Lecture 1: Introduction to nonlinear time series analysis and ordinal analysis

Prof. Cristina Masoller

Universitat Politècnica de Catalunya

cristina.masoller@upc.edu

www.fisica.edu.uy/~cris

6-th International Winter School on Data Analytics

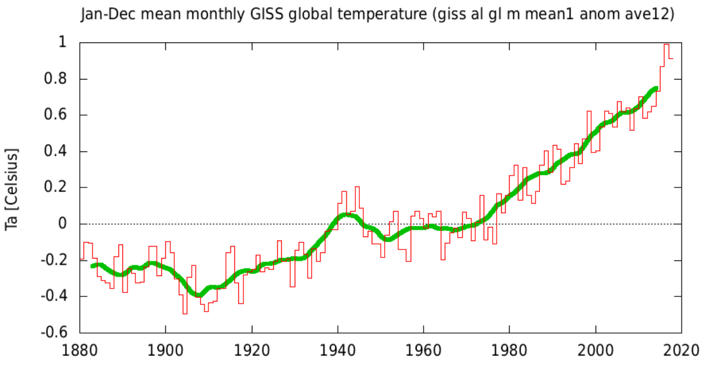
Nizhny Novgorod, October 21-22, 2021

Methods of time series analysis

- Return maps
- Distribution of data values
- Autocorrelation
- Statistical significance and surrogates
- Attractor reconstruction, Lyapunov exponents, and fractal dimension
- Symbolic methods
- Information theory measures: entropy and complexity

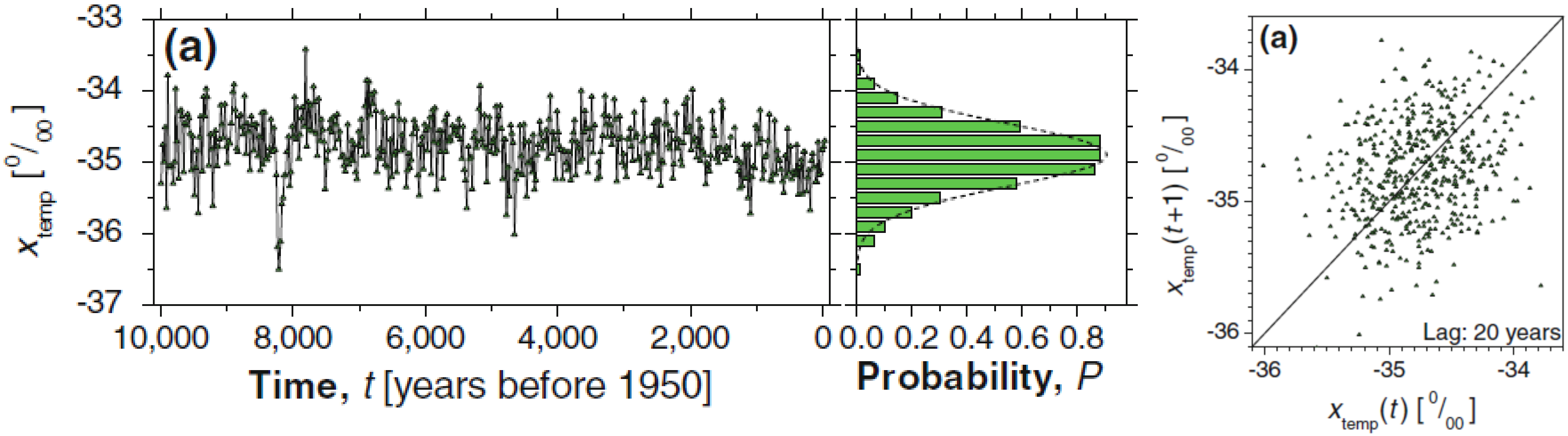
To begin with the analysis of a time series

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



- 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

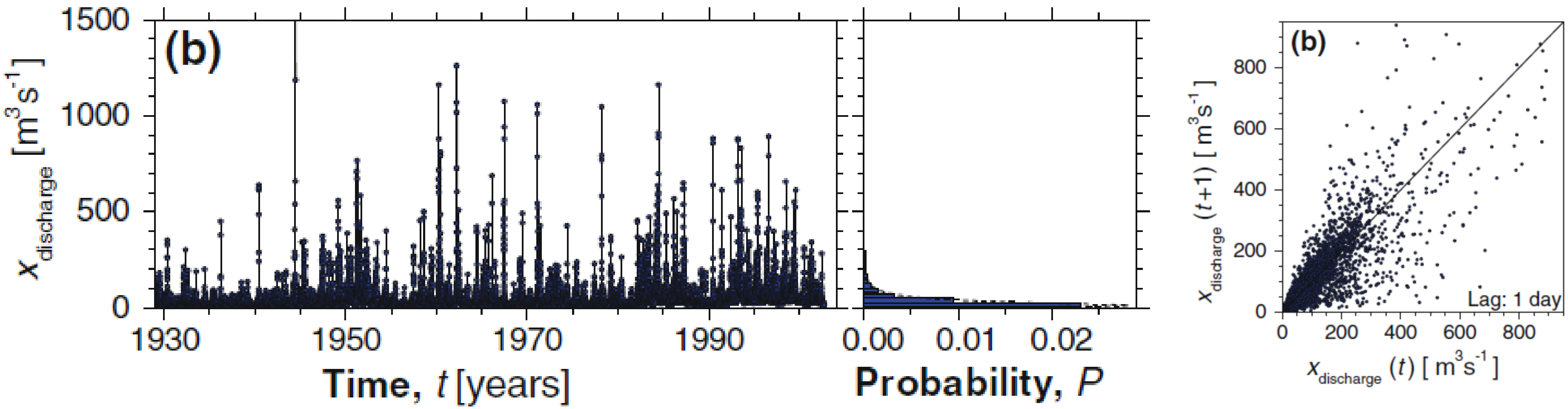
First example of a geophysical time series



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.

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

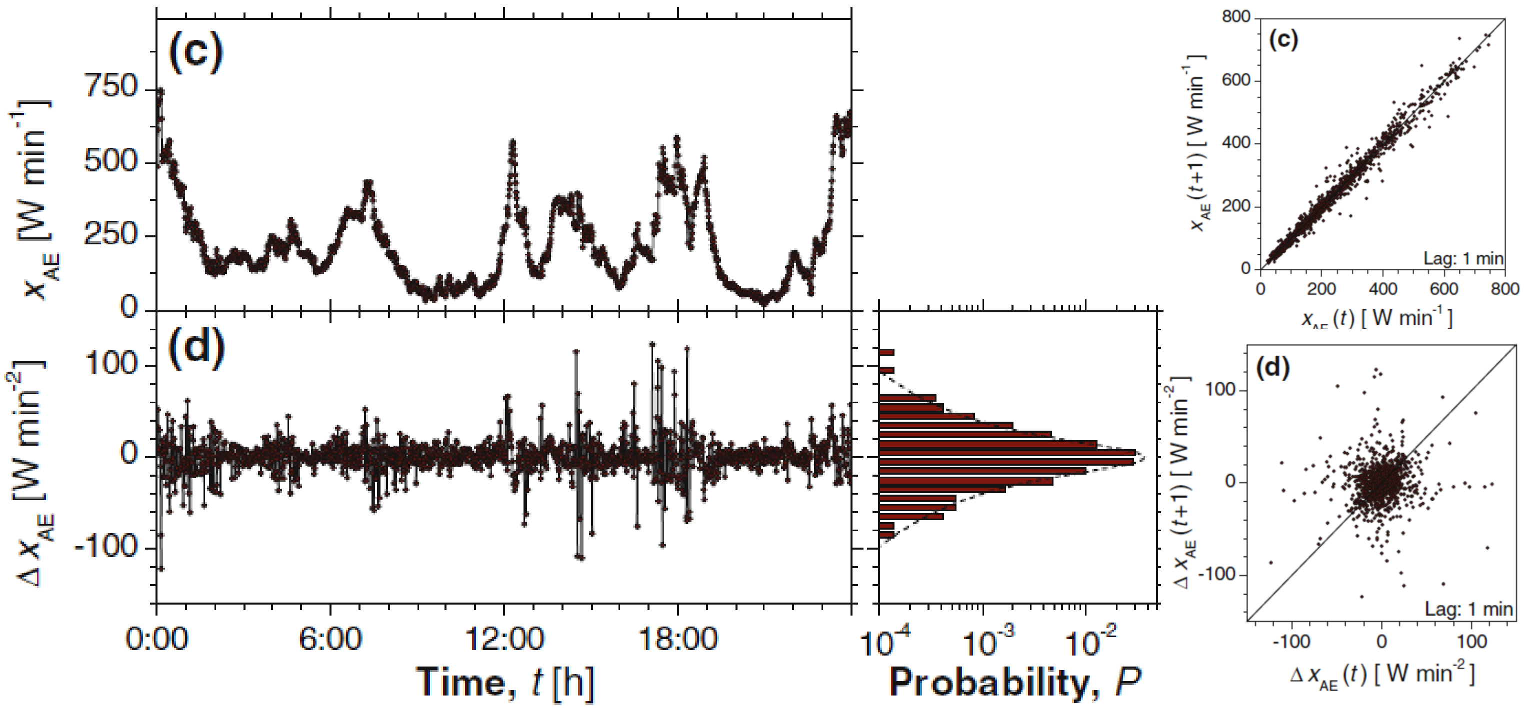
Second example



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

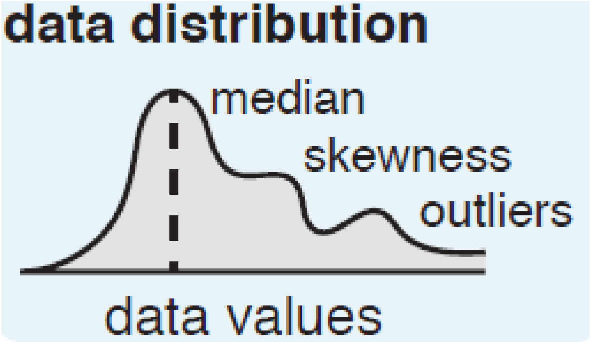
A. Witt and B. D. Malamud, Surv. Geophys. 34, 541 (2013).

Third example



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

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



How to characterize the distribution of data values?

- **Mean** (expected value of *X*): $\mu = \langle x(t) \rangle$

$$\mathrm{E}[X] = \sum_{i=1}^{k} x_i\, p_i = x_1 p_1 + x_2 p_2 + \cdots + x_k p_k$$

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

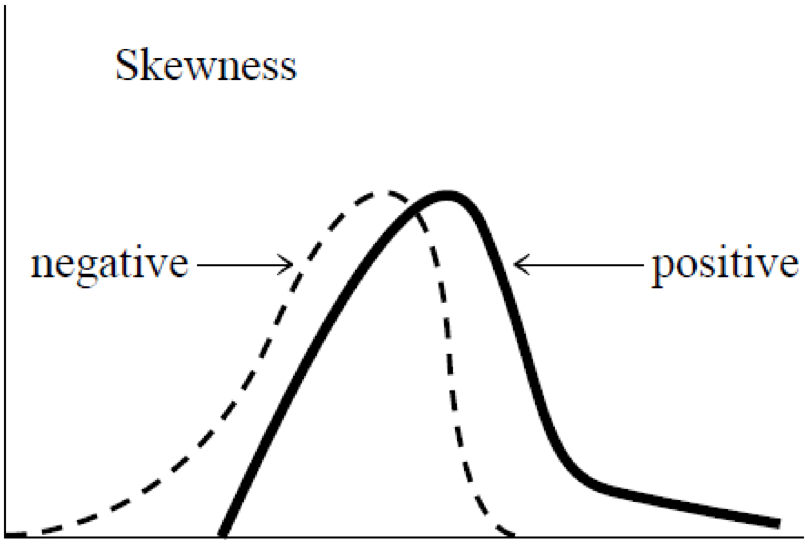
- **Skewness**: "measures" the asymmetry of the distribution

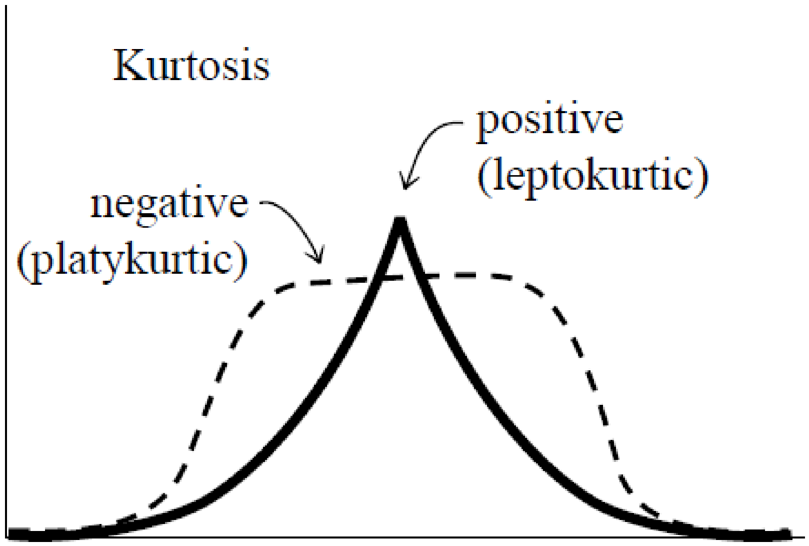
$$Z = \frac{X-\mu}{\sigma} \qquad S = \mathrm{E}[Z^3]$$

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

$$K = \mathrm{E}[Z^4]$$

- **Coefficient of variation**: normalized measure of the width of the distribution. $C_v = \sigma / |\mu|$

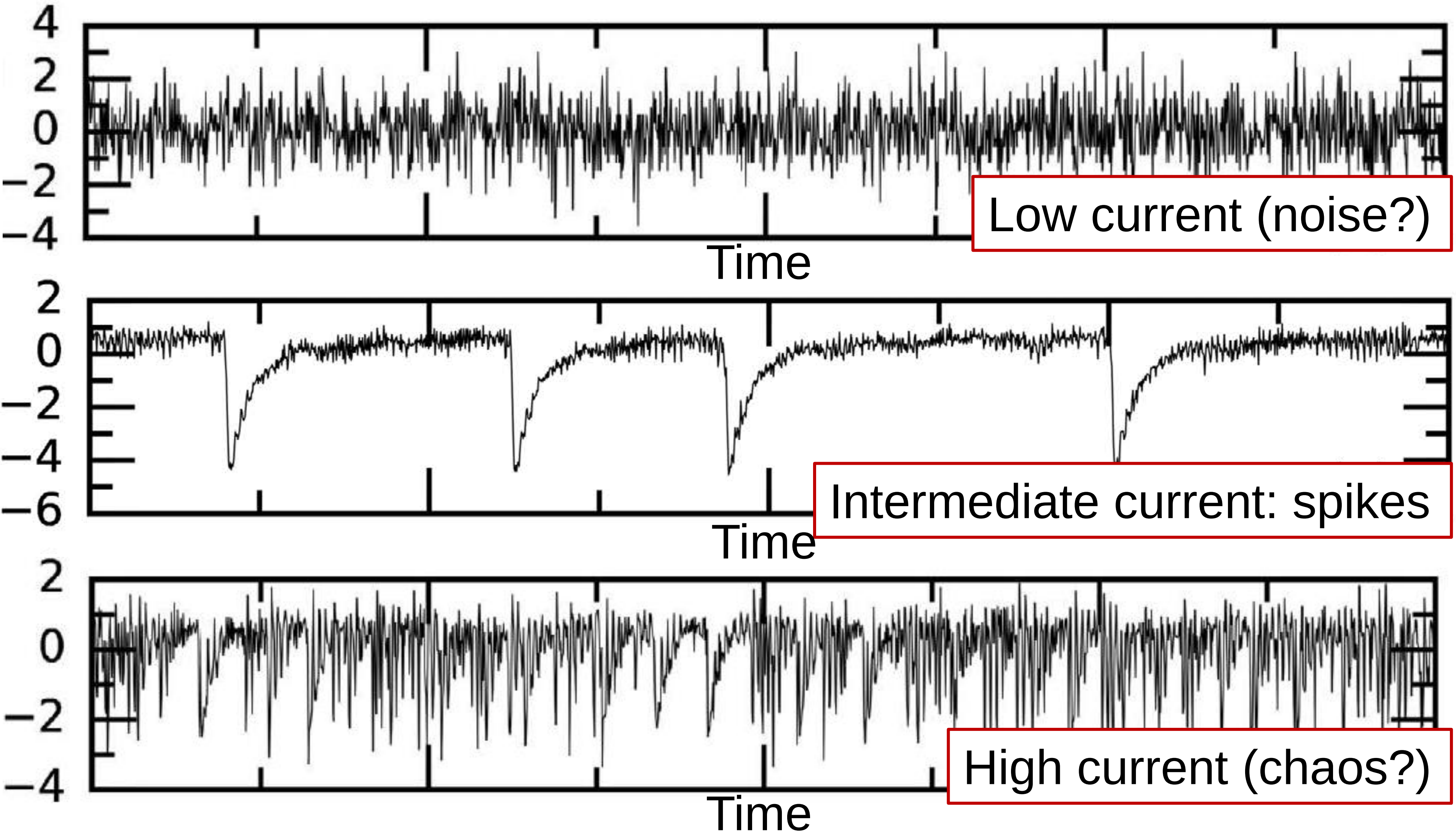




- **K<3:** the distribution produces fewer and less extreme outliers than the normal distribution. An example is the uniform distribution.
- **K=3:** Normal Gaussian
- **K>3:** the tail approaches zero more slowly than a Gaussian, and therefore produces more outliers than the normal distribution. An example is the Laplace distribution.

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

Example: intensity emitted by a diode laser with feedback, as the pump current increases (let see a video).

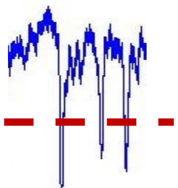


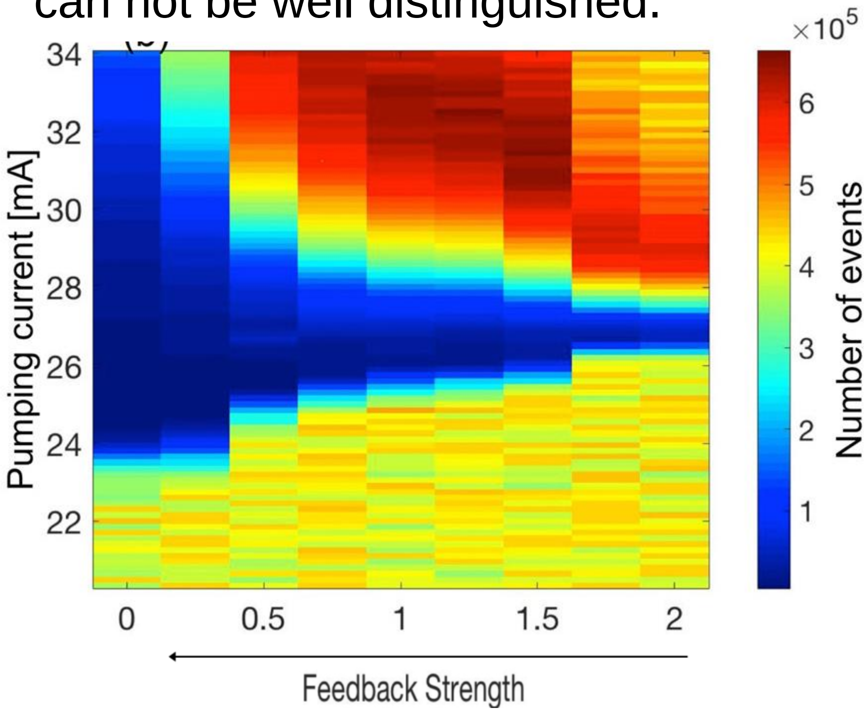
Video: how complex optical signals emerge from noisy fluctuations

Can we distinguish *quantitatively* the three regimes?

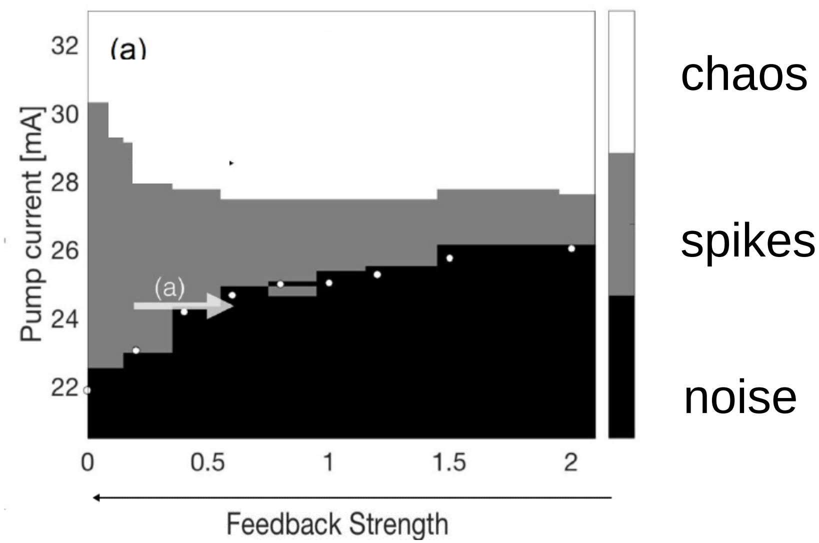
We recorded a large number of time series varying two experimental parameters: feedback strength and laser current.

First method: **Count** the number of "spikes" (# of times the intensity falls below a threshold).
Problem: chaos and noise can not be well distinguished.





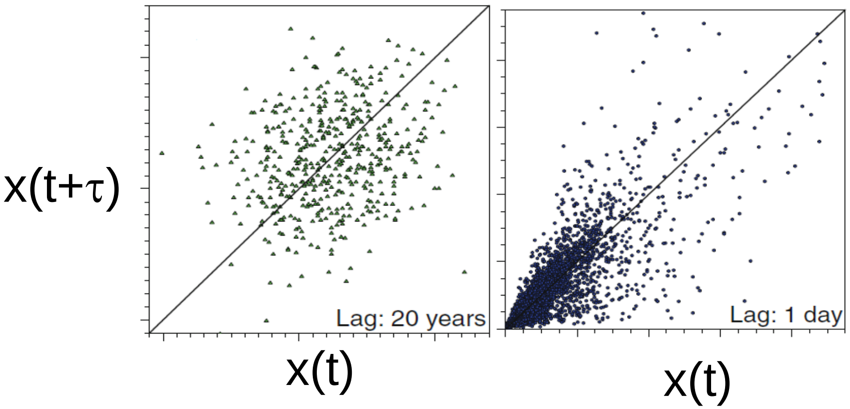
Second method: Analyze the distribution of intensity values
If K in 3-3.3 (Gaussian dist.) $\Rightarrow$ Noise
Else
If σ increases with the pump current (keeping feedback constant) $\Rightarrow$ Spikes
Else $\Rightarrow$ Chaos



[Panozzo et al, Chaos 27, 114315 (2017)]

Autocorrelation function (ACF)

For two geophysical time series

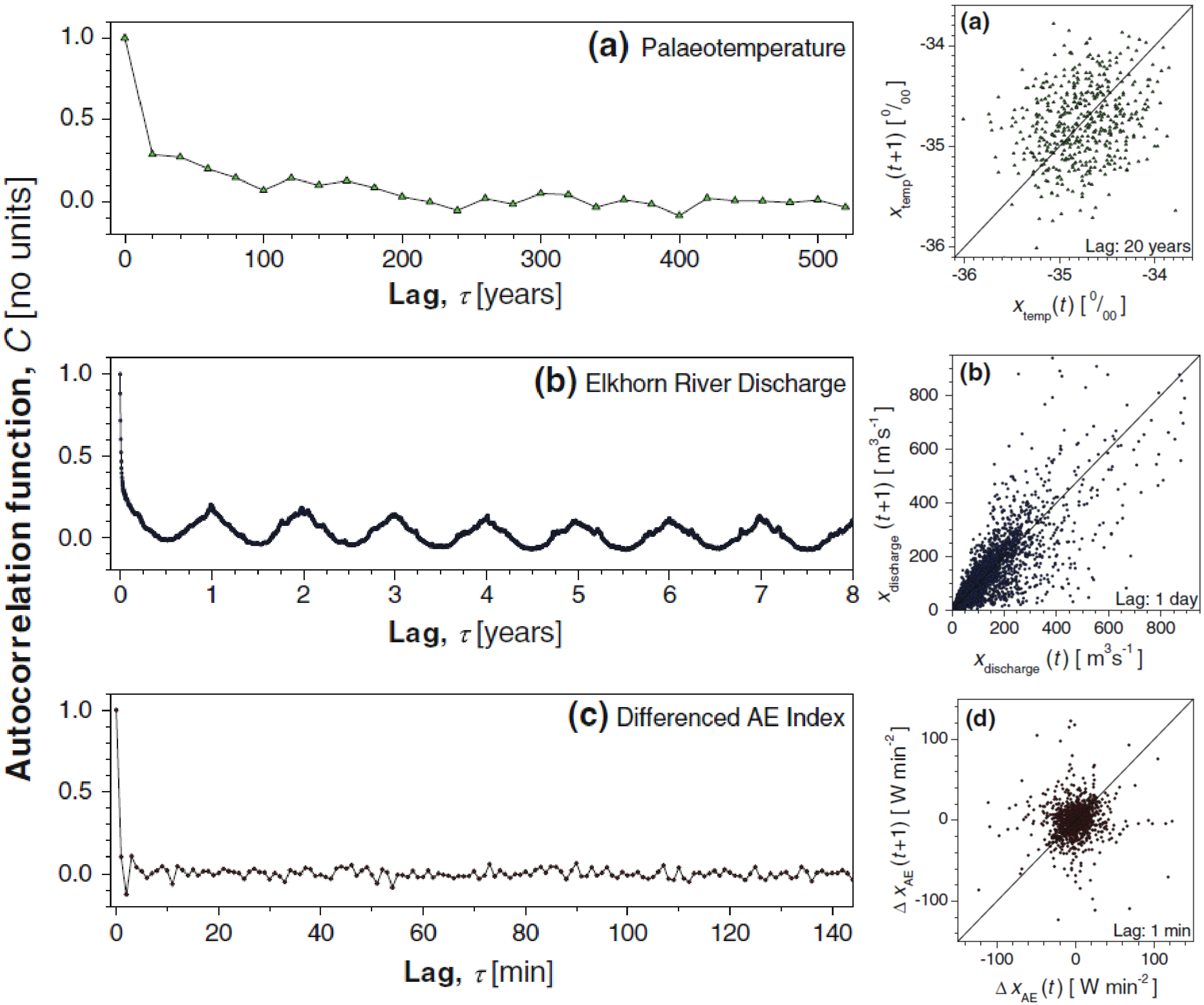


- The return map allows us to see if x(t) and x(t+τ) are "*correlated*".
- How to **quantify**?

- The autocorrelation function $C(\tau) = \dfrac{\langle [x(t)-\mu][x(t+\tau)-\mu] \rangle}{\sigma^2}$
- By definition: *C*(0)=1

- **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.

Back to the three examples of geophysical time series



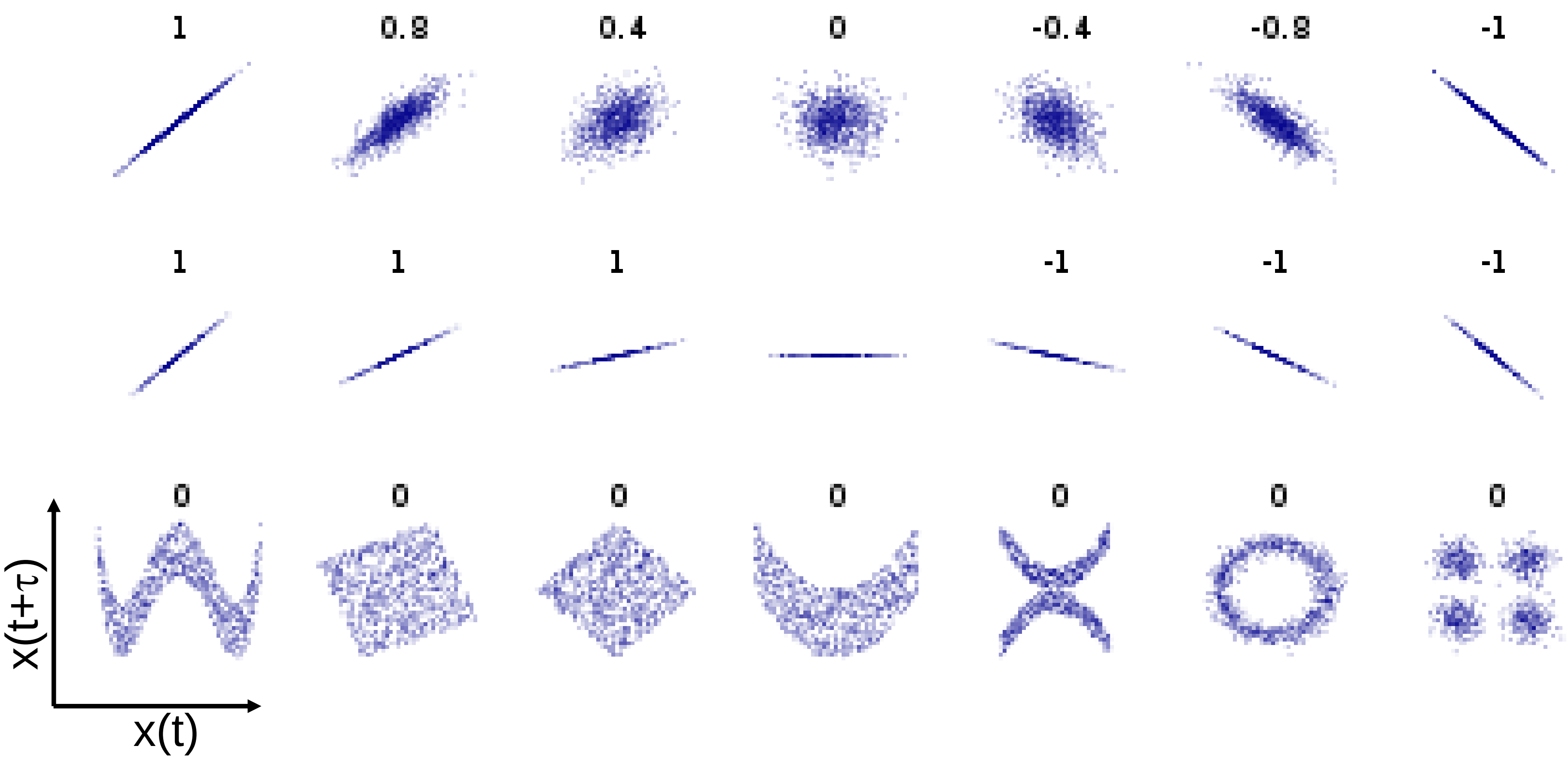
Slow decay:
long-range
correlations.

Rapid decay:
short-range
correlations.

$$|C(\tau)| \leq \kappa_0 \exp(-\kappa\tau)$$

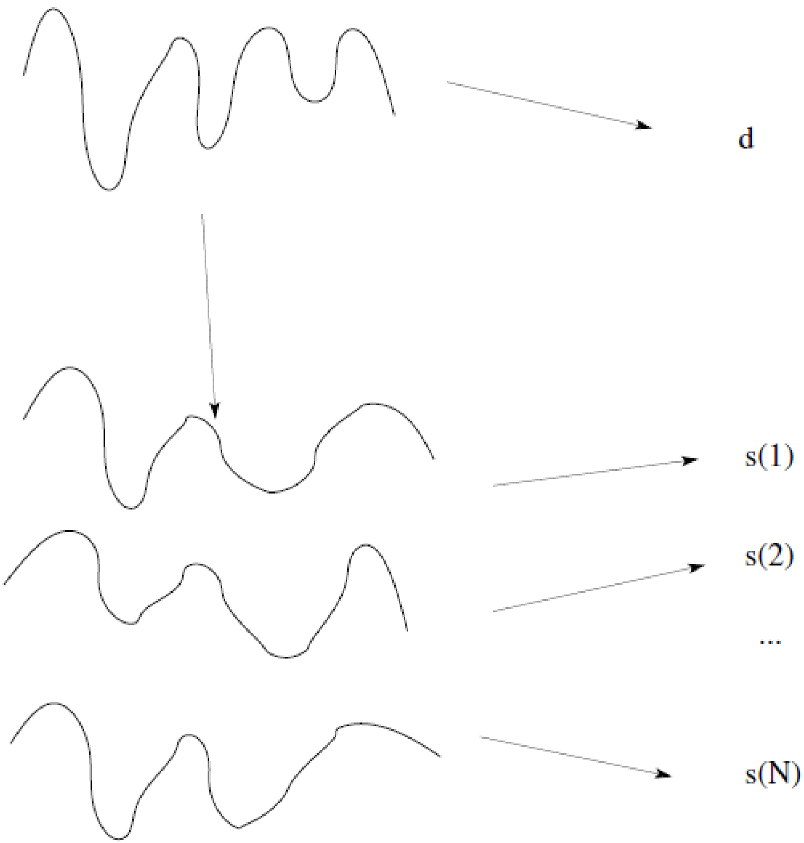
A. Witt and B. D. Malamud, Surv. Geophys. 34, 541 (2013).

Problem with the ACF: it only detects linear correlations between two data points ⇒ it is important to analyze nonlinear correlations and higher order correlations.



Source: wikipedia

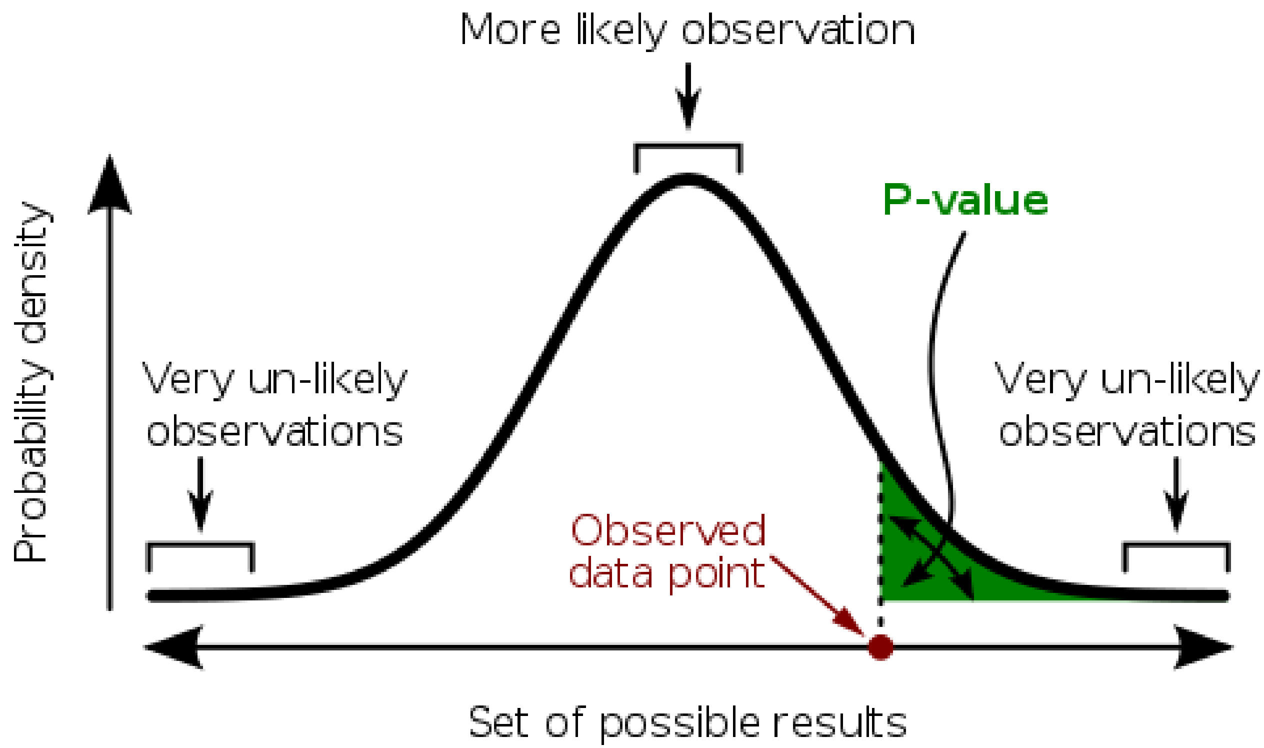
Significance analysis: the method of surrogate data



- 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.

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

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).

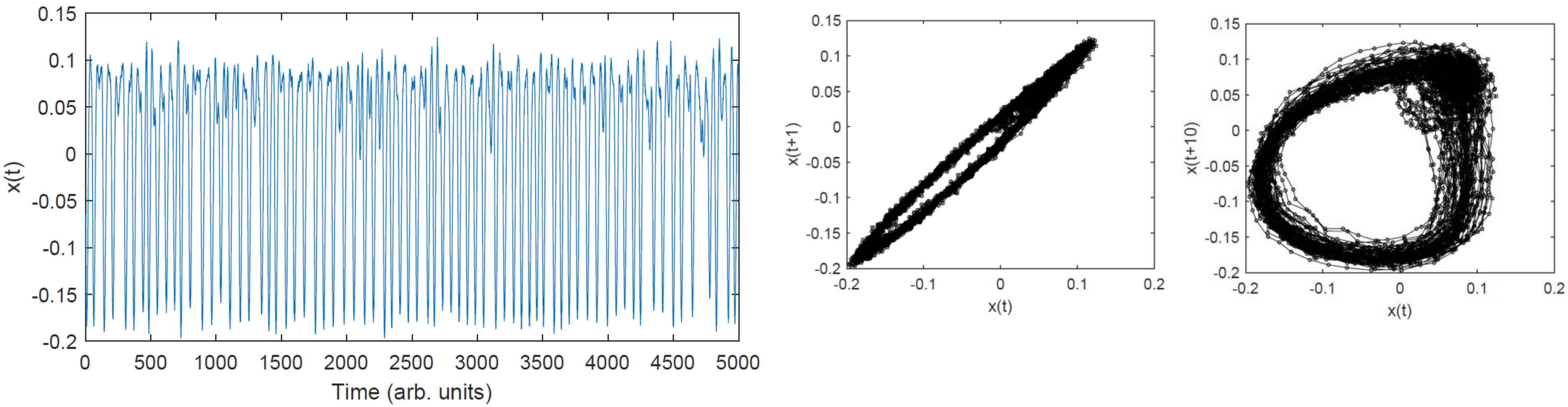
Methods of time series analysis

- Return maps
- Distribution of data values
- Autocorrelation
- Statistical significance and surrogates
- Attractor reconstruction, Lyapunov exponents, and fractal dimension
- 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

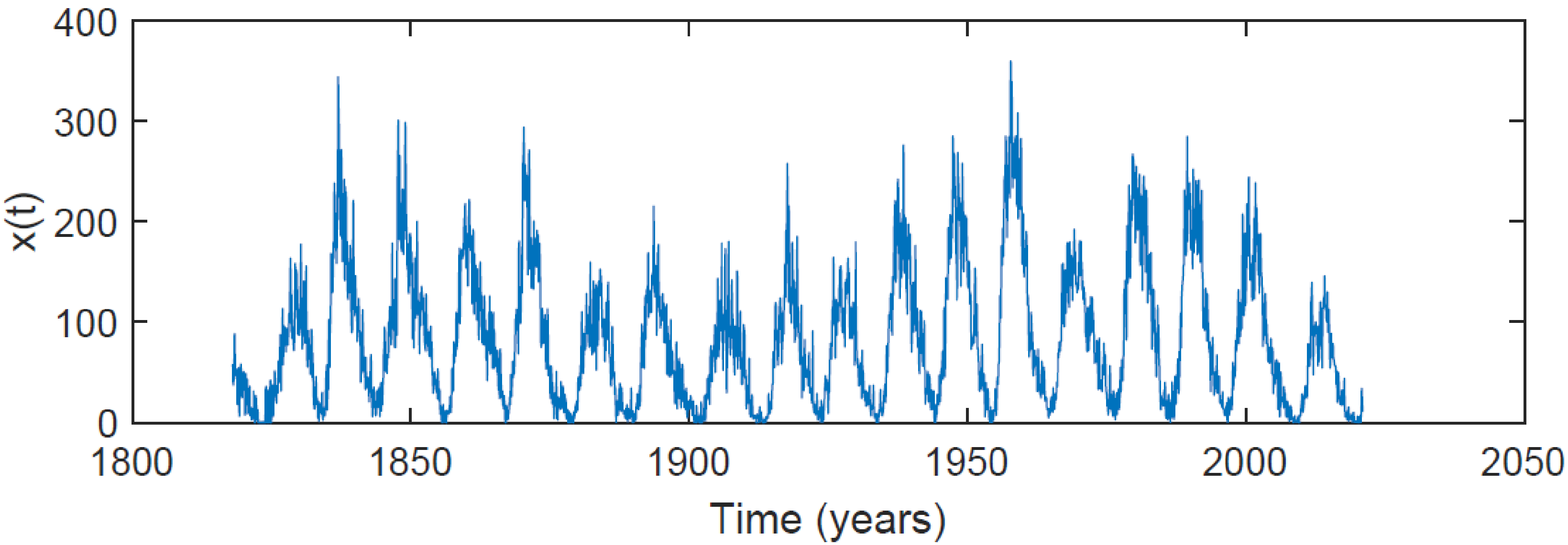
Why we want to reconstruct the phase space of a system from an observed (scalar) time series?

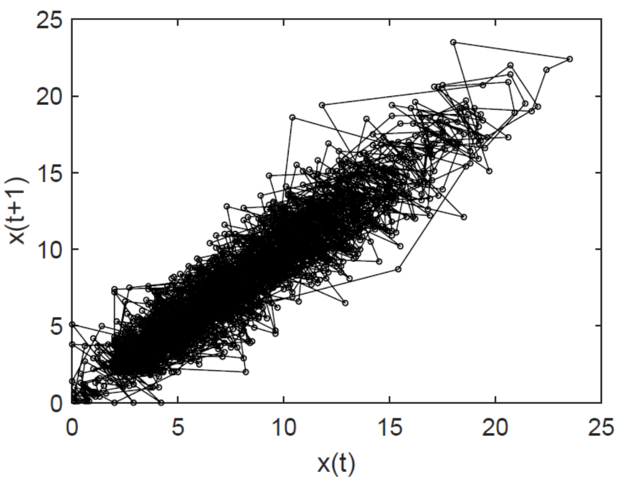
- Real systems are in general high-dimensional and we can only measure a few (hopefully relevant) variables.
- Models are too complex and have many parameters: reconstructing the phase space may allow to understand the effect of different parameters.

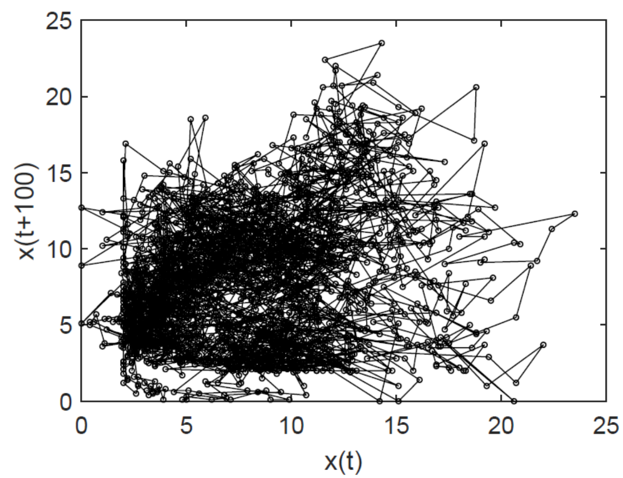
Example: the intensity emitted by a diode laser with optical feedback



A popular time series: monthly mean total sunspot number







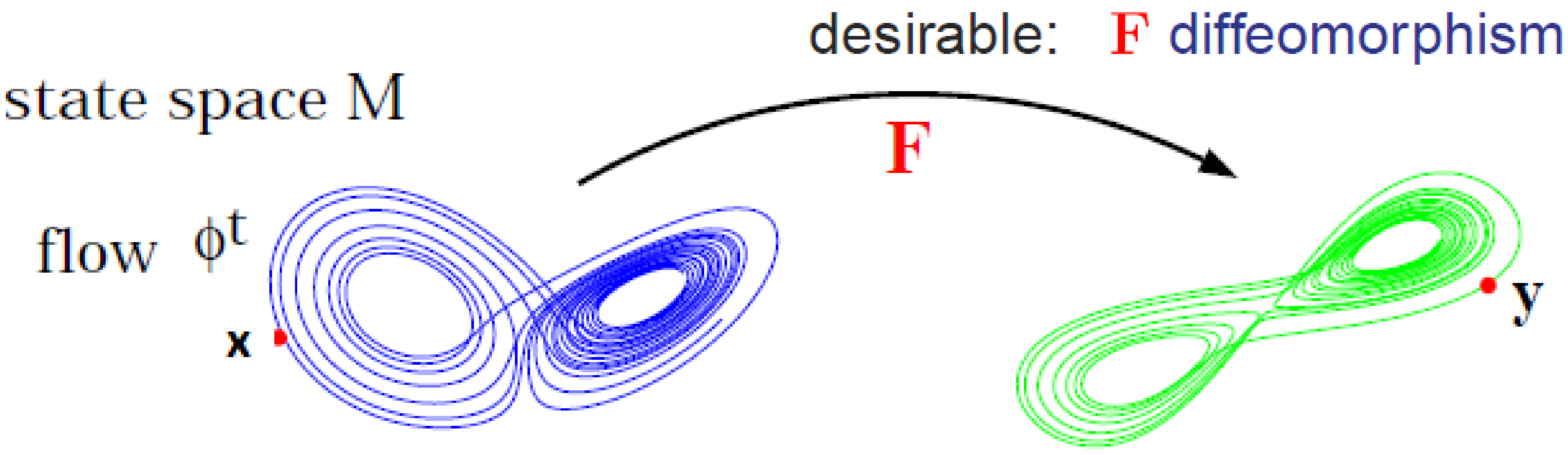
http://sidc.oma.be/silso/infosnmtot

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

Observed time series S = {s(1), s(2), … s(t) … }

Attractor reconstruction: "embed" the time series in a phase-space of dimension **d** using delay τ coordinates

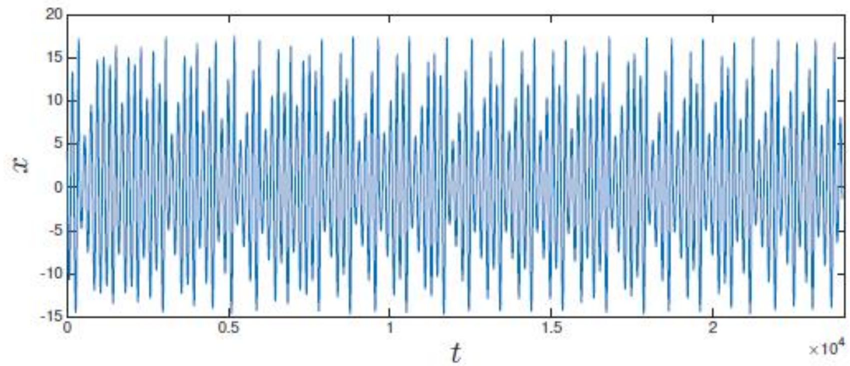
$$y(t) = \big(s(t), s(t+\tau), \ldots , s(t+(d-1)\tau)\big)$$



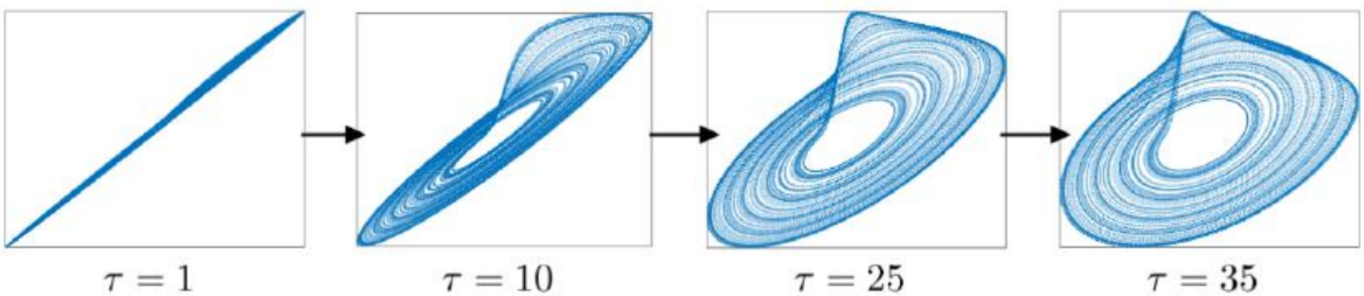
Adapted from U. Parlitz (Gottingen)

F.Takens, Lecture Notes in Mathematics 366 (1981)
Sauer et al., J. Stat. Phys. 65 (1991) 579

Reconstruction using delay coordinates



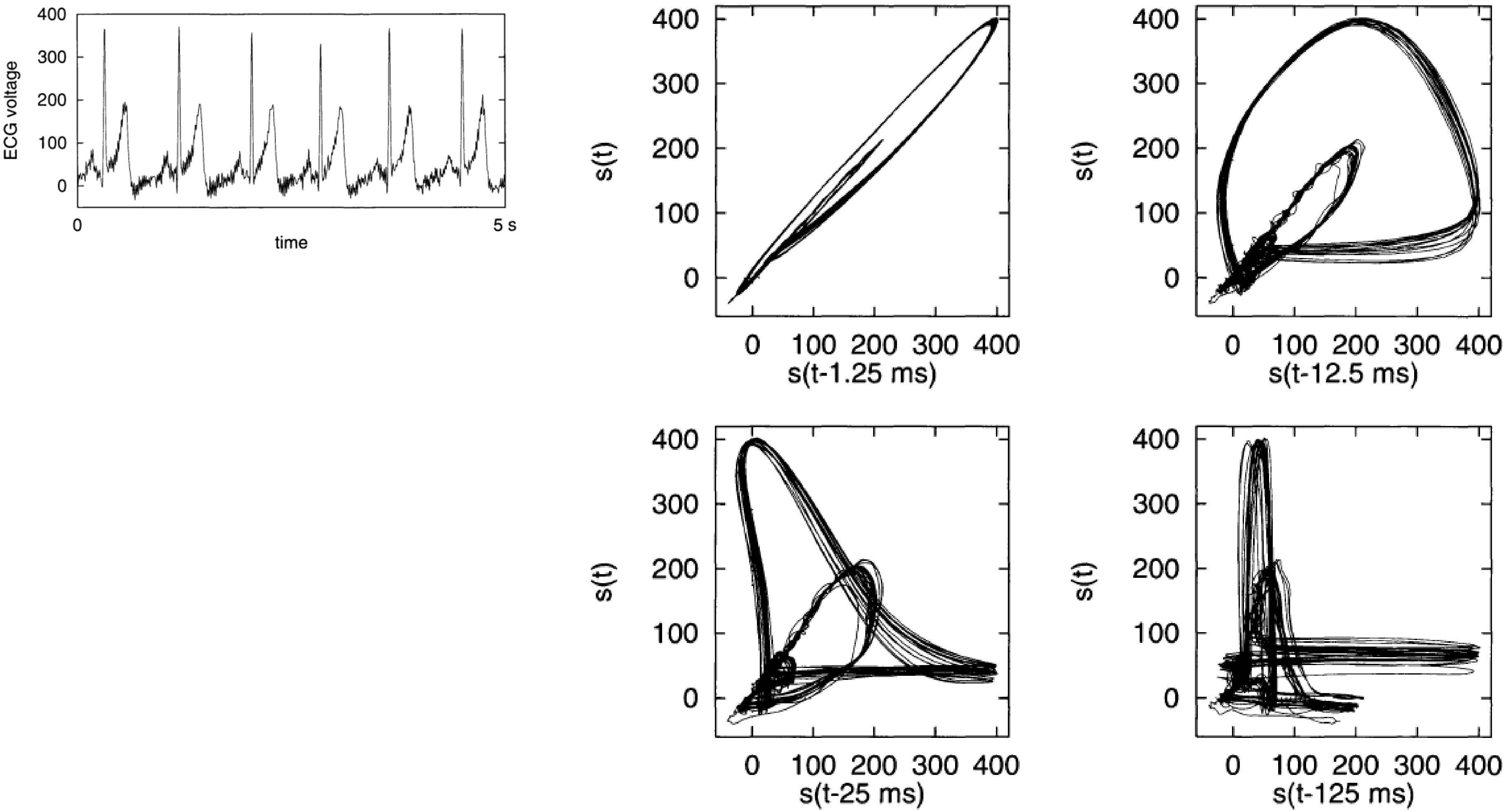
$$\mathbf{X}(t_i) = \begin{bmatrix} x(t_i) \\ x(t_i + \tau) \\ \vdots \\ x(t_i + (d-1)\tau) \end{bmatrix}$$



A problem: how to chose the embedding parameters (lag **τ**, dimension ***d***)

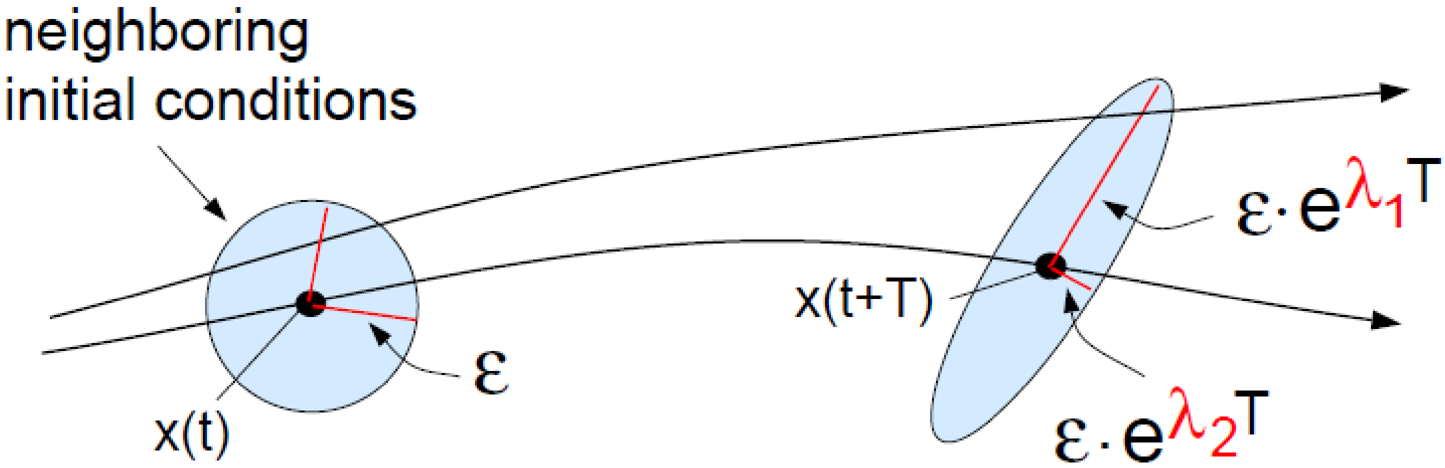
Bradley and Kantz, CHAOS 25, 097610 (2015)

Example: 2D representation of a human ECG signal



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

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.

Adapted from U. Parlitz (Gottingen)

Steps to compute the maximum LE

- 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 is averaged over the attractor, which 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)

On the interpretation of the maximum Lyapunov exponent: a word of warning!

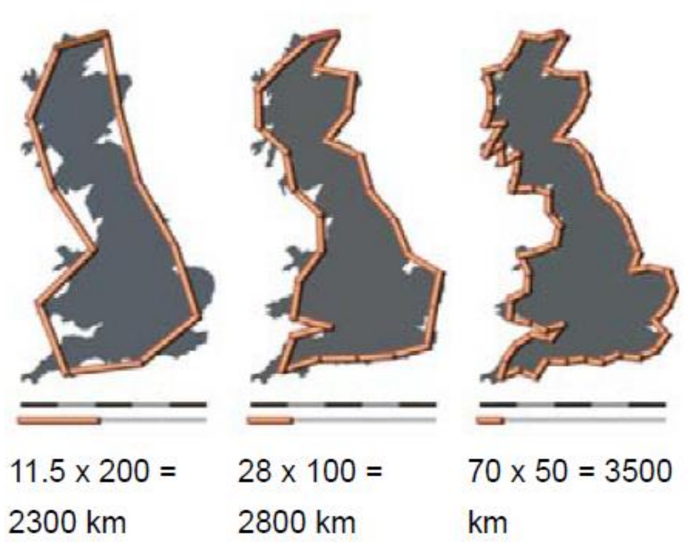
- The algorithm returns λ in the fastest expansion direction.
- The algorithm always returns a positive number!
- This is a main problem when computing the LE of noisy data.

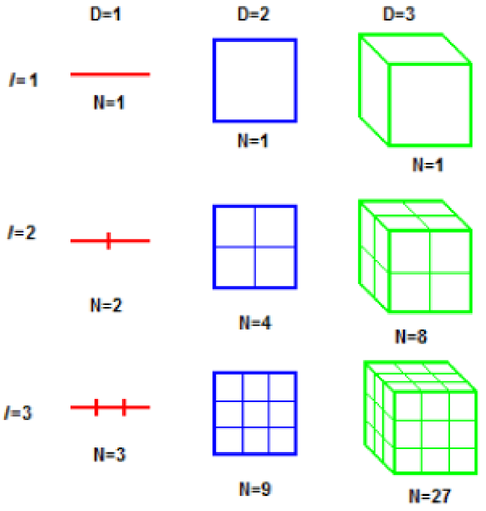
Every time series analysis algorithm returns a number of any time series. But is it useful?

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

Fractal dimension

- Example: the fractal dimension of a coastline quantifies how the number of scaled measuring sticks required to measure the coastline changes with the scale applied to the stick.

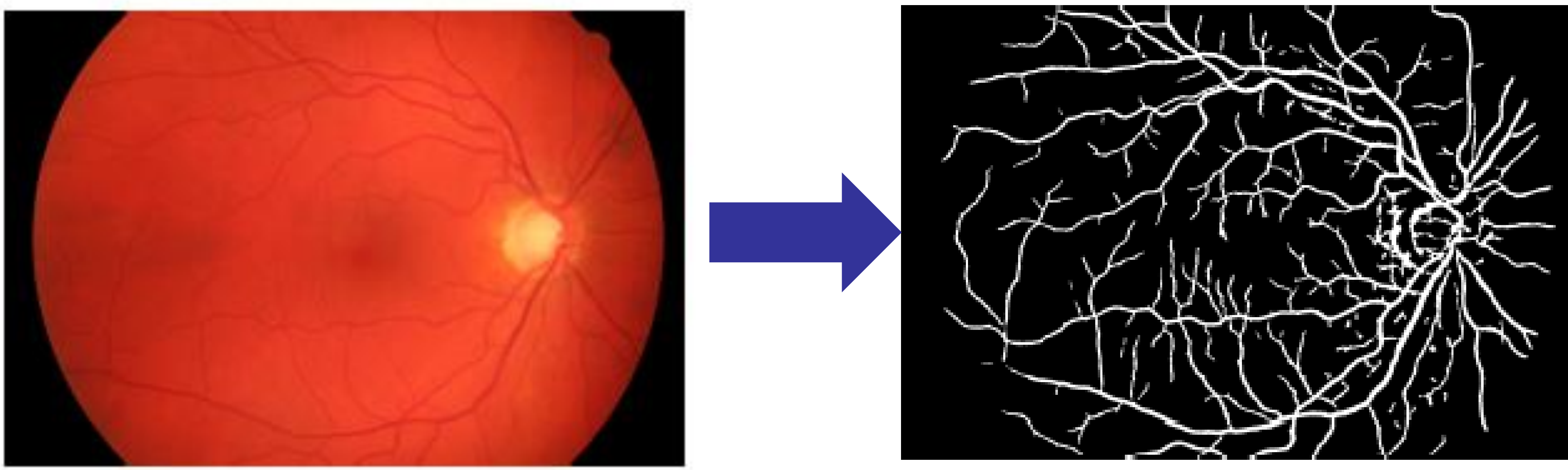




- Fractal dimension: $N \propto \epsilon^{-D} \longrightarrow D_0 = \lim_{\epsilon \to 0} \frac{\log N(\epsilon)}{\log \frac{1}{\epsilon}}$

Source: Wikipedia

Application of fractal analysis



The fractal dimension of the blood vessels in the normal human retina is about 1.7 while it tends to increase with the level of diabetic retinopathy.

P. Amil et al., PLoS ONE 14, e0220132 (2019).

Grassberger-Procaccia correlation dimension algorithm

- Another very popular method for detecting chaos in real-world data.
- Fractal dimension (box counting dimension): $D_0 = \lim_{\epsilon \to 0} \frac{\log N(\epsilon)}{\log \frac{1}{\epsilon}}$
- **Problem**: for time-series analysis, D_0 does not distinguish between frequently and unfrequently visited boxes.
- An alternative: the *correlation dimension*, based on calculating the number of pairs of points with distance between them < ε.

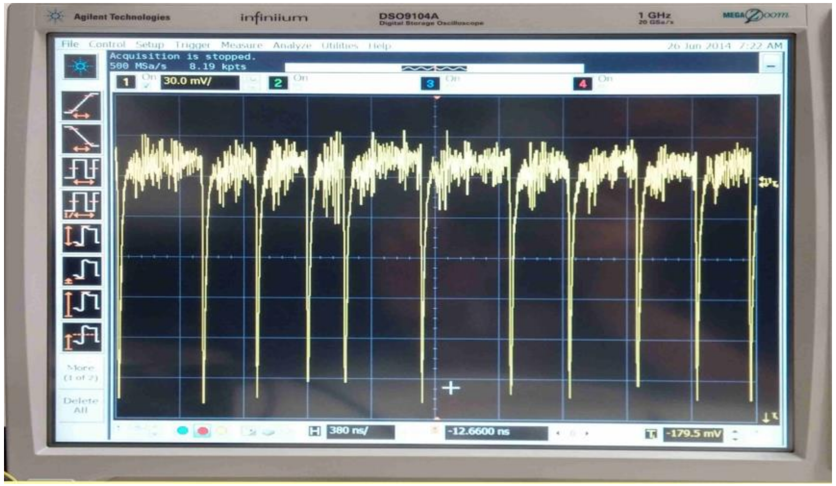
Further reading:
P. Grassberger and I. Procaccia, "Measuring the Strangeness of Strange Attractors". Physica D vol. 9, pp.189, 1983.
L. S. Liebovitch and T. Toth, “A fast algorithm to determine fractal dimensions by box counting,” Physics Letters A, vol. 141, pp. 386, 1989.

Methods of time series analysis

- Return maps
- Distribution of data values
- Correlation and Fourier analysis
- Autocorrelation
- Statistical significance 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

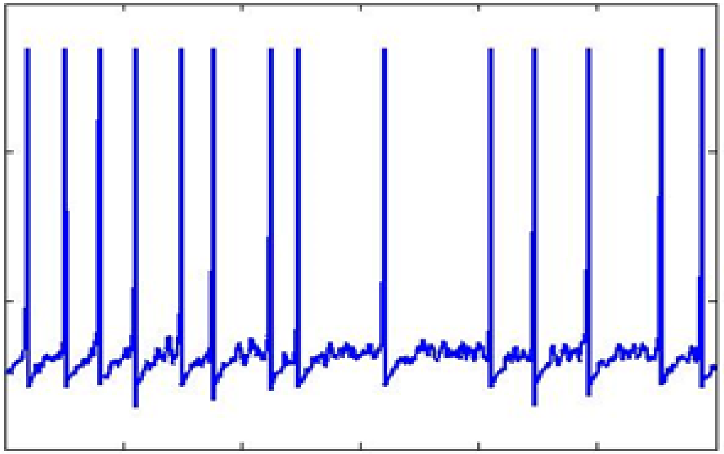
Can lasers mimic real neurons?

Laser spikes



Time (μs)

Neuronal spikes



Time (ms)

- Are there statistical similarities?
- A popular technique: define spike times via “threshold crossings” and analyze the statistical properties of the sequence of inter-spike-intervals (ISIs). Data compression!
- Results should be robust to small variations of the threshold.

Symbolic analysis

- The time series $\{x_1, x_2, x_3, \ldots\}$ is transformed (using an appropriated **rule**) into a sequence of symbols $\{s_1, s_2, \ldots\}$
- 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, 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:

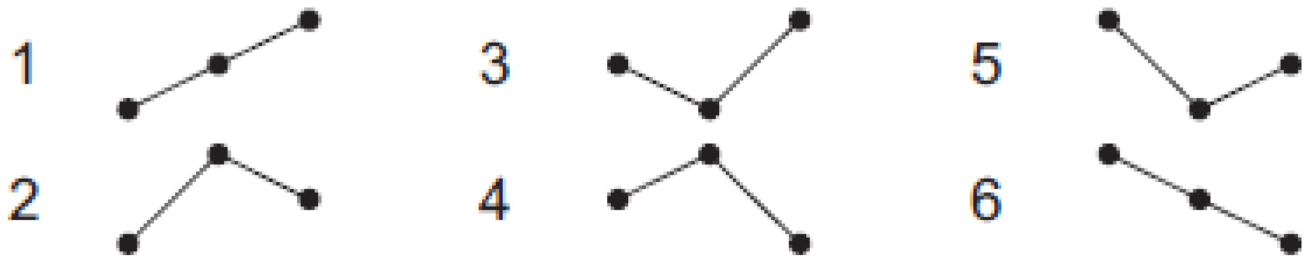
  if $x_i > x_{th1} \Rightarrow s_i = 0$;

  else if $x_i < x_{th2} \Rightarrow s_i = 2$;

  else $(x_{th2} < x_i < x_{th1}) \Rightarrow s_i = 1$.

Ordinal analysis: threshold-less method to define symbols

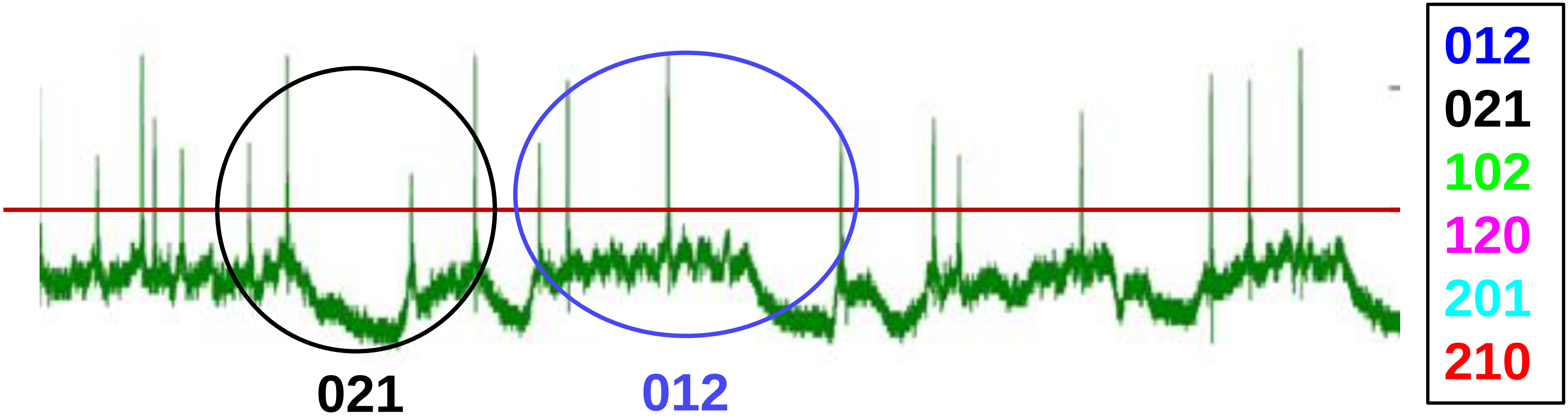
- Consider a time series $x(t)=\{\ldots x_i, x_{i+1}, x_{i+2}, \ldots\}$
- Which are the possible order relations among three consecutive data points?

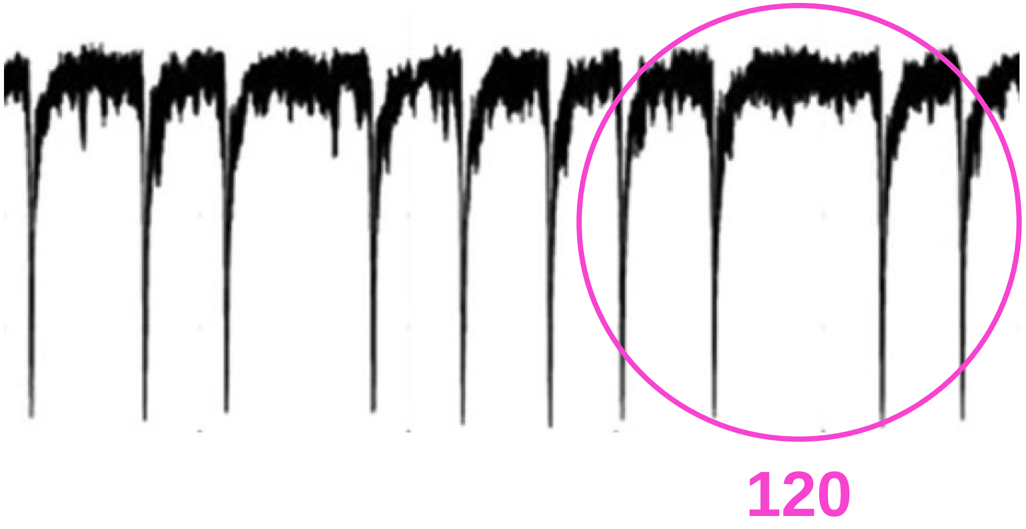


- Count how many times each “ordinal pattern” appears.
- Advantages: allows to identify temporal structures & is robust to noise.
- Drawback: information about actual data values is lost.

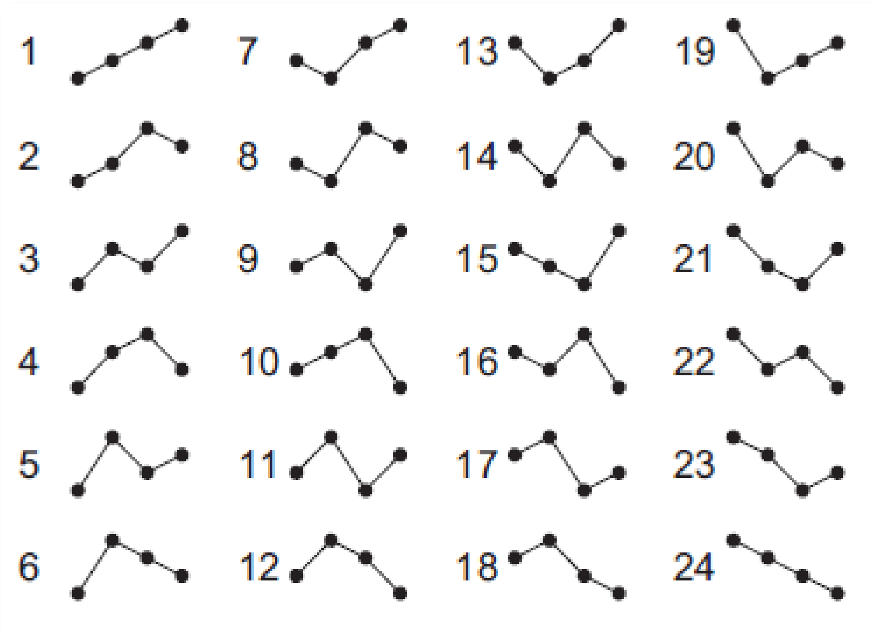
Bandt and Pompe, Phys. Rev. Lett. 88, 174102 (2002)

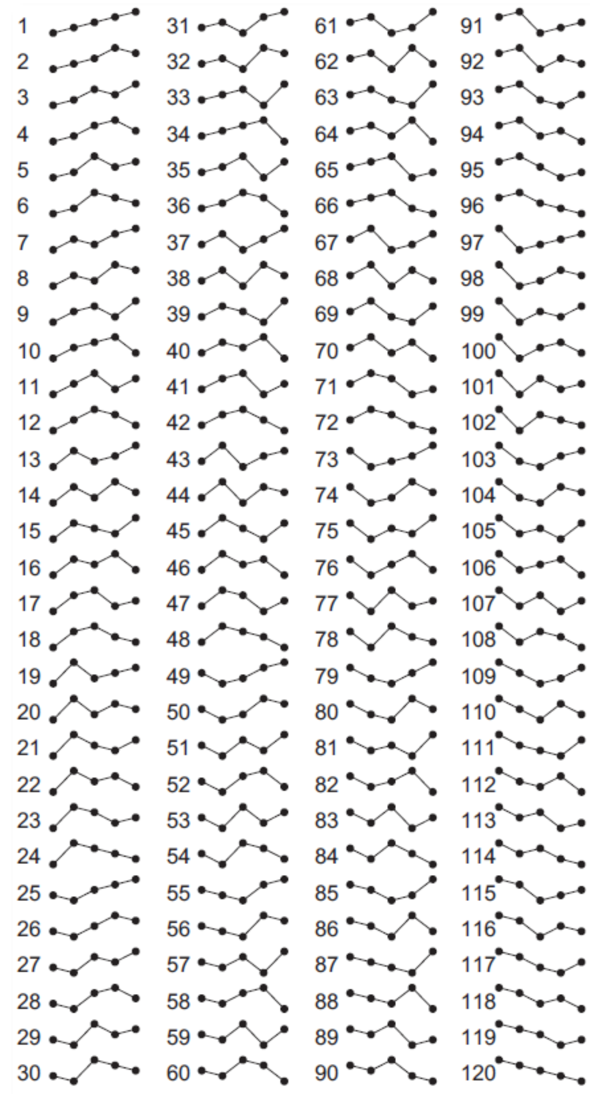
Analysis of D=3 patterns in spike sequences





The number of ordinal patterns increases as D!





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

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

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

Comparison between the two rules to define symbols

Threshold transformation:

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

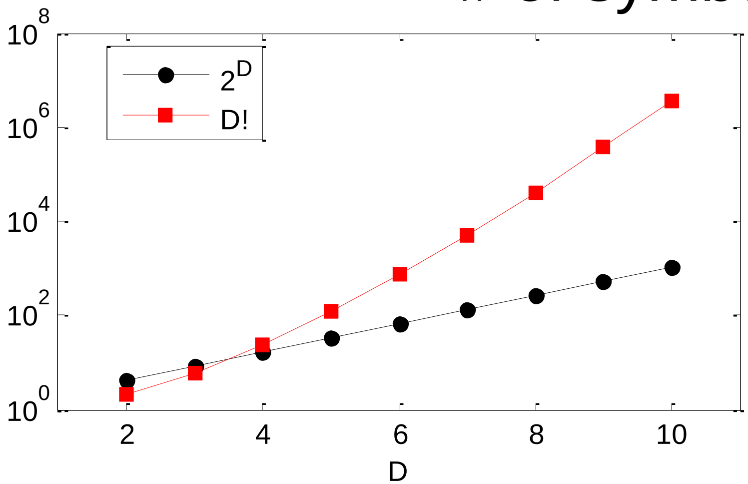
- Advantage: keeps information about the magnitude of the values.
- Drawback: how to select an adequate threshold (“partition” of the phase space).
- # of symbols: 2^D

Ordinal transformation:

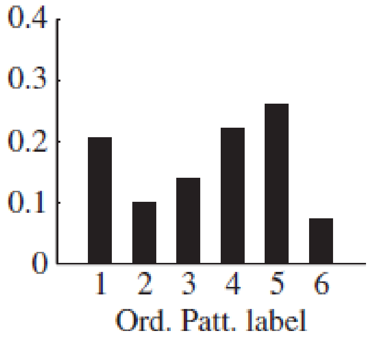
if $x_i > x_{i-1} \Rightarrow s_i = 0$; else $s_i = 1$

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

Number of symbols



Are the D! ordinal patterns equally probable?



- **Null hypothesis**:

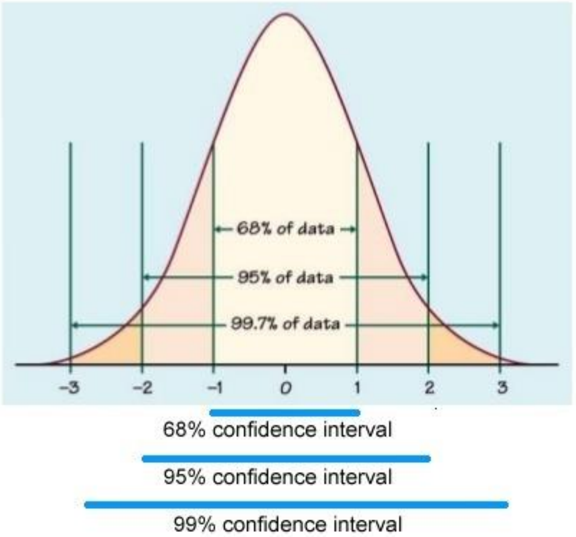
  $p_i = p = 1/D!$ for all $i = 1 \dots D!$

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

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

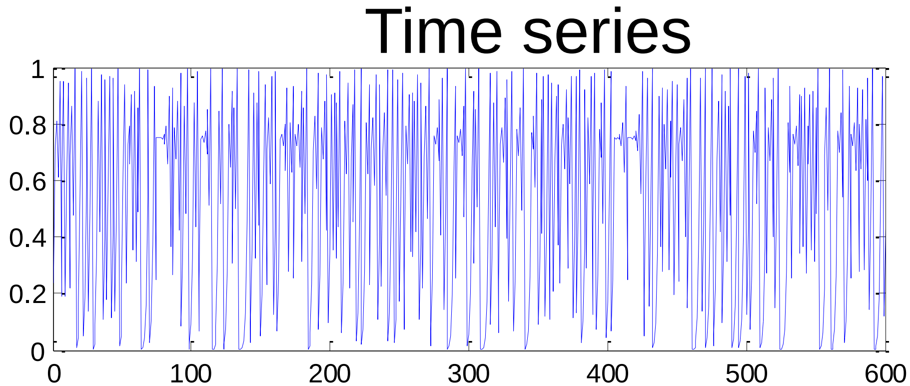
- Else

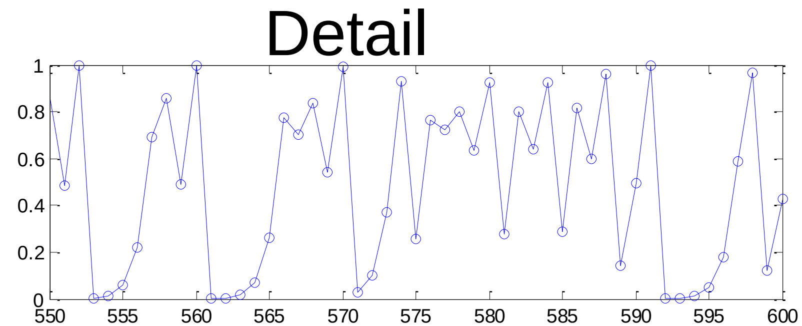
  We **fail to reject** the NH with 99.74% confidence level.

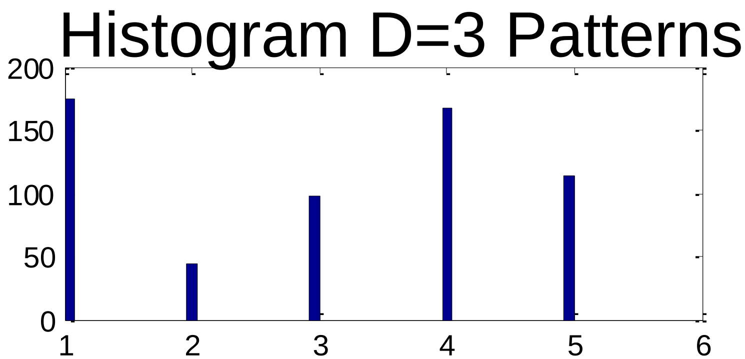


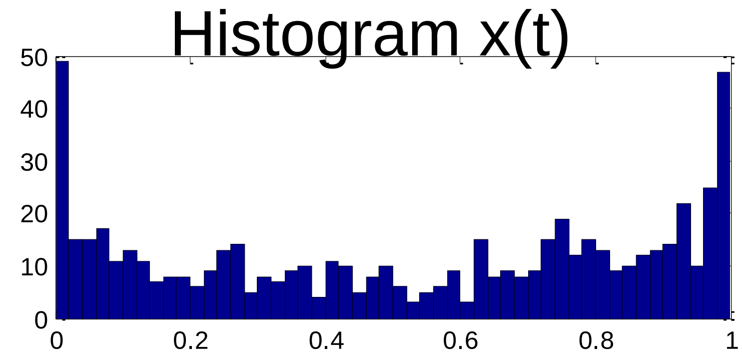
Logistic map $x(i+1) = r\,x(i)[1 - x(i)]$

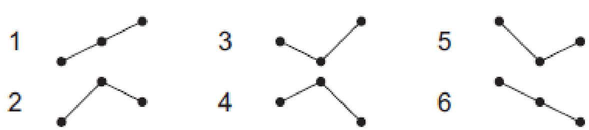
r=3.99











↑ **forbidden**

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

How to detect longer temporal correlations?

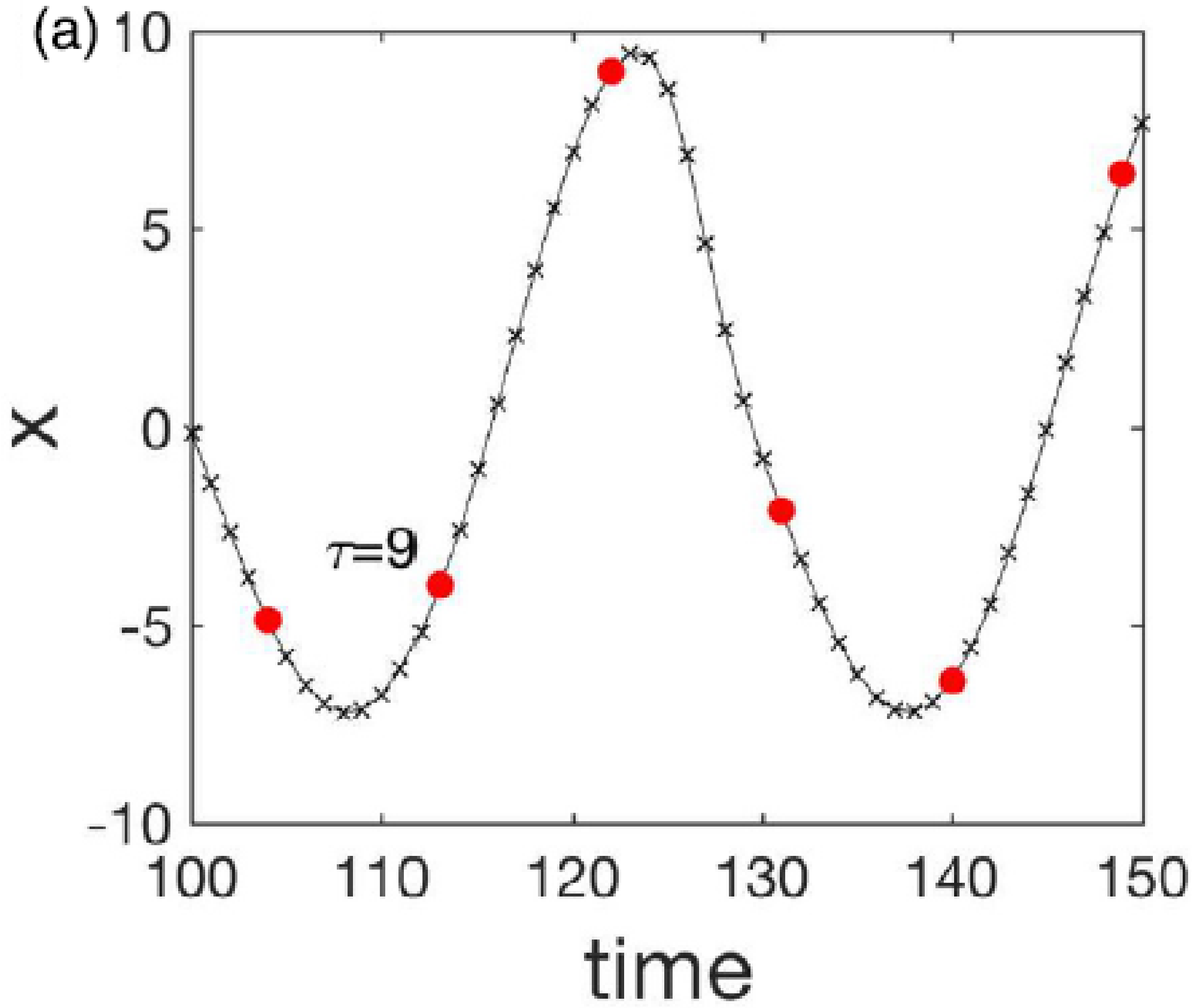
$$[\ldots\, x(t), x(t+1), x(t+2), x(t+3), x(t+4), x(t+5)\ldots]$$

- Problem: number of patterns increases as D!.

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

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

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



What to do if two values are exactly equal?
Which is the pattern?

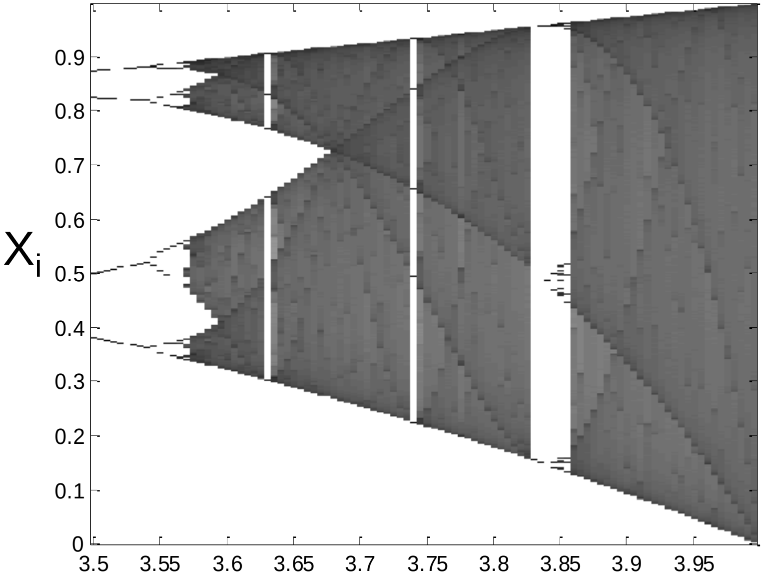
Several possible solutions, a simple one is to add a <u>very small</u> amount of noise: $x(t) = x(t) + \xi$.

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

"normal" and "ordinal" bifurcation diagram of the Logistic map with D=3

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

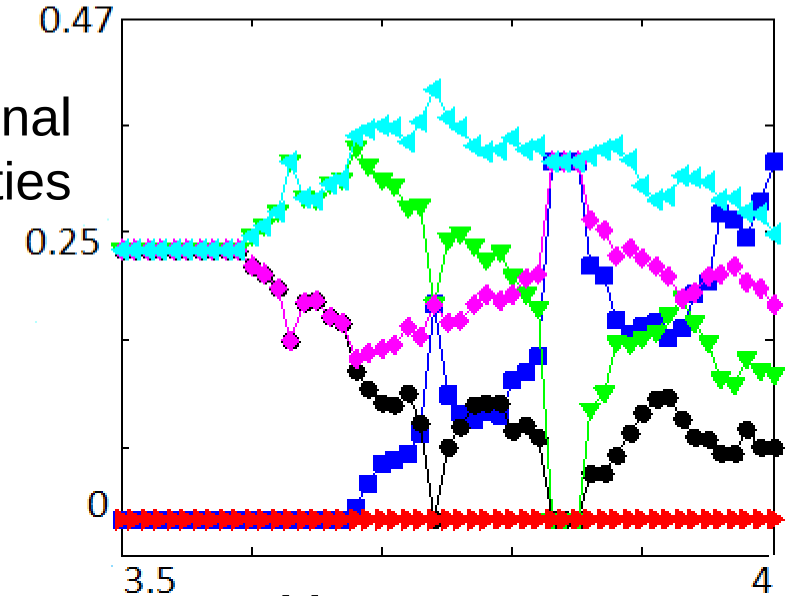
Normal bifurcation diagram



Map parameter

Ordinal bifurcation diagram

Ordinal Probabilities



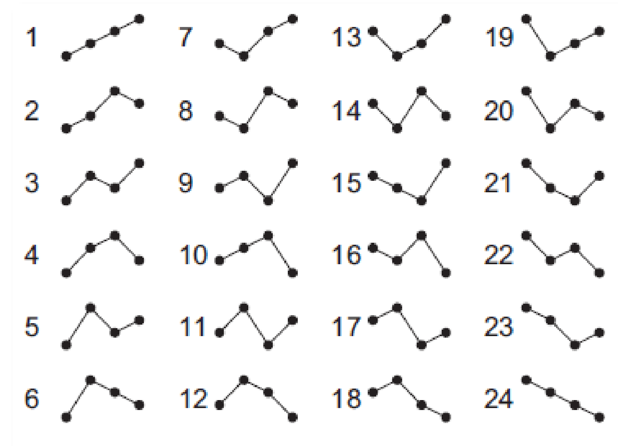
Map parameter, r

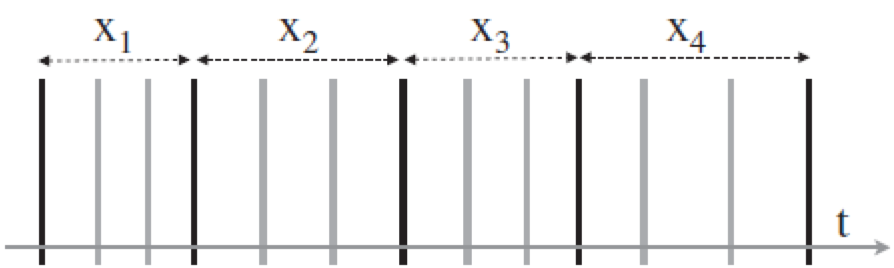
012 021 102 120 201 210

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)]()

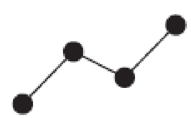




World length (wl): 4

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 ;
```

How to quantify unpredictability and complexity?

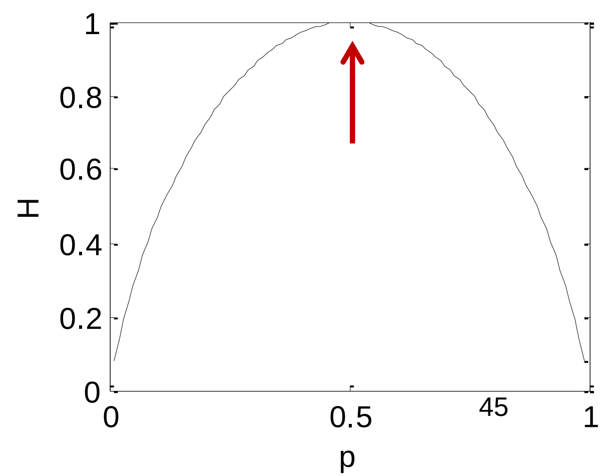
Methods of time-series analysis

- Return maps
- Distribution of data values
- Autocorrelation
- Statistical significance and surrogates
- Attractor reconstruction: Lyapunov exponents and fractal dimensions
- 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

Information measure: Shannon entropy

- The time-series is described by a set of probabilities $\sum_{i=1}^{N} p_i = 1$
- **Shannon** entropy: $$H = -\sum_{i=1}^{N} p_i \ln p_i$$
- Interpretation: "*quantity of* ***surprise*** *one should feel upon reading the result of a measurement*" Faser and Swinney (1986)
- Simple 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).

  ⇒ p=0.5: Maximum **unpredictability.**

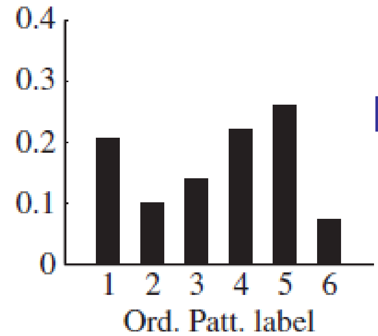


Permutation entropy

- Entropy computed from ordinal probabilities.
- Number of probabilities = # of ordinal patterns (D!)

Time series

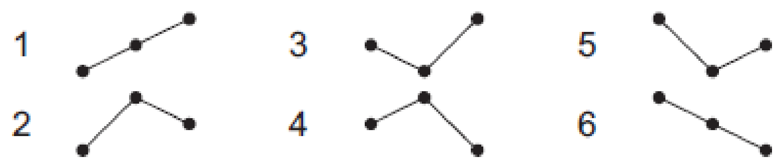
Ordinal probabilities



Permutation entropy

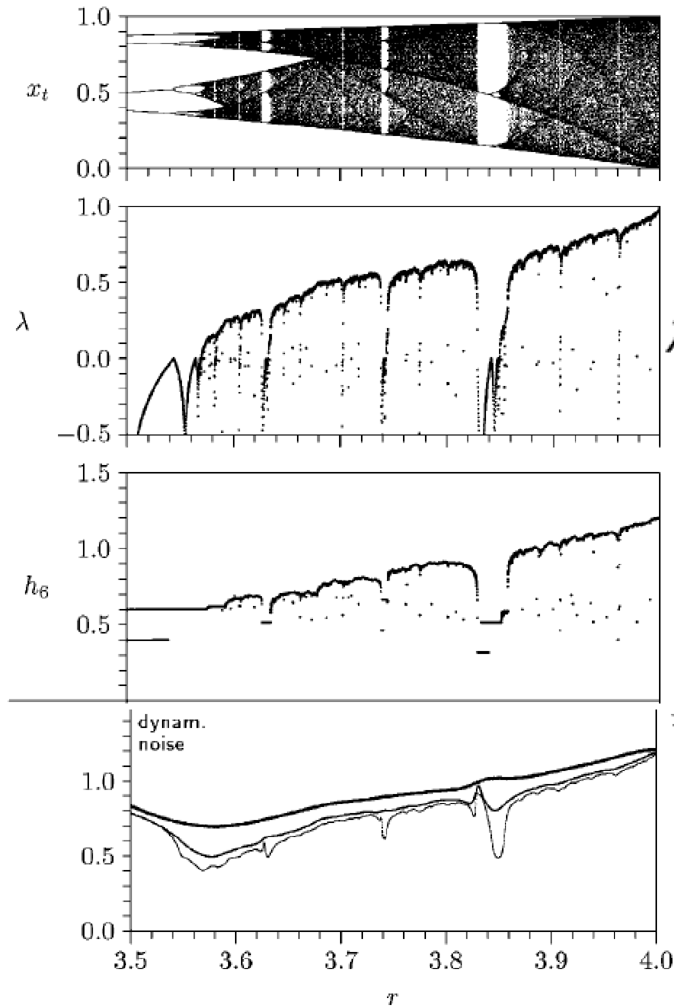
$$H = -\sum_{i=1}^{N} p_i \ln p_i$$

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



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:

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

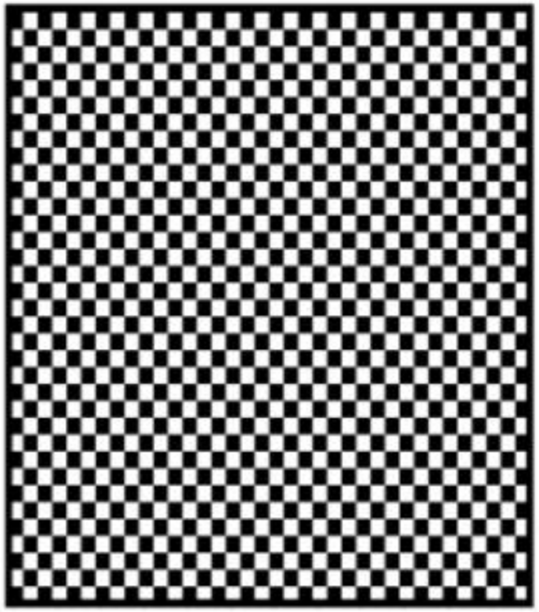
Robust to noise

Bandt and Pompe
Phys. Rev. Lett. 2002

The entropy measures the degree of 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

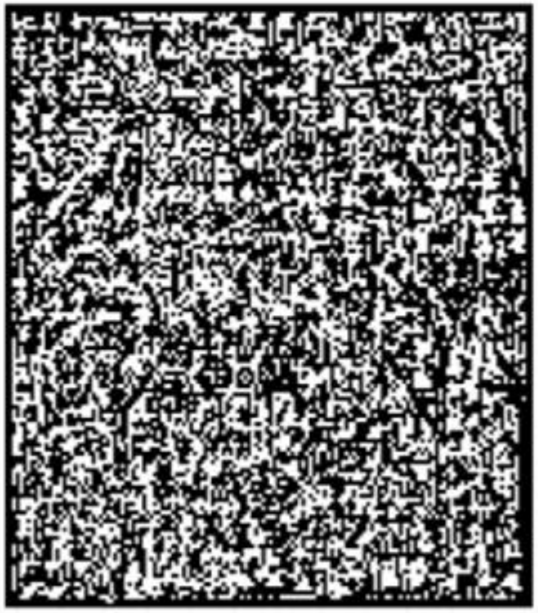


H = 0
C = 0

Chaos

H ≠ 0
C ≠ 0

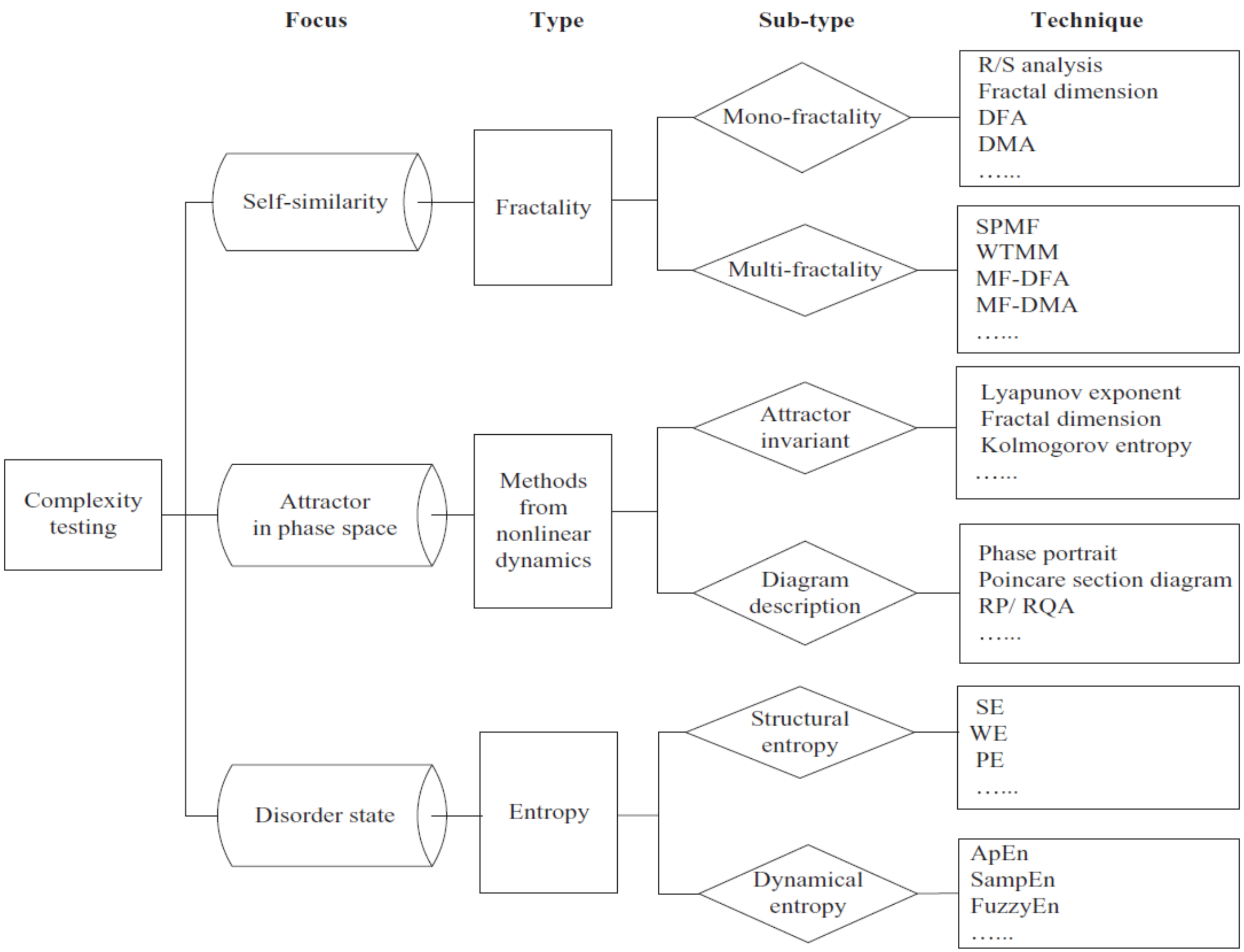
Disorder



H = 1
C = 0

Source: O. A. Rosso

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

Kolmogorov complexity

- The complexity of an object is a measure of the computability resources needed to specify the object.

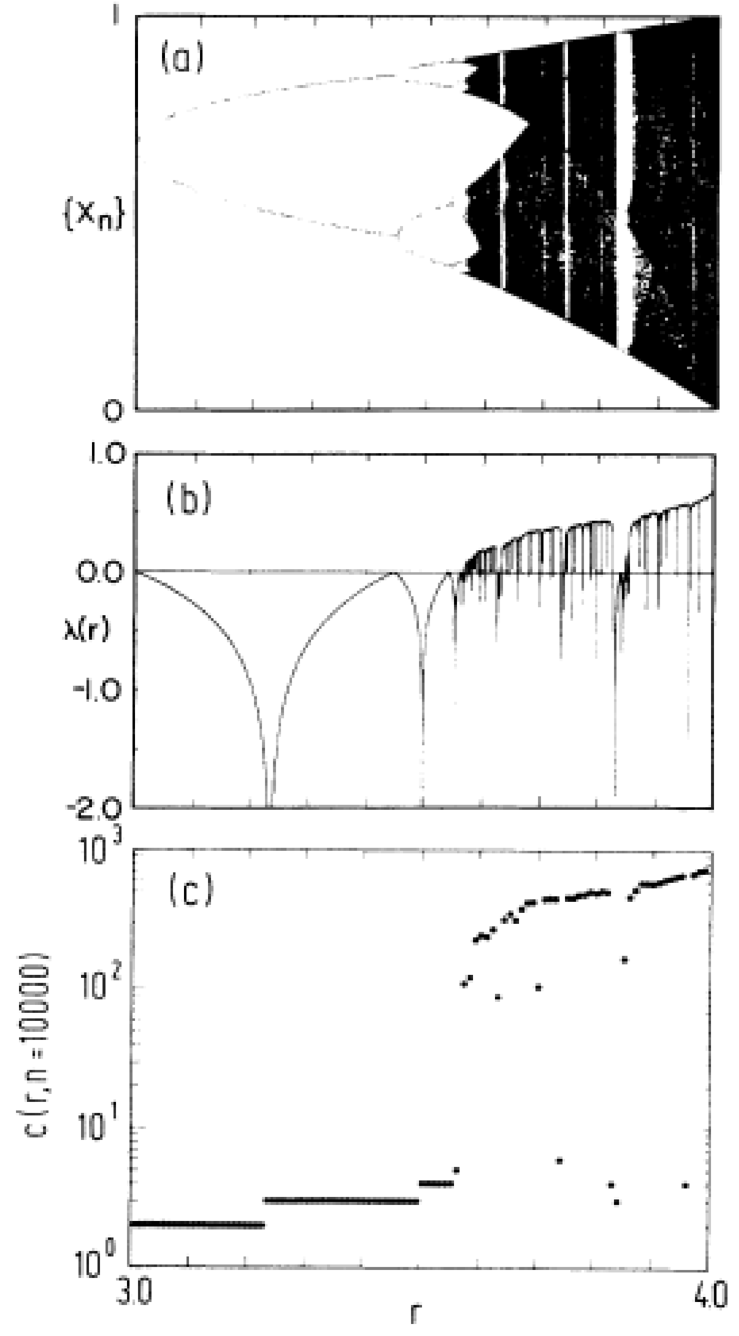
  Example: Let's consider 2 strings of 32 letters:

abababababababababababababababab

4c1j5b2p0cv4w1x8rx2y39umgw5q85s7

- The first string has a short description: "ab 16 times".
- The second has no obvious description: complex or random?
- The ***Lempel & Zip complexity*** is an estimation of the Kolmogorov complexity.

Lempel & Zip complexity of the Logistic Map



Kaspar and Schuster, Phys Rev. A 1987

<cristina.masoller@upc.edu>

Universitat Politecnica de Catalunya

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

UNIVERSITAT POLITÈCNICA DE CATALUNYA BARCELONATECH