Data-based prediction and causality inference of nonlinear dynamics

Huanfei Ma¹, Siyang Leng² & Luonan Chen²,³,*

¹School of Mathematical Sciences, Soochow University, Suzhou 215006, China; ²Institute of Industrial Science, The University of Tokyo, Tokyo 153-8505, Japan; ³Key Laboratory of Systems Biology, Innovation Center for Cell Signaling Network, Institute of Biochemistry and Cell Biology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai 200031, China

Email: hfma@suda.edu.cn, syleng@sat.t.u-tokyo.ac.jp, lnchen@sibs.ac.cn

Received August 15, 2017; accepted October 27, 2017; published online January 12, 2018

Abstract Natural systems are typically nonlinear and complex, and it is of great interest to be able to reconstruct a system in order to understand its mechanism, which cannot only recover nonlinear behaviors but also predict future dynamics. Due to the advances of modern technology, big data becomes increasingly accessible and consequently the problem of reconstructing systems from measured data or time series plays a central role in many scientific disciplines. In recent decades, nonlinear methods rooted in state space reconstruction have been developed, and they do not assume any model equations but can recover the dynamics purely from the measured time series data. In this review, the development of state space reconstruction techniques will be introduced and the recent advances in systems prediction and causality inference using state space reconstruction will be presented. Particularly, the cutting-edge method to deal with short-term time series data will be focused on. Finally, the advantages as well as the remaining problems in this field are discussed.

Keywords nonlinear system, prediction, causality inference, time series data

MSC(2010) 92B05, 57R40, 37D45

Citation: Ma H F, Leng S Y, Chen L N. Data-based prediction and causality inference of nonlinear dynamics. Sci China Math, 2018, 61: 403–420, https://doi.org/10.1007/s11425-017-9177-0

1 Introduction

In the big data era, various types of time-series data are increasingly accumulated every day in various disciplines and in various scopes (see [14,18,56,75,87]). An outstanding problem is how to make system-level study based on the vast amount of data. Particularly, based on the measured/observed time series data, how to reconstruct the system and further study the mechanism behind the phenomenon has become an attracting problem and a hot topic in interdisciplinary science. Despite of the widely used approaches employing linear models (see [29]), which are mainly based on correlation and regression, the natural systems usually show nonlinear nature: the components within one system intertwine with each other, and the dynamics of each component cannot be separated from one another. Such characteristic property distinguishes nonlinear systems from the linear stochastic systems, and nonlinear methods are thus more favorable in reconstructing complex systems.
Since the embedding theory proposed by Takens in 1980s (see [93]), the state space reconstruction (SSR) technique has been developed and applied in nonlinear time-series analysis. The key idea of the SSR stems from the non-separability of the nonlinear system. The property of non-separability implies that the dynamics of one observable variable could contain all the dynamical information of the whole system. As a model-free method, the SSR based methods have been attracting persistent interest and under rapid development in the recent decades (see [30, 42]). In this review, we will not cover the whole topics of system reconstructions using SSR methods, which can be referred to some excellent reviews (see [30, 34, 42, 82, 97]), but focus on the forefront of the recent advances in predicting future dynamics and inferring causality using SSR framework. Particularly, the majority of SSR based methods usually require sufficiently long time-series data or a large number of samples, while the real data sets always have limited length due to the experimental restrictions or costs. We will introduce in detail the recently developed methods aiming to deal with short-term time series data.

This review is organized in the following manner. In Section 2, we first introduce the principle of various embedding theorems, which constitute the theoretical foundation of SSR. Based on the embedding theory, the framework of SSR as well as several related aspects in real applications is presented. In Section 3, the developments of predicting nonlinear system using SSR in the recent decades are reviewed. Particularly, how to make predictions from short-term high-dimensional time series data will be focused and introduced in detail. Section 4 presents the directional interaction detection methods, i.e., the causality inference methods using SSR framework. Both the recently proposed convergent cross mapping (CCM) method, using convergent of mutual predictions, and cross map smoothness (CMS) method, particularly designed for short-term data, are introduced in detail. We also speculate on several open problems, for example, conditional causality and causal relations in delayed systems in this section. Conclusions and discussions can be found in Section 5.

2 State space reconstruction

2.1 Embedding theory

Consider a general dynamical system in the form of

\[
\frac{dx}{dt} = F(x),
\]

(2.1)

where \(x = [x_1, x_2, \ldots, x_n]^T\) is the system’s state vector and \(F = [f_1, f_2, \ldots, f_n]^T\) with each \(f_i(\cdot)\) a nonlinear function. Starting from the initial value \(x(0) = x_0\), the trajectory \(x(t)\) describes the evolution of the system along time. After transient dynamics, the system’s bounded trajectory will converge to an attractor, i.e., a bounded and invariant set \(A\). Here, the diverging situations are not under consideration.

For a discrete system, the dynamical system is in the form of

\[
x(t + 1) = F(x(t)),
\]

(2.2)

where \(F\) is a nonlinear map, and \(x(t), t = 0, 1, 2, \ldots\) is the trajectory of the system, and the attractor can be defined analogously.

We consider the dynamics constrained in a compact manifold \(M\) containing the attractor \(A\). For both continuous and discrete cases, the time evolution corresponding with an initial value \(x_0 \in M\) can be denoted uniformly by \(x(t) = F_t(x_0)\): in the discrete case time \(t \in \mathbb{N}\) and \(F_t = (F)^t\), while in the continuous case time \(t \in \mathbb{R}\) and \(F_t(x_0)\) is the integral curve through \(x_0\).

Intuitively, to study the property of the attractor in the state space, the values of all the system’s state components \(x_i(t), i = 1, 2, \ldots, n\) along time should be known. However, the embedding theory implies that from a few or even one single time series observed from the system, it is possible to study the dynamical property of the whole system, i.e., the system can be reconstructed by observations. Actually, it was first discovered by Packard et al. [69] that using only one single coordinate of Rössler system one can reconstruct a faithful state-space representation of the dynamics in the original 3-dimensional space.
To be specific, the dynamical system of interest in [69] can be expressed in the form of (2.1) as

\[
\frac{dx}{dt} = -(y + z), \\
\frac{dy}{dt} = x + 0.2y, \\
\frac{dz}{dt} = 0.4 + xz - 5.7z,
\]

and with time series of all the state components, the system’s attractor can be obtained, as shown in Figure 1(a) under \(x - y - z\) coordinates. As suggested by Packard et al. [69], by sampling the time series of only single coordinate \(x(t)\), a variety of three independent coordinate sets can be obtained such as \([x(t), x(t - \tau), x(t - 2\tau)]\) and \([x(t), \dot{x}(t), \ddot{x}(t)]\), where \(\tau\) is a sampling interval and \(\dot{x}\) and \(\ddot{x}\) are the first order and second order differences, respectively. The attractors in such coordinates are shown in Figures 1(b) and 1(c), respectively. The geometry property of the attractors in Figures 1(b) and 1(c) is quite similar to that in Figure 1(a). As a matter of fact, the dimensionality and the positive characteristic exponent can be further computed for each of these attractors and the results confirm that the attractors obtained in Figures 1(b) and 1(c) can be regarded as reconstructions of the attractor obtained in Figure 1(a).

Such discovery was theoretically confirmed by Takens in his seminal work [93], in which the famous Takens embedding theory was proposed.

**Theorem 2.1** (Takens embedding theory I). Let \(\mathcal{M}\) be a compact manifold of dimension \(d\). For pairs \((\phi, h)\), \(\phi : \mathcal{M} \to \mathcal{M}\) a smooth diffeomorphism and \(h : \mathcal{M} \to \mathbb{R}\) a smooth function, it is a generic property that the map \(\psi_{\phi, h} : \mathcal{M} \to \mathbb{R}^L\), defined by \(\psi_{\phi, h}(x) = (h(x), h(\phi(x)), \ldots, h(\phi^{L-1}(x)))\) is an embedding where the integer \(L = 2d + 1\).

Here, an embedding of \(\mathcal{M}\) means a smooth diffeomorphism from \(\mathcal{M}\) onto its image \(\psi(\mathcal{M})\), i.e., a smooth one-to-one map which has a smooth inverse. The number \(L = 2d + 1\) is the dimension of the space for the reconstructed attractor.

Actually, the embedding theory could be stated in other forms, such as derivative embedding and generalized embedding theorems (see [17, 93, 99]) stated as follows.

**Theorem 2.2** (Takens embedding theory II). Let \(\mathcal{M}\) be a compact manifold of dimension \(d\). For pairs \((\phi, h)\), \(\phi : \mathcal{M} \to \mathcal{M}\) a smooth diffeomorphism and \(h : \mathcal{M} \to \mathbb{R}\) a smooth function, it is a generic property that the map \(\psi_{\phi, h} : \mathcal{M} \to \mathbb{R}^L\), defined by \(\psi_{\phi, h}(x) = (h(x), \frac{d}{dx}, h(\phi(x)), \ldots, \frac{d^{L-1}}{dx^{L-1}} h(\phi(x)))\) is an embedding where the integer \(L = 2d + 1\).

**Theorem 2.3** (Generalized embedding theory). Let \(\mathcal{M}\) be a compact manifold of dimension \(d\) and a set of \(L\) observation functions \(\{h_1, \ldots, h_L\}\), where \(h_k : \mathcal{M} \to \mathbb{R}\) is a smooth function and \(L = 2d + 1\). Then it is a generic property of all possible \(\{h_k\}\) that the map \(\psi_{\phi, h} : \mathcal{M} \to \mathbb{R}^L\), defined by \(\psi_{(h_k)}(x) = (h_1(x), h_2(x), \ldots, h_L(x))\) is an embedding.

It is clear that here the dimension \(d\) is important for the embedding theory. However, the dimension \(d\) is usually much higher than the essential dimension of the attractor \(A\). Therefore, the embedding theory was later generalized into fractal dimensions by Sauer et al. [78]. Let \(d_A\) be the box-counting dimension of the attractor \(A\), which could be a non-integer and could be much smaller than the integer dimension \(d\).
of the manifold $\mathcal{M}$ containing $\mathcal{A}$. Then for almost every smooth measurement function $h$, the maps defined in the Takens embedding theorems are still an embedding on each compact subset of a smooth manifold contained within $\mathcal{A}$ with $L > 2d_A$.

Particularly, starting from one dimensional observed time series, a delay coordinate map could be introduced.

**Definition 2.4 (Delay coordinate).** If $\phi$ is a flow on a manifold $\mathcal{M}$, $\tau$ is a positive delay, and $h : \mathcal{M} \to \mathbb{R}$ is a smooth function, define the delay coordinate map $F(h, \phi, \tau) : \mathcal{M} \to \mathbb{R}^L$ by

$$F(h, \phi, \tau)(x) = (h(x), h(\phi_{-\tau}(x)), h(\phi_{-2\tau}(x)), \ldots, h(\phi_{-(L-1)\tau}(x))).$$

According to the fractal embedding theorem, the delay coordinate map is generically an embedding as long as $L > 2d_A$. Thus it is straightforward to see that when the observation function $h$ is simply chosen as the coordinate function $x_i$, the delay coordinate map $x(t) \to [x_i(t), x_i(t-\tau), \ldots, x_i(t-(L-1)\tau)]$ with sampling interval $\tau$ is generically an embedding.

The one-to-one property of embedding is essential to the reconstruction of the original system because the state of a deterministic dynamical system, as well as its future evolution, is completely specified by a point in the state space. Therefore, the delay coordinate map actually sets up a bridge between the dynamics of the original system and the dynamics in the reconstructed space. Thus the delay coordinate $[x_i(t), x_i(t-\tau), \ldots, x_i(t-(L-1)\tau)]$ is a state space reconstruction of the original system, which implies that the one dimensional observed time series can contain all the information of the dynamics of the whole original system. Moreover, the embedding theory implies that the dynamics in the reconstructed space is topologically conjugated with the dynamics in the original state space (see [42]). Thus, the nonlinear dynamic properties such as Lyapunov exponent, entropy, and even the unstable periodic orbit (UPO) are preserved from the original system to the reconstructed system, which forms the cornerstone of the state space reconstruction technique.

### 2.2 Issues related to state space reconstruction

When applying delay coordinate SSR to real data sets, the information of the original nonlinear system is generally unknown a priori, and thus there are two key factors to be determined for successful SSR application: one is the embedding dimension $L$ and the other is the delay $\tau$. Dozens of methods for estimating optimal dimension $L$ and delay $\tau$ have been developed in the past few decades based on time series data (see [8,21,25,43,44,46,47,50] and [51,70,77,94]).

Theoretically, any $\tau > 0$ could be used in the delay coordinate SSR process, but in practice, the delay coordinates will be strongly correlated with too small $\tau$ while the reconstructed attractor can fold over on itself with too large $\tau$. Hence, one wants the smallest $\tau$ that maximizes the independence of the coordinates of the embedding vector. Therefore, several representative methods have been proposed to choose optimal $\tau$ by computing a statistic that measures the independence of $\tau$-separated points in the time series. As $\tau$ increases, the first zero of the autocorrelation function of the time series, the first minima of the average mutual information, or that of the correlation sum (see [21,25,51]) could be candidates of optimal $\tau$. Actually, such $\tau$ maximizes general forms of independence, and consequently can be adopted in the delay coordinate.

Analogous to the case of choosing $\tau$, the embedding dimension $L$ should be neither too small nor too large. In theory, $L$ should be sufficiently large so as to satisfy the condition of $L > 2d_A$; however in practice, low dimensional reconstruction is more favorable. On the other hand, it is noticed that $L > 2d_A$ is a sufficient but not necessary condition. Thus the optimal $L$ should be chosen as the smallest value that affords a topologically correct result. To this, the false near neighbor (FNN) algorithm by Kennel et al. [44] is a feasible method generally accepted. In an FNN-based method, one tries to construct a delay coordinate embedding with embedding dimension $k$, computes each point’s near neighbors, and repeats such process with increased embedding dimension $k+1$. If any of the geometry relations changes, i.e., some neighbor in $k$ dimension is no longer a neighbor in $k+1$ dimension, then it is a false near neighbor.
and it indicates that the dynamics are not properly unfolded under embedding dimension $L = k$. Thus, the value appearing on the elbow of the decreasing FNN curve as $k$ increases can be a candidate for optimal embedding dimension.

In the recent years, it has been noted that the product $L \times \tau$ is the factor that really matters. Thus the both factors can be estimated at the same time and a unified approach has been proposed (see [70]).

Furthermore, when applying SSR to practical measured time series, noise is unavoidable. The theory of state space reconstruction in the presence of noise was first considered in [11], which incorporated the observational noise and estimation error into the framework. Delay coordinate embedding theory for stochastic systems was later systematically studied by Stark et al. [85, 86].

3 Prediction of nonlinear dynamics

Predicting nonlinear dynamics using state space reconstruction method has a long history, which can be dated back to 1969 when Lorenz [58] proposed the method of analogues in the state space to forecast atmosphere. Later the idea was shown to be applicable in the reconstructed state spaces by Pikovsky [72].

The nonlinear but smooth vector field or map in the state space implies that the evolution trajectories starting from two close states will still be close to each other in short time, and thus the information of unknown future value can be approximated by the future values of its near neighbors, as sketched in Figure 2(a). The embedding theory guarantees that such property also holds for states in delay coordinate reconstructed state space. This has inspired the development of making predictions for nonlinear time series using nearest neighbors.

3.1 Nearest neighbors methods

The fact that close states will evolve to close future states implies that using near neighbors on the attractor can patch local information to approximate the true future values. Based on such local patching idea, a large number of creative strategies have been fruitfully proposed to predict future dynamics, including several influential works (see [10, 19, 90]). The research on this topic has even spawned a competition on nonlinear time series prediction (see [98]).

To be specific, considering a state $\mathbf{x}(t) \in \mathbb{R}^d$ on the attractor $\mathcal{A}$ and its $k$ nearest neighbors $\mathbf{x}(t_i), i = 1, 2, \ldots, k$ where $t_i < t$, the future value $\mathbf{x}(t + \tau)$ can be locally obtained by the values of $\mathbf{x}(t_i + \tau)$. In [19], the zeroth-order approximation method was proposed where $\mathbf{x}(t + \tau)$ was simply approximated by $\mathbf{x}(t' + \tau)$, i.e.,

$$\mathbf{x}(t + \tau) \approx \mathbf{x}(t' + \tau), \quad (3.1)$$

where $\mathbf{x}(t')$ was the nearest neighbor of $\mathbf{x}(t)$ on the attractor $\mathcal{A}$. In [10], the relation

$$\mathbf{x}(t + \tau) = f(\mathbf{x}(t)) \quad (3.2)$$
was assumed to be a polynomial and fitted by the $k$-nearest neighbors $x(t_i + \tau) = f(x(t_i)), i = 1, 2, \ldots, k$. The simplex method was proposed in [90], where $x(t + \tau)$ was approximated by the weighted average of its nearest neighbors’ future values in the form of

$$x(t + \tau) = \sum_{i=1}^{k} \omega_i x(t_i + \tau), \quad (3.3)$$

where $\omega_i$ was set as an exponential weight to its original distance $\|x(t) - x(t_i)\|$ and the number of nearest neighbors $k$ was generally set to be larger than the reconstructed dimension $d$ such that $k > d$. When the future value $x(t + \tau)$ is obtained, the prediction of each component $x_i(t + \tau)$ is realized. Particularly, in the case of delay coordinate reconstruction, with the time series data of one dimensional observed variable $x(t)$, the one step prediction of $x(t + \tau)$ is realized, which is the first component of the reconstructed state vector $x(t) = [x(t), x(t - \tau), \ldots, x(t - (L - 1)\tau)]$ in the reconstructed space.

These methods have inspired many data-driven analysis and have been fruitfully reported in the study of multi-disciplines from population study to ecological and biological data even to physiological data and climate data (see [5,15,20,26,28,30,37,40], [42,45,57,66,81,83,89,95] and [98,101,104]).

The research of using local information provided by nearest neighbors to make predictions has been active and attracting even to this day. In order to leverage information from multiple time series observations and overcome the curse of dimensionality, Ye and Sugihara [104] proposed a new method in 2016 with name multiple view embedding (MVE) which was claimed to be able to make predictions for noisy high-dimensional systems. According to the generalized embedding theorem stated in Theorem 2.3, with multiple time series observations composed of $n$ variables and considering $l$ lags for each variable, as many as $m$ possible reconstructed attractors could be constructed by $L$-dimensional variable combinations. Here, the number of possible reconstructions could be calculated as $m = \binom{n}{L} - \binom{n(l-1)}{L}$, where $(\binom{n}{L})$ is the number of possible combinations picking $L$ out of $nl$ coordinates, and it excludes the “invalid” reconstructions where all $L$ coordinates have positive lags with number $(\binom{n(l-1)}{L})$.

For a specific variable $y$, each reconstruction could be used to make a one-step prediction using nearest neighbors in the way $y_{t+1} \approx y_{nn^i(t)+1}$, where $nn^i(t)$ is the time index of the nearest neighbor in the $i$-th reconstructed attractor. Though each single prediction may have reduced precision due to the limited data length and noise, the number of predictions could be fairly large. The MVE method firstly ranks all these predictions according to in-sample forecast accuracy, and then the MVE forecast is defined as a simple average of the top $k$ reconstructions in the way

$$\hat{y}_{t+1} = \frac{1}{k} \sum_{i=1}^{k} y_{nn^i(t)+1}, \quad (3.4)$$

where $k$ is empirically set to $k = \sqrt{m}$.

The base of information leverage by MVE follows from two aspects. Large number of attractors could be reconstructed due to high-dimensional time series data, and each reconstructed attractor filters the dynamical information of the whole system in a different way. Therefore, combining the information from multiple reconstructed attractors could be advantageous in the manner of (3.4).

Despite of the rapid development of the nearest neighbors based methods in the recent years, one issue on whether nearest neighbors really reflect the local information near the true target is essential to the success of applying such methods. In Figures 2(b) and 2(c), two attractors are reconstructed for the same system by different lengths of observed time series. Obviously, when the observed time series is not long enough, the points on the reconstructed attractor are sparse and thus the nearest neighbors actually cannot provide local information of future dynamics. Thus, in order to get the true nearest neighbors, the attractor is necessary to be sufficiently sampled and the time series should be long enough. However, real data sets always have a limited length due to the experiments restriction or cost. Therefore, such classic methods are generally not suitable to be used in dealing with short-term data.

On the other hand, the short-term data are usually accompanied with high-dimensional measurements, i.e., the number of state variables $n$ is much larger than the length of measurements $m$. Representative
Figure 3 Illustrations of information in different types of data sets, from [61] with permission. (a) A short-term low-dimensional time series has insufficient information to reconstruct the attractor for making predictions. (b) A long-term low-dimensional time series has sufficient information to reconstruct the attractor based on the embedding theory. Many existing methods are based on such a scheme. (c) A short-term high-dimensional time series also contains sufficient information, similar to the long low-dimensional time series, which can be used to reconstruct dynamics or the attractor and thereby make predictions on the time series for some selected variables.

Examples of such data are high-throughput biological data, e.g., microarray data and gene sequencing data, where tens of thousands of gene probes can be measured simultaneously, but only some tens or even fewer successive time points can be measured for expression of each gene (see [73]) due to various restrictions. Collective dynamics of different complex networks (see [88]) such as social networks and biological networks also show high-dimensional but short time course features, where a large number of coupled dynamical systems evolve via various topological structures but only limited time points can be recorded. As a matter of fact, such kind of high-dimensional biological data has emerged in the big data era, from microscopic gene expression data to mesoscopic neural activity data, even to macroscopic ecological and atmosphere data (see [14, 18, 35, 56, 87]).

In such high-dimensional nonlinear systems, the variables are not independent but intertwined with each other, and the interactions or correlations between them contain rich information contents, which may reflect the accumulated dynamical features of the target system, as illustrated in Figure 3.

3.2 Inverse embedding method

Prediction of short-term high-dimensional data was first addressed in [61], in which Ma et al. proposed
a new SSR based method aiming to transform the spacial information into a temporal domain and thus made it possible to predict short-term high-dimensional data in a nonlinear manner. As illustrated in Figure 4(a), the classic SSR methods start with a single long-term time series and reconstruct a shorter but higher dimensional time series using delay coordinates. By contrast, the method developed in [61] uses the embedding theory in the inverse way, i.e., starting with short high-dimensional time series and recovering the corresponding longer low-dimensional time series. Theoretically, this new idea can be viewed as an inverse direction of the embedding theory and thus transfers the spacial/interaction information into temporal domain. The process of the inverse embedding method is sketched in Figure 4(b), and the mapping $\Phi$ between the original high dimensional attractor $A$ and the reconstructed attractor $\mathcal{M}$ using delayed coordinate, as the key to the method, can be formulated in the form as shown in Figure 5, where the shaded zones stand for the unknown/predicted values.

Figure 4 Basic ideas of inverse embedding, from [61] with permission. (a) Transformation between short-term high-dimensional time series and long-term scalar (or low-dimensional) time series. (b) Sketch of inverse embedding method to predict the dynamics based on short-term high-dimensional time series data.
A simple way to construct the map $\Phi$ is solving it as a linear map, which can be obtained in a very efficient way and can be applied to the prediction of many real systems. However, generally the map $\Phi$ is an unknown nonlinear vector function and the constraints are limited according to Figure 5. In order to derive the map in an accurate manner so that the transformation can be carried out, it is firstly expanded into a basis functional space, i.e., the power expansion, the RBF (Radial Basis Function) network, or the wavelet expansion. Considering the short-term property, the training problem will be generally categorized into an underdetermined problem where there are more coefficients to be determined than the number of independent equations. Note that in many situations, especially with a proper choice of basis functional space, the coefficients in the expanding series can be sparse, and thus the recently developed idea of compressive sensing (see [9]) can be used to solve this problem.

Using the compressive sensing technique, an algorithm to solve the map $\Phi$ as well as to make predictions was proposed in [61], which was shown to be particularly effective for short-term high-dimensional biological data sets, for instance, microarray data or RNA-seq data, where time points as long as order $O(10)$ were available while thousands of probes were measured and several successful predictions could be made using inverse embedding method. The method has been further considered in the process of big biological data (see [49]) and predicting early-warning signals of critical transitions (see [53]).

To compare the inverse embedding method and nearest neighbors based method, Figure 6 shows a benchmark test with both methods. A coupled Lorenz system with 15 variables is considered and two
criteria for prediction accuracy are depicted, i.e., the correlation $\rho$ and the rooted mean squared error (RMSE) between the test series and the predicted series. In Figures 6(a) and 6(b), it is clearly shown that the inverse embedding method can reach quite a good prediction accuracy with training data of less than 100 time points, while the nearest neighbor based method is still in the convergence process even with length of 400. It validates that the inverse embedding method avoids finding nearest neighbors, and thus can deal with short-term data. In Figures 6(c) and 6(d), the robustness of the prediction methods against noise deterioration is also tested. It is shown that the nearest neighbors based method is quite robust against noise, which is mainly attributed to the averaging process of several nearest neighbors. However, compared with the nearest neighbors based method, the inverse embedding method is sensitive to the noise. This is mainly caused by the compressive sensing algorithm adopted in [61], which may enlarge the noise in the process of making parameters sparse. As a matter of fact, though the inverse embedding method transforms the interactions or correlations between variables as spacial information into temporal domain to make predictions, such information actually has redundancy, which is also confirmed in [104]. Therefore, how to dig out such rich redundant information to combat noise could be future work for the inverse embedding method.

4 Causality detection for nonlinear systems

4.1 Causal interactions

The interactions between variables within one complex system are essential to understand the mechanism behind the experimental phenomenon, and thus are the foundation of mathematical modelling in a system level. Particularly, data driven interaction inference is in an urgent need in order to model various levels of complex networks. The simplest measure of interaction is the correlation between time series data from two variables $x$ and $y$. However, correlation only provides the concomitant covariation between $x$ and $y$ and it has been shown that even a perfect correlation between two observed variables is neither necessary nor sufficient to imply a causal relationship (see [91]), especially when it is recognized that nonlinear dynamics are ubiquitous. Thus, compared to non-directional relationship as correlation, the directional interaction inference is more desired.

Intuitively, the directional interaction from $x$ to $y$ could be understood as the causal effect from the dynamics of $x$ to the dynamics of $y$, and thus the concept of causality can be introduced. Though there still has been no universally accepted definition of causality, various measures of causality or causal relations have been reported and extensively studied. Among them the Granger causality (GC), proposed by Granger [23,24], the 2003 Nobel laureate in economy, is undoubtedly a widely accepted definition and it is also a framework to detect causal relationship from time series data. In the sense of Granger causality, the definition of causality comes from the idea by Wiener in 1956 (see [100]): “For two simultaneously measured signals, if we can predict the first signal better by using the past information from the second one than by using the information without it, then we call the second signal causal to the first one”. The standard test of GC developed by Granger is based on a linear regression model and later generalized into nonlinear regression (see [1,63]). Moreover, the original Granger causality is defined in temporal domain, and the concept could be transformed into frequency domain (see [7,27,41]). The Granger causality has been widely accepted in multiple disciplines, especially in economical fields and neural signal analysis, and fruitful applications have been reported in the past decades (see [4,22]). However, one key assumption in Granger causality is the separation of effect from one variable to another, i.e., the effect of one variable could be separated and removed from the dynamics of the other, which is generally not true in real situations. Particularly, in complex systems when nonlinearity is dominated, every variable could carry all the system’s information, as indicated by Takens’ embedding theory, and thus the nonlinear systems are generally non-separative (see [91]). Moreover, the Granger causality is defined for stochastic process, and could not be applied to deterministic systems directly.
From the viewpoint of information-flow, when $x$ drives $y$, the dynamics information of $x$ is injected into the dynamics of $y$, and thus if one can discover such footprint, the driving effect from $x$ to $y$ could be detected. Based on this idea, the state space reconstruction framework provides another choice to consider causality in a purely nonlinear style. Following the embedding theorems, each attractor reconstructed by SSR from delayed coordinates of univariate gives one-to-one relation to the original system’s attractor, and thus the two attractors reconstructed by $x$ and $y$, respectively are also one-to-one mapped with each other. Explicitly, such one-to-one map can be set up by the time index, i.e., $\phi : x(t) \leftrightarrow y(t)$, $x(t) \in M_x, y(t) \in M_y$, where $t$ is the time index as shown in Figure 7(a), and such map $\phi$ is named as a cross map. Moreover, for one particular point $y(t_0) \in M_y$ and the counterpart $x(t_0) \in M_x$, the local neighborhood of $y(t_0)$ can be defined on the attractor as $N_\epsilon(y(t_0)) = \{y(y_{t_0}) : \|y(t_{t_0}) - y(t_0)\| < \epsilon\}$ and the mutual neighbors of $x(t_0)$ could be further defined as $\{x(t_{y_{t_0}}) : y(t_{y_{t_0}}) \in N_\epsilon(y(t_0))\}$, as illustrated in Figure 7(a). The mutual neighbors of $y(t_0)$ could be analogously defined as shown in Figure 7(b).

Intuitively, if $x$ and $y$ are causally linked, the mutual neighbors will have some specific properties due to the information injection. As a matter of fact, in the early development of SSR based causality detection, there are two controversial explanations on how to use such property to define the direction of causality.

In [48], Le Van Quyen et al. claimed that if $x$ drives $y$ but $y$ has no driving force to $x$, there will be a generalized synchronization relationship between the time series of $x(t)$ and $y(t)$ such that $y = F(x)$ but not vice versa (e.g., no injectivity). Therefore, using the next step of mutual neighbors of $y(t_n) \in M_y$, one can make one-step prediction for $y(t_{n+1})$ in the way $y(t_{n+1}) = \frac{1}{N} \sum_{j: x(t_{jn}) \in N_\epsilon(x(t_{n+1}))} y(t_{j_{n+1}})$, where $N$ is the number of nearest neighbors in the $\epsilon$-neighborhood. In such a manner, the cross prediction error $\text{Pred}_{x \rightarrow y} = \sum \|y(t_{n+1}) - y(t_{n+1})\|$ can be used to distinguish the causal relation between $x$ and $y$: if the cross prediction error from $x$ to $y$ is much smaller than the inverse direction, i.e., $\text{Pred}_{x \rightarrow y} \ll \text{Pred}_{y \rightarrow x}$, strong evidence of unidirectional causality $X \rightarrow Y$ is implied.

On the other hand, Schiff et al. [79] proposed another explanation. If $x$ and $y$ are unidirectionally...
coupled, i.e., information flows only from \( x \) to \( y \), the reconstructed attractor \( \mathcal{M}_x \) will present the dynamics in \( x \) alone, while the attractor \( \mathcal{M}_y \) reconstructed by \( y \) will represent the combined dynamics of both \( x \) and \( y \). Thus, \( x \) will be predictable by cross prediction using mutual neighbors. Obviously, this is completely contradictory in interpreting the differences in cross prediction errors.

This issue has not been resolved until Sugihara et al. [91] proposed the convergent cross mapping (CCM) method. In [91], the convergence of cross prediction error was firstly stressed as the length of time series data increases, and it was clearly shown that convergence was essential to distinguish the direction of causal link. If the time series data is not long enough, the variance of the cross prediction errors will be quite large and the confidence area of such errors for both directions would overlap with each other, so that it is not possible to distinguish the two directions, as illustrated in [91, Figure S5]. As the length of time series data increases, the reconstructed manifold becomes denser and the highly resolved attractor improves the accuracy of prediction based on nearest neighbors, and thus convergence is observed. With the concept of convergence, both theoretical analysis and benchmark test confirm that the direction indicated in [79] is correct, i.e., if \( x \) drives \( y \), then one can use the information of \( y \) to estimate \( x \), and vice versa.

Based on the SSR causality detection framework, a lot of applications have been fruitfully reported in multiple disciplines, especially in neurophysiological signal processing, electroencephalographic (EEG) signal processing, brain network reconstruction, ecological data analysis, cardiorespiratory data analysis, and even climate change study (see [2,3,6,13,16,31,34,36], [38,39,52,64,67,68,71,74] and [84,92,96,102]).

### 4.3 Cross map smoothness method

The convergence concept revealed in [91] actually coincides with the argument proposed in [59,61] as shown in Figures 2(b) and 2(c), i.e., using nearest neighbors on the reconstructed attractor as local approximations essentially relies on the density of points on the reconstructed attractor. Only sufficiently long-term time series can bring faithfully reconstructed attractor with dense points on it and consequently the nearest neighbors can be used as local approximations. As shown in [91, Figure S5], to realize convergence, a typical length of order \( O(10^4) \) was required for a benchmark coupled nonlinear system. Such requirement is usually not practical for typical biological data, e.g., high throughput microarray or RNA-seq data for gene expressions of a biological process. Moreover, on some occasions even though long-term data can be measured, only short (recent) pieces can correctly reflect the causal relation between subsystems due to the nonstationary and fast switching property of the concerned systems. Therefore, it is in urgent need of developing new methods to detect causality based on short-term data or a small number of samples.

In [59], Ma et al. proposed a new method named cross map smoothness (CMS) which was also based on the SSR technique but was able to infer nonlinear causality from short-term time series data. The base of CMS method lies on the observation that the nearest neighbor based methods actually measure the cross map’s local property around each point on the reconstructed attractor, leaving the global information unused. To be specific, if \( x \) causally influences \( y \), the mutual neighbors of \( \mathbf{x}(t_0) \) can be used to predict the next value of \( \mathbf{x} \), implying that the nearest neighbors of \( \mathbf{y}(t_0) \) are mapped to close states of \( \mathbf{x}(t_0) \), i.e., the cross map \( \Phi_{yx} : \mathcal{M}_y \rightarrow \mathcal{M}_x \) is locally smooth around \( \mathbf{y}(t_0) \), as shown in Figure 7(a). On the reverse direction, if \( y \) has no influence over \( x \), the image of \( \mathbf{x}(t_0)’s \) neighborhood under the cross map \( \Phi_{xy} : \mathcal{M}_x \rightarrow \mathcal{M}_y \) is not necessarily the neighborhood of \( \mathbf{y}(t_0) \), and thus the cross map \( \Phi_{xy} \) is not necessarily smooth around \( \mathbf{x}(t_0) \), as shown in Figure 7(b). If the cross map is locally smooth around each point on the attractor, it is further globally smooth on the whole attractor, and vice versa, as illustrated in Figures 7(c) and 7(d). Thus, the global smoothness of \( \Phi_{yx} \) and \( \Phi_{xy} \) can be built from local properties. Moreover, as the coupling strength increases, information becomes more distinct between the causally influenced variables. As a result, the reconstructed attractors will contain stronger historical information from the causes. Thus, finding mutual nearest neighbors is equivalent to measuring the smoothness of the cross map, and the relative smoothness indicates the relative strength of causative effectiveness, i.e., the smoothness of \( \Phi_{yx} \) indicates the strength of causative effectiveness from \( x \) to \( y \).
Figure 8  Sketch of the cross map smoothness learned by a neural network (NN), from [59] with permission. (a) and (b) Illustrations for the neural network’s approximation ability for smooth map and unsmooth map. (c) and (d) The prediction error (or the smoothness of $\Phi$) for cases in (a) and (b), respectively, where the leave-one-out scheme is used to calculate errors. (e) Assume that $x$ causally influences $y$, the information of $x$ has been encoded in $M_y$ and consequently $\Phi : M_y \rightarrow M_x$ maps a neighborhood of $y$ to a neighborhood of $x$, implying $\Phi_{yx}$ is smooth. Thus a neural network $N$ can be trained to approximate the map based on the measured data on $M_x$ and $M_y$. (f) Assume that $y$ has no impact on $x$, then $M_x$ has no information from $y$. Training a neural network to approximate the unsmooth map $\Phi : M_x \rightarrow M_y$ will fail.

Figure 9  Inferring motifs in directional complex network. (a) Fan-out pattern. (b) Fan-in pattern. (c) Cascading pattern

To quantify such smoothness without finding nearest neighbors, an efficient algorithm CMS was designed in [59]. The idea of CMS lies on the fact that any smooth map can be approximated by a neural network $N$ while training a neural network to approximate an unsmooth map will fail with large training errors, as illustrated in Figures 8(a) and 8(d). Moreover, the relative training error reflects the relative smoothness, and thus it can be a measure of the relative strength of causative effectiveness. The sketch of the cross map smoothness (CMS) with the neural network (NN) method is illustrated in Figures 8(e) and 8(f). The test on the benchmark system carried out in [59] implied that the CMS method could deal with time series data with length around $O(10)$, which greatly reduced the length requirement of $O(10^3)$ by nearest neighbor based methods.

It is stressed that the SSR methods, including both the nearest neighbor based methods and the cross map smoothness method, are designed to detect causality with weak to moderate coupling in nonlinear systems. For too weak coupling, the footprints from the driving force are too low to be measured, while for strong coupling, various kinds of synchronization will appear (see [65]), and causality cannot be defined between synchronized dynamics. Actually, the problem of synchronization, particularly transient synchronization, between two time series will cause problems in causality detection from short-term data, which should be avoided in practical applications.
4.4 Indirect causality

The causality detection from time series is mainly defined in a pair-wise style, which is the base of network topology inference, but there is still a gap between the existing methods and the network topology reconstruction. Though there are some discussions for the existing methods (see [59, 91]), the detection of complex causality patterns remains uncomplete in the literature. As indicated in [62], three types of motifs are generic in network inference: fan-out pattern, fan-in pattern, and cascading pattern, as shown in Figure 9. Intuitively, in the fan-out pattern, multiple variables are driven by one common source and consequently they are correlated and have the same footprints from the common source, as shown in Figure 9(a). Both the CCM method (see [91]) as well as CMS method (see [59]) have been shown to work on this pattern, and no significant causality has been detected between the driven nodes. This result confirms the difference between correlation and nonlinear causal relation. Moreover, Hirata and Aihara designed a method specially to deal with common driving source in [32] based on the recurrence plot technique, a technique closely related to SSR technique in nonlinear time series analysis. This was demonstrated to work well in climate and brain network reconstruction (see [33]). While the fan-in pattern shows no possibility to generate spurious relation, the detected causal strengths are actually weaker as reported in [59]. The third motif is the cascading pattern in which $x$ drives $y$ and $y$ drives $z$ but $x$ has no direct driving force over $z$, as illustrated in Figure 9(c). In this pattern, since the information of $x$ has also been injected into the dynamics of $z$ via the intermediate variable $y$, the spurious indirect causality from $x$ to $z$ could be detected by using the SSR methods pair-wisely. Compared to such detected indirect causal relation, it is more desirable to be able to distinguish the direct causal relationship from the indirect causal relationship in the network reconstruction. To the best knowledge of the authors, there is no SSR methods specially designed to deal with this issue to date, which leaves space for future work in the field.

4.5 Causality with delay

In the concept of causality, it is generally accepted that “the cause occurs before the effect” (see [24, 100]). This implies that the influence of one event on another cannot be instantaneous but delayed after some specific time. As a matter of fact, in the real world, time delays are intimately related to causation, which may result from the transmission time, switching speed, memory effect, or other effects. A real-world example is protein expression regulated by transcription factors, where the regulation process goes through a cascade of transcriptions and translations that take from minutes to hours to exert their influences. However, the SSR methods, for example CCM, actually assume that the time series are on the same footing through the implicit assumption that the directed influence occurs instantaneously, i.e., the detected causality is from $x(t)$ to $y(t)$. Thus, application of the original SSR methods to detect causality may lead to erroneous results. Recently, there are several efforts to incorporate time delay into the CCM scheme. In [103], Ye et al. considered a single time shift in the underlying time series so as to distinguish time-delayed causal interactions by CCM method. However, it is found that the simple incorporation of time shift into CCM framework is problematic. Specifically, the detected time delay by CCM may drift under different embedding dimensions (see [60]), which is definitely confusing. An analogous method was also independently proposed in [80], but it also suffered from the same problem. Quite recently, this problem has been fixed in [60], where a new score cross map evaluation (CME) was designed to substitute the old score in the CCM scheme, and it was reported to be capable of detecting causality with single, multiple, and even distributed time delays. Despite of such progresses in incorporating time delay into SSR based methods, there are still theoretical gaps to be filled. When considering nonlinear systems with time delays, actually one is dealing with infinite-dimensional dynamical systems. However, the classic Takens’ embedding theorem is proposed for finite-dimensional systems. Among the few studies on this topic, Robinson [76] has set up an infinite-dimensional Takens theorem, which partially gives theoretical foundation for applying SSR method to time delayed systems and sheds light on this problem.
5 Conclusion

The world is ruled by physical laws and each factor is evolving, together with other factors, under various governing laws. Thus the world is complex and no factor could be really independent from each other, but all the factors are intertwined by the dynamics of others. The state space reconstruction methods, stemming from the embedding theorem, have brought light to the analysis of such complex and nonlinear world. Particularly, the SSR methods provide a bridge between the observed time series data and the dynamics of the whole hidden original system, making it possible to reconstruct the original system by simply measuring some observed data. Moreover, all these processes could be done in a model-free style and no equations governing the system are required a priori. The SSR methods could work as a framework and be effective in various ways of reconstructing the system such as determining dimensionality, characteristic exponent, steady states, unstable periodic orbits, etc. In this review, we focus on two data based applications of the SSR methods: prediction and causality inference for nonlinear systems. Despite the long developing history and fruitful results, these two fields have been still active and attracting even to these days, and new methods keep on getting developed in the recent years. Particularly, in this review, we present in detail the forefront methods to deal with short-term time series data, and several perspectives for future work have been provided. Such developments would create an indispensable tool to big data analysis and systems reconstruction. Also, these types of methods can be applied to the problems of inferring dynamical network structures (see [55,105]) and detecting dynamical criticality of nonlinear systems before the abrupt changes of system states (see [12,53,54]).

Acknowledgements This work was supported by the National Key Research and Development Program of China (Grant No. 2017YFA0505500), Japan Society for the Promotion of Science KAKENHI Program (Grant No. JP15H05707), and National Natural Science Foundation of China (Grant Nos. 11771010, 31771476, 91530320, 91529303, 91439103 and 81471047).

References

1. Ancona N, Marinazzo D, Stramaglia S. Radial basis function approach to nonlinear Granger causality of time series. Phys Rev E (3), 2004, 70: 056221
2. Ansari-Asl K, Senhadji L, Bellanger J J, et al. Quantitative evaluation of linear and nonlinear methods characterizing interdependencies between brain signals. Phys Rev E (3), 2006, 74: 031916
3. Arnhold J, Grassberger P, Lehnertz K, et al. A robust method for detecting interdependencies: Application to intracranially recorded EEG. Phys D, 1999, 134: 419–430
4. Blinowska K J, Kuś R, Kamiński M. Granger causality and information flow in multivariate processes. Phys Rev E (3), 2004, 70: 050902
5. Boyce M S. Population viability analysis. Annu Rev Ecol Syst, 1992, 23: 481–497
6. Breakspear M, Terry J. Detection and description of non-linear interdependence in normal multichannel human EEG data. Clin Neurophysiol, 2002, 113: 735–753
7. Brovelli A, Ding M, Ledberg A, et al. Beta oscillations in a large-scale sensorimotor cortical network: Directional influences revealed by Granger causality. Proc Natl Acad Sci USA, 2004, 101: 9849–9854
8. Buzug T, Pfister G. Optimal delay time and embedding dimension for delay-time coordinates by analysis of the global static and local dynamical behavior of strange attractors. Phys Rev A (3), 1992, 45: 7073–7084
9. Candes E J, Romberg J K, Tao T. Stable signal recovery from incomplete and inaccurate measurements. Comm Pure Appl Math, 2006, 59: 1207–1223
10. Casdagli M. Nonlinear prediction of chaotic time series. Phys D, 1989, 35: 335–356
11. Casdagli M, Eubank S, Farmer J D, et al. State space reconstruction in the presence of noise. Phys D, 1991, 51: 52–98
12. Chen L, Liu R, Liu Z, et al. Detecting early-warning signals for sudden deterioration of complex diseases by dynamical network biomarkers. Sci Rep, 2012, 2: 342
13. David O, Guillemain I, Sallet S, et al. Identifying neural drivers with functional MRI: An electrophysiological validation. PLoS Biol, 2008, 6: e315
14. de Jong H. Modeling and simulation of genetic regulatory systems: A literature review. J Comput Biol, 2002, 9: 67–103
15. Deyle E R, Fogarty M, Hsieh C H, et al. Predicting climate effects on Pacific sardine. Proc Natl Acad Sci USA,
16 Deyle E R, Maher M C, Hernandez R D, et al. Global environmental drivers of influenza. Proc Natl Acad Sci USA, 2016, 113: 13081–13086
17 Deyle E R, Sugihara G. Generalized theorems for nonlinear state space reconstruction. PLoS ONE, 2011, 6: e18295
18 Fan J, Han F, Liu H. Challenges of big data analysis. Natl Sci Rev, 2014, 1: 293–314
19 Farmer J, Sidorowich J. Predicting chaotic time series. Phys Rev Lett, 1987, 59: 845–848
20 Finkenstädt B, Kühbier P. Forecasting nonlinear economic time series: A simple test to accompany the nearest neighbor approach. Empir Econ, 1995, 20: 243–263
21 Fraser A M, Swinney H L. Independent coordinates for strange attractors from mutual information. Phys Rev A (3), 1986, 33: 1134
22 Geweke J. Inference and causality in economic time series models. In: Handbook of Econometrics, vol. 2. Amsterdam: Elsevier, 1984, 1101–1144
23 Granger C. Investigating causal relations between econometric models and cross-spectral methods. Econometrica, 1969, 37: 424–438
24 Granger C. Time series analysis, cointegration, and applications. Amer Econ Rev, 2004, 94: 421–425
25 Grassberger P, Procaccia I. Measuring the strangeness of strange attractors. Phys D, 1983, 9: 189–208
26 Grassberger P, Schreiber T, Scharnhorst C. Nonlinear time sequence analysis. Int J Bifurcat Chaos, 1991, 1: 521–547
27 Guo S, Wu J, Ding M, et al. Uncovering interactions in the frequency domain. PLoS Comput Biol, 2008, 4: e1000087
28 Haefner J W. Modeling Biological Systems: Principles and Applications. New York: Springer Science & Business Media, 2005
29 Hamilton J D. Time Series Analysis, vol. 2. Princeton: Princeton University Press, 1994
30 Hegger R, Kantz H, Schreiber T. Practical implementation of nonlinear time series methods: The TISEAN package. Chaos, 1999, 9: 435–435
31 Heskamp L, Meel-van den Abeelen A, Lagro J, et al. Convergent cross mapping: A promising technique for cerebral autoregulation estimation. Int J Clin Neurosci Ment Health, 2014, 1: S20
32 Hirata Y, Aihara K. Identifying hidden common causes from bivariate time series: A method using recurrence plots. Phys Rev E (3), 2010, 81: 016203
33 Hirata Y, Amigó J M, Matsuzaka Y, et al. Detecting causality by combined use of multiple methods: Climate and brain examples. PLoS ONE, 2016, 11: e0158572
34 Hlaváčková-Schindler K, Paluš M, Vejmelka, M, et al. Causality detection based on information-theoretic approaches in time series analysis. Phys Rep, 2007, 441: 1–46
35 Hodgkin A L, Huxley A F. A quantitative description of membrane current and its application to conduction and excitation in nerve. Bull Math Biol, 1990, 52: 25–71
36 Huffaker R, Earne A. Empirically testing for dynamic causality between promotions and sales beer promotions and sales in England. In: International European Forum on System Dynamics and Innovation in Food Networks. Innsbruck-Igls: International European Forum, 2014, 270–274
37 Ikeguchi T, Aihara K. Estimating correlation dimensions of biological time series with a reliable method. J Intell Fuzzy Syst, 1997, 5: 33–52
38 Jamsek J, Stefanovska A, Mc Clintock P V. Nonlinear cardio-respiratory interactions revealed by time-phase bispectral analysis. Phys Med Biol, 2004, 49: 4407–4425
39 Jiang J J, Huang Z G, Huang L, et al. Directed dynamical influence is more detectable with noise. Sci Rep, 2016, 6: 24088
40 Judd K. Nonlinear state estimation, indistinguishable states, and the extended Kalman filter. Phys D, 2003, 183: 273–281
41 Kamiński M, Ding M, Truccolo W A, et al. Evaluating causal relations in neural systems: Granger causality, directed transfer function and statistical assessment of significance. Biol Cyber, 2001, 85: 145–157
42 Kantz H, Schreiber T. Nonlinear Time Series Analysis, vol. 7. Cambridge: Cambridge University Press, 2004
43 Kember G, Fowler A. A correlation function for choosing time delays in phase portrait reconstructions. Phys Lett A, 1993, 179: 72–80
44 Kennel M B, Brown R, Abarbanel H D. Determining embedding dimension for phase-space reconstruction using a geometrical construction. Phys Rev A (3), 1992, 45: 3403–3411
45 Kurths J, Ruzmaikin A. On forecasting the sunspot numbers. Sol Phys, 1990, 126: 407–410
46 Lai Y C, David L, Hayden R. An upper bound for the proper delay time in chaotic time-series analysis. Phys Lett A, 1996, 218: 30–34
47 Lai Y C, Lerner D. Effective scaling regime for computing the correlation dimension from chaotic time series. Phys D, 1998, 115: 1–18
48 Le Van Quyen M, Martinerie J, Adam C, et al. Nonlinear analyses of interictal EEG map the brain interdependences in human focal epilepsy. Phys D, 1999, 127: 250–266
49 Li Y, Chen L. Big biological data: Challenges and opportunities. Genomics Proteomics Bioinform, 2014, 12: 187–189
50 Liebert W, Pawelzik K, Schuster H. Optimal embeddings of chaotic attractors from topological considerations. Europhys Lett, 1991, 14: 521–526
51 Liebert W, Schuster H. Proper choice of the time delay for the analysis of chaotic time series. Phys Lett A, 1989, 142: 107–111
52 Lindegren M, Checkley D M, Rouyer T, et al. Climate, fishing, and fluctuations of anchovy and sand in the California Current. Proc Natl Acad Sci USA, 2013, 110: 13672–13677
53 Liu R, Chen P, Aihara K, et al. Identifying early-warning signals of critical transitions with strong noise by dynamical network markers. Sci Rep, 2015, 5: 17501
54 Liu R, Wang X W, Aihara K, et al. Early diagnosis of complex diseases by molecular biomarkers, network biomarkers, and dynamical network biomarkers. Med Res Rev, 2014, 34: 455–478
55 Liu X, Wang Y, Ji H, et al. Personalized characterization of diseases using sample-specific networks. Nucleic Acids Res, 2016, 44: e164
56 Lockhart D J, Winzeler E A. Genomics, gene expression and DNA arrays. Nature, 2000, 405: 827–836
57 Longtin A. Nonlinear forecasting of spike trains from sensory neurons. Int J Bifurcat Chaos, 1993, 3: 651–661
58 Lorenz E N. Atmospheric predictability as revealed by naturally occurring analogues. J Atmospheric Sci, 1969, 26: 636–646
59 Ma H, Aihara K, Chen L. Detecting causality from nonlinear dynamics with short-term time series. Sci Rep, 2014, 4: 7464
60 Ma H, Leng S, Tao C, et al. Detection of time delays and directional interactions based on time series from complex dynamical systems. Phys Rev E (3), 2017, 96: 012221
61 Ma H, Zhou T, Aihara K, et al. Predicting time series from short-term high-dimensional data. Int J Bifurcat Chaos, 2014, 24: 1430033
62 Marbach D, Prill R J, Schaffer T, et al. Revealing strengths and weaknesses of methods for gene network inference. Proc Natl Acad Sci USA, 2010, 107: 6286–6291
63 Marinazzo D, Pellicoro M, Stramaglia S. Kernel method for nonlinear Granger causality. Phys Rev Lett, 2008, 100: 144103
64 McBride J C, Zhao X, Munro N B, et al. Sugihara causality analysis of scalp EEG for detection of early Alzheimer’s disease. Neuroimage Clin, 2015, 7: 258–265
65 McCracken J M, Weigel R S. Convergent cross-mapping and pairwise asymmetric inference. Phys Rev E (3), 2014, 90: 062903
66 McGowan J A, Deyle E R, Ye H, et al. Predicting coastal algal blooms in southern California. Ecology, 2017, 98: 1419–1433
67 Müller A, Kraemer J F, Penzel T, et al. Causality in physiological signals. Physiol Meas, 2016, 37: R46
68 Nollo G, Faes L, Antonini R, et al. Assessing causality in normal and impaired short-term cardiovascular regulation via nonlinear prediction methods. Philos Trans R Soc Lond Ser A Math Phys Eng Sci, 2009, 367: 1423–1440
69 Packard N H, Crutchfield J P, Farmer J D, et al. Geometry from a time series. Phys Rev Lett, 1980, 45: 712–716
70 Pecora L M, Moniz L, Nichols J, et al. A unified approach to attractor reconstruction. Chaos, 2007, 17: 013110
71 Pereda E, Quiroga R Q, Bhattacharya J. Nonlinear multivariate analysis of neurophysiological signals. Prog Neurobiol, 2005, 77: 1–37
72 Pikovsky A. Discrete-time dynamic noise filtering. Sov J Commun Technol Electron, 1986, 31: 81–85
73 Quakenbush J. Microarray data normalization and transformation. Nat Genet, 2002, 32: 496–501
74 Quian Quiroga R, Kraskov A, Kreuz T, et al. Performance of different synchronization measures in real data: A case study on electroencephalographic signals. Phys Rev E (3), 2002, 65: 041903
75 Rienecker M M, Suarez M J, Gelaro R, et al. Merra: Nasa’s modern-era retrospective analysis for research and applications. J Climate, 2011, 24: 3624–3648
76 Robinson J C. Takens’ embedding theorem for infinite-dimensional dynamical systems. Nonlinearity, 1999, 12: 1263–1275
77 Rosenstein M T, Collins J J, De Luca C J. Reconstruction expansion as a geometry-based framework for choosing proper delay times. Phys D, 1994, 73: 82–98
78 Sauer T, Yorke J, Casdagli M. Embedology. J Stat Phys, 1991, 65: 579–616
79 Schiff S J, So P, Chang T, et al. Detecting dynamical interdependence and generalized synchrony through mutual prediction in a neural ensemble. Phys Rev E (3), 1996, 54: 6708–6724
80 Schumacher J, Wunderle T, Fries P, et al. A statistical framework to infer delay and direction of information flow from measurements of complex systems. Neural Comput, 2015, 27: 1555–1608
81 Sello S. Solar cycle forecasting: A nonlinear dynamics approach. Astron Astrophys, 2001, 377: 312–320
82 Small M. Applied Nonlinear Time Series Analysis: Applications in Physics, Physiology and Finance, vol. 52. Singapore: World Scientific, 2005
83 Smith L A. What might we learn from climate forecasts? Proc Natl Acad Sci USA, 2002, 99: 2487–2492
84 Stam C J. Nonlinear dynamical analysis of EEG and MEG: Review of an emerging field. Clin Neurophysiol, 2005, 116: 2266–2301
85 Stark J, Broomehead D, Davies M, et al. Takens embedding theorems for forced and stochastic systems. Nonlinear Anal, 1997, 30: 5303–5314
86 Stark J, Broomehead D, Davies M, et al. Delay embeddings for forced systems, II: Stochastic forcing. J Nonlinear Sci, 2003, 13: 519–577
87 Stein R R, Bucci V, Toussaint N C, et al. Ecological modeling from time-series inference: Insight into dynamics and stability of intestinal microbiota. PLoS Comput Biol, 2013, 9: e1003388
88 Strogatz S H. Exploring complex networks. Nature, 2001, 410: 268–276
89 Sugihara G. Nonlinear forecasting for the classification of natural time series. Philos Trans R Soc Lond Ser A Math Phys Eng Sci, 1994, 348: 477–495
90 Sugihara G, May R. Nonlinear forecasting as a way of distinguishing chaos from measurement error in time series. Nature, 1990, 344: 734–741
91 Sugihara G, May R, Ye H, et al. Detecting causality in complex ecosystems. Science, 2012, 338: 496–500
92 Tajima S, Yanagawa T, Fujii N, et al. Untangling brain-wide dynamics in consciousness by cross-embedding. PLoS Comput Biol, 2015, 11: e1004537
93 Takens F. Detecting strange attractors in turbulence. In: Lecture Notes in Mathematics, vol. 898. Berlin-New York: Springer, 1981, 366–381
94 Theiler J. Spurious dimension from correlation algorithms applied to limited time-series data. Phys Rev A (3), 1986, 34: 2427–2432
95 Tsonis A A, Elsner J B. Nonlinear prediction as a way of distinguishing chaos from random fractal sequences. Nature, 1992, 358: 217–220
96 van Nes E H, Scheffer M, Brovkin V, et al. Causal feedbacks in climate change. Nature Clim Change, 2015, 5: 445–448
97 Wang W X, Lai Y-C, Grebogi C. Data based identification and prediction of nonlinear and complex dynamical systems. Phys Rep, 2016, 644: 1–76
98 Weigend A, Gershenfeld N. Time Series Prediction: Forecasting the Future and Understanding the Past. In: Proceedings of the NATO Advanced Research Workshop on Comparative Time Series Analysis. Upper Saddle River: Addison-Wesley Publishing, 1994
99 Whitney H. Differentiable manifolds. Ann of Math (2), 1936, 37: 645–680
100 Wiener N. The theory of prediction. Mod Math Engr, 1956, 1: 125–139
101 Wu J, Loucks O L. From balance of nature to hierarchical patch dynamics: A paradigm shift in ecology. Q Rev Biol, 1995, 70: 439–466
102 Ye H, Beamish R J, Glaser S M, et al. Equation-free mechanistic ecosystem forecasting using empirical dynamic modeling. Proc Natl Acad Sci USA, 2015, 112: 1569–1576
103 Ye H, Doyle E R, Gilarranz L J, et al. Distinguishing time-delayed causal interactions using convergent cross mapping. Sci Rep, 2015, 5: 14750
104 Ye H, Sugihara G. Information leverage in interconnected ecosystems: Overcoming the curse of dimensionality. Science, 2016, 353: 922–925
105 Zhao J, Zhou Y, Zhang X, et al. Part mutual information for quantifying direct associations in networks. Proc Natl Acad Sci USA, 2016, 113: 5130–5135