Physica A 445 (2016) 1–17

Contents lists available at ScienceDirect

Physica A

journal homepage: www.elsevier.com/locate/physa

Finding metastable states in real-world time series with recurrence networks

I. Vega *, Ch. Schütte, T.O.F. Conrad

Mathematics Institute, FU Berlin, Germany

HIGHLIGHTS

- Recurrence-based method for identifying metastable states in real-world time series.
- We introduce the concept of metastability in the field of recurrence networks.
- Metastable states are identified with a fuzzy partitioning of recurrence networks.
- We show new relations between embedding parameters, recurrence rate and entropy.
- Our results are robust to considerable noise and missing data points.

ARTICLE INFO

Article history:
Received 21 October 2014
Received in revised form 11 October 2015
Available online 28 October 2015

Keywords:
Recurrence quantification analysis
Metastability
Non-linear dynamics
Recurrence threshold

ABSTRACT

In the framework of time series analysis with recurrence networks, we introduce a self-adaptive method that determines the elusive recurrence threshold and identifies metastable states in complex real-world time series. As initial step, we introduce a way to set the embedding parameters used to reconstruct the state space from the time series. We set them as the ones giving the maximum Shannon entropy of the diagonal line length distribution for the first simultaneous minima of recurrence rate and Shannon entropy. To identify metastable states, as well as the transitions between them, we use a soft partitioning algorithm for module finding which is specifically developed for the case in which a system shows metastability. We illustrate our method with a complex time series example. Finally, we show the robustness of our method for identifying metastable states. Our results suggest that our method is robust for identifying metastable states in complex time series, even when introducing considerable levels of noise and missing data points.

1. Introduction

The need to understand the dynamics of complex data coming from the biological, the financial, the environmental or the medical fields, has promoted the development of many visualization and analysis methods. According to Ref. [1], some of the linear methods – such as PCA or Classical Multi-dimensional Scaling – and non-linear methods – such as Stochastic Neighbor Embedding or Isomap – used for this purpose, can have some drawbacks, like not preserving both local and global scale properties of complex data or depending on many undetermined parameters. These problems can leave a large part of the analysis open to subjective interpretation.

An alternative approach that gives information about the local, medium and global scales in high-dimensional, non-linear time series, is recurrence analysis via recurrence plots and recurrence networks [2,3]. A recurrence plot is a tool to visualize

* Corresponding author.
E-mail address: iliusi@mi.fu-berlin.de (I. Vega).

http://dx.doi.org/10.1016/j.physa.2015.10.041

phase space trajectories which provides dynamical information of even high-dimensional data sets [4]. And by removing the diagonal from a recurrence plot, it can be represented as a graph. Such representation is called a recurrence network, although there are several other definitions of recurrence network [5]. The theoretical foundations of these constructions are explained in Section 2.

One of the difficulties of computing a recurrence plot is selecting its recurrence threshold. The recurrence threshold is a parameter that controls how close two phase space trajectories, or state space vectors, should be in order to consider them as neighbors. How to set an appropriate recurrence threshold for real-world time series is a problem long discussed [6–9]. This problem originates from some common properties of real-world time series: having a non-necessarily uniform probability distribution, frequently having noise or missing some measurement points. Though for deterministic processes, there are standard procedures [10] to set the recurrence threshold.

Additionally, a real-world time series could show *metastability*, which can be briefly described as the existence of different time scales in which the system either seems to be in equilibrium if it transits between so-called metastable states, or not (see Section 2). For a time series showing metastability, we would like the recurrence threshold to be able to distinguish the different metastable states in the system.

In this paper we introduce a method for the identification of metastable states in real-world time series based on recurrence networks. Our method is broadly divided into three parts, explained in Section 3. The first part is devoted to the estimation of the parameters necessary to reconstruct the state space from a time series. These parameters are set in terms of the simultaneous first minima of two recurrence quantification analysis (RQA) measures: the Shannon entropy of the frequency distribution of diagonal line lengths and the recurrence rate. In the second part of our method, we compute an appropriate recurrence threshold for the construction of a recurrence network. The idea behind this computation is that a robust recurrence threshold should be located in a range of values whose associated recurrence networks have similar modular structure. This means, that the number of modules and the number of nodes in every module is similar. In the last part, we identify metastable states in the time series, as well as the transitions between them.

The performance of our method is illustrated in Section 4, where we apply it to a time series showing metastability. In Section 5 we validate its ability to identify metastable states in a robust way.

To the best of our knowledge, this is the first time the concept of metastability is introduced in the analysis of real-world time series with recurrence networks.

Given that we construct a filtration of recurrence thresholds for every pair of embedding parameters within a range, and that we analyze the modular structure of its associated recurrence networks, the selected recurrence threshold and embedding parameters may be different for every realization of a time series for a given process. Interestingly, the results of analyzing a real-world time series with our method are robust to the addition of noise and missing data points.

2. Background

A state space trajectory is a series of states describing the evolution of a system. Since a deterministic dynamical system can be defined by its state space and an evolution operator, the time evolution of the state space trajectory explains the dynamics of the system.

The state space trajectory of a chaotic dynamical system can be reconstructed from a time series using embedding techniques based on the Takens' theorem of embedding [11]. The main idea behind this theorem is to reconstruct a diffeomorphic manifold M' from time-delayed observations on a manifold, M.

Takens' theorem has also been extended to fractal sets by Sauer, Yorke and Casdagli [12]. Stark et al. [13–15] have extended it to include random dynamical systems like systems where weak noise is added, input–output systems [16] and irregularly sampled systems.

In general, a delay embedding requires setting two parameters: the embedding delay, τ, and the embedding dimension, m. Different selections of embedding parameters will reconstruct state spaces with different topologically similarity to the original phase space of the system.

For a time series u_i of length N, associated to a deterministic dynamical system, the $N^* = N - \tau(m-1)$ state space vectors reconstructed, $\mathbf{x}_i$, are given by

$$\mathbf{x}_i = (u_i, u_{i+\tau}, \ldots, u_{i+(m-1)\tau}), \quad \text{for } i = 0, \ldots, N^*. \tag{1}$$

For a deterministic dynamical system, different geometrical, dynamical and topological tests can be used [17] to set the embedding dimension. The geometrical tests, like the computation of fractal dimensions or false nearest neighbors (FNN) [18], indicate the variations in distance between two close points when the embedding dimension increases. The dynamical tests, like the implementation of predictability tests or the estimation of Lyapunov exponents, are used to select the embedding that provides a unique future for every data point. The topological tests look for the embedding dimension m that avoids intersections of stable periodic orbits. For an extended discussion on these tests, see Sauer et al. [12] and Adachi [19].

For deterministic dynamical systems, in order to set the embedding delay, one must guarantee that the vector built from all the ith entries of the state space trajectories is linearly independent from the vector built from all jth entries of the state space trajectories, for all $i \neq j$. This implies that for periodic time series, the embedding delay cannot be a multiple of the period. For an extended discussion on this topic, see Abarbanel [20]. A standard procedure for deterministic processes is to set