Physics Letters A 380 (2016) 2359–2365

Contents lists available at ScienceDirect

Physics Letters A

www.elsevier.com/locate/pla

When univariate model-free time series prediction is better than multivariate

Masayoshi Chayama [a,*], Yoshito Hirata [a,b]

[a] Graduate School of Mathematical Informatics, The University of Tokyo, Tokyo 113-8656, Japan
[b] Institute of Industrial Science, The University of Tokyo, Tokyo 153-8505, Japan

ARTICLE INFO

Article history:
Received 8 February 2016
Received in revised form 13 May 2016
Accepted 15 May 2016
Available online 18 May 2016
Communicated by C.R. Doering

Keywords:
Delay coordinates
Solar irradiance
Time series prediction
Multivariate prediction

ABSTRACT

The delay coordinate method is known to be a practically useful technique for reconstructing the states of an observed system. While this method is theoretically supported by Takens' embedding theorem concerning observations of a scalar time series, we can extend the method to include a multivariate time series. It is often assumed that a better prediction can be obtained using a multivariate time series than by using a scalar time series. However, multivariate time series contains various types of information, and it may be difficult to extract information that is useful for predicting the states. Thus, univariate prediction may sometimes be superior to multivariate prediction. Here, we compare univariate model-free time series predictions with multivariate ones, and demonstrate that univariate model-free prediction is better than multivariate one when the prediction steps are small, while multivariate prediction performs better when the prediction steps become larger. We show the validity of the former finding by using artificial datasets generated from the Lorenz 96 models and a real solar irradiance dataset. The results indicate that it is possible to determine which method is the best choice by considering how far into the future we want to predict.

1. Introduction

Embedding methods are employed to reconstruct attractors from a time series. In particular, they are often used for making predictions for a time series over a short period. It is known that a chaotic dynamical system can be reconstructed from a scalar time series of observations of the state of a dynamical system under the condition stated in Takens' theorem [1,2], which is that the embedding dimension is larger than twice the dimension for the underlying dynamics. It is commonly assumed that a multivariate time series can provide a better future prediction than a scalar time series [3]. For example, weather forecasts are based on many types of data that are observed in various locations. Actually, if a distance between 2 points becomes smaller, a correlation coefficient of time series data of solar irradiance becomes higher [4]. Therefore, we commonly presume that we can obtain a better prediction by using a multivariate time series than by using a scalar time series. However, multivariate time series contains various types of information, and it is difficult to extract information that is useful for predicting the states [5]. Therefore, we investigate the conditions under which univariate time series prediction is superior to multivariate time series prediction. We use model-free prediction based on embedding techniques [6,7,5]. There are some existing methods for predicting a multivariate time series. In particular, those described in [8–10] adopt a linear approach. In this paper, we compare multivariate predictions with univariate predictions by using a nonlinear approach.

The remainder of this paper is organized as follows. In Section 2, we discuss the theoretical framework of delay coordinates and Takens' embedding theorem. In addition, we discuss our prediction methods, which are based on the analogue method. In Section 3, we compare our predictions on artificial datasets generated from the Lorenz 96 models [11] and a real dataset of solar irradiance. In Section 4, we discuss our results and conclude the paper.

2. Material and methods

2.1. Takens' embedding theorem

Takens' embedding theorem [1,2] provides information regarding the hidden status of a dynamical system. Using this theorem, we can reconstruct its attractor from a limited number of obser-

* Corresponding author.
E-mail address: tea-mountain@sat.t.u-tokyo.ac.jp (M. Chayama).

http://dx.doi.org/10.1016/j.physleta.2016.05.027
0375-9601/

vations. This theorem can be applied in the domain of time series predictions.

First, we consider a dynamical system in which Takens' embedding theorem can be applied. Suppose that we have a dynamical system $x(t) \in \mathbb{R}^m$ and an observational time series $s(t) \in \mathbb{R}$ given as follows:

$$\frac{dx(t)}{dt} = f(x(t)), \tag{1}$$

$$s(t) = g(x(t)), \tag{2}$$

where $f : \mathbb{R}^m \to \mathbb{R}^m$ represents the underlying dynamics, $g : \mathbb{R}^m \to \mathbb{R}$ is an observation function, and $T \in \mathbb{R}$ denotes a time delay. Then, we define a function $G_{d,T} : \mathbb{R}^m \to \mathbb{R}^d$ for obtaining delay coordinates as follows:

$$G_{d,T}(x(t-(d-1)T)) = (s(t-(d-1)T), s(t-(d-2)T), ..., s(t-T), s(t)), \tag{3}$$

where $d \in \mathbb{N}$ represents the embedding dimension. Takens demonstrated that if $d \geq 2m+1$, then it holds that $G_{d,T}$ and the derivative of $G_{d,T}$ are one-to-one, namely, the maps are injective on the attractor of f.

2.2. Analogue method

The analogue method was first proposed by Lorenz [6]. We define a set $N_{d,T}(t)$ of indices for the K nearest neighbors at a time t. Then, a direct prediction for p steps in the future is formulated as follows:

$$\hat{s}(t+p) = \frac{1}{|N_{d,T}(t)|} \sum_{l \in N_{d,T}(t)} s(l+p) = h(G_{d,T}(x(t-(d-1)T)), p), \tag{4}$$

where p denotes the number of prediction steps. Furthermore, we consider an iterative prediction. In the iterative prediction, we use one step prediction $\hat{s}(t+1)$ and feed it as an input repeatedly p times to generate p steps ahead prediction.

2.3. Multivariate method

In this paper, we consider a multivariate version of prediction. We suppose that a multidimensional dynamical system $\mathbf{x}(t)$ and an observed time series $\mathbf{s}(t)$ are defined by

$$\mathbf{x}(t) = (x_1(t), x_2(t), ..., x_m(t)) \in \mathbb{R}^m, \tag{5}$$

$$\mathbf{s}(t) = \left(s_{m_1}(t), s_{m_2}(t), ..., s_{m_l}(t)\right) \in \mathbb{R}^I. \tag{6}$$

In particular, these are given as

$$\frac{d\mathbf{x}(t)}{dt} = f(\mathbf{x}(t)), \tag{7}$$

$$\mathbf{s}(t) = g(\mathbf{x}(t)), \tag{8}$$

where $f : \mathbb{R}^m \to \mathbb{R}^m$ represents the underlying dynamics, and $g : \mathbb{R}^m \to \mathbb{R}^I$ is an observation function. Now, we define a different version of the function to obtain delay coordinates. By following [2,12], let $G_{d,T} : \mathbb{R}^m \to \mathbb{R}^{Id}$ be a function defined as follows:

$$\begin{aligned} G_{d,T}(\mathbf{x}(t-(d-1)T)) &= (s_1(t-(d-1)T), s_1(t-(d-2)T), ..., s_1(t-T), s_1(t), \\ &\quad s_2(t-(d-1)T), s_2(t-(d-2)T), ..., s_2(t-T), s_2(t), \\ &\quad ..., \\ &\quad s_I(t-(d-1)T), s_I(t-(d-2)T), ..., s_I(t-T), s_I(t)), \end{aligned} \tag{9}$$

where $T \in \mathbb{R}$ represents a time delay and $d \in \mathbb{R}$ a parameter for a delay dimension. According to [13], if $dI \geq 2m+1$, then the generic property holds that $G_{d,T}$ is one-to-one, and the derivative of $G_{d,T}$ is also one-to-one on the attractor of f. We define a set $N_{I,d,T}(t)$ of K nearest neighbors for a time t. Then, a prediction for p steps into the future is given as follows:

$$\hat{\mathbf{s}}(t+p) = \frac{1}{|N_{I,d,T}(t)|} \sum_{l \in N_{I,d,T}(t)} \mathbf{s}(l+p), \tag{10}$$

where p denotes the number of prediction steps.

3. Results and discussion

3.1. Lorenz 96 models

In this section, we present some results from numerical experiments by using artificial data. We compare univariate time series predictions with multivariate time series predictions by using five types of time series data. Dataset (i) was constructed from the Lorenz 96 I model [11] without observational noise, while dataset (ii) was constructed from the same model and contains observational noise, which was generated by 10% white noise. This value means the ratio between the standard deviations of noise-free signal and noise. In addition, dataset (iii) was constructed from the same model and contains dynamical noise. Datasets (iv) and (v) were constructed from the Lorenz 96 II model [11], and the latter contains observational noise generated by 10% white noise. The equations of the Lorenz 96 I model are given as follows:

$$\frac{dx_i}{dt} = x_{i-1}(x_{i+1} - x_{i-2}) - x_i + F \tag{11}$$

for $i = 1, 2, ..., n$, where the index i is cyclic. We set $F = 8$ and $n = 10$. In addition, the time step for integrating differential equations is fixed as $dt = 0.02$ if not mentioned. We observe x_i $(i = 1, 2, ..., n)$ to generate a time series. A given time series consists of the coordinate $x_i(t)$ with i fixed. We set $T = 0.02$ except for the discussion in Section 4.2. The equations of the Lorenz 96 II model are given as follows:

$$\frac{dx_i}{dt} = x_{i-1}(x_{i+1} - x_{i-2}) - x_i + F - \frac{hc}{b} \sum_{j=1}^{m} y_{i,j} \tag{12}$$

$$\frac{dy_{i,j}}{dt} = cby_{i,j+1}(y_{i,j-1} - y_{i,j+2}) - cy_{i,j} + \frac{hc}{b} x_i \tag{13}$$

for $i = 1, 2, ..., n$ and $j = 1, 2, ..., m$, where the index i and j are cyclic. We set $n = 8$ and $m = 4$. In addition, the time step for integrating differential equations is fixed as $dt = 0.002$. Moreover, the coefficients are set to $h = 1$ and $b = c = 10$. We observe $y_{i,j}$ $(i = 1, 2, ..., n,\ j = 1, 2, ..., m)$ to generate a time series. A given time series consists of the coordinate $y_{ij}(t)$ with i and j fixed as well as x_i for i fixed. We set $T = 0.002$. The dataset which contains dynamical noise are calculated by solving some stochastic differential equations. We calculate approximate solutions by the Euler–Maruyama method [14]. Dataset (iii) is given by the following equations:

$$x_{i+1}(t+dt) \approx x_i(t) + dt(x_{i-1}(t)(x_{i+1}(t) - x_{i-2}(t)) - x_i(t) + F) + \sqrt{dt} z_i(t), \tag{14}$$

$$z_i(t) \sim N(0, \sigma_i^2), \tag{15}$$

where $z_i(t)$ means a Gaussian random variable of mean 0 and standard deviation σ_i for the ith point at time t.