Discrimination of regimes of dynamical systems using measures of correlation

A. Papana, D. Kugiumtzis
General Department, Faculty of Technology, Aristotle University of Thessaloniki.

Abstract

Linear and nonlinear measures of correlation are compared in distinguishing different regimes of a dynamical system on the basis of time series observation. The linear measures are the autocorrelation $r(\tau)$ for certain delays and the sum of autocorrelations (Portmanteau) until a maximum delay, $Q(\tau_{max})$. Similarly, as nonlinear measures we consider the mutual information $I(\tau)$ on the same delays and the sum of $I(\tau)$ for the same maximum delay, $M(\tau_{max})$. We also examine whether the discrimination of the dynamical regimes can be achieved equally well, or even better, using as measures the *p*-values of the surrogate data test for nonlinearity (where the test statistics are the nonlinear measures). In this way, the *p*-value can be considered as an indirect measure of departure from linear correlations.

To assess the three correlation measure types, we make Monte Carlo simulations on well-known linear and nonlinear systems varying their complexity by monitoring control parameters of the system (AR model, Mackey Glass). We also apply the correlation measures to multi-channel EEG from epileptic patients. For each record and channel we create time series of the same length from successively overlapping segments and calculate the measures from these time series. We check if there is discrimination between EEG record from late preictal state (1-2 minutes before seizure onset) and from early preictal state (some hours before seizure onset), where the patient is considered to have a more regular cerebral operation.

Measures

Measures of linear correlation

- **Autocorrelation $r(\tau)$**, for τ =1, 5, 10, 20, 30:

$$r(\tau) = \frac{\Sigma(x_t - \bar{x})(x_{t-\tau} - \bar{x})}{\Sigma(x_t - \bar{x})^2}$$

- **Sum of τ_{max}=40 first Autocorrelations $Q(\tau_{max})$**:

$$Q(\tau_{max}) = \sum_{\tau=1}^{\tau_{max}} r(\tau)$$

Measures of nonlinear correlation

- **Mutual information $I(\tau)$** for τ = 1, 5, 10, 20, 30 *(measures linear and nonlinear correlations)*:

$$I(\tau) = I(x_i, x_{i-\tau}) = \Sigma p_{X_i, X_{i-\tau}}(x_i, x_{i-\tau}) \log \frac{p_{X_i, X_{i-\tau}}(x_i, x_{i-\tau})}{p_{X_i}(x_i) p_{X_{i-\tau}}(x_{i-\tau})}$$

$p_{X_i}, p_{X_{i-\tau}}, p_{X_i, X_{i-\tau}}$: marginal and joint probability distribution

- **Sum of τ_{max}=40 first Mutual information $M(\tau_{max})$**:

$$M(\tau_{max}) = \sum_{\tau=1}^{\tau_{max}} I(\tau)$$

Measures of nonlinear deviation from linearity

- ***p – values*** from *surrogate data test for nonlinearity*:

➢ H_0: original time series comes from a linear stochastic process.
➢ Construction of *K* surrogate data sets (time series with the same autocorrelations & the same marginal distribution as these of the original time series).
➢ Test statistic *q*: $I(\tau)$, τ = 1, 5, 10, 20, 30 & $M(40)$.
➢ q_0: statistic value from the original time series,
➢ $q_1, q_2, \ldots, q_K$: statistic values from *K* surrogates
➢ H_0 is rejected if q_0 doesn't belong in the distribution of $q_1, q_2, \ldots, q_K$.
➢ Test decision is assigned parametrically using the significance:

$$s = \frac{|q_o - \bar{q}|}{s_q}$$

➢ *p*-values are defined as: $p = 2(1 - \Phi(s))$
(Φ: cumulative distribution function of the standard normal distribution)

Surrogate data captures only the linearity of each time series. Thus p < 0.05 indicates the presence of nonlinearity in the time series.

Data for simulations

For the simulations we used 1000 realizations from well-known systems with time series lengths: *n* = 1000, 2000, 4000.

- **AR(9) model under quadratic and cubic transform.**

$$s_t = 6.9627 + 1.2064 s_{t-1} - 0.4507 s_{t-2} - 0.1747 s_{t-3} + 0.1974 s_{t-4} - 0.1366 s_{t-5} + 0.0268 s_{t-6} + 0.0128 s_{t-7} - 0.0312 s_{t-8} + 0.2123 s_{t-9} + e_t$$

quadratic transform $AR(9)^2$: $x_t = s_t^2$ cubic $AR(9)^3$: $x_t = s_t^3$

- **Mackey-Glass with Δ = 30, 70, 100**

$$\dot{s} = \frac{0.2 s(t-\Delta)}{1 + [s(t-\Delta)]^{10}} - 0.1 s(t)$$

We also added noise of 20% and 50% of the standard deviation of the data.

Real Data

- **Multi-channel EEG** from 8 patients. A 25 channel system was used in 6 patients and 63 channel system in 2 patients.

Names of Pre-ictal record	Prepre-ictal records
25 channels	
1. A	-
2. B	Bs, Bl
3. C	Cs
4. D	Dl
5. E1, E2	El
6. F	Fs, Fl
63 channels	
7. G1, G2	Gl
8. H	-

s: 1 hour before seizure
l: many hours before seizure

Brain regions	25 channels	63 channels
LF (left frontal)	1,3,5,11,20	1,3,5,7,9,11,13,15,21,27 29, 31
RF (right frontal)	2,4,6,12,21	2,4,6,8,10,12,14,16,22, 28, 30,32
MI (middle)	17,18,19	55 - 63
OC (occipital)	13,14,15,16	37, 38, 41- 54
RT(right temporal)	8,10,23,25	18,20,24,26,34,36,40
LT (left temporal)	7,9,22,24	17,19,23,25,33,35,39

Brain regions for patients with 25 channels

Methods

❑ **For the simulated data,** the discrimination of two different states of the systems was evaluated with **ROC curves** (Receiver Operating Characteristics Curves). Naming one state as "positive" and the other "negative", ROC makes a graphical representation of the trade-off between the sensitivity (true positive rate) and specificity (false negative rate). The plot shows the sensitivity on the X axis and (1 - specificity) on the Y axis. The curve always goes from point (0,0) to (1,1) varying a threshold cut-off: (0,0) is where the classifier finds no positives (all values are classified to state "positive"), (1,1) where everything is classified as positive (all values are classified to state "negative"). Thus for each cut-off the cumulative function of the distribution regarding each state is calculated. The **area under the ROC** is a convenient way of comparing classifiers. A random classifier has an area of 0.5 (no distinction of the two states for any cut-off value), while an ideal classifier has an area of 1 (absolute distinction for the whole range of cut-off values).

❑ **For EEG data** the discrimination between pre-ictal & prepre-ictal state was evaluated using comparison tests (for ROC, distributions could not be estimated reliably due to small sample size):

- **t-test for means:** We test if the time series of values of measures (from overlapping segments) from pre-ictal and prepre-ictal state have the same mean. We suppose that each time series of values of measures is independent and follows normal distribution and that the standard deviations are the same. If the two time series {x}, {y} have lengths n_x, n_y and standard deviation s, we use the statistic:

$$T = \frac{\bar{x} - \bar{y}}{s\sqrt{\frac{1}{n_x} + \frac{1}{n_y}}} \sim t_{n_x + n_y - 2}$$

- **Wilcoxon rank sum test**

We test if the time series of values of measures (from overlapping segments) from pre-ictal and prepre-ictal state have the same medians. We suppose that the two samples are independent and their distributions have the same general form. Values are ranked in increasing order. For the least sum of ranks W we have:

$$Z_W = \frac{W - \mu_W}{s_W} \sim N(0,1)$$

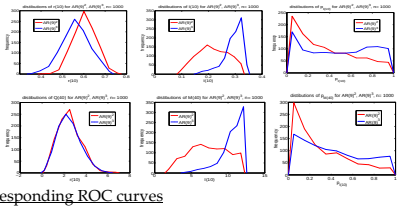
$$\text{όπου } \mu_W = \frac{n_x(n_x + n_y + 1)}{2}, \; s_w = \sqrt{\frac{n_x n_y (n_x + n_y + 1)}{12}}$$

n_x, n_y: length of time series with the minimum and maximum sum of ranks.

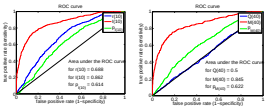
Results from simulations

- Discrimination between quadratic and cubic transform of time series generated by AR(9) (n=1000) for the measures: r(10), $I(10)$, $p_{I(10)}$ & Q(40), M(40), $p_{M(40)}$.

Distributions of each measure from $AR(9)^2$ and $AR(9)^3$.

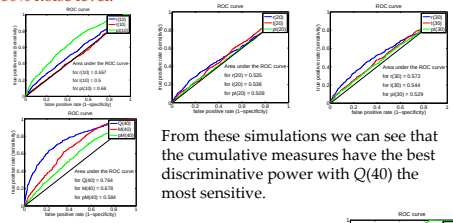


Corresponding ROC curves



From these simulations we can see that the measure with the best discriminate power is $I(10)$ & M(40).

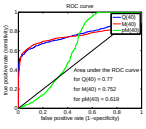
- Discrimination of Mackey Glass with Δ=70 & 100, n =1000 with 50% noise level.



From these simulations we can see that the cumulative measures have the best discriminative power with Q(40) the most sensitive.

- Discrimination of $AR(9)^2$ & Mackey Glass with Δ=30, n=1000.

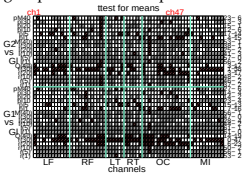
We can see that the cumulative measures have quite good discriminative power and linear and nonlinear measures perform equally well .



Results from EEG data

For patients with 63 channels: Discrimination between pre-ictal & prepre-ictal state with t-test.
Prepre-ictal records: many hours before seizure.

The statistical significance of discrimination is displayed for two factors: channels and measures. In the following figure in the x-axis are the channels and in the y-axis the measures grouped for each comparison.



Gray: the values of measures are greater for 2 min before seizure than many hours before,
Black: The opposite of Gray,
White: no discrimination between the two states.

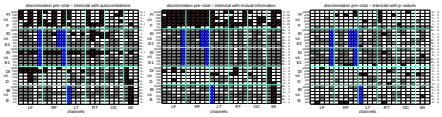
In the right side of the plot are the sums of gray - black cells.

There is a clear discrimination between the two states for most channels and measures with an exception of *p*-values. In most cases, the values of measures for 2 min before seizure are greater than for many hours before seizure.

For patients with 25 channels:

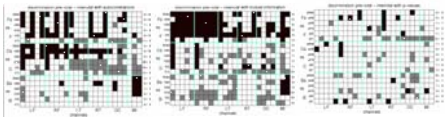
- Discrimination between pre-ictal & prepre-ictal state with t-test. Prepre-ictal records: many hours before seizure.

For these records there is also a clear discrimination between different states and even *p*-values can be used as a measure of discrimination.



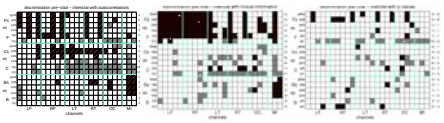
- Discrimination between pre-ictal & prepre-ictal state with t-test. Prepre-ictal records: 1 hour before seizure.

As expected, for the prepre-ictal records 1 hour before seizure there is not such a clear discrimination as for the many hours before seizure above.



- Discrimination between pre-ictal & prepre-ictal state with Wilcoxon rank sum test.

For the same records we applied Wilcoxon rank sum test in order to compare with the results from t-test. We can see that there are insignificant differences in the results.



Conclusions

From the simulations we can see that the cumulative measures Q(40) and M(40) outstand in performance, while *p*-values turned out to have poor discriminative power. Correlation measures can discriminate between pre-ictal and prepre-ictal states, especially when examining prepre-ictal records from many hours before seizure onset. The two discrimination tests (t-test & Wilcoxon rank sum test) for means and medians have no significant differences.

References

[1] "Mormann" Darbellay G. (1998). An Adaptive Histogram Estimator for the Mutual Information, *UTIA Research Report*, no. 1889, Acad. Sc., Prague
[2] Granger C. and J. Lin (1994). Using the Mutual Information Coefficiente to Identify Lags in Nonlinear Models, *Journal of Time Series Analysis*, 15, 4, 371 -384.
[3] Kantz H. and T. Schreiber (1997). Nonlinear Time Series Analysis. *Cambridge University Press*, Cambridge.
[4] Kugiumtzis D.(2002). Statistically transformed autoregressive process and surrogate data test for nonlinearity. *Physical Review* E66, 025201(R).
[5] "Chaos in Brain?" Moddemeijer R. (1999). A Statistic to estimate the Variance of the Histogram-Based Mutual Information Estimator on Dependent Pairs of Observations, Signal Processing, 75, 51-63.