Significance for a recurrence based transition analysis

Norbert Marwan, Stefan Schinkel and Jürgen Kurths

Interdisciplinary Center for the Dynamics of Complex Systems
University of Potsdam, 14415 Potsdam, Germany
Email: marwan@agnld.uni-potsdam.de

Abstract—The recurrence of states is a fundamental behaviour of dynamical systems. A modern technique of nonlinear data analysis, the recurrence plot, visualises and analyses the recurrence structure and allows us to detect transitions in the system's dynamics by using recurrence quantification analysis (RQA). In the last decade, the RQA has become popular in many scientific fields. However, a sufficient significance test was not yet developed.

We propose a statistical test for the RQA which is based on bootstrapping of the characteristic small scale structures in the recurrence plot. Using this test we can present confidence bounds for the detected transitions and, hence, get a more reliable result. We demonstrate the new technique on marine dust records from the Atlantic which were used to infer climate changes in Africa for the last 4 Ma.

1. Introduction

Recurrence plots (RPs) and recurrence quantification (RQA) [1] are widely accepted methods for data analysis in various disciplines, like life science [2, 3, 4, 5], engineering [6, 7, 8] earth science [9, 10, 11] or finance and economy [12, 13]. Based on RPs, we can study, e.g., complex system's dynamics, transitions or synchronisation [3, 14, 15, 16]. The investigation of transitions in the system's dynamics is based on changes in the system's recurrence structure. The different aspects of recurrences can be measured by measures of complexity, which are also known as *recurrence quantification analysis* (RQA) [1]. Although these measures are often applied to real data and interpreted as indicators of changes in the system, up to now there are no means to statistically validate the results. Statistical tests were suggested for the validation of interrelation and synchronisation analysis using bivariate extensions of RPs [17, 18]. These tests use certain surrogates (AR models, twin surrogates) to test against the null-hypothesis. However, these are special cases of a recurrence based analysis and are not applicable for our purpose to detect transitions. In this letter we propose a technique which calculates the confidence level for the most important RQA measures. Using this method we are able to provide a significance statement for detected transitions in the systems dynamics based on RQA. We illustrate this approach on a climate proxy time series (marine dust deposits), which were used to infer climate variability in the past.

2. Recurrence based detection of transitions

A recurrence plot tests for the pairwise closeness of all possible pairs of states $(\vec{x}_i, \vec{x}_j)$ ($i = 1 \dots N$, N as the number of time points or measurements) in an m-dimensional phase space,

$$\mathbf{R}_{i,j} = \Theta\left(\varepsilon - \|\vec{x}_i - \vec{x}_j\|\right), \qquad (1)$$

with Θ as the Heaviside function and ε as a threshold for spatial closeness, which is given by the norm $\|\cdot\|$ (e.g. maximum or Euclidean norm) [1]. The binary recurrence matrix $\mathbf{R}$ contains the value one for all close pairs $\|\vec{x}_i - \vec{x}_j\| < \varepsilon$. From a univariate timeseries the phase space trajectory can be reconstructed using time delay embedding [19].

Similar evolving epochs of the phase space trajectory cause diagonal structures parallel to the main diagonal. The length of such diagonal line structures depends on the dynamics of the system (periodic, chaotic, stochastic). Therefore, the frequency distribution $P(l)$ of line lengths l can be used to characterise the system's dynamics. Several RQA measures are based on this distribution $P(l)$. Here we focus only on the measure *determinism* (DET), which is the ratio of the recurrence points forming diagonal structures,

$$DET = \frac{\sum_{l=l_{\min}}^{N} l\,P(l)}{\sum_{l=1}^{N} l\,P(l)}. \qquad (2)$$

We use a minimal length $l_{\min}$ for the definition of a diagonal line [1].

Slowly changing states, as occuring during laminar phases (intermittency), cause vertical structures in the RP. Therefore, the distribution $P(v)$ of line lengths v is used to quantify the laminar phases occuring in a system. Similar to the measure DET, we define the ratio of the recurrence points forming vertical structures,

$$LAM = \frac{\sum_{v=v_{\min}}^{N} v\,P(v)}{\sum_{v=1}^{N} v\,P(v)}, \qquad (3)$$

and call this measure *laminarity* (LAM) [1].

In order to study the time-dependent behaviour of a system or data series, we compute these RQA measures using a moving window. The window has size W and is moved with a step of s over the data in such a way that succeeding windows overlap with $W - s$, thus providing time-dependent measures $DET(t)$ and $LAM(t)$ with $t = W/2, 3W/2, 5W/2, \dots, N - W/2$. The number of windows

N_W covering the data is floor-rounded $N_W = (N - W + s)/s$. This technique was successfully applied to detect chaos-period transitions [15], chaos-chaos transitions [3] or different kinds of transitions between strange non-chaotic behaviour and periodic or chaos [20]. It is applicable on real world data, as demonstrated for the study of cardiac variablity [21], brain activity [5], changes in finance markets [13] or thermodynamic transitions in corrosion processes [7]. However, all these applications miss a clear significance statement or require repeated measurements to allow for statistical testing.

3. Confidence intervals of univariate timeseries

In order to perform a statistical inference for the RQA measures, we propose a bootstrapping approach [22]. The bootstrap is a statistical tool that allows for estimating the precision of *any* sample statistics (mean, median, $P(l)$ or $P(v)$) by randomly resampling (*with* replacement) from the observed data.

Since the basis of the RQA measures are the frequency distributions $P(l)$ or $P(v)$ of the diagonal and vertical recurrence lines, we will bootstrap these distributions. For the sake of simplicity, we only consider $P(l)$, but the same logic applies to $P(v)$.

For each of the moving window t ($t = W/2, 3W/2, 5W/2, \ldots, N - W/2$), i.e. for different time points, we have a local distribution $P_t(l)$. However, we will use all local distributions for bootstrapping in order to get an overall distribution over the entire region of interest in the recurrence plot, which is covered by the moving windows. This means, we bootstrap from the unification

$$\hat{P}(l) = \bigcup_t P_t(l) \qquad (4)$$

of the local distributions. We draw n recurrence structures (i.e. diagonal lines) from $\hat{P}(l)$. The number n of drawings is the mean number of recurrence structures contained in the local distributions $P_t(l)$,

$$n = \frac{1}{N_W} \sum_{t=W/2}^{N-W/2} \sum_{l=l_{\min}}^{N} P_t(l). \qquad (5)$$

From the resulting empirical distribution $P^*(l)$, we compute the corresponding RQA measure, in our case DET, Eq. (2). Repeating this procedure B times (e.g. $B = 5,000$), provides a test distribution for DET, say $F(DET)$. $F(DET)$ provides a robust estimate for the system's overall behaviour as captured by the complexity measures. To this baseline of the system we can later compare any occuring transitions.

Calculating the α-quantiles of the distribution $F(DET)$, we derive the confidence intervals of DET which can be used to statistically infer whether the changes of $DET(t)$, and thus the observed transitions, are statistically significant.

4. Illustrative example

In this section we illustrate the proposed statistical test on a signal with chaos-period and chaos-chaos transitions. We use a modified logistic map with mutual transitions [15]

$$x_{i+1} = a(i)\, x(i) \left(1 - x(i)\right) \qquad (6)$$

with the control parameter a in the range [3.9200 3.9325] with increments of $\Delta a = 0.000001$. Using this intervall we find for a = [3.92221 3.92227] a period-7 window, for a = [3.93047 3.93050] a period-8 window and at a broad range around $a = 3.928$ intermittency (Fig. 1A).

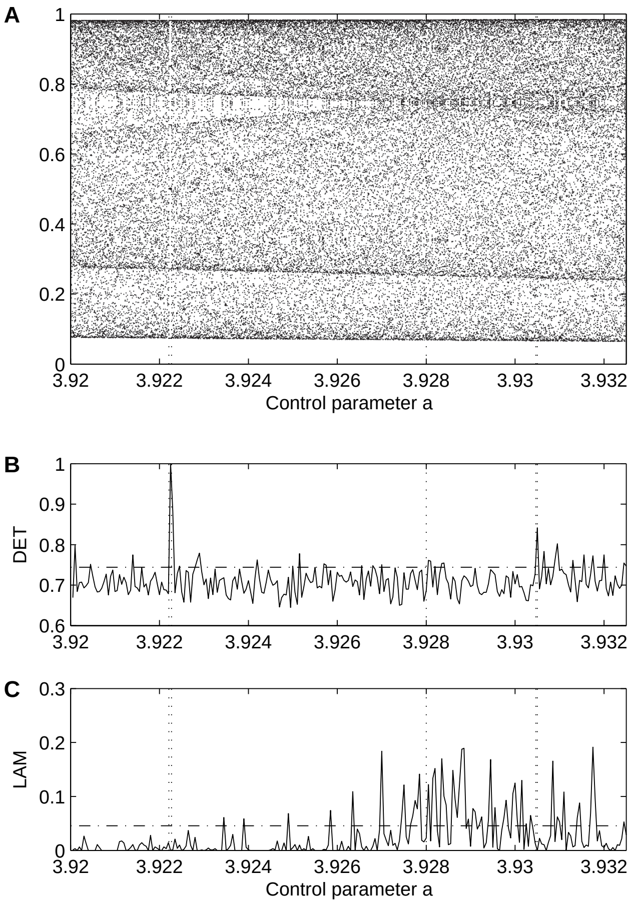

Figure 1: (A) Logistic map with chaos-period and chaos-chaos transitions for control parameter a = [3.9200 3.9325] and corresponding RQA measures (B) DET and (C) LAM. For a = [3.92221 3.92227] we have a period-7 window, for a = [3.93047 3.93050] a period-8 window and at a broad range around $a = 3.928$ intermittency (marked with dotted lines). 99% confidence bounds are shown as dash-dotted lines.

Next we compute the RQA measures DET and LAM from this data series (no embedding) using windows of size $W = 200$ and with a step size of $s = 50$. The threshold ε is chosen for each window separately in order to preserve a constant recurrence rate of 5%. As a line structure we consider each line with a length of at least two points, i.e. $l_{\min} = v_{\min} = 2$.

The measure DET shows for the periodic windows at $a = [3.92221\ 3.92227]$ and $a = [3.93047\ 3.93050]$ maxima (Fig. 1B) [3]. The periodic behaviour of the system causes only long diagonal lines, resulting in high values of DET. In contrast, LAM shows high values only for the region of intermittency around $a = 3.928$ (Fig. 1C). In this region, the system has slowly changing, laminar states [3].

For the proposed bootstrapping approach, we use 5,000 resamplings in order to construct the empirical distributions $F(DET)$ and $F(LAM)$. We have found that this number of resamplings is sufficient. The parameters of the resulting empirical distributions are already converged. As expected, the distributions $F(DET)$ and $F(LAM)$ follow normal distributions (Fig. 2). As the 99%-quantile we find for DET $q_{0.99} = 0.75$ and for LAM $q_{0.99} = 0.05$. These values provide the 99% confidence level for DET and LAM. Thus, the two maxima of DET in the periodic windows are significant on a 99% level ($p < 0.01$; Fig. 1B). For LAM we find several significant high values of 99% significance in the region of intermittency around $a = 3.928$ (Fig. 1C). This is due to the longer range of intermittent behaviour in this region of the control parameter a.

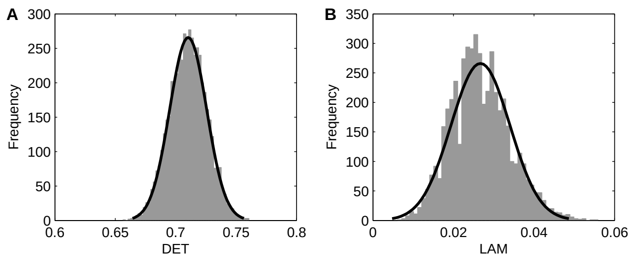

Figure 2: Empirical distributions for DET and LAM derived from bootstrapping recurrence structures. These distributions follow normal distributions (a fitted normal distribution shown by the black line).

Using these RQA measures we have shown that we are able to detect the chaos-period and chaos-chaos transitions with high significance. This is an improvement of the findings discussed in [3, 15].

5. Application on a marine dust record

Longterm variation in eolian dust deposits is highly related with terrestrial vegetation and may be used as a proxy for a changing climate (wet, dry). Therefore, marine dust records can be used to infer epochs of a drier climate in the past. In particular, a marine record from the Ocean Drilling Programme (ODP) derived from a drilling in the Atlantic, ODP site 659, was used to infer changes in the African climate during the last 4.5 Ma (Fig. 3A) [23]. The author claimed that the African climate has shifted towards more arid but variable conditions at 2.8, 1.7 and 1.0 Ma. However, a new debate about climate transitions at these times recently arose because of their importance for the hominin evolution in Africa [24]. This debate challenges for a reliable test and enhanced analysis tools for the detection of such transitions. Therefore, we apply the RQA and the proposed significance test on the dust flux record of the ODP site 659 [23].

We used a time delay embedding with dimension $m = 3$ and delay $\tau = 2$. The threshold is chosen to preserve a constant recurrence rate of 5%. The bootstrapping is performed using 5,000 resamplings. We are interested in the 95% confidence interval.

The RQA measures DET and LAM reveal significant high values between 4.2 and 4.0 Ma, 3.6 and 3.4 Ma, 2.6 and 2.4 Ma. Around 1.1 Ma only DET is significantly increased and around 2.9 Ma only LAM is significantly increased. Since 0.6 Ma, both measures increase again significantly (Fig. 3B, C).

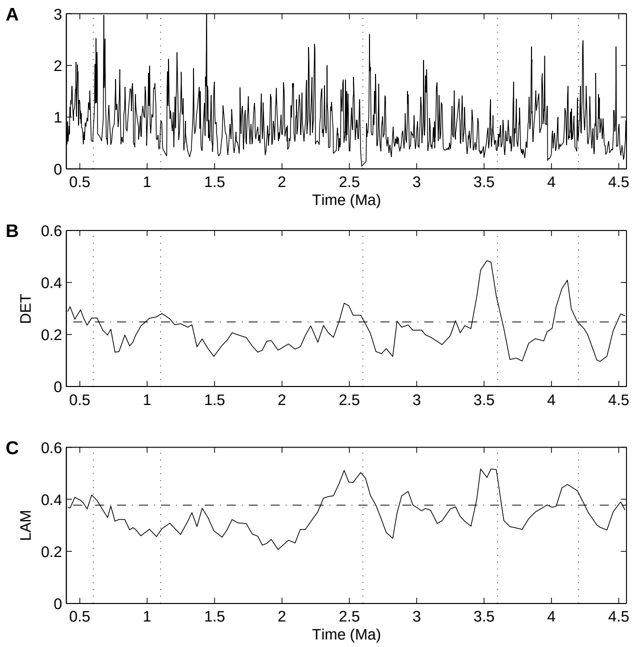

Figure 3: (A) Dust flux record of ODP site 659, and corresponding (B) DET and (C) LAM measures (95% confidence bounds are shown).

Based on the significant increase of the DET measure we can infer that especially during the epochs 4.2 to 4.0 Ma and 3.6 to 3.4 Ma the climate was behaving more regular. The increase of LAM at 4.2, 3.6, 2.6 and 0.6 Ma indicates transitions at these times in the African climate regime, as exhibited by an intermittency behaviour. These time epochs differ obviously from the climate changes proposed by deMenocal [23]. However, deMenocal was just testing for changes in the frequencies and not in the dynamics. The linear methods he used (evolutionary power spectra) are not able to detect dynamical transitions.

These epochs found using RQA coincide with the occurences of lakes in East Africa and with important hominin evolution steps [24].

6. Conclusions

By bootstrapping the smale-scale structures of recurrence plots, we were able to provide confidence levels for the recurrence quantification analysis. We have shown that the RQA reveals chaos-period and chaos-chaos transitions in the logistic map with statistical significance. Moreover, applying this approach on a palaeo-climate proxy record, we found transitions in the climate regime, which may have caused significant influences on the African climate and, thus, on the hominin evolution.

Acknowledgments

This study was supported by grants from project MAP AO-99-030 (contract #14592) of the Microgravity Application Program/Biotechnology from the Human Spaceflight Program of the European Space Agency (ESA), the Deutsche Forschungsgemeinschaft (German Research Foundation) in the Research Group FOR 868, and the COST Action BM0601.

References

[1] N. Marwan, M. C. Romano, M. Thiel, and J. Kurths, Physics Reports **438**, 237 (2007).

[2] J. P. Zbilut *et al.*, Cell Biochemistry and Biophysics **36**, 67 (2002).

[3] N. Marwan, N. Wessel, U. Meyerfeldt, A. Schirdewan, and J. Kurths, Physical Review E **66**, 026702 (2002).

[4] C. J. Stam, Clinical Neurophysiology **116**, 2266 (2005).

[5] S. Schinkel, N. Marwan, and J. Kurths, Cognitive Neurodynamics **1**, 317 (2007).

[6] J. M. Nichols, S. T. Trickey, and M. Seaver, Mechanical Systems and Signal Processing **20**, 421 (2006).

[7] L. S. Montalbán, P. Henttu, and R. Piché, International Journal of Bifurcation and Chaos **17**, 3725 (2007).

[8] A. K. Sen, R. Longwic, G. Litak, and K. Górski, Mechanical Systems and Signal Processing **22**, 362 (2008).

[9] N. Marwan, M. H. Trauth, M. Vuille, and J. Kurths, Climate Dynamics **21**, 317 (2003).

[10] T. K. March, S. C. Chapman, and R. O. Dendy, Geophysical Research Letters **32**, 1 (2005).

[11] N. V. Zolotova and D. I. Ponyavin, Solar Physics **243**, 193 (2007).

[12] J. Belaire-Franch, D. Contreras, and L. Tordera-Lledó, Physica D **171**, 249 (2002).

[13] F. Strozzi, J.-M. Zaldívar, and J. P. Zbilut, Physica A **312**, 520 (2002).

[14] J.-P. Eckmann, S. O. Kamphorst, and D. Ruelle, Europhysics Letters **5**, 973 (1987).

[15] L. L. Trulla, A. Giuliani, J. P. Zbilut, and C. L. Webber Jr., Physics Letters A **223**, 255 (1996).

[16] M. C. Romano, M. Thiel, J. Kurths, I. Z. Kiss, and J. Hudson, Europhysics Letters **71**, 466 (2005).

[17] N. Marwan and J. Kurths, Physics Letters A **302**, 299 (2002).

[18] M. Thiel, M. C. Romano, J. Kurths, M. Rolfs, and R. Kliegl, Europhysics Letters **75**, 535 (2006).

[19] N. H. Packard, J. P. Crutchfield, J. D. Farmer, and R. S. Shaw, Physical Review Letters **45**, 712 (1980).

[20] E. J. Ngamga *et al.*, Physical Review E **75**, 036222 (2007).

[21] J. P. Zbilut, N. Thomasson, and C. L. Webber Jr., Medical Engineering & Physics **24**, 53 (2002).

[22] B. Efron and R. J. Tibshirani, *An Introduction to the Bootstrap* (Chapman & Hall/CRC, Boca Raton, London, New York, Washington DC, 1998).

[23] P. B. deMenocal, Science **270**, 53 (1995).

[24] M. H. Trauth *et al.*, Journal of Human Evolution **53**, 475 (2007).