School on Applications of Nonlinear Systems to Socio-Economic Complexity, Oct. 17 – Oct. 22 2022

Nonlinear time series analysis

Cristina Masoller

Departamento de Fisica

Universitat Politecnica de Catalunya

Class 1: From dynamical systems to complex systems

Class 2: Univariate time series analysis

Class 3: Univariate time series analysis

Class 4: Bivariate and multivariate analysis

UNIVERSITAT POLITÈCNICA DE CATALUNYA BARCELONATECH

Campus d'Excel·lència Internacional

ICTP SAIFR International Centre for Theoretical Physics South American Institute for Fundamental Research

cristina.masoller@upc.edu @cristinamasoll1

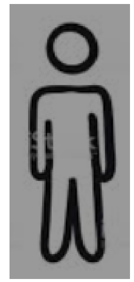
Relevant problems in time series analysis

- Is the signal stationary? Transient chaos?
- Is the signal just noise? Signatures of nonlinearity?
- Can we “reconstruct” the “effective” phase space of a system from observed data?
- Can the signal be predicted? Which is the prediction horizon?
- Is a system approaching a bifurcation point (“tipping point”)?
- Are two (or more) systems (partially) synchronized?
- Are two (or more) systems interdependent? Causal interactions? Coupling delays?
- Can we forecast how failures in one system will propagate to other systems?

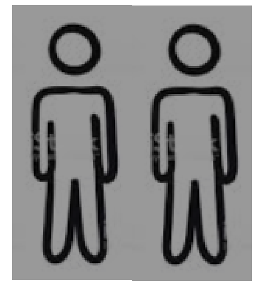
cristina.masoller@upc.edu @cristinamasoll1

Methods of time series analysis are classified as:

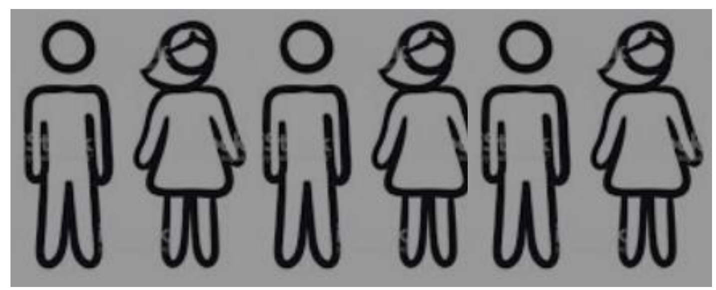
- Univariate analysis



- Bivariate analysis



- Multivariate analysis



cristina.masoller@upc.edu @cristinamasoll1

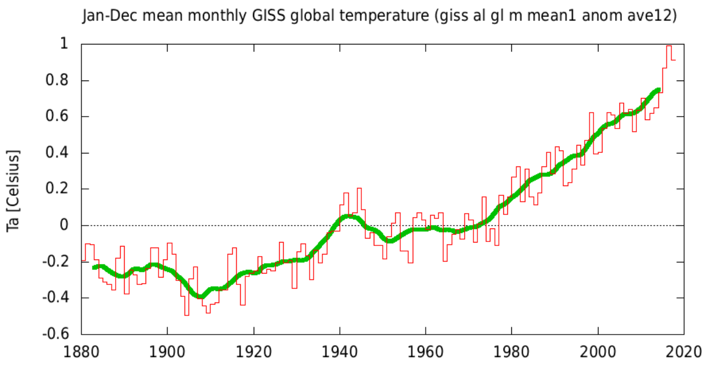
Univariate time series analysis

- Return maps
- Distribution of data values
- Autocorrelation and Fourier analysis
- Stochastic models and surrogates
- Attractor reconstruction, Lyapunov exponents, and fractal dimension
- Symbolic analysis
- Information theory measure: entropy
- Network representation of a time-series
- Spatio-temporal representation of a time-series

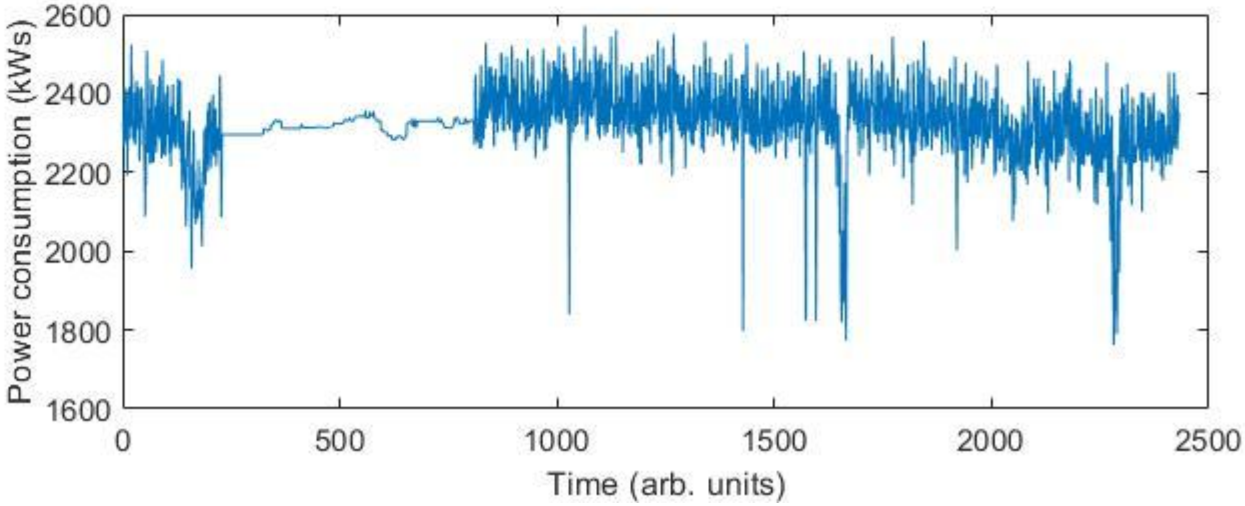
cristina.masoller@upc.edu @cristinamasoll1

To begin with the analysis of a time series

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



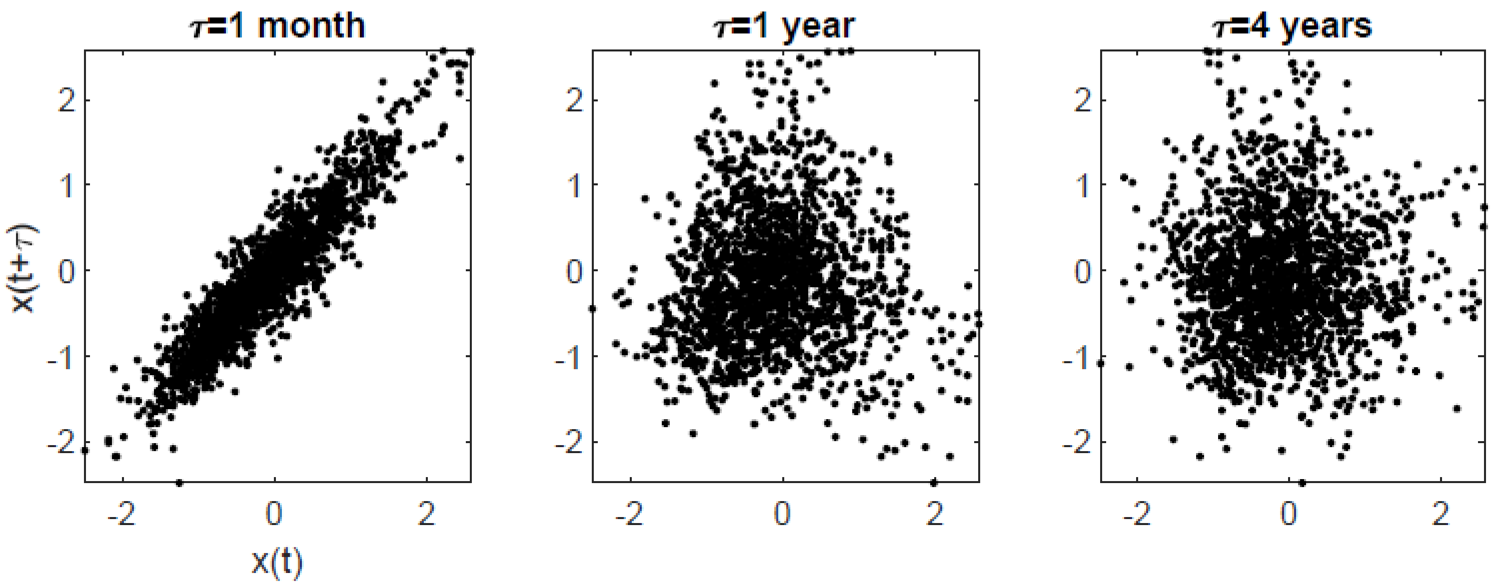
- First step: **plot the data**.



- Next: examine simple properties.

cristina.masoller@upc.edu @cristinamasoll1

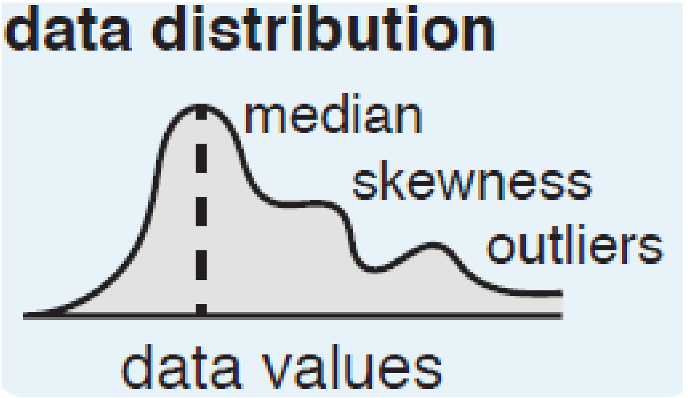
Return Map: plot of x_i vs. $x_{i+\tau}$



cristina.masoller@upc.edu @cristinamasoll1

Analyze the distribution of data values

- Plot the distribution of data points (histogram).
- Next: examine simple properties



Mean (expected value of *X*, E[X])

$$\mu = \langle x(t) \rangle$$

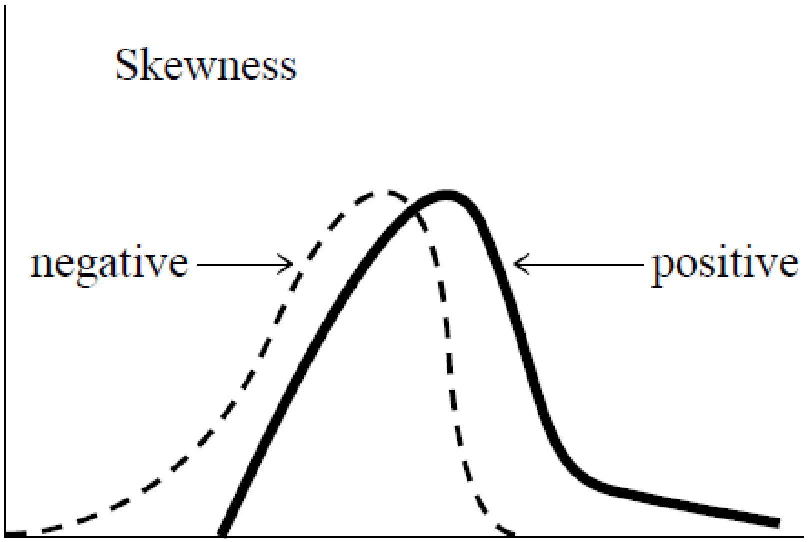
Variance: $\sigma^2 = Var\,(X) = E[(X-\mu)^2]$

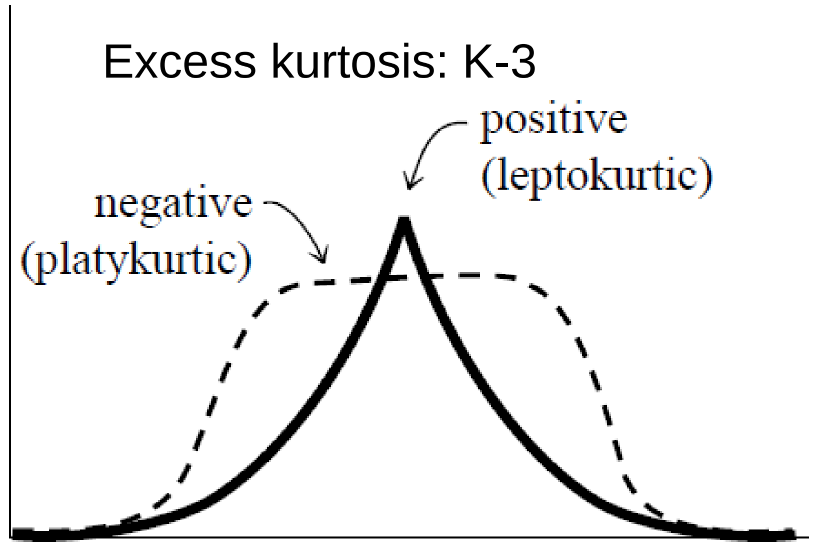
Skewness: $S = E[Z^3]$ $\quad Z = \frac{X-\mu}{\sigma}$

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

Coefficient of variation: $C_v = \sigma / |\mu|$

cristina.masoller@upc.edu @cristinamasoll1





- **K<3:** the distribution produces <u>fewer and less extreme</u> “outliers” than the normal distribution. Example: the uniform distribution.
- **K=3:** Normal Gaussian
- **K>3:** the tail approaches zero more slowly than a Gaussian, and therefore produces more outliers. Example: Laplace distribution.

Press WH et al. Numerical recipes: the art of scientific computing (Cambridge University Press)

cristina.masoller@upc.edu @cristinamasoll1

Example: "fat tail" in the distribution of financial data

Scaling behaviour in the dynamics of an economic index

Rosario N. Mantegna & H. Eugene Stanley

Center for Polymer Studies and Department of Physics,
Boston University, Boston, Massachusetts 02215, USA

THE large-scale dynamical properties of some physical systems depend on the dynamical evolution of a large number of nonlinearly coupled subsystems. Examples include systems that exhibit self-organized criticality[1] and turbulence[2,3]. Such systems tend to exhibit spatial and temporal scaling behaviour—power-law behaviour of a particular observable. Scaling is found in a wide range of systems, from geophysical[4] to biological[5]. Here we explore the possibility that scaling phenomena occur in economic systems—especially when the economic system is one subject to precise rules,

NATURE · VOL 376 · 6 JULY 1995

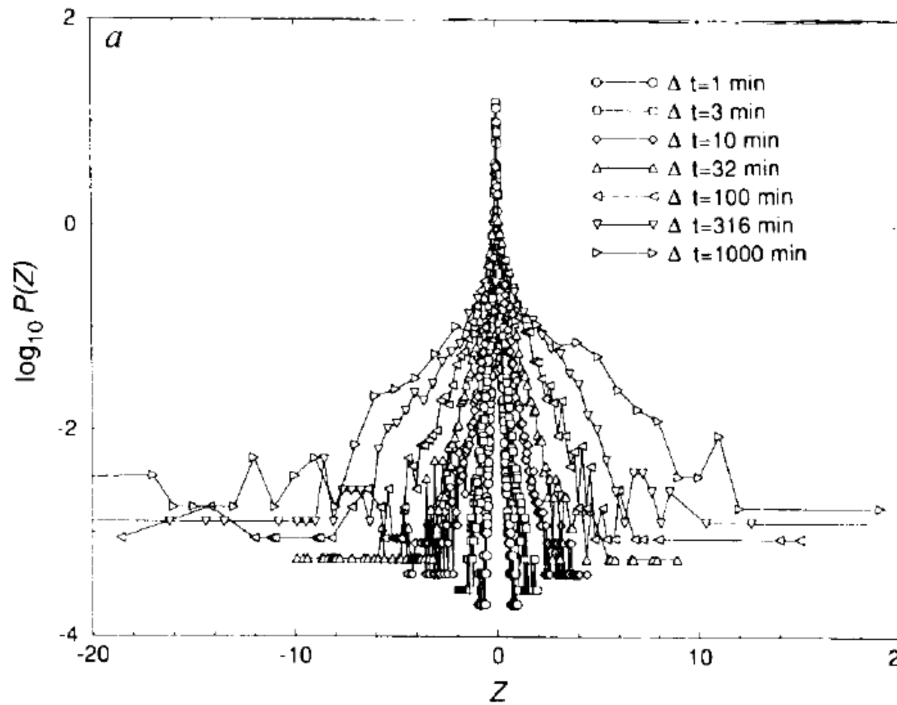
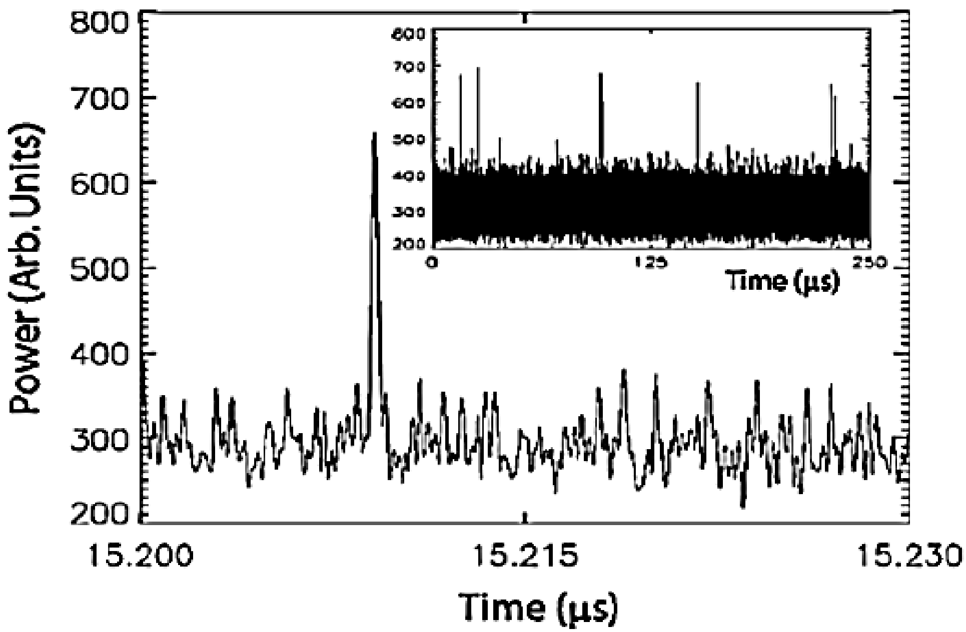


FIG. 1 *a*, Probability distributions *P*(*Z*) of the S&P 500 index variations *Z*(*t*) observed at time intervals Δ*t*, which range from 1 to 1,000 min. By

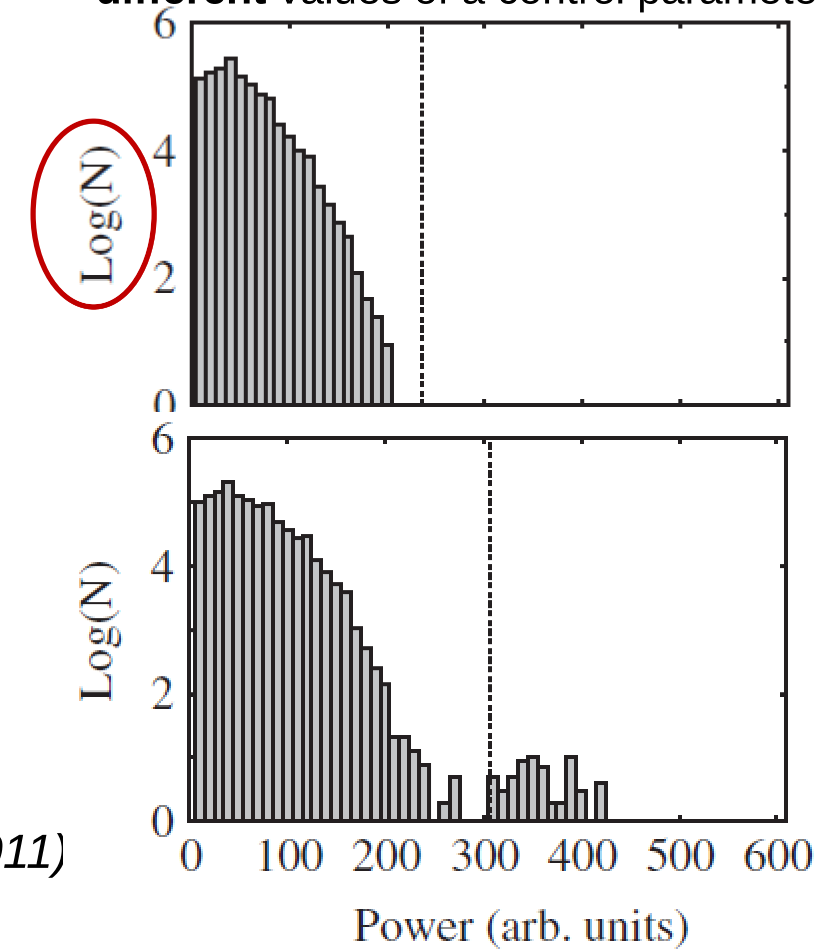
cristina.masoller@upc.edu @cristinamasoll1

Systems displaying long tailed distributions? Controlled experiments are often difficult.

Example: the intensity emitted by a chaotic laser



Intensity histograms for **slightly different** values of a control parameter



Optical rogue wave if the pulse height is: $I > \langle I \rangle + 8\sigma$

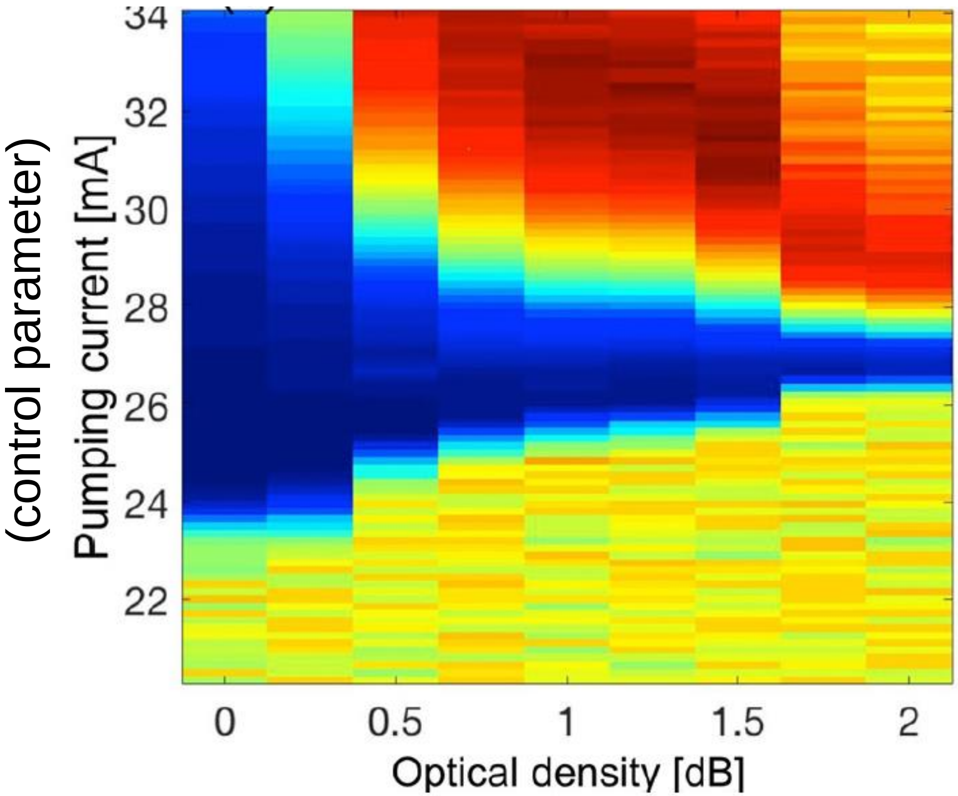
Bonatto et al. Phys. Rev. Lett. 107, 053901 (2011)

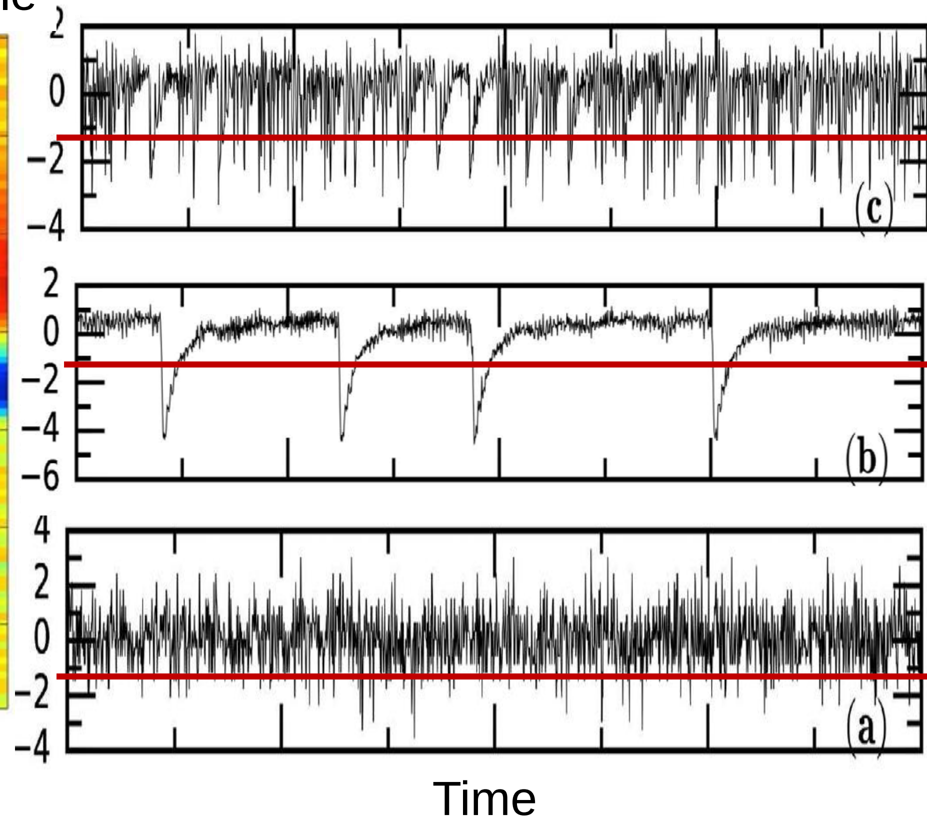
cristina.masoller@upc.edu @cristinamasoll1

Application: counting the number of extreme values allows to distinguish different dynamical regimes

$$Z = \frac{X - \mu}{\sigma}$$

Number of events (below -1.5σ) in log scale

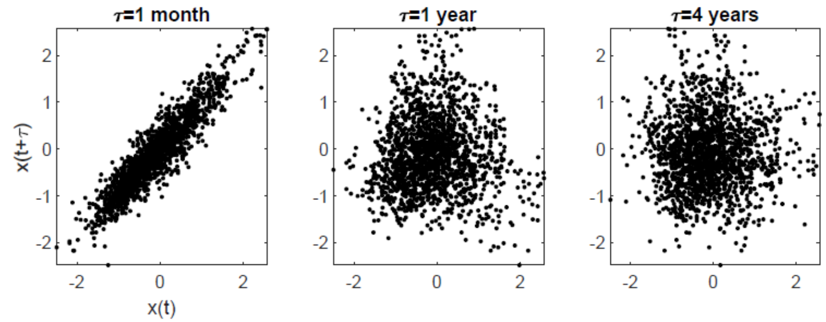




Panozzo et al, Chaos 27, 114315 (2017)

cristina.masoller@upc.edu @cristinamasoll1

3. Autocorrelation function (ACF)



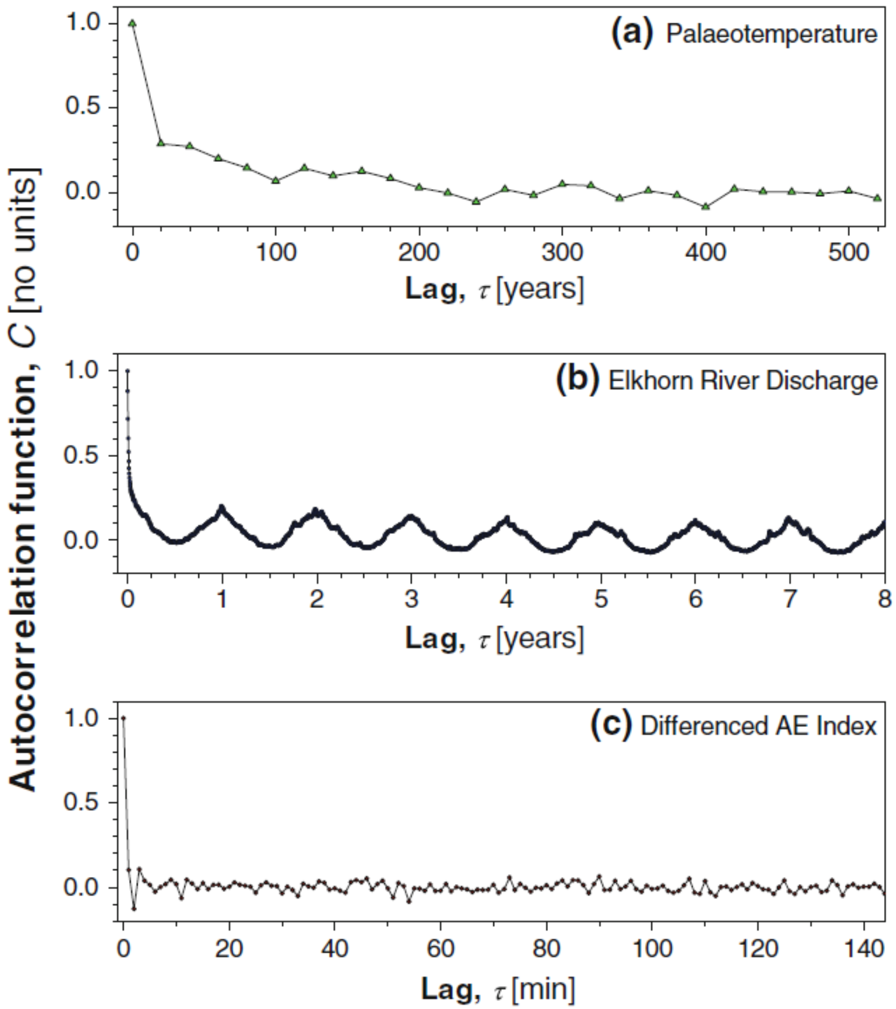
- The return map allows us to see if x(t) and x(t+τ) are "*correlated*".
- But how to **quantify**?

$$C(\tau)=\frac{\langle [x(t)-\mu][x(t+\tau)-\mu]\rangle}{\sigma^2}$$

- C(0)=1
- For a **stationary process** (μ, σ constant in time): C(τ) = C(-τ)
- **C(τ)=0** indicates that x(t) and x(t+τ) are **uncorrelated**.
- **C(τ)>0** indicates **persistence**: large values tend to follow large ones, and small values tend to follow small ones, on average (more of the time than if the time series were uncorrelated).
- **C(τ)<0** indicates **anti-persistence**: large values tend to follow small ones and small values tend to follow large ones.

cristina.masoller@upc.edu @cristinamasoll1

Examples of autocorrelation functions



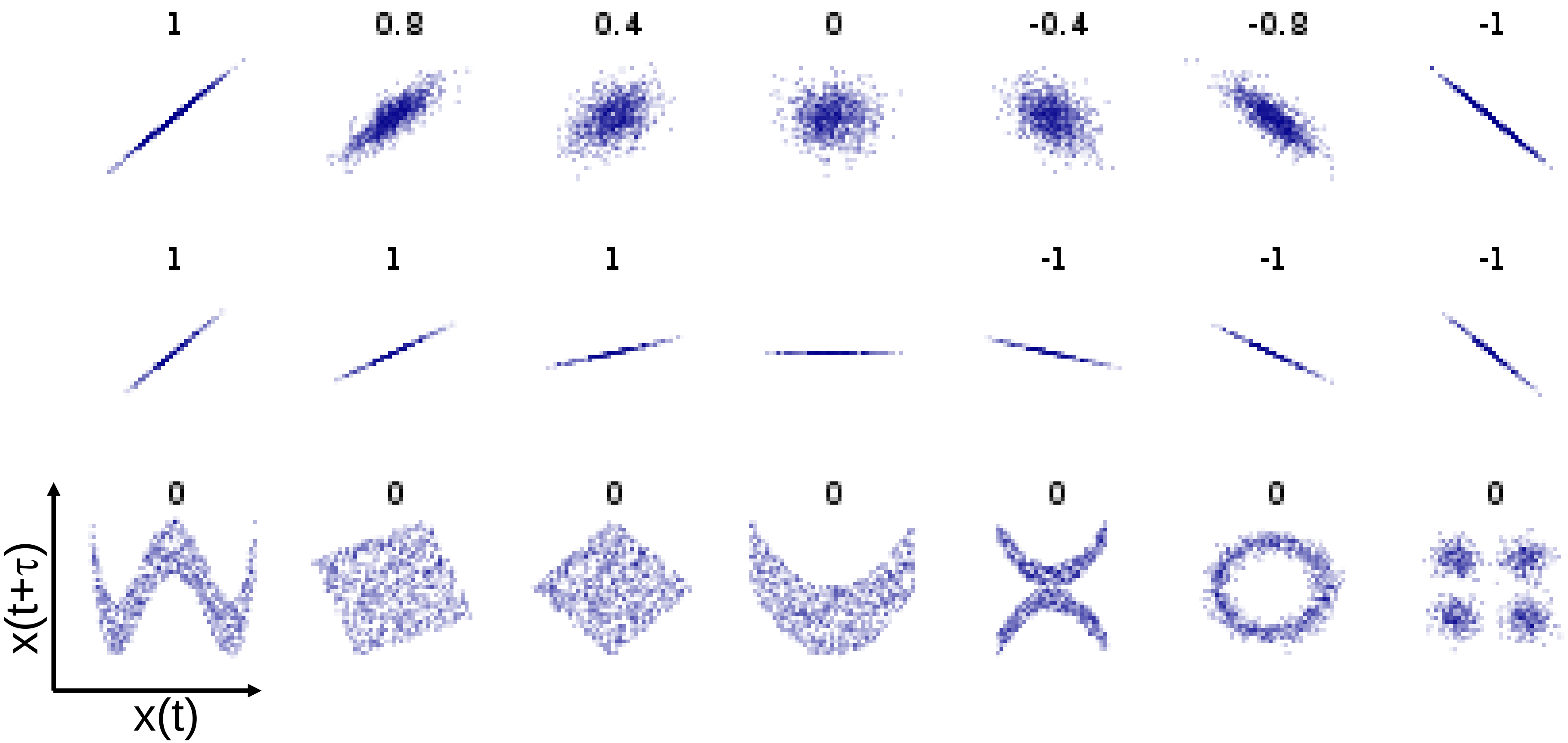
Slow decay: **long-range** correlations.

Rapid decay: **short-range** correlations.

A. Witt and B. D. Malamud, Surv. Geophys. 34, 541 (2013).

cristina.masoller@upc.edu @cristinamasoll1

Problem with the ACF: only detects linear correlations



Source: Wikipedia

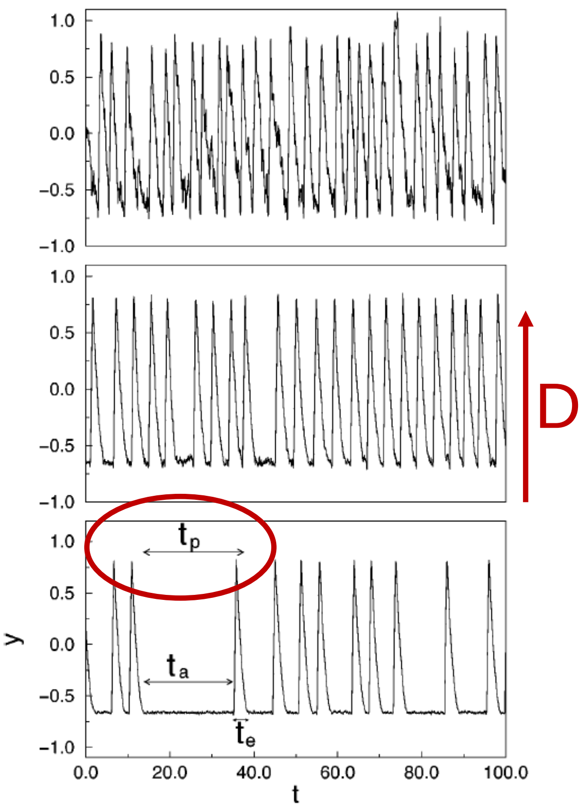
cristina.masoller@upc.edu @cristinamasoll1

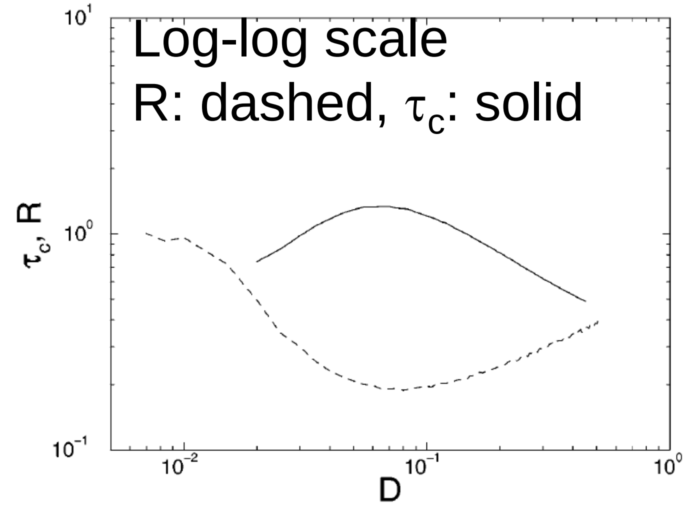
Application of variance and autocorrelation: quantification of coherence resonance

Fitz Hugh–Nagumo model (excitable system)

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

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





Coefficient of variation <u>of the distribution of inter-spike-intervals</u>

$$R_p = \frac{\sqrt{\mathrm{Var}\,(t_p)}}{\langle t_p \rangle}$$

Correlation time $\tau_c = \int_0^\infty C^2(t)\,dt$

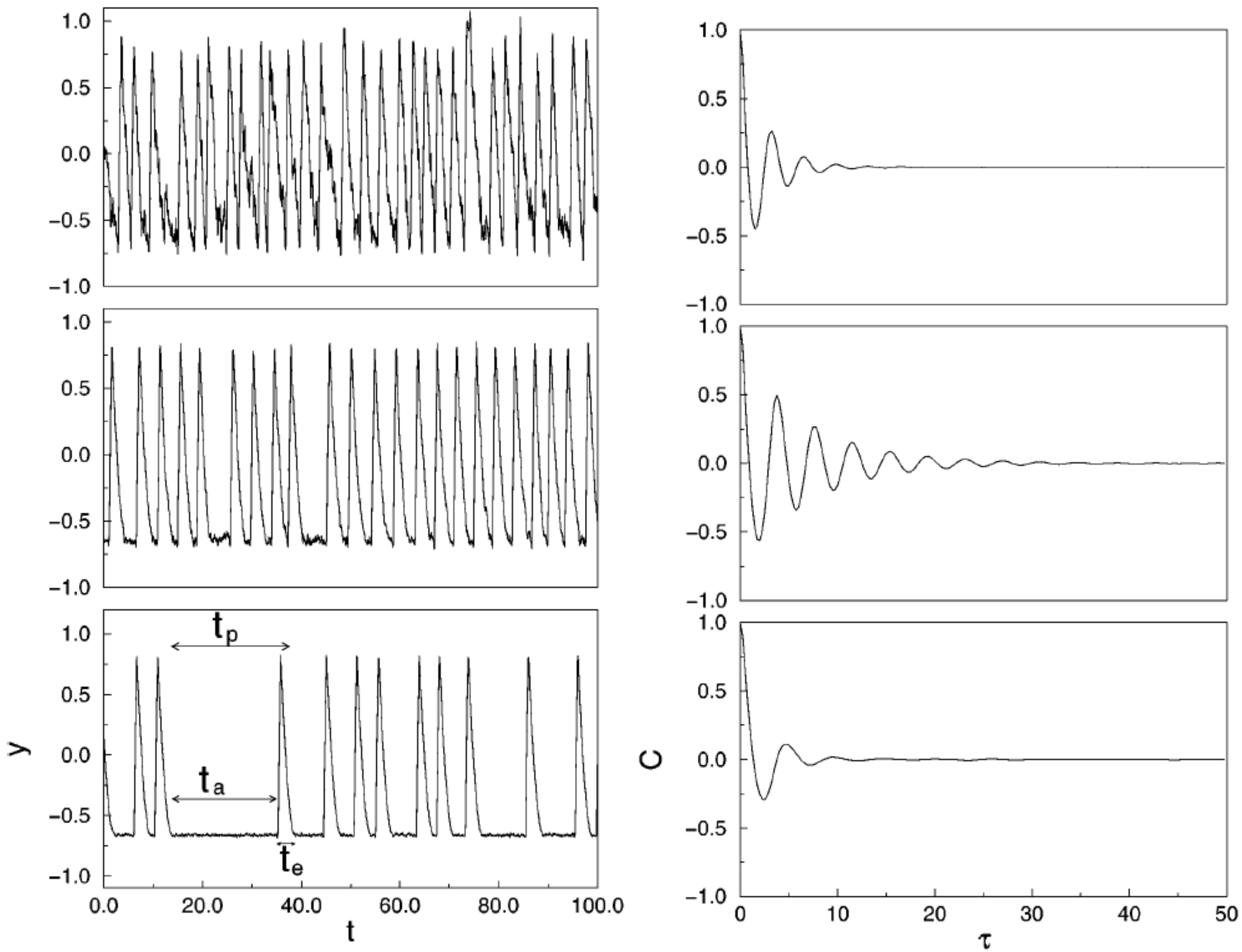
ACF $C(\tau) = \frac{\langle \tilde{y}(t)\tilde{y}(t+\tau)\rangle}{\langle \tilde{y}^2 \rangle}, \qquad \tilde{y} = y - \langle y \rangle.$

Pikovsky and Kurths PRL 1997

cristina.masoller@upc.edu @cristinamasoll1

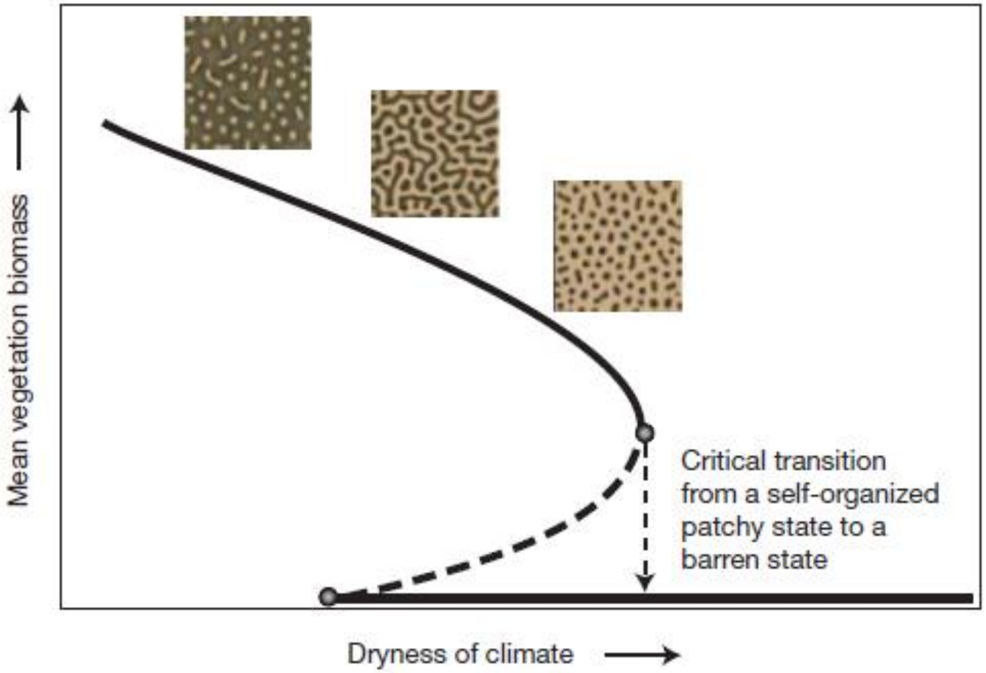
How does the autocorrelation function look like?

$$C(\tau) = \frac{\langle \tilde{y}(t)\tilde{y}(t+\tau)\rangle}{\langle \tilde{y}^2\rangle}, \qquad \tilde{y} = y - \langle y \rangle.$$

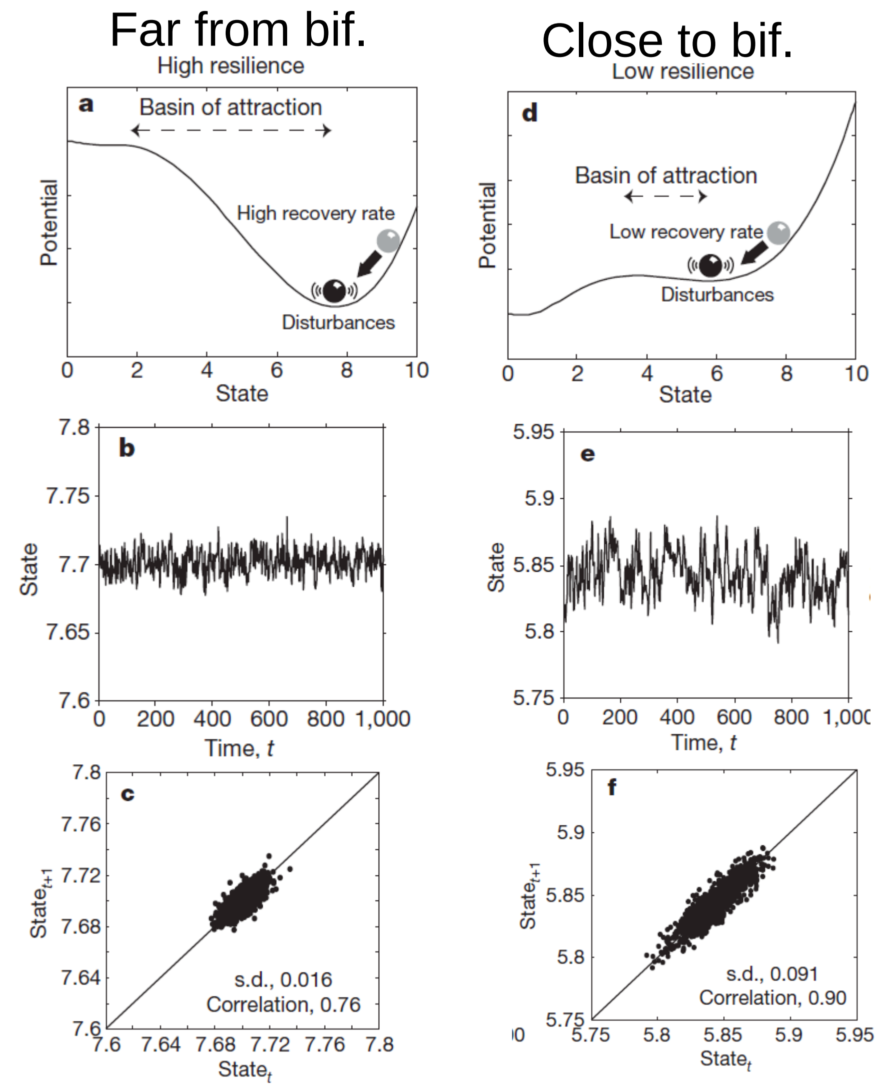


cristina.masoller@upc.edu @cristinamasoll1

Another application of variance and autocorrelation: "early warning signals" of critical transitions



M. Scheffer et al., Nature 461, 53 (2009)



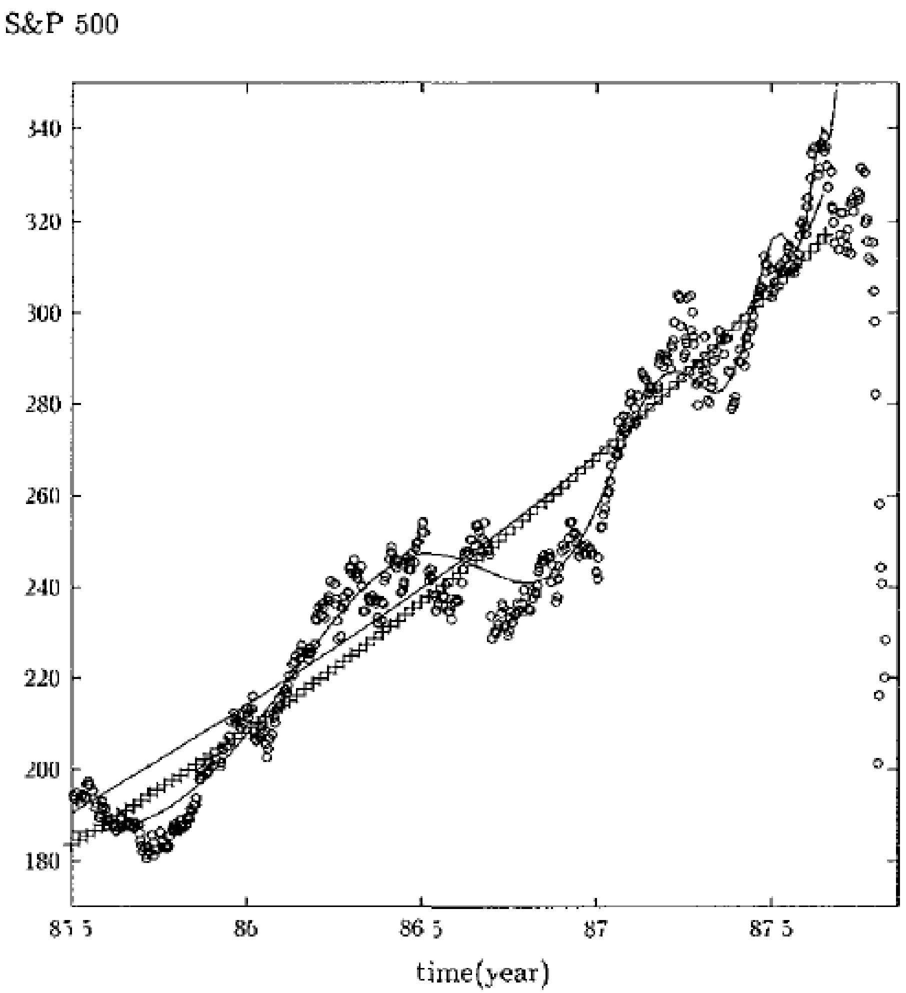
⇒ An increase in variance and autocorrelation can indicate an approaching "tipping point" or critical transition.

cristina.masoller@upc.edu @cristinamasoll1

The crash of October 1987

Evolution of the New York stock exchange index S&P500 from July 1985 to the end of October 1987 (557 trading days).

One can observe well-defined oscillations before the bubble ends in the crash.



Sornette et al. J. de Physique I 6, 167 (1996).
Sornette, Physics Reports 378, 1 (2003).

cristina.masoller@upc.edu @cristinamasoll1

4. Discrete Fourier Transform

- $x = \{x_0, x_1, \ldots x_{N-1}\}$ is described as a superposition of waves.

$$x_n = \frac{1}{N} \sum_{k=0}^{N-1} X_k \cdot e^{i \frac{2\pi}{N} kn} \qquad X_k = \sum_{n=0}^{N-1} x_n \cdot e^{-\frac{i2\pi}{N} kn}$$

- The DFT(x) is the set of complex numbers $X=\{X_0, X_1, \ldots X_{N-1}\}$.
- The frequencies are $f_k = k/(N\Delta)$ with Δ = sampling time (time interval between x_i and x_{i+1}) and N = # of data points.
- Important property: the Fourier transform is linear.

  If X=DFT(x) and Y=DFT(y) $\Rightarrow$ aX_K+bY_K = DFT($ax + by$).
- Important property: If x is a real signal: $X_K = (X_{N-k})^*$
- The Fast Fourier Transform (FFT) algorithm applied to x returns the DFT, i.e., the set of complex numbers $X = \{X_0, X_1, \ldots X_{N-1}\}$.

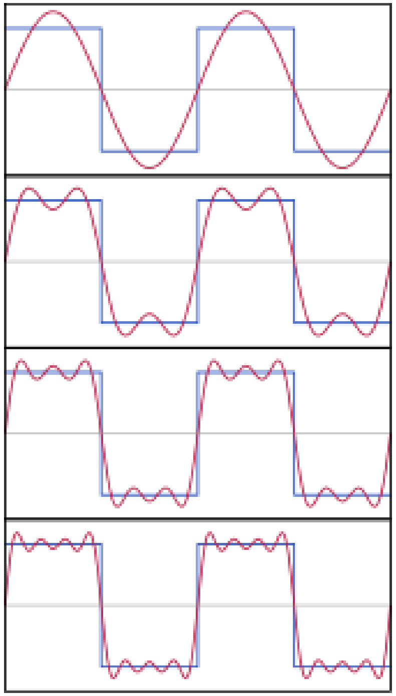
cristina.masoller@upc.edu @cristinamasoll1

Examples

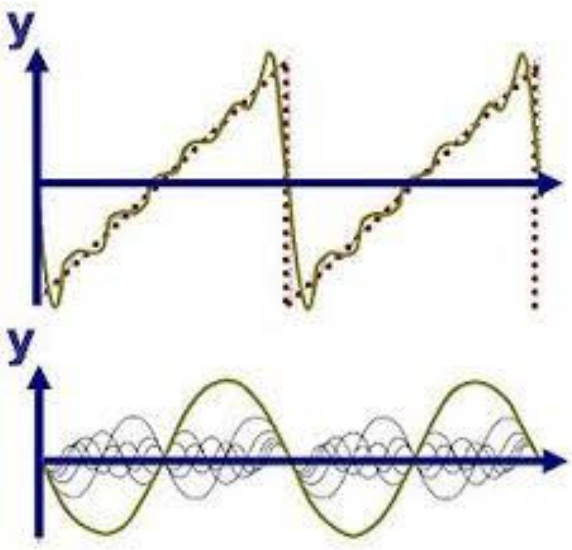
$$x_n = \frac{1}{N}\sum_{k=0}^{N-1} X_k \cdot e^{i\frac{2\pi}{N}kn}$$

X=DFT(x)

The sum of the first four terms of the Fourier series (with specific phases) gives almost a square wave.

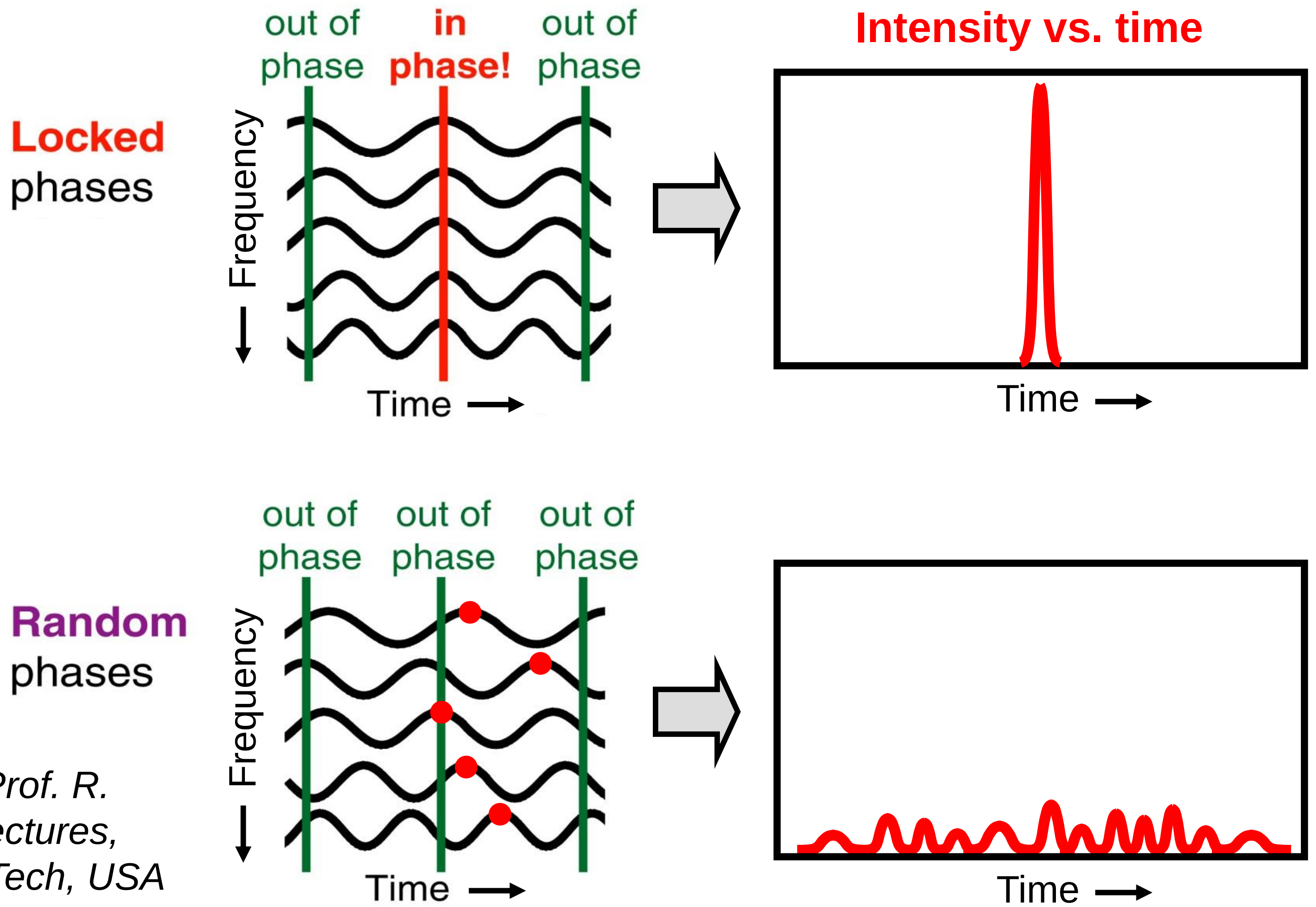


The sum of many terms of the Fourier series (with specific phases) gives a triangular signal.



cristina.masoller@upc.edu @cristinamasoll1

The DFT are complex numbers that contain information of the amplitude and phases of the waves in the Fourier series



Source: Prof. R. Trebino lectures, Georgia Tech, USA

cristina.masoller@upc.edu @cristinamasoll1

Power spectral density (PSD)

The PSD is the set of real numbers that give the “strength” of each frequency component:

$$\mathrm{PSD}(x) = \{|X_0|^2, |X_1|^2, \ldots\}$$

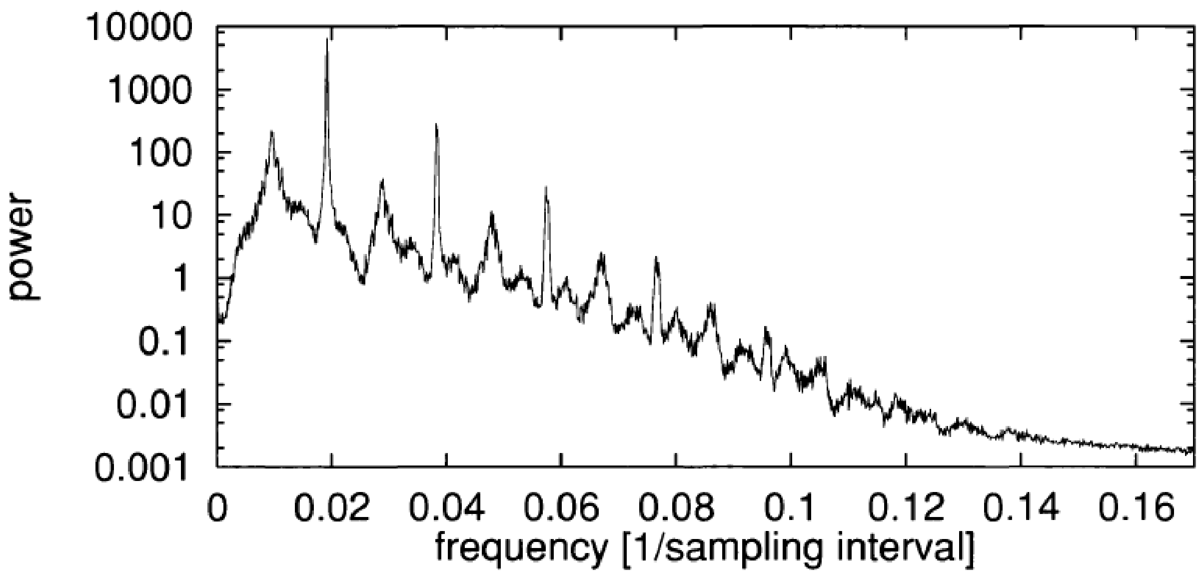


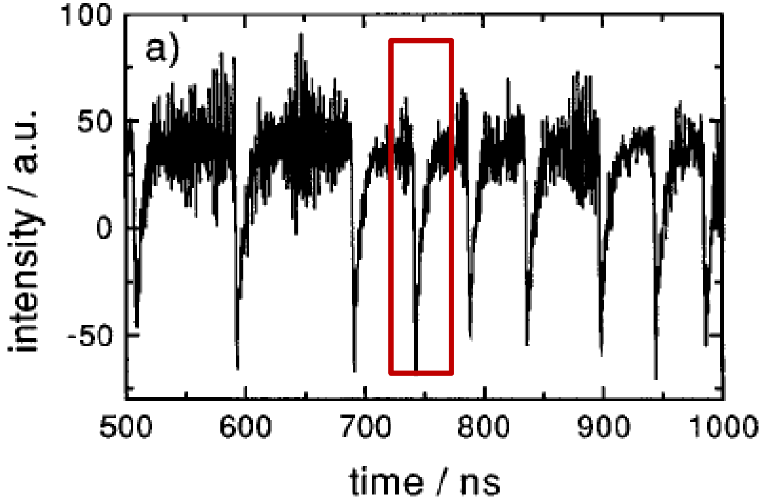
Figure 2.2 Lower part of the spectrum of a time series from the nonlinear feedback laser (Appendix B.10). The sharp peaks form a series of harmonics; the broad peak at low frequencies is a sub-harmonic.

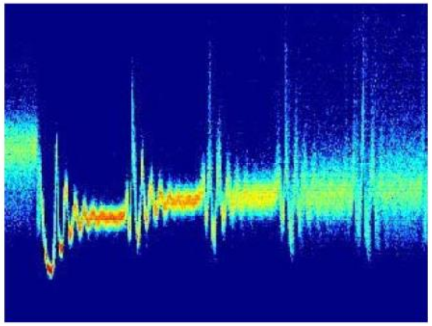
H. Kantz and T. Schreiber, Nonlinear Time Series Analysis, Cambridge University Press (2004)

cristina.masoller@upc.edu @cristinamasoll1

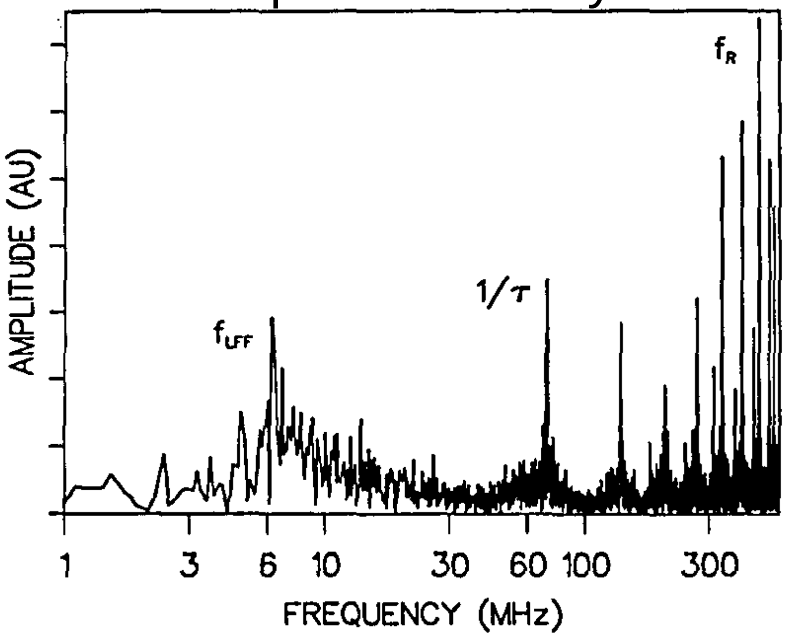
Example: Fourier analysis of a time series with oscillations with different time-scales

x(t) vs. time





Power spectral density

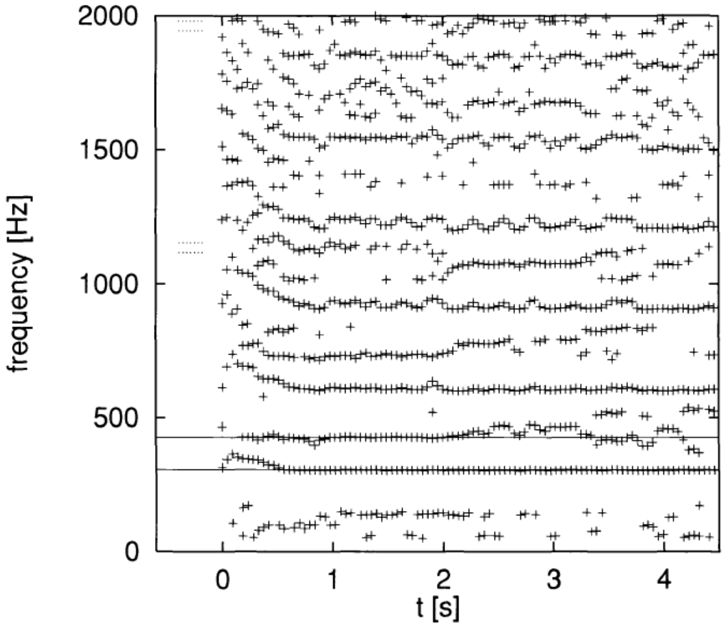


M. Sciamanna (PhD Thesis 2004).
Langley et al, Opt. Lett. 19, 2137 (1994).

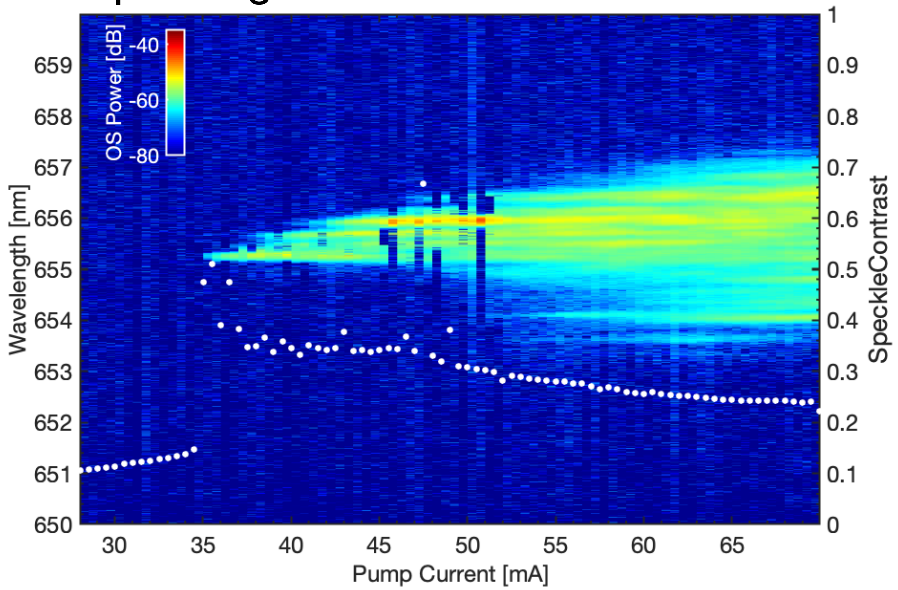
cristina.masoller@upc.edu @cristinamasoll1

Spectrogram: variation of the PSD with time (or with a control parameter)

Spectrogram of voice data



Spectrogram of a chaotic laser



M. Duque, J. Tiana (UPC 2022)

The spectrogram resolves both time and frequency information, but with **limited resolution Δt, Δf** (when, in time, a given frequency will occur? which is the precise value of the frequency at a given time?).

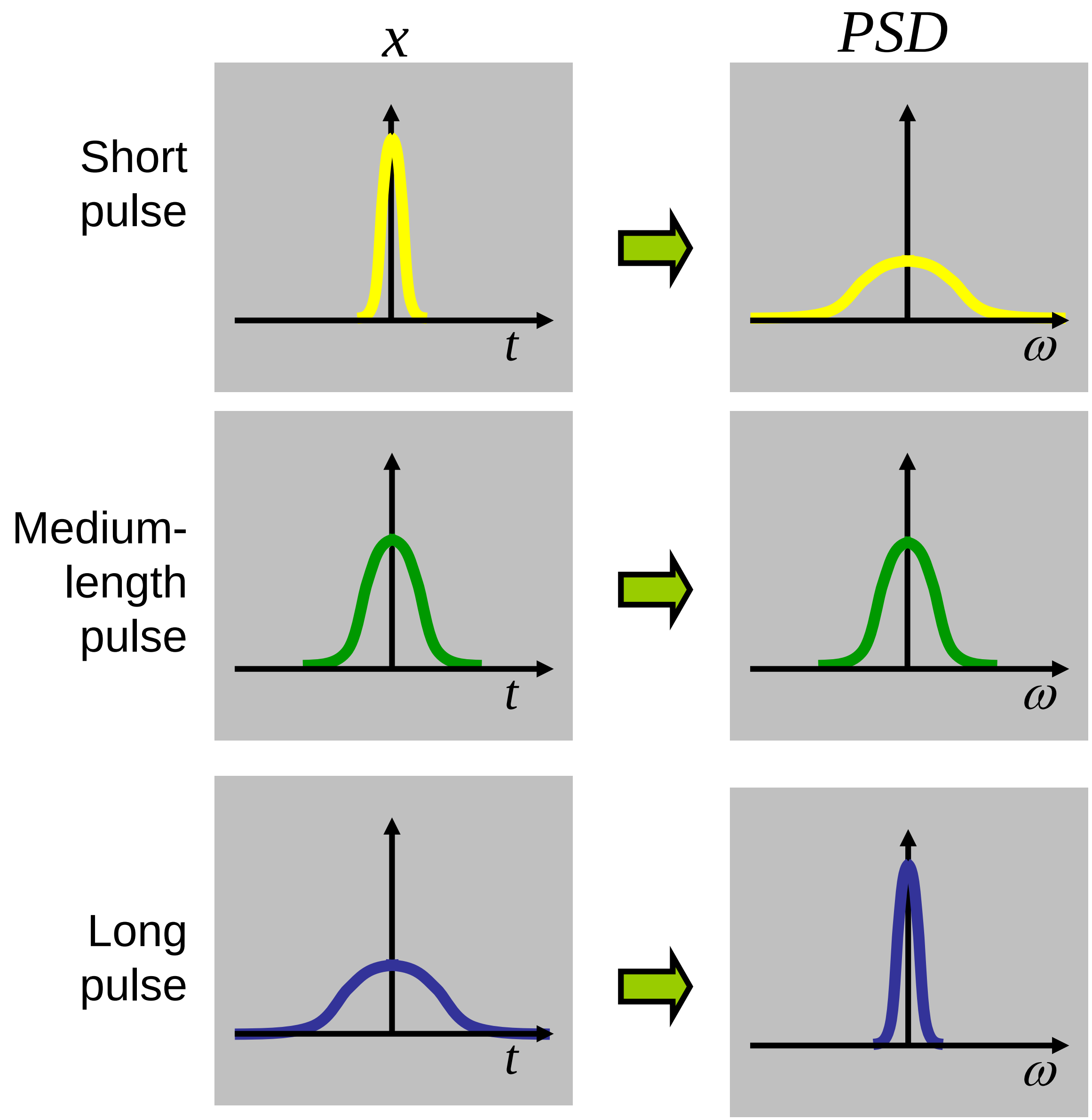
H. Kantz and T. Schreiber, Nonlinear Time Series Analysis, Cambridge University Press (2004).

cristina.masoller@upc.edu @cristinamasoll1

The shorter a pulse, the broader the power spectrum

$$\Delta t \bullet \Delta\nu \gtrsim 0.3$$

(This is the essence of the Uncertainty Principle of Quantum Mechanics)



Source: *Prof. R. Trebino lectures, Georgia Tech, USA*

cristina.masoller@upc.edu @cristinamasoll1

Power spectral density and autocorrelation function

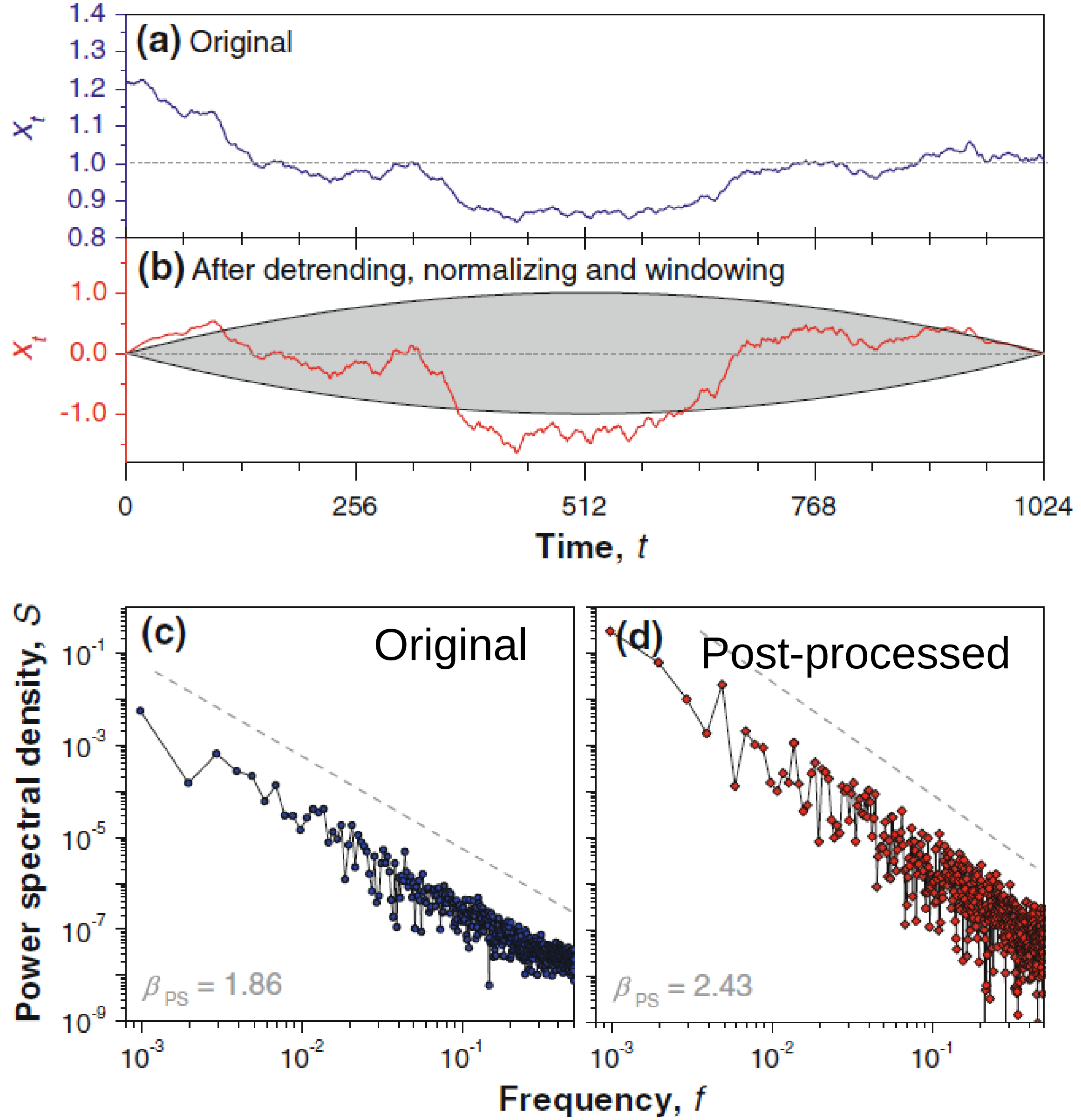
- *Wiener-Khinchin theorem*: if x is stationary, PSD(x) is the Fourier transform of the autocorrelation function, ACF C(τ).

- Long-range temporal correlations: power-law decay of the PSD.

- The Fourier transform is designed for ‘circular’ time series (i.e. the last and first values in the time series ‘follow’ one another). When $|x_N - x_1|$ is large (non-stationary time series) a more precise estimation of the PSD is obtained after “detrending” and “windowing”.

cristina.masoller@upc.edu @cristinamasoll1

Example



Witt and Malamoud, Surv. Geophys. 34, 541 (2013).

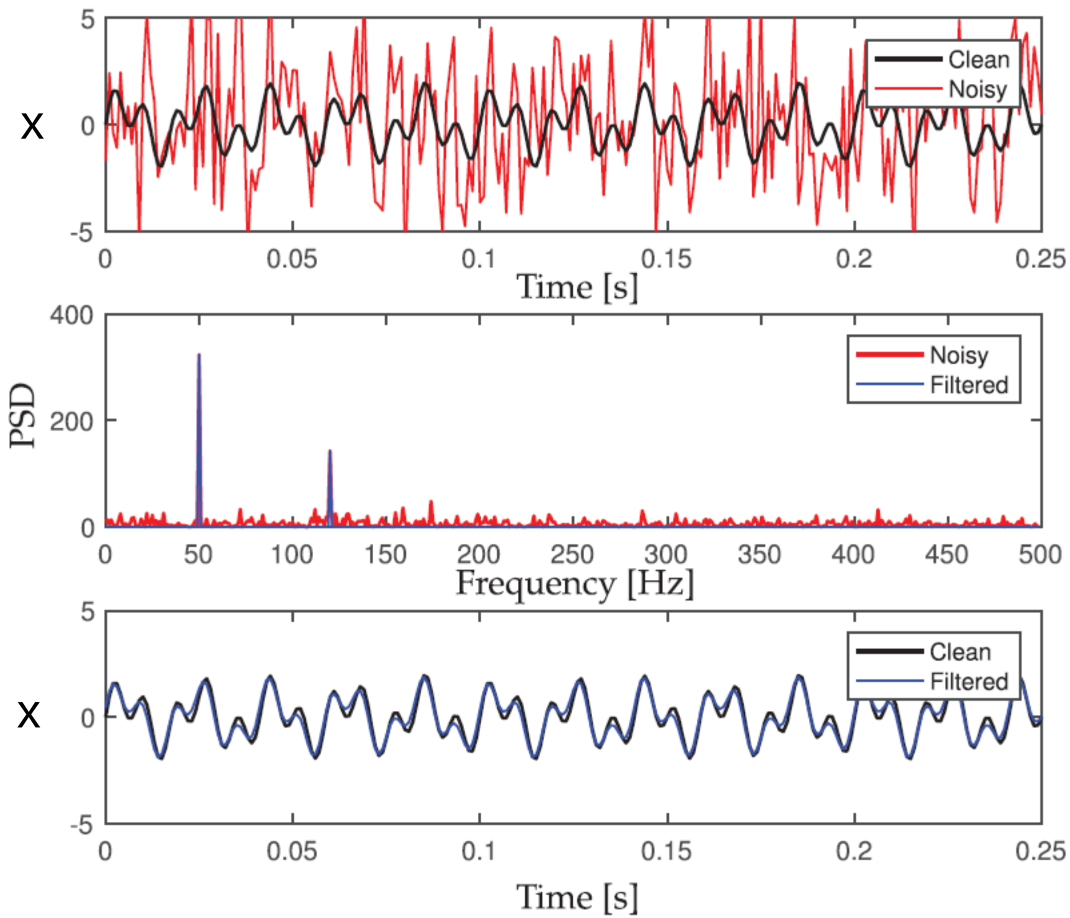
cristina.masoller@upc.edu @cristinamasoll1

The FFT is often used for "de-noising" a time series

In this example noise is added to a signal given by a sum of two sine waves.

In the Fourier domain, dominant peaks may be selected and the noise filtered.

The de-noised signal is obtained by inverse Fourier transforming the two dominant peaks.



Brunton and Kutz, Data-Driven Science and Engineering, Cambridge University Press 2019

cristina.masoller@upc.edu @cristinamasoll1

Univariate time series analysis

- Return maps
- Distribution of data values
- Autocorrelation and Fourier analysis
- **Stochastic models and surrogates**
- Attractor reconstruction, Lyap. exponents, fractal dimension
- Symbolic analysis
- Information theory measure: entropy
- Network representation of a time-series
- Spatio-temporal representation of a time-series

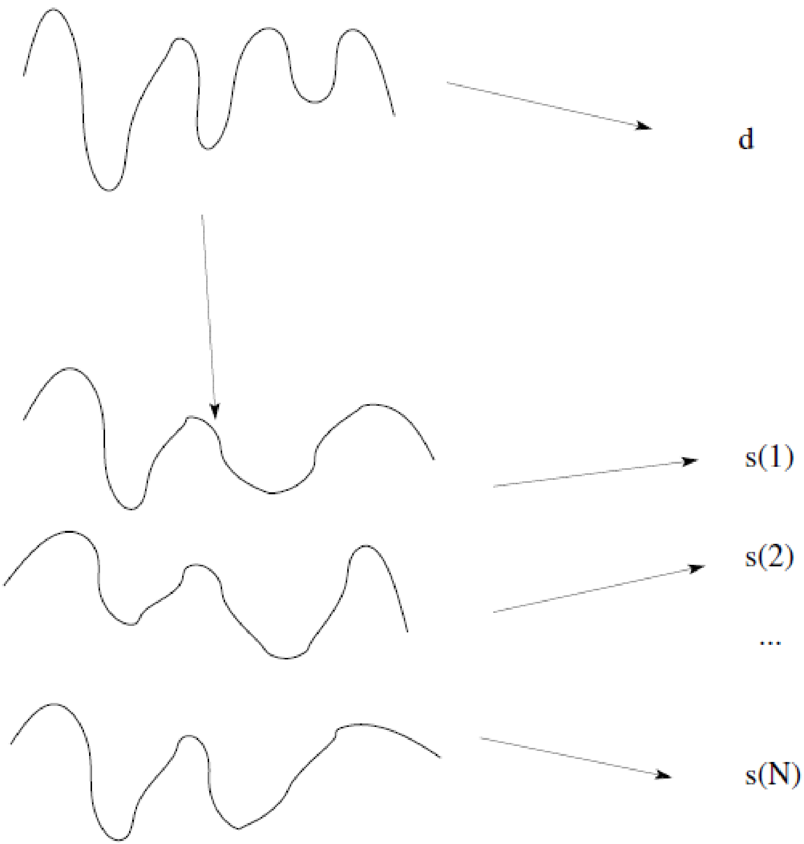
Is a time series generated by a stochastic process?

Is it generated by a nonlinear process?

Is my model good for my data?

cristina.masoller@upc.edu @cristinamasoll1

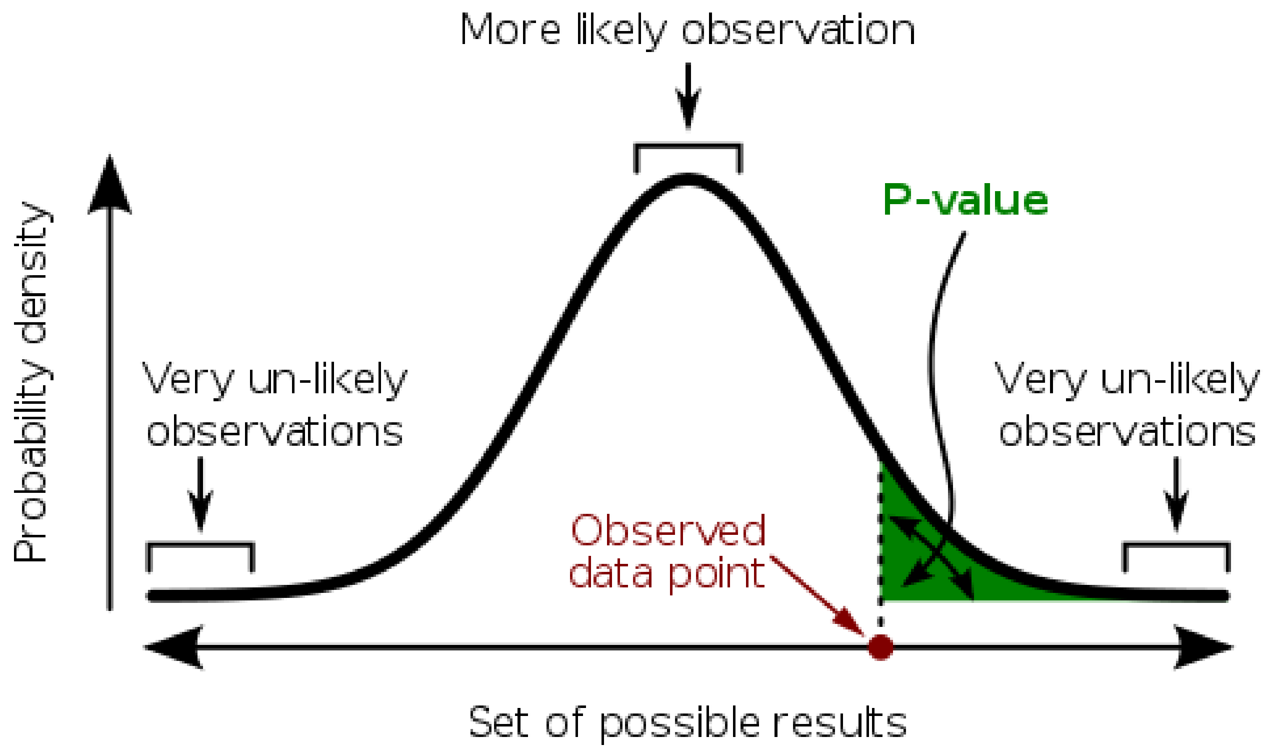
Null hypothesis testing: the method of surrogate data



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

- We have one real time series.
- Generate a set of "surrogate" time series that are "similar" to the original and consistent with the **null hypothesis** (NH) that we want to test.
- Measure an statistical property: "*d*" in the original series and "*s(i)*" in the surrogate 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



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

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

cristina.masoller@upc.edu @cristinamasoll1

Example: the process that generates a time series is nonlinear?

- As all linear systems are time-reversible, time-irreversibility indicates nonlinearity.

$$\alpha = \frac{1}{N} \sum (s_{n+1} - s_n)^3$$

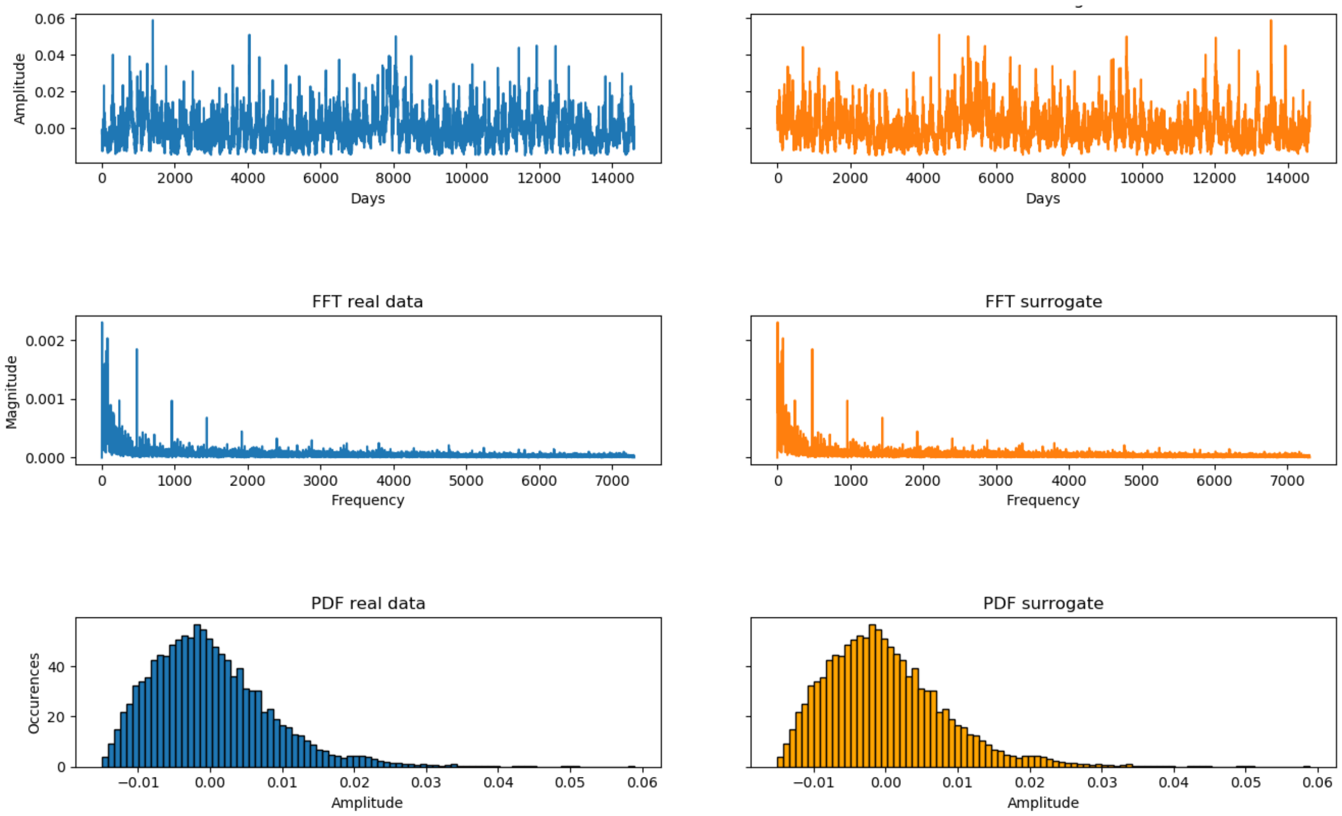
- Necessary but not sufficient condition (some nonlinear systems are time-reversible).

- Surrogates needed that preserve main properties of a time series (e.g., distribution of values, Fourier spectrum).

cristina.masoller@upc.edu @cristinamasoll1

Iterative amplitude adjusted Fourier transform (IAAFT) surrogates

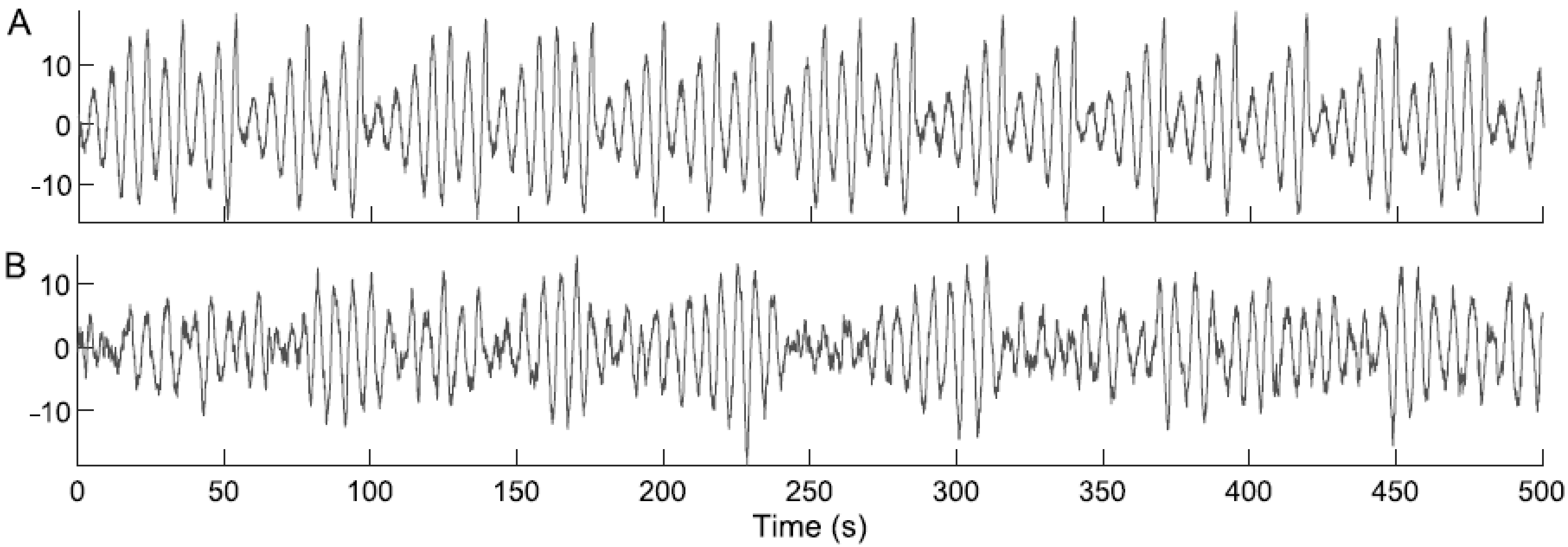
Real data Surrogate data



T. Schreiber and A. Schmitz, Physica D 142, 346 (2000).
R. Silini PhD Thesis, UPC 2022

cristina.masoller@upc.edu @cristinamasoll1

Example of 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

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

A proper surrogate test detects nonlinearity in A (reject NH) but not in B (fail to reject NH).

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

cristina.masoller@upc.edu @cristinamasoll1

Methods of univariate time series analysis

- Return maps
- Distribution of data values
- Autocorrelation and Fourier analysis
- Surrogates
- Attractor reconstruction, Lyapunov exponents, and fractal dimension
- Symbolic analysis
- Information theory measures: entropy and complexity
- Network representation of a time-series
- Spatio-temporal representation of a time-series

cristina.masoller@upc.edu @cristinamasoll1

Why we want to “reconstruct” the phase space of a system from an observed (scalar) time series?

- Real systems are in general high-dimensional and we can only measure a few (hopefully relevant) variables.

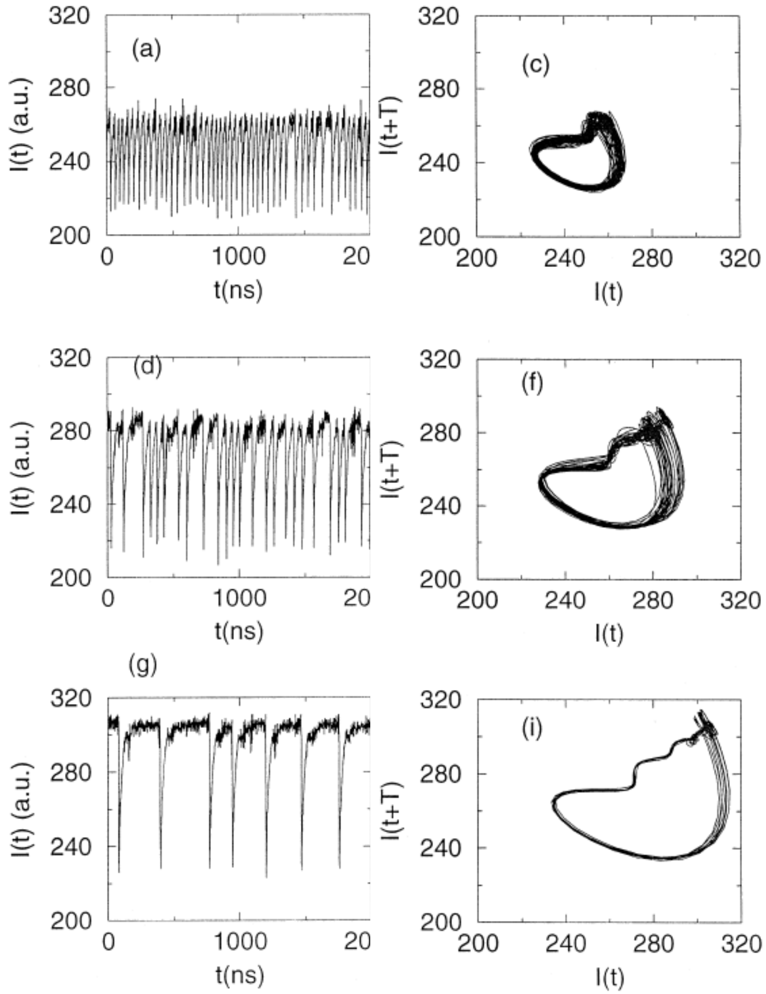
- Models are complex and have many parameters.

- Reconstructing the phase space allows to compare models and experiments and to understand the effect of parameters.

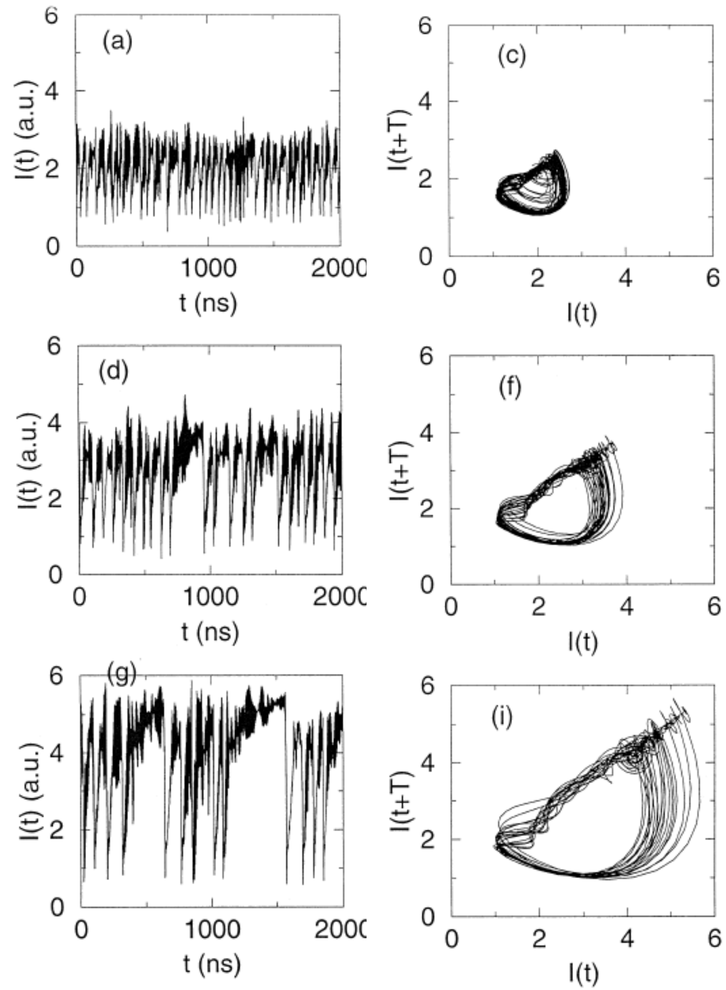
cristina.masoller@upc.edu @cristinamasoll1

Example: 2D representation of an attractor from a scalar time series

Experiments



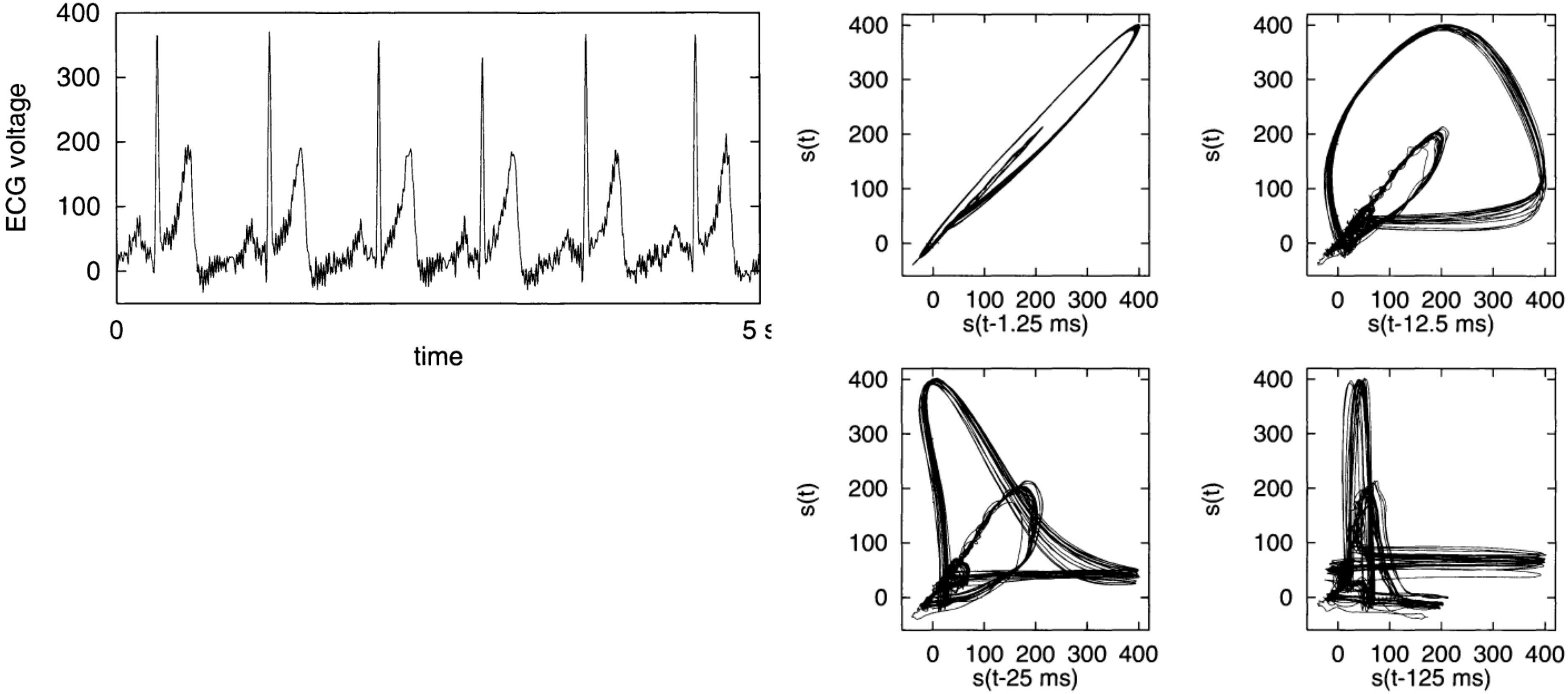
Simulations



C. Masoller et al. Opt. Comm. 157, 115 (1998)

cristina.masoller@upc.edu @cristinamasoll1

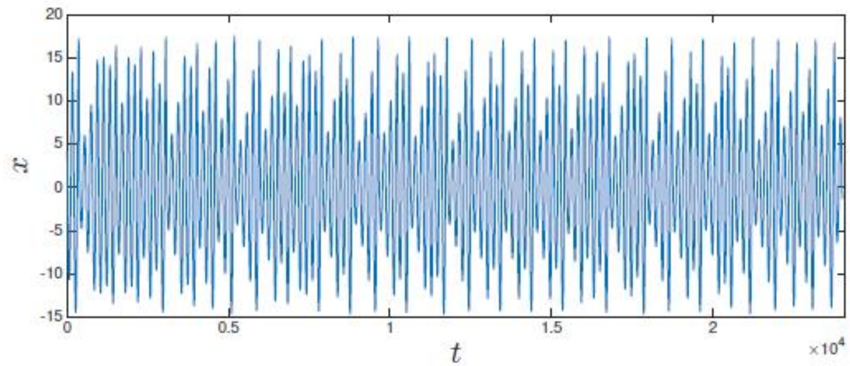
Example: 2D representation of a human ECG signal



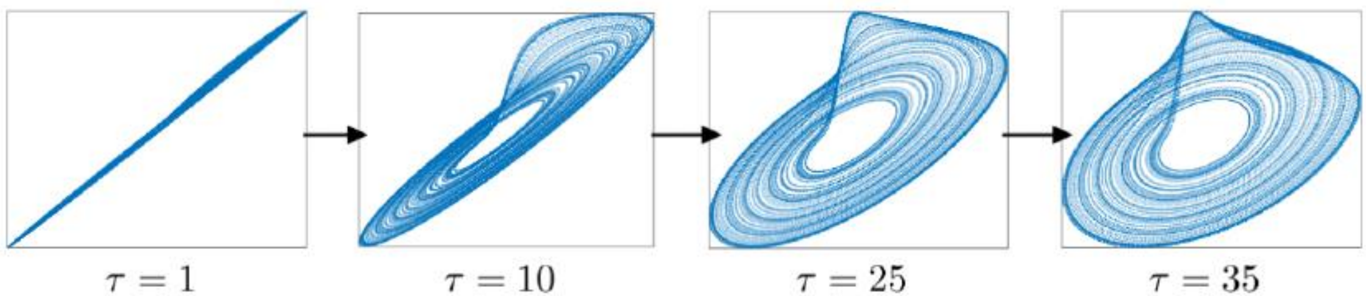
H. Kantz and T. Schreiber,
Cambridge University Press 2003

cristina.masoller@upc.edu @cristinamasoll1

Attractor reconstruction using delay coordinates ("Takens" method)



$$\mathbf{X}(t_i) = \begin{bmatrix} x(t_i) \\ x(t_i + \tau) \\ \vdots \\ x(t_i + (d-1)\tau) \end{bmatrix}$$



Bradley and Kantz, CHAOS 25, 097610 (2015)

Problem: how to chose the embedding parameters (lag τ, dimension d)?

cristina.masoller@upc.edu @cristinamasoll1

How to chose the lag τ and the dimension *d* ?

- τ is chosen such that x(t) and x(t+τ) are *uncorrelated*
  $\rightarrow$the first zero of ACF C(τ) (or where |C(τ)| is minimum).
  What if there is no zero or no minimum?

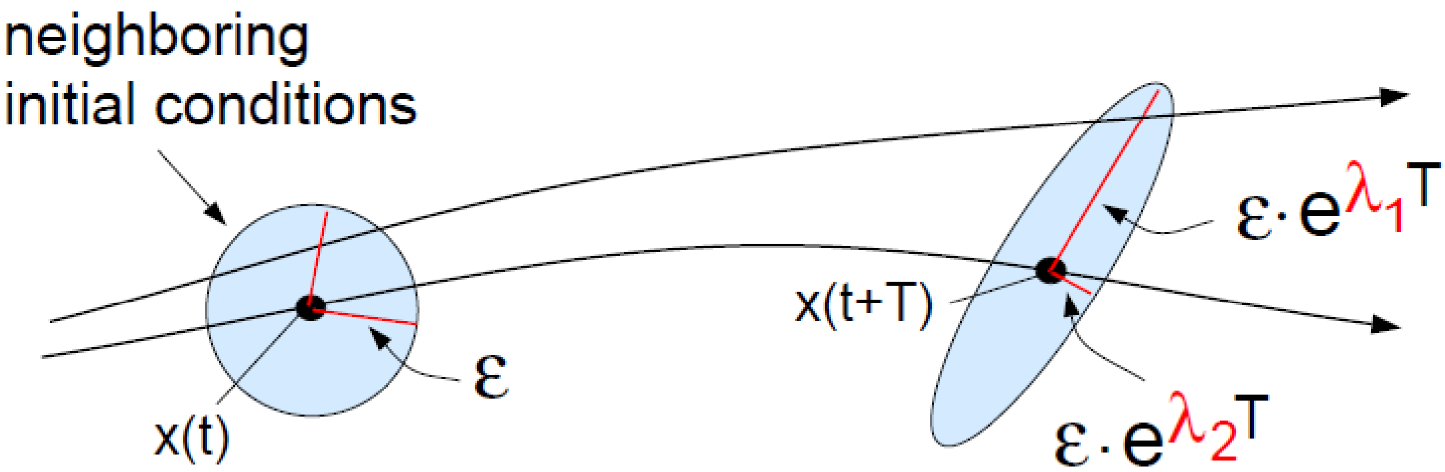
- ***d*** is often estimated with the ***false nearest neighbors*** technique that examines how close points in phase space remain close as the dimension is increased.

- Complicated technique, "practical" approach: try different ***d*** values.

After reconstructing the attractor, we can characterize the attractor calculating the Lyapunov exponents and the fractal dimension.

cristina.masoller@upc.edu @cristinamasoll1

Lyapunov exponents: measure how fast neighboring trajectories diverge.



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

G. Datseris and U. Parlitz, Nonlinear dynamics: a concise introduction interlaced with code, Springer (2022).

cristina.masoller@upc.edu @cristinamasoll1

Steps to compute the maximum Lyapunov Exponent

- Initial distance $\delta_{\mathrm{I}} = |s_i - s_j|$
- Final distance $\delta_{\mathrm{F}} = |s_{i+T} - s_{j+T}|$
- Local *exponential* grow $\lambda^*_{\mathrm{local}} = \frac{1}{T} \log(\delta_{\mathrm{F}}/\delta_{\mathrm{I}})$
- The rate of grow averaged over the attractor gives λ_{max}

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

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

cristina.masoller@upc.edu @cristinamasoll1

A word of warning on the interpretation of the maximum Lyapunov exponent!

- The algorithm returns λ in the fastest expansion direction.
- The algorithm always returns a positive number!

- This is a problem when computing the LE of noisy data.

Every time series analysis algorithm returns a number.

But is it useful?

Significance testing (with appropriate surrogates) is needed.

F. Mitschke and M. Damming, Chaos vs. noise in experimental data, Int. J. Bif. Chaos 3, 693 (1993)

cristina.masoller@upc.edu @cristinamasoll1

Grassberger-Procaccia correlation dimension

- Another very popular method for detecting chaos in real-world data.
- Fractal dimension (box counting dimension): $D_0 = \lim_{\epsilon \to 0} \frac{\log N(\epsilon)}{\log \frac{1}{\epsilon}}$
- **Problem**: for time-series analysis, D_0 does not distinguish between frequently and unfrequently visited boxes.

- **Correlation dimension**: count the number of data points with distance between them <ε *[P. Grassberger and I. Procaccia, Physica D 9, 189 (1983)].*

- Other definitions are based on Information Theory.

cristina.masoller@upc.edu @cristinamasoll1

Hands-on exercise 2: Linear analysis of a time series

Use your own data or download Niño index 3.4 from
https://www.esrl.noaa.gov/psd/gcos_wgsp/Timeseries/Data/nino34.long.anom.data

1. Plot the time series.
2. Plot the return map using different lags.
3. Plot the distribution of data values, calculate μ, σ, S, and K.
4. Calculate the ACF.
5. Calculate the PSD.
6. Compare with the ACF & PSD of Gaussian white noise.

cristina.masoller@upc.edu @cristinamasoll1

Methods of univariate time series analysis

- Return maps
- Distribution of data values
- Autocorrelation and Fourier analysis
- Surrogates
- Attractor reconstruction, Lyapunov exponents, and fractal dimension
- Symbolic analysis
- Information theory measures: entropy and complexity
- Network representation of a time-series
- Spatio-temporal representation of a time-series

cristina.masoller@upc.edu @cristinamasoll1

Sequence of data points ⇒ sequence of symbols

- Use a "rule" to transform $\{x_1, x_2, x_3, \ldots\} \Rightarrow \{s_1, s_2, \ldots\}$
  Example: if $x_i > x_{th} \Rightarrow s_i = 0$; else $s_i = 1$

- Symbols are taken from an "**alphabet**" of possible symbols.

- Then consider "blocks" of D symbols ("**patterns**" or "**words**").
- All the possible words form the "**dictionary**".

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

cristina.masoller@upc.edu @cristinamasoll1

Threshold rule: "partition" of the phase space

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

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

cristina.masoller@upc.edu @cristinamasoll1

Advantages and drawbacks of symbolic analysis

Advantages:

1) No need to "reconstruct" the attractor.
2) Different ways to process the data and/or different rules to define "symbols" may uncover different properties of the time series.

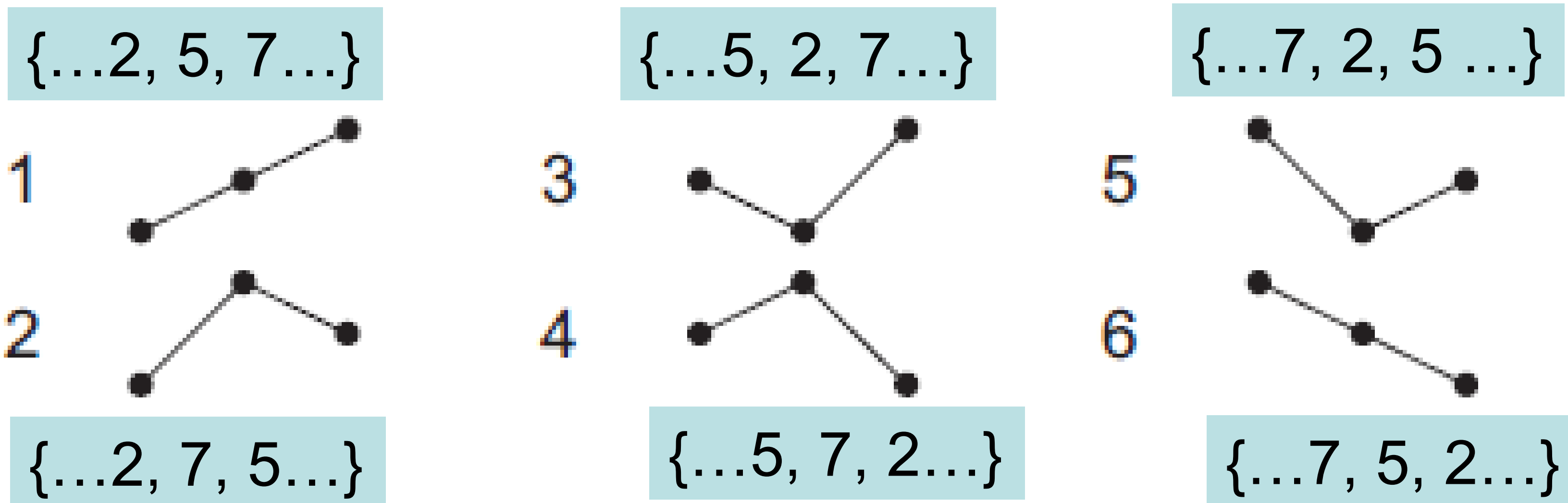
Drawbacks:

1) Most rules have "hyper-parameters" and results often depend on these hyper-parameters.
2) The length of the time series can limit the number of different symbols that can be be used.

cristina.masoller@upc.edu @cristinamasoll1

Ordinal analysis: threshold-less rule

- Which are the possible order relations among three (D=3) consecutive data points in $x(t)=\{\ldots x_i, x_{i+1}, x_{i+2}, \ldots\}$?



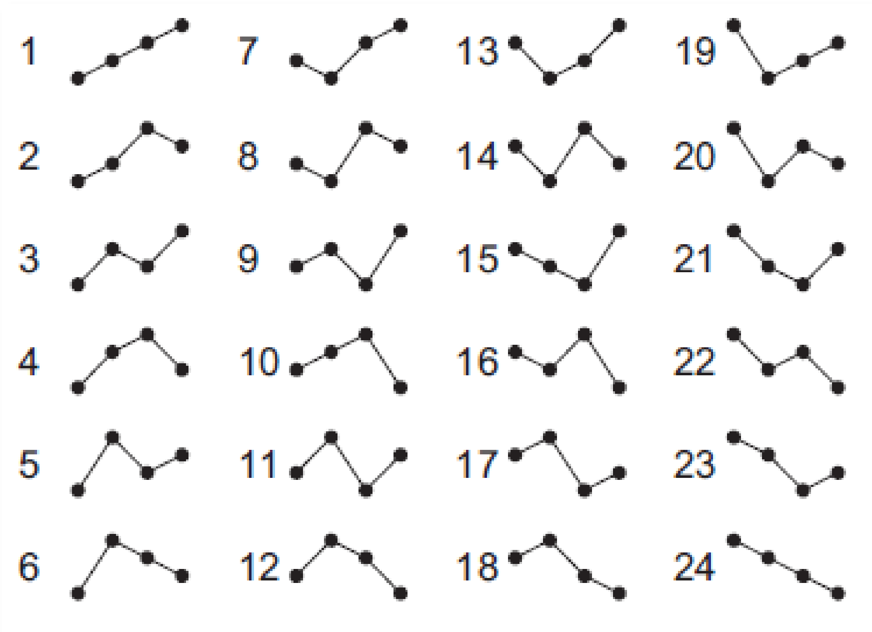
- Count how many times each “ordinal pattern” appears.
- Advantages: allows to identify temporal structures & is robust to noise.
- Drawback: information about actual data values is lost.

Bandt and Pompe, Phys. Rev. Lett. 88, 174102 (2002)

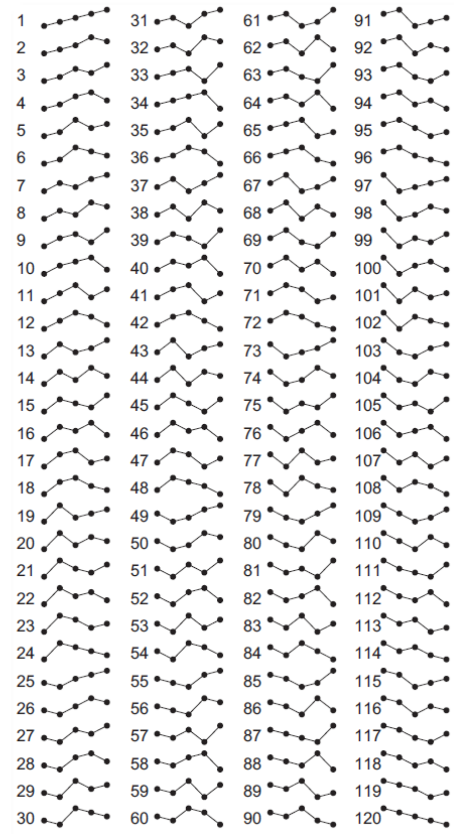
cristina.masoller@upc.edu @cristinamasoll1

The number of patterns increases as D!

D=4



D=5



- A problem for short time series
- How to select optimal D?

  it depends on:
  - The length of the time series
  - The length of the correlations

What if two values are equal? Which is the pattern? A simple solution is to add, to one of the values, a small random perturbation.

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

cristina.masoller@upc.edu @cristinamasoll1

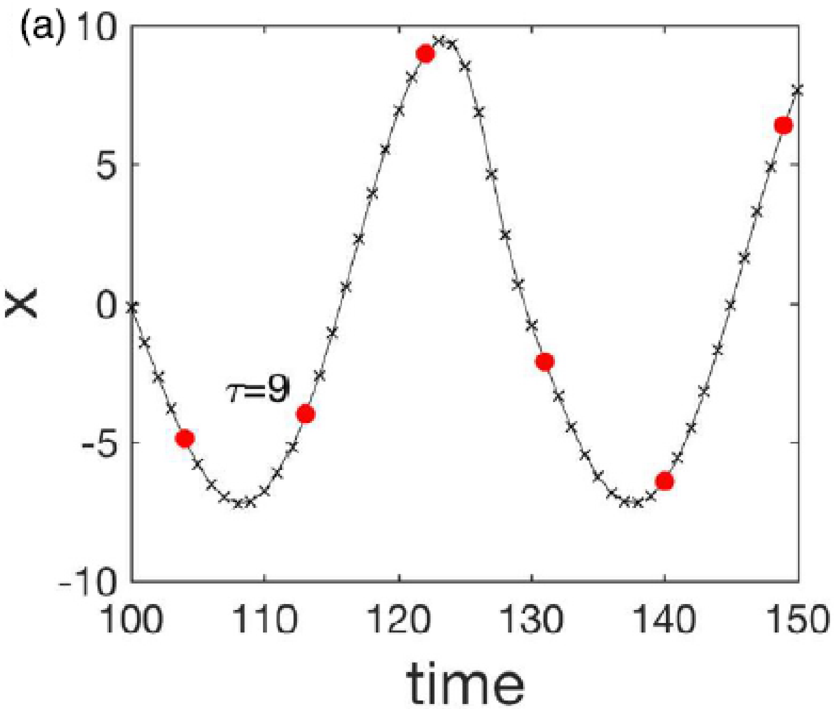
Symbolic analysis can be applied to "raw" data, or to "compressed" / filtered / preprocessed data.

How to "compress"?

First method: Select the temporal resolution (sampling time) to define the symbols.

Example: climatological time series (monthly sampled)

- Consecutive months:
  $[\dots x_i(t),\, x_i(t+1),\, x_i(t+2)\dots]$
- Consecutive years:
  $[\dots x_i(t),\dots\, x_i(t+12),\dots\, x_i(t+24)\dots]$



Y. Zou et al. Phys. Rep. 787, 1 (2019)

cristina.masoller@upc.edu @cristinamasoll1

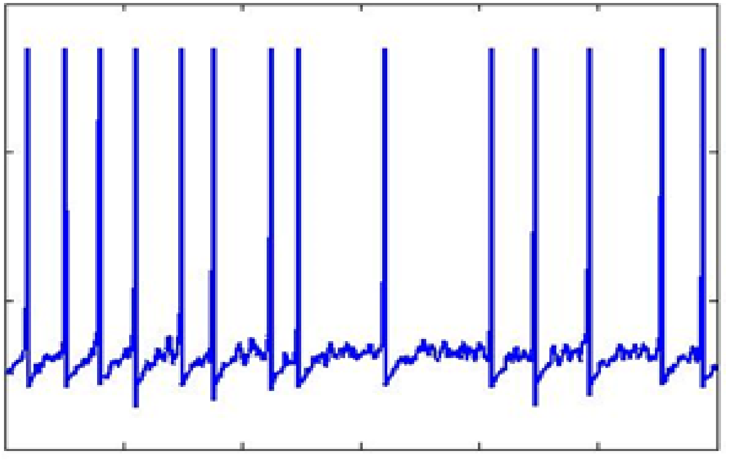
How to “compress” / pre-process the data?

2) Define “events” and analyze the sequence of time-intervals between events (inter-event-times): $\boldsymbol{\Delta T=\{\Delta T_0, \Delta T_1, \Delta T_2, \ldots\}}$

Laser spikes

Time

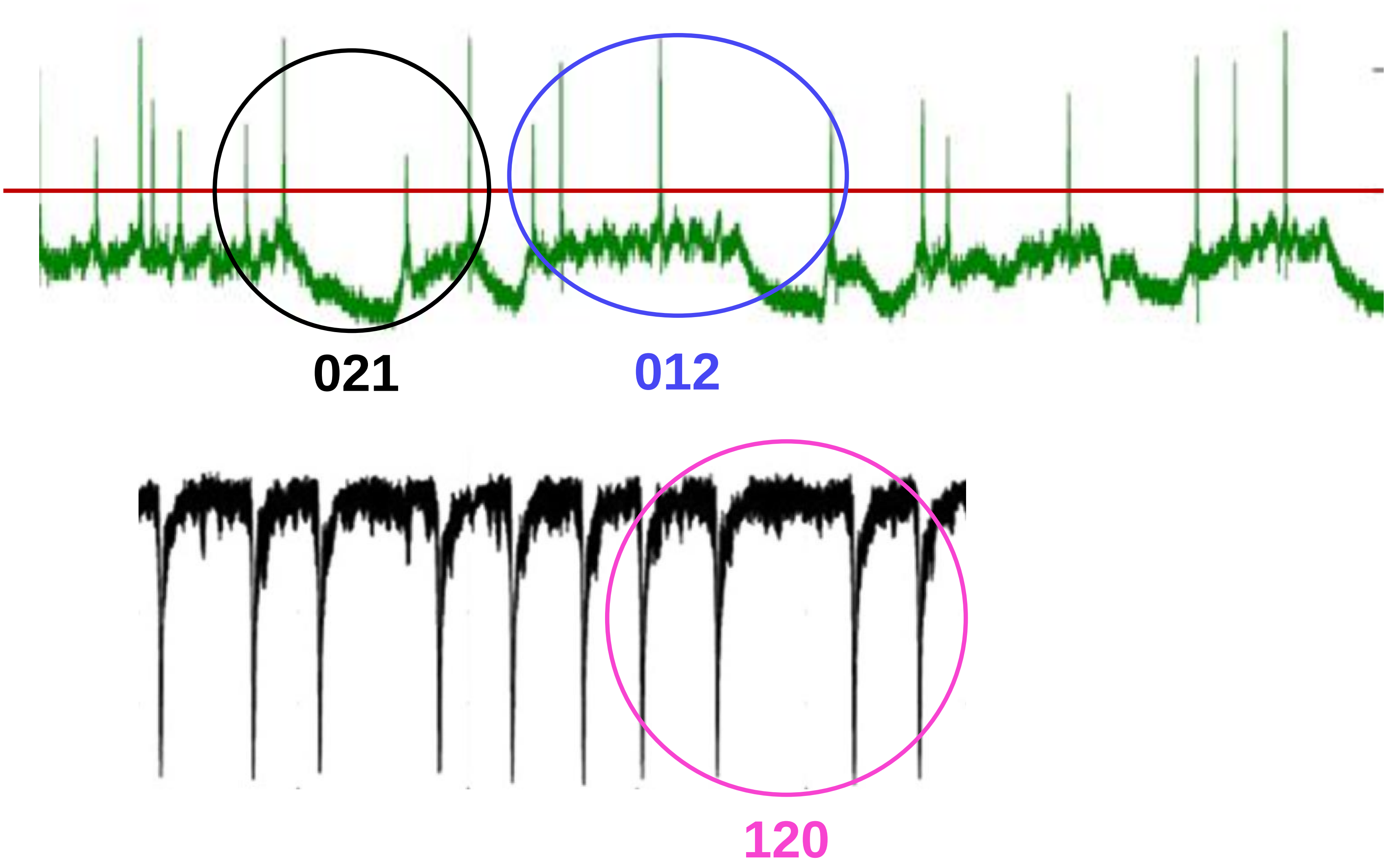
Neuronal spikes



Time

cristina.masoller@upc.edu @cristinamasoll1

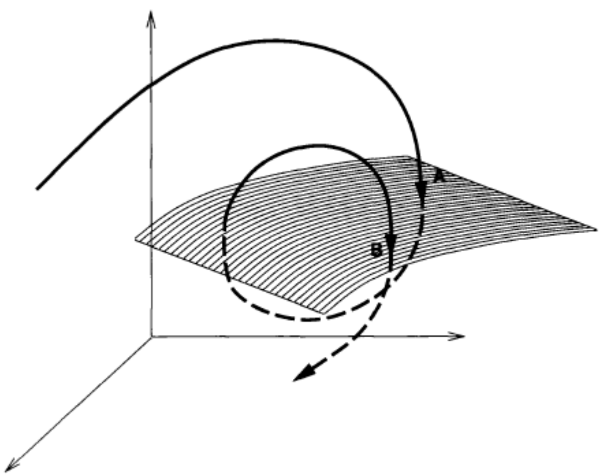
Analysis of D=3 patterns in spike sequences



cristina.masoller@upc.edu @cristinamasoll1

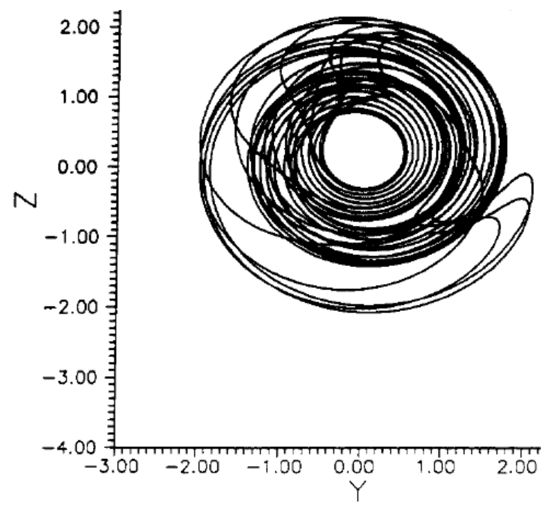
How to “compress” / pre-process the data?

3) Define a Poincare section and analyze the crossings

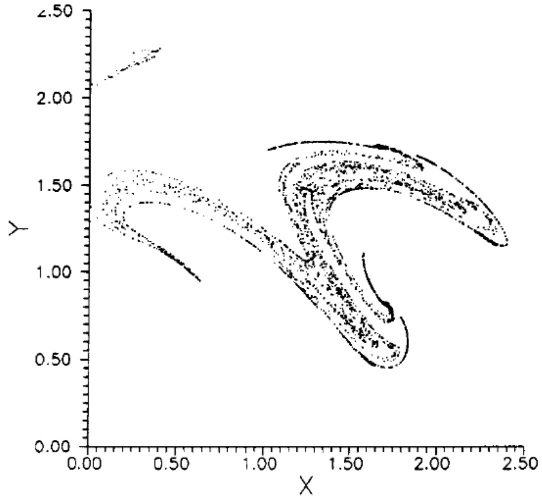


H. Kantz and T. Schreiber,
Cambridge University Press 2003

2D projection of 3D chaotic attractor



X, Y when the Z=0

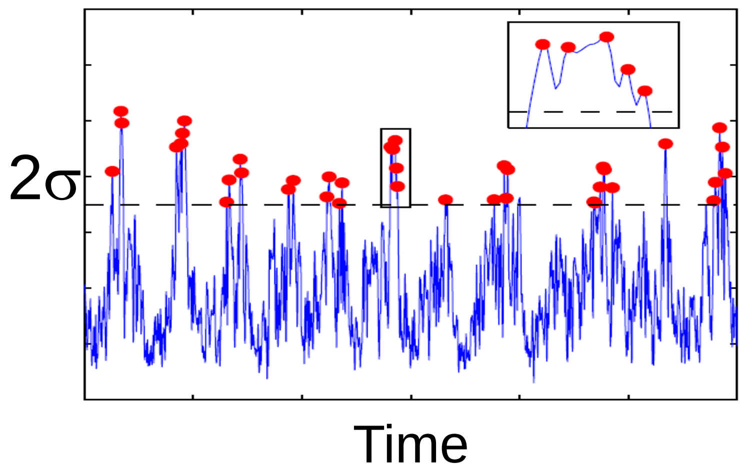


C. Masoller et al, Phys. Lett. A 167, 185 (1992).

cristina.masoller@upc.edu @cristinamasoll1

How to “compress” / pre-process the data?

4) Analyze only the extreme values (either in a given time interval, or above a certain threshold).



cristina.masoller@upc.edu @cristinamasoll1

Comparison between two rules to define symbols

Threshold transformation:

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

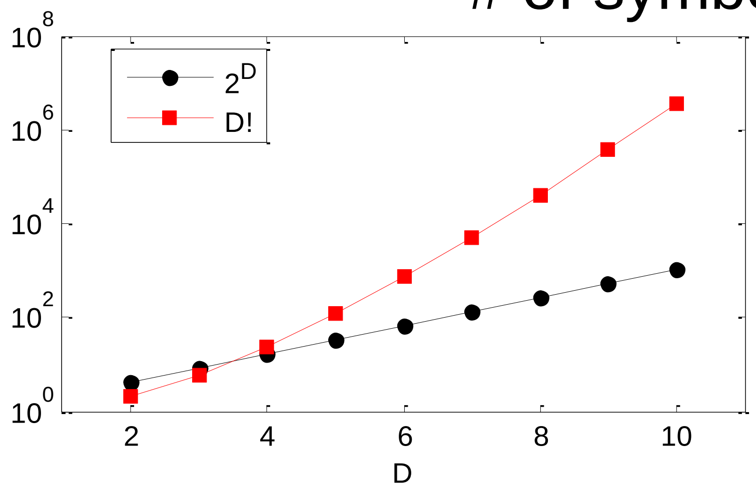
- Advantage: keeps information about the magnitude of the values.
- Drawback: how to select an adequate threshold (“partition” of the phase space).
- # of symbols: 2^D

Ordinal transformation:

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

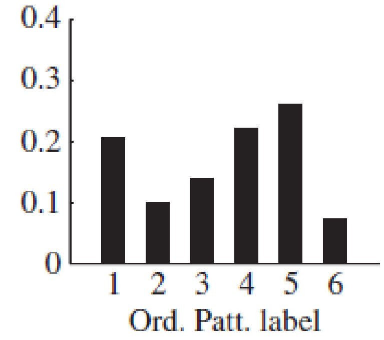
- Advantage: no need of threshold; keeps information about the temporal order in the sequence of values
- Drawback: no information about the actual data values
- # of symbols: D!

Number of symbols in log scale



cristina.masoller@upc.edu @cristinamasoll1

Are the D! ordinal patterns equally probable?



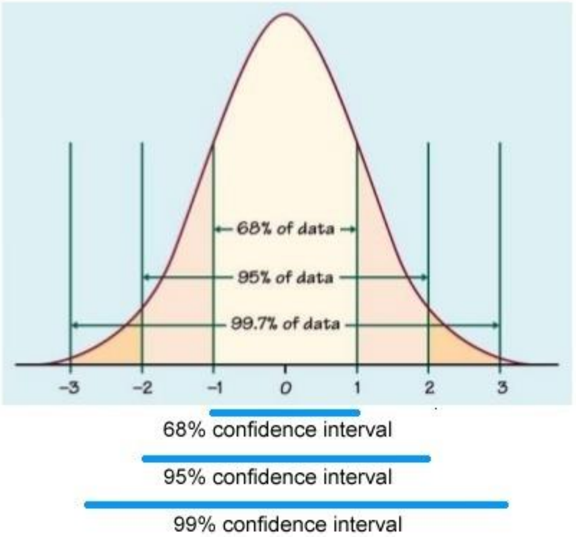
- **Null hypothesis**:

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

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

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

- Else, we **fail to reject** the NH with 99.74% confidence level.



cristina.masoller@upc.edu @cristinamasoll1

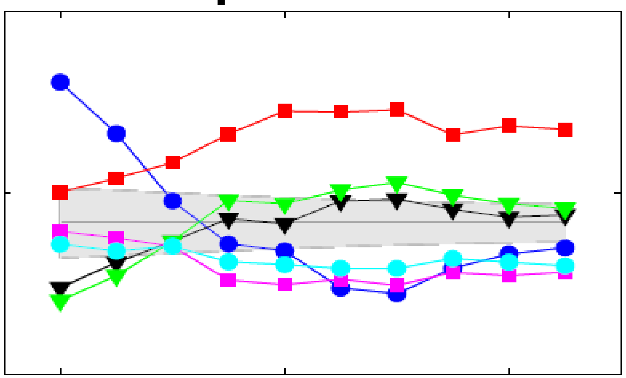
Applications

Symbolic analysis gives information about the presence of more/less expressed patterns in data that can be used to:

- compare real data – model data, fit model parameters
- classification

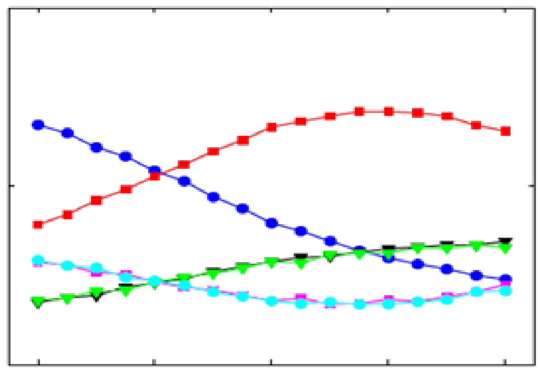
Examples:

Laser spike intervals



Experimental parameter

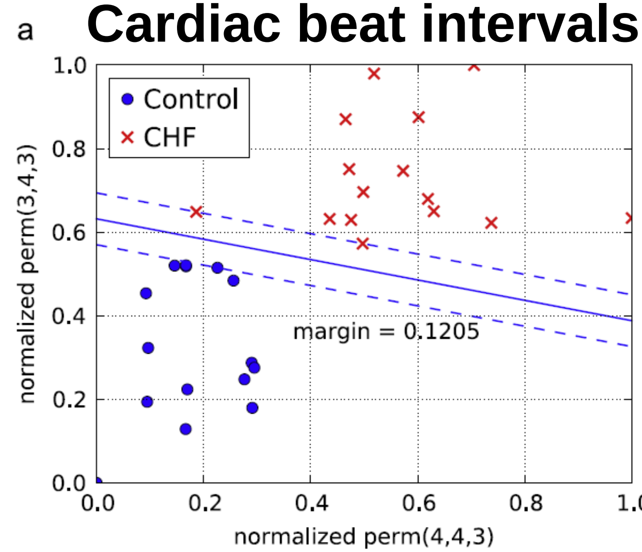
Circle map



Control parameter

A. Aragoneses et al, Sci. Rep. 4, 4696 (2014)

Cardiac beat intervals



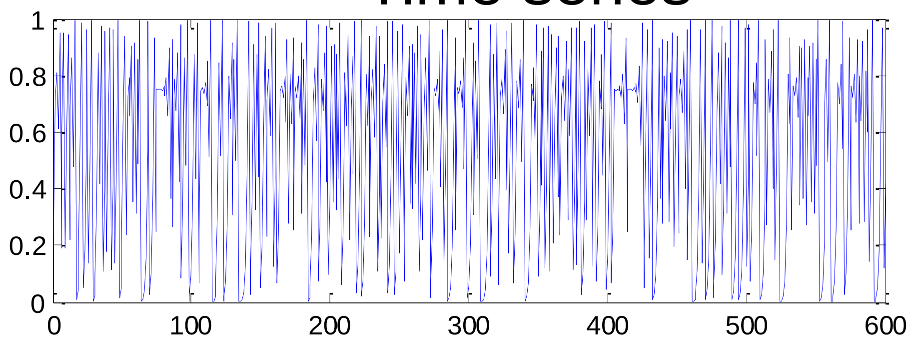
U. Parlitz et al. Comp. in Bio. and Med. 42, 319 (2012)

cristina.masoller@upc.edu @cristinamasoll1

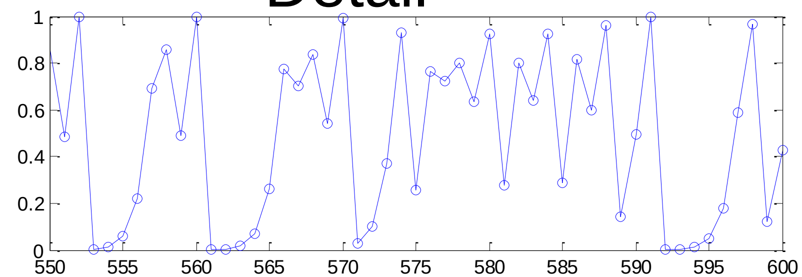
Example: chaotic time series generated with the Logistic map

$$x(i+1) = r\, x(i)[1 - x(i)] \quad r=3.99$$

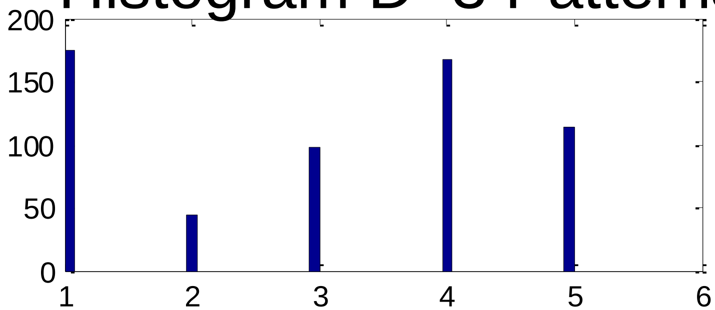
Time series



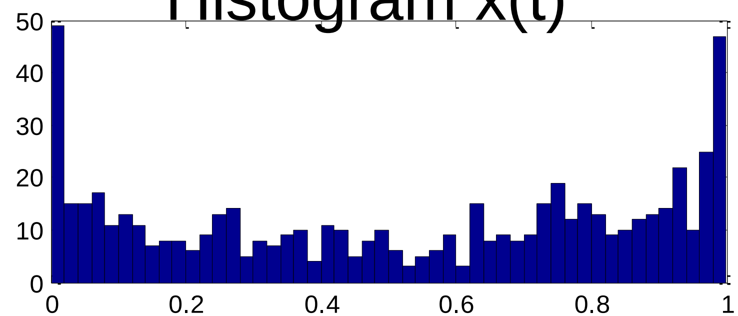
Detail

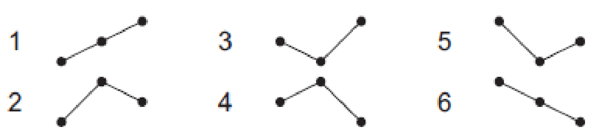


Histogram D=3 Patterns



Histogram x(t)



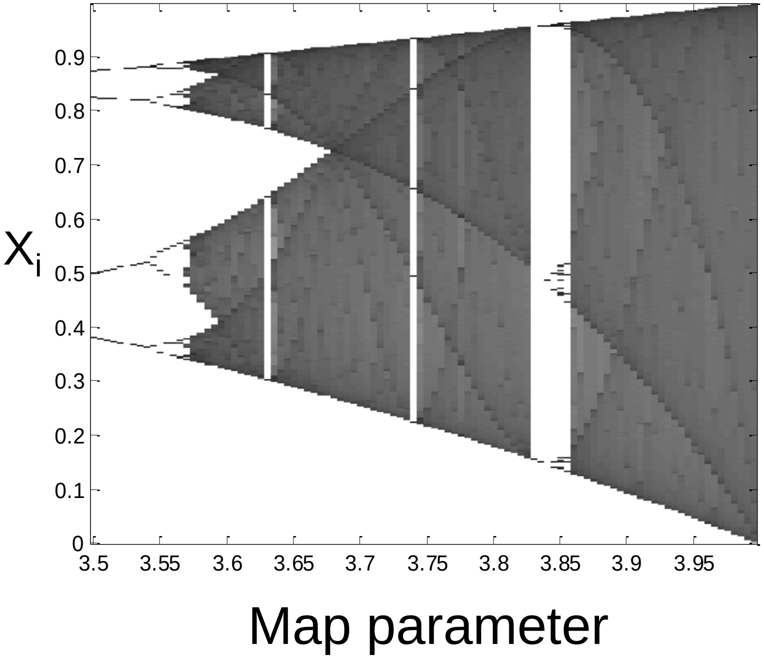


↑ **forbidden**

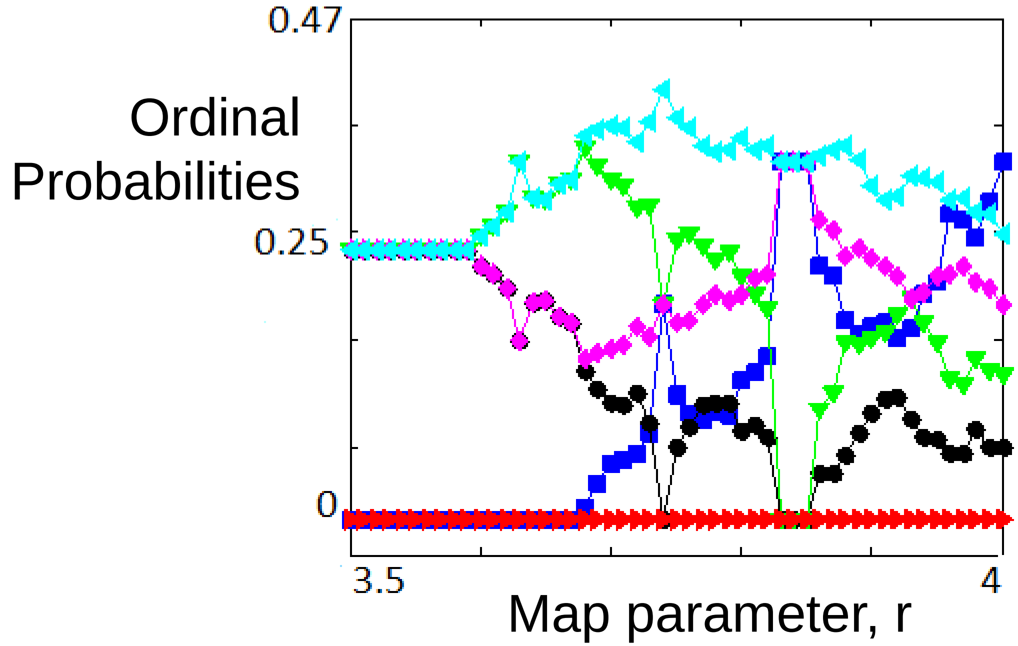
cristina.masoller@upc.edu @cristinamasoll1

“Normal” and “Ordinal” bifurcation diagrams of the Logistic map

Normal bifurcation diagram



Ordinal diagram with D=3



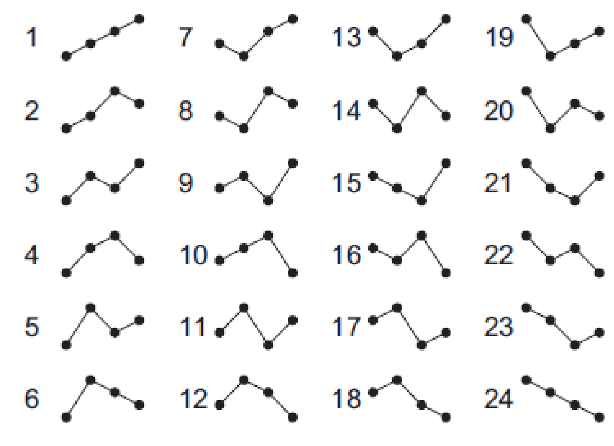
012 021 102 120 201 210

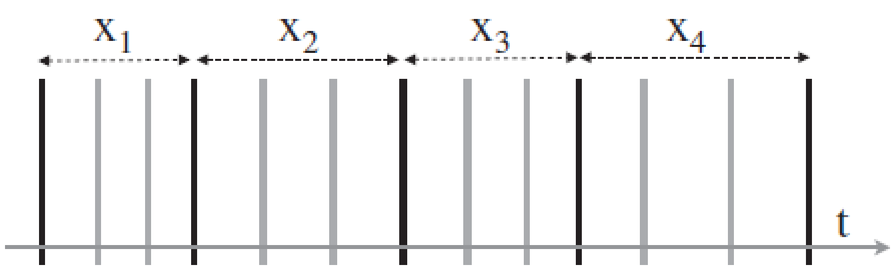
Pattern **210** is always forbidden; pattern **012** is more probable as r increases

cristina.masoller@upc.edu @cristinamasoll1

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)]

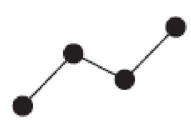




World length (wl): 4

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 ;
```


cristina.masoller@upc.edu @cristinamasoll1

Methods of univariate time series analysis

- Return maps
- Distribution of data values
- Autocorrelation and Fourier analysis
- Surrogates
- Attractor reconstruction, Lyapunov exponents, and fractal dimension
- Symbolic analysis
- Information theory measure: entropy
- Network representation of a time-series
- Spatio-temporal representation of a time-series

cristina.masoller@upc.edu @cristinamasoll1

Entropy (disorder) and information

High entropy low information

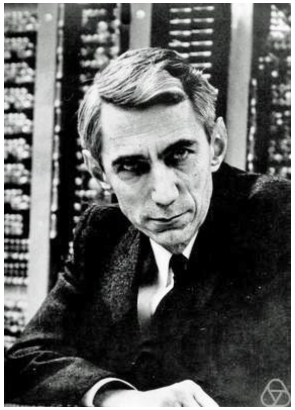
Low entropy high information

https://imgur.com/gallery/Otg97

cristina.masoller@upc.edu @cristinamasoll1

Shannon entropy

$$\sum_{i=1}^{N} p_i = 1 \qquad \boxed{H = -\sum_{i=1}^{N} p_i \ln p_i}$$



Claude Shannon

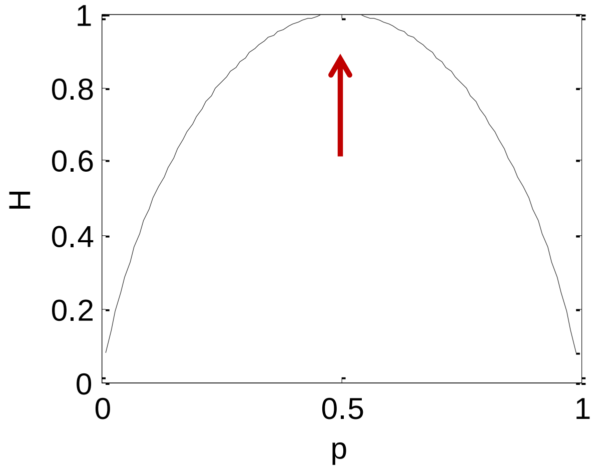
- Interpretation: "*quantity of* ***surprise*** *one should feel upon reading the result of a measurement* ".

- Example: a random variable takes values 0 or 1 with probabilities:

  p(0) = p, p(1) = 1 − p.

  H = −p ln(p) − (1 − p) ln(1 − p).

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



C. Shannon, "A Mathematical Theory of Communication", Bell System Technical Journal. 27 (3): 379–423 (1948). Bell System Technical Journal. 27 (4): 623–656 (1948).

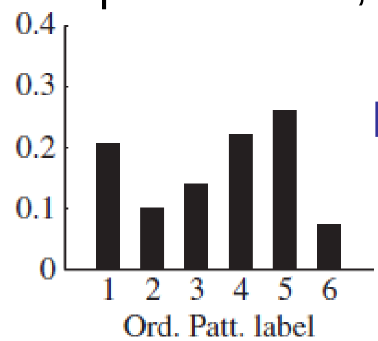
cristina.masoller@upc.edu @cristinamasoll1

Permutation entropy

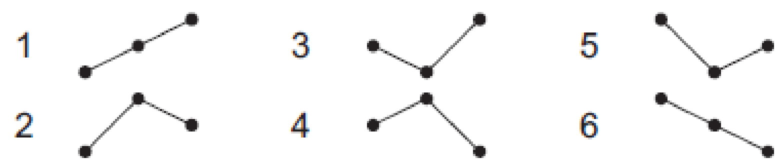
Shannon entropy of a time series, computed from the probabilities of the ordinal patterns.

Time series

Ordinal probabilities, D=3



$$\sum_{i=1}^{N} p_i = 1$$



Permutation entropy, D=3

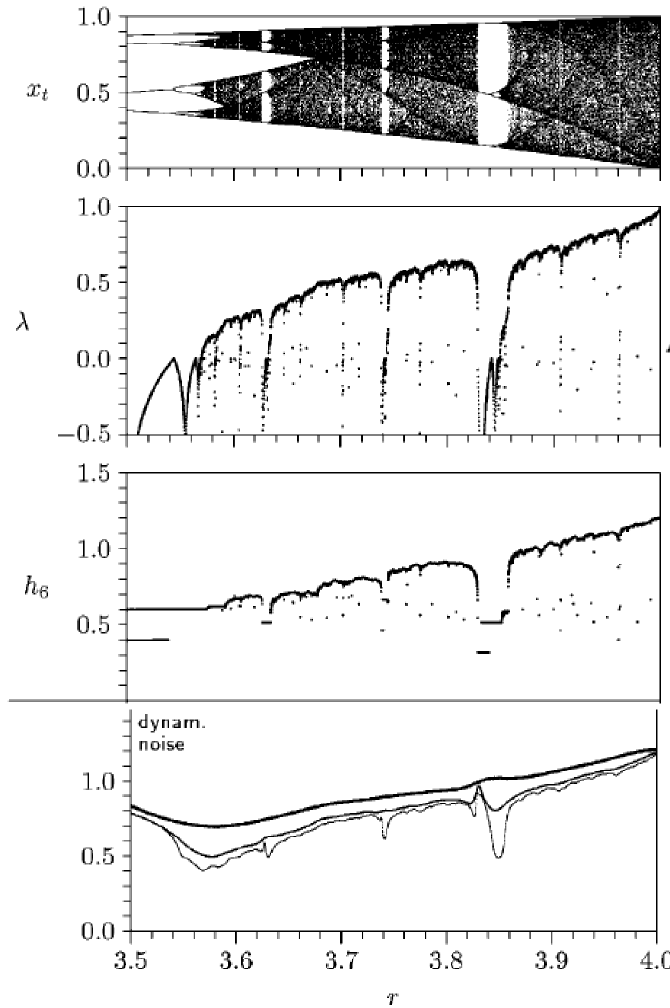
$$H = -\sum_{i=1}^{N} p_i \ln p_i$$

Bandt and Pompe, Phys. Rev. Lett. 88, 174102 (2002)

cristina.masoller@upc.edu @cristinamasoll1

Permutation entropy (PE) of the Logistic map

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



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

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

Entropy per symbol:

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

Robust to noise

Bandt and Pompe, Phys. Rev. Lett. 88, 174102 (2002)

cristina.masoller@upc.edu @cristinamasoll1

Permutation entropy analysis of financial data

Physica A 391 (2012) 4342–4349

Contents lists available at SciVerse ScienceDirect

Physica A

journal homepage: www.elsevier.com/locate/physa

On the efficiency of sovereign bond markets

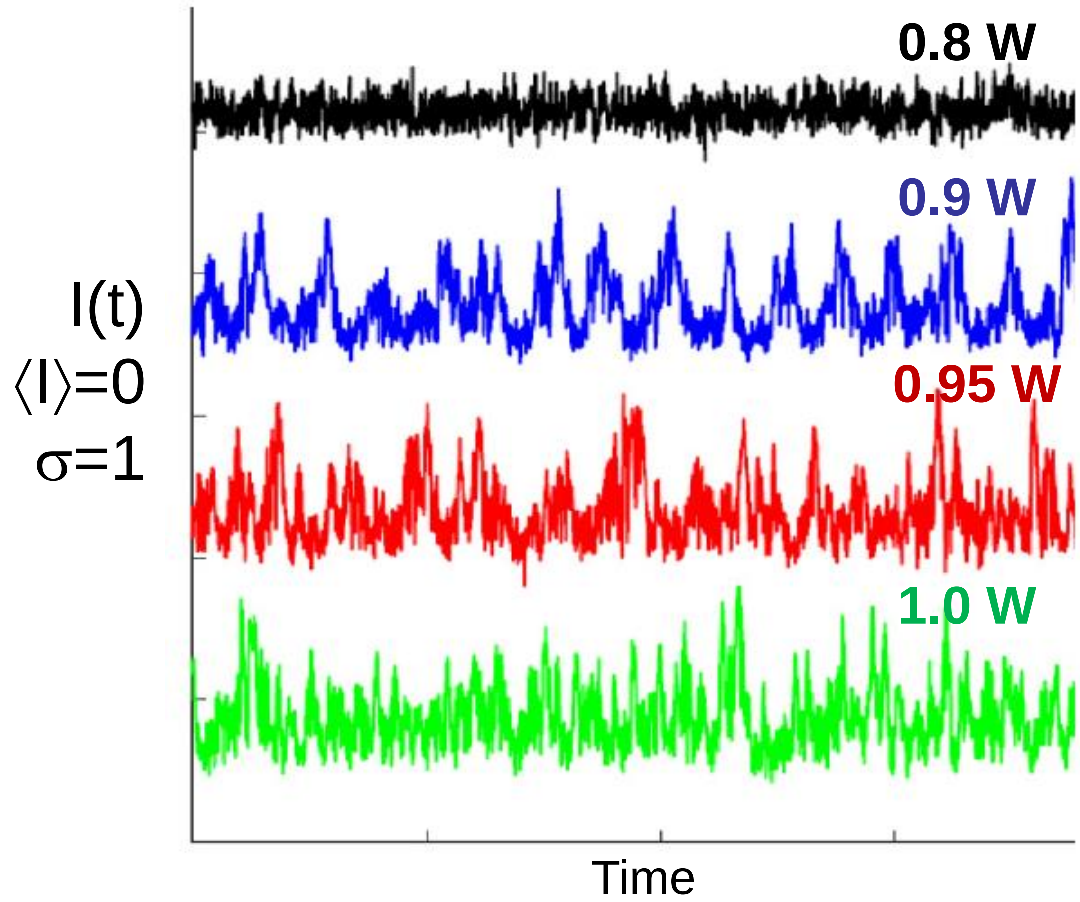
Luciano Zunino[a,b,*], Aurelio Fernández Bariviera[c], M. Belén Guercio[c], Lisana B. Martinez[c], Osvaldo A. Rosso[d,e]

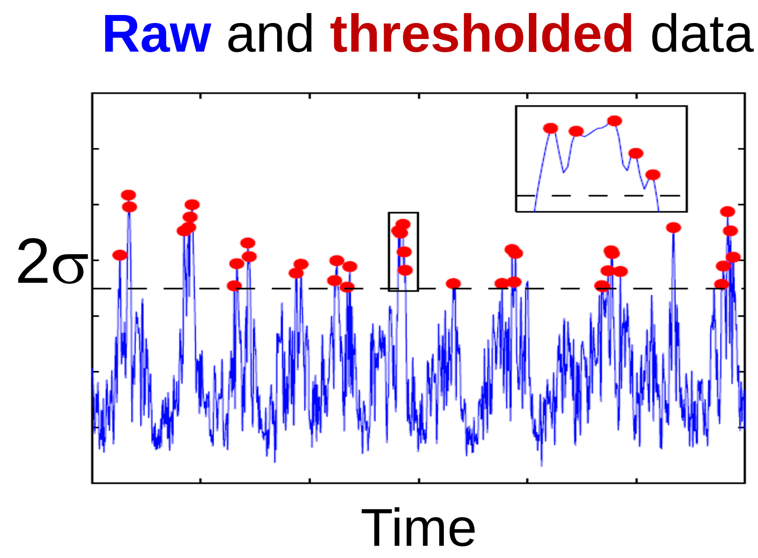
[a] Centro de Investigaciones Ópticas (CONICET La Plata - CIC), C.C. 3, 1897 Gonnet, Argentina
[b] Departamento de Ciencias Básicas, Facultad de Ingeniería, Universidad Nacional de La Plata (UNLP), 1900 La Plata, Argentina
[c] Department of Business, Universitat Rovira i Virgili, Av. Universitat 1, 43204 Reus, Spain
[d] LaCCAN/CPMAT - Instituto de Computação, Universidade Federal de Alagoas, BR 104 Norte km 97, 57072-970 Maceió, Alagoas, Brazil
[e] Laboratorio de Sistemas Complejos, Facultad de Ingeniería, Universidad de Buenos Aires. 1063 Av. Paseo Colón 840, Ciudad Autónoma de Buenos Aires, Argentina

market dynamics. We conclude that the classification derived from the complexity-entropy causality plane is consistent with the qualifications assigned by major rating companies to the sovereign instruments. Additionally, we find a correlation between permutation entropy, economic development and market size that could be of interest for policy makers and investors.

cristina.masoller@upc.edu @cristinamasoll1

Application to the Laminar → Turbulence transition in a fiber laser as the pump power increases





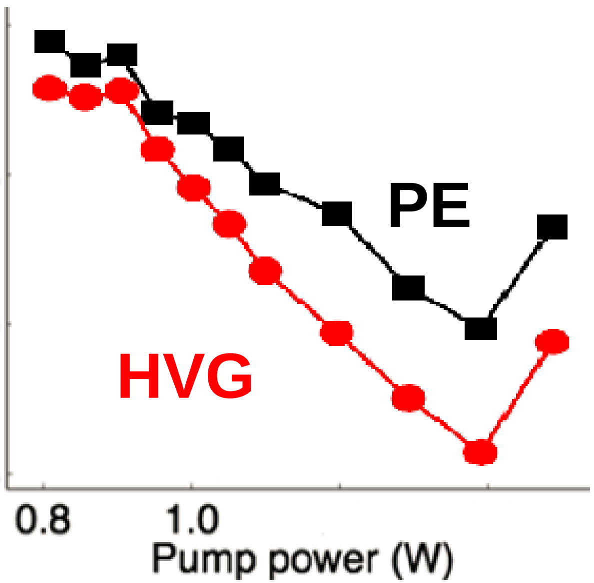
E. G. Turitsyna et al., Nat. Phot. 7, 783 (2013).
A. Aragoneses et al., Phys. Rev. Lett. 116, 033902 (2016).
L. Carpi and C. Masoller, Phys. Rev. A 97, 023842 (2018).

cristina.masoller@upc.edu @cristinamasoll1

The variation of the entropy with a parameter depends on the set of probabilities used.

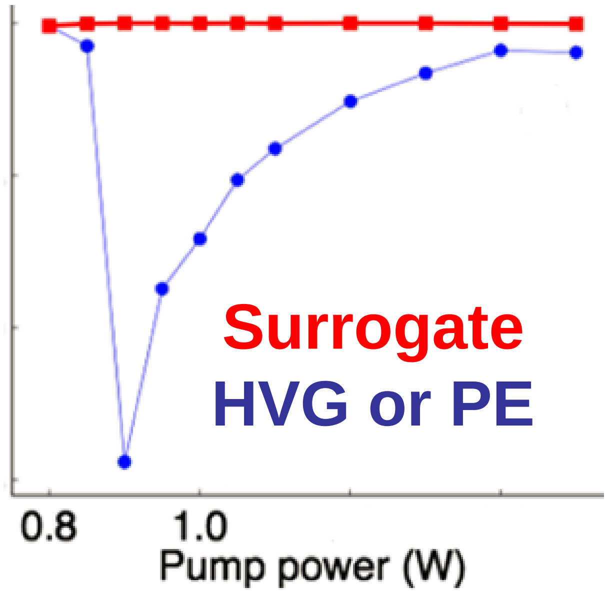
$$H = -\sum p_i \ln p_i$$

Probabilities of symbols in "raw" data



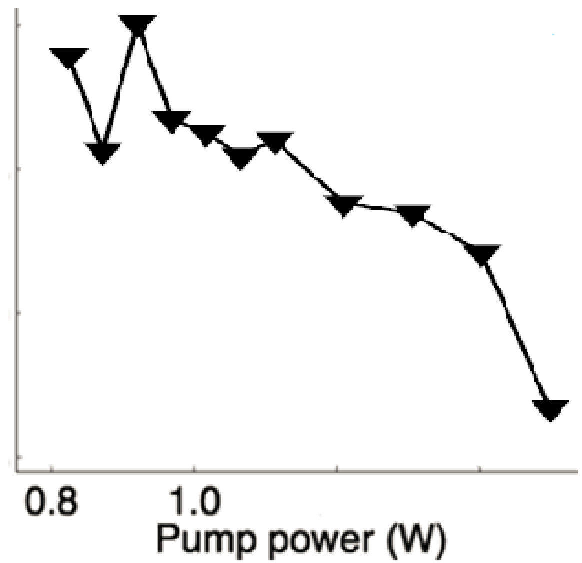
(HVG: horizontal visibility graphs –is another symbolic method)

Probabilities of symbols in "thresholded" data



(the abrupt transition is robust to variations of the threshold)

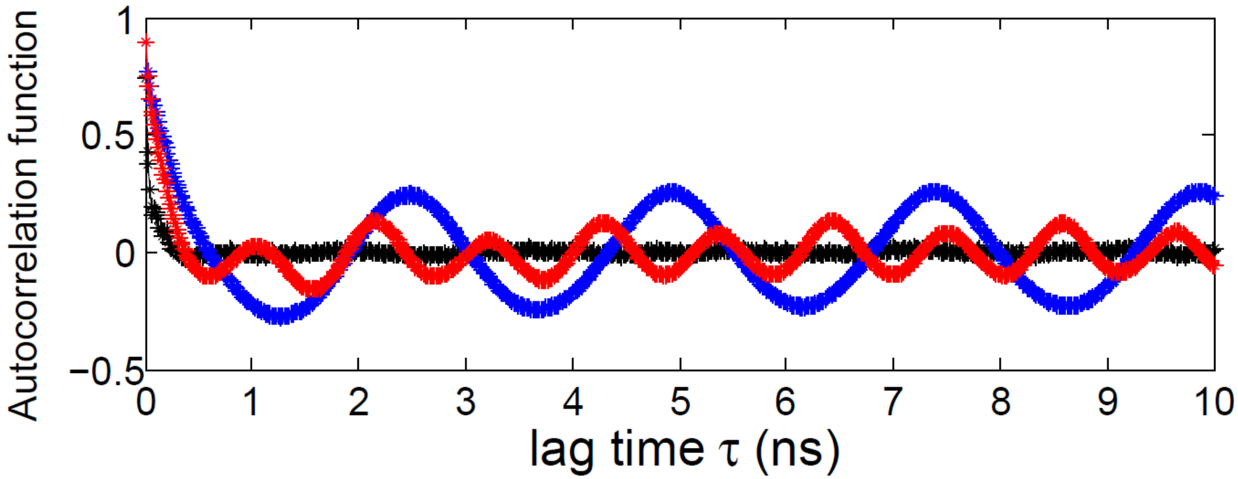
Probabilities of "raw" intensity values



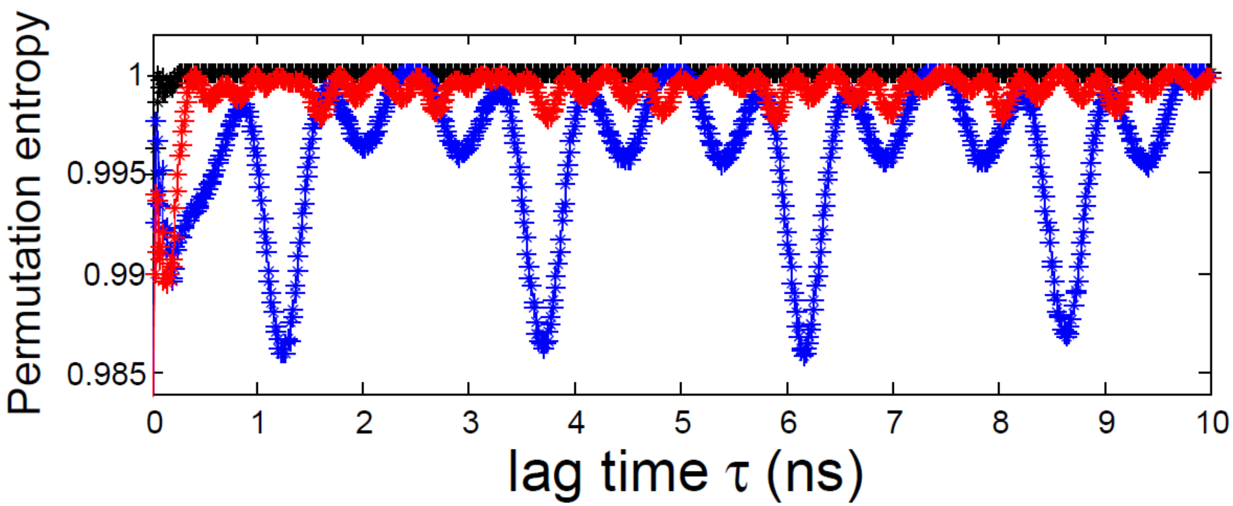
A. Aragoneses et al., Phys. Rev. Lett. 116, 033902 (2016).

cristina.masoller@upc.edu @cristinamasoll1

Varying the sampling time (τ) can identify ``hidden'' periodicities in raw data



$$C(\tau) = \frac{\langle [I_i - \langle I \rangle][I_{i-\tau} - \langle I \rangle] \rangle}{\sigma^2}$$



$$\{I_i, I_{i+\tau}, I_{i+2\tau}, \ldots\}$$

$$H = -\sum_i p_i \log_2 p_i$$

A. Aragoneses et al., Phys. Rev. Lett. 116, 033902 (2016).

cristina.masoller@upc.edu @cristinamasoll1

Renyi entropy, generalized dimension, and multifractality

$$H_q = \frac{1}{1-q} \ln\left(\sum_{i=1}^{N} p_i{}^q \right)$$

In the limit $q \to 1$ $H_q \to H$

A set is multi-fractal if H_q changes with q.
Most chaotic sets are multi-fractal.

$$D_q = \lim_{\varepsilon \to 0} \frac{-H_q(\varepsilon)}{\varepsilon}$$

ε: size of the bin for defining the probabilities

If q=0: $H_0 = \ln N \Rightarrow D_0 = -\lim_{\varepsilon \to 0} (\ln N)/\varepsilon$

D_0 is the box-counting dimension: # of occupied boxes $N \propto (1/\varepsilon)^D$

If $q=1$: Information dimension $D_1 = \lim_{\varepsilon \to 0} \frac{-H(\varepsilon)}{\varepsilon}$

cristina.masoller@upc.edu @cristinamasoll1

Methods of time-series analysis

- Return maps
- Correlation and Fourier analysis
- Stochastic models and surrogates
- Distribution of data values
- Attractor reconstruction: Lyapunov exponents and fractal dimensions
- Symbolic methods
- Information theory measure: entropy
- Network representation of a time-series
- Spatio-temporal representation of a time-series

cristina.masoller@upc.edu @cristinamasoll1

Example: ordinal network

We use ordinal patterns as the nodes of the network and the transition probabilities (TPs) $\alpha \rightarrow \beta$ define the links.

- Adjacency matrix:

  $w_{ij} = TP(i \rightarrow j)$

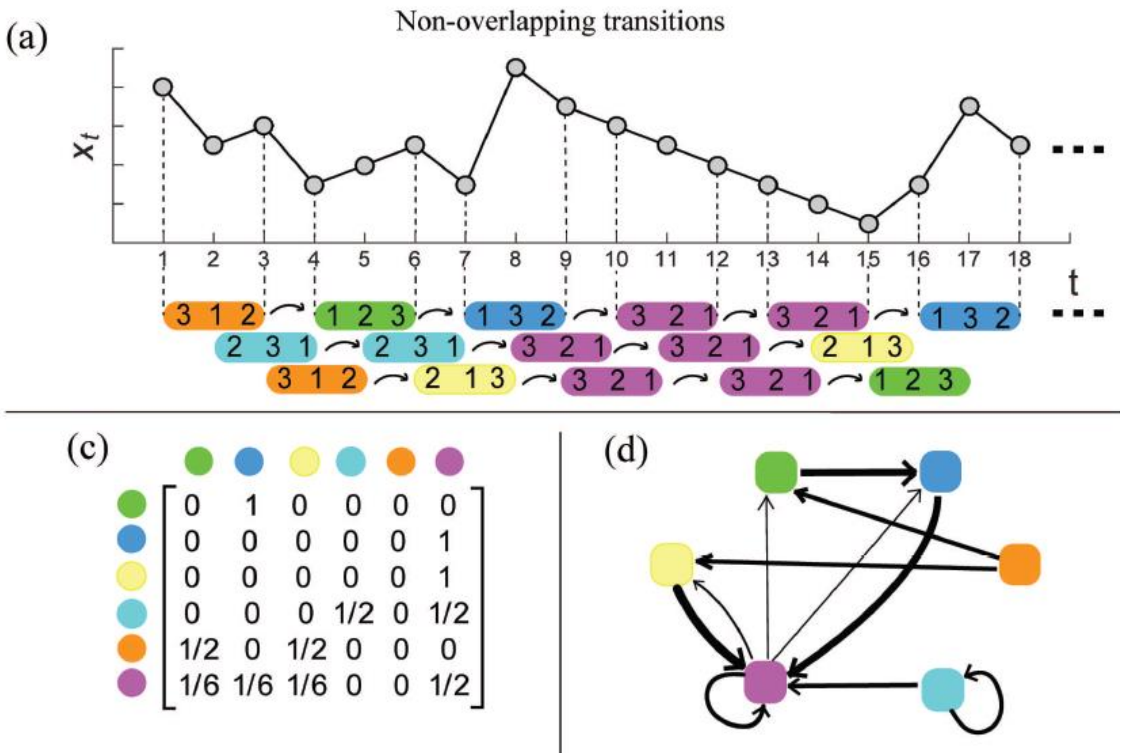
- In each node *i*:

  $\sum_j w_{ij} = 1$

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

  $(\sum_i p_i = 1)$

  $\Rightarrow$ Weighted and directed network



F. Olivares et al., Chaos 30, 063101 (2020).

cristina.masoller@upc.edu @cristinamasoll1

Network-based diagnostic tools

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

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

- Node entropy (entropy of each node, computed from the transition probabilities) and the average

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

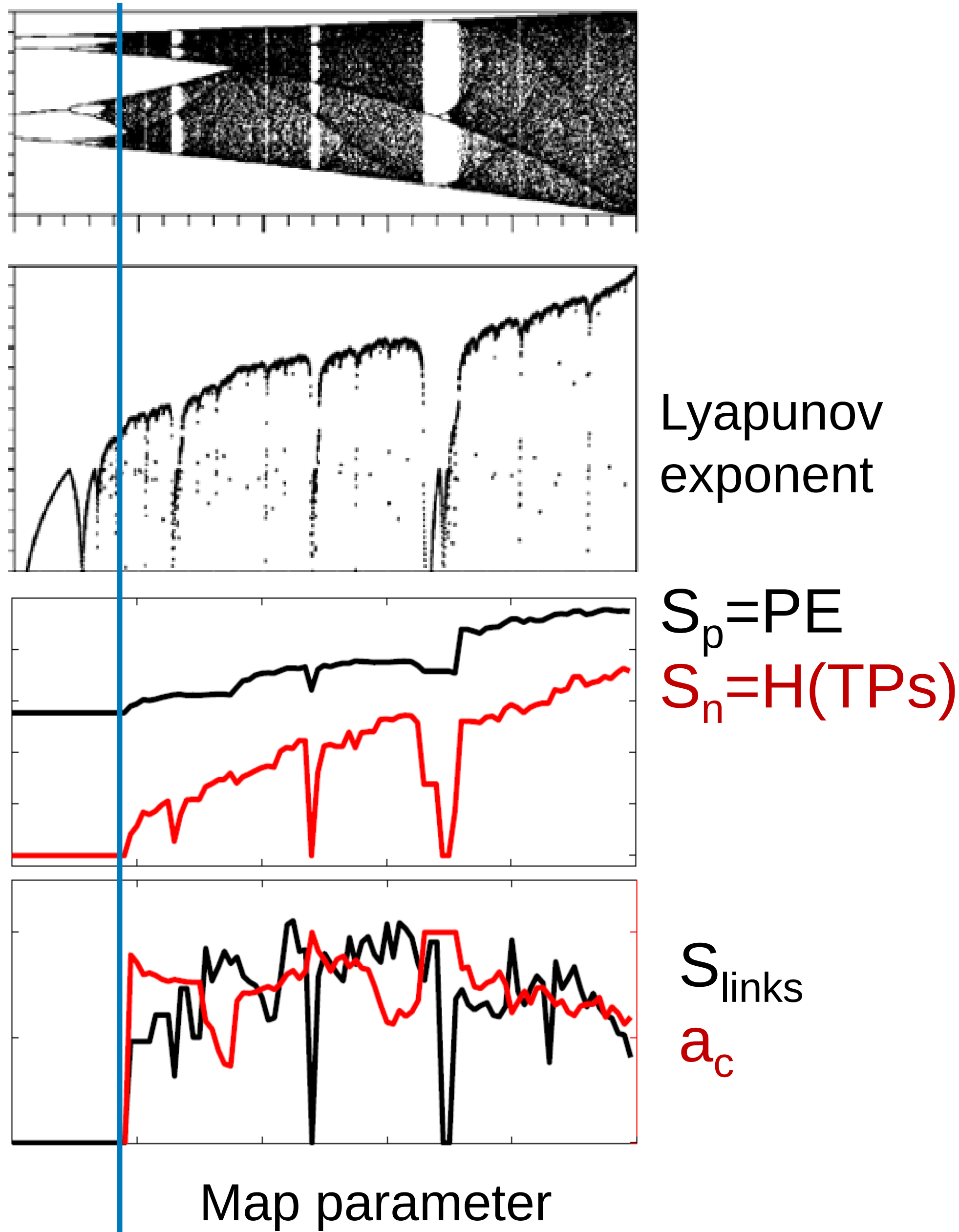
- Asymmetry coefficient: normalized difference of transition probabilities, P('01'→ '10') - P('10'→ '01'), etc.

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

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

cristina.masoller@upc.edu @cristinamasoll1

Example: ordinal analysis of the Logistic map with patterns of length D=4

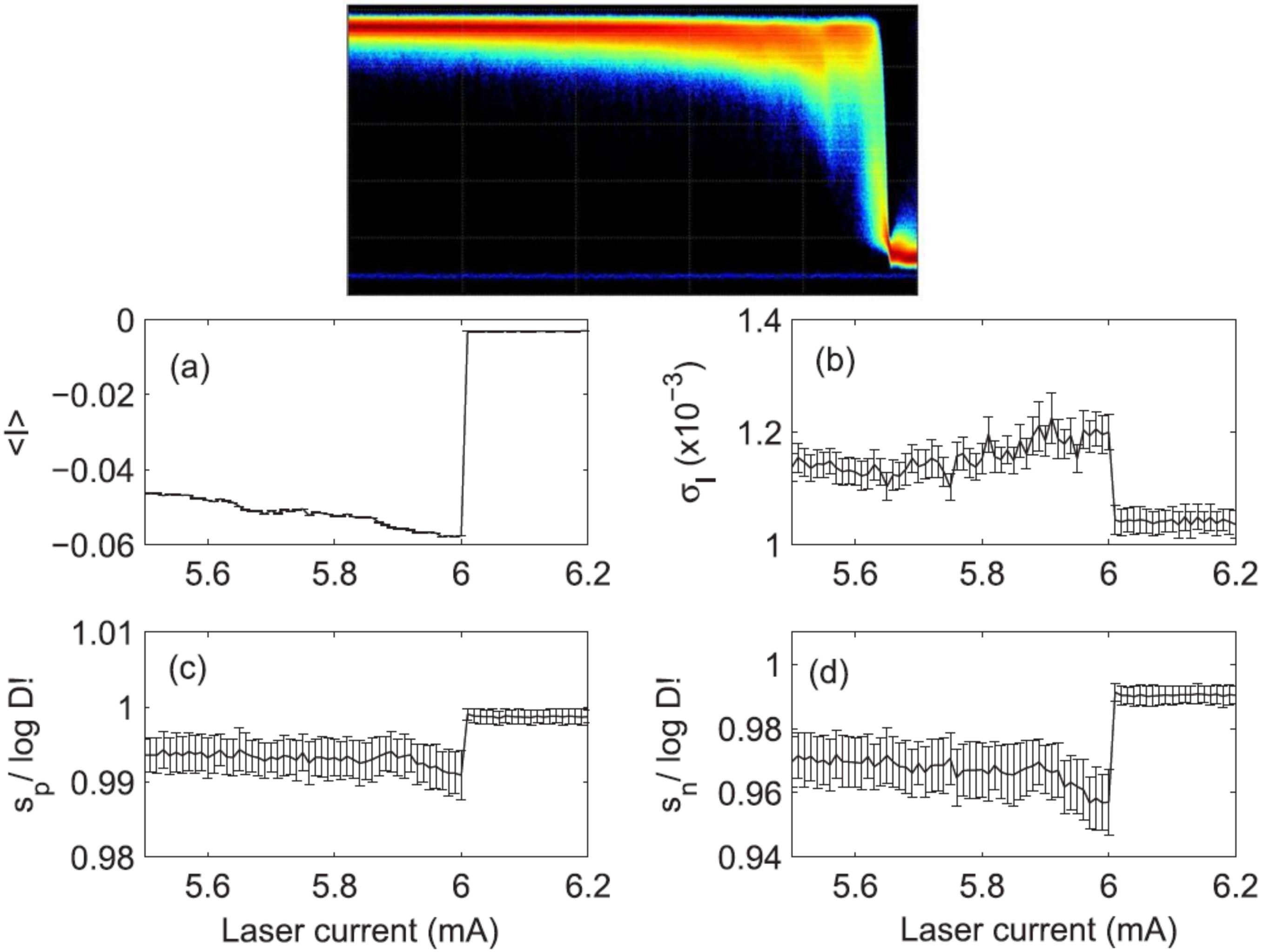


Both entropies detect the merging of four branches, which is not a "bifurcation" and is not detected by the Lyapunov exponent.

C. Masoller et al, New J. of Phys. 17, 023068 (2015)

cristina.masoller@upc.edu @cristinamasoll1

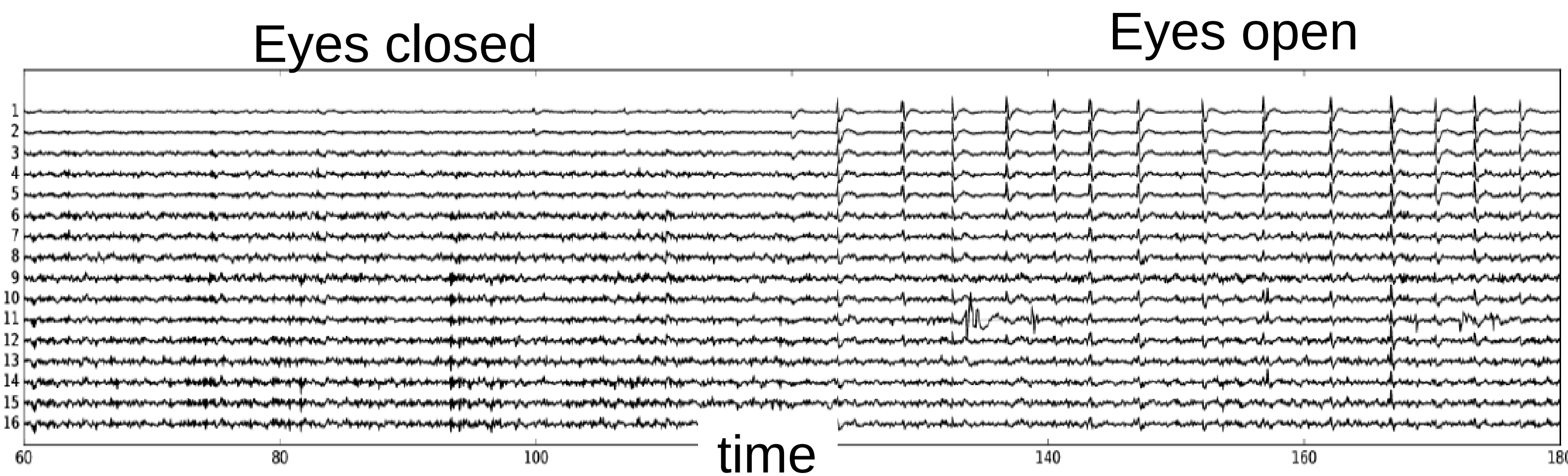
The entropies can anticipate an abrupt transition

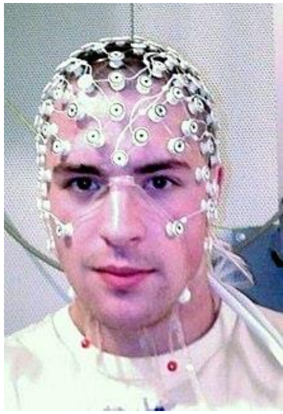


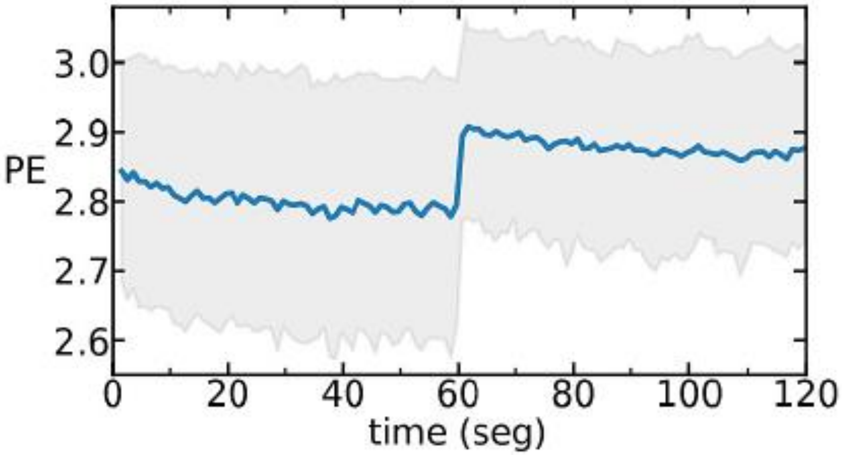
C. Masoller et al, New J. of Phys. 17, 023068 (2015)

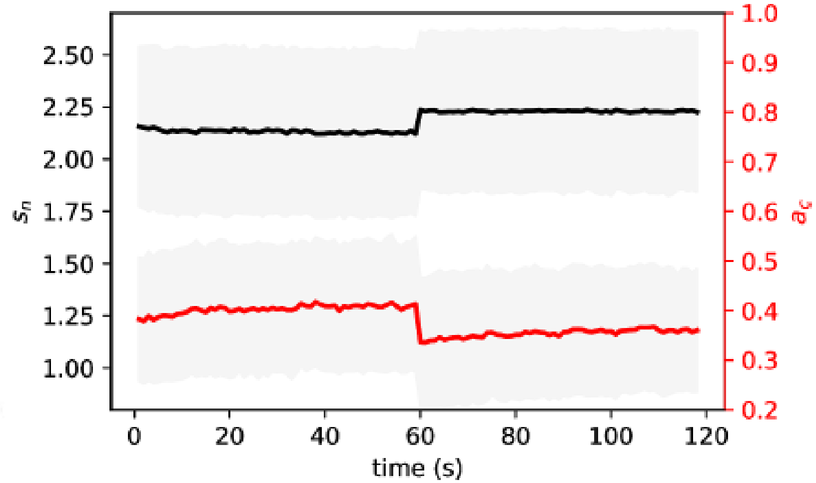
cristina.masoller@upc.edu @cristinamasoll1

Application to EEG recordings during eyes closed- eyes open transition (physionet dataset, 160 healthy subjects)





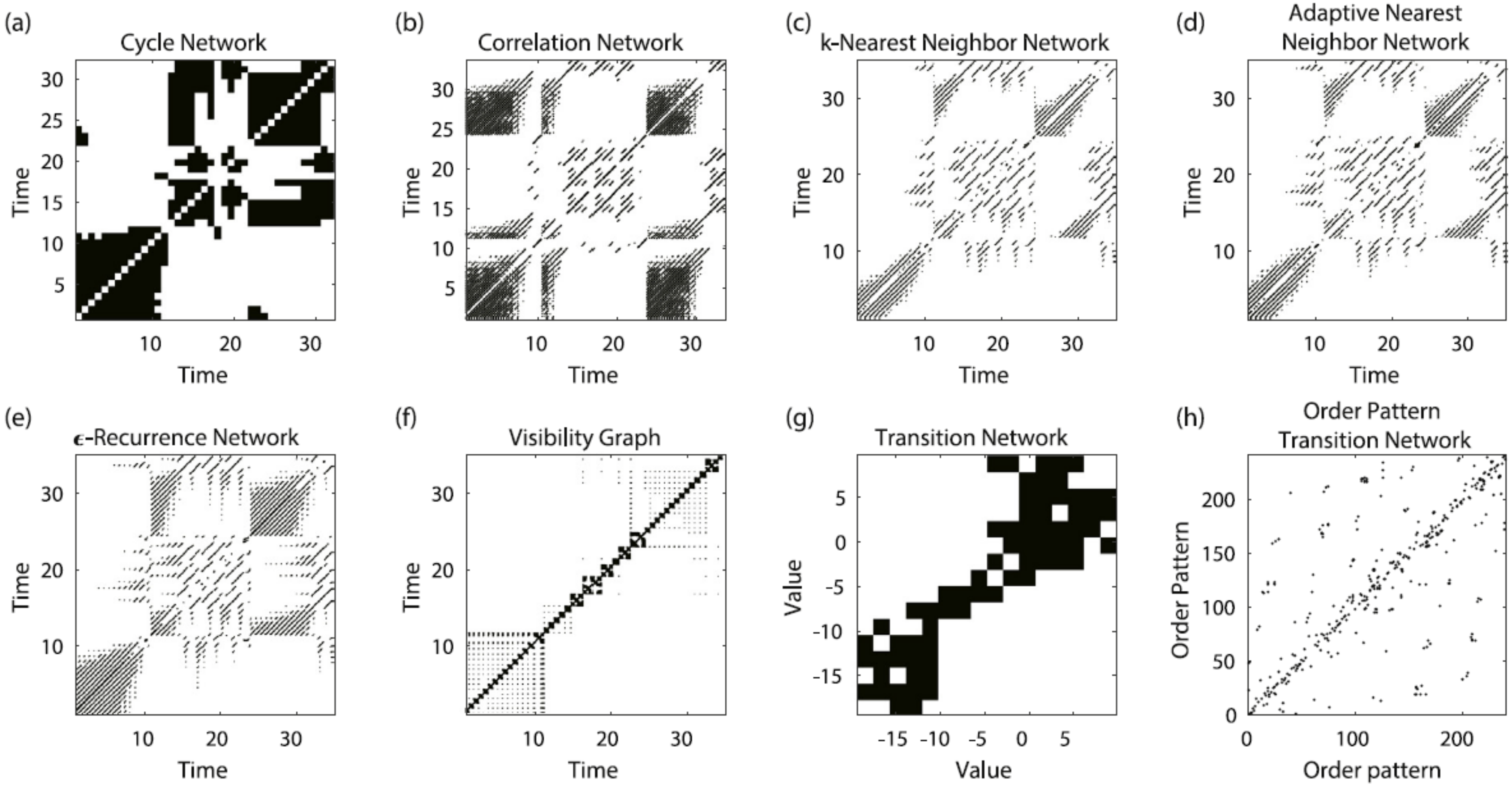




C. Quintero-Quiroz et al., Chaos 28, 106307 (2018).

cristina.masoller@upc.edu @cristinamasoll1

There are many other ways to represent a time-series as a network.



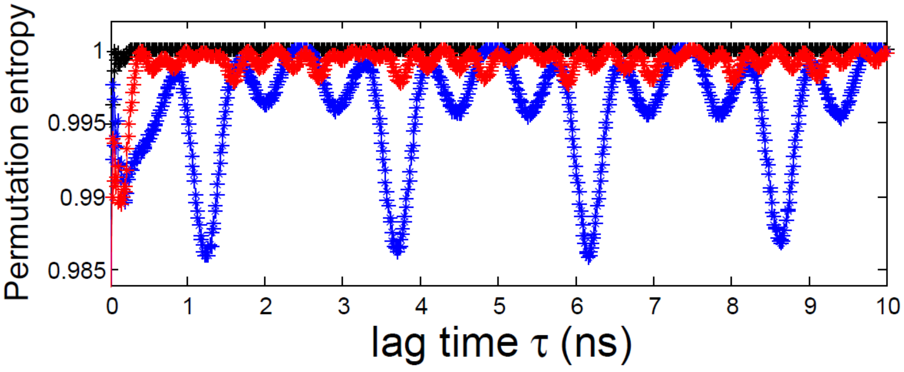
<u>Adjacency matrices</u> corresponding to different types of networks constructed from the x-coordinate of the Lorenz system.

Y. Zou et al, Physics Reports 787, 1 (2019)

cristina.masoller@upc.edu @cristinamasoll1

Methods of univariate time series analysis

- Return maps
- Distribution of data values
- Autocorrelation and Fourier analysis
- Surrogates
- Attractor reconstruction, Lyapunov exponents, and fractal dimension
- Symbolic analysis
- Information theory measure: entropy
- Network representation of a time-series
- Spatio-temporal representation of a time-series



cristina.masoller@upc.edu @cristinamasoll1

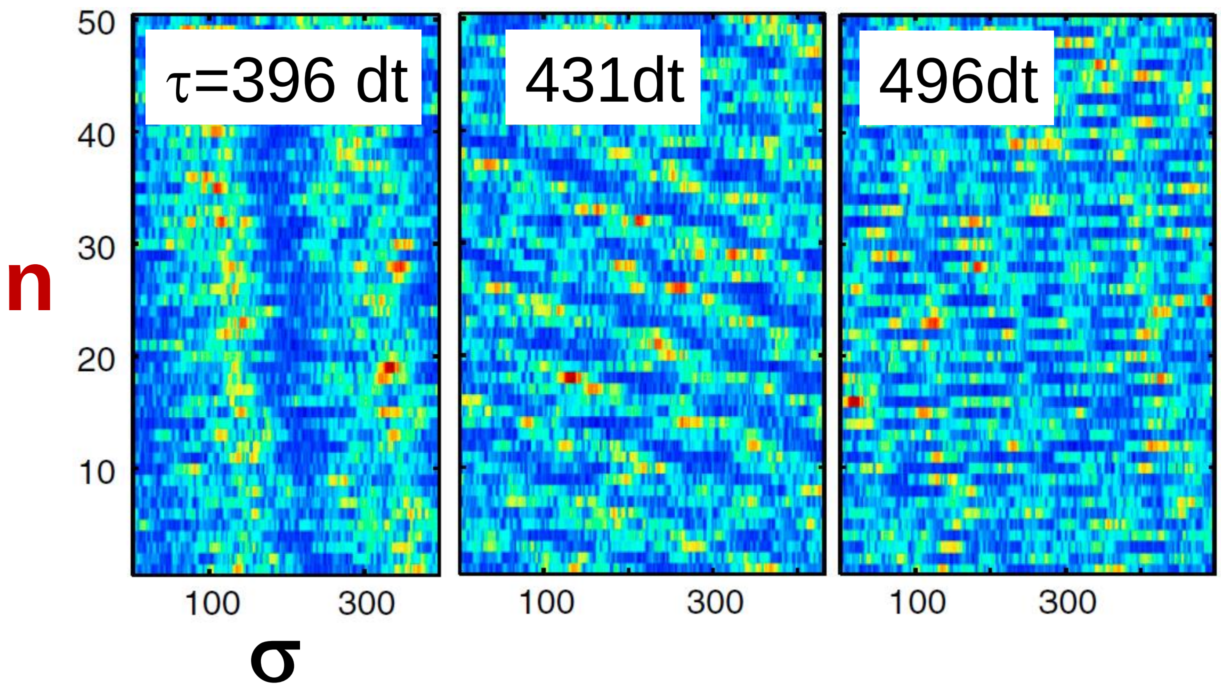
The space-time representation of a time series: a convenient way to visualize the dynamics

$$\{I_1, I_2, \ldots I_\tau, I_{\tau+1}, \ldots\} \Rightarrow \begin{bmatrix} \cdots & \cdots & \cdots & \cdots \\ I_{2\tau+1} & I_{2\tau+2} & \cdots & I_{3\tau} \\ I_{\tau+1} & I_{\tau+2} & \cdots & I_{2\tau} \\ I_1 & I_2 & \cdots & I_\tau \end{bmatrix}$$

$\sigma \longrightarrow$

Color scale: I_i

↑ n



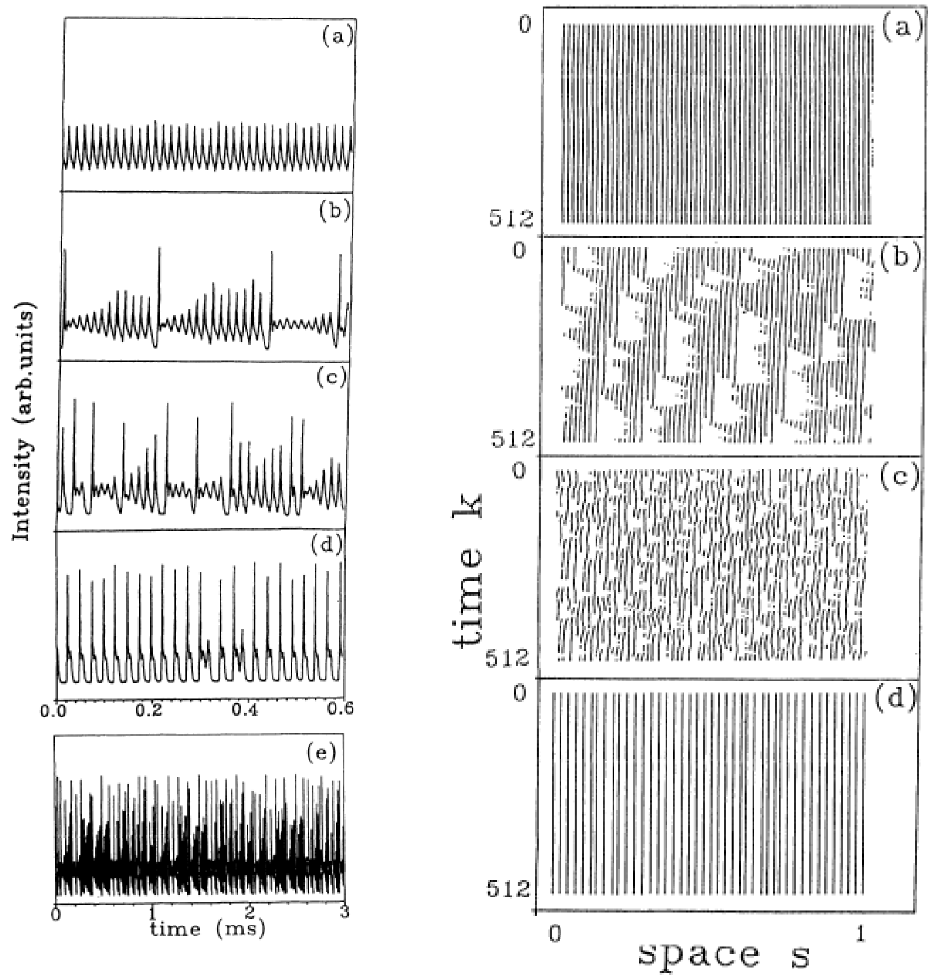
cristina.masoller@upc.edu @cristinamasoll1

RAPID COMMUNICATIONS

PHYSICAL REVIEW A VOLUME 45, NUMBER 7 1 APRIL 1992

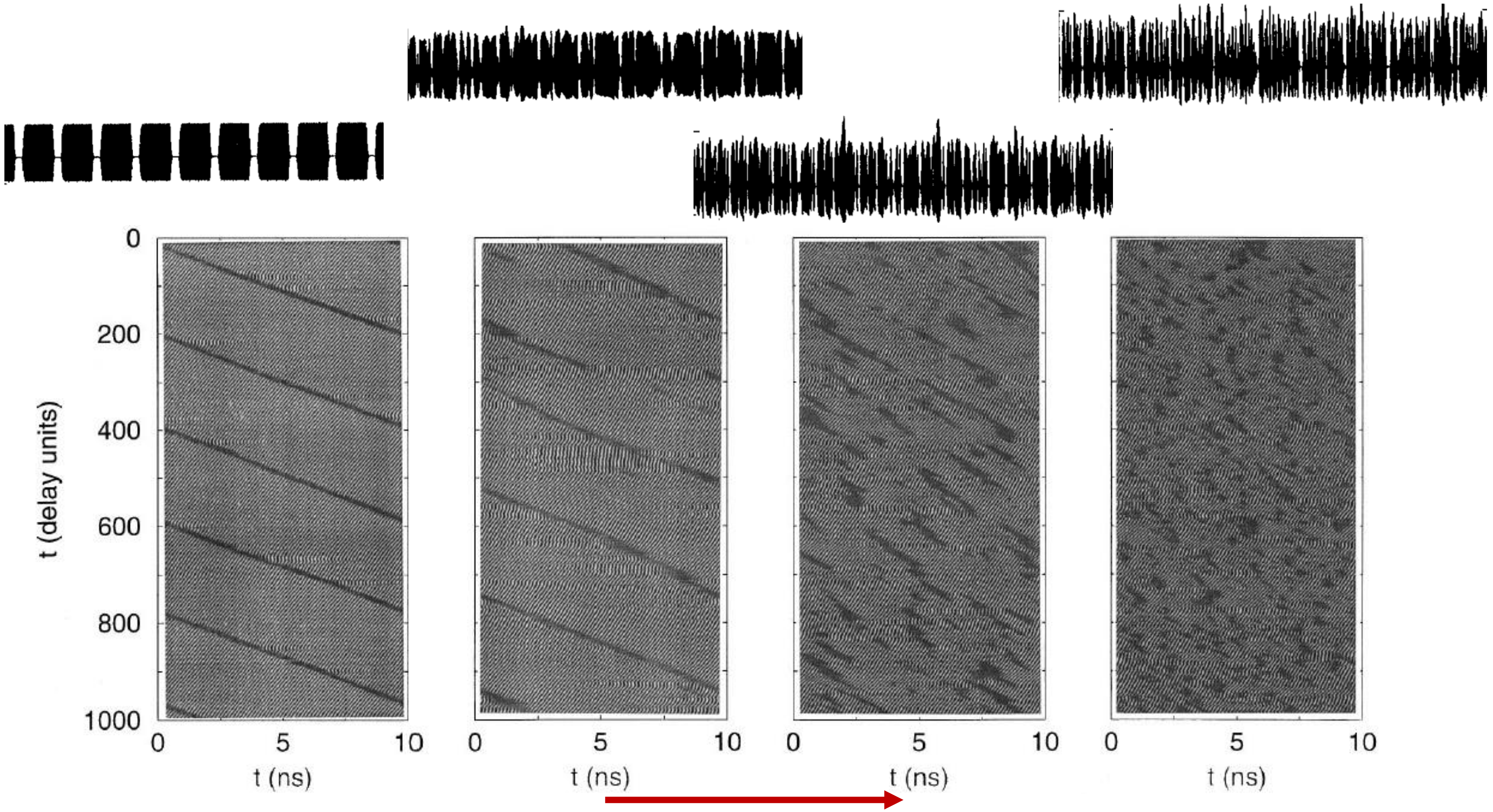
Two-dimensional representation of a delayed dynamical system

F. T. Arecchi,* G. Giacomelli, A. Lapucci, and R. Meucci
Istituto Nazionale di Ottica, Largo E. Fermi 6, 50125 Firenze, Italy
(Received 31 July 1991; revised manuscript received 10 December 1991)



82

Another example: spatio-temporal representation of quasi-periodic and chaotic time series



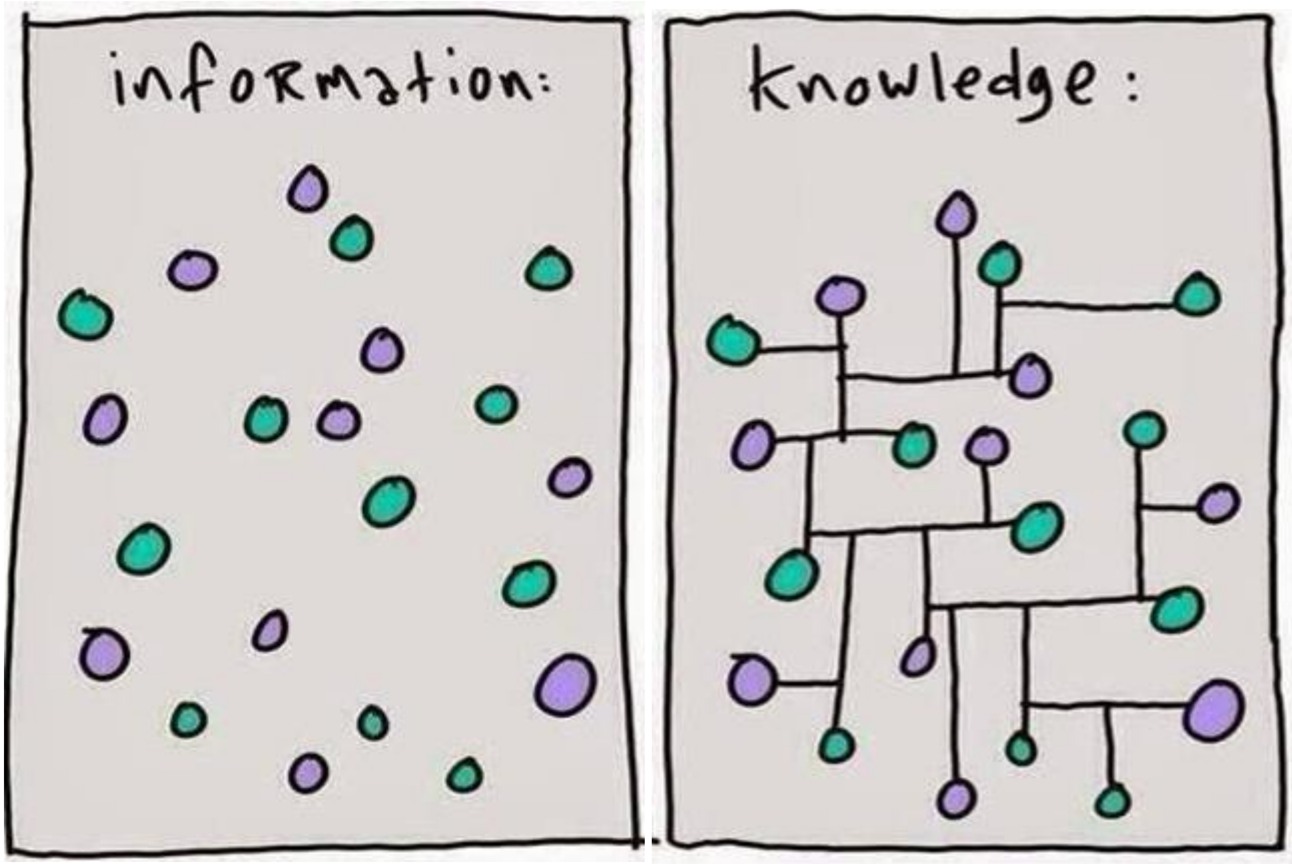
C. Masoller, Chaos 7, 455 (1997)

cristina.masoller@upc.edu @cristinamasoll1

And many more time series analysis methods

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

cristina.masoller@upc.edu @cristinamasoll1



cristina.masoller@upc.edu @cristinamasoll1

Time series analysis + complex systems ⇒ Big Data

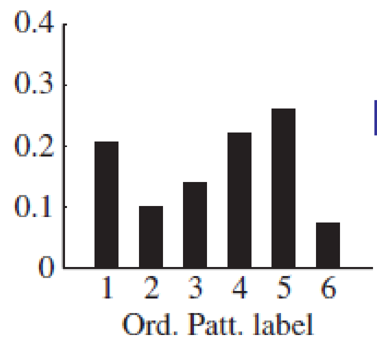
- For a given time series, by using different methods of analysis we obtain a large number of "features", M.
- Examples of "features":
  - Statistical properties (mean value, standard deviation, etc.),
  - Fourier properties (main frequencies),
  - Fractal dimension, Lyapunov exponent, etc. etc.
- If we have a large number of time series to analyze (N), we end up with a very large number of features (N x M).
- What to do?
  - Dimensionality reduction
  - Machine learning

cristina.masoller@upc.edu @cristinamasoll1

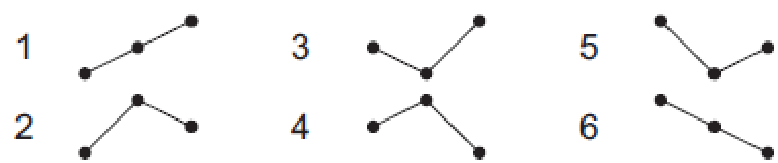
Permutation entropy: a form of "dimensionality reduction"

Time series

Ordinal probabilities



$$\sum_{i=1}^{N} p_i = 1$$

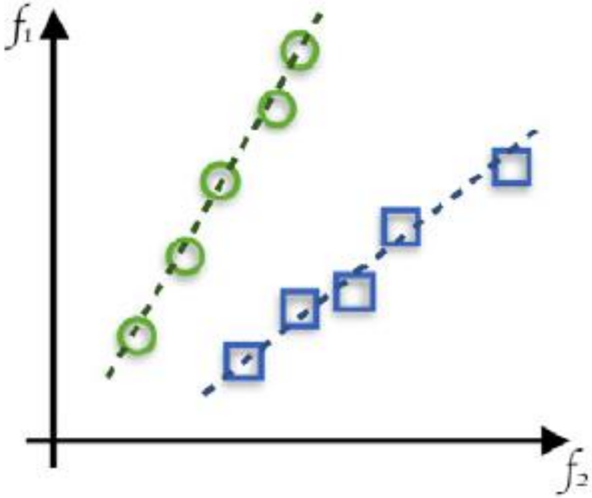


Permutation entropy

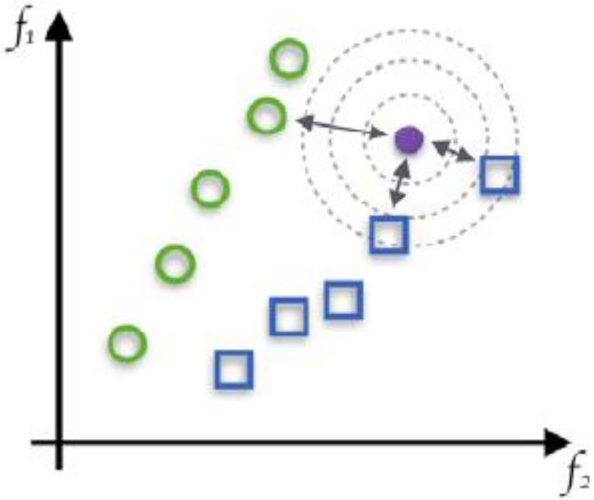
$$H = -\sum_{i=1}^{N} p_i \ln p_i$$

cristina.masoller@upc.edu @cristinamasoll1

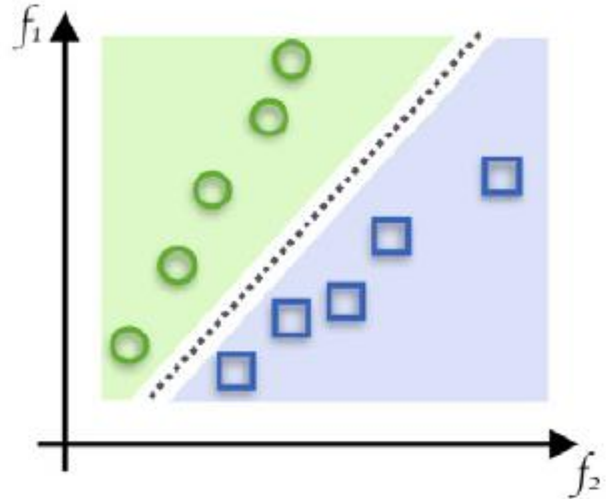
Main types of machine learning classification algorithms



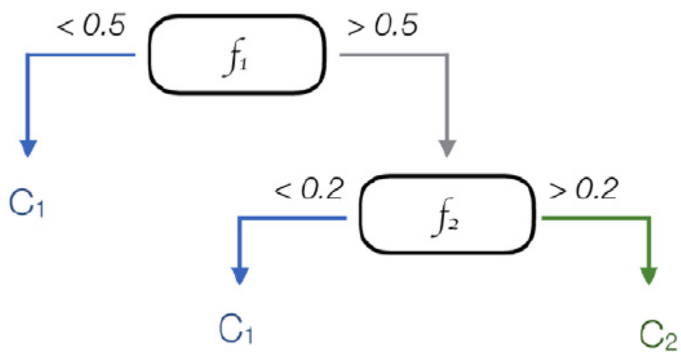
Regression



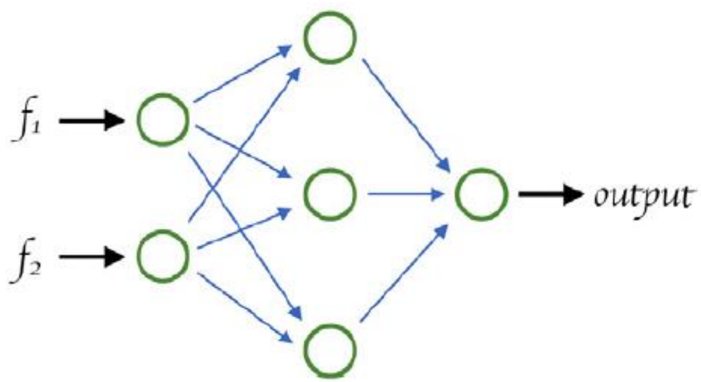
kNN



Support Vector Machine



Decision Tree



ANN

M. Zanin et al, Physics Reports 635, 1 (2016).

cristina.masoller@upc.edu @cristinamasoll1

Take home messages

- Linear methods and nonlinear methods are useful for investigating complex signals.
- Different methods provide *complementary* information (i.e., “features” that encapsulate different properties of a signal).
- Appropriate statistical “surrogate” tests are needed to determine if the numerical values are significant.

Holger Kantz: “*Every data set bears its own difficulties: data analysis is never routine*”

H. Kantz and T. Schreiber, Nonlinear Time Series Analysis, Cambridge University Press (2004)

Software implementation: https://www.pks.mpg.de/tisean/

G. Datseris and U. Parlitz, Nonlinear dynamics: a concise introduction interlaced with code, Springer (2022).

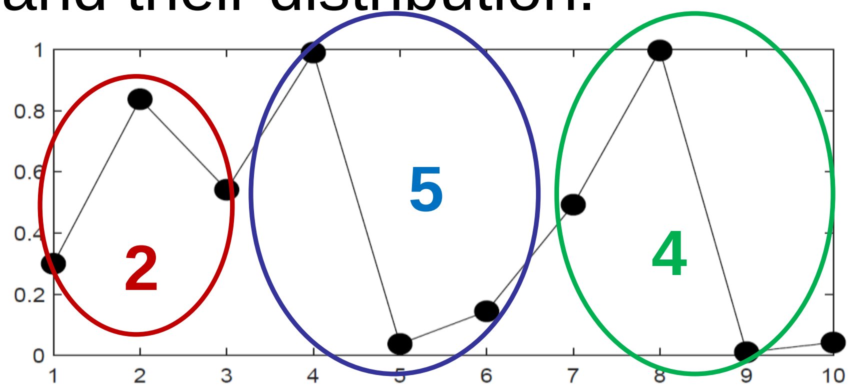
http://www.pik-potsdam.de/~donges/pyunicorn/

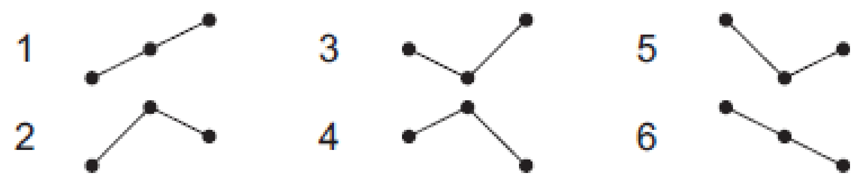
cristina.masoller@upc.edu **@cristinamasoll1**

Hands-on exercise 3: Ordinal analysis of the logistic map

(you can also use your own data)

1. Test the “ordinal pattern” program with some examples.
2. Calculate the probabilities of the 6 ordinal patterns of length D=3 for the logistic map with r=3.99.
3. Calculate the ordinal bifurcation diagram: r in (3.5,4) with D=3.
4. For r=3.99 generate two trajectories starting from very similar initial conditions and calculate the sequence of ordinal patterns and their distribution.





cristina.masoller@upc.edu @cristinamasoll1