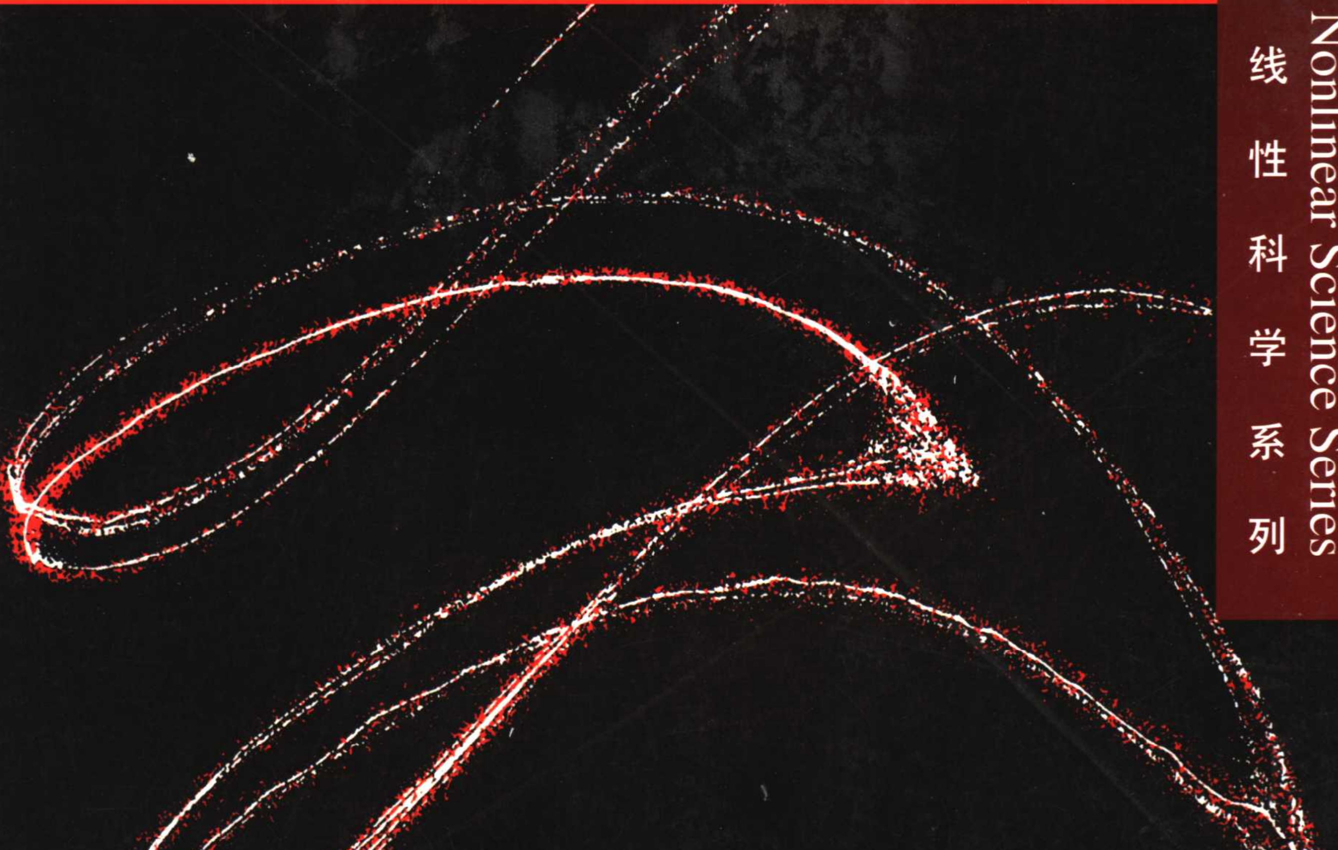
Nonlinear Time Series Analysis

非线性时间序列分析

Holger Kantz
Thomas Schreiber 著

Cambridge Nonlinear Science Series

剑桥非线性科学系列



清华大学出版社
CAMBRIDGE UNIVERSITY PRESS

(京)新登字 158 号

书　名:非线性时间序列分析(影印版)
作　者:Holger Kantz, Thomas Schreiber
出版者:清华大学出版社(北京清华大学学研大厦,邮编 100084)
　　　　http://www.tup.tsinghua.edu.cn
印刷者:清华大学印刷厂
发行者:新华书店总店北京发行所
开　本:787×1092 1/16　印张:20
版　次:2000 年 6 月第 1 版　2001 年 3 月第 3 次印刷
书　号:ISBN 7-302-03906-2/O・240
印　数:1501~3500
定　价:39.00 元

Deterministic chaos offers a striking explanation for irregular behaviour and anomalies in systems which do not seem to be inherently stochastic. The most direct link between chaos theory and the real world is the analysis of time series from real systems in terms of nonlinear dynamics. This book provides experimentalists with methods for processing, enhancing, and analysing measured signals using these methods; and for theorists it demonstrates the practical applicability of mathematical results.

The framework of deterministic chaos constitutes a new approach to the analysis of irregular time series. Traditionally, nonperiodic signals have been modelled by linear stochastic processes. But even very simple chaotic dynamical systems can exhibit strongly irregular time evolution without random inputs. Chaos theory offers completely new concepts and algorithms for time series analysis which can lead to a thorough understanding of the signals. The book introduces a broad choice of such concepts and methods, including phase space embeddings, nonlinear prediction and noise reduction, Lyapunov exponents, dimensions and entropies, as well as statistical tests for nonlinearity. Related topics such as chaos control, wavelet analysis, and pattern dynamics are also discussed. Applications range from high-quality, strictly deterministic laboratory data to short, noisy sequences which typically occur in medicine, biology, geophysics, and the social sciences. All the material discussed is illustrated using real experimental data.

This book will be of value to any graduate student and researcher who needs to be able to analyse time series data, especially in the fields of physics, chemistry, biology, geophysics, medicine, economics, and the social sciences.

Preface

The paradigm of deterministic chaos has influenced thinking in many fields of science. As mathematical objects, chaotic systems show rich and surprising structures. Most appealing for researchers in the applied sciences is the fact that deterministic chaos provides a striking explanation for irregular behaviour and anomalies in systems which do not seem to be inherently stochastic.

The most direct link between chaos theory and the real world is the analysis of time series from real systems in terms of nonlinear dynamics. On the one hand, experimental technique and data analysis have seen such dramatic progress that, by now, most fundamental properties of nonlinear dynamical systems have been observed in the laboratory. On the other hand, great efforts are being made to exploit ideas from chaos theory in cases where the system is not necessarily deterministic but the data displays more structure than can be captured by traditional methods. Problems of this kind are typical in biology and physiology but also in geophysics, economics, and many other sciences.

In all these fields, even simple models, be they microscopic or phenomenological, can create extremely complicated dynamics. How can one verify that one's model is a good counterpart to the equally complicated signal that one receives from nature? Very often, good models are lacking and one has to study the system just from the observations made in a single time series, which is the case for most nonlaboratory systems in particular. The theory of nonlinear dynamical systems provides new tools and quantities for the characterisation of irregular time series data. The scope of these methods ranges from invariants such as Lyapunov exponents and dimensions which yield an accurate description of the structure of a system (provided the data are of high quality) to statistical techniques which allow for classification and diagnosis even in situations where determinism is almost lacking.

This book provides the experimental researcher in nonlinear dynamics with methods for processing, enhancing, and analysing the measured signals. The theorist will be offered discussions about the practical applicability of mathematical results. The time series analyst in economics, meteorology, and other fields will find inspiration for the development of new prediction algorithms. Some of the techniques presented here have also been considered as possible diagnostic tools in clinical research. We will adopt a critical but constructive point of view, pointing out ways of obtaining more meaningful results with limited data. We hope that everybody who has a time series problem which cannot be solved by traditional, linear methods will find inspiring material in this book.

Dresden and Wuppertal
November 1996

Acknowledgements

If there is any feature of this book that we are proud of, it is the fact that almost all the methods are illustrated with real, experimental data. However, this is anything but our own achievement – we exploited other people's work. Thus we are deeply indebted to the experimental groups who supplied data sets and granted permission to use them in this book. The production of every one of these data sets required skills, experience, and equipment that we ourselves do not have, not forgetting the hours and hours of work spent in the laboratory. We appreciate the generosity of the following experimental groups:

NMR laser. Our contact persons at the Institute for Physics at Zürich University were Leci Flepp and Joe Simonet; the head of the experimental group is E. Brun. (See Appendix C.2.)

Vibrating string. Data were provided by Tim Molteno and Nick Tufillaro, Otago University, Dunedin, New Zealand. (See Appendix C.3.)

Taylor–Couette flow. The experiment was carried out at the Institute for Applied Physics at Kiel University by Thorsten Buzug and Gerd Pfister. (See Appendix C.4.)

Atrial fibrillation. This data set is taken from the MIT-BIH Arrhythmia Database, collected by G. B. Moody and R. Mark at Beth Israel Hospital in Boston. (See Appendix C.6.)

Human ECG. The ECG recordings we used were taken by Petr Saparin at Saratov State University. (See Appendix C.7.)

Foetal ECG. We used noninvasively recorded (human) foetal ECGs taken by John F. Hofmeister at the Department of Obstetrics and Gynecology, University of Colorado, Denver CO. (See Appendix C.7.)

Phonation data. This data set was made available by Hanspeter Herzel at the Technical University in Berlin. (See Appendix C.8.)

Autonomous CO_2 laser with feedback. The data were taken by Riccardo Meucci and Marco Ciofini at the INO in Firenze, Italy. (See Appendix C.10.)

Human posture data. The time series was provided by Steven Boker and Bennett Bertenthal at the Department of Psychology, University of Virginia, Charlottesville VA. (See Appendix C.9.)

We used the following data sets published for the Santa Fe Institute Time Series Contest, which was organised by Neil Gershenfeld and Andreas Weigend in 1991:

Human breath rate. The data we used is part of data set B of the contest. It was submitted by Ari Goldberger and coworkers. (See Appendix C.5.)

NH_3 laser. We used data set A and its continuation, which was published after the contest was closed. The data was supplied by U. Hübner, N. B. Abraham, and C. O. Weiss. (See Appendix C.1.)

During the composition of the text we asked various people to read all or part of the manuscript. The responses ranged from general encouragement to detailed technical comments. In particular we thank Peter Grassberger, James Theiler, Daniel Kaplan, Ulrich Parlitz, and Martin Wiesenfeld for their helpful remarks. Members of our research groups who volunteered to comment on some of the computer programs are Rainer Hegger, Andreas Schmitz, Marcus Richter, and Frank Schmüser.

Finally, we acknowledge support by the Sonderforschungsbereich 237 of the Deutsche Forschungsgemeinschaft.

Last not least we acknowledge the support by Simon Capelin from Cambridge University Press and the excellent help in questions of style and English grammar by Sheila Shepherd.

Contents

Part 1

Basic topics

Chapter 1

Introduction: Why nonlinear methods?

You are probably reading this book because you have an interesting source of data and you suspect it is not a linear one. Either you positively know it is nonlinear because you have some idea of what is going on in the piece of world that you are observing or you are led to suspect that it is because you have tried linear data analysis and it has failed.[1]

Linear methods interpret all regular structure in a data set, such as a dominant frequency, as linear correlations (to be defined in Chapter 2 below). This means, in brief, that the intrinsic dynamics of the system are governed by the linear paradigm that small causes lead to small effects. Since linear equations can only lead to exponentially growing or periodically oscillating solutions, all irregular behaviour of the system has to be attributed to some random external input to the system. Now, chaos theory has taught us that random input is not the only possible source of irregularity in a system's output: nonlinear, chaotic systems can produce very irregular data with purely deterministic equations of motion. Of course, a system which has both, nonlinearity *and* random input, will most likely produce irregular data as well.

Although we have not yet introduced the tools we need to make quantitative statements, let us look at a few examples of real data sets. They represent very different problems of data analysis where one could profit from reading this book since a treatment with linear methods alone would be inappropriate.

1. Of course you are also welcome to read this book if you are not working on a particular data set.

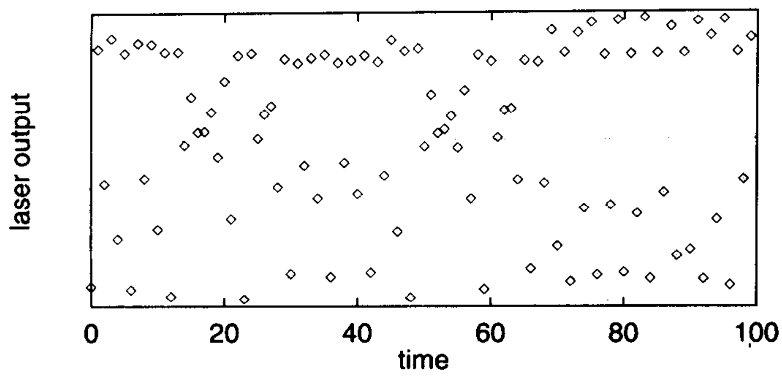


Figure 1.1 100 successive measurements of the laser intensity of an NMR laser. The time unit here is set to the measurement interval.

Example 1.1 (NMR laser data) In a laboratory experiment carried out by Flepp, Simonet & Brun at the Physics Department of the University of Zürich, a Nuclear Magnetic Resonance laser is operated under such conditions that the amplitude of the (radio frequency) laser output varies irregularly over time. From the set-up of the experiment it is clear that the system is highly nonlinear and random input noise is known to be of very small amplitude compared to the amplitude of the signal. Thus it is not assumed that the irregularity of the signal is just due to input noise. In fact, it has been possible to model the system by a set of differential equations which does not involve any random components at all; see Flepp *et al.* (1991). Appendix C.2 contains more details about this data set.

Successive values of the signal appear to be very erratic, as can be seen in Fig. 1.1. Nevertheless, as we shall see later, it is possible to make accurate forecasts of future values of the signal using a nonlinear prediction algorithm. Fig. 1.2 shows the mean prediction error depending on how far into the future the forecasts are made. Quite intuitively, the further into the future the forecasts are made, the larger will the uncertainty be. After about 35 time steps the prediction becomes worse than when just using the mean value as a prediction (horizontal line). We have fitted the growth by an exponential with an exponent of 0.3, which is indicated as a dashed line. We used the simple prediction method which will be described in Section 4.2. For comparison we also show the result for the best *linear* predictor that we could fit to the data (crosses). We observe that the predictability due to the linear correlations in the data is much weaker than the one due to the deterministic structure, in particular for short prediction times. The predictability of the signal can be taken as a signature of the deterministic nature of the system. See Section 2.5 for details on the linear prediction method used. The nonlinear structure which leads to the short-term predictability in the data set is not apparent in a representation such as Fig. 1.1. We can, however, make it visible by plotting each data point versus its

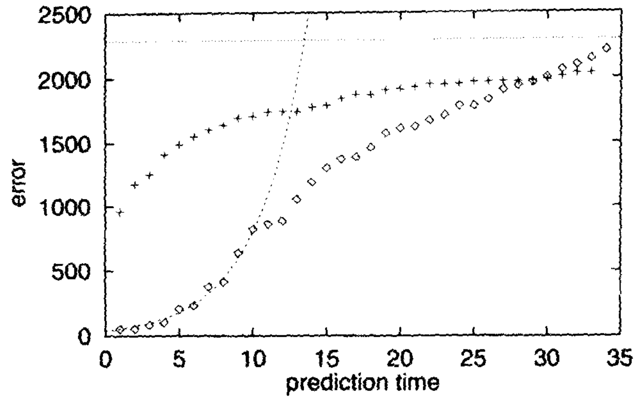


Figure 1.2 The average prediction error (in units of the data) for a longer sample of the NMR laser output as a function of the prediction time. For an explanation of the different symbols see the text of Example 1.1.

predecessor, as has been done in Fig. 1.3. Such a plot is called a *phase portrait*. This representation is a particularly simple application of a basic tool which will often be used in nonlinear time series analysis, the *time delay embedding*. This concept will be formally introduced in Section 3.2. In the present case we just need a data representation which is printable in two dimensions. □

Example 1.2 (Human breath rate) One of the data sets used for the Santa Fe Institute time series competition in 1991–92 [Weigend & Gershenfeld (1993)] was provided by A. Goldberger from Beth Israel Hospital in Boston [Rigney *et al.* (1993); see also Appendix C.5]. Out of several channels we selected a 16 min record of the air flow through the nose of a human subject. A plot of the data segment we used is shown in Fig. 1.4.

In this case only very little is known about the origin of the fluctuations of the breath rate. The only hint that nonlinear behaviour plays a role comes from the data itself: the signal is not compatible with the assumption that it is created by a Gaussian random process with only linear correlations (possibly distorted by a nonlinear measurement function). This we show by creating an artificial data set which has exactly the same *linear* properties but has no further determinism built in. This data set consists of random numbers which have been rescaled to the distribution of the values of the original (thus also mean and variance are identical) and filtered so that the power spectrum is the same. (How this is done, and further aspects, are discussed in Chapter 4 and Section 7.1.) If the measured data are properly described by a linear process we should not find any significant differences from the artificial ones.