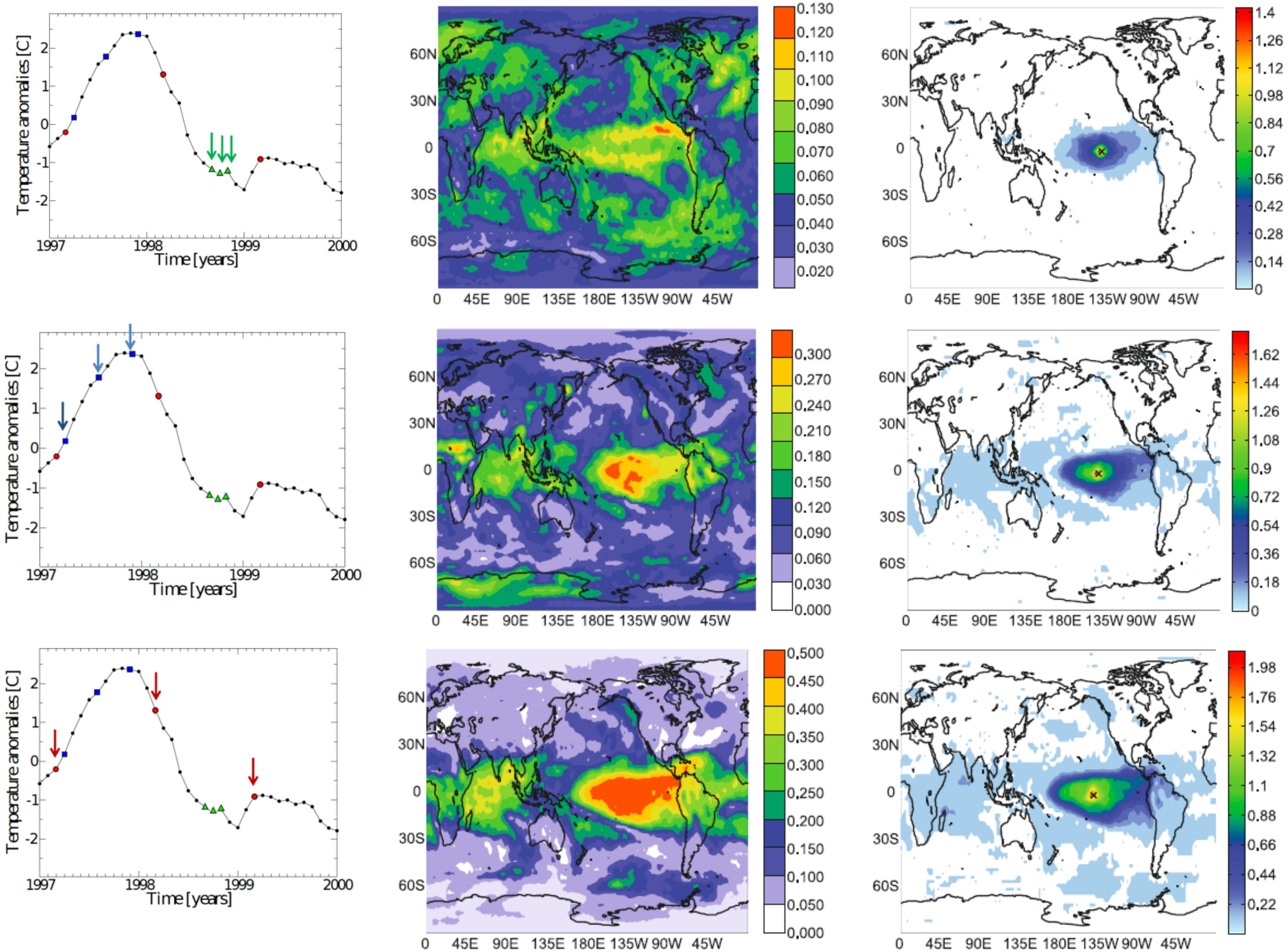


color-code indicates the MI values (only significant values)



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

Influence of the time-scale of the pattern



Longer time-scale ⇒ increased connectivity

Take home messages

- Multivariate analysis uncovers inter-relationships in datasets.

- Different similarity measures are available for inferring the connectivity of a complex system from observations.

- Different measures can uncover different properties.

- Thresholding, hidden variables, hidden “nodes” and non-stationarity can make difficult or impossible to infer the network.

- Many applications and challenges!

Acknowledgments

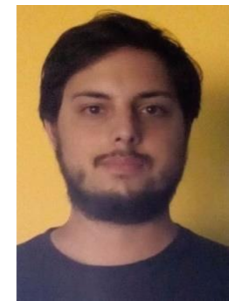










- Maria Masoliver, Pepe Aparicio Reinoso (*neurons*)
- Carlos Quintero, Jordi Tiana, Came Torrent (*laser lab*)
- Andres Aragoneses, Laura Carpi (*data analysis, networks*)
- Ignacio Deza, Giulio Tirabassi, Dario Zappala, Marcelo Barreiro (*climate*)
- Pablo Amil (*biomedical images*)

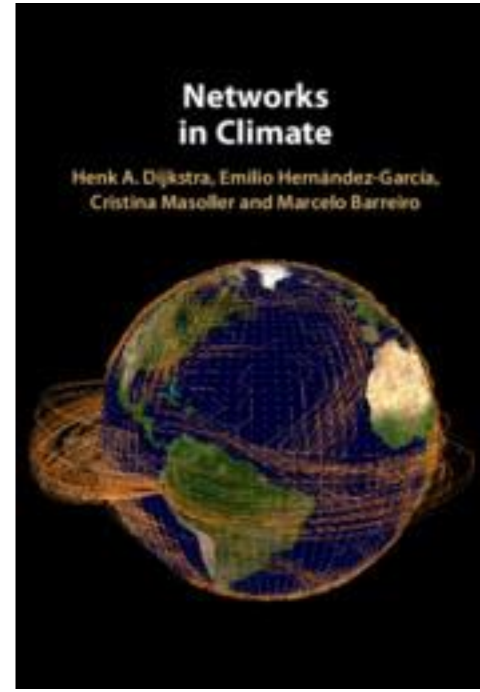
"LINC"
Learning about Interacting Networks in Climate

GOBIERNO DE ESPAÑA
MINISTERIO DE ECONOMÍA Y COMPETITIVIDAD

ICREA

References

- 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)
- G. Tirabassi et al., Ecological Complexity 19, 148 (2014)
- G. Tirabassi et al., Sci. Rep. **5** 10829 (2015)
- G. Tirabassi and C. Masoller, Sci. Rep. 6:29804 (2016)
- T. A. Schieber et al, Nat. Comm. 8, 13928 (2017)



Cambridge University Press 2019

<cristina.masoller@upc.edu>

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