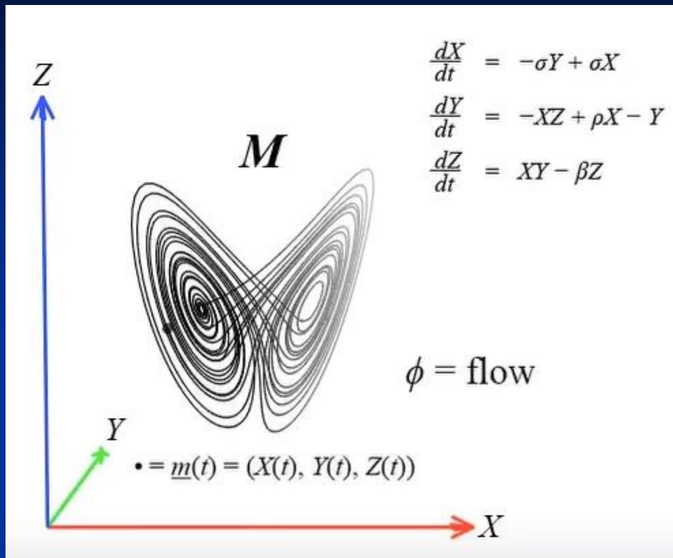
Causality detection methods for time series analysis

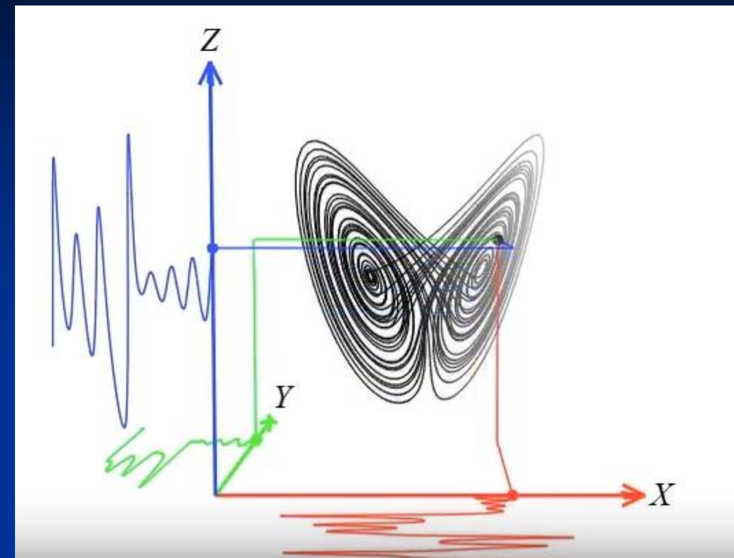
T. Craciunescu[a], A. Murari[b,c], E. Peluso[d], M. Gelfusa[d], M. Lungaroni[d], P. Gaudio[d]

[a]*National Institute for Laser, Plasma and Radiation Physics, Magurele-Bucharest, Romania,*
[b]*CCFE, Culham Science Centre, Abingdon, Oxon, OX14 3DB, UK*
[c]*Consorzio RFX, Padova, Italy*
[d]*University of Rome "Tor Vergata", Rome, Italy*

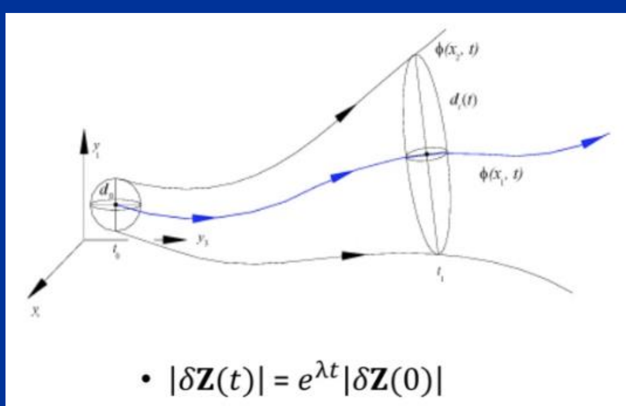
Time series



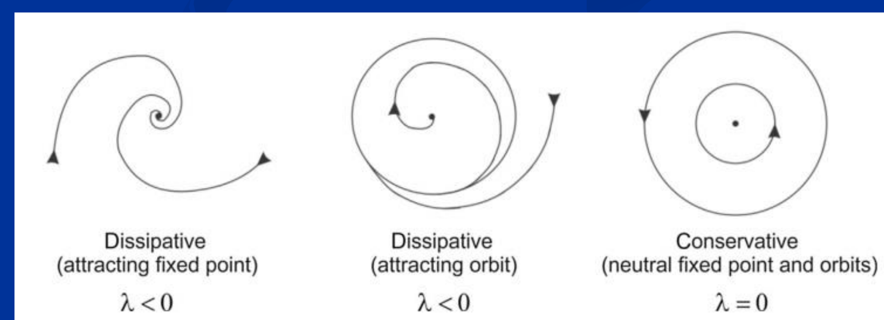
Coupled dynamical system (Lorenz)
Manifold M – set of all trajectories



Time series – projection of the manifold on the state bases



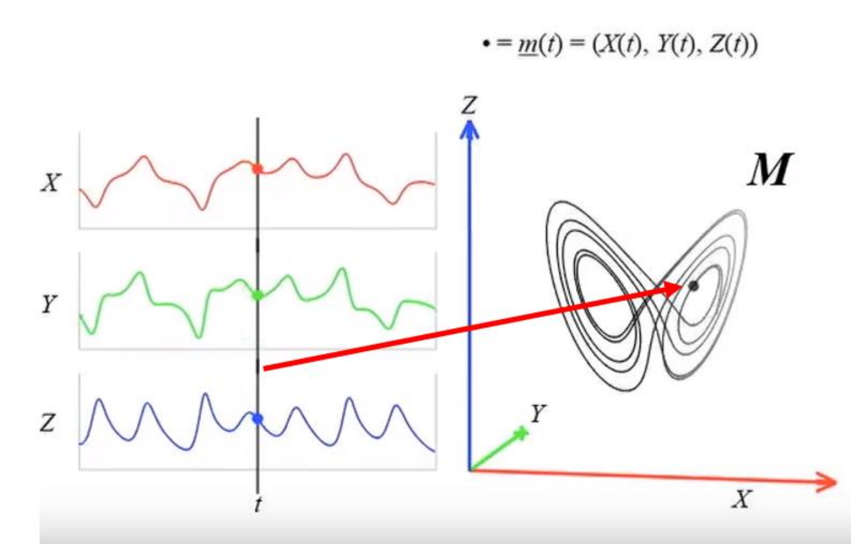
Lyapunov exponent



$\Lambda > 0$, the orbit is unstable and chaotic

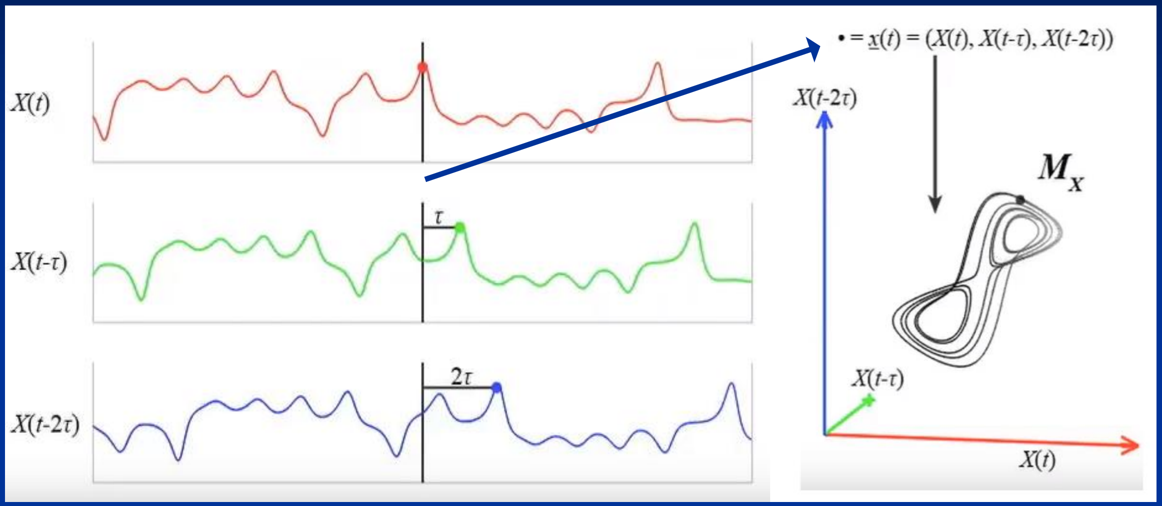
Nearby points, no matter how close, will diverge to any arbitrary distance

The manifold can be reconstructed if all time series are available



Takens theorem (1981)

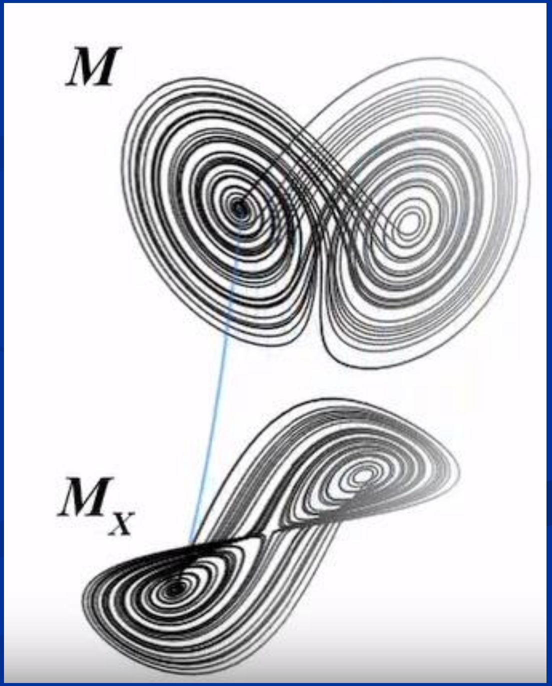
Reconstructing a shadow of the original manifold simply by looking at one of its time series projections



Copies of variable X displaced by τ

The reconstruction preserves the essential mathematical properties of the system:

- topology of the manifold
- Lyapunov exponents.



one-to-one correspondence

Embedding dimension

"false neighbours"

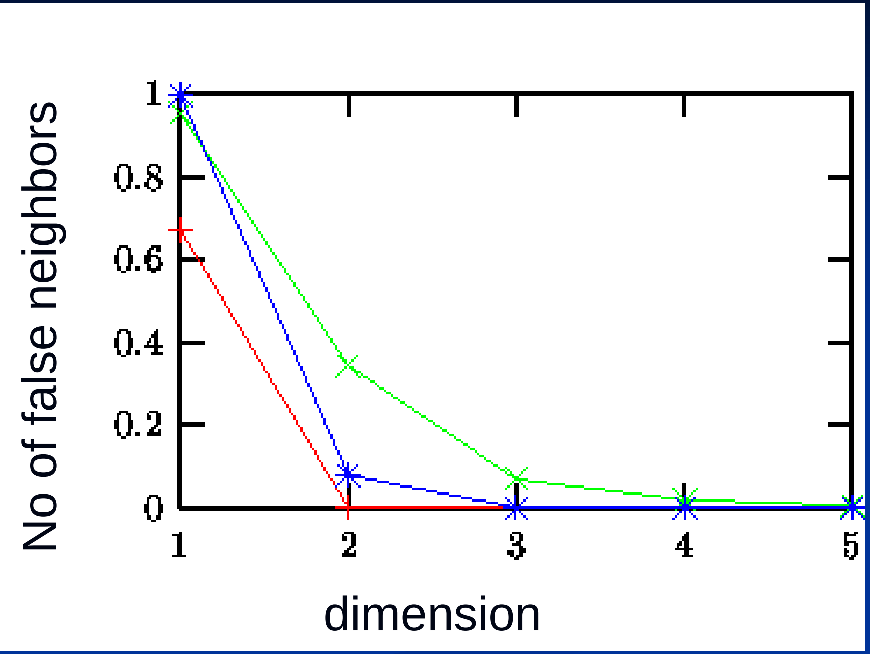
- For each point in the time series look for its nearest neighbors

➢ Some of them appear as neighbors because the geometric structure of the attractor has been **projected down to a smaller space**

$$R_i = \frac{\left\| x_i^{m+1} - x_j^{m+1} \right\|}{\left\| x_i^m - x_j^m \right\|}$$

If $R_i > \varepsilon$

then from dimension **d** to **(d + 1)**



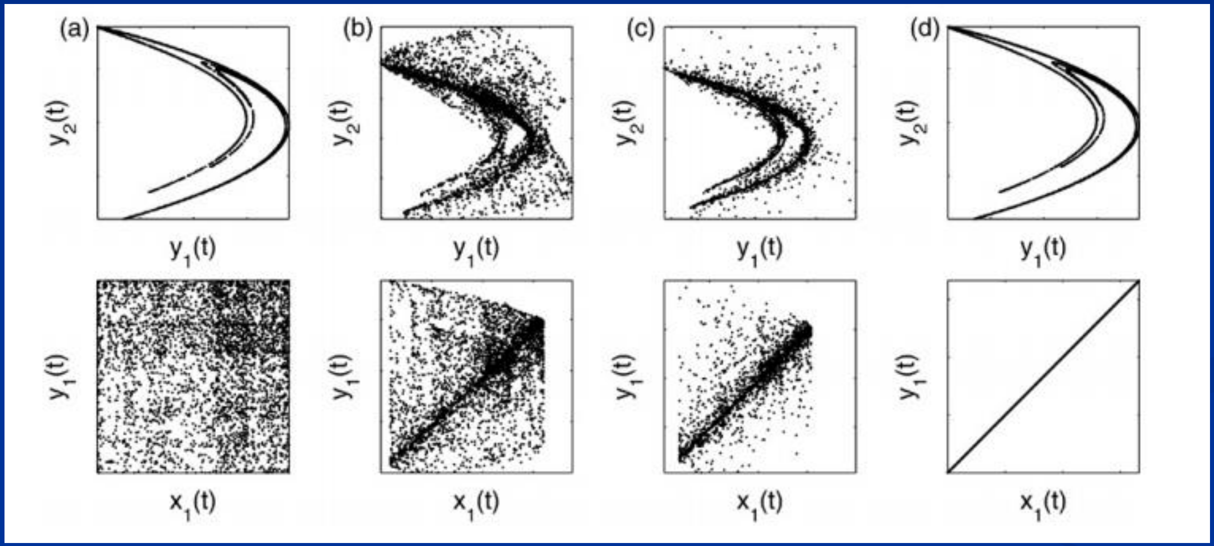
➢ If the embedding dimension is too small to unfold the attractor, not all the points that lie close to each other will be real neighbors

M.B. Kennel et al, Phys. Rev. A 45, 3403 (1992)

Coupling and synchronization

- common phenomena that occur in nature, e.g. in biological, physiological systems, engineered systems, social perception, epidemiology, econometrics

Coupled Hénon systems



Chaotic systems → implies the rapid decorrelation of nearby orbits due to their high sensitivity on initial conditions

Synchronization a rather counter-intuitive phenomenon

driver

$$x_1[n+1] = 1.4 - x_1^2[n] + 0.3x_2[n]$$
$$x_2[n+1] = x_1[n]$$

responder

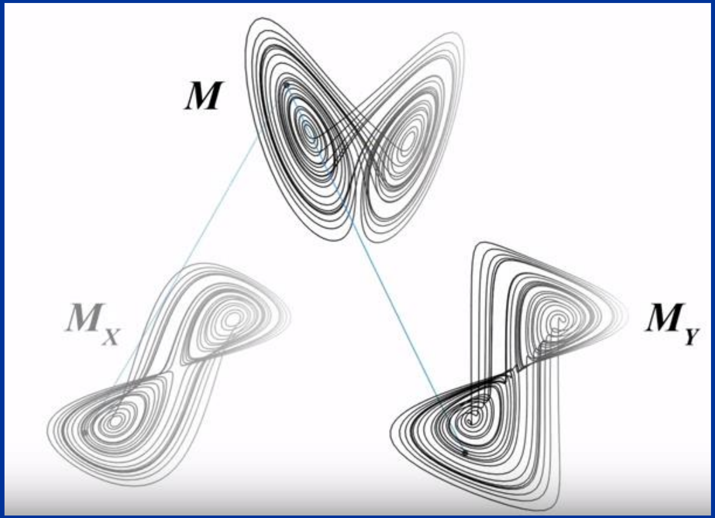
$$y_1[n+1] = 1.4 - (Cx_1[n]y_1[n] + (1-C)y_1^2[n]$$
$$y_2[n+1] = y_1[n]$$

T. Kreuz et al. / Physica D 225 (2007) 29–42

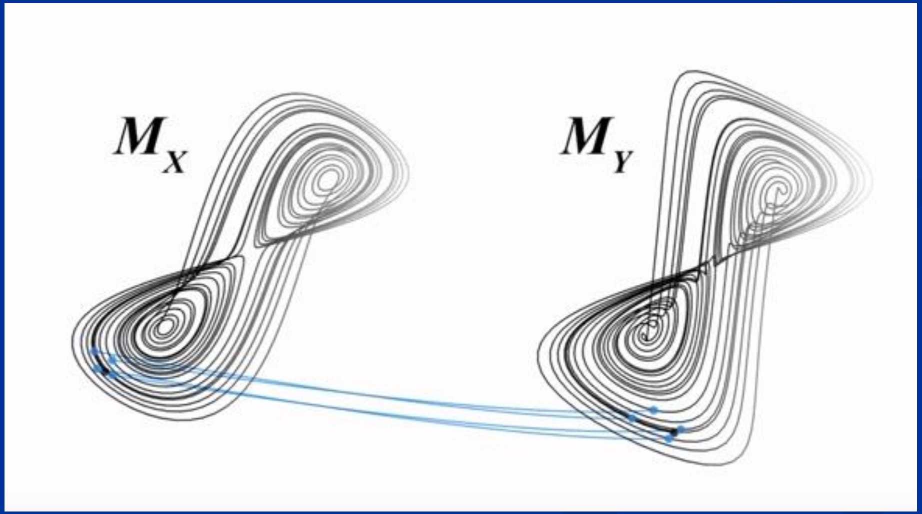
Cross Convergent Maps (CCM)

Two time series are related by a causal relation if they belong to the same dynamical system

M – original manifold
M_X – manifold created from time series X
M_Y – manifold created from time series Y
M_X, M_Y - diffeomorphic to the original **M**



As **M_X** and **M_Y** maps one-to-one to the original manifold **M** then they should map one-to-one to each other

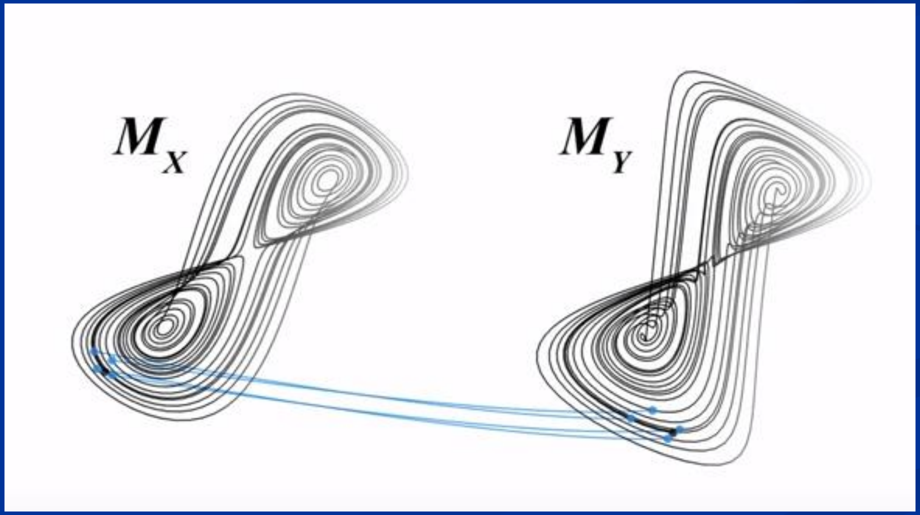


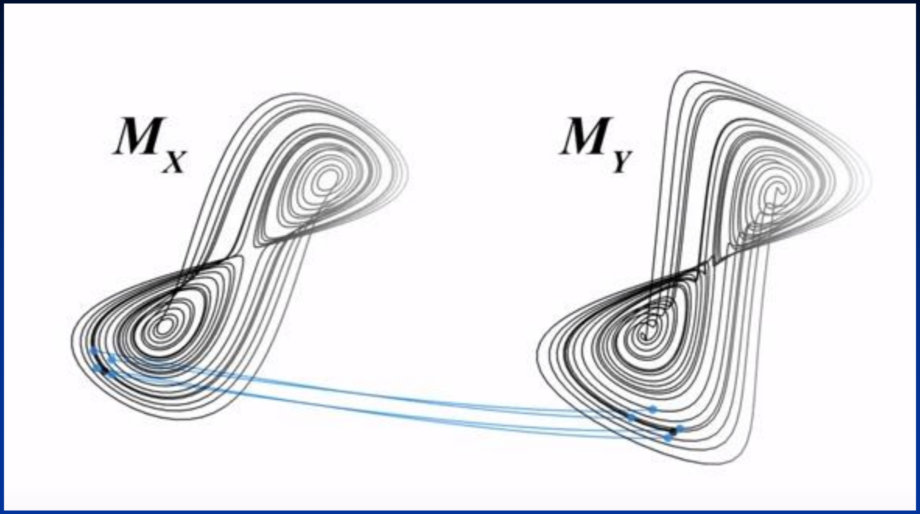
Nearest neighbors on **M_X** should correspond to nearest neighbors on **M_Y**

G. Sugihara et al., Science 26 Oct 2012

- CCM - determine how good is the correspondence between local neighbourhoods on $\mathbf{M}_x$ and local neighbourhoods on $\mathbf{M}_y$.

- determining the embedding dimension **E**
- for each time t, x(t) is the corresponding value in one time series and y(t) in the other
- for each x(t), the **E+1** neighbours are found
- $\mathbf{t_1, t_2, \ldots, t_{E+1}}$ be the time indices of the nearest neighbors of **x(t)**, ordered from closest to farthest





- Then <u>these time indices</u> are used to construct a putative neighborhood on the manifold of system **y**
- <u>Estimated</u> value from the locally weighted mean of the **(E + 1) y(t_i)** values:

$$y^{est}|M_x = \sum w_i y(t_i)$$

where $\boldsymbol{w_i}$ is a weighting factor whose definition is based on the distance between ***x(t)*** and its ***ith*** neighbour on the manifold M_x.

$$w_i = \frac{u_i}{\sum u_j}, \qquad u_i = exp\left(-\frac{d\big(x(t) - x(t_i)\big)}{d\big(x(t) - x(t_1)\big)}\right)$$

d - distance (Euclidean, Geodesic, etc)

The difference between the estimated values $y^{est}(t)$ and the actual values $y(t)$ is evaluated by the Pearson correlation coefficient:

$$\rho = \frac{COV(Y^{est}, Y)}{\sigma(Y^{est})\sigma(Y)}$$

When looking for the signature of $\boldsymbol{X}$ in $\boldsymbol{Y}$'s time series:

→ A large value of $\boldsymbol{\rho}$ indicates that points on the $\boldsymbol{Y}$ manifold can be used to identify nearby points on M_x

→ This happens only if $\boldsymbol{X}$ is causally influencing $\boldsymbol{Y}$

Recurrence Plots

A recurrence plot (RP) is a plot showing the times at which the phase space trajectory of a dynamical system visits roughly the same area in the phase space

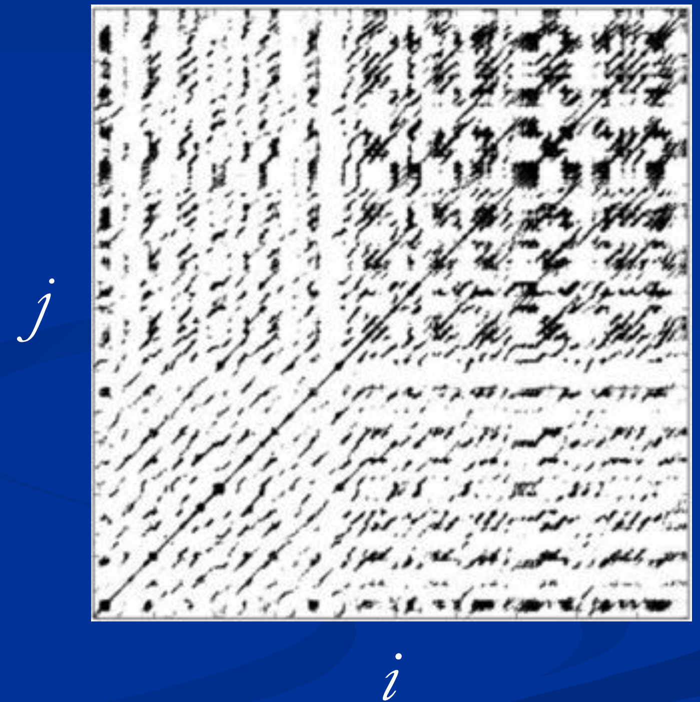
RPs are based on the following matrix representation:

$$R_{ij} = \Theta\left(\varepsilon - \left\|\overrightarrow{x_i} - \overrightarrow{x_j}\right\|\right), \quad i,j=1,\ldots,N$$

$\overrightarrow{x_i}, \overrightarrow{x_j}$ - points in phase space at time i and j

ε is a predefined threshold

$\Theta(x)$ is the Heaviside function

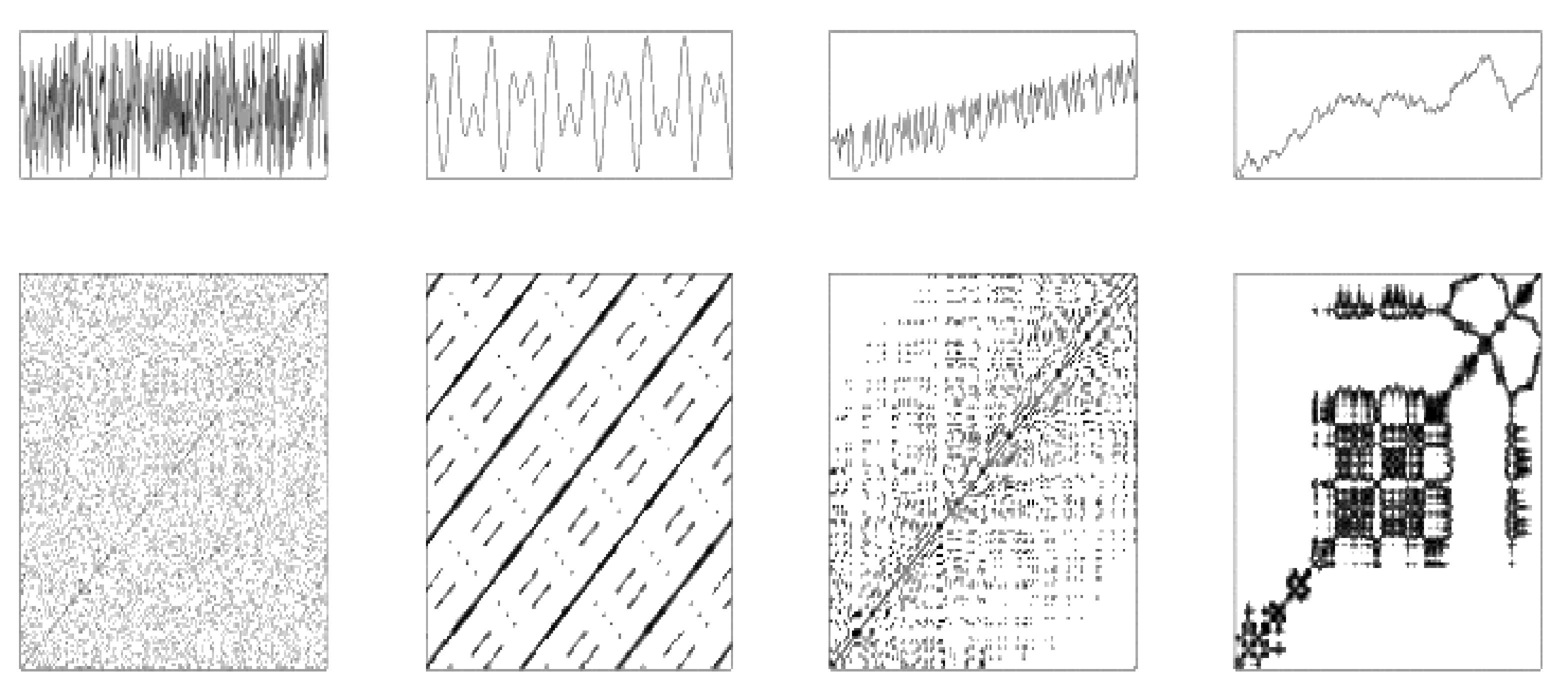


- RP depict the collection of pairs of times at which the trajectory returns sufficiently close the same place

N. Marwan et al., Physics Reports, 438(5-6), 237-329.

Recurrence Plots

- Measure the recurrences of the trajectories
- Instrument for visualizing the behavior of trajectories in phase space



uncorrelated stochastic data (white noise)

harmonic oscillation with two frequencies

chaotic data with linear trend (logistic map)

auto-regressive process

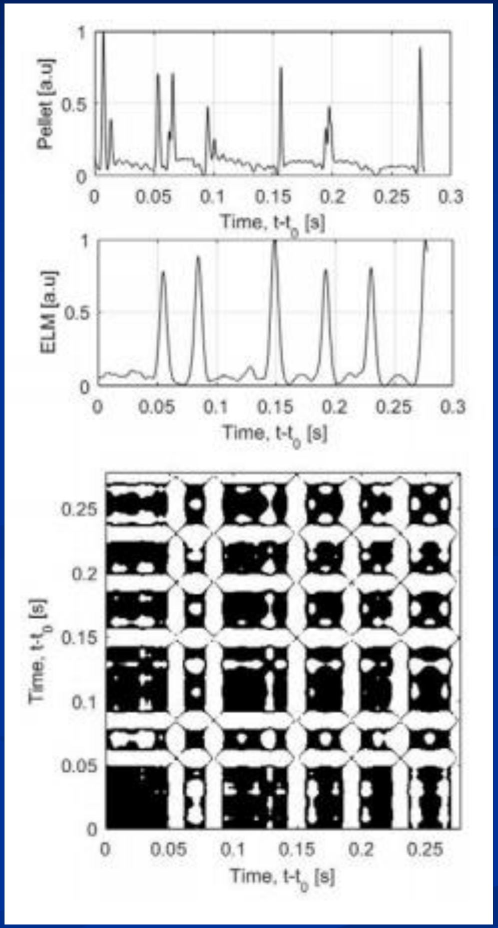
Joint recurrence plots (JRP)

Hadamard product of the recurrence plots of the considered sub-systems

$$JR(i,j) = \Theta\left(\varepsilon_x - \left\|\vec{x_i} - \vec{x_j}\right\|\right) \cdot \Theta\left(\varepsilon_y - \left\|\vec{y_i} - \vec{y_j}\right\|\right)$$

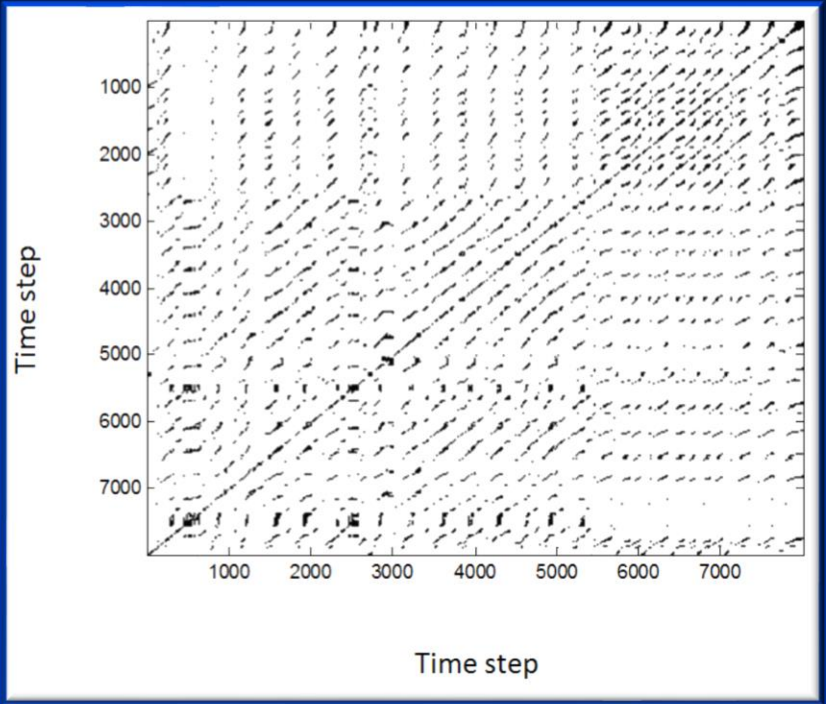
- Compare the simultaneous occurrence of recurrences in two (or more) systems
- Joint recurrence plots can be used to detect phase synchronisation.

RQA measures can be defined also for JRP



JRP between ELMs and pellets

Recurrence quantification analysis (RQA)



- The ***entropy of the diagonal lengths***:

$$ENTR = \sum_{l=l_{min}}^{N} p(l) \cdot \ln[p(l)]$$

- Gives a measure of how much information is needed to recover the system and it reflects the complexity of the RP with respect to the diagonal lines.

- Single isolated points corresponds to states with a rare occurrence, do not persist of they are characterized by high fluctuation

- Vertical/Horizonthal lines corresponds to states which do not change significantly during a certain period of time

- Diagonal lines occur when the trajectory visit the same region at different times and a segment of the trajectory runs parallel to another segment.

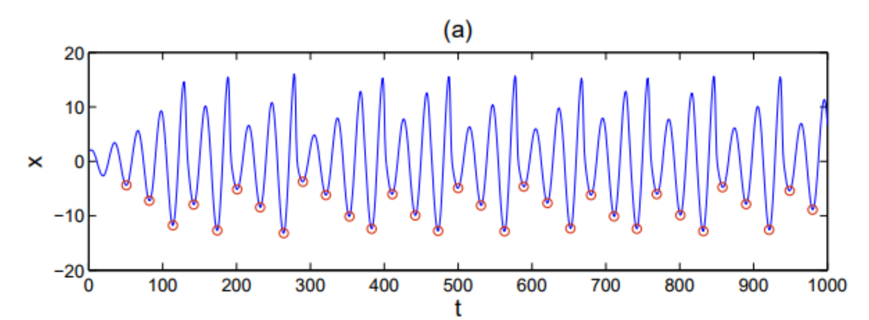
- Long diagonal structures corresponds to similar time evolution of the two processes.

J.P. Zbilut et al., Physics Letters A. **171** (3–4): 199–203

Complex networks

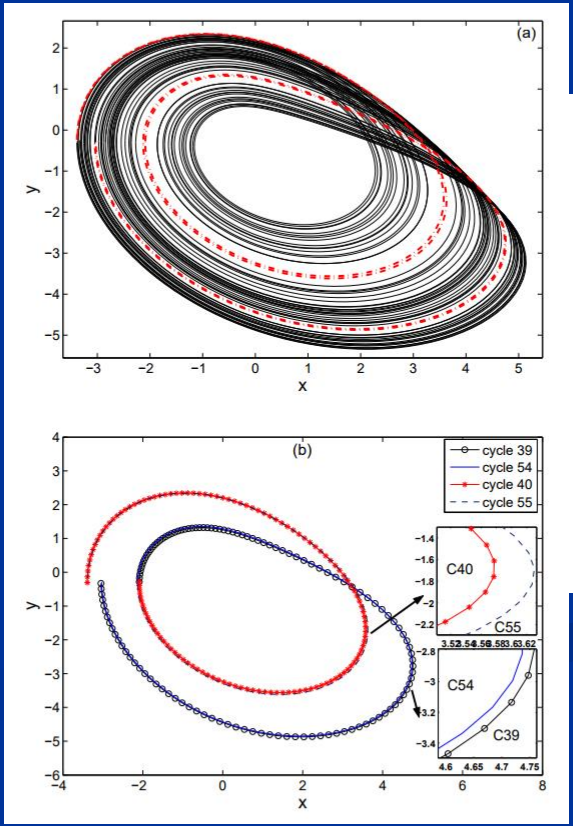
Transformation from the time domain to the network domain

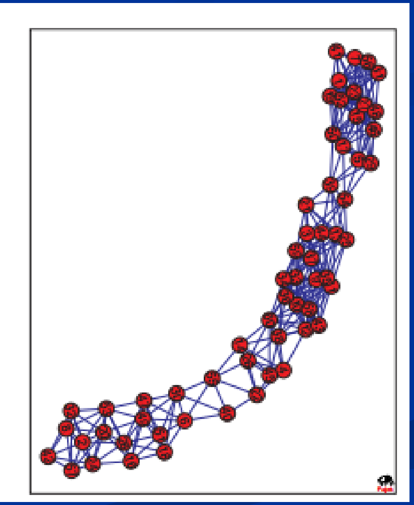
→ allows for the dynamics of the time series to be studied via the organization of the network.



- Time series is divided into m disjoint cycles usually segmentation by the local min/max
- Each cycle - a basic node of a graph
- A network representation is achieved by connecting:

  → cycles with phase space distance less than a predetermined value

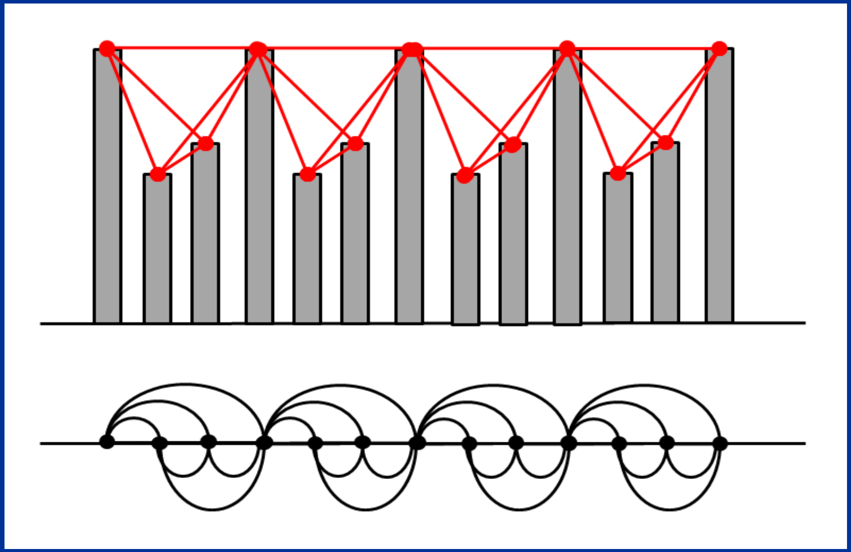




J. Zhang, M. Small, PRL 96, 238701 (2006)
J. Zhang et al., Physica D, 237-22 (2008) 1856

Visibility networks

Considering a representation of time series by using vertical bars and seeing this representation as a landscape, every bar in the time series is linked with those that can be seen from the top of the bar



- <u>Connected:</u> each node sees at least its nearest neighbors (left and right).
- <u>Undirected:</u> no direction defined in the links.

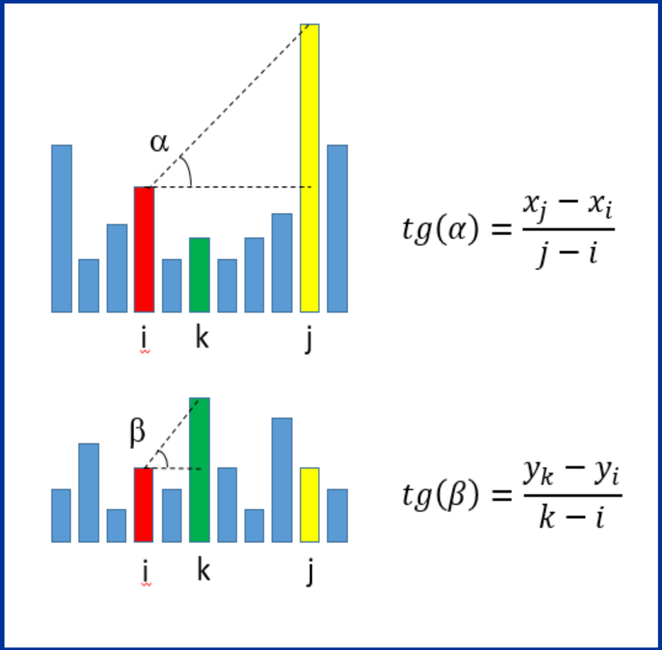
L. Lacassa et al., PNAS 105-13 (2008) 4972

Cross-Visibility networks

Transforming a set of two time series $\{x_i\}$ and $\{y_i\}$

$\rightarrow$ in order to reveal their coupling

- node is created for each time series point
- connection between two nodes *i* and *j* if:



$$y_k \geq y_i + \frac{x_j - x_i}{j-i}(k - i), \qquad i < \forall k < j$$

$tg(\alpha) > tg(\beta)$ for any $i < k < j$

visibility from the top

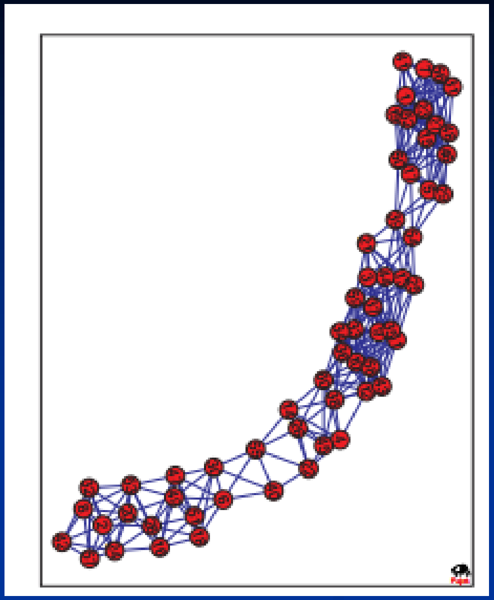
$$y_k \leq y_i + \frac{x_j - x_i}{j - i}(k - i), \qquad i < \forall k < j$$

$tg(\alpha) < tg(\beta)$ for any $i < k < j$

visibility from the bottom

A. Mehraban et al., EPL, 103-5, (2015) 50011

By means of specific transformations from the time domain to the network domain, the dynamics of time series can be studied via the organization of the networks by using topological statistics measures



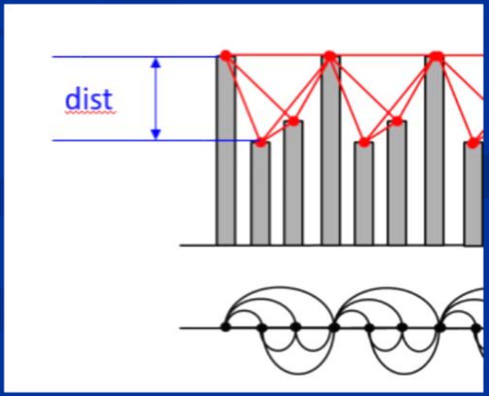
Adjacency matrix

- indicates if pairs of nodes are connected or not

$$A_{ij} = \begin{cases} 1, & if\ nodes\ i\ and\ j\ are\ connected \\ 0, & otherwise \end{cases}$$

Modified Adjacency Matrix (MAM)

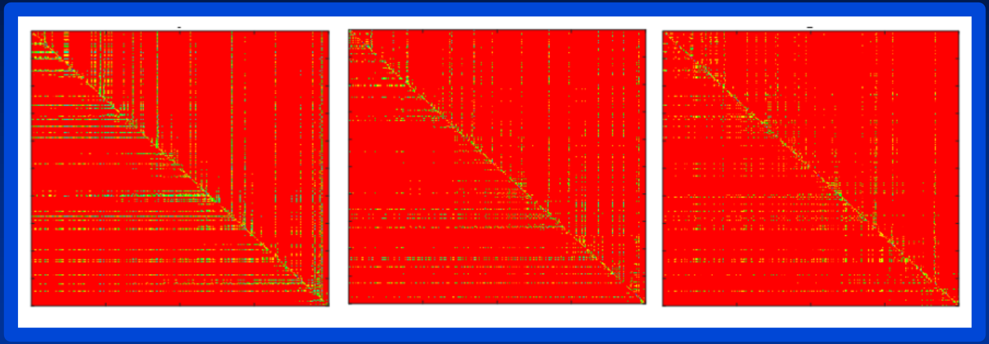
❖ by weighting the connections with the between the height of two connected nodes in the time series:



$$A_{ij} = \begin{cases} dist(y_i - y_j), & if\ visibility\ graph\ conditions\ are\ satified \\ 0, & otherwise \end{cases}$$

Entropy **2017**, *19*(10), 569

➢ **MAM can be represented as an image**



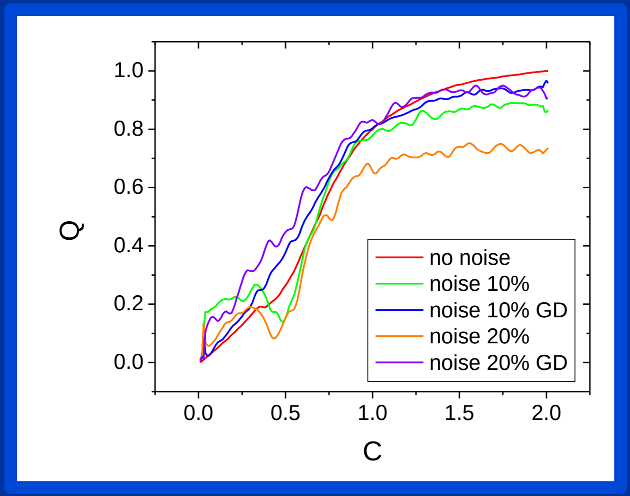
For a higher synchronisation the number of links in the network become lower leading to lower entropy images

Image entropy

$$H(X) = -\sum_i P_i \log P_i$$

Synchronization measure

$$Q = -H(CVN)$$



Q evolution for the coupled Hénon systems

Transfer Entropy

- Infer causality between two processes by measuring the information flow between them in terms of transfer entropy

- Measures the uncertainty reduction in inferring the future state of a process by learning the (current and past) states of another process

Entropy

- *discrete random variable* X *probability mass function:* $p(x) = Prob(X = x)$

$$H(X) = -\sum_{x} p(x) \log p(x)$$

approximates the minimal binary description length L of the random variable X

T. Schreiber, Phys. Rev. Lett. 85 (2000) 461–464

❖ *random variables* X, Y $\begin{cases} \text{joint distribution:} \quad p(x,y) = Prob(X = x, Y = y) \\ \text{conditional distribution:} \quad p(x|y) = Prob(X = x| \, Y = y) \end{cases}$

Joint Entropy:

$$H(X,Y) = -\sum_{x,y} p(x,y) \log p(x,y)$$

Conditional Entropy:

$$H(X|Y) = \sum_{y} p(y) \mathrm{p(x|y)}$$

Mutual Information :

$$I(X;Y) = H(X) - H(X|Y) = H(Y) - H(Y|X)$$

- measures the reduction of uncertainty of *X (Y)* given full information about *Y (X)*
- "measures their deviation from independence"

- X and Y are fully dependent: $H(X|Y) = H(Y|X) = 0$, $I(X;Y) = H(X) = H(Y)$
- X and Y are fully independent: $H(X|Y) = H(X)$ and $H(Y|X) = H(Y)$, $I(X;Y) = 0$
- in general: $0 \leq I(X;Y) \leq \min[H(X), H(Y)].$

❖ *Stochastic stationary processes*

$\{X_t\}$

Entropy Rate:

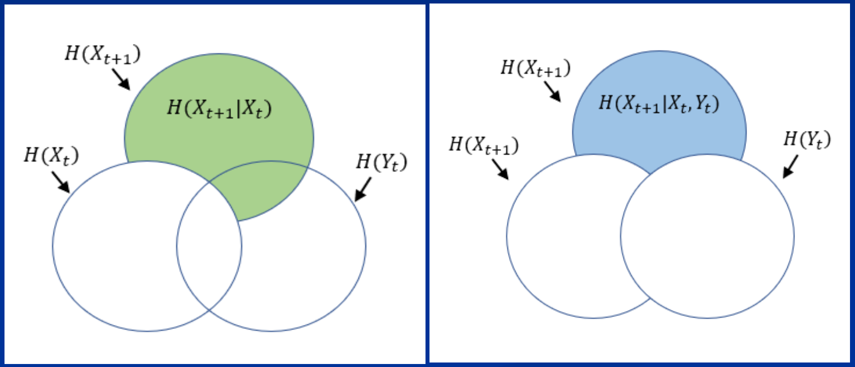
$$H(X) = \lim_{t\to\infty} H(X_t|X_{t-1}, X_{t-2}, \dots, X_1)$$

$$H(X) = \lim_{t\to\infty} H(X_t|X_{t-1})$$ *Markovian process*

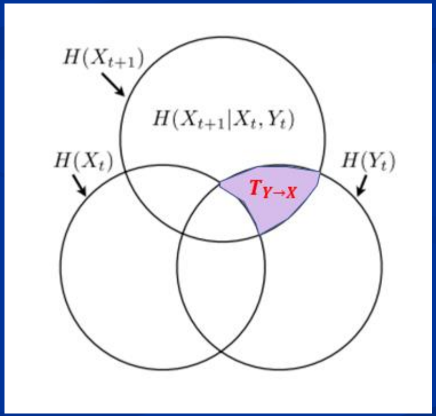
❖ *2 stochastic processes* $\{X_t\}, \{Y_t\}$

Transfer Entropy (from Y to X)

$$T_{Y\to X} = H(X_{t+1}|X_t) - H(X_{t+1}|X_t, Y_t)$$



- measures the reduction of uncertainty about X_{t+1}:
  - due to the information of the past states of **Y**
  - in addition to the information of the past states of **X**



J. Sun et al., Physica D 267 (2014) 49

Measures the extra information provided by Y_t (in addition to X_t) in the determination of X_{t+1}

$T_{Y\to X}$ does not necessarily equal to $T_{X\to Y}$

- *infer directionality of information flow and causality*

➢ *Application to relevant fusion diagnostic time series*

- ***Please follow the next presentation***