Commun Nonlinear Sci Numer Simulat 54 (2018) 84–99

Contents lists available at ScienceDirect

Commun Nonlinear Sci Numer Simulat

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

Research paper

Recurrence network measures for hypothesis testing using surrogate data: Application to black hole light curves

Rinku Jacob[a], K.P. Harikrishnan[a,*], R. Misra[b], G. Ambika[c]

[a] Department of Physics, The Cochin College, Cochin-682002, India
[b] Inter University Centre for Astronomy and Astrophysics, Pune-411007, India
[c] Indian Institute of Science Education and Research, Pune-411008, India

ARTICLE INFO

Article history:
Received 26 August 2016
Revised 20 May 2017
Accepted 22 May 2017
Available online 25 May 2017

Keywords:
Recurrence networks
Hypothesis testing
Nonlinear time series analysis
Black hole light curves

ABSTRACT

Recurrence networks and the associated statistical measures have become important tools in the analysis of time series data. In this work, we test how effective the recurrence network measures are in analyzing real world data involving two main types of noise, white noise and colored noise. We use two prominent network measures as discriminating statistic for hypothesis testing using surrogate data for a specific null hypothesis that the data is derived from a linear stochastic process. We show that the characteristic path length is especially efficient as a discriminating measure with the conclusions reasonably accurate even with limited number of data points in the time series. We also highlight an additional advantage of the network approach in identifying the dimensionality of the system underlying the time series through a convergence measure derived from the probability distribution of the local clustering coefficients. As examples of real world data, we use the light curves from a prominent black hole system and show that a combined analysis using three primary network measures can provide vital information regarding the nature of temporal variability of light curves from different spectroscopic classes.

1. Introduction

Detecting deterministic nonlinearity in real world data contaminated by different types of noise is a highly nontrivial problem. It is still one of the major challenges in nonlinear time series analysis [1], though several methods and measures have been suggested over the years to address this long standing issue [2,3]. A generally accepted procedure to detect any nontrivial behavior in a time series is the method of surrogate data [4], for a statistical hypothesis testing, though there are other ways reported in literature to probe nonlinearity of time series without employing surrogates, under certain conditions. Examples are methods related to time-directed network properties of visibility graphs for testing the time-reversal asymmetry [5,6]. The method of surrogate data involves generating an ensemble of surrogates from the data. A specific null hypothesis is assumed for the data that there is no nontrivial character associated with it. The data and the surrogates

* Corresponding author.
E-mail addresses: rinku.jacob.vallanat@gmail.com (R. Jacob), kp.hk05@gmail.com, kp_hk2002@yahoo.co.in (K.P. Harikrishnan), rmisra@iucaa.in (R. Misra), g.ambika@iiserpune.ac.in (G. Ambika).

http://dx.doi.org/10.1016/j.cnsns.2017.05.018
1007-5704/

are then subjected to the same analysis sensitive to this nonlinear measure. One then tries to statistically reject the null hypothesis for the data by comparing the results for the data and the surrogates [4], with certain confidence level.

In the present analysis, we assume a specific null hypothesis that the data is generated from a linear stochastic process and no nonlinearity is associated with it. We generate a set of surrogate data which are compatible with the null hypothesis of a linear stochastic process. We then use certain measures derived by transforming the time series to a complex network as discriminating measures (as explained below) and try to reject the null hypothesis for the data.

Though the method of surrogate analysis is very popular, there are also many challenges associated with it [7]. For example, generation of proper surrogate data is very important for the success of hypothesis testing. The method to generate surrogate data was initially introduced by Theiler et al. [4] with the Amplitude Adjusted Fourier Transform (AAFT) surrogates. These surrogates are capable of testing the null hypothesis that the data come from linear as well as nonlinear static transformation of a linear stochastic process. An improved version of the AAFT algorithm has been suggested by Schreiber and Schmitz [8,9] using an iterative scheme called the IAAFT surrogates, which is reported to be more consistent to test null hypothesis [7]. Recently, Nakamura et al. [10] have proposed a surrogate generation method called Truncated Fourier Transform (TFT) [11]. However, the surrogate data generated by this method are influenced by a cut-off frequency. In addition, there are also some other types of surrogate data testing reported in the literature, such as, cycle shuffle surrogates [12], surrogates for testing pseudoperiodic time series [13] and even recurrence based surrogates [14], with each scheme found useful in particular contexts. In this work, we apply the IAAFT scheme to generate surrogate data using the TISEAN package [15].

The second major factor in the surrogate analysis is the choice of a discriminating measure that is sensitive to the nonlinearities associated with the data. In many cases, the correlation dimension D_2 and the correlation entropy K_2 have been used as the discriminating measures [16] as they can be directly computed from the time series by the delay embedding method [17]. However, the number of data points should be sufficiently large for a proper computation of these measures. In the paper by Theiler et al. [4], a time reversal asymmetric statistic was introduced which required relatively short time series for computation. In this paper, we consider the use of recurrence network (RN) measures for hypothesis testing, under varying conditions of noise. One obvious advantage of these measures is that they can be computed with reasonable accuracy even when the time series is short (say < 5000 data points) [18]. Recently, Subramaniyam et al. [18,19] have used the RN measures for the analysis of EEG data and have shown that these measures can provide insights into the structural properties of EEG in normal and pathological states. Very recently, we have shown that the RN measures can characterize the structural changes in a chaotic attractor contaminated by white and colored noise [20]. Here our aim is to highlight their effectiveness as a tool for hypothesis testing in noisy environment, especially when colored noise is involved. We specifically show that the characteristic path length is very useful in this regard. Moreover, we also present a unique advantage of network based measures for analysis in that the degree distribution as well as the distribution of the local clustering coefficient of the RN provides important information regarding the dimension or the number of variables required to model the underlying system. We specifically derive a convergence factor using the standard Kullback–Leibler measure to identify the dimension beyond which the distributions tend to converge. Details regarding the construction of the RN and the various network measures used in this paper are discussed in the next section.

We use a time series from the Lorenz attractor as prototype to illustrate the effectiveness of using RN measures as discriminating statistic. We add different percentages of white and colored noise to the standard Lorenz attractor time series and the surrogates to get a quantitative estimate of how much noise can swamp the inherent nonlinear behavior and how to fix the threshold of the statistical measure to discriminate between nonlinearity and noise. As examples of real world data involving colored noise, we analyse light curves from the prominent black hole system GRS1915+105. This black hole system is considered to be unique with the light curves falling into 12 spectroscopic classes [21], whose details are discussed in Section 4. The system appears to randomly flip in X-ray intensity variations and these observed intensity variations averaged over all energy bands are grouped into 12 different states. Earlier analysis [22] using the measures D_2 and K_2 has strongly indicated deterministic nonlinearity for light curves in 5 of the 12 classes. Here we show that analysis using the network measures can provide more exact information regarding the dimensionality of the underlying system as well as the nature of noise contamination in different states.

Finally, it is also important to share some thoughts as to why the proposed method based on RN works. From a conceptual point of view, Donges et al. [23] have shown that all RN properties can be analytically derived from the system's invariant density. In that case, if we generate IAAFT surrogates from a univariate time series of a nonlinear deterministic system, then by definition, the surrogates will leave the probability distribution invariant. However, the particular phase relationship between different parts of the reconstructed attractor changes. Hence it is important to clarify that the success of the proposed method is based on taking IAAFT surrogates from univariate time series and then use embedding of this surrogate time series, instead of taking the original multivariate time series and their IAAFT surrogates. The latter possibility may result in some distinctly different behavior.

Our paper is organized as follows: In the next section, we discuss the details regarding the construction of the RN and the computation of the network measures to be used for hypothesis testing. In Section 3, we do surrogate analysis on synthetic data from the Lorenz attractor as well as data obtained by adding different percentages of white and colored noise to the Lorenz data. Analysis of the real world data from the black hole system is presented in Section 4. Conclusions are drawn in Section 5.