

# Nonlinear time series analysis:

## Univariate analysis

Cristina Masoller

Universitat Politècnica de Catalunya, Terrassa, Barcelona, Spain

Cristina.masoller@upc.edu  
[www.fisica.edu.uy/~cris](http://www.fisica.edu.uy/~cris)



UNIVERSITAT POLITÈCNICA  
DE CATALUNYA  
BARCELONATECH

*Campus d'Excel·lència Internacional*

## ■ 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.

# Methods of time-series analysis

- Return maps
- Distribution of data values
- Correlation and Fourier analysis
- Lyapunov and fractal dimensions
- Stochastic models and surrogates
- Symbolic methods
- Information theory measures: entropy and complexity
- Network representation of a time-series
- Spatio-temporal representation of a time-series
- Instantaneous phase and amplitude

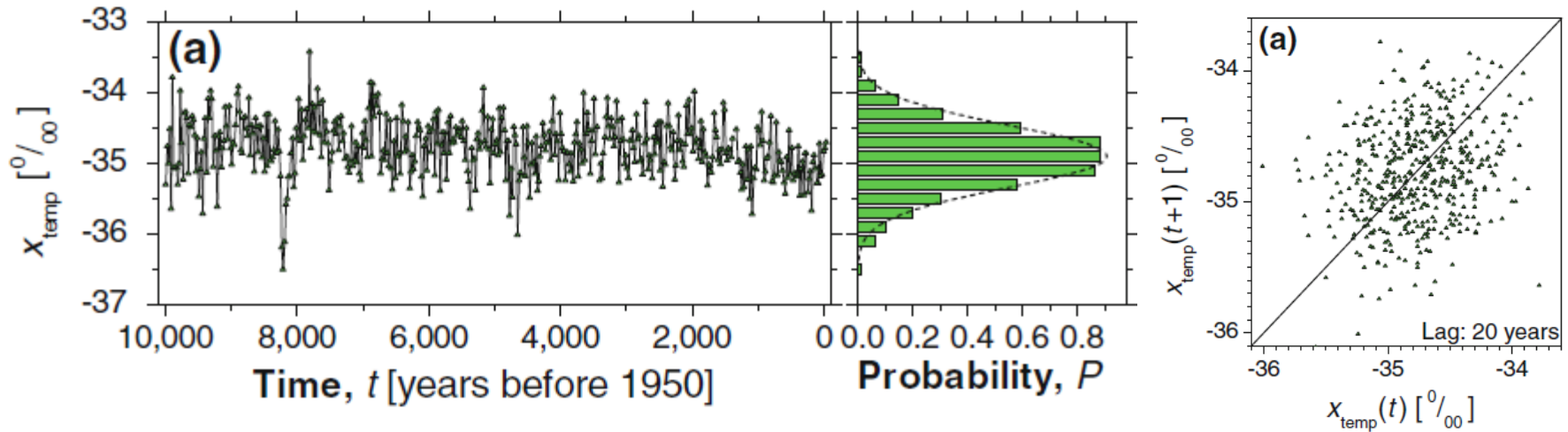
# To begin with

$$X = \{x_1, x_2, \dots, x_N\}$$

- First step: Look at the data.
- Examine simple properties:
  - Return map (plot of  $x_i$  vs.  $x_{i+\tau}$ )
  - Distribution of data values
  - Auto correlation
  - Fourier spectrum

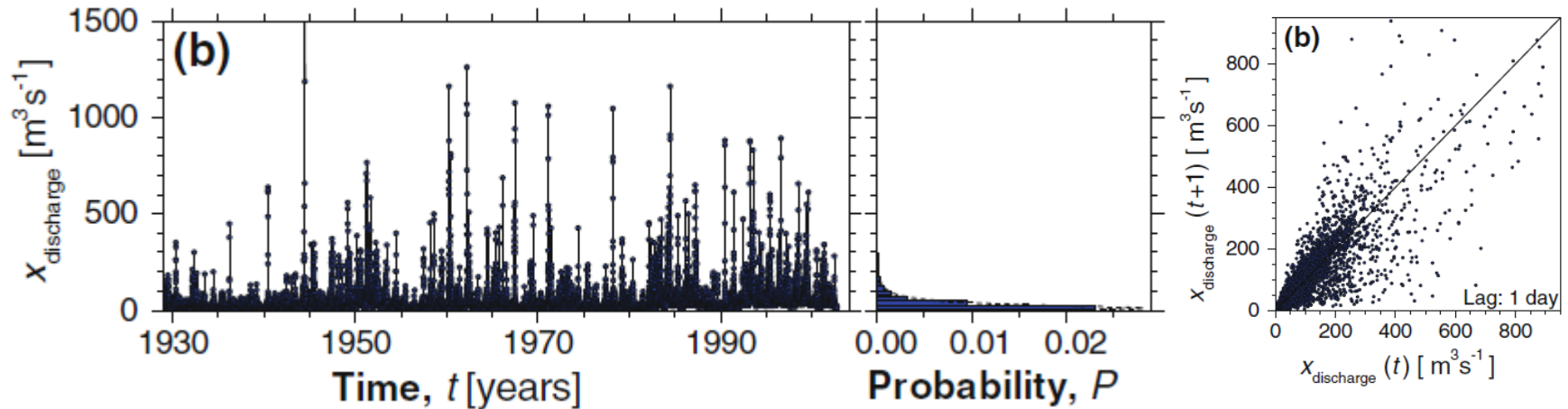


# First example of a geophysical time series



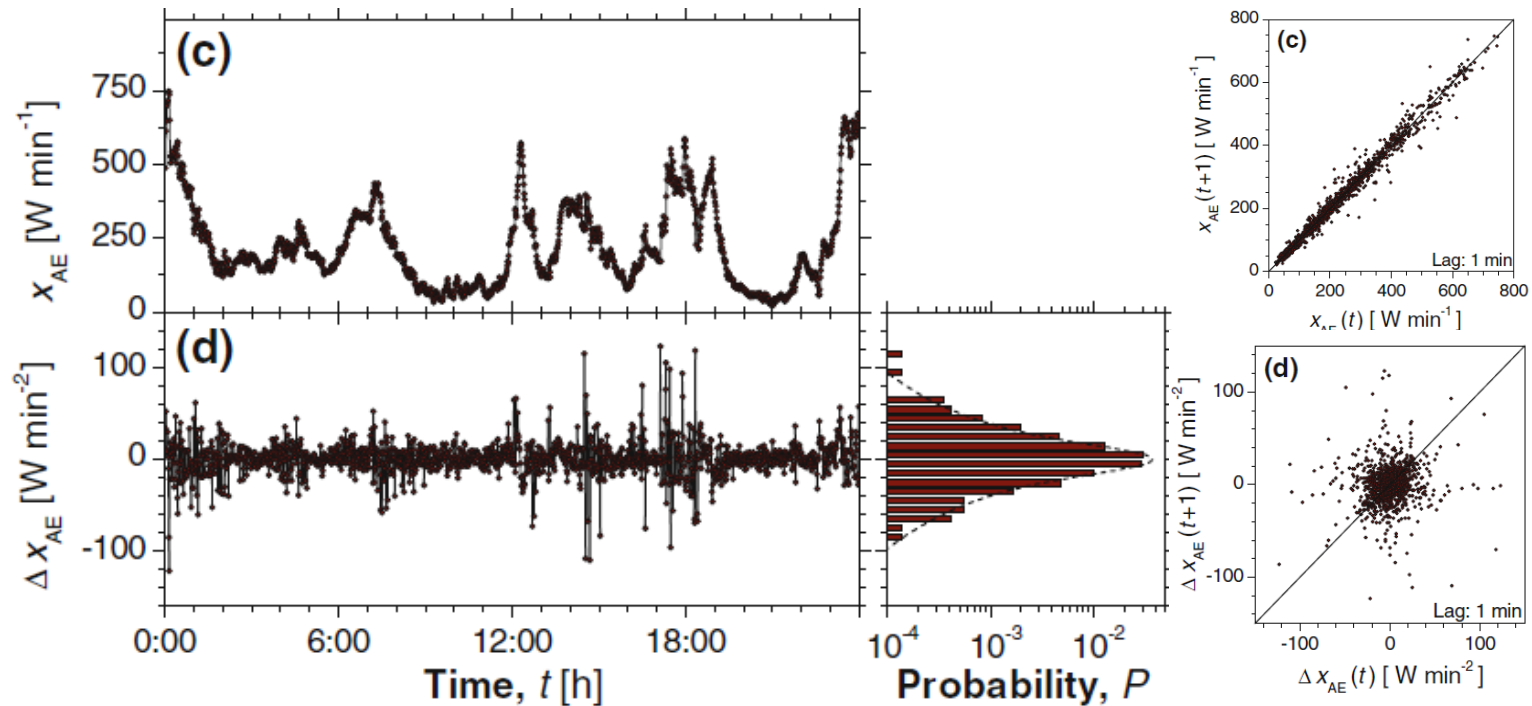
Bi-decadal oxygen isotope data set d18O (proxy for palaeotemperature) from Greenland Ice Sheet Project Two (GISP2) for the last 10,000 years with 500 values given at 20 year intervals.

## Second example



Discharge of the Elkhorn river (at Waterloo, Nebraska, USA) sampled daily for the period from 01 January 1929 to 30 December 2001.

# Third example



The geomagnetic auroral electrojet (AE) index sampled per minute for the 24 h period of 01 February 1978 and the differenced index:  $\Delta x_{AE}(t) = x_{AE}(t) - x_{AE}(t - 1)$

# Autocorrelation function

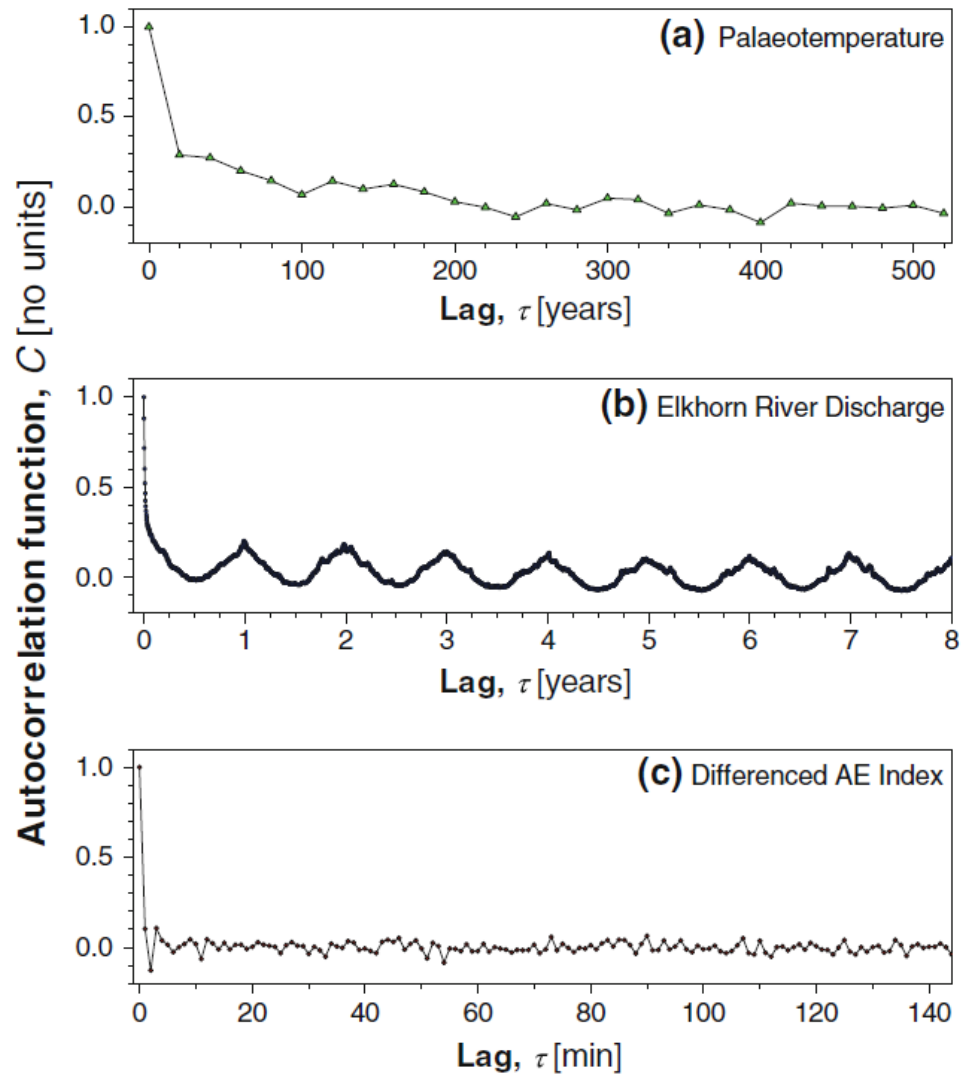
- The return maps reveal different degrees of correlations between consecutive data values.
- How to quantify?
- Linear tool: the autocorrelation function

$$C(\tau) = \frac{1}{\sigma_x^2} \frac{1}{N - \tau} \sum_{t=1}^{N-\tau} (x_t - \bar{x})(x_{t+\tau} - \bar{x}) \quad C(0) = 1$$

## Autocorrelation: Quantifies persistence (memory)

- Persistence: large values tend to follow large ones, and small values tend to follow small ones.
- Anti-persistence: large values tend to follow small ones and small values large ones.
- Short-range correlations: values are correlated with one another at short lags in time  $|C(\tau)| \leq \kappa_0 \exp(-\kappa\tau)$
- Long-range correlations: values are correlated with one another at very long lags in time (all or almost all values are correlated with one another)

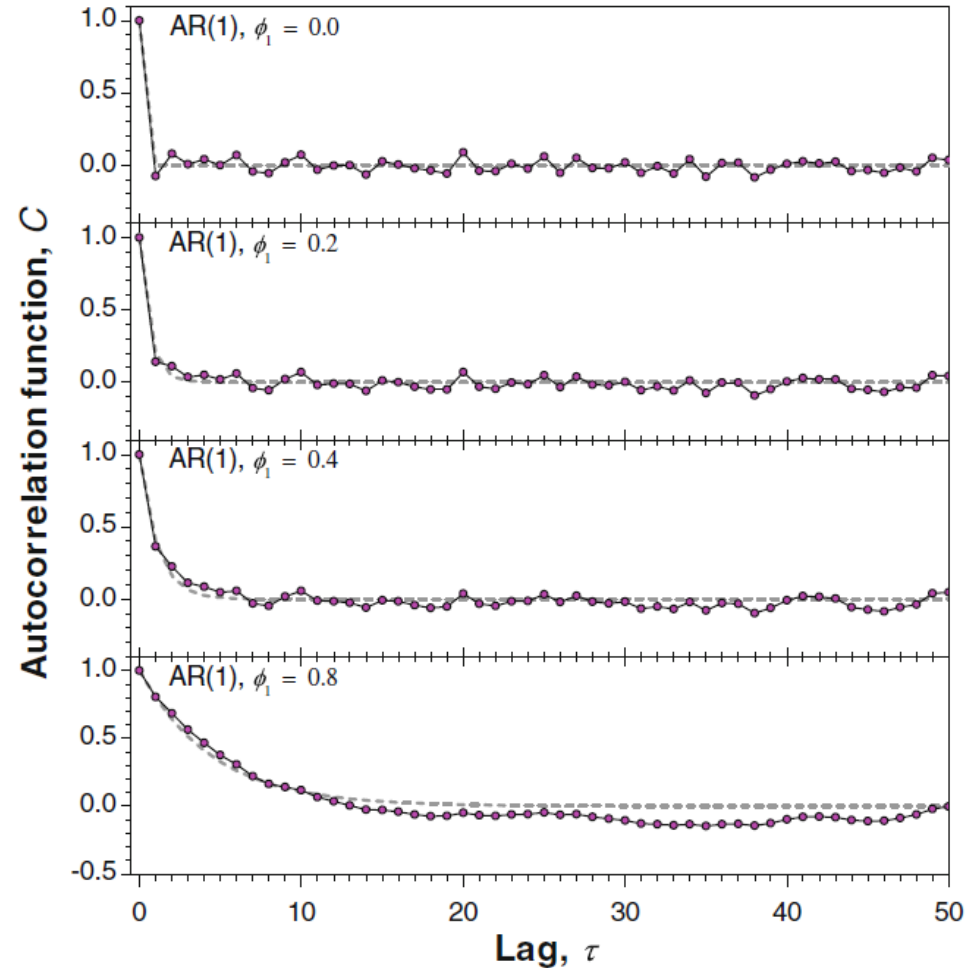
# Back to the three examples of geophysical time series



# Synthetic model: first-order autoregressive (AR(1)) process

$$x_t = \phi_1 x_{t-1} + \varepsilon_t$$

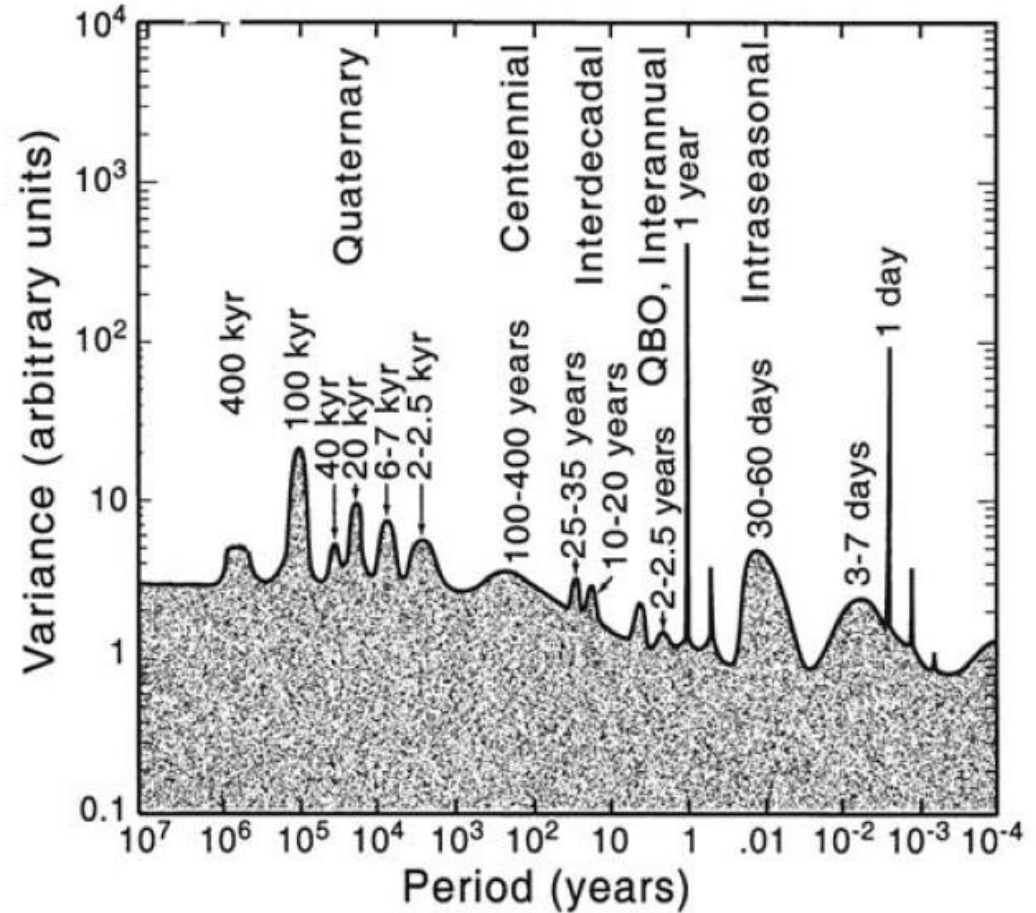
$-1 < \phi_1 < 1$       White noise



# Spectral analysis

- Climate phenomena occurs with a wide range of time-scales
  - hours to days,
  - months to seasons,
  - decades to centuries,
  - and even longer...

An “artist’s representation” of the power spectrum of climate variability (Ghil 2002).





# Power spectral density (PSD)

- Discrete Fourier transform  $X_k = \Delta \sum_{t=1}^N x_t e^{2\pi i t k / N}$ ,  $k = 1, 2, \dots, N$   
 $\Delta$  = sampling interval  
 $X_k$  = complex Fourier coeff. associated to freq.  $f_k = k / (N\Delta)$

- PSD:  $S_k = \frac{2|X_k|^2}{N\Delta}$ ,  $k = 1, 2, \dots, \frac{N}{2}$

- How to calculate? Fast Fourier Transform (FFT)

- Further reading:

Press WH et al. *Numerical recipes in C/Fortran: the art of scientific computing* (Cambridge University Press)

# Numerical technicalities

- Discrete Fourier transform  $X_k = \Delta \sum_{t=1}^N x_t e^{2\pi i t k / N}$ ,  $k = 1, 2, \dots, N$

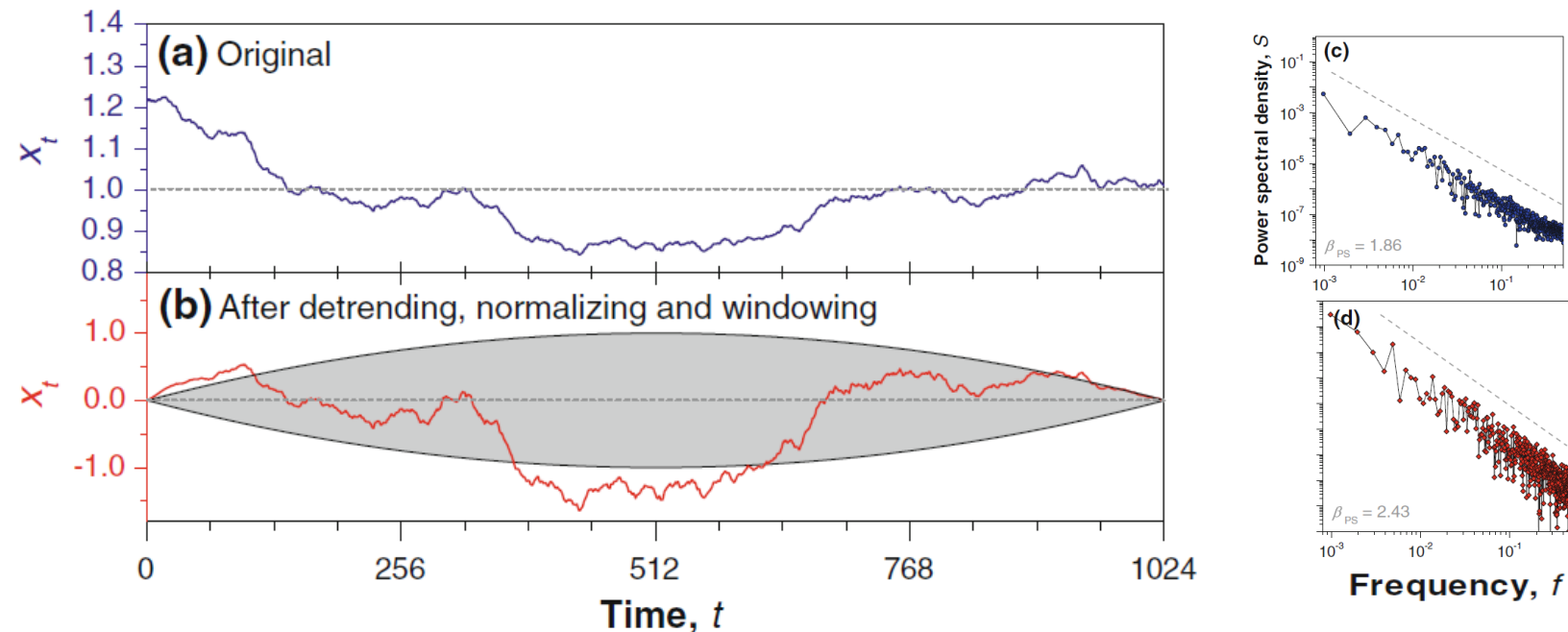
is designed for ‘circular’ time series (i.e. the last and first values in the time series ‘follow’ one another).

- Large values of  $|x_N - x_1|$  (typical in non-stationary time series) can result in spurious artifacts.
- It is recommended to normalize to  $\mu=0$ ,  $\sigma=1$  and to remove the trend (“*detrend*”) before computing the FFT.
- Simple ways to detrend:
  - Take the best-fit straight line to the time series and subtract it from all the values.
  - Connect a line from the first point to the last point and subtract this line from the time series, forcing  $x_N = x_1 = 0$

# Pre-processing to reduce the impact of the initial and final data points

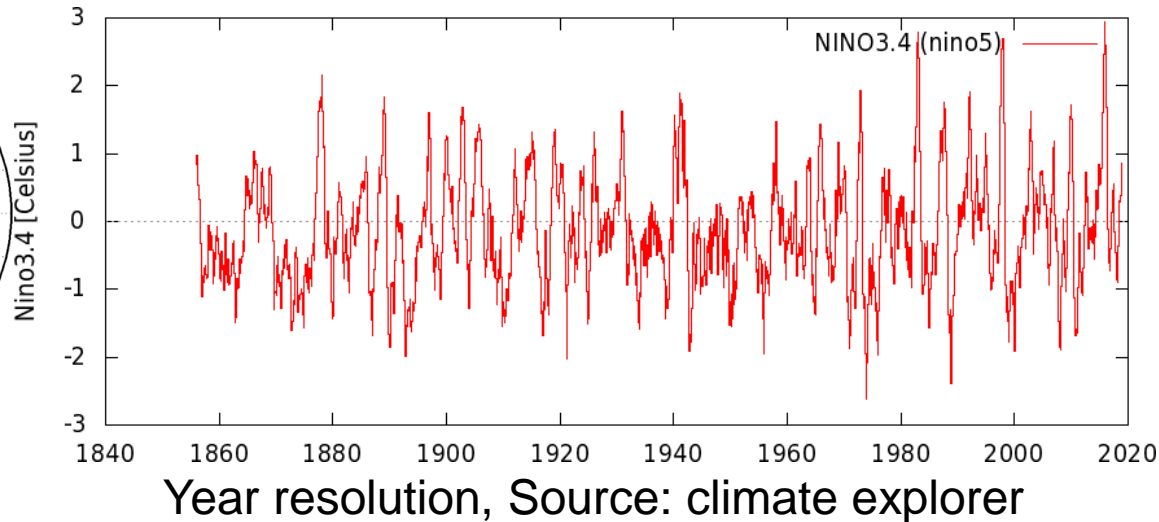
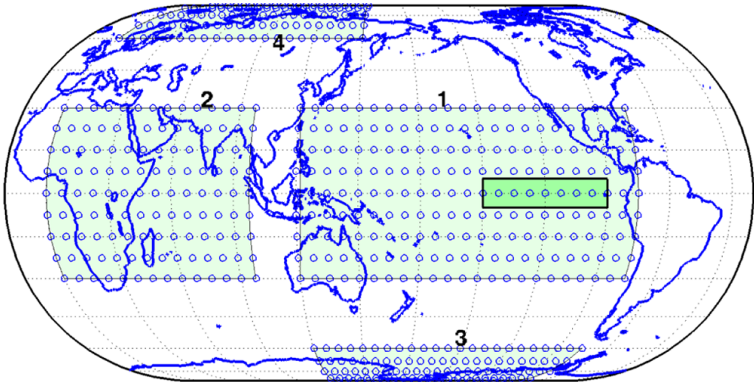
- Choose a “weight shape” and multiply each value of the time series by its weight.

$$w_t = 1 - \left( \frac{t - (N/2)}{N/2} \right)^2, \quad t = 1, 2, \dots, N.$$



# Example: El Niño 3.4 index

## Sea surface temperature (SST) anomaly in the eastern tropical Pacific Ocean

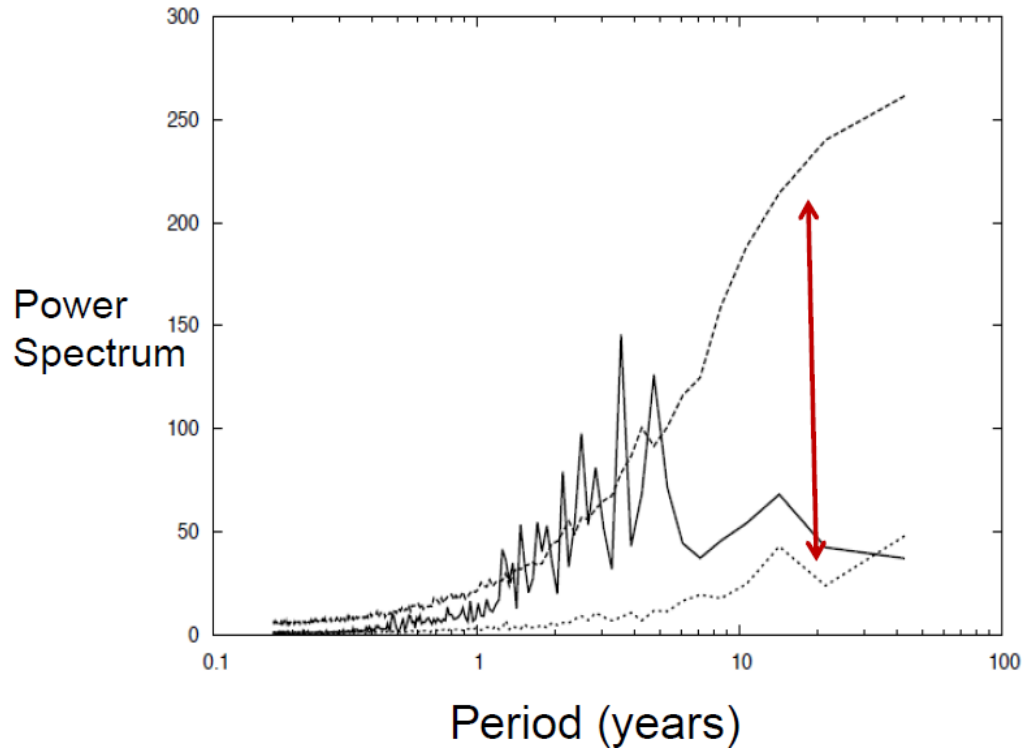


- From 5°N to 5°S and from 170°W to 120°W
- An El Niño (La Niña) event is identified if the 5-month running-average of the NINO3.4 index exceeds +0.4°C (-0.4°C) for at least 6 consecutive months.

Month resolution:

[https://www.esrl.noaa.gov/psd/gcos\\_wgsp/Timeseries/Data/nino34.long.anom.data](https://www.esrl.noaa.gov/psd/gcos_wgsp/Timeseries/Data/nino34.long.anom.data)

# Fourier analysis of NINO3.4 index (monthly resolution)



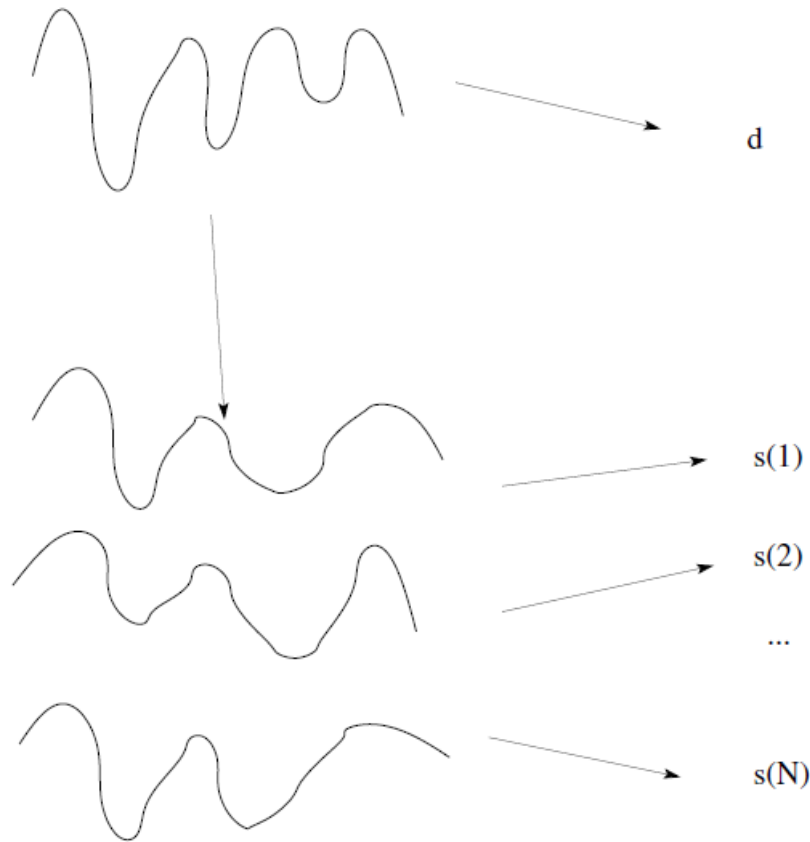
Are the peaks “significant”?

**Surrogates:** 100 stochastic AR(1) processes with same autocorrelation at lag 1 and variance as NINO3.4

➡ significant peaks at 3-4 years

Source: [G. Tirabassi PhD thesis](#)

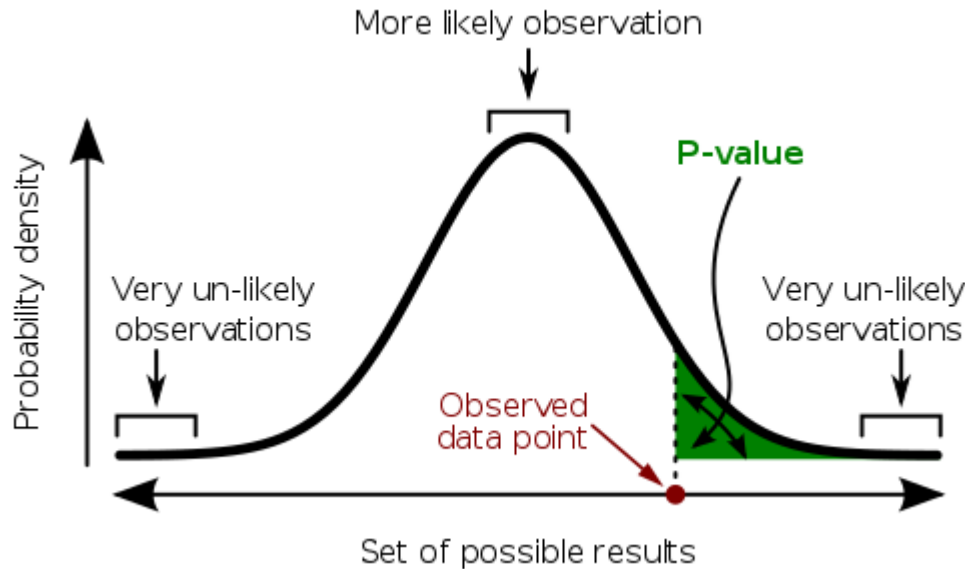
# The method of surrogate data



Taken from  
M. Small, *Applied Nonlinear Time  
Series Analysis* (World Scientific,  
2005)

- Real observed time series.
- Generate an ensemble of “surrogate” time series that are both “similar” to the original and also consistent with the specific **null hypothesis** (NH) that we want to test.
- Measure an statistical property: “ $d$ ” in the original series and “ $s(i)$ ” in the ensemble time series.
- Is “ $d$ ” consistent with the distribution of “ $s(i)$ ” values?
  - No! we **reject** the NH.
  - Yes! we **fail to reject** the NH.

# p value

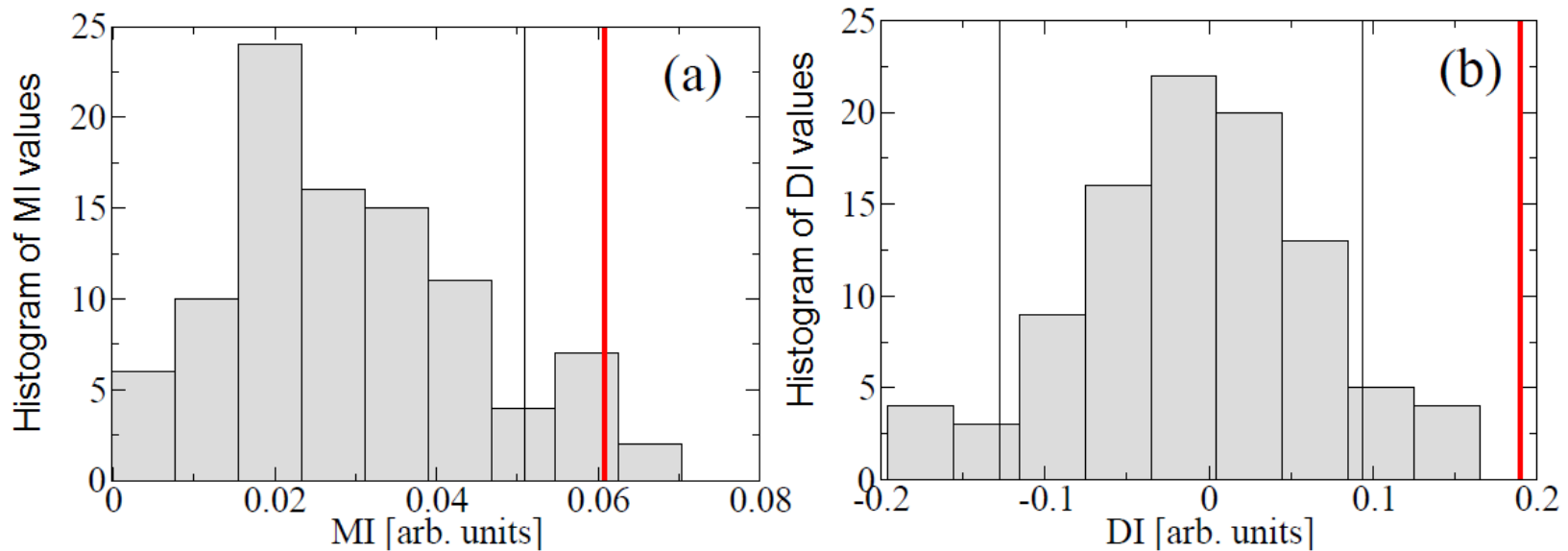


Source: wikipedia

The p-value only measures the compatibility of an observation with a hypothesis, not the truth of the hypothesis.

Altman, N. and Krzywinski, M. *Interpreting P values*. Nature Methods 14, 213 (2017).

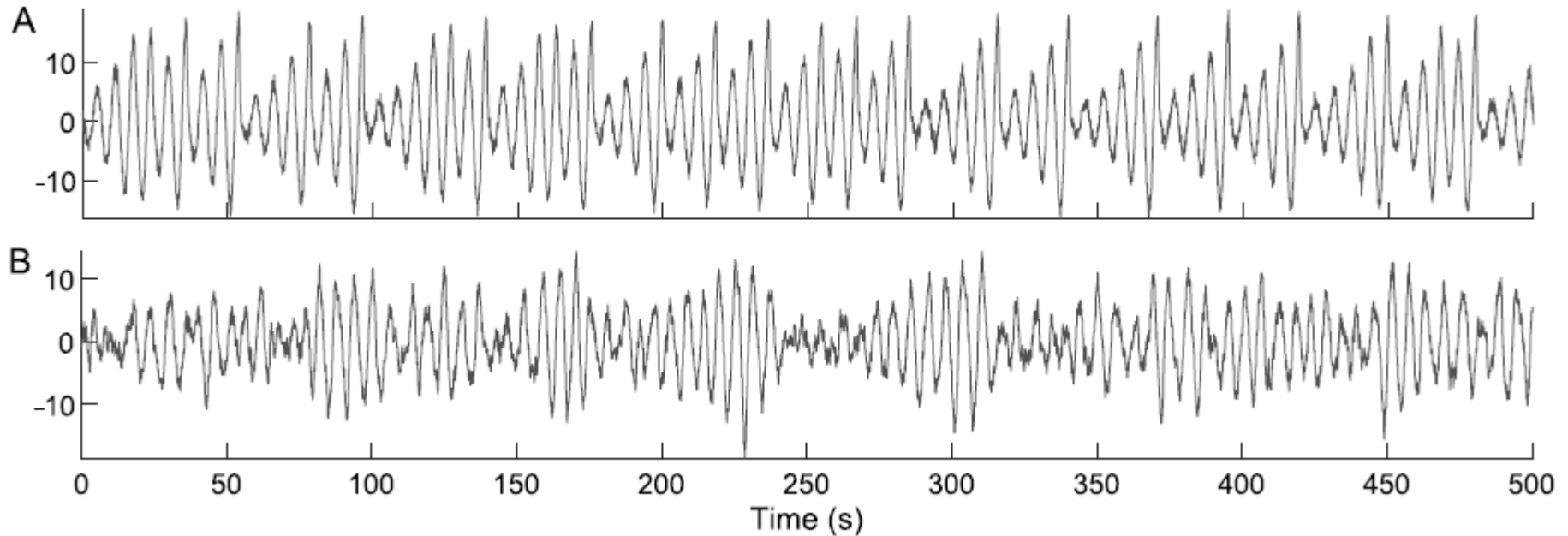
# Example



- Red: MI and DI values computed from original data
- Histogram: MI and DI values computed from surrogates
- Think lines: significance thresholds (mean  $\pm$  3  $\sigma$ )
- In both cases the NH (no MI, no DI) is rejected.



# Surrogate test for nonlinearity



A: Rossler with  
 $a = 0.165$ ,  $b = 0.2$   
and  $c = 10$

$$\dot{x} = -y - z$$

$$\dot{y} = x + ay$$

$$\dot{z} = b + z(x - c),$$

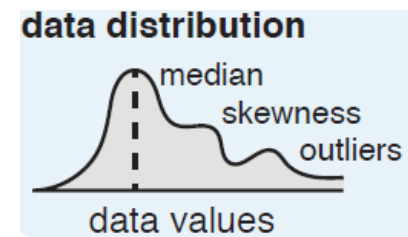
B: High-order (linear) autoregressive process

$$s_n = 1.625s_{n-1} + 0.284s_{n-2} + 0.355s_{n-3} + \eta_n - 0.960\eta_{n-1}$$

Proper surrogates detect nonlinearity in  
A (reject NH) but not in B (fail to reject NH)

G. Lancaster et al, “*Surrogate data for hypothesis testing of physical systems*”, Physics Reports 748 (2018) 1–60

# How to characterize the distribution of data values?



- **Mean** (expected value of  $X$ ):  $\mu = E[X] = \sum_{i=1}^k x_i p_i = x_1 p_1 + x_2 p_2 + \dots + x_k p_k$

- **Variance**:  $\sigma^2 = \text{Var}(X) = E[(X - \mu)^2]$

- **Skewness**: “measures” the asymmetry of the distribution

$$Z = \frac{X - \mu}{\sigma} \quad S = E[Z^3]$$

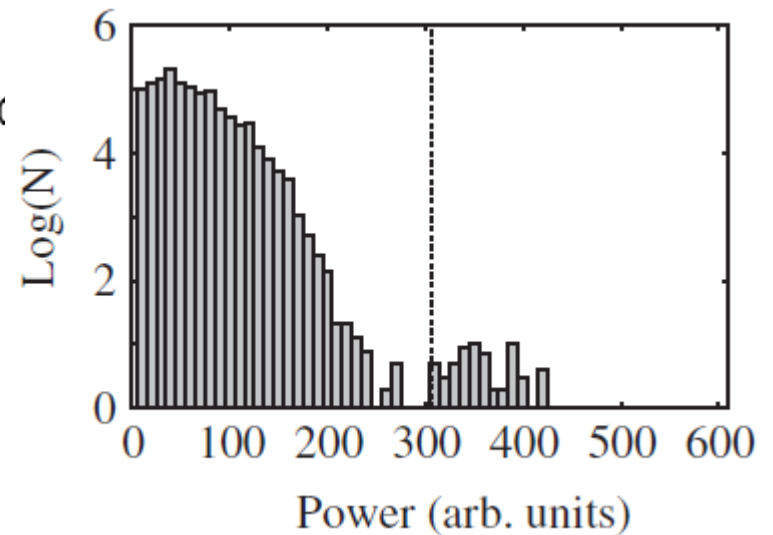
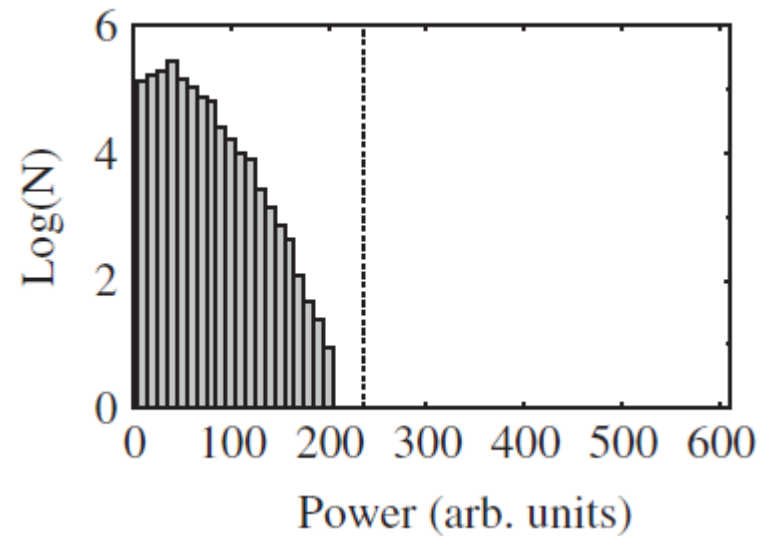
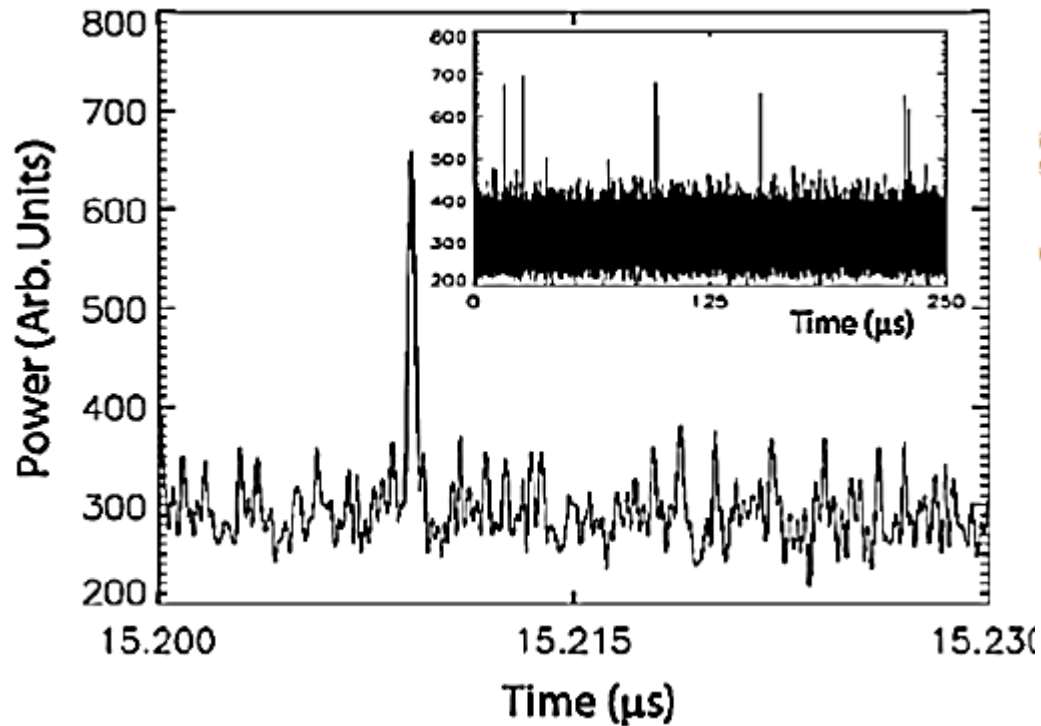
- **Kurtosis**: measures the “tailedness” of the distribution. For a normal distribution  $K=3$ .

$$K = E[Z^4]$$

- **Coefficient of variation**: normalized measure of the width of the distribution.

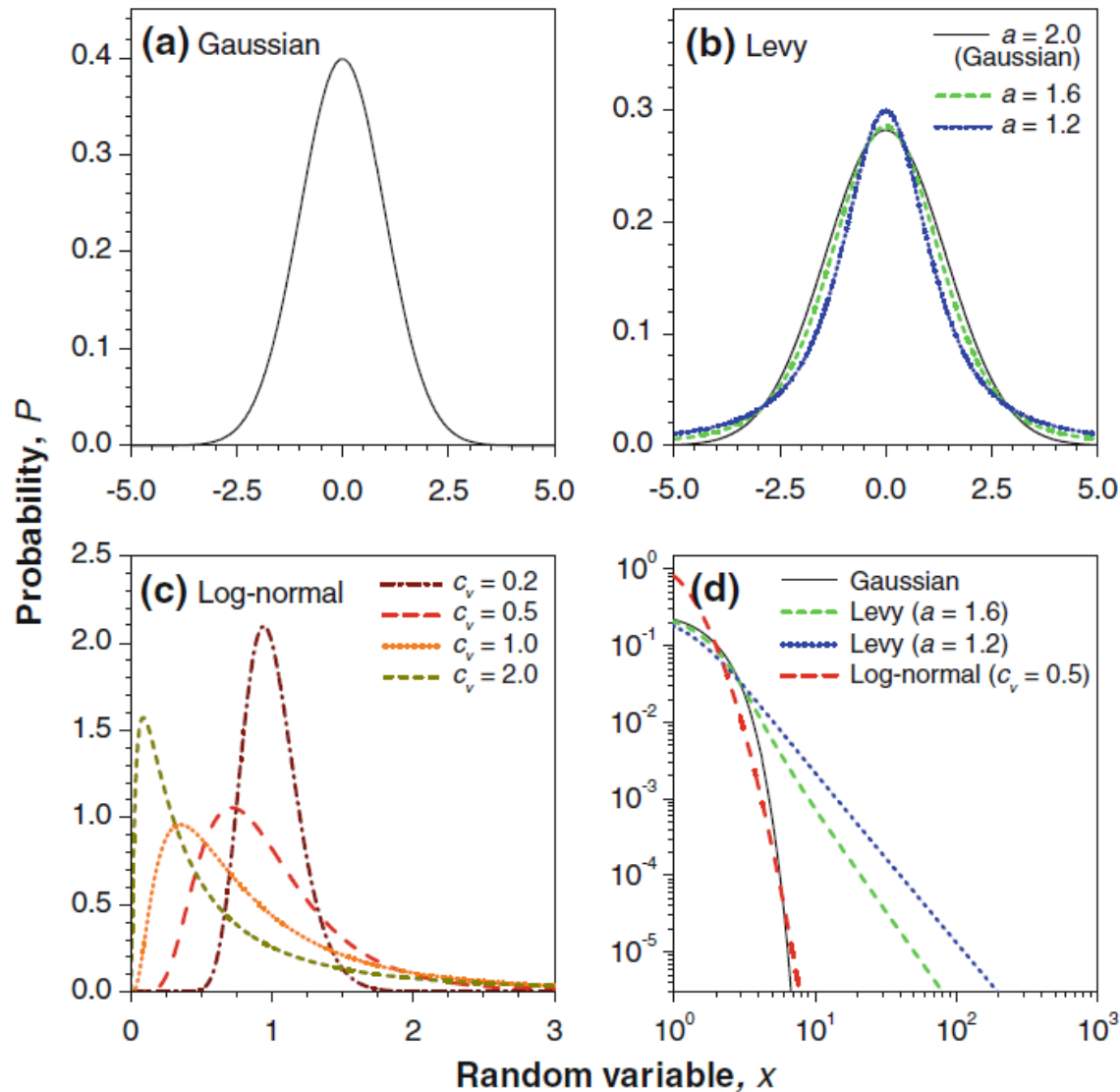
$$C_v = \sigma / |\mu|$$

# Long tailed distribution? Outliers?



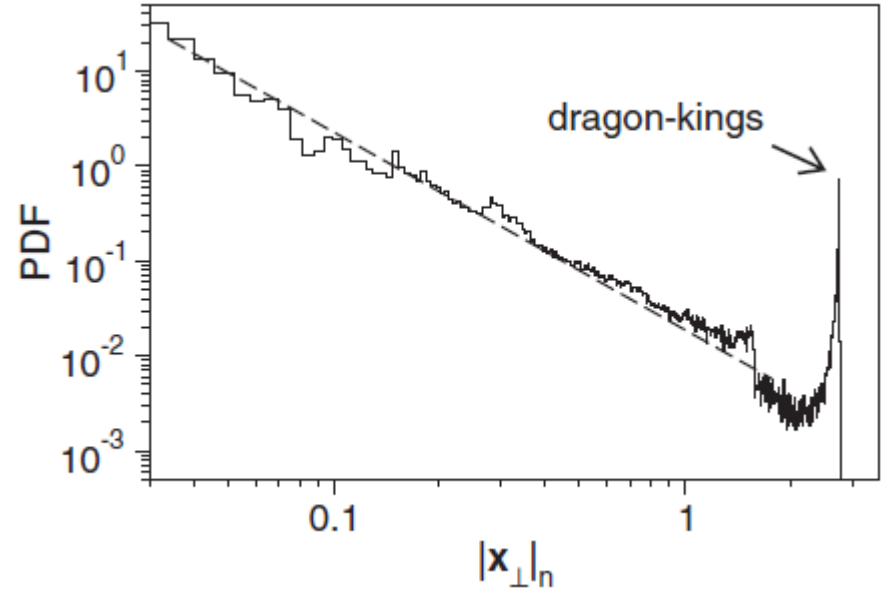
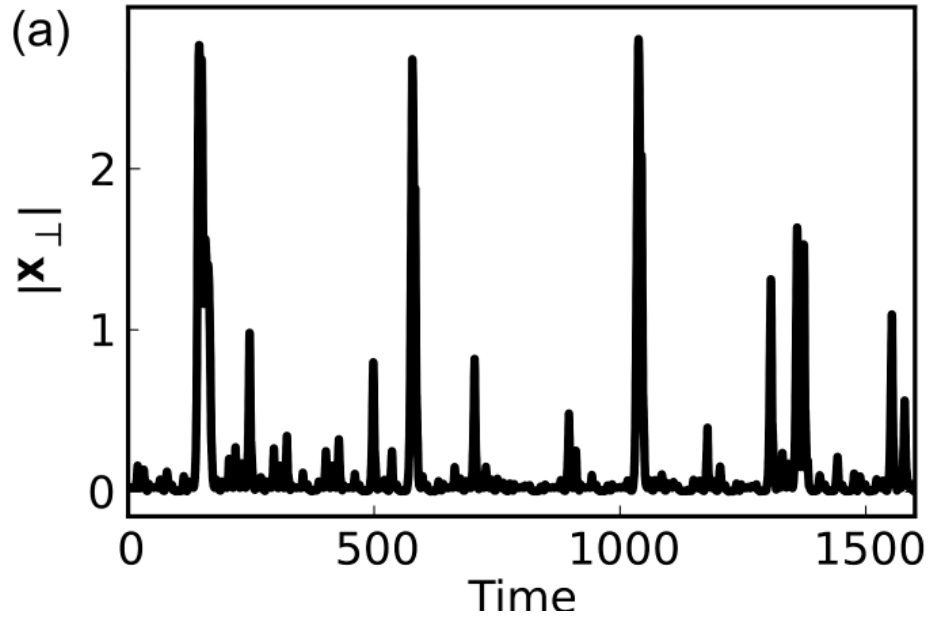
Bonato et al. PRL 107, 053901 (2011)

# Comparison of probability distributions



	Gaussian	Levy	Log-normal
<i>One-point probability distribution</i>	$P(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$	Closed expressions are only known for (Zolotarev 1986): $a = 1.0$ (Cauchy distribution), $1.5$ (Holtsmark distribution), $2.0$ (Gaussian distribution, see previous column)	$P(x) = \frac{1}{x\sqrt{2\pi\sigma^2}} e^{-\frac{(\ln x - \mu)^2}{2\sigma^2}}$
<i>Range</i>	$-\infty < x < \infty$	$-\infty < x < \infty$	$0 \leq x < \infty$
<i>Parameters</i>	$\mu$ : mean value ( $-\infty < \mu < \infty$ ) $\sigma$ : standard deviation ( $\sigma > 0$ )	$a$ : exponent ( $1 \leq a \leq 2$ )	$\mu$ : mean value of the logarithm of the time series values $x_t$ , $1 \leq t \leq N$ ( $-\infty < \mu < \infty$ ) $\sigma$ : standard deviation of the logarithm of the time series values ( $\sigma > 0$ )
<i>Mean value</i>	$\bar{x} = \mu$	$\bar{x} = 0$	$\bar{x} = e^{\mu + \frac{1}{2}\sigma^2}$
<i>Standard deviation</i>	$\sigma_x = \sigma$	Not defined	$\sigma_x = e^{\mu + \frac{1}{2}\sigma^2} \sqrt{e^{\sigma^2} - 1}$
<i>Symmetry properties</i>	Symmetric with respect to $x = \mu$	Symmetric with respect to $x = 0$	Asymmetric with the coefficient of variation: $c_v = \frac{\sigma_x}{\bar{x}} = \sqrt{e^{\sigma^2} - 1}$
<i>Tail properties</i>	Thin-tailed	Heavy-tailed: $P(x) \sim \frac{1}{ x ^{a+1}}$ for $ x  \rightarrow \infty$ , the smaller the exponent $a$ , the heavier the tail	Thin-tailed
<i>Comments</i>		Collapses to a Gaussian for $a \rightarrow 2$	Collapses to a Gaussian for $c_v \rightarrow 0$

# “Dragon kings”: extreme outliers

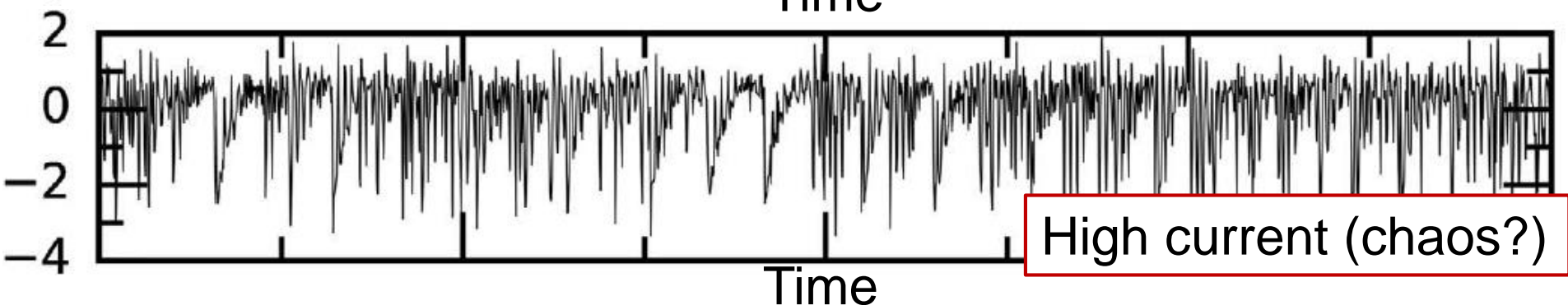
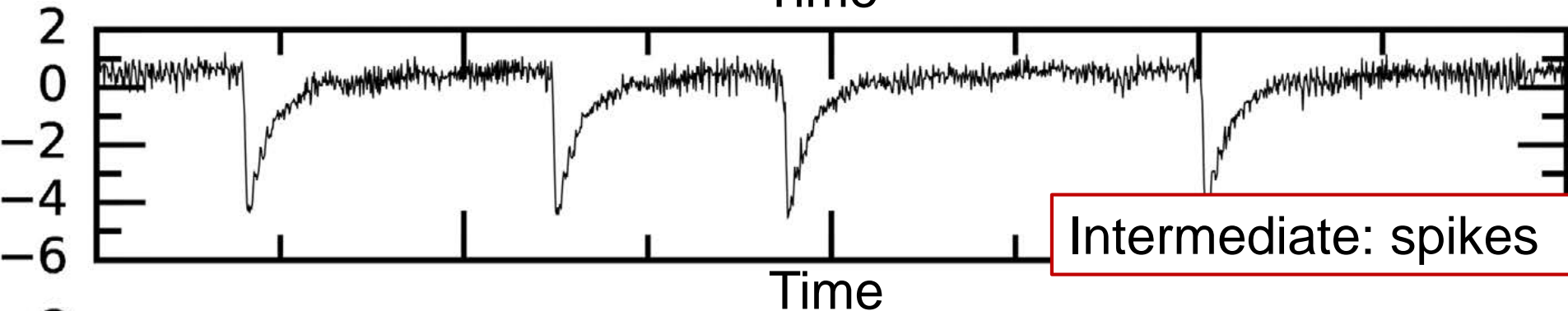
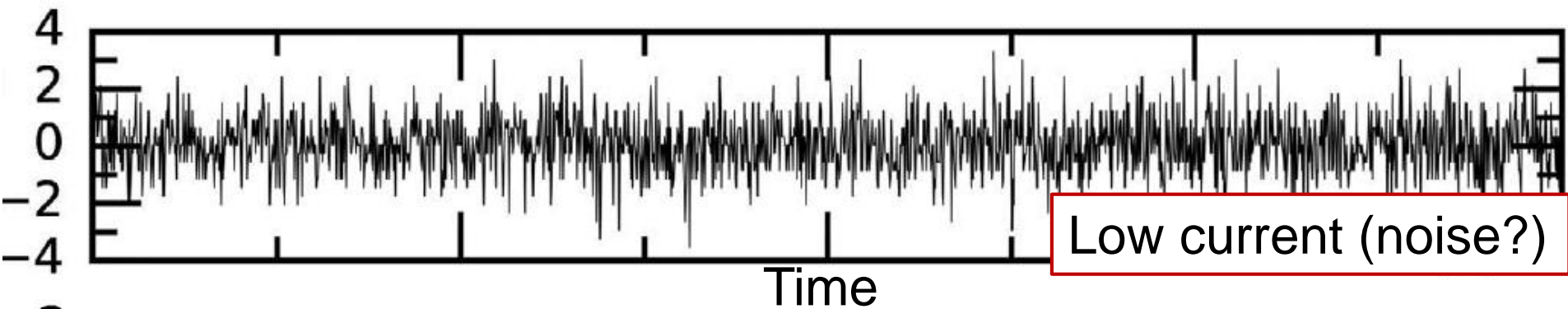


Hugo L. D. de S. Cavalcante et al PRL 111, 198701 (2013)

# Challenges in the analysis of the distribution of data values

- How to “bin” data (linear vs. logarithmic bins, how to choose the size of the bins)
- Number of data points in each bin (minimum average 10)
- How to fit the distribution (assuming a statistical model, the *maximum likelihood estimator* selects the parameter values which give the observed data the largest probability).
- Is the fit appropriated for the data? Many *goodness of fit* tests.

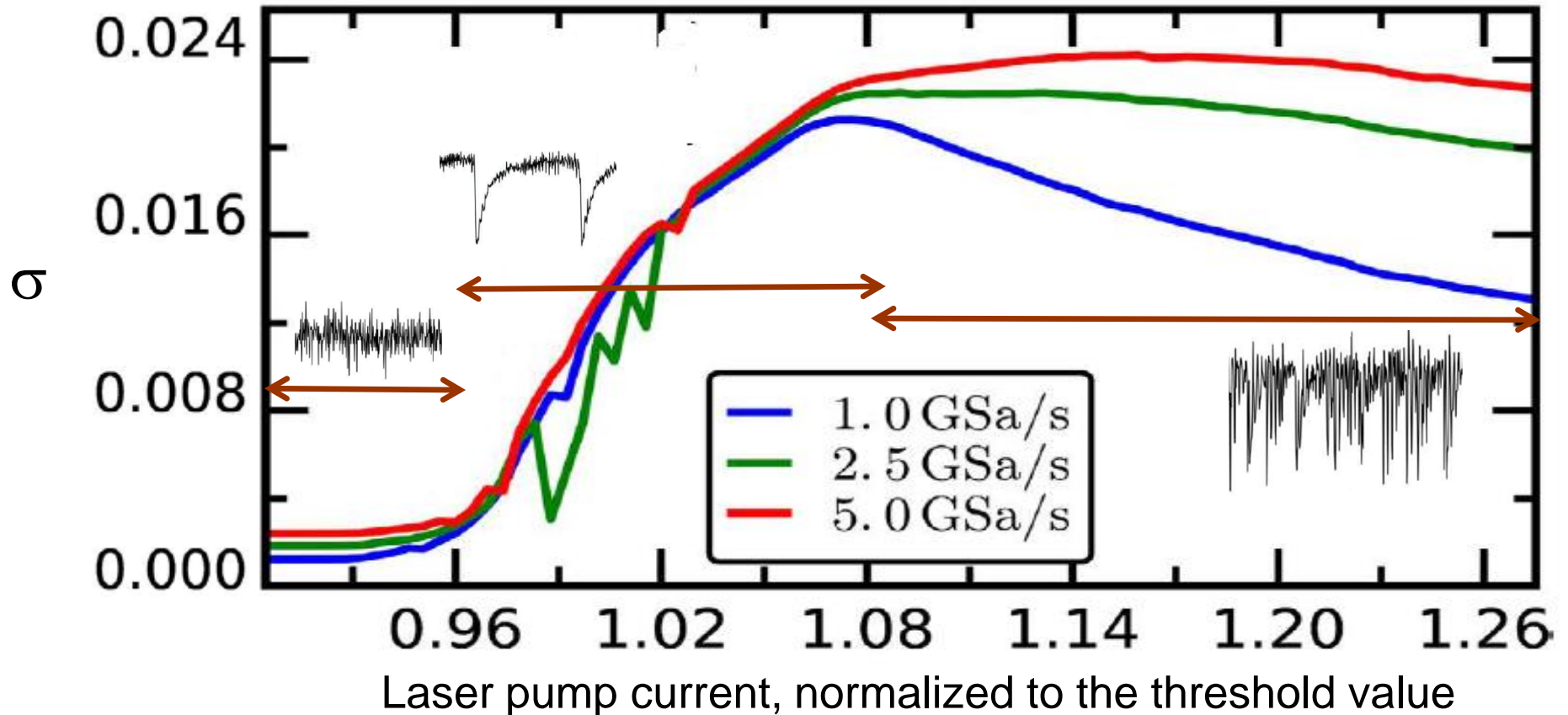
# Example: intensity emitted by a diode laser with feedback, as the pump current increases



Can different regimes be identified? With what reliability?



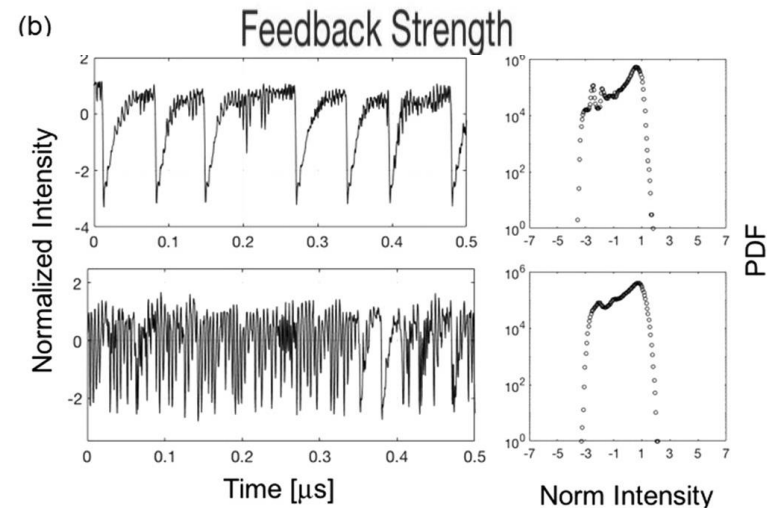
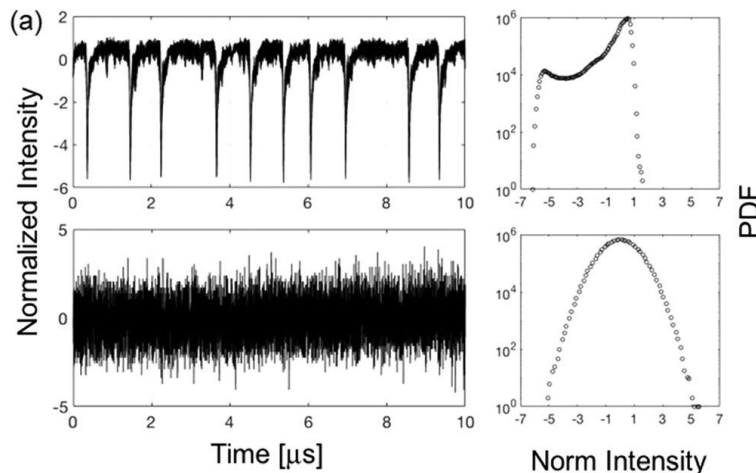
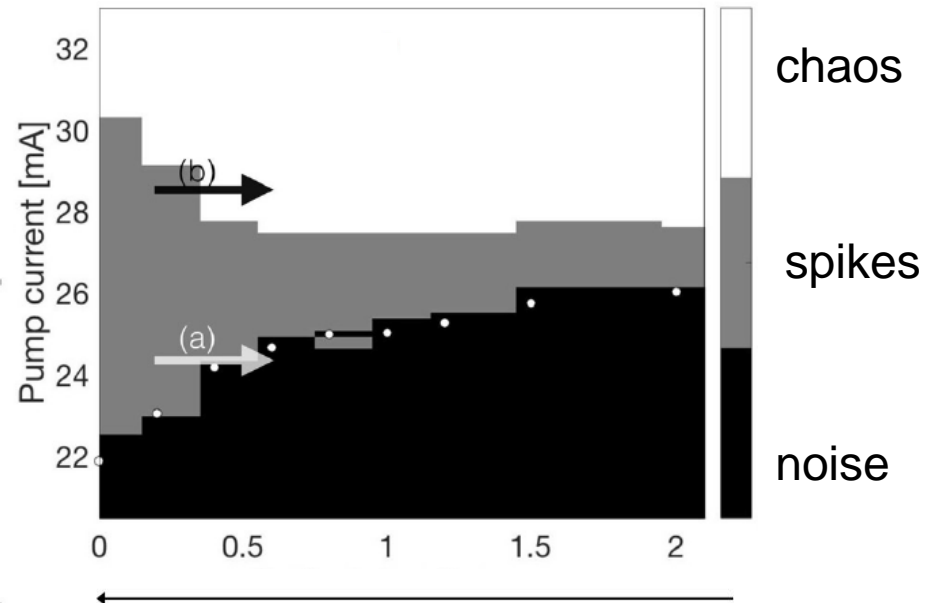
# Standard deviation of the intensity time series, $\sigma$ , recorded using three different sampling rates



# Classification criteria

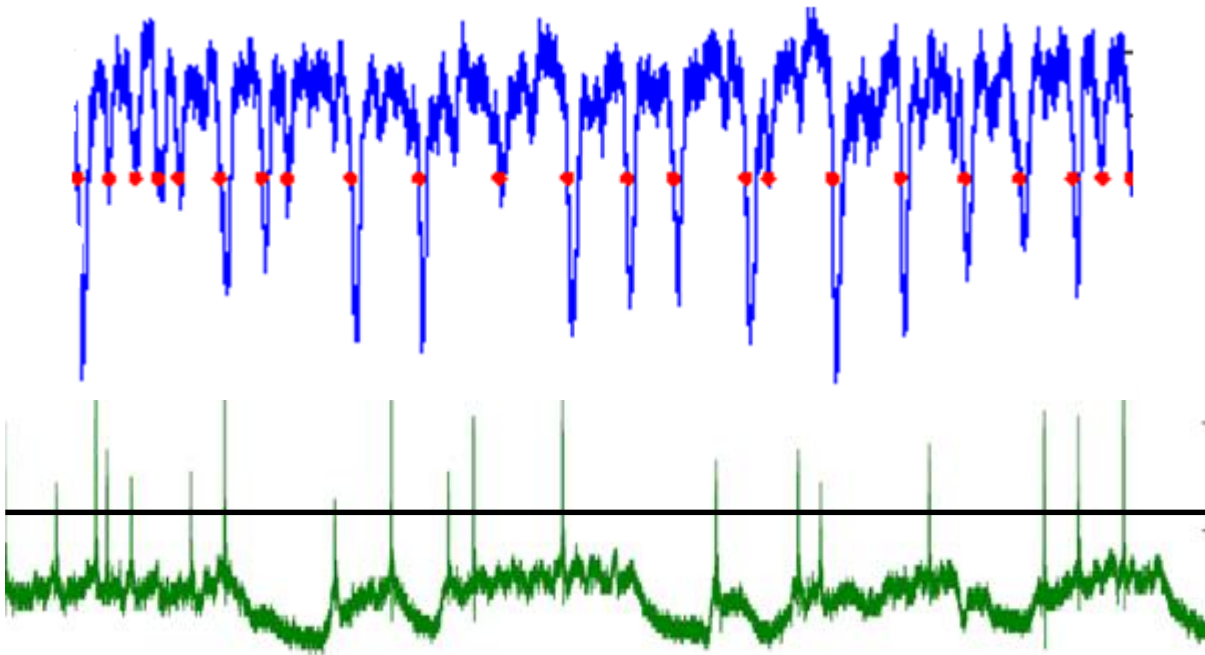
$I = \{i_1, i_2, \dots, i_N\}$  laser “raw” intensity recorded for different pump currents and feedback strengths

- If the skewness and kurtosis of the distribution of values are “consistent” with a Gaussian → noise
- Else, if  $\sigma$  increases with the pump current → spikes
- Else → chaos



# From time series to sequence of events

## Threshold crossing ``events``

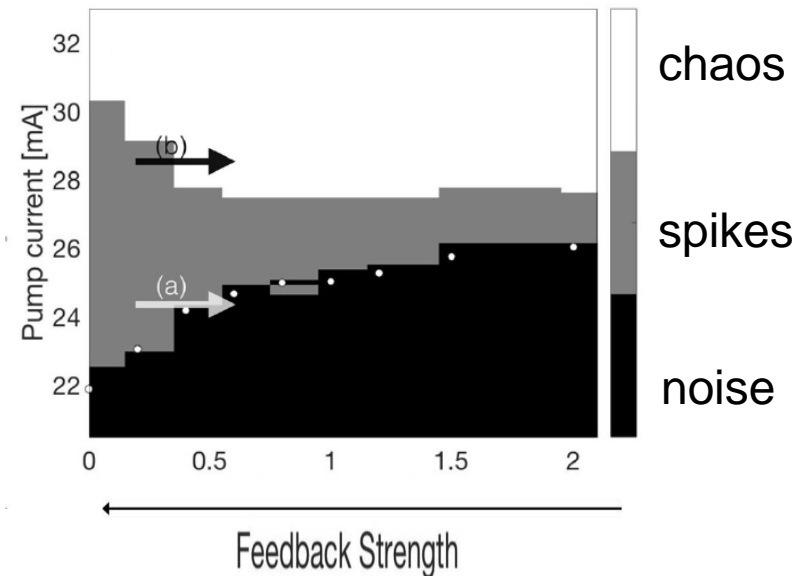
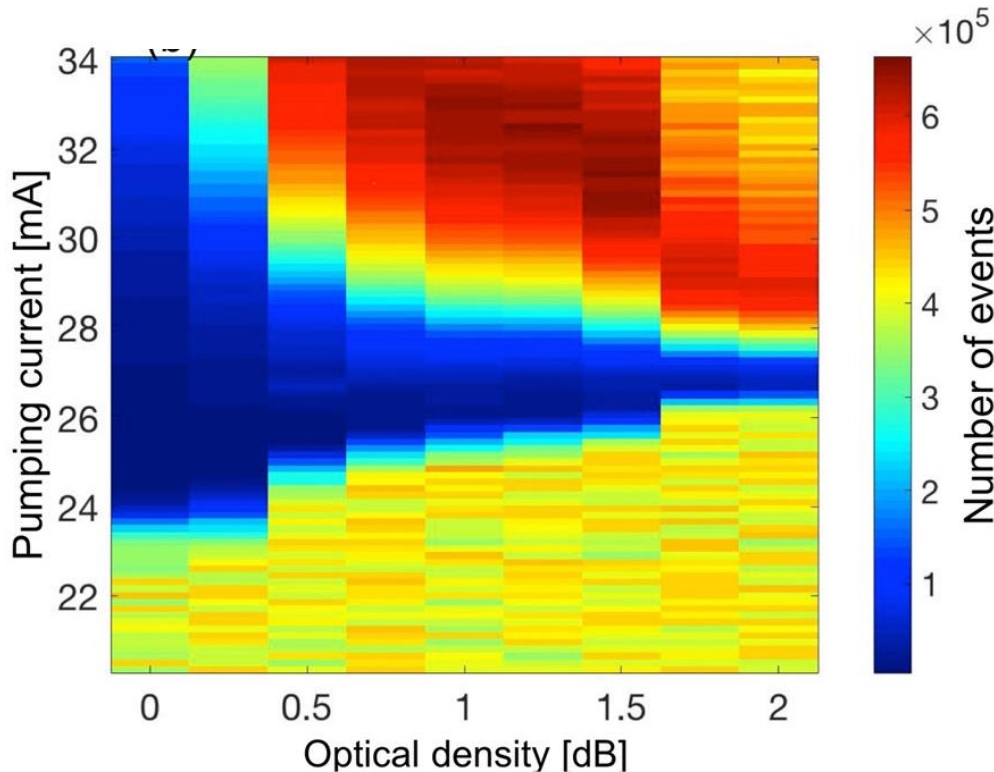
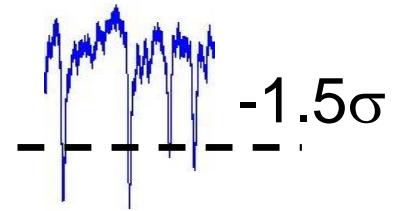


inter-spike-intervals (ISIs):  $\Delta T_i = t_{i+1} - t_i$

- Problems:
  - How to select the threshold?
  - Threshold dependent results?

“Features” that persist for a wide range of thresholds are “true” features.

**Example: by counting the number of times the laser intensity falls below a threshold we can distinguish the three regimes.**



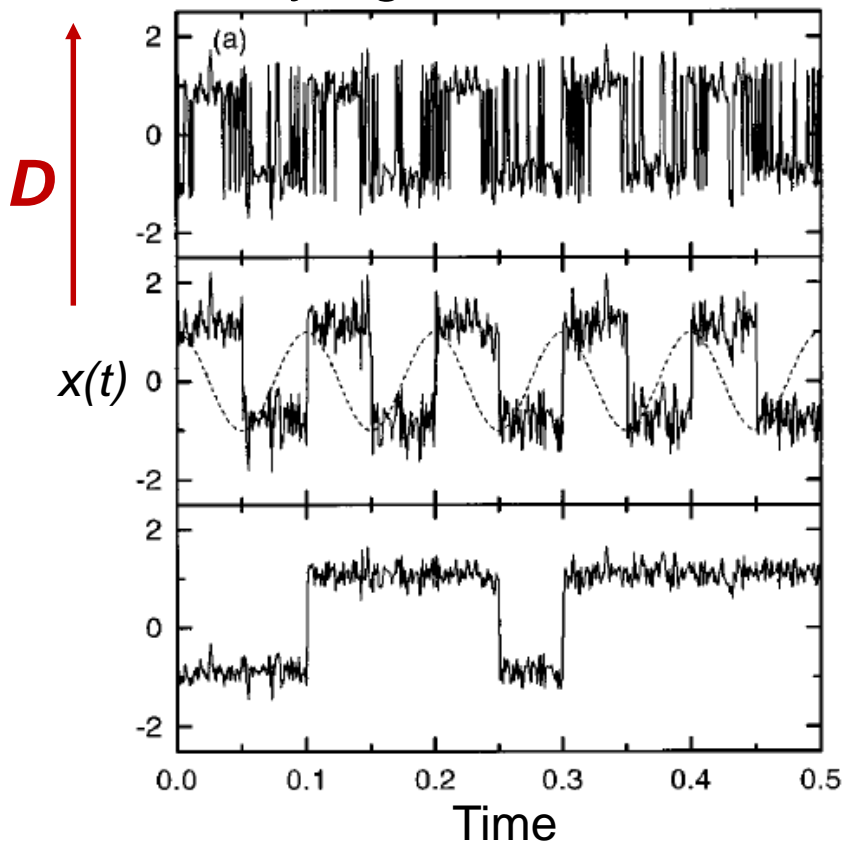
# **Stochastic and coherence resonances**

# Bistable system with sinusoidal forcing and noise

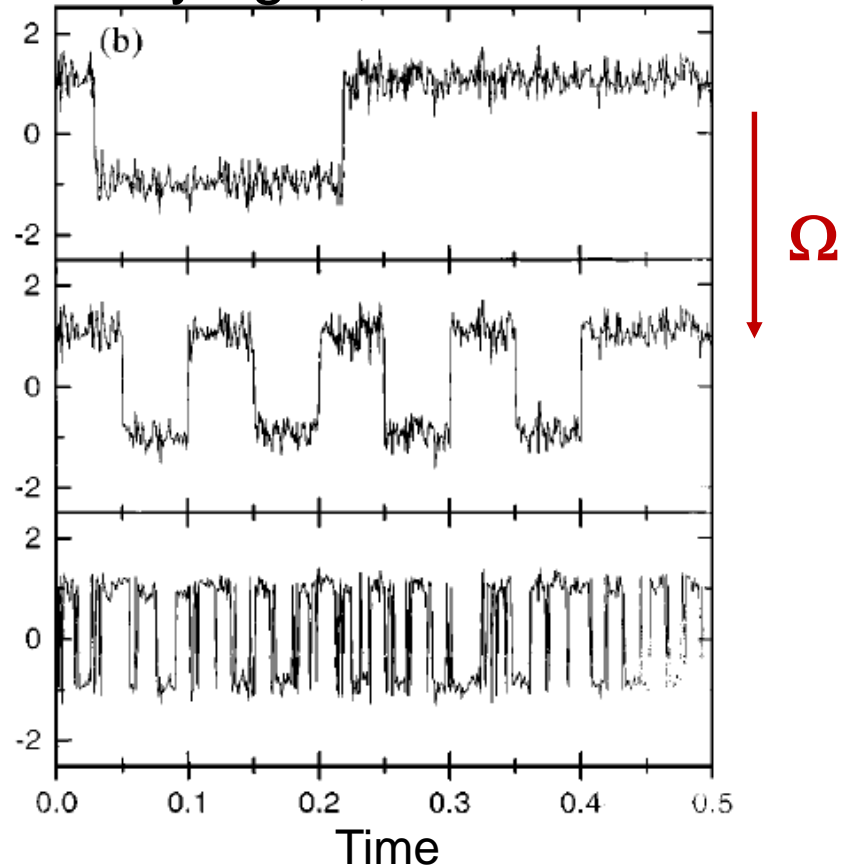
$$\dot{x}(t) = -V'(x) + A_0 \cos(\Omega t + \varphi) + \xi(t)$$

$$V(x) = -\frac{a}{2} x^2 + \frac{b}{4} x^4 \quad \langle \xi(t) \xi(0) \rangle = 2D \delta(t)$$

Varying  $D$ ;  $\Omega$  constant

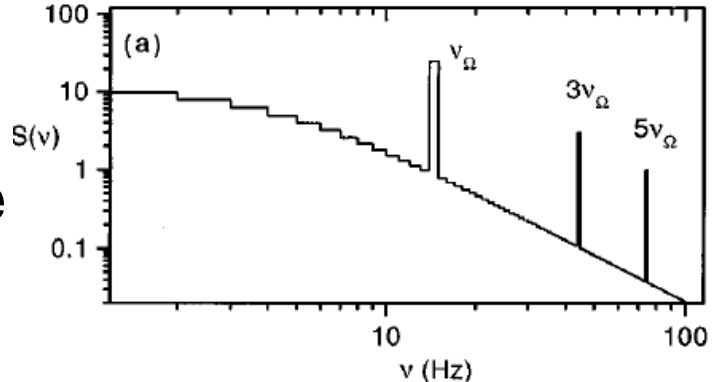


Varying  $\Omega$ ;  $D$  constant



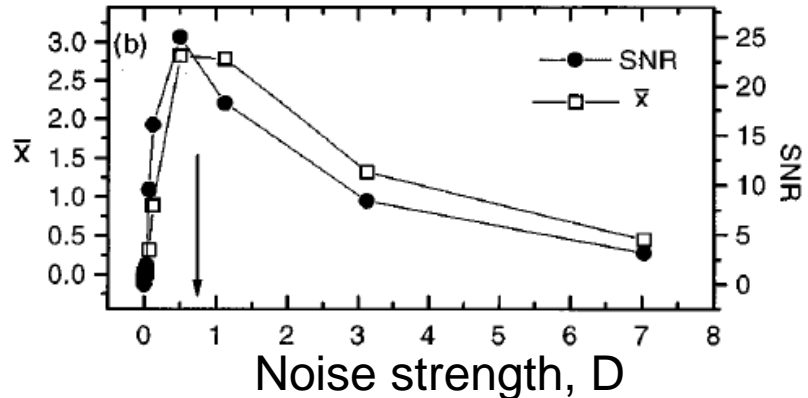
# Quantification of stochastic resonance with spectral analysis

- Phase-averaged power spectral density: average over many realizations of the noise and average over the input initial phase  $\varphi$ .

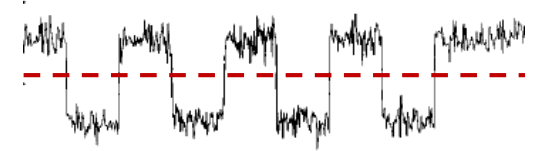


- Signal to noise ratio at  $\Omega$

$$SNR(\Omega) = 2 \left[ \lim_{\Delta\omega \rightarrow 0} \int_{\Omega - \Delta\omega}^{\Omega + \Delta\omega} S(\omega) d\omega \right] / S_N(\Omega)$$

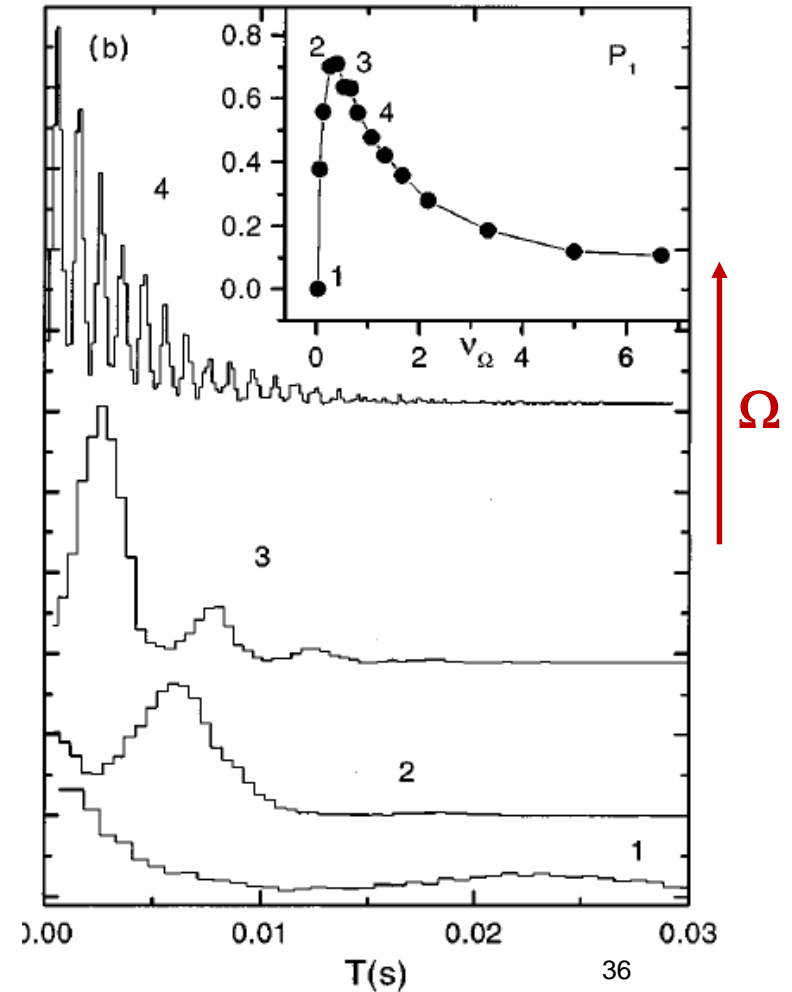
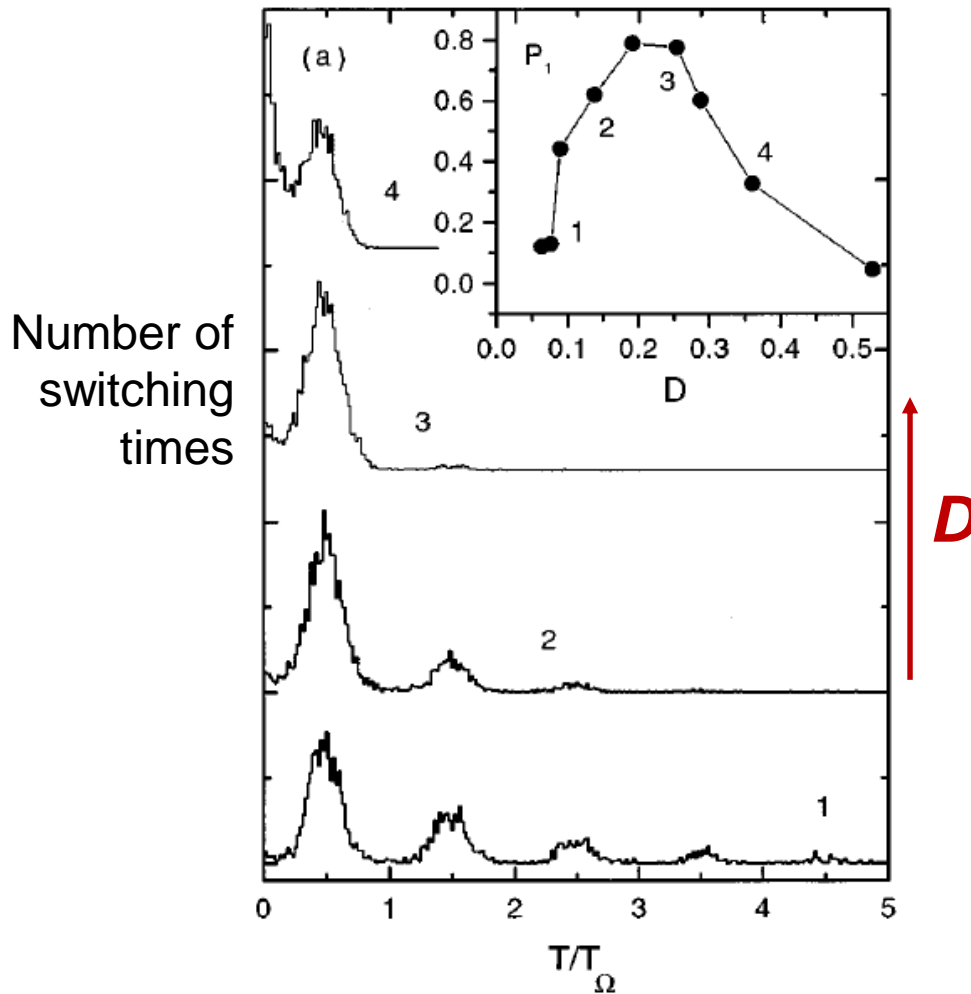


# Quantification of stochastic resonance with the *distribution of residence times*



Strength of the  $n$ th peak: area under the peak

$$P_n = \int_{T_n - \alpha T_\Omega}^{T_n + \alpha T_\Omega} N(T) dT \quad 0 < \alpha \leq \frac{1}{4}$$



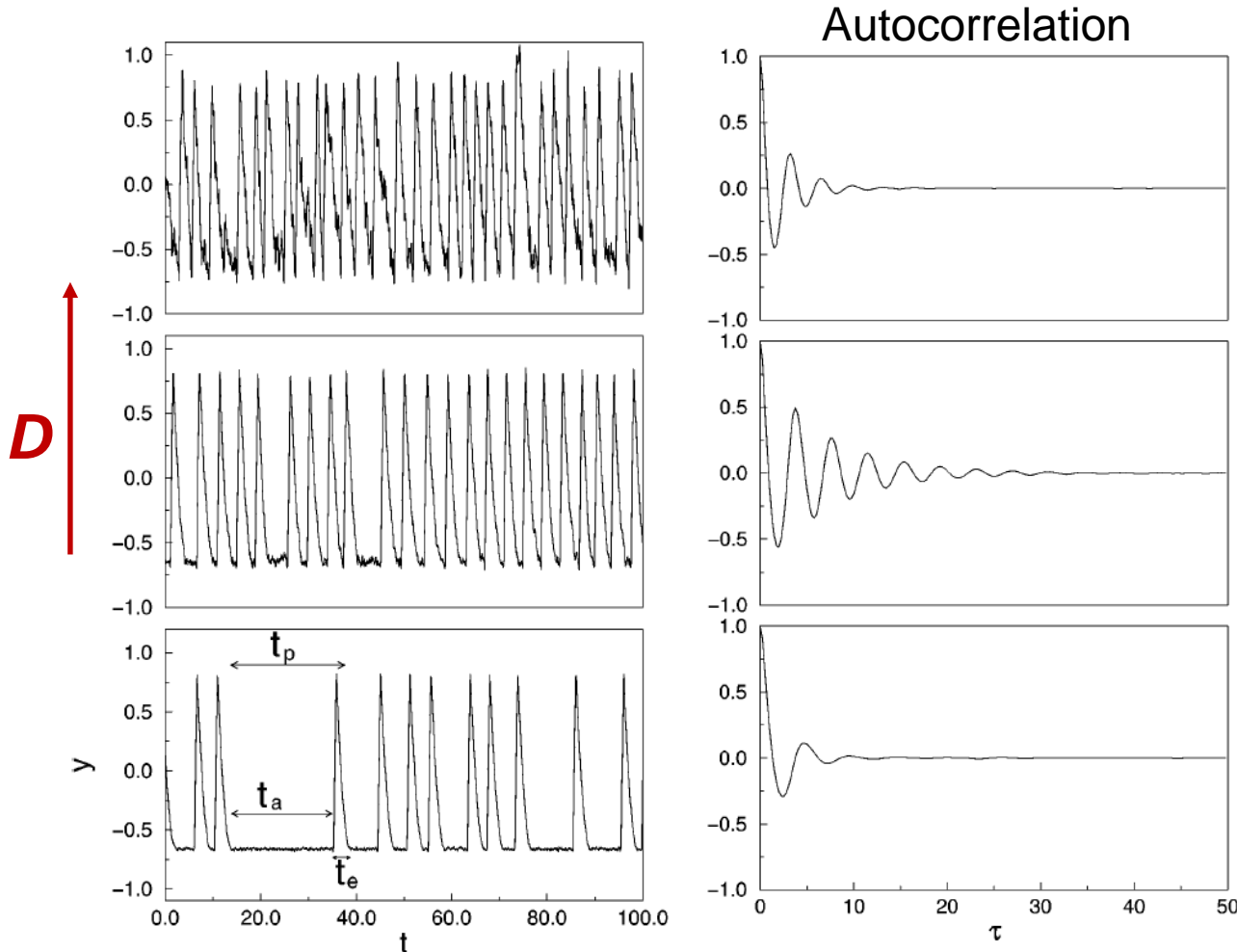


# Coherence resonance

$$\varepsilon \frac{dx}{dt} = x - \frac{x^3}{3} - y,$$

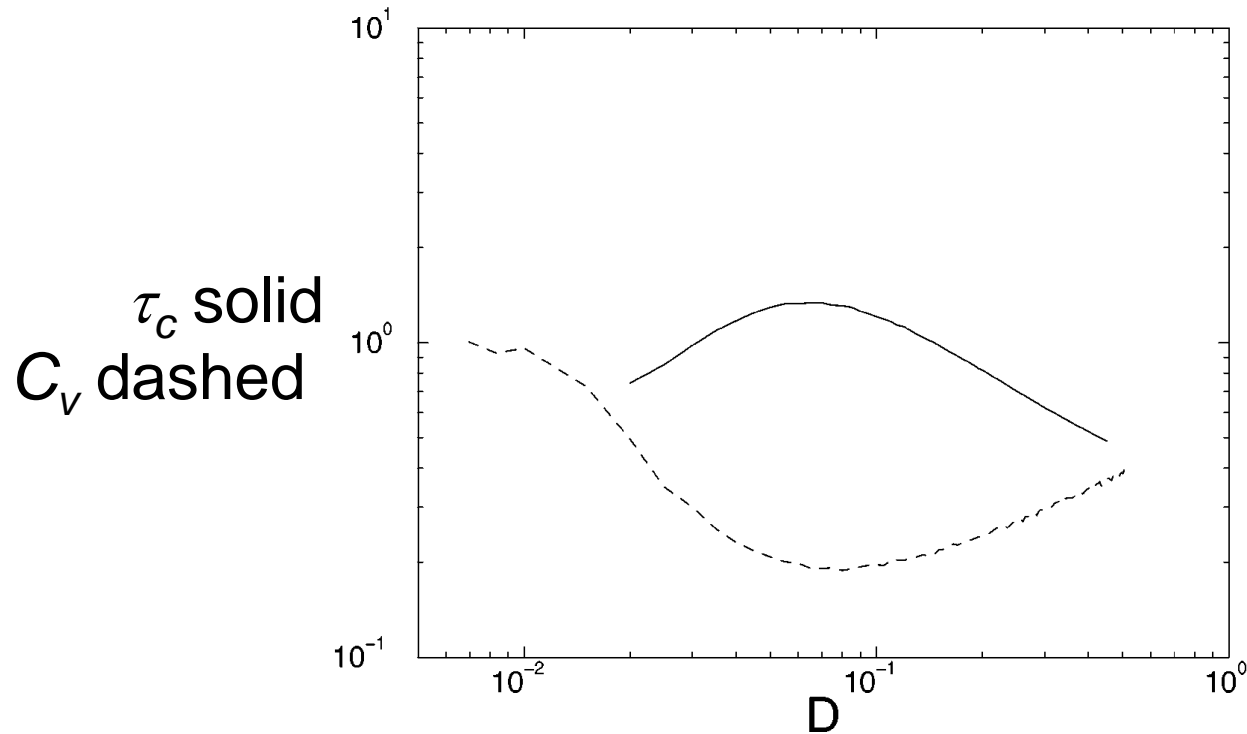
$$\frac{dy}{dt} = x + a + D\xi(t)$$

- Fitz Hugh–Nagumo model ( $\varepsilon=0.01$ ,  $a=1.05$ )



# How to quantify coherence resonance?

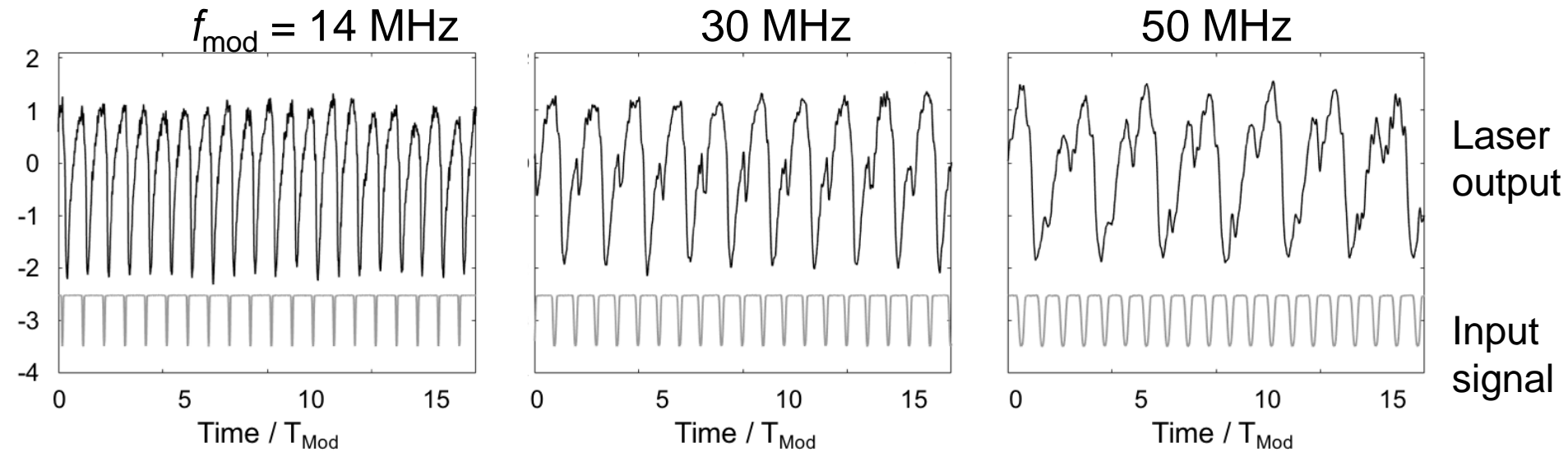
- Characteristic correlation time  $\tau_c = \int_0^\infty C^2(t) dt$
- Coefficient of variation of inter-spike-interval distribution  $C_V = \sigma/\mu$



# Locking

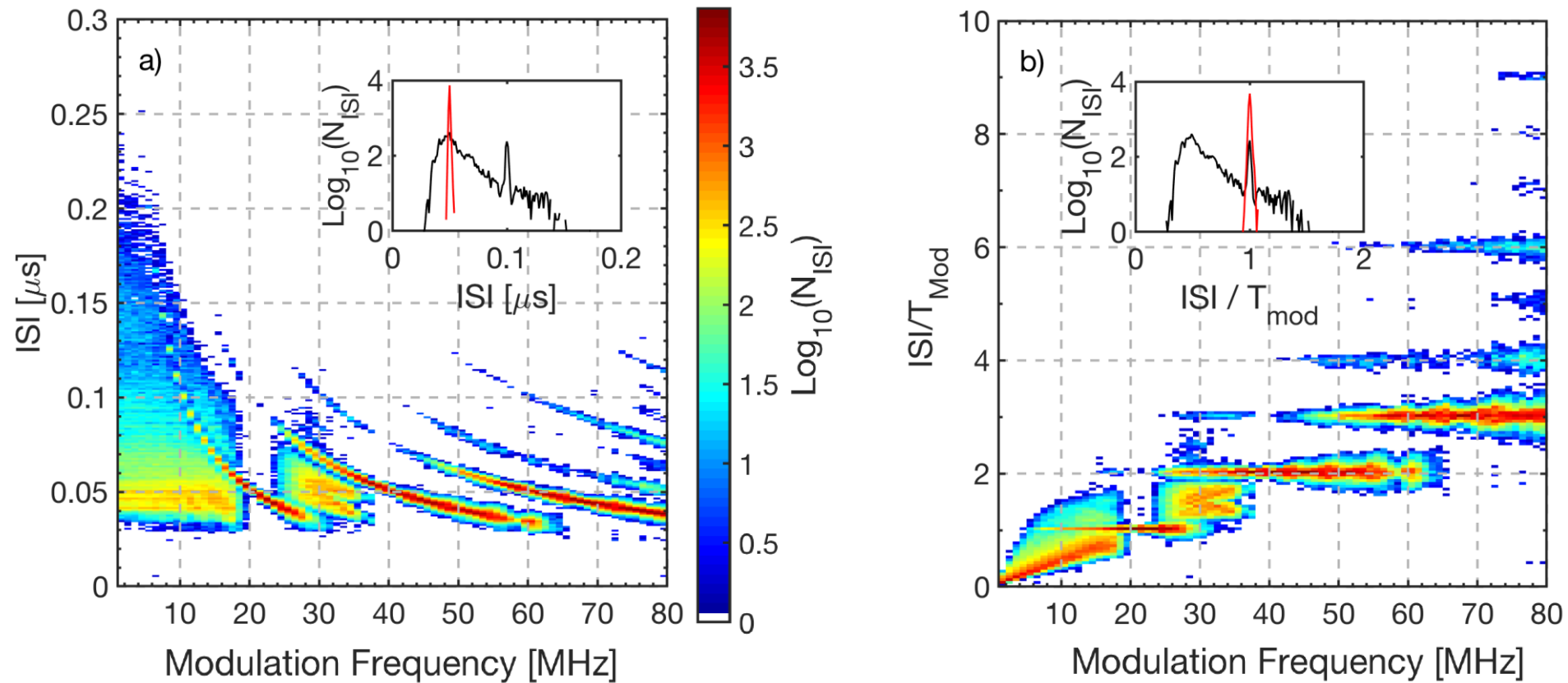
The frequency of an oscillator is controlled by an external signal

# Experimental data: the laser pump current is modulated with a small-amplitude sinusoidal signal



[Tiana et al, Opt. Express 26, 323041 \(2018\)](#)

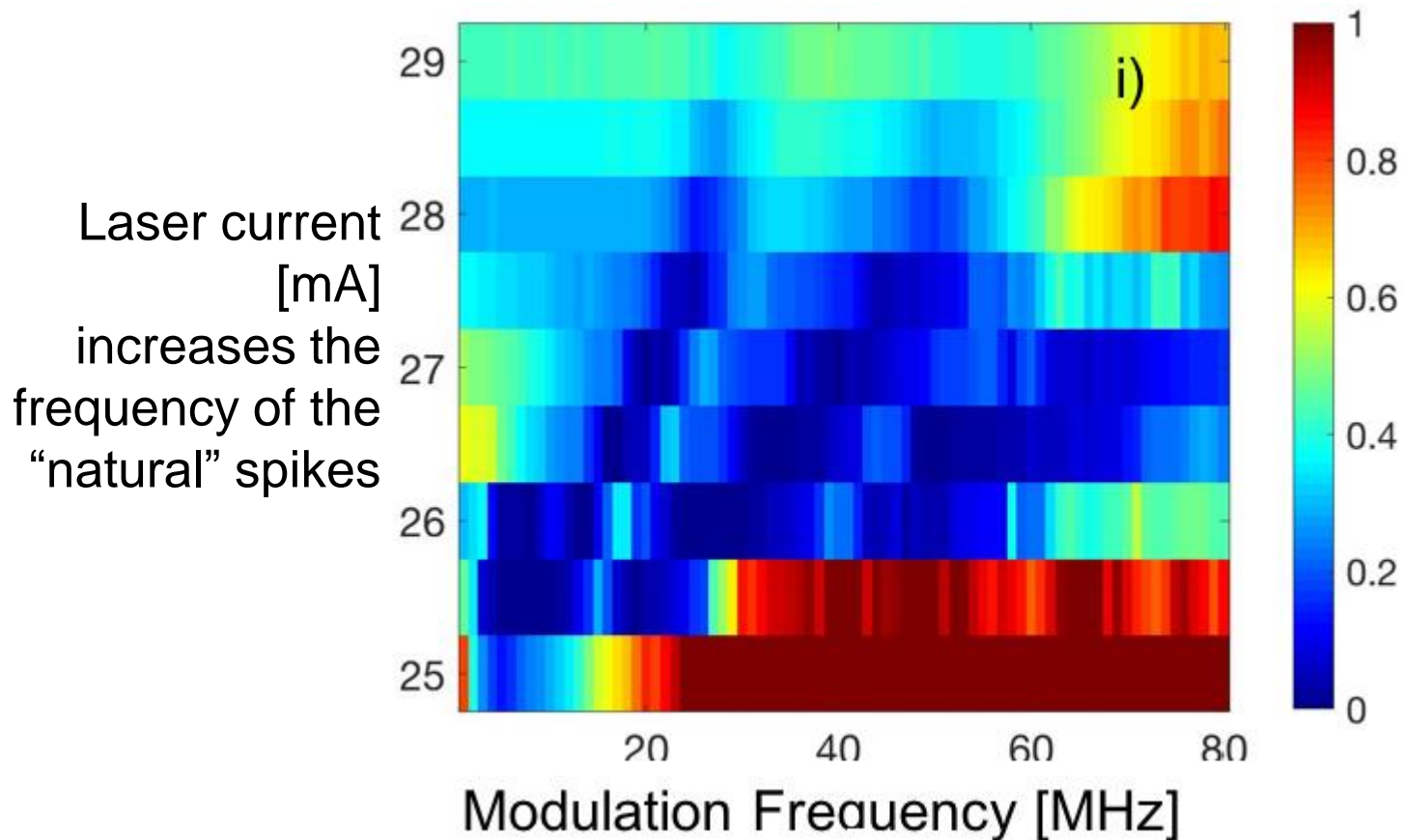
# How does the distribution of inter-spike intervals look like?



[Tiana et al, arXiv:1806.08950v1 \(2018\)](https://arxiv.org/abs/1806.08950v1)

# How to quantify the degree of locking?

Coefficient of variation of the ISI distribution  $C_V = \sigma/\mu$



[Tiana et al, Opt. Express 26, 323041 \(2018\)](#)

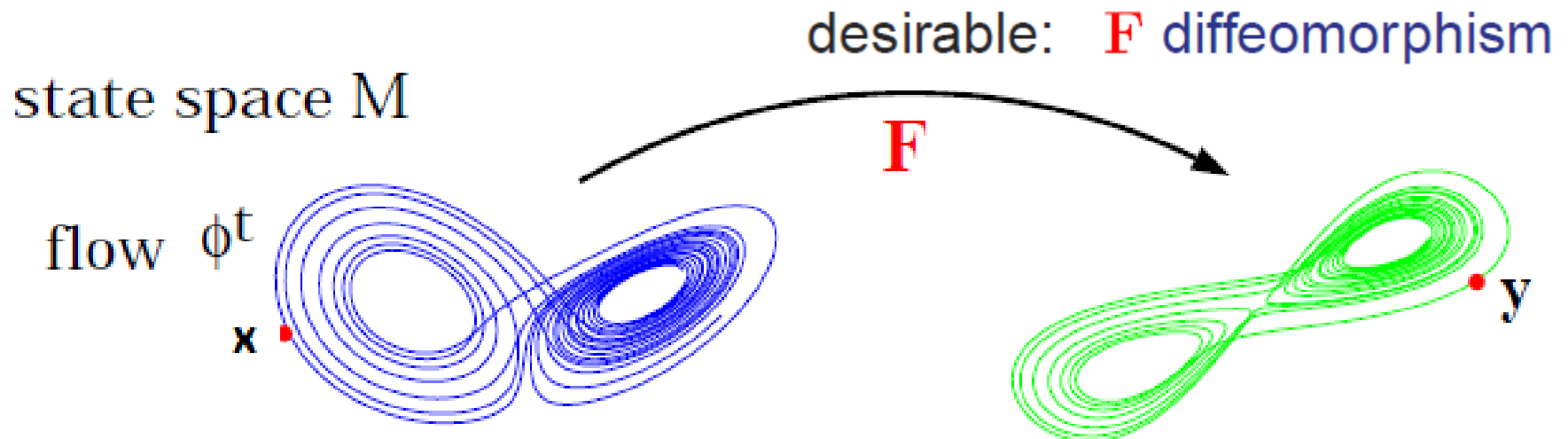
**Chaos**

# How to identify (and quantify) chaos in observed data?

Observed time series  $S = \{s(1), s(2), \dots, s(t) \dots\}$

Attractor reconstruction: “embed” the time series in a phase-space of dimension  $d$  using delay  $\tau$  coordinates

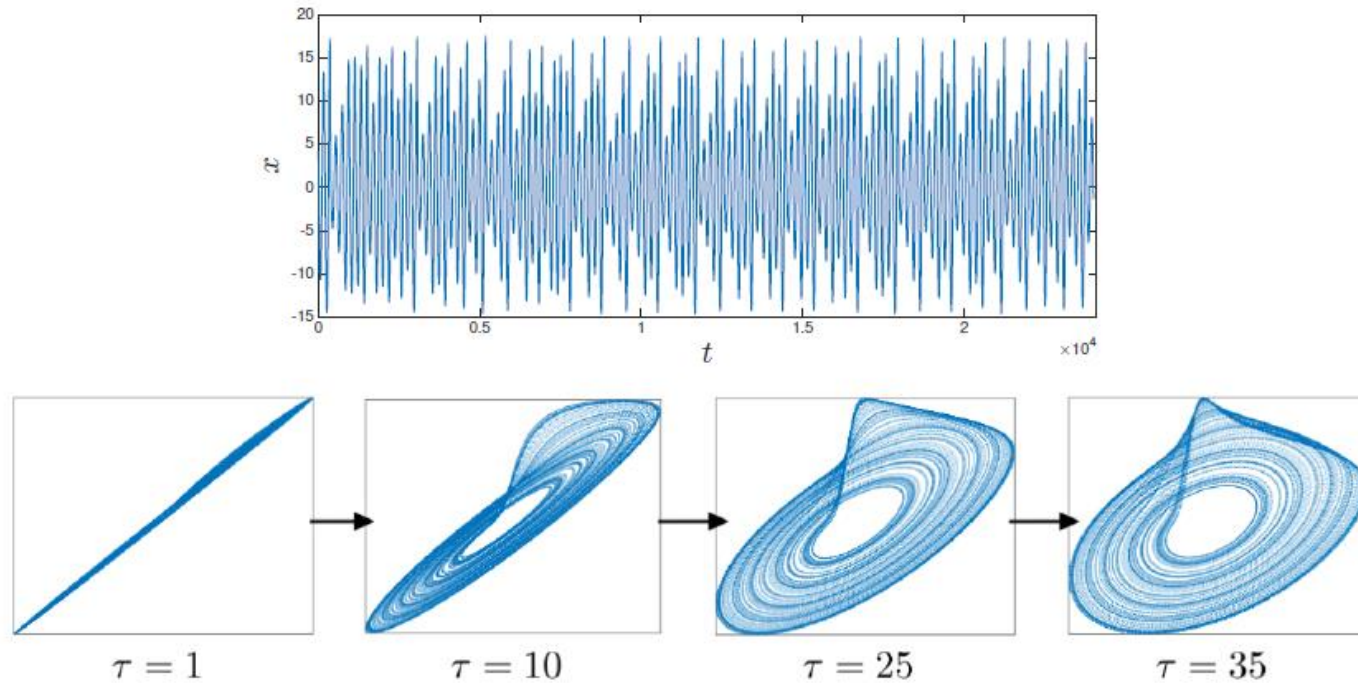
$$y(t) = (s(t), s(t+\tau), \dots, s(t+(d-1)\tau))$$



Adapted from U. Parlitz (Gottingen)



# Reconstruction using delay coordinates



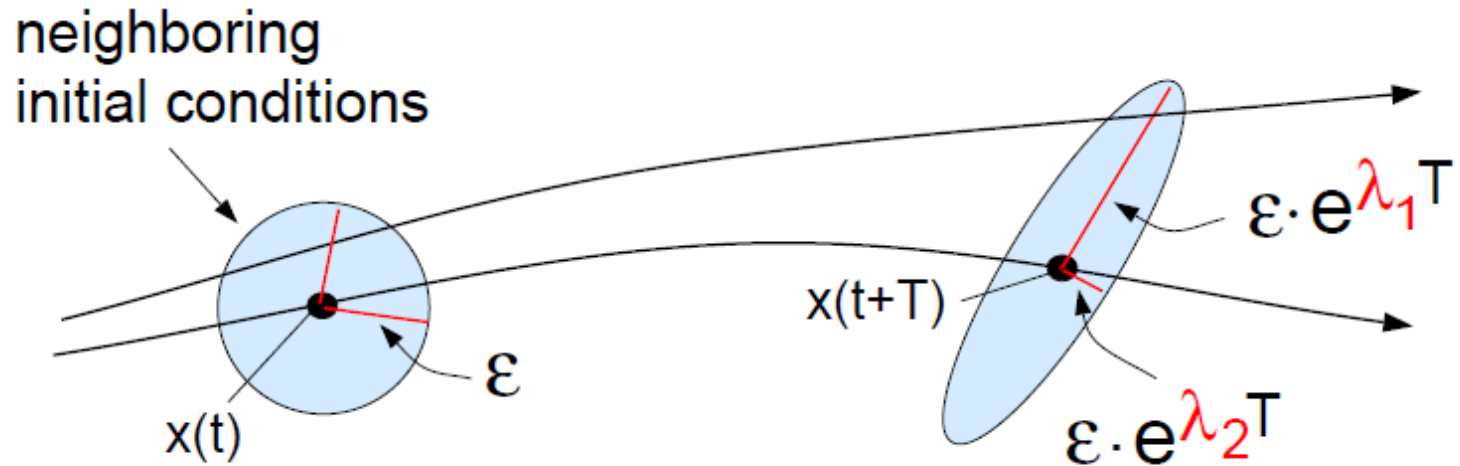
A problem: how to choose the embedding parameters  
(lag  $\tau$ , dimension  $d$ )

## How to choose the lag $\tau$ and the dimension $d$

- $\tau$  is chosen to maximize the spread of the data in phase space: the first zero of the autocorrelation function (or where  $|C(\tau)|$  is minimum)
- $d$  is often estimated with the *false nearest neighbors* technique that examines how close points in phase space remain close as the dimension is increased.
- Points that do not remain close are ‘false’ neighbors.
- The number of false neighbors decreases as the embedding dimension is increased.
- The first dimension for which the number of false neighbors decreases below a threshold provides the estimated  $d$ .

After reconstructing the attractor, we can characterize the TS by the **fractal dimension** and the **Lyapunov exponent**.

# Lyapunov exponents: measure how fast neighboring trajectories diverge



- A stable fixed point has negative  $\lambda$ s (since perturbations in any direction die out)
- An attracting limit cycle has one zero  $\lambda$  and negative  $\lambda$ s
- A chaotic attractor has at least one positive  $\lambda$ .

## Steps to compute the maximum LE

- Initial distance  $\delta_I = |s_i - s_j|$
- Final distance  $\delta_F = |s_{i+T} - s_{j+T}|$
- Local *exponential* grow  $\lambda_{\text{local}}^* = \frac{1}{T} \log(\delta_F / \delta_I)$
- The rate of grow is averaged over the attractor, which gives  $\lambda_{\text{max}}$

**A. Wolf, J.B. Swift, H.L. Swinney and J.A. Vastano, Physica D 16 (1985)**

A very popular method for detecting chaos in experimental time series.

## A word of warning!

- The rate of grow depends on the direction in the phase space.
- The algorithm returns the value in the direction with fastest expansion.
- Therefore, the algorithm always returns a positive number!
- This is a problem when computing the LE of noisy data.
- First we need to test nonlinearity in the time series.

### Further reading:

- F. Mitschke and M. Damming, *Chaos vs. noise in experimental data*, Int. J. Bif. Chaos 3, 693 (1993)
- A. Pikovsky and A. Politi, *Lyapunov Exponents* (Cambridge University Press, 2016)

# Grassberger-Procaccia correlation dimension algorithm

- Another very popular method for detecting chaos in real-world data.
- Fractal dimension (box counting dimension):  $D_0 = \lim_{\epsilon \rightarrow 0} \frac{\log N(\epsilon)}{\log \frac{1}{\epsilon}}$
- Problem: for time-series analysis, no distinction between frequent and unfrequently visited boxes.

- An alternative: the *correlation dimension*.  
 $g$  is the number of pairs of points with distance between them  $< \epsilon$ .

For any set of  $N$  points in an  $m$ -dimensional space

$$\vec{x}(i) = [x_1(i), x_2(i), \dots, x_m(i)], \quad i = 1, 2, \dots, N$$

then the correlation integral  $C(\epsilon)$  is calculated by:

$$C(\epsilon) = \lim_{N \rightarrow \infty} \frac{g}{N^2} \quad \boxed{C(\epsilon) \sim \epsilon^\nu}$$

Further reading:

P. Grassberger and I. Procaccia, "Measuring the Strangeness of Strange Attractors".  
Physica D vol. 9, pp.189, 1983.

L. S. Liebovitch and T. Toth, "A fast algorithm to determine fractal dimensions by box counting," Physics Letters A, vol. 141, pp. 386, 1989.

# **Symbolic methods to identify patterns and structure in time series**

# Are there statistical similarities between neuronal spikes and optical spikes?

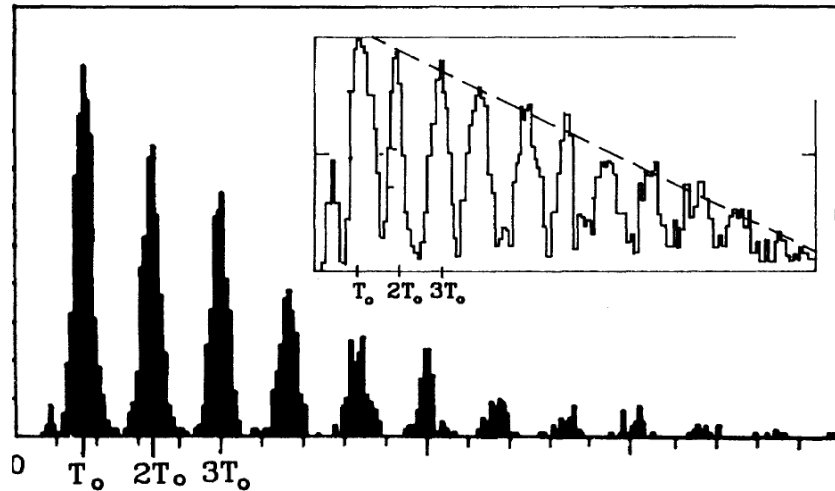
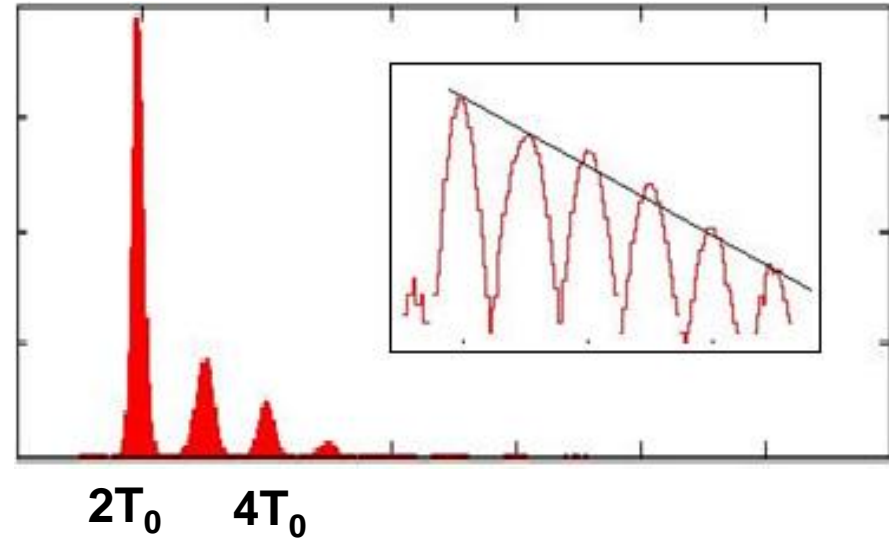


FIG. 1. (a) An experimental ISIH obtained from a single auditory nerve fiber of a squirrel monkey with a sinusoidal 80-dB sound-pressure-level stimulus of period  $T_0 = 1.66$  ms applied at the ear. Note the modes at integer multiples of  $T_0$ . Inset:

A. Longtin et al PRL (1991)



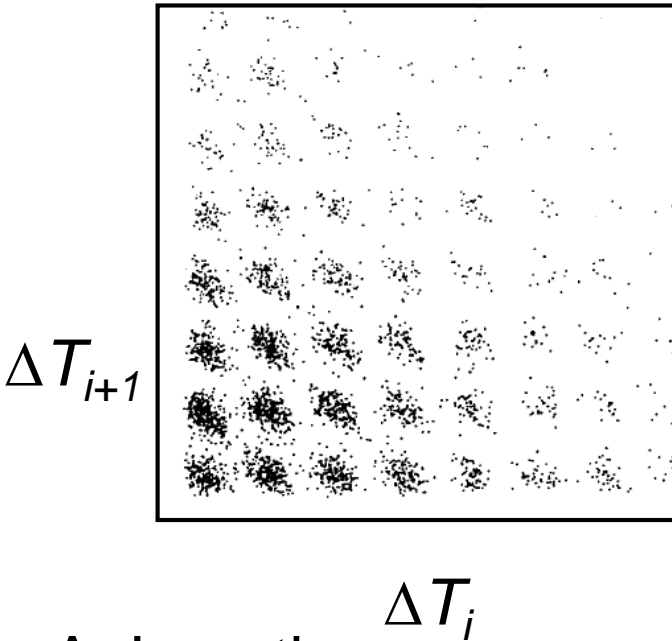
Experimental data when the laser current is modulated with a sinusoidal signal of period  $T_0$ .

[A. Aragonese et al](#)  
[Optics Express \(2014\)<sup>52</sup>](#)



# Return maps of inter-spike-intervals

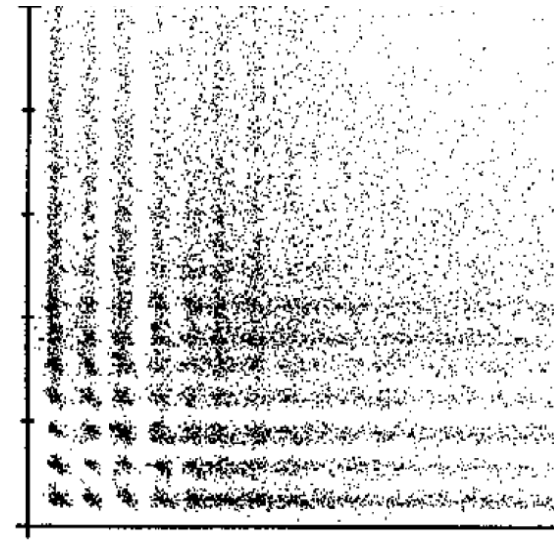
Neuronal ISIs



A. Longtin

Int. J. Bif. Chaos (1993)

Laser ISIs



M. Giudici et al PRE (1997)

[A. Aragonese et al  
Optics Express \(2014\)](#)

HOW TO IDENTIFY TEMPORAL ORDER?  
MORE/LESS EXPRESSED PATTERNS?

# Symbolic analysis

- The time series  $\{x_1, x_2, x_3, \dots\}$  is transformed (using an appropriated **rule**) into a sequence of symbols  $\{s_1, s_2, \dots\}$
- taken from an “**alphabet**” of possible symbols  $\{a_1, a_2, \dots\}$ .
- 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:  
if  $x_i > x_{th1} \Rightarrow s_i = 0$ ;  
else if  $x_i < x_{th2} \Rightarrow s_i = 2$ ;  
else ( $x_{th2} < x_i < x_{th1}$ )  $\Rightarrow s_i = 1$ .

## Alternative rule

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

also transforms a time-series into a sequence of 0s and 1s  
**without using a threshold**

- “words” of  $D$  letters are formed by considering the **order relation** between sets of  $D$  values  $\{\dots x_j, x_{j+1}, x_{j+2}, \dots\}$ .

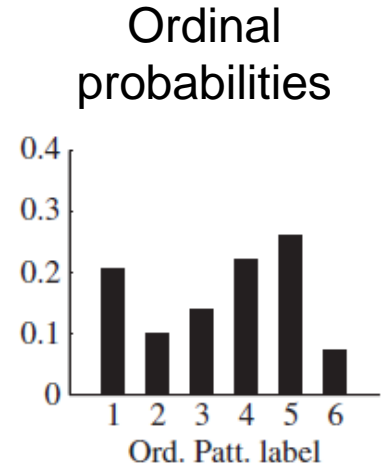
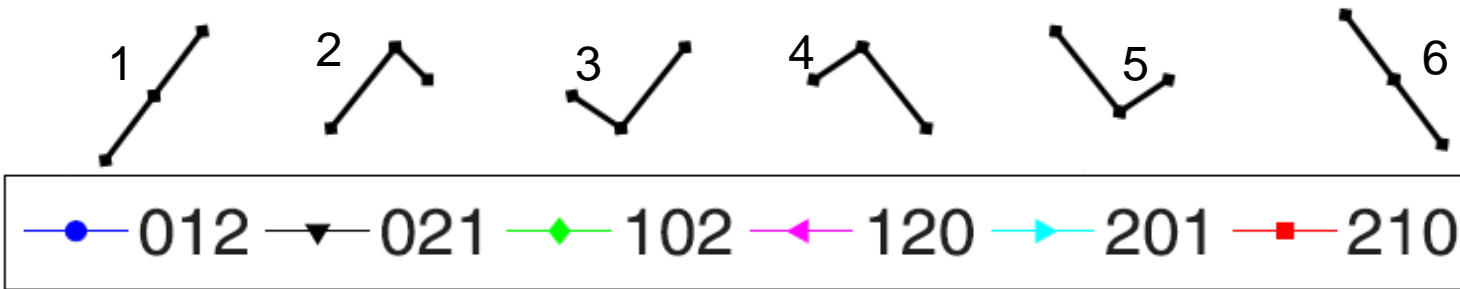


[Bandt and Pompe PRL 88, 174102 \(2002\)](#)

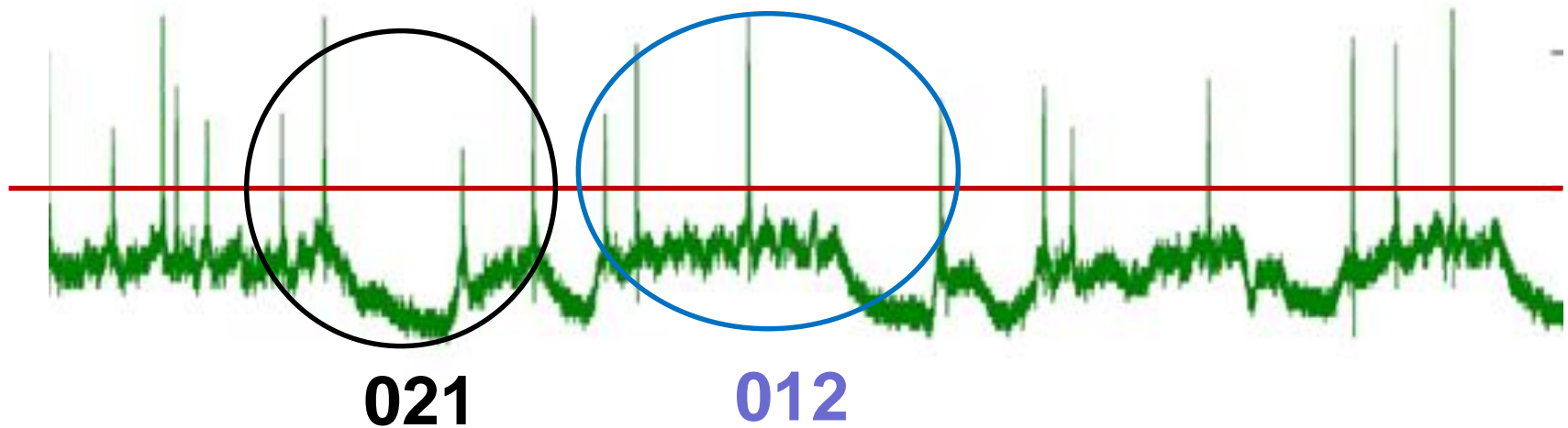
# Relative order of **three** consecutive intervals

$$l_i = t_{i+1} - t_i$$

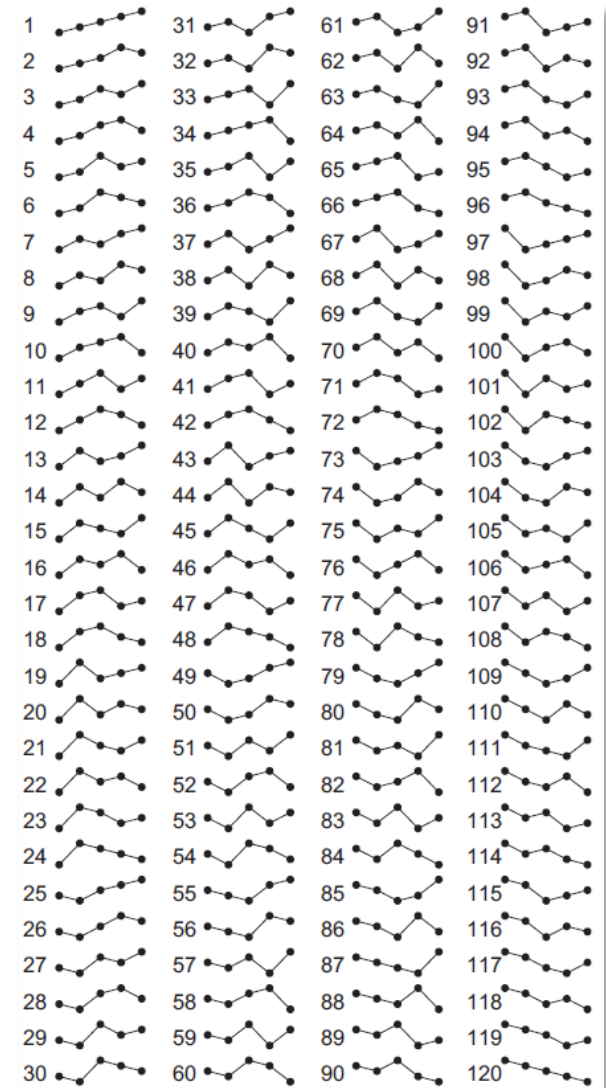
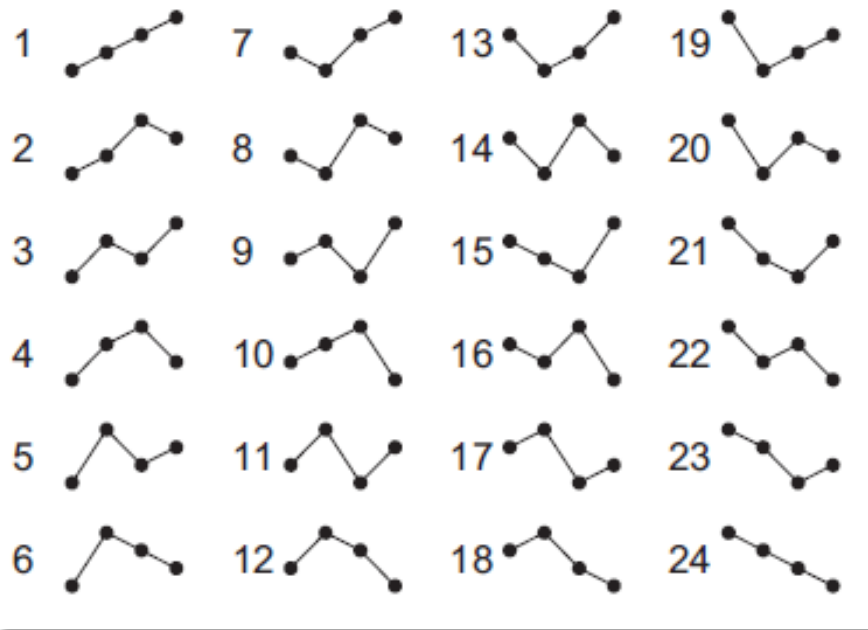
$$\{\dots l_i, l_{i+1}, l_{i+2}, \dots\}$$



Example: (5, 1, 7) gives “102” because  $1 < 5 < 7$



# 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

# Comparison

## Threshold transformation:

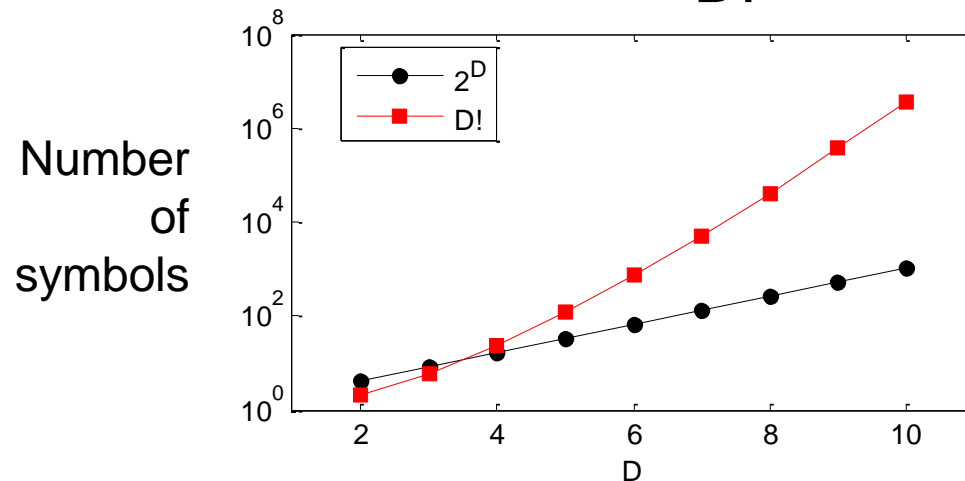
if  $x_i > x_{th} \Rightarrow s_i = 0$ ; else  $s_i = 1$

- 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:

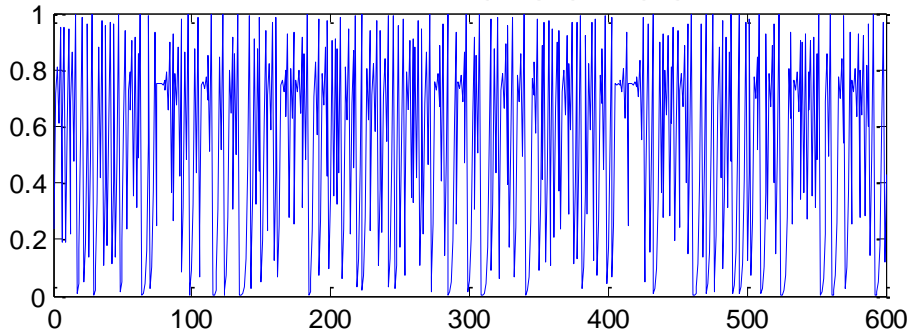
if  $x_i > x_{i-1} \Rightarrow s_i = 0$ ; else  $s_i = 1$

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

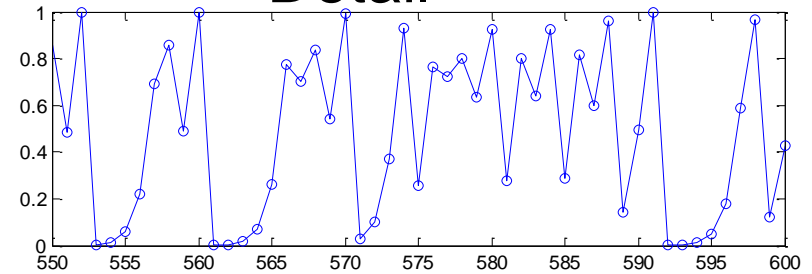


# Logistic map $x(i+1) = r x(i)[1 - x(i)]$

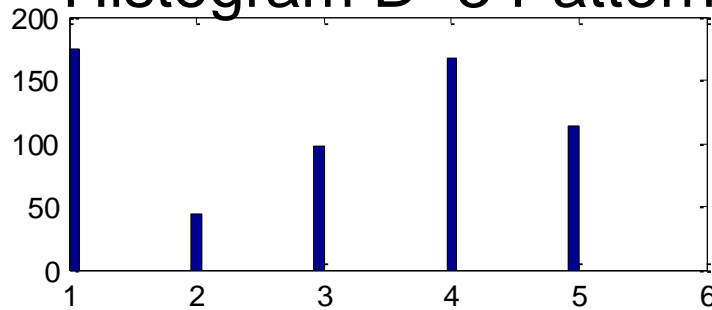
## Time series



## Detail

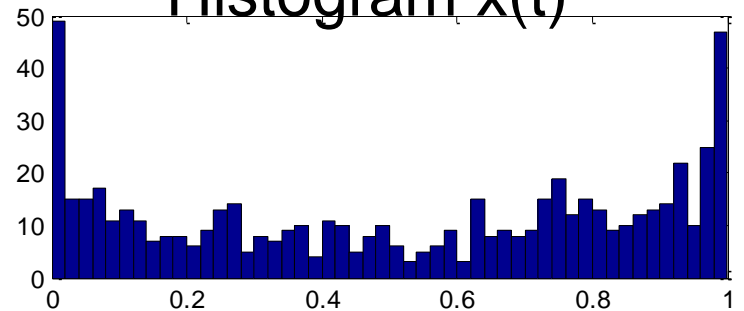


## Histogram D=3 Patterns



↑  
**forbidden**

## Histogram $x(t)$

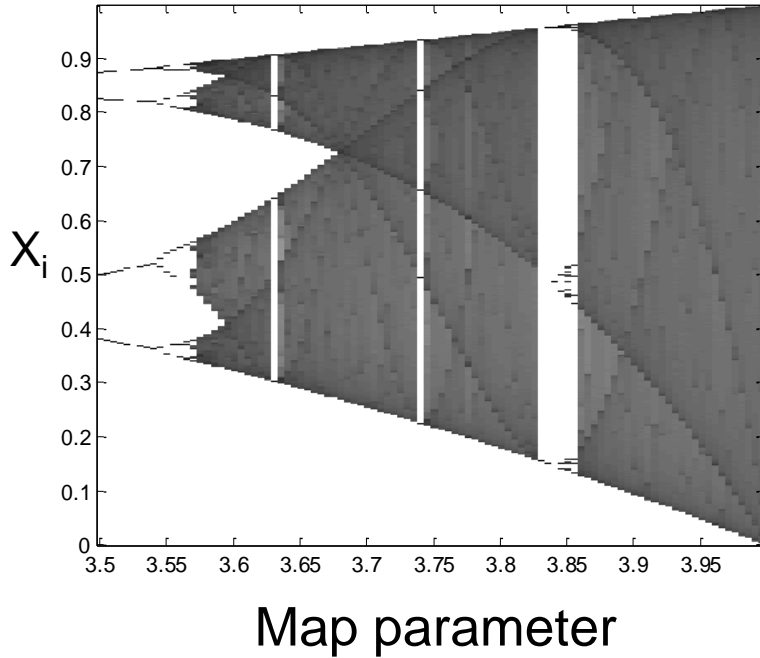


Ordinal analysis yields information about more and less expressed patterns in the data

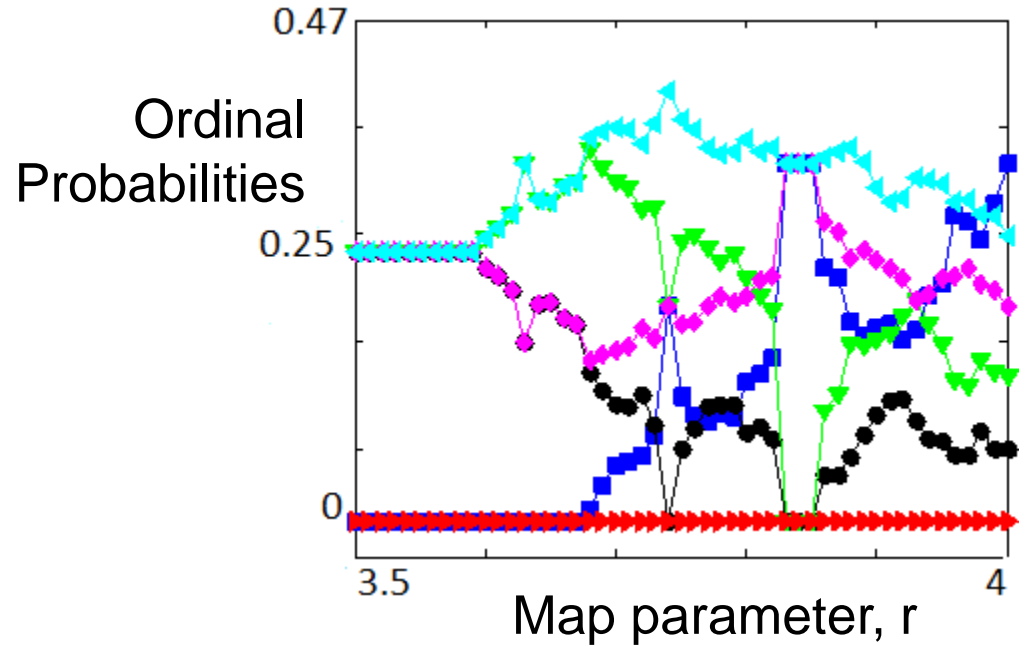


# Ordinal analysis yields complementary information

## Normal bifurcation diagram



## Ordinal bifurcation diagram (class and homework)

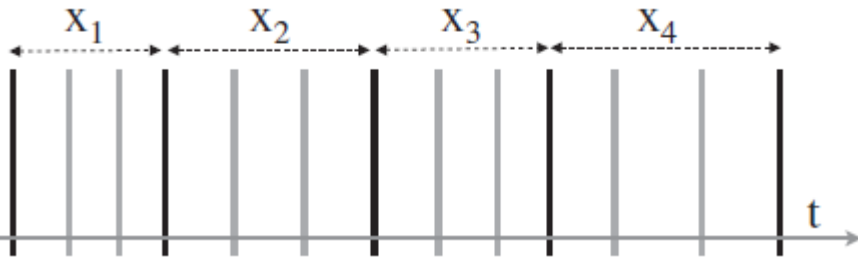
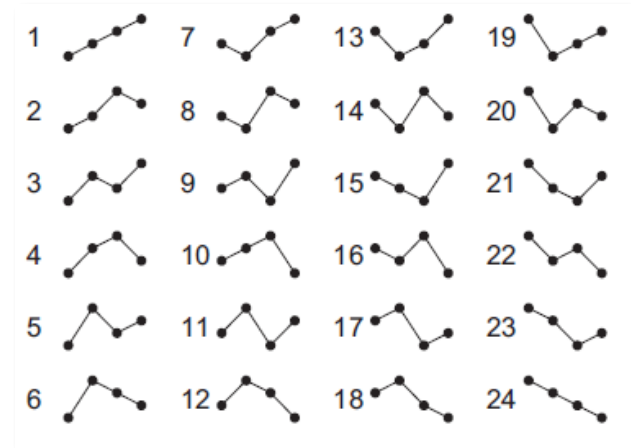


**012 021 102 120 201 210**

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

# 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 ;
```

# Are the $D!$ ordinal patterns equally probable?

- **Null hypothesis:**

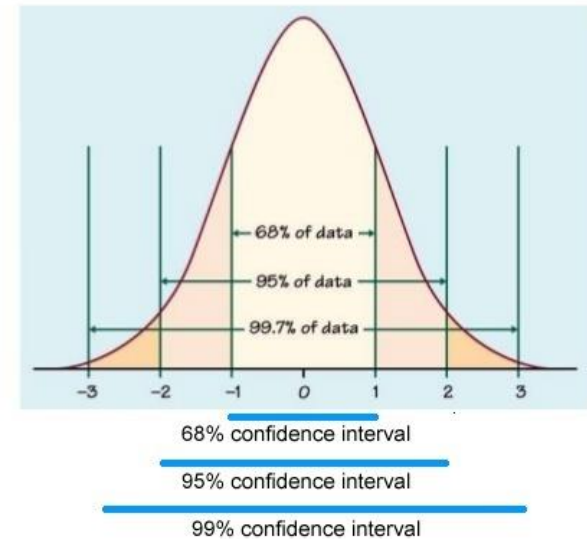
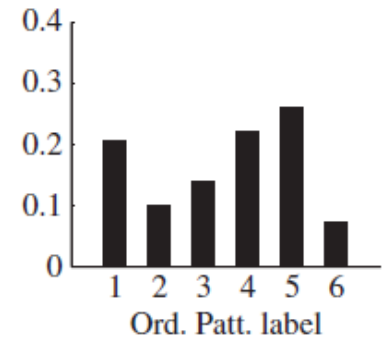
$$p_i = p = 1/D! \quad \text{for all } i = 1 \dots D!$$

- 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

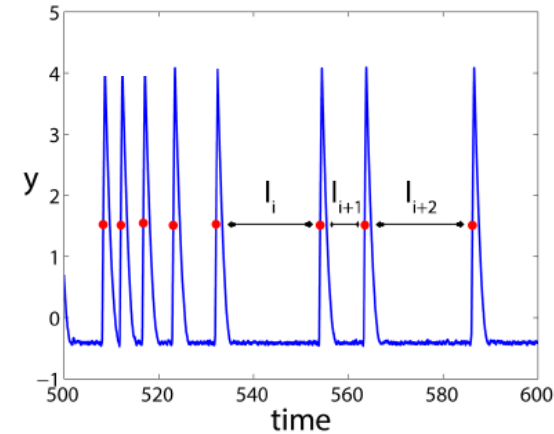
We **fail to reject** the NH with 99.74% confidence level.



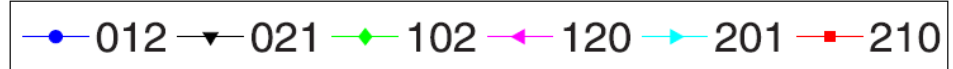
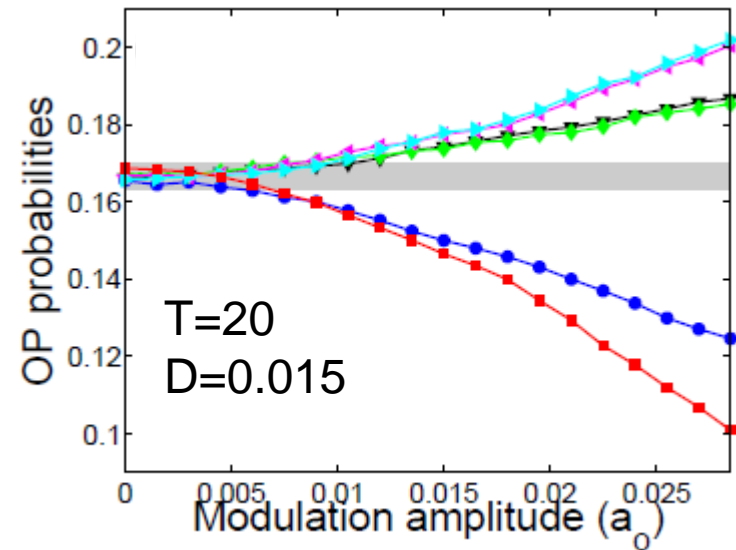
# Example: spikes generated by the FitzHugh-Nagumo model

$$\epsilon \frac{dx}{dt} = x - \frac{x^3}{3} - y,$$

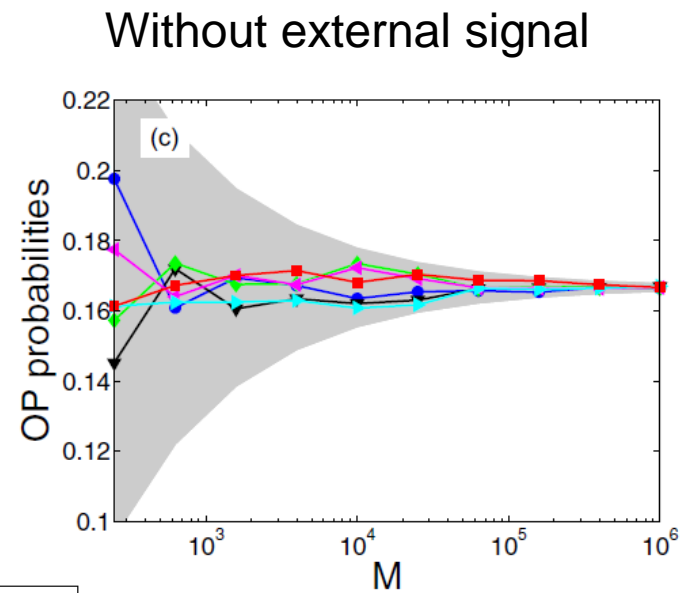
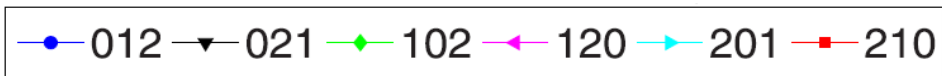
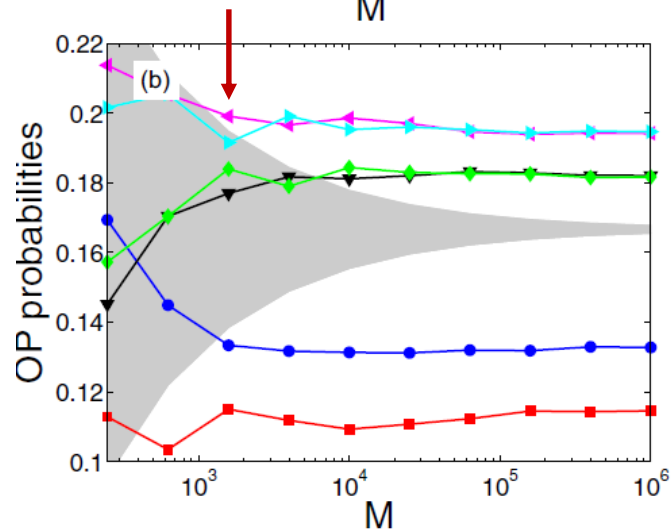
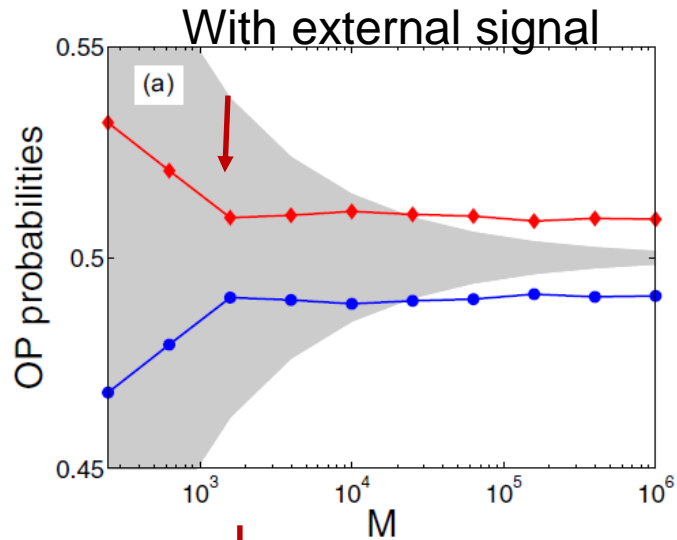
$$\frac{dy}{dt} = x + a + a_0 \cos(2\pi t/T) + D\xi(t),$$



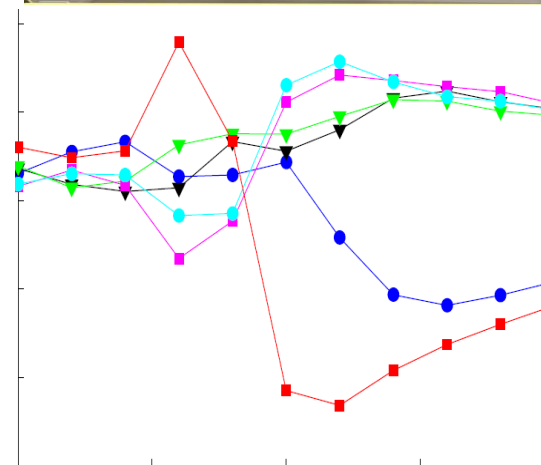
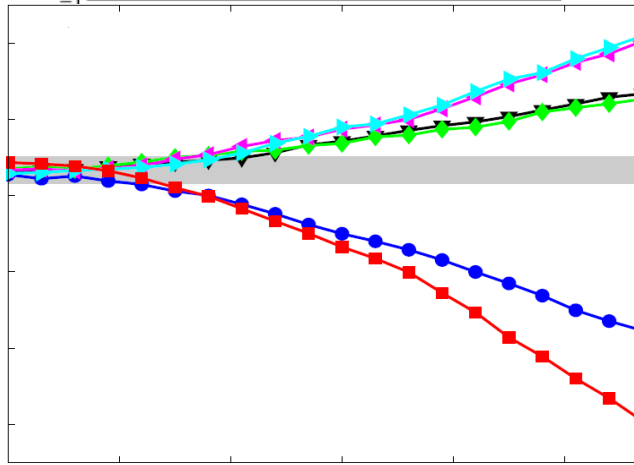
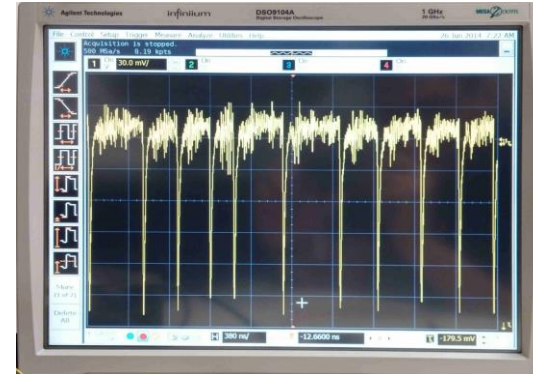
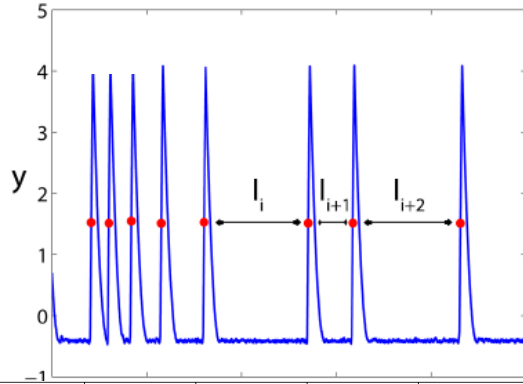
- Gaussian white noise and subthreshold signal:  $a_0$  and  $T$  such that spikes are **noise-induced**.
- Time series with  $M=100,000$  ISIs simulated ( $a=1.05$ ,  $\epsilon=0.01$ ).
- Gray region: significance analysis with  $3\sigma$  confidence level.



# Data requirements: influence of the number of patterns



# Side note: comparison with the laser spikes

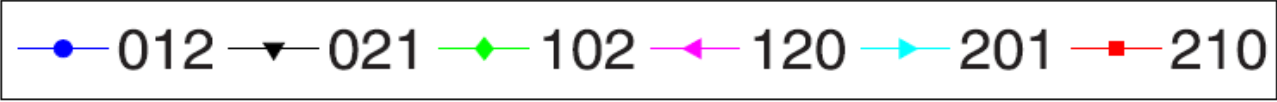


Modulation amplitude

Modulation amplitude

[J. M. Aparicio-Reinoso et al PRE 94, 032218 \(2016\)](#)

[A. Aragonese et al, Sci. Rep. 4, 4696 \(2014\)](#)



**Detecting longer correlations**

## Using longer words

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

- But long time series will be required to estimate the probabilities of the fast growing number of words in the dictionary (D!).
- Solution: a **lag  $\tau$**  allows considering long time-scales without having to use words of many letters

[...  $x(t)$ ,  $x(t + 2)$ ,  $x(t + 4)$ ,...]

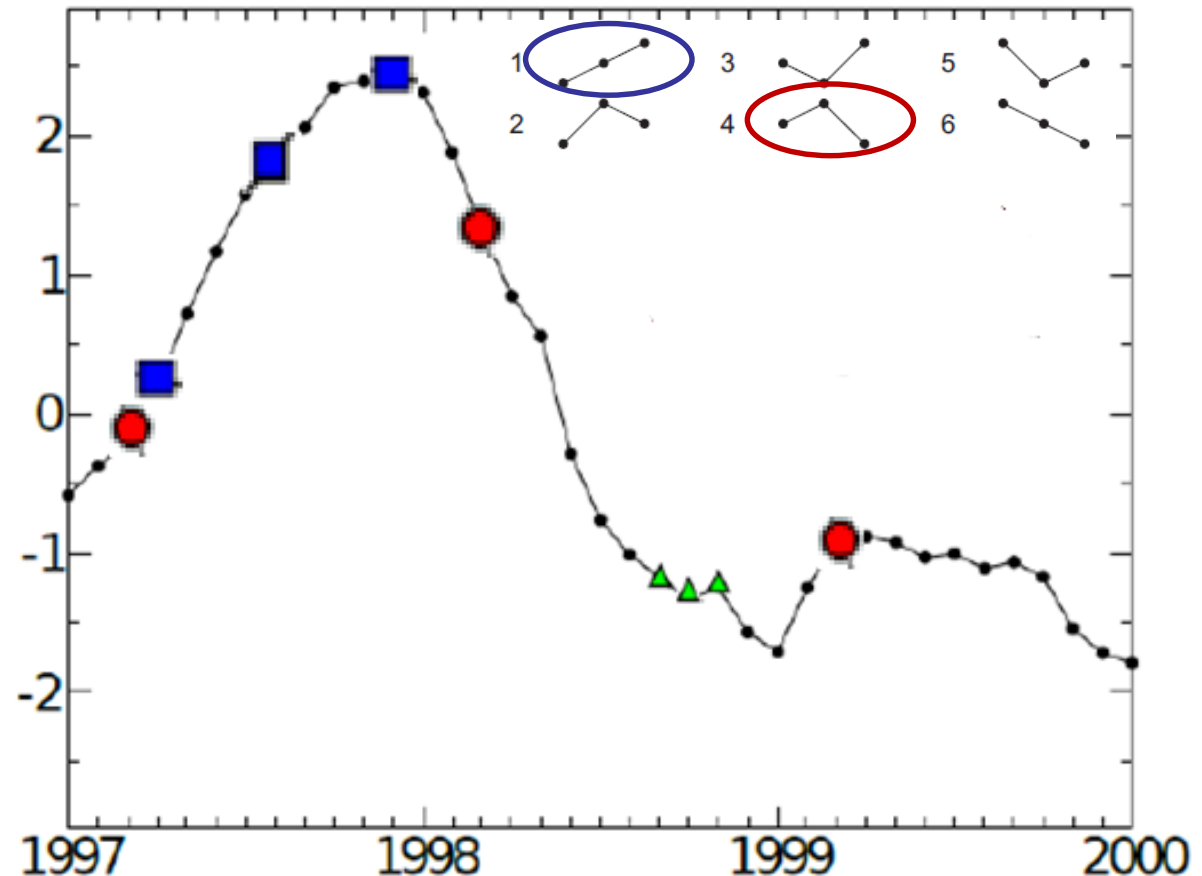
- Example: climatological data (monthly sampled)
  - Consecutive months: [... $x_i(t)$ ,  $x_i(t + 1)$ ,  $x_i(t + 2)$ ...]
  - Consecutive years: [... $x_i(t)$ ,...  $x_i(t + 12)$ ,...  $x_i(t + 24)$ ...]
- **Varying  $\tau$  = varying temporal resolution (sampling time)**



# Ordinal patterns can be defined using a lag time between the data points (varying the effective “sampling time”)

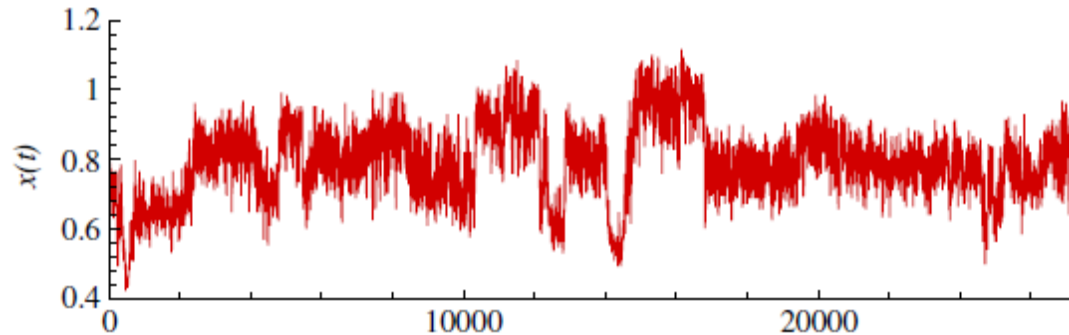
- **Green** triangles: intra-seasonal pattern,
- **blue** squares: intra-annual pattern
- **red** circles: inter-annual pattern

Example: el Niño index, monthly sampled

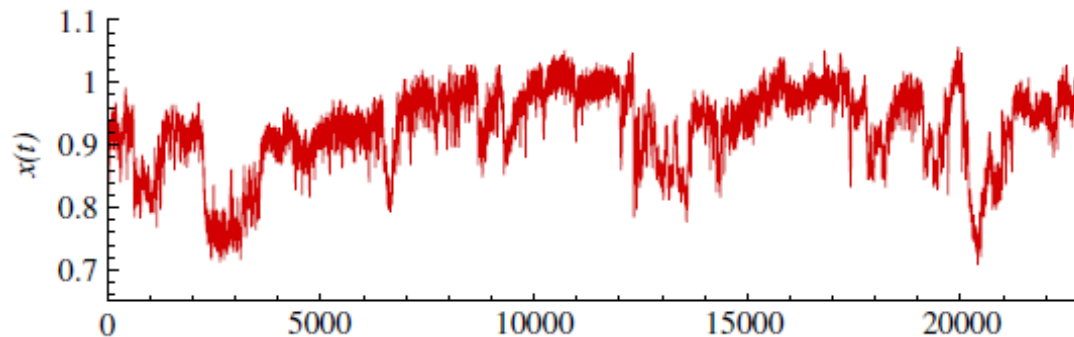


# Example of application of ordinal analysis to ECG-signals

## Time series of **inter-beat intervals**



healthy  
subject

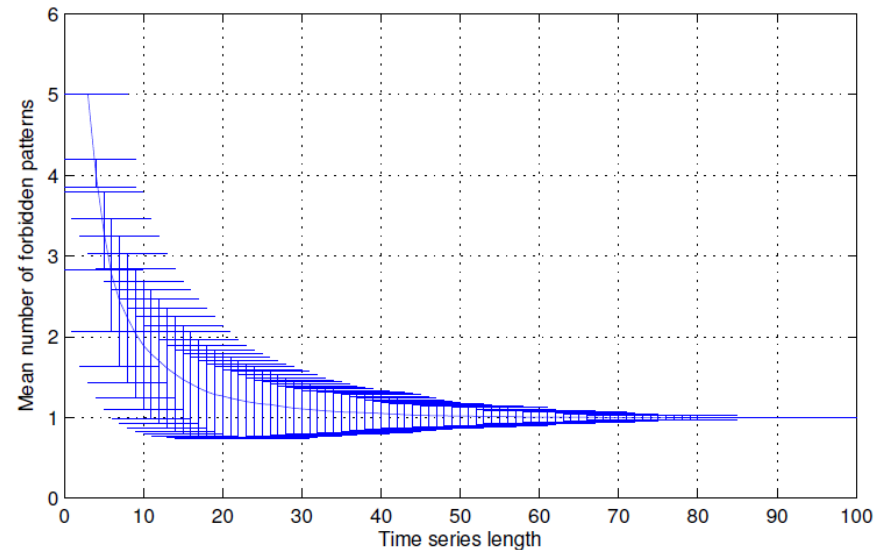


congestive  
heart failure



# Forbidden or missing patterns?

- Deterministic systems (such as the logistic map) can have **forbidden** patterns, while stochastic systems have **missing** patterns (unobserved patterns due to the finite length of the time series).
- Number of missing  $D=3$  patterns found in 1000 time series generated with the logistic map ( $r=4$ ), as a function of the length of the series.

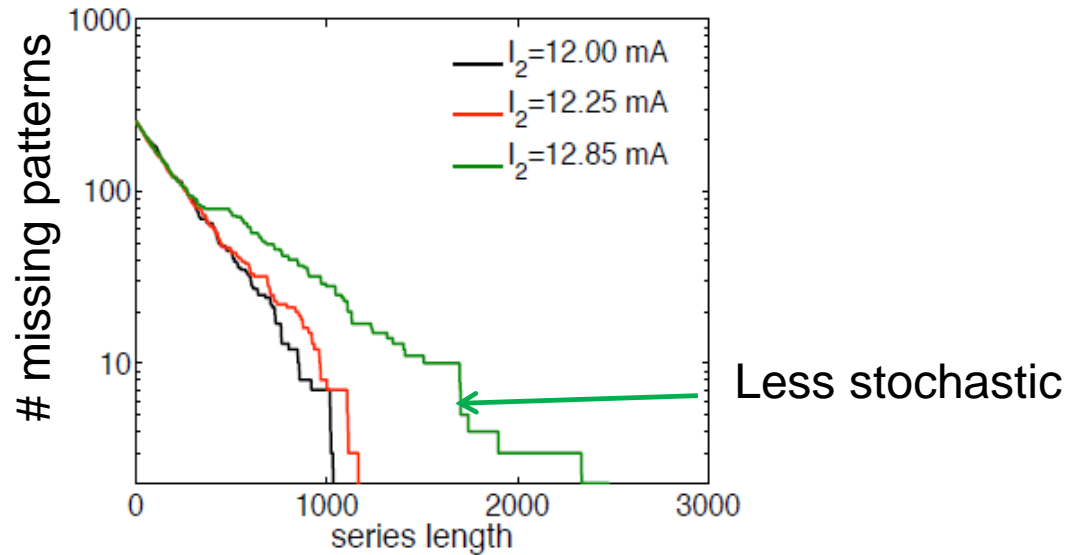


210

# Can this be useful for distinguishing noise from chaos?

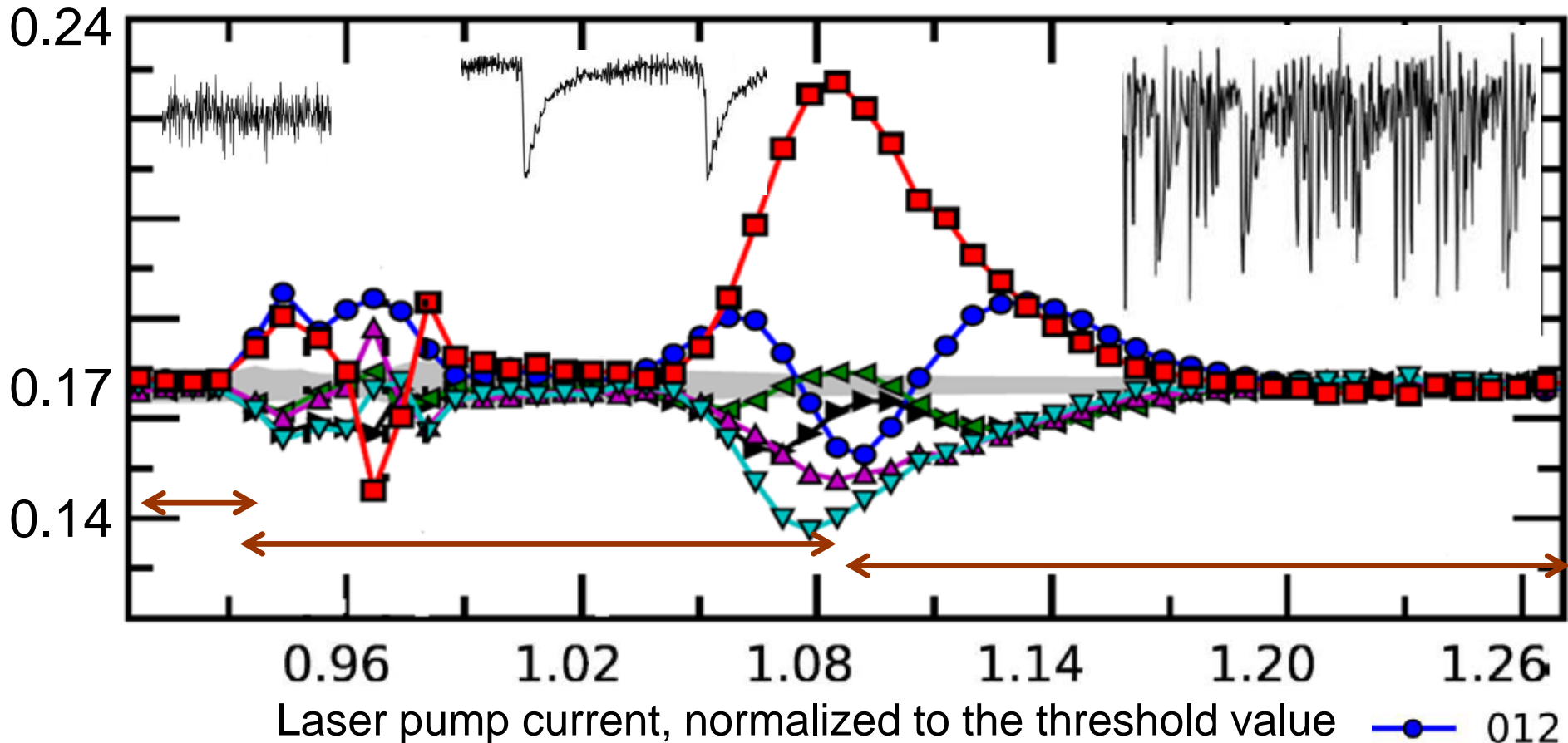
- It can provide a way to quantify the degree of stochasticity.

- Example:  
experimental data is transformed into a sequence of 0s and 1s, the patterns are formed by 8 symbols; the number of possible patterns is  $2^8 = 256$ .



J. Tiana-Alsina et al, Phil. Trans. Royal Soc. A 368, 367 (2010)

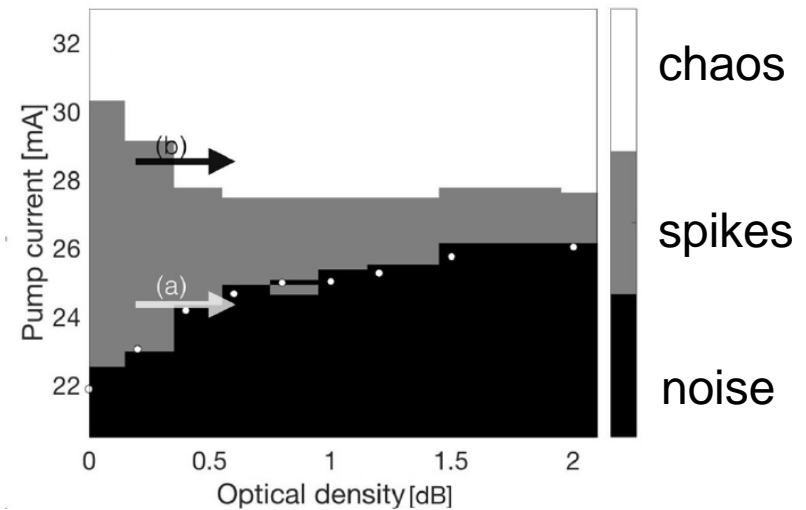
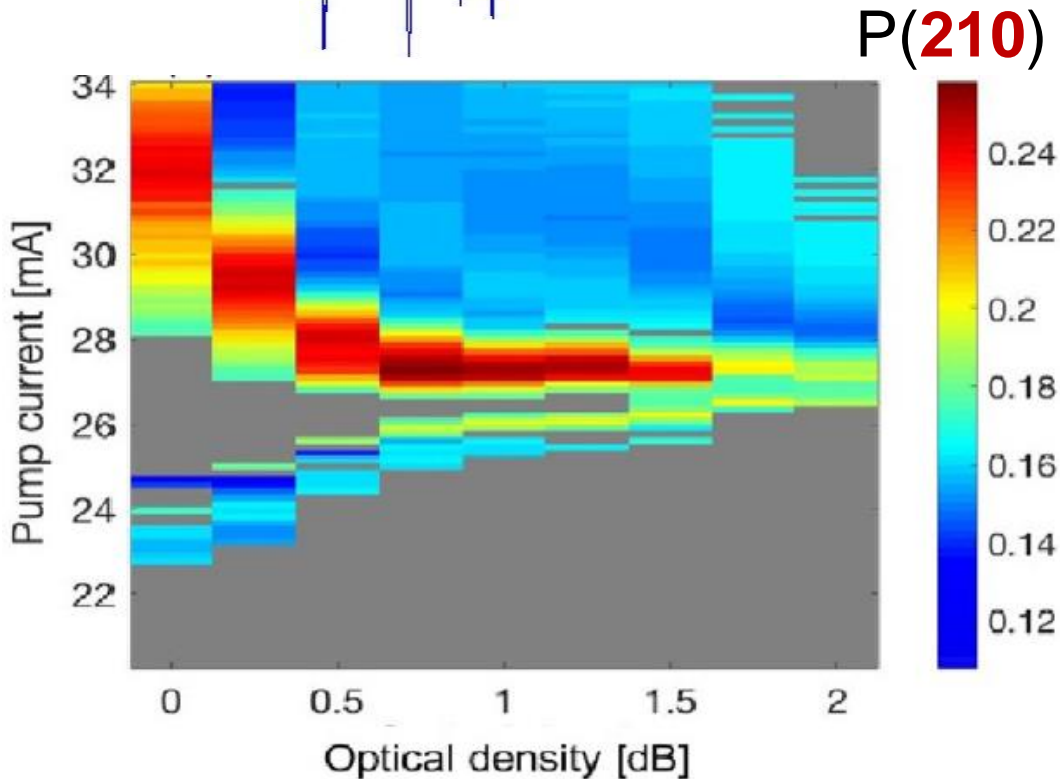
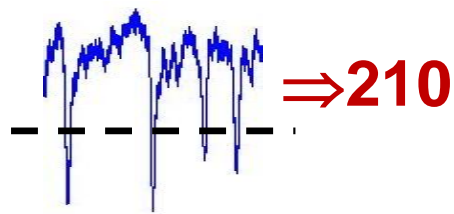
# Ordinal analysis identifies the onset of different dynamical regimes, but fails to distinguish “noise” and “chaos”



A probability value in the gray region is consistent with  $p = 1/6 \cong 0.17$  with 99.7% confidence level.

[C. Quintero-Quiroz et al, Scientific Reports \(2016\)](#)

# Another way to “classify” the laser intensity time series into noise, spikes or chaos

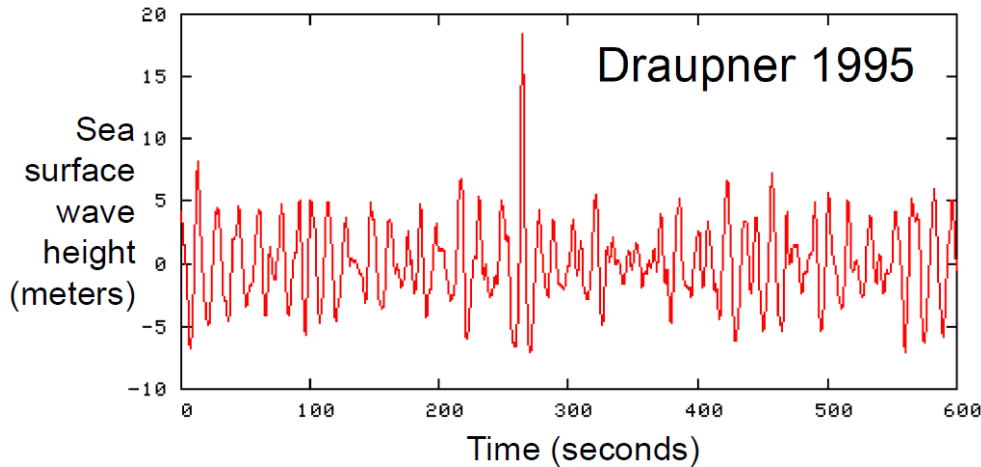


Gray region:  $P(210)$  is consistent with  $1/6$  with  $3\sigma$  confidence level.

**Another example of application  
of ordinal analysis:  
prediction of extremes**

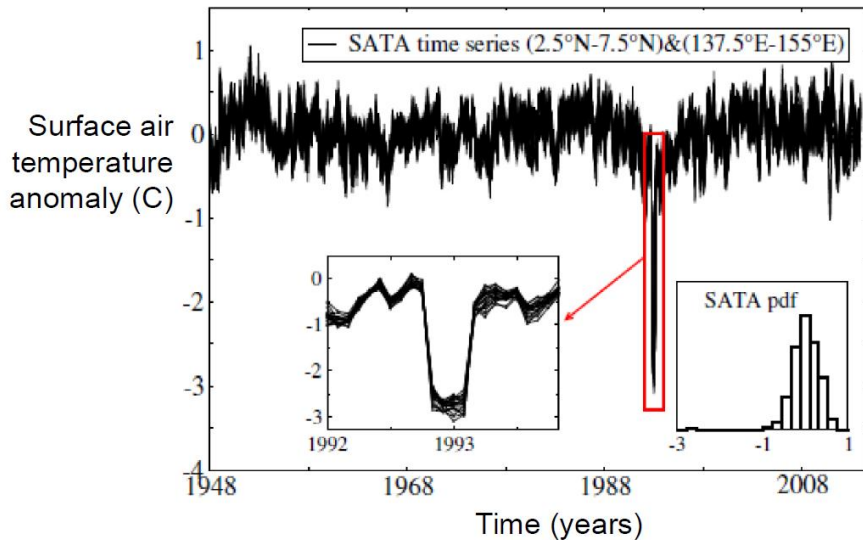


# Extreme events in nature



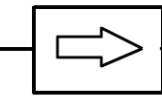
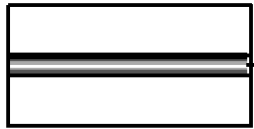
Optical chaos: provides the opportunity to

- Advance our understanding of the mechanisms that generate extreme fluctuations
- advance predictability.



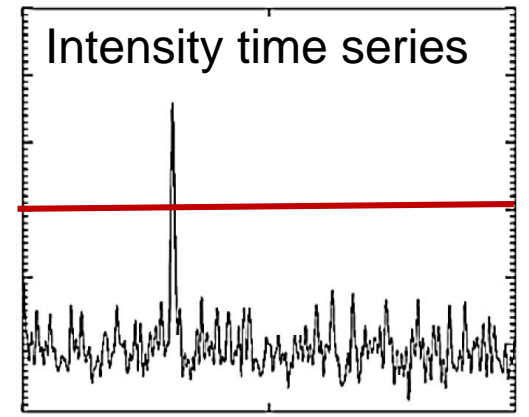
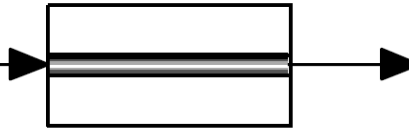
# “Deterministic” optical rogue waves

Master Laser



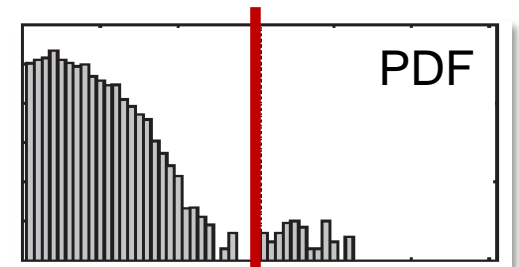
Isolator

Slave Laser



Parameters:

- Injection ratio
- Frequency detuning (controlled via the pump current)

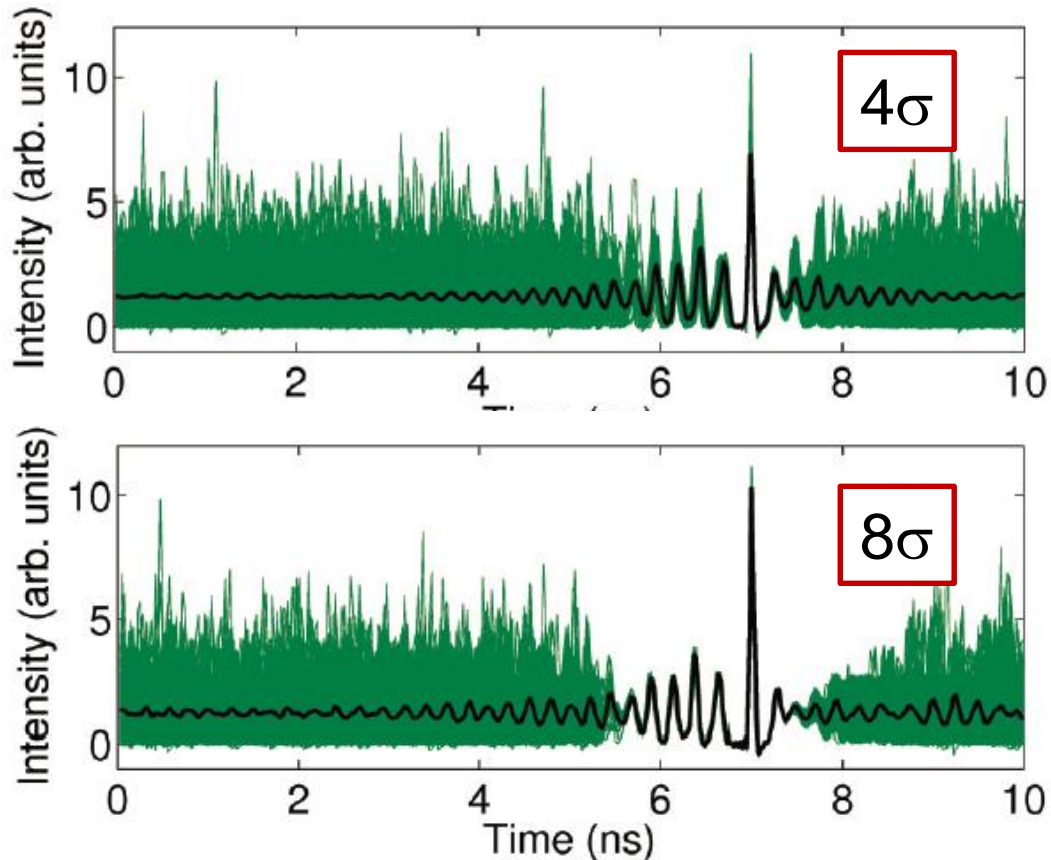


Rogue wave: pulse above

$$\langle A \rangle + 6-8 \sigma$$

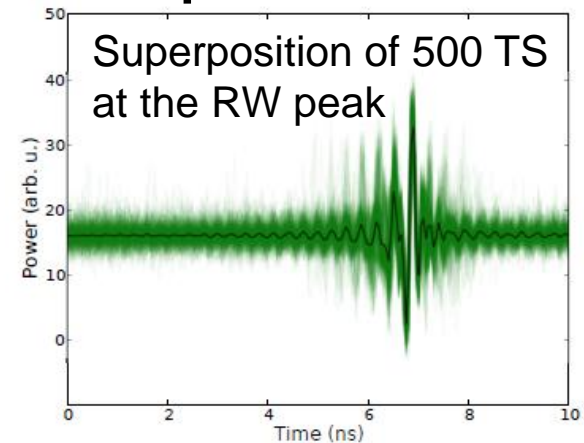
# RW predictability

## Deterministic simulations



Superposition of 50 time-series at the RW peak

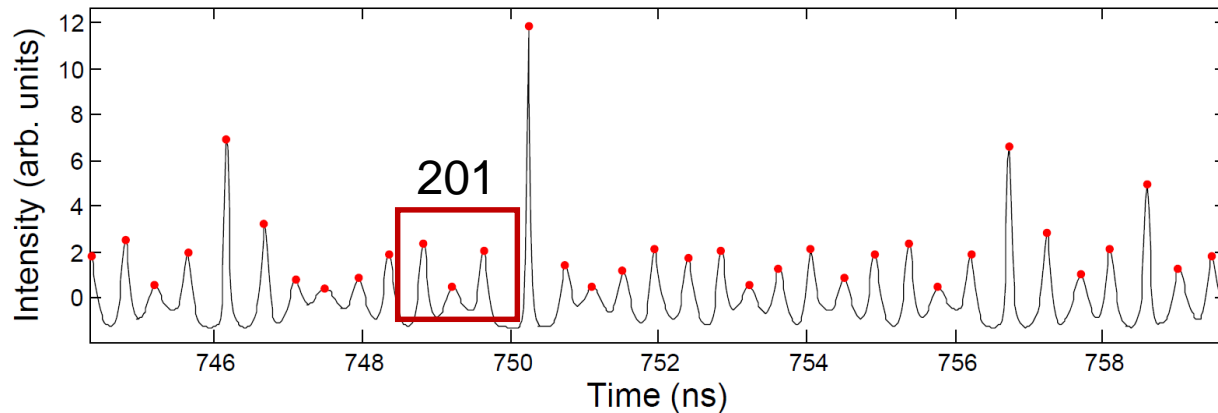
## Experiments



⇒ Well-defined oscillation pattern anticipates extreme pulses.

[J. Zamora-Munt et al, PRA 87, 035802 \(2013\)](#)

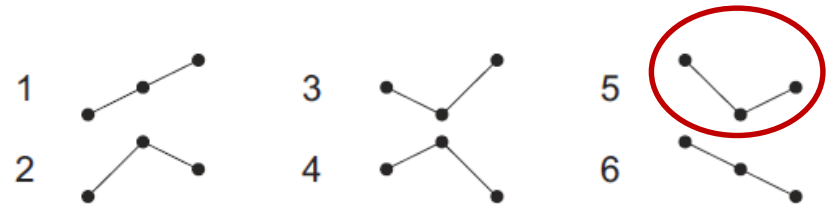
# Ordinal analysis to detect the “pattern” that tends to precede the extreme pulses



- Consider the sequence of intensity peak heights (red dots):

$$\{\dots, l_i, l_{i+1}, l_{i+2}, \dots\}$$

- Possible order relations of three consecutive values:

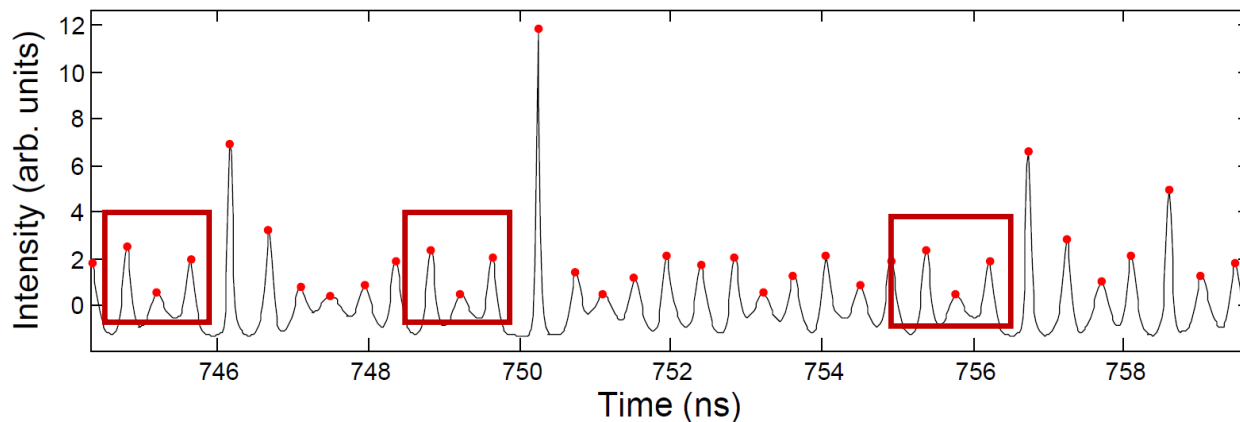
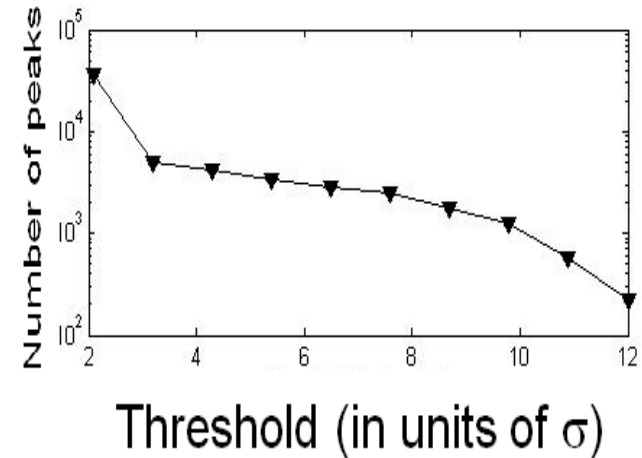
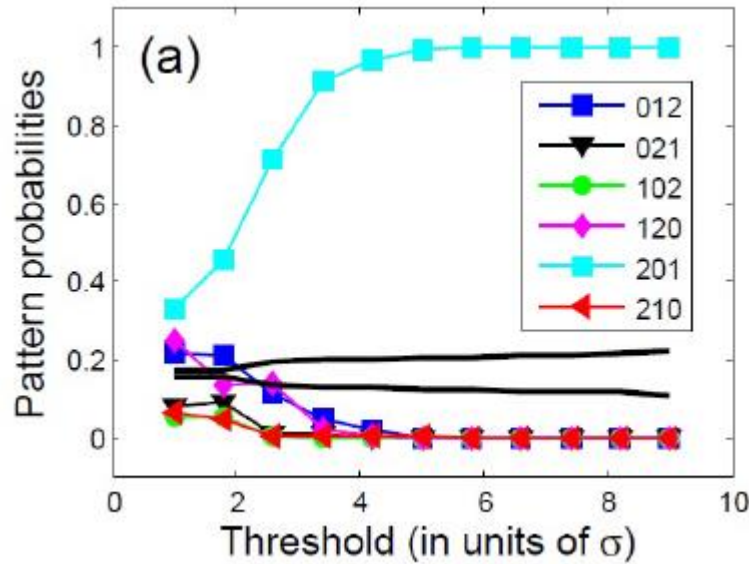


We calculate the probability of the pattern that occurs before each high pulse:

If  $l_i > \mathbf{TH}$ , we analyze the pattern defined by  $(l_{i-3}, l_{i-2}, l_{i-1})$

# Results: deterministic simulations

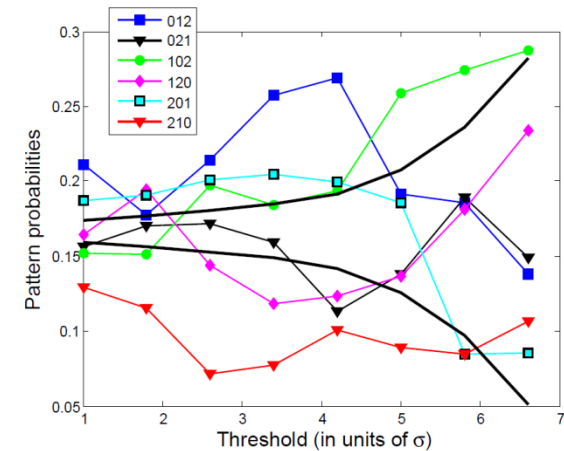
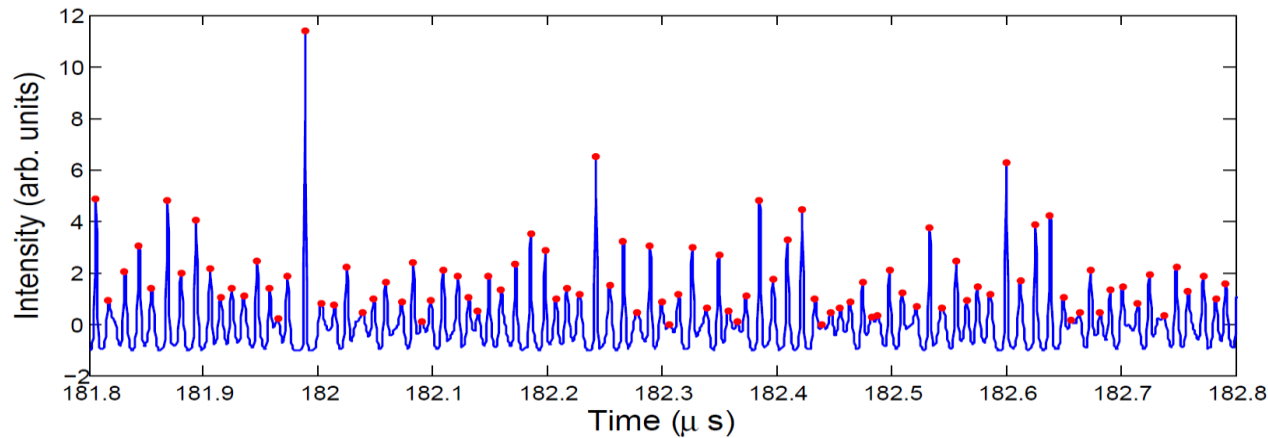
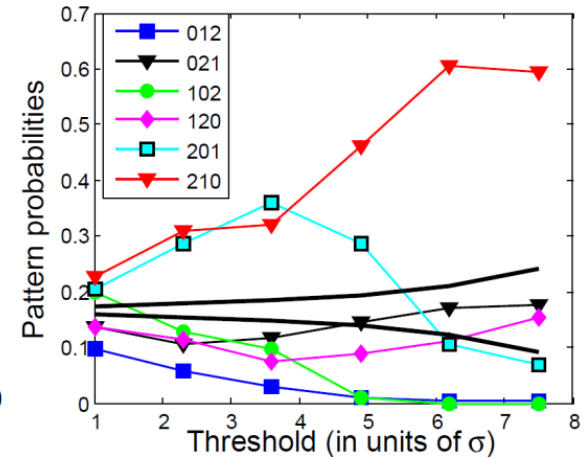
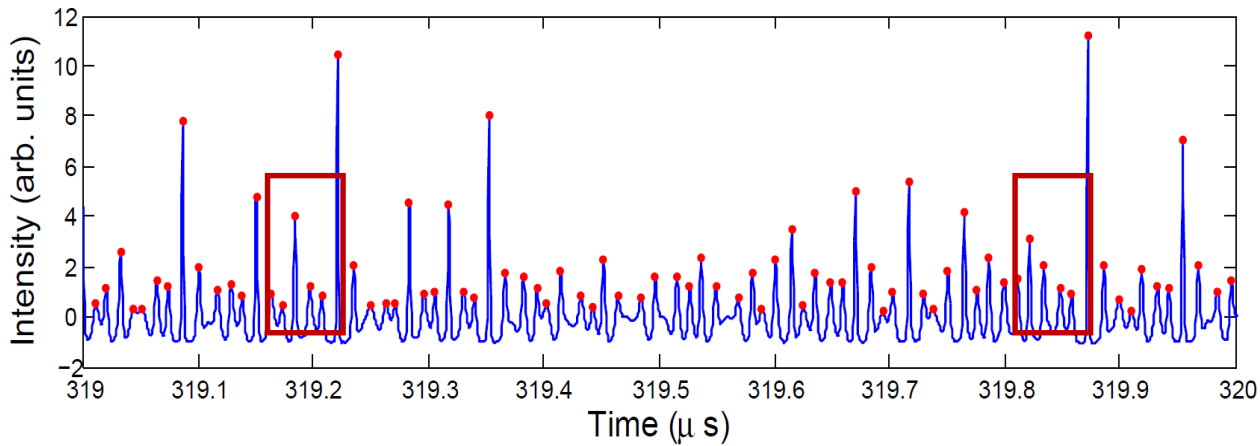
Black lines:  
99% confidence  
 $p_i = 1/6 \forall i$



- $P(201)=1$  if  $TH > 6$
- Problem:  $P(201) \neq 0$  if  $TH < 6$  (pattern 201 also anticipates some small pulses)  $\Rightarrow$  false alarms (false positives)

# Including, in the simulations, noise and a small modulation of a parameter

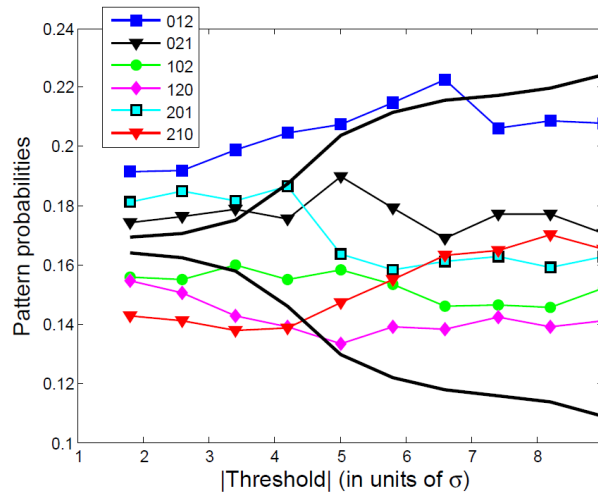
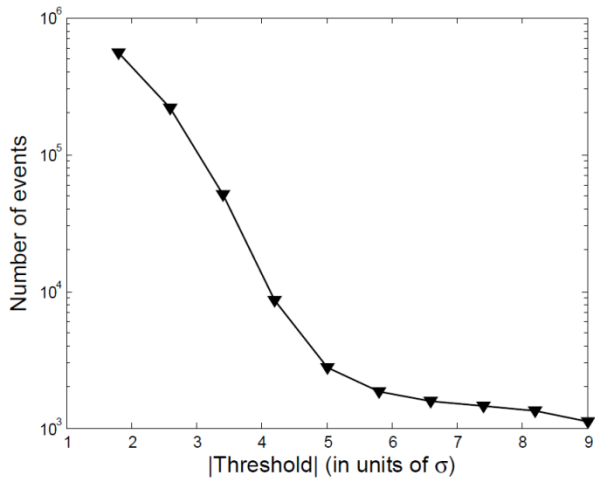
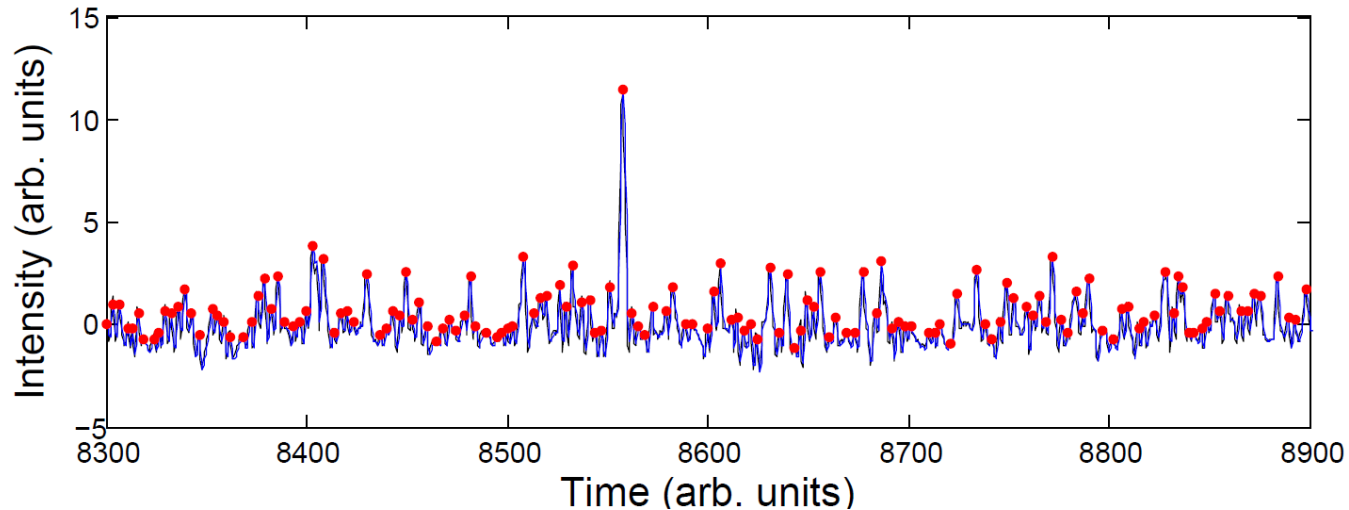
## ■ Two different modulation frequencies



In the first case: 210 is a “good” warning.

⇒ “early warning pattern” varies with parameters and might not exist.

# Analysis of experimental data

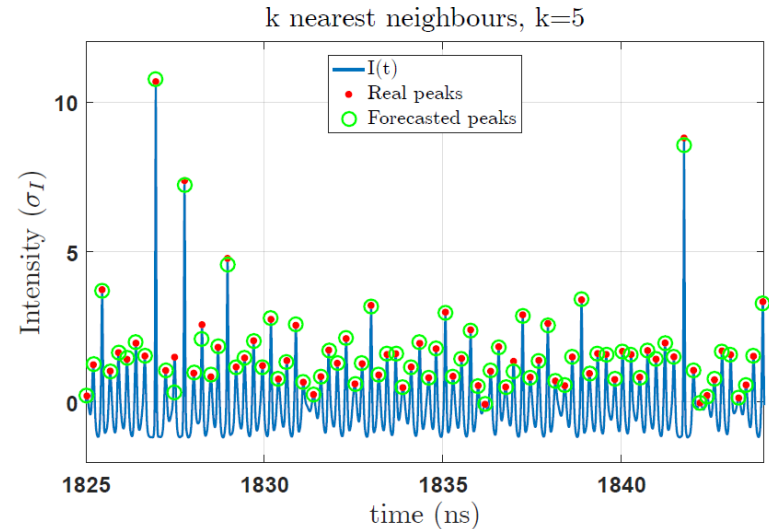
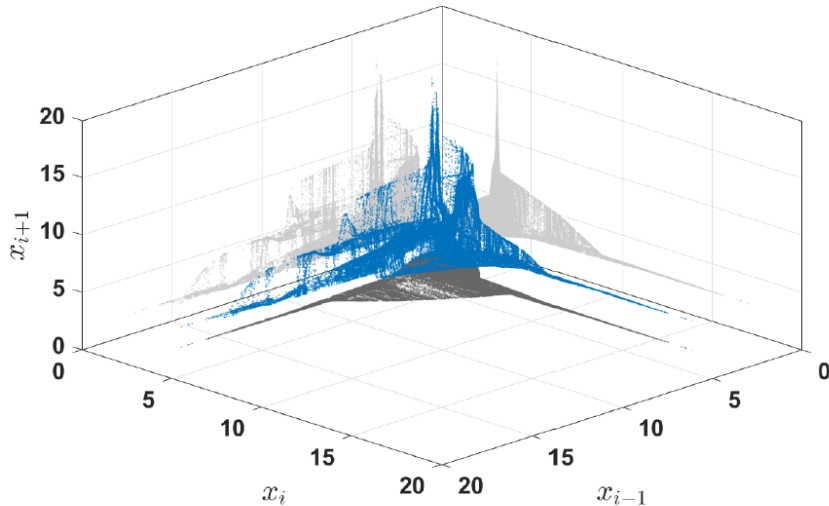


Way to improve the “early warning”:

- Filter noise
  - Longer patterns
- $\{\dots I_i, I_{i+1}, I_{i+2}, I_{i+3}, \dots\}$

# A well-known prediction tool

- Estimate the function  $f(x_{i-2}, x_{i-1}, x_i) = x_{i+1}$



Source: Pablo Amil

- Main challenges: short datasets and/or high dimensional data.
- Standard approach: neural networks.
- New development: “reservoir computing”.
- Read more:
  - S. Birkholz et al., “*Predictability of Rogue Events*”, PRL 114, 213901 (2015)
  - J. Pathak et al, “*Model-Free Prediction of Large Spatiotemporally Chaotic Systems from Data: A Reservoir Computing Approach*”, PRL 120, 024102 (2018)



**How to quantify unpredictability  
and complexity?**

# Information measure: Shannon entropy

- The time-series is described by a set of probabilities  $\sum_{i=1}^N p_i = 1$

- **Shannon** entropy:  $H = -\sum_i p_i \log_2 p_i$

- Interpretation: “*quantity of **surprise** one should feel upon reading the result of a measurement*”

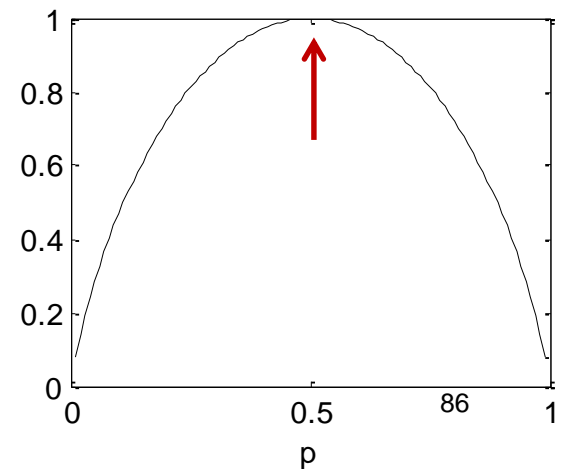
K. Hlavackova-Schindler et al, Physics Reports 441 (2007)

- Simple example: a random variable takes values 0 or 1 with probabilities:  $p(0) = p$ ,  $p(1) = 1 - p$ .

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

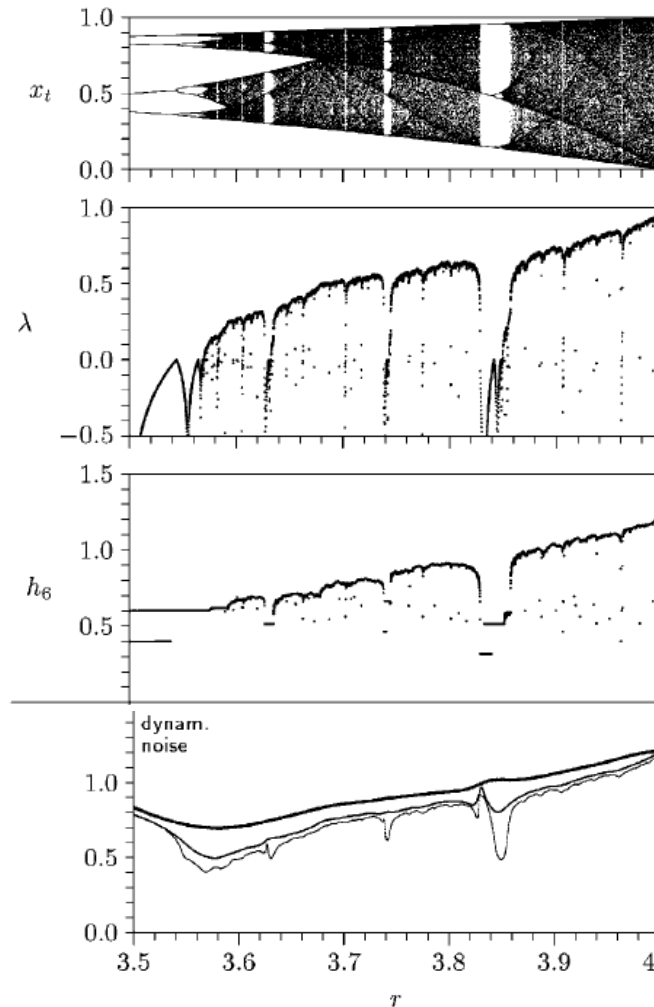
⇒  $p=0.5$ : Maximum **unpredictability**.  $\Xi$

Shannon entropy computed from ordinal probabilities: **Permutation Entropy**



# Permutation entropy and Lyapunov exponent

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



$$|\delta_n| \approx |\delta_0| e^{n\lambda}$$

$$\lambda = \lim_{n \rightarrow \infty} \left\{ \frac{1}{n} \sum_{i=0}^{n-1} \ln |f'(x_i)| \right\}$$

Entropy per symbol:

$$h_n = H(n)/(n - 1)$$

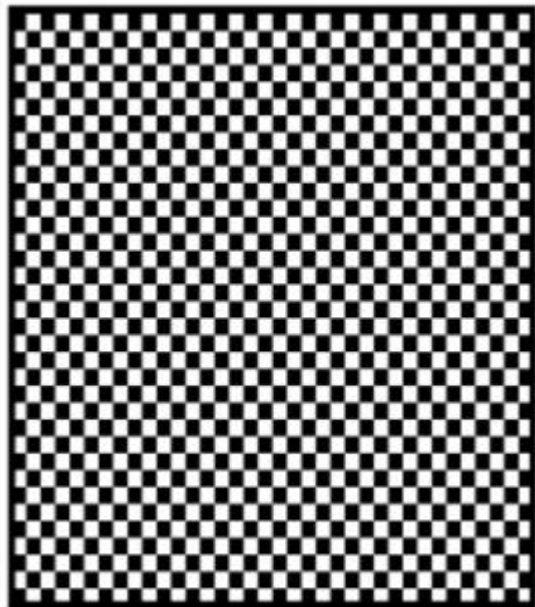
Robust to noise

*Bandt and Pompe  
Phys. Rev. Lett. 2002*

**Entropy: measures unpredictability or disorder.  
How to quantify Complexity?**

We would like to find a quantity “C” that measures **complexity**, as the entropy, “H”, measures **unpredictability**, and, for low-dimensional systems, the Lyapunov exponent measures **chaos**.

**Order**



$$H = 0$$

$$C = 0$$

**Chaos**



$$H \neq 0$$

$$C \neq 0$$

**Disorder**



$$H = 1$$

$$C = 0$$

*“A useful complexity measure needs to do more than satisfy the boundary conditions of vanishing in the high- and low-entropy limits.”*

*“Maximum complexity occurs in the region between the system’s perfectly ordered state and the perfectly disordered one.”*

## Information measures

- Assuming that we know the set of probabilities  $P=[p_i, i=1, Nbin]$  that characterizes a time series, several information measures have been proposed, a few popular ones are:

Shannon entropy  $I[P] = S_S[P] = -\sum_i p_i \ln p_i$

Tsallis entropy  $I[P] = S_T^q[P] = \frac{1}{q-1} \left[ 1 - \sum_i p_i^q \right]$

Renyi entropy  $I[P] = S_R^q[P] = \frac{1}{1-q} \ln \left[ \sum_i p_i^q \right]$

Fischer information  $I[P] = F[P] = F_0 \ln \left[ \sum_{i=1}^{Nbin-1} \left( p_{i+1}^{1/2} - p_i^{1/2} \right)^2 \right]$

## Normalization

$$H[P] = \frac{I[P]}{I_{\max}}$$

$$0 \leq H[P] \leq 1$$

where  $I_{\max} = I[P_e]$  and  $P_e$  is the equilibrium probability distribution (that maximizes the information measure).

Example: if  $I[P] = \text{Shannon entropy}$

then  $P_e = [p_i = 1/Nbin \text{ for } i=1, Nbin]$

and  $I_{\max} = \ln(Nbin)$

## Disequilibrium Q

Measures the “distance“ from  $P$  to the equilibrium distribution,  $P_e$

$$Q[P] = Q_0 D[P, P_e]$$

where  $Q_0$  is a normalization constant such that

$$0 \leq Q[P] \leq 1$$



# Distance between two probability distributions $P$ and $P_e$

Euclidean  $D_E[P, P_e] = \|P - P_e\|_E = \sum_i (p_i - p_{i,e})^2$

Kullback  $D_K[P, P_e] = K[P|P_e] = I[P_e] - I[P]$

Jensen divergence  $D_J[P, P_e] = \frac{K[P|P_e] + K[P_e|P]}{2}$

**Read more:** S-H Cha: *Comprehensive Survey on Distance/Similarity Measures between Probability Density Functions*, Int. J of. Math. Models and Meth. 1, 300 (2007)

# Statistical complexity measure C

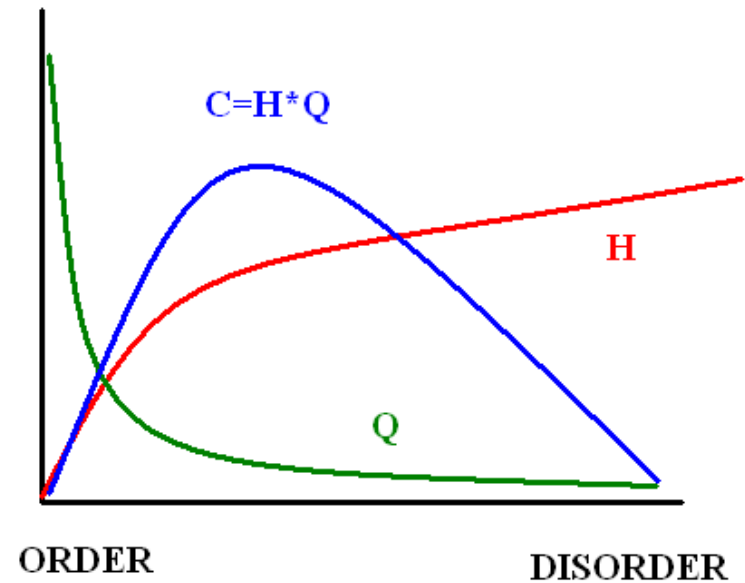
A family of complexity measures can be defined as:

$$C[P] = H_A[P] \cdot Q_B[P]$$

where

A = S, T, R (Shannon, Tsallis, Renyi)

B = E, K, J (Euclidean, Kullback, Jensen)



$$C_{LMC}[P] = H_S[P] \cdot Q_E[P]$$

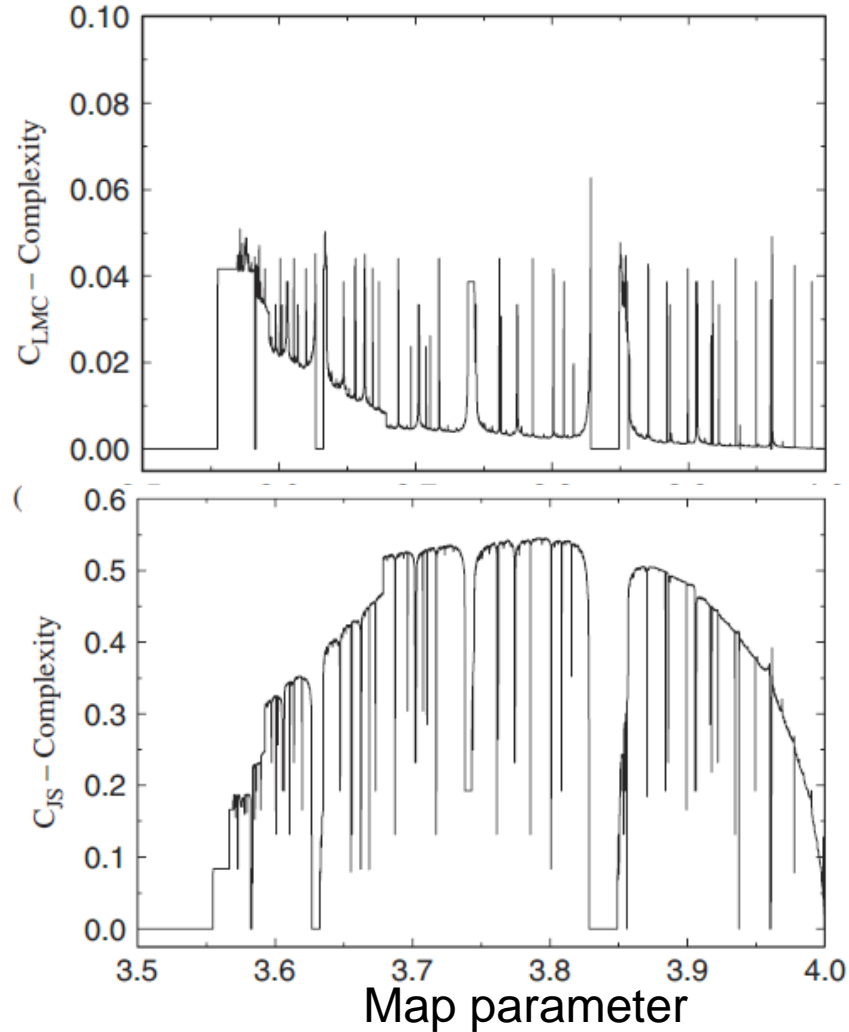
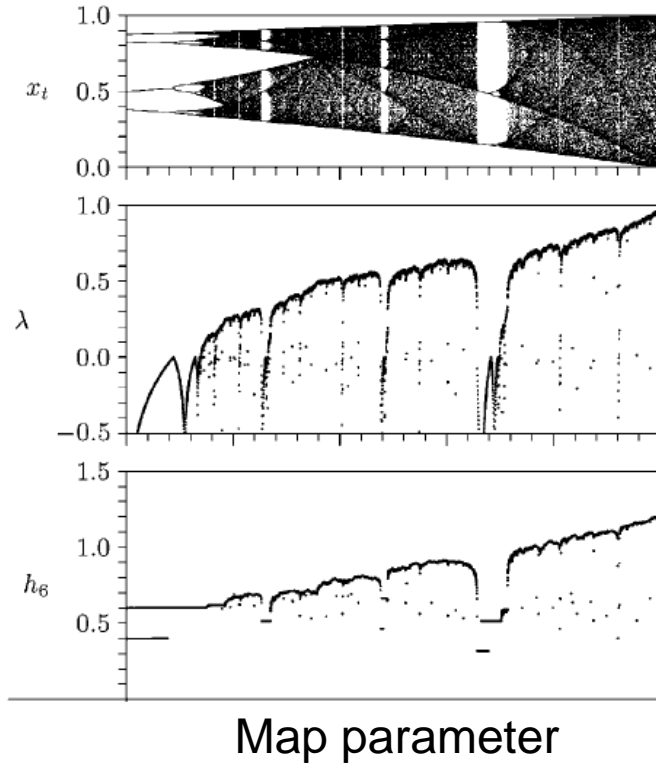
Lopez-Ruiz, Mancini & Calbet, Phys. Lett. A (1995).  
Anteneodo & Plastino, Phys. Lett. A (1996).

$$C_{MPR}[P] = H_S[P] \cdot Q_J[P]$$

Martín, Plastino & Rosso, Phys. Lett. A (2003).

# The complexity of the Logistic Map

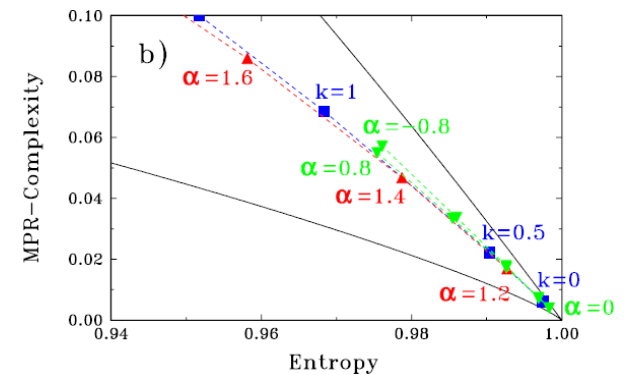
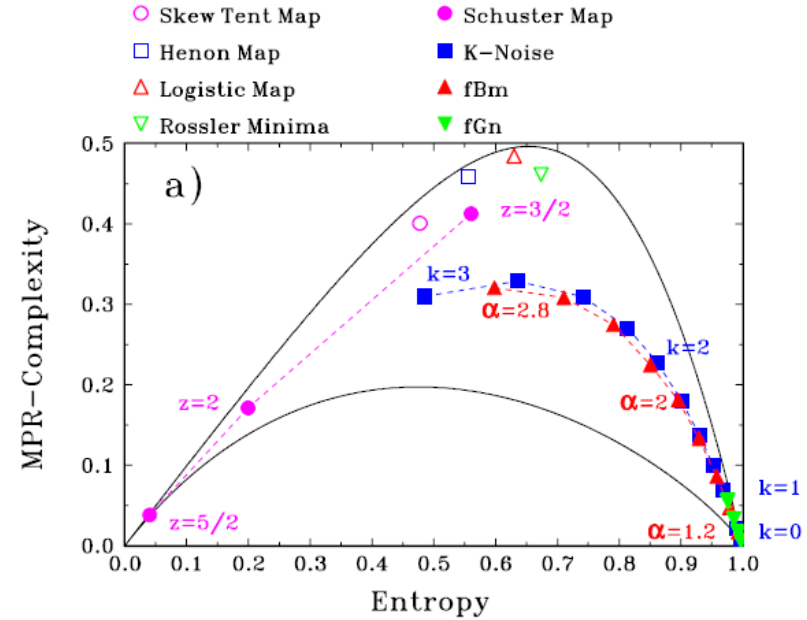
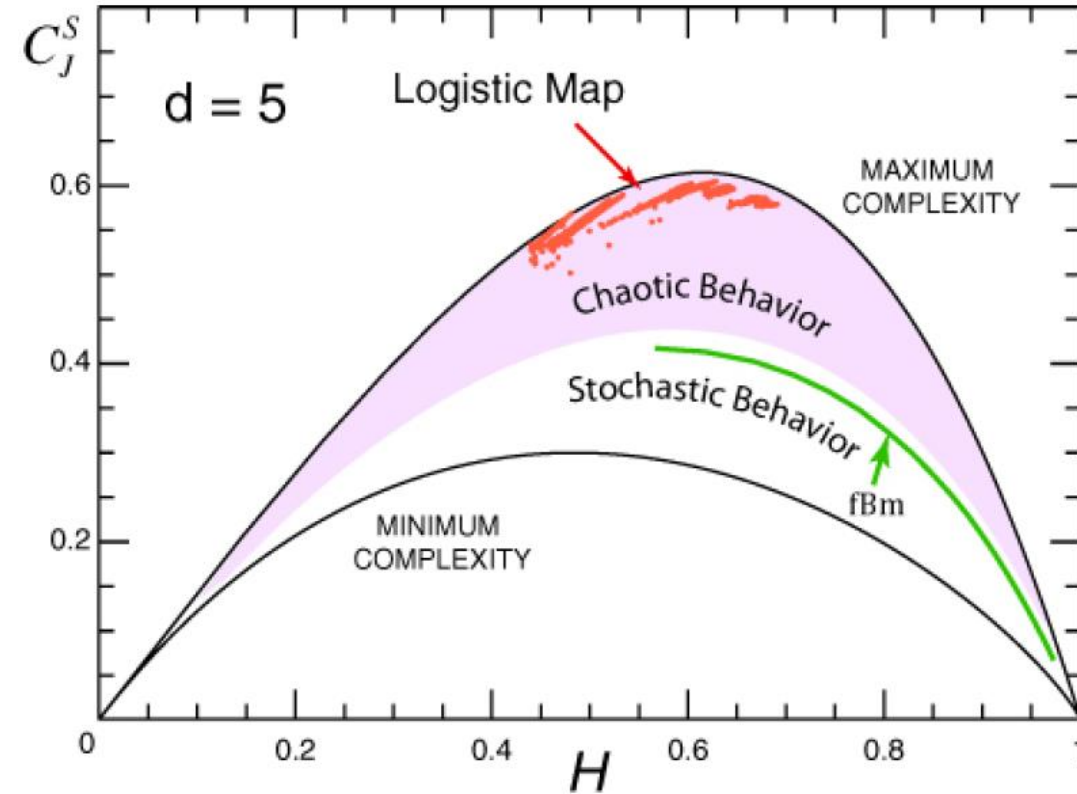
$$x(i+1) = r x(i)[1-x(i)]$$



Euclidian distance

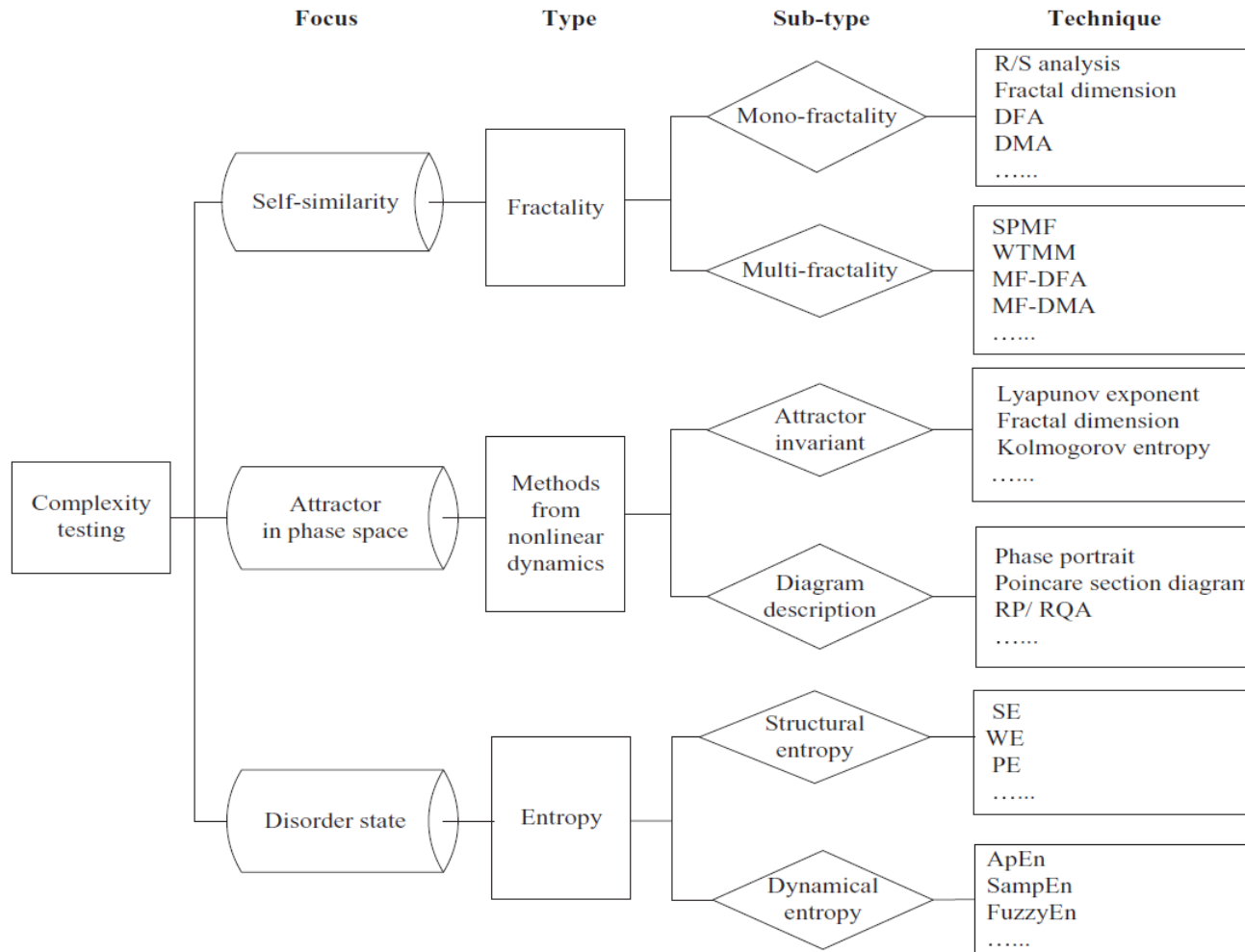
Jensen distance

# The (entropy, complexity) plane: a useful tool to distinguish noise from chaos



O. A. Rosso et al, Phys. Rev. Lett. 99, 154102 (2007)

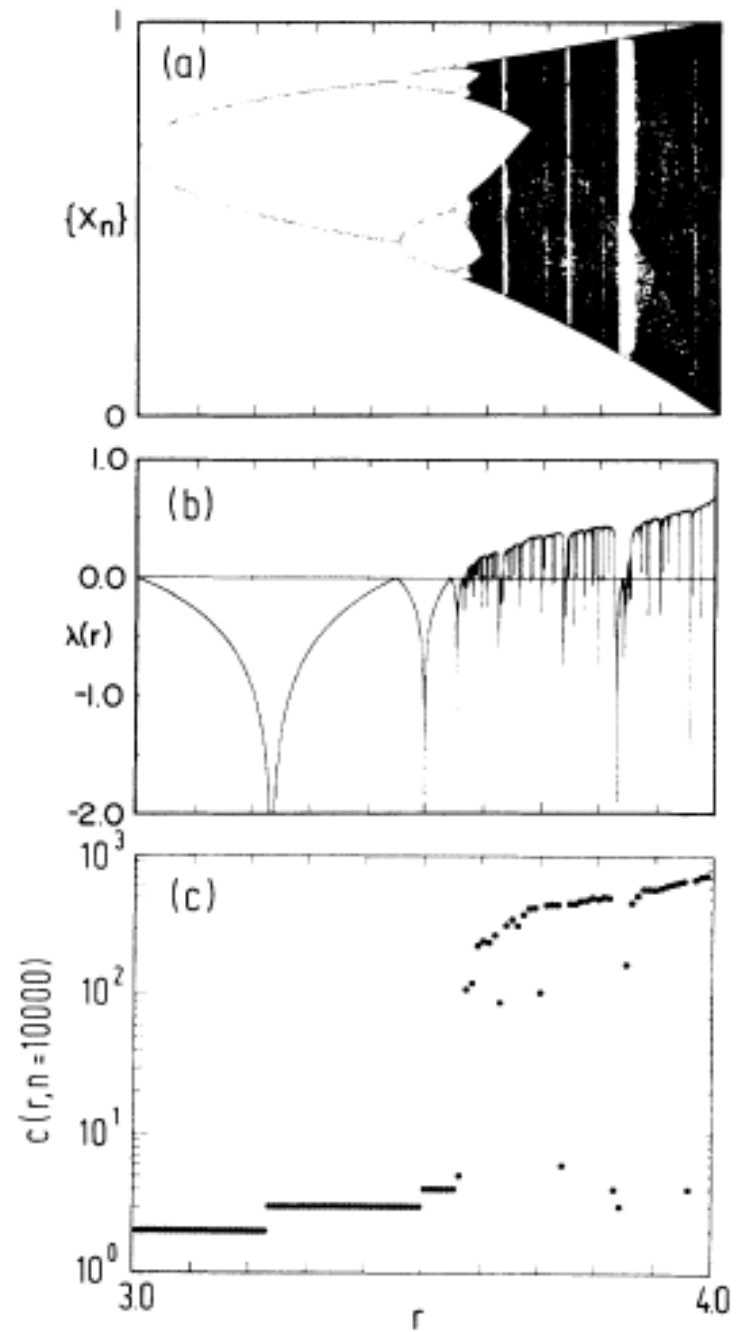
# Many complexity measures have been proposed



Further reading: L. Tang et al, “Complexity testing techniques for time series data: A comprehensive literature review”, Chaos, Solitons and Fractals 81 (2015) 117–135



# Lempel & Zip complexity of the Logistic Map



*Kaspar and Schuster, Phys Rev. A 1987*

# Application to climate data

- How to quantify
  - the degree of unpredictability of Surface Air Temperature (SAT) anomaly, and
  - the degree of nonlinear response of SAT to solar forcing?
- Proposed measures
  - Shannon entropy of SAT anomaly distribution of values (no symbols)
  - Distance from SAT to insolation (incoming solar radiation)



## Data

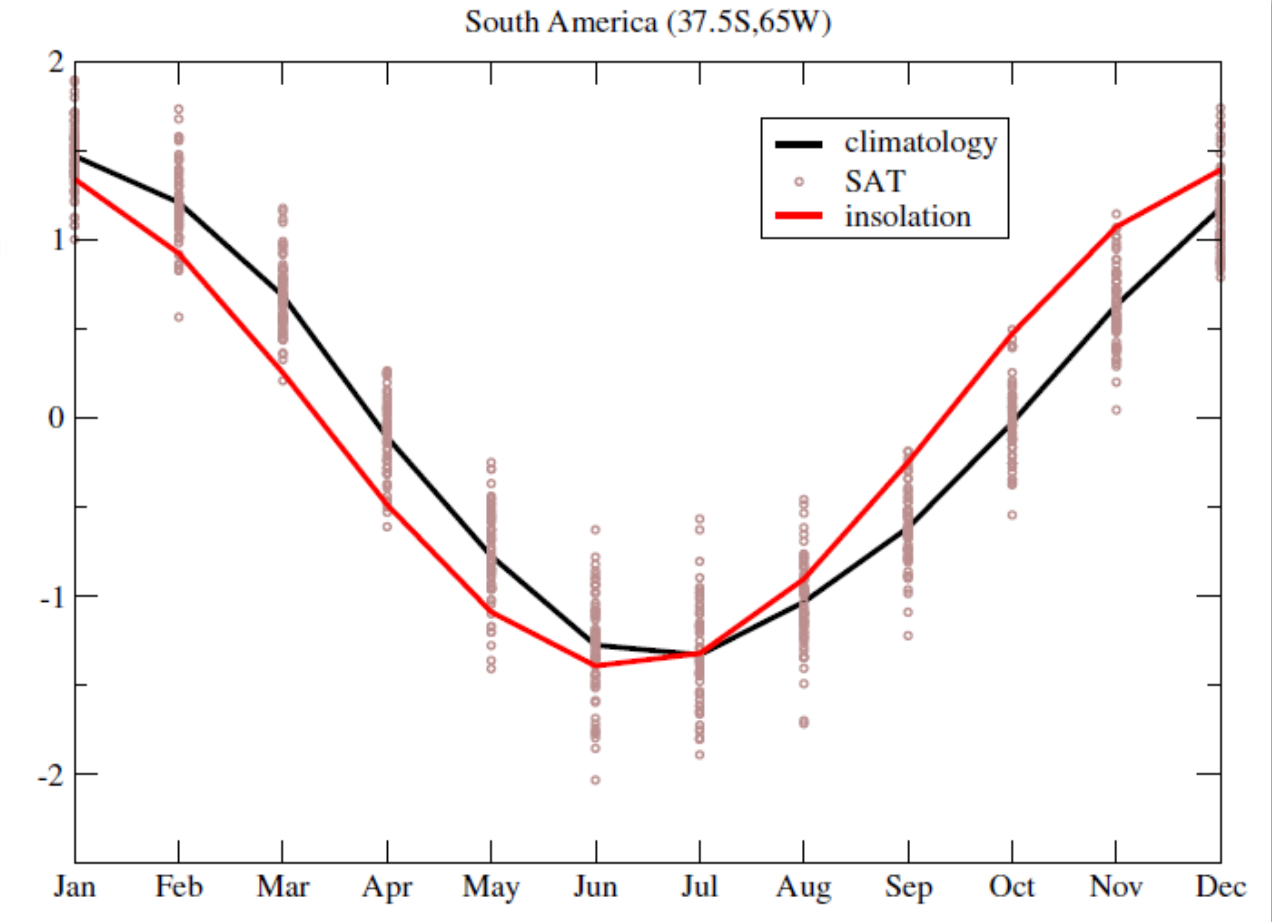
- Monthly mean SAT from two **reanalysis** (NCEP CDAS116 and ERA Interim17). The spatial resolution is  $2.5^\circ/1.5^\circ$  and cover the time-period [1949–2015]/[1980–2014] respectively.
  - NCEP CDAS1:  $N=10224$  time series of length  $L = 792$
  - ERA Interim:  $N = 28562$  and  $L = 408$ .
- Freely available.
- Reanalysis = run a sophisticated model of general atmospheric circulation and feed it with the available experimental data, in the different points of the Earth, at their corresponding times (data assimilation).
- This restricts the solution of the model to one as close to reality as possible in regions where data is available, and to a solution physically “plausible” elsewhere.

# Raw SAT, climatology and insolation

Climatology:

$$c_i(t) = \frac{1}{Y} \sum_{n=0}^Y y_i(t + 12n)$$

Monthly data,  
Y number of years

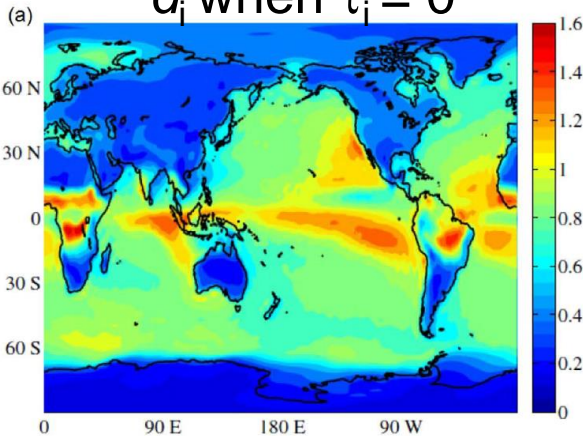


# Nonlinear response: distance between climatology and insolation

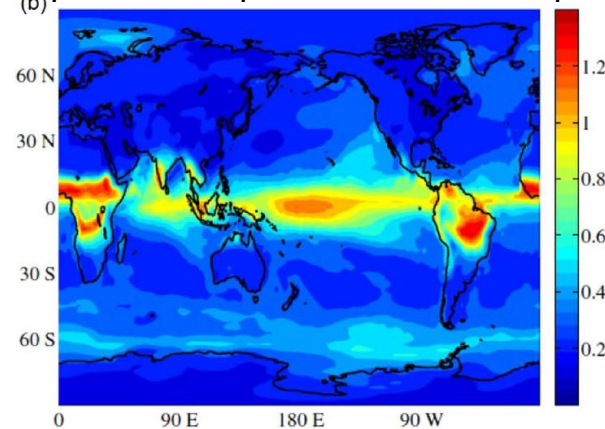
$$d_i = \sum_t |y_i(t + \tau_i) - x_i(t)|$$

$x$  (insolation) and  $y$  (climatology) are both normalized to  $\mu=0$  and  $\sigma=1$

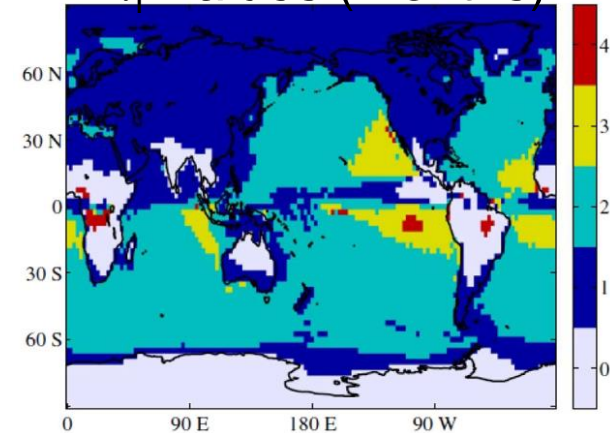
$d_i$  when  $\tau_i = 0$



$d_i$  when  $\tau_i$  minimizes  $d_i$



$\tau_i$  values (months)



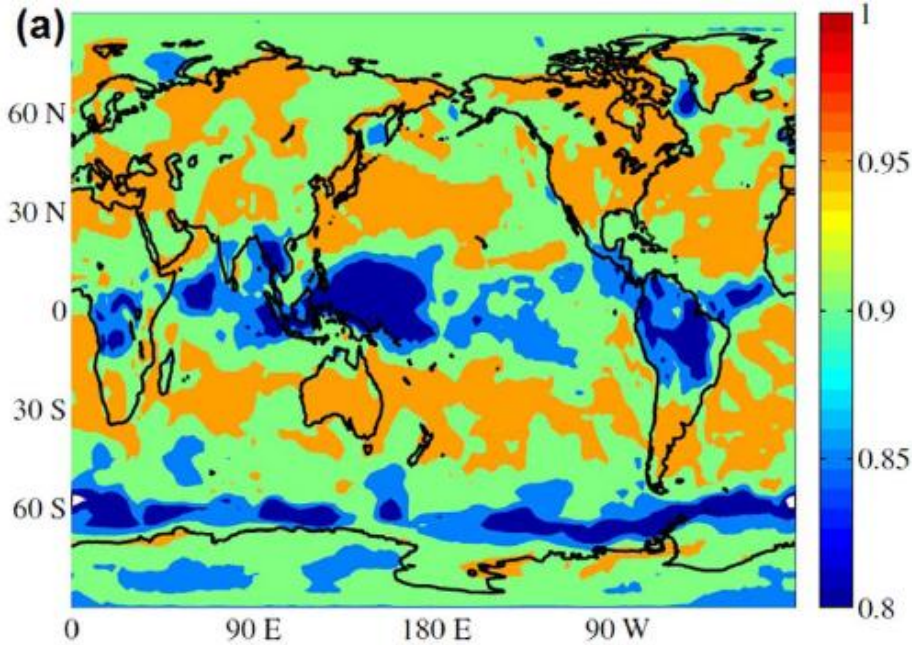
[F. Arizmendi, M. Barreiro and C. Masoller, "Identifying large-scale patterns of unpredictability and response to insolation in atmospheric data", Sci. Rep. 7, 45676 \(2017\).](#)

# Unpredictability of SAT anomaly

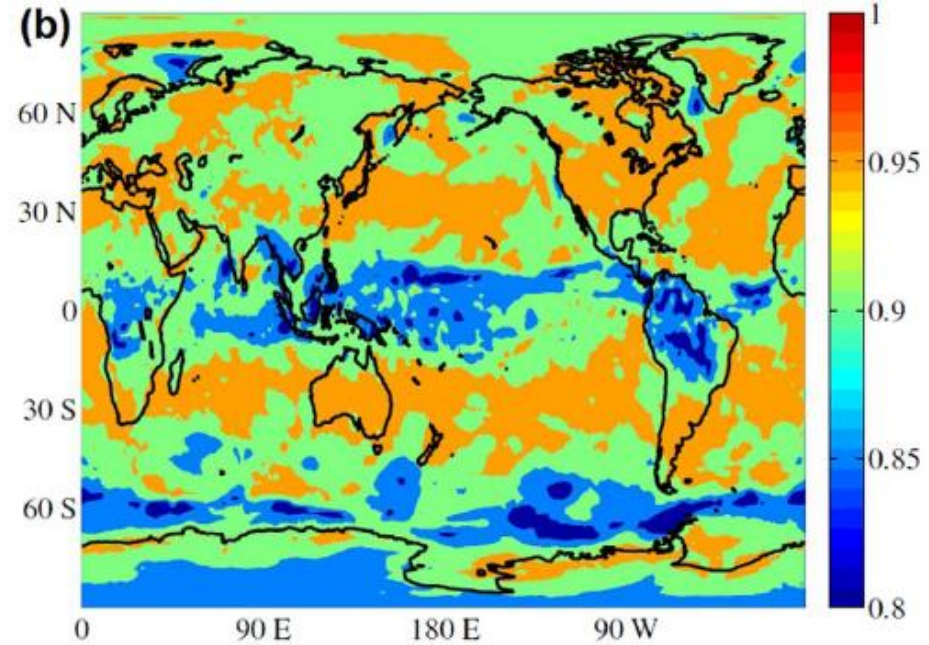
$$z_i(t) = y_i(t) - c_i(t)$$

*Shannon Entropy*: SAT anomaly is normalized to  $\mu=0$  and  $\sigma=1$

NCEP CDAS1



ERA Interim

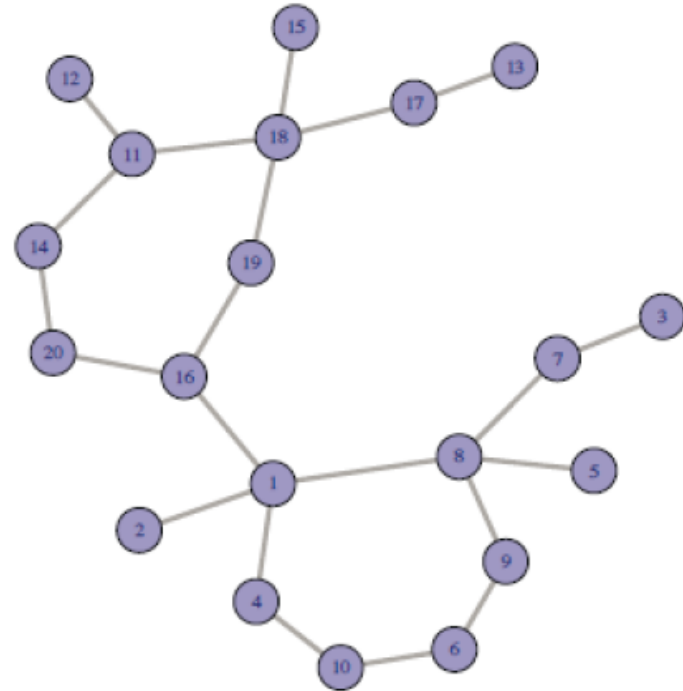


Differences are due to the presence of ***extreme values*** in one re-analysis but not in the other.

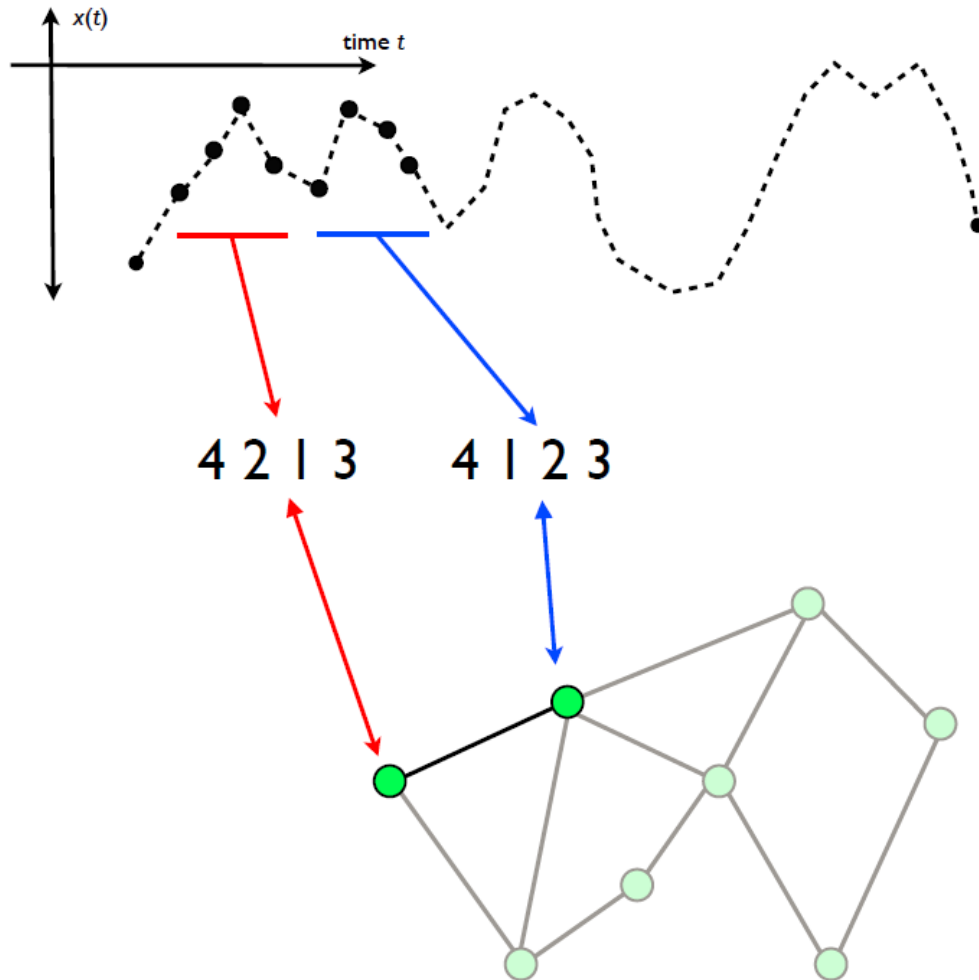
# Mapping a time series into a network

# 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 time series analysis)



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



- In each node  $i$ :

$$\sum_j w_{ij}=1$$

- Weigh of node  $i$ : the probability of pattern  $i$

$$(\sum_i p_i=1)$$

⇒ **Weighted and directed network**



# 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)

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

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

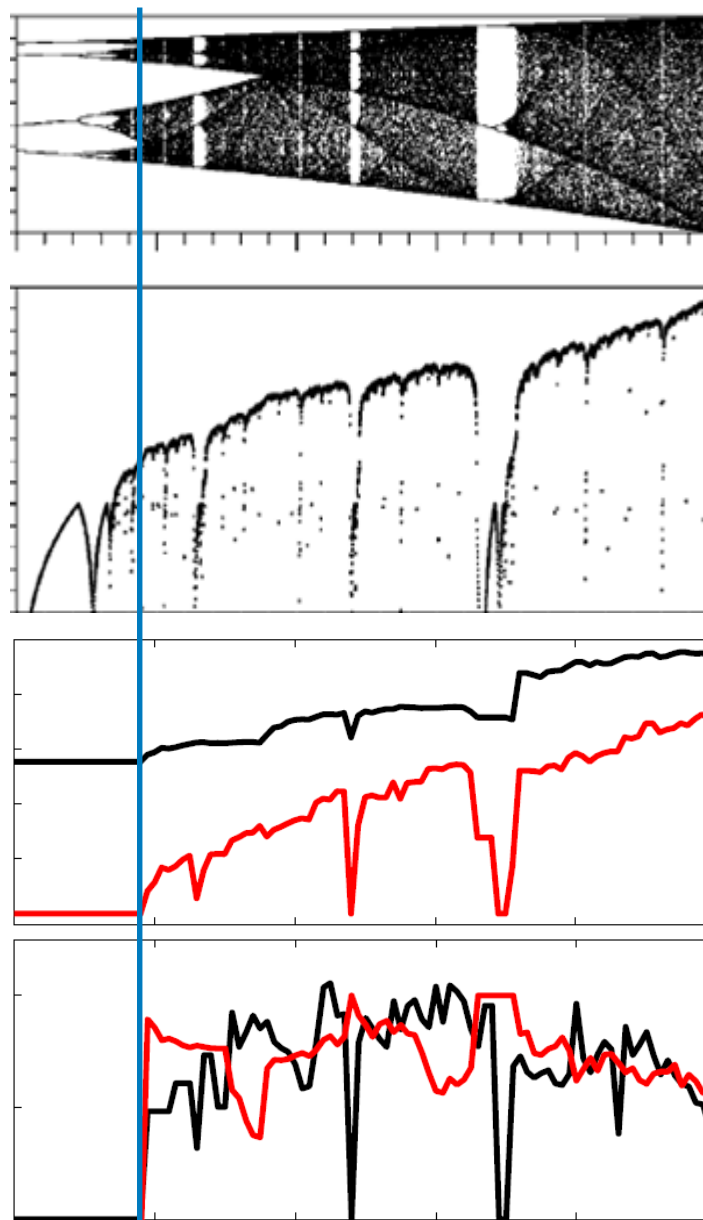
$$a_c = \frac{\sum_i \sum_{j \neq i} |w_{ij} - w_{ji}|}{\sum_i \sum_{j \neq i} (w_{ij} + w_{ji})} \quad \begin{array}{l} (0 \text{ in a fully symmetric network;} \\ 1 \text{ in a fully directed network)} \end{array}$$



# A first test with the Logistic map

$D=4$

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



Lyapunov exponent

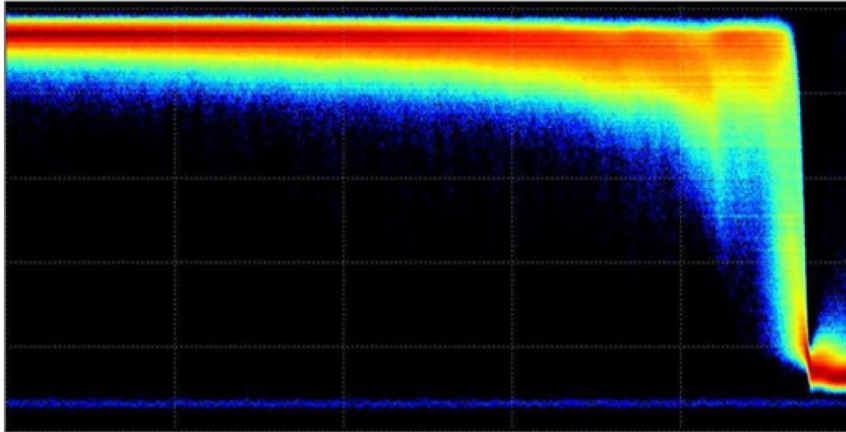
$S_p = PE$   
 $S_n = S(TPs)$

$S_{links}$   
 $a_c$

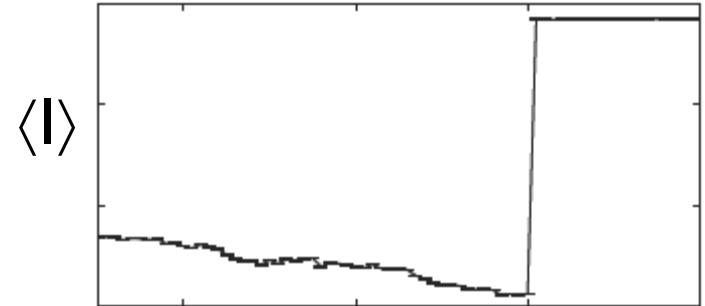
Map parameter

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

As the laser current increases



Time



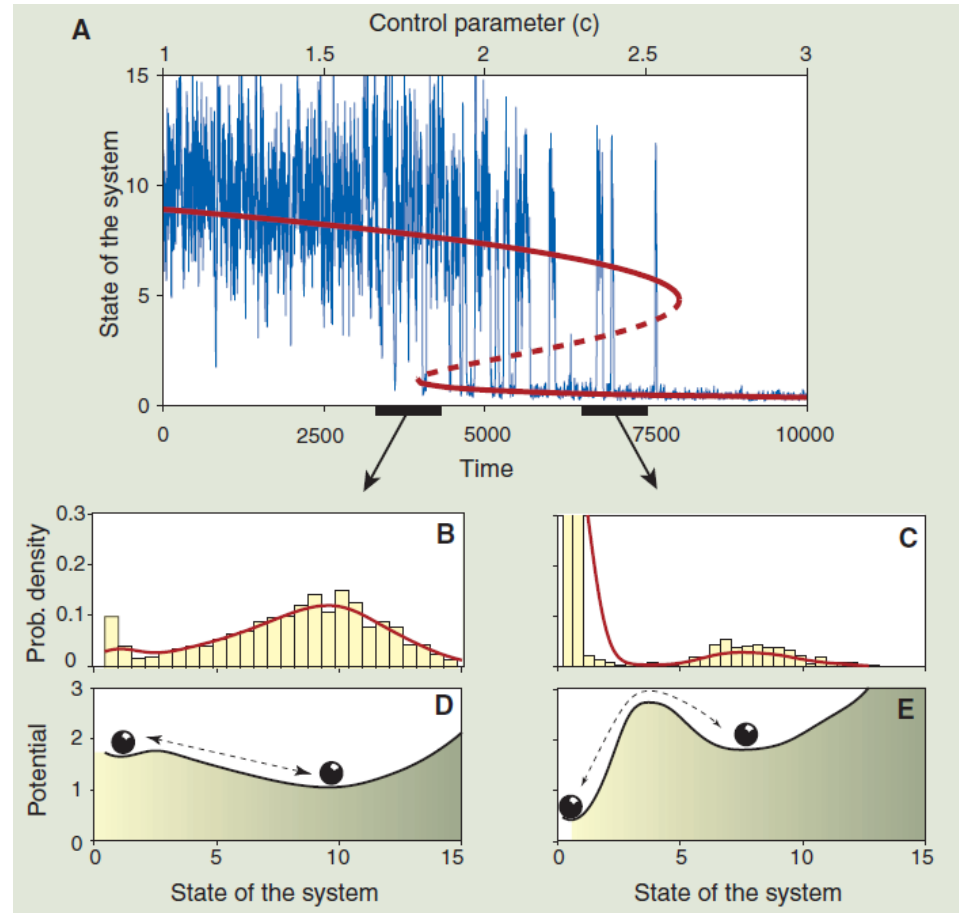
Control parameter  
(laser current)

**Apply the ordinal network method  
to detect an *early warning signal*  
of a *critical transition***

# Approaching a “tipping point”

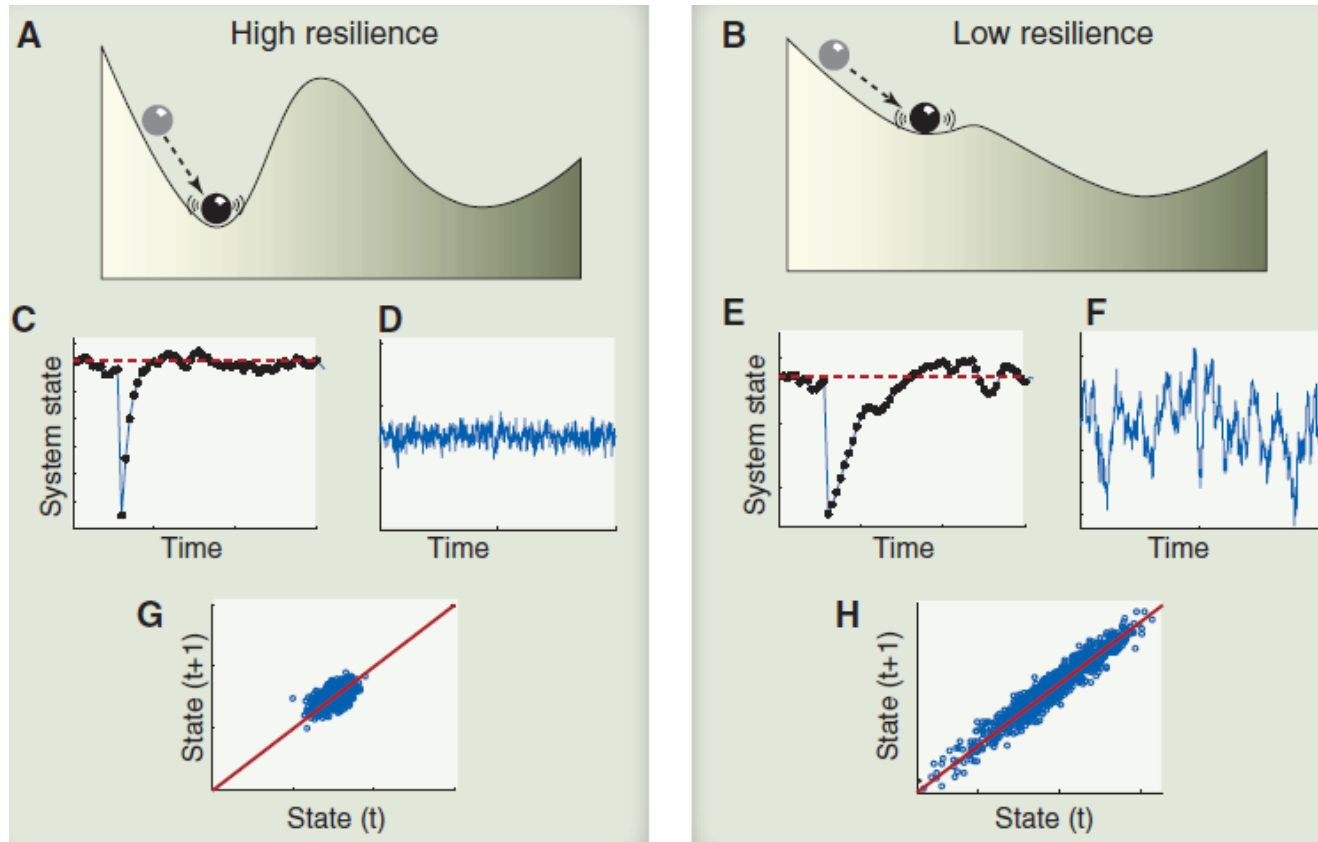


Source: E. Hernandez-Garcia (IFISC)



Scheffer et al. Science 338, 344 (2012)

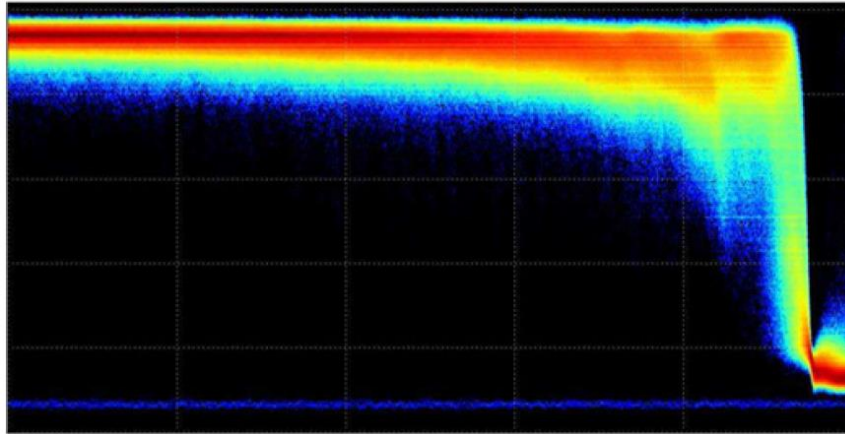
# Early warning indicators



- Approach to a bifurcation point → eigenvalue with 0 real part  
→ long recovery time of perturbations
- Critical Slowing Down: increase of autocorrelation and variance

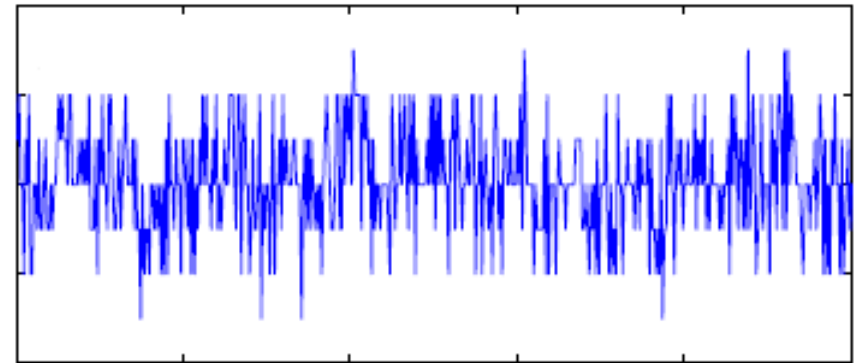
# Apply the ordinal network method to laser data

As the laser current increases



Time

Intensity @ constant current



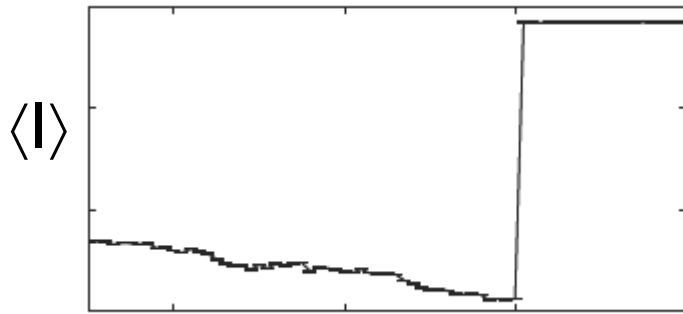
Time

- 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.

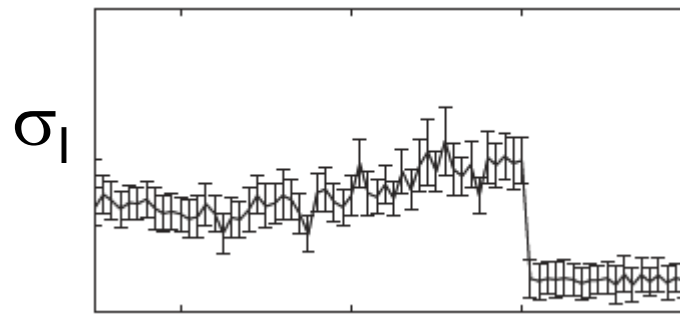
Is it possible to anticipate the switching?

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**



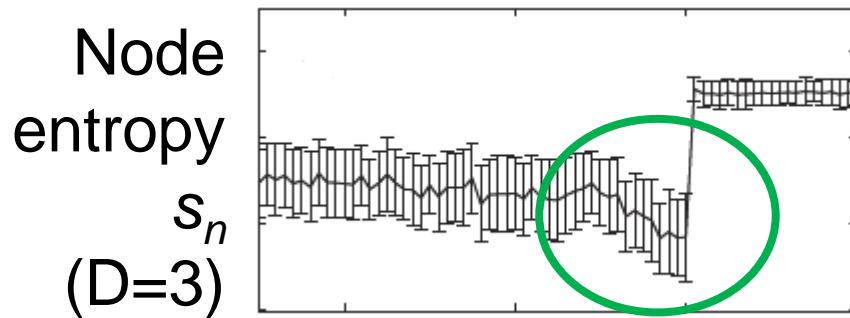
Laser current



Laser current

$L=1000$   
100 windows

$\Rightarrow$  **No  
warning**

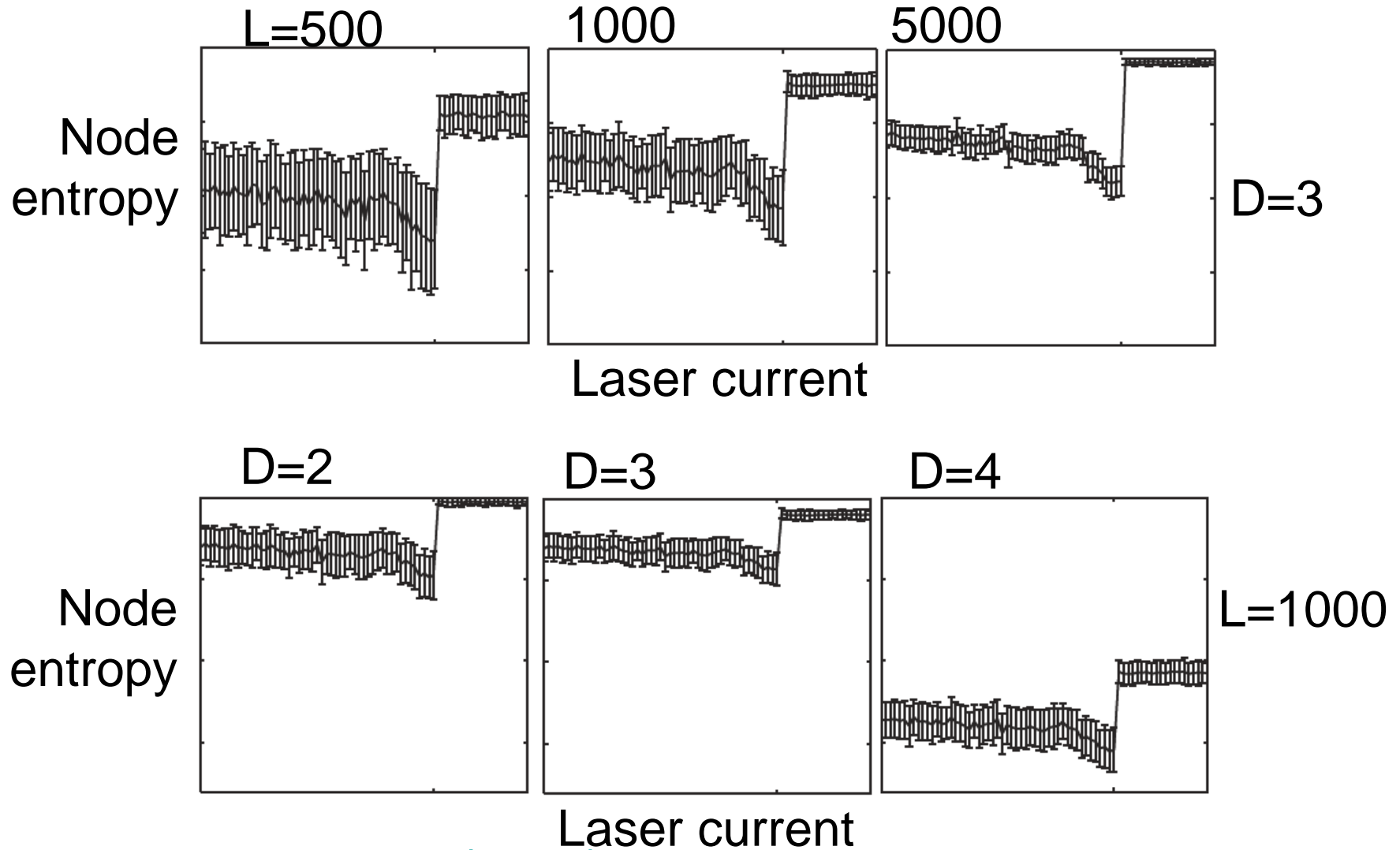


Laser current

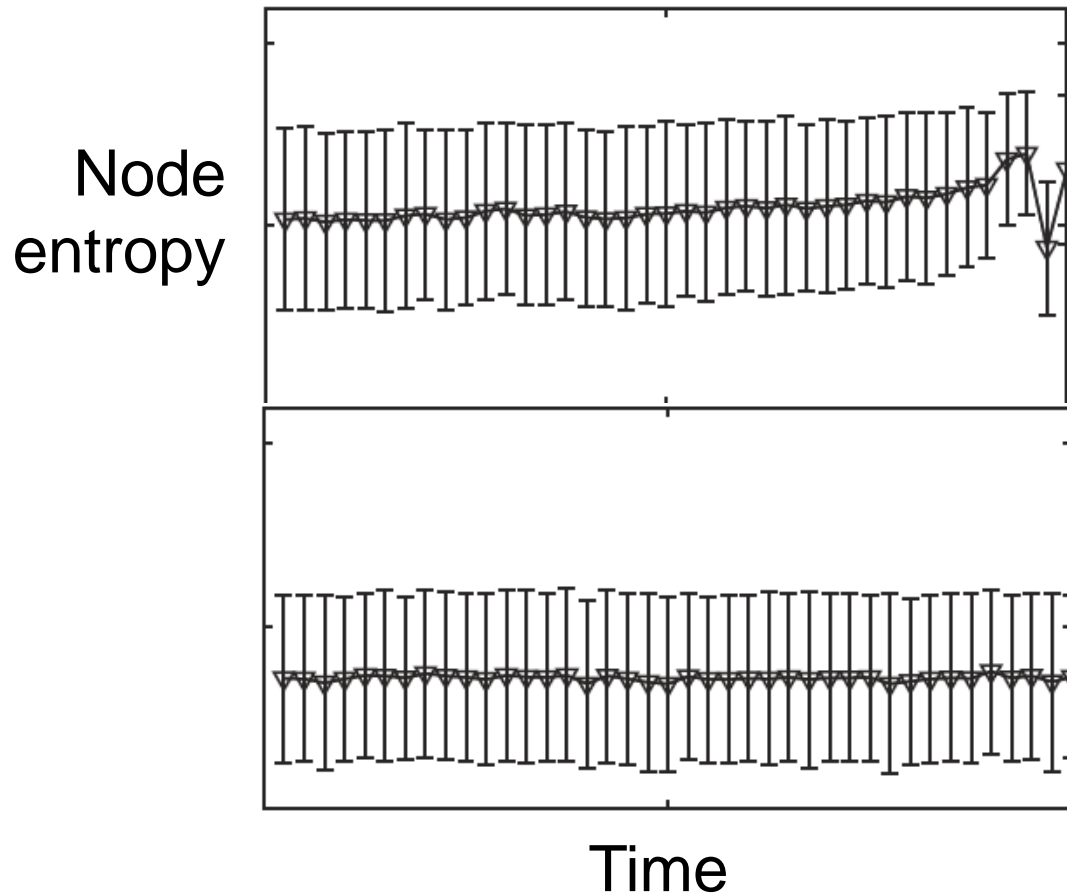
**Early warning**

$\Rightarrow$  Deterministic mechanisms  
must be involved.

The warning is robust with respect to the length of the pattern **D** and the length of the window **L**



**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.



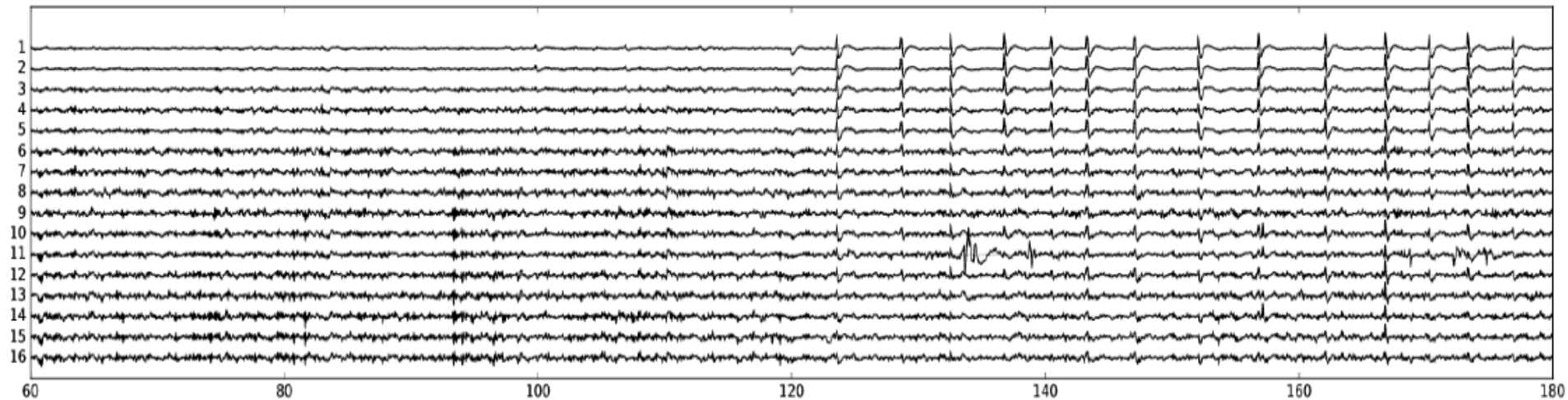
## Second application of the ordinal network method: distinguishing *eyes closed* and *eyes open* brain states

Analysis of two EEG datasets

	BitBrain	PhysioNet
	DTS1	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

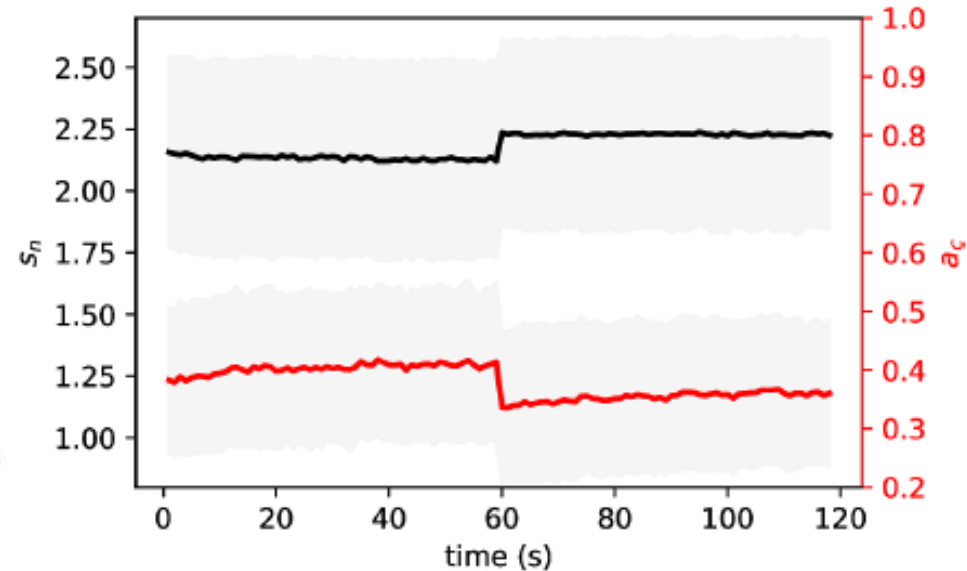
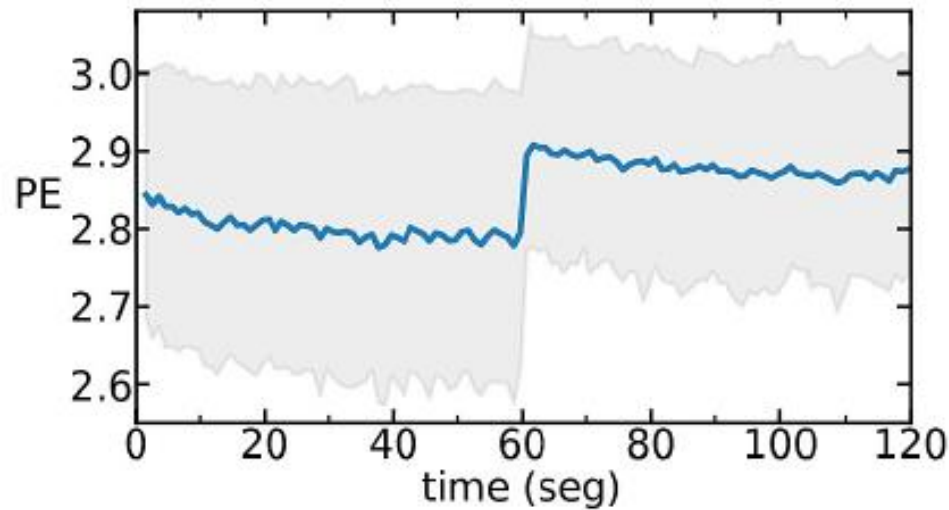
Eye closed

Eye open



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

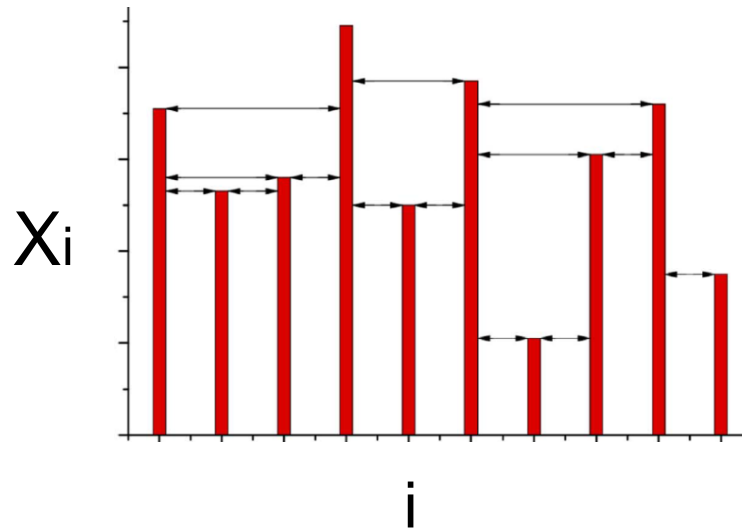
# “Randomization”: the entropies increase and the asymmetry coefficient decreases



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

⇒ **Unweighted and undirected graph**

Parameter free!

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

## Exercise

Consider the following time series:

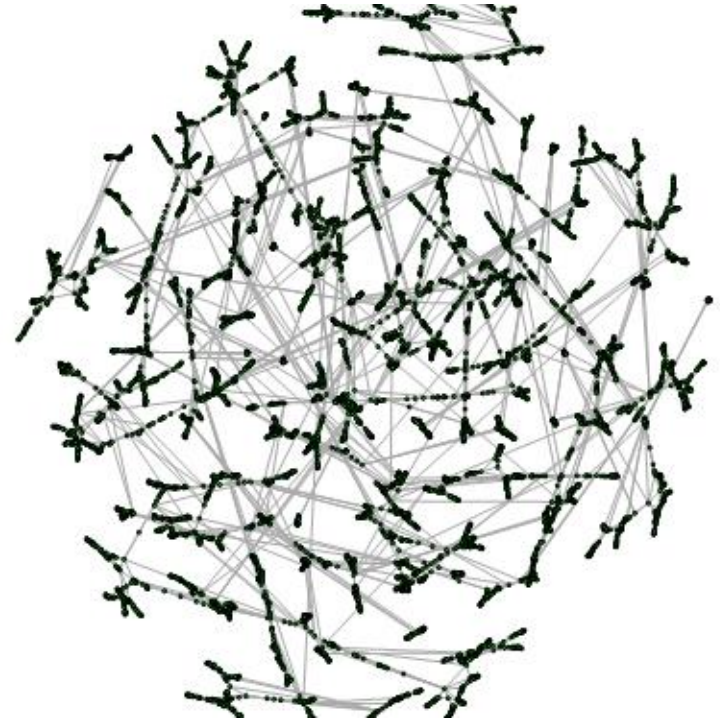
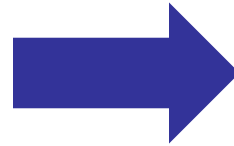
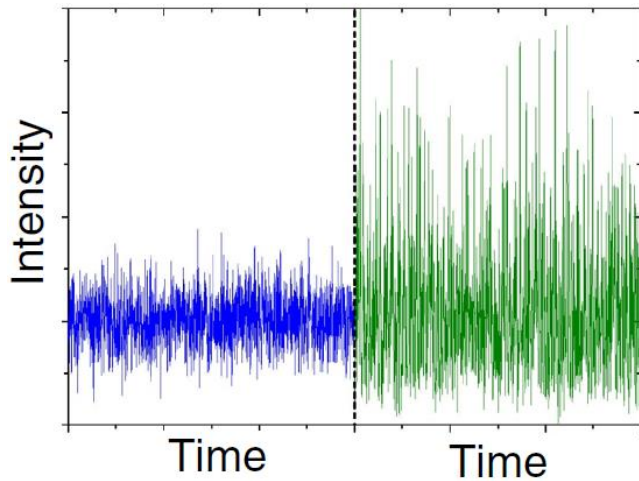
$$x = \{0.71, 0.53, 0.56, 0.89, 0.50, \\ 0.77, 0.21, 0.6, 0.72, 0.35\}.$$

How many links does each data point have?

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

# Example: the intensity emitted by a fiber laser

Low → High pump power

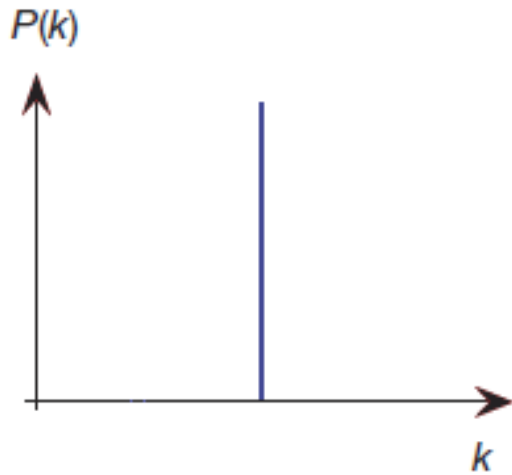
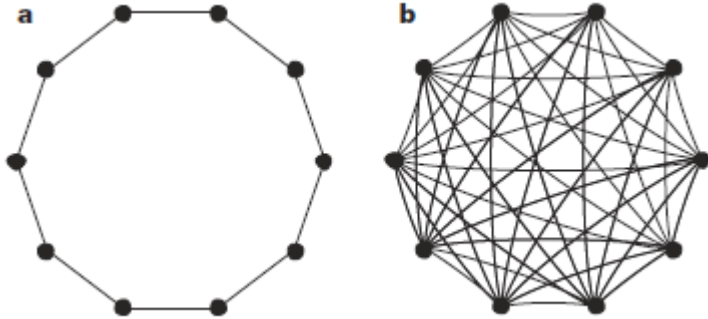


Transition  
“Laminar” → “Turbulent” regimes

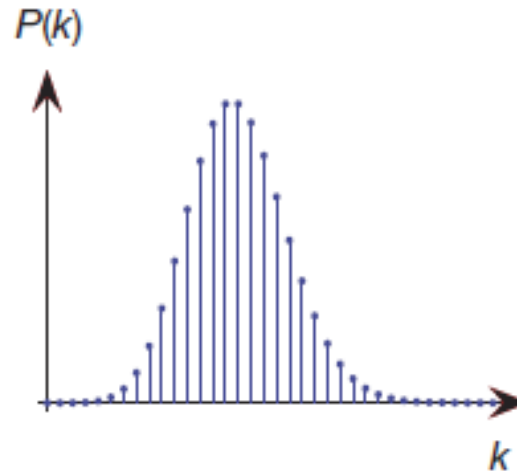
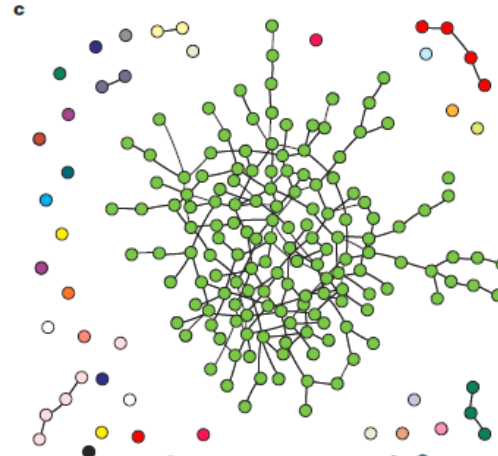
How to characterize the HV graph?

# The degree distribution: usual way to characterize a graph

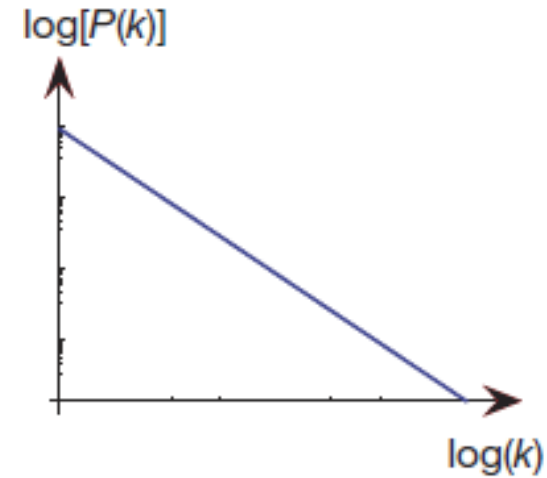
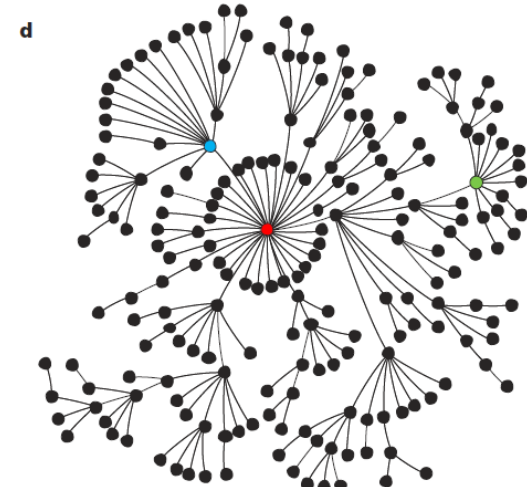
Regular



Random



Scale-free



# How is the degree distribution of the HV graph that represents a time-series?

- Exponential:  $P(k) = A \exp(-\lambda k)$

where

Uncorrelated noise  $\lambda = \lambda_c = \ln(3/2)$

Chaotic dynamics  $\lambda < \lambda_c$

Correlated noise  $\lambda > \lambda_c$

L. Lacasa and R. Toral, “*Description of stochastic and chaotic series using visibility graphs*”, Phys. Rev. E 2010.

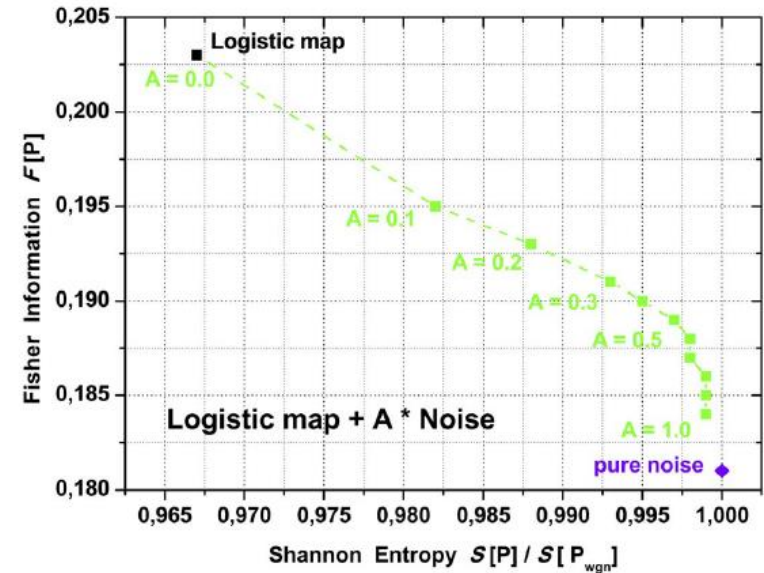
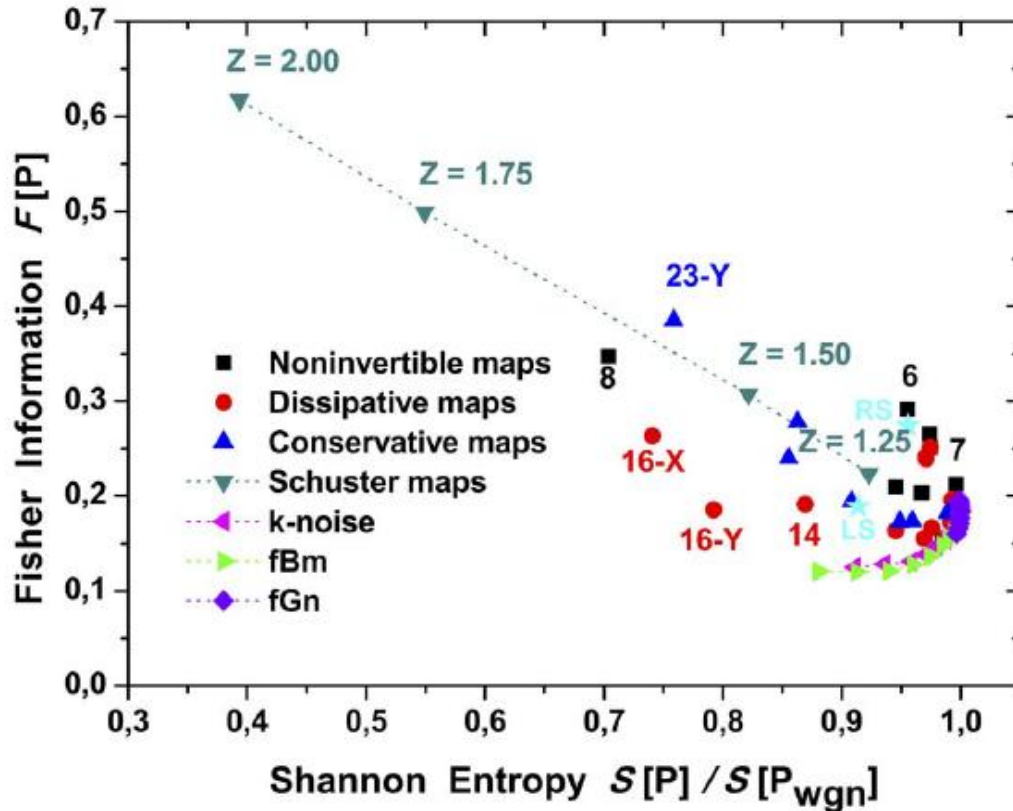
- Universal feature?
- No. But the plane (Shannon Entropy, Fisher Information) seems to work in localizing chaos and noise in different regions.

Gomez-Ravetti et al, “*Distinguishing Noise from Chaos: Objective versus Subjective Criteria Using Horizontal Visibility Graph*”, PLoS ONE 2014



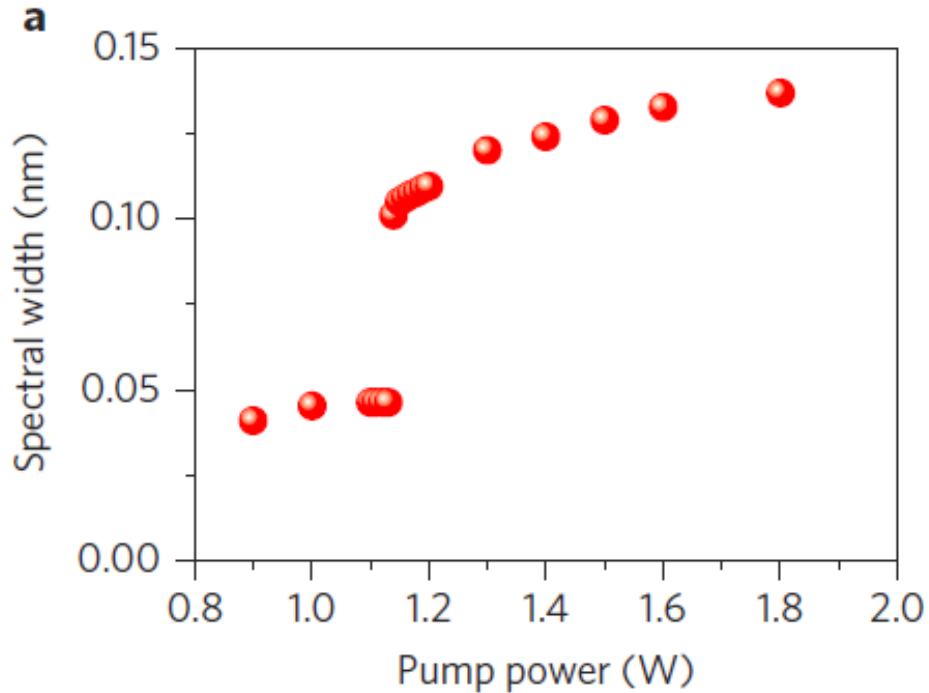
# Distinguishing noise and chaos

Shannon entropy and Fisher information are computed from the degree distribution of the HV graph,  $P(k)$

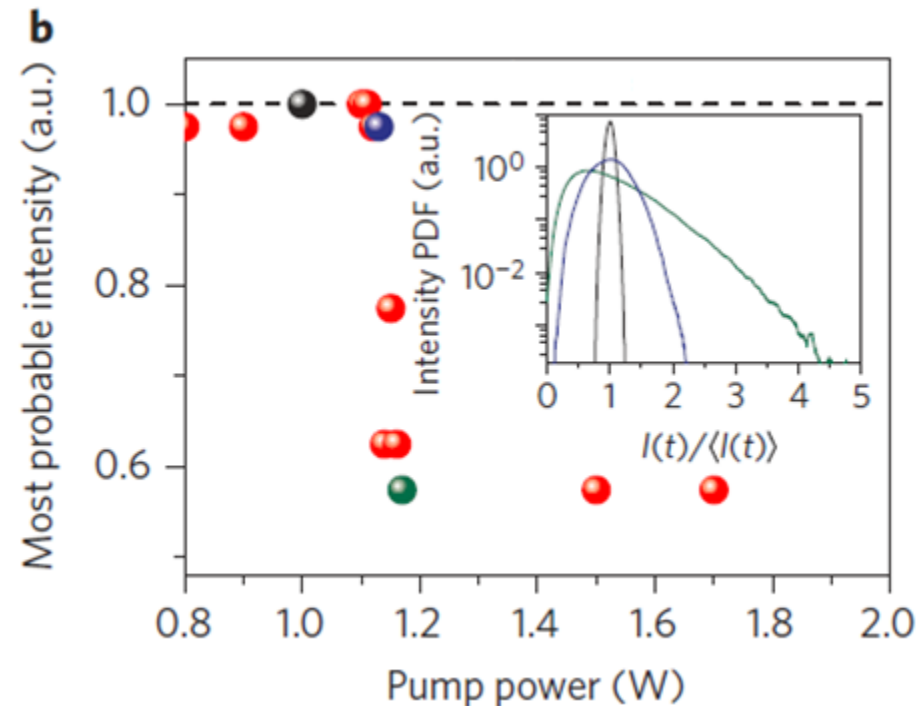


The entropy is normalized to the entropy of gaussian white noise.

# Back to the transition to optical turbulence in a fiber laser



Sharp increase of the spectral width (number of excited modes)

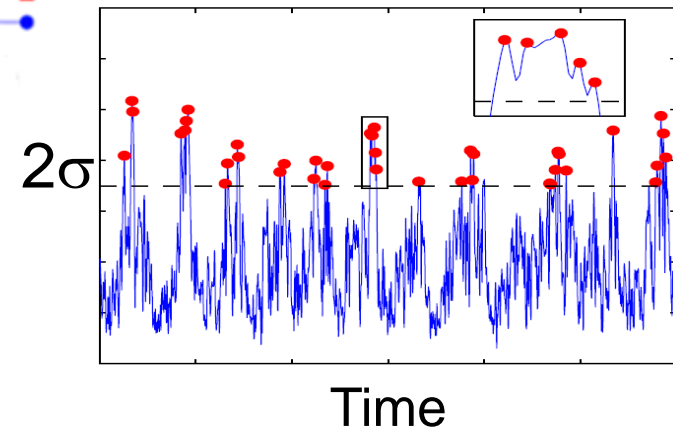
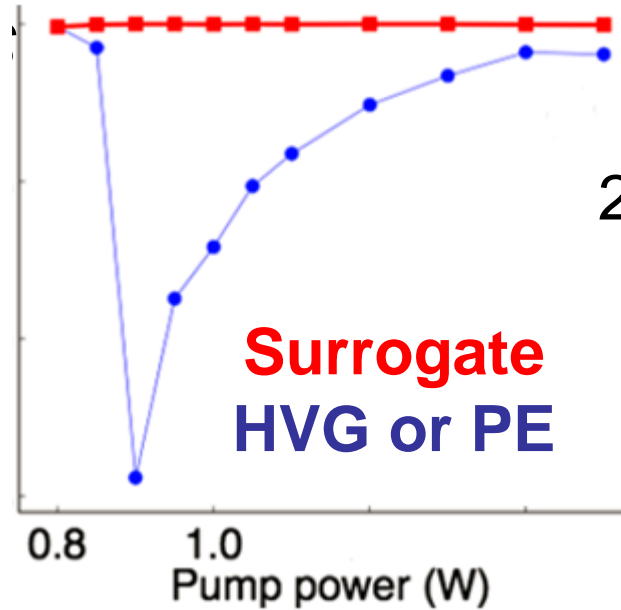
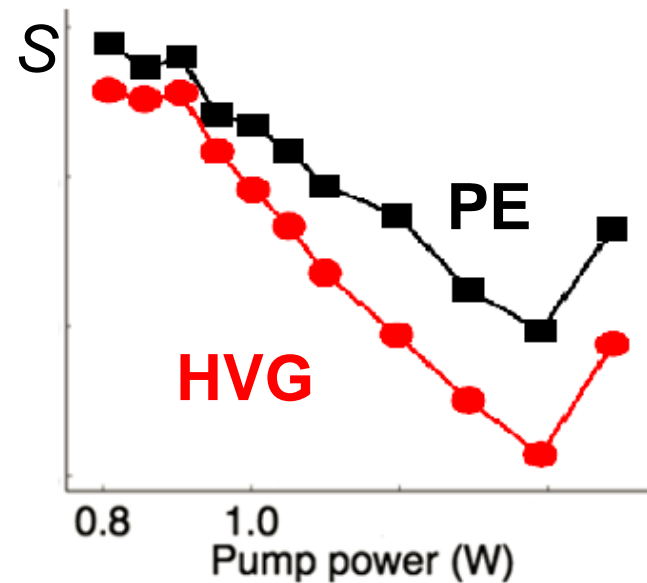


Sharp decrease of the most probable intensity value

# Different ways of calculating the **entropy $S$** uncover gradual or sharp Laminar $\rightarrow$ Turbulence transition

“Raw” data

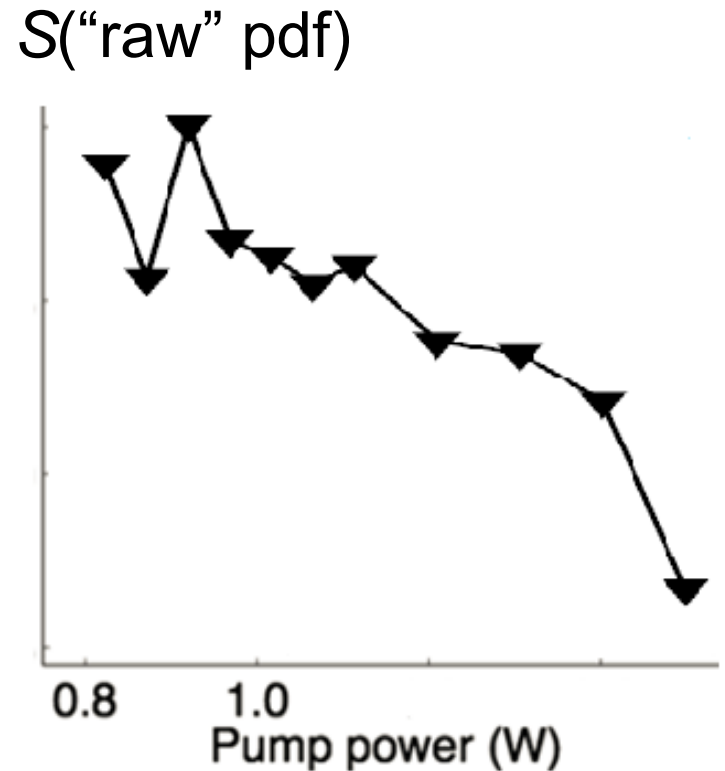
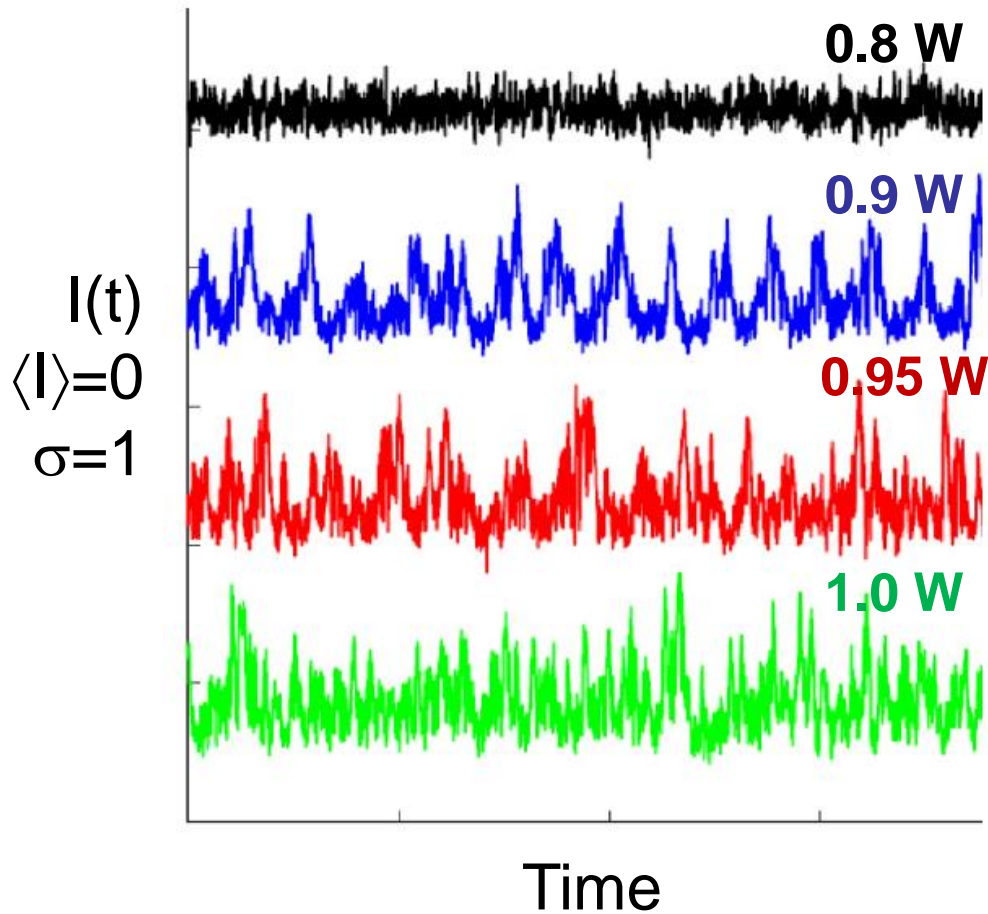
“Thresholded” data



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

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

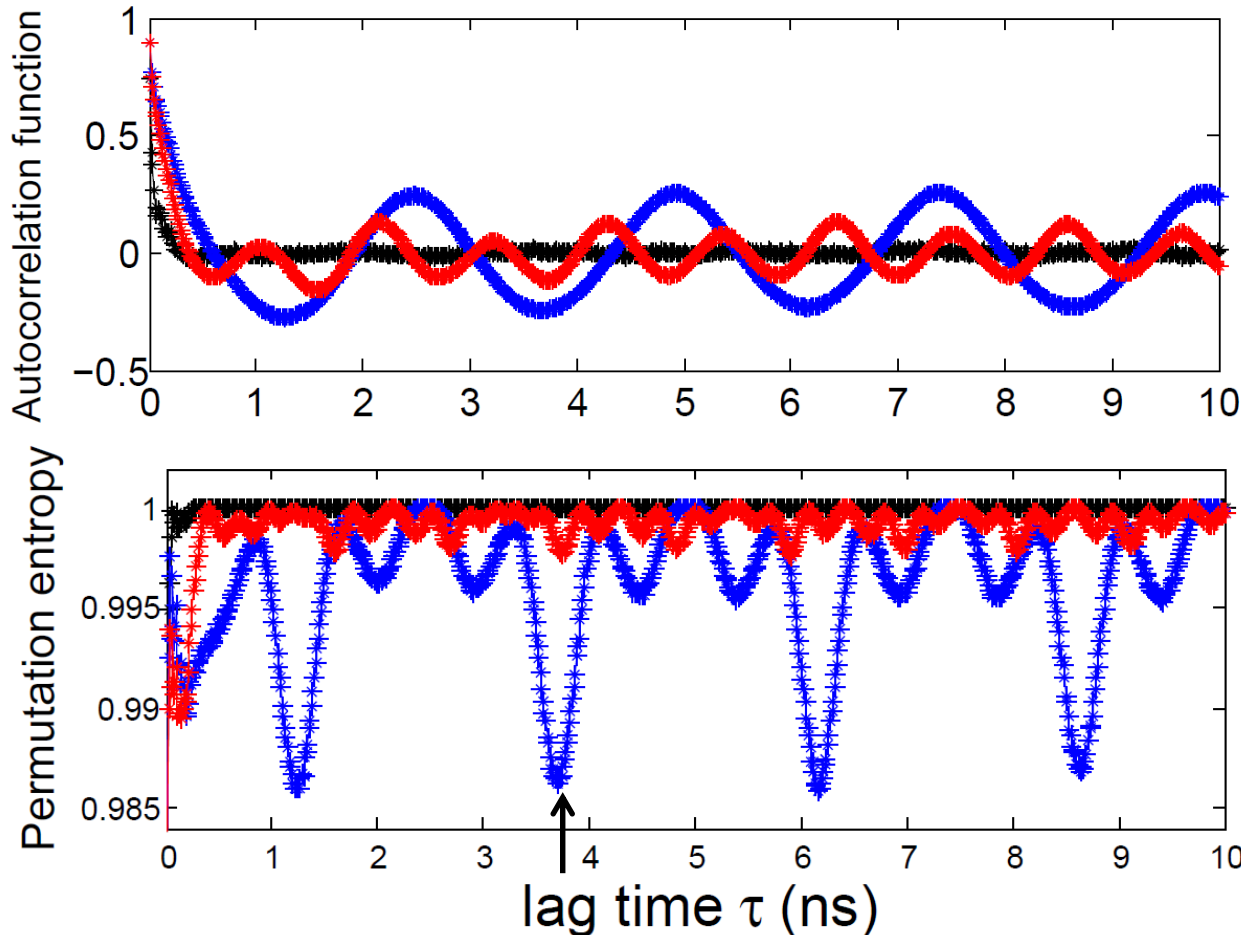
The “usual” entropy (from the distribution of all the intensity values) uncovers a different gradual transition



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

# Changing the sampling time $\tau$ identifies “hidden” time scales in the dynamics, undetected by correlation analysis

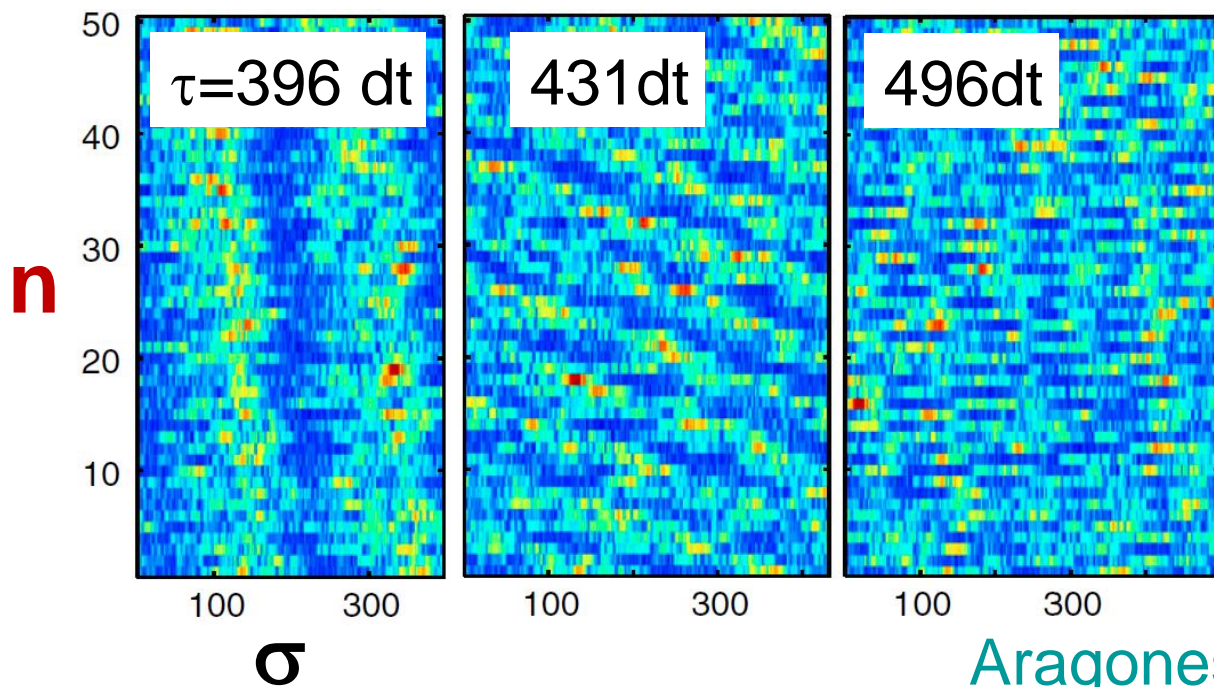
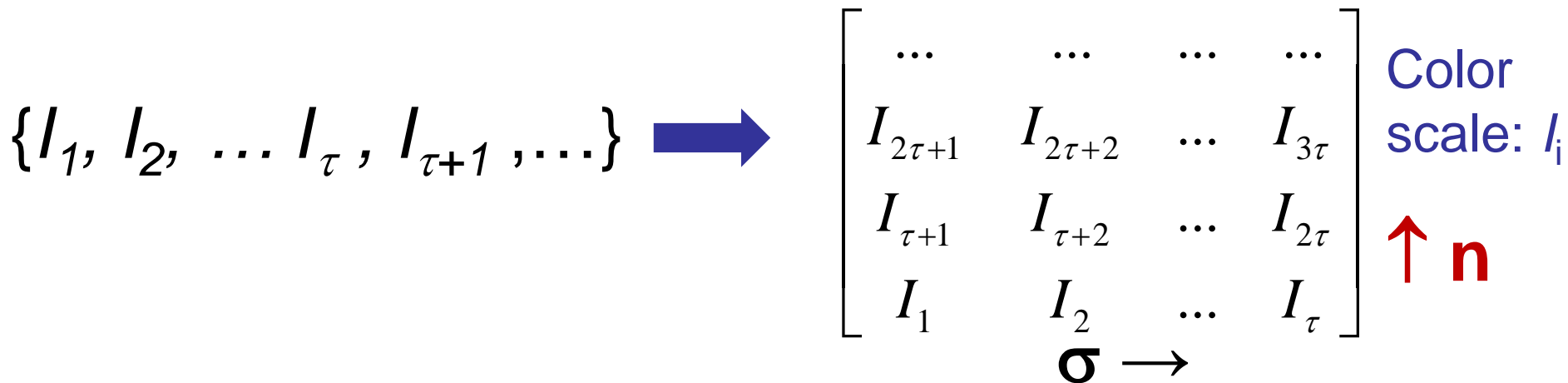
$$\{I_i, I_{i+\tau}, I_{i+2\tau}, \dots\};$$



Below transition  
at transition  
above transition

# **Space-time representation of a time series**

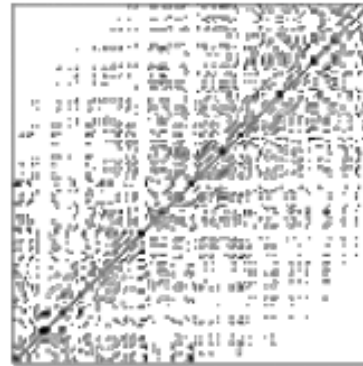
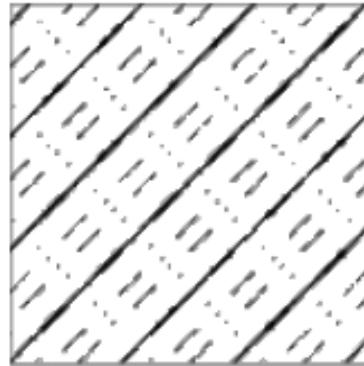
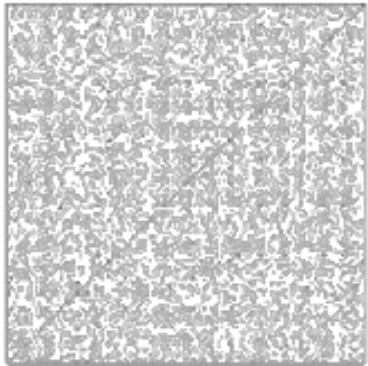
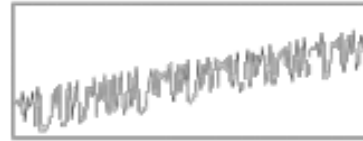
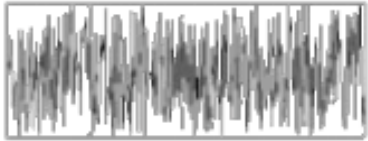
# The space-time representation of the intensity time series: a convenient way to visualize the dynamics





# Recurrence plots: another way to “visualize” a time series

- Embed the time series (find the embedding dimension and lag)
- Construct a binary matrix:  $A_{ij}=1$  if  $x_i$  and  $x_j$  are “close”, else  $A_{ij}=0$
- Plot  $A_{ij}$



Further reading: R.V. Donner, M. Small, et al. “*Recurrence-based time series analysis by means of complex network methods*”, *Int. J. Bif. Chaos* 21, 1019 (2011). 132



# Class and homework

- Investigate el Niño3.4 index (or a time-series of your own interest)

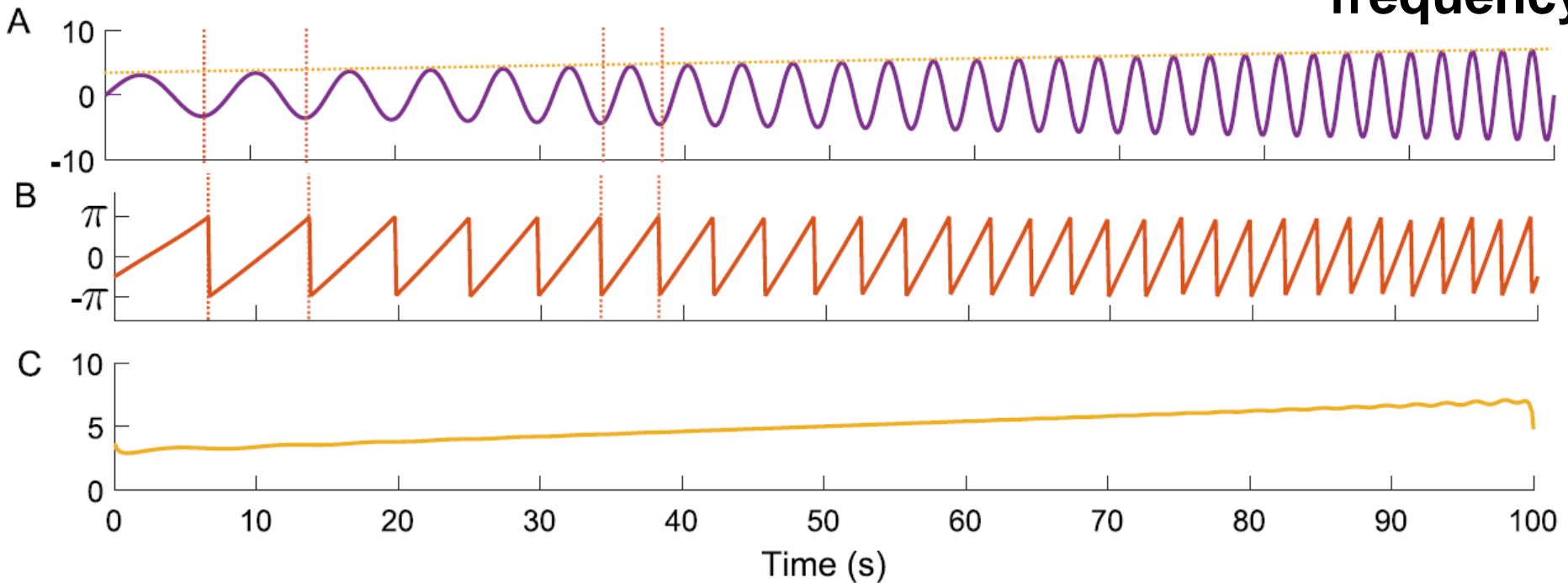
El Niño index (monthly resolution since 1870):

[https://www.esrl.noaa.gov/psd/gcos\\_wgsp/Timeseries/Data/nino34.long.anom.data](https://www.esrl.noaa.gov/psd/gcos_wgsp/Timeseries/Data/nino34.long.anom.data)

- Compute the ordinal probabilities and the permutation entropy for different values of  $D$  and  $\tau$ .
- Compare the “raw data” permutation entropy and the “thresholded data” permutation entropy.
- Analyze the influence of the threshold.
- How does the space-time representation looks like?

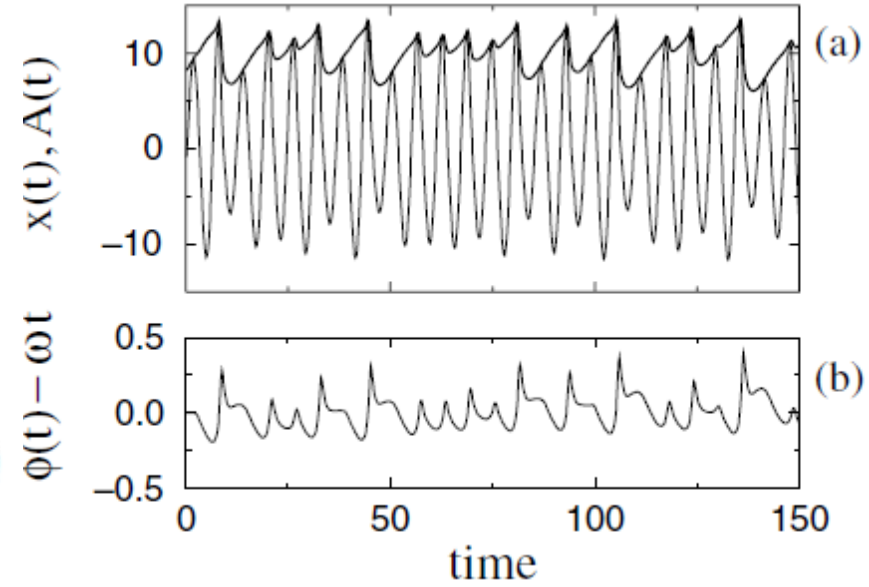
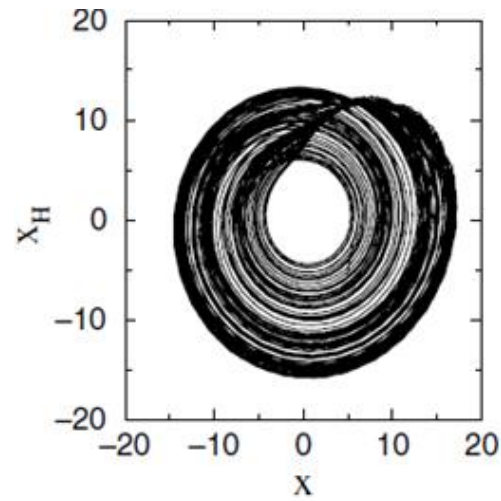
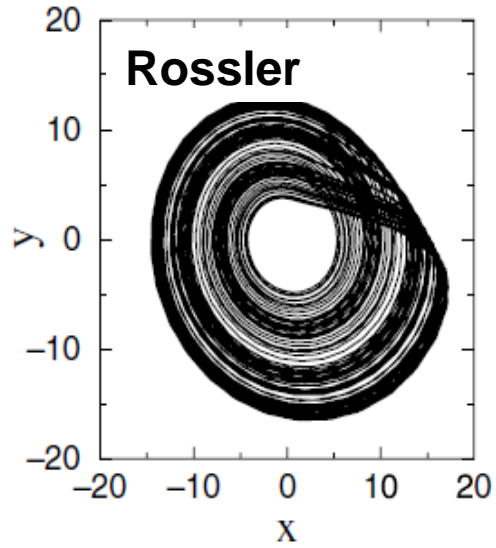
# **Extracting phase and amplitude information**

# Example: a sine wave with increasing amplitude and frequency



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

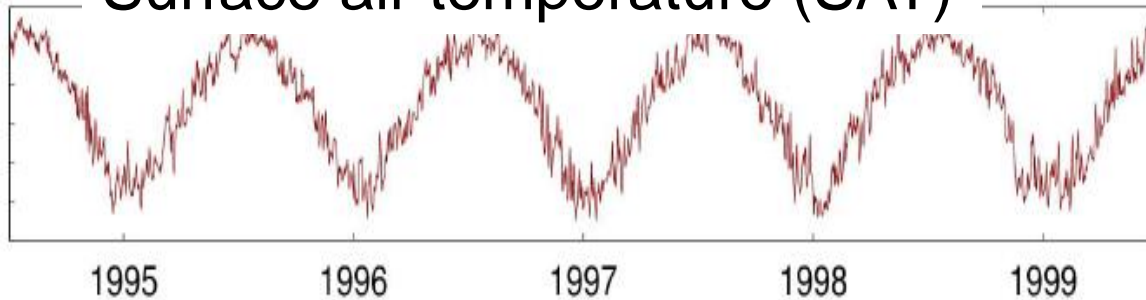
## Second example



# Third example

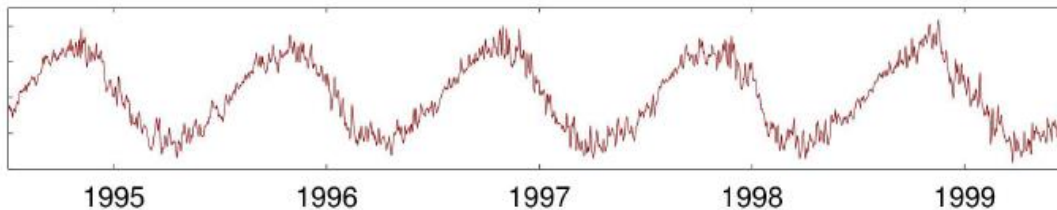
Surface air temperature (SAT)

x

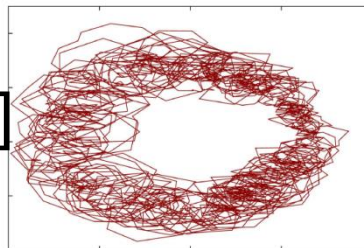


■  $HT[\sin(\omega t)] = \cos(\omega t)$

HT[x]



$y = HT[x]$



x

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

# Hilbert transform

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

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

A word of warning:

Although formally  $a(t)$  and  $\varphi(t)$  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,  $\omega(t)=d\varphi/dt$** , coincides with the dominant frequency in the power spectrum.

# 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.

# El Niño/La Niña-Southern Oscillation (ENSO)

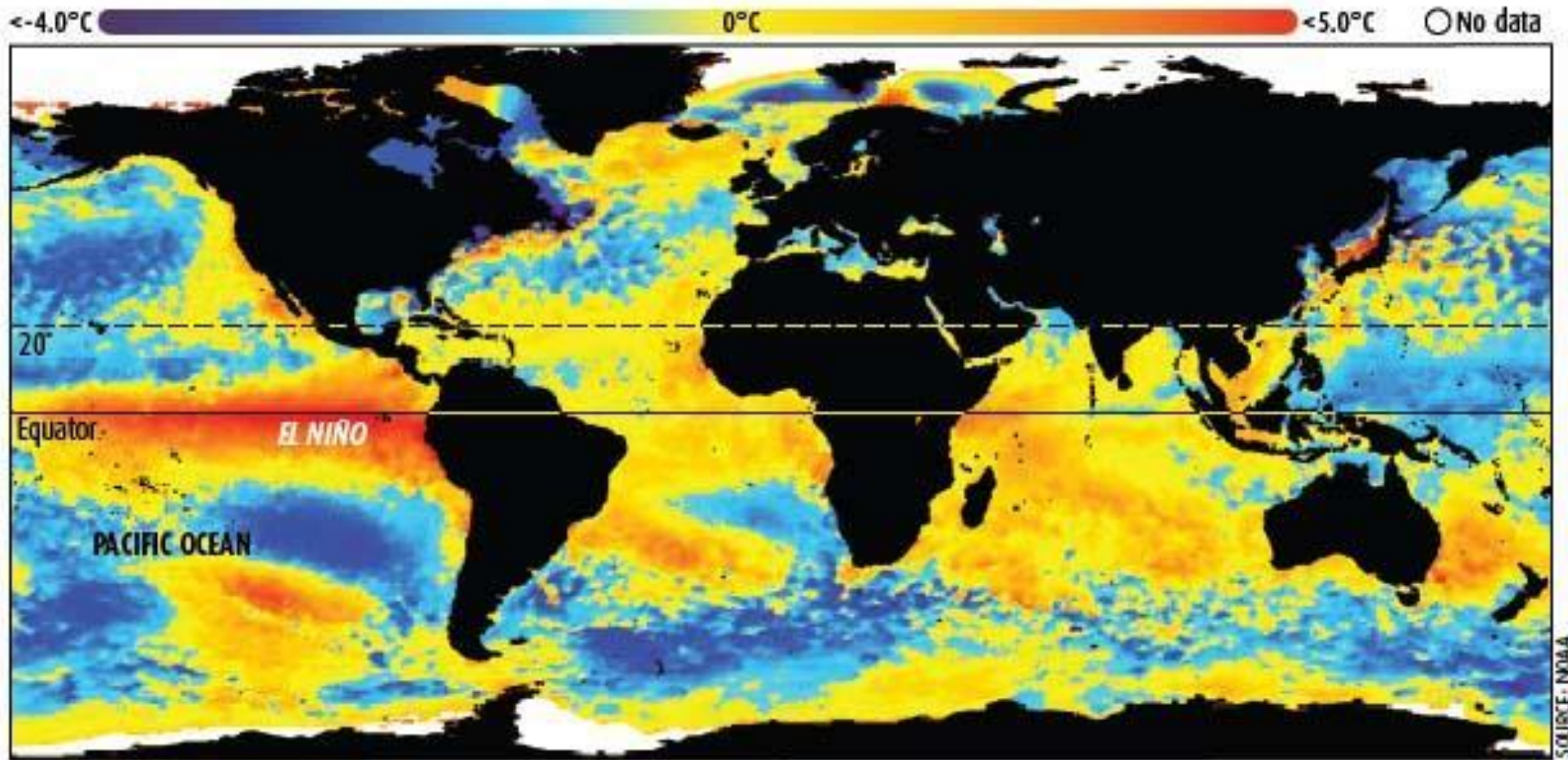
Is the most important climate phenomena on the planet

- Occurs across the tropical Pacific Ocean with  $\approx$  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.



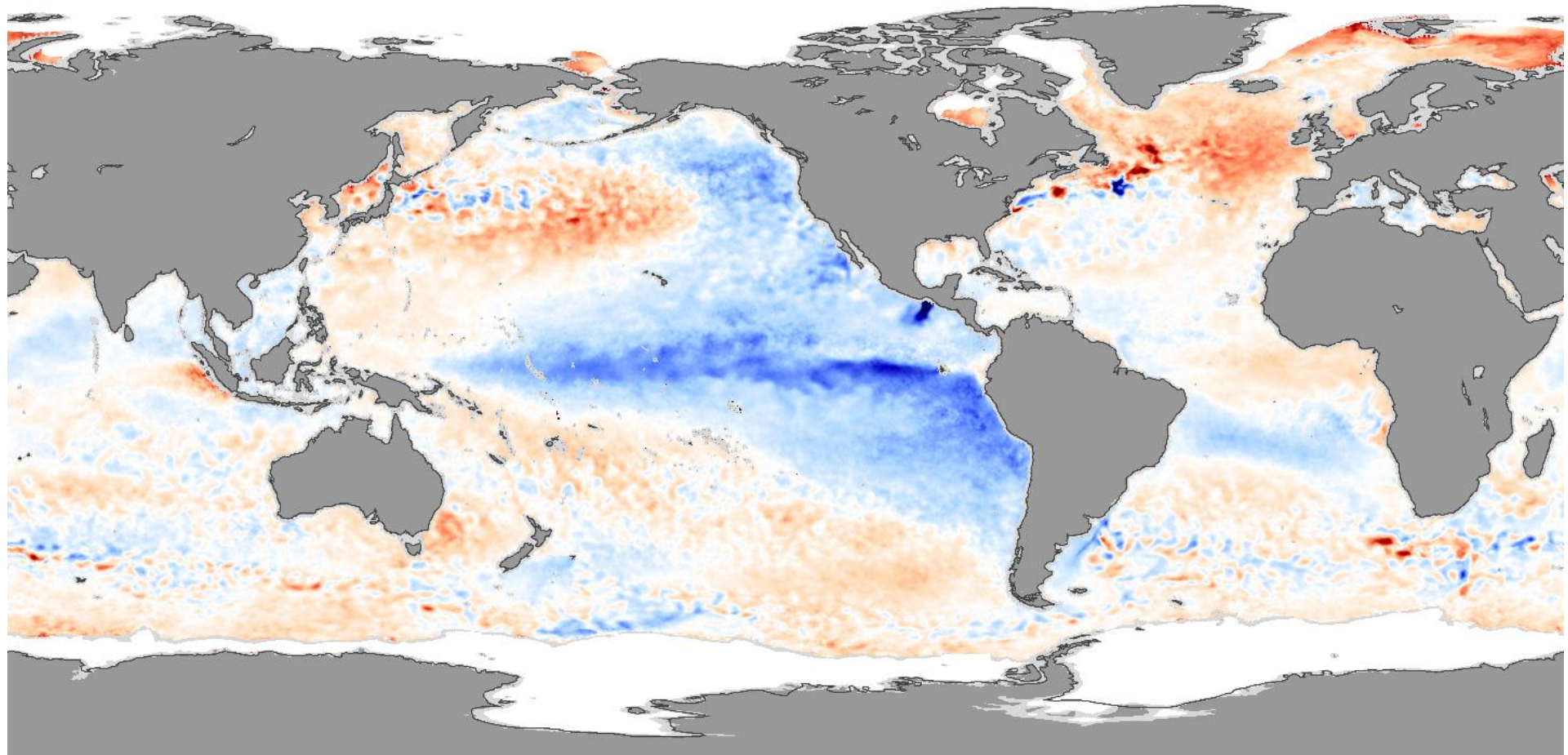
# January 1998: El Niño

## How ocean surface temperature differed from average



Satellite imagery of sea surface temperatures shows the strong El Niño that helped make 1998 one of the hottest years on record (Image: National Oceanic and Atmospheric Administration, US).

# Temperature anomalies during La Niña (November 2007)

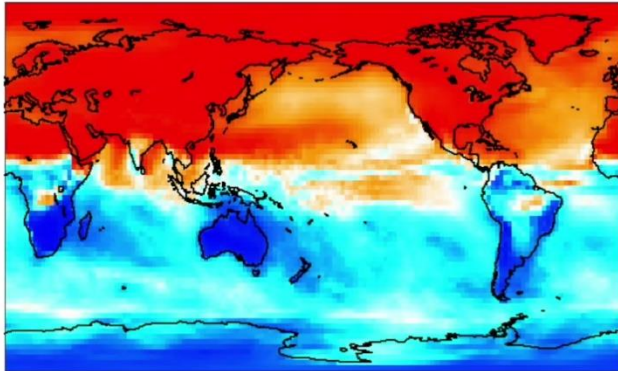




# Hilbert phase dynamics: temporal evolution of the cosine of the phase

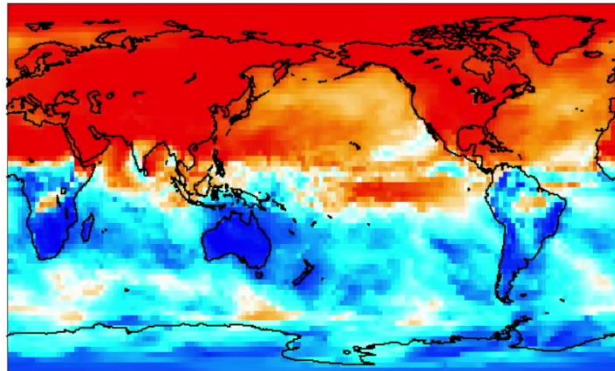
Typical year

1 July



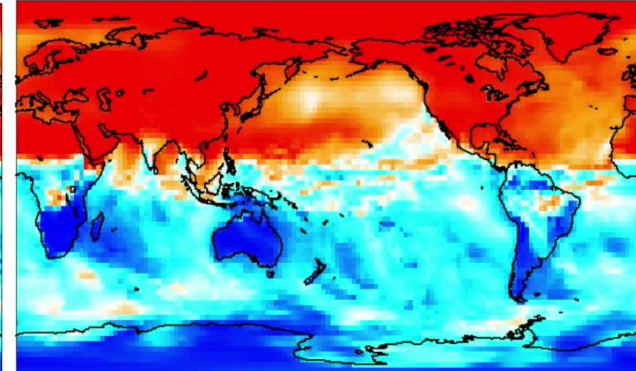
El Niño year

1 July



La Niña year

1 July



# Changes in Hilbert amplitude and frequency detect interdecadal variations in surface air temperature (SAT)

## The data:

- Spatial resolution  $2.5^{\circ} \times 2.5^{\circ} \Rightarrow 10226$  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,  $\sigma_a$ ,  $\sigma_{\omega}$

## Relative decadal variations

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

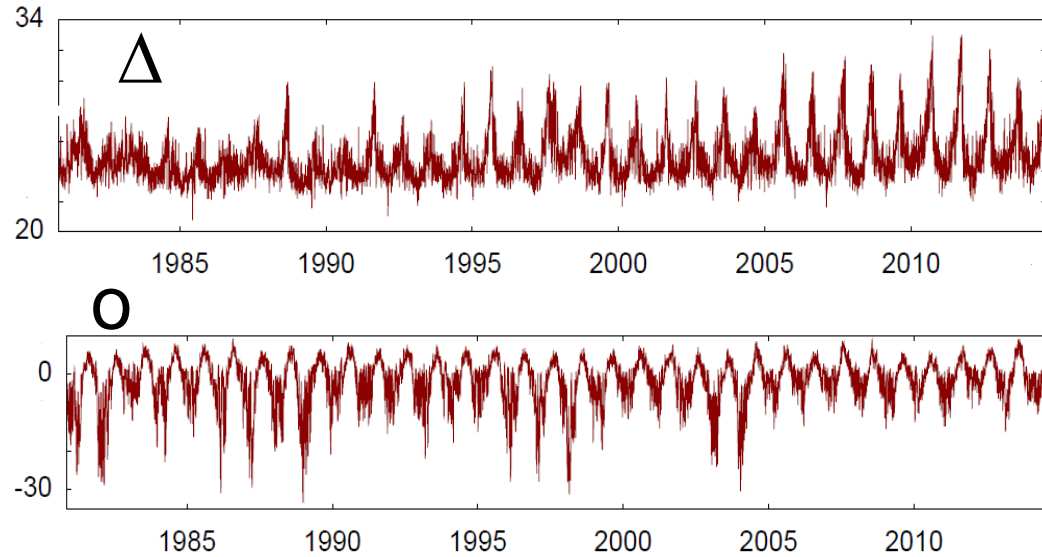
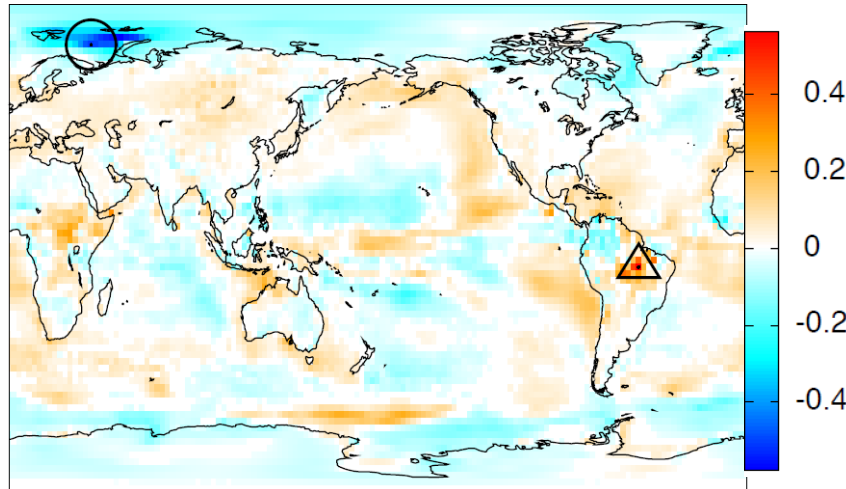
Relative variation is considered significant if:

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

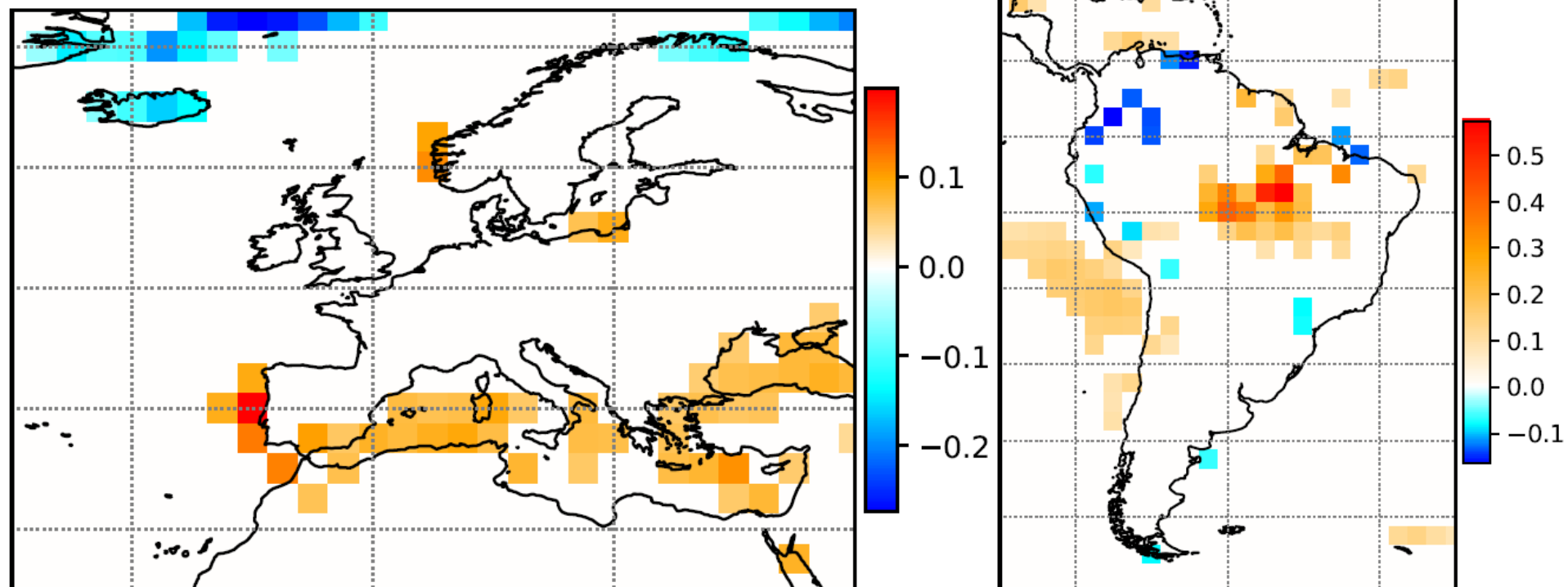
100 “block” surrogates

[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 \(2018\)](#)

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

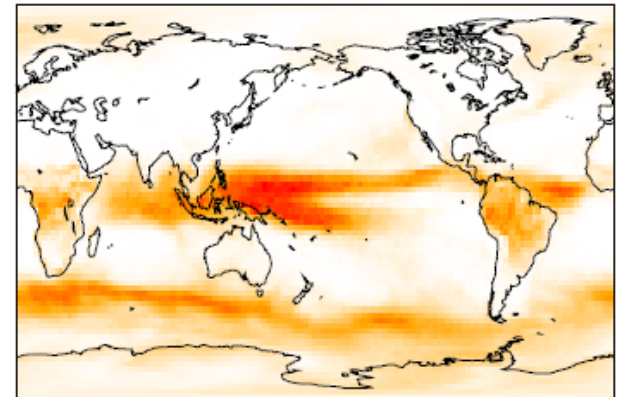


- **Decrease of precipitation:** the solar radiation that is not used for evaporation is used to heat the ground.
- **Melting of sea ice:** during winter the air temperature is mitigated by the sea and tends to be more moderated.

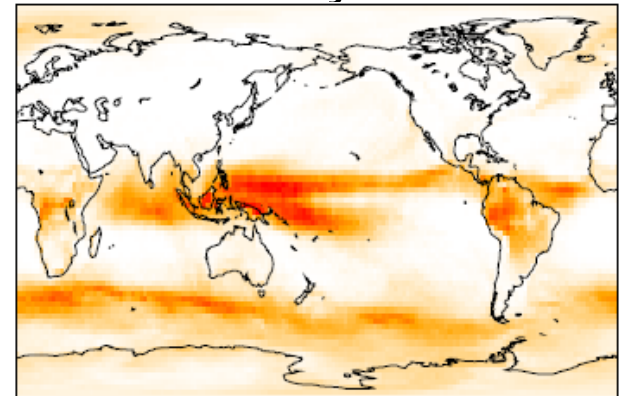


# Relative change of time-averaged Hilbert frequency consistent with a **north shift and enlargement of the intertropical convergence zone (ITCZ)**

First ten years



Last ten years





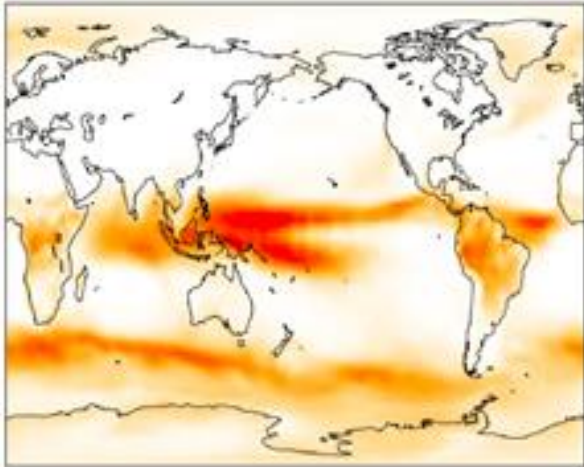
**Hilbert analysis combined  
with temporal averaging:  
another way to uncover  
temporal regularity in data**

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

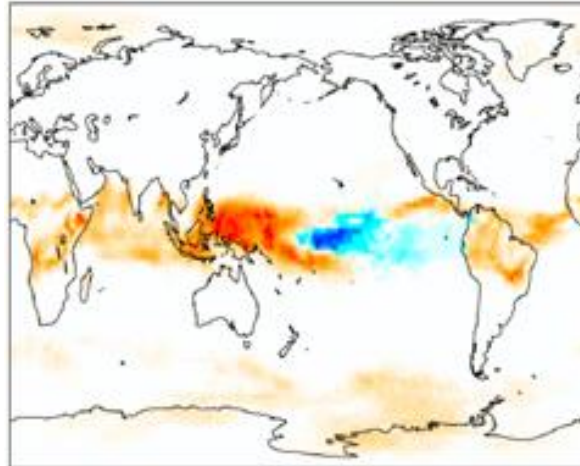
SAT  $\rightarrow$  average in a time window  $\rightarrow$  Hilbert

The colorcode shows the mean frequency (red fast, blue slow)

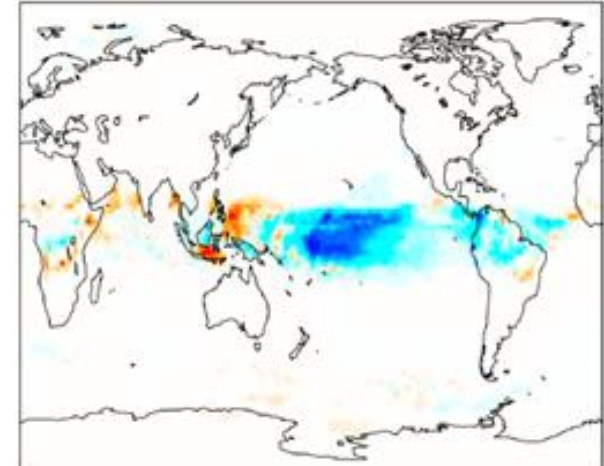
No filter



1 month



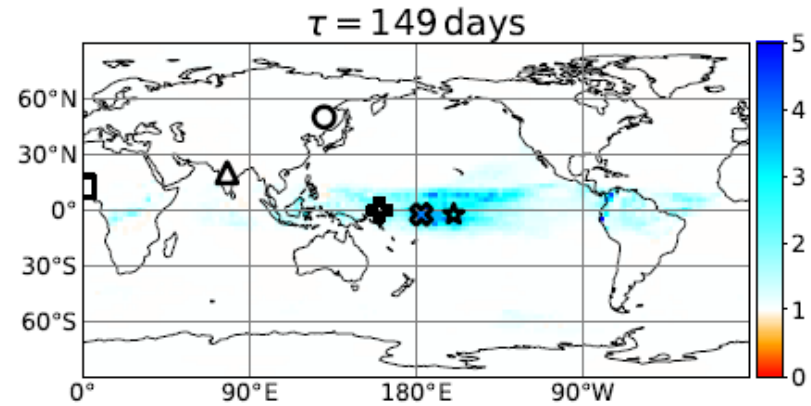
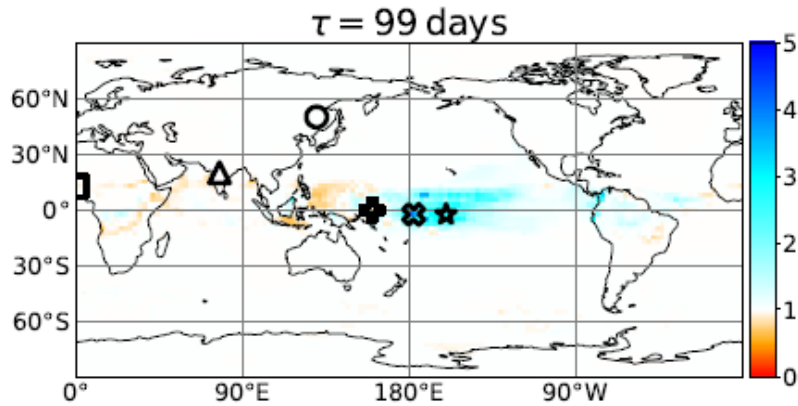
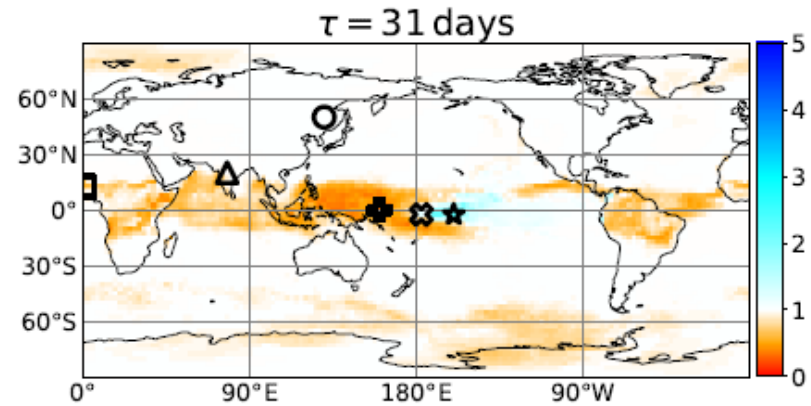
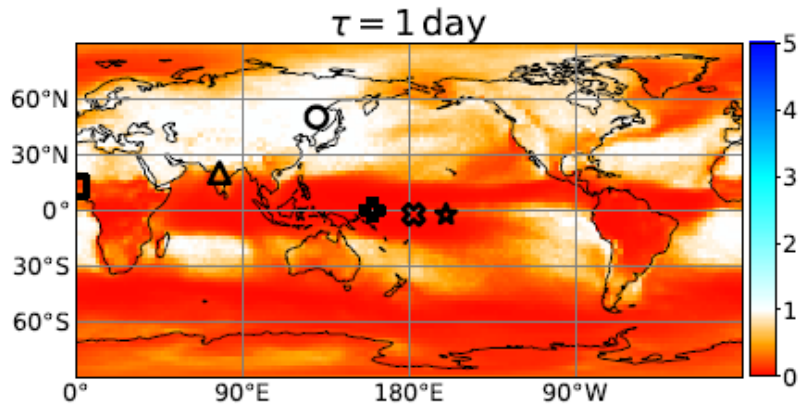
3 months



How does the phase dynamics depend on the length of the averaging window?

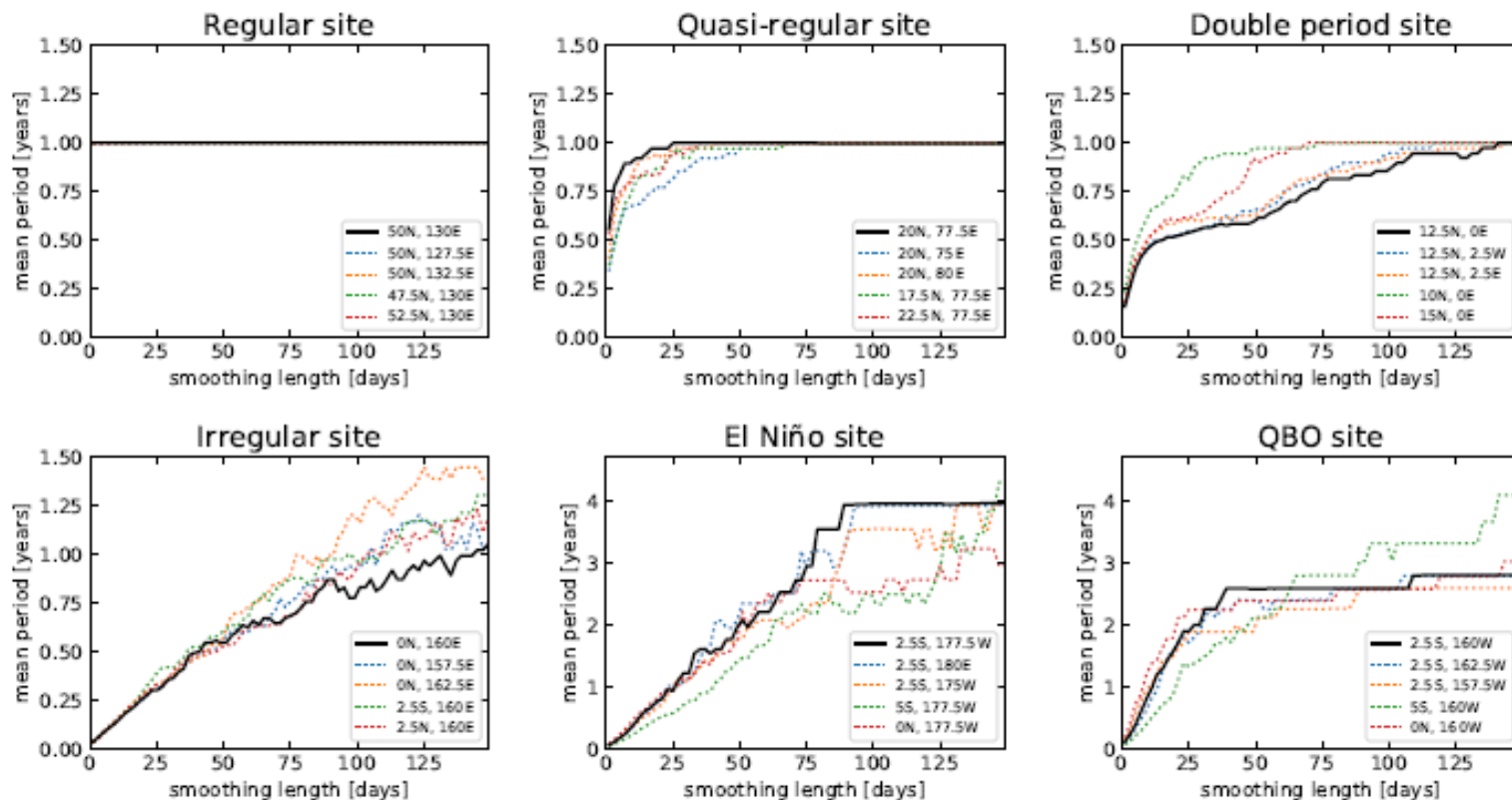
# Mean rotation period (in years)

$$\bar{T}_j = 2\pi(\Delta t / \Delta\varphi_j)$$



We will investigate some geographical locations: regular (circle), quasi-regular (triangle), double period (square), irregular (plus), El Niño (cross), quasi-biennial oscillation (QBO, star)

# Variation of mean rotation period with the smoothing length



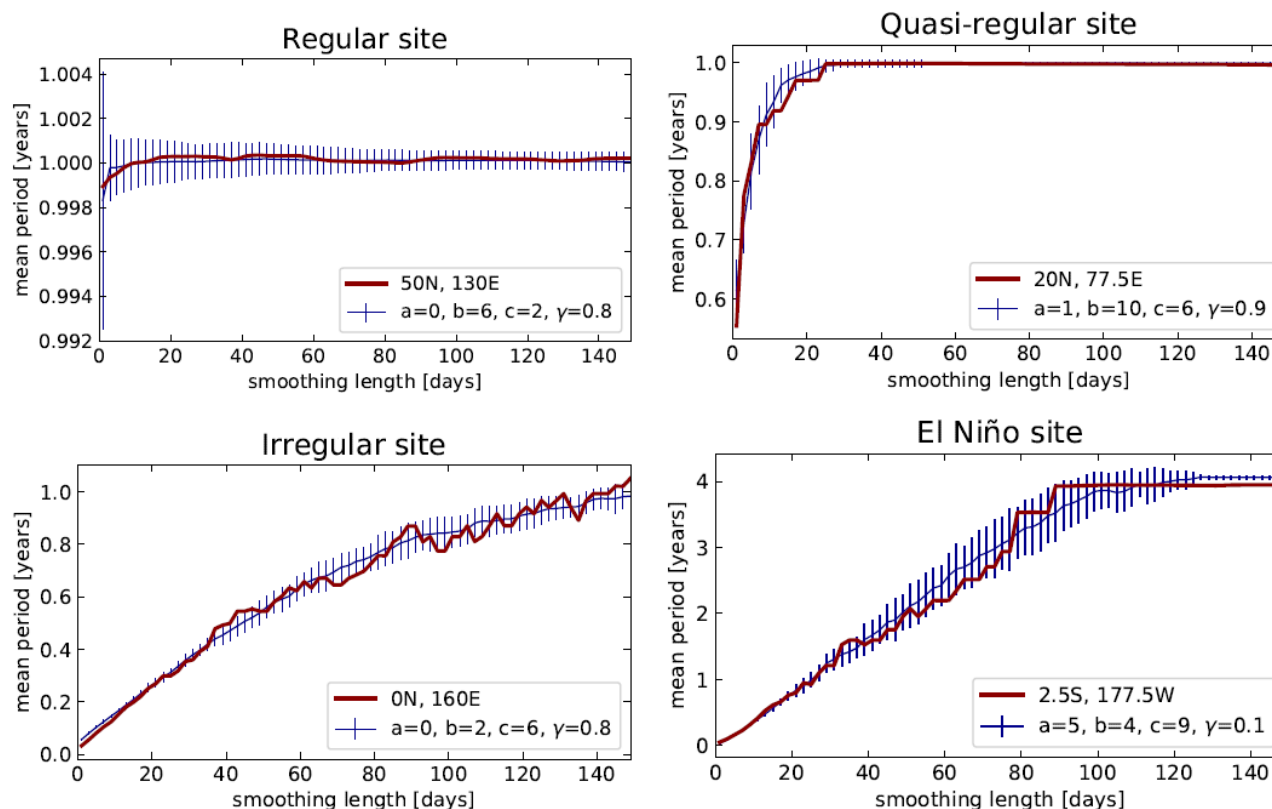
- Regular, quasi-regular and double period sites: a plateau at 1 year is found when increasing the averaging window (sooner or latter)
- Irregular site: no plateau.
- In El Niño and QBO sites: plateau at 4 and  $\approx 2.5$  years respectively.

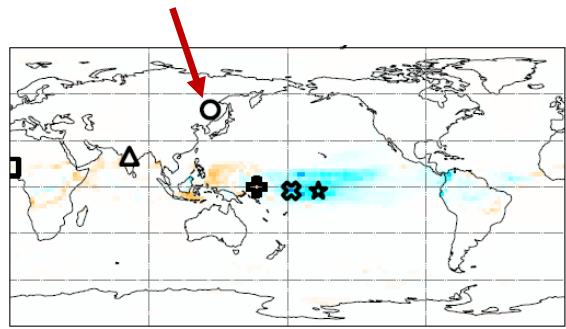
Can we understand the variation of  $T$  with  $\tau$  ?

# Simple model: two sinusoidal oscillations (the annual cycle and a slower oscillation) and autoregressive noise

$$S(t) = \frac{a\sqrt{2} \cos(\omega_{sl}t) + b\sqrt{2} \cos(\omega_y t) + c\varepsilon(t)}{\sqrt{a^2 + b^2 + c^2}}$$

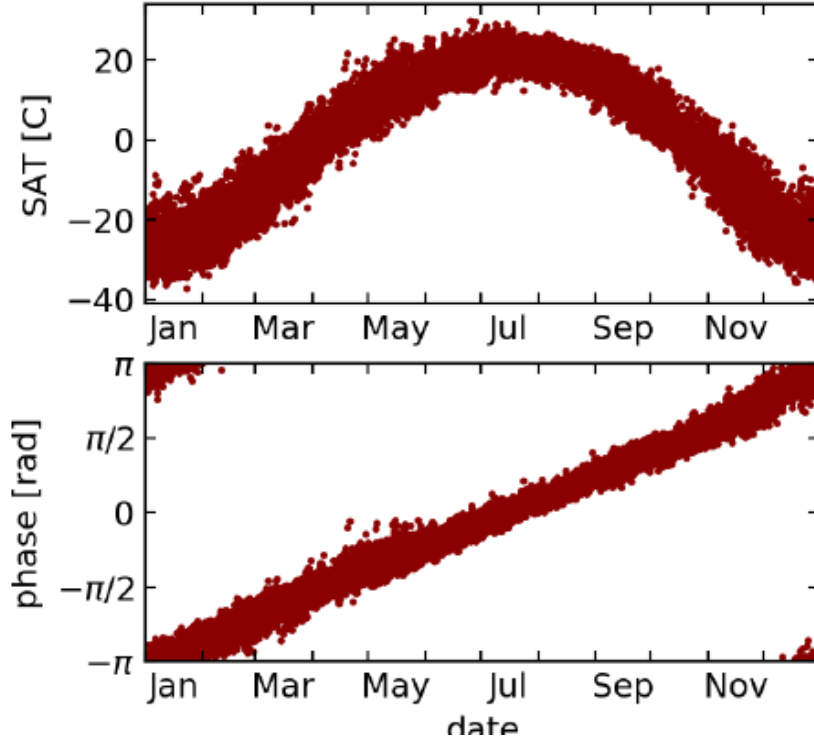
We generate synthetic data. We know the underlying equations and parameters and so can check the validity of our method.





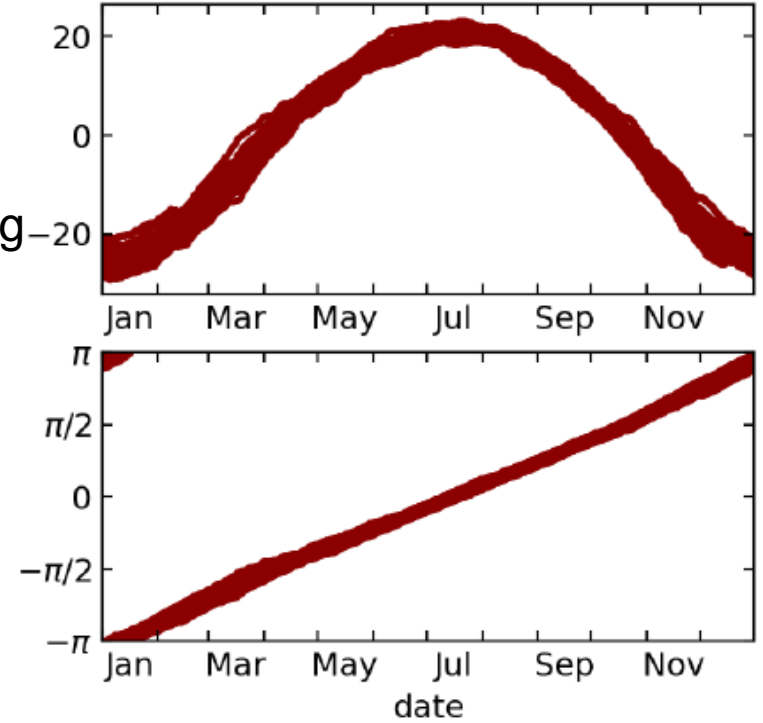
# Phase-date relation: Regular site

Regular site



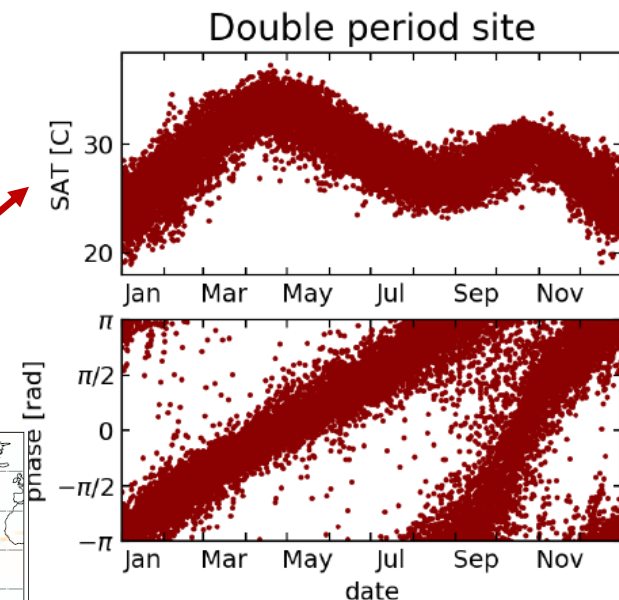
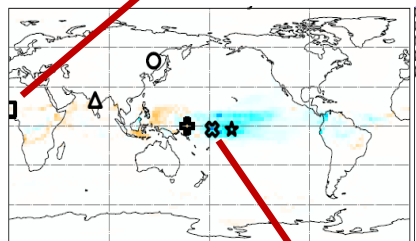
→  
Pre-filtering  
SAT time series in a  
moving window of  
 $\tau=41$  days

Regular site

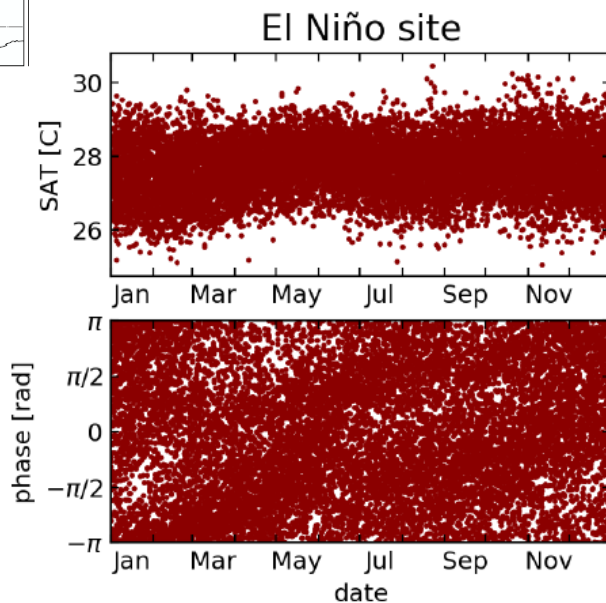
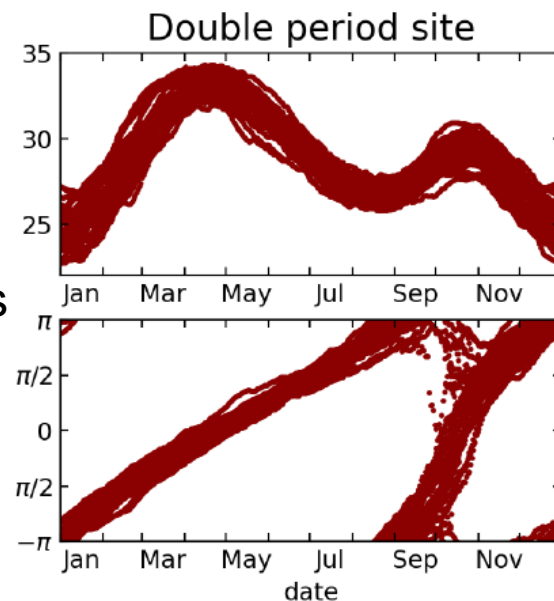




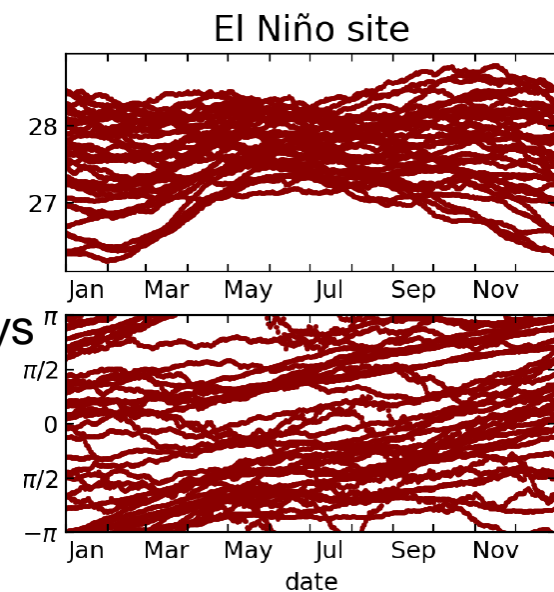
In other sites: with appropriate pre-processing (using a window length  $\tau$  in a “plateau”) we uncover hidden regularity



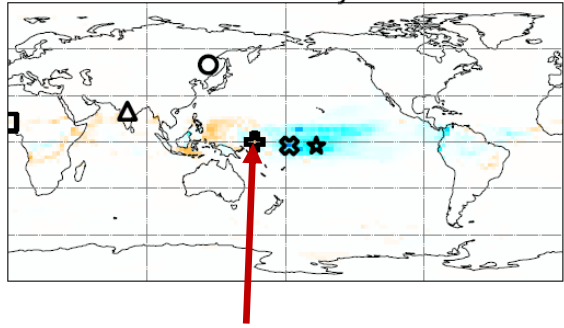
$\tau=41$  days



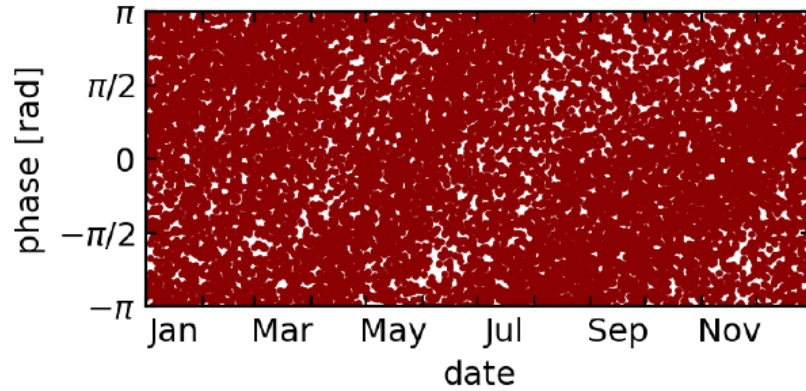
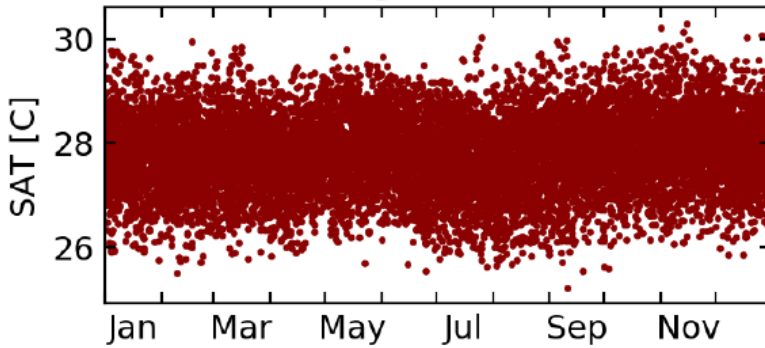
$\tau=101$  days



# But in irregular site no plateau

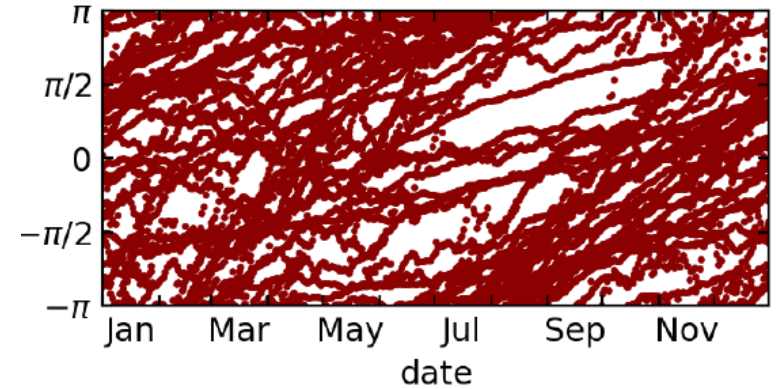
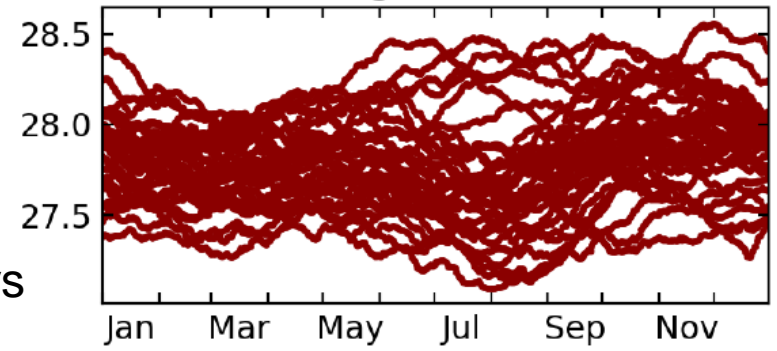


Irregular site



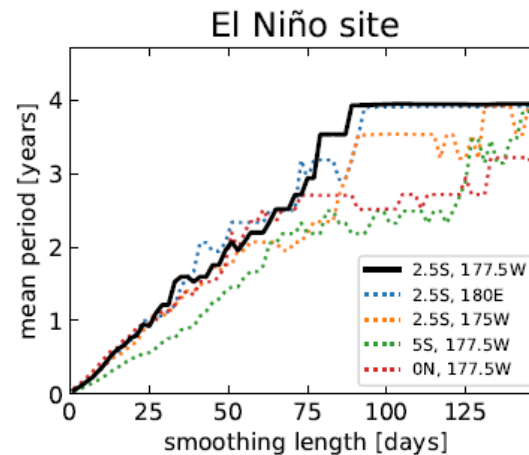
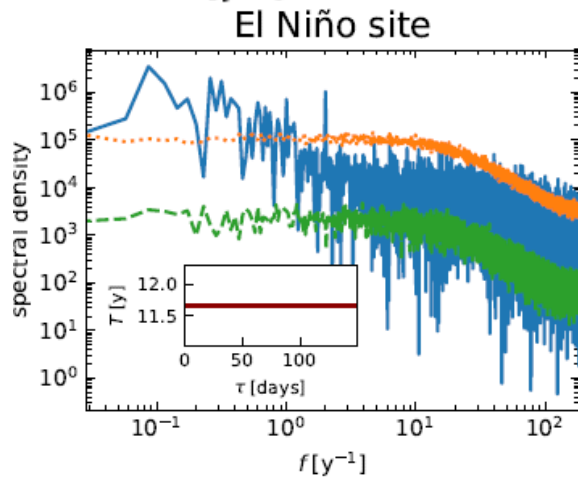
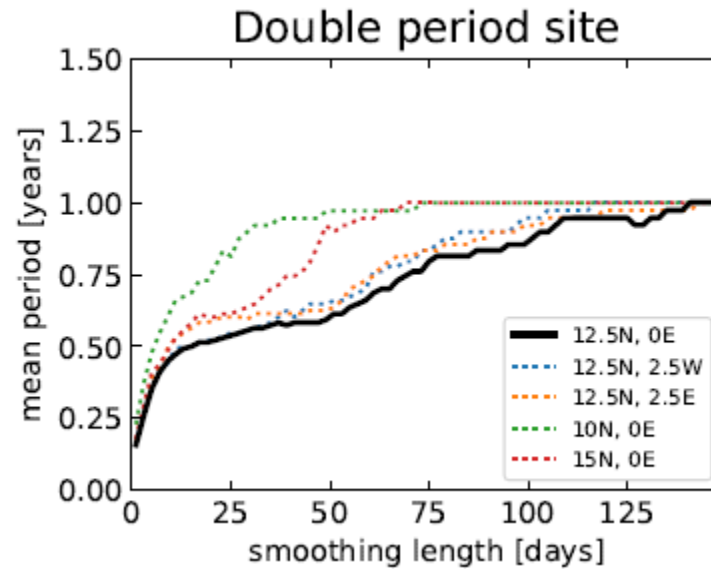
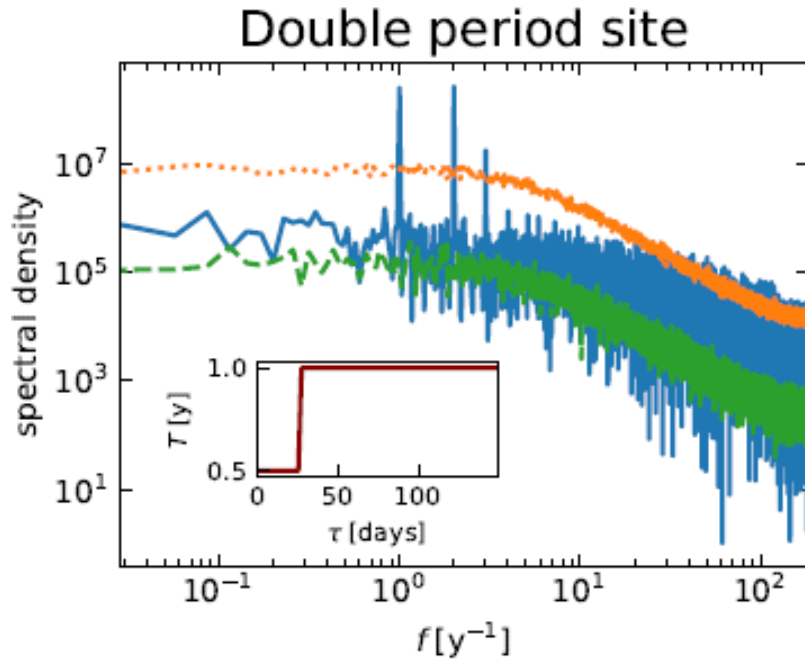
→  
 $\tau=101$  days

Irregular site



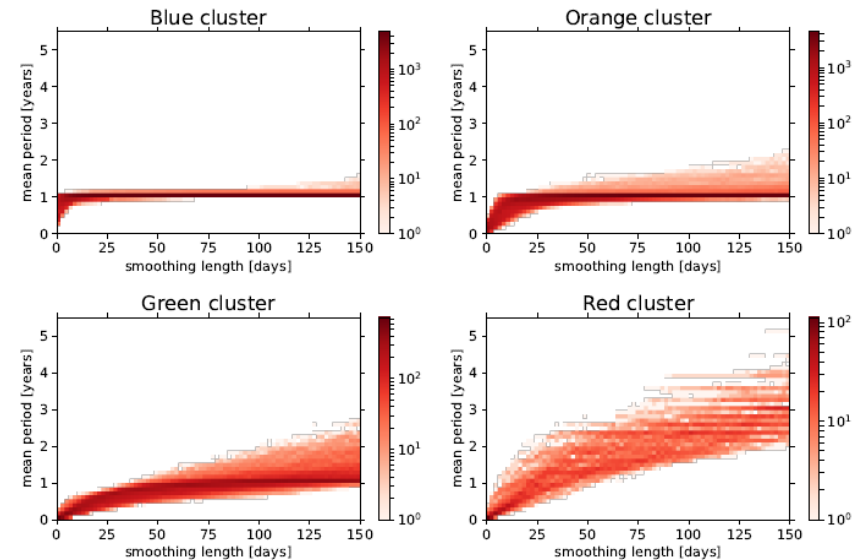
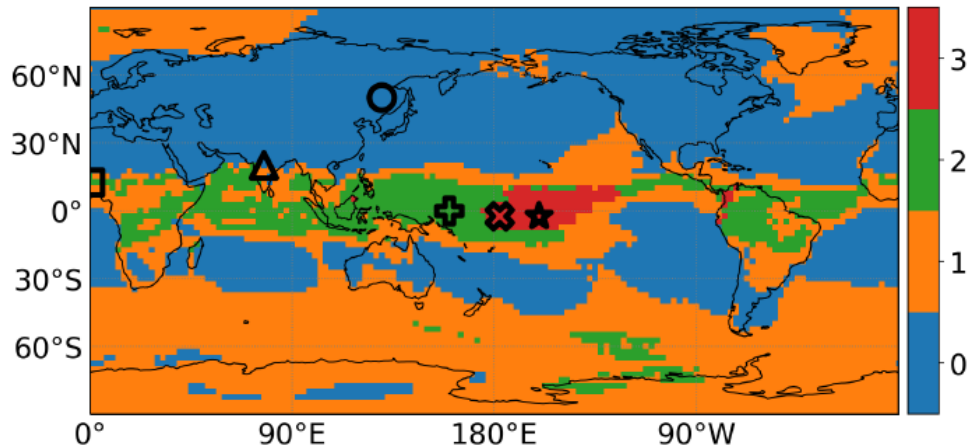


# Can we detect non-trivial underlying periodicities with Fourier analysis?



No. The position of the main peak in the Fourier spectrum does not change, or changes abruptly when increasing the length of the averaging window.

# Classification of SAT dynamics using *k*-means clustering



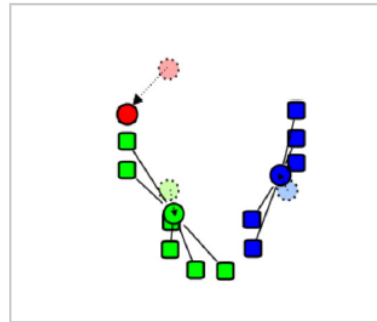
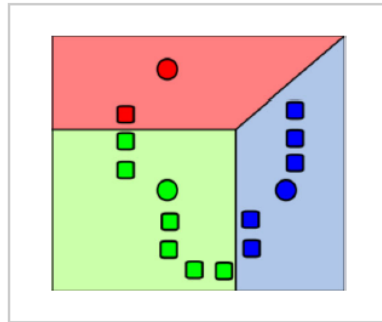
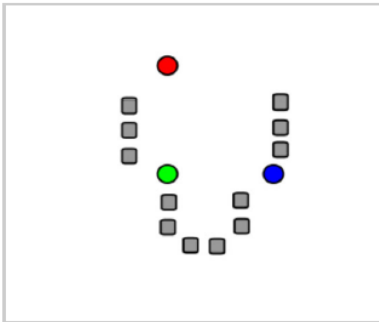
- **Blue cluster:** regions dominated by the seasonal cycle and large temperature variations
- **Orange cluster:** regions with fast dynamics that are dominated by the annual cycle only after smoothing with  $\tau > 20$  days, which may reflect the importance of subseasonal variability
- **Green cluster:** regions of low temperature variability, whose spatial structure is closely related to the mean rainfall pattern
- **Red cluster:** relatively weak annual cycle and influenced by El Niño and the QBO and thus shows slow dynamics

# *k*-means clustering

- *k*-means clustering aims to partition  $n$  observations into  $k$  clusters in which each observation belongs to the cluster with the nearest mean.
- The goal is to minimize the within-cluster sum of squares

$$\sum_{i=1}^k \sum_{\mathbf{x} \in \mathcal{S}_i} \|\mathbf{x} - \boldsymbol{\mu}_i\|^2$$

Demonstration of the standard algorithm



Steps 2 and 3 are repeated until convergence has been reached.

Source: Wikipedia

1.  $k$  initial "means" (in this case  $k=3$ ) are randomly generated within the data domain (shown in color).
2.  $k$  clusters are created by associating every observation with the nearest mean. The partitions here represent the Voronoi diagram generated by the means.
3. The centroid of each of the  $k$  clusters becomes the new mean.

- NP-hard problem, many free and licensed algorithms available.

# And many more TS analysis methods

- Wavelets
- Detrended fluctuation analysis
- Sample entropy, approximate entropy
- Multifractality
- Topological data analysis
- Etc. etc.

# There are **MANY** methods that return “features” that characterize time series.

- How to compare different methods?
- How to identify similar time-series?
- How to identify similar methods?
  - HCTSA: highly comparative time-series analysis
  - From each TS extracts over **7700 features**

B. D. Fulcher, N. S. Jones: *Automatic time-series phenotyping using massive feature extraction*. Cell Systems 5, 527 (2017).



system → time-series dataset → massive feature extraction using hctsa → statistical learning

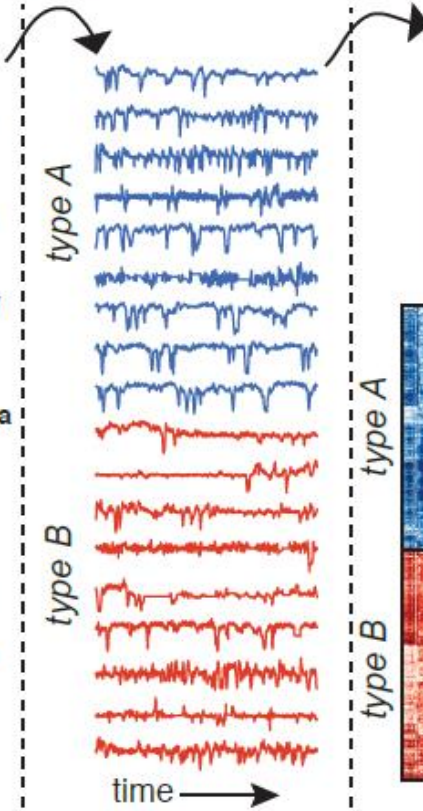
Systems producing time-series phenotype data

What analysis should I use to find differences between phenotypes A and B?

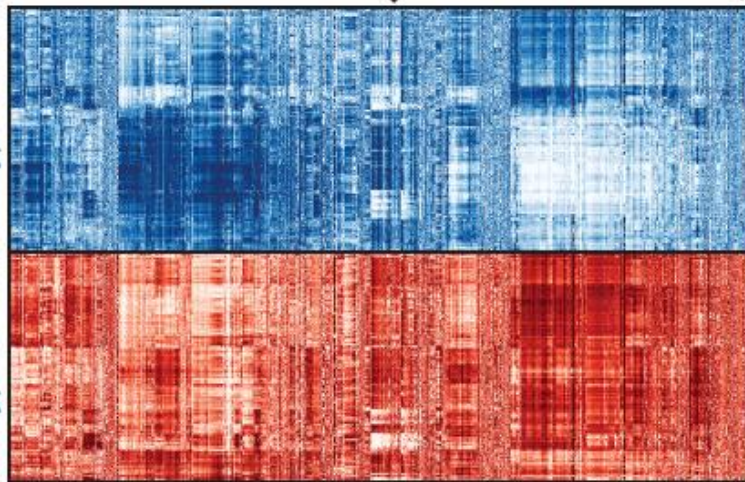
Use hctsa to compare over 7700 time-series features

Extract **interpretable insights** to diagnose disease, deduce gene function, etc.

- model organism movement data
  - genotype A
  - genotype B
- electrophysiological measurements
  - schizophrenia
  - healthy control
- speech recordings
  - Parkinson's Disease
  - healthy control

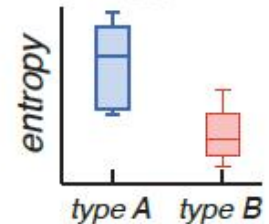


<p><b>data distribution</b></p> <p>median skewness outliers data values</p>	<p><b>correlation properties</b></p> <p>autocorrelation automutual information power spectral properties time-series entropy fluctuation analysis</p>
<p><b>model fitting</b></p> <p>linear autoregressive &amp; nonlinear models model parameters goodness of fit</p>	<p><b>others</b></p> <p>stationarity embedding dimension network properties</p>

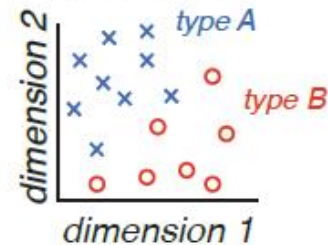


> 7700 features

discriminative features?



low dimensional structure?



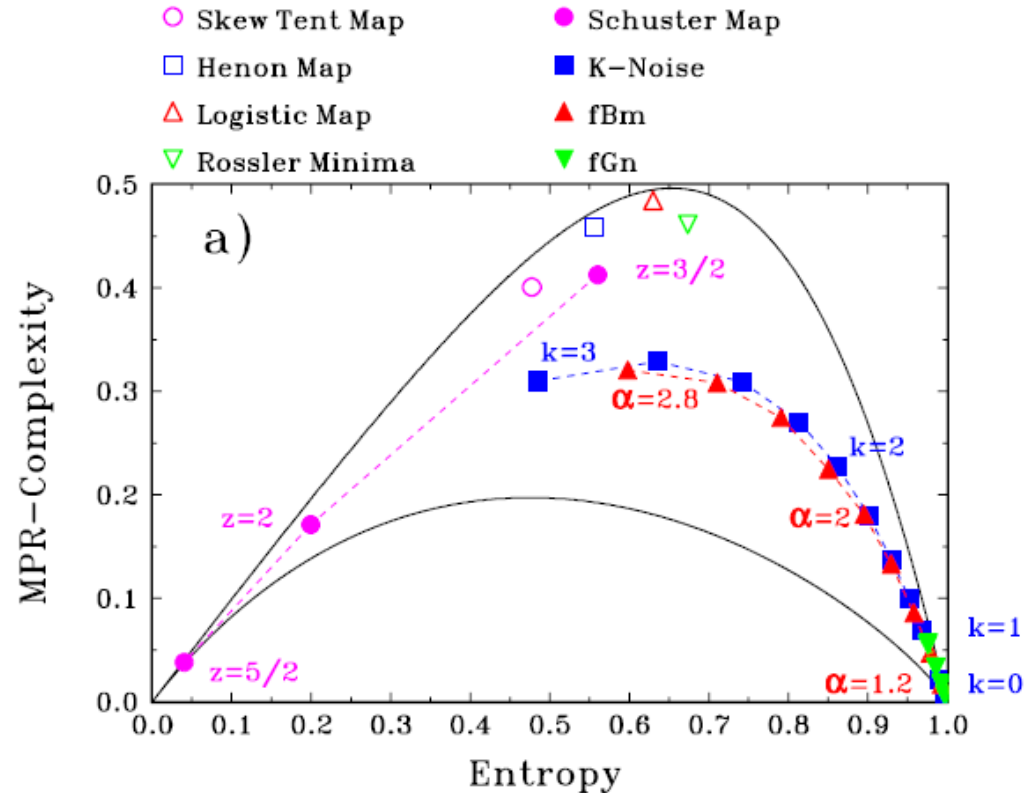
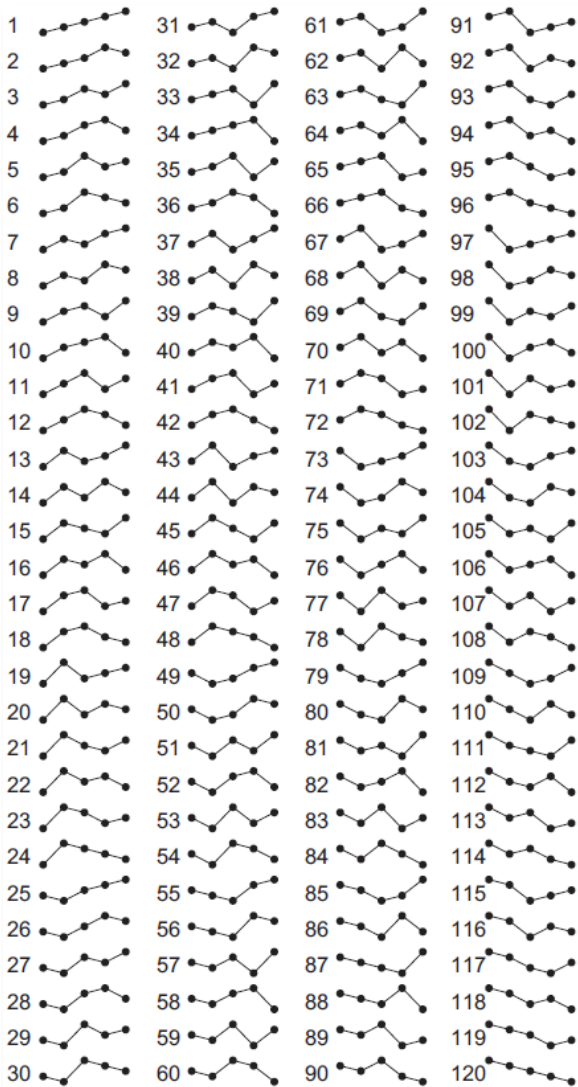
Matlab code: [www.github.com/benfulcher/hctsa](http://www.github.com/benfulcher/hctsa)

**What to do with more than  
7700 features?**

Let's take a step back.

Assuming we have computed 120 ordinal probabilities.

What we do next?

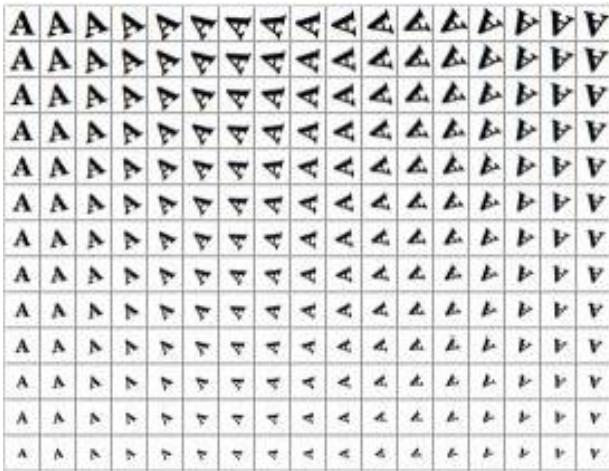


O. A. Rosso et al, Phys. Rev. Lett. 99, 154102 (2007)

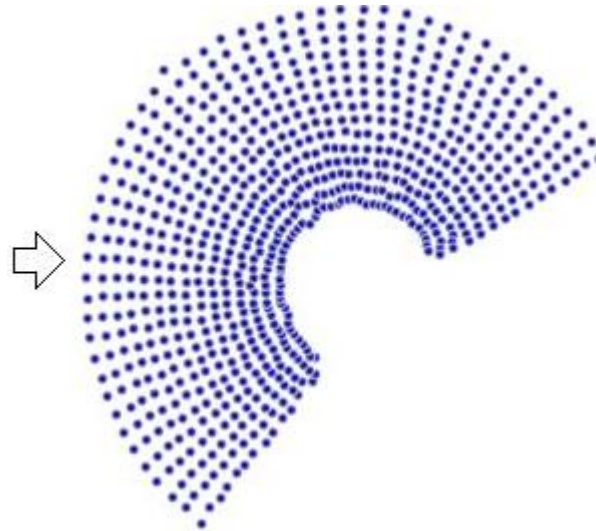


# Nonlinear dimensionality reduction (NLDR)

- Many methods for reducing high-dimensional data to a small number of dimensions.
- Example:



Each image of the letter 'A' has 32x32 pixels = vector of 1024 pixel values



NLDR reduces the data into just two dimensions (rotation and scale)

Source:  
Wikipedia

- A popular method: *ISOMAP*
- A linear method: Principal Component Analysis (PCA)

## Take home messages

- Symbolic analysis, network representation, spatiotemporal representation, etc., 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...”

# References

- R. Hegger, H. Kantz and T. Schreiber, *Nonlinear Time Series Analysis* (book and software [TISEAN](#)).
- [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 \(2016\)](#)
- [J. M. Aparicio-Reinoso et al PRE 94, 032218 \(2016\)](#)
- [F. Arizmendi, Barreiro and Masoller, Sci. Rep. 7, 45676 \(2017\)](#)
- [M. Panozzo et al, Chaos 27, 114315 \(2017\)](#)
- [Zappala, Barreiro and Masoller, Entropy 18, 408 \(2016\)](#)
- [Zappala, Barreiro and Masoller, Earth Syst. Dynam. 9, 383 \(2018\)](#)

<crisrina.masoller@upc.edu>

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