Finding recurrence networks’ threshold adaptively for a specific time series

D. Eroglu¹,², N. Marwan¹, S. Prasad³, and J. Kurths¹,²,⁴

¹Potsdam Institute for Climate Impact Research, Potsdam, Germany
²Department of Physics, Humboldt University of Berlin, Berlin, Germany
³Institute of Earth and Environmental Science, Potsdam University, Potsdam, Germany
⁴Institute of Complex Systems and Mathematical Biology, University of Aberdeen, Aberdeen, UK

Correspondence to: D. Eroglu (eroglu@pik-potsdam.de)

Received: 4 April 2014 – Published in Nonlin. Processes Geophys. Discuss.: 24 April 2014
Revised: 17 September 2014 – Accepted: 2 October 2014 – Published: 11 November 2014

Abstract. Recurrence-plot-based recurrence networks are an approach used to analyze time series using a complex networks theory. In both approaches — recurrence plots and recurrence networks — a threshold to identify recurrent states is required. The selection of the threshold is important in order to avoid bias of the recurrence network results. In this paper, we propose a novel method to choose a recurrence threshold adaptively. We show a comparison between the constant threshold and adaptive threshold cases to study period–chaos and even period–period transitions in the dynamics of a prototypical model system. This novel method is then used to identify climate transitions from a lake sediment record.

1 Introduction

Recurrence-based approaches have taken an important place in dynamical system analysis. Related approaches have been used for several decades. The basis of this analysis is finding recurrent points on a trajectory in the phase space of a dynamical system. The first recurrence-based analysis method was introduced by Poincaré as the method of the first recurrence times (Poincaré, 1890). A Poincaré recurrence is the sequence of time intervals between two visits of a trajectory at the same interval (or volume, depending on the dimension of the trajectories).

Among the different approaches of investigating dynamical properties by recurrence, the recurrence plot (RP) is a multifaceted and powerful approach to study different aspects of dynamical systems. RPs were first introduced as a visualization of recurrent states of phase space trajectories (Eckmann et al., 1987), but then enriched by different quantification techniques for characterizing dynamical properties, regime transitions, synchronization, and so on (Marwan et al., 2007). In the study of complex systems, one of the most important issues is finding dynamical transitions or regime changes. Transitions in the dynamics can be detected by different RP-based measures, which, in general, are very useful to study complex, real-world systems (Trulla et al., 1996; Marwan et al., 2002; Donges et al., 2011). Examples of their successful application in real-world systems can be found in life science (Riley et al., 1999; Marwan et al., 2002; Neuman et al., 2009; Carrubba et al., 2012), earth science (Marwan et al., 2003; Matcharashvili et al., 2008; Donges et al., 2011), astrophysics (Asghari et al., 2004; Zolotova et al., 2009), and others (Marwan, 2008).

The measures defined by the RP framework, called “recurrence quantification analysis” (RQA), are based on point density and on the length of diagonal and vertical line structures visible in the RP, being regarded as alternative measures to quantify the complexity of physical systems. In order to uncover their time-dependent behavior, RQA measures are often computed by applying a sliding window on the time series, which then can be used to identify dynamical transitions, such as period–chaos transitions (Trulla et al., 1996) or chaos–chaos transitions (Marwan et al., 2002).

Another popular method for analyzing complex systems is the complex network approach (Watts and Strogatz, 1998; Boccaletti et al., 2006). Complex network measurements are useful for investigating and understanding the complex
behavior of real-world systems, such as social, computer (Newman, 2002), or brain networks (Singer, 1999). The adjacency matrix of a complex network explains the structure of the system and thus determines the links between the nodes of a network. For unweighted and undirected networks, the adjacency matrix is binary and symmetric, hence very similar to an RP. In our previous work, we have shown that time series can be analyzed by complex networks by identifying the RP by the adjacency matrix of a network (Marwan et al., 2009; Donner et al., 2011), forming so-called recurrence networks (RNs). Complex network measures applied on RNS have been used to investigate real-world systems, such as the climate system (Donges et al., 2011) or the cardiorespiratory system (Ramírez Ávila et al., 2013). RNs have been shown to be more sensitive for the detection of periodic–chaos or chaos–periodic regime transitions than some of the standard RQA measures (Zou et al., 2010; Marwan, 2011).

Although recurrence-based methods are powerful tools to study complex systems, they come with an important, non-trivial issue (Marwan, 2011). To identify recurrences, a spatial distance (or volume, depending on the dimension of the system) in the phase space is usually used, and a sufficient closeness between the trajectories is determined by applying a so-called recurrence threshold $\epsilon$ to the distances (Marwan et al., 2007; Donner et al., 2010). Several approaches for selecting a meaningful threshold value have been suggested (Marwan et al., 2007; Schinkel et al., 2008; Donges et al., 2012). Of particular interest are such methods that help one to overcome the problem of sliding-window-based analyses of systems with varying amplitude fluctuations (as coming from different dynamical regimes or non-stationarities); e.g., those based on normalizing time series or fixing recurrence density. However, in real-world applications, time series are not usually smooth all the time. When considering the time series by an RN representation, extreme points (very high jumps or falls in the fluctuation of time series) in the time series could break the connected components in the network since the distance between an extreme point and other points would be larger than the threshold value. The normalization method would then result in non-optimal recurrence thresholds biasing the recurrence analysis.

In this work, we will suggest a novel method of an adaptive threshold selection based on the network’s spectral properties (Boccaletti et al., 2006). We will present a comparison between the constant and the adaptive threshold approach for detecting certain regime transitions (chaos–periodic or periodic–chaos). Finally, we will demonstrate the novel approach for analyzing lake-sediment-based paleoclimate variation.

2 Recurrence plots, recurrence networks, and the adaptive threshold

In the $m$ dimensional phase space reconstruction of a time series, a state is considered to be recurrent if its state vector falls into the $\epsilon$ neighborhood of another state vector. Formally, for a given trajectory $x_i \ (i = 1, \ldots, N, x_i \in \mathbb{R}^m)$, the recurrence plot $R$ is defined as

$$R_{i,j}(\epsilon) = \Theta(\epsilon - \|x_i - x_j\|), \ i, j = 1, \ldots, N,$$

(1)

where $N$ is the trajectory length, $\Theta(\cdot)$ is the Heaviside function, and $\| \cdot \|$ is the norm of the adopted phase space (Marwan et al., 2007). Thus, $R_{i,j} = 1$ if stated when times $i$ and $j$ are recurrent, and $R_{i,j} = 0$ otherwise. The trajectory in the phase space can be reconstructed via time delay embedding from a time series $\{u_i\}_{i=1}^N$ (Packard et al., 1980):

$$x_i = (u_i, u_{i+\tau}, \ldots, u_{i+(m-1)\tau}),$$

(2)

where $m$ is the embedding dimension and $\tau$ is the embedding delay. The embedding dimension $m$ can be found by false nearest neighbors and the delay $\tau$ by mutual information or auto-correlation (Kantz and Schreiber, 1997).

The main diagonal of the RP, $R_{i,i} = 1$, represents the line of identity (LOI). As we have mentioned, the RP is a symmetric, binary matrix. The structures formed by line segments, which are parallel to the LOI in an RP, characterize typical dynamical properties. We observe homogeneously distributed recurrence points if the dynamics are white noise. If the system is deterministic, diagonal line segments, which are parallel to the LOI, will dominate. The dynamics are related to the length of the diagonal line segments; chaotic dynamics cause mainly short line segments, but, conversely, regular (periodic) dynamics cause long line segments. The RQA quantifies this relation and can be used to detect transitions in the system’s dynamics (Trulla et al., 1996; Marwan et al., 2007).

Recurrence networks are based on the recurrence matrix, Eq. (1), which is an $N \times N$ matrix, where $N$ is the length of the phase space trajectory (the number of time steps). We now consider these time steps as nodes of a network; if the nodes are sufficiently close to each other – in other words, if the space vectors are neighbors –, there is a link between them. In network theory, connections between network nodes can be described with the adjacency matrix $A$, with $A_{i,j} = 1$ if there is a link between nodes $i$ and $j$; otherwise, $A_{i,j} = 0$. To obtain the adjacency matrix from the recurrence matrix, we discard self-loops in the recurrence matrix; i.e.,

$$A_{i,j} = R_{i,j} - \delta_{i,j},$$

(3)

where $\delta_{i,j}$ is the Kronecker delta ($\delta_{i,j} = 1$ if $i = j$, otherwise $\delta_{i,j} = 0$).

The number of links at the $i$th node (the degree) is given by $k_i = \sum_j A_{i,j}$. In this paper, we use the eigenvalue spectrum
of the Laplacian matrix $L$ to find an adaptive threshold $\epsilon_c$, where $L_{i,j} = \delta_{i,j} k_i - A_{i,j}$.

The crucial point in the paper is choosing the adaptive threshold for calculating the RN. A threshold for recurrence-based methods should be sufficiently small (Marwan et al., 2007; Donner et al., 2010; Donges et al., 2012). Too small $\epsilon$ causes very sparsely connected RN with many isolated components; too large $\epsilon$ results in an almost completely connected network. For data sets that are not smooth, choosing a reasonable, small threshold could nevertheless result in unconnected recurrence network components. These unconnected components would cause problems for some complex network measures, since some of them need a connected network to be computed for the entire network. For example, even if we have just one node that is not connected to the network, the average path length will always be infinite for the entire network. An even more important motivation for avoiding isolated components in the RN is that the RN provides a large amount of information about the dynamics of the underlying system, although it contains only binary information. This has been demonstrated by reconstructing time series from RPs (Thiel et al., 2004; Hirata et al., 2008). The condition for reconstructing a time series from an RP is that all points are connected by their neighborhoods; i.e., there are no isolated components. By applying recurrence measures, we would like to quantify the dynamics encoded by the RN. This can be ensured by the above-mentioned condition.

To find a sufficiently small threshold $\epsilon$ that fulfills the desired condition of connected neighborhoods, we will use the connectivity properties of the network. In particular, we choose the value for $\epsilon$ that is the smallest one for the RN to be connected. In order to find such an adaptive threshold, we start from very small values of the threshold and vary the $\epsilon$ parameter until we get a connected network. In order to apply this approach efficiently, we use an iterative bisection method in the simulations. The connectivity of a network can be measured by the second-smallest eigenvalue $\lambda_2$ of the Laplacian matrix. If the network is connected, $\lambda_2 > 0$ (Boccaletti et al., 2006). We choose the adaptive threshold value as the minimum value of the sequence of thresholds $T = T_i, T_{i+1}, \ldots$ when the second minimum eigenvalue $\lambda_2$ is positive:

$$\epsilon_c = \min(T) \text{ with } T = \{T_i | \forall i : \lambda_2(T_i) > 0\}.$$  

(4)

Values $\epsilon$ below the critical value $\epsilon_c$ are indicating the existence of unconnected components in the RN (Fig. 1). After that critical threshold, $\lambda_2$ becomes positive and, if we increase the threshold even more, the connectivity of the RN increases. By choosing the critical point $\epsilon_c$ as the recurrence threshold, we ensure that the RN will be connected by the smallest threshold possible.

**Figure 1.** Variation of the second-smallest eigenvalue of the Laplacian $\lambda_2$, due to changing threshold value, using the logistic map as an illustrative example (control parameter $a = 4.0$). $\lambda_2 = 0$ for thresholds below a critical value $\epsilon_c$, indicating the existence of unconnected components in the RN. For $\epsilon > \epsilon_c$, there are no unconnected components in the RN anymore. The adaptive threshold value for this time series is $\epsilon_c \approx 0.19$.

### 3 Applications

#### 3.1 Logistic map

As a first application, we compare some RN measures for first using the adaptive and then the constant threshold approach by analyzing the logistic map:

$$x_{i+1} = ax_i(1-x_i).$$  

(5)

It is one of the most popular iterated maps, which has different regimes for different control parameters $a$. The detection of the transitions of the logistic map between these different regimes was previously studied with RP and RN (Trulla et al., 1996; Marwan et al., 2009). The logistic map shows interesting dynamics in the range of the control parameter $a \in [3.5, 4.0]$, which is studied here with a step size of $\Delta a = 0.0005$; for example, periodic and chaotic regimes, bifurcations, and inner and outer crises occur. We compute a time series of length $N = 5000$ for each value of $a$. In order to discard transients, we delete the first 2000 values, resulting in time series consisting of 3000 values that have been used for all analyses of the logistic map in this paper.

For the constant threshold selection method, we use the recurrence rate method to choose a threshold value: a threshold is selected in such a way that the recurrence rate (RR) is constant even for different time series with different dynamics (e.g., different values of $a$) (Marwan et al., 2007). In this paper, we use $RR = 5\%$ arbitrarily for further analysis.

Now we compute the RNs by using the given threshold selection techniques $\epsilon$ and $\epsilon_c$ for each control parameter $a$. We then calculate transitivity, $T$, and betweenness centrality (BC) as the complex network measures in order to detect the transitions from periodic–chaotic, chaotic–periodic states, bifurcations, and inner and outer crisis. The network transitivity is given by
Figure 2. (a) Lyapunov exponent and transitivity using (b) an adaptive threshold and (c) a constant threshold for the logistic map. Dashed lines show certain bifurcation points before the chaotic regime.

![Graph showing Lyapunov exponent and transitivity](image)

The average betweenness centrality of network is

\[ BC = \frac{1}{N} \sum_{v} \sum_{s \neq v \neq t} \frac{\sigma_{st}(v)}{\sigma_{st}}, \]  

(7)

where \( \sigma_{st} \) is the total number of shortest paths from node \( s \) to node \( t \), and \( \sigma_{st}(v) \) is number of the paths that pass through \( v \).

As mentioned in the previous chapter, not all complex network measures can be applied to a disconnected network. However, it would cause problems for computing the measures on RNs calculated by using the constant threshold technique, since the network could be disconnected. For instance, to compute the average shortest path length or assortativity for an entire network, the network must be connected. Disconnected nodes of the network could be discarded from the calculation, but, in this case, we would lose information. In the adaptive threshold case, we could calculate all these measurements on the entire network, since the selection of the adaptive threshold ensures that the recurrence network is connected.

Both threshold selection methods could detect transitions between dynamical regimes (periodic–chaos or chaos–periodic). Transitivity gives large values for the chaotic regime and small values for periodic. In the betweenness centrality case, it is contrary to transitivity; large values for periodic and small values for the chaotic regimes. Although the constant threshold selection detects the periodic windows (chaos–period transitions) more sharply than the adaptive threshold case, the transitivity, \( T_{\text{constant}} \), and betweenness centrality, \( BC_{\text{constant}} \), for the constant threshold selection case (in the constant threshold case, in general, the threshold is arbitrarily chosen by \( RR = 5\% \)) cannot distinguish between different periodic dynamics; i.e., it cannot detect certain bifurcation points, such as for period doublings at \( a \approx 3.544, 3.564, \) and 3.84, for example. Conversely, in the adaptively chosen threshold case, \( T_{\text{adaptive}} \) and \( BC_{\text{adaptive}} \) are sensitive to these bifurcations (Figs. 2 and 3). Thus, using the adaptive threshold also allows the detection of period–period transitions (i.e., the study of bifurcation points, where the maximal Lyapunov exponent remains non-positive).

### 3.2 Application to paleoclimate record

The study of paleoclimate variation helps in understanding and evaluating possible future climate change. Lake sediments provide valuable archives of past climate variations. In the following, we will focus on a well-dated high-resolution climate archive from paleolake Lisan located beneath the archaeological site of Masada in the Near East (Prasad et al., 2004, 2009). The sediments from the upper member were deposited (26–18 cal ka BP) when the lake reached its highest stands (Bartov et al., 2003; Torfstein et al., 2013). The sedimentary sequence contains varves...
comprising seasonally deposited primary (evaporitic) aragonite and silty detritus (Prasad et al., 2004). The pure aragonite sublaminae were precipitated from the upper layer of the lake during summer evaporation. Their formation requires inflow of HCO$_3^-$ ions into the lake from the catchment area during winter floods (Stein et al., 2003) that also bring in silty detrital material. One detrital and overlying aragonite sublaminae constitute a varve. Previous studies (Prasad et al., 2004; Torfstein et al., 2013) indicate that small ice-rafting events (denoted as a, b, c, and d), as well as prominent Heinrich events in the North Atlantic, are associated with the eastern Mediterranean arid intervals. The study of seasonal sublaminae yields evidence of decadal to century scale arid events that correlate with cooler temperatures at higher latitudes. Analyses in the frequency domain indicate the presence of periodicities centered at 1500, 500, 192, 139, 90, and 50–60 years, suggesting a solar forcing on climate (Prasad et al., 2004).

We use the yearly sampled pure aragonite proxy (CaCO$_3$) from the paleolake Lisan for our RN analysis (Fig. 4a). We use a time delay embedding with dimension $m = 3$ and delay $\tau = 2$ – these parameters have been computed by a standard procedure using false nearest neighbors and mutual information (Packard et al., 1980; Kantz and Schreiber, 1997) – for reconstructing the phase space. To detect dynamical transitions in the paleoclimate data, we adopt a sliding window of $W$ data points with a step size of $\Delta W$. RNs are computed one by one for each window of the time series. We have chosen a sampling window size of $\Delta T = 100$ years, with 90\% overlap corresponding to a time window size of $W \approx 100$ data points (since there are some gaps in the data, the number is not exactly 100). The time series’ length is $N = 7665$ and the total number of windows analyzed is $N - W / \Delta W \approx 755$.

Transitivity and betweenness centrality are then calculated within these windows (Fig. 4b and c). As we have shown for the logistic map, transitivity and betweenness centrality are both sensitive to detecting transitions. Larger values of transitivity, $T$, refer to regular behavior, whereas smaller values refer to more irregular dynamics in the considered window of the time series.

The gray shaded horizontal band in Figs. 4b and c is the confidence interval of the network measures. We apply a rather simple test in order to see whether the characteristics of the dynamics at a certain time statistically differ from the general characteristics of the dynamics. In order to apply this test, we use the following approach. We create surrogate data segments of length $W$ by drawing data points randomly from the entire time series, and we compute the RN and the network measures from such a surrogate segment. We repeat this 10 000 times and have an empirical test distribution of transitivity, $T$, and BC. A confidence interval is then estimated from these distributions by their 0.05 and 0.95 quantiles.

Previous studies (Prasad et al., 2004) had identified multiple climate fluctuations in the varved Lisan record and correlated them with the Greenland oxygen isotope data.
Figure 4. (a) Aragonite (CaCO$_3$) record from paleolake Lisan, (b) transitivity, and (c) betweenness centrality results of RN using the adaptive threshold. Abrupt changes in $T$ and BC indicate transitions between different climate regimes. Dry events in Lake Lisan (cooling of the higher latitudes) are marked by blue bars and two interstadial peaks (warming) by orange bars. The gray shaded band is the 90% confidence interval for the networks measures.

(indicative of temperature changes; Stuiver and Grootes, 2000) and ice-rafting events in the North Atlantic (Bond et al., 1997). The blue and orange vertical bars in Fig. 4 delineate periods of cooling and warming, respectively, in the higher latitudes that resulted in drier and wetter episodes in the eastern Mediterranean. The network measures, $T$ and BC, both indicate abrupt transitions well (Fig. 4b and c). In particular for $T$, the values jump between high and low values. $T$ reveals epochs of significantly low values at around 25.8–25.6, 25.2–25.1, 24.3–24.2, 24.0–23.9, 22.8–22.6, 22.3–22.1, 21.5–21.1, 21.7, 20.6–20.5, 20.1–19.9, 19.8–19.6, and 19.3–18.9 cal ka BP. The periods 25.8–25.6, 22.3–22.1, 21.5–21.1, and 19.3–18.9 cal ka BP correspond to the known Bond events $d$, $c$, $b$, and $a$, and the epoch between 24.3 and 23.9 cal ka BP coincides with the Heinrich H2 event. During the interstadial peaks “IS2” event at 23.8–23.7 and 23.3–23.2 cal ka BP, $T$ shows significant high values, almost reaching the value 1. BC exhibits rather similar behavior of abrupt transitions like $T$, but with opposite signs. A general observation is that low values in $T$ can be found during dry but high values during wet regimes, and that such regimes change abruptly.

A high transitivity value indicates a more regular deposition of aragonite, and thus a more regular, or even periodic, climate variability. This could be an indication for a dominant role of the (more or less periodic) solar forcing via its influence on the temperature in the higher latitudes. During phases of a colder North Atlantic, the solar forcing becomes less important, but regional climate effects become more important and dominating, causing a more complex, irregular climate variability, finally indicated by low values of $T$.

Combining the maxima of $T$ and minima of BC, we can identify the above-mentioned periods of non-regular climate dynamics. Most of these periods correspond to cold events; e.g., the Bond and Heinrich events, and the found Lisan lake events L3 to L13 (Prasad et al., 2004). Several regular periods can be identified, some of which coinciding with the warm period during the interstadial IS2. Few remaining periods of high or low regularity have not yet been identified in the literature so far and call for further investigation.

The abrupt changes in $T$ are available due to the adaptive threshold. By using a constant threshold, $T$ varies only slowly and more gradually. Defining the time points of the climate regime shifts becomes more difficult in this case.
4 Conclusions

We have represented a novel method to choose a recurrence threshold adaptively and compared it with the constant threshold selection technique. The selection of recurrence thresholds for recurrence plots and recurrence networks is a crucial step for these techniques. So far, the threshold had to be chosen arbitrarily, taking into account different criteria and application cases, as well as requiring some expertise. Here, we have proposed a novel technique to determine such a threshold value automatically depending on the time series. Such an adaptive threshold is directly derived from the topology of the recurrence network. It is selected in such a way that the recurrence network does not have unconnected components. We have discussed transitivity and betweenness centrality measures of the complex network approach. Both measures are related to the regularity of the dynamics.

Moreover, the proposed threshold selection can also be useful for the recurrence quantification analysis. A systematic investigation of the different threshold selections remains to be looked into in the future.

We have compared the novel adaptive threshold selection with the arbitrarily selected threshold by applying them to the logistic map. Although both methods distinguish the dynamical regimes clearly, the adaptively chosen threshold approach detects many more bifurcations in particular, such as period doubling. Such bifurcations are important characteristics of the dynamical systems, since these bifurcations route to chaos from periodicity.

Moreover, we have used our approach to investigate a paleoclimate proxy record from the paleolake Lisan representing the climate variability in the Near East between 27 and 18 cal ka BP. Both transitivity and betweenness centrality measures clearly identified transitions between wet and dry (and vice versa) periods by an abrupt decrease of dynamical regularity, perhaps due to a reduced solar influence. Our method identified some transitions that have not been known so far from the literature and require further investigation: e.g., by analyzing other proxy records from this region. By choosing the adaptive threshold, we have been able to identify the transitions more clearly than by using the arbitrary selected threshold approach.

Acknowledgements. We thank Reik V. Donner for fruitful discussions. Part of the work was supported by the project “Gradual environmental change versus single catastrophe – Identifying drivers of mammalian evolution” (SAW-2013-IZW-2), funded by the Leibniz Association (WGL).

Edited by: J. Davidsen
Reviewed by: two anonymous referees

References


