Nonlinear stochastic systems

Eckehard Olbrich

MPI MIS Leipzig

Potsdam SS 2008

Measurement noise and dynamical noise

Measurement noise: Here the dynamical system is still deterministic

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

but observed via a noisy channel

$$\boldsymbol{y}_n = h(\boldsymbol{x}_n) + \boldsymbol{\nu}_n \ .$$

This case is dealt with by **noise reduction** methods.

Dynamical noise:

$$\boldsymbol{x}_{n+1} = \boldsymbol{F}(\boldsymbol{x}_n, \boldsymbol{\xi}_n) \ .$$

Here we have a qualitative different dynamics. The noise term can stand for high-dimensional and/or high-entropic processes. In the following we will study under which conditions we can apply the mthods developed for deterministic systems also in this case.

Estimating the noise level

Why should we know the noise level?

- Is there any visible noise at all?
- Caracterisation of the data
- Determining the neighbourhoodsize for noise reduction and/or local constant or local linear prediction

Estimating the noise level from entropies - Differential entropy

Literature: Cover/Thomas, Elements of Information Theory, Chapter 9

- Differential entropy

$$H(\boldsymbol{x}) = -\int d\boldsymbol{x} p(\boldsymbol{x}) \ln p(\boldsymbol{x})$$

- Example: Differential entropy of a uniform distribution in $[0, a]$, $p(x) = 1/s$.

$$H(x) = \ln a$$

- Example: Gaussian distributed random variable with standard deviation σ.

$$\begin{aligned} p(x) &= \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2}{2\sigma^2}} \\ H(x) &= \frac{1}{2}(1 + \ln 2\pi\sigma^2) \end{aligned}$$

Maximum entropy property of the Gaussian distribution

- Of all probability densities $p(x)$ with a given standard deviation σ the Gaussian distribution has maximal entropy. Thus

$$H(x) \leq \frac{1}{2}(1 + \ln 2\pi\sigma^2)$$

  for an arbitrary distribution $p(x)$ with standard deviation σ.
- Therefore the differential entropy $H(x)$ provides a lower bound for the standard deviation:

$$\sigma \geq \frac{1}{\sqrt{2\pi}} e^{H(x)-1/2}$$

Transformation rule for the differential entropy

- Transformation rule for an invertible transformation $\boldsymbol{y} = \boldsymbol{f}(\boldsymbol{x})$

$$H(\boldsymbol{y}) = H(\boldsymbol{x}) + \int d\boldsymbol{x} p(\boldsymbol{x}) \ln |\boldsymbol{J}(\boldsymbol{x})|$$

because $q(\boldsymbol{y})d\boldsymbol{y} = p(\boldsymbol{x})d\boldsymbol{x}$ and therefore

$$q(\boldsymbol{y}) = p(\boldsymbol{x}) \left| \frac{d\boldsymbol{f}}{d\boldsymbol{x}}^{-1} \right|_{\boldsymbol{x}=\boldsymbol{f}^{-1}(\boldsymbol{y})} = p(\boldsymbol{x})|\boldsymbol{J}(\boldsymbol{x})|$$

Differential entropy of the noise

- Stochastic system:

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

  In the following $\boldsymbol{x}_n = (x_n, \ldots, x_{n-m})$.
- Defines a map $\boldsymbol{f} : (\boldsymbol{x}_n, \xi_n) \mapsto (\boldsymbol{x}_n, x_{n+1})$.
- Transformation rule

$$H(x_{n+1}, \boldsymbol{x}_n) = H(\boldsymbol{x}_n, \xi_n) + \langle \ln \left| \frac{\partial \boldsymbol{f}}{\partial \xi} \right| \rangle\ .$$

- Conditional entropy

$$\begin{aligned} H(x_{n+1}|\boldsymbol{x}_n) &= H(x_{n+1}, \boldsymbol{x}_n) - H(\boldsymbol{x}_n) \\ &= H(\boldsymbol{x}_n, \xi) - H(\boldsymbol{x}_n) + \langle \ln \left| \frac{\partial \boldsymbol{f}}{\partial \xi} \right| \rangle \\ &= H(\xi) + \langle \ln \left| \frac{\partial \boldsymbol{f}}{\partial \xi} \right| \rangle \end{aligned}$$

Differential entropy of the noise

- Stochastic system:

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

  In the following $\boldsymbol{x}_n = (x_n, \ldots, x_{n-m})$.
- Defines a map $\boldsymbol{f} : (\boldsymbol{x}_n, \xi_n) \mapsto (\boldsymbol{x}_n, x_{n+1})$.
- Transformation rule

$$H(x_{n+1}, \boldsymbol{x}_n) = H(\boldsymbol{x}_n, \xi_n) + \langle \ln \left| \frac{\partial \boldsymbol{f}}{\partial \xi} \right| \rangle\ .$$

- Conditional entropy

$$H(x_{n+1} | \boldsymbol{x}_n) = H(\xi) + \langle \ln \left| \frac{\partial \boldsymbol{f}}{\partial \xi} \right| \rangle$$

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

$$H(x_{n+1} | \boldsymbol{x}_n) = H(\xi_n)$$

Noise amplitude from the conditional entropy

- Additive dynamical noise: Asymptotic behaviour of the conditional entropy only depends on the noise!
- Lower bound of the effective noise amplitude

$$\sigma \geq \frac{1}{\sqrt{2\pi}} e^{H(x_{n+1}|\boldsymbol{x}_n)-1/2}$$

- Resolution dependent entropy

$$H(x_{n+1}|\boldsymbol{x}_n;\epsilon) = H(x_{n+1}|\boldsymbol{x}_n) - \ln\epsilon + \mathcal{O}(\epsilon)$$

- A similar, but less rigorous reasoning is possible for the correlation entropies based on the correlation sum

$$H^{(2)}(\boldsymbol{x};2\epsilon) = -\ln C(\boldsymbol{x};\epsilon)$$

Noise level from the correlation entropies

- Using $H(\boldsymbol{x}; 2\epsilon) = -\ln C(\boldsymbol{x}; \epsilon)$ we get for the condtional correlation entropies

$$h^{(2)}(m, \epsilon) = lnC(m, \epsilon/2) - lnC(m+1, \epsilon/2)$$

contained in the output files **.h2* of *d2*.

- The differential Renyi order $q = 2$ entropy for a probability density $p(\boldsymbol{x})$ is defined as

$$H^{(2)}(\boldsymbol{x}) = -\ln \int d\boldsymbol{x} p(\boldsymbol{x})^2$$

and therefore for the Gaussian distribution

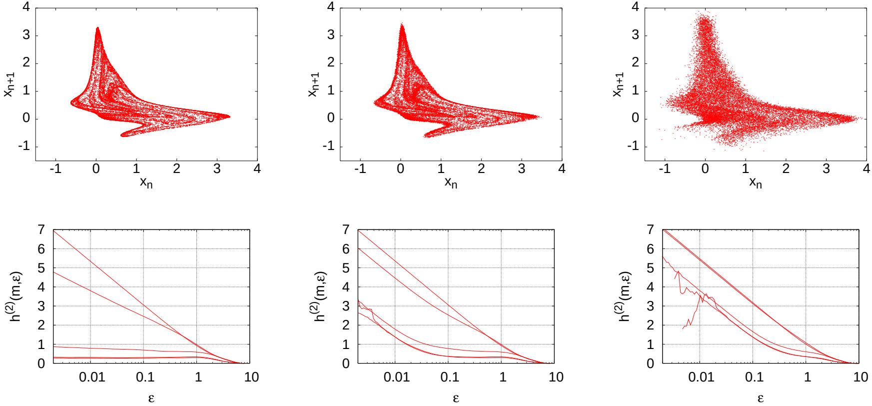
$$H^{(2)}_{Gauss} = \frac{1}{2}\ln(4\pi) + \ln\sigma$$

- Approximate estimation of the noise amplitude (only lower bound in the case of additive noise)

$$\sigma_{noise} \approx \frac{1}{\sqrt{4\pi}} e^{h^{(2)}(m,\epsilon)+\ln\epsilon}$$

Example —Exponential map

deterministic | $\sigma = 0.01$ | $\sigma = 0.1$



Estimating the noise level - measurement noise

Schreiber (1993,1997) proposed a way to estimate the noise level from the correlation dimension.
Using the maximum norm to calculate the distances, one gets

$$C(d,\epsilon) = C(m,\epsilon)[\sqrt{2}\mathrm{erf}(\epsilon/2\sigma)]^{d-m}$$

with $d > m$ being embedding dimensions and σ being the standard deviation of the (assumed) Gaussian measurement noise and $\mathrm{erf}(z) = 2/\sqrt{\pi}\int_0^z e^{-x^2}dx$.
With

$$D(m,\epsilon) = \frac{d}{\ln\epsilon}C(m,\epsilon)$$

one gets

$$D(d,\epsilon) = D(m,\epsilon) + \frac{d-m}{\sigma\sqrt{\pi}}\frac{\epsilon e^{-\epsilon^2/4\sigma^2}}{\mathrm{erf}(\epsilon/2\sigma)}$$

Estimating the noise level - measurement noise

$$D(d,\epsilon) = D(m,\epsilon) + \frac{d-m}{\sigma\sqrt{\pi}} \frac{\epsilon e^{-\epsilon^2/4\sigma^2}}{\mathrm{erf}(\epsilon/2\sigma)}$$

can be rewritten as

$$\begin{aligned} \frac{D(d,\epsilon) - D(m,\epsilon)}{d-m} &= \frac{\epsilon}{\sigma\sqrt{\pi}} \frac{\epsilon e^{-\epsilon^2/4\sigma^2}}{\mathrm{erf}(\epsilon/2\sigma)} \\ &= g(\epsilon/\sigma) \end{aligned}$$

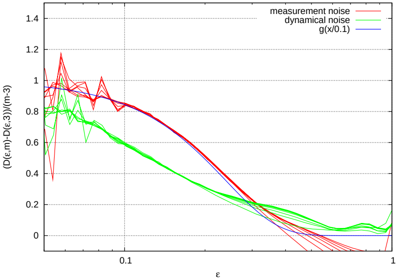
with

$$g(x) = \frac{1}{\sqrt{\pi}} \frac{x e^{-x^2/4}}{\mathrm{erf}(x/2)}$$

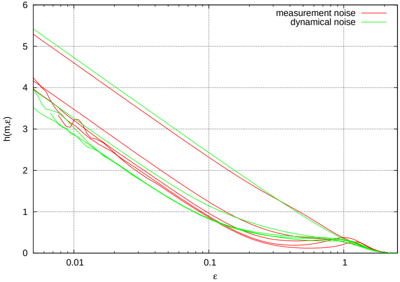
The noise level can then be estimated by fitting the function $g(\epsilon/\sigma)$ to $\frac{D(d,\epsilon) - D(m,\epsilon)}{d-m}$.

Dynamical vs. Measurement noise

Dimension



Entropies



Linear vs. Nonlinear models

If a noisy time series is given one might be interested, whether it is useful to apply non-linear methods or whether linear methods are sufficient.

- Local linear vs. global linear (*lfo-ar*)
- Surrogate data tests: Testing against a (in principle arbitrary) null hypothesis by producing an ensemble of data according to this null hypothesis.
- Null hypothesis for the test against "Linearity": linear stochastic system observed with a nonlinear measurement function (allowing for Non-Gaussian distribution of the data)

$$\begin{aligned} \mathbf{x}_n &= \mathbf{A}\mathbf{x}_{n-1} + \boldsymbol{\xi}_n \\ \mathbf{y}_n &= \mathbf{C}\mathbf{x}_n + \boldsymbol{\nu}_n \\ s_n &= h(\mathbf{y}_n) \end{aligned}$$

Surrogate data test

1. Generate an ensemble of surrogate data sets according to your null hypothesis
2. Evaluate a test statistics on your original data set and on the surrogate data
3. If the value on your data is significantly different from the values on the surrogate data sets you can reject the null hypothesis.

Possible test statistics:

- Prediction error (*predict* — local constant prediction)
- Dimension
- Entropies or entropy based quantities
- time reversal asymmetry (*timerev*)

$$\frac{\langle (x_n - x_{n-d})^3 \rangle}{\langle (x_n - x_{n-d})^2 \rangle}$$

FFT surrogates by phase randomization

1. Estimate the discrete Fourier transform (DFT) of your data, usually done by FFT
$$x_k = a_k exp^{i\phi_k}$$
2. Randomize the phases ϕ_k according to a uniform distribution
3. Transform back into time domain
4. Surrogates have the same sample spectrum estimate as the original data.
5. *TISEAN: surrogates -S -i0*

Problems:

- If the marginal distribution of the data is not a Gaussian, the surogates have a different distribution than the original data $\Rightarrow$ Null hypothesis involves Gaussianity.

Amplitude adjusted Fourier (AAFT) surrogates

After randomizing the phases the amplitudes are readjusted to match the exact distribution:

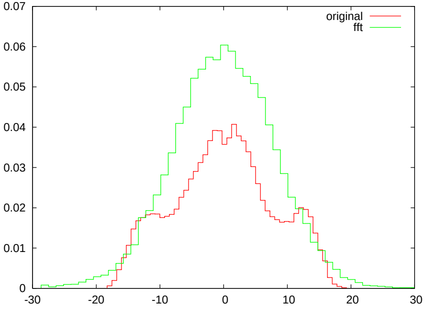
1. Rank order both your data and the FFT surrogate
2. Replace in the surrogate each data point with the data point of the same rank.
3. *TISEAN: surrogates -i1*

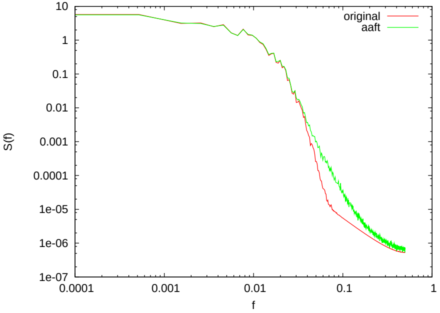
Problem: This amplitude adjustment does not conserve the spectrum.

Iterated amplitude adjusted Fourier (IAAFT) surrogates

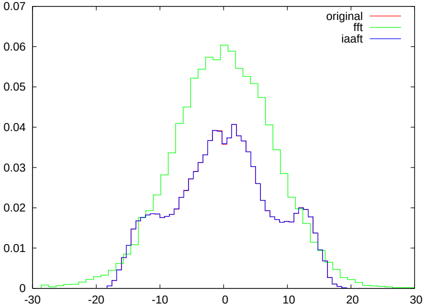
- Produce FFT surrogates and save the Fourier amplitudes
- Perform the amplitde adjustment
- Make the Fourier transform, replace the Fourier amplitudes a_k by the amplitudes from step 1. and transform back
- Iterate 2. and 3. until no improvement can be seen.
- *TISEAN: surrogates*, option *-S* generates surrogates with the exact spectrum estimate, without with the exact distribution.

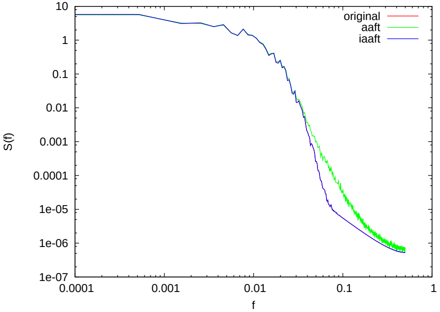
Example - Lorenz



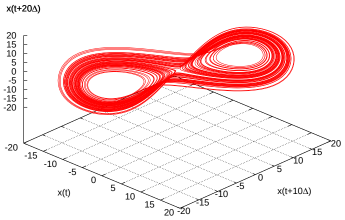


Example - Lorenz

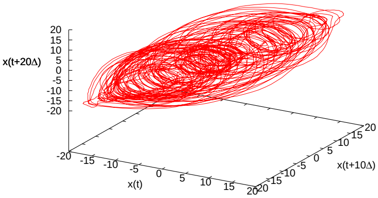




original

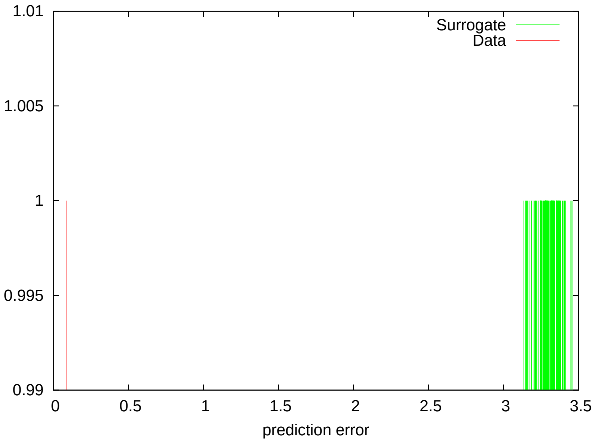


surrogate



Example - Lorenz

Test statistics: prediction error for local constant prediction, $m = 5$, $d = 10$ with *predict*

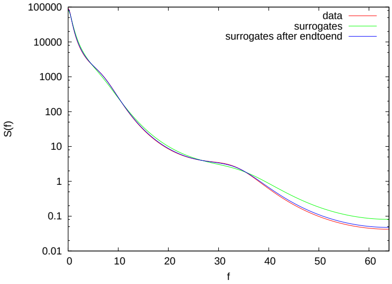


Problem whith the actual implementation

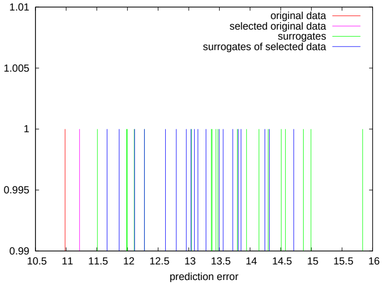
- Surrogate data have approximately the same FFT based sample estimate of the power spectrum. This might be a bad estimate of the power spectrum of the underlying process (remember windowing)
- Possible improvement: use *endtoend* to minimize discontinuities in the periodic continuation
- This does not help in all situations. There might be problems in particular in the high frequency range, which might be reflected in the predictability.

Data set generated by an AR(5) process.

Spectrum estimated by fitting an AR(5) model



Prediction errors



A stochastic proess is called Markovian if for $t > t_k$

$$p(x_t|x_{t_1}, x_{t_2}, \ldots, x_{t_n}) = p(x_t|x_{t_1})$$

i.e. conditioned on a state in the past at t_1, the actual state is independent on all states at $t < t_1$. It can be tested by the Chapmann-Kolmogorov condition

$$p(x, t|x_{t_2}) = \int dx_{t_1} p(x_t|x_{t_1}) p(x_{t_1}|x_{t_2})$$

for arbitrary t_1,$t > t_1 > t_2$. Or by the conditional mutual information

$$MI(X_t : X_{t_2}|X_{t_1}) = \int dx dx_{t_1} dx_{t_2} p(x_t, x_{t_1}, x_{t_2}) \log \frac{p(x_t|x_{t_1}, x_{t_2})}{p(x_t|x_{t_1})} = 0 \; .$$

Fokker-Planck equation

For a certain class of Markov processes the evolution of the the phase space density $\rho(\boldsymbol{x})$ evolves according to a Fokker-Planck equation

$$\frac{\partial \rho(\boldsymbol{x}, t)}{\partial t} = -\frac{\partial}{\partial \boldsymbol{x}} \boldsymbol{D}^{(1)} \rho(\boldsymbol{x}, t) + \frac{\partial^2}{\partial \boldsymbol{x}^2} \boldsymbol{D}^{(2)} \rho(\boldsymbol{x}, t) \ .$$

Equivalent Langevin equation:

$$\frac{d\boldsymbol{x}}{dt} = \boldsymbol{F}(\boldsymbol{x}) + \boldsymbol{G}(\boldsymbol{x})\boldsymbol{\Gamma}$$

with $\boldsymbol{\Gamma}$ being a Gaussian white noise process

$$\langle \Gamma_k(t)\Gamma_j(t')\rangle = 2\delta_{k,j}\delta(t - t') \ .$$

Fokker-Planck equation

For a certain class of Markov processes the evolution of the the phase space density $\rho(\boldsymbol{x})$ evolves according to a Fokker-Planck equation

$$\frac{\partial \rho(\boldsymbol{x}, t)}{\partial t} = -\frac{\partial}{\partial \boldsymbol{x}} \boldsymbol{D}^{(1)} \rho(\boldsymbol{x}, t) + \frac{\partial^2}{\partial \boldsymbol{x}^2} \boldsymbol{D}^{(2)} \rho(\boldsymbol{x}, t) \ .$$

Equivalent Langevin equation:

$$\frac{d\boldsymbol{x}}{dt} = \boldsymbol{F}(\boldsymbol{x}) + \boldsymbol{G}(\boldsymbol{x})\boldsymbol{\Gamma}$$

- The drift term (Stratonovich interpretation)

$$D_i^{(1)} = F_i + \sum_{k,j} G_{kj} \frac{\partial}{\partial x_k} G_{ij}$$

is related to the deterministic part.

Fokker-Planck equation

For a certain class of Markov processes the evolution of the the phase space density $\rho(\mathbf{x})$ evolves according to a Fokker-Planck equation

$$\frac{\partial \rho(\mathbf{x}, t)}{\partial t} = -\frac{\partial}{\partial \mathbf{x}} \mathbf{D}^{(1)} \rho(\mathbf{x}, t) + \frac{\partial^2}{\partial \mathbf{x}^2} \mathbf{D}^{(2)} \rho(\mathbf{x}, t) \ .$$

Equivalent Langevin equation:

$$\frac{d\mathbf{x}}{dt} = \mathbf{F}(\mathbf{x}) + \mathbf{G}(\mathbf{x})\mathbf{\Gamma}$$

- Drift term

$$D_i^{(1)} = F_i + \sum_{k,j} G_{kj} \frac{\partial}{\partial x_k} G_{ij}$$

- The diffusion tensor

$$D_{ij}^{(2)} = \sum_k G_{ik} G_{jk}$$

is related to the (state dependent) noise amplitudes.

Drift and diffusion terms as expectation values

In the limit $\Delta t \to 0$ one gets

$$\begin{aligned} \boldsymbol{D}^{(1)}(\boldsymbol{x}) &= \lim_{\Delta t \to 0} \frac{1}{\Delta t} \langle \boldsymbol{x}(\Delta t) - \boldsymbol{x}(0) | \boldsymbol{x}(0) = \boldsymbol{x} \rangle \\ &= \lim_{\Delta t \to 0} \frac{1}{\Delta t} E[\boldsymbol{x}(\Delta t) - \boldsymbol{x}(0) | \boldsymbol{x}(0) = \boldsymbol{x}] \end{aligned}$$

i.e. expectation value of $\boldsymbol{x}(\Delta t) - \boldsymbol{x}(0)$ conditioned on $\boldsymbol{x}(0)$ and

$$D^{(2)} ij(\boldsymbol{x}) = \lim_{\Delta t \to 0} \frac{1}{2\Delta t} E[(\boldsymbol{x}(\Delta t) - \boldsymbol{x}(0))_i (\boldsymbol{x}(\Delta t) - \boldsymbol{x}(0))_j | \boldsymbol{x}(0) = \boldsymbol{x}] \, .$$

The drift term can be estimated robustly from data - in fact it is the local constant predictor.
$\Rightarrow$ In the case of additive dynamical noise we can use *lzo-run* for prediction and with the option -% for generating data.

Finite time corrections for the diffusion term

For the estimation of the diffusion tenor finite Δt corrections have to be taken into account. Several proposals were made:
Firsts order corrections: Ragwitz and Kantz (PRL 2001)

$$\begin{aligned}
\boldsymbol{D}^{(2)}(\boldsymbol{x}) &\approx \frac{1}{2\Delta t}(E[(\boldsymbol{x}(\Delta t)-\boldsymbol{x}(0))(\boldsymbol{x}(\Delta t)-\boldsymbol{x}(0))^T|\boldsymbol{x}(0)=\boldsymbol{x}] \\
&\quad -(\Delta t)^2\boldsymbol{D}^{(1)}\boldsymbol{D}^{(1)\,T}) \\
&= \frac{1}{2\Delta t}Cov[(\boldsymbol{x}(\Delta t)-\boldsymbol{x}(0))(\boldsymbol{x}(\Delta t)-\boldsymbol{x}(0))^T|\boldsymbol{x}(0)]
\end{aligned}$$

Higher orders: See again Ragwitz and Kantz (2001), but also the comment by Friedrich et al. (2002) and the reply of the authors.