Testing for nonlinearity in nonstationary time series: A network-based surrogate data test

M. C. Mallika [*]
Department of Futures Studies, University of Kerala, Kariavattom, Kerala 695 581, India

S. Suriya Prabhaa [†]
Department of Futures Studies, University of Kerala, Kariavattom, Kerala 695 581, India

K. Asokan [‡]
Department of Mathematics, College of Engineering, Trivandrum, Kerala 695 016, India

K. S. Anil Kumar [§]
University of Kerala, Palayam, Thiruvananthapuram, Kerala 695 034, India

T. R. Ramamohan [‖]
Department of Chemical Engineering, M. S. Ramaiah Institute of Technology, MSR Nagar, Bangalore 560 054, India

K. Satheesh Kumar [¶]
Department of Futures Studies, University of Kerala, Kariavattom, Kerala 695 581, India

(Received 12 September 2021; accepted 17 November 2021; published 30 November 2021)

The classical surrogate data tests, which are used to differentiate linear noise processes from nonlinear processes, are not suitable for nonstationary time series. In this paper, we propose a surrogate data test that can be applied on both stationary time series as well as nonstationary time series with short-term fluctuations. The method is based on the idea of constructing a network from the time series, employing a generalized symbolic dynamics method introduced in this work, and using any one of the several easily computable network parameters as discriminating statistics. The construction of the network is designed to remove the long-term trends in the data automatically. The network-based test statistics pick up only the short-term variations, unlike the discriminating statistics of the traditional methods, which are influenced by nonstationary trends in the data. The method is tested on several systems generated by linear or nonlinear processes and with deterministic or stochastic trends, and in all cases it is found to be able to differentiate between linear stochastic processes and nonlinear processes quite accurately, especially in cases where the common methods would lead to false rejections of the null hypothesis due to nonstationarity being interpreted as nonlinearity. The method is also found to be robust to the presence of experimental or dynamical noise of a moderate level in an otherwise nonlinear system.

DOI: 10.1103/PhysRevE.104.054217

I. INTRODUCTION

Nonstationarity is a pervasive and persistent challenge for modeling and forecasting real-world data ranging from economic to ecological systems and political time series to machine learning as most theories are developed based on the assumption that the phenomenon under investigation is stationary and fluctuating around a time-independent mean [1–6]. It has been argued that all systems are nonstationary at some scale which often corresponds to macrosystems spanning broad spatial and temporal scales [7–10] For example, ecological processes at large scales vary within spatial or temporal domains, potentially leading to complications for cross-scale inference and prediction [7,11–13]. Nevertheless, natural resources planning and management were carried out till recently on the assumption that the climate system is stationary over a period of the Holocene epoch [5]. In the area of macroeconomic studies, nonstationary models could explain large deviations such as long-running booms or recessions. On the statistical front, classical techniques such as canonical variate analysis or principal component analysis are not good at monitoring nonstationary processes. On the machine learning front, this adversity has led to the development of a modified reinforcement learning mechanism to incorporate varying temporal components [14].

A stationary process fluctuates around a time-independent mean value with constant variance. A nonstationary process oscillates around a deterministic or a stochastic trend. Most

[*]mallikasasi@gmail.com
[†]suriyaprabhaa4@gmail.com
[‡]asokank@cet.ac.in
[§]ksanilksitm@gmail.com
[‖]trr@msrit.edu
[¶]kskumar@keralauniversity.ac.in

often, one is interested in the nature of such rapid fluctuations rather than the time-dependent trend to develop forecasting models.

Surrogate data analysis is an important step in the investigation of time series exhibiting random-like fluctuations. Such fluctuations could result from either a stochastic system or a nonlinear deterministic system with chaotic behavior. However, the time series generated by a deterministic chaotic system and a linear Gaussian process exhibit some common characteristics, especially in the behavior of the autocorrelation function, power spectrum, and Lyapunov exponents, which makes the task of distinguishing between them difficult. The surrogate data test, formulated initially by Theiler *et al.* [15], is a useful tool in identifying whether the given time series originated from a linear stochastic process or not.

The surrogate data test follows a hypothesis testing procedure. The first step is to specify a null hypothesis that the given data are a random series with some of the data's observed features. One then generates a set of surrogate data by transforming the given data in such a way as to preserve any characteristics enshrined in the null hypothesis but destroying any possible nonlinear features of the data. This process makes the surrogates consistent with the assumed null hypothesis. Then one calculates one or more test statistics from the distribution of surrogates and the original series. If the values of the test statistics differ significantly, the null hypothesis is rejected, giving further confirmation that the series under consideration has a deterministic character.

The classical methods of the surrogate data test are all designed to work only with stationary series. They cannot, in general, apply if the original series has nonstationary features in addition to random fluctuations. The very nonstationary nature of the data may lead to large differences in the values of the test statistics for the original and surrogates, but in the test, this may be wrongly interpreted as being due to nonlinearity. Nonstationary data with random fluctuations are common, and it is essential to extend these methods to encompass such series as well. A few recent studies in this direction attempted to modify the classical methods to suit the nonstationary character of the data [16–18]. In this paper, we propose a network-based surrogate method to test nonlinearity in the short-term fluctuations even when there are long-term trends in the data. In this method, the test statistic is computed from a special type of network constructed from the series and surrogates which pick up only the short-term variations in the data. Hence the method can be applied no matter whether the short-term random-like variations are stationary or around a long-term trend.

II. EXISTING METHODS

The most commonly used surrogate data algorithms for testing nonlinearity are the Fourier transform (FT) method, the amplitude-adjusted Fourier transform (AAFT) method, and the iterative AAFT (IAAFT) methods. These methods differ from one another in the null hypothesis they test and in the construction of the surrogates. In the FT method the surrogates are created by first taking the Fourier transform of the given series, then randomizing the phases, and finally taking the inverse Fourier transform [15]. This process does not change the power spectrum (periodogram) and hence preserves the autocorrelation if the original series is stationary, but the phase randomization destroys any nonlinearity in the data. The resulting surrogates can be considered completely linear and well suited for the null hypothesis that a linear Gaussian process generated the data.

The AAFT algorithm [15] is an improvement of the FT method and corresponds to the null hypothesis that the series is a stationary Gaussian linear process possibly transformed by a monotonic nonlinear static function. In this method, the original values are first scaled to a Gaussian distribution (Gaussianization) and then transformed by the FT algorithm. Finally, the data set is de-Gaussianized by scaling back to the original values. The surrogates produced by this process have the same amplitude distribution and similar (but not necessarily exact) spectrum as the original series.

The IAAFT is an iterated version of the previous algorithm where the steps are repeated several times until the spectrum is sufficiently similar and the amplitudes are identical to the original values [19]. The merits and demerits of these methods have been widely discussed [20,21], and in the literature, most researchers advocate using IAAFT since it generates surrogates with autocorrelation functions closer to that of the original series.

The preceding methods of surrogate test assume that the original series is stationary, and applying them to nonstationary data may lead to incorrect results. When the data are not stationary, the nonstationarity itself could lead to significant differences in the values of discriminating statistics of the data and the set of surrogates, so one cannot judge whether the null hypothesis is being rejected due to nonlinearity or nonstationarity [22]. A few existing methods extend the classical surrogate data tests to include a possible nonstationarity in the data. Nakamura and Small [16] proposed the small shuffle surrogate (SSS) method in which surrogates are obtained by shuffling the data index on a small scale, thus destroying local structures but maintaining the overall global behavior. However, this method tests only against the null hypothesis that the short-term irregular fluctuations are independent and identically distributed random variables and cannot indicate whether the data are linear or nonlinear. Nakamura *et al.* [17] introduced the truncated Fourier transform surrogates (TFTS) method for data with or without long trends, which test against the null hypothesis that a stationary linear system generates the short-term irregular fluctuations. Their method could be applied to data in which the power spectrum has peaks in the lower frequency domain—corresponding to the global trend—which could be distinguished from the power behavior in the higher frequency domain corresponding to the irregular fluctuations. In this case, the surrogates are generated by the IAAFT method with phase randomization restricted to the higher frequency domain only. The right choice of the cutoff frequency is very crucial for this method to be successful. Lucio *et al.* [18] use a modified IAAFT algorithm in which the trend in the series is removed just before taking the Fourier to transform and put back immediately after the inverse transform. Thus the surrogates preserve the nonstationarity in addition to all that the IAAFT is guaranteed to preserve. Lucio *et al.* [18] have also discussed the advantages of their method compared to the naive approach of initially detrending the original series, then generating the surrogates, and finally adding the trend back to the surrogates. This naive approach performs poorly in reproduction of linear structure, histogram, and local nonstationary similarity. Recently wavelet-transform-based surrogates have been proposed [23,24], to test series with possible nonstationarity. These methods use the wavelet transform and resampling of wavelet coefficients [25] or combine the wavelet transform technique with IAAFT [23] to produce surrogates that match the specific null hypothesis. However, these methods are more computationally intensive than the traditional surrogate methods.

III. OUR NETWORK-BASED METHOD

When applied to nonstationary data, the traditional surrogate tests fail to distinguish nonlinearity from nonstationarity as the IAAFT algorithm almost always alters the long-term trends. So when these methods are extended to include possible nonstationary trends in the data, additional care must be taken to maintain the long-term trends in the surrogates to ensure that the nonstationarity does not influence the difference in the values of the discriminating statistics. In the proposed method, the time series and the surrogates are transformed into appropriate networks, and one of the many network characteristics is used as the discriminating statistic. The network construction automatically removes long-term trends in both the given time series and its surrogates while preserving the short-term variations, and consequently the test statistic picks up only the short-term variations. This eliminates the need for additional steps to maintain long-term trends while constructing surrogates. This means that we can use the IAAFT algorithm without any modification to create surrogates even if there is a long-term trend in the original data, since the test statistic captures only the short-term oscillations, no matter whether the surrogate algorithm alters the trend or not. We require only that when there is a frequency overlap of the trend and the fluctuation, one of the amplitudes of either the trend or the fluctuations dominate.

Thus we can use the same null hypothesis as in the IAAFT method without any assumption of stationarity, namely, that *the given time series is generated by a linear Gaussian stochastic process, possibly altered by a static nonlinear measurement function.* We now describe the method of generating the network from a given time series used in this work.

A. Network from time series

An area of recent origin in time series analysis is the use of complex networks to represent time series and employing the statistical properties of the network to quantify the underlying dynamics of the time series. Many such representations are possible depending on how the nodes and edges of the network are defined and constructed from the time series. An earlier study in this direction was due to Zhang and Small [26] and Zhang *et al.* [27], who generated complex networks from pseudoperiodic time series by representing each cycle as a node and connecting them by an edge if the phase space distance or correlation coefficient between them is less than a threshold value. Yang and Yang [28] and Gao and Jin [29] used the correlation matrices of time series to construct the network. Other methods include using individual observations as nodes and applying a visibility condition for connecting nodes [30], using phase space vectors and their neighborhood relations for defining networks [31], or using the recurrence matrix of time series for defining adjacency in networks [32]

We use a generalized version of the method of Freitas *et al.* [33] to construct the network from a time series using symbolic dynamics. The first step in this method is to replace every data point in the time series by a 0 or 1 depending upon if it is less or greater than the median of time series values, as depicted in Fig. 1(a). From the binary series thus obtained, M bits are selected at a time, by a window of length M bits, and converted into a decimal number. The window now slides from the beginning to the end of the binary series shifting by one position at a time, thus generating a sequence of decimal numbers. The distinct decimal numbers of the series are considered as the nodes of a network. The edge between two nodes is decided by the adjacency of the corresponding decimal number in the series, with multiple occurrences of the same adjacency pair constituting only one edge. Loops constituted by the occurrence of the same number twice in adjacent positions are also permitted. Henceforth we refer to this network as the SDN network. With $M = 10$, Freitas *et al.* [33] have demonstrated that the structural properties of the SDN network can capture the dynamical properties of a chaotic time series.

B. Generalized SDN

We recently demonstrated that a generalization of the SDN by increasing the number of partitions of the range of series can remarkably improve the efficiency of the network in capturing the dynamics of the series [34]. In a two-bit encoding, we partition the range of the time series into four segments demarcated by the three quartiles Q_1, Q_2, Q_3 as depicted in Fig. 1(b). A time-series value lying below Q_1 will be represented by 00 and a value greater than or equal to Q_1 and less than Q_2 by 01 and so on. The symbolic series will be twice the length of the original series. As in single-bit encoding, we use a sliding window of M bits, moving by two positions (same as the encoding size) at a time, converting the binary series into a decimal series and constructing the corresponding network. We can extend this procedure to n-bit encoding by dividing the range of the time series into 2^n partitions, and the corresponding network shall be denoted by GSDN. Our previous analysis showed that a network with six-bit encoding could capture the bifurcation diagram of the logistic map with remarkable accuracy [34].

C. Generalized SDN for nonstationary series

The generalization described in the previous subsection can be effectively applied to stationary series. However, for nonstationary series, the network will be affected by both the trend as well as the fluctuations around it. We propose a further generalization of the above two-bit encoding method for such a series by replacing the median line Q_2 with the trend curve. The lines Q_1 and Q_3 are replaced by the translations of the trend curve by a perpendicular distance $d/2$ on either side

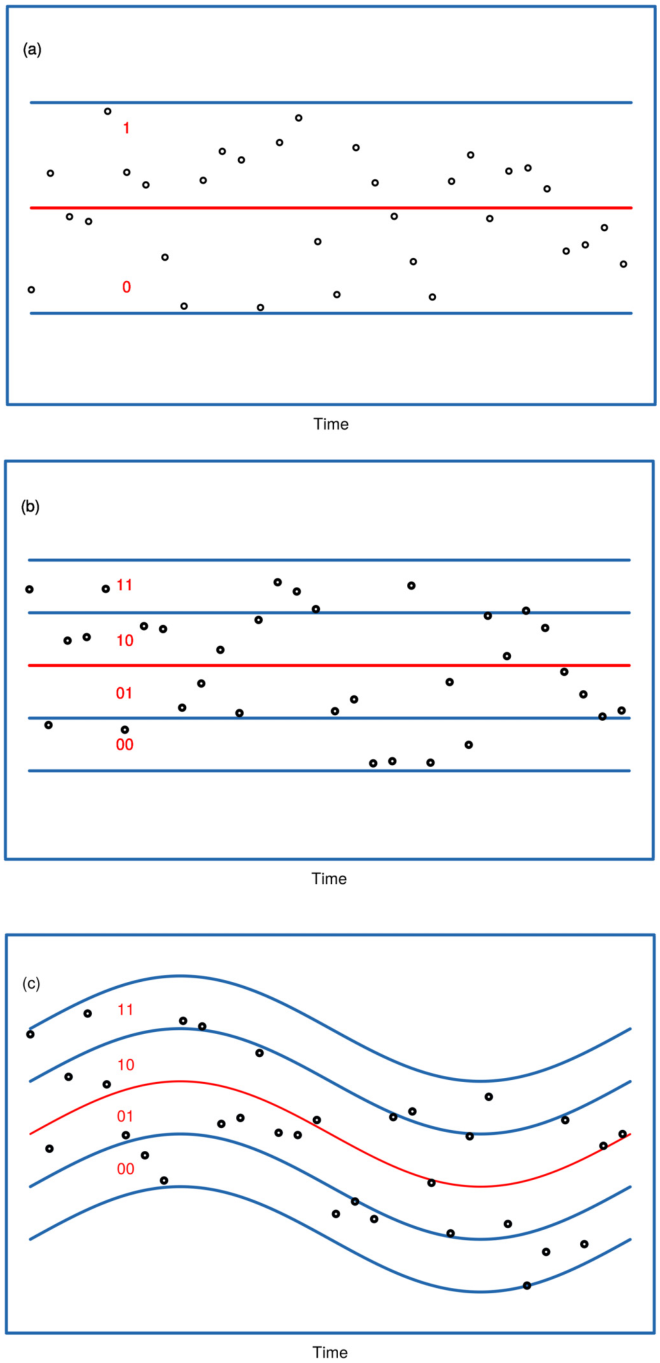

FIG. 1. Diagram showing (a) one-bit encoding for stationary time series, (b) two-bit encoding for stationary time series, and (c) two-bit encoding for nonstationary time series. The time series points are represented by the circles, and lines and curves demarcate the partition of the range.

of it where d is the maximum of perpendicular distances of the points of the nonstationary time series from the trend curve, creating four bands as depicted in Fig. 1(c). Each time series point is then represented by 00, 01, 10, or 11 as it lies in the lowest or the next higher band and so on. This method may be extended to n-bit encoding to generate the binary series similarly. From the binary series, the generation of the decimal series using the sliding window and subsequent construction of the network proceeds as in the previous method. In our analysis, we found that six-bit encoding with $M = 10$ gives the most accurate results. While the variation of M does not affect the results significantly, the number of bits by which we move the window each time is significant, and the network captures the underlying dynamics best when the width of the shift is equal to n in n-bit encoding [34]. We refer to the network constructed this way by GSDNNS.

D. The discriminating statistics

The discriminating power of the test statistics for detecting nonlinearity has also been the focus of investigation for many researchers [20,21,35,36]. The time-series-induced network provides a host of alternative test statistics by virtue of its structural properties [37]. The following are the network metrics which we have used as test statistics in this work.

The graph density (GD) is the ratio of the number of edges to the maximum possible number of edges in the network. For a network with N nodes and E edges, this is given by

$$GD = \frac{E}{N(N-1)/2}. \tag{1}$$

The transitivity T of a graph is a measure of the relative number of triangles in the graph, compared to the maximum possible number of triangles:

$$T = \frac{3 \times \text{observed number of triangles in the network}}{\text{maximum possible number of triangles in the network}}. \tag{2}$$

A path between two nodes in a network is the sequence of connected edges joining the given nodes. The number of edges in a path is its length. Among the paths between two nodes, the one with the smallest length is called the shortest path. The diameter of a network is the length of the longest of all shortest paths in a network. We have also considered the average path length and the mean, median, and standard deviation of the network's degree distribution as test statistics in this analysis.

E. Surrogate generation and calculation of test statistics

The systematic procedure for generating the surrogates and calculating the test statistics is as follows.

For a given time series $\{s_t\}$ for $t = 1, 2, 3, \ldots, n$ we first remove the end mismatch, by subtracting the end-to-end line:

$$g_t = s_1 + \frac{s_n - s_1}{n-1}(t-1). \tag{3}$$

The resulting time series $x_t = s_t - g_t$ is then used to generate a set of surrogates using the IAAFT algorithm. The surrogates will generally have a different trend than the original series, but this will not cause any serious problem if there is no frequency overlap of the trend and fluctuations or if the amplitude of either the trend or the fluctuations dominates. Removing the end mismatch is crucial in avoiding frequency spillover between stationary oscillations and the altered trend of a surrogate [38].

We next generate the GSDNNS networks of both the original time series x_t and their surrogates, and for this we approximate the trend curve [cf. the red curve in Fig. 1(c)] by

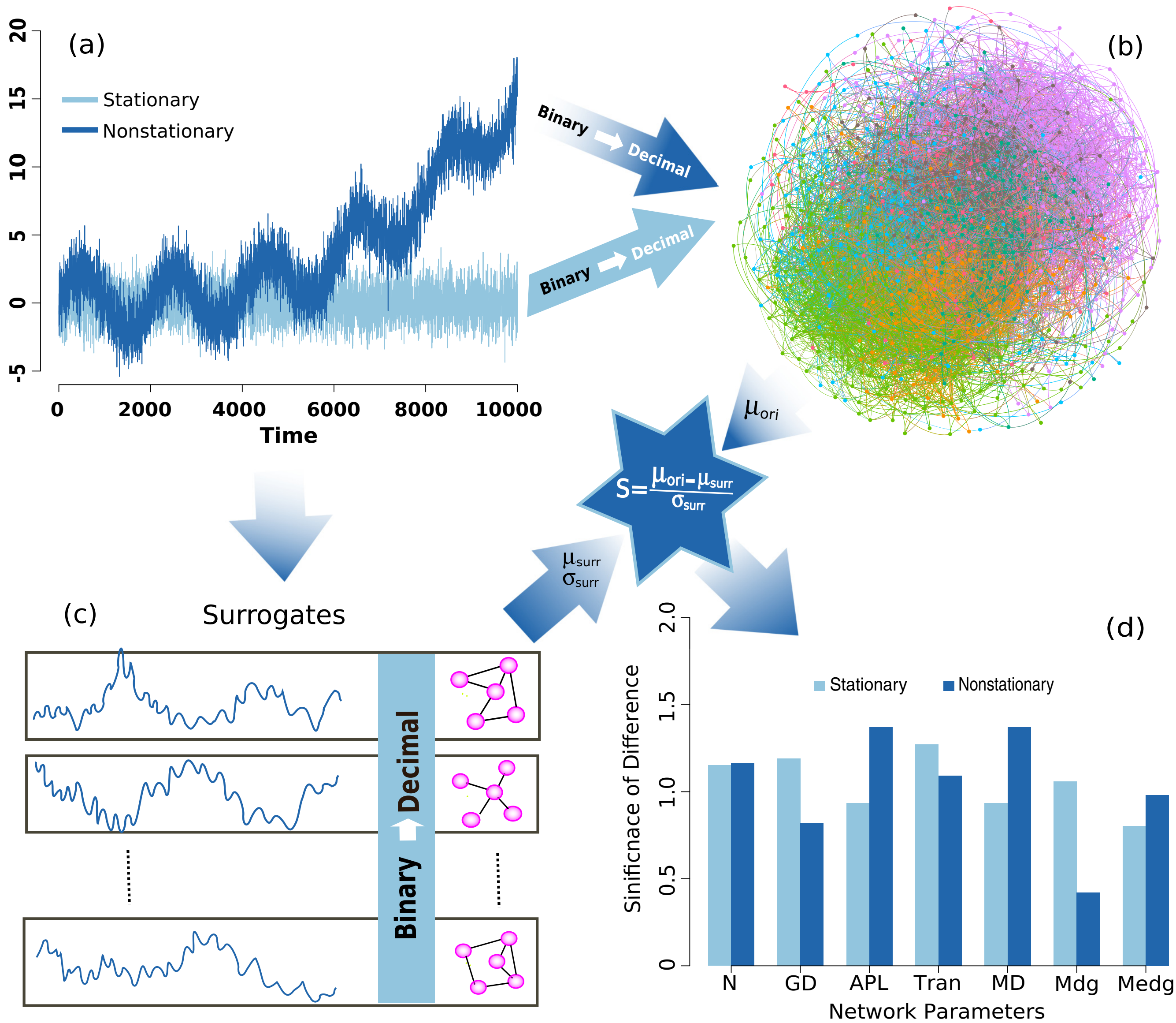

FIG. 2. (a) Stationary autoregressive process process AR(1) and the trend $t^3 + \sin(t)$ added series. The ACF, PCF, and ADF tests show that the curve with trend added is nonstationary. (b) The GSDNNS network of the trend-added series. The different colors show the various clusters. Both series produce identical networks. (c) Schematic diagram of surrogates and their networks. We have conducted the surrogate data test for both series by generating 40 surrogates for both an AR(1) process and its trend-added series. (d) The significance of difference S computed for the AR(1) process and its trend-added series using various network parameters. The values of S for most of the structural properties of the network are almost identical. The values S of both series are less than two, showing no significant difference between the original series and their surrogates for the AR(1) process and its trend-added series. The notations used in the x-axis labels are N, number of nodes; GD, graph density; APL, average path length; Tran, transitivity; MD, mean distance; Mdg, mean degree; and Medg, median degree.

a sequence of line segments. We partition the time series x_t into a number of segments, and for each time series segment, the trend is approximated by fitting a straight line. The binary series is constructed for each segment based on the trend line and is concatenated to form the binary series of the whole series x_t. The generation of decimal series from the binary series and the construction of the network then proceeds as described earlier. The same procedure is also applied to each of the surrogates to generate networks using the same segment length in the piecewise approximation of the trend. The decimal series picks up the oscillations with respect to the trend line, and hence the effect of the trend is removed in the decimal series and in the network. The segment length should be chosen to be smaller than the timescale of the trend and larger than the timescales of the stationary oscillations. Such a choice would always be possible if there is a significant difference between the timescales of trend and oscillations.

Once the GSDNNS networks for both the time series and the surrogates are constructed, we can use any of the several network metrics mentioned earlier as a test statistic. For any

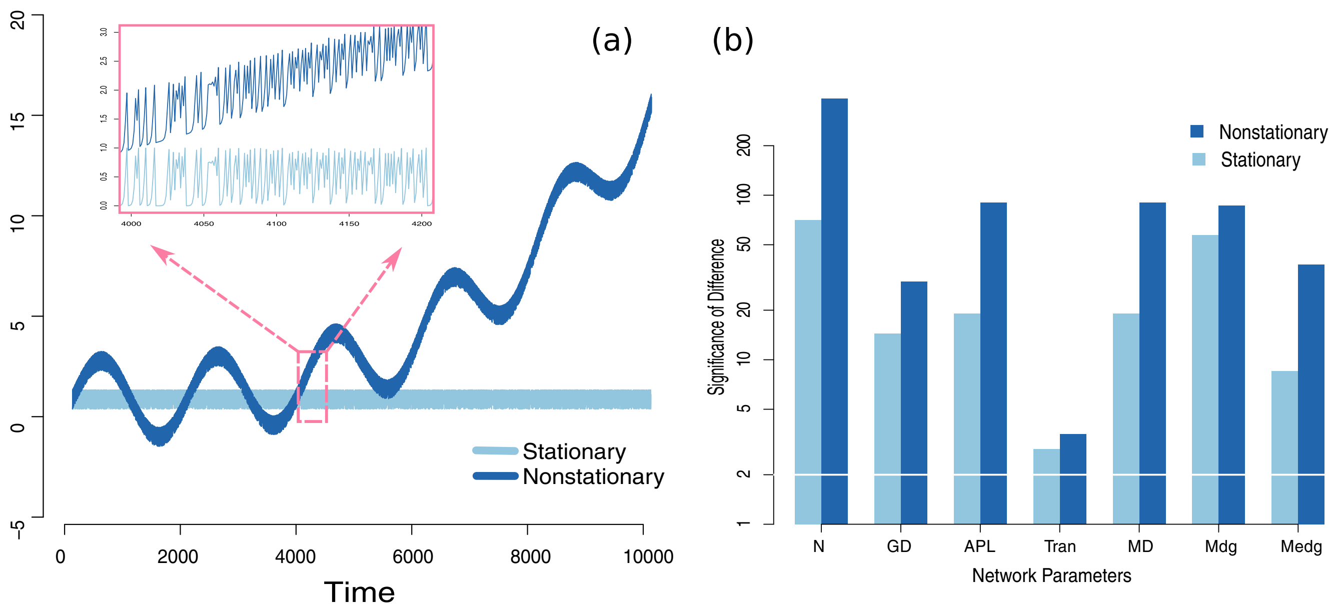

FIG. 3. (a) The stationary and the trend $t^3 + \sin(t)$ added nonstationary time series of the logistic map. (b) The significance of difference S computed with 40 surrogates of both time series. A value of S greater than 2 indicates that the null hypothesis can be rejected.

choice of the discriminating statistic μ, a measure of the significance of difference for accepting or rejecting the null hypothesis is given by

$$S = \frac{\mu_{\text{ori}} - \mu_{\text{surr}}}{\sigma_{\text{surr}}},$$

where μ_{surr} and σ_{surr} are the mean and standard deviation of μ calculated from the networks of the surrogates, respectively, and μ_{ori} is the value of μ as computed from the network of the original data. If the value of S is larger than 2, then the null hypothesis may be rejected at a 5% level of significance.

IV. RESULTS AND DISCUSSION

The network-based method for the surrogate test was applied to several kinds of time series data with different types of trends and short-term oscillations.

First, we considered a stationary AR(1) series of 10 000 data points generated by the AR model

$$x_t = x_{t-1}/2 + a_t,$$

where a_t is a Gaussian series. The resulting series is linear, Gaussian, and stationary since the unique root two lies outside the unit circle. We then added a trend $f(t) = t^3 + \sin(t)$ to this time series to produce a nonstationary time series. The autocorrelation function (ACF), partial autocorrelation function (PACF), and ugmented Dickey-Fuller (ADF) tests confirmed that the AR(1) process added with a trend is nonstationary and the AR(1) process is stationary. We then subtracted the line connecting the endpoints of the series to remove any end mismatch using Eq. (3) as discussed in the previous section. From the resulting series, 40 surrogates were generated using the IAAFT method. The GSDNNS networks were then constructed for both the original and the surrogate series, using segments of 50 data points for the piecewise linear approximation of the trend line and applying a six-bit encoding to generate the binary series as described in the previous section. The process of constructing the binary series based on segmental trend curve eliminates the possibility of manifesting the impact of the global nonstationarity on the constructed GSDNNS network. Several network parameters were computed from the networks of both the original nonstationary series as well as the surrogates, and the significance of difference S was calculated in each case. For comparison, the procedure was repeated for the stationary AR(1) series (without the trend), computing the significance of difference S for all the network parameters as in the nonstationary series. A schematic diagram of these procedures, including the results of the computations for both the stationary and nonstationary cases, is shown in Fig. 2. Figure 2(b) shows the GSDNNS network of the nonstationary series with the various colors representing the different clusters in the network. It is clear from Fig. 2(d) that the significance of difference S is smaller than two for all the network metrics computed for both the stationary and nonstationary series, showing that the null hypothesis can be accepted. This confirms that the fluctuations are generated from a linear process and that the presence of the long-term trend does not affect the test results.

To test the method against a nonlinear time series with a nonstationary trend, we generated 10 000 data points from the logistic map in the chaotic regime. The surrogates and networks were generated as before, and the values of S for the various network parameters were computed. The series was then made nonstationary by adding the same trend function $f(t)$ as before, and the computations were repeated. The results are shown in Fig. 3. The value of S in all cases was found to be greater than 2, leading to rejection of the null hypothesis, thus indicating that the series is not a linear stochastic process. In both the above cases, the trend considered was deterministic. In order to check our technique for the case of stochastic

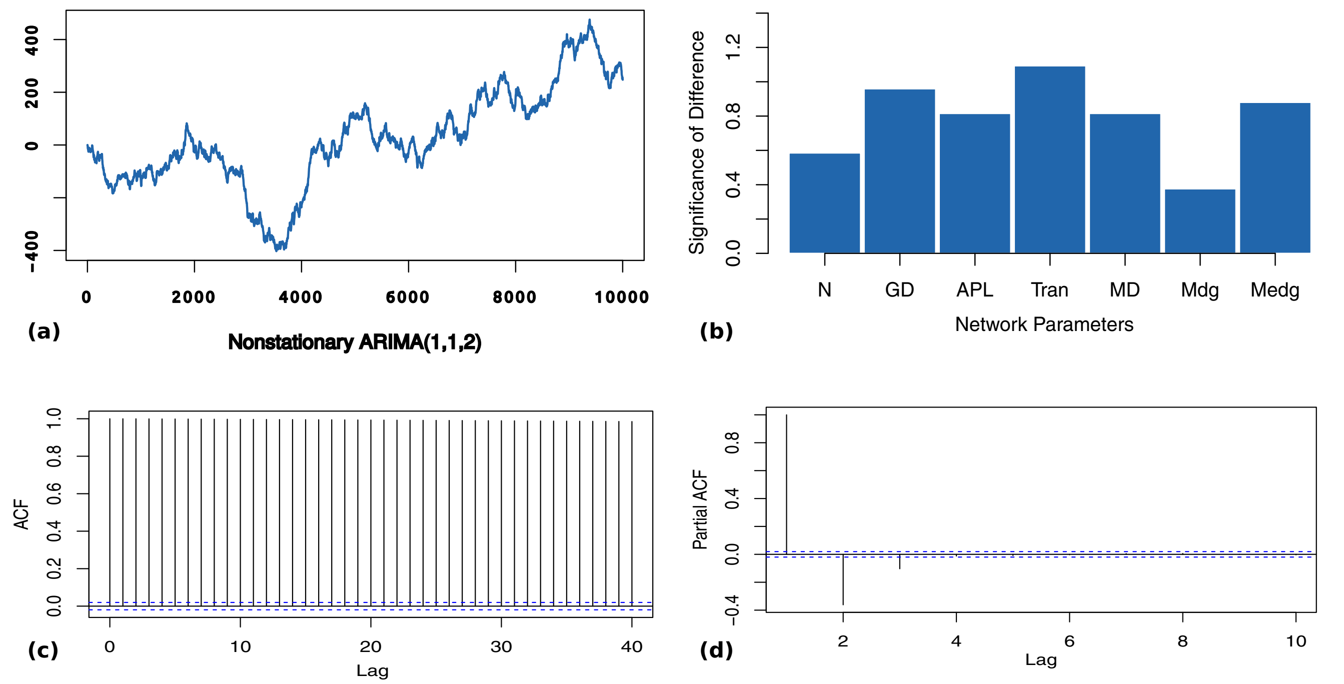

FIG. 4. (a) The nonstationary ARIMA(1,1,2) series. (b) Significance of difference S computed with 40 surrogates of the time series. The values of S being less than 2 indicate that the null hypothesis cannot be rejected. (c) The auto correlation function of the series. (d) Partial autocorrelation of the series. The augmented Dickey-Fuller test rendered P value 0.4472 indicating nonstationarity of the series.

trends, we consider a nonstationary ARIMA process and a random walk. The ARIMA(1,1,2) is a linear nonstationary process. We generated 10 000 points of the process from

$$(1-0.5)(1-B)x_t = (1+0.8B+0.3B^2)a_t. \tag{4}$$

The process is nonstationary because of its AR unit root and was so confirmed by the ACF, PACF, and ADF tests [cf. Figs. 4(c) and 4(d)]. We carried out the network-based surrogate data test as described earlier with 20 data points for the trend approximation, and the results are shown in Eq. (4). The significance of the difference lies well below the critical value, showing the linearity cannot be rejected at the 95% confidence level.

The results of applying the method on a nonstationary random walk generated by an AR(1) process with a unit root

$$x_t = x_{t-1} + a_t, \tag{5}$$

where a_t is Gaussian process are plotted in Fig. 5 along with the series generated. The computed value of the significance of the difference is seen to be less than 2 for all the network parameters, indicating at the 95% confidence level that the signal is generated by a linear process. In this case, as pointed out by Lucio *et al.* [18], the traditional methods such as AAFT or IAAFT and the TFTS algorithm of Nakamura *et al.* [17] would have led to a rejection of the null hypothesis, leading to the incorrect conclusion that the process is not linear. Thus the present method, like the method of Lucio *et al.* [18], may be applied when the underlying process has a unit root.

We generated 10 000 iterates of the Henon map corresponding to chaotic parameter values for another example of a deterministic stationary nonlinear process. The x-component of the iterates was used for the test. As shown in Fig. 6, the network-based method indicates that the significance of the difference is large enough to reject the null, suggesting that the original series is not linear. Together with the results on the logistic series, we find that the test can be used to confirm nonlinearity whether the given series is stationary or not.

A further application of the method concerns a nonlinear nonstationary stochastic series for which we consider the model

$$x_t = x_{t-1} + a_t a_{t-1}, \tag{6}$$

where $a_t \sim \mathcal{N}(0,1)$. The product term $a_t a_{t-1}$ contributes the nonlinearity of the model. The ADF test gives the P value 0.7735, indicating the nonstationarity of the data. The results of the analyses are given in Fig. 7. As expected, significant differences are observed between the original and surrogates, demonstrating that the test is successful in capturing the nonlinearity.

It may be noted that the proposed method uses the IAAFT algorithm without any modification for the construction of surrogates, except for the correction of the end mismatch of the time series before applying the algorithm. Hence it is apparent that what makes the surrogate analysis of nonstationary data possible in the present method is the construction of the decimalized network and using the network metrics as test statistics. This is confirmed by a comparison of the results obtained by crossing the surrogate technique and test statistics of the current method with those of Nakamura *et al.* [17] for typical nonstationary linear and nonlinear processes as shown in Figs. 8(a) and 8(b). Nakamura *et al.* [17] use the truncated Fourier transform (TFT) technique for generating surrogates and average mutual information for the test statistic. As is clear from these figures, in the presence of nonstationarity, the network-based test statistics succeeds in distinguishing

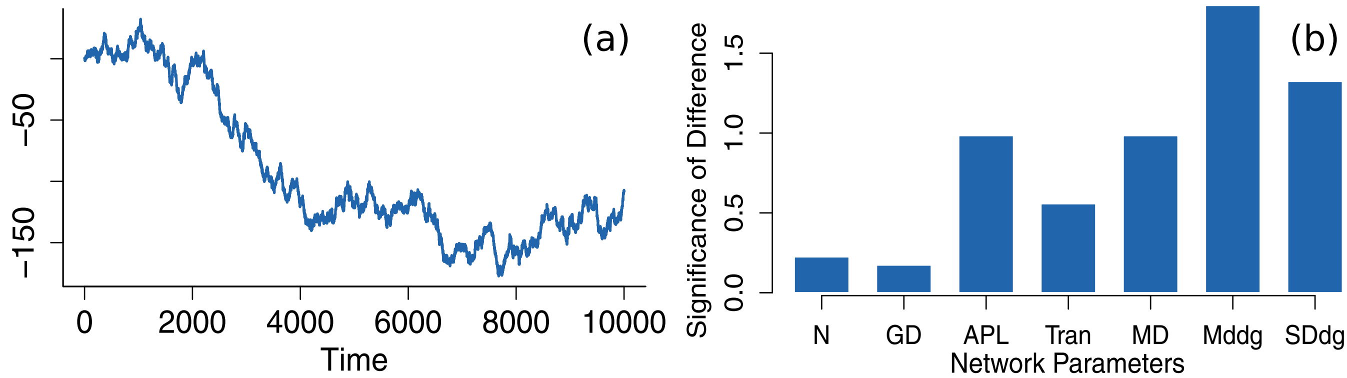

FIG. 5. (a) A random walk series. (b) Significance of difference S computed with 40 surrogates of the time series for the various network parameters. The values of S are less than 2, suggesting that the null hypothesis cannot be rejected.

nonlinear processes from linear stochastic processes when combined with the TFT surrogates as well, while average mutual information as a test statistic fares poorly when combined with plain IAAFT surrogates even after rectifying the end mismatch of the given time series.

A. Effect of noise

Another advantage of the network-based surrogate test proposed in this work is its tolerance to the presence of noise up to a moderate level in nonlinear deterministic systems. The method can detect dominant nonlinearity in a given time series even if modest levels of noise corrupt it, be it experimental or dynamical.

To explore how the presence of experimental noise in an otherwise deterministic process would affect the performance of the surrogate test, we considered the Lorenz system

$$
\begin{aligned}
dx/dt &= \sigma(y - x),\\
dy/dt &= -y + x(r - z),\\
dz/dt &= +xy - bz
\end{aligned}
\tag{7}
$$

and set the parameters at $\sigma = 10$, $b = 8/3$, and $r = 40$ for which the system is chaotic. The system was then numerically integrated, and the output was sampled at intervals of $\delta t = 0.4$ time units. The series formed by 10 000 points of the x-component x_t of the output was taken as a typical chaotic time series produced by a purely nonlinear deterministic system. By adding white noise to this series we get a model of a deterministic series affected by noise,

$$
y_t = x_t + \alpha \mathcal{N}(0, 1), \tag{8}
$$

where $N(0, 1)$ is a Gaussian noise process having zero mean and unit variance and α is a parameter representing the noise level. The results for the significance of difference at various noise levels with transitivity as test statistic are plotted in Fig. 9(a). At zero noise level, the test captures the nonlinearity as expected, and it continues to report nonlinearity for a considerable range of noise levels, indicating the robustness of the method to the presence of experimental noise.

To study the effect of dynamical noise, we again considered the Lorenz system with the same set of parametric values as before and added a noise term $\alpha \mathcal{N}(0, 1)$ to the right side of each equation in (8). The system was then numerically integrated for different values of α in the range $0 < \alpha < 1.0$, and 10 000 points of the x-component of the output, for each value of α, were selected for the analysis. The results for the significance of difference, with average path length as the test statistic, are plotted in Fig. 9(b) for various values of the noise level. It is seen that the method rejects linear-

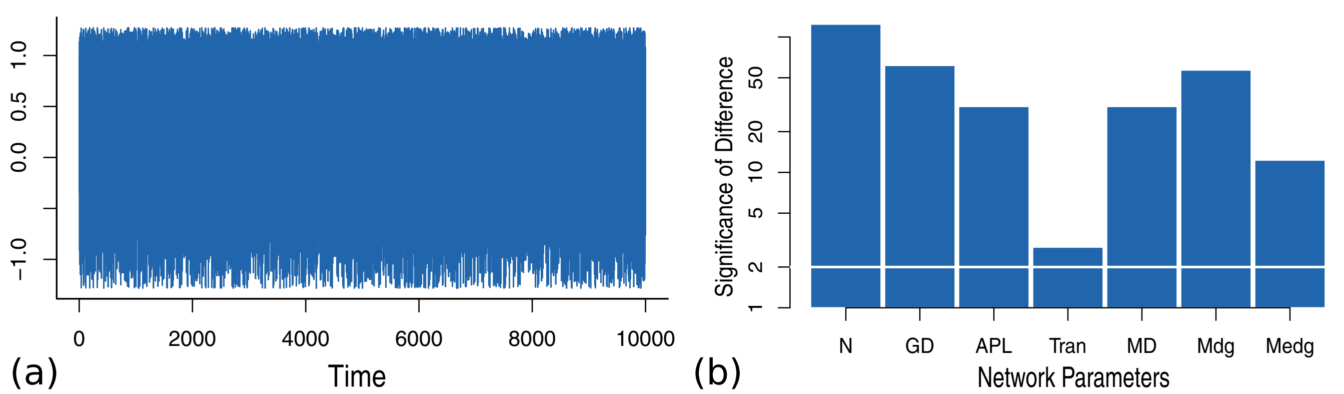

FIG. 6. (a) Stationary Henon map series. (b) Significance of difference S computed with 40 surrogates of the time series. The values of S are greater than 2, indicating rejection of the null hypothesis.

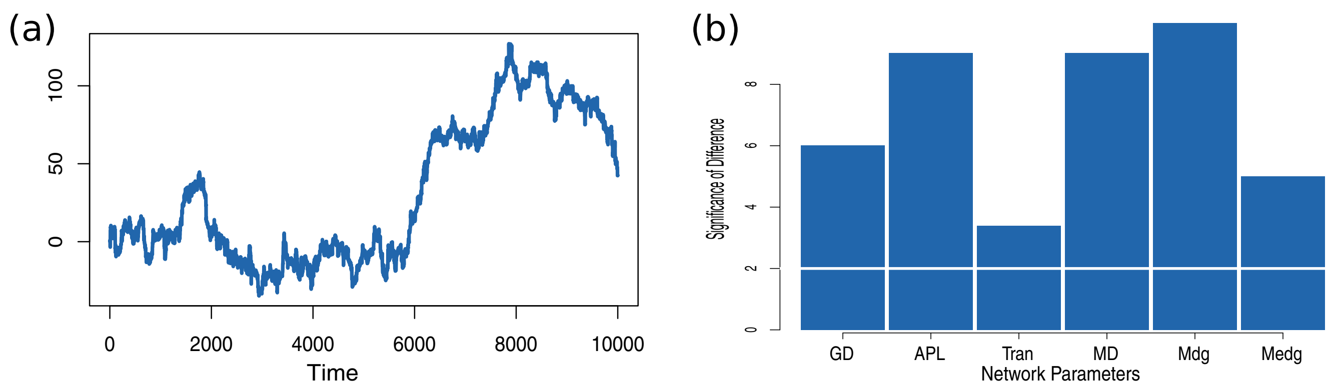

FIG. 7. (a) A realization of the nonlinear nonstationary stochastic process given by Eq. (6). (b) Significance of difference S computed with 40 surrogates of the time series. The values of S are higher than 2, indicating that the null hypothesis can be rejected.

ity in favor of the underlying nonlinearity for a reasonably large range of the noise level, indicating that the method is also robust to the presence of dynamical noise of a moderate degree.

The successful applications of the network-based method of surrogate tests suggest that the technique can be used to detect linearity or points to the nonlinearity of a given series, whether it is stationary or nonstationary with a deterministic or stochastic trend.

V. CONCLUSIONS

In this paper, we introduce a generalized symbolic dynamics approach for constructing a network from a nonstationary time series for effectively capturing the dynamics of oscillations around a long-term trend. As an application of such networks, we demonstrate that surrogate data tests with discriminating statistics based on the structural properties of such networks can correctly detect nonlinearity present in a nonstationary time series, while most of the conventional methods would confuse nonstationarity with nonlinearity. The efficacy of the proposed computational procedure has been tested on various linear, nonlinear, stationary, and nonstationary processes. The nonstationary processes include those with deterministic and stochastic trends. The advantage of the procedure is that it can be automated independent of the nature of the trend or stationarity. While the search for the suitable test statistic is a focus area for many researchers, the network induced by the time series features several structural properties which can serve as possible test statistics. The method's robustness has also been tested by considering time series corrupted by both measurement noise and dynamical noise.

ACKNOWLEDGMENTS

The authors wish to thank the campus computing facility of the University of Kerala set up under the DST-PURSE program for providing computational facilities. M.C.M. would like to acknowledge the financial assistance through the e-grants scheme of the Government of Kerala. T.R.R. wishes to

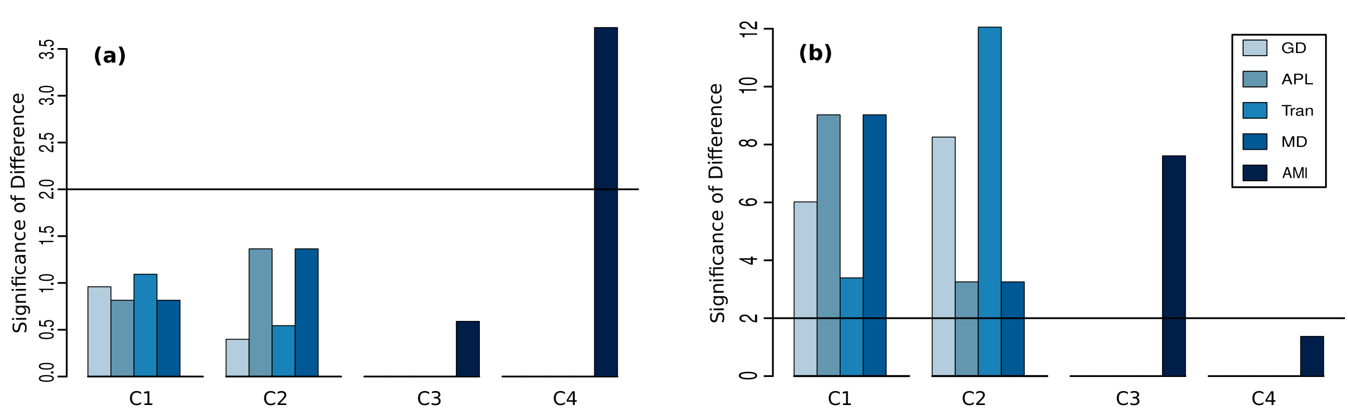

FIG. 8. (a) Comparison of different combination of surrogate methods and test statistics for the linear nonstationary system given by Eq. (4). (b) Comparison of different combination of surrogate methods and test statistics for the nonlinear nonstationary system given by Eq. (6). The combinations are given by C1, IAAFT surrogates and network-based test statistics; C2, TFT surrogates of Nakamura *et al.* [17] and the network-based test statistics; C3, TFT surrogates and average mutual information for test statistic as in Nakamura *et al.* [17]; and C4, IAAFT surrogates and average mutual information for test statistic. Noted that combination C4 fares poorly for both linear and nonlinear cases when the given data are nonstationary.

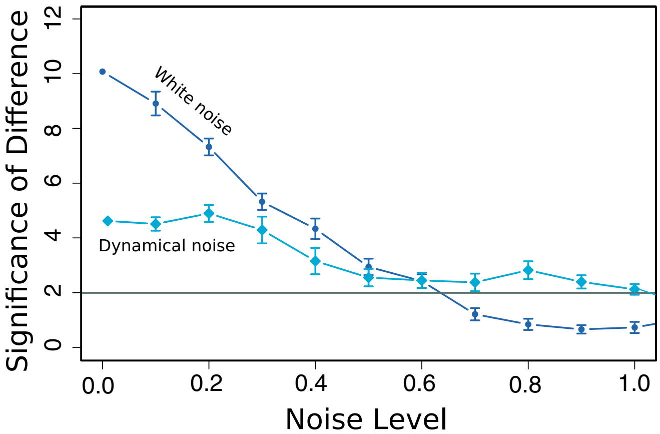

FIG. 9. Plot of significance of difference vs white and dynamical noise level for the chaotic time series of Lorenz system given by Eq. (7).

acknowledge encouragement from the Principal and the Head of the Department of Chemical Engineering of MSRIT.

[1] J. L. Castle, J. A. Doornik, and D. F. Hendry, Modelling non-stationary 'big data', Int. J. Forecast. **37**, 1556 (2021).

[2] G. Dufrénot and T. Matsuki, *Recent Econometric Techniques for Macroeconomic and Financial Data* (Springer, Cham, 2021).

[3] A. Y. Kaplan, A. A. Fingelkurts, A. A. Fingelkurts, S. V. Borisov, and B. S. Darkhovsky, Nonstationary nature of the brain activity as revealed by EEG/MEG: Methodological, practical and conceptual challenges, Signal Process. **85**, 2190 (2005).

[4] J. R. Freeman, Progress in the study of nonstationary political time series: A comment, Polit. Anal. **24**, 50 (2016).

[5] N. L. Poff, Beyond the natural flow regime? Broadening the hydro-ecological foundation to meet environmental flows challenges in a nonstationary world, Freshwater Biol. **63**, 1011 (2018).

[6] T. Chakraborty, I. Ghosh, T. Mahajan, and T. Arora, Nowcasting of COVID-19 confirmed cases: Foundations, trends, and challenges, in *Modeling, Control and Drug Development for COVID-19 Outbreak Prevention* (Springer, Cham, 2022), pp. 1023–1064.

[7] C. R. Rollinson, A. O. Finley, M. R. Alexander, S. Banerjee, K.-A. Dixon Hamil, L. E. Koenig, D. H. Locke, M. Peterson, M. W. Tingley, K. Wheeler *et al.*, Working across space and time: Nonstationarity in ecological research and application, Front. Ecol. Environ. **19**, 66 (2021).

[8] J. B. Heffernan, P. A. Soranno, M. J. Angilletta, Jr, L. B. Buckley, D. S. Gruner, T. H. Keitt, J. R. Kellner, J. S. Kominoski, A. V. Rocha, J. Xiao *et al.*, Macrosystems ecology: Understanding ecological patterns and processes at continental scales, Front. Ecol. Environ. **12**, 5 (2014).

[9] E. Wolkovich, B. Cook, K. McLauchlan, and T. Davies, Temporal ecology in the Anthropocene, Ecol. Lett. **17**, 1365 (2014).

[10] S. L. Collins, M. L. Avolio, C. Gries, L. M. Hallett, S. E. Koerner, K. J. La Pierre, A. L. Rypel, E. R. Sokol, S. B. Fey, D. F. Flynn *et al.*, Temporal heterogeneity increases with spatial heterogeneity in ecological communities, Ecology **99**, 858 (2018).

[11] T. Cornulier and V. Bretagnolle, Assessing the influence of environmental heterogeneity on bird spacing patterns: A case study with two raptors, Ecography **29**, 240 (2006).

[12] A. O. Finley, Comparing spatially-varying coefficients models for analysis of ecological data with nonstationary and anisotropic residual dependence, Methods Ecol. Evol. **2**, 143 (2011).

[13] J. A. Miller and R. Q. Hanham, Spatial nonstationarity and the scale of species–environment relationships in the Mojave Desert, California, USA, Int. J. Geog. Inf. Sci. **25**, 423 (2011).

[14] G. Papoudakis, F. Christianos, A. Rahman, and S. V. Albrecht, Dealing with non-stationarity in multi-agent deep reinforcement learning, arXiv:1906.04737 (2019).

[15] J. Theiler, S. Eubank, A. Longtin, B. Galdrikian, and J. D. Farmer, Testing for nonlinearity in time series: The method of surrogate data, Physica D **58**, 77 (1992).

[16] T. Nakamura and M. Small, Small-shuffle surrogate data: Testing for dynamics in fluctuating data with trends, Phys. Rev. E **72**, 056216 (2005).

[17] T. Nakamura, M. Small, and Y. Hirata, Testing for nonlinearity in irregular fluctuations with long-term trends, Phys. Rev. E **74**, 026205 (2006).

[18] J. H. Lucio, R. Valdés, and L. R. Rodríguez, Improvements to surrogate data methods for nonstationary time series, Phys. Rev. E **85**, 056202 (2012).

[19] T. Schreiber and A. Schmitz, Improved Surrogate Data for Nonlinearity Tests, Phys. Rev. Lett. **77**, 635 (1996).

[20] T. Schreiber and A. Schmitz, Surrogate time series, Physica D **142**, 346 (2000).

[21] G. Lancaster, D. Iatsenko, A. Pidde, V. Ticcinelli, and A. Stefanovska, Surrogate data for hypothesis testing of physical systems, Phys. Rep. **748**, 1 (2018).

[22] J. Timmer, Power of surrogate data testing with respect to nonstationarity, Phys. Rev. E **58**, 5153 (1998).

[23] C. J. Keylock, Constrained surrogate time series with preservation of the mean and variance structure, Phys. Rev. E **73**, 036707 (2006).

[24] C. J. Keylock, A wavelet-based method for surrogate data generation, Physica D **225**, 219 (2007).

[25] M. Breakspear, M. Brammer, and P. A. Robinson, Construction of multivariate surrogate sets from nonlinear data using the wavelet transform, Physica D **182**, 1 (2003).

[26] J. Zhang and M. Small, Complex Network from Pseudoperiodic Time Series: Topology Versus Dynamics, Phys. Rev. Lett. **96**, 238701 (2006).

[27] J. Zhang, X. Luo, and M. Small, Detecting chaos in pseudoperiodic time series without embedding, Phys. Rev. E **73**, 016216 (2006).

[28] Y. Yang and H. Yang, Complex network-based time series analysis, Physica A **387**, 1381 (2008).

[29] Z. Gaoa nd N. Jin, Flow-pattern identification and nonlinear dynamics of gas-liquid two-phase flow in complex networks, Phys. Rev. E **79**, 066303 (2009).

[30] L. Lacasa, B. Luque, F. Ballesteros, J. Luque, and J. C. Nuno, From time series to complex networks: The visibility graph, Proc. Natl. Acad. Sci. USA **105**, 4972 (2008).

[31] X. Xu, J. Zhang, and M. Small, Superfamily phenomena and motifs of networks induced from time series, Proc. Natl. Acad. Sci. USA **105**, 19601 (2008).

[32] N. Marwan, J. F. Donges, Y. Zou, R. V. Donner, and J. Kurths, Complex network approach for recurrence analysis of time series, Phys. Lett. A **373**, 4246 (2009).

[33] V. L. Freitas, J. C. Lacerda, and E. E. Macau, Complex networks approach for dynamical characterization of nonlinear systems, Int. J. Bifurcation Chaos **29**, 1950188 (2019).

[34] S. Suriyaprabhaa, G. Gopinath, R. Sangeerthana, S. Alfiya, P. Asha, and K. S. Kumar, Generalized symbolic dynamics approach for characterization of time series, in *Advances in Computing and Network Communications* (Springer, Singapore, 2021), pp. 53–61.

[35] T. Schreiber and A. Schmitz, Discrimination power of measures for nonlinearity in a time series, Phys. Rev. E **55**, 5443 (1997).

[36] R. Hegger, H. Kantz, and T. Schreiber, Practical implementation of nonlinear time series methods: The TISEAN package, Chaos **9**, 413 (1999).

[37] M. C. Mallika, K. Asokan, K. S. A. Kumar, T. R. Ramamohan, and K. S. Kumar, Improved tests for non-linearity using network-based statistics and surrogate data, Pramana **95**, 141 (2021).

[38] Y. Hirata, M. Shiro, and J. M. Amigó, Surrogate data preserving all the properties of ordinal patterns up to a certain length, Entropy **21**, 713 (2019).