International Journal of Bifurcation and Chaos, Vol. 24, No. 12 (2014) 1430033 (19 pages)

DOI: 10.1142/S021812741430033X

Predicting Time Series from Short-Term High-Dimensional Data

Huanfei Ma
Collaborative Research Center for Innovative Mathematical Modelling, Institute of Industrial Science, The University of Tokyo, Tokyo 153-8505, Japan
School of Mathematical Sciences, Soochow University, Suzhou 215006, P. R. China
hfma@suda.edu.cn

Tianshou Zhou
School of Mathematics and Computational Science, Sun Yat-Sen University, Guangzhou 510275, P. R. China
mcszhtsh@mail.sysu.edu.cn

Kazuyuki Aihara*
Collaborative Research Center for Innovative Mathematical Modelling, Institute of Industrial Science, The University of Tokyo, Tokyo 153-8505, Japan
aihara@sat.t.u-tokyo.ac.jp

Luonan Chen*
Key Laboratory of Systems Biology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai 200031, P. R. China
lnchen@sibs.ac.cn

Received July 15, 2014

The prediction of future values of time series is a challenging task in many fields. In particular, making prediction based on short-term data is believed to be difficult. Here, we propose a method to predict systems' low-dimensional dynamics from high-dimensional but short-term data. Intuitively, it can be considered as a transformation from the inter-variable information of the observed high-dimensional data into the corresponding low-dimensional but long-term data, thereby equivalent to prediction of time series data. Technically, this method can be viewed as an inverse implementation of delayed embedding reconstruction. Both methods and algorithms are developed. To demonstrate the effectiveness of the theoretical result, benchmark examples and real-world problems from various fields are studied.

Keywords: Prediction; short-term; high-dimensional data.

*Authors for correspondence

1. Introduction

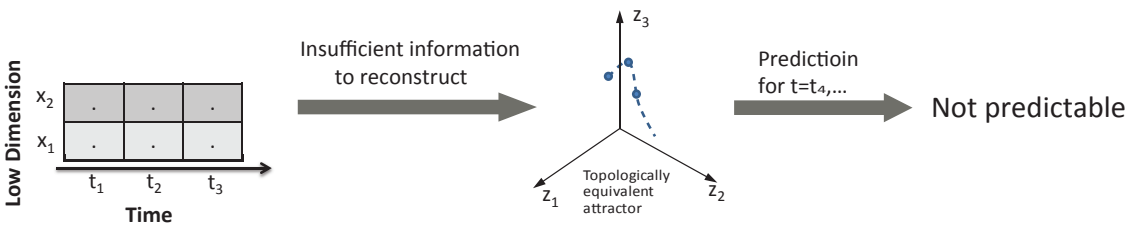
Reconstructing dynamics and making predictions based on observed time series are of great importance in various disciplines. Generally, the sources of predictability are believed to come from various kinds of information hidden in the output of a system, e.g. linear or nonlinear correlation in time-series, input–output relationship, and other statistical characteristics. Based on these sources of predictability, various methods have been proposed for reconstruction and predictions, e.g. regressive and smoothing models [Hamilton, 1994; Simonoff, 1996], nonlinear state-space reconstruction using embedding techniques [Packard *et al.*, 1980; Casdagli *et al.*, 1991; Takens, 1981; Sauer *et al.*, 1991; Stark *et al.*, 1997, 2003; Farmer & Sidorowich, 1987; Sugihara & May, 1990; Elsner & Tsonis, 1992; Abarbanel *et al.*, 1994], and even black-box modeling with neural networks [Greenwood, 1997]. Although many of these existing methods have been successfully applied to real-world systems [Weigend & Gershenfeld, 1994; Box *et al.*, 2013; Giebel *et al.*, 2011; Crone *et al.*, 2011], they all require sufficiently long measurements of time series as the system output, i.e. the observed data are required to be *long-term time series*. In other words, the predictability sources of these methods are all in a long-term temporal domain, and thus short-term time series data are generally considered to have insufficient information for such a task.

While considerably less attention has been paid to the prediction of short-term time series data, among a few studies on short-term data, Komalapriya *et al.* [2008, 2010] proposed an interesting method to reconstruct the system's dynamics by generating a single long-term time series from many short segments using the concept of recurrences in phase space. However, the observation on many short segments is necessary for one variable during different time periods, therefore this method in practice still requires combined long-term time series data to make predictions.

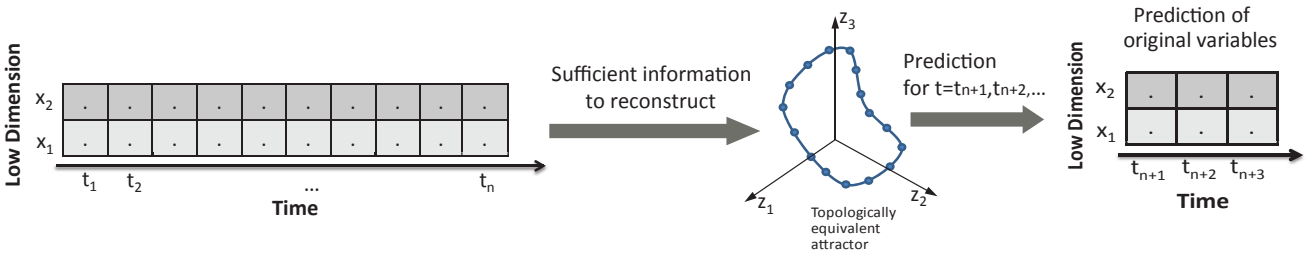
In this paper, we show that for short-term data with high-dimensional measurements, i.e. m time points of n-dimensional measured vector with $m \ll n$, it is still possible to make reliable predictions, particularly when the system has a low-dimensional attractor. Representative examples of such data are high throughput biological data, e.g. microarray data, where tens of thousands of gene probes on a chip can be measured simultaneously, but only some tens or even fewer successive time points can be measured for expression of each gene [Quackenbush, 2002] due to various restrictions. Collective dynamics of different complex networks [Strogatz, 2001] such as social networks and biological networks also shows high-dimensional but short time course characteristics, where a large number of coupled dynamical systems evolve via various topological structures but only limited time points can be recorded. Moreover, such a type of data are also prevalent in many astronomical observations, annual statistical reports, and dynamic meteorology [Strogatz, 2001; Holton & Hakim, 2013]. On the other hand, owing to strong nonstationary and irregular behavior of many real-world systems, even if a long time series is available, only recent data (thus, a short-term time series) provide valuable information for future dynamics, thereby implying the importance of the prediction based on the short-term time series.

Due to the conventional wisdom, such kinds of short-term time series data are generally believed to have insufficient temporal information for system reconstruction and predictions. On the contrary, here we note that such data have actually rich information contents due to the intertwined dynamics among a large number of measured variables, which may reflect the accumulated dynamical features of the target system in a similar way to scalar long-term time series data, as illustrated in Fig. 1. Intuitively, such intertwined relationship can be considered as a source of predictability. Thus, prediction from short-term high-dimensional time series can be achieved by the transformation from the intertwined dynamics embedded in the short-term high-dimensional data into long-term low-dimensional dynamics, as shown in Figs. 2(a) and 2(b).

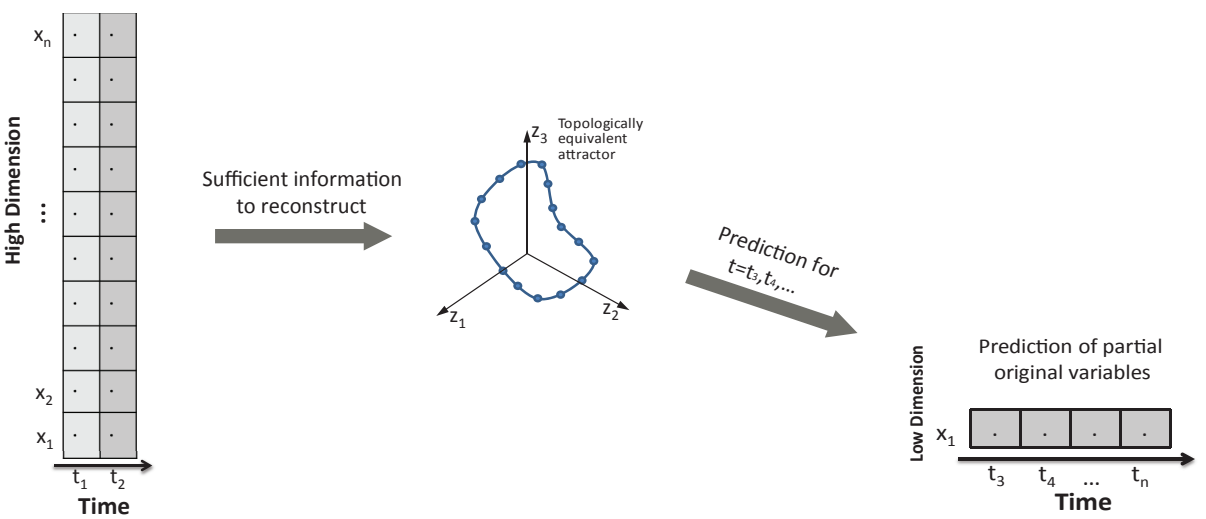
In this work, we first show that for short-term time series, the higher the dimension of measured data is, the greater is the amount of system-evolution information embedded in the data. Then, we develop a theoretical method to transform such information with the intertwined interactions among the observed multivariables into the time domain, and consequently, we transform short-term high-dimensional time series to the corresponding longer low-dimensional time series, which directly leads to the prediction of the dynamics on the selected variables with higher precision. From a mathematical viewpoint, this method can be viewed



(a)

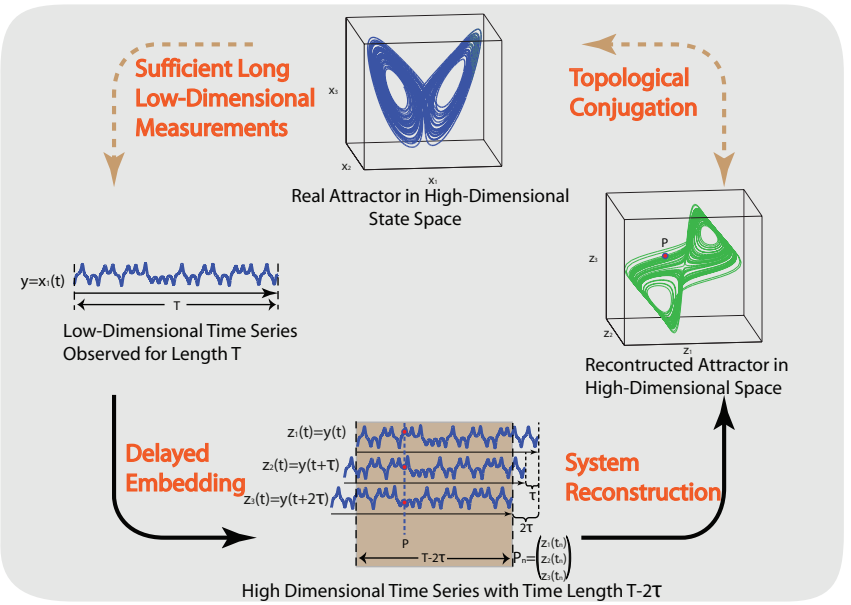


(b)

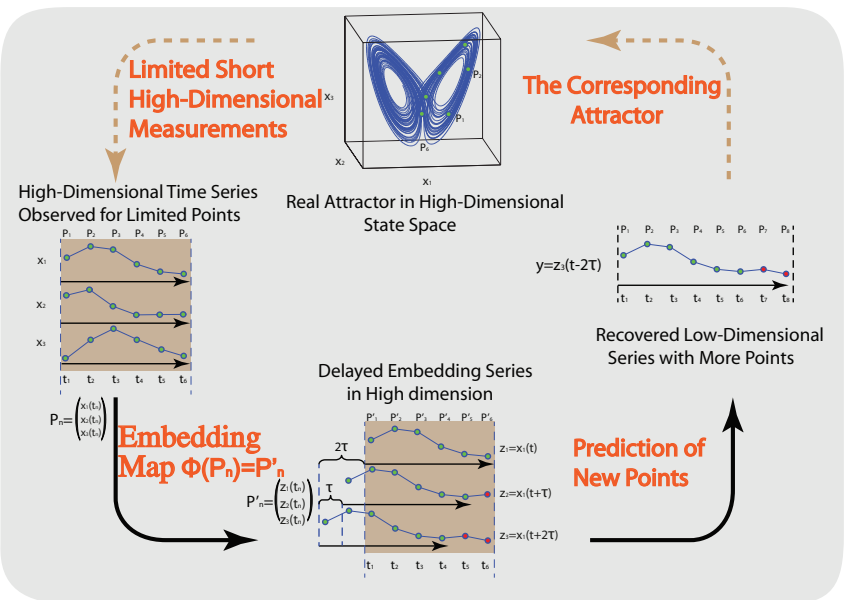


(c)

Fig. 1. Outline of the basic schemes to predict the time series by the existing method and our method. (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, which, therefore, can make the (short-term) prediction. Many existing methods are based on such a scheme. (c) A short-term high-dimensional time series also contain 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. This is the basic scheme of our method, i.e. inverse embedding theory.

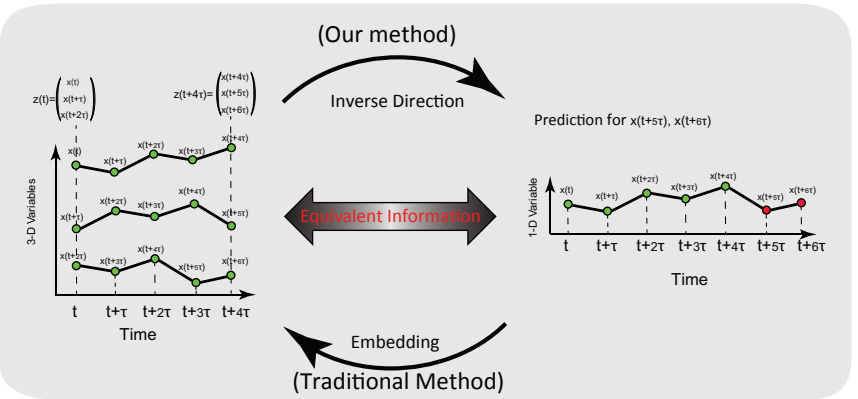


(a)



(b)

Fig. 2. Basic ideas to reconstruct nonlinear dynamics by either short-term high-dimensional data or long-term low-dimensional data. (a) Outline of the delayed embedding theory to reconstruct an attractor in a high-dimensional state space from a long-term scalar (or low-dimensional) time series. (b) Sketch of our method to predict the dynamics based on short-term high-dimensional time series data. (c) Transformation between short-term high-dimensional time series and long-term scalar (or low-dimensional) time series. That is, from the observed three-dimensional $Z(t)$ (green circles) at time points $t, \ldots, t+4\tau$, it is possible to predict the one-dimensional $x(t)$ (red circles) at the future time points $t+5\tau$ and $t+6\tau$. Note that the information of the future $x(t)$ at $t+5\tau$ and $t+6\tau$ is embedded in the observed three-dimensional $Z(t)$ at $t, \ldots, t+4\tau$ before $t+5\tau$.



(c)

Fig. 2. (*Continued*)

as an inverse of the Takens embedding theorem [Fig. 2(c)], which provides a new way to predict the dynamics of the system, in particular, from small samples or short-term time series data. We present the theoretical proof of our method for the general nonlinear forms. To demonstrate the effectiveness of the theoretical result, several numerical and real-world examples from various fields are studied.

2. Method

We consider an n-dimensional time series $\mathbf{x}(t) = (x_1(t), x_2(t), \ldots, x_n(t))^T \in \mathbb{R}^n$ with m time points measured from a system's attractor $\mathcal{A}$, namely, $\mathbf{x}_j = \mathbf{x}(t + (j-1)\tau), j = 1, 2, \ldots, m$, where τ is the sampling interval. Here, $\mathbf{x}$ is assumed to be the system's state variable vector itself, or n-dimensional observation of the unknown state variable vector $\mathbf{w}$ where the observation map $\Psi : \mathbf{w} \to \mathbf{x}$ is an embedding. The classical delayed embedding method [Kantz & Schreiber, 2004; Binder & Wissman, 2010; Pecora *et al.*, 2007] indicates that starting from a sufficiently long measured scalar variable $y(t) = h(\mathbf{x}(t))$ where h is a smooth function, it is possible to reconstruct the attractor in a high-dimensional state space (or equivalently, a trajectory with a higher dimension but shorter time length) using delay coordinates $\mathbf{z}(t) = (z_1(t), z_2(t), \ldots, z_L(t))^T \in \mathbb{R}^L$ by setting each component as $z_1(t) = y(t), z_2(t) = y(t+\tau), \ldots, z_L(t) = y(t + (L-1)\tau)$ where L is the dimension of the reconstructed system. Then from the Takens embedding theorem and the generalized embedding theorems for fractal attractors and stochastic systems [Takens, 1981; Sauer *et al.*, 1991; Stark *et al.*, 1997, 2003], the attractor of the reconstructed system with $\mathbf{z}(t)$ is one-to-one to the original one, as long as L is large enough, i.e. $L > 2d$ where d is the box counting dimension of the attractor. Moreover, there exists a smooth map $\Phi : \Phi(\mathbf{x}(t)) = \mathbf{z}(t)$ between $\mathbb{R}^n$ and $\mathbb{R}^L$ such that Φ is an embedding map on each compact subset of a smooth manifold contained in $\mathcal{A}$ [Sauer *et al.*, 1991]. From the viewpoint of time-series analysis, the delayed embedding method provides a practical way to transform long-term low-dimensional data into shorter high-dimensional data, as shown in Fig. 2(a).

We stress here that based on the embedding theory, one can actually make the transformation in the inverse way. Explicitly, without loss of generality, let us simply assume that the observed variable y is taken as the first component of the $\mathbf{x}$ system, i.e. $y(t) = h(\mathbf{x}(t)) = x_1(t)$, then we have the following matrix form:

$$\Phi : \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1m} \\ x_{21} & x_{22} & \cdots & x_{2m} \\ \vdots & \vdots & \vdots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nm} \end{pmatrix} \to \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1m} \\ x_{12} & x_{13} & \cdots & x_{1m+1} \\ \vdots & \vdots & \vdots & \vdots \\ x_{1L} & x_{1L+1} & \cdots & x_{1L+m-1} \end{pmatrix}, \quad (1)$$

where $x_{ij} = x_i(t + (j-1)\tau)$. This relationship implies that the $L + m - 1$ points of the first component $x_1(t) \in \mathbb{R}$ possess the same information

as m points of $\mathbf{x}(t) \in \mathbb{R}^n$ on the attractor in the sense of topological conjugation under the map Φ. Thus, starting from the m points of the original n-dimensional system $\mathbf{x}(t)$, one can convert these short-term high-dimensional data (m n-dimensional data points) into the long-term low-dimensional data ($m + L - 1$ one-dimensional data points) with map Φ, which is actually equivalent to the prediction of more $L - 1$ points in future for the first scalar component x_1 of state variable $\mathbf{x}$, as shown in Fig. 2(b). Clearly, the whole procedure uses the intertwined dynamics information embedded in the high-dimensional data as a source of predictability, and it can be regarded as an inverse implementation of the delayed embedding theory [Takens, 1981; Sauer *et al.*, 1991], as shown in Fig. 2(c).

We summarize the theory as follows: *Let $\mathbf{x}(t) \in \mathbb{R}^n$ be a time series generated from an n-dimensional dynamical system, and assume its attractor's box-counting dimension is d. Let L with $L > 2d$ be an integer, and let τ with $\tau > 0$ be a proper delay time. Then for the ith component $x_i \in \mathbb{R}$ and the consequent delay coordinate $\mathbf{z}(t) = (x_i(t), x_i(t+\tau), \ldots, x_i(t+(L-1)\tau))^T$, there exists a vector map $\Phi : \mathbb{R}^n \to \mathbb{R}^L$, such that $\Phi(\mathbf{x}) = \mathbf{z}$ is one-to-one and smooth. Furthermore, Φ maps $\mathbf{x}(t)$ to future states of its ith component, i.e. $x_i(t + j\tau), j = 1, 2, \ldots, L-1$ along the system's dynamics.*

Now we are at the point to derive the map Φ so that the transformation can be carried out. Since determining an arbitrary nonlinear vector function $\Phi = [\Phi_1, \Phi_2, \ldots, \Phi_L]^T : \mathbb{R}^n \to \mathbb{R}^L$ with finite constraints is too general, we will discuss the problem with expansions in specific functional spaces. Here, each component of the unknown map Φ is assumed to be expanded in a basis functional space such that

$$\Phi_i(\mathbf{x}) = \sum_{l=1}^{\infty} a_{li} g_l(\mathbf{x}), \quad i = 1, 2, \ldots, L, \qquad (2)$$

where $g_l(\mathbf{x})$ with $l = 1, 2, \ldots$ are the basic functions of specific forms. Particularly, since we consider the map restricted on the attractor, i.e. $\mathbf{x} \in \mathcal{A}$, it is reasonable to assume that each component Φ_i can be approximated locally by the expansions of finite terms, i.e.

$$\Phi_i(\mathbf{x}) \approx \sum_{l=1}^{k} a_{li} g_l(\mathbf{x}), \quad i = 1, 2, \ldots, L, \qquad (3)$$

which holds for the m measured time points $\mathbf{x} = \mathbf{x}(t + (j-1)\tau), j = 1, 2, \ldots, m$, and $a_{li}, l = 1, 2, \ldots, k$, $i = 1, 2, \ldots, L$ are the respective coefficients to be determined. Generally, the order of coefficients k will be practically large for good approximation, and m is comparatively small due to the short-term measurement. Therefore, the problem will be generally categorized into an underdetermined problem, i.e. there are more coefficients to be determined than the number of independent equations.

Even though the underdetermined problem raised above would be a difficult problem, in many situations, especially with the proper choice of a basis functional space, the number of zero or nearly zero coefficients in the expanding series can be of majority so that the coefficient vector is effectively sparse. Particularly, when the data are measured from an attractor with a low box-counting dimension, i.e. $d \ll n$, the expansion in Eq. (3) can be understood as a special form of feature extraction on a specific feature space and thus the expansion can be effectively sparse, which will be illustrated in the Result section. Therefore, the recently developed idea of compressive sensing [Candes *et al.*, 2006] provides a practical technique to solve the problem of determining the map Φ.

Generally, the problem of compressive sensing can be described as to reconstruct a sparse vector $\boldsymbol{\alpha} \in \mathbb{R}^n$ with m linear measurements of the form $\mathbf{y} = X\boldsymbol{\alpha}$, where $X = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_m]^T \in \mathbb{R}^{m \times n}$ are composed of m known test signals $\mathbf{x}_i \in \mathbb{R}^n$ and $\mathbf{y} \in \mathbb{R}^m$ are the measurements. According to the compressive sensing theory, when $m \ll n$, one can still actually recover $\boldsymbol{\alpha}$ exactly by solving the convex problem $\min_{\boldsymbol{\alpha}} \|\boldsymbol{\alpha}\|_{l_1}$ subject to $X\boldsymbol{\alpha} = \mathbf{y}$, where $\|\boldsymbol{\alpha}\|_{l_1} = \sum_{i=1}^{n} |\alpha_i|$ with the L_1 norm of vector $\boldsymbol{\alpha}$, provided that the matrix $X \in \mathbb{R}^{m \times n}$ obeys a uniform uncertainty principle [Candes *et al.*, 2006].

Thus, given m measured time-points data $\mathbf{x}_1$, $\mathbf{x}_2, \ldots, \mathbf{x}_m$ where $\mathbf{x}_j = \mathbf{x}(t + (j-1)\tau)$ and the observation vector $\mathbf{y}$ as the ith component of the state vector such that $y_j = x_{ij} = x_i(t + (j-1)\tau)$, $j = 1, 2, \ldots, L + m - 1$, we can cast the solution of Eq. (3) into the compressive sensing problem and develop the following iterative algorithm to determine Φ as well as to make predictions.

Algorithm 1.

1. Choose one proper basis functional space and expansion order k with basis functions $g_l(\mathbf{x})$, $l = 1, 2, \ldots, k$.

2. Construct

$$\hat{\mathbf{x}}_j = [g_1(\mathbf{x}_j), g_2(\mathbf{x}_j), \ldots, g_k(\mathbf{x}_j)]^T, \quad j = 1, 2, \ldots, m$$

based on $\mathbf{x}_j$ such that $\Phi_i(\mathbf{x}_j) = \hat{\mathbf{x}}_j^T \boldsymbol{\alpha}_i$.

3. **For** $i = 2, 3, \ldots, L$,

solve $\boldsymbol{\alpha}_i$ using compressive sensing technique with all the equations related to $\boldsymbol{\alpha}_i$ and y_{i+m-2} calculated in the previous step as follows:

$$\hat{\mathbf{x}}_j^T \boldsymbol{\alpha}_i = y_{i+j-1}, \quad j = 1, 2, \ldots, m-1.$$

Calculate $y_{m+i-1} = \hat{\mathbf{x}}_m^T \boldsymbol{\alpha}_i$.

End for ■

Here, $y_{m+1}, y_{m+1}, \ldots, y_{m+L-1}$ are the $L-1$ predictions for the ith component of state vector $\mathbf{x}$ solved in the $L-1$ loops, and each predicted value is added into the equations in the next loop.

3. Results

3.1. *Verification with theoretical model*

To validate our theoretical analysis we first consider a simple toy model, which is an 8-D (dimensional) oscillator in the following form:

$$\begin{aligned}
\dot{x}_1 &= \frac{1}{2}(x_1 - x_4 - 2x_2 + 2x_3) \\
&\quad - \frac{1}{4}[(x_1 - x_4)^2 + (x_2 - x_3)^2] \\
&\quad \times (x_1 - x_4 - x_2 + x_3) - \frac{1}{2}(x_1 + x_4), \\
\dot{x}_2 &= \frac{1}{2}(2x_1 - 2x_4 + x_2 - x_3) \\
&\quad - \frac{1}{4}[(x_1 - x_4)^2 + (x_2 - x_3)^2] \\
&\quad \times (x_1 - x_4 + x_2 - x_3) - \frac{1}{2}(x_2 + x_3), \\
\dot{x}_3 &= -\frac{1}{2}(2x_1 - 2x_4 + x_2 - x_3) \\
&\quad + \frac{1}{4}[(x_1 - x_4)^2 + (x_2 - x_3)^2] \\
&\quad \times (x_1 - x_4 + x_2 - x_3) - \frac{1}{2}(x_2 + x_3), \\
\dot{x}_4 &= -\frac{1}{2}(x_1 - x_4 - 2x_2 + 2x_3) \\
&\quad + \frac{1}{4}[(x_1 - x_4)^2 + (x_2 - x_3)^2] \\
&\quad \times (x_1 - x_4 - x_2 + x_3) - \frac{1}{2}(x_1 + x_4), \\
\dot{x}_i &= -x_i + \sqrt{2}x_{i-4}, \quad i = 5, 6, 7, 8,
\end{aligned} \tag{4}$$

whose attractor is a circle (i.e. 1-D) in 8-D state space expressed as

$$\begin{aligned}
\mathbf{x} &= (x_1, x_2, \ldots, x_8) \\
&= \Bigg(\frac{\sqrt{2}}{2}\cos(t), \frac{\sqrt{2}}{2}\sin(t), -\frac{\sqrt{2}}{2}\sin(t), \\
&\quad -\frac{\sqrt{2}}{2}\cos(t), \frac{1}{2}\sin(t) + \frac{1}{2}\cos(t), \\
&\quad \frac{1}{2}\sin(t) - \frac{1}{2}\cos(t), -\frac{1}{2}\sin(t) + \frac{1}{2}\cos(t), \\
&\quad -\frac{1}{2}\sin(t) - \frac{1}{2}\cos(t) \Bigg),
\end{aligned} \tag{5}$$

which is one-dimensional. As shown in Fig. 3(a), the toy model is particularly designed to produce low-dimensional dynamics with high-dimensional measures.

Assume that the time series is observed with equal intervals τ, i.e. $\mathbf{x}_1 = \mathbf{x}(t), \mathbf{x}_2 = \mathbf{x}(t+\tau), \ldots$, and simply choose the observation function as $y(\mathbf{x}(t)) = x_1(t)$; then, we have $y_1 = x_1(t), y_2 = x_1(t+\tau), \ldots$. According to the Takens embedding theorem and its generalized theorem, one can choose $L = 3$ such that $L > 2d$ and reconstruct the attractor in $\mathbb{R}^L$ space, i.e. let $\mathbf{z}_i = [y_i, y_{i+1}, y_{i+2}]^T$, such that there is an embedding map $\Phi : \mathbf{x}_i \to \mathbf{z}_i$. Let $L = 3 > 2d$, we can analytically determine the form of map Φ as

$$\Phi(\mathbf{x}) = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \cos(\tau) & -\sin(\tau) & 0 & 0 & 0 & 0 & 0 & 0 \\ \cos(2\tau) & -\sin(2\tau) & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \mathbf{x}, \tag{6}$$

and thus, if we choose power expansion as basis functions with order up to $k = 2$, the coefficients are clearly sparse, i.e.

$$\Phi_1(\mathbf{x}) = (0 \quad 1 \quad 0 \quad \cdots \quad 0)\tilde{\mathbf{x}}, \tag{7}$$

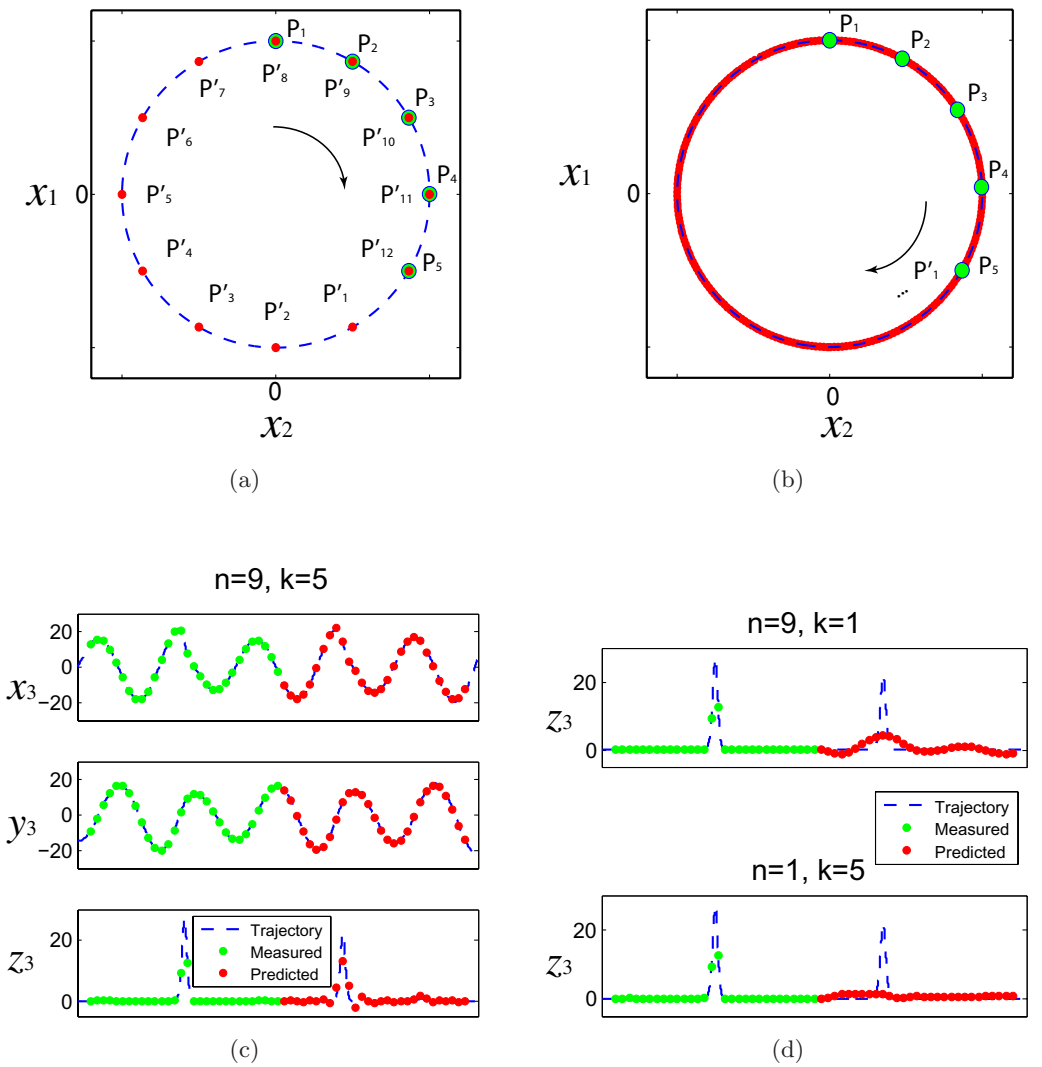


Fig. 3. (a) and (b) Validation with theoretical models. (a) Prediction for an 8-D oscillator with the period 2π, where the projection into the x_1–x_2 plane is illustrated. $P_1, P_2, \ldots, P_5$ are five measured points with time interval $\pi/6$ on the attractor, marked by green circles. $P'_1, \ldots, P'_{12}$ are predicted points, marked by red circles. (b) The same setting as in (a), except that the measured time interval is $\pi/\sqrt{40}$. (c) Prediction for a high-dimensional coupled chaotic Rössler system using power expansion technique with order 5. Only three components x_3, y_3, z_3 of the system are shown here. (d) Top: The prediction result for the nine-dimensional coupled Rössler system using a linear map, where z_3 is shown. Bottom: The prediction result with the same 5 order power expansion for one-dimensional data using only z_3 as the measured time series.

where $\tilde{\mathbf{x}} = (1, x_1, x_2, \ldots, x_8, x_1^2, x_1x_2, \ldots, x_8^2)^T$ contains 45 terms.

With $m = 5$ samples points, the embedding map Φ can be exactly determined by Algorithm 1. Moreover, the long-term evolution can be exactly predicted, as shown in Figs. 3(a) and 3(b), where the sampling interval τ is taken as $\pi/6$ and $\pi/\sqrt{40}$, respectively. Note that in the latter case, since the sampling interval has an irrational relation with the period, the predicted points are dense over the whole attractor, and the unknown system attractor is completely reconstructed.

Then to test the method on chaotic systems, let us consider a network consisting of three paradigmatic Rössler oscillators [Rössler, 1976; Menck *et al.*, 2013], in which the dynamics at node i is represented by

$$\begin{aligned} \dot{x}_i &= -y_i - z_i - K\sum_{j=1}^{N} L_{ij}x_j, \\ \dot{y}_i &= x_i + ay_i, \\ \dot{z}_i &= b + z_i(x_i - c), \quad i = 1, 2, \ldots, N, \end{aligned} \tag{8}$$

where coupling constant $K = 1$ and $L_{ij} = 1$ for all i and j. As shown in Fig. 3(c), when the parameters are set as $a = 0.1, b = 0.1, c = 14$, and $N = 3$, the trajectory of the system shows irregular dynamics due to the chaotic property, and the measured data show nonperiodic behavior. For each component of the system, we use the False Nearest Neighbors (FNNs) method [Kennel *et al.*, 1992] to determine the minimal embedding dimension and find that the minimal embedding dimension was no more than 3. We simply choose power expansion up to order $k = 5$ as the expansion of the map. The prediction result is illustrated in Fig. 3(c). Here, we stress that the measured data of every single component contain no sufficient information for the prediction, e.g. for the z_3 component, only one irregular spike is contained in the measured data. Therefore, it would definitely lead to a failure using only a part of the system state as measured data, i.e. using z_3 as single measured time series data rather than considering the nine-dimensional time series as a whole, as shown in Fig. 3(d)(bottom). Thus, we confirm that the dynamical feature of the system is hidden in the intertwined nine-dimensional data as a whole. This is the basic idea underlying successful predictions by our method, which fully considers the information hidden in the intertwined nine-dimensional variables and converts the information into low-dimensional time series with the embedding map Φ. Moreover, we also show that the proper expansion order is crucial for the success of the method. Setting $k = 1$, the map is assumed to be linear, which is definitely not a good approximation and the prediction would fail, as shown in Fig. 3(d)(top).

3.2. *Associative memory retrieval*

To investigate the mechanism of the neural computation, a wide variety of models have been proposed to realize flexible intelligent information processing such as associative memory retrieving. An artificial neural network with associative memory is a typical example of such models, and information of memory patterns is stored within synaptic connections in the networks. When distorted patterns are input into the network, the network will retrieve the stored patterns through its high-dimensional complex dynamics [Haykin, 2010].

Since the memory capacity of the network depends on the number of neurons, a large number of neurons are usually involved in the process, and each neuron will generate its own output. Therefore, the associative memory retrieval process is usually based on high-dimensional time series. On the other hand, before the network reaches the retrieved pattern, it will produce a complex dynamical output of the whole network. Thus, it is an interesting problem to examine whether it is possible to predict which pattern will be retrieved only with very short observed time series.

Here, we consider the classic Hopfield network [Hopfield, 1982] composed of $n = 120$ neurons representing 12×10 pixels. The stored patterns are eight black and white patterns as shown in Fig. 4(d), specifically designed as a benchmark [Haykin, 2010]. In the continuous-time Hopfield model, starting from a distorted pattern, the network will finally converge to one original pattern, as shown in Figs. 4(a) and 4(b).

Assuming that only $m = 15$ time points are sampled with time interval $\tau = 0.02$, it should be noted that the output time series are standard short-term high-dimensional data, i.e. $m \ll n$. Thus we first select proper basis functions. Considering the property of neural networks, we assume that the embedding map Φ could be approximated by the following linear combination of sigmoidal functions:

$$\Phi_i = \sum_{l=1}^{k} a_{li} f_l(\mathbf{x}), \tag{9}$$

and we choose the order $k = n$ with $f_l(\mathbf{x}) = f(x_l)$ for each l, where $f(x_l)$ is a sigmoidal active function. We choose $L = 30$, and we use Algorithm 1 to generate predictions for the state vector $\mathbf{x}(t)$. Using symbolic output, we can further get the predicted dynamic patterns and the final retrieved pattern, as shown in Fig. 4(c). In addition, we calculate the overlap of the dynamic output patterns with the stored pattern, where the predicted output patterns coincide with the truly stored pattern quite well, and the method retrieves the stored pattern successfully on the basis of the limited measured data, as shown in Fig. 4(e).

3.3. *Dynamics of complex networks*

The past decade has witnessed rapid growth of research interest in various kinds of complex networks over various disciplines of science. Among previously proposed research topics, collective dynamics of a complex network is especially important for understanding the networks' evolution

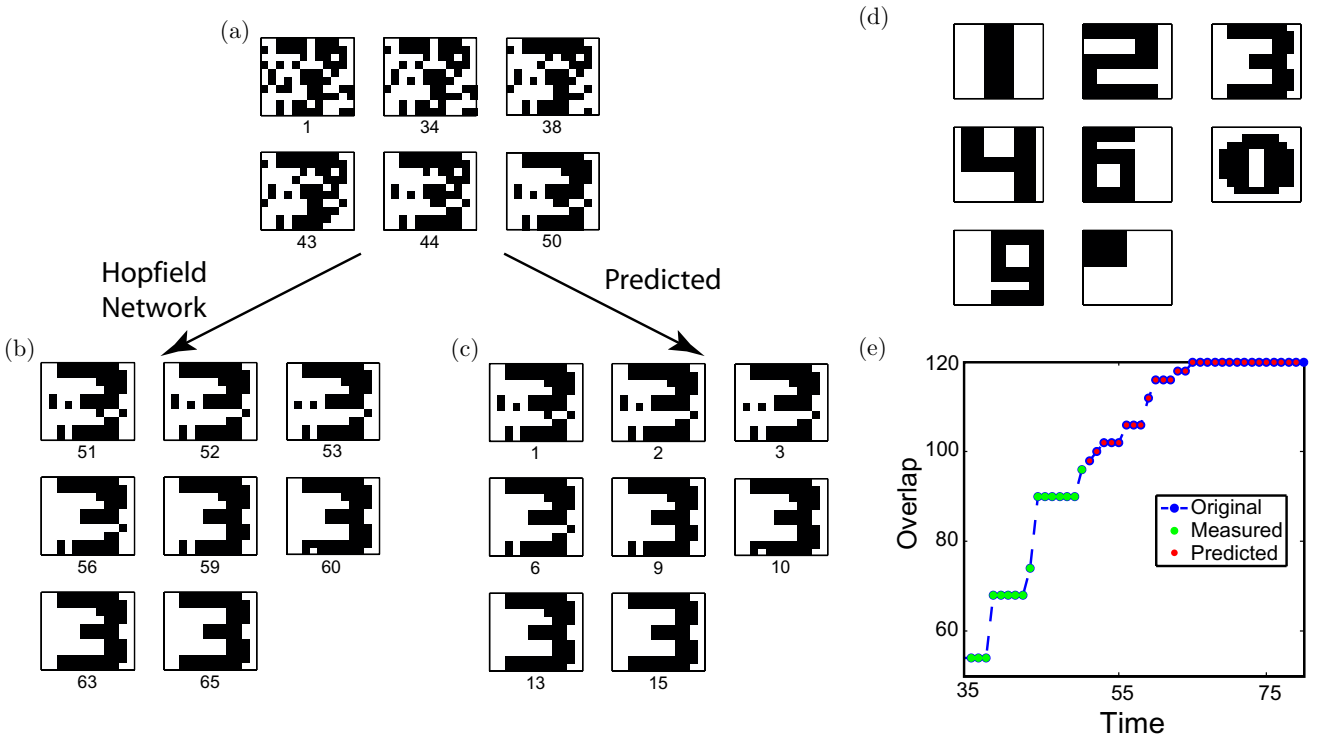


Fig. 4. Associative memory retrieval illustrations. (a) Starting from a distorted pattern with 30% pixels inverted, the dynamic output patterns during the first 50 steps of the Hopfield network are shown. (b) The dynamic output patterns during the following steps of the Hopfield network to retrieve the associative memory. (c) The predicted dynamic output patterns using our method. (d) Eight digit-like stored patterns with 10×12 pixels. (e) The overlap of the output patterns with the stored pattern.

mechanism. However, due to the large number of network nodes, considerably less attention has been paid to the dynamics of the entire network, and even less to the time series forecasting of the network behavior [Strogatz, 2001]. Therefore, it is worthwhile considering the prediction problem for the high-dimensional output time series of networks with limited observation data.

For this purpose, we consider a network consisting of n oscillator nodes coupled through different network structures. We simply represent each node as a Kuramoto-type oscillator [Kuramoto, 2003] in the following form:

$$\dot{\theta}_i = \omega_i + K \sum_{j=1}^{N} a_{ij} \sin(\theta_j - \theta_i), \quad i = 1, 2, \ldots, N, \tag{10}$$

where θ_i is the angular phase of node i, K is the coupling strength, N is the number of nodes, ω_i is the intrinsic natural frequency of node i, and a_{ij} is the element of the adjacency matrix such that $a_{ij} = 1$ if nodes i and j are connected, and $a_{ij} = 0$, otherwise. In our examples, the natural frequencies are randomly distributed in $(-1/2, 1/2)$ [Kuramoto, 2003] and the initial values of θ_i are randomly drawn from a uniform distribution in the interval $(-\pi, \pi)$, respectively, and we set the coupling strength as $K = 10$. With such setting, each isolated node exhibits periodic oscillation in the angular phase. When coupled under different network structures, the network of oscillators will exhibit various collective dynamics [Moreno & Pacheco, 2004]. Therefore, we test our method with two representative structures of complex networks, namely the scale-free network [Barabási & Albert, 1999] and the small-world network [Watts & Strogatz, 1998].

To build these networks, we use the Barabási–Albert (BA) algorithm [Barabási & Albert, 1999] and the Watts–Strogatz (WS) algorithm [Watts & Strogatz, 1998], respectively. For both situations, we choose $n = 100$ and set the average edge degree to $\langle k \rangle = 2$. The structures of the two networks are illustrated in Figs. 5(a) and 5(b), respectively.

With exactly the same parameters set except the network topology, it is observed that all the nodes of the scale-free network tend to be synchronized soon, while the small-world network tend to exhibit more complicated dynamics. To measure the

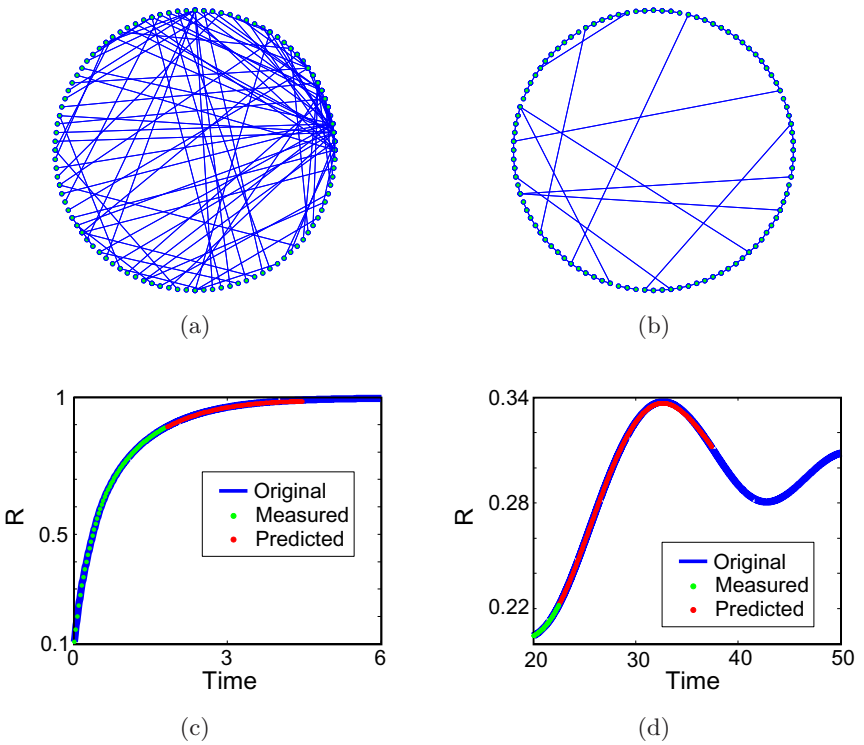


Fig. 5. (a) and (b) The structures of scale-free and small world networks with $n = 100$ oscillator nodes respectively, where both the average edge degrees are $\langle k \rangle = 2$. (c) and (d) The order parameter R for both networks to illustrate prediction results.

degree of synchronization in the networks, we define the order parameter R as

$$R(t) = \left| \frac{1}{N} \sum_{j=1}^{N} e^{\mathrm{i}\theta_j(t)} \right|, \tag{11}$$

where R with $0 \leq R \leq 1$ indicates the degree of synchronization. The dynamics of the order parameter under different structures are shown in Figs. 5(c) and 5(d), respectively.

Before applying Algorithm 1 to make predictions for such high-dimensional time series, we note that the output of the network is in the special form of angular phases for oscillators; therefore, we choose Fourier series as expansion bases, so that we could cast the nonlinear embedding map into the following form:

$$\Phi_i(\boldsymbol{\theta}) = \sum_{\mathbf{k}} c_{\mathbf{k}} e^{\mathrm{i}\mathbf{k}\boldsymbol{\theta}}, \tag{12}$$

where $\mathbf{k}$ is a lattice vector belonging to some proper set Ω. Thus, we can make predictions with such an expanded form by Algorithm 1.

For the scale-free network, using initial $m = 60$ time points measured with $\tau = 0.03$, we make predictions for $L = 90$ new points, whose order parameter $R(t) = \left| \frac{1}{N} \sum_{j=1}^{N} e^{\mathrm{i}\theta_j(t)} \right|$ is shown in Fig. 5(c), which indicates complete synchronization of the entire network. It should be noted in this case that the system's attractor is contained in the complete synchronization manifold while the observed dynamics is transient. Though the transient dynamics does not satisfy the assumption that the data are measured from the system's attractor, the prediction method still works well. Therefore, it stresses the fact that in our theory, the assumption that the data are measured from an attractor is sufficient but not necessary for the success of our method, and in some cases, even the transient dynamics can also be predicted, though not guaranteed by the theory.

For the small-world network, using $m = 50$ time points measured with $\tau = 0.05$, we further forecast the dynamics for the next $L = 400$ time points, as shown in Fig. 5(d). Both results confirm that our method performs quite well in forecasting the collective dynamics of such complex networks.

3.4. *Images series*

Images can illustrate information more straightforwardly and vividly than plain texts and numbers, while a series of moving images includes much more

information than a single one. An example is a series of medical images taken continuously during a specific time period for the same patient, which may describe the disease dynamics. Cloud images taken by a satellite every hour is another representative image series, which can carry information of winds and ocean flows that cannot be illustrated by a single snapshot. Therefore, making predictions on the basis of sequential images is an important topic.

Particularly, digital images are constituted by pixels, and the more the number of the pixels is, the better is the resolution of the picture. Therefore, every image can be regarded as a high-dimensional pixel vector in color space, and in such a manner, the successive image series can be cast into the high-dimensional time series data. Thus, we can make predictions for image series by our method, which does not require any specific knowledge about the images *a priori*.

First, we consider a toy model, where the 20×20 digital image series shows the dynamics of a Gaussian distribution with its center rotating on the unit circle, as shown in Fig. 6(a). We convert each image into a vector whose components take values in $[0, 255]$, and thus we cast these images into time series of $n = 400$ dimensional data with only $m = 10$ time points. Clearly, $m = 10 \ll n = 400$, and thus the data for such a system are short-term high-dimensional data.

It should be noted that there are only fuzzy relations between successive images, and the dynamics of every pixel seems to be governed by some unknown complicated high-dimensional system. For such types of data, we chose the Radial Basis Function (RBF) network to approximate the embedding map Φ, which is considered to have a universal approximation ability [Park & Sandberg, 1991]. Therefore, we expand each component Φ_i of the embedding map as

$$\Phi_i(\mathbf{x}) = \sum_j a_{ij}\rho_j(\mathbf{x}, c_j), \tag{13}$$

where each ρ_j is a radial basis function in the form of

$$\rho_j(\mathbf{x}, c_j) = \exp\left(\frac{-\|\mathbf{x} - c_j\|^2}{2\sigma_j^2}\right), \tag{14}$$

with c_j denoting the center for ρ_j, and σ_j denoting its width. Using the FNNs method, we find that the optimal embedding dimension for each pixel is no more than 5, thus we set $L = 5$.

With the above setting, we make predictions for every pixel using Algorithm 1 (by replacing g_l by ρ_j), and finally obtain the predicted images illustrated in Fig. 6(a), where the predicted centers of the distribution rotates along the circle as expected, and the predicted images indeed show the rotating dynamics of a Gaussian distribution.

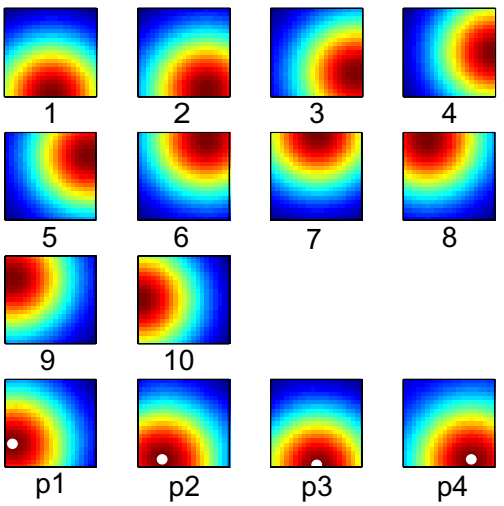
Next, we test our method with real satellite cloud image data. Here, we consider a series of 28 cloud images of the West Pacific, starting from 2012.6.16.15:32 with one image taken per hour.[1] This image series actually shows the route and development of the No. 4 Typhoon "GUCHOL", which affected China and Japan in June 2012, as shown in Fig. 6(b). We first preprocess the images and consider the local $n = 30 \times 30$ pixels around the typhoon cloud. Then, based on $m = 21$ images, we made predictions for the next eight time points. Unlike the previous cases, here we only make predictions for the center position of the typhoon (2-D), known as the typhoon eye. The predicted result is shown in the inset of Fig. 6(b), which fits well with the real trajectory of the typhoon center.

3.5. *High throughput data*

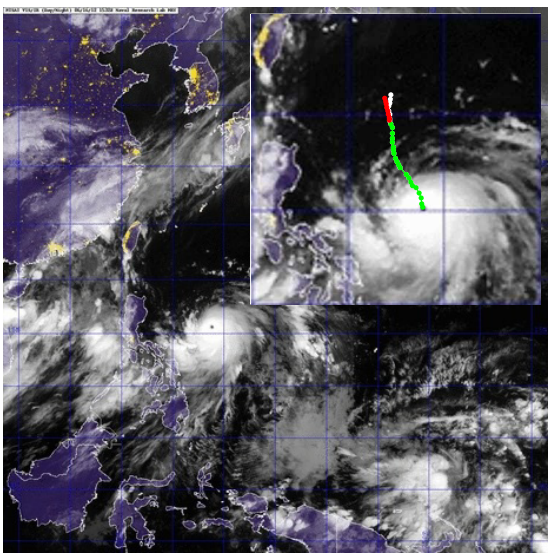
High throughput biological data are representative high-dimensional data typically with a very small number of time points. Therefore, it is of great importance to test our method with high throughput biological data. Here, we considered a set of microarray data, which was obtained by a gene expression profiling study of both miRNA and mRNA in mouse liver [Na *et al.*, 2009]. The data were a time series containing 12 time points of 46 628 probes (or gene expressions), with each probe measured every 4 hours over 48 hours.

It is believed that genes, proteins, and RNAs are not independent, but integrated in a huge and complicated regulation system [Shen-Orr *et al.*, 2002]. Therefore, in spite of the extremely high dimension and different types of probes or genes, the evolution of all these probes can be regulated by certain underlying complicated regulation dynamics, thus forming a high-dimensional dynamical system. Consequently, it is possible to use our method to convert the information in the high-dimensional data into the long-term low-dimensional temporal

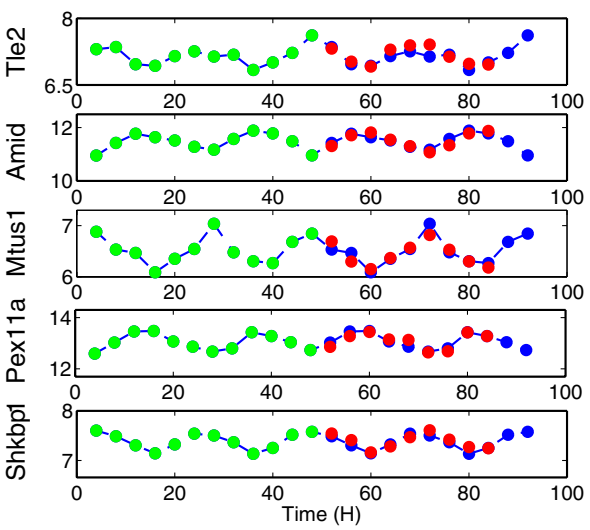
[1] http://www.nrlmry.navy.mil/NEXSAT.html.



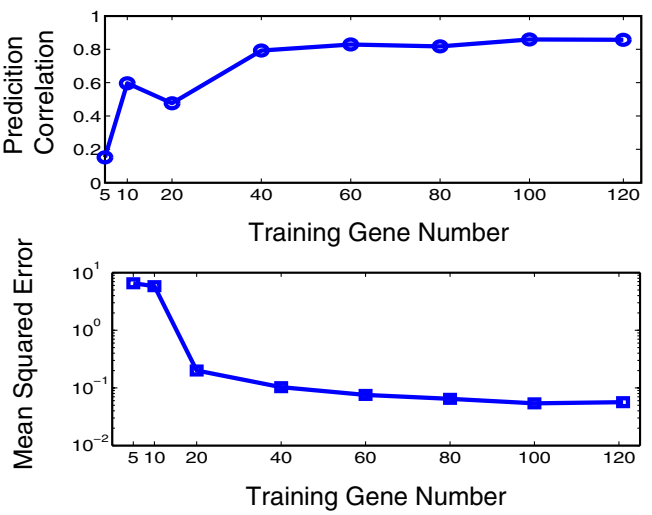
(a)



(b)



(c)



(d)

Fig. 6. (a) and (b) Sequential image time series prediction. (a) Upper series images with order numbers $1, 2, \ldots, 10$ show the dynamics of a Gaussian distribution with its center rotating on a planar circle. Lower series images with numbers $p1, p2, p3$ and $p4$ show the predicted images with centers marked with a white dot. (b) The satellite cloud image for the West Pacific area on 2012.6, where the typhoon "GUCHOL" is shown. The route of the enlarged typhoon eye is shown in the inset. Here, white dots represent the route of the typhoon eye, and green dots represent the positions of the typhoon eye used as measured data, while the red dots represent the predicted ones. (c) and (d) Prediction illustration for the microarray data using power expansion. (c) With $n = 100$ selected probes as measured data, the prediction is shown for the five probes, where the green dots denote the measured series, the red ones denote the predicted ones, and the blue ones denote the original data. The power expansion order is 2. (d) For different numbers of training genes, the average correlation coefficient between the predicted series and the original ones as well as the MSE between the predicted series and the original ones are shown.

domain and forecast the expression dynamics of a specific probe. For this purpose, we first randomly select a probe *Shkbp*1 (Sh3kbp1 binding protein 1), and compute the correlations with all the other probes. Since the probes belonging to a tightly connected regulation subnetwork can be considered to be more correlated than others, we pick 121 out of 46 628 probes, which are most correlated with *Shkbp*1. Then, we use the power expansion techniques with order up to 4 to calculate

the embedding map. With some normalization processes, the result of the prediction is shown in Fig. 6(c), where 100 related probes among the 121 correlated ones are used as measured data.

4. Discussion

4.1. *Parameters*

The theoretical analysis and the model verification imply that the proposed method is able to make precise predictions and the attractor can be well reconstructed, if the embedding map Φ can be approximately determined. The success of determining Φ depends on several characterizing parameters, such as the system dimension n, the length of measured time series m, and the length of predicted data L. In the simplest situation where the map Φ is linear, one can get the exact form of the map Φ iteratively by solving linear equations. Using linear equations theory, we can get the condition for exact reconstruction in the linear case: $m > L$ ($L \simeq 2d + 1$, and $L \in \mathbb{N}$), which is provided in Appendix A. Here, the dimension L of the reconstructed space is required to be larger than $2d$ so that the trajectory in the reconstructed space does not intersect. Actually, it is neither possible nor necessary to estimate the exact box-counting dimension for the unknown attractor where the high-dimensional time series data are measured. Thus it is a specific topic to choose a proper L, and several discussions can be found in [Takens, 1981; Sauer *et al.*, 1991]. In this paper, we apply the FNNs method [Kennel *et al.*, 1992] to find the minimal sufficient embedding dimension D from time series and select $L \geq D$. Generally, the box-counting dimension d of a high-dimensional dissipative system's attractor is much lower compared with the dimension of the original state space, i.e. $d \ll n$ [Takens, 1981; Sauer *et al.*, 1991], and as a result we have $L \ll n$. Therefore, in the linear case, a system's trajectory with high-dimensional state space (n-dimensional state space where n is large) moving on a low-dimensional attractor (with dimension d where $d \ll n$), or simply, low-dimensional dynamics with high-dimensional measurements can be exactly reconstructed with short-term ($m \simeq 2d + 1$) measurements (Appendix A).

Next, to illustrate the dependence on the available length and dimension of the measured data, we consider a small-world complex network consisting of $n = 100$ Kuramoto oscillators with the average edge degree $\langle k \rangle = 4$ as a benchmark. With such setting, the system shows irregular dynamics. First, we increase the available measured data length from $m = 5$ to $m = 100$ and calculate the length of accurate predicted time points. The result in Fig. 7(a) clearly shows that more measured data lead to better predictions. However, we stress that a saturation value around $m = 20$ indicates that $m \simeq 20$ measured data are sufficient to get an acceptable prediction result in this case. Note that since $m = 20 \ll n = 100$, the proposed method works well with short-term high-dimensional data.

Then we fix the length of available measured data and assume that only a part of the system variables can be measured, i.e. only a p-dimensional observation vector with $p < n$ can be obtained from a system with $n = 100$. With the partial dimensional data, the prediction result is summarized in Fig. 7(b), which shows that the prediction performance improves with the increase of the data dimension. This further confirms the basic idea behind our method that more information about the intertwined network in state space leads to longer predictions. However, a saturation around $p = 30$ indicates that the most closely correlated variables can provide sufficient information for predictions. Actually, this conclusion is of great significance for real data, e.g. gene chip data. For real microarray data, it is neither possible nor practical to measure every variable of the complicated regulation system, and instead it is sufficient to consider a part of the most closely correlated probes. By this reason, we test our method for the gene data [Na *et al.*, 2009] with different numbers of genes as the training set. We feed $120, 100, 80, 60, 40, 20, 10,$ and 5 records into Algorithm 1, with each selected records set contained in the previous one. To quantitatively compare the prediction result, we concatenate two 48 h periods to obtain a 96 h period due to the circadian property, and we regard the first 48 h period as training data and the second 48 h period as testing data. We introduce two indexes, i.e. one is the correlation coefficient between the predicted time series and the test data, and the second index is the mean squared error between the predicted time series and the testing data in a pointwise manner. The results are shown in Fig. 6(d), which validate the fact that about 80 most correlated genes are enough for the prediction of one gene.

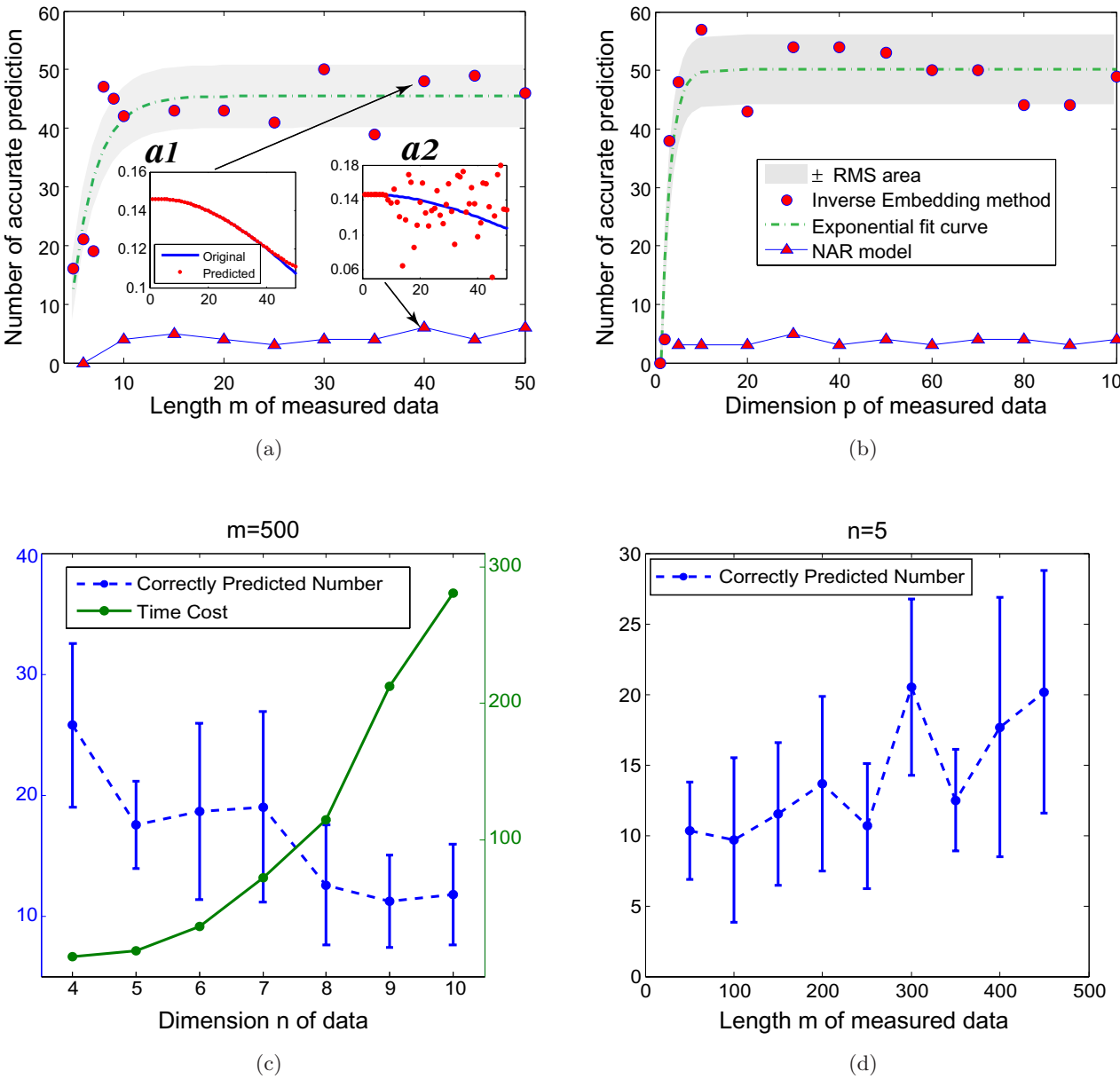


Fig. 7. (a) and (b) The number of accurate predictions, the fitted curve and the root mean square error (RMS) for both the NAR model and the proposed method for a complex network, with different numbers and dimensions of measured data. The insets of (a1) and (a2) show the prediction results for the same test set with $m = 40$ time points as measured data for both the proposed inverse embedding method and NAR method. (c) The number of accurate predictions as well as the computational time cost for networks with different dimensions using the NAR method, where the training data length is $m = 500$. (d) The number of accurate predictions with different lengths of training data for an $n = 5$ dimensional network.

4.2. *Robustness*

In practice, the map Φ is usually nonlinear with a much more complicated form, and one can hardly get the exact form of the complicated map Φ from numerical algorithms; instead, one can only obtain its approximation form, e.g. by Algorithm 1. Therefore, the predictions for a high-dimensional system usually depend on various factors such as noise, length and dimension of available data.

In particular, noise cannot be totally avoided in practical situations, and thus systems driven by stochastic processes rather than pure deterministic systems would be expected in most real applications. Though the classic embedding theories cannot directly be applied to stochastic systems due to the infinite dimensions, stochastic embedding theorem for stochastic systems by Stark *et al.* [1997, 2003] guarantees that the embedding map Φ still

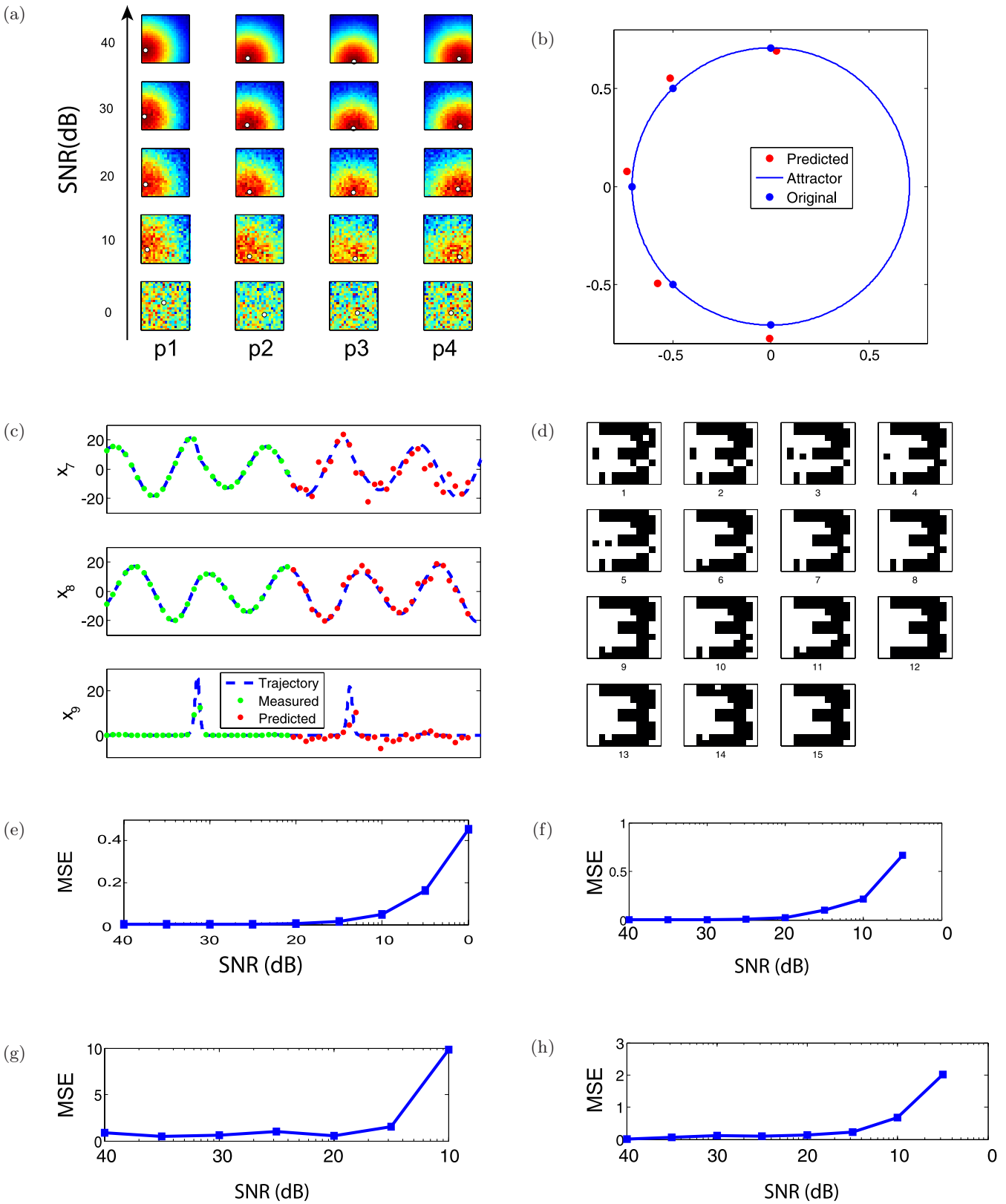


Fig. 8. (a) The prediction for image series under different levels of additive Gaussian noise, where centers are marked as white circles. (e) The MSE for the predicted center position under different levels of additive Gaussian noise. (b)–(d) The performance of the proposed method under SNR of 20 dB for the toy model, the chaotic Rössler network and the associative memory retrieval, respectively. (f)–(h) The MSE curves for the three cases of (b)–(d).

exists even when the system is driven by some independent stochastic processes, which makes Algorithm 1 theoretically effective for time series deteriorated by noise. In other words, as a generalization of the classic Takens' embedding theorem for a deterministic system, Stark's embedding theorem ensures the existence of the embedding map Φ for a stochastic system, thereby providing the theoretical foundation of our method for applying to stochastic systems. To further test the practical robustness of the proposed method, we consider the effect of additive Gaussian noise, where we use the images-series case shown in Fig. 6(a) as a benchmark. The results under different levels of the signal to noise ratio (SNR) are shown in Figs. 8(a) and 8(e), where the prediction result is very good with SNR higher than 20 dB. Other examples with various noise levels are given in Fig. 8, which also confirm that the method works well under moderate noise distortion.

4.3. *Comparison*

The above analysis indicates that the proposed method is particularly effective for predictions of high-dimensional short-term data. For such a kind of data, the conventional method based on long-term temporal information generally gets poor results due to the lack of temporal information for system reconstruction and predictions. To make a comparison, we test the regressive methods which are widely accepted as standard time series analysis methods. First, we consider the vector auto-regression (VAR) model which is the generalized form for the linear regression in high-dimensional space. The VAR model describes the evolution of a set of n variables as a linear combination of only their past values. Generally, for a system with n state variables and lag k, the VAR model requires at least $m = kn^2$ system's measured time points to determine the model. Thus, it is clear that the VAR method cannot be applied in the short-term data. Then, we consider a more general regressive method, i.e. the nonlinear auto-regression (NAR model) method, which is a nonlinear generalization for the auto-regression model. We apply the NAR method for the same short-term data set and obtain the results shown in Figs. 7(a) and 7(b). Actually, we further show that as the data dimension increases, the cost of the NAR method increases rapidly and the performance decreases as shown in Fig. 7(c), implying that the NAR method may not be suitable for high-dimensional data. Moreover, as the measured data length varies from $m = 50$ (short-term) to $m = 450$ (long-term), the result is improved in a monotone manner as the length of the training data increases, as shown in Fig. 7(d). Therefore, it confirms that the NAR method is suitable on occasions where long-term measured data are available but works poor only with short-term data.

In summary, we have introduced a method to transform short-term high-dimensional data into long-term low-dimensional data, which is equivalent to predictions of short-term data, by exploiting the information among high-dimensional variables. It can be viewed as a scheme of the transformation from the correlation (or inter-variable) information into the dynamical information. We find that the method is effective in several representative cases. Moreover, the dependence of reliability on several factors are discussed, and comparison with conventional methods are carried out.

Acknowledgments

This research is supported by the FIRST program from JSPS initiated by CSTP, the Strategic Priority Research Program of the Chinese Academy of Sciences (No. XDB13040700), and the National Natural Science Foundation of China (Nos. 61134013, 91029301, 11301366 and 91230204).

References

Abarbanel, H. D., Carroll, T., Pecora, L., Sidorowich, J. & Tsimring, L. S. [1994] "Predicting physical variables in time-delay embedding," *Phys. Rev. E* **49**, 1840.

Barabási, A.-L. & Albert, R. [1999] "Emergence of scaling in random networks," *Science* **286**, 509–512.

Binder, P.-M. & Wissman, B. [2010] "Geometry of repeated measurements in chaotic systems," *Chaos* **20**, 013106.

Box, G. E., Jenkins, G. M. & Reinsel, G. C. [2013] *Time Series Analysis*: *Forecasting and Control* (John Wiley & Sons).

Candes, E. J., Romberg, J. K. & Tao, T. [2006] "Stable signal recovery from incomplete and inaccurate measurements," *Commun. Pure Appl. Math.* **59**, 1207–1223.

Casdagli, M., Eubank, S., Farmer, J. D. & Gibson, J. [1991] "State space reconstruction in the presence of noise," *Physica D* **51**, 52–98.

Crone, S. F., Hibon, M. & Nikolopoulos, K. [2011] "Advances in forecasting with neural networks?

Empirical evidence from the NN3 competition on time series prediction," *Int. J. Forecasting* **27**, 635–660.
Elsner, J. & Tsonis, A. [1992] "Nonlinear prediction, chaos, and noise," *Bull. Amer. Meteorolog. Soc.* **73**, 49–60.
Farmer, J. D. & Sidorowich, J. J. [1987] "Predicting chaotic time series," *Phys. Rev. Lett.* **59**, 845.
Giebel, G., Brownsword, R., Kariniotakis, G., Denhard, M. & Draxl, C. [2011] "The state-of-the-art in short-term prediction of wind power: A literature overview," Tech. rep., ANEMOS.plus.
Greenwood, G. [1997] "Training multiple-layer perceptrons to recognize attractors," *IEEE Trans. Evolut. Comput.* **1**, 244–248.
Hamilton, J. D. [1994] *Time Series Analysis*, Vol. 2 (Princeton University Press, Princeton).
Haykin, S. [2010] *Neural Networks*: *A Comprehensive Foundation*, *1994* (McMillan, NJ).
Holton, J. R. & Hakim, G. J. [2013] *An Introduction to Dynamic Meteorology* (Academic Press).
Hopfield, J. J. [1982] "Neural networks and physical systems with emergent collective computational abilities," *Proc. Natl. Acad. Sci.* **79**, 2554–2558.
Kantz, H. & Schreiber, T. [2004] *Nonlinear Time Series Analysis*, Vol. 7 (Cambridge University Press).
Kennel, M. B., Brown, R. & Abarbanel, H. D. [1992] "Determining embedding dimension for phase-space reconstruction using a geometrical construction," *Phys. Rev. A* **45**, 3403.
Komalapriya, C., Thiel, M., Romano, M., Marwan, N., Schwarz, U. & Kurths, J. [2008] "Reconstruction of a systems dynamics from short trajectories," *Phys. Rev. E* **78**, 066217.
Komalapriya, C., Romano, M., Thiel, M., Marwan, N., Kurths, J., Kiss, I. & Hudson, J. [2010] "An automated algorithm for the generation of dynamically reconstructed trajectories," *Chaos* **20**, 013107.
Kuramoto, Y. [2003] *Chemical Oscillations*, *Waves*, *and Turbulence* (Courier Dover Publications).
Menck, P. J., Heitzig, J., Marwan, N. & Kurths, J. [2013] "How basin stability complements the linear-stability paradigm," *Nature Phys.* **9**, 89–92.
Moreno, Y. & Pacheco, A. F. [2004] "Synchronization of Kuramoto oscillators in scale-free networks," *EPL* (*Europhys. Lett.*) **68**, 603.
Na, Y. J., Sung, J. H., Lee, S. C., Lee, Y. J., Choi, Y. J., Park, W. Y., Shin, H. S. & Kim, J. H. [2009] "Comprehensive analysis of microRNA-mRNA co-expression in circadian rhythm," *Exp. Mol. Med.* **41**, 638–647.
Packard, N. H., Crutchfield, J. P., Farmer, J. D. & Shaw, R. S. [1980] "Geometry from a time series," *Phys. Rev. Lett.* **45**, 712.
Park, J. & Sandberg, I. W. [1991] "Universal approximation using radial-basis-function networks," *Neural Comput.* **3**, 246–257.
Pecora, L. M., Moniz, L., Nichols, J. & Carroll, T. L. [2007] "A unified approach to attractor reconstruction," *Chaos* **17**, 013110.
Quackenbush, J. [2002] "Microarray data normalization and transformation," *Nat. Genet.* **32**, 496–501.
Rössler, O. E. [1976] "An equation for continuous chaos," *Phys. Lett. A* **57**, 397–398.
Sauer, T., Yorke, J. A. & Casdagli, M. [1991] "Embedology," *J. Statist. Phys.* **65**, 579–616.
Shen-Orr, S. S., Milo, R., Mangan, S. & Alon, U. [2002] "Network motifs in the transcriptional regulation network of *Escherichia Coli*," *Nat. Genet.* **31**, 64–68.
Simonoff, J. S. [1996] *Smoothing Methods in Statistics* (Springer).
Stark, J., Broomhead, D., Davies, M. & Huke, J. [1997] "Takens embedding theorems for forced and stochastic systems," *Nonlin. Anal.*: *Th. Meth. Appl.* **30**, 5303–5314.
Stark, J., Broomhead, D. S., Davies, M. & Huke, J. [2003] "Delay embeddings for forced systems. II. Stochastic forcing," *J. Nonlin. Sci.* **13**, 519–577.
Strogatz, S. H. [2001] "Exploring complex networks," *Nature* **410**, 268–276.
Sugihara, G. & May, R. M. [1990] "Nonlinear forecasting as a way of distinguishing chaos from measurement error in time series," *Nature* **344**, 734–741.
Takens, F. [1981] "Detecting strange attractors in turbulence," *Dynamical Systems and Turbulence*, *Warwick 1980* (Springer), pp. 366–381.
Watts, D. J. & Strogatz, S. H. [1998] "Collective dynamics of small-world networks," *Nature* **393**, 440–442.
Weigend, A. & Gershenfeld, N. [1994] *Time Series Prediction*: *Forecasting the Future and Understanding the Past* (Addison-Wesley Publishing Company).

Appendix A
Conditions for Linear Map

When limited to the linear functional space, Φ can be expressed in the form of an $L \times n$ matrix A such that

$$\Phi(X) = AX = Y. \tag{A.1}$$

Here

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{L1} & a_{L2} & \cdots & a_{Ln} \end{pmatrix}, \tag{A.2}$$

which is the coefficient matrix representing the linear map Φ, and $X = (\mathbf{x}_1 \quad \mathbf{x}_2 \quad \cdots \quad \mathbf{x}_m)$ with

$\mathbf{x}_j = (x_{1j}, x_{2j}, \ldots, x_{nj})^T, j = 1, 2, \ldots, m$ as column vectors, which represent m measured state vectors at time $t_1, t_2, \ldots, t_m$. Moreover,

$$Y = \begin{pmatrix} y_1 & y_2 & \cdots & y_m \\ y_2 & y_3 & \cdots & y_{m+1} \\ \vdots & \vdots & \vdots & \vdots \\ y_L & y_{L+1} & \cdots & y_{L+m-1} \end{pmatrix}, \quad \text{(A.3)}$$

where $y_1, y_2, \ldots, y_m$ are the scalar observations of $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_m$, which we can simply assume to be one specific component of the state variable. Thus, the linear map can be rewritten in the vector form as

$$\begin{pmatrix} \mathbf{a}^1 \\ \mathbf{a}^2 \\ \vdots \\ \mathbf{a}^L \end{pmatrix} (\mathbf{x}_1 \quad \mathbf{x}_2 \quad \cdots \quad \mathbf{x}_m) = \begin{pmatrix} y_1 & y_2 & \cdots & y_m \\ y_2 & y_3 & \cdots & y_{m+1} \\ \vdots & \vdots & \vdots & \vdots \\ y_L & y_{L+1} & \cdots & y_{L+m-1} \end{pmatrix}, \quad \text{(A.4)}$$

where $\mathbf{a}^i$ represents the ith row vector of A. Note that even though the data we considered have extremely high dimension, the m measured data points are assumed to be observed from the system's attractor, which may have a much lower dimension d, i.e. the attractor lies on a set $\mathcal{A}$ of dimension d contained in the whole space $\mathbb{R}^n$ with $d \ll n$. Consequently, $L \ll n$ because L is usually taken as the smallest integer such that $L > 2d$. Furthermore, note that when the embedding map Φ is linear, there is a one-to-one linear map between $\mathcal{A}$ and $\Phi(\mathcal{A})$. Since $\Phi(\mathcal{A}) \subset \mathbb{R}^L$, the d-dimensional set $\mathcal{A}$ will lie in a subspace with dimension no more than L. Therefore, generally, in the linear case, though the measured data $\mathbf{x}_j$ is of dimension n, no more than L components are linearly independent. Therefore, without loss of generality, it is sufficient to consider only L components from the n-dimensional data $\mathbf{x}_j$, and we can rewrite Eq. (A.4) as

$$\begin{pmatrix} \tilde{\mathbf{a}}^1 \\ \tilde{\mathbf{a}}^2 \\ \vdots \\ \tilde{\mathbf{a}}^L \end{pmatrix} (\tilde{\mathbf{x}}_1 \quad \tilde{\mathbf{x}}_2 \quad \cdots \quad \tilde{\mathbf{x}}_m) = \begin{pmatrix} y_1 & y_2 & \cdots & y_m \\ y_2 & y_3 & \cdots & y_{m+1} \\ \vdots & \vdots & \vdots & \vdots \\ y_L & y_{L+1} & \cdots & y_{L+m-1} \end{pmatrix}, \quad \text{(A.5)}$$

where $\tilde{\mathbf{x}}_j \in \mathbb{R}^L$ contains the L selected components of $\mathbf{x}_j$, and $\tilde{\mathbf{a}}^i \in \mathbb{R}^L$. Note that $\tilde{\mathbf{a}}^i$ is not an n-dimensional but an L-dimensional vector.

Consequently, we can simply choose linear basis functional space for Algorithm 1 to derive the linear map, i.e. every equation to be solved in Algorithm 1 can be written as $\tilde{\mathbf{a}}^i \tilde{\mathbf{x}}_j = y_{i+j-1} = \tilde{\mathbf{a}}^{i-1} \tilde{\mathbf{x}}_{j+1}$ with $j = 1, 2, \ldots, m-1$, and thus, we can rewrite the equations in a matrix form as

$$\begin{pmatrix} -\tilde{\mathbf{x}}_2^T & \tilde{\mathbf{x}}_1^T \\ \vdots & \vdots \\ -\tilde{\mathbf{x}}_m^T & \tilde{\mathbf{x}}_{m-1}^T \end{pmatrix} \begin{pmatrix} \tilde{\boldsymbol{a}}^{i-1T} \\ \tilde{\boldsymbol{a}}^{iT} \end{pmatrix} = \mathbf{0}, \quad \text{(A.6)}$$

where $\tilde{\mathbf{a}}^{i-1}$ contains L variables determined in the previous step and $\tilde{\mathbf{a}}^i$ contains L unknown variables to be determined. Therefore, we come to the following conclusion: For

$$\tilde{X} = \begin{pmatrix} -\tilde{\mathbf{x}}_2^T & \tilde{\mathbf{x}}_1^T \\ \vdots & \vdots \\ -\tilde{\mathbf{x}}_m^T & \tilde{\mathbf{x}}_{m-1}^T \end{pmatrix}, \quad \text{(A.7)}$$

the solution of Algorithm 1 is decided according to the rank of $\tilde{X}$. To be specific, when $\text{rank}(\tilde{X}) > L$, the whole algorithm has no exact solution; when $\text{rank}(\tilde{X}) \leq L$, the whole algorithm can be solved; in particular when $\text{rank}(\tilde{X}) = L$, the linear map can be exactly solved.

Noting the definition of m and the special form of $\tilde{X}$, one can determine the linear map Φ exactly with $L + 1$ linearly independently measured points.