Nonlinear time series analysis with the TISEAN package

Eckehard Olbrich

MPI MIS Leipzig

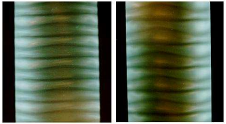
Potsdam SS 2008

Deterministic chaos

Deterministic chaos: Aperiodic irregular temporal behaviour in low-dimensional deterministic systems

Nonlinear time series analysis: A battery of methods to characterize, predict and model time series as nonlinear deterministic systems instead as linear stochastic systems as in the "classical" time series analysis

Successful applications: Electric circuits, Laser, Hydrodynamic experiments (Taylor–Couette flow), ...



Unsuccessful attempts: Climate data, EEG data, finance data, ...

The henon map

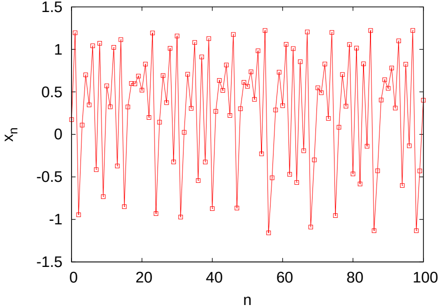
The logistic map:

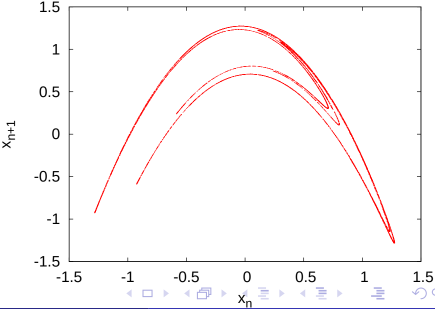
$$x_{n+1} = 1 - ax_n^2$$

Add a second degree of freedom in order to make it invertible:

$$\begin{aligned} x_{n+1} &= 1 - ax_n^2 + by_n \\ y_{n+1} &= x_n \end{aligned}$$

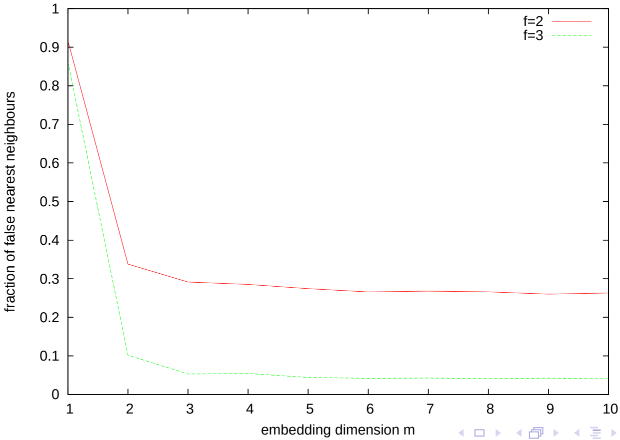
Example: 10 000 data points with *henon* - delay plot with *delay*





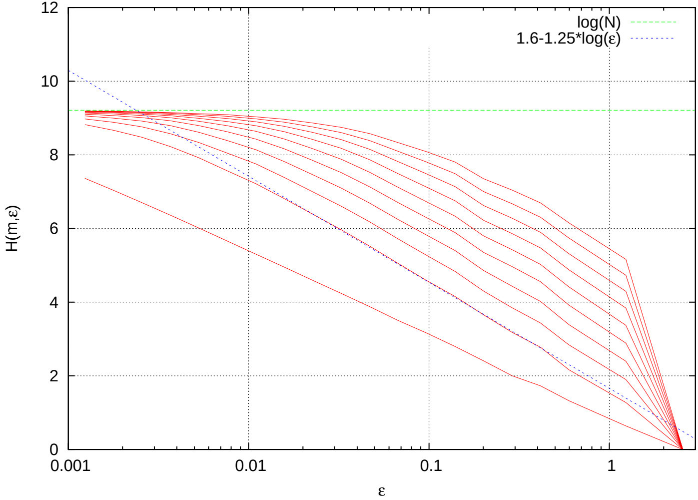
Embedding - False nearest neighbours with *false_nearest*

- Delay vector: $\mathbf{x}_n = (x_n, x_{n-1}, x_{n-2}, \ldots, x_{n-m+1})$
- False nearest neighbour: next neighbour in m dimensions, distance increases by more than the factor f in $m+1$ dimensions



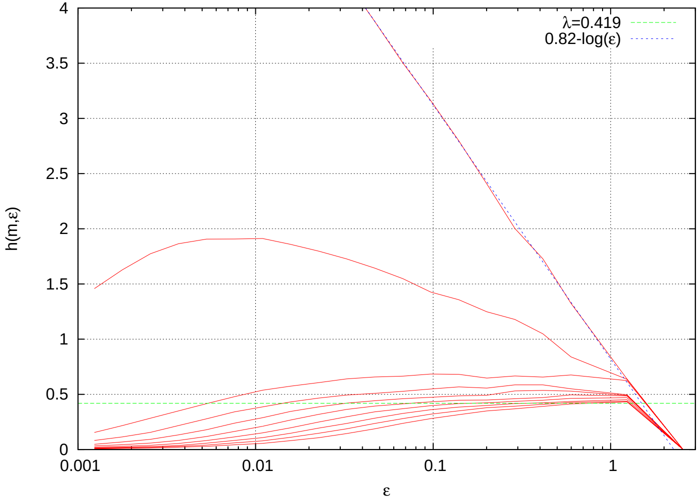
Estimating entropies with *boxcount*

Block entropies $H_m(\epsilon)$



Estimating entropies with *boxcount*

Conditional entropies $h_m(\epsilon) = H_m(\epsilon) - H_{m+1}(\epsilon)$



Correlation sum and dimension - *d2*

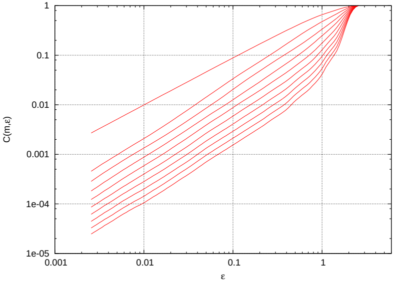
- Time series $\{x(t_0), x(t_0+\Delta), \ldots, x(t_0+N\Delta)\}$
- Delay embedding $\boldsymbol{x} = (x_t, x_{t-d\Delta}, \ldots, x_{t-(m-1)d\Delta})$
- Correlation sum

$$\begin{aligned} C(m,\epsilon) &= \frac{2}{(N-m)\cdot(N-m-1)} \sum_{i=1}^{N-m} \sum_{j=i+1}^{N} \Theta(\epsilon - ||\vec{X}_i - \vec{X}_j||) \\ &\propto \epsilon_2^D \end{aligned}$$

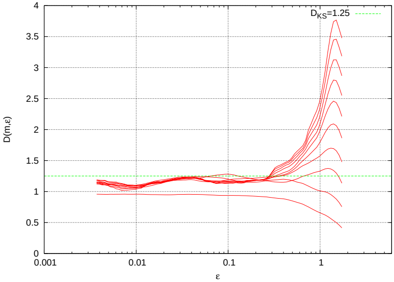
- Correlation dimension D_2

$$D_2 = \lim_{\epsilon \to 0} D_2(m,\epsilon) = \lim_{\epsilon \to 0} \frac{d \log C(m,\epsilon)}{d \log \epsilon}$$

Correlation sum and dimension - *d2*

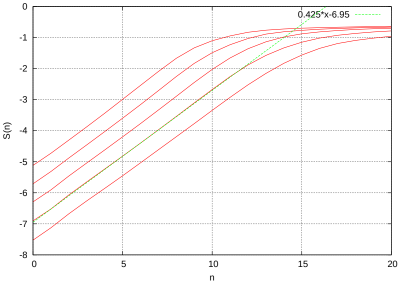


correlation sum



correlation dimension

Estimating the largest Lyapunov exponent using *lyap_k*



Stretching factor vs. time steps for different neighborhood sizes

$\lambda_{max} = 0.425$

Lyapunov exponents from the system equations:

$\lambda_1 = 0.419$, $\lambda_2 = -1.62$, $D_{KY} = 1.26$

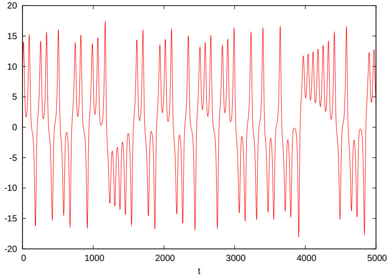
Lyapunov exponents from fitting the Jacobian with *lyap_spec*:

$\lambda_1 = 0.41$, $\lambda_2 = -1.55$, $D_{KY} = 1.27$

Flow data — the Lorenz system

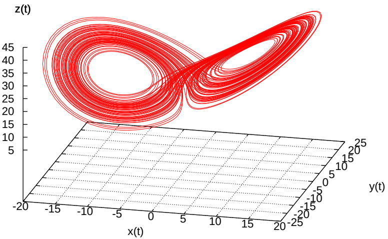
Lorenz equations:

$$
\begin{aligned}
\dot{x} &= s(-x+y) \\
\dot{y} &= -xz + rx - y \\
\dot{z} &= xy - bz
\end{aligned}
$$



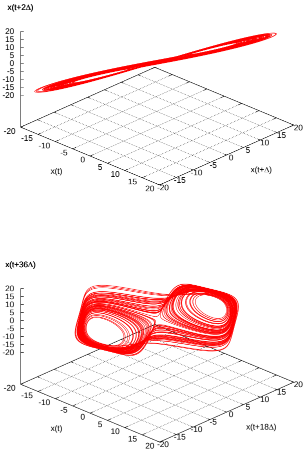
Integrated with *lorenz* at standard parameters

$$
\begin{aligned}
s &= 10 \\
r &= 28 \\
b &= 8/3
\end{aligned}
$$



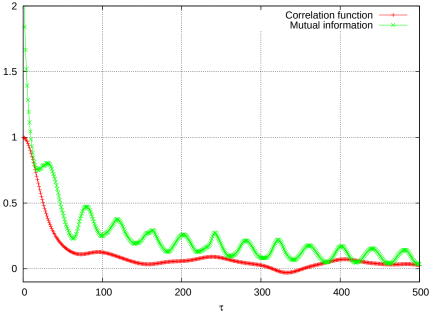
Embedding — finding the optimal delay time

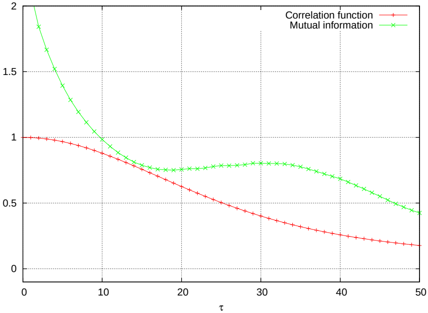
Delay embedding of the x-component



Embedding — finding the optimal delay time

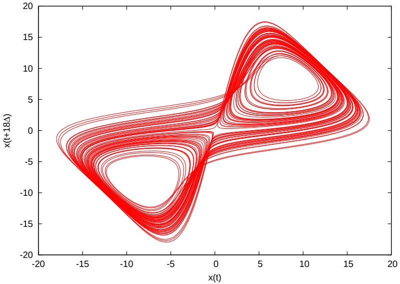
Rule of thumb: First zero of the autocorrelation function (*corr*) or first minimum of the mutual information function (*mutual*)





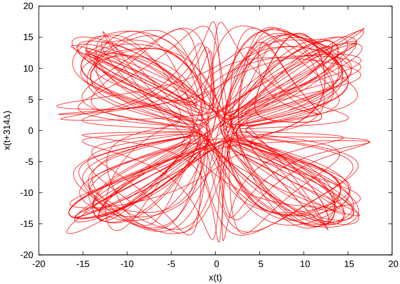
First zero of the autocorrelation function at 314 time steps. First minimum of the mutual information function at 18 time steps.

Embedding — finding the optimal delay time



Mutual information

$d = 18$



Correlation function

$d = 314$

What is an optimal state space?

One possibility - maximize the scaling range in dimension estimates.
Let us consider the correlation dimension:

$$C(\epsilon) = \frac{2}{N(N-1)} \sum_{i=1}^{N} \sum_{j=i+1}^{N} \Theta(\epsilon - ||\mathbf{x}_i - \mathbf{x}_j||)$$

Scaling range $\epsilon_l \leq \epsilon \leq \epsilon_u$
For $m > D_0$

$$C(\epsilon) \propto \epsilon^{D_2}$$

Lower end of the scaling range in the noise free case:

$$C(\epsilon_l) = \frac{2}{N(N-1)}$$

What about the upper bound?
Correlation sum approximates Renyi entropies of second order

$$-\log C_m(\epsilon) \approx H_m^{(q=2)}(\epsilon)$$

For $q = 1$ we know $H_m(\epsilon) \geq mh_{KS}$. Assuming the same for $q = 2$ and calling the corresponding entropy h_C one gets

$$\begin{aligned} -logC_m(\epsilon_u) &\geq mh_C \\ -logC_m(\epsilon_l) &\geq mh_C - D_2 \log \frac{\epsilon_l}{\epsilon_u} \end{aligned}$$

Estimating the scaling range

What about the upper bound?

$$-logC_m(\epsilon_l) \geq mh_C - D_2 \log \frac{\epsilon_l}{\epsilon_u}$$

and therefore

$$\boxed{\frac{\epsilon_u}{\epsilon_l} \leq \left(e^{-mh_C \frac{N^2}{2}}\right)^{1/D_2} .}$$

In order to get a scaling in one decade of ϵ:

$$N \geq \sqrt{2} \cdot 10^{D_2/2} \cdot e^{mh_C/2}$$

For $D_2 = 4,\ m = 5$ and $h_C = \ln 2$ we would get $N > 4525$.

Maximizing the scaling range

Maximizing the scaling range $\Rightarrow$ Minimizing $H(\epsilon_u)$.

$$H_m(\epsilon_u) = m h_{m-1}(\epsilon) + \sum_{n=1}^{m-1} n \delta h_n(\epsilon)$$

with $\delta h_n(\epsilon) = h_{n-1}(\epsilon) - h_n(\epsilon)$ and $h_0(\epsilon) = H_1(\epsilon)$. In a first approximation

$$H_m(\epsilon_u) = m h_{KS} + MI \quad MI = \lim_{\epsilon \to 0} MI(\epsilon)$$

with the mutual information $MI = \delta h_1$. For different delays τ we have

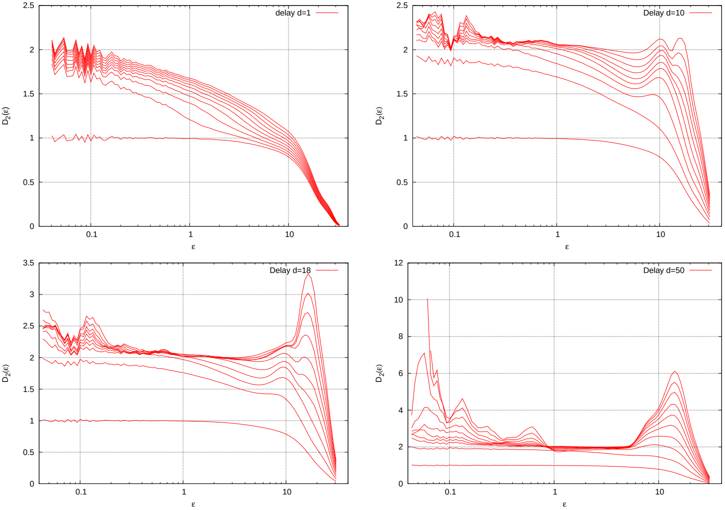
$$h_{KS} = \tau \tilde{h}_{KS}$$

thus minimizing $H_m(\epsilon)$ leads to

$$\left.\frac{dMI(\tau)}{d\tau}\right|_{\tau=\tau_{opt}} = -m\tilde{h}_{KS} \qquad \left.\frac{d^2 MI(\tau)}{d\tau^2}\right|_{\tau=\tau_{opt}} > 0$$

For $\tilde{h}_{KS} = 0$ this gives the first minimum of the mutual information.

Correlation dimension for different delays

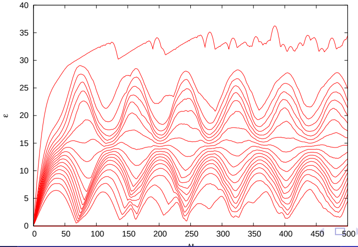


For finite data temporal correlations might lead to a based estimate of the correlation sum. Theiler correction: remove W temporal neighbours from the correlation sum

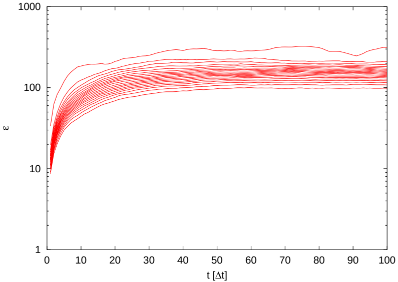
$$C(\epsilon) = \frac{2}{N - W(N - W - 1)} \sum_{i=1}^{N} \sum_{j=i+W+1}^{N} \Theta(\epsilon - ||\mathbf{x}_i - \mathbf{x}_j||)$$

Space-time separation plot: plots cumulative distribution for distances ϵ conditioned on the temporal differences.

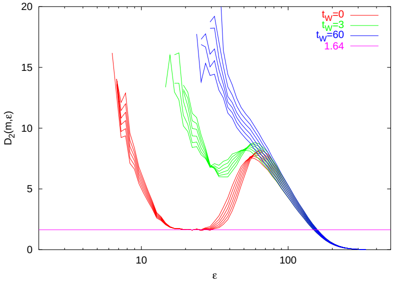
Space-time separation plot (*stp*) for the Lorenz system (m=3,d=18)



EEG — spurious dimension due to temporal correlations



Space-time separation plot



Correlation dimension

Prediction and Modeling

$$\boldsymbol{x}_{n+1} = \boldsymbol{F}(\boldsymbol{x}_{n+1})$$

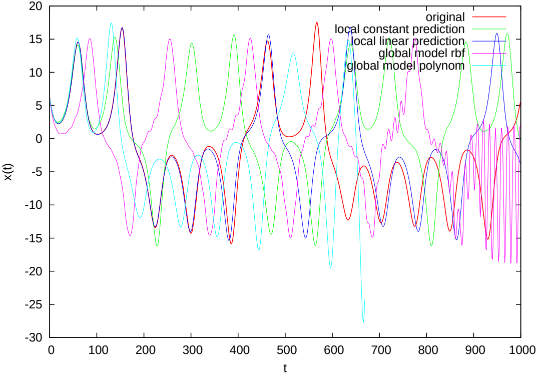
which reduces to

$$x_{n+1} = F(x_n, \ldots, x_{n-m+1})$$

in the case of delay embedding.

Local methods: Basic idea: Looking for similar events in the past and using their future for prediction.
Local constant (*lzo-run*): Using the average as prediction.
Local linear (*lfo-run*): Fitting a linear model for similar events, i.e. neighbors in phase space.

Global models: Parameterizing the function $\boldsymbol{F}$ and fitting the parameters.
Polynomials (*polynom*)
Radial basis functions (*rbf*) — related to neural networks
(not included in TISEAN)



Relative forecast errors:

lfo	1.12e-04
poly	8.28e-04
lzo	1.25e-02
rbf	4.03e-02

Polynomial fit used 4th order polynomials.