THE HONG KONG
POLYTECHNIC UNIVERSITY
香港理工大學
Pao Yue-kong Library
包玉剛圖書館

Copyright Undertaking

This thesis is protected by copyright, with all rights reserved.

By reading and using the thesis, the reader understands and agrees to the following terms:

1. The reader will abide by the rules and legal ordinances governing copyright regarding the use of the thesis.
2. The reader will use the thesis for the purpose of research or private study only and not for distribution or further reproduction or any other purpose.
3. The reader agrees to indemnify and hold the University harmless from and against any loss, damage, cost, liability or expenses arising from copyright infringement or unauthorized usage.

If you have reasons to believe that any materials in this thesis are deemed not suitable to be distributed in this form, or a copyright owner having difficulty with the material being included in our database, please contact lbsys@polyu.edu.hk providing details. The Library will look into your claim and consider taking remedial action upon receipt of the written requests.

Pao Yue-kong Library, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong

http://www.lib.polyu.edu.hk

CHARACTERISING NONLINEAR DETERMINISM FROM EXPERIMENTAL TIME SERIES DATA

by

Xiaodong Luo

A thesis submitted in partial fulfillment of the requirements for the degree of Master of Philosophy

Department of Electronic and Information Engineering
The Hong Kong Polytechnic University
December 2004

Pao Yue-kong Library
PolyU · Hong Kong

CERTIFICATE OF ORIGINALITY

I hereby declare that this thesis is my own work and that, to the best of my knowledge and belief, it reproduces no material previously published or written, nor material that has been accepted for the award of any other degree of diploma, except where due acknowledgement has been made in the text.

_____(Signed)

Xiaodong Luo (Name of Student)

ABSTRACT

This thesis includes two parts. In the first part, we will mainly introduce Takens' embedding theorem and a few nonlinear statistics which will be utilized throughout this thesis. Time delay embedding reconstruction based on Takens' embedding theorem is a powerful and vital tool to reconstruct the underlying system from a scalar time series, based on which estimation of invariant measures such as correlation dimension can be performed. In practical situations, however, we need to choose two suitable parameters, i.e., embedding dimension and time delay, to properly apply this technique. We will review some popular criteria on the choice of these two parameters, and describe a new method we have developed to choose suitable time delay for a continuous dynamical system. We compare our algorithm with several available algorithms (for example, the average mutual information criterion) and found that our algorithm has a satisfactory performance, while the implementation is simple and its computational cost is fairly low. Hence this algorithm is suitable to apply to situations such as surrogate tests, where a large amount of data might be used.

In the second part, we will introduce the technique of surrogate tests to detect possible nonlinear determinism (or nonlinearity). An irregular time series

in practice can be produced either from a stochastic process or from a nonlinear deterministic system. To understand the underlying system of the time series, the first step shall be to investigate whether the irregularity is brought by stochasticity or by nonlinearity (often chaos), only then can corresponding strategies for further analysis be properly applied. In this part, we will mainly introduce surrogate algorithms to detect nonlinear determinism for time series from unknown dynamical systems. To apply surrogate techniques on a time series for nonlinearity detection, we need to adopt a null hypothesis, which usually supposes the time series is generated by a linear stochastic process and potentially filtered by a nonlinear filter. Based on this null hypothesis, a large number of data sets (surrogates) are to be produced from the original time series, which in principle keeps the linearity of the original time series while destroying all other structures. We then calculate some nonlinear statistics (discriminating statistics), for example, correlation dimension, of both the original time series and the surrogates. If the discriminating statistics of the original time series deviate from those of the surrogates, then we can reject the null hypothesis we proposed and claim that the original time series is deterministic with a certain confidence level (depending on how many surrogates we have generated). After the detection of nonlinearity, we are also interested in examining whether the time series is pseudoperiodic or chaotic (This distinction is important in some situations, for example, cardiac disease diagnosis, where heart

rate data are believed to be chaotic for healthy patients but indicate regularity for those with congestive heart failure). We propose a further null hypothesis, i.e., we assume the time series is pseudoperiodic rather than chaotic. We will present a new algorithm to generate surrogates for pseudoperiodic time series, then by choosing the correlation dimension as the discriminating statistic, we can distinguish chaotic time series from pseudoperiodic time series. As an application of the surrogate tests, we will apply this technique to some experimental data sets.

Research works covered in this thesis

Journal papers

- X. Luo and M. Small. "Geometric measures of redundance and irrelevance tradeoff exponent to choose suitable time delays for continuous systems" Submitted to I.J.B.C (2005)

- X. Luo,T.Nakamura and M. Small. "Surrogate test to distinguish between chaotic and pseudo-periodic time series" Phys. Rev. E 71, 026230 (2005)

- T. Nakamura, X. Luo and M. Small. "Topological test for chaotic time series." Submitted to Phys. Rev. Lett. (2004)

Conference papers

- X. Luo and M. Small. "A new algorithm to choose time delay for time delay embedding reconstruction" *Dynamics Days Asia-Pacific* (Singapore, 30 June-2 July 2004)

- X. Luo, T. Nakamura and M. Small. "Surrogate generation algorithm for pseudo-periodic time series" *International Symposium on Nonlinear Theory and its Applications* (Fukuoka, Japan, 29 November-3 December 2004)

Contents

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGMENTS

First of all, of course, I must thank my parents for their support over so many years. I sincerely dedicate this dissertation to them.

As my supervisors, Prof. Chi K. Tse and Dr. Michael Small deserve my deep appreciation. I should thank Prof. Tse not only for his generous support and help, but also for his effort, as the founder and leader of our research group, to provide us research members lots of opportunities to communicate and exchange ideas with outside researchers. I should especially thank Dr. Small for his instructions on my research study. During the past two years, the frequent discussions between us greatly enhanced my knowledge and inspired many new ideas. For the similar reasons, I should also express my gratitude to Dr. Tomomichi Nakamura.

In addition, I am grateful to my colleagues and friends at the Hong Kong Polytechnic University, with whom I spent a cherishable period in Hong Kong.

Thank you everyone!

CHAPTER 1

INTRODUCTION

A sequence of measurements as time evolves is called a time series [22]. We can think of a time series as the outputs from an underlying system in terms of signal processing. However, it is often the case that we do not know the explicit descriptions of the underlying system *a priori*, hence we have to analyze the time series instead to extract information of the underlying system. Usually this is not an easy task, especially for complex systems such as the human heart, in which nonlinearities are often involved. We may find spurious results if extending well established linear analysis techniques to nonlinear fields. This fact reminds us to be more cautious with our familiar linear methods if we attempt to apply them to analyze nonlinear time series. An alternative choice is to apply techniques within the context of nonlinear dynamics and chaos theory, which will be called nonlinear time series analysis in this thesis. The hallmark of nonlinear time series analysis is to present data to be analyzed in phase space (or state space) rather than in a time or frequency domain [20]. Usually phase space is spanned by the state variables (and possibly, their derivatives). If the ensemble of data points in phase space is invariant under the dynamics, towards which neighboring states in a given basin of attraction asymptotically approach in the course of dynamic evolution [58], we call such ensembles attractors. A characteristic description of attractors of dynamical systems, especially for those with fractal structures, is the fractal dimension (or

Hausdorff dimension) [10]. In practical situations, however, due to the difficulty in calculation, an alternative statistic, namely correlation dimension, is proposed as a substitution for the Hausdorff dimension [14]. This will be introduced in a later chapter.

Since usually we do not know the system equations *a priori*, a problem naturally arising is that, how could we obtain the attractor of the underlying system in phase space only from our observations? One solution is to seek a diffeomorphism of the attractor in a reconstructed state space instead of the original one in the underlying system's original phase space. The embedding theorems proven by Whitney and Takens [59] guarantee the uniqueness and equivalence of the reconstructed attractor to the original one in an embedding space (to be introduced below). Therefore we can equivalently describe the underlying system's characteristic behavior by reconstructing the attractor from a scalar time series [1]. With the reconstructed attractors, we can then apply nonlinear techniques such as dimension analysis for further investigation.

The motivation of this thesis is to introduce the application of the surrogate method, together with the method of dimension analysis, for nonlinearity detection to our observed data. We can often observe the irregularity in measurements, which is caused either by stochasticity or by nonlinearity (often chaos) in deterministic systems. To distinguish between stochasticity and nonlinearity is important as they indicate quite different directions in research. The idea to apply the surrogate

[1] In many cases a scalar time series might be sufficient. However, in some situations multivariate time series are needed to provide more information for reconstruction.

test technique for nonlinearity detection is that, we assume a stationary irregular time series [2] is generated from a linear stochastic process (and potentially filtered by some nonlinear filter), based on this assumption we can generate many data sets (the surrogates) from the original time series, which keeps the linearity in the original time series while destroying all other structure. By comparing the correlation dimension (the discriminating statistic) of the original time series and the surrogates, we can determine whether our assumption is likely to be true. After nonlinearity in the time series is detected, a further investigation of the underlying system can be performed to determine whether the time series is possibly chaotic.

The organization of the thesis is as the following: In Chapter 2 we will introduce the embedding theorems. We will also discuss the problems on how to choose proper embedding parameters in practical situations and present some criteria on parameter choice, including one proposed by us in a recent work on the choice of suitable time delay. In Chapter 3 we will introduce the concept of correlation dimension and the calculation algorithm. Another statistic to appear is the one step local prediction error [11], arising as a reflection of reconstruction quality in Chapter 2. With the necessary knowledge introduced in the previous two chapters, in Chapter 4 we will apply the surrogate test method, with the discriminating statistic of correlation dimension, to detect nonlinearity in a time series. We will also propose a new surrogate generation algorithm for pseudoperiodic time series, which can be applied to distinguish between chaos and pseudoperiodicity in a time

[2] See Chapter 4 for the definition of stationarity.

series after the detection of nonlinearity. An application of these techniques to experimental data will also be demonstrated at the end of Chapter 4. Finally, in Chapter 5, we will have a summary of the whole thesis.

CHAPTER 2

STATE SPACE RECONSTRUCTION FOR NONLINEAR TIME SERIES ANALYSIS

2.1 Overview

In general, a finite dimensional dynamical system M can be described by the n first-order ordinary differential equations as shown in Eq. (2.1),

$$\begin{cases} \dot{\xi}_1 = F_1\left(\xi_1, \xi_2, \ldots, \xi_n\right); \\ \dot{\xi}_2 = F_2\left(\xi_1, \xi_2, \ldots, \xi_n\right); \\ \qquad\qquad \vdots \\ \dot{\xi}_n = F_n\left(\xi_1, \xi_2, \ldots, \xi_n\right). \end{cases} \tag{2.1}$$

The n dimensional space consisting of the state variables $(\xi_1(t), \xi_2(t), \ldots, \xi_n(t))^T$ is called *state space* or *phase space*, where t is the time index.

Usually we will introduce a measurement function h to observe the system M, for example, we typically obtain the experimental data sets via some apparatus. Hence we actually obtain the transformed data sets

$$\left\{ X(t) \equiv (x_1(t), x_2(t), \ldots, x_p(t))^T : X(t) = h(\Gamma(t)) \right\}, \tag{2.2}$$

where $\Gamma(t) = (\xi_1(t), \xi_2(t), \ldots, \xi_n(t))^T$. This process can be described as $h : \mathbb{R}^n \rightarrow \mathbb{R}^p$, i.e., h maps vectors $\Gamma(t)$ in n dimensional state space to vectors $X(t)$ in p dimensional space. For the case of $p = 1$, we obtain scalar measurements of

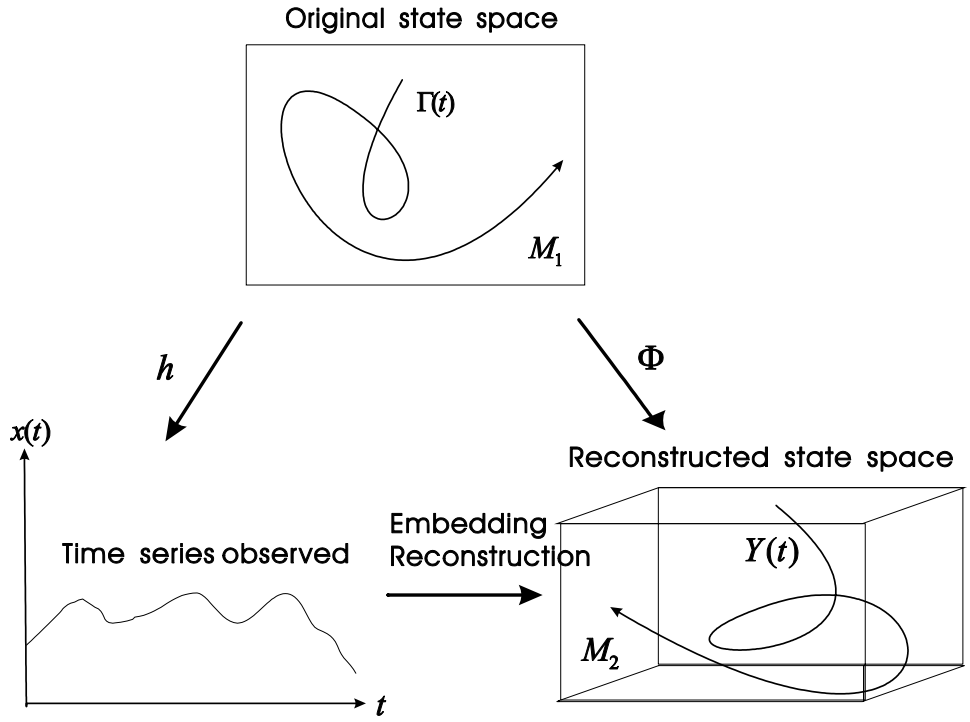

Figure 2.1: Schematic representation of the embedding reconstruction problem.

the time series, as shown in Fig. 2.1. We will only consider scalar time series throughout this thesis, the reason is that, on one hand some times we cannot measure several variables simultaneously, or the dimension n is too high to conduct a feasible measurement (for example, systems with infinite dimensions). On the other hand, thanks to Whitney's and Takens's embedding theorems [59] to be introduced below, we often need only measure one state variable of the dynamical system. With a proper embedding reconstruction of scalar time series from this state variable, the attractor thus reconstructed can sufficiently characterize the original attractor in state space. Hence, the main task left for us is to determine how to choose suitable parameters for an embedding reconstruction, which shall preserve the characteristic properties of the original dynamical system.

In the next sections, we will introduce the embedding theorems of Whitney and Takens. We will also introduce the algorithms to choose suitable parameters for the time delay embedding reconstruction, including a new algorithm we have proposed recently to choose suitable time delays.

2.2 The Embedding Theorems

The embedding reconstruction is a construction process going from a scalar time series in time domain to a multivariate attractor in phase space. As shown in Fig. 2.1, we apply a measurement function h to the dynamical system M_1 and obtain a scalar time series $\{x_i\}$. We want to reconstruct an attractor from scalar time series $\{x_i\}$ which is embedded in the manifold M_2 and can characterize the original attractor of the system M_1. This process is denoted by the map $\Phi : M_1 \to M_2$, which means $\Phi(M_1)$ is the submanifold embedded in the manifold M_2. Notice that, usually the phase space reconstructed by the time series from an unknown system is not the true (original) phase space, i.e., $\Phi(M_1) \neq M_1$. Their topological dimensions will typically also be different, i.e., $\dim(\Phi(M_1)) \neq \dim(M_1)$. Nevertheless, by properly reconstructing the phase space, we may infer the behavior of the original system from that of the reconstructed attractor. Hence in this sense we can capture the characteristic behavior of the original system.

Not all maps $\Phi : M_1 \to M_2$ can lead to a good reconstruction. To be a good reconstruction we require that a map Φ shall be one-to-one. In addition, its derivative map $D\Phi$ shall also be one-to-one, where D denotes the *derivative operator*. The reason to require Φ to be a one-to-one map is that, we expect the

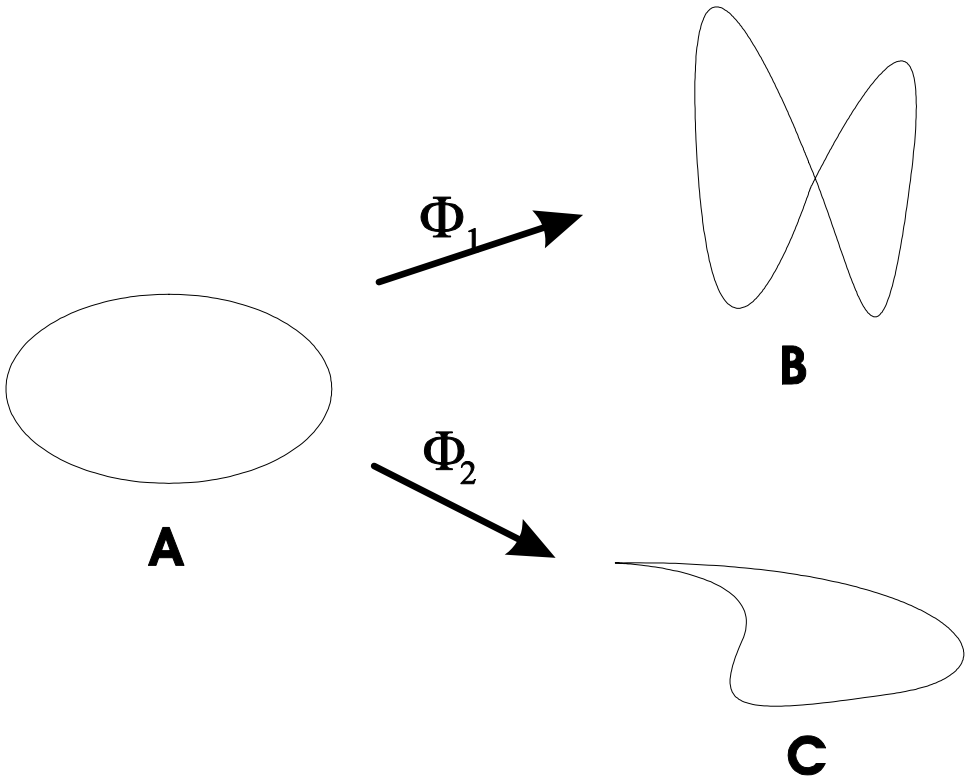

Figure 2.2: Demonstration of two situations under which the reconstructions cannot be embeddings. Attractor A is the original orbit. Attractor B is the reconstructed orbit by map Φ_1 with the appearance of self-intersection. Attractor C is the reconstructed orbit by map Φ_2 which does not preserve the differential structure of the original orbit. This graph is plotted with some modifications according to the one in [13].

dynamical system in the reconstructed space to be unique, any self-intersection appearing through a map, for example, the situation brought by the map Φ_1 in Fig. 2.2, is not desired. But uniqueness is not sufficient to guarantee the capture of the characteristic behavior of the original system, we further require the derivative map $D\Phi$ to be one-to-one, which will preserve the linearization of the dynamics at any points in the state space [20]. A map satisfying the above two conditions is called an embedding. In the following we will introduce the theorems of Whitney and Takens on embedding issues (and the corresponding strengthened versions by Sauer, Yorke & Casdagli [59]).

Suppose manifold M_1 is a d dimensional topological space $\mathbb{R}^d$, through the map $\Phi : M_1 \to M_2$ we want to embed M_1 into another manifold M_2. Whitney [59] proved a theorem stating that:

Theorem 1 (Embedding theorem of Whitney) $\left\{Embedding\ \Phi : M_1 \to \mathbb{R}^{2d+1}\right\}$ *is an open and dense set in the space of smooth maps.*

From the above statement we see that M_1 can be embedded in any manifold M_2 as long as that the topological dimension of M_2, denoted $\dim(M_2)$, $\geqslant 2d+1$.

Whitney's embedding theorem is useful in that it indicates a way through which we can achieve an embedding. However, stating in "topology language" makes this theorem difficult to be applied in practice. We might not be sure whether "an open and dense set" is a meaningful assertion in practice because observables in "an open and dense set" could still have a rather low probability to be observed. In addition, Whitney's embedding theorem requires that manifold

M_1 shall be smooth enough. This requirement is actually not compulsory. In a work of Sauer, Yorke & Casdagli [59], they proved a strengthened theorem saying that, for a compact fractal set M_1 with fractal dimension d_0, the following theorem holds,

Theorem 2 (Enhanced embedding theorem of Whitney) *Φ chosen arbitrarily from the set $\left\{\Phi : M_1 \to \mathbb{R}^{ceil(2d_0)}\right\}$ is an embedding with probability 1.*

The above statement can be interpreted as that, even for a compact fractal set M_1 (not necessary to be a smooth manifold) with fractal dimension d_0, an arbitrary map $\Phi : M_1 \to \mathbb{R}^{ceil(2d_0)}$ can almost always lead to an embedding if we choose the embedding dimension to be at least $ceil(2d_0)$, where $ceil(x)$ denotes the smallest integer larger than x.

Although Whitney's embedding theorem and its enhanced version can indeed tell us how to achieve an embedding, there is still a technical problem remaining. Recall the situation depicted in Fig. 2.1, usually we do not know the manifold M_1 exactly (otherwise there is no need to seek an embedding map), hence, we cannot directly construct the map $\Phi : M_1 \to \mathbb{R}^m$, where $m(\geqslant ceil(2d_0))$ is the embedding dimension. But notice that, actually we do have much information of the unknown system M_1 which is obtained by measuring the dynamical system via a measurement function h, however, such information ($\{x_i : x_i = h(\Gamma(t_i))\}$ in concretion), is the projection of the dynamical system M_1 onto a lower dimensional space. The information is somewhat distorted and unreliable. We need to remove the distortion and resolve the information from scalar time series. One idea is

that, as indicated in Fig. 2.1, from scalar time series $\{x_i\}$ we produce a set of vectors $\left\{y_i : y_i = \left(x_i, x_{i+\tau}, ..., x_{i+(m-1)\tau}\right)^T\right\}$ which span a subspace M_2 of $\mathbb{R}^m$, where parameter m and τ are called embedding dimension and time delay respectively. If the equivalent map $\Phi : M_1 \to M_2$ proves to be an embedding, where $\Phi = \Xi h$ with Ξ denoting the map $\Xi : \{x_i\} \to \left\{\left(x_i, x_{i+\tau}, ..., x_{i+(m-1)\tau}\right)^T\right\}$, then we have found out a convenient way to (indirectly) embed the original system M_1 into a m dimensional topological space. The process $\Xi : \{x_i\} \to \left\{\left(x_i, x_{i+\tau}, ..., x_{i+(m-1)\tau}\right)^T\right\}$ is called time delay embedding reconstruction. In 1981, Takens proved an embedding theorem related to this process (the same result is also obtained by Mañé, see Ref. [59]), which is stated as follows:

Theorem 3 (Embedding theorem of Takens) $\left\{$*Embedding* $\Phi_{\Xi h} : M_1 \to \mathbb{R}^{2d+1}\right\}$ *is an open and dense set in the space of smooth maps* (h, Ξ) [1].

It is not surprising to find that the statement of Takens' embedding theorem is almost the same as that of Whitney's embedding theorem. The meaning of Takens' embedding theorem is that, through this theorem, we know we have a good chance to form an embedding $\Phi : M_1 \to M_2$ within certain regions in the space of smooth maps (h, Ξ) by generating a vector field $\left\{\left(x_i, x_{i+\tau}, ..., x_{i+(m-1)\tau}\right)^T\right\}$. But this theorem also has the same limitations as those mentioned above. Again, Sauer, Yorke & Casdagli [59] proved an enhanced version of this theorem:

Theorem 4 (Enhanced embedding theorem of Takens) $\Phi_{\Xi h}$ *chosen arbitrarily from the set* $\left\{\Phi_{\Xi h} : M_1 \to \mathbb{R}^{ceil(2d_0)}\right\}$ *is an embedding with probability* 1.

[1] Actually, the map Γ, which governs the evolution trajectory of manifold M_1, is implicitly included without affecting our discussions. See [13] for more details.

Similarly, we see that, via map $\Xi : \{x_i\} \rightarrow \left\{\left(x_i, x_{i+\tau}, ..., x_{i+(m-1)\tau}\right)^T\right\}$, in ideal situations $\Phi_{\Xi h}$ will almost always be an embedding provided that embedding dimension $m \geqslant ceil(2d_0)$.

We give some final remarks on the embedding issues without involving the details. For more information, please see the works mentioned below and the references therein.

- *The requirement that $m \geqslant ceil(2d_0)$ is a sufficient but not necessary condition. In some situations, we can still obtain an embedding map $\Phi_{\Xi h}$ even only condition $m \geqslant ceil(d_0)$ satisfies.*

- *Time delay embedding reconstruction $\Xi : \{x_i\} \rightarrow \left\{\left(x_i, x_{i+\tau}, ..., x_{i+(m-1)\tau}\right)^T\right\}$ is not the only method to construct an embedding. Available alternatives include, for example, filtered time delay embedding [42], derivative embedding [34], interspike interval embedding [41] and their hybrids.*

2.3 Algorithms to Choose Embedding Parameters

As indicated above, Takens' embedding theorem is a powerful tool which allows us to conveniently conduct an embedding of the original system based only on a set of scalar observations. Unfortunately, this theorem only works under the ideal situations, where we assume the data set obtained is infinitely long and is measured with infinite precision. However, such assumptions cannot be true in practical situations. We have to accept the fact that the data set we obtained is only an approximation to that from the original system, which contains only a

part of the whole evolution states and is contaminated by noise from digitization process, quantization process and so on. Nevertheless, time delay reconstruction is still a useful tool based on which many nonlinear analysis techniques are developed, although there are some differences from that in the ideal situation. For example, originally from (the enhanced) Takens' embedding theorem we learn that for an arbitrary time delay τ, the map $\Phi_{\Xi h}$ will be an embedding with probability 1 as long as the embedding dimension m is large enough. However in practical situations, too small or too large time delays can no longer lead to an embedding for a chaotic system [2], as we shall introduce below. Hence we have to be careful to choose the time delay τ. Moreover, as often encountered in reality, we cannot know *a priori* the fractal dimension of an unknown system, therefore we also have to investigate to what extent an embedding dimension m can be considered large enough to produce an embedding. Consequently, we will divide our following discussions into two parts. One is focused on the algorithms to choose proper time delay. The other, of course, is dedicated to the introduction to the algorithms to determine suitable embedding dimension.

2.3.1 Algorithms to Choose Time Delay

[2] In reality, no measurement function will be smooth enough, hence the conditions for Takens' embedding theorem actually cannot rigorously hold, i.e., time delay reconstruction cannot be an embedding in the strictest mathematical sense. But this does not mean time delay reconstruction is not useful, instead it is a valid representative of the underlying system from the discrete observations we have obtained. For convenience, we will still call a proper time delay reconstruction as an embedding, but readers should bear in mind this limitation.

The Criteria of Autocorrelation Function and Other Higher Order Statistics

An often used criterion at the early development stage of nonlinear time series analysis is the criterion of the second order autocorrelation function ($SOAC$). The autocorrelation function for a scalar time series $\{x_i\}$ is defined by

$$A(\tau) = \langle x_i x_{i+\tau} \rangle - \langle x_i \rangle^2 , \tag{2.3}$$

where $\langle x_i \rangle$ denotes the expectation of data set $\{x_i\}$. Initially the criterion is to choose τ as the time delay when $A(\tau)$ first drops to zero, since when $A(\tau) = 0$, elements x_i and $x_{i+\tau}$ are linearly uncorrelated. Some authors [2] found that although this algorithm works well in some situations (as an example, the Rössler system), it may fail to yield a suitable time delay for some dynamical systems, for instance, the Lorenz system [26]. They suggested an alternative criterion to choose the time delay at the moment when $A(\tau)$ first drops to $1/e$ of its initial value $A(0)$. This alternative method is more robust to obtain the suitable time delay for dynamical systems (such as the Lorenz system), however, no evidence shows that $1/e$ is a universal factor of the autocorrelation function criterion to yield suitable time delays for, if not all, most of the dynamical systems.

As an extension of the above idea, Albano *et al.* [3] proposed to take the time at the consistent extrema (say, local minima or maxima) of several different higher order autocorrelation functions, for example $\langle x_i x_{i+\tau} x_{i+2\tau} \rangle$ and $\langle x_i^2 x_{i+\tau} \rangle$, as the candidates of suitable embedding windows [3]. The practical performance of

[3] Embedding window is usually defined as $\tau_w = (m-1)\tau$ or $\tau_w = m\tau$. Therefore if we know the embedding dimension m, we know the time delay as well.

this criterion proves to be good, but its theoretical basis is not strongly grounded.

The Criterion of Average Mutual Information

In Eq. (2.3) the *SOAC* $A(\tau)$ is a linear measure of the correlation between subsets $\{x_i\}_{i=1}^{N-\tau}$ and $\{x_{i+\tau}\}_{i=1}^{N-\tau}$ of a time series $\{x_i\}_{i=1}^{N}$ [4], therefore for a nonlinear dynamical system, there is a potential blindness of function $A(\tau)$ to detect the nonlinear correlation between subsets$\{x_i\}_{i=1}^{N-\tau}$ and $\{x_{i+\tau}\}_{i=1}^{N-\tau}$. With this consideration, Fraser and Swinney [12] applied an important statistic, namely *mutual information*, to measure the nonlinear correlation. The definition goes as follows: Let $p(x)$ denote the probability distribution of a data set $\{x_i\}$, according to Shannon's information theory, we can calculate the *information entropy*

$$H(x) \equiv -\sum_i p(x_i) \log p(x_i). \tag{2.4}$$

Mutual information $I(x, y)$ between two data sets $\{x_i\}$ and $\{y_i\}$ is defined by [37]

$$I(x, y) \equiv H(x) + H(y) - H(x, y), \tag{2.5}$$

where $H(x, y)$ is the joint information entropy between $\{x_i\}$ and $\{y_i\}$ with a joint probability distribution $p(x, y)$, i.e.,

$$H(x, y) \equiv -\sum_i p(x_i, y_i) \log p(x_i, y_i). \tag{2.6}$$

Mutual information $I(x, y)$ can tell us how much information we can learn about the data set $\{y_i\}$ if we already know the data set $\{x_i\}$. For a scalar time series, we shall instead calculate the *average mutual information (AMI)* $I(x_i, x_{i+\tau})$

[4] Previously, we ignore the data set size N of time series $\{x_i\}_{i=1}^{N}$ since it does not affect our discussion, for convenience we will continue to use the form $\{x_i\}$ later if it causes no confusion.

between subsets $\{x_i\}_{i=1}^{N-\tau}$ and $\{x_{i+\tau}\}_{i=1}^{N-\tau}$ of the original time series $\{x_i\}_{i=1}^{N}$. The criterion is to choose time delay at the first local minimum of $I(x_i, x_{i+\tau})$, where a local minimum nonlinear correlation (or dependence) between subsets $\{x_i\}_{i=1}^{N-\tau}$ and $\{x_{i+\tau}\}_{i=1}^{N-\tau}$ can be achieved. Mutual information is a valuable concept, it provides us with a proper tool to deal with time series in a nonlinear way. But the minor disadvantages of this algorithm are that, the calculation of *AMI* involves a complex implementation because of the difficulty in calculating joint probability distribution $p(x_i, x_{i+\tau})$. As a common problem, both the length of a data set and the presence of noise will have great influences on the computation of joint probability distribution. Hence, it is natural to expect that the *AMI* algorithm will not be very robust for small noisy data sets, as observed in [30]. In addition, it has been found that for some dynamical systems, the *AMI* measure will decay monotonically and show no minima at all, in which cases we cannot select any time delay through this criterion. We will propose a possible remedy for this situation later, which is based on the idea of tradeoff between redundancy and irrelevance.

The Criterion of Fill-factor

When examining the reasons for the failure of the *SOAC* criterion, we see some comments such as "a zero of the autocorrelation indicates on average linear independence of $x(t)$ and $x(t+\tau)$, but not necessarily of $x(t)$ and $x(t+2\tau)$" and " The deeper reason for this failure is the fact that this approach does not employ any state space information" [5]. Indeed, usually the information-like criteria such the

[5] See ref. [13] (p87) and references therein for more details.

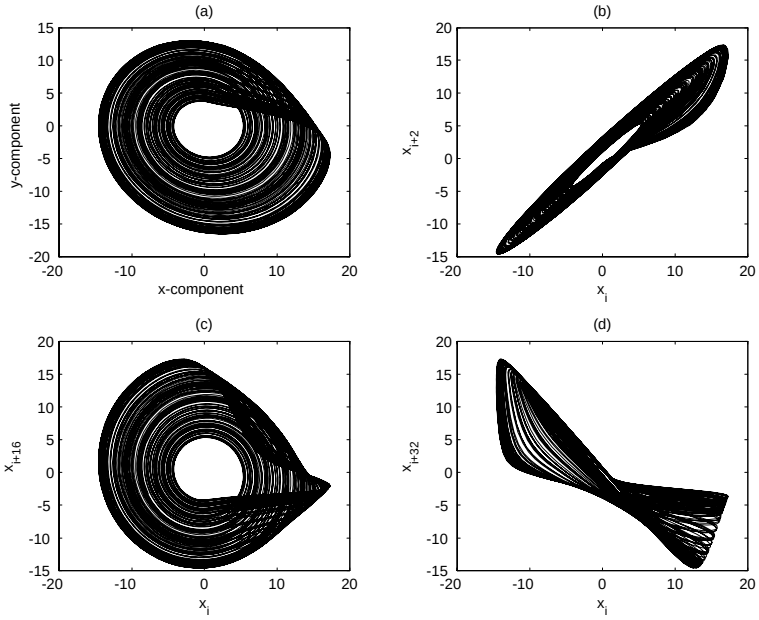

Figure 2.3: (a) Projection of the Rössler attractor onto the x-y plane; (b), (c) and (d): Reconstructed attractor of the Rössler system in two dimensional embedding space $x_{i+\tau}$ vs x_i with $\tau = 2$, $\tau = 16$ and $\tau = 32$ respectively.

SOAC and the *AMI* do not utilize any geometric information of the reconstructed attractors in embedding spaces [6]. Comparatively, Buzug and Pfister [6] proposed the *fill-factor algorithm* to determine a suitable time delay by examining the reconstructed attractors expansion in embedding spaces. The authors noticed that if a dynamical system has only one unstable focus, choosing too small or too large delay time will lead to the collapse of the reconstructed attractor in embedding space. Therefore their criterion is to choose time delay when the volume of the reconstructed attractor in hyperspace is maximized. Take the Rössler system (See

[6] This situation is understandable. In fact, the information criteria tend to determine a suitable time delay τ without involving embedding dimension m as the parameter. Comparatively, criteria utilizing reconstruction information in embedding spaces will inevitably take into account the affect of different embedding dimensions, thus such criteria are often related to choosing suitable embedding windows, as to be shown below.

Eq. (2.18) for system equations) as an example, as shown in Fig. 2.3, panel (a) indicates the projection of the Rössler system onto the $x-y$ plane, while panel (b), (c) and (d) demonstrate reconstructed attractors in two dimensional embedding space (e.g., $x_{i+\tau}$ vs. x_i) with different time delays, from which we see that for time delay $\tau = 2$ or 32 ("improper" choices), the corresponding attractors have less volumes than that of the attractor with time delay $\tau = 16$.

To implement this algorithm, we need to randomly choose $m+1$ delay vectors on the reconstructed attractor in metric space $\mathbb{R}^m$ and calculate the volume $V(m,\tau)$ of the corresponding hyper-parallelepiped, repeat this procedure for many times and we obtain an average $\langle V(m,\tau)\rangle$ of these volumes. The fill-factor is defined by

$$\gamma(m,\tau) \equiv \log_{10} \frac{\langle V(m,\tau)\rangle}{(\frac{1}{2}(\max(x_i) - \min(x_i)))^m}, \tag{2.7}$$

where $\max(x_i)$ and $\min(x_i)$ denote the maximum and the minimum of time series $\{x_i\}$ respectively. This is a simple and effective method to choose time delay for systems with only one unstable focus, however, it will fail to yield significant delays for dynamical systems with more than one unstable focus, for example, the Lorenz system.

The Criterion of Integral Local Deformation

Realizing the limitation of the fill-factor algorithm, Buzug and Pfister [6] suggested an alternative algorithm, namely *integral local deformation* (ILD), to overcome the difficulties we introduced previously. The essential idea of this algorithm

is that, for a deterministic system, false neighbours [7] will usually no longer be neighbours as time evolves ahead, while true adjacent points will have approximately the same direction of evolution within a short period. A good reconstruction shall (on average) have a minimum deviation between evolution directions of adjacent points. The detail in implementing this algorithm is presented below.

We first randomly choose a delay vector $\overrightarrow{X_i} = \left(x_i, x_{i+\tau}, \ldots, x_{i+(m-1)\tau}\right)^T$ as a reference point in m dimensional embedding space, then search and determine the ensemble $\left\{\overrightarrow{X_j} : \left\|\overrightarrow{X_j} - \overrightarrow{X_i}\right\| < \epsilon\right\}$, which consists of the delay vectors falling within a hyper-sphere of radius ϵ centered by $\overrightarrow{X_i}$, where $\|\cdot\|$ denotes the metric distance in Euclidean space. The *mass center* $\overrightarrow{M_i}$ of the ensemble can thus be calculated (with equal weight for each element $\overrightarrow{X_j}$), and the distance d_i between the reference point $\overrightarrow{X_i}$ and the mass center $\overrightarrow{M_i}$ will be $d_i = \left\|\overrightarrow{X_i} - \overrightarrow{M_i}\right\|$. After p step evolution, the reference point $\overrightarrow{X_i}$ and the original ensemble $\left\{\overrightarrow{X_j} : \left\|\overrightarrow{X_j} - \overrightarrow{X_i}\right\| < \epsilon\right\}$ become $\overrightarrow{X_{i+p}}$ and $\left\{\overrightarrow{X_{j+p}} : \left\|\overrightarrow{X_j} - \overrightarrow{X_i}\right\| < \epsilon\right\}$ respectively, then we need to locate the new mass center position $\overrightarrow{M_{i+p}}$ of the new ensemble $\left\{\overrightarrow{X_{j+p}} : \left\|\overrightarrow{X_j} - \overrightarrow{X_i}\right\| < \epsilon\right\}$. Therefore the corresponding distance d_{i+p} turns to be $\left\|\overrightarrow{X_{i+p}} - \overrightarrow{M_{i+p}}\right\|$. We denote the difference between distance d_{i+p} and d_i as $\Delta d(m, \tau, p, i) = d_{i+p} - d_i$. The average *integral local deformation* is defined by

$$\delta(m, \tau, p) = \frac{\left\langle \sum_{q=1}^{p} \left(\Delta d(m, \tau, q-1, i) + \Delta d(m, \tau, q, i)\right) \right\rangle}{2(\max(x_i) - \min(x_i))}, \tag{2.8}$$

[7] For the concept of false neighbour, see a full description in the next subsection, where we will use the criterion of false nearest neighbour to choose suitable embedding dimension.

where operator $\langle\cdot\rangle$ denotes expectation taken over i, which means we shall choose different reference vector for many times, and functions $\max(x_i)$ and $\min(x_i)$ have the same meaning as that described in Eq. (2.7).

If we choose to observe one step evolution, then we can define a simpler measure of the average local deformation, which governs the relative growth of distances between reference points and the corresponding mass centers in the logarithm scale, i.e.,

$$\delta(m,\tau) = \left\langle \ln \frac{d_{i+1}}{d_i} \right\rangle . \tag{2.9}$$

The advantage of the ILD algorithm is that, by examining the evolution of the reference points and their neighbours in local flows, the statistic $\delta(m,\tau,p)$ (or $\delta(m,\tau)$) can provide us dynamical information of the underlying system, however, the corresponding computational cost will substantially increase compared to other algorithms. In addition, as we are to show below [27], the algorithm based on Eq. (2.9) might be more suitable to determine the embedding window rather than to determine embedding dimension m and delay time τ separately.

The Criterion of Redundancy and Irrelevance Trade-off Exponent

Recently we proposed a new method [27] to choose time delay based on the concepts of redundancy and irrelevance [7], which is very effective in computation but can achieve roughly equivalent performance to the previous criteria.

In a very recent paper Cellucci and coworkers [8] state their viewpoint on previous embedding methods as: *A circular logic has resulted in which embedding*

criteria are assessed by an adjudicating criterion which is itself an embedding criterion. Following this viewpoint, we learn that the best embedding criterion might differ under different adjudicating criteria. Hence we would like to elucidate that we do not seek the best embedding criterion for different adjudicating criteria, instead we attempt to seek time delays through which the delay reconstructions can equivalently characterize the original dynamical systems (to be verified later).

Suppose we apply a smooth measurement function h to measure a manifold M and obtain a N data point scalar time series $\{x_i\}_{i=1}^{N}$. We want to reconstruct an attractor in m dimensional embedding space $\mathbb{R}^m$ from the scalar time series which characterizes the original one on manifold M. According to Takens' embedding theorem, a sufficiently high embedding dimension m will almost always lead to a time delay embedding reconstruction for any time delay τ [8]. In practice, however, the situation will be different. Due to the non-smoothness of the measurement function h (in the processes such as digitization and quantization, for example), we can no longer achieve an embedding reconstruction in a rigorous sense. However, delay reconstruction, as a valid representative of the underlying system if properly constructed from the discrete observations, is still a useful tool in most practical cases, based on which further steps in data processing such as invariant statistic analysis (dimension, entropy etc.) can proceed. By "properly constructed", we mean that, for the ensemble of discrete observations E (obtained by measuring manifold M), the reconstruction map $\Psi : E \to U$ shall be one-to-

[8] For Takens' embedding theorem to hold, there shall be some additional requirements (see [13] for details).

one, where U is a sub-manifold embedded in $\mathbb{R}^m$. Moreover, its *derivative mapping* $D\Psi$ shall also be one-to-one, where D denotes the differentiating operator on Ψ. Note that, because of the finiteness of the time series and the presence of noise, a sufficiently high embedding dimension turns out to be necessary but not sufficient to produce a proper delay reconstruction (for convenience, we will use "delay reconstruction" or "reconstruction" instead of "proper delay reconstruction" when causing no confusion). Also unlike the ideal situation, not all of the time delays can lead to a reconstruction (to be shown below).

We shall emphasize that, although in practice some time delays no longer lead to a reconstruction, it is expected the remaining time delays shall equivalently lead to a reconstruction in the sense of characterizing the original attractor in state space, although some particular values might indeed facilitate certain analysis of a time series. Hence, even if we later obtain different time delays from different criteria, we still consider them equally as possible candidates for a reconstruction.

Now let us first consider the effects of different time delays on the reconstructed attractor in an embedding space. From a time series $\{x_i\}$, we want to construct a set of delay vectors $\overrightarrow{X_i} = (x_i, x_{i+\tau}, \ldots, x_{i+(m-1)\tau})^T$ in a m dimensional embedding space. For convenience, we confine our argument within a two-dimensional embedding space $x_{i+\tau}$ vs. x_i. Fig. 2.4 illustrates the projection of the original attractor onto the $x-y$ plane and the reconstructed attractors of the Lorenz system (see Eq. (2.17)) from the x component for three different time delays. When τ is too small, $x_{i+\tau}$ will be very close to x_i due to the continuity of a manifold. Therefore

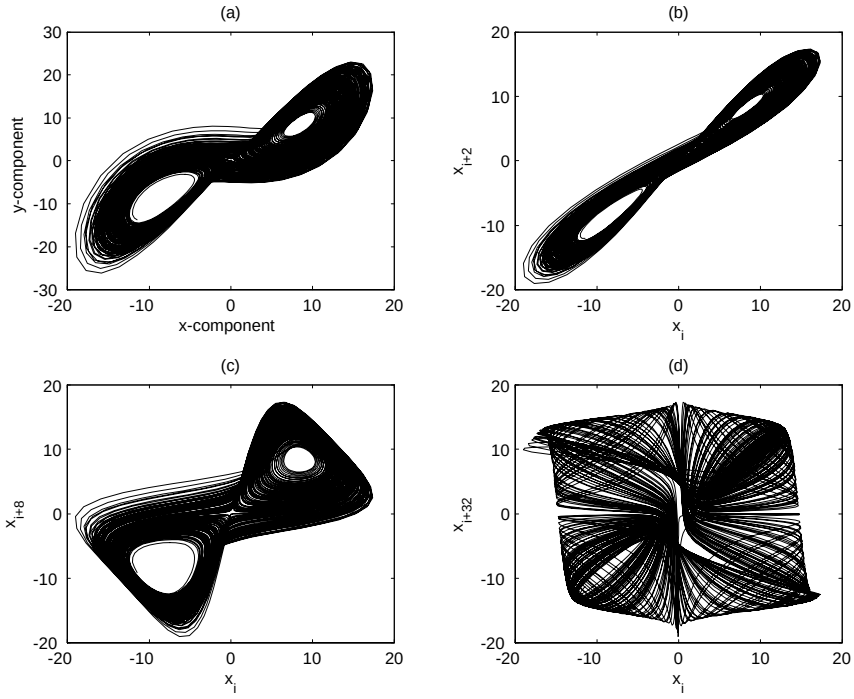

Figure 2.4: Effects of different time delays on the reconstructed attractor of the Lorenz system in two-dimensional embedding space. (a) Projection of the Lorenz attractor onto the x-y plane; (b), (c) and (d): Reconstructed attractor in embedding space $x_{i+\tau}$ vs x_i when time delay $\tau = 2$, $\tau = 8$ and $\tau = 32$ respectively.

the pair points $(x_i, x_{i+\tau})$ will distribute around the unity line $x_{i+\tau} = x_i$ as indicated in panel (b) of Fig. 2.4. But in practice, the presence of noise will lead to a noisy delay vector $\overrightarrow{X_i^n} = (x_i + \varepsilon_i, x_{i+\tau} + \varepsilon_{i+\tau})$ distributed within a "square" rather than as a point in the metric space $\mathbb{R}^2$, where ε_i denotes the noise introduced from measurement. The squares of adjacent vectors might intersect with each other. Then a delay vector $\overrightarrow{X_i^n}$, falling within a common region of these squares, can be either one of the adjacent vectors. Hence in this situation the reconstruction map Ψ is not one-to-one and cannot lead to a reconstruction. When time delay

τ is too large, say, $\tau = 32$ as adopted in panel (c) of Fig. 2.4, the reconstructed attractor is over-decorrelated and does not preserve the tangent space structure of the original attractor compared to those in panel (a) and (b). The corresponding derivative map $D\Psi$ is not one-to-one in this case [9], hence it also cannot lead to a reconstruction.

From the viewpoint of information theory, too small time delay τ means that $x_{i+\tau}$ contain mainly redundant information of x_i, which is called *redundancy [7]*. If a time delay is too large, then for chaotic systems, $x_{i+\tau}$ will be irrelevant to x_i, hence $x_{i+\tau}$ contains no information of x_i, which is known as *irrelevance [7]*. As Liebert and Schuster have argued that, we shall consider not only the effect of redundancy but also that of irrelevance in estimate of suitable time delays [24]. Therefore a tradeoff shall be achieved between redundancy and irrelevance so as to produce a reconstruction map Ψ. We define the following statistic, namely *redundancy and irrelevance tradeoff exponent (RITE)*, to measure the tradeoff,

$$RITE = \frac{\rho(x_i, x_{i+\tau}) \langle x_i^2 \rangle + (1 - \rho(x_i, x_{i+\tau})) \langle x_i \rangle^2}{\langle x_i^2 \rangle + \langle x_i \rangle^2}, \tag{2.10}$$

where $\langle \cdot \rangle$ denotes the expectation taken over time index i and

$$\rho(x_i, x_{i+\tau}) = \frac{cov(x_i, x_{i+\tau})}{var(x_i)} = \frac{\langle x_i x_{i+\tau} \rangle - \langle x_i \rangle^2}{\langle x_i^2 \rangle - \langle x_i \rangle^2}, \tag{2.11}$$

where $\rho(x_i, x_{i+\tau})$ is the *SOAC*, $cov(x_i, x_{i+\tau})$ and $var(x_i)$ are the covariance and variance function respectively for a time series $\{x_i\}$. After simplifications, we have

$$RITE = \frac{\langle x_i x_{i+\tau} \rangle}{\langle x_i^2 \rangle + \langle x_i \rangle^2}. \tag{2.12}$$

[9] For example, we can observe some break points where the directions of tangent vectors abruptly change.

Now let us interpret the meaning of Eq. (2.10). We take $\langle x_i^2 \rangle$ as the case of complete redundancy for the measure $\langle x_i x_{i+\tau} \rangle$, when time delay τ tends to zero and referring to $x_{i+\tau}$ brings no more information of x_i. Conversely, $\langle x_i \rangle^2$ is the case of complete irrelevance for the measure $\langle x_i x_{i+\tau} \rangle$, when $x_{i+\tau}$ is irrelevant and thus uncorrelated to x_i, hence $\langle x_i x_{i+\tau} \rangle$ is reduced to $\langle x_i \rangle^2$. $\rho(x_i, x_{i+\tau})$ plays the role to measure the redundancy between $x_{i+\tau}$ and x_i with a weight of $\langle x_i^2 \rangle \big/ \big(\langle x_i^2 \rangle + \langle x_i \rangle^2 \big)$, while $1 - \rho(x_i, x_{i+\tau})$ denotes the measure of irrelevance with the assigned weight of $\langle x_i \rangle^2 \big/ \big(\langle x_i^2 \rangle + \langle x_i \rangle^2 \big)$. Starting from $\tau = 0$, as time delay τ increases, the redundancy measure $\rho(x_i, x_{i+\tau})$ shall usually decrease while the irrelevance measure $1 - \rho(x_i, x_{i+\tau})$ shall increase, hence our criterion is to choose the suitable time delay at the first local minimum of *RITE*, where an optimal tradeoff between redundancy and irrelevance is deemed to be achieved according to Eq. (2.10).

If directly applying the measure of *RITE* to measure the original scalar time series $\{x_i\}$, from Eq. (2.10) we can find that, Eq. (2.12) is a constant affine transformation of the *SOAC,* therefore it will have the same performance as that of the ordinary *SOAC* algorithm since $\langle x_i^2 \rangle$ and $\langle x_i \rangle^2$ are both independent of time delay τ. When examining the performance of the *SOAC* algorithm, some authors gave some comments (cf. [13], p87 for more details) such as "a zero of the autocorrelation indicates on average linear independence of $x(t)$ and $x(t+\tau)$, but not necessarily of $x(t)$ and $x(t+2\tau)$" [24] and "The deeper reason for this failure is the fact that this approach does not employ any state space information" [40]. However, as to be indicated below, if we apply certain kind of transform on $\{x_i\}$

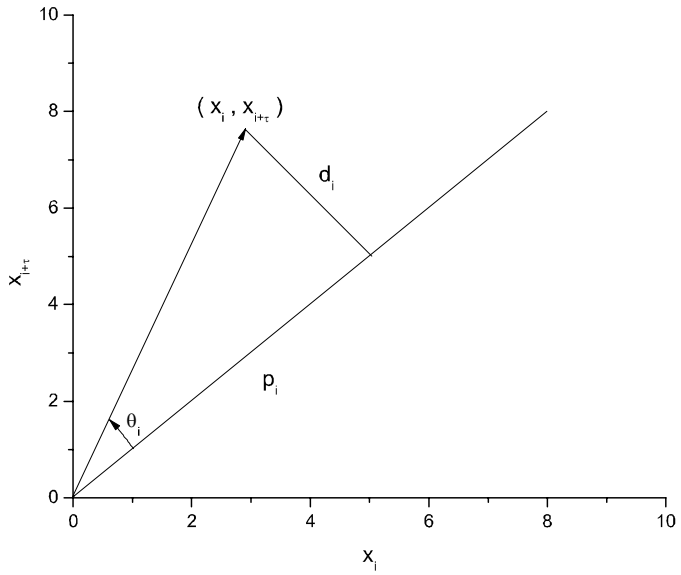

Figure 2.5: Geometric variables in the two-dimensional embedding space.

and use the new generated time series to substitute the original one, we can let the *SOAC* algorithm not only contain the information of $x_{i,}$ $x_{i+\tau}$ and $x_{i+2\tau}$ at the same time but also utilize the geometric information of the reconstructed attractor in the two-dimensional embedding space $x_{i+\tau}$ vs. x_i.

Now let us consider the possible transforms over the time series $\{x_i\}$. Let $\overrightarrow{(x_i,x_{i+\tau})}$ denote the vector from the origin to point $(x_i, x_{i+\tau})$ in the two dimensional embedding space, as shown in Fig. 2.5, we have the distance d_i of the pair points $(x_i, x_{i+\tau})$ to the identity line $x_{i+\tau} = x_i$ expressed by:

$$d_i = \frac{1}{\sqrt{2}} \left| x_{i+\tau} - x_i \right|, \tag{2.13}$$

where $|\cdot|$ denotes the distance in Euclidean space. The projection length p_i of

vector $\overrightarrow{(x_i, x_{i+\tau})}$ onto the identity line is:

$$p_i = \frac{1}{\sqrt{2}} \left| x_{i+\tau} + x_i \right| . \tag{2.14}$$

Therefore the angle between vector $\overrightarrow{(x_i, x_{i+\tau})}$ and the identity line is :

$$\theta_i = \tan^{-1} \left| \frac{x_{i+\tau} - x_i}{x_{i+\tau} + x_i} \right| . \tag{2.15}$$

From Eq. (2.13), (2.14) and (2.15), we can obtain three new time series $\{d_i\}$, $\{p_i\}$ and $\{\theta_i\}$, which consist of geometric description variables of the reconstructed attractor in the two-dimensional embedding space (other transforms are also possible of course). We apply the measure of *RITE* to these geometric variables with the same criterion aforementioned to choose suitable time delays, i.e., a suitable time delay will be chosen at the first local minimum of the geometric measures of *RITE*.

We note that if the origin in the embedding space of a time series is marginal to or even outside of the reconstructed attractor, the sensitivity of the geometric measures of *RITE* to different time delays will be significantly reduced. We therefore conduct the following smooth affine transform on the original time series $\{x_i\}$

$$y_i = \frac{x_i - \langle x_i \rangle}{\sqrt{var(x_i)}}. \tag{2.16}$$

The new time series $\{y_i\}$ shall have the same dynamical properties in the time domain as the original one has, while it takes the origin of embedding space as the center of its reconstructed attractor in a statistical sense. With this property, we prefer to study the time series $\{y_i\}$ rather than $\{x_i\}$. In addition, we will discard

the scale factor $1/\sqrt{2}$ of both Eq. (2.13) and (2.14) in all of our calculations without affecting the results.

We will study the simulation data sets from the Lorenz and Rössler systems [13]. For the Lorenz system, the system equations are:

$$\begin{cases} \dot{x} = \sigma(y - x); \\ \dot{y} = rx - y - xz; \\ \dot{z} = xy - bz. \end{cases} \tag{2.17}$$

with parameters $\sigma = 10$, $r = 28$, $c = 8/3$ and the sampling time $\Delta t_s = 0.02$ time units. For the Rössler system, the equations are:

$$\begin{cases} \dot{x} = -y - z; \\ \dot{y} = x + ay; \\ \dot{z} = b + z(x - c). \end{cases} \tag{2.18}$$

with parameters $a = 0.15$, $b = 0.20$, $c = 10.00$ and the sampling time $\Delta t_s = 0.1$ time units.

We will also apply the geometric measures of *RITE* to the sunspot record from year 1700 to year 1987 and infant respiratory data during stage 4 sleep (S4) [49]. In addition, we will calculate the time delays chosen by the criteria aforementioned, such as the *SOAC* (both ordinary and geometric measures), *ILD* and *AMI* algorithms, for the purpose of comparison. Our results are listed in Tables 2.1 and 2.2.

For the ordinary *SOAC* algorithm in Table 2.1, we use both the criteria which chooses time delay when the *SOAC* first drops to zero or $1/e$ (see [2] and references

Table 2.1: Time delays chosen by the ordinary *SOAC* and the geometric measures of *SOAC* algorithms.

Data set	Ordinary *SOAC*	Geometric measures of *SOAC*		
	(first dropping to $0/e^{-1}$)	distance	projection	angle
Lorenz	209/15	10	14	13
Rössler	15/11	16	15	15
sunspot	3/2	2	3	3
S4	7/5	7	7	6

therein). These two criteria work for the Rössler system, sunspot record and S4 data set. However, for the first-to-zero criterion, choosing time delay at 209 for the Lorenz system is improper (From Fig. 2.4 we see that time delays larger than 32 cannot reconstruct the Lorenz system properly), which can be taken as an example of failure for this criterion. While for the first-to-$1/e$ criterion, the choice of 15 is acceptable (to be shown below). Indeed, no longer requiring complete linear independence between two subsets, the first-to-$1/e$ criterion can be thought of as an effort to achieve certain tradeoff between redundancy and irrelevance, although we may not know *a priori* whether the prescribed constant factor $1/e$ is better than any other coefficients for an unknown system.

It is interesting to notice that, if we adopt the geometric variables obtained from Eq. (2.13), (2.14) and (2.15) in the ordinary *SOAC* algorithm (i.e, in Eq. (2.11) we calculate, for example, $\rho(d_i, d_{i+\tau})$ instead of $\rho(x_i, x_{i+\tau})$), it also can be a time delay selection criterion (similar to the *AMI* algorithm, time delay to be cho-

Table 2.2: Time delays chosen by the algorithms of the *ILD*, *AMI* and the geometric measures of *RITE*.

Data set	*ILD*	*AMI*	Geometric measures of *RITE*		
	τ/m		distance	projection	angle
Lorenz	9/4	8	9	12	10
Rössler	10/3	16	16	14	13
sunspot	2	3	2	3	2
S4	5/5	8	7	8	5

sen at the first local minimum) [10]. In Table 2.1, time delays chosen by this criterion are close to those chosen by the *AMI* algorithm and the geometric measure algorithms of *RITE*. However, in this criterion we do not include any tradeoff between redundancy and irrelevance, which is not desired as we have stated previously.

In Table 2.2, although the *ILD* algorithm was originally designed to determine suitable time delays τ, it might be more appropriate to utilize it to establish an embedding window $m\tau$. As indicated in the left panel of Fig. 2.6, when using Eq. (24) in Ref. [6] for calculation, for the Lorenz system the products of embedding dimensions m ($m >$correlation dimension d_c) and the corresponding time delays τ at the first local minimum of the average *ILD* are nearly a constant of 36. This conclusion also holds for data sets of the Rössler system and S4. In contrast, the

[10] Professor A. M. Albano commented that it might actually be the inclusion of the dynamics in the geometric variables (distance, projection and angle) of the two-dimensional embedding space that make it possible for a linear statistical measure such as SOAC to extract dynamical information from a scalar time series. In the future work, we could, of course, extend this idea to higher-dimensional embedding space.

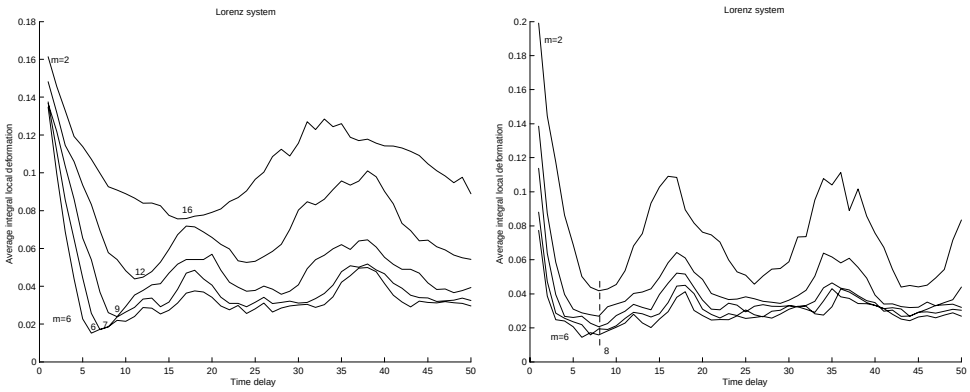

Figure 2.6: Figure in the left panel indicates the average integral local deformation vs. time delay for the time series from the Lorenz system with 9000 data points. The number of reference points is 500, radius for neighbour searching is set to 5. Embedding dimension m varies from 2 to 6 (from upper to lower) and time delay increases from 1 to 50. Figure in the right panel adopts the same parameters as the left for calculations except that the time series is shorter, consisting of only 1200 data points.

sunspot record has consistent local minima and the products of $m \cdot \tau$ do not keep constant. This still does not contradict our conclusion as the sunspot record is an extremely short time series. As shown in the right panel of Fig. 2.6, the constant embedding window will vanish when the time series from the Lorenz system is short, instead a consistent local minimum appears at $\tau = 8$.

Since embedding window $m\tau$ remains constant, different time delays will be obtained from the *ILD* algorithm for different embedding dimensions, nevertheless, we still think that the *ILD* algorithm can indicate how to obtain a proper reconstruction by choosing proper embedding dimensions. Here we will choose a

suitable embedding dimension for each data set (except for the sunspot record) under the criterion of *global False Nearest Neighbours (GFNN)* [29], then we can obtain the corresponding time delay according to the embedding window [11]. For the sunspot record we choose the time delay at the first consistent local minimum of the average *ILD*. The results are indicated in Table 2.2.

We will compare the algorithm performances below, especially for the *AMI* algorithm and geometric measure algorithms of *SOAC* and *RITE* . Take the time delays chosen by the popular *AMI* algorithm as the reference, loosely speaking, time delays selected by the geometric measures of *SOAC* and *RITE* are close to those chosen by the *AMI*. But note that direct comparison between these time delays tells us little, we have to instead evaluate whether a time delay is suitable according to the reconstruction quality, which can be reflected by some statistics related to the reconstruction. We will adopt two such statistics (both to be introduced in the next chapter with more details) in the following for reconstruction evaluation.

The first statistic is the *correlation dimension*, which shall be approximately the same for different reconstructions with properly chosen time delay and embedding dimension. We use the *Gaussian kernel algorithm (GKA)* implemented in [60] to calculate the correlation dimension. Since the experimental data sets we

[11] We must admit that this is somewhat "circular",since the choice of the suitable embedding dimension by *GFNN* algorithm in turn needs to take the suitable delay time as a parameter . In our calculation, we use the suitable delay time obtained by the *AMI* algorithm in Table 2.2 as the parameter to determine the optimal embedding dimension for each data set (except for the sunspot record).

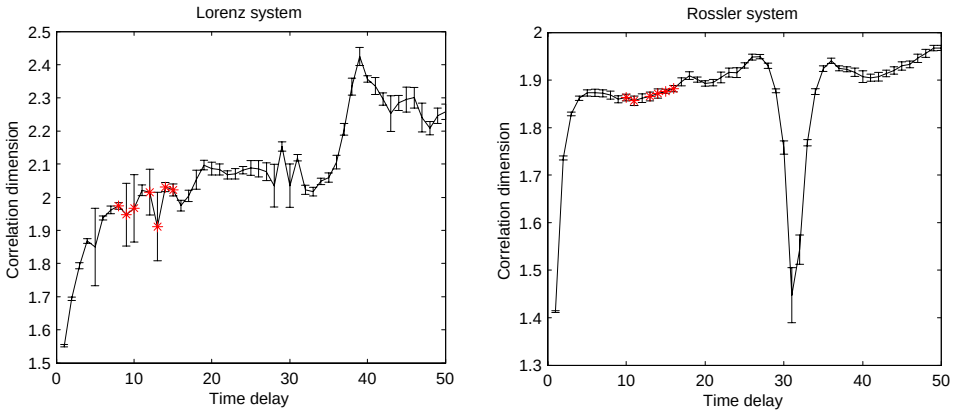

Figure 2.7: Correlation dimension vs. time delay with error bars. We adopt the *Gaussian kernel algorithm (GKA)* implemented in [60] for calculation. To speed up computation, we use only 4000 data points as references, hence for each time delay we calculate 10 times in order to evaluate the uncertainty. Embedding dimensions are 4 and 3 for the Lorenz and Rössler systems respectively, and both of the time delay ranges are from 1 to 50. For each data set, the *correlation dimensions* corresponding to time delays listed in Table 2.1 and 2.2 are marked with stars (except for the choice 209 of the Lorenz system).

have are relatively short [12], their estimated correlation dimensions are unreliable. We thus will only investigate simulation data sets from the Lorenz and Rössler systems, each with 9000 data points. In Fig. 2.7, we plot the curve of correlation dimension vs. time delay and mark correlation dimensions corresponding to the time delays in both Table 2.1 and 2.2 with stars (except for the choice 209 of the Lorenz system). As can be found, for the Lorenz system, the curve is less smooth than that of the Rössler system. Besides, the corresponding fluctuations are also

[12] The sunspot record has 288 data points while data set S4 has 1900.

larger than those of the Rössler system. One explanation for these phenomena may be that, to speed up computation, in the *GKA* we randomly adopt 4000 data points as the reference points, and the distribution of these randomly picked data points may have a greater influence on the Lorenz system because of its two-wing structure than that on the Rössler system with the same number of reference points. But roughly speaking, the marked correlation dimensions of both systems lie on plateaus as expected, which means, for those time delays we may obtain approximately the same correlation dimension. Hence the marked time delays can be thought of as equally suitable for a reconstruction.

The second statistic is the one step *local prediction error (LPE)*. As we know, local constant model [11] utilizes nearest neighbours for nonlinear prediction, when sufficiently high embedding dimension is reached, most of the effect of false nearest neighbours will be excluded. With embedding dimension and the radius of neighbour searching fixed, the *LPE* will only depend on the time delay. Hence one step *LPE* can qualitatively tell whether our choice is acceptable, since it shall achieve a tradeoff between being too small and being too large if the reconstruction is neither too redundant nor too irrelevant. In Fig. 2.8, the *LPE*s corresponding to the time delays listed in Table 2.1 and 2.2 chosen by different algorithms for each data set are marked with stars, from which we can see that they are indeed neither too small nor too large. In this sense, a certain tradeoff is achieved when choosing time delays in the above tables.

As we have mentioned before, we think all of the time delay candidates are

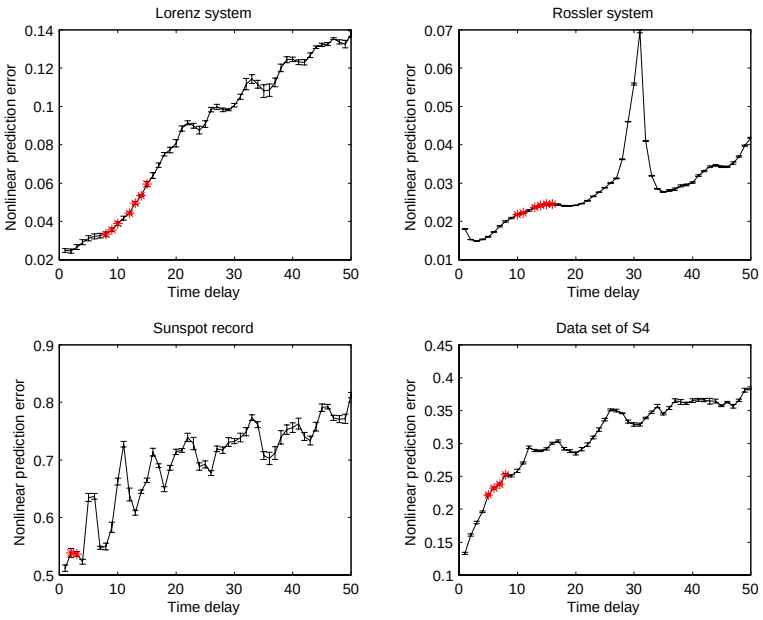

Figure 2.8: One step prediction error of local constant model vs. time delay with error bars. Note that to calculate the uncertainty of the prediction errors, we use 10 different subsets of each original data set without affecting the choice of time delays. Embedding dimensions used in the model are 4,3,3 and 5 for the Lorenz system, the Rössler system, the sunspot record and data set of S4 respectively. The ranges of time delays are all from 1 to 50 . The *LPE*s corresponding to time delays in Table 2.1 and 2.2 chosen by different algorithms for each data set are marked with stars (except for the choice 209 of the Lorenz system). We use the program *zeroth* in TISEAN package [18] for all our calculations.

equally suitable if they can lead to a proper reconstruction, no matter if they are the best choices under certain adjudicating criterion. But we still prefer geometric measure algorithms of *RITE* to those of *SOAC* because the *RITE* criterion represents the idea of tradeoff between redundancy and irrelevance. Indeed, only

Table 2.3: Time delays chosen by the geometric measures of *RITE* for the time series from the Lorenz and Rössler systems contaminated with observational Gaussian white noises .

Noise level (%)	Lorenz system			Rössler system		
	distance	projection	angle	distance	projection	angle
0	9	12	10	16	14	13
3	9	12	10	16	14	13
6	9	12	9	16	14	13
9	9	12	10	16	14	13
12	9	12	8	16	14	2

considering the linear and nonlinear dependence between two subsets, the *SOAC* and *AMI* criteria may fail for some systems. We have shown that the first-to-zero criterion of the ordinary *SOAC* does not work for the Lorenz system. As for the *AMI* criterion, we may only obtain monotonically decaying curves for some dynamical systems [13], hence there is no local minimum to choose time delay. Therefore we would like to suggest that, even for the *AMI* algorithm, it may be better to include certain tradeoff.

Now let us examine the computational cost of each algorithm listed in Table 2.1 and 2.2. Let N denote the data set size of time series $\{x_i\}$, then the *ILD* algorithm approximately requires $O(N_{ref} \times (N \ln N))$ operations on searching nearest neighbours for each embedding dimension and each time delay, where N_{ref} is the

[13] For example, some ECG observations.

number of reference points. The *AMI* algorithm needs about $O(N^2)$ operations to calculate joint probability distribution for each time delay, while the *RITE* algorithm will be faster than both of them, undergoing about $O(N)$ operations on both the transforms over the original data set and the calculations of expectation for each time delay. This is an advantage under some situations, for example, surrogate tests [52], where lots of data sets will be involved in calculation.

We will also test the robustness of the geometric measures of *RITE* against observational Gaussian white noise $N(0, \delta^2)$. The noise level is defined as the ratio of δ to δ_s, where δ_s is the standard deviation of the original scalar time series $\{x_i\}$ before the transform of Eq. (2.16). As indicated in Table 2.3, using the time delays chosen at noise level zero as the references, we find both the distance and the projection measures of *RITE* are rather robust against observational noise, noise level up to 12% still do not affect the choices of the time delay. However, the angle measure of *RITE* is more sensitive to noise. For the Lorenz system, small fluctuations of the choice appear when noise level is higher than 6%. For the Rössler system, the performance seems better. The odd choice $\tau = 2$ at noise level 12% follows our criterion suggested above, which is due to a small spike on the curve of the angle measure of *RITE* vs. time delay, while the next local minimum is exactly at the time delay $\tau = 13$. Although the robustness against observational noise of the geometric measures of *RITE* might vary from system to system, we believe in general it is satisfactory.

Up to now, *average mutual information* is the most preferred statistic used

for choosing time delay since it has a valuable physical meaning, but it requires a complicated implementation algorithm. To achieve higher accuracy, more complex implementation and more running time are needed. Also it does not deal well with short time series. The *RITE* algorithm can overcome some of the disadvantages of the *AMI* algorithm. It is much simpler to implement, and our calculations indicate that the *RITE* algorithm performs well for a variety of time series of various lengths and even with the presence of substantial noise. Although through the *RITE* algorithm we may obtain different results under different geometric criteria, we can take all of them as suitable candidates for a reconstruction if we do not seek the "best" choice for certain adjudicating criterion. We therefore feel that such a simple algorithm could be a valuable tool to choose time delay, especially for the situation such as the surrogate test, where there are many data sets needing to be reconstructed.

2.3.2 Algorithms to Choose Embedding Dimension

There are many works concerning how to choose a suitable embedding dimension for a nonlinear time series, in general, they can be included into three kinds of algorithms:

- One algorithm is based on the invariant statistics, for example, the correlation dimension and entropy etc., of the underlying system. As we know, theoretically the invariant statistics of a dynamical system shall be the same for different but equally proper embedding parameters. If we know a proper time delay, then by fixing time delay and examining the trend of convergence

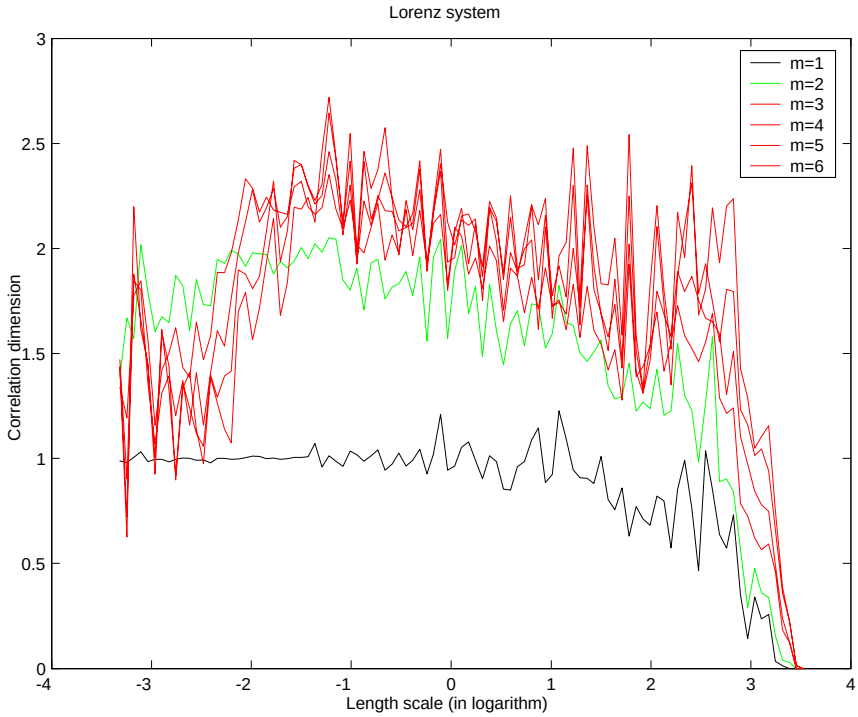

Figure 2.9: Using the correlation dimension as the invariant statistic to choose a proper embedding dimension. We prescribe a suitable time delay through the *RITE* algorithm introduced above, then calculate the local correlation dimensions at different length scales. Within a moderate scale interval (we will discuss the effect of length scale on the calculation of local correlation dimension in the next chapter), we find the curves will begin to converge when the embedding dimension is up to three. Hence we can determine that the embedding dimension shall be at least three.

of different embedding dimensions, we may choose a proper embedding dimension, although it is somewhat empirical and has to depend on our visual inspection of the situations. We use the Lorenz system as an example to calculate the (local) correlation dimensions of different embedding dimensions

with the prescribed time delay, as shown in Fig. 2.9.

- Another algorithm utilizes the *singular value decomposition (SVD)* technique. In [5] the authors define the embedding matrix E for a scalar time series $\{x_i\}_{i=1}^{N}$ as

$$E_{r\times d} = r^{-1/2}\cdot \begin{bmatrix} \overrightarrow{E_1}^T \\ \overrightarrow{E_2}^T \\ \vdots \\ \overrightarrow{E_r}^T \end{bmatrix}_{r\times d} = r^{-1/2}\cdot \begin{bmatrix} x_1 & x_2 & \cdots & x_d \\ x_2 & x_3 & \cdots & x_{d+1} \\ \vdots & \vdots & \vdots & \vdots \\ x_r & x_{r+1} & \cdots & x_n \end{bmatrix}_{r\times d}, \quad (2.19)$$

where $\overrightarrow{E_i} = \left(x_i, x_{i+1}, ..., x_{i+(d-1)}\right)^T$, $r = N-(d-1)$ and d is the embedding dimension to be chosen. Note that matrix E has a Toeplitz structure, i.e., its elements $E_{i,j}$ satisfy that $E_{i,j} = x_{i+j-1}$, and obviously $E_{i,j} = E_{m,n}$ if $i+j = m+n$.

We will discard the factor $r^{-1/2}$ since it does not affect our discussion at all. Utilizing the SVD technique, we can decompose the matrix $E_{r\times d}$ into two orthogonal matrices (U and V) and one diagonal matrix (S),

$$E_{r\times d} = V_{r\times d} S_{d\times d} U_{d\times d}, \quad (2.20)$$

where $V^TV = U^TU = I_{d\times d}$ with $I_{d\times d}$ denoting the d dimensional identity matrix.

Usually it is preferred to arrange the (non-negative) singular values in matrix S in decreasing order, i.e., given the diagonal matrix $S = diag(s_1, s_2, ..., s_d)$, we have

$$s_1 \geqslant s_2 \geqslant ...s_d \geqslant 0.$$

If $rank(E) = n$, then we have

$$s_1 \geqslant ... \geqslant s_n > s_{n+1} = ... = s_d = 0.$$

We can take n as the number of eigenvectors which span the space where the original attractor of the underlying system is embedded. Without noise, we can count the number of positive values in set $(s_1, s_2, ..., s_d)$ to determine the matrix rank, which is taken as the embedding dimension. If noise is present in the time series, however, the situation will be different. Consider the case of an i.i.d noise process $\{\xi_i\}$ with mean value 0 and standard deviation σ. Denote the new singular values as $(s_1^{'}, s_2^{'}, ..., s_d^{'})$, it can prove that [5]

$$\left(s_i^{'}\right)^2 = s_i^2 + \ \sigma^2/d \qquad i = 1, 2, ..., d. \tag{2.21}$$

The presence of i.i.d noise will not change the directions of the original eigenvectors but only increase their magnitudes, for $i > rank(E) = n$, $\left(s_i^{'}\right)^2 = \sigma^2/d$ hence they will no longer be zero. Nevertheless, we can still determine the rank of matrix E by inspecting the onset of convergence (to σ^2/d, ideally).

If the perturbation to the time series is colored noise, i.e., the noise components have certain correlations with each other, in general, the presence of noise will change both the amplitudes and the directions of the original eigenvectors, and usually we cannot get the simple relationship as that described in Eq. (2.21), instead, we have to write the relationship as

$$\left(s_i^{'}\right)^2 = g(\overrightarrow{s_1}, \overrightarrow{s_2}, ..., \overrightarrow{s_3};\ \overrightarrow{\sigma_1}, \overrightarrow{\sigma_2}, ..., \overrightarrow{\sigma_d}).$$

where $(\overrightarrow{\sigma_1}, \overrightarrow{\sigma_2}, ..., \overrightarrow{\sigma_d})$ are the eigenvectors of colored noise which span the same metric space as $(\overrightarrow{s_1}, \overrightarrow{s_2}, ..., \overrightarrow{s_3})$. Here we add an arrow for each component to highlight that the eigenvectors of signal and noise might follow different directions. Therefore it is possible that the singular values $\left(s_1^{'}, s_2^{'}, ..., s_d^{'}\right)$ might not converge. For such situations, we cannot use this technique to determine the suitable embedding dimension any more.

There are many further discussions on the failure of the SVD algorithm (see ref. [31] & [35], for example). Essentially, the SVD technique, being equivalent to the power spectrum analysis technique, is a linear approach. Hence, it is natural that it will fail to reveal the characteristic structures of nonlinear dynamical systems.

- The third algorithm is based on the *false nearest neighbour (FNN)*. We use an example to demonstrate the concept of false nearest neighbour. As shown in Fig. 2.10, we have an attractor in two dimensional space, whose trajectory is denoted T. Consider data points A, B and C on T, initially B and C are

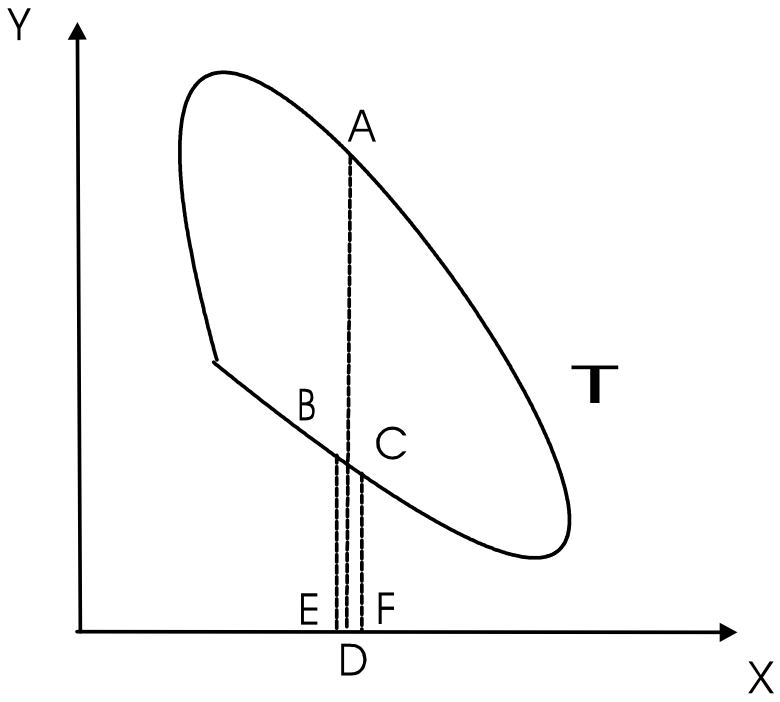

Figure 2.10: An attractor in two dimensional space. We denote the trajectory of the attractor as T. Consider data points A and B on trajectory T, initially they are too far away to be neighbours, however, if we project the whole trajectory onto the x axis, the corresponding projection points, D and E, can become neighbours (depending on how we choose the radius for nearest neighbour searching). Through this example we can see that, given a scalar time series, there could be many false neighbours if we choose an improper embedding dimension, see text for more discussions.

neighbours while A and B are not [14], however, the corresponding projections onto the x axis, D, E and F respectively, are all close enough to become pos-

[14] Whether two data points are neighbours actually depend on the maximal distance we choose between two adjacent data points (i.e., the radius for nearest neighbour searching). However, usually we will not take two data points as neighbours if they have a large evolution time gap, otherwise the trajectory we reconstructed will distort to some extent since we ignore too much trajectory information between the so-called "neighbours".

sible neighbours. Suppose the projection of trajectory T (onto the x axis) is exactly the scalar time series we obtained, and if we choose embedding dimension $m = 1$ for reconstruction, then D and E will be falsely treated as neighbours, we call such adjacent data points *false neighbours*. If we increase the embedding dimension to 2, then the false neighbours can be easily distinguished. Hence we can design a criterion to choose the suitable embedding dimension, i.e., we randomly select some data points, count the number of their nearest neighbours[15] for each embedding dimension. For suitable embedding dimensions, the number of their nearest neighbours will approximately be constant, however, if the embedding dimension is improper, the false nearest neighbours will occupy a substantial portion of the whole. Hence by comparing the variation of the number of nearest neighbours for different embedding dimension, we can select a suitable embedding dimension.

2.4 Brief Summary

In this chapter we have introduced the embedding theorems, especially Takens' embedding theorem, for dynamical systems. Takens' embedding theorem proves that a time delay reconstruction from a scalar time series will almost always be a diffeomorphism (as a sub-manifold embedded in space $\mathbb{R}^m$) of the original attractor if we choose a sufficiently high embedding dimension m. Hence time delay embedding reconstruction is a convenient and powerful method to represent the

[15] In practice, we use the concept of the i-th nearest neighbour. An i-th nearest neighbour of a data point A is the data point whose distance to data point A is the i-th shortest among all of the data points (except A).

underlying system of scalar observations, however, since the theorem only considers the ideal conditions (no noise, infinite precision etc.), in practice not all of the time delays are equally good even we have chosen a suitable embedding dimension. Hence after introducing the theorem, we also provided some criteria for parameter selection, including the RITE algorithm which we proposed recently for the choice of suitable time delay .

CHAPTER 3

SOME USEFUL STATISTICS IN NONLINEAR TIME SERIES ANALYSIS

In this chapter we will discuss a few nonlinear statistics, including both the correlation dimension and the local prediction error. The main reason to introduce these two statistics is that, throughout this thesis, we will utilize them as the characteristic measures of the dynamical systems or as the discriminating statistics in the surrogate tests, with which we can either quantitatively estimate the systems' characteristic behaviors (see Chapter 3) or distinguish different dynamical systems (see Chapter 4 and 5). There are also many other nonlinear statistics, for example, information entropy and Lyapunov exponent etc., but since no application of these measures will be involved in this thesis, detailed discussion of the statistics will not be presented.

3.1 Introduction to Correlation Dimension: Definition and Calculation

An attractor of a chaotic system in state space is often called strange attractor. To measure the strangeness of the attractor, one may use the statistic such as fractal (Hausdorff) dimension [10]. The difficulty in adopting fractal dimension as the characteristic measure of a strange attractor is that, it is often impractical to calculate the fractal dimension from a scalar time series whenever the (fractal) dimensionality is larger than two, as pointed out by Grassberger and Procaccia [14]. They proposed an alternative measure, namely correlation dimension (or

correlation exponent), by considering nonlinear correlations between points of the reconstructed attractor. Given a scalar time series $\{x_i\}$, by choosing proper embedding dimension and time delay, we can reconstruct the underlying system and obtain a set of delay vectors $\left\{\overrightarrow{X_i} = \left(x_i, x_{i+\tau}, \ldots, x_{i+(m-1)\tau}\right)^T\right\}$. Grassberger and Procaccia [14] defined a measure called correlation integral as follows,

$$C(m,r) = \lim_{N\to\infty} \frac{1}{N^2} \sum_{i,j=1}^{N} \theta(r - \left|\overrightarrow{X_i} - \overrightarrow{X_j}\right|), \tag{3.1}$$

where N is the element number of set $\left\{\overrightarrow{X_i} = \left(x_i, x_{i+\tau}, \ldots, x_{i+(m-1)\tau}\right)^T\right\}$, r is the threshold of interpoint distance (length scale), $\theta(x)$ is the Heaviside function satisfying

$$\theta(x) = \begin{cases} 0, x < 0; \\ \frac{1}{2}, x = 0; \\ 1, x > 0. \end{cases} \tag{3.2}$$

and $|\cdot|$ denotes the metric norm (in calculation, we adopt the Euclidean norm).

It is assumed that, $C(m,r)$ follows a power law of r for small enough r with infinite embedding dimension, i.e., in ideal situations (no noise, infinite data, for example), we have

$$\lim_{\substack{r\to 0 \\ m\to\infty}} C(m,r) \sim r^{d_c}, \tag{3.3}$$

where d_c is the correlation dimension. From Eq. (3.3) we can estimate the correlation dimension by

$$d_c = \lim_{\substack{r\to 0 \\ m\to\infty}} \frac{\log C(m,r)}{\log r}. \tag{3.4}$$

Hence in order to calculate d_c, we need first to estimate the correlation integral defined in Eq. (3.1). We call this estimate the correlation sum when N is finite. Given $C(m,r)$, we can estimate the correlation dimension d_c through many conventional parameter estimation methods. Since we are more interested in the estimation of the correlation sum, no details in parameter estimation (i.e., estimating d_c according to the behavior of $C(m,r)$) will be discussed. Readers are referred to, for example,[13], for more details.

To estimate $C(m,r)$ in Eq. (3.1), a straightforward way is to adopt the box-counting algorithm, i.e., we use hypercubes with edge length r to cover the reconstructed attractor, then index the non-empty hypercubes by number $1, 2, \ldots, M(r)$. By counting the number of points μ_i on the reconstructed attractor which fall within the i-th non-empty hypercubes, roughly we have [14],

$$C(m,r) \simeq \frac{1}{N^2} \sum_{i=1}^{M(r)} \mu_i^2 . \tag{3.5}$$

Recalling the methods of estimation of the density of a data set, we will find the box-counting algorithm is similar to the histogram algorithm, i.e., both algorithms will need to partition the region where the data sets locate on and then count the number of data points in each partition. The difference is that, density estimation calculates the distribution of data points while the correlation sum estimates the distribution of interpoint distances. Nevertheless, in the spirit of

distribution estimation, some techniques in density estimation can be applied to the correlation sum estimation, for example, Diks applied the kernel density estimation (KDE) technique [47] to calculate the correlation sum [9]. First we write Eq. (3.1) equivalently in the integration form

$$C(m,r) = \int \rho(\overrightarrow{X};m)d\overrightarrow{X} \int \rho(\overrightarrow{Y};m)\theta(r - \left|\overrightarrow{X} - \overrightarrow{Y}\right|)d\overrightarrow{Y}, \tag{3.6}$$

where $\rho(x)$ is the *probability density function* (PDF) of variable x. To introduce the technique of kernel density estimation, Diks substituted function $\theta(x)$ by a kernel function $w(x)$, hence the correlation integral can be generalized as

$$T(m,r) = \int \rho(\overrightarrow{X};m)d\overrightarrow{X} \int \rho(\overrightarrow{Y};m)w(\left|\overrightarrow{X} - \overrightarrow{Y}\right|/r)d\overrightarrow{Y}. \tag{3.7}$$

By adopting the Gaussian kernel function

$$w(x) = e^{-x^2/4}, \tag{3.8}$$

we have

$$T(m,r) = \int \rho(\overrightarrow{X};m)d\overrightarrow{X} \int \rho(\overrightarrow{Y};m)e^{-\left|\overrightarrow{X}-\overrightarrow{Y}\right|^2/4r^2}d\overrightarrow{Y}. \tag{3.9}$$

Also under the assumption of the power law behavior of $T(m,r)$ as $r \to 0$, $m \to \infty$ [1], Eq. (3.9) can be shown to be scaled as

$$T(m,r) \sim e^{-Km\tau}m^{-d_c/2}r^{d_c}. \tag{3.10}$$

[1] It is shown that, for any kernel function $w(x)$ which monotonically decreases in x for $x \geq 0$, if it also satisfies that $\lim_{r\to 0} r^{-p}w(x/r) = 0$ (pointwisely) for $x > 0$ and $p \geq 0$, then the corresponding correlation integral will follow a power law behavior [9].

Parameter K is called the correlation entropy, which needs to be estimated from the behavior of $T(m,r)$ just as the correlation dimension d_c does. In practice, Diks suggested to extract parameters K and d_c from $T(m,r)$ through the method of Schouten [36].

To approximate the correlation integral, we discretize Eq. (3.7) and take and average over the interpoint distribution. The estimate $\hat{T}(m,r)$ turns out to be

$$\hat{T}(m,r) = \frac{1}{N_p} \sum_{i,j \neq i} \psi_{i,j}, \tag{3.11}$$

where N_p is the number of delay vector pairs $(\overrightarrow{X}_i, \overrightarrow{X}_j)$ used an

$$\psi_{i,j} = \exp(-\left|\overrightarrow{X}_i - \overrightarrow{X}_j\right|^2 / 4r^2). \tag{3.12}$$

Yu et al. [60] devised an efficient implementation of the Gaussian kernel algorithm (GKA) to estimate the correlation integral, for example, by excluding the repetition in scanning interpoint distance $\left|\overrightarrow{X}_i - \overrightarrow{X}_j\right|$ and adopting efficient binning strategies to take an average over interpoint distribution, the authors substantially decrease the computational cost in calculating the summation of the integral core $\left\{\psi_{i,j}\right\}$. In this thesis we will adopt their implementation to calculate the correlation dimension unless otherwise specified.

Remark 5 *Note in the above discussions, we always assume the correlation integral $C(m,r) \sim r^{d_c}$ for m fixed, $r \to 0$. However, in some situations (for example, a Cantor-like data set), this assumption does not necessarily hold. Judd [16] suggested to use the relation $C(m,r) \sim \varphi(r) r^{d_c}$ for m fixed, $r \to 0$. For practical*

purposes, he used the first order polynomial to approximate the function $\varphi(r)$, i.e., we have $C(m,r) \sim (a_0 + a_1 r) \cdot r^{d_c}$. Hence, besides parameter d_c, there are two more parameters a_0 and a_1 that need to be extracted.

Remark 6 *The correlation dimension discussed above is obtained by fitting the global slope of the curve $C(m,r)$ for fixed embedding dimension m. If we want to inspect the local slope (local correlation dimension) of $C(m,r)$ at different length scales, another formula is adopted, i.e., $d_c^l(m,r) = \partial \log C(m,r) / \partial \log r$. Take the curves in Fig. 2.9 as an example, the typical behavior of $d_c^l(m,r)$ is that, at very large length scale, due to the boundary (or finite size) of the reconstructed attractor of the Lorenz system, the local correlation dimension will drop to zero. At very small length scale, the presence of measurement noise will conceal the power law. In addition, if length scale r is close to $\min(\left|\overrightarrow{X}_i - \overrightarrow{X}_j\right|)$, statistical fluctuations will appear when counting the number of pair points $(\overrightarrow{X}_i, \overrightarrow{X}_j)$, therefore the corresponding local correlation dimension will be unstable.*

3.2 Introduction to One Step Local Prediction Error

First let us consider the prediction problem of a chaotic time series $\{x_i\}$. In principle, long-term prediction for a chaotic time series is impossible. However, short-term prediction is still feasible. For example, if currently we know the values $\{x_i : i = 1, 2, ..., k\}$ and we want to predict the future value x_{k+1} (one step prediction), one idea for (local) prediction could be that [11], first we use time delay embedding reconstruction to reconstruct the underlying system from the scalar time series $\{x_i\}$, then we can obtain a set of vectors $\{\overrightarrow{x_i} : \overrightarrow{x_i} = (x_i, x_{i-\tau}, ..., x_{i-(m-1)\tau})\}$.

We could examine the evolution history of variable $\overrightarrow{x}$, if two vectors, for example, $\overrightarrow{x_p}$ and $\overrightarrow{x_k}$ are very close, then due to the determinism in the chaotic time series, the values of the next evolution step, $\overrightarrow{x_{p+1}}$ and $\overrightarrow{x_{k+1}}$ are possibly also close if the sampling time is not very large (also depending on the local Lyapunov exponent of course). Therefore if we know the value x_{p+1}, we can infer the value of x_{k+1} from $\overrightarrow{x_{p+1}}$ (see below discussion). In practice, however, due to the presence of noise, some states of the observations might be indistinguishable from the true states of the underlying system, which brings the uncertainty in prediction [17]. Therefore, in order to increase the stability of prediction, we do not compare $\overrightarrow{x_k}$ with only one of the closest neighbours, instead, we will search a number of nearest neighbours. We have two approaches to obtain the ensemble of them. One is that we first specify a proper radius ϵ [20]. We take the vectors falling within the sphere $\{\overrightarrow{x_i} : \|\overrightarrow{x_i} - \overrightarrow{x_k}\| < \epsilon, \overrightarrow{x_i} \neq \overrightarrow{x_k}\}$ as the neighbour ensemble, where $\|\cdot\|$ denotes the metric norm, usually we adopt the Euclidean norm, i.e., $\|\overrightarrow{x_i} - \overrightarrow{x_k}\| = \left\{ \sum_{j=1}^{m} (x_{i-(j-1)\tau} - x_{k-(j-1)\tau})^2 \right\}^{1/2}$. The other approach is to select q nearest neighbours $\{\overrightarrow{x_{I_s}} : s = 1, 2, ..., q, q \geqslant m + 1\}$, where m is the embedding dimension and $\overrightarrow{x_{Is}}$ is the *s-th* nearest neighbour of $\overrightarrow{x_k}$ [2] [11]. For convenience in discussion, we denote the neighbour ensemble as $U = \{\overrightarrow{x_{n_i}} : i = 1, 2, ..., n | \overrightarrow{x_k}\}$, where n is the number of nearest neighbours. We can construct a local predictor from U [11], for example, the simplest way, called local constant prediction or zeroth-order prediction, is to let the prediction value $\hat{x}_{k+1}$ be the average value of x_{n_i+1}, i.e.,

[2] The *s-th* nearest neighbour of $\overrightarrow{x_k}$ is defined as the vector $\overrightarrow{x_{I_s}}$ among the set $\{\overrightarrow{x_i}\}$, which satisfies that the distance rank $rank(\|\overrightarrow{x_{I_s}} - \overrightarrow{x_k}\|) = s$, where function $rank(\cdot)$ denotes the ordinal number of a value in a list (in the increasing order).

$\hat{x}_{k+1} = \sum_{i=1}^{n} w_{n_i} \cdot x_{n_i+1} \Big/ \sum_{i=1}^{n} w_{n_i}$, where x_{n_i+1} is the next scalar value of x_{n_i} in time, or equivalently $x_{n_i+1} \in \{x_{n_i+1} : \overrightarrow{x_{n_i}} \in U\}$, and w_{n_i} is the corresponding weight. A superior approach is to utilize the information of $\{(\overrightarrow{x_{n_i}}, x_{n_i+1}) : i = 1, 2, ..., n\}$, we can use an autoregressive model to fit $(\overrightarrow{x_k}, x_{k+1})$ linearly (local linear prediction), i.e., we have $\hat{x}_{k+1} = a_k \cdot \overrightarrow{x_k} + b_k$, where a_k and b_k are vector coefficient and scalar coefficient respectively, which are obtained by fitting $\{(\overrightarrow{x_{n_i}}, x_{n_i+1}) : i = 1, 2, ..., n\}$ (for example, using the singular value decomposition technique, see [11]).

After we obtain the prediction value $\{\hat{x}_i\}$ through the above methods, we can define the (one step) local prediction error as

$$e = \left\{ \sum_i (\hat{x}_i - x_i)^2 \right\}^{1/2} . \tag{3.13}$$

3.3 Brief Summary

In this chapter we have introduced two statistics, the correlation dimension and the local prediction error. The correlation dimension indicates the strangeness of a strange attractor, while the local prediction error describes the local predictability of a dynamical system. Both of these statistics can characterize the system's behavior, therefore they are often adopted in many applications.

CHAPTER 4

SURROGATE TESTS ON IRREGULAR TIME SERIES

4.1 Overview

Constrained-realization surrogate tests (surrogate tests in short) [52] are examples of Monte Carlo hypothesis tests [13]. Take the surrogate test of nonlinearity in a time series [52] as an example, we first need to adopt a null hypothesis, which usually supposes the time series is generated by a linear stochastic process and potentially filtered by a nonlinear filter. Based on this null hypothesis, a large number of data sets (surrogates) are to be produced based on the original time series. Two approaches are available to create surrogates. One is the traditional bootstrap method, which usually uses the explicit model extracted from the original time series to generate surrogate data. This is a parametric method in essence, the distributions of the discriminating statistics (defined in previous chapters) are most likely dependent on the parameters of the extracted model and hence it might lead to spurious conclusions on the original data set. An alternative approach called constrained realizations [54] is more suitable for our purpose. In general, it avoids fitting any model, instead depending on the null hypothesis, this method ensures that the surrogates keep some properties which are supposed to exist in the original time series according to the null hypothesis. Surrogates thus generated are known as constrained-realization surrogates. In case of nonlinearity detection, we want

the surrogates to keep the linearity of the original time series but destroy all other structures. After the generation of the surrogates, we then calculate some non-linear statistics (discriminating statistics), for example, the correlation dimension, of both the original time series and the surrogates. If the discriminating statistic of the original time series deviates from those of the surrogates, we can reject the null hypothesis we proposed and claim that the original time series is more likely to be deterministic than to be stochastic with certain confidence level (depending on how many surrogates we have generated, to be shown later).

In general, to apply surrogate techniques to test if a time series possesses the property P we are interested in, we first select a null hypothesis, which assumes the time series instead has a property Q opposite to P. We then devise a corresponding algorithm to produce surrogates from the observed data set. In principle, these surrogates shall preserve the potential property Q while destroying all others. The next step is to choose a suitable discriminating statistic, which shall be an invariant measure for both the surrogates and the original time series if the null hypothesis is true. Hence if the discriminating statistic of the original time series distinctly deviates from the distribution of the discriminating statistic of the surrogates, the null hypothesis is unlikely to be true, or in other words, the time series is more likely to possess the property P than Q. In this way, we can assess the statistical significance of our calculations through surrogate tests even when we have only a very limited amount of observations. Such assessments are important because in many practical situations statistical fluctuations are inevitable due to the presence

of noise, hence the surrogate test is a proper tool to evaluate the reliability of our results in a statistical sense. Note that, one can also apply direct detection techniques, for examples, some authors proposed direct techniques to detect the nonlinearity in a time series (see, for example, [21]). However, a direct test usually will not give out the confidence level. Consider the following situation, if we find that the (maximal) Lyapunov exponent of a time series is 0.01, it may be difficult for us to tell whether the time series is chaotic, or the time series is pseudoperiodic, but the presence of noise causes the Lyapunov exponent to be slightly larger than zero. In such situations, we would like to suggest adopting the surrogate test (if possible) rather than the direct test.

In this chapter we will focus on two topics. First we will apply the surrogate test technique to detect the nonlinear determinism in irregular time series. In general, an irregular time series is generated either from a (pure) stochastic process or from a deterministic nonlinear system [1]. To understand the underlying system of the time series, the first step shall be to investigate whether the irregularity is caused by stochasticity or nonlinearity (often chaos), then corresponding strategies for further analysis can be properly developed and applied. After the detection of nonlinearity, we are also interested in examining whether the time series is pseudoperiodic or chaotic [2], we will also employ the surrogate test technique to distinguish chaos and pseudoperiodicity in the time series.

[1] In this sense, determinism in an irregular time series often means nonlinearity.

[2] In many cases, an irregular nonlinear deterministic time series is either pseudoperiodic or chaotic.

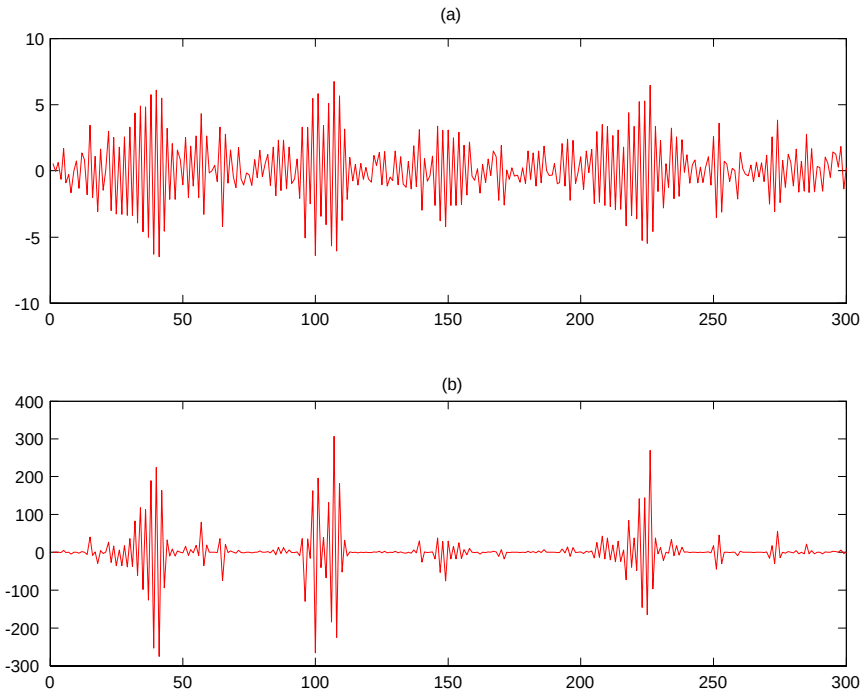

Figure 4.1: (a) The waveform of an AR(1) process; (b) The waveform of data set from the AR(1) process after the measurement of a cubic function.

4.2 Surrogate Tests to Detect Nonlinearity

This section will be arranged in the order of the surrogate test procedures. As we have introduced, to apply the surrogate technique, we first need a null hypothesis. Based on the null hypothesis we then design the surrogate generation algorithm. After calculating the discriminating statistic for both the surrogates and the original time series, according to certain criterion, we shall examine if the null hypothesis can be rejected or not, and give out the corresponding interpretations of the results.

4.2.1 Null Hypotheses

Note that in our discussions we will always assume the time series is second order stationary. Loosely speaking, by second order stationary we mean that the time series has constant mean value, finite second order autocovariance, while its covariance function only depends on the time difference between two subsets. Concretely, if a time series $\{x_i\}$ is second order stationary, we have

- $E(x_i)$ is a constant, where E denote the expectation operator.
- $|var(x_i) = E(x_i^2) - E^2(x_i)| < \infty$.
- $cov(x_i, x_{i+\tau}) = cov(x_0, x_\tau)$ for any indices i, where $cov(x, y) = E(xy) - E(x)E(y)$ is the covariance function between data set x and y.

Wold proved that any stationary stochastic process $\{x_i\}$ can be expressed in a linear representation ([4], p47), for example,

$$x_i = \mu + \sum_{j=1}^{\infty} \phi_j \cdot x_{i-j} + \sigma \cdot \xi_i \tag{4.1}$$

with $\sum_{j=1}^{\infty} \left|\phi_j\right| < \infty$, where ξ_i are independent random variables with standard deviations σ, μ is a constant determined by the mean of time series $\{x_i\}$ and the coefficients ϕ_j.

Since irregular time series are usually generated by nonlinear deterministic systems or stochastic processes, if we know that a stationary time series is not produced from a linear stochastic process, then it is highly likely that it comes from a nonlinear deterministic system, hence we say the time series contains nonlinearity.

However, a problem we shall note is that, we might introduce nonlinearity to a linear stochastic time series during the measurement process. Let us use an example to demonstrate this point. Suppose $\{x_i\}$ is a linear stochastic time series from the $AR(1)$ process in Eq. (4.2), whose waveform is indicated in the upper panel of Fig. 4.1.

$$x_i = -0.9 \cdot x_{i-1} + \xi_i. \tag{4.2}$$

If we use a cubic function $f(x) = x^3$ to measure $\{x_i\}$, we will obtain the distorted time series $\{y_i : y_i = f(x_i)\}$. By observing the distorted time series in the lower panel of Fig. 4.1, we might be deceived that the data are obtained from a low dimensional intermittent chaotic system. The nonlinearity is introduced by the measurement function, although the original time series are generated from a linear stochastic process.

With the above considerations, three null hypothesis for nonlinearity detection can be proposed in hierarchy [13]

- NH1: The time series $\{x_i\}$ are i.i.d noise of with unspecified mean and variance.

- NH2: The time series $\{x_i\}$ are generated via $AR(p)$ process

$$X_i = \mu + \sum_{j=1}^{p-1} \phi_j \cdot X_{i-j} + \sigma \cdot \xi_j, \tag{4.3}$$

where μ,$\{\phi_j\}$ and σ have the same meaning as those in Eq. 4.1.

- NH3: The time series $\{x_i\}$ are produced in essence from an $AR(p)$ process described by Eq. (4.3), but the final data set we obtained is the output filtered by a static, monotonic and time-independent nonlinear filter, i.e.,

$$x_i = f(s_i), \tag{4.4}$$

where $\{s_i\}$ are the underlying signals satisfying Eq. (4.3) and f represents a class of nonlinear filters satisfying the above constraints.

4.2.2 Surrogate generation algorithm

Given a time series (see panel (a) of Fig. 4.2), to produce surrogates of NH1, we only need to randomly permute the original time series (panel (b) of Fig. 4.2). If the original time series really follows an i.i.d distribution, then the surrogates shall also keep the same distribution. If the original time series is not i.i.d noise, then shuffling shall destroy any existing temporal relations.

Now we consider how to generate constrained-realization surrogates of NH2. It is well known that the autocorrelation functions of a time series is related to its Fourier coefficients (moduli) [4] [3], if we keep the power spectrum of the surrogates the same as that of the original time series, we also preserve the autocorrelation functions and hence the linear properties of the original time series during the generation of surrogates. By replacing the Fourier phases of the original time series

[3] Autocorrelation function can also be related to parameters of AR(p) model. But due to the finite size of data set, we will usually find somewhat different parameter values from the original time series and spurious results might therefore appear, as shown by Theiler and Prichard [54]. This is also the reason not to apply parametric bootstrap methods in this context.

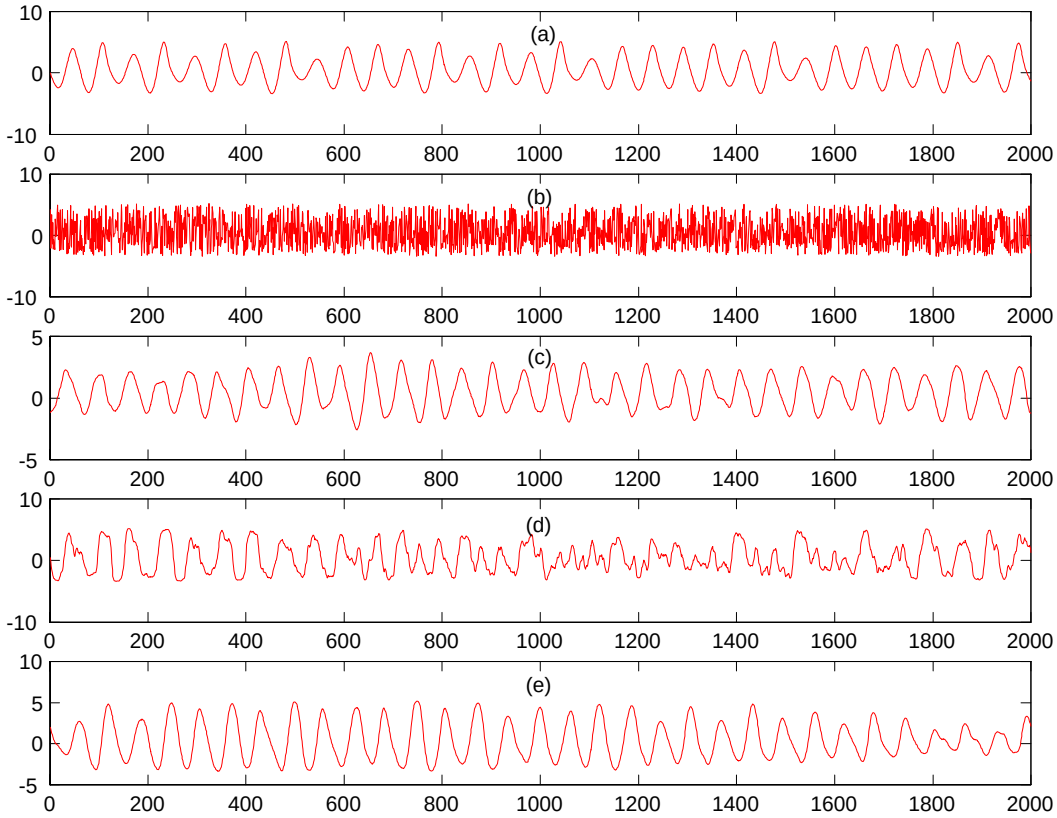

Figure 4.2: (a)Original time series; (b) Surrogates corresponding to NH1; (c) Surrogates corresponding to NH2, generated by FTPR algorithm; (d) Surrogates corresponding to NH3, generated by AAFT algorithm; (e) Surrogates corresponding to NH3, generated by iAAFT algorithm. Realization algorithms are described in the text.

with random phases, the potential nonlinearities are destroyed. Thus a sketch is formed to generate surrogates of NH2 as follows [13]:

1. Apply Fourier transform to the original time series.

2. Record the moduli of the Fourier coefficients of the original time series.

3. Replace the phases of the Fourier coefficients with random values uniformly drawn from interval $(-\pi, \pi]$.

4. Use the original moduli and the new phases to form new Fourier coefficients and conduct an inverse Fourier transformation. Data set thus generated is the required surrogate.

This surrogate generation procedure is called Fourier transformation phase randomization (FTPR in short). For example, see panel (c) of Fig. 4.2.

We now discuss the surrogate generation procedure of NH3. Notice that surrogates generated by the FTPR procedure approximately follow Gaussian distribution as a consequence of the central limit theorem. But after passing through a nonlinear filter, the distribution will typically be distorted, as the example shown in Eq. (4.2). According to the observations of Rapp *et al.* [39], applying FTPR procedure to time series with non-Gaussian distribution might result in spurious rejections of the null hypothesis. Thus for NH3, we need to conduct a transformation on the original time series so that the new time series follows a Gaussian distribution. Then the FTPR can be safely applied to the new time series to generate surrogates. An inverse transformation will be applied on the surrogates of

the new time series and we thus obtain the surrogates of the original time series. The procedure is summarized by the following steps [13]:

1. Produce a random data set drawn from the normal Gaussian distribution.

2. Rank the original time series, i.e., indicate which data point of the original time series is the largest, which is the second largest one, and so on.

3. Arrange the random data set so that the new random data set has exactly the same rank order as the original time series. The new data set is a "Gaussianized" version of the original time series after transformation.

4. By applying the FTPR, we can obtain a surrogate of the new random data set in step 3.

5. Rank the interim surrogate in step 4.

6. Arrange the original time series so that it has exactly the same rank order as that of the surrogate generated in step 4. The arranged time series in this step is the final surrogate of the original time series.

We call this procedure amplitude-adjusted Fourier-transform (AAFT) [13]. Again the example is shown in panel (d) of Fig. 4.2.

Later, Schreiber and Schmitz [43] noticed that the power spectra of the surrogates generated by AAFT algorithm will be slightly biased toward a white spectrum comparing to the power spectrum of the original time series. The reason is that,

when arranging the original time series according to the rank order of the surrogates generated in step 4 of the AAFT, we will distort the power spectra somehow. As a remedy, Schreiber and Schmitz suggested to replace the moduli of the Fourier coefficients of the surrogates generated by the AAFT algorithm with those of the original time series but to preserve the phases of the Fourier coefficients. In this way, the surrogates shall have exactly the same power spectra as that of the original time series but their distributions shall differ from the Gaussian distribution, hence, another AAFT shall be performed. We have to repeat the AAFT procedure and the moduli replacement until some criterion is reached, for example, the difference between the Fourier moduli of the surrogates and those of the original time series does not change after a certain number of iterations. But note that the final step in iteration shall be moduli replacement so that the surrogates we finally obtained have the same power spectra as the original time series. Schreiber and Schmitz call the whole procedure as iterative amplitude-adjusted Fourier-transform (iAAFT). The whole procedure is outlined in the following together with the example shown in panel (e) of Fig. 4.2 [13]:

1. Perform Fourier transformation on the original time series and keep the moduli.

2. Generate an AAFT surrogate of the original time series.

3. Conduct a Fourier transform on the AAFT surrogate to calculate its Fourier moduli and phases. Keep the phases of the surrogate but replace the moduli with those of the original time series.

4. Conduct an inverse Fourier transform and we obtain a new surrogate. Rank this new surrogate.

5. Arrange the original time series so that it has exactly the same rank order as that of the new surrogates in step 4, the arranged time series is the interim surrogate

6. Repeat the above operations (from step 2 to step 5) on the interim in step 5 until a certain criterion is reached, for example, the criterion can be that the deviation of the Fourier moduli of the surrogate from those of the original time series does not vary for 10 consecutive iterations [13].

Remark 7 *It is well known that the implementation of discrete Fourier transform, e.g., FFT, assumes that the data set is drawn from one period of a periodic orbit. In practice, however, a data set usually will not satisfy this condition without careful selection. Spurious high frequencies might be introduced due to the endpoints mismatch of the data set. Theiler and Rapp [55] call this phenomenon "wraparound artifact ". Because of the essential role of Fourier transform in generating surrogates, this problem will exist in the above algorithms, that is, FTPR, AAFT and iAAFT. Remedies proposed can be found in [55], [45] and references therein.*

4.2.3 Choice of Discriminating Statistics, Criterion for Rejection of Null Hypothesis and The Corresponding Interpretation of The Results

After the generation of the surrogates, we need to choose a discriminating statistic for both the surrogates and the original time series. In principle, this statistic shall

keep invariant for all data set if our null hypothesis is true, however, if our null hypothesis is false, it shall turn out to be different for the surrogates and the original time series. In our calculations, we adopt the correlation dimension as the discriminating statistic. Theoretically, the correlation dimension of a stochastic process is infinite [4]. If our null hypotheses (NH1-NH3) are true, then the surrogates and the original time series are actually different realizations of the same stochastic process (for NH3, the final data will be filtered by a nonlinear filter). Therefore the correlation dimensions of the surrogates and the original time series will be statistically the same. If our null hypothesis is not true, it is very possible that the time series is from a nonlinear deterministic system. For a nonlinear (deterministic) time series, applying all of the above three surrogate generation algorithms will destroy the temporal sequence of the original time series, the surrogates obtained will almost always have different structures from the original time series. Hence in this sense, the correlation dimension is a suitable candidate for the discriminating statistic.

After we calculate the correlation dimensions of both the surrogates and the original time series, we need a criterion to judge whether we can reject our null hypotheses or not. Here we adopt the ranking criterion [56]. The idea of this criterion is that, let Q_0 denote the statistic of the original time series and $\{Q_1, Q_2, ..., Q_{N_s}\}$ denote the statistics of the surrogate $1, 2, ...N_s$ respectively. If Q_0 obviously falls outside the distribution of $\{Q_1, Q_2, ..., Q_{N_s}\}$, then we expect that $Q_0 < \min\{Q_1, Q_2, ..., Q_{N_s}\}$

[4] In practice since we can only generate a time series with finite amplitude and finite length, it is of course impossible to actually obtain an infinite correlation dimension .

or $Q_0 > \max\{Q_1, Q_2, ..., Q_{N_s}\}$. However, consider the inverse problem, that is, if $Q_0 < \min\{Q_1, Q_2, ..., Q_{N_s}\}$ or $Q_0 > \max\{Q_1, Q_2, ..., Q_{N_s}\}$, can we know for sure that Q_0 does not follow the distribution of $\{Q_1, Q_2, ..., Q_{N_s}\}$? Under the ranking criterion, the answer is no. For the distribution $\{Q_0, Q_1, Q_2, ..., Q_{N_s}\}$, we can obtain a corresponding rank set $\{r_0, r_1, r_2, ..., r_{N_s}\}$, where $r_i = rank(Q_i)$ with function $rank(\cdot)$ denoting the ordinal number of a value in a list in the increasing order, for example, for $\min\{Q_0, Q_1, Q_2, ..., Q_{N_s}\}$, its rank is 1, while for $\max\{Q_0, Q_1, Q_2, ..., Q_{N_s}\}$, its rank is $N_s + 1$, etc. The ranking criterion assumes that, if $\{Q_0, Q_1, Q_2, ..., Q_{N_s}\}$ follows the same distribution, then for each element Q_i, its rank r_i has an equal probability $(1/N_s + 1)$ to be any integer number between 1 and $N_s + 1$. Therefore, in a surrogate test, if we find the statistic of the original time series $Q_0 < \min\{Q_1, Q_2, ..., Q_{N_s}\}$ or $Q_0 > \max\{Q_1, Q_2, ..., Q_{N_s}\}$, under the ranking criterion, we have two alternative interpretations. One is that Q_0 does not follow the distribution of $\{Q_1, Q_2, ..., Q_{N_s}\}$. The other one is that, Q_0 follows the same distribution as that of $\{Q_1, Q_2, ..., Q_{N_s}\}$, however, it happens that Q_0 is the smallest or the largest one. Therefore in the surrogate test, if we reject a null hypothesis according to the condition $Q_0 < \min\{Q_1, Q_2, ..., Q_{N_s}\}$ or $Q_0 > \max\{Q_1, Q_2, ..., Q_{N_s}\}$, it is possible that we will falsely reject the null hypothesis. The probability of a false rejection is $1/(N_s + 1)$ for a one-sided test (the statistic of the original time series cannot be larger (or smaller) than those of the surrogates), or $2/(N_s + 1)$ for a two-sided test (the statistic of the original time series can be either larger or smaller than those of the surrogates).

Another problem to be discussed here is the interpretations of the results. As we introduced previously, the surrogate test does not tell us which property a time series has, rather it tells us which property a time series lacks by rejecting the corresponding null hypothesis. If we fail to reject a null hypotheses, our null hypothesis might be true, however, we shall also note the possibility that our discriminating statistic is not powerful enough to detect the difference between the surrogates and the original time series, in such situations, we shall have to seek more powerful statistics [44]. Successful rejection of null hypotheses in most of the cases can provide accurate information, but we shall still be careful when interpreting the results. As we mentioned previously, for NH3 we actually restrict the distortion function to be static,monotonic and time-independent for the purposes of simplicity and feasibility, but sometimes other nonlinear filtering functions might also lead to the rejection of our null hypothesis [20].

4.2.4 An Example

We use an example to demonstrate the above procedures. The data set is generated from the Rössler system according to Eq. (2.18), the corresponding parameters are set to be $a = 0.395$, $b = 2$, $c = 4$ and the sampling time $\Delta t_s = 0.1$ time unit, which makes the Rössler system exhibit one-band chaotic behavior. We plot the time series in panel (a) of Fig. 4.3.

Using the corresponding surrogate generation algorithm, we generate 100 surrogates for each null hypothesis test (NH1-NH3). We adopt the GKA (chapter 3) to calculate the correlation dimension for the original time series and the surro-

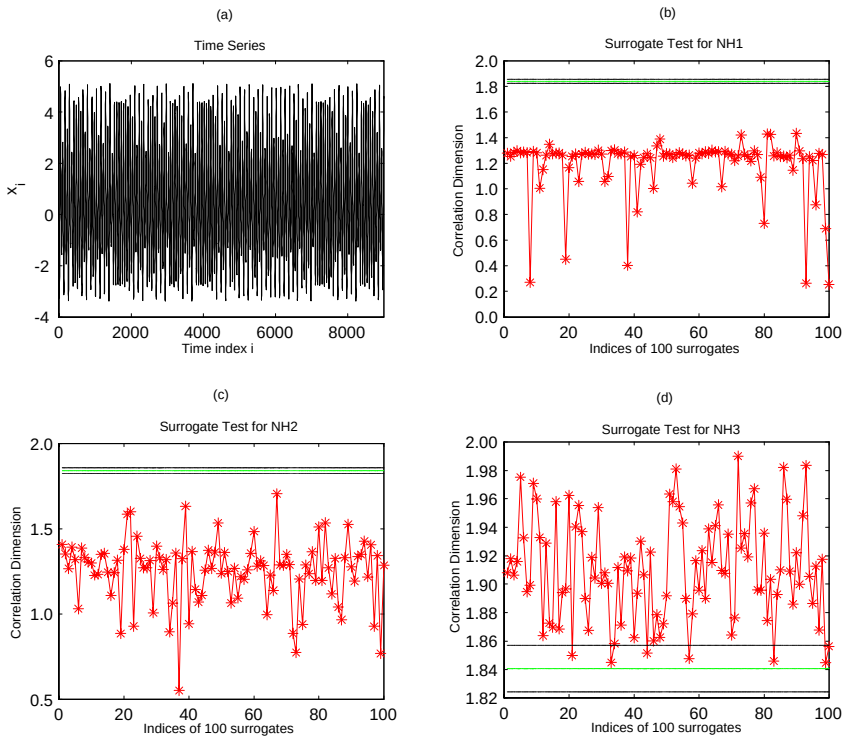

Figure 4.3: (a) Waveform of the time series. (b) Surrogate test for NH1. The abscissa is the indices of 100 surrogates and the ordinate is the corresponding correlation dimensions. The middle line is the mean correlation dimension of the original time series calculated 100 times using the GKA, the upper and lower lines denote the correlation dimensions twice the standard deviation away from the mean value and the asterisks connected with lines indicate the correlation dimensions of 100 surrogates. (c) Surrogate test for NH2. The meaning of the lines and asterisks is the same as that in panel (b). (d)Surrogate test for NH3. The meaning of the lines and asterisks is the same as that in panel (b).

gates. Through the false nearest neighbour criterion (chapter 2), we choose the embedding dimension $m = 3$ for all data sets. We also adopt the RITE algorithm (chapter 2) to choose the time delay. Note that to speed up the calculation, only

2000 data points are used as the reference points for the GKA. There are some statistical fluctuations even for the same data set when calculating its correlation dimension, therefore for the original time series, we will calculate 100 times to estimate the mean correlation dimension and the standard deviation. As shown in panel $(b), (c)$ and (d) of Fig. 4.3 [5], in each panel there are three lines parallel to the abscissa. The middle line denote the estimation of the mean correlation dimension of the original time series, while the upper and lower lines indicate the positions twice the standard deviation away from the mean value. For the surrogates, however, we will calculate their correlation dimensions only once to save time. We plot the correlation dimensions of the surrogates marked with asterisks and connected with solid lines in each panel.

Using the ranking criterion introduced previously, we can reject all of the null hypothesis (as we expect). By rejecting NH1, we can see that the time series is very unlikely to be an i.i.d noise realization (the probability of a false rejection is 2%, the same for the following tests). By rejecting NH2, we can claim it is almost impossible that the time series is from a linear stochastic process. By rejecting NH3, we also exclude most of the probability that the time series is from certain linear stochastic process which passes through a monotonic, static and time-invariant nonlinear filter. Together with these three hierarchy surrogate test, we may conclude that the time series is very unlikely to be generated from a linear stochastic process. Instead, it is more likely that there exists (nonlinear)

[5] Note for all panels, the correlation dimensions of the original time series are the same.

determinism instead.

4.3 Surrogate Tests to Distinguish between Pseudoperiodicity and Chaos

4.3.1 Null Hypothesis for A New Surrogate Generation Algorithm for Pseudoperiodic Time Series

In this section we will employ the surrogate tests to distinguish between pseudoperiodicity and chaos. For a nonlinear irregular time series, in most cases it is either pseudoperiodic or chaotic. By pseudoperiodic time series we mean a representative of a periodic orbit perturbed by dynamical noise, or contaminated by observational noise, or with the combination of both types of noise, whose states within one cycle are largely independent of those within the previous cycles for a given cycle length. Note that chaotic and pseudoperiodic time series often look similar, we might not be able to distinguish them from each other only through visual inspections, quantitative techniques will be needed instead. Once again we recommend use of the surrogate test technique in such situations.

Initially, to generate surrogates for pseudoperiodic time series, Theiler [53] proposed the cycle shuffled algorithm. The idea is to divide the whole data set into some segments and let each segment contain exactly an integer number of cycles. The surrogates are obtained by randomly shuffling these segments, which will preserve the intracycle dynamics but destroy the intercycle ones by randomizing the temporal sequence of the individual cycles. The difficulty in applying this algorithm is that it requires *a priori* knowledge of the precise periodicity, otherwise shuffling the individual cycles might lead to spurious results [50]. Therefore, in or-

der to avoid introducing contradictory results, we will not apply it to the surrogate tests in all of the following computations.

Later, Small *et.al* [50] proposed the pseudoperiodic surrogate (PPS) algorithm from another viewpoint. They first apply the time delay embedding reconstruction [59] to the original data set, then utilize a method based on local linear modelling techniques to produce surrogate data which approximate the behavior of the underlying dynamical system. As the authors pointed out, this algorithm works well even with very large dynamical noise, but it may incorrectly reject the null hypothesis if the intercycles of the pseudoperiodic orbit have a linear stochastic dependence induced by colored additive observational noise.

In this section we will propose a new surrogate generation algorithm [28] for continuous dynamical systems, which properly copes with linear stochastic dependence between the cycles of the pseudoperiodic orbits. The null hypothesis to be tested is that the stationary data set is pseudoperiodic with noise components which are (approximately) identically distributed and uncorrelated for sufficiently large temporal translations. Note that the constraints of the noise components in our null hypothesis are stronger than those of Theiler's algorithm, which requires the noise distribution only periodically depends on the phase of the signal. However, under our hypothesis, we can produce the surrogates in a simple way through the algorithm to be described below. In addition, a large scope of noise processes often encountered in practical situations, including (but not limited to) linear colored additive observational noise described by the $ARMA(p, q)$ model [4], match

the above constraints.

4.3.2 Surrogate Generation Algorithm

Let $\{x_i\}_{i=1}^{N}$ be a data set with N observations (the form $\{x_i\}$ will be adopted instead for convenience when there is no ambiguity), where x_i means the observation measured at time $t_i = i \cdot \Delta t_s$ with Δt_s denoting the sampling time. We assume $\{x_i\}_{i=1}^{N}$ is stationary and can be decomposed into the deterministic components and the noise components, which are approximately independent of each other. Similar to the surrogate test idea of time shifting to desynchronize two data sets [38], we also assume the noise components (approximately) follow an identical distribution and are uncorrelated for sufficiently large temporal translations (or time shifts). According to the null hypothesis we proposed in the previous section, the deterministic components are periodic, then we can write a data point x_i as $x_i = p_i + n_i$, where p_i and n_i denote the periodic component and the noise component respectively. In many cases, we can set $E(p_i) = E(n_i) = 0$ where E is the expectation operator. Since $\{p_i\}$ are roughly independent of $\{n_i\}$, we have that the autocovariance $var(x_i) = var(p_i) + var(n_i)$. Let

$$y_i^{\tau} = \alpha x_i + \beta x_{i+\tau} = (\alpha p_i + \beta p_{i+\tau}) + (\alpha n_i + \beta n_{i+\tau}) \tag{4.5}$$

with $i = 1, 2, \ ..., N - \tau$, where coefficients α and β satisfy $\alpha^2 + \beta^2 = 1$ and parameter τ is the temporal translation between subsets $\{x_i\}_{i=1}^{N-\tau}$ and $\{x_{i+\tau}\}_{i=1}^{N-\tau}$, then the autocovariance function $var(y_i^{\tau}) = var(\alpha p_i + \beta p_{i+\tau}) + var(\alpha n_i + \beta n_{i+\tau})$. Consider the noise component, if τ is sufficiently large, under our hypothesis, n_i and $n_{i+\tau}$ are uncorrelated. Also note $\{n_i\}$ and $\{n_{i+\tau}\}$ are drawn from (approximately)

the same distribution, we have $var(\alpha n_i + \beta n_{i+\tau}) = var(n_i)$. For the deterministic component, if we require the translation τ to satisfy $cov(p_i, p_{i+\tau}) = 0$, then $var(\alpha p_i + \beta p_{i+\tau}) = var(p_i)$. Hence by choosing a suitable temporal translation, the noise levels of $\{y_i^\tau\}$, defined by $(var(\alpha n_i + \beta n_{i+\tau})/var(y_i^\tau))^{1/2}$, will be the same as that of $\{x_i\}_{i=1}^N$, i.e., $(var(n_i)/var(x_i))^{1/2}$. The reason to preserve the noise level is that, the presence of noise will affect the calculation of the correlation dimension, hence we would like to let the surrogates and the original time series (roughly) have the same noise level in order to make the results be more reliable.

The above deduction leads to the central idea of our surrogate algorithm. From Eq. (4.5), we note that if $\{p_i\}$ is periodic, the nonconstant deterministic components $\{\alpha p_i + \beta p_{i+\tau}\}$ shall also be periodic. In addition, $\{x_i\}_{i=1}^N$ and $\{y_i^\tau\}$ shall have the same noise level if a suitable translation τ is selected. Therefore by randomizing the coefficient α or β, we can generate many data sets $\{y_i^\tau\}$ as the surrogates of $\{x_i\}_{i=1}^N$. Note that $\{p_i\}$ and $\{\alpha p_i + \beta p_{i+\tau}\}$ have the same degree-of-freedom, if both of them are periodic, their correlation dimensions [14] will theoretically be the same. Now let us consider the noise components. Although the noise components $\{\alpha n_i + \beta n_{i+\tau}\}$ may have a different distribution from that of $\{n_i\}$, the noise level is preserved after the transform in Eq. (4.5). As Diks [9] has reported, the Gaussian kernel algorithm (GKA) can reasonably estimate the correlation dimensions of noisy data sets with different noise distributions. This implies that, under the same noise level, the correlation dimensions of $\{x_i\}_{i=1}^N$ and $\{y_i^\tau\}$, calculated by the GKA, shall statistically be the same if $\{x_i\}_{i=1}^N$ and $\{y_i^\tau\}$ are

both pseudoperiodic (and satisfy the constraints we imposed). In contrast, if $\{p_i\}$ is chaotic, its linear combination, $\{\alpha p_i + \beta p_{i+\tau}\}$, may have a new dynamical structure with a different correlation dimension from that of $\{p_i\}$, hence by adopting the correlation dimension as the discriminating statistic we might detect this difference.

We shall also note that, for an unstable periodic orbit, even a small dynamical noise might drive the resultant orbit far away from the original position after a sufficiently long time, and the pseudoperiodicity might be lost. In such situations, our algorithm might fail to work. Nevertheless, we suggest applying our algorithm as the first step in pseudoperiodicity test. This algorithm is computationally fast and in principle deals well with a large scope of observational noise (comparatively, the PPS algorithm will sometimes fail for colored observational noise). If we can reject the null hypothesis proposed previously, the time series under test is possibly chaotic or pseudoperiodic perturbed by dynamical noise. Then we can adopt the PPS algorithm for further tests, which will work well even under a large amount of dynamical noise. If the corresponding null hypothesis, i.e., the time series is pseudoperiodic perturbed by dynamical noise, can be rejected again, then we may claim the time series is likely to be chaotic.

We now consider several computational issues in our algorithms. As described in Eq. (4.5), to generate the surrogates $\{y_i^\tau\}$, we select two subsets of $\{x_i\}_{i=1}^{N}$–$\{x_i\}_{i=1}^{N-\tau}$ and $\{x_{i+\tau}\}_{i=1}^{N-\tau}$–multiply them by the coefficients α and β respectively and then add them together. We shall emphasize that choosing the temporal translation τ is a crucial issue for our algorithm. From one aspect, we require

the translation τ to satisfy the condition $cov(p_i, p_{i+\tau}) = 0$. The reason is that we want to keep the noise level for the original time series and the surrogates. In addition, we want the deterministic components $\{\alpha p_i\}$ to be orthogonal to $\{\beta p_{i+\tau}\}$ for arbitrary coefficients α and β, otherwise the projection of $\{\alpha p_i\}$ onto $\{\beta p_{i+\tau}\}$ might counteract $\{\beta p_{i+\tau}\}$ under some situations, for example, if $p_i \approx -p_{i+\tau}$ and $\alpha = \beta$, the deterministic components $\{\alpha p_i + \beta p_{i+\tau}\}$ will almost vanish while the noise components $\{\alpha n_i + \beta n_{i+\tau}\}$ remain. In this situation, the correlation dimensions calculated would actually be those of the noise components instead of the deterministic components, which will certainly cause the false rejection of the null hypothesis. From another aspect, we require τ to be sufficiently large to guarantee the decorrelation between the noise components. However, we expect $\{x_i\}_{i=1}^{N-\tau}$ and $\{x_{i+\tau}\}_{i=1}^{N-\tau}$ will have at least some overlaps to make use of the information of the whole data set $\{x_i\}_{i=1}^{N}$ (i.e., τ shall not exceed $N/2$). In addition, it is recommended the length of a data set shall not be too short in order to appropriately calculate its correlation dimension, which also implies τ shall not be too large.

Moreover, from Eq. (4.5) we note that the coefficient ratio $|\alpha/\beta|$ shall not be too large or too small, otherwise $\{y_i^\tau\}$ will be very close to $\{x_i\}_{i=1}^{N-\tau}$ or $\{x_{i+\tau}\}_{i=1}^{N-\tau}$, which will lead to approximately the same correlation dimensions of $\{x_i\}_{i=1}^{N}$ and $\{y_i^\tau\}$ regardless of the dynamical behavior of $\{x_i\}_{i=1}^{N}$, and thus decrease the discriminating power of the correlation dimension. To avoid this situation, we require $\{x_i\}_{i=1}^{N-\tau}$ and $\{x_{i+\tau}\}_{i=1}^{N-\tau}$ to contribute equally when producing the surrogates. Thus, ideally, we expect $|\alpha/\beta| \simeq 1$ and coefficient $|\alpha|$ (we let $\beta = \sqrt{1-\alpha^2}$) to be

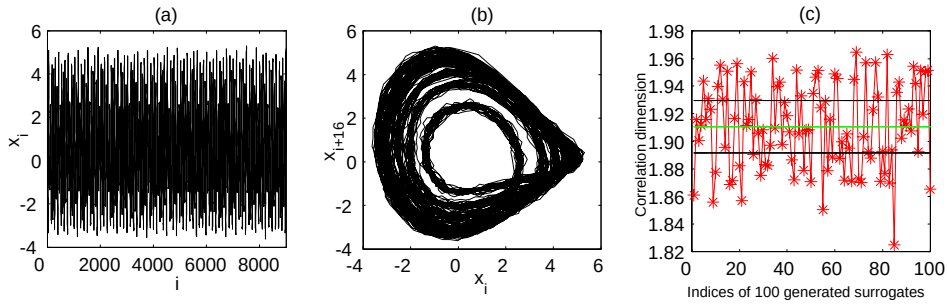

Figure 4.4: (a) Pseudoperiodic time series contaminated by observational noise; (b) State space x_{i+n} vs. x_i of the pseudoperiodic time series from the Rössler system with $n = 16$; (c) Surrogate test for pseudoperiodic time series based on our new algorithm. The meaning of the lines and asterisks is the same as that in panel (b) of Fig. 4.3.

uniformly drawn from an interval $[p, q]$, which satisfies

$$p^2 + q^2 = 1, 0 < p < q \text{ and } q - 1/\sqrt{2} = 1/\sqrt{2} - p \tag{4.6}$$

so that probability $\Pr(|\alpha/\beta| \geqslant 1) = \Pr(|\beta/\alpha| \geqslant 1)$, therefore $\{x_i\}_{i=1}^{N-\tau}$ and $\{x_{i+\tau}\}_{i=1}^{N-\tau}$ will equally affect the generation of the surrogates $\{y_i^\tau\}$. Unfortunately, condition $q - 1/\sqrt{2} = 1/\sqrt{2} - p$ can only be achieved when $p = q = 1/\sqrt{2}$, we have to require $q - 1/\sqrt{2} \simeq 1/\sqrt{2} - p$ instead. In our calculations we let α be uniformly drawn from the interval $[-0.8, -0.6] \cup [0.6, 0.8]$ and $\beta = \sqrt{1 - \alpha^2}$, which can satisfy our requirements well. But note that other choices are of course also available.

4.3.3 Examples

We now use two examples to demonstrate the application of our algorithm. The data sets are also generated from the Rössler system according to Eq. (2.18). We

fix parameters $b = 2$, $c = 4$ and the sampling time $\Delta t_s = 0.1$ time unit. For the first example, parameter $a = 0.39095$, which makes the Rössler system exhibit limit cycle behavior (period 6). We then introduce 5% observational noise into the periodic time series and plot it in panel (a) of Fig. 4.4 [6]. For comparison (to the second example), we also plot the corresponding attractor in two dimensional embedding space in panel (b) of Fig. 4.4. Although Gaussian white observational noise is the most common choice in this situation, in order to demonstrate the ability of our surrogate algorithm to deal with colored noise, we will instead adopt the noise generated from the $AR(1)$ process $\xi_{i+1} = 0.8\xi_i + \epsilon_i$ with the variable ϵ following the normal Gaussian distribution $N(0, 1)$, which is the more difficult case due to the correlation between noise components. However, we note that, Gaussian white noise and other colored noises satisfying the constraints in our null hypothesis (for example, those modelled by the $ARMA(p, q)$ processes), can be dealt with in the same way.

To produce surrogate data, first we shall choose a suitable temporal translation. Since it is impractical to separate noise from signal completely, in general it is difficult to estimate the correlation decay time between noise components. Fortunately, to decorrelate noise components, all temporal translations are equivalent as long as they are large enough. In addition, in many real situations, it is often possible to observe the background noise and thus estimate the decay time. In this example, we think the $AR(1)$ noise will be uncorrelated when the temporal transla-

[6] For the case that time series is driven by small dynamical noise, see examples in Ref. [28] for more details.

tion is larger than 50 (in units of the sampling time Δt_s). As another requirement, temporal translation satisfying $cov(p_i, p_{i+\tau}) = 0$ is desired. In practice, of course, this requirement is generally impractical due to digitization and quantization in the sampling process. Recall the discussion in the previous section, by letting $\alpha^2 + \beta^2 = 1$, we have $var(\alpha p_i + \beta p_{i+\tau}) = var(p_i) + 2\alpha\beta \cdot cov(p_i, p_{i+\tau})$. Function $cov(p_i, p_{i+\tau}) \neq 0$ means we do not preserve the noise level. However, under the null hypothesis of pseudoperiodicity, there shall always be some temporal translations to make $cov(p_i, p_{i+\tau}) \sim 0$, hence the noise level will not deviate from the original one too much. Besides, according to Eq. (4.5), we generate the surrogates by uniformly drawing coefficient α from interval $[-0.8, -0.6] \cup [0.6, 0.8]$ ($\beta = \sqrt{1-\alpha^2}$ is always kept positive), the noise level of the surrogates will fluctuate around that of the original one due to the alternative signs of product $\alpha\beta$. Therefore, $cov(p_i, p_{i+\tau}) \neq 0$ will only cause some fluctuations to the calculated correlation dimension because of the fluctuations of noise level, however, generally such fluctuations will not affect our conclusion if we can select a temporal translation τ to let $cov(p_i, p_{i+\tau}) \sim 0$. Since we have assumed the noise components are roughly independent of the deterministic components, then $cov(x_i, x_{i+\tau}) = cov(p_i, p_{i+\tau})$ for a large enough temporal translation (to decorrelate noise components), therefore in order to let $cov(p_i, p_{i+\tau}) \sim 0$, we can equivalently require $cov(x_i, x_{i+\tau}) \sim 0$. In the first example, there are many temporal translations satisfying the two constraints discussed above, i.e., $\tau > 50$ and $cov(x_i, x_{i+\tau}) \sim 0$. To pick a value from all these candidates, we first select an interval $[100, 150]$, then search the temporal transla-

tion which makes the absolute value $|cov(x_i, x_{i+\tau})|$ be the minimum (most close to zero) among all translations $100 \leqslant \tau \leqslant 150$. Note that the choice of the interval $[100, 150]$ is arbitrary, except that we have to make sure that the lower bound of the interval is large enough to decorrelate noise components, and that there exists temporal translations to let $cov(x_i, x_{i+\tau}) \sim 0$ within the interval. After selecting the temporal translation, by randomizing the coefficient α we will generate 100 surrogates according to Eq. (4.5).

To recover the underlying system from the scalar time series, two parameters, i.e., embedding dimension and time delay, shall be properly chosen to apply the technique of time delay embedding reconstruction. We still use the false nearest neighbour criterion to determine the global optimal embedding dimension, in the first example, the embedding dimension is chosen at $m = 4$. Again we use the RITE algorithm to choose the time delay since it is more convenient to be used for the surrogate tests. The correlation dimensions of all data sets calculated by the GKA are indicated in panel (c) of Fig. 4.4, from which we can see that the mean value of the correlation dimension of the original time series falls within the dimension distribution of the surrogates, therefore we cannot reject our null hypothesis, in other words, the time series is possibly pseudoperiodic as we expect.

In the second example, we set parameter a in Eq. (2.18) to be $a = 0.395$. The Rössler system exhibits single-band chaotic behavior. We also introduce the same $AR(1)$ noise as that in the first example into the obtained time series, which, together with the corresponding reconstructed attractor in two dimensional em-

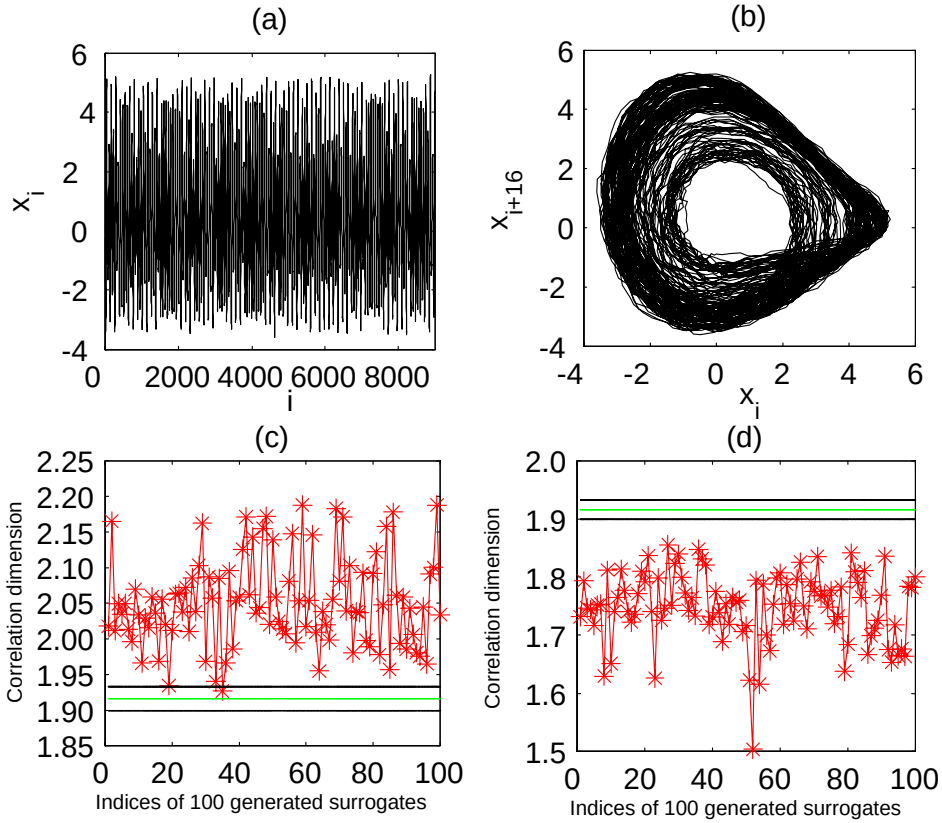

Figure 4.5: (a) Chaotic time series contaminated by observational noise; (b) State space x_{i+n} vs. x_i of the chaotic time series from the Rössler system with $n = 16$; (c) Surrogate test for chaotic time series based on our new surrogate generation algorithm. The meaning of the lines and asterisks is the same as that in panel (b) of Fig. 4.3. (d) Surrogate test for chaotic time series based on the PPS algorithm. The meaning of the lines and asterisks is the same as that in panel (b) of Fig. 4.3.

bedding space, are plotted in panel (a) and (b) of Fig. 4.5 respectively.

Similar to the first example, to generate the surrogates, we search for a suitable temporal translation within the interval $[100, 150]$. By uniformly drawing a value from interval $[-0.8, -0.6] \cup [0.6, 0.8]$ for the coefficient a ($\beta = \sqrt{1-\alpha^2}$), we produce

100 surrogates from Eq. (4.5).

The surrogate test based on our new algorithm is indicated in panel (c) of Fig. 4.5, from which we see that the mean value of the correlation dimension of the original time series falls outside the dimension distribution of the surrogates. Therefore under the ranking criterion we can reject our hull hypothesis, the probability of a false rejection is 2%. Note that, to exclude the possibility that the time series is produced by a periodic orbit driven by large dynamical noise, we also perform a surrogate test based on the PPS algorithm, which deals well with dynamical noise. The corresponding test results are indicated in panel (d) of Fig. 4.5, from which we can also reject the null hypothesis that the time series is pseudoperiodic driven by dynamical noise. Together with these two surrogate tests, we can claim that the time series in the second example is more likely to be chaotic than pseudoperiodic.

4.4 Application to Experimental Data

From the above discussions we have seen the generic procedure to apply the surrogate test technique. Here we will apply this technique to two experimental data sets to demonstrate the whole procedures of nonlinearity and chaos detection. Note that, for the detection of chaos, since the cycle shuffled algorithm sometimes will lead to spurious results, we will not apply it to the surrogate tests in order to avoid introducing contradictory results. Of course, in principle one can apply the procedures to other data in the same way.

The first data set, as shown in left panel of Fig. 4.6, are the observations

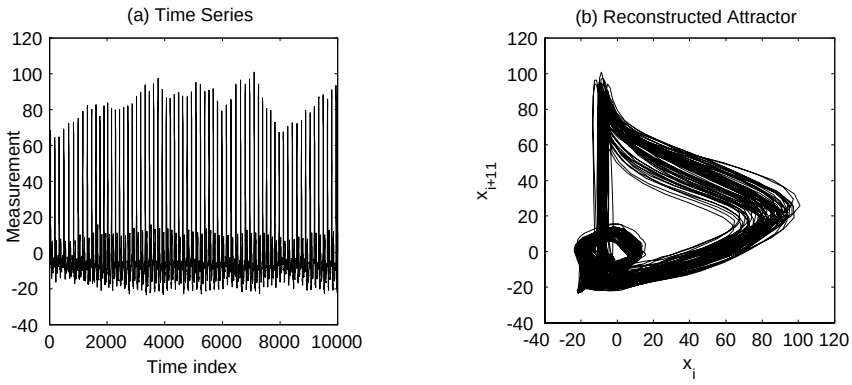

Figure 4.6: (a) The observations of human pulse; (b) The reconstructed attractor in two dimensional embedding space.

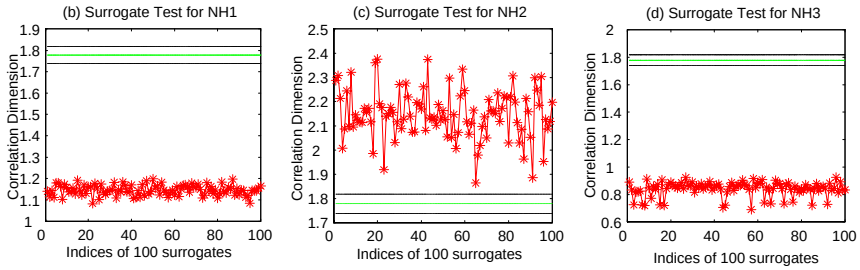

Figure 4.7: Surrogate test to detect nonlinearity in human pulse data. (a) Surrogate test for NH1. The meaning of the lines and asterisks is the same as that in panel (b) of Fig. 4.3; (b) Surrogate test for NH2. The meaning of the lines and asterisks is the same as that in panel (b) of Fig. 4.3; (c) Surrogate test for NH3. The meaning of the lines and asterisks is the same as that in panel (b) of Fig. 4.3.

of human pulse, measured from the forefinger of the right hand at a sampling frequency of 150Hz. The corresponding reconstructed attractor in two dimensional embedding space is indicated in the right panel of Fig. 4.6. Using the nearest neighbour criterion, the suitable embedding dimension is chosen at $m = 7$, while for the suitable time delay, we find $\tau = 11$ through the RITE algorithm.

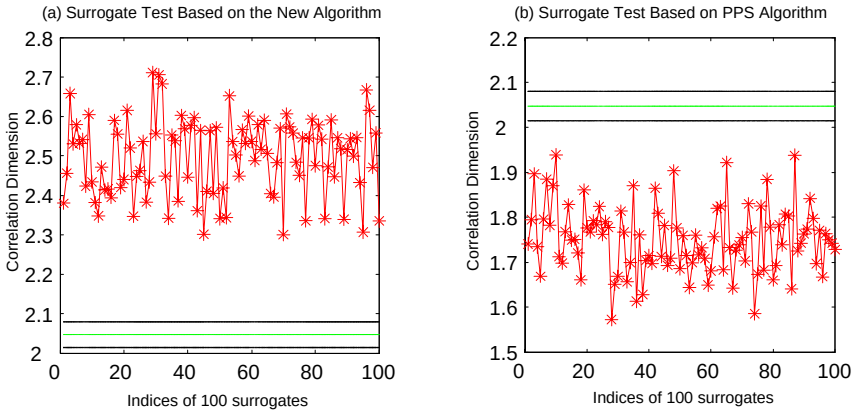

Figure 4.8: Surrogate test to detect chaos in human pulse data. (a) Surrogate test for pseudoperiodicity based on the new algorithm. The meaning of the lines and asterisks is the same as that in panel (b) of Fig. 4.3; (b) Surrogate test for pseudoperiodicity based on the PPS algorithm. The meaning of the lines and asterisks is the same as that in panel (b) of Fig. 4.3.

Since we are going to inspect the nonlinearity in the pulse data through surrogate test technique, the first step is to perform the hierarchical tests for null hypotheses NH1, NH2 and NH3 presented previously, which aims to examine whether the time series is generated from a potentially filtered linear stochastic process. Using the associated surrogate generation algorithms introduced above, we produce 100 surrogates for each test. We still adopt the correlation dimension as the discriminating statistic. To calculate the correlation dimension of the surrogates, we keep the embedding dimension $m = 7$ but choose a suitable time delay for each surrogate via the RITE algorithm. The corresponding results are shown in Fig. 4.7, from which we can see that, for the surrogate test for each null hypothesis, the mean correlation dimension of the original time series falls outside the distribution

of the correlation dimensions of the surrogates. Using the ranking criterion we can therefore reject all of the null hypotheses, in other words, the pulse data is most likely to be generated from a nonlinear deterministic system (human heart).

After the detection of nonlinearity, the further step is to investigate whether the time series is possibly chaotic or pseudoperiodic. Usually the pulse of a healthy subject will have some regular patterns, an irregular pulse means disorder or even heart disease (such as sinus arrhythmia, atrial fibrillation etc, cf. [19]). We first apply the surrogate test based on the our new algorithm to the pulse data. By observing the (second order linear) autocorrelation function of the pulse, we let the temporal translation be selected from interval $[100, 180]$, and coefficient a is uniformly drawn from interval $[-0.8, -0.6] \cup [0.6, 0.8]$. Finally we generate 100 surrogates based on the algorithm described in the last section. The correlation dimension is kept as the discriminating statistic. In calculations, we let embedding dimension $m = 7$ and time delay be chosen by the RITE algorithm. The result of the surrogate test is indicated in panel (a) of Fig. 4.8. As we can find, the correlation dimensions of the surrogates are all larger than the mean value of the original time series, therefore we can reject our null hypothesis. We would also like to include an additional test based on the PPS algorithm in case that the pulse data contains large dynamical noise. We generate 100 surrogates for test, and choose the correlation dimension as the discriminating statistic. The result of test is plotted in panel (b) of Fig. 4.8. Once again we can reject the null hypothesis using the ranking criterion. Therefore, together with these two tests

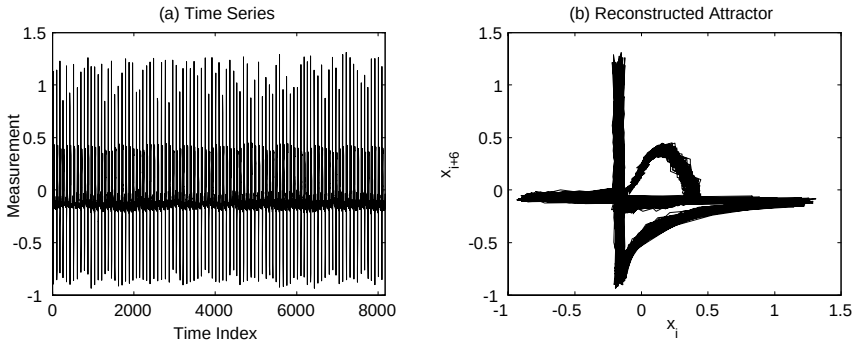

Figure 4.9: (a) The observations of ECG data; (b) The reconstructed attractor in two dimensional embedding space.

on pseudoperiodicity, we can claim that the pulse data is very likely to be chaotic rather than pseudoperiodic.

The second data set, as shown in left panel of Fig. 4.9, are the observations of human electrocardiogram (ECG), measured from human chest at a sampling frequency of 150Hz. The corresponding reconstructed attractor in two dimensional embedding space is indicated in the right panel of Fig. 4.9. We again apply the nearest neighbour criterion to choose the suitable embedding dimension at $m = 10$, while for the suitable time delay, we find $\tau = 6$ through the RITE algorithm.

Similar to the procedure for nonlinearity detection in the first data set, we generate 100 surrogates for NH1, NH2 and NH3 and then calculate the correlation dimensions of both the surrogates and the original time series. The results are indicated in Fig. 4.10. For the surrogate test of each null hypothesis (NH1-NH3), the mean correlation dimension of the original time series falls outside the distribution of the correlation dimensions of the surrogates. Using the ranking criterion we can reject all of the null hypotheses, hence, the ECG data is most

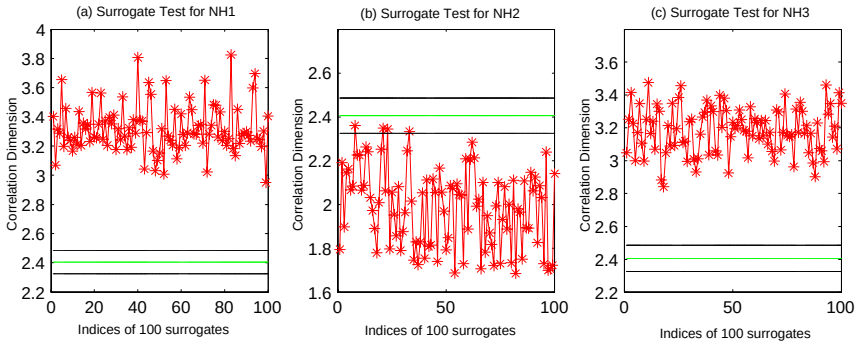

Figure 4.10: The surrogate test to detect nonlinearity in human ECG data. (a) Surrogate test for NH1. The meaning of the lines and asterisks is the same as that in panel (b) of Fig. 4.3; (b) Surrogate test for NH2. The meaning of the lines and asterisks is the same as that in panel (b) of Fig. 4.3; (c) Surrogate test for NH3. The meaning of the lines and asterisks is the same as that in panel (b) of Fig. 4.3.

likely to be (nonlinear) deterministic than (linear) stochastic.

To detect chaos in the data set, we choose a temporal translation from interval $[100, 200]$ (by observing the linear autocorrelation function). Coefficient a is uniformly drawn from interval $[-0.8, -0.6] \cup [0.6, 0.8]$ again. We first generate 100 surrogates through our algorithm. The corresponding correlation dimensions of the surrogates and the surrogates are indicated in panel (a) of Fig. 4.11. As we can find, the mean correlation dimension of the original time series falls within the dimension distribution of the surrogates, therefore we cannot reject our null hypothesis. We also conduct an additional test based on the PPS algorithm to verify the results. We generate 100 surrogates for the test again, the corresponding correlation dimensions of the surrogates and the original time series are plotted in panel (b) of Fig. 4.11. The second test results are consistent with first one, that is

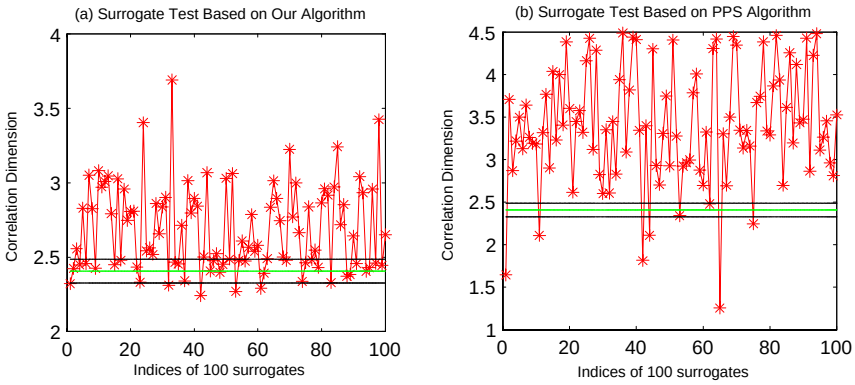

Figure 4.11: Surrogate test to detect chaos in human ECG data. (a) Surrogate test based on our new algorithm. The meaning of the lines and asterisks is the same as that in panel (b) of Fig. 4.3; (b) Surrogate test based on the PPS algorithm. The meaning of the lines and asterisks is the same as that in panel (b) of Fig. 4.3.

once again we cannot reject the null hypothesis using the ranking criterion.

4.5 Brief Summary

In this chapter we have introduced the surrogate test technique. To apply this technique, the first step is to propose a hypothesis describing the property that we are interested in. The other associated procedures in a surrogate test shall also include the design of the surrogate generation algorithm and the choice of discriminating statistic. In principle, the generated surrogates shall preserve the potential property proposed by our null hypothesis while destroying all others. Accordingly, the discriminating statistic shall be an invariant for the surrogates and the original time series if the null hypothesis is true, meanwhile, if the null hypothesis is false, the discriminating statistic shall have the power to detect the

difference between the surrogates and the original time series.

As an application, we adopt this technique to detect the nonlinearity in irregular time series, which is often the preparative step to investigate the underlying system of the data. We also applied the surrogate test technique to distinguish between pseudoperiodicity and chaos, as a natural issue to proceed after the positive identification of nonlinearity in a signal.

CHAPTER 5

CONCLUSION

In this thesis we have studied the application of the surrogate test technique, together with the technique of dimension analysis, to detect nonlinearity and chaos in irregular time series. By generating many surrogates from the original time series, it can provide us with a confidence level even when we only have a very limited amount of data. But note that, the surrogate test does not directly tell us whether a time series possesses a property that we are interested in, instead, it can only tell us that the data does not possess a property described by our assumption. We should be careful to interpret the result when we cannot reject the null hypothesis. Besides the possibility that our null hypothesis is true, we shall also notice other possibility, for example, the discriminating statistic is not powerful enough, or, our null hypothesis is not generic enough to cover the general situations.

The main contribution of this thesis lies in two aspects. One is that, in Chapter Two we proposed an efficient algorithm to choose suitable time delay [27], which is especially useful for dimension analysis in a surrogate test, where a large amount of data sets might be involved. The other is that, in Chapter Four we designed a surrogate generation algorithm for pseudoperiodic time series [28], which can work well under a range of colored observational noise. This algorithm, together with the PPS algorithm, can provide a schematic surrogate test to detect chaos in a

time series.

As we discussed previously, when applying the FFT in the generation of surrogates for NH1-NH3, we actually implicitly assume the time series is periodic with a cycle length equal to the length of the time series, it is not true of course. Recently we proposed an alternative surrogate generation algorithm [32], which aims to not only avoid adopting the FFT but also generate the surrogates in a more effective way. The main idea is that, given a stationary time series $\{x_i\}_{i=1}^N$, we presume that it is generated from a linear stochastic process but might be filtered by a nonlinear filter (that is, we actually adopt the same null hypotheses, i.e., NH1, NH2 and NH3). First let us consider the surrogate generation for NH1 and NH2, if $\{x_i\}_{i=1}^N$ is from a stationary linear stochastic process, then usually it can be modelled by an $ARMA$ process [4]. Note for data set $\left\{y_i^\tau : y_i^\tau = \alpha x_i + (1-\alpha^2)^{1/2} x_{i+\tau},\ i = 1, 2, ..., N-\tau\right\}$, if temporal translation τ is large enough to decorrelate x_i and $x_{i+\tau}$, then $\{y_i^\tau\}_{i=1}^{N-\tau}$ and $\{x_i\}_{i=1}^N$ will follow the same $ARMA$ process (but note they are different realizations for the same stochastic process), therefore, $\{y_i^\tau\}_{i=1}^{N-\tau}$ can be treated as the surrogate for both NH1 and NH2. Randomizing coefficient α, we can obtain many surrogates. To produce the surrogates for NH3, we adopt the same way as that of the AAFT algorithm, that is, given a stationary time series $\{x_i\}_{i=1}^N$, we generate a "Gaussianized" version $\{G_i\}_{i=1}^N$of $\{x_i\}_{i=1}^N$, based on NH1 and NH2, we produce an interim surrogate $\{y_i^\tau\}_{i=1}^{N-\tau}$ of $\{G_i\}_{i=1}^N$, arrange $\{x_i\}_{i=1}^{N-\tau}$ so that $\{x_i\}_{i=1}^{N-\tau}$ and $\{y_i^\tau\}_{i=1}^{N-\tau}$ have the same rank order, then we obtain the final surrogate $\{s_i^\tau\}_{i=1}^{N-\tau}$.

REFERENCES

[1] H. D. I. Abarbanel, R. Brown, J. J. Sidorowich and L. S. Tsimring. The analysis of observed chaotic data in physical systems. *Rev. Mod. Phys. 65, 1331* (1993).

[2] A. M. Albano, J. Muench, C. Schwartz, A. I. Mees and P. E. Rapp. Singular value decomposition and Grassberger-Procaccia algorithm. *Phys. Rev. A 38, 3017* (1988).

[3] A. M. Albano, A. Passamante and M. E. Farrell. Using higher-order correlations to define an embedding window. *Physica D 54, 85* (1991).

[4] G. E. P. Box, G. M. Jenkins and G. C. Reinsel. *Time Series Analysis: Forecasting and Control* 3rd Ed.. (Prentice-Hall, 1994).

[5] D. S. Broomhead, and G. P. King. Extracting qualitative dynamics from experimental data. *Physica D 20, 217* (1986).

[6] T. Buzug and G. Pfister. 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 45, 7073* (1992).

[7] M. Casdagli, S. Eubank, J. Farmer, and J. Gibson. State space reconstruction in the presence of noise. *Physica D 51, 52* (1991).

[8] C. J. Cellucci, A. M. Albano, and P. E. Rapp. Comparative study of embedding methods. *Phys. Rev. E* 67, 066210 (2003).

[9] C. Diks. Estimating invariant of noisy attractors. *Phys. Rev. E* 53, 4263 (1996).

[10] K. J. Falconer. *Fractal geometry: Mathematical foundations and applications* (Wiley, New York, 1990).

[11] J. D. Farmer, and J. J. Sidorowich. Predicting chaotic time series. *Phys. Rev. Lett.* 59, 845 (1987).

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

[13] A. Galka. *Topics in Nonlinear Time Series Analysis with Implications for EEG Analysis* (World Scientific, 2000).

[14] P. Grassberger and I. Procaccia. Characterization of strange attractors. *Phys. Rev. Lett.* 50, 346 (1983).

[15] P. Grassberger and I. Procaccia. Measuring the strangeness of the strange attractor. *Physica D 9, 189* (1983).

[16] K. Judd. An improved estimator of dimension and some comments on providing confidence intervals. *Physica D* 56, 216 (1992); Estimating dimension from small samples. *Physica D* 71, 421 (1994).

[17] K. Judd and L. Smith. Indistinguishable states I: Perfect model scenario. *Physica D* 151, 125 (2001).

[18] R. Hegger, H. Kantz, and T. Schreiber. Practical implementation of nonlinear time series methods: The TISEAN package. *Chaos 9, 413* (1999).

[19] A. R. Houghton and D. Gray. *Making Sense of the ECG: A Hands-on Guide.* Oxford University (1997).

[20] H. Kantz and T. Schreiber. *Nonlinear Time Series Analysis* (Cambridge, 1997).

[21] D. T. Kaplan and L. Glass. Direct test for determinism in a time series. *Phys. Rev. Lett.* 68, 427 (1992).

[22] D. Kaplan and L. Glass. *Understanding Nonlinear Dynamics* (Springer, New York, 1995).

[23] G. Kember and A. C. Fowler. A correlation function for choosing time delays in phase portrait reconstructions. *Phys. Lett. A 179, 72* (1993).

[24] W. Liebert and H. Schuster. Proper choice of the time delay for the analysis of chaotic time series. *Phys. Lett. A 143, 107* (1989).

[25] W. Liebert, K. Pawelzik and H. G. Schuster. Optimal embeddings of chaotic attractors from topological considerations. *Europhys. Lett. 14, 521* (1991).

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

[27] X. Luo and M. Small. Geometric measures of redundancy and irrelevance tradeoff exponent to choose suitable time delays for continuous systems (2005) Submitted to *IJBC*, in review

[28] X. Luo, T. Nakamura and M. Small. Surrogate Test to distinguish between chaotic and pseudoperiodic time series. *Phys. Rev. E*, in press (2005)

[29] M. B. Kennel, R. Brown and H. D. I. Abarbanel. Determining embedding dimension for phase-space reconstruction using a geometrical construction. *Phys. Rev. A 45, 3403* (1992).

[30] J. M. Martinerie, A. M. Albano, A. I. Mees and R. E. Rapp. Mutual information, strange attractors, and the optimal estimation of dimension. *Phys. Rev. A 45, 7058* (1992).

[31] A. I. Mees, P. E. Raap, and L. S. Jennings. Singular-value decomposition and embedding dimension. *Phys. Rev. A 36, 340* (1987).

[32] T. Nakamura, X. Luo and M. Small. Topological test for chaotic time series. (2004) Submitted to *Phys. Rev. Lett.*, in review

[33] E. Ott. *Chaos in dynamical systems* 2nd Ed., p18-19 (Cambridge, 2002).

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

[35] M. Paluš and I. Dvořák. Singular-value decomposition in attractor reconstruction: Pitfalls and precautions. *Physica D 55, 221* (1992).

[36] W. H. Press, S. A. Teukolski, W. T. Vetterling and B. P. Flannery. *Numerical recipes in C: The art of scientific computing*, 2nd ed. (Cambridge University Press, Cambridge, 1992).

[37] D. Prichard and J. Theiler. Generalized redundancies for time series analysis. *Physica D 84, 476* (1995).

[38] R. Quian Quiroga, A. Kraskov, T. Kreuz, and P. Grassberger. Performance of different synchronization measures in real data: A case study on electroencephalographic signals. *Phys. Rev. E* 65, 041903 (2002).

[39] P. E. Rapp, A. M. Albano, I. D. Zimmerman and M. A. Jiménez-Montaño. Phase-randomized surrogates can produce spurious identifications of non-random structure. *Phys. Lett. A 192,27* (1994).

[40] M. T. Rosenstein, J. J. Collins and C. J. De Luca. Reconstruction expansion as a geometry-based framework for choosing proper delay times. *Physica D 73, 82* (1994).

[41] T. Sauer. Reconstruction of dynamical systems from interspike intervals. *Phys. Rev. Lett. 72, 3811* (1994).

[42] T. Sauer. Time series prediction by using delay coordinate embedding. In A. S. Weigend and N. A. Gershenfeld, editors, *Time series prediction: Forecasting the future and understanding the past*, Perseus Books, Reading, MA, 1994.

[43] T. Schreiber and A. Schmitz. Improved surrogate data for nonlinearity tests. *Phys. Rev. Lett. 77, 635* (1996)

[44] T. Schreiber and A. Schmitz. Discrimination power of measures for nonlinearity in a time series. *Phys. Rev. E 55, 5443* (1997).

[45] T. Schreiber. Constrained randomization of time series data. *Phys. Rev. Lett. 80, 2105* (1998).

[46] T. Schreiber and A. Schmitz. Surrogate time series. *Physica D 142, 346* (2000).

[47] B. W. Silverman. *Density estimation for statistics and data analysis* (Chapman and Hall, London 1986).

[48] M. Small and K. Judd. Correlation dimension: A pivotal statistic for non-constrained realizations of composite hypothesis in surrogate data analysis. *Physica D 120, 386* (1998).

[49] M. Small, K. Judd, M. Lowe, and S. Stick. Is breathing in infants chaotic? Dimension estimates for respiratory patterns during quiet sleep, *Journal of Applied Physiology 86, 359* (1999).

[50] M. Small, D. J. Yu and R. G. Harrison. A surrogate test for pseudo-periodic time series data. *Phys. Rev. Lett. 87, 188101* (2001).

[51] J. Theiler. Estimating fractal dimension. *J. Opt. Soc. Am. A 7,1055* (1990).

[52] J. Theiler, S. Eubank, A. Longtin, B. Galdrikian and J. D. Farmer. Testing for nonlinearity in time series: the method of surrogate data. *Physica D 58,77* (1992).

[53] J. Theiler. On the evidence for how-dimensional chaos in an epileptic electroencephalogram. *Phys. Lett. A* 196, 335 (1995).

[54] J. Theiler and D. Prichard. Constrainded-realization Monte-Carlo method for hypothesis testing. *Physica D 94, 221* (1996).

[55] J. Theiler and P. E. Rapp. Re-examination of the evidence for low dimensional nonlinear structure in the human electroencephalogram. *Electroenc. Clin. Neurophys. 98, 213* (1996).

[56] J. Theiler and D. Prichard. Using 'Surrogate Data' to calibrate the actual rate of false positives in tests for nonlinearity in time series. *Fields Inst. Commun. 11, 99* (1997).

[57] C. D. Wagner and P. B. Persson. Chaos in the cardiovascular system: An update. *Cardiovasc. Res. 40, 257* (1998).

[58] E. W. Weisstein. *Concise Encyclopedia of Mathematics* (CRC Press LLC., 1999).

[59] H. Whitney. Differentiable manifolds. *Ann. Math. 37, 645* (1936); F. Takens. Detecting Strange Attractors in Turbulence, *Lecture Notes in Math. Vol. 898* (Springer, New York, 1981); R. Mañé. On the dimension of the compact invariant sets of certain non-linear maps, *Lecture Notes in Math. Vol. 898* (Springer, New York, 1981); T. Sauer, J. A. Yorke, and M. C. Casdagli. Embeddology. *J. Stat. Phys., 65, 579* (1991).

[60] D. J. Yu, M. Small, R. G. Harrison, and C. Diks, Efficient implementation of the Gaussian kernel algorithm in estimating invariants and noise level from noisy time series data, *Phys. Rev. E* 61, 3750 (2000)