CHIN. PHYS. LETT. Vol. 34, No. 5 (2017) 050501

Chaos Identification Based on Component Reordering and Visibility Graph *

Sheng-Li Zhu(朱胜利), Lu Gan(甘露)**

School of Electronic Engineering, University of Electronic Science and Technology of China, Chengdu 611731

(Received 6 January 2017)

The identification between chaotic systems and stochastic processes is not easy since they have numerous similarities. In this study, we propose a novel approach to distinguish between chaotic systems and stochastic processes based on the component reordering procedure and the visibility graph algorithm. It is found that time series and their reordered components will show diverse characteristics in the 'visibility domain'. For chaotic series, there are huge differences between the degree distribution obtained from the original series and that obtained from the corresponding reordered component. For correlated stochastic series, there are only small differences between the two degree distributions. For uncorrelated stochastic series, there are slight differences between them. Based on this discovery, the well-known Kullback–Leible divergence is used to quantify the difference between the two degree distributions and to distinguish between chaotic systems, correlated and uncorrelated stochastic processes. Moreover, one chaotic map, three chaotic systems and three different stochastic processes are utilized to illustrate the feasibility and effectiveness of the proposed method. Numerical results show that the proposed method is not only effective to distinguish between chaotic systems, correlated and uncorrelated stochastic processes, but also easy to operate.

PACS: 05.45.Tp, 74.40.De, 05.45.Pq DOI: 10.1088/0256-307X/34/5/050501

In recent years, the analysis of nonlinear time series is still an active research field.[1−3] In this field, the identification between chaotic systems and stochastic processes is very fundamental and it is essential if one intends to model the associated phenomenon and to determine the relevant quantifiers.[4] However, it is not an easy task since chaotic systems and stochastic processes share several properties, e.g., a wide-band power spectrum, a delta-like autocorrelation function, and an irregular behavior of the measured signals.[5] In fact, this similarity has made it possible to replace stochastic processes by chaotic systems in many applications. In the last few decades, numerous methods have been proposed to distinguish chaotic systems from stochastic processes.[4−9]

Recently, the visibility graph (VG) method, which was introduced by Lacasa *et al.*,[10] has attracted extensive attention.[11−14] The VG technique is a statistical algorithm that maps a time series into a network or graph according to a simple geometric criteria. It has been shown that the VG not only inherits several properties of time series into its structure, but also captures the dynamical fingerprints of the process of generating time series. Specifically, periodic time series can be converted into regular graphs, fractal time series can be mapped into scale-free networks, and random series can be converted into random graphs.[10,15] As a result, methods of complex network theory can be exploited to characterize time series.

Generally, given a time series $\{x_t\}_{t=1}^N$ of length N, the associated VG contains N nodes (each point represents a node), two arbitrary data values (i.e., two nodes) (t_i, x_i) and (t_j, x_j) will have connection in the associated VG if every data value (t_l, x_l) between them (i.e., $t_i < t_l < t_j$) fulfills the following geometric criteria

$$x_l < x_i + (x_j - x_i)\frac{t_l - t_i}{t_j - t_i}. \tag{1}$$

It should be noted that the VG derived from any time series is always connected, undirected and invariant under affine transformations[10] (i.e., invariant under rescaling of both horizontal and vertical axes). Two illustrative examples for VG are given in Figs. 1(a) and 1(b).

The VG method makes it possible to investigate the time series in the 'graph domain' or the 'visibility domain'[16] since the structure of time series is inherited in the associated VG. Here we attempt to distinguish chaotic systems from stochastic processes in the visibility domain. However, it is not an easy task if we just use the VG technique. In fact, some information of time series has been lost due to the simpleness of geometric criteria of the VG method. Hence, the distinction between the degree distributions obtained from chaotic series and those derived from stochastic series sometimes becomes very subtle. In this work, we show that the identification between chaotic systems and stochastic processes can be performed successfully in the visibility domain with the help of the reordered component (or reordered series).[17] A reordered component is generated by a very simple procedure called component reordering. Actually, the component reordering procedure can be implemented as follows: (1) obtaining the first component $\{x_t^{\rm f}\}_{t=1}^{N-\tau} = \{x_t\}_{t=1}^{N-\tau}$ and the second component $\{x_t^{\rm s}\}_{t=1}^{N-\tau} = \{x_t\}_{t=\tau+1}^{N}$ of the original series $\{x_t\}_{t=1}^N$, where τ is the embedding delay. (2) Sorting the first component $\{x_t^{\rm f}\}_{t=1}^{N-\tau}$ with an ascending order, letting $t_{\rm new}$ be the new subscript after sorting. (3) Obtaining the new time series $\{z_t\}_{t=1}^{N-\tau}$ with $z_t = x_{t_{\rm new}}^{\rm s}$.

*Supported by the National Natural Science Foundation of China under Grant No U1530126.
**Corresponding author. Email: ganlu@uestc.edu.cn

Time series $\{z_t\}_{t=1}^{N-\tau}$ derived from the above steps is called the reordered component since it is actually a reordered version of the second component. The component reordering procedure is based on the incomplete two-dimensional reconstruction of the phase-space.[17] Notice that the embedding delay τ is also important for the component reordering procedure, we will fix it to $\tau = 1$, since the structure information captured by the incomplete reconstruction is considerable under this value.

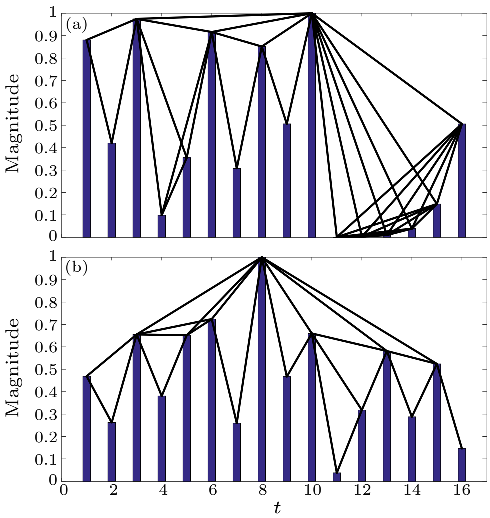

Fig. 1. The VG obtained from (a) logistic series and (b) Gaussian white noise. There are 16 data points that are used. The Gaussian white noise is normalized to $[0, 1]$ before deriving the VG.

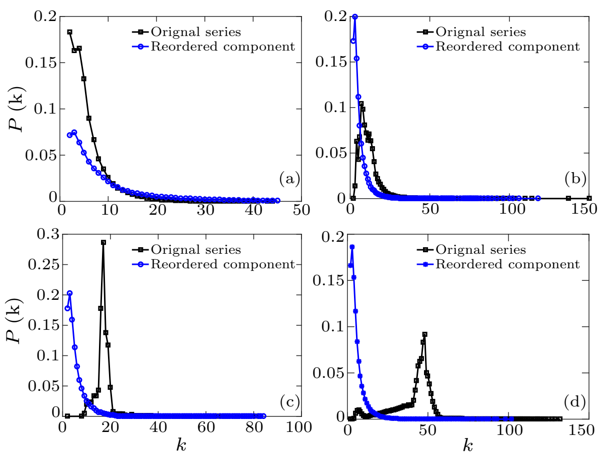

Fig. 2. Regular plot of the degree distribution of a VG associated to a time series from (a) logistic map, (b) Lorenz system, (c) Rössler system and (d) HyperRössler system. The value of k represents the degree of the node, and $P(k)$ indicates the probability that the degree of a node is k. The squares and circles represent the distribution of the original time series and the associated reordered component, respectively. Here 10^5 data points are used.

For chaotic series, there is a striking similarity between the incomplete two-dimensional reconstructed phase-space and the corresponding reordered component.[17] Moreover, the information captured by the incomplete reconstructed phase-space will be transferred to the reordered component.[17] Since stochastic processes arise from an infinite-dimensional attractor,[18] the incomplete reconstructed phase-space cannot capture any useful information of the phase-space of stochastic processes.

We will obtain two degree distributions by applying the VG method to the time series under study and its reordered component. For the purpose of convenience, let Q be the degree distribution derived from the original series and P be the degree distribution derived from the corresponding reordered component.

Basically, the results of the component reordering procedure are relatively diverse for different kinds of series. For a chaotic series, the reordered component is affected by two aspects. One is obtained from the information captured by the incomplete reconstruction of the phase-space and the other is obtained from the dependence among the data. Thus the reordering procedure will produce a new time series which is quite different from the original one and the corresponding Q and P will be significantly different. As for the series generated by correlated stochastic processes, the reordered component is only affected by the correlation among the data, which will lead to small differences between the corresponding Q and P. As for the series generated by uncorrelated stochastic processes, e.g., the Gaussian white noise, the reordering procedure cannot change anything because of the independence among the data, and Q and P are almost the same. In this work we consider the following four chaotic systems. (1) The logistic map defined by

$$x_{t+1} = r x_t (1 - x_t), \tag{2}$$

where $r = 4$ is chosen to make the map operate in the chaotic regime. (2) The x-coordinate of the Lorenz system

$$\begin{cases} \dot{x} = & \sigma y - \sigma\, x, \\ \dot{y} = & r x - y - x z, \\ \dot{z} = & -b z + x y, \end{cases} \tag{3}$$

where $\sigma = 10$, $r = 28$ and $b = 8/3$ are chosen to make the system operate in the chaotic regime. (3) The x-coordinate of the Rössler system

$$\begin{cases} \dot{x} = & -y - z, \\ \dot{y} = & x + a y, \\ \dot{z} = & b + z(x - c), \end{cases} \tag{4}$$

where $a = 0.15$, $b = 0.2$ and $c = 10$ are chosen to make the system operate in the chaotic regime. (4) The x-coordinate of the Hyper–Rössler system:

$$\begin{cases} \dot{x} = & -y - z, \\ \dot{y} = & x + a y + w \\ \dot{z} = & b + x z \\ \dot{w} = & -c z + d w, \end{cases} \tag{5}$$

where $a = 0.25$, $b = 3$, $c = 0.5$, $d = 0.05$ and $d = 0.05$ are chosen to make the system operate in the chaotic regime.

We will deal with the following three stochastic processes. (1) The Gaussian white noise generated by the 'randn.m file' in the Matlab. (2) The pink noise with f^{-1} power spectral. (3) The AR(1) model is defined as follows:

$$x_{t+1} = 0.3x_t + \varepsilon(t), \tag{6}$$

where $\varepsilon(t) \sim iid$ and $\mathcal{N}(0,1)$.

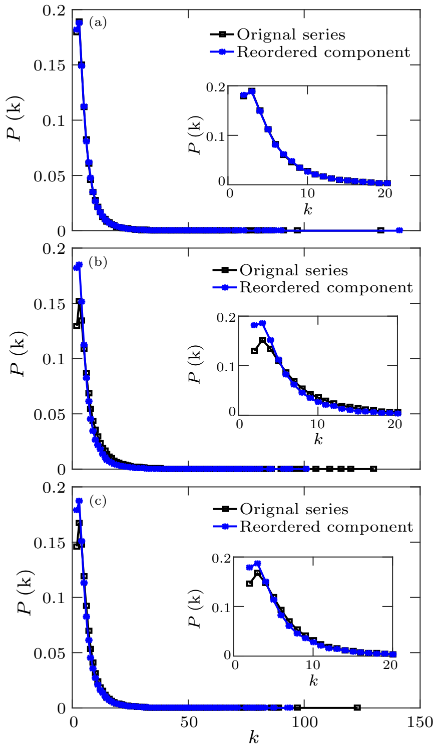

Fig. 3. Regular plot of the degree distribution of a VG associated to (a) Gaussian white noise, (b) $1/f$ noise and (c) AR(1). The insets show the enlargement for $k \in [0, 20]$ to observe more clearly. The value of k represents the degree of the node and $P(k)$ indicates the probability that the degree of a node is k. The squares and circles represent the distribution of the original time series and the associated reordered component, respectively. Here 10^5 data points are used.

To investigate the difference between the original series and its reordered component in the visibility domain for chaotic systems, in Fig. 2 we plot Q and P obtained from four chaotic series of 10^5 data points. As can be easily observed from Fig. 2, there are significant differences between Q and P. In the case of logistic series, the probability of the nodes with small degrees is decreased in the reordered component compared with that of the original series. For the Lorenz time series, there is a peak in Q, whereas it disappears in P. The similar results can be found for the Rössler and the HyperRössler series.

We also plot Q and P obtained from the stochastic series and the results are shown in Fig. 3. On the one hand, for the uncorrelated stochastic process (Gaussian white noise), Q and P are almost the same, as can be seen in Fig. 3(a). On the other hand, for correlated stochastic processes ($1/f$ noise and AR(1)), there are small differences between Q and P, as can be seen in Figs. 3(b) and 3(c). The results shown in Figs. 2 and 3 are in good agreement with the above discussions and demonstrate that the identification between chaotic systems, uncorrelated and correlated stochastic processes can be performed well in the visibility domain.

Although the differences between Q and P can be perceived by human eyes, it would be better to identify such differences from a statistical point of view. To quantify the differences between Q and P, the well-known Kullback–Leible divergence (KLD),[19] which is denoted as $D_{\rm KLD}(Q||P)$, is used in the present study. For discrete probability distributions Q and P, their KLD $D_{\rm KLD}(Q||P)$ is defined as follows:

$$D_{\rm KLD}(Q||P) = \sum_i Q(i) \ln \frac{Q(i)}{P(i)}. \tag{7}$$

The KLD is a semi-distance (i.e., non-symmetric) which is zero if and only if $Q = P$ and positive otherwise. It is widely used to measure the similarity (distance) between two distributions.[20] It should be noted that the KLDs obtained from different degree distributions are always larger than zero due to the finite data length. Nevertheless, a relatively large KLD means that the two distributions are quite different. As a technical remark, notice that $D_{\rm KLD}(Q||P)$ diverges if Q and P have different supports (i.e., if $Q(i) = 0$, $P(i) \neq 0$ or $Q(i) \neq 0$, $P(i) = 0$ for some value i). To solve this problem, a common procedure[21] is to introduce a bias of order $\mathcal{O}(1/N^2)$, where N is the length of the time series. In other words, we replace all vanishing frequencies with $1/N$, and normalize the frequency histogram appropriately.

To test the validity and effectiveness of the proposed method, $D_{\rm KLD}(Q||P)$ obtained from different time series with diverse data lengths are calculated and the results are listed in Table 1. It should be pointed out that these results are derived from 100 initial conditions and are averaged.

It can be observed from Table 1 that the KLDs for diverse time series are significantly different. For chaotic series, the values of KLDs are always far larger than zero, which shows the significant differences between P and Q. For the Gaussian white noise, the values of KLDs are quite small in all cases, which means that the corresponding P and Q are almost the same. As for correlated stochastic processes, the values of KLDs are larger than those of uncorrelated processes, but much smaller than those of chaotic series. The results listed in Table 1 demonstrate that the proposed method is effective to distinguish between chaotic systems, uncorrelated and correlated stochastic processes.

In general, the surrogate data technique is broadly applied to generate new numerical samples for identification of deterministic or stochastic processes.[22] The surrogate data share with time series under study certain properties and also fulfill a certain null hypothesis. Here we also study the performance of the surrogate data in the visibility domain. Let S be the degree distribution obtained from the surrogate data of the underlying time series. The KLD $D_{\rm KLD}(Q||S)$ obtained from different time series with diverse data lengths are calculated and the results are listed in Table 2. Note that the results listed in Table 2 are obtained from 100 surrogate series and are averaged.

Table 1. The KLDs between P and Q derived from diverse time series with different data lengths.

	Logistic	Lorenz	Rössler	HyperRössler	Gaussian	$1/f$ noise	AR(1)
$N = 1 \times 10^5$	0.1648	0.3354	1.3231	2.1193	3.0314×10^{-4}	0.0152	0.0040
$N = 5 \times 10^5$	0.1715	0.3344	1.3255	2.0546	6.2955×10^{-5}	0.0149	0.0036
$N = 1 \times 10^6$	0.1743	0.3334	1.3227	2.0619	4.7104×10^{-5}	0.0151	0.0034

Table 2. The KLDs between S and Q derived from diverse time series with different data lengths.

	Logistic	Lorenz	Rössler	HyperRössler	Gaussian	$1/f$ noise	AR(1)
$N = 1 \times 10^5$	0.0040	0.0802	0.3723	0.9341	2.6766×10^{-4}	4.0611×10^{-4}	2.9125×10^{-4}
$N = 5 \times 10^5$	0.0041	0.0787	0.3770	0.9609	9.1836×10^{-5}	9.2994×10^{-5}	8.5574×10^{-5}
$N = 1 \times 10^6$	0.0041	0.0784	0.3845	0.9611	3.5122×10^{-5}	6.1828×10^{-5}	4.1496×10^{-5}

From Table 2 we can observe that the values of KLDs derived from chaotic series are much larger than those of stochastic processes, which means that the surrogate data are also effective to distinguish chaotic systems from stochastic processes. However, the values of KLDs are relatively small for both correlated and uncorrelated stochastic processes, which means that the surrogate data cannot distinguish between uncorrelated and correlated stochastic processes in the visibility domain.

Compared with the surrogate data, the reordered components have several advantages. First, the calculation process of the reordered component is simpler than that of surrogate series. Secondly, the reordered component still preserves the information about the phase-space of the original series, while the surrogate data does not. Finally, we can distinguish between chaotic systems, correlated and uncorrelated processes by using the reordered component and the VG method, whereas we only can distinguish between chaotic systems and stochastic processes when the surrogate data and the VG technique are used.

In summary, the component reordering procedure and the VG method have been used to distinguish between chaotic systems, uncorrelated and correlated stochastic processes. For chaotic series, the reordered component is affected not only by the structure information captured by the incomplete reconstruction of the phase-space but also by the deterministic dependence among the series. Thus the original series and its reordered component are quite different in the visibility domain. As for the uncorrelated stochastic processes, the component reordering procedure has no effect on them since there is no correlation among the data. Thus the original series and its corresponding reordered component will show the same degree distribution. For correlated stochastic processes, the reordered components are only affected by the correlation among the data. Therefore, the original series and their reordered components will show small differences in the visibility domain. Based on the above principles, the KLD is used to measure the differences between the degree distribution obtained from the original series and that derived from the corresponding reordered component. Moreover, one chaotic map, three chaotic systems and three stochastic processes are exploited to test the feasibility and effectiveness of the proposed method. Numerical results show that our method is effective to distinguish between chaotic systems, uncorrelated and correlated stochastic processes. The successful identification by using the component reordering procedure and the VG method may aid in understanding of the underlying chaotic systems.

References

[1] Qu H, Ma W T, Zhao J H and Chen B D 2013 *Chin. Phys. Lett.* **30** 110505
[2] Gao Z K, Hu L D and Jin N D 2013 *Chin. Phys.* B **22** 050507
[3] Kong D R and Xie H B 2011 *Chin. Phys. Lett.* **28** 090502
[4] Rosso O A, Olivares F and Plastino A 2015 *Papers Phys.* **7** 070006
[5] Rosso O, Larrondo H, Martin M, Plastino A and Fuentes M 2007 *Phys. Rev. Lett.* **99** 154102
[6] Kaplan D T and Glass L 1992 *Phys. Rev. Lett.* **68** 427
[7] Ortega G J and Louis E 1998 *Phys. Rev. Lett.* **81** 4345
[8] Jeong J, Gore J C and Peterson B S 2002 *IEEE Trans. Biomed. Eng.* **49** 1374
[9] Amigó J M, Kocarev L and Szczepanski J 2006 *Phys. Lett.* A **355** 27
[10] Lacasa L, Luque B, Ballesteros F, Luque J and Nuno J C 2008 *Proc. Natl. Acad. Sci. USA* **105** 4972
[11] Flanagan R and Lacasa L 2016 *Phys. Lett.* A **380** 1689
[12] Jiang W, Wei B Y, Zhan J, Xie C H and Zhou D Y 2016 *Comput. Ind. Eng.* **101** 260
[13] Bhaduri A and Ghosh D 2016 *Front. Physiol.* **7** 44
[14] Telesca L, Lovallo M, Aggarwal S, Khan P and Rastogi B 2016 *Pure Appl. Geophys.* **173** 125
[15] Donner R V and Donges J F 2012 *Acta Geophys.* **60** 589
[16] Luque B, Lacasa L, Ballesteros F and Luque J 2009 *Phys. Rev.* E **80** 046103
[17] Zhu S L and Gan L 2016 *Phys. Rev.* E **94** 052210
[18] Lacasa L and Toral R 2010 *Phys. Rev.* E **82** 036120
[19] MacKay D J C 2003 *Inf. Theory Inference Learning Algorithms* (Cambridge: Cambridge University Press) p 37
[20] Cheng Y, Hua X, Wang H, Qin Y and Li X 2016 *Entropy* **18** 381
[21] Roldán É and Parrondo J M 2012 *Phys. Rev.* E **85** 031129
[22] Theiler J, Eubank S, Longtin A, Galdrikian B and Farmer J D 1992 *Physica* D **58** 77