

RECURRENCE QUANTIFICATION ANALYSIS

auto-RQA of categorical & continuous time series

Recurrence Quantification Analysis

STORY 1

“Jort en An vragen aan Jan of ze met de wandelwagen mogen rijden en het mag van Jan en ze gaan er in en ze rijden heel snel. Ze zien een boom en de wandelwagen gaat kapot. Ze komen weer bij. Jan maakt de wandelwagen weer”

MLU: 3.70

woorden: 47

STORY 2

“Papa zit in de bank en papa werkt in de tuin die maakt een kar de kinderen. Papa maakt een kar van de kinderen en de kinderen en de kinderen tegen de boom en de kar is kapot en de kinderen huilen en de kinderen zijn blij”

MLU: 3.68

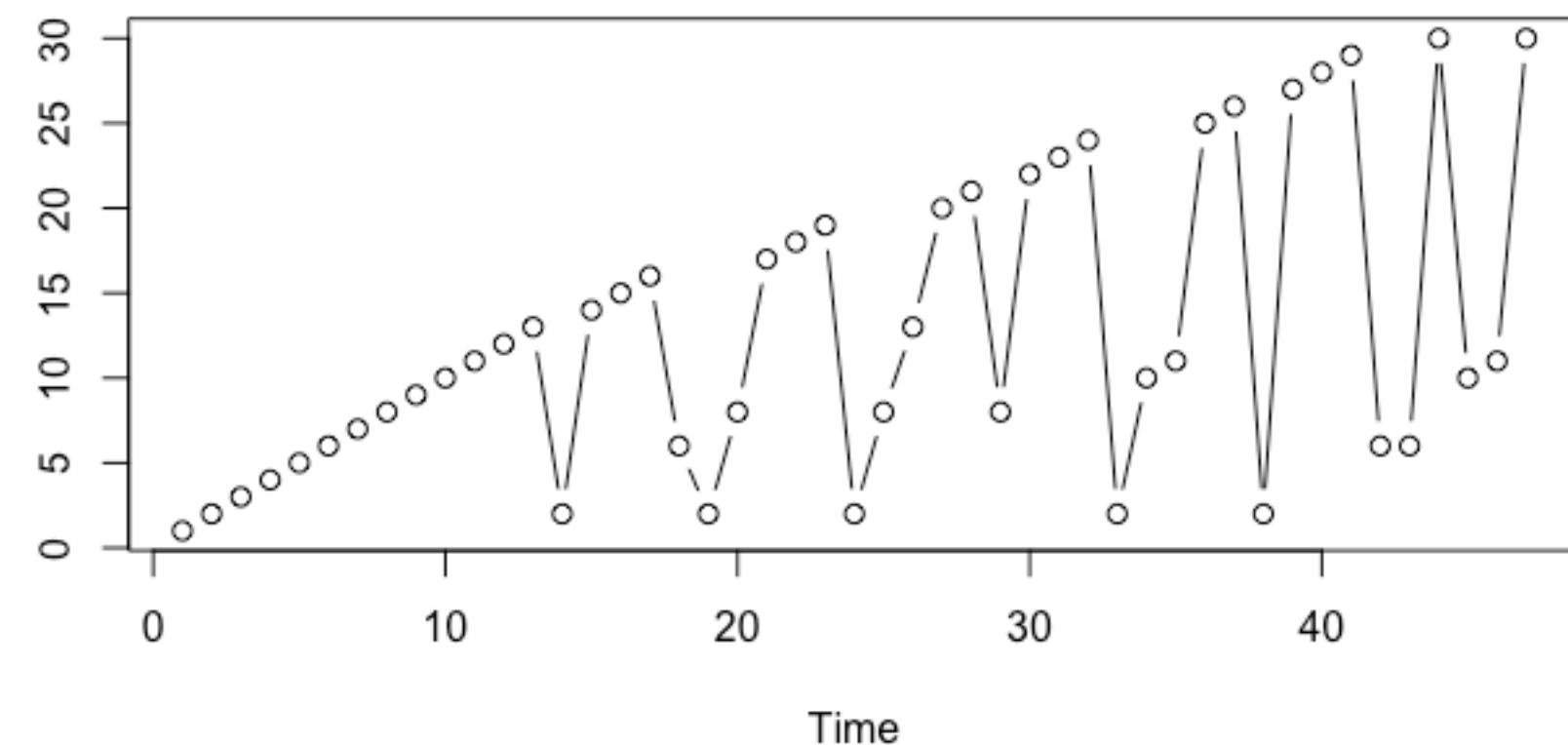
woorden: 47

Inter-rater reliability of “quality” is ok, but “why”?

Recurrence Quantification Analysis: Nominale Tijdseries

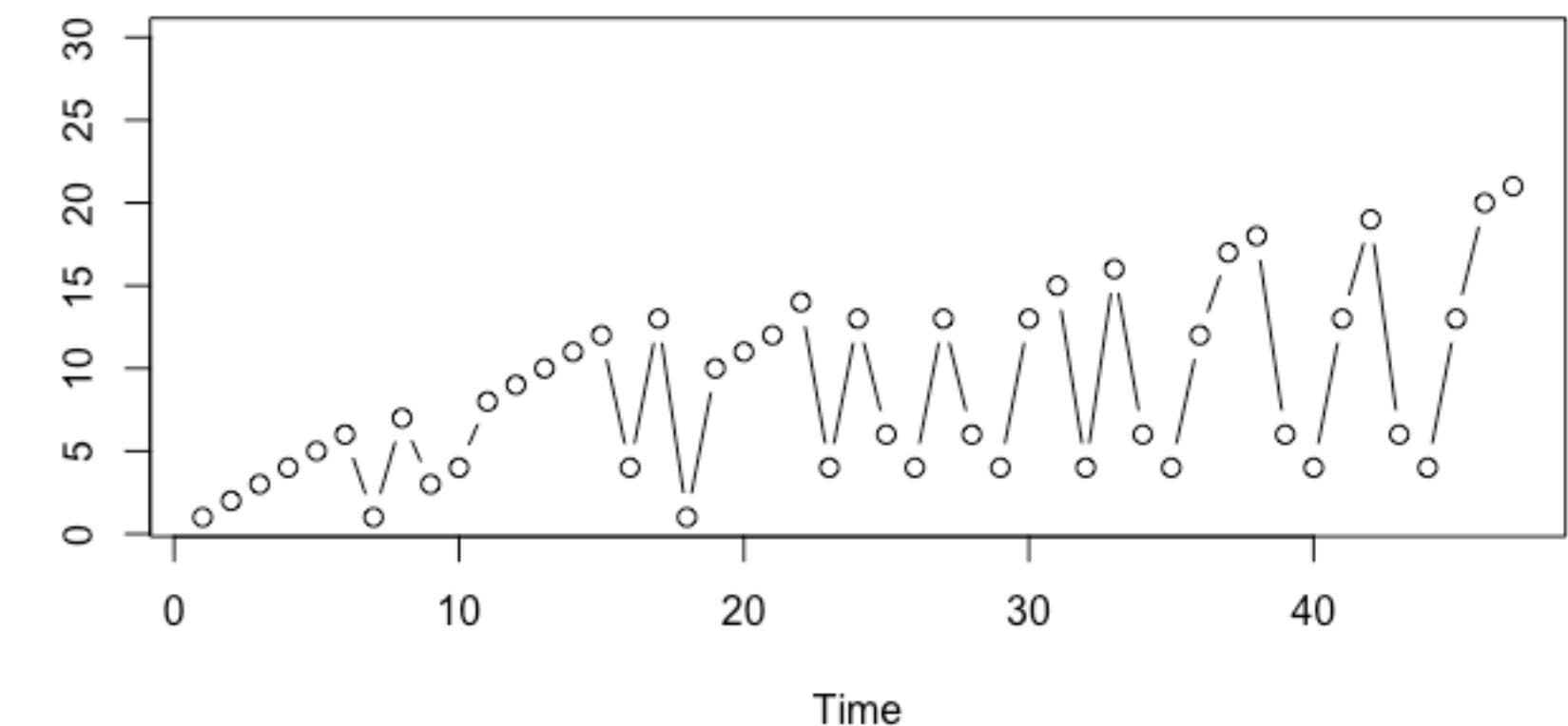
STORY 1

“1 2 3 4 5 6 7 8 9 10 11 12 13
2 14 15 16 6 2 8 17 18 19 2 8 13
20 21 8 22 23 24 2 10 11 25 26 2
27 28 29 6 6 30 10 11 30”

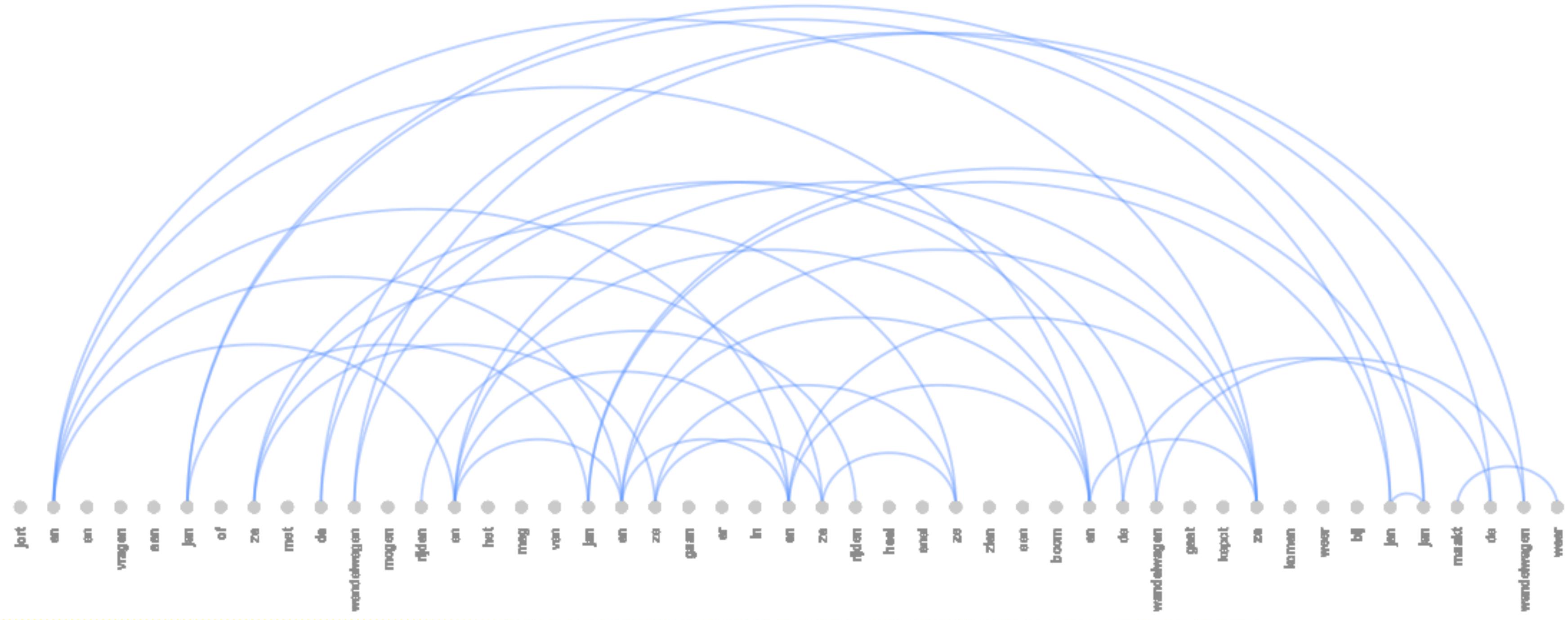


STORY 2

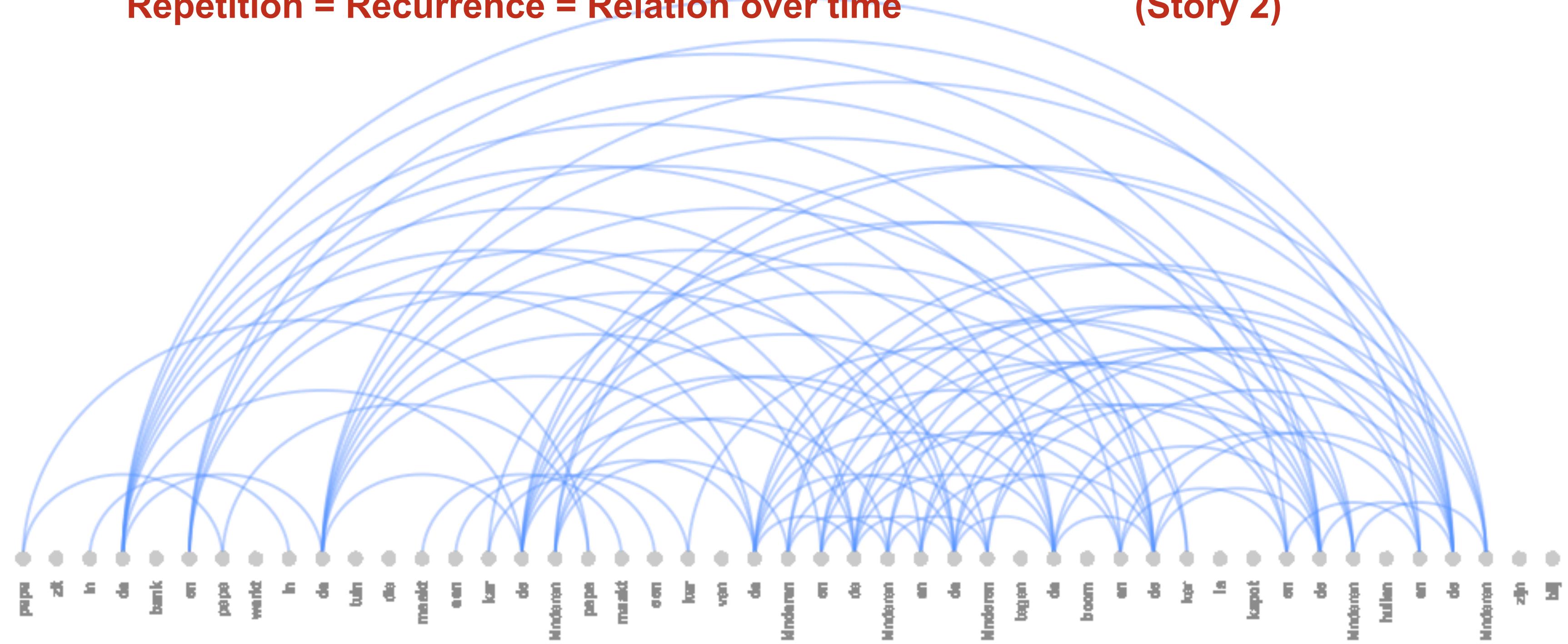
“1 2 3 4 5 6 1 7 3 4 8 9 10
11 12 4 13 1 10 11 12 14 4 13 6
4 13 6 4 13 15 4 16 6 4 12 17 18
6 4 13 19 6 4 13 20 21”



Repetition = Recurrence = Relation over time (Story 1)



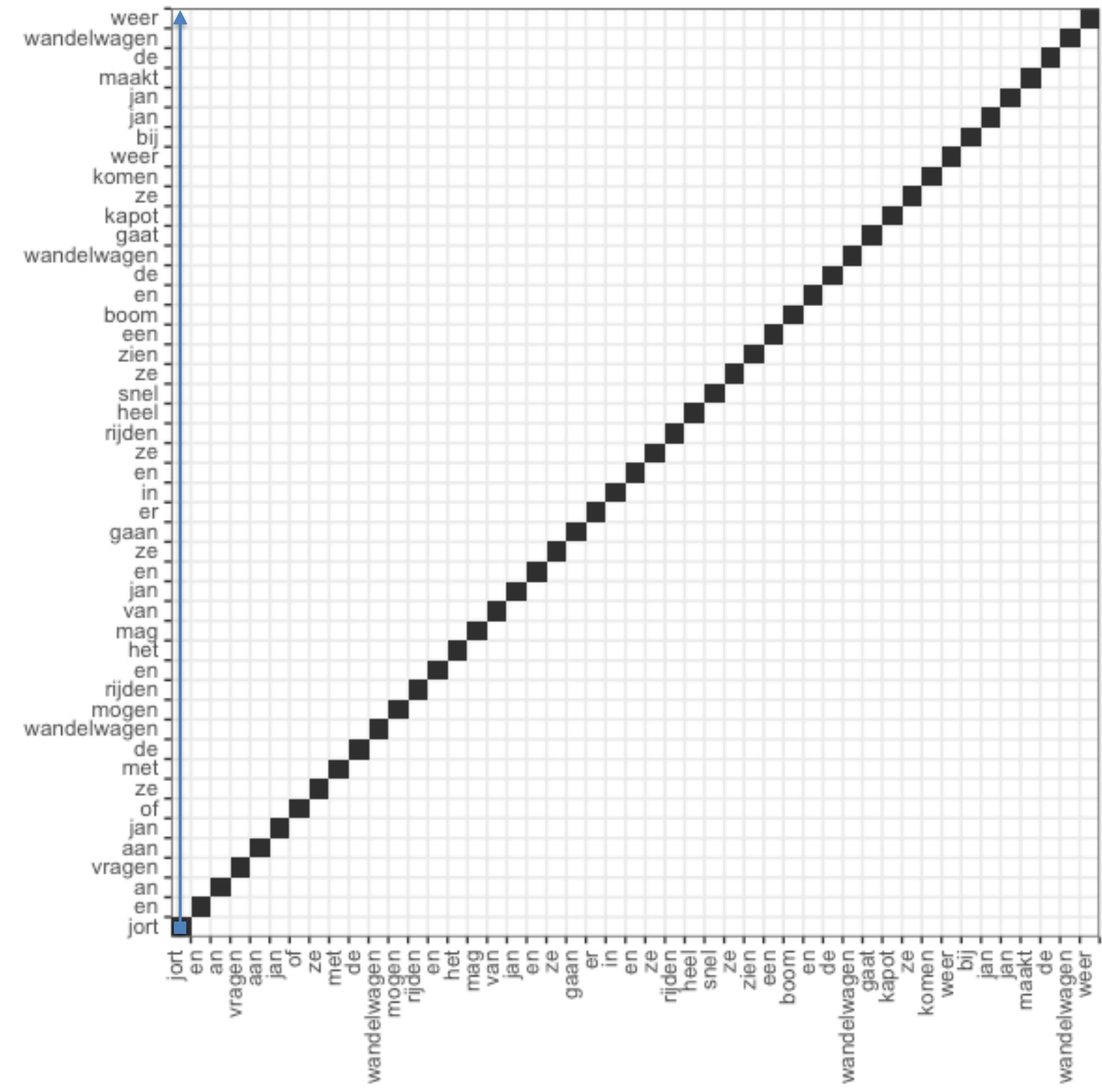
Repetition = Recurrence = Relation over time



Recurrence Plot

Place a dot when a word is recurring

'jort' (0 keer)

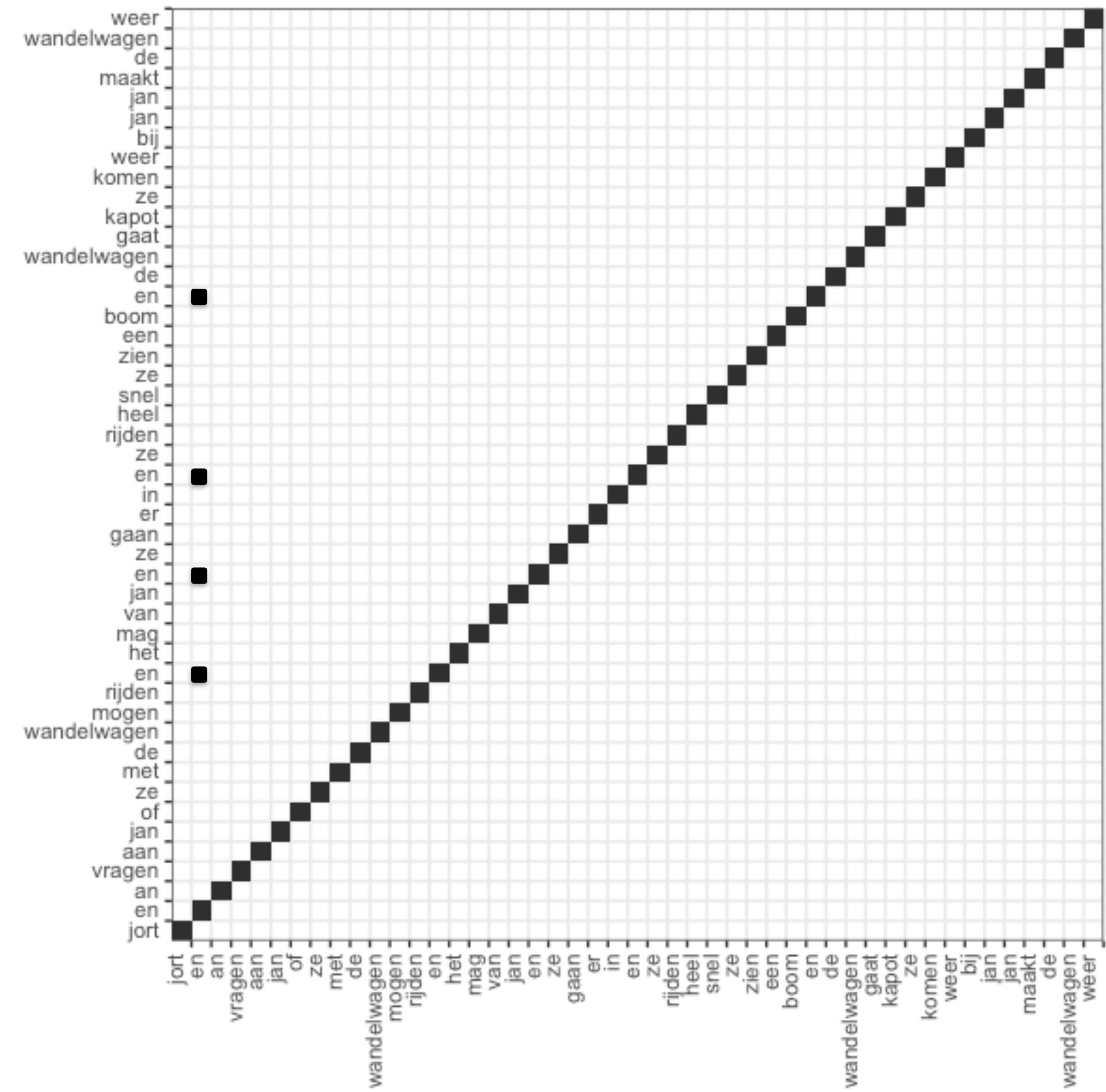


Recurrence Plot

Place a dot when a word is recurring

'jort' (0 keer)

'en' (4 keer)



Recurrence Plot

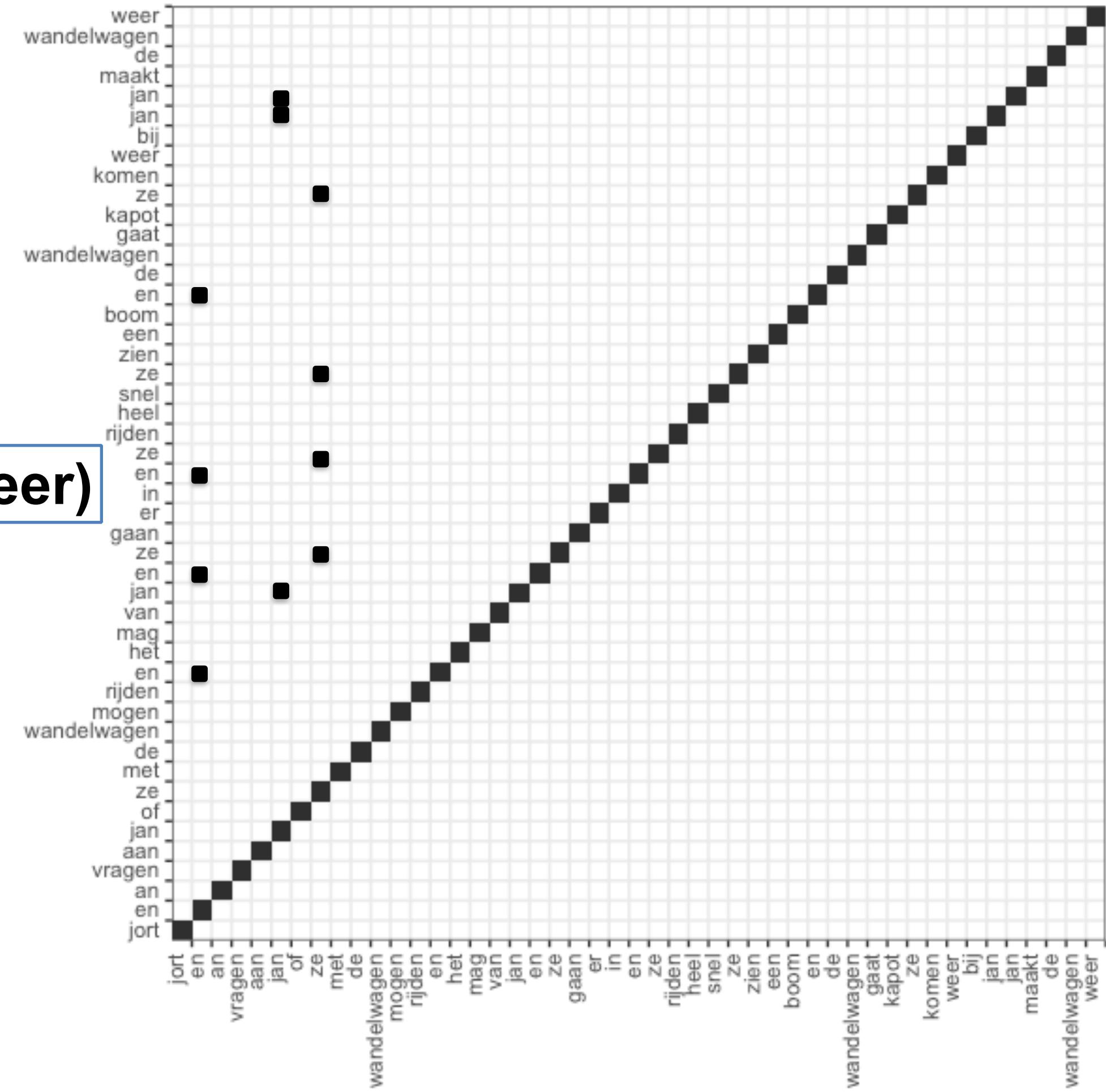
Place a dot when a word is recurring

'jort', 'an', 'vragen', 'aan', 'of' (0 keer)

'en' (4 keer)

'jan' (3 keer)

'ze' (4 keer)



Recurrence Matrix / Recurrence Plot

Recurrence Quantification Analysis

auto-Recurrence: Symmetric recurrence plot around the LOS (Line of Synchronisation)

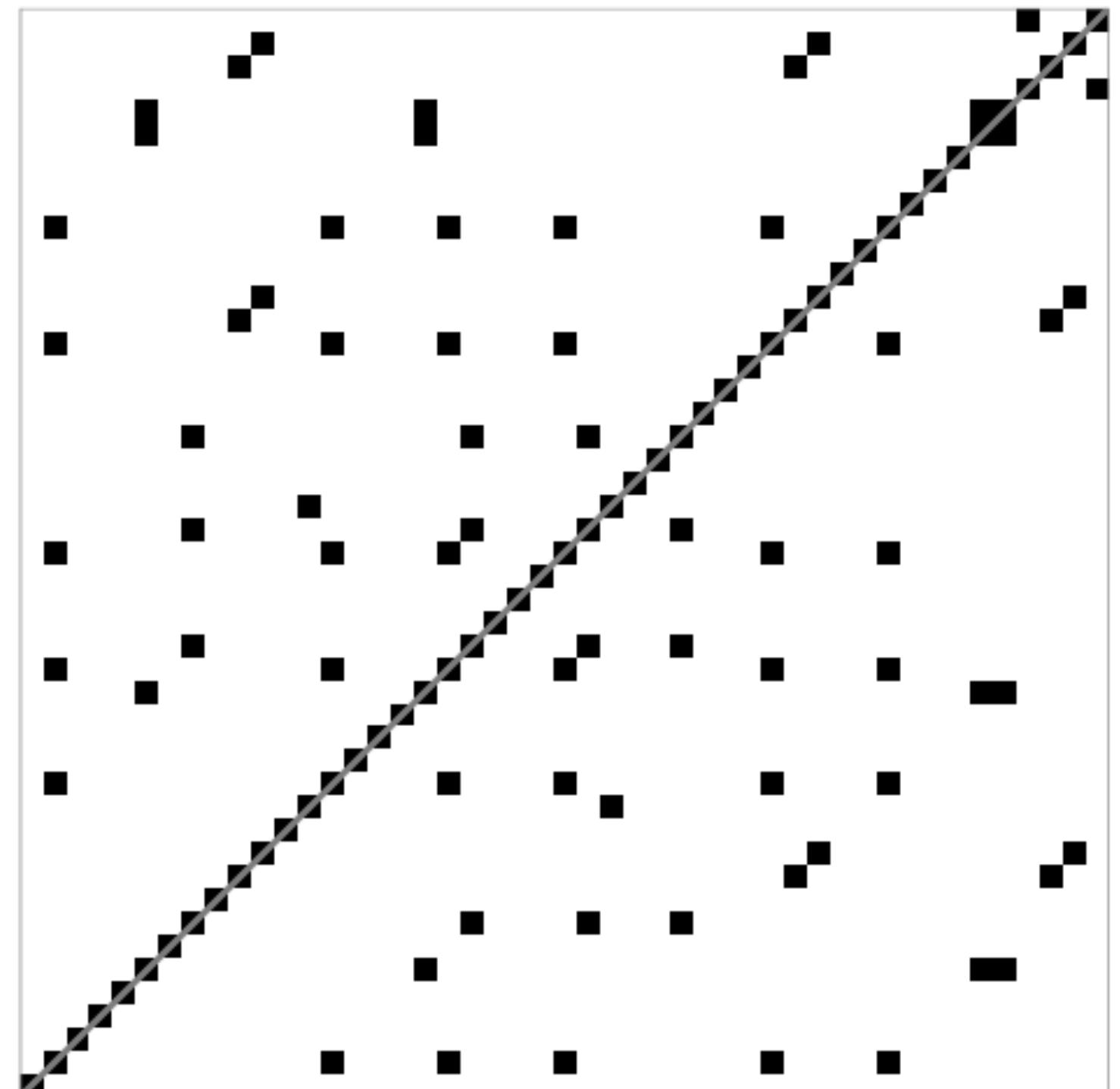
Categorical (nominal): 1 point = repetition of a category

Quantify patterns of recurrences:

Recurrence Rate (RR): Proportion actual recurrent points on maximum possible recurrent point (minus the diagonal):

$$70 / (47^2 - 47) = 0.032 \text{ (3.2\%)}$$

$$35 / ((47^2 - 47) / 2) = 0.032 \text{ (3.2\%)}$$



Recurrence Matrix / Recurrence Plot

Recurrence Quantification Analysis

Diagonal lines → repetition of any pattern:
“de wandelwagen” is recurring 2 times

Determinism (DET): proportion recurrent points that lie on a diagonal line

$$8 / 70 = 0.114 \text{ (11.4\%)}$$

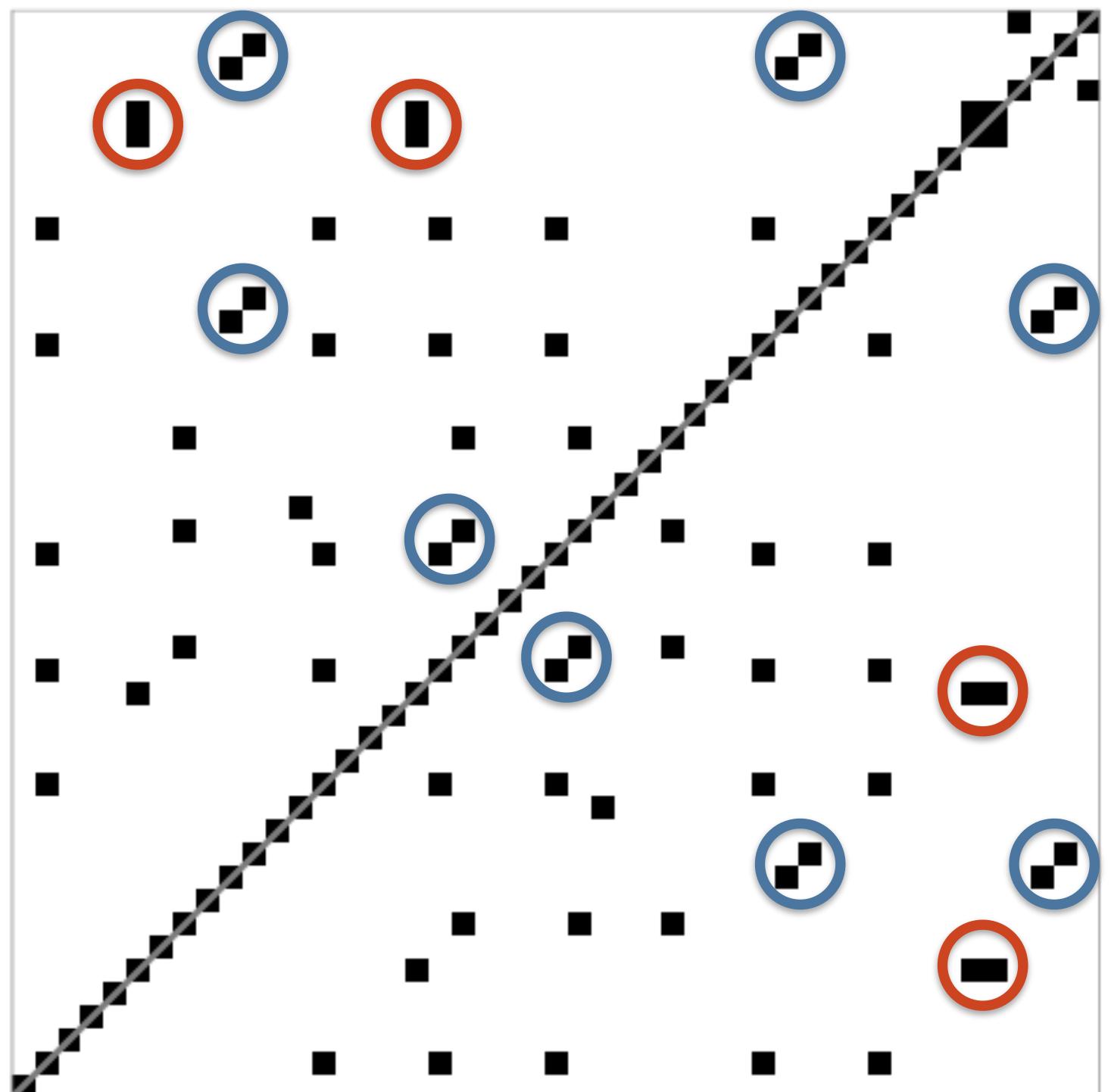
$$4 / 35 = 0.114 \text{ (11.4\%)}$$

Vertical lines → recurrence of exactly the same value:
“jan jan”

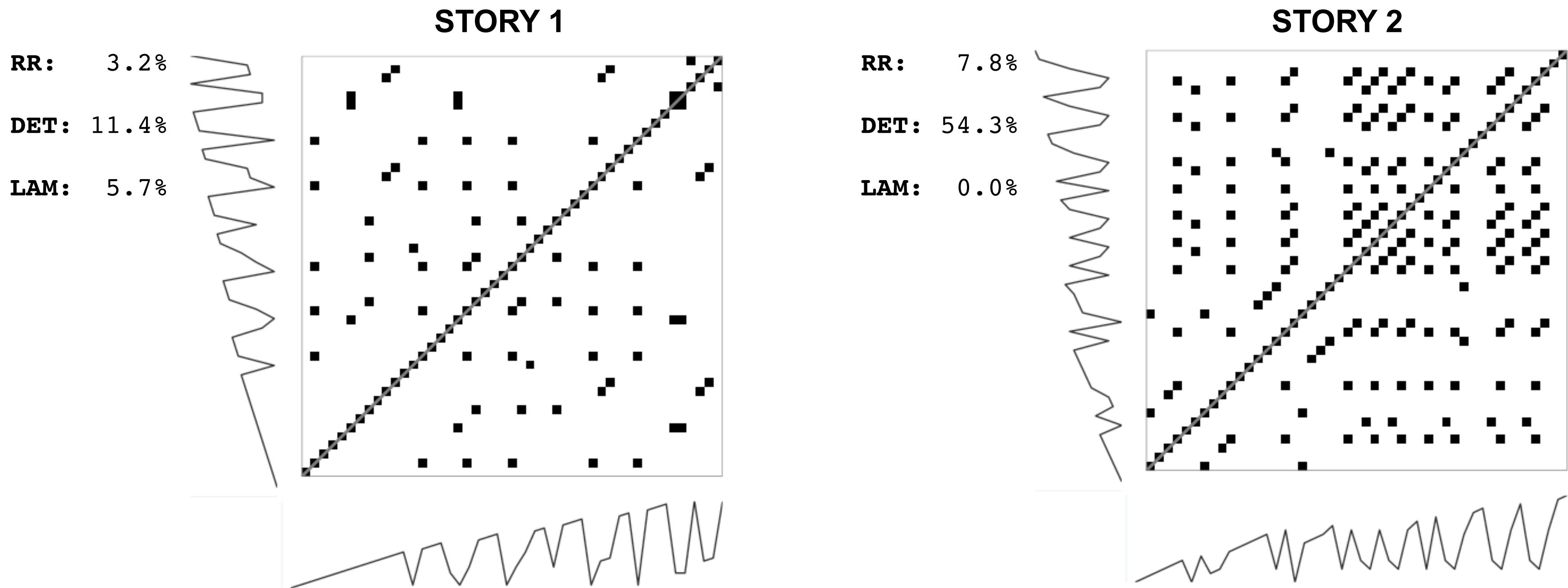
Laminarity (LAM): proportion recurrent points that lie on a vertical line

$$4 / 70 = .057 \text{ (5.7\%)}$$

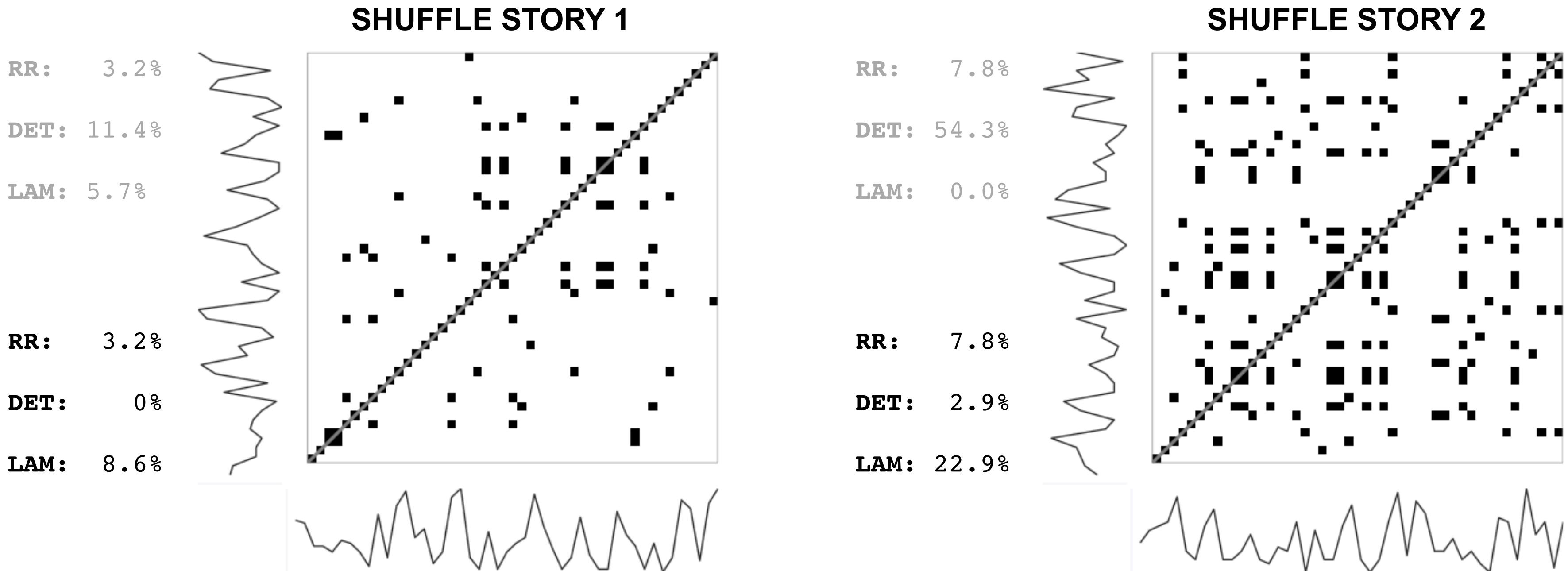
$$2 / 35 = .057 \text{ (5.7\%)}$$



Recurrence Quantification Analysis



Recurrence Quantification Analysis



Executive functions? RQA analysis of the RNG task

Oomens, W., Maes, J. H., Hasselman, F., & Egger, J. I. (2015). A time series approach to random number generation: using recurrence quantification analysis to capture executive behavior. *Frontiers in Human Neuroscience*, 9

Executive control:

“be as random
as you can”



Vignette:

R manual or: <https://fredhasselman.github.io/casnet/index.html>

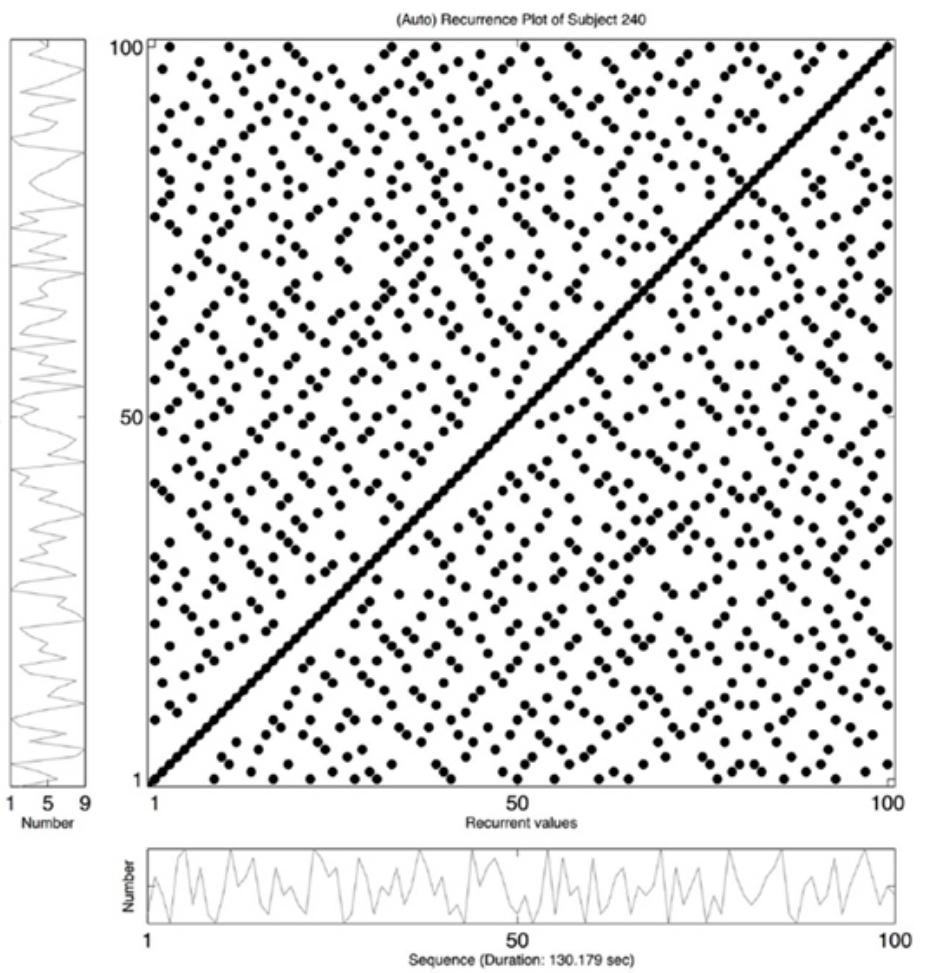
Behavioural Science Institute
Radboud University Nijmegen



A

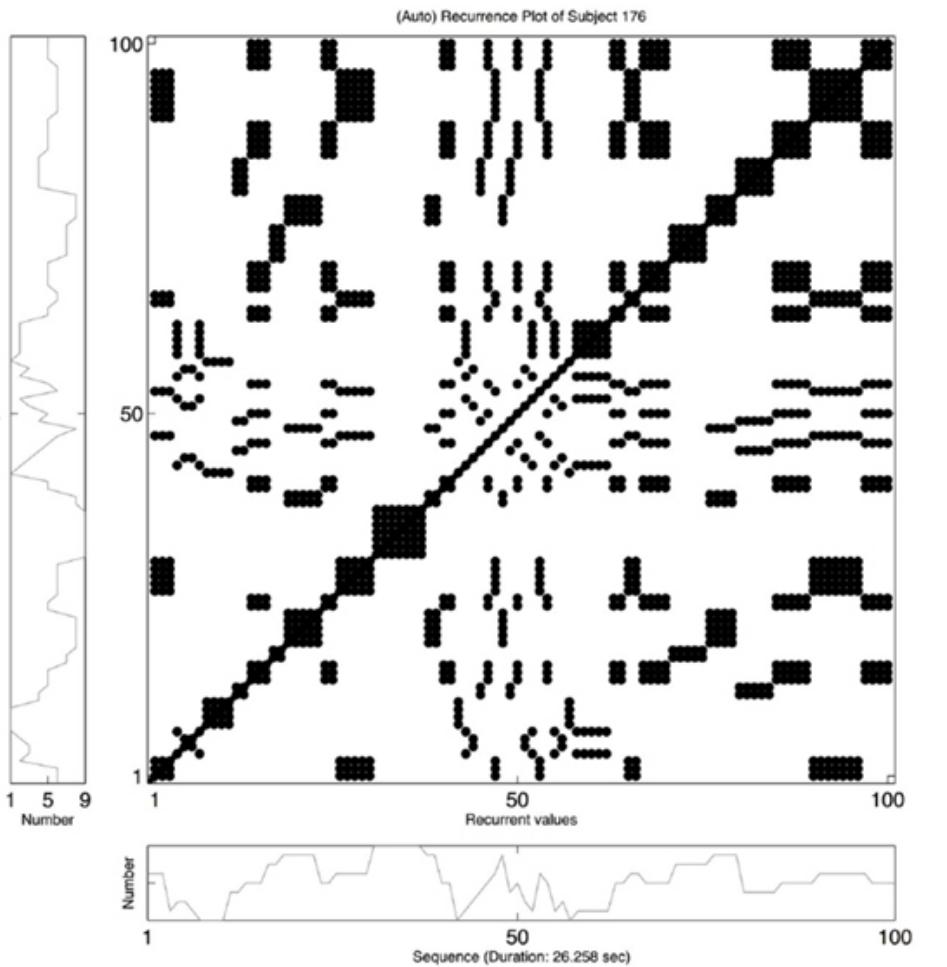
RQA measures:

REC = 0.105
DET = 0.185
Lmn = 2.04
Lmx = 3
ENT = 0.176
LAM = 0
Vmn = NaN
Vmx = 0

**C**

RQA measures:

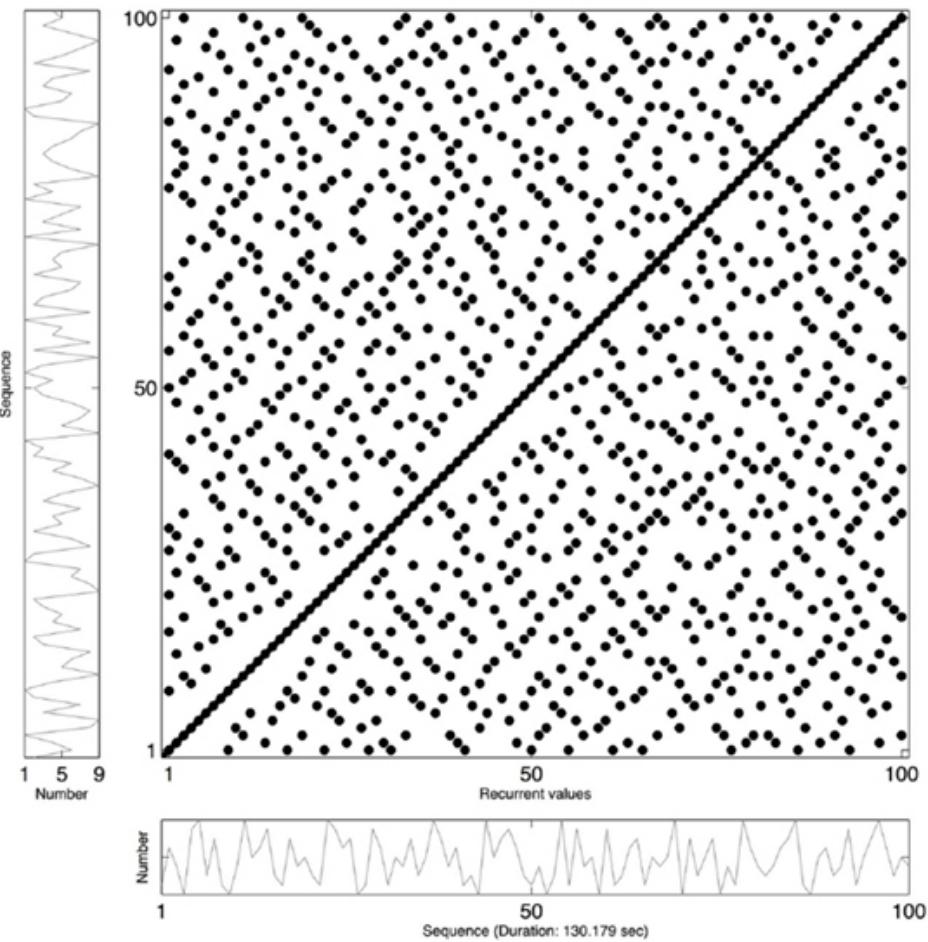
REC = 0.138
DET = 0.654
Lmn = 2.65
Lmx = 7
ENT = 1.1
LAM = 0.82
Vmn = 3.3
Vmx = 7



A

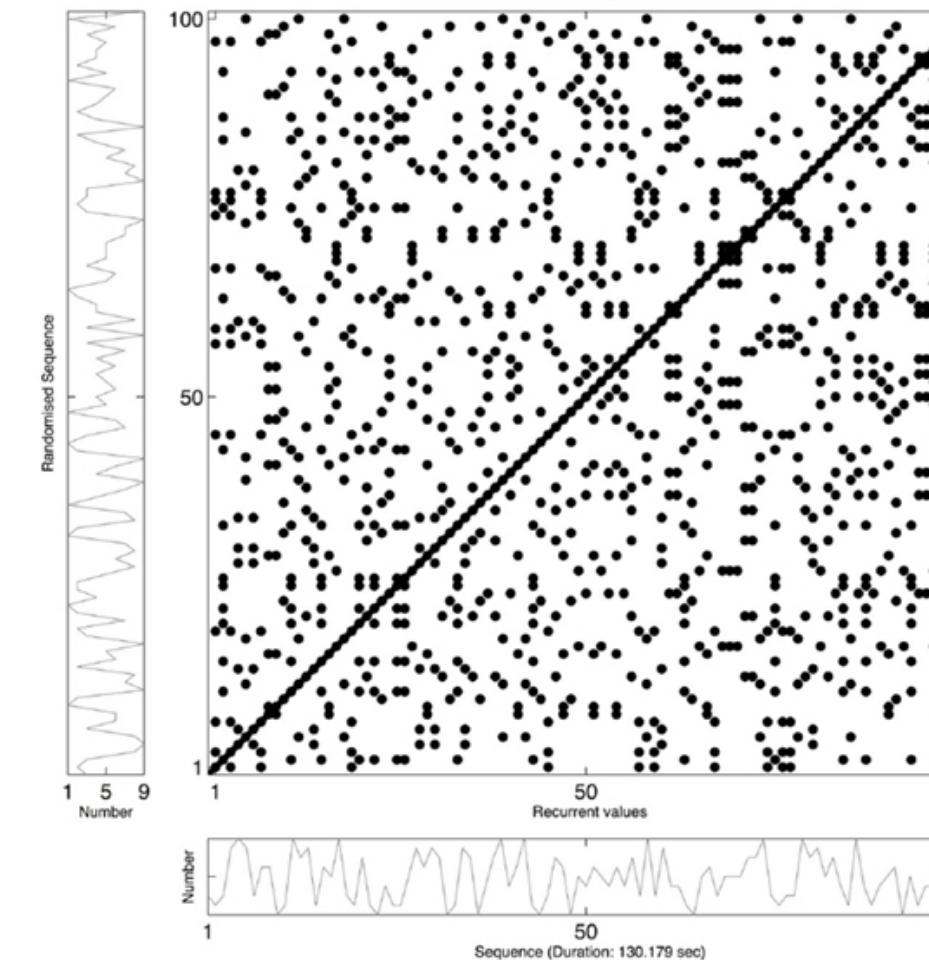
RQA measures:

REC = 0.105
 DET = 0.185
 Lmn = 2.04
 Lmx = 3
 ENT = 0.176
 LAM = 0
 Vmn = NaN
 Vmx = 0

**B**

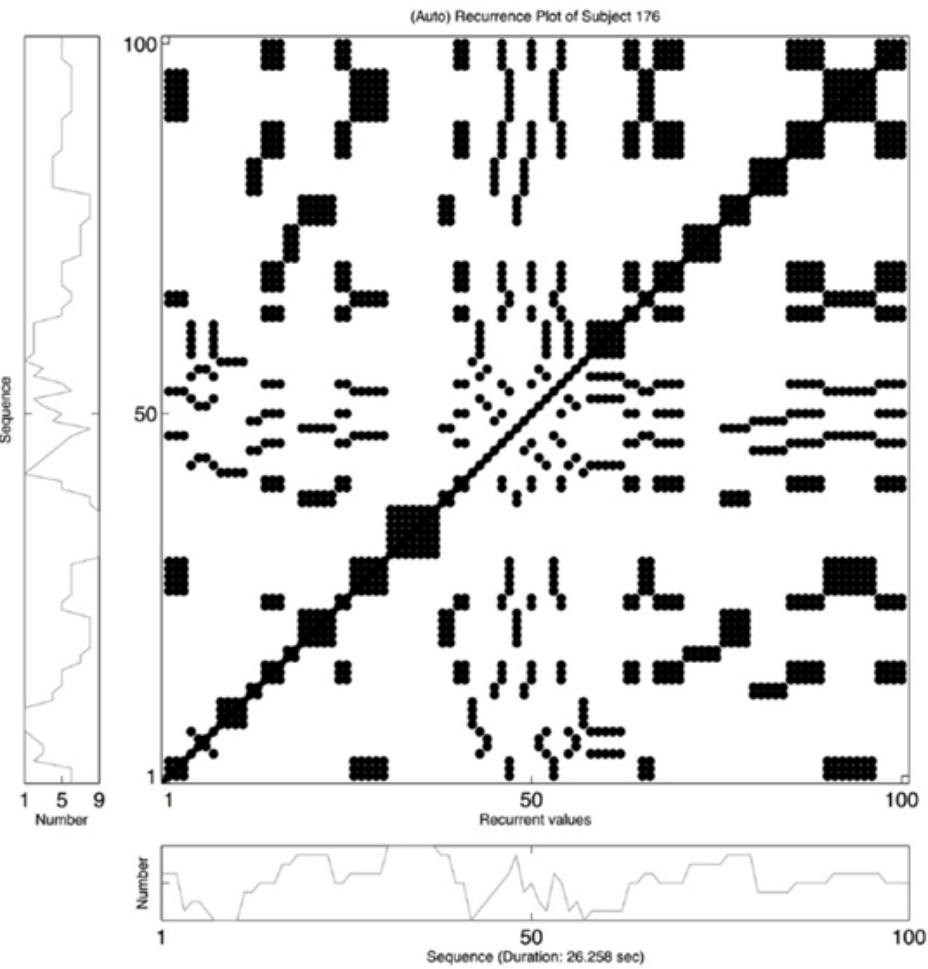
RQA measures:

REC = 0.105
 DET = 0.177
 Lmn = 2.14
 Lmx = 3
 ENT = 0.404
 LAM = 0.149
 Vmn = 2.12
 Vmx = 3

**C**

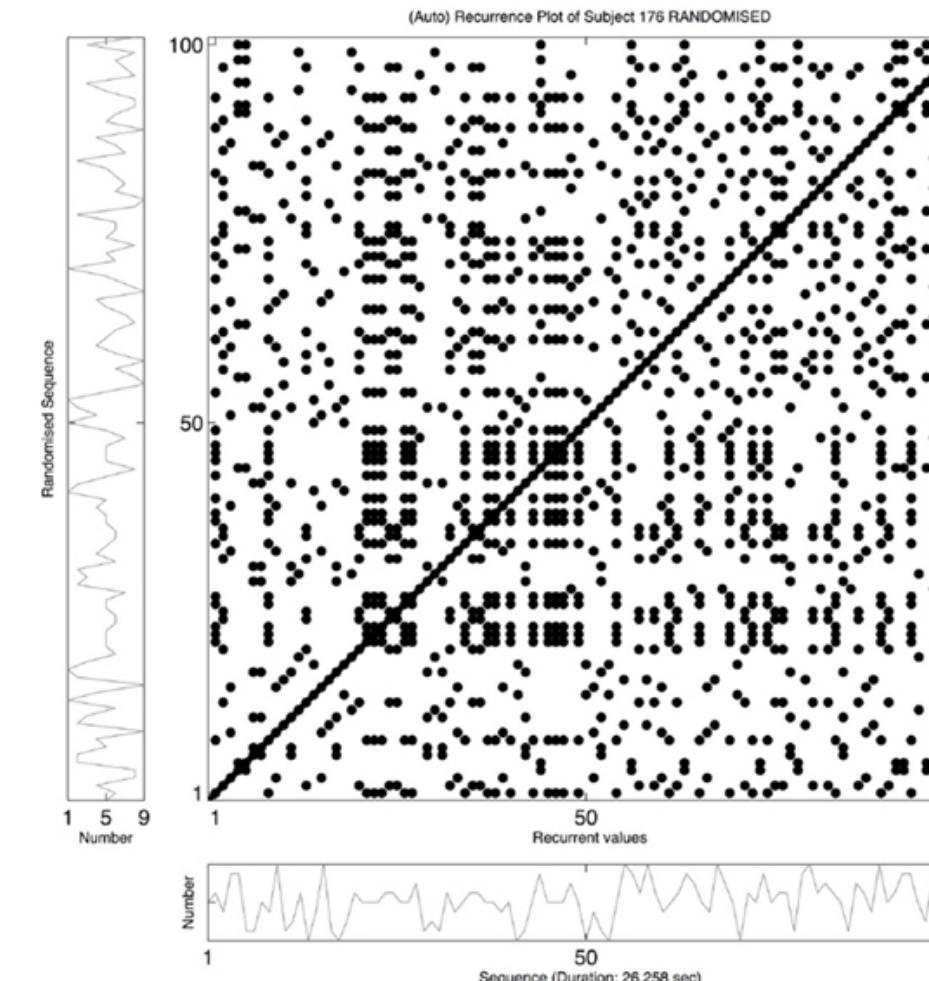
RQA measures:

REC = 0.138
 DET = 0.654
 Lmn = 2.65
 Lmx = 7
 ENT = 1.1
 LAM = 0.82
 Vmn = 3.3
 Vmx = 7

**D**

RQA measures:

REC = 0.138
 DET = 0.234
 Lmn = 2.22
 Lmx = 5
 ENT = 0.569
 LAM = 0.285
 Vmn = 2.25
 Vmx = 3



N=181

	Updating	Inhibition of prepotent responses	Output inhibition	Undefined
Redundancy	0.792			
RNG2		0.859		
RG median	-0.785			
RG mean	-0.586			
Coupon	0.830			
Adjacency		0.874		
TPI		-0.844		
Runs		0.478	0.769	
RNG		0.874		
Phi 2			0.876	
Phi 3			0.811	
Phi 4			0.691	
Phi 6	0.423	-0.569		
Phi 5	0.445		0.462	0.521
Phi 7	0.631			
RG mode	-0.475			
Eigenvalues	3.409	3.844	2.729	1.201
% of variance	21.304	24.026	17.059	7.508

Output is sorted by size and a cut-off value of 0.4 was used.

N=242

	Updating	Inhibition of prepotent responses	Output inhibition	Undefined
Redundancy	0.782			0.432
RNG2	0.713		0.478	
RG median	-0.674			-0.486
RG mean	-0.652			-0.461
Coupon	0.630			0.515
Adjacency			0.885	
TPI			-0.828	
Runs			0.791	
RNG	0.593		0.645	
Phi 2				0.879
Phi 3				0.719
Phi 4				0.570
Phi 6				0.803
Phi 5				0.637
Phi 7				0.634
RG mode				-0.546
Eigenvalues	3.200		2.817	2.392
% of variance	19.998		17.607	14.949
				19.045

Output is sorted by size and a cut-off value of 0.4 was used.

	Inhibition of prepotent responses	Updating
Averaged diagonal	0.963	
Longest diagonal	0.922	
Determinism	0.917	
Entropy	0.839	
Laminarity		0.918
Trapping time		0.878
Recurrence rate		0.486
Eigenvalues	3.487	1.857
% of variance	49.818	26.523

Output is sorted by size and a cut-off value of 0.4 was used.

	Inhibition of prepotent responses	Updating
Averaged diagonal	0.957	
Entropy	0.937	
Longest diagonal	0.852	
Determinism	0.730	
Laminarity		0.861
Trapping time		0.765
Recurrence rate		0.712
Eigenvalues	3.086	1.948
% of variance	44.085	27.830

Output is sorted by size and a cut-off value of 0.4 was used.

Phase Space Reconstruction

continuous time series

Quantifying Complex Dynamics

scale-free / fractal
highly correlated / interdependent
nonlinear / maybe chaotic
result of multiplicative interactions

Takens' (1981) Embedding Theorem tells us that a (strange) attractor can be recovered ("reconstructed") from observations of a single component process of a complex interaction-dominant system.



How to study interaction-dominant systems

As you know in a **coupled system** the time evolution of one variable depends on other variables of the system. This implies that one variable contains information about the other variables (of course depending upon the strength of coupling and maybe the type of interaction)

So given the Lorenz system ...

$$dX/dt = \delta \cdot (Y - X)$$

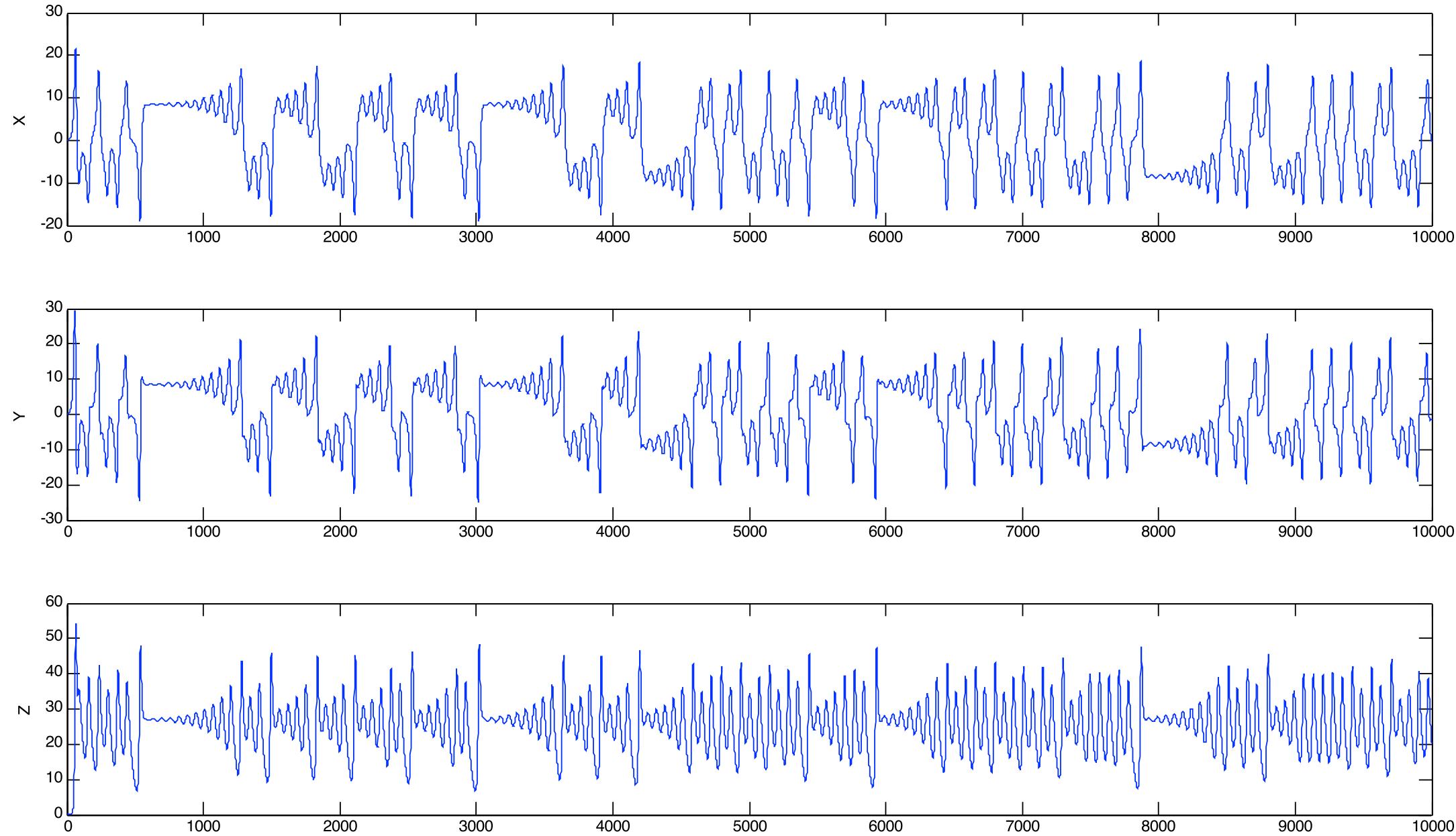
$$dY/dt = r \cdot X - Y - X \cdot Z$$

$$dZ/dt = X \cdot Y - b \cdot Z$$

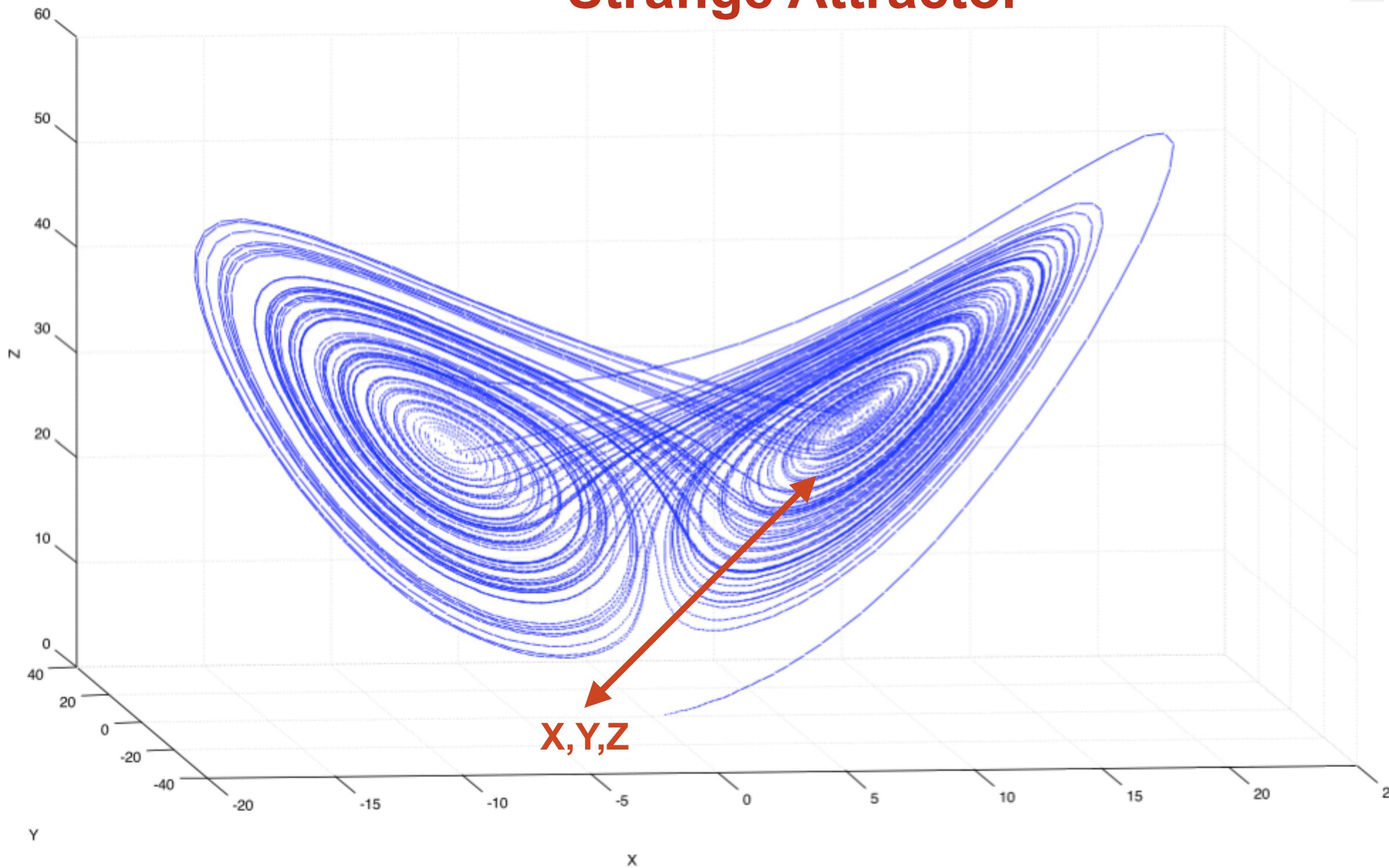
Takens' theorem suggests that we should be able to reconstruct the highly chaotic “butterfly” attractor by just using $X(t)$ [or $Y(t)$ or $Z(t)$] ...



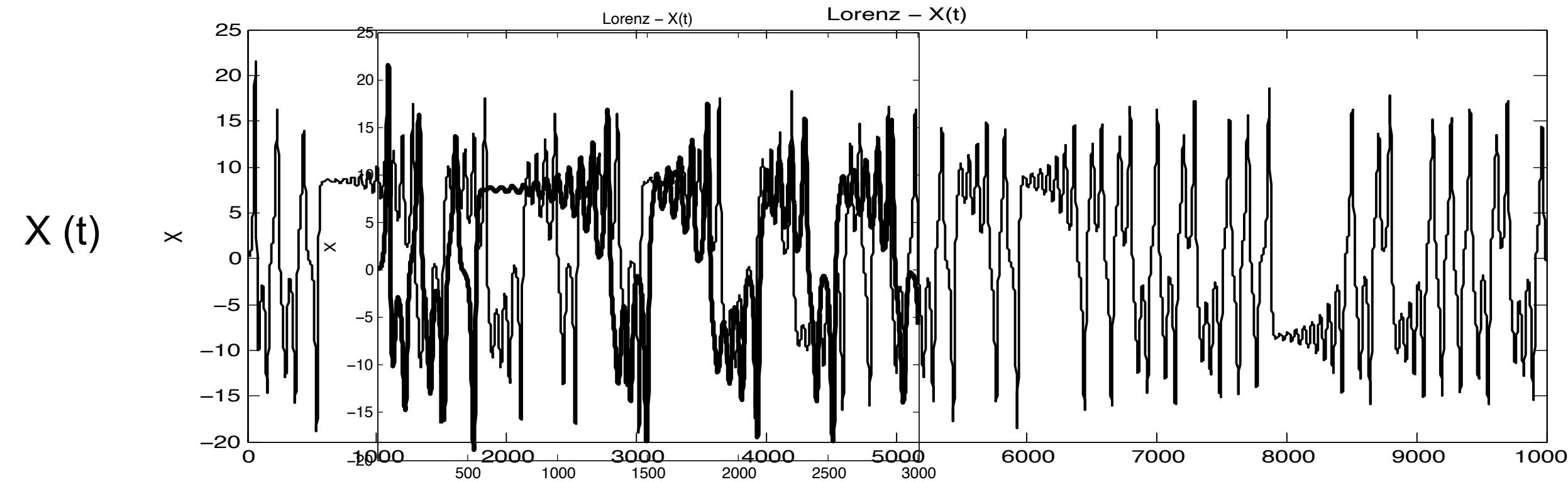
Lorenz system – Time series of X, Y and Z



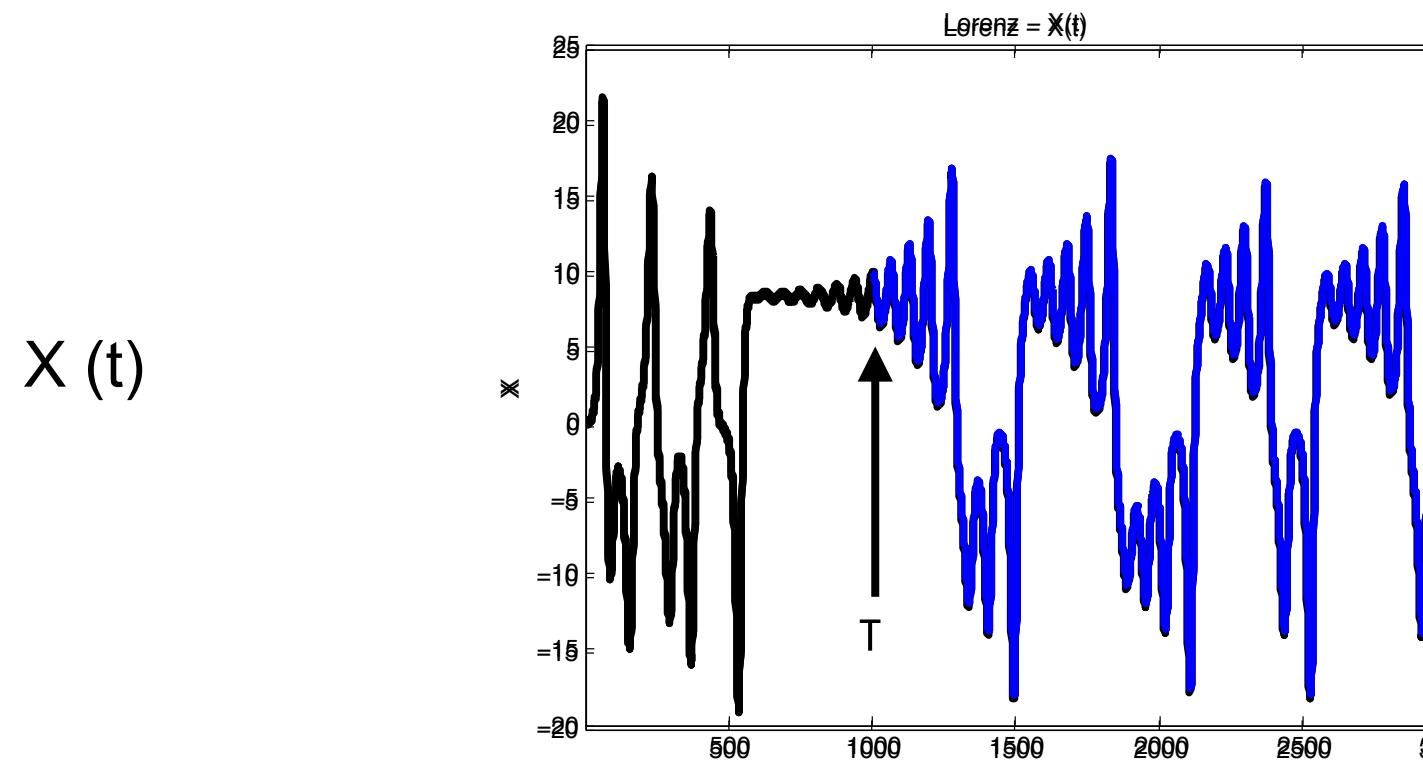
Lorenz system – X,Y,Z State space Strange Attractor



Creating surrogate dimensions using the method of delays



Creating surrogate dimensions using the method of delays



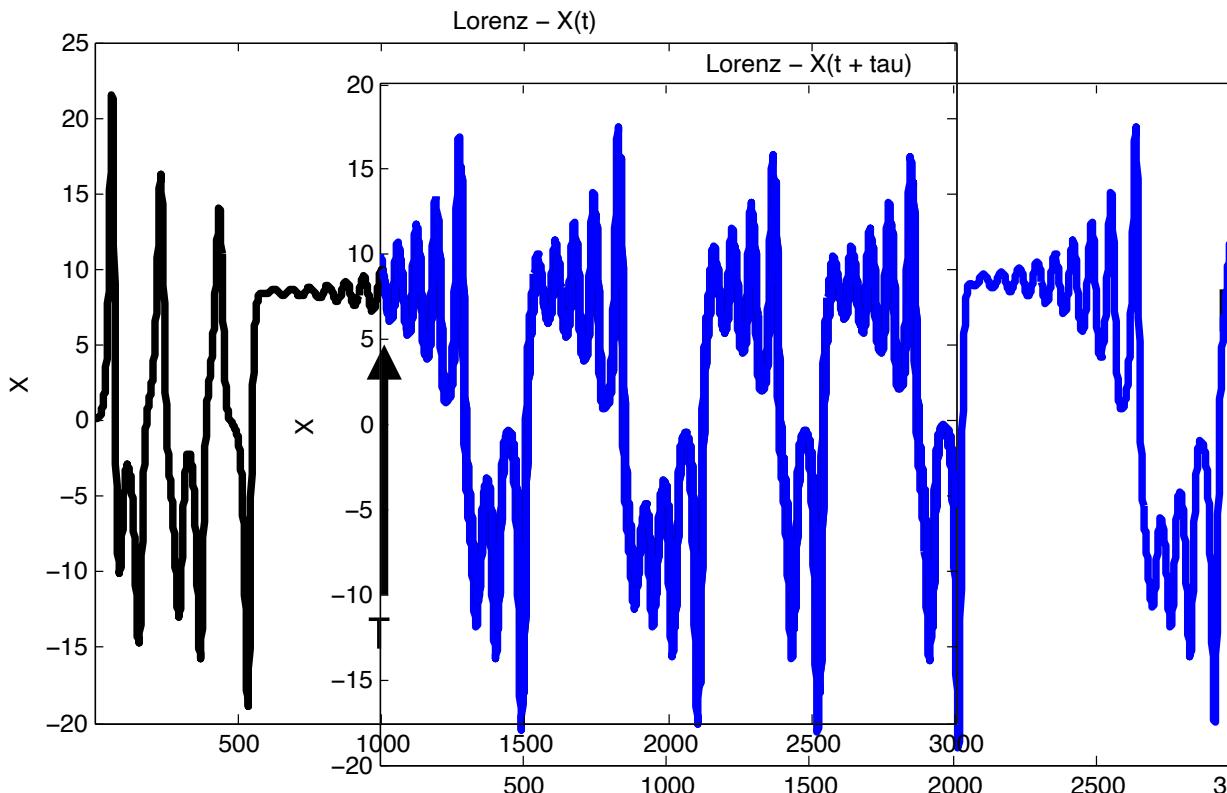
Let's take our embedding delay
or lag to be:

$$T = 1000$$



Creating surrogate dimensions using the method of delays

$X(t)$



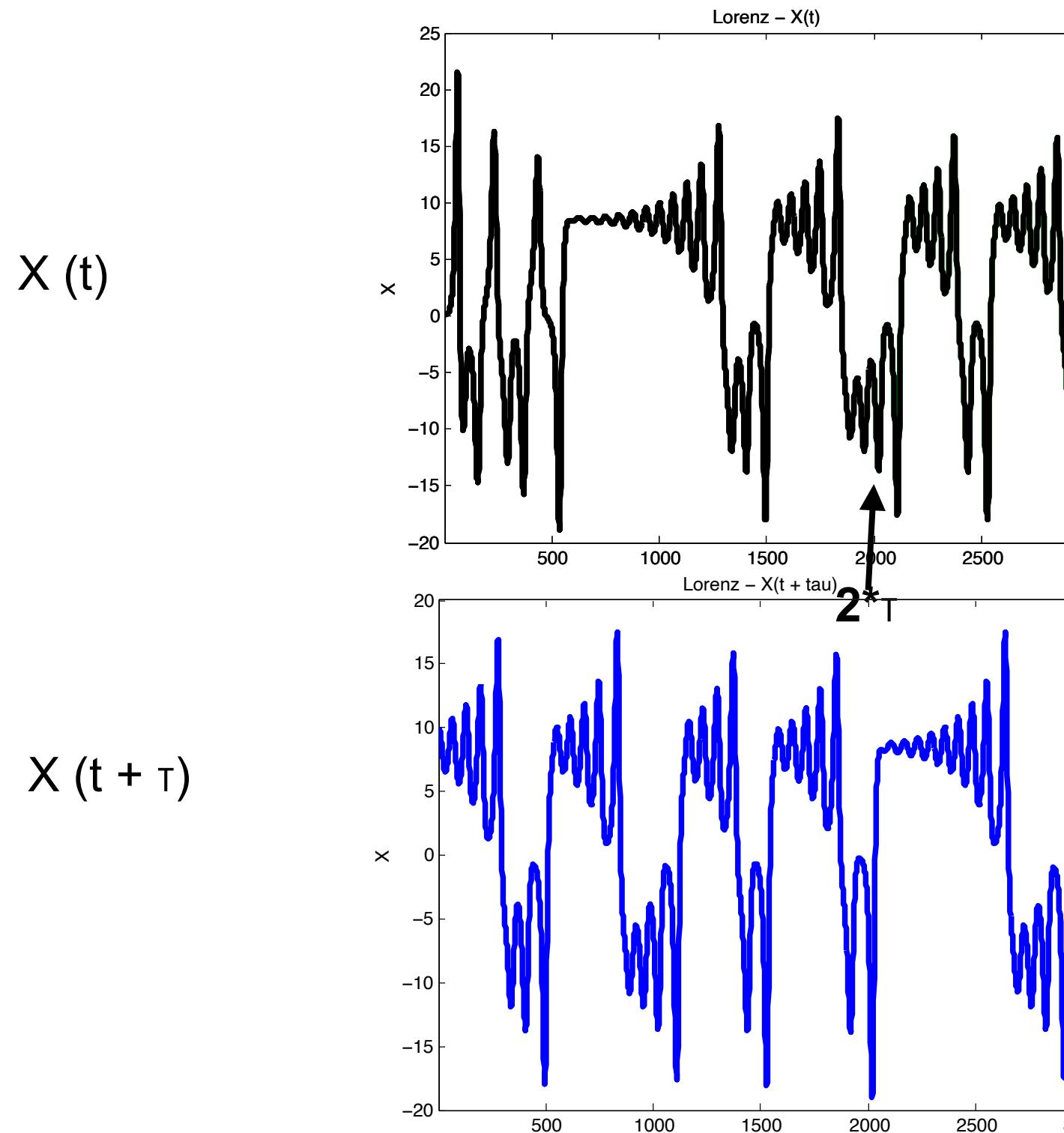
Let's take our embedding delay
or lag to be:

$$\tau = 1000$$

$X(t + \tau)$

Data point $1 + \tau$ [$X(t) = 1001$]
becomes data point 1 for this
dimension

Creating surrogate dimensions using the method of delays



Let's take our embedding delay
or lag to be:

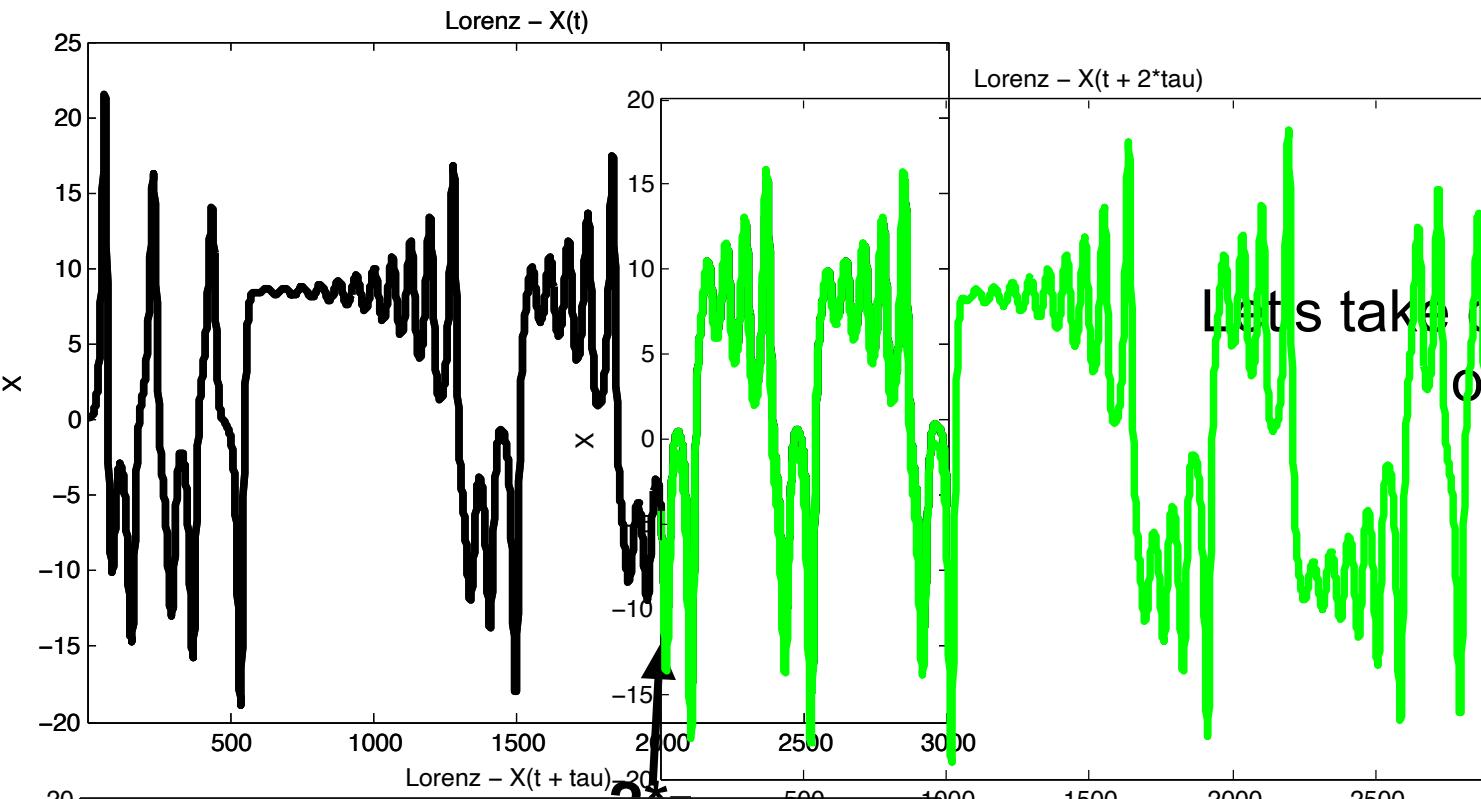
$$T = 1000$$

Data point $1 + T$ [$X(t) = 1001$]
becomes data point 1 for this
dimension



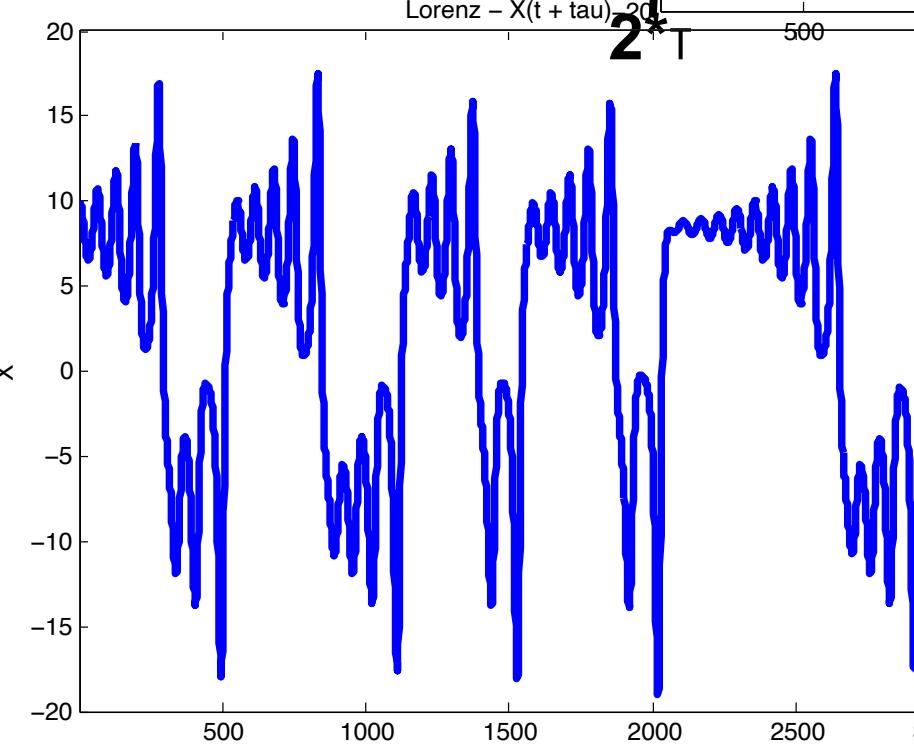
Creating surrogate dimensions using the method of delays

$X(t)$



$X(t + \tau)$

$X(t + 2\tau)$



Data point $1 + T [X(t) = 1001]$
becomes data point 1 for this
dimension

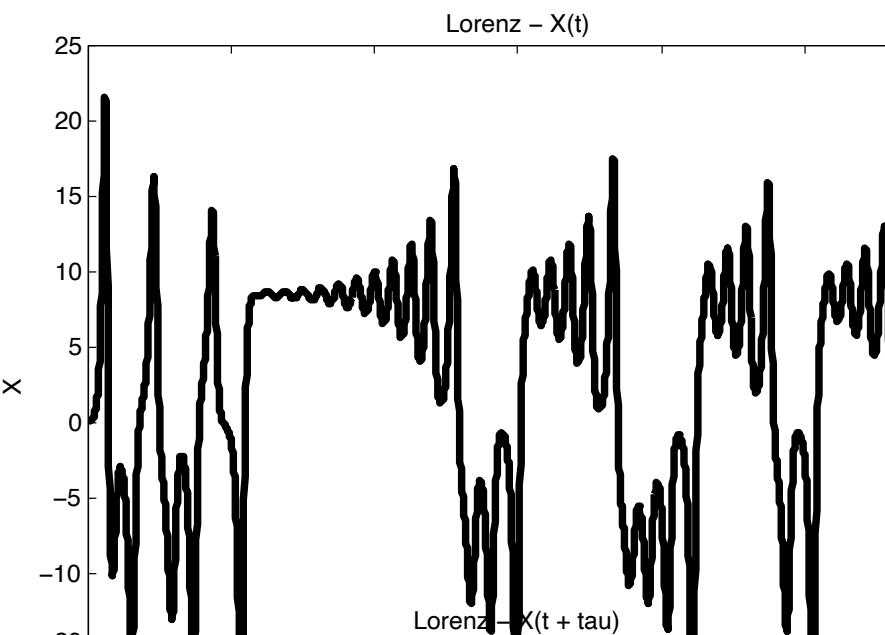
Data point $1 + 2T [X(t) = 2001]$
becomes data point 1 for this

dimension
Behavioural Science Institute
Radboud University Nijmegen

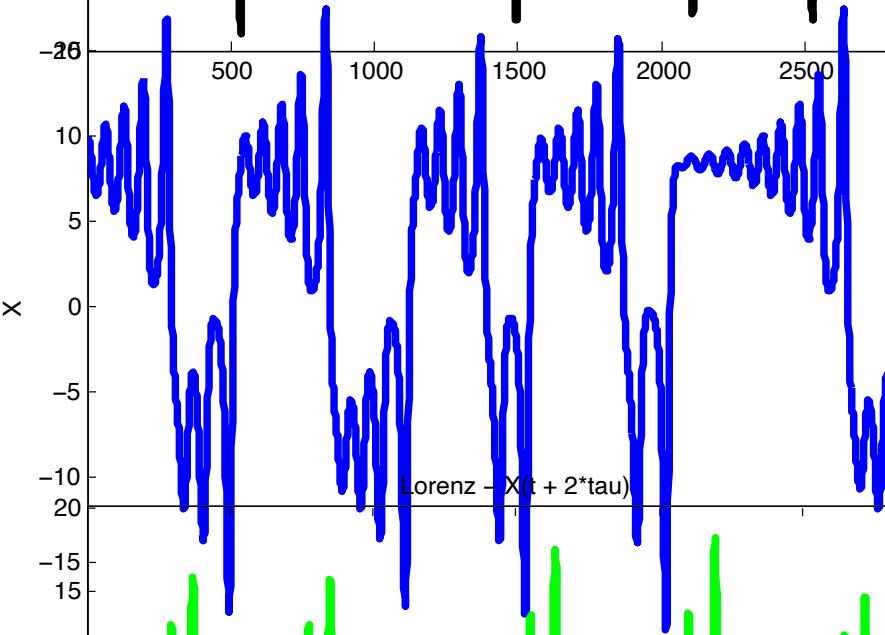


Creating surrogate dimensions using the method of delays

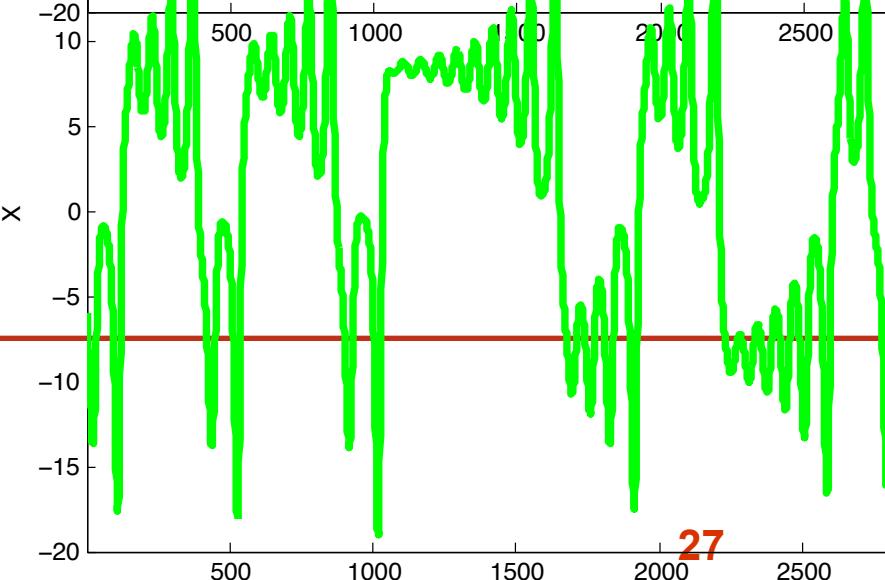
$X(t)$



$X(t + \tau)$



$X(t + 2\tau)$



The embedding lag reflects
the point in the time series at
which we are getting
new information about the system...

In theory any lag can be used,
everything is interacting...

We are looking for the lag which
is optimal, gives us maximal new
information about the temporal structure in
the data...

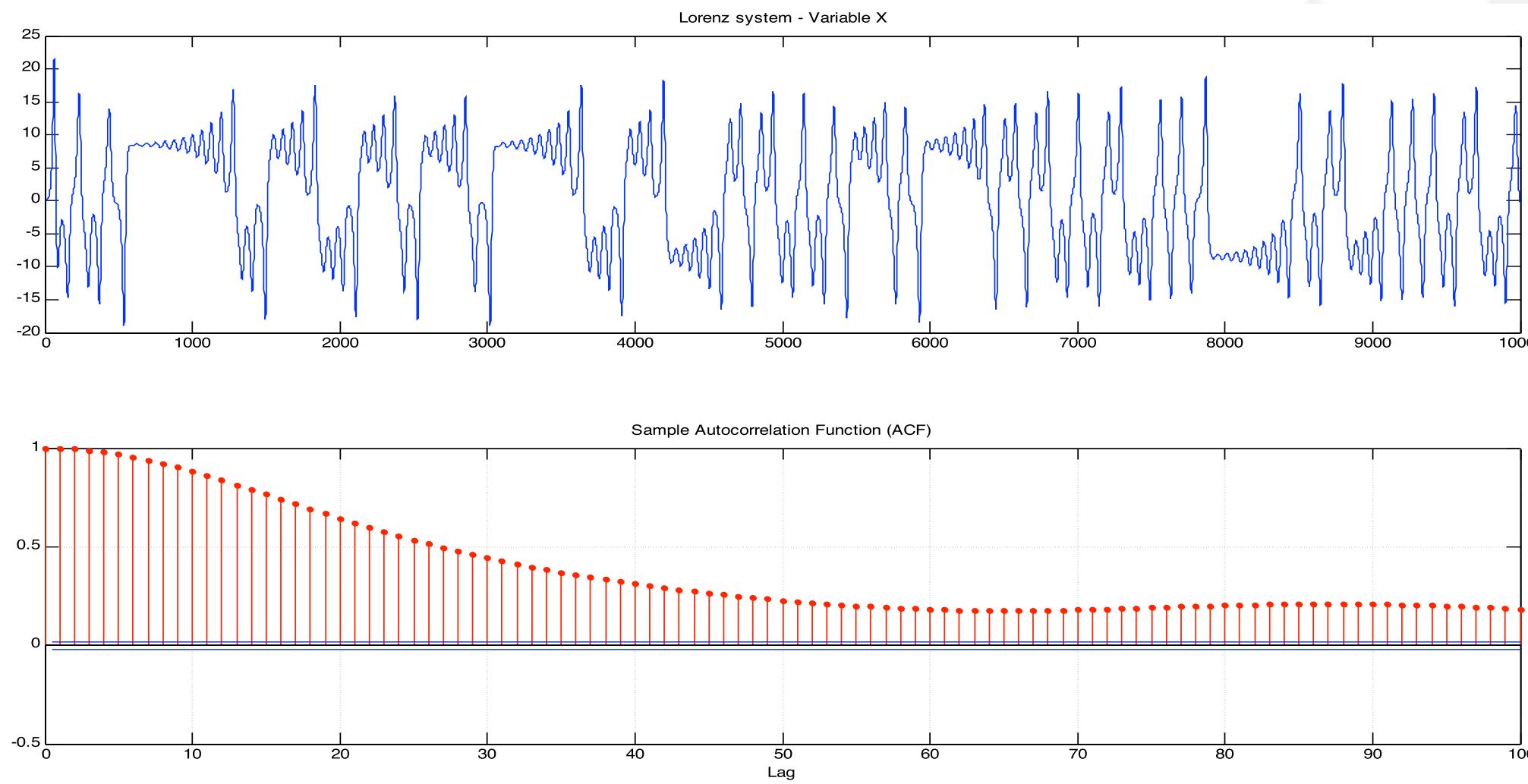
Intuitively:
Where the autocorrelation is zero

We are creating a return plot to
examine the systems' state space!

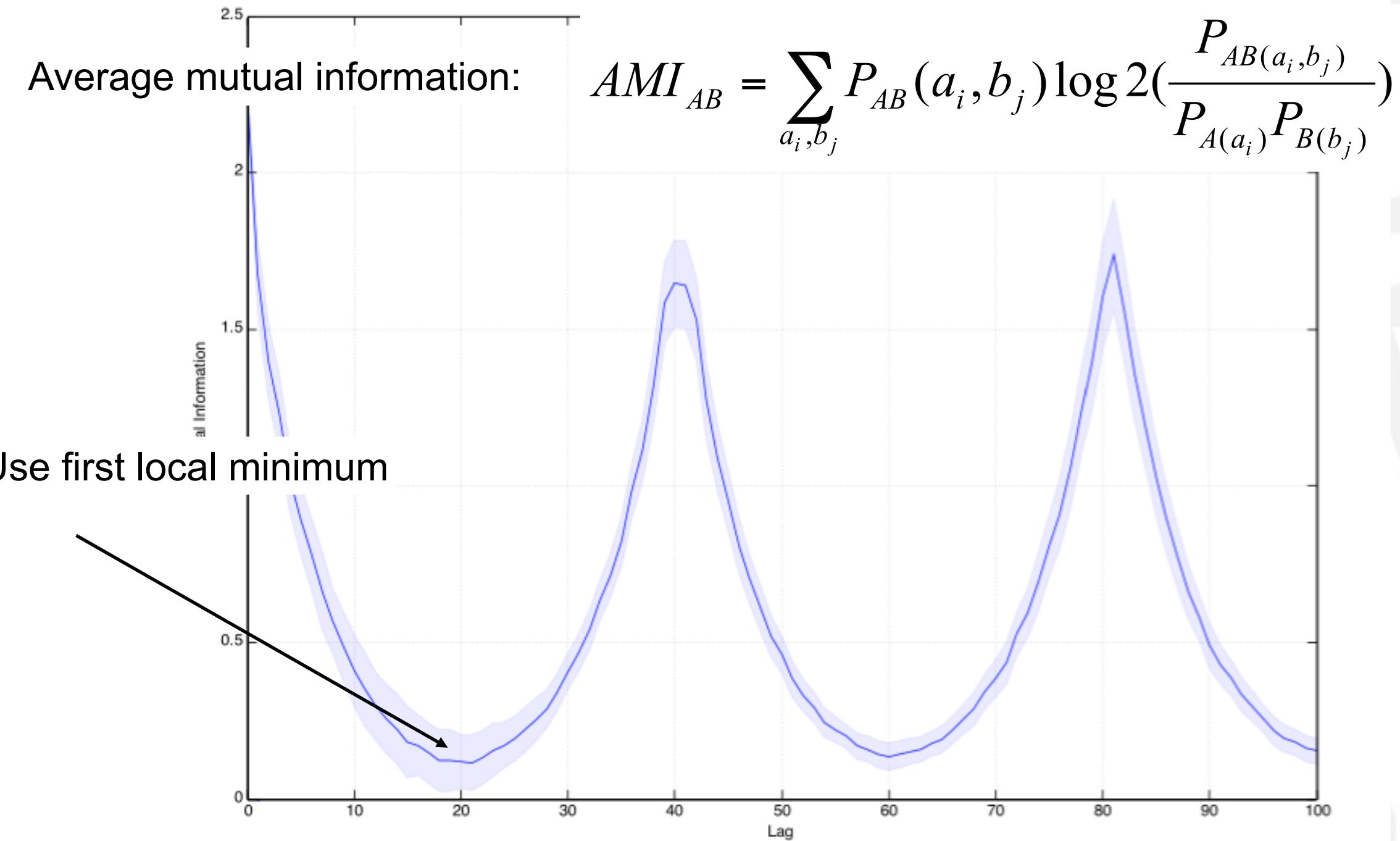


How to determine embedding lag?

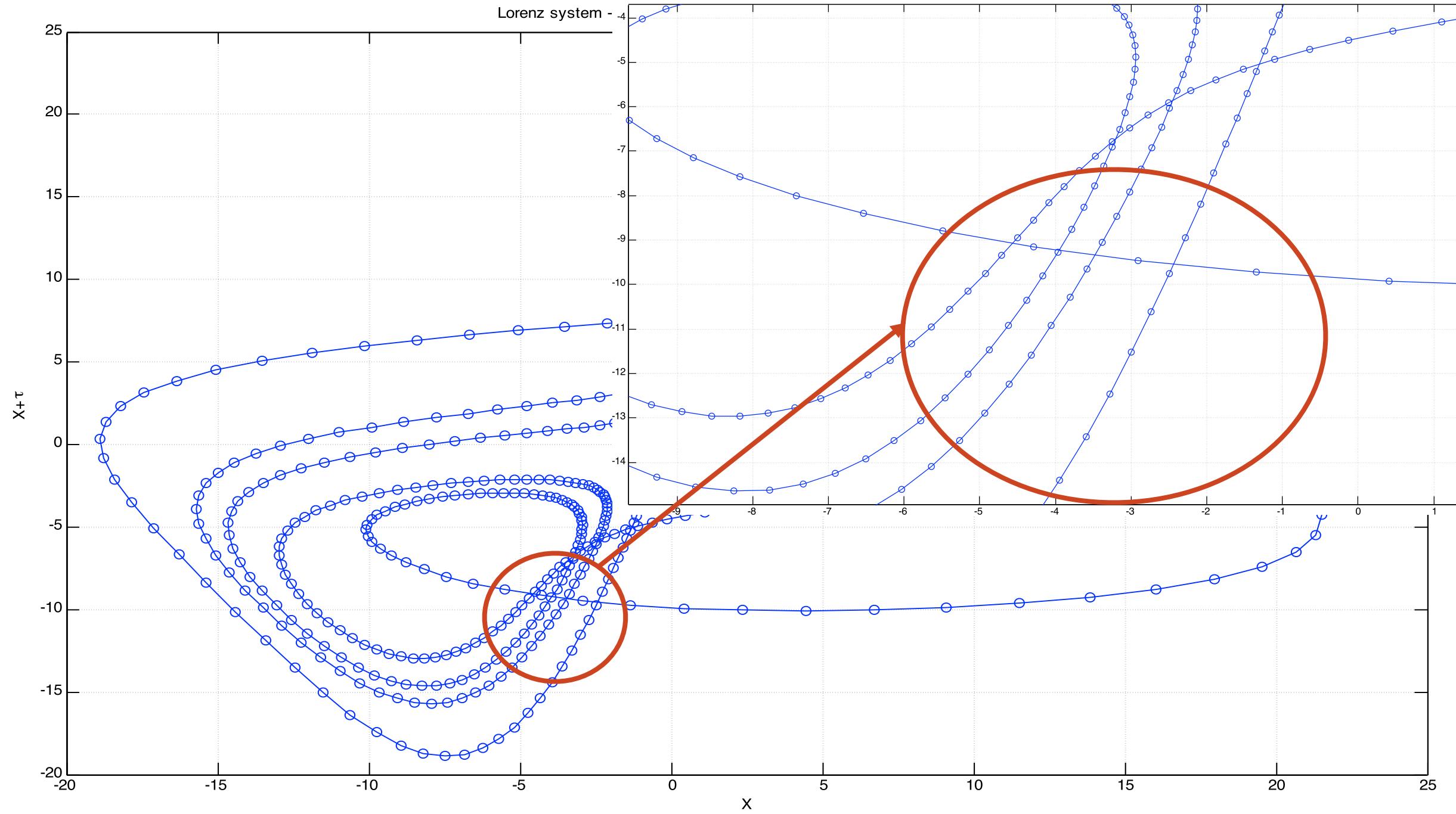
- We saw that the autocorrelation function is not very helpful when you are dealing with long range correlations in the data.



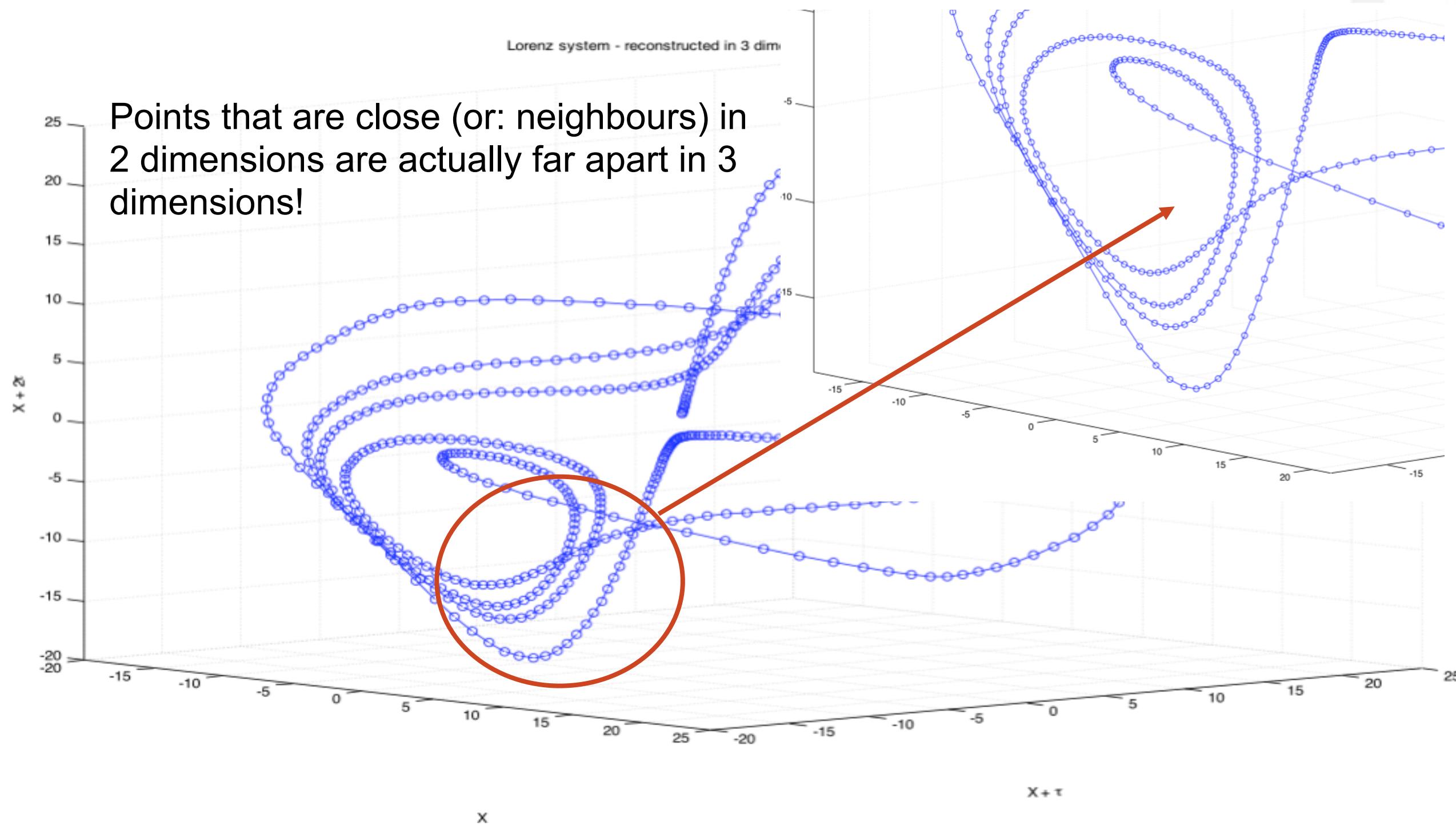
Lorenz system – Determine embedding lag



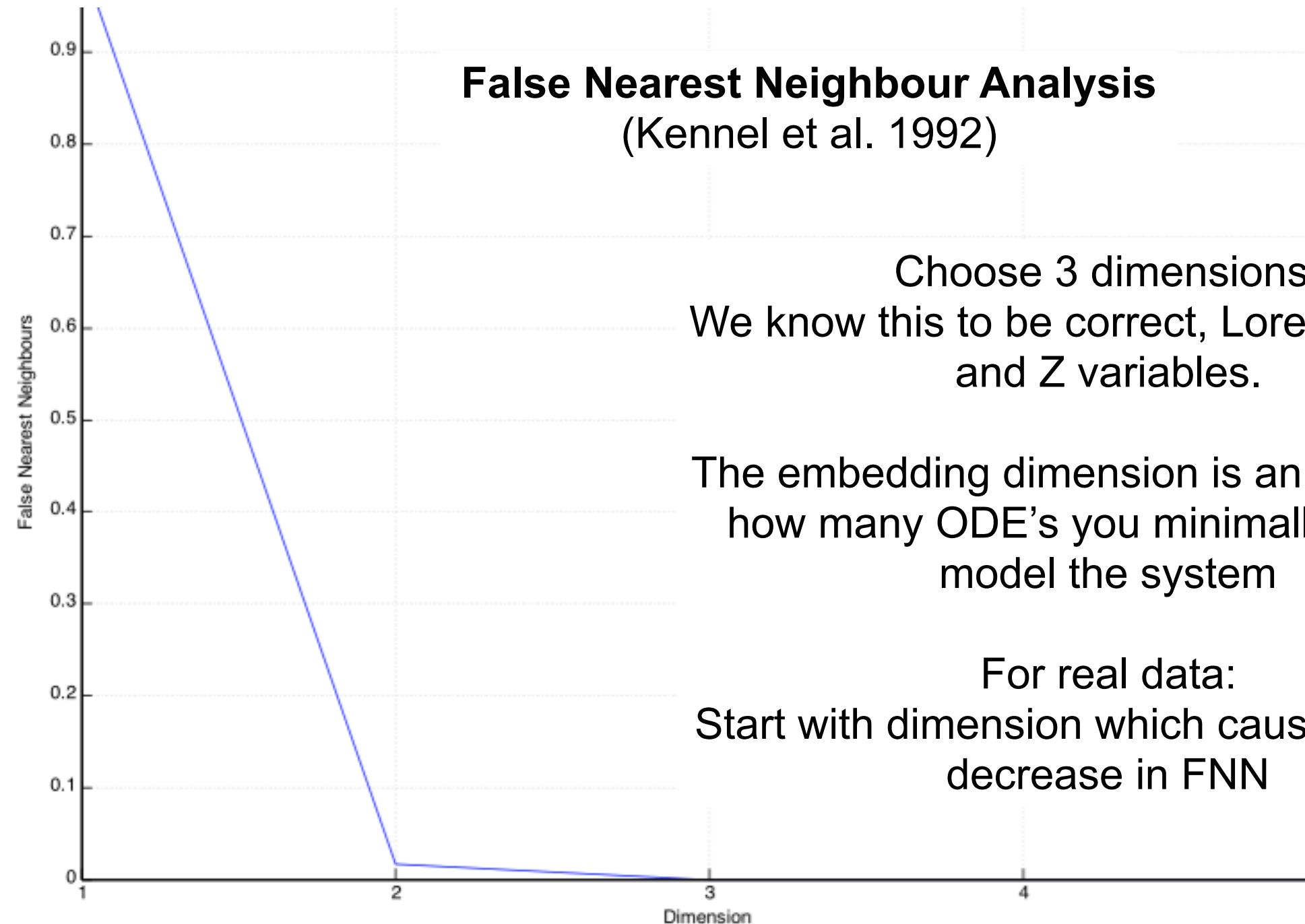
How many dimensions? Determine *embedding dimension* (m)



Lorenz system – Determine embedding dimension



Lorenz system – Determine embedding dimensions



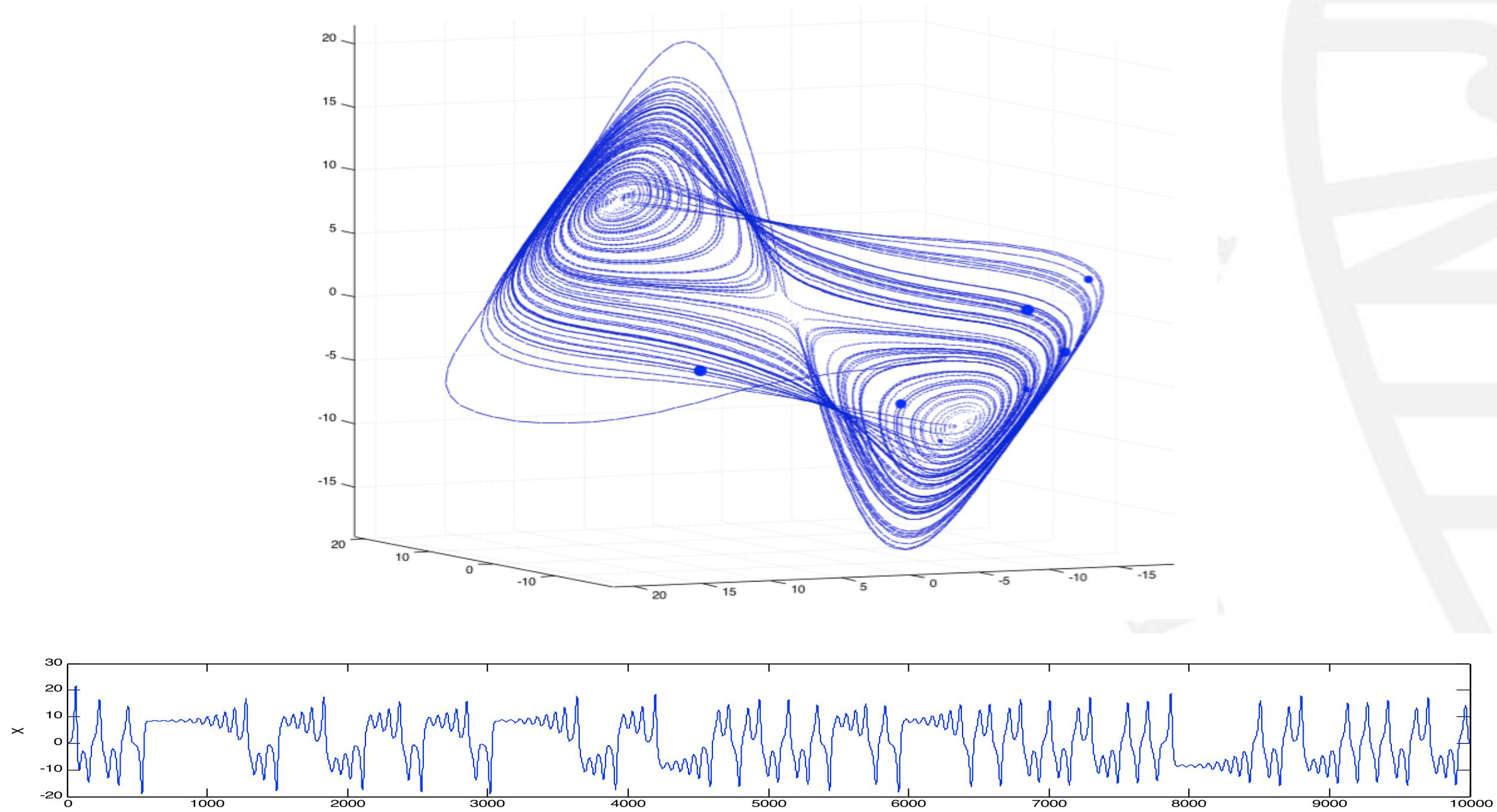
False Nearest Neighbour Analysis
(Kennel et al. 1992)

Choose 3 dimensions!
We know this to be correct, Lorenz has X, Y
and Z variables.

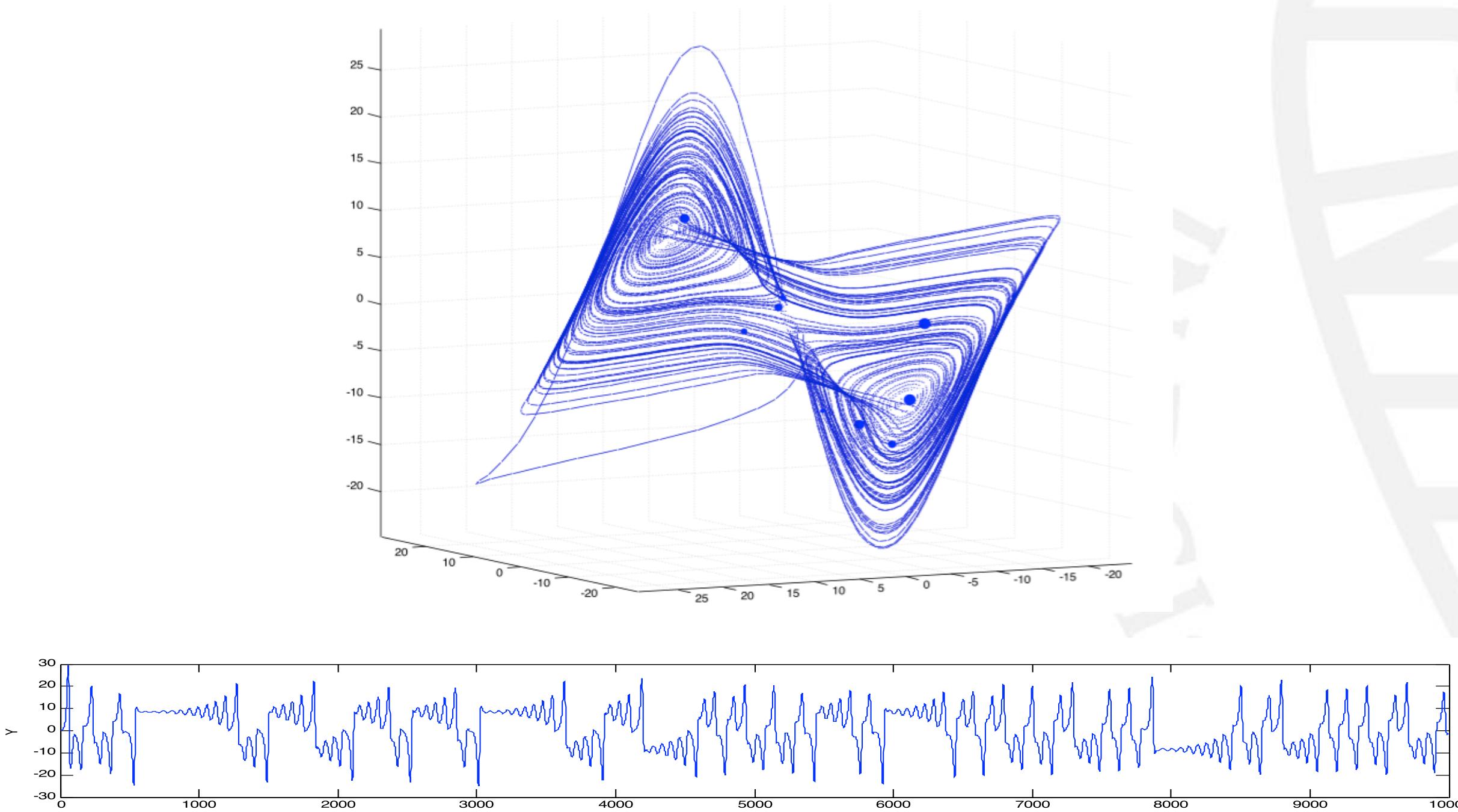
The embedding dimension is an estimate of
how many ODE's you minimally need to
model the system

For real data:
Start with dimension which causes greatest
decrease in FNN

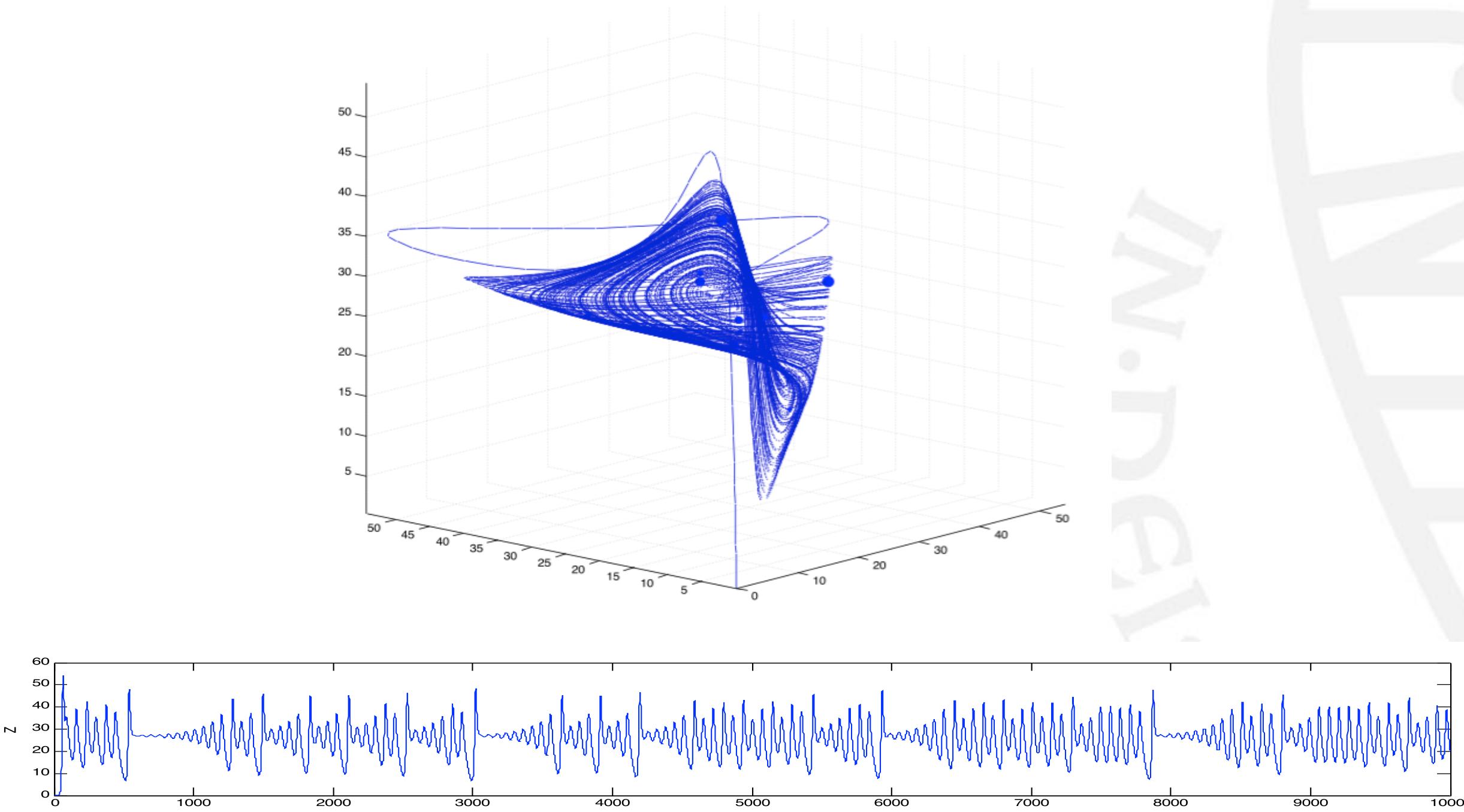
Lorenz system – Reconstruct phase space using X



Lorenz system – Reconstruct phase space using Y



Lorenz system – Reconstruct phase space using Z



Isn't that amazing?

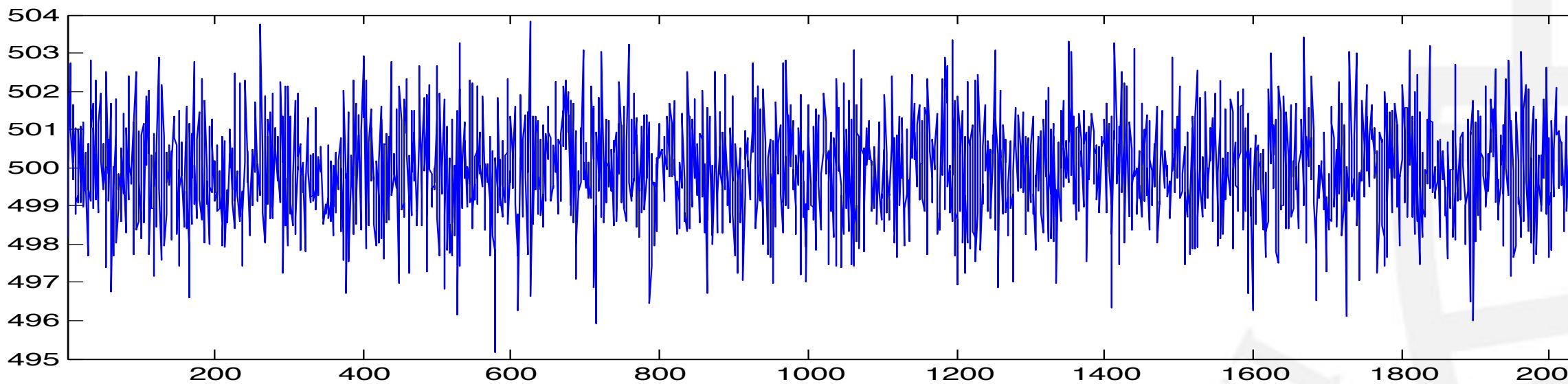
- Take a moment to realise what we just did:
- The state space (defined by X,Y and Z) of a complex, nonlinear chaotic system was reconstructed to a phase space (lag plot) of 3 surrogate dimensions $X, X_{t+\tau}, X_{t+2\tau}$
- **You only need to measure one variable of a system!!**
... because “everything is interacting”...
We exploit (and need) the dependencies in the data!

The length of your data set needs to be long enough to create the surrogate dimension.

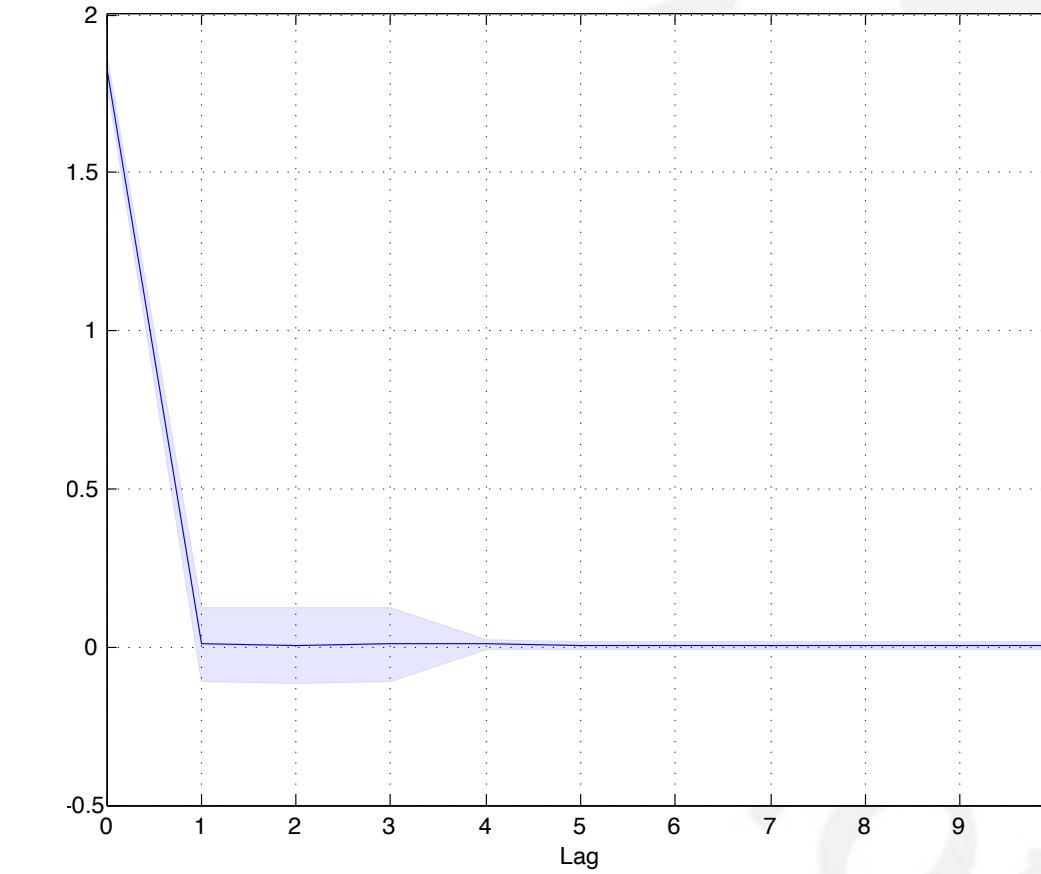
- The reconstruction process **does not make many assumptions about the data**. You can also try to reconstruct a phase space from a random variable.
(What will happen?)



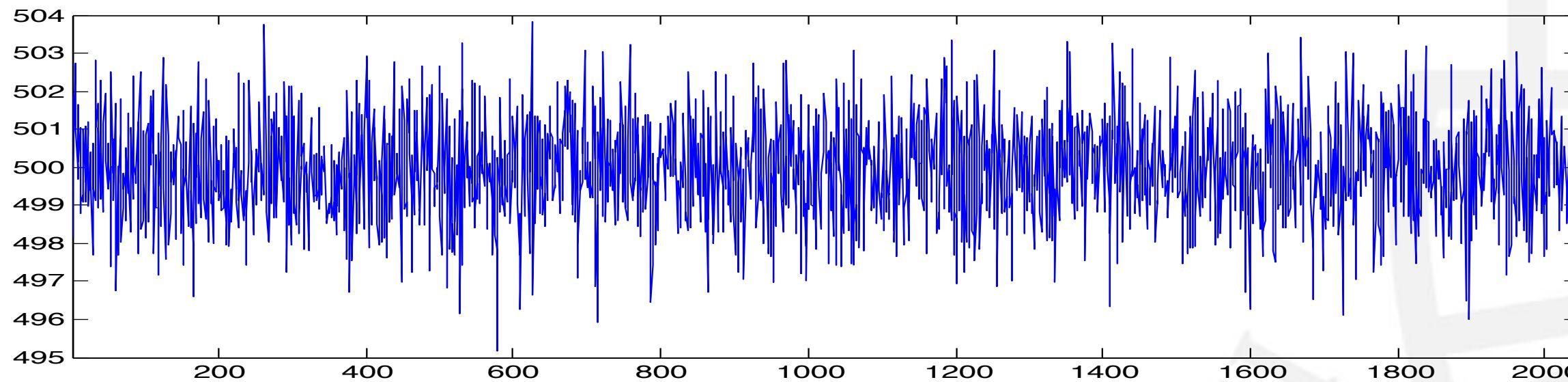
Suppose we have measured a true IID variable



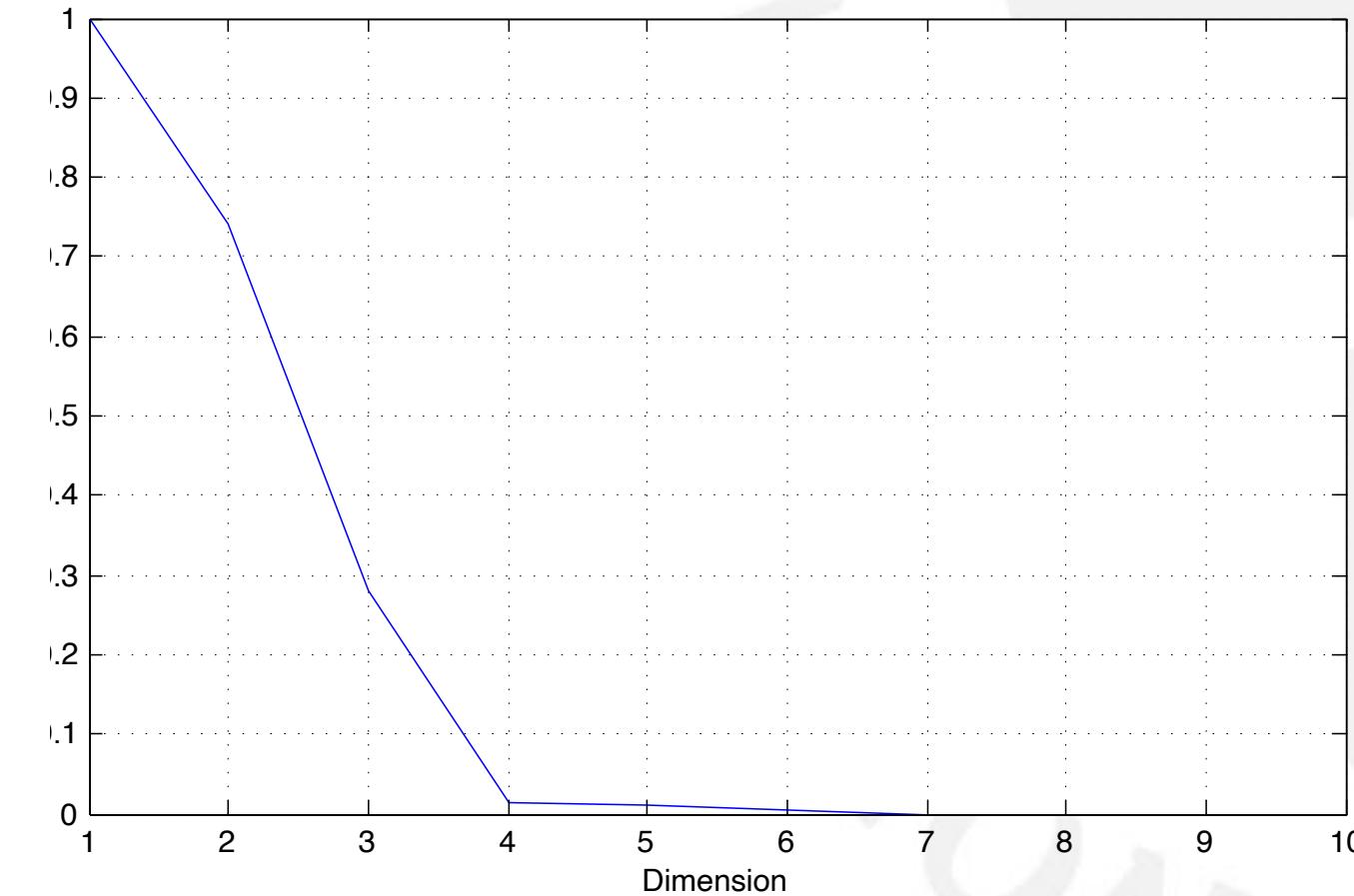
- Determine the embedding lag:
- Lag = 1?



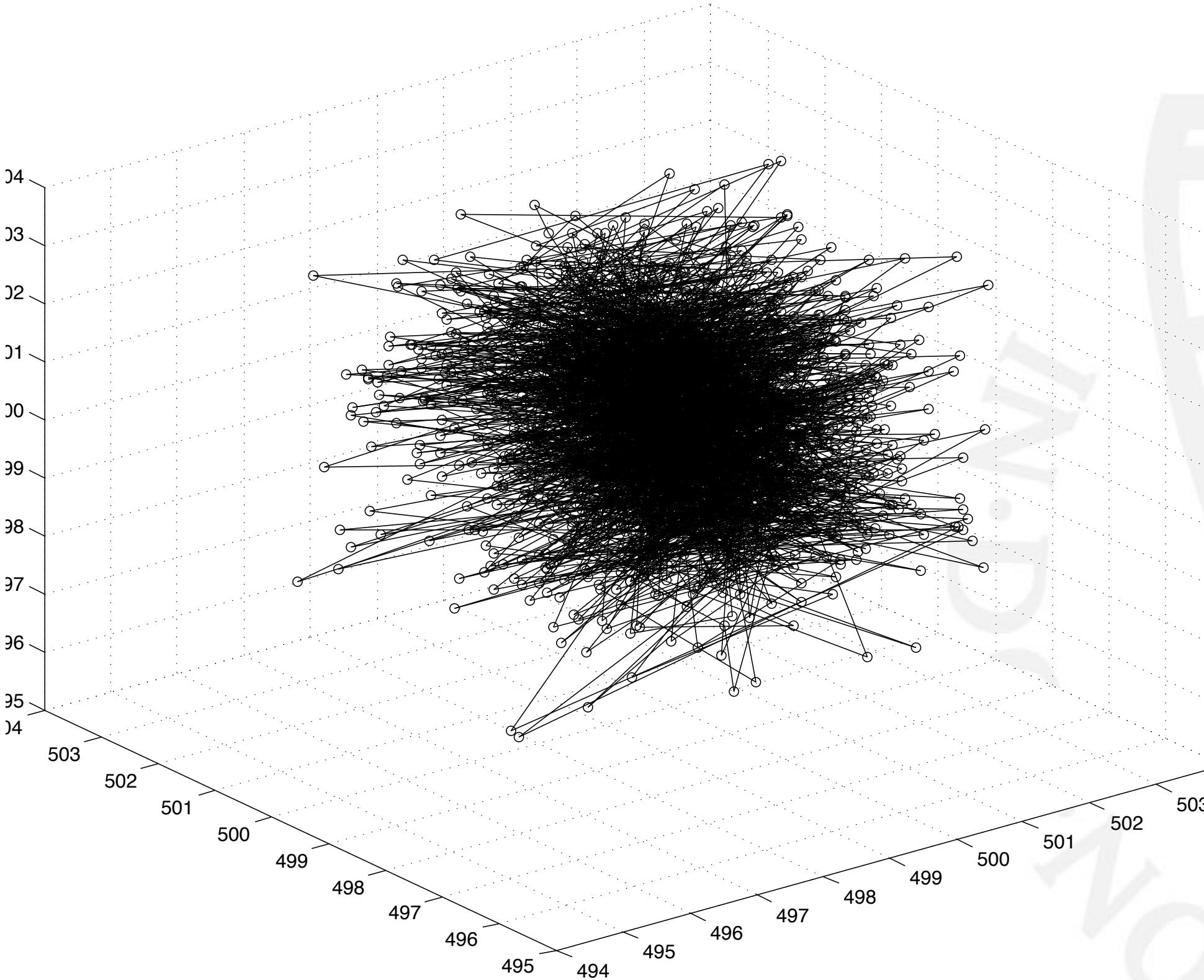
Suppose we have measured a true IID variable



- Determine the embedding dimension:
- Dimension = 4,5,6,7?



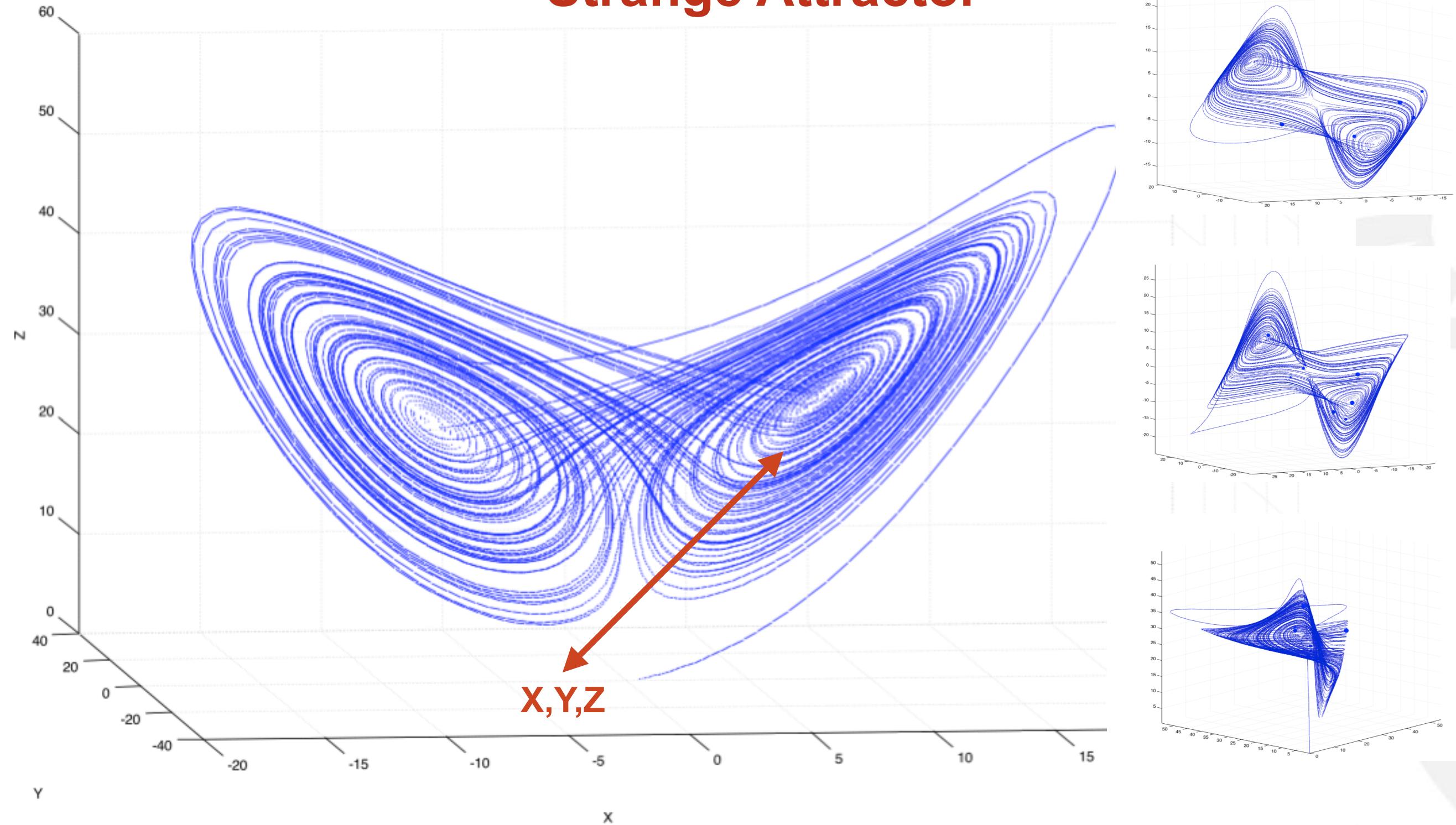
Suppose we have measured a true IID variable



Not so amazing?

- The reconstructed attractor is '*Topologically equivalent*' not exactly the same!!! (compare to random cloud of points)
The exact lag is not that important, it is just a way to optimize the reconstruction
- If you are working with 'real' data from psychological experiments you will find that the dimensionality needed to describe the system is usually 10 dimensions or higher... No visual inspection anymore!
- Solution: Quantify the dynamic behaviour of the system in state space in terms of periodicity, randomness, etc. This remains similar to the original dynamics even if the attractor is not reconstructed exactly the same way (the reconstructed attractor is still much more constrained than all the states theoretically possible).
- (Cross) Recurrence Quantification Analysis!

Lorenz system – X,Y,Z State space Strange Attractor



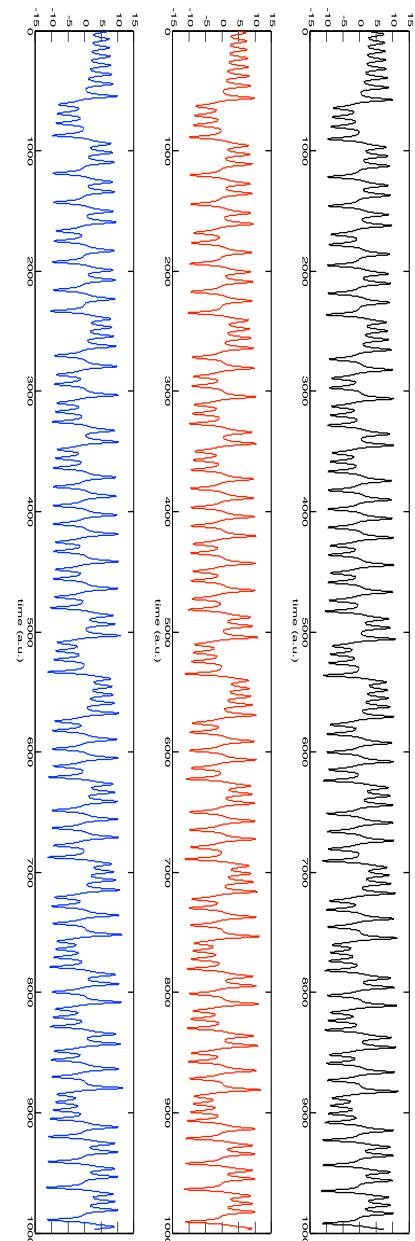
Topological Equivalence (~Homeomorphic)



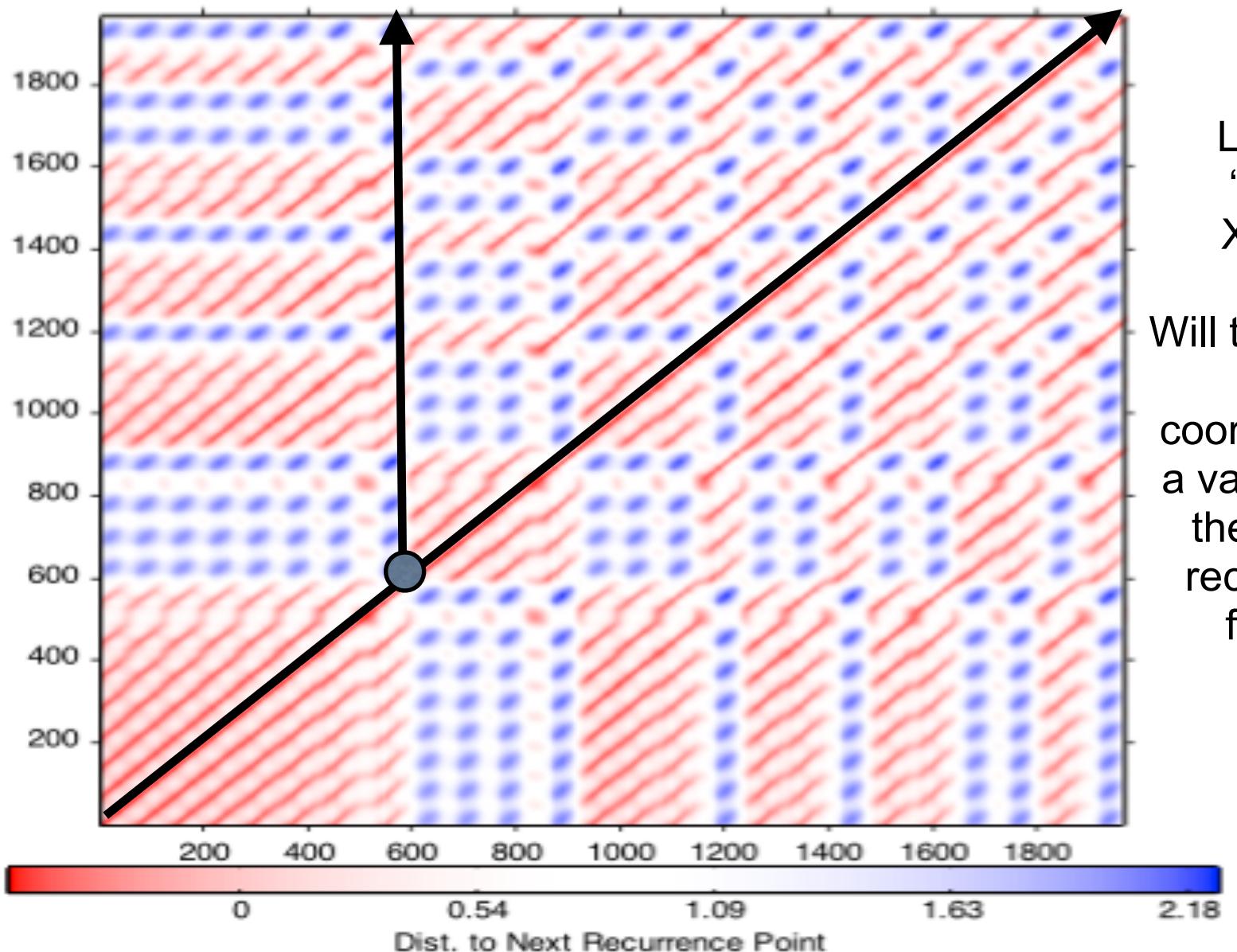
To which of these “stars”



Recurrence Quantification

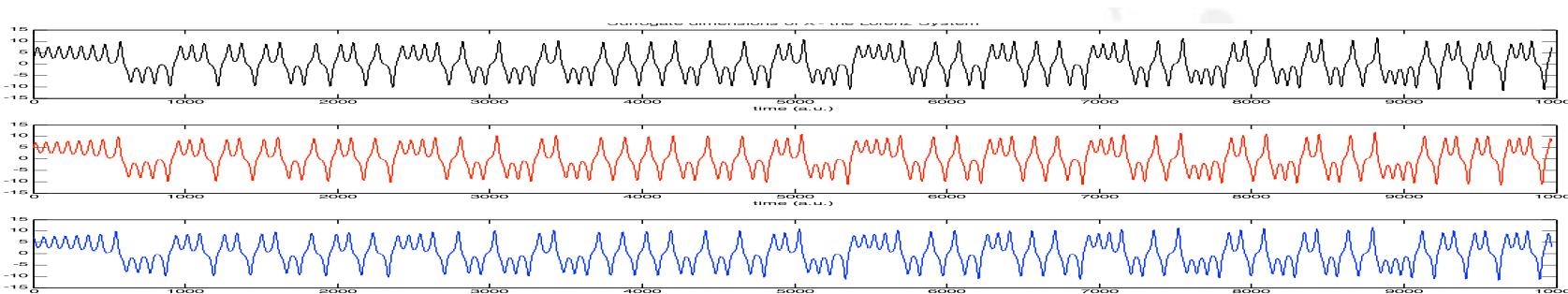


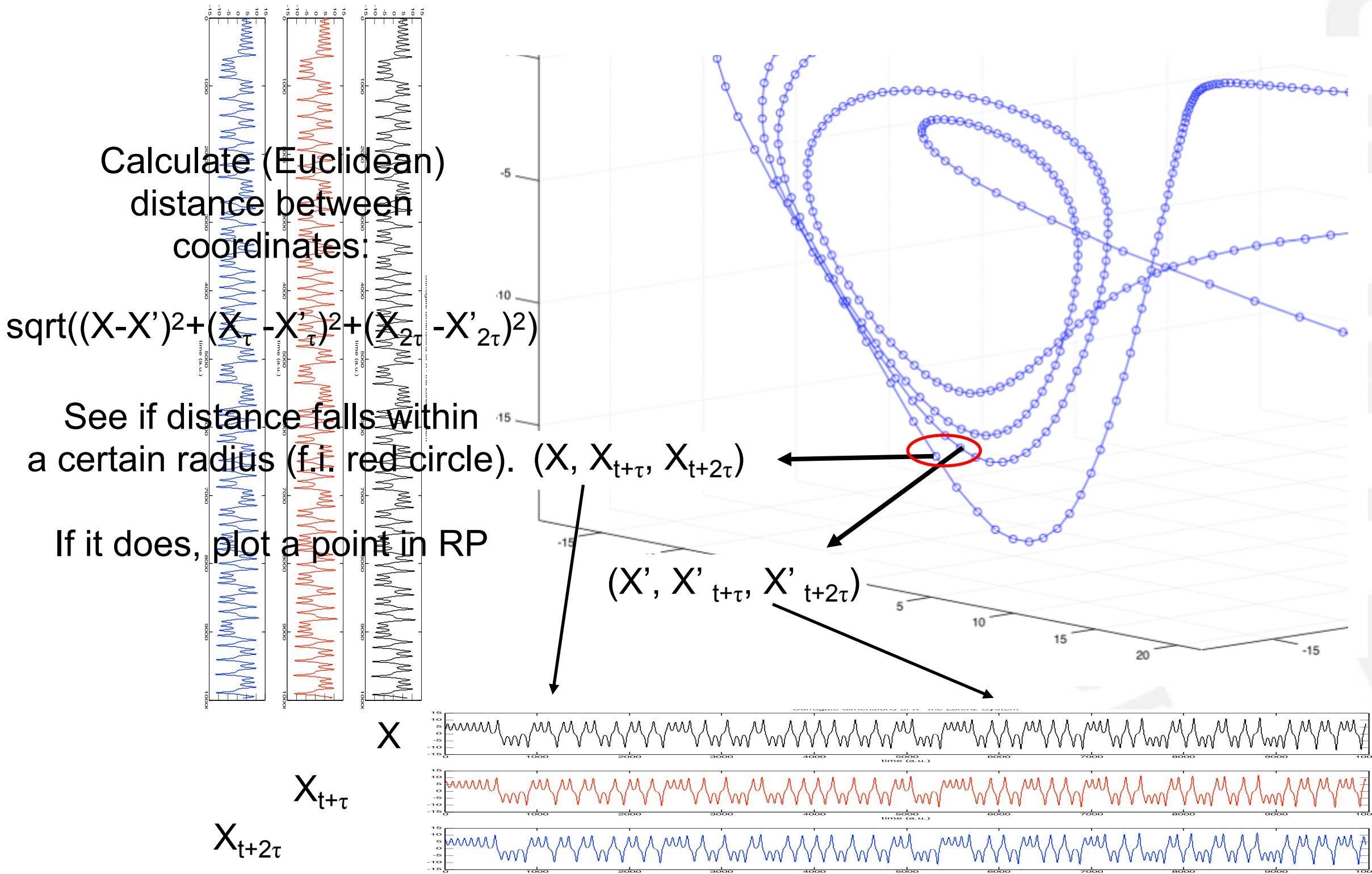
X
 $X_{t+\tau}$
 $X_{t+2\tau}$

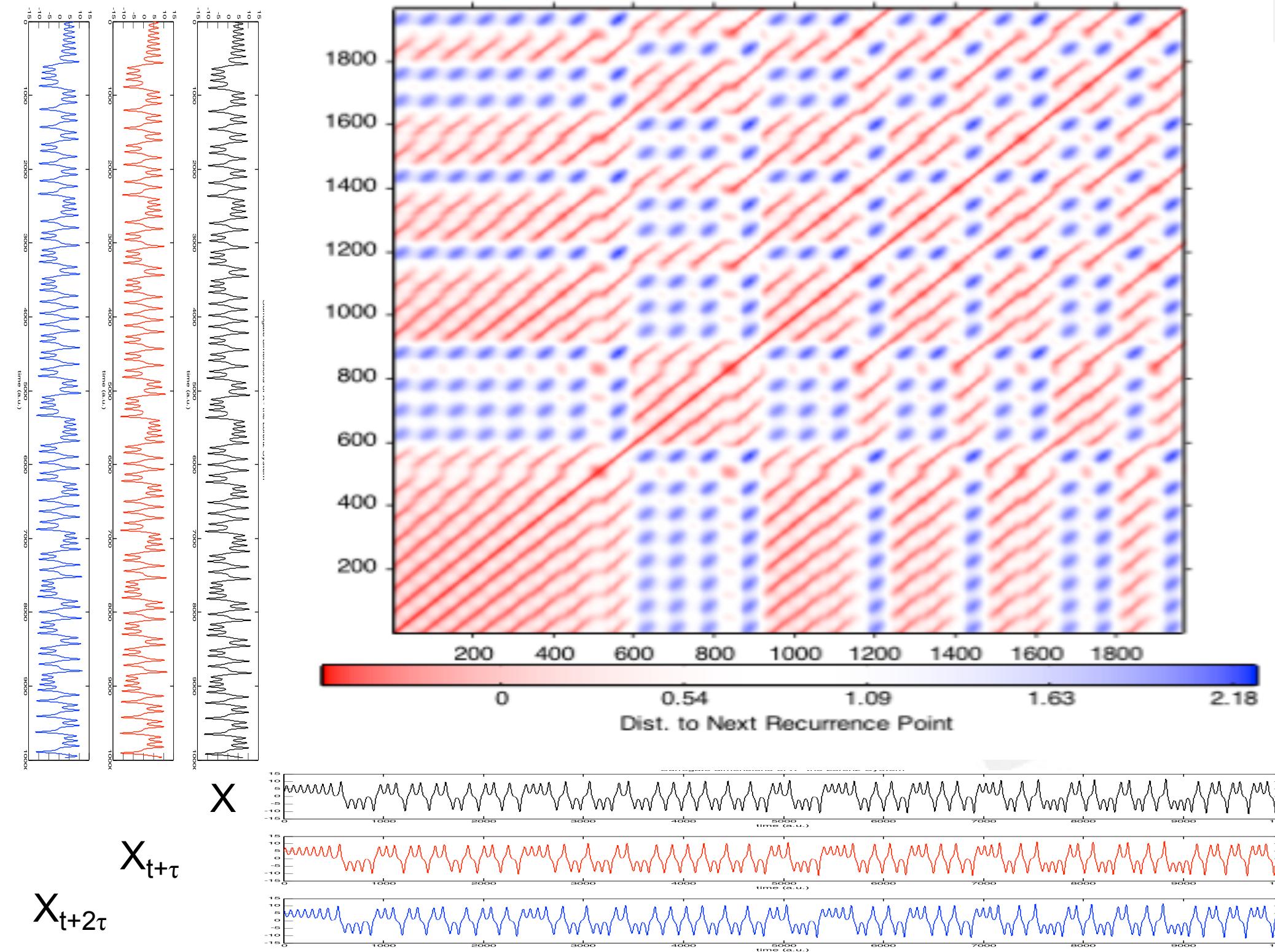


Looking
“up” at
 $X(600)$:

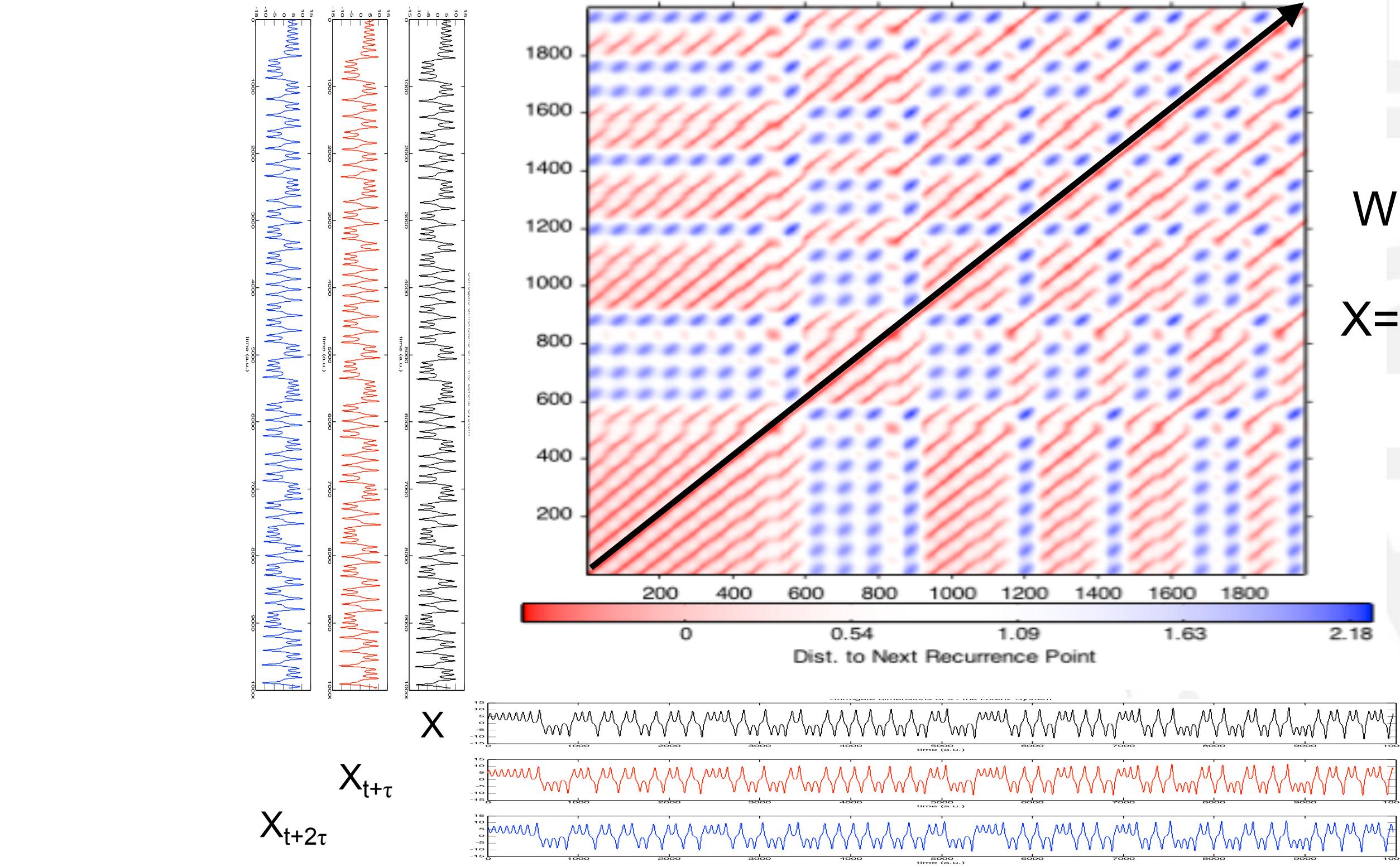
Will the current
 X, Y, Z
coordinate (or
a value within
the radius)
recur in the
future?

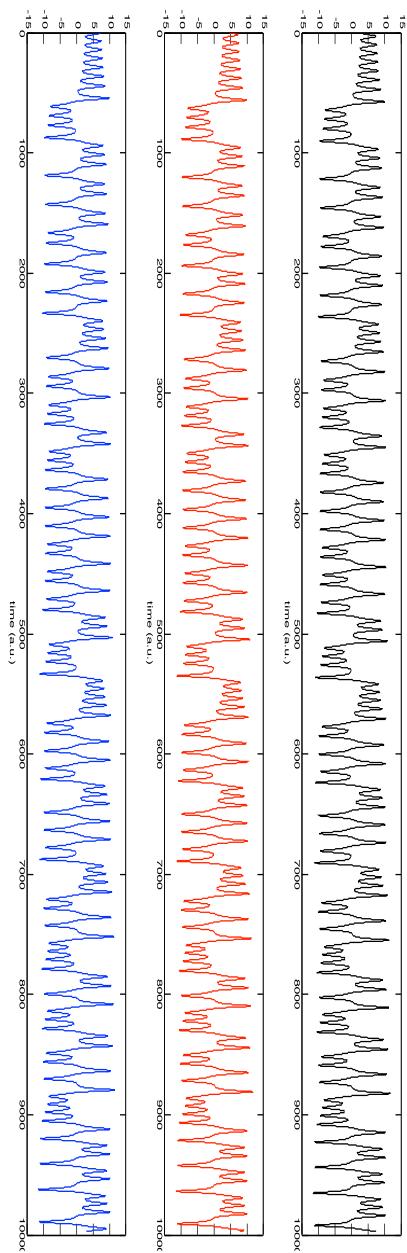






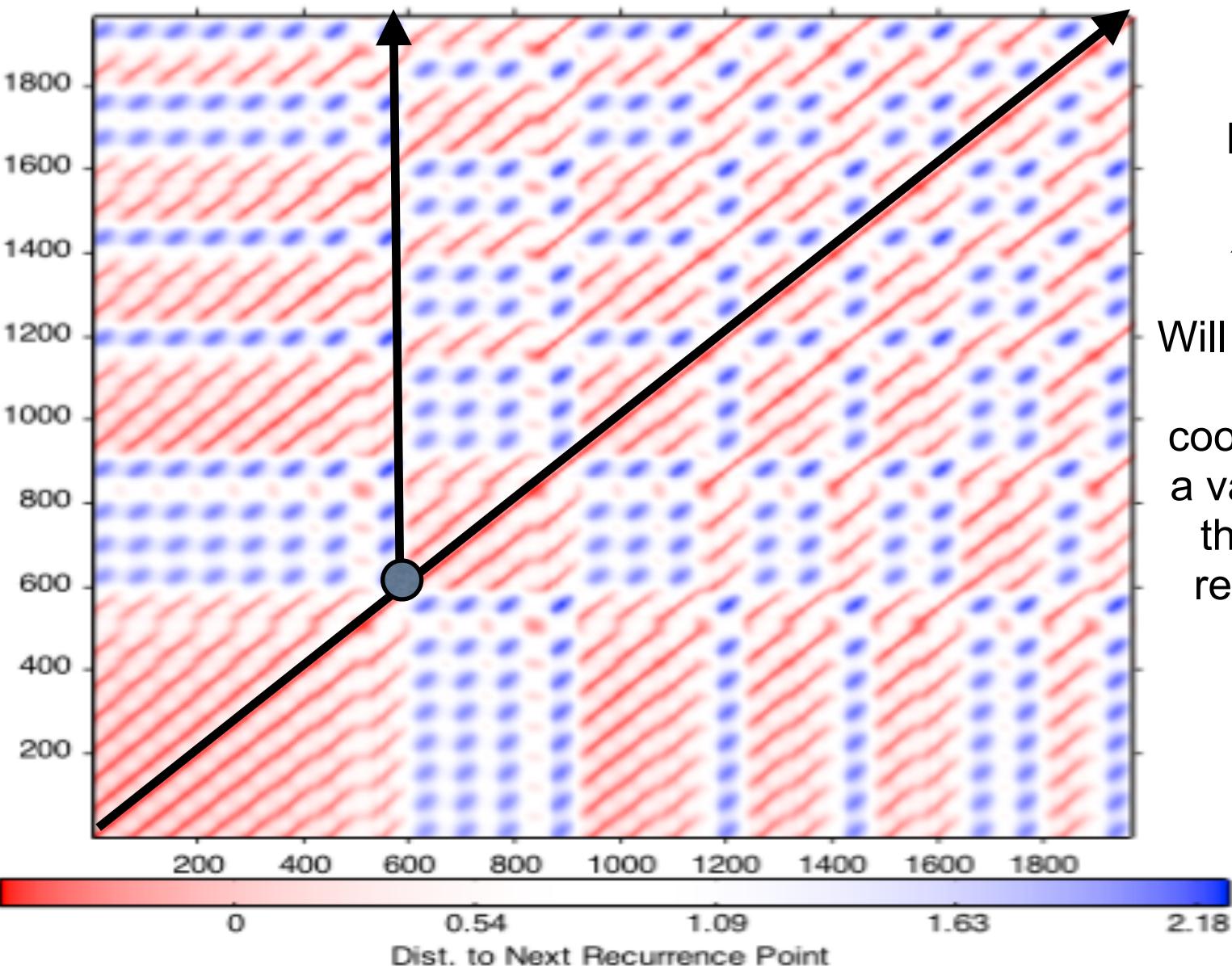
Where
is
 $X=X(t)$?



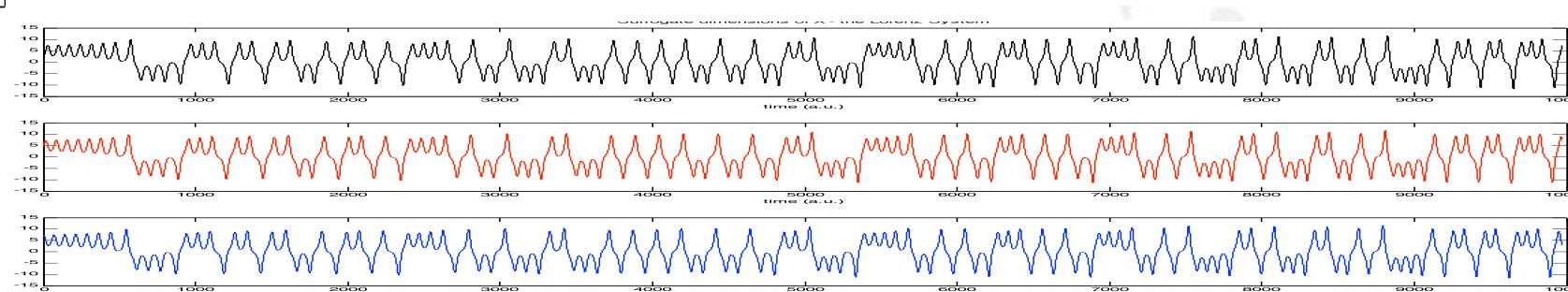


$X_{t+\tau}$

$X_{t+2\tau}$



Looking
“up” at
 $X(600)$:
Will the current
 X, Y, Z
coordinate (or
a value within
the radius)
recur in the
future?



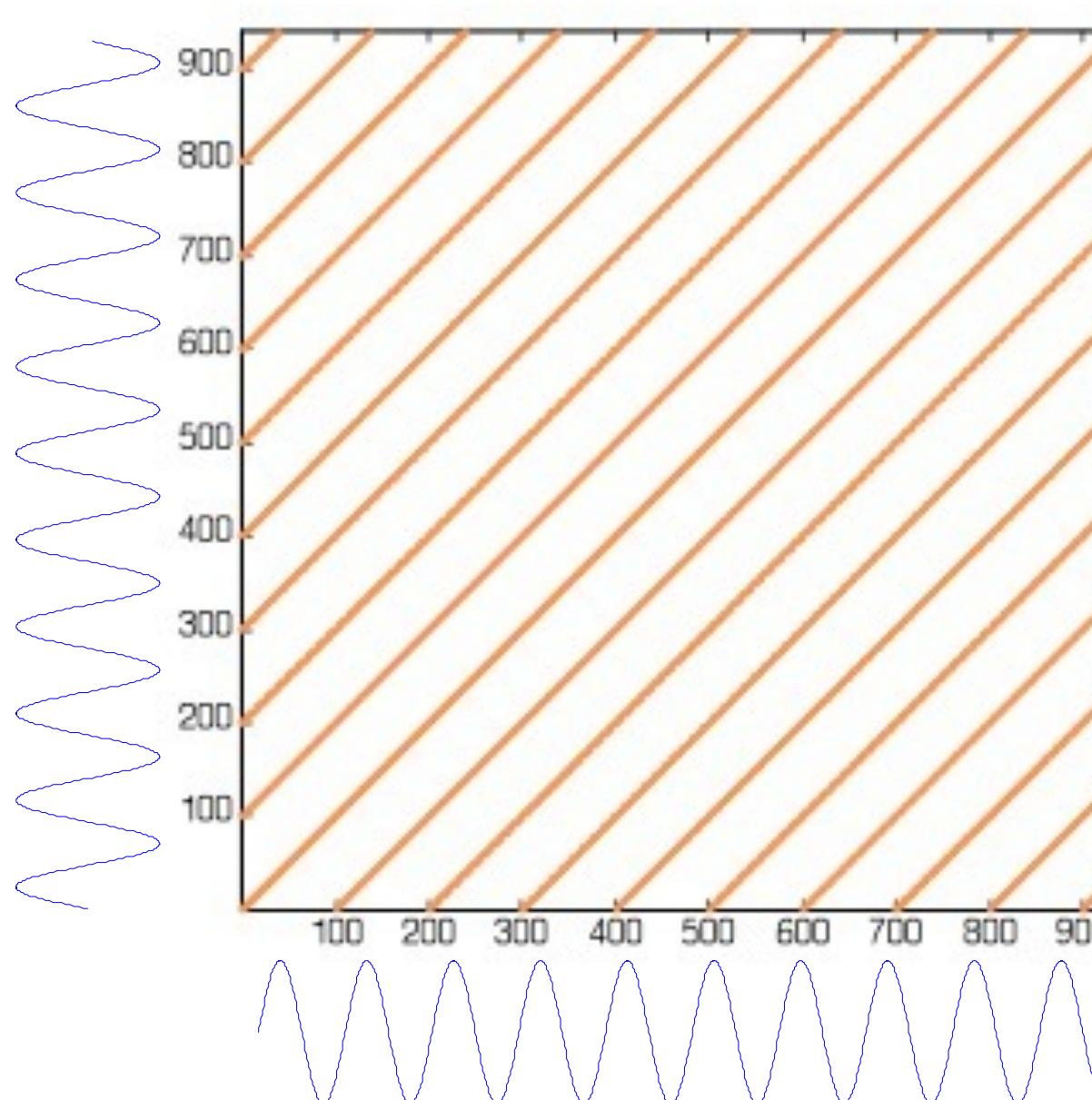
Quantifying Recurrence

Shockley 2007

%REC =

Sine

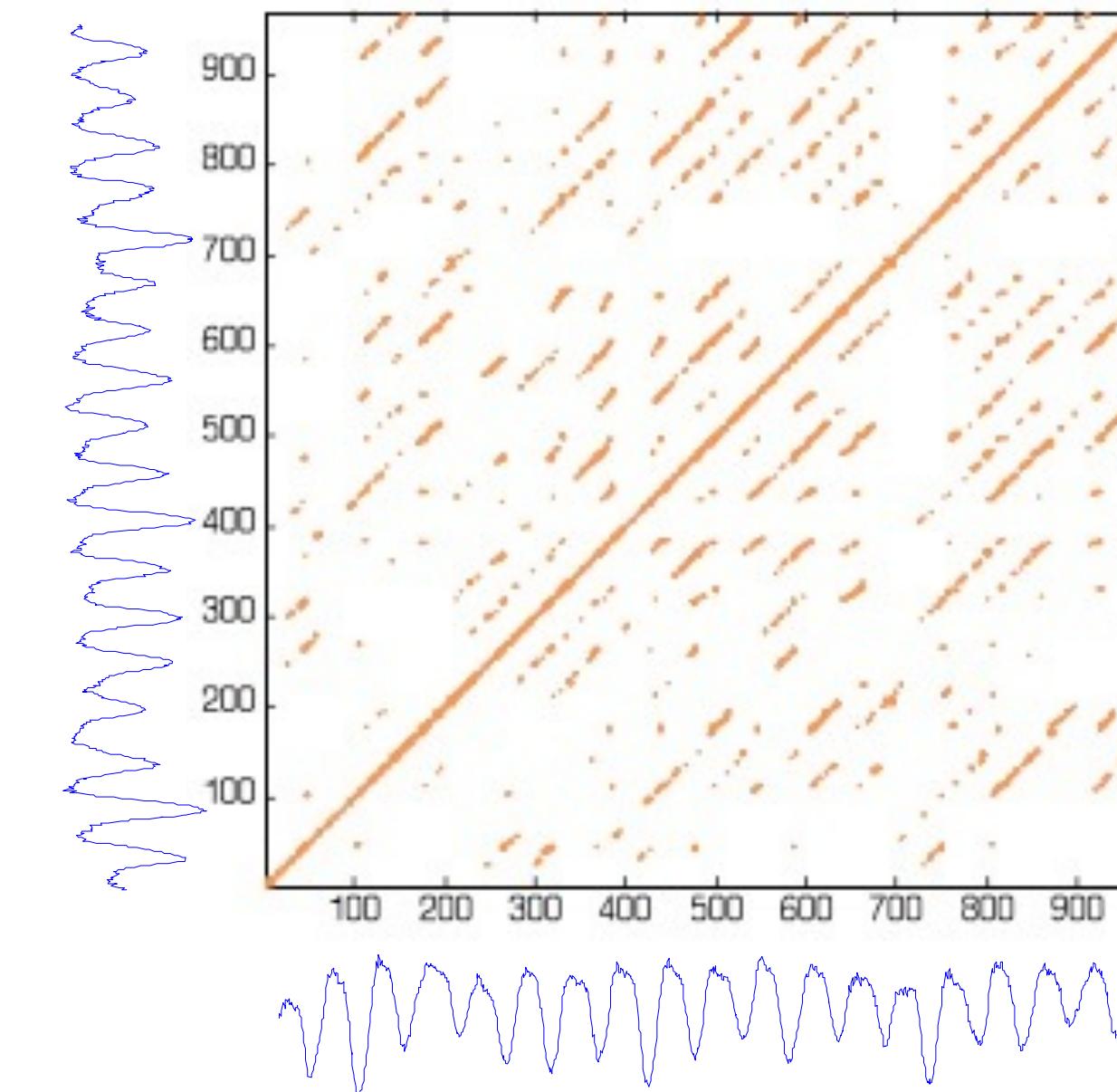
%REC = 2.9



Number of recurrent points
Total number of locations

$\times 100$

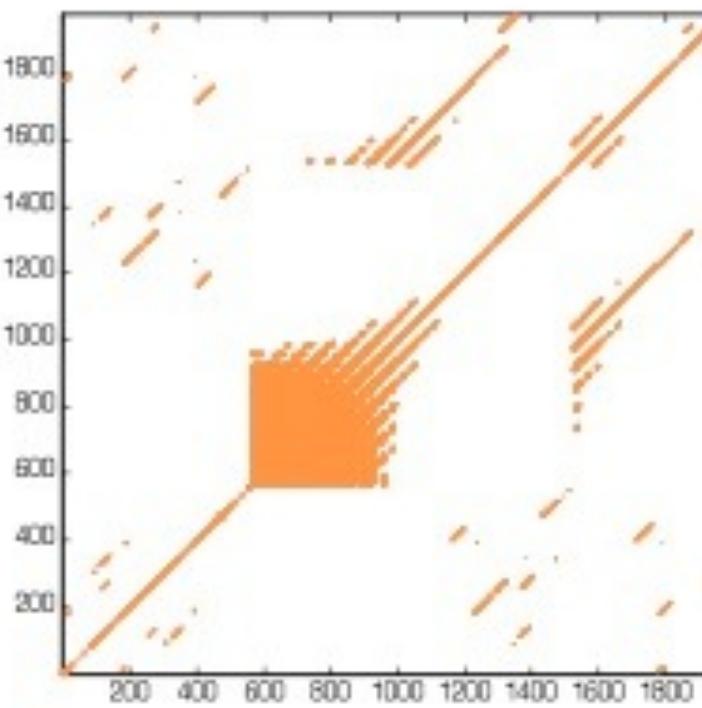
Limb oscillation to a metronome
%REC = .72



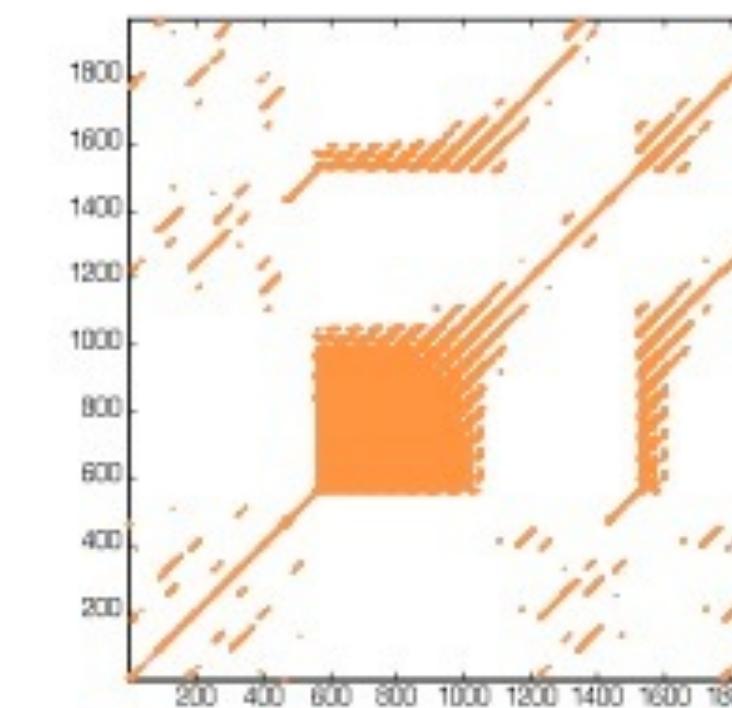
- Note that %REC is the number of points in phase space that recur, relative to all possible points that could recur. It is influenced by the radius you choose!
- When comparing groups or subjects: keep %REC constant.

Note how the recurrence plot changes with changes in radius

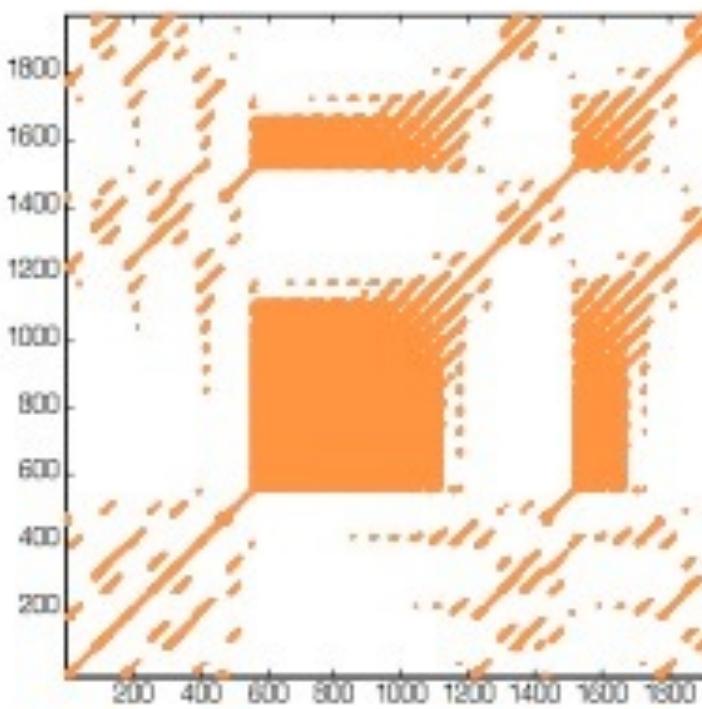
Radius = 3



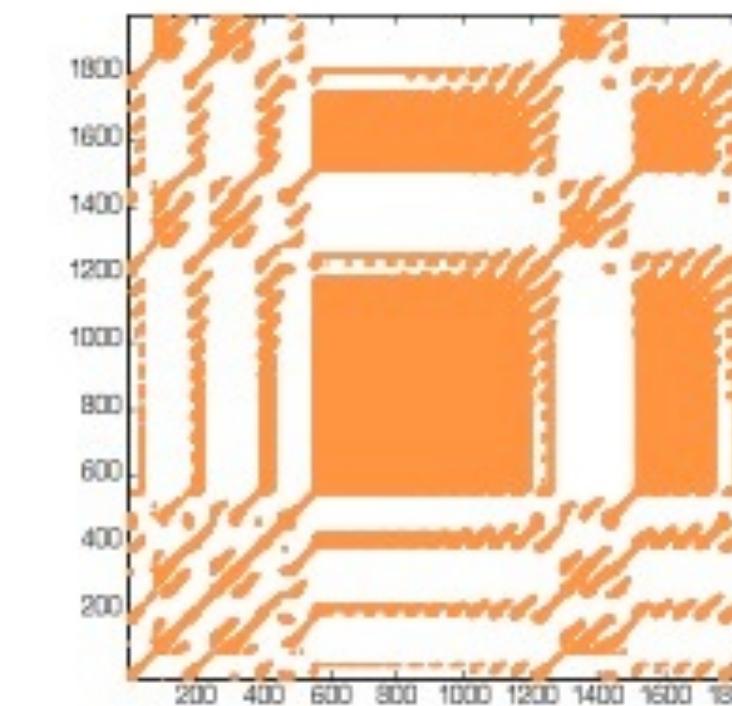
Radius = 5



Radius=10

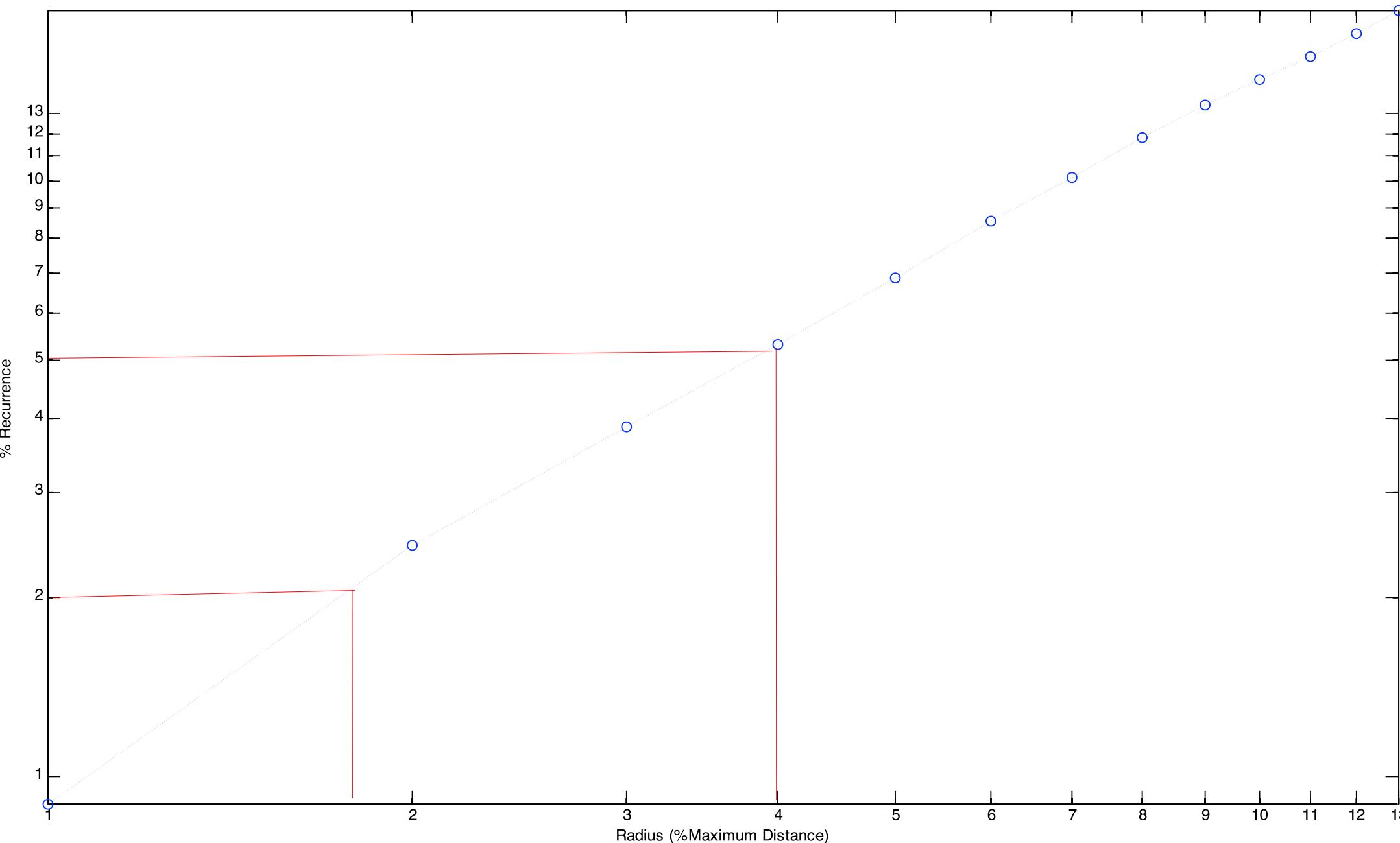


Radius=20



Shockley 2007

Is there a prescription for picking your radius?



- Take a radius size where the relationship is linear, between 2% - 5% for behavioural data.
- The exact value of the radius is not important, just keep it constant when comparing different time series

%DETERMINISM

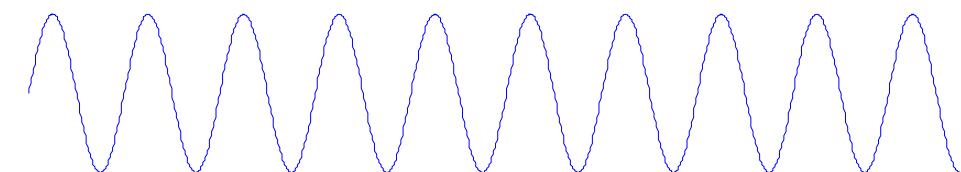
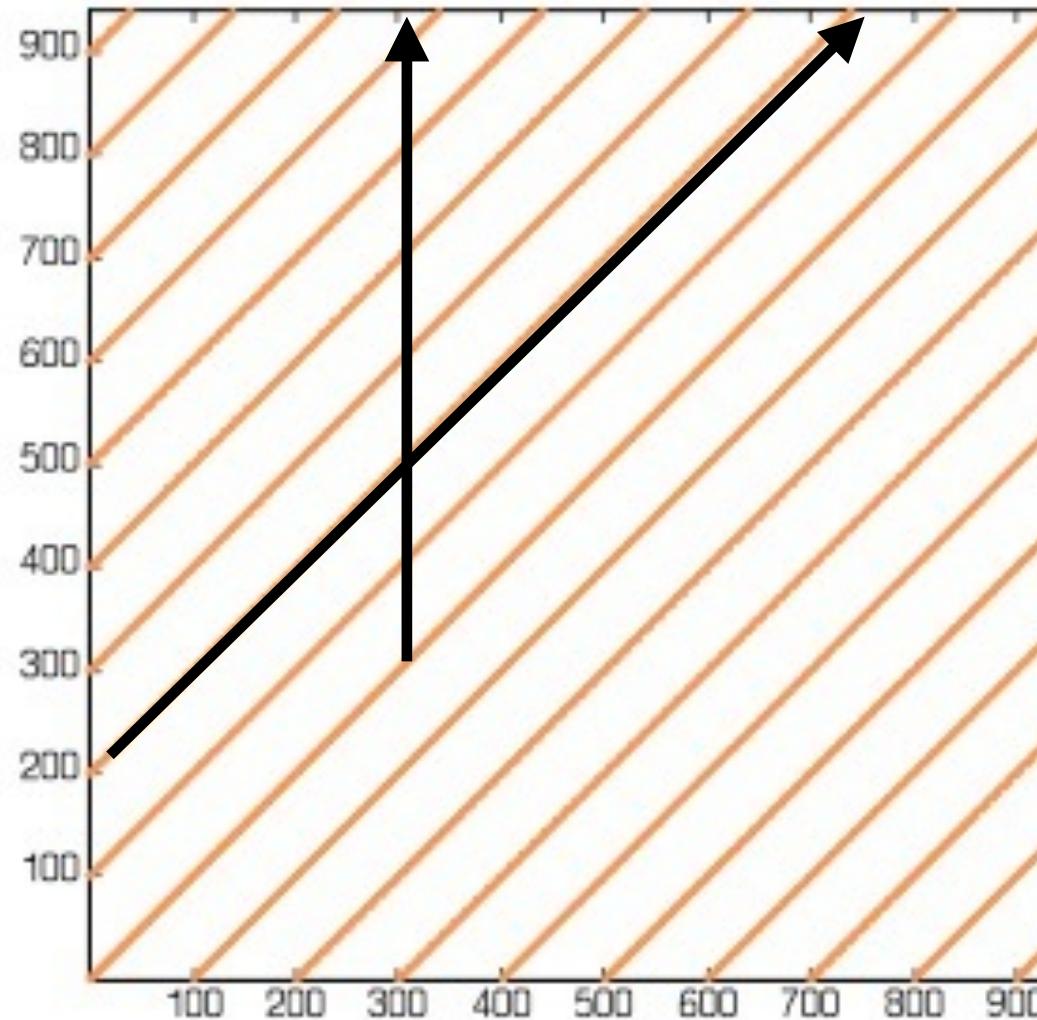
Indexes how “patterned” the data are.

Does the system return to the same region of phase space for a longer period of time?

$$\%DET = \frac{\text{Number of recurrent points forming diagonal line}}{\text{Total recurrent points}} \times 100$$

Sine

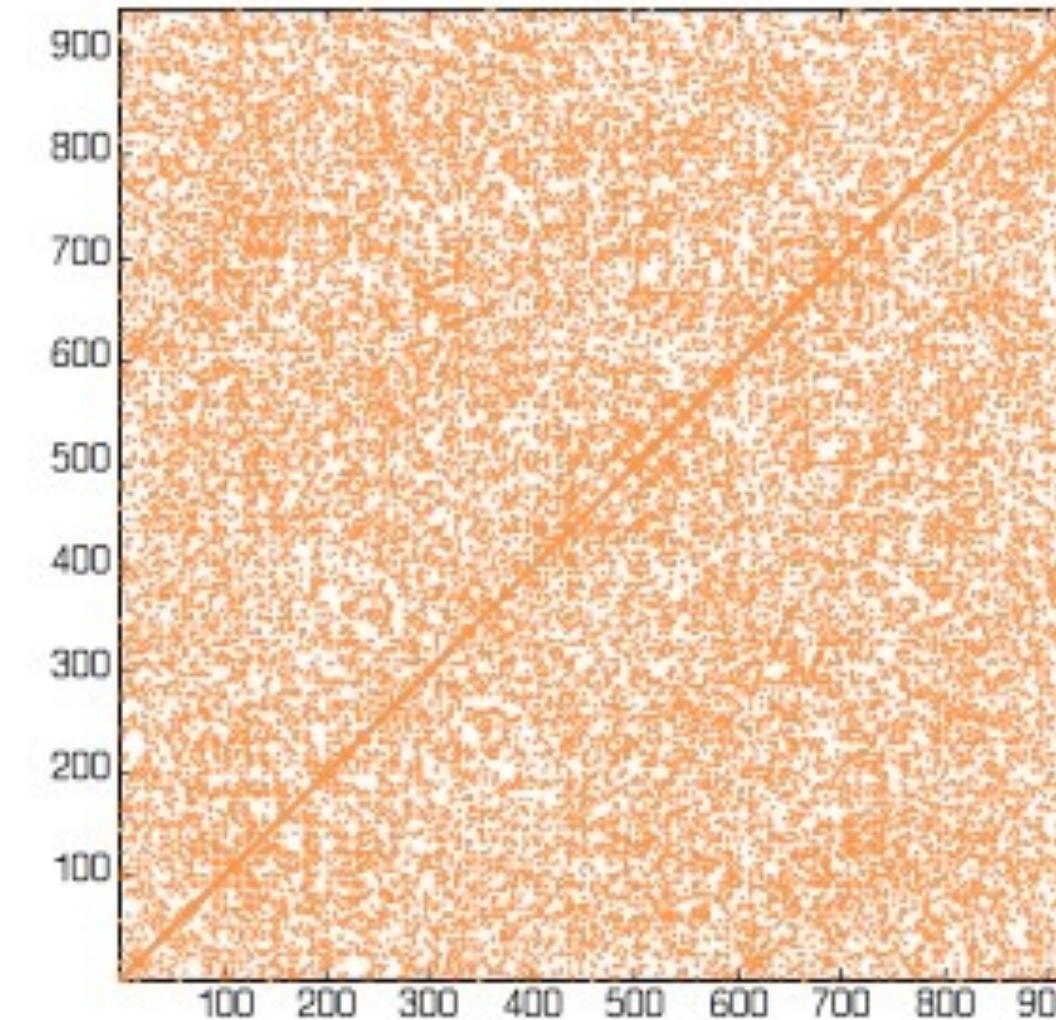
$$\begin{aligned}\%REC &= 2.9 \\ \%DET &= 99.8\end{aligned}$$



51

White Noise

$$\begin{aligned}\%REC &= 2.9 \\ \%DET &= 5.4\end{aligned}$$



MAXLINE

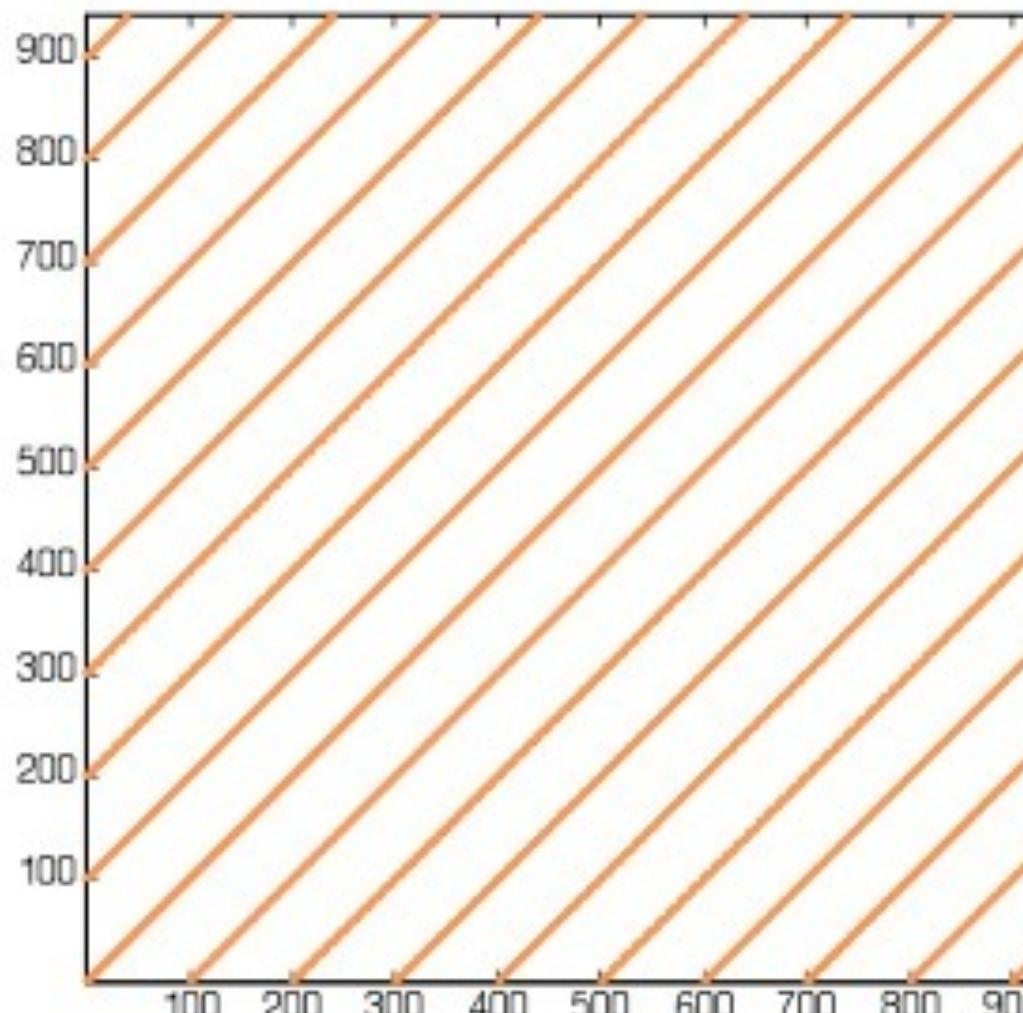
How long the system can maintain a recurring pattern ~ “Stability”

MAXLINE = The longest sequence of recurring points

Sine

%REC = 2.9

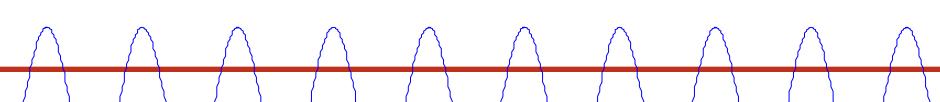
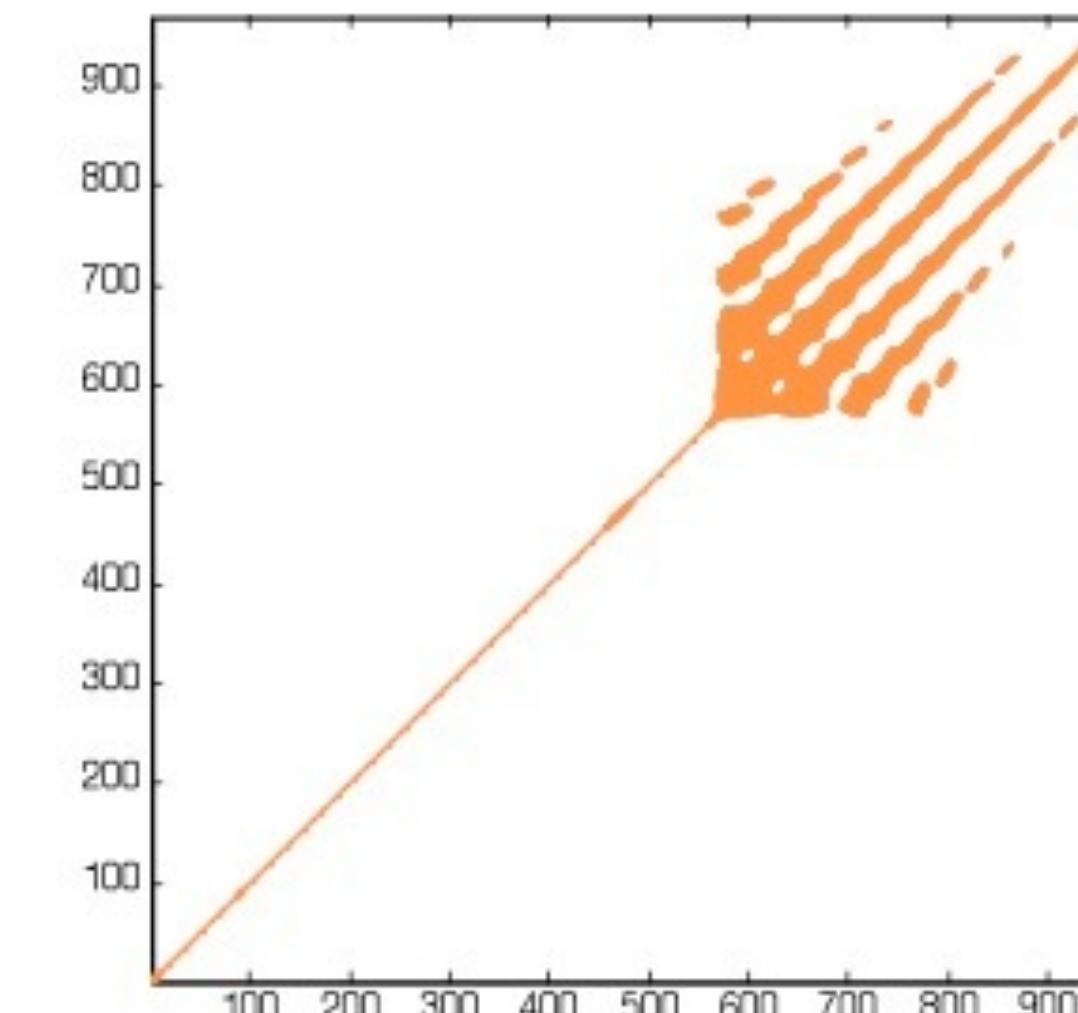
MAXLINE = 938



Lorenz

%REC = 2.9

MAXLINE = 410



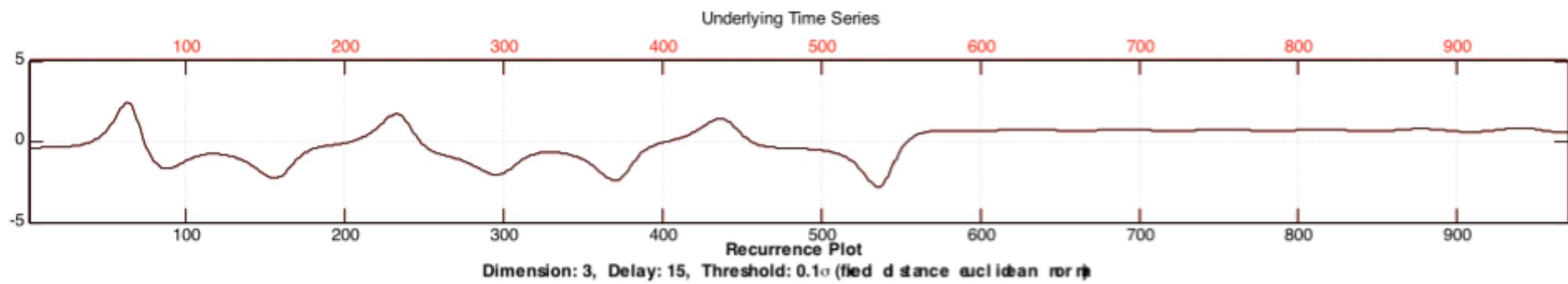
1/maxline = Divergence (Thought to be an estimate of largest Lyapunov exponent)

RQA measures

- %REC or RR (recurrence rate)
- %DET (is the data from a deterministic process or random?)
- MAXLINE (maximal diagonal line length)
- DIV (divergence, $1/\text{maxline}$, suggested estimate of largest Lyapunov exponent)
- Average LINE (average diagonal line length)
- ENTROPY (complexity of deterministic structure)
- TREND (is the data stationary?)
- %LAM (laminarity, points on vertical lines, connected to Laminar phases)
- TT (Trapping Time, average length of vertical lines: How long the system stays in a specific state)
- Create your own...

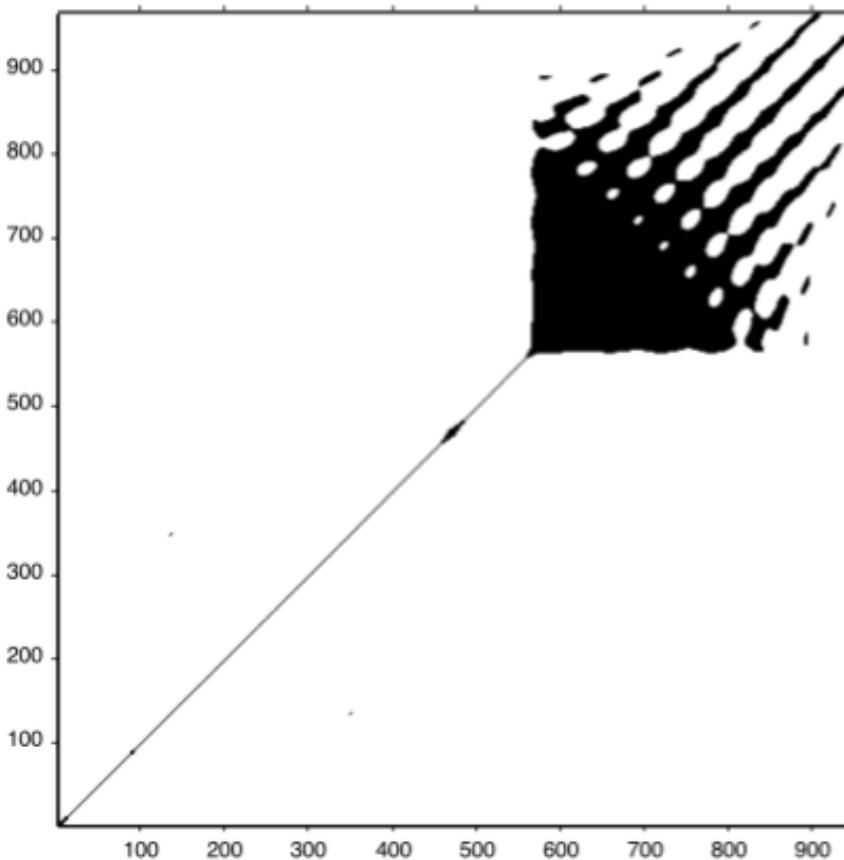


How to decide these values have meaning?

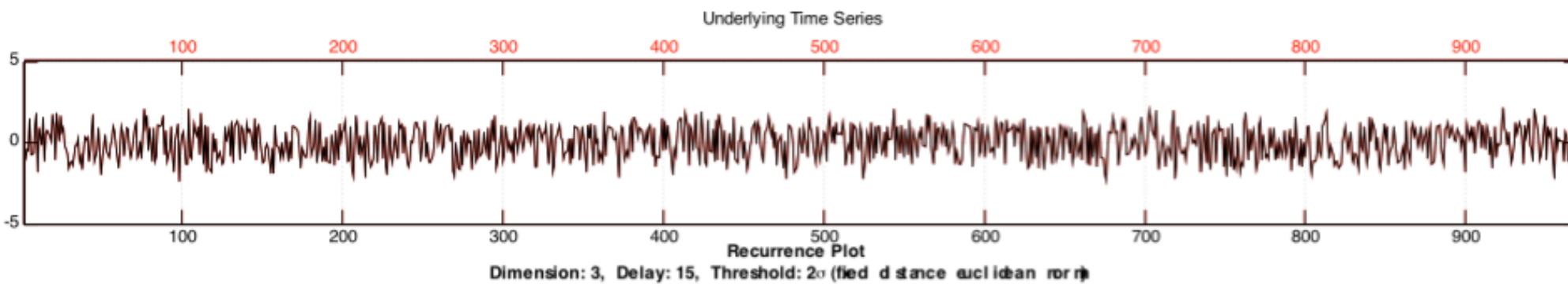


Recurrence Plot
Dimension: 3, Delay: 15, Threshold: 0.10 (Euclidean distance)

Original:
%REC = 7%
%DET = 100%
Av. LINE = 58
ENTROPY = 4.34

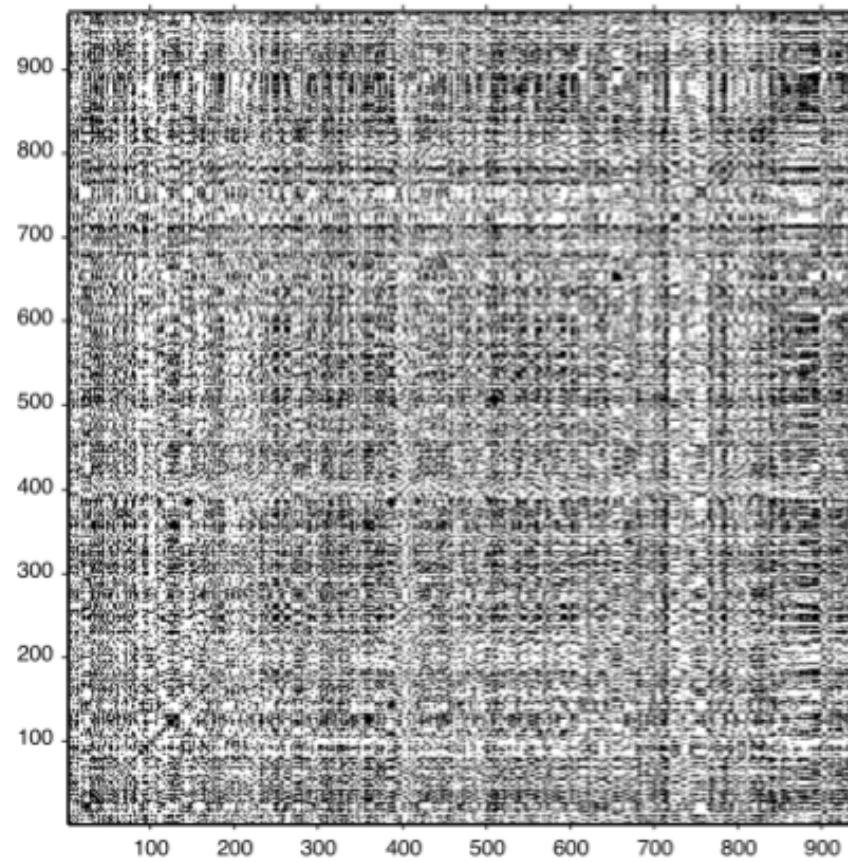


How to decide these values have meaning?



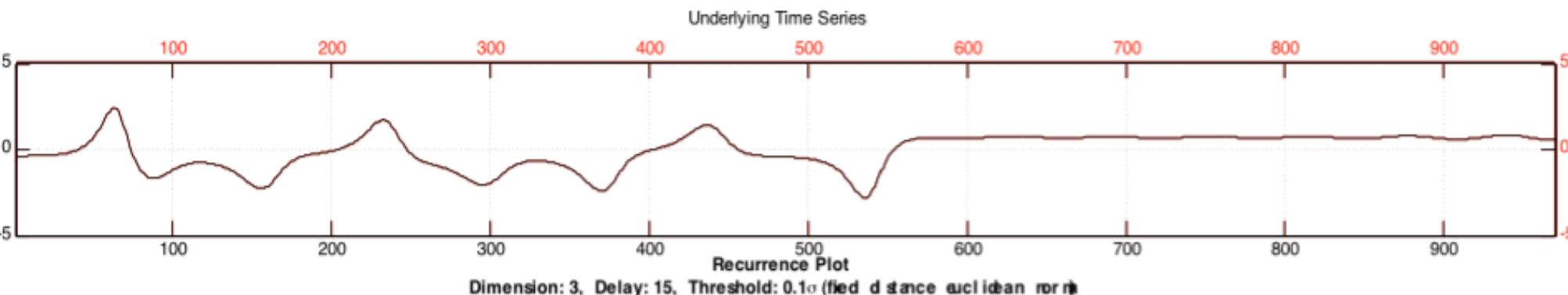
Original:
%REC = 7%
%DET = 100%
Av. LINE = 58
ENTROPY = 4.34

Shuffled:
%REC = 7%
%DET = 14%
Av. LINE = 2.1
ENTROPY = 0.25



Or use a surrogate

Recurrence Plots - Software

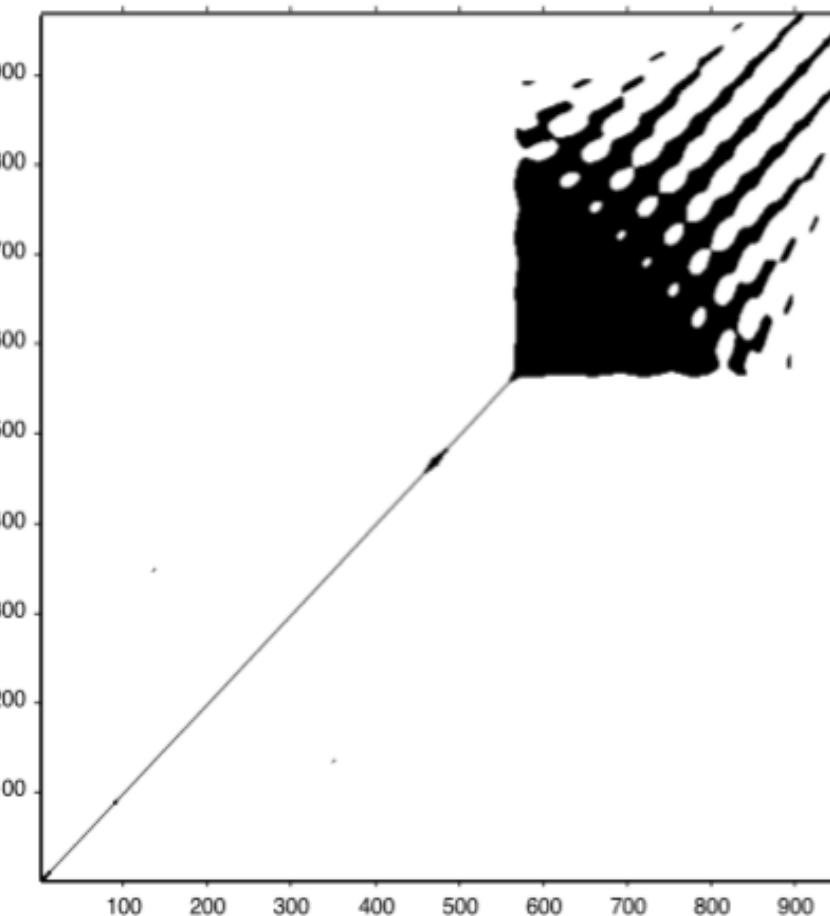


Recurrence plots come in many flavors, check:

<http://www.recurrence-plot.tk/>

By Norbert Marwan

Also links to a great Matlab Toolbox!



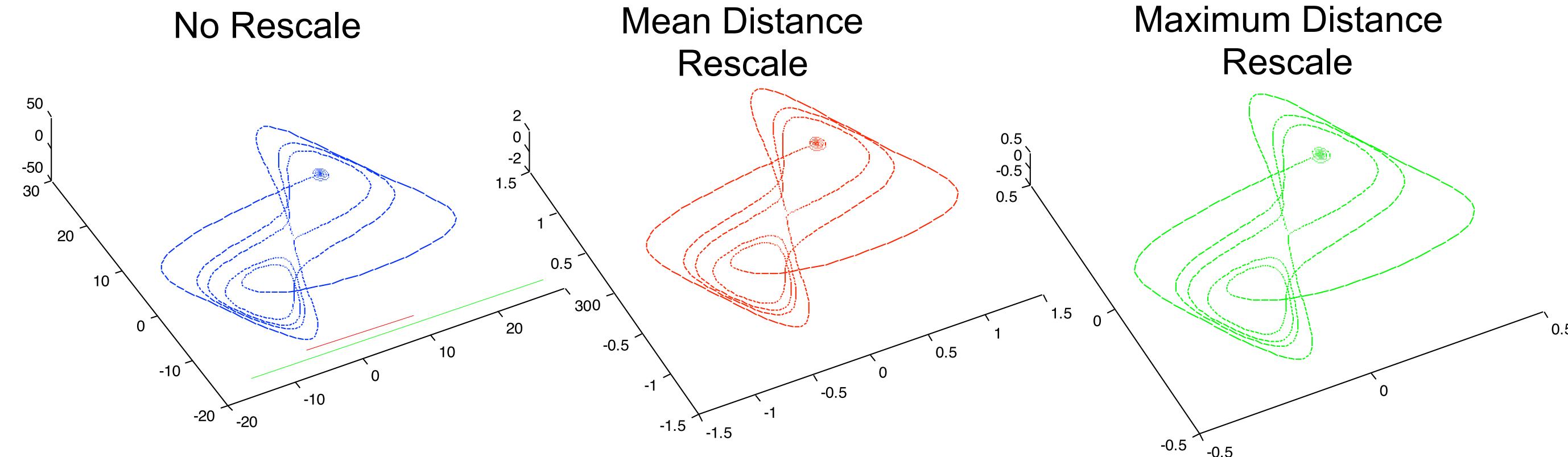
Command line software from Webber & Zbilut:

<http://homepages.luc.edu/~cwebber/>



Generally it is a good idea to re-scale your data relative to either the mean or maximum distance separating points in reconstructed phase space.

This way data is scaled to itself which allows comparisons across data sets.



Maximum distance re-scaling recommended

Webber, C.L., Jr., & Zbilut, J.P. (2005). Recurrence quantification analysis of nonlinear dynamical systems. In: *Tutorials in contemporary nonlinear methods for the behavioral sciences*, (Chapter 2, pp. 26-94), M.A. Riley, G. Van Orden, eds. Retrieved June 5, 2007 <http://www.nsf.gov/sbe/bcs/pac/nmbs/nmbs.pdf>

General Recipe for Recurrence Quantification with toolbox:

- Decide which **lag** to use:

Calculate the Average Mutual Information for a range of lags (*crqa_parameters*).

Take the lag where AMI reaches its first minimum. This is the lag at which least is known about $X(t+\tau)$ given $X(t)$, so we can create surrogate dimensions which give most new information about the system.

- Decide which **embedding dimension** to use:

Calculate how many False Nearest Neighbours you loose by adding a dimension (*crqa_parameters*). Take the embedding dimension with the lowest % of nearest neighbours (or start with the dimension which gives the greatest decrease of neighbours).

- Decide which type of **rescaling** you want to use:

Plot your timeseries: Lots of outliers? Use Mean Distance. Otherwise: Max Distance.

Calculate the max distance in reconstructed phasespace, after lag and embedding are known using *max(repmat(y,emDim,emLag))*, divide by this value.

- Decide which **radius / threshold** to use:

Use *rp_plot* to show unthresholded (without radius) plots use *crqa_radius* to find a radius

- Run **RQA** (*crqa_c!*) with these parameters! Or use *crqa_rp*

- Compare to shuffled data (*shuffle, surrogates*)

Radius:
3.629

RP_N:
202508

RR:
0.05

DET:
0.999

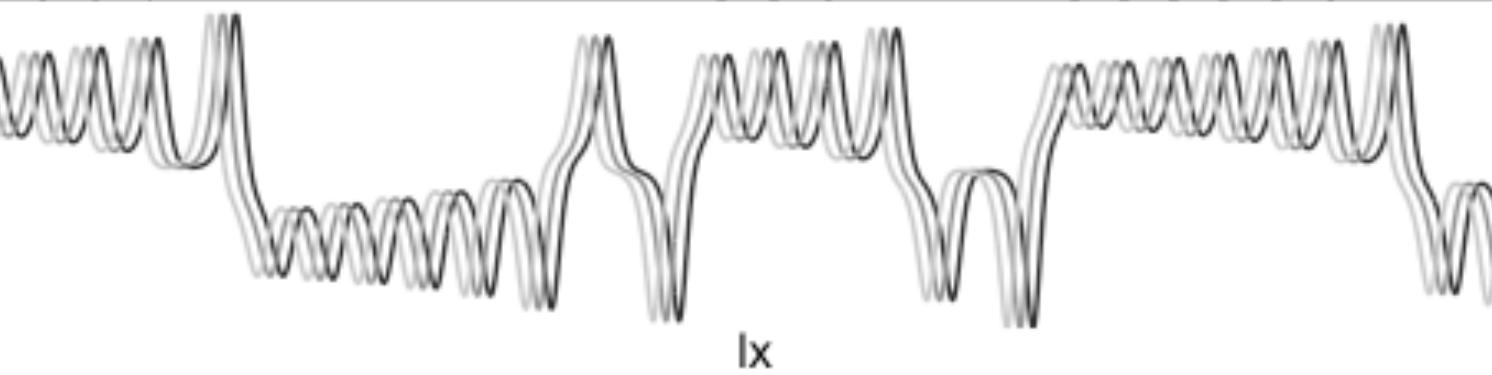
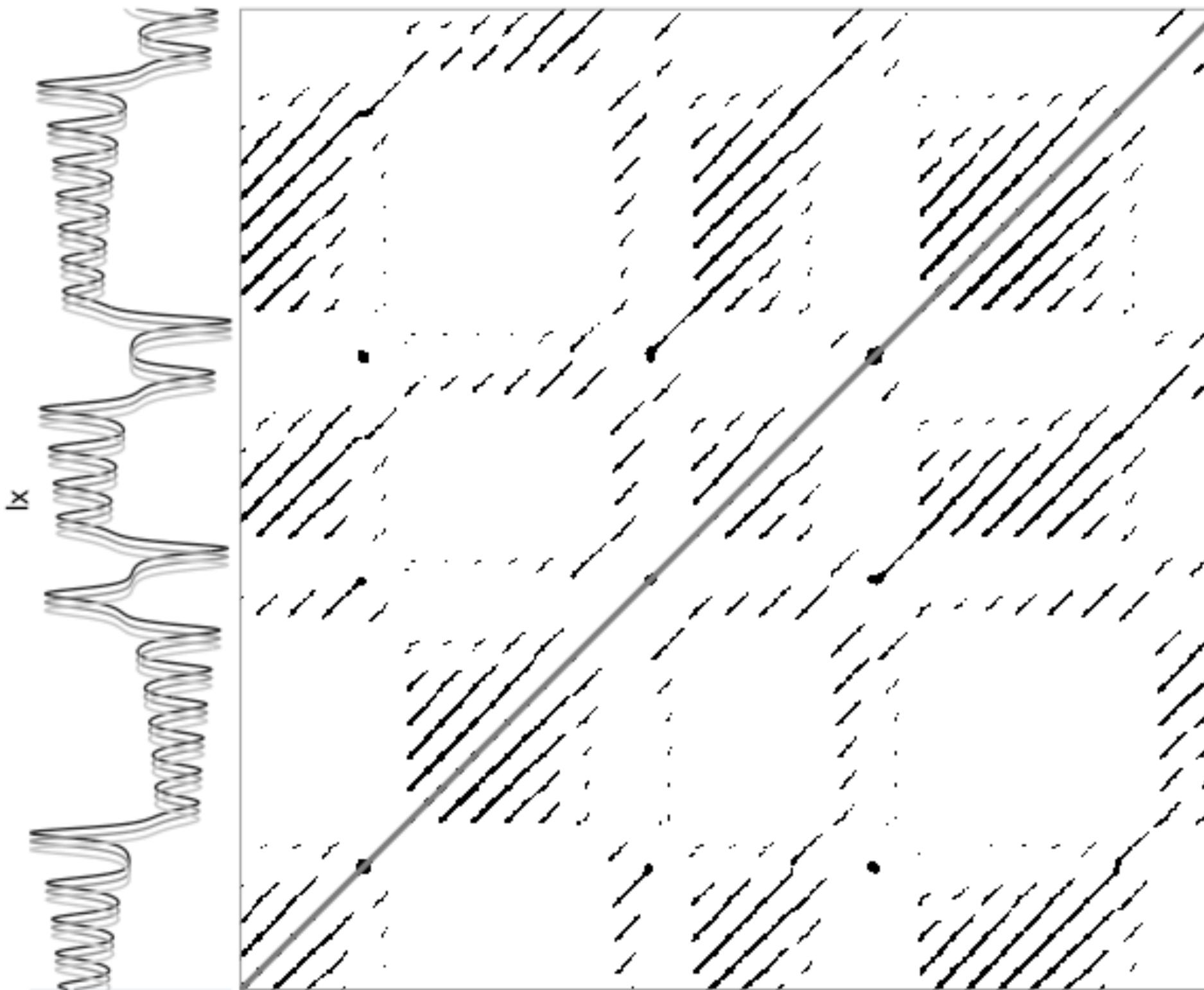
MEAN_dl:
24.689

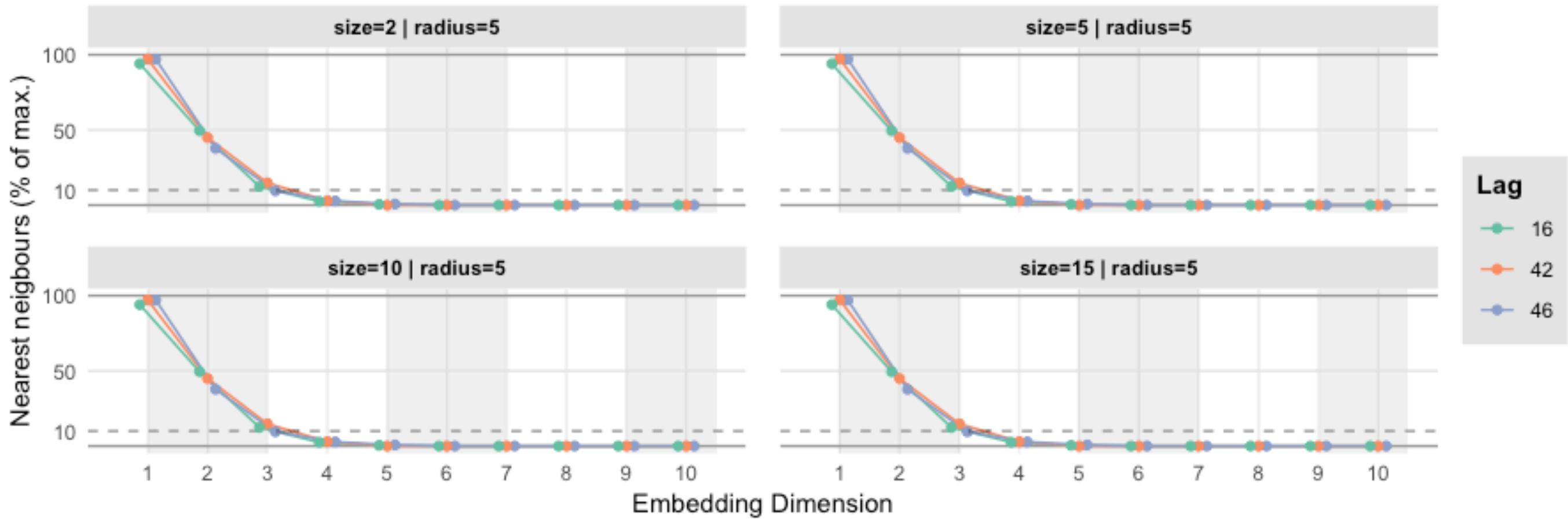
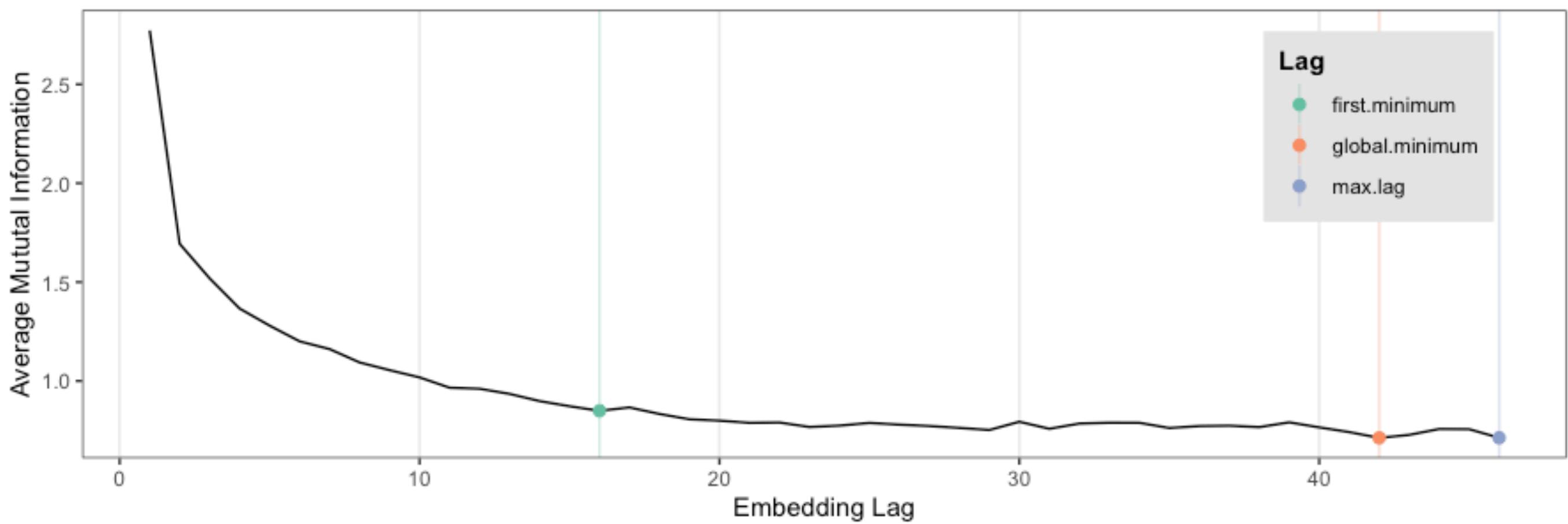
ENT_dl:
3.761

LAM_vl:
0.999

TT_vl:
9.212

ENT_vl:
2.761





Note that:

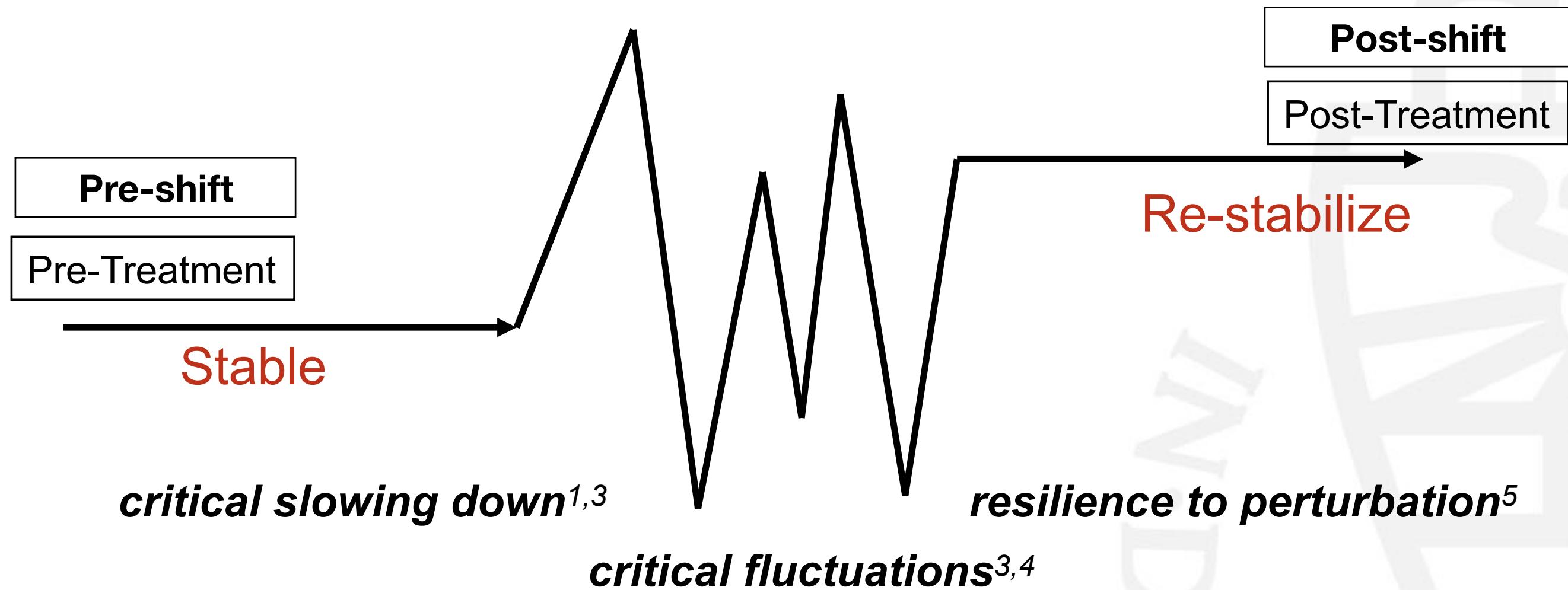
Recurrence values **will** change with changes in the parameters

The safest bet for behavioural data:

- Do recurrence calculations with one set of parameters for all of your data sets.
- Then, do this again with another set of parameters and make sure the overall results pattern the same way.
- Then, you can be sure that your results are not artefacts of your parameter selection



Period of Destabilization



- increase in recovery and switching time after perturbation
- increase in variance, autocorrelation, long-range dependence
 - increase in occurrence and diversity of unstable states
- increase in the entropy of the distribution of state occurrences

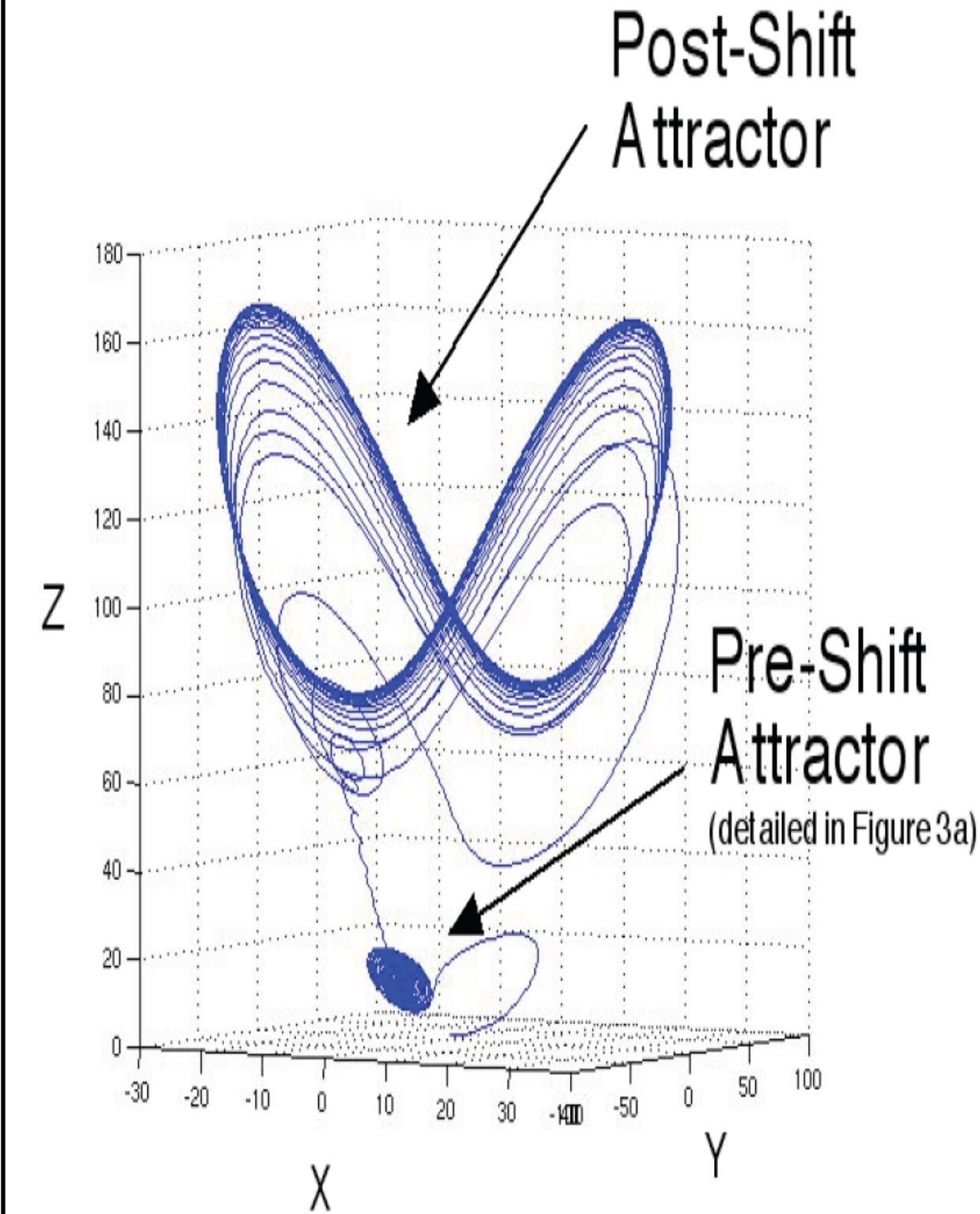
¹Scholz JP, Kelso JAS, Schöner G. (1987). Nonequilibrium phase transitions in coordinated biological motion: critical slowing down and switching time. *Physics Letters A* 123, 390–394.

²Scheffer M, Bascompte J, Brock W A, Brovkin V, Carpenter SR, Dakos V, Held H, van Nes EH, Rietkerk M, Sugihara G. (2009). Early-warning signals for critical transitions. *Nature* 461, 53–9.

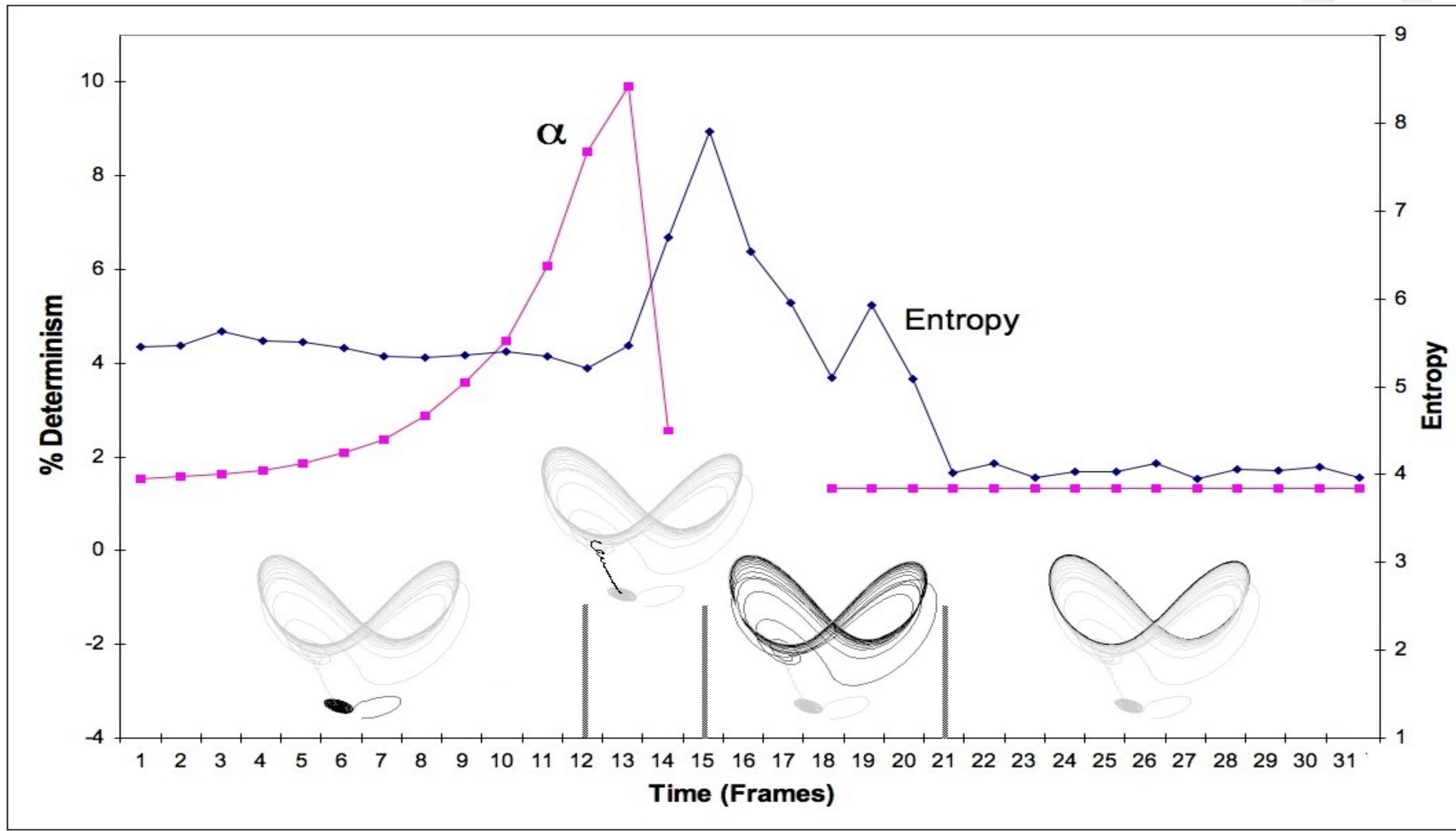
³Stephen DG, Dixon JA, Isenhower RW. (2009). Dynamics of representational change: Entropy, Action and Cognition. *JEP: Human Perception and Performance* 35, 1811–1832.

⁴Schiepek G, Strunk G. (2010). The identification of critical fluctuations and phase transitions in short term and coarse-grained time series ... *Biological cybernetics* 102, 197–207.

1. If we can reconstruct the state space of a complex dynamical system from one observable dimension....
2. If we can quantify the attractor dynamics in this state space...
3. Direct measurements of physical observables in humans should tell us something about the the dynamics of the unobservable cognitive system
4. Could we predict insight in problem solving from a phase transition in phase space reconstructed from hand movements?



Lorenz system – Transitions in phase space



Insight as a phase transition

- Stephen, D.G., Dixon, J.A., & Isenhower, R.W. (2009). Dynamics of representational change: Entropy, action, and cognition. *JEP: HPP*.

Gear Domain

- Gear systems problems
- Solve problem any way they wish
- Code strategies
 - Force-tracing
- Gear system does not move
- Force-tracing actions create information about the system
- Discovery of Alternation



Insight as a phase transition

Optotak

100 Hz sampling rate, 4 markers Velcro-ed to forefinger

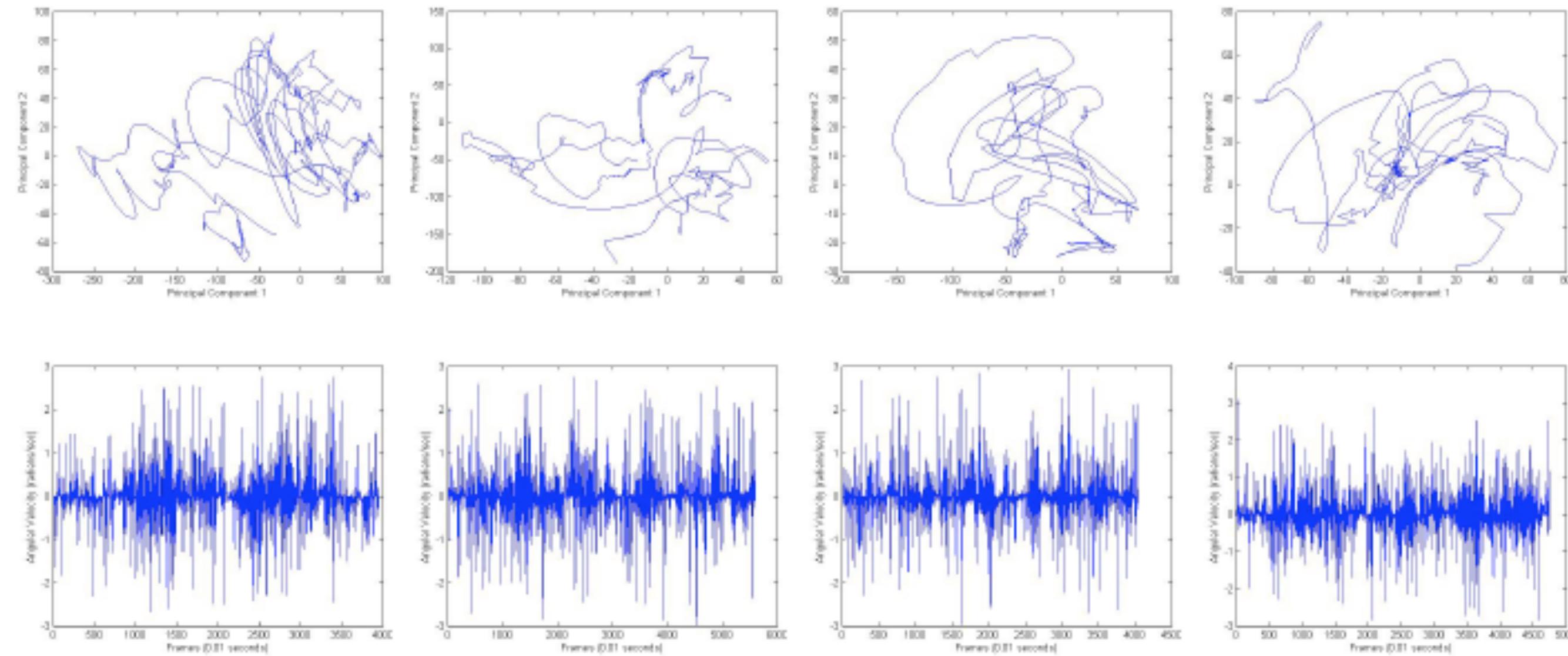
Markers emit infrared light

2 markers for
left camera



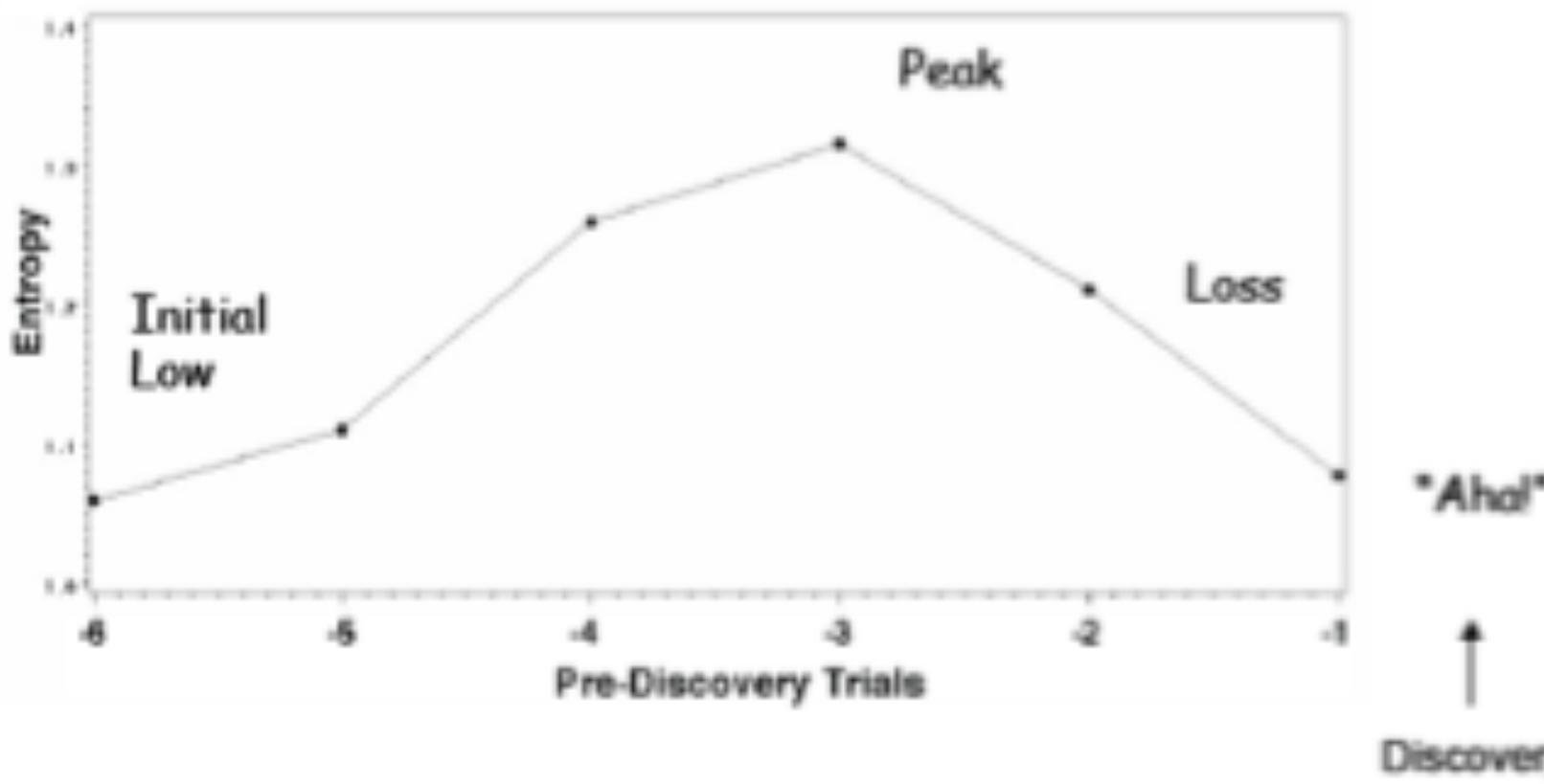
2 markers for
right camera

Angular velocity of finger movements

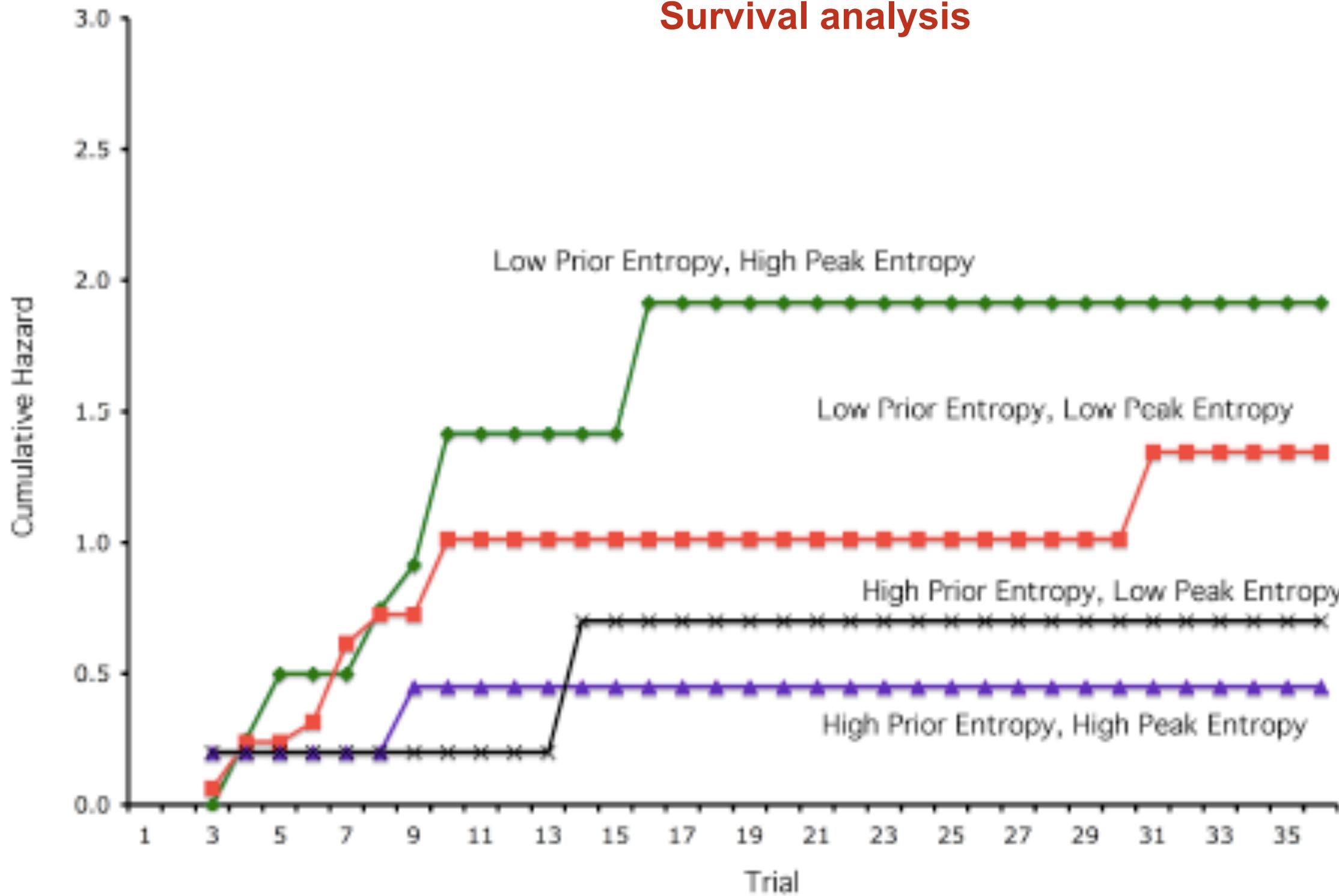


Insight as a phase transition

Entropy, Pre-Discovery



Survival analysis



1. Assumption: Noise / Entropy drives the structural change
2. Hypothesis: Increase noise, this will lead to an earlier discovery of the rule
3. Additional condition: increase noise by making the gear problems shift position on the screen



