Feature Extraction for Deterministic Dynamical Systems

A chaos theory perspective on predictive maintenance

Michael Small, Ayham Zaitouny and Débora Corrêa

Outline

Chaos:

Why Chaos? What data?

Embedding

Tipping points

Surrogates

Further reading

Recurrence:

Embedding review

Recurrence Plots

Recurrence Quantification analysis (RQA)

RPs to detect change points

Complex Network Analysis of Time Series

Recap: Phase Space and Recurrence Plots

Ordinal Partition Networks

Entropy measures

Application to data

Further reading

Part I

Chaos in predictive maintenance

Why Chaos? What data?

What is Chaos?

- The basic task ion predictive maintenance is to preempt change in the system behaviour
- Achieved via either stochastic (i.e. Bayesian or frequentist) or deterministic (machine learning or chaotic) models
- Build a model a check when the model stops working $\implies$ system is changing

Example (sensitivity to initial conditions)

https://www.youtube.com/watch?v=n-mpifTiPV4

Example (chaos in a mechanical double pendulum)

https://www.youtube.com/watch?v=U39RMUzCjiU

Bernoulli map — the simplest example of Chaos

Simplest example of chaos (and specifically sensitive dependence on initial conditions):

$$x_{n+1} = f(x_n) = \begin{cases} 2x_n & 0 \leq x_n < 0.5 \\ 2x_n - 1 & 0.5 \leq x_n \leq 1 \end{cases}$$

Hence:

- This map is bounded by (and onto) $[0, 1]$
- $f'(x) = 2$ a.e. and so local separation increases as 2^n.
- No $x_0 \in \mathbb{R} \setminus \mathbb{Q}$ is periodic $\implies$ chaos a.e.
- Every $x_0 \in \mathbb{Q}$ is periodic $\implies$ periodic orbits are dense

Note, that this means that while almost every random initial condition leads to chaos, every initial condition chosen by a computer is periodic (with period 1).

Logistic growth

In continuous time,

$$x'(t) = \lambda x(t)(1 - x(t))$$

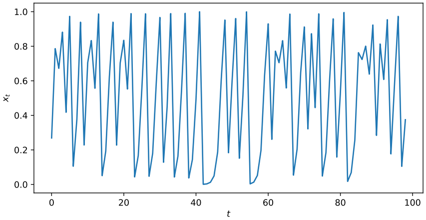
is a simple paradigmatic model of population growth with capacity. In discrete time,

$$x_{t+1} = \lambda x_t(1 - x_t)$$

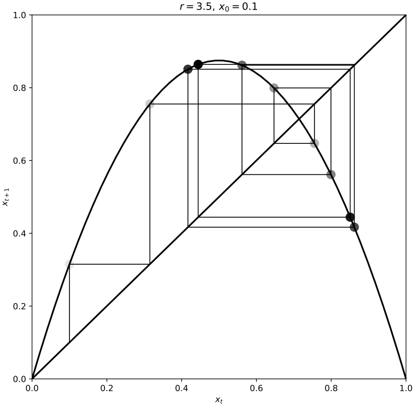
becomes chaotic and is (perhaps) one of the simplest examples of non-trivial practical chaotic dynamics. There are two ways of understanding chaotic dynamical systems like this:

- the *bifurcation diagram*; and,
- *phase space*.

Phase space

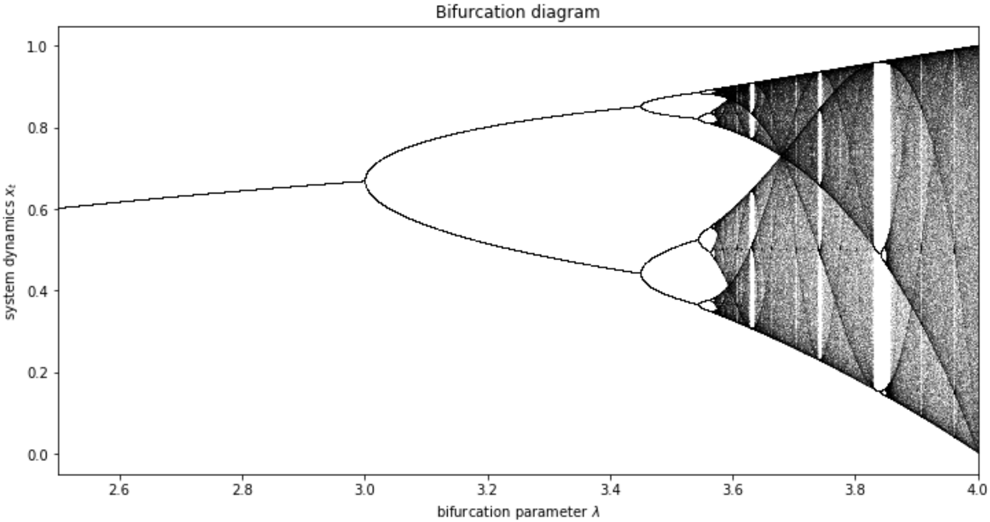


the time series



phase space

Bifurcation



the bifurcation diagram

Data Science Transforming Maintenance

THE UNIVERSITY OF WESTERN AUSTRALIA

TCSG The Complex Systems Group

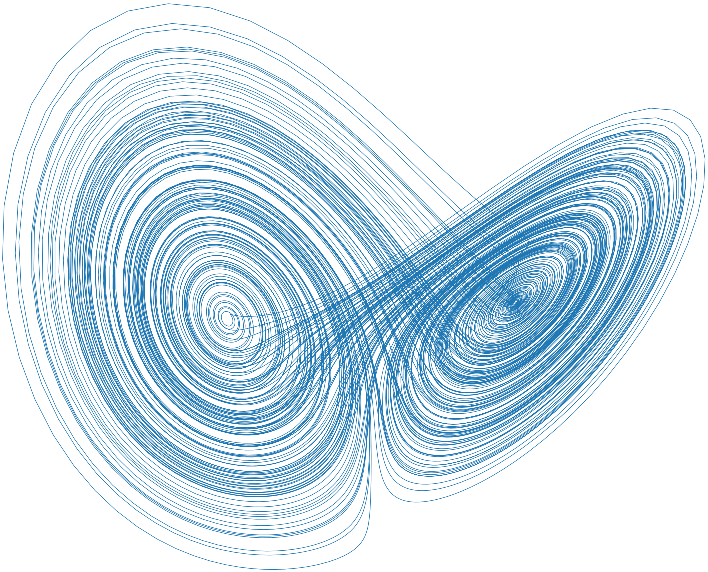
The Lorenz "butterfly"

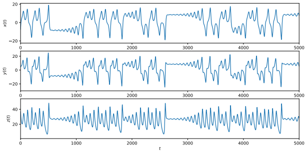
Lorenz proposed a three dimension set of ordinary differential equations to model convection forced via a temperature gradient:

$$\begin{aligned} \frac{dx}{dt} &= -\sigma x + \sigma y \\ \frac{dy}{dt} &= -xz + rx - y \\ \frac{dz}{dt} &= xy - bz \end{aligned}$$

where (it turned out), for $\sigma = 10$, $r = 28$, and $b = \frac{8}{3}$, the system is bounded and aperiodic. The variable x represents intensity of convective motion, y temperature gradient, and z is the degree of nonlinearity in the temperature profile.

Lorenz in phase space

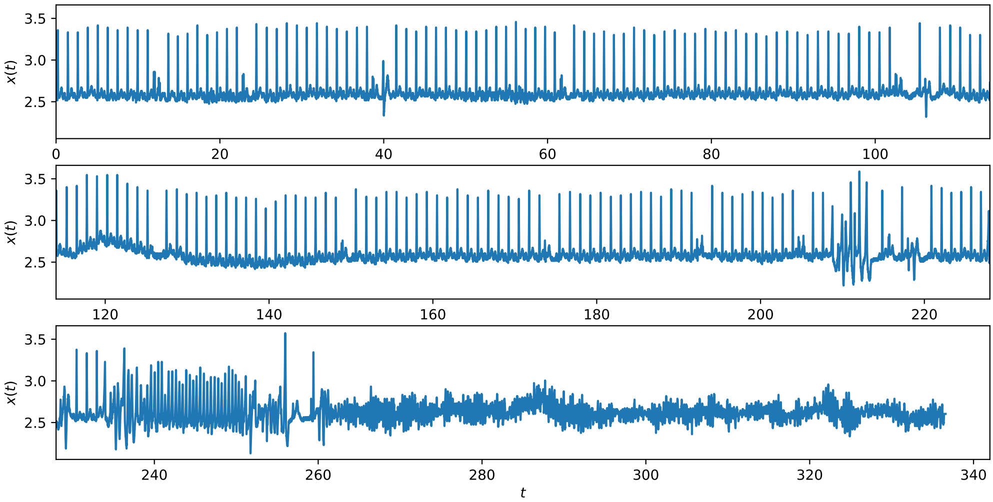




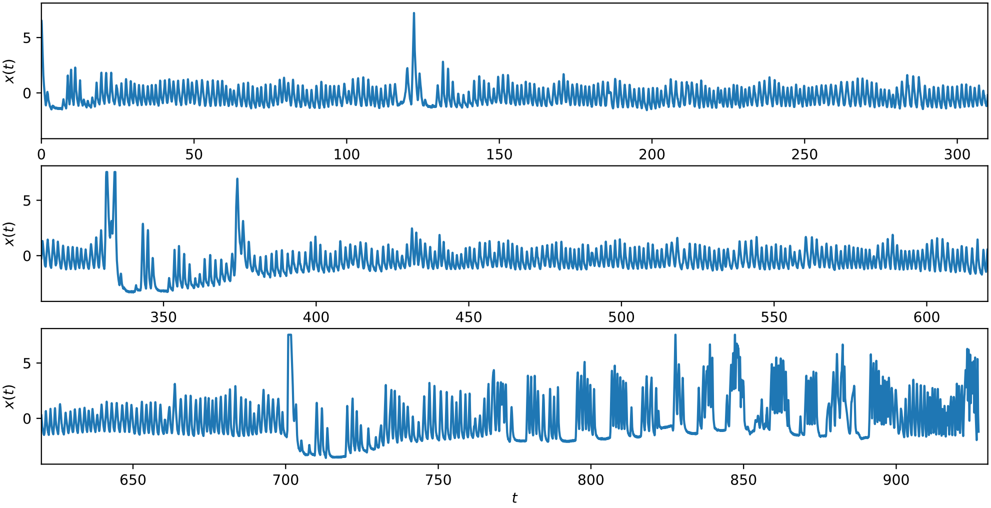
Data for predictive modelling

- What are the characteristics of a “good” data set?
- What sort of data are generated in industry ?

Electrocardiogram during a heart attack



Inductance Plethysomnography during onset of Cheynes-Stokes respiration

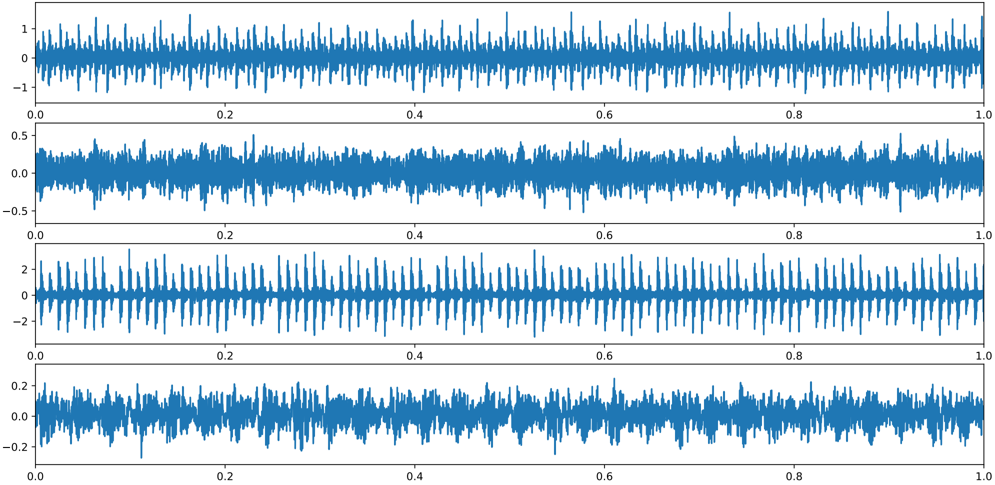


Data Science Transforming Maintenance

THE UNIVERSITY OF WESTERN AUSTRALIA

TCSG The Complex Systems Group

Mechanical bearing failure



Open data available from csegroups.case.edu

Embedding

An Experiment

Let $\Phi(z_0) = \{\phi_t(z_0) | t \in T\}$ be the *trajectory* for an initial condition z_0[1]

Let $h : \mathcal{M} \to \mathbb{R}$ be a *measurement function* and suppose that we can only observe

$$x_i = h(\phi_{i\kappa}(z_0))$$

for $i \in \mathbb{Z}^+ \cup \{0\}$. We call κ the *sampling rate* of our experiment.

What can the *time series* $\{x_i\}_{i=1}^N$ tell us about ϕ?

[1] We can, of course, think about trajectories backwards in time as well — and to do so is more typical. Moreover, we have yet to prove that $\Phi(z_0)$ is well defined, but suppose for now that it is.

Data Science Transforming Maintenance THE UNIVERSITY OF WESTERN AUSTRALIA TCSG The Complex Systems Group

Takens' Embedding Theorem (1981)

The map

$$x_i \mapsto (x_i, x_{i+1}, \ldots x_{(i+m-1)}) =: v_i$$

is an embedding of a compact manifold with dimension $d \in \mathbb{Z}^+$ ($m = 2d + 1$) if $h : \mathcal{M} \to \mathbb{R}$ is C^2 and "generic"[2]. Moreover, evolution of v_i are diffeomorphic to the dynamics of ϕ.

Corollary (Sauer, York, and Casdagli, 1981)

Takens' embedding theorem also holds for $d \in \mathbb{R}^+$ with the condition that $m \geq 2d + 1$.

Hence, a fractal attractor $\mathcal{A} \subset \mathcal{M}$ can be *reconstructed* from a time series generated from a trajectory lying on that attractor.

[2] Sufficiently well coupled between the d-dimensional variables and the theorem is then true almost always.

Delay reconstruction of experimental data

Suppose that a time series, as defined above, is the output of a deterministic and stationary (autonomous) dynamical system.

Techniques to reconstruct the underlying attractor are now well established[3] and widely used.

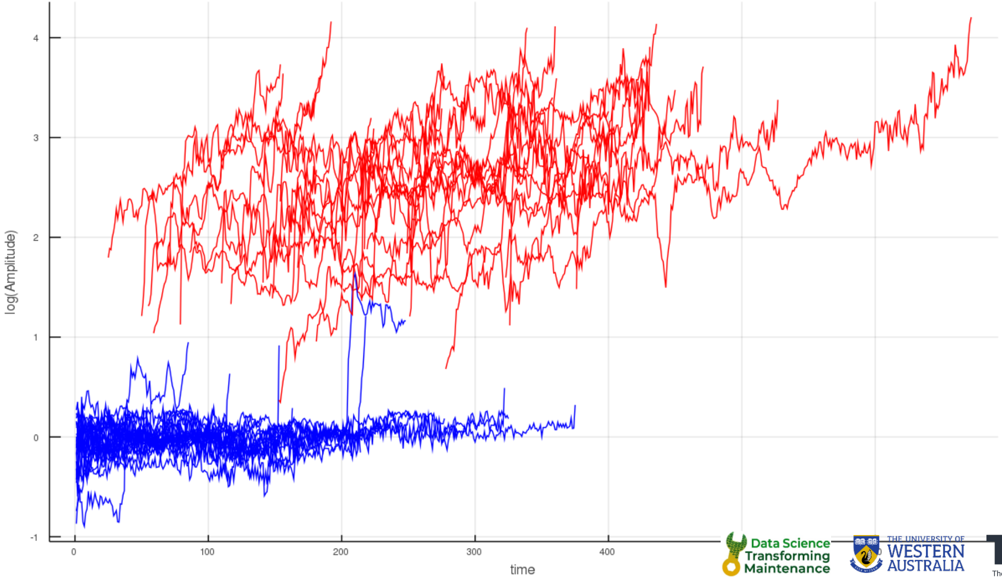
This is usually achieved by estimating two parameters: *embedding dimension* m, and *embedding lag* τ and applying the map

$$x_i \mapsto (x_i, x_{i-\tau}, \ldots x_{(i-(m-1)\tau)}) =: v_i$$

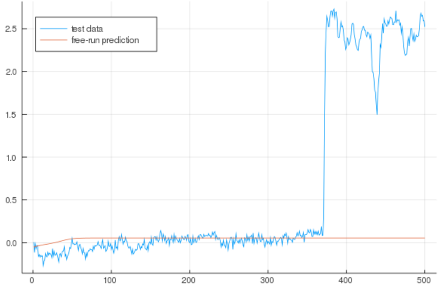
[3] See M. Small Applied Nonlinear Time Series Analysis, World Scientific, 2003.

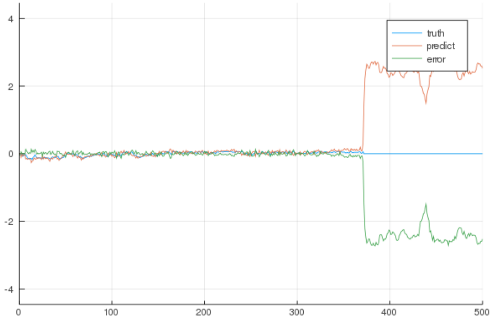
Tipping points

Some maintenance data: Component failure



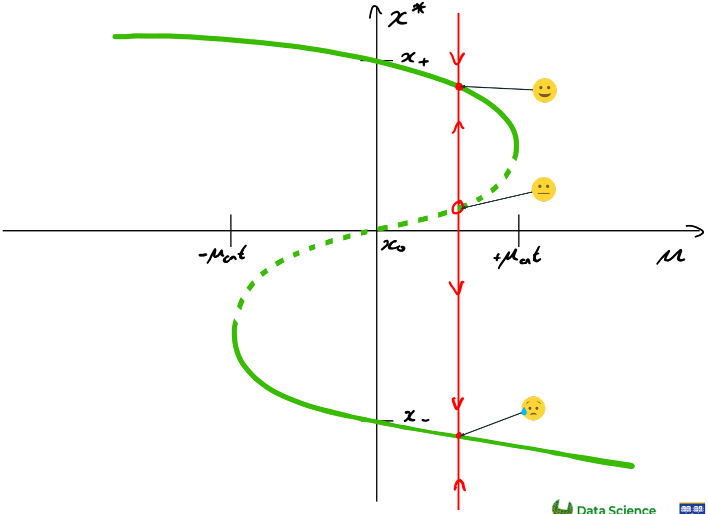
Model it



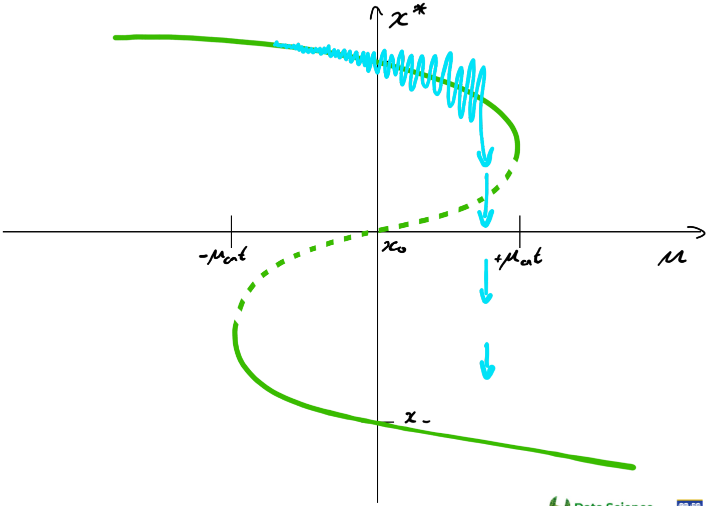


Data Science Transforming Maintenance
THE UNIVERSITY OF WESTERN AUSTRALIA
TCSG The Complex Systems Group

Hysteresis and a tipping point



Hysteresis and a tipping point



The tipping point model of dynamics

The signature of tipping point transitions are

- loss of stability
- increase sensitivity to noise
- increase error in predictive model (built on historical data)

Surrogates

Surrogate data

- How can we know that an estimated statistical quantity (hopefully an invariant of the underlying dynamical system) has been reliably estimated from data?
- What does the estimated correlation dimension (Lyapunov exponent, etc.) actually mean?
- What would we expect for boring (linear noise) data?
- Are nonlinear time series techniques warranted by the data? Or would linear methods suffice?
- Which class of model (linear vs. nonlinear, radial basis functions, etc.) are warranted by, and produce results consistent with, the data?

Hypothesis testing

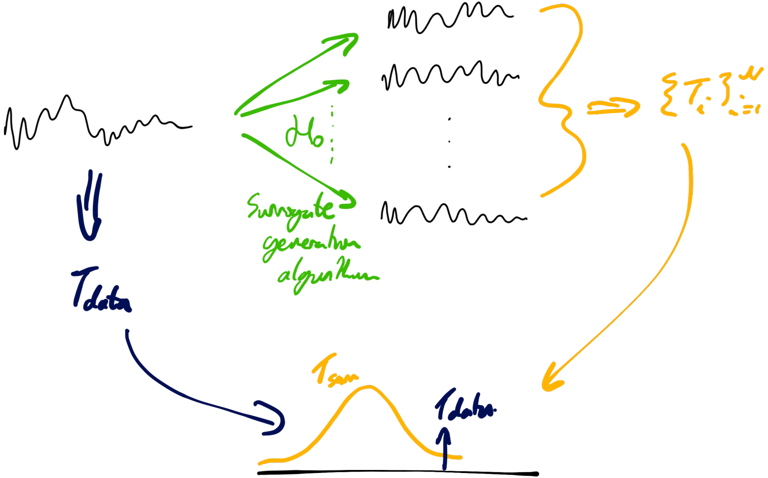
Definition

Hypothesis testing in statistics is the idea that one can compute some statistical value from observed data, compare that value to some theoretical distribution and conclude that an underlying (null) hypothesis is either *rejected* or that we *fail to reject* said hypothesis.

Definition

A *surrogate data hypothesis test* consists of two components — (i) an *algorithm* with which to generate surrogate (that is, randomised) realisations s_t from an observed time series x_t, and (ii) a *test statistic* $d(\cdot)$. One the computes values of the test statistic for the data $d_* := d(\{x_t\}_t)$ and for the ensemble of (many) surrogates $d_k := d(\{s_t\}_t)$. Comparing the statistic values for the data d_* to the distribution generated from the surrogates $\{d_k\}_k$, one can then make the decision to “reject” or “fail to reject”.

The surrogate algorithm



Example (Theiler's "Algorithm 0")

Let $\{x_t\}_{t=1}^{N}$ be an observed time series, and then an *algorithm 0* surrogate realisation $\{s_t\}_{t=1}^{N}$ is obtained by resampling, with replacement: choose a bijection $\tau : \mathbb{Z}_n \mapsto \mathbb{Z}_n$ then $s_t = z_{\tau(t-1)+1}$ for $t = 1, 2, 3, \ldots, N$.

Note:

- s_t is a random reordering (shuffling) of x_t
- s_t is not correlated with $s_{t-\ell}$ $\forall \ell$.

Example (applied to linear noise)

Suppose that x_t is a realisation of an autoregressive linear noise process and that the autocorrelation with lag one $\rho(1) = E(x_t x_{t-1}) =: \rho_1 \neq 0$. Let $d(\cdot) = E(x_t x_{t-1})$, then $d_* = \rho_1$ and $E(d_k) = 0$ (the d_k follow a normal distribution with mean 0 and variance scaling with $\frac{1}{k}$). For k sufficiently large one can reject the hypothesis that x_t is independent and identically distributed noise.

Algorithm 0 tests the hypothesis that the observed data is *independent and identically distributed noise*.

Example ("Algorithm 1": Fourier transform surrogates)

Let X_n be the discrete Fourier transform of x_t (which we denote $\mathcal{F}(\{x_t\}) = \{X_n\} = \{R_n e^{-i\pi\theta_n + \phi_n}\}$). The Fourier transform surrogate is the inverse Fourier transform of X_n with the complex phases randomised (pairwise, to ensure that the result is real) over 2π:

$$s_t = \mathcal{F}^{-1}\left(\{R_n e^{-i\pi\theta_n + \phi_n + \psi_n}\}_n\right)$$

where ψ_n are uniform random numbers on $[0, 2\pi)$.

Algorithm 1 tests the hypothesis that the observed data is *linearly filtered noise* (NOTE: this means that we should suppose that the x_t have a Gaussian distribution).

Inherent in the Fourier transform are all the assumption about linearity and periodicity. When these assumption are violated, one expects to reject the hypothesis, however there is no guarantee that something loopy might not happen. Moreover, this scheme works when the input data is (close enough to) Gaussian distributed. Otherwise, what is being tested should be trivially rejected and the following test is better.

Example ("Algorithm 2": Amplitude adjusted Fourier transform (AAFT) surrogates)

Generate N realisations of a Gaussian distribution $\{g_t\}$ and reorder $\{g_t\}$ so that it has the same rank distribution as $\{x_t\}$ — i.e. if x_i is the m_i-th largest among all the observations $\{x_t\}$ then g_i should be the m_i-th largest of the $\{g_t\}$ $\forall i$. Generate an Algorithm 1 surrogate of $g_t - \hat{g}_t$. Create the surrogate s_t by reordering the original data to have the same rank distribution as $\hat{g}_t$. I.e. $s_t = x_{\tau(t-1)+1}$ where the permutation τ is chosen so that when $\hat{g}_i$ is the m_i-th largest among all the observations $\{\hat{g}_t\}$ then s_i should be the m_i-th largest of the $\{s_t\}$

Algorithm 2 tests the hypothesis that the observed data is *monotonic static nonlinear transform of linearly filtered noise* This is useful when the data is clearly not Gaussian, but one suspects the underlying cause to be a linear stochastic process. To address this, we rescale the data to be Gaussian, apply Algorithm 1, and then rescale the output to match the original data.

- Algorithm 0 preserves, in the surrogates, the probability distribution of the data
- Algorithm 1 preserves the linear correlation structure (and Fourier power spectrum) of the data
- Algorithm 2 preserves, approximately, both probability distribution and power spectrum

Testing your data with surrogates

example....

Further reading

Further reading

- B. Thorne, T. Jüngling, M. Small and M. Hodkiewicz. "Parameter Extraction with Reservoir Computing: Nonlinear Time Series Analysis and Application to Industrial Maintenance" Chaos (2021), in press.
- D.C. Corrêa, J.M. Moore, T. Jüngling and M. Small. "Constrained Markov order surrogates" Physica D **406** (2020): 132437. pdf
- X. Peng, M. Small, Y. Zhao, J.M. Moore. "Detecting and Predicting Tipping Points" International Journal of Bifurcations and Chaos **29** (2019): 1930022 pdf
- K. Sakellariou, T. Stemler and M. Small. "Markov modelling via ordinal partitions: A novel paradigm for network-based time series analysis" Physical Review E **100** (2019), 062307 pdf

Data Science Transforming Maintenance
THE UNIVERSITY OF WESTERN AUSTRALIA
TCSG The Complex Systems Group

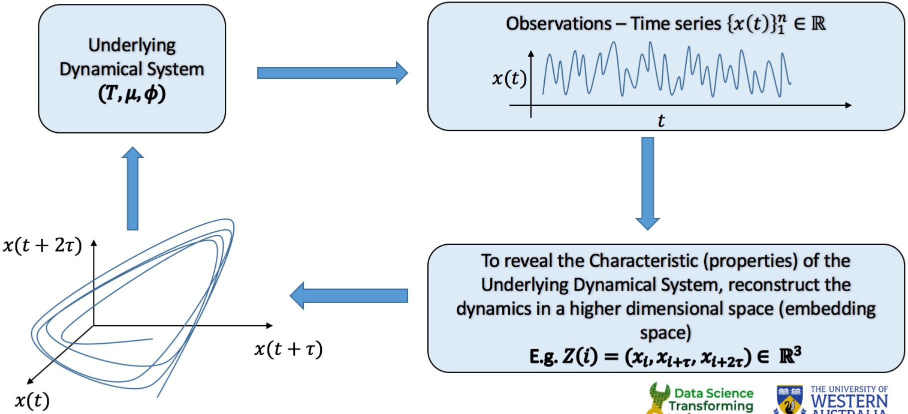
Part II

Pattern Recognition and Change Point Detection Using Recurrence and Dynamical Systems

Embedding review

Embedding Definition

For a time series $\{x(t)\}_1^N$ we define an embedded sequence with embedding dimension m and time–lag τ as $\{Z(k) = (x(k), x(k+\tau), x(k+2\tau), \ldots, x(k+(m-1)\tau))\}_1^M$.



Embedding Parameters

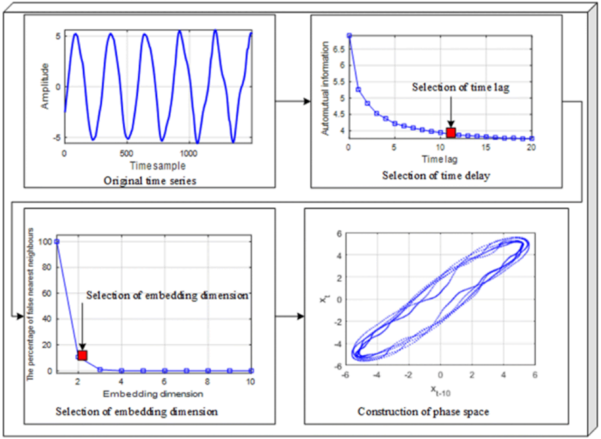
The embedding parameters can be selected by one of the methods[4]:

- For τ:
  - First zero Auto-correlation
  - First minimum Mutual Information
- For m:
  - False Nearest Neighbours (FNN): For each m we count the points that they are close (neighbours) to $\mathbf{x(i)}$, but they are not neighbours to $\mathbf{x(i+1)}$

[4]There are other methods addressed by M. Small, *Applied nonlinear time series analysis: applications in physics, physiology and finance*, Vol. 52 (World Scientific, 2005)

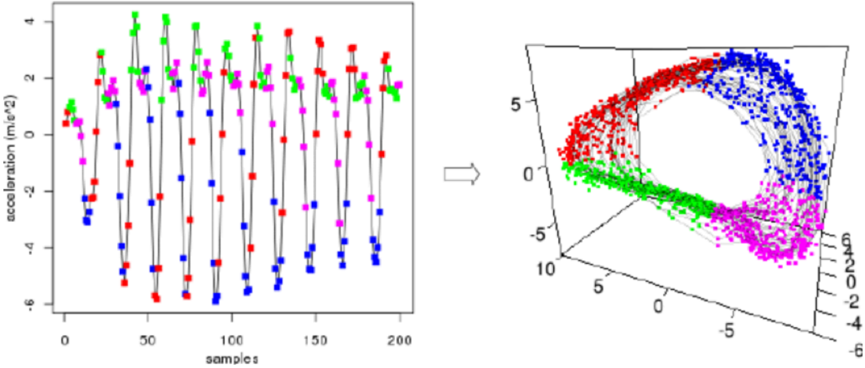
Embedding Parameters

Aydin, İ., Karakose, M. & Akin, E. A new method for time series classification using multi-dimensional phase space and a statistical control chart. *Neural Comput & Applic* **32,** 7439–7453 (2020). https://doi.org/10.1007/s00521-019-04270-1



Embedding Example

Frank, J., Mannor, S. and Precup, D., 2011, September. Activity recognition with mobile phones. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases* (pp. 630-633). Springer, Berlin, Heidelberg.



Recurrence Plots

Why Recurrence Plots?

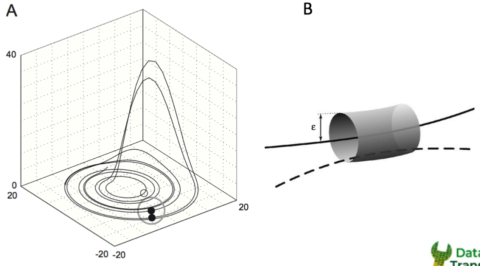
- Recurrences is originally introduced by Henry Poincare in 1890. Later in 1987, Eckmann et al. introduced the method of RPs to visualise the recurrences of dynamical systems.
- Recurrence is a fundamental characteristic of many dynamical systems.
- Recurrence event is defined when two states of the system pass close to each other in different times.
- RPs is a visualisation tool associated with quantitative analysis for time series of nonlinear dynamical systems.
- It provides an alternative and powerful mathematical framework to study time series data and extract features of different systems or patterns.

RPs definition

For a time series (embedded or multivariate) $\mathbf{x}_{i\,i=1}^{\;N} \in \mathbb{R}^d$, the Recurrence Plot matrix is defined as follows:

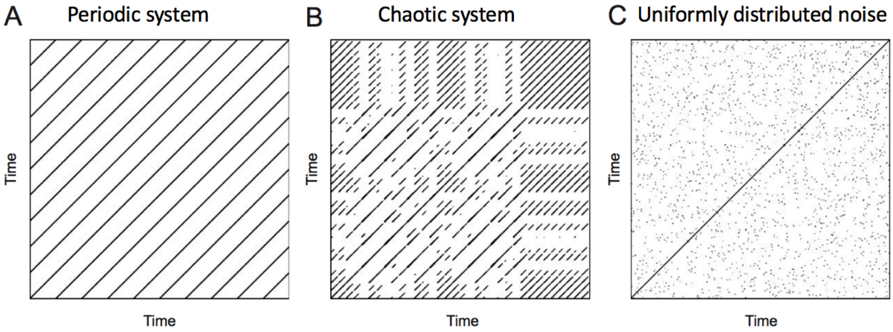
$$RP_{ij} = \begin{cases} 1 & \text{if } \|\mathbf{x}_i - \mathbf{x}_j\| < \varepsilon \\ 0 & \textit{otherwise} \end{cases}$$

where $\|\cdot\|$ is some norm (distance) metric, e.g. Euclidean distance. ε is the recurrence threshold, could be topological or metric.



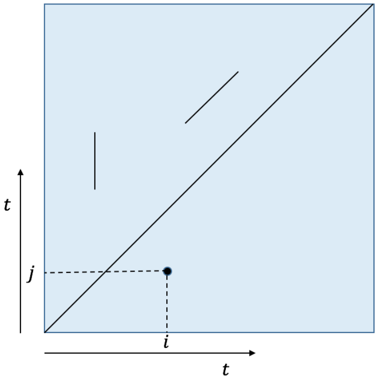
N. Marwan et al. / Physics Reports 438 (2007) 237 – 329

Examples of RPs of some dynamics (patterns)



N. Marwan et al. / Physics Reports 438 (2007) 237 – 329

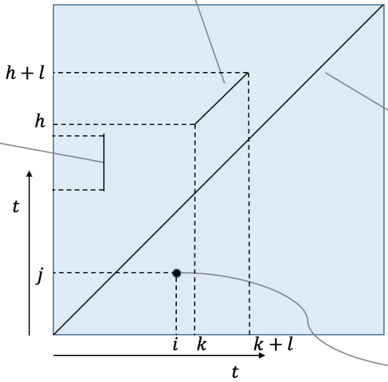
Some structures in RPs



Some structures in RPs

Diagonal lines occur when a segment of the trajectory runs almost in parallel to another segment for l time units

Vertical lines occur when a state doesn't change or change very slowly (trap behaviour)



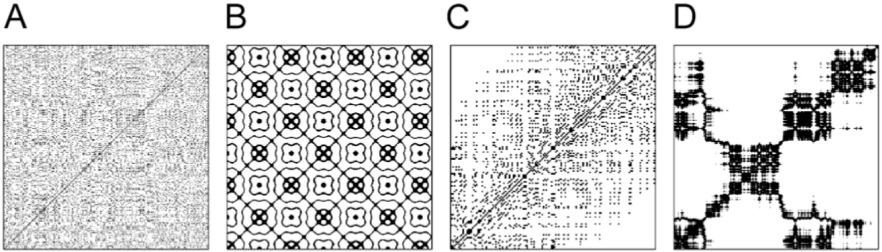
Main diagonal; all points = 1

$$RP_{ij} = 1\ iff\ \vec{x_i} \approx \vec{x_j}$$

(close to each other in the phase space)

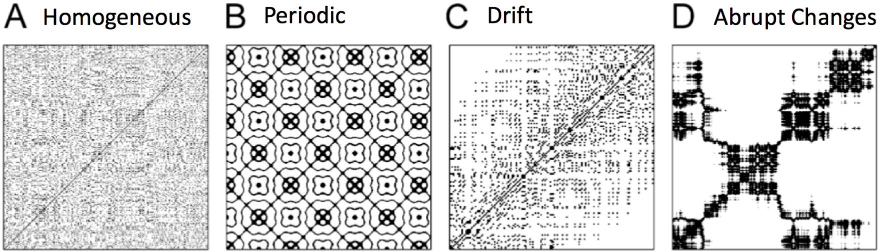
The structure of the RPs reflects important characteristic of the dynamic

N. Marwan et al. / Physics Reports 438 (2007) 237 – 329



The structure of the RPs reflects important characteristic of the dynamic

N. Marwan et al. / Physics Reports 438 (2007) 237 – 329



(White noise – stationary system)

(Super-positioned harmonic oscillations)

(Non-stationary system – fading to upper left and lower right corners)

(White areas – bands)

Typical patterns in RPs

Pattern	meaning
Homogeneity	The system is stationary
Fading to the upper left and lower right corners	The system is non-stationary and contains a trend or a drift
Disruptions (white bands)	The system is non-stationary and abrupt changes may have occurred
Periodic/quasi periodic patterns	The system is periodic, the time distance between the periodic pattern (e.g. diagonal lines) corresponds to the period. Different distances between the patterns reveal quasi-periodic system
Single isolated points	Strong fluctuation in the system
Diagonal segments	The evolution of the system is similar at different epochs. The system could be deterministic. If these diagonal segments beside single isolated points, this means the system could be chaotic.
Vertical or horizontal segments	Some states don't change or change slowly for some time, indication of trap-behaviour (laminar states)

Recurrence Quantification analysis (RQA)

Why RQA?

- RQA provides quantitative measures of the complexity of the time series and the underlying system.
- The measures are based on the density, diagonal lines or vertical lines of the RPs.
- These RQA measures reflect the system's characteristic (properties).
- An alternative and powerful framework to extract features of the system, which can be used for further analysis (e.g. predictive maintenance).

RQA measures

Recurrence Rate (*RR*): measures the density of recurrence points in the RP. It corresponds with probability that a specific state will recur. ($\sim$ correlation sum).

$$RR = \frac{1}{N^2} \sum_{i,j=1}^{N} RP_{ij}$$

Where N is the length of the time series.

RQA measures

Determinism (DET): is the percentage of the recurrence points which form diagonal lines of minimal length l_{min} in the RP.

$$DET = \frac{\sum_{l=l_{min}}^{N} lP(l)}{\sum_{l=1}^{N} lP(l)}$$

where $P(l)$ is the frequency distribution of length l of the diagonal lines (i.e. it counts how many instances have length l).
DET is related with the predictability of the dynamical system, for example:

- Random system of white noise $\Rightarrow$ RP has almost only single dots and very few diagonal lines $\Rightarrow$ smaller DET.
- Deterministic process $\Rightarrow$ RP has very few single dots but many long diagonal lines $\Rightarrow$ larger DET.

Average diagonal line length (L): is the average length of the diagonal lines.

$$L = \frac{\sum_{l=l_{min}}^{N} lP(l)}{\sum_{l=l_{min}}^{N} P(l)}$$

It is related with the predictability time of the dynamical system (i.e. the average time that two segments of the system's trajectory are close to each other).

Divergence (DIV): is the inverse of the maximal diagonal line length L_{max}.

$$DIV = \frac{1}{L_{max}}$$

where $L_{max} = \max(\{l_i\}_{i=1}^{N_l})$, where N_l is the total number of diagonal lines in the RP. DIV is related with the positive Lyapunov exponent of the dynamical system (i.e. faster trajectory segments diverge $\Rightarrow$ shorter are the diagonal lines $\Rightarrow$ larger DIV).

RQA measures

Entropy ($ENTR$): is the Shannon entropy of the probability of the diagonal line lengths $p(l)$.

$$ENTR = -\sum_{l=l_{min}}^{N} \rho(l) ln(\rho(l))$$

where $\rho(l) = \frac{P(l)}{\sum_{l=l_{min}}^{N} P(l)}$ (the probability that a diagonal line has length l).
$ENTR$ reflects the complexity of the RP in respect of the diagonal lines.

Ratio: is the ratio between *DET* and *RR*.

$$RATIO = N^2 \frac{\sum_{l=l_{min}}^{N} lP(l)}{(\sum_{l=1}^{N} lP(l))^2}$$

It can be used to uncover transitions in the dynamics.

Laminarity (*LAM*): is the percentage of the recurrence points which form vertical lines of minimal length v_{min} in the RP.

$$LAM = \frac{\sum_{v=v_{min}}^{N} vP(v)}{\sum_{v=1}^{N} vP(v)}$$

where $P(v)$ is the frequency distribution of length v of the diagonal lines (i.e. it counts how many vertical lines have length v).

LAM is related with the amount of laminar states in the system (i.e. the states which are trapped for some time).

RQA measures

Trapping Time (*TT*): is the average length of the vertical lines.

$$TT = \frac{\sum_{v=v_{min}}^{N} vP(v)}{\sum_{v=v_{min}}^{N} P(v)}$$

TT is related with the laminarity time of the dynamical system (i.e. how long the system remains in a specific state).

RQA measures

Maximal vertical line length (v_{max}): is the longest time in which a state is trapped.

$$v_{max} = \max(\{v_l\}_{l=1}^{N_v})$$

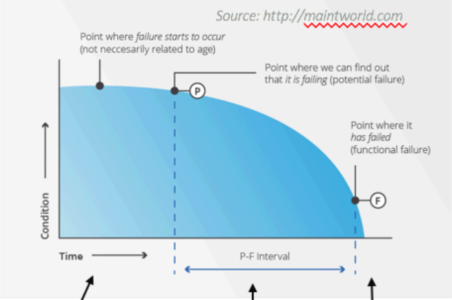
where N_v is the total number of vertical lines in the RP.

RPs to detect change points

Motivation

Source: http://lase.de

Source: http://tenova.com



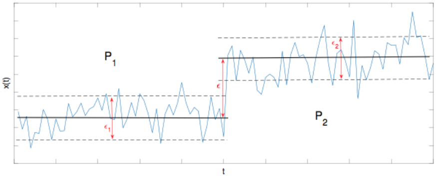
Predictive

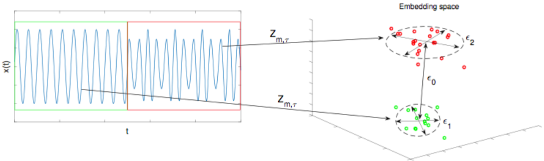
Proactive

Reactive

Change points (tipping points)

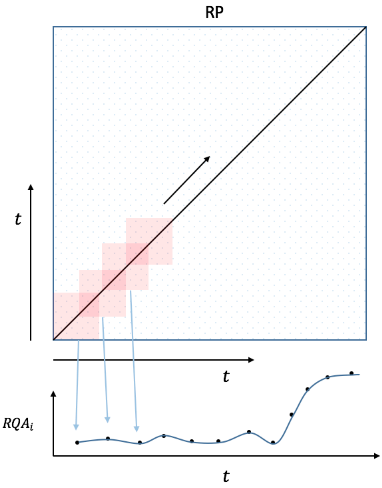
- Tipping points are the critical points at which an important change in the system occurs or is derived (e.g. a bifurcation value of a system parameter).
- Locating or detecting these tipping points is of high importance to many fields – including predictive maintenance.
- Two types of transition:
  - State–transition
  - Dynamic–transition





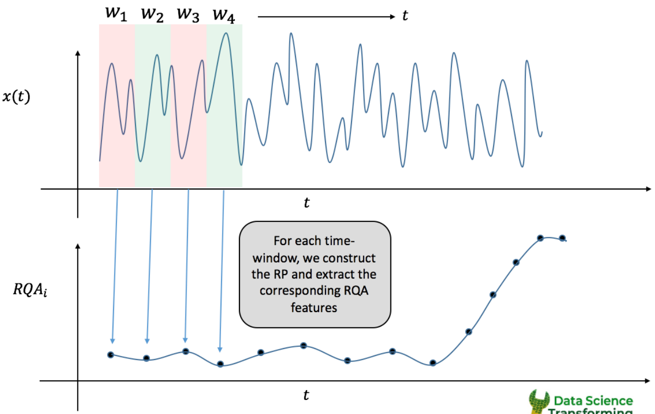
Time–dependent RQA

Approach 1 – sliding window: Instead of computing the RQA measures of the entire RP, RQA measures can be computed in small windows sliding over the RP along the main diagonal ⇒ This provides time–dependent RQA measures.



Time–dependent RQA

Approach 2 – windowing the time series: Alternatively, in this approach, the time series is segmented into sequential windows. For each window, a RP is created and associated RQA measures are computed ⇒ This provides time–dependent RQA measures.



Quadrant Scan

Definition: From the recurrence plot matrix RP, we construct a sequence $QS(k)$ by counting the ratio of the density of points of those that are in the quadrants with $i, j < k$ or $i, j > k$ versus the density of whole points in all quadrants of RP.

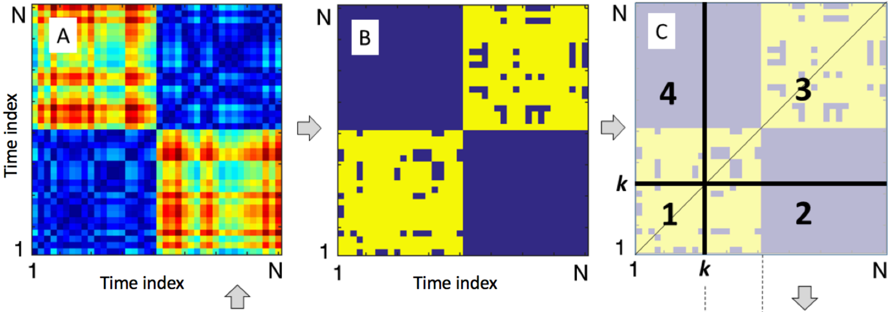
$$QS(k) = \frac{D_{1,3}}{D_{1,3} + D_{2,4}},$$

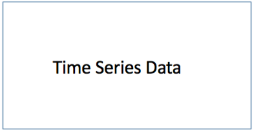
where $D_{1,3}$ is the density of the points in quadrants 1 and 3 while $D_{2,4}$ is the density in quadrants 2 and 4. They are defined as follows:

$$\begin{aligned}
D_{1,3} &= \frac{\sum_{i,j<k} RP_{ij} + \sum_{i,j>k} RP_{ij}}{(k-1)^2 + (N-k)^2} \\
D_{2,4} &= \frac{\sum_{i<k,j>k} RP_{ij} + \sum_{i>k,j<k} RP_{ij}}{(k-1) \times (N-k) \times 2}
\end{aligned}$$

Maxima (peaks) of $QS(k)$ correspond to transitions in the system. This is proved in the following theorem. The value of $QS(k)$ is between 0 and 1.

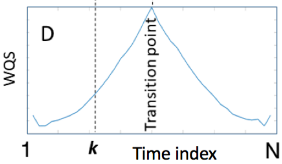
Quadrant Scan





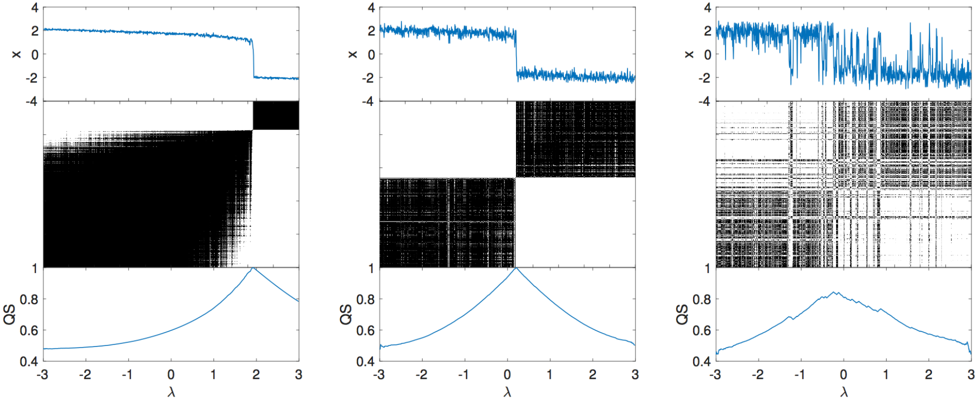
Workflow:

A. distance matrix
B. recurrence plot
C. quadrant scan (e.g. for point k)
D. quadrant scan profile



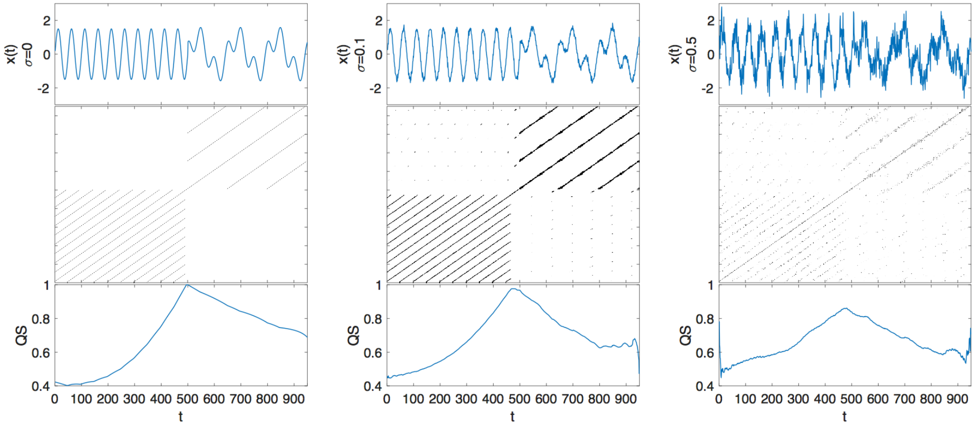
Example of detecting state–transition (type 1)

Noisy Stochastic System:

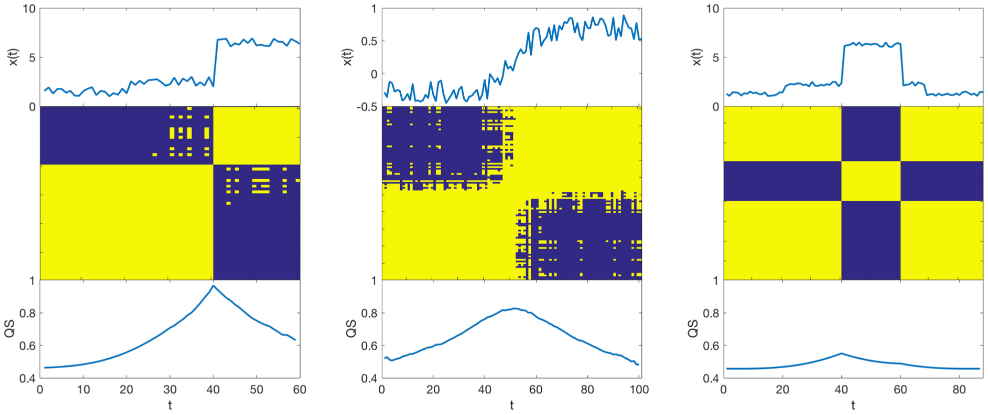


Example of detecting dynamic–transition (type 2)

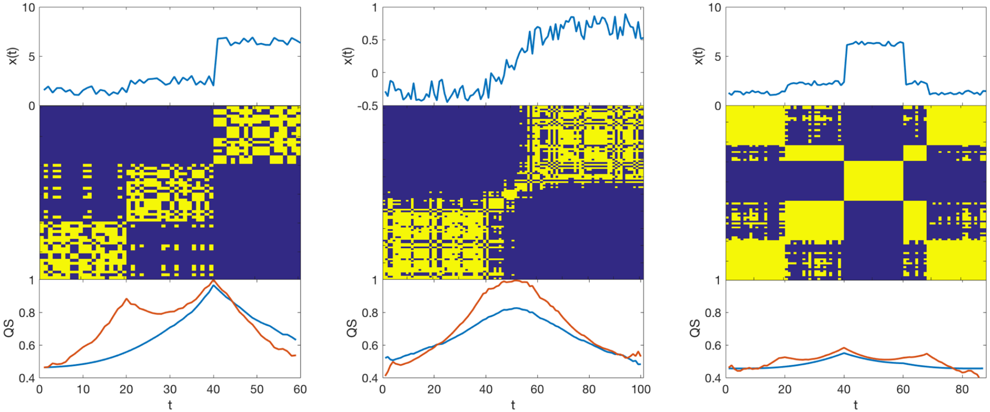
Periodic Stochastic System:



Different scenarios

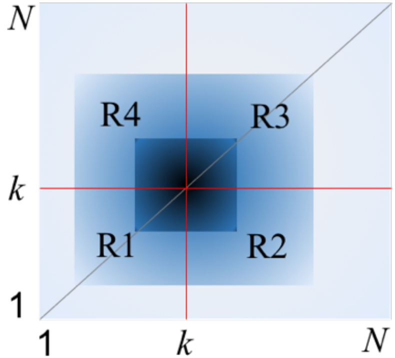


Different scenarios

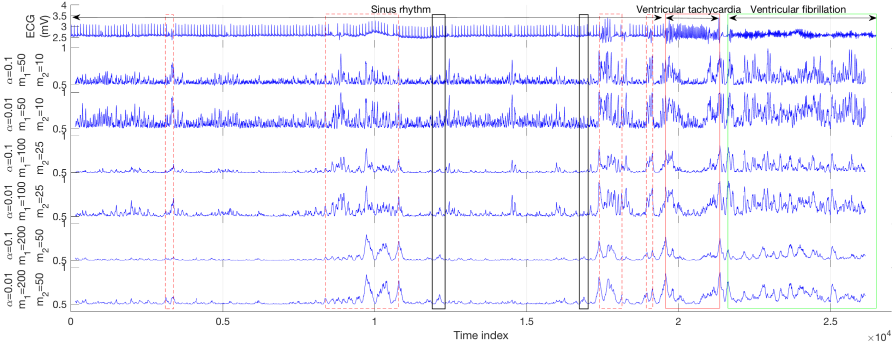


Weighted Quadrant Scan

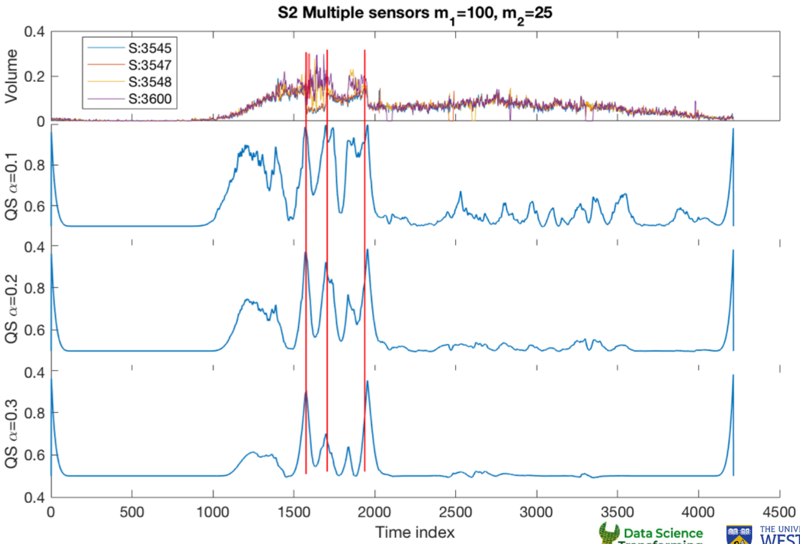
- To detect local transitions, we assign higher weightings to closer points.
- At time k, the dark region represents higher weighted points and the light region for lower weighted points.



Real–world application 1: ECG data

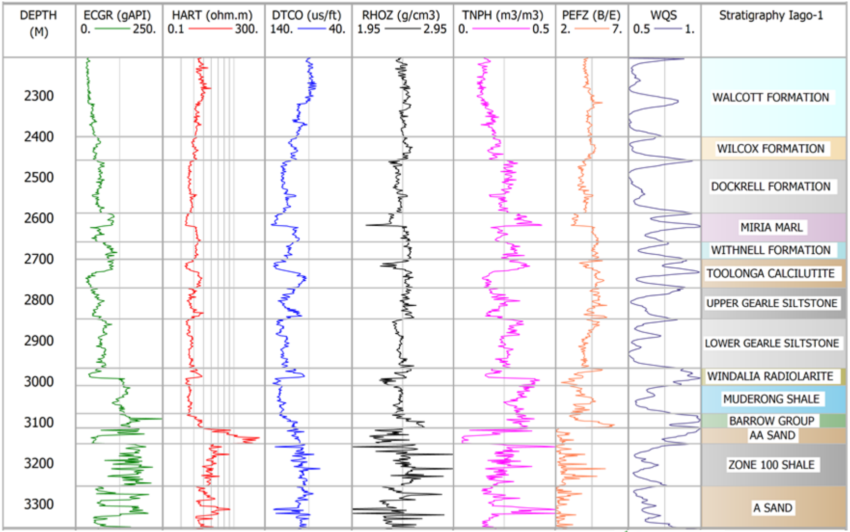


Real–world application 2: Traffic data



Data Science Transforming Maintenance
THE UNIVERSITY OF WESTERN AUSTRALIA
TCSG The Complex Systems Group

Real–world application 3: Geological data



Part III

Complex Networks Analysis of Time Series

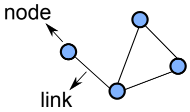
Complex Network Analysis of Time Series

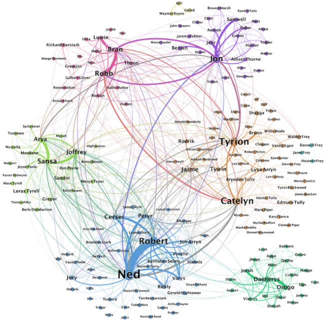
What is a Complex Network?

A mathematical representation of a real-world complex system.

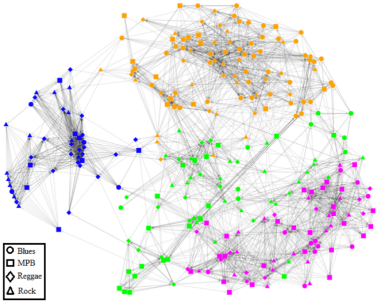
Defined by a graph $G = (V, E)$ where V is the set of vertices/nodes and E the set of edges/links that connect two nodes.

It describes non-trivial relationships of interconnected entities.





Network of Thrones

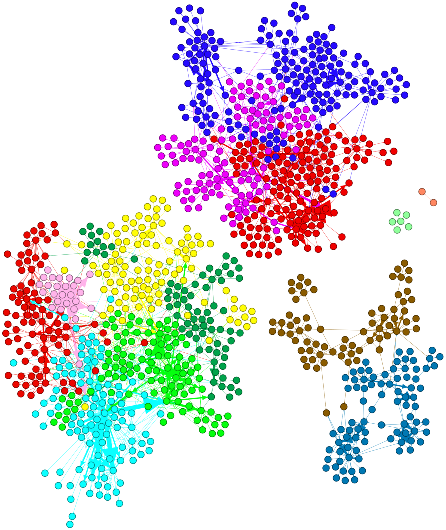


Music network

Network of Thrones: `https://networkofthrones.wordpress.com`

Music Network: Corrêa, D. C., Levada, A. L., & Costa, L. D. F. (2011). Finding Community Structure in Music Genres Networks. In ISMIR (pp. 447-452).

In a industrial context



A network can represent the inter-relations of sub-systems, components and/or assets in a plant.

-> But that is not the type of network we will talk about today.

Network from a time series

Represent the dynamical features of a time series in an alternative domain, so that tools of that domain can be applied to measure properties of interest.

Nodes are representation of the states of a dynamical system.

Links are transactions of one state region to another.

Network properties <=> system properties (track the dynamics).

Network of state transitions - Markov state process to infer properties of the dynamics.

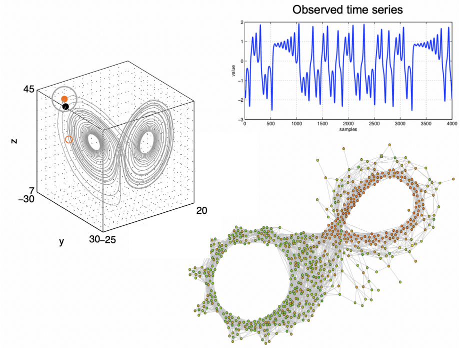


Image from Donner, R. V., Zou, Y., Donges, J. F., Marwan, N., & Kurths, J. (2010). Recurrence networks—a novel paradigm for nonlinear time series analysis. New Journal of Physics, 12(3), 033025.

Recap: Phase Space and Recurrence Plots

Dynamical Systems

A dynamical system is a triple $(T, \mathcal{M}, \phi)$:

$$\phi : U \subset T \times \mathcal{M} \to \mathcal{M}$$

Consider T as time and $\mathcal{M}$ as space. The operator $\phi_t(z)$ defines the evolution of a point $z \in \mathcal{M}$ forwarded by time t.

- $\phi_0(z) = z$ for all $z \in \mathcal{M}$
- $\phi_{t+s}(z) = \phi_t \circ \phi_s(z)$ for all $z \in \mathcal{M}$ and $t, s \in T$

Hence, $\phi(.)$ is the dynamical evolution operator.

Dynamical Systems (Cont.)

Let $\Phi(z_0) = \{\phi_t(z_0) | t \in T\}$ be the trajectory for an initial condition z_0
Let $h : \mathcal{M} \to \mathbb{R}$ be a measurement function and suppose we can only observe x_i

$$x_i = h\big(\phi_{ik}(z_0)\big)$$

for $i \in \mathbb{Z}^+ \cup \{0\}$. We call k the sampling rate of our experiment.

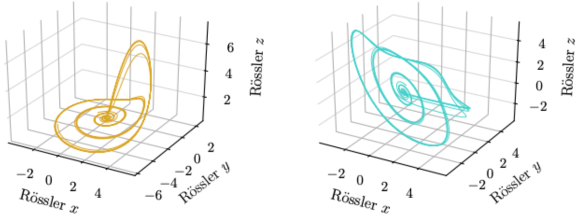
What can the time series $\{x_i\}_{i=1}^N$ tell us about ϕ?

State space reconstruction using time delay embedding

The way Nonlinear Time Series Analysis deals with it is by means of Taken's Embedding Theorem:

$$x_i \to \left(x_i, x_{i-\tau}, \ldots, x_{(i-(m-1)\tau}\right) =: v_i$$

A fractal attractor $\mathcal{A} \subset \mathcal{M}$ can be reconstructed from a time series generated from a trajectory lying on that attractor.

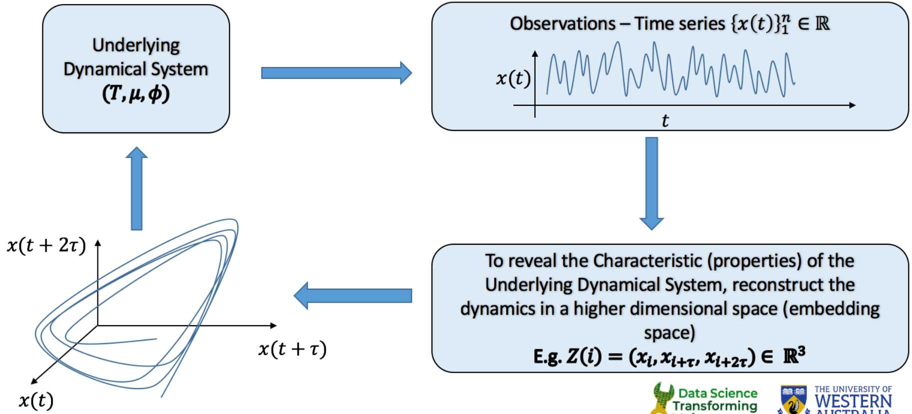


Require the estimation of the embedding dimension m, and the embedding lag τ.

Images from Goswami, B. (2019). A Brief Introduction to Nonlinear Time Series Analysis and Recurrence Plots. Vibration, 2(4), 332-368.

State space reconstruction using time delay embedding

For a time series $\{x(t)\}_1^N$ we define an embedded sequence with embedding dimension m and time–lag τ as $\{Z(k) = (x(k), x(k+\tau), x(k+2\tau), \ldots, x(k+(m-1)\tau))\}_1^M$.

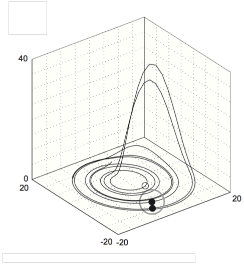


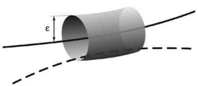
Recurrence Plots

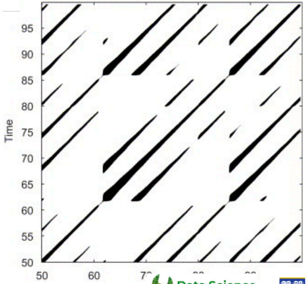
For a time series (embedded or multivariate) ${\mathbf{x}_i}_{i=1}^{N} \in \mathbb{R}^d$, the Recurrence Plot matrix is defined as follows:

$$RP_{ij} = \begin{cases} 1 & \text{if } \|\mathbf{x}_i - \mathbf{x}_j\| < \varepsilon \\ 0 & otherwise \end{cases}$$

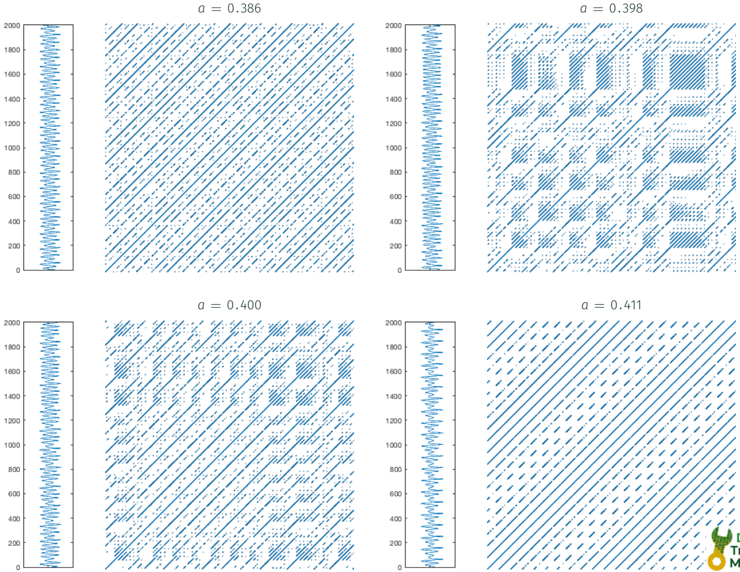
where $\|\cdot\|$ is some norm (distance) metric, e.g. Euclidean distance. ε is the recurrence threshold, could be topological or metric.





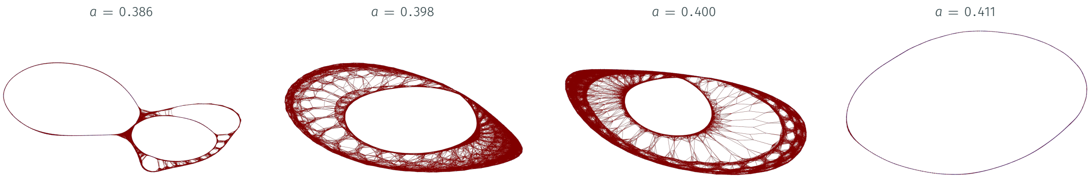


Recurrence Plots (Cont).

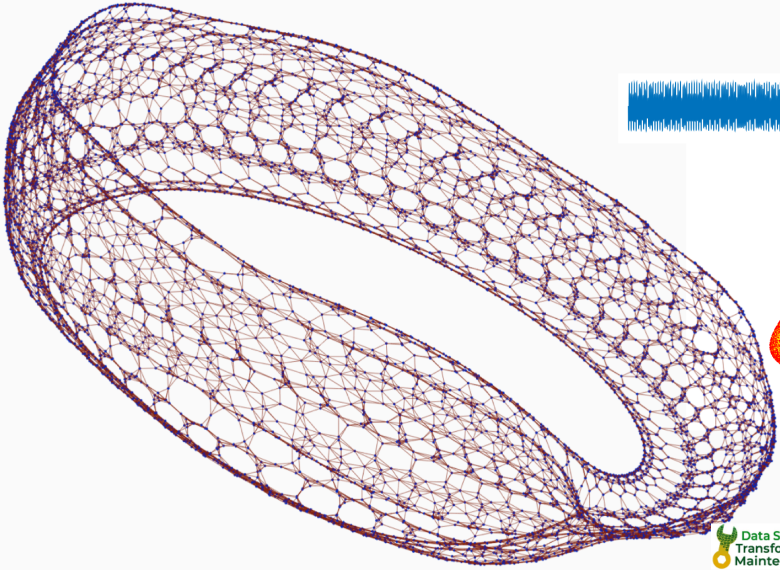


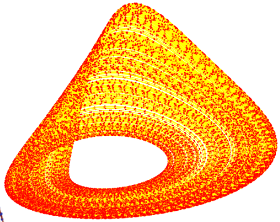
Recurrence Networks

How can we bring *networks* in on the act? We can treat $A \equiv R(\epsilon)$ as an adjacency matrix:



Recurrence Networks (Cont.)



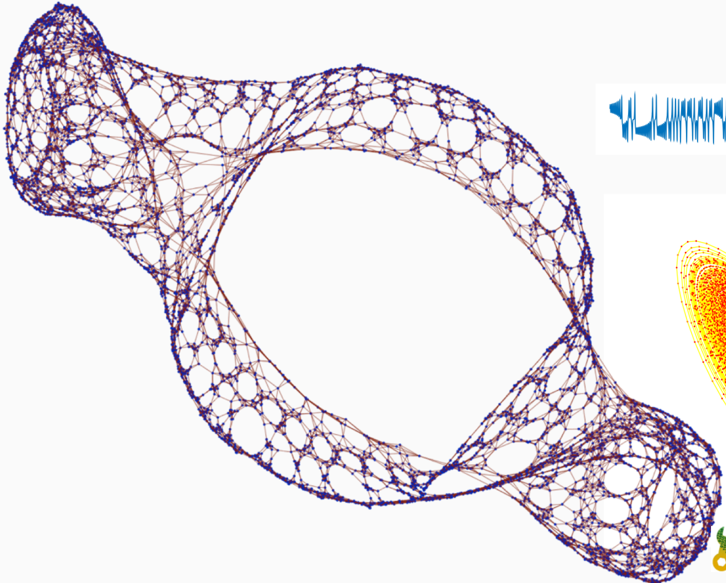


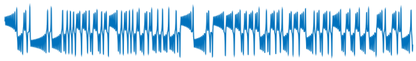
Data Science Transforming Maintenance

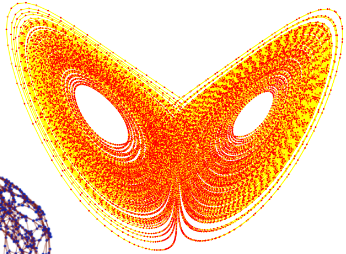
THE UNIVERSITY OF WESTERN AUSTRALIA

TCSG The Complex Systems Group

Recurrence Networks (Cont.)







Complex Network representation of time series

Method	Vertex	Edge	Directedness
Proximity networks			
Cycle networks	Cycle	Correlation or phase space distance between cycles	undirected
Correlation networks	State vector	Correlation coefficient between state vectors	undirected
Recurrence networks			
k-nearest neighbor networks	State (vector)	Recurrence of states (fixed neighborhood mass)	directed
adaptive nearest neighbor networks	State (vector)	Recurrence of states (fixed number of edges)	undirected
ε-recurrence networks	State (vector)	Recurrence of states (fixed neighborhood volume)	undirected

Method	Vertex	Edge	Directedness
Visibility graphs			
natural visibility graphs	Scalar state	Mutual visibility of states	undirected
horizontal visibility graphs	Scalar state	Horizontal mutual visibility of states	undirected
Transition networks			
threshold based networks	Phase space partition	Temporal succession	directed
ordinal pattern networks	Ordinal patterns	Temporal succession	directed

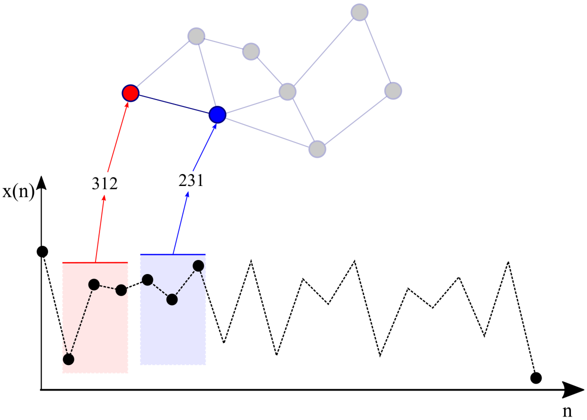
Zou, Y., Donner, R. V., Marwan, N., Donges, J. F., & Kurths, J. (2019). Complex network approaches to nonlinear time series analysis. Physics Reports, 787, 1-97.

Ordinal Partition Networks

Ordinal Partition Networks

Construct a graph with each vertex associated with a permutation π_j, and connect vertices if the corresponding permutations occur in succession.

We are creating a transition network of ordinal patterns.



Ordinal Patterns

Let $\{x(n), n = 1, 2, 3, ...N\}$ denote a scale time series. Consider a embedding dimension m and time-lag τ.

The reconstructed series can be denoted by:

$$\mathbf{X(i)} = \{x(i), x(i + 1\tau), x(i + 2\tau), ..., x(i + (m - 1)\tau)\}$$

For $\tau = 1$, simply:

$$\mathbf{X(i)} = \{x(i), x(i + 1), x(i + 2), ..., x(i + m - 1)\}$$

$\mathbf{X(i)}$ can be mapped to a symbol $(\pi_1, \pi_2, ..., \pi_m)$ where $\pi_k \in \{1, 2, ...m\}$ and

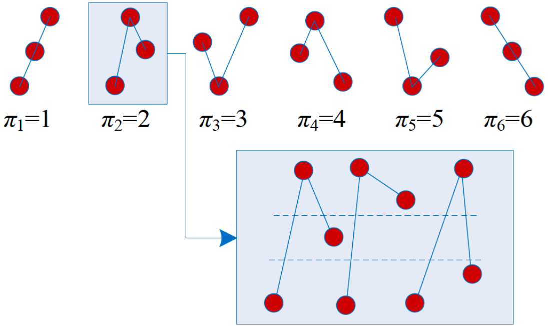
$\pi_k \neq \pi_l \iff k \neq l$ such that $\pi_k < \pi_l \iff x(k) > x(l), \forall x(k), x(l) \in \mathbf{X(i)}$.

Let's define the order pattern of $\mathbf{X(i)}$ as π_j. We can have $m!$ of such patterns.

Ordinal Patterns (Cont.)

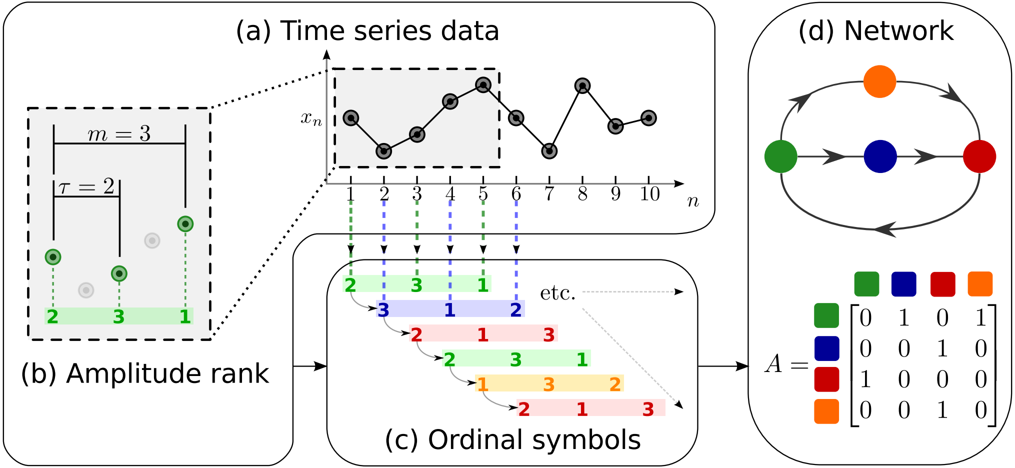
Example, let's consider $m = 3$. We can have six possible patterns:

$\{\pi_1, x_1 \le x_2 \le x_3\}$ $\{\pi_2, x_1 \le x_3 \le x_2\}$ $\{\pi_3, x_2 \le x_1 \le x_3\}$ $\{\pi_4, x_3 \le x_1 \le x_2\}$ $\{\pi_5, x_2 \le x_3 \le x_1\}$ $\{\pi_6, x_3 \le x_2 \le x_1\}$.



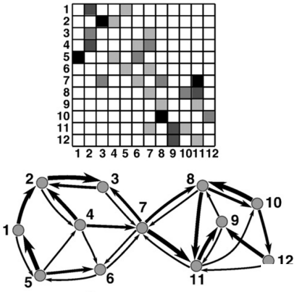
A unique set **s** of all ordinal symbol sequences present in the time series is obtained by giving each pattern π_j a value. Therefore we obtain a pattern series $\{s(i), i = 1, 2, ..., N - m + 1\}$ for $s(i) \in \mathbf{s}$.

Yan, B., He, S., & Sun, K. (2019). Design of a network permutation entropy and its applications for chaotic time series and EEG signals. Entropy, 21(9), 849.

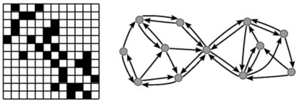

(a) Time series data
x_n
1 2 3 4 5 6 7 8 9 10
n
$m = 3$
$\tau = 2$
2 3 1
(b) Amplitude rank
etc.
(c) Ordinal symbols
(d) Network
$A = \begin{bmatrix} 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$

Types of Networks

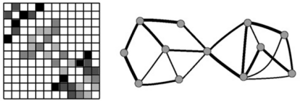
weighted directed networks



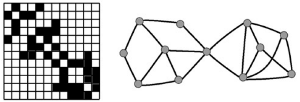
binary directed networks



weighted undirected networks



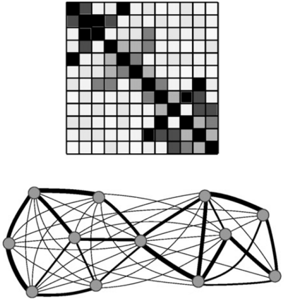
binary undirected networks

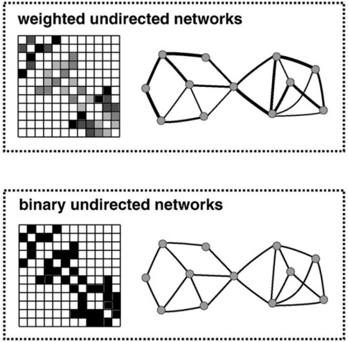


From Rubinov, M., & Sporns, O. (2010). Complex network measures of brain connectivity: uses and interpretations. Neuroimage, 52(3), 1059-1069.

Types of Networks (Cont.)

weighted undirected networks

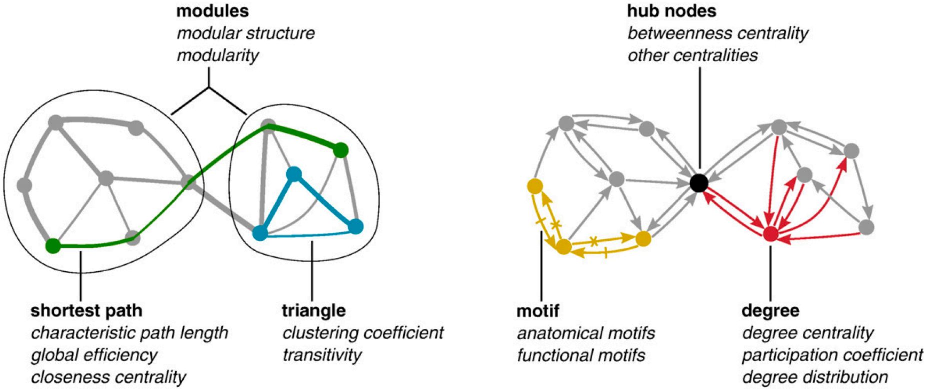




From Rubinov, M., & Sporns, O. (2010). Complex network measures of brain connectivity: uses and interpretations. Neuroimage, 52(3), 1059-1069.

Network Analysis

Describe important properties of complex systems/time-series by characterising topologies of their respective network representations.



From Rubinov, M., & Sporns, O. (2010). Complex network measures of brain connectivity: uses and interpretations. Neuroimage, 52(3), 1059-1069.

Complex Network Analysis - examples of popular measures

Measures of Segregation:

- Clustering Coefficient
- Transitivity
- Modularity

Measures of Resilience:

- Degree Distribution
- Average neighbour degree
- Assortativity coefficient

Measures of Integration:

- Characteristic path length
- Shortest path length
- Global efficiency

Measures of Centrality:

- Closeness centrality
- Betweenness centrality
- Degree centrality
- Eigenvector centrality and variations.

Other concepts:

- Network motifs (patterns of local connectivity)
- K-cycles
- Spectral properties

A good survey on the topic is Costa, L. D. F., Rodrigues, F. A., Travieso, G., & Villas Boas, P. R. (2007). Characterization of complex networks: A survey of measurements. Advances in physics, 56(1), 167-242.

Link Density

Consider $l = \sum_{i,j,i \neq j} A_{ij}$ the number of links in the graph excluding self-loops (links from a node to itself).

For undirected graphs:

$$\rho(G) = \frac{l}{\frac{N(N-1)}{2}} = \frac{2l}{N(N-1)}$$

For directed graphs:

$$\rho(G) = \frac{l}{N(N-1)}$$

where $N = |V|$ is the number of nodes.

The link density is 0 if G has no edges and 1 if G is a complete graph.

Degree Distribution and Average Degree

We take the degree sequence of a graph and count how many vertices have degree k.

We can convert this into a probability p_k, or $p(k)$, which is the probability that a randomly chosen vertex has degree k. Empirically,

$$p(k) = \frac{\#\text{ of vertices with degree } k}{N}.$$

The average degree $\langle k \rangle$ is the first moment of this distribution

$$\langle k \rangle = \sum_{k=0}^{N-1} k p_k.$$

Clustering Coefficient

How densely connected are vertices in a (localized) area of the network?

The clustering coefficient of a vertex calculates the probability for its neighbours to also be each others neighbours.

Essentially, an estimate of the local density of triangle structures in the network.

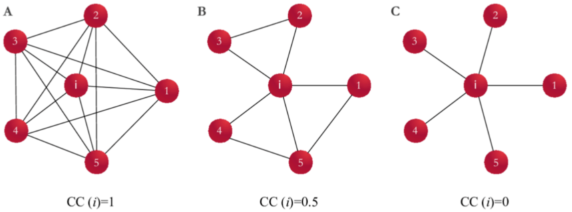


Image from de Arruda, G. F., Rodrigues, F. A., & Moreno, Y. (2018). Fundamentals of spreading processes in single and multilayer complex networks. Physics Reports, 756, 1-59.

Clustering Coefficient (Cont.)

Consider a vertex i with degree k_i

$$CC(i) = \frac{\Delta_i}{k_i(k_i - 1)/2},$$

where Δ_i is the number of triangles (3-cycles) containing vertex i.

The denominator is the total *possible* number of vertex pairs within vertex i's neighbourhood. By convention if $k_i < 2$ then $CC(i) = 0$.

A global network summary of this quantity is the average clustering coefficient given by

$$CC(G) = \frac{\sum_i C(i)}{N}.$$

where N is the number of nodes in the graph.

Transitivity or Global Clustering Coefficient

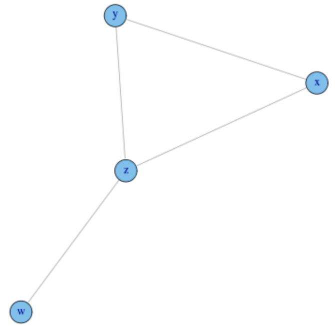
This is the ratio of the number of triangles (actually three times the number) in the network (Δ) divided by the number of triads.

A triad centred at a vertex v is a subgraph of three vertices and two edges where v is incident with the two edges.

$$C_T = \frac{3\Delta}{\#\text{triads}} = \frac{3\Delta}{\sum_i \binom{k_i}{2}}.$$

Transitivity provides information similar to the average

clustering coefficient but treats each triangle, in some sense, more equally.



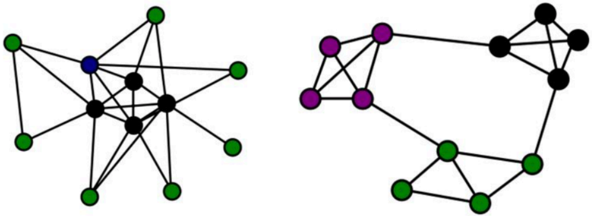
Closeness centrality

The mean distance (in terms of shortest path length) from a vertex to all other vertices.

If we denote the shortest path length between vertex i and vertex j by d_{ij} then the mean shortest path length from vertex i to all other vertices in the network is (considering self-loops):

$$\ell_i = \frac{1}{n} \sum_j d_{ij},$$

It represents how efficiently information is transmitted in the network. We can average over all vertices to have a measure for the whole network.



Closeness centrality (Cont.)

Defined to be the reciprocal of ℓ_i, i.e.,

$$C_i = \frac{1}{\ell_i} = \frac{n-1}{\sum_{j \neq i} d_{ij}}.$$

We can average over all vertices and obtain a closeness centrality value for the network as a whole.

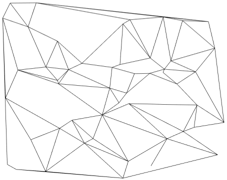
Higher values of closeness indicate higher centrality.

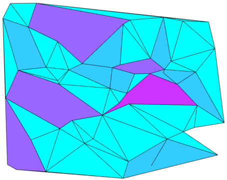
k-cycle distribution

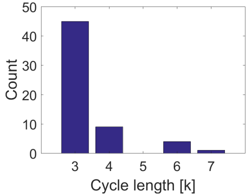
A cycle of length k is a path from a specified node to itself that consist of k links.

Equivalent to self-loops when $k = 1$, bidirectional links when $k = 2$, triangles when $k = 3$, squares when $k = 4$, etc...

-> Minimal cycle basis of the network[5].





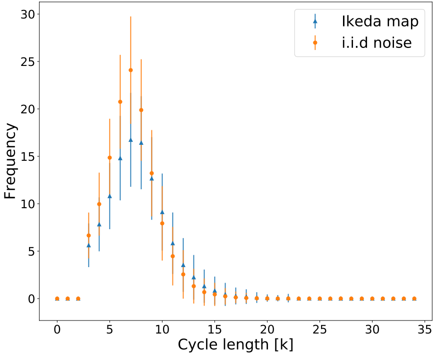


This distribution help to distinguish between different types of dynamics.

[5] Paton, K. An algorithm for finding a fundamental set of cycles of a graph. Comm. ACM 12, 9 (Sept 1969), 514-518.

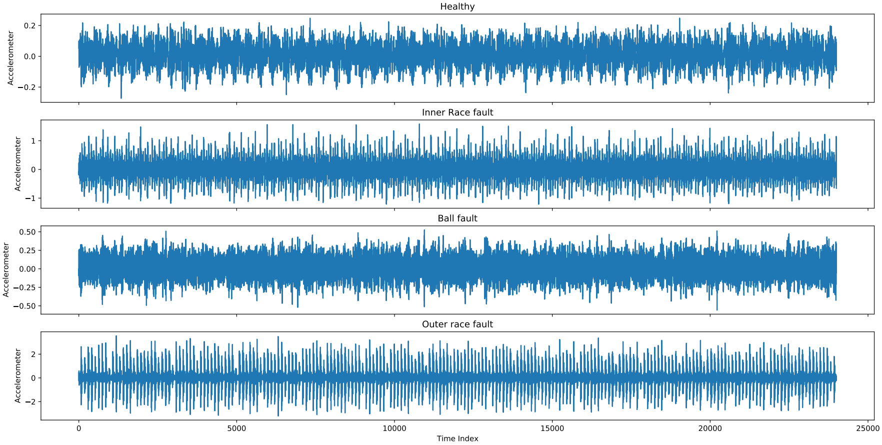
Noise vs chaos? k-cycle distributions: iid noise and Ikeda map

1000 realisations of i.i.d. $N(0,1)$ noise of length 2000 and 1000 orbits of length 2000 of the Ikeda map with different initial conditions.



Kolmogorov-Smirnov test shows the two distributions are different for all k-cycles, $3 \leq k \leq 18$ except for $k = 9$.

Practice time - Characterising normal and faulty modes of bearing data



Practice: Download the notebook “Ordinal Partition Networks - Part 1”. Have a look on what it does and work on the proposed exercises.

Entropy measures

Entropy measures

Quantifiers reflecting the complexity of the dynamics of the time series.

Quantifiers of the randomness in the ordinal pattern dynamics of a time series.

Quantifiers of the average uncertainty of the random variable given by the ordinal patterns.

Quantifiers of the variability or level of “surprise” in the ordinal pattern dynamics.

We will discuss three entropy measures[6]:

- Permutation Entropy;
- Conditional Permutation Entropy;
- Global and Local Node Entropy.

[6] Several other have been proposed. A good reference is Anand, K., & Bianconi, G. (2009). Entropy measures for networks: Toward an information theory of complex topologies. Physical Review E, 80(4), 045102.

Data Science Transforming Maintenance
THE UNIVERSITY OF WESTERN AUSTRALIA
TCSG The Complex Systems Group

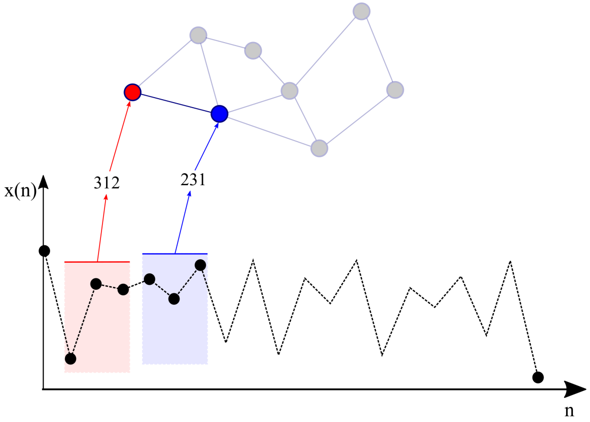
Let’s remember our Ordinal Partition Network

We construct a graph with each vertex associated with a permutation π_j, and connect vertices if the corresponding permutations occur in succession.

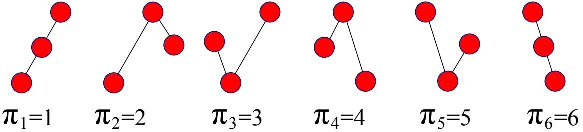
We are creating a transition network of ordinal patterns.

We are interested in quantifying the transitions between permutations that are characterising our dynamics.

The result is a weighted and directed graph.



Permutation Entropy (PE)[8]



Given the probability distribution $p(\pi_j)$:

$$p(\pi_j) = \frac{\#\{s(i) | i \leq N - m + 1; s(i) = \pi_j\}}{N - m + 1}$$

Bandt and Pompe PE is defined as the Shannon entropy of this probability distribution:

$$h^{PE(x^N, m)} = -\sum_{i=1}^{m!} p(\pi_j) log(p(\pi_j))$$

$h^{PE} \approx log(m!)$ accounts for random behaviour; while $h^{PE} \approx 0$ indicates regular dynamics[7].

[7] The logarithms are in base 2, so entropy will be measured in bits.

[8] Bandt, C., & Pompe, B. (2002). Permutation entropy: a natural complexity measure for time series. Physical review letters, 88(17), 174102.

Permutation Entropy (Cont.)

As the maximum value of PE is $log(m!)$ a normalised version of permutation entropy is also used:

$$H(h^{PE}) = \frac{h^{PE}}{log(m!)}$$

The values of $H(h^{PE})$ is now in the interval $[0, 1]$.

To properly estimate the ordinal probability distribution, $m! \ll N$.

Conditional Entropy of ordinal patterns[9]

Extension of the PE that quantifies the local uncertainty of each state:

$$h^{CPE} = \sum_i \left(-p_i \sum_j p_{i,j} log(p_{i,j}) \right)$$

where p_i is the relative frequency of each symbol in **s**, and $p_{i,j}$ is an element of the stochastic matrix P defined as the probability of a transition from symbol (state) $s(i)$ to $s(j)$ estimated from S:

$$p_{i,j} = \frac{a_{i,j}}{\sum_j a_{i,j}}$$

where $a_{i,j}$ are the elements of the adjacency matrix A.

[9] Unakafov, A. M., & Keller, K. (2014). Conditional entropy of ordinal patterns. Physica D: Nonlinear Phenomena, 269, 94-102.

Global and Local Node Entropy

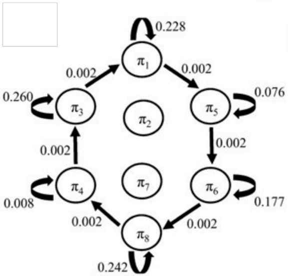
Conditional entropy excluding self-loops.

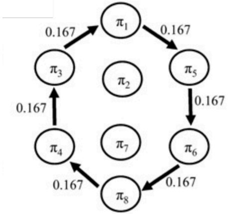
$$p^T_{i,j} = \begin{cases} 0 & \text{if } \ i = j \\ \dfrac{a_{i,j}}{\sum_{j,j\neq i} a_{i,j}} & \text{if } \ i \neq j. \end{cases}$$

$$h^{GNE} = \sum_i p_i h_i^{LNE}$$

where h_i^{LNE} is the local node entropy defined as:

$$h_i^{LNE} = -\sum_j p^T_{i,j} \log p^T_{i,j}.$$





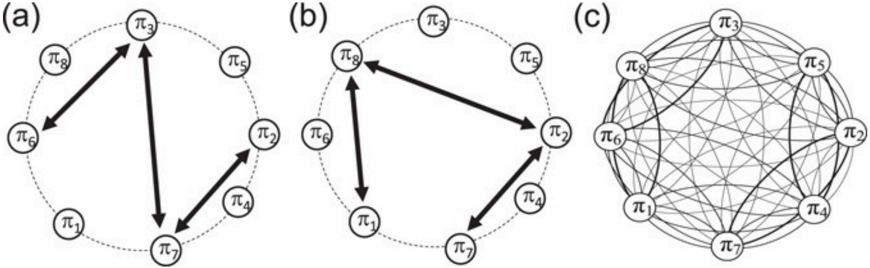
Pictures from Zhang, J., Zhou, J., Tang, M., Guo, H., Small, M., & Zou, Y. (2017). Constructing ordinal partition transition networks from multivariate time series. Scientific reports, 7(1), 1-13.

Data Science Transforming Maintenance
THE UNIVERSITY OF WESTERN AUSTRALIA
TCSG The Complex Systems Group

Forbidden Patterns

Properties related to the absence of particular ordinal patterns (missing sequences that do not occur in a deterministic time series) can also characterise the dynamics of the system.

Useful to discriminate between deterministic and stochastic dynamics.

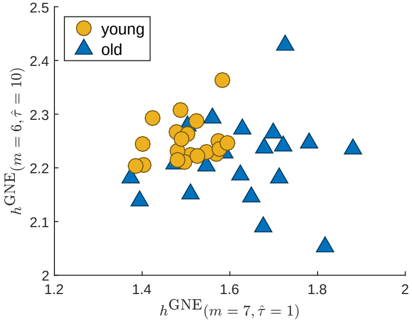


Pictures from Zhang, J., Zhou, J., Tang, M., Guo, H., Small, M., & Zou, Y. (2017). Constructing ordinal partition transition networks from multivariate time series. Scientific reports, 7(1), 1-13.

Application to data

Electrocardiogram (RR-intervals)

Global node entropy for different time scale (τ) and m.

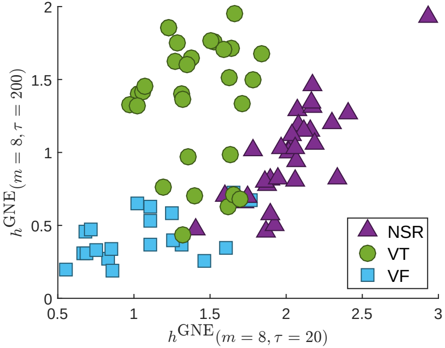


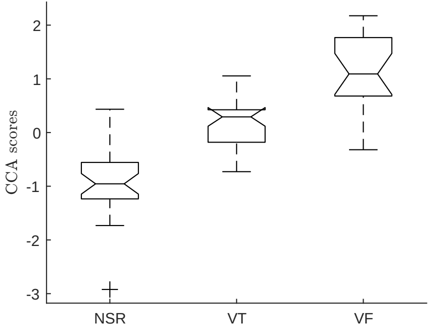
Motivation: Investigate alterations in the fractal scaling of cardiac interbeat interval dynamics that are related to age.

McCullough, M., Small, M., Iu, H. H. C., & Stemler, T. (2017). Multiscale ordinal network analysis of human cardiac dynamics. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 375(2096), 20160292.

Ventricular Arrhythmia (Electrocardiogram)

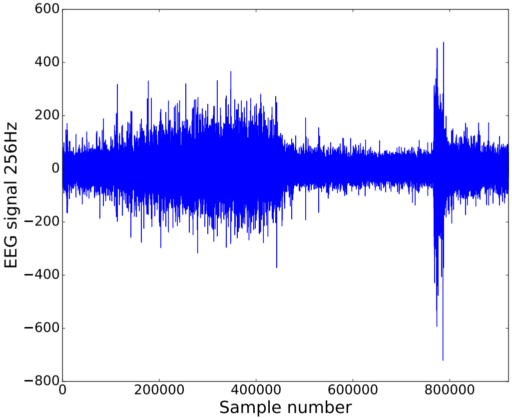
Global node entropy for different time scale (τ) and m.

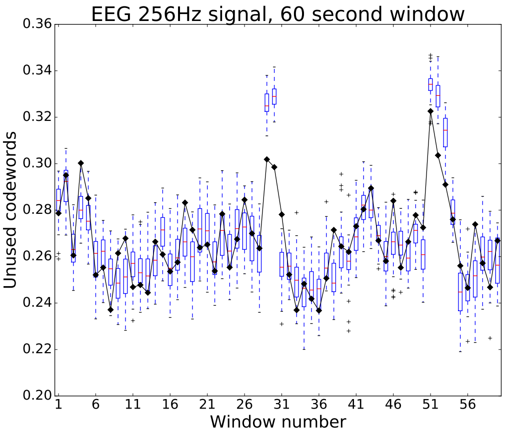




McCullough, M., Small, M., Iu, H. H. C., & Stemler, T. (2017). Multiscale ordinal network analysis of human cardiac dynamics. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 375(2096), 20160292.

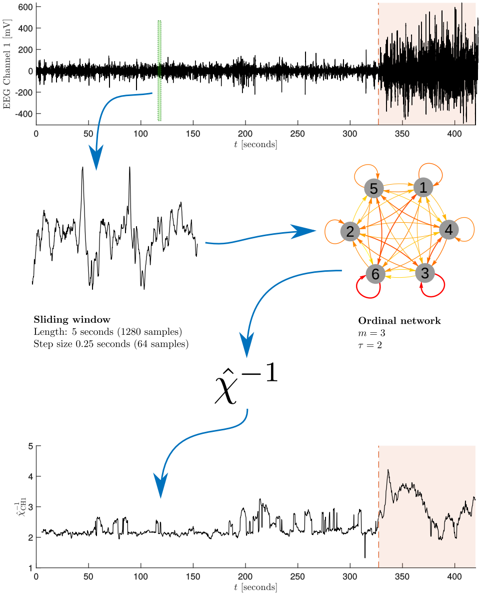
Onset of change in EEG scans

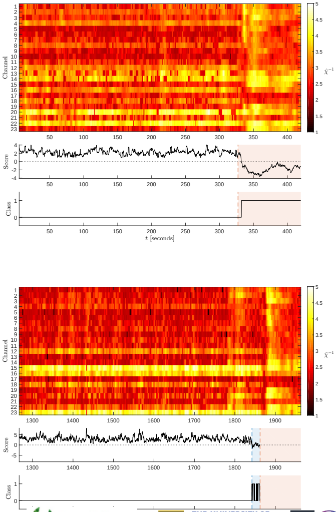




Data available at *`https://physionet.org/physiobank/database/chbmit`*

Epileptic seizure detection (Electroencephalogram)





Practice time

Part 1:

Download the notebook “Ordinal Partition Networks - Part 2”. Have a look on what it does and work on the proposed exercises.

Practice time

Part 2:

Download the data “df.csv”. It contains several sensor measurements and several failure events along time. However, only some of the sensors contain relevant dynamic information anticipating the events.

Practice suggestion: Apply the techniques discussed today to identify which sensors are relevant for detecting changes before the event happens.

Further reading

References

McCullough, M., Small, M., Iu, H. H. C., & Stemler, T. (2017). Multiscale ordinal network analysis of human cardiac dynamics. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 375(2096), 20160292.

Zou, Y., Donner, R. V., Marwan, N., Donges, J. F., & Kurths, J. (2019). Complex network approaches to nonlinear time series analysis. Physics Reports, 787, 1-97.

Rubinov, M., & Sporns, O. (2010). Complex network measures of brain connectivity: uses and interpretations. Neuroimage, 52(3), 1059-1069.

Bandt, C., & Pompe, B. (2002). Permutation entropy: a natural complexity measure for time series. Physical review letters, 88(17), 174102.

Yan, B., He, S., & Sun, K. (2019). Design of a network permutation entropy and its applications for chaotic time series and EEG signals. Entropy, 21(9), 849.

Unakafov, A. M., & Keller, K. (2014). Conditional entropy of ordinal patterns. Physica D: Nonlinear Phenomena, 269, 94-102.

Data Science Transforming Maintenance
THE UNIVERSITY OF WESTERN AUSTRALIA
TCSG The Complex Systems Group

References

Zhang, J., Zhou, J., Tang, M., Guo, H., Small, M., & Zou, Y. (2017). Constructing ordinal partition transition networks from multivariate time series. Scientific reports, 7(1), 1-13.

Sakellariou, K., Stemler, T., & Small, M. (2019). Markov modeling via ordinal partitions: An alternative paradigm for network-based time-series analysis. Physical Review E, 100(6), 062307.

Zou, Y., Donner, R. V., Marwan, N., Donges, J. F., & Kurths, J. (2019). Complex network approaches to nonlinear time series analysis. Physics Reports, 787, 1-97.

Amigó, J. M., Zambrano, S., & Sanjuán, M. A. (2007). True and false forbidden patterns in deterministic and random dynamics. EPL (Europhysics Letters), 79(5), 50001.

Unakafov, A. M., & Keller, K. (2014). Conditional entropy of ordinal patterns. Physica D: Nonlinear Phenomena, 269, 94-102.

Thank you!

Contact:

Prof. Michael Small
michael.small@uwa.edu.au

Dr. Ayham Zaitouny
ayham.zaitouny@uwa.edu.au

Dr. Débora Corrêa
debora.correa@uwa.edu.au

Acknowledgement

This work is supported by the Australian Research Council through the Industrial Transformation Training Centre for Transforming Maintenance through Data Science (grant number IC180100030) funded by the Australian Government and our industry partners.