The Hong Kong Polytechnic University

Department of Electronic and Information Engineering

Modelling Nonlinear Dynamics of Time Series with Applications

Zhao Yi

A thesis submitted in partial fulfilment
of the requirements for the
Degree of Doctor of Philosophy

September 2006

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 or diploma, except where due acknowledgement has been made in the text.

_____ (Signed)

YI ZHAO (Name of student)

ABSTRACT

Nature, as human beings observe it, brings forth rich and colorful phenomena which are recorded as time series. We usually wish to understand the underlying dynamics hidden under these time series. An efficient way of capturing dynamics is by time series modelling. As it became obvious that nonlinear dynamics abounds in the world, nonlinear modelling techniques were greatly studied and developed.

This thesis describes building optimal nonlinear models based on an information theoretic criterion to investigate the underlying dynamics of various time series, especially cardiovascular time series (electrocardiograph and pulse data). The purpose of this research is to determine whether these techniques may be used to extend our understanding of the human cardiovascular system. We wish to build the relationship between the ECG and pulse data, and then classify distinct dynamics from recording of cardiac data.

The major problem endemic to either linear or nonlinear models with a large number of parameters is overfitting. The usual methods of avoiding this problem are to avoid fitting the data too precisely, but these techniques can not determine the exact model size directly. We propose an alternative, information theoretic criterion to determine the optimal models. When applied to time series prediction these optimal models both generalize well and accurately capture the underlying nonlinear dynamics.

The preceding optimal modelling techniques have been employed to model blood pressure propagation from the human wrist to the fingertip. We apply the well-known surrogate data method to model residuals, and conclude that there is no significant difference between pulse waveforms measured on the lateral artery

(wrist) and fingertip. We then investigate the deterministic dynamics of ECG and pulse data, and the relationship between them. The surrogate data method against the null hypothesis of linear noise, does not provide sufficient evidence to confirm the existence of deterministic dynamics in them. We present a recently suggested pseudo-periodic surrogate method to determine whether they are consistent with deterministic chaos. Algorithmic complexity is proposed as the robust test statistic of the surrogate data method. Short-term prediction from ECG to pulse data and vice versa by our optimal models are also described. The results indicate that bounded aperiodic determinism exists in both ECG and pulse data. We conclude that the human ECG data contains information of the human body that pulse data does not have or can not replicate.

The feasibility and utility of complexity and the surrogate data method for identification of nonlinear dynamics in noisy experimental data is also studied in this thesis. To provide an additional independent test system we apply these techniques to the international effort to experimentally observed gravitational waves. We propose complexity to detect the simulated gravitational data contaminated with strong noise. Complexity is proved to be robust and very sensitive to the existence of gravitational waves. The surrogate data method is used to examine the deterministic dynamics in the noisy gravitational data and attach statistical significance to the results estimated by complexity.

ACKNOWLEDGMENTS

First of all, I would like to gratify my supervisors, Dr. Michael Small and Prof. Chi Kong Tse for their excellent supervision, invaluable guidance, and generous support. I gratefully acknowledge Dr. Small's immense instructions and advices during the past three years. I also wish to thank Dr. Small for regularly discussing with me and organizing the group meeting. Without his illumination I would have never realized that nonlinear time series analysis is so fascinating and significant. It has been a wonderful experience working with him. I especially thank Professor Chi Kong Tse for providing us a platform to conveniently discuss and exchanges ideas with researchers outside. Without my supervisors' support I would not have worked on a collaborative project in The University of Western Australia. Their kindness will always be remembered.

I wish to thank Dr. Chunnong Zhao, Dr. Li Ju, Prof. David Blair, Dr. David Coward and Eric Howell in the Gravity Wave Group of The University of Western Australia. I thank Dr. Zhao for giving me a basic grasp of the gravitational detectors and helping me to facilitate the life in Perth. I am grateful for the opportunity to work in the Gravity Wave Group and thank Dr. Ju and Prof. Blair for their trusts and supports. For the simulated gravitational data in Chapter 5 I am indebted to Dr. David Coward and Eric Howell.

I wish to thank Dr. Tomomichi Nakamura, Dr. David Walker, Xiaodong Luo, Jie Zhang, and Junfeng Sun in our research group. Communication and discussion with them broadened my knowledge and inspired many new ideas. In particular, I wish to thank Dr. Nakamura, when asked for helps, for providing abundant materials. I also thank Dr. Walker for many helpful explanation on Kalman filters and Markov chain Monte Carlo. In addition, I would like to extend my gratitude to

the colleagues in the Applied Nonlinear Circuits and Systems Research Group, and many friends I have met in the department and student hostel.

I gratefully acknowledge the Research Committee of The Hong Kong Polytechnic University for the financial support of my study.

Finally, my sincere appreciation goes to my wife and my parents for their persistent support, encouragement, and understanding. I thank them for encouraging this endeavor, for believing that it was worthwhile, and for telling me the truth all through. I am deeply grateful.

PUBLICATIONS

Journal Papers

1. Yi Zhao and Michael Small, "Evidence consistent with deterministic chaos with human cardiac data: surrogate and nonlinear dynamical modeling," *International journal of bifurcation and chaos*, in press.

2. Yi Zhao, Michael Small, David Coward, Eric Howell, Chunnong Zhao, Li Ju and David Blair, "Identifying deterministic signals in simulated gravitational wave data: algorithmic complexity and the surrogate data method," *Classical and Quantum Gravity*, vol. 23, pp. 1801-1814, 2006.

3. Yi Zhao and Michael Small, "Minimum description length criterion for modeling of chaotic attractors with multi-layer perceptron networks," *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications*, vol. 53, no. 3, pp. 722-732, 2006.

4. Yi Zhao, Chunnong Zhao, Li Ju, Michael Small and David Blair, "Telemetry system driven by radiation power for use in gravitational wave detectors," *Review of Scientific Instruments*, vol. 76, no. 8, pp. 084503, 2005.

5. Yi Zhao and Michael Small, "Equivalence between "feeling the pulse" on the human wrist and pulse pressure wave at fingertip," *International Journal of Neural Systems*, vol. 15, no. 4, pp. 1-10, 2005.

Conference Papers

1. Yi Zhao and Michael Small, "Identifying deterministic chaos in human cardiac time series," *The 2006 International Symposium on Nonlinear Theory and its Applications (NOLTA 2006)*, September 2006, Bologna, Italy.

2. Yi Zhao and Michael Small, "Deterministic propagation of blood pressure waveform from human wrists to fingertips," *The Fifth International Conference on Intelligent Data Engineering and Automated Learning (IDEAL 2004)*, August 2004, Exeter, UK (embodied in *Lecture Notes in Computer Science*, vol. 3177, pp. 142-147, 2004).

CONTENTS

LIST OF FIGURES

LIST OF TABLES

1. INTRODUCTION

1.1 Generalities

Nonlinear dynamics describes a system, whose data points are distributed in a geometrical space over a course of time and whose mathematical description is not simply linear. A dynamical system is a pair (M^D, f), where M^D is the manifold of the system and f is an evolution function that maps a point of the phase space back into the phase space (for a nonlinear dynamical system, f is a nonlinear function). A trajectory of the dynamical system is the sequence generated by iteration of f with an initial condition ($x_0 \in M^D$) [30].

When a nonlinear dynamical system is deterministic, the current state is determined only by the previous states. Chaotic dynamics describes the behavior of the certain deterministic nonlinear dynamical system, which appears to be disorderly and is sensitive to initial condition. This behavior, that is unexpected or ignored in linear dynamics, is counterintuitive but very interesting, and therefore attracts much attention not only from mathematics [1] [2] and physics [3], but also from other sciences [4] [5]. One example is the profile of works on nonlinear dynamics in the biomedical literature indexed by PubMed [1] since the first use of "chaos" in its present meaning [29], as illustrated in Fig. 1.1.

In this thesis, we are concerned with modelling nonlinear dynamics of time series. We develop modelling techniques to determine models with adequate generalization for various time series prediction. We then investigate two specific applica-

[1] PubMed is a search service of scientific journals developed by the United States National Library of Medicine (NLM). It provides access to over 16 million citations from MEDLINE (the NLM's premier bibliographic database)and other life science journals.

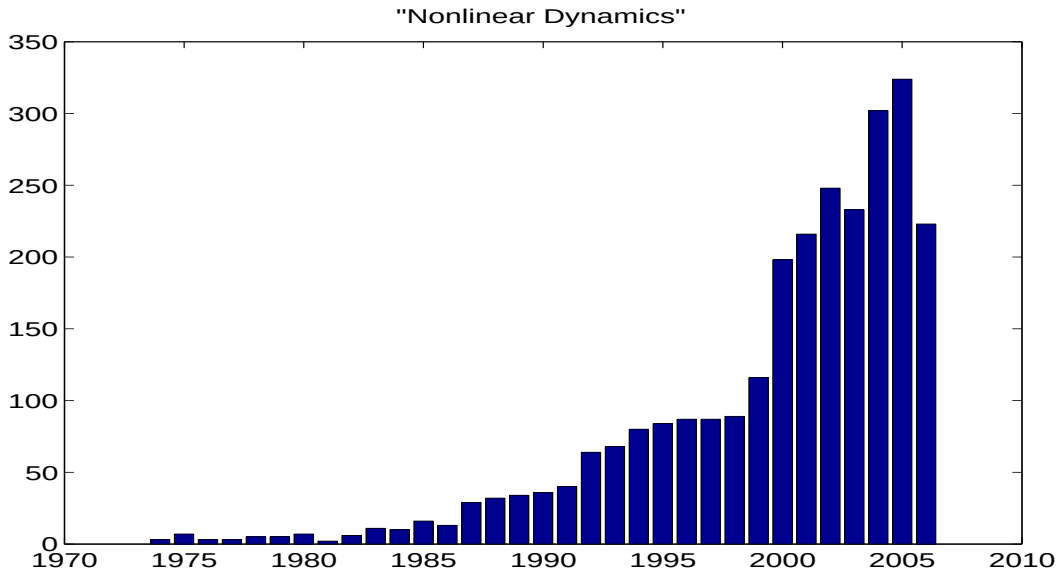


Fig. 1.1: The number of publications over years on applications of nonlinear dynamics. The plot is based on keywords searching of "nonlinear dynamics" (in the title or abstract) in the journals cited by PudMed. The first histogram counts all publications over the period from 1964 to 1974; the last histogram only counts the publications up to July 2006. Note that some of all these publications may refer to "nonlinear dynamics" in different context.

tions: the human cardiac data and simulated gravitational waves. The first kind of time series appears to be dense periodic structures while gravitational waves which are composed of many individual transient bursts appear to be stochastic noise. The two kinds of time series embody two distinct underlying nonlinear dynamics, which we are interested in.

To tackle time series surrogate data hypothesis testing and nonlinear modelling are both useful. However, the aims of hypothesis testing and modelling are different. Modelling will try to capture the nonlinear dynamics hidden in the data and provide information about features that may or may not be in the data: the features one deduces from the model are attributable to that particular description of the data. Surrogate data methods provide a rigorous test of whether or not the data is consistent with the data produced by a particular class of linear (or nonlinear) models. One can then be assured that the data are suitable to that particular class

of models, or not.

In this thesis we propose to combine nonlinear modelling with the surrogate data method so as to carry out the following two tasks: (i) to test model residuals against the hypothesis of random noise so as to confirm if the model accurately captures the specific nonlinear dynamics; (ii) to estimate the possible class of models that may be amenable to such data and exclude possibility of stochastic data to be modelled.

The major results of this thesis concern: (i) generalization of the method of minimum description length to produce optimal models for various time series prediction; (ii) the application of optimal models and surrogate data method to human cardiac data; (iii) the application of the nonlinear surrogate data method as a form of hypothesis testing to human cardiac data; (iv) developments of test statistics (complexity) of existing surrogate data methods to provide robust criteria; (v) the application of complexity and the surrogate data method to detect deterministic gravitational data from noise.

We demonstrate that optimal models determined by the minimum description length can make accurate prediction for four typical time series: the Rössler system and the Ikeda map, both with the addition of dynamic noise, and two experimental recording of the chaotic laser data and electrocardiogram data. All the optimal models can provide adequate generalization and avoid overfitting. We show that the prediction errors of optimal models from pulse on the wrist to fingertip are found to be consistent with random noise, and the pulse signal on the wrist is equivalent to that on the fingertip. We employ the pseudo-periodic surrogate data method demonstrate that the electrocardiograph (ECG) and pulse data of healthy human are inconsistent with periodic data driven by uncorrelated noise, but consistent with the deterministic chaos. We investigate the predictability from ECG to pulse data and vice versa by using our optimal models. Prediction from ECG to pulse data can exactly follow the pulse data while the prediction from pulse to ECG data is poor. We thus conclude that the ECG data contains more deterministic components than the pulse data. In addition , application of the transfer entropy [77] between ECG

and pulse also shows higher information flow from ECG to pulse, which is consistent with the conclusion taken in prediction analysis. We prove that complexity is higher sensitive to the detection of the weak gravitational data compared to popular alternatives, such as the fast Fourier transform algorithm (FFT). The surrogate data method with complexity as the test statistic shows that the noisy data containing weak gravitational waves is less likely random data. In comparison with the chaotic laser data and human normal cardiac data, little periodic structure was found in the gravitational data but a lot of burst signals were revealed. Conventional test statistics, like correlation dimension, do not show good performance in identifying this data, especially when the data is contaminated with strong noise. It is motivated to investigate new techniques to identify the existence of deterministic gravitational waves or reject the possible random noise. We demonstrate that complexity and the surrogate data method are good options.

1.2 Literature Review

The literature review is organized under several categories according to the application of nonlinear dynamics techniques to various time series. Each category will be presented in each following chapter. We aim to provide a clear historical overview of how the research in this area have been developed, followed by our contributions. This section is to serve as an introduction to the detailed literature review and contributions discussed in the remaining chapters. Here we concentrate only on the main points.

The concept of nonlinear dynamics becomes popular in recent decades. However, modelling nonlinear dynamics is still challenging since overfitting is a serious problem in building either linear or nonlinear models. Overfitting means that a model may fit training data very well, but fail to generalize on novel data from the same dynamics. A number of works have discussed methods for improving generalization capability or avoiding overfitting. However, most studied improving the parameters of the fixed model [6] [7] [10] [37] but did not directly estimate the opti-

mal model size for one specific application. Our approach is to develop an general information theoretic method to determine the optimal model size. This method is motivated from minimum description length [11] [13]. Modifications to facilitate the calculation of this methods have been proposed [16]. However, few works applied this method to model selection of neural networks for time series prediction [17]. Our work focuses on this subject and generalize the previous work in [17].

We then use the optimal modelling technique to study human pulse signals, which are widely used to examine unhealthiness in tradition Chinese medicine (TCM) [61]. We wish to model the relationship between pulses measured on two places in order to examine whether there is any significant difference between these two measurements. The surrogate data method [42] is used to test model residuals and verify if a built model is able to predict well. Few works have been done to address this problem since pulse diagnosis, especially feeling the pulse in traditional Chinese medicine has not yet been accepted by the modern medicine.

However, there are a lot of attempts to investigate the possibility of chaotic cardiac data. Some works revealed that the ECG has a finite noninteger correlation dimension and positive Lyapunov exponent [52] [53] [54]. Notably, hypothesis testing for the presence of chaos in ECG data has been proposed recently [55] [57]. But most results are based on the surrogate data method to test for linear noise. Actually it has been found that the normal ECG and pulse data usually exhibit strong periodicities, which definitely are not linear noise. Also, few attempts to determine the presence of chaos in pulse data are found in the literature. In this thesis a pseudo-periodic surrogate data method [47] [62] was used to generally test both complete ECG and pulse data against the hypothesis of a periodic orbit with uncorrelated noise.

Correlation dimension estimations are not applicable to field measurement [55] [63] but algorithmic complexity [74] is proved to be robust to the noisy signals, and sensitive to the intrinsic deterministic dynamics [67]. We, therefore, employ complexity as the test statistic of the surrogate data method. Modelling the relationship (i.e. prediction analysis) between pulse measurements on the human wrist and fin-

gertip was originally proposed by us [68] [70]. Obviously, few nonlinear dynamical differences between those signals should be expected and then conclusion of this work is straightforward. Therefore, it is more significant to extend our modelling techniques to model relationship between ECG and pulse data, which record distinct information of the human heart. No similar work has been found in the literature. We will investigate the predictability between ECG and pulse data and deterministic dynamics in them. Transfer entropy, which identified higher information flow from the human cardiac system to respiratory system than vice versa [77], is thus applied to quantify the information flow between the ECG and pulse data of the human cardiac system. We have found that result of transfer entropy is consistent with the result of prediction analysis.

In contrast to cardiac time series, the cumulative gravitational wave is composed of individual transient events [98] and distinct from periodic and pseudo-periodic waveforms. The collapse of massive objects can change the spacetime around and then these spacetime changes propagate as gravitational waves (GW). The cumulative gravitational wave from transient GW sources at cosmological distances is commonly described as a stochastic background because of the temporal randomness of the individual events. Moreover, the gravitational waves is too tiny to readily detect from background noise. Several long-baseline laser interferometer GW detectors have been, or are nearly, constructed, which are expected to provide a practical detection rate [87]–[91]. Techniques developed from the domain of signal processing have been used to detect GW bursts from noise [99] [101] [103], including a set of improved filter techniques [104] [105]. Here we propose the techniques developed from the field of nonlinear time series analysis to detect GW signals in presence of Gaussian white or coloured noise. Complexity is independent of characteristics of the signal, and its calculation does not require prior knowledge of the signal. We also find that its signal noise ratio (SNR) performance is better than spectral estimation alone. The surrogate data method is employed to determine with a certain probability whether a particular value of complexity indicates the presence of deterministic dynamics. This technique attaches statistical significance to the results

of complexity. Complexity, the surrogate data method, and their combination are demonstrated to be potentially available methods to detect the gravitational wave.

1.3 Outline of the Thesis

Main contents of this thesis are organized into four chapters, which present the areas of our investigation.

Specifically, Chapter 2 presents a general information theoretic method to determine optimal models for prediction of four typical time series. All the optimal models estimated by this information theoretic method can capture the nonlinear dynamics of each time series and make the accurate prediction. We discuss main information theoretic methods and outline developments of our minimum description length method in comparison with the previous works.

Chapter 3 investigates the relationship between measurements of pulse on the wrist and fingertip using the surrogate data method the optimal models estimated by the preceding minimum description length. Some physiological aspects of this application are also discussed.

Chapter 4 presents a thorough investigation into the problem of deterministic chaos in the human cardiac data. A new surrogate data method is used to provide the evidence that the human cardiac data is consistent with deterministic chaos. We then investigate the relationship between the ECG and pulse data. Difference between ECG and pulse data, and the medical consideration are discussed.

Chapter 5 presents application of techniques (complexity and the surrogate data method) developed from the domain of nonlinear time series analysis to identify gravitational waves contaminated with strong Gaussian noise.

Finally, the last chapter (Chapter 6) presents conclusions of the preceding four chapters, and underlines the primary contribution of this thesis. We also outline plans for the further research.

2. DETERMINATION OF OPTIMAL NEURAL NETWORKS FOR TIME SERIES PREDICTION

2.1 Introduction

In this chapter, we propose an information theoretic criterion, minimum description length (MDL) for model selection in order to build optimal models for time series prediction, and compare this algorithm with others. We found that the model which minimizes the description length (DL) of models both generalizes well and accurately capture the underlying nonlinear dynamics.

The biological motivation for artificial neural networks is a massive highly connected array of nonlinear excitatory "neurons". It is natural to build artificial neural networks with a fairly large number of neurons. This creates a statistically ill-posed problem as the number of model parameters can easily exceed the number of available data. Therefore, it is very easy for the resultant model to overfit. However, the critical issue in developing a neural network is the generalizability of that network: the network should make small prediction errors on novel inputs. Elementary discussion of overfitting was represented in [6]. Geman, Bienenstock, and Doursat illustrated a rigorous approach based on the bias/variance trade-off to improve network training. That is, underfitting produces excessive bias in the outputs, whereas overfitting produces excessive variance [7]. The most common methods to avoid overfitting are: (i) early stopping [37]; and (ii) statistical regularization techniques, such as weights decaying [9] and Bayesian learning [10]. For early stopping, there is one limitation: we must be careful not to choose a training algorithm that converges too rapidly. With this method, the choice of the validation set is also difficult. The validation set should be representative of all points in the training set. The weights

decaying method involves modifying the performance function to $msereg = \mu\ mse + (1-\mu)\ msw$, where mse is the mean sum of squares of the network errors, msw is the mean sum of squares of the network weights and biases, and μ is the performance ratio.

But the problem with this method is that it is difficult to select the optimum value for μ. If μ is too large, the networks may overfit; if μ is too small, the network will not adequately fit the training set. For Bayesian learning the optimal regularization parameter, μ is determined in an automated fashion [10]. Another feature of Bayesian learning is that it provides a measure of how many network parameters are effectively used by the network, i.e. how many parameters are not required or have little significance to the networks. Hence, using this method we usually build a larger network with many "wasteful" parameters. The smaller optimal network cannot be determined by this method.

A common feature of all these methods is that they optimize parameters of the known neural network but they cannot determine the optimal number of neurons (model size) in the network for a specific application. We propose an alternative approach, which estimates exactly how many neurons of the feedforward multi-layer neural network are required for the specific time series prediction. The criterion is an approximation to the minimum description length (MDL) [11]. The MDL principle rooted in the theory of algorithmic complexity [12] and mainly developed by J. Rissanen with a series of papers starting with [13], who proposed the issue of model selection as a problem in data compression. Several modifications to the MDL for improving its performance have been proposed [14] [15] and [16] [17].

There are other typical methods applicable to model selection. Akaike proposed an information criterion (AIC) based on a weighted function of the fit of a maximum log-likelihood model and the number of independent parameters adjusted to maximize the likelihood [18]. The motivation of AIC and its assumptions was the subject of some discussion in [13]–[19]. From a practical point of view, the AIC tends to overfit the data [20], [21]. Wallace et al developed the minimum message length (MML) [22]. As in the MDL, MML chooses the hypothesis minimizing the

code-length of the data but the codes are quite different from those in MDL. The Bayesian information criterion (BIC), also known as Schwarz's information criterion (SIC) is equivalent to the MDL criterion [23] [19] [11]. Recently Xu developed the Bayesian Ying Yang (BYY) information criterion for model selection [24] – [26].

The MDL criterion is related to other well-known model selection criteria. The AIC and BIC perform best for linear models; for nonlinear models description length style information criterion is better. Our choice of the MDL in particular is based on the fact that it is robust and relatively easy to estimate. The MDL has the great advantage of relatively small computational costs.

There have been relatively few attempts at applying the minimum description length principle to neural networks model selection for time series prediction. Judd and Mees have successfully applied the minimum description length principle to the selection of optimal radial basis function networks [16]. More recently, Small and Tse employed this method to yield a model fitting algorithm to neural network architectures [17]. Our current work is to develop the method of minimum description length to estimate optimal models among all the standard neural networks candidates. It is a generalization of [17] in the three major areas:

1. Each neural network in [17] is built by the expansion of the previous (smaller) model, i.e. the larger new neural network is formed by adding one new candidate radial basis function to the old model. In the current work all the neural networks are built independently, which guarantees that there is not any connection among them. Although it is never possible to find the global MDL, by allowing models of different sizes to be built independently of one another, MDL is achieved in a broader search of parameter space. In [17] models are built by successively expanding the previous (smaller) model: in doing so, one searches a much more restricted area of parameter space.

2. In this work we count the contribution to the description length of not only linear parameters of the neural network, but also nonlinear parameters of it. The penalty for model parameters, $M(k)$, then involves a variable representing

the effective number of parameters within one neuron. However, in [17] the authors ignore the nonlinear parameters in a neuron, and hence $M(k)$ is only a rough approximation.

3. The specific training algorithm used in [17] is borrowed from statistical approximation theory and restricted to those models (there is no flexibility in choosing fitting techniques). Therefore, it is not possible to systematically compare the model performance to standard neural network techniques. Networks in the current research can be implemented with the existing training algorithms, such as basic gradient descent, conjugate gradient algorithm, and Levenberg-Marquardt algorithm. The MDL criterion is applicable to select the optimal model size of neural networks trained by the existing training algorithms. So it provides a general methodology for model selection. In other words the current MDL criterion is independent of training algorithms and robust.

As its title implies, this thesis is involved in modelling approaches (i.e. models). However, overfitting has long been recognized as a problem endemic in modelling. Therefore, before modelling time series we have to know how to select the optimal models avoiding overfitting and verify these models. This is the target of this chapter. In the following we then continue employing the selected optimal models to study dynamics of human cardiac data that we pay more attention to. In a word, selection of optimal models is the groundwork of the overall research program.

In Section 2.2 we review the minimum description length criterion and modelling algorithms we employ. Section 2.3 demonstrates a systematic study of the performance of this method with both computational and experimental time series, the Rössler system, Ikeda map, chaotic laser data, and human electrocardiograph data.

2.2 The Minimum Description Length Criterion

If an observed data is independent and identically distributed random numbers then there is no good predictive model, and the most compact description of this data is simply to describe the observations themselves. If an observed data is (for example) a realization of some autoregressive process, then the best (most compact) description of this data is to describe the initial few observations, the autoregressive structure, and the further stochastic perturbations (that is, the model prediction errors). Finally, if the observed data represent a deterministic system (the chaotic Ikeda map for example) then the most compact description of this data is to describe the initial values, the parameters of the Ikeda map, and small corrections. Note that for chaotic systems and certainly most real world data the perfect model can not make perfect prediction, and therefore we must admit the necessity of model prediction errors.

2.2.1 Backpropagation Neural Networks

A neural network is a type of predictive model that is particularly good at describing the behavior of complex systems with an acceptable accuracy. They can help with a wide range of problems. Among a large number of potential modelling regimes a neural network model was found to perform best [27]. For practical application of neural networks, most neural networks adopt the backpropagation algorithm and its variants. The backpropagation algorithm is the major algorithm for training feedforward networks and embody the pith of neural networks.

The backpropagation algorithm is proposed by generalizing the Widrow-Hoff learning rule [28] to multiple-layer networks and nonlinear differentiable transfer functions. Input vectors and the corresponding output vectors are paired to train a network until a pre-defined function is best approximated. Two-layer feedforward networks with a hidden layer of sigmoid neurons, and a linear output layer are capable of approximating any nonlinear continuous function with high exactness, provided that the hidden layer contains sufficient neurons.

We thus restrict our interest to multi-layer backpropagation networks with a single hidden layer of sigmoid neurons followed a linear output. For inputs $(x_{t-1}, x_{t-2}, ..., x_{t-d})$ this neural network f can be mathematically expressed as

$$f(x_{t-1}, x_{t-2}, ..., x_{t-d}) = b_0 + \sum_{i=1}^{k} \upsilon_i \phi \left(\sum_{j=1}^{d} \omega_{i,j} x_{t-j} + b_i \right) \tag{2.1}$$

where υ_i, ω_j, $b_i \in R$, k is the model size (i.e. the number of sigmoid neurons) and d is the number of inputs.

For multi-layer networks, ϕ is usually selected to be a bounded monotonically increasing function. We choose the tan-sigmoid transfer function $\phi(x) = (1 - e^{2x})/((1 + e^{2x})$. The typical construction of a two-layer feedforward neural network is illustrated as follows in Fig. 2.1 .

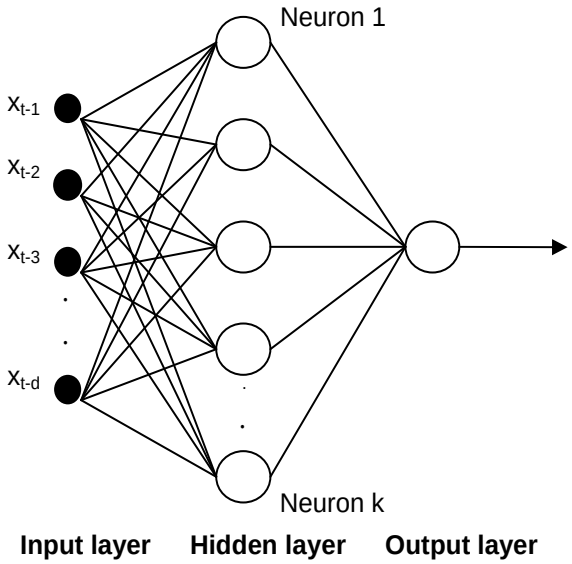


Fig. 2.1: The basic schematic of the two-layer feedforward neural network. Selection of the optimal neural network is the selection of the number of neurons in our experiments.

Typically, a neural network will consist of a very large number of nonlinear "neurons". It is very easy to build a model with small in-sample prediction error, but it is much more difficult to build a model which generalizes well. Hence, the

problem is how to determine the good model for specific applications, i.e. estimating the sufficient number of neurons. To address this problem, we take the method of minimum description length to determine directly the optimal model size of the feedforward multi-layer network for time series prediction.

2.2.2 To Calculate Minimum Description Length

The basic principle of minimum description length is to estimate both the cost of specifying the model parameters and the associated model prediction errors.

Let z be data of the model; Λ represents the vector of all the model parameters. In [16] the description length of the data with respect to a particular model Λ, is expressed as

$$L(z, \Lambda) = L(z|\Lambda) + L(\Lambda) \tag{2.2}$$

$L(z|\Lambda)$ represents the description length of the model prediction error; $L(\Lambda)$ is the description length of the model parameters. Let $M(k)$ and $E(k)$ be the cost of describing the model parameters and the cost of describing the model prediction error. The description length of the data with respect to this model is then given by the sum[1]

$$D(k) = M(k) + E(k) \tag{2.3}$$

Intuitively, typical behaviors of $E(k)$ and $M(k)$ are that if the model size k increases $M(k)$ increases and $E(k)$ decreases. However, the penalty for introducing more parameters finally outweighs their contribution to the description of the data. So the optimal models should balance the model error against the model size so as to prevent potential overfitting or underfitting. The MDL principle states that the optimal model is the one that for which $D(k)$ is minimal. Fig. 2.2 plots the expected behavior of description length as a function of model size k.

Let $\mathbf{y} = \{y_i\}_{i=1}^{N}$ be a time series of N measurements; let $f(y_{i-1}, y_{i-2}, ..., y_{i-d}; \Lambda_k)$ $(i = d+1, ..., N)$ be a scalar function of d variables that is completely described

[1] Calculation of $M(k)$ and $E(k)$ depends on the specific encoding selected for the model and for rational numbers. We use the optimal encoding of floating point numbers described by Rissanen [11].

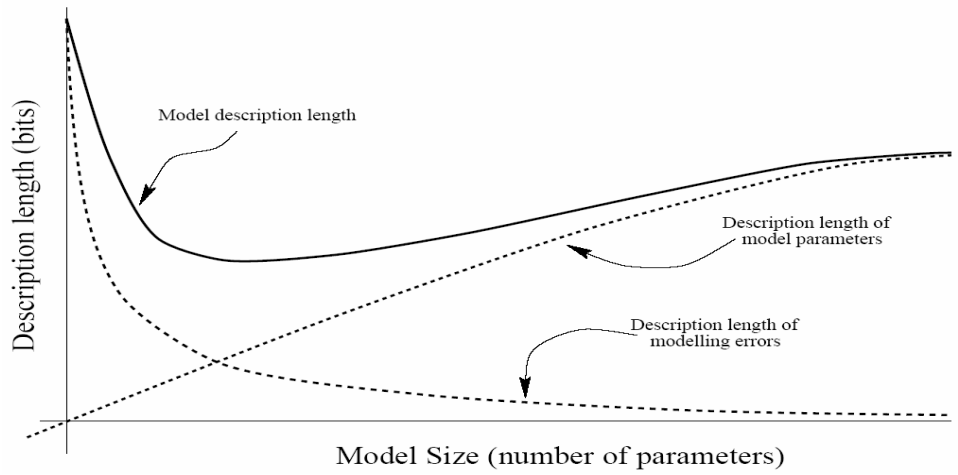


Fig. 2.2: As the model size increases the description length of modelling parameters increases while the description length of modelling error decreases. However, much larger models contribute little to the decreasing of modelling error (i.e. the decreasing description length of modelling error). Finally there should be a minimal point of model description length, which estimates the optimal model. (Figure courtesy of [29])

by the k parameters $\Lambda_k = (\lambda_1, \lambda_2, ..., \lambda_k)$. Define the prediction error e_i by $e_i = f(y_{i-1}, y_{i-2}, ..., y_{i-d}; \Lambda_k) - y_i$. For any Λ_k the description length of the model $f(.; \Lambda_k)$ is given by the description length of the k parameters Λ_k [17]

$$M(k) = L(\Lambda_k) = \sum_{i=1}^{k} \ln \frac{\gamma}{\delta_i} \tag{2.4}$$

where γ is a constant and represents the number of bits required in the exponent of the floating point representation. Typically, $1 \leq \gamma \leq 32$, and $\gamma = 32$ is more than adequate for nearly all purposes, and smaller values can be chosen if desired [16]. In Eq. 2.4, δ_i is interpreted as the optimal precision of the parameters $\lambda_i (i = 1, ..., k)$ and $(\delta_0, \delta_1, ..., \delta_k)$ are defined as the solution of

$$\left(Q \begin{bmatrix} \delta_0 \\ \delta_1 \\ \vdots \\ \delta_k \end{bmatrix} \right)_i = \frac{1}{\delta_i} \tag{2.5}$$

where

$$Q = D_{\Lambda_k \Lambda_k} E(k) \tag{2.6}$$

Q is the second derivative of $E(k)$ with respect to the model parameters Λ_k, and $(\cdot)_i$ denotes the ith element of the vector $(\cdot)$ [16].

$E(k)$ is the negative logarithm of the likelihood of the errors $\mathbf{e} = \{e_i\}_{i=1}^{N-d}$ under the assumed distribution of those errors

$$E(k) = -\ln \ Prob(\mathbf{e}|\Lambda_k) \tag{2.7}$$

where $Prob(\mathbf{e}|\Lambda_k)$ is the probability density function of $\mathbf{e}$ given Λ_k.

For the general case of an unknown distribution of errors, the situation is rather complicated. However, assuming that the model errors are Gaussian distributed with mean zero and standard deviation, $\sigma^2 = \sum_{i=1}^{N-d} \frac{e_i^2}{N-d}$, we obtain

$$E(k) = \frac{N-d}{2} + \ln(\frac{2\pi}{N-d})^{(N-d)/2} + \ln(\mathbf{e}^T\mathbf{e})^{(N-d)/2} \tag{2.8}$$

In principle we may calculate description length by solving Eq. 2.5 to yield the precision δ_j, substituting it to Eq. 2.4 and 2.7 to calculate the description length of the model $M(k)$, and the description length of the model prediction error $E(k)$. Note that for large k a computational bottleneck results from ensuring that the matrix Eq. 2.6 yields a solution to Eq. 2.5 [17]. Substituting Eq. 2.8 to Eq. 2.6, we obtain

$$\begin{aligned}
Q &= D_{\Lambda_k \Lambda_k} \left(\tfrac{N-d}{2} + \ln \left(\tfrac{2\pi}{N-d} \right)^{(N-d)/2} + \ln \left(\sum_{i=1}^{N-d} e_i^2 \right)^{(N-d)/2} \right) \\
&= D_{\Lambda_k \Lambda_k} \left(\ln \left(\sum_{i=1}^{N-d} e_i^2 \right)^{(N-d)/2} \right) \\
&= D_{\Lambda_k \Lambda_k} \left(\tfrac{N-d}{2} \ln(\mathbf{e}^T\mathbf{e}) \right) \\
&= \tfrac{N-d}{2} D_{\Lambda_k \Lambda_k} (\ln((f(\mathbf{y};\Lambda_k) - \mathbf{y})^T (f(\mathbf{y};\Lambda_k) - \mathbf{y})))
\end{aligned}$$

The above equation reveals that if $f(\mathbf{y};\Lambda_k)$ is a linear function, the computation of Q is straightforward [16], but when it is nonlinear the computation of Q becomes considerably more complicated and the solution of Eq. 2.5 becomes substantially more difficult. This full nonlinear approximation to description length was realizable

(for radial basis models), but it was also rather slow, and provided marginal benefit [31].

The model parameters of the neural network described by Eq. 2.1, Λ_k, are $\{b_0, b_i, \upsilon_i, \omega_{i,j} | i = 1, ..., kj = 1, ..., d\}$. Of these parameters, the weights υ_i and the offset b_0 are all linear, the remaining parameters $\omega_{i,j}$ and $b_i (i = 1, ..., k; j = 1, ..., d)$ are nonlinear. The tan-sigmoid transfer function, $\phi(x)$, is approximately linear in the region of interest, so we suppose that the precision of the nonlinear parameters is similar to that of the linear one in this region. We, therefore, employ the linear parameters, the weights υ_i, to calculate the precision of the model δ_j, which makes it computationally feasible to solve Eq. 2.5 and 2.6.

One may usually expect that not all the parameters are significant for a particular neuron and some parameters are effectively "irrelevant" to the calculation of description length. To account for the contribution of all linear and nonlinear parameters to $M(k)$ we define $n_p(i)$ as the effective number of parameters of the ith neuron contributed to the description length of the neural network. In previous research for the ith neuron $n_p(i)$ is just equal to one although this neuron contains one linear parameter and a few of nonlinear parameters [17]. They merely computed the contribution of the linear parameter to $M(k)$.

Thus, we have

$$M(k) = \sum_{i=1}^{k} n_p(i) \ln \frac{\gamma}{\delta_i} \tag{2.9}$$

where δ_i is the relative precision of the weight $\upsilon_i (i = 1, ..., k)$.

In general, $n_p(i)$ will be variable with respect to different neurons but in order to make the problem tractable we make one further approximation. We consider that $n_p(i)$ is fixed for all i and then replace $n_p(i)$ with n_p. So for Eq. 2.9 there exists one parameter, n_p, which makes the following approximation,

$$n_p \sum_{i=1}^{k} \ln \frac{\gamma}{\delta_i} \approx \sum_{i=1}^{k} n_p(i) \ln \frac{\gamma}{\delta_i} \tag{2.10}$$

However, the exact value of n_p is difficult to estimate and we, therefore, approximate n_p with $\widehat{n_p}$ using the embedding dimension calculated by false nearest

neighbors (FNN) [32].

We, therefore, obtain

$$M(k) \approx \widehat{n_p} \sum_{i=1}^{k} \ln \frac{\gamma}{\delta_i} \tag{2.11}$$

In Eq. 2.4, we take $\widehat{n_p} = 1$, but for neural networks used in Section 2.2.1 $\widehat{n_p} \in (1, d+2)$. $d+2$ is the total number of both linear and nonlinear parameters in one neuron. The motivation for this approximation is described in the following section.

This approximation is necessary to make the calculation of description length computationally feasible. It is difficult to analytically evaluate the accuracy of the approximation. Probably it is not possible to evaluate the accuracy of this approximation in general. For specific cases, the accuracy of this approximation can be quantified indirectly by evaluating the performance of the resultant models.

For nonlinear optimization problems (such as fitting a nonlinear model to time series data), it is not possible to guarantee a global optimal solution without performing an exhaustive search. For a continuous (and in-fact fairly high-dimensional) parameter space, one can only ever achieve a local model. We cannot consider that the models estimated are necessarily the best especially comparing with other more complicated models, but we can conclude that this technique can estimate the optimal one from the available neural networks, and provide improved modelling results.

2.2.3 False Nearest Neighbors

For a deterministic system, we can establish a phase space for the system such that specifying a point in this space specifies the state of the system, and vice versa. This is the problem of phase space reconstruction [33] which is technically obtained by the delay time and embedding dimension. Vectors in the embedding space are formed from time delayed values of the scalar measurements $y(n)$ [32]: $y(n) = (x(n), x(n+\tau), ..., x(n+(m-1)\tau)$, where m is called the embedding dimension, and τ is the delay time. Fig. 2.3 describes the phase space reconstruction of a time series. Note that for the purpose of modelling the lag of the input vector is equal to one and the number of input vectors m is approximately the length of one orbit

(i.e. pseudo period) in our experiments.

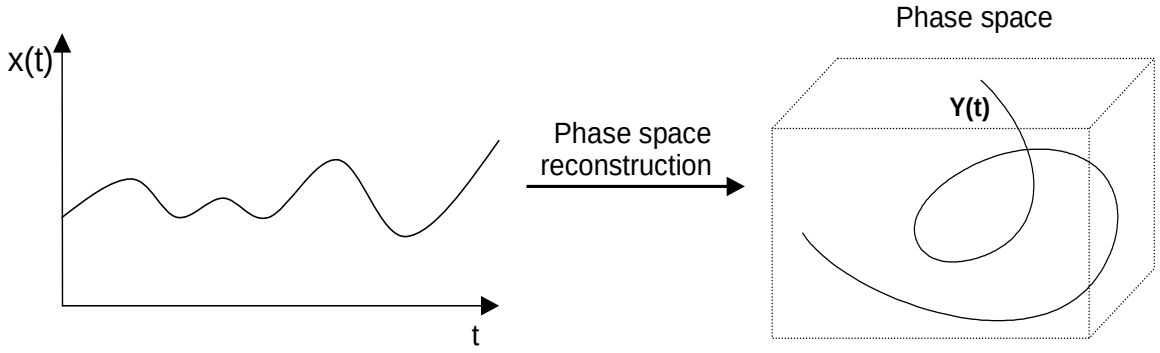


Fig. 2.3: An observed time series, $x(t)$, is reconstructed to $Y(t)$ in the phase space.

Embedding dimension represents the dimension of the phase space required to specify a point in that phase space, i.e., it represents the effective number of inputs for a model. According to Eq. 2.1 the effective number of parameters is closely relative to the effective number of inputs. So the main point to consider here is that we try to determine how many coordinates of a high dimensional embedding space provide significant useful information.

FNN is widely used to determine the number of degrees of freedom required to unambiguously unfold the dynamics, i.e. the FNN embedding dimension ensures that trajectories in the phase space do not cross. Conversely, an embedding dimension larger than that suggested by FNN introduces additional redundancy: that is, trajectories that should be close are also sparsely distributed. When we consider the model building problem it has the same requirement that the parametric representation of coordinates in phase space should be sufficient to distinguish between points that represent disparate trajectories, but should still allow similar trajectories to be close to one another.

Good representation of dynamics depends crucially on appropriate embedding and reconstruction: one simply cannot build a good model without considering the embedding dimension at the same time [34]. Alternatively, one found that the best embedding for the purpose of modelling depends on the construction of the model [35]. So using FNN to choose $\widehat{n_p}$ is equivalent to making the correct choice of embedding reconstruction parameters of the modelling problem, i.e. the minimal

embedding dimension represents the maximal effective number of model parameters which is defined as $\widehat{n_p}$ in Section 2.2.1.

The idea of the false nearest neighbors algorithm as follows [32]: each point $\overrightarrow{R_i} = (x(n), x(n+\tau), ..., x(n+(m-1)\tau))$ in the time series looks for its rth nearest neighbor $\overrightarrow{R_j}$ in a m-dimensional space. Calculate their distance $R_m^2(n,r) = \sum_{k=0}^{m-1} [x_i(n+k\tau) - x_j(n+k\tau)]^2$. When embedding dimension increases from m to $m+1$, one calculates the distance between $\overrightarrow{R_i}$ and the same rth nearest neighbors $R_{m+1}^2(n,r) = R_m^2(n,r) + [x_i(n+m\tau) - x_j(n+m\tau)]^2$. The criterion for designating a neighbor as false is that $\left(\frac{R_{m+1}^2(n,r) - R_m^2(n,r)}{R_m^2(n,r)}\right)^{1/2} = \frac{|x_i(n+m\tau) - x_j(n+m\tau)|}{R_m(n,r)} > R_{tol}$, where R_{tol} is a given heuristic threshold.

For $R_{tol} = 5$ the false nearest neighbors can be clearly identified in our application. The determination of which neighbors are true and which are false is usually insensitive to the threshold one chooses once the number of data is sufficient to nicely populate the attractor [36].

According to the usual criterion, one decides how many points can be counted as false nearest neighbors. The percentage of false nearest neighbors decreases with increasing embedding dimension. If one applies FNN to clean data from a chaotic system, one usually expects that the percentage of false nearest neighbors will drop from nearly 100% in dimension one to strictly zero when m is reached. Furthermore, it will remain zero from then on, since once the attractor is unfolded, it remains unfolded [36]. If the signal is contaminated by noise, the noise may be sufficient to always produce some false nearest neighbors. So the percentage of false nearest neighbors will not reach zero but remain approximately stable after the minimal embedding dimension. Fig. 2.4 presents the process to select the embedding dimension for the four kinds of time series used in this chapter.

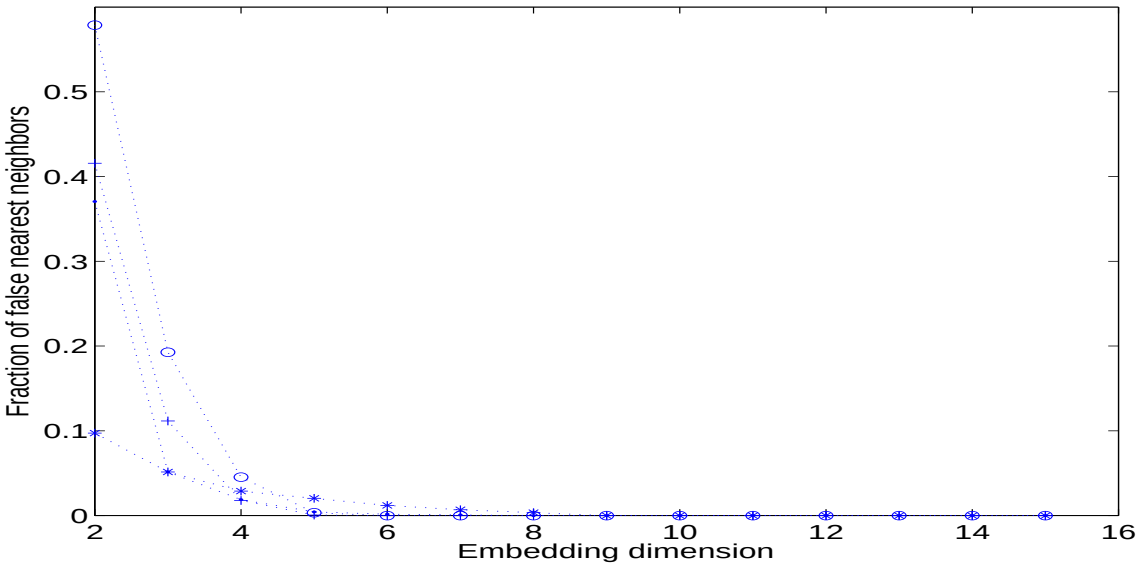


Fig. 2.4: The fraction of false nearest neighbors, for the Rössler sytem $(-+-)$, the Ikeda map $(-\circ-)$, the laser data $(-*-)$, and the ECG data $(-\cdot-)$, decreases with the increasing embedding dimension and finally reaches zero.

2.2.4 Nonlinear Curvefitting

Researchers want to represent empirical data by using a model based on mathematical equations. With the correct model, they can determine important characteristics of the data. The models fitting the data depend on adjustable parameters. The goal of curve fitting is to find the model parameter values so that the built model can fit the data very well. To perform fitting, we need to define a function (such as least-square error) that measures the errors between the data and the fitted data generated by the model. This function is then minimized to the smallest possible value with respect to the model parameters. The parameters that minimize the errors are the best-fitting parameters. In general, the measured systems are nonlinear. Nonlinear curvefitting also seeks for those parameters that minimize the deviations between the observed values and the fitted values. In nonlinear models, however, various iteration to estimate the parameters are necessary. Given an initial guess of model parameter values x, the method tries to find coefficients x that "best-fit" the

observed output data according to the following formula,

$$\min_x \frac{1}{2} |F(x, xdata) - ydata|^2 = \frac{1}{2} \sum_i (F(x, xdata_i) - ydata)^2 \tag{2.12}$$

where $xdata$ is the input data, $ydata$ is the observed output data, and $F(x, xdata)$ is a defined vector function.

For our experiments, all DL curves get a minimal point, i.e. the minimum description length, but there is a problem that these DL curves fluctuate dramatically and it is very likely to affect the estimation of the true minimal point. That is because that the different neural networks are constructed independently in the current work and the perturbation in structural parameters (e.g., associated with estimation noise) of each network is not related to that of successive networks. This correspondingly results in fluctuation of DL curves but the potential changing tendency of DL curves still exists. The curve fitting procedure is intended to smooth noise and provide a more accurate estimation of the actual minimum. Our idea to apply nonlinear fitting can be explained in the following three steps:

1. Define a function, which takes a value k (the number of hidden neurons) and parameter vector to estimate description length. In Eq. 2.8 since N is the number of measurements the first and second terms are constant, and the form of $E(k)$ is consistent with that of a decreasing exponential function. So we use $ae^{-bk}(a > 0, b > 0)$ to describe $E(k)$. Referring to Eq. 2.11, n_p and γ are constant; δ_i is the relative precision of the weights of the ith neuron, and for different neurons it is close to a constant. So $M(k)$ can be regarded appropriately as the linear function about the number of neurons k. Thus define $ck + d$ to describe $M(k)$. Actually d is also required to compensate for an arbitrary constant which is missing from the computation of description length in Eq. 2.9 and 2.11. So the defined function is

$$F(a, b, c, d; k) = ae^{-bk} + ck + d \tag{2.13}$$

where k is the number of hidden neurons; a, b, c, and d are the required coefficients. The above function is the empirical approximation of the true tendency of DL curves.

2. According to the least-square sense, $\min_{a,b,c,d} \sum_{i=1}^{k} (F(a,b,c,d;i) - D(i))^2$, we solve Eq. 2.13 and obtain the vector (a,b,c,d). Note that $D(i)$ is the original description length of neural networks.

3. Substitute a, b, c, and d, which are achieved in Step 2, and then determine the solution of k which make Eq. 2.13 minimal.

The fitted DL curve provides a smooth estimation of the original curve and aims to catch the true minimum. When expending the additional effort of nonlinear curvefitting we find that the estimate of k is much more robust, especially for complicated experimental data.

2.3 Nonlinear Modelling via Minimum Description Length

We present the application of our minimum description length method in conjunction with nonlinear curvefitting to two test systems, the Rössler system (Section 2.3.1) and the Ikeda map (Section 2.3.2), both with the addition of dynamic noise, and two experimental recording of the chaotic laser data (Section 2.3.4) [37] and electrocardiogram data (Section 2.3.5) [38].

2.3.1 The Rössler system

The first computational simulation is a reconstruction of the Rössler system with dynamic noise. The equations of the Rössler system are [39]:

$$\begin{cases} x(t) = -y(t) - z(t) \\ y(t) = x(t) + a * y(t) \\ z(t) = b + z(t) * [x(t) - c] \end{cases} \tag{2.14}$$

with parameters a=0.15, b=0.20, c=10.00, and we take the sampling time, t_s, to be 0.5. The x-component data is converted into three vectors, $x(t)$, $x(t+3)$, and $x(t+5)$ to reconstruct the phase space, as shown in Fig. 2.5 for reference.

By dynamic noise we mean that the Gaussian noise is added to the x-component data prior to prediction of the succeeding state at the interval of 0.5s, i.e., we

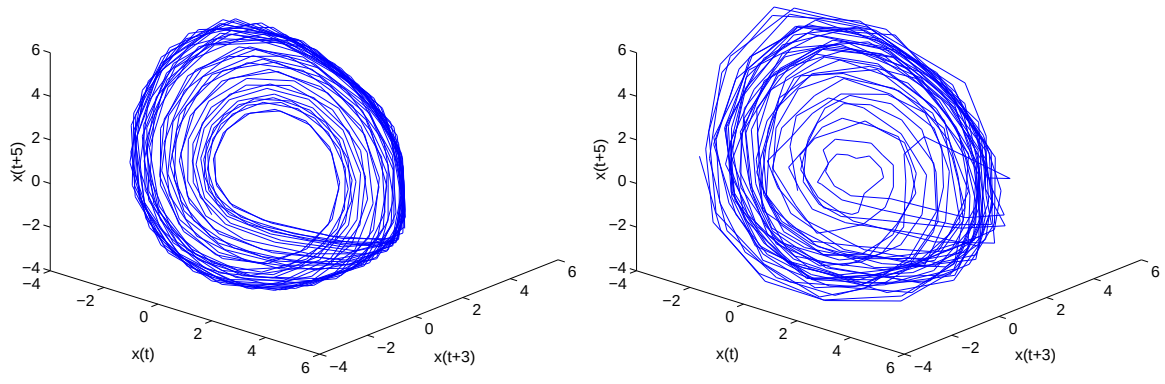
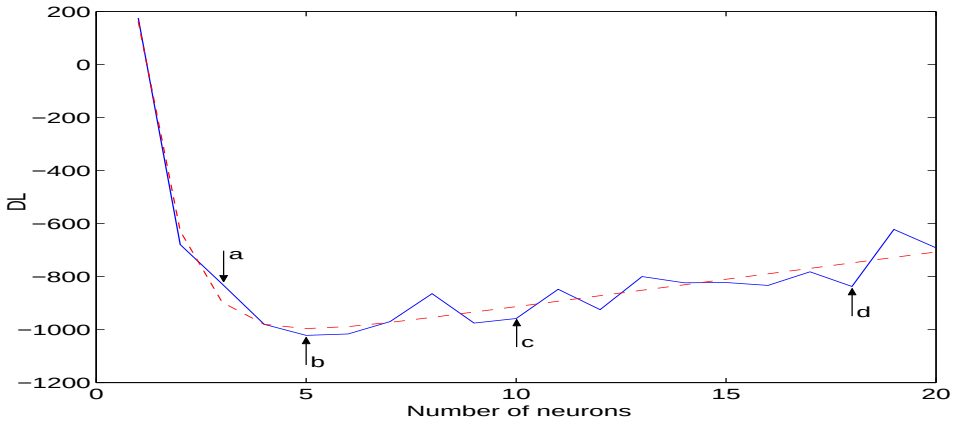


Fig. 2.5: The Rössler systems without noise (left panel) and with added dynamic noise (right panel).

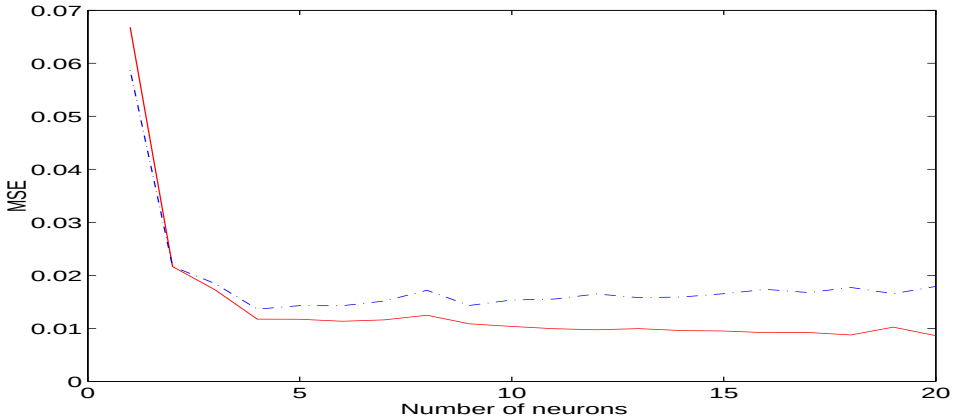
integrate the ODE equation 2.14 with the step size of 0.5 and then use the integrated results as initial data for the next step. The standard deviation of the Gaussian noise is set at 9% of the standard deviation of the data. We generate 2000 data points of which 1600 points are selected as a training set for the neural network and the rest are used as a test set. We calculate description length of each network constructed with from one to twenty neurons, i.e. D(1), D(2), ..., D(20). Results are shown in Fig. 2.6 with mean square errors of these networks.

Referring to Fig. 2.6(a) we find both the DL curve and fitted curve attain the same minimal point, which denotes that for this application the optimal number of neurons estimated is five. Note that mean square error of the test set in Fig. 2.6(b) starts increasing at the fourth point. This reflects the fact that for the large networks the errors between test data and its prediction become large gradually and these networks tend to overfit. The mean square error is one of criteria to measure the performance of models. It can be used to imply the possible overfitting of models but it do not estimate the exact position of overfitting and accordingly do not guarantee that the minimal mean square error estimates the optimal model. However, our method successfully estimate the optimal model, which is consistent with the trend of the MSE of testing data.

We select four networks with different numbers of neurons to perform free-run



a



b

Fig. 2.6: Description Length of the noisy Rössler data (solid line) in the top panel has the minimal point at five (i.e. five neurons), and the fitted curve (broken line) attains the minimum at the same point. In the bottom panel the solid line represents mean square error of training set and the dotted line represents that of test data.

prediction for the test set. The prediction data is converted into three vectors, $x(t)$, $x(t+3)$, and $x(t+5)(t \in [1, 395])$, as shown in Fig. 2.7.

The network that consists of five neurons can accurately capture the dynamics of the Rössler system but with less or more neurons the network is apt to underfit or overfit, respectively.

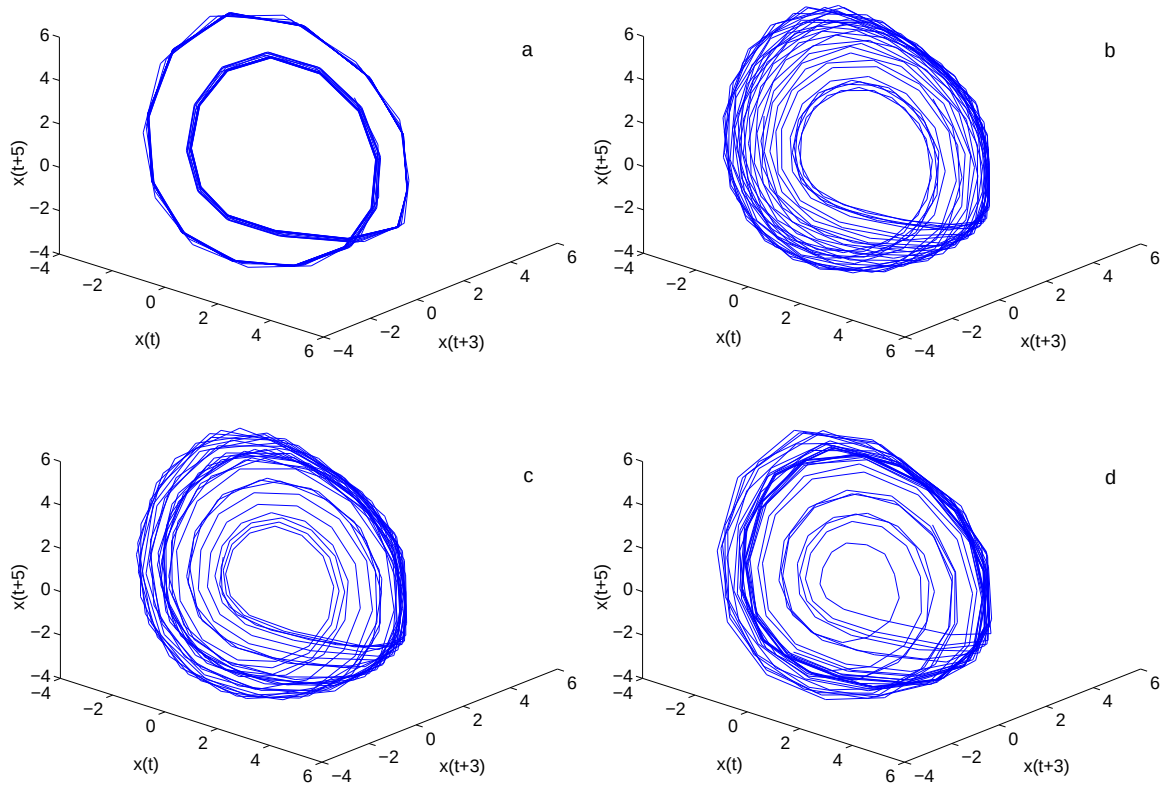


Fig. 2.7: Four Rössler systems of four hundred free-run prediction points predicted by the networks with 3, 5, 10, and 18 neurons, which are labeled as a, b, c, and d in Fig. 2.6(a).

2.3.2 The Ikeda Map

The second computational simulation is the reconstruction of the Ikeda map with dynamic noise. We add the dynamic noise to the Ikeda Map in the same way. The standard deviation of the noise is set at 30% of the standard deviation of the data. The equations of the Ikeda map are given by:

$$\begin{cases} x(t) = 1 + \mu(x(t)cos\tau - y(t)sin\tau) \\ y(t) = \mu(x(t)sin\tau + y(t)cos\tau) \end{cases} \tag{2.15}$$

where μ is equal to 0.7, and $\tau = 0.4 - 0.6/(1 + x(t)^2 + y(t)^2)$.

Fig. 2.8 presents the reconstrunction the x-component with two vectors, $x(t)$ and $x(t+1)$. We generate 1000 points of this map, of which 600 points are selected

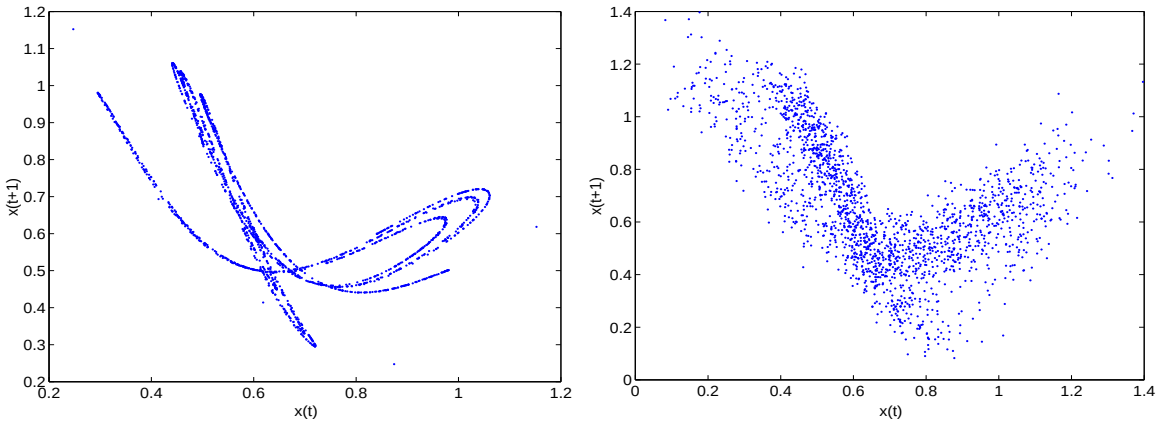


Fig. 2.8: The Ikeda maps without noise (left panel) and with added dynamic noise (right panel).

as a training set for the neural network and the rest are used as test data. Fig. 2.9 presents description length and mean square error of the training data and test data.

Although the original DL curve in Fig. 2.9(a) attains the minimum, it fluctuates somewhat dramatically, which we do not expect. However, the fitted curve is smooth and attains the same minimum point as the DL curve. We found that mean square error of the test set also begins to increase at the certain point.

Fig. 2.10 presents four Ikeda maps, where the free-run prediction data is converted into $x(t)$ and $x(t+1)(t \in [1, 399])$.

We observe the network that contains three neurons can predict the best result. The attractor in Fig. 2.10(b) is almost identical to that of the true Ikeda map. Whereas with fewer neurons, the network can not fit either the training set or test set well, and with more neurons the neural network is apt to overfit: the attractors in Fig. 2.10(d) are extremely "noisy". So description length can estimate the correct number of neurons in the neural network (the optimal network) for the reconstruction of the Ikeda map.

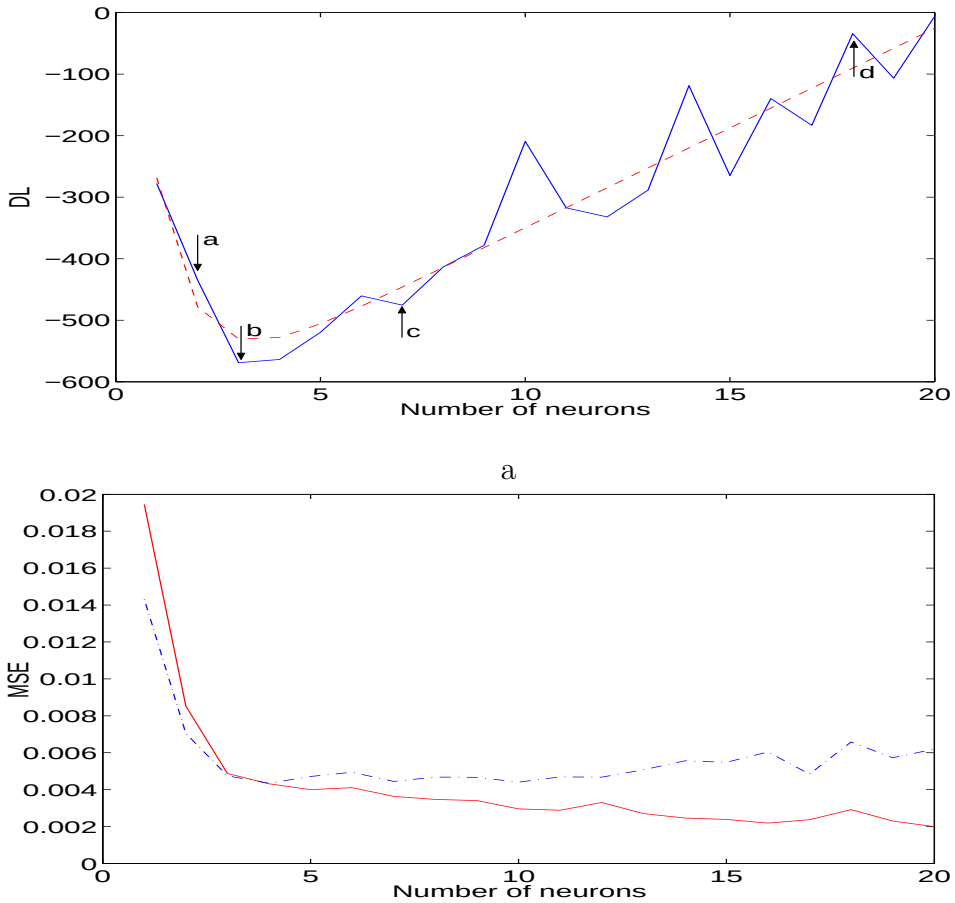


Fig. 2.9: Description Length of the noisy Ikeda data (solid line) in the top panel points out the minimum at three (i.e. three neurons), and the fitted curve (broken line) also shows the minimum at three. The solid line and dotted line in the bottom panel are mean square errors of training set and test set respectively.

2.3.3 Comparative Experiments

In the preceding subsections we demonstrated that the method of description length can determine the optimal networks for two specific applications. However, we were not sure whether alternative methods such as a Bayesian learning algorithm or early stopping can train the networks to attain good or even better results. To address this question we adopt these standard methods to avoid the overfitting of selected networks for the same applications.

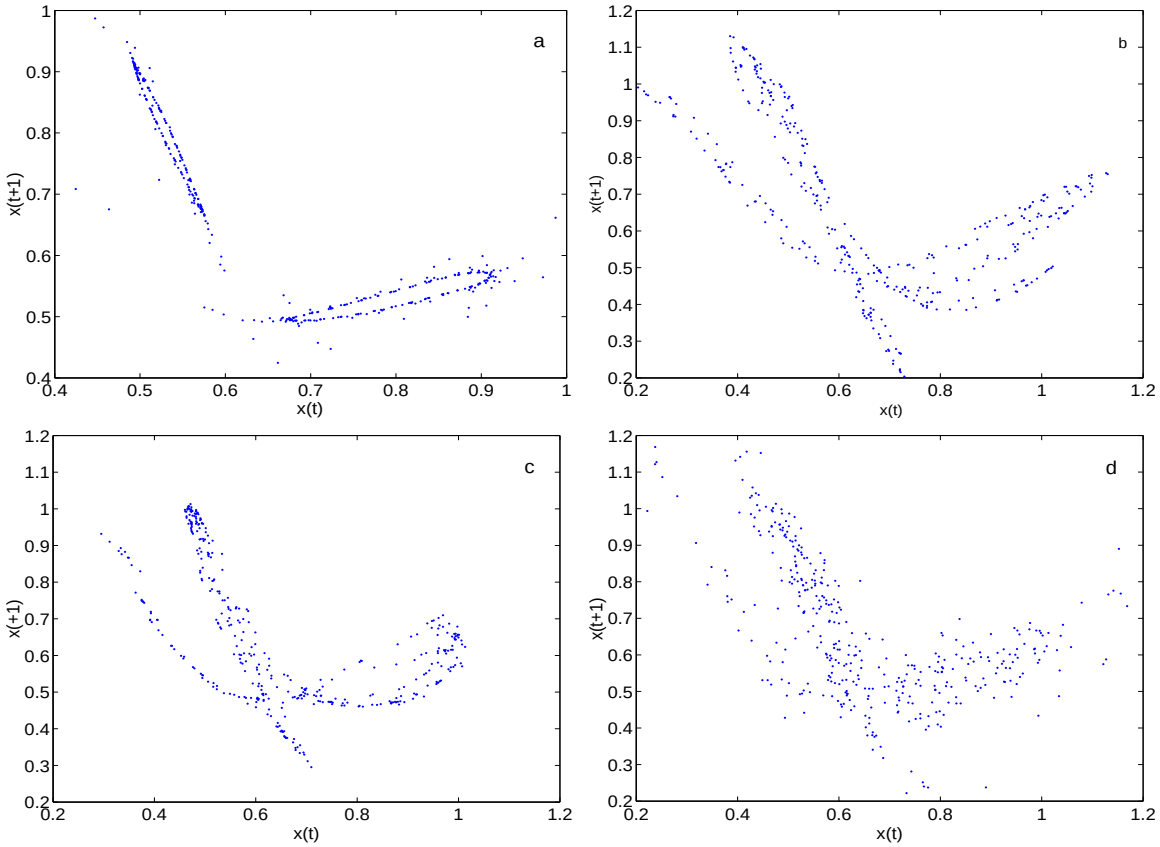


Fig. 2.10: Four Ikeda maps of four hundred free-run prediction points predicted by the networks with 2, 3, 7, and 18 neurons, which are labeled as a, b, c, and d in Fig. 2.9(a).

As mentioned in Section 3.1, the Bayesian learning algorithm [114] provides a measure of how many network parameters (weights and biases) are being effectively used by the network. This effective number should remain approximately the same, no matter how large the total number of parameters in the network becomes. This assumes that the network has been trained for a sufficient number of iterations to ensure convergence. So we apply the Bayesian learning algorithm to train large networks whose neurons are between eighteen and thirty. Such neurons are more than adequate for the previous time series prediction. For the Rössler system we utilize the same 1600 point training set and 400 point test set; for the Ikeda map the same 600 point training set and 400 point test set are used. We also ensure sufficient iterations during training neural networks. Typical results are shown in

Fig. 2.11.

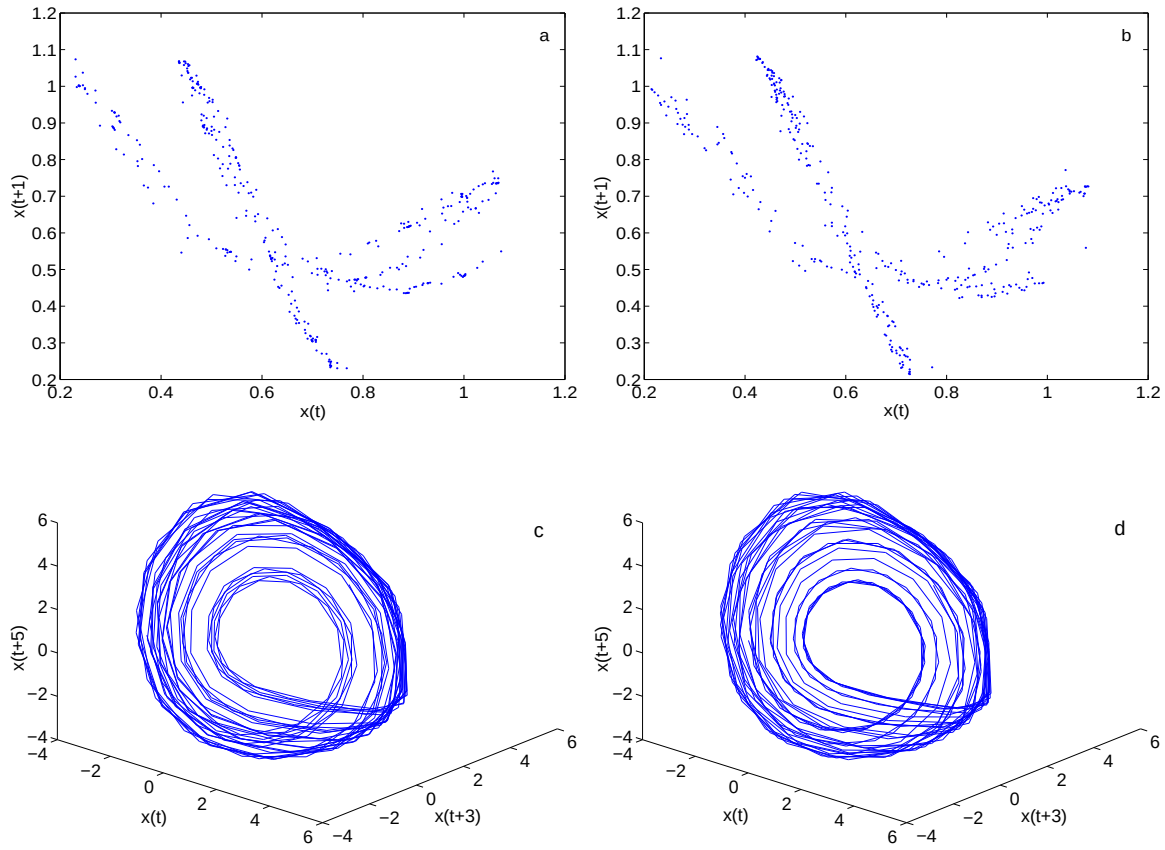


Fig. 2.11: Two Ikeda maps (a, b) of four hundred free-run data points predicted by the networks with 18 and 25 neurons trained by the Bayesian learning algorithm; two Rössler systems (c, d) of four hundred free-run data points predicted by the networks with 20 and 30 neurons trained by the same algorithm.

Referring to these figures we found that for either the Ikeda map or Rössler system the large networks can capture the basic underlying dynamics. Without the Bayesian learning algorithm these large networks are very likely to overfit. We agree that the Bayesian learning algorithm improves the generalization of the networks. However, comparing with Fig. 2.7(b) and 2.10(b) the optimal networks estimated by the minimum description length can capture the dynamics more precisely, which indicates that the optimal networks possess better generalization than the networks trained by the Bayesian learning algorithm.

For the method of early stopping we also repeat the experiments and construct

the attractors of the free-run prediction, but these attractors are considerably poor. One of the primary reasons is that we do not know how to choose the validation set which should be representative of all points in the training set. Based on our experiences, correct choice of the validation set is vital but very difficult to achieve.

In the following we apply our information theoretic method to two kinds of experimental data.

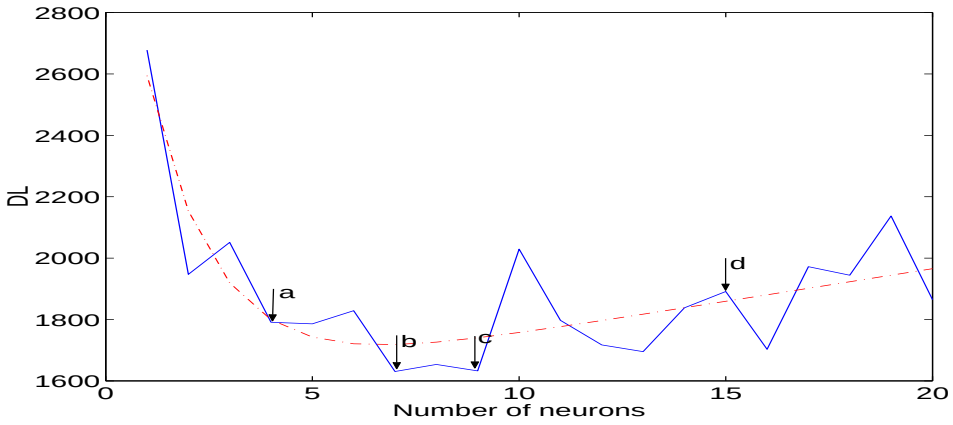
2.3.4 Chaotic Laser Data

Experimental data, such as of the chaotic laser data and ECG data that are natural phenomena are more complicated than the above computational data. These experimental data are more difficult to predict accurately. So we wish to substantiate the utiligy of this information theoretic method to experimental data from two practical systems: experimental recording of the chaotic laser data utilized in the 1992 Santa Fe time series competition and human ECG measurement, both of which are the focus of considerable attempts to model dynamics.

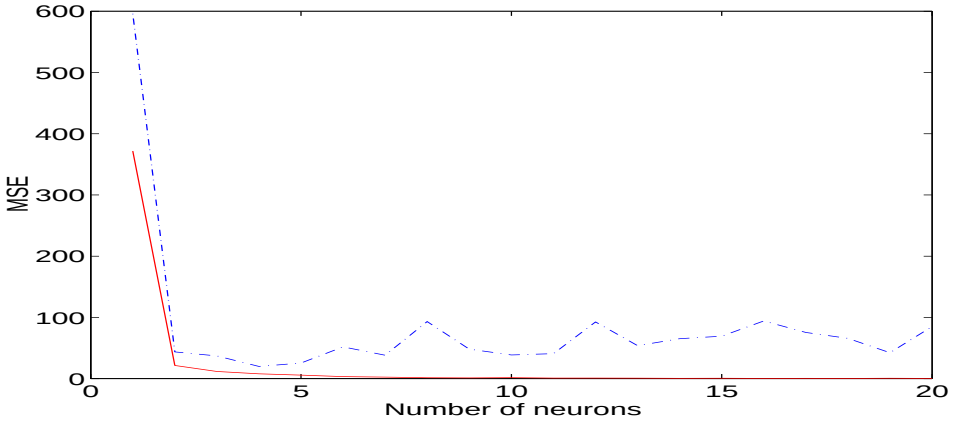
We select the 1000 laser data points, of which 600 points are used as the training data and the rest are the test data. Description length and mean square error of every neural network constructed with from one to twenty neurons are shown in Fig. 2.12.

The MSE curve of test data still starts to increase at certain point (the fourth point), but comparing to the former two MSE curves in Fig. 2.6(b) and 2.9(b) the increasing trend is not obvious. Fig. 2.13 presents 400 free-run prediction points predicted by networks with 4, 7, 9, and 15 neurons.

Although all the prediction of selected neural networks can not exactly follow the original laser data the network with seven neurons captures the underlying dynamics of the laser data best based on a qualitative comparison of free-run dynamics. It is not surprising that the network with nine neurons can obtain relatively good prediction, because DL curves indicate a range for selecting neurons, including the minimal point. It suggests that networks with neurons in this range can provide



a



b

Fig. 2.12: Description length of the laser data (solid line) in the top panel attains the minimal point at seven (i.e. seven neurons), and the fitted curve (broken line) attains the minimum at same point. The solid line is mean square error of training set with mean square error of test data (dotted line) in the bottom panel.

good fitting for the data with high probability, and the network with neurons corresponding to the minimum of the DL curve (fitted curve) can provide adequate generalization and capture the dynamics of time series most exactly.

2.3.5 Human Electrocardiograph Data

ECG data (during sinus rhythm) was collected by a unique data collection facility established in the coronary care unit (CCU) of the Royal Infirmary of Edinburgh [38]. From 3000 data points, we utilize 2550 points to build networks and the other

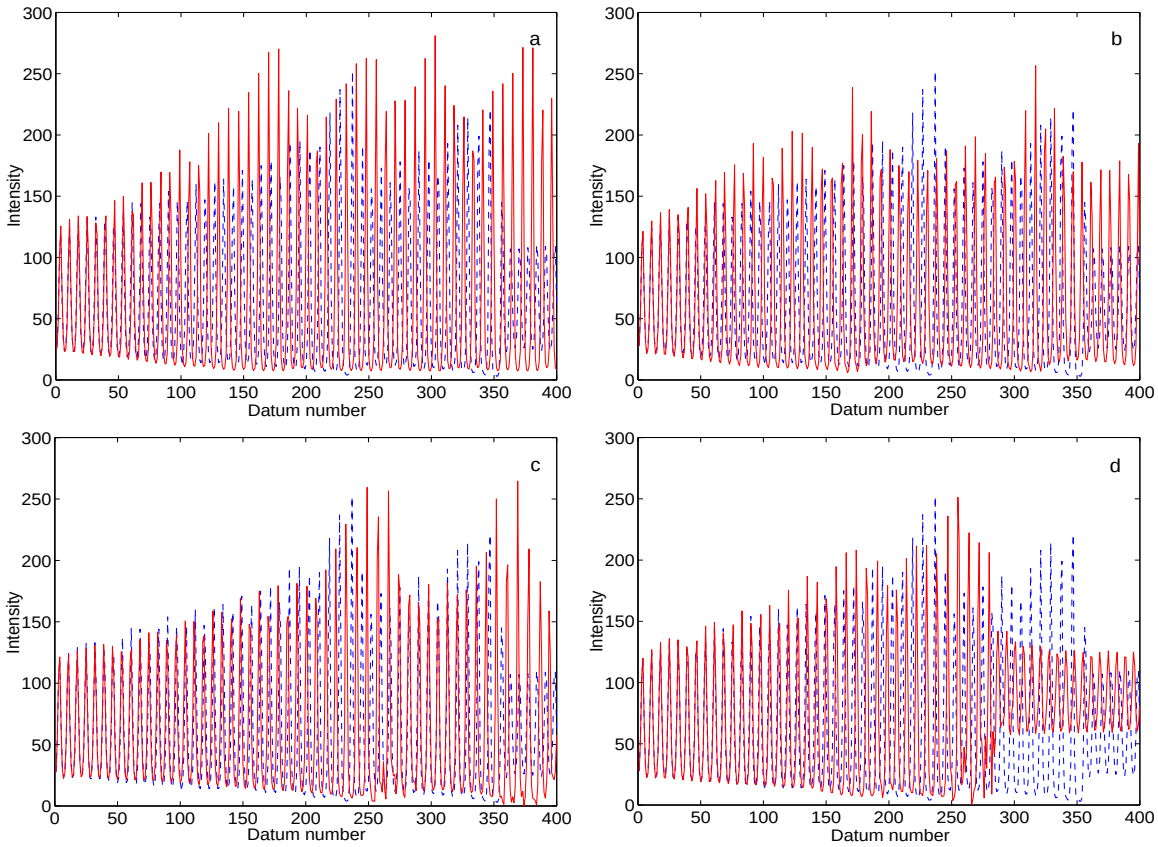


Fig. 2.13: The orignal chaotic laser data (broken line) and its free-run prediction (solid line) obtained by neural networks with 4, 7, 9, and 15 neurons, which are labeled as a, b, c, and d in Fig. 2.12(a).

450 to test them. Note that among the previous three simulation examples, every false nearest neighbors curve will drop to zero and remain at zero, but for ECG data the process is similar to that of contaminated signal (i.e. it does not reach zero). So the original ECG data is pre-processed by a low-pass Chebyshev Type II filter with normalized cutoff frequency 40 Hz in order to remove any noise at 50 Hz, and we then apply false nearest neighbors to calculate the filtered data. The power of the filtered data is 99.79% of the power of the original data. Fig. 2.14 describes the description length and mean square error for this application.

The DL curve suggests the optimal number of neurons is sixteen. But the fitted curve estimates that the optimal number of neurons is eight. In this experiment the MSE of test data cannot help us determine which networks are susceptible

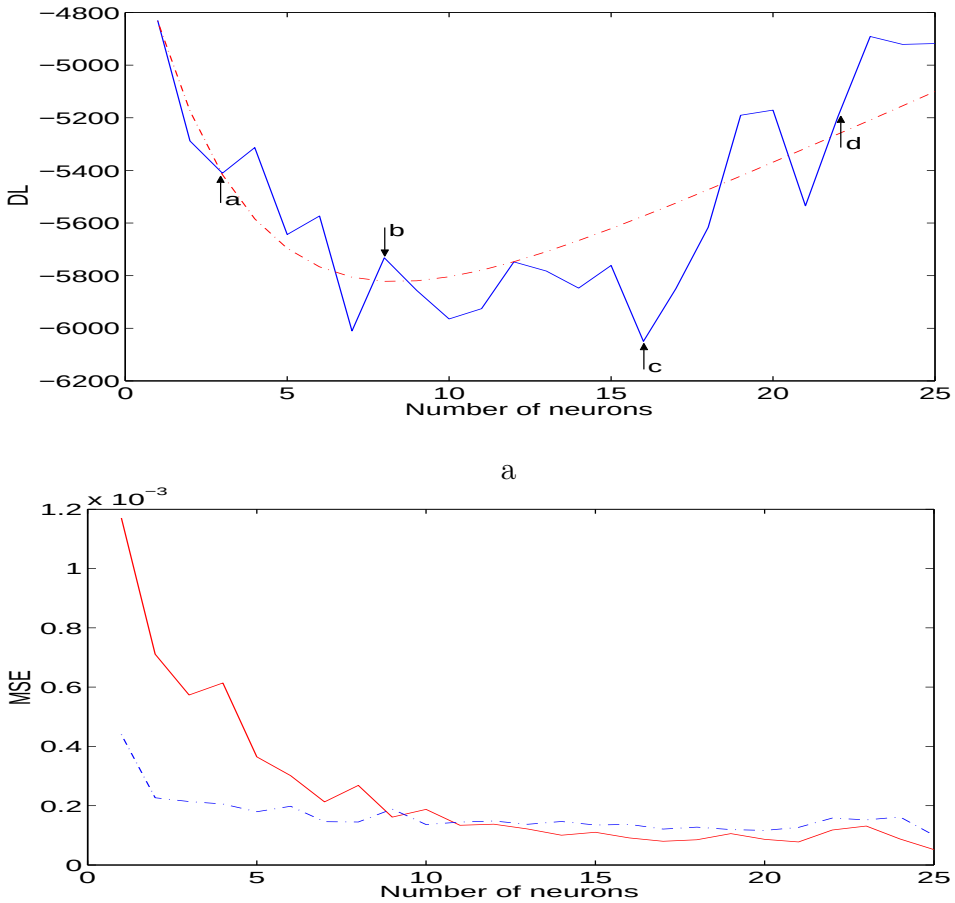


Fig. 2.14: Description Length of ECG data (solid line) and its fitted curve (broken line) in the top panel estimate the minimal point at the sixteenth and eighth point respectively. The solid line is mean square error of training data with mean square error of test data (dotted line) in the bottom panel.

to overfitting. So we deliberately select free-run prediction of the network with eight neurons and the networks with sixteen neurons, three neurons, and twenty-two neurons for comparison, as shown in Fig. 2.15.

Networks with eight neurons can predict the test data best, but the network with sixteen neurons predicts poorly. So we conclude that the optimal number of neurons is eight, i.e. the network with eight neurons can provide adequate fitting in this experiment, and networks with more neurons, such as sixteen neurons, overfit. Although referring to the DL curve we may select the wrong networks, the fitted

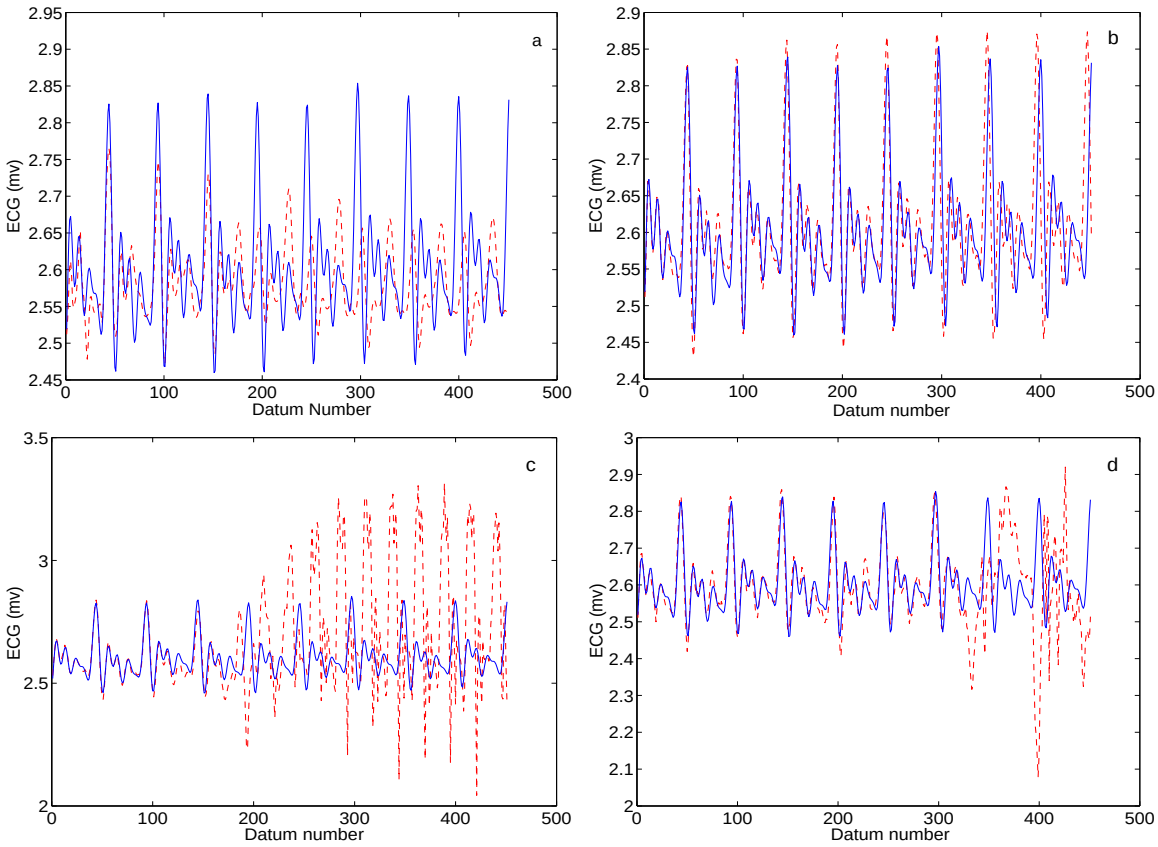


Fig. 2.15: The orignal ECG data (solid line) and its free-run prediction (broken line) for four networks with 3, 8, 16, and 22 neurons, which are labeled as a, b, c, and d in Fig. 2.14(a).

curve captures the true changing tendency of the description length of networks and correctly estimate the optimal neural networks. In the above experiment we noticed that the DL curve actually estimates a range for selecting neurons, not only the minimum. In this experiment we found that the networks with seven and nine neurons also obtained relatively good results.

2.4 Conclusion

The information theoretic approach described in this chapter is used to select the model size of neural networks. This algorithm determines the optimal model for a specific application in terms of the minimum description length. However, DL

curves fluctuate somewhat dramatically preventing an accurate estimation of the true description length. We thus employ nonlinear curvefitting to approximate the true DL curve. We demonstrate the application of the minimum description length method and nonlinear curvefitting technique to four typical time series predictions. Description length can directly estimate optimal numbers of neurons for the Ikeda map and the Rössler system, as does nonlinear curvefitting. For chaotic laser data and ECG data, description length with the help of nonlinear curvefitting can also estimate optimal networks. In all experiments, these optimal networks can provide adequate generalization and capture the dynamics very well.

Regularization techniques optimize parameters of the known neural network to improve the generalization of this neural network. They do not determine how many neurons of the neural network are sufficient to a specific application. However, the minimum description length criterion can determine the optimal model size directly. We found that optimal models estimated by this criterion consist of a small number of neurons. In despite of this, such models generalize well (have good prediction error performance on new data) and also capture the underlying dynamics (the deterministic attractor of the model is the same, or similar, to that of the data). Furthermore, we observe that such models actually perform better than larger models built by using standard methods.

Note that the models discussed here are time-invariant (i.e. stationary) systems. For time-varying systems, the unknown time-varying parameter may affect model building in an arbitrary way, and can therefore be modelled with an arbitrary number of models. If the instantaneous values of the time-varying parameter are known, then it can be added as an additional model parameter (an exogenous input). Modelling such processes with MDL type model selection technique is possible [40].

In theory the MDL approach is also applicable to neural networks with multiple hidden layers since it aims to calculate the cost of model parameters and the cost of model errors. If we apply this approach to such models the cost of model parameters would contribute more to the description length of the model. However, the trend of

the DL curve that $M(k)$ increases and $E(k)$ decreases with increasing k still exists and then the minimum description length estimates the optimal model. However, in practice the computational cost of MDL of complicated models would be very high or even not feasible. Conventional techniques, like regularization may be more practical in this case.

The algorithm to train neural networks is the Levenberg-Marquardt algorithm [41], which is a very general learning function. When applied to neural networks trained with different training algorithms the minimum description length method will estimate different optimal model sizes. This is easily understood. The same four experiments have been repeated to estimate optimal model sizes of neural networks with different training algorithms. We found that the MDL criterion can also select optimal numbers of neurons for the preceding four application. The method of minimum description length thus appears to be robust for neural networks whose training algorithms are different.

In the next chapter we apply optimal neural networks determined by MDL to model the dynamics of human blood propagation pressure signals so as to solve a practical problem.

3. APPLICATION OF OPTIMAL MODELS TO BLOOD PRESSURE PROPAGATION

3.1 Introduction

Traditional Chinese medicine practitioners (TCMP) always feel the pulse on a patient's wrists during diagnosis. This procedure has been routine in traditional Chinese medicine for thousands of years. It is both convenient and easy for TCMP to feel the pulse on the wrist. But is there any other significant advantage for feeling the pulse on the wrist? What about feeling the pulse at other locations, such as carotid artery or fingertips [1], is there any significant difference other than the signal intensity? Divisions of feeling pulse on the wrist is described in Fig. 3.1 for reference. Similarly, one may ask the same questions for other cardiac data, such as human ECG data. Does ECG data collected from different parts of one person reflect the same result? More generally, is there any distinction between measuring ECG and lateral arterial pulse. In this chapter we focus on the first question whether there is any significant difference between measuring the pulse on the wrist and other carotid artery.

To answer this question, we take the approach of surrogate data to make a decision. One may apply it to determine whether an observed time series has a statistically significant deterministic component. But the surrogate data method alone cannot separate the noise and deterministic component. We need to identify the deterministic component of blood propagation between two different pulse waveforms. To achieve this, we apply optimal neural networks described in the last chapter

[1] Modern doctors more often measure patients' pulse at their fingertips by electronic devices if necessary.

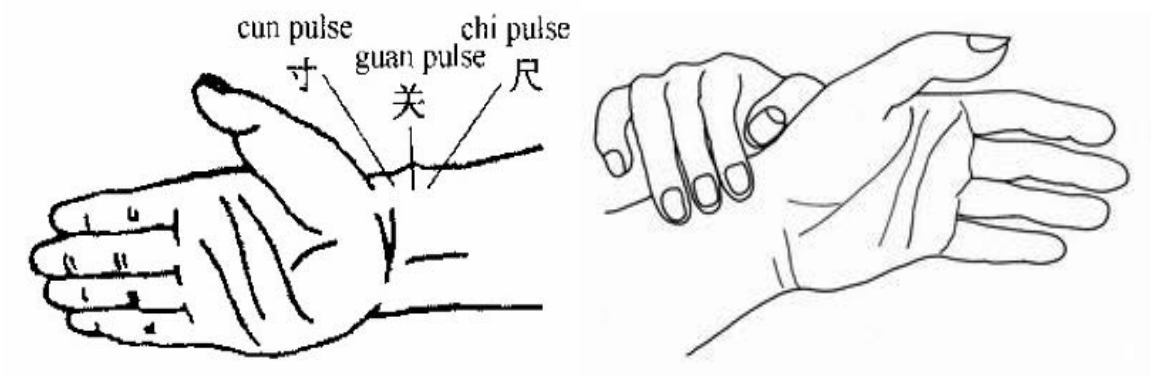


Fig. 3.1: Traditional Chinese medicine practitioners usually locate three fingers (the first finger, middle finger, and third finger) on the patient's wrist, as listed in this figure (Image courtesy of [43]).

to model the nonlinear transformation (assumed there is nonlinear transformation) from the wrist to fingertip.

3.2 The Surrogate Data Method

The surrogate Data method, suggested and implemented in [42], has been widely applied in the literature. The rationale of the surrogate data method is to generate an ensemble of artificial surrogate data (surrogates for brevity) that are both "like" the original data and consistent with some null hypothesis. One then applies some test statistics to both the surrogates and the original data. The concept of the surrogate data test is visualized in Fig. 3.2 for the case of the null hypothesis of a linear noise.

Let $\{x_t\}_{t=1}^{N} = \{x_1, x_2, x_3, ..., x_N\}$ be a time series of N measurements, which is abbreviated as $\{x_t\}$. For each class of dynamical system Φ, i.e. a hypothesis, one generates an aggregation of N surrogates $\{s_t\}$ $(n = 1, 2, 3, ..., N)$, which is consistent with both the data $\{x_t\}$ and the class of dynamical systems being tested. One then calculates the test statistic, $f(\cdot)$ for both original data and surrogates. If $f(\{x_t\})$ is distinct from the aggregation of $f(\{s_t\})$, one can reject the class of dynamical systems Φ (the hypothesis) as the likely origin of $\{x_t\}$; but if $f(\{x_t\})$ is typical of

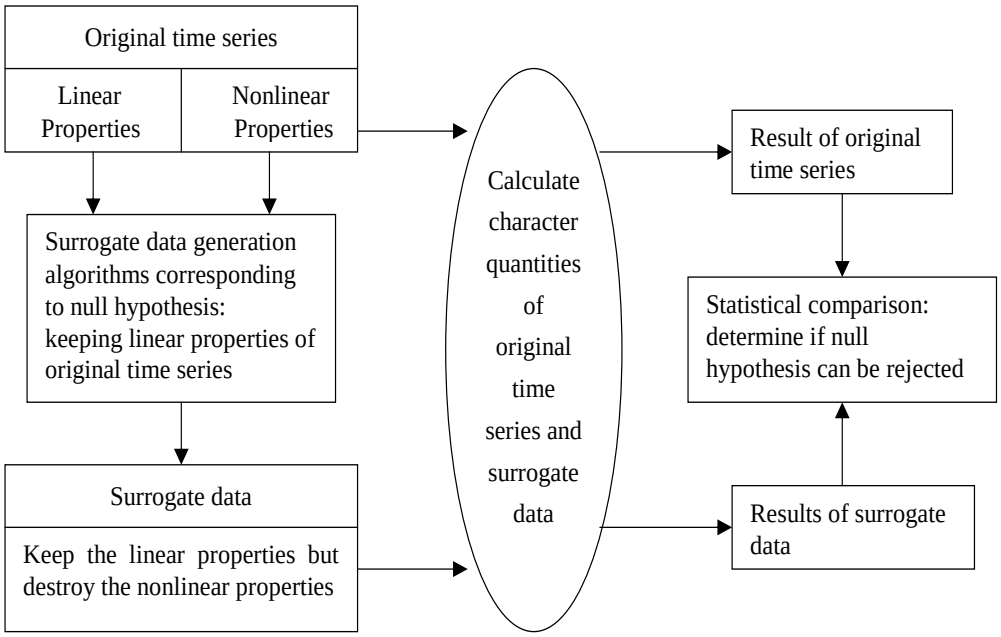


Fig. 3.2: Framework representation of the surrogate data method for the case of the null hypothesis of a linear process (Figure based on [44]).

$f(\{s_t\})$, the class of dynamical system Φ cannot be rejected. Note that failure to reject a specific class of dynamical systems does not mean one can accept that class as the likely origin of data. There may still exist a different statistic which is able to discriminate between original and surrogates.

3.2.1 Null Hypotheses of the Surrogate Data Method

Commonly employed null hypotheses include [42]:

1. NH0: The data X_t is independent and identically distributed (i.i.d.) noise of with unspecified mean and variance, e.g., Gaussian white noise.

2. NH1: The data X_t is linearly filtered noise generated via an AR process,

$$X_t = \mu + \sum_{i=1}^{p-1} \phi_i X_{t-i} + \sigma \cdot \xi_t \tag{3.1}$$

   where μ, ϕ_i, and σ are unspecified parameters and ξ_t are i.i.d noises.

3. NH2: The data X_t is static monotonic nonlinear transformation [2] of linearly filtered noise generated by the above formula.

So far these are the most important null hypothesis for the surrogate data test, but it is also possible to formulate other null hypothesis, like the hypothesis of periodic orbits with uncorrelated noise described in the next chapter.

3.2.2 Generation of Surrogate Data Sets

Surrogate generation algorithms are originally illustrated in [42]. There are three algorithms to generate surrogates, known as Algorithm 0, Algorithm 1, and Algorithm 2 corresponding to the above three null hypotheses.

1. Algorithm 0: shuffles the order of the data. Such shuffling will destroy any temporal correlation. In essence such surrogates are random data consistent with the same probability distribution as the original. For the hypothesis NH0, Algorithm 0 is adopted to generate surrogates.

2. Algorithm 1: surrogate data produced by this algorithm are linearly filtered noise. To generate these surrogates one employs the discrete Fourier transform (DFT) of the data and shuffles (or randomizes) the phases of the complex conjugate pairs. The surrogates are the inverse discrete Fourier transform. By shuffling the phases but maintaining the amplitude of the complex conjugate pairs the surrogate will have the same power spectrum as the data, but will have no nonlinear determinism.

3. Algorithm 2: amplitude adjusted Fourier transform (AAFT) algorithm. Surrogates generated by this algorithm are static monotonic nonlinear transformations of linearly filtered noise. One rescales values of the original data so that they are gaussian, and then apply Algorithm 1 to generate the surrogates

[2] A static filter is a function g, such that $y_t = g(x_t)$ does not depend on previous or future values, $x_{t\pm i}$, or on derivatives of x_t.

that have the same power spectra as the rescaled data. The algorithm aims to preserve both the power spectrum and probability distribution of the data.

The Algorithm 1 and Algorithm 2 (particularly Algorithm 2) are hampered by technical issues related to the Fourier transformation. If the original time series is stationary and adequately long Algorithm 1 can work well without limitation [42]. Note that for the real data we need to analyze the stationarity of the data before applying the surrogate data method with one of the two algorithms to it. Otherwise, non-stationary data would increase false rejections of the given hypotheses. Surrogates generated by Algorithm 2 usually fail to keep exactly the same power spectra as the original and such systematic errors can result in high false rejections [45] [76]. Solution to technical problems of this algorithm are also available in the same literature but require great computational cost. In addition, there are other algorithms to produce surrogate data, such as the pseudo-periodic surrogate (PPS) algorithm [62] and cycle-shuffled surrogate algorithm [48]. The surrogate data method with these algorithms tests the hypothesis that an observed time series is periodic orbit driven by uncorrelated noise. We will discuss the PPS algorithm and cycle-shuffled surrogate algorithm at length in the next chapter.

3.2.3 Test Statistics of the Surrogate Data Method

To test the hypothesis of surrogate data we must select an appropriate statistic. There are many discriminating statistics, such as correlation dimension, complexity, Lyapunov exponent etc. In this chapter we select correlation dimension as the discriminating statistic. Correlation dimension provides a simple way to distinguish a random signal from a strange (possibly chaotic) set and characterize the strange attractor. For example, in principle a random data has an "infinite" correlation dimension. Intuitively, an orbit of random noise does not have any spatial structure. In contrast, the correlation dimension for a closed curve (for example, a periodic orbit) is 1, and for a two-dimensional surface is 2. A strange attractor can have a correlation dimension that is not an integer.

Correlation dimension proposed by Grassberger and Procaccia [49] is intended to compute nonlinear correlation between points of the reconstructed attractors in phase space. Given a scaler time series $\{x_i\}_i$, by selecting proper embedding dimension, d_e and delay time, τ we can obtain a set of vectors $\left\{X_i = (x_i, x_{i+\tau}, ..., x_{i+(d_e-1)\tau})\right\}$. The correlation integral is defined as follows [49],

$$C(\varepsilon) = \lim_{N\to\infty} \frac{1}{N^2} \sum_{i,j=1}^{N} \theta\left(\varepsilon - |X_i - X_j|\right) \tag{3.2}$$

where N is the total number of points in the time series of x_i, ε is the threshold of interpoint distance (i.e. length scale), and $\theta(x)$ is the Heaviside function with $\theta(x) = 1$ for positive x, and 0 otherwise.

For a limited range of ε the correlation integral takes the form $C(\varepsilon) \propto \varepsilon^{d_c}$ where d_c is the correlation dimension, and hence

$$d_c = \lim_{\varepsilon\to 0} \frac{\log C(\varepsilon)}{\log \varepsilon}. \tag{3.3}$$

The Grassberger-Proccacia algorithm [49] is not robust for the finite time series or noisy time series. We, therefore, adopt the Gaussian Kernel Algorithm (GKA) [50] [51] to estimate correlation integral in Eq. 3.3, called the Gaussian kernel correlation integral $T_{d_e}(\varepsilon)$.

$$\begin{aligned} T_{d_e}(\varepsilon) &= \int d\overrightarrow{x}\, \rho_m(\overrightarrow{x}) \int d\overrightarrow{y}\, \rho_m(\overrightarrow{y}) e^{-\left|\overrightarrow{x}-\overrightarrow{y}\right|^2/4\varepsilon^2} \\ &\backsim e^{-d_e K\tau} \left(\tfrac{\varepsilon}{\sqrt{d_e}}\right)^{d_c} \quad \text{for } \varepsilon \to 0,\ d_e \to \infty \end{aligned}$$

where the parameter K is called the correlation entropy, which is estimated from the behavior of $T_{d_e}(\varepsilon)$ as the correlation dimension d_c does.

To compute the correlation integral in practice, we discretize the above equation and take the average over the inter-point distribution. The approximated estimation is given by [51]

$$\widehat{T}_{d_e}(\varepsilon) = \tfrac{1}{N_p} \sum_{i,j\neq i} \exp(-\left|\overrightarrow{x_i} - \overrightarrow{x_j}\right|^2 / 4\varepsilon^2).$$

By calculating the above two equations we can obtain the GKA correlation dimension estimation, d_c.

3.3 Equivalence between Pulse Measured on Human Wrists and Fingertips

We apply the surrogate data method to the model of blood pressure propagation from the wrist to the fingertip for six healthy candidates. The device measuring their pulse signals is PowerLab 4/25 of ADInstruments. The sampling rate is set to 100 Hz and resolution is 16 bits. Note that we collect the pulse data on both the wrist and fingertip (forefinger) at the same time. Pulse data of one candidate measured on the wrist and fingertip is presented in Fig. 3.3. There is a short time delay ($40ms$) between them since it takes certain time to propagate blood from the wrist to the fingertip.

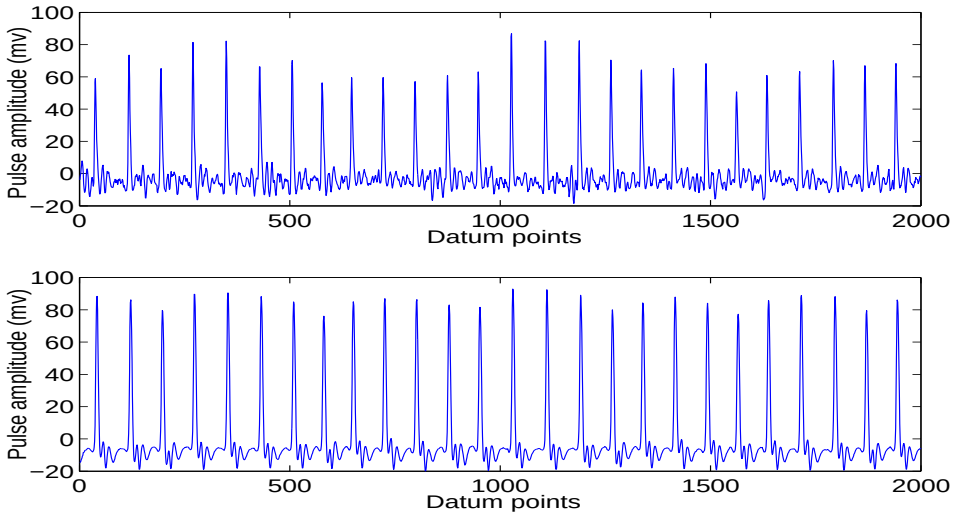


Fig. 3.3: Pulse data measured on the wrist (top panel) and the fingertip (bottom panel) of one subject.

From Fig. 3.3 we can find differences between the two waveforms. But, are there "significant" differences that are related to the dynamics of the data, or just observational noise? So we need to figure out whether there are differences on their intrinsic dynamics.

3.3.1 Application of the Surrogate Data Method to Pulse Data

The procedure to confirm deterministic propagation of blood pressure from human wrists to fingertips can be explained in the following three steps. Fig. 3.4 describes this procedure using the framework.

1. We utilize backpropagation neural networks with different numbers of neurons to model blood pressure propagation from the wrist to fingertip. When building models, neural networks use pulse data on the wrist to make one-step prediction of pulse data on the fingertip so as to try to capture the relationship between them. We select 2600 data points to build neural networks with another 420 data points as the test set. The Levenberg-Marquardt algorithm [41] is used to train neural networks.

2. After building models, we employ minimum description length with the help of nonlinear curvefitting to determine the optimal network. The optimal networks is then applied to make the prediction for the test set. Fig. 3.5 presents the DL curve and its fitted curve of neural networks with neurons from 1 to 23 for one case.

3. Finally we generate 30 surrogates for the one-step prediction error of the optimal model. The given hypothesis is NH0, i.e. the one-step error is i.i.d noise. We then calculate correlation dimension of the error and its surrogates with embedding dimension from 2 to 10.

The prediction of selected networks with 5, 8, 13, and 17 neurons listed in Fig. 3.5 for the test data is presented in Fig. 3.6. The network that is made up of thirteen neurons provides the best prediction among these four predictions.

For all the six volunteers the test pulse data on the fingertip has high correlativity with the corresponding prediction of the optimal network, as shown in Table 3.1. That is, the prediction obtained by the optimal networks is almost the same as the test data. Therefore, all optimal networks estimated by MDL accurately model the blood pressure propagation from the wrist to fingertip for all the cases.

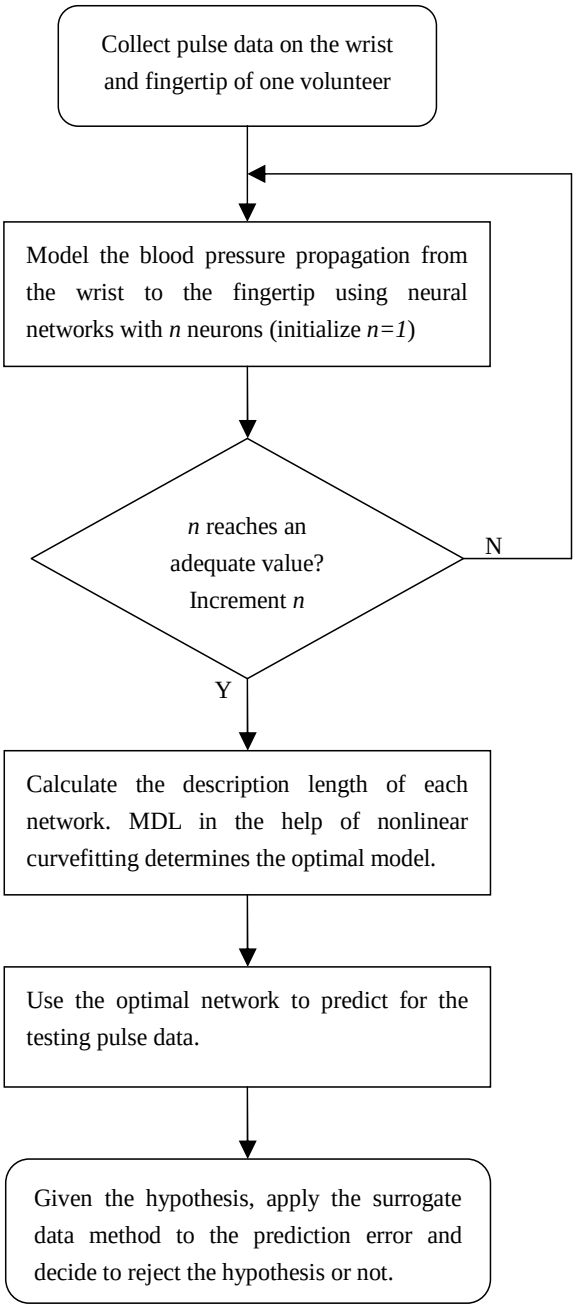


Fig. 3.4: Combination of neural networks, MDL, and the surrogate data method to confirm the deterministic blood pressure propagation from the wrist to the fingertip.

Based on whether the correlation dimension of the original data is out of or in the distribution of correlation dimension of surrogates, we decide whether to reject or fail to reject the given hypothesis. If we can reject it, we have to consider it is very likely that there is other influence, such as some dynamics in the prediction

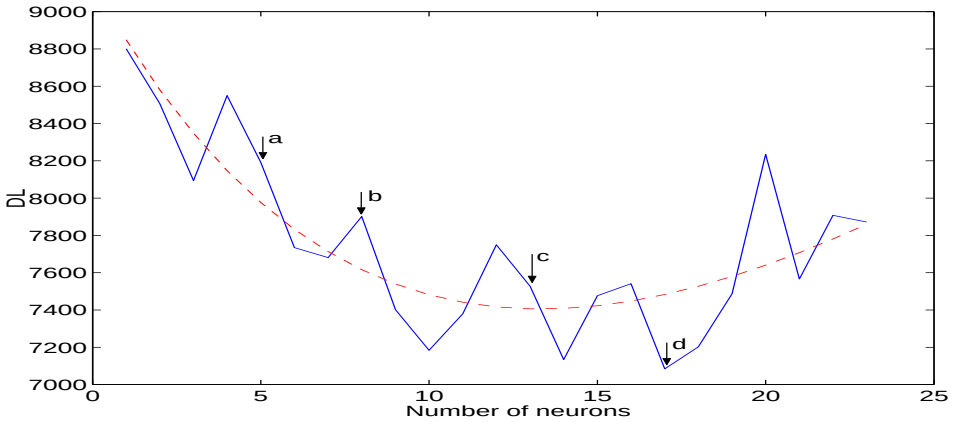


Fig. 3.5: Description length of pulse data (solid line) gets the minimal point at seventeen (i.e. seventeen neurons), but its fitted curve (dashed line) estimates the minimum at thirteen.

Tab. 3.1: Cross-correlation coefficient between the test pulse data on the fingertip and their corresponding prediction made by optimal networks for six volunteers.

Subject	Subject 1	Subject 2	Subject 3	Subject 4	Subject 5	Subject 6
Cross coefficient	0.9924	0.9956	0.9915	0.9934	0.9962	0.989

error, i.e. there is difference between feeling pulse on the wrist and on the fingertip. Typical results for the above case are illustrated in Fig. 3.7, where x-axis represents embedding dimension (d_e) and y-axis represents correlation dimension (d_c).

One can find that the correlation dimension of the original error stays in the range of the mean plus or minus one standard deviation between $d_e = 3$ and $d_e = 8$. Furthermore, most of them are close to the average, which means the correlation dimension of the original data is close to the center of the distribution of correlation dimension of surrogates, and then the original data cannot be distinguished from the results of the surrogates. Consequently we can not reject the given hypothesis of i.i.d noise, i.e. we cannot reject that there is no dynamic noise between the pulse data on the wrist and the fingertip. Note that the behavior of correlation dimension with embedding dimension higher than eight becomes unstable. The correlation dimension of surrogates at such embedding dimension, or higher, fluctuates dra-

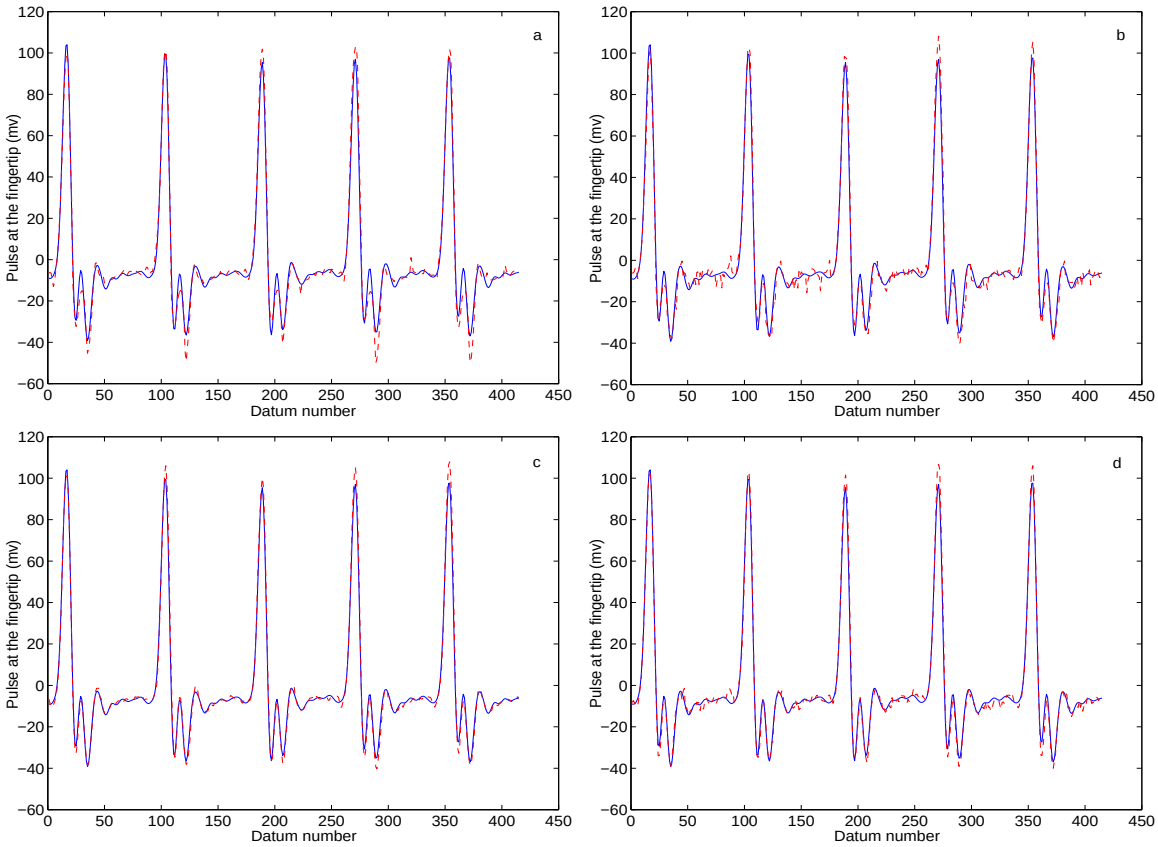


Fig. 3.6: The test pulse data (solid line) and its prediction (dotted line) obtained by networks with 5, 8, 13, and 17 neurons, as illustrated in (a), (b), (c), and (d) respectively.

matically between 0.8 and 2.6. Increasing embedding dimension yields increasing correlation dimension. But for embedding dimension sufficiently large correlation dimension will be equal to the true embedding dimension, and any further increase of embedding dimension should not change the value of correlation dimension any more. So the proper embedding dimension should be lower than eight in this experiment. It seems likely that for higher embedding dimension correlation dimension calculated by GKA algorithm fails to converge as expected. As we know, the model residual is just random noise (or regarded as a deterministic signal contaminated with strong noise). Although the GKA algorithm [51] is robust to small noisy data, the reconstructed vectors of the model residual with large embedding dimension are far away from the maximum noise level the GKA algorithm can tolerate. There-

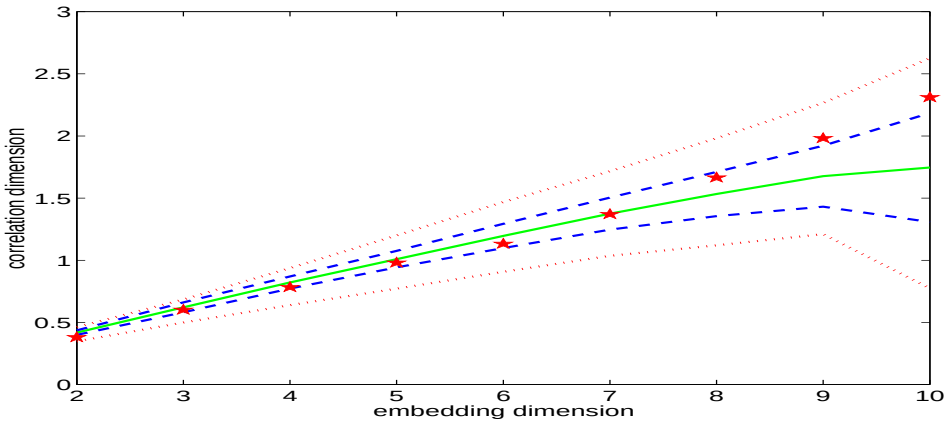


Fig. 3.7: Stars are correlation dimension of the original prediction error obtained by the optimal model for embedding dimension from 2 to 10; the solid line is the mean correlation dimension of 30 surrogates at each embedding dimension; two dashed lines denote the mean plus one standard deviation and minus one standard deviation; two dotted lines are maximum and minimum correlation dimension among these surrogates.

fore, the correlation dimension estimated by this algorithm keeps ascending with increasing dimension rather than steady.

3.3.2 Comparative Experiments

It, however, may be possible that the optimal models failed to distinguish between data and surrogates even though significant differences exist in the blood propagation from the wrist to the fingertip. To address this problem, we build another further experiment in which we examine the prediction of pulse data on the finger with (deterministic but independent) observational "noise". The "noise" is one component of the Rössler system described by the same equation as Eq. 2.14, where $a = 0.15$, $b = 0.20$, $c = 10.00$, and the sampling time t_s is 0.5.

We add x-component data of the Rössler system to the pulse data measured on the fingertip of the same volunteer, and then repeat the preceding experiment. We still select 2600 data to build neural networks with another 420 data as the test set. The magnitude of x-component data is set at 5.6% of the magnitude of the pulse

data on the fingertip; the standard deviation of the variants is set at 10% of the standard deviation of the pulse data on the fingertip. The dynamic noise, therefore, is considerably small comparing to this pulse data. Fig. 3.8 shows the pulse data of the fingertip with and without the added "noise".

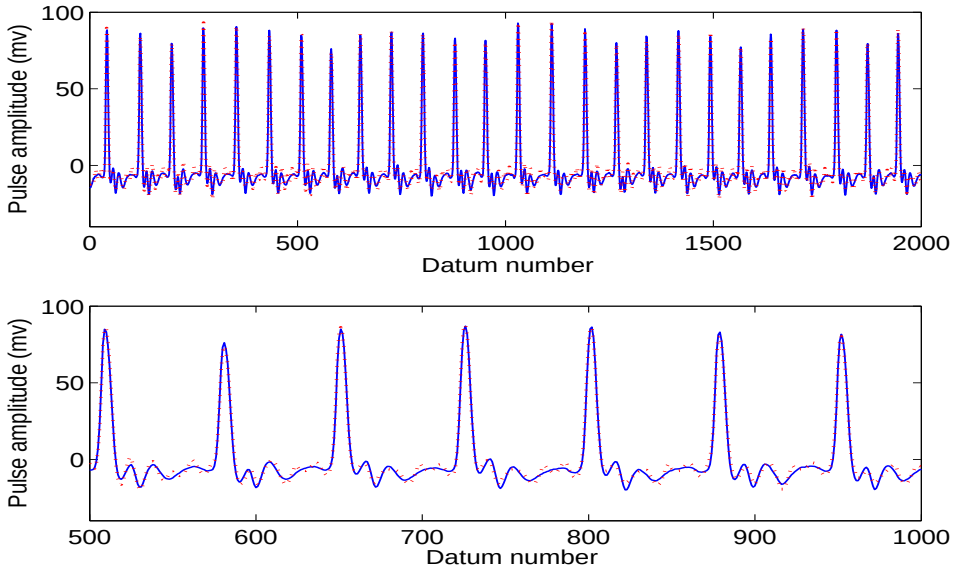


Fig. 3.8: The pulse data of the fingertip without noise (solid line) and with noise (dotted line) in the top panel. The part of them are enlarged in the bottom panel so as to illustrate their difference.

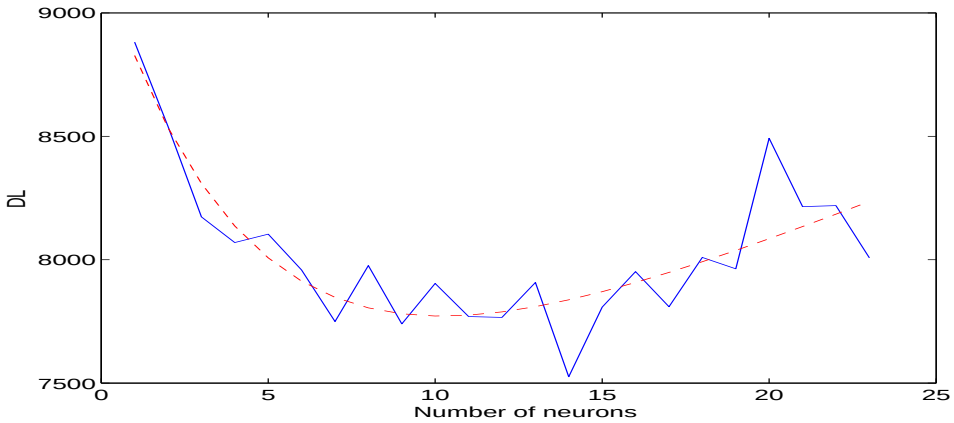


Fig. 3.9: The description length curve (solid line) and fitted curve (dashed line) get the minimal point at fourteen and at ten respectively. We select the trained neural network with 10 neurons as the optimal model.

The DL curve and fitted one are presented in Fig. 3.9. The one-step prediction of the optimal network with 10 neurons, as estimated in Fig. 3.9, is plotted in Fig. 3.10. Relevant results about correlation dimension for this pulse data contaminated with the noise are presented in Fig. 3.11, which highlights the deviation between the correlation dimension of the original data and surrogates. Correlation dimension of the original error is even far away from the range bounded by the minimal and maximal correlation dimension of surrogates. We can reject the given hypothesis that the original error is i.i.d noise. Since we artificially added the Rössler dynamic data to the pulse data on the fingertip, such dynamics should exist in the prediction error. This is consistent with our expectation. Again, for $d_e > 8$ the GKA fails to converge properly.

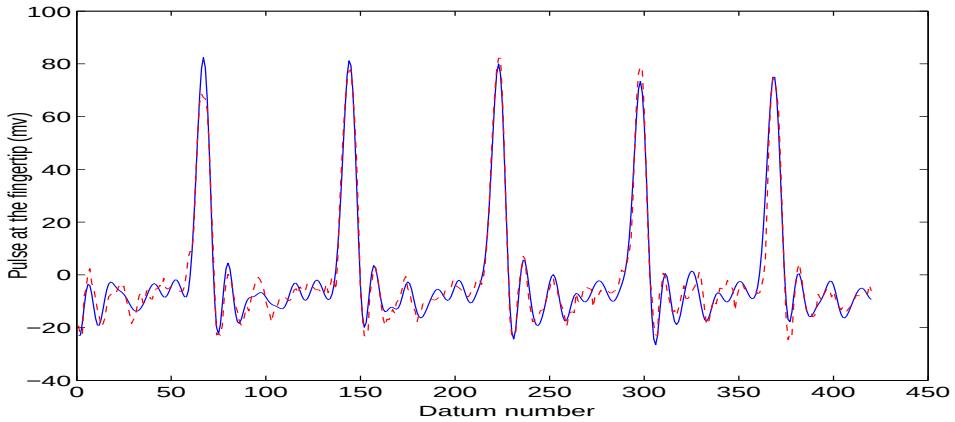


Fig. 3.10: The test pulse data contaminated with "noise" (solid line) and its prediction (dotted line) of the optimal network.

We, therefore, conclude that if there is certain significant difference, i.e. some dynamic noise between pulse data on the wrist and the fingertip, the optimal model can identify them, and then the surrogate data method can substantiate the existence of this deterministic dynamics in the model residual. On the other hand if there is no difference between model prediction and data, the surrogate data method will show corresponding results, as presented in Fig. 3.7.

We can not exclude the possibility that there exists small dynamic noise between pulse data on the wrist and fingertip which is weaker than the Rössler data

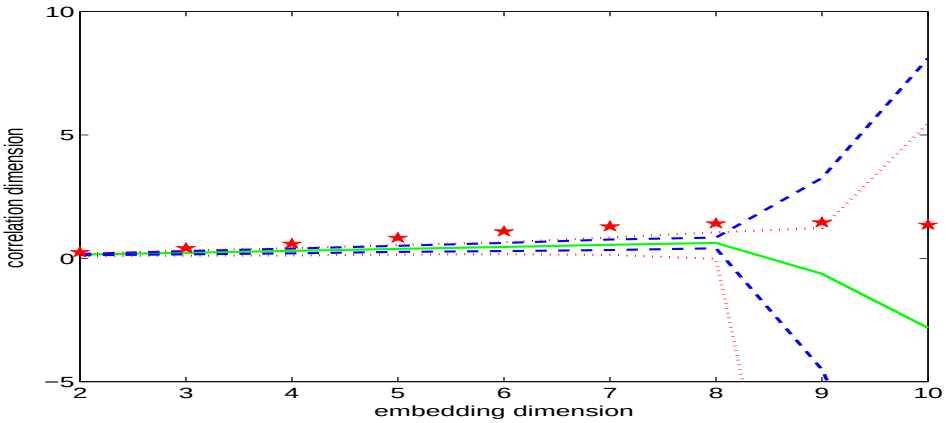


Fig. 3.11: Stars are correlation dimension of the original prediction error of the optimal model for embedding dimension from 2 to 10. Properties of the rest lines are the same as those in Fig. 3.7. Numerical problems with GKA are evident for d_e>8.

added and so can not be detected by our methods. However, we consider that this small dynamic noise also may not be detected by the traditional Chinese practitioners' hands, i.e. such dynamic noise has been ignored or does not efficiently reflect the patient's symptoms. Hence, even if this dynamic noise exists it is out of either practitioners' interest or our research interest.

As shown in Table 3.2, for all the case the correlation dimension of the prediction errors of the optimal models stays in the middle of distribution of its surrogates, and then we can not reject the hypothesis that the residual is i.i.d. noise. Consequently, we cannot reject that there is no significant difference between the pulse data on the wrist and the fingertip.

3.4 Conclusion

A procedure to capture the deterministic propagation of blood pressure from the human wrist to the fingertip is proposed in this chapter. In the preceding work we found that MDL can estimate optimal models for different kinds of nonlinear time series, including ECG data. Superficially, pulse data is extremely similar to ECG

Tab. 3.2: Results of all six volunteers' pulse data. d_c is correlation dimension of the original prediction error obtained by optimal networks, $< d_c >$ is the mean of correlation dimension of all surrogates, and σ is the standard deviation of them. d_e (embedding dimension) is selected from 5 to 8.

Subject \ d_e	**5**	**6**	**7**	**8**
Volunteer 1	d_c =0.981 $< d_c >$=0.9213 σ =0.1031	d_c =1.132 $< d_c >$=1.0877 σ =0.1495	d_c =1.371 $< d_c >$=1.2578 σ =0.1964	d_c =1.665 $< d_c >$=1.4288 $\sigma = 0.2459$
Volunteer 2	d_c =1.149 $< d_c >$=1.0801 σ =0.1155	d_c =1.298 $< d_c >$=1.2826 σ =0.1656	d_c =1.429 $< d_c >$=1.4618 $\sigma = 0.1800$	d_c =1.583 $< d_c >$=1.6493 σ =0.2296
Volunteer 3	d_c =0.975 $< d_c >$=0.9264 σ =0.1102	d_c =1.211 $< d_c >$=1.1094 σ =0.1411	d_c =1.486 $< d_c >$=1.2859 σ =0.1828	d_c =1.754 $< d_c >$=1.4619 σ =0.2317
Volunteer 4	d_c =1.083 $< d_c >$=0.9449 σ =0.1586	d_c =1.304 $< d_c >$=1.0873 σ =0.1992	d_c =1.504 $< d_c >$=1.2476 σ =0.2663	d_c =1.742 $< d_c >$=1.3763 σ =0.3052
Volunteer 5	d_c =0.590 $< d_c >$=0.5838 σ =0.092	d_c =0.706 $< d_c >$=0.6896 σ =0.1272	d_c =0.816 $< d_c >$=0.7924 σ =0.1586	d_c =0.944 $< d_c >$=0.8955 σ =0.1968
Volunteer 6	d_c =0.485 $< d_c >$=0.489 σ =0.111	d_c =0.562 $< d_c >$=0.5890 σ =0.1392	d_c =0.592 $< d_c >$=0.6867 σ =0.1574	d_c =0.699 $< d_c >$=0.7647 σ =0.2140

data, so we can employ MDL to select the optimal model for pulse data. Afterward we apply the surrogate data method to the residual of the optimal model (i.e. the prediction error). It estimates correlation dimension for this prediction error and its surrogates under the given hypothesis (NH0): the prediction error is consistent with i.i.d noise. According to results of all volunteers, we cannot reject that the prediction error is i.i.d noise. We, therefore, conclude that with the test statistic at our disposal, pulse measurements on the fingertip and wrist are indistinguishable.

For comparison we repeat the experiment to the same pulse data of the fingertip with the addition of observational "noise" (the data of the Rössler system) under the same hypothesis. In this experiment we reject the hypothesis that the prediction

error is i.i.d noise. This result implies that once deterministic deviation exists in the pulse data on the fingertip (in other words, there is significant difference between the pulse data on the wrist and finger), our techniques also can detect the deterministic deviation.

Results of the above two experiments indicate that there is no significant difference between pulse waveforms measured on the lateral artery (wrist) and the fingertip. Although the conclusion may appear somewhat straightforward we originally proposed these methods to examine the relationship between one simple cardiac diagnosis in modern medicine and "feeling the pulse [3]" in traditional Chinese medicine. In the following chapter we extend to examine not only nonlinear dynamics of pulse data but also that of ECG data, and further investigate the relationship between them.

In addition, the combination of neural networks, minimum description length, and the surrogate data method is the main methodological contribution of the current work. We feel that this technique is important and will be applicable to a wide variety of real world data. We choose to illustrate this method with a physiological system (in this and next chapter).

[3] As feeling the pulse, traditional Chinese medicine practitioners usually put three fingers (like three sensors) on the patient's wrist. The diagnosis procedure is more complicated than our measurement on pulse data but they do eventually feel the blood propagation pressure.

4. INVESTIGATION OF DETERMINISTIC CHAOS IN HUMAN CARDIAC TIME SERIES

4.1 Introduction

Whether or not the human cardiac system is chaotic has long been a subject of interest in the application of nonlinear time series analysis. Techniques developed from the domains of nonlinear dynamics have been applied to study cardiac systems. Notably, estimation of dynamic invariants from time series has been employed with varying success to characterize the dynamics of a wide variety of electrocardiogram signals [52] [53] [54]. This reveals that the ECG has a finite noninteger correlation dimension and positive Lyapunov exponent. But these invariants, such as saturation of correlation or the existence of positive Lyapunov exponent, alone do not offer sufficient evidence to confirm the presence of deterministic chaos [55].

More significantly, statistical tests for the presence of determinism in ECG data have been proposed in recent years. Govindan et al. applied the surrogate data method and Lyapunov exponent to measure ECG data for several normal and pathological cases [55]. They suggested that both the correlation dimension calculated by Grassberger-Procaccia (GP) algorithm [49] and Lyapunov exponent [56] should be treated with suspicion. However, they still adopted them and selected correlation dimension as the test statistic. In [55] results are only shown for the single data set and one typical surrogate (i.e. no distribution of surrogates is provided). Small et al. [57] adopted the surrogate data method to show that human ECG recording during normal rhythm, ventricular tachycardia (VT) and ventricular fibrillation (VF) are not linear process. Zhang et al applied complexity in conjunction with the surrogate data method to VT and VF signals collected from dogs [58].

However, they did not emphasize statistical hypothesis testing and they presented the various transformation of complexity to these abnormal ECG data under the assumption that they were chaotic. A number of papers have also studied characteristics of VF in various animal models. But whether VF is chaotic or not is still in debate [59] and almost all VF signals used were from animals since acquisition of human VF data was quite difficult [38].

Meanwhile, relatively few attempts to determine the presence of determinism in blood pressure propagation (pulse data) are found in the literature. Eyal et al. have investigated the nonlinear properties of the blood pressure signals of hypertensive rats [60]. The relative lack of studies of blood pressure signals is possibly due to the fact that measuring pulse data ("feeling the pulse") has not been accepted in Western medicine. Up to the present pulse measurement is still widely used in modern medicine. But pulse measurement in either western medicine or modern medicine is not as important as pulse measurement (i.e., so-called feeling the pulse) in traditional Chinese medicine. It has been routine during diagnosis in traditional Chinese medicine for centuries [61]. It is therefore also important to establish whether the activity of pulse pressure propagation observed via pulse data is deterministic process (and possibly confirm to the deterministic origin of ECG data).

Most results reported in the literature are based on the surrogate data method which tests an observed ECG time series against the hypotheses of: 1) independent and identically distributed noise; 2) linearly filtered noise; and 3) a monotonic nonlinear transformation of linearly filtered noise. The three hypotheses are all some forms of the linear noise process (despite a possible static nonlinear filter). One can confidently reject the three hypotheses that the ECG data are linear noise. But one cannot make a decision on whether the ECG and pulse data are periodic waveforms with uncorrelated noise, which are consistent with the deterministic process. Actually ECG and pulse data usually exhibit strong periodicities and definite nonlinearity (such as observable QRS complex) during sinus rhythm. So it is natural to not only study whether the ECG and pulse signals of the healthy human are

deterministic or consistent with a stochastic process, but to examine whether they are noise driven periodic orbits or even pseudo-periodic deterministic chaos.

In the current work we apply a new surrogate data method, pseudo-periodic surrogate with the hypothesis of a periodic orbit with uncorrelated noise to both complete ECG and pulse data. This method has been outlined in [47] [62], where the application of the PPS method is limited to a single case. Furthermore, correlation dimension estimations are not applicable to field measurement since about 2% of noise in the data can destroy all nontrivial self-similarity, even if they come from a deterministic source plus observational noise [63]. For uncorrelated noise the correlation dimension estimated from the GP algorithm converges to a constant value [64]. In particular, ECG signals that are measured with surface electrodes are usually contaminated with noise. If filters in Fourier space are used to eliminate the noise, this will also smooth the signal and alter the structure of the QRS complex. So one option is to employ techniques of nonlinear noise reduction, like phase space projection [65] to remove noise. Another significant option, which we use in this work, is to replace correlation dimension with a more robust test statistic, algorithmic complexity. Algorithmic complexity aims to measure the regularity of the finite specified sequence, which can be employed to search for determinism in otherwise apparently random data. It has been successfully used in studies of electroencephalogram (EEG) [66]. In an application with low SNR Zhao and coworkers found that complexity is robust to the noisy signals, and very sensitive to the intrinsic deterministic dynamics [67]. In addition, algorithmic complexity has the great advantage of small computational cost and is well suited for real-time implementation.

Finally, we employ neural networks with adequate generalization, which are determined by the methods proposed and validated in the previous chapters (Chapter 2 and 3), to perform the one-step prediction between ECG and pulse data. This aims to measure the predictability from ECG data to pulse data and vice versa and confirm that ECG can determine pulse data. Application of transfer entropy to ECG and pulse data also indicates a stronger information flow from ECG to pulse

signal than vice versa.

4.2 The Pseudo-periodic Surrogate Data Method

The surrogate data method described in Section 3.2, which identifies an observed time series against three common kinds of hypotheses, does not provide sufficient evidence to confirm the existence of deterministic chaotic dynamics in cardiac time series. These methods fail to exclude all but the most trivial hypothesis of linear noise. Moreover, the two surrogate generation algorithms, Algorithm 1 and Algorithm 2, are hampered by technical issues related to the Fourier transformation [45]. Hence, the common surrogate techniques have a very limited utility when applied to a time series with a strong pseudo-periodic behavior. We present a recently suggested algorithm to determine whether these signals are consistent with periodic orbits driven by uncorrelated noise. Certainly we cannot exclude all other alternatives but our test is certainly stronger than those applied previously.

4.2.1 The PPS Algorithm to Generate Surrogate Data Sets

The PPS algorithm provides an entirely new way to generate surrogates. This algorithm can be described as follows [47]:

1. Let $\{x(i)\}_{i=1}^{N}$be a scaler time series of N points. Referring to embedding dimension d_e and delay time τ we obtain the phase space reconstruction of this time series [72]: $y(i) = (x(i), x(i-\tau), x(i-2\tau), ..., x(i-(d_e-1)\tau))$, where $i = (d_e-1)\tau + 1, ..., N$. Let us define$\{y(i)\}_{i=1}^{\check{N}}(\check{N} = N - (d_e-1)\tau)$ for the brevity.

2. Set $t = 1$ and randomly choose an initial condition $s(1)$, where $s(1) \in \{y(i), i = 1, ..., \check{N}\}$.

3. We then choose a neighbor of $s(t)$ $(t = 1)$, $s(r) \in \{y(i), i = 1, ..., \check{N}\}$ with the probability $\text{Prob}(s(r) = y(i)) \propto \exp\frac{-\|y(i)-s(t)\|}{\rho}$. ρ is the noise radius,

which determines the deviation of the neighbor $s(r)$ from the original one $s(t)$. According to the set probability, we conversely deduce the value of the noise radius.

4. Set $s(t+1) = s(r+1)$ and increase t. Repeat the procedure from Step 3 until $i = N$.

The vector time series $\{s(t), t = 1, ..., N\}$ is a stochastic trajectory on the attractor approximated by $\{y(i)\}_{i=1}^{\check{N}}$. It, therefore, follows approximately the same vector field as the data, but is contaminated with dynamic noise [47]. When the noise radius increases gradually, the dynamic noise introduced by the PPS algorithm will obliterate the fine dynamics and then the strong dynamics. For example, chaotic time series, such as the chaotic Rössler data and chaotic Chua circuit data, can be destroyed by the small noise added by the PPS method while the periodic orbits have to be obliterated by the larger noise. Different dynamics which exists in chaotic and periodic data lead to distinct trends of their surrogates produced by the PPS method with increasing noise radius. Finally, for either time series the surrogates generated by the PPS algorithm with much larger noise radius are just random noise. Consequently, the trends of surrogates for chaotic and periodic time series, generated by the PPS algorithm, are distinguishable.

The PPS algorithm has three parameters: the embedding dimension d_e, delay time τ, and the noise level ρ. The embedding dimension and embedding lag are easily obtained by the False nearest neighbors (FNN) algorithm [32] and second order autocorrelation (SOAC) [73]. The most important point is the right selection of the noise level regarding to the given hypothesis of periodic orbits with uncorrelated noise. If the value of ρ is too large the generated surrogates are simply temporally uncorrelated random data; if the ρ is greatly small the surrogates are almost identical to the part of the original data. With a moderate noise radius periodic dynamics can be preserved and the generated surrogates are consistent with periodic orbits with uncorrelated noise. In [47] it was determined by the maximal number of short segments in a representative surrogate that are identical to the data for n successive data points. This criterion has to introduce a new parameter, n, and how to select

the value of n is also open to the problem. So the problem of the PPS algorithm is that one cannot exactly estimate the noise level regarding to the given hypothesis. We just provide a reasonable scope (e.g. $0.3 \sim 0.7$), which is applicable to produce surrogates consistent with this given hypothesis. Hence, in the current work we more emphasize the trends of ECG and pulse data, indicated by the PPS method with probabilities from 0 to 1.0 than results obtained by the PPS method with a certain probability.

To apply the surrogate method we must select an appropriate statistic. The usual statistics, Lyapunov exponent and correlation dimension, are rather sensitive to noise or include many parameters that affect not only the computational speed but also the veracity of results. Here we employ *normalized* complexity [74] as the criterion to distinguish between the original data and surrogates. For the purposes of comparison with previous published results ([47] and [62]) we also apply the correlation dimension to the surrogates and original data.

4.2.2 The Algorithm of Complexity

A sequence S of length n is fully specified by $S = (s_1, s_2, ..., s_n)$ where each s_i is one of d symbols, $s_i \in B = \{b_1, b_2, ..., b_d\}$. For example, for the binary case $B = \{0, 1\}$, that is, S only comprises zeros and ones. For the general case, B is an alphabet of d ($d \geq 2$) symbols. Let $c(n)$ be the counter of the novel sub-sequences in the sequence S; P and Q denote two sequences which are substrings of S; PQ is the concatenation of P and Q; and $PQ-$ means that the last digit of the concatenation of P and Q has been deleted. Let $v(PQ-)$ denote the set of all the substrings of $PQ-$.

The procedure to compute algorithmic complexity of the sequence $S = \{s_i\}_{i=1,2,...,n}$ where $s_i \in B$ can be described as follows:

Firstly initialize $c(n) = 1$, $P = s_1$, and, $Q = s_2$. So $PQ- = s_1$. If $Q \in v(PQ-)$, leave P unchanged and update $Q = s_2 s_3$; if $Q \notin v(PQ-)$, add one to $c(n)$, update $P = s_1 s_2$, and $Q = s_3$.

Continue the previous step. Let us assume $P = s_1 s_2 ... s_j$ and $Q = s_{j+1}$. If $Q \in v(PQ-)$, leave P unchanged and update $Q = s_{j+1}s_{j+2}$, and then check whether Q belongs to $v(PQ-)$. Repeat the previous procedure, until $Q \notin v(PQ-)$. So $c(n) = c(n) + 1$. If $Q = s_{j+1}s_{j+2}...s_{j+k}$ at this time, and then P is updated to $P = s_1 s_2 ... s_j s_{j+1} s_{j+2} ... s_{j+k}$ whereas Q is updated to s_{j+k+1}. Thus complete the calculation of complexity of $c(n)$ of the sequence S until Q reaches the last string of $S = \{s_i\}_{i=1,2,...,n}$.

Lempel and Ziv [74] have demonstrated that for a sequence of length n consisting of d symbols

$$c(n) < \frac{n}{(1-2\frac{(1+\log_d \log_d(dn))}{\log_d(n)})\log_d(n)}$$

Since $\frac{(1+\log_d \log_d(dn))}{\log_d(n)} \to 0$ as n$\to \infty$ we deduce that $c(n)$ is bounded above by $\frac{n \log d}{\log n}$ and complexity of a random sequence of length n with an alphabet of d symbols is precisely $\frac{n \log d}{\log n}$. Therefore, it is usually more useful to define the *normalized* complexity as

$$C(n) = \frac{c(n)}{n} \log_d(n), \tag{4.1}$$

which is between zero and one (for a random sequence we expect that its *normalized* complexity is approximately one). In what follows, when we describe complexity, we mean the *normalized* complexity (Eq. 4.1).

Consequently, the *normalized* complexity only has two parameters, the length of the sequence n and the number of symbols d of which the sequence is composed. We found the sensitivity of complexity to the data did not significantly change for different d and n.

To compute complexity of an experimental time series $\{x_i\}$ we need to adopt some encoding scheme f to convert it to the sequence. For example, in the binary case $B = \{0, 1\}$ the time series is converted to the sequence of zeros and ones. The standard encoding method is to partition the observed data into d bins in terms of either equal size or equal probability. In this research the observed data is partitioned into three symbols 0, 1, and 2, where the probability of each symbol

occurring is constrained to be equal (i.e. 1/3). Fig. 4.1 illustrates the ECG data and its partitioned symbols.

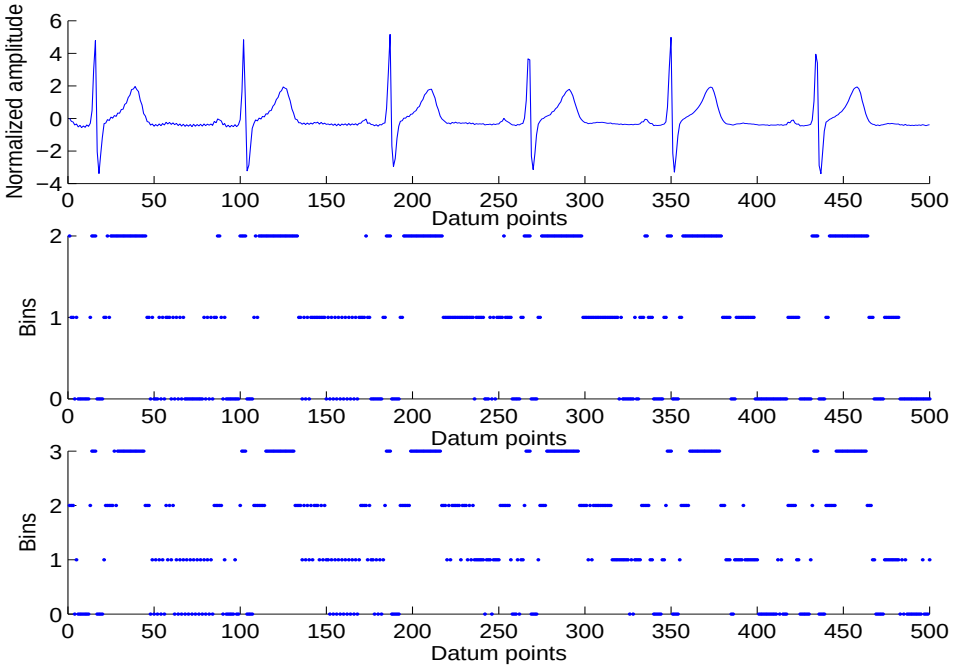


Fig. 4.1: The original ECG data (top panel) is partitioned into 3 bins (middle panel) and 4 bins (bottom panel). In both cases each bin holds the same probability.

Complexity may be variant for normal cardiac data of different subjects. So are other test statistics, such as correlation dimension estimations for different individual normal ECG data. Nevertheless, complexity of such data is distinct from complexity of random noise and periodic data. All the hypotheses employed in this work are random noise (independently distributed noise) and periodic orbits with uncorrelated noise. Hence, complexity is able to determine whether ECG or pulse data satisfies with these given hypotheses.

4.3 Evidence for Deterministic Chaos in Human Cardiac Data

In this section we demonstrate the results obtained by applying the surrogate data method to human cardiac data. Section 4.3.1 presents the application of the PPS method to both ECG and pulse data. To compare the performance of complexity as

the test statistic, we also employ correlation dimension calculated by the Gaussian Kernel Algorithm [51] as an alternative in section 4.3.2. Unlike the GP algorithm, the GKA can model the underlying attractor as a deterministic time series with added noise. In section 4.3.3 we apply the cycle shuffled surrogate data method to test the same experimental data as a comparison. The cycled shuffled surrogate data method is an optional approach to test the data against the hypothesis of periodic determinism.

4.3.1 Application of the PPS Method

Acquisition of ECG and pulse data of healthy subjects (2 female and 5 male volunteers) was done in the morning. The subjects lay supine on a bed in a quiet and relaxed situation to minimize internal and external influences. None of them was receiving any form of medication. The measurement device was PowerLab 4/25 of ADInstruments. The sampling rate is set to 100 Hz and resolution is 16 bits. Note that we collected each volunteer's surface ECG data and pulse data on the fingertip simultaneously.

Due to the fact that the amplitude of ECG data (μV) is greatly different from the amplitude of pulse data (mV) we normalized both kinds of data. We employ the PPS method with the probability of $0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8$, and 0.9 to generate 50 surrogates for ECG (1000 points) and pulse data (1000 points) of each volunteer, and then calculate complexity of the data and surrogates. Fig. 4.2 presents typical surrogates of the ECG and pulse data used in this section, which are generated by the PPS with relatively small probability. We aim to observe the relationship between the data and surrogates for certain probability (noise level) and pay more attention to the changing trends of complexity of surrogates. Typical results of one volunteer's ECG and pulse data are presented in Fig. 4.3.

The surrogate constructed with the smaller probability should be more similar to the original time series than the surrogates constructed with larger probabilities. However, the surrogate constructed with the very small value of probability usually contains the repeated waveform because such surrogate data infrequently jump

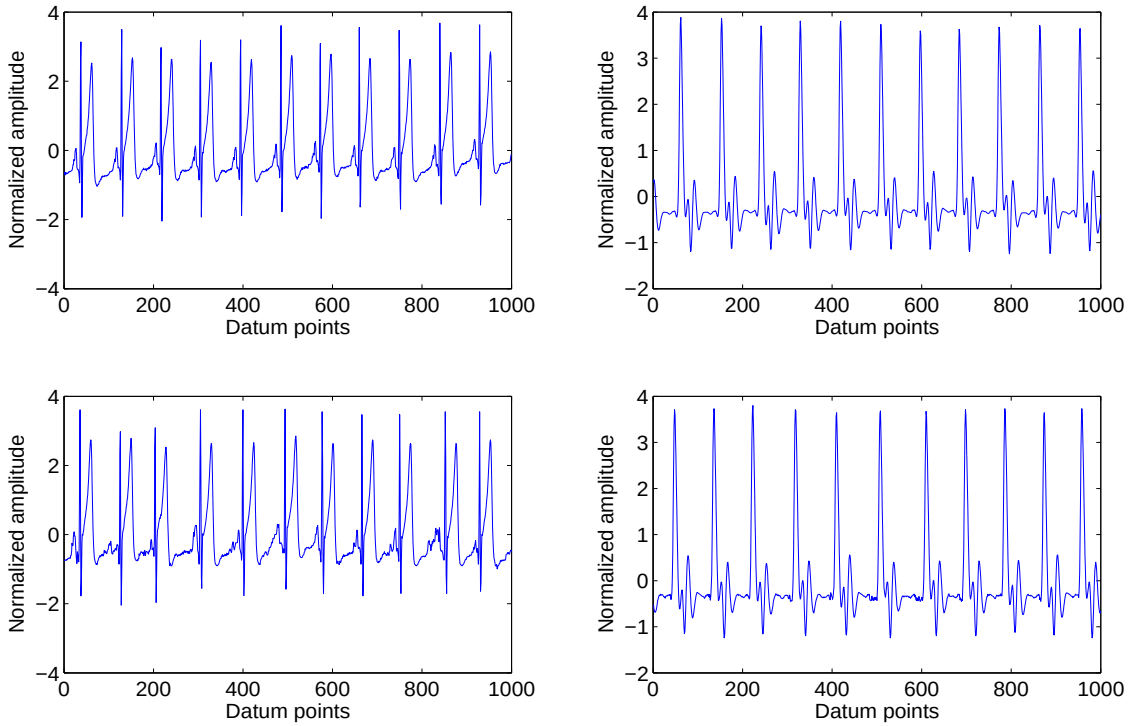
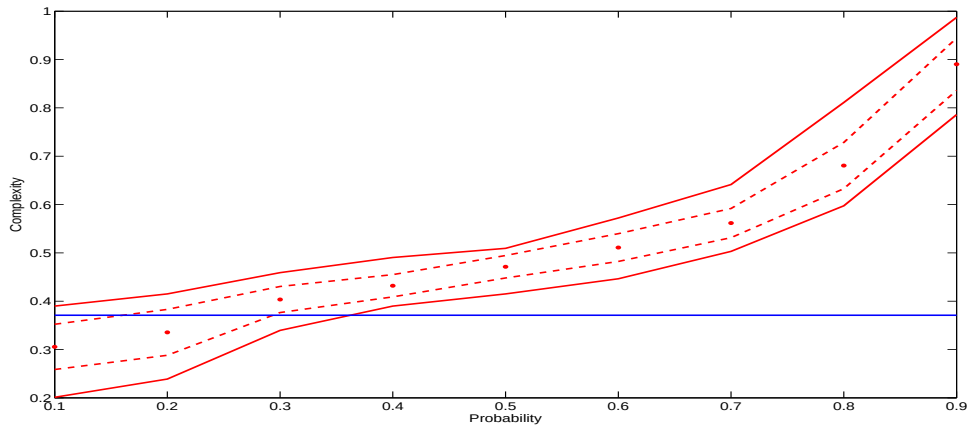
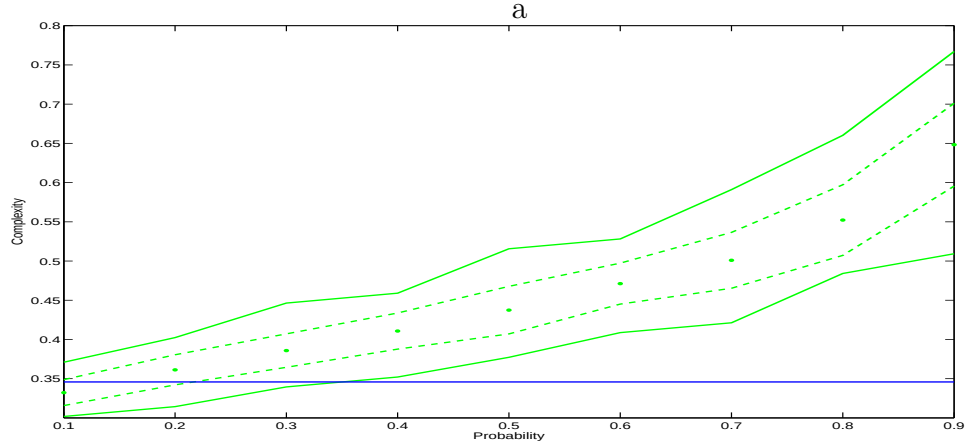


Fig. 4.2: ECG and pulse data and their representative surrogates. The left and right panels in top depict representative ECG and pulse recording respectively. The two panels in bottom show their PPS surrogates respectively. Qualitatively the data and surrogates are indistinguishable.

from one trajectory to another in the phase space (i.e. they may follow a complete trajectory and then repeat the part of this trajectory). The surrogate, therefore, constructed with the very small probability contains long-term periodicity. So for both ECG and pulse data the mean complexity of surrogates generated by the PPS algorithm with the probability of 0.1 is lower than complexity of the original one. With increasing noise radius (i.e. increasing probability) the surrogate data more frequently jumps among different trajectories and less repeats itself. Complexity of surrogates with the probability of 0.2 or 0.3 is closer to complexity of the original data since these surrogates reproduce all the intercycle dynamics in the data. Certainly, if the probability (noise radius) is too large the surrogates are equivalent to the i.i.d noise. The moderate range of probability in regard to the hypothesis of periodic orbits in our experiments is between 0.3 and 0.7. In Fig. 4.3a and 4.3b



a



b

Fig. 4.3: (a) The straight line is complexity of the ECG data; the dots are the mean complexity of 50 surrogates generated by the PPS with the probability from 0.1 to 0.9; two dashed lines denote the mean plus one standard deviation (the upper line) and minus one standard deviation (the lower line); two solid lines are the maximum and minimum complexity among the 50 surrogates. (b) Complexity of the pulse data. Properties of the lines and markers are defined the same as (a).

when the probability is larger than 0.3, the complexities of the data begin to be far from the distribution of the surrogates. That is, some fine dynamics other than the periodic structures have been destroyed by the PPS algorithm with the noise radius (indicated by the probability of 0.4). The generated surrogates at this noise level contain periodic structures and the added dynamic noise. We therefore reject this given hypothesis and conclude that the ECG and pulse data are not strictly periodic

determinism with dynamics noise. This result indicates that pseudo-periodic deterministic chaos may exist in both ECG and pulse data. When the noise level continue increasing the periodic structures are then obliterated by larger added noise. Certainly the subsequent generated surrogates are not consistent with periodic orbits with uncorrelated noise so the given hypothesis of periodic orbits with uncorrelated noise is not applicable any more to these surrogates.

Tab. 4.1: Results of application of the PPS method to ECG data for all subjects. c in each row is complexity of ECG data; $< c >$ is the mean complexity of 50 surrogates for each case; σ is the standard deviation; c_{min} and c_{max} represent the minimum and maximum complexity among these surrogates.

		Probability								
		0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
Subject 1 *c=0.378*	<c>	0.375	0.410	0.424	**0.451**	**0.472**	**0.514**	**0.547**	**0.669**	**0.861**
	σ	0.023	0.021	0.028	**0.020**	**0.025**	**0.026**	**0.031**	**0.049**	**0.059**
	c_{max}	0.421	0.446	0.477	**0.515**	**0.547**	**0.578**	**0.610**	**0.767**	**0.981**
	c_{min}	0.316	0.358	0.365	**0.410**	**0.414**	**0.465**	**0.478**	**0.572**	**0.750**
Subject 2 *c=0.369*	<c>	0.314	0.339	0.393	**0.442**	**0.470**	**0.506**	**0.572**	**0.674**	**0.883**
	σ	0.044	0.050	0.036	**0.022**	**0.026**	**0.026**	**0.035**	**0.048**	**0.055**
	c_{max}	0.382	0.428	0.459	**0.504**	**0.528**	**0.560**	**0.660**	**0.811**	**0.968**
	c_{min}	0.207	0.239	0.301	**0.396**	**0.402**	**0.446**	**0.522**	**0.591**	**0.750**
Subject 3 *c=0.421*	<c>	0.378	0.417	0.455	**0.496**	**0.528**	**0.571**	**0.642**	**0.792**	**0.872**
	σ	0.026	0.025	0.022	**0.025**	**0.024**	**0.024**	**0.037**	**0.038**	**0.045**
	c_{max}	0.422	0.465	0.497	**0.553**	**0.591**	**0.622**	**0.748**	**0.868**	**0.962**
	c_{min}	0.314	0.365	0.396	**0.446**	**0.484**	**0.516**	**0.578**	**0.679**	**0.736**
Subject 4 *c=0.434*	<c>	0.380	0.415	0.456	**0.506**	**0.546**	**0.585**	**0.613**	**0.665**	**0.794**
	σ	0.027	0.033	0.036	**0.029**	**0.029**	**0.026**	**0.030**	**0.027**	**0.029**
	c_{max}	0.438	0.516	0.553	**0.560**	**0.622**	**0.660**	**0.717**	**0.742**	**0.843**
	c_{min}	0.337	0.340	0.385	**0.446**	**0.490**	**0.528**	**0.534**	**0.616**	**0.736**
Subject 5 *c=0.415*	<c>	0.371	0.413	0.446	0.471	**0.505**	**0.528**	**0.566**	**0.669**	**0.847**
	σ	0.023	0.024	0.022	0.022	**0.019**	**0.018**	**0.025**	**0.042**	**0.059**
	c_{max}	0.416	0.453	0.484	0.522	**0.553**	**0.572**	**0.610**	**0.780**	**0.956**
	c_{min}	0.296	0.346	0.383	0.412	**0.459**	**0.484**	**0.478**	**0.578**	**0.698**
Subject 6 *c=0.462*	<c>	0.446	0.475	0.500	**0.523**	**0.555**	**0.578**	**0.599**	**0.635**	**0.801**
	σ	0.025	0.024	0.022	**0.025**	**0.023**	**0.023**	**0.032**	**0.024**	**0.037**
	c_{max}	0.490	0.534	0.547	**0.578**	**0.604**	**0.635**	**0.666**	**0.685**	**0.880**
	c_{min}	0.396	0.427	0.446	**0.466**	**0.509**	**0.484**	**0.522**	**0.591**	**0.723**
Subject 7 *c=0.458*	<c>	0.434	0.470	0.492	**0.514**	**0.535**	**0.568**	**0.619**	**0.626**	**0.791**
	σ	0.029	0.024	0.028	**0.023**	**0.024**	**0.027**	**0.038**	**0.043**	**0.047**
	c_{max}	0.471	0.522	0.564	**0.560**	**0.594**	**0.654**	**0.689**	**0.783**	**0.874**
	c_{min}	0.365	0.409	0.434	**0.477**	**0.503**	**0.516**	**0.515**	**0.567**	**0.653**

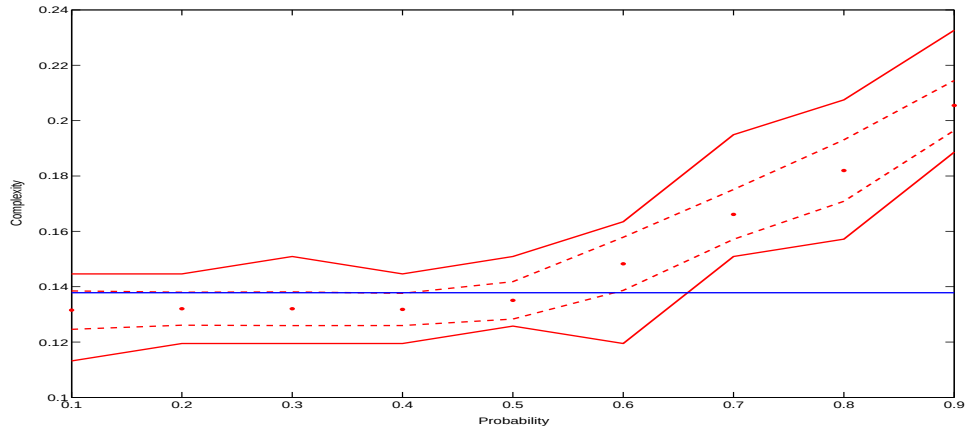
Tab. 4.2: Results of application of the PPS method to pulse data for all subjects. c in each row is complexity of pulse data; $< c >$ is the mean complexity of 50 surrogates for each case; σ is the standard deviation of them; c_{min} and c_{max} represent the minimum and maximum complexity among these surrogates.

		Probability								
		0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
Subject 1 *c=0.365*	<c>	0.330	0.365	0.411	**0.455**	**0.480**	**0.495**	**0.528**	**0.575**	**0.673**
	σ	0.021	0.029	0.029	**0.032**	**0.039**	**0.037**	**0.038**	**0.053**	**0.067**
	c_{max}	0.370	0.452	0.509	**0.509**	**0.604**	**0.560**	**0.616**	**0.685**	**0.830**
	c_{min}	0.289	0.308	0.352	**0.390**	**0.415**	**0.390**	**0.453**	**0.460**	**0.553**
Subject 2 *c=0.346*	<c>	0.332	0.359	0.382	**0.416**	**0.436**	**0.467**	**0.509**	**0.558**	**0.658**
	σ	0.021	0.018	0.021	**0.022**	**0.022**	**0.030**	**0.039**	**0.052**	**0.060**
	c_{max}	0.384	0.402	0.421	**0.478**	**0.484**	**0.534**	**0.622**	**0.710**	**0.817**
	c_{min}	0.296	0.314	0.321	**0.365**	**0.384**	**0.402**	**0.434**	**0.465**	**0.541**
Subject 3 *c=0.333*	<c>	0.297	0.317	0.346	**0.373**	**0.400**	**0.436**	**0.476**	**0.527**	**0.675**
	σ	0.020	0.019	0.018	**0.016**	**0.021**	**0.021**	**0.021**	**0.036**	**0.046**
	c_{max}	0.333	0.352	0.383	**0.409**	**0.459**	**0.490**	**0.578**	**0.641**	**0.830**
	c_{min}	0.245	0.217	0.283	**0.358**	**0.358**	**0.384**	**0.415**	**0.465**	**0.509**
Subject 4 *c=0.277*	<c>	0.263	0.306	0.333	**0.373**	**0.401**	**0.434**	**0.465**	**0.490**	**0.518**
	σ	0.021	0.018	0.019	**0.020**	**0.020**	**0.027**	**0.018**	**0.024**	**0.025**
	c_{max}	0.314	0.358	0.371	**0.408**	**0.452**	**0.484**	**0.503**	**0.566**	**0.597**
	c_{min}	0.201	0.258	0.270	**0.321**	**0.365**	**0.352**	**0.434**	**0.440**	**0.465**
Subject 5 *c=0.265*	<c>	0.238	0.267	0.293	**0.321**	**0.359**	**0.396**	**0.431**	**0.481**	**0.648**
	σ	0.026	0.014	0.015	**0.019**	**0.022**	**0.023**	**0.032**	**0.032**	**0.071**
	c_{max}	0.289	0.296	0.321	**0.364**	**0.402**	**0.453**	**0.509**	**0.565**	**0.823**
	c_{min}	0.182	0.233	0.253	**0.282**	**0.295**	**0.346**	**0.358**	**0.427**	**0.515**
Subject 6 *c=0.300*	<c>	0.290	0.311	0.336	**0.372**	**0.404**	**0.442**	**0.504**	**0.599**	**0.698**
	σ	0.015	0.019	0.021	**0.022**	**0.025**	**0.041**	**0.042**	**0.061**	**0.070**
	c_{max}	0.333	0.352	0.364	**0.409**	**0.453**	**0.515**	**0.591**	**0.741**	**0.849**
	c_{min}	0.251	0.264	0.270	**0.314**	**0.352**	**0.328**	**0.409**	**0.472**	**0.541**
Subject 7 *c=0.354*	<c>	0.369	0.394	0.406	**0.439**	**0.457**	**0.514**	**0.546**	**0.590**	**0.699**
	σ	0.024	0.026	0.026	**0.029**	**0.031**	**0.048**	**0.048**	**0.052**	**0.051**
	c_{max}	0.443	0.452	0.465	**0.490**	**0.528**	**0.610**	**0.685**	**0.723**	**0.873**
	c_{min}	0.314	0.327	0.342	**0.389**	**0.409**	**0.415**	**0.446**	**0.503**	**0.572**

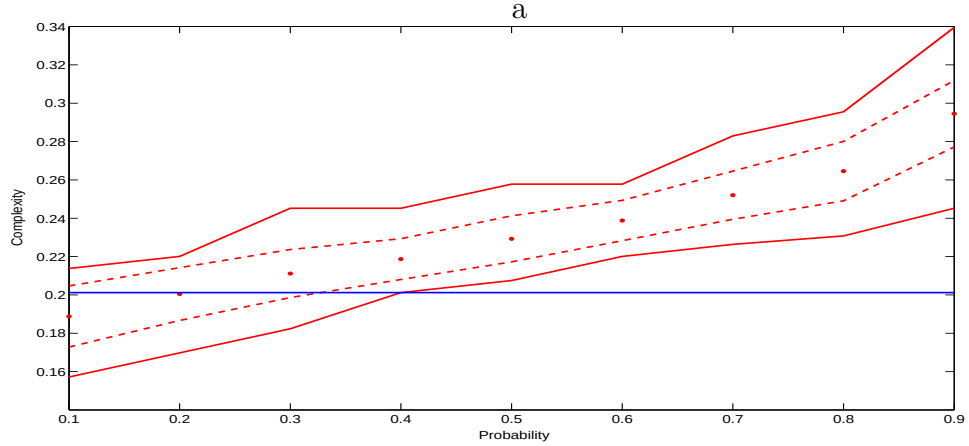
Tables 4.1 and 4.2 summarize results of application of the PPS method to all the seven subjects' ECG and pulse data. In both tables the bold font in each row means that at this probability or higher complexity of one subject's data is distinguished from those of the corresponding PPS surrogates. Note that for all the subjects the mean complexity of surrogates is lower than complexity of the original data in the two tables. Consequently, the results obtained in this section are representative and reproducible. Another significant result is that for all cases complexity of ECG data is higher than that of pulse data of the same subject. This is consistent with our later conclusion in Section 4.4 that the ECG data comprises certain deterministic components which the pulse data can not replicate or does not contain. We will discuss this conclusion at length in Section 4.4.

By way of illustration we demonstrate the application of this algorithm to periodic sine, chaotic Rössler data, and periodic and chaotic Chua circuit data. The period of the sine data is the same as that of the preceding ECG data. We also normalized the four kinds of data. Results for these data sets are illustrated in Fig. 4.4 and 4.5.

According to Fig. 4.3, 4.4, and 4.5, we observe that the trends of surrogates of ECG and pulse data are consistent with those of chaotic Rössler data and chaotic Chua data, but the trends of the periodic, approximately stable for the probability lower than 0.7, is not. In particular, for the probability of 0.5 complexity of the periodic is in the middle of the distribution of surrogates. We thus can not reject the hypothesis that they are a periodic orbit, as we expect. Consequently, the trends of the surrogates' complexity in Fig. 4.4 and 4.5, which change with the probabilities (noise levels), clearly reflect distinction of periodic and chaotic orbits and are the criterion to distinguish chaotic and periodic data. More significantly, the trends of ECG and pulse data indicate that human cardiac data is consistent with the deterministic chaos.



a

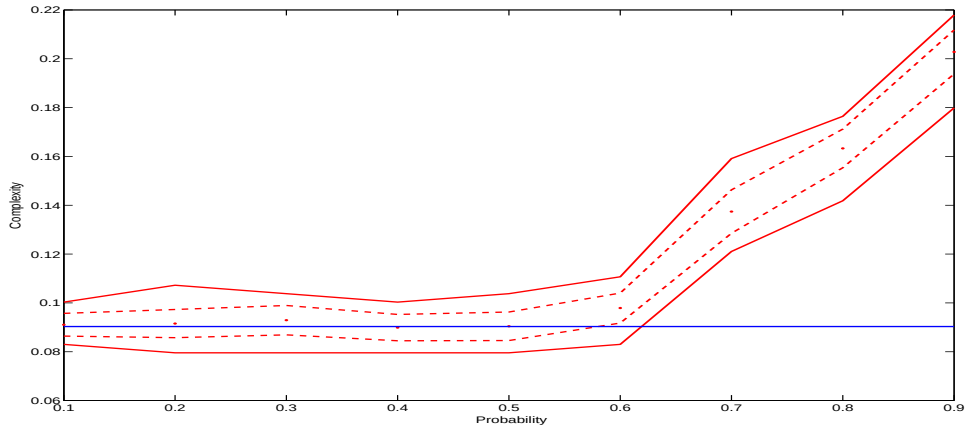


b

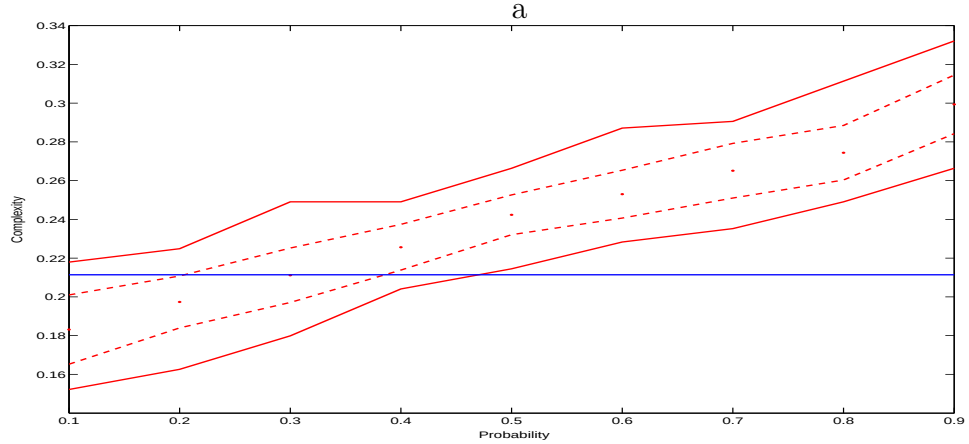
Fig. 4.4: (a) The straight line is complexity of the sine data; the dots are the mean complexity of 50 surrogates generated by the PPS with the probability from 0.1 to 0.9; two dashed lines denote the mean plus one standard deviation (the upper line) and minus one standard deviation (the lower line); two solid lines are the maximum and minimum complexity among the 50 surrogates. (b) Complexity of the chaotic Rössler data. Properties of the lines and markers are defined the same as (a).

4.3.2 Application of Correlation Dimension as a Test Statistic

For comparison of the previous results [47] we apply the correlation dimension to the preceding ECG and pulse data and their surrogates. The selected probability of PPS algorithm is 0.4, which means that most surrogates are qualitatively similar to the data. In both cases the hypothesis cannot be rejected with 100% probability, as shown in Fig. 4.6. The evidence from correlation dimension to reject the given



a

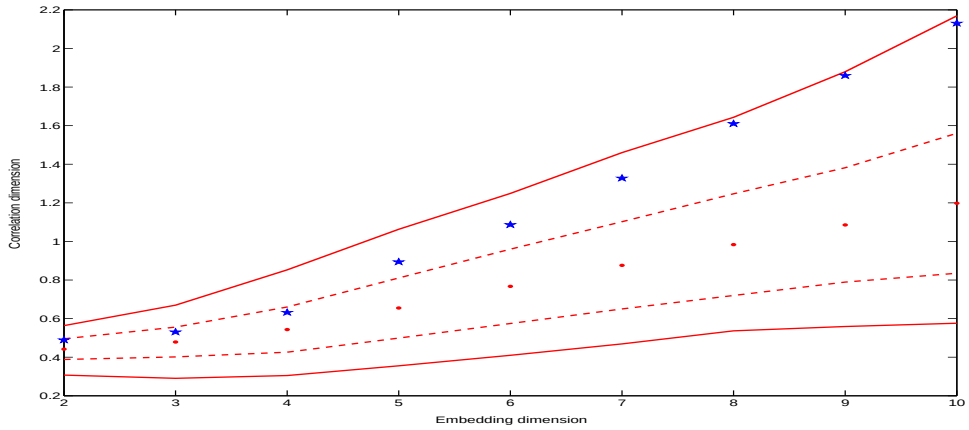


b

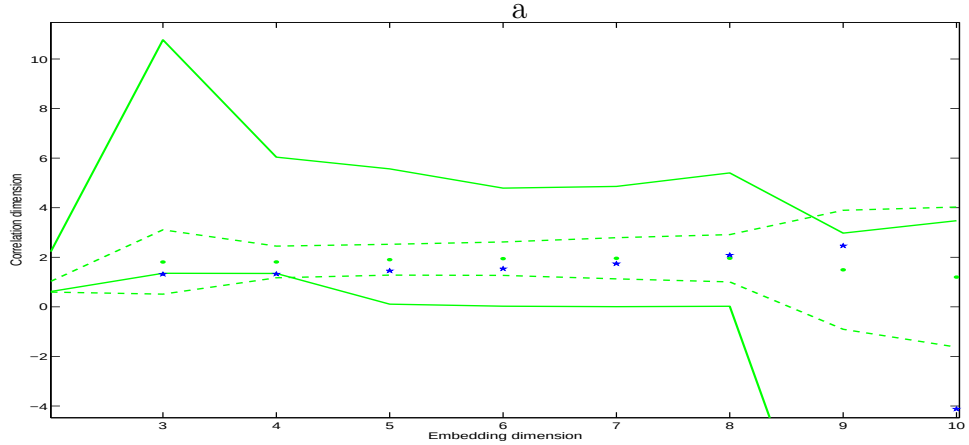
Fig. 4.5: (a) The straight line is complexity of the periodic Chua data; the dots are the mean complexity of 50 surrogates between probabilities from 0.1 to 0.9; two dashed lines denote the mean plus one standard deviation (the upper line) and minus one standard deviation (the lower line); two solid lines are the maximum and minimum algorithmic complexity among the 50 surrogates. (b) Complexity of the chaotic Chua data. Properties of the lines and markers are defined the same as (a).

hypothesis is weak.

This conflict does not mean that the results in Fig. 4.6 contradict those in [47] (which was limited to a single specific case). First of all, the ECG data themselves are incomparable. The two data sources (subjects) are different and measured under different situation. Secondly, how to choose the suitable probabilities (noise levels) completely depends on the personal estimation of maximal segments in the surro-



a



b

Fig. 4.6: (a) correlation dimension of the ECG data (stars) for embedding dimension from 2 to 10; the mean correlation dimension (dots) for 50 surrogates for embedding dimension from 2 to 10; the mean plus the standard deviation (the upper dashed line) and minus the standard deviation (the dashed lower line); the maximum and minimum correlation dimension (two solid lines) among these surrogates. (b) Correlation dimension of pulse data (stars) for embedding dimension from 2 to 10. Properties of lines and markers are the same as (a).

gates. Therefore, results obtained by the PPS algorithm with a single probability (noise level) may not be adequate to make a decision. Finally, the algorithms to calculate correlation dimension are different. But either correlation dimension is sensitive to parameters, which are required to calculate it.

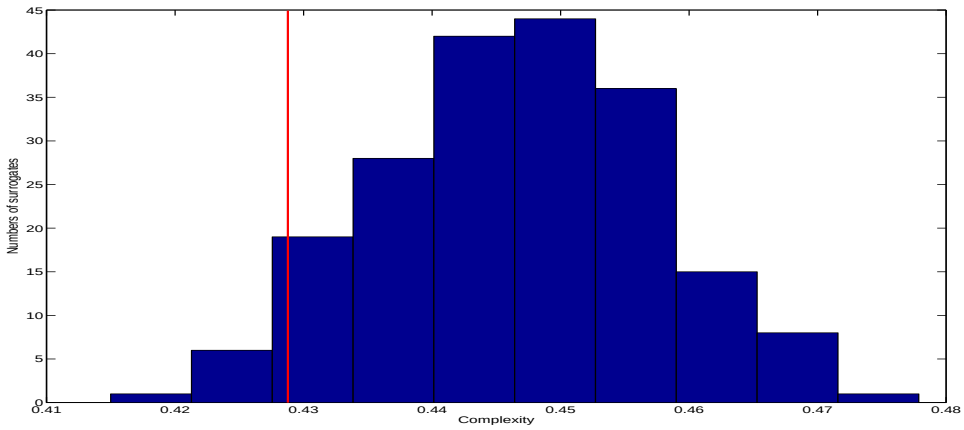
4.3.3 Application of the Cycle Shuffled Surrogate Data Method

An alternative surrogate generation technique, the cycle shuffled surrogate method, was proposed by Theiler [48] to test the hypothesis of periodic structures between cycles in a time series with strong periodic components. Instead of shuffling the individual data points in a time series, one shuffles the individual cycles. The surrogates generated by shuffling in this way should destroy any structures with a period longer than the cycle length. Theiler applied this technique to an epileptic electroencephalogram (EEG) [48].

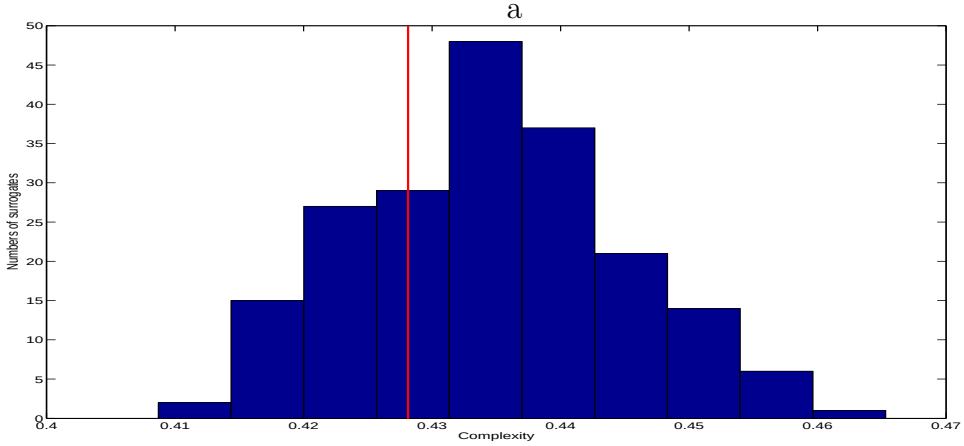
The application of the cycle shuffled surrogate method to ECG and pulse data is presented in Fig. 4.7. We break the cycles at peaks, which are convenient places to break the cycles. The data are the same ECG and pulse data (1000 points) used in Section 4.3.1 plus the successive 1000 point data, and the step size of the data is set to 2 so as to keep the same length of data to calculate as previous experiments and introduce more cycles.

Referring to Fig. 4.7 we fail to reject the hypothesis that ECG and pulse data are periodic signals. Actually the degree to which the cycle shuffled surrogate method randomizes the data depends on the number of cycles in the data. For this 1000 point data with double numbers of cycles (22 cycles) this approach still can not completely randomize the cycles, which makes the surrogates keep some deterministic structures. We observe that until the data is extended with 45 cycles this surrogate method can randomize the data to high degree. Fig. 4.8 illustrates the relation of ECG data and one typical surrogate. Pulse data and its surrogates are analogous to the contents in this figure.

This method suffers from two drawbacks. (1) One has to identify a convenient place to break the cycles, which inevitably produces discontinuities (at the reassembled points) or non-stationarity in surrogates. (2) The degree to which this surrogate method randomizes the data relies on the number of cycles in the data [75]. One can try to employ more data, if possible, but the longer data is more difficult to preserve the continuity and stationarity of surrogates. Even for the short data,



a



b

Fig. 4.7: (a) The distribution of correlation dimension for 200 cycle shuffled surrogates of the ECG data. The x-axis of the solid line represents correlation dimension of the ECG data. (b) The distribution of correlation dimension for 200 cycle shuffled surrogates of the pulse data. The x-axis of the solid line represents correlation dimension of the pulse data.

this is also possible. In Fig. 4.9 the surrogates of ECG and pulse data generated by the cycle shuffle algorithm are non-stationary and broken at some points. This technique thus appears to be limited to the specific time series that has convenient places to break the cycles and abundant cycles without non-stationarity, and therefore is not applicable to general time series. So the cycle shuffled surrogate method lacks generality.

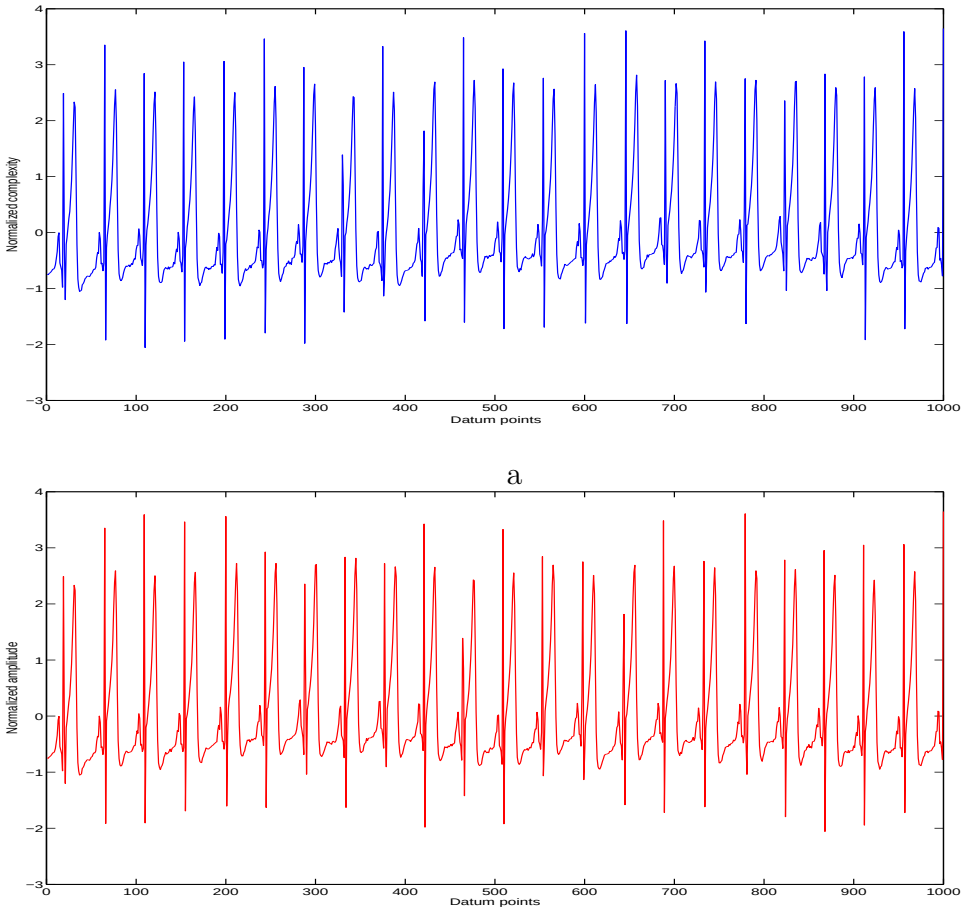


Fig. 4.8: The ECG data (a) and one typical cycle shuffled surrogate (b). The cycle shuffled method can not fully randomize the data with relatively few cycles.

4.4 Prediction Analysis from ECG to Pulse Data

For both ECG and pulse data, short-term prediction from one to the other would be impossible if one or both are independent or stochastic, but it is possible to make the prediction between them if they exhibit deterministic components and if these deterministic components are related. In addition, ECG measurement is often used in modern medicine while "feeling the pulse" is a routine during diagnosis in traditional Chinese medicine. By prediction between ECG and pulse data we can build the relation between them and further determine whether there is significant difference between ECG and pulse data other than the signal intensity or whether

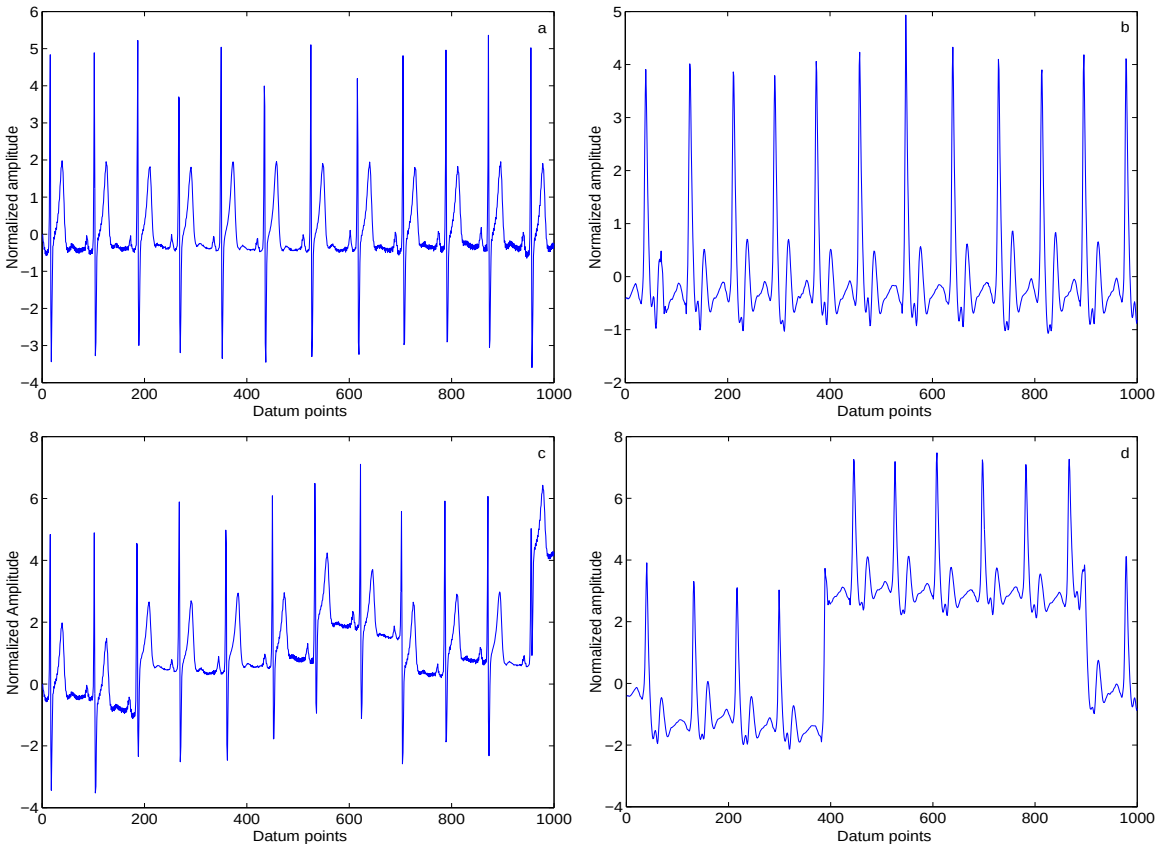


Fig. 4.9: The ECG data (a) and one of its cycle shuffled surrogates (c); the pulse data (b) and one of its cycle shuffled surrogates (d).

they can reflect equivalent symptoms. In the previous work we found that pulse measurement on the wrist and fingertip are equivalent. So the distance between them does not affect pulse propagation through the hand (or contributes little to this pulse propagation rather than pulse intensity). In this section we extend prediction analysis to model the relationship between ECG and pulse data, and we find that ECG can determine pulse data.

Neural networks with different numbers of neurons are used to make one-step prediction of pulse data from ECG. The first 3600 data points are used to train the neural networks with the remaining 800 points as the test set. The training algorithm is still the Levenberg-Marquardt training algorithm. After building models we employ the method of minimum description length [69] to determine the optimal model with adequate generalization. Fig. 4.10 presents the pulse data of Subject

one and its one-step prediction obtained by the optimal model. For all the seven subjects the test pulse data has high correlation with the corresponding prediction made by optimal neural networks, as illustrated in Table 4.3.

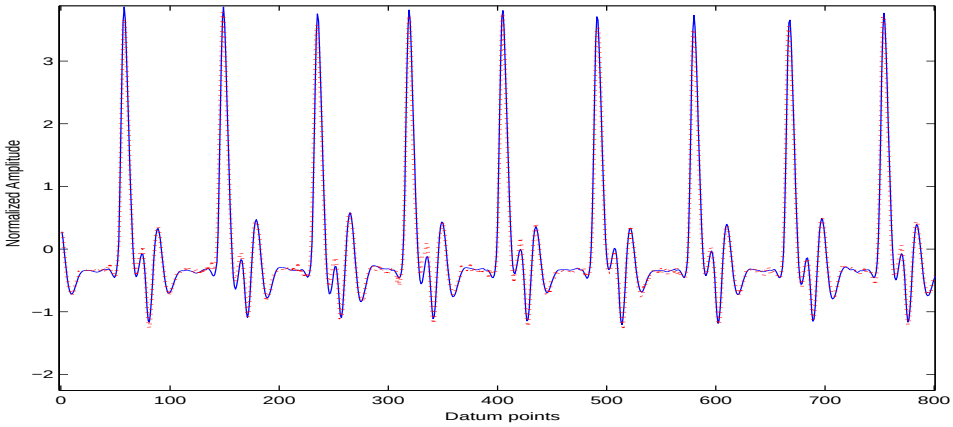


Fig. 4.10: The pulse data (solid line) and its one-step prediction (dotted line). Prediction from ECG data exactly follows the pulse waveform.

Tab. 4.3: Cross-correlate coefficient between the pulse data and corresponding prediction for all subjects.

Subject	Subject 1	Subject 2	Subject 3	Subject 4	Subject 5	Subject 6	Subject 7
Cross Coefficient	0.997	0.996	0.993	0.992	0.995	0.993	0.990

Finally, we apply surrogate data methods to test whether the prediction error is just random noise, i.e. to make sure the ECG data can exactly capture dynamics of the pulse data. Surrogate data methods we employ are the small shuffled surrogate data method [76] and the common surrogate data method with the given hypothesis of NH0. We apply complexity and correlation dimension as test statistics to each surrogate data method. The surrogate data method with the hypothesis of NH0 is used to test whether the data is consistent with i.i.d. noise while the small shuffled surrogate data method is designed to determine whether the data is consistent with independently distributed noise, i.e. not necessarily identically distributed noise. The small shuffled surrogate method does not globally randomize the data but does

so locally, that is, this method can control how far one data point is moved on average and preserve variable volatility.

In the residual, relatively large amplitudes periodically appear (e.g. corresponding to every peak of the pulse data). Obviously, this residual is not i.i.d. noise and the surrogate data method with the hypothesis of NH0 is not applicable to this residual, but we are not sure whether it is consistent with another random noise, the independently distributed noise. Results of the application of the two surrogate data methods on the model residual are illustrated in Fig. 4.11. In this figure Panel (a) and (b) demonstrate the application of the small shuffled surrogate data method to the prediction error; Panel (c) and (d) demonstrate the application of the surrogate data method with the null hypothesis of NH0 to the same error. The x-axis of the solid line in Fig. 4.11 (a) denotes complexity of the error. Referring to this figure, we confirm that both ECG and pulse data conform to deterministic processes and the ECG data can replicate the deterministic process of pulse data.

However, we still need to test the causal relationship between ECG and pulse data. Therefore, we repeat the above procedures to predict from pulse data to ECG data. But the prediction from pulse data cannot follow ECG data. Even if we try more training data, longer training iteration time, and different training algorithms, the prediction is still poor. That is because the model residuals are not just random, i.e. there are certain deterministic information in such residuals. Consequently, we conclude that the ECG data comprises certain deterministic components, which the pulse data can not replicate or does not contain.

4.5 Measurement of Transfer Entropy in Human Cardiac Data

Transfer entropy [77] is used to quantify the exchange of information between two systems for both directions. Let $X_i^{(k)}$ and $Y_i^{(k)}$ be two discrete random processes with k-dimension vector state $x_i^{(k)} = (x_i, ..., x_{i-k+1})$ and $y_i^{(k)} = (y_i, ..., y_{i-k+1})$ respectively.

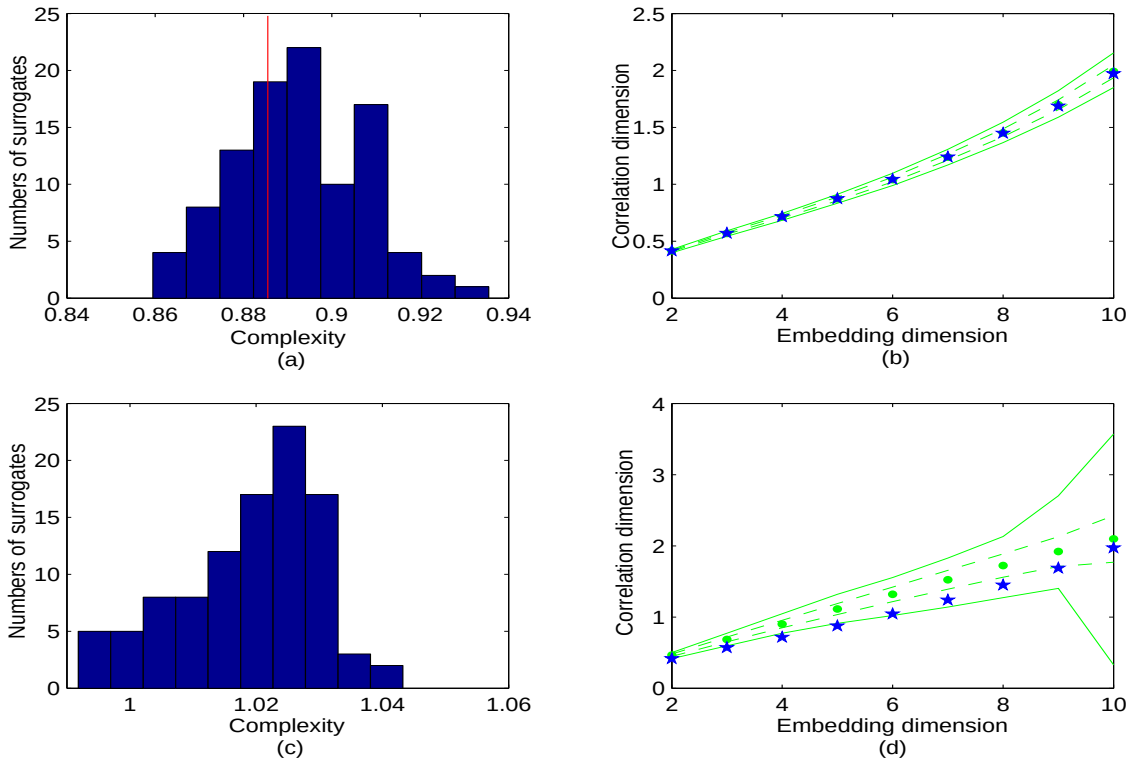


Fig. 4.11: (a) The histogram is the distribution of complexity for 100 small shuffled surrogates; (b) stars are correlation dimension of the same error for embedding dimension from 2 to 10. Properties of lines and markers in (b) and (d) are the same as the same as Fig. 4.6. (c) presents the distribution of complexity for 100 Algorithm 0 surrogates; (d) presents correlation dimension of this error and its 100 Algorithm 0 surrogates.

Transfer entropy from $X_i^{(k)}$ to $Y_i^{(k)}$ is defined as [78],

$$T_{X\to Y} = \sum_{x_{i+1},x_i^{(k)},y_j^{(l)}} p(x_{i+1}, x_i^{(k)}, y_j^{(l)}) \log \frac{p(x_{i+1}|x_i^{(k)}, y_j^{(l)})}{p(x_{i+1}|x_i^{(k)})} \tag{4.2}$$

where $p(\cdot)$ represents the probability distribution and $p(x_{i+1}|x_i^{(k)}, y_j^{(l)})$ is the transition probabilities.

Mutual information [79] is widely applied to quantity the overlapped information of two systems. However, the mutual information function is symmetric and fails to detect any directional information. In contrast to the mutual information, the transfer entropy is non-symmetric function and can identify the direction of

information exchange between two systems [77].

We thus measure transfer entropy between ECG and pulse data, as illustrated in Fig. 4.12. The transfer entropy indicates a stronger flow of information from ECG to pulse data than vice versa over a significant range of resolution. This result supports the conclusion taken from the preceding prediction analysis (Section 4.4).

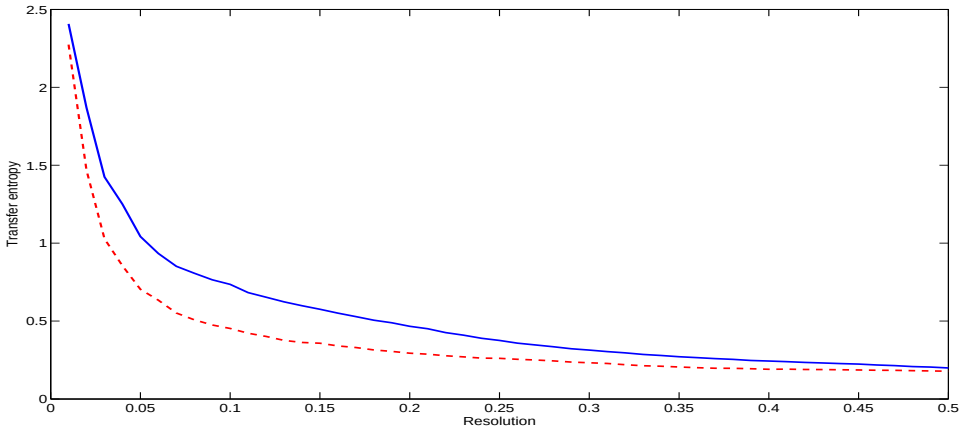


Fig. 4.12: Transfer entropy from ECG to pulse (solid line) and transfer entropy from pulse to ECG (dashed line) for a ten-minute cardiac time series of one subject.

Although they both stem directly from the human heart, ECG and pulse data are measured deterministically through different processes, which in turn depend on distinct influences. These dependency altogether constitutes the two different deterministic processes. The human ECG data may provide additional information of the human cardiovascular system that the pulse does not. Conversely, pulse measurement ("feeling the pulse" in TCM) does not contain additional information not present in the ECG. It is sufficient to measure ECG. Of course, TCM practitioners extract information from many sources when making a diagnosis [1] and this information may not be adequately reflected in electro-mechanical (using a condenser microphone) measurement of pulse on the human fingertip.

[1] Diagnosis-observation (of the patient's complexion, expression, movements, tongue, etc.), auscultation and olfaction, interrogation, and feeling pulse are four ways of diagnosis-observation in TCM. Among them feeling pulse is the most important.

4.6 Conclusion

In this chapter, seven healthy human ECG and pulse data have been subjected to a variety of tests designed to demonstrate the determinism in the cardiac system. The PPS method is employed to test ECG and pulse data against the hypothesis of periodic orbits with uncorrelated noise. A robust nonlinear measure, algorithmic complexity, is employed as the test statistic of the PPS method. Of the alternative test statistics, correlation dimension is also applied to the same cardiac time series and their surrogates. However, complexity remains simpler and requires less user's adjustable parameters. In contrast to complexity, correlation dimension can not converge at the stable value for high embedding dimension.

The experimental results indicate that healthy human ECG and pulse data are both inconsistent with periodic orbits with uncorrelated noise. Note that we do not deny one possibility that the human ECG and pulse data are generated by a periodic system with correlated stochastic elements. However, by combining results of ECG and pulse data with periodic and chaotic reference data we found that the changing tends of ECG and pulse data are both consistent with those of chaotic Rössler and Chua data and significantly different form those of periodic data, which indicates the presence of deterministic chaos in these cardiac data. Certainly these results can not constitute a definitive proof of chaos in health human cardiac output signals but they are found to be consistent with chaotic dynamics.

We have described the prediction analysis from ECG to pulse data and vice versa, which indicates that both ECG and pulse data conform to the deterministic processes. We found that human ECG and pulse data do not exactly conform to the same deterministic process. Meanwhile, the transfer entropy between ECG and pulse quantifies higher exchange of information from ECG to pulse activity. We conclude that ECG data contains more deterministic information than the pulse data, as the results of prediction between them suggest. Hence, medical diagnosis may benefit either from measuring ECG as in standard western medicine or using more sensitive data collection devices for pulse measurement.

Based on the results in this chapter we notice that in contrast with correlation dimension, complexity can be employed to search for deterministic components in otherwise apparently random data. Unlike complexity, correlation dimension is powerless to distinguish the data contaminated with large-level noise (i.e. very low signal noise ratio) and stochastic noise. For the following research we are collaboration closely with the Australian International Gravitational Research Center in the School of Physics of The University of Western Australia. The research subject is the simulated gravitational waves, which are always submerged in strong background noise. These simulated gravitational waves are provided by our partners in the Australian International Gravitational Research Center. We intend to employ nonlinear dynamics methods (the surrogate data method and complexity) to detect these weak deterministic signals.

5. IDENTIFYING DETERMINISTIC SIMULATED GRAVITATIONAL WAVES: ALGORITHMIC COMPLEXITY AND THE SURROGATE DATA METHOD

5.1 Introduction

5.1.1 What is Gravitational Waves

Sir Isaac Newton is the the founder of modern physics. One of his most significant achievements is to describe the law of gravity. Newton noticed that all objects are attracted to each other, and the strength of the attraction increases for large masses, but decreases as the objects move farther apart [80].

But Newton's law of gravity suffers from serious flaws. For examples, according to Newton's theory, information can travel across the universe instantaneously, named by action at a distance, i.e. some information can travel faster than light [81]. Einstein described that objects are attracted to each other because heavy objects bend spacetime and other objects follow the shortest path through this curved spacetime [82]. Even the light has to travel along the curved spacetime taking the shortest path between two objects. Intuitively, the spacetime looks like a stretchy fabric. Fig. 5.1 illustrates the curved spacetime surrounding a massive object in space. If one massive object collapses (or changes its mass), the spacetime round it will be disturbed. These spacetime changes propagates as gravitational waves. As they spread through space, gravitational waves cause the change of spacetime. That is, the shape of an object stretches or contracts when gravitational waves pass through it. The gravitational waves that scientists can detect under the current situation are from sources, such as binary neutron stars, supernovae, and colliding

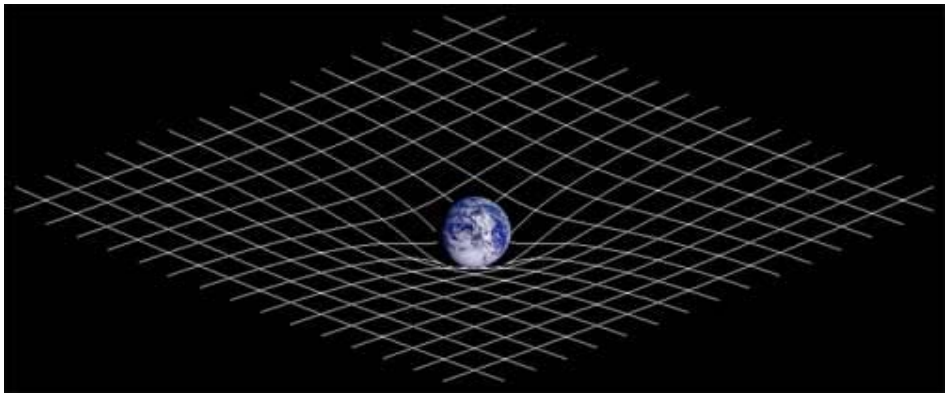


Fig. 5.1: A massive object changes the spacetime around (Image courtesy of [82]).

black holes [1] [83].

The strength of the gravitational wave determines the change of an object in shape [84]. It is measured by strain, the fractional change in distance, called h. Currently, there are two kinds of GW detectors: resonant mass detectors [85] [86] and laser interferometers [87]–[91]. We will omit the working principles of these two detectors and their detection since they are out of the topic of this thesis.

5.1.2 Simulating a GW Background from Cosmological Supernova

Three long-baseline laser interferometer GW detectors have been, or are nearly, constructed. The US LIGO (Laser Interferometer Gravitational-wave Observatory) has started observation with two 4-km arm detectors situated at Hanford, Washington, and Livingston, Louisiana; the Hanford detector also contains a 2-km interferometer. The Italian/French VIRGO project is commissioning a 3-km baseline instrument at Cascina, near Pisa. There are detectors being developed at Hannover (the German/British GEO project with a 600-m baseline, which had its first test runs in 2002) and near Perth (the Australian International Gravitational Observatory, AIGO, initially with an 80-m baseline). A detector at Tokyo (TAMA, 300-m baseline) has been in operation since 2001. The astrophysical detection rates are

[1] Gravitational waves are only information emitted by the black hole. By capturing these gravitational waves scientists may understand the nature of black holes.

expected to be low for the current interferometers but second-generation observatories with high optical power are in the early stages of development; these 'advanced' interferometers have target sensitivities that are predicted to provide a practical detection rate.

Our interest is in developing new signal processing methods for detecting the GW background generated by transient events throughout the Universe, in particular supernova. For an assumed local GW transient source rate density, r_0, methods have been developed to simulate the GW amplitude and temporal distribution of cosmic transient GW events e.g. supernova [92] – [95]. The simulation provides a tool to model the signature of the GW signal comprised of many unresolved GW transients in interferometric data.

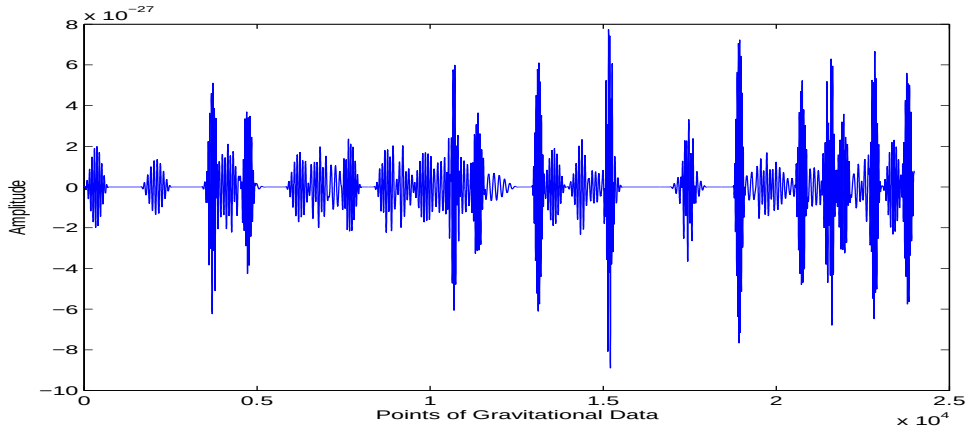
With assumption that interferometer noise is Gaussian and stationary gravitational waves [2] are simulated through procedures described in [92] [94]. Meanwhile, there are some significant efforts to model realistic non-stationary noise for the interferometric data in VIRGO group [96] and LIGO group [97].

Fig. 5.2(a) shows a simulated GW signal from supernova throughout the Universe; the other shorter section – Fig. 5.2(b), shows the individual event. The cumulative signal from transient GW sources at cosmological distances is commonly described as a stochastic background because of the temporal randomness of the individual events.

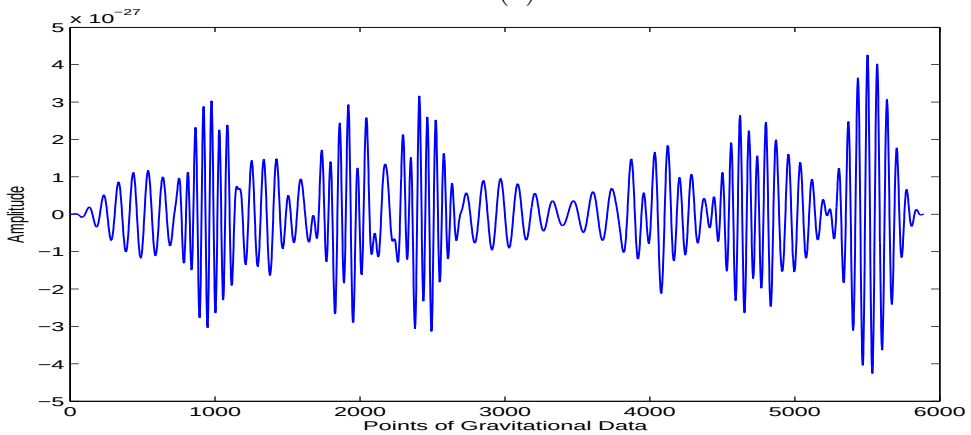
5.1.3 GW Data Detection

Techniques developed from the domain of signal processing have been verified to detect GW bursts from the noise. The standard matched filter technique is the op-

[2] For illustrative purposes, we use here a highly simplified input waveform, $h(t)$ — a quasi-monochromatic damped sinusoid of characteristic rest-frame frequency 1 kHz and duration 10 ms, with a maximum dimensionless strain amplitude of 7×10^{-24} at a fiducial distance of 10 Mpc. The waveform duration is approximately that of strongest GW emission of a DFM Type I (regular collapse) waveform, corresponding roughly to the ringdown phase. In addition to the input waveform, we assume an all-sky cumulative core-collapse rate of about $25s^{-1}$.



(a)



(b)

Fig. 5.2: The simulated GW background signal from core-collapse supernova occurring throughout the universe using the procedure developed by Coward, Burman and Blair [92]. For definiteness, the simulation uses a GW waveform, from a set of 78, obtained from the simulations of Zwerger & Müller 1997 [98]. We use this waveform as a representative example for a transient GW event. Panel (a) shows simulated cumulative GW data and (b) shows one individual transient event.

timal solution when the detected waveforms buried in stationary and Gaussian noise are known. Due to the same reason this optimal method may be impossible in practice. Recently feasible filtering methods are developed in European, American, and Japanese groups. Flanagan and Hugues [99] have described in detail the features for GW detection from the three different phases of coalescence events. Anderson et al. [100] [101] further developed the excess power method to detect gravitational

bursts of unknown waveform. Time-frequency detection algorithms are also proposed to identify short bursts of gravitational radiation [102] [103]. Meanwhile, a set of practical filters with high robustness are designed to detect gravitational burst signals, and the performances and efficiencies of these filters are also studied [104] [105] [106]. Beauville et al. firstly compared search methods for GW bursts using LIGO and VIRGO simulated data [107]. These improved filter techniques can obtain the optimal signal noise ratio only in term of certain known characteristics of the signal, such as signal duration and bandwidth. In addition, the outputs of the filters named SF and ALF are distorted comparing with the input signal in [106].

We utilize the estimation of algorithmic complexity to detect GW signals in presence of Gaussian white or coloured noise. Algorithmic complexity of a sequence, is a measure of the extent to which the given sequence resembles a random one [74]. In essence, it measures the regularity of the specified finite sequence, and its behavior is distinct for deterministic and for random sequences. Therefore, algorithmic complexity is sensitive to intrinsic deterministic components of the signal. Notably, the calculation of complexity is independent of characteristics of the signal, and its performance on SNR is better than spectral estimation. Algorithmic complexity does not require prior knowledge of these underlying signals, nor does it require fixed characteristics (eg. period). It is robust to noise and may be applied in conjunction with existing filtering methods as a further improvement to detection performance. Hence, comparison to existing techniques is largely irrelevant as this method will actually augment these existing methods.

But estimating complexity is not sufficient to make a decision on the data: one cannot determine with certainty (or even probability) that a particular value of complexity indicates the presence of deterministic dynamics. To address this problem we employ the surrogate data method (a form of statistical hypothesis testing). Algorithmic complexity provides a quantitative measure of deterministic dynamics in time series. The surrogate data method may be employed to benchmark these statistical results and compare them to the results expected for various types of pure noise processes. Significantly, VIRGO group has described the surrogate

data method with the hypothesis of NH1 to test for the nonlinearity of the data of VIRGO interferometer [108]. According to this hypothesis we can reject the data is not linear noise but it may be insufficient to determine whether the data contain nonlinearity. Moreover, our application of the surrogate data method is different. Here we adopt the surrogate data method with hypotheses of NH0 and NH1 so as to attach statistical significance to the results of algorithmic complexity and also ensure the data distinguished by complexity is not linear filtered noise. For our choice of test statistics, we propose to replace the popular statistic, correlation dimension with complexity. Since correlation dimension estimations are sensitive to the noise, even uncorrelated noise they are not applicable to field measurement [63].

5.1.4 Encoding Schemes of Algorithmic Complexity

The performance of algorithmic complexity is closely related to the encoding scheme selected. One can generate a new encoding scheme easily but there is no criterion to determine the right encoding scheme for the given time series. Very probably one can obtain better performance of algorithmic complexity by using some other encoding scheme. In this work the observed data are partitioned into three symbols 0, 1, and 2 in term of the same probability. We also have tried to convert this data to sequences of two, four, and five symbols and calculate their complexity respectively. Among these results complexity in term of three symbols are most sensitive to the SNR. But we have not found a way to select the suitable numbers of bins for this encoding scheme. Before computing this symbolic sequence, we first employ a numerical filter to smooth the data. The filtered data is achieved through the equation, $x_{filter}(i) = (x(i) + x(i+1) + x(i+2))/3 \;\; i = 1....n$, where n is the length of time series $\{x_i\}$. We note that the numerical filter is not necessary for calculation of complexity but it can improve the sensitivity of complexity to the noisy data. We may expect better performance of this method by applying advanced filtering techniques. This also indicates that combination of filtering techniques and algorithmic complexity may improve the performance of the application of one of them to detect GW data in the presence of strong noise.

Real interferometric data may contain some short-time pulses (glitches) probably produced by malfunction in the detector. The method of complexity is robust to the effect of the glitches. First of all, by using our encoding scheme of three symbols (0, 1, 2) and two symbols (0,1) to convert this data, the converted sequence can be affected by these glitches with the probability of only 66.7% and 50% respectively. Even if the converted sequence has been disturbed by the glitches these abnormal segments contribute little to the calculation of complexity of the whole sequence. So the glitches do not substantially change complexity of measured signals. Due to the same reason complexity does not identify the Gaussian burst signal contaminated with strong noise as well as the filtering techniques in [106]. Complexity of the noisy data and strong noise itself are almost the same. Hence, this method is expected to perform well for the background signals and complicated waveforms.

5.2 Identification of GW Transients

In Section 5.2.1 and 5.2.2 we apply our estimation of algorithmic complexity and the surrogate data method to detection of the background from different level noise, including both Gaussian white and coloured noise; in the section of 5.2.3 we try to localize each single event in certain strong noise by using complexity. There are several forms of SNR, such as Eq. 24 in [110], Eq. 1 in [107], and Eq. 29 in [111]. For consistency, we choose for our definition of SNR the square root of the ratio of the power spectral densities of the signal and noise integrated over all frequencies.

To compare the performance of this method to existing techniques, we also calculate linear autocorrelation as an alternative criterion. The numerical results indicate that application of algorithmic complexity can distinguish deterministic GW data from both the white and coloured noise. By comparison autocorrelation failed to identify the GW signals in the presence of coloured noise. Our computational scheme is illustrated as follows:

(1) Add different levels of Gaussian white noise to the GW data (from the very low SNR to high SNR) and then calculate complexity of the sum. Higher SNR

indicates that the GW data is dominant and the calculated complexity is closer to complexity of the noise-free GW data; lower SNR means that the noise is dominant and the calculated complexity is closer to complexity of the noise. By varying the SNR we can test how much noise this complexity measure is able to overcome.

(2) Verify the sensitivity of complexity to the GW data with different levels of coloured (that is, linearly filtered) noise. The coloured noise is generated by Gaussian white noise plus the same noise with the certain delay time. We add the coloured noise to the original data in the same way as step (1).

(3) Apply the surrogate data method to the same data. For the data contaminated with white noise, we employ the surrogate data method to determine whether to reject the hypothesis, NH0, that the data can be described as i.i.d. noise. (For white noise, this hypothesis is true; for the GW data it is false.) At each SNR we generate 30 surrogates and calculate complexity for the surrogates and the original data. According to the results, we can determine the minimum SNR for which complexity can present meaningful results.

(4) For the data contaminated with coloured noise, we apply the hypothesis, NH1, that this generated surrogate data is linearly filtered noise. We then repeat the above step. Rejection of the first case (step (3)) indicates that the data exhibits temporal correlation. This is true for both linearly filtered noise and deterministic signals such as the GW data. Rejection of the second case (step (4)) indicates that the data is also inconsistent with simple linear noise.

(5) To compare the results of this test with more standard (albeit linear) statistics, we make further comparative experiments (following the procedure in step (3) and (4)) using the autocorrelation as the discriminating statistic in the surrogate test. That is, autocorrelation is used to detect the deterministic dynamics in the place of complexity.

5.2.1 Sensitivity of Algorithmic Complexity to GW Data

For both a long data section (23988 points) and a short data section (5892 points) the sensitivity of algorithmic complexity is shown in Fig. 5.6 and 5.7 respectively.

We observe that algorithmic complexity can even identify the difference between the data and white noise at the SNR of 0.08 (see Fig. 5.6 (a)) and 0.14 (see Fig. 5.7 (a)), and also the difference between the data and the coloured noise at the SNR of 0.05.

We take the standard deviation of the detector noise to be [106]:

$$\sigma = h_{\mathrm{rms}}\sqrt{f_{\mathrm{o}}/2f_{\mathrm{c}}}\,, \tag{5.1}$$

with h_{rms}, the root-mean-squared value of the advanced LIGO noise curve at f_{c}, the source frequency of a supernovae, and f_{o}, the sampling frequency. We assume the noise is white, Gaussian with zero-mean and take $h_{\mathrm{rms}} \sim 1.4 * 10^{-23}$ corresponding to the most sensitive region of the Advanced LIGO noise curve ($f_{\mathrm{c}} = 100$Hz). So the Gaussian detector noise of the Advanced LIGO is the standard normal Gaussian noise (mean = 0, standard deviation = 1) multiplied by h_{rms}.

The SNR for the same data in the presence of Gaussian detector noise of advanced LIGO is 10^{-6}, which is beyond the capability of the current approach. In fact for such a low SNR it is unlikely that any method can successfully distinguish signal from noise — without making significant assumption about the signal or demanding substantially more data. However, a combination of advanced filtering techniques and complexity with an effective encoding scheme may provide further improved sensitivity.

Fig. 5.3 and 5.4 illustrate the long GW data contaminated with white noise (0.16, i.e. $-16dB$) and coloured noise (0.10, i.e. $-16dB$) and their spectrum respectively. Either noisy GW data appears to be random noise. But their SNRs are still greatly larger than the minimal SNR that complexity is sensitive to. It is difficult to determine whether the GW data contributes these spectrums. Fig. 5.5 is the spectrum of the long GW data. We can not find any part matched with

the original spectrum in Fig. 5.3 and 5.4. Therefore, algorithmic complexity has superior power to detect determinism compared to the standard of FFT, especially for long and non-stationary data set. The power spectrum estimated by FFT seeks to describe the signal as the average frequency content over the entire data length. For longer samples with several pulses of different frequency, this approach does not make sense. For short samples, there is insufficient data. However, complexity algorithm looks for patterns in the data, and can detect such patterns even if the dominant frequency changes.

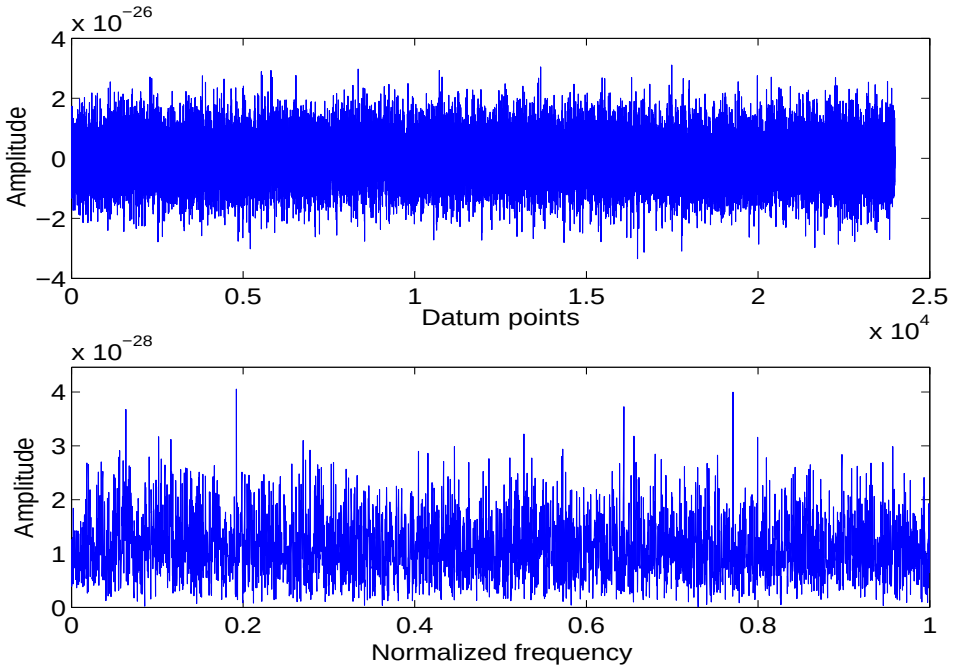


Fig. 5.3: The long GW data contaminated with white noise (top panel) and its spectrum (bottom panel).

In the following we not only determine to what extent complexity of these data is significantly different from that of random noise, but also test whether more sophisticated coloured noise could contribute to the observed difference in complexity. To address both questions we employ the surrogate data method to figure them out.

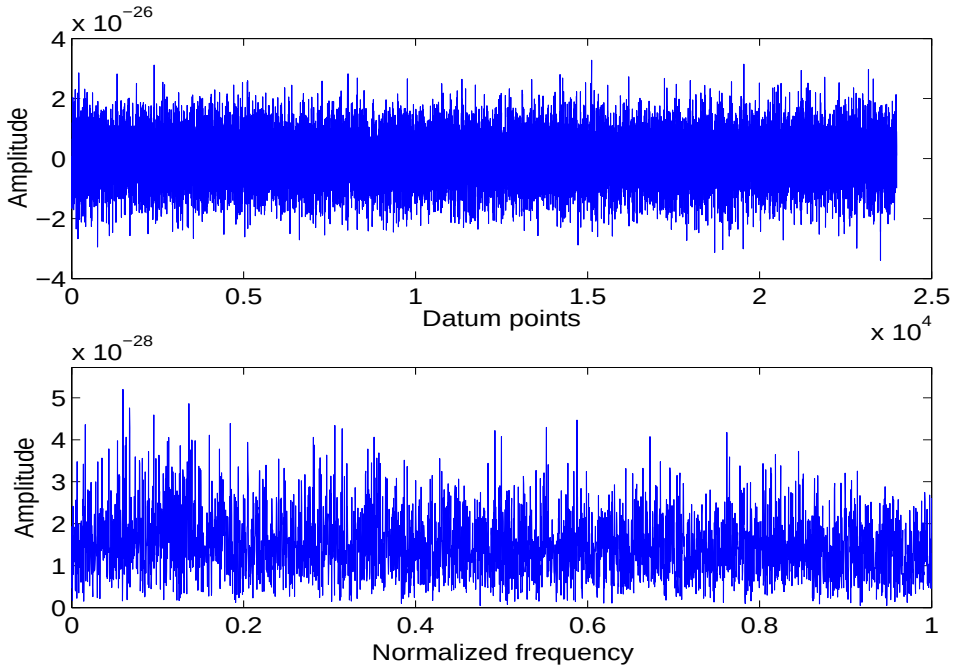


Fig. 5.4: The long GW data contaminated with coloured noise (top panel) and its spectrum (bottom panel).

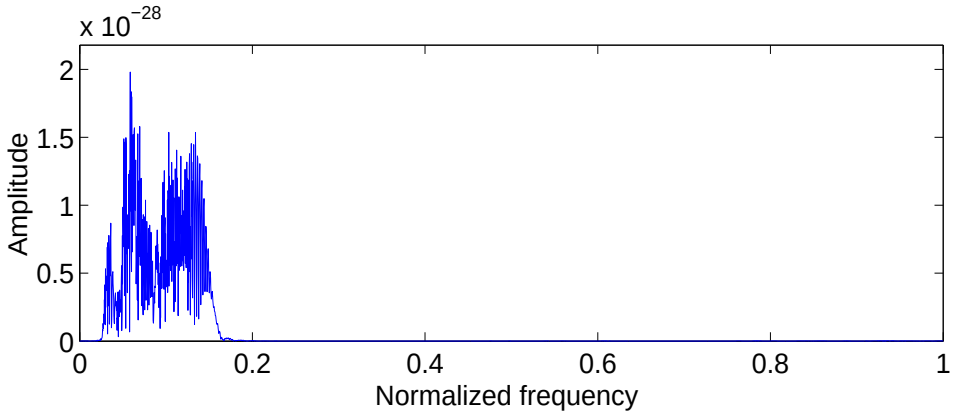
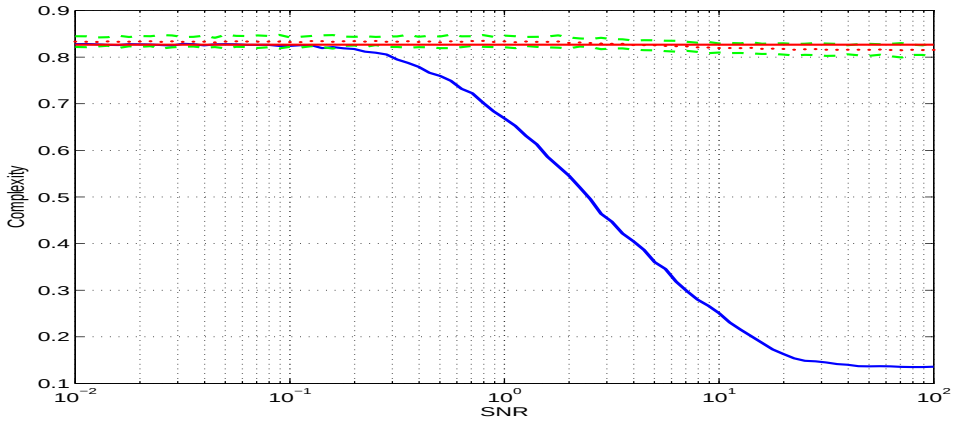


Fig. 5.5: The spectrum of the original long GW data. By observation we can not recognize its profile in Fig. 5.3 and 5.4.

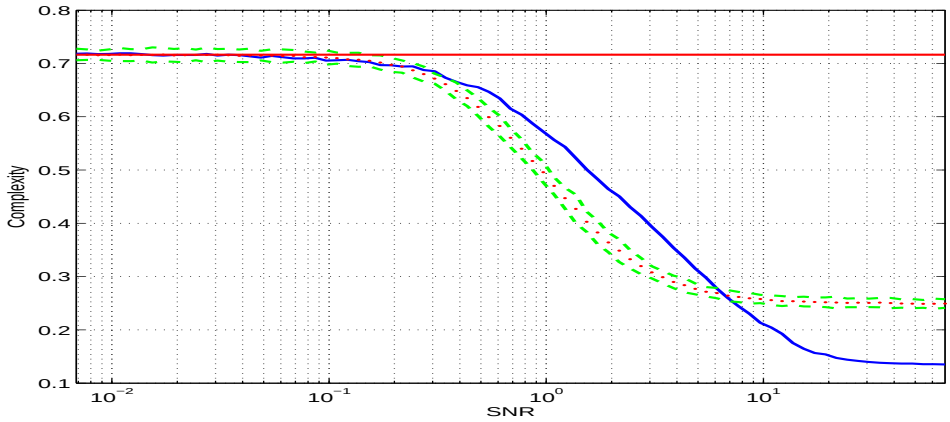
5.2.2 Application of the Surrogate Data Method

For the case of added white noise, the given hypothesis is that the contaminated signal is i.i.d. noise and the surrogate generation algorithm, Algorithm 0 is required; for added coloured noise the given hypothesis is that the contaminated noise is

linear noise and Algorithm 1 is used to generate surrogate data. We then calculate complexity for both surrogates and the original data to make a decision.



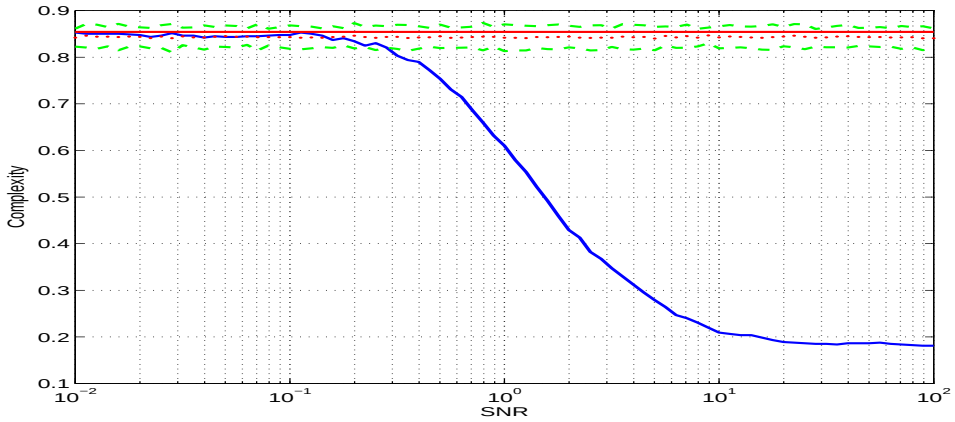
(a)



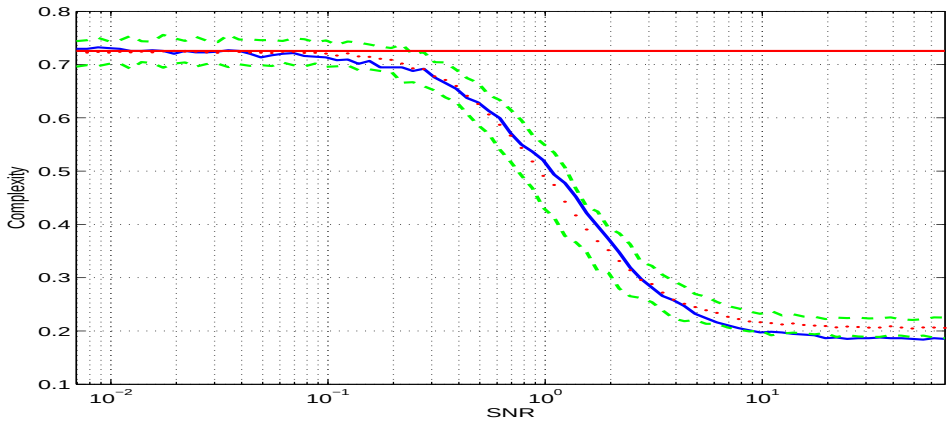
(b)

Fig. 5.6: (a) The solid line is complexity of the original data (23988 points) contaminated with white noise for each SNR; the straight solid line is complexity of white noise added; dots are the mean complexity of 30 surrogates for every SNR; two dashed lines denote the mean plus and minus three standard deviation. (b) The solid line is complexity of the same original data contaminated with coloured noise for each SNR. The straight solid line is complexity of coloured noise added. Properties of the other lines and markers are defined the same as (a).

In Fig. 5.6(a) algorithmic complexity of the long original data is significantly different from the range of complexity of surrogates for the data whose SNR is equal to or higher than 0.18 with 99.97% probability. The ensemble of surrogates



(a)



(b)

Fig. 5.7: (a) The solid line is complexity of the original data (5892 points) contaminated with white noise for each SNR; the straight solid line is complexity of white noise added; dots are the mean complexity of 30 surrogates for every SNR; two dashed lines denote the mean plus three standard deviation (the upper line) and the mean minus three standard deviation (the lower line). (b) The solid line is complexity of the same original data contaminated with coloured noise for each SNR. The straight solid line is complexity of coloured noise added. Properties of the other lines and markers are defined the same as (a).

conforms to a Gaussian distribution. Fig. 5.8 illustrates the typical results that for the above four kinds of surrogates their distribution are all approximate to the Gaussian distribution. We also conclude that the lowest SNR for which the surrogate method can identify the non-random structure using complexity as a discriminating

statistic is 0.18. Note that if the range is reduced to twofold standard deviation the surrogate method can identify the SNR of 0.06 with 60% probability. Fig. 5.7(a) presents the analogous results for the short data contaminated with white noise. When making a decision on the contaminated data in Fig. 5.6(b) the surrogate data method can reject the hypothesis for the original data whose SNR is higher than the SNR of 0.31 with 99.97% probability and the SNR of 0.24 with 60% probability. Although there is crossing at the SNR of 6.6 in the figure, it is of no significance as the GW data is dominant.

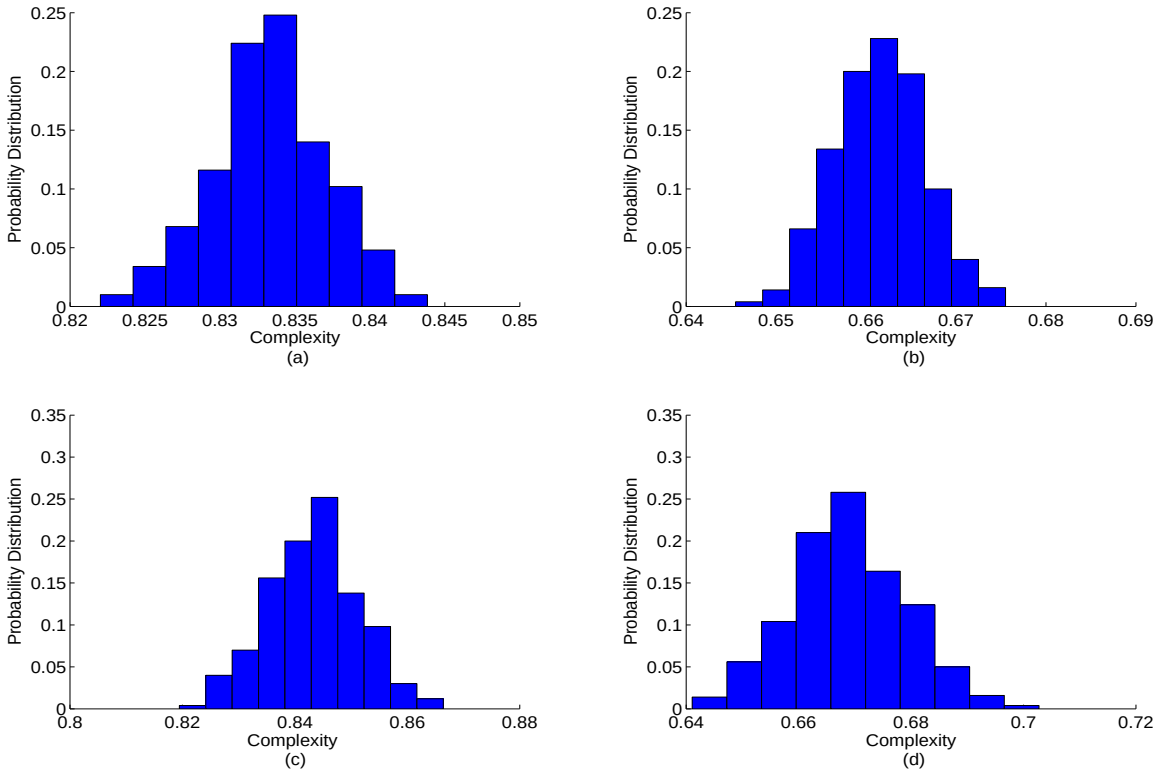


Fig. 5.8: (a) Distribution of surrogates of the long GW data contaminated with white noise; (b) distribution of surrogates of the same long data contaminated with coloured noise; (c) distribution of surrogates of the short GW data contaminated with white noise; (d) distribution of surrogates of the same short data contaminated with coloured noise. The SNR of the original data in (a), (b), and (c) is the minimal SNR that complexity can identify for such data.

For the short GW data, as shown in Fig. 5.7(b), complexity of surrogates follows that of the original GW data, but this does not mean that complexity fails

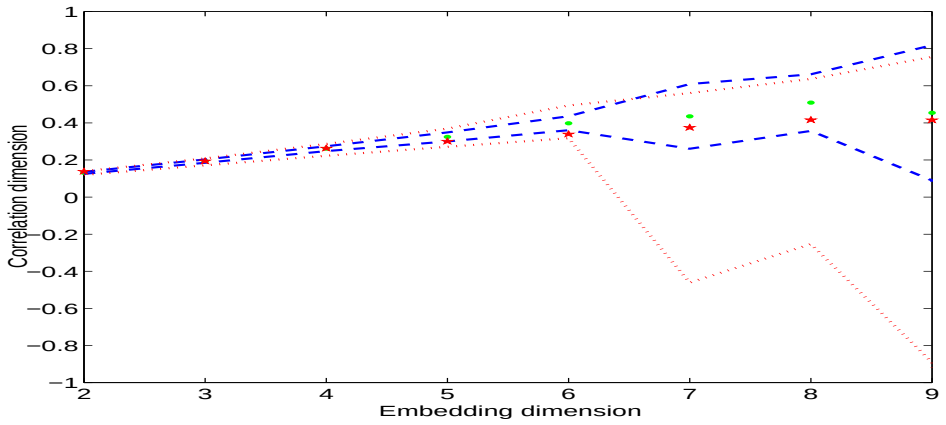
to work as a discriminating statistic. Actually a feature of the surrogate generation algorithm (Algorithm 1) leads to this results. When applying Algorithm 1, the phases of the complex conjugate pairs are shuffled to generate the surrogates. The short data is approximately pseudo-periodic time series and surrogates generated by this algorithm are similar to the original data. Certainly complexity of the surrogates is close to complexity of the original. So this is a negative result — but not an unexpected one. Also, complexity can obtain better performance on longer data sets contaminated by either noise.

We also apply correlation dimension used in Chapter 3 to the long GW data contaminated with white noise (0.79, i.e. -2dB) and coloured noise(0.54, i.e. -2dB) ,respectively, and their surrogates. From Fig. 5.9 we notice that correlation dimension cannot distinguish the deterministic GW data from the background noise. Obviously, this noise level (the SNR of -2dB) exceeds the maximum noise level that correlation dimension estimations can tolerate. Moreover, correlation dimension still fails to converge with increasing embedding dimension.

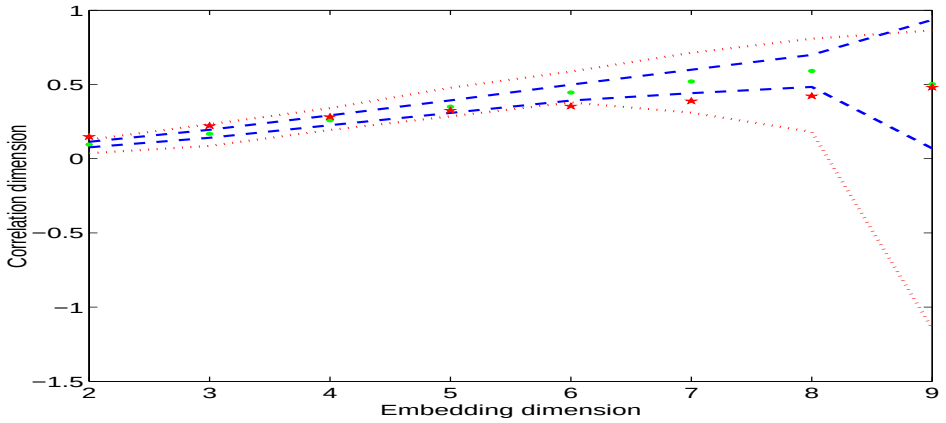
The autocorrelation function is commonly used for two primary purposes: to detect non-randomness in data and to identify an appropriate linear time series model if the data is not random [112]. If the autocorrelation is used to detect non-randomness, the first (lag 1) autocorrelation is usually sufficient. If the autocorrelation is used to identify an appropriate model of time, the autocorrelation is usually plotted over a range of lags. Since we are not, at this stage, concerned with modelling, the first autocorrelation (lag 1) is selected.

Accordingly, we employ the autocorrelation function to repeat all the previous experiments for the same data. We found that the sensitivity of the autocorrelation to GW data is similar but not superior to complexity. However, when applied to the data contaminated with coloured noise the autocorrelation is clearly deficient. The relevant results are shown in Fig. 5.10.

Referring to Fig. 5.10(b) we cannot reject the hypothesis that the contaminated noise is linear noise. In fact the contaminated data consists of the deterministic signal (GW data), which is not linear noise. Consequently, although the autocor-



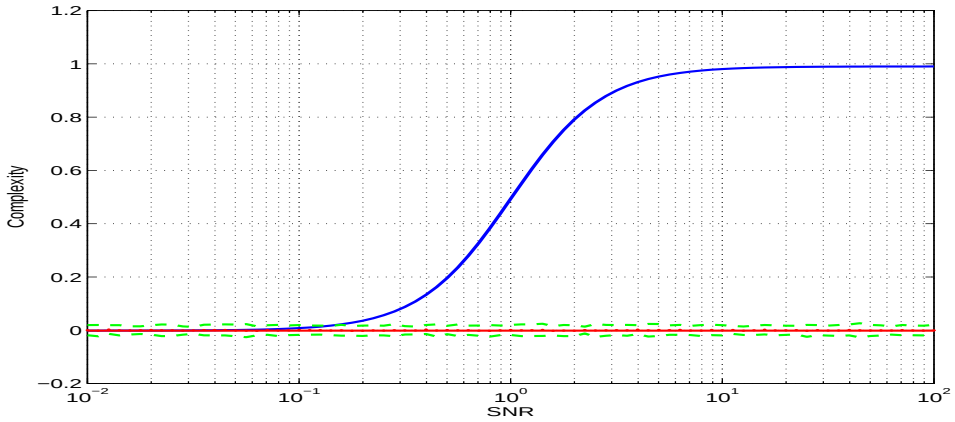
(a)



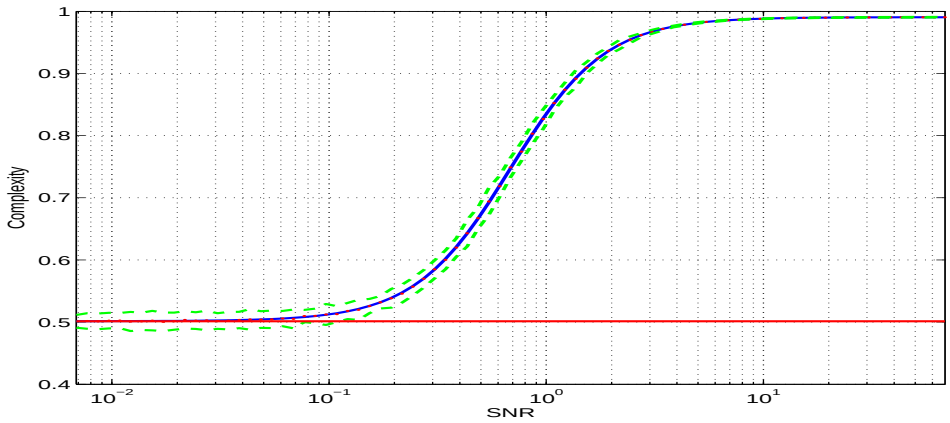
(b)

Fig. 5.9: The top panel is the results of GW data with the white noise; the bottom one is the results of GW data with the coloured noise. In both panels, stars are correlation dimension of original noisy GW data with embedding dimension from 2 to 9; dots are the mean correlation dimension of 30 surrogates at each embedding dimension; two dashed lines are the mean plus and minus the standard deviation, and two dotted lines are the maximum and minimum correlation dimension among these 30 surrogates.

relation can differentiate either contaminated data from the random noise as well as algorithmic complexity, it is of no use in the presence of coloured noise. Note that this is not surprising as the autocorrelation is mathematically equivalent to the power spectrum. Weakness of the FFT based power spectrum estimation is also evident in the autocorrelation calculation. For the short GW data contaminated



(a)



(b)

Fig. 5.10: (a) The solid line is the autocorrelation of the original data (23988 points) contaminated with white noise for each SNR; the straight line is autocorrelation of the white noise added; dots are the mean complexity of 30 surrogates for each SNR; two dashed lines denote the mean plus and minus three standard deviation. (b) The solid line is the autocorrelation of the same original data contaminated with coloured noise for each SNR; the straight line is autocorrelation of the coloured noise added. Properties of the other lines and markers are defined the same as (a).

with either white noise or coloured noise autocorrelation is nearly identical to those of the long data.

5.2.3 *Localization of GW Data Using Algorithmic Complexity*

Finally, we now turn to the problem of locating a short deterministic burst in a noisy signal. In order to estimate the location of GW data in the noisy signal, we apply a moving window to the noisy data to calculate complexity of the data in the window. The window size we selected is 1000 points, and the moving step size is 200 points. The noisy data are the same five bursts selected from the long GW data, which are then contaminated with white noise. Its SNR is 0.24. The results are presented in Fig. 5.11.

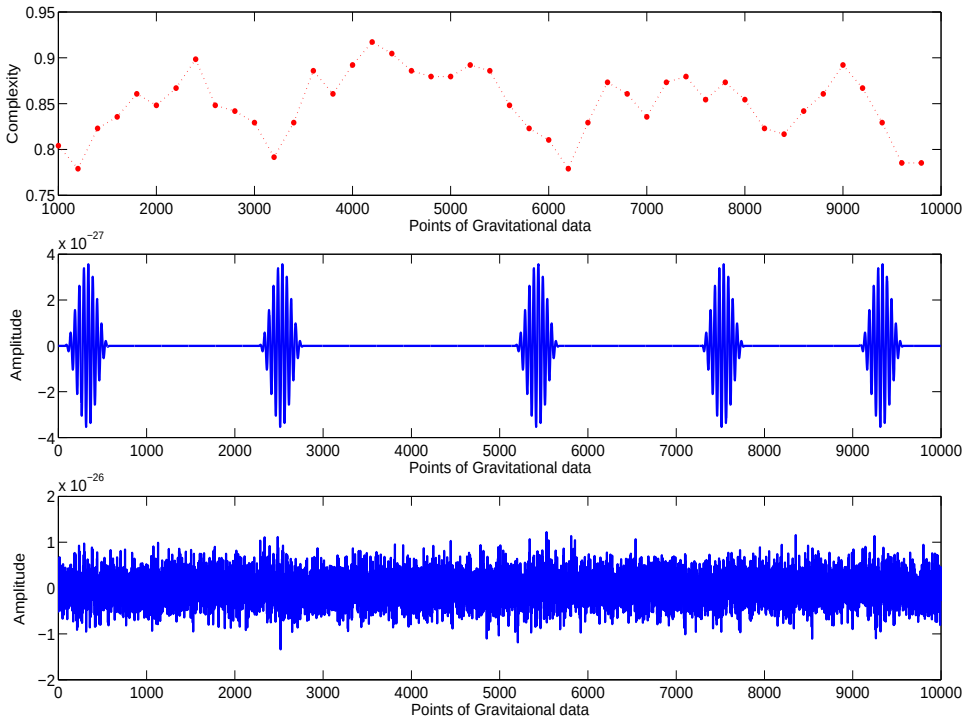


Fig. 5.11: Complexity of the data (dots) in the each window (top panel), the simulated GW data for reference (middle panel), and the corresponding noisy data (bottom panel).

We notice that the local minima (top panel) indicate the existence of GW data in the corresponding windows, and the local minima of complexity match the location of GW data in the noisy data.

5.3 Improvements on Gravitational Waves Detectors

In collaboration with the Australian International Gravitational Research Center of The University of Western Australia we designed a telemetry system driven by radiation power so as to prevent electrical wiring from mechanically short circuiting in the high performance vibration isolators for gravitational wave detection [71]. This work aims to improve data measurement at a higher SNR, which could greatly facilitate the following identification of possible GW signals. The material presented in this section can be regarded as the supplement of identification of gravitational waves in Chapter 5. Moreover, this work demonstrates that future improvement in gravitational wave detection hardware will facilitate improved measurement and allow techniques such as the methods presented earlier in Chapter 5 to be usefully applied for gravitational wave detection.

The laser interferometric gravitational wave detector aims to directly detect gravitational waves by measuring the relative arm length variation induced by gravitational waves. To reach such extremely high sensitivity, very good isolation against seismic noise is required. While this has been achieved by suspending each mirror (acting as test mass) at the end of a multistage pendulum, the large amplitude pendulum motions must be accurately controlled to bring the interferometer towards its optimal working configuration.

One frequently used control system in the interferometric gravitational wave detectors is comprised of motion sensing by a shadow sensor and actuation by a system of coils and magnet [113]. The shadow sensor consists of a blade fixed to the test mass which cuts part of the light path between a light-emitting diode and photodiode. The blade position determines the photocurrent. The feedback control signal is derived from an analog circuit or digital embedded controller to either damp the motion or to position the isolator. It requires a large number of wires to transmit motion sensing signals and actuation signals between the isolator in vacuum and the controller in air. Especially for the control of the test mass, all these wires have to be deliberately wired through the vibration isolator stages to minimize the seismic

coupling through the wires. Here we present experimental results of our telemetry actuation on two stage pendulums, which aims to replace all wires and realize truly wireless communication for the test mass control.

Our system consists of two subsystems, the embedded system on the reference platform inside the vacuum tank and the monitor system outside in air. In order to realize communication between the two subsystems over a relatively short distance, we adopt infrared with a wavelength of 850nm 900nm to transfer all signals. Infrared is highly adapted to point-to-point communication over short distances. The telemetry system and the schematic of the two stage isolator are shown in Fig. 5.12.

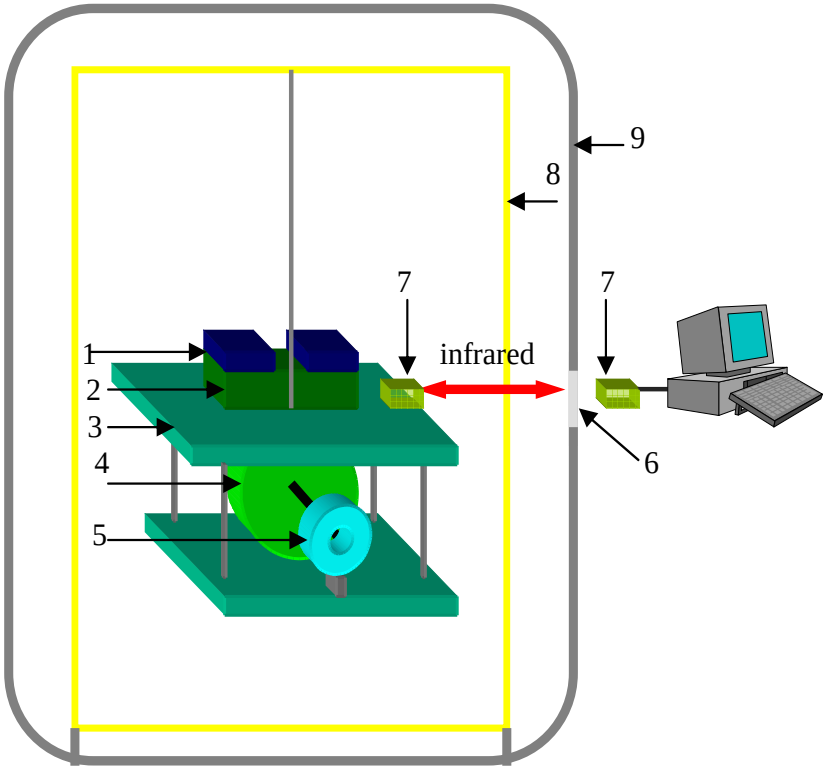


Fig. 5.12: The schematic of the telemetry system and the two-stage isolator. 1 photo silicon arrays; 2 the embedded system; 3 reference mass; 4 test mass; 5 shadow sensor and coil; 6 glass window; 7 emitting/receiving diodes; 8 frame; 9 vacuum tank. Photo silicon arrays, shadow sensor and coil, and diodes are connected to the embedded system with wires. The embedded system is installed on the reference mass.

The shadow sensor and the coil actuator are connected to the embedded controller with wires, which is mounted on the reference mass platform. Two photosilicon arrays are also mounted on the reference mass stage to provide power to the embedded system. The infrared transceiver of the embedded system points to that of the monitor system in air. The shadow sensor collects the one-dimension motion signals of the test mass relative to the reference mass. The actuator (coil/magnet) with feedback keeps the test mass stable in that direction. The embedded system executes all the following tasks: analog to digital (A/D) conversion, computing, transmitting data of motion sensing and receiving gain signal though infrared diodes, and digital to analog (D/A) conversion. There are two types of signal required to transmit between the two subsystems: sampled motion-sensing signal that is transmitted all the time, and the value of gain that is transmitted as required.

Sampled data is transmitted to the monitor system outside through the infrared communication. The monitor system is connected to the serial interface of a computer. We adopt LabView to design the interface software, which reads and saves the data received and plots the waveform of motion-sensing signal in real time. Fig. 5.13 displays the interface of this software monitoring the real-time position of the test mass. We can also send the value of gain, which adjusts the amplitude of the feedback, to the embedded system inside through this interface software.

Using the SRS780 spectrum analyzer we measure the transfer function of the embedded system including A/D, D/A, and microcontroller, which is presented in Fig. 5.14. The analog circuits in the sensors and actuators are much faster than the microcontroller and will not affect performance of the control loop. Their transfer functions are not included in the plots. The control band width of 45Hz is also sufficient for the local control of the isolator. The telemetry system is configured to apply velocity damping to the pendulum. We can remotely change the damping gain. Fig. 5.15 shows the pendulum ring down curve measured with ($gain > 1$) and without ($gain = 0$) damping. It is obvious that the test mass motion is damped effectively with damping on.

This prototype telemetry system demonstrates the feasibility of using infrared

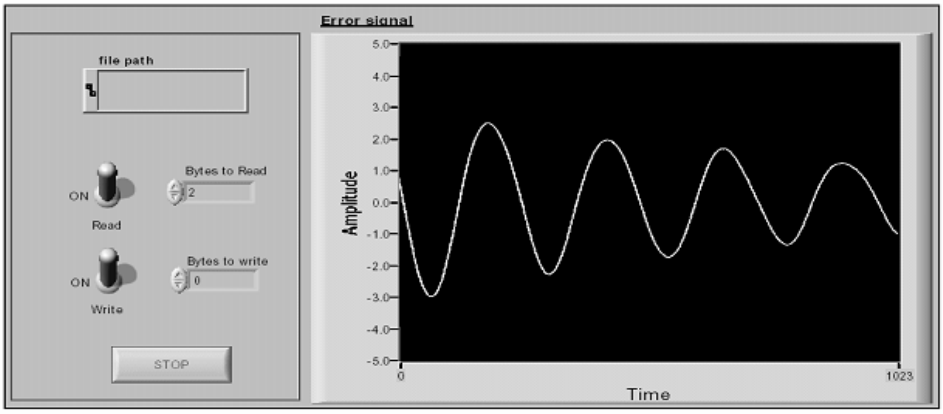


Fig. 5.13: Interface of the monitor software. The READ button is to enable the software to receive the data from the interface of the computer. The WRITE button is to enable the software to transmit the gain, which can be modified in the small window besides it, to the embedded system. The large window is to show the real-time waveform that represents the movement of the test mass.

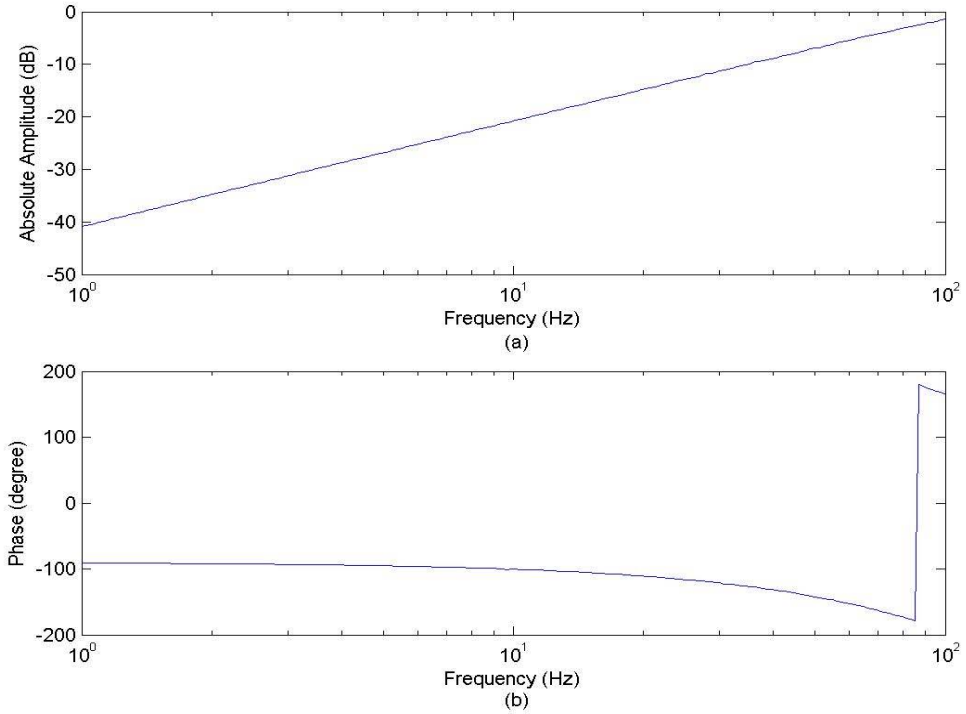


Fig. 5.14: Bode plots of the embedded system (A/D, D/A, and microcomputer). The phase delay increases with the signal frequency due to the time delay.

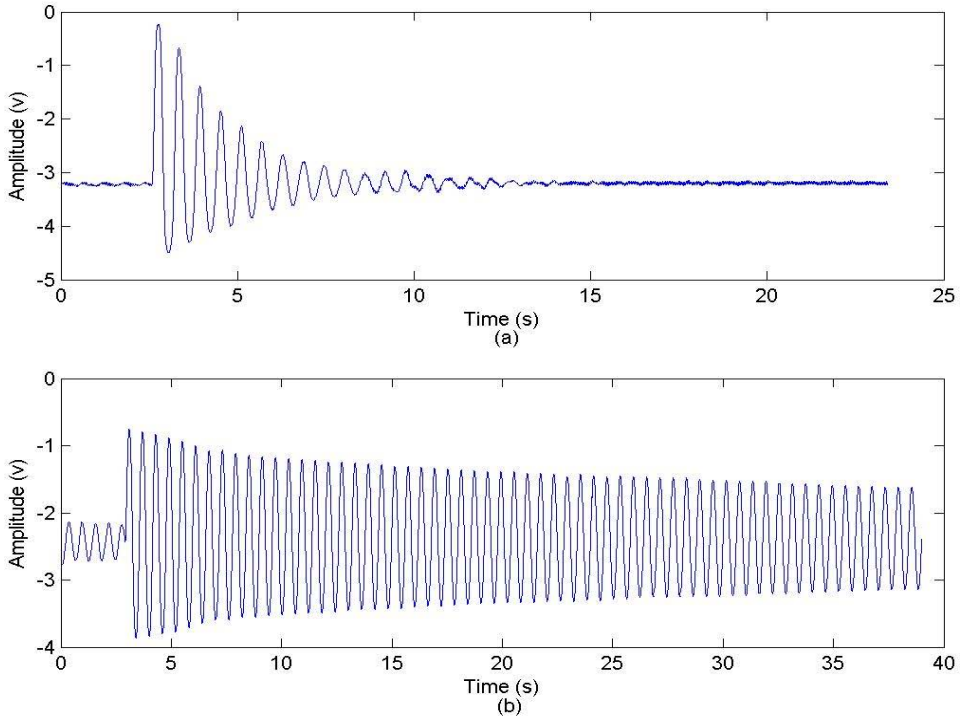


Fig. 5.15: (a) The ring down curve of the test mass damped by the embedded system; (b) the ring down curve of the test mass without damping.

wireless communication and radiation power supplies to avoid noise coupling through wiring in advanced vibration isolators for gravitational wave detection. In our experiments the telemetry system controls the test mass in one dimension. It is feasible to extend the current system to control several degrees of freedom almost simultaneously. If higher control bandwidth or more control channels are required we may consider more powerful controllers or alternative configurations.

5.4 Conclusion

We have described an alternative method, algorithmic complexity to identify deterministic dynamics (simulated GW data) in the presence of substantial background noise. It can identify the existence of GW data contaminated with strong white or coloured noise better than other common methods, such as the FFT. Complexity used as the test statistic of the surrogate data method is more robust than autocor-

relation for the surrogate data method.

To provide a statistical benchmark for the results, we find it necessary to employ the method of surrogate data, which aims to determine whether the data contains statistically significant deterministic dynamics. For the surrogate data method presented here, we employ complexity as a test statistic. In the literature, however, correlation dimension is a common and popular choice. But, because correlation dimension is rather sensitive to noise, it is not a good choice to identify small dynamics contaminated with substantial noise. In contrast, the experimental results we present show that the surrogate data method with complexity as a test statistic can make the correct decision even for the relatively low SNR data. The final minimal SNR is determined by the sensitivity of the surrogate data method to the noisy data.

Like all statistical estimation problems, with more data our estimation of algorithmic complexity will improve. In fact, for very long data sets the sensitivity of algorithmic complexity to deterministic dynamics with a low signal to noise level will be much better. Moreover, unlike linear statistical methods, algorithmic complexity is also relatively robust to non-stationary (that is parametric changes) in the underlying deterministic signal. The essential, and important feature of algorithmic complexity is that it differentiates between deterministic patterns and random perturbations: irrespective of the precise origins of those signals. Hence, with sufficient computational resources, algorithmic complexity may offer a viable alternative for the detection of persistent deterministic dynamics hidden under substantial noise — even when the exact form of both the noise and the deterministic dynamic are unknown and may change with time.

In practical applications, due to the limitation of current interferometer technology, the data is usually contaminated with unknown noise sources, which may contain GW signals. It is possible to generate Gaussian noise with the same energy as the noisy data, and apply complexity to both this data set and the noisy data. If complexity of the data is smaller than that of the noise we could utilize the surrogate data method to assess the level of deterministic dynamics. Algorithmic complexity,

therefore, a potentially available method to detect deterministic dynamics in GW data where the signal power is significantly smaller than the noise power.

6. CONCLUSIONS

In this chapter, we summarize the major contributions of the whole research project, review works in each sub-project, and discuss the future research plans.

6.1 Contributions of the Thesis

This thesis reports two major contributions. Firstly, we have developed the current information criterion (minimum description length) with application of feedforward multi-layer neural networks to model nonlinear dynamics. By systematizing and approximating the critical parameters in such networks our minimum description length criterion is applicable to estimate the optimal networks regardless of training algorithms. The thesis presents demonstration of our methods for typical computational and experimental time series (including cardiac time series). Secondly, the optimal modelling techniques are then used for two specific applications from two physical systems (human cardiovascular dynamics and the dynamics of gravitational waves). The objective is to apply nonlinear modelling with optimal neural networks and the surrogate data method to characterize and understand their underlying dynamics. The surrogate data method further verifies the performance of the optimal models by analyzing model residuals. The two contributions are closely related to each other. In terms of the first contribution, we develop our modelling techniques to study several specific problems, which constitutes the second contribution. Meanwhile, in virtue of the second contribution, we also examine the performance of modelling techniques obtained in the first part.

Specifically, the contributions include the following:

1. A general methodology of selecting optimal neural networks for diverse time series has been introduced. All the neural networks are built independently, which guarantees that there is no connection between them. By allowing models of different sizes to be built independently, MDL is achieved through a broader search of parameter space. The improved formula of $M(k)$ more accurately counts the description length of the neural network parameters. A variable representing the effective number of parameters within one neuron is proposed in calculation of description length for model parameters, $M(k)$. Moreover, the current MDL criterion is robust and independent of training algorithms.

2. Combination of neural networks, minimum description length, and the surrogate data method is proposed in this thesis. Neural networks are employed to build the relationship between two kinds of time series, and the minimum description length estimates optimal neural networks which provide adequate generalization and accurately capture this relationship. Finally, the surrogate data method confirms the relationship between them by testing for model residuals against the given hypothesis. The combination in this manner is applicable to a wide variety of real world data.

3. We pay particular attention to the development of suitable test statistics for surrogate data methods to field measurement. The practical data through a measurement process is usually somewhat contaminated with noise. Even small uncorrelated noise can destroy all nontrivial self-similarity, which correlation dimension estimates. In contrast algorithmic complexity is robust and much more sensitive to the inherent deterministic components hidden by the strong noise. It has no connection with slection of embedding dimension and delay time (i.e. independent of phase space reconstruction) and is well suited for real-time implementation.

4. The surrogate data method with the test statistic of complexity is thoroughly investigated for identification of gravitational waves. This technique attaches the statistical (i.e. general) significance to the result of the single data set.

To assist in the future collection of suitable experimantal data, a telemetry system driven by radiation power is designed for use in the high performance vibration isolators for gravitational waves detection.

In summary, there are five kinds of surrogate data methods used in this thesis: 1) the common surrogate data methods with the hypothesis of NH0, 2) the common surrogate data methods with the hypothesis of NH1, 3) the pseudo-periodic surrogate data method, 4) the cycle-shuffled surrogate data method and 5) the small-shuffled surrogate data method. The objective criterion of selecting one of them is which property of the observed data we wish to investigate. For example, in Chapter 3 we were interested in whether model residuals are random noise, i.e. identical and independently distributed noise in the strict definition so we applied the common surrogate data method with the corresponding hypothesis, NH0 to those data. Obviously, all the other surrogate data methods are not applicable to this test. In Chapter 4, we were interested in the periodic behavior of ECG and pulse data and then the available methods we are familiar with are the pseudo-periodic surrogate data method and cycle shuffled surrogate data method. Hence, we applied both methods to those data in comparison. We found that the PPS method worked better than another method and was suitable to the generic case. Note that the selection of interesting properties of the observed data is a subjective process.

6.2 Summary

We began this thesis by introducing the contents of this thesis, including the research subjects, techniques used, and major results. We gave a brief comprehensive literature review on the previous and current developments of nonlinear dynamics modelling and its applications.

In Chapter 2, we employed the description length method for selecting optimal neural networks in modelling four typical time series, the Rössler system, the Ikeda map, chaotic laser data, and human ECG data. Our improved method is a significant generalization of previous attempts at applying description length to model selection

of neural networks for time series prediction. Independent construcion of each neural network leads to the fluctuation of the original description length curve. It is very likely to select the sub-optimal model estimated by the original minimum, especially for the practical time series. We proposed nonlinear curvefitting to catch the true trend of the original curve and obtain the smooth one. We found that the minimum description length estimated by the fitted curve can determine the correct optimal model but the original minimal point may make the mistake on model selection. To demonstrate the effectiveness of the proposed methods, we compared them to the techniques of Bayesian regularization and early stopping. We applied the two techniques to build models of the Rössler system and Ikeda map, and compared the results. We found that the Bayesian algorithm and early stopping can somewhat improve the generalization of models but cannot make the model work as well as the optimal models estimated by our methods.

In Chapter 3, we investigated the relationship between measurements of pulse waveform on the wrist and fingertip by using our optimal neural networks modelling and the surrogate data method. Neural networks with different neurons are used to make the prediction from the pulse on the wrist to the counterpart on the fingertip. Among these models, we selected the optimal model estimated by the minimum description length of the fitted curve. Based on the previous results we are inclined to accept that this optimal model exactly captures the nonlinear dynamics between the two kinds of pulse signals. We then applied the surrogate data method with the given hypothesis of NH0 to confirm the residual of the optimal model is consistent with the i.i.d. noise. There are two significant consequences of such results. Firstly, we confirmed that the optimal model can accurately model the relation between pulse waveforms on the wrist and the fingertip, i.e., our information theoretic criterion works well on model selection for this time series prediction. Secondly, we concluded that it was equivalent to measure the pulse waveforms on the wrist and fingertip. For comparison we artificially added the observational "noise" (x-component data of the Rössler system) to the pulse data of the fingertip under the same hypothesis and repeated the experiment. At this time the optimal model kept

the added dynamics in its residual, which was tested by the surrogate data method. This result indicated that once deterministic deviation exists in the pulse data on the fingertip (in other words, there is significant difference between the pulse data on the wrist and fingertip), combination of the surrogate technique and optimal model modelling can identify it.

In Chapter 4, we extended to investigate the relationship between the ECG data and pulse data on the fingertip. Different from pulse measurements on the wrist and fingertip, ECG and pulse are measured in distinct ways, although they both stem directly from the human heart. An electrocardiogram records the electrical activity of the heart whereas pulse is the measurement of blood pressure propagation. We employed the pseudo-periodic surrogate method to confirm they both do not conform with periodic orbits. In this work we proposed complexity as the test statistic of the surrogate data method. Because ECG signals that are measured with surface electrodes are usually contaminated with noise, and correlation dimension estimations are sensitive to noise. However, complexity is robust to the noisy signals, and very sensitive to the intrinsic deterministic dynamics. For the PPS method we focused on adjusting its noise level over the full-range scale (i.e. the probability from zero to one). We thus can observe the trends of ECG and pulse data with increasing noise levels. When we applied this method to periodic and chaotic Rössler data, and periodic and chaotic Chua circuit data, the trends of ECG and pulse data are consistent with those of chaotic Rössler data and Chua circuit data, rather than those of periodic ones. Such evidence indicates the human normal cardiac data (ECG and pulse) is consistent with deterministic chaos. For the prediction from ECG to the pulse, the optimal model can exactly capture the underlying dynamics of pulse data (by using the small shuffled surrogate data method the model residual is validated to be consistent with independently distributed noise) while the prediction from pulse to ECG data, even the result of the optimal model, is poor. Hence, we concluded that ECG and pulse both conform to deterministic processes but the ECG data comprises certain deterministic components, which the pulse data can not replicate or does not contain. In addition, application of transfer entropy showed higher

information flow from ECG to pulse signal than vice versa. This result supported the conclusion taken from the previous prediction modelling.

In Chapter 5, we presented our methods of modelling nonlinear dynamics to an interesting time seiries, simulated gravitational waves. We demonstrated complexity and the surrogate data method to identify the deterministic gravitational signals contimated with strong white or coloured noise. Complexity can identify the existence of simulated gravitational waves in much lower SNR than the popular methods, such as FFT. The surrogate data method was used to attach statistical significance to the result of complexity. Meanwhile, the surrogate data method with complexy as its test statistic also determined whether complexity indicated significant nonlinear determinism in this data. But autocorrelation failed to distinguish the GW data contaminated with coloured noise and its surrogates. Subsequently, we employed complexity to locate GW bursts in a noisy GW data. We found that local minima of complexity exactly matched the location of these bursts. These methods we proposed can be seen as an improvement to the sensitivity of GW data detection. We also made improvements on the high performance vibration isolators for gravitational wave detection so as to prevent electrical wiring disturbance. The telemetry system we designed was fed with radiation power, and conveyed information by infrared communication so it was completely isolated from wire connection outside and consequently avoided noise coupling through wiring in the advanced vibration isolators.

6.3 Future Work

We succeeded in developing the minimum description length method to select model size of feedforward multi-layer neural networks. Such optimal neural networks estimated by MDL can provide adequate fitting and exactly capture the nonlinear dynamics of diverse time series. We then employ our nonlinear modelling in conjunction with several surrogate data methods (including the PPS method and small

shuffled surrogate data method) to human cardiac time series [1], and conclude on these experimental results. Based on the achievements of these works we intend to substantially extend the studies on model selection, cardiac time series, and interferometric data.

6.3.1 Nonlinear Dynamics in "Feeling the Pulse"

Using models to interpret the relationship between ECG and pulse signals is significant but not adequate. Because pulse signals are used for various unhealthy diagnoses in TCM and ECG signals are mainly measured for cardiac medical diagnosis in modern medecine. A deeper explanation of why the ECG signals can dominate the pulse signals will be the focus of the near future research plan. This work would draw much more appreciation from wider communities.

This research plan will also lay the foundation to provide theoretical evidence to confirm the utility of feeling the pulse and the certain relationship between cardiac measurements (including ECG) of modern medicine and feeling the pulse of traditional Chinese medicine. We intend to understand the correspondence between the various distinct pulses described in traditional Chinese medicine so as to classify their dynamics or identify aberrant cardiovascular condition. The ultimate aim of this research is to develop an expert system which can provide diagnosis and prognosis of a patient's physiological state based on analysis of either pulse waveform or cardiac electrical potential, such as ECG.

6.3.2 Minimum Description Length in Broader Nonlinear Modelling

Another significant research direction is to extend this project to study and model gravitational waves. At this moment the gravitational wave data has been used as a test system for our methods. We intend to model the dynamics of gravitational waves. According to the characteristic shape of gravitational data we propose

[1] Biomedicine is one of preferred selection but the combination is also applicable to other real time series.

wavelets, rather than standard neural networks to model it. This then raises the problem of how to determine the optimal wavelet model. We preliminarily propose the mothod of minimum description length to select optimal models in term of the past successful experiences. This work will involve improvements of nonlinear modelling.

6.3.3 Investigation of Nonlinear Dynamical Methods in Real Interferometric Data

In addition, the gravitational data used in the current work is simulated data. It has proved that complexity possesses ability to identify the existence of weak gravitational data contaminated with noise. Compared to existing techniques, algorithmic complexity has the great advantage of small computational cost and is well suited for real-time implementation in practical situation. We consider to apply our techniques to real interferometric data when they are available and keep on improving the sensitivity of these techniques.

BIBLIOGRAPHY

[1] M. Henon, "A two-dimensional mapping with a strange attractor," *Communications in Mathematical Physics*, vol. 50, pp. 69-77, 1976.

[2] A.I. Mees, eds. *Nonlinear dynamics and statistics*, Boston: Birkhäuser, 2001.

[3] H.G. Solari, M.A. Natiello, and G.B. Mindlin, *Nonlinear dynamics : a two-way trip from physics to math*, Bristol: Institute of Physics, 1996.

[4] A. Kazunori, *Nonlinear dynamics and chaos in semiconductors*, Phailadelphia : Institute of Physics, 2001.

[5] A. Beuter, eds. *Nonlinear dynamics in physiology and medicine*, New York : Springer, 2003.

[6] M. Smith, *Neural Networks for Statistical Modeling.* Boston: International Thomson Computer, 1996.

[7] S. Geman, E. Bienenstock, and R. Doursat, "Neural networks and the bias/variance dilemma," Neural Computation, vol. 4, pp. 1-58, 1992.

[8] A. Weigend, "On overfitting and the effective number of hidden units," in *Proceedings of the 1993 Connectionist Models Summer School*, pp. 335-342, 1994.

[9] J.E. Moody, S.J. Hanson, and R.P. Lippmann, "The effective number of parameters: an analysis of generalization and regularization in nonlinear learning systems," *Advances in Neural Information Processing Systems*, vol. 4, pp. 847-854, 1992.

[10] D.J.C. MacKay, "Bayesian interpolation," *Neural Computation*, vol. 4, no. 3, pp 415-447, 1992.

[11] J. Rissanen, *Stochastic complexity in statistical inquiry*. Singapore: World Scientific, 1989.

[12] M. Li and P. Vitányi, *An Introduction of Kolmogorov and Its Applications (second edition)*. New York: Springer-Verilog, 1997.

[13] J. Rissanen, "Modeling by the shortest data description," *Automatica*, vol. 14, pp. 465-471, 1978.

[14] L. Xu, "Advances on BYY harmony learning: information theoretic perspective, generalized projection geometry, and independent factor auto-determination," *IEEE Transactions on Neural Networks*, vol. 15, no. 4, pp. 885-902, 2004.

[15] X.M. Gao, S. J. Ovaska,, M. Lehtokangas, and J. Saarinen, "Modeling of speech signals using an optimal neural network structure based on the PMDL principle," *IEEE Transactions on Speech and Audio Processing*, vol. 6, no. 2, pp. 177-180, 1998.

[16] K. Judd and A. Mees, "On selecting models for nonlinear time series," *Physica D*, vol. 82, pp. 426-444, 1995.

[17] M. Small and C.K. Tse, "Minimum description length neural networks for time series prediction," *Physical Review E*, vol. 66, pp. 066701, 2002.

[18] H. Akaike, "A new look at the statistical model identification," *IEEE Transactions on Automatic Control*, vol. AC-19, no. 6, 1974.

[19] M. Stone, "Comments on model selection criteria of Akaike and Schwarz," *Journal of the Royal Statistical Society, B*, vol. 41, no. 2, pp. 276-278, 1979.

[20] J. Rice, "Bandwidth choice for nonparametric regression," *Annals of Statistics*, vol. 12, no. 4, pp. 1215-1230, 1984.

[21] Z. Liang, R.J. Jaszczak, and R.E. Coleman, "Parameter estimation of finite mixtures using the EM algorithm and information criteria with application to medical image processing," *IEEE Transactions on Nuclear Science*, vol. 39, no. 4, pp. 1126-1133, 1992.

[22] C. Wallace and D. Boulton, "An information measure for classification," *Computing Journal*, vol. 11, pp. 185-195, 1968.

[23] A. Barron and J. Rissanen, "The minimum description length principle in coding and modeling," *IEEE Transactions on Information Theory*, vol. 44, pp. 2743-2760, 1998.

[24] X. Hu and L. Xu, "A comparative investigation on subpace dimension determination," *Neural Networks*, vol. 17, pp. 1051-1059, 2004.

[25] L. Xu, "Ying-Yang learning," *The Handbook of Brain Theory and Neural Networks (2nd edition)*. Massachusettes: MIT, 2002.

[26] L. Xu, "Bayesian Ying Yang learning (II): a new mechanism for model selection and regularization," *Intelligent Technologies for Information Analysis, Springer*, pp. 661-706, 2004.

[27] E.A. Wan, *Time Series Prediction: Forecasting the Future and Understanding the Past*, vol. XV of *Studies in the Sciences of Complexity*, edited by A. Weigend and N. Gershenfeld, MA: Addison-Wesley Reading, pp. 195-217, 1993.

[28] D.E Rumelhart, G.E. Hinton, and R.J. Williams, "Learning internal representations by error propagation,"vol. 1 of *Parallel Data Processing* edited by D. Rumelhart and J. McClelland, MA: M.I.T. Press, pp. 318-362, 1986.

[29] M. Small, "Nonlinear dynamics in infant respiration," PhD thesis, The University of Western Australia, Department of Mathematics, Perth, Australia, 1998.

[30] T. Nakamura, "Modeling nonlinear time series using selection methods and information criteria," PhD thesis, The University of Western Australia, Departement of Mathematics, Perth, Australia, 2003.

[31] M. Small and K. Judd, "Comparison of new nonlinear modeling techniques with applications to default repiration," *Physica D*, vol. 117, pp. 283-298, 1998.

[32] M.B. Kennel, R. Brown, and H.D.I. Abarbanel, "Determining embedding dimension for phase-space reconstruction using a geometrical construction," *Physical Review A*, vol. 45, pp. 3403, 1992.

[33] H. Kantz and T. Schreiber *Nonlinear Time Series Analysis*, New York: Cambridge University, pp. 30-36, 2004.

[34] K. Judd and A. Mees, "Embedding as a modeling problem," *Physica D*, vol. 120, pp. 273-286, 1998.

[35] M. Small and C.K. Tse, "Optimal embedding: a modeling paradigm," *Physcia D*, vol. 194, pp. 283-296, 2004.

[36] H.D.I. Abarbanel, *Analysis of observed chaotic data*, New York: Springer, pp. 39-45, 1996.

[37] A.S. Weigend and N.A. Gershenfeld, eds. *Time series prediction: forecasting the future and understanding the past*, MA: Addison-Wesley, 1994.

[38] M. Small, D. Yu, N. Grubb, J. Simonotto, K. Fox, and R.G. Harrison, "Automatic identification and recording of cardiac arrhythmia," *Computers in Cardiology*, vol. 27, pp. 355-358, 2000.

[39] O.E. Rössler, "An equation for continous chaos," *Physics Letter A*, vol. 57, pp. 397, 1976.

[40] M. Small, D.J. Yu, and R.G. Harrison, " Observation of a period doubling bifurcation during onset of human ventricular fibrillation," *International Journal of Bifurcation and Chaos*, vol. 13, no. 33, pp. 743-754, 2003.

[41] M.T. Hagan and M. Menhaj, "Traning feedforward networks with the marquardt algorithm," *IEEE Transactions on Neural Networks*, vol. 5, no. 6, pp. 989-993, 1994.

[42] J. Theiler, S. Eubank, A. Longtin, B. Galdrikian, and J.D. Farmer, "Testing for nonlinearly in time series: The method of surrogate data," *Physica D*, vol. 58, pp. 77-94, 1992.

[43] http://small.eie.polyu.edu.hk; www.koreanmedicine.org/english/lecture_2_4.html

[44] A. Galka, *Topics in nonlinear time series analysis with implications for EEG analysis*. Singapore: World Scientific, pp. 222, 2000.

[45] T. Schreiber and A. Schmitz, "Improved surrogate data for nonlinearity tests," *Physical Review Letters*, vol. 77, no. 4, pp. 635-638, 1996.

[46] T. Nakamura, X. Luo, and M. Small, "Testing for nonlinearity in time series without the Fourier transform," *Physical Review E*, vol. 72, pp. 055201(R), 2005.

[47] M. Small, D.J. Yu, and R.G. Harrison, "A surrogate test for pseudo-periodic time series data," *Physical Review Letters*, vol. 87, pp. 188101, 2001.

[48] J. Theiler, "On the evidence for low-dimensional chaos in an epileptic electroencephalogram," *Physics Letter A*, vol. 196, pp. 335-341, 1995.

[49] P. Grassberger and I. Procaccia, "Measuring the strangeness of strange attractors," *Physica D*, vol. 9, pp. 189, 1983.

[50] C. Diks, "Estimating invariants of noisy attractors," *Physical Review E*, vol. 52, no. 5, pp. R4263-R4266, 1996.

[51] 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," *Physical Review E*, vol. 61, pp. 3750, 2000.

[52] F. Ravelli and R. Antolini, "Complex dynamics underlying the human electrocardiogram," *Biological cybernetics*, vol. 67, pp. 57-65, 1992

[53] A. Casseleggio and A. Carana, "Correlation dimension estimation from electrocardiograms," *Chaos Solitons and Fractals*, vol. 5, pp. 713-726, 1995.

[54] M.I. Owis, A.H. Abou-Zied, Abou-Bakr M. Youssef, and Y.M. Kadah, "Study of features based on nonlinear dynamical modeling in ECG arrhythmia detection and classification," *IEEE Transaction on Biomedical Engineering*, vol. 49, no. 7, pp. 733-736, 2002.

[55] R.B. Govindan, K. Narayanan, and M.S. Gopinathan, "On the evidence of deterministic chaos in ECG: Surrogate and predictability analysis," *Chaos*, vol. 8, no. 2, pp. 495-502, 1998.

[56] A. Wolf, J.B. Swift, H.L. Swinney, and J.A. Vastano, "Determining Lyapunov exponents from a time series," *Physica D*, vol. 16, pp. 285-317, 1985.

[57] M. Small, D.J. Yu, J. Simonotto, R.G. Harrison, N. Grubb, and K.A.A, "Uncovering non-linear structure in human ECG recongdings," *Chaos Solitons and Fractals*, vol. 13, pp. 1755-1762, 2002.

[58] H.X. Zhang, Y. Zhu, and Y. Xu, "Complexity information based analysis of pathological ECG rhythm for ventricular tachycardia and ventricular fibrillation," International Journal Bifurcation and Chaos, vol. 12, no. 10, pp. 2293-2303, 2002.

[59] M.J. Reed, G.R. Clegg, and C.E. Robertson, "Ananlysing the ventricular fibrillation waveform," *Resuscitation*, vol. 57, pp. 11-20, 2003.

[60] S. Eyal, Y. Almog, O. Oz, S. Eliash, S. Akselrod, "Nonlinear dynamics applied to blood pressure control," *Autonomic Neuroscience: Basic and Clinical*, vol. 89, pp. 24-30, 2001.

[61] B. Flaws, *The Secret of Chinese Pulse Diagnosis*, Columbia: Blue Poppy Press, 1995.

[62] M. Small and C.K. Tse, "Applying the method of surrogate data to cyclic time series," *Physica D*, vol. 164, pp. 187-201, 2002.

[63] H. Kantz and T. Schreiber, "Human ECG: nonlinear deterministic versus stochastic aspects," *IEE Proceeding of Science, Measurement and Technology*, vol. 145, no. 6, pp. 279-284, 1998.

[64] J.P. Eckmann and D. Ruelee, "Fundamental limitations for estimating dimensions and Lyapunov exponents in dynamical systems," *Physica D*, vol. 56, pp. 185-187, 1992.

[65] T. Schreiber and D.T. Kaplan, "Nonlinear noise reduction for electrocardiograms," *Chaos*, vol. 6, pp. 87, 1996.

[66] N. Radhakrishnan and B.N. Gangadhar, "Estimating regularity in epileptic seizure time-series data," *IEEE Engineering in Medicine and Biology Magazine*, vol. 17, no. 3, pp. 89-94, 1998.

[67] Y. Zhao, M. Small, D. Coward, E. Howell, C. Zhao, L. Ju, and D.G. Blair, "Identifying deterministic signals in simulated gravitational wave data: algorithmic complexity and the surrogate data method," *Classical and Quantum Gravity*, vol. 23, pp. 1801-1814, 2006.

[68] Y. Zhao and M. Small, "Equivalence between "feeling the pulse" on the human wrist and pulse pressure wave at fingertip," *International Journal of Neural Systems*, vol. 15, no. 4, pp. 277-286, 2005.

[69] Y. Zhao and M. Small, "Minimum description length criterion for modeling of chaotic attractors with multi-layer perceptron networks," *EEE Transactions on Circuits and System Part I: Fundamental Theory and Applications*, vol. 53, no. 3, pp. 722-731, 2006.

[70] Y. Zhao and M. Small, "Deterministic propagation of blood pressure waveform from human wrists to fingertips," *Lecture notes in Computer Science*, vol. 3177, pp. 142-147, 2004.

[71] Y. Zhao, C. Zhao, L. Ju, M. Small, and D. Blair, "Telemetry system driven by radiation power for use in gravitational wave detectors," *Review of Scientific Instruments*, vol. 76, no. 8, pp. 084503, 2005.

[72] F. Takens, "Detecting Strange Attractors in Turbulenc," *Lecture Notes on Mathematica*, vol. 898, pp. 366, 1981.

[73] A.M. Albano, J. Muench, C. Schwartz, A.I. Mees, and P.E. Rappet, "Singular value decomposition and the Grassberger-Procaccia algorithm," *Physical Review A*, vol. 38, pp. 3017, 1988.

[74] A. Lempel and J. Ziv, "On the Complexity of Finite Sequences," *IEEE Transaction on Information Theory*, vol. 22, no. 1, pp. 75-81, 1976.

[75] M. Small, *Applied nonlinear time series analysis*, New York: World Scientific press, pp. 125-128, 2005.

[76] T. Nakamura and M. Small, "Small-shuffle surrogate data: Testing for dynamics in fluctuating data with trends ," *Physical Review E*, vol. 72, pp. 056216, 2005.

[77] T. Schreiber, "Measuring information transfer," *Physical Review Letter*, vol. 85, pp. 461-464, 2000.

[78] A. Kaiser and T. Schreiber, "Information transfer in continous processes," *Physica D*, vol. 166, pp. 43-62, 2002.

[79] C.E. Shannon and W. Weaver, *The mathematical theory of informations*. Urbana: University of Illinois, 1949.

[80] http://www.einsteinathome.org/gwaves/predict/index.html.

[81] http://www.einsteinathome.org/gwaves/predict/flaws.html.

[82] http://www.einsteinathome.org/gwaves/predict/general.html.

[83] http://www.einsteinathome.org/gwaves/sources/index.html.

[84] http://www.einsteinathome.org/gwaves/predict/space.html.

[85] Z. A. Allen *et al.*, "First Search for Gravitational Wave Bursts with a Network of Detectors", *Physical Review Letters*, vol. 85, pp. 5046C5050, 2000.

[86] M. Cerdonio, L. Conti, J.A. Lobo, A. Ortolan, L. Taffarello, and J. P. Zendri, "Wideband Dual Sphere Detector of Gravitational Waves," *Physical Review Letters*, vol. 87, no. 3, pp. 0301102, 2001.

[87] B. Abbott et al.,"Detector description and performance for the first coincidence observations between LIGO and GEO.," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 517, pp. 154-179, 2004.

[88] D. Sigg, "Commissioning of LIGO detectors," *Classical and Quantum Gravity*, vol. 21, pp. S409, 2004.

[89] F. Acernese et al., "Status of VIRGO ," *Classical and Quantum Gravity*, vol. 21, pp. S385, 2004.

[90] B. Willke et al., "Status of GEO 600," *Classical and Quantum Gravity*, vol. 21, pp. S417, 2004.

[91] R. Takahashi et al., "Status of TAMA300," *Classical and Quantum Gravity*, vol. 21, pp. S403, 2004.

[92] D.M. Coward, R.R. Burman, and D.G. Blair, "The stochastic background of gravitational waves from neutron star formation at cosmological distances," *Monthly Notices of the Royal Astronomical Society*, vol. 324, pp. 1015, 2001.

[93] D.M. Coward, R.R. Burman, and D.G. Blair, "Statistical characteristics of a stochastic background of gravitational waves from neutron star formation at cosmological distances," *Classical and Quantum Gravity*, vol. 19, pp. 1303, 2002.

[94] D.M. Coward, H.P.M. Putten, and R.R. Burman, "Proposed gravitational wave background from black hole-torus systems," *The Astrophysical Journal*, vol. 580, pp. 1024, 2002.

[95] E. Howell, D.M. Coward, R.R. Burman, and D.G. Blair, "Fast temporal evolution of a cosmic gravitational wave background spectrum," *Classical and Quantum Gravity*, vol. 22, pp. 723, 2005.

[96] E. Cuoco, G. Calamai, L. Fabbroni, G. Losurdo, M. Mazzoni, R. Stanga, and F. Vetrano, "On-line power spectra identification and whitening for the noise in interferometric gravitational wave detectors," *Classical and Quantum Gravity*, vol. 18, pp. 1727, 2001.

[97] S. Mukherjee, "Interferometric data modelling: issues in realistic data generation," *Classical and Quantum Gravity*, vol. 21, pp. S1783, 2004.

[98] T. Zwerger, E. Müller, "Dynamics and gravitational wave signature of axisymmetric rotational core collapse," *Astronomy and Astrophysics*, vol. 320, pp. 209-227, 1997.

[99] E.E. Flanagan and S.A. Hughes, "Measuring gravitational waves from binary black hole coalescences. I. Signal to noise for inspiral, merger, and ringdown," *Physical Review D*, vol. 57, pp. 4535-4565, 1998.

[100] W.G. Anderson, P.R. Brady, J.D.E. Creighton, and E.E. Flanagan, "A power filter for the detection of burst sources of gravitational radiation in interferometric detectors," *International Journal of of Modern Physics D [Gravitation; Astrophysics and Cosmology]*, vol. 9, pp. 303, 2000.

[101] W.G. Anderson, P.R. Brady, J.D.E. Creighton, and E.E. Flanagan, "Excess power statistic for detection of burst sources of gravitational radiation," *Physical Review D*, vol. 63, pp. 042003, 2001.

[102] W.G. Anderson and R. Balasubramanian, "Time-frequency detection of gravitational waves," *Physical Review D*, vol. 60, pp. 102001, 1999.

[103] J. Sylvestre, "Time-frequency detection algorithm for gravitational wave bursts," *Physical Review D*, vol. 66, pp. 102004, 2002.

[104] N. Arnaud, F. Cavalier, M. Davier, and P. Hello, "Detection of gravitational wave bursts by interferometric detectors," *Physical Review D*, vol. 59, pp. 082002, 1999.

[105] T. Pradier, N. Arnaud, M. Bizouard, F. Cavalier, M. Davier, and P. Hello, "Efficient filter for detecting gravitational wave bursts in interferometric detectors," *Physical Review D*, vol. 63, pp. 042002, 2001.

[106] N. Arnaud, M. Barsuglia, M. Bizouard, V. Brisson, F. Cavalier, M. Davier, P. Hello, S. Kreckelbergh, E.K. Porter, and T. Pradier, "Comparison of filters for detecting gravitational wave bursts in interferometric detectors," *Physical Review D*, vol. 67, pp. 062004, 2003.

[107] F. Beauville et al, "A first comparison of search methods for gravitational wave bursts using LIGO and Virgo simulated data," *Classical and Quantum Gravity*, vol. 22, pp. S1293-S1301, 2005.

[108] The VIRGO collaboration, "Data analysis methods for non-Gaussian, nonstationary and nonlinear features and their application to VIRGO," *Classical and Quantum Gravity*, vol. 20, pp. S915-S924, 2003.

[109] F. Kaspar and H.G. Schuster, ",Easily calculable measure for the complexity of spatiotemporal patterns" *Physical Review A*, vol. 36, pp. 842-848, 1987.

[110] M. Maggiore, "Stochastic backgrounds of gravitational waves," *Physics Reports*, vol. 331, pp. 283-367, 2000.

[111] K.S. Thorne, *Three hundred years of gravitation* London: Cambridge University Press), pp. 368, 1989.

[112] G.E.P. Box and G.M. Jenkins, *Time Series Analysis: Forecasting and Control* CA: Holden-Day, 1976.

[113] D. Shoemaker, R. Schilling, L. Schnupp, W. Winkler, K. Maischberger, and A. Rudiger, "Noise behavior of the Garching 30-meter prototype gravitational-wave detector," *Physical Review D*, vol. 38, pp. 423, 1987.

[114] F.D. Foresee and M.T. Hagan, "Gauss-Newton approximation to Bayesian regularization," *Proceedings of the 1997 International Joint Conference on Neural Networks*, pp. 1930-1935, 1997.