Incomplete phase-space method to reveal time delay from scalar time series

Shengli Zhu and Lu Gan*

Center for Cyber Security, School of Electronic Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China

(Received 10 August 2016; revised manuscript received 26 September 2016; published 9 November 2016)

A computationally quick and conceptually simple method to recover time delay of the chaotic system from scalar time series is developed in this paper. We show that the orbits in the incomplete two-dimensional reconstructed phase-space will show local clustering phenomenon after the component reordering procedure proposed in this work. We find that information captured by the incomplete two-dimensional reconstructed phase-space is related to the time delay τ_0 present in the system, and will be transferred to the reordered component by the procedure of component reordering. We then propose the segmented mean variance (SMV) from the reordered component to identify the time delay τ_0 of the system. The proposed SMV shows clear maximum when the embedding delay τ of the incomplete reconstruction matches the time delay τ_0 of the chaotic system. Numerical data generated by a time-delay system based on the Mackey-Glass equation operating in the chaotic regime are used to illustrate the effectiveness of the proposed SMV. Experimental results show that the proposed SMV is robust to additive observational noise and is able to recover the time delay of the chaotic system even though the amount of data is relatively small and the feedback strength is weak. Moreover, the time complexity of the proposed method is quite low.

DOI: 10.1103/PhysRevE.94.052210

I. INTRODUCTION

Delay phenomena, which are due to the finite signal propagation speed or the memory effects, are ubiquitous in various systems, including nonlinear optics [1,2], biology [3,4], chemistry [5,6], and climatology [7,8]. It is found that even a very simple time-delay chaotic system can produce highly complex dynamics with a lot of degree of freedom [9], which makes such systems very attractive. We can find a lot of relevant applications based on the delay phenomena in nonlinear optics, for example, the chaotic radar [10] and lidar [11], the optical chaos encryption [12], rainbow refractometry [13], and ultrahigh-speed physical random number generation [14]. The time delay is important for chaos communication since the dynamics of such delayed chaotic systems can be identified and modeled once their time delay is recovered [15,16]. Consequently, the identification of time delay present in chaos communication systems would weaken their security and confidentiality [17,18]. Besides, it is necessary to determine whether there are time delays present in the scalar time series if one wants to develop suitable models for simulation and forecasting purposes.

For the reasons aforementioned, it is very necessary to study the time-delay signature present in the chaotic system. However, the great challenge is that the corresponding underlying equations or even the relevant governing mechanisms are often unknown and the time series that is always contaminated by noise is insufficient sometimes in the study of nonlinear dynamical systems. There were a lot of approaches proposed to recover the time delay τ_0 of the system from recorded time series, e.g., the autocorrelation function (ACF) and the delayed mutual information (DMI) [19,20], the filling factor analysis [21], extrema statistics [22,23], information theory methods [24,25], the practical criterion [26], and the permutation entropy and the permutation statistical complexity (C_{JS}) [27,28]. The aim of this work is to recover the time delay present in the time-delay chaotic system when the underlying equation is not known and the amount of data is relatively small. Generally speaking, the phase-space reconstruction is a fundamental tool for chaotic time-series analysis. Nonetheless, it is inappropriate to apply this technique to a time-delay system since even a first-order delay differential equation can possess high-dimensional chaotic dynamics [29], and we cannot directly reconstruct the phase-space of such a system since the phase-space of such a system has to be regarded as infinite-dimensional [23]. Recently, a new technique, which is called the incomplete reconstruction of the dynamics [30], gives us new insight into the way to capture the structural information of the dynamics. In the present paper, a new method based on the information captured by the incomplete reconstruction of the dynamics will be introduced to recover the time delay of the system.

Before that, we propose a simple procedure called component reordering, in order to show the local clustering phenomenon of the orbits of the chaotic system in the incomplete two-dimensional reconstructed phase-space. We find that information captured by the incomplete two-dimensional reconstructed phase-space not only is related to the time delay present in the system, but also can be transferred to the reordered component by this procedure. Then, in order to recover the time delay of the system, the segmented mean variance (SMV) is derived from the reordered component. The proposed SMV will show pronounced maximum when the embedding delay τ of the incomplete reconstruction is close to the time delay τ_0 of the time-delay system. Numerical data generated from a time-delay system based on the Mackey-Glass system operating in the chaotic regime are used to illustrate the validity of the proposed SMV. A series of successful time delay identifications demonstrate that the structural information captured by the incomplete two-dimensional reconstructed phase-space is enough to recover

*ganlu@uestc.edu.cn

the time delay of the system from the scalar time series analysis. The proposed method is easy to operate with a small amount of computation, and it also has a good robustness against additive observational noise. Most importantly, it can recover the time delay of the system even though the amount of data is relatively small and the feedback strength of the system is weak.

The present paper is structured as follows. In Sec. II, the delay-coordinate reconstruction is briefly introduced first, then the component reordering procedure is developed; after that the local clustering phenomenon of the chaotic time series is described by utilizing the numerical data generated by the Hénon map and the Mackey-Glass equation; finally, the segmented mean variance (SMV) is proposed to recover the time delay present in the scalar time series. In Sec. III, the feasibility and reliability of the proposed SMV are first checked by utilizing the numerical time series generated by the time-delay systems based on the well-known Mackey-Glass equation, then the effects of the additive observational noise, data length, small time delay, and feedback strength on the proposed SMV are tested. At last, a simple comparison of time consumptions between the C_{JS}, the ACF, and the proposed SMV for different data lengths is obtained. In Sec. IV, some brief conclusions are given.

II. THE LOCAL CLUSTERING PHENOMENON AND THE SEGMENTED MEAN VARIANCE

A. Delay-coordinate reconstruction

The reconstruction of the phase-space plays a very important role in chaotic time series analysis since the structure of phase-space is very helpful, and it is also a fundamental tool for nonlinear time series analysis. The widely used method to reconstruct the phase-space of the dynamics of the chaotic system is the delay-coordinate reconstruction proposed by Takens *et al.* [31,32]. Specifically, let $\{x_n\}_{n=1}^{N}$ be a scalar time series of length N, and let m be the embedding dimension and τ be the embedding delay. Then the m-dimensional reconstructed phase-space of this time series consists of the following vectors:

$$\mathbf{V}_n = [x_n, x_{n+\tau}, x_{n+2\tau}, \ldots x_{n+(m-1)\tau}], \tag{1}$$

where $n = 1,2,\ldots,N-(m-1)\tau$. Though the embedding dimension m and the embedding delay τ are very essential for the delay-coordinate reconstruction, good value for them are not easy to be estimated due to the data length, noise, nonstationarity, algorithm parameters, and the like [33] in practice. About estimating good value for the embedding dimension m and the embedding delay τ, please see Refs. [33,34] and the references therein.

Actually, the full structure of the dynamics of the chaotic system is not always necessary, and a partial knowledge of the dynamics is helpful for data analysis purpose sometimes. In Ref. [30], it was found that forecast models that utilized the incomplete reconstruction of the dynamics can obtain accurate predictions of the future course of the dynamics. And in Ref. [35], the approach based on the incomplete two-dimensional reconstructed phase-space was successfully applied to distinguish between noise and chaotic signals. In the present paper, we will show that we can recover the time delay τ_0 present in the chaotic time series based on the information captured by the incomplete two-dimensional reconstructed phase-space.

The incomplete reconstruction means that the choice of embedding dimension m does not satisfy the conditions of Takens' theorem [31]. It should be noted that the full reconstructed phase-space can be obtained for low-dimensional chaotic time series if the embedding dimension m and the embedding delay τ satisfy the conditions of the Takens' theorem. The value of embedding dimension m for the reconstruction of phase-space of the dynamics is set to 2 in this paper, which means that the phase-space is incompletely reconstructed for high-dimensional chaotic systems.

B. The component reordering procedure

When the embedding dimension m is fixed, another parameter, the embedding delay τ, will have a major impact on the incomplete two-dimensional reconstruction of the phase-space. Obviously, the structure captured by the incomplete two-dimensional reconstructed phase-space is quite diverse if different embedding delay τ is used. Based on it, we will propose an efficient method to recover the time delay of the system from chaotic time series in the present paper. Before that, a simple but important procedure, which is called the component reordering, will be presented first. To the best of our knowledge, this is the first time that the component reordering procedure is introduced.

Let $\{x_n\}_{n=1}^{N}$ be a scalar time series of length N and τ be the embedding delay of the incomplete two-dimensional reconstruction. Then the *component reordering procedure* is implemented as follows:

(1) Obtaining the first component $\{x_n^f\}_{n=1}^{N-\tau} = \{x_n\}_{n=1}^{N-\tau}$ and the second component $\{x_n^s\}_{n=1}^{N-\tau} = \{x_n\}_{n=\tau+1}^{N}$ of the incomplete two-dimensional reconstructed phase-space,

(2) Sorting the first component $\{x_n^f\}_{n=1}^{N-\tau}$ with ascending order, let n_{new} be the new subscript after sorting. Let $z_n = x_{n_{\mathrm{new}}}^s$, then $\{z_n\}_{n=1}^{N-\tau}$ will be a new time series.

We will call $\{z_n\}_{n=1}^{N-\tau}$ the *reordered component* for the reason that it is actually a reordered version of the second component $\{x_n^s\}_{n=1}^{N-\tau}$. The *component reordering procedure* can be better described with a simple example; suppose that we have a short time series $\{x_n\}_{n=1}^{7} = \{1.1,7.1,6.1,2.3,4.5,5.3,8.2\}$ and we set the embedding delay $\tau = 1$, then we can obtain the first component $\{x_n^f\}_{n=1}^{6} = \{1.1,7.1,6.1,2.3,4.5,5.3\}$ and the second component $\{x_n^s\}_{n=1}^{6} = \{7.1,6.1,2.3,4.5,5.3,8.2\}$ of the original time series. By sorting the first component with ascending order, we can get the new subscript $n_{\mathrm{new}} = 1,4,5,6,3,2$; correspondingly, we can obtain the reordered component $\{z_n\}_{n=1}^{6} = \{7.1,4.5,5.3,8.2,2.3,6.1\}$ by utilizing the new subscript.

C. The local clustering phenomenon

After obtaining the reordered component, we have observed that there is a special relationship between the incomplete two-dimensional reconstructed phase-space and its corresponding reordered component. To better describe this relationship, we

first consider the following equation for Hénon map [36]:

$$x_{n+1} = y_n + 1 - ax_n^2$$
$$y_{n+1} = bx_n. \tag{2}$$

The typical values $a = 1.4$ and $b = 0.3$ are chosen to produce a deterministic chaotic time series. The Hénon map is used here because it has a low-dimensional attractor and the relationship can be described more clearly. It should be stressed that the discrete chaotic map itself is not the focus of this paper.

In Fig. 1 we plot the phase-space, the incomplete two-dimensional reconstructed phase-space, and the corresponding reordered component of Hénon map. From Figs. 1(a) and 1(b) it can be observed that the incomplete reconstruction recovers all of the structure of this system since the delay-coordinate reconstruction of this data with $m = 2$ and $\tau = 1$ is indeed the actual map just scaled. However, the main point of Fig. 1 concerns the similarity between Figs. 1(b) and 1(c), which represent the incomplete two-dimensional reconstructed phase-space and its corresponding reordered component, respectively. By studying the reordered component, we also find that the values of the adjacent data of the reordered component are very close when the similarity is well displayed, as if these data are locally clustered by the component reordering procedure. We will call that the *local clustering phenomenon* in this paper. In other words, the local clustering phenomenon contains two meanings: the similarity between the incomplete two-dimensional reconstructed phase-space and its corresponding reordered component and the clustering of the data in the reordered component.

The similarity between the incomplete two-dimensional reconstructed phase-space and the corresponding reordered component of Hénon map is perfectly displayed since the phase-space of Hénon map is well reconstructed. It should be noted that the similarity exists when other values of τ are used and disappears when τ is large enough for Hénon map, which we do not plot in this paper. It seems that the local clustering phenomenon depends on the embedding delay τ of the incomplete reconstruction. However, for a time-delay chaotic system, which can produce highly complex dynamics—the incomplete two-dimensional reconstructed phase-space cannot capture the full structure of the dynamics—does the local clustering phenomenon still exist? To answer this question, we consider the well-known Mackey-Glass equation [3]:

$$\frac{dx}{dt} = \frac{\alpha x(t-\tau_0)}{1 + x^{\gamma}(t-\tau_0)} - x, \quad \alpha, \gamma > 0, \tag{3}$$

where τ_0 is the time delay feedback, α is the feedback strength, γ is the degree of nonlinearity, and t is a dimensionless time. The typical values τ_0=60, $\alpha = 2$, and $\gamma = 10$ are chosen to make the system operate in the chaotic regime. For the purpose of obtaining the Mackey-Glass time series, the fourth-order Runge-Kutta method [37] is used to numerically integrate the equation from Eq. (3), and the integration step and sampling step are $\Delta t = 0.01$ and $\delta t = 0.2$ time units, respectively. The time delay present in the Mackey-Glass time series is 300 ($\tau_0/\delta t = 300$) under these parameters.

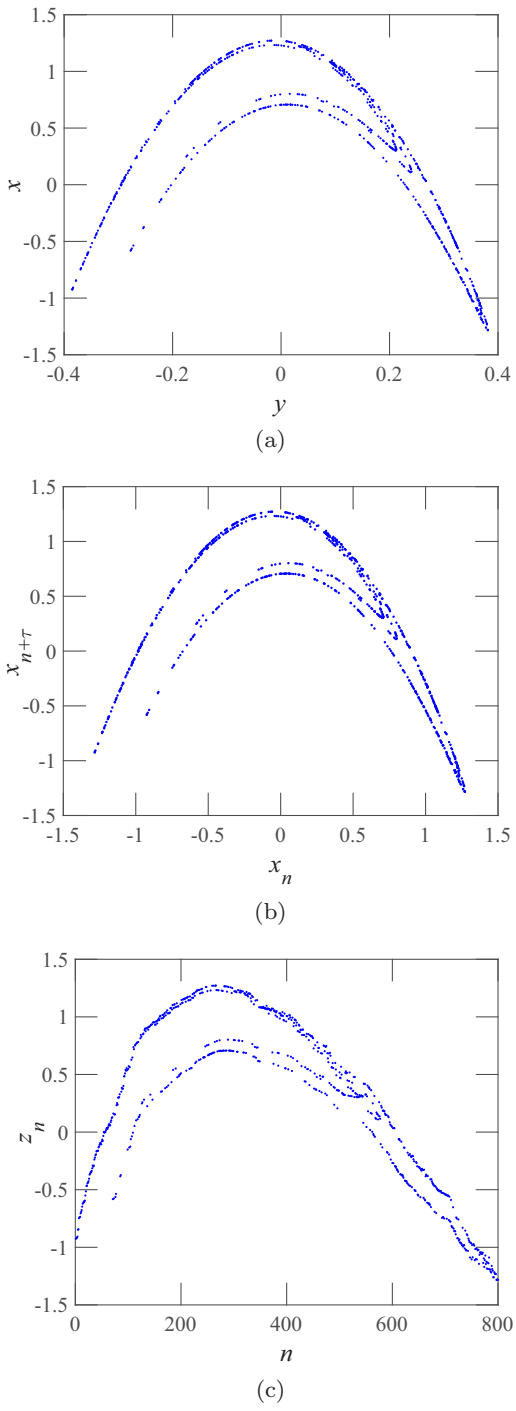


FIG. 1. (a) Phase-space of Hénon map; (b) incomplete two-dimensional reconstructed phase-space of Hénon map (embedding dimension $m = 2$ and embedding delay $\tau = 1$); (c) Reordered component with $\tau = 1$. $N = 800$ data points are used.

In Fig. 2 we plot the incomplete two-dimensional reconstructed phase-space with embedding delay $\tau = 300$ and its corresponding reordered component of the Mackey-Glass time series. Besides, we also plot the reordered component with embedding delay $\tau = 600$ and $\tau = 133$ in Fig. 2, while their corresponding incomplete reconstructed phase-space are not plotted for the sake of brevity. From Figs. 2(a) and 2(b) it can be observed that the local clustering phenomenon still

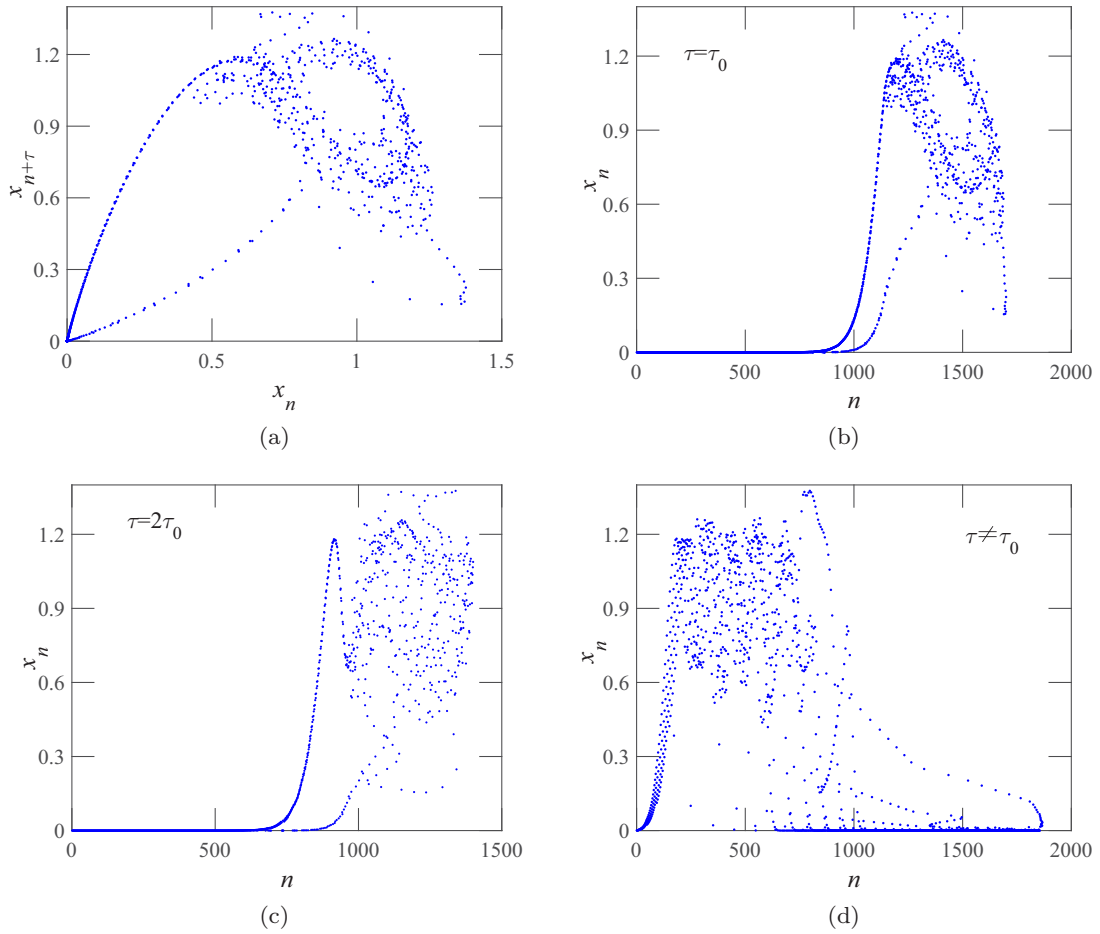


FIG. 2. (a) Incomplete two-dimensional reconstructed phase-space of Mackey-Glass time series (embedding dimension $m = 2$ and embedding delay $\tau = 300$) and reordered component with embedding delay, (b) $\tau = 300$, (c) $\tau = 600$, (d) $\tau = 133$.

exists, though it is poorly displayed (compared with the Hénon scenario).

The local clustering phenomenon illustrates that the information captured by the incomplete two-dimensional reconstructed phase-space is transferred to the corresponding reordered component by the procedure of component reordering. Hence, we can extract some important features of the dynamics of the time-delay chaotic system from the corresponding reordered component. It is also more convenient to study the reordered component rather than the incomplete reconstructed phase-space since the dimension of the phase-space is reduced.

As shown in Fig. 2, the reordered component has three different forms when different embedding delays τ of the incomplete reconstruction are considered, and these forms are related to the time delay τ_0 of the system. The relationships between the embedding delay τ of the incomplete reconstruction and the time delay τ_0 present in the system are described as follows:

$$\begin{aligned} &\text{I } \tau = \tau_0, \\ &\text{II } \tau = n\tau_0, n = 2,3,\ldots, \\ &\text{III } \tau \neq n\tau_0, n = 1,2,3,\ldots. \end{aligned} \tag{4}$$

In case I, i.e., the embedding delay τ of the incomplete reconstruction is equal to the time delay τ_0 of the system, the amount of information captured by the incomplete reconstructed phase-space is the most, as shown in Fig. 2(a). Accordingly, the information contained in the reordered component is the most and the local clustering phenomenon is well displayed, as shown in Fig. 2(b). In case II, the reordered component also shows some fundamental structure of the dynamics of the system since the structural information captured by the incomplete two-dimensional reconstructed phase-space is also considerable, as shown in Fig. 2(c). In case III, however, the reordered component acts like random time series for the reason that the incomplete two-dimensional reconstructed phase-space cannot capture any structural information of the dynamics of the system, as shown in Fig. 2(d). It should be noted that there is no local clustering phenomenon at this time. From Figs. 2 and 1 we also observe that the local clustering phenomenon of the Mackey-Glass time series is not demonstrated as well as that of the Hénon map, the reason is that the incomplete two-dimensional reconstructed phase-space cannot capture the full dynamics of the Mackey-Glass system.

D. The segmented mean variance

As described above, for a time-delay chaotic system, the property of the reordered component is closely related to the relationship between the embedding delay τ of the incomplete

reconstruction and the time delay τ_0 of the system. We find that the local clustering phenomenon is well displayed when the embedding delay τ of the incomplete reconstruction is equal to the time delay τ_0 present in the system since the structural information captured by the incomplete two-dimensional reconstructed phase-space is the most in this situation. Inspired by the local clustering phenomenon, we present a simple approach to recover the time delay τ of the system from the reordered component in this paper. This method, which will be called the segmented mean variance (SMV), is based on the calculation of the mean and variance of the reordered component. The calculation procedure of the SMV is described as follows:

(1) Dividing the reordered component $\{z_n\}_{n=1}^{N-\tau}$ into L groups: $Z_l = \{z_{(l-1)K+1}, z_{(l-1)K+2}, \ldots, z_{lK}\}, l = 1,2,\ldots,L$, with $K = \lfloor N/L \rfloor$ being the amount of data of each group and $\lfloor P \rfloor$ denoting an integer less than or equal to P;

(2) Calculating the mean $\hat{\mu}_l$ and the variance $\hat{\sigma}_l^2$ of each group $Z_l, l = 1,2,\ldots,L$,

$$\hat{\mu}_l = \frac{1}{K}\sum_{k=1}^{K} z_{(l-1)K+k}, \qquad \hat{\sigma}_l^2 = \frac{1}{K-1}\sum_{k=1}^{K}\left(z_{(l-1)K+k} - \hat{\mu}_l\right)^2; \tag{5}$$

(3) Computing the mean of $\hat{\mu}_l$ and $\hat{\sigma}_l^2, l = 1,2,\ldots,L$,

$$\hat{\mu} = \frac{1}{L}\sum_{l=1}^{L}\hat{\mu}_l, \hat{\sigma}^2 = \frac{1}{L}\sum_{l=1}^{L}\hat{\sigma}_l^2; \tag{6}$$

(4) Calculating the variance of $\hat{\mu}_l, l = 1,2,\ldots,L$,

$$\hat{\sigma}_0^2 = \frac{1}{L-1}\sum_{l=1}^{L}(\hat{\mu}_l - \hat{\mu})^2; \tag{7}$$

(5) Then the SMV is obtained as follows:

$$\mathrm{SMV} = \frac{K}{\hat{\sigma}^2}\hat{\sigma}_0^2. \tag{8}$$

According to the relationships described in Eq. (4), the SMV will show three different types of values. In the first case, the values of data of each group $Z_l, l = 1,2,\ldots,L$ are almost the same because of the well presented local clustering phenomenon, therefore, the mean of segmented variances $\hat{\sigma}^2$ is of small value. Meanwhile, the values of data of different groups are different, then the segmented means $\hat{\sigma}_0^2$ is of large value. A small value of $\hat{\sigma}^2$ and a large value of $\hat{\sigma}_0^2$ will lead to a large SMV in this situation. It should be noted that the parameter K in Eq. (8) is just a multiplication factor, which makes the SMV follow F distribution if the time series $\{x_n\}_{n=1}^{N}$ is a Gaussian white noise [35].

In the second case, though the local clustering phenomenon is not presented as well as that of the first case, the incomplete two-dimensional reconstructed phase-space also can capture some fundamental information of the dynamics of the system. Though the value of the SMV is large enough, it is still much smaller than that of the first case. In the last case, there is no helpful information contained in the reordered component, and the local clustering phenomenon disappears. The reordered component acts like a random time series. Thus, the mean of segmented variances $\hat{\sigma}^2$ is of large value and the variance of segmented means $\hat{\sigma}_0^2$ is of small value, so the value of the proposed SMV is relatively small. Above all, we can recover the time delay present in the system according to the values of the SMV.

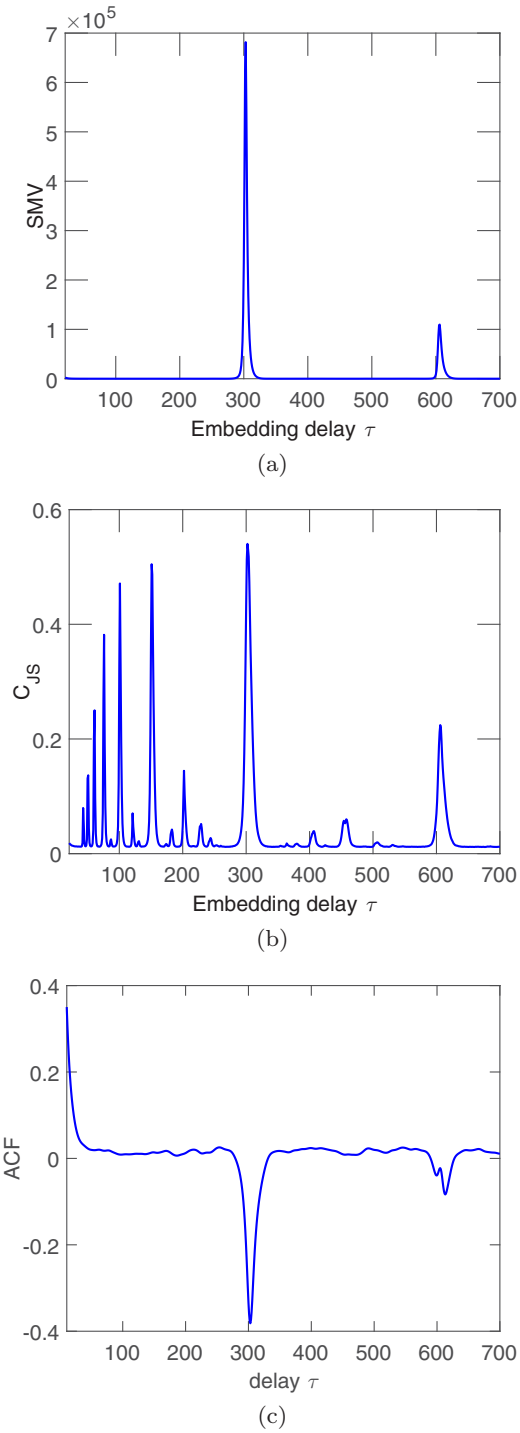


FIG. 3. (a) SMV as a function of the embedding delay τ for number of data segments $L = 5$. (b) The C_{JS} as a function of the embedding delay τ for embedding dimension $m = 8$. (c) The ACF of the Mackey-Glass time series. $N = 10^6$ data points are used.

III. NUMERICAL RESULTS AND DISCUSSIONS

As stated in Sec. II, the values of the proposed SMV are relatively different according to the relationship between the embedding delay τ of the incomplete reconstruction and the time delay τ_0 of the system, which means that we can identify the time delay present in the scalar time series according to the values of the proposed SMV. In this section, some experiments will be given in order to check the effectiveness and reliability of the proposed SMV. Besides, for the purpose of comparison, the permutation statistical complexity (C_{JS}) [27] and the ACF are used as the gold standards. The ACF is a conventional and widely used method to determine the time delay present in the time series, and it will be evaluated by exploiting the autocorr.m file in MATALB of release 2016*a*. The C_{JS} is a relatively new method proposed to identify the time delay of the system and its performance is relatively good [27]. Note that the calculation of C_{JS} is related to the phase-space reconstruction [38]; it also needs two parameters, namely the embedding delay τ and embedding dimension m when evaluating the C_{JS}. Generally speaking, to obtain a reliable statistics when evaluating the C_{JS}, the length N of the time series and the embedding dimension m should satisfy the condition $N \gg m!$ [39], or they should satisfy the condition $N \geqslant 5m!$ [40] at least. In the following experiments, the embedding dimensions for C_{JS} are chosen according to the latter condition. About this approach, please see Refs. [38] and [27] for details. It should be pointed out that the starting points of the C_{JS} and proposed SMV are totally different, though they both are related to the phase-space reconstruction.

Numerical data used in the following simulations are generated by the time-delay systems based on the Mackey-Glass equation. First of all, the effectiveness of the SMV will be tested by calculating the SMV as a function of the embedding delay τ of the incomplete reconstruction, and the effect of the number of data segments L on the proposed SMV will be discussed. Then, the effect of additive observational noise, data length, small value of time delay, and feedback strength on the proposed SMV will be checked. Finally, the time complexity of the proposed SMV is obtained for different data lengths.

A. The effectiveness of the SMV

In order to check the effectiveness of the proposed method, we calculate the SMV as a function of the embedding delay τ of the incomplete reconstruction for the Mackey-Glass time series generated by Eq. (3). The results are shown in Fig. 3(a). It can be clearly observed that the proposed SMV has well-defined and sharp maxima when the embedding delay τ of the incomplete reconstruction is close to the time delay τ_0 of the system, i.e., for τ near $300(\tau_0/\delta t = 300)$. In this situation, the structural information of the system captured by the incomplete two-dimensional reconstructed phase-space is the most and the local clustering phenomenon is well presented. Consequently, the value of the SMV is the largest. Moreover, the proposed SMV also shows clear peak when the embedding delay τ of the incomplete reconstruction is approximately double the time delay τ_0 of the system, i.e., for τ near $600(\tau_0/\delta t = 600)$, while the value of the proposed SMV is much smaller in the case of other embedding delays. Thus, the results shown in Fig. 3(a) are perfectly consistent with the discussion in Sec. II.

From Fig. 3(a) it can also be observed that there is a light time-delay overestimation. This time-delay overestimation can be attributed to the internal response time or inertia of the Mackey-Glass system [27]. It is difficult to estimate the inertia accurately, and most of the methods proposed to recover the time delay present in the recorded time series [41,42] are affected by the inertia. We also calculate the C_{JS} and the ACF for the same Mackey-Glass time series because they are also affected by the inertia, and for comparison purpose the results are shown in Figs. 3(b) and 3(c). It can be observed from Fig. 3(b) that there are a lot of spurious peaks in the C_{JS}, while these spurious peaks do not appear in our approach and the ACF.

The number of data segments L is important for calculation of the proposed SMV. It should be pointed out that L should satisfy the following condition:

$$2 \leqslant L \leqslant \lfloor (N - \tau_{\max})/2 \rfloor, \tag{9}$$

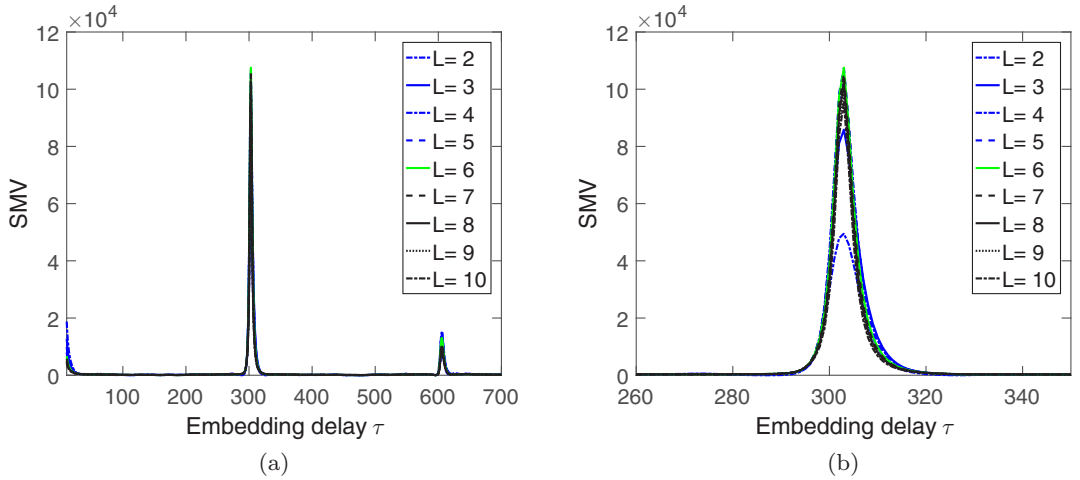


FIG. 4. (a) SMV as a function of embedding delay τ for the numbers of data segments $2 \leqslant L \leqslant 10$. (b) Enlargement near the time delay τ_0 of the system in order to observe more clearly the effect of the number of data segments L on the SMV. $N = 2 \times 10^5$ data points are used. The number of data segments associated with the different curves from top to bottom are $L = 6,5,7,4,8,9,10,3,2$.

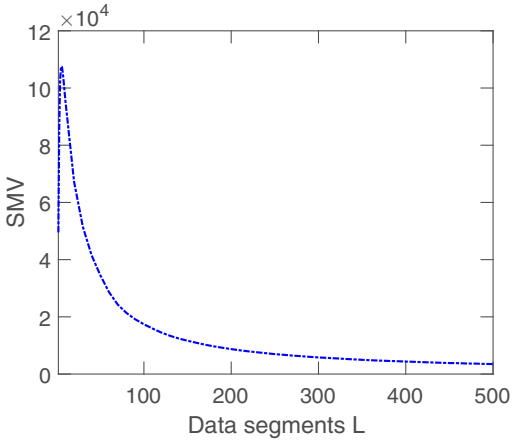


FIG. 5. SMV as a function of number of data segments L for embedding delay $\tau = 303$. The maximum value of the SMV occurs when $L = 6$. $N = 2 \times 10^5$ data points.

where $\tau_{\max}$ is the largest embedding delay of the incomplete reconstruction one chooses and N is the data length. From Eq. (7) we can find that $L \neq 1$ is obvious, and if $L > \lfloor (N - \tau_{\max})/2 \rfloor$, the amount of data in each group is $K = 1$; the calculation of SMV is meaningless in this situation. It is obvious that the values of the proposed SMV are different if we choose different numbers of data segments L. In order to check the effect of L on the SMV, in Fig. 4(a) we plot the SMV as a function of embedding delay τ of the incomplete reconstruction for different numbers of data segments L, and in Fig. 4(b) we plot the enlargement near the time delay τ_0 the system in order to observe more clearly. From Fig. 4, it can be observed that the values of the SMV corresponding to $L = 5$ and $L = 6$ are bigger than that of other data segments L. To better explain this, the proposed SMV as a function of the number of data segments L for $\tau = 303$ (because of the time delay overestimation) is plotted in Fig. 5. It can be observed that the SMV is maximized when $L = 6$ in this situation (the time series is generated from Eq. (3) and $N = 2 \times 10^5$ data points are used), while the corresponding number of data segments may not be $L = 6$ for other data lengths and time series derived from other time-delay chaotic systems. Nevertheless, the proposed SMV can always recover the time delay of the system as long as the number of data segments L one chooses satisfies the condition in Eq. (9). From now on, the number of data segments L will be fixed to 5 for the sake of uniformity in this article.

In practical applications, there are more than one time delay in the system sometimes. In order to test the performance of the SMV in this case, we consider the generalized Mackey-Glass equation [43] with two time delays:

$$\frac{dx}{dt} = \frac{1}{2}\sum_{k=1}^{2} \frac{\alpha x(t - \tau_{0,k})}{1 + x^{\gamma}(t - \tau_{0,k})} - x. \tag{10}$$

To obtain the numerical data, the same integration method and parameters ($\alpha = 2, \gamma = 10$) as in the single time delay case are used. In Fig. 6(a) we plot the SMV as a function of the embedding delay τ of the incomplete reconstruction in the case of a generalized Mackey-Glass system with two time delays ($\tau_{0,1} = 60$, $\tau_{0,2} = 96$). The proposed SMV shows obvious peaks when the embedding delays of the incomplete reconstruction are close to the time delays of the system, i.e., $\tau \sim 300\,(\tau_{0,1}/\delta t = 300)$ and $\tau \sim 480\,(\tau_{0,2}/\delta t = 480)$. Similar to the case of one time delay, there is also a slight time

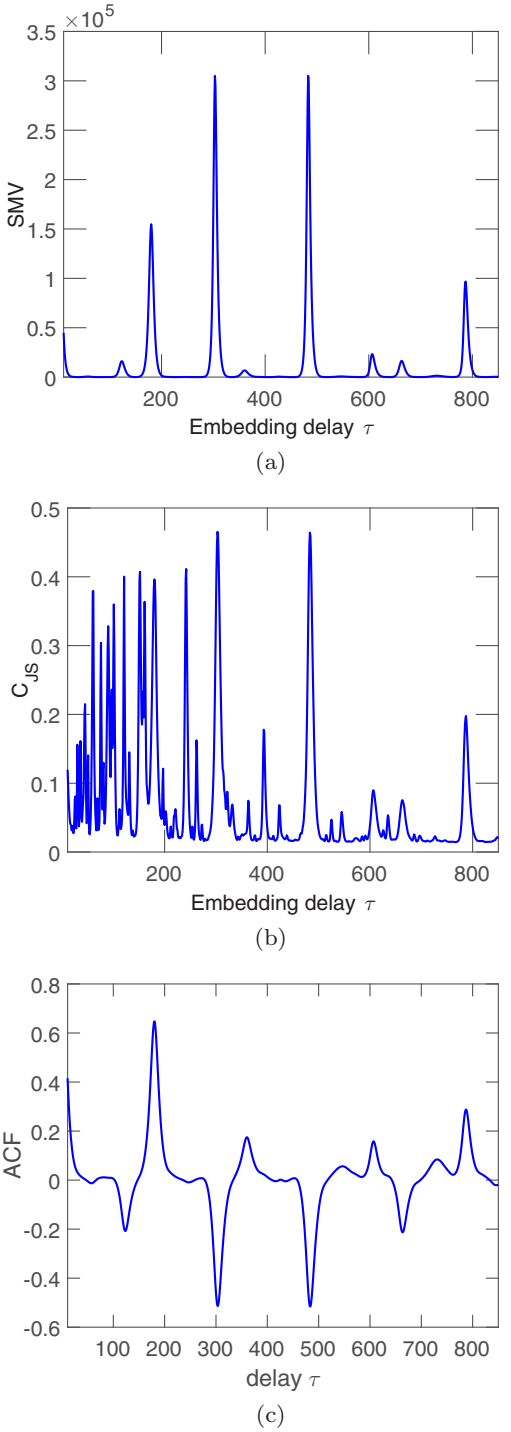


FIG. 6. (a) SMV and (b) the C_{JS} as a function of the embedding delay τ for a Mackey-Glass system with time delays $\tau_{0,1} = 60$ and $\tau_{0,2} = 96$. (c) The ACF obtained from the same Mackey-Glass time series. The number of data segments for the SMV is $L = 5$; the embedding dimension for C_{JS} is $m = 8$. $N = 1 \times 10^6$ data points are used.

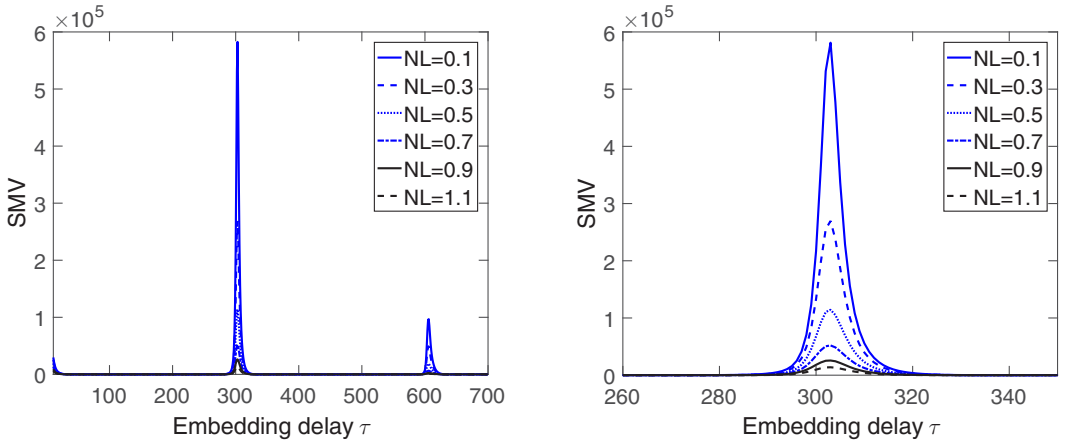


FIG. 7. (a) SMV as a function of the embedding delay τ of the incomplete reconstruction for different levels of the observational noise. The noise levels associated with the different curves ($NL = 0.1,0.3,0.5,0.7,0.9,1.1$) increases from top to bottom. The number of data segments is $L = 5$ and $N = 10^6$ data points are used. (b) Enlargement near the time delay τ_0 of the system in order to observe more clearly the effect of the NL on the SMV.

delay overestimation. From Fig. 6(a) it can also be seen that the SMV shows peaks when the embedding delay is close to $\tau = 180(480 - 300 = 180)$ and $\tau = 780(480 + 300 = 780)$, but less pronounced. The occurrence of these peaks is similar to the intermodulation in inverters; their presence do not affect the estimation of the time delays of the system. As a comparison, the C_{JS} and the ACF are also calculated from the same time series, the results are shown in Figs. 6(b) and 6(c). From Fig. 6(b) we can also observe the intermodulation phenomena, besides, a lot of spurious peaks appeared on the left side of the time delays of the system, which has a bad effect on the identification of the time delays of the system. For the ACF, the intermodulation phenomenon also exists and it has a bad effect on the recovery of the time delays present in the system since the amplitude of the peak associated to $\tau = 180$ is too large, as can be seen from Fig. 6(c). Thus, it can be concluded that our method still works well in the case of two time delays, and its performance is better than that of the C_{JS} and the ACF.

B. The effect of an additive observational noise

The next goal of this paper is to analyze the effect of an additive observational noise on the proposed method. It is meaningful since the experimental time series is always contaminated by observational noise in practice. For this purpose, a Gaussian white noise with different noise levels is added to the numerical data generated by the Mackey-Glass system with one time delay. The noise level (NL) is defined as the ratio of the standard deviation of the noise and the standard deviation of the original signal.

In Fig. 7(a) we plot the SMV as a function of embedding delay τ of the incomplete reconstruction for different levels of observational noise, and in Fig. 7(b) we plot the enlargement near the time delay τ_0 of the system in order to observe more clearly. It can be observed that the proposed method is very robust in the presence of the observational noise. The computation of the SMV is based on a comparison of all the points in the first component of the incomplete two-dimensional reconstructed phase-space, and the relationship among the points of the first component can not be completely destroyed by the observational noise. Hence, the proposed SMV can recover the time delay of the system in the presence of observational noise.

C. The effect of data length

In practice, the data we need may not be always sufficient, and most of methods fail to recover the time delay τ_0 of the system due to the data shortage. Thus, it is of significance to study the effect of data length on the proposed SMV. The effect of data length on the C_{JS} and the ACF is also studied for the purpose of comparison. The embedding dimension $m = 6$ is considered when calculating the C_{JS} in the case of a small amount of data. The observational noise with $NL = 0.2$ is added to the Mackey-Glass time series. Figure 8 compares the results obtained for the SMV, the C_{JS}, and the ACF for two different data lengths $N = 50\,000$ and $N = 10\,000$. From Figs. 8(a) and 8(d) it can be seen that our approach is able to recover the time delay τ_0 of the system successfully in all cases. The amplitude of the peak associated with the time delay τ_0 of the system just becomes lower when the amount of data is of small value. The C_{JS} can easily reveal the correct time delay τ_0 in the first case, as shown in Fig. 8(b), whereas it fails to recover the time delay of the system when $N = 10\,000$, as shown in Fig. 8(e). It should be noted that other embedding dimensions ($m = 3, 4, 5$) for C_{JS} cannot recover the time delay τ_0 of the system in the second case. Actually, the C_{JS} cannot recover the time delay τ_0 any more when the amount of data is relatively small, i.e., for $N < 10\,000$. As for the ACF, it can identify the time delay of the system in the case of a large amount of data, as shown in Fig. 8(c). Its performance becomes worse when the data length is of relatively small value, as shown in Fig. 8(f). It should be pointed out that the correct position of the time delay is the trough [the red point shown in Figs. 8(c) and 8(f)], which is surrounded by two large peaks around $\tau = 300$. These two large peaks will affect the recovery of the

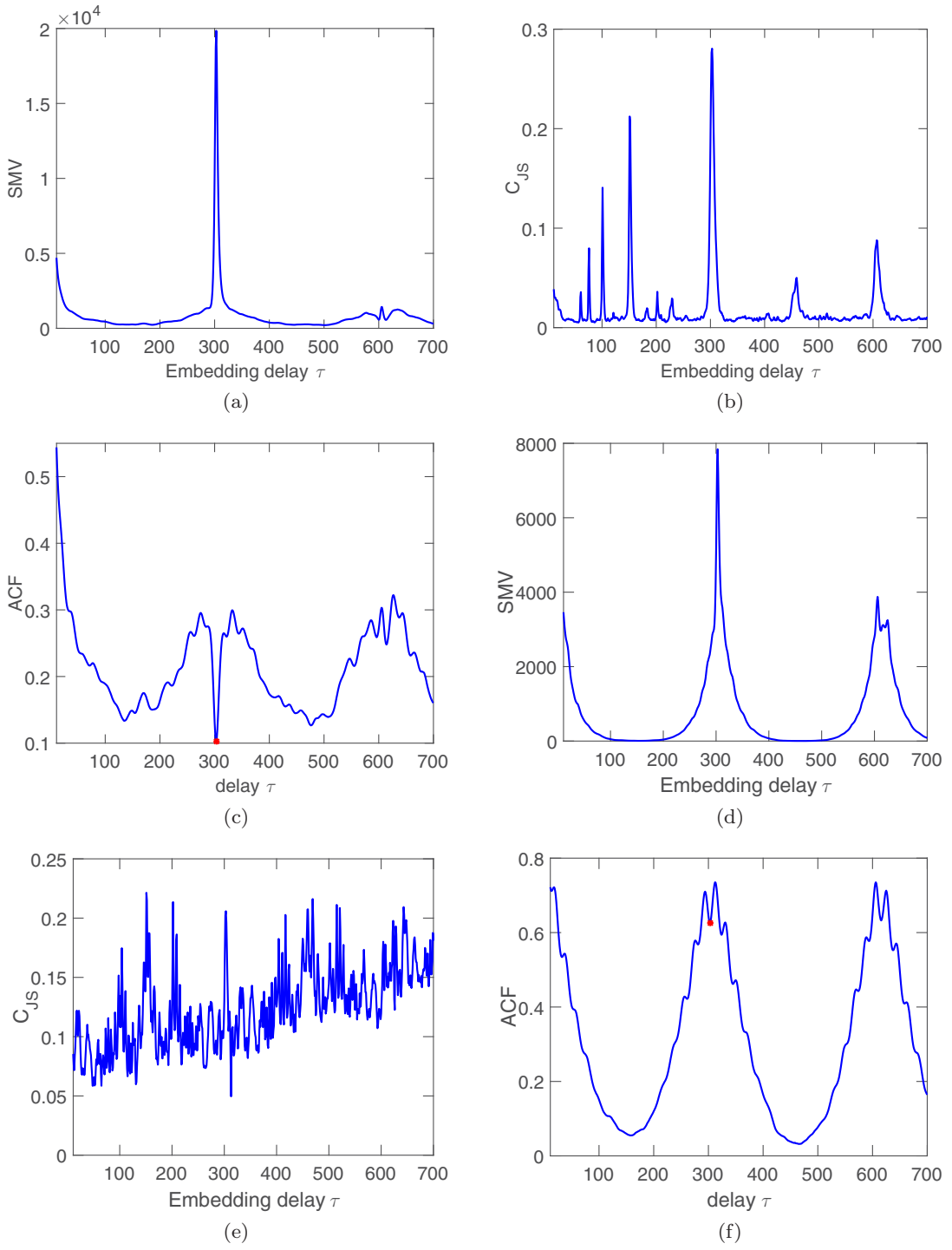


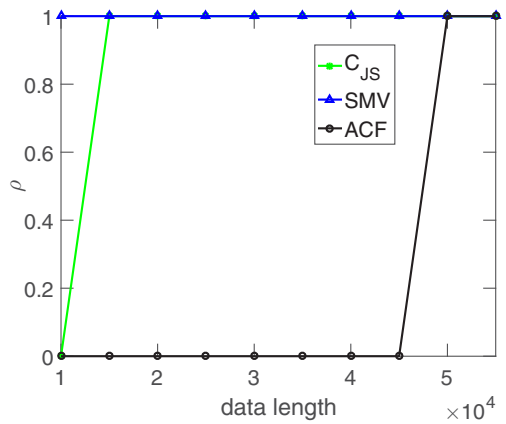
FIG. 8. Comparison between the SMV, the C_{JS}, and the ACF for the Mackey-Glass time series with different data lengths and $NL = 0.2$: (a), (b), and (c) $N = 50\,000$, and (d), (e), and (f) $N = 10\,000$. The red point in (c) and (f) corresponds to the value of ACF of the correct time delay $\tau = 303$ (because of the overestimation). The number of data segments L for the proposed SMV is equal to 5. The embedding dimensions for the C_{JS} in (b) and (e) are both 6.

true time delay present in the time series for the ACF. As a matter of fact, one cannot identify the correct time delay in this situation.

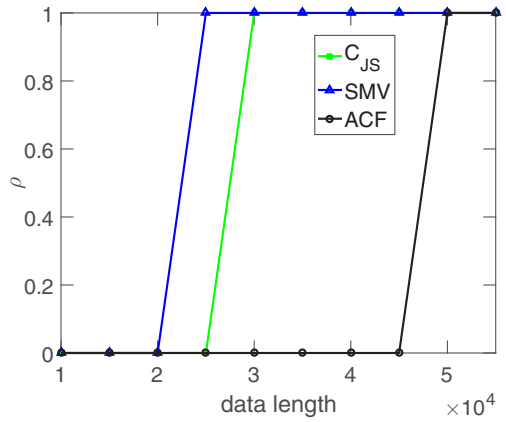
In order to better describe the effect of data length on these three methods and for the purpose of comparison, we will exploit the location of the embedding delay τ associated with the largest SMV to verify the accuracy of time delay identification in the present paper. In consideration of time delay overestimation, a vicinity $W(\tau_0)$ of the time delay τ_0 of the system is defined as [44]

$$W(\tau_0) = [\tau_0 - \varepsilon \times \tau_0, \tau_0 + \varepsilon \times \tau_0], \tag{11}$$

where ε is the mismatch coefficient, and ε is set to 4% in this paper. The time delay is still considered to be estimated



(a)



(b)

FIG. 9. Identification rate ρ as a function of data length for the proposed SMV, the C_{JS}, and the ACF: (a) $NL = 0$, (b) $NL = 1.2$. The number of data segments L for the SMV is equal to 5. The embedding dimension m for the C_{JS} is 6, and the simulation is carried out with 1000 Monte Carlo experiments and is averaged.

successfully once the embedding delay τ associated with the largest SMV locates in the vicinity $W(\tau_0)$. The identification rate ρ is defined as the ratio of the number of successful identifications and the number of trials. The simulation is carried out with 1000 Monte Carlo experiments and is averaged; the results are depicted in Fig. 9. Note that the results for ACF are obtained by searching the smallest value of the ACF, in other words, the estimation is still believed to be correct if the delay associated with the smallest value of the ACF locates in the vicinity $W(\tau_0)$.

Observe in Fig. 9(a) that the proposed SMV is able to recover the time delay of the system in all situations. For the C_{JS}, it performs well when data length is greater than 10 000, yet it fails to recover the time delay of the system when $N = 10\,000$. Besides, it can be seen from Fig. 9(a) that the identification rate of the ACF is zero when the data length is smaller than 50 000, which means that the ACF cannot identify the time delay of the system in the case of small amount of data. Hence, for large time delay, the proposed SMV performs better than the C_{JS} and the ACF when the size of data without noise pollution is small.

In Fig. 9(b) we compare the identification rates of the proposed method, the C_{JS}, and the ACF obtained from a Mackey-Glass time series contaminated by the observational noise with $NL = 1.2$. It can be seen in Fig. 9(b) that the proposed SMV still works well when the data length is larger than 20 000, yet it cannot reveal the time delay present in the time series when $N = 10\,000$ and $N = 15\,000$ due to the effect of higher level of the observational noise. The C_{JS} performs well in the case of large amount of data, i.e., for $N > 25\,000$, and it has no ability to identify the time delay of the system when $N = 25\,000$ while the SMV does. As for the ACF, it can recover the time delay when the data length is greater than 45 000, but it fails to determine the correct time delay of the system in other situations. From Fig. 9(b) it can be concluded that the performance of the SMV for large time delay is better than that of the C_{JS} and the ACF in the case of small amount of data and higher level of observational noise.

D. The effect of small value of time delay

The results obtained from above simulations show that the proposed method is an effective method to recover large time delay (e.g., $\tau_0 = 60$) of the system. In this subsection we will check the effect of small value of time delay on the proposed SMV. To this end, in Fig. 10 we plot the SMV as a function of the embedding delay τ of the incomplete reconstruction in the case of a Mackey-Glass system with small time delay ($\tau_0 = 20$, $\tau_0 = 10$, and $\tau_0 = 5$). It can be observed from Fig. 10 that the SMV shows sharp and well-defined peak when the embedding delay τ of the incomplete reconstruction is close to the time delay τ_0 of the system in all situations, which means that the proposed SMV also can recover small time delay present in the time series.

E. The effect of feedback strength

It is well known that the recovery of the time delay would be difficult if the feedback strength is small, it was also pointed out that the identification of the time delay of the system would be impossible when the optical feedback of the chaotic semiconductor laser was weak [42]. In order to check the performance of the proposed SMV in this severe time-delay identification scenario, numerical simulation of the data generated by Eq. (3) with the same parameters ($\tau_0 = 60$, $\gamma = 10$) but low feedback strength ($\alpha = 1.4$) is analyzed. Moreover, additive Gaussian white noise with $NL = 0.2$ is added. Meanwhile, the C_{JS} and the ACF are also used to identify the time delay of the system for the purpose of comparison. The results are shown in Fig. 11.

We can see from Fig. 11(a) that the proposed SMV shows pronounced maximum when the embedding delay τ of the reconstruction is close to the time delay of the system, which means that the proposed method can recover the time delay of the system even though the feedback strength is small. Moreover, we can also observe from Fig. 11(a) that the amplitude of the peak associated with $\tau = 600$ becomes larger than that of the peak associated with $\tau = 600$ in Fig. 8(a), this is due to the effect of the low feedback strength. As for the C_{JS}, it fails to recover the time delay in this situation, as shown in

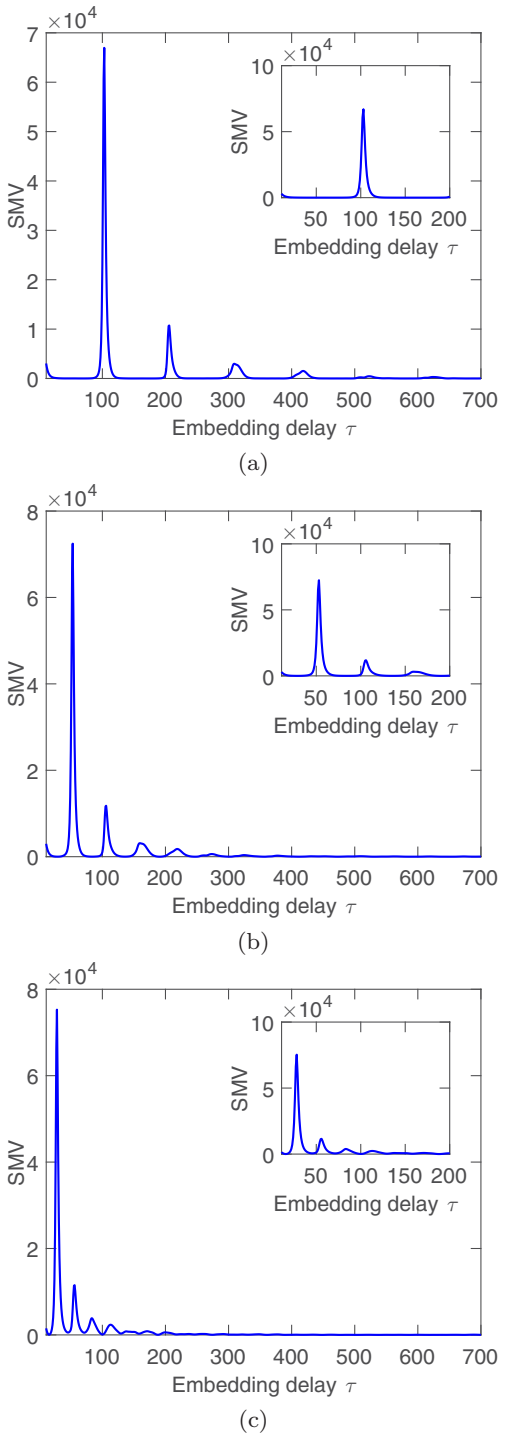


FIG. 10. The SMV as a function of the embedding delay τ for a Mackey-Glass system with (a) $\tau_0 = 20$, (b) $\tau_0 = 10$, and (c) $\tau_0 = 5$. The insets show the enlargement near the time delay τ_0 of the system in order to observe more clearly. The number of data segments for the SMV is $L = 5$; $N = 1 \times 10^5$ data points are used.

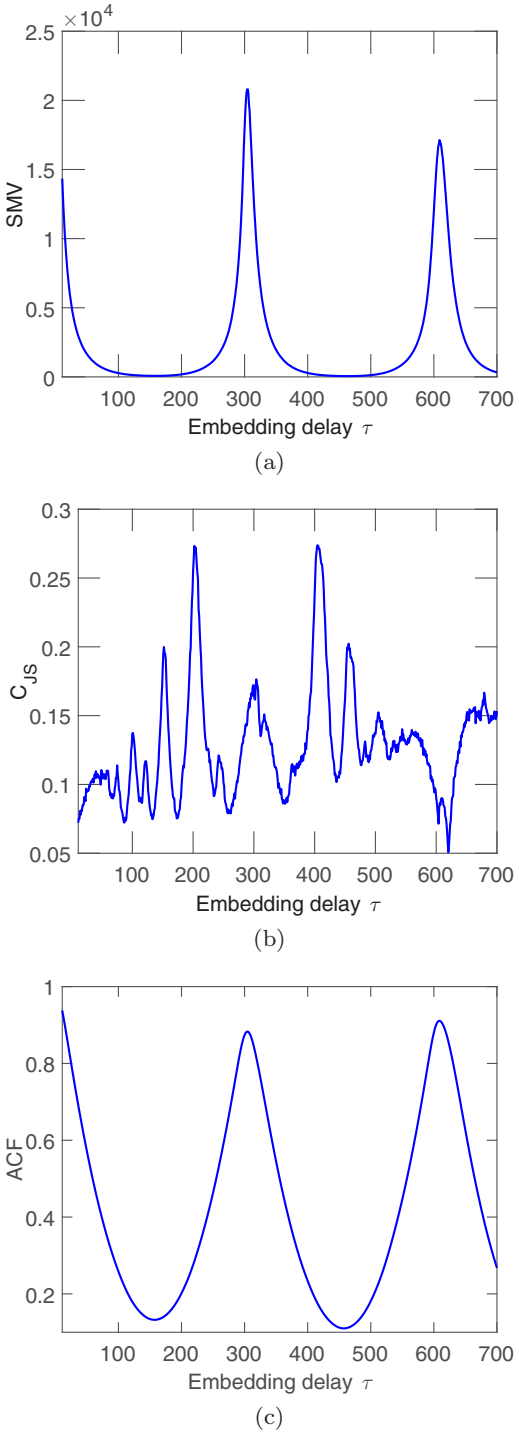


FIG. 11. Comparison between the SMV, the C_{JS}, and the ACF with low feedback strength ($\alpha = 1.4$) and noise level $NL = 0.2$: (a) the SMV, (b) the C_{JS}, and (c) the ACF. The number of data segments for the SMV is $L = 5$, the embedding dimension for C_{JS} is $m = 6$. $N = 15\,000$ data points are used.

Fig. 11(b). The ACF also shows obvious peak yet less sharper than that of the SMV around $\tau = 300$, which means that the ACF also can identify the time delay of the system in the case of small feedback strength, as can be seen in Fig. 11(c). It should be stressed that the original trough shown in Fig. 3(c) is turned into a peak in Fig. 11(c) because of the low feedback strength.

From Fig. 11 we can conclude that the performances of the proposed SMV and the ACF are much better than that of the C_{JS} in the case of weak feedback strength.

TABLE I. Time consumptions of the C_{JS}, the ACF, and the SMV with different data lengths (unit:s).

	$N = 600$	$N = 3600$	$N = 25\,200$	$N = 201\,600$
C_{JS}	1.5166×10^{-4}	5.0204×10^{-4}	0.0033	0.2471
ACF	1.1347×10^{-4}	3.7658×10^{-4}	0.0022	0.1583
SMV	7.5021×10^{-4}	2.5334×10^{-4}	0.0011	0.0705

F. The time complexity

The time complexity is also an important consideration in practical applications, especially when the amount of data is relatively large. From Sec. II we know that the calculation of the SMV is quite simple since one only need to compute the mean and variance of the data. In contrast, the computation of the C_{JS} is somewhat complicated when the embedding dimension m is of large value. In order to compare the time complexity of these three approaches, the average time consumptions via 200 Mente Carlo runs of the SMV, the ACF, and the C_{JS} for different data lengths are shown in Table I, the data lengths N we choose here is corresponding to the embedding dimension $m = 5, 6, 7, 8$ for C_{JS} subjected to $N = 5m!$. It should be pointed out that in each run 50 delays are evaluated for all three methods. We can observe from Table I that the time consumptions of the proposed SMV and the C_{JS} are the smallest and the largest, respectively. It is found that the time consumptions of the SMV are lower than that of the ACF and the C_{JS} in all situations, and the difference becomes obvious when the amount of data is of large value. The simulations are run on the computer with a 3.40 GHz Intel Core i7-2600K CPU and an 8.00 GB RAM. The release of MATLAB is 2016*a*.

IV. CONCLUSIONS

Time-delay chaotic systems are widely used in practice because of their high degree of nonlinearity and complex dynamics. In such systems, time delay always plays a vital role and can provide additional information about the relationship between different components. The recovery of time delay present in the time series is one of the key problems in the study of time-delay chaotic systems. However, time-delay identification is not an easy task due to the shortage of a prior knowledge, the small amount of data, and the effect of noise. In this paper, a computationally quick and conceptually simple approach is introduced to deal with this task. Before that, we propose an important procedure called the component reordering to show the local clustering phenomenon of the chaotic system based on the incomplete two-dimensional reconstruction of dynamics of the system. We find that the amount of information captured by the incomplete two-dimensional reconstructed phase-space is associated with the time delay present in the time-delay chaotic system. Furthermore, the structural information of the dynamics will be shown in a one-dimensional time series (the reordered component) because of the local clustering phenomenon. An application of this phenomenon is to distinguish between chaotic signal and Gaussian noise [35]. In this paper, we have shown that the local clustering phenomenon can be used to recover the time delay present in the time series. For this purpose, a statistic SMV is developed from the reordered component.

Numerical data generated by the time-delay systems based on the well-known Mackey-Glass equation are used to test the effectiveness and reliability of the proposed method. Numerical results show that the proposed SMV is robust to additive observational noise and is able to recover the time delay of the chaotic system with small feedback strength. What's more, it is found that the performance of the proposed method is also good in the case of small amount of data contaminated by a large amount of observational noise. The time complexity of the proposed SMV is also quite low.

ACKNOWLEDGMENTS

This work was supported by the National Natural Science Foundation of China (Grant No. U1530126).

[1] K. Ikeda, Opt. Commun. **30**, 257 (1979).
[2] R. Lang and K. Kobayashi, IEEE J. Quantum Electron. **16**, 347 (1980).
[3] M. C. Mackey and L. Glass, Science **197**, 287 (1977).
[4] A. Longtin, J. G. Milton, J. E. Bos, and M. C. Mackey, Phys. Rev. A **41**, 6992 (1990).
[5] I. R. Epstein, Int. Rev. Phys. Chem. **11**, 135 (1992).
[6] M. R. Roussel, J. Phys. Chem. **100**, 8323 (1996).
[7] E. Tziperman, L. Stone, M. A. Cane, and H. Jarosh, Science **264**, 72 (1994).
[8] A. J. Clarke, X. Liu, and S. Van Gorder, J. Clim. **11**, 987 (1998).
[9] I. Fischer, O. Hess, W. Elsäßer, and E. Göbel, Phys. Rev. Lett. **73**, 2188 (1994).
[10] F.-Y. Lin and J.-M. Liu, IEEE J. Quantum Electron. **40**, 682 (2004).
[11] F.-Y. Lin and J.-M. Liu, IEEE J. Sel. Top. Quantum Electron. **10**, 991 (2004).
[12] R. Vicente, J. Dauden, P. Colet, and R. Toral, IEEE J. Quantum Electron. **41**, 541 (2005).
[13] M. Peil, I. Fischer, W. Elsäßer, S. Bakić, N. Damaschke, C. Tropea, S. Stry, and J. Sacher, Appl. Phys. Lett. **89**, 091106 (2006).
[14] T. E. Murphy and R. Roy, Nature Photonics **2**, 714 (2008).
[15] R. Hegger, M. J. Bünner, H. Kantz, and A. Giaquinta, Phys. Rev. Lett. **81**, 558 (1998).
[16] C. Zhou and C.-H. Lai, Phys. Rev. E **60**, 320 (1999).
[17] V. Udaltsov, L. Larger, J. Goedgebuer, A. Locquet, and D. Citrin, J. Opt. Technol. **72**, 373 (2005).
[18] M. C. Soriano, P. Colet, and C. R. Mirasso, IEEE Photonics Technol. Lett. **21**, 426 (2009).
[19] V. S. Udaltsov, J.-P. Goedgebuer, L. Larger, J.-B. Cuenot, P. Levy, and W. T. Rhodes, Phys. Lett. A **308**, 54 (2003).
[20] R. M. Nguimdo, M. C. Soriano, and P. Colet, Opt. Lett. **36**, 4332 (2011).

[21] M. J. Bünner, A. Kittel, J. Parisi, I. Fischer, and W. Elsäßer, Europhys. Lett. **42**, 353 (1998).
[22] B. P. Bezruchko, A. S. Karavaev, V. I. Ponomarenko, and M. D. Prokhorov, Phys. Rev. E **64**, 056216 (2001).
[23] M. D. Prokhorov, V. Ponomarenko, A. Karavaev, and B. Bezruchko, Physica D **203**, 209 (2005).
[24] Y.-C. Tian and F. Gao, Physica D **108**, 113 (1997).
[25] R. K. Azad, J. S. Rao, and R. Ramaswamy, Chaos Solitions Fractals **14**, 633 (2002).
[26] M. Siefert, Phys. Rev. E **76**, 026215 (2007).
[27] L. Zunino, M. C. Soriano, I. Fischer, O. A. Rosso, and C. R. Mirasso, Phys. Rev. E **82**, 046212 (2010).
[28] M. C. Soriano, L. Zunino, O. A. Rosso, I. Fischer, and C. R. Mirasso, IEEE J. Quantum Electron. **47**, 252 (2011).
[29] J. D. Farmer, Physica D **4**, 366 (1982).
[30] J. Garland and E. Bradley, Chaos **25**, 123108 (2015).
[31] F. Takens, *Detecting Strange Attractors in Turbulence* (Springer, Berlin, 1981), pp. 366–381.
[32] N. H. Packard, J. P. Crutchfield, J. D. Farmer, and R. S. Shaw, Phys. Rev. Lett. **45**, 712 (1980).
[33] E. Bradley and H. Kantz, Chaos **25**, 097610 (2015).
[34] J. Garland, R. G. James, and E. Bradley, Phys. Rev. E **93**, 022221 (2016).
[35] S.-L. Zhu and L. Gan, Acta Phys. Sin. **65**, 070502 (2016).
[36] M. Hénon, Commun. Math. Phys. **50**, 69 (1976).
[37] T. C. Gard, *Introduction to Stochastic Differential Equations* (Marcel Dekker, New York, 1988).
[38] O. A. Rosso, H. A. Larrondo, M. T. Martin, A. Plastino, and M. A. Fuentes, Phys. Rev. Lett. **99**, 154102 (2007).
[39] M. Staniek and K. Lehnertz, Int. J. Bifurcat. Chaos **17**, 3729 (2007).
[40] M. Matilla-García and M. R. Marín, J. Econometrics **144**, 139 (2008).
[41] D. Rontani, A. Locquet, M. Sciamanna, and D. Citrin, Opt. Lett. **32**, 2960 (2007).
[42] D. Rontani, A. Locquet, M. Sciamanna, D. S. Citrin, and S. Ortin, IEEE J. Quantum Electron. **45**, 879 (2009).
[43] H. Voss and J. Kurths, Phys. Lett. A **234**, 336 (1997).
[44] S. Xiang, W. Pan, B. Luo, L. Yan, X. Zou, N. Jiang, L. Yang, and H. Zhu, Opt. Commun. **284**, 5758 (2011).