

DETECTION OF SIGNALS IN CHAOS

By

XIAO BO LI, B. Eng., M. Eng.

A Thesis

Submitted to the School of Graduate Studies
in Partial Fulfilment of the Requirements
for the Degree
Ph.D. of Engineering

McMaster University

©Copyright by Xiao Bo Li, August 1995

PH.D. OF ENGINEERING (1995)
(Electrical and Computer Engineering)

MCMASTER UNIVERSITY
Hamilton, Ontario

TITLE: Detection of Signals in Chaos

AUTHOR: Xiao Bo Li
B. Eng. (National University of Def Tech, China)
M. Eng. (University of Electronic Sci & Tech of China)

SUPERVISOR(S): Dr. Simon Haykin, Professor
Department of Electrical and Computer Engineering

NUMBER OF PAGES: xviii, 148

DETECTION OF SIGNALS IN CHAOS



Abstract

A new method for the detection of signals in “noise”, which is based on the premise that the “noise” is chaotic with at least one positive Lyapunov exponent, is presented. The method is naturally rooted in nonlinear dynamical systems, and relies on neural networks for its implementation.

We first present a theoretical basis for methods of modeling the underlying dynamics of a chaotic system using a time series. The subject matter selected for this part of the thesis is written with emphasis on experimental studies of chaos. Specifically, we discuss the issues involved in the reconstruction of chaotic dynamics, attractor dimensions, and Lyapunov exponents. We describe procedures for the estimation of the correlation dimension and the Lyapunov exponents. The method of false nearest neighbors analysis for finding the minimum embedding dimension is described. The need for an adequate data length is stressed.

In the second part of the thesis we apply the chaos-based method to the radar detection of a small target in sea clutter. To justify the applicability of the new method to this problem, we clearly need to show in a convincing way that sea clutter is indeed the result of a chaotic dynamical system. We do this by presenting the results of a detailed experimental study using surface-truthed real-life data collected by means of an instrument-quality radar at different geographic locations. Specifically, we show that (1) sea clutter has a finite correlation dimension, (2) the largest Lyapunov exponent of sea clutter is positive, and (3) sea clutter is locally predictable. Most importantly, we show that both the correlation dimension and the largest Lyapunov exponent are essentially invariant to the choice of radar signal component used to construct the time series, and that the correlation dimension and Lyapunov exponent do not appear to change appreciably with sea state or

with geographic locations. These results suggest that there may exist a universal chaotic structure responsible for the generation of sea clutter. Perhaps the most dramatic result presented in the thesis is the fact that this prior information (*i.e.*, the knowledge that sea clutter exhibits chaotic behavior) can be exploited to build a chaos-based detector operating on amplitude information only (as in a noncoherent marine radar), realizing a performance comparable to that of a "conventional" receiver using coherent radar data (*i.e.*, both amplitude and phase). This result points to the potential of trading off sophisticated but inexpensive computer software for expensive microwave hardware. Lastly, we show experimentally that a chaos-based coherent detector can provide a further improvement in radar detection performance.

Acknowledgement

First of all, I would like to sincerely thank my supervisor, Dr. Simon Haykin, for his support and patient guidance during my graduate studies. Simon has not only provided invaluable feedback, he has also given me the latitude I've needed to develop as a researcher. I would also like to thank the members of my supervisory committee, Dr. P. Yip and Dr. J. M. Chadam, for their helpful comments and advice. I am indebted to Mr. V. Kezys, Mr. B. Currie, and Dr. G. Jones for their help in collecting the experimental data used in this study, to Dr. F. Pineda of the Applied Physics Laboratory, Johns Hopkins University, for providing his software for the computation of the correlation dimension.

To all the researchers and staff of the Communications Research Laboratory, I thank you for the enjoyable time we spent and many interesting discussions we had during my study. To Dr. Andrew Ukrainec, I especially thank you for the help and guidance on computer you so freely offered me. To Mr. Hugh Pasika, I thank you for reading through the first draft of the thesis.

To my parents, brother and sister who have contributed to my successes over the years through their confidence and belief in me, I thank you.

Last but not least, I'd like to thank my wife, Weihui, who patiently supported and encouraged me through my studies.

Contents

List of Tables	x
List of Figures	xviii
1 Introduction	1
1.1 Problem Description	2
1.2 Dissertation Overview	5
2 Nonlinear Dynamical Systems and Chaos	7
2.1 Nonlinear Dynamical Systems	7
2.2 Strange Attractors and Chaos	8
2.3 How Can Chaos Arise?	13
3 Chaos-based Hypothesis Testing	15
3.1 Introduction	15
3.2 Detection Hypothesis	16
3.3 Detection Principle	17
3.4 Computer Experiments	19
3.5 Summary	23

4	Characterization of a Chaotic Process	25
4.1	Introduction	25
4.2	Reconstruction of Chaotic Dynamics	25
4.3	Attractor Dimension	30
4.3.1	Hausdorff Dimension	30
4.3.2	Correlation Dimension	32
4.3.3	False Nearest Neighbors Analysis	35
4.4	Lyapunov Characteristic Exponents	37
4.4.1	Definition	37
4.4.2	Lyapunov Exponents from Dynamical Equations	38
4.4.3	Lyapunov Exponents from a Time Series	40
4.5	Summary	45
5	Experimental Database Description	47
5.1	Introduction	47
5.2	Experiments at Cape Bonavista, Newfoundland	49
5.3	Experiments at Dartmouth, Nova Scotia	51
5.4	Experiments at Lake Ontario	52
5.5	Summary	52
6	Chaotic Characterization of Sea Clutter	53
6.1	Introduction	53
6.2	Correlation Dimension of Sea Clutter	54
6.3	False Nearest Neighbors Analysis	62

6.4	Lyapunov Exponents	67
6.4.1	Largest Lyapunov Exponent	67
6.4.2	Lyapunov Spectrum	71
6.5	Kolmogorov Entropy	74
6.6	Estimation of Sea Clutter Dynamics	76
6.6.1	A Neural Network Model Based on Noncoherent Data	78
6.6.2	Temporal Extent of Local Predictability	85
6.6.3	A Neural Network Model Based on Coherent Data	85
6.7	New Strange Attractors	88
6.8	Summary	99
7	Detection of Small Targets in Sea Clutter	101
7.1	Introduction	101
7.2	CFAR Detectors	102
7.3	Chaos-Based Noncoherent Detector	103
7.4	Chaos-Based Coherent Detector	106
7.5	Effect of Noise on Receiver Performance	112
7.6	Summary	115
8	Chaotic Modeling Using Modular Network	117
8.1	Introduction	117
8.2	Modular Network	118
8.3	Chaotic Modeling of Sea Clutter Using Modular Network	123
8.4	Summary	131

CONTENTS

ix

9 Summary and Recommendations

133

Appendix

137

Bibliography

139

List of Tables

1.1	Comparison of properties of various processes	3
5.1	Radar and sea state parameters for the Cape Bonavista database	50
5.2	Radar and sea state parameters for the Dartmouth database	51
6.1	Lyapunov spectrum for different sea states, where the sea state is classified by wave height	71
6.2	The evolving values of the Lyapunov spectrum shown in Figure 6.21	72
6.3	The Lyapunov exponents of sea clutter for defferent embedding dimensions.	73
6.4	The Lyapunov dimension of sea clutter verse the embedding dimension . . .	74
8.1	Comparison of local predictability of different types of gating network structure in the modular network for sea clutter modeling	129
8.2	Comparison of local predictability of a multilayer perceptron network with a modular neural network for sea clutter modeling	130

List of Figures

2.1	Time waveforms of Equation (2.3), in which the solid curves in both parts of the figure correspond to the initial state of $[0.9, 0.8]$ and the dashed curves correspond to the initial state of $[0.9 + 10^{-12}, 0.8]$	10
2.2	A phase portrait, plot of $x(n)$ versus $y(n)$, of (2.3).	11
3.1	Signal model for the detection of signals in chaos. Φ is a chaotic dynamical system for generating a chaotic process, s is the signal to be detected, and f is the observation function (receiver)	17
3.2	A schematic representation of signal extraction from chaotic background. The dashed curve is an assumed chaotic trajectory and the solid curve is the trajectory of a signal added to the chaotic system.	18
3.3	Detection of a transient signal in a chaotic background. a) The transient signal; b) the x -component of the Ikeda chaotic system in the form of time series; c) the transient signal added to the chaotic sequence of b); d) the output evaluation, which clearly shows the presence of the transient signal.	21
3.4	Detection of a sinusoid in a chaotic background. a) The amplitude spectrum of the sinusoid; b) the amplitude spectrum of the x -component of the Ikeda chaotic system; c) the amplitude spectrum of the sinusoid added to the chaotic sequence of b); d) the amplitude spectrum of the resulting output. This figure shows not only the presence of the sinusoid, but also the frequency of the detected signal.	22

4.1	Illustration of the procedure used to calculate Lyapunov exponents from dynamical equations.	41
4.2	A schematic representation for estimating the largest Lyapunov exponent from a time series.	44
5.1	The IPIX radar trailer and antenna used for collection of the database reported in this chapter.	48
6.1	Mutual information versus time delay for sea clutter data. The minimum mutual information dimension occurs at $\tau = 5$	55
6.2	Mutual information dimension versus time delay for sea clutter data. The minimum mutual information dimension occurs at $\tau = 5$	55
6.3	Correlation dimension analysis for sea clutter data set $B318_{-20}$, where N is the length of the data set used for the computation. The solid curves correspond to the raw sea clutter data, and the dashed curves correspond to the data after being smoothed.	56
6.4	Correlation dimension analysis for sea clutter data set $B98_{-35}$, where N is the length of the data set used for the computation. The solid curves correspond to the raw sea clutter data, and the dashed curves correspond to the data after being smoothed.	57
6.5	Correlation dimension analysis for sea clutter data set $B318_{-20}$, where N is the length of the data set used for the computation. The dashed curves correspond to the sea clutter data, while the solid curves correspond to a filtered noise process (surrogate data) with exactly the same Fourier spectrum as that of the original clutter data.	59
6.6	Correlation dimension of a uniformly distributed random process used to generate the "surrogate data" used in Figure 6.6	59
6.7	Correlation dimension analysis for different received data. A data set of 30000 samples is used for each computation.	60

6.8	Comparison of different polarization on the computation of correlation dimension. Each data set used for the computations contains 30000 samples.	61
6.9	The effects of geographic locations on the computation of correlation dimension. Each curve corresponds to a sea clutter data set of 30000 samples, collected at a specific location. All three data sets were collected for roughly the same wave height.	62
6.10	The relationship between the estimated correlation dimension and sea state, with the sea state being classified roughly by wave height. (a) The results correspond to the database collected at Cape Bonavista, Newfoundland (1989). The solid curve shows the mean value of 38 estimates computed for each sea state, and each error bar shows the standard error (twice the standard deviation) in the estimation of the mean; (b) the results correspond to the database collected at Dartmouth, Nova Scotia (1993); once again, the solid curve shows the mean value of 14 estimates computed for each sea state, and each error bar has the same meaning as that in (a)..	63
6.11	False nearest neighbor analysis of a data set collected at the Cape Bonavista with wave height of 4.59 <i>ft.</i>	64
6.12	False nearest neighbor analysis of a data set collected at the Cape Bonavista with wave height of 6.70 <i>ft.</i>	65
6.13	False nearest neighbor analysis of a data set collected at the Cape Bonavista with wave height of 8.30 <i>ft.</i>	65
6.14	False nearest neighbor analysis of a data set collected at the Dartmouth with wave height of 4.90 <i>ft.</i>	66
6.15	False nearest neighbor analysis of a data set collected at the Dartmouth with wave height of 6.05 <i>ft.</i>	66
6.16	False nearest neighbor analysis of a data set collected at the Dartmouth with wave height of 7.45 <i>ft.</i>	67

6.17 Lyapunov exponents of the same sea clutter data used for the computations of Figure 6.7. The solid curve refers to I-channel data, the dashed curve refers to Q-channel data, and the dash-dotted curve refers to amplitude data. Each data set contains 30000 samples.	68
6.18 Lyapunov exponent of different polarization data used for the computation of Figure 6.8	69
6.19 Lyapunov exponents of the same sea clutter data used for the computations of Figure 6.9. Each curve refers to a sea clutter data set collected at a specific location. All three data sets were collected for roughly the same wave height.	69
6.20 The relationship between the estimated largest Lyapunov exponent and sea state, with the sea state being classified roughly by wave height. Traces are as defined in Figure 6.10	70
6.21 Lyapunov spectrum computed by the method based on 3rd-order polynomial fit to build the Jacobian	72
6.22 Kolmogorov entropy of sea clutter. The corresponding data is picked up from data set B218 of table 5.1. It has length of 35000 samples.	76
6.23 A multilayer perceptron structure used for reconstruction of the underlying dynamics of sea clutter, based on noncoherent radar data. It consists of an input layer with 50 source nodes, two hidden layers with 80 and 55 neurons, respectively, and one output neuron.	79
6.24 Recursive prediction result of the multilayer perceptron model shown in Figure 6.23. The solid curve refers to the original sea clutter waveform. The dashed curve refers to the recursive predicted waveform, for which the first 50 points of the sea clutter set (not shown in the figure) are used as the initial starting point.	81
6.25 Recursive prediction result of the multilayer perceptron model shown in Figure 6.23. Traces are as defined in Figure 6.24.	82

6.26	Recursive prediction result of the multilayer perceptron model shown in Figure 6.23. Traces are as defined in Figure 6.24.	82
6.27	Recursive prediction result of the multilayer perceptron model shown in Figure 6.23. Traces are as defined in Figure 6.24.	83
6.28	This figure shows the sensitivity of the recursive prediction process to a change in the neural network design. The network used for this experiment consists of an input layer with 45 source nodes, two hidden layers with 80 and 55 neurons, respectively, and one output neuron.	84
6.29	A comparison of both the AR model and the multilayer perceptron (MLP) model for reconstruction of the underlying dynamics of sea clutter.	85
6.30	A multilayer perceptron structure used for reconstruction of the underlying dynamics of sea clutter, based on coherent radar data. It consists of an input layer with 60 source nodes, (with 30 source nodes used for the in-phase component and 30 other source nodes used for the quadrature component), two hidden layers with 80 and 55 neurons, respectively, and two output neurons to provide one-step predictions of the in-phase and quadrature components.	87
6.31	Recursive prediction result of the multilayer perceptron model shown in Figure 6.30 for a testing data chosen from a sea state with wave height of 3.78 <i>ft</i> . The solid curves refer to the original sea clutter waveforms, in which one is the in-phase component and the other is the quadrature component. The dashed curves refer to the recursive predicted waveforms, for which the first 30 points of the complex-valued sea clutter data (not shown in the figure) are used as the initial starting point.	89
6.32	Recursive prediction result of the multilayer perceptron model shown in Figure 6.30 for a testing data chosen from a sea state with wave height of 4.54 <i>ft</i> . Traces are as defined in Figure 6.31.	90
6.33	Recursive prediction result of the multilayer perceptron model shown in Figure 6.30 for a testing data chosen from a sea state with wave height of 5.95 <i>ft</i> . Traces are as defined in Figure 6.31.	91

- 6.34 Recursive prediction result of the multilayer perceptron model shown in Figure 20 for a testing data chosen from a sea state with wave height of 8.30 *ft*. Traces are as defined in Figure 6.31. 92
- 6.35 (a) The phase portrait of a natural clutter data set; (b) the phase portrait of an artificial data set generated by the neural network model shown in Figure 17, for which the first 50 points of the clutter data (used in (a)) are used as the initial starting point. Each data set consists of 500 points. Each point of these phase portraits is defined by a sample value of the pertinent time series and its delayed version corresponding to a time delay $\tau = 5$ 94
- 6.36 (a) The power spectral density (PSD) of a natural clutter data set; (b) the power spectral density of the artificial data set generated by the neural network model shown in Figure 17, for which the first 50 points of the clutter data (used in (a)) are used as the initial starting point. Each data set consists of 102,400 points. The area between the two dashed curves in (a) and (b) shows 95% confidence for the estimation of the power spectrum. 95
- 6.37 (a) The probability density function of the data generated by the multilayer perceptron model of Figure 17 (star points) and that of the theoretical *K*-distribution (solid curve). (b) A comparison of the normalized amplitude moments of the data generated by the multilayer perceptron model of Figure 17 (dashed curve) with those of the theoretical *K*-distribution (solid curve). 96
- 6.38 (a) a phase portrait of a natural clutter data set; (b) a phase portrait of an artificial data set generated by the neural network model shown in Figure 6.30, for which the first 30 complex-valued points of the natural clutter data are used to initiate the model. 97
- 6.39 (a) The amplitude spectrum of a natural clutter data; (b) The amplitude spectrum of a clutter data generated by the neural network model shown in Figure 6.30. 98

7.1	A chaos-based noncoherent detector used for the detection of a small target in sea clutter. It consists of a neural network predictor and a “conventional” noncoherent CFAR processor.	104
7.2	Detection performance of the chaos-based noncoherent detector. The dashed curve is the ROC of the chaos-based noncoherent detector. The solid curve shows the ROC of a “conventional” Doppler CFAR detector for comparison. It can be seen that the chaos-based noncoherent detector has comparable performance to the Doppler CFAR detector.	105
7.3	A chaos-based coherent detector used for the detection of a small target in sea clutter. It consists of a neural network predictor and a “conventional” Doppler CFAR processor.	108
7.4	(a) Amplitude spectrum of sea clutter data produced by a coherent radar. (b) Amplitude spectrum of the prediction error produced by the neural network predictor of Figure 7.3 for the same clutter data.	109
7.5	(a) Amplitude spectrum produced by a coherent radar data set containing both target signal and clutter. (b) Amplitude spectrum of the prediction error produced by the neural network predictor of Figure 7.3 for the same target plus clutter data.	110
7.6	Detection performance of the chaos-based coherent detector. The dashed curve is the ROC of the chaos-based coherent detector. The solid curve shows the ROC of a “conventional” Doppler CFAR detector for comparison. It can be seen that the chaos-based coherent detector is superior to the Doppler CFAR detector.	111
7.7	Detection performance on the receiver noise for the chaos-based noncoherent detector, where the solid line corresponds to a false alarm rate of 10^{-3} and the dashed line corresponds to a false alarm rate of 10^{-2}	114
7.8	Detection performance on the receiver noise for the chaos-based coherent detector, where the solid line corresponds to a false alarm rate of 10^{-3} and the dashed line corresponds to a false alarm rate of 10^{-2}	114

- 8.1 Block diagram of a modular network; the outputs of the expert networks are mediated by a gating network. 119
- 8.2 Recursive prediction result of the modular network shown in figure 8.1 for a testing data set chosen from a sea state with wave height of 3.78 *ft*. The solid curves refer to the original sea clutter waveforms, in which one is the in-phase component and the other is the quadrature component. The dashed curves refer to the recursive predicted waveforms, for which the first 30 points of the complex-valued sea clutter data (not shown in the figure) are used as the initial starting point. 125
- 8.3 Recursive prediction result of the modular network shown in figure 8.1 for a testing data set chosen from a sea state with wave height of 4.54 *ft*. Traces are as defined in Figure 8.2. 126
- 8.4 Recursive prediction result of the modular network shown in figure 8.1 for a testing data set chosen from a sea state with wave height of 5.95 *ft*. Traces are as defined in Figure 8.2. 127
- 8.5 Recursive prediction result of the modular network shown in figure 8.1 for a testing data set chosen from a sea state with wave height of 8.30 *ft*. Traces are as defined in Figure 8.2. 128

Chapter 1

Introduction

Traditionally, signal detection has been based on statistical decision theory approached from the viewpoint of hypothesis testing. Specifically, a choice must be made between two hypotheses through the use of some decision-making criterion. For example, the problem of detecting a signal process in additive noise is usually solved by using Bayes' hypothesis-testing procedure, which is based on a statistical characterization of the noise. Given that we have a signal that is corrupted by a form of additive "noise" known to be chaotic, we may raise a fundamental question: How can this prior information be exploited to enhance the detection of a signal of interest? In this dissertation we develop a new detection strategy rooted in chaos theory. The power and flexibility of the new method are demonstrated by applying it for the detection of a small target in sea clutter.

The central claims of this dissertation are:

- Prior information that the additive "noise" is known to be chaotic may be exploited to enhance the detection of a signal of interest.
- It is demonstrated experimentally that sea clutter is indeed the result of a chaotic dynamical system. The characteristics of sea clutter as a chaotic process are shown to be essentially independent of environmental conditions.
- An universal neural network-based perceptron system able to capture the underlying dynamics of sea clutter globally is shown to be feasible.
- A chaos-based noncoherent detector operating on amplitude information only is shown

to perform at least as well as a conventional coherent detector.

- A chaos-based coherent detector is shown to be superior to a conventional coherent detector.

1.1 Problem Description

Signals (encompassing desired signals as well as interfering signals) have been partitioned traditionally into two broadly defined classes:

- Deterministic signals, by which we mean the class of signals whose waveforms can be defined precisely for all instants of time.
- Stochastic processes, by which we mean ensembles of random waveforms with each ensemble being defined by an underlying probability distribution.

This classification, however, overlooks another important class of signals, known as chaotic signals, that shares attributes with both deterministic signals and stochastic processes. Specifically, a chaotic signal generally has a very irregular waveform, but it is generated by a deterministic mechanism [75]. A comparison of various properties of a chaotic process with that of a conventional deterministic process and that of a random process is given in Table 1.1, in which the terms of both attractor dimension and Lyapunov exponent are described in great detail in Chapter 3.

Chaos is the very complex behavior of a dynamic system that is both nonlinear and deterministic. It represents a powerful notion, permitting the use of a simple deterministic system to explain highly irregular fluctuations exhibited by many physical phenomena encountered in nature. The first description of a chaotic process was made in 1963 by Lorenz [60], who developed a system of coupled nonlinear differential equations known as the Lorenz attractor to model weather patterns. Since then, the study of chaos has attracted the attention of mathematicians, physicists, and scientists from many other disciplines.

Some novel engineering applications of chaos have recently been reported in the literature, which may be grouped under two broadly defined categories [38, 71]:

Type of Process	Attractor Dimension	Lyapunov exponents	Predictability
Deterministic	finite	all ≤ 0	long-term predictable
Chaotic	fractal	some > 0	locally predictable
Random	infinite	some > 0	unpredictable

Table 1.1: Comparison of properties of various processes

- *Synthesis of chaotic signals*, exploiting the noise-like waveform of a chaotic signal; this class of applications includes *signal masking* and *spread-spectrum communications*.
- *Analysis of chaotic signals*, exploiting the fact that some physical phenomena permit the use of a chaotic model.

The subject matter of this thesis falls under the second category. In particular, we describe a new strategy for the detection of signals in chaos, and illustrate its application to a radar clutter-dominated ocean environment. The idea that sea clutter (*i.e.*, radar backscatter from an ocean surface) may be modeled as a chaotic process, was first reported by Leung and Haykin [56].

Given that we have a signal that is corrupted by a form of additive “noise” known to be chaotic, we may raise the following fundamental questions:

1. How can this prior information be exploited to enhance the detection of a signal of interest?
2. What if, for practical reasons, we are limited to a time series corresponding to a single component (*e.g.*, amplitude of a received radar signal)?

The answer to question 2 lies in Takens’ embedding theorem [76, 93], which states that it is possible to reconstruct the underlying dynamics of a chaotic system from a time series of a single component.

The answer to question 1 rests on the premise that we have several realizations of two different time series: time series I corresponding to the chaotic component acting alone, and time series II corresponding to the signal plus the chaotic component. We may then proceed in two stages. First, a particular realization of time series I is used to construct a nonlinear predictive model for a deterministic characterization of the chaotic background; for example, some form of a supervised neural network can be used to build the predictive model. Second, with the predictive model at hand, the statistical hypothesis testing procedure may be applied to the prediction error produced by the model. The performance of the whole system is then tested with realizations of time series I (not used before) and time series II. The ideas described here constitute the essence of the new chaos-based detection method.

To illustrate the power of the chaos-based detection method, we choose a difficult detection task. Specifically, we consider the radar detection of a small target (*e.g.*, a small piece of ice, called a growler, with a radar cross section of about 1 square meter or less) embedded in an ocean environment. In such an environment, the detection problem is dominated by the unavoidable presence of sea clutter that is the result of an underlying nonlinear dynamical system.

Sea clutter, referring to unwanted radar echoes from an ocean surface, has a long history of being modeled as a stochastic process. Some successful models, such as log-normal, Weibull and K-distribution, have been widely used by radar researchers and designers. In this context we may raise the following fundamental question: What is the physical or mathematical justification for the use of a stochastic model to describe sea clutter? Among the stochastic models mentioned, the semi-empirical K-distribution [49] is the only one that addresses the issue from a physical point of view. Even the K-distribution model, however, does not fully address the question that we have posed.

Notwithstanding this long history, it is our belief that sea clutter is basically chaotic in nature. In particular, we present strong experimental evidence, based on surface-truthed real-life radar data recorded with an instrument-quality radar, to support this assertion.

A chaotic description of sea clutter has several important implications. First, it means that we can build a predictive model for sea clutter with a much smaller number

of degrees of freedom than would be possible with a stochastic approach. Second, we may exploit the chaotic nature of sea clutter by invoking Takens' embedding theorem [76, 93]. Specifically, as explained previously, we only require a single component (*e.g.*, amplitude) to reconstruct the underlying nonlinear dynamics responsible for the generation of sea clutter. This raises another provocative question in the context of an ocean environment: Is it possible to exploit prior information that sea clutter is chaotic so as to enhance the detection performance of a noncoherent marine radar (providing amplitude information only) significantly enough to become comparable to that of a "conventional" coherent marine radar that supplies both amplitude and phase information? We will present experimental results that answer this question in the affirmative. This is important from a practical point of view. Basically, it means that, given a noncoherent marine radar, we may be able to exchange the relatively inexpensive but highly sophisticated computer software involved in the design of a chaos-based detector for the expensive microwave hardware involved in building a coherent marine radar for a conventional Doppler processor. Third, the use of the chaotic description of sea clutter may improve the detection performance of a "conventional" coherent radar, since a coherent radar can provide more information about the target of interest, and thereby further enhance the target indication for chaos-based detection.

1.2 Dissertation Overview

Chapter 2 gives an introductory review of nonlinear dynamical systems and chaos. A definition of a chaotic system is presented and discussed.

Chapter 3 presents a chaos-based hypothesis testing procedure, which is based on the existence of at least one positive Lyapunov exponent in a chaotic dynamical system. The chaos-based hypothesis testing method is demonstrated by computer experiments.

In Chapter 4, we discuss how to characterize chaos from a time series. We start with a discussion of the reconstruction of dynamics from a time series, which is built around Takens' embedding theorem. Then, we consider the notion of *attractor dimension*, and describe methods for the estimation of correlation dimension and minimum embedding dimension from a time series. In this chapter, we also describe another important notion.

namely, the *Lyapunov exponent*, which quantifies the sensitivity of a chaotic process to initial conditions. Algorithms for estimating Lyapunov spectrum and the largest Lyapunov exponent are described. The need for an adequate data length is stressed.

In Chapter 5, the experimental database used in this research is described in detail.

Chapter 6 presents chaotic characterization of sea clutter. We present experimental results on the correlation dimension analysis, the false nearest neighbors analysis, and Lyapunov exponent of sea clutter using real-life radar data, discuss the estimation of the underlying dynamics of sea clutter and demonstrate its plausibility using a neural network.

Detection philosophies and design strategies are described in Chapter 7. Two different chaos-based detectors are described, one of which is based on the noncoherent radar data and the other is based on the coherent radar data. A conventional Doppler CFAR detector is also discussed. Performance evaluations of these detectors are given. As well, comparisons with the conventional Doppler CFAR detector are presented and discussed. Finally practical considerations for each of the detectors are presented.

Chapter 8 studies the search of a global neural network model for the chaotic modeling of sea clutter. A modular network structure is proposed, and an algorithm for training the network is presented. Recursive predictions for testing data are performed.

Finally, Chapter 9 concludes by summarizing the results and contributions of this research, indicating future directions that may be taken.

Chapter 2

Nonlinear Dynamical Systems and Chaos

2.1 Nonlinear Dynamical Systems

A dynamical system is characterized by a set of related variables whose values change with time in a predictable manner, in principle, provided that the external influences acting on the system are all known exactly (*i.e.*, the system is deterministic). A model most naturally suited for the purpose of describing the dynamics of a nonlinear system is the so-called *state-space model*. According to this model, we think in terms of a set of state variables whose values (at any particular instant of time) are supposed to contain sufficient information to predict the future evolution of the system. Let $x_1(t), x_2(t), \dots, x_N(t)$ denote the state variables of the system, where continuous time t is the independent variable and N is the order of the system. For convenience of notation, these state variables are collected into an N -by-1 vector $\mathbf{x}(t)$, called the *state vector* of the system. The dynamics of a large class of nonlinear dynamical systems, assumed to be autonomous, may then be cast in the form of a system of first-order differential equations written in terms of the state vector $\mathbf{x}(t)$ as follows

$$\frac{d}{dt}\mathbf{x}(t) = F(\mathbf{x}(t)) \quad (2.1)$$

where the vector function F is, in general, nonlinear. We may think of $d\mathbf{x}/dt$ as a “velocity” vector, not in a physical but rather abstract sense. Then, according to (2.1), we may refer to the vector function $F(\mathbf{x})$ as a velocity vector field or, simply, a *vector field*.

For the state-space equation (2.1) to have a solution and for the solution to be unique, we have to impose certain restrictions on the vector function $F(\mathbf{x})$. For a solution to exist, it is sufficient that $F(\mathbf{x})$ be continuous in all of its arguments. However, this restriction by itself does not guarantee uniqueness of the solution. To do so, we have to impose a further restriction known as the *Lipschitz condition*. Let $\|\mathbf{x}\|$ denote the *norm* or *Euclidean length* of the vector \mathbf{x} . Let \mathbf{x} and \mathbf{u} be a pair of vectors in an open set \mathcal{M} in a normal vector (state) space. Then, according to the *Lipschitz condition*, there exists a constant K such that

$$\|F(\mathbf{x}) - F(\mathbf{u})\| \leq K \|\mathbf{x} - \mathbf{u}\| \quad (2.2)$$

for all \mathbf{x} and \mathbf{u} in \mathcal{M} . A vector function $F(\mathbf{x})$ that satisfies (2.2) is said to be *Lipschitz*, and K is called the *Lipschitz constant* for $F(\mathbf{x})$.

It is very informative to view the state-space equation (2.1) as describing the *motion* of a point in an N -dimensional state space, commonly referred to as the *phase space* of the system. The phase space can be a *Euclidean space* or a subset thereof. But then, it can also be a non-Euclidean space such as a circle, a sphere, a torus, or some other *differentiable manifold*. The phase space is important, because it provides us with a visual/conceptual tool for analyzing the dynamics of a nonlinear system described by (2.1). It does so by focusing our attention on *global characteristics* of the motion rather than detailed aspects of analytic or numeric solutions of the equation.

In this thesis, we are concerned with dissipative dynamical systems. Such a system is characterized by the convergence of its trajectories in phase space onto *manifolds* of lower dimensionality. By a “manifold” we simply mean the generalization of a surface or a curve in \mathbb{R}^N , it is usually referred as the attractor of the phase space. The manifold may consist of a single point in phase space, in which case we speak of a *point attractor*. Alternatively, it may be in the form of a periodic orbit, in which case we speak of a *limit cycle*.

2.2 Strange Attractors and Chaos

Nonlinear dynamical systems have the capability of exhibiting a chaotic behavior that is highly complex. This form of dynamic behavior arises by virtue of a type of attractors

called strange attractors.

In a nonlinear dynamical system, when the orbits (or trajectories) of an attractor with neighboring initial conditions tend to move apart with increasing time, the system is said to possess a *strange attractor* and the system itself is said to be *chaotic*. In other words, a fundamental property that makes an attractor “strange” is its *sensitive dependence on initial conditions*. Sensitivity in this context means that if two identical nonlinear dynamical systems are started with slightly different initial conditions, namely \mathbf{x} and $\mathbf{x} + \epsilon$ where ϵ is a very small quantity, then their states will diverge from each other in phase space and their separation will increase exponentially on the average.

As an example, consider a nonlinear dynamical system described by the following pair of coupled difference equations:

$$\begin{aligned}x(n+1) &= \frac{1}{2} \sin(2.3\pi x(n)(2x(n) + 0.56)) + \frac{1}{2} \cos(2\pi x(n)) \\y(n+1) &= 0.89x(n)\end{aligned}\tag{2.3}$$

An evaluation of this equation started at $[x(0), y(0)] = [0.9, 0.8]$ is shown by the solid lines of Figure 2.1. Consider next a new initial condition that deviates from $[x(0), y(0)]$ by a very small amount $\epsilon = 10^{-12}$ at $x(0)$ only. Then, restarting the evaluation of (2.3) at $[x(0) + \epsilon, y(0)]$, we get two distinct and very different waveforms shown by the dashed lines in Figure 2.1. A phase portrait of this system, a plot of $x(n)$ versus $y(n)$, is shown in Figure 2.2. The attractor described herein is strange, which is exemplified by sensitive dependence on its initial conditions.

For a formal definition of a chaotic dynamical system, we may refer to Newhouse [68, 69, 70], Farmer *et al.* [23], Holmes [44] and Guckenheimer *et al.* [36]. In this section we present a refinement of Newhouse’s definition that better addresses our specific problem. Without loss of generality, we confine our attention to discrete dynamical systems.

Consider an m -dimensional smooth manifold \mathcal{M} and a twice differentiable mapping $\phi: \mathcal{M} \rightarrow \mathcal{M}$. It is assumed that ϕ is one-to-one and the inverse map $\phi^{-1}: \phi(\mathcal{M}) \rightarrow \mathcal{M}$ is also twice differentiable. Let $\mathbf{x} \in \mathcal{M}$. Consider the positive orbit of \mathbf{x} , denoted by

$$O_+ = \{\mathbf{x}, \phi(\mathbf{x}), \phi^2(\mathbf{x}), \dots\}\tag{2.4}$$

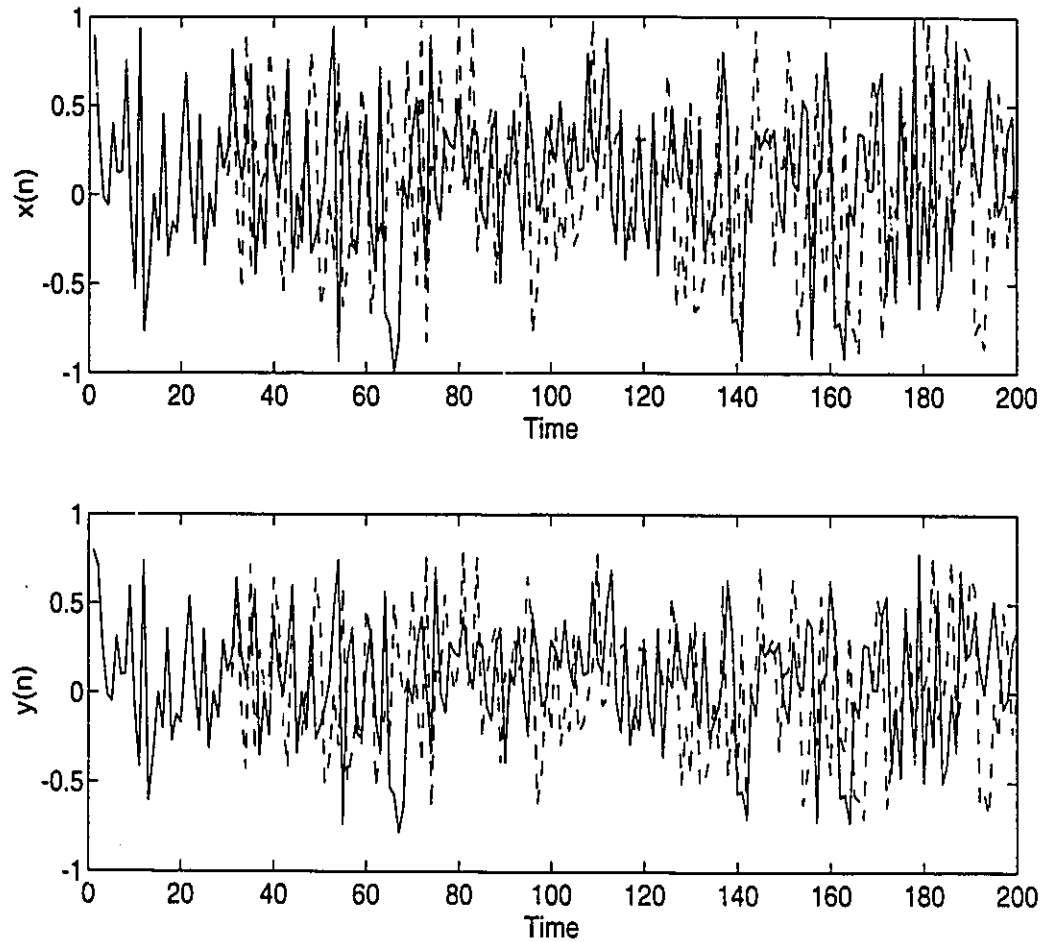


Figure 2.1: Time waveforms of Equation (2.3), in which the solid curves in both parts of the figure correspond to the initial state of $[0.9, 0.8]$ and the dashed curves correspond to the initial state of $[0.9 + 10^{-12}, 0.8]$.

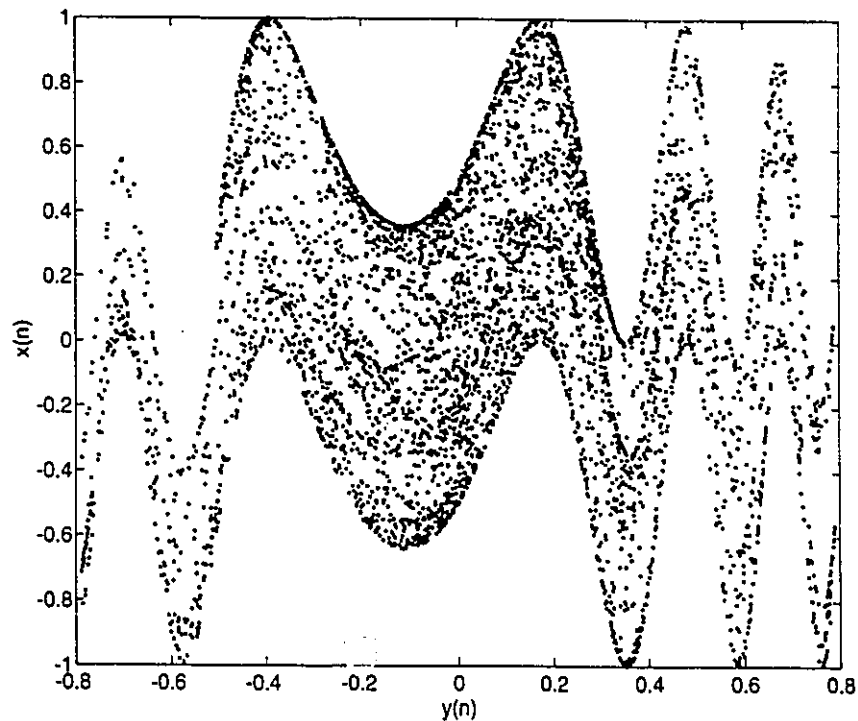


Figure 2.2: A phase portrait, plot of $x(n)$ versus $y(n)$, of (2.3).

and the attractor set called ω -limit set:

$$\begin{aligned} \omega(\mathbf{x}) &= \omega(\mathbf{x}, \phi) \\ &= \{ \mathbf{y} \in \mathcal{M} : \text{there is a sequence } n_1 < n_2 < \dots \text{ such that } \phi^{n_i}(\mathbf{x}) \rightarrow \mathbf{y} \text{ as } i \rightarrow \infty \} \end{aligned} \quad (2.5)$$

Note that if $O_+(\mathbf{x})$ is bounded, then $\omega(\mathbf{x})$ is a closed, bounded set, and $\phi(\omega(\mathbf{x})) = \omega(\mathbf{x})$. In general, we are interested in describing $\omega(\mathbf{x}, \phi)$ for as many \mathbf{x} and ϕ as possible. A simple situation arises when each $\omega(\mathbf{x}, \phi)$ is a periodic orbit for $\mathbf{x} \in \mathcal{M}$. Then, each initial state tends toward a periodic orbit. Such a ϕ is not chaotic. A particular situation in which ϕ may be called chaotic is when there are uncountably many points \mathbf{x} for which $\omega(\mathbf{x}, \phi)$ is also uncountable [68, 23, 44].

As mentioned previously, a fundamental property of chaotic systems is “sensitive dependence on initial conditions.” This could be formally defined as follows [68, 44, 36]. Let $d(\mathbf{x}, \mathbf{y})$ denote the distance between \mathbf{x} and \mathbf{y} . A point \mathbf{x} exhibits sensitive dependence on initial conditions if there are an $\alpha > 0$ and a constant $C > 0$ such that for any $\varepsilon > 0$

and any positive integer $n > 0$, there is a point \mathbf{y} such that the following two conditions are satisfied:

$$\begin{aligned} d(\mathbf{x}, \mathbf{y}) &< \varepsilon \\ d(\phi^j(\mathbf{x}), \phi^j(\mathbf{y})) &\geq C e^{\alpha j} d(\mathbf{x}, \mathbf{y}) \quad 1 \leq j \leq n \end{aligned}$$

The infinitesimal version of this statement can be more easily defined [69, 12]. Specifically, for sensitive dependence on initial conditions, there has to be a vector \mathbf{v} tangential to \mathcal{M} at \mathbf{x} such that

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log \|T_{\mathbf{x}} \phi^n \mathbf{v}\| > 0 \quad (2.6)$$

where $T_{\mathbf{x}} \phi^n$ is the derivative of ϕ^n with respect to \mathbf{x} , and $\|\cdot\|$ is the norm of a vector. The condition on \mathbf{x} and \mathbf{v} described in (2.6) is frequently referred to as “ (\mathbf{x}, \mathbf{v}) has a positive Lyapunov exponent”. The notion of a Lyapunov exponent will be discussed in great detail in Chapter 4.

Newhouse [70] has summarized the above discussion, giving the following definition of a chaotic system:

A nonlinear dynamical system described by $\phi : \mathcal{M} \rightarrow \mathcal{M}$ is said to be chaotic if there is a closed bounded subset \mathcal{V} of \mathcal{M} with non-empty interior such that

1. ϕ maps \mathcal{V} into its interior.
2. There are uncountably many points \mathbf{x} in \mathcal{V} for which $\omega(\mathbf{x})$ is uncountable and \mathbf{x} has a positive Lyapunov exponent.

In most practical situations, we only have a time series describing some observations of a dynamical system. To this end, we offer the following definition that builds on Newhouse’s definition:

A bounded deterministic dynamical system with at least one positive Lyapunov exponent is a chaotic system. A chaotic signal is an observation of the system.

By “deterministic” we mean that the system can be described completely by a set of rules;

these rules may be expressed as nonlinear differential or difference equations. More specifically, the system is at least locally (i.e., short-time) predictable and has an attractor whose dimension is finite and generally *fractal* (i.e., noninteger).

This definition has practical significance. It actually paves the way for testing the chaotic behavior of a given nonlinear dynamical system. First of all, we check the boundedness of the given system. In most practical problems, the boundedness is obvious. To test its deterministic property, we have to do two things: estimate the dimension of the system, and check its predictability. Generally speaking, a deterministic system has a finite dimension. The predictability is the most important property of a deterministic system. It is very difficult to confirm the predictability in most practical problems since we need to construct a model, and then confirm that the model is indeed the reconstruction of the underlying dynamics of the system. Fortunately, the use of neural networks provides a practical procedure to perform the modeling task [55, 45, 80, 39]. The predictability is then confirmed by *recursive prediction*, a topic that we will discuss in detail in Chapter 4. After both the bounded and deterministic properties of the system have been confirmed, the existence of at least one positive Lyapunov exponent would have to be established. This criterion plays a key role in the characterization of a process as chaotic, because the Lyapunov exponent of a map provides a measure of sensitive dependence of the map on initial conditions. It actually gives the average rate of divergence or convergence of the dynamical system along the principal axes of its phase space. Details of how to extract Lyapunov exponents from experimental data will be given in Chapter 4. Thus, a decision on whether the nonlinear dynamical system is chaotic or not can be made by invoking the modification of Newhouse's definition described above.

2.3 How Can Chaos Arise?

There are two basic ways in which a chaotic process can arise:

1. The process is purposely designed to be chaotic
2. The underlying physical phenomenon responsible for the generation of the process is naturally chaotic.

Examples of the first category include the logistic map, the Henon map, the Lorenz attractor, the Rossler attractor, and also the map described in (2.3). Examples of the second category include heart rhythm [31], turbulence in fluids [1, 2], and radar backscatter from an ocean surface [56], to mention just a few.

In determining the suitability of a model for the mathematical description of a physical phenomenon, there are four important criteria that we should keep in mind [38]:

1. The model should be rooted in the underlying physics of the process.
2. It should be possible to estimate the parameters of the model from measured data (observations).
3. The model should replicate in some sense the process which produced the data by being able to predict the future behavior of the process.
4. The model should permit the use of a minimum number of degrees of freedom (i.e., independent variables) for its physical implementation.

Unfortunately, in the literature, not enough attention has been paid to the selection of a model that builds on all of these four criteria together. Much too often in the past, a stochastic process has been selected as the model for a physical phenomenon. This is done not on physical grounds, but rather because of mathematical convenience, or in an *ad hoc* manner. This, in turn, has two serious consequences. First, physical attributes of the phenomenon are overlooked. Second, the use of a stochastic process as the mathematical model introduces unnecessary degrees of freedom. Herein lies the advantage of a nonlinear dynamical system over a stochastic process as the mathematical model in that it utilizes a minimum number of degrees of freedom. The important point that we wish to make here is that for some specific physical phenomena the use of a chaotic model not only provides the natural mathematical model, but also does so with the minimum number of degrees of freedom.

Chapter 3

Chaos-based Hypothesis Testing

3.1 Introduction

Traditionally, signal detection has been based on a statistical decision theory approach utilizing hypothesis testing. Specifically, a choice must be made between two hypotheses through the use of some decision-making criterion. The problem of detecting a signal process in additive noise may be simply formulated as follows:

$$\begin{aligned} \text{Hypothesis } H_1 : \quad & y(t) = s(t) + w(t), \quad 0 \leq t \leq T \\ \text{Hypothesis } H_0 : \quad & y(t) = w(t), \quad 0 \leq t \leq T \end{aligned} \quad (3.1)$$

where $y(t)$ is the observed signal, $s(t)$ is the target signal of interest, and $w(t)$ is the additive noise component. This decision-making problem is usually solved by using Bayes' hypothesis-testing procedure [97], which is based on a statistical characterization of the noise $w(t)$.

In this chapter, we present a new detection strategy rooted in chaos theory. The strategy is based on the existence of at least one positive Lyapunov exponent in a chaotic dynamical system responsible for the generation of a "noise" process $w(t)$. The basic idea is as follows: Since a chaotic system is a deterministic nonlinear dynamical system, we may invoke Takens' embedding theorem to build a model that represents a reconstruction of the system responsible for generating $w(t)$. When the observed signal $y(t)$ contains a target component $s(t)$ as in the first line of (3.1), $y(t)$ does not match the model any more; hence, the existence of the target component $s(t)$ is signified by the presence of a corresponding

perturbation at the output of the model. This new detection strategy has the potential of outperforming the traditional detection strategy by exploiting the “prior” information that the noise process $w(t)$ is in fact chaotic.

3.2 Detection Hypothesis

Let \mathcal{M} be an m -dimensional compact manifold, and ϕ be a chaotic dynamical system on \mathcal{M} , which is a map $\phi: \mathcal{M} \rightarrow \mathcal{M}$. An observation on the dynamical system is a smooth function $f: \mathcal{M} \rightarrow \mathfrak{R}$. Consider a trajectory $c(t)$ of ϕ , which is started from an initial state $c(t_0)$. Then an observation may be expressed as:

$$y(t) = f(c(t)) \quad (3.2)$$

In a signal detection problem, f acts as the receiver. It is reasonable to assume that f is the identity map, in which case we may set $y(t) = c(t)$. For convenience, we will use $c(t)$ as the observation process of the dynamical system ϕ .

Consider now another signal being added to the system, as shown in Figure 3.1; it may be a periodic deterministic signal, a transient signal, or another chaotic signal. We may then identify two kinds of chaotic detection procedures:

1. Deterministic signal in a chaotic background:

$$\begin{aligned} \text{Hypothesis } H_1 : & \quad y(t) = c(t) + s(t), \quad 0 \leq t \leq T \\ \text{Hypothesis } H_0 : & \quad y(t) = c(t), \quad 0 \leq t \leq T \end{aligned} \quad (3.3)$$

2. Classification of two chaotic dynamical systems:

$$\begin{aligned} \text{Hypothesis } H_1 : & \quad y(t) = c_1(t), \quad 0 \leq t \leq T \\ \text{Hypothesis } H_0 : & \quad y(t) = c_0(t), \quad 0 \leq t \leq T \end{aligned} \quad (3.4)$$

where $c_0(t)$ and $c_1(t)$ are observations of the two chaotic systems. Most particularly, the two chaotic systems may be of the same structure, but there are some changes in their free parameters.

Some work has been reported on the classification model [35, 102, 8, 64]. In this chapter, we focus on model 1.

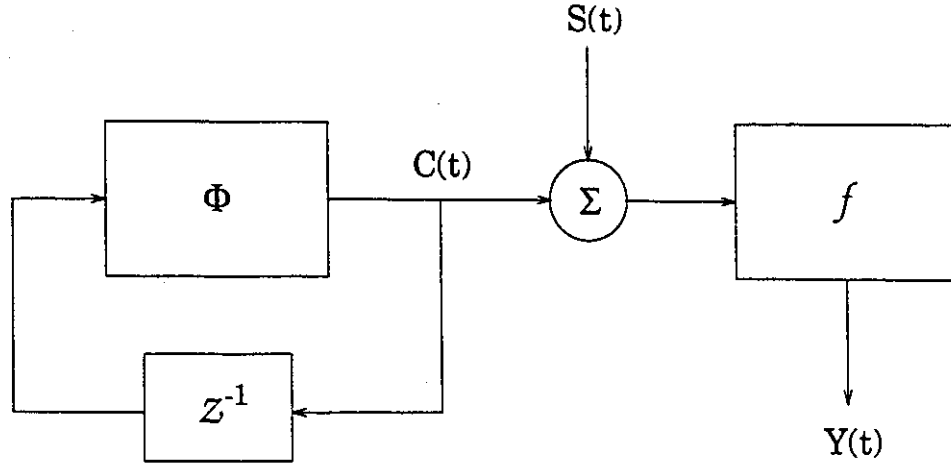


Figure 3.1: Signal model for the detection of signals in chaos. Φ is a chaotic dynamical system for generating a chaotic process, s is the signal to be detected, and f is the observation function (receiver)

3.3 Detection Principle

Let ϕ be a chaotic dynamical system on \mathcal{M} . Consider an infinitesimally small m -dimensional sphere, an initial state of ϕ , in \mathcal{M} . During its evolution, the sphere will become distorted, but being infinitesimal, it will remain ellipsoidal. Denote the principal axes of this ellipsoid by $c_i(t) : i = 1, \dots, m$. We may then approximately write

$$c_i(t) \approx c_i(0)e^{\lambda_i t} \quad (3.5)$$

where λ_i is the i^{th} Lyapunov exponent of the system, and $c_i(0)$ is an initial state corresponding to $c_i(t)$ of the i^{th} principal axis of the ellipsoid.

Suppose that a signal $s_i(t)$ is added to the process $c_i(t)$, modifying the observations as follows:

$$y_i(t) = c_i(t) + s_i(t) \quad (3.6)$$

$$y_i(0) = c_i(0) + s_i(0) \quad (3.7)$$

Let there be an identical chaotic system ϕ at the receiver; then the output of the system is

$$\hat{y}_i(t) \approx y_i(0)e^{\lambda_i t} \quad (3.8)$$

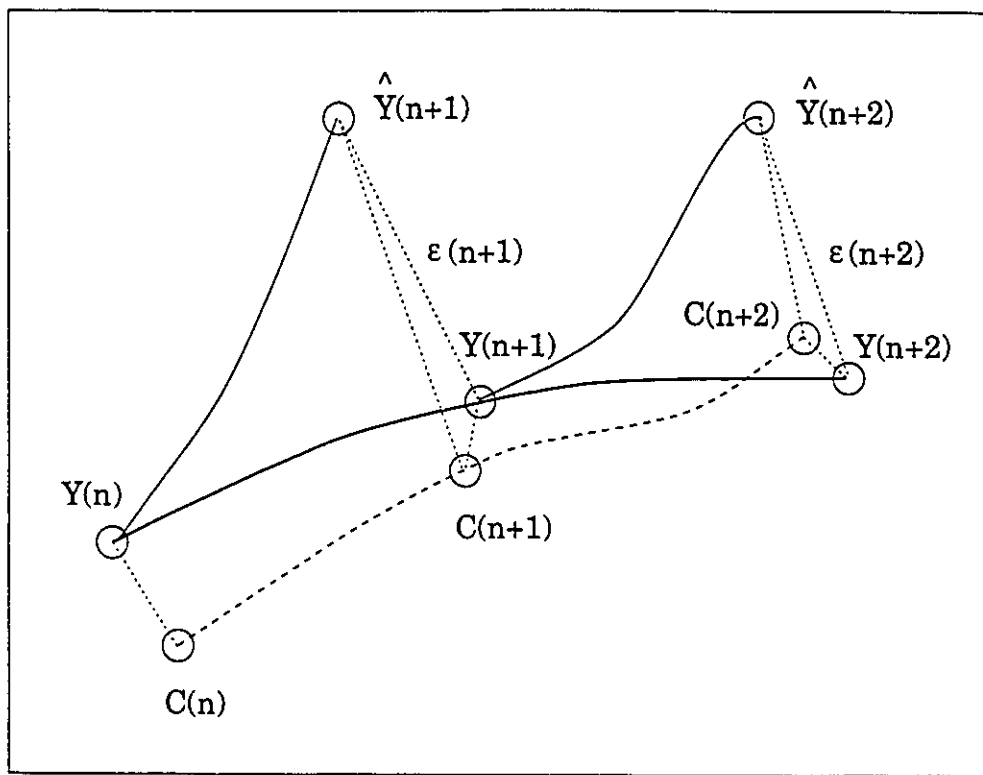


Figure 3.2: A schematic representation of signal extraction from chaotic background. The dashed curve is an assumed chaotic trajectory and the solid curve is the trajectory of a signal added to the chaotic system.

Assuming that the signal is known to change slowly, we may set $s_i(t) \approx s_i(0)$ for a small range of time t , in which case we may define the difference between $\hat{y}_i(t)$ and $y_i(t)$ as follows:

$$\varepsilon_i(t) = \hat{y}_i(t) - y_i(t) \quad (3.9)$$

Substituting (3.5) to (3.8) into (3.9), we get

$$\varepsilon_i(t) \approx s_i(0)(e^{\lambda_i t} - 1) \quad (3.10)$$

If $\lambda_i = 0$, which corresponds to a periodic deterministic system, we have $\varepsilon_i(t) = 0$, and the signal component cannot be extracted in the manner described here. If $\lambda_i < 0$ for all i , we have a point attractor, which is of no interest to us here. If, however, $\lambda_i > 0$ for some i , which corresponds to a chaotic system, then $e^{\lambda_i t} > 1$, and so (3.10) can be rewritten as

$$s_i(0) \approx \frac{\varepsilon_i(t)}{e^{\lambda_i t} - 1} \quad (3.11)$$

According to (3.11), the signal component may be extracted, given that the system is chaotic with a positive Lyapunov exponent.

The formula of (3.11) cannot be used to extract the signal component from a chaotic time series exactly, since a strange attractor results from a stretching and folding process; the stretching expands the volume occupied by the phase space of the dynamical system, while the folding contracts it. Nevertheless, (3.11) provides the basis of a method to detect the existence of a signal component, which is still a highly valuable attribute to have. The idea may be described schematically as in Figure 3.2, where the dashed line represents a trajectory of the underlying chaotic system. When a signal component is present, the observation process is shown by the solid line in the figure. It is represented by $y(t_0), y(t_1), y(t_2), \dots$. Suppose, next, there is an identical chaotic system at the receiver. Then, applying the received signal to the latter chaotic system, we get an evaluation process $\hat{y}(t_1), \hat{y}(t_2), \dots$. The reconstruction error process $\varepsilon(t_1), \varepsilon(t_2), \dots$, defined in accordance with (3.9), indicates the existence of the signal component. In what follows we present two computer experiments to illustrate how the chaos-based detection strategy works.

3.4 Computer Experiments

(a) Transient signal in chaos

Consider the well-known *Ikeda chaotic system* [74]. It is described by the following system of coupled difference equations:

$$\begin{aligned}x(n) &= 1 + \mu[x(n-1)\cos(\nu) - y(n-1)\sin(\nu)] \\y(n) &= \mu[x(n-1)\cos(\nu) + y(n-1)\sin(\nu)] \\ \nu &= 0.4 - \frac{6}{1 + x(n-1)^2 + y(n-1)^2}\end{aligned}\tag{3.12}$$

where $\mu = 0.9$ and $n = 1, 2, \dots, N$. Suppose that a transient signal $s(n)$, shown in Figure 3.3(a), is added to the x component of the system output. Figure 3.3(b) shows a plot of the x component of the chaotic system versus discrete time n . Suppose also that the received signal $r(n)$ is defined by the sum $x(n) + s(n)$, as shown in Figure 3.3(c). In this experiment we wish to detect $s(n)$ from the received signal $r(n)$, assuming a signal-to-noise ratio (SNR) of about -60 dB. Suppose next that we have a dynamical system identical to that of (3.12) at the receiver; its implementation is described by

$$\begin{aligned}\hat{x}(n) &= 1 + \mu[r(n-1)\cos(\nu) - \hat{y}(n-1)\sin(\nu)] \\ \hat{y}(n) &= \mu[r(n-1)\cos(\nu) + \hat{y}(n-1)\sin(\nu)] \\ \nu &= 0.4 - \frac{6}{1 + r(n-1)^2 + \hat{y}(n-1)^2}\end{aligned}\tag{3.13}$$

The processed output is defined by $e(n) = \hat{x}(n) - r(n)$. Figure 3.3(d) shows the output for the conditions described here. This figure clearly indicates the existence of the added transient signal, despite the fact that the signal-to-noise ratio used in the computer experiment is very low. The reason why the chaos-based detector is able to perform so well is because of the prior information that the additive "noise" is chaotic, which is exploited in designing the receiver.

(b) Periodic signal in chaos

Reconsider the same problem described in the computer experiment (a). The only difference now is that the signal to be detected is periodic. Without loss of generality, we may suppose that the periodic signal is a sinusoid, as shown by

$$s(n) = a \sin(\omega n \tau)\tag{3.14}$$

where τ is the sampling period, ω is the angular frequency, and a is the amplitude of the signal. The amplitude spectrum of the signal, defined as the magnitude of its Fourier

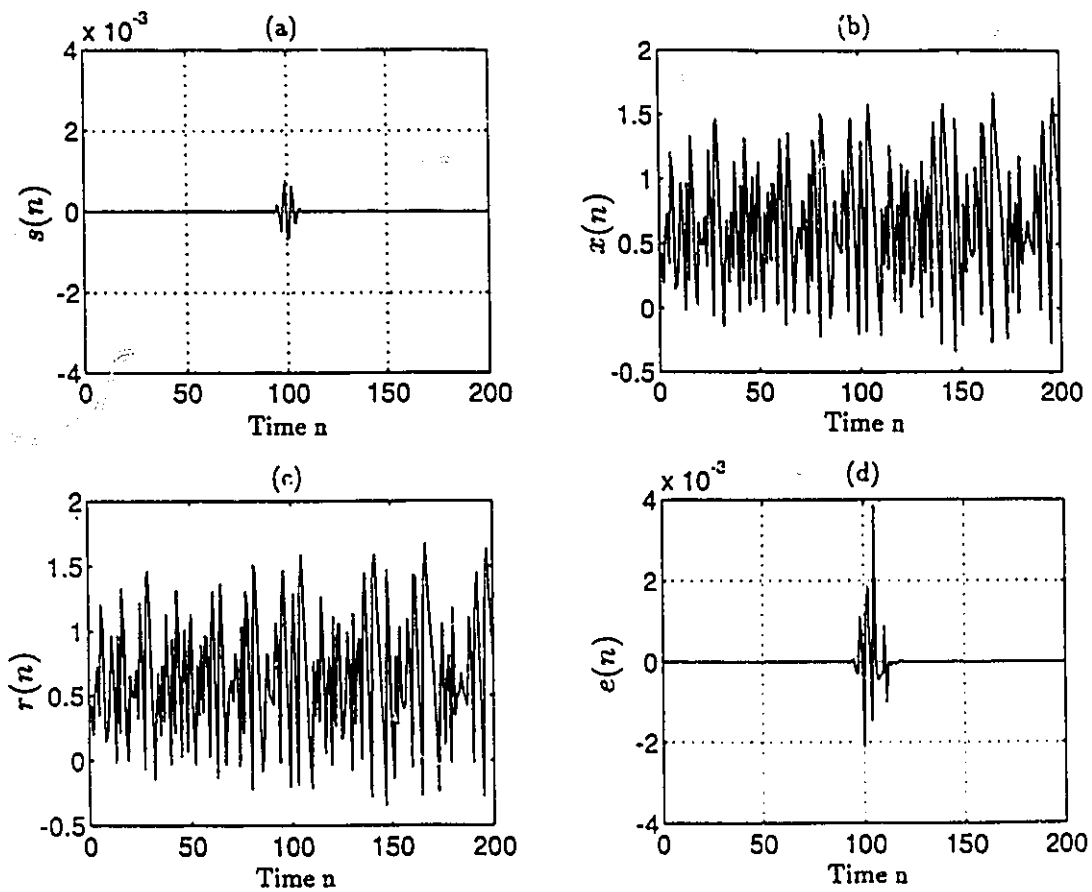


Figure 3.3: Detection of a transient signal in a chaotic background. a) The transient signal; b) the x -component of the Ikeda chaotic system in the form of time series; c) the transient signal added to the chaotic sequence of b); d) the output evaluation, which clearly shows the presence of the transient signal.

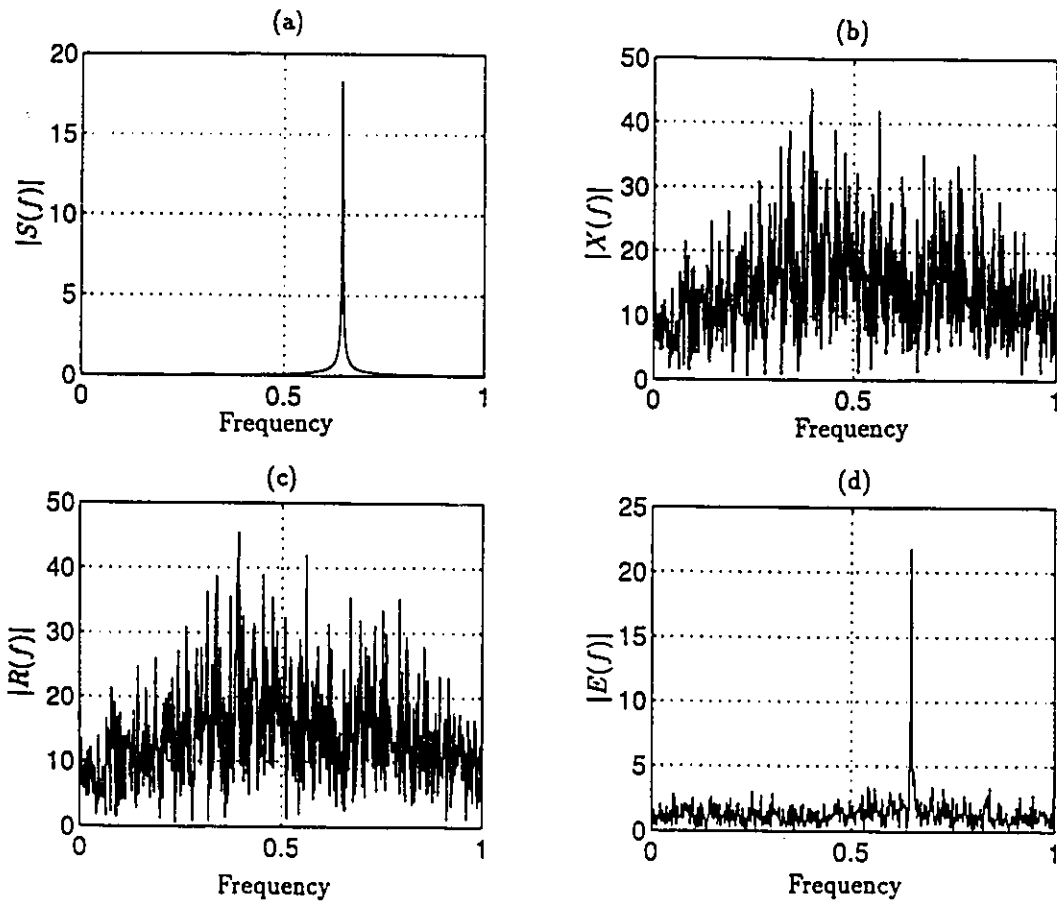


Figure 3.4: Detection of a sinusoid in a chaotic background. a) The amplitude spectrum of the sinusoid; b) the amplitude spectrum of the x -component of the Ikeda chaotic system; c) the amplitude spectrum of the sinusoid added to the chaotic sequence of b); d) the amplitude spectrum of the resulting output. This figure shows not only the presence of the sinusoid, but also the frequency of the detected signal.

transform, is shown in Figure 3.4(a). Figure 3.4(b) shows the amplitude spectrum of $x(n)$. The received signal is $r(n) = x(n) + s(n)$, assuming a signal-to-noise ratio of about -50 dB. Its amplitude spectrum is shown in Figure 3.4(c). As before, the receiver is implemented by using (3.13), and the receiver output is $e(n) = \hat{x}(n) - r(n)$. To detect the sinusoidal signal, we use the fast Fourier transform (FFT) method, since it gives an optimal solution for an unknown sinusoid in additive “white” noise. Figure 3.4(d) shows the amplitude spectrum of the receiver output $e(n)$. In this case, the chaos-based receiver indicates not only the presence of the added sinusoidal signal, but also the frequency of the detected signal.

3.5 Summary

In this chapter, we have presented a chaos-based hypothesis testing procedure. It is based on the existence of at least one positive Lyapunov exponent in a chaotic dynamical systems. Two computer experiments are presented to demonstrate the chaos-based hypothesis testing method. Based on the two experiments described above, we may make the following observations:

- A chaos-based receiver can detect extremely weak signals buried in a chaotic background by exploiting the prior knowledge that the additive “noise” is chaotic.
- The receiver preserves the inherent characteristics of the target signal that are essential for its detection.

Chapter 4

Characterization of a Chaotic Process

4.1 Introduction

We have discussed in Chapter 2 that a dynamical system is chaotic if 1) it is bounded, 2) it is deterministic, and 3) it has at least one positive Lyapunov exponent. In most practical problems, we may have only one or two observations, recorded in the form of time series, of an unknown system. The boundedness of the system is mostly obvious. The question then is how to demonstrate that the system responsible for the generation of the time series is deterministic? And how to extract the Lyapunov exponents from the recorded time series?

We start this chapter by discussing the Takens' embedding theorem, which provides a theoretical basis for methods of modeling the underlying dynamics of a chaotic system using a time series.

4.2 Reconstruction of Chaotic Dynamics

An important issue in the study of dynamical systems is that of *model reconstruction* or *reconstruction of dynamics*. Specifically, given a time series made up of a physical observation of a nonlinear dynamical system known to be chaotic, how do we reconstruct the underlying dynamics of the system? More specifically, what if the time series is from a

single component of the system (e.g., amplitude of a received radar signal)? At first glance, such a proposition may appear to be ambitious and perhaps overoptimistic. Fortunately, Takens [93] and Packard *et al.* [76] independently proposed the idea of using sequences of measurements to gain geometric information of the underlying dynamics. The proposition was put on a firm theoretical basis by Takens [93], and the result is commonly known as *Takens' embedding theorem*. This theorem provides the mathematical basis of the dynamic reconstruction problem. It states that model reconstruction of a nonlinear dynamical system using just *one* observation of the system should succeed to a certain extent, and the reconstruction is independent of which signal component is used. Naturally, the model reconstruction as described here cannot be expected to produce a model exactly the same as the original system. Nevertheless, the reconstruction will have the same behavior as the original system up to a diffeomorphism [93, 76]. *Diffeomorphic equivalence* establishes the equivalence between the original nonlinear dynamical system and the reconstructed system to within a smooth curvilinear change of coordinates. More formally, we may state the following [93]:

Theorem (Takens): Let \mathcal{M} be a compact manifold of dimension m . For pairs (ϕ, f) , $\phi : \mathcal{M} \rightarrow \mathcal{M}$ denoting a smooth diffeomorphism, and $f : \mathcal{M} \rightarrow \mathcal{R}$ denoting a smooth function, it is a generic property that the map $\Phi_{(\phi, f)} : \mathcal{M} \rightarrow \mathcal{R}^{2m+1}$ defined by:

$$\begin{aligned}\Phi_{(\phi, f)}(\mathbf{x}) &= (f(\mathbf{x}), f(\phi(\mathbf{x})), \dots, f(\phi^{2m}(\mathbf{x}))) \\ &\equiv (y(0), y(1), \dots, y(2m)),\end{aligned}\tag{4.1}$$

is an embedding; here 'smooth' means that at least twice differentiable and continuous, and $y(i) = f(\phi^i(\mathbf{x}))$.

The 'generic' property mentioned here implies the following:

- If \mathbf{x} is a point with period k of ϕ with $k \leq 2m + 1$, then all eigenvalues of $\frac{d\phi^k}{d\mathbf{x}}$ are different from each other and different from unity. In other words, the trajectory is not duplicated even in a small region around the point \mathbf{x} .
- Any two different fixed points of ϕ are not in the same level of y . That is, the map is guaranteed to be injective.

To see how this theorem can be of practical use, consider an m -dimensional compact manifold \mathcal{M} ; a chaotic dynamical system defined on \mathcal{M} is described by the difference equation:

$$\mathbf{x}(n+1) = \phi(\mathbf{x}(n)) \quad (4.2)$$

where ϕ is a nonlinear function. A trajectory of the dynamical system, started at an initial state $\mathbf{x}(0)$, is then a sequence of points $\{\mathbf{x}(i) = \phi^i(\mathbf{x}(0)), i = 1, 2, \dots\}$. Suppose we now have a smooth observable map, say f , which is described by a sequence of observations $\{y(i) = f(\mathbf{x}(i)), i = 0, 1, \dots\}$. Takens' embedding theorem says with $D_E \geq 2m+1$, the map $\Phi: \mathcal{M} \rightarrow \mathcal{R}^{D_E}$ defines a corresponding trajectory, in which each point defined by

$$\Phi(i) = \{y(i), y(i+1), \dots, y(i+D_E-1)\} \quad (4.3)$$

is a state of a related dynamical system $\psi: \mathcal{R}^{D_E} \rightarrow \mathcal{R}^{D_E}$, and which is equivalent to $\phi: \mathcal{M} \rightarrow \mathcal{M}$, up to a diffeomorphism, where D_E is the embedding dimension. Thus we have

$$\Phi(i+1) = \psi(\Phi(i)) \quad (4.4)$$

or, equivalently,

$$\{y(i+1), y(i+2), \dots, y(i+D_E)\} = \psi(y(i), y(i+1), \dots, y(i+D_E-1)) \quad (4.5)$$

Equation (4.5) states that each component of the left-hand side is determined by the components on the right-hand side. Hence, there exists an F such that

$$y(i+N) = F(y(i), y(i+1), \dots, y(i+D_E-1)) \quad (4.6)$$

where F is a nonlinear function that predicts the next entry in the time series. Takens' embedding theorem is an existence theorem in that it assures us that this function does exist, but does not tell us how to find it. In principle, then, we may approximate the reconstructed phase space by using a time series observed from the original system [93, 76].

Typically, a time series is a series of recorded values of continuous observations of a dynamical system at regular, discrete time intervals. Assume that we are given an observation $y(t)$, where $t \in [0, T]$. In most experiments, t is a continuous variable. We record this observation $y(t)$ at time intervals Δt , corresponding to a sampling rate of $1/\Delta t$; thus, we get a discrete sequence $y(i) = y(i\Delta t)$ of real numbers, which is a one-dimensional time

series. Before using Takens' embedding theorem to reconstruct the underlying dynamics, we first have to determine a characteristic time $\tau = k\Delta t$, called the *time delay*, after which the system loses an essential part of its information about its previous state. Without loss of generality, we may set $\Delta t = 1$; then (4.3) becomes

$$\Phi(i) = \{y(i), y(i + \tau), \dots, y(i + (D_E - 1)\tau)\} \quad (4.7)$$

The first zero-crossing of the autocorrelation is often used to determine the delay [55]. Average mutual information provides a more refined estimate of the delay for the reconstruction of underlying dynamics [25, 28].

The average mutual information for a time series is defined as:

$$I(k) = \sum_{i=1}^N P(y(i), y(i+k)) \log_2 \left[\frac{P(y(i), y(i+k))}{P(y(i))P(y(i+k))} \right] \quad (4.8)$$

where $P(y(i))$ is the probability of observing $y(i)$, $P(y(i+k))$ is the probability of observing $y(i+k)$, and $P(y(i), y(i+k))$ is the joint probability of observing both $y(i)$ and $y(i+k)$. N is the total number of observed data. The first local minimum of the average mutual information $I(k)$ plotted versus k has been shown to be a good prescription for selecting the optimum time delay τ for chaotic reconstruction [25, 28].

Pineda and Sommerer [79] proposed another method for estimating the time delay τ based on the definition of generalized dimensions of a dynamical system. To choose a delay τ for dynamical reconstruction, we may embed the data in one and two dimensions, and compute the *mutual information dimension*

$$D_\mu = 2D_{I1} - D_{I2} \quad (4.9)$$

where D_{I1} is the information dimension for embedding dimension $D_E = 1$, and D_{I2} is the information dimension for embedding dimension $D_E = 2$. Next, the mutual information dimension is minimized with respect to τ .

Since Takens' embedding theorem plays a key role in the estimation of attractor dimension, Lyapunov exponents, and the underlying dynamics of the given system, the determination of the time delay provides an important step in performing the above estimations. Typically, having chosen τ , we perform the dimension analysis for increasing

values of the embedding dimension D_E and look for a plateau in the plot of an estimation of the attractor dimension versus D_E .

There are two important observations about the reconstruction method that should be noted. First, Takens' embedding theorem states that the reconstruction of the underlying dynamics of a chaotic system can be implemented directly by using the observation data. The only problem that we may face here is the construction of the map F . In some situations, a direct construction is possible, because it is simple and some prior information is given. However, this is not the case in general. For example, in the sea clutter modeling problem considered later in this thesis, we have no existing knowledge on the modeling strategy; hence, no prior assumption should be imposed to bias the model. The derivation of F is another major and important part of the dynamical reconstruction. Fortunately, the use of a neural network makes the problem resolvable in an experimental sense [55, 45, 80]. A neural network is ideally suited to reconstruct the nonlinear dynamics of a chaotic process. We make this assertion for three reasons:

- A neural network is inherently a nonlinear device, and nonlinearity is basic to the existence of a chaotic process.
- A neural network lends itself to training through the use of a set of examples representative of the chaotic process.
- A neural network is a universal approximator in that, in theory, it can approximate any continuous input-output mapping to a desired precision.

The use of neural networks for model reconstruction will be discussed in detail in Section 6.5.

Takens' embedding theorem states that, in principle, model reconstruction using just one observation component should succeed to a certain extent, and the reconstruction is independent of which component is used [76]. This point is of great practical significance. It implies that we may choose any observation component of the process to perform the dynamic reconstruction. To be specific, consider a coherent radar system in which the received (observed) signal $x(t)$ consists of an in-phase component $x_I(t)$ and a quadrature component $x_Q(t)$. Typically, both $x_I(t)$ and $x_Q(t)$ are continuous and differentiable functions of time,

and do therefore lend themselves to the application of Takens' embedding theorem. According to this theorem, dynamic reconstruction of the process responsible for generating $x(t)$ may be possible up to a diffeomorphism by using $x_I(t)$ or $x_Q(t)$ alone. Consider next the case of a noncoherent radar system, which operates on an amplitude signal denoted by $x_A(t)$. Here too we find that $x_A(t)$ is typically continuous and differentiable with respect to time, and is therefore suitable for the application of Takens' embedding theorem. That is, the dynamic reconstruction of the process responsible for generating $x(t)$ may be possible up to a diffeomorphism by using the amplitude component $x_A(t)$. In the context of radar detection, for example, the latter statement implies that the performance of a noncoherent radar (providing amplitude information only) may be improved significantly enough via dynamic reconstruction to be comparable to that of a "conventional" coherent radar. This observation was first demonstrated experimentally by Li and Haykin [59].

Throughout the discussion, we have assumed that the system admits a finite dimensional model, and this dimension determines the number of variables involved in the process. Determining the attractor dimension of a nonlinear dynamical system is covered in the next section.

4.3 Attractor Dimension

Not only does the concept of attractor dimension help to select the correct embedding dimension D_E , but it also serves as a means of illuminating the complexity of an attractor structure. In this section, we first describe the well-known Hausdorff dimension; we then go on to discuss the so-called correlation dimension, which lends itself much more readily to computation for a time series of finite length.

4.3.1 Hausdorff Dimension

In 1919, Hausdorff [37] first introduced the notion of dimension analysis to nonlinear dynamical systems, and gave the definition of attractor dimension, which has become known as the *Hausdorff dimension*. This definition is based on the Hausdorff measure of a compact space. It can be stated as follows:

Let \mathcal{M} be an m -dimensional nonempty compact metric space. We define the diameter of \mathcal{M} , denoted by $\|\mathcal{M}\|$, as the largest distance between any two points x and y in \mathcal{M} , that is

$$\|\mathcal{M}\| = \sup_{x,y \in \mathcal{M}} \|x - y\| \quad (4.10)$$

Let C_i denote a countable collection of subsets of \mathcal{M} with diameter ε_i , where $0 < \varepsilon_i \leq \delta$. Let $N(C_i)$ be the minimum number of C_i needed to cover \mathcal{M} , that is $\mathcal{M} \subset \bigcup_{i=1}^{N(C_i)} C_i$. We define

$$\Gamma_H^d(\delta) = \inf_{C_i} \sum_{i=1}^{N(C_i)} \varepsilon_i^d \quad (4.11)$$

The d -dimensional Hausdorff measure of the compact space \mathcal{M} is then defined as

$$\Gamma_H^d = \lim_{\delta \rightarrow 0} \Gamma_H^d(\delta) \quad (4.12)$$

In general, it can be shown that Γ_H^d is positive infinite if d is less than some critical value, denoted by D_H , and is zero if d is greater than D_H . The critical value D_H is called the Hausdorff dimension of the compact space \mathcal{M} [74].

Unfortunately, the Hausdorff dimension is a mathematical definition that does not lend itself to numerical implementation. Other definitions of the attractor dimension of a nonlinear dynamical system include the *capacity dimension* (also known as the *box-counting dimension*), *information dimension*, *correlation dimension*, and *Lyapunov dimension*. Parker and Chua [75] have given a good summary of these dimensions in their tutorial paper. However, we will focus on the correlation dimension, since it has emerged as the most popular one used to perform the estimation of an attractor dimension from practical time series for the following reasons: First, the ergodic characterization of a dynamical system involves partitioning the phase space \mathcal{M} into small subspaces M_1, M_2, \dots, M_K , and the probability of observing the orbit falling in the k^{th} subspace is given by N_k/N , where $k = 1, 2, \dots, K$ and N_k is the number of times the orbit falls in the k^{th} subspace out of N points in the time series. It has been shown in [33] that for finite N , there is a bias introduced in the calculation of the attractor dimension by using the estimate of the probability, except for the correlation dimension. Second, most dimension estimation procedures are based on the assumption that the deviations of the data points from the least squares line

are Gaussian-distributed random fluctuations, which arises because a finite sample of points are used to represent the attractor. This kind of statistical analysis has been performed only on the correlation dimension so far [15], where consistency and asymptotic Gaussianity of the correlation dimension have been demonstrated. Third, the minimum number of data points needed for the calculation of the correlation dimension may be specified [3, 21]. The last point is of practical significance, since the determination of attractor dimension using experimental data usually requires many data points; it provides a lower bound on the data length needed for a reliable estimation of the correlation dimension.

4.3.2 Correlation Dimension

Consider a time series $y(1), y(2), y(3), \dots$ generated in the phase space \mathcal{M} . The proportion of points that lie within a ball of radius r is equal to the m -dimensional normalized volume of that ball, so that the normalized volume of the whole space \mathcal{M} is equal to one. The volume of a ball of radius r is proportional to r^D as $r \rightarrow 0$. Accordingly, after fixing a point $y(0)$ in \mathcal{M} , we may write

$$P_r = \lim_{n \rightarrow \infty} \left\{ \frac{\text{Number of integers } k \leq n \text{ such that } |y(0) - y(k)| \leq r}{n} \right\} \quad (4.13)$$

Equation (4.13) defines the proportion of the sequence $y(0), y(1), y(2), \dots$ that lies in the ball of radius r centered at the point $y(0)$. Then, since for a D -dimensional space the proportion P_r is proportional to r^D , that is,

$$P_r \sim r^D \quad (4.14)$$

we may define the dimension D as follows [55, 94]:

$$D = \lim_{r \rightarrow 0} \left\{ \frac{\ln(P_r)}{\ln(r)} \right\} \quad (4.15)$$

where \ln is the natural logarithm.

Next we modify the above definition to match the one introduced in [34]. The definition we have just given is localized around the point $y(0)$. The dimension we seek is global in the sense that it must apply to the whole space. We therefore replace P_r with its average over the whole space, as shown by

$$A_r = \lim_{n \rightarrow \infty} \left\{ \frac{\text{Number of integers } k, m \leq n \text{ such that } |y(m) - y(k)| \leq r}{n^2} \right\} \quad (4.16)$$

The definition of Grassberger and Procaccia [34] for the correlation dimension is given by

$$D_c = \lim_{r \rightarrow 0} \left\{ \frac{\ln(A_r)}{\ln(r)} \right\} \quad (4.17)$$

On the basis of this definition, Grassberger and Procaccia [34] developed an algorithm for the numerical computation of the correlation dimension from a time series. A description of this algorithm follows.

Given a time series $\{y(i) : i = 1, 2, \dots, N\}$, we wish to reconstruct a D_E -dimensional phase portrait, which is a subspace of \mathcal{R}^{D_E} . By the use of Takens' embedding theorem, a state on the reconstructed attractor is given by

$$\Phi(i) = \{y(i), y(i + \tau), \dots, y(i + (D_E - 1)\tau)\} \quad (4.18)$$

where τ is the time delay, and D_E is the embedding dimension. It is in the nature of most strange attractors that pairs of trajectories that are initially close will eventually become temporally uncorrelated. However, both trajectories will still be on the attractor and remain spatially correlated. This cumulative correlation is defined as [34]

$$C(\epsilon) = \frac{1}{N^2} \sum_{k=1}^N \sum_{\substack{j=1 \\ j \neq k}}^N H(\epsilon - \|\Phi(k) - \Phi(j)\|) \quad (4.19)$$

where H is the Heaviside function

$$H(x) = \begin{cases} 0 & \text{if } x \leq 0 \\ 1 & \text{otherwise} \end{cases} \quad (4.20)$$

The norm $\|\Phi(k) - \Phi(j)\|$ in (4.19) may be any kind of norm measure. In practice, the Euclidean norm is used frequently, as it provides a convenient method to calculate the distance between any two points. Our calculations are based on the Euclidean norm.

The relationship between the correlation dimension and the cumulative correlation function is based upon the power law described in (4.14), which is derived from the geometric considerations discussed previously. The result is valid for "small enough" values of ϵ . A determination of the correlation dimension is found by plotting $C(\epsilon)$ versus ϵ on a log-log graph. The particular region in which the power law is obeyed usually appears as a straight line; the slope of that line is found by making a least-squares fit to this part of the curve. The slope so computed is an estimate of the correlation dimension. If the sequence of

estimates converges as the embedding dimension D_E increases, then we have an estimate of the attractor dimension D_c .

The original algorithm developed by Grassberger and Procaccia [34] is unfortunately very slow to compute, since the running time scales at least as $O(N^2)$ for the calculation of cumulative correlation in each embedding. To improve the computational efficiency of the correlation dimension, Pineda and Sommerer [79] have introduced the *box-counting algorithm* to estimate the generalized dimensions from a time series, and developed a *fast algorithm* to implement the computation. The algorithm is briefly described as follows: For a given time series, we form an m -dimensional vector series, in which each vector is a state point of the attractor in the reconstructed phase space. The first step of the basic algorithm is to scale and truncate the floating-point values into unsigned r -bit integers; then for each value, we interleave the r bits of the m r -bit integers to form a single mr -bit long integer. The tricky point of the algorithm happens at this step. That is, the box size is restricted to be of the form $\varepsilon = 2^{-k}$; we then only need to examine the first km bits of each interleaved integer to count the number of occupied boxes at each level of resolution. This can be performed efficiently by a *radix sort algorithm* that requires only $O(N)$ times. The procedure is repeated for each of the r levels of resolution, thereby getting an estimate of the attractor dimension.

Care must be exercised in computing the correlation dimension, especially for real-life data. The number of data points used in the experiment is a very crucial factor for obtaining reliable results in that the number of data must be large enough to allow the orbit to go through a neighborhood sufficiently many times. More specifically, the size of the data set cannot be smaller than a *fundamental limit* derived by Eckmann and Ruelle [21] for estimating attractor dimensions and Lyapunov exponents in dynamical systems. Based on Grassberger-Procaccia's algorithm for the correlation dimension, Eckmann and Ruelle present a relationship between the attractor dimension and the size of the data set as follows:

$$D_c \leq \frac{2 \log N}{\log(1/\rho)} \quad (4.21)$$

where $\rho = \varepsilon/L \ll 1$, L is the diameter of the reconstructed attractor, which is defined by $L = \max(|\Phi(k) - \Phi(j)|)$, and ε is the maximum mutual distance defined in (4.19). Equation (4.21) provides a lower bound on the size of the data set, N , needed for estimating the

correlation dimension from a time series by the use of the Grassberger-Procaccia algorithm. It is imperative that we adhere to this limit for experimental studies of chaos using a time series of finite length if the results obtained are to be reliable. A similar formula about the relationship between the attractor dimension and the size of the data set used to estimate the correlation dimension has been proposed by Theiler as follows [95]:

$$N \geq \sqrt{2\rho^D} \quad (4.22)$$

where D is the fractal dimension of the attractor.

4.3.3 False Nearest Neighbors Analysis

The method of false nearest neighbors [3, 54] determines the minimum embedding dimension required to reconstruct the state space of a dynamical system from observed data with time delay embedding. The basic idea of the method is that in passing from dimension d to dimension $d + 1$, one can differentiate between those points on the reconstructed trajectory $\Phi(i)$ with time delay embedding that are “true” neighbors and other points on the trajectory that are “false” neighbors. A false neighbor is a point in the data set that is a neighbor solely because we are viewing the trajectory in too small an embedding space. In other words, in an embedding dimension that is too small to unfold the attractor, not all points that lie close to one another will be neighbors because of the dynamics. Some will actually be far from each other and simply appear as neighbors because the geometric structure of the attractor has been projected down onto a smaller space. When we have achieved a large enough embedding space, all neighbors of every point on the trajectory in the multivariate phase space will be true neighbors.

Given a time series $\{y(i) : i = 1, 2, \dots, N\}$, we reconstruct a d -dimensional phase space by using time delay embedding. A state vector on the reconstructed phase space is given by

$$\Phi(i) = \{y(i), y(i + \tau), \dots, y(i + (d - 1)\tau)\} \quad (4.23)$$

where τ is the time delay. Consider the nearest neighbor $\Phi^n(i)$ of $\Phi(i)$ with nearness in the sense of some distance function, *i.e.*, Euclidean distance. Then the square of the Euclidean

distance between the point $\Phi(i)$ and the neighbor $\Phi^n(i)$ is

$$R_d^2(i, n) = \sum_{k=0}^{d-1} [y(i+k\tau) - y^{(n)}(i+k\tau)]^2 \quad (4.24)$$

$R_d(i, n)$ is presumably small when one has lots of data, and for a data set with N entries, this distance is on the order of $\frac{1}{N^{\frac{1}{d}}}$ [3]. In going from the phase space of dimension d to a phase space of dimension $d+1$ by time delay embedding, we add a $(d+1)$ th coordinate onto each of the state vectors $\Phi(i)$. This new coordinate is just $y(i+d\tau)$. Thus the Euclidean distance, measured in the phase space of dimension $d+1$, between $\Phi(i)$ and the same nearest neighbor $\Phi^n(i)$ determined in the phase space of dimension d is changed due to the new coordinates $y(i+d\tau)$ and $y^{(n)}(i+d\tau)$ to

$$R_{d+1}^2(i, n) = R_d^2(i, n) + [y(i+d\tau) - y^{(n)}(i+d\tau)]^2 \quad (4.25)$$

If $R_{d+1}^2(i, n)$ is large, we can presume it is because the near neighborliness of the two points is due to the projection from some higher-dimensional attractor down to dimension d . A natural criterion for catching embedding errors is that the increase in distance between $y(i+d\tau)$ and $y^{(n)}(i+d\tau)$ is large when going from dimension d to dimension $d+1$. A false nearest neighbor of $y(i+d\tau)$ is declared if

$$\frac{|y(i+k\tau) - y^{(n)}(i+k\tau)|}{R_d(i, n)} > R_T \quad (4.26)$$

where R_T is some threshold. In practice, it is found that for $R_T \geq 10$ the false neighbors are clearly identified [3, 54].

The method of false nearest neighbors described above supposes that data used for the computation of the distance between $y(i+d\tau)$ and $y^{(n)}(i+d\tau)$ has an infinite accuracy. But with noisy data, which is the case in most practical situations, all points have "near neighbors" that do not move apart very much as the embedding dimension is increased. As it happens, the fact that points are nearest neighbors does not mean they are close on a distance scale set by the approximate size R_A of the attractor, where R_A is the estimation of the standard deviation of the data. If the nearest neighbor to $\Phi(i)$ is not close, so $R_d(i, n) \approx R_A$, then the distance $R_{d+1}(i, n)$ will be $2R_d(i, n)$ [3]. This suggests as a second criterion for falseness of nearest neighbors if

$$R_d(i, n) \leq 2R_A \quad (4.27)$$

The decision for falseness of nearest neighbor is made when either of the two criteria described above is designated as false.

4.4 Lyapunov Characteristic Exponents

The determination of Lyapunov characteristic exponents is important in the analysis of a possibly chaotic system, since Lyapunov exponents not only show qualitatively the sensitive dependence on initial conditions but also give a quantitative measure of the average rate of separation or attraction of nearby trajectories on the attractor. In this section, we discuss the extraction of Lyapunov exponents of a nonlinear dynamical system from both the dynamical equations and time series. We begin by presenting a definition of the Lyapunov characteristic exponents for a dynamical system.

4.4.1 Definition

To determine the stability of motion of a dynamical system in an n -dimensional phase space, the asymptotic rate of a small perturbation of a trajectory is investigated by means of average exponential growth or decay rates called Lyapunov exponents. A positive Lyapunov exponent describes an exponentially increasing separation of nearby trajectories in a certain direction. This then leads to the sensitive dependence of the dynamics on initial conditions, which is one of the most characteristic features of deterministic chaos. For bounded systems, this situation is possible only if a folding property of a strange attractor is involved.

Given a dynamical system in an n -dimensional phase space, suppose we monitor the long-term evolution of an infinitesimal n -sphere of initial conditions. The sphere will evolve into an n -dimensional ellipsoid due to the locally deforming nature of the flow. The i th one-dimensional Lyapunov exponent is then defined in terms of the length of the ellipsoid principal axis $l_i(t)$ as follows [12]:

$$\lambda_i = \lim_{t \rightarrow \infty} \frac{1}{t} \log \frac{l_i(t)}{l_i(0)} \quad (4.28)$$

where the λ_i are ordered from largest to smallest. Thus the Lyapunov exponents are related to the expanding or contracting nature of principal axes in phase space. Since the orientation

of the ellipsoid changes continuously as it evolves, the directions associated with a given exponent vary in a complicated way through the attractor. We cannot therefore speak of a well-defined direction associated with a given exponent.

4.4.2 Lyapunov Exponents from Dynamical Equations

Let an m -dimensional compact manifold \mathcal{M} be the state space of a dynamical system. The system on \mathcal{M} is a nonlinear differentiable map $\phi: \mathcal{M} \rightarrow \mathcal{M}$, which can be conveniently expressed by the following difference equation:

$$\mathbf{x}(k) = \phi(\mathbf{x}(k-1)) = \phi^k(\mathbf{x}(0)) \quad (4.29)$$

Let $\mathbf{v}(0)$ denote an initial perturbation of a generic point $\mathbf{x}(0)$, and ε be a constant that is small enough. Consider the separation of trajectories of the unperturbed and perturbed points after n iterations:

$$\begin{aligned} \|\phi^k(\mathbf{x}(0)) - \phi^k(\mathbf{x}(0) + \varepsilon\mathbf{v}(0))\| &= \|D\phi^k(\mathbf{x}(0))\mathbf{v}(0)\varepsilon\| + \mathcal{O}(\varepsilon^2) \\ &= \left\| \left(\prod_{i=0}^{k-1} D\phi(\mathbf{x}(i)) \right) \mathbf{v}(0)\varepsilon \right\| + \mathcal{O}(\varepsilon^2) \end{aligned} \quad (4.30)$$

where $D\phi(\mathbf{x}(i))$ is the Jacobian, that is, the $m \times m$ matrix of partial derivatives of ϕ evaluated at the point $\mathbf{x}(i)$. For a one-dimensional system the average exponential rate of separation is

$$\lim_{k \rightarrow \infty} \frac{1}{k} \log \left| \prod_{i=0}^{k-1} \phi'(\mathbf{x}(i)) \right| = \lim_{k \rightarrow \infty} \frac{1}{k} \sum_{i=1}^{k-1} \log |\phi'(\mathbf{x}(i))| \quad (4.31)$$

In general, the occurrence of products of noncommutative matrices in (4.30) renders the existence and evaluation of the limits nontrivial. It has been shown, however, by Oseledec [73] that analogous quantities can be defined and evaluated according to the following theorem [20, 73, 81]

Theorem (Oseledec): Let ρ be a probability measure on a space \mathcal{M} , and $\phi: \mathcal{M} \rightarrow \mathcal{M}$, a measure preserving map such that ρ is ergodic. Let also $D: \mathcal{M} \rightarrow \mathcal{D}$, the space of $m \times m$ matrices, be a measurable map such that

$$\int \rho(d\mathbf{x}) \log^+ \|D(\mathbf{x})\| < \infty \quad (4.32)$$

where $\mathbf{x} \in \mathcal{M}$, $\log^+ u = \max(0, \log u)$. Define the matrix $D\phi^k(\mathbf{x}) = \prod_{i=0}^{k-1} D(\phi^i(\mathbf{x}))$, $\phi^0(\mathbf{x}) = \mathbf{x}$, then for ρ -almost all \mathbf{x} , the following limits exist:

$$K = \lim_{k \rightarrow \infty} [D\phi^k(\mathbf{x})^* D\phi^k(\mathbf{x})]^{\frac{1}{2k}} \quad (4.33)$$

where the asterisk denotes the adjoint operation.

Let the distinct eigenvalues of this matrix be denoted by $\{d_i : i = 1, \dots, m\}$. The logarithms of the number d_i , that is, the Lyapunov exponent λ_i , can be ordered as $\lambda_1 > \lambda_2 > \dots > \lambda_m$, and the subspaces of the tangent space of ϕ at \mathbf{x} corresponding to all eigenvalues $\leq \exp \lambda_i$ can be labeled by E_i . Then the following holds [20, 73, 81]

Theorem: For ρ -almost all \mathbf{x}

$$\lambda_i = \lim_{k \rightarrow \infty} \frac{1}{k} \log \|D\phi^k(\mathbf{x})v\|, \quad \text{for } v \in E_i \setminus E_{i+1} \quad (4.34)$$

On the basis of this theorem, an efficient algorithm for the evaluation of these Lyapunov exponents was proposed by Benettin *et al* [12]. The method can be described in principle as follows. In the tangent space of ϕ at x_0 , an arbitrary complete set of orthonormal basis vectors $\{v_1^0, \dots, v_m^0\}$ is chosen. For $j = 1, \dots, k$ subsequent steps, the following calculations are then performed. First, we determine the vectors

$$w_i^j = D\phi(\mathbf{x}(j-1))v_i^{j-1} \quad (4.35)$$

where $i = 1, \dots, m$. By orthogonalization, these vectors can be restricted to the subspaces M_i , and stretching factors d_i^j are obtained by

$$d_1^j = \|w_1^j\| \quad \text{and} \quad d_k^j = \|\hat{w}_k^j\| \quad (4.36)$$

with

$$\hat{w}_k^j = w_k^j - \sum_{l=1}^{k-1} \langle v_l^j, w_k^j \rangle v_l^j \quad (4.37)$$

where $k = 2, \dots, m$ and

$$v_l^j = \frac{\hat{w}_l^j}{d_l^j} \quad (4.38)$$

The Lyapunov exponents are finally obtained from

$$\lambda_i = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=0}^{n-1} \log(d_i^j) \quad (4.39)$$

The computational procedures of Lyapunov spectrum for a two-dimensional dynamical system can be illustrated by Figure 4.1. In this figure, each vector v_1 tends to relax in the direction that permits the largest stretching. After some transient steps, the first vectors v_1^j will point in the direction of fastest separation.

Unfortunately, this method cannot be applied directly to experimental data, since when we work with a time series we usually do not know the underlying dynamical equations responsible for generating the time series. The next subsection will discuss how to extract Lyapunov exponents of a chaotic system from experimental data.

4.4.3 Lyapunov Exponents from a Time Series

Typically, experimental data consist of discrete measurements of a single observable. The well-known technique of phase-space reconstruction with delay coordinates makes it possible to obtain from such a time series an attractor whose Lyapunov spectrum is identical to that of the original attractor [93, 76].

Suppose that we are given a time series $y(n)$, which is an observation of an m -dimensional chaotic dynamical system $\phi : \mathcal{M} \rightarrow \mathcal{M}$. A D_E -dimensional phase space \mathcal{R}^{D_E} is reconstructed by using Takens' embedding theorem [93, 76]. Then, a state point on the reconstructed phase space is given by

$$\Phi(n) = \{y(n), y(n + \tau), \dots, y(n + (D_E - 1)\tau)\} \quad (4.40)$$

where τ is the time delay discussed in section 4.2, and D_E is the embedding dimension. Without loss of generality, we suppose $\tau = 1$. Thus $\Phi(n), n \in [1, \infty)$ defines a trajectory on the reconstructed phase space, and an equivalent dynamical system $\psi : \mathcal{R}^{D_E} \rightarrow \mathcal{R}^{D_E}$ then is

$$\Phi(n + 1) = \psi(\Phi(n)) \quad (4.41)$$

To calculate the Lyapunov exponents, $D\psi(\Phi(n))$ should be known at every point $\Phi(n)$. When we incorporate ideas put forward by Echmann and Ruelle [20], the Lyapunov exponents can be approximately calculated from the information contained in the dynamical map if we replace the linearized flow $D\psi$ by a suitable linear approximation $\tilde{D}\psi$. To construct this approximation at $\Phi(k)$, the images of various points in the neighborhood of $\Phi(k)$

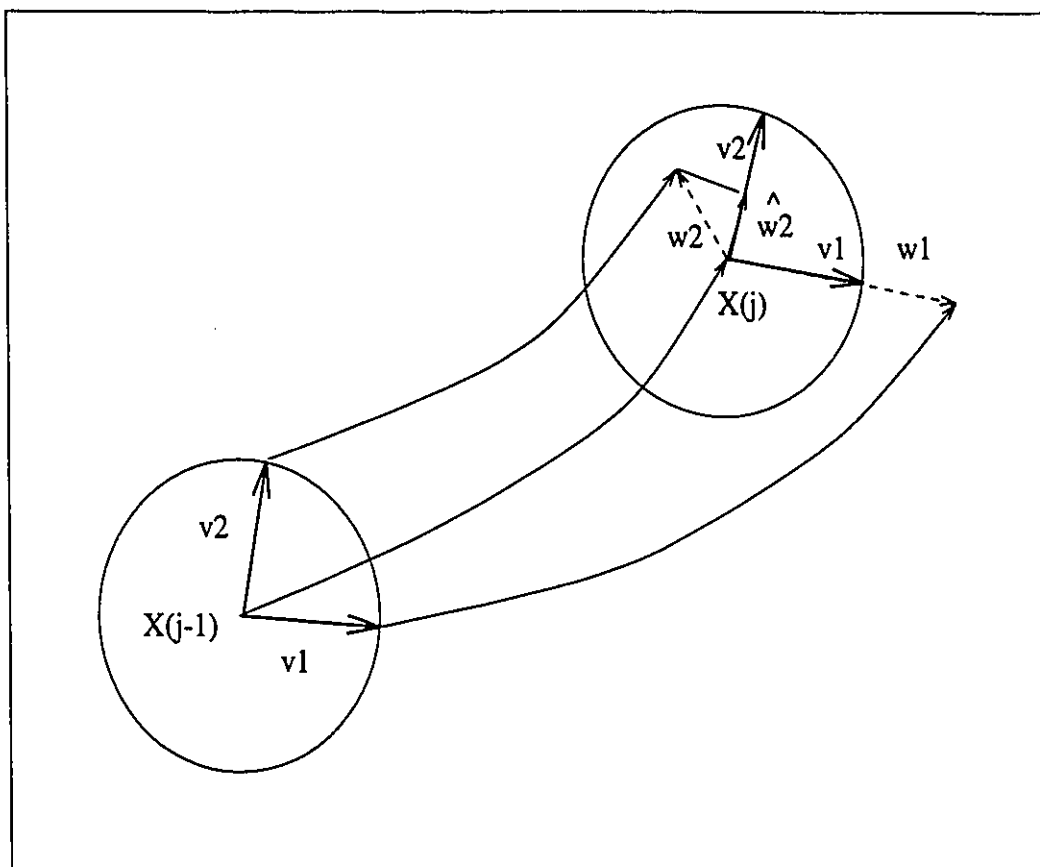


Figure 4.1: Illustration of the procedure used to calculate Lyapunov exponents from dynamical equations.

are investigated. Let $\Phi_l(k)$ be a point within a small ball of radius r around $\Phi(k)$. Then define

$$\mathbf{v}_i^k = \Phi_l(k) - \Phi(k) \quad (4.42)$$

and

$$\mathbf{w}_i^{k+1} = \Phi_l(k+1) - \Phi(k+1) \quad (4.43)$$

The relation

$$\mathbf{w}_i^{k+1} = \tilde{D}\psi(\Phi(k))\mathbf{v}_i^k \quad (4.44)$$

determines the matrix $\tilde{D}\psi(\Phi(k))$ if D_E linearly independent vectors \mathbf{w}_i are given. If the corresponding points $\Phi_l(k)$ can be found within a neighborhood of $\Phi(k)$, where a linear approximation can be justified, the constructed $\tilde{D}\psi(\Phi(k))$ will be a good approximation of $D\psi(\Phi(k))$. In that case $\tilde{D}\psi(\Phi(k))$ can be used for the evaluation of Lyapunov exponents as outlined in the previous section.

In most practical applications, usually only a limited set of noisy data points is available. The distribution of points will be rather inhomogeneous since a smooth distribution of points along the unstable direction but a singular distribution along the stable direction of the manifold is expected. To get more precise information about the stable directions, we need to investigate the mapping of L neighboring points of $\Phi(k)$, where L is usually considerably larger than the embedding dimension D_E . The matrix $\tilde{D}\psi(\Phi(k))$ is then determined from a least-square fit in the following way:

$$\tilde{D}\psi(\Phi(k)) = A(\Phi(k))B(\Phi(k))^\dagger \quad (4.45)$$

where \dagger denotes the Moore-Penrose pseudo-inverse operation, and

$$A(\Phi(k)) = W(\Phi(k))V(\Phi(k))^T \quad \text{and} \quad B(\Phi(k)) = V(\Phi(k))V(\Phi(k))^T \quad (4.46)$$

with

$$W(\Phi(k)) = [\mathbf{w}_1^{k+1}, \mathbf{w}_2^{k+1}, \dots, \mathbf{w}_L^{k+1}] \quad \text{and} \quad V(\Phi(k)) = [\mathbf{v}_1^k, \mathbf{v}_2^k, \dots, \mathbf{v}_L^k] \quad (4.47)$$

Because of the limited number of data points available, a compromise always has to be made between choosing a sufficient number of data points within a small ball of radius r to reflect the geometric structure of the attractor in all direction and selecting only nearest

neighbors such that a linear approximation of ψ can be justified. The method described above is proposed by Stoop and Meier [90]. Note that in the above method a “local linear” approach is applied to build the Jacobian matrix, as formulized by (4.42), (4.43) and (4.44). To get more accurate estimates of the Lyapunov exponents, it may be necessary to use higher (*i.e.*, third)-order polynomial fit to build the Jacobian matrix [3, 13].

Wolf’s Method

To justify a given time series to be possibly chaotic, in most practical cases, it is only necessary to evaluate the largest Lyapunov exponent. Wolf *et al.* [101] have developed an algorithm for estimating the largest positive Lyapunov exponent from experimental data, which is started from the definition given previously and well suited for the purpose of the research reported in this thesis.

Again, suppose that we are given a time series $y(n)$. A D_E -dimensional phase space is reconstructed by using Takens’ embedding theorem. Then, a state point on the reconstructed phase space is given by

$$\Phi(i) = \{y(i), y(i+1), \dots, y(i+D_E-1)\} \quad (4.48)$$

Thus $\Phi(n)$ defines a trajectory on the reconstructed phase space. To estimate the largest Lyapunov exponent, we can in effect monitor the long-term evolution of a single pair of nearby orbits. The reconstructed attractor, though defined by a single trajectory, can provide points that may be considered to lie on different trajectories.

A schematic representation of the procedure is shown in Figure 4.2. The solid line in this figure signifies a *fiducial trajectory*. We denote an initial point on the fiducial trajectory as $\Phi(t_0) = \{y(t_0), y(t_0 + \tau), \dots, y(t_0 + [D_E - 1]\tau)\}$. Consider the nearest neighbor of the initial point, and denote the corresponding distance between these two points as $L(t_0)$, which we call the *initial length*. Suppose that the time series $y(t)$ is long enough, so that for each state point on the fiducial trajectory there exists a neighbor such that $L(t_i) < r$, $i = 0, 1, \dots$, where r is the radius of a sphere on \mathcal{R}^{D_E} . It is assumed that the radius r is much smaller than the extension of the reconstructed attractor. After one mean orbital period, the initial length will have evolved to length $L'(t_1)$, which may be larger than r . We then replace the

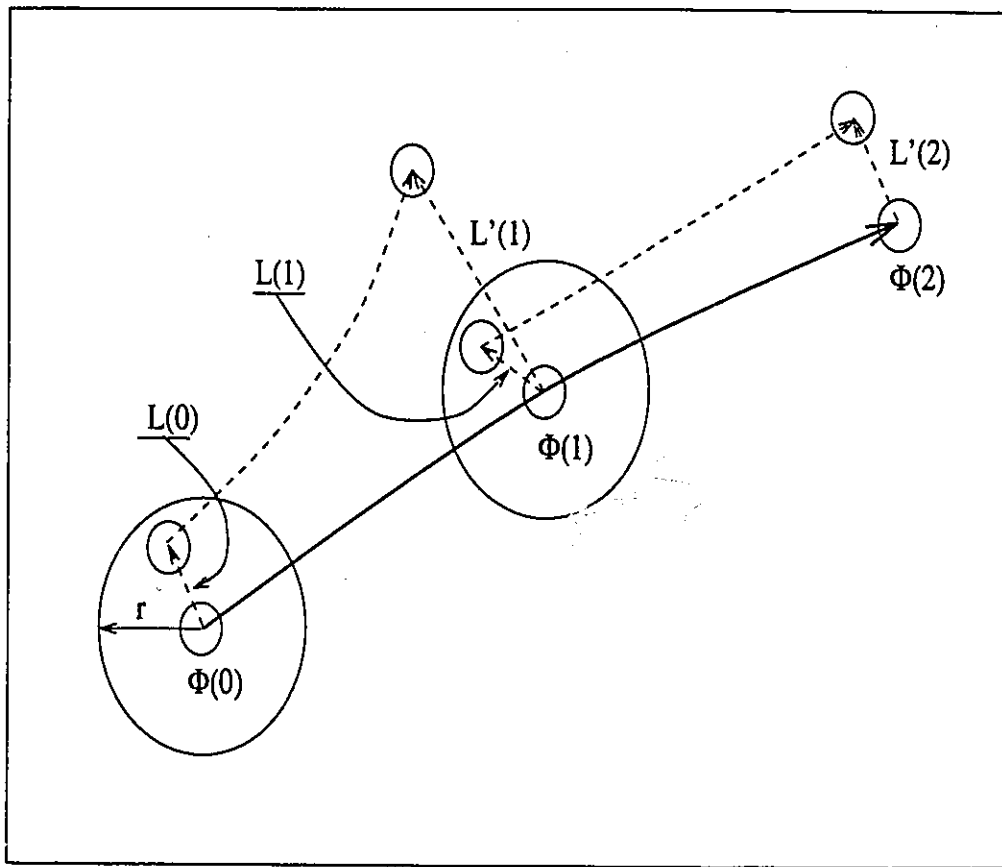


Figure 4.2: A schematic representation for estimating the largest Lyapunov exponent from a time series.

non-fiducial data point with a point closer to the fiducial point $\Phi(t_1)$, for which $L(t_1) < r$, in nearly the same direction as the original vector. This procedure is repeated until all the points of the trajectory are exhausted. The largest Lyapunov exponent, λ_1 , is then estimated by using the formula [101]

$$\lambda_1 = \frac{1}{t_K - t_0} \sum_{k=1}^K \log \frac{L'(t_k)}{L(t_{k-1})} \quad (4.49)$$

where K is the total number of replacement steps. The important point of this algorithm is that, through a simple replacement procedure that attempts to preserve orientation and minimize the size of the replacement vectors, we can monitor the long-term behavior of a single principal-axis vector, and thus get an estimate of the largest Lyapunov exponent of the system.

Wolf's algorithm can be easily extended to estimate $\lambda_1 + \lambda_2$ by considering two nearest neighbors of the initial fiducial point, and replacing the initial length in the preceding algorithm by the area determined by these three points. But, as pointed out in [3, 101], Wolf's method almost always gives accurate values for only the largest Lyapunov exponent λ_1 due to inevitable numerical errors.

4.5 Summary

In this chapter, we have described methods to extract chaotic parameters from a given time series. We started this chapter by describing the reconstruction of chaotic dynamics from observation.

Dimension analysis is commonly used to distinguish a given dynamical system to be whether deterministic or random. In this chapter, we presented the definition of Hausdorff dimension, then described the correlation dimension that is well suited to our research. Various methods to evaluate the correlation dimension from a given time series are given and discussed. The method of false nearest neighbors analysis for finding the minimum embedding dimension for chaotic reconstruction of a given time series was presented. The need for an adequate data length was stressed.

Lastly in this chapter, we discussed the estimation of Lyapunov exponents, which

is an important characteristic in the analysis of a possibly chaotic system.

Chapter 5

Experimental Database Description

5.1 Introduction

The Communications Research Laboratory (CRL) of McMaster University is committed to a program of research to develop improved marine radar techniques for the detection and classification of icebergs and sea ice. In response to the equipment requirements of this program, the CRL has developed a research oriented radar system called IPIX radar [41]. Figure 5.1 shows the IPIX radar trailer and the antenna. This radar has enabled us to make real measurements of growler and sea backscatter at X-band. The database used for the study was collected by using the system.

The IPIX radar has the following features. It is dual-polarized and fully coherent. It has two identical receivers: one is connected to the vertical polarized antenna feed and the other to the horizontal polarized antenna feed. Each receiver provides the in-phase and quadrature output components for subsequent processing. This property is useful in our study of nonlinear dynamical systems, since it provides a basis for comparison to the single-component dynamic reconstruction method described in Chapter 4. The IPIX radar has built-in calibration equipment, which permits quantitative measurements to be made with high confidence, and a digital control system which allows experimental parameters to be easily varied. These provisions make the radar measurements highly reliable.

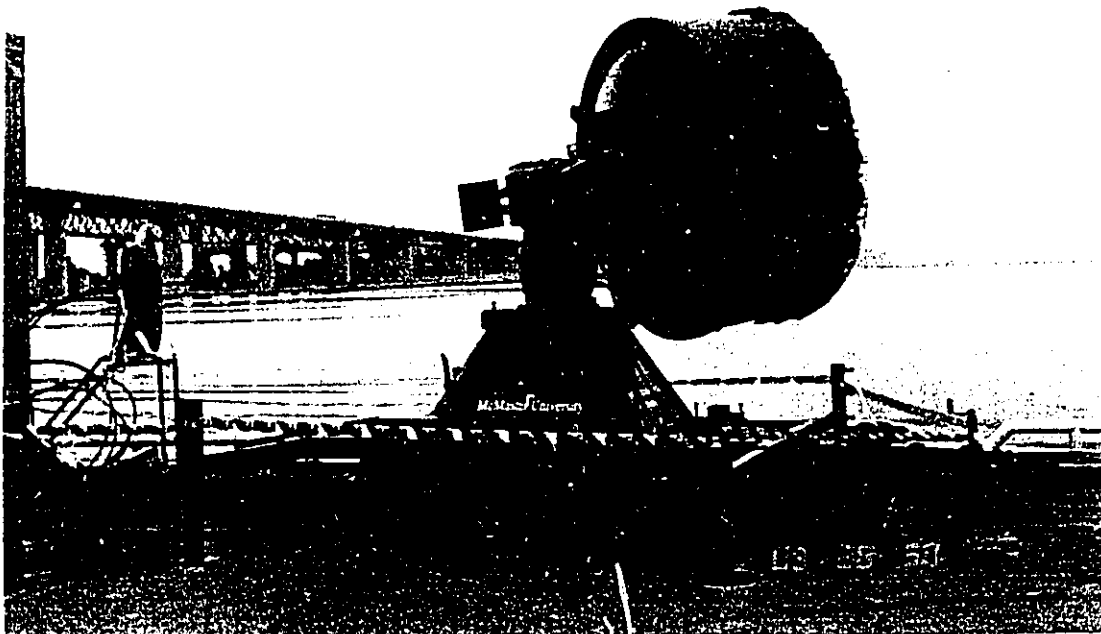
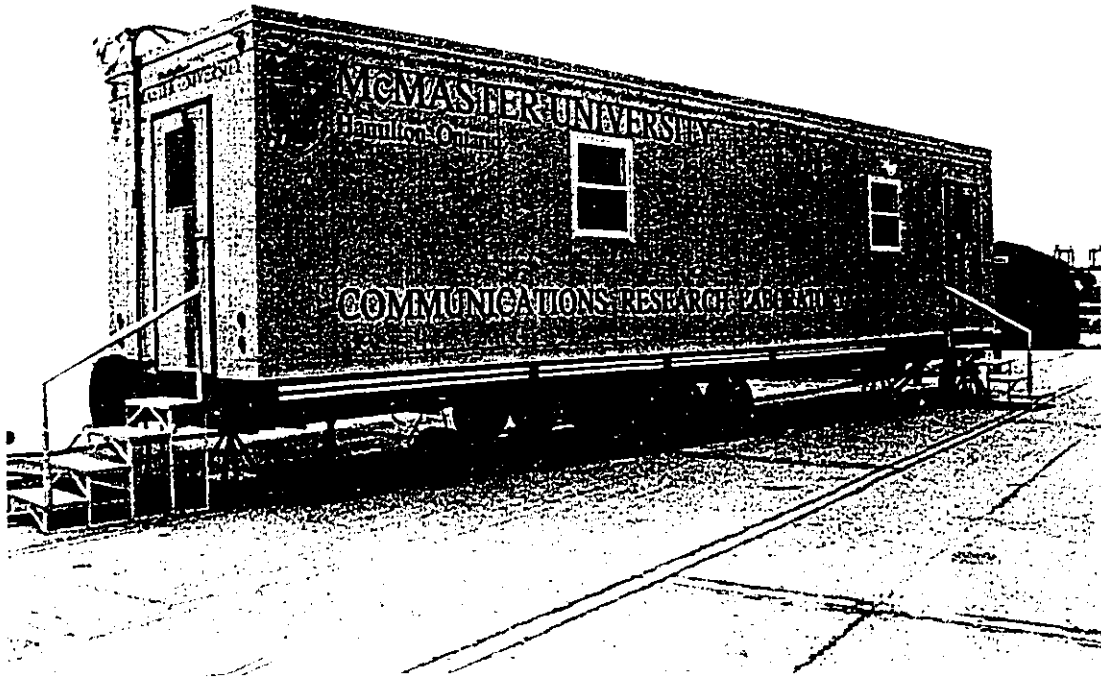


Figure 5.1: The IPIX radar trailer and antenna used for collection of the database reported in this chapter.

The database described in this chapter contains data collected in three different locations with different sea states.

5.2 Experiments at Cape Bonavista, Newfoundland

The IPIX radar system was first installed at Cape Bonavista, Newfoundland, Canada, in May 1989. Floating ice targets of all sizes commonly pass within radar range of this site during the Spring season and a wide variety of sea states can be observed throughout the year. The radar antenna was mounted at a height of about 25 meters above the water, similar to that of a typical ship-mounted radar. During the field trip of June 1989, various radar data, including returns from icebergs, fishing vessels and sea clutter, in various sea states, were collected. Calibrations were performed, which allowed the processing software to relate digitized signal values to an absolute power level at the receiver input.

Table I describes the radar and weather parameters of some data collected on the trip. The relevant parameters are: ID (identification of each data pack), TYPE (target type: c=clutter and g=growler), POL (transmit receive polarization, h=horizontal and v=vertical), PRF (pulse repetition frequency in Hz), PULSE (pulse width in nano-second), RSTART (distance in meters from the radar to the first sample collected), WSPEED (wind speed in knots, which were provided by the Atmospheric Environment Service), and WHEIGHT (wave height in feet, which were provided by a nondirectional waverider).

For each data package, there are 38 range gates; in each range gate there is an I-channel data set and a Q-channel data set, each consisting of at least 40,000 samples from consecutive scans.

The data sets used in the studies represented here were all obtained with the IPIX radar operating in a dwelling mode, that is, with the antenna pointing toward a patch of the ocean surface along a fixed direction. The sea states for the data sets listed in Table I vary from near calm to rough.

ID	DATE	TYPE	POL	PRF	PULSE	RSTART	WSPEED	WHEIGHT
B79	6/6/89	c	hh, hv	2000	200	3000	13	3.78
B86	6/6/89	c	hh, hv	2000	200	3000	13	3.51
B95	7/6/89	c	hh, hv	2000	200	3300	8	5.74
B98	7/6/89	g	hh, hv	2000	200	6420	8	4.59
B115	7/6/89	g	hh, hv	2000	200	4080		5.00
B218	9/6/89	c	hh, hv	2000	200	7980	11	7.48
B251	10/6/89	c	hh, hv	2000	200	4800	10	5.95
B318	10/6/89	c	hh, hv	2000	200	5700	27	8.30
B362	10/6/89	c	hh, hv	2000	200	7950	12	6.98
B379	10/6/89	c	hh, hv	2000	200	6030	4	6.70

Table 5.1: Radar and sea state parameters for the Cape Bonavista database

ID	DATE	TYPE	POL	PRF	PULSE	RSTART	WHEIGHT
D6s5	6/11/93	c	hh, hv	2000	200	2100	7.45
D7s4	7/11/93	c	hh, hv	2000	200	2500	6.05
D12s4	12/11/93	c	hh, hv	2000	200	2100	5.28
D12s15	14/11/93	c	hh, hv	2000	200	2100	6.70
D14sa0	14/11/93	c	hh, hv	2000	200	2100	4.90
D17sa0	17/11/93	c	hh, hv	2000	200	2100	5.60
D17sa1	17/11/93	c	hh, hv	2000	200	2100	4.25

Table 5.2: Radar and sea state parameters for the Dartmouth database

5.3 Experiments at Dartmouth, Nova Scotia

Further field experiments at Dartmouth, Nova Scotia, Canada, were carried out in November 1993, using a refined version of the IPIX radar. The radar antenna was mounted at a height of about 30 meters above the water level and within 15 meters of the shore. The radar system was maintained with the same parameters as those for the trip to Bonavista. Each data package shown in Table 2 includes 14 range gates with both I-channel and Q-channel samples. In each channel about 260,000 samples are recorded per experiment. The radar system was again operating in the dwelling mode. The sea states for the data shown in Table 2 fall in the range of moderate to rough.

5.4 Experiments at Lake Ontario

The IPIX radar was also set up at Fifty Point Conservation Area, Stoney Creek, Ontario, in May 1993. Its antenna was mounted approximately 10 meters above the water level and within 100 meters of the shore. Since the lake state does not change as much as an ocean environment, we will use only one data set, in which the radar parameters have the same values as those described earlier. The data set has a length of about 250,000 samples. The wave height for the data set was about 4.9 feet.

5.5 Summary

In this chapter, a database used for the chaotic analysis of sea clutter and the detection of small target in sea clutter is described. The database contains data collected at two locations on Canada's east coast, overlooking the ocean, and one near McMaster, overlooking Lake Ontario. At each location, data sets were collected over a number of surface conditions, with radar antenna fixed, staring at a particular patch of the water surface.

Chapter 6

Chaotic Characterization of Sea Clutter

6.1 Introduction

Sea clutter refers to radar backscatter from a sea surface. It is a very complicated process that is highly dependent on the sea state. To apply the new chaos-based detection method described in Chapter 3 to the radar detection of a small target in sea clutter, we first have to demonstrate that sea clutter is indeed chaotic. To do this, we use the radar database described in Chapter 5, and present results on correlation dimension analysis and false nearest neighbors' analysis, the extraction of Lyapunov exponents from sea clutter data by applying the methods described in Chapter 4. The validity of the software for performing these estimations is demonstrated by applying it to a chaotic time series generated by the Lorenz system; the results of these computations are presented in the Appendix at the end of the thesis.

We study the effects of various factors on the extraction of chaotic characteristics of sea clutter. These factors include: data length, different channels, background noise, radar site locations, and sea states. After the relationships between these parameters and the chaotic model have been confirmed, we build a model that estimates the chaotic dynamics of sea clutter. The reconstruction of sea clutter dynamics not only provides confirmation of local predictability, which is an important property of a chaotic system, but also gives a predictive model for subsequent application of the signal detection strategy.

This chapter is organized as follows. In Section 6.2, we present results on the correlation analysis using the radar database described in Chapter 5. In this section, various data sets are used to demonstrate that the dynamical mechanism responsible for the generation of sea clutter is deterministic and global. In Section 6.3, we present experimental results on the computation of Lyapunov exponents to further demonstrate the chaotic nature of sea clutter. The experimental results of false nearest neighbors' analysis for finding the minimum embedding dimension for the reconstruction of chaotic dynamics of sea clutter are presented in Section 6.4. In Section 6.5, we study the estimation of sea clutter dynamics. Two neural network structures are proposed and implemented. The local predictability of sea clutter based on the models is studied. In Section 6.6, a method for defining a new strange attractor is presented. This chapter is concluded by a summary in Section 6.7.

6.2 Correlation Dimension of Sea Clutter

In our database, each sea clutter data set has at least 40,000 samples. To choose a time delay τ for dynamical reconstruction, we calculate the mutual information from time delayed time series of sea clutter using the method described in Section 4.2. A plot of the mutual information versus time delay is shown in Figure 6.1. We see that the minimum mutual information occurs at a time delay $\tau = 5$. To further confirm the result of Figure 6.1, we embed the data in one and two dimensions, and minimize the mutual information dimension D_μ with respect to τ , as described in Section 4.2. Figure 6.2 shows a plot of D_μ versus τ . The minimum of D_μ is seen to occur at $\tau = 5$. For the above computations, a data file of total 40,000 samples was used. Thus, calculations of the correlation dimension, false nearest neighbors, the Lyapunov exponents, and estimation of the sea clutter dynamics are all based on a time delay $\tau = 5$.

A fast algorithm developed by Pineda and Sommerer [79] was used to calculate the correlation dimension of sea clutter. Figures 6.3 through 6.10 show the results of the correlation dimension analysis for data packages listed in the last chapter.

First, the effect of the data length on the estimation of the correlation dimension is studied. It is found that for a data set longer than 10,000 samples, the estimation of

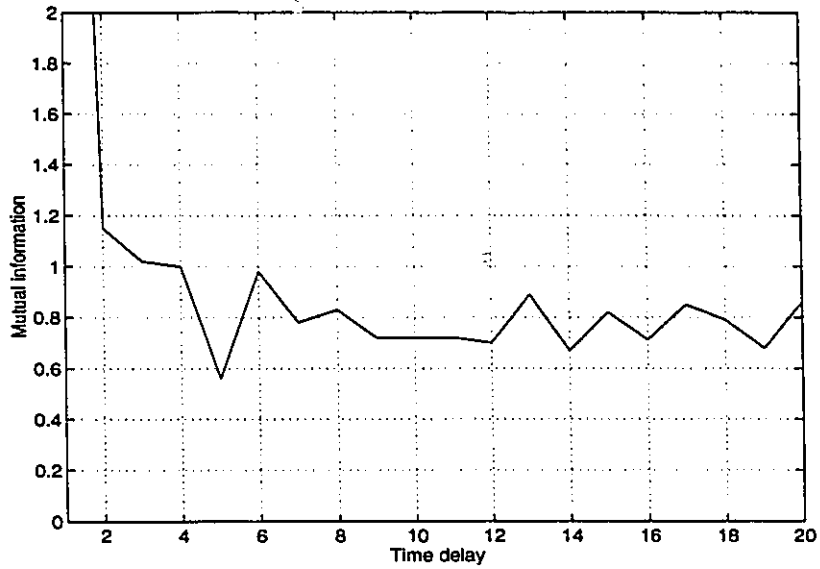


Figure 6.1: Mutual information versus time delay for sea clutter data. The minimum mutual information dimension occurs at $\tau = 5$.

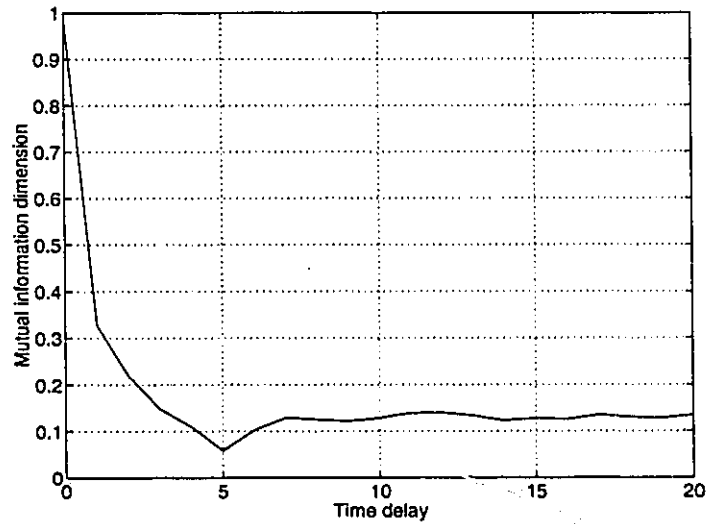


Figure 6.2: Mutual information dimension versus time delay for sea clutter data. The minimum mutual information dimension occurs at $\tau = 5$.

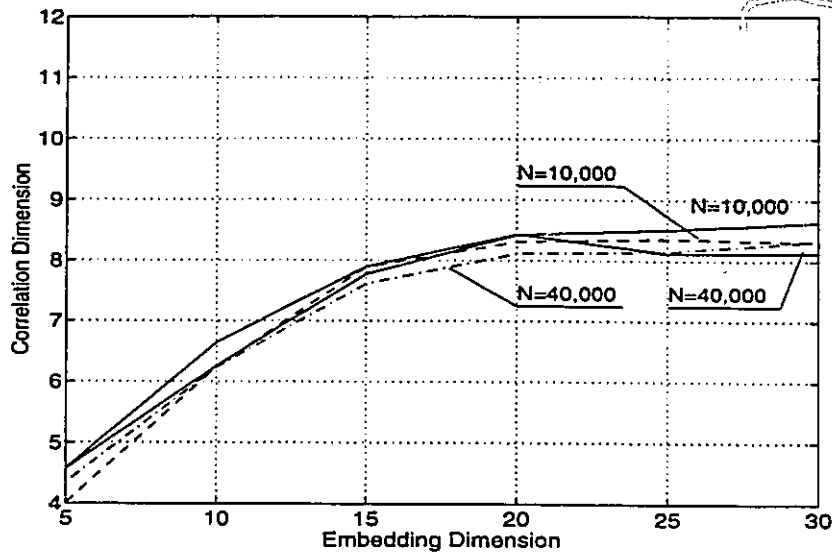


Figure 6.3: Correlation dimension analysis for sea clutter data set $B318_20$, where N is the length of the data set used for the computation. The solid curves correspond to the raw sea clutter data, and the dashed curves correspond to the data after being smoothed.

correlation dimension for sea clutter may be considered to be reliable. Figure 6.3 shows four curves of the correlation dimension as a function of the embedding dimension, corresponding to the data set $B318_20$. The two solid curves are the results for the original sea clutter data, in which one corresponds to $N = 10,000$ and the other corresponds to $N = 40,000$, where N is the total number of data points used to implement the computation. The two dashed curves included in Figure 6.3 are the results obtained for the same data set smoothed by a filter, more on which will be discussed later. The results are almost the same for the four cases. We believe that these results are reliable in light of the fundamental limit of equation (4.21). If we set $\rho = 0.1$, which refers to the maximum value of the ratio of possible noise power to received sea clutter power, then a data set with $N = 40,000$ samples satisfies the lower bound for estimating a correlation dimension $d_c \leq 9.4$. Figure 6.3 shows that the correlation dimension for data set $B318_20$ is about 8.4; hence, we are well within the Eckmann-Ruelle limit.

To test the influence of noise on the estimation of correlation dimension, we choose two sea clutter data sets, $B318_20$ and $B98_35$. $B318_20$ has a higher clutter-to-noise ratio than that of $B98_35$, as the scattering distance (radar range) for $B98_35$ is much larger

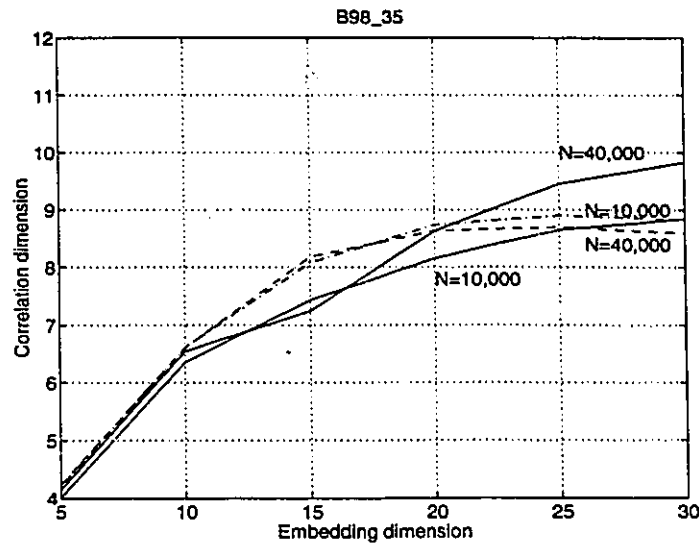


Figure 6.4: Correlation dimension analysis for sea clutter data set *B98_35*, where N is the length of the data set used for the computation. The solid curves correspond to the raw sea clutter data, and the dashed curves correspond to the data after being smoothed.

than that for *B318_20*. A smoothing filter of the form $\hat{x}(n) = \sum_{i=1}^3 a_i x(n+i-2)$ with $a_i = 1/3$ was used to smooth the raw sea clutter data. The dashed lines in Figures 6.3 and 6.4 are results based on the smoothed data sets, and the solid lines are the results obtained from raw data sets. For the almost noiseless data set *B318_20*, the results obtained from both the smoothed and the original data are essentially the same, as shown in Figure 6.3. But for the more noisy data set *B98_35*, the estimation of correlation dimension of the nonsmoothed data set does not appear to converge with increased embedding dimension as shown by the two solid curves in Figure 6.4. In particular, the two solid curves in Figure 6.4 show that the longer the data set is, the larger the correlation is. When, however, the data set is smoothed, the estimation of the correlation dimension converges as the embedding dimension is increased, and the results for both $N = 10,000$ and $N = 40,000$ are almost the same as those shown by the two dashed curves in Figure 6.4. The reason why the solid curves in Figure 6.4 diverge is because the effect of receiver noise is much more pronounced in that particular data set; and a random system has, theoretically, infinite dimension.

Care must be exercised in choosing a smoothing filter for a chaotic time series. Reference [10] reported that the Lyapunov dimension increases in filtered chaotic signal since

a filter with limited bandwidth introduces another negative Lyapunov exponent and that exponent may make the Lyapunov dimension increase. This is not the case in our application. Generally speaking, a chaotic time series has a relatively wide band spectrum. If the smoothing filter is chosen to have a sufficient bandwidth so that all frequency components of the chaotic signal fall in the pass-band of the filter, then the filter cannot affect the characterization of the chaotic signal. In our application, the frequency components of sea clutter concentrate mostly in the low frequency region (see Figure 6.37), and the smoothing filter chosen for the application is actually a simple low-pass filter with a bandwidth much higher than that of the sea clutter component.

To compare the correlation dimension of sea clutter with that of *filtered noise*, we created a new randomized time series with exactly the same Fourier spectrum as the original sea clutter time series. Specifically, we take the Fourier transform of the original time series, randomize the phase, and then compute the inverse Fourier transform of the "surrogate data" so generated, as described in [72, 95, 98]. Figure 6.5 shows the results of this calculation. The dashed lines pertain to the original sea clutter data set *B318_20*, while the solid lines pertain to the randomized version of this data set. For the randomized data set (i.e., filtered noise) there is no saturated region even though it has exactly the same Fourier spectrum as the original data set. The longer the data set, the larger the computed correlation dimension. It appears that the time delay has little effect on the computation for the randomized data set. It is expected if a random noise process has infinite length, then the computed correlation dimension would increase linearly with respect to the embedding dimension. This is precisely the result shown in Figure 6.5 for the randomized data set. In this experiment, the phase is chosen to be uniformly distributed in $[0, 2\pi]$. It is generated by computation means. A result of the correlation dimension of this process is shown in Figure 6.6. Clearly, there is no saturated region.

As mentioned at the beginning of this section, our radar system is a coherent system. It provides both I-channel and Q-channel data. Let these components be denoted by $y_I(n)$ and $y_Q(n)$ respectively. We may thus express the complex baseband form of the received radar signal as follows:

$$y(n) = y_I(n) + jy_Q(n) \quad (6.1)$$

where $j = \sqrt{-1}$. Correspondingly, we may express the amplitude (envelope) component of

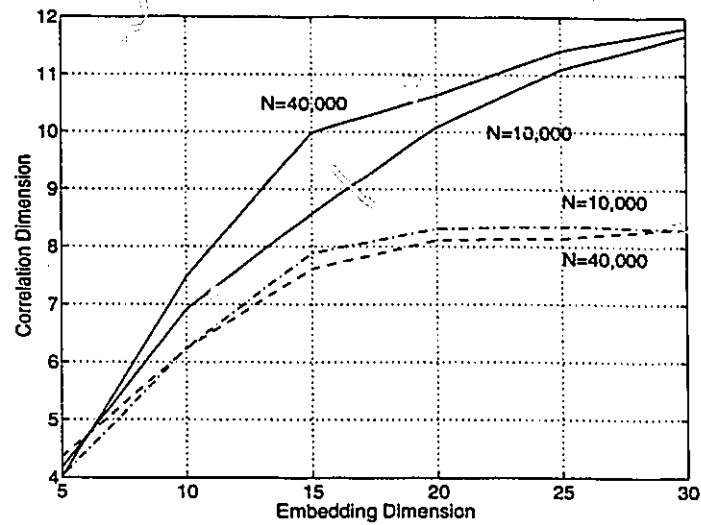


Figure 6.5: Correlation dimension analysis for sea clutter data set *B318_20*, where N is the length of the data set used for the computation. The dashed curves correspond to the sea clutter data, while the solid curves correspond to a filtered noise process (surrogate data) with exactly the same Fourier spectrum as that of the original clutter data.

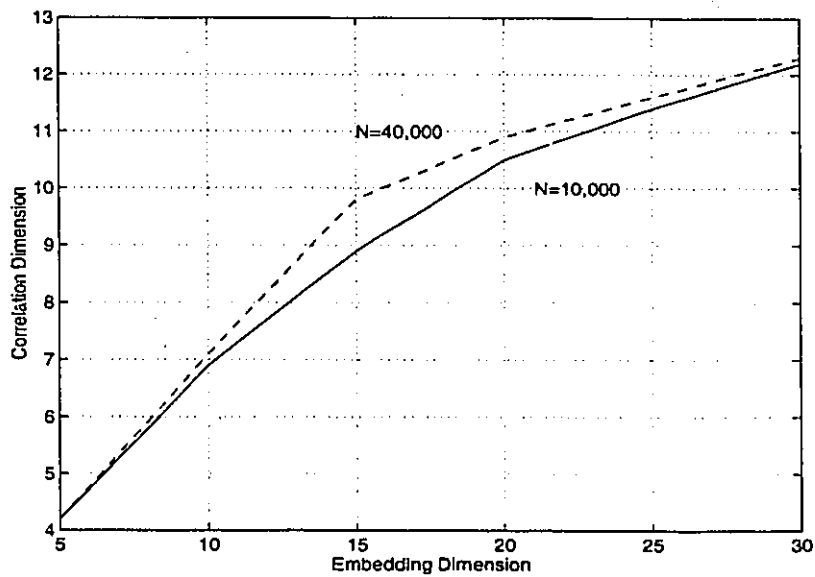


Figure 6.6: Correlation dimension of a uniformly distributed random process used to generate the "surrogate data" used in Figure 6.6

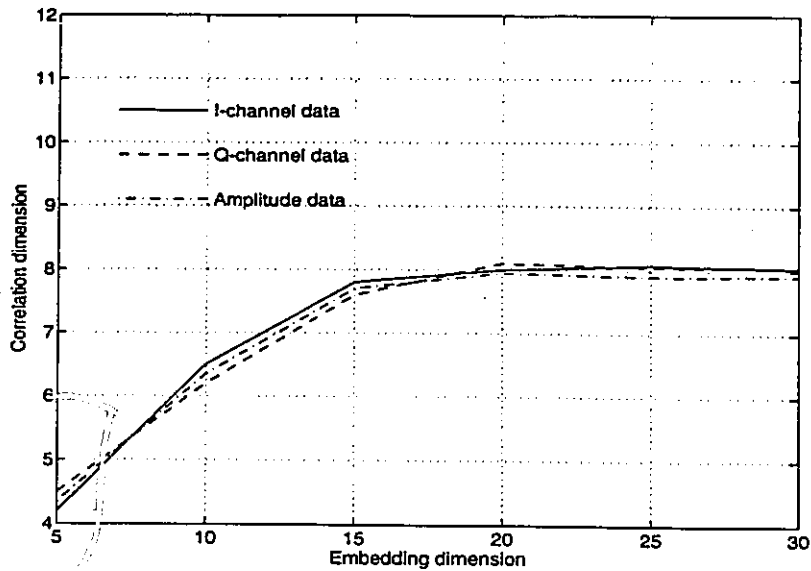


Figure 6.7: Correlation dimension analysis for different received data. A data set of 30000 samples is used for each computation.

the received radar signal as

$$y_A(n) = \sqrt{y_I^2(n) + y_Q^2(n)} \quad (6.2)$$

Hence, we have three different sea clutter components to consider. Figure 6.7 shows the result of applying the correlation dimension analysis to these three components. The solid curve corresponds to the I-channel data, the dashed curve corresponds to the Q-channel data, and the dash-dotted curve corresponds to the amplitude data. The resulting values of the correlation dimension are all very close to each other. The results presented here provide convincing evidence that a single component (in-phase, quadrature, or amplitude) is sufficient for correlation dimension analysis, in accordance with Takens' embedding theorem.

Antenna polarization is also another important feature to be considered in radar system design. The IPIX radar system is dual-polarized. It provides both the like-polarized (*i.e.*, HH) data and the cross-polarized (*i.e.*, HV) data simultaneously. It is therefore desirable to study the effect of polarization on correlation dimension analysis. Figure 6.8 shows the results of this analysis for different polarizations. It can be seen from this figure that the results for both HH and HV polarizations are close to each other. In other words, the correlation dimension of sea clutter is essentially invariant to the choice of radar polarization.

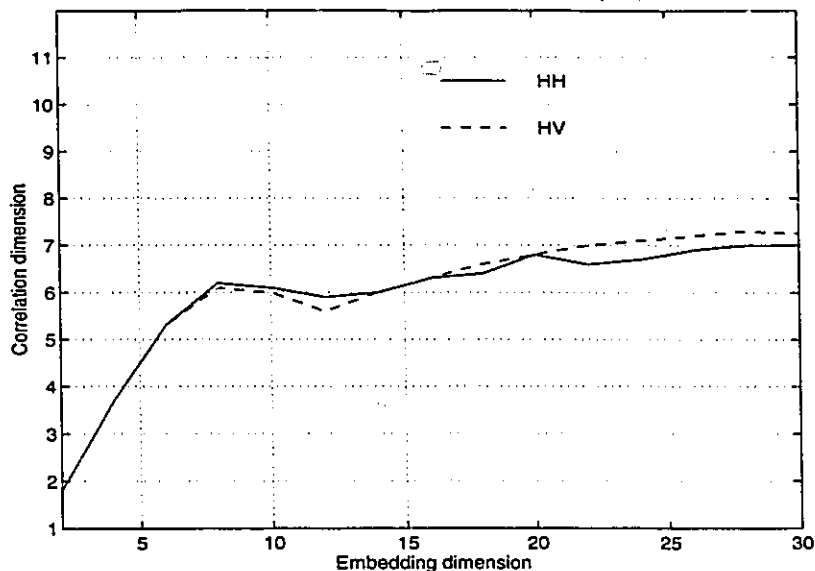


Figure 6.8: Comparison of different polarization on the computation of correlation dimension. Each data set used for the computations contains 30000 samples.

The next step in the correlation dimension analysis is to study the effects of geographic location of the observations and sea state at the time of observations. Figure 6.9 shows three estimations of correlation dimension versus the embedding dimension, with each curve corresponding to a sea clutter data set collected at a specific location. All three time series were collected for roughly the same wave height. The results show that all three curves first reach their saturated region at the embedding dimension of about 20. Defining the average value in the saturated region to be an estimate of the correlation dimension for the specific data set, we find that the correlation dimensions shown in Figure 6.9 are 8.21, 7.68 and 7.45, respectively. For all practical purposes, we can say that they are sufficiently close to each other. Figure 6.10 shows the relationship between the estimated correlation dimension and the sea state, with the sea state being roughly classified by wave height. The results shown in Figure 6.10 (a) correspond to the data collected at Cape Bonavista, Newfoundland (1989), listed in Table 1. The solid curve in the figure shows the mean value based on 38 estimates of the correlation dimension for each sea state, and the error bar shows the corresponding error, more precisely, twice the standard deviation, in the estimation of the mean. The results in Figure 6.10 (b) correspond to the data collected at Dartmouth, Nova Scotia (1993), and listed in Table 2. In this case, the solid curve is the

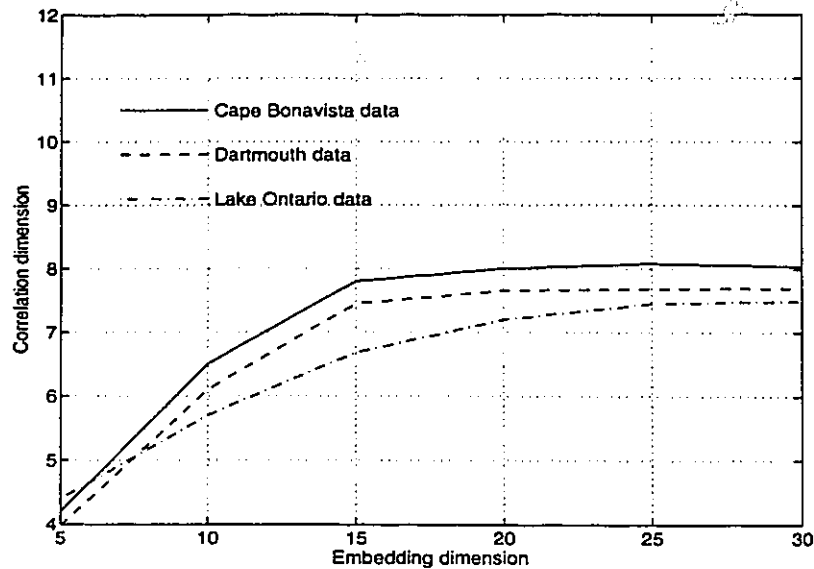


Figure 6.9: The effects of geographic locations on the computation of correlation dimension. Each curve corresponds to a sea clutter data set of 30000 samples, collected at a specific location. All three data sets were collected for roughly the same wave height.

mean value of the correlation dimension for each sea state, based on 14 estimates; error bars are as previously defined. We can draw two conclusions from Figure 6.10. First, it seems that the correlation dimension of sea clutter does not vary appreciably with the sea state. Second, comparing Figures 6.10 (a) and (b), we find that the correlation dimension of sea clutter does not vary with radar site.

The results of the correlation dimension analysis for sea clutter presented in this subsection tell us that sea clutter has a finite and fractal correlation dimension somewhere between 7 and 9. Moreover, the correlation dimension of sea clutter is essentially invariant with respect to the choice of radar channel (in-phase, quadrature, amplitude, polarization), radar site, and sea state.

6.3 False Nearest Neighbors Analysis

We have also applied the method of false nearest neighbors described in Section 4.3.3 to measure the minimum embedding dimension for the reconstruction of sea clutter dynamics. As our sea clutter data were recorded from field experiments by the IPIX radar system

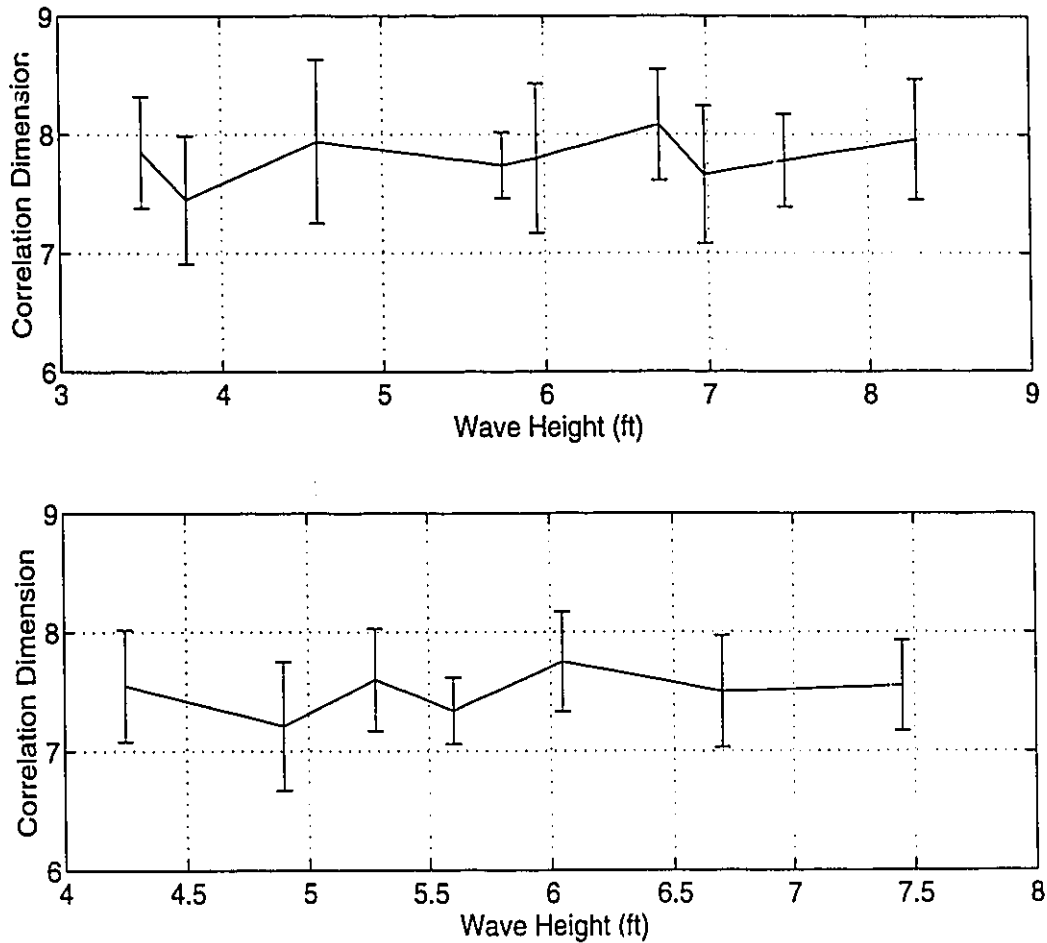


Figure 6.10: The relationship between the estimated correlation dimension and sea state, with the sea state being classified roughly by wave height. (a) The results correspond to the database collected at Cape Bonavista, Newfoundland (1989). The solid curve shows the mean value of 38 estimates computed for each sea state, and each error bar shows the standard error (twice the standard deviation) in the estimation of the mean; (b) the results correspond to the database collected at Dartmouth, Nova Scotia (1993); once again, the solid curve shows the mean value of 14 estimates computed for each sea state, and each error bar has the same meaning as that in (a).

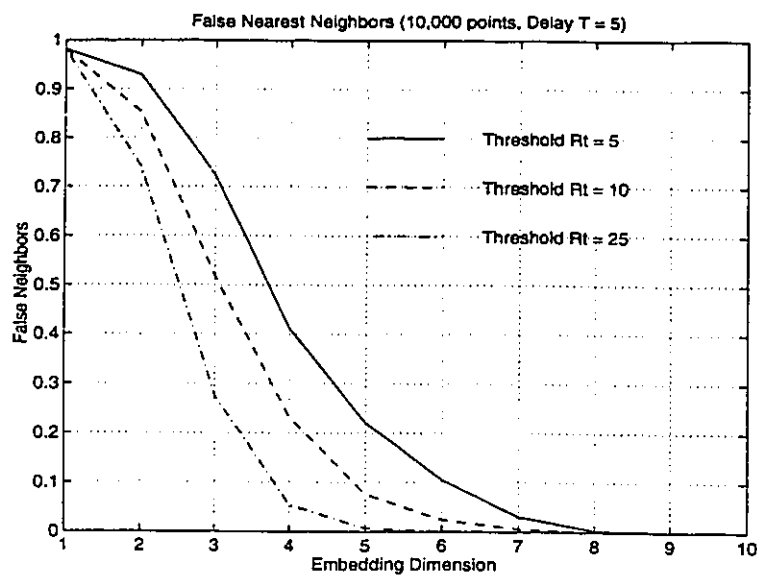


Figure 6.11: False nearest neighbor analysis of a data set collected at the Cape Bonavista with wave height of 4.59 ft..

described in Chapter 5, the data are inevitably contaminated by some form of noise. The quantization noise of the digitized data is another factor that affects the quality of our sea clutter data. To get a more reliable and conclusive result on the false nearest neighbors analysis for sea clutter, both criteria of Section 4.3.3 are used alternatively in our algorithm.

The results, plots of the probability of false nearest neighbors versus embedding dimension, are shown in figure 6.11 to 6.16; each figure includes the use of three thresholds for the criterion of (4.26) and the standard deviation of the sea clutter data for the criterion of (4.27), on an alternate basis. As the embedding dimension increases, a false nearest neighbor is declared when either of those two criteria is designated as false. These results pertain to sea clutter data collected at two different sites and various sea states. The figures show that the probability of false nearest neighbors is essentially zero after the embedding dimension $D_E = 9$. In other words, the embedding dimension chosen for the reconstruction of sea clutter dynamics has to be at least 9. The results obtained by the method of false nearest neighbors are thus seen to be consistent with the Grassberger-Procaccia algorithm for the correlation dimension of sea clutter.

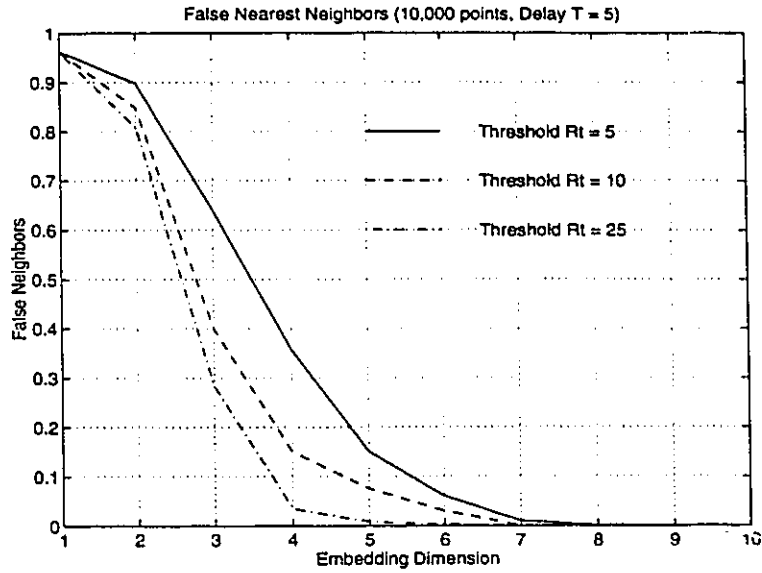


Figure 6.12: False nearest neighbor analysis of a data set collected at the Cape Bonavista with wave height of 6.70 ft..

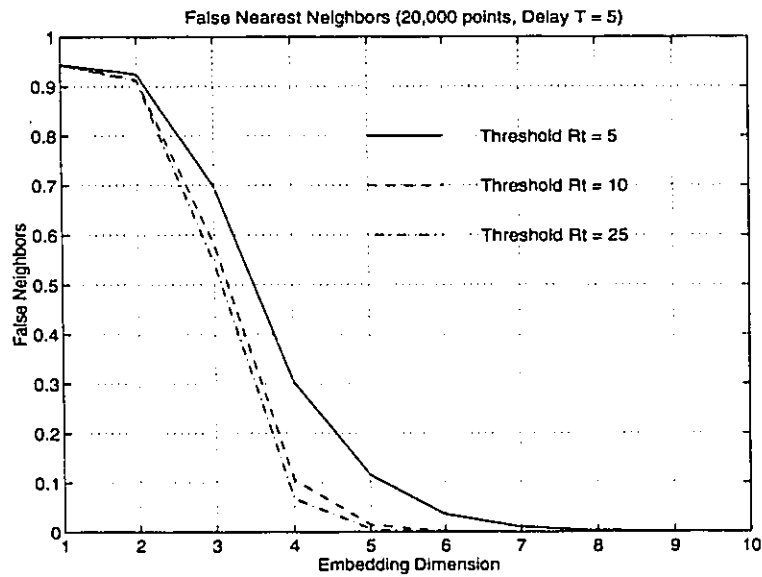


Figure 6.13: False nearest neighbor analysis of a data set collected at the Cape Bonavista with wave height of 8.30 ft..

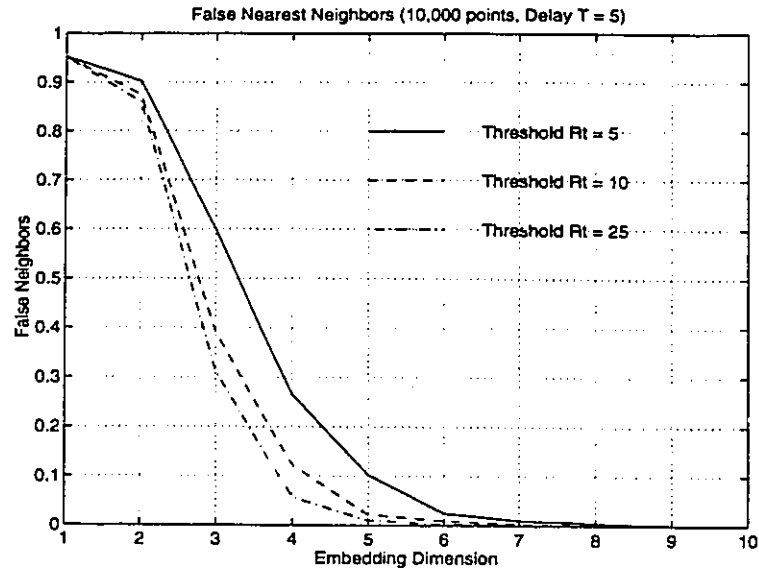


Figure 6.14: False nearest neighbor analysis of a data set collected at the Dartmouth with wave height of 4.90 *ft.*.

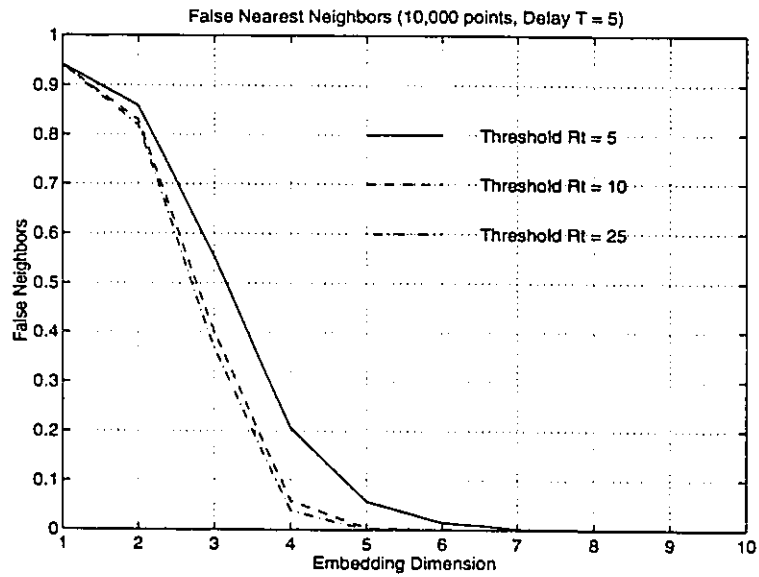


Figure 6.15: False nearest neighbor analysis of a data set collected at the Dartmouth with wave height of 6.05 *ft.*.

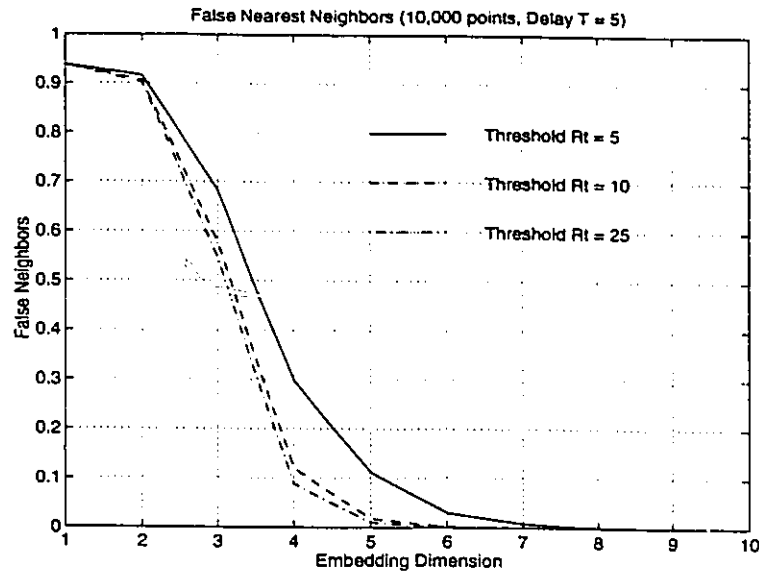


Figure 6.16: False nearest neighbor analysis of a data set collected at the Dartmouth with wave height of 7.45 ft..

6.4 Lyapunov Exponents

6.4.1 Largest Lyapunov Exponent

To calculate the largest Lyapunov exponent of sea clutter, the algorithm developed by Wolf *et al.* [101] was first employed to sea clutter data. For this calculation, a mutual time delay $\tau = 5$ and an embedding dimension of $D_E = 10$ were chosen, for which D_E is slightly larger than the correlation dimension as shown in Figures 6.3 to 6.4. Figure 6.17 shows the estimations of the largest Lyapunov exponent of the data sets used for the computation of Figure 6.7. The solid curve in the figure refers to the I-channel data, the dashed curve refers to the Q-channel data, and the dash-dotted curve refers to the amplitude of the same data set. Defining the average value of the saturated region to be an estimate of the largest Lyapunov exponent of the data set, we find that the values shown in Figure 6.17 are almost identical in the range of 0.035 to 0.042, confirming that a single component of sea clutter is adequate for estimation of the Lyapunov exponent. This is in perfect accord with Takens' embedding theorem.

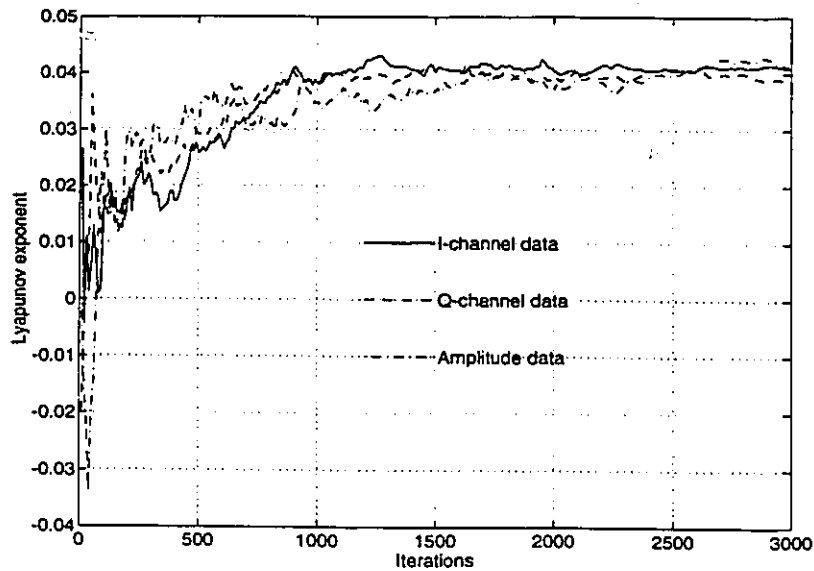


Figure 6.17: Lyapunov exponents of the same sea clutter data used for the computations of Figure 6.7. The solid curve refers to I-channel data, the dashed curve refers to Q-channel data, and the dash-dotted curve refers to amplitude data. Each data set contains 30000 samples.

The effects of transmitter and receiver polarizations on the extraction of the Lyapunov exponent were also checked. Figure 6.18 shows the estimations of the largest Lyapunov exponent as a function of iteration time using the data sets that produced Figure 6.8. The solid curve in the figure refers the HH polarization data; and the solid one refers the HV polarization data. The results show that the largest Lyapunov exponent for both polarizations lies in the range of 0.03 to 0.04. Here again, we see that they are close to each other.

As with the results presented in last section, the effects of geographic location of the observations and the sea state at the time of observation on the Lyapunov exponent were studied. Figure 6.24 shows three estimates of the largest Lyapunov exponent as a function of iteration time on the data sets used for the computation of Figure 6.9. All three curves in their saturated region lie in the range of 0.03 to 0.04. They are close to each other, despite the fact that the data sets used for this computation were collected at different locations and at different times.

Figures 6.20(a) and (b) show the relationship between the estimated Lyapunov

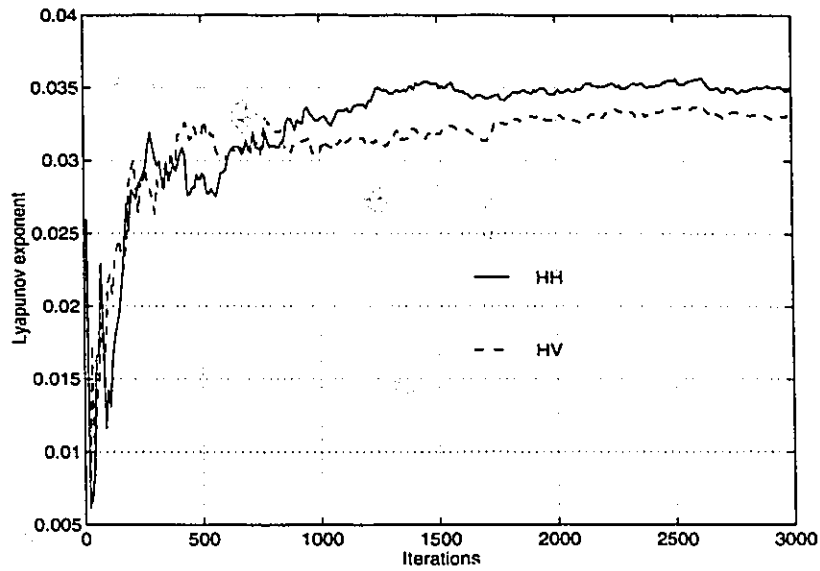


Figure 6.18: Lyapunov exponent of different polarization data used for the computation of Figure 6.8

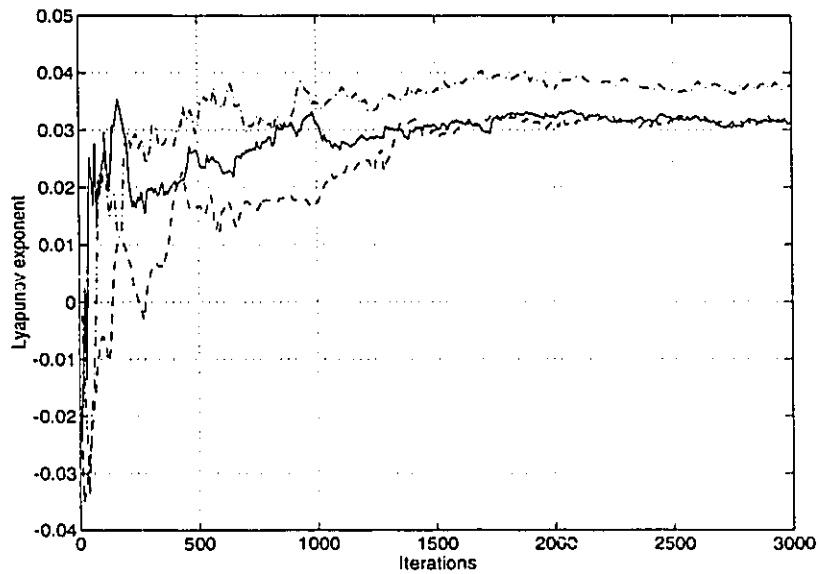


Figure 6.19: Lyapunov exponents of the same sea clutter data used for the computations of Figure 6.9. Each curve refers to a sea clutter data set collected at a specific location. All three data sets were collected for roughly the same wave height.

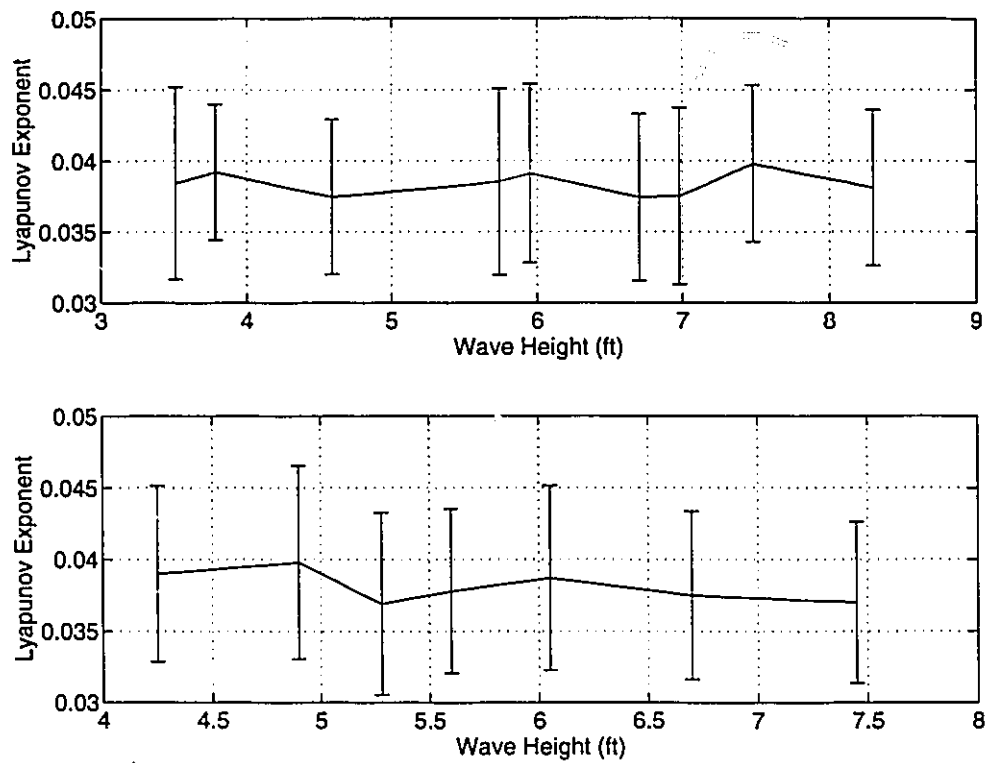


Figure 6.20: The relationship between the estimated largest Lyapunov exponent and sea state, with the sea state being classified roughly by wave height. Traces are as defined in Figure 6.10

Sea State	3.78	4.54	5.95	8.30
λ_1	0.0358 ± 0.006	0.0345 ± 0.004	0.0368 ± 0.005	0.0371 ± 0.005
λ_2	0.000 ± 0.000	0.000 ± 0.000	0.000 ± 0.000	0.000 ± 0.000
λ_3	-0.0071 ± 0.005	-0.009 ± 0.005	-0.091 ± 0.004	-0.0082 ± 0.005
λ_4	-0.152 ± 0.046	-0.164 ± 0.050	-0.158 ± 0.050	-0.172 ± 0.068
λ_5	-0.226 ± 0.084	-0.231 ± 0.080	-0.252 ± 0.078	-0.230 ± 0.081
λ_6	-0.301 ± 0.098	-0.321 ± 0.100	-0.312 ± 0.090	-0.309 ± 0.098
λ_7	-0.401 ± 0.112	-0.424 ± 0.104	-0.422 ± 0.106	-0.429 ± 0.099
λ_8	-0.589 ± 0.120	-0.601 ± 0.100	-0.610 ± 0.110	-0.592 ± 0.100
λ_9	-0.892 ± 0.214	-0.901 ± 0.120	-0.920 ± 0.102	-0.895 ± 0.120
λ_{10}	-1.542 ± 0.342	-1.620 ± 0.380	-1.583 ± 0.421	-1.518 ± 0.401

Table 6.1: Lyapunov spectrum for different sea states, where the sea state is classified by wave height

exponent and the sea state. All the data sets used for the computation of this figure are exactly the same as those for the computation of Figure 6.10. Again, the largest Lyapunov exponent of sea clutter is positive and almost invariant with respect to geographic locations and sea state, providing further confirmation for the chaotic nature of sea clutter.

6.4.2 Lyapunov Spectrum

To calculate the complete Lyapunov spectrum of the embedded system of sea clutter, we applied two procedures based on the ideas described in Section 4.4.3. For this computation, four kinds of specific sea states, varied from near calm to rough, were chosen. In each sea state, classified by wave height, we chose 10 data sets to the computation. The embedding dimension was chosen to be 10.

First, the results of the complete Lyapunov spectrum, obtained by the method based on a local linear approach to build the Jacobian matrix, are summarized in Table 6.1. In this table, the first row lists the sea states classified by wave height; the subsequent rows list the corresponding Lyapunov exponents. Consider, for example, $\lambda_1 = 0.0358 \pm 0.006$ for sea state 1. The first number is the mean value of 10 estimates computed for that sea state, and the second number is the standard deviation in the estimation of the mean. The resulting largest Lyapunov exponent by this method is in close agreement with the Wolf's algorithm.

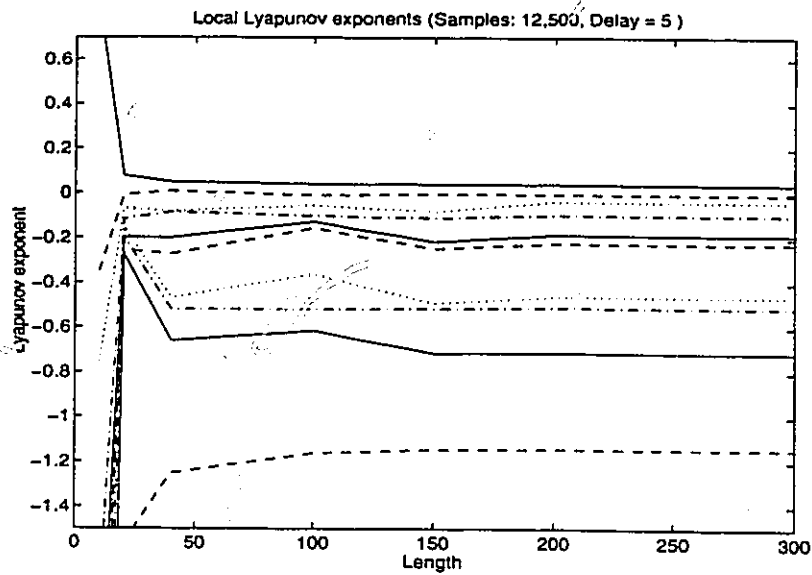


Figure 6.21: Lyapunov spectrum computed by the method based on 3rd-order polynomial fit to build the Jacobian

To further substantiate the validity of the results on the largest Lyapunov exponent of sea clutter presented previously, we have used the method described in [13], using a higher (*i.e.*, third)-order polynomial fit to build the Jacobian matrix for a data picked up from data set B218 of Table 5.1. The results of the latter computation are presented in Figure 6.21. The evolving values of the Lyapunov exponents are summarized in Table 6.2; on the whole, these results further substantiate the final values summarized in Table 6.1.

The results shown above indicate that the first Lyapunov exponent of sea clutter is

Length	10	20	40	100	150	200	300
λ_1	0.8129	0.0772	0.0495	0.0408	0.0408	0.0406	0.0405
λ_2	-0.3480	-0.0084	0.0086	-0.0092	-0.0037	-0.0037	-0.0037
λ_3	-0.7557	-0.0689	-0.0824	-0.0548	-0.0827	-0.0366	-0.0386
λ_4	-1.8895	-0.1171	-0.0854	-0.1021	-0.1106	-0.0971	-0.0981
λ_5	-2.5475	-0.1999	-0.2000	-0.1279	-0.2168	-0.1854	-0.1863
λ_6	-3.1362	-0.2566	-0.2720	-0.1571	-0.2479	-0.2235	-0.2230
λ_7	-3.1845	-0.1534	-0.4681	-0.3626	-0.4917	-0.4615	-0.4721
λ_8	-4.2345	-0.1998	-0.5190	-0.5188	-0.5165	-0.5104	-0.5108
λ_9	-6.5482	-0.2731	-0.6590	-0.6151	-0.7152	-0.7124	-0.7134
λ_{10}	-13.2502	-1.5705	-1.2511	-1.1616	-1.1462	-1.1437	-1.1442

Table 6.2: The evolving values of the Lyapunov spectrum shown in Figure 6.21

	$D_E = 5$	$D_E = 6$	$D_E = 7$	$D_E = 8$	$D_E = 9$	$D_E = 10$
λ_1	0.0432	0.0412	0.0408	0.0409	0.0410	0.0405
λ_2	0.0054	0.0028	0.0034	-0.0008	-0.0019	-0.0037
λ_3	-0.0151	-0.0436	-0.0386	-0.0401	-0.0396	-0.0386
λ_4	-0.5837	-0.2571	-0.2487	-0.2375	-0.1874	-0.0981
λ_5	-1.0547	-0.7267	-0.6802	-0.5607	-0.2833	-0.1863
λ_6		-1.2043	-0.8804	-0.7823	-0.5140	-0.2230
λ_7			-1.2183	-0.8973	-0.7063	-0.4721
λ_8				-1.1742	-0.8854	-0.5108
λ_9					-1.1632	-0.7134
λ_{10}						-1.1442

Table 6.3: The Lyapunov exponents of sea clutter for different embedding dimensions.

indeed positive, and the Lyapunov spectrum of the embedded system does not appear to change with sea state.

Two interesting observations can be made from the above computations:

- The second Lyapunov exponent of sea clutter appears to be zero. This may imply that sea clutter is governed by a coupled system of nonlinear differential equations [3, 6].
- The sum of the Lyapunov exponents is negative, implying that the sea clutter is describable as a flow.

To further substantiate this observation, we have embedded the same sea clutter data used for the computation of Figure 6.21 in a system with varying dimension of 5, 6, 7, 8, 9, 10, respectively, and used the method described in [6] with a third-order polynomial fit to build the Jacobian matrix. The corresponding Lyapunov spectrum of the embedded systems are shown in Table 6.3. The results presented here provide further evidence for the above mentioned observations.

With computational results of the Lyapunov spectrum of sea clutter at hand, we go on to compute the Lyapunov dimension, or Kaplan-Yorke dimension, which can be defined as

$$D_L = K - \frac{\sum_{i=1}^K \lambda_i}{\lambda_{K+1}} \quad (6.3)$$

	$D_E = 5$	$D_E = 6$	$D_E = 7$	$D_E = 8$	$D_E = 9$	$D_E = 10$
Lyapunov dimension	3.058	3.002	3.023	3.000	2.987	2.953

Table 6.4: The Lyapunov dimension of sea clutter verse the embedding dimension

where $\lambda_1 > \lambda_2 > \dots$ are the Lyapunov exponents of the system under study, and K is determined such that $\sum_{i=1}^K \lambda_i > 0$ and $\sum_{i=1}^{K+1} \lambda_i < 0$. The Lyapunov dimension was originally defined by Kaplan and Yorke [53], as a lower bound on the fractal dimension.

Table 6.4 shows computational results of the Lyapunov dimension for sea clutter corresponding to the results on Lyapunov spectrum reported in Table 6.3. As the embedding dimension is varied from 5 to 10, the Lyapunov dimension remains around 3. This result is smaller than the correlation dimension reported in Section 6.2, which is in accord with Kaplan and Yorke [53].

6.5 Kolmogorov Entropy

The Kolmogorov entropy of an attractor can be considered as a measure for the rate of information loss along the attractor or as a measure for the degree of predictability of points along the attractor, given an arbitrary initial point. In general, a positive, finite entropy is considered as conclusive proof that the time series and its underlying dynamic phenomenon are chaotic [86]. A zero entropy represents a constant or a regular, cyclic phenomenon that can be represented in the state space by a fixed point, a periodic attractor, or a multiperiodic attractor. An infinite entropy refers to a stochastic phenomenon.

Given a nonlinear dynamical system, we consider two trajectories on its attractor, which are initially very close. The distance between two nearby points on different trajectories is less than a specified maximum value l_0 . According to [34, 86], the separation of the nearby points is assumed to be exponential, and the time interval ΔT required for two initially nearby points to separate by a distance larger than l_0 will be exponentially distributed according to

$$C(\Delta T) \sim e^{-K\Delta T} \quad (6.4)$$

where K is the Kolmogorov entropy.

Typically, an experimental time series is a record of continuous observation of a dynamical system in time interval Δt , corresponding to a sampling rate of $\frac{1}{\Delta t}$. In this case, ΔT in (6.4) is replaced by $b\Delta t$, where $b = 1, 2, \dots$. The distribution function of b can be defined as

$$C(b) = e^{-Kb\Delta t} \quad (6.5)$$

On the basis of this definition, a maximum likelihood estimate of the Kolmogorov entropy is derived in [86] as follows:

$$K_{ML} = -\frac{1}{\Delta t} \ln\left(1 - \frac{1}{\bar{b}}\right) \quad (6.6)$$

with

$$\bar{b} = \frac{1}{M} \sum_{i=1}^M b_i \quad (6.7)$$

which is the average value of the b 's in the sample (b_1, b_2, \dots) , with an effective sample size of M . It has been shown that, as $M \rightarrow \infty$ and $l_0 \rightarrow 0$, $K_{ML} \rightarrow K$ [86].

In practical use of the Kolmogorov entropy estimate of (6.7), l_0 is recommended to be the average absolute deviation of the time series under study, and M is the number of full cycles after the first passage through the average of the signal [85]. Obviously, the longer the time series, the larger the value of M , thus, a more accurate estimate of the Kolmogorov entropy is expected.

The Reactor Research Foundation of the Netherland [85] has developed a software package for the estimation of the Kolmogorov entropy, based on the method described above. The software is applied to our sea clutter data. A result is shown in Figure 6.22. The cutoff length in the figure refers to l_0 , and the normalized embedding length refers to the average cycle time. As the cutoff length approaches the value 2, the estimated Kolmogorov entropy converges to the range between 0.04 and 0.05, which, in light of the computational results on the Lyapunov spectrum reported in last section, appears to indicate that the Kolmogorov entropy of sea clutter is not significantly different from its own largest Lyapunov exponent. This gives further evidence supporting our observation that the second Lyapunov exponent is likely to be zero.

As discussed in [86, 87], 50 to 200 numbers of points in each full cycle are needed to obtain a reliable estimation of the Kolmogorov entropy. In other words, an over sampled time series is recommended [85]. In the computation of Kolmogorov entropy for sea clutter,

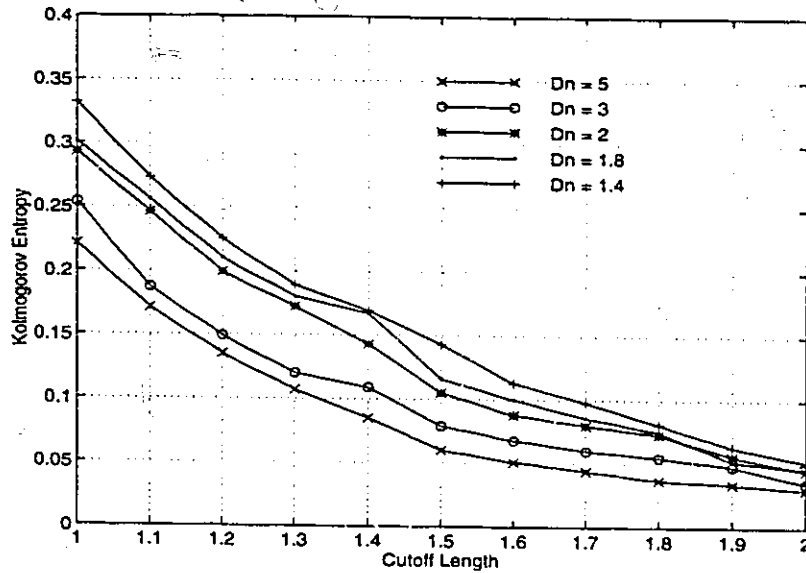


Figure 6.22: Kolmogorov entropy of sea clutter. The corresponding data is picked up from data set B218 of table 5.1. It has length of 35000 samples.

an average number of points of about 10 per cycle is found, which is smaller than the recommended value of 50. This suggests that a further modification of the radar system used to collect the sea clutter data is needed such that the sampling rate is higher enough.

6.6 Estimation of Sea Clutter Dynamics

The next issue we wish to consider is application of the chaos-based detection method described in Chapter 3 to small target detection in sea clutter. To do this, however, we first need to build a model to reconstruct the underlying dynamics from sea clutter time series.

Dynamic reconstruction is an *inverse problem*. To be specific, suppose that we are given a time series $\{y(i) : i = 1, \dots, N\}$. A D_E -dimensional phase portrait can be reconstructed by using Takens' embedding theorem. If it is assumed that the time series is an observation of a chaotic system, then there must exist a nonlinear function F , such that

$$y(i + D_E\tau) = F(y(i), y(i + \tau), \dots, y(i + (D_E - 1)\tau)) \quad (6.8)$$

where τ is the time delay. This is actually a nonlinear prediction problem. To reconstruct the unknown dynamics of the system, two operations are performed:

1. Choose a nonlinear function $\hat{F} : R^N \rightarrow R^1$ that approximates the unknown function F .
2. Optimize the approximating function \hat{F} by adjusting its free parameters to minimize the mean-square value of the difference between the sample $y(n+N)$ generated by F and the output $\hat{y}(n+N)$ produced by \hat{F} in response to the input samples $y(n), y(n+1), \dots, y(n+N-1)$.

The solution to the inverse problem described herein is recognized to be actually that of building a *predictive model*. To do so, we propose to *learn* from the sea clutter data in a supervised manner. Certain classes of neural networks are well suited to such a learning process¹.

Neural networks have the inherent ability to learn a mathematical mapping by adaptation in response to examples of the mapping's action. In the prediction problem for a chaotic time series, the network learns input sequences and produces an approximation to an unknown mapping for a deterministic system on its attractor. Indeed, a multilayer perceptron trained with the back-propagation algorithm is known to be a universal approximator [45]. Typically, we are presented with a set of examples $(\mathbf{x}(1), \mathbf{y}(1)), (\mathbf{x}(2), \mathbf{y}(2)), \dots$ of the action of a function $g: R^N \rightarrow R^M$, where $\mathbf{y}(k) = g(\mathbf{x}(k))$. The requirement is to design a neural network that implements an approximation to the function g ; such a network is called a *mapping neural network*. For the situation described in (6.8), $g = F$, and the neural network is described by \hat{F} .

The most commonly used mapping neural networks employ an underlying model and adjust the free parameters of the network to retrieve useful information from the inputs. The back-propagation algorithm used to train a multilayer perceptron [84, 39] is by far the most popular algorithm in use today, even though it has the shortcoming of a slow rate of convergence during the training session. Once the training is done, the multilayer perceptron operates entirely in a feed forward fashion and can therefore be very fast. In this paper, we use a multilayer perceptron trained with the back-propagation algorithm to perform the modeling task.

¹Other techniques reported in the literature for modeling chaotic processes include the use of a family of hidden Markov models [19, 26], functional interpolation [89], and nonlinear Kalman filters [65].

As mentioned previously, we believe that our clutter data have a high clutter-to-noise ratio, measured with respect to the additive front-end receiver noise. The presence of noise has the effect of forcing the actual low-dimensional behavior back into a high dimensional space. To guard against this effect, we choose the embedding dimension to be an integer which is just slightly larger than the correlation dimension. Our experimental results show that the reconstructed system has essentially the same characteristics as the original sea clutter. Herein lies an important feature of the dynamic reconstruction method described in this paper. Specifically, the results presented in Chapter 6 on correlation dimension analysis and false nearest neighbors analysis provide the mathematical basis for determining the number of input samples used to feed the mapping neural network, thereby establishing a highly valuable link between the nonlinear dynamics of the sea clutter process and the design of a multilayer perceptron model.

6.6.1 A Neural Network Model Based on Noncoherent Data

A noncoherent radar receiver provides amplitude (envelope) information only. Thus, given the input samples $y_A(n-1), y_A(n-2), \dots, y_A(n-N)$, the one-step prediction computed by the neural network model may be expressed as follows:

$$\hat{y}_A(n) = \hat{F}(y_A(n-1), y_A(n-2), \dots, y_A(n-N)) \quad (6.9)$$

where the function \hat{F} is the approximation to the unknown function F , and N is the number of data points used to do the prediction.

A multilayer perceptron used for the model reconstruction of sea clutter based on noncoherent data is shown in Figure 6.23. It consists of an input layer of source nodes, two hidden layers, and an output layer with a single neuron providing the prediction. To be sure of achieving a satisfactory performance, the size of the input layer, N , must be chosen properly. In this context, the following two guidelines are noted:

1. The embedding dimension D_E should be an integer larger than the largest possible value assumed by the correlation dimension D_c .
2. The neighbors of each input sample used to reconstruct the phase space are employed in the prediction process, since sea clutter is observed to be locally correlated. This,

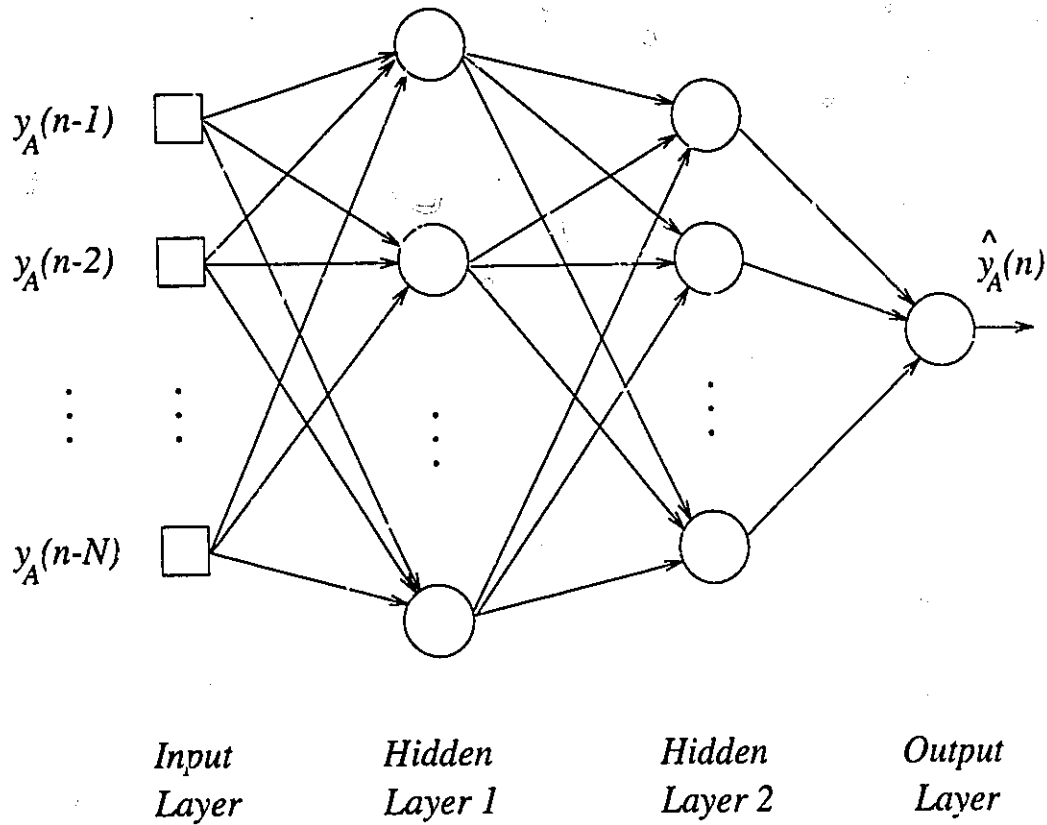


Figure 6.23: A multilayer perceptron structure used for reconstruction of the underlying dynamics of sea clutter, based on noncoherent radar data. It consists of an input layer with 50 source nodes, two hidden layers with 80 and 55 neurons, respectively, and one output neuron.

in turn, suggests using the time delay τ determined in accordance with the notion of mutual information, as described in Section 4.2.

The first guideline is in agreement with the choice of embedding dimension described in [3], in which D_E is 1 to 2.5 times D_c . Accordingly, for the case of a multilayer perceptron model trained on a one-dimensional time series, we may propose a lower bound on N , the size of the model's input layer, as follows:

$$N \geq \tau D_E \quad (6.10)$$

In practice, it is inadvisable to choose a value for N much larger than the bound defined in (6.10), as the effect of additive (receiver) noise would then become a serious issue.

Returning to the task at hand, the correlation dimension for sea clutter lies in the range 7 to 9; we may therefore choose $D_E = 10$. The choice of this value for the embedding dimension is further supported by the results of false nearest neighbors analysis presented in Section 6.3. Next, we note from Figure 6.2 that $\tau = 5$. To satisfy (6.10), we may thus choose $N = 50$ as the size of the input layer of the multilayer perceptron model for sea clutter.

Since our modeling problem is actually one of function approximation, we choose two hidden layers, with the first hidden layer consisting of 80 neurons and the second hidden layer consisting of 55 neurons; this is done to improve the approximation accuracy [17] in performing the modeling task. The output layer consists of a single neuron since the dynamic reconstruction involves a single component, namely, amplitude.

After the multilayer perceptron has been trained with examples consisting of sea clutter alone, its synaptic weights and bias terms are all fixed. However, the resulting model can only be said to have captured the underlying dynamics of sea clutter and is therefore a candidate for the application at hand, if and only if the following two conditions are satisfied:

1. During training the prediction error is reasonably small in a statistical sense; this condition ensures that the model is operating at a global minimum on the error performance surface or a local minimum close enough to a global minimum.
2. The generalization ability of the model has been successfully demonstrated by testing its predictive performance on clutter data not seen before.

For our application, *recursive (iterated) prediction* provides a well-proven method for testing the generalization performance of the neural network model [16]. Specifically, recursive prediction uses new input samples (not in the training set) to start the prediction, and then uses each predicted point to dispense with one of the input samples. This operation is continued until no new input data are needed. At that point, the neural network predictor begins to operate in an autonomous fashion. Ordinarily, this is a very difficult task, since there is no new information supplied to the network, except for the initial starting point. Figures 6.24 to 6.27 show the results of recursive prediction for four different

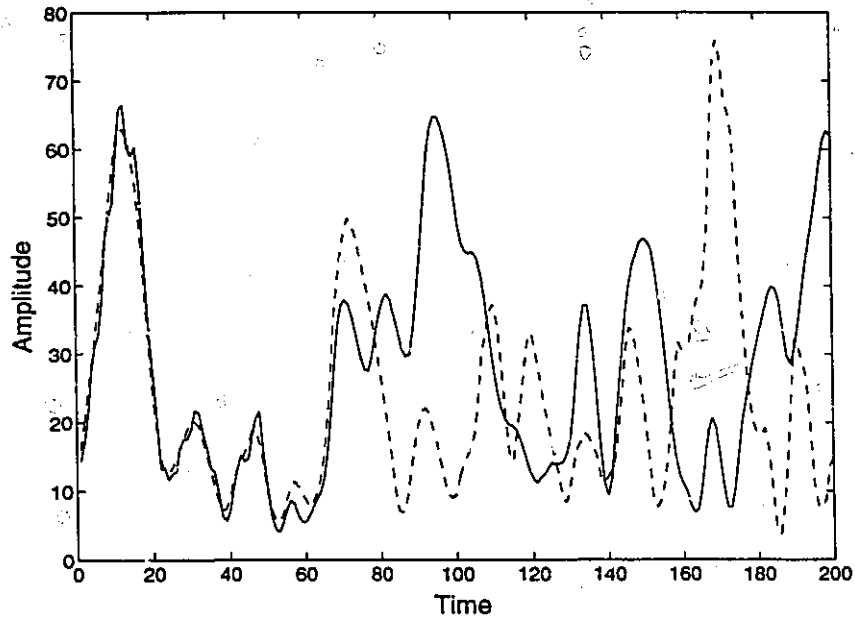


Figure 6.24: Recursive prediction result of the multilayer perceptron model shown in Figure 6.23. The solid curve refers to the original sea clutter waveform. The dashed curve refers to the recursive predicted waveform, for which the first 50 points of the sea clutter set (not shown in the figure) are used as the initial starting point.

sets of noncoherent radar data. The solid lines are the original sea clutter waveform. Using the first 50 points of the test set (not shown in the figure) as the initial starting point, we get the predicted waveform shown as the dashed curve in these figures. For about the first 50 points shown in Figure 6.24, both curves match fairly well, and thereafter they diverge. This result confirms that sea clutter produced by a noncoherent radar is indeed locally predictable, and also demonstrates sensitivity to initial conditions, both of which are characteristics attributable to a chaotic process.

Chaos-based detection is actually a model-based detection procedure. It is implemented by analyzing the one-step prediction error as described in Chapter 3. To further substantiate the significance of the multilayer perceptron model trained by the use of actual sea clutter (amplitude) data, we evaluate the statistics of the prediction error by the use of Pearson's rule [50].

Let μ_1 be the mean of a random process, and μ_2, μ_3 , and μ_4 be the second, third, and fourth moments of the process about the mean μ_1 , respectively. Define the skewness

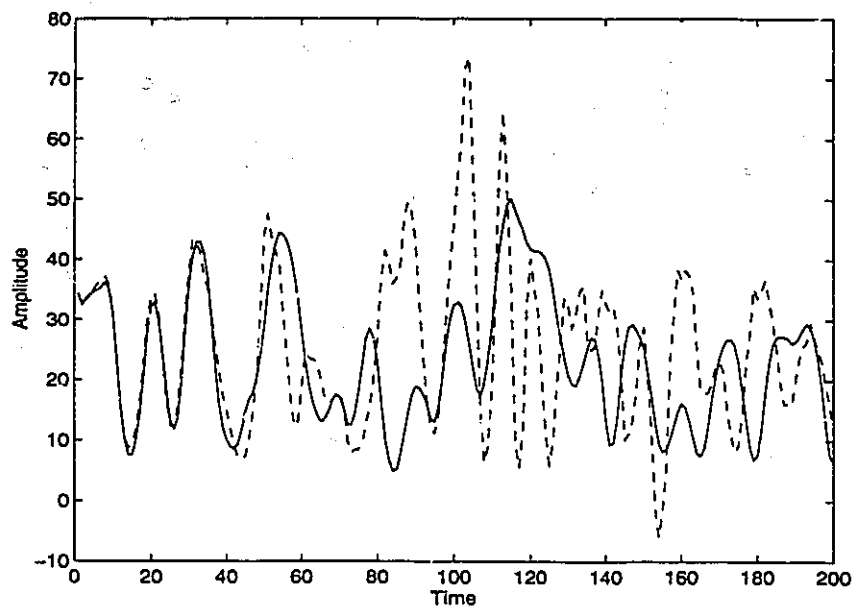


Figure 6.25: Recursive prediction result of the multilayer perceptron model shown in Figure 6.23. Traces are as defined in Figure 6.24.

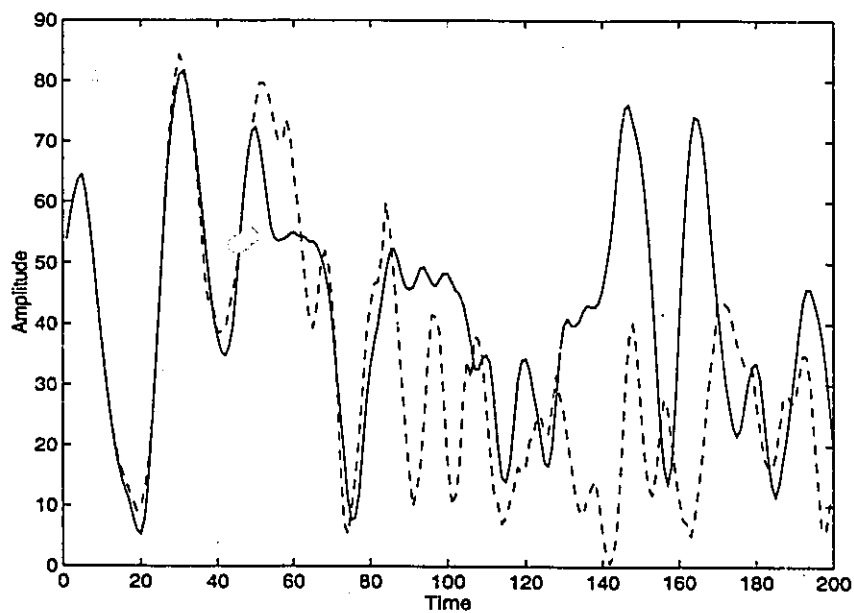


Figure 6.26: Recursive prediction result of the multilayer perceptron model shown in Figure 6.23. Traces are as defined in Figure 6.24.

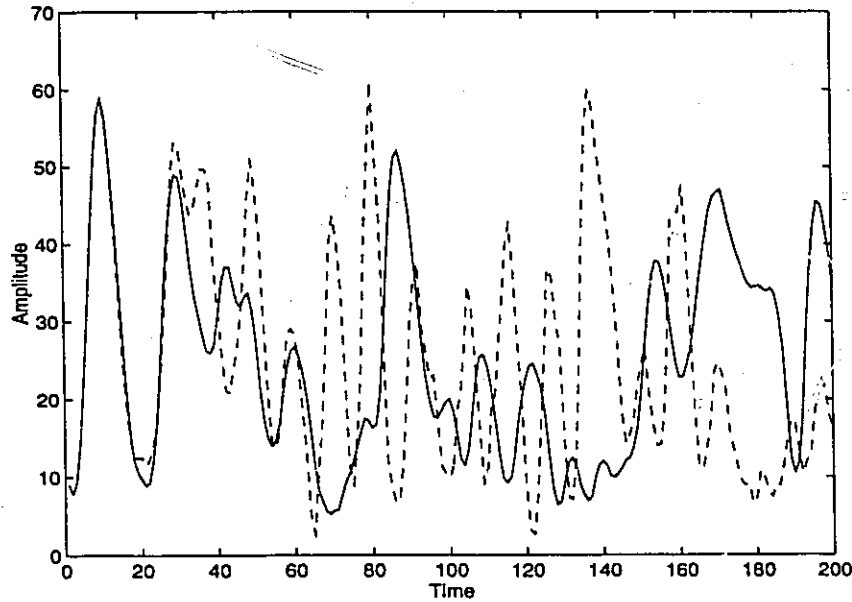


Figure 6.27: Recursive prediction result of the multilayer perceptron model shown in Figure 6.23. Traces are as defined in Figure 6.24.

β_1 and the kurtosis β_2 of the random process as follows:

$$\beta_1 = \frac{\mu_3^2}{\mu_2^3} \quad (6.11)$$

$$\beta_2 = \frac{\mu_4}{\mu_2^2} \quad (6.12)$$

Then the distribution types of the random process are classified by the use of these parameters. Applying (6.11) and (6.12) to the one-step prediction error of the multilayer perceptron model of sea clutter, we find that typically $\mu_1 = 0.01$, $\beta_1 = 0.043$, and $\beta_2 = 3.42$. Thus according to Pearson's rule, the underlying distribution is, for all practical purposes, Gaussian for which $\beta_1 = 0$ and $\beta_2 = 3$.

The empirical formula (6.10) provides a lower bound on the size of the input layer of the multilayer perceptron model. It is inadvisable to use a size N much larger than this bound, as the effect of receiver noise would then become more pronounced. But what if N is smaller? Figure 6.28 shows the result of a recursive prediction performed by a multilayer perceptron model with $N = 45$, which is smaller than the bound defined in (6.10) for $D_E = 10$ and $\tau = 5$. Except for this difference, the model has two hidden layers with 80 neurons in the first one and 55 neurons in the second one, and a single output neuron, as

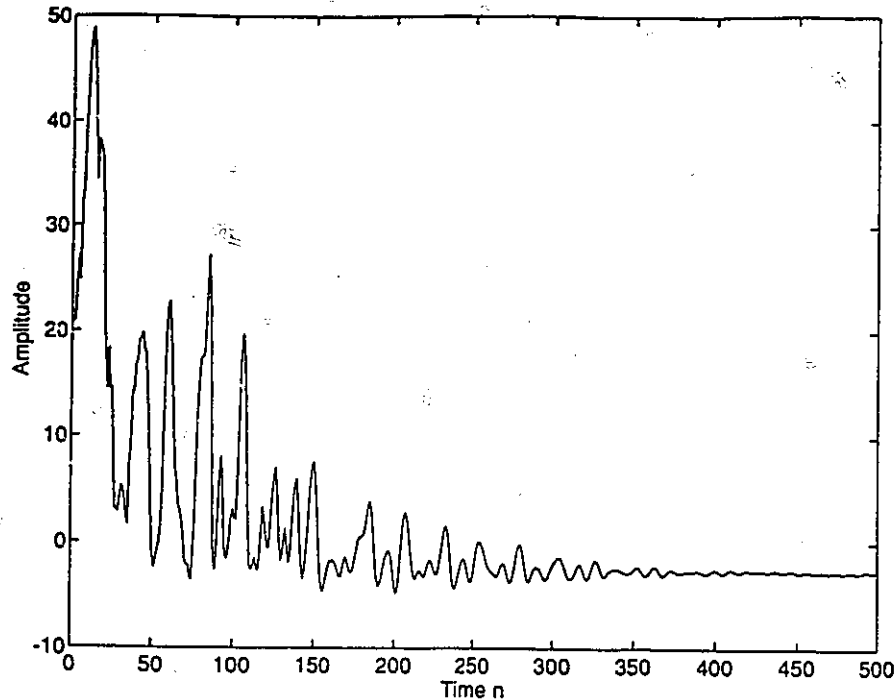


Figure 6.28: This figure shows the sensitivity of the recursive prediction process to a change in the neural network design. The network used for this experiment consists of an input layer with 45 source nodes, two hidden layers with 80 and 55 neurons, respectively, and one output neuron.

before. Moreover, the model is trained with the same data set used to obtain the result shown in Figure 6.24, and the recursive prediction process is used to test the model after completing the training session. There is a dramatic difference between the results shown in Figures 6.24 and 6.28. In particular, when the size of the input layer of the multilayer perceptron model is not large enough, the model behaves as a fixed point attractor and therefore fails to capture the underlying dynamics of the sea clutter. Clearly, the choice of a design exemplified by Figure 6.28 is unacceptable.

To emphasize the need for a nonlinear predictive model, Figure 6.29 shows the recursive prediction results using an autoregressive (AR) model and the multilayer perceptron model for reconstruction of the underlying dynamics of sea clutter. Both models use 50 delay taps. Clearly, the AR model, which is linear, completely fails to capture the underlying dynamics of sea clutter.

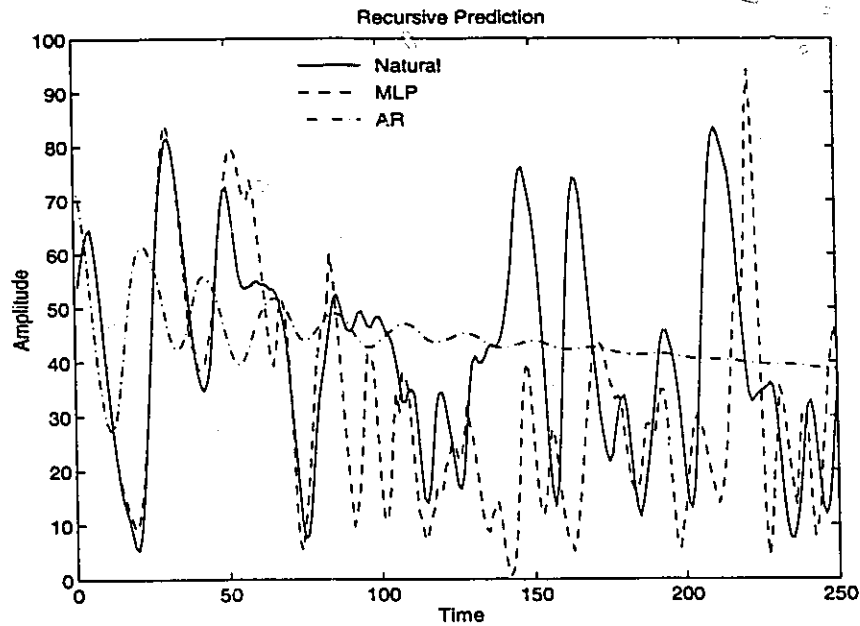


Figure 6.29: A comparison of both the AR model and the multilayer perceptron (MLP) model for reconstruction of the underlying dynamics of sea clutter.

6.6.2 Temporal Extent of Local Predictability

Farmer and Sidorowich [24] have studied the local prediction problem in chaotic systems. Specifically, two distinct regimes are identified. In the first regime the prediction is limited by noise. This occurs when the typical spacing between data points is approximately

$$\epsilon \simeq N^{-\frac{1}{D}} < S^{-1} \quad (6.13)$$

where N is the number of data points used to do the prediction, D is the attractor dimension, and S is the signal-to-noise ratio. The second regime occurs when $\epsilon > S^{-1}$, and the prediction is limited by the number of data points. In the latter case, we find that the *local prediction time*, denoted by T , is inversely proportional to the largest Lyapunov exponent λ_1 , as shown by ²

$$T \simeq \frac{1}{\lambda_1} \ln\left(\frac{\sigma_T}{\sigma_0}\right) \quad (6.14)$$

²Equation (6.14) is a special case of a formula proposed by Farmer and Sidorowich [24]; see also [16].

where the parameter σ_T is the normalized standard deviation of the prediction error at time T of the recursive prediction process; it is defined by

$$\sigma_T^2 = \frac{\frac{1}{T} \sum_{i=0}^{T-1} (\hat{y}(n+i) - y(n+i))^2}{\frac{1}{N} \sum_{n=1}^N (y(n) - m_y)^2} \quad (6.15)$$

where N is the total number of the time series $y(n)$, and m_y is the mean value of $y(n)$ over the observation period. The parameter σ_0 is the normalized standard deviation of the prediction error at the start of the recursive prediction process; it is defined by (6.15) with $T = 0$. Equation (6.14) follows from the fact that in a chaotic system under iteration, errors grow exponentially at a rate equal to the largest Lyapunov exponent of the system. It is also of interest to note that σ_0 is a measure of the precision with which the initial state of the attractor is determined by the predictive model.

For the experimental study presented in this thesis, the second regime applies since the clutter-to-noise ratio is known to be high. For $\sigma_T/\sigma_0 = 10$ and $\lambda_1 = 0.04$, the use of (6.14) yields $T \simeq 58$. This result is in close agreement with the "sample" value $T \simeq 50$ noted from Figures 6.24 to 6.27.

6.6.3 A Neural Network Model Based on Coherent Data

A coherent radar receiver provides two independent observations about the ocean surface for exactly the same sea state. The observations are represented by the I-channel component and Q-channel component of the received signal. To test the local predictability of the complex-valued sea clutter produced by such a radar, we may use the multilayer perceptron model depicted in Figure 6.30. Experimentation with this model shows that we need an input layer whose size is slightly larger than that of the model described in Figure 6.23; a size of 60 source nodes was found to be satisfactory for our application. Correspondingly, the output layer of the model in Figure 6.30 consists of two neurons, which provide predictions of the in-phase and quadrature components of the sea clutter. As for the case of noncoherent radar data, the model has two hidden layers with 80 neurons in the first and 55 neurons in the second.

The model of Figure 6.30 is trained with the clutter data using the back-propagation algorithm, and after the learning session is completed, the synaptic weights and bias terms

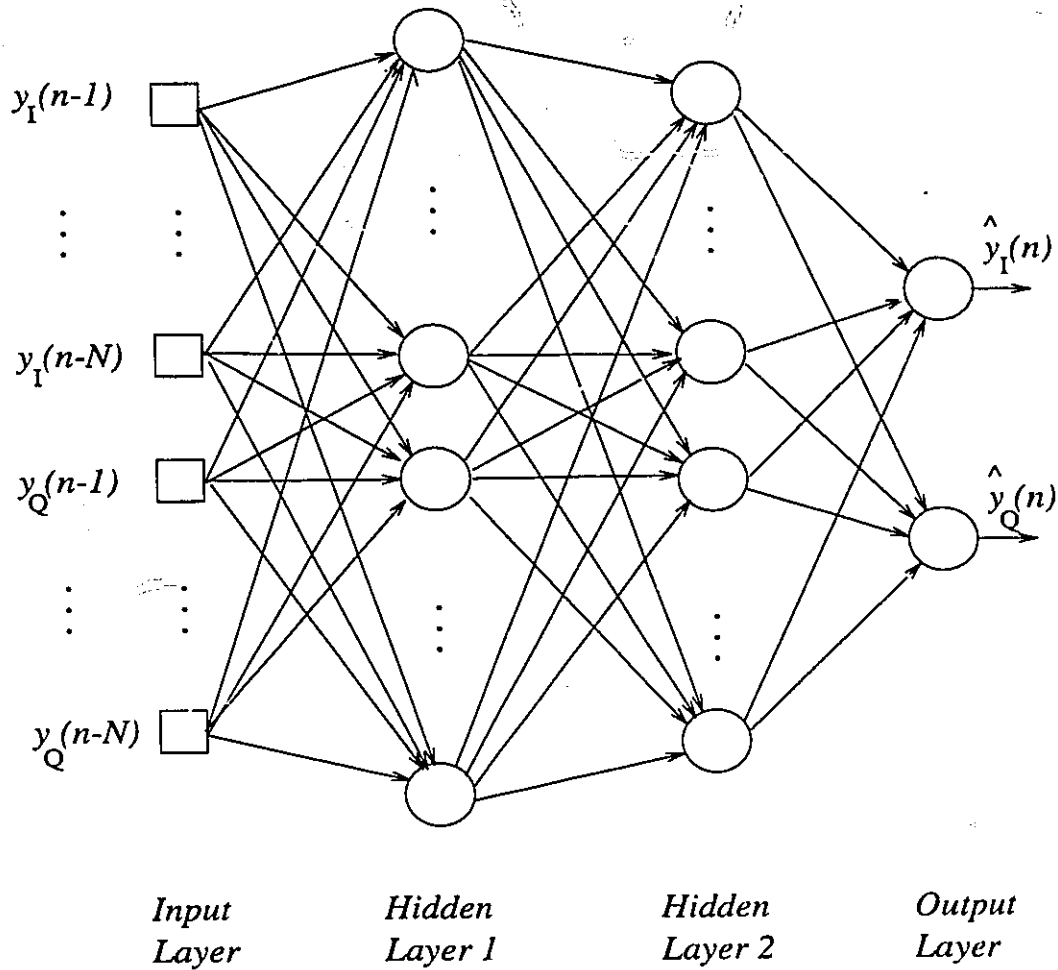


Figure 6.30: A multilayer perceptron structure used for reconstruction of the underlying dynamics of sea clutter, based on coherent radar data. It consists of an input layer with 60 source nodes, (with 30 source nodes used for the in-phase component and 30 other source nodes used for the quadrature component), two hidden layers with 80 and 55 neurons, respectively, and two output neurons to provide one-step predictions of the in-phase and quadrature components.

of the network are all fixed. The results of the recursive prediction performed by the network are shown by the dashed curves of Figures 6.31 to 6.34, where (for convenience of presentation) we have omitted the sequences of 30 complex-valued samples used to initialize the network. The solid curves shown in Figures 6.31 to 6.34 are the actual in-phase and quadrature components of the sea clutter. The results shown here confirm that both the in-phase and quadrature components of sea clutter produced by a coherent marine radar are locally predictable.

To calculate the local prediction time of a multi-dimensional chaotic model, the largest Lyapunov exponent λ_1 in (6.14) should be replaced with the *Kolmogorov entropy* [88]. The Kolmogorov entropy is defined as the sum of the positive Lyapunov exponents of the chaotic process under study [20]. From Figures 6.31 to 6.34 we note that the local prediction time T of the two-dimensional predictive model of Figure 6.30 approximately equals 50, which, in light of (6.14), appears to indicate that the Kolmogorov entropy for sea clutter is not significantly different from its own largest Lyapunov exponent. This is in close agreement with the results reported in Section 6.4, which show that there is only one positive Lyapunov exponent in the embedded system of sea clutter.

6.7 New Strange Attractors

Traditionally, strange attractors (*e.g.*, the Lorenz attractor) are defined by a set of coupled nonlinear differential (or difference) equations. The neural network models described in Subsections 6.3.1 and 6.3.3 provide another approach for defining new strange attractors.

Consider first the multilayer perceptron model of Figure 6.23 for sea clutter. The strange attractor, or the “clutter map” to be more precise, embodied in this multilayer perceptron is defined by the following set of coupled nonlinear difference equations:

$$\begin{aligned} x_j(n) &= \phi(\sum_{i=1}^N w_{ji}^{(1)} y_A(n-i) + b_j^{(1)}), \quad j = 1, 2, \dots, J \\ z_k(n) &= \phi(\sum_{j=1}^J w_{kj}^{(2)} x_j(n) + b_k^{(2)}), \quad k = 1, 2, \dots, K \\ \hat{y}_A(n) &= \sum_{k=1}^K w_{ok}^{(3)} z_k(n) + b_o^{(3)} \end{aligned} \quad (6.16)$$

where $y_A(n-1), y_A(n-2), \dots, y_A(n-N)$ are the input amplitude data, and the $\hat{y}_A(n)$ is the one-step prediction at time n . The synaptic weights $w_{ji}^{(1)}$ and bias term $b_j^{(1)}$ refer

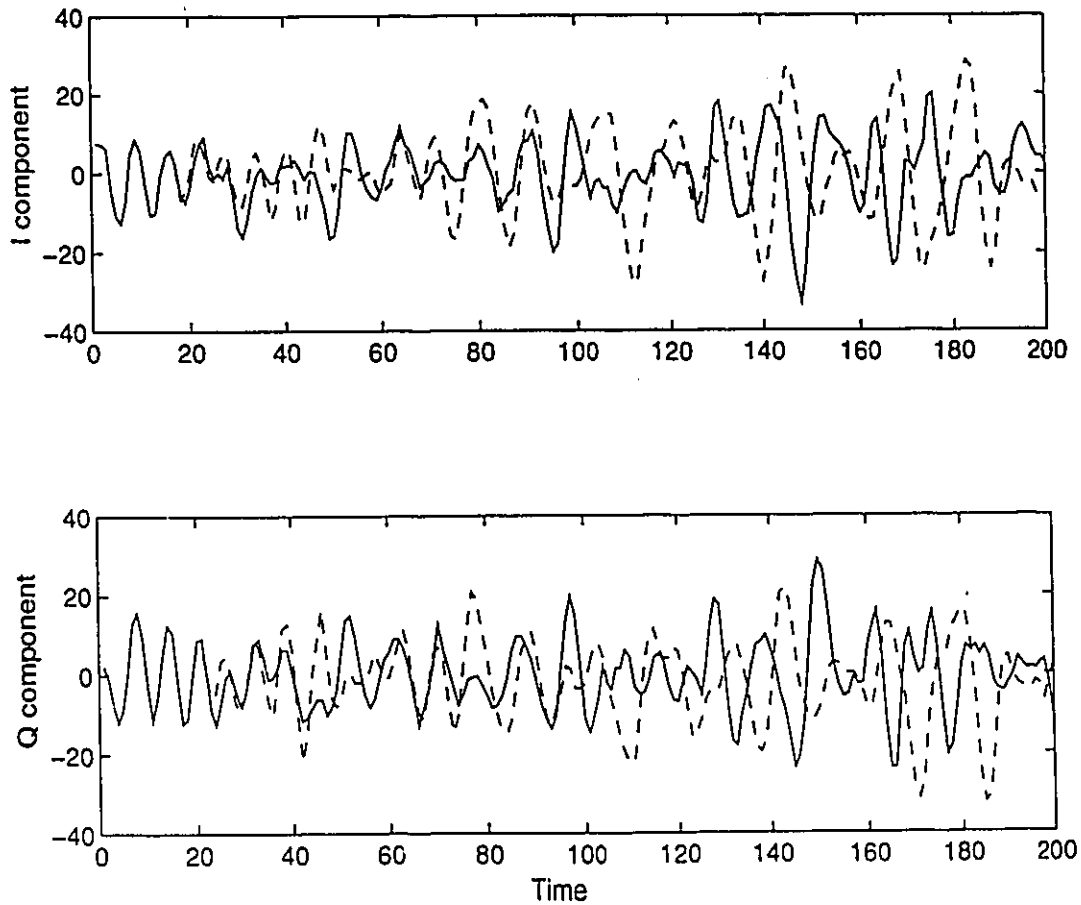


Figure 6.31: Recursive prediction result of the multilayer perceptron model shown in Figure 6.30 for a testing data chosen from a sea state with wave height of 3.78 *ft*. The solid curves refer to the original sea clutter waveforms, in which one is the in-phase component and the other is the quadrature component. The dashed curves refer to the recursive predicted waveforms, for which the first 30 points of the complex-valued sea clutter data (not shown in the figure) are used as the initial starting point.

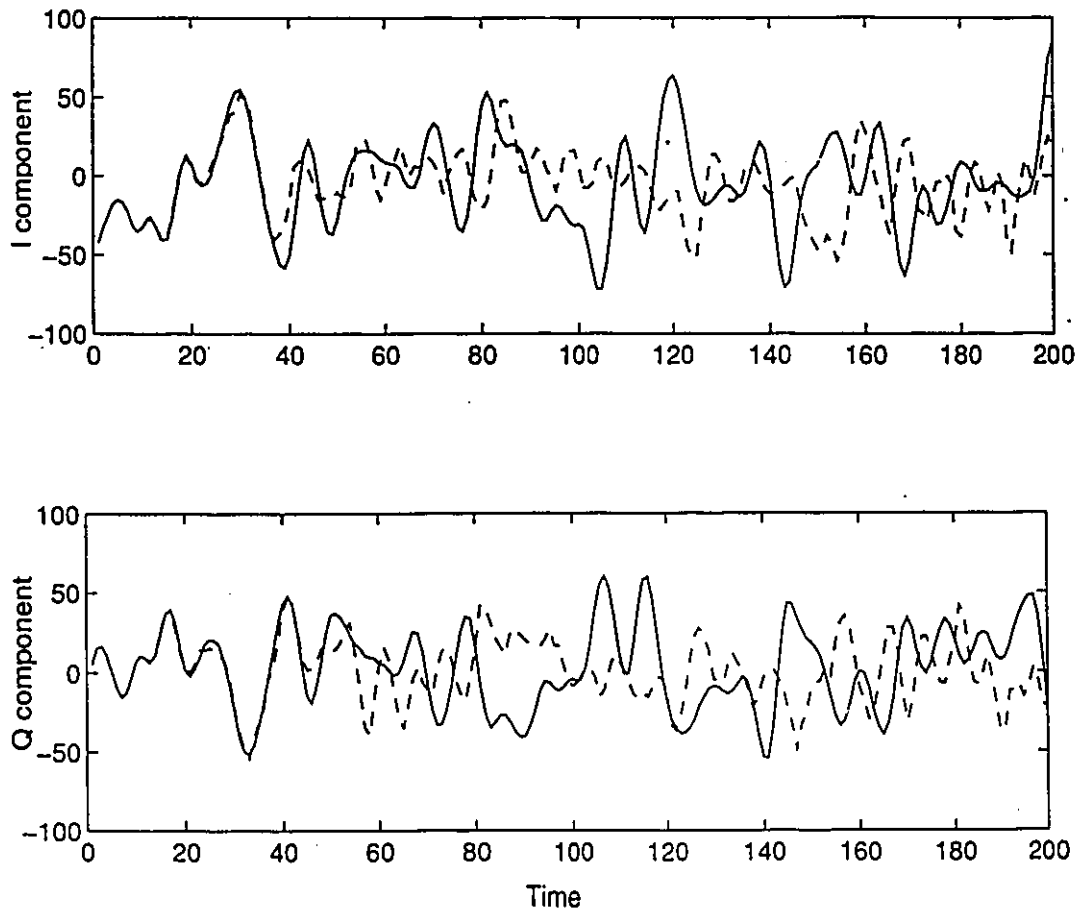


Figure 6.32: Recursive prediction result of the multilayer perceptron model shown in Figure 6.30 for a testing data chosen from a sea state with wave height of 4.54 *ft*. Traces are as defined in Figure 6.31.

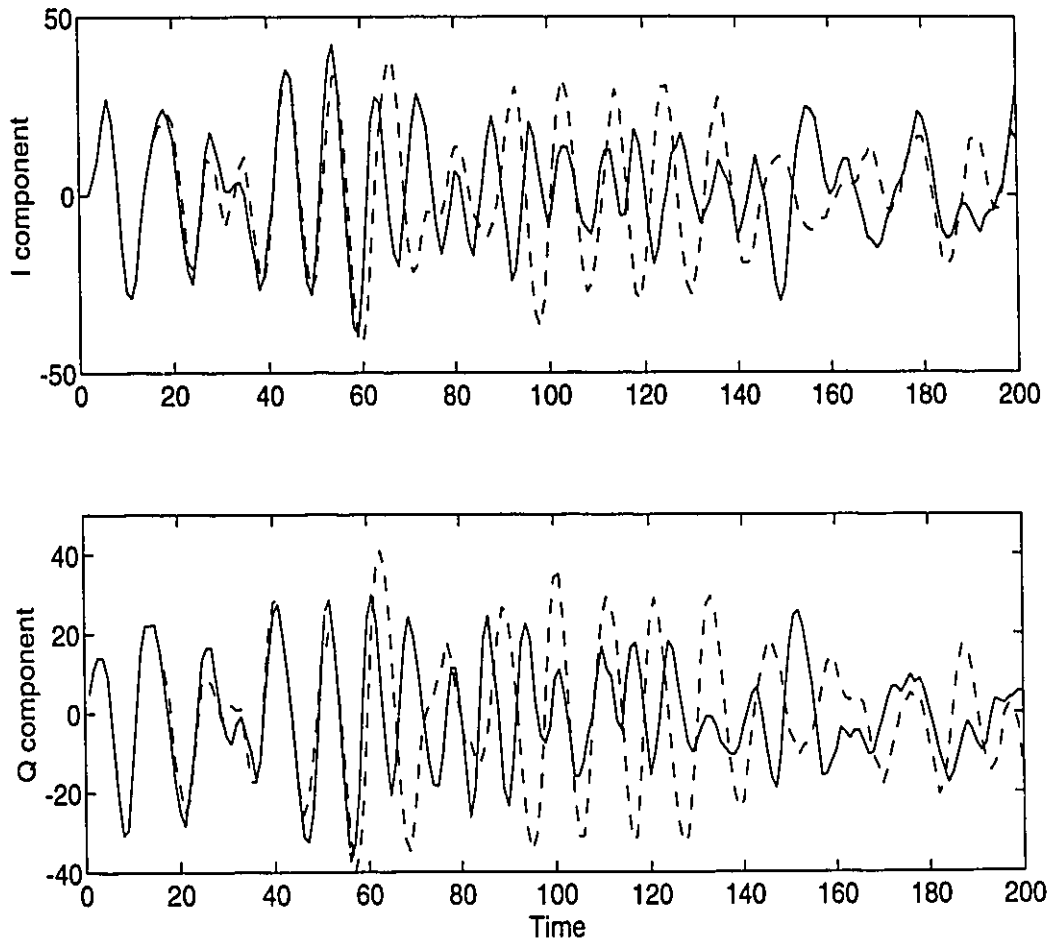


Figure 6.33: Recursive prediction result of the multilayer perceptron model shown in Figure 6.30 for a testing data chosen from a sea state with wave height of 5.95 ft. Traces are as defined in Figure 6.31.

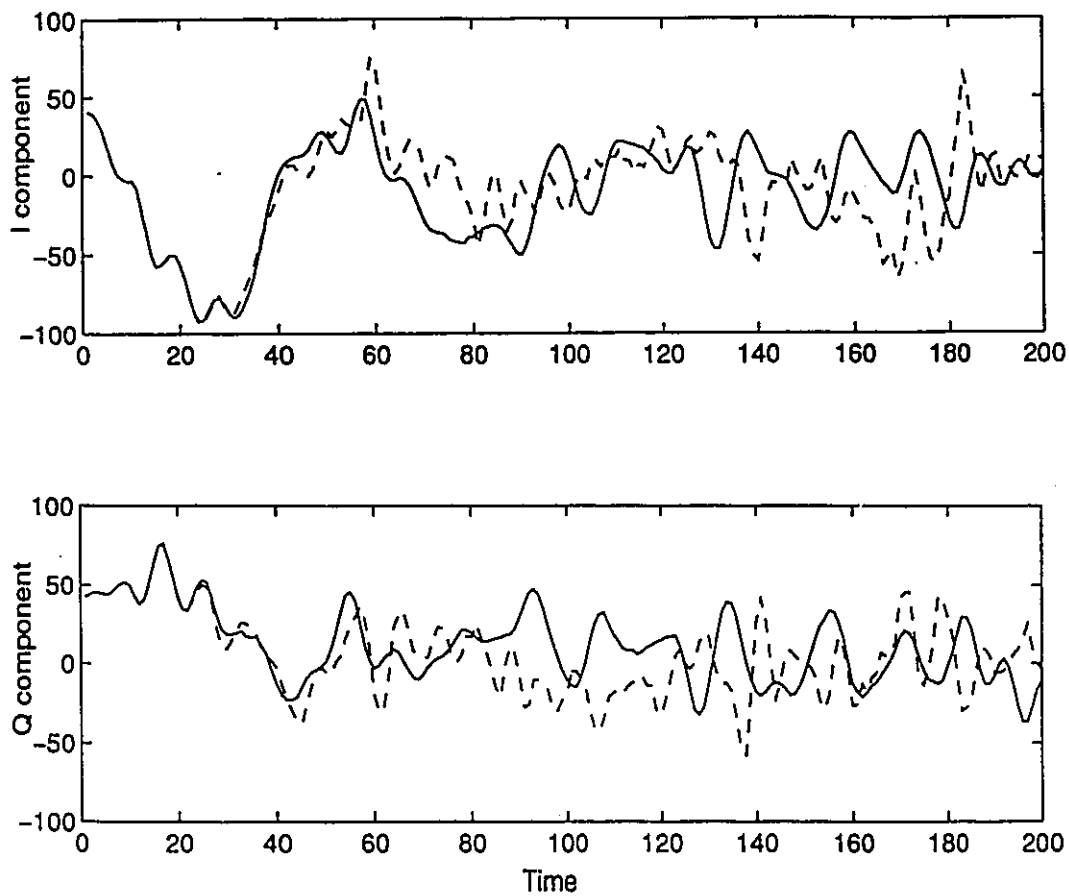


Figure 6.34: Recursive prediction result of the multilayer perceptron model shown in Figure 20 for a testing data chosen from a sea state with wave height of 8.30 *ft*. Traces are as defined in Figure 6.31.

to neuron j in the first hidden layer; the synaptic weights $w_{kj}^{(2)}$ and bias term $b_k^{(2)}$ refer to neuron k in the second hidden layer; and the synaptic weights $w_{ok}^{(3)}$ and bias term $b_o^{(3)}$ refer to the only neuron in the output layer. Note that the output layer is linear, which is the customary practice in the design of nonlinear regression models. It is further assumed that all the neurons in the two hidden layers of the model use the same sigmoid nonlinearity described by the logistic function:

$$\phi(v) = \frac{1}{1 + \exp(-v)} \quad (6.17)$$

The parameters of the model described in (6.16) are determined in accordance with the back-propagation algorithm acting on the sea clutter (amplitude) data.

The reconstructed dynamics built into the predictive model exhibits a phase portrait similar to that of the original dynamics of the sea clutter process, as displayed in Figure 6.35, where (a) refers to natural clutter data and (b) refers to artificial data generated by the model.

To further demonstrate the practical validity of the clutter map described here, we may compare the power spectral density (PSD) of the computer-generated sea clutter with that of the actual sea clutter. Figure 6.36 (a) shows the PSD of an actual sea clutter data set. Figure 6.36 (b) shows the PSD of an artificial data set generated by the neural network model of Figure 6.23, for which the first 50 points of the clutter data (used in Figure 6.36 (a)) were used as the initial starting point. The power spectrum estimation was carried out by averaging 100 different segments, each consisting of 1,024 points. The solid curves in Figure 6.36 represent the mean of the power spectrum, and the area between the two dashed curves shows the 95 percent confidence for the estimation of the power spectrum. Comparing the results presented in Figures 6.36 (a) and (b), we see that there is very good agreement between the natural and artificial clutter data sets.

Thus, the multilayer perceptron model of Figure 6.23 may be viewed as a *sea clutter simulator*. To further emphasize the significance of this model, we present in Figure 6.37(a) the amplitude distribution of a computer-generated data set, and compare it with the theoretical K -distribution. The solid curve in the figure shows the probability density function of a theoretical K -distribution with a shape parameter of 2.2 and a scale parameter of 1.25 [48, 99]. The dashed curve shows the histogram of 100,000 data points generated by

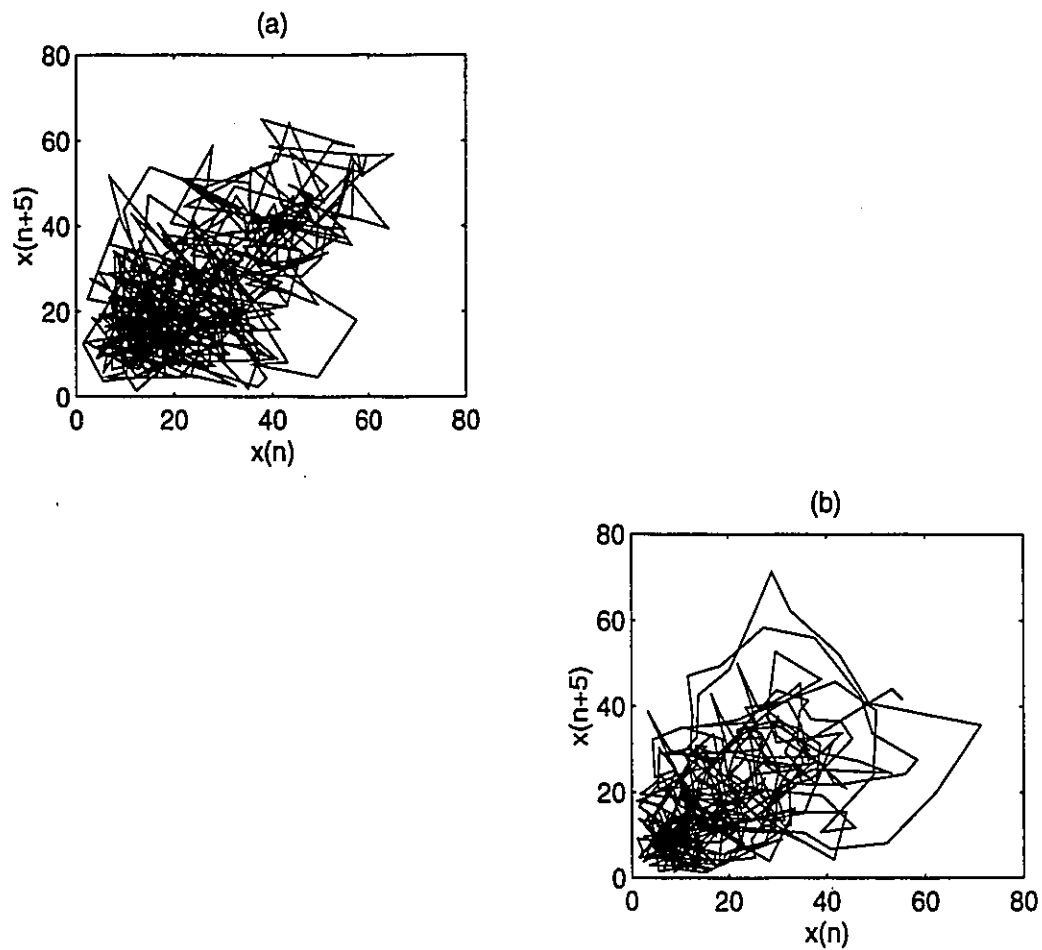


Figure 6.35: (a) The phase portrait of a natural clutter data set; (b) the phase portrait of an artificial data set generated by the neural network model shown in Figure 17, for which the first 50 points of the clutter data (used in (a)) are used as the initial starting point. Each data set consists of 500 points. Each point of these phase portraits is defined by a sample value of the pertinent time series and its delayed version corresponding to a time delay $\tau = 5$.

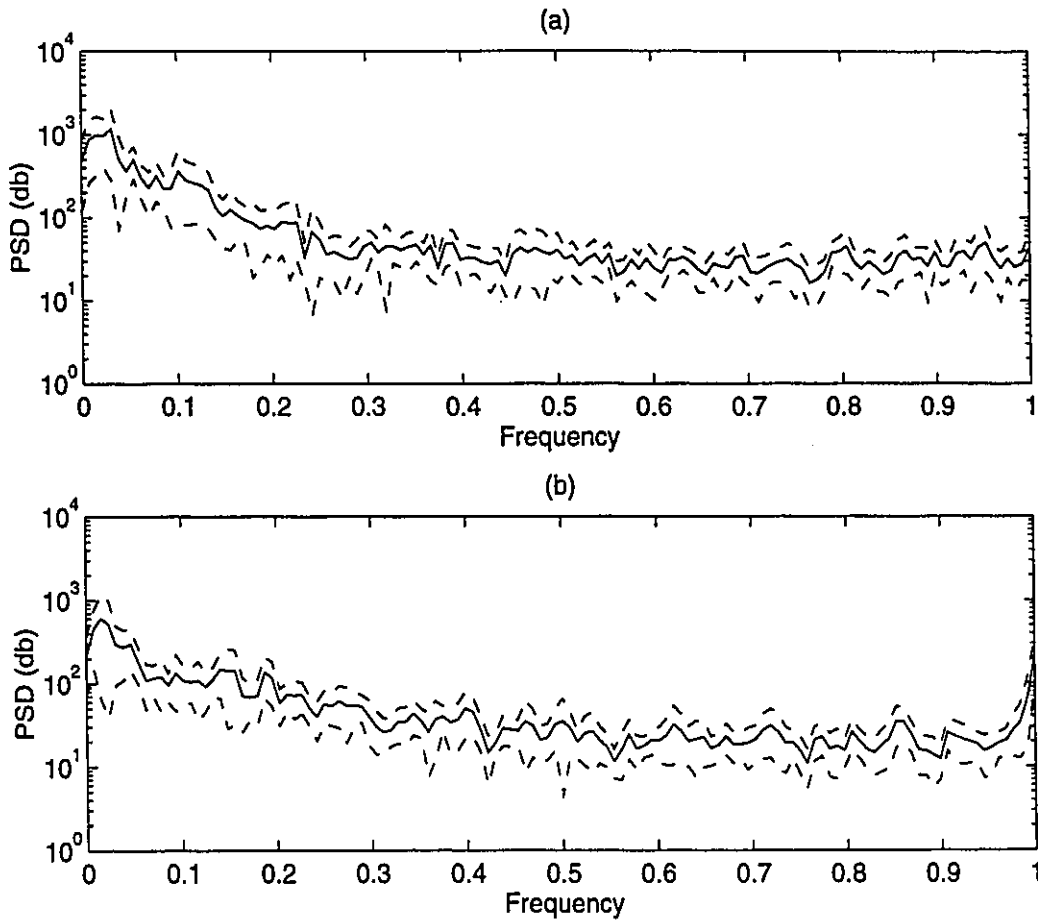


Figure 6.36: (a) The power spectral density (PSD) of a natural clutter data set; (b) the power spectral density of the artificial data set generated by the neural network model shown in Figure 17, for which the first 50 points of the clutter data (used in (a)) are used as the initial starting point. Each data set consists of 192,400 points. The area between the two dashed curves in (a) and (b) shows 95% confidence for the estimation of the power spectrum.

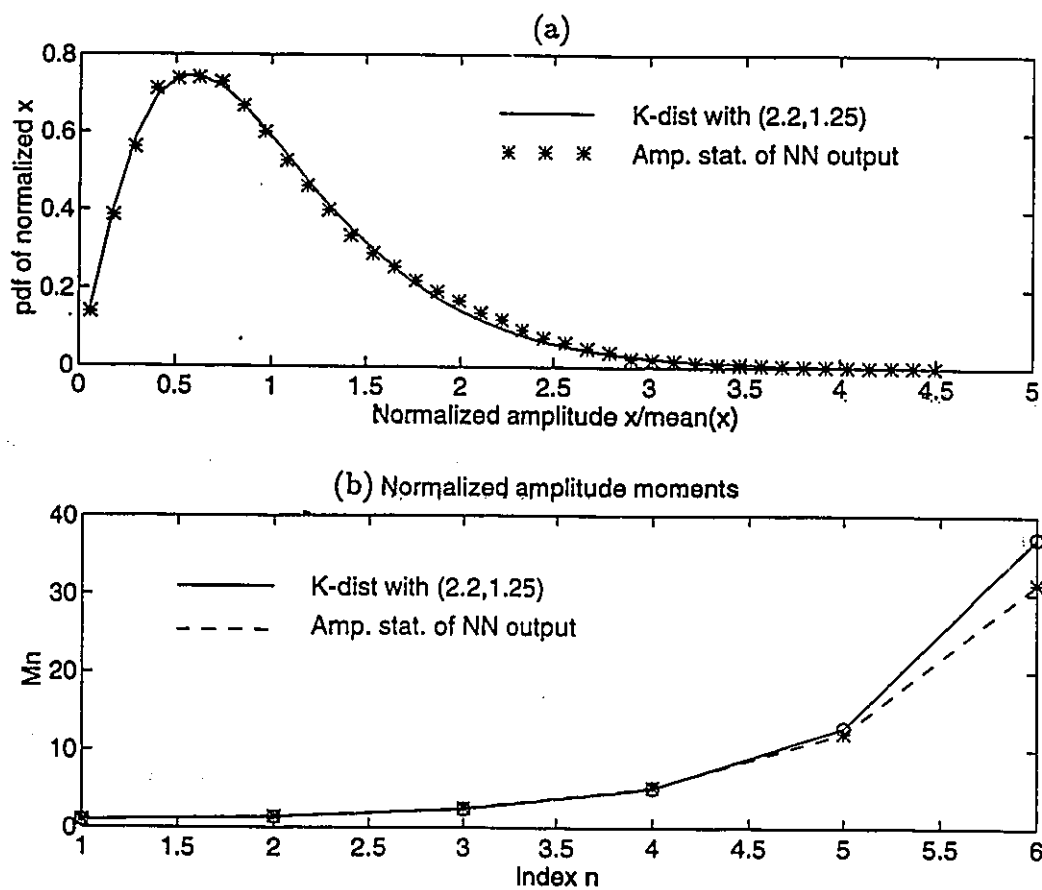


Figure 6.37: (a) The probability density function of the data generated by the multilayer perceptron model of Figure 17 (star points) and that of the theoretical K -distribution (solid curve). (b) A comparison of the normalized amplitude moments of the data generated by the multilayer perceptron model of Figure 17 (dashed curve) with those of the theoretical K -distribution (solid curve).

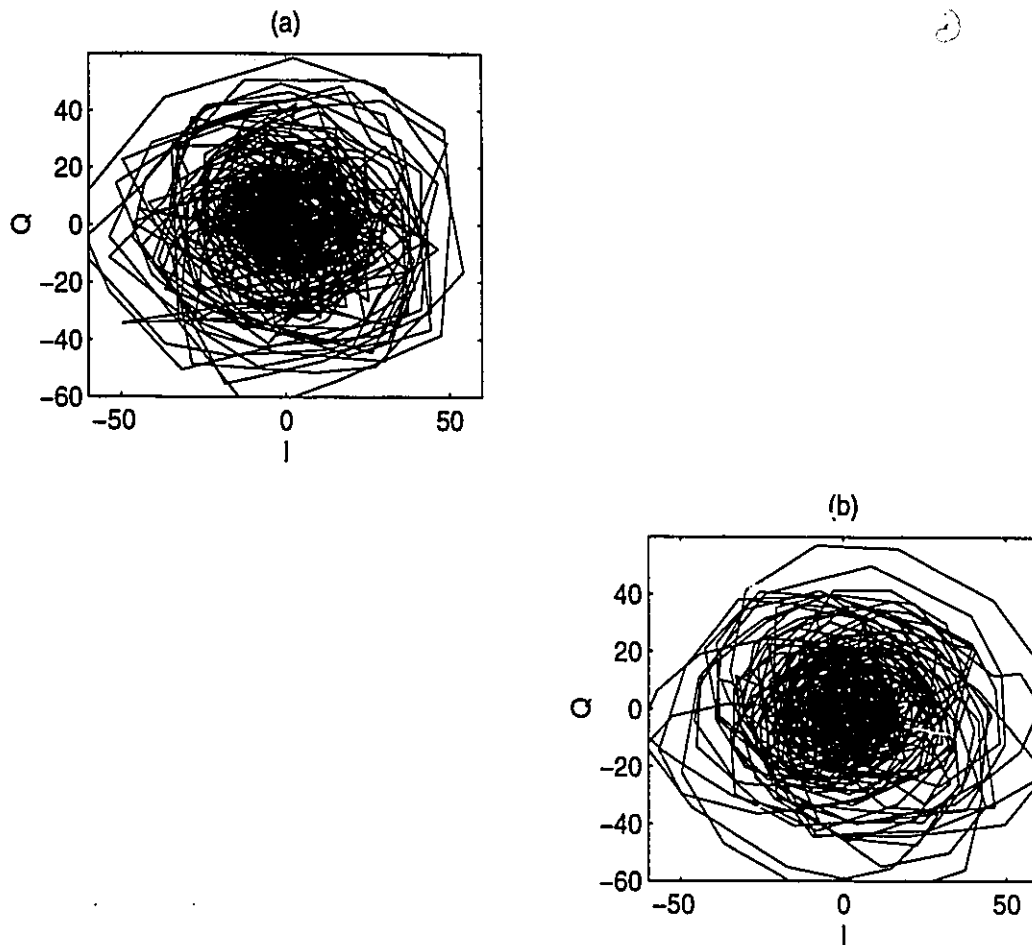


Figure 6.38: (a) a phase portrait of a natural clutter data set; (b) a phase portrait of an artificial data set generated by the neural network model shown in Figure 6.30, for which the first 30 complex-valued points of the natural clutter data are used to initiate the model.

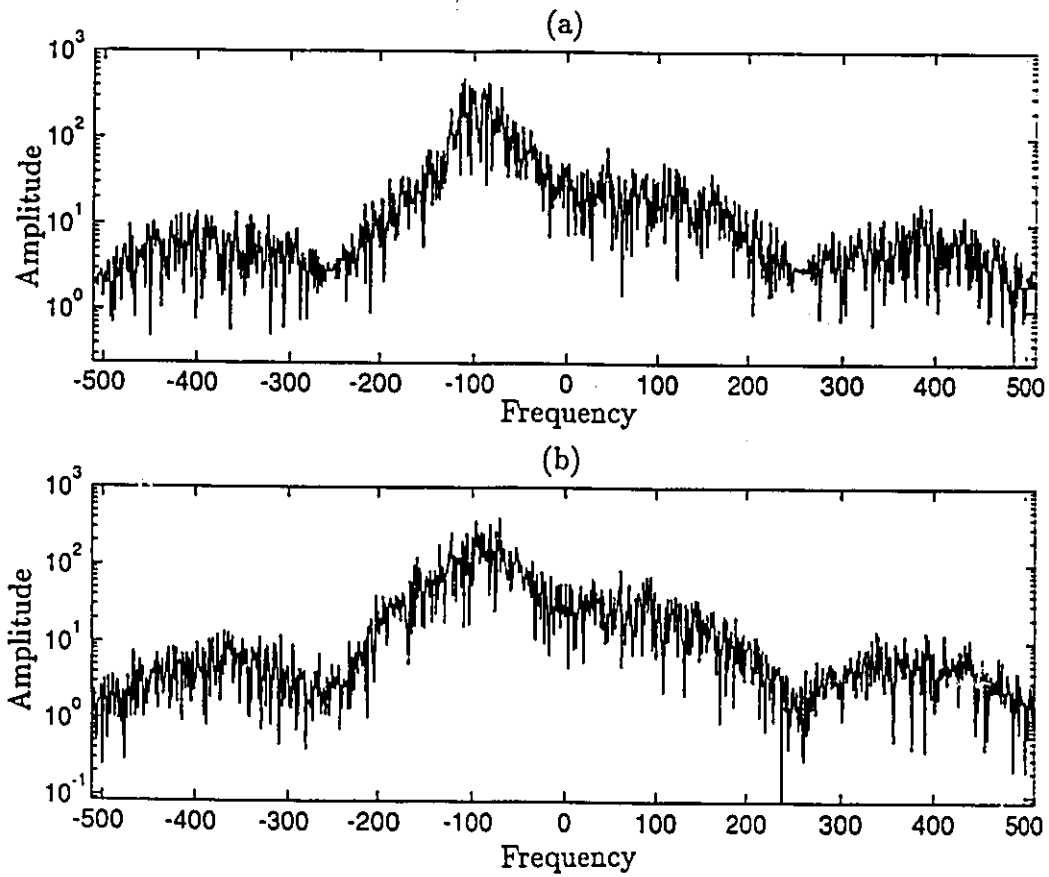


Figure 6.39: (a) The amplitude spectrum of a natural clutter data; (b) The amplitude spectrum of a clutter data generated by the neural network model shown in Figure 6.30.

the multilayer perceptron model. The agreement between these results is further illustrated in Figure 6.37(b), where we show the normalized amplitude moments, which are considered to be important parameters used to distinguish a K -distributed random process from other distributions. The solid curve in the figure corresponds to the theoretical K -distribution of part (a). The dashed curve corresponds to the data generated by the multilayer perceptron model. Here again the agreement is very good. The point that we wish to emphasize here is that a “deterministic” model described by the multilayer perceptron of Figure 6.23 is capable of generating a random-like “chaotic” clutter waveform that exhibits the statistical characteristics that are usually associated with actual radar data.

The clutter map defined by Equation (6.16) pertains to real-valued amplitude data. In a similar manner, we may define a complex-valued clutter map, based on the multilayer perceptron predictor of Figure 6.30 that involves the use of in-phase and quadrature components. The reconstructed dynamics built into the predictive model exhibits a similar phase portrait to that of the original dynamics of sea clutter, as shown in Figure 6.38, where (a) refers to the natural sea clutter data and (b) refers to artificial data generated by the model. Figure 6.39 shows the amplitude spectrum of a complex-valued received radar signal and that of an artificial data generated by the model shown in Figure 6.30. There is a very good agreement between the natural and artificial clutter data sets, providing the basis of a sea clutter simulator for use with a coherent radar.

6.8 Summary

In this chapter, a detailed experimental study using ireal-life radar data was presented. It demonstrates that sea clutter is essentially a chaotic process. The experimental evidence for this assertion is summarized as follows. Irrespective of whether the in-phase, quadrature, or amplitude (envelope) component is used, or whether like- or cross-polarization is employed, the following statements can be made:

- sea clutter is bounded;
- sea clutter has a finite and fractal correlation dimension lying in the range of 7 to 9; an embedding dimension of 10 for reconstruction of the underlying dynamics is

confirmed by the false nearest neighbors analysis;

- the largest Lyapunov exponent of sea clutter is positive (lying in the range between 0.03 and 0.045, normalized with respect to the pulse repetition frequency), indicating a strong dependence on initial conditions;
- sea clutter is locally (short-time) predictable.

The experimental evidence presented in this chapter is considered to be complete in light of what we know about chaos theory. It provides solid confirmation that sea clutter is chaotic. Moreover, the local predictability of sea clutter has been demonstrated experimentally by building a nonlinear predictive model in the form of a multilayer perceptron trained with the back-propagation algorithm³. The generalization capability of the model has been demonstrated by performing recursive (iterated) prediction, and local prediction time of the model has been shown to be inversely proportional to the largest Lyapunov exponent of sea clutter. Several experiments have shown that a model so designed has the ability to capture the underlying dynamics of sea clutter in accordance with Takens' embedding theorem. In particular, artificial data generated by the model undergoing recursive prediction exhibits characteristics very similar to those of natural sea clutter. In building the model, careful attention has been given to the selection of an appropriate size for the input layer, taking account of two specific parameters: the embedding dimension D_E , and the characteristic delay time τ .

The fact that sea clutter is chaotic may be exploited by trading off highly sophisticated but relatively inexpensive software for fairly expensive microwave hardware. Specifically, a chaos-based receiver using amplitude information only (as in a noncoherent marine radar) is capable of providing a detection performance comparable to that of a "conventional" Doppler-based receiver using both amplitude and phase information (as in a coherent marine radar). The chaos-based detector is built around the nonlinear predictive model of sea clutter. Details about design and performance evaluation of chaos-based detectors are discussed in the next chapter.

³It is feasible to use other supervised neural network models such as a radial basis function (RBF) network for the prediction [39, 43, 16].

Chapter 7

Detection of Small Targets in Sea Clutter

7.1 Introduction

The experimental results presented in the previous chapter provide convincing evidence that sea clutter is indeed the result of a chaotic dynamical system. Most importantly, we have computational models for reconstruction of the underlying dynamics of sea clutter process, which may be used to deal with noncoherent and coherent radar. In this chapter, we demonstrate experimentally how this prior information, namely, the chaotic property of sea clutter, can be exploited in radar detection. Our goal is to detect a small target (*e.g.*, a small piece of ice with a cross section of about 1 square meter or less) embedded in a background of sea clutter, using the chaos-based (or model-based) detection theory described in Chapter 3. First, we wish to demonstrate that in such an environment it is indeed possible to design a chaos-based detector using noncoherent radar data which can realize a performance comparable to that of a conventional receiver using coherent data. Then, we wish to demonstrate experimentally that a chaos-based detector using coherent radar data can provide a further improvement in receiver performance.

This chapter is organized as follows: In Section 7.2, a review of the constant false-alarm rate (CFAR) processors in radar signal processing is given. The Doppler CFAR processor is discussed in this section since it will be used as a frame of reference in comparing

different detection strategies. In Section 7.2, we describe the design of a chaos-based noncoherent detector, evaluate its detection performance, and compare it with the conventional Doppler CFAR processor. In Section 7.3, the design of a chaos-based coherent detector is described. The detection performance is evaluated and compared with the conventional Doppler CFAR processor. In Section 7.5, the effect of noise on the detection performance of both detectors is studied. Finally, a summary is given.

7.2 CFAR Detectors

The basic concept of the constant false alarm rate (CFAR) technique is that the amplitude of a test cell is compared to that of a set of reference cells. If the test cell amplitude is "similar" to those of the reference cells, a target-absent decision is made. If the test cell amplitude is not similar, a target-present decision is made. The CFAR processor is commonly used in present radar systems [22, 67]. It prevents saturation of the detector due to increases in clutter or noise by adapting the detection threshold in step with the changing clutter or noise conditions. Likewise, in the case that the clutter or noise decreases, it lowers the threshold thereby increasing the detectability of weak targets which otherwise would be missed. The operation of CFAR systems has been widely studied [22, 67]. The basic idea is to estimate a sufficient statistic of the clutter, and to use this estimate to control a detection threshold so that a constant false alarm rate is obtained.

CFAR detectors are either noncoherent or coherent processors. The noncoherent CFAR schemes all involve adaptively setting the decision threshold in some manner. These processors can be subdivided into two groups: the spatial CFAR processors and the temporal CFAR processors. The spatial CFAR's involve using returns from neighboring cells to determine the required detection threshold for a given probability of false alarm. On the other hand, the temporal CFAR's use scan-to-scan processing at each resolution cell in order to determine the threshold. Clutter maps, two-dimensional maps (azimuth *vs* range) where the mean amplitude in each cell is determined by successive scans, are used for comparison to each test cell. This method is particularly susceptible to self-masking by targets in the radar cell, and therefore the clutter map cell size is generally larger than the radar resolution cell size. Additional drawbacks result from making the clutter map cells too large as well.

Other noncoherent methods include nonlinear receivers, an example of which is the binary integrator.

Coherent CFAR techniques are also divided into two main categories: coherent clutter maps and spectral methods. However, only the spectral methods are applicable for sea clutter; this is due to the uniform phase distribution. Therefore, coherent clutter maps will not be considered any further in this discussion. Spectral methods (also called Doppler CFAR) operate as follows. For each of several range cells, the spectrum is computed using either Doppler filter banks (FFT) or other spectrum estimators. Then each Doppler bin for the test range cell is compared with the Doppler bins at neighboring range cells in order to determine whether or not a target is present. In this way, CFAR is achieved by setting the adaptive threshold for the test cell based upon the corresponding Doppler bins of neighboring cells. Not only is a decision made, but also the Doppler frequency of the target is determined.

7.3 Chaos-Based Noncoherent Detector

Given a fully coherent system, the amplitude component of the received radar signal is defined by

$$y_A(t_k) = \sqrt{y_I^2(t_k) + y_Q^2(t_k)} \quad (7.1)$$

where y_I is the I-channel component, and y_Q is the Q-channel component. Both components are of zero mean. To extract the target information from the received data by using the neural network model described in Figure 17, the prediction error is defined as:

$$\varepsilon(t_k) = y_A(t_k) - \hat{y}_A(t_k) \quad (7.2)$$

where $y_A(t_k)$ is the actual value of the received signal at time t_k , and $\hat{y}_A(t_k)$ is the prediction on the basis of a set of previous samples. When $y_A(t_k)$ consists of sea clutter alone, the prediction error $\varepsilon(t_k)$ is small at the output; but when a growler (*c.g.*, small piece of ice) is present in sea clutter, $\varepsilon(t_k)$ assumes a relatively large amplitude. On the basis of these observations, we can now describe the chaotic detection algorithm based on the neural network model described previously for sea clutter as follows:

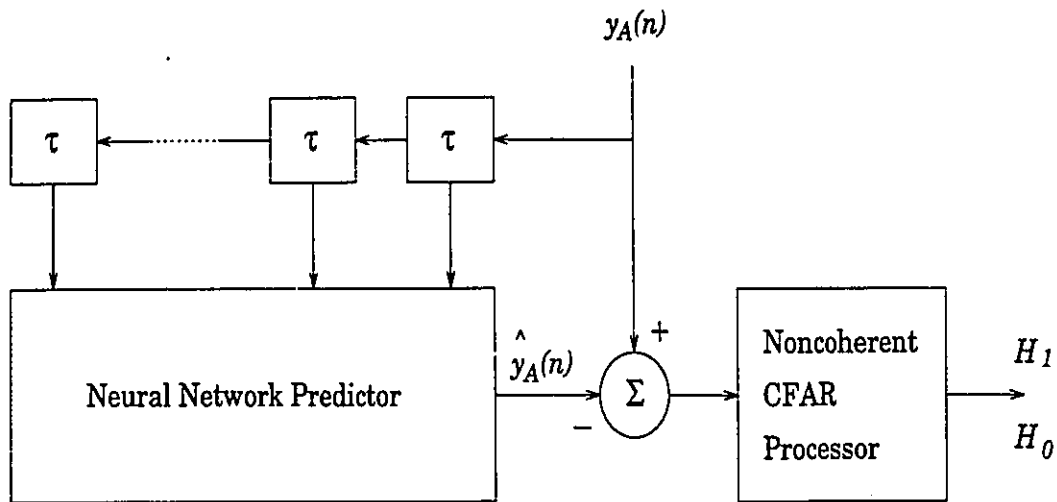


Figure 7.1: A chaos-based noncoherent detector used for the detection of a small target in sea clutter. It consists of a neural network predictor and a “conventional” noncoherent CFAR processor.

Step 1: For a received signal $y_A(t)$, where $t \in [0, T]$, calculate the prediction $\hat{y}_A(t)$ produced by the neural network model; hence, compute the prediction error $\varepsilon(t)$ using (7.2).

Step 2: Integrate the error in the observation interval, and determine its mean-squared value

$$R_\varepsilon = \frac{1}{T} \int_0^T \varepsilon^2(t) dt \quad (7.3)$$

Step 3: Compare R_ε with a threshold for a given probability of false alarm; if it exceeds the threshold, a target is declared to be present; otherwise, the algorithm decides that no target is present.

The chaos-based detection algorithm is illustrated by the block diagram of Figure 7.1. The new detection scheme differs from a conventional noncoherent radar detector in that a neural network predictor, which was trained on clutter only data described in last chapter, is employed as a preprocessor and the detection is based on the prediction error.

As discussed in section 6.3.1, the chaos-based detection algorithm described here is designed to transform the detection of a target in unknown colored noise (a term used in the traditional sense) into that of detecting an unknown signal in additive white Gaussian noise,

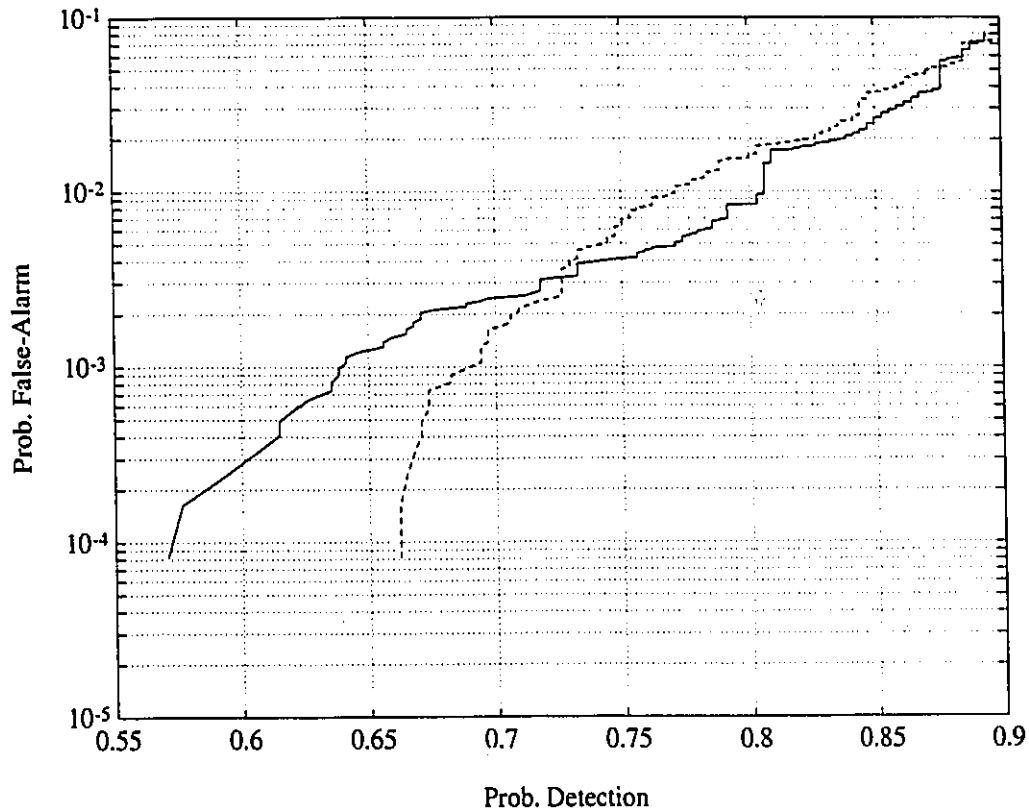


Figure 7.2: Detection performance of the chaos-based noncoherent detector. The dashed curve is the ROC of the chaos-based noncoherent detector. The solid curve shows the ROC of a “conventional” Doppler CFAR detector for comparison. It can be seen that the chaos-based noncoherent detector has comparable performance to the Doppler CFAR detector.

which is a communication theorist’s dream. In effect, the learnable part of the information contained in sea clutter (*i.e.*, the chaotic component) is transplanted in the mapping neural network, providing an output that is essentially a sample function of white Gaussian noise.

The detection results are shown in Figure 7.2. The dashed line of this figure is a sample Receiver Operating Curve (ROC) of the noncoherent chaos-based detector. For the calculation of the ROC, more than 12,000 independent data sets are used, each data set consists of 150 time-series samples. For the purpose of comparison, the “conventional” Doppler CFAR algorithm is employed to calculate the ROC shown by the solid curve in

Figure 7.2 using the same data sets. This time, however, each data set includes 256 sample pairs of both I-channel and Q-channel. Comparing the two curves of Figure 7.2, we note the following:

- A noncoherent chaos-based detector has a performance comparable to that of a conventional Doppler CFAR processor by virtue of the fact that the former uses dynamic reconstruction (based on amplitude information) to exploit the chaotic property of sea clutter.

7.4 Chaos-Based Coherent Detector

According to Takens' embedding theorem, a single signal component is sufficient for the dynamic reconstruction of a chaotic system. In a coherent radar system, the received signal consists of two components, namely, the in-phase and quadrature components. These two components provide independent observations of the ocean environment for exactly the same sea state; most importantly, they provide Doppler information about the target signal that is absent from a noncoherent radar. However, in so far as dynamic reconstruction of the sea clutter is concerned, the in-phase and quadrature components, individually or collectively, provide essentially the same results. It is important to note that in light of the computer experiments reported in Section 3.3, we expect to find that the dynamic reconstruction model preserves the Doppler information contained in the target signal.

Let the received signal of a coherent radar be denoted by

$$y(t_k) = y_I(t_k) + jy_Q(t_k) \quad (7.4)$$

where y_I is the in-phase component, y_Q is the quadrature component, and $j = \sqrt{-1}$. Applying the previous samples $y(t_{k-1}), y(t_{k-2}), \dots, y(t_{k-N})$ of the received signal to the neural network predictor described in Figure 19, we get the one-step prediction

$$\hat{y}(t_k) = \hat{y}_I(t_k) + j\hat{y}_Q(t_k) \quad (7.5)$$

Accordingly, the complex-valued prediction error is defined as

$$\begin{aligned} \varepsilon(t_k) &= y(t_k) - \hat{y}(t_k) \\ &= \varepsilon_I(t_k) + j\varepsilon_Q(t_k) \end{aligned} \quad (7.6)$$

where

$$\varepsilon_I(t_k) = y_I(t_k) - \hat{y}_I(t_k) \quad (7.7)$$

$$\varepsilon_Q(t_k) = y_Q(t_k) - \hat{y}_Q(t_k) \quad (7.8)$$

Figure 7.3 shows a block diagram of the chaos-based coherent detector. It consists of a neural network predictor operating on the in-phase and quadrature components of the received signal, followed by a Doppler CFAR processor. The only difference between this receiver and the “conventional” coherent receiver is that a neural network predictor is employed in the former case under the assumption that sea clutter is the result of deterministic chaos. Note that in the design of this predictor presented in subsection 6.3.2, a multilayer perceptron with real-valued synaptic weights is used. Alternatively, we may use a multilayer perceptron with complex-valued synaptic weights for the predictor, in which case we would have to use the complex back-propagation algorithm [30, 18, 40] to train the predictor. In any event, the feature that distinguishes the chaos-based coherent detector of Figure 7.3 from the chaos-based noncoherent detector of Figure 7.1 is that in the coherent case the predictor handles complex-valued inputs and produces complex-valued outputs, whereas in the noncoherent case both the input and output are real-valued.

Figures 7.4 and 7.5 present some experimental results to demonstrate the predictive capability of the chaos-based coherent detector of Figure 7.3:

1. Figure 7.4(a), corresponding to a data set of Table 1 with sea clutter only, shows the amplitude spectrum of a complex-valued received radar signal. Figure 7.4(b) shows the amplitude spectrum of the complex-valued prediction error produced at the output of the neural network of Figure 7.3. The whitening property of the predictor is clearly visible in Figure 7.4(b).
2. Figure 7.5(a), corresponding to a data set of Table 5.1 with target plus clutter, shows the amplitude spectrum of the new complex-valued received radar signal. Figure 7.5(b) shows the amplitude spectrum of the prediction error. As expected, the target signal is much more clearly visible in the output than in the input of the predictor.

Finally, Figure 7.6 shows the detection performance of the chaos-based coherent detector. The dashed curve in the figure is a sample Receiver Operating Curve of the

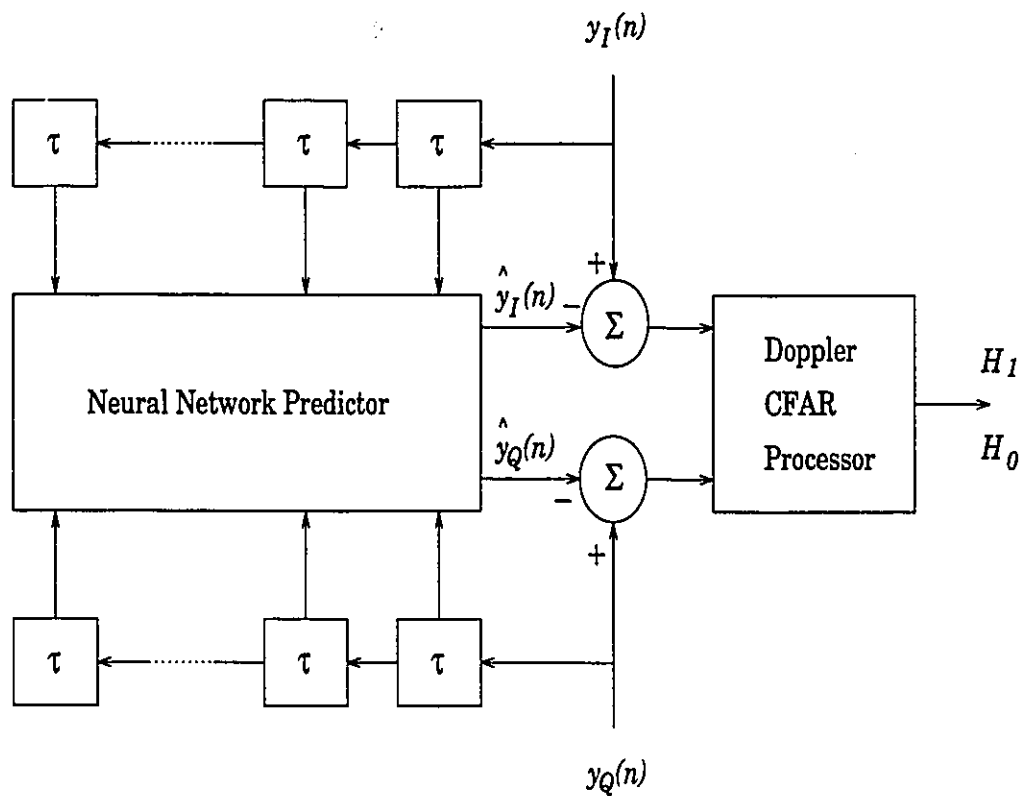


Figure 7.3: A chaos-based coherent detector used for the detection of a small target in sea clutter. It consists of a neural network predictor and a “conventional” Doppler CFAR processor.

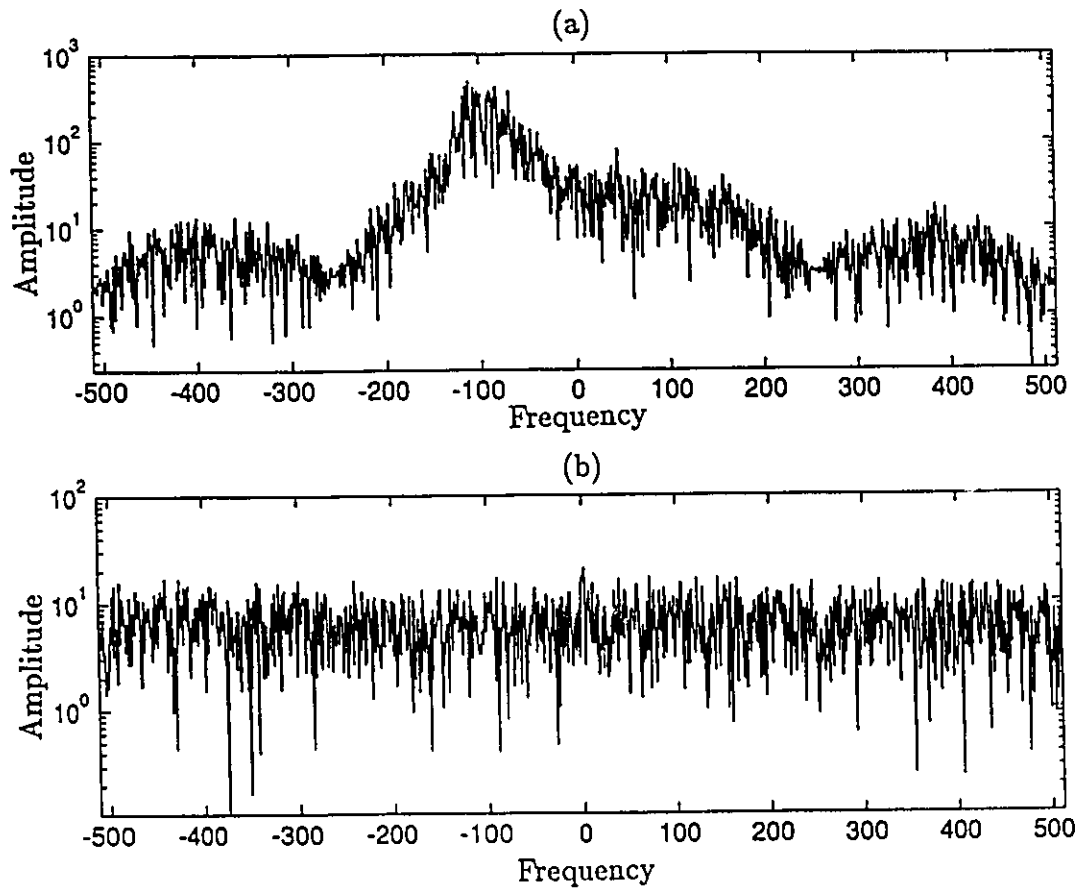


Figure 7.4: (a) Amplitude spectrum of sea clutter data produced by a coherent radar. (b) Amplitude spectrum of the prediction error produced by the neural network predictor of Figure 7.3 for the same clutter data.

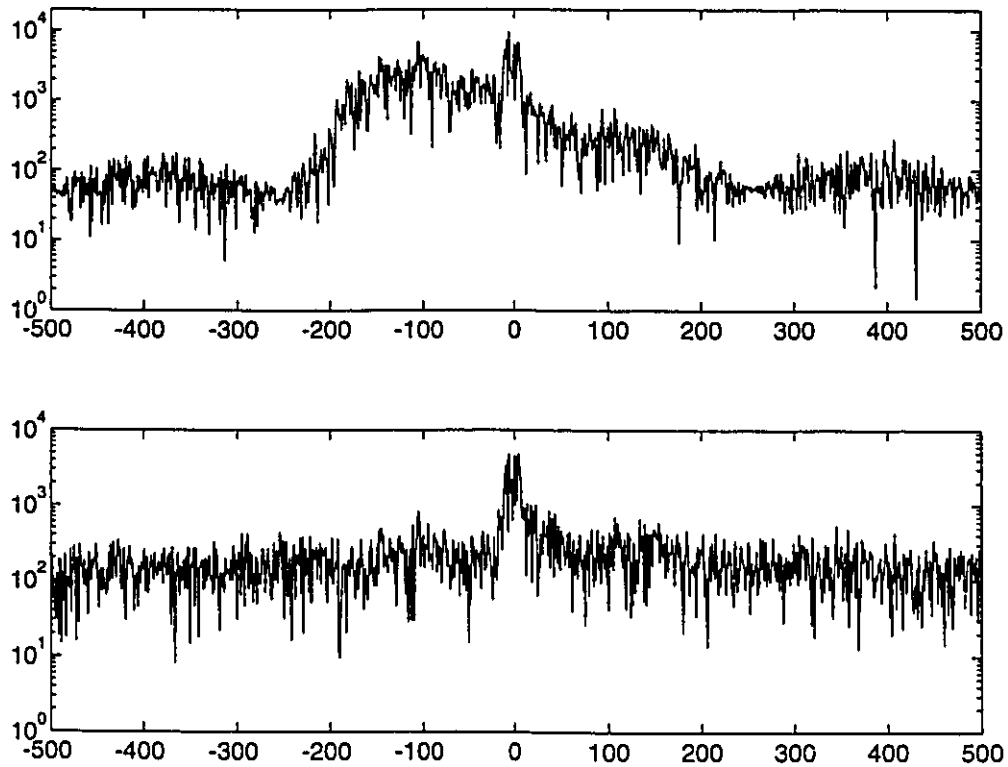


Figure 7.5: (a) Amplitude spectrum produced by a coherent radar data set containing both target signal and clutter. (b) Amplitude spectrum of the prediction error produced by the neural network predictor of Figure 7.3 for the same target plus clutter data.

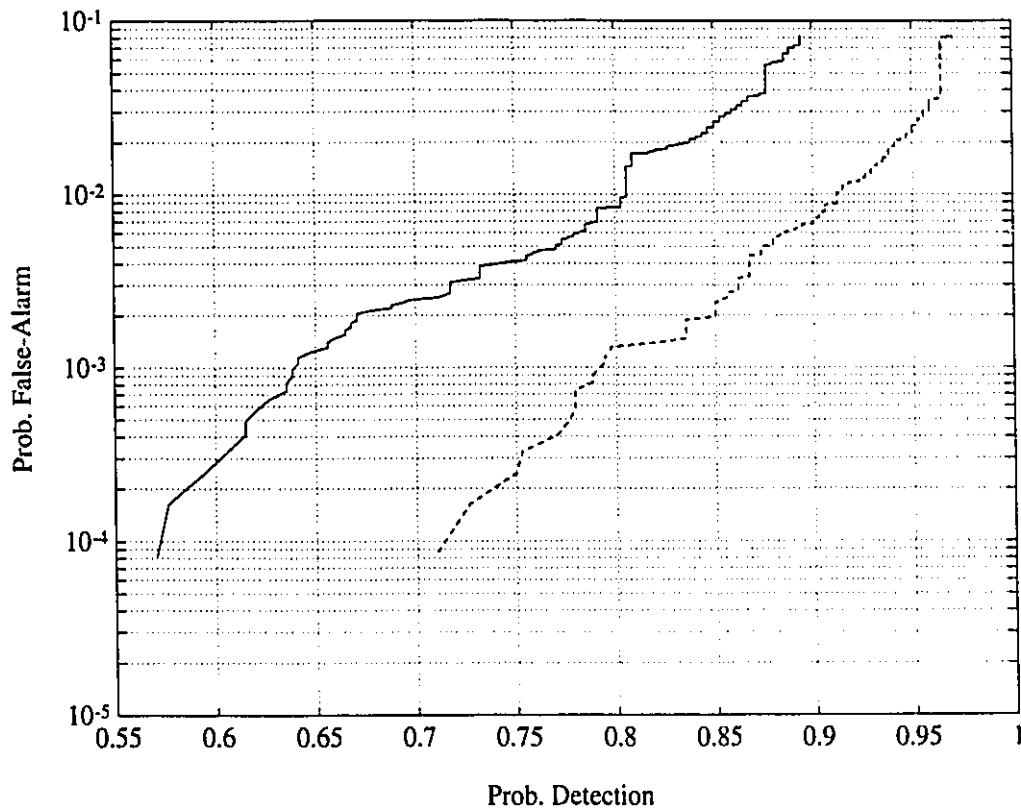


Figure 7.6: Detection performance of the chaos-based coherent detector. The dashed curve is the ROC of the chaos-based coherent detector. The solid curve shows the ROC of a "conventional" Doppler CFAR detector for comparison. It can be seen that the chaos-based coherent detector is superior to the Doppler CFAR detector.

chaos-based detector of Figure 7.3, while the solid curve is the Receiver Operating Curve of the conventional Doppler CFAR detector. The data used for the computations are exactly the same as those used for the computations of Figure 7.2. From the results shown in Figure 7.6, the following points are noted:

- A chaos-based coherent detector is superior to a conventional Doppler CFAR detector by virtue of using the prior information that sea clutter is chaotic. But, it is more complicated than a conventional Doppler CFAR detector since a nonlinear predictor is included in the receiver.
- The coherent version of a chaos-based detector is superior to its noncoherent version by virtue of the fact that the coherent system preserves the Doppler information about the target signal, which is not available to the noncoherent system.

7.5 Effect of Noise on Receiver Performance

In much of the material presented in this chapter so far the effect of the additive receiver noise has been ignored. This is justifiable due to the fact that the clutter-to-noise ratio for the data sets used in the study is known to be high. In practice, this condition is not always satisfied. Given that the sea clutter permits the use of a chaotic model, it is important to evaluate the sensitivity of the chaos-based detector described herein to additive front-end receiver noise, which is always present in the received signal.

As mentioned previously, we believe that our sea clutter data have a high clutter-to-noise ratio of about 35 *dB*, measured with respect to the additive front-end noise ratio. To evaluate the sensitivity of our chaos-based detectors to noise in the received signal, a series of subjective tests were performed on the chaos-based detectors with different noisy test data. These noisy test data consist of the high clutter-to-noise ratio data described in Chapter 5, to which computer-generated white noise of different power levels is added. The clutter-to-noise ratio of the resulting data is estimated to be 32 *dB*, 28 *dB*, 24 *dB*, and 20 *dB*, respectively.

The evaluation procedures can be summarized as follows:

1. Generate white Gaussian noise of different power levels with respect to the average power of sea clutter data.
2. Add the generated white noise to both the clutter data only and the data of clutter plus target; each data set corresponds to an individually generated noise component.
3. Apply the resulting data of the last step to the chaos-based detector described previously as a new received signal, set a constant-false-alarm rate, and measure the probability of detection.

The results of this experiment are shown in Figures 7.7 and 7.8. The x -axis of these figures corresponds to the clutter-to-noise ratio (CNR), and the y -axis of those figures is the probability of detection. Figure 7.7 presents the results corresponding to the chaos-based noncoherent detector, whereas the solid curve corresponds to a false-alarm rate (FAR) set to be 10^{-3} , and the dashed curve corresponds to a false-alarm rate of 10^{-2} . Figure 7.8 presents the results corresponding to the chaos-based coherent detector. Here again the solid curve corresponds to a false-alarm rate of 10^{-3} , and the dashed curve corresponds to a false-alarm rate of 10^{-2} .

From these two figures, the following observations are made:

- As the noise power in the input signal of the receiver is increased, the detection probability of the chaos-based detectors is progressively reduced.
- For a false-alarm rate of 10^{-3} or 10^{-2} , the probability of detection of the chaos-based noncoherent detector is less than 0.5 when the clutter-to-noise ratio is less than 29 dB or 24 dB respectively, which is practically useless.
- For a false-alarm rate of 10^{-3} or 10^{-2} , the probability of detection of the chaos-based coherent detector is always greater than 0.67 when the clutter-to-noise ratio is greater than 20 dB .

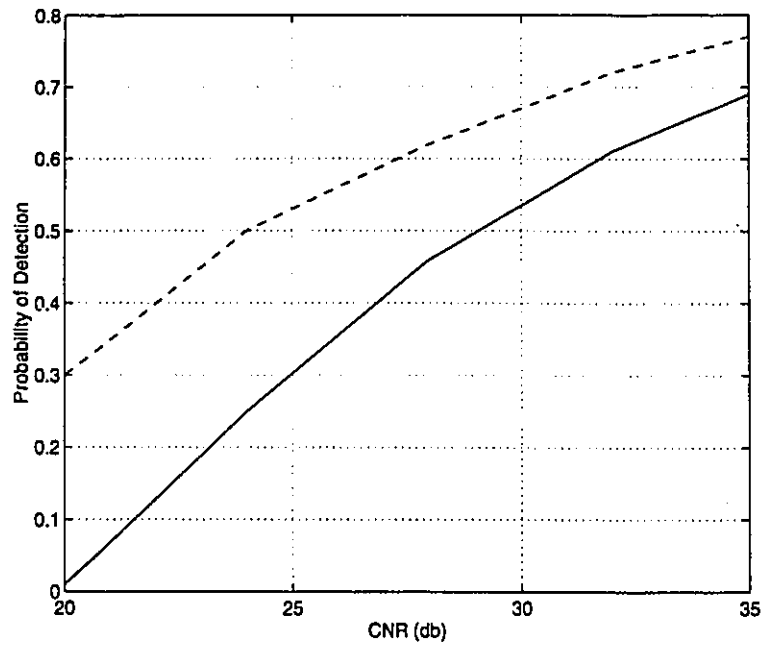


Figure 7.7: Detection performance on the receiver noise for the chaos-based noncoherent detector, where the solid line corresponds to a false alarm rate of 10^{-3} and the dashed line corresponds to a false alarm rate of 10^{-2} .

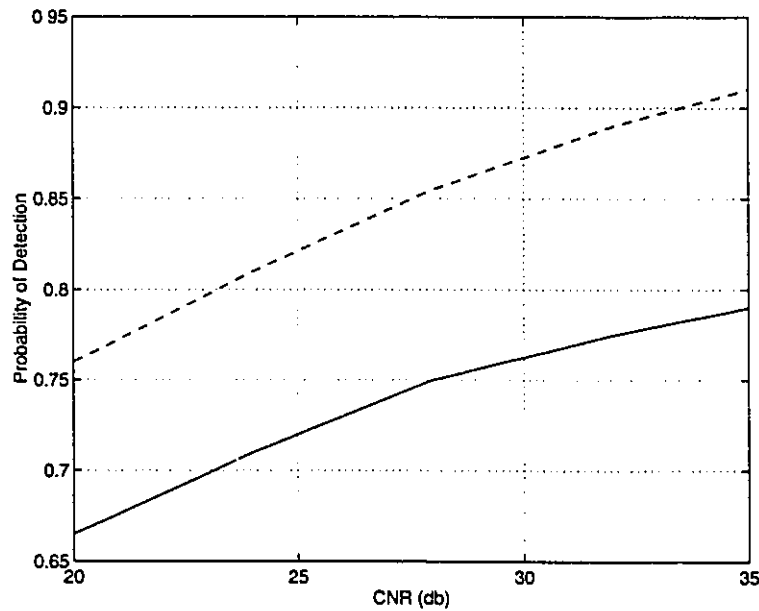


Figure 7.8: Detection performance on the receiver noise for the chaos-based coherent detector, where the solid line corresponds to a false alarm rate of 10^{-3} and the dashed line corresponds to a false alarm rate of 10^{-2} .

7.6 Summary

In this chapter, the design of two chaos-based detectors for the detection of small target in sea clutter has been described. The detection performance of these detectors has been evaluated by using real radar data, and compared with the conventional Doppler CFAR processor. From the results reported in this chapter, the following summary is made:

- A chaos-based noncoherent detector operating on amplitude information only is able to perform in a manner comparable to a conventional coherent CFAR detector for the detection of small targets in sea clutter by virtue of using the prior information that sea clutter is chaotic.
- A chaos-based coherent detector is superior to both the conventional coherent detector and the chaos-based noncoherent detector.
- A chaos-based noncoherent detector is much more sensitive to additive front-end receiver noise than a chaos-based coherent one.

Chapter 8

Chaotic Modeling Using Modular Network

8.1 Introduction

Earlier we remarked on the possible existence of a “global” predictive model for reconstruction of the chaotic dynamics of sea clutter. For such a model to be of practical value, it should be capable of dealing with a wide variety of sea states, which would be exploited to provide a further improvement in detection performance. A multilayer perceptron trained using the back-propagation algorithm is a universal approximator [45]. But it has a shortcoming of continuing to modify the synaptic weights that have already developed useful properties for a task when it is applied to a new task that is incompatible with the previous one [92]. Consequently, after learning the second task, the multilayer perceptron is no longer able to perform the first task. We may of course alternate the training process between tasks, and carry on in this way until the multilayer perceptron eventually learns both tasks. However, the price paid for it is a prolonged learning process. The generalization performance of the multilayer perceptron may also be affected by the blocked presentation of incompatible training data [39].

To overcome the shortcomings of the multilayer perceptron for reconstruction of the chaotic dynamics of sea clutter, we may adopt a “modular” approach for designing the predictive model. Specifically, we envision a number of multilayer perceptrons, each of which is trained using the back-propagation algorithm for a particular range of sea states,

and designed to capture the underlying chaotic dynamics that applies essentially to that particular range of sea states. The individual outputs of the multilayer perceptrons are then mediated by means of a *gating network* [39, 46]. In effect, the multilayer perceptrons compete with each other in an unsupervised manner for the right to produce the one-step prediction of the sea clutter; that right is decided by the gating network. The modular network described here is well suited for solving nonlinear regression problems [39, 46, 47].

This chapter is organized as follows. In Section 8.2, we present a modular network structure, and describe a learning algorithm for the specific model. In Section 8.3, we apply the modular network for the chaotic modeling of sea clutter, and compare the recursive prediction performance with the multilayer perceptron model described in Chapter 6. The chapter concludes in Section 8.4 with a summary of results.

8.2 Modular Network

A specific modular network structure is shown in Figure 8.1. The structure consists of K supervised modules called *expert networks*, and an integrating unit called a *gating network* that performs the function of a mediator among the expert networks.

Let the training examples be denoted by an input vector \mathbf{x} of dimension m and desired response vector \mathbf{d} of dimension q . The input vector \mathbf{x} is applied to the expert networks and the gating network simultaneously. Let \mathbf{y}_i denote the output vector of the i th expert network, let g_i denote the activation of the i th output neuron of the gating network, and let \mathbf{y} denote the output vector of the whole modular network. We may then write

$$\mathbf{y} = \sum_{i=1}^K g_i \mathbf{y}_i \quad (8.1)$$

The goal of the learning algorithm used to train the modular network of Figure 8.1 is to model the probability distribution of the set of training patterns $\{\mathbf{x}, \mathbf{d}\}$. We assume that the patterns $\{\mathbf{x}, \mathbf{d}\}$ used to do the training are generated by a number of different regressive processes. Then a desired response vector \mathbf{d} is generated by the selected rule i according to the regressive process

$$\mathbf{d} = \mathbf{F}_i(\mathbf{x}) + \varepsilon_i \quad (8.2)$$

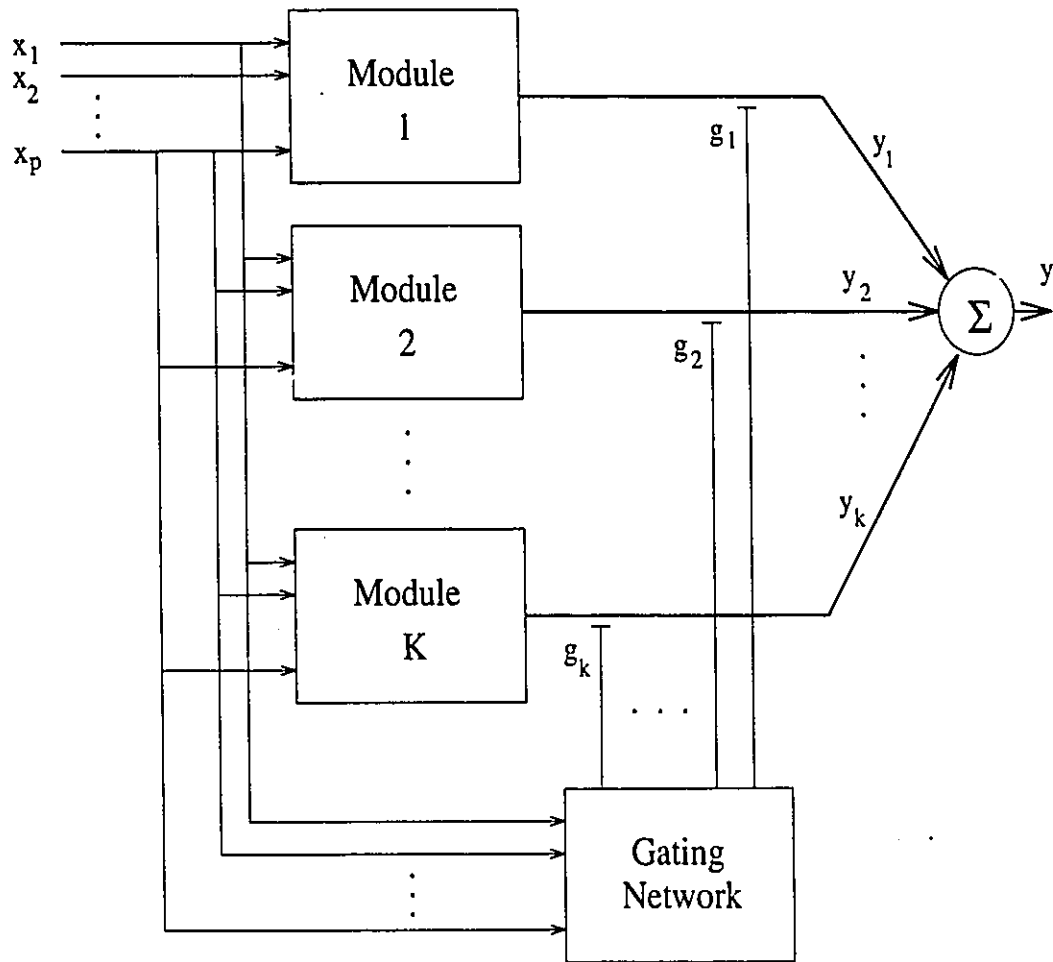


Figure 8.1: Block diagram of a modular network; the outputs of the expert networks are mediated by a gating network.

where $i = 1, \dots, K$, $F_i(\mathbf{x})$ is a deterministic, vector-valued function of the input vector \mathbf{x} , and ε_i is a random vector. For simplicity, it may be assumed that the random vector ε_i is Gaussian-distributed with zero mean and identity covariance matrix \mathbf{I} . To do the learning, the expert networks and the gating network of Figure 8.1 are all trained simultaneously.

For this purpose, an associative Gaussian mixture model is proposed by Jacobs and Jordan [46, 47]. In this model, the probability distribution of the desired response vector \mathbf{d} is treated as a linear combination of K different multivariate Gaussian distributions for a given input vector \mathbf{x} , that is

$$f(\mathbf{d}|\mathbf{x}) = \frac{1}{(2\pi)^{\frac{d}{2}}} \sum_{i=1}^K g_i \exp\left(-\frac{1}{2}\|\mathbf{d} - \mathbf{y}_i\|^2\right) \quad (8.3)$$

where $\|\cdot\|$ denotes the Euclidean norm of the enclosed vector; \mathbf{y}_i is the output vector of the i th expert network, and it is viewed as the conditional mean of a multivariate Gaussian distribution; g_i is the activation of the i th output neuron of the gating network, which is defined as follows [14]:

$$g_i = \frac{\exp(u_i)}{\sum_{j=1}^K \exp(u_j)} \quad (8.4)$$

where u_i is the output of the i th output neuron of the gating network.

Let \mathbf{w} denote the synaptic weights of all the expert networks arranged appropriately, and let vector \mathbf{g} denote the activations of all the output neurons in the gating network. The conditional probability density function $f(\mathbf{d}|\mathbf{x})$ of (8.3) can be viewed as a likelihood function, with the whole synaptic weight \mathbf{w} and the activation vector \mathbf{g} playing the roles of unknown parameters. Thus a log-likelihood function can be defined as follows:

$$l(\mathbf{w}, \mathbf{g}) = \ln \sum_{i=1}^K g_i \exp\left(-\frac{1}{2}\|\mathbf{d} - \mathbf{y}_i\|^2\right) \quad (8.5)$$

where it is understood that \mathbf{y}_i depends on \mathbf{w}_i , all weights of the i th expert network. Then $l(\mathbf{w}, \mathbf{g})$ can be viewed as an objective function, the maximization of which yields maximum-likelihood estimates of all the free parameters of the modular network in Figure 8.1, represented by the synaptic weights of the different expert networks and those of the gating network.

To assist in the formulation of the learning algorithm for the modular network of Figure 8.1, we may define the *a posteriori probability* associated with the output of the i th

expert network as [46]

$$h_i = \frac{g_i \exp(-\frac{1}{2}\|\mathbf{d} - \mathbf{y}_i\|^2)}{\sum_{i=1}^K g_i \exp(-\frac{1}{2}\|\mathbf{d} - \mathbf{y}_i\|^2)} \quad (8.6)$$

Differentiating (8.3) with respect to the output vector \mathbf{y}_i of the i th expert network, we get

$$\frac{\partial l}{\partial \mathbf{y}_i} = h_i(\mathbf{d} - \mathbf{y}_i), \quad i = 1, 2, \dots, K \quad (8.7)$$

Equation (8.7) states that, during the training process, the synaptic weights of the i th expert network are adjusted to correct the error between the output vector \mathbf{y}_i and the desired response vector \mathbf{d} , but in proportion to the *a posteriori* probability h_i that the i th expert network generated the training pattern in current use [46].

Suppose now that each expert network consists of a single layer of neurons. Let $\mathbf{y}_i^{(m)}$ denote the output vector of the m th neuron in the i th expert network and $\mathbf{w}_i^{(m)}$ denote the corresponding synaptic weight vector, then

$$\mathbf{y}_i^{(m)} = \mathbf{x}^T \mathbf{w}_i^{(m)} \quad (8.8)$$

To modify the synaptic weights of all expert networks, differentiating (8.5) with respect to the synaptic weights and using chain rule, we get

$$\frac{\partial l}{\partial \mathbf{w}_i^{(m)}} = \frac{\partial l}{\partial \mathbf{y}_i} \frac{\partial \mathbf{y}_i}{\partial \mathbf{w}_i^{(m)}} \quad (8.9)$$

Differentiating (8.8) with respect to the synaptic weights $\mathbf{w}_i^{(m)}$, and substituting the result and (8.7) into (8.9), we get

$$\frac{\partial l}{\partial \mathbf{w}_i^{(m)}} = h_i e_i^{(m)} \mathbf{x} \quad (8.10)$$

where $e_i^{(m)} = d^{(m)} - y_i^{(m)}$ is the error signal produced at the output of the m th neuron in the i th expert network.

To maximize the log-likelihood function l with respect to the synaptic weights of the different expert networks, we may use gradient ascent in weight space. In particular, we modify the synaptic weight vector $\mathbf{w}_i^{(m)}$ by applying a small adjustment $\Delta \mathbf{w}_i^{(m)}$, defined by

$$\Delta \mathbf{w}_i^{(m)} = \eta \frac{\partial l}{\partial \mathbf{w}_i^{(m)}} \quad (8.11)$$

where η is a small learning-rate parameter. Thus, using $w_i^{(m)}(n)$ to denote the value of the synaptic weight $w_i^{(m)}$ at iteration n of the learning algorithm, the updated value of this synaptic weight vector at iteration $n + 1$ is computed in accordance with the recursion

$$\begin{aligned} w_i^{(m)}(n+1) &= w_i^{(m)}(n) + \Delta w_i^{(m)}(n) \\ &= w_i^{(m)}(n) + \eta h_i^{(m)}(n) e_i^{(m)}(n) \mathbf{x} \end{aligned} \quad (8.12)$$

This is the desired recursive formula for adapting the expert networks of the modular architecture shown in Figure 8.1 [39].

To adapt the gating network, substituting (8.4) into (8.5), and rewrite the log-likelihood function as

$$l = \ln \sum_{i=1}^K \exp(u_i) \exp\left(-\frac{1}{2} \|\mathbf{d} - \mathbf{y}_i\|^2\right) - \ln \sum_{i=1}^K \exp(u_i) \quad (8.13)$$

The partial derivative of the log-likelihood function l with respect to the output u_i of the i th neuron in the gating network is therefore

$$\frac{\partial l}{\partial u_i} = h_i - g_i \quad (8.14)$$

Equation (8.14) states that the synaptic weights of the i th output neuron in the gating network are adjusted such that the activation of the network moves toward the corresponding *a posteriori* probability h_i .

Suppose again that the gating network consists of a single layer of neurons. Let $\mathbf{a}_i(n)$ be the value of the synaptic weight vector of the i th output neuron in the gating network at iteration n of the learning algorithm. Applying the same method described above, we may update the synaptic weight \mathbf{a}_i in accordance with the following

$$\mathbf{a}_i(n+1) = \mathbf{a}_i(n) + \eta(h_i(n) - g_i(n))\mathbf{x} \quad (8.15)$$

Note that in the preceding algorithms we suppose that the expert networks and the gating network are all single layer perceptrons. For the case of the expert networks and the gating network being multilayer perceptrons, the well-known back-propagation learning algorithm can be employed to upgrade the synaptic weights of each hidden layer.

8.3 Chaotic Modeling of Sea Clutter Using Modular Network

In Chapter 7, we demonstrated that the chaos-based noncoherent detector, which is based on a multilayer perceptron trained with noncoherent radar data, is more sensitive to receiver front-end noise than the chaos-based coherent detector, which is based on a multilayer perceptron trained with coherent radar data. To be of practical significance, the subsequent chaotic modeling of sea clutter using modular network is based on coherent radar data.

To design a modular network for the reconstruction of the chaotic dynamics of sea clutter with a wide variety of sea states, we have to divide the sea environment into a number of specific states, for each sea state we then train an expert network to capture the local dynamics. For this purpose, we choose four specific sea states, named state 1, state 2, state 3 and state 4, which correspond to wave heights of *3.78 ft*, *4.53 ft*, *5.95 ft* and *8.30 ft*, respectively. Thus, four expert networks are chosen. All have the structure as shown in Figure 6.30. That is, each expert network consists of 60 source nodes, in which 30 source nodes correspond to the I-channel components and other 30 source nodes correspond to the Q-channel components of the received radar signal. Correspondingly, the output layer of each expert network consists of two linear neurons, which provide predictions of the in-phase and quadrature components of the sea clutter for the corresponding sea state. As discussed in Chapter 7, the expert network has two hidden layers with 80 neurons in the first one and 55 neurons in the second one.

The gating network of the modular architecture shown in Figure 8.1 also has an input layer of 60 source nodes and an output layer of 4 linear neurons. Experimentation with the model of Figure 8.1 shows that we need a hidden layer for the gating network. A size of 80 neurons was found to be satisfactory for our application.

As before, the model of Figure 8.1 is trained with the mixed clutter data chosen from the database of Table 5.2, corresponding to sea state 1, state 2, state 3 and state 4, respectively. The synaptic weights corresponding to the output layers of all expert networks and the gating network are adapted using the algorithm described in the previous section. Other synaptic weights of the model are adapted using the back-propagation algorithm.

After the learning session is completed, the synaptic weights and bias terms of the four expert networks and the gating network are all fixed. The results of the recursive prediction performed by the network are shown in Figures 8.2 to 8.5. Each figure shows the result corresponding to clutter data from state 1, state 2, state 3 and state 4, respectively. In these figures, we have omitted the sequences of 30 complex-valued samples used to initialize the network. All the solid curves shown in these figures are the actual in-phase and quadrature component of the corresponding clutter data, and the dashed curves in these figures are the recursive prediction results.

To calculate the local prediction time of the modular network of Figure 8.1, based on the recursive prediction results, we define the average power σ_0 of sea clutter as follows:

$$\sigma_0 = \frac{1}{N} \sum_{i=1}^N \sqrt{y_I^2(i) + y_Q^2(i)} \quad (8.16)$$

where $y_I(i)$ is the in-phase component of the received signal and $y_Q(i)$ is the quadrature component of the received signal; both have zero mean. N is total number of the sea clutter data. Let $\hat{y}_I(i)$ be the recursive predicted in-phase component of the model, and $\hat{y}_Q(i)$ be the recursive predicted quadrature component of the model. The recursive prediction error of the model is defined in complex-valued form as follows:

$$\epsilon(i) = (y_I(i) - \hat{y}_I(i)) + j(y_Q(i) - \hat{y}_Q(i)) \quad (8.17)$$

where $j = \sqrt{-1}$. Thus we may define the average power of the error signal after the model operates L iterations as follows:

$$\sigma_\epsilon(L) = \frac{1}{L} \sum_{i=1}^L \sqrt{(y_I(i) - \hat{y}_I(i))^2 + (y_Q(i) - \hat{y}_Q(i))^2} \quad (8.18)$$

Then the recursive prediction time T_L is defined such that $\frac{\sigma_\epsilon(L)}{\sigma_0} = 10\%$. The subsequent generation performance analysis for the model is based on the calculation of the recursive prediction time.

The results shown in Figures 8.2 to 8.5 are produced by the modular network in which the gating network has a hidden layer of 80 neurons. To choose the best gating network that satisfies our application, three types of gating network structures were considered. In the first structure, the gating network has one hidden layer as described above.

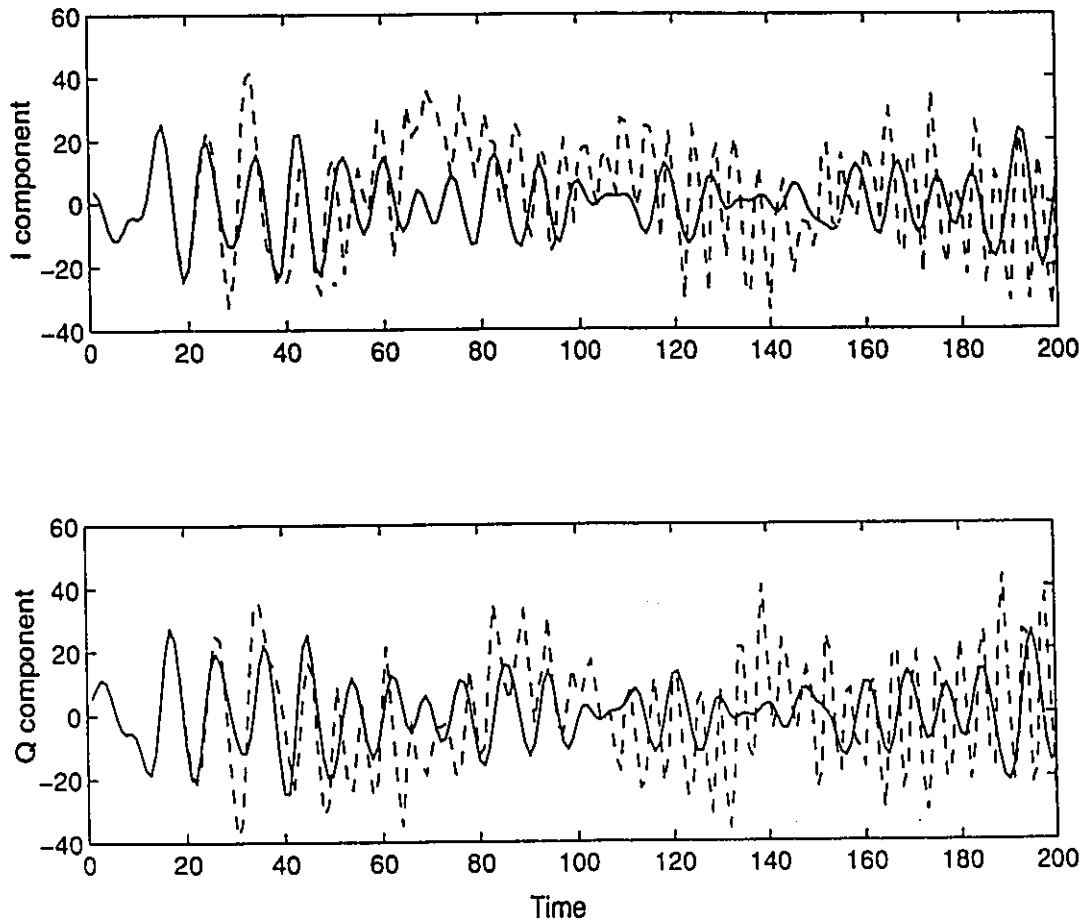


Figure 8.2: Recursive prediction result of the modular network shown in figure 8.1 for a testing data set chosen from a sea state with wave height of 3.78 ft. The solid curves refer to the original sea clutter waveforms, in which one is the in-phase component and the other is the quadrature component. The dashed curves refer to the recursive predicted waveforms, for which the first 30 points of the complex-valued sea clutter data (not shown in the figure) are used as the initial starting point.

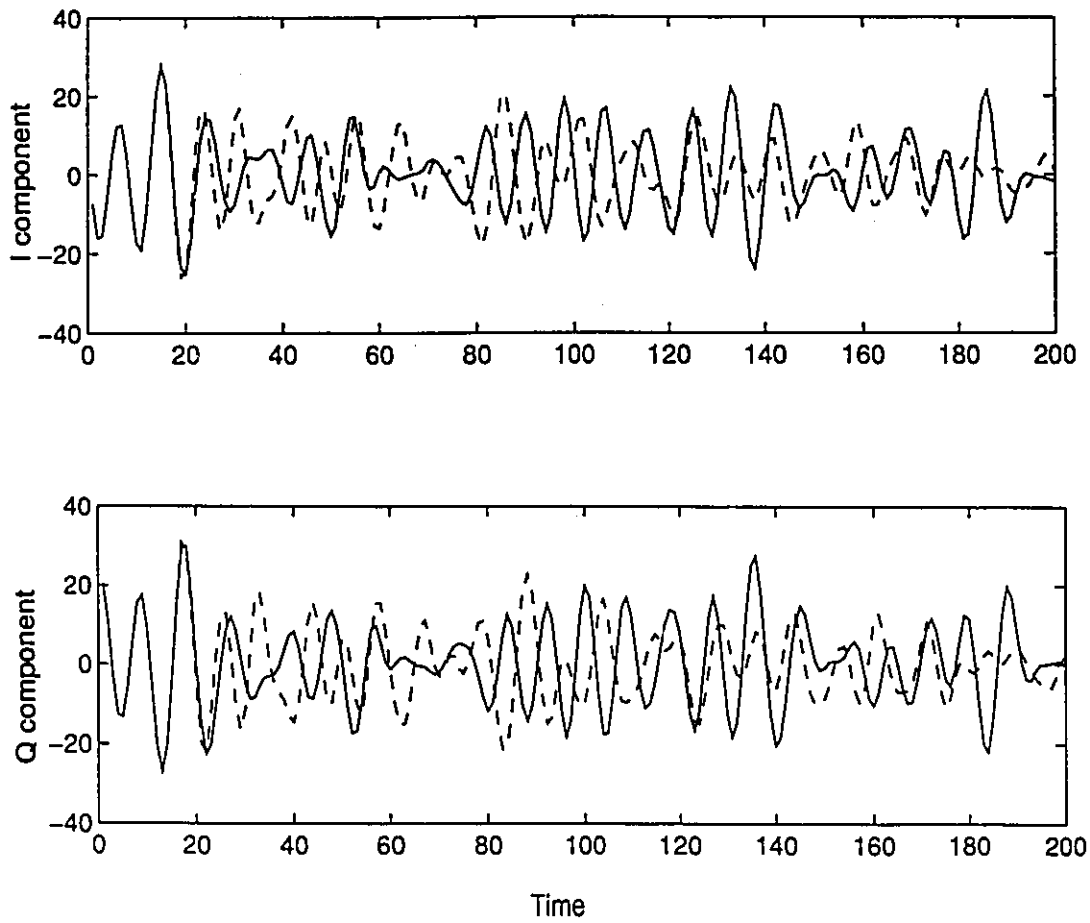


Figure 8.3: Recursive prediction result of the modular network shown in figure 8.1 for a testing data set chosen from a sea state with wave height of 4.54 *ft*. Traces are as defined in Figure 8.2.

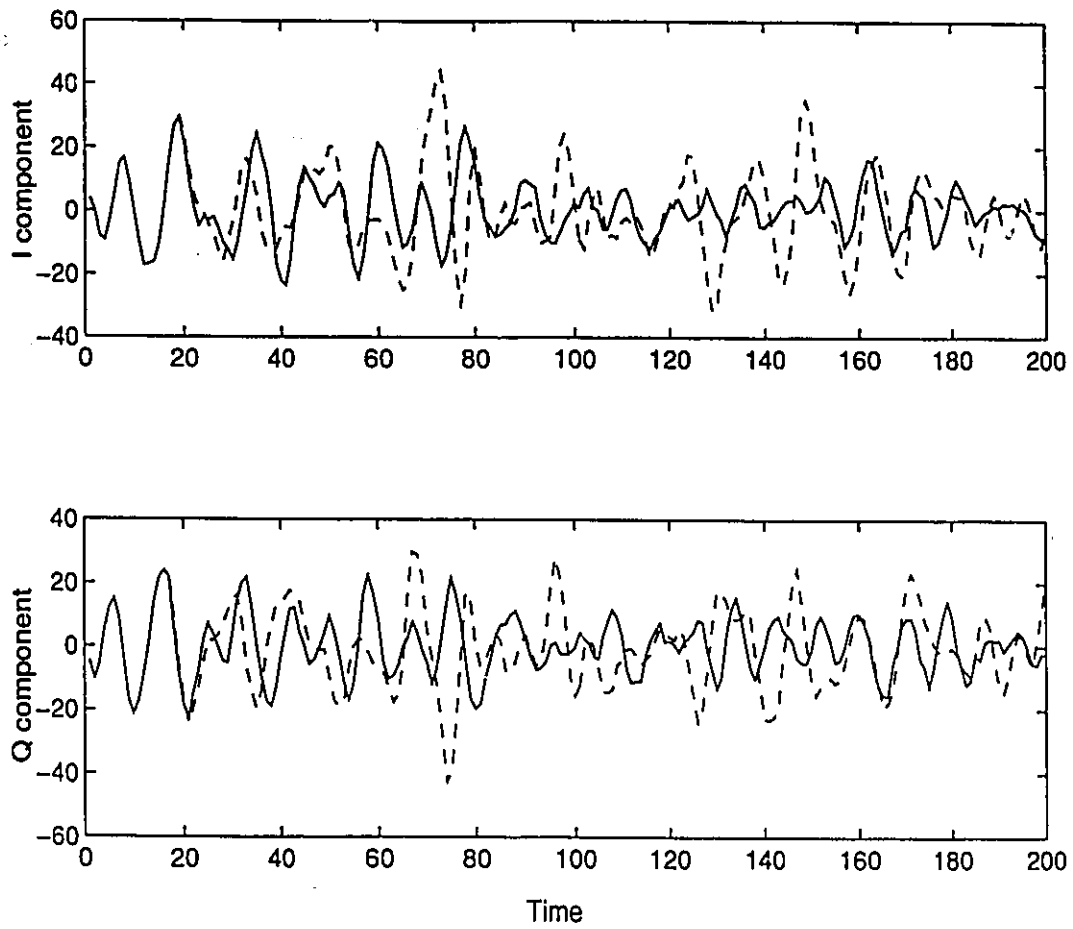


Figure 8.4: Recursive prediction result of the modular network shown in figure 8.1 for a testing data set chosen from a sea state with wave height of 5.95 *ft*. Traces are as defined in Figure 8.2.

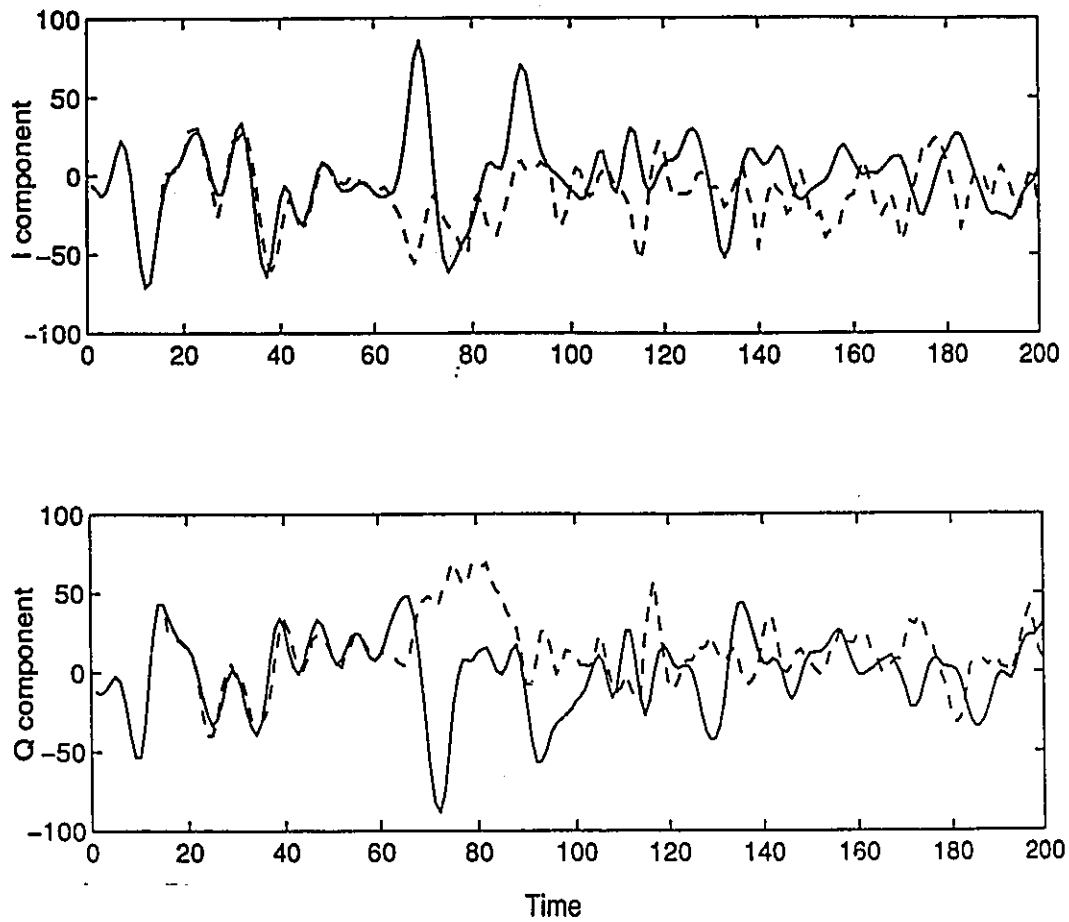


Figure 8.5: Recursive prediction result of the modular network shown in figure 8.1 for a testing data set chosen from a sea state with wave height of 8.30 *ft*. Traces are as defined in Figure 8.2.

Sea State (wave height in ft.)	State 1 3.78	State 2 4.54	State 3 5.95	State 4 8.30
Gating network without hidden layer	33 ± 5	34 ± 4	34 ± 5	34 ± 4
Gating network with one hidden layer	38 ± 4	39 ± 5	41 ± 4	40 ± 4
Gating network with two hidden layers	39 ± 5	39 ± 4	40 ± 4	40 ± 5

Table 8.1: Comparison of local predictability of different types of gating network structure in the modular network for sea clutter modeling

In the second structure the gating network has no hidden layer. And in the third structure the gating network has two hidden layers, with the first hidden layer having 80 neurons and the second hidden layer having 55 neurons. In all three cases, the expert networks have the same structure of Figure 6.30. Again, after the learning sessions are completed, the synaptic weights and bias terms of the models are all fixed. The generalization performance of the models is evaluated by comparing their recursive prediction time.

Table 8.1 shows the recursive prediction time of the three types of modular networks corresponding to three types of gating network structures just described. The first row in the table lists the sea states for which the testing data are chosen in the range of these states. The results shown in the second row to fourth row list the recursive prediction time referring to the gating network with no hidden layer, one hidden layer, and two hidden layers, respectively. There are two numbers for each recursive prediction time. The first number is the mean value of 10 estimates of the recursive prediction time computed for the particular sea state range, and the second number is the standard error (the standard deviation) in the estimation of the mean. It can be seen from Table 8.1 that the modular network in which the gating network has no hidden layer has the worst recursive prediction performance, while the prediction performance of the modular network whose gating network has two

Sea State (wave height in ft.)	State 1 3.78	State 2 4.54	State 3 5.95	State 4 8.30
Single Perceptron Network	33 ± 9	38 ± 8	51 ± 8	40 ± 7
Modular Neural Network	38 ± 4	39 ± 5	41 ± 4	40 ± 4

Table 8.2: Comparison of local predictability of a multilayer perceptron network with a modular neural network for sea clutter modeling

hidden layers is comparable to that of the modular network whose gating network has one hidden layer. Considering a gating network with two hidden layers is more complicated than a gating network with just one hidden layer, and the fact that it converges slower during the learning session, we recommend the use of a modular network whose gating network has one hidden layer for reconstruction of the chaotic dynamics of sea clutter.

Next we compare the generalization performance of the modular network to that of the multilayer perceptron of Figure 6.30. In this comparison, a modular network whose gating network has one hidden layer is used. It is trained by using a mixed clutter data selected from ranges of sea state 1, state 2, state 3, and state 4, respectively. The multilayer perceptron model of Figure 6.30 is trained by using a clutter data selected from the range of sea state 3. The computation results of the recursive prediction time of these models are shown in Table 8.2. Again, the first row in the table lists the sea states for which the testing data are chosen in the range of these sea states. The results listed in the second row and third row refer to the recursive prediction time for the multilayer perceptron model and the modular network, respectively. There are also two numbers for each recursive prediction time. The first number is the mean value of 10 estimates of the recursive prediction time computed for the particular sea state range, and the second number is the standard error in the estimation of the mean. From the results listed in Tables 8.2, the following observations are obtained:

- The multilayer perceptron model has the best generalization performance for the testing data where sea state is the same as the data used to trained it, and has an inferior generalization performance when the testing data are chosen from another sea state range.
- The modular network model has an average generalization performance that is consistent for testing data chosen from all sea state ranges.
- The standard error of the modular network in computing the recursive prediction time is smaller than that of the multilayer perceptron model for testing data drawn from different sea state ranges.

8.4 Summary

In this chapter, we have presented a modular network structure for reconstruction of the chaotic dynamics of sea clutter. An associative Gaussian mixture model was described. The modular network model was applied to do the reconstruction of the chaotic dynamics of sea clutter. The optimization of the gating network of the modular structure was discussed. Recursive prediction results are presented and compared with that of the multilayer perceptron model described in Chapter 6.

An important advantage of a modular network over a multilayer perceptron for the reconstruction of chaotic dynamics of sea clutter is that it is robust with respect to a wide variety of sea states, which could be exploited to provide a further improvement in detection performance. A shortcoming of this approach, however, is that the modular network, in which four multilayer perceptrons are employed as expert networks and one is used as the gating network, is more complicated than a multilayer perceptron model, and therefore converges much slower than the multilayer perceptron model during the training session. This shortcoming may be overcome by using parallel processing technology.

Chapter 9

Summary and Recommendations

In this thesis we have presented a detailed experimental study, using ground-truthed real-life radar data, which demonstrates that sea clutter is essentially a chaotic process. The experimental evidence for this assertion may be summarized as follows: Irrespective of whether we use the in-phase, quadrature, or amplitude (envelope) component, we may state that

- Sea clutter has a finite and fractal correlation dimension lying in the range of 7 to 9; an embedding dimension of 10 is confirmed by the false nearest neighbors analysis.
- The largest Lyapunov exponent of sea clutter is positive (lying in the range between 0.03 and 0.045, normalized with respect to the pulse-repetition frequency), indicating a strong dependence on initial conditions.
- Sea clutter is locally (short-time) predictable.

The experimental evidence presented in this thesis is considered to be complete in light of what we know about chaos theory. Indeed, it provides solid confirmation that sea clutter is chaotic, an observation that was first made in 1990 by Leung and Haykin [56].

Another important finding reported in the thesis is that we can exploit the prior information that sea clutter is chaotic by trading off highly sophisticated but relatively inexpensive software for fairly expensive microwave hardware. Specifically, a chaos-based receiver using amplitude information only (as in a noncoherent marine radar) is capable of

providing a detection performance comparable to that of a "conventional" Doppler-based receiver using both amplitude and phase information (as in a coherent marine radar). The chaos-based detector includes a nonlinear predictive model of sea clutter, the design of which is made possible by the use of a multilayer perceptron trained with the back-propagation algorithm. It is feasible, of course, to use other supervised neural network models, such as a radial basis function network for the predictor. In the case of a coherent radar, further improvement in the performance of the chaos-based detector is possible by virtue of the fact that the receiver preserves the Doppler information about the target signal. This information is not available to a noncoherent radar.

The important point that we wish to stress in this thesis is that chaos theory provides a new approach for the detection of signals in "noise" that is chaotic in nature. Indeed, much can be gained by revisiting signal detection problems, where the noise component has traditionally been described by a stochastic model, and exploring whether the use of a chaotic model is a better choice by virtue of concrete physical arguments or experimental evaluations.

An issue that has to be further studied is the computation of the Lyapunov spectrum and the Kolmogorov entropy for sea clutter. As results reported in Chapter 6 indicate that there may be only one positive Lyapunov exponent in sea clutter, and the second Lyapunov exponent is likely to be zero, it implies that sea clutter may be governed by coupled differential equations. The results, however, is still inconclusive because of the limitation of data length and the sampling rate of the radar system used to collect the sea clutter data. To get a more conclusive result on the Lyapunov spectrum and the Kolmogorov entropy, larger sea clutter data set is needed, and the radar system is suggested to be modified such that a much higher (*i.e.*, at least 5 times higher) sampling rate is achieved.

Another issue that needs to be considered is that of antenna scanning. The experimental results presented in this thesis have focused on the antenna usually staring at a patch of the ocean environment along a particular direction. In an operational marine radar, the antenna operates in a scanning mode, thereby introducing scanning modulation into the received signal. The key issue that needs to be examined here is how the inherent chaotic nature of sea clutter is affected by the antenna scanning modulation. Ordinarily,

this modulation process is deterministic, and we therefore expect to see the sea clutter process remaining chaotic. However, as the antenna rotates, radar backscatter is collected from different parts of the ocean surface, making the sea clutter component of the received signal assume a spatio-temporal character. By *spatio-temporal chaos*, we understand a deterministic nonlinear dynamical system that exhibits temporal chaos at different "lattice" sites, and that has a nontrivial incoherent structure in space [64].

Appendix

The Lorenz system is described by the following coupled differential equations:

$$\begin{aligned}\frac{dx_1(t)}{dt} &= \sigma(x_2(t) - x_1(t)) \\ \frac{dx_2(t)}{dt} &= -x_1(t)x_3(t) + rx_1(t) - x_2(t) \\ \frac{dx_3(t)}{dt} &= x_1(t)x_2(t) - bx_3(t)\end{aligned}$$

where we take $\sigma = 16$, $b = 4$, and $r = 45.92$. For these parameters the accepted value for the fractal dimension D is 2.05, and the values for the Lyapunov exponents λ_1 , λ_2 , and λ_3 are 1.50, 0.00, and -22.5, respectively.

In the following table we present a summary of the values of these parameters using the software that were actually employed for the characterization of sea clutter reported in Chapter 6. A data set of 50000 samples was used in the calculations. The results of this table are in close agreement with the above mentioned (theoretical) values for the Lorenz system. This should therefore establish confidence in the experimental results reported in Chapter 6.

Sampling time Δt	0.01
Time delay τ	0.10
Correlation dimension	2.09
Minimum embedding dimension	3
The largest Lyapunov exponent by Wolf's algorithm	$\lambda_1 = 1.482$
Lyapunov spectrum based on linear approach	$\lambda_1 = 1.450$
	$\lambda_2 = -0.002$
	$\lambda_3 = -14.988$
Lyapunov spectrum based on 3rd-order polynomial fit	$\lambda_1 = 1.510$
	$\lambda_2 = 0.003$
	$\lambda_3 = -22.300$

Bibliography

- [1] Abarbanel, H., R. Katz, J. Cembrola, T. Galib, and T. Frison, "Nonlinear analysis of high reynolds number flow over a buoyant axisymmetric body", *Physical Review D*, Vol. 49, No. 5, pp. 4003-4018, May 1994.
- [2] Abarbanel, H., R. Katz, J. Cembrola, T. Galib, and T. Frison, "High reynolds number boundary layer chaos", *Physical Review Letters*, Vol. 72, No. 15, pp. 2383-2386, April 1994.
- [3] Abarbanel, H., R. Brown, J. Sidorowich, and L. Tsimring, "The analysis of observed chaotic data in physical systems", *Reviews of Modern physics*, Vol. 65, No. 4, pp. 1331-1392, 1993.
- [4] Abarbanel, H. and M. B. Kennel, "Local false nearest neighbors and dynamical dimensions from observed chaotic data", *Physical Review E*, Vol. 47, No. 5, pp. 3057-3068, May 1993.
- [5] Abarbanel, H., R. Brown, and Matthew Kennel, "Variation of Lyapunov exponents on a strange attractor", *Journal of Nonlinear Science*, Vol. 1, pp. 175-199, 1991.
- [6] Abarbanel, H., R. Brown, and Matthew Kennel, "Local Lyapunov exponents computed from observed data", *Journal of Nonlinear Science*, Vol. 2, pp. 343-365, Sept. 1992.
- [7] Abarbanel, H. and M. Sushchik, "True Lyapunov exponents and models of chaotic data", *Int. J. of Bifurcation and Chaos*, Vol. 3, pp. 534-550, 1993.
- [8] Abarbanel, H., R. Brown and J. Kadtko, "Prediction and system identification in chaotic nonlinear systems: Time series with broadband spectra", *Physics Letters A*, Vol. 138, No. 8, pp. 401-408, 1989.

- [9] Abraham, N. B., A. M. Albano, B. Das, G. DeGuzman, S. Yong, R. S. Gioggia, G. P. Puccioni and J. R. Tredicce, "Calculating the dimension of attractors from small data sets", *Physics Letters*, Vol. 114A, No. 5, pp. 217-221, 1986.
- [10] Badii, R., G. Broggi, B. Derighetti, M. Ravani, S. Ciliberto, A. Politi and M. A. Rubio, "Dimension increase in filtered chaotic signals", *Physical Review Letters*, Vol. 60, No. 11, pp. 979-982, 1988
- [11] Baker, C. J., "K-distributed coherent sea clutter", *IEE Proceedings*, Part F, Vol. 138, No. 2, pp. 89-92, 1991.
- [12] Benettin, G., L. Galgani, A. Giorgilli and J. Strelcyn, "Lyapunov characteristic exponents for smooth dynamical systems and for Hamiltonian systems; A method for computing all of them. Part I, II", *Meccanica*, Vol. 15, No. 1, pp. 9-30, 1980.
- [13] Brown, R., P. Bryant, and H. Abarbanel, "Computing the Lyapunov exponents of a dynamical system from observed time series", *Physical Review A*, Vol. 43, pp. 2787-2806, March 1991.
- [14] Bridle, J. S., "Probabilistic interpretation of feedforward classification network outputs, with relationships to statistical pattern recognition", in *Neuro-computing: Algorithms, Architectures and Applications*, edited by F. Fogelman-Soulie and J. Hertz, NY: Springer-Verlag, pp. 227-236, 1990.
- [15] Brock, W. A., W. D. Dechert and J. A. Scheinkman, "A test for independence based on the correlation dimension", Working Paper, Dept. of Economics, University of Wisconsin, Wisconsin, 1986.
- [16] Casdagli, M., "Nonlinear prediction of chaotic time series", *Physica D*, Vol. 35, pp. 335-356, 1989.
- [17] Chester, D. L., "Why two hidden layers are better than one", *Proc. IJCNN-90*, pp. 265-268, 1990.
- [18] Clarke, T. L., "Generalization of neural networks to the complex plane", *Proc. IJCNN-90*, Vol. 2, pp. 435-440, San Diego, 1990.

- [19] Crutchfield, J. P. and K. Young, "Computation at the onset of chaos", in *Entropy, complexity, and the physics of information*, edited by Y. C. Lee, World Scientific, Singapore, 1988.
- [20] Eckmann, J. P., and D. Ruelle, "Ergodic theory of chaos and strange attractors", *Reviews of Modern Physics*, Vol. 57, No. 3, pp. 617-656, July 1985.
- [21] Eckmann, J. P., and D. Ruelle, "Fundamental limitations for estimating dimensions and Lyapunov exponents in dynamical systems", *Physica D* 56, pp.185-1187, 1992.
- [22] Farina, A. and F. A. Studer, "A review of CFAR detection techniques in radar systems", *Microwave Journal*, pp. 115-129, Sept. 1986.
- [23] Farmer, D., E. Ott and J. Yorke, "The dimension of chaotic attractor", *Physica* 7D, pp. 153-180, 1983.
- [24] Farmer, J. D. and J. J. Sidorowich, "Predicting chaotic time series", *Physical Review Letters*, Vol. 59, No. 8, pp. 845-848, August 1987.
- [25] Fraser, A. M. and H. Swinney, "Independent coordinates for strange attractors" *Phys. Rev. A* 33, pp. 1134-1140, 1986.
- [26] Fraser, A. M. and A. Dimitriadis, "Forecasting probability densities by using hidden Markov models with mixed states", in *Time series prediction: Forecasting the future and understanding the past*, edited by A. S. Weigend and N. A. Gershenfeld, pp. 265-282, Addison Wesley, 1994
- [27] Fraser, A. M., "Information and entropy in strange attractors", *IEEE Trans. on Information Theory*, Vol. 35, No. 2, pp. 245-262, March 1989.
- [28] Fraser, A. M., "Using mutual information to estimate metric entropy" in *Dimensions and entropies in chaotic systems*, edited by G. Mayer-Kress, Springer-Verlag, 1986.
- [29] Fraser, A. M., "Modeling nonlinear time series", *Proc. ICASSP-92*, Vol. 5, pp. 313-316, San Francisco, 1992.
- [30] Georgiou, G. M. and C. Koutsougeras, "Complex domain backpropagation", *IEEE Trans. Circuits and Systems Part II: Analog and Digital Signal Processing*, Vol. 39, May, 1992.

- [31] Glass, L., M. R. Guevara and A. Shrier, "Bifurcation and chaos in a periodically stimulated cardiac oscillator", *Physica* 7D, pp. 89-101, 1983.
- [32] Goldstein, H., "Sea echo in propagation of short radio waves", MIT Radiation Lab. Ser., Vol. 13, Set 6.6, edited by D.E. Kerr, McGraw-Hill, New York, 1951.
- [33] Grassberger, P., "Finite sample corrections to entropy and dimension estimates", *Physics Letters A*, Vol. 128, pp. 369-373, 1988.
- [34] Grassberger, P. and I. Procaccia, "Measuring the strangeness of strange attractor", *Physica* 9D, pp. 189-208, 1983.
- [35] Grassberger, P., R. Hegger, H. Kantz, C. Schaffrath and T. Schreiber, "On noise reduction methods for chaotic data", *Chaos* 3 (2), pp. 127-141, 1993.
- [36] Guckenheimer, J. and P. Holmes, *Nonlinear oscillations, dynamical systems, and bifurcations of vector fields*, Applied Math. Sci. Vol. 42, Springer-Verlag, 1986.
- [37] Hausdorff, F. "Dimension und äußeres Maß", *Mathematische Annalen*, Vol. 79, pp. 157, 1919.
- [38] Haykin, S., "Chaotic signal processing: New research directions and novel applications", IEEE Workshop on SSAP, Victoria, BC, Oct. 1992.
- [39] Haykin, S., *Neural networks: A comprehensive foundation*, Macmillan College Publishing Company, New York, 1994.
- [40] Haykin, S. and A. Ukrainec, "Neural networks for adaptive signal processing", in *Adaptive system identification and signal processing algorithms*, eds. N. Kalouptsidis and S. Theodoridis, Prentice Hall, pp. 512-553, 1993
- [41] Haykin, S., C. Krasnor, T. Nohara, B. Currie, and D. Hamburger, "A coherent dual-polarized radar for studying the ocean environment ", *IEEE Trans. Geoscience and Remote Sensing*, Vol. 29, pp. 189-191, 1991.
- [42] Haykin, S. and X. Li, "Detection of signals in chaos", *Proceedings of the IEEE*, Vol. 83, No. 1, pp. 95-122, 1995.

- [43] He, X. and A. Lapedes, "Nonlinear modeling and prediction by successive approximation using radial basis functions", *Physica D* 70, pp. 289-301, 1993.
- [44] Holmes, P., *New approaches to nonlinear problems in dynamics*, SIAM, 1980.
- [45] Hornik, K., M. Stinchcombe and H. White, "Multilayer feedforward networks are universal approximators", *IEEE Trans. Neural Networks*, Vol. 2, pp. 359-366, 1989.
- [46] Jacobs, R. A. and M. I. Jordan, "A competitive modular connectionist architecture", In *Advances in neural information processing systems 3*, eds. R. P. Lippmann, J. E. Moody, and D. J. Touretzky, pp. 767-773, 1991.
- [47] Jacobs, R. A. and M. I. Jordan, "Learning piecewise control strategies in a modular neural network architecture" *IEEE Trans. Systems, Man, and Cybernetics*, Vol. 23, pp. 337-345, 1993.
- [48] Jakeman, E., "On the statistics of K -distributed noise", *J. Phys. A: Math. Gen.*, Vol. 13, pp. 31-48, 1980.
- [49] Jakeman, E. and P.N. Pusey, "A model for non-Rayleigh sea echo", *IEEE Trans. Antennas and Propagation*, Vol. AP-24, No. 6, 1976.
- [50] Johnson, N. I., *Continuous univariate distributions*, Houghton Mifflin Company, Boston, 1970.
- [51] Kaplan, D. T. and L. Glass, "Coarse-grained embedding of time series: random walks, Gaussian random processes, and deterministic chaos", *Physica D* 64, pp. 431-454, 1993.
- [52] Kaplan, D. T. and L. Glass, "Direct test for determinism in a time series", *Phy. Rev. Let.*, Vol. 68, No. 4, pp. 427-430, 1992.
- [53] Kaplan, J. and J. Yorke, "Functional differential equations and the approximation of fixed points", *Lecture Notes in Math. 730*, H. O. Peitgen and H. O. Walther, eds., Springer, Berlin, p. 228, 1978.
- [54] Kennel, M. B., R. Brown, and H. Abarbanel, "Determining embedding dimension for phase-space reconstruction using a geometrical construction", *Physical Review A*, Vol. 45, No. 6, pp. 3403-3411, March 1992.

- [55] Leung, H., "Chaotic modeling of sea clutter by neural network", Ph.D. thesis, McMaster University, 1991.
- [56] Leung, H., and S. Haykin, "Is there a radar clutter attractor?", *Applied Physics Letters*, Vol. 56, No. 6, pp. 592-595, 1990.
- [57] Li, X. and S. Haykin, "A new pseudo-noise generator for spread spectrum communications", *Proc. ICASSP-95*, Vol. 5, pp. 3603-3606, Detroit, Michigan, May 1995.
- [58] Li, X. and S. Haykin, "Chaotic characterization of sea clutter", *l'Onde Electrique, Special issue on radar*, SEE, France, pp. 60-65, Mar. 1994.
- [59] Li, X. and S. Haykin, "Chaotic detection of small targets in sea clutter", *Proc. ICASSP-93*, Vol. 1, pp. 237-240, Minneapolis, Minnesota, 1993.
- [60] Lorenz, E.N., "Deterministic nonperiodic flow", *Journal of Atmospheric Science*, Vol. 20, pp. 130-141, 1963.
- [61] Lyapunov, A. M., *Problème Général de la Stabilité du Mouvement* (translated from Russian), *Ann. Fac. Science, Université Toulouse* 9, pp. 203-475, 1907. Reproduced in *Ann. Math. Study*, Vol. 17, Princeton, 1947.
- [62] Mainieri, R., "On the equality of Hausdorff and box counting dimensions", *Chaos* 3 (2), pp. 119-125, 1993.
- [63] Mayer-Kress, G., "Application of dimension algorithms to experimental chaos", in *Directions in chaos*, edited by Hao Bai-Lin, World Scientific, Singapore, 1987.
- [64] Mayer-Kress, G. and K. Kaneko, "Spatiotemporal chaos and noise", *Journal of Statistical Physics*, Vol. 54, Nos. 5/6, pp. 1489-1508, 1989.
- [65] Myers, C., S. Kay and M. Richard, "Signal separation for nonlinear dynamical systems", *Proc. ICASSP-92*, Vol. 4, pp. 129-132, San Francisco, California, 1992.
- [66] Michel, O. and P. Flandrin, "An investigation of chaos-oriented dimensionality algorithm applied to AR(1) process". *Proc. ICASSP-92*, Vol. 5, pp. 317-320, San Francisco, 1992.

- [67] Minkler, G. and J. Minkler, *CFAR: The principles of automatic radar detection in clutter*, Magellan Book Company, 1990.
- [68] Newhouse, S., "Generic properties of conservative systems", in *Chaotic behavior of deterministic systems*, edited by G. Iooss *et al.*, Les Houches, XXXVI, pp. 443-452, 1981.
- [69] Newhouse, S., "Entropy and smooth dynamics, dynamical systems and chaos", Lecture notes in Physics Vol. 179, pp. 165-180, Springer-Verlag, 1983.
- [70] Newhouse, S., "Understanding chaotic dynamics" in *Chaos in Nonlinear Dynamical Systems*, edited by J. Chandra, SIAM, 1984.
- [71] Oppenheim, A. V., G. W. Wornell, S. H. Isabelle and K. M. Cuomo, "Signal processing in the context of chaotic signals", Proc. ICASSP-92, Vol. 4, pp. 117-120, San Francisco, 1992.
- [72] Osborne, A. R. and A. Provenzale, "Finite correlation dimension for stochastic systems with power-law spectra", *Physica D* 35, pp. 357-381, 1989
- [73] Oseledec, V. I., "A multiplicative ergodic theorem. Lyapunov characteristic numbers for dynamical systems", *Moscow Math. Soc.* Vol. 19, p. 197, 1968.
- [74] Ott, Edward, *Chaos in Dynamical Systems*, Cambridge University Press, 1993
- [75] Parker, T. S. and L. O. Chua, "Chaos: a tutorial for engineers", *Proceedings of the IEEE*, Vol. 75, No. 8, pp. 982-1008, Aug. 1987.
- [76] Packard, N., J. Crutchfield, J. Farmer, and R. Shaw, "Geometry from a time series", *Physical Review Letters*, Vol. 45, No. 9, pp. 712-716, 1980.
- [77] Passamante, A., T. Hediger and M. Gollub, "Fractal dimension and local intrinsic dimension", *Physical Review A*, Vol. 39, No. 7, pp. 3640-3645, 1989.
- [78] Pincus, S. M., "Approximate entropy as a measure of system complexity", *Proc. Natl. Acad. Sci. USA*, Vol. 88, pp. 2297-2301, Mar. 1991.
- [79] Pineda, F.J., and J.C. Sommerer, "A fast algorithm for estimating generalized dimensions and choosing time delays" in *Time series prediction: Forecasting the future and*

- understanding the past*, edited by A. S. Weigend and N. A. Gershenfeld, pp. 367-385, Addison Wesley, 1994.
- [80] Poggio, T. and F. Girosi, "Networks for approximation and learning", Proceedings of the IEEE, Vol. 78, No. 9, pp. 1481-1497, 1990.
- [81] Raghunathan, M. S., "A proof of Oseledec's multiplicative ergodic theorem", Israel Journal of Mathematics, Vol. 32, p. 356, 1979.
- [82] Ruelle, D., *Elements of differentiable dynamics and bifurcation theory*, Academic Press, New York, 1989.
- [83] Ruelle, D. and F. Takens, "On the nature of turbulence", Comm. Math. Phys., Vol. 20, pp. 167-192, 1971.
- [84] Rumelhart, D.E. and J.L. McClelland, Eds., *Parallel distributed processing*, Cambridge, MA, M.I.T. Press, 1986.
- [85] Schouten, J. C. and C. M. van den Bleek, *RRCHAOS: a Menu-Driven Software Package for Chaotic Time Series Analysis*, Reactor Research Foundation, Delft, The Netherlands, 1994.
- [86] Schouten, J. C., F. Takens and C. M. van den Bleek, "Maximum-likelihood estimation of the entropy of an attractor", Physical Review E, Vol. 49, No. 1, pp. 126-129, 1994.
- [87] Schouten, J. C., F. Takens and C. M. van den Bleek, "Estimation of the dimension of a noisy attractor", Physical Review E, Vol. 50, No. 3, pp. 1851-1861, 1994.
- [88] Schuster, H. G., *Deterministic Chaos: An Introduction*, p. 117, VCH, Weinheim, Germany, 1988.
- [89] Sidorowich, J. J., "Modeling of chaotic time series for prediction, interpolation, and smoothing", Proc. ICASSP-92, Vol. 4, pp. 121-124, San Francisco, California, 1992.
- [90] Stoop, R. and P. Meier, "Evaluation of Lyapunov exponents and scaling functions from time series", J. Opt. Soc. Am. B, Vol. 5, No. 5, pp. 1037-1045, 1988.
- [91] Stoop, R. and J. Parisi, "Calculation of Lyapunov exponents avoiding spurious elements", Physica D, Vol. 50, pp. 89-94, 1991.

- [92] Sutton, R. S., "Two problems with back-propagation and other steepest-descent learning procedures for networks", Proc. 8th Ann. Conf. of the Cognitive Science Society, pp. 823-831, Hillsdale, New Jersey, 1986.
- [93] Takens, F., "Detecting strange attractor in turbulence", in *Dynamical Systems and Turbulence, Warwick 1980*, edited by D.A. Rand and L.S. Young, Lecture Notes in Mathematics, Vol. 898, pp. 366-381, 1981.
- [94] Takens, F., "On the numerical determination of the dimension of an attractor", in *Dynamical systems and bifurcations*, Lecture Notes in Mathematics, Vol. 1125, pp. 99, Springer-Verlag, 1985.
- [95] Theiler, J., "Estimating the fractal dimension of chaotic time series", The Lincoln Laboratory Journal, Vol. 3, No. 1, pp. 63-86, 1990.
- [96] Theiler, J., "Statistical precision of dimension estimators", Physical Review A, Vol. 41, No. 6, pp. 3038-3051, 1990.
- [97] Van Trees, H.L., *Detection, estimation, and modulation theory*, Part I, Wiley, New York, 1968.
- [98] Theiler, J., S. Eubank, A. Longtin, B. Galdrikian, and J. D. Farmer, "Testing for nonlinearity in time series: the methods of surrogate data", Physica D 58, pp. 77-94, 1992.
- [99] Ward, K. D., C. J. Baker, and S. Watts, "Maritime surveillance radar Part 1: Radar scattering from the ocean surface", IEE Proceedings, Vol. 137, No. 2, pp. 51-62, 1990.
- [100] Watts, S., C. J. Baker, and K. D. Ward, "Maritime surveillance radar Part 2: Detection performance prediction in sea clutter", IEE Proceedings, Vol. 137, No. 2, pp. 63-72, 1990.
- [101] Wolf, A., J. B. Swift, H. L. Swinney and J. A. Vastano, "Determining Lyapunov exponents from a time series", Physica 16D, pp. 285-317, 1985.
- [102] Wright, J. and R. L. Schult, "Recognition and classification of nonlinear chaotic signals", Chaos 3 (3), pp. 295-304, 1993.

