Statistical detection of changes in the underlying dynamics of observed time series

Angeliki Papana and Dimitris Kugiumtzis
Department of Mathematical, Physical and Computational Sciences, Faculty of Engineering
Aristotle University of Thessaloniki
email: agpapana@gen.auth.gr

Abstract

In the evolution of many physical systems, such as brain potential activity, we observe changes in their dynamical regime. The objective of this paper is to find reliable statistical methods capable of detecting changes in the dynamical state. In particular, emphasis is given on the change of the level of the stochastic component in the system. We consider the surrogate data test for nonlinearity in combination with two discriminating nonlinear statistics, i.e. the mutual information and the local linear fit. If the original time series contains detectable nonlinear dynamics, a suitable nonlinear statistic should be able to discriminate the original time series from its surrogate time series, which preserve only the original linear structure and are otherwise random. Changes in the stochastic component of the observed system are detected from the changes in the statistical significance of the test, i.e. the strength of discrimination between original and surrogate data. We consider different well-known simulated systems, such as Henon map and Lorenz system, and we control the level of observational noise added to the system. The surrogate data test for nonlinearity is applied to overlapped segments of the original time series. Our aim is to compare the two test statistics for different scenarios of changes of the level of noise. We examine also the discriminating power of each statistic for a range of segment lengths. Further, we apply our procedure to pre-ictal EEG records in order to assess the power of our implementation of the surrogate data test to detect changes in the dynamical evolution of EEG that are precursors of forthcoming epileptic seizure.

1. Tools

1.1 Measures

The measures used are mutual information and Local linear fit.

- **Mutual Information I(τ)** for **a delay τ** computes the linear and nonlinear correlation of two variables x_i, $x_{i-\tau}$, of a time series x_i, i=1,…,n:

$$I(\tau) = I(x_i, x_{i-\tau}) = \sum_{x\,y} p_{\chi_i,\chi_{i-\tau}}(x,y)\log\frac{p_{x_i,x_{i-\tau}}(x,y)}{p_{x_i}(\chi)\,p_{x_{i-\tau}}(y)}$$

Where $p_{x_i}(\chi)$ is the probability of $x_i=x$, $p_{x_i,x_{i-\tau}}(\chi,y)$ is the joint probability of $\chi_i=x$ and $\chi_{i-\tau}=y$ and the sum is calculated for all possible values of χ_i, $\chi_{i-\tau}$.

- The second measure is the NRMSE from the **Local Linear Fit LLF(m)**. Local linear prediction models for each time *i* is given as:

$$\begin{aligned} x_{i+1} &= F(\mathbf{x}_i) = F(x_i, x_{i-\tau}, \ldots, x_{i-(m-1)\tau}) \\ &= a_0 + a_1 x_i + a_2 x_{i-\tau} + \cdots + a_m x_{i-(m-1)\tau} \end{aligned},$$

where m is the embedding dimension. The statistic used for the fitting error at time T ahead is:

$$\text{NRMSE}(\tau, m) = \sqrt{\frac{\frac{1}{n-m-\tau}\sum_{i=m}^{n-\tau}(x_{i+\tau}-\hat{\chi}_{i+\tau})^2}{\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2}}$$

These two measures are computed using the TISEAN software package. [1]

1.2 Surrogate data test for nonlinearity.

a. A **null hypothesis Ho** is formulated, that the data has been created by a stationary Gaussian linear process that undergoes a nonlinear static transform.

b. **Algorithm for generation of surrogate data.**

Since the null assumption is not a simple one but leaves room for free parameters (mean, autocovariance), we generate random surrogate data with the same autoccorelation and distribution as a given data set. For the generation of surrogate data consistent to H_o the algorithm **statistically transformed autoregressive process (STAP)** is used. It identifies a normal autoregressive process and a monotonic static transform, so that the stransformed realizations of this process fulfill exactly both conditions: they possess the sample autocorrelation and amplitude distribution of the given data.[2]

c. **Discriminating statistics**

The discriminating statistics used are **mutual information and Local linear fit.**

d. **Test decision**

To decide for the rejection of H_0 we compute the **significance s** for the two discriminating statistics. If q_0 is the statistic from the original time series and $q_1, \ldots, q_k$ from the surrogates then the significance is:

$$\boxed{s = \frac{|q_0 - \langle q_s \rangle|}{\sigma_q}}$$

Where <q> is the average and σ is the standard deviation of $q_1, \ldots, q_k$, customarily given in units of 'sigmas'. Significance s >1.96 suggests the **rejection** of H_o at a = 0.05.

2. Simulation

2.1 Setup

We use two well-known simulated systems, Henon map and Lorenz system. We simulate the change of system characteristics during its evolution by monitoring the level of noise added to the observed time series. To detect the changes of system characteristics we split the observed data record in overlapping segments with a given sliding window. In each overlapping segment we apply the surrogate data test with the two statistics. The objective is to investigate the performance of the two statistics under different conditions of noise level, overlapping segment length and reconstruction parameters.

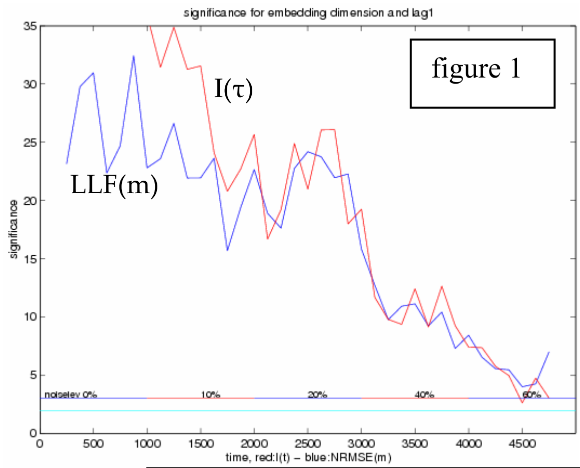
Length of original data	N
Length of overlapping segments	n
Sliding window	k = n/4

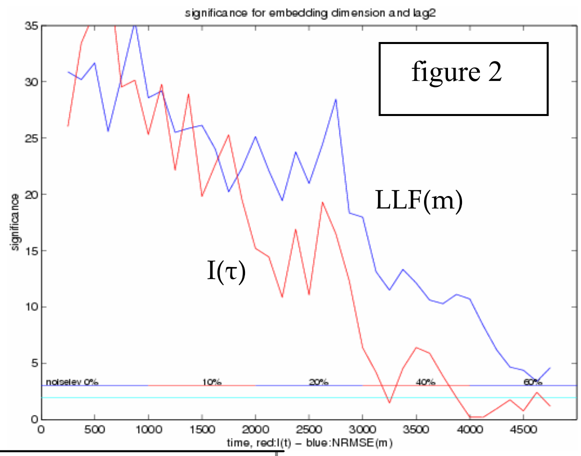
	Henon	Lorenz
τ (for I(τα))	[1,10]	[1,20]
m (for LLF(m))	[1,10]	[1,10]
N	5000	10000
N	500, 2500	2000
Noise level	10,20,40,60	10,20,40,60

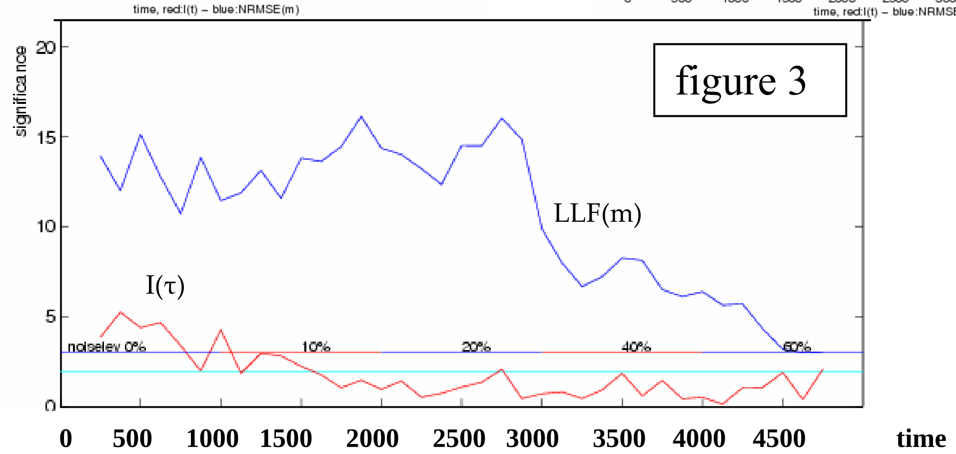
We generate the noisy time series of length N by gradually adding noise of increasing amplitudes.

2.2 Simulation results

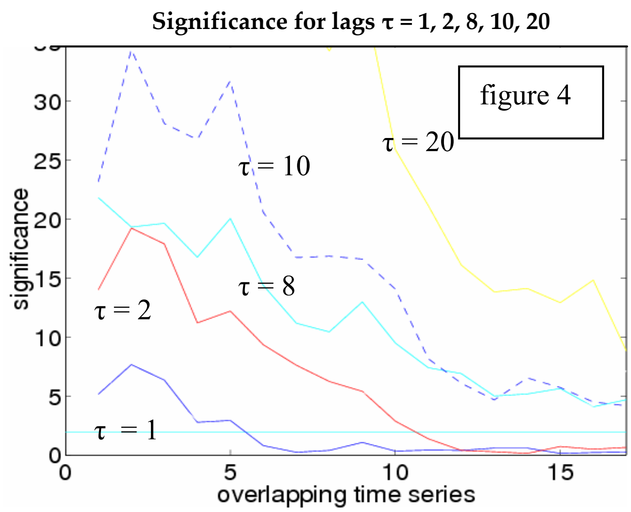
The simulations for the **Henon map** showed that LLF(m) is a more powerful statistical measure than I(τ), as the significances extracted are much higher. That means that distinction between the original time series from its surrogates is clearer when LLF is used as test statistic. We expect that significance s of the test will decrease as noise level increases. First we set **n=500**. For lag τ = 1 and embedding dimension m=1 we can see that the two statistics give similar results even for high noise levels (fig.1). However, for lag τ = 2 (and accordingly m=2 for LLF) mutual information gives much smaller significance (fig.2) and for even larger lags and I(τ) does not have any discriminating power (s < 2 for noise level >40%) . Besides the fact that for high noise amplitude, it is hard to obtain significant discrimination, we observe that local fit discriminates for unsuitably large embedding dimensions and even for sort time series, contrary to mutual information (for m=10, n=500, see fig.3). To assess how the length of the overlapping segments (n) may alter the significance of the test we set **n =2500** and repeat the same simulation. We observe that when **I(τ)** is the statistic the significance **s** is decreasing as τ increases as before (only for the first overlapping segments s for τ = 2 is larger than for τ = 1). However, s is much higher than in the first case (n=500), which means that the discrimination of the original data and the surrogates is much clearer. The same holds for LLF (m), but again s from LLF statistic reaches higher levels, which are about the same for different m.

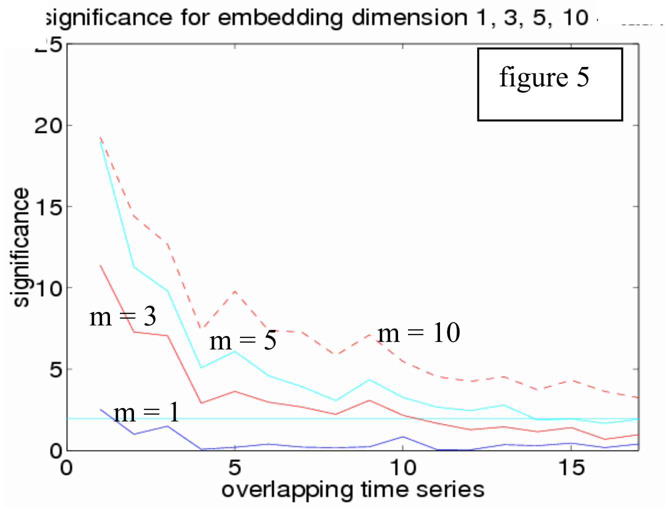




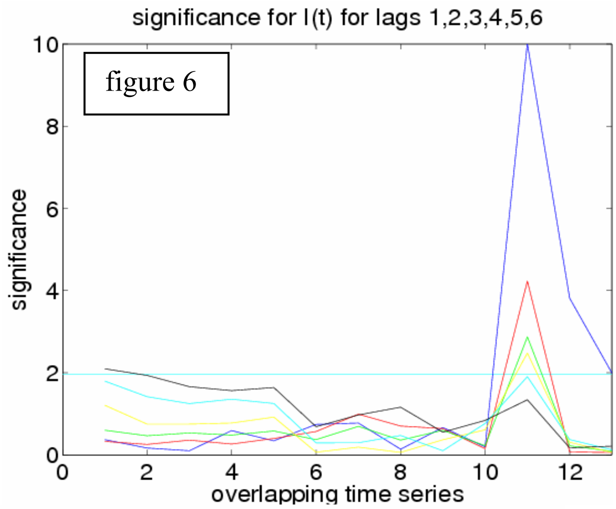


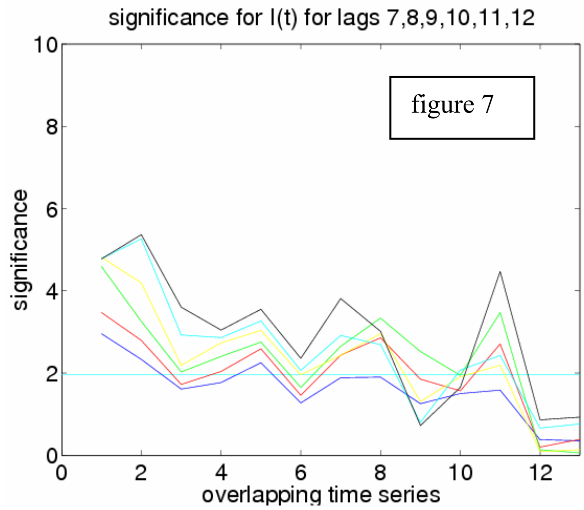
We consider the standard **Lorenz system** at chaotic regime and observe variable z with **sampling time 0,02s.** From lag τ=1 to 7 significance for I(t) is gradually increasing, from τ=8 to 14 is gradually decreasing and from τ = 15 to 20 is again increasing (fig 4). Only for τ =1 and τ = 2 mutual information does not have discriminating power for high noise levels. As the Lorenz system is a flow, we consider different delays τ when we compute LLF. For **τ=1** the significance for LLF(m) is gradually increasing for larger values of m and the increase slows down for m>4 (fig. 5). This feature holds for larger τ (we checked up to 10) but at higher levels of s that seem to rise with τ. We observe that only for m=1 and m=2 the null hypothesis is rejected and only for the last few overlapping segments which have high noise.

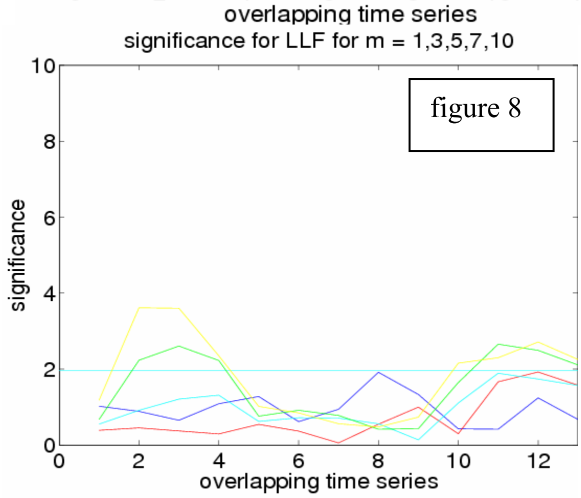




Finally, we apply this procedure to pre-ictal EEG records. Specifically we use a scalp EEG record (from left frontal lobe) of N = 12000 data (τ_s = 0.01s) from an epileptic patient and the seizure onset is at about 100s. The overlapping segments are of length n = 3000 (30s). The significance s for I(τ) increases for the first 10 overlapping segments for each lag from 1 to 6 and then has a burst at segment 11 (75s – 105s) and decreases for the last 3 overlapping segments, i.e. the dramatic change appears at the time of seizure onset (fig. 9). For increasing τ values (we checked up to 15), s rises from non significant values (s < 1.96) towards significant ones (up to s 5), showing actually a downward trend along time up to the seizure onset (see fig.6,7).







This trend was not present, at least not at the same degree, when we applied the test with **LLF** as the test statistic (see Fig.8). For example, for **τ = 5**, only for large m, we could observe that s > 1.96 for the last overlapping segments and then s< 1.96 for those up to seizure onset. Also, there was no dramatic rise of s at seizure onset.

Conclusions

The goal of this work was on the one hand to examine whether the changes in the characteristics of a dynamical system during its evolution can be detected using the statistical measures of mutual information I(τ) and local linear fit LLF(m) (monitoring the noise level and the length of overlapping segments) and on the other hand to assess and compare the strength of discrimination of these statistics. From the simulations on the Henon map we observe that for high noise levels, I(τ) has no discriminating strength (noise levels higher than 40%), whereas LLF was able to discriminate even clearer with both measures. On the other hand, from the simulation of the Lorenz system we observe that the two measures gave evidence of similar discriminating power. As for the EEG record, values of the significance of I(τ) present a clear decreasing trend for the pre-ictal period, especially for large values of τ. This indicates an increase of the stochastic component in the system, for tens of seconds prior to the seizure onset, in agreement with other findings [3]. The test does not show a clear trend of significance when LLF(m) was used as test statistic. Overall, the simulations showed that the surrogate data test equipped with suitable test statistics can detect changes in system characteristics (due to the stochastic component) then it is applied sequentially on subsequent segments of the observed signal.

References

[1] **TISEAN 2.1**, Nonlinear Time Series Analysis, Rainer Hegger, Holger Kantz, Thomas Schreiber http://www.mpipks-dresden.mpg.de/~tisean.

[2] D.Kugiumtzis, '*Statistically transformed autoregressive process and surrogate data test for nonlinearity*', Physical Review E66, 025201(R), (2002)

[3] D. Kugiumtzis, and P.G. Larson, 'Linear and Nonlinear Analysis of EEG for the Prediction of Epileptic Seizures',*Proceedings of the 1999 Workshop "Chaos in Brain?"*, World Scientific, Singapore, pp 329 – 333, 2000.