Nonlinear Analysis of Time Series
Part I: Univariate

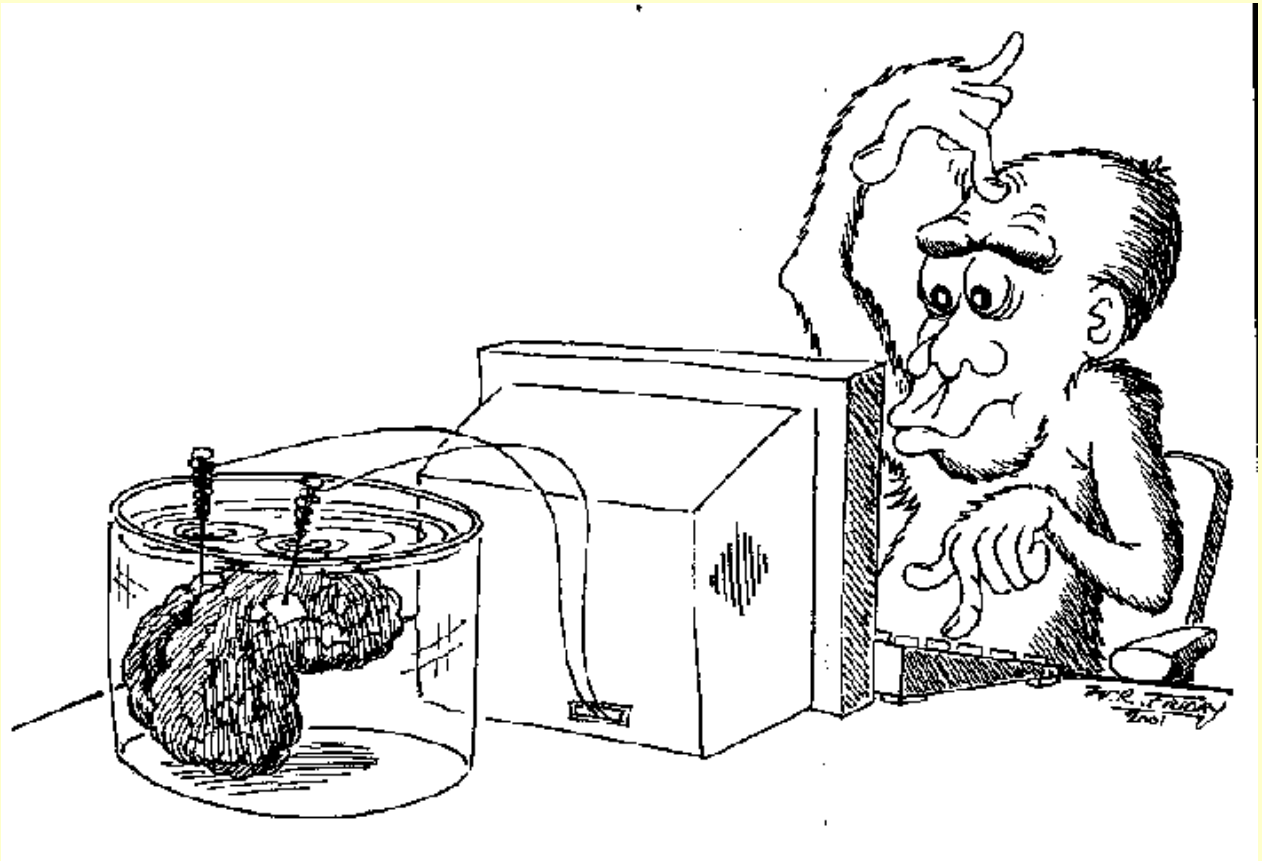
Dimitris Kugiumtzis

Department of Electrical and Computer Engineering,
Faculty of Technology, Aristotle University of Thessaloniki, Greece

e-mail: dkugiu@auth.gr
http: //users.auth.gr/dkugiu

23nd Summer School and Conference “Dynamical Systems and Complexity”,
Aristotle Univerity Camping, Kalandra, 30 August 2016

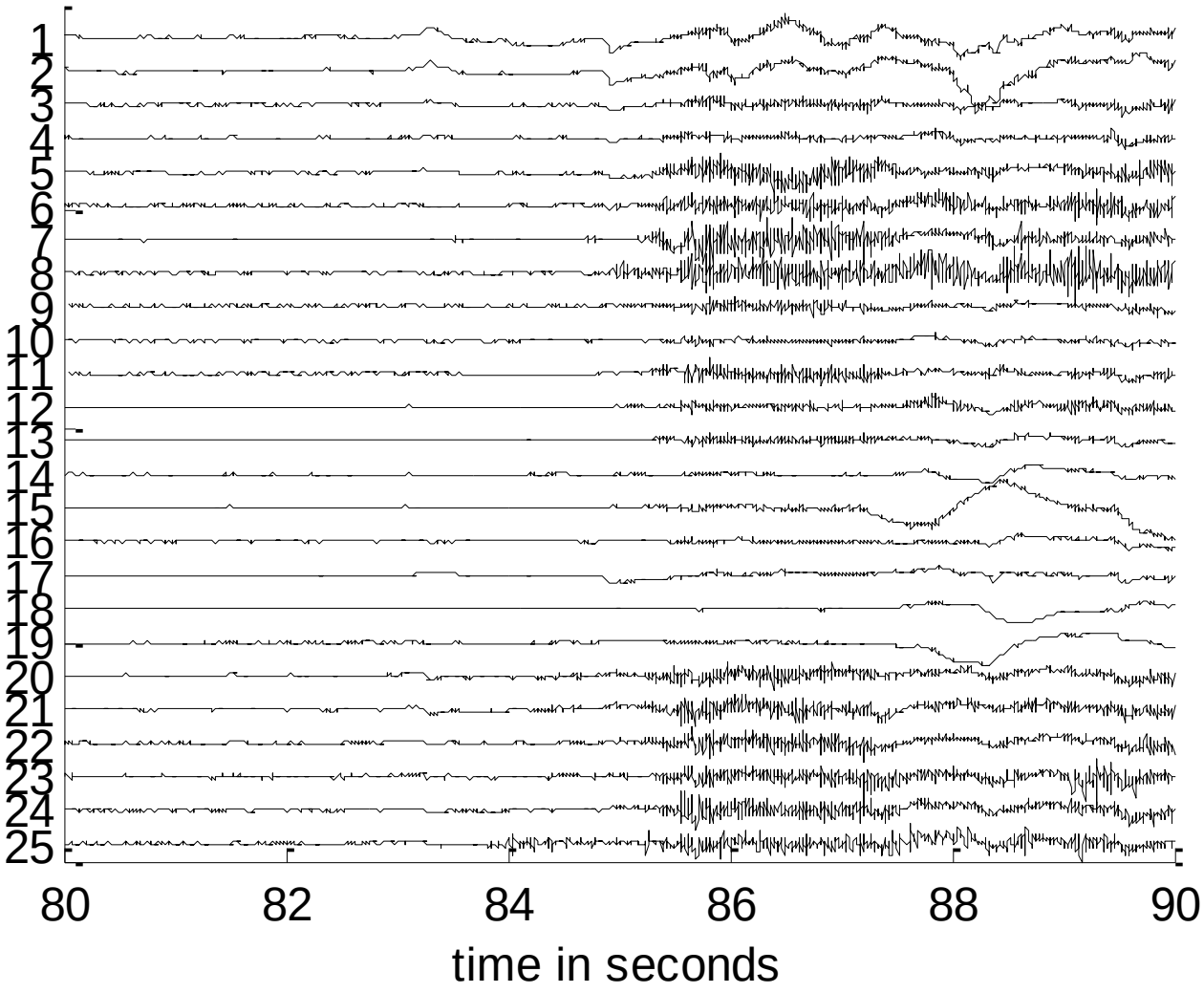
Measurements: What do I do with these?



http://www.gla.ac.uk/departments/philosophy/Undergraduate%20Resources/Honours/Honours%20Courses/JH3/Brain-in-vat.gif

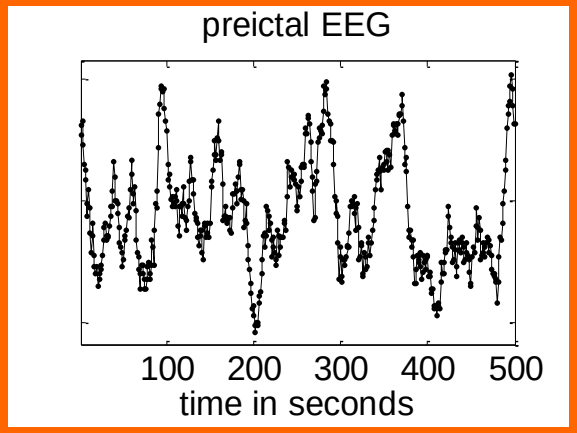
EEG measurements

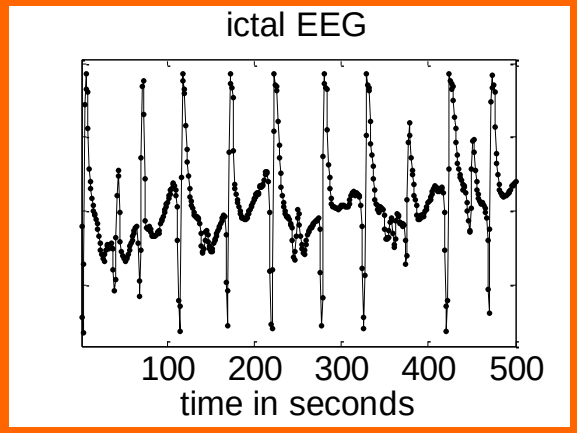
http://psg275.bham.ac.uk/bbs/symon-fac.htm



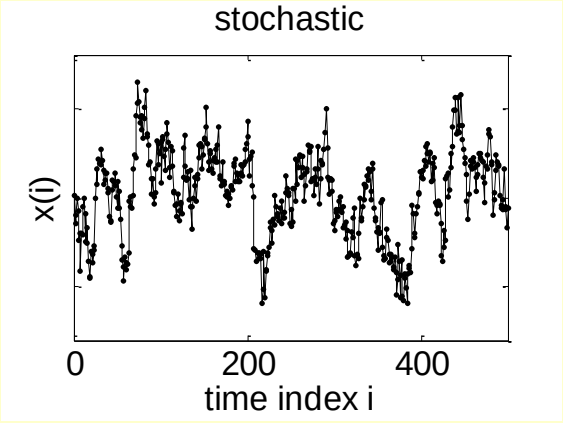
Underlying system for EEG

Real data

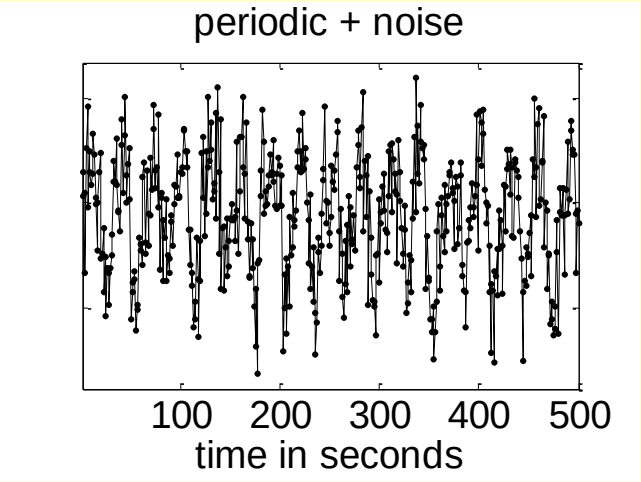


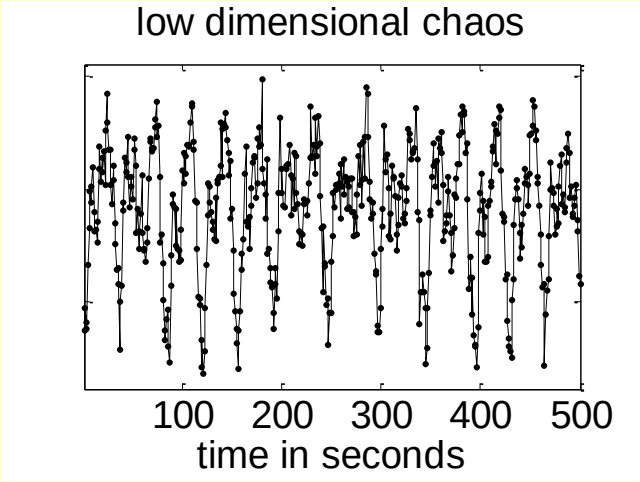


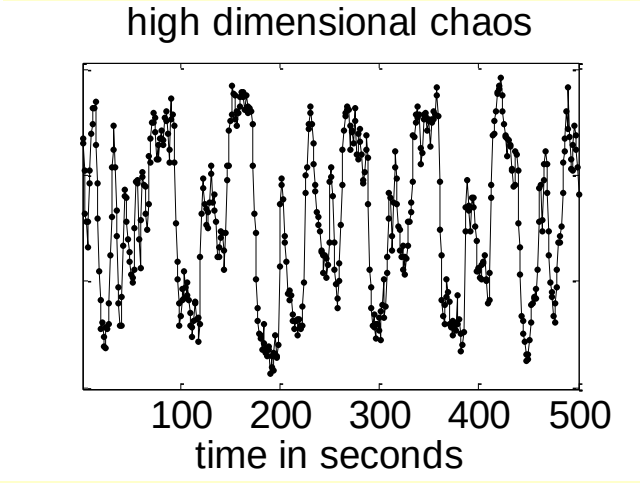
Model data (*under the stochastic perspective*)



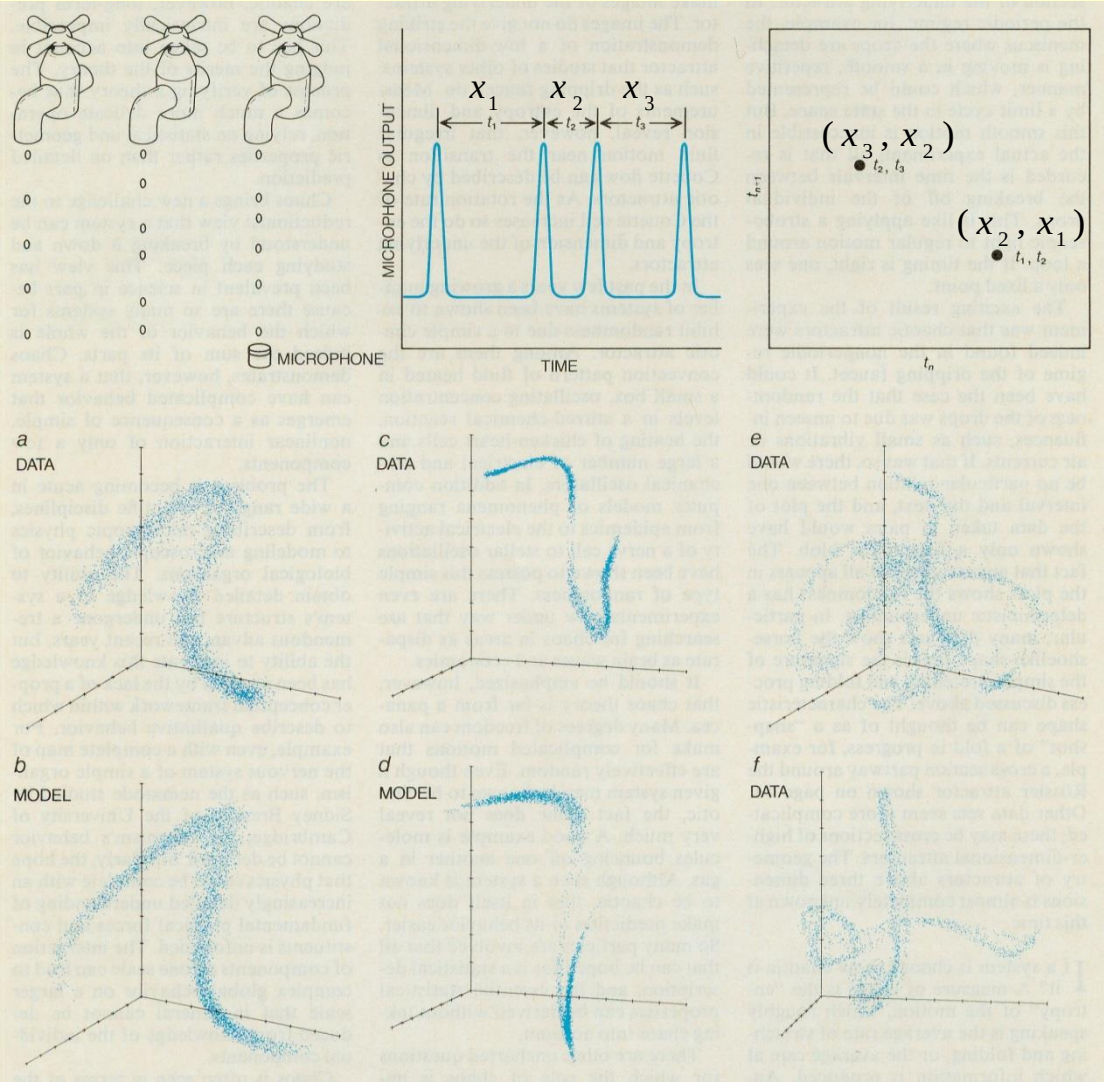
Model data (*under the deterministic perspective*)







Scientists at UC Santa Cruz found **chaos** in a **dripping water faucet**.



Crutchfield et al, Scientific American, 1986

By recording a dripping faucet and recording the periods of time, they discovered that at a certain flow velocity, the dripping no longer occurred at even times.

When they graphed the data, they found that the dripping did indeed follow a pattern.

scatter plot: (x_i, x_{i-1}) (x_i, x_{i-1}, x_{i-2})

Henon map

$$s_i = 1 - 1.4 s_{i-1}^2 + 0.3 s_{i-2}$$ chaos

Observed variable

$$x_i = s_i + e_i$$ e_i is noise

linear analysis

$\{x_t\}_{t=1}^{N}$ time series

Autocorrelation $r(\tau)$

$$r(\tau) = \frac{\frac{1}{N-\tau}\sum_{t=\tau+1}^{N}(x_t-\bar{x})(x_{t-\tau}-\bar{x})}{s_x^2}$$

$\longleftrightarrow$ power spectrum

$\updownarrow$

Autoregressive model AR(m)

$$x_{t+1} = \phi_0 + \phi_1 x_t + \cdots + \phi_m x_{t-m+1} + \varepsilon_{t+1} \qquad \varepsilon_t \sim \mathrm{iid}(0, \sigma_\varepsilon^2)$$

time-dependent coefficients

Fit / prediction statistics

NRMSE(m), CC(m)

Nonlinear correlation measure

Bi-correlation $r_3(\tau)$

$$r(\tau) = \frac{\frac{1}{N-\lambda}\sum_{t=\lambda+1}^{N}(x_t-\bar{x})(x_{t-\tau}-\bar{x})(x_{t-\lambda}-\bar{x})}{s_x^3}$$

($\lambda > \tau$)

Nonlinear correlation / information measures

Entropy information / uncertainty in variable(s)

For numerical time series, assume binning of $\{x_t\}_{t=1}^{N}$

Shannon Entropy for a discrete variable *X*

$$H(X) = -\sum_x p_X(x)\log p_X(x) = \left\langle -\log p_X(x)\right\rangle$$

… for two variables *X,Y*

$$H(X,Y) = -\sum_{x,y} p_{XY}(x,y)\log p_{XY}(x,y) = \left\langle -\log p_{XY}(x,y)\right\rangle$$

… for vector variable ***X***

$$H(\boldsymbol{X}) = -\sum_{\boldsymbol{x}} p_{\boldsymbol{X}}(\boldsymbol{x})\log p_{\boldsymbol{X}}(\boldsymbol{x}) = \left\langle -\log p_{\boldsymbol{X}}(\boldsymbol{x})\right\rangle$$

Generalization of Shannon entropy → **Tsallis entropy**

$$S_q(X) = \frac{1}{q-1}\left(1 - \sum_x \left(p_X(x)\right)^q\right)$$

Mutual information

Information on *X* from *Y* $\quad I(X,Y) = H(X) + H(Y) - H(X,Y)$

or $$I(X,Y) = \sum_{x,y} p_{XY}(x,y) \log \frac{p_{XY}(x,y)}{p_X(x)\,p_Y(y)} = \left\langle \frac{p_{XY}(x,y)}{p_X(x)\,p_Y(y)} \right\rangle$$

Binning estimates (equidistant, equiprobable, adaptive)

Kernel and Nearest neighbors estimates

$$I(X,Y) = \left\langle \frac{f_{XY}(x,y)}{f_X(x)\,f_Y(y)} \right\rangle = \sum_{t=1}^{N} \frac{f_{XY}(x_t,y_t)}{f_X(x_t)\,f_Y(y_t)}$$

$X \rightarrow x_t$

$Y \rightarrow x_{t-\tau}$

$\rightarrow$

$$I(X,Y) = I(\tau)$$

$$H(X,Y) = H(\tau)$$

$$S_q(X,Y) = S_q(\tau)$$

Mutual Information $I(\tau)$

Shannon entropy $H(\tau)$

Tsallis entropy $S_q(\tau)$

State Space Measures

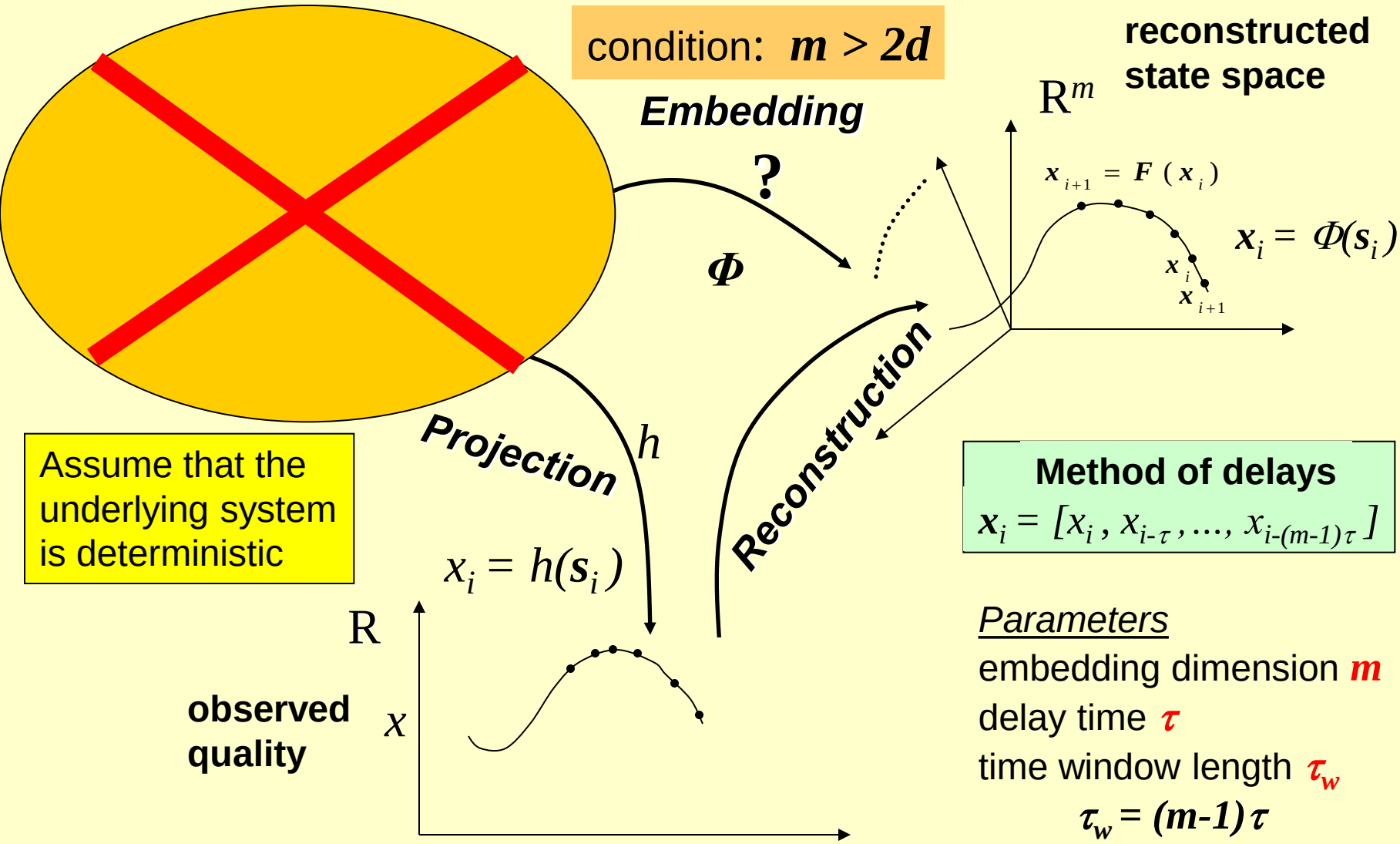
- **State space reconstruction**
  to view the complexity / stochasticity of the underlying system

- Estimation of system / attractor characteristics
  to quantify **complexity** and **dimension** of the underlying system
  - correlation dimension
  - Lyapunov exponents
  - …

- Modeling / prediction
  to **model** / **predict** the time series / underlying system

State space reconstruction

State space reconstruction (embedding)



Assume that the underlying system is deterministic

Method of delays

$x_i = [x_i, x_{i-\tau}, \ldots, x_{i-(m-1)\tau}]$

Parameters

embedding dimension m

delay time τ

time window length τ_w

$\tau_w = (m-1)\tau$

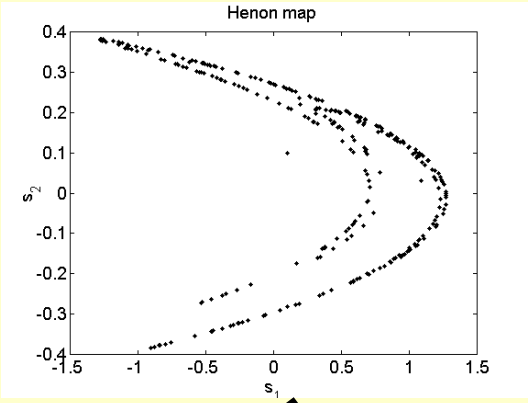
Example: Henon map (discrete)

$s(i) = 1 - 1.4\, s(i-1)^2 + 0.3 s(i-2)$

or

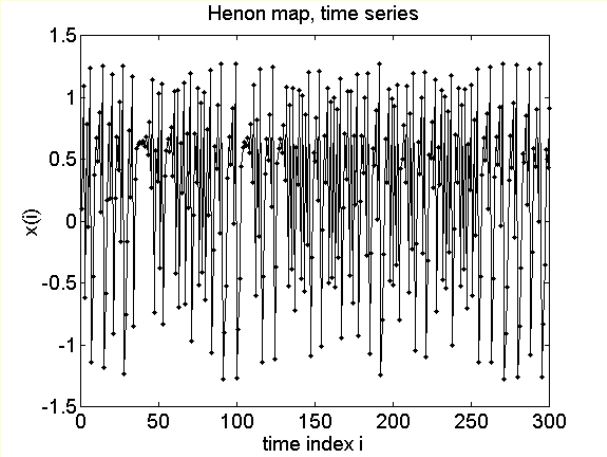
$s_1(i) = 1 - 1.4\, s_1(i-1)^2 + s_2(i-1)$

$s_2(i) = 0.3\, s_1(i-1)$



Projection

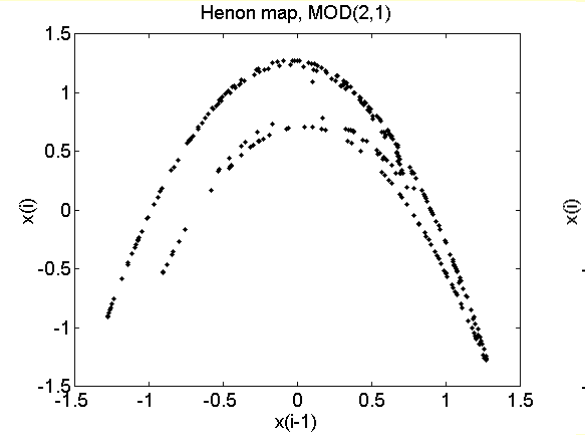
$x_i = s_1(i)$



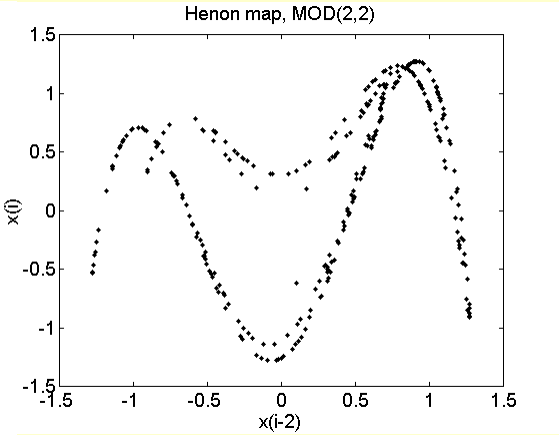
Reconstruction

Method of delays

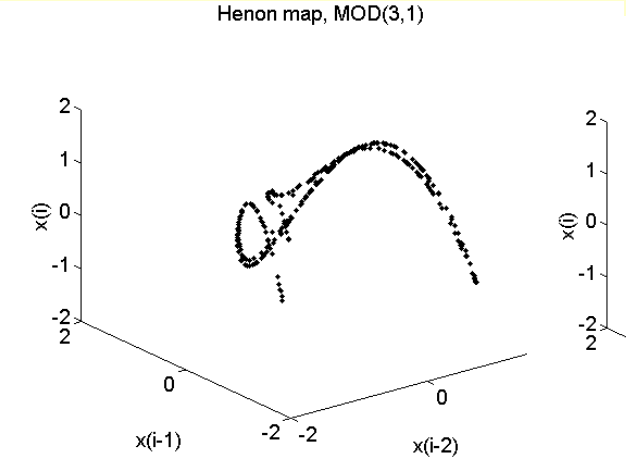
$m=2 \quad \tau=1$



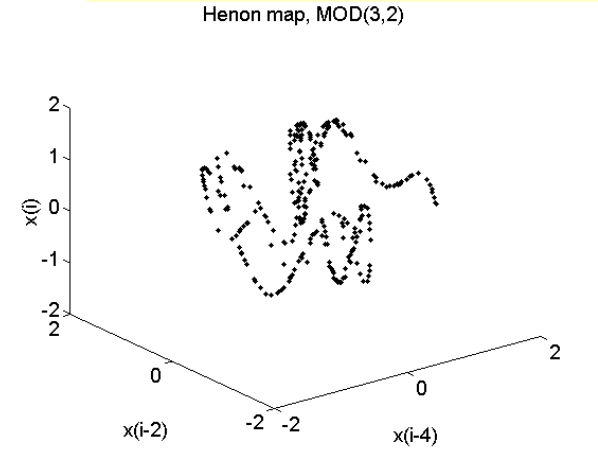
$m=2 \quad \tau=2$



$m=3 \quad \tau=1$



$m=3 \quad \tau=2$



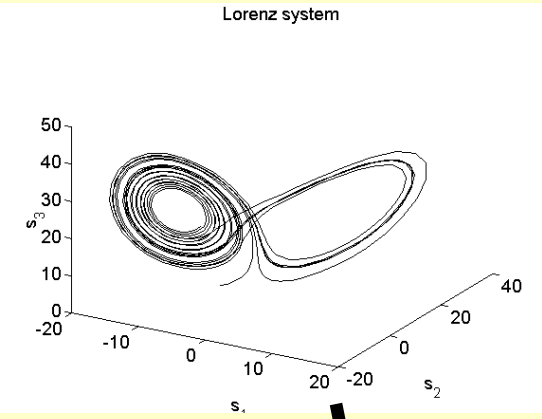
Example: Lorenz system (continuous)

$$\dot{s}_1 = -a(s_1 - s_2)$$
$$\dot{s}_2 = -s_1 s_3 + b s_1 - s_2$$
$$\dot{s}_3 = s_1 s_2 - c s_3$$

a=10, b=28, c=8/3

Optimal τ ?

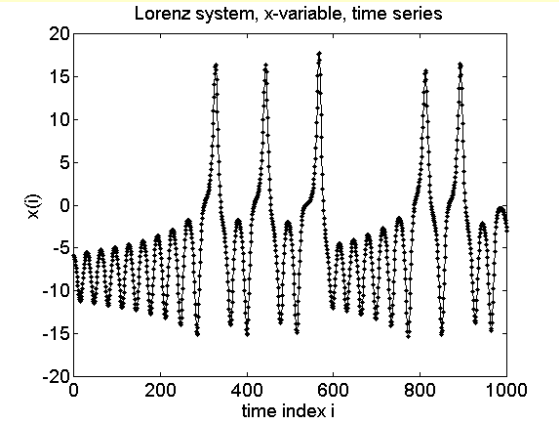
Method of delays, ***m=3***



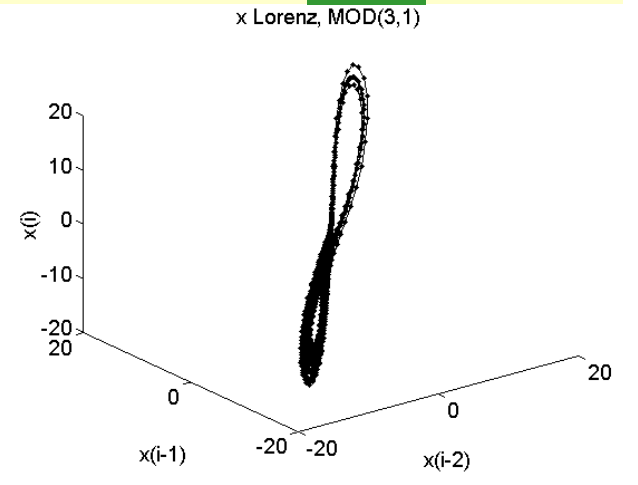
Projection

$x_i = s_1(i)$

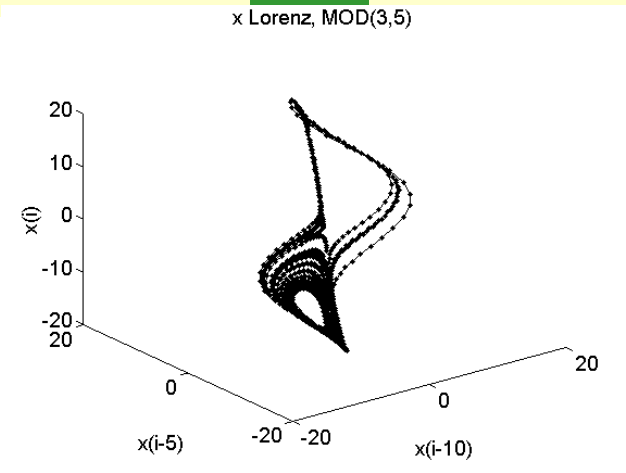
Reconstruction



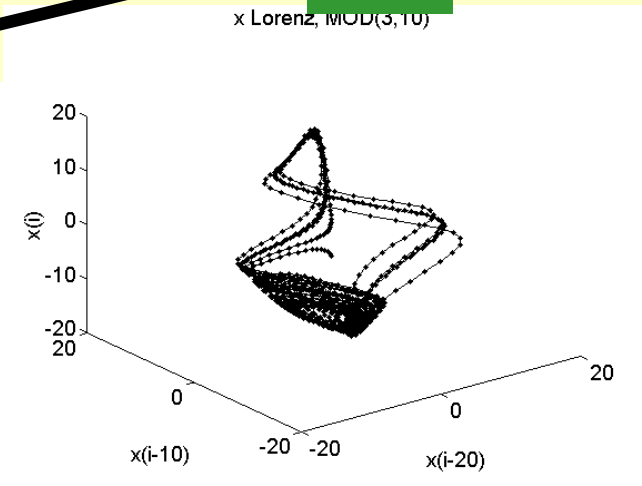
τ=1



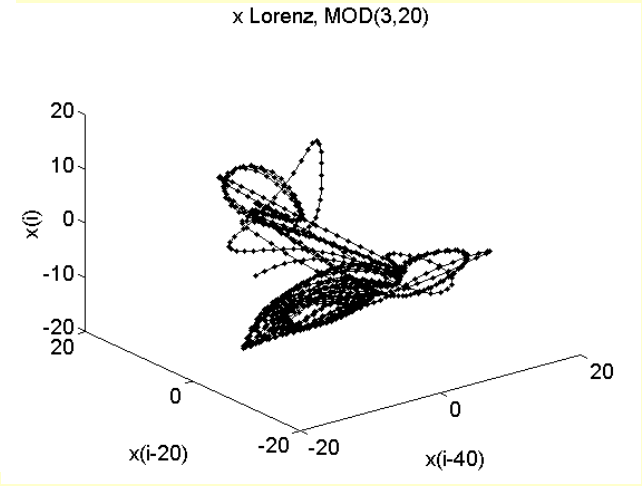
τ=5



τ=10



τ=20



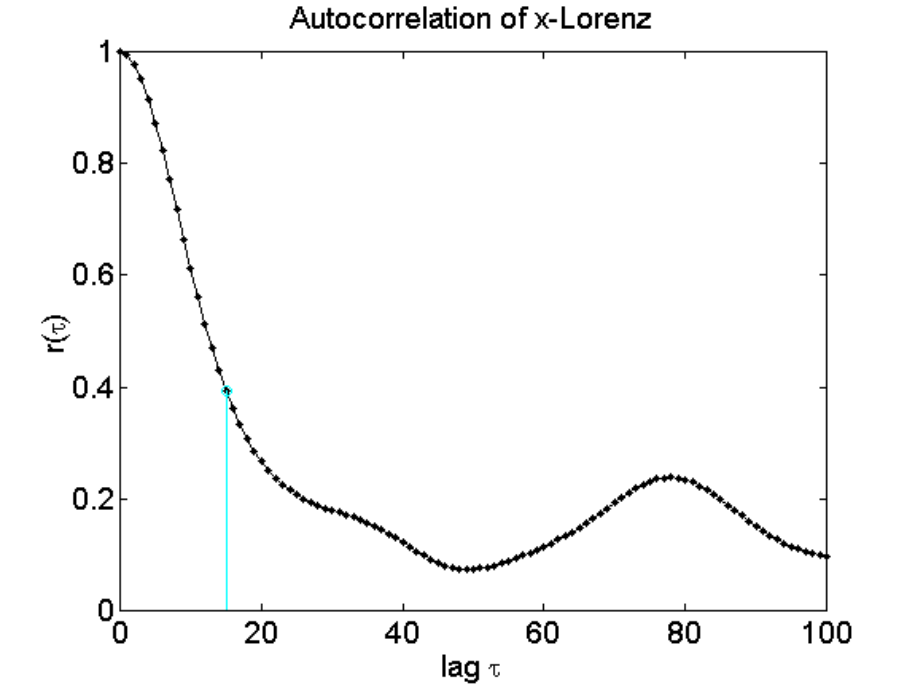
Estimation of τ

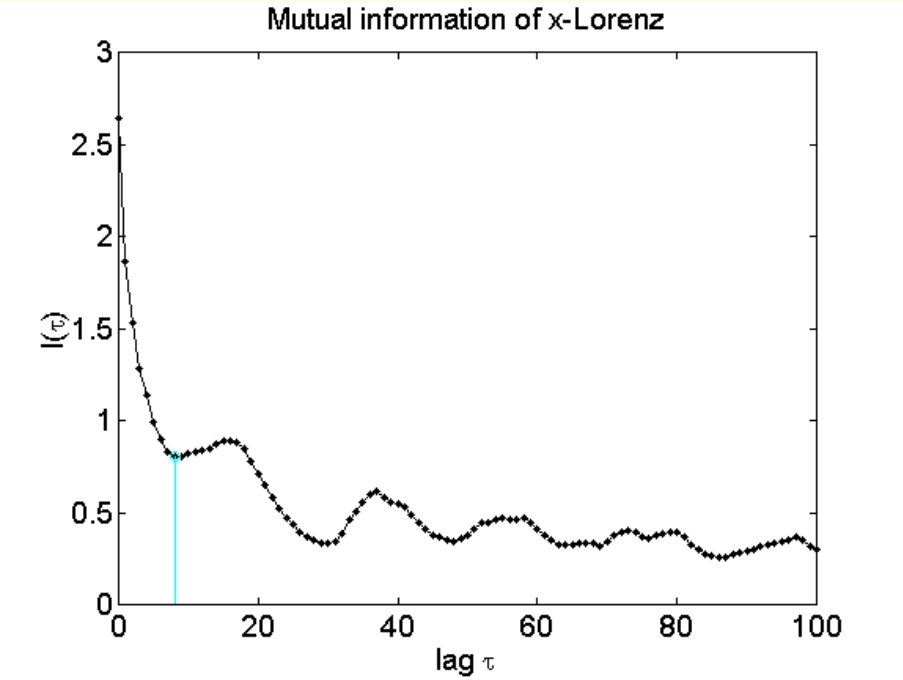
From **autocorrelation $r(\tau)$**

τ from $r(\tau) = 1/e$ or $r(\tau) = 0$

From **mutual information $I(\tau)$**

τ from the first min of $I(\tau)$





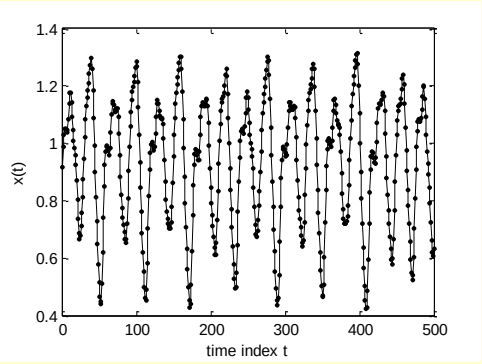
Estimation of τ - *toy models*

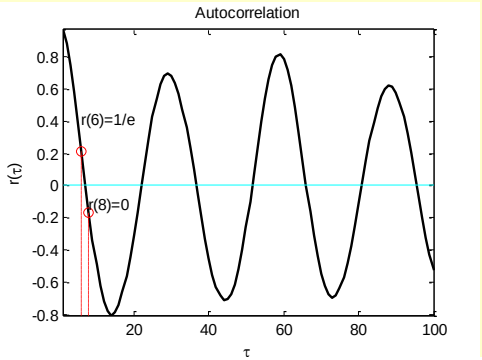
$$\dot{s}(t) = \frac{0.2\,s(t-\Delta)}{1+s(t-\Delta)^{10}} - 0.1\,s(t)$$

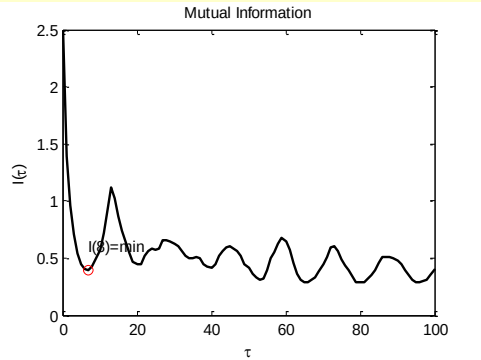
Mackey-Glass delay differential equation

Complexity increases with Δ parameter

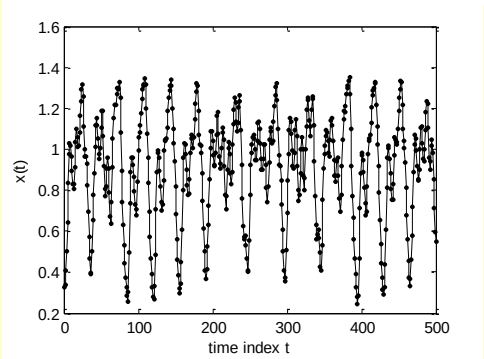
Δ=17

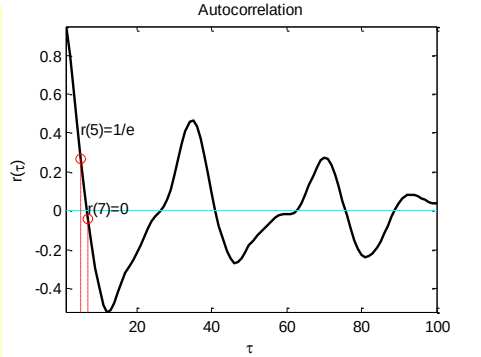


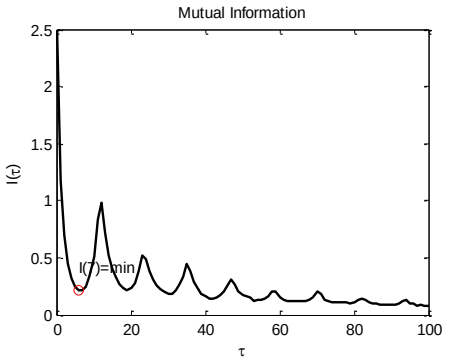




Δ=30

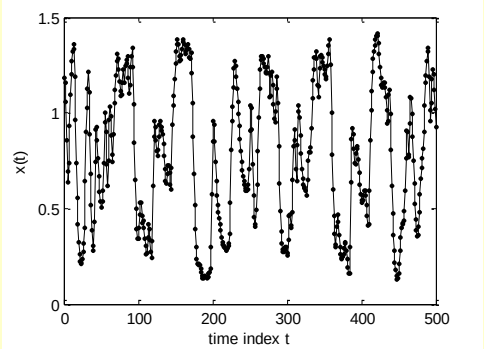


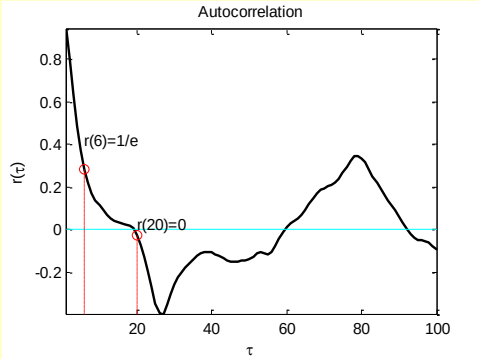


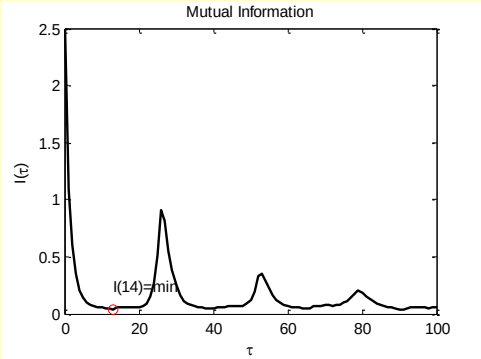


Good agreement from $r(\tau)$ and $I(\tau)$

Δ=100

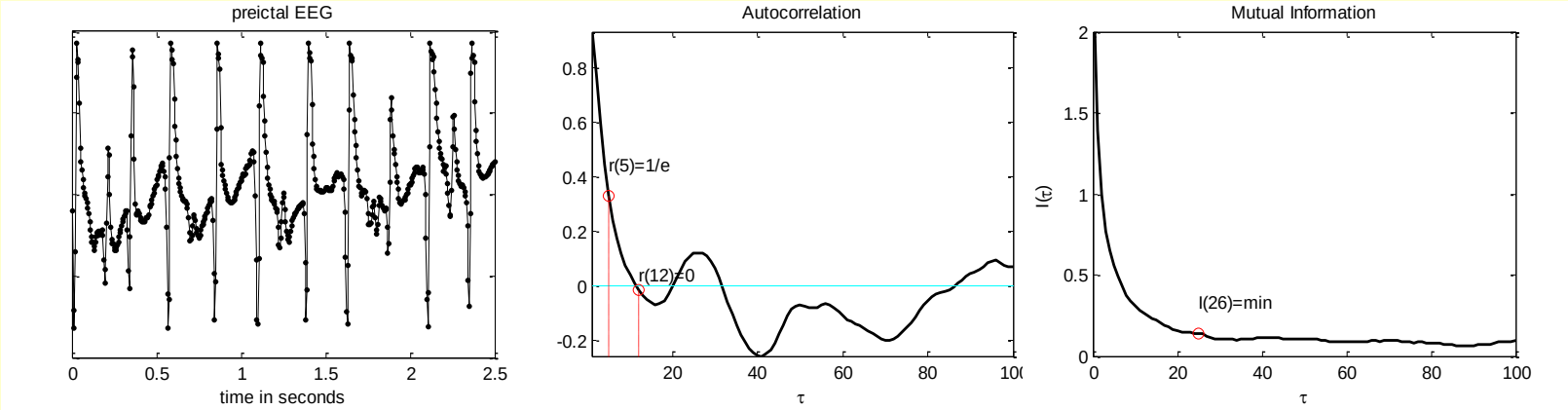




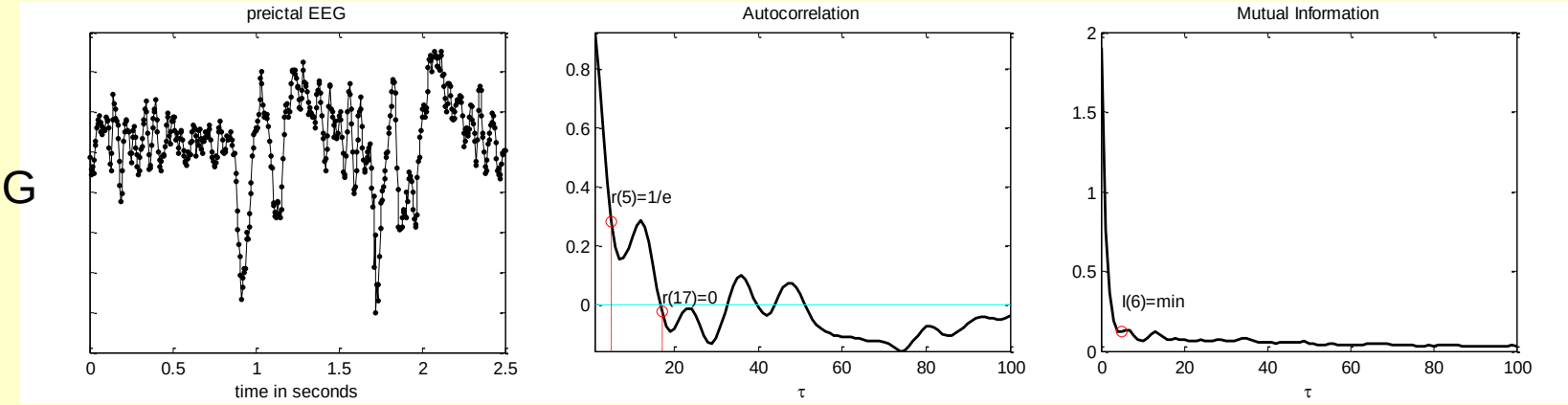


Estimation of τ - *EEG*

ictal EEG

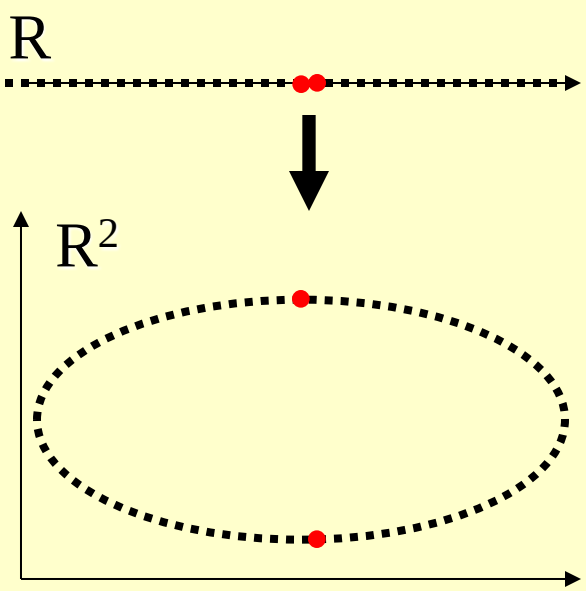


preictal EEG



No unique optimal delay time τ

Estimation of m



Optimal m ?

- If m is too small, the attractor displays self intersections
- If m is too large, then "curse of dimensionality"
- Takens' theorem: $m > 2d$, but d is not known

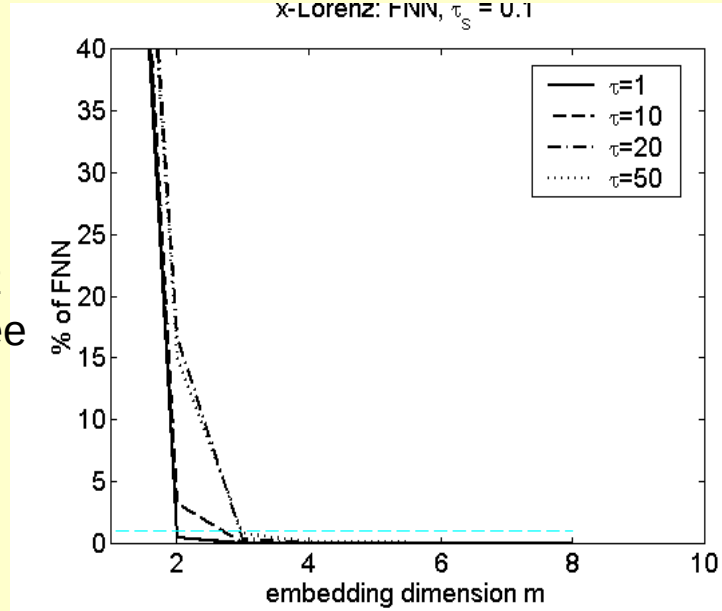
Method of **false nearest neighbors (FNN)**

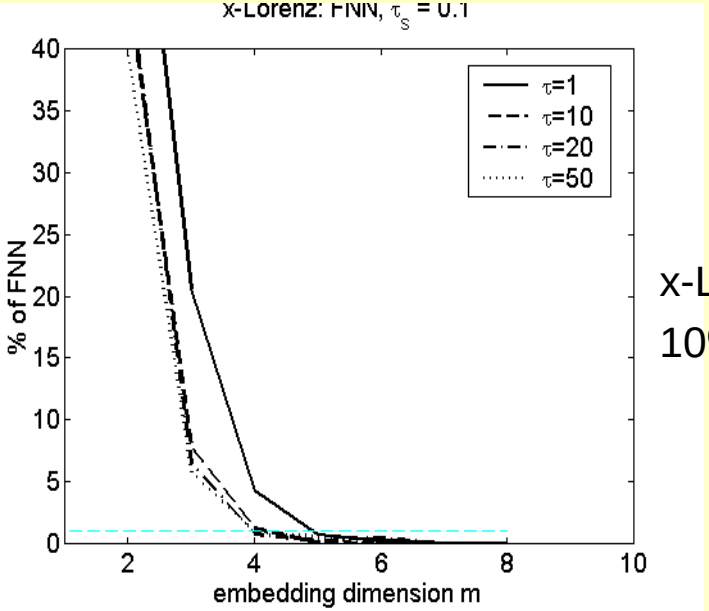
- Spatially nearby points on the attractor are either *real* neighbors due to the system dynamics or *false* neighbors due to self-intersections.
- In a higher dimension, where the self intersections are resolved, the false neighbors are revealed as they are not neighbors any more.
- An optimal m is estimated for which no false neighbors are found as the dimension increases beyond m.

Other estimates of m...

Example of estimation of m by FNN

x-Lorenz
noise-free





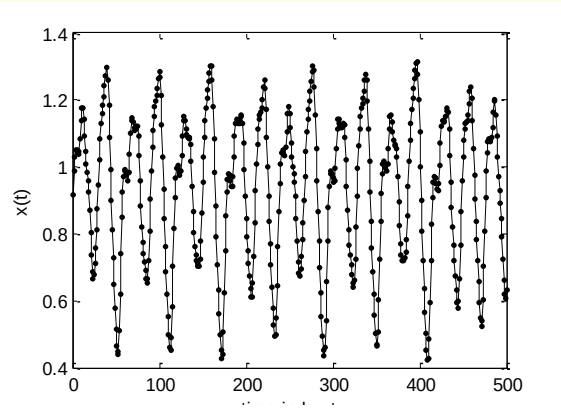
x-Lorenz +
10% noise

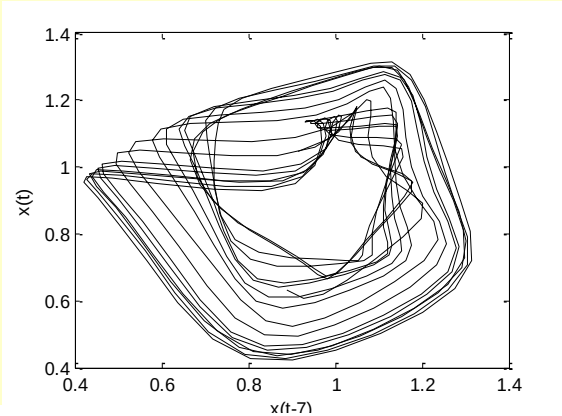
The FNN estimate of optimal m depends on

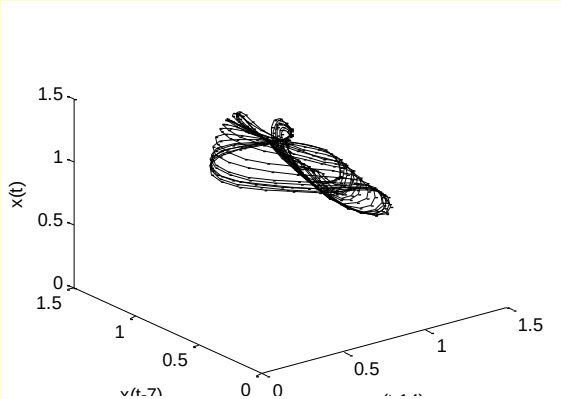
- **delay τ**
- **noise**

Dimension of attractor – toy models (Mackey-Glass)

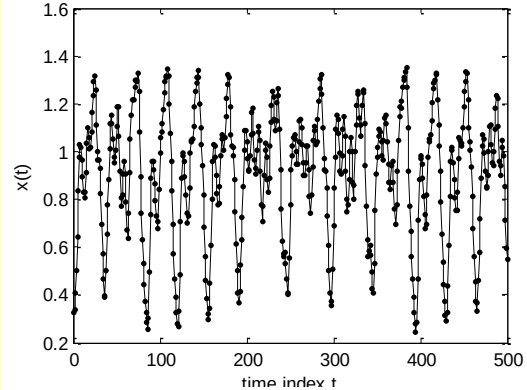
Δ=17

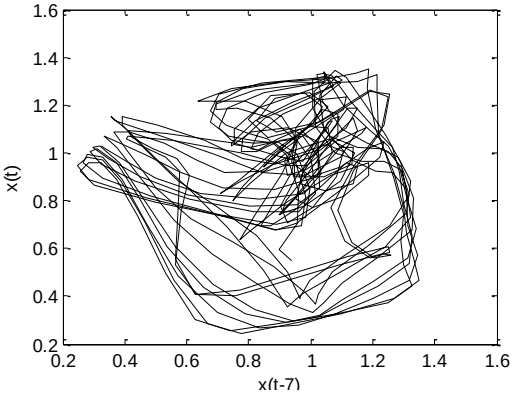


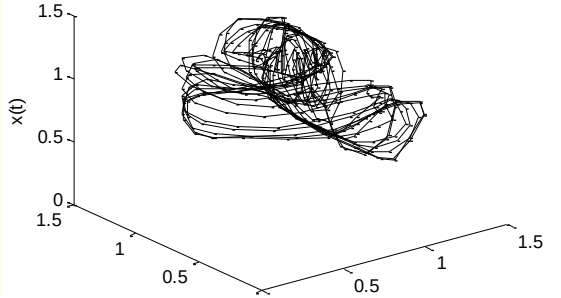




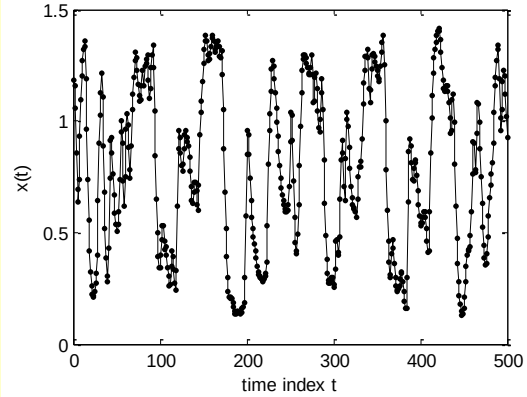
Δ=30

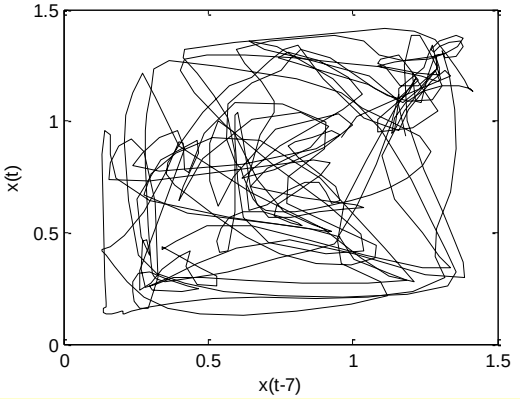


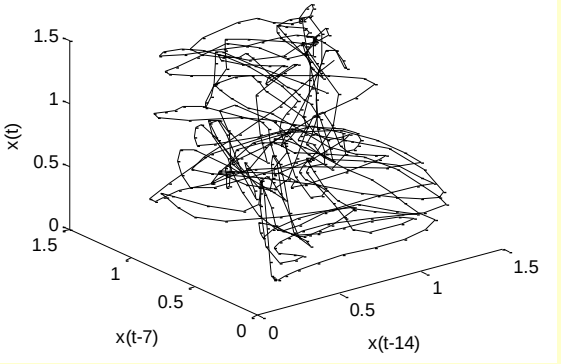




Δ=100

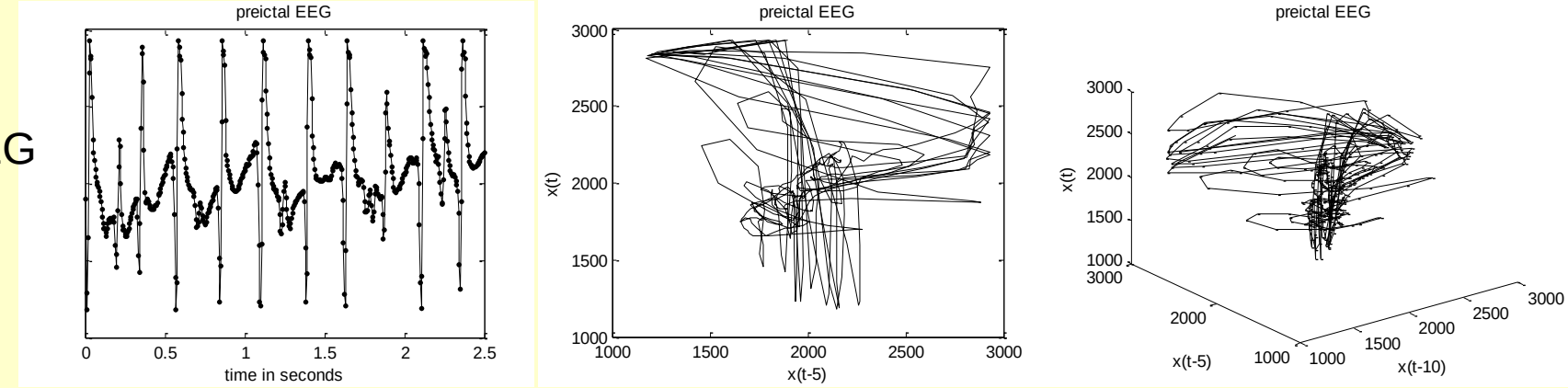




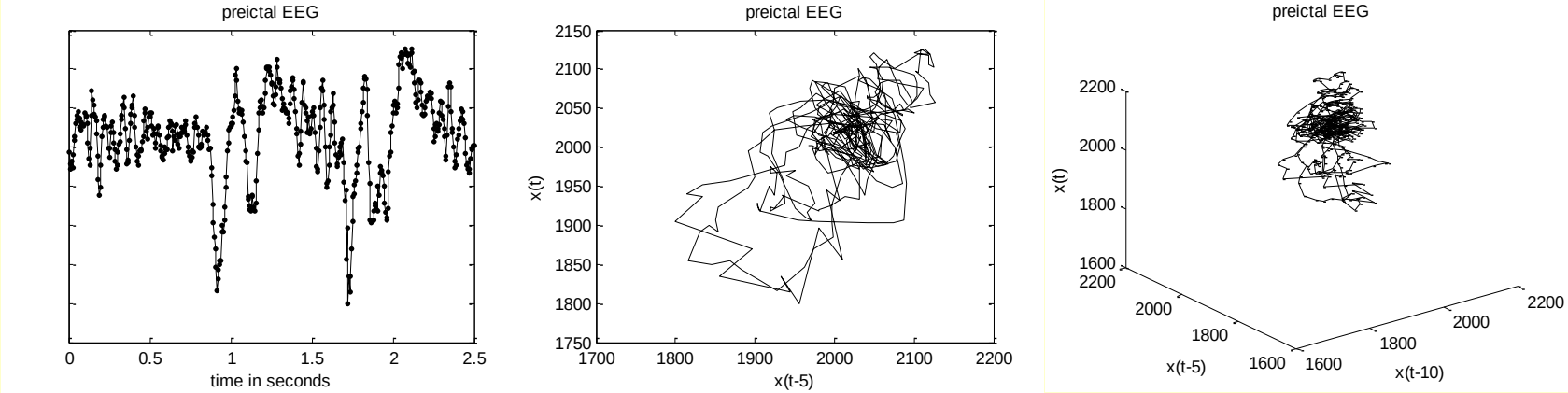


Dimension of attractor – EEG

ictal EEG



preictal EEG



What is the dimension (degrees of freedom) of the underlying system?

Correlation dimension ν

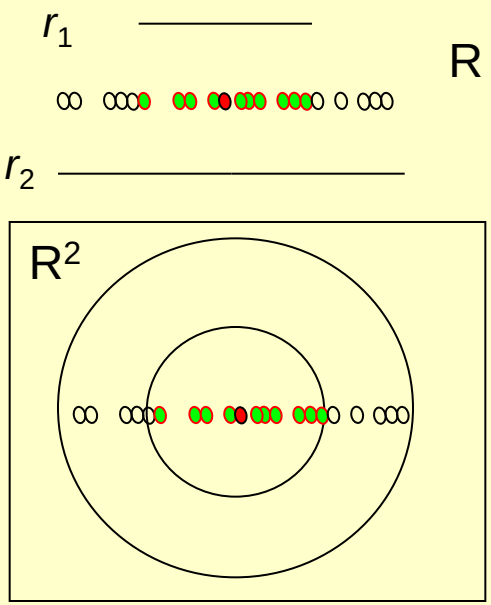
Correlation dimension characterizes the fractal structure of the attractor (self-similarity in different scales), using the density of the points of the attractor in the state space

The idea is that the "density" $p(r)$ for a typical r-ball covering part of the attractor scales with its radius like $p(r) \sim r^D$, where D is the dimension

Example: D=1

r_1=1 → interval contains 10 points

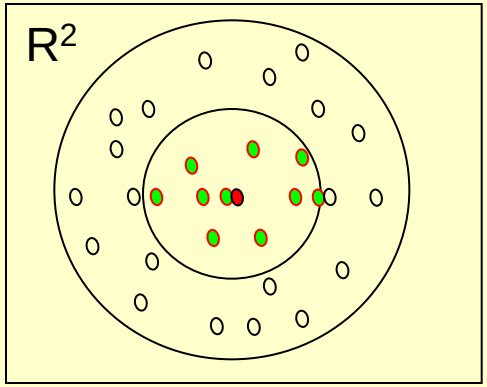
r_2=2 → interval contains 20 points



Example: D=2

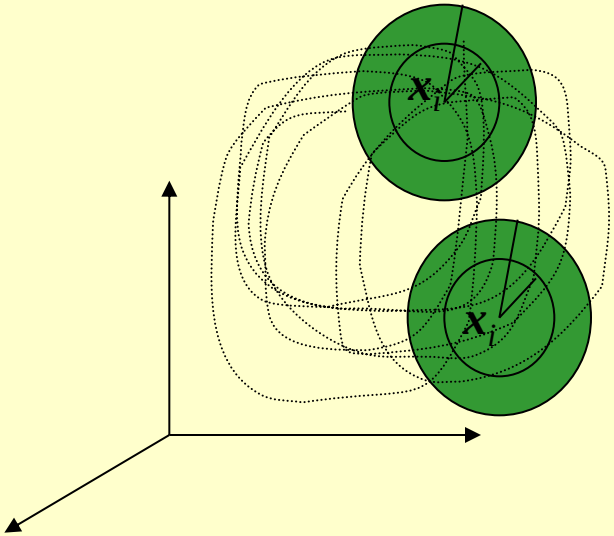
r_1=1 → circle contains 10 points

r_2=2 → circle contains 40 points



D=?

Correlation dimension ν



time series $\{x_1, x_2, \ldots, x_n\}$

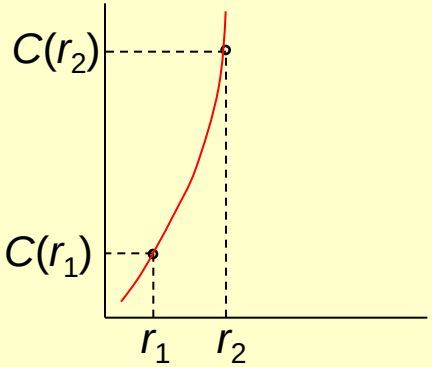
Method of delays

$$\mathbf{x}_i = [x_i, x_{i-\tau}, \ldots, x_{i-(m-1)\tau}]$$

reconstructed trajectory (attractor) $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_{n'}\}$

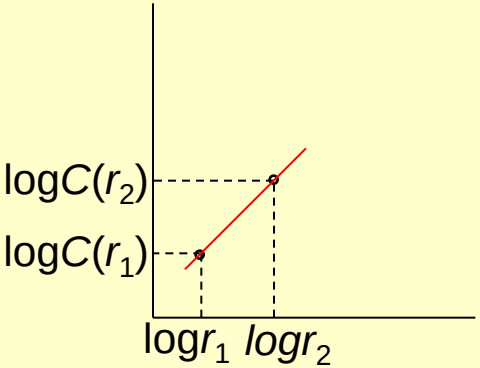
Correlation sum

$$C(r) = \frac{1}{N_{pairs}} \sum_i \sum_j \Theta(r - \|\mathbf{x}_i - \mathbf{x}_j\|)$$



Scaling law $C(r) \propto r^{\nu}$ for r small

Estimation $\nu = \dfrac{\log C(r)}{\log r}$ for a range of small r



Convergence of $\nu(m)$ as $m > d$

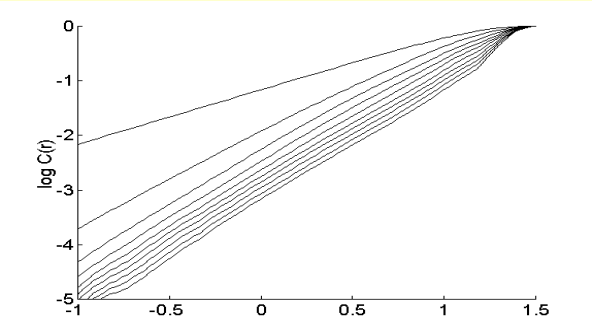
If ν small / non-integer **and** system is deterministic

$\Downarrow$

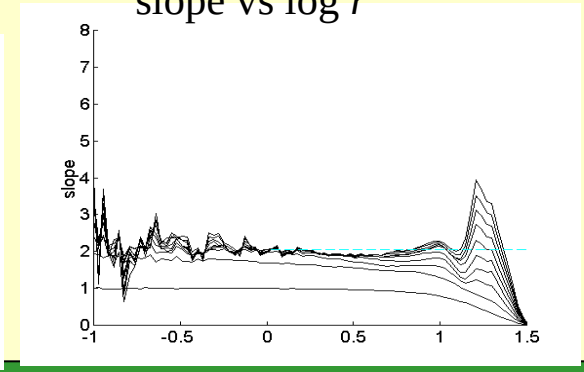
low-dimensional / fractal structure (chaos)

x-Lorenz + noise-free, τ=2

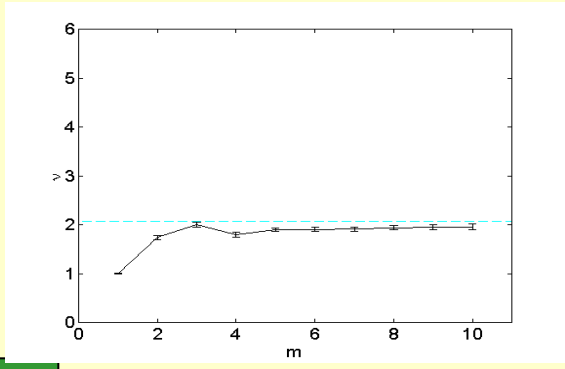
log $C(r)$ vs log r



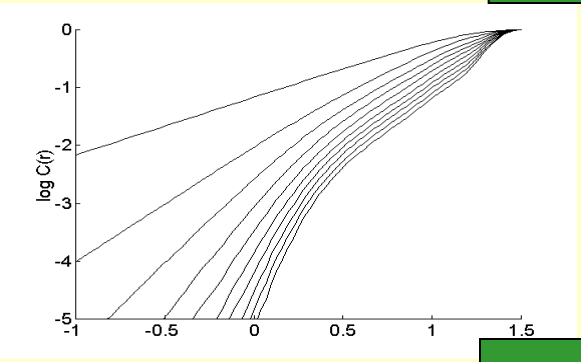
slope vs log r

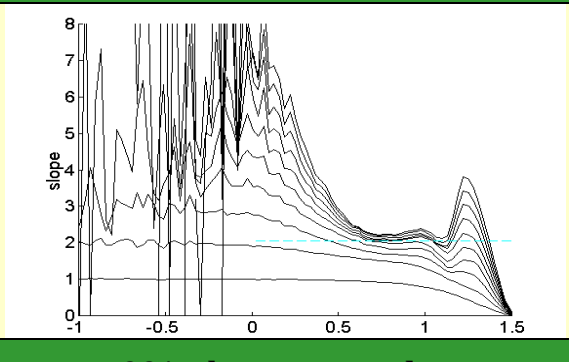


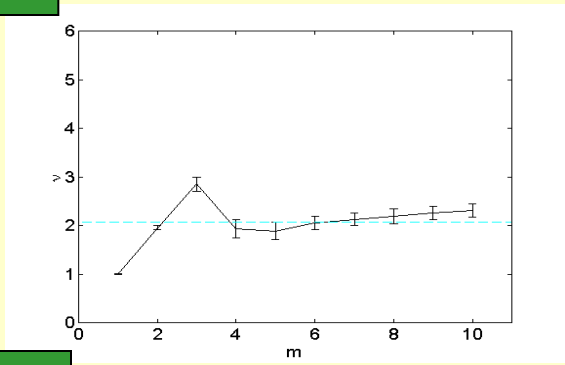
ν vs m



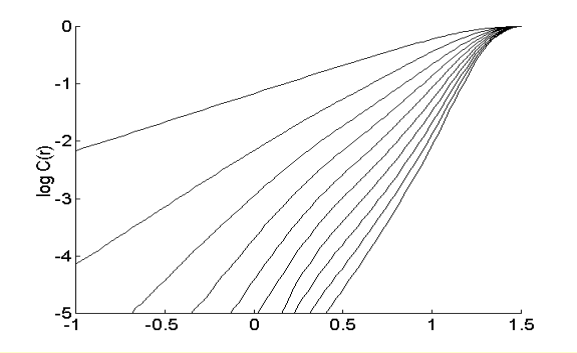
x-Lorenz + 10% observational noise, τ=2

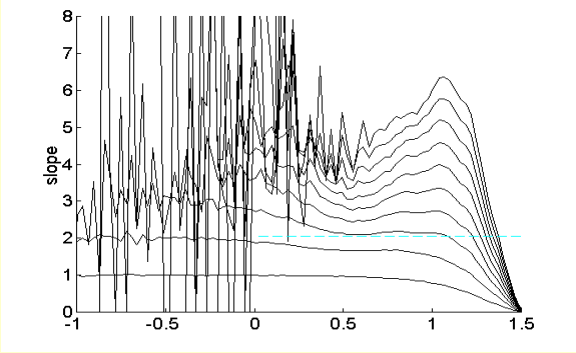


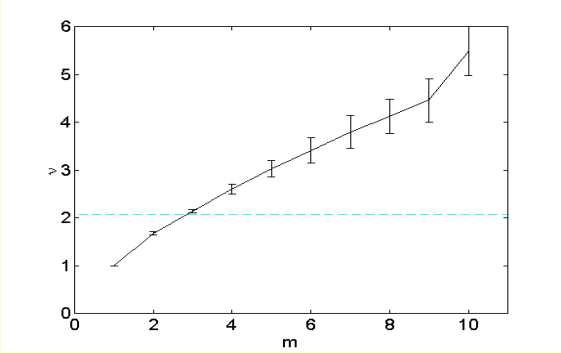




x-Lorenz + 10% observational noise, τ=10





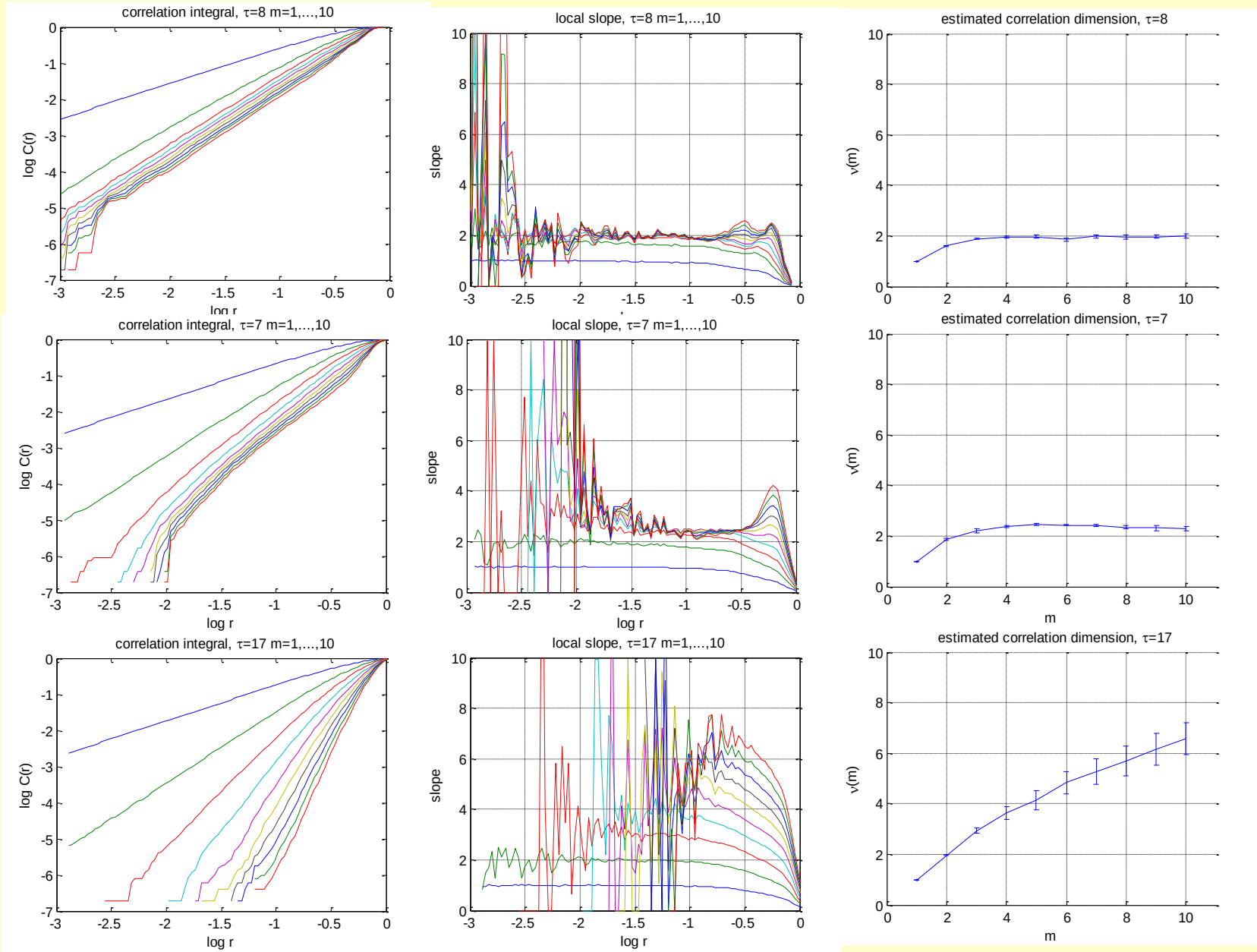


Estimation of correlation dimension – toy models (Mackey-Glass)

Δ=17

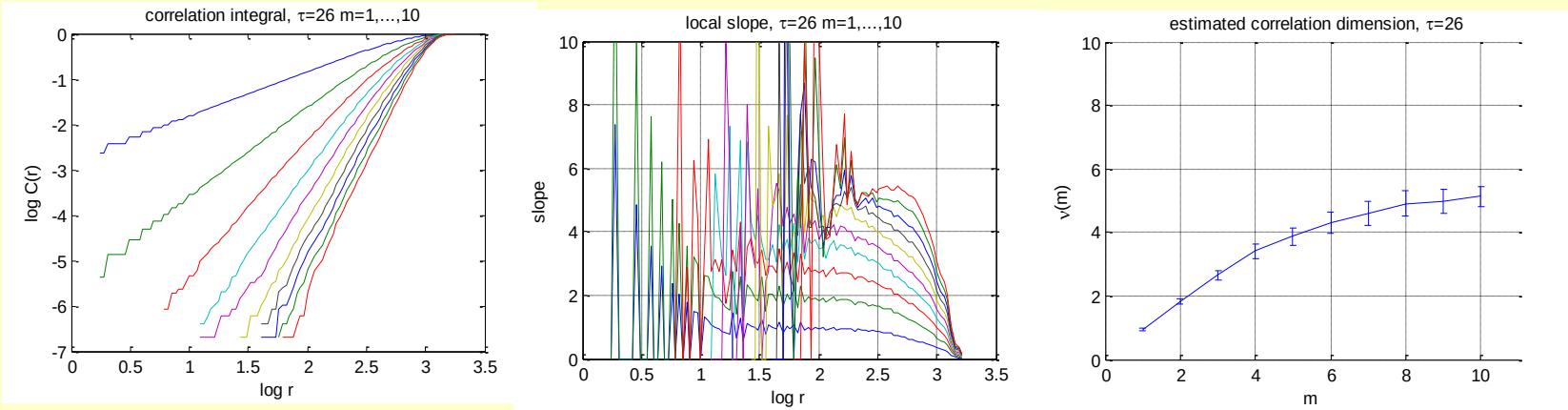
Δ=30

Δ=100

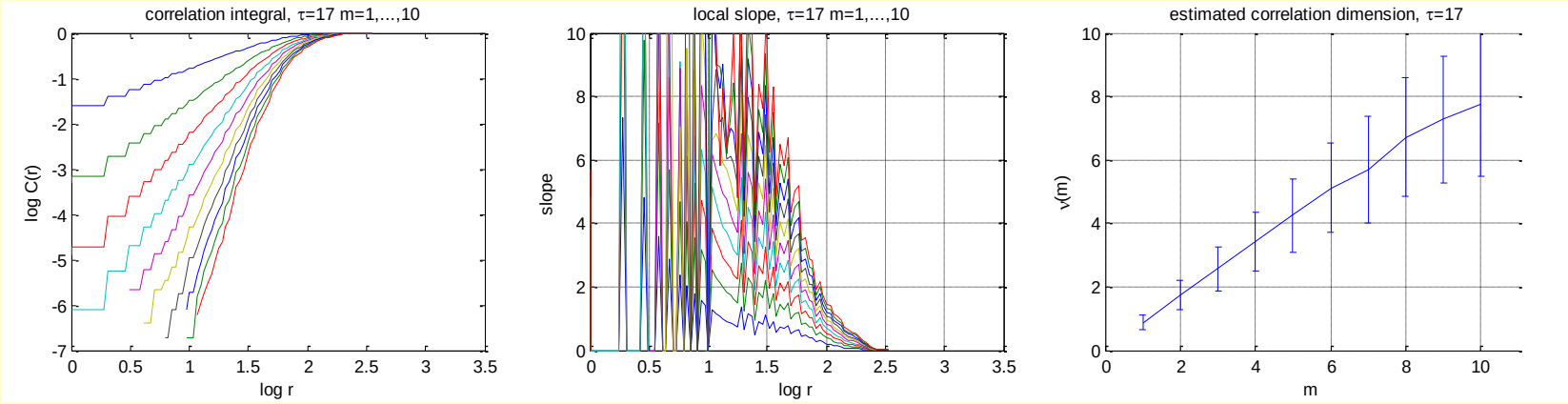


Estimation of correlation dimension – EEG

ictal EEG



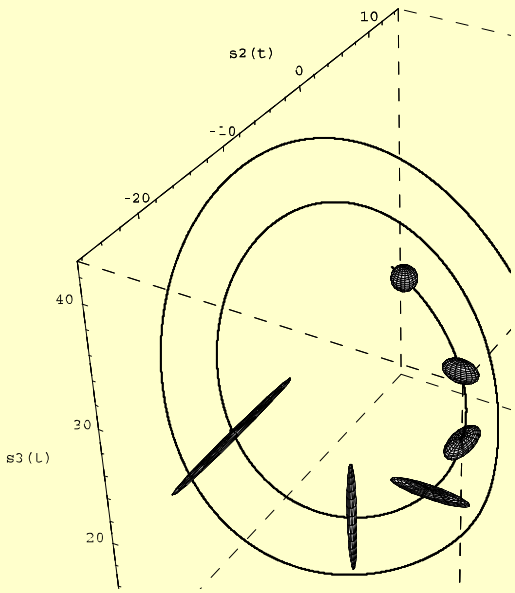
preictal EEG



No reliable estimation of correlation dimension, maybe only for ictal EEG

Lyapunov exponents

Lyapunov exponents are average rates of stretching or contraction over the attractor, in the directions of the locally decomposed state space



Lyapunov spectrum: $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_m$

$\lambda_i > 0$ ➔ stretching

$\lambda_i < 0$ ➔ contraction

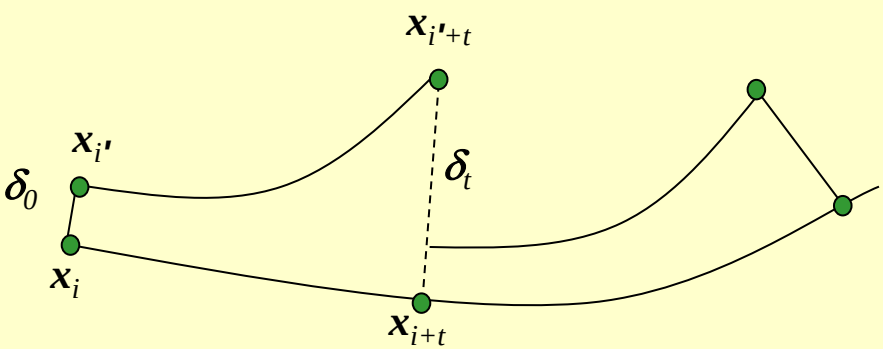
$\lambda_i = 0$ ➔ along the flow

If $\lambda_i > 0$ **and** system is deterministic

⇩

chaos

Largest Lyapunov Exponent λ_1 (LLE)



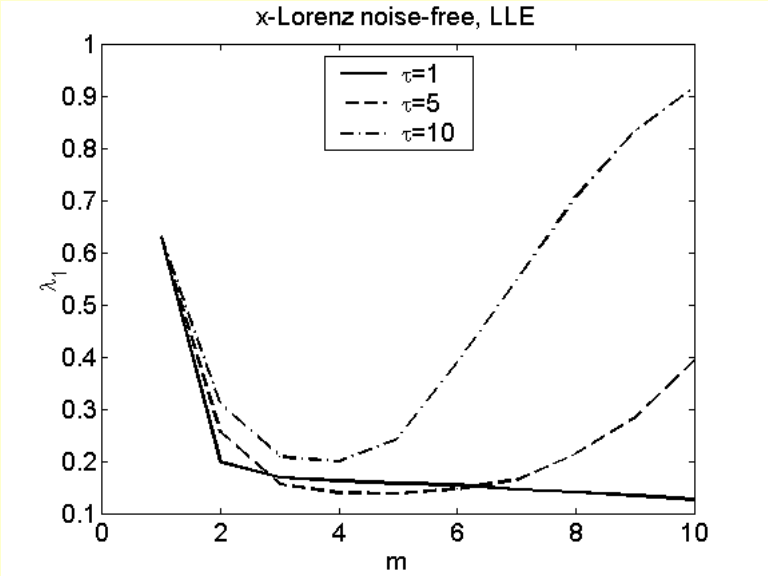
Distance $\delta_0 = x_i - x_{i'}$ small perturbation **should** grow exponentially in time

After time t: $\delta_t = x_{i+t} - x_{i'+t}$

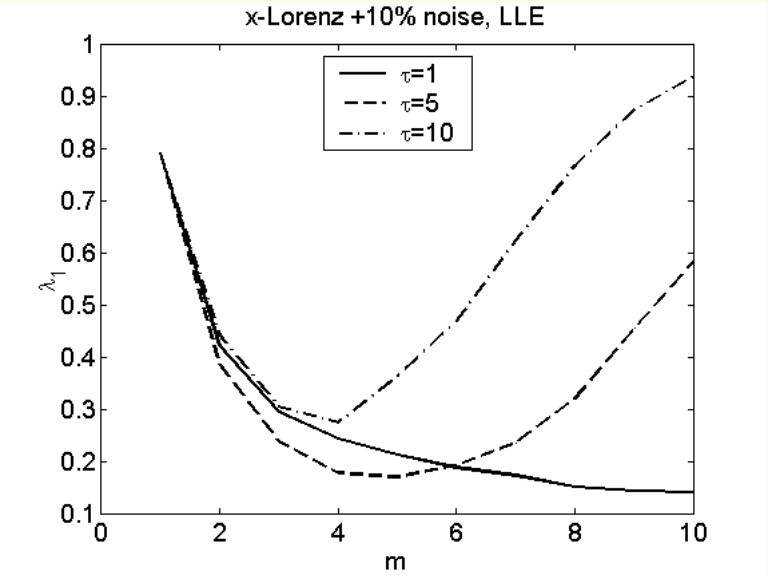
If $\delta_t \approx \delta_0 e^{\lambda_1 t}$ $\Longrightarrow$ λ_1 is LLE

Example: x-Lorenz

noise-free

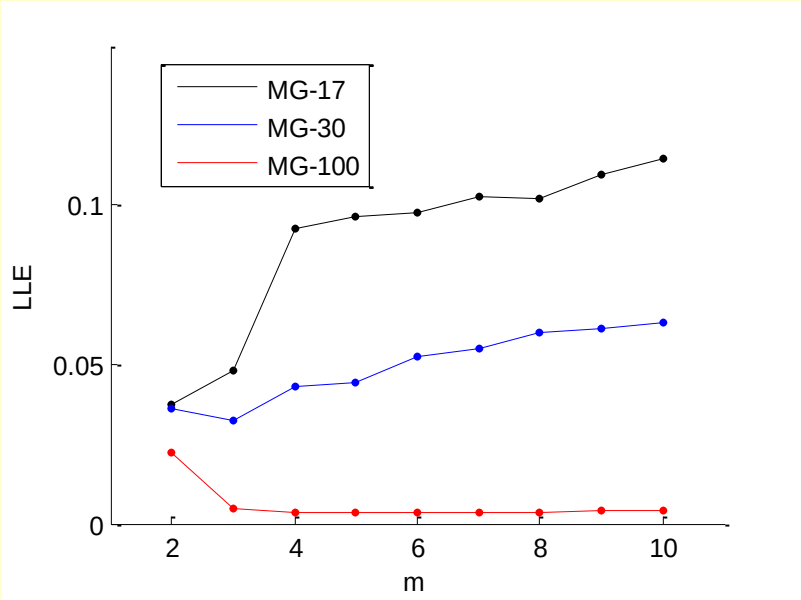


with 10%-noise

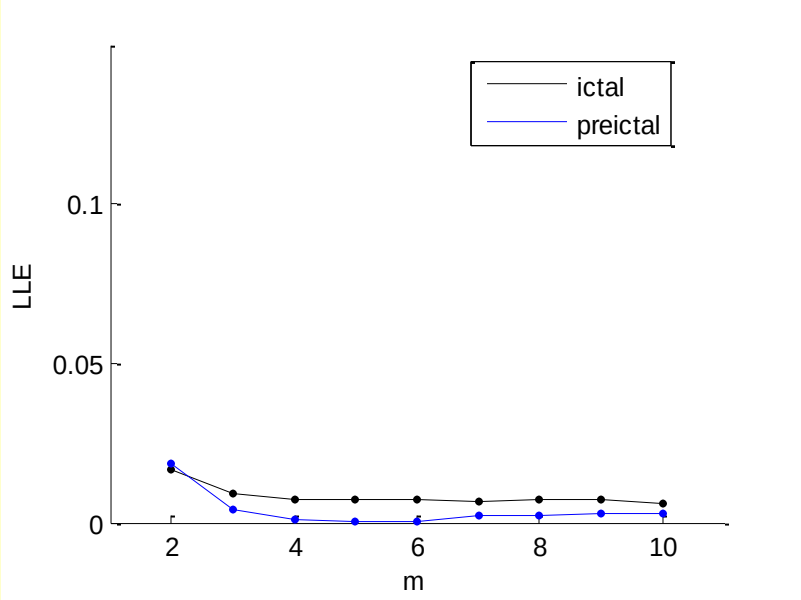


Largest Lyapunov Exponent (LLE) estimation

toy models



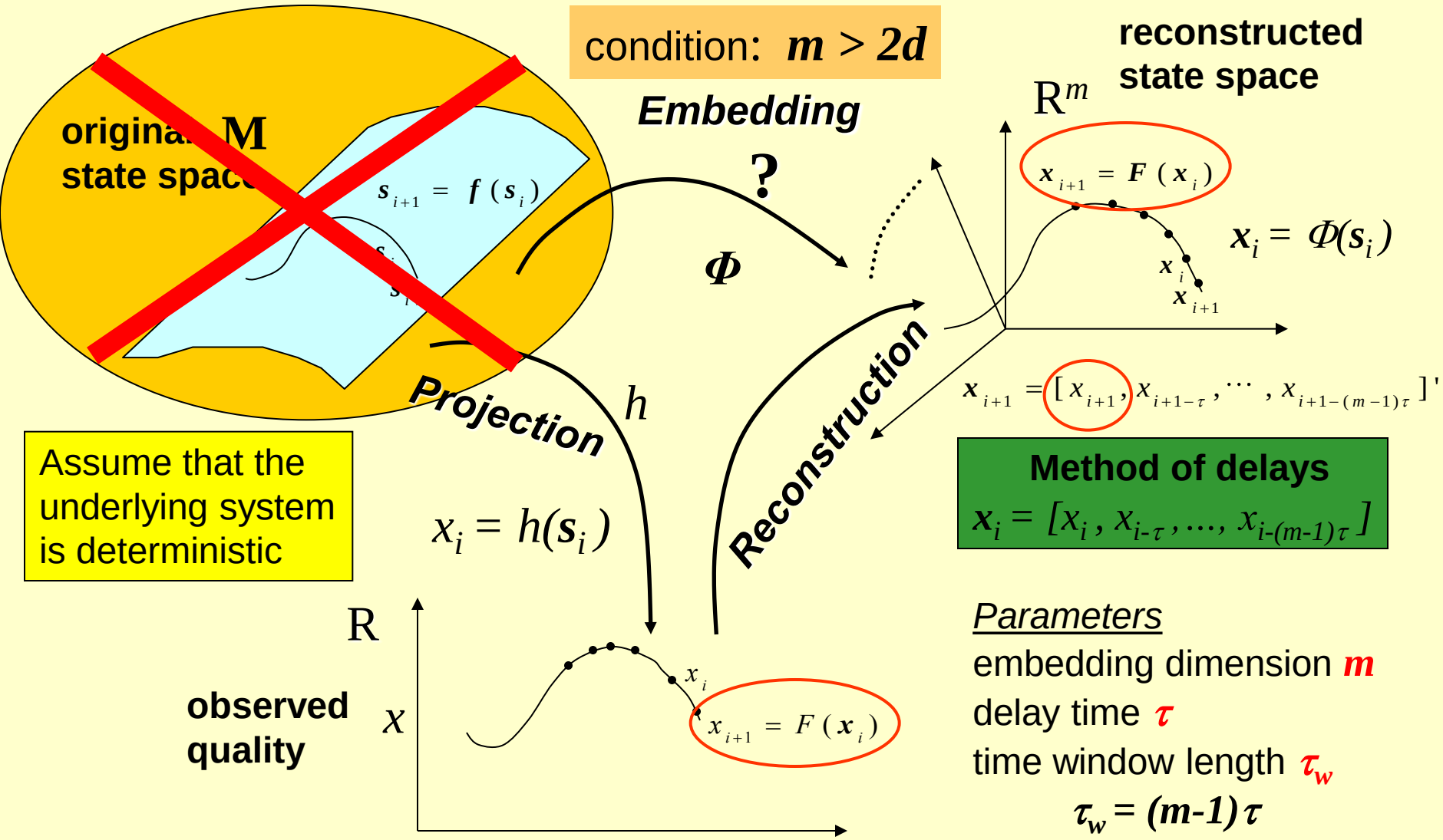
EEG



No reliable estimation of LLE

However, the characteristics of the systems (correlation dimension, LLE) can be used as indices / measures that can distinguish states of the EEG

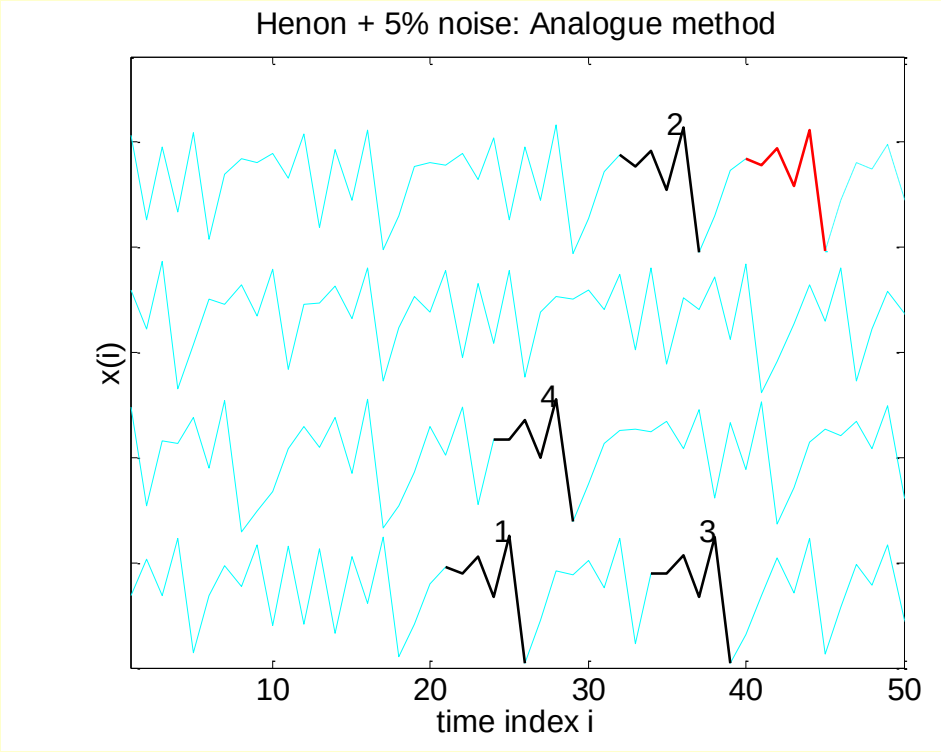
State space reconstruction (embedding)

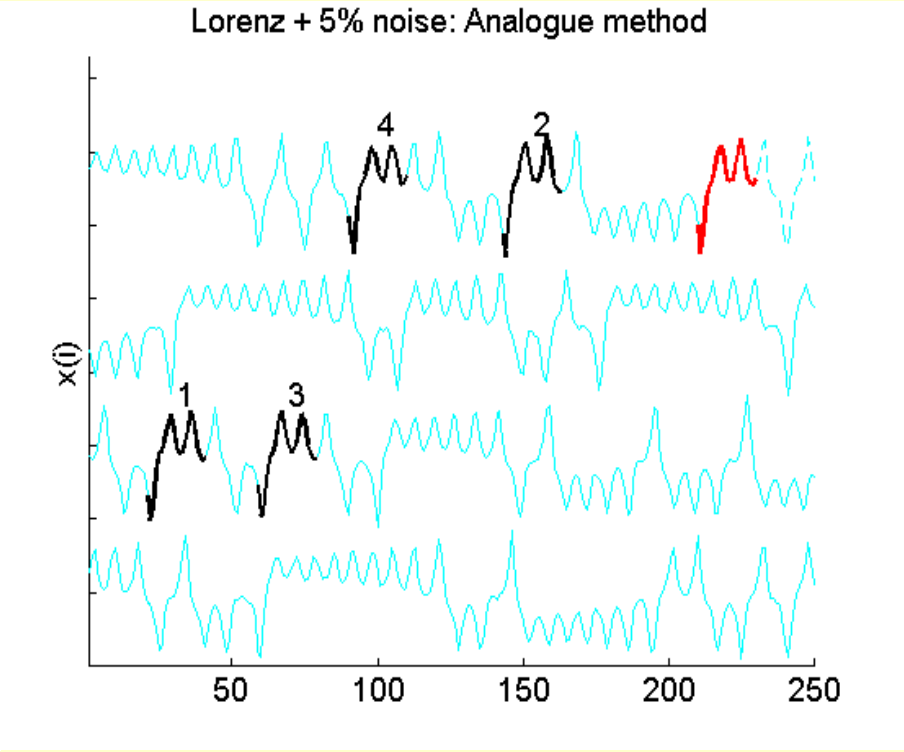


Prediction using similar past segments of the time series

given $x_1, x_2, \ldots x_i$ → predict x_{i+1} or x_{i+T}

Predict for time $i+T$ using the images T time steps ahead of the segments from the past, which are similar to the current segment





Local Prediction Models

Implementing the idea of analogue segments:
time series segments ➔ reconstructed points

segment $x_{i-(m-1)}, x_{i-(m-2)}, \cdots, x_{i-1}, x_i$

reconstructed point $\boldsymbol{x}_i = [x_i, x_{i-1}, \cdots x_{i-(m-1)}] \in R^m$

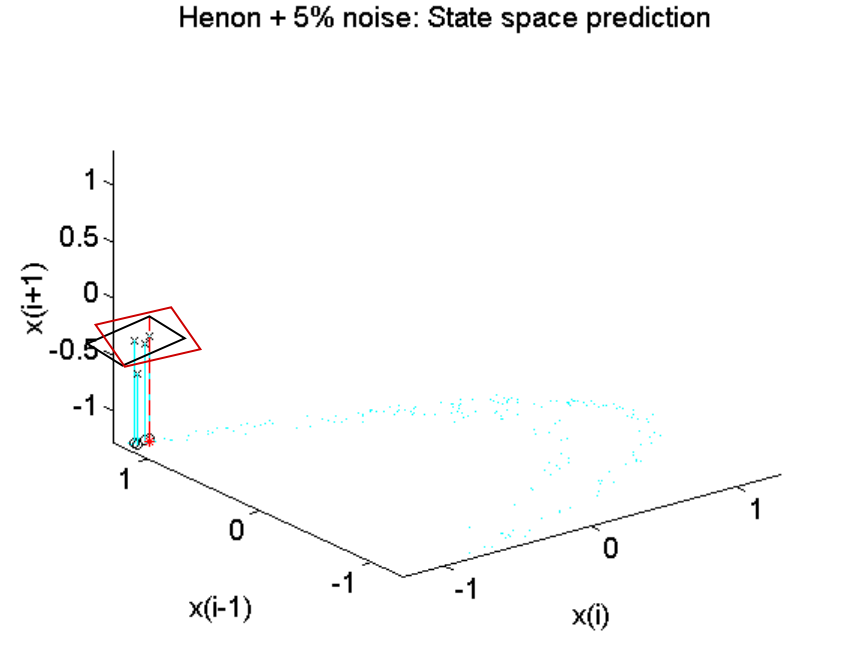
Nearest points to $\boldsymbol{x}_i$: $\{\boldsymbol{x}_{i(1)}, \boldsymbol{x}_{i(2)}, \ldots, \boldsymbol{x}_{i(K)}\}$

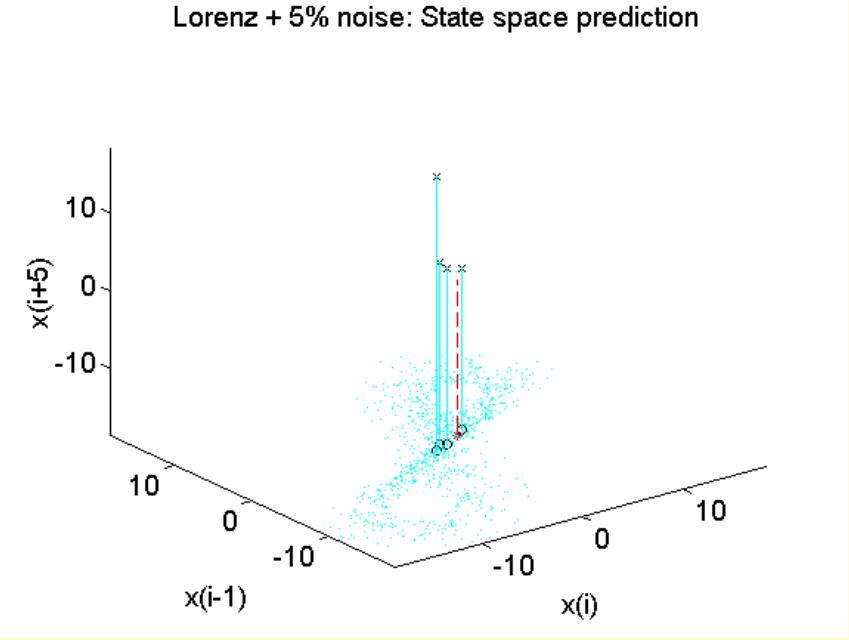
Prediction of $\boldsymbol{x}_{i+T}$ from the images of the neighbors $\{x_{i(1)+T}, x_{i(2)+T}, \ldots, x_{i(K)+T}\}$

Constant prediction: $\hat{x}_{i+T} \equiv x_i(T) = x_{i(1)+T}$

Average prediction $x_i(T) = \frac{1}{K}\sum_{j=1}^{K} x_{i(j)+T}$

Local Average Map (LAM)

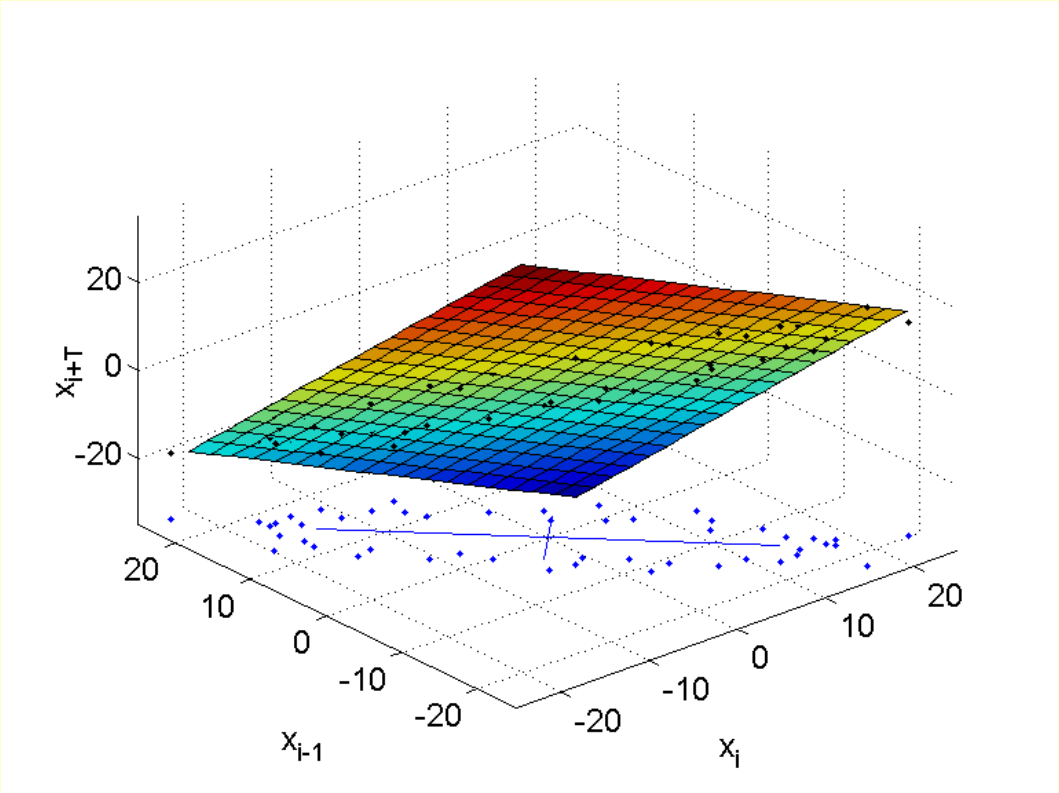




Local Linear Map (LLM)

We assume that for each point $\boldsymbol{x}_i$ the underlying system can be approximated by a linear model:

$$x_{i+1} = F(\boldsymbol{x}_i) = F(x_i, x_{i-\tau}, \dots, x_{i-(m-1)\tau})$$
$$= a_0 + a_1 x_i + a_2 x_{i-\tau} + \cdots + a_m x_{i-(m-1)\tau}$$
$$= a_0 + \boldsymbol{a}'\boldsymbol{x}_i$$



The model holds for

$$\boldsymbol{x}_{i(1)}, \boldsymbol{x}_{i(2)}, \dots, \boldsymbol{x}_{i(K)}$$

$$x_{i(1)+T} = a_0 + \boldsymbol{a}'\boldsymbol{x}_{i(1)}$$
$$\vdots$$
$$x_{i(K)+T} = a_0 + \boldsymbol{a}'\boldsymbol{x}_{i(K)}$$

Parameter estimation $a_0, a_1, \dots, a_m$ (least square method)

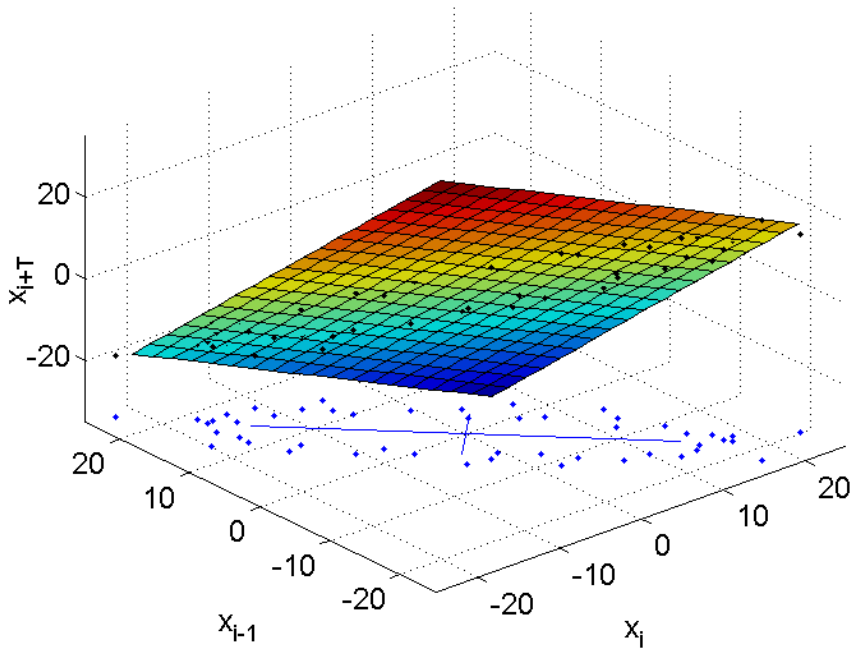
$$\min_{a_0, a_1, \dots, a_m} \sum_{j=1}^{K} \left(x_{i(j)+1} - (a_0 + \boldsymbol{a}'\boldsymbol{x}_{i(j)}) \right)^2$$

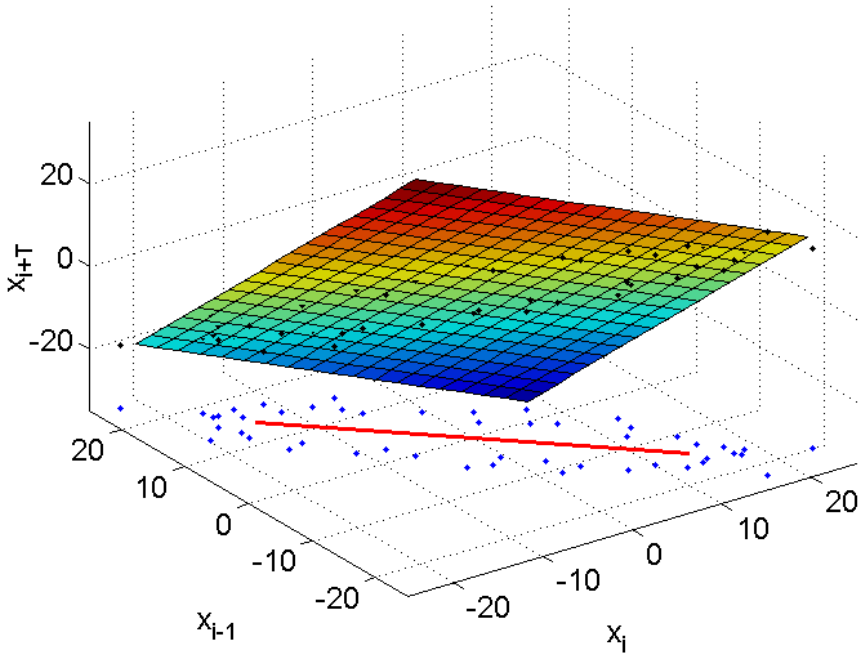
Improvement of LLM

Regularization of the ordinary least square solution of the model parameters

(principal component regression, PCR)

The parameter solution is restricted to a subspace defined by the principal components





Estimation of prediction error

Split the time series in two parts:

Predictions $\hat{x}_{N_1+1}, \cdots, \hat{x}_N$

$$\underbrace{x_1, x_2, \cdots, x_{N_1}}_{\text{Training set}}, \underbrace{x_{N_1+1}, \cdots, x_N}_{\text{test set}}$$

$$e_{i+T} = x_{i+T} - \hat{x}_{i+T}$$

$x_i(T)$

prediction error

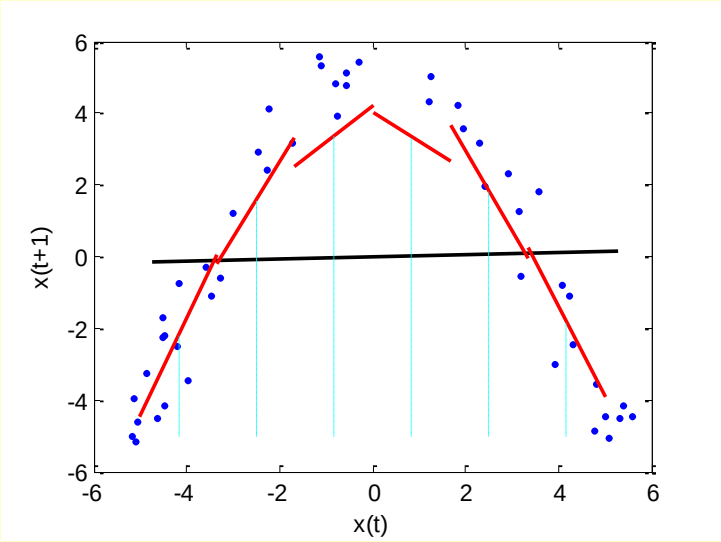
Prediction error statistic

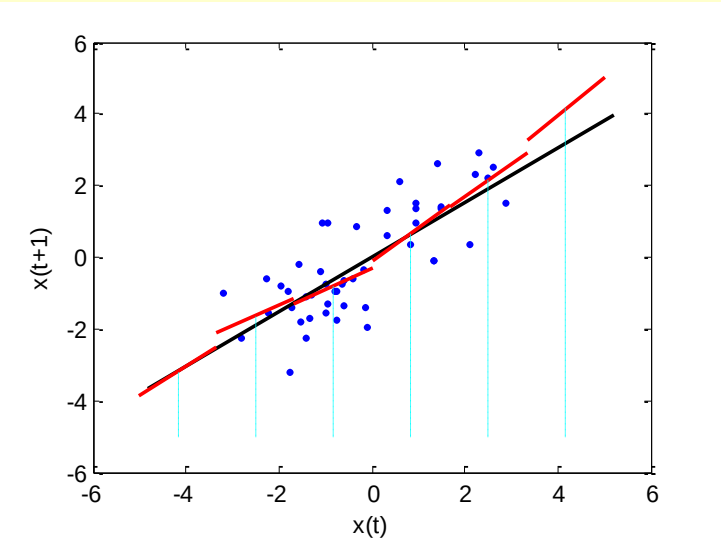
$$\text{NRMSE}(T) = \sqrt{\frac{\frac{1}{N-T-N_1}\sum_{t=N_1+1}^{N-T}\left(x_{t+T} - \hat{x}_{t+T}\right)^2}{\frac{1}{N}\sum_{i=1}^{N}\left(x_i - \bar{x}\right)^2}}$$

- Nonlinear modeling and prediction

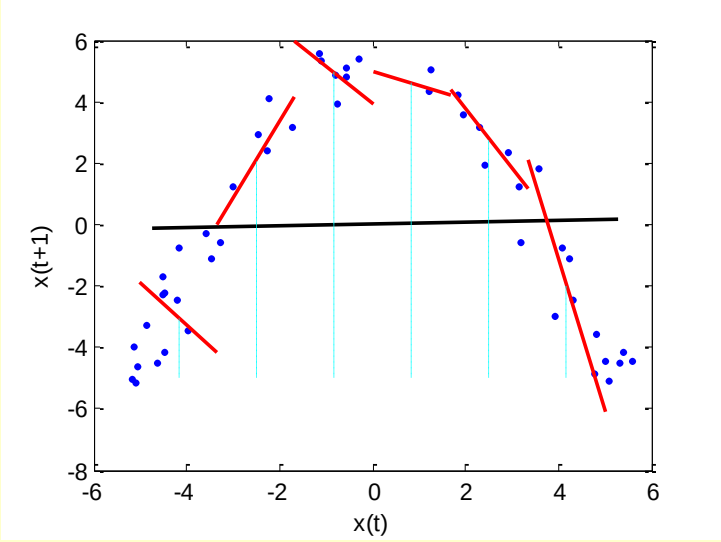
Can a nonlinear model be worse than a linear?

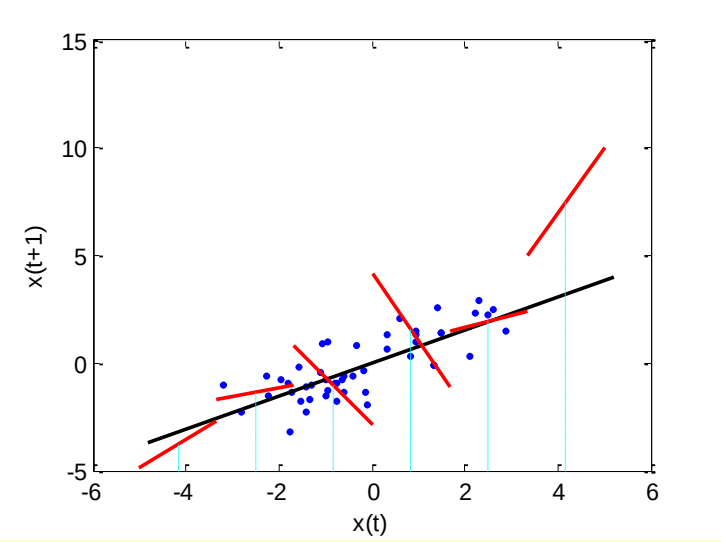
$K=20$





$K=6$





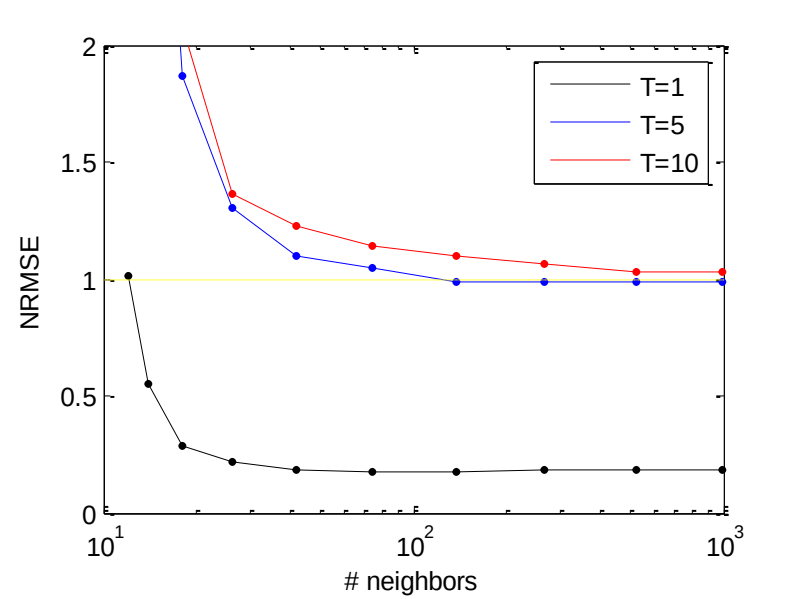
- Nonlinear modeling and prediction

Can a nonlinear model be worse than a linear?

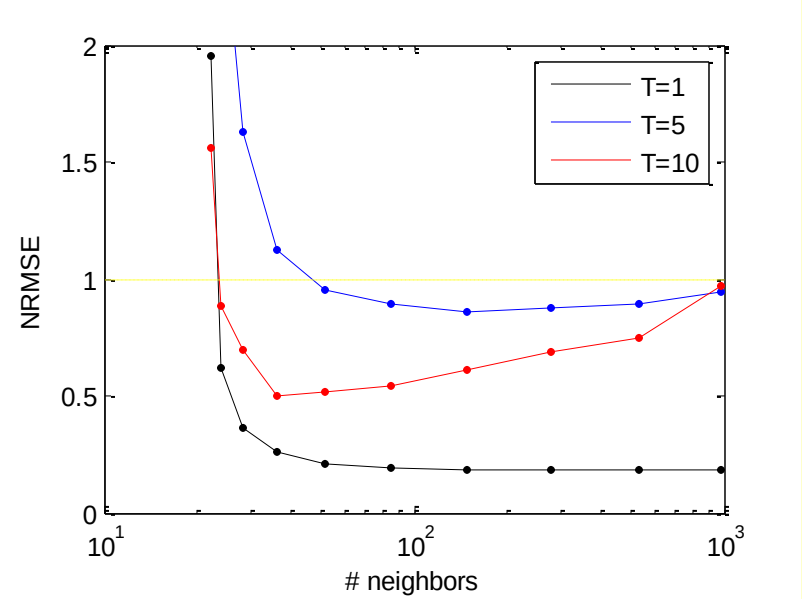
- Computation of NRMSE for increasing number of neighbors
- At the limit of the largest number of neighbors the LLM becomes … the linear autoregressive model

Example: Mackey-Glass, Δ=100 $\tau=1,\ n=1500,\ n - n_1=500$

m=10

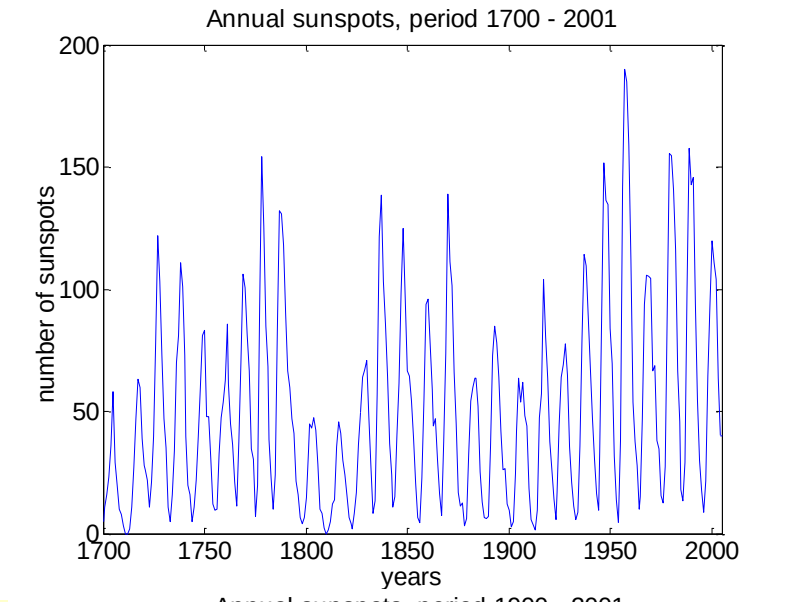


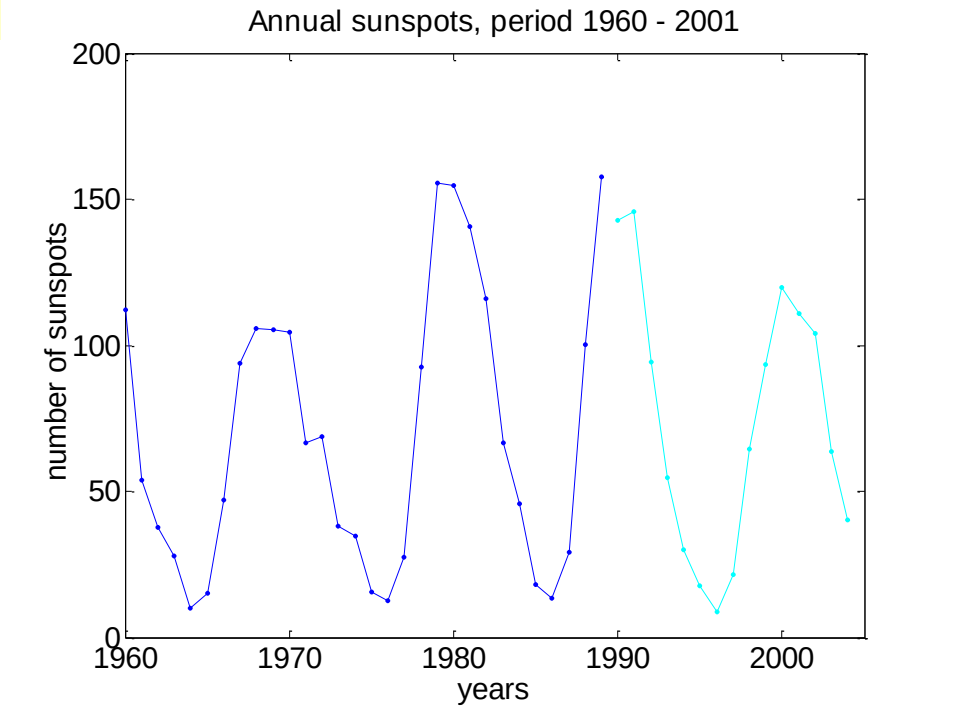
m=20

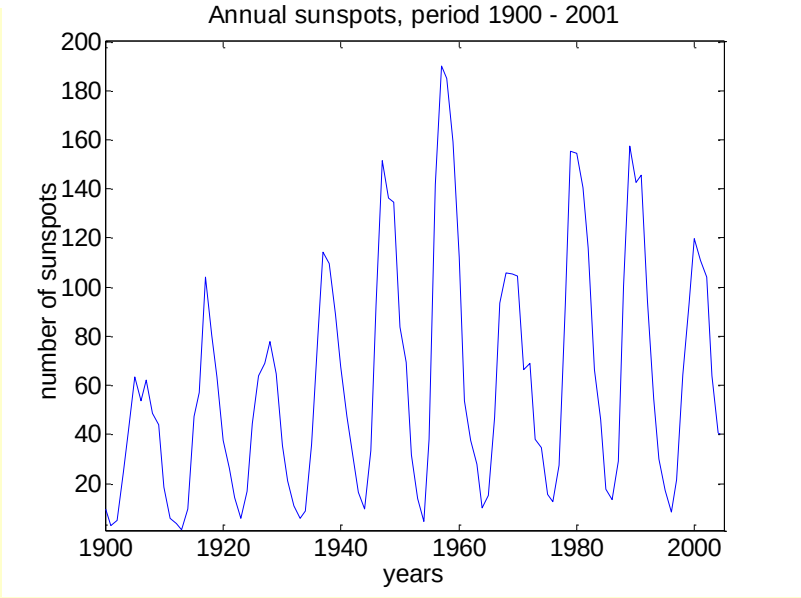


If the model parameters like m and K are not properly assigned the error with the nonlinear model can be larger than with the linear

Ηλιακές κηλίδες

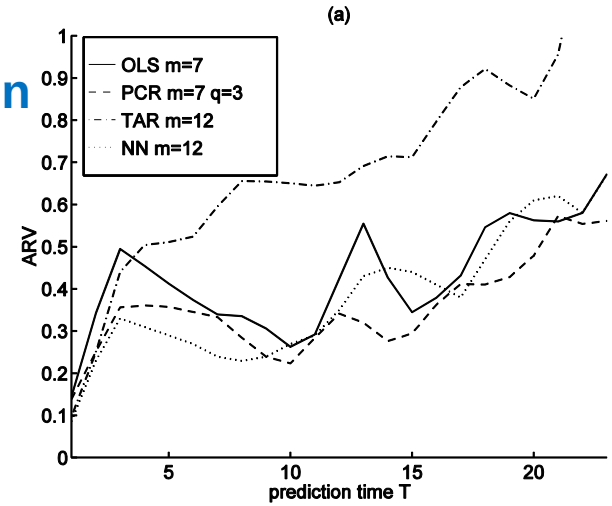


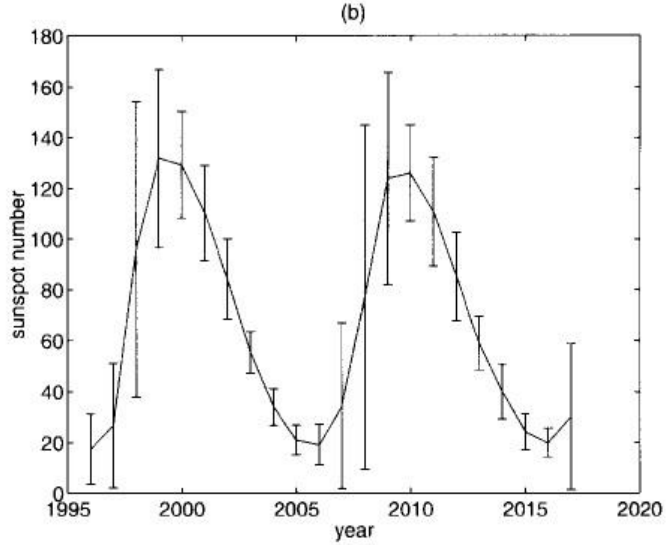




(1989)
Πόσες θα είναι οι ηλιακές κηλίδες τον επόμενο ή τα επόμενα χρόνια?

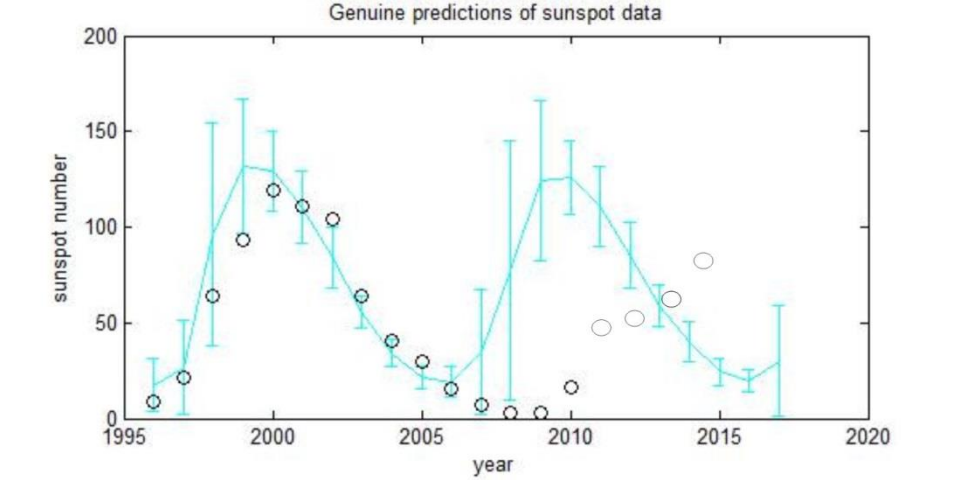
Comparison of models





Genuine prediction

Fig. 9. (a) ARV for iterative multi-step prediction with different models. The test set is the period 1921–1955. The first prediction year of the test set increases with T, e.g. for $T = 1$ the first prediction starts in 1921 and for $T = 23$ in 1944. Thus the number of predictions in the test set decreases with T. The models used in each plot are shown in the legend. (b) The genuine out-of-sample iterative prediction of the sunspot numbers up to the year 2017 based on data up to 1995 using the PCR model with $q = 3$ and $m = 7$. The bars show two times the standard deviation of the point estimate for each step.



CHAOS IN BRAIN?

Proceedings of the Workshop

University of Bonn, Germany 10 - 12 March 1999

edited by **K Lehnertz, C E Elger** *(University of Bonn, Germany)*, **J Arnhold & P Grassberger** *(NIC, Forschungszentrum Jülich, Germany)*

LINEAR AND NONLINEAR ANALYSIS OF EEG FOR THE PREDICTION OF EPILEPTIC SEIZURES

D. KUGIUMTZIS
Max-Planck Institute for Physics of Complex Systems, Nöthnitzer Str. 38, 01187 Dresden, Germany
E-mail: dimitris@mpipks-dresden.mpg.de

P. G. LARSSON
National Center for Epilepsy, Pb 900, 1303 Sandvika, Norway,
E-mail: paalgl@epilepsy.no

Linear and nonlinear methods are applied to multichannel scalp EEG records from 7 epileptic patients for the prediction of epileptic seizures of generalized tonic-clonic and complex partial type. The estimates on overlapped data segments are statistically processed in order to discriminate phases within the preictal state (comparing records long before and shortly before the onset) and find possible trends within the last preictal state. The results from the different subjects are not conclusive neither for the prognosis of the seizure (phase discrimination and trend) nor for the performance of the methods. However, in most cases, the methods could discriminate between records many hours before and shortly before the onset and in some cases detect a trend of increasing "complexity" within the last preictal state. It turned out that no method gave consistently best results and generally the linear methods performed as good as the nonlinear ones.

"… the linear methods performed as good as the non-linear ones."

EVIER Clinical Neurophysiology 116 (2005) 569–587
www.elsevier.com/locat

On the predictability of epileptic seizures

Florian Mormann[a,b,*], Thomas Kreuz[a,c], Christoph Rieke[a,b], Ralph G. Andrzejak[a,c], Alexander Kraskov[c], Peter David[b], Christian E. Elger[a], Klaus Lehnertz[a]

Abstract

Objective: An important issue in epileptology is the question whether information extracted from the EEG of epilepsy patients can be used for the prediction of seizures. Several studies have claimed evidence for the existence of a pre-seizure state that can be detected using different characterizing measures. In this paper, we evaluate the predictability of seizures by comparing the predictive performance of a variety of univariate and bivariate measures comprising both linear and non-linear approaches.

Methods: We compared 30 measures in terms of their ability to distinguish between the interictal period and the pre-seizure period. After completely analyzing continuous intracranial multi-channel recordings from five patients lasting over days, we used ROC curves to distinguish between the amplitude distributions of interictal and preictal time profiles calculated for the respective measures. We compared different evaluation schemes including channelwise and seizurewise analysis plus constant and adaptive reference levels. Particular emphasis was placed on statistical validity and significance.

Results: Univariate measures showed statistically significant performance only in a channelwise, seizurewise analysis using an adaptive baseline. Preictal changes for these measures occurred 5–30 min before seizures. Bivariate measures exhibited high performance values reaching statistical significance for a channelwise analysis using a constant baseline. Preictal changes were found at least 240 min before seizures. Linear measures were found to perform similar or better than non-linear measures.

"… Linear measures were found to perform similar or better than non-linear measures."

Primitive Brain Is 'Smarter' Than We Think

↕

Primitive methods are "smarter" than we think ?

primitive: simple measures

ScienceDaily
Your source for the latest research news

News | Articles | Videos | Images | Bo
Health & Medicine | Mind & Brain | Plants & Animals | Earth & Climate | Space & Tin

Science News Share

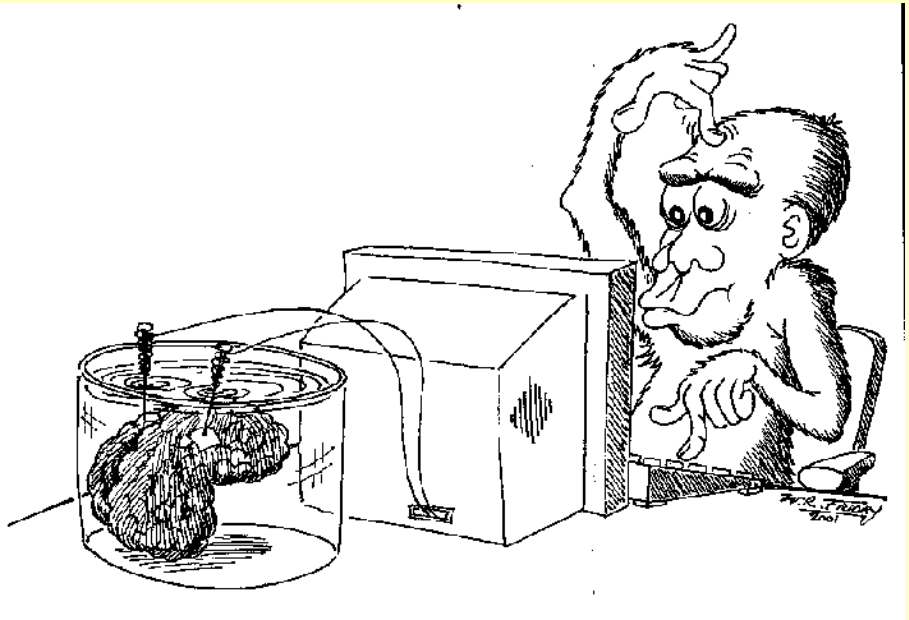
Primitive Brain Is 'Smarter' Than We Think, MIT Study Shows

ScienceDaily (Mar. 14, 2005) — Primitive structures deep within the brain may have a far greater role in our high-level everyday thinking processes than previously believed, report researchers at the MIT Picower Center for Learning and Memory in the Feb. 24 issue of Nature.

Nonlinear dynamics are appealing and we want to investigate them in physiological, financial, geological etc data, but…

- are they really there?
- can we detect them from the measurements?
- …

- is it sufficient to have time series from a single observable?



matlab toolkit

Measures of Analysis of Time Series, MATS

together with A. Tsimpiris

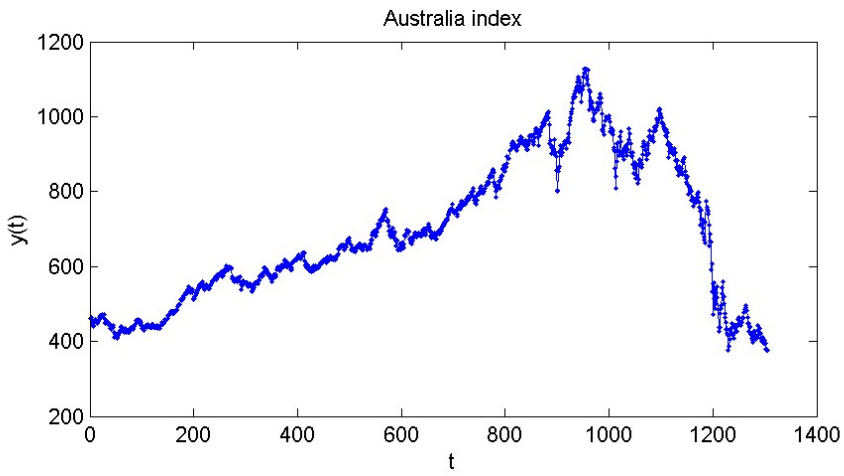
Nonlinear Analysis of Time Series
Part II: Multivariate

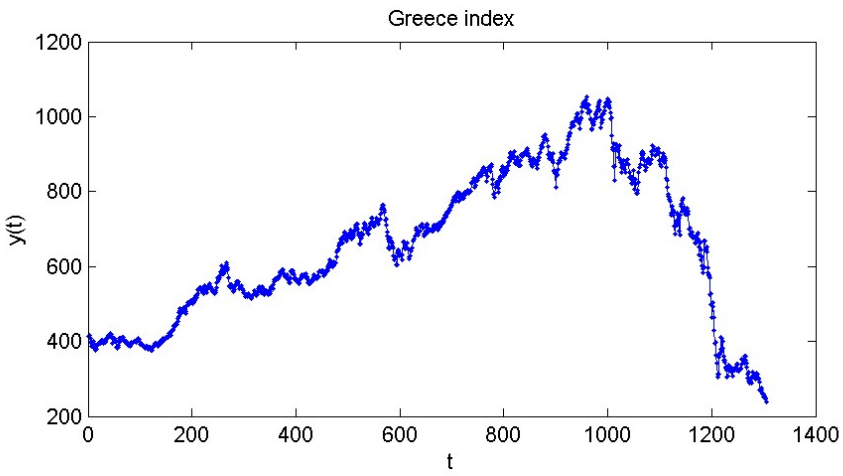
Kugiumtzis Dimitris

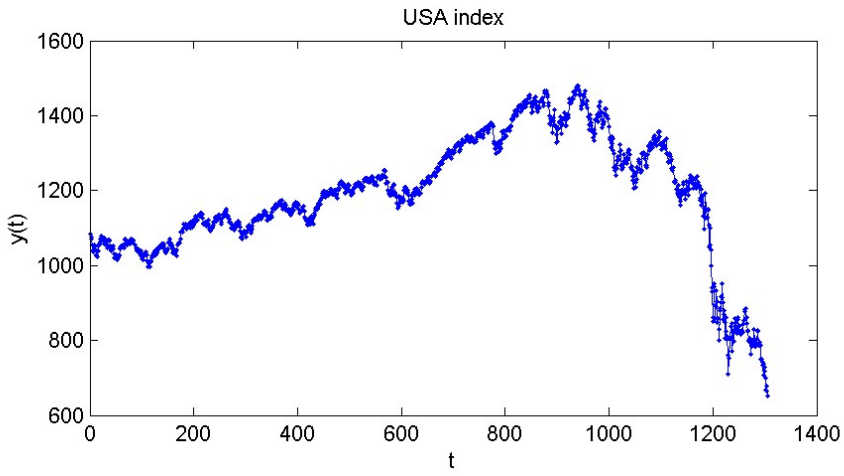
Department of Electrical and Computer Engineering,
Faculty of Engineering, Aristotle University of Thessaloniki, Greece
e-mail: dkugiu@auth.gr http:\\users.auth.gr\dkugiu

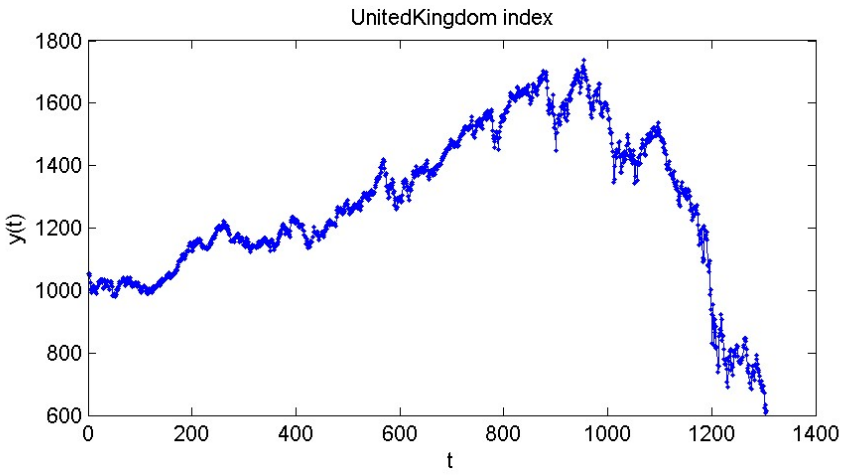
23nd Summer School and Conference "Dynamical Systems and Complexity", Aristotle Univerity Camping, Kalandra, 30 August 2016

Financial World Markets

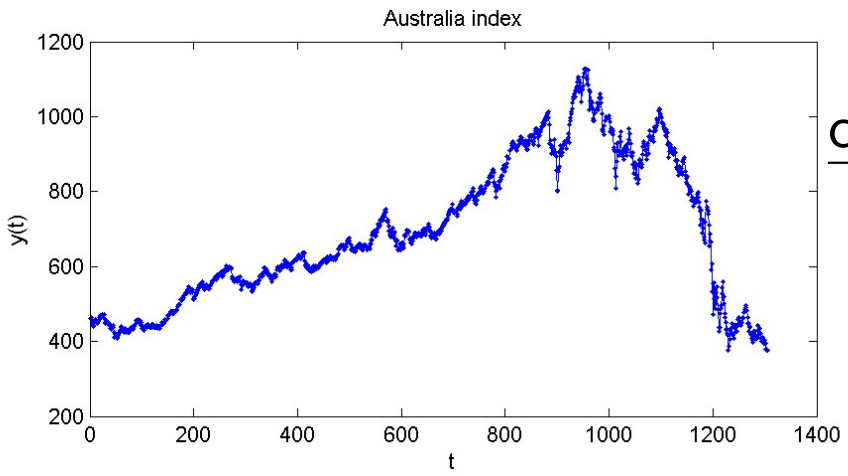




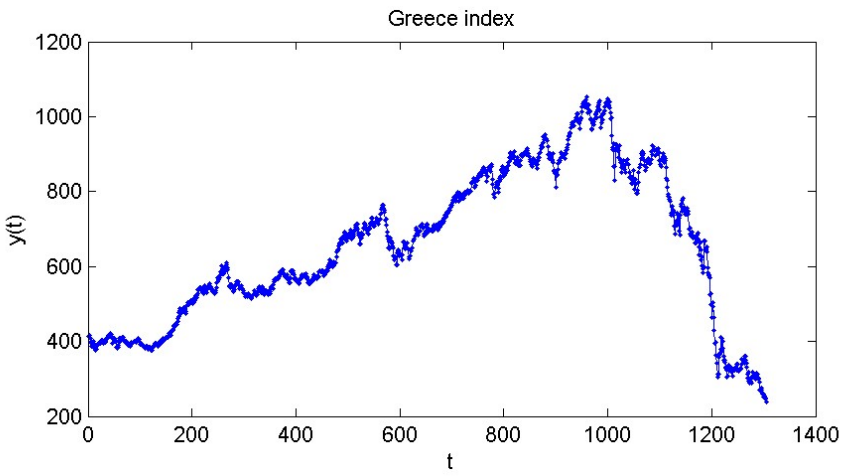


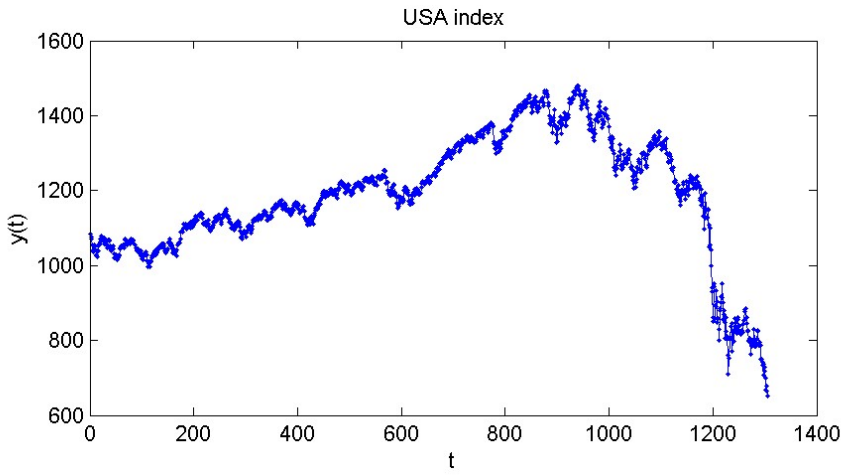


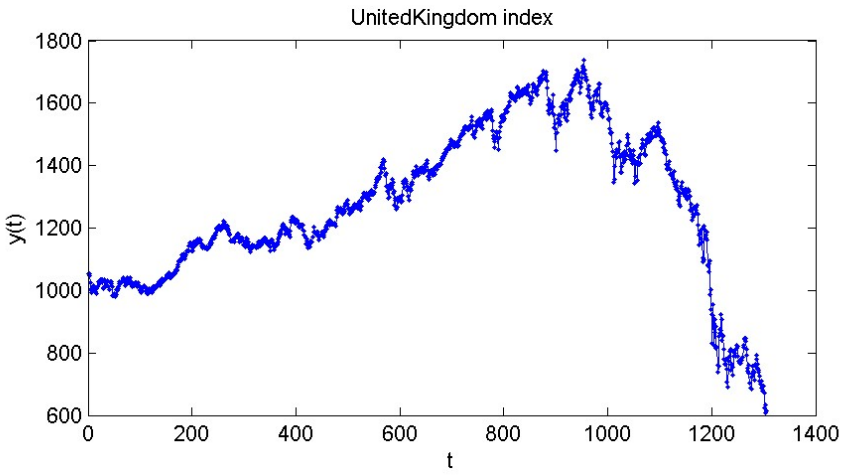
Financial World Markets



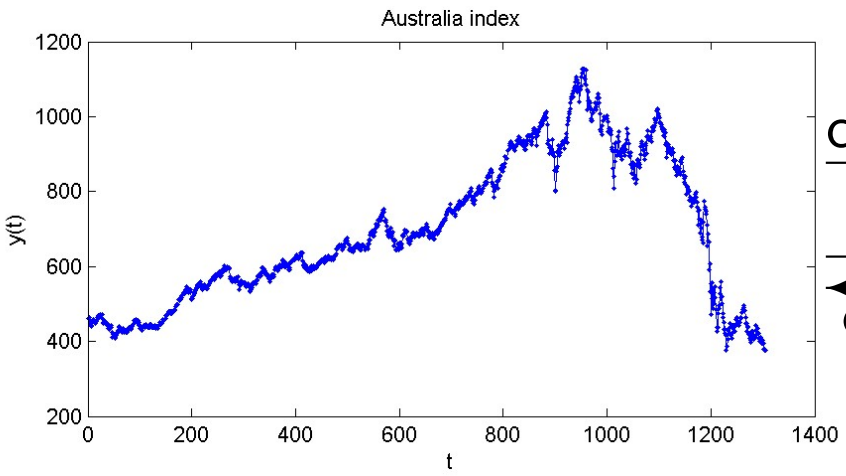
?
correlation







Financial World Markets

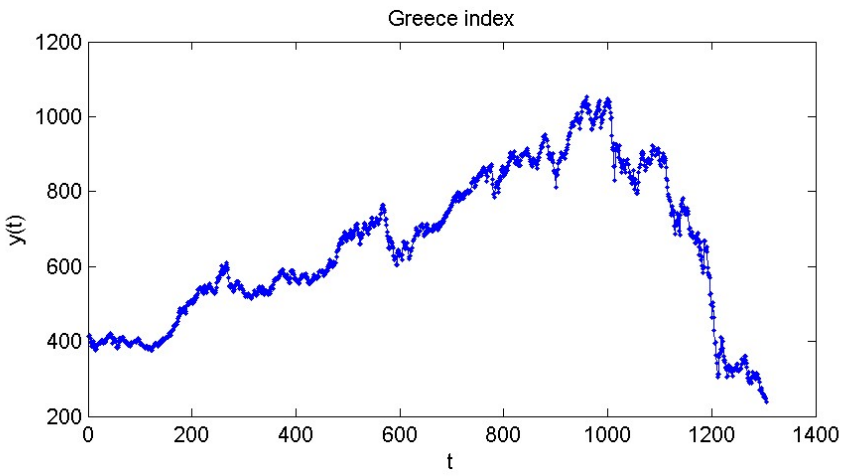


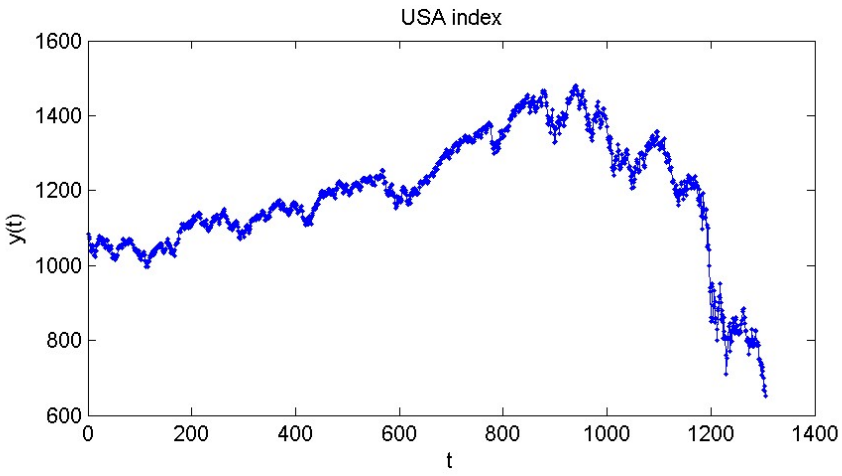
?

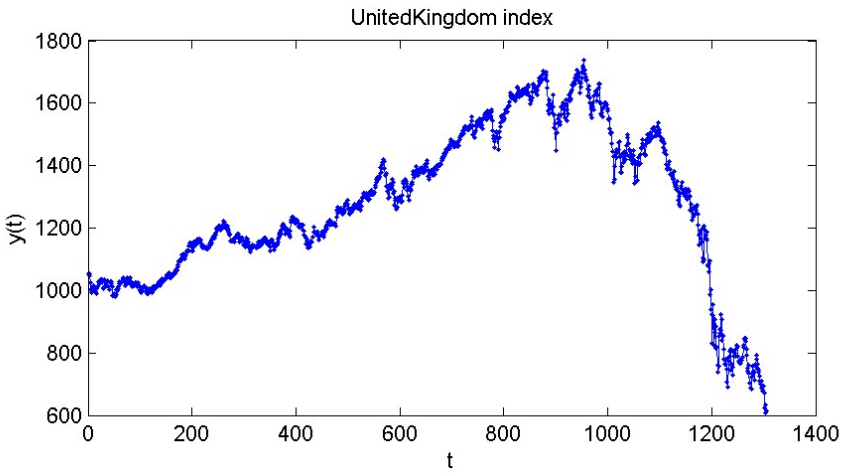
correlation

causality

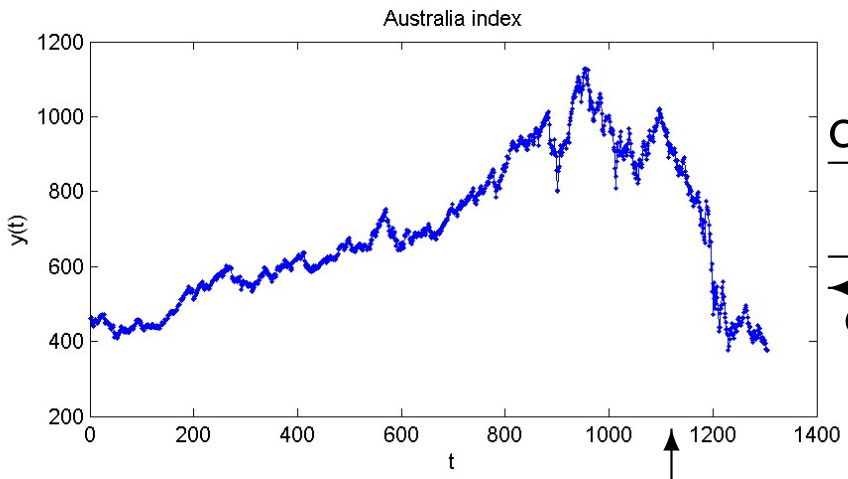
?







Financial World Markets

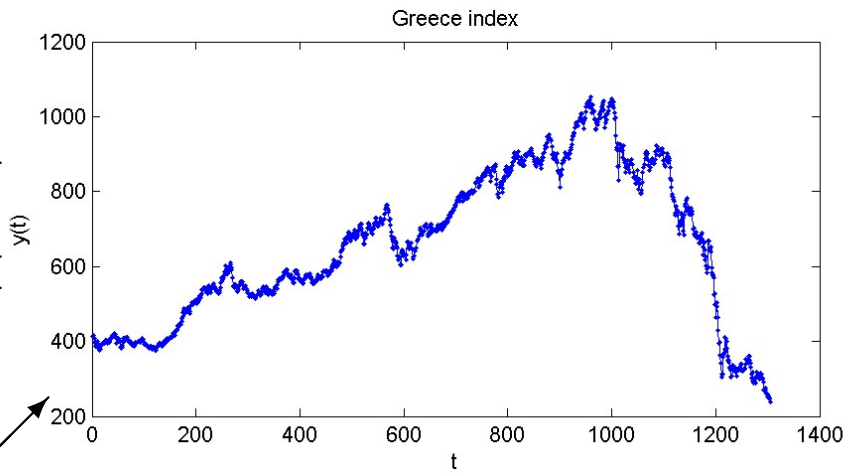


?

correlation

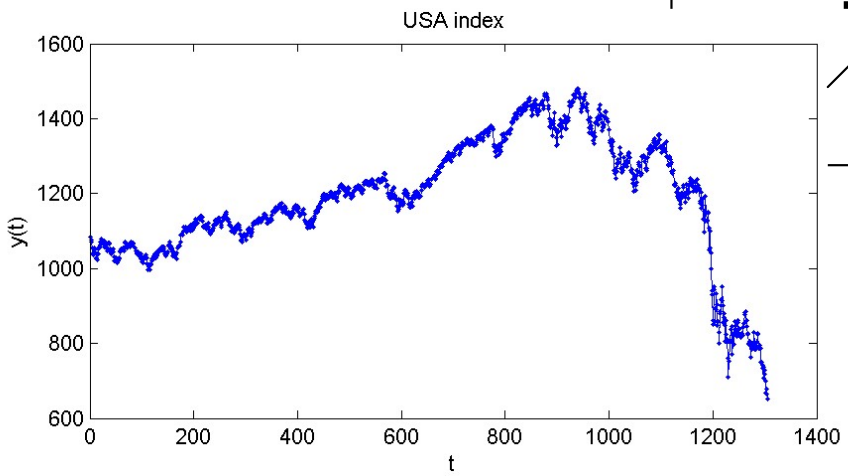
causality

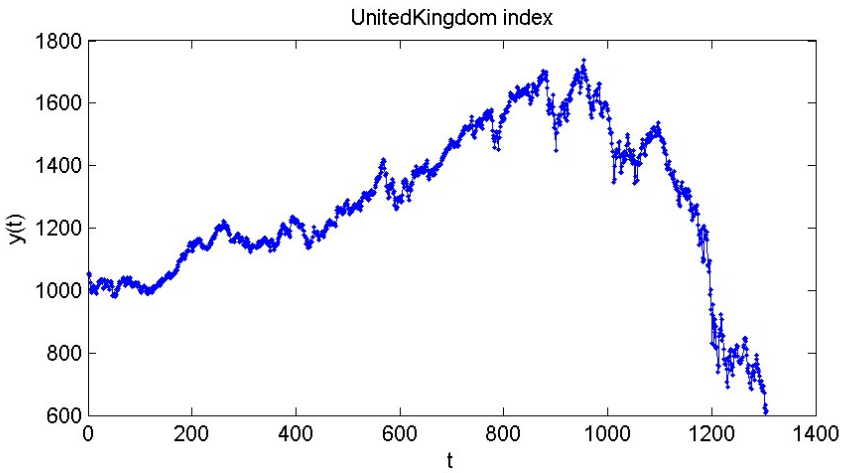
?



Does one market drive the other markets?

?

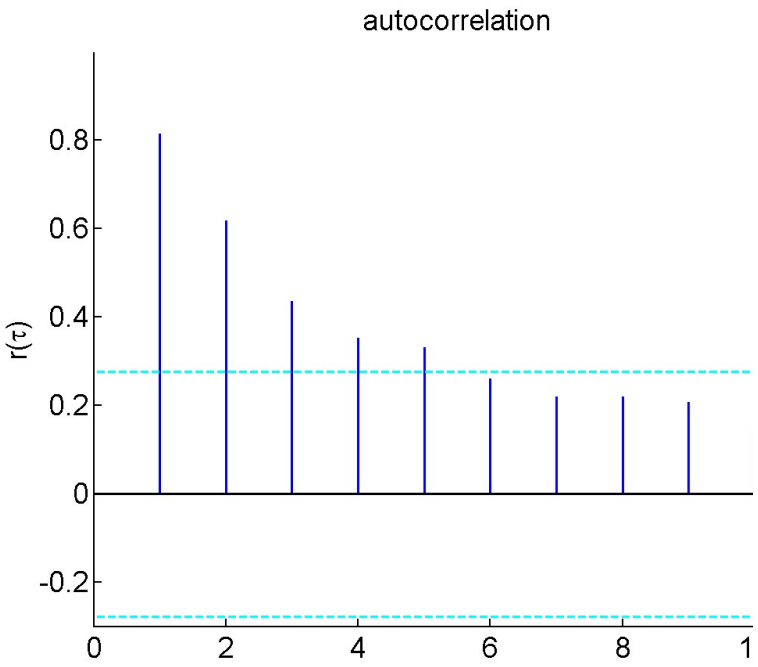


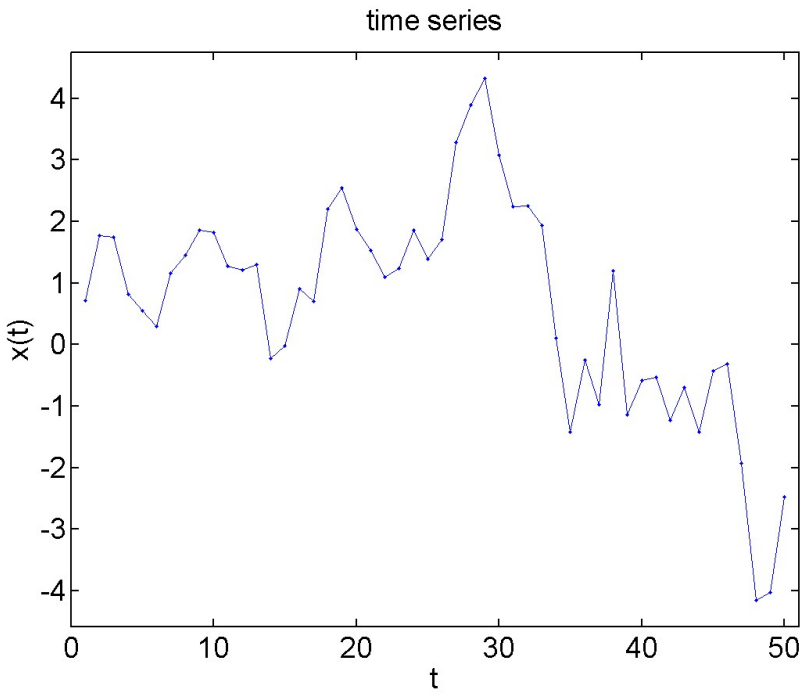


(auto)correlation $r(X_t; X_{t-\tau})$

Are X_t and X_{t-1} linearly correlated? $r(X_t; X_{t-1}) \neq 0$?

Are X_t and X_{t-2} linearly correlated? $r(X_t; X_{t-2}) \neq 0$?

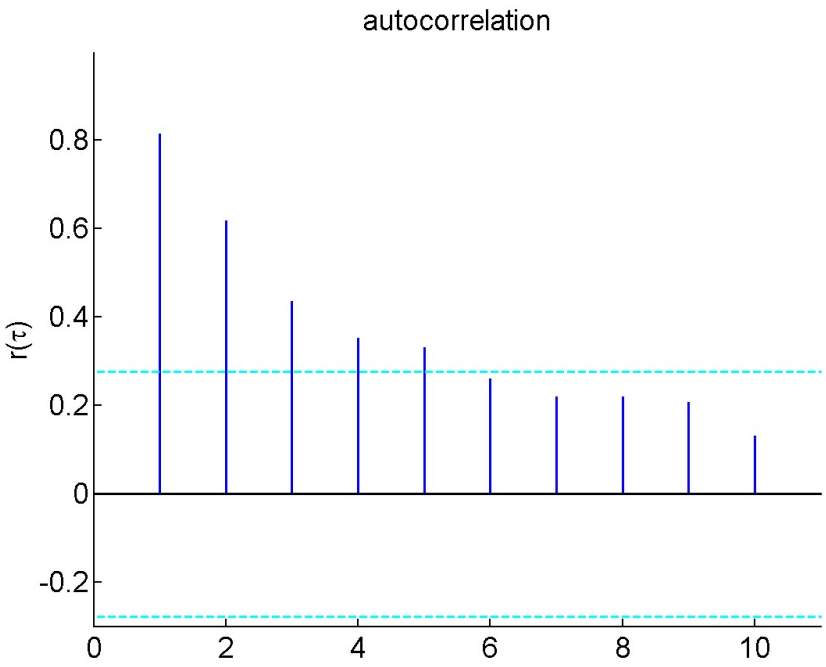




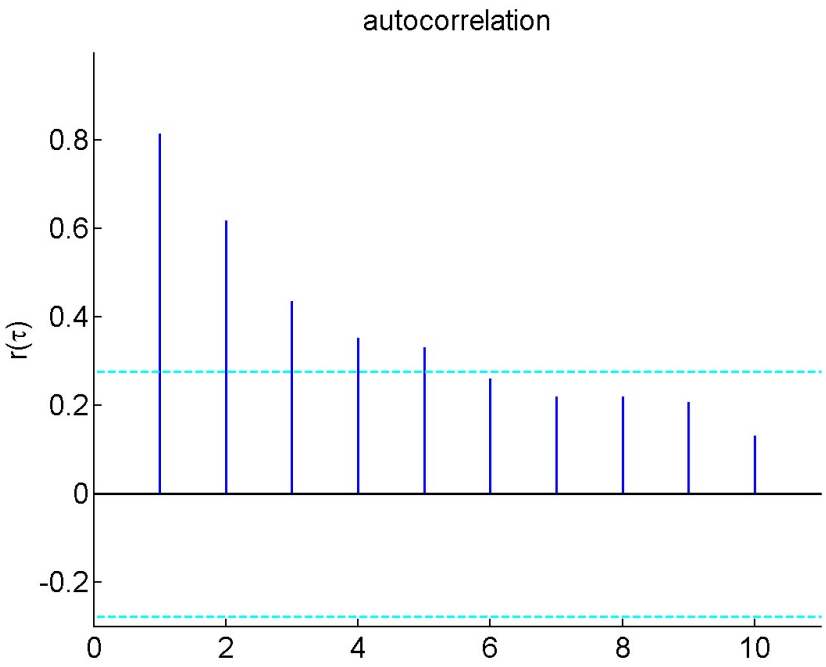
(auto)correlation $r(X_t; X_{t-\tau})$

Are X_t and X_{t-1} linearly correlated? $r(X_t; X_{t-1}) \neq 0$? Yes

Are X_t and X_{t-2} linearly correlated? $r(X_t; X_{t-2}) \neq 0$? Yes



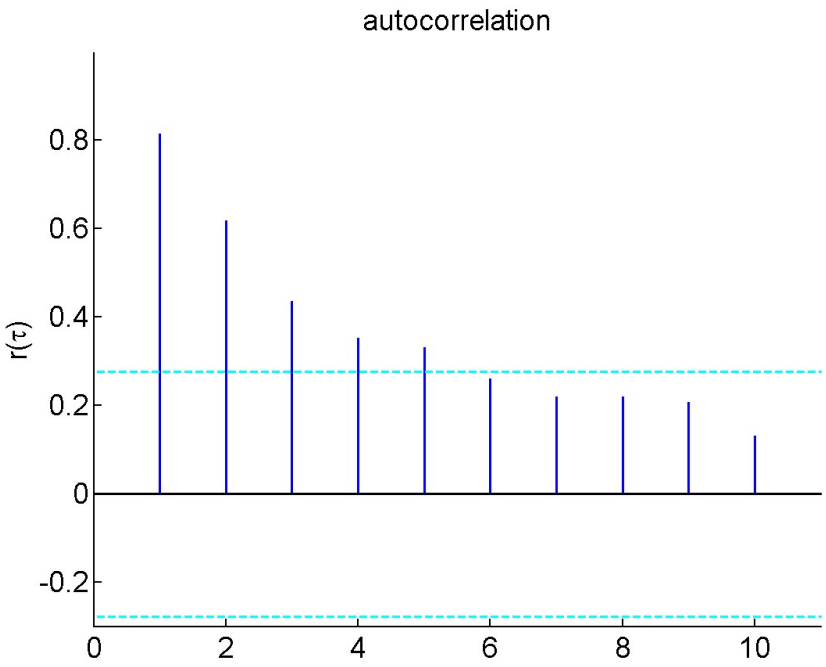
Are X_t and X_{t-2} directly linearly correlated?



Are X_t and X_{t-2} directly linearly correlated?

Are X_t and X_{t-2} linearly correlated given X_{t-1}?

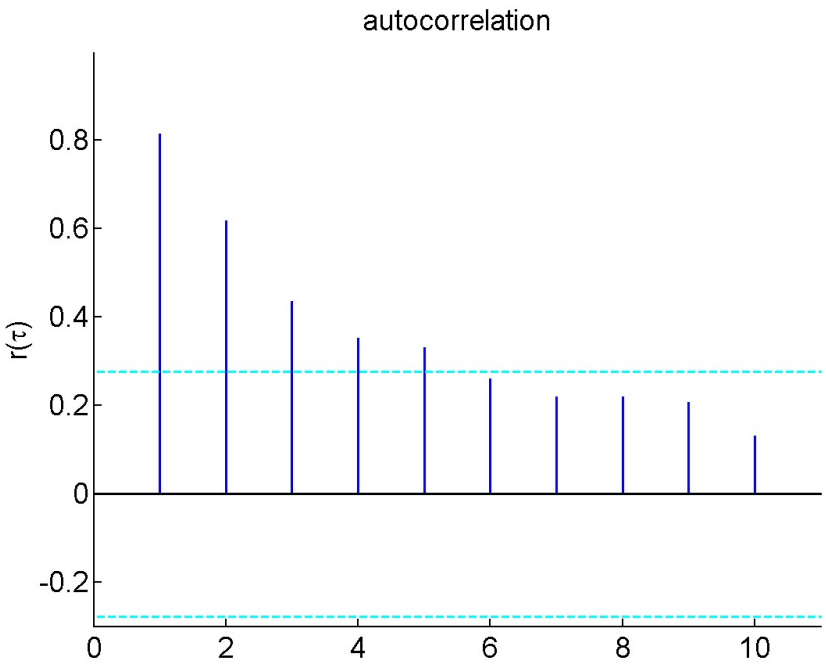
$$r(X_t; X_{t-2}|X_{t-1}) \neq 0?$$

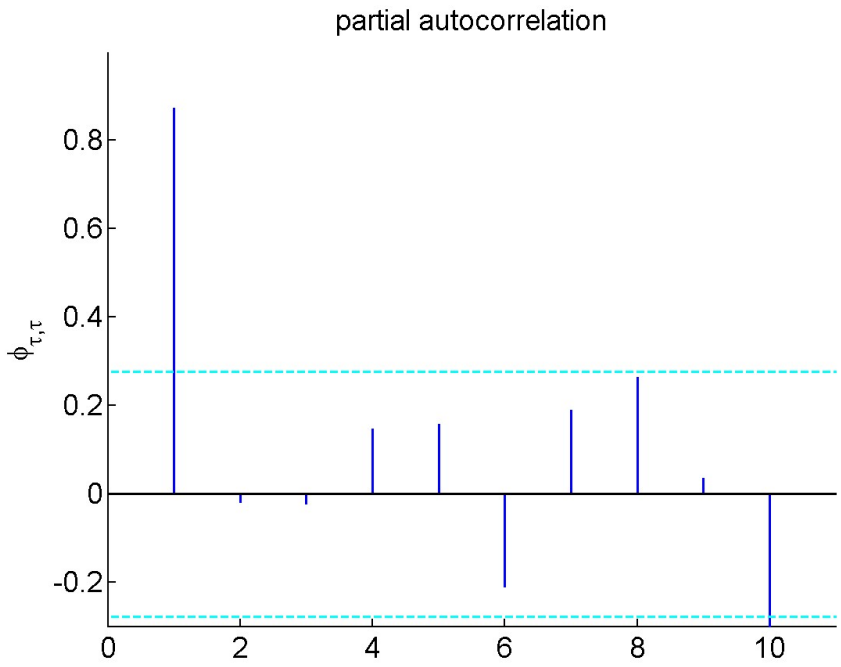


Are X_t and X_{t-2} directly linearly correlated?

Are X_t and X_{t-2} linearly correlated given X_{t-1}?

$$r(X_t; X_{t-2}|X_{t-1}) \neq 0? \quad \text{No}$$





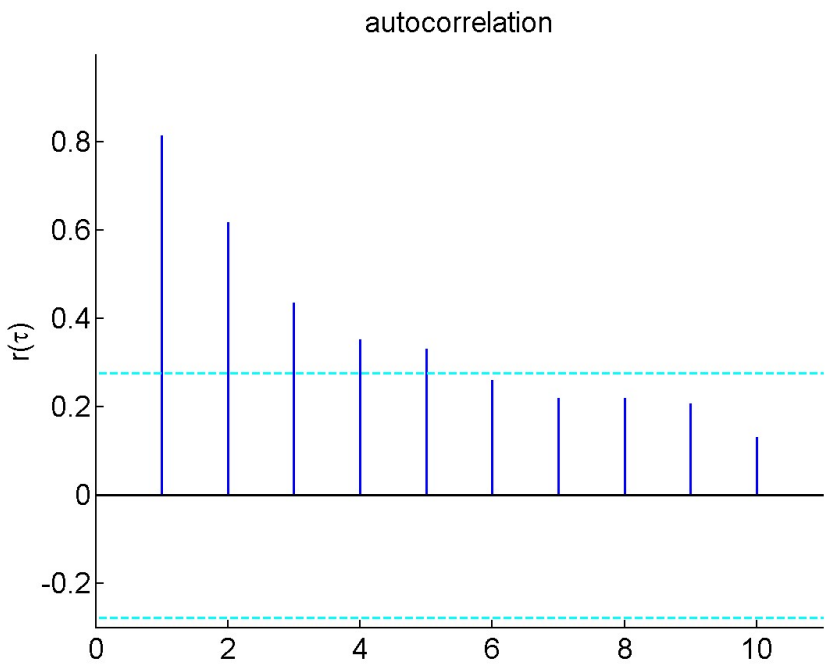
Are X_t and X_{t-2} directly linearly correlated?

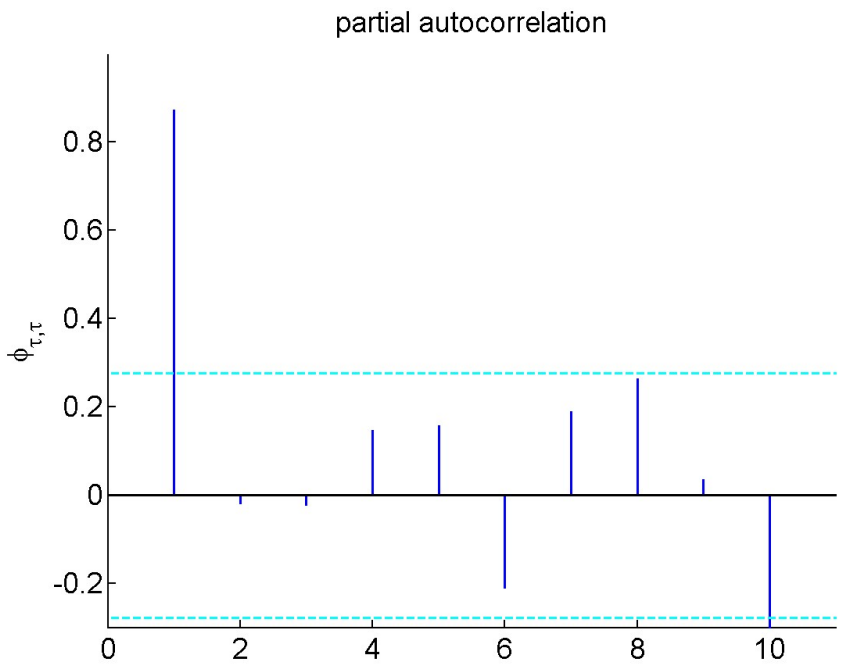
Are X_t and X_{t-2} linearly correlated given X_{t-1}?

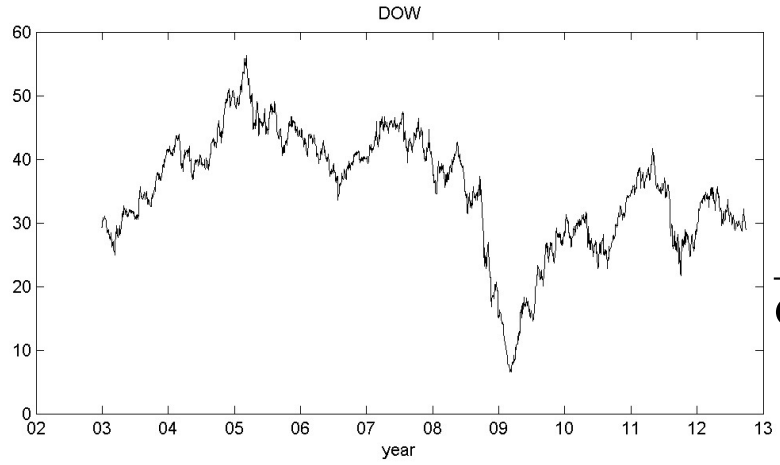
$$r(X_t; X_{t-2}|X_{t-1}) \neq 0? \quad \text{No}$$

Are X_t and X_{t-2} linearly or/and nonlinearly correlated given X_{t-1}?

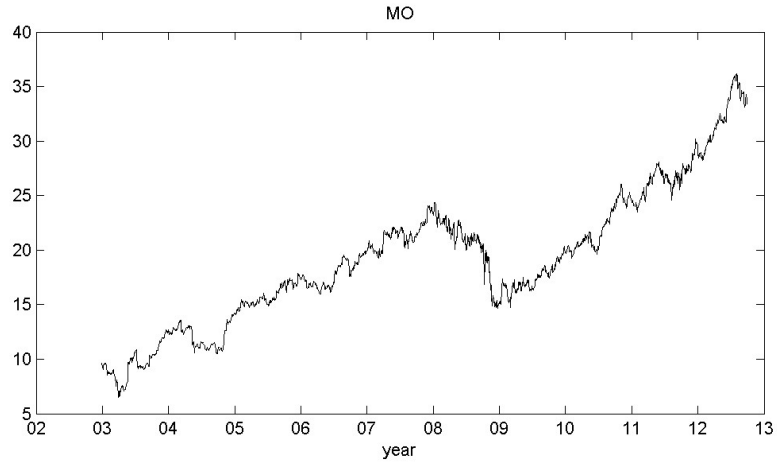
$$I(X_t; X_{t-2}|X_{t-1}) \neq 0?$$

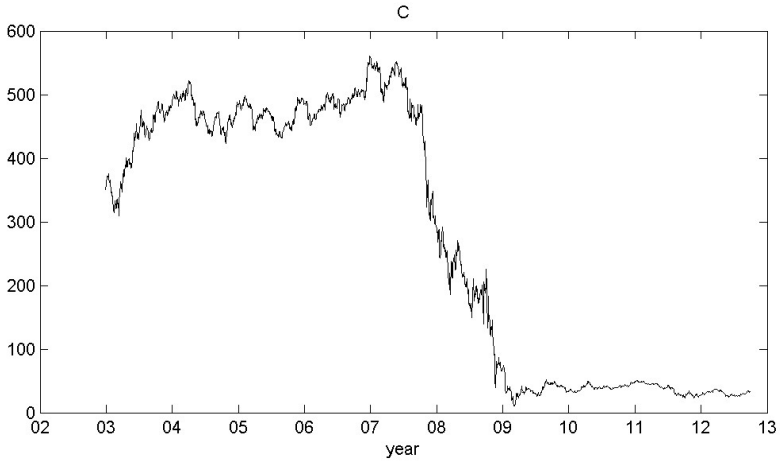


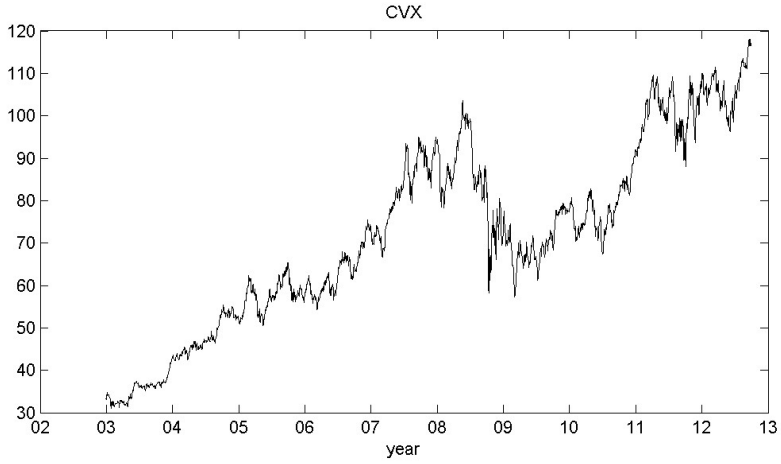




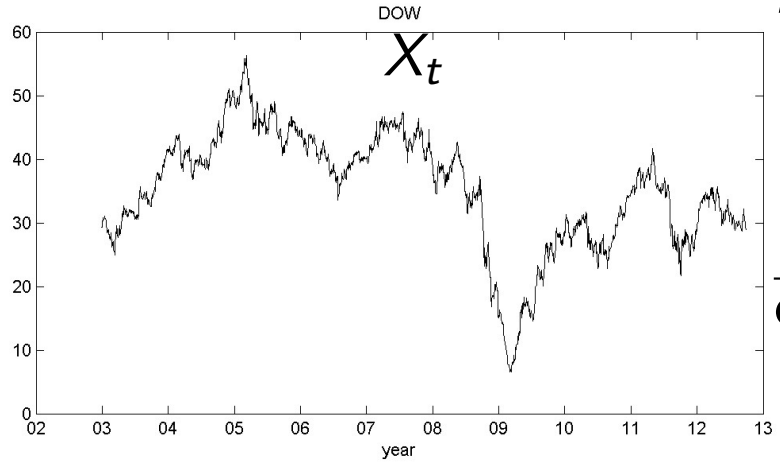
correlation
?



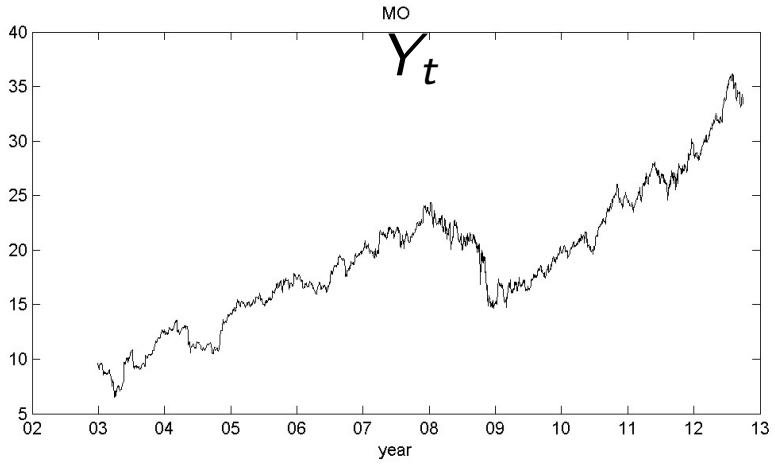


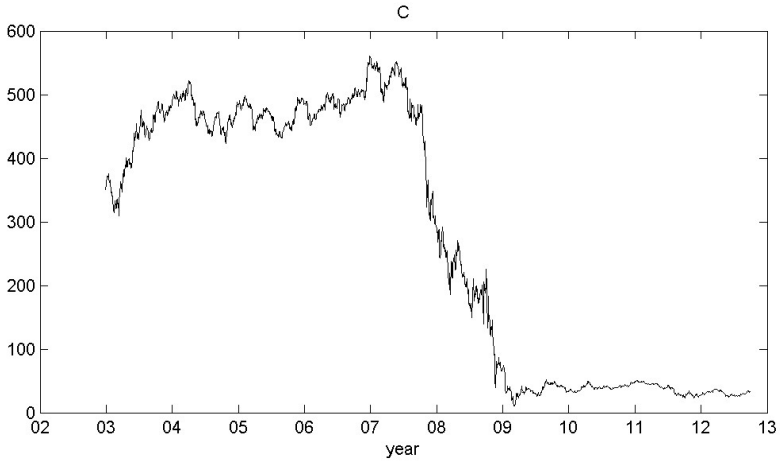


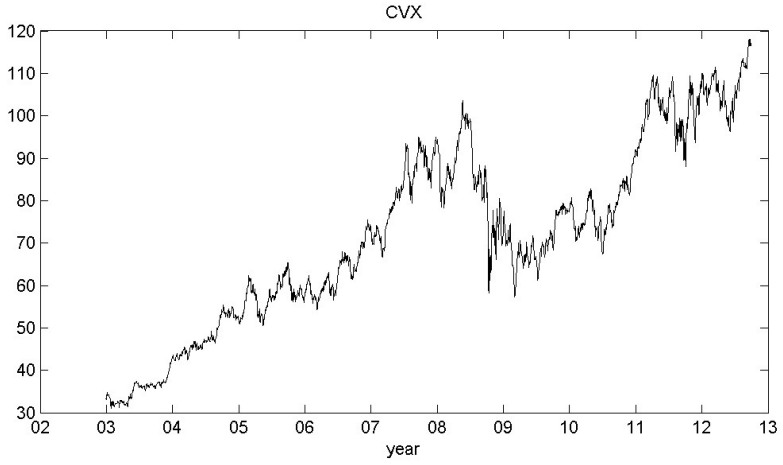
$r(X_t; Y_t)$



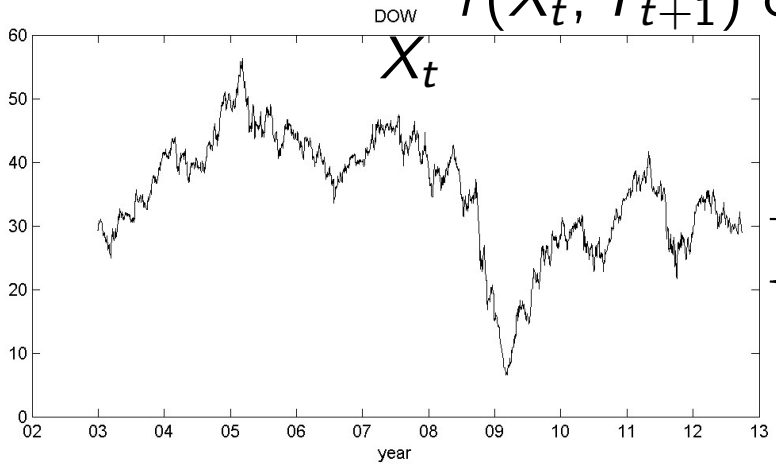
correlation
?



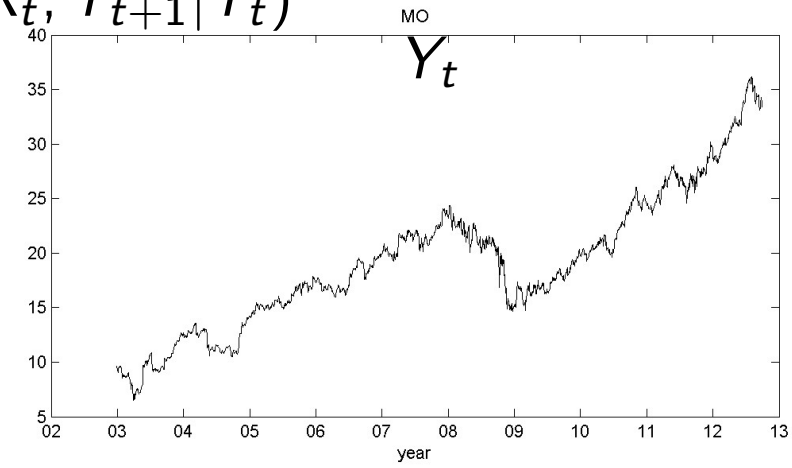


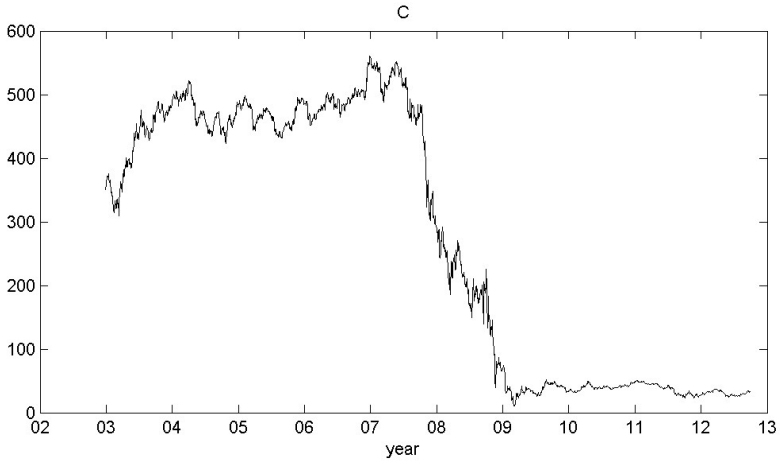


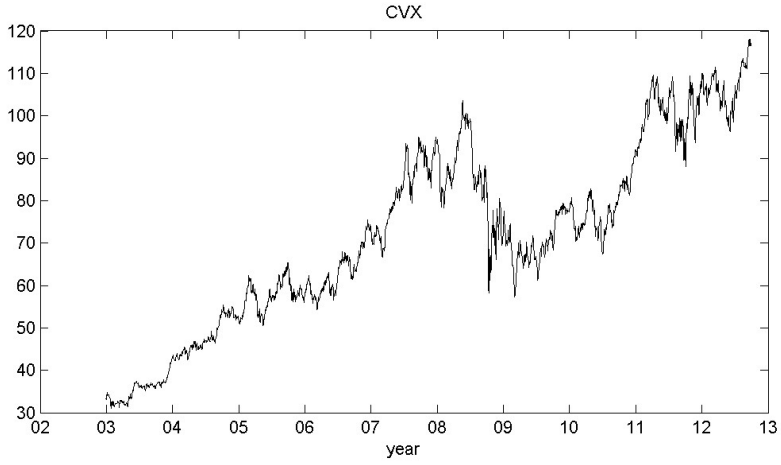
$r(X_t; Y_{t+1})$ or better $r(X_t; Y_{t+1}|Y_t)$



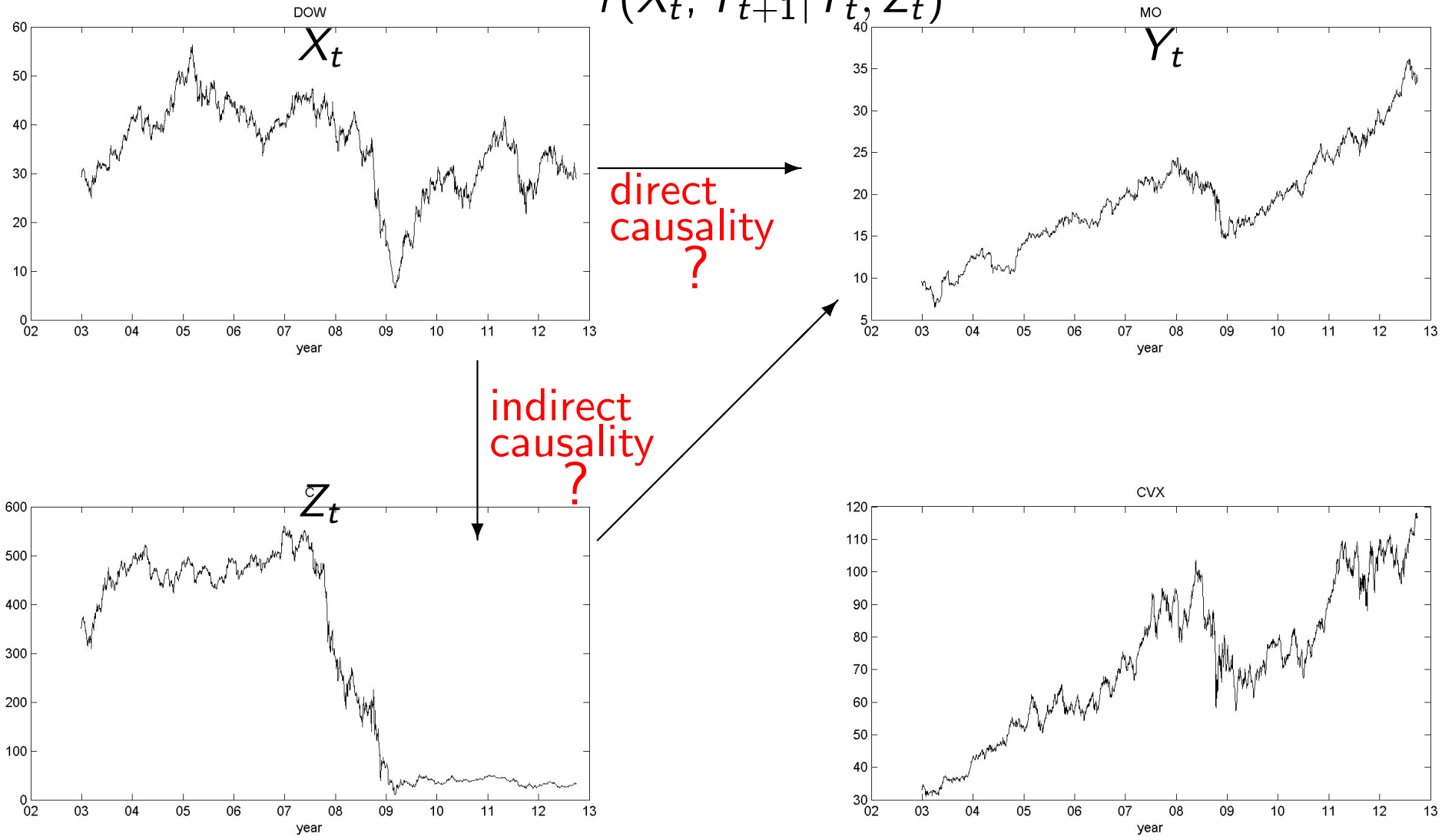
causality
?







$$r(X_t; Y_{t+1} | Y_t, Z_t)$$


DOW
X_t
year
MO
Y_t
year
direct causality ?
indirect causality ?
Z_t
year
CVX
year

Correlation measures

Linear	$X \sim Y$	$X \sim Y \vert Z$
	Cross-Correlation	Partial Correlation
	Coherence	Partial Coherence (pCOH)
Nonlinear	$X \sim Y$	$X \sim Y \vert Z$
	Mutual Information (MI)	Partial mutual information (PMI) [Frenzel & Pompe 2007]
	Mean Phase Coherence (MPC) [Mormann et al 2000]	
	Imaginary Coherence (iCOH) [Nolte et al 2004]	
	Phase Locking Value (PLV) [Lachaux et al 1999]	
	Rho index (RHO) [Tass et al 1998]	
	Phase Lag Index (PLI) [Stam et al 2007]	
	Event Synchronization (EventSync) [QuianQuiroga et al 2002]	

Correlation measures

Bivariate time series $\{x_t, y_t\}_{t=1}^{n}$

Linear correlation measures:

Estimate of cross-covariance

$$c_{XY}(\tau) = \hat{\gamma}_{XY}(\tau) = \frac{1}{n-\tau}\sum_{t=1}^{n-\tau}(x_t - \bar{x})(y_{t+\tau} - \bar{y})$$

$\bar{x}$ and $\bar{y}$ are sample means.

Estimate of **cross-correlation**:

$$r_{XY}(\tau) = \hat{\rho}_{XY}(\tau) = \frac{c_{XY}(\tau)}{c_{XY}(0)} = \frac{c_{XY}(\tau)}{s_X s_Y}$$

s_X and s_Y are sample standard deviations.

- $|r_{XY}(\tau)| \leq 1$
- $r_{XY}(\tau) = r_{YX}(-\tau)$ but $r_{XY}(\tau) \neq r_{XY}(-\tau)$

Nonlinear correlation measures:

Entropy: information from each sample of X (assume proper discretization of X)

$$H(X) = \sum_x p_X(x) \log p_X(x)$$

Mutual information: information for Y knowing X and vice versa

$$I(X,Y) = H(X)+H(Y)-H(X,Y) = \sum_{x,y} p_{XY}(x,y) \log \frac{p_{XY}(x,y)}{p_X(x)p_Y(y)}$$

For $X \to X_t$ and $Y \to Y_{t+\tau}$,
cross-delayed mutual information:

$$I_{XY}(\tau) = I(X_t, Y_{t+\tau}) = \sum_{x_t, y_{t+\tau}} p_{X_t Y_{t+\tau}}(x_t, y_{t+\tau}) \log \frac{p_{X_t Y_{t+\tau}}(x_t, y_{t+\tau})}{p_{X_t}(x_t) p_{Y_{t+\tau}}(y_{t+\tau})}$$

To compute $I_{XY}(\tau)$ make a partition of $\{x_t\}_{t=1}^n$, a partition of $\{y_t\}_{t=1}^n$ and compute probabilities for each cell from the relative frequency.

$r_{XY}(0) \neq 0$:

$\Longrightarrow$ (linear) correlation of x_t and y_t

$\Longrightarrow$ systems X and Y are correlated, $X \sim Y$

$r_{XY}(\tau) \neq 0$:

$\Longrightarrow$ (linear) correlation of x_t and $y_{t+\tau}$

$\Longrightarrow$ X effects the future of Y

$\Longrightarrow$ $X \rightarrow Y$

$r_{XY}(-\tau) \neq 0 \quad \Longrightarrow \quad Y \rightarrow X$

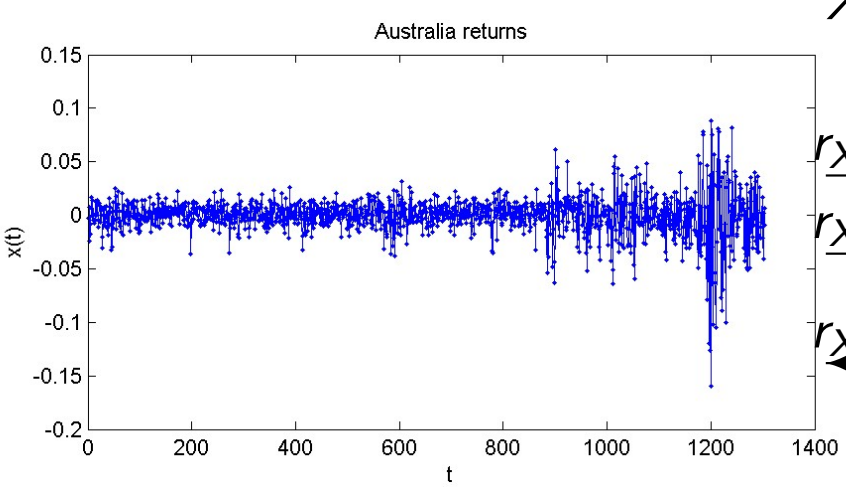
Thus $r_{XY}(\tau)$ and $I_{XY}(\tau)$ indicate the direction of interaction.

Can they also be used as causality measures?

Not the most appropriate, but they have been used in many studies

Example: Returns for USA, UnitedKingdom, Greece and Australia.

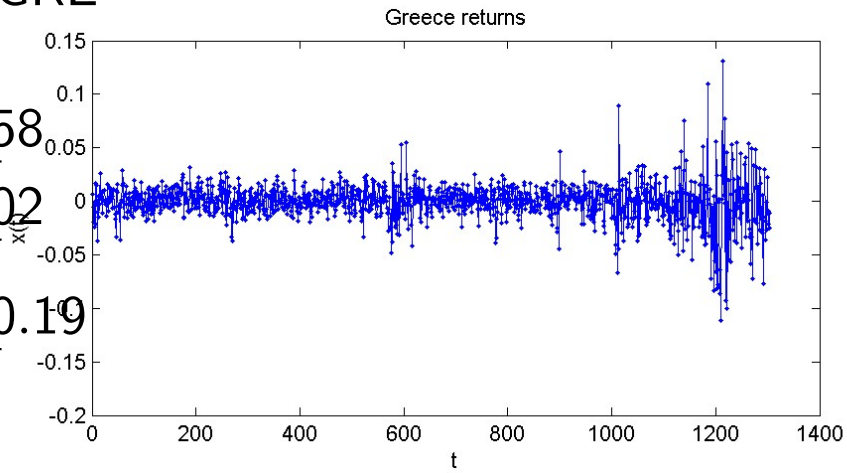
X:AUS, Y:GRE



$r_{XY}(0) = 0.58$

$r_{XY}(1) = 0.02$

$r_{XY}(-1) = 0.19$

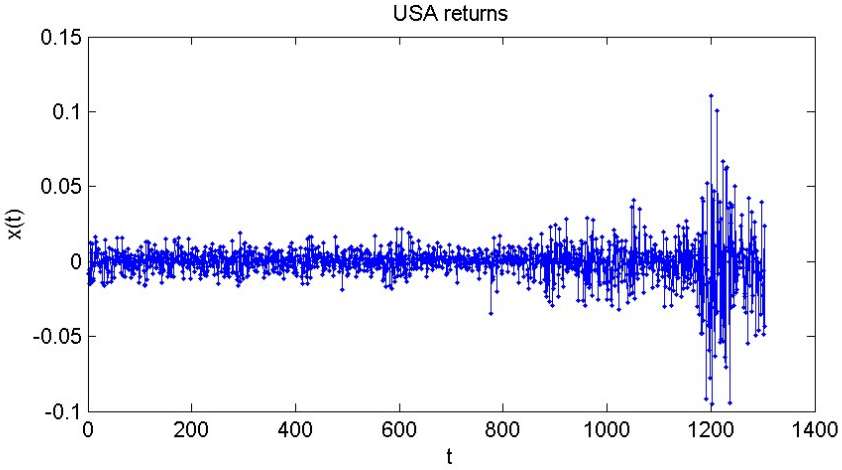


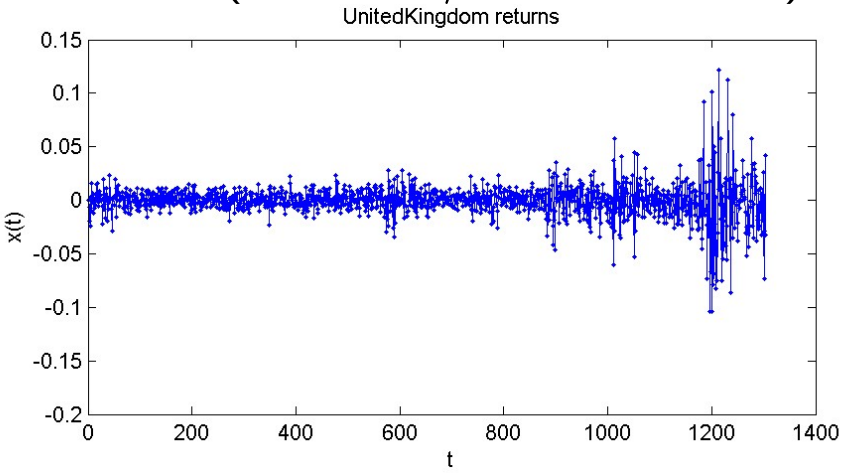
returns:

$x_t = \log(y_t) - \log(y_{t-1})$

Is the measure significant?

Can I draw a link? (directed/non-directed)





Significance randomization test for a correlation / causality measure q,
$\mathrm{H}_0 : q = 0 \qquad \mathrm{H}_1 : q \neq 0$

1. Generate M resampled (surrogate) time series, each by shifting the original observations with a random time step w:

   original time series: $\{x_t\} = \{x_1, x_2, \ldots, x_n\}$

   i-th surrogate time series:
   $\{x_t^{*i}\} = \{x_{w+1}, x_{w+2}, \ldots, x_n, x_1, \ldots, x_{w-1}, x_w\}$

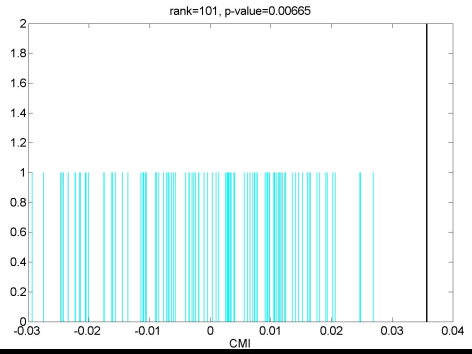
2. Compute the statistic q on the original pair, q_0, and on the M surrogate pairs, $q_1, \ldots, q_M$,

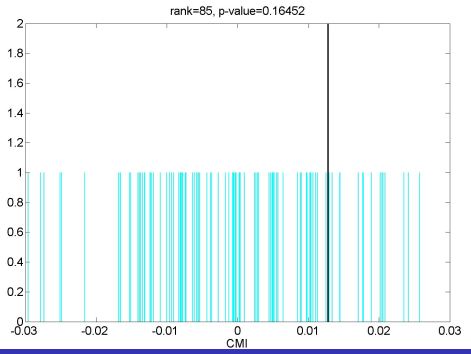
   e.g. $q_0 \equiv r_{XY}(\tau) = \mathrm{Corr}(x_t, y_{t+\tau})$ and $q_i \equiv \mathrm{Corr}(x_t^{*i}, y_{t+\tau}^{*i})$

3. If q_0 is at the tails of the empirical null distribution formed by $q_1, \ldots, q_M$, reject H_0.

   Using rank ordering: for a two-sided test, the p-value of the test is

$$2\frac{r_{q_0} - 0.326}{M+1+0.348} \quad \text{if} \quad r_{q_0} < \frac{M+1}{2}$$
$$2\left(1 - \frac{r_{q_0} - 0.326}{M+1+0.348}\right) \quad \text{if} \quad r_{q_0} \geq \frac{M+1}{2}$$





Example: Returns for USA, UnitedKingdom, Greece and Australia.
Correlation matrix for delay 1, $r_{XY}(1)$

$$R(1) = \begin{bmatrix} & 0.382 & 0.333 & 0.596 \\ 0.049 & & 0.039 & 0.303 \\ 0.096 & 0.001 & & 0.190 \\ 0.031 & -0.001 & -0.021 & \end{bmatrix}$$

Randomization significance test for $r_{XY}(1)$ ($M = 1000$)

Matrix of p-values

$$P(R(1)) = \begin{bmatrix} & 0.0013 & 0.0013 & 0.0033 \\ 0.0732 & & 0.1991 & 0.0013 \\ 0.0073 & 0.8901 & & 0.0033 \\ 0.2450 & 0.9760 & 0.4028 & \end{bmatrix}$$

Adjacency matrix

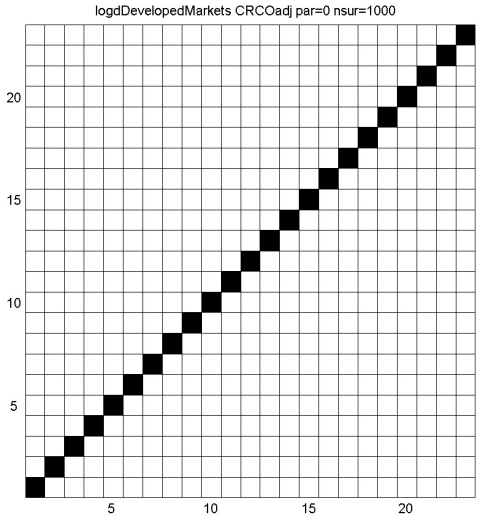
$$A = \begin{bmatrix} & 1 & 1 & 1 \\ 0 & & 0 & 1 \\ 1 & 0 & & 1 \\ 0 & 0 & 0 & \end{bmatrix}$$

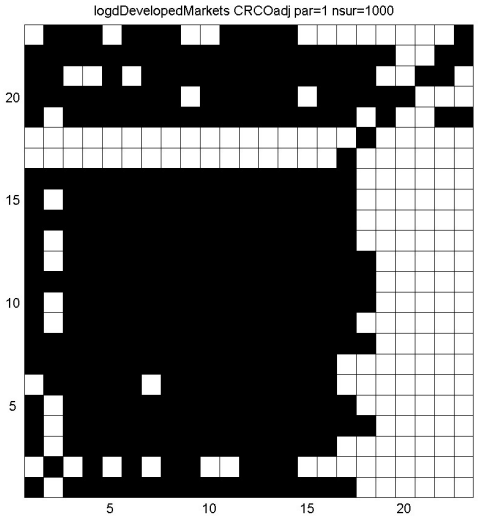
For significance level, say $\alpha = 0.05$, there may be $p < \alpha$ more often than it should be due to multiple testing.
Correction with e.g. False Discovery Rate (FDR)

Network for World Financial Markets

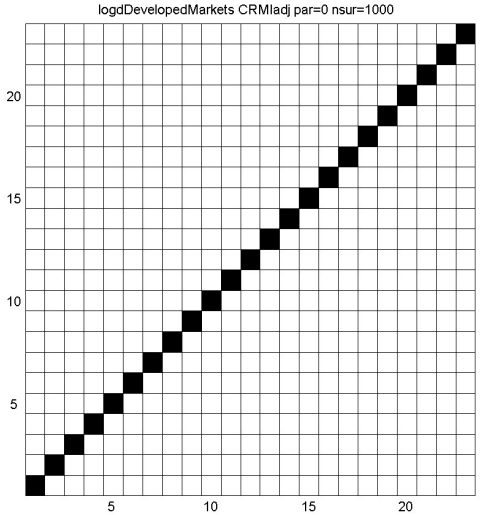
index	*market*
1	Austria
2	Belgium
3	Denmark
4	Finland
5	France
6	Germany
7	Greece
8	Ireland
9	Italy
10	Netherlands
11	Norway
12	Portugal
13	Spain
14	Sweden
15	Switzerland
16	UnitedKingdom
17	USA
18	Canada
19	Australia
20	HongKong
21	Japan
22	NewZealand
23	Singapore

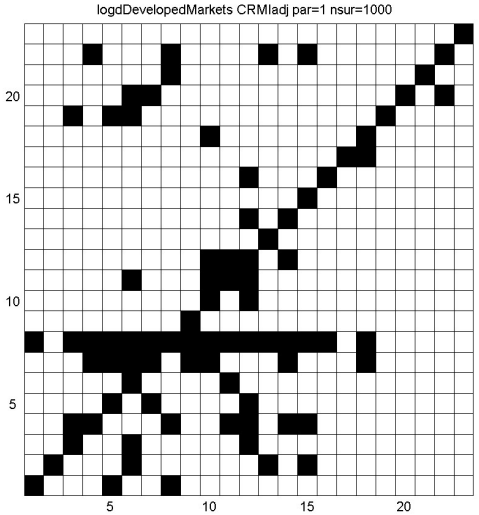
$r_{XY}(0)$ Adjacency matrix $r_{XY}(1)$



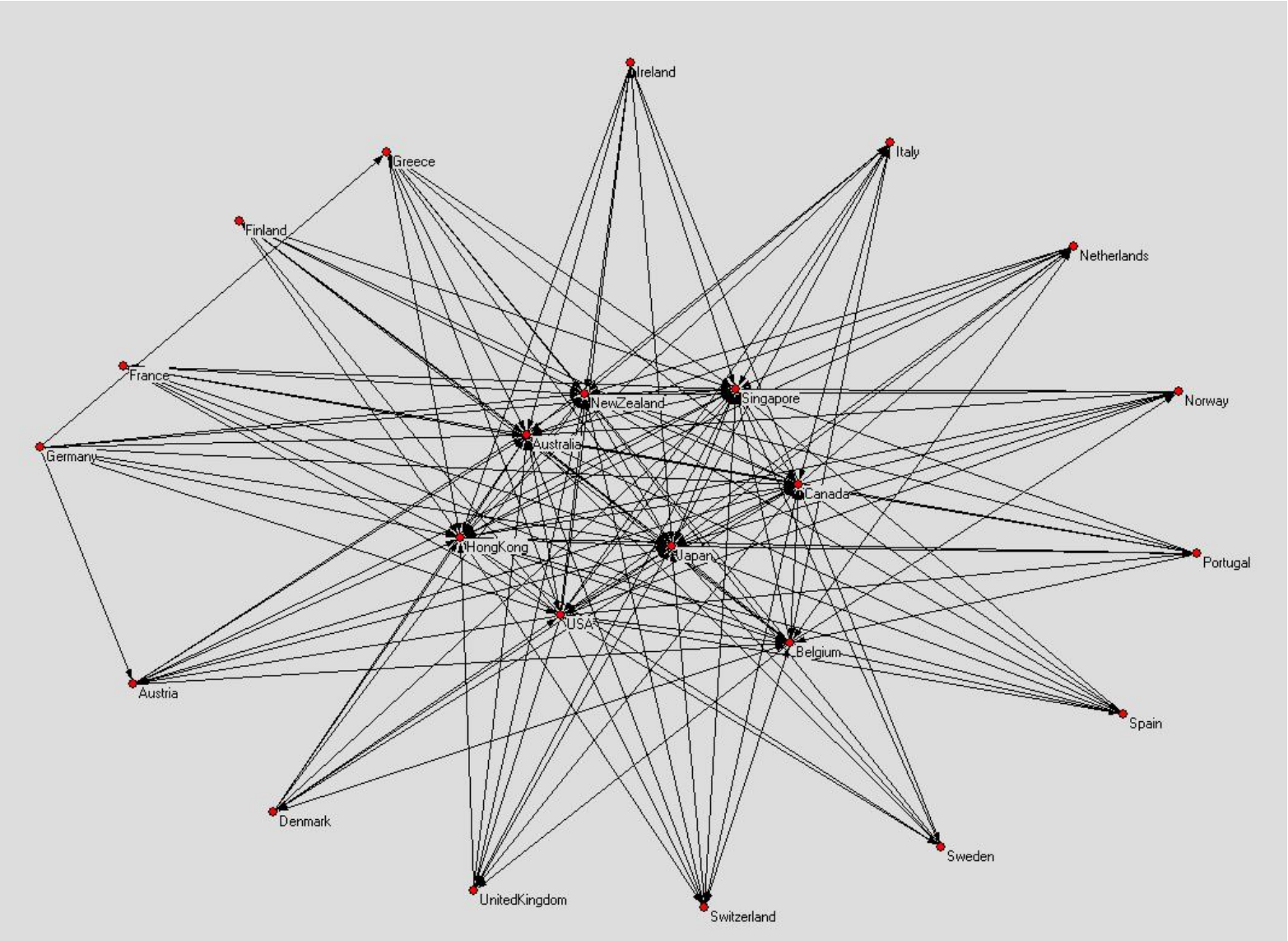


$l_{XY}(0)$ Adjacency matrix $l_{XY}(1)$





Correlation network, nodes: 23 financial markets, directed links: $r_{XY}(1)$



Granger Causality measures

Linear	$X \to Y$	$X \to Y \mid Z$
	Granger Causality Index (GCI)	Conditional / Partial GCI (CGCI/PGCI)
		[Geweke 1982; Guo et al 2008]

Granger Causality measures

Linear	$X \to Y$	$X \to Y \mid Z$
	Granger Causality Index (GCI)	Conditional / Partial GCI (CGCI/PGCI) [Geweke 1982; Guo et al 2008]
		Restricted Granger Causality Index (RGCI) [Siggiridou & Kugiumtzis, 2016]

Granger Causality measures

Linear	$X \to Y$	$X \to Y \mid Z$
	Granger Causality Index (GCI)	Conditional / Partial GCI (CGCI/PGCI) [Geweke 1982; Guo et al 2008]
		Restricted Granger Causality Index (RGCI) [Siggiridou & Kugiumtzis, 2016]
	Directed Coherence (DC) [Wang & Takigawa 1992]	Partial DC (PDC) [Baccala & Sameshima 2001]

Granger Causality measures

Linear	$X \to Y$	$X \to Y \mid Z$
	Granger Causality Index (GCI)	Conditional / Partial GCI (CGCI/PGCI) [Geweke 1982; Guo et al 2008]
		Restricted Granger Causality Index (RGCI) [Siggiridou & Kugiumtzis, 2016]
	Directed Coherence (DC) [Wang & Takigawa 1992]	Partial DC (PDC) [Baccala & Sameshima 2001]
		Restricted PDC (RPDC) [Siggiridou & Kugiumtzis]

Granger Causality measures

Linear	$X \to Y$	$X \to Y \mid Z$
	Granger Causality Index (GCI)	Conditional / Partial GCI (CGCI/PGCI) [Geweke 1982; Guo et al 2008]
		Restricted Granger Causality Index (RGCI) [Siggiridou & Kugiumtzis, 2016]
	Directed Coherence (DC) [Wang & Takigawa 1992]	Partial DC (PDC) [Baccala & Sameshima 2001]
		Restricted PDC (RPDC) [Siggiridou & Kugiumtzis]
	Directed Transfer Function (DTF) [Kaminski & Blinowska 1991]	direct Directed Transfer Function (dDTF) [Korzeniewska et al 2003]

Granger Causality measures

Linear	$X \to Y$	$X \to Y \mid Z$
	Granger Causality Index (GCI)	Conditional / Partial GCI (CGCI/PGCI) [Geweke 1982; Guo et al 2008]
		Restricted Granger Causality Index (RGCI) [Siggiridou & Kugiumtzis, 2016]
	Directed Coherence (DC) [Wang & Takigawa 1992]	Partial DC (PDC) [Baccala & Sameshima 2001]
		Restricted PDC (RPDC) [Siggiridou & Kugiumtzis]
	Directed Transfer Function (DTF) [Kaminski & Blinowska 1991]	direct Directed Transfer Function (dDTF) [Korzeniewska et al 2003]

Nonlinear	$X \to Y$	$X \to Y \mid Z$
	Phase Directionality Index (PDI) [Rosenblum & Pikovsky 2001]	

Granger Causality measures

Linear	$X \to Y$	$X \to Y \mid Z$
	Granger Causality Index (GCI)	Conditional / Partial GCI (CGCI/PGCI) [Geweke 1982; Guo et al 2008]
		Restricted Granger Causality Index (RGCI) [Siggiridou & Kugiumtzis, 2016]
	Directed Coherence (DC) [Wang & Takigawa 1992]	Partial DC (PDC) [Baccala & Sameshima 2001]
		Restricted PDC (RPDC) [Siggiridou & Kugiumtzis]
	Directed Transfer Function (DTF) [Kaminski & Blinowska 1991]	direct Directed Transfer Function (dDTF) [Korzeniewska et al 2003]

Nonlinear	$X \to Y$	$X \to Y \mid Z$
	Phase Directionality Index (PDI) [Rosenblum & Pikovsky 2001]	
	Phase Slope Index (PSI) [Nolte et al 2008]	

Granger Causality measures

Linear	$X \to Y$	$X \to Y \mid Z$
	Granger Causality Index (GCI)	Conditional / Partial GCI (CGCI/PGCI) [Geweke 1982; Guo et al 2008]
		Restricted Granger Causality Index (RGCI) [Siggiridou & Kugiumtzis, 2016]
	Directed Coherence (DC) [Wang & Takigawa 1992]	Partial DC (PDC) [Baccala & Sameshima 2001]
		Restricted PDC (RPDC) [Siggiridou & Kugiumtzis]
	Directed Transfer Function (DTF) [Kaminski & Blinowska 1991]	direct Directed Transfer Function (dDTF) [Korzeniewska et al 2003]

Nonlinear	$X \to Y$	$X \to Y \mid Z$
	Phase Directionality Index (PDI) [Rosenblum & Pikovsky 2001]	
	Phase Slope Index (PSI) [Nolte et al 2008]	
	Inderdependence measures (InderDep) [Arnhold et al 1999; Chicharro & Andrzejak 2009]	

Granger Causality measures

Linear	$X \to Y$	$X \to Y \mid Z$
	Granger Causality Index (GCI)	Conditional / Partial GCI (CGCI/PGCI) [Geweke 1982; Guo et al 2008]
		Restricted Granger Causality Index (RGCI) [Siggiridou & Kugiumtzis, 2016]
	Directed Coherence (DC) [Wang & Takigawa 1992]	Partial DC (PDC) [Baccala & Sameshima 2001]
		Restricted PDC (RPDC) [Siggiridou & Kugiumtzis]
	Directed Transfer Function (DTF) [Kaminski & Blinowska 1991]	direct Directed Transfer Function (dDTF) [Korzeniewska et al 2003]

Nonlinear	$X \to Y$	$X \to Y \mid Z$
	Phase Directionality Index (PDI) [Rosenblum & Pikovsky 2001]	
	Phase Slope Index (PSI) [Nolte et al 2008]	
	Inderdependence measures (InderDep) [Arnhold et al 1999; Chicharro & Andrzejak 2009]	
	Mean Conditional Recurrence (MCR) [Romano et al 2007]	

Granger Causality measures

Linear	$X \to Y$	$X \to Y \mid Z$
	Granger Causality Index (GCI)	Conditional / Partial GCI (CGCI/PGCI) [Geweke 1982; Guo et al 2008]
		Restricted Granger Causality Index (RGCI) [Siggiridou & Kugiumtzis, 2016]
	Directed Coherence (DC) [Wang & Takigawa 1992]	Partial DC (PDC) [Baccala & Sameshima 2001]
		Restricted PDC (RPDC) [Siggiridou & Kugiumtzis]
	Directed Transfer Function (DTF) [Kaminski & Blinowska 1991]	direct Directed Transfer Function (dDTF) [Korzeniewska et al 2003]

Nonlinear	$X \to Y$	$X \to Y \mid Z$
	Phase Directionality Index (PDI) [Rosenblum & Pikovsky 2001]	
	Phase Slope Index (PSI) [Nolte et al 2008]	
	Inderdependence measures (InderDep) [Arnhold et al 1999; Chicharro & Andrzejak 2009]	
	Mean Conditional Recurrence (MCR) [Romano et al 2007]	Prediction Improvement (PredImprov) [Faes et al 2008]

Granger Causality measures

Linear	$X \to Y$	$X \to Y \mid Z$
	Granger Causality Index (GCI)	Conditional / Partial GCI (CGCI/PGCI) [Geweke 1982; Guo et al 2008]
		Restricted Granger Causality Index (RGCI) [Siggiridou & Kugiumtzis, 2016]
	Directed Coherence (DC) [Wang & Takigawa 1992]	Partial DC (PDC) [Baccala & Sameshima 2001]
		Restricted PDC (RPDC) [Siggiridou & Kugiumtzis]
	Directed Transfer Function (DTF) [Kaminski & Blinowska 1991]	direct Directed Transfer Function (dDTF) [Korzeniewska et al 2003]

Nonlinear	$X \to Y$	$X \to Y \mid Z$
	Phase Directionality Index (PDI) [Rosenblum & Pikovsky 2001]	
	Phase Slope Index (PSI) [Nolte et al 2008]	
	Inderdependence measures (InderDep) [Arnhold et al 1999; Chicharro & Andrzejak 2009]	
	Mean Conditional Recurrence (MCR) [Romano et al 2007]	Prediction Improvement (PredImprov) [Faes et al 2008]
	Transfer Entropy (TE) [Schreiber 2000]	Partial TE (PTE) [Vakorin et al 2009; Papana et al 2011]

Granger Causality measures

Linear	$X \to Y$	$X \to Y \mid Z$
	Granger Causality Index (GCI)	Conditional / Partial GCI (CGCI/PGCI) [Geweke 1982; Guo et al 2008]
		Restricted Granger Causality Index (RGCI) [Siggiridou & Kugiumtzis, 2016]
	Directed Coherence (DC) [Wang & Takigawa 1992]	Partial DC (PDC) [Baccala & Sameshima 2001]
		Restricted PDC (RPDC) [Siggiridou & Kugiumtzis]
	Directed Transfer Function (DTF) [Kaminski & Blinowska 1991]	direct Directed Transfer Function (dDTF) [Korzeniewska et al 2003]

Nonlinear	$X \to Y$	$X \to Y \mid Z$
	Phase Directionality Index (PDI) [Rosenblum & Pikovsky 2001]	
	Phase Slope Index (PSI) [Nolte et al 2008]	
	Inderdependence measures (InderDep) [Arnhold et al 1999; Chicharro & Andrzejak 2009]	
	Mean Conditional Recurrence (MCR) [Romano et al 2007]	Prediction Improvement (PredImprov) [Faes et al 2008]
	Transfer Entropy (TE) [Schreiber 2000]	Partial TE (PTE) [Vakorin et al 2009; Papana et al 2011]
	Symbolic Transfer Entropy (STE) [Staniek & Lehnertz 2008]	Partial STE (PSTE) [Papana et al 2014]

Granger Causality measures

Linear	$X \to Y$	$X \to Y \mid Z$
	Granger Causality Index (GCI)	Conditional / Partial GCI (CGCI/PGCI) [Geweke 1982; Guo et al 2008]
		Restricted Granger Causality Index (RGCI) [Siggiridou & Kugiumtzis, 2016]
	Directed Coherence (DC) [Wang & Takigawa 1992]	Partial DC (PDC) [Baccala & Sameshima 2001]
		Restricted PDC (RPDC) [Siggiridou & Kugiumtzis]
	Directed Transfer Function (DTF) [Kaminski & Blinowska 1991]	direct Directed Transfer Function (dDTF) [Korzeniewska et al 2003]

Nonlinear	$X \to Y$	$X \to Y \mid Z$
	Phase Directionality Index (PDI) [Rosenblum & Pikovsky 2001]	
	Phase Slope Index (PSI) [Nolte et al 2008]	
	Inderdependence measures (InderDep) [Arnhold et al 1999; Chicharro & Andrzejak 2009]	
	Mean Conditional Recurrence (MCR) [Romano et al 2007]	Prediction Improvement (PredImprov) [Faes et al 2008]
	Transfer Entropy (TE) [Schreiber 2000]	Partial TE (PTE) [Vakorin et al 2009; Papana et al 2011]
	Symbolic Transfer Entropy (STE) [Staniek & Lehnertz 2008]	Partial STE (PSTE) [Papana et al 2014]
	Transfer Entropy on Rank Vectors (TERV) [Kugiumtzis 2012]	Partial TERV (PTERV) [Kugiumtzis 2013]

Granger Causality measures

Linear	$X \to Y$	$X \to Y \mid Z$
	Granger Causality Index (GCI)	Conditional / Partial GCI (CGCI/PGCI) [Geweke 1982; Guo et al 2008]
		Restricted Granger Causality Index (RGCI) [Siggiridou & Kugiumtzis, 2016]
	Directed Coherence (DC) [Wang & Takigawa 1992]	Partial DC (PDC) [Baccala & Sameshima 2001]
		Restricted PDC (RPDC) [Siggiridou & Kugiumtzis]
	Directed Transfer Function (DTF) [Kaminski & Blinowska 1991]	direct Directed Transfer Function (dDTF) [Korzeniewska et al 2003]

Nonlinear	$X \to Y$	$X \to Y \mid Z$
	Phase Directionality Index (PDI) [Rosenblum & Pikovsky 2001]	
	Phase Slope Index (PSI) [Nolte et al 2008]	
	Inderdependence measures (InderDep) [Arnhold et al 1999; Chicharro & Andrzejak 2009]	
	Mean Conditional Recurrence (MCR) [Romano et al 2007]	Prediction Improvement (PredImprov) [Faes et al 2008]
	Transfer Entropy (TE) [Schreiber 2000]	Partial TE (PTE) [Vakorin et al 2009; Papana et al 2011]
	Symbolic Transfer Entropy (STE) [Staniek & Lehnertz 2008]	Partial STE (PSTE) [Papana et al 2014]
	Transfer Entropy on Rank Vectors (TERV) [Kugiumtzis 2012]	Partial TERV (PTERV) [Kugiumtzis 2013]
	Mutual Information from Mixed Embedding (MIME) [Vlachos & Kugiumtzis 2010]	Partial MIME (PMIME) [Kugiumtzis 2013]

Linear causality measures (direct and indirect)

Idea of Granger causality $X \to Y$ [Granger 1969]:
predict Y better when including X in the regression model.

Granger Causality Index (GCI) [Brandt & Williams 2007]

Bivariate time series $\{x_t, y_t\}_{t=1}^{n}$
driving system: X, response system: Y

Model 1 (restricted, R, X absent in the model):

$$y_t = \sum_{i=1}^{p} a_i y_{t-i} + e_{R,t}$$

Model 2 (unrestricted, U, X present in the model):

$$y_t = \sum_{i=1}^{p} a_i y_{t-i} + \sum_{i=1}^{p} b_i x_{t-i} + e_{U,t}$$

$$\mathrm{GCI}_{X\to Y} = \ln \frac{\mathrm{Var}(\hat{e}_{R,t})}{\mathrm{Var}(\hat{e}_{U,t})} \qquad \mathrm{GCI}_{X\to Y} > 0 \Rightarrow X \to Y \text{ holds}$$

Parametric significance test for GCI

$\mathrm{GCI}_{X\to Y} > 0$? $\Rightarrow$ Significance test

If X does not Granger causes Y then the contribution of X-lags in the unrestricted model should be insignificant $\Rightarrow$
the terms of X should be insignificant

H_0: $b_i = 0$, for all $i = 1, \ldots, p$
H_1: $b_i \neq 0$, for any of $i = 1, \ldots, p$

Snedecor-Fisher test (F-test):

$$F = \frac{(\mathrm{SSE}^R - \mathrm{SSE}^U)/p}{\mathrm{SSE}^U/\mathrm{ndf}}$$

SSE: sum of squared errors
ndf: number of degrees of freedoms, $\mathrm{ndf} = (n-p) - 2p$,
$n-p$: number of equations,
$2p$: number of coefficients in the U-model.

Linear causality measures (direct and indirect)

Conditional Granger Causality Index (CGCI)

K time series $\{x_t, y_t\}_{t=1}^n$ and $\{\mathbf{z}_t\}_{t=1}^n = \{z_{1,t}, z_{2,t}, \ldots, z_{K-2,t}\}_{t=1}^n$
driving system: X, response system: Y,
conditioning on system Z, $Z = \{Z_1, Z_2, \ldots, Z_{K-2}\}$

Model 1 (restricted, R, X absent in the model):

$$y_t = \sum_{i=1}^{p} a_i y_{t-i} + \sum_{i=1}^{p} A_i \mathbf{z}_{t-i} + e_{R,t}$$

Model 2 (unrestricted, U, X present in the model):

$$y_t = \sum_{i=1}^{p} a_i y_{t-i} + \sum_{i=1}^{p} b_i x_{t-i} + \sum_{i=1}^{p} A_i \mathbf{z}_{t-i} + e_{U,t}$$

$$\mathrm{CGCI}_{X \to Y|Z} = \ln \frac{\mathrm{Var}(\hat{e}_{R,t})}{\mathrm{Var}(\hat{e}_{U,t})}$$

Parametric significance test for CGCI

$\text{CGCI}_{X \to Y|Z} > 0$? $\Rightarrow$ Significance test as for GCI

H_0: $b_i = 0$, for all $i = 1, \ldots, p$
H_1: $b_i \neq 0$, for any of $i = 1, \ldots, p$

$$F = \frac{(\text{SSE}^R - \text{SSE}^U)/p}{\text{SSE}^U/\text{ndf}}$$

$\text{ndf} = (n - p) - Kp$,
$n - p$: number of equations,
Kp: number of coefficients in the U-model.

Model order and embedding parameters

VAR model for Y

$$y_t = \sum_{i=1}^{p} a_i y_{t-i} + \sum_{i=1}^{p} b_i x_{t-i} + e_{U,t}$$

y_{t+1} is given in terms of $\{y_t, y_{t-1}, \ldots, y_{t-p+1}\}$ and $\{x_t, x_{t-1}, \ldots, x_{t-p+1}\}$.

$\mathbf{y}_t = [y_t, y_{t-1}, \ldots, y_{t-p+1}]$: vector of lagged Y

let the lag step be $\tau \geq 1 \quad \Rightarrow \quad \mathbf{y}_t = [y_t, y_{t-\tau}, \ldots, y_{t-(p-1)\tau}]$:

τ, p: embedding parameters (generally different for X and Y)

State space reconstruction:

$\mathbf{x}_t = [x_t, x_{t-\tau_x}, \ldots, x_{t-(m_x-1)\tau_x}]'$, embedding parameters: m_x, τ_x

$\mathbf{y}_t = [y_t, y_{t-\tau_y}, \ldots, y_{t-(m_y-1)\tau_y}]'$, embedding parameters: m_y, τ_y

y_{t+1}: future state of Y

Nonlinear causality measures (direct and indirect)

$\mathbf{x}_t = [x_t, x_{t-\tau_x}, \ldots, x_{t-(m_x-1)\tau_x}]'$, embedding parameters: m_x, τ_x
$\mathbf{y}_t = [y_t, y_{t-\tau_y}, \ldots, y_{t-(m_y-1)\tau_y}]'$, embedding parameters: m_y, τ_y
y_{t+1}: future state of Y

Transfer Entropy (TE) [Schreiber, 2000]

Measure the effect of X on Y at one time step ahead, accounting (conditioning) for the effect from its own current state

$$\begin{aligned}
\mathrm{TE}_{X\to Y} &= I(y_{t+1}; \mathbf{x}_t | \mathbf{y}_t) \\
&= H(\mathbf{x}_t, \mathbf{y}_t) - H(y_{t+1}, \mathbf{x}_t, \mathbf{y}_t) + H(y_{t+1}, \mathbf{y}_t) - H(\mathbf{y}_t) \\
&= \sum p(y_{t+T}, \mathbf{x}_t, \mathbf{y}_t) \log \frac{p(y_{t+T}|\mathbf{x}_t, \mathbf{y}_t)}{p(y_{t+T}|\mathbf{y}_t)}
\end{aligned}$$

Joint entropies (and distributions) can have high dimension!

Entropy estimates from nearest neighbors [Kraskov et al, 2004]

Nonlinear causality measures (direct and indirect)

driving system: X, response system: Y,
conditioning on system Z, $Z = \{Z_1, Z_2, \ldots, Z_{K-2}\}$
join all $K-2$ z-reconstructed vectors: $\mathbf{Z}_t = [\mathbf{z}_{1,t}, \ldots, \mathbf{z}_{K-2,t}]$

Partial Transfer Entropy (PTE) [Papana et al, 2012]

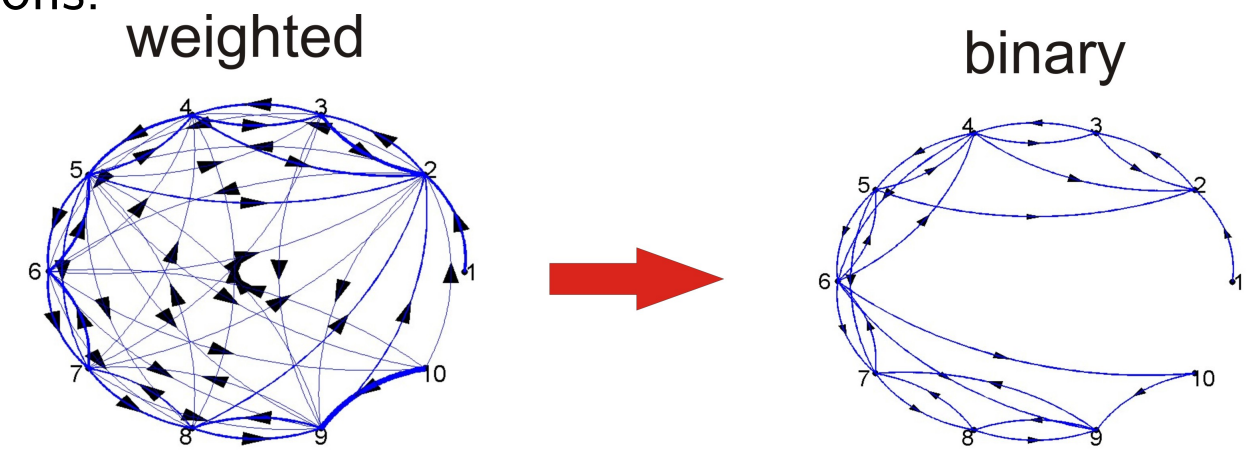
Measure the effect of X on Y at T times ahead, accounting (conditioning) for the effect from its own current state and the current state of the other variables except X.

$$\mathrm{PTE}_{X \to Y|Z} = I(y_{t+1}; \mathbf{x}_t | \mathbf{y}_t, \mathbf{Z}_t)$$
$$= H(\mathbf{x}_t, \mathbf{y}_t | \mathbf{Z}_t) - H(y_{t+1}, \mathbf{x}_t, \mathbf{y}_t | \mathbf{Z}_t) + H(y_{t+1}, \mathbf{y}_t | \mathbf{Z}_t) - H(\mathbf{y}_t | \mathbf{Z}_t)$$

Joint entropies (and distributions) can have very high dimension!

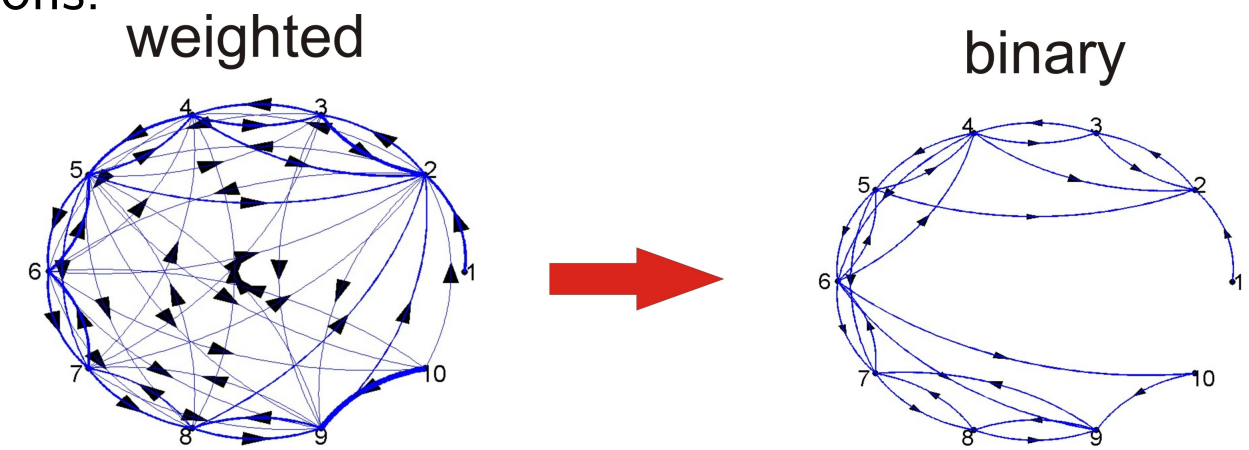
How to assess the presence of a connection?

Three possible ways to convert a network of weighted connections (the Granger causality measure) to a network of binary connections:



How to assess the presence of a connection?

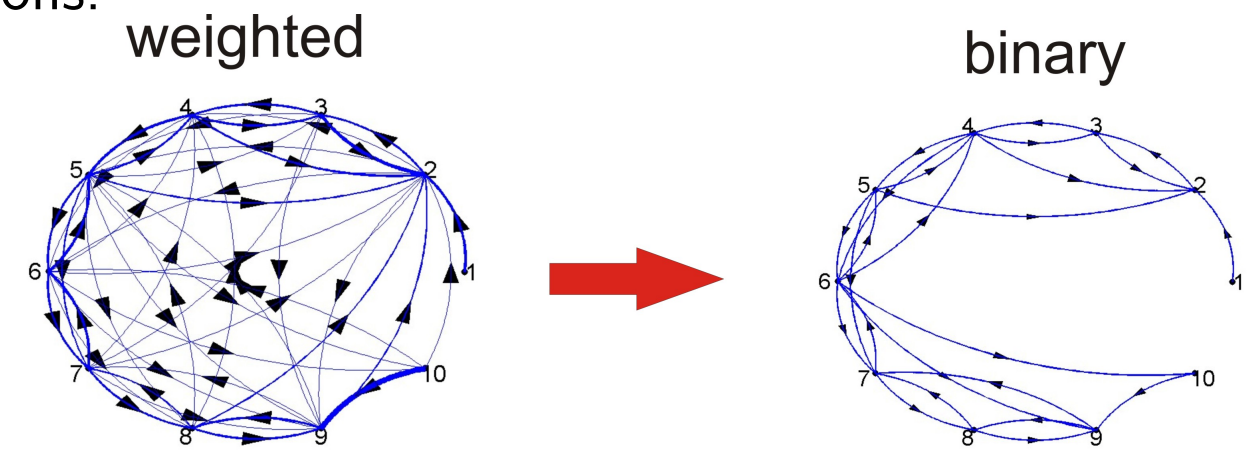
Three possible ways to convert a network of weighted connections (the Granger causality measure) to a network of binary connections:



1. Threshold on the measure magnitude, $q(i \to j) > \text{thr}$.

How to assess the presence of a connection?

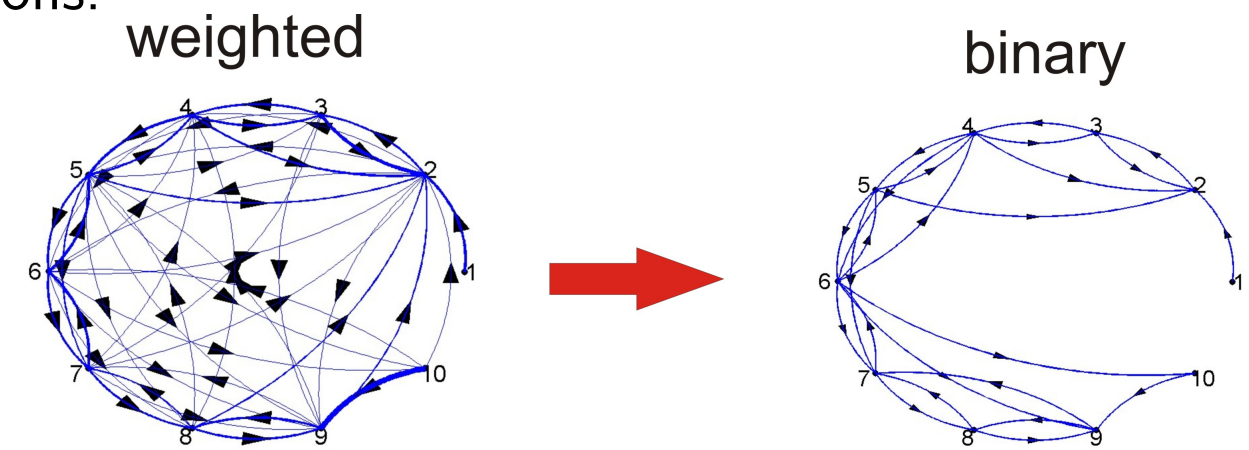
Three possible ways to convert a network of weighted connections (the Granger causality measure) to a network of binary connections:



1. Threshold on the measure magnitude, $q(i \to j) > \text{thr}$.
2. Threshold on the network density, only the $d\%$ largest $q(i \to j)$.

How to assess the presence of a connection?

Three possible ways to convert a network of weighted connections (the Granger causality measure) to a network of binary connections:



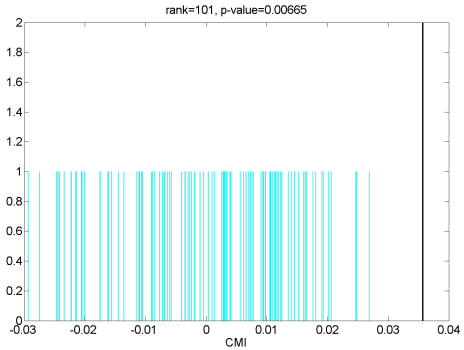
1. Threshold on the measure magnitude, $q(i \to j) > \text{thr}$.
2. Threshold on the network density, only the $d\%$ largest $q(i \to j)$.
3. Significance test on each $q(i \to j)$. Threshold, e.g. $\alpha = 0.05$ on the p-value of the test.
   Parametric or resampling test (resampling test for a nonlinear causality measure).

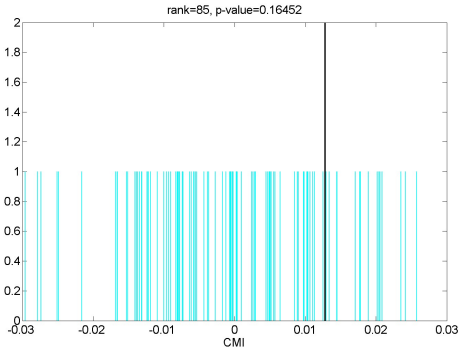
Significance resampling test

$\mathrm{H}_0 : q = 0 \qquad \mathrm{H}_1 : q \neq 0 \qquad$ for a correlation/causality measure q

1. Generate M resampled (surrogate) time series.
   Simple approach: time-shifting by a random time step w:
   original time series: $\{x_t\} = \{x_1, x_2, \ldots, x_n\}$
   i-th surrogate: $\{x_t^{*i}\} = \{x_{w+1}, x_{w+2}, \ldots, x_n, x_1, \ldots, x_{w-1}, x_w\}$
2. Compute the statistic q on the original pair, q_0, and on the M surrogate pairs, $q_1, \ldots, q_M$,
3. If q_0 is at the tails of the empirical null distribution formed by $q_1, \ldots, q_M$, reject H_0.
   Using rank ordering: for a two-sided test, the p-value of the test is

$$\begin{array}{ll} 2\frac{r_{q_0}-0.326}{M+1+0.348} & \text{if} \quad r_{q_0} < \frac{M+1}{2} \\ 2(1-\frac{r_{q_0}-0.326}{M+1+0.348}) & \text{if} \quad r_{q_0} \geq \frac{M+1}{2} \end{array}$$





The problem of multiple testing

Significance resampling test on $q(i \to j)$ for each pair (X_i, X_j).

The problem of multiple testing

Significance resampling test on $q(i \to j)$ for each pair (X_i, X_j).

$K(K-1)$ significance tests $\Longrightarrow$ Correction for multiple testing

The problem of multiple testing

Significance resampling test on $q(i \to j)$ for each pair (X_i, X_j).

$K(K-1)$ significance tests $\Longrightarrow$ Correction for multiple testing

Popular choice:

False Discovery Rate (FDR) [Benjamini & Hochberg, 1995]

- $K(K-1)$ p-values in ascending order: $p_{(1)}, p_{(2)}, \ldots, p_{(K(K-1))}$
- Rejection for the k tests with $p \leq p_{(k)}$, where $p_{(k)}$ is the largest p-value for which $p_{(k)} < k\alpha/(K(K-1))$.

The problem of multiple testing

Significance resampling test on $q(i \to j)$ for each pair (X_i, X_j).

$K(K-1)$ significance tests $\Longrightarrow$ Correction for multiple testing

Popular choice:

False Discovery Rate (FDR) [Benjamini & Hochberg, 1995]

- $K(K-1)$ p-values in ascending order: $p_{(1)}, p_{(2)}, \ldots, p_{(K(K-1))}$
- Rejection for the k tests with $p \leq p_{(k)}$, where $p_{(k)}$ is the largest p-value for which $p_{(k)} < k\alpha/(K(K-1))$.

Small p-value can only be obtained with large number of surrogates

The problem of multiple testing

Significance resampling test on $q(i \to j)$ for each pair (X_i, X_j).

$K(K-1)$ significance tests $\Longrightarrow$ Correction for multiple testing

Popular choice:

False Discovery Rate (FDR) [Benjamini & Hochberg, 1995]

- $K(K-1)$ p-values in ascending order: $p_{(1)}, p_{(2)}, \ldots, p_{(K(K-1))}$
- Rejection for the k tests with $p \leq p_{(k)}$, where $p_{(k)}$ is the largest p-value for which $p_{(k)} < k\alpha/(K(K-1))$.

Small p-value can only be obtained with large number of surrogates

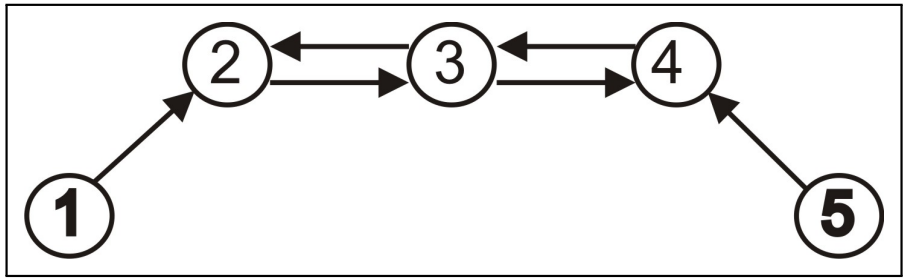
When K gets large, FDR requires **huge** M (impractical).

Example: coupled Henon maps

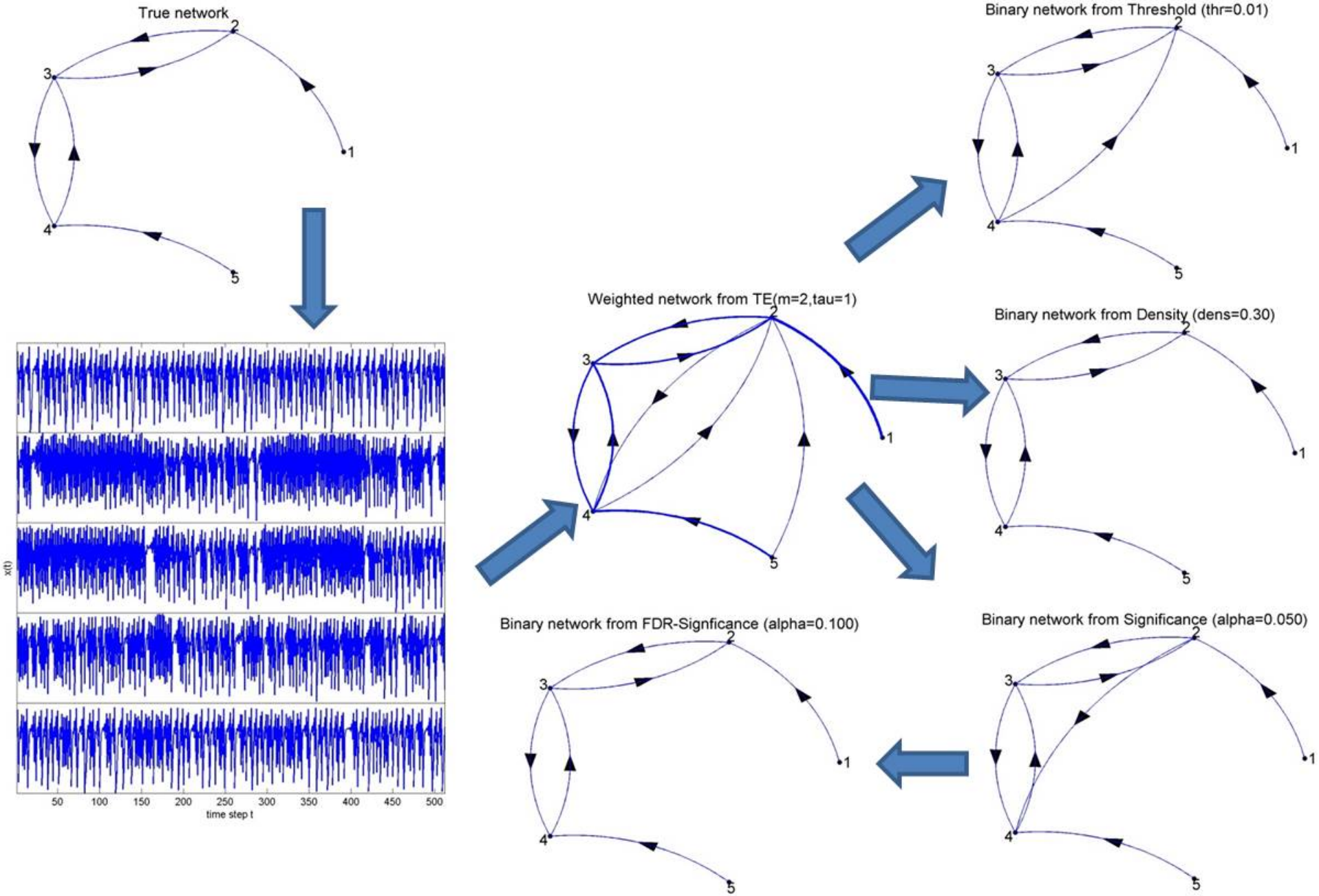
$$
\begin{aligned}
x_{1,t+1} &= 1.4 - x_{1,t}^2 + 0.3x_{1,t-1} \\
x_{i,t+1} &= 1.4 - (0.5C(x_{i-1,t} + x_{i+1,t}) + (1-C)x_{i,t})^2 + 0.3x_{i,t-1} \\
x_{K,t+1} &= 1.4 - x_{K,t}^2 + 0.3x_{K,t-1}
\end{aligned}
$$

C: coupling strength [Politi & Torcini, 1992]

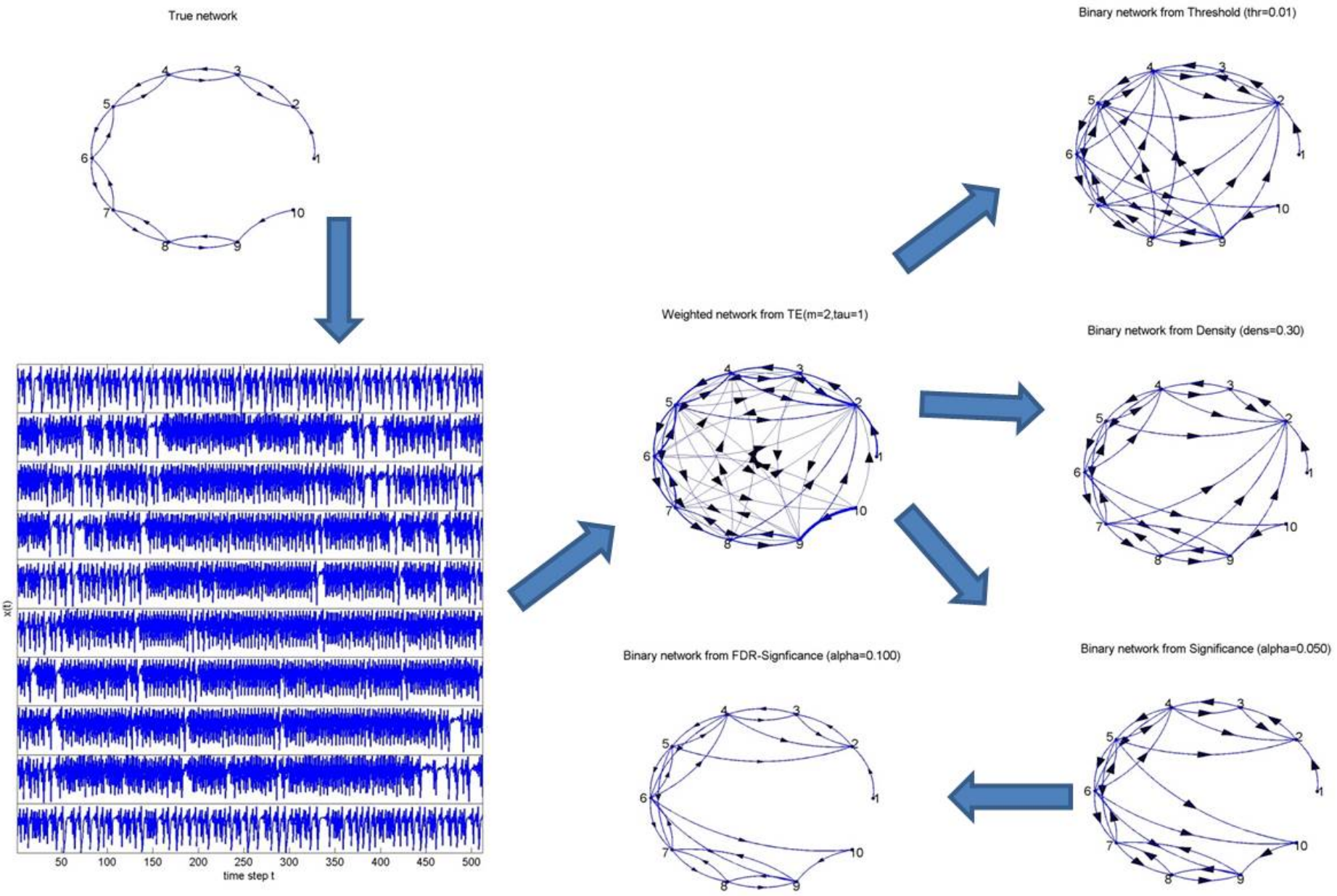
Network structure
for $K = 5$



Example, TE, $K = 5$



Example, TE, $K = 10$



What if there are many observed variables?

The **curse of dimensionality**:

• For FDR, in general $M \sim K(K-1)/\alpha$. When K gets large, **huge** M may be required (impractical).

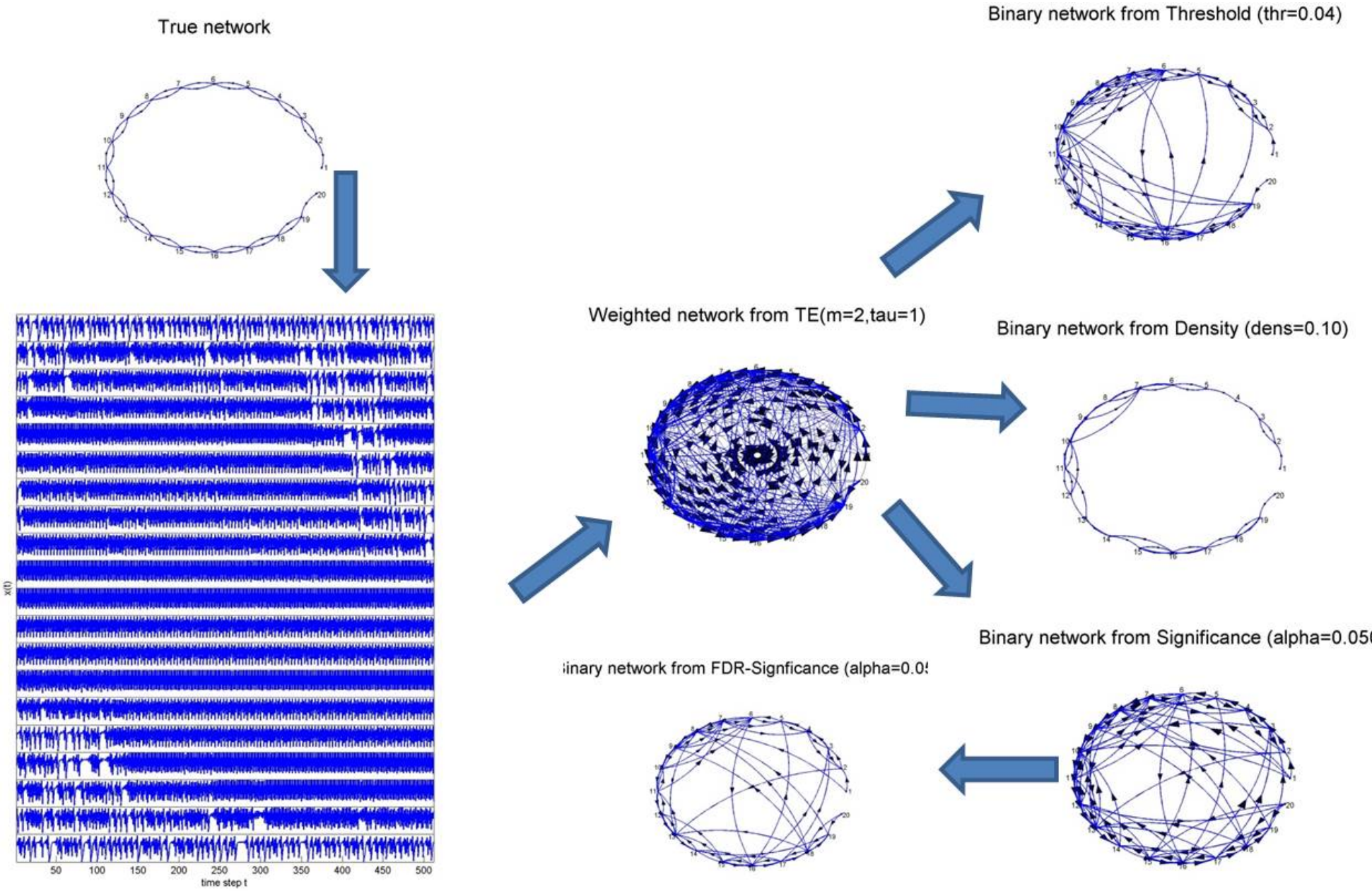
What if there are many observed variables?

The **curse of dimensionality**:

• For FDR, in general $M \sim K(K-1)/\alpha$. When K gets large, **huge** M may be required (impractical).

• For $K > 2$, bivariate measures are likely to produce false couplings (indirect connections).

Example, TE, $K = 20$



What if there are many observed variables?

The **curse of dimensionality**:

• For FDR, in general $M \sim K(K-1)/\alpha$. When K gets large, **huge** M may be required (impractical).

• For $K > 2$, bivariate measures are likely to produce false couplings (indirect connections).

What if there are many observed variables?

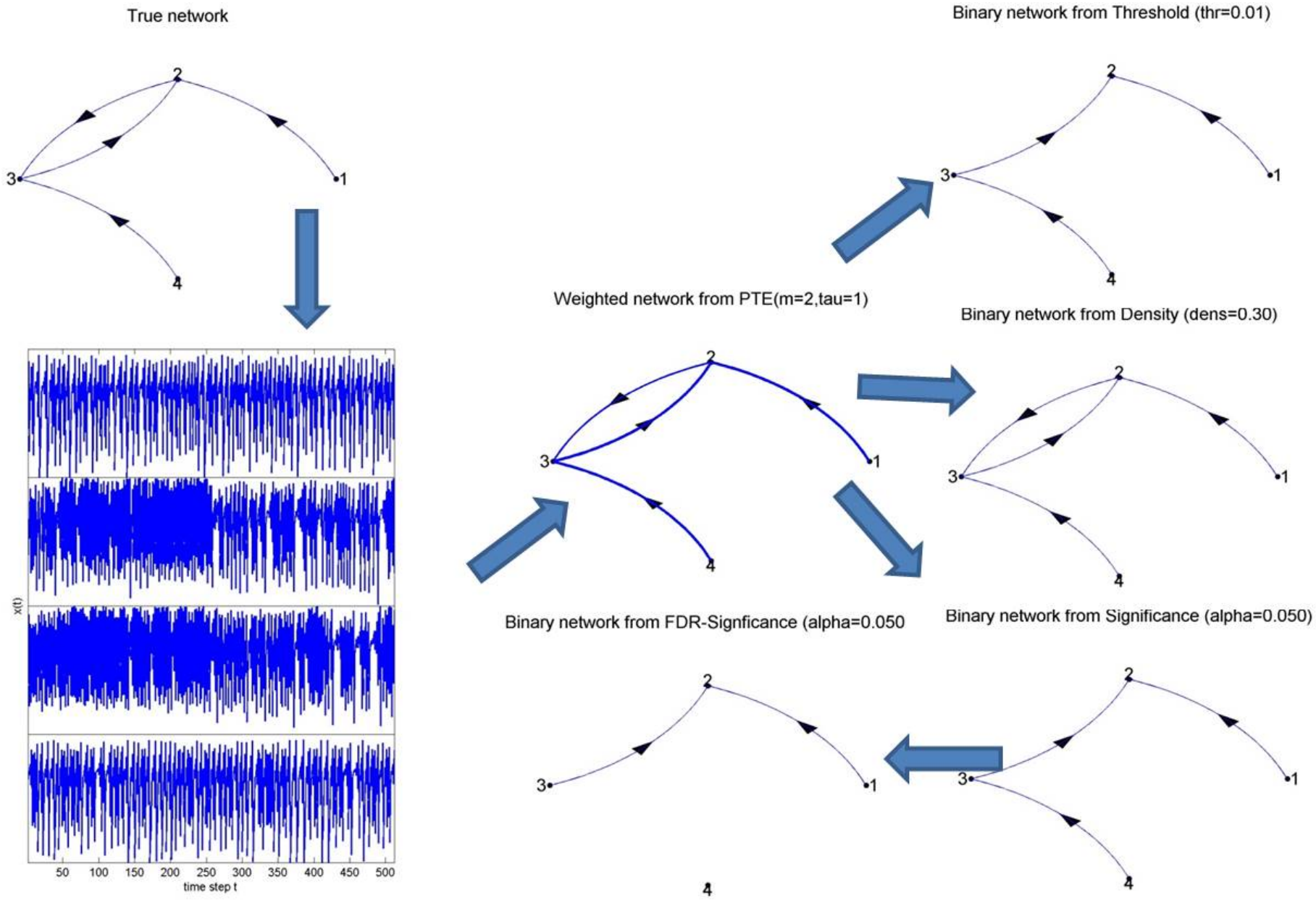
The **curse of dimensionality**:

• For FDR, in general $M \sim K(K-1)/\alpha$. When K gets large, **huge** M may be required (impractical).

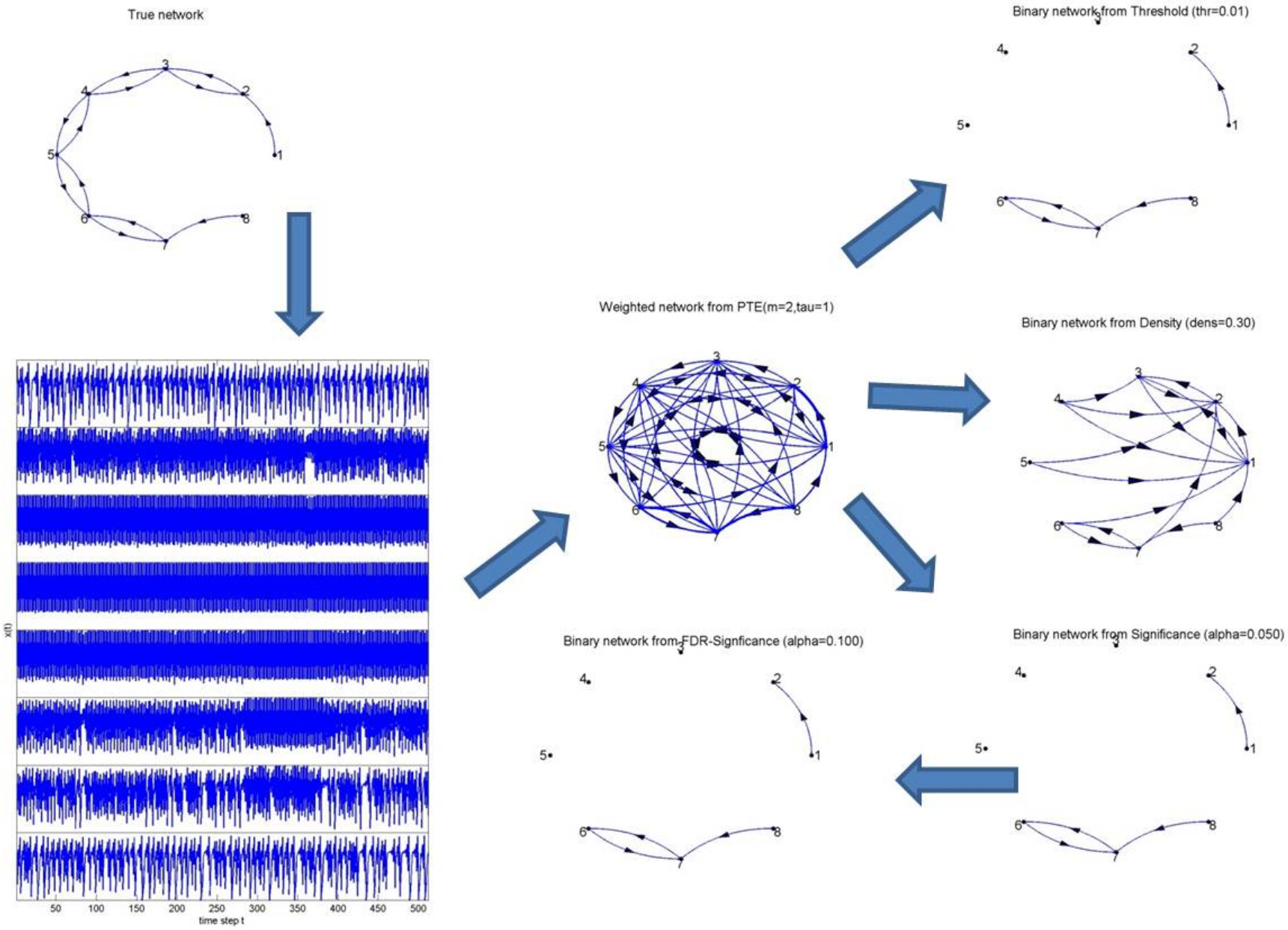
• For $K > 2$, bivariate measures are likely to produce false couplings (indirect connections).

• Multivariate measures require long time series, e.g. $\mathrm{PTE}_{X \to Y|Z} = I(y_{t+1}; \mathbf{x}_t | \mathbf{y}_t, \mathbf{Z}_t)$ requires the estimation of entropy of $[y_{t+1}, \mathbf{x}_t, \mathbf{y}_t, \mathbf{Z}_t]'$ of dimension $1 + Km$.

Example, PTE, $K = 4$



Example, PTE, $K = 8$



Granger Causality using Dimension Reduction

K time series $\{x_t, y_t\}_{t=1}^n$ and $\{\mathbf{z}_t\}_{t=1}^n = \{z_{1,t}, z_{2,t}, \ldots, z_{K-2,t}\}_{t=1}^n$
driving system: X, response system: Y,
conditioning on system Z, $Z = \{Z_1, Z_2, \ldots, Z_{K-2}\}$

Granger Causality using Dimension Reduction

K time series $\{x_t, y_t\}_{t=1}^n$ and $\{\mathbf{z}_t\}_{t=1}^n = \{z_{1,t}, z_{2,t}, \ldots, z_{K-2,t}\}_{t=1}^n$
driving system: X, response system: Y,
conditioning on system Z, $Z = \{Z_1, Z_2, \ldots, Z_{K-2}\}$

If K large with respect to n multivariate Granger causality is problematic ("the curse of dimensionality").

Granger Causality using Dimension Reduction

K time series $\{x_t, y_t\}_{t=1}^n$ and $\{\mathbf{z}_t\}_{t=1}^n = \{z_{1,t}, z_{2,t}, \ldots, z_{K-2,t}\}_{t=1}^n$
driving system: X, response system: Y,
conditioning on system Z, $Z = \{Z_1, Z_2, \ldots, Z_{K-2}\}$

If K large with respect to n multivariate Granger causality is problematic ("the curse of dimensionality").

In CGCI, the VAR model

$$y_t = \sum_{i=1}^{p} a_i y_{t-i} + \sum_{i=1}^{p} b_i x_{t-i} + \sum_{i=1}^{p} A_i \mathbf{z}_{t-i} + e_{U,t}$$

has Kp lagged variables.

Granger Causality using Dimension Reduction

K time series $\{x_t, y_t\}_{t=1}^n$ and $\{\mathbf{z}_t\}_{t=1}^n = \{z_{1,t}, z_{2,t}, \ldots, z_{K-2,t}\}_{t=1}^n$
driving system: X, response system: Y,
conditioning on system Z, $Z = \{Z_1, Z_2, \ldots, Z_{K-2}\}$

If K large with respect to n multivariate Granger causality is problematic ("the curse of dimensionality").

In CGCI, the VAR model

$$y_t = \sum_{i=1}^{p} a_i y_{t-i} + \sum_{i=1}^{p} b_i x_{t-i} + \sum_{i=1}^{p} A_i \mathbf{z}_{t-i} + e_{U,t}$$

has Kp lagged variables.

RCGCI (linear multivariate) and PMIME (nonlinear multivariate) apply dimension reduction.

RCGCI [Siggiridou & Kugiumtzis, 2016]

• A subset of all Kp lagged variables is selected using a modification of the scheme of backward-in-time selection (mBTS) [Vlachos & Kugiumtzis, 2013]

$$\mathbf{w}_y = [\mathbf{w}_{t,1}\ \mathbf{w}_{t,2}\ \ldots \mathbf{w}_{t,K}]$$

where for $k = 1, \ldots, K$

$$\mathbf{w}_{t,k} = \{X_{k,t-\tau(1)}, \ldots, X_{k,t-\tau(p_k)}\},$$

RCGCI [Siggiridou & Kugiumtzis, 2016]

• A subset of all Kp lagged variables is selected using a modification of the scheme of backward-in-time selection (mBTS) [Vlachos & Kugiumtzis, 2013]

$$\mathbf{w}_y = [\mathbf{w}_{t,1}\ \mathbf{w}_{t,2}\ \ldots \mathbf{w}_{t,K}]$$

where for $k = 1, \ldots, K$

$$\mathbf{w}_{t,k} = \{X_{k,t-\tau(1)}, \ldots, X_{k,t-\tau(p_k)}\},$$

• The sparse VAR model is

$$y_t = \sum_{k=1}^{K} \mathbf{a}_k \mathbf{w}_{t,k}^T + e_{U,t}$$

RCGCI [Siggiridou & Kugiumtzis, 2016]

- A subset of all Kp lagged variables is selected using a modification of the scheme of backward-in-time selection (mBTS) [Vlachos & Kugiumtzis, 2013]

$$\mathbf{w}_y = [\mathbf{w}_{t,1}\ \mathbf{w}_{t,2}\ \ldots \mathbf{w}_{t,K}]$$

where for $k = 1, \ldots, K$

$$\mathbf{w}_{t,k} = \{X_{k,t-\tau(1)}, \ldots, X_{k,t-\tau(p_k)}\},$$

- The sparse VAR model is

$$y_t = \sum_{k=1}^{K} \mathbf{a}_k \mathbf{w}_{t,k}^T + e_{U,t}$$

- This is the U-model

RCGCI [Siggiridou & Kugiumtzis, 2016]

- A subset of all Kp lagged variables is selected using a modification of the scheme of backward-in-time selection (mBTS) [Vlachos & Kugiumtzis, 2013]

$$\mathbf{w}_y = [\mathbf{w}_{t,1}\ \mathbf{w}_{t,2}\ \ldots \mathbf{w}_{t,K}]$$

where for $k = 1, \ldots, K$

$$\mathbf{w}_{t,k} = \{X_{k,t-\tau(1)}, \ldots, X_{k,t-\tau(p_k)}\},$$

- The sparse VAR model is

$$y_t = \sum_{k=1}^{K} \mathbf{a}_k \mathbf{w}_{t,k}^T + e_{U,t}$$

- This is the U-model

RCGCI [Siggiridou & Kugiumtzis, 2016]

• A subset of all Kp lagged variables is selected using a modification of the scheme of backward-in-time selection (mBTS) [Vlachos & Kugiumtzis, 2013]

$$\mathbf{w}_y = [\mathbf{w}_{t,1}\ \mathbf{w}_{t,2}\ \ldots \mathbf{w}_{t,K}]$$

where for $k = 1, \ldots, K$

$$\mathbf{w}_{t,k} = \{X_{k,t-\tau(1)}, \ldots, X_{k,t-\tau(p_k)}\},$$

• The sparse VAR model is

$$y_t = \sum_{k=1}^{K} \mathbf{a}_k \mathbf{w}_{t,k}^T + e_{U,t}$$

• This is the U-model

• The R-model is obtained by omitting the lagged components of X.

RCGCI [Siggiridou & Kugiumtzis, 2016]

• A subset of all Kp lagged variables is selected using a modification of the scheme of backward-in-time selection (mBTS) [Vlachos & Kugiumtzis, 2013]

$$\mathbf{w}_y = [\mathbf{w}_{t,1}\ \mathbf{w}_{t,2}\ \ldots \mathbf{w}_{t,K}]$$

where for $k = 1, \ldots, K$

$$\mathbf{w}_{t,k} = \{X_{k,t-\tau(1)}, \ldots, X_{k,t-\tau(p_k)}\},$$

• The sparse VAR model is

$$y_t = \sum_{k=1}^{K} \mathbf{a}_k \mathbf{w}_{t,k}^T + e_{U,t}$$

• This is the U-model

• The R-model is obtained by omitting the lagged components of X.

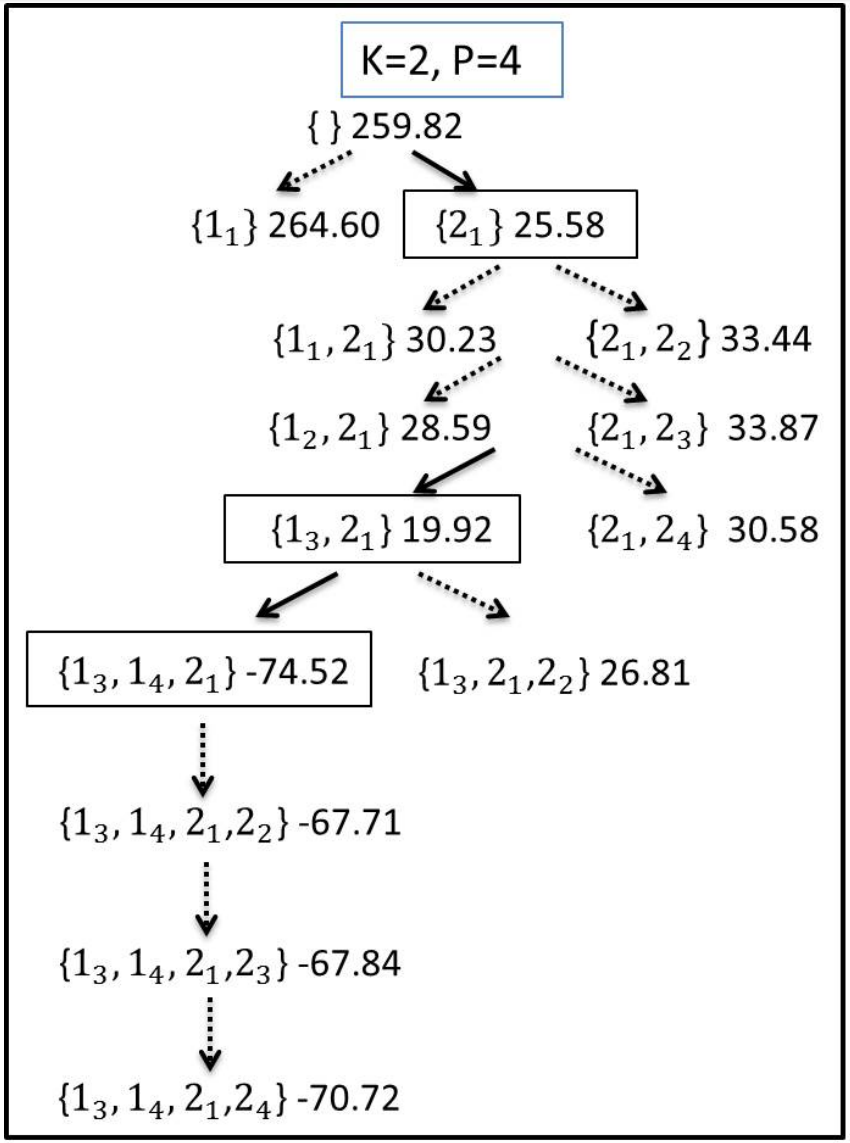
• If there are no components of X in $\mathbf{w}_y$, then RCGCI $= 0$, otherwise it is defined as for CGCI.

Example of mBTS

$$X_{1,t} = 0.4X_{1,t-1} + u_{1,t}$$
$$X_{2,t} = -0.3X_{1,t-4} + 0.4X_{2,t-1} + u_{2,t}$$
$$X_1 \rightarrow X_2$$

Progressive selection of lagged variables with mBTS:

1. $\mathbf{w}_y = [X_{2,t-1}]$
2. $\mathbf{w}_y = [X_{2,t-1}, X_{1,t-3}]$
3. $\mathbf{w}_y = [X_{2,t-1}, X_{1,t-3}, X_{1,t-4}]$



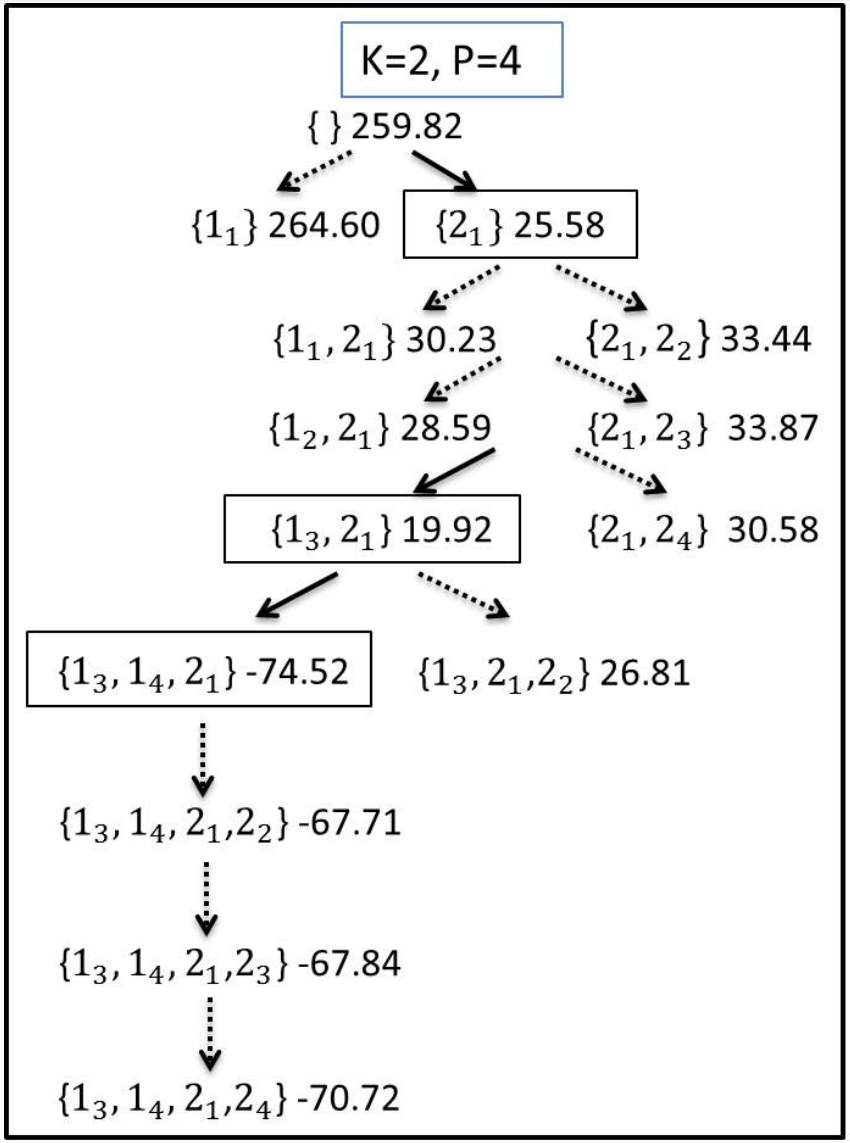
Example of mBTS

$$X_{1,t} = 0.4X_{1,t-1} + u_{1,t}$$
$$X_{2,t} = -0.3X_{1,t-4} + 0.4X_{2,t-1} + u_{2,t}$$

$X_1 \rightarrow X_2$

Progressive selection of lagged variables with mBTS:

1. $\mathbf{w}_y = [X_{2,t-1}]$
2. $\mathbf{w}_y = [X_{2,t-1}, X_{1,t-3}]$
3. $\mathbf{w}_y = [X_{2,t-1}, X_{1,t-3}, X_{1,t-4}]$



Unrestricted model for X_2

$$X_{2,t} = a_{1,1}X_{1,t-3} + a_{1,2}X_{1,t-4} + a_{2,1}X_{2,t-1} + e_{U,t}$$

Restricted model for X_2 (without X_1)

$$X_{2,t} = a_{2,1}X_{2,t-1} + e_{R,t}$$

Mutual Information from Mixed Embedding - 1

MIME applies dimension reduction and then uses conditional mutual information. The idea: [Vlachos & Kugiumtzis, PRE, 2010]

Mutual Information from Mixed Embedding - 1

MIME applies dimension reduction and then uses conditional mutual information. The idea: [Vlachos & Kugiumtzis, PRE, 2010]

1. Find a subset $\mathbf{w}_t$ of lagged variables from X and Y that explains best the future of Y, y_{t+1}.

Mutual Information from Mixed Embedding - 1

MIME applies dimension reduction and then uses conditional mutual information. The idea: [Vlachos & Kugiumtzis, PRE, 2010]

1. Find a subset $\mathbf{w}_t$ of lagged variables from X and Y that explains best the future of Y, y_{t+1}.
2. Quantify the information on Y ahead that is explained by the X-components in this subset.

Mutual Information from Mixed Embedding - 1

MIME applies dimension reduction and then uses conditional mutual information. The idea: [Vlachos & Kugiumtzis, PRE, 2010]

1. Find a subset $\mathbf{w}_t$ of lagged variables from X and Y that explains best the future of Y, y_{t+1}.
2. Quantify the information on Y ahead that is explained by the X-components in this subset.

If there are no components of X in $\mathbf{w}_t$, then MIME $= 0$.

Mutual Information from Mixed Embedding - 2

$$\mathbf{w}_t = (\underbrace{x_{t-\tau_{x1}}, x_{t-\tau_{x2}}, \ldots, x_{t-\tau_{xm_x}}}_{\mathbf{w}_t^x}, \underbrace{y_{t-\tau_{y1}}, y_{t-\tau_{y2}}, \ldots, y_{t-\tau_{ym_y}}}_{\mathbf{w}_t^y})$$

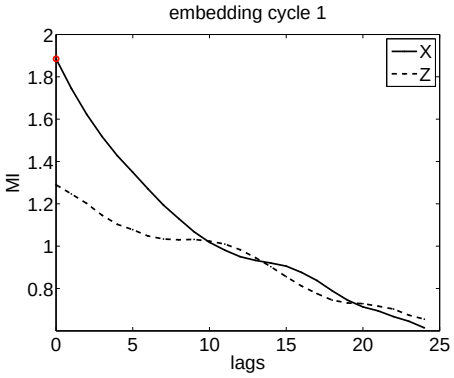
The causality measure MIME

$$R_{X\to Y} = \frac{I(y_{t+1}; \mathbf{w}_t^x \mid \mathbf{w}_t^y)}{I(\mathbf{y}_t^T; \mathbf{w}_t)}$$

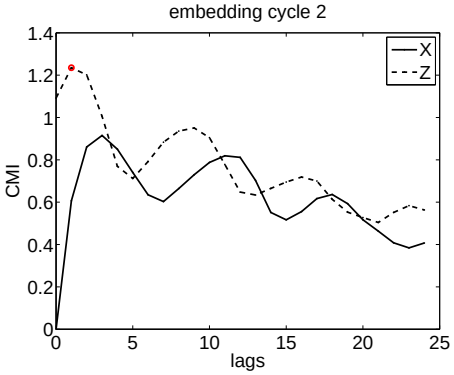
- $R_{X\to Y}$: information of Y explained only by X-components of the embedding vectors, normalized against the total mutual information (in order to give a value between 0 and 1).
- If $\mathbf{w}_t$ contains no components from X, then $R_{X\to Y} = 0$ and X has no effect on the future of Y.

Example: Embedding from X and Z variables of the chaotic Lorenz system to explain X, $W_t = \{x_t, \ldots, x_{t-24}, z_t, \ldots, z_{t-24}\}$ $\mathbf{y}_t^T = (x_{t+1}, \ldots, x_{t+5})$, $N = 10000$, sampling time $\tau_s = 0.05$

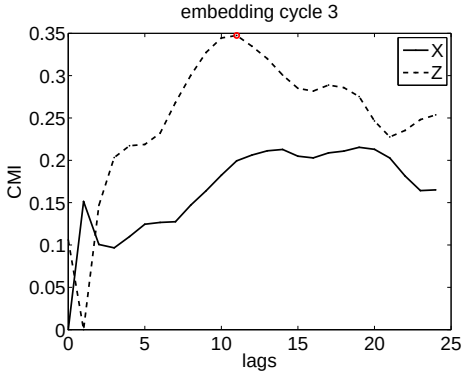
$x_t = \arg\max \left\{ I\left(\mathbf{y}_t^T; w_t\right) \right\}$, $w_t \in W_t$



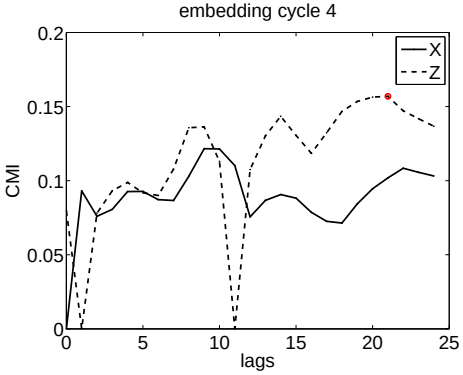
$z_{t-1} = \arg\max \left\{ I\left(\mathbf{y}_t^T; w_t \mid x_t\right) \right\}$

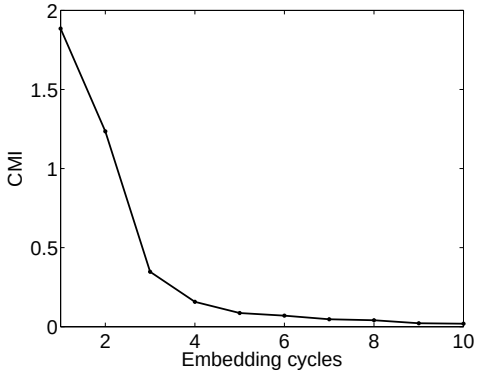


$z_{t-11} = \arg\max \left\{ I\left(\mathbf{y}_t^T; w_t \mid x_t, z_{t-1}\right) \right\}$



Too small increase in CMI





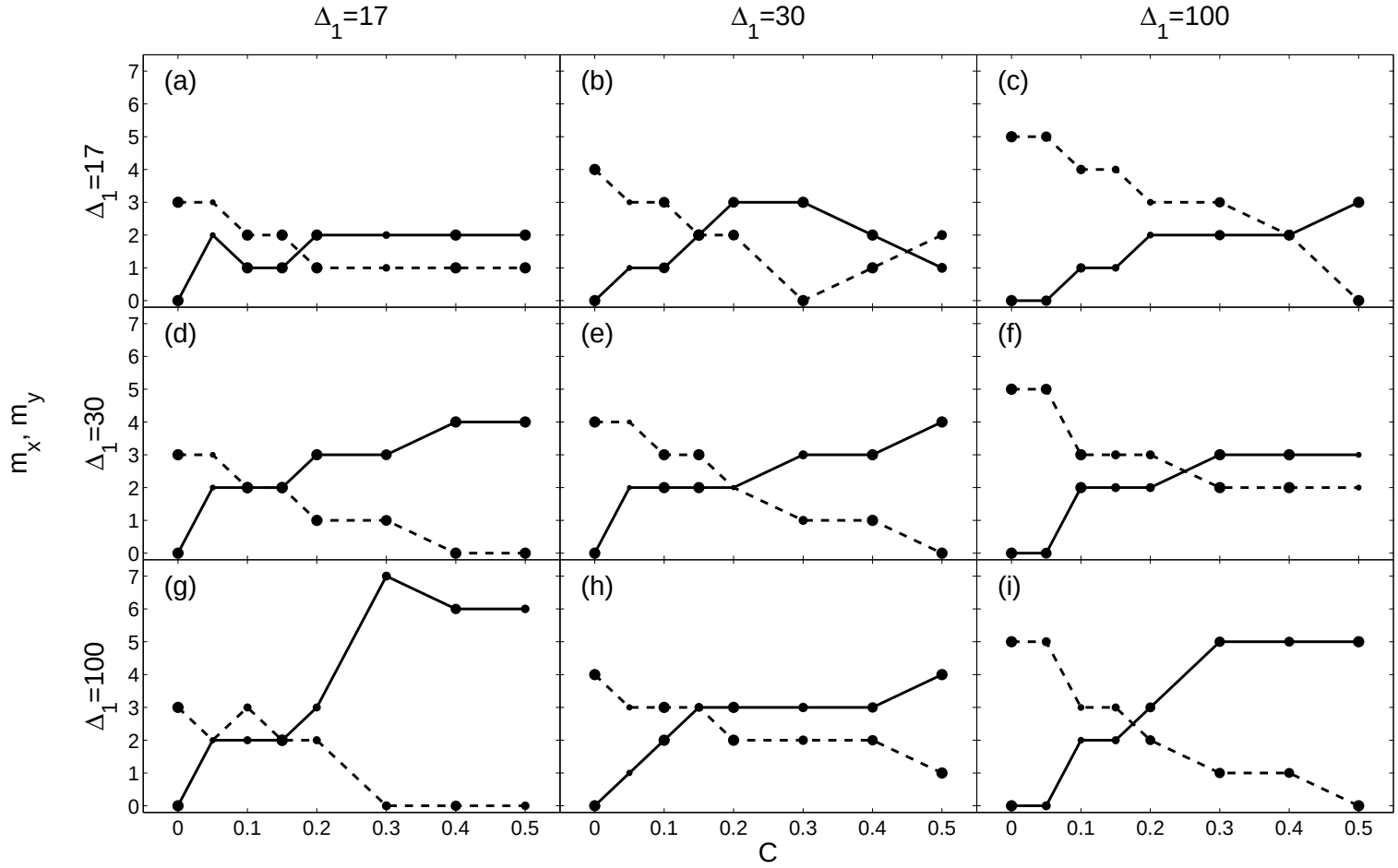
Embedding vector:

$\mathbf{w}_t = (x_t, z_{t-1}, z_{t-11})$

Example: Coupled Mackey-Glass system

$\Delta = 17, 30, 100, \qquad N = 4096$

$\mathbf{y}_t^T = \{y_{t+1}, y_{t+\tau_1}, y_{t+\tau_2}\}, \qquad L_x = L_y = 50$



solid line: driving system dashed line: response system

Partial Mutual Information from Mixed Embedding - 1

driving system: X, response system: Y,
conditioning on system Z, $Z = \{Z_1, Z_2, \ldots, Z_{K-2}\}$

The same non-uniform embedding scheme for explaining $\mathbf{y}_t^T$ from vector of lags of all $X, Y, Z_1, Z_2, \ldots, Z_{K-2}$,

$W_t =$
$\{x_t, \ldots, x_{t-L_x-1}, y_t, \ldots, y_{t-L_y-1}, z_{1,t}, \ldots, z_{1,t-L_z-1}, \ldots, z_{K-2,t-L_z-1}\}$
e.g., for $K = 3$, X, Y, Z:

$$\mathbf{w}_t = (\underbrace{x_{t-\tau_{x1}}, \ldots, x_{t-\tau_{xm_x}}}_{\mathbf{w}_t^x}, \underbrace{y_{t-\tau_{y1}}, \ldots, y_{t-\tau_{ym_y}}}_{\mathbf{w}_t^y}, \underbrace{z_{t-\tau_{z1}}, \ldots, z_{t-\tau_{zm_z}}}_{\mathbf{w}_t^z})$$

Partial Mutual Information from Mixed Embedding - 2

The non-uniform embedding vector of lags of all X, Y, Z for explaining y_{t+1}:

$$\mathbf{w}_t = (\underbrace{x_{t-\tau_{x1}}, \ldots, x_{t-\tau_{xm_x}}}_{\mathbf{w}_t^x}, \underbrace{y_{t-\tau_{y1}}, \ldots, y_{t-\tau_{ym_y}}}_{\mathbf{w}_t^y}, \underbrace{z_{t-\tau_{z1}}, \ldots, z_{t-\tau_{zm_z}}}_{\mathbf{w}_t^z})$$

Partial Mutual Information from Mixed Embedding - 2

The non-uniform embedding vector of lags of all X, Y, Z for explaining y_{t+1}:

$$\mathbf{w}_t = (\underbrace{x_{t-\tau_{x1}}, \ldots, x_{t-\tau_{xm_x}}}_{\mathbf{w}_t^x}, \underbrace{y_{t-\tau_{y1}}, \ldots, y_{t-\tau_{ym_y}}}_{\mathbf{w}_t^y}, \underbrace{z_{t-\tau_{z1}}, \ldots, z_{t-\tau_{zm_z}}}_{\mathbf{w}_t^z})$$

The causality measure PMIME

$$R_{X \to Y|Z} = \frac{I(y_{t+1}; \mathbf{w}_t^x \mid \mathbf{w}_t^y, \mathbf{w}_t^Z)}{I(y_{t+1}; \mathbf{w}_t)}$$

- $R_{X \to Y|Z}$: information on the future of Y explained only by X-components of the embedding vector (given the components of Y and Z), normalized with the mutual information of the future of Y and the embedding vector.

Partial Mutual Information from Mixed Embedding - 2

The non-uniform embedding vector of lags of all X, Y, Z for explaining y_{t+1}:

$$\mathbf{w}_t = (\underbrace{x_{t-\tau_{x1}}, \ldots, x_{t-\tau_{xm_x}}}_{\mathbf{w}_t^x}, \underbrace{y_{t-\tau_{y1}}, \ldots, y_{t-\tau_{ym_y}}}_{\mathbf{w}_t^y}, \underbrace{z_{t-\tau_{z1}}, \ldots, z_{t-\tau_{zm_z}}}_{\mathbf{w}_t^z})$$

The causality measure PMIME

$$R_{X \to Y|Z} = \frac{I(y_{t+1}; \mathbf{w}_t^x \mid \mathbf{w}_t^y, \mathbf{w}_t^Z)}{I(y_{t+1}; \mathbf{w}_t)}$$

- $R_{X \to Y|Z}$: information on the future of Y explained only by X-components of the embedding vector (given the components of Y and Z), normalized with the mutual information of the future of Y and the embedding vector.
- If $\mathbf{w}_t^Z = \emptyset$, then $R_{X \to Y|Z} = R_{X \to Y}$.

Partial Mutual Information from Mixed Embedding - 2

The non-uniform embedding vector of lags of all X, Y, Z for explaining y_{t+1}:

$$\mathbf{w}_t = (\underbrace{x_{t-\tau_{x1}}, \ldots, x_{t-\tau_{xm_x}}}_{\mathbf{w}_t^x}, \underbrace{y_{t-\tau_{y1}}, \ldots, y_{t-\tau_{ym_y}}}_{\mathbf{w}_t^y}, \underbrace{z_{t-\tau_{z1}}, \ldots, z_{t-\tau_{zm_z}}}_{\mathbf{w}_t^z})$$

The causality measure PMIME

$$R_{X \to Y|Z} = \frac{I(y_{t+1}; \mathbf{w}_t^x \mid \mathbf{w}_t^y, \mathbf{w}_t^Z)}{I(y_{t+1}; \mathbf{w}_t)}$$

- $R_{X \to Y|Z}$: information on the future of Y explained only by X-components of the embedding vector (given the components of Y and Z), normalized with the mutual information of the future of Y and the embedding vector.
- If $\mathbf{w}_t^Z = \emptyset$, then $R_{X \to Y|Z} = R_{X \to Y}$.
- If $\mathbf{w}_t$ contains no components from X, then $R_{X \to Y|Z} = 0$ and X has no direct effect on the future of Y.

Partial Mutual Information from Mixed Embedding - 3

Three main advantages of PMIME

Partial Mutual Information from Mixed Embedding - 3

Three main advantages of PMIME

- $R_{X \to Y|Z} = 0$ when no significant causality is present, and $R_{X \to Y|Z} > 0$ when it is present
  [no significance test, no issues with multiple testing!]

Partial Mutual Information from Mixed Embedding - 3

Three main advantages of PMIME

- $R_{X \to Y|Z} = 0$ when no significant causality is present, and $R_{X \to Y|Z} > 0$ when it is present
  [no significance test, no issues with multiple testing!]
- mixed embedding for all variables is formed as part of the measure [it does not require the determination of embedding parameters for each variable]

Partial Mutual Information from Mixed Embedding - 3

Three main advantages of PMIME

- $R_{X \to Y|Z} = 0$ when no significant causality is present, and $R_{X \to Y|Z} > 0$ when it is present
  [no significance test, no issues with multiple testing!]
- mixed embedding for all variables is formed as part of the measure [it does not require the determination of embedding parameters for each variable]
- inclusion of more confounding variables only slows the computation and has no effect on statistical accuracy
  [**no "curse of dimensionality"** for any dimension of Z, only slow computation time]

Partial Mutual Information from Mixed Embedding - 3

Three main advantages of PMIME

- $R_{X \to Y|Z} = 0$ when no significant causality is present, and $R_{X \to Y|Z} > 0$ when it is present
  [no significance test, no issues with multiple testing!]
- mixed embedding for all variables is formed as part of the measure [it does not require the determination of embedding parameters for each variable]
- inclusion of more confounding variables only slows the computation and has no effect on statistical accuracy
  [**no "curse of dimensionality"** for any dimension of Z, only slow computation time]

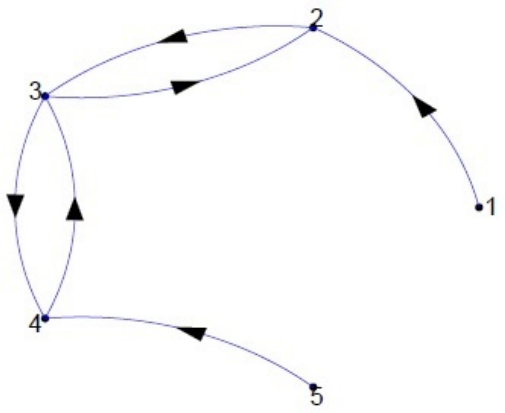
$\Rightarrow$ good candidate for causality analysis with many variables

Example: coupled Mackey-Glass

Coupled identical Mackey-Glass delayed differential equations

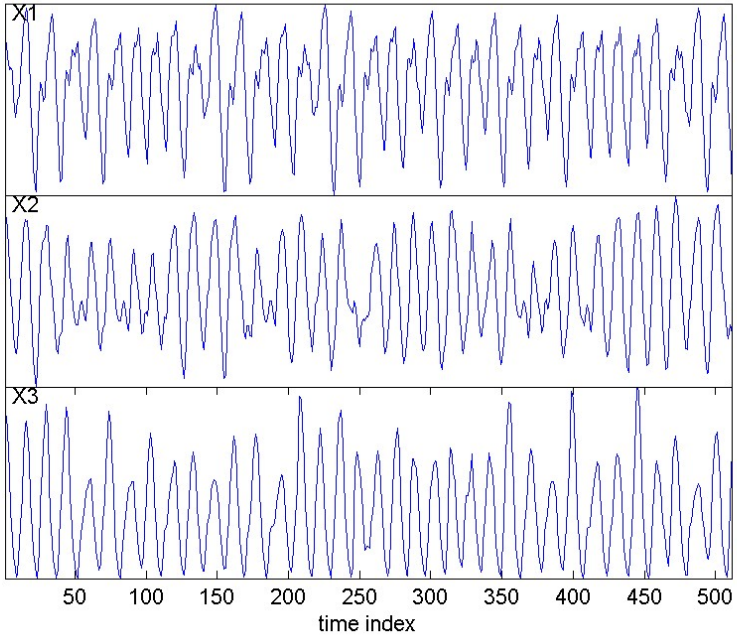
$$\dot{x}_i(t) = -0.1x_i(t) + \sum_{j=1}^{K} \frac{C_{ij}x_j(t-\Delta)}{1+x_j(t-\Delta)^{10}} \quad \text{for} \quad i = 1, \ldots, K$$

$K = 5$



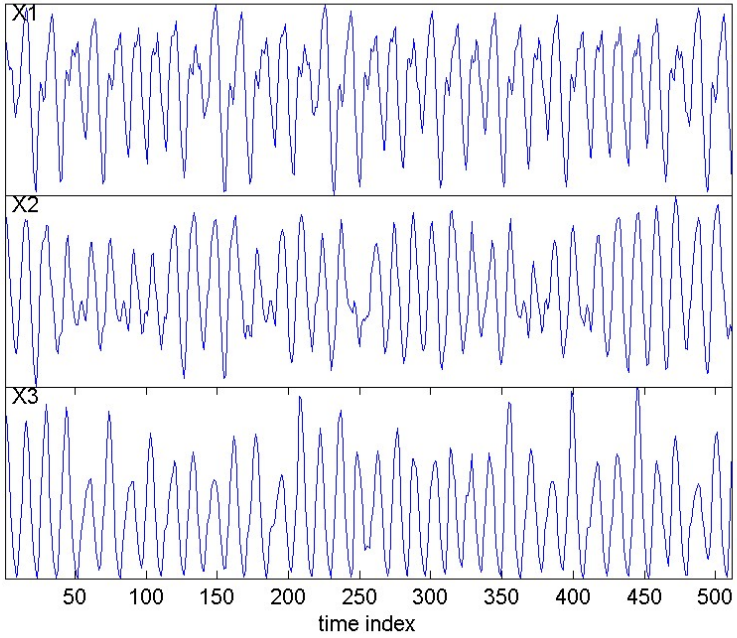
Mackey-Glass, $C = 0.2$

$\Delta = 20$

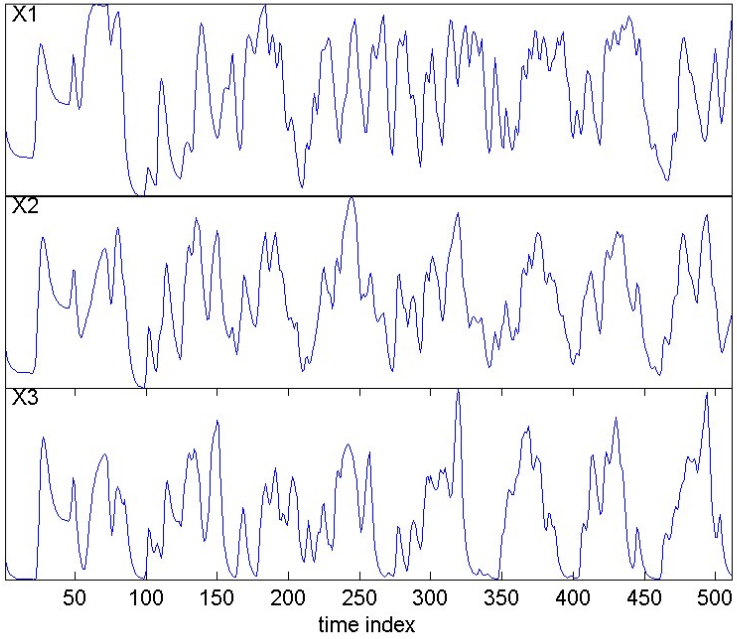


Mackey-Glass, $C = 0.2$

$\Delta = 20$

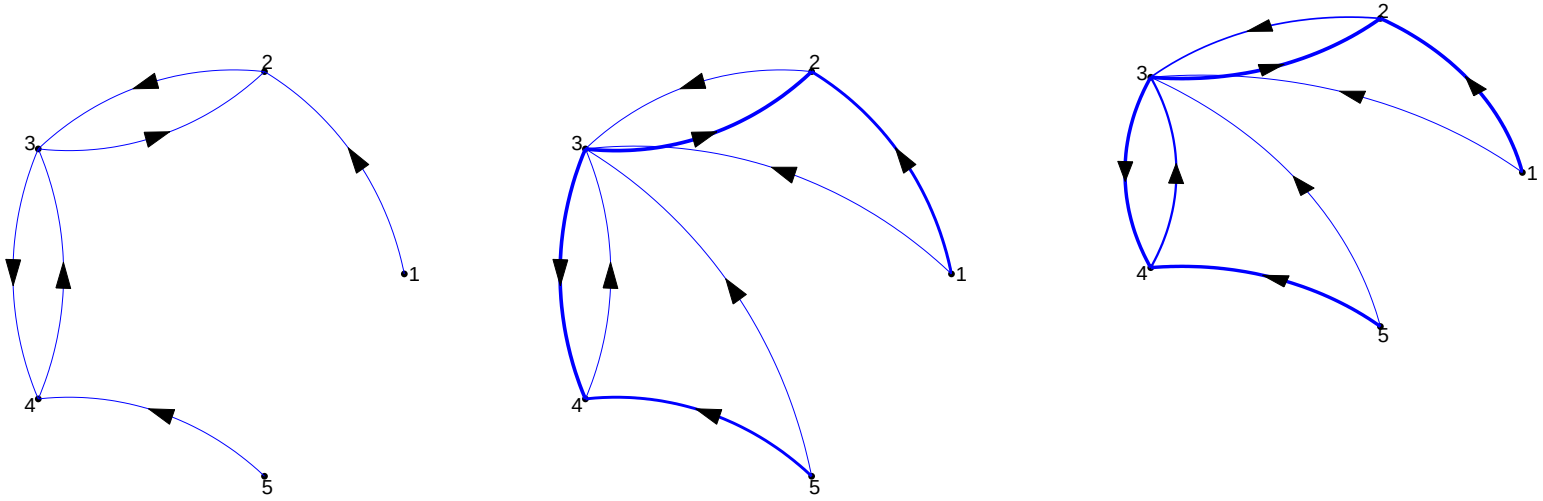


$\Delta = 100$



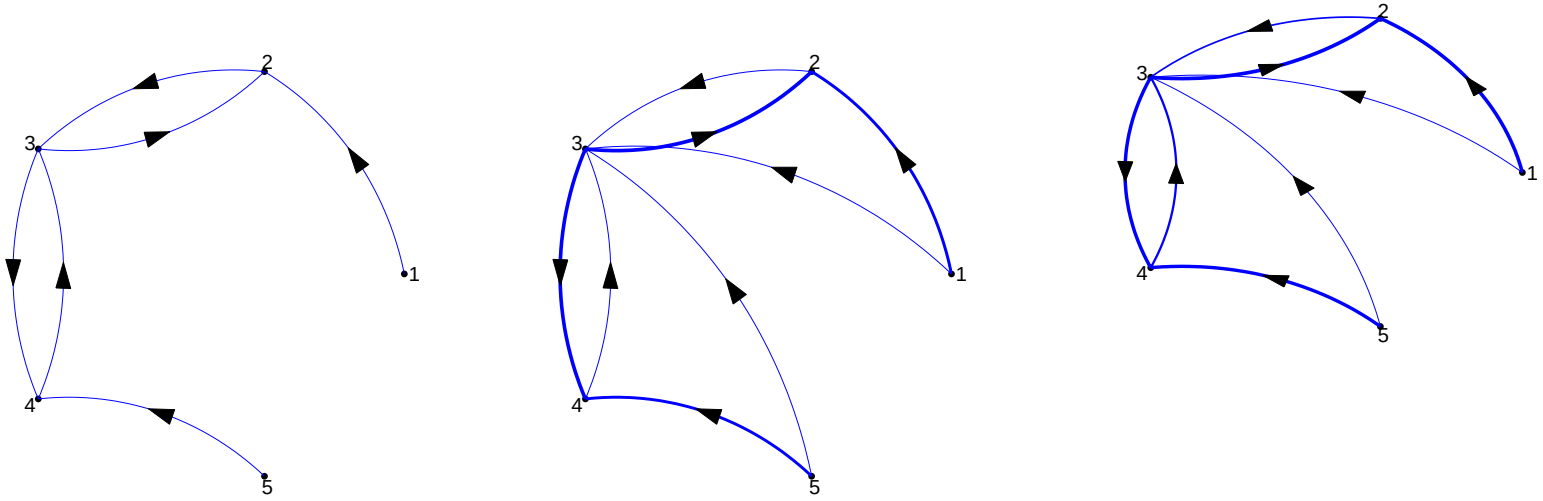
Mackey-Glass: true/estimated network [Kugiumtzis and Kimiskidis, IJNS 2015]

$K = 5$ True from PMIME ($\Delta = 20$) from PMIME ($\Delta = 100$)

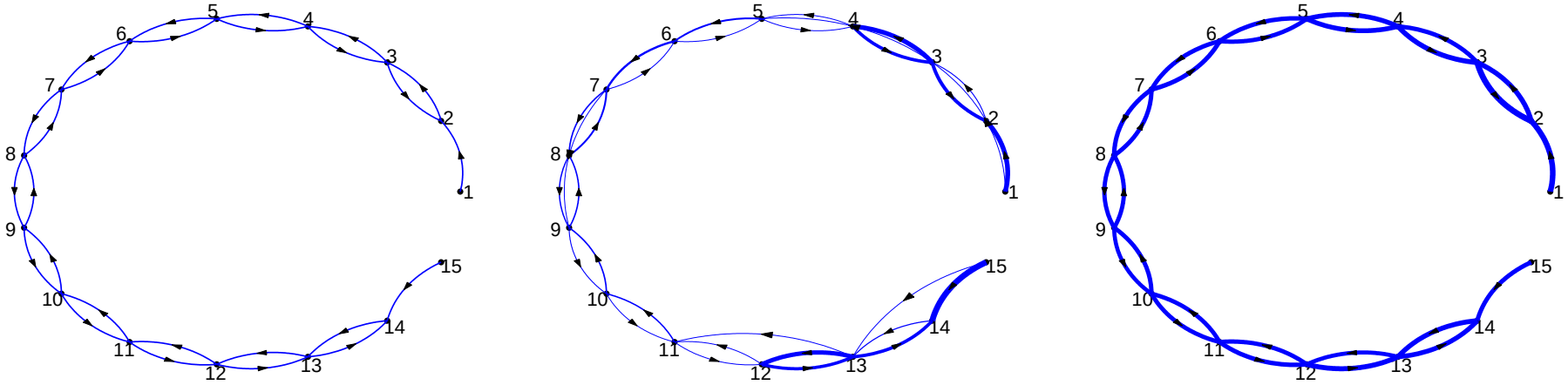


Mackey-Glass: true/estimated network [Kugiumtzis and Kimiskidis, IJNS 2015]

$K = 5$ True from PMIME ($\Delta = 20$) from PMIME ($\Delta = 100$)

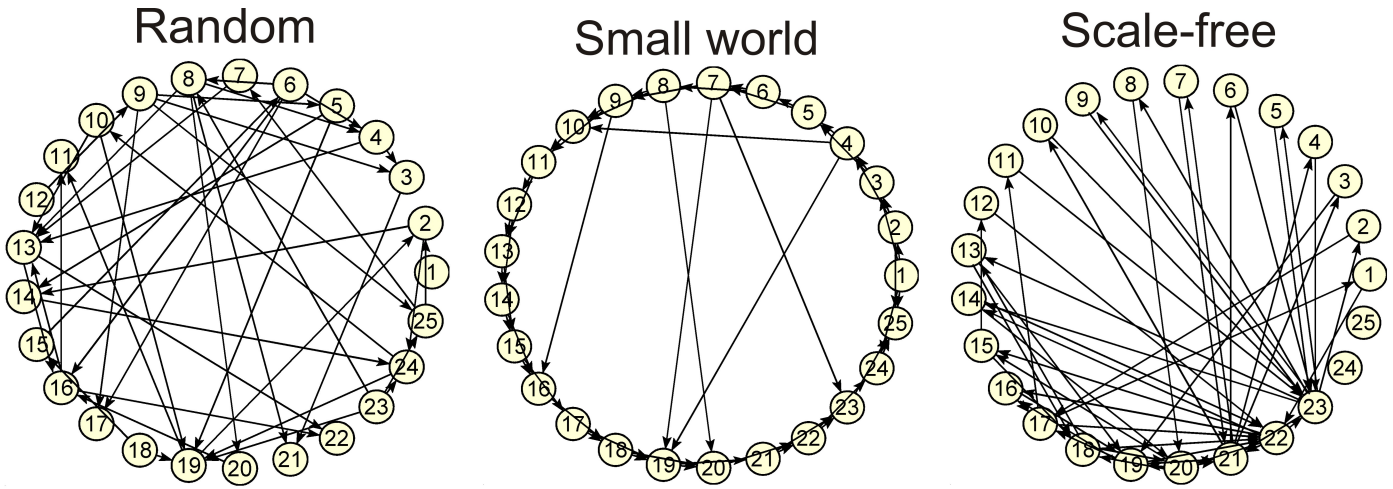


$K = 15$ True from PMIME ($\Delta = 20$) from PMIME ($\Delta = 100$)



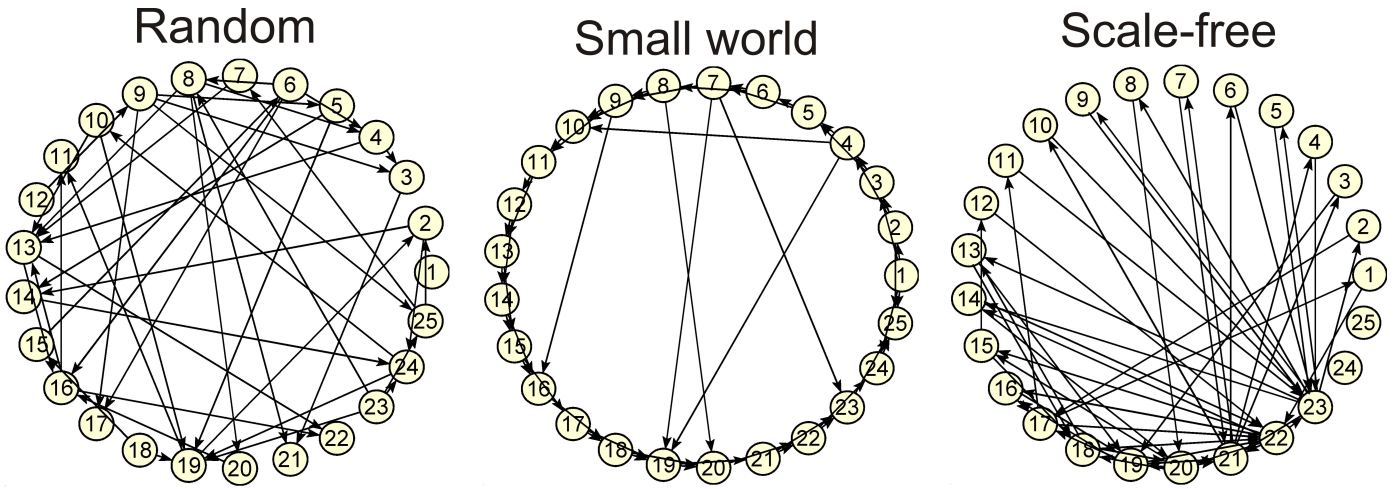
Can different network structures be detected?

Simulation: three types of networks for the generating system



Can different network structures be detected?

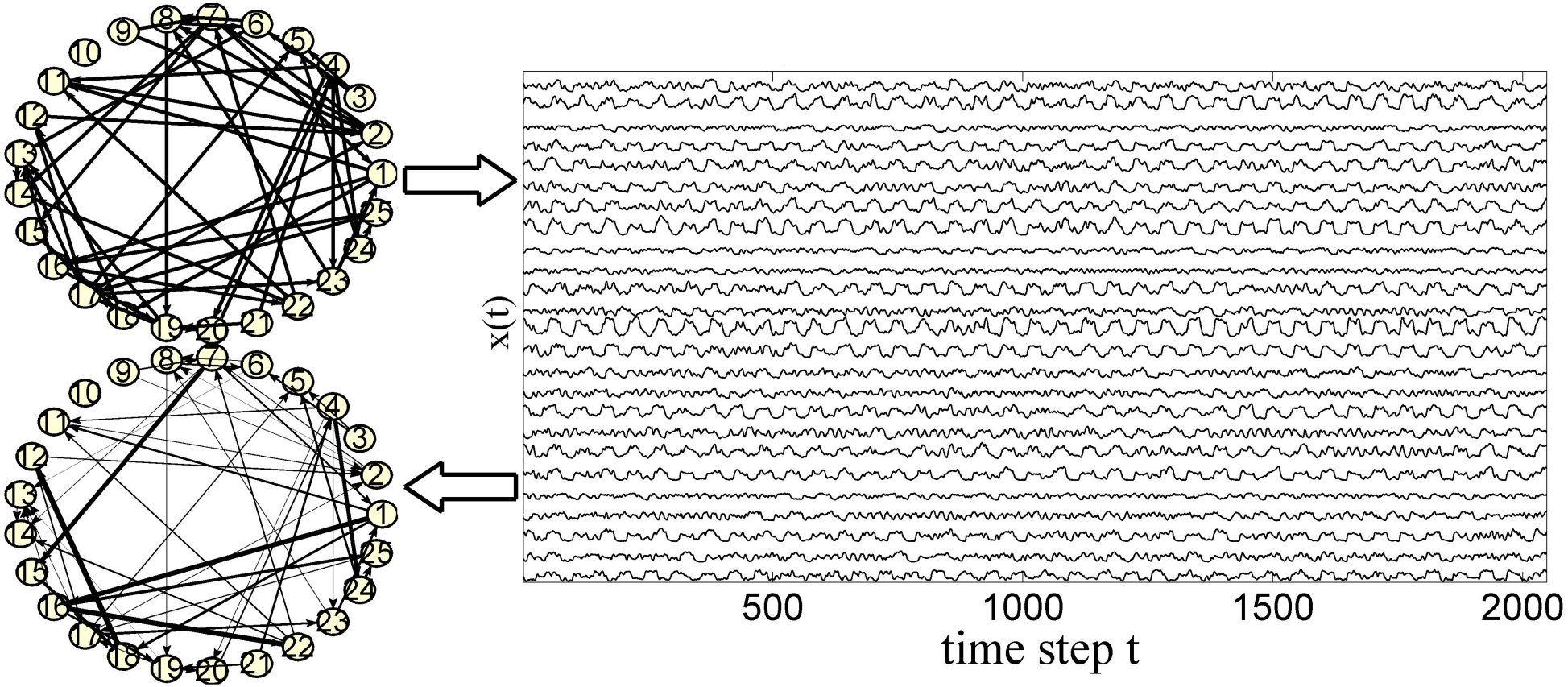
Simulation: three types of networks for the generating system



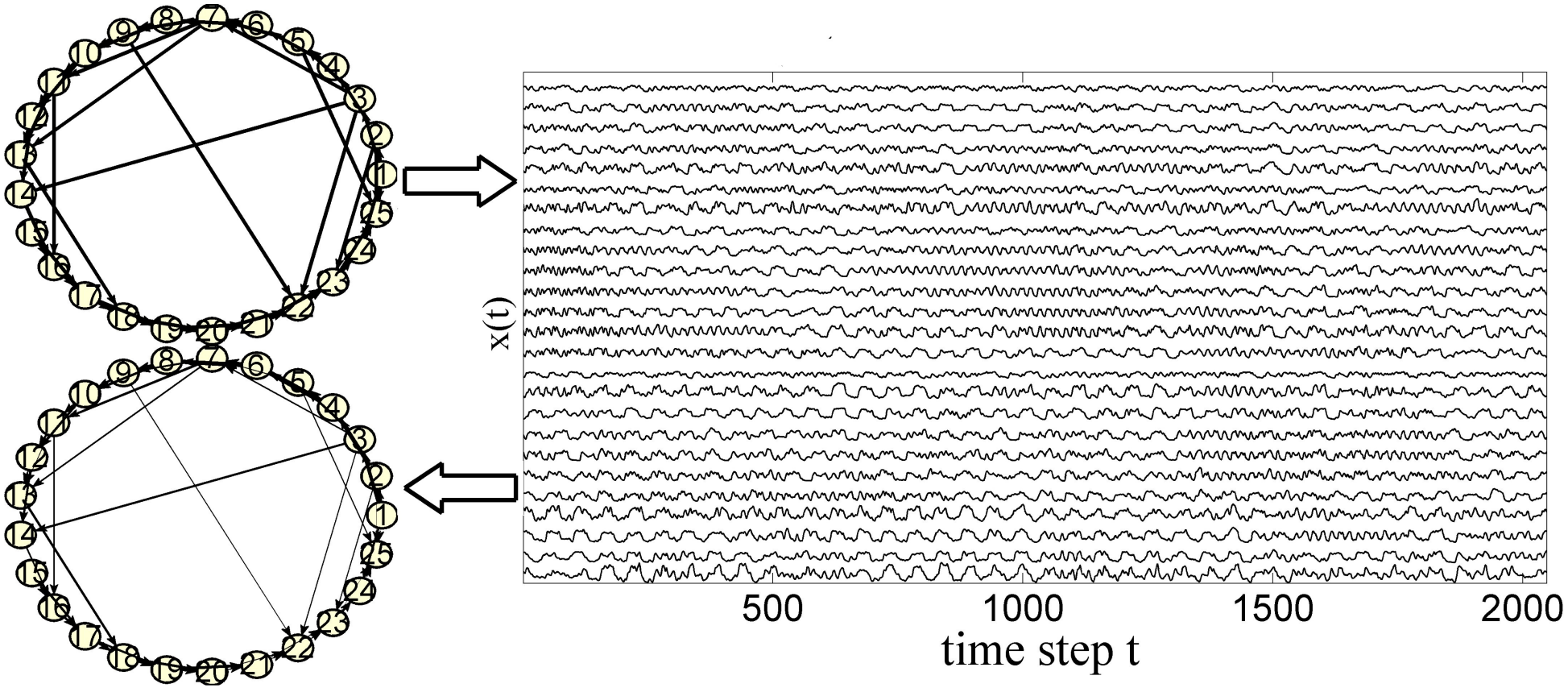
Generating system:
coupled Mackey-Glass system, $K = 25$, $\Delta = 100$, $C = 0.2$
with coupling structure defined by the network type

Causality measure: PMIME

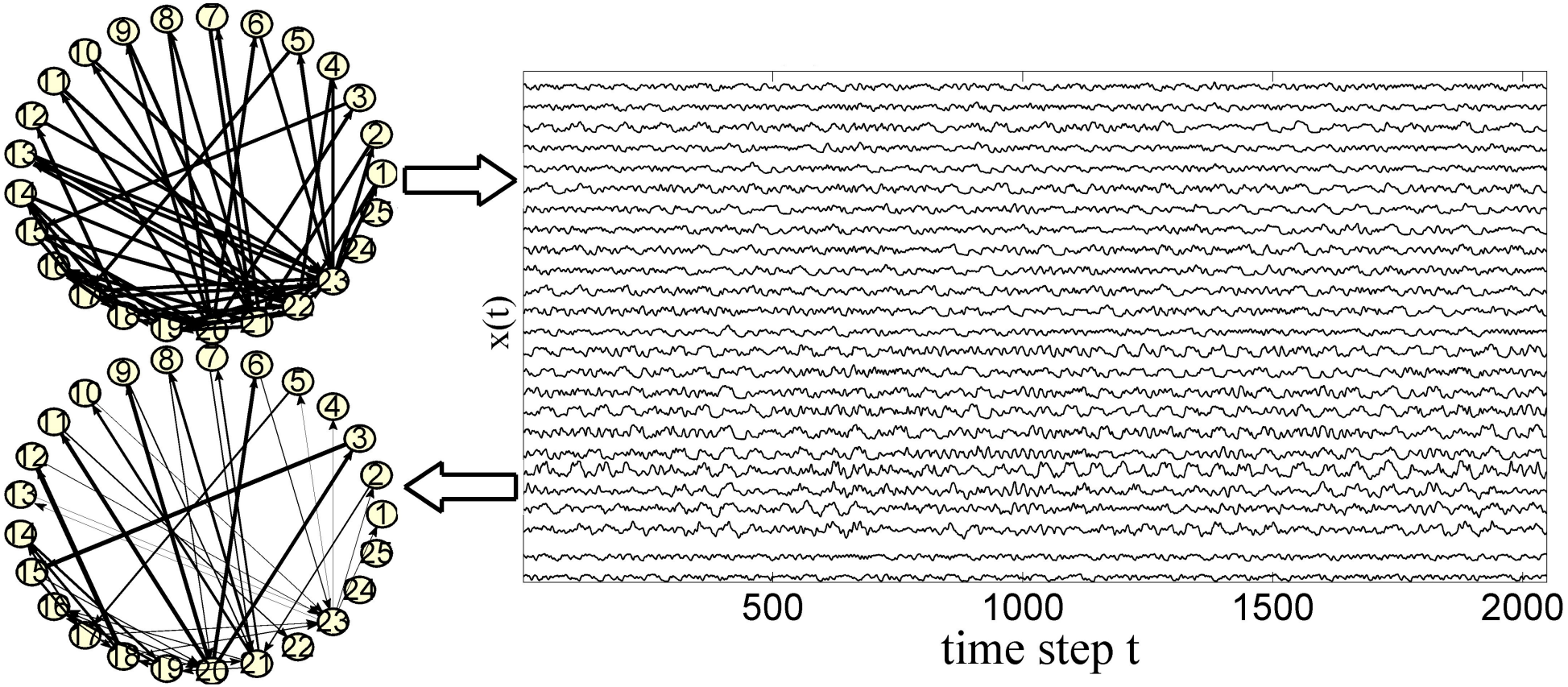
Estimation of the Random Network



Estimation of the Small-World Network



Estimation of the Scale-Free Network



Structural Change

Simulation example:

The network structure undergoes structural change at specific time points:
Random $\Rightarrow$ Small-World $\Rightarrow$ Scale-Free

Structural Change

Simulation example:

The network structure undergoes structural change at specific time points:
Random $\Rightarrow$ Small-World $\Rightarrow$ Scale-Free

Estimation of networks with PMIME at sliding windows

Structural Change

Simulation example:

The network structure undergoes structural change at specific time points:
Random $\Rightarrow$ Small-World $\Rightarrow$ Scale-Free

Estimation of networks with PMIME at sliding windows

Estimation of network characteristics on the PMIME networks

Structural Change

Simulation example:

The network structure undergoes structural change at specific time points:
Random $\Rightarrow$ Small-World $\Rightarrow$ Scale-Free

Estimation of networks with PMIME at sliding windows

Estimation of network characteristics on the PMIME networks

Structural change detection,
[Slow], **[Middle]**, **[Fast]**, **[Very fast]**

Summary

Summary

• Granger causality measures can capture the inter-dependence structure of a multivariate complex system / stochastic process.

Summary

• Granger causality measures can capture the inter-dependence structure of a multivariate complex system / stochastic process.

• Granger causality measures are good candidates to detect structural changes.

Summary

- Granger causality measures can capture the inter-dependence structure of a multivariate complex system / stochastic process.

- Granger causality measures are good candidates to detect structural changes.

- Many measures of causality:

Summary

• Granger causality measures can capture the inter-dependence structure of a multivariate complex system / stochastic process.

• Granger causality measures are good candidates to detect structural changes.

• Many measures of causality:

best: these that can capture also nonlinear and direct causal effects at the presence of many variables...

Summary

• Granger causality measures can capture the inter-dependence structure of a multivariate complex system / stochastic process.

• Granger causality measures are good candidates to detect structural changes.

• Many measures of causality:

best: these that can capture also nonlinear and direct causal effects at the presence of many variables... but practically hard to estimate reliably.

Summary

• Granger causality measures can capture the inter-dependence structure of a multivariate complex system / stochastic process.

• Granger causality measures are good candidates to detect structural changes.

• Many measures of causality:

best: these that can capture also nonlinear and direct causal effects at the presence of many variables... but practically hard to estimate reliably.

1. More advanced measures (nonlinear, direct effects) involve more (and depend more on) **free parameters**.
2. Harder to establish **statistical significance** of the measures when many variables are present (many nodes in the network). Correction for multiple testing requires many many surrogates.
3. Statistical accuracy of the direct causality measures decreases with the **number of confounding variables**.

References

Brandt PT & Williams JT (2007) *Multiple Time Series Models*, Sage Publications

Granger J (1969) "Investigating causal relations by econometric models and cross-spectral methods", *Acta Physica Polonica B*, 37: 424-438

Koutlis C & Kugiumtzis D (2016) "Discrimination of Coupling Structures using Causality Networks from Multivariate Time Series", submitted

Kraskov A, Stögbauer H & Grassberger P (2004) "Estimating Mutual Information", *Physical Review E*, 69(6): 066138

Kugiumtzis D (2013) "Direct Coupling Information Measure from Non-uniform Embedding", *Physical Review E*, 87: 062918

Kugiumtzis D & Kimiskidis VK (2015) "Direct causal networks for the study of transcranial magnetic stimulation effects on focal epileptiform discharges", *International Journal of Neural Systems*, 25: 1550006

Papana A, Kugiumtzis D & Larsson PG (2012) "Detection of direct causal effects and application in the analysis of electroencephalograms from patients with epilepsy", *International Journal of Bifurcation and Chaos*, 22(9): 1250222

Schreiber T (2000) "Measuring Information Transfer", *Physical Review Letters*, 85(2): 461-464

Siggiridou E, Koutlis C, Tsimpiris A, Kimiskidis VK & Kugiumtzis D (2015) "Causality networks from multivariate time series and application to epilepsy", in *Engineering in Medicine and Biology Society (EMBC), 37th Annual International Conference of the IEEE*, 4041-4044

Siggiridou E & Kugiumtzis D (2016) "Granger Causality in Multi-variate Time Series using a Time Ordered Restricted Vector Autoregressive Model", *IEEE Transactions on Signal Processing*, accepted, 2016

Vlachos I & Kugiumtzis D (2010) "Non-uniform state space reconstruction and coupling detection", *Physical Review E*, 82: 016207

Vlachos I & Kugiumtzis D (2013) "Backward-in-Time Selection of the Order of Dynamic Regression Prediction Model", *Journal of Forecasting*, 32(8): 685-701