

ALGORITHM DEVELOPMENT FOR MODELING AND ESTIMATION
PROBLEMS IN HUMAN EEG ANALYSIS

by

ANATOLY ZLOTNIK

Submitted in partial fulfillment of the requirements
For the degree of Master of Science

Thesis Advisor: Dr. Kenneth Loparo

Department of Electrical Engineering and Computer Science
CASE WESTERN RESERVE UNIVERSITY

June 2006

CASE WESTERN RESERVE UNIVERSITY

SCHOOL OF GRADUATE STUDIES

We hereby approve the thesis of

ANATOLY ZLOTNIK

candidate for the Master of Science degree *

Committee Chair: _____

Dr. Kenneth Loparo

Thesis Advisor

Professor,

Department of Electrical Engineering & Computer Science

Committee: _____

Dr. Wojbor Woyczynski

Professor,

Department of Statistics

Committee: _____

Dr. Narasingarao Sreenath

Associate Professor,

Department of Electrical Engineering & Computer Science

Committee: _____

Dr. Marc Buchner

Associate Professor,

Department of Electrical Engineering & Computer Science

May 2006

*We also certify that written approval has been obtained for any
proprietary material contained therein.

Copyright © 2006 by Anatoly Zlotnik
All rights reserved

Table of Contents

Table of Contents	iv
List of Tables	vii
List of Figures	viii
Acknowledgements	x
Abstract	xi
1 Introduction	1
1.1 History of EEG Analysis	2
1.2 Modern EEG Analysis	2
1.2.1 Methods	2
1.2.2 Applications	3
1.2.3 Recent Developments	3
1.3 Research Focus	4
1.3.1 EEG structure	4
1.3.2 Sleep State Identification	5
1.3.3 Neurological Development in Neonates	5
1.3.4 Seizure Onset Detection	5
1.3.5 EEG Data Sets	6
1.4 Organization of Thesis	6
2 Stable and Fractional Processes	8
2.1 Stable Random Variables	9
2.1.1 Definitions	9
2.1.2 Properties	10
2.1.3 Dependence on Parameters	11
2.1.4 Generating Pseudo-Random Samples	13
2.1.5 Approximation of Stable Densities	14
2.1.6 Statistical Parameter Estimation	17
2.1.7 Iterative Numerical Estimation	19
2.1.8 Testing Goodness of Fit	21
2.1.9 Accuracy and Time of Estimation	23
2.2 Stable Stochastic Processes	26
2.2.1 Self-Similar Processes	27

2.2.2	Stable Noises and Motions	28
2.2.3	Simulation of Fractional Stable Noises	31
2.3	Discussion and Future Work	34
3	Nonlinear Time-Series Methods	36
3.1	Spectral Analysis of Stochastic Signals	37
3.1.1	Correlation and Spectral Density	38
3.1.2	Spectral Measures	39
3.2	Dimension and Self-Affinity Measures	41
3.2.1	Fractal Dimension	42
3.2.2	Hurst Exponent	46
3.3	Phase-Space Reconstruction	47
3.3.1	Reconstruction Time Delay	50
3.3.2	Embedding Dimension	52
3.4	Phase-Space Measures	55
3.4.1	Testing for Determinism	55
3.4.2	Correlation Dimension	63
3.4.3	Kolmogorov Entropy	66
3.4.4	Lyapunov Exponents	68
3.5	Discussion and Future Work	69
4	State Detection and Identification in Physiological Signals	71
4.1	Review of Detection and Identification Methodology for Physiological Signals	72
4.2	Change-Point Detection	74
4.2.1	A-Posteriori Change Point Problem	75
4.2.2	Statistical Methods	77
4.2.3	Description of Algorithm	79
4.3	Segmentation and Clustering	82
4.3.1	Feature Extraction and Segmentation	82
4.3.2	Clustering and State Identification	84
4.4	Discussion and Future Work	86
5	Analysis of EEG Structure	88
5.1	Introduction	89
5.2	Data Sets and Assumptions	90
5.3	Linear Statistical Analysis	94
5.3.1	Correlation and Power Spectrum	95
5.3.2	Long Memory and Non-Stationarity	98
5.3.3	Distribution of Signal and Increments	101
5.3.4	Self-Similarity Properties	105
5.4	The Nonlinearity Question	108
5.5	A Model for the Normal EEG	111

5.5.1	Normal EEG as Fractional Stable Noise	112
5.5.2	Self-Organization Theory of Neuronal Activity	114
5.6	Discussion and Future Work	116
6	Automatic Sleep-State Identification in Neonates	118
6.1	Initial Feature Selection	120
6.2	Sleep State Scoring Procedure	123
6.3	Calibration of Algorithm	125
6.4	Quantification of Brain Maturation	128
6.5	Discussion and Future Work	129
7	Automatic Epileptic Seizure Onset Detection	131
7.1	Recent Computational Epilepsy Analysis	132
7.2	Quantification of Epileptiform Activity	134
7.3	Clinical Applications and Early Seizure Onset Detection . . .	139
7.4	Discussion and Future Work	140
8	Conclusions and Recommendations for Future Work	142
8.1	Methods	143
8.1.1	Stable and Fractional Processes	143
8.1.2	Nonlinear Time-Series Analysis	143
8.1.3	Segmentation and Clustering	144
8.2	Applications	145
8.2.1	EEG Analysis and Modeling	145
8.2.2	Sleep-State Detection	145
8.2.3	Epileptic Seizure Detection	146
A	List of MATLAB Codes	148
A.1	α -Stable Distributions	148
A.2	Stable and Fractional Processes	149
A.3	Time-Series Analysis	149
A.3.1	Spectral Analysis	149
A.3.2	Fractal Analysis	149
A.3.3	Attractor Reconstruction	149
A.3.4	Phase-Space Measures	150
A.4	Change-Point Detection	150
	Bibliography	151

List of Tables

3.1	Comparison of Fractal Dimension Methods	46
6.1	Features for Neonatal Sleep-State Detection	123

List of Figures

2.1	α -stable distributions	12
2.2	‘A’ vs. ‘M’ stable parametrization	13
2.3	Statistical estimation of Simulated α -stable Samples for $\alpha = 1.5$	18
2.4	Statistical estimation of Simulated α -stable Samples for $\alpha = 0.7$	18
2.5	MLE for Simulated α -stable Samples, $\alpha = 1.7$	24
2.6	MLE for Simulated α -stable Samples, $\alpha = 0.7$	25
2.7	Kernels for Linear Fractional Stable Noise	30
2.8	Balanced Linear Fractional Stable Noise	32
2.9	Balanced Compound Linear Fractional Stable Noise	33
2.10	Long Memory Property of Filtered LFSN	33
3.1	Power Spectrum of LFSN, $H=0.2$, $\alpha = 1.8$	41
3.2	Fractal Dimension of Brownian Motion	45
3.3	Lorenz and Rossler Attractors	49
3.4	Independent Delay Coordinates	52
3.5	Embedding Dimension from False Nearest Neighbors	54
3.6	Surrogate Data for Lorenz Time-Series	58
3.7	Linear Properties of Lorenz Data and Surrogate	58
3.8	Nonlinear Properties of Lorenz Data and Surrogate	61
3.9	Surrogate Data for Integrated LFSN	61
3.10	Nonlinear Properties of Integrated LFSN Data and Surrogate	62
3.11	Automatic Diagnostics for Correlation Dimension	66
3.12	Automatic Diagnostics for Kolmogorov Entropy	67
4.1	Example of Change-Point Detection	81
4.2	Example of Change-Point Detection in Subtle Case	81
4.3	Example of Change-Point Detection in Distribution	82
5.1	Epoch of Cortical EEG Data	91
5.2	International 10-20 Electrode System	92
5.3	Epoch of Neonatal Scalp EEG Data	92
5.4	Local Autocorrelation of Cortical EEG Data	95
5.5	Long-Range Autocorrelation of Cortical EEG Data	96

5.6	Change in Autocorrelation of Cortical EEG Data	96
5.7	Change in Autocorrelation of Neonatal Scalp EEG Data	97
5.8	Log of Power Spectrum of Cortical EEG	97
5.9	Log of Power Spectrum of Cortical EEG over Time	97
5.10	Autocorrelation of Large Epoch of Cortical EEG Data	98
5.11	Self-Affinity Estimators for Cortex and Scalp EEG Data	99
5.12	Cortical EEG epoch and 2-Second Moving Average	100
5.13	Density of EEG Amplitude and α -stable fit	102
5.14	Autocorrelation of Cortex and Scalp EEG	103
5.15	Density of EEG Increments and α -stable fit	103
5.16	α -Stable Parameter Estimates for EEG Increments	104
5.17	Scale parameter σ of Cortex EEG Increments with Time Step	106
5.18	Scale parameter σ of Scalp EEG Increments with Time Step .	107
5.19	Box-plots of Empirical CTM Criterion for EEG Data	109
5.20	Mutual Information of EEG data and Surrogates	109
5.21	Box-plots: Ratio of Mutual Information of EEG to that of Surrogates	110
5.22	Cortex EEG and CLFSN Signals	112
5.23	Autocorrelation and Mutual Information of EEG and CLFSN	113
5.24	Shape Parameter α of Increment Distribution for EEG and CLFSN	113
5.25	Scale Parameter σ of Increment Distribution for EEG and CLFSN	114
6.1	Box-Plots of EEG Measures for Neonatal Sleep States	121
6.2	Box-Plots of Band Power for Neonatal Sleep States	122
6.3	Neonatal Sleep-State Identification Procedure	124
6.4	Distribution of Cohen's Kappa and % Agreement for each Channel	126
6.5	Distribution of Cohen's Kappa and % Agreement for Feature Combinations	127
6.6	Sleep-State Transitions per Hour using Center Channel	128
7.1	Empirical S-measure and CTM for Ictal EEG	136
7.2	CTM and FD Measures During Seizures	137
7.3	CTM and FD Measures During Seizures, All Channels	138

ACKNOWLEDGEMENTS

I would like to thank Professor Ken Loparo for his guidance and support of this research project. I thank Professor Wojbor Woyczynski for introducing me to statistics, and for his encouragement and advice. I owe many thanks to Dr. Alexandra Piryatinska. I much appreciate the opportunity to collaborate with her on the problem of truncated Lévy flights and on change-point detection. Farhad Kaffashi and I spent many hours discussing nonlinear time-series analysis. I thank him for the insight I gained from our collaboration. I thank Dr. Vladimir Uchaikin for a solid introduction to α -stable distributions, and I also thank Dr. Gyorgy Terdik for introducing me to the theory of stochastic processes. I thank Dr. Michiel Hochstenbach for giving me a solid foundation in numerical analysis. I also thank Professor Jiayang Sun for introducing me to statistical models. I would like to express my gratitude to Professors Mark Scher and Mary Ann Werz, who provided the EEG data without which this research would not be possible. I would like as well to thank Dr. Narasingarao Sreenath for kindly agreeing to serve as a member of the defense committee.

ALGORITHM DEVELOPMENT FOR MODELING AND ESTIMATION PROBLEMS IN HUMAN EEG ANALYSIS

Abstract

by

Anatoly Zlotnik

The structure of human electroencephalographic (EEG) signals is investigated using computational methods with the objective of developing automatic algorithms for clinical applications. The EEG is examined globally as a stochastic process with alpha-stable increments, and a novel parameter estimation method is used to investigate its properties. The signal is also examined locally as a smooth, deterministic dynamical system, and the performance and applicability of methods for nonlinear feature estimation are evaluated. The results are compared to standard linear and spectral methods. A set of inter-cranial recordings from epileptic patients is used to study the local behavior of the EEG, and algorithms for detection and forecasting of seizures are considered. In addition, a set of sleep-EEG recordings from a study on neonatal development is examined. An algorithm for sleep-state identification in neonates is presented, and the assessment of brain maturation in neonates using the EEG is explored.

Chapter 1

Introduction

The electroencephalogram (EEG) records the electrical activity of the brain, and is ubiquitous in neurological research and clinical applications. Electrical potential is measured between pairs of electrodes, which can be placed on the scalp or surgically implanted to measure the EEG signal. It is known that this electrical activity is caused by the firing of single neurons, while electrode measurements reflect the combined activity of millions of cells in a large-scale multi-unit network. The result is a signal with components that vary greatly in amplitude, frequency spectrum, recurrence, and duration.

The structural complexity of this system presents issues regarding its dynamics involving memory, chaos, periodicity, and scaling. Conventional methods for time-series and dynamic systems analysis are based on assumptions of determinism or randomness, and linearity or nonlinearity, while the EEG evades such classification. Clinicians consequently analyze the EEG visually, and such analysis remains an art practiced by experts, although much research has been focused on developing computational methods to make certain examination faster or even automatic. The central theme of this work is an investigation of EEG signal structure on different time-scales, since understanding this behavior is vital for the effective application of time-series analysis methods to create consistent algorithms useful for clinical use.

1.1 History of EEG Analysis

The measurement and analysis of electrical activity in the brain has a long history beginning in the 19th century. The first recordings were made by Vladimir Pravdich-Neminsky [109] in 1912 and the term “Elektroenkephalogramm” was first used by Hans Berger [11], the discoverer of the human EEG, in 1929. A detailed account of historical aspects of the EEG is given in [94].

Great progress in the clinical interpretation of the EEG was made in the next 30 years, after which interest shifted towards automatic analysis, focusing largely on power spectrum estimation using the Fast Fourier Transform and statistical signal analysis. Pioneering work was done by Hjörth [59], who presented several metrics based on moments of the power spectral density that could be computed in real time. The results of Elul [36] in examining the Gaussianity of short EEG segments have been of special interest, and Sugimoto [130] tested the normality and stationarity of EEG segments in different sleep states. Most of the work in this area has focused on the adult population.

The limitations of frequency and time-domain techniques quickly became apparent. It is difficult to cite exact measures for many phenomena in the EEG, such as spikes or sharp waves, let alone automate them. The EEG is too complex a signal to be fully characterized by frequency analysis, and advances in nonlinear time-series methods and computer power in the 80s and 90s have caused a surge in computational EEG research.

1.2 Modern EEG Analysis

1.2.1 Methods

Established methods for computational analysis of the EEG are generally classified as nonparametric or parametric. The former can involve amplitude, interval, or period distributions, correlation functions, power spectrum, or

cross-spectrum functions. The latter involves AR or ARMA models, Kalman filtering, or segmentation analysis. Other methods include mimetic analysis, time-frequency analysis, and topographic analysis [94]. In general, the aim of computational analysis is to quantify a specific characteristic of a single- or multi-channel EEG. The process of mapping a time-series to a real number or a vector is referred to as *feature extraction*, and the derived *features* can then be used for qualitative characterization, automatic diagnostics, prediction algorithms, and other analysis.

1.2.2 Applications

The applications of automatic EEG analysis include detecting abnormality in short records, classifying abnormalities, and evaluating changes between serial records or trends over long records. Any automatic algorithm must make some decision based on diagnostics or features of the signal, which requires heuristics such as neural networks or clustering. It is important to choose the correct method for any given application. Research in EEG analysis focuses both on such heuristics as well as identification of useful signal features for specific problems. The focus of this work is on the problems of sleep-state identification in the polysomnogram (PSG), or sleep-EEG [121], and seizure onset detection [63].

1.2.3 Recent Developments

Much research has been done to determine if the EEG has elements of dynamical chaos, and whether attractor reconstruction methods can be used to develop useful diagnostics. Some works [114, 115, 43, 42, 136, 106, 90, 122, 141, 142] have focused on the dimensionality of EEG signals obtained from correlation integrals [53], with mixed results [56]. Largest Lyapunov exponents of the EEG have also been compared between sleep states [115, 147] and used in seizure prediction algorithms [64, 62]. Time-delay embedding has also been used in algorithms to quantify nonlinearity [67, 100, 113], com-

plexity [68], and similarity [80].

Wavelet and other time-frequency methods have been applied to several problems [101, 25, 125, 129, 144]. The use of heuristics, both to extract features and to make decisions, has been examined during sleep [45, 138, 76, 55] and for seizure detection [28, 41, 47, 75, 69]. Several segmentation methods have been applied to the EEG [14, 18, 73], and some time-scaling and fractal properties of EEG signals have been observed [81, 106] and noted as a useful diagnostics [38, 2].

It is important to note here that fully automated clinical EEG diagnosis is far from a reality. Computer methods are of great aid to researchers and practitioners, however any result must still be subjectively interpreted by an expert. Furthermore, no unified theory exists about the structure of the EEG as a long-range process. The theory of self-organized criticality [7, 8], which focuses on self-similarity properties, has been proposed [81]. In general, works focusing on distributional properties of, or developing stochastic models for, the EEG have been absent from the literature recently. This is the motivation for much of the work presented here.

1.3 Research Focus

1.3.1 EEG structure

Both nonlinear and stochastic behaviors have been observed in the EEG and used to discriminate states [43], though it is unclear which type of process is dominant [100, 136]. The goals are then to identify and quantify both the nonlinearity and randomness of the process. Nonlinearity assumes an underlying chaotic attractor [72], which can be quantified through attractor reconstruction by the parameters of dimensionality and Lyapunov exponents [78]. Randomness can be described by an underlying stochastic process with increments from one or more probability distributions [44].

In this work evidence is presented that the EEG is globally a stochastic

process with correlated increments from a symmetric α -stable distribution. Variation of the parameters of this distribution are examined over different time-scales, as well as other properties, including measures of determinism, chaoticity, and dimension. This sheds light on the appropriate analysis to use given observations on any particular time-scale.

1.3.2 Sleep State Identification

A novel algorithm, also to appear in [108], is presented for sleep state identification in neonates based on the methods in [18, 73]. A number of standard and novel feature extraction methods are used to create diagnostic sequences from polysomnographic recordings. The diagnostic sequences are segmented by nonparametric change-point detection and clustered to create minute-by-minute scores of neonatal sleep-EEG records. Optimal combinations of features are chosen based on the correspondence of resulting scores to those provided by a clinical neurophysiologist. The goals and materials are similar to [121].

1.3.3 Neurological Development in Neonates

Polysomnographic recordings of premature and full-term neonates are compared using the above time-series analysis methods. The results of past research on the neonatal polysomnogram [122, 120, 119] suggest that level of brain maturation of the neonate can be assessed by analysis of sleep cycles from the PSG. Sleep-cycle patterns automatically extracted from the PSG are compared among pre-term and full-term groups, indicating a statistically significant difference.

1.3.4 Seizure Onset Detection

EEG time-series from intra-cranial recordings of epileptic patients are examined to determine which features can be used to consistently detect seizure onset. Epochs from pre-ictal, ictal (containing seizure activity), inter-ictal

(between seizures), and post-ictal states are examined. Statistical tests are formulated to verify hypotheses about the qualitative transitions in the signal that occur during the onset of an epileptic seizure. The features that best reflect these transitions are then used to establish appropriate onset criteria and to develop an automatic algorithm for real-time seizure onset detection. Clinical application of the algorithm is discussed, as well as the possibility of predicting epileptic seizures.

1.3.5 EEG Data Sets

The data used for investigation includes a set of 27 intra-cranial 98-channel EEG recordings sampled at 1000 Hz from a study of 8 epileptic patients. This data set was provided by Dr. Mary Ann Werz of the Department of Neurology at the Case School of Medicine. Another set used is a set of 116 14-channel sleep-EEG recordings sampled at 64 Hz from a study of neonatal development. A comprehensive database with demographic information on the neonates is available. The data and scoring were provided by Dr. Mark Scher of the Department of Pediatrics at the Case School of Medicine.

1.4 Organization of Thesis

The thesis is structured to review the mathematical background for the methods used, to build a basis for their application, and eventually to describe possible clinical implementations. The second chapter provides an overview of stable distributions, and presents a set of novel algorithms realizing their practical use. Stable fractional processes and their simulation are also discussed. In the third chapter spectral and nonlinear measures are reviewed, as well as statistical testing for determinism, the process of attractor reconstruction, and issues related to dynamical measures and their estimation. The fourth chapter deals with segmentation, clustering, and state identification methods for analysis of physiological signals, and the problem of

statistical change-point estimation.

In the fifth chapter, the results from chapters two and three are used to analyze the structure of the EEG alongside well-known examples of dynamical systems and stochastic processes, respectively. The sixth chapter deals with the process of sleep-state identification using polysomnographic recordings. The features described in chapter three are used as diagnostics, and the methods described in chapter four are used to develop an automatic sleep-state scoring algorithm. The application of the results to the assessment of brain maturation in neonates is discussed. In the seventh chapter, qualitative observations are stated about the process of epileptic seizure onset, and statistical hypothesis testing is used to verify these statements quantitatively. A method for automatic seizure onset detection based on those observations is presented. The eighth chapter summarizes results, and includes conclusions, suggestions for clinical applications, and recommendations for future work.

Chapter 2

Stable and Fractional Processes

The theory of stable distributions and Lévy processes began in the 1920s with Paul Lévy and Alexander Khintchine, and was extended by Kolmogorov and Gnedenko [51] in the 1940's. The focus is on limit distributions of sums of independent random variables (r.v.s), generalized by the class of α -stable distributions. Such distributions do not have finite moments of order higher than parameter α , and in general do not have finite variance or an explicit form of the probability density function (p.d.f.). The stable distribution, along with the notion of a fractional stochastic process, can be used to construct models with a rich variety of variabilities and time-scaling properties.

Non-Gaussian stable distributions and fractional processes have found applications in many fields. Notably, Mandelbrot [84, 85] and Fama [39] pioneered the use of fractional Brownian motions in the 1960s spurring interest in stable and fractional processes. Recent advances in computing power have allowed highly accurate approximation of stable densities [91, 33, 96, 97], and rapid simulation of stable processes. Applications currently include telecommunications [49, 124], meteorology, physics, hydrology, biology, ecology, finance and economics [86, 66], and connections have been established to other branches of mathematics. Reviews of Lévy processes and processes with long-range correlations are given in [9], and [112], respectively.

2.1 Stable Random Variables

Stable distributions are studied here with a perspective on efficient and effective application to analysis of real data sets in engineering and the sciences. Important definitions, non-trivial properties, dependence on parameters, and simulation are reviewed. A recently proposed method for numerical density approximation is optimized and implemented to design accurate and rapid algorithms for parameter estimation and goodness-of-fit tests using the MATLAB software package [88]. This provides a complete set of computational tools necessary for estimation and modeling with stable random variables.

2.1.1 Definitions

Several useful definitions and non-trivial properties [140, 117, 148] of stable distributions are listed below.

Definition 2.1 (Stable r.v.) *A random variable X is defined as stable if for any sequence $\{X_i\}_{i=1}^n$ of independent identically distributed (i.i.d.) copies of X , there exist constants a_n and b_n such that*

$$\sum_{i=1}^n X_i \stackrel{d}{=} a_n + b_n X, \quad (2.1)$$

where “ $\stackrel{d}{=}$ ” denotes equality in distribution. The case where $a_n = 0$ for all n is called strictly stable.

Definition 2.2 (Equivalent to Definition 2.1) *A random variable X is defined as stable if it has a domain of attraction, i.e. if there is a sequence $\{Y_i\}_{i=1}^n$ of i.i.d. random variables and a sequence of pairs $\{a_n, b_n\}$ of real numbers such that*

$$\frac{1}{b_n} \sum_{i=1}^n Y_i + a_n \stackrel{d}{\Rightarrow} X, \quad (2.2)$$

where “ $\stackrel{d}{\Rightarrow}$ ” denotes convergence in distribution.

Definition 2.3 (Zolotarev's 'A' Parametrization) *A random variable X is said to have a stable distribution if there are parameters $0 < \alpha \leq 2$, $-1 \leq \beta \leq 1$, $\sigma \in \mathbb{R}^+$, and $\mu \in \mathbb{R}$ such that its characteristic function (c.f.) has the form*

$$\phi(t) = E \exp(itX) = \begin{cases} \exp \left\{ -\sigma^\alpha |t|^\alpha \left[1 - i\beta \operatorname{sign}(t) \tan \frac{\pi\alpha}{2} \right] + i\mu t \right\}, & \alpha \neq 1, \\ \exp \left\{ -\sigma |t| \left[1 + i\beta \frac{2}{\pi} \operatorname{sign}(t) \ln |t| \right] + i\mu t \right\}, & \alpha = 1. \end{cases} \quad (2.3)$$

In Definition 2.3, $X \sim S_\alpha(\sigma, \beta, \mu)$, where α , β , σ , and μ are the shape, skewness, scale, and location parameters, respectively. Note that $S_2(\sigma, 0, \mu)$, $S_{1/2}(\sigma, 1, \mu)$, and $S_1(\sigma, 0, \mu)$ are the Gaussian, Lévy, and Cauchy distributions, respectively. The case where $\beta = 0$ and $\mu = 0$ is called symmetric α -stable ($S\alpha S$). $S\alpha S$ r.v.s are strictly stable.

2.1.2 Properties

Property 2.4 (Existence of Moments) *Given $X \sim S_\alpha(\sigma, \beta, \mu)$ with $0 < \alpha \leq 2$, then*

$$\begin{aligned} E|X|^p &< \infty \quad \text{for any } 0 < p < \alpha \\ E|X|^p &= \infty \quad \text{for any } p > \alpha \end{aligned} \quad (2.4)$$

Therefore the only stable distribution with finite variance is the Gaussian, for which $\alpha = 2$. Further references to cases where $\alpha < 1$ will be omitted, since the purpose here is to model measurements of real physical phenomena, which always have finite expectation.

Property 2.5 (Summation) *Given a sequence of r.v.s $\{X_i\}_{i=1}^n$ with $X_i \sim S_\alpha(\sigma_i, \beta_i, \mu_i)$, then $\sum_{i=1}^n X_i \sim S_\alpha(\sigma, \beta, \mu)$ has parameters*

$$\sigma = \left(\sum_{i=1}^n \sigma_i^\alpha \right)^{1/\alpha}, \quad \beta = \frac{\sum_{i=1}^n \beta_i \sigma_i^\alpha}{\sum_{i=1}^n \sigma_i^\alpha}, \quad \mu = \left(\sum_{i=1}^n \mu_i \right) \quad (2.5)$$

Property 2.6 (Scaling) *Given a r.v. $X \sim S_\alpha(\sigma, \beta, \mu)$,*

$$\begin{aligned} aX &\sim S_\alpha(|a|\sigma, \text{sign}(a)\beta, a\mu) & \text{if } \alpha \neq 1 \\ aX &\sim S_1(|a|\sigma, \text{sign}(a)\beta, a\mu - \frac{2}{\pi}a \ln |a|\sigma\beta) & \text{if } \alpha = 1 \end{aligned} \quad (2.6)$$

Corollary 2.7 *Given a sequence of i.i.d. r.v.s $\{X_i\}_{i=1}^n$ with $X \sim S_\alpha(\sigma, \beta, \mu)$, then from properties 2.5 and 2.6 it follows that*

$$\begin{aligned} (\sum_{i=1}^n X_i) &\stackrel{d}{=} n^{1/\alpha} X_1 + \mu(n - n^{1/\alpha}) & \text{if } \alpha \neq 1 \quad \text{and} \\ (\sum_{i=1}^n X_i) &\stackrel{d}{=} nX_1 + \frac{2}{\pi}\sigma\beta \ln(n) & \text{if } \alpha = 1 \end{aligned} \quad (2.7)$$

Property 2.8 (Limiting Tail Behavior) *Given a r.v. $X \sim S_\alpha(\sigma, \beta, \mu)$, then*

$$\begin{cases} \lim_{\lambda \rightarrow \infty} \lambda^\alpha P\{X > \lambda\} &= C_\alpha \frac{1+\beta}{2} \sigma^\alpha, \\ \lim_{\lambda \rightarrow \infty} \lambda^\alpha P\{X < -\lambda\} &= C_\alpha \frac{1-\beta}{2} \sigma^\alpha, \end{cases} \quad (2.8)$$

where

$$C_\alpha = \begin{cases} \frac{1-\alpha}{\Gamma(2-\alpha) \cos(\pi\alpha/2)} & \text{if } \alpha \neq 1 \\ 2/\pi & \text{if } \alpha = 1 \end{cases} \quad (2.9)$$

Therefore the limiting tail probability $z_\alpha(x)$ of an $S\alpha S$ distribution decays according to $z_\alpha(x) \propto 1/x^\alpha$.

2.1.3 Dependence on Parameters

From definition 2.3, for a stable r.v. $X \sim S_\alpha(\sigma, \beta, \mu)$ the parameters, α , β , σ , and μ (shape, skewness, scale, and location) completely specify the distribution. The dependence on μ and σ is quite straightforward. The dependence of the p.d.f and c.d.f. on α , and the power-law behavior of the tails are illustrated in figure 2.1, along with the dependence of the p.d.f. on β for values of $\alpha = 0.8$ to $\alpha = 1.2$ in the ‘A’ parametrization. Note the power-law heavy-tail behavior due to α , as compared to the Gaussian case ($\alpha = 2$), as well as the offset of the mode caused by β .

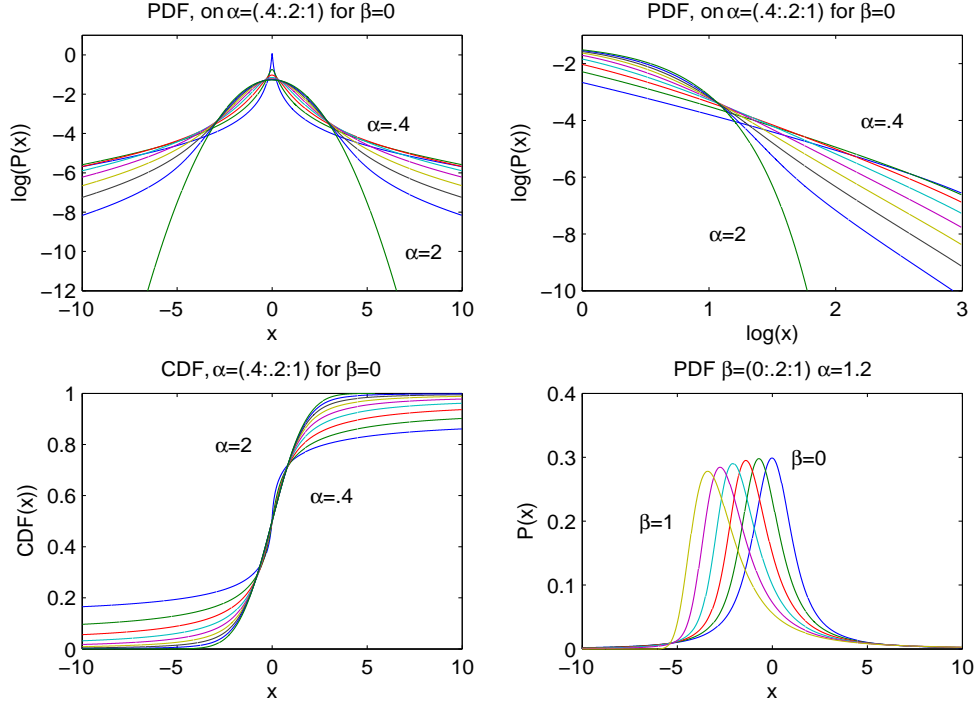


Figure 2.1: Dependence on Parameters

Another representation of stable distributions is the ‘M’ parametrization, in which the stable distribution transitions from Gaussian to Cauchy to Levy through continuous changes in the parameters α and β .

Definition 2.9 (Zolotarev’s ‘M’ Parametrization) *Another parametrization of stable r.v. $X \sim S_{\alpha}^0(\sigma, \beta^0, \mu^0)$ via the c.f. is*

$$\phi^0(t) = \begin{cases} \exp \left\{ -\sigma^{\alpha} |t|^{\alpha} (1 + i\beta \text{sign}(t) \tan \frac{\pi\alpha}{2} [(\sigma|t|)^{1-\alpha} - 1]) + i\mu_0 t \right\}, & \alpha \neq 1, \\ \exp \left\{ -\sigma |t| (1 + i\beta \frac{2}{\pi} \text{sign}(t) \ln(\sigma|t|) + i\mu_0 t) \right\}, & \alpha = 1. \end{cases} \quad (2.10)$$

The location parameters μ^0 of the ‘M’ parametrization and μ of the ‘A’ parametrization are related by $\mu = \mu^0 - \beta\sigma \tan \frac{\pi\alpha}{2}$ for $\alpha \neq 1$ and $\mu = \mu^0 - \beta\sigma \frac{2}{\pi} \ln(\sigma)$ for $\alpha = 1$.

The ‘M’ parametrization is considered more suitable to numerical compu-

tations because the mode always remains at the origin when $\mu_0 = 0$, unlike that for the ‘A’ parametrization. The continuous dependence on α and β is demonstrated in figure 2.2 below.

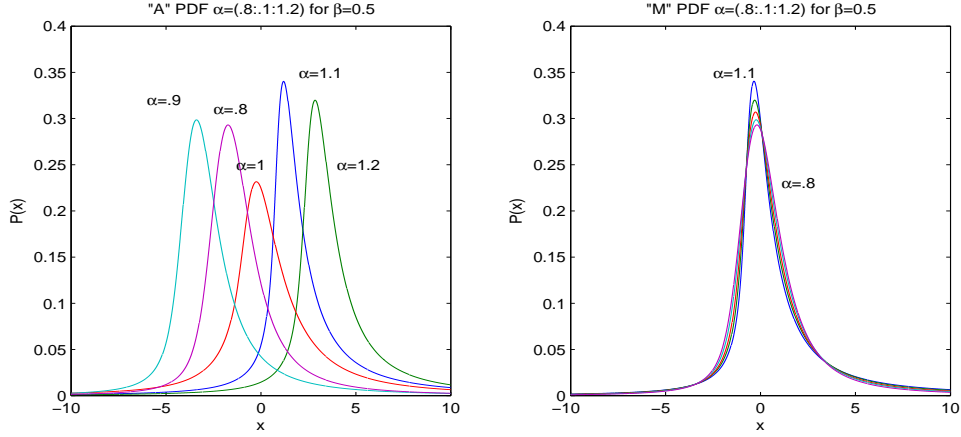


Figure 2.2: Dependence on Parametrization: ‘A’ (left), and ‘M’ (right)

2.1.4 Generating Pseudo-Random Samples

The most widely used method for generating samples from stable distributions is the Chambers-Mallows-Stuck method [21]. It is regarded as the fastest and most accurate [16], and reduces to the Box-Muller method for generating Gaussian samples [65]. The algorithm is quite straightforward:

Algorithm 2.10 (Chambers-Mallows-Stuck Method) *To generate a r.v. $X \sim S_\alpha(\sigma, \beta, \mu)$,*

1. *Generate a uniform r.v. V on $(-\pi/2, \pi/2)$ and an exponential r.v. W with mean $\mu = 1$*
2. *Compute the factors $C_{\alpha,\beta} = \frac{1}{\alpha} \arctan(\beta \tan(\frac{\pi\alpha}{2}))$ and $D_{\alpha,\beta,\sigma} = \sigma [\cos(\arctan(\beta \tan(\frac{\pi\alpha}{2})))^{-1/\alpha}$*
3.
$$Y = \begin{cases} D_{\alpha,\beta,\sigma} \frac{\sin(\alpha(V+C_{\alpha,\beta}))}{(\cos(V))^{1/\alpha}} (W^{-1} \cos(V - \alpha(V + C_{\alpha,\beta})))^{\frac{(1-\alpha)}{\alpha}}, & \alpha \neq 1 \\ \sigma \frac{2}{\pi} \left[\left(\frac{2}{\pi} + \beta V \right) \tan(V) - \beta \log \left(\frac{\frac{2}{\pi} W \cos(V)}{\frac{2}{\pi} + \beta V} \right) \right], & \alpha = 1 \end{cases}$$
4. *Evaluate $X = Y + \mu$*

2.1.5 Approximation of Stable Densities

A major obstacle to the use of stable distributions for modeling real data is the lack of an explicit form for the probability density function $f(x)$, which is central to most estimation and design techniques. One approach is to approximate the density of $S_\alpha(\sigma, \beta, \mu)$ by numerically evaluating the inverse Fourier transform of the c.f. $\phi(t)$ over a finite interval:

$$f(x) = \frac{1}{2\pi} \cdot \int_{-\infty}^{\infty} e^{-itx} \cdot \phi(t) dt \approx \frac{1}{2\pi} \cdot \int_{-a}^a e^{-itx} \cdot \phi(t) dt \quad (2.11)$$

Several methods based on the discrete Fourier transform (DFT) have been presented in [33, 96] for strictly stable r.v.s. The method implemented here is based on the midpoint integration rule of Menn and Rachev [91]:

$$\text{Midpoint Rule (MPR): } \int_a^b f(x) dx \approx f\left(\frac{b-a}{2}\right) \cdot (b-a) \quad (2.12)$$

The construction is as follows:

Algorithm 2.11 (FFT-based Density Approximation) *Given a c.f. $\phi(t)$, the corresponding p.d.f. approximation $\hat{f}(x)$ is given by:*

1. Define $N = 2^m \in \mathbb{N}$, $a \in \mathbb{R}^+$, and $h = \frac{2a}{N}$. Define a grid on $[-a, a]$ by $\{t_j = -a + jh\}_{j=0}^N$, with midpoints $\{t_j^* = \frac{1}{2}(t_j + t_{j+1})\}_{j=0}^{N-1}$.
2. Define vector $y \in \mathbb{C}^N$ by $\{y_j = (-1)^j \phi(t_j^*)\}_{j=0}^{N-1}$ and vector $C \in \mathbb{C}^N$ by $\{C_j = ih(-1)^j e^{-i\frac{\pi}{N}j}\}_{j=0}^{N-1}$.
3. Evaluate the DFT of y : $Y = \text{DFT}(y)$.
4. For the grid $\{x_k = -\frac{N\pi}{2a} + \frac{\pi}{a} \cdot k\}_{k=0}^{N-1}$, the values of the p.d.f. are given by $\hat{f}(x_k) = \frac{1}{2\pi} C_k \cdot Y_k$.

Lemma 2.12 (Proof of Algorithm 2.11) *Using the notation in algorithm 2.11, it follows that*

$$t_j^* \cdot x_k = \left(-a + \frac{h}{2} + \frac{2a}{N}j\right) \left(-\frac{N\pi}{2a} + \frac{\pi}{a}k\right) = \frac{N\pi}{2} - \pi k - \frac{\pi}{2} + \frac{\pi k}{N} - \pi j + \frac{2\pi k}{N}j$$

Applying the midpoint rule (MPR),

$$\begin{aligned}
\int_{-a}^a e^{-itx_k} \cdot \phi(t) dt &\stackrel{MPR}{\approx} \frac{1}{2\pi} \sum_{j=0}^{N-1} \phi(t_j^*) \cdot e^{-it_j^* x_k} \cdot (t_{j+1} - t_j) \\
&= \frac{2a}{N} \sum_{j=0}^{N-1} \phi(t_j^*) \cdot e^{-i\left(\frac{N\pi}{2} - \pi k - \frac{\pi}{2} + \frac{\pi k}{N} - \pi j + \frac{2\pi k}{N} j\right)} \\
&= \frac{2a}{N} \underbrace{e^{-i\frac{N\pi}{2}}}_{=1} e^{i\pi k} \underbrace{e^{i\frac{\pi}{2}}}_{=1} e^{-i\frac{\pi k}{N}} \sum_{j=0}^{N-1} \underbrace{e^{i\pi j}}_{=(-1)^j} \phi(t_j^*) \cdot e^{-i\frac{2\pi k}{N} j} \\
&= ih(-1)^k e^{-i\frac{\pi k}{N}} \cdot Y_k = C_k \cdot Y_k
\end{aligned} \tag{2.13}$$

The c.f. $\phi(t)$ is sampled on the grid $\{t_j\}_{j=0}^{N-1}$ on the domain $t \in [-a, a - \frac{2a}{N}]$, so algorithm 2.11 will approximate $f(x)$ on the grid $\{x_j\}_{j=0}^{N-1}$ on the domain $x \in [-c, c - \frac{\pi}{a}]$, where $c = N\pi/2a$. This follows from elementary properties of the Discrete Fourier Transform (see [35]). This concludes the proof.

Error analysis for the evaluation of standardized stable densities by algorithm 2.11 is discussed in [91]. Specifically, the errors $\epsilon_b(a, N)$, due to truncation of the domain of integration, $\epsilon_c(a, N)$, due to applying numerical integration, and $\epsilon_d(a, N)$, due to interpolation of the estimate, are addressed. These errors depend on the integration bound a , and the number of grid points, N . The factor ϵ_b will subsequently play a significant role, and is examined in additional detail here.

Lemma 2.13 (Integration Domain Truncation Error) *Given a tolerance ϕ_{tol} , a stable c.f. $\phi_X(t)$ for $X \sim S_\alpha(\sigma, \beta, \mu)$ satisfies $|\phi_X(t)| \leq \phi_{tol}$ for all $|t| \geq a$ where $a = \frac{1}{\sigma}(-\log(\phi_{tol}))^{1/\alpha}$. Furthermore, the truncation error $\epsilon_b(a, N)$ for algorithm 2.11 applied to X is bounded by $\frac{1}{\pi\alpha} \Gamma\left(\frac{1}{\alpha}, a^\alpha\right)$, where $\Gamma(a, z)$ is the upper-tail incomplete gamma function.*

Proof: Observe that

$$|\phi_X(t)| = \left| \exp\left\{-\sigma^\alpha |t|^\alpha + i\sigma^\alpha |t|^\alpha \beta \text{sign}(t) \tan \frac{\pi\alpha}{2}\right\} \right| = e^{-\sigma^\alpha |t|^\alpha} \tag{2.14}$$

Solving for a in $\phi_{tol} = e^{-\sigma^\alpha |t|^\alpha}$ gives $a = \frac{1}{\sigma}(-\log(\phi_{tol}))^{1/\alpha}$. Now, the error $\epsilon_b(a, N)$ can be bounded as follows. Setting $\omega = \tan(\frac{\pi\alpha}{2})$,

$$\begin{aligned}
|\epsilon_b| &= \frac{1}{2\pi} \left| \int_{-a}^{-\infty} \phi(t) \cdot e^{-itx} dt + \int_a^\infty \phi(t) \cdot e^{-itx} dt \right| \\
&= \frac{1}{2\pi} \left| \int_{-\infty}^{-a} e^{-\sigma^\alpha |t|^\alpha} e^{-i\sigma^\alpha (\beta |t|^\alpha \omega - |t|x)} dt + \int_a^\infty e^{-\sigma^\alpha t^\alpha} e^{i\sigma^\alpha (\beta |t|^\alpha \omega - |t|x)} dt \right| \\
&\leq \frac{1}{\pi} \int_a^\infty e^{-\sigma^\alpha t^\alpha} \underbrace{|\cos(\sigma^\alpha \beta |t|^\alpha \omega - |t|x)|}_{\leq 1} dt \\
&\leq \frac{1}{\pi} \int_a^\infty e^{-\sigma^\alpha t^\alpha} dt = \frac{1}{\pi\alpha} \Gamma\left(\frac{1}{\alpha}, a^\alpha\right) \quad (\leq e^a/\pi \text{ for } \alpha \geq 1)
\end{aligned} \tag{2.15}$$

The final result is obtained from Mathematica [146].

An analytical examination of the error $\epsilon_c(a, N)$ due to discretization of the integral is problematic, since the c.f. $\phi_X(t)$ oscillates for $\beta \neq 0$ in a nonlinear manner. Such error analysis has not been performed, and remains a major obstacle. Heuristically, it is clear that a greater value of N decreases $\epsilon_c(a, N)$. The error $\epsilon_d(a, N)$ only arises if values of $f_X(x)$ are required for x that do not coincide with the grid values. Spline interpolation can be used in that case with negligible error. In addition, because the heavy-tail behavior of stable distributions results in many ‘outliers’, the following series expansion, derived in [12], is useful for computing extreme probabilities:

Lemma 2.14 (Bergström Expansion) *For a standardized stable r.v. with $1 < \alpha \leq 2$, the p.d.f $f_{\alpha,\beta}(x)$ has an asymptotic series expansion for $x \rightarrow \infty$:*

$$f(x) = \frac{1}{\pi} \sum_{k=1}^{\infty} c \cdot a_k(cx)^{-k\alpha-1} \tag{2.16}$$

where

$$b = \frac{2}{\pi} \arctan(-\beta \tan(\pi\alpha/2)) \tag{2.17}$$

$$c = \cos(\pi b/2)^{1/\alpha} \tag{2.18}$$

$$a_k = \frac{(-1)^{k-1}}{k} \Gamma(1+k\alpha) \cdot \sin\left(\frac{k\pi}{2}(\alpha+b)\right) \tag{2.19}$$

2.1.6 Statistical Parameter Estimation

The estimation of parameters for samples with stable densities is a very well-studied problem. An excellent review of statistical estimators for this purpose can be found in [40]. One method that requires minimal computation is the log-moments estimator of Zolotarev [148]:

Algorithm 2.15 (Zolotarev Estimator) *Given a set of N i.i.d. samples $\{X_i\}_{i=1}^N$ from a centered r.v. $X_i \sim S_\alpha(\sigma, \beta, 0)$,*

1. *Perform the transformations $U_i = \text{sign}(X_i)$ and $V_i = \log|X_i|$*
2. *Compute $\hat{\theta} = E[U]$, $\hat{\tau} = E[V]$, and $\hat{\nu} = \frac{6}{\pi^2} \text{Var}V - \frac{3}{2} \text{Var}U + 1$*
3. *Transform to obtain the estimates*

$$\begin{aligned}\hat{\alpha} &= 1/\sqrt{\hat{\nu}} \\ \hat{\beta} &= \hat{\theta} \max(1, 1/(2\sqrt{\hat{\nu}} - 1)) \text{sign}(1 - 1/\hat{\nu}) \\ \hat{\sigma} &= \begin{cases} \exp\{\hat{\tau}/\sqrt{\hat{\nu}} - \mathbb{C}(1 - 1/\sqrt{\hat{\nu}})\} & \text{if } \nu \neq 1 \\ \exp\{\hat{\tau} + \log \cos(\pi\hat{\theta}/2) - \log(\pi/2)\} & \text{if } \nu = 1 \end{cases}\end{aligned}$$

The statistics $\hat{\theta}$ and $\hat{\tau}$ are consistent and unbiased.

Other procedures for statistical estimation of stable random variables exist, some of which are given in [95] and [40]. It is desirable to have maximum precision in the estimate of parameters in many applications. By property 2.4, a minimum-variance estimator is not possible, but a minimum α -order moment estimator is. Some further examination of Zolotarev's estimator follows.

Although the statistics $\hat{\theta}$ and $\hat{\tau}$ are unbiased, numerical simulations indicate that the estimators $\hat{\alpha}$, $\hat{\sigma}$ and $\hat{\beta}$ are biased in an undetermined way. Figure 2.3 shows the estimates by Zolotarev's method for samples of $X_i \sim S_{1.5}(2, .4, 0)$ generated by the Chambers-Mallows-Stuck method of

sizes $N = 1000, 2000, 4000, 8000, 16,000, 32,000, 64,000$, and $128,000$. Figure 2.3 shows the estimates by Zolotarev's method for samples of $X_i \sim S_{0.7}(2, .4, 0)$.

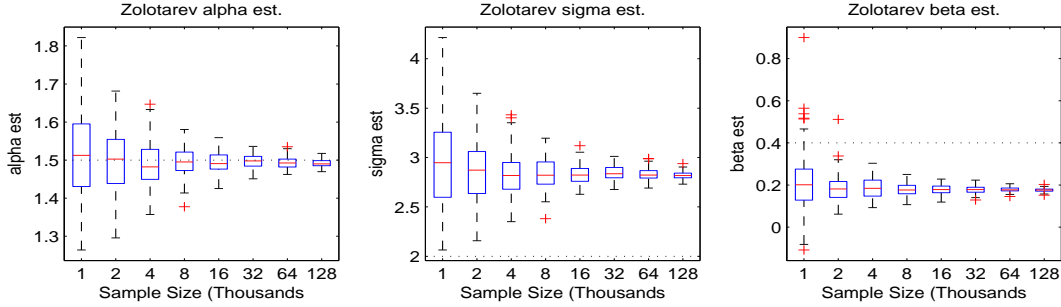


Figure 2.3: Statistical estimation of Simulated α -stable Samples of $X_i \sim S_{1.5}(2, .4, 0)$: $\hat{\alpha}$ (left), $\hat{\sigma}$ (center), $\hat{\beta}$ (right)

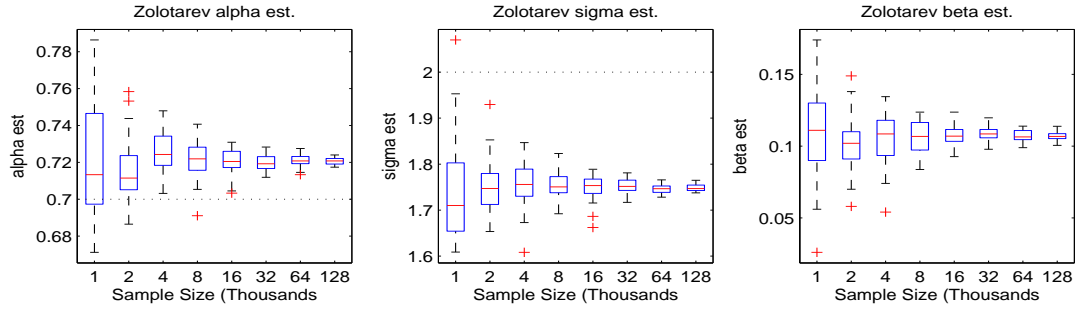


Figure 2.4: Statistical estimation of Simulated α -stable Samples of $X_i \sim S_{.7}(2, .4, 0)$: $\hat{\alpha}$ (left), $\hat{\sigma}$ (center), $\hat{\beta}$ (right)

Although there is a clear bias in the estimates, it occurs in a consistent manner that depends on the parameters. Numerical simulation shows that the estimate $\hat{\sigma}$ is unbiased if $\sigma = 1$, and that $\hat{\sigma}^{1/\hat{\alpha}}$ provides a better estimate of σ . An unbiased estimate of σ can be obtained iteratively as follows.

Algorithm 2.16 (Unbiased σ Estimation) *A set of N i.i.d. samples*

$\{X_i\}_{i=1}^N$ from a centered r.v. $X_i \sim S_{\alpha}(\sigma, \beta, 0)$ can be standardized and an

unbiased estimate of σ can be obtained as follows.

1. Define $\{X_i^0\}_{i=1}^N$ by $X_i^0 = X_i$ and let $\sigma_0 = \hat{\sigma}_0$ and $\alpha_0 = \hat{\alpha}_0$ where $\hat{\sigma}_0$ and $\hat{\alpha}_0$ are obtained by algorithm 2.15 applied to $\{X_i^0\}_{i=1}^N$.
2. For $j = 0, 1, 2, \dots$, re-scale the sample by $X_i^{j+1} = X_i^j / \sigma_j^{(1/\alpha_j)}$ and obtain $\sigma_{j+1} = \hat{\sigma}_{j+1}$ and $\alpha_{j+1} = \hat{\alpha}_{j+1}$, using algorithm 2.15 for $\{X_i^{j+1}\}_{i=1}^N$.
3. Repeat step 2 until $\sigma_k - \sigma_{k-1}$ is within some tolerance at step k ; less than 10 steps will achieve convergence
4. The final sample $\{X_i^k\}_{i=1}^N$ will be standardized, with $\sigma_k = 1$. $\hat{\sigma}$ for the original sequence $\{X_i\}_{i=1}^N$ is given by $\hat{\sigma} = \tilde{\sigma} = \prod_{i=0}^k \sigma_i$

2.1.7 Iterative Numerical Estimation

In order to correct bias in a statistical estimator, a maximum likelihood estimation (MLE) method for stable distributions has been suggested [98] in the form of an inverse problem. The implementation presented here is designed to automatically resolve convergence issues, which are related to the accuracy issues discussed in section 2.1.5, and is therefore useful for implementation in automatic algorithms.

Definition 2.17 (Log-Likelihood Function) For the p.d.f. $f(x)$ of a centered stable r.v. $X \sim S_\alpha(\sigma, \beta, 0)$ and i.i.d. samples $\{X_i\}_{i=1}^N$, the log-likelihood function (l.l.f.) is defined by:

$$L(\alpha, \sigma, \beta) \triangleq \sum_{i=1}^N \log f(X_i; \alpha, \sigma, \beta) \quad (2.20)$$

The supremum of the l.l.f. over the parameter space yields the maximum likelihood (ML) estimates $(\alpha^*, \sigma^*, \beta^*)$:

$$(\alpha^*, \sigma^*, \beta^*) = \sup_{\alpha, \sigma, \beta} L(\alpha, \sigma, \beta) \quad (2.21)$$

If the samples $\{X_i\}_{i=1}^N$ have a stable distribution, the l.l.f. is convex and the maximum can be found using numerical techniques. Because the gradient of this function cannot be calculated, a black-box routine must be used. The `fminsearch` function in the MATLAB Optimization Toolbox [88] implements a Nelder-Mead simplex method suitable for such problems.

Algorithm 2.18 (Iterative Numerical MLE) *Given data vector $\{X_i\}_{i=1}^N$,*

1. *The sample $\{X_i\}_{i=1}^N$ is standardized by division by a scaling factor $\tilde{\sigma}$ obtained using algorithm 2.16, and a starting point $y^0 = (\alpha^0, \sigma^0, \beta^0)$ is obtained by algorithm 2.15, where $\sigma^0 = 1$*
2. *Integration range a is obtained for $\phi_{tol} = 10^{-10}$ by*

$$a = (-\log(\phi_{tol}))^{(1/\alpha^0)}$$

and the grid size $N = 2^m$ is selected by the user.

3. *Global simplex minimization routine `fminsearch.m` is used with the l.l.f. $L(\alpha, \sigma, \beta)$ as the objective function. At each evaluation, the density $\hat{f}(x)$ is approximated on the domain $[-c, c]$, where $c = N\pi/2a$, by algorithm 2.11. The l.l.f. is evaluated by spline-interpolation at $X_i \in [-c/2, c/2]$ using $\hat{f}(x)$, and by the asymptotic Bergström expansion given in algorithm 2.14 for $X_i < -c/2$ or $X_i > c/2$. The standardized sample is used in this step.*
4. *The optimization terminates at $y^* = (\alpha^*, \sigma^*, \beta^*)$ when a user-set tolerance is reached, and the final estimate $(\hat{\alpha}, \hat{\sigma}, \hat{\beta})$ is given by $(\alpha^*, \sigma^* \tilde{\sigma}, \beta^*)$*

The optimization routine is very sensitive to the initial conditions, most importantly because the initial σ estimate largely determines the integration range a . Standardization by algorithm 2.16 eliminates this issue. Convergence of the routine depends on reasonably accurate evaluation of the p.d.f. using the same a throughout the procedure. The estimate may be refined by running it again using its output as the initial estimate.

2.1.8 Testing Goodness of Fit

Given a vector of samples $\{X_i\}_{i=1}^N$ and an estimate of its distribution $F(x)$, it is important to test the goodness-of-fit (g.o.f.). Two powerful g.o.f. tests are the Kolmogorov-Smirnov (KS) test and the Anderson-Darling (AD) test, which are described in [30] and [27, 5], respectively. They are based on empirical distribution function (e.d.f.) statistics that compare the distribution of $\{F(X_i)\}_{i=1}^N$ to the uniform distribution. The AD test is designed to give more weight to the tails. These tests are implemented here to test the accuracy of algorithm 2.18. The null hypothesis H_0 is that the elements of a sample $\{X_i\}_{i=1}^N$ (assumed to be i.i.d.) from a r.v. X are distributed according to $X \sim S_\alpha(\sigma, \beta, 0)$. More on hypothesis testing is given in [30].

Algorithm 2.19 (Kolmogorov-Smirnov g.o.f. Test) *Given a sample*

$\{X_i\}_{i=1}^N$, a c.d.f. $F(x)$, and significance level α_{KS} ,

1. The e.d.f. statistic is calculated by $D = \max_i |i/N - F(X_i)|$
2. The p-value is given by $p_{KS}(D) = K(D)$ where $K(\cdot)$ is the KS-distribution:

$$K(x) = 1 + 2 \sum_{j=1}^{\infty} (-1)^j \exp(-2x^2 j^2) \quad (2.22)$$

3. If $p_{KS} > 1 - \alpha_{KS}$ the hypothesis that the sample is distributed with c.d.f. $F(x)$ is rejected at significance level $100(\alpha_{KS})\%$.

Algorithm 2.20 (Anderson-Darling g.o.f. Test) *Given a sample $\{X_i\}_{i=1}^N$,*

a c.d.f. $F(x)$, and significance level α_{AD} ,

1. The e.d.f. statistic A^2 is calculated:

$$A^2 = -N - \frac{1}{N} \sum_{i=1}^N (2i-1) (\log(F(X_i)) + \log(1 - F(X_{n+i-1}))) \quad (2.23)$$

2. The Anderson-Darling statistic D_α for a significance level α is given by a table [27], as computation of the AD-distribution is problematic [87].
3. If $A^2 > D_\alpha$ the hypothesis that the sample is distributed with c.d.f. $F(x)$ is rejected at significance level $100(\alpha_{AD})\%$.

Algorithm 2.21 (Approximation of Stable c.d.f.) For a stable r.v.

$X \sim S_\alpha(\sigma, \beta, 0)$ the c.d.f. approximation $\{\hat{F}(X_i) = P(X < X_i)\}_{i=1}^N$ can be obtained on a set of points $\{X_i\}_{i=1}^N$, sorted in ascending order, by:

1. The density $\hat{f}(x)$ is computed according to 2.11 on $x \in [-c, c]$, using step 2 in 2.18 to insure accuracy
2. Using a cumulative vector integration routine, such as `cumsimpsum.m` available at the MATLAB file exchange [88], numerically integrate $\hat{f}(x_k)$ for $k = 2^{m-2} \dots 2^{m-1} + 2^{m-2} + 1$ to obtain the vector $\{\tilde{F}_j\}$ for $j = 1 \dots 2^{m-1} + 2$ with values at $\{\tilde{x}_j = x_{j+2^{m-2}-2}\}$.
3. If $X_i \in [-c/2, c/2] \forall i$, interpolate the pair $(\{\tilde{x}_j\}, \{\tilde{F}_j\})$ at $\{X_i\}_{i=1}^N$ to obtain $\{F_i\}_{i=1}^N$. This can be done with cubic-spline interpolation by `spline.m`. Define $\hat{F}(X_i) = F_i + Q$ where $Q = P(X < X_1)$ is the lower tail probability, which can be obtained using property 2.8.
4. If values of X_i exist outside of $[-c/2, c/2]$, define $q1 = \arg \max_i \{X_i < -c/2\}$ and $q2 = \arg \min_i \{X_i > c/2\}$. Interpolate the pair $(\{\tilde{x}_j, \tilde{F}_j\})$ at $\{X_i\}_{i=q1+1}^{q2-1}$ to obtain $\{F_i\}_{i=q1+1}^{q2-1}$. Compute tail probabilities $\{F_i = P(X < X_i)\}_{i=1}^{q1}$ and $\{F_i = P(X > X_i)\}_{i=q2}^N$ using property 2.8. Define the c.d.f. estimate as

$$\hat{F}(X_i) = \begin{cases} F_i & \text{for } i = 1 \dots q1 \\ F_{q1} + F_i & \text{for } i = q1 + 1 \dots q2 - 1 \\ F_{q1} + F_{q2-1} + F_{q2} - F_i & \text{for } i = q2 \dots N \end{cases} \quad (2.24)$$

The values given by the Bergström expansion and the vector integration at the upper tail combine seamlessly with negligible error for practical application for values of $\alpha > 1$.

2.1.9 Accuracy and Time of Estimation

It has been demonstrated through simulation studies that maximum-likelihood estimators are by far the most accurate of all available for stable distributions. However, previous numerical MLE algorithms require several minutes to terminate, even for samples fewer than 50,000. An important factor is the proper selection of the integration range a . Different choices of a result in different values for $\hat{f}(x)$, the l.l.f. $L(\alpha, \sigma, \beta)$, and therefore the objective function. The parameter a must remain constant throughout the algorithm for the optimization to converge. The method for choosing a used here results in convergence of the algorithm for samples generated using 2.10 for all permissible (α, σ, β) where $\alpha \geq 1$ using Zolotarev's 'M' parametrization. Furthermore, the samples pass goodness-of-fit tests as described above for the estimated distributions. The algorithm also converges for samples where $\alpha < 1$, though its performance is less consistent in such cases. Reasonable estimates can be obtained in cases where $\alpha > 0.5$.

To give an impression of the accuracy, efficiency, consistency, and precision of the algorithm, simulations are performed for several values of α and β involving the symmetric and skewed cases. The Chambers-Mallows-Stuck method is used to generate 100 samples each of sizes $N = 1000, 2000, 4000, 8000, 16,000$, as well as $32,000, 64,000, \text{ and } 128,000$ if $\alpha > 1$, in Zolotarev's 'M' parametrization. A grid size of $N = 2^{11}$ is used. Box-plots of the MLE estimator values for increasing sample sizes are shown in for each simulation, as well as Box-plots for Zolotarev's estimator for comparison. The conver-

gence of the variance of the estimates, the computation time on a 3 GHz dual core workstation, and the mean p-values for KS-tests of the estimates are illustrated. This is shown for $X \sim S_{1.7}(2, -0.2, 0)$ below:

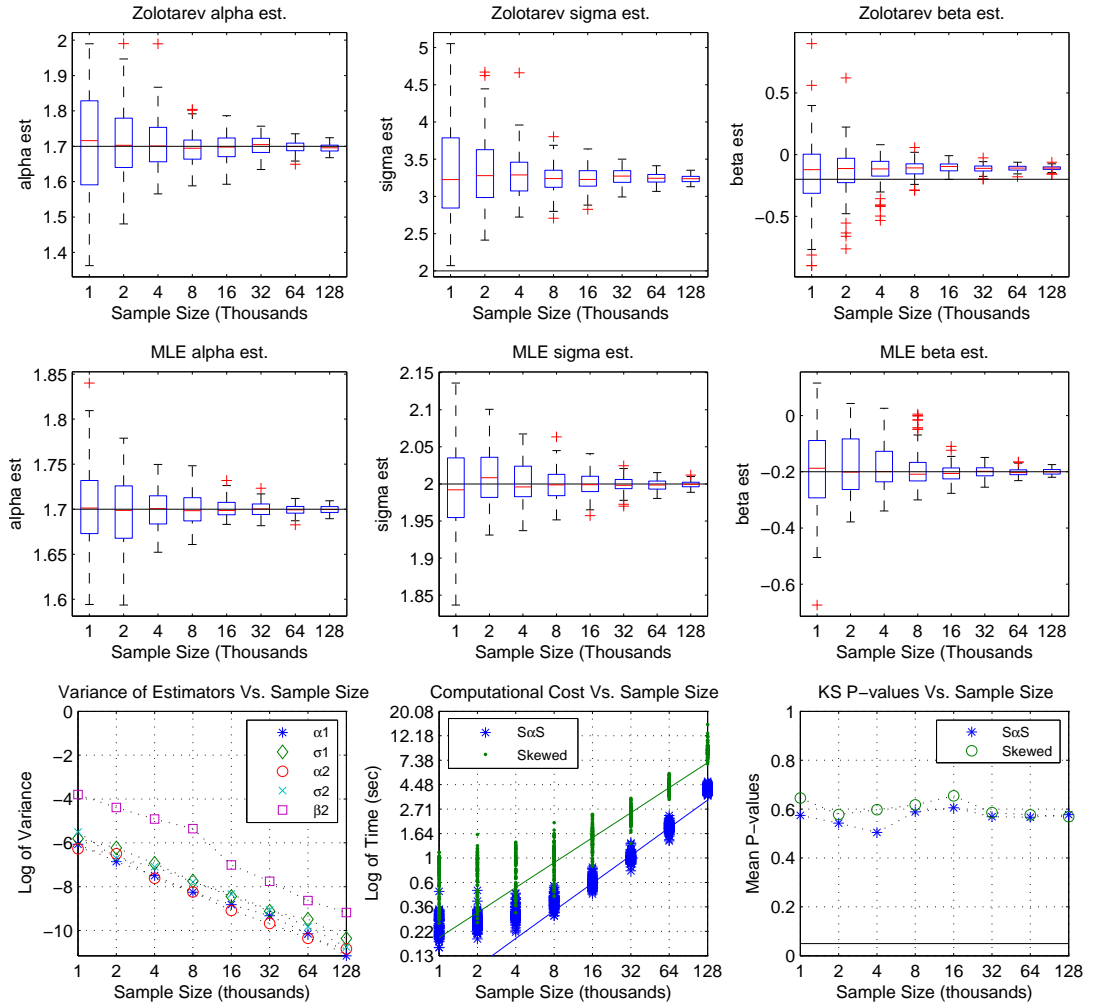


Figure 2.5: MLE for Simulated α -stable Samples, $X \sim S_{1.7}(2, -0.2, 0)$: Box-plots by Zolotarev's method (top), Box-plots for MLE method (center), and convergence of estimates (bottom left), computational cost (bottom center), mean p-values (complementary) for KS-test (bottom right)

Note that for $\alpha = 1.7$ in figure 2.5, the procedure takes under 8-10 seconds for a sample of size $N = 120,000$, and under 3-5 seconds if it is $S\alpha S$. The estimates are consistent and unbiased, and the samples pass the Kolmogorov-Smirnov g.o.f. test formulated above with low p-values.

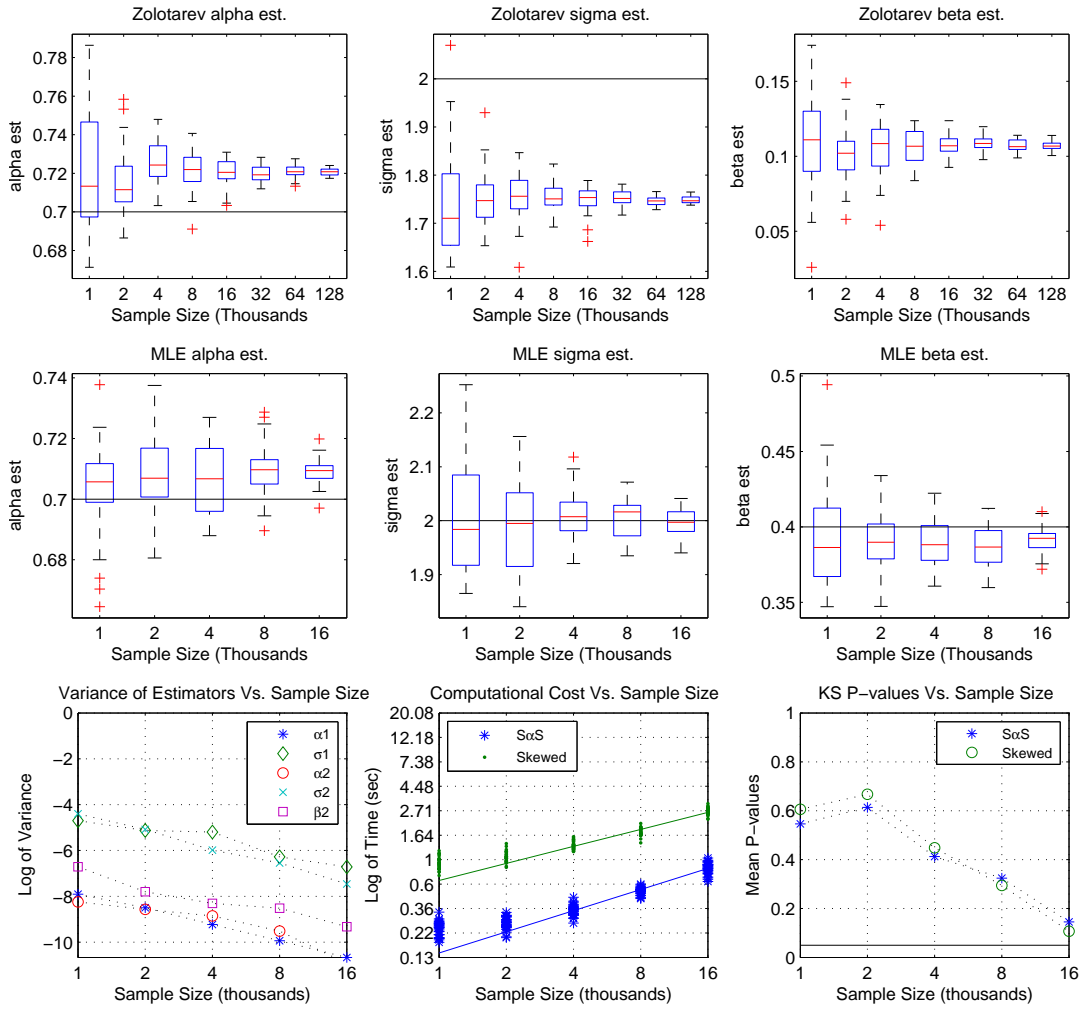


Figure 2.6: MLE for Simulated α -stable Samples, $X \sim S_{0.7}(2, 0.4, 0)$: Box-plots by Zolotarev's method (top), Box-plots for MLE method (center), and convergence of estimates (bottom left), computational cost (bottom center), mean p-values (complementary) for KS-test (bottom right)

The performance of the algorithm for $\alpha < 1$ is demonstrated in figure 2.6. There is little bias in the results, which are acceptable and even pass the KS-test. The p-values increase for larger samples due to the error in approximating the p.d.f., and possibly due to bias in the Chambers-Mallows-Stuck simulation method. In general, for real data that does in fact fit the α -stable model, the MLE algorithm provides a rapid way to obtain consistent and unbiased parameter estimates.

2.2 Stable Stochastic Processes

The class of fractional stochastic processes with stable increments includes motions with a wide variety of scaling and volatility properties. Two examples are fractional Brownian motion and Levy processes, which have applications in many fields. While fractional Brownian motion always has a normal distribution, Levy processes have heavy-tailed distributions. A review of the Levy processes, representations, and applications in physical sciences can be found in [9], and a review of fractional Brownian motion can be found on page 25 of [112].

The stochastic processes considered here are linear and have the notable property of self-similarity. Two cases of the α -stable self-similar stochastic processes were applied to modeling traffic in broadband networks in [49], and have potential for modeling applications in other fields, including biology. The general concepts of long range correlations and time-scaling have been examined extensively in time-series from the applied sciences. Relevant

examples include [105, 81, 147] as well as pp. 286-392 of [112]

2.2.1 Self-Similar Processes

Self-Similarity is the fractal-like property of invariance over different space and time-scales. A self-similar stochastic process keeps the same distribution on all time-scales. The processes considered here also have stationary increments.

Definition 2.22 (H -sssi Process) *A process $X(t)$, $t \in \mathbb{R}$, is self-similar with index H if, for $a > 0$, $X(at) \stackrel{d}{=} a^H X(t)$. The process is H -sssi if it is self-similar with index H and has stationary increments.*

The increments of H -sssi processes can display long-range dependence or long memory properties, evidenced by a covariance function that decays to zero slowly like a power function. H -sssi processes include fractional Brownian motion (fBm), linear fractional stable motion (LFSM), log-fractional stable motion (log-FSM), and Levy α -stable motion. These motions can be represented as integrations of corresponding noises. Detailed derivations and discussions of stable processes and their properties can be found in [117]. In general, two parameters describe a linear fractional stable process, H , which determines the time-scaling, and α , which characterizes the ‘volatility’ of the independent increments used to generate process.

2.2.2 Stable Noises and Motions

Two cases of particular interest are the LFSM and log-FSM, which are extensions of fractional Brownian motion and Levy motion to the α -stable case, respectively. Other extensions exist, however these forms are best suited to simulation.

Definition 2.23 (Linear Fractional Stable Motion) *The linear fractional stable motion (LFSM) is the stochastic process $\{L_{\alpha,H}(a, b; t), -\infty < t < \infty\}$ with*

$$L_{\alpha,H}(a, b; t) \triangleq \int_{-\infty}^{\infty} f_{\alpha,H}(a, b; t, x) M(dx), \quad (2.25)$$

where

$$\begin{aligned} f_{\alpha,H}(a, b; t, x) &= a \left(((t-x)_+)^{H-1/\alpha} - ((-x)_+)^{H-1/\alpha} \right) \\ &+ b \left(((t-x)_-)^{H-1/\alpha} - ((-x)_-)^{H-1/\alpha} \right) \end{aligned} \quad (2.26)$$

and where $a, b \in \mathbb{R}$, $|a| + |b| > 0$, $0 < \alpha < 2$, $0 < H < 1$, $H \neq 1/\alpha$, and

$$z_+ \triangleq \begin{cases} z, & z \geq 0 \\ 0, & z < 0 \end{cases}, \quad z_- \triangleq \begin{cases} 0, & z \geq 0 \\ -z, & z < 0 \end{cases}$$

Here M is an α -stable random measure on \mathbb{R} with Lebesgue control measure and skewness intensity $\beta(x)$, $-\infty < x < \infty$, and $\beta(\cdot) = 0$ if $\alpha = 1$. The LFSM is called balanced if $a = b$ and anti-balanced if $a = -b$.

Definition 2.24 (Log-Fractional Stable Motion) *The log-fractional stable motion (log-FSM) is an extension of the LFSM to the case where $H =$*

$1/\alpha$. For $1 < \alpha < 2$, $\beta(\cdot) = \beta$ constant, and M as given in definition 2.23, the log-FSM is given for $-\infty < t < \infty$ by

$$\Lambda_{\alpha,1/\alpha}(t) = \int_{-\infty}^{\infty} (\ln|t-x| - \ln|x|) M(dx) \quad (2.27)$$

Lemma 2.25 (Fractional Stable Noise) *The linear fractional stable noise (LFSN) is a stationary sequence $\{Y_j\}$ where*

$$Y_j(a, b, \alpha, H) = L_{\alpha,H}(a, b, ; j+1) - L_{\alpha,H}(a, b, ; j) \quad (2.28)$$

and the log-fractional stable noise (log-FSN) is a stationary sequence $\{\Upsilon_j\}$ where

$$\Upsilon_j(a, b, \alpha, 1/\alpha) = \Lambda_{\alpha,1/\alpha}(j+1) - \Lambda_{\alpha,1/\alpha}(j) \quad (2.29)$$

$\{Y_j\}$ or $\{\Upsilon_j\}$ is said to have long-range dependence when $H > 1/\alpha$ and negative dependence when $H < 1/\alpha$.

Note that long-range dependence is possible only when $\alpha > 1$ since $H \in (0, 1)$. This is explained conceptually by the fact that when $\alpha \leq 1$, the measure M generates many large values causing the integrand in 2.23 to diverge.

Another possibility not previously presented in the literature is to generate a noise that combines two scaling properties H_1 and H_2 using the same α -stable measure M . Such a process can be defined as a *Compound Linear Fractional Stable Noise* as follows.

Definition 2.26 (Compound Linear Fractional Stable Noise) *A Compound Linear Fractional Stable Noise (CLFSN) can be defined as in 2.23,*

with kernel $f(\cdot)$ given by

$$f_{\alpha,H}(a,b;t,x) = a \left(((t-x)_+)^{H_1-1/\alpha} - ((-x)_+)^{H_1-1/\alpha} \right) + b \left(((t-x)_-)^{H_2-1/\alpha} - ((-x)_-)^{H_2-1/\alpha} \right). \quad (2.30)$$

The compounding is said to be symmetric if $H_1-1/\alpha = -(H_2-1/\alpha)$.

The kernels for one-sided, well-balanced, anti-balanced, and symmetric-compound balanced linear fractional stable motions are shown below, for $H = .2$ and $\alpha = 1.8$, and $d = H - 1/\alpha$. The significance of the compound LFSN is explained in the next section.

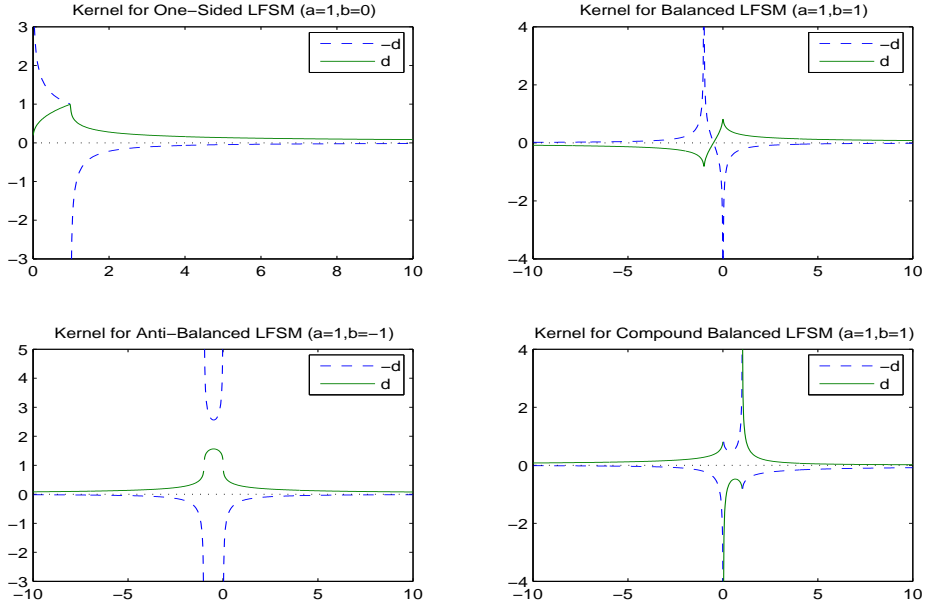


Figure 2.7: Kernels for one-sided (top right), well-balanced (top left), anti-balanced (bottom left), and symmetric-compound well-balanced (bottom right) linear fractional stable noise

2.2.3 Simulation of Fractional Stable Noises

Simulation of fractional stable noise can be done by discretization of the integral equation. This introduces “low-frequency” error due to truncation of the limits of integration and “high-frequency” error due to summation over a discrete grid. This integration is effectively a linear convolution of a random vector with a kernel. The method for simulating fractional noise is as follows:

Algorithm 2.27 (Simulation of Fractional Noise and Motion)

Fractional noise $\{Z_j(M, f(\tau, \cdot))\}_{j=1}^J$ with kernel $f(\tau, x)$, truncated at N and with discretization n (time-step $1/n$), and driven by measure M is simulated by

$$Z_j(d, f(\cdot)) = \sum_{u=-Nn}^{Nn} f\left(\frac{u}{n}\right) \epsilon_{u-j}, \quad j = 1, \dots, J \quad (2.31)$$

where $\{\epsilon_j\}$ is a sequence of i.i.d. α -stable random numbers. The corresponding fractional motion $\{X(t_j)\}_{j=1}^J$ is approximated by

$$X\left(j\frac{\tau}{m}\right) = \left(j\frac{\tau}{m}\right)^{-H} \sum_{j=1}^J Z_j, \quad j = 1, \dots, J \quad (2.32)$$

The above algorithm can be easily implemented by computing $\bar{X} = \bar{K}\bar{R}$, where \bar{K} is the discretized kernel, and \bar{R} is a random matrix. Specifically, \bar{R} is formed from J columns from a non-symmetric Toeplitz matrix generated using two vectors of pseudo-random α -stable numbers from algorithm 2.10. Filtering, by moving average or by a Gaussian windowed method is suggested to remove high frequency errors.

The balanced LFSN with $a = b = 1$ is simulated here for $H = 0.2$, $T = 200$, $\tau = 1$, with memory $N = 1000$ and time-step $n = 0.02$. It is driven by $S\alpha S$ noise with $\alpha = 1.8$ and $\sigma = 1$. The process is then filtered by a $2m$ -point moving average filter, which, in conjunction with the LFSN kernel, creates a nonlinear temporal correlation structure driven by impulsive α -stable random noise. Both signals and details are shown in figure 2.8.

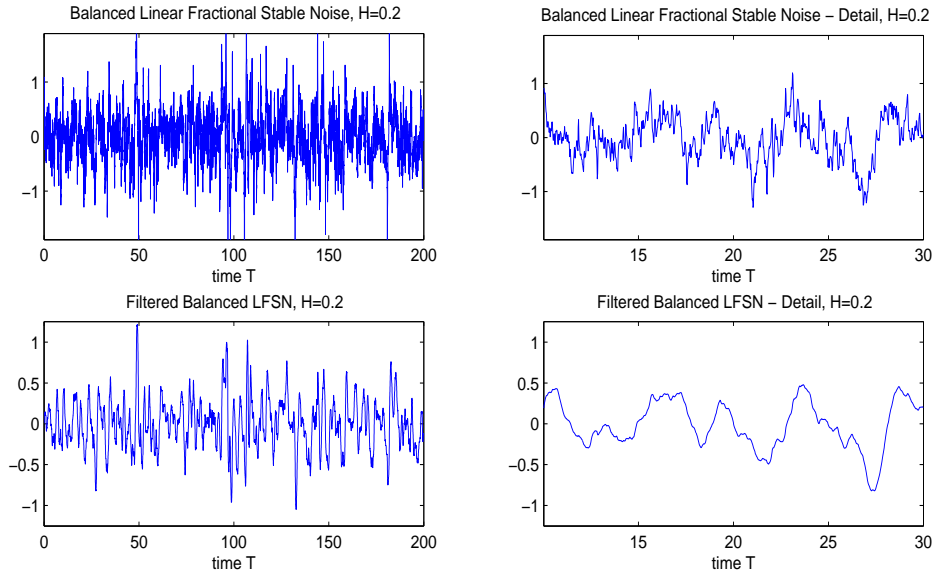


Figure 2.8: Balanced Linear Fractional Stable Noise (top left) and detail (top right), and Filtered LFSN (bottom left) and detail (bottom right), $H = 0.2$, $\alpha = 1.8$

Then, a balanced symmetric-compound LFSN, as defined in 2.26, with the same parameters as the balanced LFSN above is simulated. It is similarly filtered, and the signals and details are shown in figure 2.9.

The correlation and long-memory properties of the balanced LFSN and CLFSN and their filtrations are illustrated in figure 2.10. Note that for the filtered CLFSN process the autocovariance has a first zero-crossing for very

large lag, indicating the presence of highly positively correlated long-range memory.

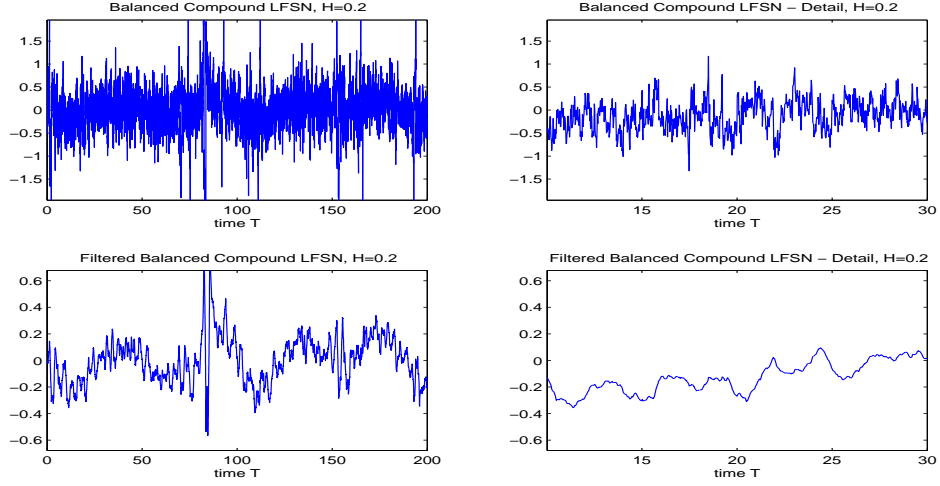


Figure 2.9: Balanced Compound Linear Fractional Stable Noise (top left) and detail (top right), and Filtered LFSN (bottom left) and detail (bottom right), $H = 0.2$, $\alpha = 1.8$

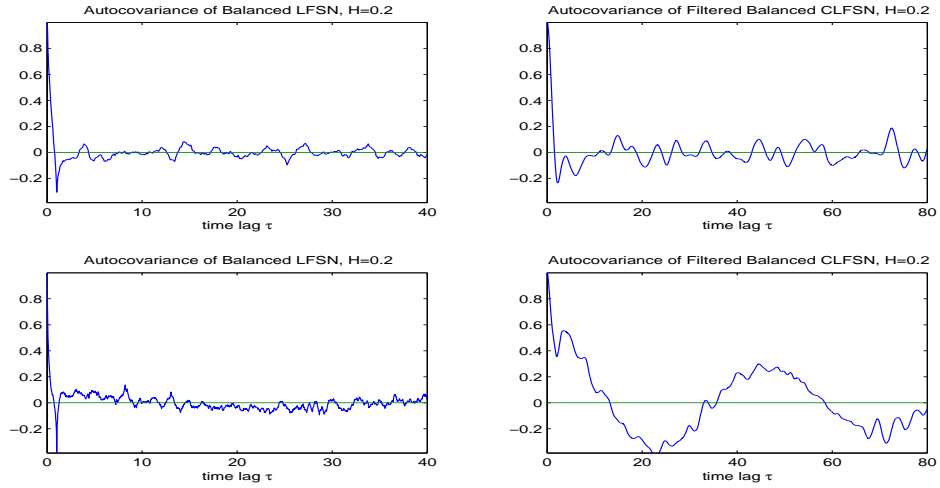


Figure 2.10: Autocovariance of Balanced LFSN (top left) and its filtration (top right), and of Balanced Compound LFSN (bottom left) and its filtration (bottom right), $H = 0.2$, $\alpha = 1.8$

The multi-scaling property of the low-frequency and high-frequency correlations in the compound process are due to the fact that the kernel is generated by two different values of H . Also, notice that the autocorrelation of the unfiltered signal appears similar to the kernel, and the filtering removes this anomaly, which is due to the numerical method used to generate the process.

2.3 Discussion and Future Work

In this chapter an algorithmic basis is developed for modeling with and estimation of random samples from stable distributions with high speed and accuracy. This is done by minimizing the error in numerical evaluation of the α -stable p.d.f. using the fast Fourier transform. The algorithms presented here can be used for simulation of random samples, and estimation and testing of the parameters. Fractional stable noises and motions and their simulation are also discussed, and the compound linear fractional stable noise is introduced. These processes will serve as models in subsequent chapters, where numerical estimation procedures are used to adapt them to physical processes, specifically the EEG.

Further work in the area of simulation and estimation of stable fractional processes and noises is required in order that they can be used as effective models for real data. A comparison of accuracy and efficiency with respect to other recent estimation techniques for α -stable random variables presented in [15, 137, 89] is necessary. In addition, the performance and applicability of

Fractional ARIMA and ARFIMA models needs to be investigated, as well as techniques for parameter estimation for such models. More investigation into the properties created by compounding scaling properties is also warranted.

The numerical methods used here may be extended to apply other probabilistic and statistical models to real data, particularly those that involve distributions with no explicit analytical probability density function. Examples of such models include the class of Smoothly-Truncated Lévy distributions [107] and fractional stable distributions [139], which arise in models of anomalous diffusion in statistical physics.

Chapter 3

Nonlinear Time-Series Methods

Nonlinear time-series analysis is an evolving field that has experienced many major developments in the past decades while the assumptions of linearity and stationarity underlying conventional time-series analysis were gradually abandoned. Though the mathematics involved have become much more complicated, simple nonlinear models can capture behaviors that could not be observed in a linear framework. Unlike the traditional linear case, there is no unified theory or conceptual framework for analysis of nonlinear time-series. Most of the works in this area are dispersed throughout the literature, though two notable collections of modern methods and ideas are [111] and [72]. The former focuses on linear systems driven by a stochastic inputs, while the latter assumes a deterministic nonlinear system.

The purpose in this chapter is to examine methods for feature extraction from short time-series. One focus is to examine the time-series as a self-similar random fractal in the sense of Hausdorff [24, 118], Mandelbrot [85], or Hurst [52, 31, 71, 70, 104] and to estimate power-law scaling parameters,

or otherwise analyze time-scaling [82, 105] or correlation [93, 131] properties. Another approach is to assume in the sense of Takens the presence of an underlying discrete [134, 50] or continuous [132, 4] dynamical system. Various ways exist to characterize such systems from time-series data that generally involve estimating optimal embedding parameters for the data [53, 46, 78, 74] and ‘reconstructing’ the underlying attractor [79, 3]. A variety of measures can then be estimated to quantify the ‘dimensionality’ [79, 133, 127], ‘determinism’ [67, 22], ‘entropy’ [55], or ‘chaoticity’ [145, 19, 116] of the system. These approaches are considered here in light of their utility and performance for extracting information from short one-dimensional time series. The intended application is algorithm development for estimation and detection problems related to the EEG.

3.1 Spectral Analysis of Stochastic Signals

The traditional methods of time-series analysis examine the basic statistical and spectral properties of a signal. The essential assumption is ergodicity, a term originally from statistical mechanics and a subject of deep mathematical study. Sufficient requirements are stationarity and linearity, in particular the convergence of sample means and autocovariances to appropriate limits in a mean square sense. A compilation of definitions, detailed derivations, and some methods can be found in [110]. The relevant framework for spectral analysis of short time-series is summarized below, and some estimators are presented:

3.1.1 Correlation and Spectral Density

Definition 3.1 (Wide-sense Stationarity) *A stochastic process $X(t)$ is called wide-sense stationary (WSS) if its mean is constant and its autocovariance $R_X(t, t + \tau)$ and autocorrelation function (a.c.f.) $\rho_X(t, t + \tau)$ depend only on τ :*

$$\begin{aligned} E[X_t] &= \mu_X & \text{var}(X(t)) &= \sigma_X^2 \\ \text{cov}[X(t)X(t + \tau)] &= R_X(\tau) & \rho_X(\tau) &= R_X(\tau)/R_X(0) \end{aligned} \quad (3.1)$$

Definition 3.2 (Power Spectrum) *The power spectrum (p.s.) $S_X(f)$ of a WSS process $X(t)$ is given by*

$$S_X(f) \triangleq \lim_{T \rightarrow \infty} E \left\{ \frac{1}{2T} |\hat{X}_T(f)|^2 \right\}, \quad \text{where} \quad \hat{X}_T(f) = \frac{1}{\sqrt{2\pi}} \int_{-T}^T X(t) e^{-i2\pi f t} dt \quad (3.2)$$

Definition 3.3 (Power Spectral Density) *The power spectral density (p.s.d.), or normalized power spectrum $s_X(f) = S_X(f)/\sigma_X^2$ is related to the autocorrelation $\rho(\tau)$ by*

$$s_X(f) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \rho_X(\tau) e^{-i2\pi f \tau} d\tau \quad (3.3)$$

When $s_X(f)$ exists, $s_X(f)df$ is interpreted as the average (over all realizations) of the proportion of total power in components of $X(t)$ with frequencies between f and $f + df$ [110].

Definition 3.4 (Periodogram) *Given a series of observations $\{X_j\}_{j=1}^N$, a function called the periodogram $I_N(f)$ is defined for $-1/2 \leq f \leq 1/2$ by*

$$I_N(f) = \frac{2}{N} \left| \sum_{k=1}^N X_k e^{-i2\pi f k} \right|^2 \quad (3.4)$$

If $X_j = X(j\tau)$ for a WSS process $X(t)$ sampled at $f_s = 1/\tau$ Hz, a consistent and unbiased estimator $\{\hat{s}_j\}_{j=1}^N$ of the p.s.d. of $X(t)$ is given by $\hat{s}_j = I_N((j-1)/(2N))/\sigma_X^2$ at frequencies $f_j = (j-1)f_s/(2N)$ (in Hz).

The periodogram can be rapidly evaluated using the FFT if $N = 2^m$, and can be ‘smoothed’ if the sequence $\{X_j\}_{j=1}^N$ is split into multiple segments of equal length M and the results for each I_M are averaged.

3.1.2 Spectral Measures

Many measures, or features, exist for extracting information from time-series through the power spectral density, and several are applied to the EEG in [43]. The simplest is *band power*, or the proportion of the total energy in components of the signal in *frequency band* $f \in [f_1, f_2]$.

Definition 3.5 (Band Power) *The band power of a signal in frequency band $f \in [f_1, f_2]$ is defined as*

$$s_X(f_1, f_2) \triangleq \int_{f_1}^{f_2} s_X(f) df \quad (3.5)$$

Definition 3.6 (Spectral Edge) *The $E_s\%$ spectral edge is defined as the frequency below which $E_s\%$ of the power is accumulated:*

$$\int_0^{EDGE} s_X(f) df = E_s \int_0^\infty s_X(f) df \quad (3.6)$$

Definition 3.7 (Spectral Moment) *The n^{th} spectral moment η_X^n is defined as*

$$\eta_X^n \triangleq \int_{-\infty}^\infty (2\pi f)^n s_X(f) df \quad (3.7)$$

Lemma 3.8 (Time-Domain Analysis) *The even spectral moments η_X^n for $n = 2, 4, \dots$ can be obtained in the time-domain by*

$$\eta_X^n = - \left. \frac{d^n R_X(\tau)}{d\tau^n} \right|_{\tau=0} = E \left[\frac{d^{n/2} X(t)}{dt^{n/2}} \right]^2 \quad (3.8)$$

The spectral moments were used in [59] to define parameters that give a valid description of an EEG signal provided that the signal has a symmetric p.d.f. with only one maximum.

Definition 3.9 (Hjorth Parameters) *The Hjorth parameters are:*

$$\begin{aligned} \text{Activity, } A &\triangleq \sigma_X^2 \\ \text{Mobility, } M &\triangleq (\eta_X^2)^{1/2} \\ \text{Complexity, } C &\triangleq (\eta_X^4 / \eta_X^2 - \eta_X^2)^{1/2} \end{aligned} \quad (3.9)$$

Hjorth's descriptors are notable for their low computational cost (due to Lemma 3.8), however large errors may be introduced in the parameter C . Another frequency-domain measure is spectral entropy, which quantifies the dispersion of power throughout the frequency spectrum.

Definition 3.10 (Spectral Entropy) *The spectral entropy is defined as*

$$SEN = - \int_0^\infty s_X(f) \log(s_X(f)) df \approx - \frac{1}{N} \sum_{j=1}^N \hat{s}_X(f) \log(\hat{s}_X(f)) \quad (3.10)$$

Spectral analysis of time-series is convenient for its low computational cost, making it suitable for practical applications, especially in real-time. The power spectrum of linear fractional stable noise, described in section 2.2.2, is shown in figure 3.1 as an example, with the 1-3 Hz frequency band and the spectral edge marked.

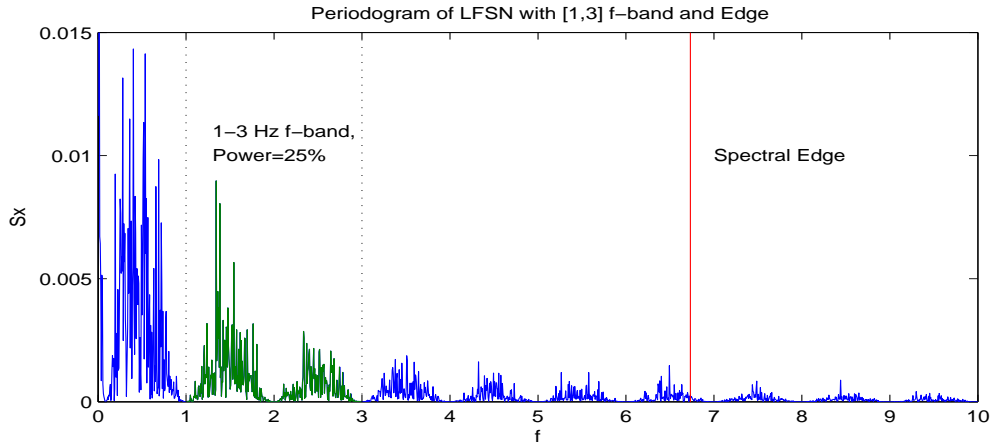


Figure 3.1: Power Spectrum of LFSN, $H=0.2$, $\alpha = 1.8$, with 1-3 Hz band power and spectral edge

3.2 Dimension and Self-Affinity Measures

One method for the analysis of time-series classifies the sampled signal as a section or profile of a fractal structure. This approach is based on geometric and topological principles from fractal theory, which is concerned with self-affine sets defined by deterministic recurrence-relations on metric spaces.

Self-similar structures are thus a subset of the self-affine class [134]. An introduction to fractal theory can be found in [10]. For a known fractal structure it is of interest to quantify how densely the set occupies the metric space in which it resides, known as its *fractal dimension* D . It is also useful to quantify how its structure is retained across different scales, a property called *self-affinity* and measured by the Hurst parameter H , first examined by Hurst in [61]. Thus fractal dimension is a local property, yet for a self-affine structure it is retained globally, resulting in $D + H = n + 1$ for a self-affine surface in n -dimensional space [52]. Furthermore for real, non-degenerate processes, $0 \leq H \leq 1$ and $n \leq D \leq n + 1$. Thus for a one-dimensional fractional Brownian motion, $D = 2 - H$, however nothing is known about the relationship between D and H in general [24]. Researchers in physiology often use the relation $D + H = 2$ [31], while in reality a general physiological signal is not globally self-affine. This results in the ‘crossover’ phenomena discovered by Peng et. al. in [104]. Recently, stochastic models have been proposed that separate D and H [52]. Several recent methods for estimating the fractal dimension for time-series are reviewed here, as well as a method for quantifying self-affinity.

3.2.1 Fractal Dimension

This section is concerned with computation of the fractal dimension of one-dimensional objects, or profiles. Many approaches exist, including box-dimension, Minkowski-Bouligand dimension, and intersection dimension [34].

Several methods for estimating the fractal dimension of signals are given in [118], and accuracy, consistency, biases, and computational cost are examined. Other methods are suggested in [24] and [58]. The focus here is on estimation using the p.s.d., and the use of more precise methods based on the variogram and curve length. The definition of fractal dimension used here is the Minkowski-Bouligand dimension [34]:

Definition 3.11 (Fractal Dimension) *The Minkowski cover $A(\epsilon)$ of a set A in \mathbb{R}^n is defined as $A(\epsilon) = \{y : y \in B_\epsilon(x), x \in A\}$, where $B_\epsilon(x)$ is a n -ball of radius ϵ centered at x . Denoting by $|E|_2$ the area of a region $E \in \mathbb{R}^2$, the Minkowski-Bouligand dimension is given by*

$$D_{MB} = \lim_{\epsilon \rightarrow 0} \left(2 - \frac{\log |A(\epsilon)|_2}{\log \epsilon} \right) \quad (3.11)$$

Lemma 3.12 *Equation 3.11 suggests a power law where $|A(\epsilon)|_2 \sim \epsilon^{2-D}$, so that from $\log |A(\epsilon)|_2 = \lambda \log(\epsilon)$, one can obtain $D \approx 2 - \lambda$*

This fact suggests the following technique presented in [58] for estimating the fractal dimension D of a time-series, where the set A is taken as a given ordered set in \mathbb{R} .

Algorithm 3.13 (Fractal Dimension - Curve Length) *To estimate the fractal dimension of a time-series $\{X(j)\}_{j=1}^N$ the curve length can be used as an approximation for the Minkowski cover as follows.*

1. *Construct a set of k time-series $\{X_k^1, X_k^2, \dots, X_k^k\}$ defined by*

$$X_k^m = \{X(m), X(m+k), X(m+2k), \dots, X(m + \lfloor (N-m)/k \rfloor \cdot k)\}$$

2. Define the length of the curve X_k^m by

$$L_m(k) = \frac{1}{k} \sum_{i=1}^{\lfloor (N-m)/k \rfloor} |X(m+ik) - X(m+(i-1)k)|$$

3. Average over the $L_m(k)$ to get $L(k) = (1/k) \sum_{m=1}^k L_m(k)$

4. Estimate the slope λ of $\log(L(k)) = \lambda \log(k)$, and then $D = 2 - \lambda$.

Another method to estimate D for a profile relies on the variogram.

Definition 3.14 (Variogram) *The variogram is defined as*

$$\gamma(\tau) = E[(X(t) - X(t + \tau))^2] = 2(R_X(0) - R_X(\tau)) \quad (3.12)$$

Proposition 3.15 *Let $A(\epsilon)$ be the Minkowski cover of a small section of the curve $X(t) \in \mathbb{R}$ representing a realization of an ergodic stochastic process. Then the expected curve-length estimate of the local Minkowski cover is $E[|A(\epsilon)|_2] \approx E[X(t + \epsilon) - X(t)]$. The square of the expected area in \mathbb{R}^2 of the local Minkowski cover is then*

$$E[|A(\epsilon)|_2]^2 \approx E[(X(t + \epsilon) - X(t))^2] = \gamma(\epsilon) \quad (3.13)$$

From 3.11, this suggests $\gamma(\epsilon) \approx |A(\epsilon)|_2^2 \propto \epsilon^{2(2-D)}$. It follows that finding the slope λ of $\log(\gamma(\epsilon)) = \lambda \log(\epsilon)$, the estimate $D \approx 2 - \lambda/2$. Refer to [24].

Proposition 3.15 provides a simple method for estimating the fractal dimension using the variogram:

Algorithm 3.16 (Fractal Dimension - Variogram) *To estimate the fractal dimension of a time-series $\{X(j)\}_{j=1}^N$ using the variogram method,*

1. Compute the variogram $\gamma(j)$
2. Estimate the slope λ of $\log(\gamma(k)) = \lambda \log(k)$, and then $D = 2 - \lambda/2$.

Lemma 3.17 For an autocovariance function $R_X(\tau) \propto |\tau|^\lambda$ as $t \rightarrow 0$, the power spectral density follows the law $S_X(f) \propto |f|^{-(\lambda+1)}$ as $f \rightarrow \infty$ [24].

Algorithm 3.18 (Fractal Dimension - Power Spectrum) To estimate the fractal dimension of a time-series $\{X(j)\}_{j=1}^N$ using the power spectrum,

1. Compute the power spectrum $S_X(f)$
2. Estimate the slope λ of $\log(S_X(f)) = \lambda \log(f)$, and then $D = 2.5 + \lambda/2$.

The use of the three methods discussed above is illustrated for estimating the fractal dimension D for Brownian motion, approximated by the cumulative sum of a white noise process.

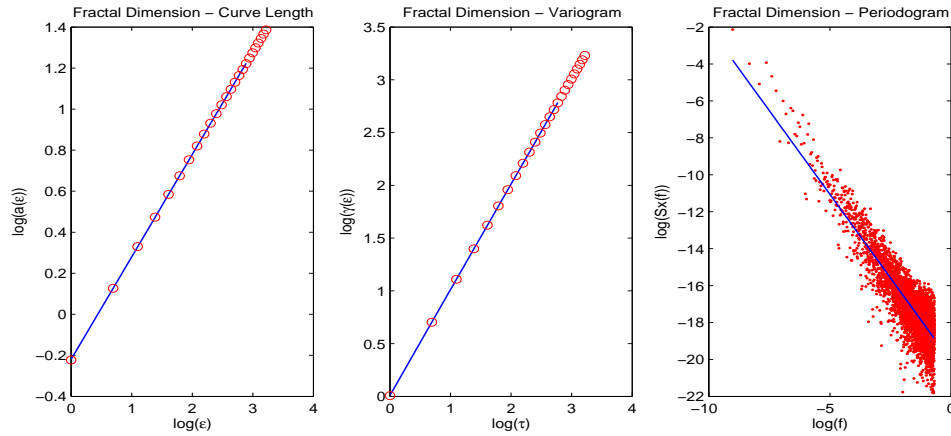


Figure 3.2: Fractal Dimension of Brownian Motion by curve length (left), variogram (center), and periodogram (right) estimation

	Actual	Curve Length	Variogram	Periodogram
Smooth	1	1.0000 ± 0.0001	1.0004 ± 0.0001	1.6025 ± 0.0003
Brownian	1.5	1.5004 ± 0.0085	1.5003 ± 0.0071	1.6030 ± 0.0003
Noise	2	1.9999 ± 0.0024	1.9999 ± 0.0020	2.5014 ± 0.0080

Table 3.1: Comparison of Fractal Dimension Methods

The means and standard deviations of the estimators for D of white noise, Brownian motion, and a smooth curve (e.g. sinusoid on $[0, 1]$) using each method are shown in table 3.1 for 100 trials using 10000 points.

The two methods based on the Minkowski-Bouligand dimension are consistent and unbiased, while the periodogram method cannot distinguish between Brownian Motion and a straight line, and gives an unreasonable value for white noise. Of these methods, the variogram method is the most accurate and consistent, and will be used for the subsequent analysis in later chapters.

3.2.2 Hurst Exponent

The Hurst exponent H is a measure of the self-affinity of a structure, which is a generalization of self-similarity discussed in section 2.2.1. In this section, the Hurst exponent of one-dimensional objects, or profiles is discussed. The standard approach to evaluating H has been to estimate D and assume global self-affinity to obtain $H = 2 - D$. One recently developed approach for estimating H independent of D uses scaled windowed variance (SWV) methods, which are reviewed in [20]. The most widely used method is detrended fluctuation analysis (DFA), presented in [105] and examined also in

[70, 71].

Algorithm 3.19 (De-trended Fluctuation Analysis) *To estimate the Hurst parameter of a time-series $\{X(j)\}_{j=1}^N$ by de-trended fluctuation analysis,*

1. *For $n = 2, 4, \dots, N/2, N$, partition $\{X(j)\}_{j=1}^N$ into non-overlapping windows of size n . Remove the linear least-squares fit from each window and define $SD(n)$ as the standard deviation of the resulting time-series.*
2. *Detect the first linear region of the plot of $\log(SD(n))$ vs. $\log(n)$. The slope of the plot in this region is related to the Hurst parameter H .*

In addition to self-affinity, the Hurst parameter H also reflects the presence of long-range correlations in a time-series, which is the focus in [105]. In general, H is much more difficult to estimate than D , especially for non-Gaussian or impulsive signals, such as those discussed in chapter 2. The exact relationship between the DFA and H is not known. Other estimators are reviewed in [102], in particular the Whittle estimator.

3.3 Phase-Space Reconstruction

The time-series analysis methods discussed in the next two sections rely on the fact that irregularity in a signal may be caused by deterministic non-linearity as well as randomness. Long-range correlations may be present because an underlying dynamical system is generating quasi-periodic oscillations. With this assumption, the time-series is a measurement of the output of a differentiable dynamical system defined on an m -dimensional manifold. A natural initial approach to testing for and quantifying the existence of a

quasi-periodic oscillator is to study sequences sampled from one-dimensional outputs of dynamical systems governed by known equations. Over the past decades the advances in nonlinear theory along with increases in computing power have resulted in methods for numerically estimating the dimension [53] and Lyapunov exponents [145], which are measures of chaos, for known dynamical systems. These methods have subsequently been applied to real data that exhibit behavior suggesting the presence of nonlinear chaos. Because the correlation structure of many physiological signals appears to be nonlinear, these and other nonlinear measures have been readily applied to medical data.

Output measurements of the underlying attractor from real data consist of one-dimensional time-series, so it is necessary to reconstruct the phase-space of the attractor from these available observations. This is done by time-delay embedding, or reconstruction.

Definition 3.20 (Delay Reconstruction) *The delay reconstruction*

$\{\vec{X}_\tau^m(j)\}_{j=1}^{N-(m-1)\tau}$ *from a series of observations $\{X(j)\}_{j=1}^N$ is defined by*

$$\vec{X}_\tau^m(j) = \{X(j), X(j + \tau), X(j + 2\tau), \dots, X(j + (m - 1)\tau)\} \quad (3.14)$$

where m is the embedding dimension and τ is the embedding delay.

The choice of parameters m and τ is important for the attractor reconstruction, and as expected, these parameters interact. Before continuing, it is useful to present two famous dynamical systems in \mathbb{R}^3 that are often used as examples for delay reconstruction.

Example 1 (Lorenz Attractor) *The Lorenz attractor is defined by*

$$\begin{aligned}\dot{x} &= \sigma(y - x) \\ \dot{y} &= -xz + \rho x - y \\ \dot{z} &= xy - \beta z\end{aligned}\tag{3.15}$$

Example 2 (Rossler Attractor) *The Rossler attractor is defined by*

$$\begin{aligned}\dot{x} &= -y - z \\ \dot{y} &= x + ay \\ \dot{z} &= b + z(x - c)\end{aligned}\tag{3.16}$$

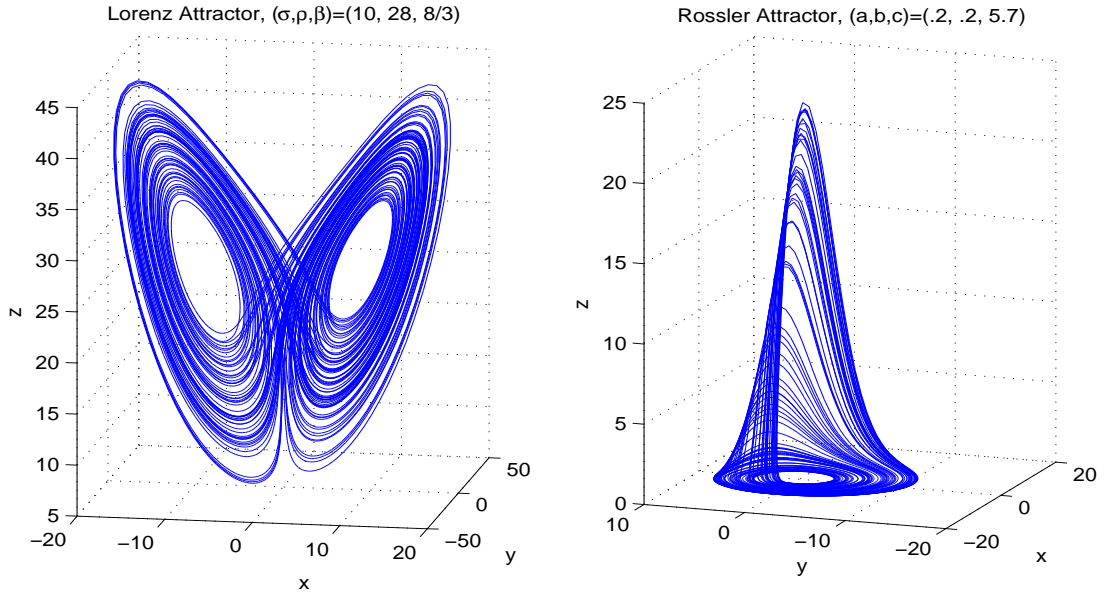


Figure 3.3: Lorenz Attractor with $(a, b, c) = (.2, .2, 5.7)$, $t \in [0, 100]$ (left) and Rossler Attractor for $(\sigma, \rho, \beta) = (10, 28, 8/3)$, $t \in [0, 500]$ (right)

A review of many widely used discrete and continuous dynamical systems as well some of their properties can be found in [127].

3.3.1 Reconstruction Time Delay

The criteria for selecting the reconstruction delay τ are thoroughly discussed in [46]. The basic idea is to choose a delay such that the elements of each state vector $\vec{X}_\tau^m(j)$ are as independent as possible. Ideally, the elements of the vectors should be linearly uncorrelated, and also share as little information as possible. The objective is to reconstruct the attractor using a minimum embedding dimension m and a delay τ such that the embedding vectors contain the maximum possible information about the state of the attractor. Thus the more sophisticated measures of *entropy* and *mutual information* are introduced, which are discussed in the context of message transmission. For a dynamical system, the ‘message’ is the value of a state variable, and because attractors are ergodic with asymptotic distributions, the probability of a certain ‘message’ at any given time converges to a limit. In the context of a time-series, the information contained in neighboring points is the quantity of interest.

Definition 3.21 (Information Entropy) *Consider a set of possible messages $\mathbb{S} \equiv \{s_j\}$, each associated with a probability $P_s: s_j \mapsto P(s = s_j)$ that a message s is received, with $\sum_j P_s(s_j) = 1$. Then the information entropy $H(\mathbb{S})$ of the system is*

$$H(\mathbb{S}) \triangleq \sum_j P_s(s_j) \log \left[\frac{1}{P_s(s_j)} \right] = - \sum_j P_s(s_j) \log(P_s(s_j)) \quad (3.17)$$

Furthermore, given two related systems \mathbb{Q} and \mathbb{S} , the conditional entropy is

$$H(\mathbb{Q}|\mathbb{S}) \triangleq - \sum_{j,k} P_{sq}(s_j, q_k) \log \left[\frac{P_{sq}(s_j, q_k)}{P_s(s_j)} \right] = H(\mathbb{S}, \mathbb{Q}) - H(\mathbb{S}) \quad (3.18)$$

where

$$H(\mathbb{S}, \mathbb{Q}) \triangleq - \sum_{j,k} P_{sq}(s_j, q_k) \log [P_{sq}(s_j, q_k)] \quad (3.19)$$

The entropy $H(\mathbb{S})$ can be interpreted as the average quantity of information about a system gained from a measurement of s . The conditional entropy $H(\mathbb{Q}|\mathbb{S})$ can be considered as the average uncertainty in a measurement of q given that s is known. Additionally, mutual information is explained more directly as the amount of information that can be obtained about q given a measurement of s .

Definition 3.22 (Mutual Information) *The mutual information $I(\mathbb{Q}, \mathbb{S})$, equivalent to $I(\mathbb{S}, \mathbb{Q})$, of two systems \mathbb{Q} and \mathbb{S} is defined as*

$$I(\mathbb{Q}, \mathbb{S}) \triangleq H(\mathbb{Q}) - H(\mathbb{Q}|\mathbb{S}) = H(\mathbb{Q}) + H(\mathbb{S}) - H(\mathbb{S}, \mathbb{Q}) \quad (3.20)$$

Rigorous definitions and derivations of entropy and information in the context of dynamical systems can be found in [46].

To choose an optimal reconstruction delay, the measures of mutual information and autocovariance may be used. The time-delay τ is chosen as the first zero-crossing or first minimum of either of these measures. The evaluation of the optimal time-delay is illustrated in figure 3.4 for the Lorenz and Rossler attractors from the first state variable using both methods.

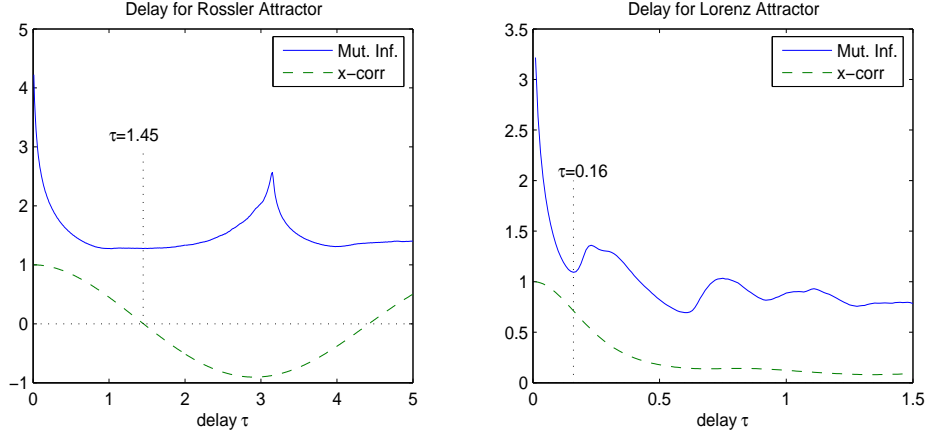


Figure 3.4: Independent Delay Coordinates for Rossler (left) and Lorenz (right) Attractors using Mutual Information and Autocorrelation

The two metrics may yield different delay values. In the case of the Rossler attractor, the mutual information is nearly invariant on the interval $\tau \in [1, 1.5]$, while a clear zero-crossing in the autocorrelation is evident. However for the Lorenz attractor, there are no zero-crossings or clear local minima of the autocorrelation function on the relevant domain, while the mutual information has a clear first minimum. It appears that mutual information is a superior metric to autocorrelation because it can reflect the nonlinear behavior of an attractor. This is the same result found by Fraser [46] for other attractors.

3.3.2 Embedding Dimension

The standard method for selecting an embedding dimension for a time-delay reconstruction is that of *false nearest neighbor* (FNN) analysis, a technique discussed in [74] and also in [72]. The idea is to choose a dimension for which the temporal evolution of nearby embedding vectors does not diverge too

dramatically. In addition, the dimension must not be so large to permit the embedding vectors to contain redundant information, which may “confuse” the algorithm.

Definition 3.23 (Nearest Neighbor) *For a given delay embedding*

$\{\vec{X}_\tau^m(j)\}_{j=1}^{N-(m-1)\tau}$, *the nearest neighbor $\vec{X}_\tau^m(j)$ to a state vector $\vec{X}_\tau^m(j)$ in the m -dimensional embedding space is a vector $\vec{X}_\tau^m(k)$ such that*

$$k = \arg \min_i \|\vec{X}_\tau^m(i) - \vec{X}_\tau^m(j)\| \quad (3.21)$$

Kennel and Abarbanel [74] give two tests for false nearest neighbors. The first examines if increasing the dimension from m to $m+1$ causes a significant increase in the distance between coordinates that are nearest neighbors in \mathbb{R}^m . The second compares the distance between those coordinates in \mathbb{R}^{m+1} to the size of the attractor.

Proposition 3.24 (False Nearest Neighbor Test) *If $\vec{X}_\tau^m(j)$ and $\vec{X}_\tau^m(j)$ are nearest neighbors, they are false nearest neighbors if*

$$\frac{|\tilde{X}(j + m\tau) - X(j + m\tau)|}{\|\vec{X}_\tau^m(j) - \vec{X}_\tau^m(j)\|} > R_{tol} \quad \text{or} \quad \frac{\|\vec{X}_\tau^{m+1}(j) - \vec{X}_\tau^{m+1}(j)\|}{\sigma} > A_{tol} \quad (3.22)$$

where σ is the standard deviation of $\{X(j)\}_{j=1}^N$

The tolerances $R_{tol} = 10$ and $A_{tol} = 2$ are reasonable values. This test can be applied to choose a minimum embedding dimension for which a sufficiently low ratio of nearest neighbors are false. It involves finding the nearest

neighbor for each state vector, and is computationally very intensive. The proportion of false nearest neighbors for 10000 data points of the first state variable sampled at $\tau = 0.01$ is shown in figure 3.5 for the Rossler and Lorenz systems for several reconstruction dimensions.

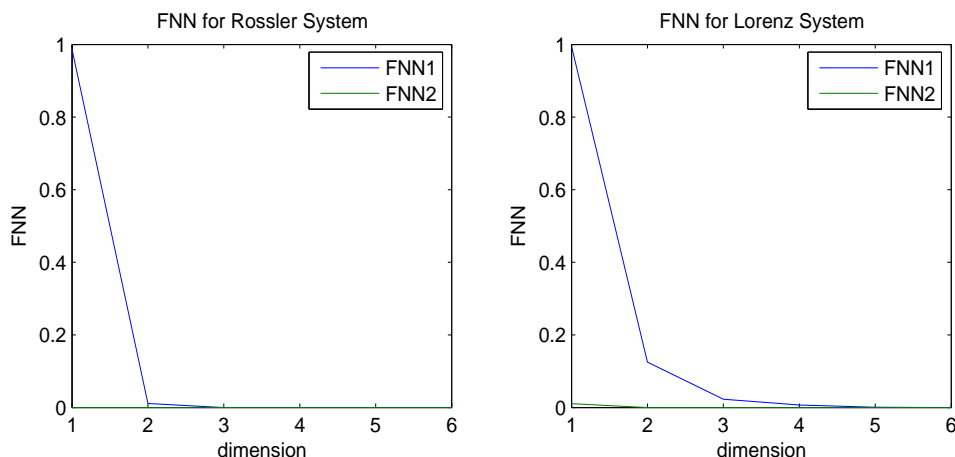


Figure 3.5: Embedding Dimension for Rossler (left) and Lorenz (right) Attractors via False Nearest Neighbors

The choice of embedding dimension is a difficult theoretical matter. Takens' theorem states that a dimension of $m = 2D + 1$ yields an embedding of a compact manifold with dimension D . However, there are data sets [72] from which the attractor may be reconstructed in spaces between D and $2D$. Thus unless the dimension of the state space of the underlying nonlinear dynamical system is known a priori for an observed data set, it is problematic to choose a reconstruction dimension in a non-arbitrary way. In practice, a reasonable value is $2\hat{D} + 1$, where \hat{D} is the reconstruction dimension for which the proportion of false nearest neighbors is near 2%.

3.4 Phase-Space Measures

Section 3.3 summarized the basic process for attractor reconstruction from a time-series. Assuming that the time-series is in fact an observation of a low-dimensional nonlinear dynamical system, and that the state reconstruction is a reasonable representation of the attractor in the sense of Takens' theorem, several useful measures can be extracted from this data to describe the system. A natural first step is to define a metric to quantify the determinism of the data and to test the validity of the model. The system can be further described by its correlation dimension, a measure of its complexity. Another useful measure is entropy, which quantifies how well the system retains information. The Lyapunov exponents characterize how information is retained or lost by the system by examining how evolution trajectories change due to perturbations of the system state. The procedures for estimation of these measures from time-series is reviewed here. Furthermore, an attempt is made to develop algorithms for automatic evaluation of these measures from experimental data so as to minimize the need for human interpretation and to make the results of such methods less subjective and more consistent.

3.4.1 Testing for Determinism

Given time-series data observed from a continuous signal, before initiating some nonlinear analysis of the data it is important to test the hypothesis that nonlinear dynamics are present in the signal at a level significant enough to warrant examination. The basis for such a test is surrogate data analysis,

which is discussed in [135].

Definition 3.25 (Surrogate Data Test) *To test for a property P_1 in a time series $\{X(j)\}_{j=1}^N$ using surrogate data:*

1. *The null hypothesis H_0 is that the property P_1 is absent from $\{X(j)\}_{j=1}^N$*
2. *A set of M sequences $\{\{Y_k(j)\}_{j=1}^N\}_{k=1}^M$ are generated, such that the properties that guarantee consistency with H_0 are shared with $\{X(j)\}_{j=1}^N$. A value of $M = 30$ is sufficient.*
3. *A measure Z of P_1 is applied to the original data to get Z_0 and to the surrogates to obtain $\{Z_k\}_{k=1}^M$. If P_1 is present, then $E[Z_{k \neq 0}]$ should be different from Z_0 .*
4. *A student t -test is used to test H_0 , i.e. $E[Z_{k \neq 0}] = Z_0$*
5. *A result of $H = 0$ indicates that absence of P_1 cannot be rejected, and a result of $H = 1$ means that P_1 is present.*

To implement the test for determinism, surrogate time-series similar to the original in distribution, autocorrelation, and spectral density but otherwise random must be generated. A measure that quantifies determinism is then applied to the original and the surrogates. If the difference in the outcomes is statistically significant, then the hypothesis that the original data is deterministic cannot be rejected.

Two outstanding issues are how, if possible, one can generate the appropriate surrogate data, and which measure should be used that will effectively differentiate stochastic and deterministic data. These issues are discussed in [77], where the conclusion is that the results of the surrogate data testing, on simulated and real data sets, are dependent on the method being evaluated and its parameters, the algorithm for generating the surrogate data, and the type of the original data.

Various solutions to address these issues are examined in [123], which aims for the design of more robust surrogate data generation methods, iterative or otherwise. In that work, an efficient algorithm for generating iteratively refined surrogate data (IRSD) is presented, which preserves the autocorrelation, distribution, and power spectrum of the original signal. This technique preserves all linear stochastic properties in the data, but no non-linear or deterministic properties.

Algorithm 3.26 (Iteratively Refined Surrogate Data) *Given a time series $\{X(j)\}_{j=1}^N$, a surrogate data set that preserves linear temporal correlations and distribution can be generated by the following:*

1. Evaluate the magnitude $\{S(j)\}_{j=1}^N$ of the DFT of $\{X(j)\}_{j=1}^N$
2. Let $\{C(j)\}_{j=1}^N$ be the data set $\{X(j)\}_{j=1}^N$ sorted in ascending order
3. Define $\{R_0(j)\}_{j=1}^N$ by a random re-arrangement of $\{X(j)\}_{j=1}^N$
4. Let $\{Y(j)\}_{j=1}^N$ be the phase of the DFT of $\{R_i(j)\}_{j=1}^N$, and define $\{V(j)\}_{j=1}^N$ by $V(j) = Y(j)S(j)$
5. Compute the inverse DFT $\{Z(j)\}_{j=1}^N$ of $\{V(j)\}_{j=1}^N$
6. Compute $\{R_{i+1}(j)\}_{j=1}^N$ by replacing the rank ordering of $\{Z(j)\}_{j=1}^N$ with $\{C(j)\}_{j=1}^N$:

$$R_{i+1}(j) = C(\text{rank}(Z(j)))$$
7. Repeat steps 4 to 6 until $\sum_{j=1}^N (R_{i+1}(j) - R_i(j)) = 0$

This heuristic converges in a finite number of steps to a limit $\{R_\infty(j)\}_{j=1}^N$. The distribution is preserved since the result is a rearrangement of $\{X(j)\}_{j=1}^N$, and the linear correlation structure is retained since the frequency spectrum

$\{S(j)\}_{j=1}^N$ is invariant. The surrogate data will therefore have similar continuity, smoothness, and statistical properties as the original signal, with none of the nonlinear or deterministic aspects. More details may be found in [123].

As an example, 50000 points sampled at $\tau = .01$ are taken from the first state of the Lorenz system and used to generate a random surrogate. Segments from both sequences are shown below:

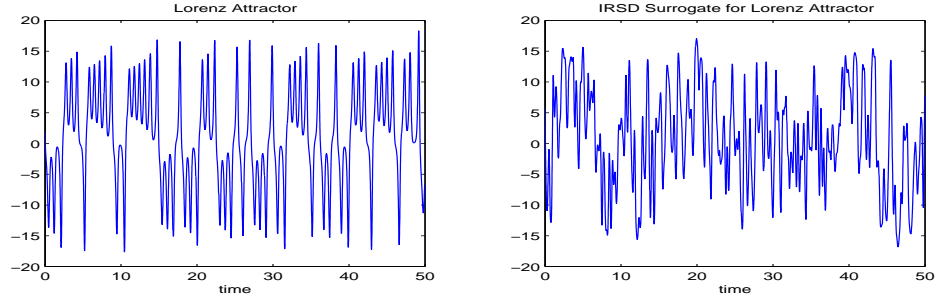


Figure 3.6: First State of Lorenz Attractor (left) and Iteratively Refined Surrogate Data (right)

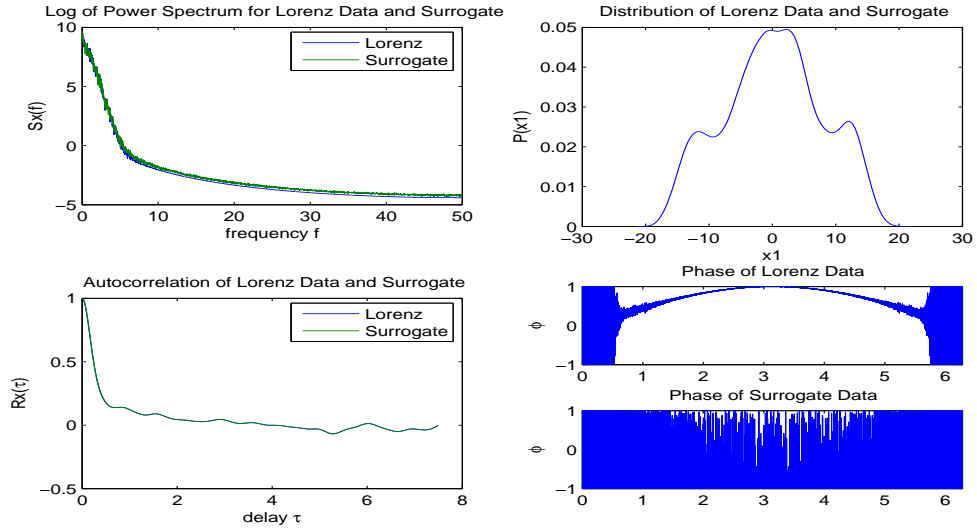


Figure 3.7: Log of Power Spectrum (top left), Distribution (top right), Autocorrelation (bottom left), and Comparison of Phase (bottom right) of Lorenz Data and Surrogate

All linear statistical properties of the signal are preserved. Note that the only difference in linear properties is the phase, as expected.

The above technique produces a signal with the same linear statistical properties as the original without preserving any nonlinear features. The surrogate data test is based on the idea that if a nonlinear feature is not different statistically between the original and the surrogate, the hypothesis that the original signal is dominated by a linear stochastic process cannot be rejected. The remaining issue is what measure to use.

The assumption here is that the theory of stochastic models involving nonlinear temporal relationships is not yet well enough developed for application in these algorithms. Testing for nonlinearity in time-series sampled from continuous physical processes should therefore be aimed to test for continuous deterministic dynamics. Any continuous physical process appears smooth if it is sampled at a high enough rate, so what is being examined must be more subtle than continuity, smoothness, or complexity. The measure used needs to quantify the variance of the relative temporal evolution of trajectories in the phase space. Such invariance would indicate nonlinear determinism, as opposed to the high variance in temporal evolutions that is present in linear stochastic signals.

A simple approach to such a measure is presented in [67], which involves a central tendency measure (CTM) that quantifies the variation in angle of successive directional vectors in the phase-space trajectory. The use of this measure with a surrogate data test provides a rapid, effective, and robust

way to test for determinism.

Algorithm 3.27 (Central Tendency Measure) *The central tendency measure (CTM) of a time-series $\{X(j)\}_{j=1}^N$ is given by the following:*

1. *Reconstruct the phase space $\{\vec{X}_\tau^m(j)\}_{j=1}^N$*
2. *Compute angles $\{R(j)\}_{j=1}^{N-1}$ between successive vectors $\vec{X}_\tau^m(j)$ where*

$$R(j) = \frac{\vec{X}(j+1) \cdot \vec{X}(j)}{|\vec{X}(j+1)| |\vec{X}(j)|}$$

3. *The CTM is given by*

$$CTM = \frac{1}{N-2} \sum_{j=1}^{N-2} \sqrt{(R(j+2) - R(j+1))^2 + (R(j+1) - R(j))^2} \quad (3.23)$$

Remark 3.28 (Second Order Difference Plot) *A second order difference plot (SODP) of a time-series of angles of directional vectors is a plot of $(R(j+2) - R(j+1))$ vs. $(R(j+1) - R(n))$. The CTM is a measure of the dispersion of the SODP that is independent of the variance of the signal.*

Note that if the CTM is used for a surrogate data test for determinism, a result of $E[Z_{k \neq 0}] < Z_0$ will also result in acceptance of H_0 , since the surrogates appear less random than the original.

The difference between a time-series observation of the Lorenz system, which is a continuous deterministic signal, and its surrogate, a stochastic signal with the same linear statistical properties, is illustrated in figure 3.8, using the same 50000 points:

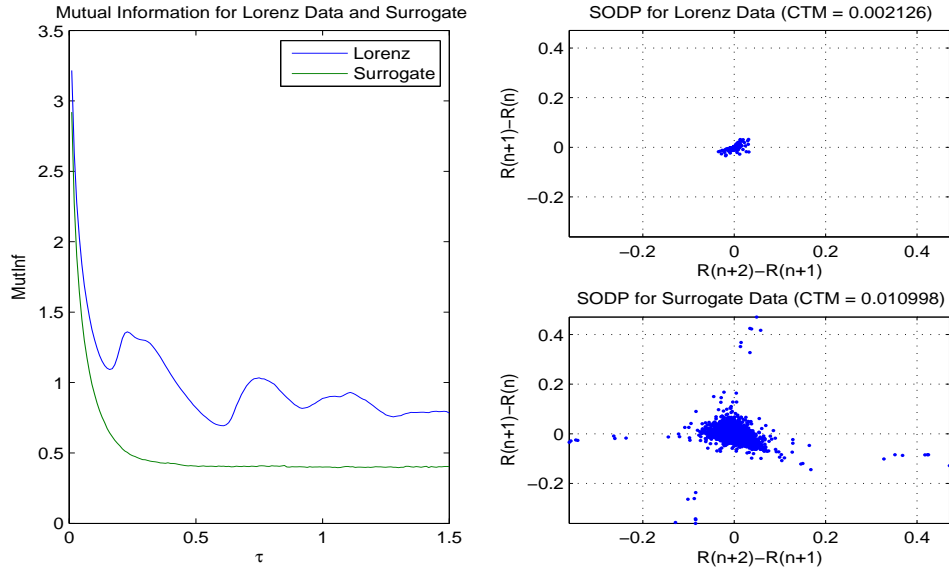


Figure 3.8: Mutual Information (left) and SODP plots (right) of Lorenz Data and Surrogate

Alternately, for a linear stochastic signal, specifically the filtered balanced compound linear fractional stable noise discussed in chapter 2, the results for the original and surrogate are much more similar. Note that since the signal has been smoothed by a moving average filter, so that its form is consistent with observations of a continuous process.

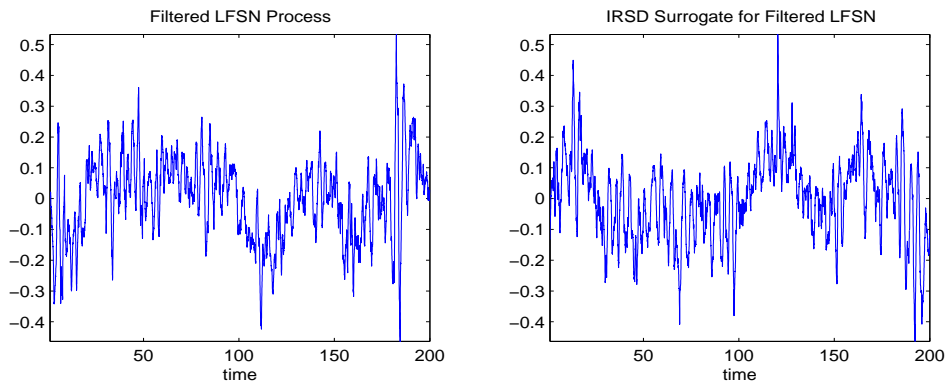


Figure 3.9: Integrated LFSN (left) and Iteratively Refined Surrogate Data (right)

In figure 3.10, note that the CTM is actually greater for the original data than the surrogate in this case, indicating that a stochastic process dominates the signal.

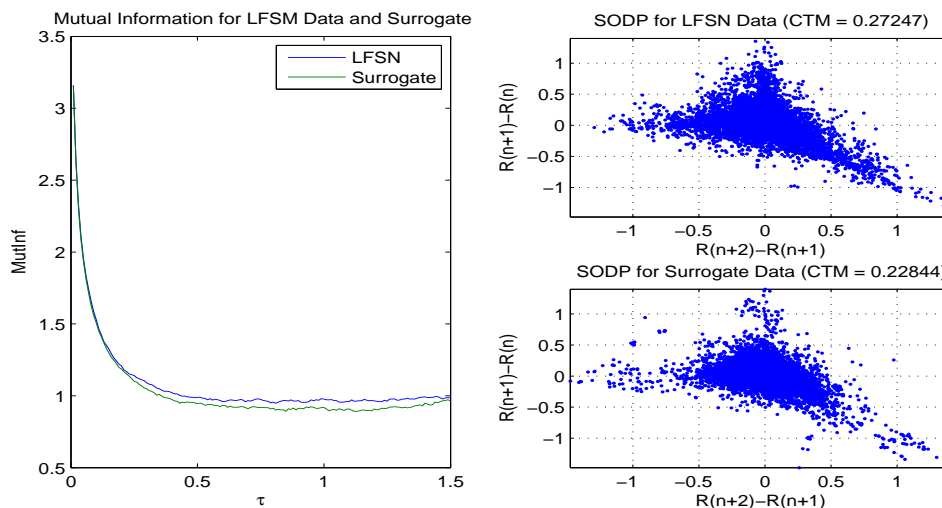


Figure 3.10: Mutual Information (left) and SODP plots (right) of Integrated LFSN Data and Surrogate

Observe that the mutual information of the original and surrogate signals is far more similar for integrated LFSN than in the case of the Lorenz data. The mutual information plots validate the results of the CTM-based method.

It is important to note that the CTM measure is a heuristic, and the effects of linear correlations in the signal on this measure are not known. Also, filtering, interpolation, and noise will affect the measure in undetermined ways. The phase-space reconstruction procedure may also affect the outcome, as well as the method used to generate the surrogate data. These issues are typical of surrogate data analysis, as described in [77]. The authors of [67] suggest that the test can be made more robust by replacing the student t-test

by an empirical test for a change in magnitude in the CTM. The authors suggest that the criterion $S_{CTM} = CTM_{surr}/CTM_{orig}$ can be considered, where CTM_{surr} is the CTM of the surrogates and CTM_{orig} is the CTM of the original. If S is significantly higher than 1, there are grounds on which to reject the null hypothesis. That is, for a value of $S > 2$, a deterministic model should be considered, values of $1.3 < S < 2$ could go either way, and values of $S < 1.3$ are inconclusive.

Another approach for detecting determinism, using nonlinear autoregressive filtering (NAR) to estimate Lyapunov exponents is presented in [22].

3.4.2 Correlation Dimension

One measure of the complexity of a fractal structure on a manifold in \mathbb{R}^n is the correlation dimension. The connection to time-series analysis is that the temporal evolution of a signal, under proper transformation, can represent a path on a manifold in \mathbb{R}^n that is defined either by a deterministic law or as a random walk. The complexity of this manifold is rigorously defined by the box-counting dimension D , given by covering the state-space in which the attractor lies by hypercubes of edge length ϵ , and counting the proportion of cubes containing a piece of the attractor as $\epsilon \rightarrow 0$. The correlation dimension gives a more convenient method to define and compute an estimate of complexity given an optimal phase-space reconstruction for the attractor as defined in section 3.3.

The correlation dimension was first defined as a measure of the complex-

ity of a chaotic attractor in [53], and is commonly referred to as D_2 . The measure is derived from the correlation integral, which is the distribution of the Euclidean distance between coordinates in the phase-space $\{\vec{X}_\tau^m(j)\}_{j=1}^\infty$. Furthermore, it well-known that D_2 is related to the box-counting dimension. Further reading on the correlation dimension and the related computational issues can be found in [79, 133, 127].

Definition 3.29 (Correlation Integral) *The correlation integral in embedding dimension m is*

$$C^m(r) \triangleq \lim_{N \rightarrow \infty} \frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N \theta(r - \|\vec{X}_\tau^m(j) - \vec{X}_\tau^m(i)\|) \equiv \int_0^r d^d r' c(\vec{r}') \quad (3.24)$$

Definition 3.30 (Correlation Dimension) *It is well-known that $C^m(r)$ behaves according to a power law for small r , and the correlation dimension D_2 is defined by the relation $C^m(r) \propto r^{D_2}$.*

Remark 3.31 (Theiler Window) *A correction was proposed by Theiler in [134] to mitigate the effects of local autocorrelation on the global value of $C^m(r)$ for small values of r . Coordinate vectors within W points of one another are to be eliminated from the integral calculation:*

$$C^m(r) \triangleq \lim_{N \rightarrow \infty} \frac{2}{(N+1-W)(N-W)} \sum_{j=W}^N \sum_{i=1}^{N-W} \theta(r - \|\vec{X}_\tau^m(j) - \vec{X}_\tau^m(i+j)\|) \quad (3.25)$$

Algorithm 3.32 (Correlation Dimension) *Given a phase-space embedding $\{\vec{X}_\tau^m(j)\}_{j=1}^{N-(m-1)\tau}$ of a time-series $\{X(j)\}_{j=1}^N$, the correlation dimension estimate D_2 can be obtained by the following:*

1. *The correlation integral $\{C_k^m = C^m(r_k)\}_{k=1}^K$ is computed for an exponentially spaced set of values $\{r_k\}_{k=1}^K$, with, for instance, $K = 300$.*
2. *The two-point derivative approximation of the sequence $\log(C_k^m)$ is used to select the longest section of the plot $\log(C_k^m)$ vs. $\log(r_k)$ for which the slope is statistically invariant, defined by k_1 and k_2 . The algorithm used for automatic statistical segmentation of sequences is discussed in chapter 4.*
3. *Linear regression is used to fit a line to the plot $\log(C_k^m(r))$ vs. $\log(r)$. The slope of this line gives the estimate \hat{D}_2*

This algorithm is applied to the Lorenz attractor data used previously to demonstrate the automatic implementation of the correlation dimension calculation. The Lorenz attractor is notable because its correlation integral fits the power-law model very well. 20000 points of the first state of the Lorenz attractor sampled at $dt = 0.01$ are examined. The value of $\tau = 16$, the first minimum of mutual information, is used for embedding delay, and $W = 65$, the first minimum of autocorrelation, is used as the Theiler correction. The results are shown in figure 3.11. The algorithm clearly detects the proper linear region in the correlation integral plot. The result of the method is $\hat{D}_2 = 2.04$, which differs slightly from the result of 2.05 obtained in [53] due to the Theiler correction and to proper selection of the regression region.

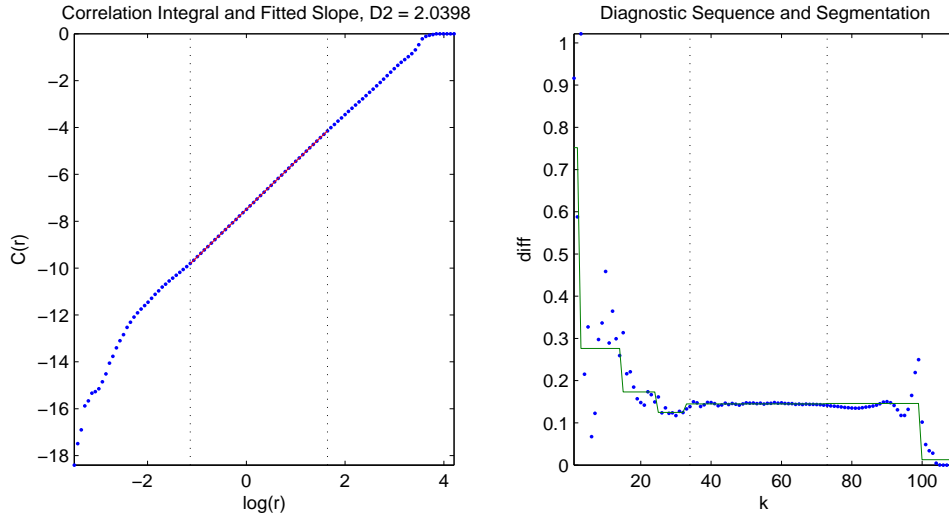


Figure 3.11: Correlation Integral and Fitted Linear Approximation (left) and Diagnostic Sequence for Identification (right)

It is important to note that the correlation dimension is highly sensitive to noise in the signal and linear correlations in the signal. The correlation integral measures the distribution of magnitude changes in the attractor, and is an essentially probabilistic measure. Therefore in the presence of random noise, the correlation dimension may reflect the properties of the noise rather than any determinism.

3.4.3 Kolmogorov Entropy

The Kolmogorov entropy is a measure of how the system retains information. This measure was applied to physiological signals in [43], along with other nonlinear measures to extract information that could not be obtained using linear techniques. An approximation to the Kolmogorov entropy using the measure D_2 was suggested by the authors of [53]. This approximation is referred to as K_2 -entropy, for which a rigorous definition involves a very

sophisticated topological formulation. It suffices to say that the measure quantifies the average increase in entropy, or randomness, of a process on a compact manifold in \mathbb{R}^n due to temporal evolution of the phase-space trajectories. The formulation of K_2 as an estimate of Kolmogorov entropy is intuitive as a heuristic in addition to being mathematically well-justified. The difficulty of its objective application is, as in the case of correlation dimension, a consequence of implementation issues.

Definition 3.33 (K_2 (Kolmogorov) Entropy) *The Kolmogorov Entropy estimate K_2 is given by*

$$K_2(r) = \log \left[\frac{C^m(r)}{C^{m+1}(r)} \right] \quad (3.26)$$

for a large value of m , such as 20 or 30.

The outstanding issue is finding a region of r for which the values of K_2 are statistically invariant.

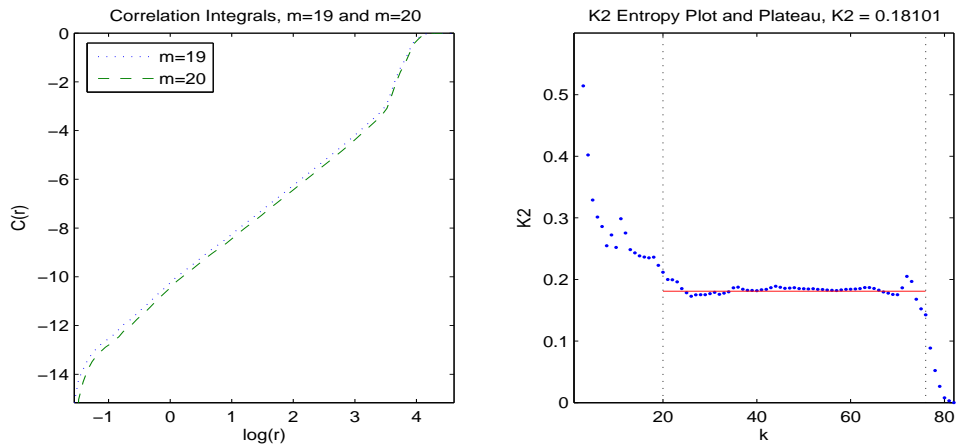


Figure 3.12: Correlation Dimension of Lorenz attractor for $m=19$ and $m=20$ (left) and Plateau Identification for Kolmogorov Entropy (right)

The technique is similar to algorithm 3.32, for which the detection algorithm will be described in the subsequent chapter. The algorithm is applied to 20000 points of the first state of the Lorenz attractor sampled at $dt = 0.01$, and the values of $\tau = 16$ and $W = 65$ are again used.

This technique automatically computes the required correlation integrals, entropy sequence, and detects the plateau on which the result is invariant. The value of K_2 entropy obtained for the Lorenz data is $K_2 = 0.1810$. Again, the Lorenz system is notable for the rapid convergence of its correlation integral to a power law, so the method requires testing on diverse signals.

3.4.4 Lyapunov Exponents

Another nonlinear measure that has been applied to physiological signals is the largest Lyapunov exponent. The Lyapunov spectrum of a dynamical system quantifies the sensitivity to perturbation of the system by evaluating the divergence of trajectories given a small perturbation in the initial conditions. Two approaches to computing the Lyapunov spectrum of a chaotic attractor were presented in [145] and [19], and the approach to extracting this measure from short time-series is found in [116]. Two notable applications to physiological signals are [55] and [43]. The meaning of numerical measures for time-series generated by solving nonlinear systems is straightforward, while it is not so clear when the time-series is a random process.

Definition 3.34 (Largest Lyapunov Exponent) *Let $\vec{X}_0(t)$ denote the initial distance between two nearby coordinates in the phase-space of an at-*

tractor, and by $\vec{X}(s)$ the distance between them after an evolution in time of $s - t$. Then the largest Lyapunov exponent L_1 of the attractor is computed by

$$L_1 = \frac{1}{t_m - t_0} \sum_{k=1}^m \log \frac{\vec{X}(t_k)}{\vec{X}_0(t_{k-1})} \quad (3.27)$$

The techniques for actual implementation of this method require sophisticated construction, and the results depend on the technique and parameters. In addition, the interpretation of the method applied to random time-series is not well-defined.

3.5 Discussion and Future Work

The fields of fractal and nonlinear time-series analysis encompass a wide array of methods for extracting information from indeterminate data. However, the results of such methods require substantial interpretation and are difficult to automate for the analysis of a large number of time-series segments. In this chapter, recent methods for statistical characterization of the dominant structure of a time-series were discussed. The significance and estimation of scaling, complexity, dimensionality, memory, and chaotic properties are examined. Most notably, an attempt is made to automate the interpretation of nuances in the application of these methods, with a focus on algorithm development for hierarchical analysis of large data sets.

Future work in this area must involve the testing and adaptation of the phase-space based methods for application to differently structured signals. Also, the estimation of fractal dimension and the Hurst parameter in the

presence of non-stationarity is an area of active current research. Though power-law scaling can be verified easily in simulated data, the behavior of physical systems is much more complex. The methods discussed above can be integrated and augmented to characterize richer behaviors and extract the maximum amount of information from measurement data.

Chapter 4

State Detection and Identification in Physiological Signals

The previous chapters focused on the extraction of diverse forms of information, referred to as *features*, from time-series data. The estimation of specific parameters, tests for choosing a proper qualitative model, and quantification of some absolute and relative properties were reviewed. This chapter introduces a framework for using those methods for automatic state detection and identification problems related to physiology and biomedicine. The chosen methodology and validation techniques are based on a broad review of methods in the literature. Certain algorithms described here are also applied to automate the computation of certain measures introduced in chapter 3.

A central premise in this chapter is that the features extracted from measurements of a physiological system can be highly variable in any given state. Therefore, discrimination among different states is problematic using features of only one epoch of an observed time-series, even though those features may

have statistically significant differences between the states in general. In order to overcome this difficulty, one can look a-posteriori for transitions in the features of a time-series. These transitions will then separate the time-series into segments, with invariant properties, for which one can assume that the system remains in one state. This process is referred to as segmentation. Separate states can thus be clearly detected and identified, and their properties can be evaluated. The process of grouping segments of a time-series with similar properties to correspond to one of a set of states is called clustering.

The problem of state transition detection can be simplified to the estimation of the most likely transition points, referred to as change-points, in the mean of a diagnostic sequence derived from a time-series of system observations. This is accomplished in this chapter using an algorithm for non-parametric statistical change-point detection based on a variant of the Kolmogorov-Smirnov statistic. That algorithm and standard algorithms for data partitioning are then applied to the problem of time-series segmentation and clustering. The use of this technique for state discovery is also discussed.

4.1 Review of Detection and Identification Methodology for Physiological Signals

The general state detection and identification problem in computational biomedicine can be formulated as follows.

Definition 4.1 (State Detection and Identification Problem) *Consider a physiological system in a clinical setting with two or more well-defined*

and qualitatively observable/identifiable pathological or regular states. Given sampled observations of measurable physiological signals from that system, develop an algorithmic method that automatically identifies and detects the clinical states.

The formulation here is intentionally vague, because such problems may involve a-posteriori, real-time, or predictive requirements, and could require the discovery of clinically unobservable states. Solutions are developed so that the output on training data corresponds to clinical identifications for that data. The diverse methods used to address such problems are reviewed below.

A first and very important step is, given a time-series of measurements and state classifications, to identify features of the data that exhibit statistically significant differences among states. A comparison of features based on different qualitative models is necessary to rigorously validate the relevance of chosen measures, and support clinically meaningful results. A straightforward way to do this is to separate the time-series into segments corresponding to clinical states, extract features, and compare the results from the diagnostic groups using analysis of variance (ANOVA). This analysis is used in [43] to identify features that are statistically significant for separation of sleep states. The notion of statistical significance of chosen features is mentioned in [90] with regards to seizure detection, with emphasis on the caveat that measures must also be medically significant.

Once features are chosen, they are applied to time-series to identify physiological states. This can be done using neural networks, as done for sleep-state identification [54, 55], or for detecting seizure onset [125]. Such methods determine optimal linear subspaces that divide the feature space into regions corresponding to each state. Other methods involve quantification of similarity between epochs and some training epoch [80], or dynamical entrainment of a feature at critical groups of electrode sites [62]. Heuristics and genetic algorithms are also used [28, 126].

Because the data in question is a time-series, temporal factors are important. Therefore robust identification requires the ability to segment time-series data into stationary or invariant epochs, and identify statistically significant temporal change-points in the features. One method employs a-posteriori change-point identification by iterative optimization using a model-based cost function [14]. A very robust and versatile approach involves non-parametric statistical methods [18]. Such methods are applied to sleep analysis in [73]. Time-frequency transforms can be used for segmentation as well, as in [25].

4.2 Change-Point Detection

Change-Point detection refers to the identification of significant changes in the properties of a time-series. This is accomplished by identifying statistically significant changes in the mean of a diagnostic sequence derived from the time-series. This section concerns the non-parametric statistical change-

point detection methods of Brodsky [17] and their implementation for physiological signals following [18, 73].

4.2.1 A-Posteriori Change Point Problem

Definition 4.2 (Piecewise Stationary Random Sequence) *Consider a piecewise constant function $f(\cdot)$ defined on the interval $[0, 1]$ by*

$$f(x) = \sum_{k=1}^K a_k \Theta(x - b_k) \quad \text{where} \quad \Theta(x) = \begin{cases} 1 & \text{for } x \geq 0 \\ 0 & \text{for } x < 0 \end{cases}$$

is the heaviside function, $a_k \in \mathbb{R}$ and $b_k \in [0, 1]$. Then a piecewise stationary random sequence $\{X(j)\}_{j=1}^N$ is defined by $X(j) = f(j/N) + \xi(j)$, where $\{\xi(j)\}_{j=1}^N$ is a sequence of zero-mean random variables. The points $\{j^(k)\}_{k=1}^K$ are the change-points of the sequence, defined by $j^*(k) = \lfloor b_k N \rfloor$.*

In definition 4.2, the sequence $\{X(j)\}_{j=1}^N$ is a random vector generated by a known $f(\cdot)$ and by the distribution of a known r.v. ξ such that $\{X(j) - f(j/N)\}_{j=1}^N \stackrel{d}{=} \xi$, where “ $\stackrel{d}{=}$ ” denotes equality in distribution. The detection of change-points is the inverse problem, where $f(\cdot)$ is unknown and needs to be estimated from a realization of the random sequence $\{X(j)\}_{j=1}^N$.

Regarding the sequence $\{\xi(j)\}_{j=1}^N$, it is assumed that it satisfies the mixing condition in [103], i.e. independence of time, has a finite moment generating function in some neighborhood of the origin, and there exists a limit

$$\lim_{n \rightarrow \infty} \frac{1}{n} E \left[\left(\sum_{j=1}^n \xi(j) \right)^2 \right] = \sigma^2 < \infty. \quad (4.1)$$

Observe that if $\{\xi(j)\}_{j=1}^N$ are i.i.d., then $\sigma^2 = \text{Var}(\xi)$.

Definition 4.3 (Change-Point Detection Problem) *Assume that $\{X(j)\}_{j=1}^N$ is a piecewise stationary sequence as in definition 4.2. Then provided that one or more change-points are present, the sequence of change-points $\{j^*(k)\}_{k=1}^K$ is to be estimated. Further, the piecewise constant function $f(\cdot)$ that best corresponds to $\{X(j)\}_{j=1}^N$ is to be recovered by forming a sequence $\{\hat{f}(j)\}_{j=1}^N$ from $\{X(j)\}_{j=1}^N$ and $\{j^*(k)\}_{k=1}^K$ such that $f(j/N) = \hat{f}(j)$.*

The solution developed by Brodsky [17] and others to this problem requires lengthy and complex mathematical derivations. However, the resulting technique, based on the *Brownian bridge*, makes sense heuristically as well.

Definition 4.4 (Brownian Bridge) *Define by W the Wiener process, or the Brownian motion process, on the space of continuous functions on $[0, 1]$, and with the well-known properties $E[W_t] = 0$ and $E[W_s W_t] = s$ if $s \leq t$. Then, the Brownian bridge W^O is defined by $W_t^O = W_t - tW_1$, with*

$$\begin{aligned} E[W_t^O] &= 0 \\ E[W_s^O W_t^O] &= s(1-t) & \text{if } s \leq t, \\ E[(W_t^O - W_s^O)^2] &= (t-s)(1-(t-s)) & \text{if } s \leq t \end{aligned} \quad (4.2)$$

Observe that W_t^O is ‘tied-down’ such that $W_0^O = W_1^O = 0$ with probability 1.

Furthermore, the Brownian bridge has the notable property that

$$F_{KS}(C) \triangleq P\left(\sup_t |W_t^O| \leq C\right) = 1 + 2 \sum_{k=1}^{\infty} (-1)^k \exp\{-2k^2 C^2\} \quad (4.3)$$

where $C > 0$ and $F_{KS}(\cdot)$ denotes the Kolmogorov-Smirnov distribution.

Further information on the properties of the Brownian bridge and other stochastic processes can be found in [13]. The presence of a change-point in a sequence is detected within the framework of hypothesis testing based on this model.

4.2.2 Statistical Methods

Proposition 4.5 (Test for Presence of Change-Points) *For a random sequence $\{X(j)\}_{j=1}^N$ as in definition 4.2, the null (homogeneity) hypothesis H_0 is that the sequence is stationary with a constant mean, and has no change-points. Consider the family of statistics*

$$Y_N(j, \delta) = \left[\left(1 - \frac{j}{N} \right) \frac{j}{N} \right]^\delta \left[\frac{1}{j} \sum_{k=1}^j X(k) - \frac{1}{N-j} \sum_{k=j+1}^N X(k) \right] \quad (4.4)$$

where $\delta \in [0, 1]$ and $j \in [1, N-1]$. It is clear that if $E[X] = m_1$ for $j \leq j^*$ and $E[X] = m_2 \neq m_1$ for $j \geq j^*$, then $\arg \max_j |Y_N(j, \delta)| = j^*$. Then for some given threshold C (to be defined later), if $|Y_N(j, \delta)| \geq C$ then H_0 is rejected and j^* is selected as a change-point of $\{X(j)\}_{j=1}^N$.

Equation 4.4 is a variant of the Kolmogorov-Smirnov statistic, used for testing the equality of distributions of two samples. At each point j in the time series, $Y_N(j, \delta)$ is the difference in means of $\{X(k)\}_{k=1}^j$ and $\{X(k)\}_{k=j+1}^N$, weighted by a function $[(1 - j/N)(j/N)]^\delta$.

Proposition 4.6 (Convergence to Brownian Bridge) *It can be shown [17, 26] using functional limit theorems for random sequences [103]*

that the statistic $\sqrt{N} \cdot Y_N(\lfloor Nt \rfloor, 1)$ converges, with $N \rightarrow \infty$, as a random process to σW_t^O . Here W^O is the Brownian bridge described in definition 4.4, and σ is defined by equation 4.1 and the mixing conditions in [103].

Remark 4.7 (False Detection Probability and Threshold) From proposition 4.6, it follows that the threshold C for the test in proposition 4.5 and the false detection probability P_{fa} of the test are related by

$$\begin{aligned} P_{fa} &= \lim_{N \rightarrow \infty} P \left(\max_j \sqrt{N} |Y_N(j, \delta)| > C \right) = \\ &= P \left(\sup_t |W_t^O| > C/\sigma \right) = 1 - F_{KS}(C/\sigma) \end{aligned} \quad (4.5)$$

or $C = \sigma F_{KS}^{-1}(1 - P_{fa})/\sqrt{N}$, where F_{KS} is the Kolmogorov-Smirnov distribution and N is the length of the test sequence.

The test in proposition 4.5 can therefore be calibrated by adjusting the parameters P_{fa} and δ (in equation 4.4). It is shown in [17] that a value of $\delta = 0$ minimizes the probability of false tranquility, i.e. ignoring an existing change point, and that $\delta = 1$ minimizes the probability of false detection, i.e. identifying a nonexistent change-point. A value of $\delta = 0.5$ guarantees minimum estimation error for the location of a change-point. The parameter P_{fa} determines the sensitivity of the test since it defines the threshold. A low value of P_{fa} , i.e. 0.01, will detect only very significant changes in the sequence $\{X(j)\}_{j=1}^N$, while a higher value of P_{fa} , i.e. 0.1, will make the test more sensitive.

4.2.3 Description of Algorithm

The detection technique described in [18] and [73] consists of three stages, also detailed here. Given a time-series $\{X(j)\}_{j=1}^N$ and a sensitivity defined by P_{fa} (a default value is 0.1), estimation of change-points is as follows.

Algorithm 4.8 (Stage 1: Preliminary Estimation)

1. Compute threshold C for the given P_{fa} and $\sigma = 1$ using remark 4.7
2. Compute the statistic $|Y_N(j, 1)|$ and set $j_1 = \arg \max_j |Y_N(j, 1)|$.
3. Split $\{X(j)\}_{j=1}^N$ into two sequences, $Z_1(j) = X(j)$: $1 \leq j \leq j_1$, and $Z_2(j) = X(j)$: $j_1 < j \leq N$. Form the sequence $\{Z(j)\}_{j=1}^N$ defined by

$$Z(j) = \begin{cases} Z_1(j) - \mu_1 & \text{for } j \leq j_1 \\ Z_2(j) - \mu_2 & \text{for } j > j_1 \end{cases}$$

where μ_1 is the mean of Z_1 and μ_2 is the mean of Z_2 . Set $\hat{\sigma} = \text{std}(Z)$

4. If $|Y_N(j_1, 1)| \leq C\hat{\sigma}/\sqrt{N}$, the homogeneity hypothesis is accepted, and the procedure is completed. If $|Y_N(j_1, 1)| \geq C\hat{\sigma}/\sqrt{N}$, j_1 is designated as a candidate change-point, and the procedure continues with step 5
5. Repeat steps 1 to 4 for both Z_1 and Z_2 until the homogeneity hypothesis is accepted for all sub-segments.
6. Stage 1 will output an ordered set $\{j_k\}_{k=1}^{K_1}$ of K_1 candidate change-points

Algorithm 4.8 can be implemented rapidly using recursion.

Algorithm 4.9 (Stage 2: Rejection of Doubtful Change-Points)

Given an ordered set of candidate change-points $\{j_k\}_{k=1}^{K_1}$,

1. Separate $\{X(j)\}_{j=1}^N$ into K_1 sub-samples X_k as follows:

$$\begin{aligned} X_1 &= X(j) : & 1 \leq j \leq j_1 + (n_2 - n_1)/2 \\ X_k &= X(j) : & j_{k-1} + (j_k - j_{k-1})/2 \leq j \leq j_k + (j_{k+1} - j_k)/2, \\ & & \text{for } 2 \leq k \leq K_1 - 1 \\ X_{K_1} &= X(j) : & j_{K_1-1} + (j_{K_1} - j_{K_1-1})/2 \leq j \leq N \end{aligned}$$

2. Perform the homogeneity test for each sub-segment analogous to steps 1 to 4 in algorithm 4.8 using a lower false alarm probability, i.e. $P_{fa}/10$, and $\delta = 0$. If the homogeneity hypothesis is accepted for X_k , then j_k is removed from the set of candidate change-points.
3. Stage 2 will output an revised set $\{j_k\}_{k=1}^{K_2}$ of $K_2 \leq K_1$ candidate change-points

Algorithm 4.10 (Stage 3: Final Estimation of Change-Points)

1. Separate $\{X(j)\}_{j=1}^N$ into K_2 sub-samples X_k as in step 1 of algorithm 4.9 using the revised set of candidate change-points $\{j_k\}_{k=1}^{K_2}$
2. For each sample X_k , compute $|Y_N(j, 1/2)|$, and set $j_k^* = \arg \max_j |Y_N(j, 1/2)|$ as the final estimate of the location of the k^{th} change point
3. Stage 3 will output the final set $\{j_k^*\}_{k=1}^K$ of $K = K_2$ locations of change-points in the sequence $\{X(j)\}_{j=1}^N$
4. The estimate $\{\hat{f}(j)\}_{j=1}^N$ of the function $f(\cdot)$ that generates $\{X(j)\}_{j=1}^N$ can be estimated by generating the sub-samples

$$\begin{aligned} X_1 &= X(j) : 1 \leq j \leq j_1^* \\ X_k &= X(j) : j_{k-1}^* \leq j \leq j_k^* \text{ for } 2 \leq k \leq K-1 \\ X_K &= X(j) : j_{K-1}^* \leq j \leq N \end{aligned}$$

and computing

$$\hat{f}(j) = \begin{cases} \bar{X}_1 & \text{for } 1 \leq j \leq j_1^* \\ \bar{X}_k & \text{for } j_{k-1}^* \leq j \leq j_k^* \\ \bar{X}_K & \text{for } j_{K-1}^* \leq j \leq j_K^* \end{cases} \quad (4.6)$$

where \bar{X}_k is the arithmetic mean of X_k .

The results of algorithms 4.8 through 4.10 optimally estimate the solution to the change-point detection problem as formulated in definitions 4.2 and 4.3. This entire procedure has been combined and implemented in MATLAB, as part of this work, to rapidly detect change points in i.i.d. piecewise sequences. Algorithm 4.8 is implemented using recursion, so that computational time

for the procedure is not an issue. A simple illustration of the output for a piecewise stationary signal is shown in figure 4.1.

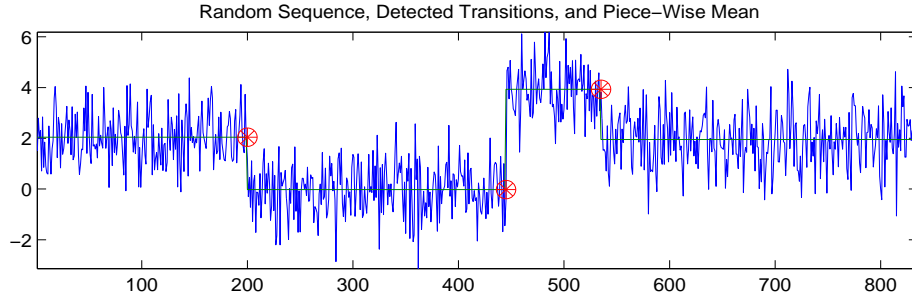


Figure 4.1: Example of Change-Point Detection: Random Sequence, Piece-wise Mean, and Change-Points (red stars)

This algorithm is quite powerful, because it can detect subtle transitions that may not be evident visually but are still statistically significant, as shown in figure 4.2.

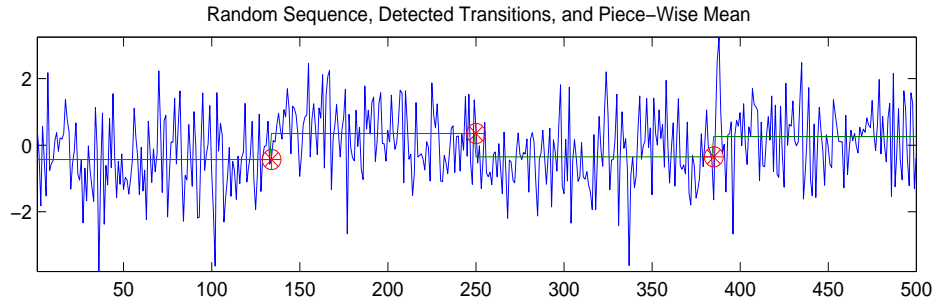


Figure 4.2: Example of Change-Point Detection in Subtle Case: Random Sequence, Piecewise Mean, and Change-Points (red stars)

In addition, changes in distribution can be detected by finding changes in the mean of a diagnostic sequence obtained by squaring the centered signal.

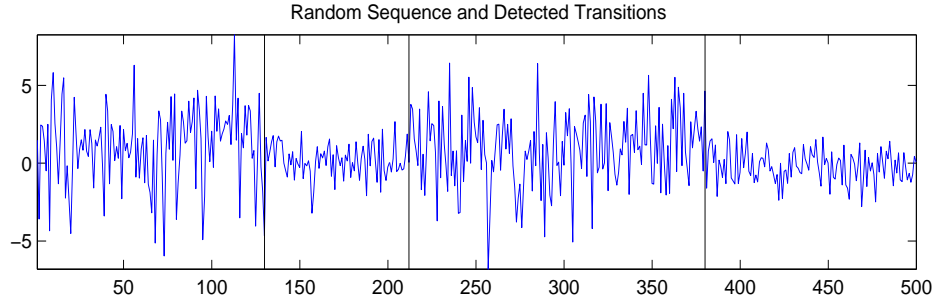


Figure 4.3: Example of Change-Point Detection in Distribution: Random Sequence and Change-Points (lines)

4.3 Segmentation and Clustering

This section reviews the methodology for separating physiological signals into epochs in which observable properties remain invariant, which is referred to as *segmentation*. This is done by feature extraction and change-point detection. Also, the procedure of grouping invariant segments into groups belonging to the same state, referred to as *clustering*, is examined.

4.3.1 Feature Extraction and Segmentation

An effective way to examine the variation of a feature for a time-series is to partition the time series into sub-segments and evaluate that feature for each sub-segment. This process is called *feature extraction*, or the creation of a *diagnostic sequence*. An important first step is selection of the size of the sub-segments, a rather subjective choice that based on several criteria. First, evaluation of the feature examined must be meaningful for that amount of data. Second, the variability of the feature for segments of that size within a given state are not so great as to confound the identification of transitions.

Third, changes on that temporal scale are relevant for the application.

For example, clinical examination of the EEG is conducted on the scale of minutes, and a state can be defined by a clinical examiner with relative certainty for 30-120 seconds depending on the application. Thus, for the problem of sleep-state identification a practical sub-segment length is 30-60 seconds.

The next step is the extraction of diagnostic sequences.

Algorithm 4.11 (Extraction of Diagnostic Sequences) *Consider a feature $G(\cdot)$ that is defined by some mapping of a vector $\{x_j\}_{j=1}^n$ to a scalar g , i.e. $G : \{x_j\}_{j=1}^n \mapsto g$. Given a time-series $\{X(j)\}_{j=1}^N$ sampled at a frequency f_s , a diagnostic sequence with period T_g for a given feature is obtained:*

1. *Define the sub-sequence length by $L = f_s T_g$*
2. *Define a sequence of sub-sequences X_i , $i = 1, 2, \dots, M$, by*

$$X_i(j) := X(j + L(i - 1)), 1 \leq j \leq L$$

where $M = \lfloor N/L \rfloor$ is the length of the diagnostic sequence

3. *Extract the feature for each sub-sequence by $g_i = G(\{X_i(j)\}_{j=1}^L)$ to obtain a diagnostic sequence $\{g_i\}_{i=1}^M$*

The same procedure can be done for a family of features, and a diagnostic sequence can be formed in turn from some function of those features.

The problem of segmentation of physiological signals is, assuming that a feature is stationary within a given state, to separate the signal into segments for which the given feature is stationary. Once a feature $G(\cdot)$ that can discriminate among states in the sequence is found, i.e. by testing for statistical significance between features extracted from data in known states,

the process is straightforward. Algorithm 4.11 is used to create a diagnostic sequence $\{g_i\}_{i=1}^M$, which is in turn processed by algorithms 4.8 to 4.10 to detect change-points. The change-points in the diagnostic sequence determine transition points in the original data. The results of the procedure should be verified visually to determine if meaningful segmentation has been achieved, i.e. that detected transitions correspond to changes in clinical state.

4.3.2 Clustering and State Identification

If meaningful segmentation is achieved for a data set, the invariant segments of the data can then be separated into groups corresponding to observed states S_j . This can be done by a clustering or partitioning algorithm applied to the piecewise constant sequence $\{\hat{f}(j)\}_{j=1}^M$ extracted from the diagnostic sequence $\{g_i\}_{i=1}^M$ after change-point detection, in the sense of definitions 4.2 and 4.3. Consider the K-means clustering algorithm:

Algorithm 4.12 (K-means Clustering) *A set of N data points $\{x_i\}_{i=1}^N$ can be clustered into K disjoint subsets S_j containing N_j data points by minimizing the sum-of-squares criterion*

$$J = \sum_{j=1}^K \sum_{i \in S_j} |x_i - \mu_j|^2$$

where μ_j is the geometric centroid of the data points in S_j . Starting with a random assignment of the data points to K sets, the centroid for each set is computed, and every point is re-assigned to the cluster whose centroid is

closest to that point. This is repeated until there is no further change in the assignment of data points.

The K-means algorithm does not in general achieve a global or even local minimum of J over the assignments, however it is rapid and easily implementable. A version of the algorithm is given in [57], which is implemented in MATLAB as `kmeans.m` [88].

Clustering algorithms will yield far more consistent results for the extracted piecewise sequence \hat{f} than for the diagnostic sequence $\{g_i\}_{i=1}^M$ for any feature, by construction. Within a given invariant segment the diagnostic sequence will likely be variable, and may even take values that are in the range observed in other states. The sequence \hat{f} is by definition constant within an invariant segment, so that values of the points \hat{f}_i in the same state S_j will be less variable, and thus closer to the centroid μ_j . This leads to the following procedure for identification problems in physiology.

Algorithm 4.13 (State Identification) *Consider a system with K known states S_k , for which a feature $G(\cdot)$ extracted from sub-segments of some observed time-series $\{X(j)\}_{j=1}^N$ exhibits statistically significant differences among the K states. Assume $\{X(j)\}_{j=1}^N$ is sampled with frequency f_s and the feature $G(\cdot)$ is meaningful for periods of length T_g time units. Then the invariant epochs and corresponding states can be identified by the following procedure, known as state identification.*

1. *Use algorithm 4.11 to create a diagnostic sequence $\{g_i\}_{i=1}^M$ of the feature $G(\cdot)$ for sub-segments $\{X_i(j)\}_{j=1}^L$, with $L = f_s T_g$*

2. Use algorithms 4.8 to 4.10 to detect change-points in $\{g_i\}_{i=1}^M$ and create a piecewise-constant sequence $\{\hat{f}_i\}_{i=1}^M$ of means of $\{g_i\}_{i=1}^M$ within the invariant segments
3. Apply algorithm 4.12 to partition $\{\hat{f}_i\}_{i=1}^M$ into K groups corresponding to states S_k , to produce a sequence of state identifications $\{\hat{s}_i\}_{i=1}^M$ corresponding to each sub-segment $\{X_i(j)\}_{j=1}^L$, where s_i is an integer and $s_i \in [1, K]$
4. If the elements of a subsequence $\{\hat{s}_i\}_{i=a}^b$, $a < b$, are of the same value $k \in [1, K]$, then the sub-sequence $\{X\}_{j=1}^r$ of the original time-series given by

$$\tilde{X}(j) = X(j + (a - 1)L), \quad 1 \leq j \leq (b - a + 1)L$$

corresponds to system state S_k

5. It follows that the system is in the state S_k for the time period $t \in [(a - 1)T_g, bT_g]$, measured from the start of observation.

The procedure in algorithm 4.11 can also be done for a set of features $\{G_k(\cdot)\}_{k=1}^K$ and a set of diagnostic sequences $\{g_i^k\}_{i=1}^M$, for $k = 1, 2, \dots, K$. These sequences are then processed in step 2 separately, and in step 3 together as a set of vector coordinates.

4.4 Discussion and Future Work

The procedure detailed in algorithm 4.13 provides a powerful tool for automatically making decisions about the state of a system for which states are well-defined. The methodology creates a framework for automatic application of the feature extraction methods described in chapters 2 and 3. This will be applied in chapter 6 for automatic sleep-state identification in neonates. Furthermore, the change-point detection algorithm presented in

section 4.2 can be applied to a wide variety of a-posteriori detection problems, and extended to the multi-dimensional case.

In addition to state identification, the procedure in algorithm 4.13 can be applied with little modification to determine what states are present in a system, in the case that they are not observable qualitatively. This procedure is called *state discovery*. Further investigation of non-parametric statistical methods for change-detection, decision, and a-priori prediction methods is warranted. Application to other areas of science and engineering surely exist, for example in traffic modeling, process control, array signal processing, and pattern recognition.

Chapter 5

Analysis of EEG Structure

A fundamental issue in EEG analysis that has remained unresolved concerns the nature of electrical activity in the brain. Specifically, the matter in question is whether EEG signals reflect a stochastic or a deterministic process. A great deal of literature has been devoted to examining the properties of the EEG for both academic and clinical purposes. Much work has also been done concurrently to develop models intended to produce signals structured similarly to the EEG. A review of the literature is given in chapter 1, though it represents a small sampling of the large body of work done in this field.

Despite all of the investigation into the properties of the EEG, there is no consensus regarding a comprehensive theory explaining its nature. The issue is whether stochastic or deterministic components are dominant in the EEG. It remains a matter of ongoing research. The aim in this chapter is to attempt to solve that fundamental question. The results are due to examination of newly obtained 1000 Hz measurements directly from the surface of the cortex of human patients. The findings clarify questions about the nature

of the signal, which in turn guides the development of appropriate analysis techniques for specific clinical applications.

5.1 Introduction

The models for the EEG that have been presented previously are derived from very different branches of mathematics and the physical sciences. Some models, such as that in [99], are based on the theories of neural fields and pattern and wave formation. This theory centers on an analytical approach used to examine, in a probabilistic sense, the electrical fields produced given certain spatial and temporal distributions and correlations in neuronal firing rates. Recent models use a deterministic dynamical system [60] or Kalman filtering [48] to simulate the EEG. Other models, such as that of [92], are based on ideas from statistical physics, with the assumption that the EEG is a stochastic process.

Fundamental analysis of properties of the EEG on short time-scales was done in [36] and [130], with the conclusions that a hypothesis of normality and stationarity on short intervals cannot be rejected. The analysis was done in each case using scalp measurements at low sample rates, and the conclusions are not applicable over time ranges greater than a few seconds. In the last decade, some investigation has been done on the clinical applicability in EEG analysis of nonlinear time-series methods, for example [114, 115]. The presence of nonlinear dynamics in the EEG has also been examined in [136] and [100]. In both cases, the authors concluded that their findings were not

consistent with the hypothesis of low-dimensional nonlinear chaos.

In his work, Palūs [100] concludes that a theory combining deterministic oscillations with a stochastic process may be the best explanation for the structure of the EEG signal. The use of measures based on dynamical systems theory is somewhat criticized, and models based on nonlinear stochastic processes are suggested. Furthermore, Palūs strongly advises that a study of the EEG should include an analysis using linear methods and hypothesis tests using surrogate data to demonstrate the relevance and necessity of any nonlinear methods used to quantify the phenomena being examined.

The investigation in this chapter is very much in agreement with the suggestions of Palūs. The focus is on making qualitative conclusions about the structure of the data, not clinical findings. The first step is a standard linear statistical analysis of the data, which is extended to include an examination of memory, stationarity, distributional, and scaling properties of the signal. Next a hypothesis test is conducted to test for determinism, decide whether any nonlinear analysis is necessary or beneficial, and conclude which analysis may be relevant. Finally, an explanation for the behavior of the EEG signal, that is consistent with the theories of self-organized systems and neuronal networks, is presented, and future work is discussed.

5.2 Data Sets and Assumptions

The data used for investigation includes a set of 46 intra-cranial 98-channel EEG recordings sampled at 1000 Hz from a study of 8 epileptic patients.

These recordings are of measurements from electrodes directly on the cortex of the patient. Another set used consists of 116 14-channel sleep-EEG recordings sampled at 64 Hz from a study of neonatal development. These recordings are measured from electrodes on the scalp. The data were provided by Dr. Mark Scher of the Department of Pediatrics and Dr. Mary Ann Werz of the Department of Neurology at the Case School of Medicine.

The 98-channel data is filtered using a 200 Hz low-pass filter. Grids of electrodes in meshes are implanted directly onto the cortex of the patient by a neurosurgeon in normal and pathological areas.

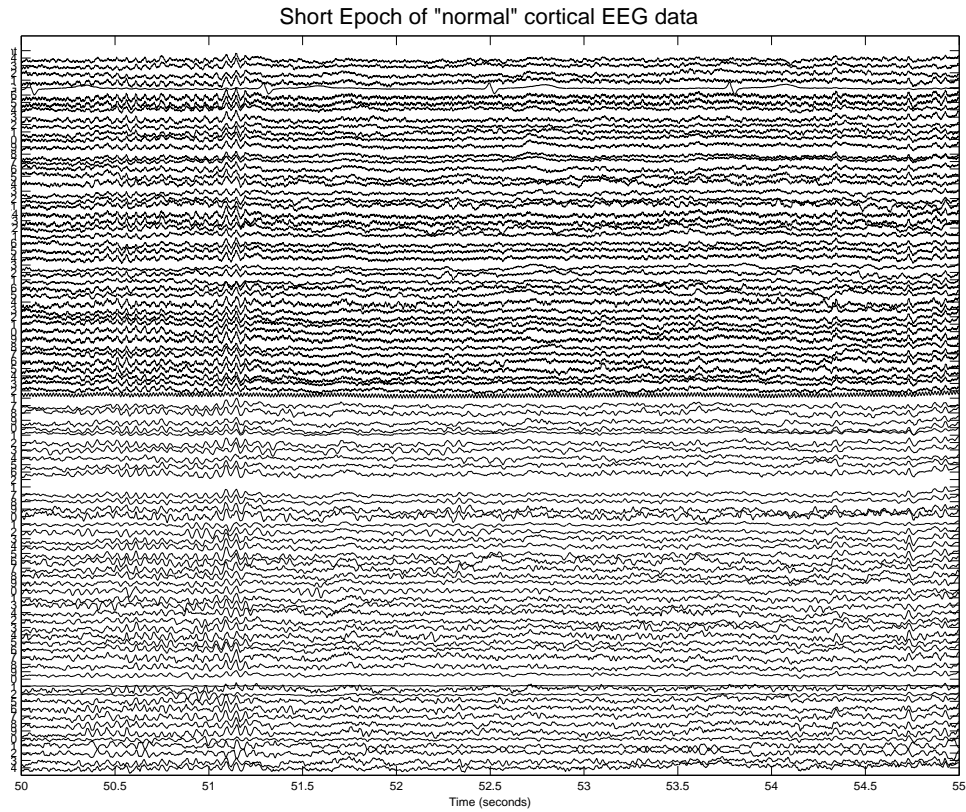


Figure 5.1: Epoch of Cortical EEG Data, 5 Second Interval

The 14-channel data is filtered using an anti-aliasing 32 Hz low-pass filter and a 60 Hz notch filter, and the electrodes are arranged in the international 10-20 system, shown in figure 5.2.

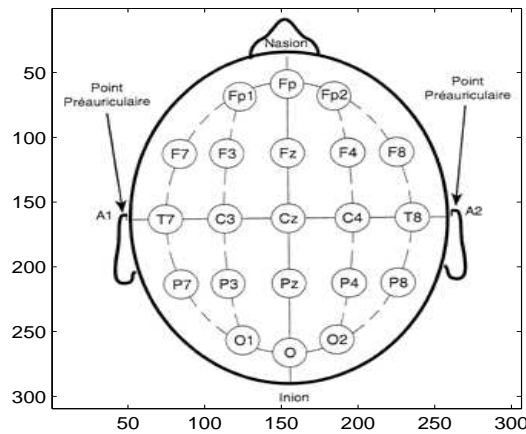


Figure 5.2: International 10-20 Electrode System (Image obtained from http://www.sfar.org/sfar_actu/ca97/html/ca97_002/97_02-3.gif)

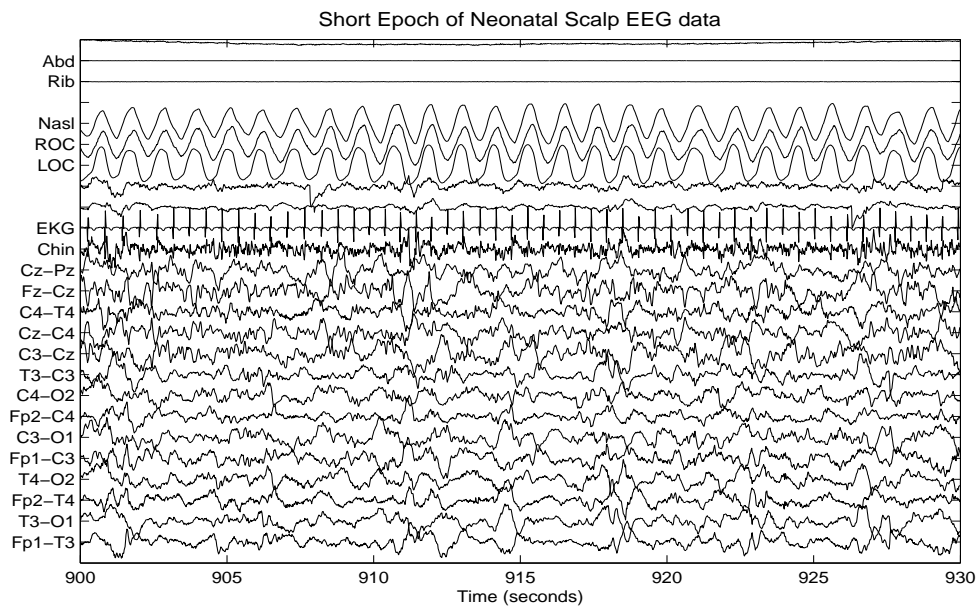


Figure 5.3: Epoch of Neonatal Scalp EEG Data, 30 Second Interval

Note that the cortical data contains a great deal of information that is very difficult to interpret visually. The scalp data is sampled at a much lower rate and contains fewer channels, but is still difficult to assess. A practical way to interpret the data in a format that is more readily viewed is to examine a set of diagnostic sequences. Once a feature is related to a clinical state or event, viewing a diagnostic sequence is a more practical way to understand the behavior of EEG records over time.

The assumptions are as follows. First, the EEG is the measurement of a collection of continuous processes, i.e. the electrical potential between pairs of locations on the cortex or scalp, sampled at some frequency f_s . Because the change in electrical potential on the scale in question is a continuous process, the digital EEG recording approximates a smooth, continuous signal for a sufficiently high sample rate. It is clear that for a high enough sample rate on some small interval, any continuous signal will appear deterministic, and for a large, arbitrary sample rate, a deterministic signal may appear random. It follows that analysis must be done on the data after re-sampling at a relevant time scale, and if the sampling rate is too low, analysis that requires a high rate must be foregone.

It is unlikely that a healthy individual would volunteer for unnecessary brain surgery, so it is evident that EEG recordings measured directly from the cortex can only be obtained for pathological cases. Thus an analysis of the structure of a normal cortical EEG cannot be undertaken in the strict sense. To overcome this obstacle, the EEG of pathological subjects is as-

sumed “normal” in the absence of clinically observed pathological activity. That is, the EEG of an epileptic patient resembles normal EEG activity in the absence of a seizure, and will be subsequently referred to as normal. A more precise classification is “non-ictal”, referring to the absence of seizure activity.

A final assumption is that the electrical potentials measured from the scalp are significantly distorted with respect to the subtle electrical activity measurable from the cortex. It is assumed that the 1000 Hz cortical data, albeit from pathological subjects, best reflects the structure of the EEG, and the results are verifiable using scalp EEG data.

5.3 Linear Statistical Analysis

A great deal of established clinical precedent exists for visual identification of various behaviors in the EEG. These behaviors are referred to as wave morphology, and include evaluation of spatial distribution, reactivity, various rhythms and frequencies, spikes, cyclic alternating patterns, and evoked potentials. Such visual analysis is not discussed here. Automatic computer analysis in this field has been largely limited to time-frequency analysis, heavily dependent on the Fast Fourier Transform. An excellent review by Lopes da Silva of the traditional theory and clinical practice is given in chapter 61 of [94]. In this section the linear statistical properties of the EEG are examined, using methods including time-frequency analysis.

5.3.1 Correlation and Power Spectrum

Local and long-range correlations are examined using the methods in section 3.1, as well as the distribution of power in the frequency spectrum. The local ($t \in [0, 1]$ seconds) a.c.f. of normal cortical EEG data is shown in figure 5.4. It is computed using 10 seconds (10000 samples) for each channel.

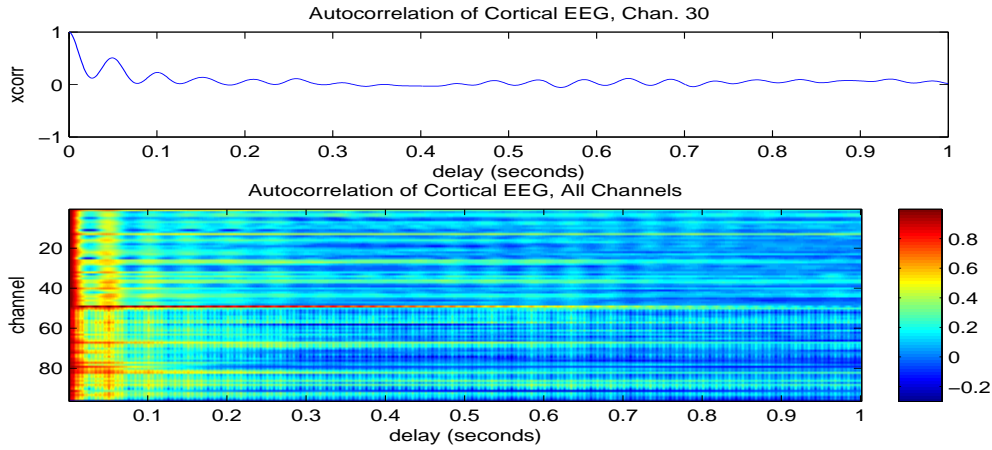


Figure 5.4: Local (1 sec.) Autocorrelation of Cortical EEG Data: Channel 30 (top) and All channels (bottom)

The bottom plot in figure 5.4 is a convenient way to display the value of a large family of functions. The domain is time, the range is the set of channels, and the color of the plot represents magnitude as indicated in the color bar at right. The channels numbered 50 and greater are contaminated by measurement noise due to equipment error, which was not removed by filtering. The long-range ($t \in [1, 10]$ seconds) a.c.f. is shown in figure 5.5. It is computed using 30 seconds (30000 samples) for each channel.

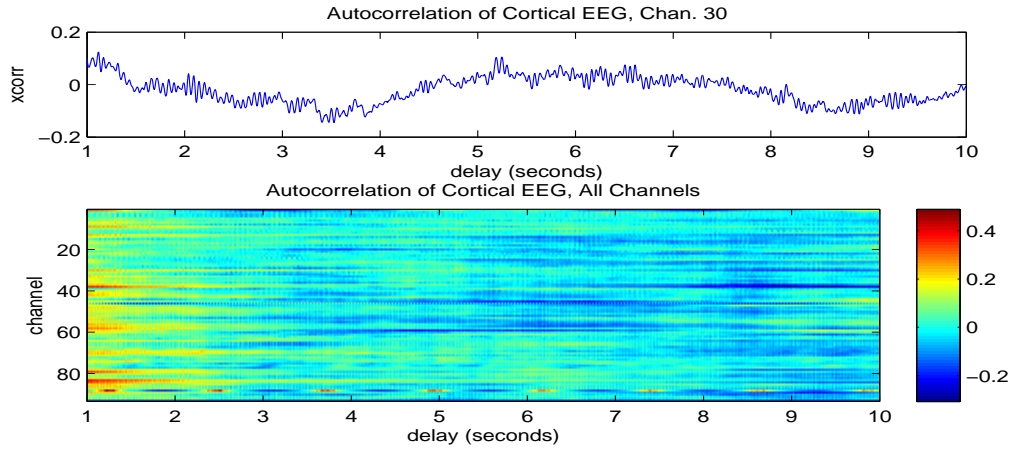


Figure 5.5: Local (10 sec.) Autocorrelation of Cortical EEG Data: Channel 30 (top) and All channels (bottom)

The local a.c.f. of the EEG is a highly variable property, as shown in figure 5.6. 10 second epochs are used to approximate the a.c.f. every 2 seconds. The domain is time, the range is the delay, and the color is a.c.f. value.

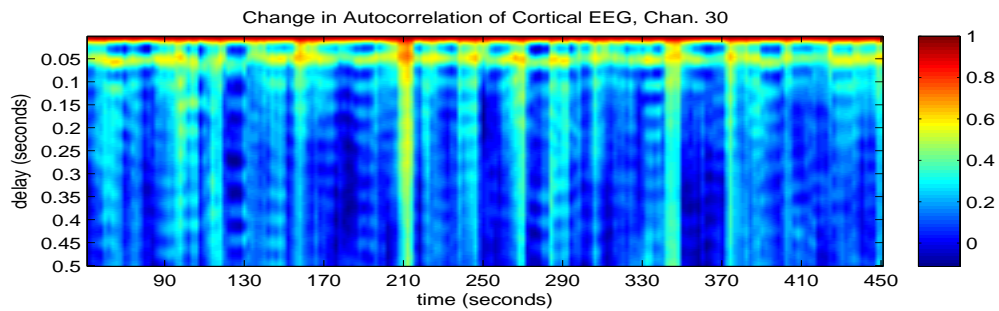


Figure 5.6: Change in Autocorrelation of Cortical EEG Data

The same variability is evident in the neonatal data, for which the same analysis is shown in 5.7 using 10 second segments to estimate the a.c.f. for each second.

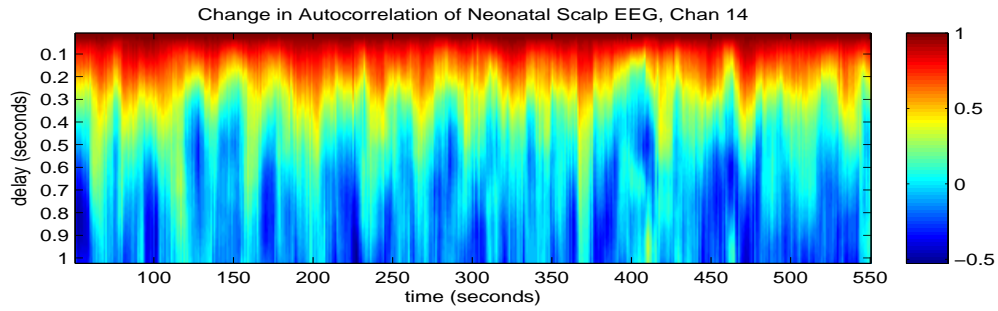


Figure 5.7: Change in Autocorrelation of Neonatal Scalp EEG Data

The periodogram estimate of the spectral density for 10 seconds of cortical EEG data is shown in figure 5.8. Note the spike at 60 Hz in channels 50 and greater, which is due to electrical interference during data collection.

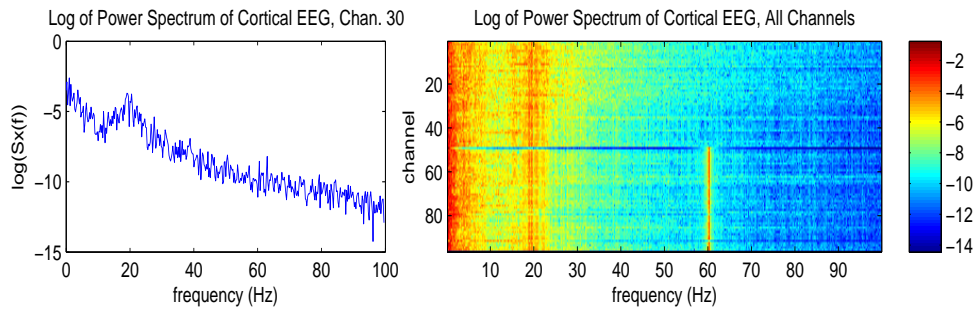


Figure 5.8: Log of Power Spectrum of Cortical EEG

Figure 5.9 shows the log of the power spectrum, estimated every 2 seconds using a 30 second epoch, over a period of 400 seconds.

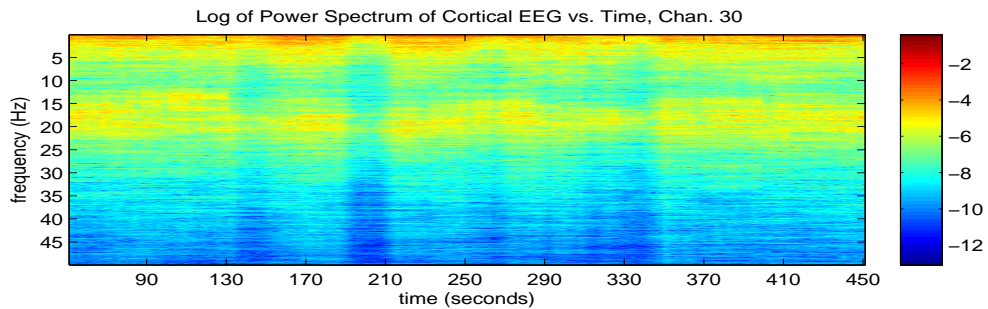


Figure 5.9: Log of Power Spectrum of Cortical EEG over Time

The power spectrum estimate varies dynamically, just as the a.c.f., because the two are directly related. The power is concentrated in the lower frequencies ($f < 30$ Hz). It is important to note that the spectral density can vary greatly depending on the clinical state of the patient, i.e. various sleep states, conscious activities, and pathological conditions.

5.3.2 Long Memory and Non-Stationarity

The presence of long memory has been verified in EEG recordings for various states. In [81], a.c.f.s of very long epochs of band-filtered EEG are examined. Notable scaling properties are observed in different frequency bands, and statistically significant correlations are found at delays of up to 200 seconds for certain frequencies.

Such an analysis is not undertaken here. It suffices to examine the a.c.f. of one channel of a long (100 second) epoch of cortical EEG data (figure 5.10) to see that the signal has long memory properties.

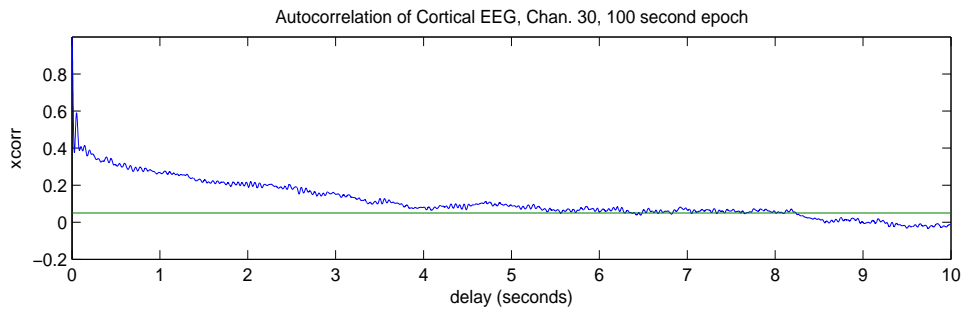


Figure 5.10: Autocorrelation of Large Epoch of Cortical EEG Data

The long memory property can be verified quantitatively by applying the self-affinity measures described in section 3.2. The Hurst parameter

H is approximated by estimating the fractal dimension D and computing $H = 2 - D$ (call this the FD method), and also by applying the DFA method. This is done for 30 50-second epochs of seizure-free cortical EEG data and for 30 100-second epochs of neonatal scalp EEG data. Box-plots of the measures are shown for both groups in figure 5.11.

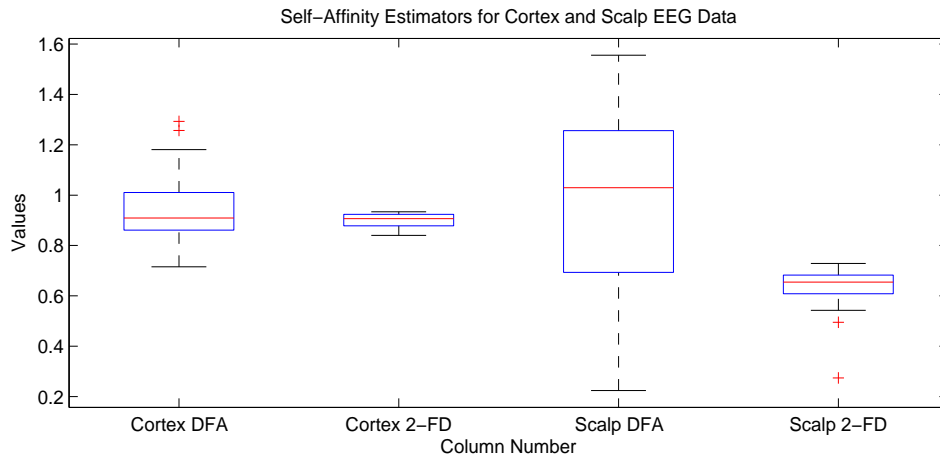


Figure 5.11: Self-Affinity Estimators for Cortex and Scalp EEG Data

Recall that the Hurst parameter, with permissible values $H \in [0, 1]$, is near 1 for high self-similarity, and near 0 for low self-similarity. Similar behavior on various scales indicates the presence of long memory. The results support the hypothesis that the EEG is a long memory process, since both the DFA and FD results are close to 1 for the cortical data. Their means and standard deviations are 0.946 ± 0.136 and 0.900 ± 0.028 , respectively. The DFA of the scalp data samples has a standard deviation of 0.37, and is not significant, while the FD measure remains a consistent estimator, with mean and standard deviation of 0.633 ± 0.086 . The latter is closer to 0.5, indicating low memory. This suggests that subtle long-memory processes are reflected

in cortical measurements that are not present in scalp measurements. These conclusions are consistent with the paradigm of brain activity as a complex self-organized network, where long-memory and hierarchical self-similarity are inherent.

One property that the EEG lacks is stationarity, as shown in figure 5.12. The signal is non-stationary in mean, variance, and as seen in section 5.3.1, correlation and spectral power characteristics as well.

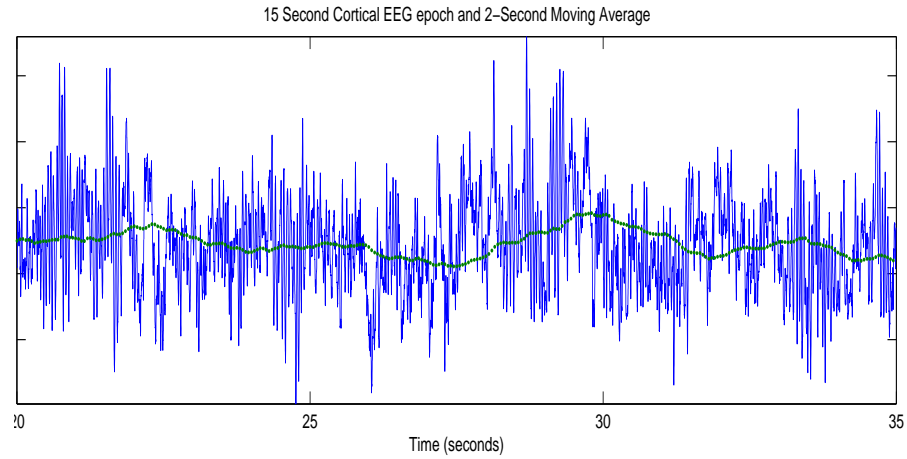


Figure 5.12: 15 Second Cortical EEG epoch (line) and 2-Second Moving Average (dots)

Non-stationarity renders certain measures difficult to quantify, for example the Hurst parameter, and is the reason for the high variability in DFA. The analysis of long-range dependence in the presence of non-stationarity is addressed in [131]. However, the performance of estimators based on the variogram appear to resolve this issue to some extent, and are far less computationally intensive. Furthermore, the non-stationarity may be a regular feature of the EEG, indicating multi-scaling of self-similarity properties. This

conjecture is supported by the high power in the low-frequency band.

5.3.3 Distribution of Signal and Increments

When considering a stochastic process as a model for real data, important properties to examine are the distributions of the signal and of increments in the signal. The term “increments” here refers to a change in amplitude for a given temporal difference. The fundamental work of testing for normality and stationarity in the amplitude of EEG signals has been done by Elul [36] and Sugimoto [130]. They concluded that short, 2-second, epochs of the EEG are Gaussian and stationary. Over longer epochs the distribution of the EEG is irregular and highly variable, so little further investigation has been done in that direction.

The hypothesis examined in this section is that, because the EEG is the measurement of a complex self-organizing system, its increments are distributed with, approximately, an α -stable distribution. The tails of the actual distribution will be truncated because the EEG is bounded. The amplitude distribution of the EEG is not expected to be α -stable, because the functions of the brain are not completely random. Also, since the EEG is a continuous process, the amplitude cannot jump instantaneously.

The increments of the signal should be close to $S\alpha S$ (symmetric α -stable). This assumption is due to the fact that cortical measurements reflect the number of neurons firing near the electrode site at any given time. The electrical potential at a site on the cortex is thus a sum of a large and random

number of random variables reflecting local potentials in small groups of neurons, which are in turn sums of potentials in single neurons, which are also random. This assumption relates well to the paradigm of brain activity as interactions within a complex self-organized network.

Using the estimation methods in chapter 2, the parameters of the α -stable distribution are fitted to a 40 second (40000 point) epoch of cortical EEG data, and 625 seconds (40000 points) of scalp EEG data. The Kolmogorov-Smirnov density estimate, α -stable fit and parameters, and Gaussian fit are shown in figure 5.13.

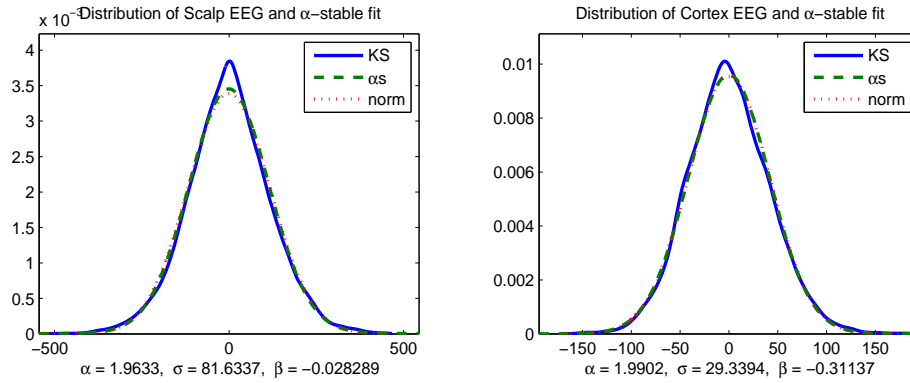


Figure 5.13: Distribution of EEG Amplitude and α -stable fit: Scalp (left) and Cortex (right)

The α -stable fits are only slightly better by visual inspection than the Gaussian fit, and do not pass goodness-of-fit (g.o.f.) tests as described in section 2.1.8 at a 5% significance level. The empirical density estimate appears almost triangular, and the results are similar for longer epochs.

Now, the increments of the same data used in figure 5.13 are fitted to the α -stable distribution. The time step used for computing differences is the

first minimum or zero-crossing of the a.c.f. for an epoch, so that the data are as linearly uncorrelated as possible (see section 3.3.1). The time step used is 0.027 seconds (27 points) for the cortical data, and 0.422 seconds (27 points) for the scalp data. The a.c.f.s for the epochs are shown in figure 5.14.

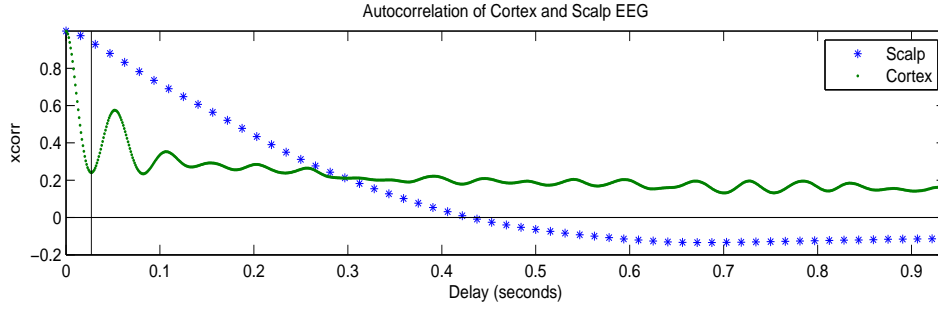


Figure 5.14: Autocorrelation of Cortex (dots) and Scalp (stars) EEG Data

The first minimum of the a.c.f. for the EEG measured from the cortex is marked, as well as the first zero-crossing of the a.c.f. for the scalp EEG.

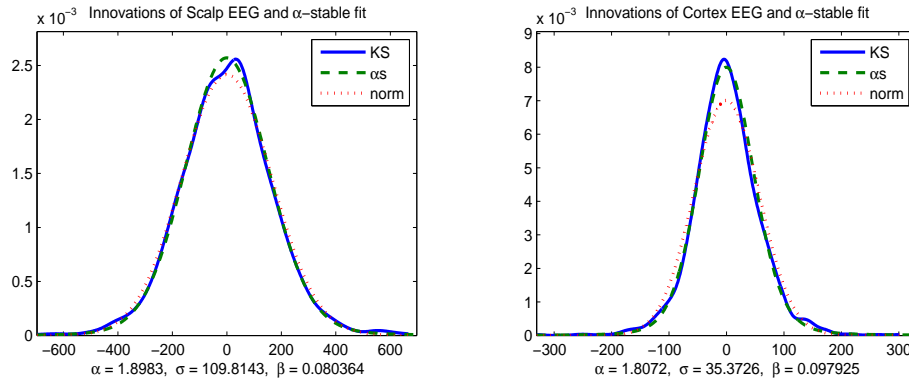


Figure 5.15: Distribution of EEG Increments and α -stable fit: Scalp (left) and Cortex (right)

Upon visual inspection, the α -stable distribution is a good fit for the distribution of increments in the cortical EEG, and is considerably different

from the normal fit, since $\alpha = 1.8$. That fit also passes the Kolmogorov-Smirnov and Anderson-Darling g.o.f. tests at a 5% significance level. The fit for the scalp EEG, however does not, which is surprising because the autocorrelation appears more comparable to that of a random process than that of the cortical EEG. It is nevertheless a good model in that case as well.

The 0.027-second increments of 45 non-overlapping, 50-second, seizure-free epochs of cortical EEG data from 4 patients are fitted with α -stable distributions, and each passes g.o.f. tests successfully at a 5% significance level. Box-plots of the estimates of α and σ are shown in figure 5.16.

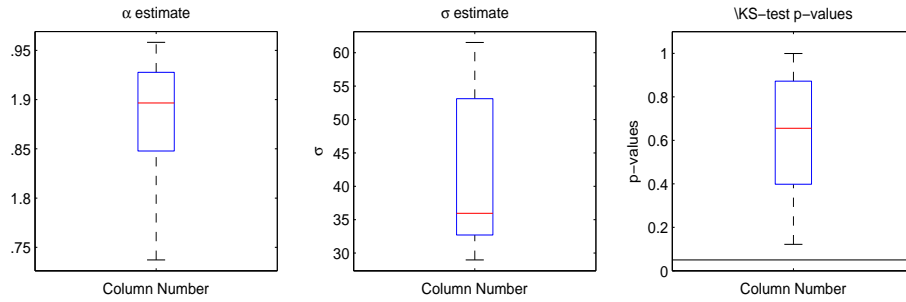


Figure 5.16: α -Stable Parameter Estimates for EEG Increments: α (left) and σ (right)

Note that the shape parameter α is 1.896, which is considerably different from the Gaussian case ($\alpha = 2$). It can be concluded with confidence that the increments of the cortical EEG are α -stable with truncated tails. This leads to the conjecture that α -stable impulsive noise could play a significant role in the generation of neuronal activity.

5.3.4 Self-Similarity Properties

The scaling properties of the EEG were briefly discussed in section 5.3.2. The high values of the DFA and fractal dimension estimators of the Hurst parameter indicated high self-similarity, and thus long memory. This section examines whether the EEG is indeed locally self-similar by examining the scaling of the distribution of increments for increasing steps in time. Specifically, the change in the scale parameter σ of the α -stable parameter estimate can be examined.

Consider a stochastic process $X(t)$ that is self-similar with stationary, symmetric, α -stable increments (recall definition 2.22). That is, the increments $\Delta X(h) \triangleq (X(t+h) - X(t))$ satisfy $\Delta X(h) \stackrel{d}{=} S_\alpha(\sigma_0, 0, 0)$. Such a process also has the property $\Delta X(nh) \stackrel{d}{=} n^{1/\alpha} \Delta X(h)$, as follows.

$$\begin{aligned} \Delta X(nh) &= X(t+nh) - X(t) \\ &= \sum_{i=1}^n (X(t+ih) - X(t+(i-1)h)) \\ &\stackrel{d}{=} \sum_{i=1}^n \Delta X_i(h) \\ &\stackrel{d}{=} n^{1/\alpha} \Delta X(h) \end{aligned} \tag{5.1}$$

where $\Delta X_i(h) \stackrel{d}{=} \Delta X(h)$ are i.i.d., and the last step is due to corollary 2.7.

The increments $\Delta X(h)$ are H -sssi with self-similarity index $H = 1/\alpha$.

It was shown in section 5.3.3 that the cortical EEG has α -stable increments for time-steps of about 0.027 seconds. The following analysis investigates the scaling properties of the EEG signals over small time-intervals, and examines whether a 20-second epoch of cortical EEG is locally H -sssi. The scale parameter σ , as estimated by algorithm 2.18, is shown for time

increments from 0.001 to 0.1 seconds. In addition, the same analysis is done for an H -sssi process simulated by cumulative summation of a sequence of α -stable random numbers with the same shape parameter $\alpha \approx 1.9$ as the EEG data. The relationship expected for the H -sssi model, $\sigma/\sigma_0 = n^{1/\alpha}$, due to equation 5.1, is shown for comparison. The same result is also displayed on a logarithmic scale, for which the relationship between the time increment and scale should be linear, along with a regression fit. The results are given in figure 5.17.

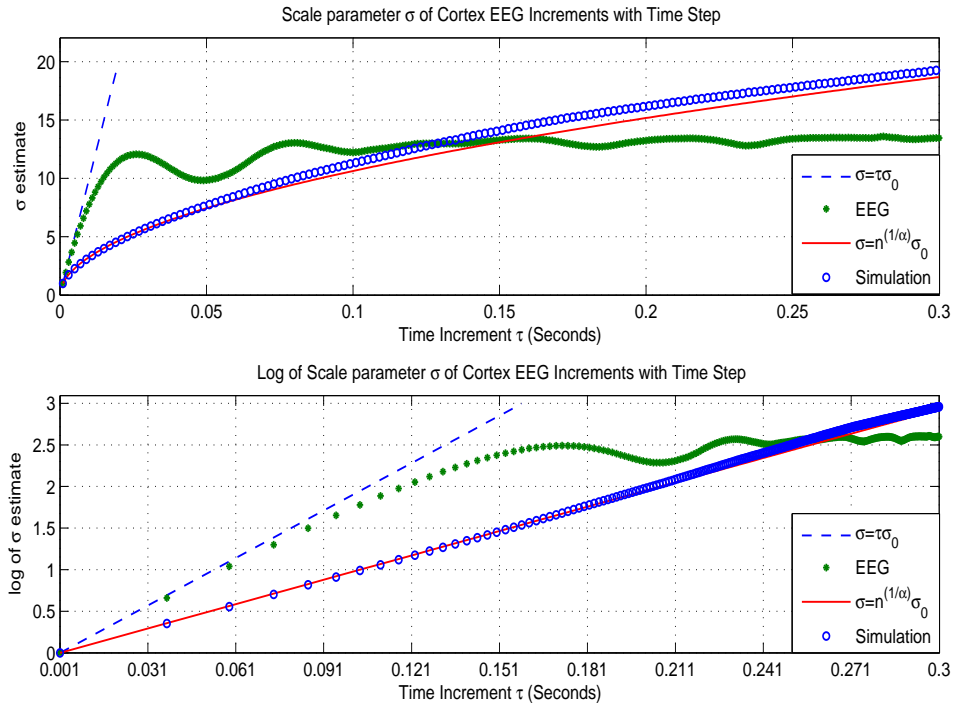


Figure 5.17: Scale parameter σ of Cortex EEG Increments and Simulation with Time Step (top) and on log scale (bottom)

It is evident that the scaling of the EEG increments as measured from the Cortex does not follow a power law. For small time increments the

scaling is actually linear, which indicates a value of $H = 1$, however the self-similarity index decreases quickly to $H = 0$, since the signal is bounded. Therefore one may conclude that the H -sssi model is not appropriate for the EEG locally or globally. The change in the index H leads to the conjecture that a nonlinear law governs the local behavior of the EEG, and that the nonlinearity is stochastic, because as shown in section 5.3.3, the distribution of increments is α -stable. Scaling properties are also examined in the scalp EEG of neonates, and the results are shown in figure 5.18.

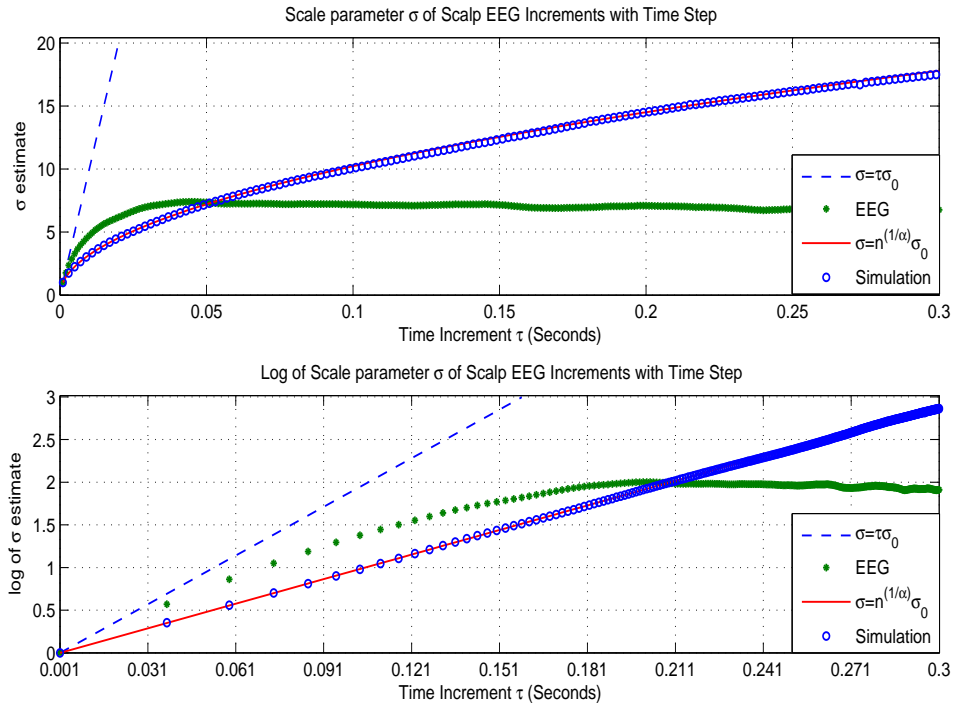


Figure 5.18: Scale parameter σ of Scalp EEG Increments and Simulation with Time Step (top) and on log scale (bottom)

From these results it follows that the EEG is not strictly self-similar. However, the self-affinity may be present in other observable features, as

suggested in [81].

5.4 The Nonlinearity Question

In this section an attempt is made to examine the fundamental question of the presence and nature of nonlinearity in the EEG in the framework of hypothesis testing. In literature related to electrical bio-physiology, “nonlinearity” is often used in the context of modeling the behavior of a process as a deterministic, low-dimensional nonlinear dynamical system. Theiler [136] examined the justification for this model to represent the EEG without conclusive results, and suggests that the paradigm of a dynamical system is too narrow to explain its behavior. It is difficult to demonstrate the relevance and necessity of nonlinear measures for real data. Theiler suggests that nonlinear measures may not be informative even when used to examine relative differences or changes.

The question of whether measures of nonlinear deterministic signal properties are relevant in EEG analysis can be investigated by testing whether nonlinear determinism is present in the EEG. The method presented in section 3.4.1 was shown to be very sensitive to the presence of any determinism in a signal. The surrogate data test for determinism using the central tendency measure (CTM) is applied to the same 45 epochs of EEG that were used in section 5.3.3 to examine the distribution of increments. An embedding dimension of $m = 7$ is used, and the time-delay τ is the first zero-crossing or minimum of the autocorrelation, which is computed auto-

matically for each segment. Both the student t-test and empirical criterion $S_{CTM} = CTM_{surr}/CTM_{orig}$ are used for the test. Box-plots of the criterion S_{CTM} for the cortex and scalp EEG epochs are shown in figure 5.19.

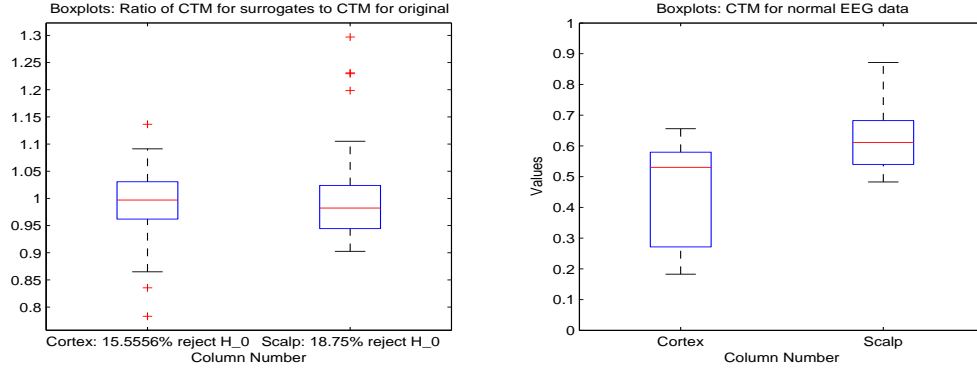


Figure 5.19: Box-plots of Empirical CTM Criterion for EEG Data (left) and CTM values (right); cortex (left) and scalp (right) for each plot

The null hypothesis H_0 that there is no determinism is rejected in only 15% of the cases for cortex data and 18% of the cases for scalp data. That is, in 85% of cortical EEG epochs examined, the hypothesis that the signal was produced randomly could not be rejected. This is evidence that the normal EEG is in general stochastic.

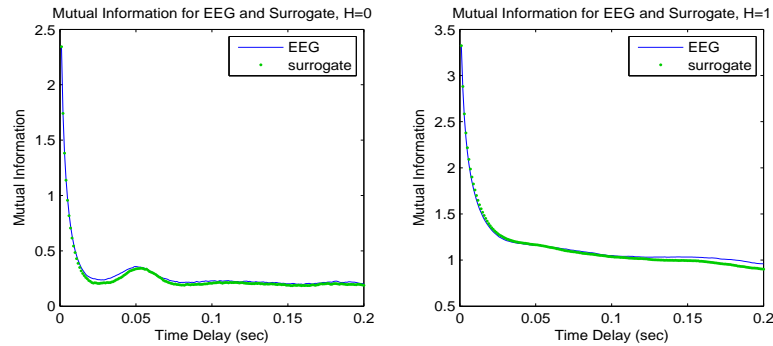


Figure 5.20: Mutual Information of EEG data and Surrogate Data: Accept H_0 (left) and Reject H_0 (right)

The CTM technique, like the correlation dimension, is also sensitive to random noise in unknown ways. To support the above results, the mutual information for EEG data and corresponding surrogates is compared in figure 5.20. Even in the case where determinism cannot be rejected using the surrogate data test with the CTM, the mutual information for the EEG data is not very different from that of surrogate data. This is shown for all the EEG epochs examined by computing $S_{MI} = MI_{EEG}(\tau)/MI_{surr}(\tau)$ for delay values $\tau \in [0.001, 0.2]$. The distribution of these values is shown in figure 5.21.

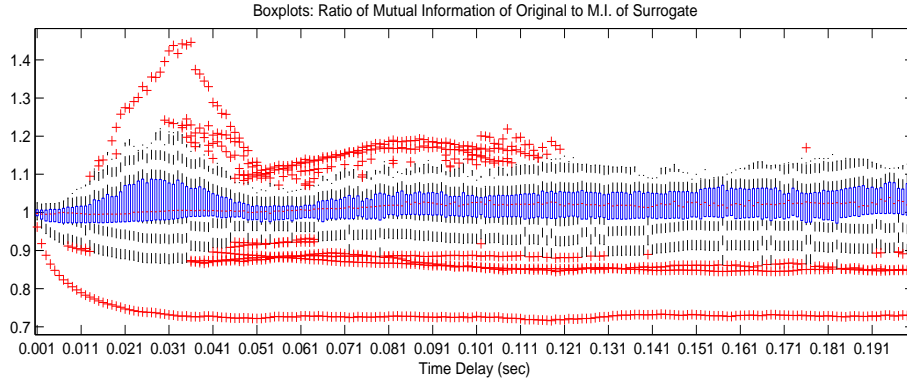


Figure 5.21: Box-plots: Ratio of Mutual Information of EEG to that of Surrogates

The results show that the mutual information of the EEG data and the surrogate data is similar in the majority of cases for all delay values. This is strong additional evidence that any given epoch of the EEG is generated by a random process, and not simply the output of a deterministic dynamical system.

An analysis of nonlinearity in EEG recordings was done using alternate

methods in [113], focusing on discerning low-dimensional nonlinear determinism from non-stationarity, and similar results were obtained.

5.5 A Model for the Normal EEG

Various models have been developed for generating signals similar to the electrical activity of the brain, as referenced in 5.1. The emphasis in these works is on reproducing the background rhythms, rather than on evoked potentials and more complex processes. An alternative model for the same process is presented in this section, using the results of the investigations in this chapter.

It was established in section 5.3.1 that activity of the EEG falls primarily in the low-frequencies, but that the spectrum is highly variable with time. In section 5.3.2, it was shown that the EEG has long memory and is a non-stationary process. In section 5.3.3, it was demonstrated that the increments of the EEG are from the α -stable distribution, and fit the model of impulsive noise. In section 5.3.4, it was shown that the EEG increments have complex scaling properties that do not fit the power-law model. Finally, the hypothesis that there is very little determinism present in the EEG is given strong support in section 5.4, suggesting that the best model is a stochastic process.

The model presented below is based on the filtered linear fractional stable noise (LFSN) stochastic process described in section 2.2.2. Recall that in this process, impulsive α -stable shot noise is passed through two linear filters, first a kernel composed of a difference of power laws, and second a moving average

filter to remove discretization errors. The EEG and filtered LFSN share the properties listed above. The properties of EEG epochs and simulated filtered LFSN processes are compared in the next section.

5.5.1 Normal EEG as Fractional Stable Noise

Recall the compound linear fractional stable noise (CLFSN) in definition 2.26. A realization of the symmetric balanced CLFSN process ($a = b = 1$) with $H_1 = 0.2$ and $H_2 = 2/\alpha - H_1 = 0.911$ is generated for time $T = 200$ seconds using the parameters $\alpha = 1.8$, $\sigma = 1$, memory $N = 1000$, time discretization $n = 0.02$. It is re-scaled in time and plotted with a 10-second epoch of normal EEG measured from the cortex in figure 5.22. Both are normalized and centered, and the signal details are plotted as well.

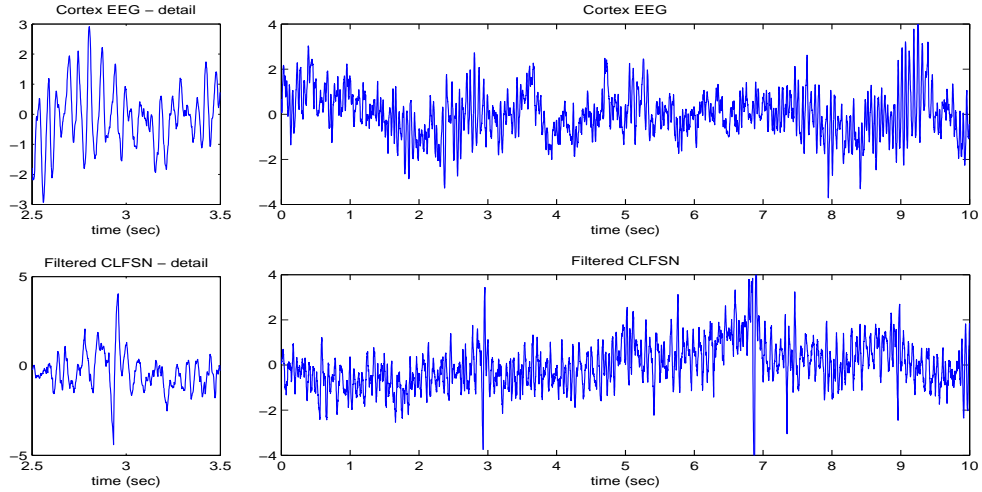


Figure 5.22: Cortex EEG (top) and CLFSN (bottom) Signals and details (left)

The autocorrelation and mutual information of the two signals are very

similar locally, though the CLFSN in this example has greater long-range correlation. They are compared in figure 5.23.

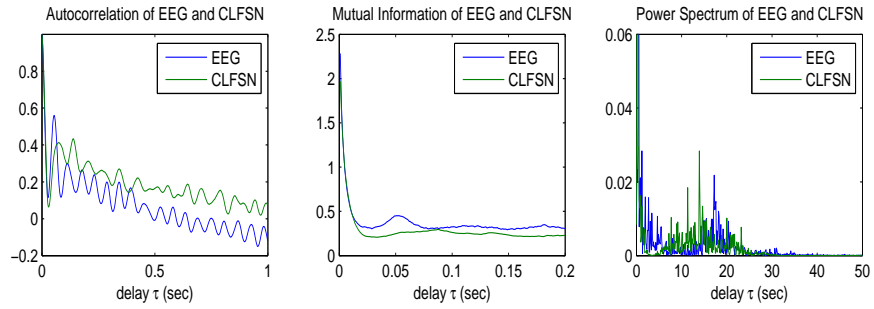


Figure 5.23: Autocorrelation (left), Mutual Information (center), and Power Spectral Density (right) of cortex EEG and CLFSN

The shape parameter α of the distribution of increments on different time scales is compared for the EEG and CLFSN in figure 5.24. The shape parameter of increments the CLFSN and EEG vary on different time scales, and the increments of the EEG are normally distributed ($\alpha = 2$) on some scales.

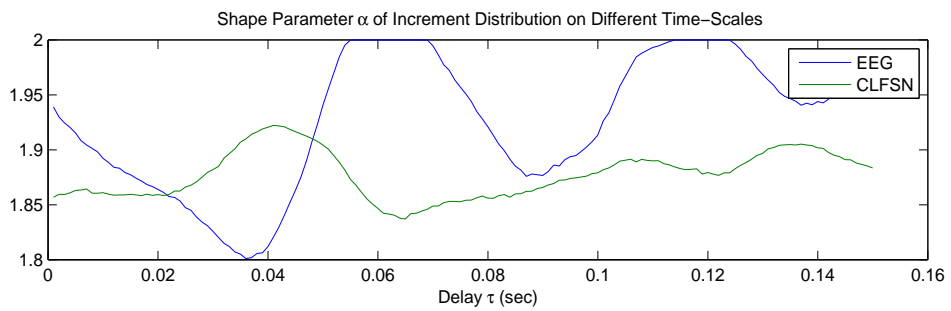


Figure 5.24: Shape Parameter α of Increment Distribution for EEG and CLFSN

The scaling of the two signals on increasing time-scales is compared by examining the scale parameter σ of the distribution of increments. This is

shown in figure 5.25. The fractal dimension is 1.061 ± 0.01 for the EEG sampled at 1000 Hz and 1.131 ± 0.02 for the CLFSN.

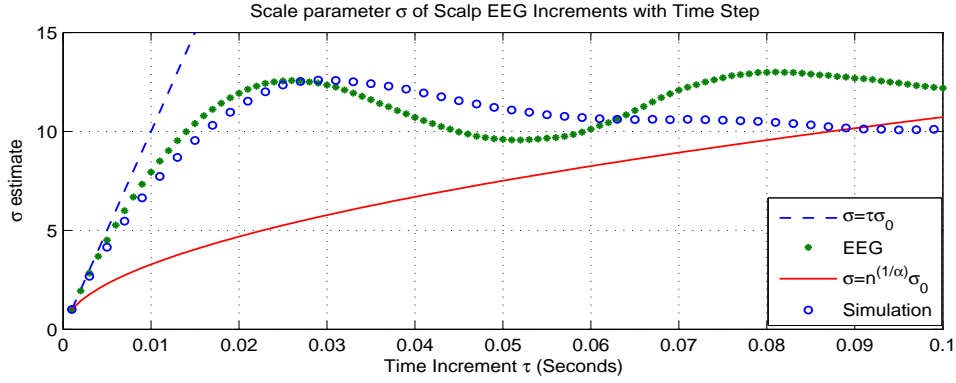


Figure 5.25: Scale Parameter σ of Increment Distribution for EEG and CLFSN

The LFSN shares several important properties with the EEG, specifically that the increments are α -stable. Also, the autocorrelation, mutual information, and scaling properties are very similar on small time-scales, and comparable on larger scales. The LFSN is suggested as a good model for normal EEG activity.

5.5.2 Self-Organization Theory of Neuronal Activity

The idea that the electrical activity of the brain is the manifestation of a self-organized network of single neurons dates back to the work of Norbert Wiener [143]. Wiener pioneered the use of power spectrum estimation in electrophysiology and further suggested that nonlinear phenomena play a large role in the functions of the brain and nervous system.

The advances in nonlinear theory and computer power in the past decades has made the examination of such phenomena in the EEG possible. The

EEG fluctuates in amplitude and frequency, but the nature of these oscillations remains unclear. Another feature is the spontaneous synchronization of oscillations in EEG signals, which has been observed during sleep [45], and in an extreme form is the symptom of epileptiform activity [83]. It is suggested that such spontaneous synchronization of neuronal activity may be correlated [81] over long time-scales. The authors of that report also suggest that the theory of self-organized criticality [7, 8] may be an explanation for this phenomenon.

The theory of self-organized criticality explains the evolution of complex systems through local interactions, in space and time, towards so-called critical state. The apparently spontaneous emergence of these states is in fact correlated over long time-scales. The state is termed ‘critical’ because, although very sensitive to small changes, the system reaches an equilibrium at that state. The qualitative concept of criticality provides a conceptual connection between the notions of nonlinear dynamics, spatial self-similarity, and α -stable noise, which have been observed in complex physical systems. A quantitative basis for this connection is presented in [8].

The theory of self-organized criticality was proposed for the EEG in [81], and some scaling properties were examined to support this hypothesis. The distributional analysis in this chapter supports that hypothesis as well, and the fractional stable noise model reflects the properties of the phenomenon of criticality.

The notion of spontaneous self-organization can be used to understand

applied problems in EEG analysis. The emergence of pathological states can be considered as the result of a spontaneous departure from a normal, critical state. This paradigm can be used to explain the phenomenon of epilepsy, and possibly other pathologies.

5.6 Discussion and Future Work

In this chapter the structure of the EEG signal was examined. The linear statistical properties of autocorrelation and power spectrum were investigated for some example EEG epochs, as well as their temporal variability. The amplitude distribution of the EEG was found to be irregular, while the distribution of increments was found to be α -stable. Goodness-of-fit tests verified this model in every case. The EEG was found to have long memory and self-similarity properties. The presence of nonlinearity was investigated, and the hypothesis that low-dimensional nonlinear determinism is present in the EEG signal was found to be questionable and unlikely. A new type of fractional stable noise process, the filtered compound LFSN, was suggested as a model for the EEG, with which its properties were found to be comparable, notably the distribution of increments, scaling, and non-stationarity. It is the first EEG model to use α -stable noise, and is supported by its connection to the theory of self-organized criticality.

Future work in analysis of the properties of the EEG may be founded qualitatively on the theory of self-organized criticality. Further testing for the presence of nonlinear phenomena is suggested, specifically in long-range

temporal correlations, and the emergence of spontaneous synchronized oscillations. Any model for the EEG must be validated by testing whether its increments satisfy the α -stable model. The fractional stable noise model should be investigated further, and parameter estimation techniques developed to calibrate the model. Applied problems in EEG analysis should be connected with proper qualitative categorization of the phenomena in question and computational methods to quantify their presence or emergence.

Chapter 6

Automatic Sleep-State Identification in Neonates

The problem of sleep-state identification has been examined in detail for adults, and known results are described in [94]. Sleep cycles are well developed in adults, and can be identified effectively by clinicians and to some extent by power spectrum analysis. The sleep dynamics of neonates are substantially more complex, because they are in the process of development and maturation. It follows that certain information about the neural maturity of a neonate can be inferred given the level of maturation of its sleep cycle, as suggested in [119].

Information about sleep cycles depends on the definition and identification of various clinical sleep states. For adults, defined states are rapid eye movement (REM), slow-wave sleep (SWS), and two types of non-REM sleep (NREM1 and NREM2). These states are defined and identified in practice from polysomnographic (PSG) recordings by clinical experts, among whom there can be substantial variability for the same record. A study was con-

ducted [29] to quantify this variability by using 196 recordings and independent sleep-state scorings from two European sleep laboratories. The authors gave the agreement between the scores of two clinicians in identifying sleep states as 78.4% for REM, 71.6% for SWS, 63.3% for NREM2, and 40.3% for NREM1. Because the sleep-cycle of the neonate is in development, it is unlikely that agreement would be any better in this case. The defined states for mature neonates are two active states: mixed frequency (MF) and low-voltage irregular (LVI); and two quiet sleep states: high voltage slow (HVS) and tracé alternant (TA). When scoring PSG records, clinicians also specify awake, transitional, and indeterminate states.

In order to consistently quantify brain maturation or identify pathology using the temporal characteristics of sleep cycles, a consistent scoring technique must be developed. Automatic sleep-state detection from the EEG by computational methods was proposed in [128] for this purpose, and more modern approaches are proposed in [121]. Any automatic methods must be compared to existing clinical practices, described in [32].

In this chapter, a novel algorithm, to appear in a similar form in [108], is presented for sleep state identification in neonates based on the methods in [18, 73]. The techniques of feature extraction, segmentation and clustering described in section 4.3 are applied to neonatal PSG recordings. The features used are chosen based on applicability and robustness, and the method is verified by measuring the agreement of the resulting scores with those provided by Dr. Mark Scher of the Case School of Medicine.

6.1 Initial Feature Selection

The choice of features can be conducted empirically, or by testing statistical significance as in [43] using analysis of variance (ANOVA). Feature extraction involves the processing of many short epochs of EEG data, so simple, rapid, and robust measures are desired. The measures chosen for analysis in the neonate sleep-EEG application are band power (BP), Hjorth parameters of mobility (HM) and complexity (HC), spectral entropy (SE), fractal dimension (FD, by the variogram method), and central tendency (CTM). These measures are analyzed for significance in distinguishing sleep states visually and by ANOVA.

The measures are applied to all one-minute epochs of EEG data from the MF, LVI, HVS, and TA states as identified by a clinician for 19 full-term neonates. Box-plots of these measures are shown for epochs of the center channel, Cz-Pz in the international 10-20 system, in figure 6.1, and box-plots of power in low-frequency bands are shown in figure 6.2.

The box-and-whisker plots (box-plots) in figures 6.1 and 6.2 are used to visualize the distribution of each feature in each state, and are interpreted as follows. The box is the domain of the middle two quartiles of the data, and the line that divides it is the median estimate. The notches on the median line indicate a 90% confidence interval for the median. The whiskers give the domain of the upper and lower quartiles, and the plus marks indicate outliers.

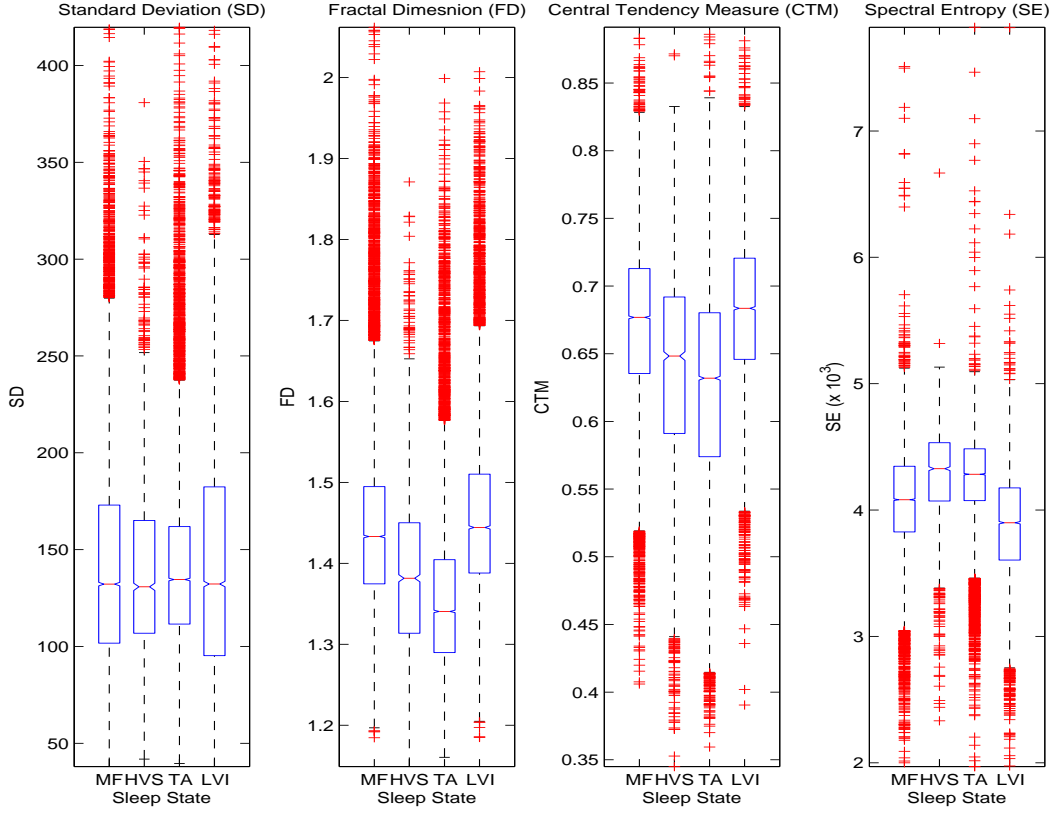


Figure 6.1: Box-Plots of EEG Measures for Neonatal Sleep States, Center Channel (Cz-Pz), Left to Right: Standard Deviation, Fractal Dimension, Central Tendency Measure, Spectral Entropy

It can be seen in figure 6.1 that the confidence intervals of the medians are distinguishable for all of the features, even the standard deviation, so each feature is statistically significant. However, this has limited utility in clinical applications because very short segments, or even single points, of the diagnostic sequence must be classified as belonging to a given state. ANOVA, which tests for differences in distribution, is not effective in identifying useful features. A more robust empirical criterion is to select features for which distributions in distinct states have limited overlap in their distributions.

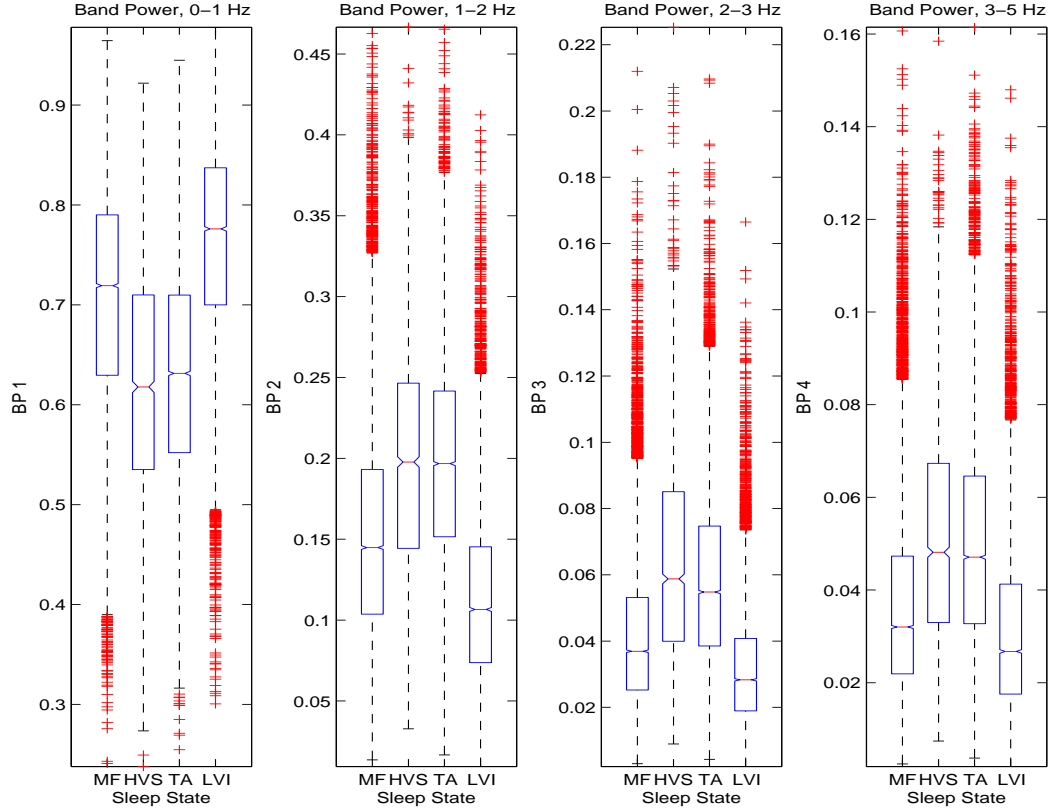


Figure 6.2: BBox-Plots of Band Power for Neonatal Sleep States, Center Channel (Cz-Pz): 0-1 Hz, 1-2 Hz, 2-3 Hz, and 3-5 Hz

Every feature examined has significant overlap in distribution within the quiet sleep states, HVS and TA, and within the active sleep states, LVI and MF. All the features except standard deviation are significantly different between active and quiet sleep, however. Therefore these features can be used to distinguish active and quiet sleep, but not necessarily between the sub-states within them. Table 6.1 lists the features and their difference for active sleep compared to quiet sleep, in addition to p-values and F-statistics for one-way analysis of variance (ANOVA). The p-values for ANOVA should be

	Feature	active \Rightarrow quiet	p-value	F-statistic
SD	standard deviation	N/A	0.0121	3.64
FD	fractal dimension	higher	0	2141.88
CTM	central tendency meas.	higher	0	1734.95
SE	spectral entropy	lower	0	1296.44
BP1	0-1 Hz Power	higher	0	2900.92
BP2	1-2 Hz Power	lower	0	3136.38
BP3	2-3 Hz Power	lower	0	2919.94
BP4	3-5 Hz Power	lower	0	1813.53
HM	Hjorth mobility	N/A	3.5×10^{-11}	17.23
HC	Hjorth complexity	N/A	0.0198	3.29

Table 6.1: Features for Neonatal Sleep-State Detection

zero for a significant feature, and the F-statistic can be used to choose a preliminary set of the most useful features. The features selected are fractal dimension (FD), central tendency measure (CTM), spectral entropy (SE), 0-1 Hz band power (BP1), and 1-2 Hz band power (BP2). The Hjorth parameters and standard deviation are not included, as they provide no usable information. Also, the frequency bands above 2 Hz are not included because they provide redundant information. The subsequent analysis is limited to separating active and quiet sleep.

6.2 Sleep State Scoring Procedure

The automatic sleep-state scoring procedure is implemented following the framework given in algorithm 4.13 for automatic state identification. The features FD, CTM, SE, BP1, and BP2 are used to create diagnostic sequences from each channel of each scored EEG recording with 1-minute epochs, analogous to the minute-by-minute clinical score provided.

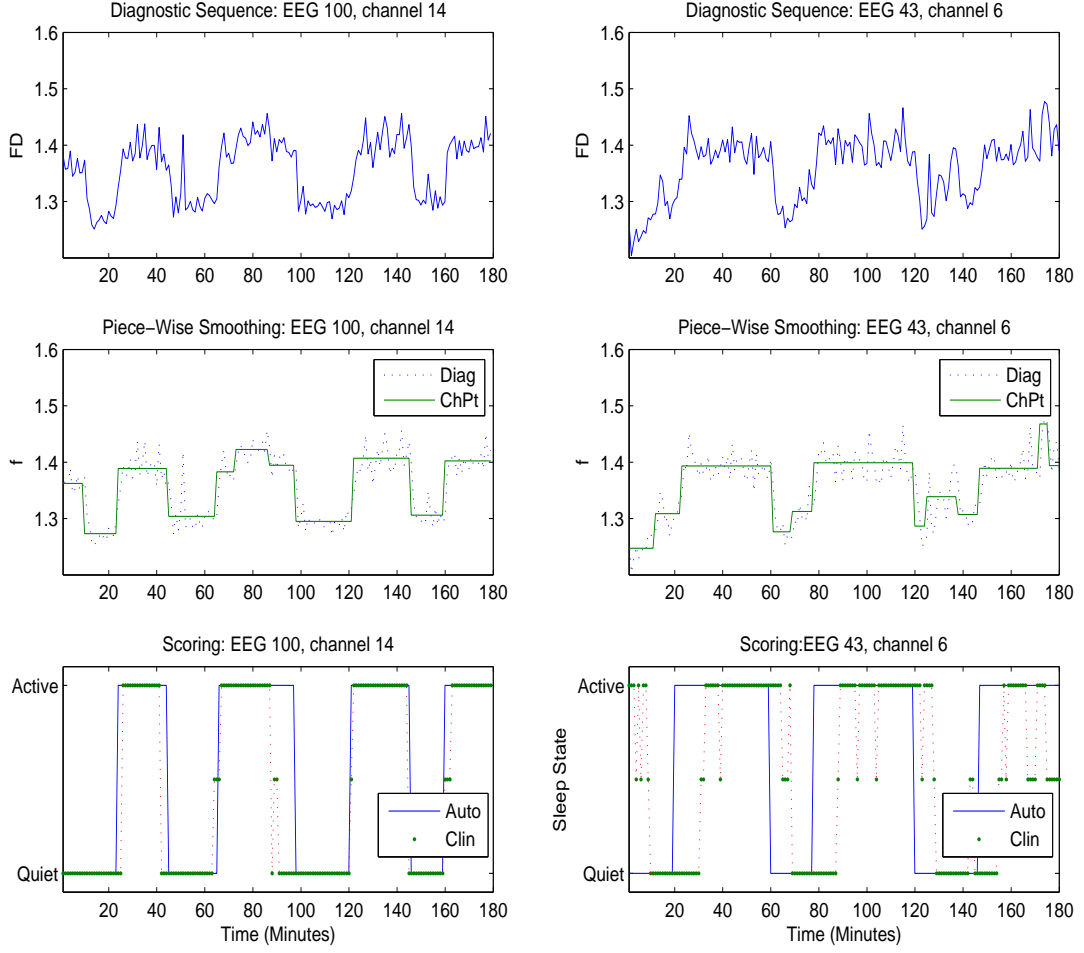


Figure 6.3: Neonatal Sleep-State Identification Procedure: FD Diagnostic Sequence (top), Change-Point Detection (center), and Clustering (bottom)

In the change-point detection step, a high false alarm probability $P_{fa} = 0.3$ is used, because the aim of segmenting the diagnostic sequence is a piece-wise ‘smoothing’ that makes the clustering algorithm more efficient and robust. The procedure is illustrated in figure 6.3 for the FD feature for two neonates with well-developed sleep cycles, and the clinical score is shown for comparison.

6.3 Calibration of Algorithm

The aim of automatic sleep-state detection is to achieve the same result in discriminating active and quiet sleep as a clinician. After automatic scoring using the selected features is accomplished, the resulting score must be compared to the clinical score, and the results used to both calibrate and then also validate the algorithm. Calibration is accomplished by selecting the channel or channels and combination of features for which a measure of agreement is greatest. In addition to % agreement, the authors of [29] use Cohen's Kappa [23], which is a statistic that measures the agreement between scorers or raters. Cohen's Kappa is widely used because it takes into account the probability that agreement is due to chance.

Definition 6.1 (Cohen's Kappa) *Cohen's Kappa is defined as*

$$\kappa = \frac{P(a) - P(e)}{1 - P(e)} \quad (6.1)$$

where $P(a)$ is the relative observed agreement among raters and $P(e)$ is the probability that the agreement is due to chance. Consider two scores $\{A_i\}_{i=1}^N$ and $\{B_i\}_{i=1}^N$ that can take values in $\{S_j\}_{j=1}^K$. Then $P(a)$ is given by

$$P(a) = \frac{1}{N} \sum_{j=1}^K \sum_{i=1}^N (\mathbb{I}\{A_i = B_i\} \cdot \mathbb{I}\{A_i = S_j\}) \quad (6.2)$$

and $P(e)$ is given by

$$P(e) = \frac{1}{N^2} \sum_{j=1}^K \left[\left(\sum_{i=1}^N \mathbb{I}\{A_i = S_j\} \right) \cdot \left(\sum_{i=1}^N \mathbb{I}\{B_i = S_j\} \right) \right] \quad (6.3)$$

where $\mathbb{I}\{\cdot\}$ is the indicator function, given by

$$\mathbb{I}\{expression\} = \begin{cases} 1 & \text{if } expression = true \\ 0 & \text{if } expression = false \end{cases} \quad (6.4)$$

The sleep-state detection algorithm is applied to 19 full-term and 16 pre-term neonates 15 times for each channel. Because the randomly seeded K-means clustering algorithm is used, the algorithm does not always converge to the same final result, so the average performance is of interest. Cohen's Kappa is computed for agreement between the algorithm and epochs identified as active or quiet in the clinical score, for all the runs on each channel.

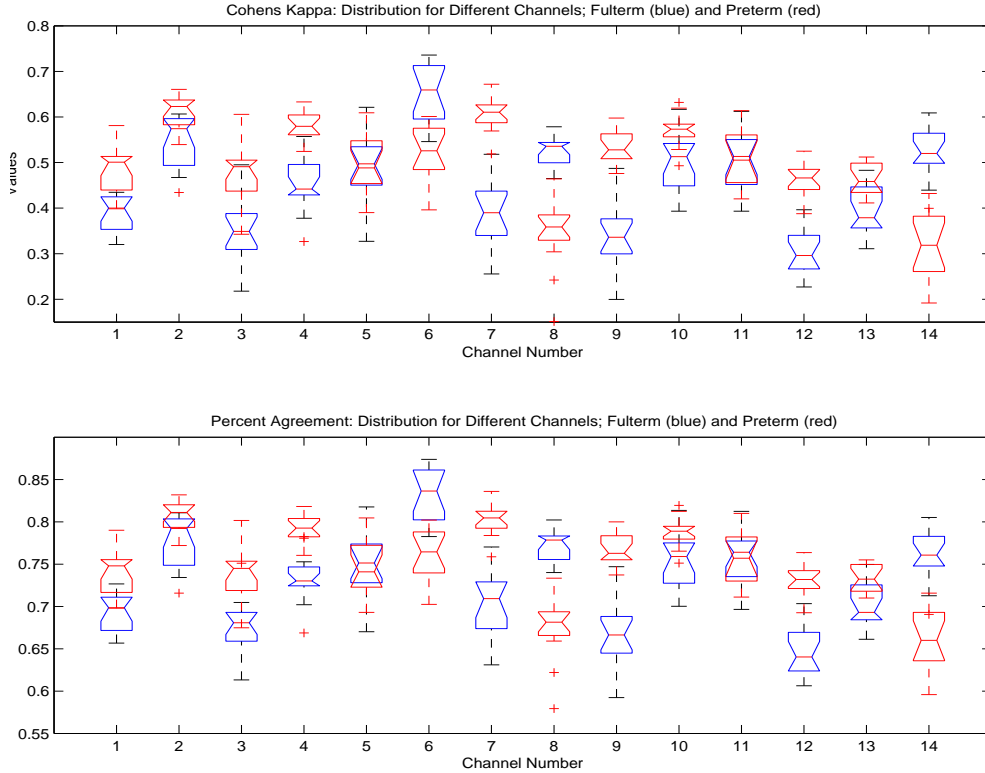


Figure 6.4: Distribution of Cohen's Kappa (top) and % Agreement (bottom) for each Channel; full-term (blue) and pre-term (red)

Box-plots of the distribution of Kappa and % agreement are shown in figure 6.4. The best agreement for the full-term group was obtained for channel 6, which is the potential measured between the locations C3 and O1 in the international 10-20 convention (see section 5.2). This is the potential across the left parietal lobe. The value of Cohen’s Kappa is $\kappa = 0.66$ (84% agreement) for the full-term group, and $\kappa = 0.53$ (77% agreement) for the pre-term group. This is consistent or better than the inter-scorer variability for adults reported in [29].

Next, Cohen’s Kappa and % agreement is computed for the scores produced using several combinations of available features.

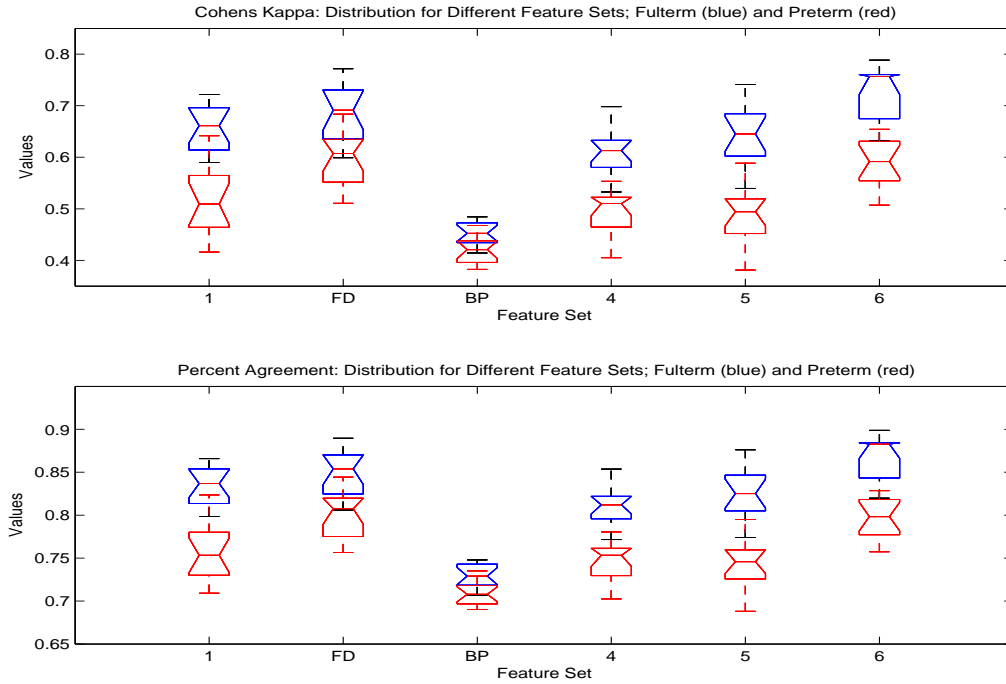


Figure 6.5: Distribution of Cohen’s Kappa (bottom) and % Agreement (top) for Feature Combinations

The Combinations are {FD, CTM, SE, BP1, BP2}, {FD}, {BP1}, {FD, BP1}, {FD, CTM, SE, BP1, BP2, BP3, BP4, BP5}, and {FD, SE, BP1, BP2, BP3}, and the distribution of agreement for each combination is shown in figure 6.5. The values of Cohen’s Kappa using fractal dimension (FD) are $\kappa = 0.70$ (86% agreement) for the full-term group and $\kappa = 0.60$ (81% agreement) for the pre-term group. These values are $\kappa = 0.73$ (87% agreement) for the full-term group and $\kappa = 0.68$ (85% agreement if the worst-case outlier is removed). These values of Kappa and agreement are better than the inter-scorer agreement reported in [29].

6.4 Quantification of Brain Maturation

It is suggested in [119] that the neural development of neonates can be assessed using sleep measures. A simple feature that can be used to quantify the temporal patterns of a sleep sleep cycle is the number of transitions between the active to the quiet sleep states observed per hour of sleep. This feature can be used to compare the sleep cycles of two groups.

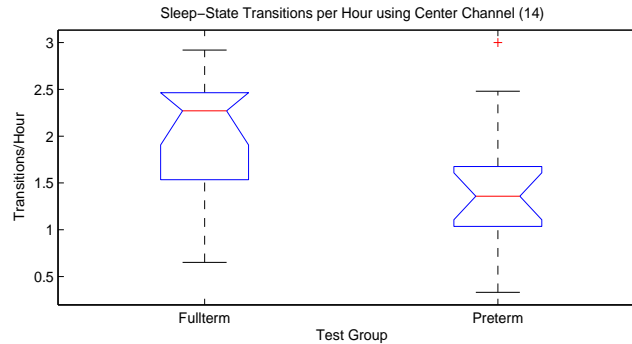


Figure 6.6: Sleep-State Transitions per Hour using Center Channel (Cz-Pz): Full-term (left) and Pre-term (right)

Figure 6.6 shows this comparison for the full-term and pre-term neonate groups. An ANOVA for the sleep-cycle transition rate using a maturity factor with full-term and pre-term values results in a p-value of 0.0135 and F-statistic of 6.89. The full-term group had on average 2.3 state transitions per hour and the preterm group had on average 1.4 state transitions per hour. There is clearly a difference in distributions for the two groups, but it is questionable whether a predictive model can be constructed from these results. Further examination on additional data sets is necessary before a model can be developed. Other features that quantify the differences between sleep cycles may be more informative.

6.5 Discussion and Future Work

In this chapter the state identification methodology introduced in chapter 4 was applied to develop an algorithm for sleep-state identification in neonates. The selection of significant features and implementation of the algorithm is discussed. The calibration of the procedure to best match the clinical scores is done using Cohen's Kappa for measuring agreement between scores. The highest agreement was found using the fractal dimension of potentials across the left parietal lobe. The agreement of the automatic scoring algorithm with the clinical score for this case is on average as good or better than the agreement between independent human clinical scores reported in the literature.

The results of this algorithm can be applied to automatically reconstruct

the sleep-cycles of neonates from their EEG using only one channel of measurement. The regularity of sleep-cycles can be used to assess brain maturity of neonates in a clinical setting. Epochs of the EEG without state transitions can be investigated to search for other significant properties. Other more sensitive features can be examined in order to discriminate between all four clinical neonatal sleep states. The methodology used can also be applied to other problems involving complex systems.

Chapter 7

Automatic Epileptic Seizure Onset Detection

Epilepsy is a serious brain disorder that affects 1% of the population, and it is estimated that over 50 million have the disease worldwide. The manifestations of this disease include seizures, which are characterized by synchronous neuronal firing. Such events can have cognitive, sensory, and motor symptoms that can vary greatly in type and severity, and can also result in serious brain injury. A comprehensive review of the modern medical knowledge and clinical practice in the area of epilepsy is contained in [37].

Over the past decade interest has grown in the use of computational methods to detect and potentially forecast the onset of epileptic seizures using EEG recordings. A review of the state of the art is given in [63], and the growth in the number of publications in this field is commented on in [28]. Many works have appeared recently that propose methods for the detection of epileptic seizures, and some suggest that the onset of seizures can be predicted. A review of these is given in section 7.1.

This chapter examines the phenomenon of epilepsy and the results are used to suggest which measures may best be used to discriminate seizure activity from the normal EEG. Measures that best reflect the phenomenological change in the nature of the signal are used as onset criteria in an algorithm that realizes rapid seizure detection, and the possibility of forecasting seizure onset is discussed as well. It is important to note that this investigation has been conducted using rare, recently obtained clinical EEG data measured directly from the cortex of epileptic patients at a high 1000 Hz sample rate. This study is notable, in that previous research on epilepsy has been based on scalp data.

7.1 Recent Computational Epilepsy Analysis

Recent work in the field of computational EEG analysis that focuses on epilepsy approaches the problems of seizure detection, prediction, and control. A large amount of work is focused on the detection of epileptic activity and identifying EEG signal features that change during seizure onset. All of the works surveyed used inter-cranial EEG data measured from the scalp of patients.

Various ways in which seizures arise are investigated in [83]. Self-organized neural networks are used for detection in [47], and clustering algorithms for multi-channel data are used in [75]. The fractal dimension has been observed to change considerably during epileptic seizures [38]. Wavelet analysis is used in [129] and also in [125] along with machine learning algorithms for detec-

tion. System-theoretic methods are used to identify changes in the power spectrum of the EEG that can be used to detect seizures in [126]. The effectiveness of linear and nonlinear methods for detecting seizures are compared in [90], and the SLEX method for time-frequency estimation is suggested as a possibility in [25]. In addition, the authors of [56] suggest that the correlation dimension is not a reliable measure of epileptiform activity.

Several research groups have developed methods claimed to predict the onset of seizures. In [62], the authors observed that the largest Lyapunov exponents of short epochs of the EEG gradually decreased in critical regions of the brain leading up to a seizure. The forecasting algorithm that they present evaluates the ‘entrainment’ between ‘critical’ electrode sites by comparing largest Lyapunov exponents for signals measured at those sites. That work is based on the observation that the largest Lyapunov exponents of the EEG decrease gradually leading up to seizure onset [64].

A genetic feature selection heuristic is used in [28] for early seizure onset detection. The primary features used include estimators of fractal dimension, spectral energy, entropy, and wavelet packets, as well as secondary features obtained by arithmetic and statistical operations applied to the primary features. A wavelet-based entropy measure is also used as a feature in [101]. A measure of complexity of the EEG is investigated in [68] for use in seizure onset detection. Complexity is found to decrease gradually leading up to a seizure, and then to drop markedly just before clinical onset.

All of the reviewed prediction methods are patient-specific, in that they

require EEG recordings that contain seizures to calibrate the thresholds for the required decision structures.

7.2 Quantification of Epileptiform Activity

The aims of much of the work discussed in section 7.1 can be separated into two distinct but related goals. The first goal, called *feature identification*, is to discover features of the signal, such as complexity or power spectrum, that change significantly at, or leading up to, seizure onset (referred to as the *ictal* state). The second goal, called *decision structure development*, is to construct algorithmic measures that quantify changes in those features in single or multiple channel EEG recordings and make decisions about the onset of a seizure. Works such as [28] combine the above goals by selecting features based on the performance of a decision structure for a given data set.

The investigation here is restricted to feature identification, with the goal of selecting a signal feature for which some measure can be rapidly computed to quantify epileptiform activity. A consensus has been reached throughout the literature [63] that such activity is characterized by synchronization of neuronal firing. The scope of such synchrony in the brain is related directly to the severity of the seizure.

Convincing evidence has been presented for the presence a low-dimensional nonlinear dynamics in the EEG during epileptic seizures [6, 64]. Consequently, the features that have been proposed to quantify ictal activity are

dimension and chaos, measured by the correlation dimension and largest Lyapunov exponents, respectively. The paradigm of a low-dimensional nonlinear dynamical system is too narrow to describe the EEG, as shown in [136], [100] and in chapter 5 of this work.

Recall that it was shown in chapter 5 that the hypothesis that deterministic nonlinear dynamics are present in the normal EEG is questionable, and it was suggested that the normal EEG is best modeled as a stochastic process. It then follows that epileptiform activity can be related to a change in the qualitative nature of the EEG from a stochastic process to a low-dimensional deterministic nonlinear system. At seizure onset, a region of the brain spontaneously re-organizes its functions from complex asynchronous activity that is similar to fractional stable noise to a simpler, synchronized firing that has a structure similar to a low-dimensional nonlinear system. This hypothesis is justified conceptually by the fact that the action potential of a neuron is a nonlinear deterministic process [1], as is the electrical activity of neurons firing in concert.

To validate this hypothesis quantitatively, the test for determinism from section 3.4.1 can be applied to ictal EEG epochs. EEG recordings from 5 patients containing 11 seizures are examined. The surrogate data test for determinism using the CTM is performed on 6 consecutive 10-second (10000 point) EEG epochs beginning at clinical onset for each seizure. A channel with significant ictal activity is chosen in each case. Recall that the null hypothesis of the test is that the signal is stochastic. For 83% of the epochs

the null hypothesis is rejected, meaning that the signal is deterministic in those cases. Box-plots of the empirical S_{CTM} measure and the CTM are shown in figure 7.1. The CTM of non-ictal epochs from figure 5.19 is included for comparison.

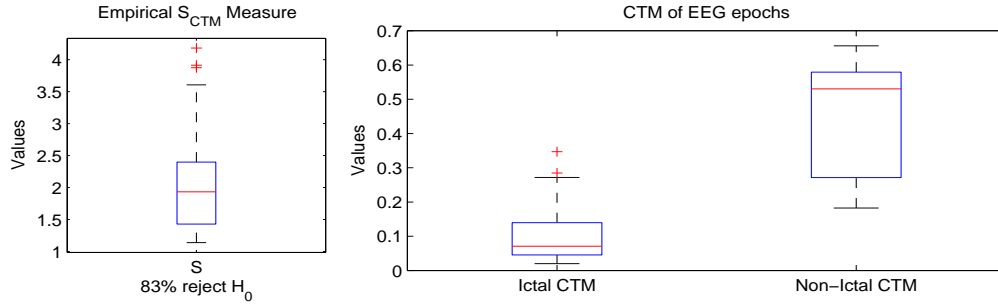


Figure 7.1: Boxplots of Empirical S-measure (left) and CTM (center) for Ictal EEG Epochs and CTM for Non-Ictal Epochs (right)

These results verify the hypothesis that the ictal EEG is most likely a deterministic signal. Furthermore, there is an average difference of an order of magnitude between the CTM of the ictal and non-ictal EEG. It follows that the CTM itself can be used as a feature to identify ictal activity. The fractal dimension (FD), a rapidly computable measure, has also been found to identify ictal activity [38]. That measure quantifies the ‘smoothness’ of a time-series, so its use can also be justified qualitatively. To illustrate the use of CTM and FD to identify seizure activity, diagnostic sequences for both measures are computed for all channels of several EEG epochs containing seizures. 5-second epochs are used, and time-delay embedding is performed using $\tau = 0.03$ and $m = 7$. The results of two channels with marked seizure activity are shown in figure 7.2 for patient 1.

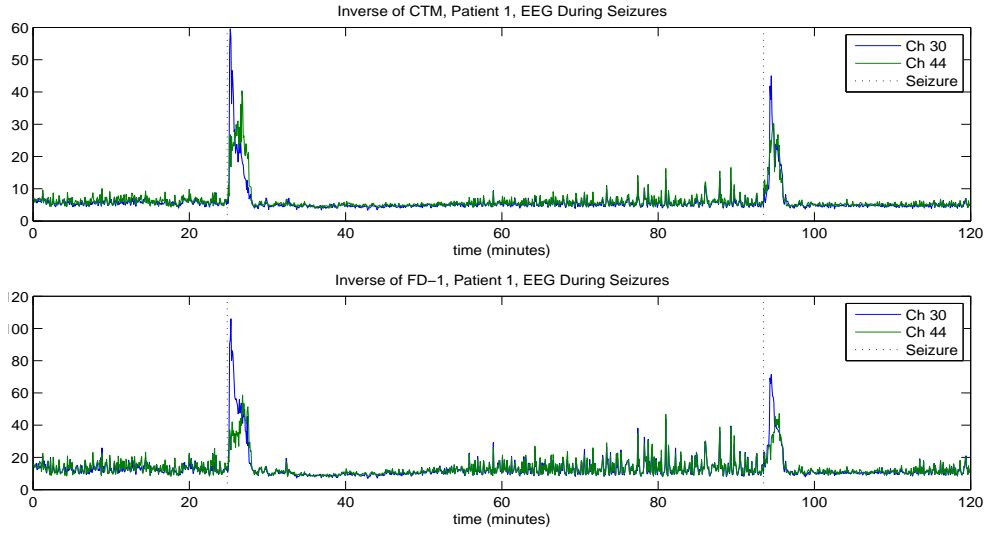


Figure 7.2: CTM (top) and FD (bottom) Measures During Seizures

The transformed measures $1/CTM$ and $1/(FD-1)$, for which changes are most significant, are shown. The changes in both the CTM and FD features are very clearly visible at the same time as clinical onset. Furthermore, changes in the features are visible during the time leading up to the second seizure.

The decrease in fractal dimension and increase in determinism leading up to seizure onset is similar to that of the observations of the largest Lyapunov exponents [62] and signal complexity [68], which is the basis for forecasting algorithms presented in the literature. All of these factors point to a gradual increase in the synchronization of neuronal activity throughout the brain before electrographic seizure onset.

Strong evidence was presented to support the hypothesis that the normal state of the EEG is random. It is possible that anomalous neuronal

synchronization begins to occur spontaneously in pathological regions of the brain, which in turn triggers control mechanisms that attempt to stabilize the resulting oscillations. This would explain the ‘spikes’ in the CTM and FD during the inter-ictal period in figure 7.2. It is also possible that eventually the feedback mechanism over-compensates and begins to contribute to the pathological entrained oscillations, resulting in the spread of ictal activity throughout the brain. The ictal state then lasts until the neurons have exhausted their stored energy and are no longer able to fire synchronously. Thereafter the neurons self-organize to resume normal ‘autonomous’ activity.

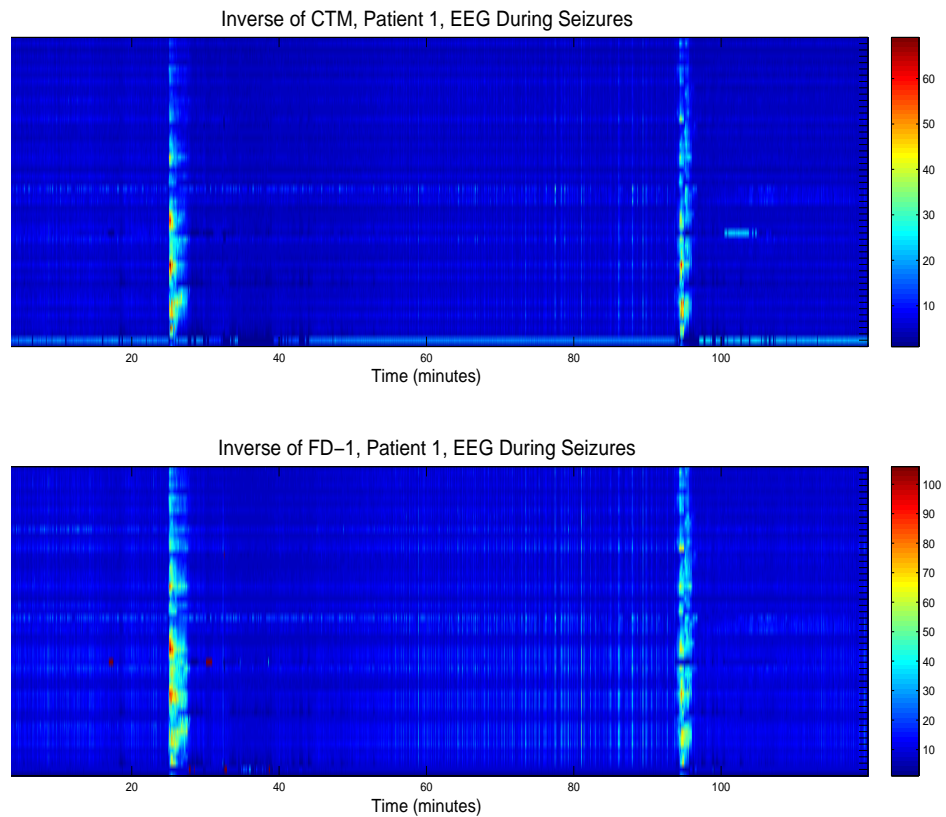


Figure 7.3: CTM (top) and FD (bottom) Measures During Seizures

The localization of ictal activity in the brain can be determined by viewing diagnostic sequences for all channels at once, as shown in 7.3. The red areas in the upper plot of the CTM indicate the channels corresponding to the most pathological regions of the brain. Similar results were observed in other EEG recordings containing seizures. Therefore seizure identification and detection simply require checking whether the CTM or FD measures have increased past an easily-calibrated non-patient-specific threshold for any given channel.

7.3 Clinical Applications and Early Seizure Onset Detection

As noted in section 7.1, several methods for early detection, or prediction, of seizure onset using the scalp EEG have been proposed. In order to implement and test such a technique using the features suggested in section 7.2, a *decision structure* must be developed. One such example is the “T-statistic” described in [62].

The requirements for a decision structure are to identify critical electrode sites, quantify the probability of onset at an given time using a-priori information, forecast the time until onset if it is likely, and positively identify onset if it occurs. All of this should be accomplished while minimizing the use of patient-specific information or any training sets, in order to be adaptable for application in clinical trials.

Furthermore, for an early seizure onset detection algorithm to be clinically applicable, the feature extraction and decision process must be simple enough

to operate in real-time. The fractal dimension and the central tendency measure are well-suited for this purpose because they change consistently in the presence of ictal activity and do not require parametric analysis. Their computation is also very rapid, with CTM and FD requiring about .015 and .005 seconds, respectively, on a 3 GHz machine for 5000 points. A simplified nonparametric decision structure similar to the “T-statistic” could be implemented using these features to achieve the objectives mentioned above. Such a framework would be efficient enough to process data from several channels in an embedded microprocessor.

In the case of severe epileptic patients, grids of electrodes are surgically implanted directly onto the cortex for the purpose of diagnosis. The objective of this procedure is to identify the most pathological area of the brain for surgical resection. This can be greatly aided by the use of diagnostic sequences created by computational feature extraction, such as figure 7.3. For the two seizures that occur during that epoch, the CTM has the greatest magnitude in the most pathological areas, as indicated by an epilepsy specialist who worked with the patient.

7.4 Discussion and Future Work

The qualitative changes that occur in the EEG signal during seizure onset were discussed using the results of previous work and new results from chapter 5. The spontaneous synchronization of neuronal activity was used to explain the onset of epilepsy. Synchronization between individual neurons

cannot be measured, and synchronization between EEG channels cannot be attributed to that phenomenon. However, signal features that reflect synchronization, such as fractal dimension and measures of determinism, can be used to identify epileptiform activity. Such measures, whose relevance can be explained qualitatively and demonstrated quantitatively by hypothesis testing, are better suited to robust algorithm development that minimizes parametrization requirements.

In order to implement the features suggested here, an efficient decision structure that requires little computation time should be developed to work in conjunction with feature estimation algorithms. Using a decision structure with few parameters and a simple design, a real-time seizure onset detection and forecasting device can be designed. Using such a device, researchers could enter the next phase in epilepsy research, discovering electrical signals with which pathological areas of the brain can be stimulated to prevent or counteract ictal activity.

Chapter 8

Conclusions and Recommendations for Future Work

This thesis develops several non-standard mathematical and algorithmic approaches for modeling and analysis of the human EEG. Because the behavior of the EEG cannot be explained by any single narrow mathematical model, especially in pathological cases, multiple time-series analysis methodologies are used, and the literature in each area is reviewed. The notions of stable stochastic processes, nonlinear deterministic dynamical systems, random fractals, and self-organizing networks are used to create a unifying model for EEG behavior using statistical tests for hypothesis validation. This approach is found to be useful in investigating epilepsy. In addition, a state-identification methodology is developed and applied to the neonatal EEG sleep-state identification problem. The methods and applications presented in the thesis are reviewed in detail in sections 8.1 and 8.2, and suggestions for future work are included. The integration of very diverse mathematical

concepts to more completely characterize complex systems such as the EEG is novel, and many directions for future work can be found.

8.1 Methods

8.1.1 Stable and Fractional Processes

A computational toolbox is developed for data analysis and modeling using α -stable distributions. Novel parameter estimation and goodness-of-fit testing algorithms are developed. Self-similar, stable, and fractional stochastic processes are examined. The compound linear fractional stable noise, which combines correlation structures on two separate time scales, is introduced.

Additional work is necessary to compare the parameter estimation algorithm to other recent algorithms, and to develop more efficient estimation methods. Stable and fractional stochastic processes should be investigated further, focusing on simulation methods, discovering new types of processes, and modeling physical phenomena.

8.1.2 Nonlinear Time-Series Analysis

Various methods for time-series analysis are explored. Standard methods for processing of linear stochastic signals are reviewed. Methods for fractal time-series analysis are compared, and nonlinear time-series methods based on phase-space reconstruction are reviewed. The types of signals are rigorously classified, i.e. deterministic or stochastic and linear or nonlinear, using hypothesis testing with surrogate data.

Further investigation of methods for estimating the fractal properties of objects of dimension greater than $n = 2$ is warranted. Also, existing phase-space based methods for nonlinear time-series analysis should be applied to different types of signals to determine what properties they quantify, and how they can be modified to remain invariant with respect to other properties. The evaluation of most phase-space based measures is computationally intensive, therefore new, more efficient methods to use time-delay embedding to extract useful information should be developed.

8.1.3 Segmentation and Clustering

A rapid a-posteriori change point detection algorithm based on the Brownian bridge process is implemented to detect statistically significant transitions in random sequences. The process of feature extraction to create diagnostic sequences from time-series data is reviewed. An algorithm for state identification problems using diagnostic sequences is presented. Pre-processing of diagnostic sequences by change-point detection and smoothing is used to make the decision structure based on K-means clustering more robust.

Change-point detection is a problem that arises in many engineering applications. Novel ways to apply the algorithm in signal processing applications may be found, and it can be adapted to problems in more than one dimension. In addition, the “smoothing” effect of the algorithm can be combined with intelligent clustering methods to develop more robust state-identification algorithms.

8.2 Applications

8.2.1 EEG Analysis and Modeling

An investigation was conducted into the general structure of the normal EEG signal. Most notably, the increments of the EEG were found to be a good fit to an α -stable distribution. Furthermore, the notion of low-dimensional deterministic nonlinear dynamics in the normal EEG was found to be questionable. An alternate model, based on a linear stochastic process with long-range correlated α -stable increments, was proposed, and was shown to share many important properties with the EEG signal. In addition, this model is in agreement with the theory of self-organized criticality, which was recently proposed to explain EEG activity. Because the α -stable distribution is the result of a limiting sum of random variables, it is an intuitive model for EEG activity, which is the result of the combined electrical potentials of billions of nerve cells.

It is suggested that any model for the EEG should result in an α -stable distribution of the increments. Models that combine nonlinear and stochastic components should be investigated, because explaining EEG activity appears to require such a combination.

8.2.2 Sleep-State Detection

The state identification methodology described in chapter 4 was used to develop an automatic algorithm to identify sleep-states in neonates from the EEG. This algorithm can be viewed as an automatic sleep-scoring technique

that extracts the sleep-cycle from the EEG, identifying transitions between active and quiet sleep. The number of transitions per hour are found to be related to the maturity of the neonate. This is in accord with clinical findings relating the regularity of the sleep-cycle to brain maturation in neonates.

These results are based on an investigation of 35 EEG recordings out of the 116 in the NIH study for which clinical scores were available. The analysis should be applied to the other EEG recordings to test whether differences in the number of state transitions per hour between full-term and pre-term populations are consistent with the findings. The number of state transitions per hour, and other simple measures of maturity derived from the automatic sleep-cycle score, can be used to develop a predictive statistical model that can be tested for clinical relevance. Additional data from EEG studies using higher sample-rates can be used to verify the results and increase the predictive specificity of the model.

8.2.3 Epileptic Seizure Detection

The phenomenon of seizure onset was investigated using EEG recordings measured directly from the cortex of severe epileptic patients. The EEG signal was found to change qualitatively from a predominantly linear stochastic signal to a nonlinear deterministic signal during seizure onset. The central tendency measure, used to quantify determinism, was applied to 5-second epochs of the EEG recordings and shown to clearly identify the onset of epileptic seizures. Furthermore, the CTM was greatest in magnitude in the

most pathological areas of the brain. In addition, the fractal dimension of the signal was shown to have possible predictive properties. The theory of self-organized criticality was suggested to explain seizures as the overcompensation of feedback mechanisms applied by the brain to suppress epileptiform activity.

This work focuses on discovering signal features that best identify epileptiform activity. The fractal dimension and central tendency measure accomplish this, and are extremely simple and can be rapidly computed, compared with published methods. The next step is to design a robust and rapid decision structure based on non-parametric statistics to quantify the likelihood of onset of a seizure. Together with feature extraction methods, such a decision structure can be used to forecast seizures and intervene to prevent or mitigate seizures and their symptoms.

Appendix A

List of MATLAB Codes

A.1 α -Stable Distributions

ASadtest.m Anderson-Darling test for goodness-of-fit of α -stable parameters to a data vector

AScdf.m Cumulative distribution function of the α -stable distribution

ASestinit.m α -stable parameter estimation by Zolotarev's method of log-moments with corrections for feasibility

ASkstest.m Kolmogorov-Smirnov test for goodness-of-fit of α -stable parameters to a data vector

ASmle.m Numerical maximum likelihood estimation of α -stable parameters

ASpdf.m Probability density function of the α -stable distribution

astail.m Tail probabilities of the α -stable distribution

bergstrom.m Asymptotic Bergström expansion for evaluating the probability density function for extreme domain values

CharFun.m Characteristic function of the α -stable distribution

ksdist.m Complementary Kolmogorov-Smirnov distribution

KSinverse.m Inverse of the Kolmogorov-Smirnov distribution

LLfun.m Log-likelihood objective function for Numerical MLE

randas.m Simulation of α -stable random samples by the Chambers-Mallows-Stuck method

A.2 Stable and Fractional Processes

clfsn.m Simulation of the compound linear fractional stable noise process

clfsnkf.m Kernel function for the compound linear fractional stable noise process

lfsn.m Simulation of the linear fractional stable noise process

lfsnkf.m Kernel function for the linear fractional stable noise process

logfsnkf.m Kernel function for the log-fractional stable noise process

A.3 Time-Series Analysis

A.3.1 Spectral Analysis

pgram.m Spectrum estimation by the periodogram method

bandpower.m Power in a given frequency band from the periodogram

SPedge.m Spectral edge from the periodogram

Hjorth.m Hjorth's parameters in the time domain

SPent.m Spectral entropy from the periodogram

A.3.2 Fractal Analysis

FrDim.m Fractal dimension by the variogram method

FrDimH.m Fractal dimension by the curve-length method

A.3.3 Attractor Reconstruction

MutInf.m Auto-mutual information function

NfalseNN.m The number of false nearest neighbors

Statevec.m Reconstruction of the phase-space from a time-series

A.3.4 Phase-Space Measures

CorrInt.m Correlation integral

SlopeDet.m Detects the appropriate linear region of a log-log plot for consistent estimation of the correlation dimension

KolmEnt.m Kolmogorov-Sinai entropy

CTM.m Central tendency measure

Lyap.m Largest Lyapunov exponent

A.4 Change-Point Detection

Palarm.m Detects change-points of a random sequence and returns an estimate of the piece-wise constant function and the transition (change-point) locations

Bibliography

- [1] ABBOT, L., AND KEPLER, T. *Statistical Mechanics of Neural Networks*. Springer, Berlin, 1990, ch. Model Neurons: From Hodgkin-Huxley to Hopfield, pp. 5–18.
- [2] ACCARDO, A., AFFINITO, M., CARROZZI, M., AND BOUQUET, F. Use of the fractal dimension for the analysis of electroencephalographic time-series. *Biological Cybernetics* 77 (1997), 339–350.
- [3] ALBANO, A., MUENCH, J., SCHWARTZ, C., MEES, A., AND RAPP, P. Singular-value decomposition and the grassberger-procaccia algorithm. *Physical Review A* 38, 6 (1988), 3017–3026.
- [4] ALI, M., AND SAHA, L. Local lyapunov exponents and characteristics of fixed/periodic points embedded within a chaotic attractor. *Journal of Zhejiang University SCIENCE* 6A, 4 (2005), 296–304.
- [5] ANDERSON, T., AND DARLING, D. A test of goodness of fit. *Journal of the American Statistical Association* 49, 268 (1954), 765–769.
- [6] BABLOYANTZ, A., AND DESTEXHE, A. Low-dimensional chaos in an instance of epilepsy. *Proceedings of the National Academy of Sciences of the United States of America* 82, 10 (1986), 3513–3517.
- [7] BAK, P., TANG, C., AND WIESENFELD, K. Self-organized criticality: An explanation of $1/f$ noise. *Physical Review Letters* 59, 4 (1987), 381–384.
- [8] BAK, P., TANG, C., AND WIESENFELD, K. Self-organized criticality. *Physical Review A* 38, 1 (1988), 364–374.
- [9] BARNDORFF-NIELSEN, O. E., MIKOSCH, T., AND RESNICK, S. I., Eds. *Lévy Processes: Theory and Applications*. Birkhäuser, Boston, 2001.
- [10] BARNSLEY, M. *Fractals Everywhere*. Academic Press, 1988.
- [11] BERGER, H. Über das elektroencephalogramm des menschen. *Archiv für Psychiatrie und Nervenkrankheiten* 87 (1929), 527.
- [12] BERGSTRÖM, H. On some expansions of stable distributions. *Arkiv för Matematik* (1952), 375–378.
- [13] BILLINGSLEY, P. *Convergence of Probability Measures*. Wiley and Sons, Inc., New York, 1968.

- [14] BISCAY, R., LAVIELLE, M., GONZÁLEZ, A., CLARK, I., AND VALDÉS, P. Maximum a posteriori estimation of change points in the eeg. *International Journal of Bio-Medical Computing* 38 (1995), 189–196.
- [15] BODENSCHATZ, J., AND NIKIAS, C. Maximum-likelihood symmetric α -stable parameter estimation. *IEEE Transactions on Signal Processing* 47, 5 (1999), 1382–1384.
- [16] BORAK, S., HÄRDLE, W., AND WERON, R. Stable distributions. Crc 649 discussion papers, Collaborative Research Center 649, Humboldt University, Berlin, Mar. 2005.
- [17] BRODSKY, B., AND DARKHOVSKY, B. *Nonparametric Methods in Change-Point Problems*. Kluwer Academic Publishers, Dordrecht, 1993.
- [18] BRODSKY, B., DARKHOVSKY, B., KAPLAN, A., AND SHISHKIN, S. A nonparametric method for the segmentation of the eeg. *Computer Methods and Programs in Biomedicine* 60 (1999), 93–106.
- [19] BROWN, R., BRYANT, P., AND ABARBANEL, H. Computing the lyapunov spectrum of a dynamical system from an observed time-series. *Physical Review A* 43, 6 (1991), 2787–2806.
- [20] CANNON, M., PERCIVAL, D., CACCIA, D., RAYMOND, G., AND BASSINGTHWAIGHTE, J. Evaluating scaled windowed variance methods for estimating the hurst coefficient of time series. *Physica A* 241 (1997), 606–626.
- [21] CHAMBERS, J., MALLOWS, C., AND STUCK, B. A method for simulating stable random variables. *Journal of the American Statistical Association* 71, 354 (1976), 340–344.
- [22] CHON, K., KANTERS, J., IYENGAR, N., COHEN, R., AND HOLSTEIN-RATHOU, N. Detection of chaotic determinism in stochastic short time series. In *Proceedings -1 19th International Conference* (Nov 1997), IEEE/EMBS, pp. 275–277.
- [23] COHEN, J. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement* 20 (1960), 37.
- [24] CONSTANTINE, A., AND HALL, P. Characterizing surface smoothness via estimation of effective fractal dimension. *Journal of the Royal Statistical Society B* 56, 1 (1994), 97–113.

- [25] CRANSTOUN, S., E. A. Time-frequency spectral estimation of multichannel eeg using the auto-slex method. *IEEE Transactions on Biomedical Engineering* 49, 9 (2002).
- [26] CSÖRGÖ, S., AND HORVÁTH, L. Nonparametric methods for change-point problems. In *Handbook of Statistics* 7. 1998, pp. 403–425.
- [27] D’AGOSTINO, R., AND STEPHENS, M., Eds. *Goodness-of-Fit Techniques*, vol. 68 of *Statistics: textbooks and monographs*. Marcel Dekker, 1986.
- [28] D’ALLESSANDRO, M., E. A. Epileptic seizure prediction using hybrid feature selection over multiple intracranial eeg electrode contacts: A report of four patients. *IEEE Transactions on Biomedical Engineering* 50, 5 (2003), 603–615.
- [29] DANKER-HOPFE, H., E. A. Interrater reliability between scorers from eight european sleep laboratories in subjects with different sleep disorders. *Journal of Sleep Research* 13 (2004), 63–69.
- [30] DENKER, M., AND WOYCZYNSKI, W. *Introductory Statistics and Random Phenomena*. Birkhauser Verlag, 1998.
- [31] DEPETRILLO, P., SPEERS, D., AND RUTTIMANN, U. Determining the hurst exponent of fractal time series and its application to electrocardiographic analysis. *Computers in Biology and Medicine* 29 (1999), 393–406.
- [32] DIETRICH, K., E. A. Principles and practices of neurodevelopmental assessment in children. *Environmental Health Perspectives* 113, 10 (2005), 1437–1446.
- [33] DOGANOGLU, T., AND MITTNIK, S. An approximation procedure for asymmetric stable paretian densities. *Computational Statistics* 13 (1998), 463–475.
- [34] DUBUC, B., QUINIOU, J., ROQUES-CARMES, C., TRICOT, C., AND ZUCKER, S. Evaluating the fractal dimension of profiles. *Physical Review A* 39, 3 (1989), 1500–1512.
- [35] DUHAMEL, P., AND VETTERLI, M. Fast fourier transforms: A tutorial review and a state of the art. *Signal Processing* 19, 259–299.
- [36] ELUL, R. Gaussian behavior of the electroencephalogram: Changes during performance of mental task. *Science, New Series* 164, 3877 (1969), 328–331.

- [37] ENGEL, J., AND PEDLEY, T., Eds. *Epilepsy: A Comprehensive Textbook*. ch. 77.
- [38] ESTELLER, R., E. A. Fractal dimension characterizes seizure onset in epileptic patients. *IEEE International Conference on Acoustics Speech and Signal Processing 4* (1999), 2343–2346.
- [39] FAMA, E. F., AND ROLL, R. Some properties of symmetric stable distributions. *J. Amer. Statist. Assoc.* *63* (Sept. 1968), 817–836.
- [40] FAN, Z. *Estimation Problems for distributions with heavy tails*. PhD thesis, Georg-August-Universität zu Göttingen, 2001.
- [41] FAUL, S., BOYLAN, G., CONNOLLY, S., MARNANE, L., AND LIGHTBODY, G. An evaluation of automated neonatal seizure detection methods. *Clinical Neurophysiology* *116* (2005), 1533–1541.
- [42] FELL, J., MANN, K., RÖSCHKE, J., AND GOPINATHAN, M. Non-linear analysis of continuous ECG during sleep i: Reconstruction. *Biological Cybernetics* *82* (2000), 477–483.
- [43] FELL, J., RÖSCHKE, J., MANN, K., AND SCHÄFFNER, C. Discrimination of sleep stages: a comparison between spectral and nonlinear eeg measures. *Electroencephalography and clinical Neurophysiology* *98* (1996), 301–410.
- [44] FELLER, W. *An Introduction to Probability Theory and Its Applications*, vol. 2. Wiley, New York, 1971.
- [45] FERRI, R., RUNDO, F., OLIVIERO, B., TERZANO, M., AND STAM, C. Dynamics of the eeg slow-wave synchronization during sleep. *Clinical Neurophysiology* *116* (2005), 2783–2795.
- [46] FRASER, A., AND SWINNEY, H. Independent coordinates for strange attractors from mutual information. *Physical Review A* *33*, 2 (1986), 1134–1140.
- [47] GABOR, A. Seizure detection using a self-organized neural network: validation and comparison with other detection strategies. *Electroencephalography and Clinical Neurophysiology* *107* (1998), 27–32.
- [48] GALKA, A., YAMASHITA, O., OZAKI, T., BISCAY, R., AND VALDÉS-SOSA, P. A solution to the dynamical inverse problem of eeg generation using spatiotemporal kalman filtering. *NeuroImage* *23* (2004), 435–453.

- [49] GALLARDO, J. R., DIMITRIOS, M., AND OROZCO-BARBOSA, L. Use of α -stable self-similar stochastic processes for modeling traffic in broadband networks. *Performance Evaluation* 40 (2000), 71–98.
- [50] GIONA, M., LENTINI, F., AND CIMAGALLI, V. Functional reconstruction and local prediction of chaotic time series. *Physical Review A* 44, 6 (1991), 3496–3502.
- [51] GNEDENKO, B., AND KOLMOGOROV, A. *Limit Distributions for Sums of Independent Random Variables*. Addison Wesley, Reading, MA, 1954.
- [52] GNEITLING, T., AND SCHLATHER, M. Stochastic models that separate fractal dimension and the hurst effect. *Society for Industrial and Applied Mathematics* 46, 2 (2004), 269–282.
- [53] GRASSBERGER, P., AND PROCACCIA, I. Characterization of strange attractors. *Physical Review Letters* 50, 5 (1983), 346–349.
- [54] GRÖZINGER, M., E. A. Online detection of rem sleep based on the comprehensive evaluation of short adjacent eeg segments by artificial neural networks. *Prog. Neuro-Psychopharmacol. and Biol. Psychiat.* 21 (1997), 951–963.
- [55] GRÖZINGER, M., FELL, J., AND RÖSCHKE, J. Neural net classification of rem sleep based on spectral measures as compared to nonlinear measures. *Biological Cybernetics* 85 (2001), 335–341.
- [56] HARRISON, M., OSORIO, I., FREI, M., ASURI, S., AND LAI, Y. Correlation dimension and integral do not predict epileptic seizures. *Chaos* 15, 3 (2005).
- [57] HARTIGAN, J., AND WONG, M. Algorithm as 136: A k -means clustering algorithm. *Applied Statistics* 28, 1 (1979), 100–108.
- [58] HIGUCHI, T. Approach to an irregular time series on the basis of the fractal theory. *Physica D* 31 (1988), 277–283.
- [59] HJORTH, B. Eeg analysis based on time domain properties. *Electroencephalography and Clinical Neurophysiology* 29 (1970), 306–310.
- [60] HSU, M., AND HSU, D. A piece-wise harmonic langevin model of eeg dynamics: Theory and application to eeg seizure detection. *Neurocomputing* 65–66 (2005), 469–474.

- [61] HURST, H. Long-term storage capacity of reservoirs. In *Transactions of the American Society of Civil Engineers* (1951), vol. 116, pp. 770–808.
- [62] IASEMIDIS, L., E. A. Adaptive epileptic seizure prediction system. *IEEE Transactions on Biomedical Engineering* 50, 5 (2003), 616–627.
- [63] IASEMIDIS, L. Epileptic seizure prediction and control. *IEEE Transactions on Biomedical Engineering* 50, 5 (2003), 549–558.
- [64] IASEMIDIS, L., PRINCIPE, J., AND SACKELLARES, J. Measurement and quantification of spatio-temporal dynamics of human epileptic seizures. In *Nonlinear Signal Processing II: Dynamic Analysis and Modeling*, M. Akay, Ed. IEEE Press, 2001, pp. 294–318.
- [65] JANICKI, A. *Numerical and Statistical Approximation of Stochastic Differential Equations with Non-Gaussian Measures*. H. Steinhaus Center for Stochastic Methods, Wrocław, 1996.
- [66] JANICKI, A. W., POPOVA, I., RITCHKEN, P. H., AND WOYCZYNSKI, W. A. Option pricing bounds in a α -stable security market. *Communications In Statistics - Stochastic Models* 13, 4 (1997), 817–840.
- [67] JEONG, J., GORE, J., AND PETERSON, B. Detecting determinism in short time-series, with an application to the analysis of a stationary eeg recording. *Biological Cybernetics* 86 (2002), 335–342.
- [68] JIA, W., E. A. An epileptic seizure prediction algorithm based on second-order complexity measure. *Physiological Measurement* 26 (2005), 609–625.
- [69] JIN, S., JEONG, J., JEONG, D., KIM, D., AND KIM, S. Nonlinear dynamics of the eeg separated by independent component analysis after sound and light stimulation. *Biological Cybernetics* 86 (2002), 395–401.
- [70] KANTELHARDT, J., KOSCIELNY-BUNDE, E., REGO, H., HAVLIN, S., AND BUNDE, A. Detecting long-range correlations with detrended fluctuation analysis. *Physica A* 295 (2001), 441–454.
- [71] KANTELHARDT, J., ZSCHIEGNER, S., KOSCIELNY-BUNDE, E., HAVLIN, S., BUNDE, A., AND STANLEY, H. Multifractal detrended fluctuation analysis of nonstationary time-series. *Physica A* 316, 87 (2002).

- [72] KANTZ, H., AND SCHREIBER, T. *Nonlinear Time Series Analysis*, 2 ed. Cambridge University Press, New York, NY, 2003.
- [73] KAPLAN, A., J., R., DARKHOVSKY, B., AND FELL, J. Macrostructural eeg characterization based on nonparametric change point segmentation: application to sleep analysis. *Journal of Neuroscience Methods* 106 (2001), 81–90.
- [74] KENNEL, M., BROWN, R., AND ABARBANEL, H. Determining embedding dimension for phase-space reconstruction using a geometrical construction. *Physical Review A* 45, 6, 3403.
- [75] KLATCHKO, A., RAVIV, G., WEBBER, W., AND LESSER, R. Enhancing the detection of seizures with a clustering algorithm. *Electroencephalography and clinical Neurophysiology* 106 (1998), 52–63.
- [76] KOPRINSKA, I., PFURTSCHELLER, G., AND FLOTZINGER, D. Sleep classification in infants by decision tree-based neural networks. *Artificial Intelligence in Medicine* 8 (1996), 387–401.
- [77] KUGIUMTZIS, D. Lack of robustness of the surrogate data test. Tech. rep., Max-Planck Institute for Physics of Complex Systems, 1999.
- [78] KUGIUMTZIS, D., LILLEKJENDLIE, B., AND CHRISTOPHERSEN, N. Chaotic time series, part 1: Estimation of invariant properties in state space. *Modeling, Identification and Control* 15, 4 (1994), 205–224.
- [79] KUGIUMTZIS, D., LILLEKJENDLIE, B., AND CHRISTOPHERSEN, N. Chaotic time series, part 2: System identification and prediction. *Modeling, Identification and Control* 15, 4 (1994), 225–243.
- [80] LE VAN QUYEN, M., MARTINERIE, J., BAULAC, M., AND VARELA, F. Anticipating epileptic seizures in real time by a non-linear analysis of similarity between eeg recordings. *Computational Neuroscience: Neuroreport* 10 (1999), 2149–2155.
- [81] LINKENKAER-HANSEN, K., NIKOULINE, V., PALVA, J., AND ILMONIEMI, R. Long-range temporal correlations and scaling behavior in human brain oscillations. *The Journal of Neuroscience* 21, 4 (2001), 1370–1377.
- [82] LIU, Y., E. A. Statistical properties of the volatility of price fluctuations. *Physical Review E* 60, 2 (1999), 1390–1400.
- [83] LOPES DA SILVA, F., E. A. Dynamical diseases of brain systems: Different routes to epileptic seizures. *IEEE Transactions on Biomedical Engineering* 50, 5 (2003), 540–548.

- [84] MANDELBROT, B., AND VAN NESS, J. W. Fractional brownian motions, fractional noises and applications. *SIAM Review* 10 (1968), 422–437.
- [85] MANDELBROT, B., AND WALLIS, J. Computer experiments with fractional gaussian noises part 3: Mathematical appendix. *Water Resources Research* 5 (1969), 260–267.
- [86] MANTEGNA, R., AND STANLEY, E. Scaling behaviour in the dynamics of an economic index. *Nature* 376 (1995), 46–49.
- [87] MARSAGLIA, G., AND MARSAGLIA, J. Evaluating the anderson-darling distribution. *Journal of Statistical Software* 9 (2004).
- [88] MATHWORKS. Matlab 7 file exchange. The Mathworks, Inc., 2005. <http://www.mathworks.com/matlabcentral/>.
- [89] MAYMON, S., FRIEDMAN, J., AND MESSER, H. A new method for estimating parameters of a skewed alpha-stable distribution. *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing* 6 (2000).
- [90] MCSHARRY, P., SMITH, L., AND TARASSENKO, L. Comparison of predictability of epileptic seizures by a linear and a nonlinear method. *IEEE Transactions on Biomedical Engineering* 50, 5 (2003), 628–633.
- [91] MENN, C., AND RACHEV, S. Calibrated fft-based density approximations for α -stable distributions. *Computational Statistics & Data Analysis* 52 (2006). forthcoming.
- [92] MUNCK, J., VIJN, P., AND LOPES DA SILVA, F. A random dipole model for spontaneous brain activity. *IEEE Transactions on Biomedical Engineering* 39, 8 (1992), 791–804.
- [93] NEIMAN, A., SCHIMANSKY-GEIER, L., MOSS, F., SHULGIN, B., AND COLLINS, J. Synchronization of noisy systems by stochastic signals. *Physical Review E* 60, 1 (1999), 284–292.
- [94] NIEDERMEYER, E., AND LOPES DA SILVA, F. *Electroencephalography: Basic Principles, Clinical Applications, and Related Fields*, 4 ed. Waverly, Baltimore, MD, 1999.
- [95] NIKIAS, C., AND SHAO, M. *Signal Processing with Alpha-Stable Distributions and Applications*. Wiley, New York, 1995.

- [96] NOLAN, J. P. Numerical calculation of stable densities and distribution functions. *Communications in Statistics - Stochastic Models* 13, 759–774.
- [97] NOLAN, J. P. Parameter estimation and data analysis for stable distributions. In *Proceedings of the IEEE 31st Asilomar Conf. on Signals, Systems, and Computers, Monterey, USA* (1998).
- [98] NOLAN, J. P. Maximum likelihood estimation and diagnostics for stable distributions. Tech. rep., Dept. Math. Statist., Amer. Univ., Washington, DC, June 1999.
- [99] NUNEZ, P. The brain wave equation: A model for the eeg. *Mathematical Biosciences* 21 (1974), 279.
- [100] PALŮS, M. Nonlinearity in normal human eeg: cycles, temporal asymmetry, nonstationarity and randomness, not chaos. *Biological Cybernetics* 75 (1996), 389–396.
- [101] PAUL, J., E. A. Prediction of ptz-induced seizures using wavelet-based residual entropy of cortical and subcortical field potentials. *IEEE Transactions on Biomedical Engineering* 50, 5 (2003), 640–648.
- [102] PAXSON, V. Fast, approximate synthesis of fractional gaussian noise for generating self-similar network traffic. *Computer Communication Review* 27, 5 (1997), 5–18.
- [103] PELIGRAD, M. Invariance principles for mixing sequences of random variables. *Annals of Probability* 10, 4 (1982), 968–981.
- [104] PENG, C., HAVLIN, S., STANLEY, H., AND GOLDBERGER, A. Quantification of scaling exponents and crossover phenomena in nonstationary heartbeat time-series. *Chaos* 5, 1 (1995), 82–87.
- [105] PENG, C., E. A. Long-range anticorrelations and non-gaussian behavior of the heartbeat. *Physical Review Letters* 70, 9 (1993), 1343–1346.
- [106] PEREDA, E., GAMUNDI, A., RIAL, R., AND GONZÁLEZ. Non-linear behaviour of human eeg: fractal exponent versus correlation dimension in awake and sleep stages. *Neuroscience Letters* 250 (1998), 91–94.
- [107] PIRYATINSKA, A. *Inference for the Lèvy models and their applications in medicine and statistical physics*. PhD thesis, Case Western Reserve University, Cleveland, OH, January 2005.

- [108] PIRYATINSKA, A., TERDIK, G., WOYCZYNSKI, W., LOPARO, K., SCHER, M., AND ZLOTNIK, A. Statistical analysis of eeg-sleep patterns of neonates. To appear, 2006.
- [109] PRAVDICH-NEMINSKY, V. Ein versuch der registrierung der elektrischen gehirnerscheinungen (in german). *Zbl Physiol* 27 (1913), 951–960.
- [110] PRIESTLEY, M. *Spectral Analysis and Time Series*, vol. 1 of *Probability and Mathematical Statistics*. Academic Press, 1981.
- [111] PRIESTLEY, M. *Non-linear and Non-stationary Time Series Analysis*. Academic Press, London, 1988.
- [112] RANGARAJAN, G., AND DING, M., Eds. *Processes with Long-Range Correlations*, vol. 621 of *Lecture Notes in Physics*. Springer, 2003.
- [113] RIEKE, C., E. A. Discerning nonstationarity from nonlinearity in seizure-free and preseizure eeg recordings from epilepsy patients. *IEEE Transactions on Biomedical Engineering* 50, 5 (2003), 634–639.
- [114] RÖSCHKE, J., AND ALDENHOFF, J. The dimensionality of human’s electroencephalogram during sleep. *Biological Cybernetics* 64 (1991), 307–313.
- [115] RÖSCHKE, J., FELL, K., AND MANN, K. Non-linear dynamics of alpha and theta rhythm: correlation dimensions and lyapunov exponents from healthy subject’s spontaneous eeg. *International Journal of Psychophysiology* 26 (1997), 251–261.
- [116] ROSENSTEIN, M., COLLINS, J., AND DE LUCA, C. A practical method for calculating largest lyapunov exponents from small data sets. *Physica D* 65 (1993), 117–134.
- [117] SAMORODNITSKY, G., AND TAQQU, M. *Stable non-Gaussian random processes: stochastic models with infinite variance*. Chapman & Hall, New York, 1994.
- [118] SCHEPERS, H., VAN BEEK, J., AND BASSINGTHWAIGHTE, J. Four methods to estimate the fractal dimension from self-affine signals. *IEEE Engineering in Medicine and Biology* 11 (1992), 57–64.
- [119] SCHER, M., STEPPE, D., AND BANKS, D. Postnatal adaptation of brain function in full-term neonates as assessed by eeg sleep analyses. *Sleep* 18, 7 (1995), 531–535.

- [120] SCHER, M., STEPPE, D., DAHL, R., ASTHANA, S., AND GUTHRIE, R. Comparison of eeg sleep measures in healthy full-term and preterm infants at matched conceptional ages. *Sleep* 15, 5 (1992), 442–448.
- [121] SCHER, M., TURNBULL, J., LOPARO, K., AND JOHNSON, M. Automated state analyses: Proposed applications to neonatal neurointensive care. *Journal of Clinical Neurophysiology* 22 (2005), 256–270.
- [122] SCHER, M., WAISANEN, H., LOPARO, K., AND JOHNSON, M. Prediction of neonatal state and maturational change using dimensional analysis. *Journal of Clinical Neurophysiology* 22 (2005), 159–165.
- [123] SCHREIBER, T., AND SCHMITZ, A. Surrogate time series. *Physica* 142, 3 (2000), 346.
- [124] SHAO, M., AND NIKIAS, C. Signal processing with fractional lower order moments: stable processes and their applications. *Proceedings of the IEEE* 81, 7 (1993), 986–1010.
- [125] SHOEB, A., EDWARDS, H., CONNOLLY, J., BOURGEOIS, B., TREVES, T., AND GUTTAG, J. Patient-specific seizure onset detection. In *Proceedings of the 26th Annual International Conference of the IEEE EMBS* (Sept. 2004), pp. 419–422.
- [126] SINHA, A., RICHOUX, W., AND LOPARO, K. A system theoretic state description for temporal transitions in the electroencephalogram data of severe epileptic patients. *43rd Conference on Decision and Control* (2004).
- [127] SPROTT, J. Improved correlation dimension calculation. *International Journal of Bifurcation and Chaos* 11, 7 (2001), 1865–1880.
- [128] STOFFER, D., SCHER, M., RICHARDSON, G., DAY, N., AND COBLE, P. A walsh-fourier analysis of the effects of moderate maternal alcohol consumption on neonatal sleep-state cycling. *Journal of the American Statistical Association* 83, 404 (1988), 954–963.
- [129] SUBASI, A. Epileptic seizure detection using dynamic wavelet network. *Expert Systems with Applications* 29 (2005), 343–355.
- [130] SUGIMOTO, H., ISHII, N., IWATA, A., AND SUZUMURA, N. On the stationarity and normality of the electroencephalographic data during sleep stages. *Computer Programs in Biomedicine* 8 (1978), 224–234.
- [131] TEVEROVSKY, V., AND TAQQU, M. Testing for long-range dependence in the presence of shifting means or a slowly declining trend,

- using a variance-type estimator. *Journal of Time-Series Analysis* 18 (1997), 279–303.
- [132] THEILER, J. Spurious dimension from correlation algorithms applied to limited time-series data. *Physical Review A* 34, 3 (1986), 2427–2432.
 - [133] THEILER, J. Efficient algorithm for estimating the correlation dimension from a set of discrete points. *Physical Review A* 36, 9 (1987), 4456–4462.
 - [134] THEILER, J. Estimating fractal dimension. *Journal of the Optical Society of America* 7, 6 (1990), 1055–1073.
 - [135] THEILER, J., EUBANK, S., LONGTIN, A., GALDRIKIAN, B., AND FARMER, J. Testing for nonlinearity in time series: the method of surrogate data. *Physica D* 58 (1992), 77–94.
 - [136] THEILER, J., AND RAPP, P. Re-examination of evidence for low-dimensional, nonlinear structure in the human electroencephalogram. *Electroencephalography and clinical Neurophysiology* 98 (1996), 213–212.
 - [137] TSIHRINTZIS, G., AND NIKIAS, C. Fast estimation of the parameters of alpha-stable impulsive interference. *IEEE Transactions on Signal Processing* 44, 6 (1996), 1492–1503.
 - [138] TURNBULL, J., LOPARO, K., JOHNSON, M., AND SCHER, M. Automated detection of tracé alternant during sleep in healthy full-term neonates using discrete wavelet transform. *Clinical Neurophysiology* 112 (2001), 1893–1900.
 - [139] UCHAIKIN, V., AND SAENKO, V. Simulation of random vectors with isotropic fractional stable distributions and calculation of their probability density functions. *Journal of Mathematical Sciences* 112, 2 (2002).
 - [140] UCHAIKIN, V. V., AND ZOLOTAREV, V. M. *Chance and stability: Stable distributions and their applications*. VSP, Utrecht, 1999.
 - [141] WAISANEN, H., LOPARO, K., AND SCHER, M. Neuroplasticity in neonates, part 1: Computational approaches to dimensional analysis of eeg. Tech. rep., Case Western Reserve University, Cleveland, OH, 2005.
 - [142] WAISANEN, H., LOPARO, K., AND SCHER, M. Neuroplasticity in neonates, part ii: Application of correlation dimension to dimensional

analysis of neonatal eeg. Tech. rep., Case Western Reserve University, Cleveland, OH, 2005.

- [143] WIENER, N. *Cybernetics: Control and Communication in the Animal and the Machine*. MIT Press, Cambridge, 1961.
- [144] WILSON, S., SCHEUER, M., EMERSON, R., AND GABOR, A. Seizure detection: evaluation of the reveal algorithm. *Clinical Neurophysiology* 115 (2004), 2280–2291.
- [145] WOLF, A., SWIFT, B., SWINNEY, H., AND VASTANO, J. Determining lyapunov exponents from a time-series. *Physica D* 16 (1985), 285–317.
- [146] WOLFRAM. Mathematica 5.2. Wolfram Research, Inc., 2005. Champaign, Illinois.
- [147] YOSHIKAWA, Y., AND YOSHIFUMI, Y. Non-linear dynamics in heart rate variability in different generations. *Bulletin of Toyohashi Sozo College* 7 (2003), 63–78.
- [148] ZOLOTAREV, V. M. *One-dimensional Stable Distributions*, vol. 65. Trans. Math. Monographs, AMS, Providence, RI, 1986.