Mean local autocovariance provides robust and versatile choice of delay for reconstruction using frequently sampled flowlike data

Jack Murdoch Moore,[1,2,*] David M. Walker,[2] and Gang Yan[1,3,4]

[1]*School of Physical Science and Engineering, Tongji University, Shanghai, 200092, P. R. China*

[2]*Department of Mathematics and Statistics, University of Western Australia, Crawley, Western Australia 6009, Australia*

[3]*Shanghai Institute of Intelligent Science and Technology, Tongji University, Shanghai, 200092, P. R. China*

[4]*Center for Excellence in Brain Science and Intelligence Technology, Chinese Academy of Sciences, Shanghai, 200031, P. R. China*

(Received 24 April 2019; revised manuscript received 5 November 2019; published 28 January 2020)

The first step in nonlinear time-series analysis can be selecting a delay for reconstruction. The most popular choices of this delay are the first zero of the autocovariance and the first minimum of the mutual information. An advantage of the first method arises from the robustness to noise of the autocovariance function, while an advantage of the second is that the first minimum of the mutual information provides a useful choice of delay for a wide range of nonlinear systems. We propose a method to choose a delay for frequently sampled flowlike data based on a mean local autocovariance function and compare its performance to methods based on the autocovariance and the mutual information. In addition, we compare the novel method to an established method based on cross-validatory mean-square errors of predictors corresponding to different choices of delay. The mean local autocovariance combines the versatility of the mutual information with some of the robustness to noise of the autocovariance.

DOI: 10.1103/PhysRevE.101.012214

I. INTRODUCTION

There exist few real-world systems for which the assumption of linearity could be expected rigorously to be correct. Nonlinear time-series analysis proceeds without making the limiting assumption of linearity, and has found employment in a wide range of areas. Applications of nonlinear time-series analysis include separating the heartbeat of a fetus from that of its host [1, pp. 193–196], predicting epileptic seizures [2], exploring the mechanisms underlying sunspots [3] and geological features [4], and characterizing fluctuations in commodity prices [5,6].

Nonlinear time-series analysis of a real scalar time series of M data $(s_i)_{i=1}^{M} \subseteq \mathbb{R}$ often begins with uniform time-delay reconstruction. This proceeds according to a time-delay map given by

$$\boldsymbol{v}_i = (s_i, s_{i+\tau}, \ldots, s_{i+(m-1)\tau})^T,$$

where m is called the reconstruction dimension and τ is called the delay, and leads to $M-(m-1)\tau$ delay vectors $(\boldsymbol{v}_i)_{i=1}^{M-(m-1)\tau} \subseteq \mathbb{R}^m$. Takens [7] (see also Sauer *et al.* [8]) proved that, in some sense, the delay τ does not typically matter. Specifically, when the time-series arises from a dynamical system on a compact n-dimensional manifold $\mathcal{N}$, as long as the reconstruction dimension m is greater than or equal to $2n+1$, almost all observation functions and delays τ will typically lead to a time-delay map which embeds $\mathcal{N}$ in $\mathbb{R}^m$ [7].

This result only applies directly to systems which evolve and can be observed without any extraneous stochastic or dynamical disturbances, conditions which cannot be guaranteed in observations of experimental or real-world systems. However, Stark [9] identified a wide range of conditions under which, assuming that observations are independent of deterministic external influences which unfold on a compact p-dimensional manifold $\mathcal{P}$, a time-delay reconstruction of dimension $2(n+p)+1$ or greater embeds in $\mathbb{R}^m$ the manifold $\mathcal{M} \times \mathcal{P}$. Further, Stark *et al.* [10] showed that for systems which are stochastically, rather than deterministically, forced, or which are observed amid noise, under a wide range of conditions the dimensional requirement for reconstruction dimension is the same as in the case without disturbances: $m \geqslant 2n+1$.

Only for a time series of infinite length can the results of Takens [7], Stark [9], or Stark *et al.* [10] provide direct support that a time-delay map will faithfully represent geometry. In practice, for finite data length M, if the delay τ is too small, then successive coordinates of a delay vector are strongly correlated, so that most delay vectors deviate little from the straight line in $\mathbb{R}^m$ along which all coordinates would be equal [11]. However, when the delay τ is too large successive coordinates can practically be independent [12], which leads to diffuse and confusing structure.

There exist numerous means [1,13,14] of choosing a delay τ for uniform delay reconstruction from frequently sampled flowlike data. Two of the most popular each involve choosing the first delay τ which locally minimizes the mean redundancy between pairs of successive coordinates. The first of these involves minimizing locally the linear redundancy by choosing τ as close as possible to the first zero of the autocovariance

*jackmoore@tongji.edu.cn

function

$$C\big[(s_i)_{i=1}^{M}\big](\tau) = \frac{1}{M-\tau}\sum_{i=1}^{M-\tau}[(s(i)-\bar{s}_{[1,M-\tau]}) \times (s(i+\tau)-\bar{s}_{[\tau+1,M]})],$$

where $\bar{s}_{[k,l]}$ is the mean $\bar{s}_{[k,l]} = \frac{1}{l-k+1}\sum_{j=k}^{l} s(j)$. The second [15] involves minimizing locally the mean redundancy of the time series after it has in some way been discretized [16]. For $i = 1, 2, \ldots, M$, let a_i denote the state to which the datum s_i is discretized, so that the full discretized sequence is $(a_i)_{i=1}^{M}$. The second approach involves choosing the smallest τ which is a local minimum of the mutual information

$$I(\tau) = \frac{1}{M-\tau}\sum_{i=1}^{M-\tau}\log\left[\frac{P_{0,\tau}(a_i, a_{i+\tau})}{P_0(a_i)P_\tau(a_{i+\tau})}\right],$$

where, for $j = 0, \tau$, $P_j(a)$ is the probability of occurrences of the symbol a in the symbolic time series $(a_{i+j})_{i=1}^{M-\tau}$, and $P_{0,\tau}(a, b)$ is the probability of occurrences of the ordered pair (a, b) in the time series $[(a_i, a_{i+\tau})]_{i=1}^{M-\tau}$.

The method of uniform delay reconstruction has received criticism [17,18]. A more general recipe for delay reconstruction in dimension m is nonuniform delay reconstruction given by

$$\boldsymbol{v}_i = (s_{i+\tau_1}, s_{i+\tau_2}, \ldots, s_{i+\tau_m})^T,$$

where $0 \leqslant \tau_1 < \tau_2 < \ldots < \tau_m$. For fixed reconstruction dimension m, this method of delay reconstruction has m free parameters. In contrast, a uniform delay reconstruction contains only one free parameter, τ. There is little reason to believe that a uniform delay reconstruction would generally be optimal. In addition, the method of mutual information is sometimes criticized [17] because discretization of continuous data often involves a tuning parameter; for example, a histogram bin width.

Chan and Tong [19] (see also Cheng and Tong [20]) describe an approach to identify coordinates for a nonuniform time-delay reconstruction through cross-validation of λ-step ahead predictors $\hat{F} : \mathbb{R}^m \to \mathbb{R}$ given by

$$\hat{F}(\boldsymbol{v}) = \frac{\sum_{i=1}^{M-(K-1+\lambda)} s_{i+K-1+\lambda}\,\phi\left(\frac{\boldsymbol{v}-\boldsymbol{v}_i}{h}\right)}{\sum_{i=1}^{M-(K-1+\lambda)}\phi\left(\frac{\boldsymbol{v}-\boldsymbol{v}_i}{h}\right)},$$

where $\phi : \mathbb{R}^m \to \mathbb{R}$ is a kernel function, $h \in \mathbb{R}$ is a bandwidth, K is the length of the reconstruction window, and terms with which to predict $s_{i+K-1+\lambda}$ are sought from among $s_i, s_{i+1}, \ldots, s_{i+K-1}$. They consider the leave-one-out cross-validatory mean-square error, which with prediction horizon λ is given by

$$V_\lambda(\boldsymbol{T}) = \frac{\sum_{i=1}^{M-(K-1+\lambda)}[s_{i+K-1+\lambda} - \hat{F}^{(i)}(\boldsymbol{v}_i)]^2}{M-(K-1+\lambda)},$$

where $\boldsymbol{T} = (\tau_1, \tau_2, \ldots, \tau_m) \in \mathbb{R}^m$ is the vector of delays, which satisfy $0 \leqslant \tau_1 < \tau_2 < \ldots < \tau_m < K$, and for $i = 1, 2, \ldots, M-K$, $\hat{F}^{(i)} : \mathbb{R}^m \to \mathbb{R}$ given by

$$\hat{F}^{(i)}(\boldsymbol{v}) = \frac{\sum_{\substack{j=1\\ j\neq i}}^{M-(K-1+\lambda)} s_{j+K-1+\lambda}\,\phi\left(\frac{\boldsymbol{v}-\boldsymbol{v}_j}{h}\right)}{\sum_{\substack{j=1\\ j\neq i}}^{M-(K-1+\lambda)}\phi\left(\frac{\boldsymbol{v}-\boldsymbol{v}_j}{h}\right)}$$

is a version of the predictor which is not trained to predict the value $s_{i+K-1+\lambda}$. For a stationary bounded sequence $(s_i)_{i=1}^{M}$ the cross-validatory method has an advantage in terms of mathematical rigour [19]. For fixed $K \in \mathbb{Z}^+$, as M grows large, if $h = M^{-\mu(m)/m}$ for some $\mu(m)$ increasing in m then the probability approaches one that the cross-validatory method identifies the indices $0 \leqslant \tau_1 < \tau_2 < \ldots < \tau_m < K$ defining any least-squares optimal set of regressors $s_{i+\tau_1}, s_{i+\tau_2}, \ldots, s_{i+\tau_m}$ for $s_{i+K-1+\lambda}$.

This cross-validatory approach can also be adapted to the simultaneous selection of uniform delay and reconstruction dimension [21]. This procedure allows simultaneous selection of a delay and a reconstruction dimension, which has the potential to provide better results then performing these tasks sequentially. Of course, the best choice of embedding parameters could also depend on the purpose to which the reconstruction will be applied [1].

Despite its limitations and the availability of a mathematically rigorous means to choose simultaneously dimension and delay(s), the process of first choosing a delay and subsequently choosing a reconstruction dimension suitable for this delay remains a popular choice for nonlinear time-series analysis [22–26]. In part this may be because of its simplicity and ease of implementation, or due to its exposition in important textbooks [1,12]. Also, a researcher in nonlinear time-series analysis might be reluctant to have the effectiveness of a method rely on implementation of precisely the optimal nonuniform delay reconstruction. Hence, this limitation of uniform time-delay reconstruction is coupled with a strength; many people, wary of overfitting, would be willing to sacrifice a significant amount of performance to reduce the number of parameters of a model from m to 1.

The two most popular methods of selecting a uniform delay have relative strengths and weaknesses, which will be outlined now and illustrated in Sec. III. Mutual information allows selection of useful delays for systems for which the method of autocovariance fails. However, the autocovariance is far more robust to noise than is the mutual information. We consider a function, the mean local autocovariance, which combines the versatility of the mutual information function with some of the robustness to noise of the autocovariance function. Furthermore, the method avoids tuning parameters. We compare the mean local autocovariance method not only to the autocovariance and mutual information methods, but also to versions of the cross-validatory method applicable to choosing uniform delay reconstructions.

II. METHODS

A. Mean local autocovariance

Underlying the proposed method is the principle that if consecutive nonoverlapping intervals of duration τ are independent then, in particular, two data separated by duration τ

will be independent. That is, independence of these intervals would imply that τ satisfied the property for which we seek a zero of the autocovariance or a minimum of the mutual information. Also, over a short interval 2τ, the dynamics could be approximated linearly, so that independence of consecutive nonoverlapping intervals of data $(s_i)_{i=j}^{j+\tau-1}$ and $(s_i)_{i=j+\tau}^{j+2\tau-1}$ could be sought by seeking linear independence. This would involve seeking a root of the autocovariance $C[(s_i)_{i=j}^{j+2\tau-1}](\tau)$ of the short time series $(s_i)_{i=j}^{j+2\tau-1}$.

The discussion of the preceding paragraph considers only a particular segment of the time series. The best delay τ for the time series as a whole is sought by calculating the mean of the autocovariances at delay τ of all intervals of length 2τ,

$$L(\tau) = \frac{1}{M-2\tau+1} \sum_{j=1}^{M-2\tau+1} C\big[(s_i)_{i=j}^{j+2\tau-1}\big](\tau),$$

called the mean local autocovariance. The proposed method involves seeking the first root of the mean local autocovariance $L(\tau)$ which satisfies $\tau > 1$ and at which there is a negative slope.

We note that the mean local autocovariance will not usually exactly satisfy $L(\tau) = 0$ for any integer $\tau > 1$. Usually, there will be a delay τ^* for which $L(\tau^*) > 0$ and $L(\tau^* + 1) < 0$. The mean local autocovariance method involves choosing the delay $\tau = \tau^*$ or $\tau = \tau^* + 1$ for which $L(\tau)$ has a value closer to zero. Such a delay τ will be referred to as the first interior root of the mean local autocovariance.

The condition $\tau > 1$ arises because the mean local autocovariance always satisfies $L(1) = 0$. It can be hard to tell whether transitions from positive to negative sign which occur for small τ are meaningful and will lead to appropriate delays. Hence, this method is only suitable for determining reconstruction delays for frequently sampled flowlike data. Although roots at which there is a negative slope which occur for small delay τ might not be meaningful, we do not complicate our heuristic by attempting to distinguish further between small and large τ.

The stipulation of a negative slope requires some explanation. For a delay τ much smaller than the delay which leads to maximal independence we would expect similarity between one sequence of τ data and the next, and hence, a positive value of $L(\tau)$. If the first root which satisfies $\tau \geqslant 2$ has a positive slope, then for small delay $\tau \geqslant 2$ the mean local autocovariance must be negative. This could be the case: (1) because the delay which minimizes redundancy is less than two, or (2) because of noise, a short sample size or other sources of imprecision or inaccuracy. When we know that we are dealing with frequently sampled flows we would not expect the optimal delay to be $\tau = 1$ and, ruling out the first case, we presume that noise has impacted the mean local autocovariance in the small τ part of its domain. The requirement of a negative slope increases the probability that the mean local autocovariance method selects a meaningful delay for frequently sampled flowlike data, but this extra condition might not be appropriate when the data are not known *a priori* to arise from a frequently sampled flow.

A different method to identify the mean local autocovariance function, which is more relevant where the delay τ is large, is related to the total squared error involved in using a translation and possibly a reflection (by which we mean a multiplication by "-1") of each segment of τ consecutive observations $(s_i)_{i=j}^{j+\tau-1}$ to describe the next segment of τ consecutive observations $(s_i)_{i=j+\tau}^{j+2\tau-1}$. The optimal translation would map the mean of each (possibly reflected) segment to the mean of the next, and the resultant summed square error would be

$$\begin{aligned} S_j &= \sum_{i=j}^{j+\tau-1} (\alpha \times \hat{s}_{[j+\tau,j+2\tau-1]}(i+\tau) - \hat{s}_{[j,j+\tau-1]}(i))^2 \\ &= \sum_{i=j}^{j+\tau-1} [(\hat{s}_{[j+\tau,j+2\tau-1]}(i+\tau))^2 + (\hat{s}_{[j,j+\tau-1]}(i))^2 \\ &\quad - 2\alpha \times \hat{s}_{[j,j+\tau-1]}(i)\hat{s}_{[j+\tau,j+2\tau-1]}(i)], \end{aligned}$$

where $\hat{s}_{[j,k]}(i) \triangleq s(i) - \bar{s}_{[j,k]}$ is the difference from the mean over the interval $[j, k]$ and $\alpha = -1$ when a reflection has taken place, while otherwise $\alpha = +1$. If the time series is stationary with variance ${\sigma_{\rm TS}}^2$ and τ is not too small, then each of the first two terms in the summand can be approximated by $\tau{\sigma_{\rm TS}}^2$. Hence, the total summed squared error in this case is

$$\begin{aligned} S &= \sum_{j=1}^{M-2\tau+1} S_j \\ &= 2\tau(M-2\tau+1){\sigma_{\rm TS}}^2 - 2(M-2\tau+1)\alpha \times L(\tau) \\ &= 2\tau(M-2\tau+1){\sigma_{\rm TS}}^2 - |2(M-2\tau+1)L(\tau)|, \end{aligned}$$

where the final equality arises from allowing minimization of S to dictate whether or not to include a reflection in all the transformations of one segment to the segment which follows. This sum involves $\tau(M-2\tau+1)$ error terms, and so, subject to the assumptions and approximations we have used, the mean-square error of these transformations is

$$\frac{1}{\tau(M-2\tau+1)} S = 2{\sigma_{\rm TS}}^2 - \tfrac{2}{\tau}|L(\tau)|.$$

Thus, attempting to minimize the redundancy between successive segments by maximizing the mean-square error incurred by these transformations is equivalent—subject to our approximations—to minimizing $\frac{1}{\tau}|L(\tau)|$, and this will be minimized at a root of the mean local autocovariance.

The preceding sketch did not motivate, but instead was motivated by, our initial investigation of the mean local autocovariance as a means to identify delays for reconstruction. Other similar variational approaches might feel like more natural ways to minimize redundancy. For example, instead of choosing a single $\alpha \in \{-1, +1\}$ a distinct least-squares optimal factor $\alpha_j \in \{-1, +1\}$ could be chosen for each $j = 1, 2, \ldots, M-2\tau+1$. This approach could suggest minimizing not the sum of covariances but the sum of absolute values of covariances. In turn, choosing a least-squares optimal sequence $\alpha_1, \alpha_2, \ldots, \alpha_{M-2\tau+1} \in \mathbb{R}$ could suggest minimization of a sum of squares of covariances. Our limited investigation of the functions identified via these arguably more obvious variational approaches suggested that they did not exhibit roots—they are, after all, sums of nonnegative terms—and such minima as they exhibited were not stable with respect to

observational noise. We will not present detailed results from these heuristics.

B. Calculation of mutual information

Following Abarbanel [12], Kantz and Schreiber [1], and Small [13], we discretize the time series $(s_i)_{i=1}^M$ with bins of identical width which partition the interval $[\min[(s_i)_{i=1}^M], \max[(s_i)_{i=1}^M]]$. If this region is spanned by B bins, then the joint probability $P_{0,\tau}$ must be estimated for B^2 distinct symbols. Although methods have been developed which can estimate entropies with less bias [27], the histogram approach seems popular in nonlinear time-series analysis [1,12]. There exist sophisticated ways to select a good bin width with which to estimate a probability distribution [28]. However, Kantz and Schreiber [1] suggest that since the variation of mutual information is of more interest than its absolute value, the most critical requirement when using a histogram to estimate the mutual information is to choose bins wide enough that $I(\tau)$ varies stably with τ. For simplicity, but fairly arbitrarily, the width of each bin is chosen such that the mean number of data per estimate is about eight. Specifically, B is chosen as $B = \lceil\sqrt{M/8}\rceil$, where, for $z \in \mathbb{R}$, $\lceil z\rceil$ denotes the smallest integer greater than or equal to z.

C. Calculation of cross-validatory mean-square error

Following Chan and Tong [19], we linearly scale the data such that it has unit variance before calculating the leave-one-out cross-validatory mean-square error (CV). Since we are focusing on uniform delay reconstruction, following Giannerini and Rosa [21], we consider only sets of delays of the form $\tau_i = (i-1)\tau$, $i = 1, 2, \ldots, m$. We calculate the cross-validatory error for two distinct prediction horizons λ which will be detailed later (Sec. II E), for each combination of dimension $m = 1, 2, \ldots, 8$ and uniform delay $\tau = 1, 2, \ldots, \tau_{\max}$, where, as detailed in Sec. II D, $\tau_{\max}$ depends on the system under consideration. Following Chan and Tong [19], we use a "data-driven" bandwidth h. Specifically, we consider a geometric sequence of 120 potential bandwidths beginning at 2^{-20} and ending at 2^{10}, with each candidate bandwidth separated by a factor of $2^{1/4}$ from that which precedes it. We note that the total number of bandwidths is similar to the number, 100, considered by Chan and Tong [19]. Furthermore, the logarithmic interval between successive bandwidths is similar to the mean logarithmic separation of the bandwidths which Cheng and Tong [20] list when they describe their mean-square error calculation. The final bandwidth used for each dimension m is that which minimizes the cross-validatory error. Again following Chan and Tong [19], we employ a Gaussian kernel $\phi : \mathbb{R}^m \to \mathbb{R}$, $\boldsymbol{v} \mapsto \exp(-\frac{1}{2}\|\boldsymbol{v}\|_2{}^2)$.

D. Data

The autocovariance, mutual information, cross-validatory and and mean local autocovariance methods for choosing delays are evaluated on three well-established chaotic flows as well as a quasiperiodic system with a tunable parameter a. Also considered is a ± 10 mV electrocardiogram (ECG), sampled at 250 Hz and a resolution of 12 bits [29], comprising the first $M = 5\,000$ data of the first channel of the digitized recording "chf10" obtained from the BIDMC Congestive Heart Failure Database [30,31]. The chaotic flow data comprised the x coordinates of the Rössler [32] system

$$\begin{pmatrix}\dot{x}\\ \dot{y}\\ \dot{z}\end{pmatrix} = \begin{pmatrix}-z-y\\ x+ay\\ b+z(x-c)\end{pmatrix},$$

with $a = 0.15$, $b = 0.20$ and $c = 10.0$, sampled with step size $\Delta t = 0.1$; the Lorenz [33] system

$$\begin{pmatrix}\dot{x}\\ \dot{y}\\ \dot{z}\end{pmatrix} = \begin{pmatrix}\sigma(y-x)\\ -y+rx-xz\\ -bz+xy\end{pmatrix}$$

with $r = 28$, $b = 8/3$ and $\sigma = 10$, sampled with step size $\Delta t = 0.01$; and the Moore and Spiegel [34] system

$$\begin{pmatrix}\dot{x}\\ \dot{y}\\ \dot{z}\end{pmatrix} = \begin{pmatrix}y\\ z\\ -z-(T-R+Rx^2)y-Tx\end{pmatrix}$$

with, following Sprott [35, p. 77], $T = 6$ and $R = 20$, sampled with step size $\Delta t = 0.01$. The final, quasiperiodic flow comprised a series of values $x = ax_1 + (1-a)x_2$ where $0 \leqslant a \leqslant 1$ and for $i = 1, 2$,

$$\begin{pmatrix}\dot{x}_i\\ \dot{y}_i\end{pmatrix} = \omega_i\left[\begin{pmatrix}-y_i\\ x_i\end{pmatrix} + \left(1-x_i{}^2-y_i{}^2\right)\begin{pmatrix}x_i\\ y_i\end{pmatrix}\right],$$

where $\omega_1 = 2\pi$ is the slower angular frequency and $\omega_2 = 20$ is the faster angular frequency. For $i = 1$ or 2, in the absence of noise, the trajectory described by the coordinates x_i, y_i tends towards motion along the unit circle with constant speed ω_i. Hence, in the absence of noise, $x = ax_1 + (1-a)x_2$ tends towards a quasiperiodic signal, and when $a = 0$ or $a = 1$ this signal is periodic with period 1 or $\pi/10$, respectively. This flow is sampled with step size $\Delta t = 0.01$, and so when $a = 0$ or 1 the period of x is $T_1 \triangleq 100$ sampling periods or $T_2 \triangleq 10\pi \approx 31$ sampling periods, respectively.

Each system was integrated using a Runge-Kutta (4, 5) method [36]. The three chaotic systems were integrated from the initial condition $(1, 0, 0)^T$, while the limit cycle was integrated from an initial condition each coordinate of which was chosen independently and uniformly at random from a Gaussian distribution with mean zero and unit variance. To allow convergence onto the attractor, in each case the first 10 000 values were discarded. Time series of length $M = 500$, 5 000, and 50 000 were considered.

The optimal delay according to each method was selected from among $\tau = 1, 2, \ldots, \tau_{\max}$, where $\tau_{\max} = 20$, 30, 80, 50 and 25 for data from the Rössler [32] system, the Lorenz [33] system, the Moore and Spiegel [34] system, the quasiperiodic system and ECG, respectively.

1. Noise

To each set of flow (but not ECG) data was added Gaussian observational noise with mean zero and standard deviation equal to $\epsilon = 0\%$, 10% or 50% of the population standard deviation of the original time series.

Since dynamic noise led to numerical issues or divergence for the Rössler [32] and Moore and Spiegel [34] systems, this was added only to the Lorenz [33] and quasiperiodic systems. Gaussian dynamic noise was added, at a relative level $\epsilon = 5\%$, 10%, or 20% and independently to each of coordinate, in the following manner. To generate a time series of length M with dynamic noise of level ϵ, first, the system was integrated from the usual initial condition for $M + 10\,000$ time steps, and the first 10 000 data for each coordinate was discarded. The population standard deviation σ_i of each coordinate i was calculated. Next, the system was integrated from the same initial condition for $M + 10\,000$ time steps, adding to each coordinate i, after each time step, Gaussian noise with mean zero and standard deviation $\epsilon\sigma_i$. Dynamic noise was added only after observation.

E. Evaluation

In the absence of noise the first minimum of the mutual information is a trusted means to choose a sensible delay for uniform delay reconstruction. Although we do exhibit and examine reconstructions which arise from competing delays, to some extent the assessment of different methods is based on how well they can maintain, in the presence of noise, a value of delay which is close to that obtained with clean data and using the mutual information method.

For the quasiperiodic attractor, at least when $a = 0$ or 1, we can identify an unambiguously optimal value for the delay. A delay equal to a quarter of the period—$\tau = 25$ for $a = 0$ and $\tau = 5\pi/2 \approx 7.85 \approx 8$ when $a = 1$—leads to a circular embedding which leads to wide separation of distinct regions of state space. As a increases from 0 to 1 we might naively expect an appropriate delay to increase monotonically from $\tau = 8$ to $\tau = 25$.

To facilitate comparison, before they were plotted the values of the autocovariance, mutual information, cross-validatory error and mean local autocovariance were scaled linearly. Specifically, for each distinct noise level ϵ and length M of data the values of the autocovariance, mutual information, and mean local autocovariance were individually scaled linearly to span an interval of unit length. Because only their minima, and not their roots, were sought, before they were plotted the mutual information and cross-validatory error were also shifted by a constant such that they spanned the interval [0, 1].

Three of the methods which we consider could be thought of as arising through restricting attention to reconstruction dimension $m = 2$. Hence, for the fourth method—the cross-validatory method—in some cases we consider restricting the reconstruction dimension to $m = 2$. However, we do also present $V_\lambda(\tau)$ versus τ for the dimension for which the minimal cross-validatory error is achieved. We write $\hat{m}$ for the error-minimizing reconstruction dimension. For the three chaotic flows and for ECG data we consider predictions either $\lambda = 1$ steps ahead or a number of steps λ equal to the reconstruction delay selected via the mean local autocovariance method when no noise is added by us. For the quasiperiodic data we consider predicting a number of steps ahead given by the integer closest to a quarter of the period corresponding to linear interpolation between the frequencies of the noiseless signals for $a = 0$ and $a = 1$,

$$\lambda = \mathrm{nint}\left(\frac{1}{4}\left\{\Delta t\left[a + (1-a)\frac{10}{\pi}\right]\right\}^{-1}\right), \tag{1}$$

where nint$(\cdot)$ denotes rounding to the nearest integer.

III. RESULTS AND DISCUSSION

A. Flow data

Figure 1 pertains to $M = 5\,000$ observations of the Rössler [32] system made amid $\epsilon = 0\%$, 10% or 50% added Gaussian measurement noise. Figures 1(a), 1(b) and 1(e) show that three of the methods, autocovariance, mutual information, and mean local autocovariance, produce similar results for this time series, which has a strong almost-periodic component. The delays suggested by the autocovariance and mean local autocovariance methods, $\tau = 15$ and $\tau = 14$, respectively, are stable with respect to noise level. It is hard to discern from the figure, but the mutual information method suggests delay $\tau = 15$ for $\epsilon = 0\%$ and 50%, and suggests $\tau = 14$ for $\epsilon = 10\%$. It would be difficult to infer confidently from Fig. 1(c), but for a clean time series the minimum of the cross-validatory error for prediction horizon $\lambda = 1$ (for the error-minimizing reconstruction dimension $\hat{m} = 2$) occurs at time delay $\tau = 2$. In the presence of observational noise $\epsilon = 10\%$ or $\epsilon = 50\%$ the location of the minimum of V_1 for $m = \hat{m}$ is instead $\tau = 5$. For reconstruction dimension $m = 2$ the minimum of V_1 is less stable as nonvanishing noise levels vary: its location changes from $\tau = 19$ to $\tau = 1$ as observational noise increases from $\epsilon = 10\%$ to $\epsilon = 50\%$. Figure 1(d) shows that, when considering the longer prediction horizon $\lambda = 14$ and reconstruction dimension $m = 2$, the minimum of the cross-validatory error increases from $\tau = 5$ to 12 and through to 15 as the level of observational noise increases from $\epsilon = 0\%$ to 10% and through to 50%. The figure also shows that for $m = \hat{m}$ the minimum of V_{14} remains stable at $\tau = 4$ as the noise level varies.

Figure 1(f) shows the Rössler [32] attractor reconstructed [37] in dimension $m = 2$ from clean observations using delays $\tau = 4$ and 14. The delay $\tau = 14$ is suggested by the mean local autocovariance method, while $\tau = 4$ is suggested by cross-validation from reconstruction in the error-minimizing dimension $\hat{m}$ when using a prediction horizon $\lambda = 14$. The figure shows that under $\tau = 14$, the delay suggested by the mean local autocovariance method, the two-dimensional reconstruction of the attractor occupies a larger fraction of its convex hull than is the case under $\tau = 4$. This will tend to allow clearer separation of distinct points and distinct trajectories in state space reconstructed in two dimensions. However, $\tau = 14$ does lead to an additional intersection in $m = 2$ dimensions (evident at the top right of the reconstruction) which $\tau = 4$ avoids. Hence the delay $\tau = 4$ may be more appropriate than $\tau = 14$.

Figure 2 presents results for $M = 5\,000$ observations of the Lorenz [33] system made amid $\epsilon = 0\%$, 10% or 50% added Gaussian observational noise. Figures 2(b)–2(e) show that three of the methods, mutual information, cross-validation and mean local autocovariance, can produce from clean observations of the x-coordinate of this system the same choice of

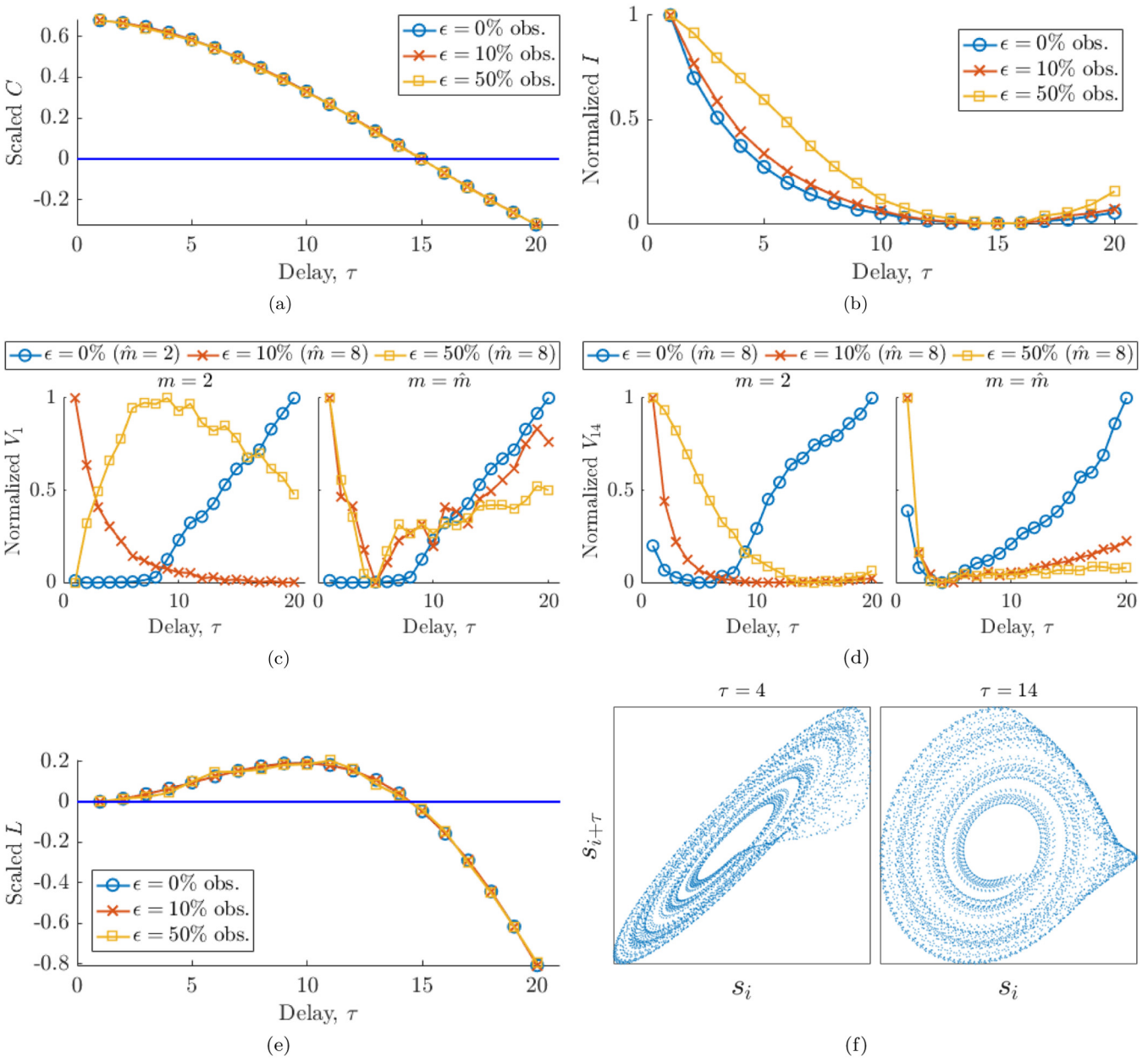


FIG. 1. For $M = 5\,000$ observations of the Rössler [32] system made amid $\epsilon = 0\%$, 10%, or 50% added Gaussian measurement noise, the variation with delay τ of: (a) autocovariance (C), (b) mutual information (I), (c) cross-validatory mean-square error with prediction horizon $\lambda = 1$ (V_1) for reconstruction dimension (left) $m = 2$ and (right) $m = \hat{m}$, (d) cross-validatory mean-square error with $\lambda = 14$ (V_{14}) for (left) $m = 2$ and (right) $m = \hat{m}$, and (e) mean local autocovariance (L). For each distinct noise level ϵ, each function has been scaled linearly such that it spans an interval of unit length. In addition, for each distinct noise level ϵ, I, and V_λ have been shifted to span [0, 1]. (f) Time-delay reconstructions of a clean time series using the delays $\tau = 4$ and $\tau = 14$ suggested by the minimum of the cross-validatory error, using prediction horizon $\lambda = 14$ with the optimal reconstruction dimension $\hat{m}$, and the first interior root of the mean local autocovariance, respectively.

delay: $\tau = 17$. The first interior root of the mean local autocovariance is stable as observational noise is increased to $\epsilon = 50\%$. In contrast, it is hard to discern minima in the mutual information produced from noisy observations. Close inspection might reveal that for $\epsilon = 10\%$ the first minima is at $\tau = 20$ and for $\epsilon = 50\%$ the first minima is at $\tau = 15$. Figure 2(c) shows that when reconstruction dimension restricted to $m = 2$, the minimum of the cross-validatory error using prediction horizon $\lambda = 1$ is fairly stable as observational noise increases from $\epsilon = 0\%$ to $\epsilon = 10\%$; the minimum shifts from $\tau = 17$ to $\tau = 20$. However, if observational noise increases to $\epsilon = 50\%$ or if the error-minimizing reconstruction dimension $m = \hat{m}$ is used then the minimum shifts dramatically, into the interval $1 \leqslant \tau \leqslant 4$. When using instead prediction horizon $\lambda = 17$ and with reconstruction dimension restricted to $m = 2$, as Figure 2(d) suggests, the minimum of the cross-validatory error remains at $\tau = 16$ or 17 as the level of observational noise varies. If the error-minimizing reconstruction dimension

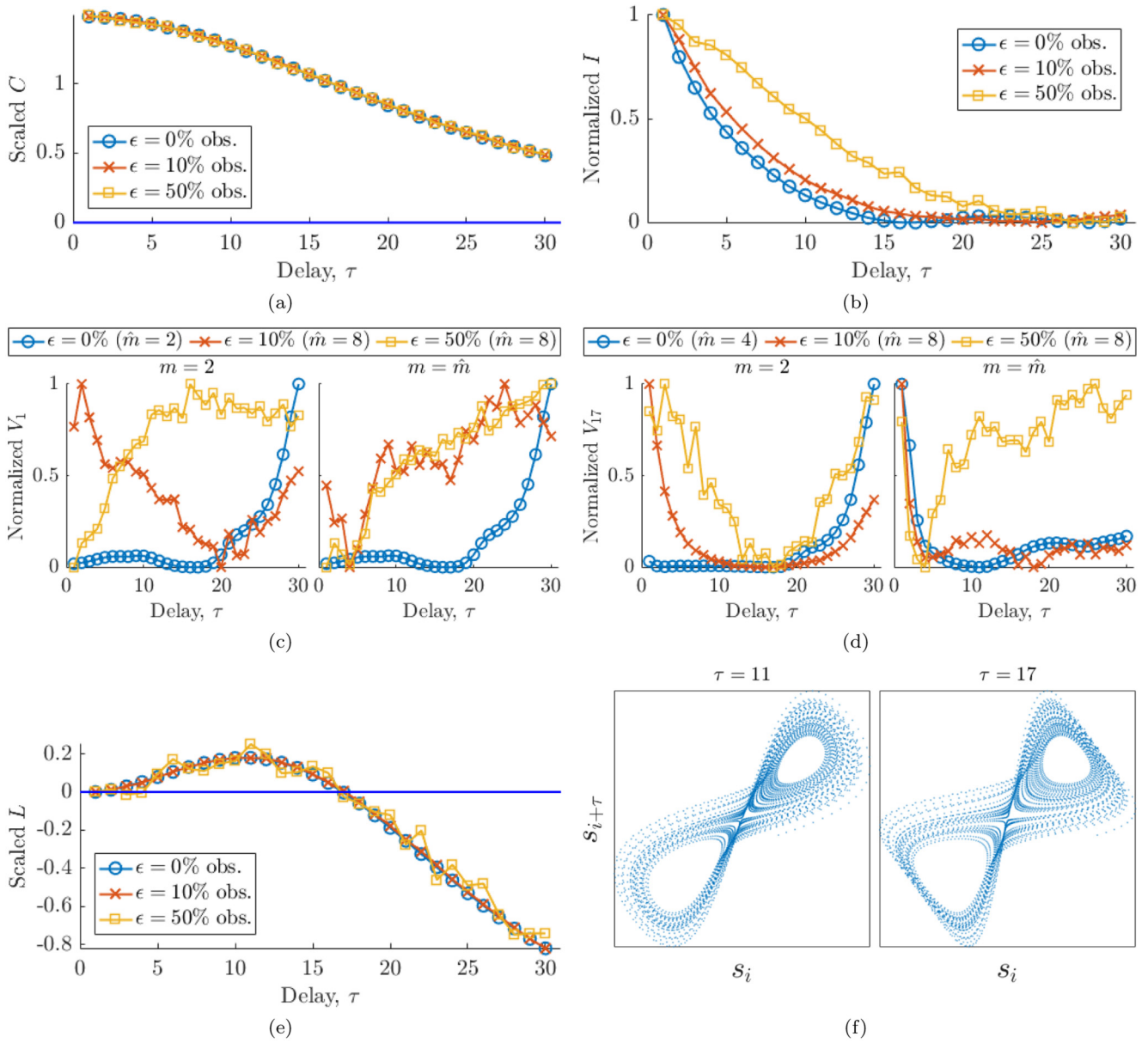


FIG. 2. (a–e) Figures analogous to Figs. 1(a)–1(e), but for $M = 5\,000$ points from the Lorenz [33] system and using, in (d), prediction horizon $\lambda = 17$. (f) Time-delay reconstructions of a clean time series using the delays $\tau = 11$ and $\tau = 17$ suggested by the minimum of the cross-validatory mean-square error using prediction horizon $\lambda = 17$ with the optimal reconstruction dimension $\hat{m}$ and the first interior root of the mean local autocovariance, respectively.

$m = \hat{m}$ is used, then the minimum shifts dramatically, from $\tau = 11$ to 18 to 4 as the noise level increases from $\epsilon = 0\%$ to 10% to 50%. Figure 2(a) shows that the autocovariance curve is admirably stable with respect to noise, but that it does not achieve a root (and thus suggest a delay) over $0 \leqslant \tau \leqslant 30$.

The Lorenz [33] attractor reconstructed in dimension $m = 2$ from clean observations using delays, $\tau = 11$ and 17, appears in Fig. 2(f). The delay $\tau = 17$ was identified using the mutual information and mean local autocovariance methods as well as several incarnations of the cross-validation method, while $\tau = 11$ was the minimum of the cross-validatory error using clean observations, a prediction horizon $\lambda = 17$ and the error-minimizing reconstruction dimension $\hat{m}$. The figure suggests that $\tau = 17$ unfolds the Lorenz [33] attractor acceptably throughout most of the two-dimensional reconstruction space. However, the figure also suggests that the smaller delay more clearly separates distinct points which appear at the top right and bottom left of the portrait. Thus, $\tau = 11$ may correspond to a time delay better than $\tau = 17$.

Figure 3 relates to $M = 5\,000$ observations of the Moore and Spiegel [34] system made amid $\epsilon = 0\%$, 10%, or 50% added Gaussian measurement noise. Figure 3(a) shows that the autocovariance curve is impressively stable with respect to noise, but that it does not achieve a root (and thus suggest a delay) over $0 \leqslant \tau \leqslant 80$. As Fig. 3(b) shows, the curve of mutual information is less stable with respect to noise, and even for the curve of mutual information produced from clean

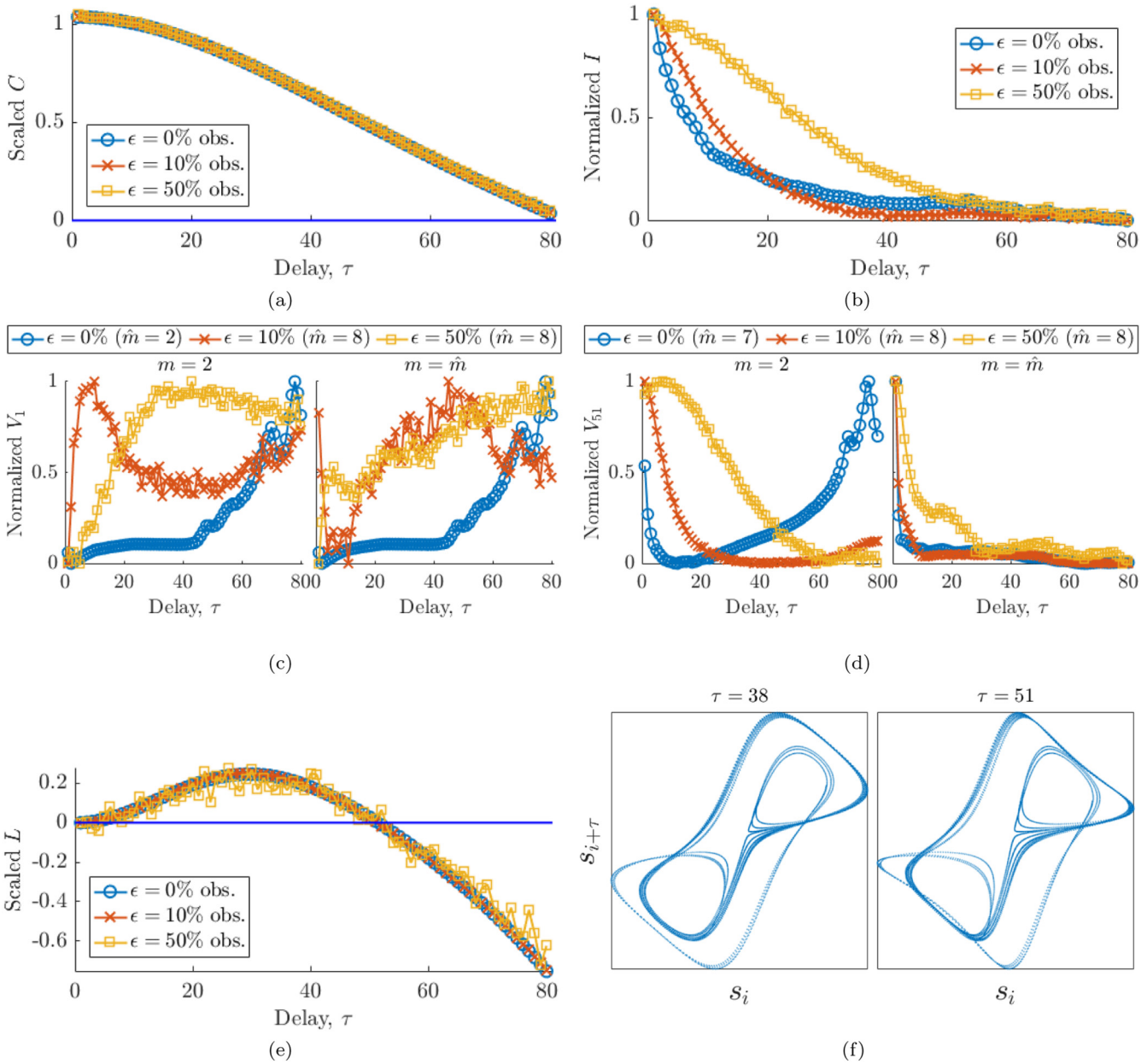


FIG. 3. (a–e) Figures analogous to Figs. 1(a)–1(e), but for $M = 5\,000$ points from the Moore and Spiegel [34] system and using, in (d), prediction horizon $\lambda = 51$. (f) Time-delay reconstructions of a clean time series using delays $\tau = 38$ and $\tau = 51$ suggested by the first minimum of the mutual information and the first interior root of the mean local autocovariance, respectively.

observations it is hard to identify by eye the delay suggested by the mutual information method. However, for clean data the mutual information exhibits subtle local minima at $\tau = 38, 42, 45, 51, 58, 64, 70, 77,$ and 80. Figure 3(c) illustrates that with reconstruction dimension restricted to $m = 2$, the minimum of the cross-validatory error with prediction horizon $\lambda = 1$ is fairly stable as observational noise is added; the minimum shifts from $\tau = 2$ for clean observations to $\tau = 1$ for noise levels of either $\epsilon = 10\%$ or $\epsilon = 50\%$. However, if the error-minimizing reconstruction dimension $m = \hat{m}$ is employed, then the minimum is less stable. Specifically, as observational noise increases from $\epsilon = 0\%$ to 10% then increases again to $\epsilon = 50\%$ the delay increases from $\tau = 2$ to 11 then decreases to $\tau = 1$. As Fig. 3(d) shows, when using the larger prediction horizon $\lambda = 51$ and reconstruction dimension $m = 2$, the minimum of the cross-validatory error increases from $\tau = 12$ to 38 to 59 as the level of noise increases from $\epsilon = 0\%$ to 10% to 50%. When the error-minimizing dimension is used instead, the minimum of the cross-validatory error is more stable, and increases from $\tau = 65$ to 72 to 78 as noise changes from $\epsilon = 0\%$ to 10% to 50%. Figure 3(e) reveals that the choice of delay suggested by the mean local autocovariance remains stable as the level of observational noise increases from $\epsilon = 0\%$ to 10%; for each of these noise levels the mean local autocovariance method suggests the delay $\tau = 51$. For noise $\epsilon = 50\%$ the mean local autocovariance has several zeros in the interval shown, but the general trend remains similar to that for clean data and, after smoothing, the mean

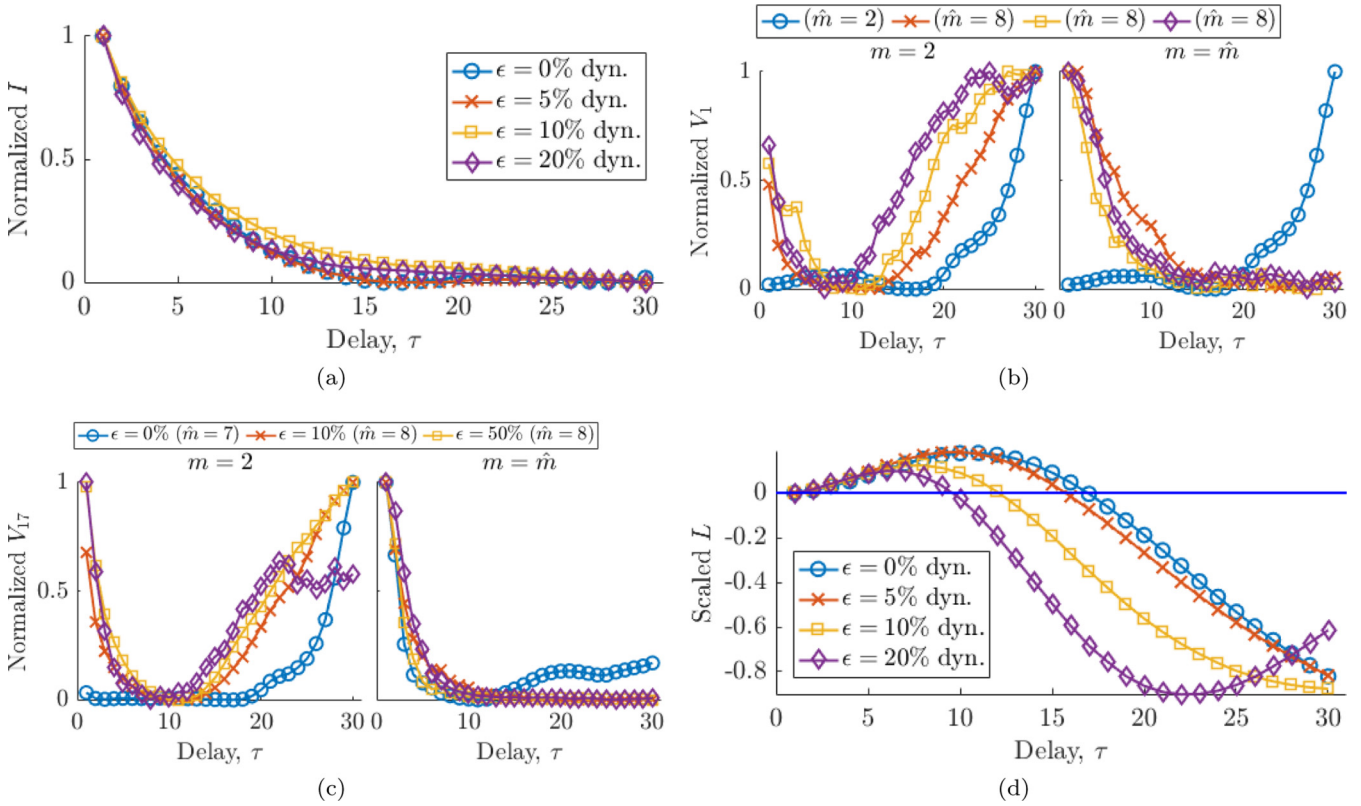


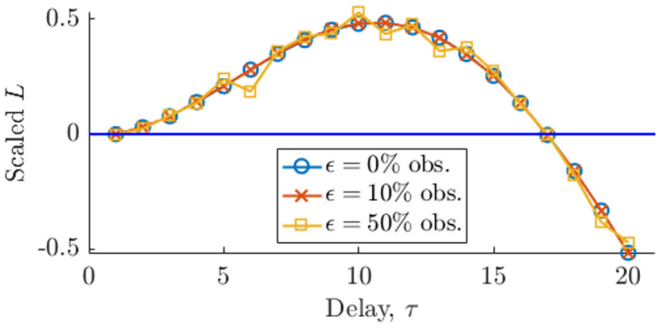
FIG. 4. Figures analogous to Figs. 1(b)–1(e), respectively, but for $M = 5\,000$ observations of the Lorenz [33] system with $\epsilon = 0\%$, 5%, 10%, or 20% added Gaussian dynamic noise and using, in (c), prediction horizon $\lambda = 17$.

local autocovariance method could suggest a delay similar to that which it suggests for clean data.

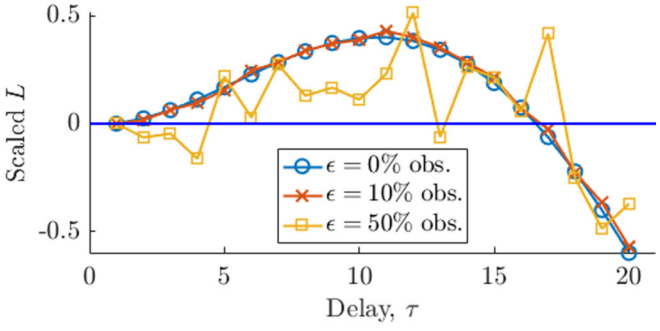
Figure 3(f) presents the Moore and Spiegel [34] attractor reconstructed in dimension $m = 2$ from clean observations using delays $\tau = 38$ and 51. For the delay, $\tau = 51$, obtained using the mean local autocovariance method, the majority of the attractor is unfolded clearly and its structure becomes more easily interpretable. However, when the delay is decreased to $\tau = 38$—the delay suggested by the mutual information and also by the cross-validatory method in the presence of $\epsilon = 10\%$ observational noise when using a prediction horizon $\lambda = 51$ and reconstruction dimension $m = 2$—unnecessary intersections disappear. Hence, a delay smaller than that suggested by the mean local autocovariance method may lead to better reconstruction in two dimensions.

Figure 4 pertains to dynamical instead of observational noise. The figure shows the mutual information, cross-validatory error, and the mean local autocovariance calculated from $M = 5\,000$ observations of the Lorenz [33] system made in the presence of $\epsilon = 0\%$, 5%, 10%, or 20% added Gaussian dynamic noise. It would not be easy to discern reliably from Fig. 4(a), but as noise increases from $\epsilon = 0\%$ to 5% the delay suggested by the mutual information method increases from $\tau = 17$ to $\tau = 18$. For the higher level of dynamic noise $\epsilon = 10\%$ the minimum of the mutual information occurs at the upper boundary of the domain considered, $\tau = 30$, while for $\epsilon = 20\%$ a delicate minimum appears at $\tau = 26$. Figure 4(a) reveals that although the mutual information is, in some sense, quite stable with respect to dynamic noise, its subtle minima are easily obscured. Figure 4(b) shows that when the cross-validatory error is restricted to $m = 2$ and while considering a prediction horizon $\lambda = 1$, its minimum changes monotonically as noise increases, from $\tau = 17$ to 13 to 11 and through to 7 as ϵ increases from 0% to 5% to 10% and through to 20%. Increasing levels of dynamic noise correspond to shorter correlation times. In contrast, when using the error-minimizing reconstruction dimension $m = \hat{m}$, the location of the minimum of the cross-validatory error does not consistently decrease with noise level. As the level of dynamic noise increases the minimum moves from $\tau = 17$ to 27 to 28 and, finally, back to 27. Figure 4(c) shows that similar patterns are present when instead a longer prediction horizon $\lambda = 17$ is employed. When embedding dimension is restricted to $m = 2$, the first interior zero of the cross-validatory error changes monotonically as dynamic noise increases; from $\tau = 17$ to $\tau = 11$ to $\tau = 11$ through to $\tau = 8$ as ϵ increases from 0% to 5% to 10% through to 20%. However, when using the error-minimizing reconstruction dimension $m = \hat{m}$, as the level of dynamic noise increases the minimum moves from $\tau = 11$ to 23 to 30 and, finally, decreases to 26.

Figure 5 shows how trends in the mean local autocovariance vary with data length M for observations of the Lorenz [33] system made amid $\epsilon = 0\%$, 10%, or 50% added Gaussian observational noise. From Fig. 5(a) it can be seen that, when $M = 50\,000$ observations are available, the trends of the mean local autocovariance remain stable as noise increases to $\epsilon = 50\%$, and the delay suggested by the mean local autocovariance method remains fixed at $\tau = 17$. Figure 5(b) shows that even when only $M = 500$ observations are available, mean local autocovariance and the delay which it



(a)



(b)

FIG. 5. Mean local autocovariance (L) versus τ for (a) $M = 50\,000$ and (b) $M = 500$ observations of the Lorenz [33] system made amid $\epsilon = 0\%$, 10%, or 50% added Gaussian observational noise. For each distinct noise level ϵ and data length M, the mean local autocovariance function has been scaled linearly such that it spans an interval of unit length.

suggests remain stable for observational noise up to $\epsilon = 10\%$. For the higher noise level $\epsilon = 50\%$ there are several roots over $0 \leqslant \tau \leqslant 30$, but the broad trend remains stable and, after smoothing, the mean local autocovariance method could suggest a delay similar to that which it suggests for clean data.

Figure 6 relates to delays selected using different methods for $M = 500$ observations of the quasiperiodic system made amid $\epsilon = 0\%$, 10%, 20%, or 50% added Gaussian observational noise for values of the parameter a in the interval $0 \leqslant a \leqslant 1$. When $a = 0$ or 1 the signal is sinusoidal with quarter-period $5\pi/2 \approx 8$ or 25 sampling intervals, respectively, and so a robust method for selecting reconstruction delay would presumably interpolate in some way between $\tau = 8$ when $a = 0$ and $\tau = 25$ when $a = 1$. Figure 6(a) shows that the autocovariance method does just this, although the interpolation is not monotonic but curiously peaked around $a = 0.7$ or 0.8. Figure 6(b) reveals that in the presence of noise the mutual information method appropriately selects a delay close to $\tau = 8$ when $a = 0$, but on average suggests a delay of only about $\tau = 10$ when $a = 1$. Without noise, the mutual information method interpolates from an average delay of $\tau = 2$ when $a = 0$ to a delay of $\tau = 3$ when $a = 1$. In Fig. 6(c) are shown the delays suggested by the cross-validatory method using a prediction horizon λ defined by Eqn. (1) and the error-minimizing embedding dimension $m = \hat{m}$. Applied to clean data, this method interpolates from a mean delay of $\tau = 7.1$ when $a = 0$ to $\tau = 15.8$ when $a = 1$, and exhibits a distinct peak at $a = 0.3$. For noisily observed data the technique suggests average delays higher than $\tau = 8$ when $a = 0$. However, when $a = 1$ the technique suggests for noise levels $\epsilon = 20\%$ and 50%, on average, delays $\tau = 25.5$ and 28.3, respectively, which are close to the target $\tau = 25$. Finally, as can be seen from Fig. 6(d), for noise levels 20% or less, the mean local autocovariance method interpolates correctly and monotonically from $\tau = 8$ to an average delay close to $\tau = 25$. For the highest noise level, $\epsilon = 50\%$, roots of the mean local autocovariance which occur for low delay τ [Fig. 5(b)] mean that the mean local autocovariance method suggests an average delay of only $\tau = 2.9$ when $a = 1$.

Figure 7 relates instead to dynamical noise. Specifically, the figure presents delays selected for $M = 500$ observations of the quasiperiodic system for $0 \leqslant a \leqslant 1$ made amid $\epsilon = 0\%$, 10%, 20%, or 50% added Gaussian dynamical noise. The results are quite similar to those obtained for observational noise and presented in Fig. 6. Figure 7(a) shows that the autocovariance method successfully moves from $\tau = 8$ when $a = 0$ to $\tau = 25$ when $a = 1$, although there is a surprising peak around $a = 0.7$ or $a = 0.8$. Figure 7(b) reveals that in the presence of noise the mutual information method appropriately selects a delay close to $\tau = 8$ when $a = 0$, but on average suggests a delay much lower than $\tau = 25$ when $a = 1$. Figure 7(c) presents delays suggested by the cross-validatory method using a prediction horizon λ defined by Eq. (1) and the error-minimizing embedding dimension $m = \hat{m}$. For noisy data the technique suggests average delays higher than $\tau = 8$ when $a = 0$ and higher than $\tau = 25$ when $a = 1$. Figure 7(d) shows that for dynamical noise of levels 10% or less, the mean local autocovariance method on average interpolates monotonically from $\tau = 8$ to an average delay close to $\tau = 25$. For the highest level of dynamical noise, $\epsilon = 20\%$, the mean delay suggested by the mean local autocovariance method consistently lies below the delay suggested for less noisy data.

B. Electrocardiogram (ECG) data

Figure 8 relates to the $M = 5\,000$ ECG data described in Sec. II D. In Fig. 8(a) the data are shown plotted against time. Figures 8(b)–8(d) show the variation with delay τ of the autocovariance, mean local autocovariance, mutual information, and cross-validatory error. Careful inspection of the figures may reveal that two of the methods, autocovariance and mutual information, suggest the same choice of delay; the first zero of the autocovariance and the first minimum of mutual information coincide at $\tau = 18$. The cross-validatory and mean local autocovariance methods suggest quite different choices of delay. The first interior root of the mean local autocovariance occurs at $\tau = 4$, while the minimum of the cross-validatory error for prediction horizon $\lambda = 1$ and reconstruction dimension $m = \hat{m} = 2$ occurs at $\tau = 2$. When using the prediction horizon $\lambda = 4$, the global minimum of the cross-validatory error is located at $\tau = 3$ when using reconstruction dimension $m = 2$ and at $\tau = 24$ when using the error-minimizing reconstruction dimension $m = 6$. However, when using the error-minimizing reconstruction dimension $m = 6$ and the prediction horizon $\lambda = 4$, the cross-validatory error does have another clear local minimum at delay $\tau = 4$,

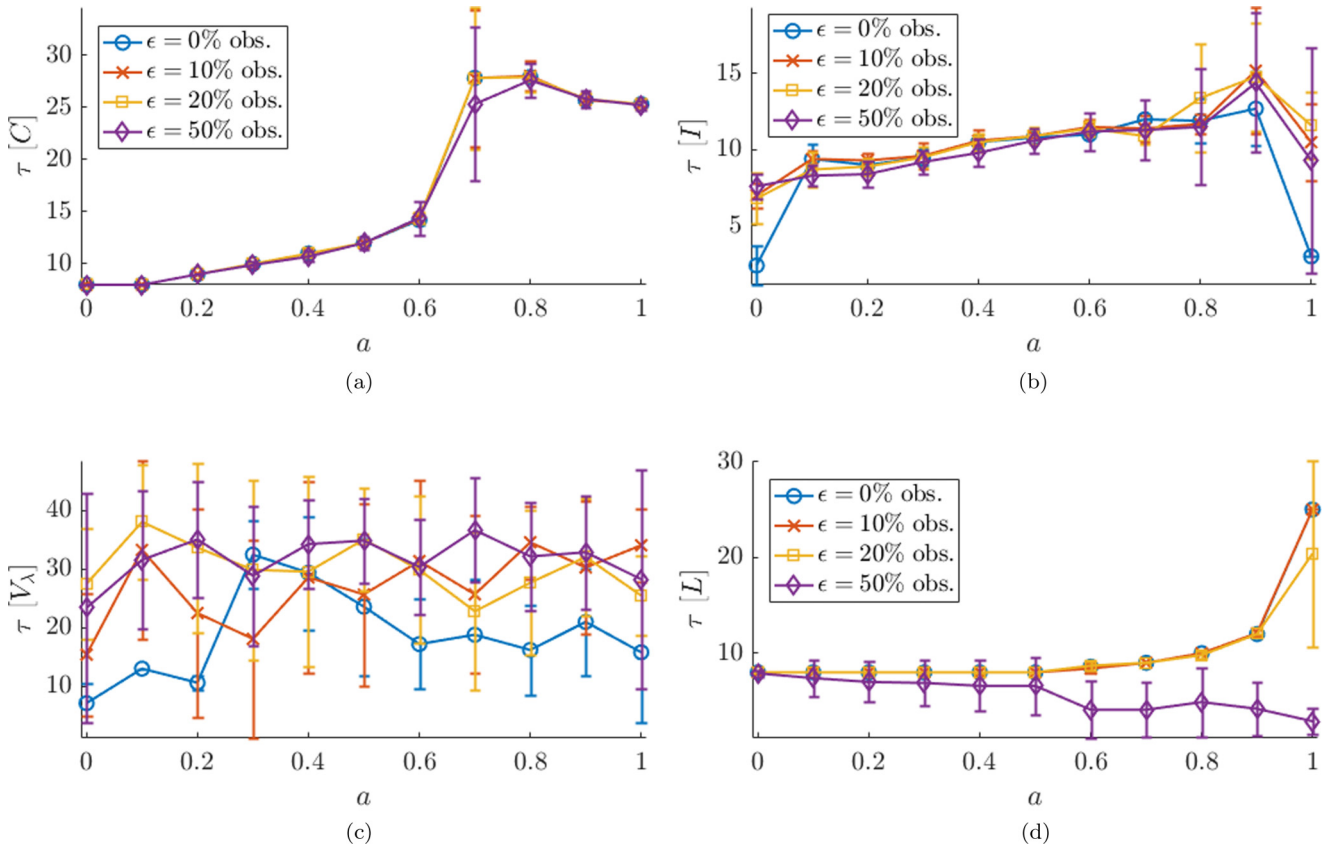


FIG. 6. Delays chosen using the methods of (a) autocovariance (C), (b) mutual information (I), (c) cross-validation using the error-minimizing reconstruction dimension $m = \hat{m}$ and prediction horizon given by Eq. (1), and (d) mean local autocovariance (L), based upon $M = 500$ observations of the quasiperiodic system with parameter $0 \leqslant a \leqslant 1$, made amid $\epsilon = 0\%$, 10%, 20%, or 50% added Gaussian observational noise. Markers and error bars show, respectively, the mean and sample standard deviation over 10 independently and randomly chosen initial conditions.

and this occurs in a region where the cross-validatory error is varying less erratically than around the global minimum $\tau = 24$.

Figure 8(e) shows the ECG time series reconstructed in dimension $m = 2$ using the delays, $\tau = 2$, 4, and 18, suggested by the cross-validatory method with prediction horizon $\lambda = 1$ and reconstruction dimension $m = 2$, the novel mean local autocovariance method and the established autocovariance and mutual information methods. Bradley and Kantz [14] mention that it is impossible to unfold efficiently all aspects of ECG dynamics using a single delay reconstruction. Figure 8(e) may suggest that $\tau = 18$ makes accessible an aspect of the ECG dynamics different from that made accessible by $\tau = 2$ or $\tau = 4$. However, for $\tau = 18$ the two-dimensional reconstruction exhibits extended regions of self-intersection; one oriented horizontally, and the other vertically. These particularly overt overlaps are avoided when using the delays, $\tau = 2$ and $\tau = 4$, prescribed by the cross-validatory and mean local autocovariance methods, respectively. Of these two delays, $\tau = 4$ appears to spread out points more effectively.

C. Discussion

The autocovariance curve proved robust to high levels of observational noise. However, it seemed more useful for suggesting a delay for the Rössler [32] data than for the Lorenz [33], Moore and Spiegel [34], or ECG data. The mutual information method provided useful results for noiseless versions of the Rössler [32] and Lorenz [33] chaotic flows. However, its subtle minima are easily obscured by either observational or dynamic noise. The autocovariance and mutual information methods each suggested a surprisingly high value for the delay with which to reconstruct an ECG time series.

The cross-validatory method produced useful results for some systems and levels of noise, but in other cases, suggested delays which were less intuitive. Some issues should be expected when using a prediction horizon $\lambda = 1$ because of the high sampling rate with which the chaotic flows were observed. Modulo certain conditions, the cross-validatory method is mathematically guaranteed to identify correct parameters in the limit of an infinite number of observations, but when data are oversampled good results may require a long time series. Indeed, the cross-validatory method with prediction horizon $\lambda = 1$ led to more intuitive results for the Rössler [32] system (at least when observed amid noise) and for the ECG times eries, each of which featured more full oscillations over 5 000 observations than did data from the Lorenz [33] or Moore and Spiegel [34] systems.

The high sampling rate might explain the low error-minimizing reconstruction dimension, $\hat{m} = 2$, evidenced in

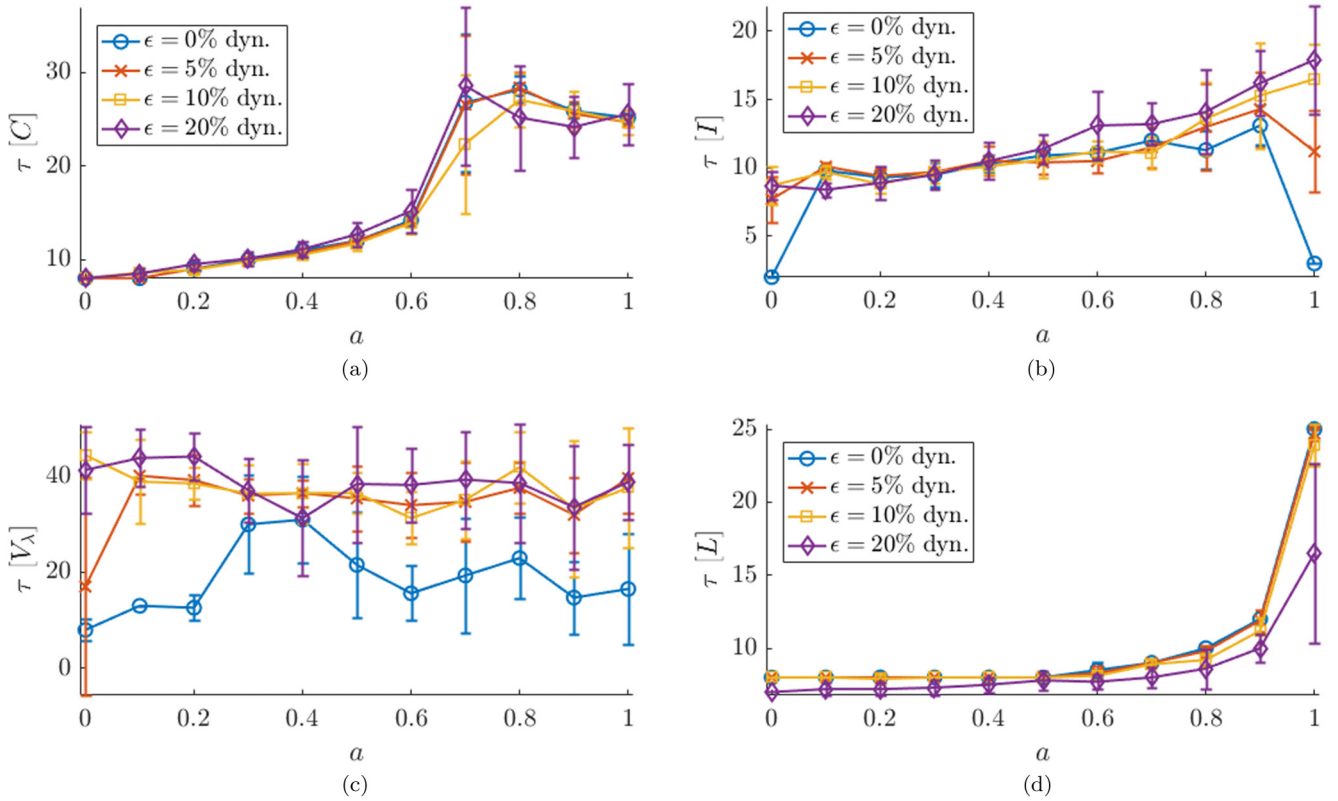


FIG. 7. Figures analogous to Figs. 6(a)–6(d), but for $\epsilon = 0\%$, 5%, 10%, or 20% dynamic noise.

the legends of Figs. 1–3(c), for noiseless data from each chaotic flow when considering prediction horizon $\lambda = 1$. A linear approximation to the dynamics should allow accurate prediction for sufficiently short time scales. Presumably, a two-dimensional reconstruction and high sampling rate allows, across the majority of each attractor, a reasonable linear approximation to each of the chaotic flows which we consider.

With prediction horizon chosen using the mean local autocovariance method, the cross-validatory method identified useful reconstruction delays. In fact, reconstruction of state space showed that the lower delays suggested by the cross-validatory method for clean observations allowed clearer separation of distinct states of the Rössler [32] and Lorenz [33] systems, and for the Rössler [32] system the suggestions for the delay remained stable as observational noise was added. However, the delay prescribed by the same method for the Lorenz [33] and Moore and Spiegel [34] systems varied, sometimes erratically, as noise level increased.

Chan and Tong [19] actually applied the cross-validatory method to dimension estimation [20], and selecting reconstruction delays may not be the precise purpose intended for their framework. In fact, Chan and Tong [19, pp. 252–253] suggest downsampling inconveniently finely sampled data with a delay chosen using the mutual information method [15], which is what led us to use the mean local autocovariance method to choose prediction horizons λ.

The cross-validatory method may have benefited from the consideration of higher embedding dimensions. Figures 1, 2, and 3 show that for the chaotic flow data, regardless of the time series or exact noise level, in the presence of noise the error-minimizing reconstruction dimension is $\hat{m} = 8$, the maximum reconstruction dimension considered, suggesting that reconstruction dimensions which were still higher might have allowed lower cross-validatory error.

Such high values for the error-minimizing embedding dimension $\hat{m}$ in the presence of noise may be surprising, but are not entirely unreasonable. Stark [38] explored the idea that selecting the parameters of a time-delay reconstruction implicitly involves choosing the part of the observations which we wish to describe as deterministic and which part we are willing to attribute to noise. In the case of dynamic noise, choosing a reconstruction dimension $m = 8$ presumably represents an attempt to model deterministically the high-dimensional noise. Stark [9] showed that n-dimensional dynamics for which observations are subject to external disturbances with p-dimensional dynamics can be reconstructed in dimension $2(n + p) + 1$. Since this upper bound increases with the dimension p of the external disturbance, in the presence of high-dimensional noise it might not be inappropriate to embed in a dimension as high as possible.

Two periodic systems (the special cases of a quasiperiodic system with parameter a for $a = 0$ and $a = 1$) had known and unambiguously optimal delays for time-delay reconstruction, at least in $m = 2$ dimensions and in the absence of noise. Of the four methods considered—autocovariance, mutual information, cross-validation, and mean local autocovariance—only the autocovariance and mean local autocovariance methods could reliably achieve the known optimal delays. The autocovariance method could more robustly retain the optimal delays in the presence of noise, but, unlike the mean local

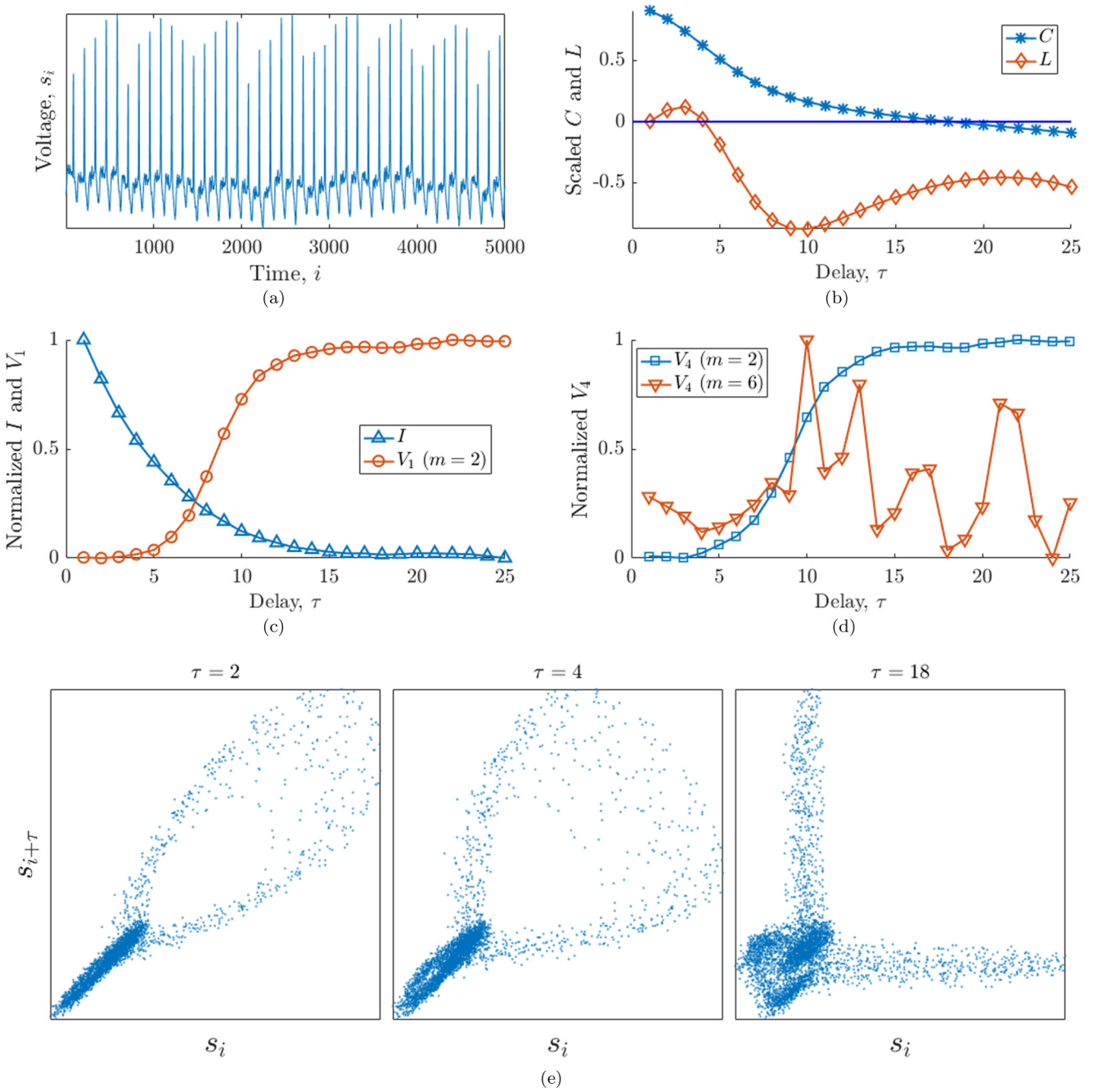


FIG. 8. (a) An ECG time series of length $M = 5\,000$, and the corresponding (b) scaled autocovariance (C) and mean local autocovariance (L), (c) scaled and shifted mutual information (I) and cross-validatory mean-square error with prediction horizon $\lambda = 1$ (V_1) using the error-minimizing reconstruction dimension $m = 2$, and (d) scaled and shifted cross-validatory error with prediction horizon $\lambda = 4$ (V_4) using reconstruction dimension $m = 2$ and the error-minimizing reconstruction-dimension $m = 6$. (e) Time-delay reconstruction in reconstruction dimension $m = 2$ using the delay $\tau = 2$ suggested by cross-validation with prediction horizon $\lambda = 1$, the delay $\tau = 4$ suggested by the mean local autocovariance, and the delay $\tau = 18$ suggested by the autocovariance and mutual information.

autocovariance method, interpolated between the two optimal delays in a somewhat surprising, nonmonotonic way.

For clean data from chaotic flows, whenever the firmly established autocovariance or mutual information methods suggested an unambiguous delay in the interval considered, the mean local autocovariance suggested a similar delay. However, it appeared that reconstruction in $m = 2$ dimensions could sometimes benefit from a delay lower than suggested by these three methods, but often suggested by one of the incarnations of the cross-validatory method. In the presence of varying levels of observational or dynamic noise, the mean local autocovariance method reliably provided reasonable suggestions for the delay with which to reconstruct three chaotic flows. Furthermore, the mean local autocovariance method suggested a useful delay for uniform reconstruction of an ECG time series.

IV. CONCLUSION

Uniform delay reconstruction is a fundamental step in nonlinear time-series analysis. The autocovariance, mutual information, and cross-validatory mean-square error are functions important in choosing the delay used to define a reconstruction of frequently sampled flowlike data. Our investigation of three chaotic flows showed that the mutual information could be used to determine appropriate delays for noiseless systems for which the autocovariance method failed. However, the autocovariance was more robust to observational noise than was the mutual information. Under certain conditions the cross-validatory method is guaranteed to identify appropriate reconstruction parameters in the limit of a large number of observations, and this theoretical result was reflected in advantageous delays suggested for clean observations of the chaotic flows which we considered and also for noisy observations of the Rössler [32] system. For other frequently and noisily sampled chaotic flows which we considered, the cross-validatory method may have required longer time series to yield results which were robust to noise.

We proposed choosing the delay for uniform delay reconstruction of frequently sampled flowlike data as the zero of a function which is easy to define and calculate and which we referred to as the mean local autocovariance. The mean local autocovariance is more robust to noise than is the mutual information, and more versatile than the autocovariance method. Furthermore, the delay identified using the mean local autocovariance revealed structure in an ECG time series more clearly than did the delay reached via autocovariance and mutual information. Not only does the mean local autocovariance provide a means of selecting delays which is versatile and robust to noise, but it also avoids tuning parameters and is simple to implement. Thus the mean local autocovariance method shares advantageous features of the methods of delay selection most widely represented in the nonlinear time-series analysis literature.

ACKNOWLEDGMENTS

We thank the anonymous reviewers for helpful comments. J.M.M. and G.Y. acknowledge the financial support provided by the National Natural Science Foundation of China (Grants No. 11875043 and No. 61803047), Shanghai Science and Technology Committee (Grant No. 18ZR1442000), the Fundamental Research Funds for the Central Universities, and the National Youth 1000 Talents Program of China.

[1] H. Kantz and T. Schreiber, *Nonlinear Time Series Analysis*, Vol. 7 (Cambridge University Press, Cambridge, 2004).

[2] L. Iasemidis, S. Sabesan, N. Chakravarthy, A. Prasad, and K. Tsakalis, Brain dynamics and modeling in epilepsy: Prediction and control studies, in *Complex Dynamics in Physiological Systems: From Heart to Brain* (Springer, Berlin, 2009), pp. 185–214.

[3] C. Letellier, L. A. Aguirre, J. Maquet, and R. Gilmore, Evidence for low dimensional chaos in sunspot cycles, Astron. Astrophys. **449**, 379 (2006).

[4] A. Ord and B. Hobbs, Quantitative measures of deformed rocks: The links to dynamics, J. Struct. Geol. **125**, 74 (2019).

[5] L. Mastroeni, P. Vellucci, and M. Naldi, A reappraisal of the chaotic paradigm for energy commodity prices, Energy Econ. **82**, 167 (2019).

[6] C. A. T. Cortez, J. Coulton, C. Sammut, and S. Saydam, Determining the chaotic behavior of copper prices in the long-term using annual price data, Palgrave Commun. **4**, 8 (2018).

[7] F. Takens, Detecting strange attractors in turbulence, in *Dynamical Systems and Turbulence, Warwick 1980* (Springer, Berlin, 1981), pp. 366–381.

[8] T. Sauer, J. A. Yorke, and M. Casdagli, Embedology, J. Stat. Phys. **65**, 579 (1991).

[9] J. Stark, Delay embeddings for forced systems. I. Deterministic forcing, J. Nonlin. Sci. **9**, 255 (1999).

[10] J. Stark, D. S. Broomhead, M. Davies, and J. Huke, Delay embeddings for forced systems. II. Stochastic forcing, J. Nonlin. Sci. **13**, 519 (2003).

[11] N. H. Packard, J. P. Crutchfield, J. D. Farmer, and R. S. Shaw, Geometry from a Time Series, Phys. Rev. Lett. **45**, 712 (1980).

[12] H. D. I. Abarbanel, *Analysis of Observed Chaotic Data* (Springer-Verlag, Berlin, 1996).

[13] M. Small, *Applied Nonlinear Time Series Analysis: Applications in Physics, Physiology and Finance*, Vol. 52 (World Scientific, Singapore, 2005).

[14] E. Bradley and H. Kantz, Nonlinear time-series analysis revisited, Chaos **25**, 097610 (2015).

[15] A. M. Fraser and H. L. Swinney, Independent coordinates for strange attractors from mutual information, Phys. Rev. A **33**, 1134 (1986).

[16] This approach is not precisely that of Fraser and Swinney [15], but an adaptation which is easier to implement and which seems to have become the default means to use mutual information to choose a uniform time delay [1,12].

[17] S. P. Garcia and J. S. Almeida, Nearest neighbor embedding with different time delays, Phys. Rev. E **71**, 037204 (2005).

[18] L. M. Pecora, L. Moniz, J. Nichols, and T. L. Carroll, A unified approach to attractor reconstruction, Chaos **17**, 013110 (2007).

[19] K. S. Chan and H. Tong, *Chaos: A Statistical Perspective* (Springer Verlag, Berlin, 2001).

[20] B. Cheng and H. Tong, Orthogonal projection, embedding dimension and sample size in chaotic time series from a statistical perspective, Philos. Trans. R. Soc. London A **348**, 325 (1994).

[21] S. Giannerini and R. Rosa, Assessing chaos in time series: Statistical aspects and perspectives, Studies Nonlin. Dynam. Econ. **8**, 11 (2004).

[22] D. Walker, T. Stemler, and M. Small, Time-series network induced subgraph distance as a metonym for dynamical invariants, Europhys. Lett. **124**, 40001 (2018).

[23] K. Harezlak and P. Kasprowski, Searching for chaos evidence in eye movement signals, Entropy **20**, 32 (2018).

[24] A. Santuz, A. Ekizos, N. Eckardt, A. Kibele, and A. Arampatzis, Challenging human locomotion: Stability and modular organisation in unsteady conditions, Sci. Rep. **8**, 2740 (2018).

[25] V. J. D. Vieira, S. C. Costa, S. L. Correia, L. W. Lopes, W. C. d. A. Costa, and F. M. de Assis, Exploiting nonlinearity of the speech production system for voice disorder assessment by recurrence quantification analysis, Chaos **28**, 085709 (2018).

[26] D. Panagoulia and E. I. Vlahogianni, Recurrence quantification analysis of extremes of maximum and minimum temperature patterns for different climate scenarios in the Mesochora catchment in Central-Western Greece, Atmos. Res. **205**, 33 (2018).

[27] Z. Zhang and M. Grabchak, Bias adjustment for a nonparametric entropy estimator, Entropy **15**, 1999 (2013).

[28] H. Shimazaki and S. Shinomoto, A method for selecting the bin size of a time histogram, Neural Comput. **19**, 1503 (2007).

[29] The sampling, resolution, and voltage information were available at https://physionet.org/physiobank/database/chfdb/.

[30] D. S. Baim, W. S. Colucci, E. S. Monrad, H. S. Smith, R. F. Wright, A. Lanoue, D. F. Gauthier, B. J. Ransil, W. Grossman, and E. Braunwald, Survival of patients with severe congestive heart failure treated with oral milrinone, J. Am. Coll. Cardiol. **7**, 661 (1986).

[31] A. L. Goldberger, L. A. N. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng, and H. E. Stanley, PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals, Circulation **101**, e215 (2000).

[32] O. E. Rössler, An equation for continuous chaos, Phys. Lett. A **57**, 397 (1976).

[33] E. N. Lorenz, Deterministic nonperiodic flow, J. Atmos. Sci. **20**, 130 (1963).

[34] D. W. Moore and E. A. Spiegel, A thermally excited nonlinear oscillator, Astrophys. J. **143**, 871 (1966).

[35] J. C. Sprott, *Elegant Chaos: Algebraically Simple Chaotic Flows* (World Scientific, Singapore, 2010).

[36] L. F. Shampine and M. W. Reichelt, The MATLAB ODE suite, SIAM J. Sci. Comput. **18**, 1 (1997).

[37] The reconstruction dimension $m = 2$ is too low for an embedding of any of the systems herein considered but is useful for illustration.

[38] J. Stark, Delay reconstruction: Dynamics versus statistics, in *Nonlinear Dynamics and Statistics*, edited by A. I. Mees (Springer, Berlin, 2001), pp. 81–103.